Proceedings
of the International Conference
on Computational Methods

(Vol. 10, 2023)

14th ICCM, 6th-10th August 2023

Editors: **G. R. Liu, Nguyen-Xuan Hung**

ICCM2023

Proceedings of the International Conference on
Computational Methods (Vol. 10, 2023)

14th ICCM, 6th-10th August 2023

at Ho Chi Minh City, Vietnam

Edited by

Guirong Liu
University of Cincinnati, USA

Nguyen-Xuan Hung
HUTECH University, Vietnam

Proceedings of the International Conference on Computational Methods, Vol.10, 2023

This volume contains full papers accepted by the 14th ICCM, 6th-10th August 2023.

First Edition, August 2023

International Standard Serial Number: ISSN 2374-3948 (online)

Papers in this Proceedings may be identically cited in the following manner: Author names, *Paper title, Proceedings at the 14th ICCM2023, 6th-10th August 2023, at Ho Chi Minh City, Vietnam*, Eds: Guirong Liu, Nguyen-Xuan Hung, ScienTech Publisher.

Note: The papers/data included in this volume are directly from the authors. The editors are not responsible of the inaccuracy, error, etc. Please discuss with the authors directly, if you have any questions.

Published by
Scientech Publisher LLC, USA
https://www.sci-en-tech.com/

WELCOME MESSAGE

Dear Colleagues and Friends,

We are delighted to extend our heartfelt welcome to you for the 14th International Conference on Computational Methods (ICCM2023), which is hosted by Ho Chi Minh City University of Technology (HUTECH University) in Ho Chi Minh City, Vietnam from August 6 to 10, 2023. The ICCM conference was founded by Professor Guirong Liu in 2004 and organized for the first time in Singapore. Since then, the conference has evolved into a series of global discussions conducted in numerous nations. Due to the Covid-19 pandemic, the ICCMs 2020, 2021, and 2022 were organized virtually. Fortunately, we have the chance to resume our regular organization this year to ensure the continuity of important top-notch research and to foster academic discussion within our community.

The ICCMs are now widely recognized as an international venue for academic and industrial researchers to share knowledge, gain new perspectives, and push the limits of computational methodologies. The conference's subjects reflect multidisciplinary studies on contemporary approaches as well as the application of artificial intelligence in a variety of technological fields. Throughout the conference days, international specialists will present captivating keynote addresses, panel discussions, and technical sessions.

Authors at the conference have presented extremely meaningful discoveries covering both theoretical and practical aspects, highlighting emerging trends. All submissions have undergone a rigorous peer-review process, and accepted papers will be presented orally during the event, while exceptional contributions will be considered for publication in a special issue of prestigious journals.

In addition to educational presentations, this conference provides an excellent opportunity for crucial conversations, networking, and connections. The sharing of perspectives and ideas is essential for excellent future research and ground-breaking developments. We encourage you to engage with other participants, cultivate profound friendships, and create incredible possibilities for future collaborations.

We would like to express our sincere thanks to each member of the Organizing Committee, the International Scientific Committee, and all other supporters who have worked tirelessly to make this conference a reality. We would also like to express our deepest gratitude to the international reviewers for their diligent work in examining submitted manuscripts and abstracts.

Furthermore, we want to express our acknowledgment of your valuable contributions to the ICCM2023 conference. We are delighted with your participation in this event and eagerly looking forward to your continued involvement in future ICCM conferences.

Finally, we have confidence that this conference will be a promising platform for learning, sharing, inspiration, and growth in the limitless potential of computational methods.

I appreciate your time, and now let's start the 14th International Conference on Computational Methods!

Professor Hung Nguyen-Xuan   **Professor Guirong Liu**
Conference Chairman   Honorary Conference Chairman
CIRTECH Institute, HUTECH University, Vietnam   University of Cincinnati, USA

ORGANIZATION COMMITTEES

Conference Chairman
Nguyen-Xuan Hung, HUTECH University, Vietnam

Honorary Chairman
Guirong Liu, University of Cincinnati, United States

International Co-Chairs
Magd Abdel-Wahab, Ghent University, Belgium
Stephane P.A. Bordas, Luxembourg University, Luxembourg
Vu-Hieu Nguyen, Paris-East Créteil University, France
Jaehong Lee, Sejong University, South Korea
Hua Li, Nanyang Technological University, Singapore
Tuan Ngo, University of Melbourne, Australia
Hiroshi Okada, Tokyo University of Science, Japan
Timon Rabczuk, Bauhaus University Weimar, Germany
Dia Zeidan, German Jordanian University, West Asia

Local Co-Chairmen
Canh Van Le, International University-VNU-HCMC
Hung Quoc Nguyen, Vietnam-German University, Vietnam
Kien Trung Nguyen, HUTECH University, Vietnam

Hutech Secretary General
Viet Van Pham, HUTECH University
Yen Hai Dinh, HUTECH University

Local Organizing Committee
Anh Ngoc Lai, Binh Anh Tran, Bang Quang Tao, Cuong Huu Ngo, Chien Hoang Thai, Hai Van Luong, Hieu Van Nguyen, Long Minh Nguyen, Nhon Thanh Nguyen, Lieu Bich Nguyen, Phuc Hong Pham, Phuc Van Phung, Phuong Tran, Phuoc Trong Nguyen, Thanh Trung Bui, Tinh Quoc Bui, Nam Van Hoang, Nghi Van Vu, Son Hoai Nguyen, Thanh Dinh Chau, Trung Thoi Nguyen, Truong Van Vu, Viet Duc La

International Scientific Advisory Committee
(sorting by last name)

Addessi Daniela (Italy)	Chen Songying (China)	Gan Yixiang (Australia)
Armas Rafael Montenegro (Spain)	Chen Weiqiu (China)	Gao Wei (Australia)
Atroshchenko Elena (Australia)	Chen Zhen (USA)	Gao Xiao-Wei (China)
Birken Philipp (Sweden)	Cheng Yuan (Singapore)	Gravenkamp Hauke (Germany)
Bui Ha (Australia)	Cheng Yumin (China)	Gu Yuantong (Australia)
Bui Tinh Quoc (Japan)	Cui Fangsen (Singapore)	Gupta Murli (USA)
Bui Trung Thanh (Vietnam)	Das Raj (Australia)	Herrera Ismael (Mexico)
Chen Bin (China)	Dias-da-Costa Daniel (Australia)	Hou Shujuan (China)
Chen Jeng-Tzong (Taiwan)	Dong Leiting (China)	Huang Yu (China)
Chen Jiye (UK)	Dulikravich George (USA)	Jabareen Mahmood (Israel)
Chen Lei (Singapore)	Fan Chia-Ming (Taiwan)	Jacobs Gustaaf (USA)
Chen Shaohua (China)	Fu Zhuojia (China)	Jiang Chao (China)
Chen Shunhua (China)	Gan Buntara S. (Japan)	Jin Feng (China)

Kanayama Hiroshi (Japan)
Kang Zhan (China)
Kougioumtzoglou Ioannis (USA)
Le Van Canh (Vietnam)
Lee Chin-Long (New Zealand)
Lee Heow Pueh (Singapore)
Lee Ik-Jin (South Korea)
Lenci Stefano (Italy)
Li Bing (China)
Li Chenfeng (UK)
Li Eric (UK)
Li Qing (Australia)
Li Wei (China)
Li Weihua (Australia)
Li Yue-Ming (China)
Liu Moubin (China)
Liu Yan (China)
Liu Yinghua (China)
Lombardi Domenico (UK)
Mammoliti Francesco (Italy)
Miller Karol (Australia)
Natarajan Sundararajan (India)
Nguyen Anh Dong (Vietnam)
Nguyen Duc Dinh (Vietnam)
Nguyen Giang (Australia)
Nguyen The Hung (Vietnam)
Ni Bingyu (China)
Nithiarasu Parumal (UK)
Niu Yang-Yao (Taiwan)
Nogueira Xesús (Spain)
Noto Francesco (Italy)
Ogino Masao (Japan)
Orlande Helcio (Brazil)

Onishi Yuki (Japan)
Papadrakakis Manolis (Greece)
Peng Qing (Saudi Arabia)
Picu Catalin (USA)
Quek Jerry Sinsin (Singapore)
Rabczuk Timon (Germany)
Ramaiah Pattabhi (India)
Reali Alessandro (Italy)
Reddy Daya (South Africa)
Sadowski Tomasz (Poland)
Saitoh Takahiro (Japan)
Shen Lian (USA)
Shen Luming (Australia)
Shioya Ryuji (Japan)
Son Gihun (South Korea)
Song Chongmin (Australia)
Stefanou George (Greece)
Su Cheng (China)
Sun Waiching Steve (USA)
Tadano Yuichi (Japan)
Tan B.C. Vincent (Singapore)
Tian Zhao-Feng (Australia)
Trung Thoi Nguyen (Vietnam)
Tsubota Ken-Ichi (Japan)
Wan Decheng (China)
Wang Cheng (China)
Wang Dongdong (China)
Wang Hu (China)
Wang Jie (China)
Wang Lifeng (China)
Wang Lihua (China)
Wang Wenquan (China)
Wang Yue-Sheng (China)

Wu Bin (Italy)
Wu Wei (Austria)
Xiang Zhihai (China)
Xiao Feng (Japan)
Xiao Jinyou (China)
Xie Mike (Australia)
Xu Chao (NPU, China)
Xu Chao (ZJU, China)
Xu Xiangguo George (Singapore)
Yang Chun Charles (Singapore)
Yang Judy (Taiwan)
Yang Qingcheng (China)
Yang Qingsheng (China)
Yang Zhenjun (China)
Yao Jianyao (China)
Ye Hongling (China)
Ye Qi (China)
Yeo Jingjie (USA)
Yu Chengxiang Rena (Spain)
Zhang Chuanzeng (Germany)
Zhang Guiyong (China)
Zhang Jian (China)
Zhang Lihai (Australia)
Zhang Xiong (China)
Zhang Yixia Sarah (Australia)
Zhang Zhao (China)
Zhang Zhe (China)
Zhao Liguo (UK)
Zheng Hui (China)
Zhong Zheng (China)
Zhou Anna (Australia)
Zhou Kun (Singapore)
Zhuang Zhuo (China)

TABLE OF CONTENTS

Low-Order Triangular Elements for Thick Shell Analysis

†J. Petrolito and D. Ionescu

Department of Engineering, La Trobe University, Australia.

†Corresponding author: j.petrolito@latrobe.edu.au

Abstract

This paper discusses the development of four, low-order triangular elements for the analysis of thick and thin shell structures. The elements combine triangular plane stress elements with triangular thick plate bending elements. The theoretical formulations used for the elements are the free formulation, the deviatoric strain formulation and the template formulation. Examples are given in the paper to demonstrate the accuracy that is achieved with the proposed elements.

Keywords: Shell analysis, thick shells, shear deformation, finite elements.

Introduction

Shells are among the most demanding structures to analyse, because of their complicated geometry and behaviour. Due to their importance in structural engineering, reliable and efficient methods are needed for their analysis and design, and there is a vast literature in this area [1–4]. Analytical solutions are only available for simple geometries and loadings [1], and the solution of practical shell problems normally requires a numerical approach. The finite element method is by far the preferred method, due to its generality and ability to deal with complex geometries [5]. The literature on finite elements for shells is also vast, and many shell elements have been developed [4–7]. Research in this area is continuing, with new elements constantly being developed [8–14].

The two basic load-carrying modes that need to be modelled in a shell are the membrane (plane stress) mode and the transverse bending (plate bending) mode. Two fundamental approaches to the development of shell finite elements are possible [4, 5]. One option is to use curved finite elements to try to closely model the shell geometry. In this approach, the two load-carrying modes are also coupled in the formulation. However, the development of these elements is generally complex, and numerical integration is required to derive the element matrices. The other option is to use flat shell elements where the two load-carrying modes are treated separately, and the load-carrying coupling is achieved by approximating the curved geometry using flat elements that meet at different orientations [5].

Flat shell elements are generally simpler to formulate, because elements for plane stress and plate bending can be combined to form a shell element. However, these elements also need care in their development, due to such issues as element locking and spurious mechanisms, and many techniques have been used to alleviate these problems [4, 5].

The current paper discusses some new low-order thick shell elements. The plate-bending elements are based on Mindlin's theory, which includes shear deformation, and their formulation avoids shear locking by design. These elements are combined with plane stress elements, which

include in-plane rotations as element variables, to produce several flat shell elements. The applications and advantages of the elements are presented and compared. Examples are given in the paper to demonstrate the accuracy that is achieved with the proposed elements.

Element Formulation

All elements considered in this paper are three-node flat triangular elements with six degrees of freedom per node, making a total of eighteen degrees of freedom for each element. The elements are combinations of plane stress and plate bending elements, each with nine degrees of freedom.

Plane Stress Elements

Fig. 1 shows the geometry and degrees of freedom for the triangular plane stress elements. The elements have three degrees of freedom at each node, making a total of nine degrees of freedom. The element variables are the displacements in the x and y directions, u and v, and the drilling freedom, θ_z.
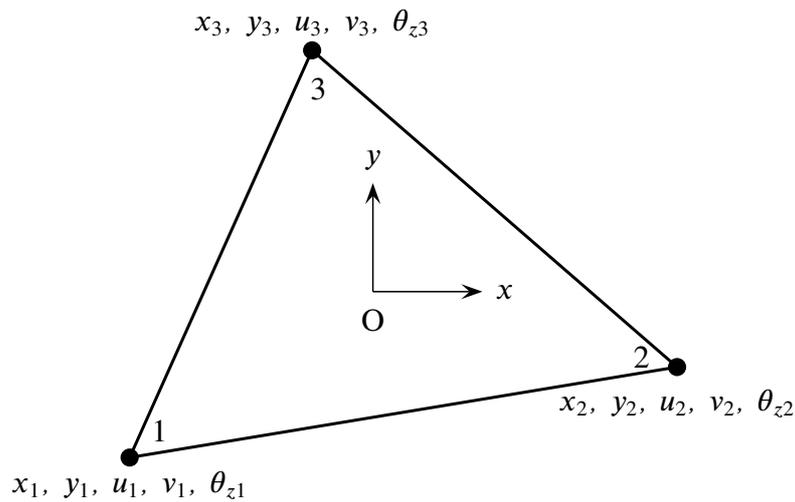
Figure 1. Triangular plane stress element (origin is at centre of element).

The simplest possible element that could be used for the plane stress component of the shell element is the constant strain triangle [5]. This element ignores the drilling freedoms, and hence special treatment may be required to avoid a potential global singularity for the shell element [5]. In addition, this element is not very accurate. Hence, it is not used in this paper, and instead we consider two alternative elements, namely the free formulation plane stress element [15] (denoted TPSF9 here) and the optimum plane stress element (denoted OPT in the reference and TPSO9 here) that is based on the template formulation [16]. Both elements include the drilling freedoms. Complete details of the two elements are given in [15, 16].

It is worth noting that the template formulation is a very flexible formulation that leads to elements with variable parameters. Felippa [16] has shown that many well-known plane stress elements are examples of the general formulation for particular values of the parameters. However, the free formulation element is not an example of the general template.

Plate Bending Elements

Fig. 2 shows the geometry and degrees of freedom for the triangular plate bending elements. The elements have three degrees of freedom at each node, making a total of nine degrees of freedom. The element variables are the transverse displacement, w, and the rotations about the x and y axes, θ_x and θ_y.

$x_3,\ y_3,\ w_3,\ \theta_{x3},\ \theta_{y3}$

$x_2,\ y_2,\ w_2,\ \theta_{x2},\ \theta_{y2}$

$x_1,\ y_1,\ w_1,\ \theta_{x1},\ \theta_{y1}$
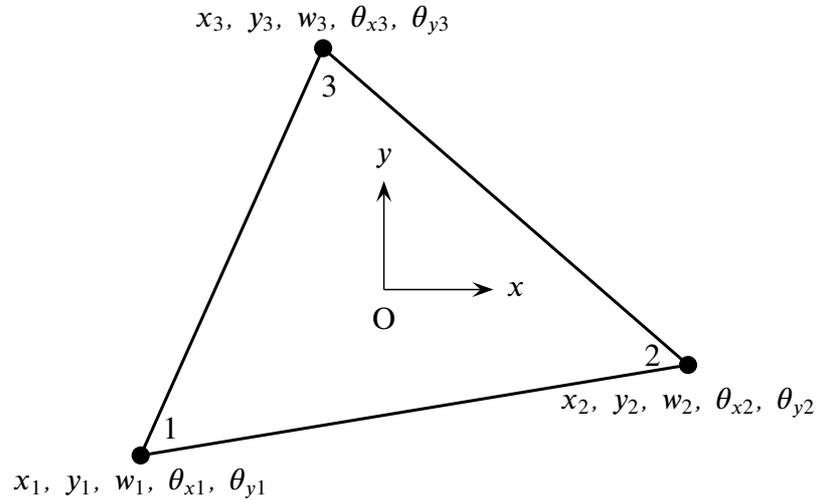
Figure 2. Triangular plate bending element (origin is at centre of element).

The plate bending elements use Mindlin's theory [17] as their theoretical foundation, and they can be used to analyse both thick and thin plates. In particular, their design avoids the shear locking problem automatically in the limit as the plate thickness goes to zero. In this paper, we use elements TF9 and TD9 from [18] for the plate bending components. These elements are derived from the free formulation and the deviatoric strain formulation, respectively. Complete details of the two elements are given in [18].

Shell Elements

The triangular shell elements that are discussed in this paper are combinations of the different plate bending and plane stress elements. They are denoted by the nomenclature T(B)-(S)18, where T denotes a triangular element, B refers to the plate bending element, S refers to the plane stress element and all elements have eighteen degrees of freedom. Hence, the elements that are discussed are denoted as follows.

1. TF-F18: Free formulation plate bending (TF9) and plate stress elements (TPSF9).

2. TF-O18: Free formulation plate bending element (TF9) and optimum plane stress element (TPSO9).

3. TD-F18: Deviatoric strain formulation plate bending element (TD9) and free formulation plane stress element (TPSF9).

4. TD-O18: Deviatoric strain formulation plate bending element (TD9) and optimum plane stress element (TPSO9).

Examples

We consider three examples to illustrate the behaviour of the proposed elements. Shear deformation is included in the plate bending elements, and the shear correction factor was taken as $k = 5/6$. The distributed load on the element was modelled using a consistent load vector [5].

Fig. 3 shows a cantilever beam that is made from a z-section with a uniform thickness of $0.1\,\text{m}$ [19]. The beam is fully clamped at $x = 0\,\text{m}$ and subjected to a torque of $1.2\,\text{MNm}$ at $x = 10\,\text{m}$, which is modelled by a pair of vertical loads $F = 0.6\,\text{MN}$ as shown. The material properties are $E = 210\,\text{GPa}$ and $v = 0.3$.
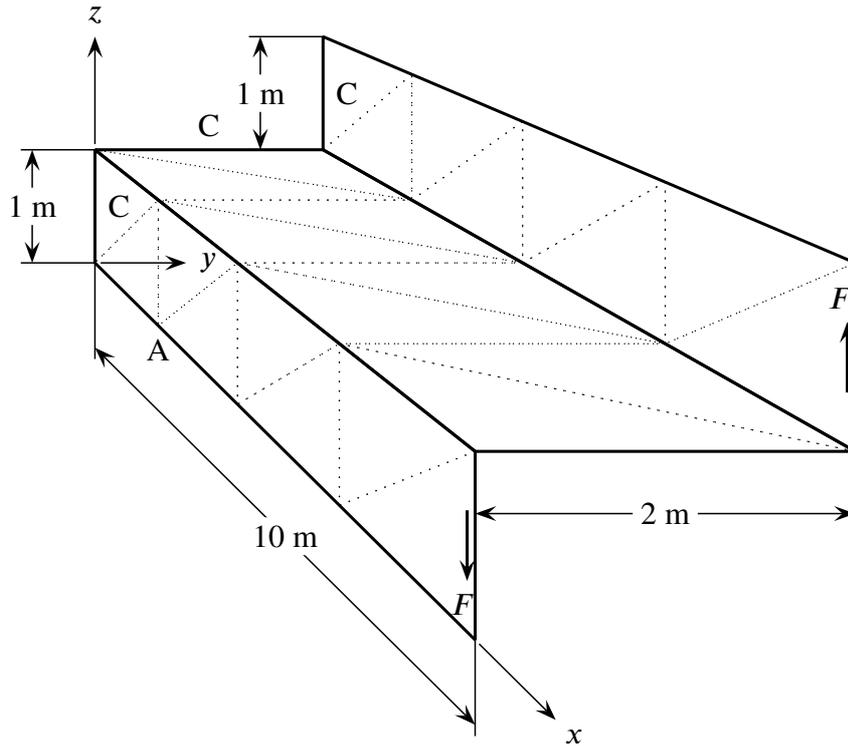
Figure 3. Z-section cantilever subjected to end torque with typical mesh (C = clamped).

The target result for this example is the midplane axial stress, σ_x, at point A, which is located at $x = 2.5\,\text{m}$. The value quoted in [19] is $\sigma_x = -108\,\text{MPa}$, while $\sigma_x = -111.2\,\text{MPa}$ has been suggested as a more accurate value [20]. The beam was analysed using a uniform mesh of $16N \times 2N \times 2N$ elements, in the x, y and z directions, respectively, and the results are given in Table 1. All elements converge rapidly for this problem, with only slight differences between the results for the different elements. A value of $N \geq 8$ is an acceptable mesh for the problem.

Table 1. Compressive stress at point A in MPa for z-section cantilever.

N	TF-F18	TF-O18	TD-F18	TD-O18
2	102.3	106.2	102.8	106.7
4	106.7	108.0	106.9	108.2
8	108.5	109.0	108.6	109.1
16	109.4	109.6	109.5	109.7
32	109.9	109.9	110.0	110.0
64	110.2	110.2	110.3	110.3

Fig. 4 shows a quarter section of a cylindrical shell roof. The radius of the cylinder is $R = 25$ m and the shell has a uniform thickness of 0.25 m [19]. The roof is supported with rigid diaphragms that restrain the displacements in the x and y directions. It is subjected to a uniform gravity load in the $-z$ direction of magnitude 90 N/m^2. The material properties are $E = 432$ MPa and $\nu = 0$.

Figure 4. Cylindrical shell roof subjected to gravity load with typical mesh (S = symmetry condition; RD = rigid diaphragm condition).

The target result for this example is the displacement in the z direction at point A. The value quoted in [19] is $w = -0.3024$ m, while $w = -0.3020$ m has been suggested as a more accurate value [20]. The roof was analysed using a uniform mesh of $2N \times 2N$ elements, and the results are given in Table 2.

Table 2. Displacement at point A in m for cylindrical shell roof.

N	TF-F18	TF-O18	TD-F18	TD-O18
2	0.3215	0.3389	0.3826	0.4027
4	0.2926	0.3025	0.3025	0.3138
8	0.2983	0.3016	0.3011	0.3046
16	0.3004	0.3013	0.3015	0.3024
32	0.3009	0.3011	0.3014	0.3015
64	0.3010	0.3011	0.3012	0.3010

There are two sources of error for this problem. The first error is due to the approximation of the curved shell by a series of flat elements, while the second error is the usual approximation error due to the use of finite elements. Hence, sufficient elements are required to accurately model the curved geometry independently of the approximation error. Once again, all elements provide accurate results with increasing N, with $N \geq 16$ being an acceptable mesh for the problem.

Fig. 5 shows a hyperbolic paraboloid shell. The surface of the shell is given by $z = -x^2 + y^2$, for $-1/2 \leq x \leq 1/2$ and $-1/2 \leq y \leq 1/2$, and the shell has a uniform thickness, t. The shell is clamped on the side $x = -1/2$ and the other sides are free to move. It is subjected to a uniform gravity load in the $-z$ direction of magnitude $8000\,\text{N/m}^3$. The material properties are $E = 200\,\text{GPa}$ and $v = 0.3$.
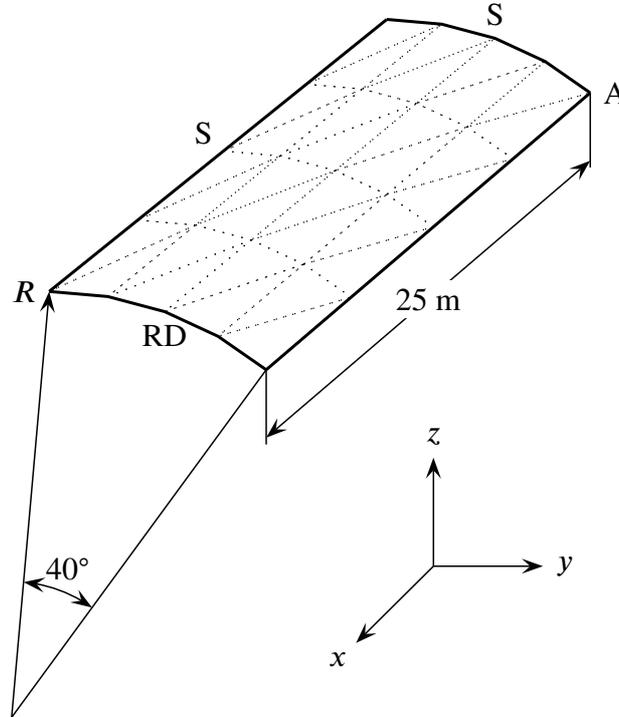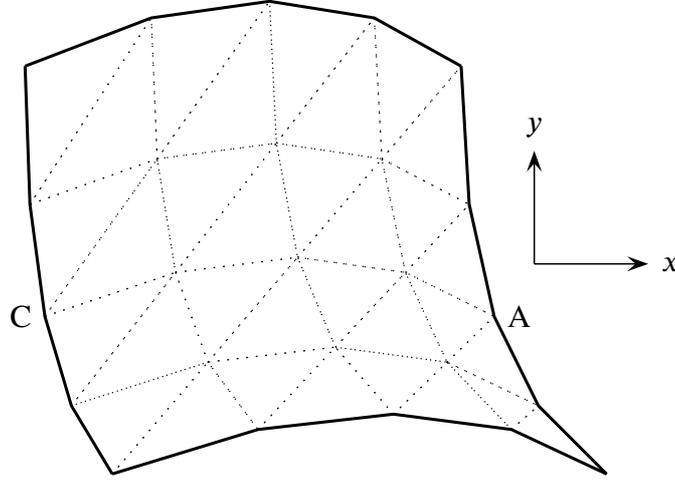

Figure 5. Hyperbolic paraboloid shell subjected to gravity load with typical mesh (C = clamped).

The target results for this example are the displacement in the z direction at point A, with coordinates $(1/2, 0)$, and the strain energy for different values of the shell thickness, t. The example was originally discussed in [4], and the reference values in [4, 20] are used for comparison. The shell was analysed using a uniform mesh of $2N \times 2N$ elements, and the results are given in Tables 3 to 5.

This is a challenging example for shell elements, particularly as the shell becomes thinner. Once again, sufficient elements are required to accurately model the curved geometry. The accuracy achieved by the elements for a given value of N decreases with decreasing thickness, and finer meshes are required for acceptable results. While $N \geq 64$ is an acceptable mesh for $t = 1/100$, $N \geq 128$ is required for accurate results for $t = 1/1000$ and $t = 1/10000$. In addition, elements TF-O18 and TD-O18 provide more accurate results as the thickness decreases. These results confirm the challenging nature of this example with decreasing thickness.

Table 3. Displacement at point A, $10^5 w$, in m and strain energy, $10^3 U$, in Nm for hyperbolic paraboloid shell ($t = 1/100$; reference values for $10^5 w$: 9.336 [4], 9.350 [20]; reference values for $10^3 U$: 1.679 [4], 1.682 [20]).

	w				U			
N	TF-F18	TF-O18	TD-F18	TD-O18	TF-F18	TF-O18	TD-F18	TD-O18
8	6.105	7.381	6.497	7.860	1.094	1.320	1.165	1.407
16	8.516	8.714	8.750	8.952	1.522	1.559	1.565	1.603
32	9.061	9.089	9.143	9.171	1.625	1.631	1.640	1.646
64	9.236	9.241	9.264	9.269	1.659	1.660	1.665	1.666
128	9.301	9.302	9.312	9.313	1.672	1.672	1.674	1.675
256	9.326	9.326	9.331	9.331	1.677	1.677	1.678	1.678

Table 4. Displacement at point A, $10^3 w$, in m and strain energy, $100U$, in Nm for hyperbolic paraboloid shell ($t = 1/1000$; reference values for $10^3 w$: 6.394 [4], 6.400 [20]; reference values for $100U$: 1.101 [4], 1.103 [20]).

			w				U	
N	TF-F18	TF-O18	TD-F18	TD-O18	TF-F18	TF-O18	TD-F18	TD-O18
8	0.214	0.471	0.216	0.476	0.038	0.082	0.038	0.083
16	1.777	3.135	1.825	3.240	0.306	0.539	0.315	0.558
32	5.213	5.807	5.320	5.929	0.895	0.998	0.915	1.020
64	6.208	6.266	6.254	6.312	1.068	1.078	1.076	1.086
128	6.348	6.353	6.363	6.369	1.093	1.094	1.096	1.097
256	6.382	6.382	6.386	6.387	1.099	1.099	1.100	1.100

Table 5. Displacement at point A, $10w$, in m and strain energy, $100U$, in Nm for hyperbolic paraboloid shell ($t = 1/10000$; reference values for $10w$: 5.299 [4], 5.314 [20]; reference values for $100U$: 8.987 [4], 9.015 [20]).

			w				U	
N	TF-F18	TF-O18	TD-F18	TD-O18	TF-F18	TF-O18	TD-F18	TD-O18
8	0.002	0.005	0.002	0.005	0.004	0.009	0.004	0.009
16	0.026	0.064	0.026	0.064	0.045	0.109	0.045	0.109
32	0.318	0.756	0.320	0.761	0.539	1.278	0.542	1.287
64	2.453	3.723	2.477	3.768	4.152	6.304	4.195	6.385
128	4.881	5.125	4.905	5.151	8.272	8.687	8.315	8.733
256	5.258	5.277	5.267	5.286	8.916	8.948	8.932	8.964

Conclusions

This paper has discussed the development of four, low-order triangular elements for the analysis of thick and thin shell structures. Two formulations are considered for both the plane stress and plate bending components for the shell elements. The element interpolations for the plate bending component are the solutions of the governing equations of Mindlin's theory, and locking cannot occur in the thin shell limit with these elements. All formulations lead to reliable and accurate elements. Three examples were considered in the paper to demonstrate the accuracy that is achieved with the proposed elements.

References

[1] Timoshenko, S. and Woinowsky-Krieger, S. (1959) *Theory of Plates and Shells*, second edn., McGraw-Hill, New York.
[2] Libai, A. and Simmonds, J. (1998) *The Nonlinear Theory of Elastic Shells*, second edn., Cambridge University Press, Cambridge.
[3] Reddy, J. (2007) *Theory and Analysis of Elastic Plates and Shells*, second edn., Taylor and Francis, Philadelphia.
[4] Chappelle, D. and Bathe, K. (2011) *The Finite Element Analysis of Shells - Fundamentals*, second edn., Springer-Verlag, Berlin.

[5] Zienkiewicz, O., Taylor, R. and Fox, D. (2014) *The Finite Element Method for Solid and Structural Mechanics*, vol. 2, seventh edn., Butterworth-Heinemann, Oxford.

[6] Yang, H., Saigal, S., Masud, A. and Kapania, R. (2000) A survey of recent shell finite elements, *International Journal for Numerical Methods in Engineering*, **47**, 101–127.

[7] Mackerle, J. (2002) Finite and boundary-element linear and nonlinear analyses of shells and shell-like structures - a bibliography (1999-2001), *Finite Elements in Analysis and Design*, **38**, 765–782.

[8] Kiendl, J., Marino, E. and De Lorenzis, L. (2017) Isogeometric collocation for the Reissner-Mindlin shell problem, *Computer Methods in Applied Mechanics and Engineering*, **325**, 645–665.

[9] Ko, Y., Lee, Y., Lee, P. and Bathe, K. (2017) Performance of the MITC3+ and MITC4+ shell elements in widely-used benchmark problems, *Computers and Structures*, **193**, 187–206.

[10] Ludwig, T., Huehne, C. and De Lorenzis, L. (2019) Rotation-free Bernstein-Bezier elements for thin plates and shells - development and validation, *Computer Methods in Applied Mechanics and Engineering*, **348**, 500–534.

[11] Lavrencic, M. and Brank, B. (2021) Hybrid-mixed low-order finite elements for geometrically exact shell models: Overview and comparison, *Archives of Computational Methods in Engineering*, **28**, 3917–3951.

[12] Zou, Z., Hughes, T., Scott, M., Sauer, R. and Savitha, E. (2021) Galerkin formulations of isogeometric shell analysis: Alleviating locking with Greville quadratures and higher-order elements, *Computer Methods in Applied Mechanics and Engineering*, **380**, 10.1016/j.cma.2021.113757.

[13] Krysl, P. (2022) Robust flat-facet triangular shell finite element, *International Journal for Numerical Methods in Engineering*, **123**, 2399–2423.

[14] Krysl, P. (2022) Explicit dynamics of shells with a flat-facet triangular finite element, *International Journal for Numerical Methods in Engineering*, **123**, 4545–4561.

[15] Bergan, P. and Felippa, C. (1985) A triangular membrane element with rotational degrees of freedom, *Computer Methods in Applied Mechanics and Engineering*, **50**, 25–69.

[16] Felippa, C. (2003) A study of optimal membrane triangles with drilling freedoms, *Computer Methods in Applied Mechanics and Engineering*, **192**, 2125—2168.

[17] Mindlin, R. (1951) Influence of rotatory inertia and shear on flexural motion of isotropic plates, *Journal of Applied Mechanics*, **23**, 31–38.

[18] Petrolito, J. and Ionescu, D. (2022) Alternative formulations for finite elements for thick plate analysis, in G. Liu and N. Hung (editors), *Proceedings, 13th International Conference on Computational Methods*, 1–12, ScienTech Publisher.

[19] NAFEMS (1990) *The Standard NAFEMS Benchmarks*, NAFEMS, Glasgow.

[20] Krysl, P. and Chen, J. (2023) Benchmarking computational shell models, *Archives of Computational Methods in Engineering*, **30**, 301—315.

Filtering spurious eigenmodes in electromagnetic cavities discretized by energy-orthogonal finite elements

Francisco J. Brito

Departamento de Ingeniería Industrial, Universidad de La Laguna, Spain

Presenting and corresponding author: fjbrito@ull.edu.es

Keywords: Energy-Orthogonal Formulation, Electromagnetic Cavities, Spurious Eigenmodes

1. Introduction

1.1 Governing equation

An electromagnetic cavity as discussed in this contribution will be a source-free domain Ω enclosed by a perfect electric wall Γ and filled with a loss-free homogeneous isotropic linear dielectric medium. The time-harmonic electromagnetic fields in this domain will be solutions to the Maxwell's equations (the complex phasor notation is adopted) [1],

$$
\begin{aligned}
\nabla \cdot \varepsilon \mathbf{E} = 0 \quad (a), &\qquad \nabla \cdot \mu \mathbf{H} = 0 \quad (b) \\
\nabla \times \mathbf{E} = i\omega\mu\mathbf{H} \quad (c), &\quad \nabla \times \mathbf{H} = -i\omega\varepsilon\mathbf{E} \quad (d)
\end{aligned} \; ; \quad \mathbf{E} = \tilde{\mathbf{E}}\exp(-i\omega t) \quad \mathbf{H} = \tilde{\mathbf{H}}\exp(-i\omega t) \quad (1)
$$

where: $\tilde{\mathbf{E}}$ ($\tilde{\mathbf{H}}$) complex amplitude of electric (magnetic) field intensity $\mathbf{E}$ ($\mathbf{H}$); ε (μ), permittivity (permeability) of the medium; $\omega = 2\pi/T$, circular frequency; T, period; t, time;

The essential boundary conditions on the perfect electric wall will be,

$$
\mathbf{n} \times \mathbf{E} = \mathbf{0} \quad (a), \quad \mathbf{n} \cdot \mathbf{H} = 0 \quad (b) \tag{2}
$$

where $\mathbf{n}$ is the unit vector normal to the surface Γ.

We will deal with time-harmonic electromagnetic fields directly in terms of either the electric field $\mathbf{E}$ or the magnetic field $\mathbf{H}$. For this, it is necessary to derive from Maxwell's equations, which involve both electric and magnetic fields, the governing differential equation involving only either field. Then, from Eq. (1c) and Eq. (1d), is produced the vector wave equation,

$$
q^{-1}\nabla \times \nabla \times \mathbf{U} - \omega^2 p\mathbf{U} = \mathbf{0} \tag{3}
$$

where: $\mathbf{U}$, either $\mathbf{E}$ or $\mathbf{H}$; if $\mathbf{U} = \mathbf{E}$, then $q = \mu$ and $p = \varepsilon$; if $\mathbf{U} = \mathbf{H}$, then $q = \varepsilon$ and $p = \mu$.

If the electric field $\mathbf{E}$ is solved by Eq. (3), then the magnetic field $\mathbf{H}$ will be induced by the Maxwell's equation Eq. (1c). Alternatively, if the magnetic field $\mathbf{H}$ is solved by Eq. (3), then the electric field $\mathbf{E}$ will be induced by the Maxwell's equation Eq. (1d). The Maxwell's equations Eq. (1a) and Eq. (1b) are implicit in Eq. (3) whenever $\omega \neq 0$.

In this contribution the energy of the induced field will be an important quantity, which is computed by the equation,

$$
E = \tfrac{1}{2}q^{-1}\omega^{-2}\int_{\Omega} \mathrm{Im}(\nabla \times \mathbf{U}) \cdot \mathrm{Im}(\nabla \times \mathbf{U})dV \tag{4}
$$

1.2 Energy-orthogonal finite element formulation

The functional equivalent to the source-free vector wave equation Eq. (3) and suitable natural boundary conditions will be [2][3],

$$F(\mathbf{U}) = \int_\Omega \left[q^{-1}(\nabla \times \mathbf{U}) \cdot (\nabla \times \mathbf{U}) - \omega^2 p\mathbf{U} \cdot \mathbf{U} \right] d\Omega \tag{5}$$

From Eq. (5), by considering the standard discretization by the finite element method, it will be produced the following matrix equation,

$$q^{-1}\mathbf{K}\mathbf{u} - \omega^2 p\mathbf{M}\mathbf{u} = \mathbf{0}; \quad \mathbf{u} = \tilde{\mathbf{u}}\exp(-i\omega t) \tag{6}$$

where: $\mathbf{K}$ ($\mathbf{M}$), global stiffness (mass) matrix (the structural notation is adopted); $\tilde{\mathbf{u}}$, column matrix containing the complex amplitude of the nodal values of $\mathbf{U}$ (either $\mathbf{E}$ or $\mathbf{H}$).

At element level,

$$\mathbf{K}^e = \int_{\Omega e} \mathbf{B}^T \mathbf{B} dV \quad (a), \quad \nabla \times \mathbf{U} = \mathbf{B}\mathbf{u}^e \quad (b) \tag{7}$$

If the matrix $\mathbf{B}$ is partitioned into mean and deviatoric components,

$$\mathbf{B} = \bar{\mathbf{B}} + \mathbf{B}_d; \quad \bar{\mathbf{B}}V^e = \int_{\Omega e} \mathbf{B} dV, \quad \mathbf{B}_d = \mathbf{B} - \bar{\mathbf{B}} \tag{8}$$

the matrix Eq. (7a) will be decomposed as addition of basic and higher order components,

$$\mathbf{K}^e = \mathbf{K}_b^e + \mathbf{K}_h^e; \quad \mathbf{K}_b^e = \bar{\mathbf{B}}^T \bar{\mathbf{B}} V^e, \quad \mathbf{K}_h^e = \int_{\Omega e} \mathbf{B}_d^T \mathbf{B}_d dV \tag{9}$$

In this case the element stiffness matrix is formulated in energy-orthogonal form [4]. The decomposition in Eq. (9) holds for the complete model,

$$\mathbf{K} = \mathbf{K}_b + \mathbf{K}_h; \quad \mathbf{K}_b = \sum_e \mathbf{K}_b^e, \quad \mathbf{K}_h = \sum_e \mathbf{K}_h^e \tag{10}$$

By considering Eq. (4), Eq. (7) and Eq. (10), the energy of the induced field at the discretized domain will be computed, and also will be its basic and higher order components,

$$E = \tfrac{1}{2} q^{-1}\omega^{-2} \text{Im}[\mathbf{u}^T]\mathbf{K}\,\text{Im}[\mathbf{u}]; \quad E = E_b + E_h, \quad E_{b(h)} = \tfrac{1}{2} q^{-1}\omega^{-2} \text{Im}[\mathbf{u}^T]\mathbf{K}_{b(h)}\,\text{Im}[\mathbf{u}] \tag{11}$$

For a standing-wave field the column matrix containing the amplitude of nodal values of $\mathbf{U}$ will be a real-values vector in Eq. (6). Then, from Eq. (11), the period-averaged energy of the induced field at the discretized domain and its basic and higher order components will be,

$$\bar{E} = \tfrac{1}{4} q^{-1}\omega^{-2}\tilde{\mathbf{u}}^T\mathbf{K}\tilde{\mathbf{u}}; \quad \bar{E} = \bar{E}_b + \bar{E}_h, \quad \bar{E}_{b(h)} = \tfrac{1}{4} q^{-1}\omega^{-2}\tilde{\mathbf{u}}^T\mathbf{K}_{b(h)}\tilde{\mathbf{u}} \tag{12}$$

1.3 Scope of research

An electromagnetic cavity can support modes of free oscillation at an infinite number of discrete frequencies. These standing-wave fields and natural frequencies are solutions of the eigenproblem associated with the double-curl operator Eq. (3). As it is well known, if the above operator is discretized with nodal finite elements then the eigenspectrum computed by Eq. (6) will be severely polluted with spurious eigenmodes, the spurious eigenvalues having the same order of magnitude as the physical eigenvalues and not easily distinguished from them. Enlightening discussions about the origin of the spurious eigenmodes, including references to relevant papers, can be found in some books, see references [5][6]. Identifying the spurious eigenmodes by the energy-orthogonal formulation is the main objective of this contribution which extends the author's previous research about this topic [7].

2. Dispersion analysis

A physical cavity eigenmode can be represented by the superposition of plane harmonic waves travelling in several directions [8]. For these fundamental wave solutions, by a dispersion

analysis, the behavior of the energy of the induced field will be researched in an unbounded domain discretized by a regular mesh of finite elements. Two different isoparametric finite elements with consistent mass matrix are considered: the hexahedron with twenty nodes and brick geometry HE20, Fig. 1, and the pentahedron with fifteen nodes and right prismatic isosceles geometry PE15, Fig. 2 [9]. The nodal lattice formed by the finite element assemblage has four and five nodes per unit cell, respectively. Different meshes with the same element volume can be produced by selecting the aspect ratio parameter, $0 < \gamma \leq 1$; the distortion-Z parameter, $\alpha > 0$; and the skew angle, $0 \leq \beta < 90º$ (only HE 20 element).
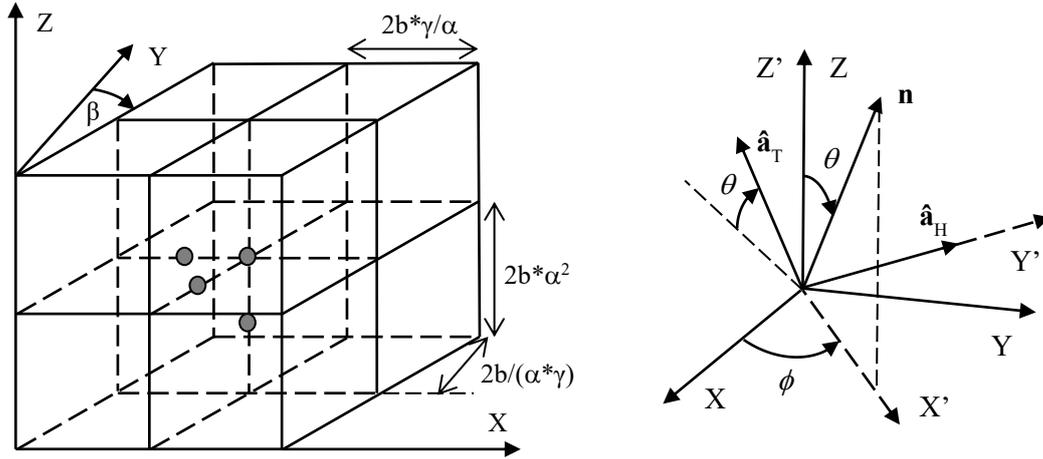
Figure 1. Electromagnetic domain discretized by a regular mesh of HE20 elements and unit cell with four nodes. Wave normal and polarization vectors

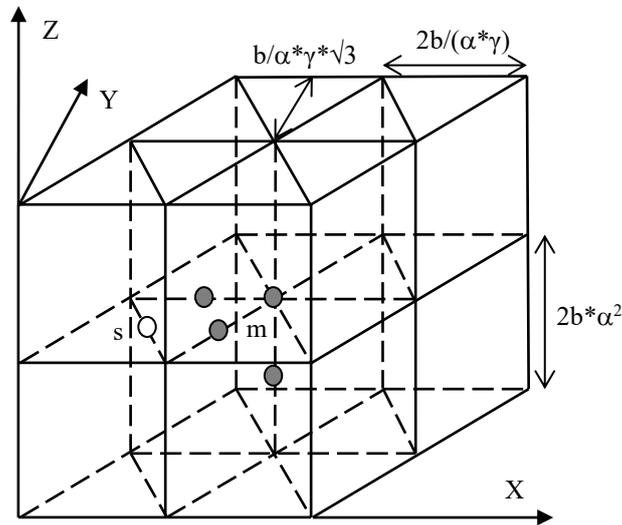**Figure 2. Electromagnetic domain discretized by a regular mesh of PE15 elements and unit cell with four master nodes (m) and one slave node (s)**

For plane harmonic waves,

$$\mathbf{U} = \tilde{\mathbf{u}}(\mathbf{r})\exp(-i\omega t), \quad \mathbf{r} \in \Omega; \quad \tilde{\mathbf{u}}(\mathbf{r}) = A\hat{\mathbf{a}}\exp(ik\mathbf{n}\cdot\mathbf{r}) \tag{13}$$

where: $\tilde{\mathbf{u}}$, complex amplitude of the field $\mathbf{U}$ (either $\mathbf{E}$ or $\mathbf{H}$); A, amplitude wave; $\mathbf{n}(\phi, \theta)$, wave normal; $\hat{\mathbf{a}}$, polarization vector; $k = 2\pi/\lambda = \omega/v$, wave number; λ, wavelength; $v = (\varepsilon\mu)^{-1/2}$, phase speed of the continuum; ϕ, azimuthal angle, $0 \le \phi \le 180^\circ$; θ, polar angle, $0 \le \theta \le 180^\circ$, Fig. 1.

Two different polarizations $\hat{\mathbf{a}} \cdot \mathbf{n} = 0$ will be considered: T-waves and H-waves, Fig. 1,

$$\hat{\mathbf{a}}_T = (-\cos\phi\cos\theta, -\sin\phi\cos\theta, \sin\theta) \quad (a), \quad \hat{\mathbf{a}}_H = (-\sin\phi, \cos\phi, 0) \quad (b) \tag{14}$$

For a plane harmonic wave Eq. (13), the density of period-averaged energy of the induced field can be computed by the equation [1],

$$\bar{E}_0 = \tfrac{1}{4}q^{-1}v^{-2}A^2 \tag{15}$$

2.1 Characteristic equations

The characteristic equations can be found assuming harmonic waves Eq. (13) with different amplitudes in each node of the unit cell,

$$\tilde{\mathbf{u}} = A_j\hat{\mathbf{a}}\exp(ik\mathbf{n}\cdot\mathbf{r}), \quad j = 1,\ldots,N \tag{16}$$

where: N, number of nodes per unit cell.

Inserting the solutions Eq. (16) into Eq. (6), the characteristic equation for each node of the unit cell is yielded by equilibrium of nodal forces into the direction of polarization [10],

$$q^{-1}\mathbf{F}_K \cdot \hat{\mathbf{a}} - \omega^2 p\mathbf{F}_M \cdot \hat{\mathbf{a}} = 0 \tag{17}$$

where: $\mathbf{F}_K$, nodal force associated to the global stiffness matrix; $\mathbf{F}_M$, nodal force associated to the global mass matrix.

By considering Eq. (17) for each node of the unit cell, a homogeneous system of N algebraic equations is formed for the unknown amplitudes,

$$(\mathbf{a}(m,\phi,\theta,\mathbf{p}) + \varpi^2\mathbf{b}(m,\phi,\theta,\mathbf{p}))\mathbf{A} = 0 \quad \Leftrightarrow \quad \mathbf{Z}\mathbf{A} = 0 \tag{18}$$

$$m = bk/\pi = 2b/\lambda \quad (a), \quad \varpi = (2b/v)\omega \quad (b) \tag{19}$$

where: m, dimensionless wave number, $0 < m < 1$; b, half of the element size; ϖ, dimensionless frequency of the discretized domain; $\mathbf{p} = (\gamma, \alpha, \beta)$ form parameter for HE20 mesh, or $\mathbf{p} = (\gamma, \alpha)$ form parameter for PE15 mesh.

In this procedure, the global stiffness and mass matrices have been expressed in the form,

$$\mathbf{K} = (2b)\mathbf{K}^0 \quad (a), \quad \mathbf{M} = (2b)^3\mathbf{M}^0 \quad (b) \tag{20}$$

2.2 Dispersion equations

The system of homogeneous algebraic equations given in Eq. (18) has a non-trivial solution only if the matrix $\mathbf{Z}$ is singular; that is, $\det[\mathbf{Z}] = 0$. Then it is yielded the following polynomial equation which is called the characteristic frequency equation for the plane wave propagation,

$$\sum_{r=0}^{N} c_r(m,\phi,\theta,\mathbf{p})\varpi^{2r} = 0, \quad c_N = 1 \tag{21}$$

It is an important fact that the N zeroes of a polynomial of degree $N \ge 1$ with complex coefficients depend continuously upon the coefficients [11]. Thus, sufficiently small changes in the coefficients of a polynomial can lead only to small changes in any zero. However, if the zeros are numerically computed, there is no simple way to define a function which takes the N coefficients (all but the leading 1) of a monic polynomial of degree N to the N zeroes of the

polynomial, since there is no natural way to define an ordering among the N zeroes. In the case of the HE20 mesh, for which the polynomial Eq. (21) is quartic, the above difficulty has been overcome by computing the zeroes in closed form as functions of its coefficients. Then, the components

$$\varpi_k = \varpi_k(m, \phi, \theta, \mathbf{p}), \quad k = 1, \ldots, 4 \tag{22}$$

will be continuous functions precisely defined. They are called the dispersion equations.

Substituting Eq. (22) into Eq. (18), the wave amplitudes corresponding to the nodes of the unit cell are yielded for each dispersion equation.

To find the zeros of a polynomial equation as functions of its coefficients beyond the quartic equation is a very difficult mathematical problem [12]. Then, for the mesh PE15 which has five nodes per unit cell, if the zeros of Eq. (21) are numerically computed, the above mentioned ordering difficulty could be a problem. In this case, by considering the initial condition,

$$\lim_{m \to 0} \varpi_1(m, \phi, \theta, \mathbf{p}) = 0 \tag{23}$$

it is proposed to compute the first dispersion equation by a reduced unit cell obtained by a procedure of exact dynamic condensation [13].

Assume that the total nodes at the unit cell are categorized as master nodes (m) and slave nodes (s), where the number of master nodes is four and the number of slave nodes is one, Fig 2. With this arrangement, the system of characteristic equations Eq. (18) may be partitioned as

$$\begin{pmatrix} \mathbf{Z}_{mm}(q) & \mathbf{Z}_{ms}(q) \\ \mathbf{Z}_{sm}(q) & \mathbf{Z}_{ss}(q) \end{pmatrix} \begin{pmatrix} \mathbf{A}_m \\ \mathbf{A}_s \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad q = \varpi^2 \tag{24}$$

From Eq. (24) the relation of wave amplitudes between the master and slave nodes may be obtained as

$$\mathbf{A}_s = -\mathbf{Z}_{ss}^{-1}(q)\mathbf{Z}_{sm}(q)\mathbf{A}_m \tag{25}$$

Then, by back-substituting, the system of characteristic equations of the reduced unit cell is obtained as

$$\mathbf{Z}_R(q)\mathbf{A}_m = \mathbf{0}, \quad [\mathbf{K}_R(q) + q\mathbf{M}_R(q)]\mathbf{A}_m = \mathbf{0} \tag{26}$$

Then, from Eq. (26), the reduced form of the characteristic frequency equation is obtained

$$q^4 + c_{3R}q^3 + c_{2R}q^2 + c_{1R}q + c_{0R} = 0, \quad c_{rR}(m, \phi, \theta, \mathbf{p}, q), \quad r = 0, \ldots, 3 \tag{27}$$

The first dispersion equation is computed by the following iterative procedure:

Set by Eq. (23) the initial value $q_{i1} = 0$.

Do for $0 < m < 1$, step Δm
1. Compute the coefficients $c_{rR}(m, \phi, \theta, \mathbf{p}, q_{i1})$ of Eq. (27).
2. Compute the zero q_1 of the quartic polynomial Eq. (27).
Do while $(\left| (q_1 - q_{i1})/q_1 \right| \geq \delta)$
1. Refresh initial value $q_{i1} = q_1$.
2. Compute the coefficients $c_{rR}(m, \phi, \theta, \mathbf{p}, q_{i1})$ of Eq. (27).
3. Compute the zero q_1 of the quartic polynomial Eq. (27).
End do

3. Substituting q_1 into Eq. (26) to compute the wave amplitudes at the master nodes.

4. Substituting q_1 into Eq. (25) to compute the wave amplitudes at the slave nodes.

5. Refresh initial value $q_{i1} = q_1$ for the next step.

End do

The range of dimensionless wave number values where each dispersion equation represents the propagation of electromagnetic waves in the discretized medium will be called the physical branch of the dispersion equation. In order to determine the physical branches the following constraint conditions are imposed,

$$A_1 = 1, \quad A_j(m,\phi,\theta,\mathbf{p}) > 0 \quad j = 2,\ldots,N \quad (a) \quad \left(\partial \varpi / \partial m\right)_{\phi,\theta,\mathbf{p}} > 0 \quad (b) \tag{28}$$

In molecular physics, condition Eq. (28a) is called the restriction of the lattice spectrum to the acoustical branch [14]. The preliminary constraint condition $\dim[\mathbf{N}(\mathbf{Z})] = 1$, over the dimension of the null space of matrix $\mathbf{Z}$ for the HE20 mesh, or $\dim[\mathbf{N}(\mathbf{Z}_R)] = 1$ for the PE15 mesh, must be imposed in order to Eq. (28a) would be a meaningful constraint condition. The constraint condition Eq. (28b) imposes that the phase velocity and the group velocity have the same sign. In this research, only the first physical branch will be selected for the analysis.

2.3 Energy of the induced field

From Eq. (11), Eq. (19b) and Eq. (20a) the density of period-averaged energy of the induced field will be computed

$$\bar{E} = \tfrac{1}{2} q^{-1} v^{-2} \varpi^{-2} \int_0^1 \mathrm{Im}[\tilde{\mathbf{u}}^T \exp(-i2\pi\tau)] \, \mathrm{Im}[\tilde{\mathbf{F}}^0 \exp(-i2\pi\tau)] d\tau \tag{29}$$

where: $\tau = t/T$, $0 \le \tau \le 1$, dimensionless time; $\mathbf{F}^0 = \tilde{\mathbf{F}}^0 \exp(-i2\pi\tau)$, column matrix of forces at the nodes of the unit cell.

The density of energy Eq. (29) is partitioned as addition of basic and higher order components. Each one of these energy components is then partitioned at the unit cell as addition of the component associated with the vertex node and the one associated with the mid-side nodes,

$$\bar{E} = \bar{E}_b + \bar{E}_h; \quad \bar{E}_b = \bar{E}_b^V + \bar{E}_b^M, \quad \bar{E}_h = \bar{E}_h^V + \bar{E}_h^M \tag{30}$$

Then, the percentages of basic and higher order energy can be defined

$$e_b(m,\phi,\theta,\mathbf{p}) = \bar{E}_b/\bar{E} \quad (a), \quad e_b^V = \bar{E}_b^V/\bar{E} \quad (b), \quad e_b^M = \bar{E}_b^M/\bar{E} \quad (c) \tag{31}$$

$$e_h(m,\phi,\theta,\mathbf{p}) = \bar{E}_h/\bar{E} \quad (a), \quad e_h^V = \bar{E}_h^V/\bar{E} \quad (b), \quad e_h^M = \bar{E}_h^M/\bar{E} \quad (c) \tag{32}$$

From Eq. (29) and Eq. (15), the percentage of energy error associated with the spatial discretization can be defined as,

$$e_P(m,\phi,\theta,\mathbf{p}) = \bar{E}/\bar{E}_0 - 1 \tag{33}$$

A mapping between percentage of energy error Eq. (33) and higher order energy Eq. (32a) can be also computed,

$$e_P = e_P(e_h,\phi,\theta,\mathbf{p}) \tag{34}$$

2.4 Numerical research

For the wave polarizations selected in Eq. (14), the indicators Eq. (31) to Eq. (33) are computed versus dimensionless wave number for different meshes and directions of wave propagation.

Five meshes having the same element volume will be considered for each of the elements analyzed. Specifically, for HE20 element, Fig. 1:

Q1: square section; $\gamma = 1$, $\beta = 0$, $\alpha = 1$.
Q2: rectangular section with aspect ratio 1:2; $\gamma = 1/\sqrt{2}$, $\beta = 0$, $\alpha = 1$.
Q3: skewed section; $\gamma = 1$, $\beta = 45°$, $\alpha = 1$.
Q4: distorted in the Z-direction with aspect ratio 2:1; $\gamma = 1$, $\beta = 0$, $\alpha = 2^{1/3}$.
Q5: distorted in the Z-direction with aspect ratio 1:2; $\gamma = 1$, $\beta = 0$, $\alpha = 1/2^{1/3}$.

and, for PE15 element, Fig. 2:

T1: triangular section of equilateral geometry; $\gamma = 1$, $\alpha = 1$.
T2: triangular section of right geometry; $\gamma^2 = 1/(3)^{1/2}$, $\alpha = 1$.
T3: triangular section with angle of 30° opposite to the base; $\gamma^2 = 1/(3)^{1/2} \, tg75°$, $\alpha = 1$.
T4: distorted in the Z-direction with aspect ratio 2:1; $\gamma = 1$, $\alpha = 2^{1/3}$.
T5: distorted in the Z-direction with aspect ratio 1:2; $\gamma = 1$, $\alpha = 1/2^{1/3}$.

The dispersion analysis is numerically carried out by a step of $\pi/36$ for the azimuthal and polar angles, and a step of 1/10000 for the dimensionless wave number.

For the mesh Q1 and T-waves, the indicators Eq. (32a) and Eq. (34) are plotted in Fig. 3 for three directions of wave propagation with the same azimuthal angle. Both for the mesh Q1 and the mesh T1, the percentages of vertex and mid-side basic energy, Eq. (31b) and Eq. (31c), and the percentages of vertex and mid-side higher order energy, Eq. (32b) and Eq. (32c), are also plotted for T-waves and H-waves, respectively, Fig. 4 and Fig. 5.

It is displayed in Fig. 3 how the percentage of higher order energy vanishes as dimensionless wave number goes to zero; that is, as the mesh is refined and in the limit of long waves. It is also displayed how the percentage of energy error vanishes as the percentage of higher order energy goes to zero. Then, it is inferred that the percentage of energy error vanishes as dimensionless wave number goes to zero.
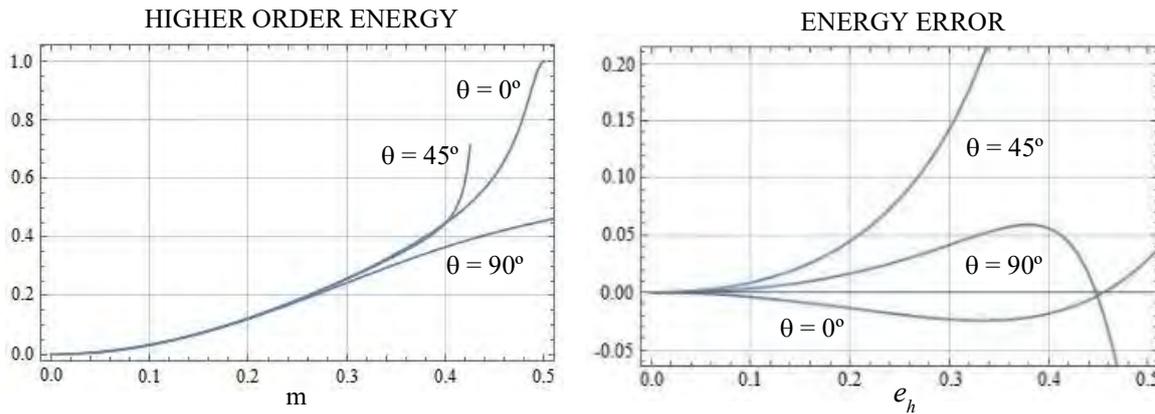
Figure 3. Percentage of higher order energy versus dimensionless wave number and percentages of energy error versus higher order energy. Test Q1 and T-waves, $\phi = 45°$

It can be concluded that, given the mesh, in the limit of long wavelength, although the energy density of the induced field does not vanish, its higher order component does vanish. Similarly, given the wavelength, as the solution converges on account of mesh refinement, the energy density is increasingly dominated by its basic component. The above behavior of the higher

order energy as dimensionless wave number goes to zero is a consequence that the induced field inside each element becomes uniform.

Figure 4. Vertex and mid-side components of the basic and higher order energy versus dimensionless wave number. Test Q1 and T-waves, $\phi = 45°$

Nevertheless, the higher order energy density vanishes as a cancellation at the unit cell of the component associated with the vertex node and the one associated with the mid-side nodes, which do not vanish but are equal and opposite in sign. A sign characteristic is also satisfied by the vertex and mid-side components of the basic energy density, Fig. 4 and Fig. 5,

$$e_b^V < 0 \quad (a), \quad e_b^M > 0 \quad (b), \quad e_h^V < 0 \quad (c), \quad e_h^M > 0 \quad (d) \tag{35}$$

Both for the HE20 meshes and for the PE15 meshes, the numerical research reveals that the sign characteristic Eq. (35) does not depend on the direction of wave propagation, direction of wave polarization, and mesh parameters.

For the HE20 meshes the convergence values of the vertex and mid-side energies also do not depend on the direction of wave propagation, direction of wave polarization, and mesh parameters, Table 1. Nevertheless, for the PE15 meshes these convergence values do depend on the propagation polar angle. Limit values for selected polar angles are shown in Table 2.

A discrete averaged relationship between the percentage higher order energy, the percentage energy error and the dimensionless wave number is investigated. Three reference values of the percentage higher order energy are selected, Table 3.
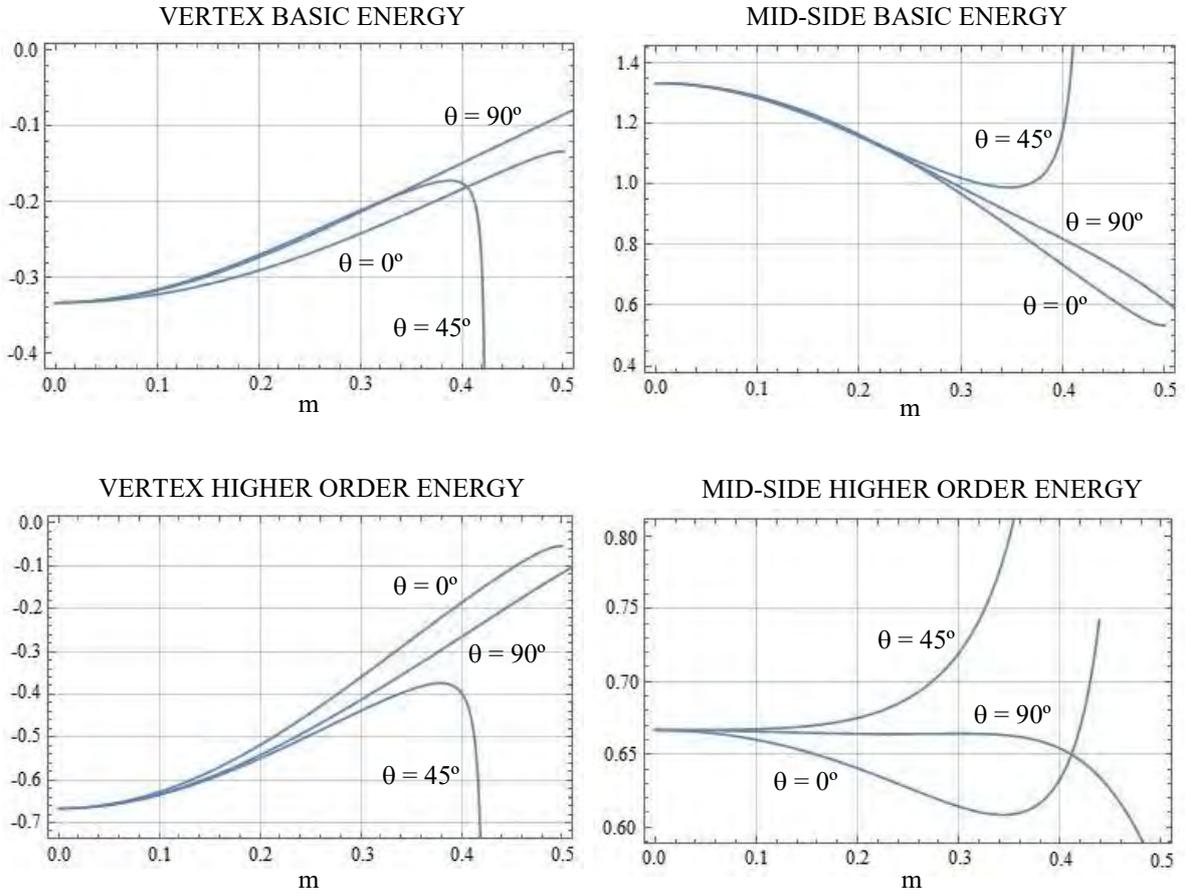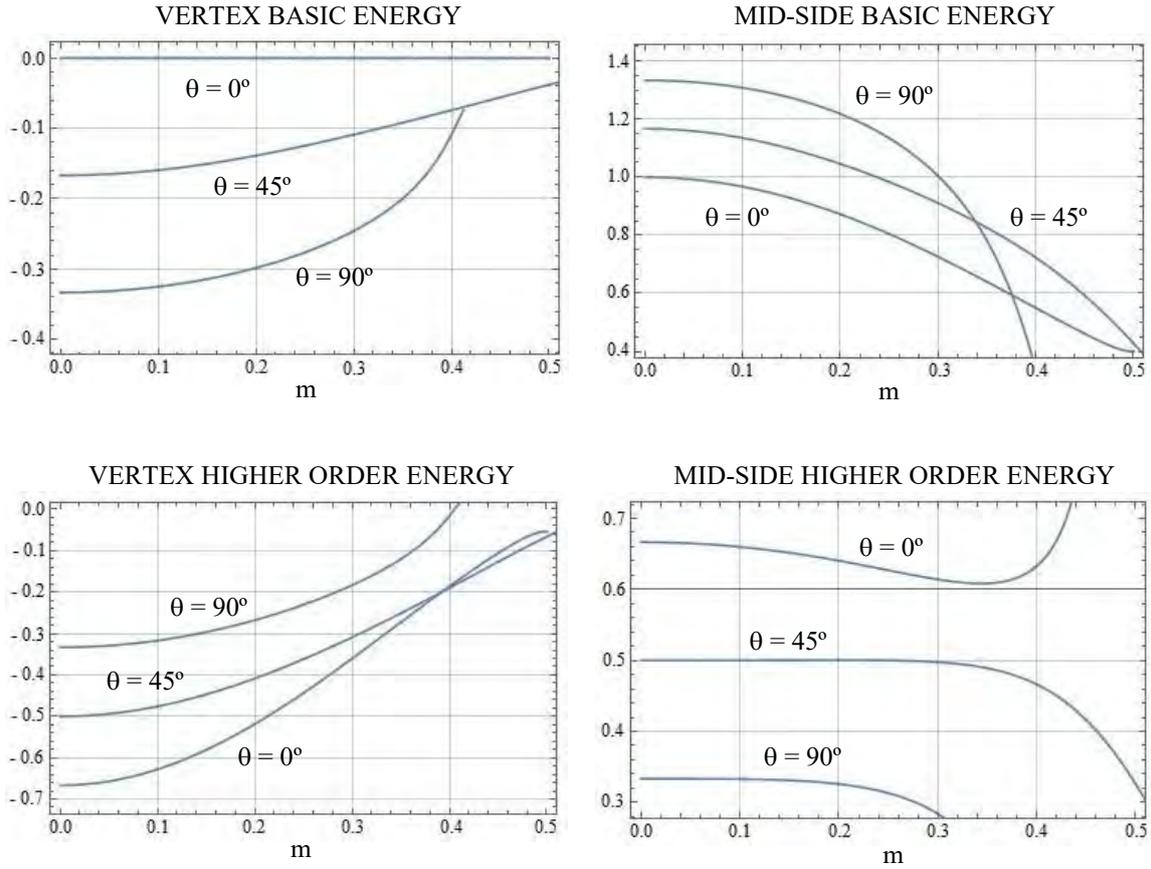
Figure 5. Vertex and mid-side components of the basic and higher order energy versus dimensionless wave number. Test T1 and H-waves, $\phi = 30°$

Table1. Convergence values of the vertex and mid-side energies. HE20 meshes

$\lim_{m \to 0} e_b^V$	$\lim_{m \to 0} e_b^M$	$\lim_{m \to 0} e_h^V$	$\lim_{m \to 0} e_h^M$
-1/3	4/3	-2/3	2/3

Table 2. Convergence values of the vertex and mid-side energies for selected propagation polar angles. PE15 meshes

$\theta°$	$\lim_{m \to 0} e_b^V$	$\lim_{m \to 0} e_b^M$	$\lim_{m \to 0} e_h^V$	$\lim_{m \to 0} e_h^M$
0	0	1	-2/3	2/3
45	-1/6	7/6	-1/2	1/2
90	-1/3	4/3	-1/3	1/3

Table 3. Reference values of the percentage higher order energy

	e_{h1}	e_{h2}	e_{h3}
HE20	0.05	0.10	0.20
PE15	0.0375	0.075	0.15

Then, by Eq. (32a) and Eq. (34), the related reference values of dimensionless wave number and percentage energy error are computed,

$$m_{1,2,3} = m_{1,2,3}(\phi,\theta,\mathbf{p}), \quad e_{P1,2,3} = e_{P1,2,3}(\phi,\theta,\mathbf{p}) \tag{36}$$

The mean value of each reference dimensionless wave number, and the rms value of each reference percentage energy error are computed on the range of propagation angles,

$$\bar{m}_{1,2,3}(\mathbf{p}) = \frac{1}{\pi^2}\int_0^\pi\int_0^\pi m_{1,2}\,d\phi\,d\theta \quad (a) \qquad e_{P1,2,3}^{RMS}(\mathbf{p}) = \sqrt{\frac{1}{\pi^2}\int_0^\pi\int_0^\pi |e_{P1,2}|^2\,d\phi\,d\theta} \quad (b) \tag{37}$$

Consistent with the discrete analysis carried out, the integrals in Eq. (37) are numerically computed. The results are displayed in Table 4 and Table 5 for the meshes and wave polarizations selected. Mesh-averaged values are also displayed.

Table 4. Mean values of reference dimensionless wave numbers and reference percentages of energy error computed over the range of azimuthal and polar angles. HE20 meshes

		$\bar{m}_1$	e_{P1}^{RMS}	$\bar{m}_2$	e_{P2}^{RMS}	$\bar{m}_3$	e_{P3}^{RMS}
MESH Q1	T	0.1249	0.001206	0.1791	0.005342	0.2603	0.026728
	H	0.1250	0.001022	0.1795	0.004454	0.2618	0.021399
MESH Q2	T	0.1217	0.001223	0.1745	0.005481	0.2537	0.028941
	H	0.1218	0.001317	0.1748	0.006427	0.2530	0.158606
MESH Q3	T	0.1189	0.001142	0.1705	0.005208	0.2477	0.059839
	H	0.1190	0.006372	0.1706	0.050212	0.2478	0.784589
MESH Q4	T	0.1079	0.002067	0.1544	0.011792	0.2132	0.633950
	H	0.1082	0.000912	0.1553	0.003876	0.2267	0.017785
MESH Q5	T	0.1361	0.001413	0.1952	0.006806	0.2824	0.135914
	H	0.1362	0.001100	0.1956	0.004801	0.2855	0.022919
Averaged values		0.1220	0.001777	0.1749	0.010440	0.2532	0.189067

Table 5. Mean values of reference dimensionless wave numbers and reference percentages of energy error computed over the range of azimuthal and polar angles. PE15 meshes.

		$\bar{m}_1$	e_{P1}^{RMS}	$\bar{m}_2$	e_{P2}^{RMS}	$\bar{m}_3$	e_{P3}^{RMS}
MESH T1	T	0.1275	0.001162	0.1826	0.005081	0.2644	0.024514
	H	0.1278	0.004064	0.1835	0.013707	0.2682	0.088720
MESH T2	T	0.1256	0.001314	0.1798	0.005766	0.2605	0.029494
	H	0.1258	0.008778	0.1806	0.026270	0.2633	0.140245
MESH T3	T	0.1238	0.001413	0.1772	0.006245	0.2565	0.053418
	H	0.1241	0.007427	0.1782	0.025523	0.2591	0.152633
MESH T4	T	0.1070	0.001993	0.1528	0.018430	0.2053	0.623055
	H	0.1074	0.002985	0.1541	0.010043	0.2251	0.062824
MESH T5	T	0.1432	0.001444	0.2051	0.006202	0.2971	0.032186
	H	0.1435	0.005326	0.2061	0.018155	0.3017	0.112612
Averaged values		0.1256	0.003591	0.1800	0.013542	0.2601	0.131970

It is verified that the first reference value of percentage higher order energy roughly correspond (in an averaged sense) to eight elements per wavelength, the second reference value would correspond to six elements per wavelength, and the third reference value would correspond to four elements per wavelength. For the first reference value, the mesh-averaged rms energy error is lower than one half per cent, an acceptable upper limit for high precision computation. For

the second reference value, the mesh-averaged rms energy error is slightly over one per cent, an acceptable upper limit for engineering computations. For the third reference value, the mesh-averaged rms energy error clearly indicates a coarse precision computation.

3. Cavity eigenmodes

The standing-wave fields and natural frequencies in an electromagnetic cavity discretized by a finite element mesh will be the solution of the eigenproblem associated to Eq. (6), which is suitably expressed in wave number form,

$$\mathbf{K}\tilde{\mathbf{u}}_j - k_j^2 \mathbf{M}\tilde{\mathbf{u}}_j = \mathbf{0}, \quad k_j^2 = \varepsilon\mu\omega_j^2; \quad j = 1,\ldots,N \tag{38}$$

The eigenproblem Eq. (38) is solved by the LAPACK routine DSPCVX [15]. From Eq. (12), for each eigenmode computed, both the basic component and higher order component of the period-averaged energy of the induced field will be computed in percentage form,

$$EB_j = \overline{E}_{b,j}\big/\overline{E}_j; \quad j = 1,\ldots,N \quad (a), \quad EH_j = \overline{E}_{h,j}\big/\overline{E}_j; \quad j = 1,\ldots,N \quad (b) \tag{39}$$

Then, both the basic energy and higher order energy will be expressed as addition of the component associated with the vertex nodes and the one associated with the mid-side nodes,

$$EB_j = EBv_j + EBm_j; \quad j = 1,\ldots,N \quad (a), \quad EH_j = EHv_j + EHm_j; \quad j = 1,\ldots,N \quad (b) \tag{40}$$

As it is well known, if the electromagnetic cavity is discretized by nodal finite elements then the eigenspectrum computed by Eq. (38) will be severely polluted with spurious eigenmodes, the spurious eigenvalues having the same order of magnitude as the physical eigenvalues and not easily distinguished from them. Additionally, the number of spurious eigenmodes computed between consecutive physical eigenmodes generally will increase if the mesh is refined [7]. In short, the spurious eigenmodes are numerical solutions of the eigenproblem that do not converge to any physical solution as the mesh is refined [16].

In this contribution, in order to distinguish between the physical eigenmodes and the spurious ones, it is proposed to select only those eigenmodes for which the basic energy component of the induced field prevails over the higher order energy component. The third reference value of percentage higher order energy displayed in Table 3 is adopted as reference. Also, the vertex and mid-side energy components must fulfill the sign characteristic deduced for these energy components by the dispersion analysis carried out for the fundamental solution of plane harmonic waves, Eq. (35). Then, it will be proposed as candidates for physical eigenmodes those finite element eigenmodes for which,

$$EB_j > 0.80 \quad for \quad HE20; \quad EB_j > 0.85 \quad for \quad PE15 \tag{41a}$$

$$EBv_j < 0 \quad (i), \quad EBm_j > 0 \quad (ii), \quad EHv_j < 0 \quad (iii), \quad EHm_j > 0 \quad (iv) \tag{41b}$$

In order to test the proposed criterion three cavities for which the eigenproblem Eq. (3) has analytical solution are analyzed. In each case, the eigenproblem Eq. (38) is solved for the electric field ($\mathbf{U} = \mathbf{E}$) by applying the boundary condition Eq. (2a).

3.1 Test problem 1. Rectangular cavity

The geometry of the cavity, discretized by a HE20 mesh, is illustrated in Fig. 6. The eigenmodes up to $N = 1200$ are computed, being the modal wave number versus the number of mode displayed in a stick diagram, Fig. 6.

Next, if the percentage basic energy versus the number of mode is displayed, the finite element eigenmodes fulfilling the criterion Eq. (41a) are clearly highlighted, Fig. 7. These eigenmodes

are listed in Table 6 and all of them fulfill the criterion Eq. (41b). They will be called the filtered eigenmodes up to $Nf = 25$.
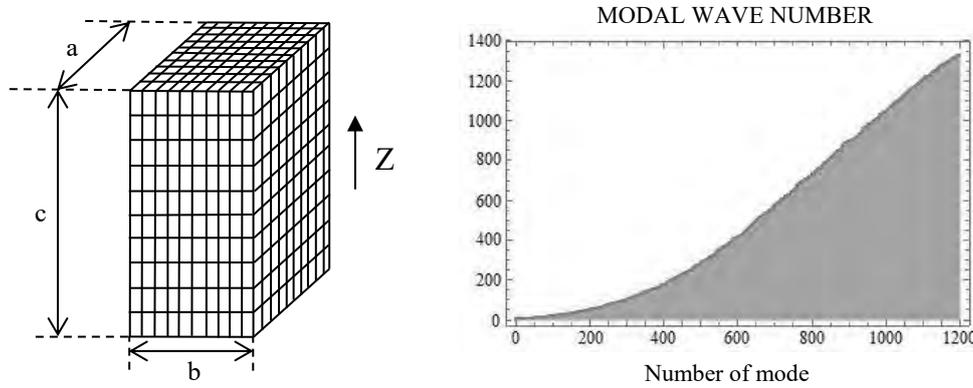

Figure 6. Rectangular cavity. Dimensions: a = 0.0075 m; b = 0.0050 m; c = 0.01 m. Stick diagram of modal wave number versus number of mode

Table 6. Rectangular cavity discretized by HE20 finite elements. Filtered eigenmodes versus number of mode

|	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
1	670	523.602	-0.325187	1.308900	-0.649818	0.666105	0.983713
2	783	702.486	-0.325188	1.308900	-0.649818	0.666105	0.983713
3	819	755.150	-0.325187	1.308900	-0.649818	0.666105	0.983713
4	820	755.202	-0.314793	1.274683	-0.623591	0.663700	0.959890
5	858	817.893	-0.322269	1.297931	-0.646678	0.671016	0.975662
6	859	817.893	-0.322064	1.297726	-0.646351	0.670689	0.975662
7	891	888.627	-0.313276	1.273208	-0.625110	0.665177	0.959932
8	901	894.810	-0.316248	1.276098	-0.622136	0.662286	0.959850
9	960	982.400	-0.313053	1.265031	-0.627935	0.675958	0.951978
10	961	982.405	-0.309634	1.261730	-0.622538	0.670442	0.952096
11	992	1031.824	-0.301090	1.222696	-0.578753	0.657148	0.921606
12	1003	1047.272	-0.314380	1.274282	-0.624004	0.664102	0.959902
13	1004	1047.313	-0.301581	1.238362	-0.600855	0.664073	0.936781
14	1033	1093.373	-0.313640	1.265613	-0.625269	0.673295	0.951974
15	1034	1093.378	-0.310619	1.262677	-0.620499	0.668441	0.952058
16	1054	1133.133	-0.298207	1.219982	-0.581636	0.659861	0.921776
17	1104	1208.038	-0.298804	1.212731	-0.588082	0.674155	0.913927
18	1105	1208.077	-0.293441	1.207771	-0.579662	0.665331	0.914330
19	1115	1221.320	-0.301915	1.230845	-0.607645	0.678715	0.928930
20	1116	1221.329	-0.298212	1.227317	-0.601891	0.672787	0.929104
21	1145	1261.441	-0.283715	1.183584	-0.561299	0.661431	0.899868
22	1172	1295.442	-0.316758	1.276594	-0.621629	0.661793	0.959836
23	1174	1295.952	-0.303236	1.224715	-0.576616	0.655136	0.921480
24	1198	1324.739	-0.316432	1.276277	-0.621953	0.662108	0.959845
25	1199	1326.526	-0.283779	1.154325	-0.518735	0.648188	0.870546
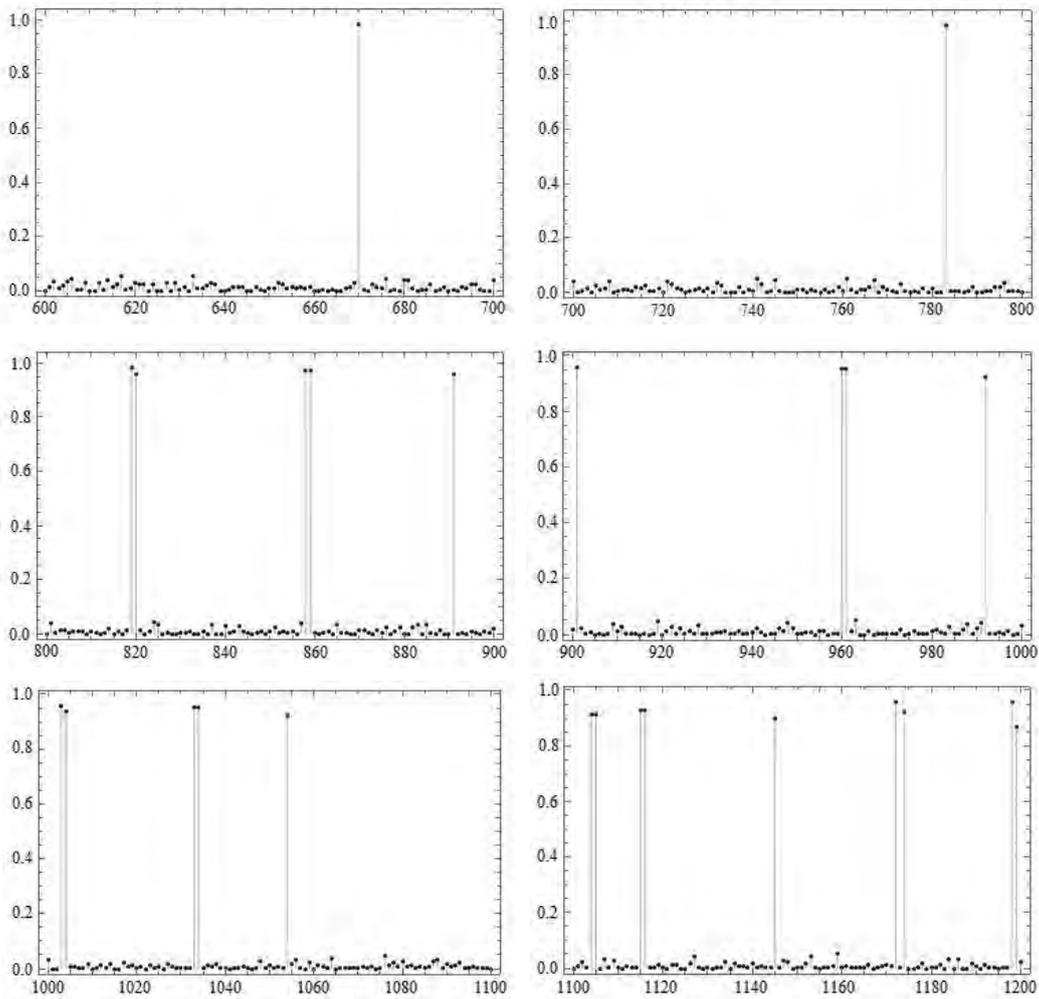
Figure 7. Rectangular cavity. Stick diagram of percentage basic energy versus number of mode highlighting the filtered eigenmodes

From Table 6 it is verified that the vertex and mid-side energy components of the filtered eigenmodes versus the number of mode have values in accordance with the ones exhibit by these energy components for the fundamental solution of plane harmonic waves versus dimensionless wave number, Fig. 4 and Table 1.

The exact wave numbers up to $Ne = 25$ are also computed and listed in the Table 7 [1][17].

Table 7. Rectangular cavity. Exact TE and TM wave numbers with respect to the Z axis

k_{mnp} (1/m)	Mode	k_{mnp} (1/m)	Mode
523.599	TE_{101}	1047.197	TE_{202} TM_{210}
702.481	TE_{011}	1093.306	TE_{211} TM_{211}
755.145	TE_{102} TM_{110}	1132.717	TE_{013}
817.887	TE_{111} TM_{111}	1207.687	TE_{113} TM_{113}
888.577	TE_{012}	1221.232	TE_{212} TM_{212}
894.726	TE_{201}	1260.993	TE_{203}
982.358	TE_{112} TM_{112}	1295.312	TE_{021} TE_{301}
1031.370	TE_{103}	1324.612	TE_{104} TM_{120}

From Table 6 and Table 7 it is verified that the wave numbers of the filtered eigenmodes are an accurate approximation of the exact wave numbers. Then, it can be inferred that the filtered eigenmodes will be physical eigenmodes. This set of physical eigenmodes has an upper limit of percentage higher order energy around the second reference value defined in Table 3, which roughly corresponds to six elements per wavelength, Table 4. Then, it can be concluded that the above set of eigenmodes is accurately captured at least from the engineering point of view.

The numerical research has revealed that a spurious eigenmode polluting the HE20 mesh is characterized by its induced field having a small percentage of basic energy, lower than about ten per cent, as it can be verified in Fig. 7; therefore, the induced field at the element will be far from a uniform field, on the contrary, it will fluctuate around a small mean value. A spurious eigenmode moreover generally does not fulfill the sign characteristic Eq. (41b).

3.2 Test problem 2. Coaxial cavity

The geometry of the cavity, discretized by a PE15 mesh, is illustrated in Fig. 8. The eigenmodes up to $N = 1740$ are computed. The finite element eigenmodes fulfilling the criterion Eq. (41a) are clearly highlighted in Fig. 9. These eigenmodes are listed in Table 8 and all of them fulfill the criterion Eq. (41b). They will be the filtered eigenmodes up to $Nf = 25$.
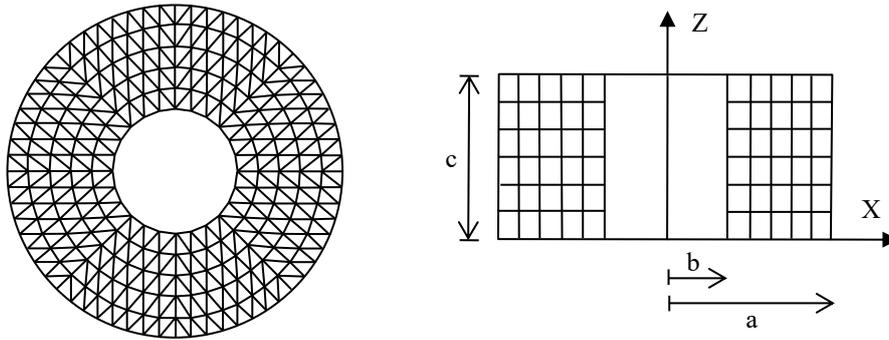

Figure 8. Coaxial cavity. Dimensions: a = 0.012 m; b = 0.0048 m; c = 0.012 m

Table 8a. Coaxial cavity discretized by PE15 finite elements. Filtered eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
1	1114	261.81309	-0.0000004	0.9737565	-0.6516177	0.6778616	0.9737561
2	1180	288.89345	-0.0569441	1.0309873	-0.5759083	0.6018651	0.9740432
3	1181	288.89345	-0.0569441	1.0309873	-0.5759083	0.6018651	0.9740432
4	1318	353.36193	-0.1411792	1.1139676	-0.4830186	0.5102302	0.9727884
5	1319	353.36193	-0.1411792	1.1139676	-0.4830186	0.5102302	0.9727884
6	1433	431.00207	-0.3227638	1.2983996	-0.3122760	0.3366402	0.9756357
7	1434	431.34128	-0.1952180	1.1643963	-0.4224754	0.4532971	0.9691782
8	1435	431.34128	-0.1952180	1.1643963	-0.4224754	0.4532971	0.9691782
9	1474	448.48012	-0.3223031	1.2975804	-0.3113305	0.3360532	0.9752773
10	1475	448.48012	-0.3223031	1.2975805	-0.3113305	0.3360532	0.9752773
11	1568	496.71406	-0.3209277	1.2946294	-0.3085137	0.3348120	0.9737017
12	1569	496.71406	-0.3209277	1.2946294	-0.3085137	0.3348120	0.9737017
13	1590	504.24895	-0.2290334	1.1827787	-0.3902869	0.4365415	0.9537453
14	1615	512.51472	-0.2172881	1.1792377	-0.3723524	0.4104029	0.9619496
15	1616	512.51704	-0.2276042	1.1931728	-0.3863308	0.4207621	0.9655686

Table 8b. Coaxial cavity discretized by PE15 finite elements. Filtered eigenmodes versus number of mode

|	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
16	1625	519.18375	-0.2376362	1.1921561	-0.3888445	0.4343246	0.9545199
17	1626	519.29722	-0.2332008	1.1848104	-0.3891978	0.4375882	0.9516097
18	1627	519.29722	-0.2332008	1.1848104	-0.3891978	0.4375882	0.9516097
19	1639	524.02262	-0.0000073	0.9046323	-0.5541856	0.6495606	0.9046250
20	1674	538.78235	-0.2419406	1.1951002	-0.3785361	0.4253764	0.9531597
21	1675	538.78235	-0.2419406	1.1951002	-0.3785361	0.4253764	0.9531597
22	1721	561.44017	-0.2423589	1.1916725	-0.3480536	0.3987400	0.9493135
23	1722	561.44017	-0.2423589	1.1916725	-0.3480535	0.3987400	0.9493135
24	1733	566.60745	-0.3186614	1.2888503	-0.3038885	0.3336997	0.9701889
25	1734	566.60745	-0.3186614	1.2888503	-0.3038885	0.3336997	0.9701889
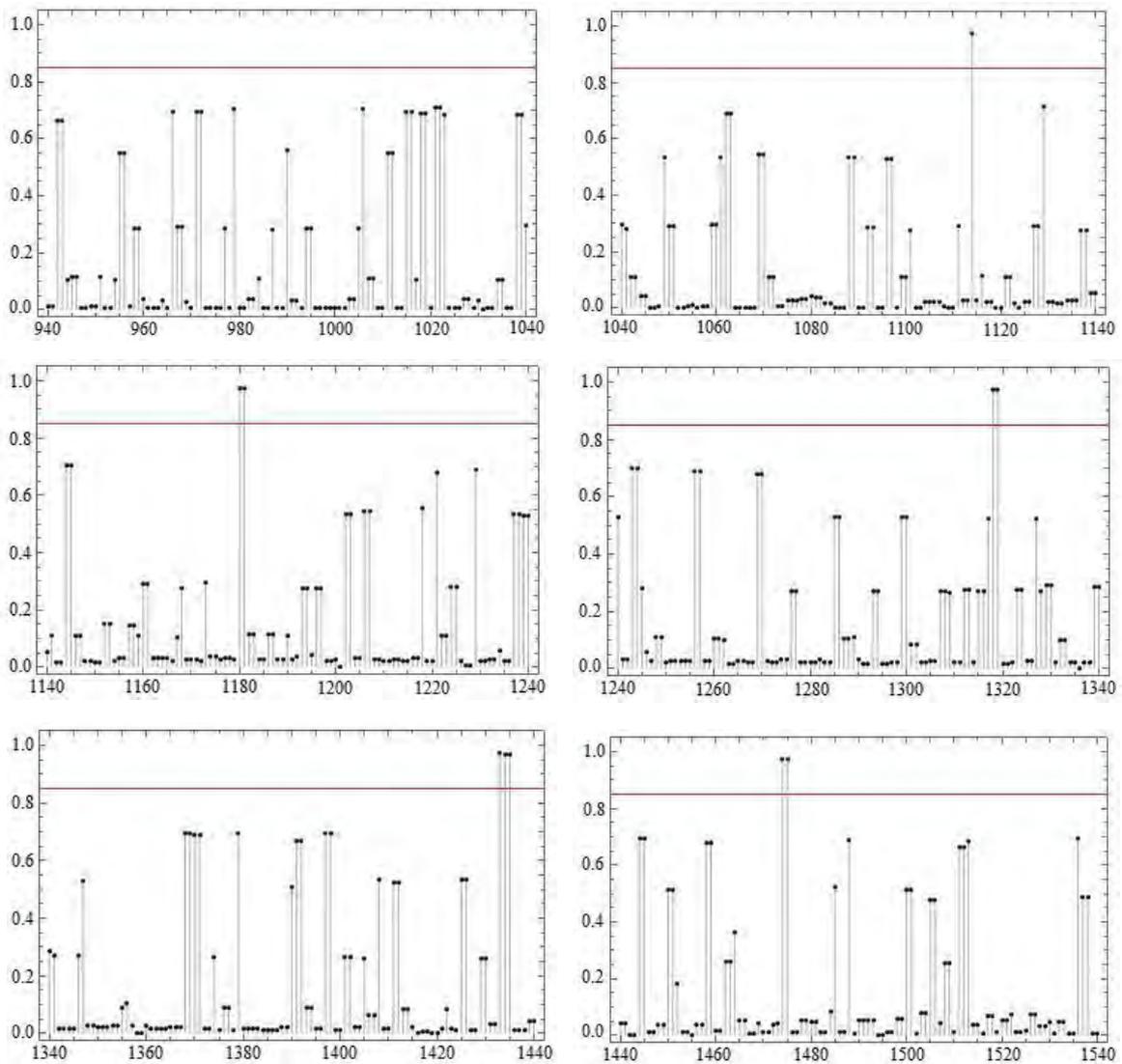
Figure 9a. Coaxial cavity. Stick diagram of percentage basic energy versus number of mode highlighting the filtered eigenmodes
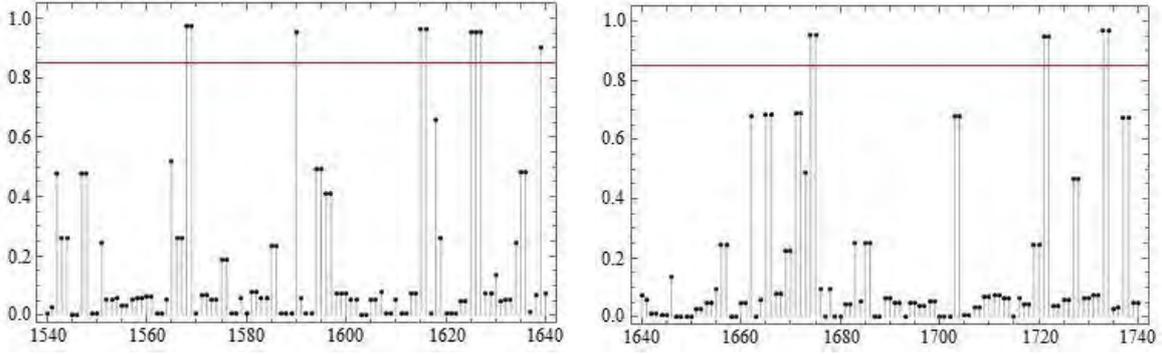
Figure 9b. Coaxial cavity. Stick diagram of percentage basic energy versus number of mode highlighting the filtered eigenmodes

The exact wave numbers up to $Ne = 15$, which are related to the filtered eigenmodes, are listed in Table 9 [17].

Table 9. Coaxial cavity. Exact TE and TM wave numbers with respect to the Z axis

k_{mnp} (1/m)	Mode	k_{mnp} (1/m)	Mode
261.79939	TE_{001}	505.02964	TM_{011}
288.77748	TE_{101}	513.06967	TE_{401}
353.05621	TE_{201}	520.07385	TE_{011} TM_{111}
430.93213	TE_{301}	523.59878	TE_{002}
431.87500	TM_{010}	539.45191	TE_{111}
449.37500	TM_{110}	561.99459	TM_{211}
497.29167	TM_{210}	566.66667	TM_{310}

From Table 8 and Table 9 it is verified that the wave numbers of the filtered eigenmodes are an accurate approximation of the exact wave numbers. Then, it can be inferred that the filtered eigenmodes will be physical eigenmodes. This set of physical eigenmodes has an upper limit of percentage higher order energy around the second reference value defined in Table 3, which roughly corresponds to six elements per wavelength, Table 5. Then, it can be concluded that the above set of eigenmodes is accurately captured at least from the engineering point of view.

It must be remarked how, by considering the null value of vertex basic energy in Table 2 for Z-propagation, the TEM modes (TE_{001} and TE_{002}) can be easily identified in Table 8.

Differently to the HE20 mesh, the numerical research has revealed that the spurious eigenmodes polluting the PE15 mesh have induced fields with percentage basic energy within a wide range of values. Then, as the modal order increases, by applying the criterion Eq. (41a), non-spurious eigenmodes could be rejected, but such eigenmodes would be so far from convergence that they could be also described as spurious in a different meaning of the word.

3.3 Test problem 3. Spherical cavity

The geometry of the cavity is illustrated in Fig. 10. Since the spherical cavity has three planes of symmetry, which are *YZ*, *XZ* and *XY*, the eigenmodes can be computed by idealizing one eighth of it, and applying combinations of boundary conditions on the three planes for symmetric field (S) and for antisymmetric field (A). The combinations (SSS), (AAA), (SSA) and (AAS) are considered. The domain is discretized by ten radial layers of finite elements which include 720 HE20, 540 PE15, and 60 TE10 (tetrahedron with ten nodes) as fillers. The

first four radial layers of finite elements are showed (discretizing the half of the domain) to highlight how the finite element mesh is built up, Fig. 10.
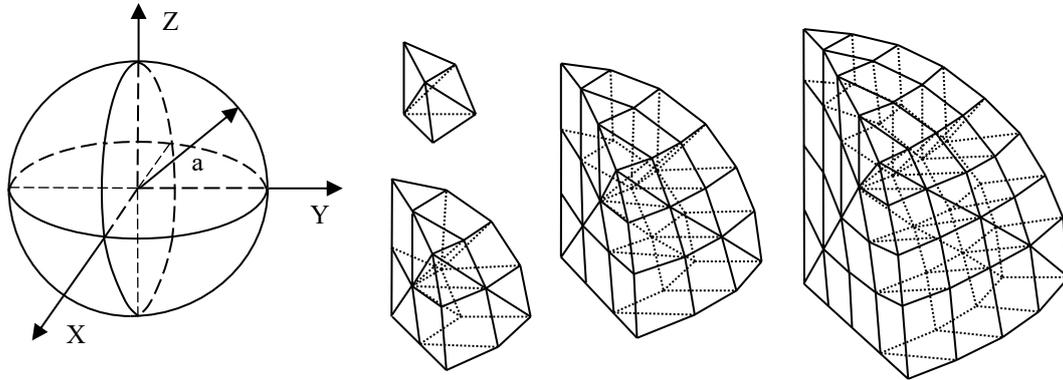

Figure 10. Spherical cavity. Dimension: a = 0.01 m.

It must be remarked that the sign characteristic Eq. (41b) is also satisfied with eigenmodes computed by TE10 meshes, as just it has been verified in the context of solid dynamics [18].

The finite element eigenmodes (SSS, AAA, SSA, AAS) up to N = (1600, 1350, 1500, 1400) are computed. The eigenmodes fulfilling the criterion Eq. (41a) for the HE20 element are displayed in Table 10 to Table 13, and all of them fulfill the criterion Eq. (41b). As a particular case, the filtered SSA eigenmodes are clearly highlighted in Fig. 11.

The exact wave numbers up to Ne = 35, which are related to the filtered eigenmodes, are listed in the Table 14 [1][17][19]. Note that there are numerous degeneracies (same resonant wave number) among the modes, since k is independent of m.

Table 10a. Spherical cavity. Filtered SSS eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
1	754	387.384299	-0.302114	1.295441	-0.586042	0.592714	0.993328
2	755	387.441125	-0.290915	1.264502	-0.571678	0.598091	0.973587
3	947	606.776953	-0.295625	1.270581	-0.588495	0.613539	0.974956
4	948	606.783806	-0.297206	1.257160	-0.552545	0.592591	0.959954
5	949	606.857319	-0.308993	1.279956	-0.570653	0.599690	0.970963
6	1062	745.022728	-0.237073	1.082067	-0.441865	0.596871	0.844994
7	1063	745.039659	-0.276679	1.229694	-0.538410	0.585395	0.953015
8	1122	821.826974	-0.267351	1.224214	-0.545801	0.588938	0.956863
9	1123	821.896359	-0.301638	1.248023	-0.563234	0.616850	0.946384
10	1124	821.977650	-0.269379	1.101628	-0.514883	0.682633	0.832249
11	1125	822.014327	-0.298057	1.255870	-0.563266	0.605453	0.957813
12	1216	936.546747	-0.268040	1.181125	-0.512951	0.599866	0.913085
13	1217	936.606460	-0.215482	1.030645	-0.413789	0.598626	0.815163
14	1268	997.991568	-0.268099	1.175321	-0.531269	0.624047	0.907223
15	1298	1034.408650	-0.230311	1.043027	-0.441846	0.629130	0.812716
16	1299	1034.498216	-0.271219	1.179335	-0.530504	0.622387	0.908116
17	1300	1034.655401	-0.278137	1.152650	-0.491009	0.616496	0.874513
18	1301	1034.666719	-0.280139	1.148566	-0.513753	0.645326	0.868426
19	1312	1043.215409	-0.237258	1.081272	-0.483294	0.639280	0.844014

Table 10b. Spherical cavity. Filtered SSS eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
20	1337	1072.924179	-0.244034	1.094331	-0.439639	0.589343	0.850297
21	1414	1167.352597	-0.255025	1.152959	-0.519346	0.621413	0.897934
22	1415	1167.413535	-0.271717	1.176351	-0.528547	0.623913	0.904634
23	1475	1240.999845	-0.230146	1.040062	-0.424664	0.614748	0.809917
24	1476	1241.078882	-0.258709	1.127494	-0.500348	0.631563	0.868785
25	1482	1245.642304	-0.242435	1.099160	-0.462748	0.606023	0.856725
26	1483	1245.660483	-0.255526	1.131347	-0.524081	0.648260	0.875822
27	1484	1245.927494	-0.242620	1.076850	-0.449693	0.615463	0.834229
28	1486	1246.126101	-0.274263	1.159214	-0.515269	0.630318	0.884951
29	1532	1298.938219	-0.252723	1.141120	-0.493509	0.605113	0.888397

Table 11. Spherical cavity. Filtered AAA eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
1	709	497.762426	-0.296651	1.285735	-0.576482	0.587399	0.989083
2	771	576.879945	-0.293265	1.273396	-0.560940	0.580809	0.980131
3	877	714.662940	-0.265139	1.126469	-0.509212	0.647882	0.861329
4	878	714.694758	-0.308369	1.282440	-0.602177	0.628106	0.974071
5	951	819.091890	-0.278541	1.229658	-0.567065	0.615948	0.951117
6	952	819.121344	-0.275475	1.220532	-0.553431	0.608374	0.945057
7	953	819.171230	-0.284923	1.224568	-0.569372	0.629726	0.939645
8	993	873.134942	-0.280323	1.228281	-0.533347	0.585389	0.947958
9	1020	910.533563	-0.262072	1.200210	-0.496050	0.557912	0.938138
10	1021	910.567491	-0.267379	1.216843	-0.506641	0.557177	0.949464
11	1033	928.308814	-0.268110	1.172908	-0.510220	0.605422	0.904798
12	1035	928.431115	-0.299128	1.252166	-0.560386	0.607348	0.953038
13	1122	1052.430230	-0.251175	1.159661	-0.548662	0.640176	0.908486
14	1123	1052.605721	-0.294673	1.227963	-0.528425	0.595135	0.933290
15	1125	1052.700838	-0.280147	1.201945	-0.517660	0.595862	0.921799
16	1173	1120.320273	-0.255679	1.140031	-0.497321	0.612969	0.884351
17	1189	1140.138624	-0.247061	1.141646	-0.509381	0.614797	0.894585
18	1191	1140.244013	-0.265301	1.148120	-0.528032	0.645213	0.882819
19	1192	1140.371787	-0.256701	1.065811	-0.456230	0.647121	0.809109
20	1218	1172.492423	-0.232025	1.043612	-0.443776	0.632189	0.811587
21	1219	1172.497518	-0.258982	1.152165	-0.514151	0.620968	0.893183
22	1245	1208.363449	-0.209955	1.024692	-0.379631	0.564894	0.814737
23	1301	1280.871486	-0.243118	1.123412	-0.498044	0.617751	0.880293
24	1304	1281.336961	-0.252801	1.132739	-0.479293	0.599355	0.879938
25	1305	1281.453642	-0.225606	1.053407	-0.412717	0.584916	0.827801

Table 12a. Spherical cavity. Filtered SSA eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
1	596	274.666771	-0.285236	1.275772	-0.560493	0.569957	0.990536
2	807	497.833259	-0.301056	1.256410	-0.609581	0.654226	0.955354
3	874	576.899024	-0.274037	1.234801	-0.580916	0.620153	0.960763
4	905	612.314450	-0.255685	1.218202	-0.521903	0.559386	0.962517
5	983	714.674404	-0.291752	1.264539	-0.592389	0.619602	0.972787
6	984	714.712675	-0.302411	1.253497	-0.591859	0.640773	0.951086

Table 12b. Spherical cavity. Filtered SSA eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
7	985	714.814105	-0.309843	1.280700	-0.569315	0.598457	0.970858
8	1069	819.022507	-0.190978	1.009661	-0.396281	0.577598	0.818682
9	1070	819.188414	-0.280431	1.227157	-0.521219	0.574493	0.946726
10	1110	873.149430	-0.257034	1.144738	-0.535775	0.648070	0.887704
11	1139	910.654906	-0.248727	1.185697	-0.534238	0.597268	0.936970
12	1155	928.316350	-0.256059	1.165044	-0.514917	0.605932	0.908985
13	1156	928.407839	-0.289100	1.225710	-0.567307	0.630696	0.936610
14	1157	928.509955	-0.303917	1.235705	-0.577115	0.645328	0.931787
15	1158	928.800733	-0.251067	1.066495	-0.433495	0.618067	0.815427
16	1164	932.859090	-0.251676	1.154518	-0.492134	0.589292	0.902843
17	1252	1052.571053	-0.267210	1.180212	-0.512638	0.599636	0.913002
18	1254	1052.739422	-0.256649	1.177796	-0.504312	0.583165	0.921147
19	1304	1120.381315	-0.253707	1.142665	-0.508977	0.620019	0.888958
20	1305	1120.389856	-0.267564	1.151074	-0.515946	0.632436	0.883510
21	1306	1120.510835	-0.278625	1.173613	-0.505553	0.610566	0.894987
22	1326	1140.234122	-0.269192	1.144622	-0.493331	0.617901	0.875430
23	1327	1140.475667	-0.268120	1.159486	-0.507480	0.616115	0.891365
24	1328	1140.573527	-0.287836	1.191289	-0.510144	0.606691	0.903453
25	1387	1208.565919	-0.263508	1.153090	-0.483288	0.593705	0.889583
26	1409	1234.940006	-0.183431	0.998421	-0.374880	0.559890	0.814990
27	1421	1251.081458	-0.228073	1.092902	-0.438575	0.573746	0.864829
28	1449	1280.840203	-0.265441	1.137084	-0.503777	0.632134	0.871643
29	1450	1281.268600	-0.230973	1.042894	-0.449363	0.637442	0.811921
30	1451	1281.586214	-0.195749	1.040012	-0.348695	0.504432	0.844263

Table 13a. Spherical cavity. Filtered AAS eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
1	672	387.376592	-0.304338	1.293828	-0.600421	0.610931	0.989489
2	729	449.778566	-0.276310	1.261222	-0.590897	0.605985	0.984912
3	854	606.697683	-0.289302	1.271606	-0.577567	0.595263	0.982304
4	855	606.787719	-0.301594	1.197660	-0.569468	0.673402	0.896066
5	924	699.489611	-0.295691	1.264237	-0.563566	0.595020	0.968546
6	925	699.500497	-0.264388	1.222423	-0.538019	0.579984	0.958035
7	961	745.099486	-0.273117	1.235906	-0.560040	0.597251	0.962789
8	985	773.385464	-0.254640	1.195658	-0.548152	0.607134	0.941018
9	1020	821.804722	-0.292596	1.257608	-0.573333	0.608321	0.965012
10	1021	821.858268	-0.287557	1.248186	-0.602757	0.642129	0.960629
11	1022	821.920897	-0.306943	1.256693	-0.561378	0.611627	0.949751
12	1107	936.566726	-0.273412	1.156617	-0.508112	0.624906	0.883206
13	1108	936.636672	-0.272270	1.198523	-0.558203	0.631949	0.926254
14	1109	936.694024	-0.264391	1.169560	-0.518274	0.613105	0.905169
15	1151	997.976828	-0.278735	1.205274	-0.538708	0.612169	0.926539
16	1179	1034.463048	-0.278732	1.182764	-0.532260	0.628228	0.904032
17	1180	1034.510778	-0.284126	1.192890	-0.547984	0.639221	0.908763
18	1181	1034.540940	-0.278853	1.163961	-0.517504	0.632396	0.885108
19	1191	1043.159410	-0.214730	1.059485	-0.422628	0.577873	0.844755
20	1192	1043.186928	-0.256032	1.152224	-0.477793	0.581601	0.896193

Table 13b. Spherical cavity. Filtered AAS eigenmodes versus number of mode

	# Mode	k (1/m)	EBv	EBm	EHv	EHm	EB
21	1211	1072.980088	-0.245178	1.152917	-0.498710	0.590971	0.907740
22	1229	1092.169254	-0.223622	1.093203	-0.465906	0.596325	0.869581
23	1285	1167.265460	-0.267757	1.138783	-0.488192	0.617166	0.871026
24	1287	1167.462416	-0.289461	1.198819	-0.522117	0.612759	0.909358
25	1288	1167.466504	-0.261162	1.150896	-0.474718	0.584985	0.889734
26	1342	1240.863390	-0.249112	1.153542	-0.526752	0.622321	0.904431
27	1344	1241.084865	-0.249948	1.074742	-0.437905	0.613111	0.824794
28	1349	1245.574057	-0.246853	1.114791	-0.479219	0.611280	0.867938
29	1350	1245.584561	-0.223958	1.083118	-0.469635	0.610474	0.859161
30	1351	1245.753593	-0.266272	1.134617	-0.510892	0.642546	0.868345
31	1394	1298.992769	-0.224378	1.044407	-0.440834	0.620805	0.820029
32	1396	1299.308124	-0.251751	1.133094	-0.489741	0.608398	0.881342
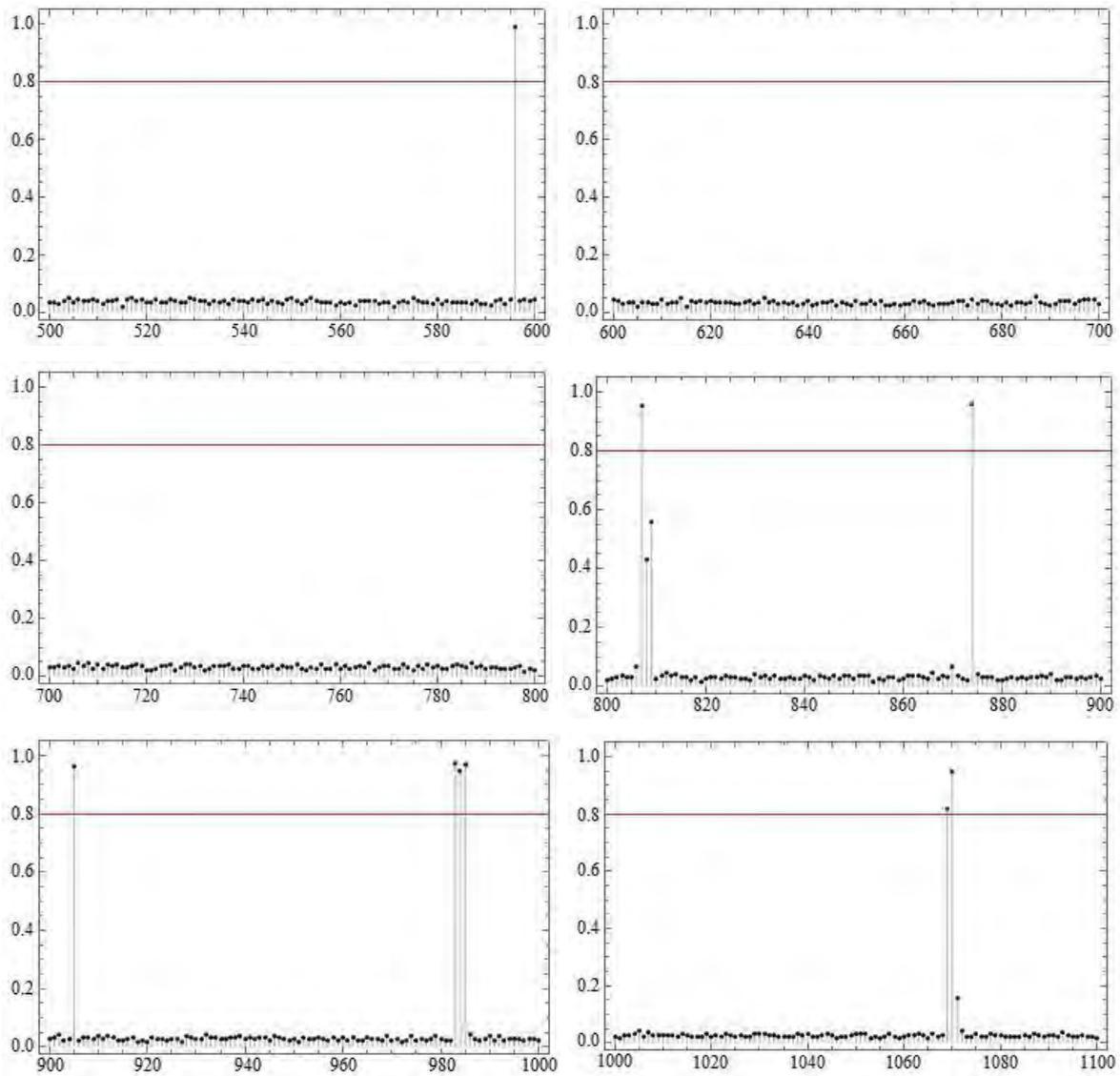
Figure 11a. Spherical cavity. Stick diagram of percentage basic energy versus number of mode highlighting the filtered SSA eigenmodes
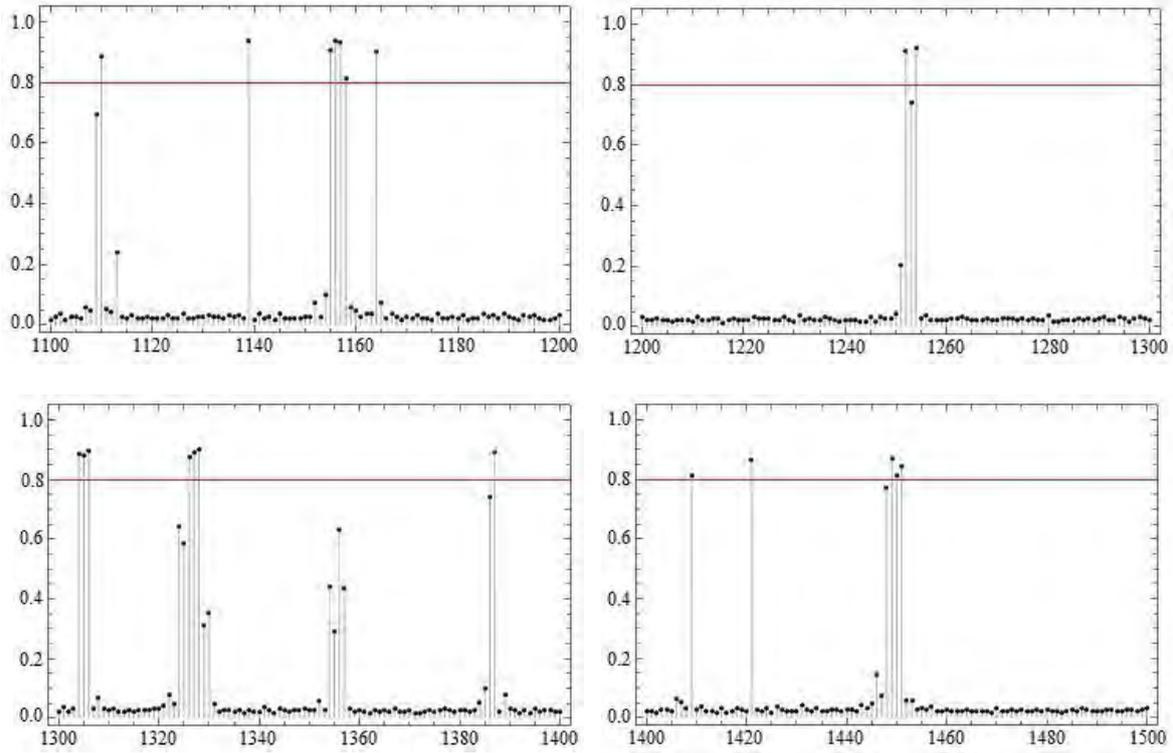
Figure 11b. Spherical cavity. Stick diagram of percentage basic energy versus number of mode highlighting the filtered SSA eigenmodes

Table 14. Spherical cavity. Exact TE and TM wave numbers respect to the radial direction (n = 1, 2, 3, …; m = 0, 1, 2, … , n; p = 1, 2, 3, …)

k_{mnp} (1/m)	Mode	k_{mnp} (1/m)	Mode	k_{mnp} (1/m)	Mode
274.37	TM_{m11}	821.09	TM_{m61}	1118.90	TM_{m52}
387.02	TM_{m21}	872.18	TM_{m32}	1139.10	TM_{m91}
449.34	TE_{m11}	909.50	TE_{m22}	1165.70	TE_{m71}
497.34	TM_{m31}	927.55	TM_{m71}	1170.49	TE_{m42}
576.35	TE_{m21}	931.66	TM_{m13}	1206.36	TM_{m33}
606.20	TM_{m41}	935.58	TE_{m51}	1232.29	TE_{m23}
611.68	TM_{m12}	996.75	TM_{m42}	1239.15	TM_{m62}
698.79	TE_{m31}	1033.53	TM_{m81}	1244.34	TM_{m101}
714.02	TM_{m51}	1041.71	TE_{m32}	1248.59	TM_{m14}
744.31	TM_{m22}	1051.30	TE_{m61}	1279.08	TE_{m81}
772.53	TE_{m12}	1071.30	TM_{m23}	1296.65	TE_{m52}
818.26	TE_{m41}	1090.41	TE_{m13}		

Similarly to the rectangular and coaxial cavity tests, from Table 10 to Table 13 and Table 14, it is also verified that the wave numbers of the filtered eigenmodes are an accurate approximation of the exact wave numbers. Then, the filtered eigenmodes will be physical eigenmodes captured with a precision so better as higher the percentage basic energy would be.

The numerical research has also revealed that the spurious eigenmodes polluting the finite element solution are generally characterized by its induced field having a small percentage of basic energy; therefore, the PE15 and TE10 elements included with the HE20 elements only

have slightly modified the behavior of the spurious eigenmodes displayed in a mesh fully made of HE20 elements.

4. Conclusions

For an electromagnetic domain discretized by energy-orthogonal finite elements, by a dispersion analysis, the behavior of the induced field energy has been researched for the fundamental solution of plane harmonic waves. Based on this analysis, for the finite element computation of the cavity eigenspectrum, it has been proposed:

A criterion based on the energy behavior to efficiently identify the physical eigenmodes between a myriad of spurious eigenmodes.

The use of a correlation between the percentage of higher order energy and the number of elements per wavelength in order to verify by this energy percentage if the physical eigenmodes computed have the precision required for applications.

References

[1] Harrington, R.F. (2001) *Time-Harmonic Electromagnetic Fields*, IEEE Press-Wiley.
[2] Silvester, P.P. and Ferrari, R.L. (1996) *Finite Elements for Electrical Engineers*, Cambridge University Press.
[3] Jin, J.M. (2014) *The Finite Element Method in Electromagnetics*, IEEE Press-Wiley.
[4] Felippa, C.A., Haugen, B. and Militello, C. (1995) From the individual element test to finite element templates: evolution of the patch test, *International Journal for Numerical Methods in Engineering* **38**, 199-229.
[5] Salazar-Palma, M. et al. (1998) *Iterative and Self-Adaptive Finite-Elements in Electromagnetic Modeling*, Artech House.
[6] Davidson, D.B. (2011) *Computational Electromagnetics for RF and Microwave Engineering*, Cambridge University Press.
[7] Brito, F.J. Filtering spurious eigenmodes in electromagnetic cavities discretized by energy-orthogonal twenty-nodes hexahedral finite elements, *Proceedings of 15th World Congress on Computational Mechanics (WCCM-XV) and 8th Asian Pacific Congress on Computational Mechanics (APCOM-VIII)*, Virtual Congress, Yokohama, Japan, 2022.
[8] Stratton, J.A. (1941) *Electromagnetic Theory*, McGraw-Hill.
[9] MacNeal, R.H. (1994) *Finite Elements: Their Design and Performance*, Marcel Dekker.
[10] Okrouhlík, M., Höschl, C. (1993) A contribution to the study of dispersive properties of one-dimensional lagrangian and hermitian elements, *Computers & Structures* **49**, 779-795.
[11] Horn, R.A. and Johnson, C.R. (1992) *Matrix Analysis*, Cambridge University Press.
[12] King, R.B. (1996) *Beyond the Quartic Equation*, Birkhäuser.
[13] Qu, Z.-Q. (2004) *Model Order Reduction Techniques: with Applications in Finite Element Analysis*, Springer.
[14] Brillouin, L. (2003) *Wave Propagation in Periodic Structures*, Dover Publications.
[15] Anderson, E. et al. (1999) *LAPACK User's Guide*, SIAM Press.
[16] Fernandes, P. and Raffetto, M. (2002) Characterization of spurious-free finite element methods in electromagnetics, *COMPEL – The International Journal for Computation and Mathematics in Electrical and Electronic Engineering* **21**, 147-164.
[17] Waldron, R.A. (1969) *Theory of Guided Electromagnetic Waves*, Van Nostrand Reinhold Company.
[18] Brito, F.J. and Militello, C. (2001) A modal error indicator based on a work-cancellation property of energy-orthogonal finite elements, *International Journal of Computational Engineering Science* **2**, 569-582.
[19] Gallagher, S. and Gallagher W.J. (1985) The spherical resonator, *IEEE Transactions on Nuclear Science* Vol. NS-32. No. **5**, 2980-2982.

Temperature-dependent interface properties of polypropylene/silicon oxide

†LinhuiHu[1], and †Lihong Liang[1]*

[1]College of Mechanicaland Electrical Engineering, Beijing University of Chemistry Technology, China.

*Presenting author:lianglh@mail.buct.edu.cn
†Corresponding author:lianglh@mail.buct.edu.cn

Abstract

Interfaces between polymer/ceramics have wide application in mechanical, electronic, and chemical engineering. For example, polypropylene (PP) is usually used as a separation between electrodes in batteries, and silicon oxide (SiO_2) is sometimes added on PP to improve the separation's properties; the interface properties between PP and SiO_2 determine the improving effect directly. Specifically, some parts (e. g. batteries) often service in a higher temperature above the room temperature due to various heat management problems and specific application environment, thus temperature-dependent interface properties attract much attention.

In this work, the interface intrinsic strength and fracture energy of PP/SiO_2 are firstly studied based on molecular dynamic tensile method at 1 K. Furthermore, the temperature effect on the interface properties is studied by increasing temperature to 100 K, 200 K, 300 K and 400 K, respectively. Interface damage is characterized quantitatively based on the interface free volume and the deformation of PP single chains. And the mechanism of temperature dependence is discovered combining with the damage characterization.

The results show that the interface strength of PP/SiO_2 is 0.257 GPa at the atomic scale at 1 K at tensile speed of 10 m/s, and the intrinsic fracture energy is 0.068 J/m^2. As the tensile temperature increases, both the interface strength and fracture energy decrease. At the room temperature, the interface strength is 0.151 GPa, and the intrinsic fracture energy is 0.063J/m^2.When the temperature is higher than the glass transition temperature 267 K of PP, the interface fracture behavior changes from a catastrophic fracture to a continuous fracture. The mechanism is related to the change of the damage. When the temperature rises from 1 K to 400 K, the interfacial free volume of PP increases by 11.6 times, and the increment of average gyration radius of all PP single chains increases by 6.6 times.

Keywords: polymer/ceramic interface, molecular dynamics, interface damage, interface strength, fracture energy

Introduction

Polymer/ceramic interface is widely used in mechanical [1][2], electronic [3][4] and chemical engineering [5][6] because of its excellent mechanical properties [7], barrier properties [8], electrical properties [9] and so on. Polypropylene (PP) is usually used as a separator between battery electrodes, and sometimes silicon oxide (SiO_2) is added to PP to improve the performance of the separator. For example, Afsharimani [3] *et al.* prepared polyvinyl alcohol (PVA) films with and without SiO_2 nanoparticles as polymer gate dielectric by spin coating method. Compared with pure PVA, PVA/SiO_2 films show larger capacitance, smaller hydrophilicity, rougher surface and considerable leakage current.

The interfacial properties between PP and SiO_2 directly determine the modification effect. Xu [10] *et al.* prepared polyamide composites containing surface-modified SiO_2 nanoparticles by

melt blending method, and studied the influence of interfacial layer structure on the thermal properties of the materials. The results show that the addition of surface modified SiO_2 nanoparticles will lead to the decrease of crystallization temperature and the increase of glass transition temperature of composites. Specifically, due to various thermal management problems and specific application environment, some components (e. g. batteries) often work at higher temperatures above room temperature, so the temperature-related interface properties have attracted widespread attention [11].

The rest of this paper is organized as follows: a numerical model of PP/SiO$_2$ interface is first established. Subsequently, the relationship between intrinsic interfacial stress and interfacial displacement under tensile load is calculated at 1 K, and the interface strength and interface fracture energy are obtained. The influence of temperature on interfacial properties is studied by raising the temperature to 100 K, 200 K, 300 K and 400 K respectively. The interface free volume and the evolution of typical PP chains near the interface are systematically analyzed and the damage mechanism of the interface is revealed. Finally, this paper gives the conclusion.

Model

The interface model of PP/SiO$_2$ was established as shown in Figure 1, based on the molecular dynamics method with the consistent valence force field [12], CLAYFF [13], the standard Lennard-Jones (L-J) potential [14] and Coulomb interaction. A uniaxial tensile load was applied to the PP surface along the z-direction at a series of temperatures between 1 K, 100 K, 200 K, 300 K and 400 K with tensile speed of 10 m/s to study the temperature effect of the interface properties.
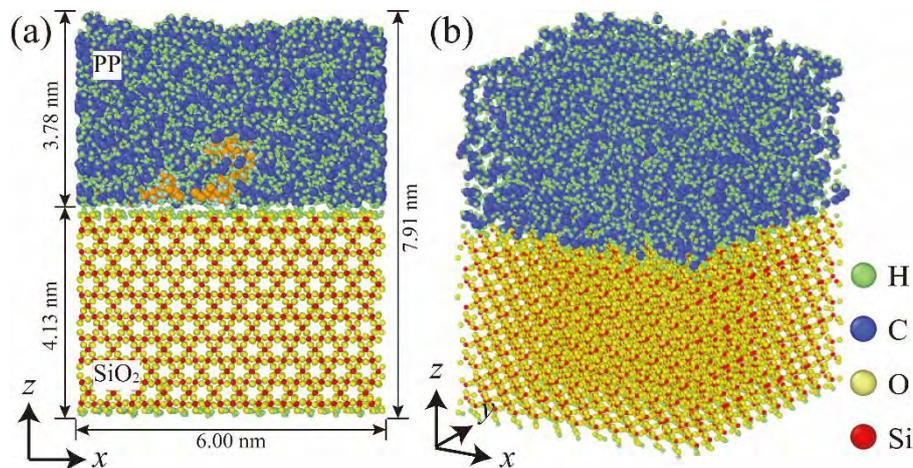
Figure 1. Interface model of PP/SiO$_2$

Results and discussions

Figure 2(a) shows the interface stress-displacement relations at different temperatures. The maximum interface stress corresponds to the interface strength, and the area under the curve corresponds to the intrinsic fracture energy of the interface. It can be seen that the interface strength decreases with increasing temperature, as shown in Figure 2(b), at the same time, the fracture energy also decreases with increasing temperature.

The relationship between the temperature and interface properties can be described by the exponential functions [15]:

$$\delta_{\alpha} = \delta_{\alpha}(T_0) \times \left(\frac{T_m - T}{T_m - T_0} \right)^{D_\delta} \qquad \text{\textbackslash* MERGEFORMAT (1)}$$

$$\Delta W = \Delta W(T_0) \times \left(\frac{T_m - T}{T_m - T_0} \right)^{D_{\Delta W}} \qquad \text{\textbackslash* MERGEFORMAT (2)}$$

where δ_{α} (T_0) and ΔW (T_0) are the interfacial stress and fracture energy at reference temperature T_0 (1 K), respectively; and T_m and T are the melting temperature and temperature, respectively. D_δ and $D_{\Delta W}$ are material constants fitted from the simulation results, where $D_\delta =$ 0.084 and $D_{\Delta W} = 0.035$. It can be seen from Figure 2(b) that the simulation results are in agreement with the predictions based on the equations.
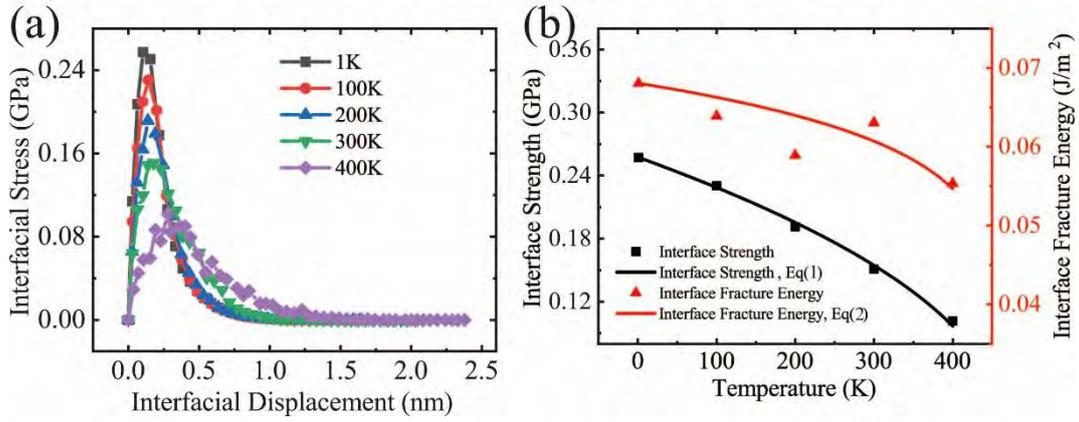
Figure 2. Effect of temperature on interfacial properties of PP/SiO₂ interface. (a) Interfacial stress–interfacial displacement curves at various temperatures; (b) variation of interface strength and fracture energy with temperature.

In order to understand the difference of interface properties, the interface damage was quantitatively characterized by the interface free volume of PP near interface. The dimensionless interface free volume parameter C_0 is expressed as

$$C_0 = \frac{V - V_0}{V} \qquad (3)$$

where V and V_0 represent the total volume and atomic occupied volume of PP. Figure 3(a) shows the interface free volume evolution at different temperatures. It is clear that when the temperature is higher than 200 K, the free volume increases obviously with tension. At the same time, the interface fracture is sudden after the interface strength is reached at 1 K and interface damage is faster, but the interface damage and fracture are continuous at 400 K as shown in Figure 3(b), different fracture modes can be found.
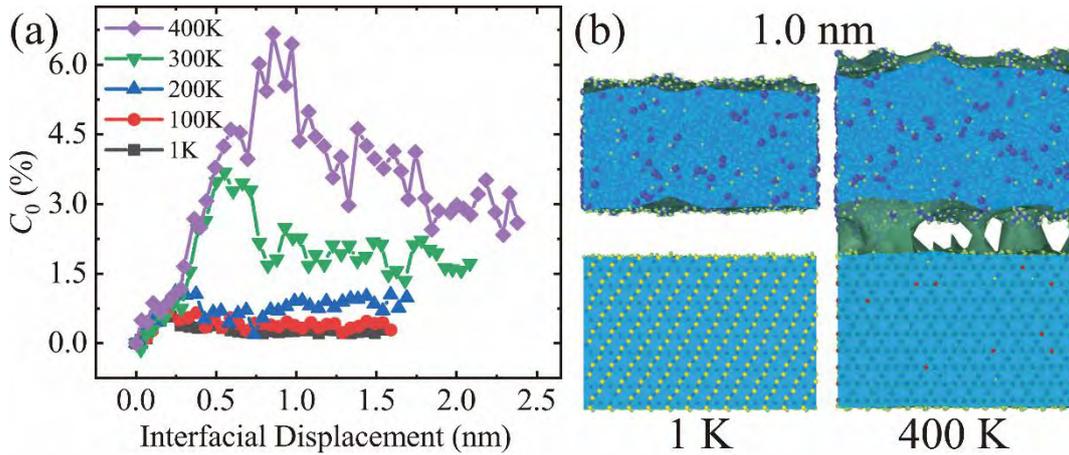
Figure 3. (a) Variation of interface free volume with interfacial displacement at different temperatures. (b) typical snapshots of the cavity in the PP/SiO$_2$ interface at the same interfacial displacement of 1.0 nm at temperatures of 1 and 400 K, respectively.

Furthermore, the average gyration radius evolution of PP single chains was studied. Figure 4(a) shows the change of average gyration radius of all PP chains at different temperatures. It can be seen that the maximum value of the average radius of gyration increases with increasing temperature, as shown in Figure 4(b). When the temperature is higher than 200 K, the average radius of gyration increases more obviously with the increase of temperature.
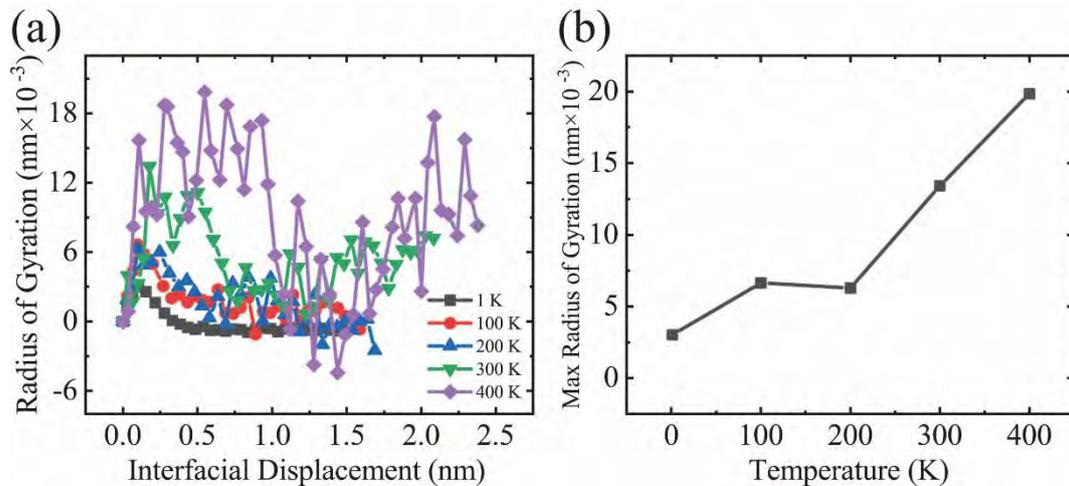
Figure 4. (a) Variation of average gyration radius with interfacial displacement at different temperatures. (b) Variation of peak value of average radius of gyration with temperature.

Conclusions

In this paper, through the tensile calculation of PP/SiO$_2$ micro-interface at different temperatures, the interface strength and the interface fracture energy are obtained. According to the evolution of polymer free volume and average radius of gyration of PP chains, the quantitative characterization of interfacial damage process is given.

With the increase of temperature from 1 K to 400 K, the interface strength and interface fracture energy of PP/SiO$_2$ decreases, the interface strength decreases from 0.257 GPa to 0.102 GPa, and the interface fracture energy decreases from 0.068 J/m^2 to 0.055 J/m^2. The free volume of PP and the average radius of gyration of PP chains increase with increasing

temperature. When the temperature is higher than the glass transition temperature of PP (267 K), the interfacial fracture behavior changes from abrupt fracture to continuous fracture, and the free volume and average radius of gyration increase obviously with tension.

References

[1] Liu, X., Zhou, X., Kuang, F., Zuo, H., Huang, J. (2021) Mechanical and Tribological Properties of Nitrile Rubber Reinforced by Nano-SiO2: Molecular Dynamics Simulation, *Tribol. Lett.* **69**, 54.

[2] Samant, A. N., Dahotre, N. B. (2009) Laser Machining of Structural Ceramics-a Review, *J. Eur. Ceram. Soc.* **29**, 969-993.

[3] Afsharimani, N., Nysten, B. (2019) Hybrid Gate Dielectrics: A Comparative Study between Polyvinyl Alcohol/SiO2 Nanocomposite and Pure Polyvinyl Alcohol Thin-Film Transistors, *Bulletin of Materials Science* **42**, 26.

[4] Muller, D., Pinheiro, G. K., Bendo, T., Gutiérrez Aguayo, A. J., Barra, G. M. O., Rambo, C. R. (2015) Synthesis of Conductive Ppy/SiO2 Aerogels Nanocomposites by in Situ Polymerization of Pyrrole, *Journal of Nanomaterials* **2015**, 658476.

[5] Paggioli, K. (2006) Centrifugally Cast Fiberglass-Reinforced Polymer Mortar Pipe for a Large-Diameter Interceptor Sewer, *Mater. Performance* **45**, 58-61.

[6] Horowitz, Y., Lifshitz, M., Greenbaum, A., Feldman, Y., Greenbaum, S., Sokolov, A. P., Golodnitsky, D. (2020) Review-Polymer/Ceramic Interface Barriers: The Fundamental Challenge for Advancing Composite Solid Electrolytes for Li-Ion Batteries, *J. Electrochem. Soc.* **167**, 160514.

[7] Arulvel, S., Reddy, D. M., Rufuss, D. D. W., Akinaga, T. (2021) A Comprehensive Review on Mechanical and Surface Characteristics of Compositesreinforced with Coated Fibres, *Surf. Interfaces* **27**, 101449.

[8] Hong, H., Kim, J. U., Kim, T. I. (2017) Effective Assembly of Nano-Ceramic Materials for High and Anisotropic Thermal Conductivity in a Polymer Composite, *Polymers* **9**, 413.

[9] Musa, F. N., Bashir, N., Ahmad, M. H., Buntat, Z., Piah, M. A. M. (2016) Investigating the Influence of Plasma-Treated SiO2 Nanofillers on the Electrical Treeing Performance of Silicone-Rubber, *Applied Sciences-Basel* **6**, 348.

[10] Xu, X., Li, B., Lu, H., Zhang, Z., Wang, H. (2007) The Interface Structure of Nano-SiO2/Pa66 Composites and Its Influence on Material's Mechanical and Thermal Properties, *Appl. Surf. Sci.* **254**, 1456-1462.

[11] Tsagkalias, I. S., Loukidi, A., Chatzimichailidou, S., Salmas, C. E., Giannakas, A. E., Achilias, D. S. (2021) Effect of Na- and Organo-Modified Montmorillonite/Essential Oil Nanohybrids on the Kinetics of the in Situ Radical Polymerization of Styrene, *Nanomaterials (Basel, Switzerland)* **11**, 474.

[12] Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Hagler, A. T. J. P.-s. F., Bioinformatics (1988) Structure and Energetics of Ligand Binding to Proteins: Escherichia Coli Dihydrofolate Reductase-Trimethoprim, a Drug-Receptor System, *Proteins* **4**, 31-47.

[13] Cygan, R. T., Liang, J.-J., Kalinichev, A. G. (2004) Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field, *The Journal of Physical Chemistry B* **108**, 1255-1266.

[14] Zojaji, M., Hydarinasab, A., Hashemabadi, S. H., Mehranpour, M. (2021) Rheological Behaviour of Shear Thickening Fluid of Graphene Oxide and Sio2 Polyethylene Glycol 400-Based Fluid with Molecular Dynamic Simulation. *Mol. Simulat.* **47**, 317-325.

[15] Zhang, M., Jiang, B., Chen, C., Drummer, D., Zhai, Z. (2019) The Effect of Temperature and Strain Rate on the Interfacial Behavior of Glass Fiber Reinforced Polypropylene Composites: A Molecular Dynamics Study, *Polymers* **11**, 1766.

Stochastic seismic response analysis of subway station structure based on optimization point selection strategy

*Yifan Fan[1,2], †Zhiyi Chen[1], and Zhiqian Liu[1]

[1]Department of Geotechnical Engineering, Tongji University, Shanghai, China
[2]Department of Architectural and Civil Engineering, City University of Hong Kong, Hong Kong, China

*Presenting author: yifan_fan@tongji.edu.cn
†Corresponding author: zhiyichen@tongji.edu.cn

Abstract

At present, researchers have recognized that the uncertainty of the ground motion and the soil property needs to be considered in the seismic analysis of underground structures. The interaction mechanism between soil and underground structures is very complex, and using refined numerical models for dynamic response analysis is an effective numerical simulation analysis method. However, due to the highly nonlinear characteristics of soil and structure, the calculation cost will significantly increase with the increase of random variables. In view of this, this paper first establishes a finite element model and a random excitation model for the soil and subway station structure system. Subsequently, a high-dimensional space optimization point selection strategy was used to couple random ground motion variables with random soil shear wave velocity variables, and a representative point set was obtained based on the idea of minimizing Generalized F-discrepancy (GF-discrepancy). Finally, the nonlinear time history analysis (NLTHA) was performed on the representative point set. The research results indicate that the influence of random excitation on the seismic response of subway station structures cannot be ignored, and additional random shear wave velocity will increase the lateral deformation of subway station structures and reduce their dynamic reliability. This study provides a reference for seismic analysis and design of underground structures.

Keywords: Subway station structure; Stochastic seismic response; Optimized point selection strategy; Seismic reliability; Seismic analysis and design

Introduction

Underground structures have become an important component of urban infrastructure. Underground structures, represented by subway stations and tunnels, serve as key hubs for urban lifelines. Underground structures exist in soil, making it difficult to repair and rebuild after earthquakes, and weak links are not easily detected. Therefore, a comprehensive analysis of the seismic response and seismic performance of underground structures is crucial for ensuring urban safety and improving urban resilience. The research on seismic analysis and design of underground structures started relatively late but has achieved relatively rich results [1-6]. With the advancement of research and technology, the research on the seismic analysis of underground structures has gone through several stages: static analysis, pseudo-static analysis, and dynamic time history analysis. At the same time, researchers also recognized that underground structures face uncertainty in ground motions and soil properties during earthquakes [7-12]. It should be emphasized that underground structures are mainly controlled by the deformation of surrounding soil during earthquakes. Whether it is the randomness of ground motions or the randomness of soil properties, the order of action is to first affect the response of the site, and then affect the seismic response of underground structures through

the dynamic interaction between soil and structure. On the one hand, the randomness of the ground motion is the main factor affecting the seismic response of underground structures. Chen and Liu [8] studied the evolution law of seismic response of a multi-story subway station structure under stochastic ground motions and verified it using shaking table tests. Hu et al. [9] adopted the probability density evolution method to study the seismic dynamic reliability of the slope with variable soil properties under random ground motions. Their conclusion confirms that the spatial variability of soil properties has adverse effects on seismic analysis. Feng et al. [10] showed that compared to uniform seismic excitation, non-uniform random seismic excitation could lead to greater internal forces and seismic energy consumption in tunnels. On the other hand, due to the heterogeneity of soil layers and the continuity of geological evolution, the randomness of the soil property also has a significant impact on the seismic response of underground structures. The research results of Li et al. [11] confirm that the spatial variability of soil strength parameters cannot be ignored in the seismic reliability evaluation of the slope. He et al. [12] incorporated the spatial variability of soil properties into the evaluation framework for the seismic reliability of underground structures. However, according to the above analysis, current research on the stochastic seismic response of underground structures mainly focuses on two aspects: (1) seismic response analysis of complex structures under a single random excitation; (2) Seismic response analysis of simple structures under multiple random excitations. Unlike underground engineering structures such as slopes or circular tunnels, subway station structures have both complex structural forms and highly nonlinear characteristics. The changes in site response have a more significant disturbance on the seismic response of subway station structures. The seismic resistance of subway station structures provides important support for ensuring the safety of urban personnel and facilities. Therefore, it is urgent to conduct the seismic response analysis of subway station structures under multiple random excitations.

The use of refined numerical models is an effective numerical simulation method for analyzing the seismic dynamic response of subway station structures. However, the nonlinearity of material and dynamic interactions greatly increases the cost of NLTHA of the soil and subway station structure system. Specifically, when randomness is included in seismic analysis, the number of examples for NLTHA will increase exponentially with the increase of random variables. Therefore, achieving a balance between calculation cost and accuracy is the foundation for solving the problem of stochastic seismic response analysis of subway station structures. The Monte Carlo simulation method is a universal method for selecting sample points in stochastic analysis problems. However, the drawback of this method is that it requires too many sample points, making it difficult to reduce the computational cost in this study. Some improved methods have been proposed successively, such as the Latin hypercube sampling method [13] and the tangent sphere point selection method [14]. However, when the number of random variables is large, the uniformity of the representative point set obtained based on these methods is poor. In order to solve the above problems, Yang et al. [15] proposed a method called the iterative screening-rearrangement method (IS-RAM) based on the idea of GF-discrepancy minimization. This method achieves the reduction of GF-discrepancy and the increment of assigned probability's uniformity by adjusting the threshold coefficient. This study intends to use this method to select the representative point set in high-dimensional space.

The novelty of this study is as follows: (1) The refined numerical model and random excitation model for the soil and subway station structure system have been established; (2) The optimized point selection strategy is used to reduce the size of the sample point set and the computational cost of stochastic analysis problems; (3) The seismic response and seismic reliability of the subway station structure under single random excitation and combined random excitations are analyzed.

The structure of this paper is as follows. The first section is the research background and significance. The second section introduces the optimized point selection strategy for high-dimensional space. The third section presents the method for establishing the random excitation model. In the fourth section, a typical two-story, three-span subway station structure in Shanghai was used as a case study to analyze the stochastic seismic response of the subway station structure. The main conclusions of this study are presented in the fifth Section.

Optimization point selection strategy

Conception of GF-discrepancy

Chen and Chan [16] proved that reducing the GF-discrepancy of the representative point set can improve the uniformity of the representative set and the accuracy of the generalized probability density evolution equation. In this study, GF-discrepancy is introduced as a criterion to measure the uniformity of the representative point set. The smaller the GF-discrepancy, the better the uniformity of the point set.
GF-discrepancy is defined as follows [16-17]:

$$D_{GF}(M_{nsel}) = \max_{1 \le i \le s} \left\{ D_{F,i}(M_{nsel}) \right\} \tag{1}$$

$$D_{F,i}(M_{nsel}) = \sup_{x \in R} \left| \tilde{F}_{n,i}(x) - F_i(x) \right| \tag{2}$$

Where $D_{F,i}(M_{nsel})$ represents the marginal deviation of the *i-th* random variable; M_{nsel} is the point set related to serial number *n*; $F_i(x)$ is the marginal CDF considering the assigned probability.

$$\tilde{F}_{n,i}(x) = \sum_{q=1}^{n} P_q \cdot I \left\{ x_{q,i} \le x \right\} \tag{3}$$

Where $X_{q,i}$ is the *i-th* component of x_q, and P_q is the assigned probability corresponding to point x_q.

Implementation steps of IS-RAM

Yang et al. [15] proposed the IS-RAM method to reduce the non-uniformity of the representative set. The idea of this method is to use the threshold coefficient *b* and iterative algorithm to filter representative point sets. The implementation steps are as follows:
(1) Using the Sobol sequence sampling method to generate an initial point set of s-dimensional uniform distribution, which is the cumulative distribution function value (CDF) distributed between (0,1);
(2) Using CDF inverse transformation to obtain a new point set and calculating the assigned probability of the new point set;
(3) Delete the points whose assigned probability is less than *b/n* (*n* denotes the number of current sample points), and calculate the assigned probability for the filtered point set again;
(4) Repeat step (3) until all sample points' assigned probability exceeds the threshold or the number of sample points meets the prescribed target value;
(5) Rearrange the sample points according to regulations to obtain the final representative point sets.

Sources of the random excitation

Non-stationary stochastic ground motion model

Liu et al. [18] proposed a non-stationary random ground motion generation method that reduces the number of random variables to 1-2. The parameters of this method are derived from China's Code for Seismic Design of Buildings [19]. The random ground motion model is as follows:
In the spectral representation method, the real non-stationary random ground motion with zero mean value can be expressed as [18]:

$$\ddot{X}_g(t) = \sum_{k=1}^{N} \sqrt{2 S_{\ddot{X}_g}(t, \omega_k) \Delta\omega} \left[\cos(\omega_k) X_k + \sin(\omega_k) Y_k \right] \tag{4}$$

$$\omega_k = k \Delta\omega \tag{5}$$

Where $\{X_k, Y_k\}$ (1,2,3…N) is a standard orthogonal random variable, $S_{\ddot{X}_g}(t, \omega_k)$ is the improved evolutionary power spectral density function of the non-stationary seismic random process $\ddot{X}_g(t)$, $\Delta\omega = 0.2 rad / s$.

Constructing the expression of random function $\{X_k, Y_k\}$ (1,2,3…N) based on the idea of the random function, based on the assumption that any two sets of orthonormal random variables are respectively functions of the fundamental random variable θ, the random function can be further written as follows:

$$X_k = \sqrt{2} \cos(k\theta + \frac{\pi}{4}) \tag{6}$$

$$Y_k = \sqrt{2} \sin(k\theta + \frac{\pi}{4}) \tag{7}$$

Where, θ is a basic random variable uniformly distributed in [-π, π].
Based on the Clough-Penzien stationary motion spectrum model [20], Deodatis [21] proposed an evolutionary power spectrum model for non-stationary motion, which was subsequently improved by Cacciola and Deodatis [22]. The evolutionary power spectral density function can be written as [22]:

$$S_{\ddot{X}_g}(t, \omega_k) = A^2(t) \bullet \frac{\omega_g^4(t) + 4\xi_g^2(t)\omega_g^2(t)\omega^2}{\left[\omega_g^2 - \omega_g^2(t) \right]^2 + 4\xi_g^2(t)\omega_g^2(t)\omega^2}$$

$$\bullet \frac{\omega^4}{\left[\omega^2 - \omega_f^2(t) \right]^2 + 4\xi_f^2(t)\omega_f^2(t)\omega^2} \tag{8}$$

$$A^2(t) = \left[\frac{t}{c} \exp\left(1 - \frac{t}{c} \right) \right]^d \tag{9}$$

Where, $S_0(t)$ is the spectral intensity factor, $A(t)$ is the intensity modulation function, c is the peak time, and d is the shape of $A(t)$. The parameters controlling the non-stationary characteristics of ground motion are listed in Eqs. (10), (11), and (12):

$$\omega_t(t) = \omega_0 - a\frac{t}{T}, \xi_g(t) = \xi_0 + b\frac{t}{T} \tag{10}$$

$$\omega_f(t) = 0.1\omega_g(t), \xi_f(t) = \xi_g(t) \tag{11}$$

$$\omega_0(t) = \omega_g + \frac{a}{2}, \xi_0 = \xi_g - \frac{b}{2} \tag{12}$$

Where, $\omega_g(t)$ and $\xi_g(t)$ are the excellent circle frequency and damping ratio of the site,

respectively, and $\omega_f(t)$ and $\xi_f(t)$ are the corresponding filtering parameters; The four parameters ω_0, ξ_0, a, and b can be determined by site classification.

The parameter $S_0(t)$ reflecting the earthquake intensity can be calculated by the following Equation:

$$S_0(t) = \frac{2\overline{a}_{max}^2}{\overline{r}^2 \pi \omega_g(t)\left(2\xi_g(t) + \dfrac{1}{2\xi_g(t)}\right)} \tag{13}$$

Where, $\overline{a}_{max}^2$ represents the mean value of PGA, and $\overline{r}$ represents the effective peak factor (which can be determined by best fitting the seismic response spectrum).

In Eq. (13), $\overline{a}_{max}^2$ and $\overline{r}$ reflect ground motion intensity, and ω_0, ξ_0, a, and b represent non-stationary frequency characteristics, c and d reflect the characteristics of earthquake intensity, and T is the earthquake duration.

The Toro randomization shear wave velocity model

Toro [23] further explained a random shear wave velocity (*Vs*) profile model in recent research to characterize the uncertainty of site response. The Toro randomization shear wave velocity model is as follows:

Assuming that V_s obeys lognormal distribution at any depth, the V_s of layer i can be expressed as follows:

$$V_s(i) = \exp\left\{\ln\left[V_{s,m}(i)\right] + X_i \cdot \sigma_{\ln V_s}\right\} \tag{14}$$

Among them, $V_{s,m}(i)$ is the mean value of $V_s(i)$, $\sigma_{\ln V_s}$ is the natural logarithm of the standard deviation of $V_s(i)$; both are calculated according to the data in Table 2; X_i (Eq. (15)) is a standard normal distribution random variable, which represents the standard deviation number between $ln[V_s(i)]$ and $ln[V_{s,m}(i)]$ in the logarithmic space.

$$X_i = \begin{cases} \varepsilon_i & i = 1 \\ \rho_{IL} \cdot X_{i-1} + \varepsilon_i \cdot \sqrt{1 - \rho_{IL}^2} & i \geq 2 \end{cases} \tag{15}$$

ε_i is a standard normal distribution random variable with a mean value of 0 and a standard deviation of 1; ρ_{IL} is the correlation coefficient between layer i and layer $i-1$, and it can be expressed as a function of depth d and layer thickness t, as shown in Eq. (16):

$$\rho_{IL}(d,t) = \left[1 - \rho_d(d)\right]\rho_t(t) + \rho_d(d) \tag{16}$$

Where $\rho_d(d)$ and $\rho_t(t)$ represent depth correlation coefficient and layer thickness correlation coefficient, respectively, their definitions are shown in Eqs. (17) and (18):

$$\rho_d(d) = \begin{cases} \rho_{200}\left[\dfrac{d + d_0}{200 + d_0}\right]^b & d \leq 200m \\ \rho_{200} & d > 200m \end{cases} \tag{17}$$

$$\rho_t(t) = \rho_0 \exp\left(\frac{-t}{\Delta}\right) \tag{18}$$

Among them, ρ_{200}, d_0, b, ρ_0, and Δ are all related parameters of the model, which are determined according to the average shear wave velocity V_{s30} in the field within 30m depth. The values of the above five parameters are ρ_{200}=0.98, d_0=0, b=0.344, ρ_0=0.99, Δ=3.9,

respectively.

Case study

Numerical model

As shown in Figure 1, the research object in this paper is a typical two-story, three-span subway station structure in Shanghai. The height of the structure is 12.37m and the width is 20.90m. The total width of the soil and subway station structural system is 230m, and the total height is 80m. To ensure consistency in lateral deformation, multi-point constraints are used on both sides of the soil. The bottom of the soil is constrained by the hinge support constraint. The soil and structure are constrained by friction with a hard contact surface, with a friction coefficient of 0.4. The subway station structure and soil are simulated by CPE4 element and CPE4R element, respectively. The rebar is simulated by T2D2 element.
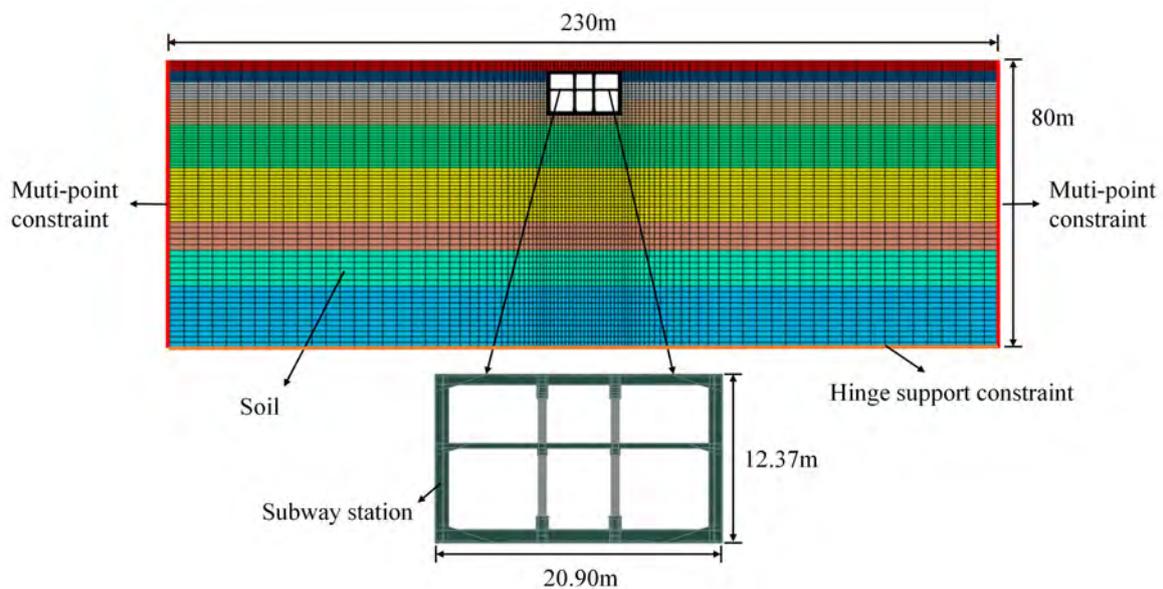
Figure 1. Numerical model

The material of the columns in the structure is C45 concrete, while the other components are C35 concrete. The concrete material is simulated by a plastic damage constitutive model (PDCM). The rebar's mechanical behavior is simulated by the idealized elastoplastic constitutive model (IEPCM). The soil mass's constitutive model follows Mohr-Coulomb yield criterion. Table 1 shows the physical and mechanical parameters of steel bars and concrete. Table 2 shows the physical and mechanical parameters of soil.

Table 1. Physical and mechanical parameters of concrete and rebar

Material	Constitutive model	Parameter type	Parameter value
Concrete	PDCM	Mass density (kg·m^{-3})	2500
		Poisson's ratio	0.2
		Elastic modulus (MPa)	31500/33500 (C35/C45)
Rebar	IEPCM	Mass density (kg·m^{-3})	7800
		Poisson's ratio	0.3
		Elastic modulus (GPa)	200
		Yield stress (MPa)	400

Table 2. Physical and mechanical parameters of soil layer

Soil layer	Soil depth (m)	Shear wave velocity (m/s)	Internal friction angle (°) / Cohesion (kPa)	Mass density (kg•m-3)
S1	3	101	25.0 / 16.9	1920
S2	6	130	11.2 / 11.0	1670
S3	11	292	16.4 / 44.0	1950
S4	18	278	31.1 / 2.0	1820
S5	30	231	15.2 / 17.0	1770
S6	45	279	28.1 / 8.0	1840
S7	53	380	31.7 / 5.0	1880
S8	63	446	35.2 / 2.0	1890
S9	80	560	42.6 / 1.0	2000

Representative point set and random excitation in this study

According to the random excitation model provided in Section 3 and the research case in this section, this study involves ten basic random variables, namely the shear wave velocity basic random variables of nine soil layers and one phase angle basic random variable that generates random ground motions. Using the optimized point selection method in Section 2, select the representative point set in a ten-dimensional space. Figure 2 shows the threshold coefficient b and the corresponding representative point set. Overall, the GF-discrepancy of the representative point set increases as the number of sample points decreases. The NLTHA cost for a single sample is around four hours. In order to achieve a compromise between calculation accuracy and cost, the size of the representative point set should be controlled within 400. The number of the preliminarily feasible representative point set is 384 (b=0.8), 274 (b=0.9), and 394 (b=1.0), respectively. However, when the number of sample points is

274, the GF-discrepancy of the representative point set is relatively large. In order to further compare the effectiveness of representative point sets with threshold coefficients of 0.8 and 0.9, Figure 3 shows the distribution of assigned probability corresponding to the two representative point sets. From Figure 3, when the threshold coefficient is 0.8, the assigned probability distribution of the representative point set is more concentrated, and the uniformity of the point set is also better. Therefore, the number of samples selected for the representative point set in this study is 384, with a threshold coefficient of b=0.8.
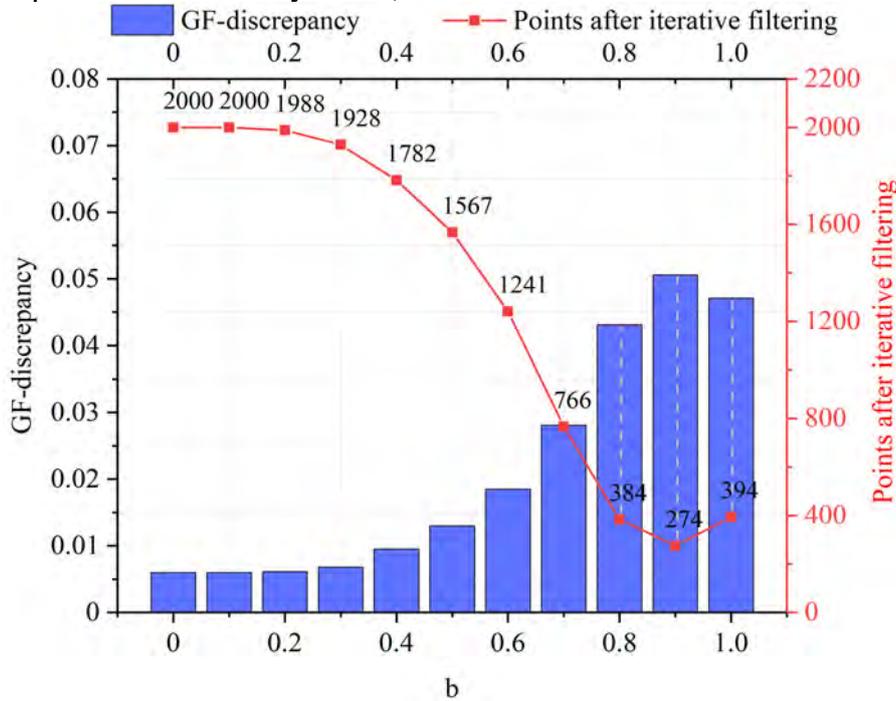
Figure 2. Iterative screening process for the representative point set
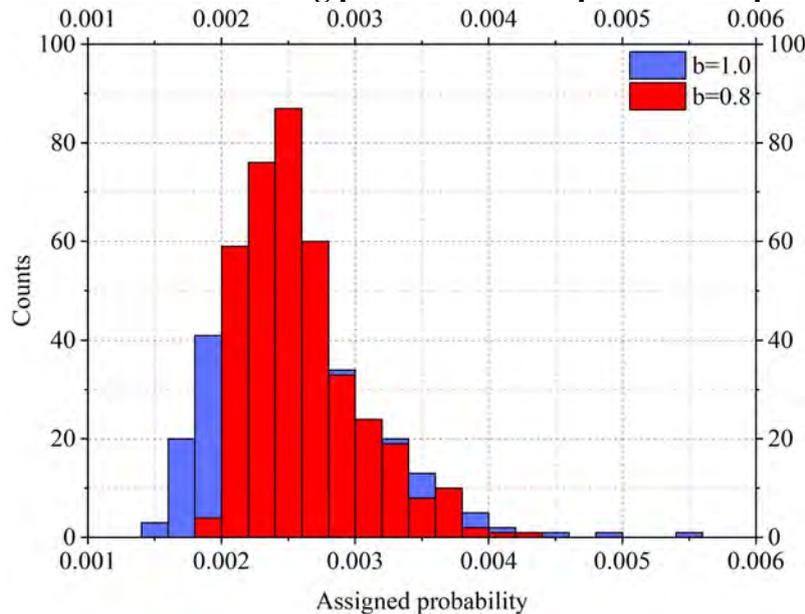
Figure 3. Assigned probability of the representative point set corresponding to different threshold coefficients

In this study, the seismic fortification intensity of the structure is 7 degrees, and the site class is IV. The maximum value of horizontal seismic impact coefficient αmax is 0.08, and the characteristic period value is 0.9s. The parameters used to generate random ground motions in this paper are as follows [19]: ω_0=11s-1, ξ_0=0.7, a=6s-1, b=0.15, c=7s, d=2, $\bar{r}$=2.65, T=30s.

Figure 4 shows a typical ground motion sample. Figure 5 shows three typical shear wave velocity profile samples.
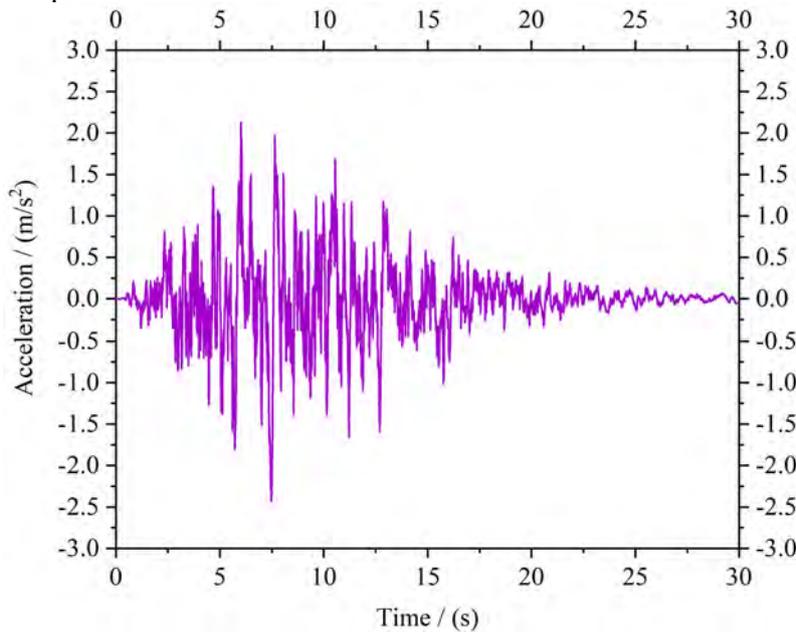
Figure 4. Typical ground motion sample
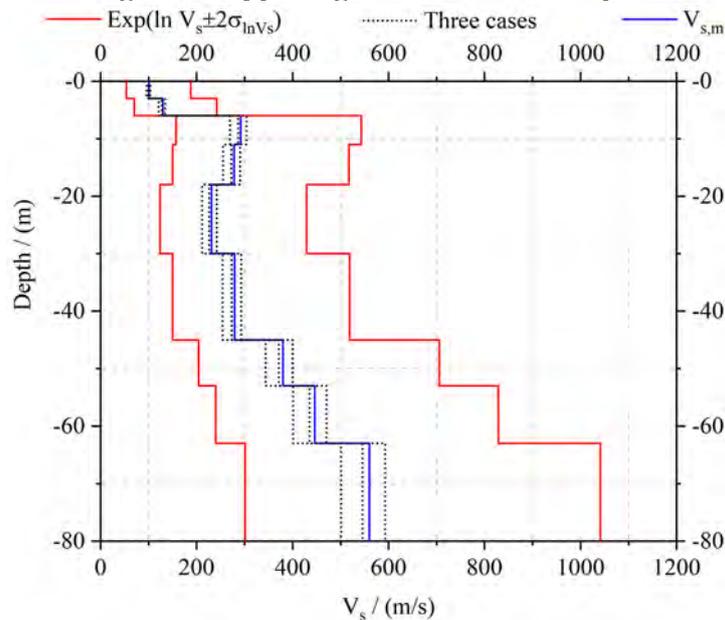
Figure 5. Typical shear wave velocity profile sample

Stochastic seismic response analysis

This section analyzes the seismic response of the subway station structure under random ground motion (Under RGM) and the combination of random ground motion and random shear wave velocity (Under RGM&RSWV). The inter-story drift ratio (IDR) is an effective performance indicator for evaluating the seismic response of underground rectangular frame structures. Table 3 provides the IDR's limit value specified in the Chinese seismic code for underground structures [24-25]. Perform NLTHA on representative point sets under two working conditions, and extract samples' IDR to characterize the seismic response level of subway station structures. Figure 6 shows the distribution of the subway station structure's seismic response under two working conditions. Under random ground motion excitation, the distribution range of IDR is between 1.56×10^{-3} and 3.22×10^{-3}. Under the combined random

excitation, the distribution range of IDR is between 2.61×10^{-4} and 6.50×10^{-3}. The interval length of the latter is approximately 3.76 times that of the former. This means that the seismic response information of subway station structures under combined random excitations is more abundant, and the seismic risk is also greater.

Table 3. The limit value of IDR in Chinese codes

Code and standard	Applicable object	Elastic state	Elastic-plastic state
Code for Seismic Design of Subway Structures (DG/TJ08-2064-2009)	Concrete frame structure	1/550 (0.0018)	Not Applicable
	Reinforced concrete frame structure	Not Applicable	1/250 (0.0040)
Standard for seismic design of underground structures (GB/T51336-2018)	Single-storey or double-story rectangular section structure	1/550 (0.0018)	1/250 (0.0040)
	Three-story and above rectangular section structure	1/1000 (0.0001)	1/250 (0.0040)
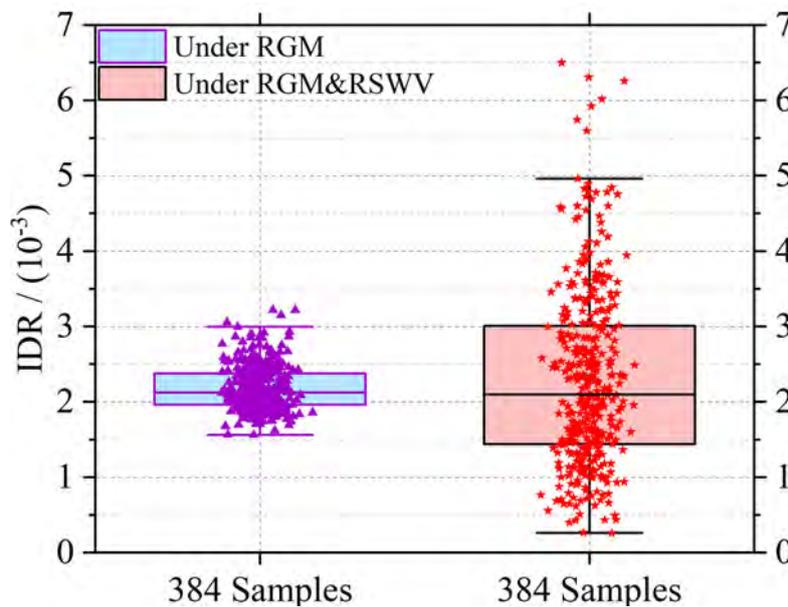
Figure 6. Distribution of IDR

Figure 7 shows the probability density function (PDF) of IDR. Under the combined random excitation, the seismic response distribution of the structure is more in line with the actual situation. This is conducive to a comprehensive analysis of the potential seismic risks of the subway station structure. According to Table 3, taking the IDR response of the structure as the seismic performance index, the probability of not exceeding the elastic or elastic-plastic state limit is defined as seismic reliability. Figure 8 shows the cumulative distribution function (CDF) of IDR, which can be used to mark the seismic reliability of the subway station structure under two working conditions. Under random seismic excitation, the elastic and elastic-plastic seismic reliability of the structure is 0.0705 and 1, respectively, without entering a plastic failure state. Under the combined random excitation, the elastic-plastic seismic reliability of the structure is 0.3916 and 1, respectively. According to the above analysis, if the randomness of soil properties is not considered, the seismic risk of the subway station will be underestimated.
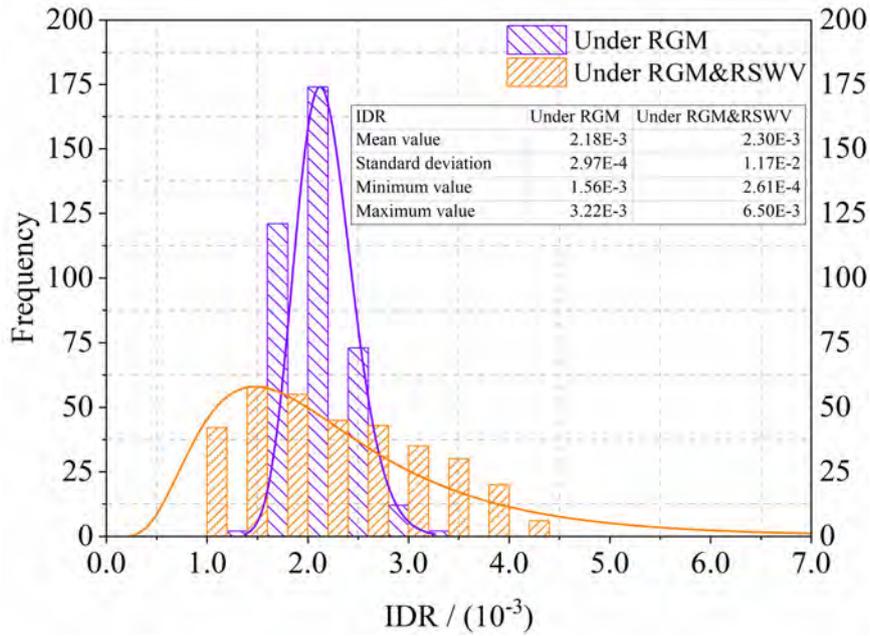
Figure 7. PDF of IDR
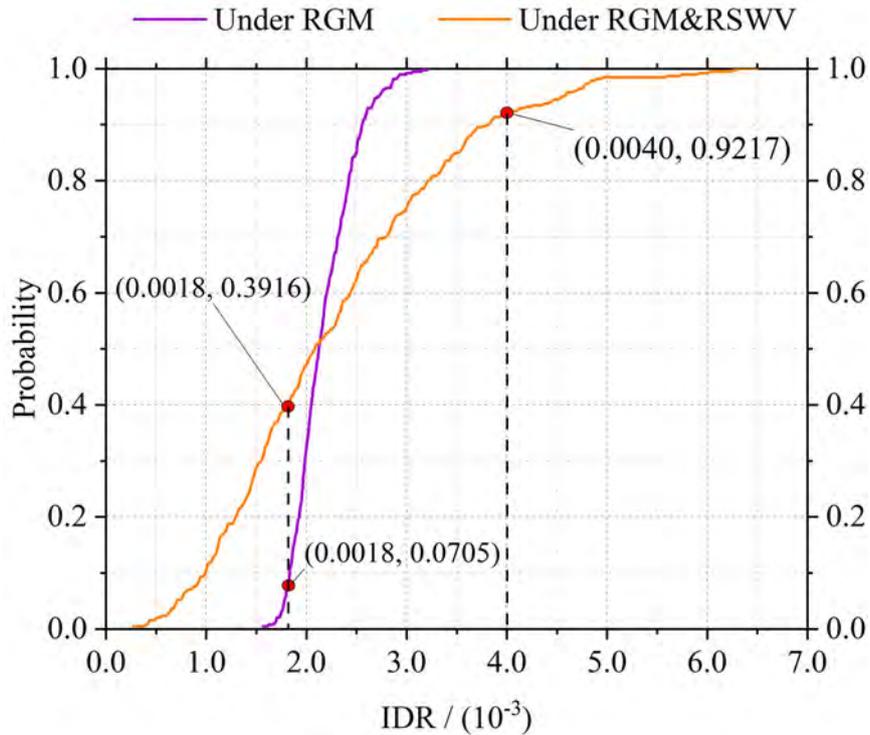
Figure 8. Seismic reliability of subway station structure

Conclusions

This study constructs the refined numerical model and the random excitation model for soil and subway station structure. The high-dimensional spaced optimized point selection method was used to couple the random seismic motion variable with the random soil shear wave velocity variable, and a representative point set for NLTHA was selected. The stochastic seismic response of the subway station structure under RGM excitation and RGM&RSWV were analyzed. The main conclusions of this study are as follows:

(1) IS-RAM significantly reduces the scale of the representative point set and the computational cost of NLTHA. In this study, the threshold coefficient *b* is 0.8 and the number of sample points is 384.

(2) In order to comprehensively analyze the seismic response of the subway station structure, both the randomness of ground motion and the randomness of soil property need to be considered. Deterministic analysis will ignore adverse working conditions.

(3) Compared to single random seismic excitation, the seismic response distribution range of the subway station structure under combined random excitation is wider. If the randomness of soil parameters is not considered, the seismic reliability of the subway station structure will be overestimated.

Acknowledgment

References

[1] Chen, Z., Fan, Y., & Jia, P. (2021). Influence of buried depth on seismic capacity of underground subway stations through performance-based evaluation. Structures, 32, 194-203.

[2] Chen, Z., Jia, P., Fan, Y., & Liu, Z. (2022). Parameter Analysis and Optimization of Friction Pendulum Bearings in Underground Stations Based on Genetic Algorithm. Journal of Earthquake Engineering, 26(15), 7814-7831.

[3] Chen, Z., Fan, Y., Huang, P., & Liu, Z. (2020). Column-wall shear ratio of subway station based on pushover method. China Civil Engineering Journal, 53(S1), 233-237. (In Chinese)

[4] Wu, W., Ge, S., Yuan, Y., Ding, W., & Anastasopoulos, I. (2020). Seismic response of subway station in soft soil: Shaking table testing versus numerical analysis. Tunnelling and Underground Space Technology, 100, 103389.

[5] Zhuang, H., Ren, J., Miao, Y., Jing, L., Yao, E., & Xu, C. (2021). Seismic performance levels of a large underground subway station in different soil foundations. Journal of Earthquake Engineering, 25(14), 2808-2833.

[6] Ma, C., Lu, D., & Du, X. (2018). Seismic performance upgrading for underground structures by introducing sliding isolation bearings. Tunnelling and Underground Space Technology, 74, 1-9.

[7] Fan, Y., Chen, Z., & Liu, Z. (2021, June). Influence of random soil parameters on seismic reliability of underground structure. In The 12th International Conference on Computational Methods (ICCM2021).

[8] Chen, Z. Y., & Liu, Z. Q. (2019). Stochastic seismic lateral deformation of a multi-story subway station structure based on the probability density evolution method. Tunnelling and Underground Space Technology, 94, 103114.

[9] Hu, H., Huang, Y., & Zhao, L. (2020). Probabilistic seismic-stability analysis of slopes considering the coupling effect of random ground motions and spatially-variable soil properties. Natural Hazards Review, 21(3), 04020028.

[10] Feng, S., Dai, S., & Lei, H. (2023). Application of an integrated nonuniform seismic random excitation method in tunnel engineering in soft soil areas. Computers and Geotechnics, 153, 105043.

[11] Li, D. Q., Wang, M. X., & Du, W. (2020). Influence of spatial variability of soil strength parameters on probabilistic seismic slope displacement hazard analysis. Engineering Geology, 276, 105744.

[12] He, Z., Xu, H., Gardoni, P., Zhou, Y., Wang, Y., & Zhao, Z. (2022). Seismic demand and capacity models, and fragility estimates for underground structures considering spatially varying soil properties. Tunnelling and Underground Space Technology, 119, 104231.

[13] Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety, 81(1), 23-69.

[14] Chen, J.B. & Li, J. (2006). Strategy of selecting points via sphere of contact in probability density evolution method for response analysis of stochastic structures. Journal of Vibration Engineering, (01),1-8. (In Chinese)

[15] Yang, J., Tao, J., Sudret, B., & Chen, J. (2020). Generalized F-discrepancy-based point selection strategy for dependent random variables in uncertainty quantification of nonlinear structures. International Journal for Numerical Methods in Engineering, 121(7), 1507-1529.

[16] Chen, J. B., & Chan, J. P. (2019). Error estimate of point selection in uncertainty quantification of nonlinear structures involving multiple nonuniformly distributed parameters. International Journal for Numerical Methods in Engineering, 118(9), 536-560.

[17] Chen, J. B., & Zhang, S. H. (2013). Improving point selection in cubature by a new discrepancy. SIAM Journal on Scientific Computing, 35(5), A2121-A2149.

[18] Liu, Z. J., Liu, W., & Peng, Y. B. (2016). Random function based spectral representation of stationary and non-stationary stochastic processes. Probabilistic Engineering Mechanics, 45, 115-126.

[19] Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). (2010). Code for seismic design of buildings (GB50011-2010). Beijing: China Architecture and Building Press. (In Chinese)

[20] Clough, R. W., & Penzien, J. (1975). Of structures. McGraw-Hill.

[21] Deodatis, G. (1996). Non-stationary stochastic vector processes: seismic ground motion applications. Probabilistic engineering mechanics, 11(3), 149-167.

[22] Cacciola, P., & Deodatis, G. (2011). A method for generating fully non-stationary and spectrum-compatible ground motion vector processes. Soil Dynamics and Earthquake Engineering, 31(3), 351-360.

[23] Toro, G. R. (2022). Uncertainty in shear-wave velocity profiles. Journal of Seismology, 26(4), 713-730.

[24] Ministry of Housing and Urban-Rural Development of the People's Republic of China (MOHURD). (2018). Standard for seismic design of underground structures (GBT51336-2018). Beijing: China Architecture and Building Press. (In Chinese)

[25] Shanghai Urban Constructions Communications (SUCC). (2010). Code for Seismic Design of Subway Structures (DG/TJ08-2064-2009). Shanghai: Shanghai Building Materials Industry Market Board. (In Chinese)

Numerical calculation of three dimensional MHD natural convection based on spectral collocation method and artificial compressibility method

Jiapeng Chang[1], †*Jingkui Zhang[1,2], Qifen Li[1], Yi Fan[1], Jiakai Zhang[1]

[1]College of Energy and Mechanical Engineering, Shanghai University of Electric Power, Shanghai 200090, China.
[2]Shanghai Non-carbon energy conversion and utilization institute，Shanghai 200240, China.

*Presenting author: zk_neu@163.com
†Corresponding author: zk_neu@163.com

Abstract

The flow and heat transfer characteristics of a three-dimensional cavity filled with a conducting fluid are investigated in this study. The three-dimensional Navier-Stokes equations and energy equation are solved directly using our self-developed algorithm, SCM-ACM, which combines the spectral collocation method (SCM) with high-precision and exponential convergence, and the artificial compression method (ACM) with easy implementation and good numerical stability. In this paper, we examine the effects of Hartmann numbers (Ha) ranging from 0 to 100, magnetic field directions, and Grashof numbers (Gr) ranging from 1×10^4 to 1×10^6 on the structure of the flow and temperature fields, with a Prandtl number (Pr) of 0.71. The results show the Grashof and Hartmann number have a significant impact on the flow and temperature structure in the middle of the cube, but little effect on that near the walls. As the Grashof number increases, a stable thermal stratification is formed at the center of the cube, and thermal boundary layers are formed near the horizontal walls. The increase in Grashof number enhances the heat transfer rate and increases the temperature difference between the upper hot fluid and the lower cold fluid in the cube. Furthermore, the increase in Grashof number enhances the convective intensity between the isothermal walls, leading to the formation of more vortices, which move towards the corners due to the combined action of centrifugal force and inertia. The Hartmann number has a stabilizing effect on the flow and weakens the heat transfer, while at higher Grashof numbers, the magnetic effect becomes more notable. When $Ha>50$, the magnetic effect is no longer significant. The magnetic field parallel to the temperature gradient (B_X) is more effective in suppressing heat transfer than the magnetic field perpendicular to the temperature gradient (B_Y). The methods and conclusions have certain theoretical guidance for the design and optimization of relevant engineering fields.

Keywords: Spectral collocation method, artificial compression method, magnetohydrodynamics, three-dimensional natural convection

Introduction

Natural convection flow is one of the important phenomenon in fluid mechanics, which has been widely observed in natural systems and engineering applications[1]–[4]. The control of the flow direction and heat transfer rate of conducting fluids by an externally applied magnetic field has attracted significant attention in the fields of thermodynamics and fluid mechanics. Consequently, magnetohydrodynamics (MHD) has experienced substantial development and application in the past decade[5], [6]. In this study, a three-dimensional cavity filled with a conductive fluid is investigated to examine the effects of the applied magnetic field intensity, Grashof number, and magnetic field orientation on the heat transfer rate and flow structure.

Natural convection heat transfer inside an enclosure is a fundamental heat transfer mechanism[7]. The heat boundary conditions, fluid properties, and intensity and directions of the magnetic field all have significant effects on the flow structure. The initial research focused on two-dimensional models. Bondareva et al. [8] studied the natural convective phase transformation process of solid pure potassium in a cavity with an internal heat source under the action of an inclined magnetic field. They concluded that the average Nusselt number (Nu_{avg}) increased with the tilt angle of the magnetic field and decreased with the Hartmann number. Saleh et al. [9] used the finite difference method to study the heat transfer in a rotating cavity under a magnetic field. They analyzed the effects of the strength and directions of the magnetic field and the rotation speed of the square cavity on heat transfer. Rudraiah et al. [10] found that when the magnetic field was relatively strong, the fluid in most areas of the shell almost stagnated. Sivasankaran et al. [11] used the finite volume method to study the influence of temperature-related characteristics of water near the density extremum on fluid flow and heat transfer under a uniform magnetic field. They found that the heat transfer rate increased with the increase of Rayleigh number (Ra) and decreased with the increase of Hartmann number.

In recent years, some scholars have conducted research on natural convection of magnetohydrodynamics in a three-dimensional cavity, benefiting from the optimization of numerical methods and the improvement of computational power. Sajjadi et al. [12], [13] studied the effects of Rayleigh number and magnetic field strength on heat transfer using the lattice Boltzmann method. They found that increasing the Rayleigh number enhances the heat transfer rate, while increasing the Hartmann number had the opposite effect. Later, Zhou et al. [14] used the same method to investigated the flow dynamics and mixed heat transfer of an Al_2O_3/water nanofluid in a cubic cavity. Their results showed that the Richardson number (Ri) has a significant impact on both the flow and temperature fields. Ellahi et al. [15] studied the heat transfer in a cavity with a pair of adiabatic horizontal and spanwise walls. They concluded that the Nusselt number (Nu) increases with the Rayleigh number. Zikanov et al. [16] investigated the effects of the magnetic field orientation on natural convection heat transfer using the finite volume method. Their results showed that the magnetic field parallel to the temperature gradient could maximally suppress heat transfer. Bouchta et al. [17] studied the heat transfer behavior of a particle-fluid mixture in a cavity with adiabatic horizontal and spanwise walls using the finite volume method and analyzed the relative positions of hot and cold walls on heat transfer. Okada et al. [18], [19] investigated the effects of different magnetic field directions and strengths using experimental and numerical simulation methods. They found that the magnetic field parallel to the temperature gradient could suppress the heat transfer rate by a factor of 10 compared to the horizontal magnetic field perpendicular to the temperature gradient.

Currently, most studies on MHD flow utilize finite element method, finite difference method, or lattice Boltzmann method. Traditional spatial discretization schemes require high spatial and temporal accuracy to directly solve the governing equations of natural convection in MHD, leading to significantly increased computational costs especially with lager Grashof number. The SCM exhibits exponential convergence and high accuracy. Therefore, employing this method for spatial discretization can achieve higher accuracy while computational costs can be greatly reduced. Additionally, the ACM possesses advantages such as numerical stability and easy convergence. Hence, in this study, we directly solve the MHD natural convection equations using the SCM-ACM. Regarding the research scope,

numerous studies assume that the horizontal and spanwise walls of the cavity are adiabatic. However, achieving such ideal thermal boundary conditions in experiments and practical engineering applications is challenging. Considering the numerical solutions as theoretical guidance for real-world engineering applications, we introduce a cubic cavity with one pair of isothermal walls, while the remaining four walls are conductive. We conducted a detailed investigation on the effect of magnetic field intensity, Grashof numbers, and magnetic field directions on the three-dimensional flow and heat transfer.

2. Mathematical Description

In this work, the natural convection flow characteristics of MHD are analyzed in a three-dimensional cavity. As shown in Fig. 1, the cavity with edge length L (L=1) is filled with viscous, incompressible, and electrically conductive fluid. The vertical walls (X=0 and X=L) are maintained at constant temperatures of T_h and T_c (T_h>T_c), respectively, while the other four walls are assumed to be perfectly conductive. The magnetic field is applied parallel to the coordinate axes, such as the magnetic field B_X in the X direction.
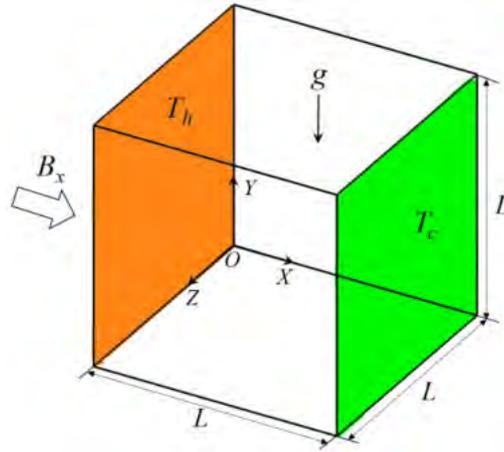
Fig. 1. The geometric model for the natural convection.

Since the Hartmann number (0-100) is not very large and the magnetic Reynolds number Re_m<<1, the induce magnetic field can be ignored. The ACM method adds $\partial p / \partial t$, $\partial \boldsymbol{u} / \partial t$ and $\partial T / \partial t$ to the continuity, momentum, and energy equations, respectively. When the solution time is long enough, the steady-state results can be obtained [20]. ACM is widely used due to its simplicity and ease of convergence when solving incompressible or weakly compressible flow equations [21]–[24].

$$\frac{\partial p}{\partial t} + c^2 \nabla \cdot \boldsymbol{u} = 0 \tag{1}$$

$$\frac{D\boldsymbol{u}}{Dt} = -\nabla \left(\frac{p}{\rho} \right) + \nabla \cdot (v \nabla \boldsymbol{u}) + \frac{1}{\rho} \boldsymbol{J} \times \boldsymbol{B} \tag{2}$$

$$\frac{\partial T}{\partial t} + \nabla \cdot (\boldsymbol{u}T) = \alpha \nabla^2 T \tag{3}$$

where p, $\boldsymbol{u}$, ρ, v, T, α and t are the pressure, velocity vector, density, kinematic viscosity, temperature, thermal diffusivity and the pseudo time, respectively. c is the artificial compressibility parameter. For all the cases in this paper, c=1.5 [25]. $\boldsymbol{J}$ is the current density, $\boldsymbol{J} = \varepsilon (\boldsymbol{u} \times \boldsymbol{B})$, where ε and $\boldsymbol{B}$ are the conductivity and magnetic field.

In order to make the obtained results universal, each variable is dimensionless through Eq. (4).

$$X = \frac{x}{L}, Y = \frac{y}{L}, Z = \frac{z}{L}, U = \frac{u}{U_0}, V = \frac{v}{U_0}, W = \frac{w}{U_0}, P = \frac{p}{\rho U_0^2}, \theta = \frac{T}{T_r} \tag{4}$$

where *x*, *y* and *z* are the Cartesian coordinates. *X*, *Y* and *Z* are the dimensionless Cartesian coordinates, respectively. *U*, *V* and *W* are the dimensionless velocities in *X*-, *Y*- and *Z*-directions. *L* and U_0 are the reference length and the reference velocity. *P* is the dimensionless pressure. θ and T_r are the dimensionless temperature and reference temperature. The ACM equations for three-dimensional natural convection in dimensionless form are as follows.

$$\frac{\partial P}{\partial t} + c^2 \left(\frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y} + \frac{\partial W}{\partial Z} \right) = 0 \tag{5}$$

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial X}(U^2) + \frac{\partial}{\partial Y}(VU) + \frac{\partial}{\partial Z}(WU) = -\frac{\partial P}{\partial X} + \frac{1}{Re}\nabla^2 U + F_x \tag{6}$$

$$\frac{\partial V}{\partial t} + \frac{\partial}{\partial X}(UV) + \frac{\partial}{\partial Y}(V^2) + \frac{\partial}{\partial Z}(WV) = -\frac{\partial P}{\partial Y} + \frac{1}{Re}\nabla^2 V + F_y \tag{7}$$

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial X}(UW) + \frac{\partial}{\partial Y}(VW) + \frac{\partial}{\partial Z}(W^2) = -\frac{\partial P}{\partial Z} + \frac{1}{Re}\nabla^2 W + F_z \tag{8}$$

$$\frac{\partial \theta}{\partial t} + \frac{\partial}{\partial X}(U\theta) + \frac{\partial}{\partial Y}(V\theta) + \frac{\partial}{\partial Z}(W\theta) = \frac{1}{PrRe}\nabla^2 \theta \tag{9}$$

where *Re* is the Reynold number, $Re = \frac{U_0 L}{v}$, . *Pr* is the Prandtl number, *Pr=v/α*. F_x, F_y and F_z are the source terms in *X*-, *Y*- and *Z*-directions. Taking the magnetic field along the *Y* direction as an example, there are $F_x = \frac{Ha^2}{Re}U$, $F_y = \frac{Ra}{PrRe^2}\theta$ and $F_z = \frac{Ha^2}{Re}W$. *Ra* is the Rayleigh number, $Ra = \frac{g\beta(T_h - T_c)L^3}{\alpha v}$, where β is the thermal expansion coefficient and *g* is the gravitational acceleration. *Ha* is the Hartmann number, $Ha = B_y L\sqrt{\frac{\varepsilon}{\mu}}$, where μ is the dynamic viscosity.

All walls are assumed to be non-slip. The two vertical walls located at *X*=0 and *X*=1 are assumed to be isothermal.

$$\theta|_{X=0} = 1, \ \theta|_{X=1} = 0 \tag{10}$$

Both horizontal and spanwise walls conduct heat.

$$\theta|_{Y=0, 1} = \theta|_{Z=0, 1} = 1 - X \tag{11}$$

The rate of heat transfer is characterized by the local Nusselt number (Nu_{local}) and the average Nusselt number, which can be obtained based on the temperature field.

$$Nu_{local} = \frac{-L}{\Delta T}\frac{\partial T}{\partial x}\bigg|_{x=0,L} \tag{12}$$

$$Nu_{avg} = \frac{1}{A}\int_A Nu_{lacal} dA = \frac{1}{A}\iint Nu_{local} dydz \tag{13}$$

3. Numerical method

3.1 Spectral collocation method

Spectral methods are widely used to solve problems related to flow, heat, and mass transfer due to their high accuracy and exponential convergence [26]–[29]. In this paper, Chebyshev-Gauss-Lobatto (CGL) collocation point pairs are chosen to discretize equations (5)-(9) spatially. Here, the X direction is considered for illustration. The first step involves selecting the configuration points. The location r_i of CGL collocation point N can be determined as follows.

$$r_i = -\cos\frac{\pi i}{N}, i = 0, \cdots, N \tag{14}$$

where r_i is the location of CGL collocation point i in X coordinate. N is the CGL collocation points in X coordinate.

The second step is to transform the computational domain. Eq. (15) is used to convert the actual physical interval $\{\mathbf{E}:(X) \in [X_{\min}, X_{\max}]\}$ into a standard spectral interval $\{\mathbf{E}:(r) \in [-1, 1]\}$. The spatial partial derivatives after the transformation are as follows.

$$r = \frac{2X - (X_{\max} + X_{\min})}{X_{\max} - X_{\min}}, r \in [-1, 1] \tag{15}$$

$$\begin{cases} \dfrac{\partial \phi}{\partial X} = \dfrac{\partial \phi}{\partial r}\dfrac{\partial r}{\partial X} = \dfrac{2}{(X_{\max} - X_{\min})}\dfrac{\partial \phi}{\partial r} \\ \dfrac{\partial^2 \phi}{\partial X^2} = \left(\dfrac{2}{X_{\max} - X_{\min}}\right)^2 \dfrac{\partial^2 \phi}{\partial r^2} \end{cases} \tag{16}$$

Thirdly, the spatial partial derivatives are replaced by truncated series. The general variable $\phi(r)$ can be approximated on collocation points by truncated series as:

$$\phi(r) \approx \sum_{j=0}^{N} \phi(r_j) h_j(r) \tag{17}$$

where $h_j(r)$ is the Lagrange interpolation polynomial, and the expression is given as follows.

$$h_j(r) = \frac{(-1)^{j+1}(1 - r^2)T_N'(r)}{c_j N^2 (r - r_j)} \tag{18}$$

$$c_j = \begin{cases} 2, & j = 0, N \\ 1, & j = 1, \cdots, N-1 \end{cases} \tag{19}$$

where $T_N'(r)$ is the first-order derivative of the Chebyshev polynomial $T_N(r)$.

$$T_N(r) = \cos\left(i\cos^{-1}(r)\right), i = 0, \cdots, N \tag{20}$$

So, the expressions of the first-order and second-order derivatives at $r=r_0$ are as follows.

$$\frac{\partial \phi(r)}{\partial r}\bigg|_{r=r_0} = \sum_{j=0}^{N} h_j^{(1)}(r_0)\phi(r_j) = \sum_{j=0}^{N} \boldsymbol{D}_{ij}^{(1)}\phi(r_j), i = 0, \cdots, N \tag{21}$$

$$\frac{\partial^2 \phi(r)}{\partial r^2}\bigg|_{r=r_0} = \sum_{j=0}^{N} h_j^{(2)}(r_0)\phi(r_j) = \sum_{j=0}^{N} \boldsymbol{D}_{ij}^{(2)}\phi(r_j), i = 0, \cdots, N \tag{22}$$

where $h^{(1)}$ and $h^{(2)}$ are the first-order and second-order derivatives of $h(r)$. $\boldsymbol{D}^{(1)}$ and $\boldsymbol{D}^{(2)}$ are the first-order and second-order coefficient matrices corresponding to the CGL collocation points. The elements of coefficient matrices are shown as follows.

$$
\boldsymbol{D}_{ij}^{(1)} = \begin{cases} \dfrac{c_i(-1)^{i+j}}{c_j(r_i - r_j)}, & i \neq j \\[4mm] \dfrac{-r_j}{2(1 - r_j^2)}, & 1 \leq i = j \leq N - 1 \\[4mm] \dfrac{2N^2 + 1}{6}, & i = j = 0 \\[4mm] -\dfrac{2N^2 + 1}{6}, & i = j = N \end{cases}
, \quad
\boldsymbol{D}_{ij}^{(2)} = \begin{cases} \dfrac{(-1)^{i+j}\left(r_i^2 + r_i r_j - 2\right)}{c_j\left(1 - r_i^2\right)\left(r_i - r_j\right)^2}, & 1 \leq i \leq N-1,\ 0 \leq j \leq N,\ i \neq j \\[4mm] -\dfrac{\left(N^2 - 1\right)\left(1 - r_i^2\right) + 3}{3\left(1 - r_i^2\right)^2}, & 1 \leq i = j \leq N - 1 \\[4mm] \dfrac{2(-1)^j\left(2N^2 + 1\right)\left(1 - r_j\right) - 6}{3c_j\left(1 - r_j\right)^2}, & i = 0,\ 1 \leq j \leq N \\[4mm] \dfrac{2(-1)^{(j+N)}\left(2N^2 + 1\right)\left(1 - r_j\right) - 6}{3c_j\left(1 + r_j\right)^2}, & i = N,\ 1 \leq j \leq N - 1 \\[4mm] \dfrac{N^4 - 1}{15}, & i = j = 0, N \end{cases}
\tag{23}
$$

Therefore, the dimensionless governing equations (5)-(9) can be discretized in space as follows.

$$
\frac{\partial P}{\partial t} + c^2\left(\sum_{l=0}^{N_x} \boldsymbol{A}_{il} U_{ljk} + \sum_{l=0}^{N_y} \boldsymbol{B}_{jl} V_{ilk} + \sum_{l=0}^{N_z} \boldsymbol{C}_{kl} W_{ijl} \right) = 0
\tag{24}
$$

$$
\begin{aligned}
&\frac{\partial U}{\partial t} + U_{ijk} \sum_{l=0}^{N_x} \boldsymbol{A}_{il} U_{ljk} + V_{ijk} \sum_{l=0}^{N_y} \boldsymbol{B}_{jl} U_{ilk} + W_{ijk} \sum_{l=0}^{N_z} \boldsymbol{C}_{kl} U_{ijl} = \\
&-\sum_{l=0}^{N_x} \boldsymbol{A}_{il} P_{ljk} + \frac{1}{Re}\left(\sum_{l=0}^{N_x} \boldsymbol{E}_{il} U_{ljk} + \sum_{l=0}^{N_y} \boldsymbol{F}_{jl} U_{ilk} + \sum_{l=0}^{N_z} \boldsymbol{M}_{kl} U_{ijl} \right)
\end{aligned}
\tag{25}
$$

$$
\begin{aligned}
&\frac{\partial V}{\partial t} + U_{ijk} \sum_{l=0}^{N_x} \boldsymbol{A}_{il} V_{ljk} + V_{ijk} \sum_{l=0}^{N_y} \boldsymbol{B}_{jl} V_{ilk} + W_{ijk} \sum_{l=0}^{N_z} \boldsymbol{C}_{kl} V_{ijl} = \\
&-\sum_{l=0}^{N_y} \boldsymbol{B}_{jl} P_{ilk} + \frac{1}{Re}\left(\sum_{l=0}^{N_x} \boldsymbol{E}_{il} V_{ljk} + \sum_{l=0}^{N_y} \boldsymbol{F}_{jl} V_{ilk} + \sum_{l=0}^{N_z} \boldsymbol{M}_{kl} V_{ijl} \right) + \frac{Gr}{Re^2}\theta - \frac{Ha^2}{Re} V_{ijk}
\end{aligned}
\tag{26}
$$

$$
\begin{aligned}
&\frac{\partial W}{\partial t} + U_{ijk} \sum_{l=0}^{N_x} \boldsymbol{A}_{il} W_{ljk} + V_{ijk} \sum_{l=0}^{N_y} \boldsymbol{B}_{jl} W_{ilk} + W_{ijk} \sum_{l=0}^{N_z} \boldsymbol{C}_{kl} W_{ijl} = \\
&-\sum_{l=0}^{N_z} \boldsymbol{C}_{kl} P_{ijl} + \frac{1}{Re}\left(\sum_{l=0}^{N_x} \boldsymbol{E}_{il} W_{ljk} + \sum_{l=0}^{N_y} \boldsymbol{F}_{jl} W_{ilk} + \sum_{l=0}^{N_z} \boldsymbol{M}_{kl} W_{ijl} \right) - \frac{Ha^2}{Re} W_{ijk}
\end{aligned}
\tag{27}
$$

$$
\frac{\partial \theta}{\partial t} + U_{ijk} \sum_{l=0}^{N_x} \boldsymbol{A}_{il} \theta_{ljk} + V_{ijk} \sum_{l=0}^{N_y} \boldsymbol{B}_{jl} \theta_{ilk} + W_{ijk} \boldsymbol{C}_{kl} \theta_{ijl} = \frac{1}{RePr}\left(\sum_{l=0}^{N_x} \boldsymbol{E}_{il} \theta_{ljk} + \sum_{l=0}^{N_y} \boldsymbol{F}_{il} \theta_{ilk} + \sum_{l=0}^{N_z} \boldsymbol{M}_{il} \theta_{ijl} \right)
\tag{28}
$$

where *A*, *B* and *C* are the first-order coefficient matrix $\boldsymbol{D}^{(1)}$ for configuration points N_x, N_y, and N_z corresponding to directions *X*-, *Y*-, and *Z*-directions, while *E*, *F* and *M* are the second-order coefficient matrices $\boldsymbol{D}^{(2)}$ for configuration points corresponding to directions *X*-, *Y*-, and *Z*-directions.

The explicit fourth-order Runge-Kutta scheme is adopted in this paper for the discretization of the time term [20], [25], which is known for its excellent performance in solving incompressible fluid flows.

To ensure obtaining the steady-state results using ACM, it is necessary to solve for a sufficiently long time to make the unsteady term tend to 0. The calculation can be stopped when the maximum residuals of all variables at two adjacent time steps satisfy Eq. (29), where ε is the tolerance and is set as $\varepsilon = \Delta t \times 10^{-4}$.

$$\max\left(\left|U^{n+1}-U^n\right|, \left|V^{n+1}-V^n\right|, \left|W^{n+1}-W^n\right|, \frac{1}{c^2}\left|P^{n+1}-P^n\right|, \left|\theta^{n+1}-\theta^n\right|\right) < \varepsilon \quad (29)$$

3.2 Grid independence test

We employ four sets of grid nodes as 22×22×22, 32×32×32, 42×42×42, and 52×52×52, to investigate grid test. For the selected case (*Ha*$_Y$=10, *Gr*=1×10^5, and *Pr*=0.71), after attaining convergence of the iterative results, we compare the average Nusselt numbers on the cold wall under different grid schemes. As depicted in Fig. 2, a significant variation in the average Nusselt numbers is observed when the number of nodes increases from 22^3 to 42^3. However, the average Nusselt number only changes by 0.055% when the number of nodes increases from 42^3 to 52^3. Taking the computational efficiency and accuracy into account, we select the 42^3-node configuration for spatial discretization.
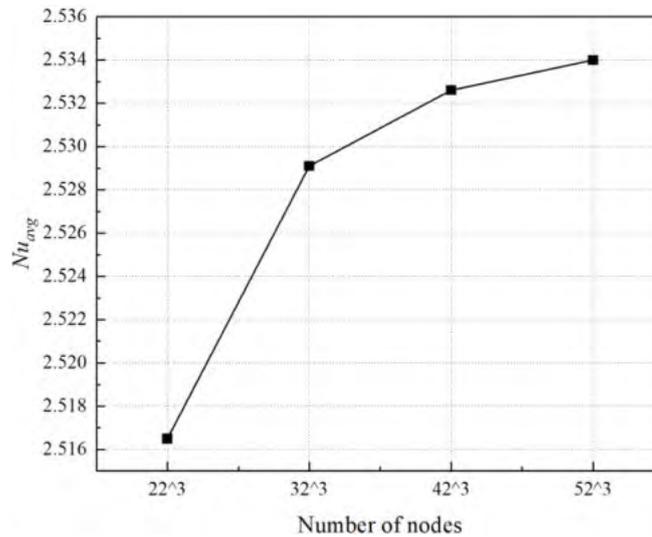
Fig. 2 The average Nusselt number on the cold wall at different number of nodes.

3.3 Code validation

The code validation is divided into two parts. First, the natural convection is verified without considering the magnetic field. We set the same thermal boundary conditions as those by Fusegi et al. [30], Tric et al. [31], Saitoh et al. [32], and Luo et al. [33]. Table 1 shows the calculated average Nusselt number of the hot wall, and our results exhibit good agreement with the existing results.

Table 1 The average Nusselt number on the hot wall of natural convection in a three-dimensional cavity with different *Ra* for *Pr*=0.71.

Ra	Fusegi et al.	Tric et al.	Saitoh et al.	Luo et al.	Present results
10^4	2.100	2.054	2.0624	2.057	2.0582
10^5	4.361	4.337	4.3665	4.337	4.3423
10^6	8.770	8.641	8.6973	8.648	8.6547

Next, we compare our results with those by Singh et al. [34] with the magnetic field condition, and Fig. 2 shows excellent agreement between our results and those in the literature.
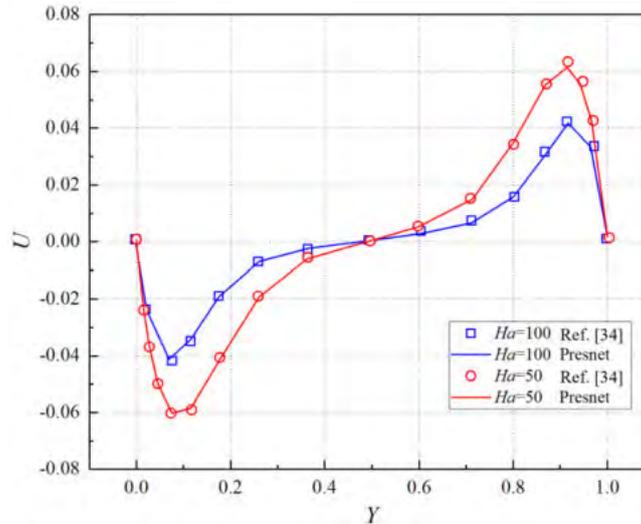
Fig. 3. The distribution of velocity (*U*) alone the line (0.5, *Y*, 0.5) for *Ha*=50 and 100.

4. Results and discussion

In this study, the SCM-ACM numerical method is employed to solve the natural convection of MHD in a three-dimensional cavity with two isothermal walls and the remaining four walls being linear temperature distribution. The effects of Grashof numbers, magnetic field intensity and directions on flow and heat transfer are comprehensively analyzed through the examination of streamlines, isotherms and heat transfer rates.

4.1 Effects of Grashof number on natural convection

We firstly analyze the three-dimensional temperature isosurfaces. Fig. 4 shows that temperature isosurfaces gradually become curved with the increase of Grashof number. When $Gr=1\times10^6$, the isothermal surfaces become approximately parallel to the horizontal walls near the center of the cubic cavity, with little change near the vertical walls.

Then, the flow field and temperature field on the plane *Z*=0.5, *Y*=0.5 and *X*=0.5 are analyzed. Fig. 5 shows the isotherms on the main circulation surface (*Z*=0.5), where it is observed that the increase in Grashof number results in temperature stratification at the center of the cubic cavity. This leads to an increase in temperature gradient along the height direction, and subsequently, an increase in heat transfer rate. Moreover, as the Grashof number increases, the boundary layers near the two isothermal walls become thinner, and convective heat transfer has become the dominant mode of heat transfer. Since the horizontal surfaces are all thermally conductive, the boundary layers here develop faster.

Since the flow field and temperature field are coupled together, Fig. 6 shows that the streamlines at plane Z=0.5 also changes greatly. When the Grashof number (Gr=1×10⁴) is relatively small, the natural convection is not strong enough to form more than one primary vortex. With the increase in the Grashof number to $1×10^5$, the natural convection strengthens, and the number of vortices increases to two. When the Grashof number further increases to $1×10^6$, the two vortices move towards the corners under the driving force of natural convection and eventually settle in the top-left and bottom-right corners. This is because the corners are the convergence points of the heat convection, providing enough space and conditions for the vortices to stay. In contrast, the center area is the divergence point of the heat convection, and the streamlines will spread out, making it unfavorable for the vortices to exist. The inertia force that causes the vortex to move towards the center and the centrifugal force that causes the vortex to move towards the corners reach a balance, resulting in the vortexes concentrating in the left upper and right lower corners.

Fig. 7 shows that at the plane Y=0.5 the increase in Grashof number makes the isotherms approaching the isothermal surfaces (X=0 and X=1), and the boundary layers of the isothermal surfaces become thinner, leading to an increase in heat transfer rate.

According to the streamlines at the plane Y=0.5 in Fig. 8, it can be observed that the number of streamline intersection points increases from two to four when the Grashof number increases from $1×10^4$ to $1×10^5$. Further increasing the Grashof number to $1×10^6$, the flow intensity inside the cube continues to increase, and the four streamline intersection points gradually move to the four corners. This phenomenon occurs because the vortex structure can be better maintained in the corners of the cube, where there exists a stable pressure difference that makes it easier for the flow to form vortices.

Fig. 9 shows that in the cube, the hot fluid is located in the upper part, while the cold fluid is in the lower part. The phenomenon is caused by a combination of temperature gradients and gravity. As the Grashof number increases from $1×10^4$ to $1×10^6$, the convective strength increases, causing the upper part temperature to rise from 0.63 to 0.69 and the lower part temperature to decrease from 0.37 to 0.29. In other words, the temperature difference between the upper and lower parts of the cube becomes greater. This is because at higher Grashof numbers, the momentum and heat transfer become more intense, leading to more intense fluid motion and temperature distribution changes. The streamline in Fig. 10 changes greatly with the increase of Grashof number.

Fig. 11 demonstrates the effects of the Grashof number and Hartman number on heat transfer. The average Nusselt number on the hot wall increases with the Grashof number, and a larger Hartman number has a greater impact on the average Nusselt number. The average Nusselt number increases by 326.6 % when the Grashof number increases from 10^4 to 10^6 at Ha=0, while the increase is only 104.8 % at Ha=100.

Fig. 12 shows that the local Nusselt number increases with Grashof number and decreases with Hartman number. However, the position corresponding to the maximum value of local Nusselt number is not affected by changes in Hartman number and Grashof number. It should be noted that under strong magnetic fields (i.e., Ha=50 and 100), Fig. 12 (c) and Fig. 12 (d) show that the local Nusselt number decreases with increasing Grashof number at a height of 0.6-1.0 on the hot wall. This phenomenon is attributed to the fact that as the Grashof number increases, the effects of natural convection become more significant. While, the influence of the magnetic field also becomes stronger.

Fig. 13 illustrates that the velocity *V* distribution along the *X* direction is consistent under various Grashof numbers. Initially, the velocity is zero due to the no-slip boundaries. Subsequently, the velocity increases and reaches its maximum as a result of upward flow along the hot wall. Finally, the velocity decreases, and at the center of the cavity (*X*=0.5), the velocity is zero. The magnitudes of velocities in the left and right halves of the cubic cavity are equal but opposite in direction. Under the same Hartman number, as the Grashof number increases, the velocity near the wall increases while the velocity in the middle region of the cubic cavity decreases. For *Ha*=10, the maximum velocity corresponding to *Gr*=1×10^6 is approximately 1.6 times that of *Gr*=1×10^4. This value can reach 7.5 and 10.5 for *Ha*=50 and 100, respectively. Therefore, we conclude that the velocity difference under different Grashof numbers becomes more pronounced with an increase in Hartman number.

(a) *Gr*=1×10^4 (b) *Gr*=1×10^5 (c) *Gr*=1×10^6
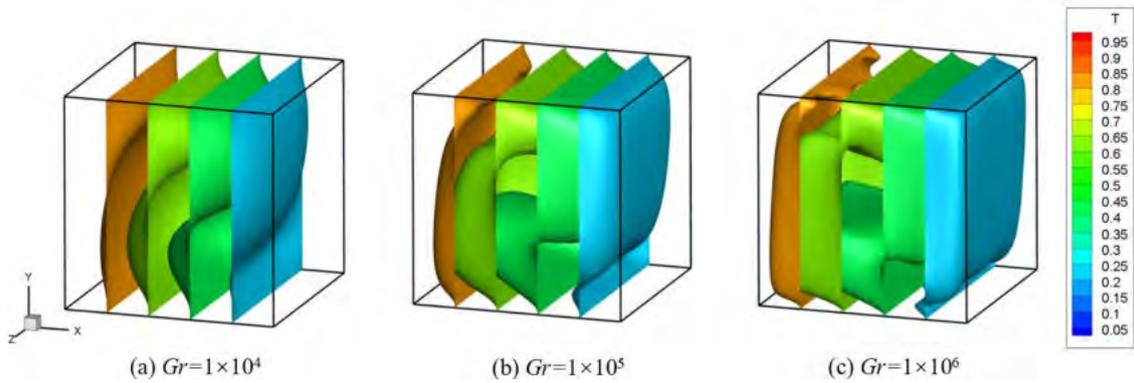
Fig. 4. The distribution of three-dimensional temperature isosurfaces for (a) *Gr*= 1×10^4, (b) *Gr*=1×10^5, (c) *Gr*=1×10^6.
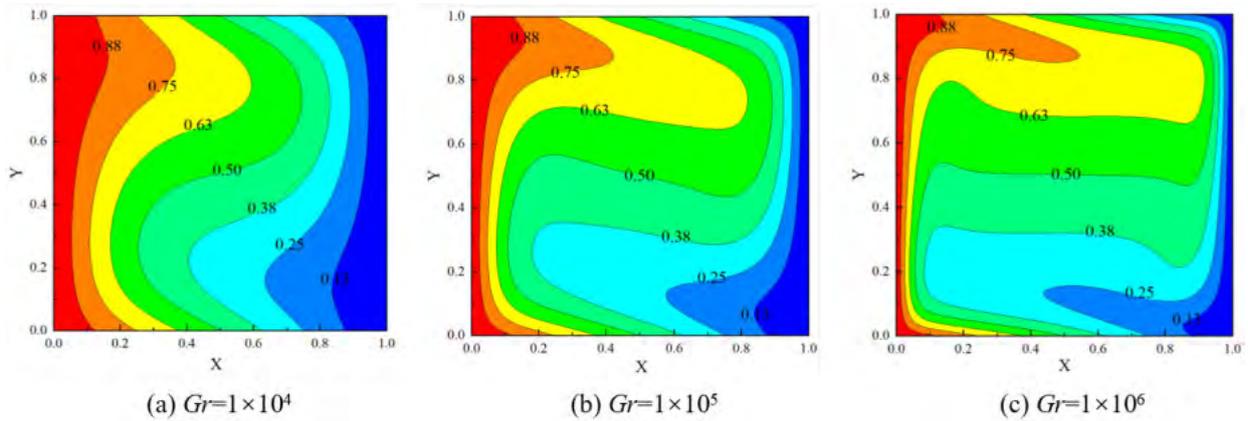
(a) *Gr*=1×10^4 (b) *Gr*=1×10^5 (c) *Gr*=1×10^6

Fig. 5. The distribution of temperature on the plane *Z*=0.5 for (a) *Gr*=1×10^4，(b) *Gr*=1×10^5 and (c) *Gr*=1×10^6.
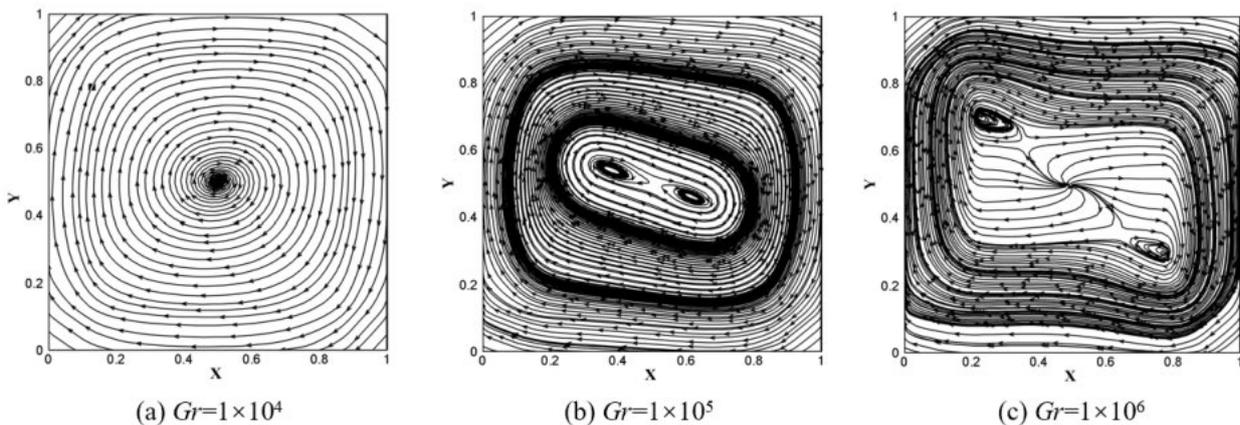
(a) *Gr*=1×10^4 (b) *Gr*=1×10^5 (c) *Gr*=1×10^6

Fig. 6. The distribution of streamline on the plane *Z*=0.5 for (a) *Gr*=1×10⁴，(b) *Gr*=1×10⁵ and (c) *Gr*=1×10⁶.
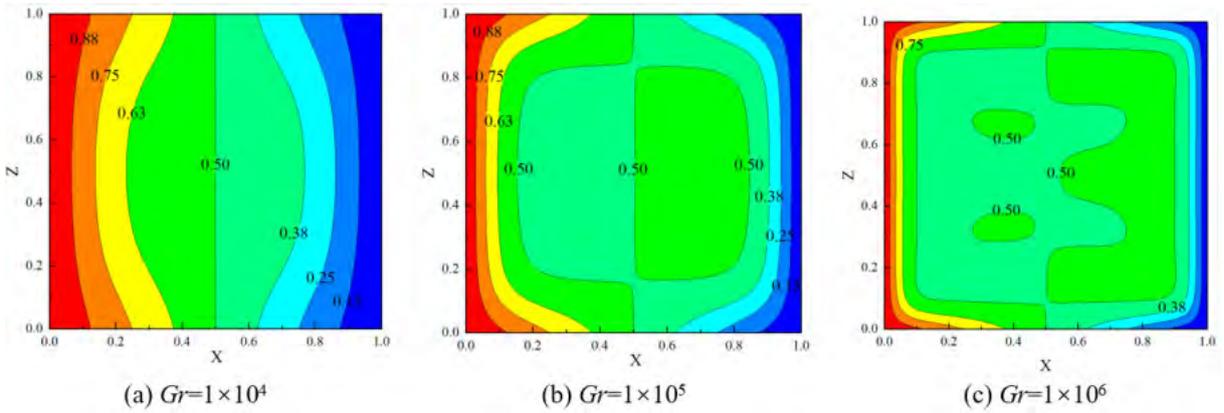
(a) *Gr*=1×10⁴

(b) *Gr*=1×10⁵

(c) *Gr*=1×10⁶

Fig. 7. The distribution of temperature on the plane *Y*=0.5 for (a) *Gr*=1×10⁴，(b) *Gr*=1×10⁵ and (c) *Gr*=1×10⁶.
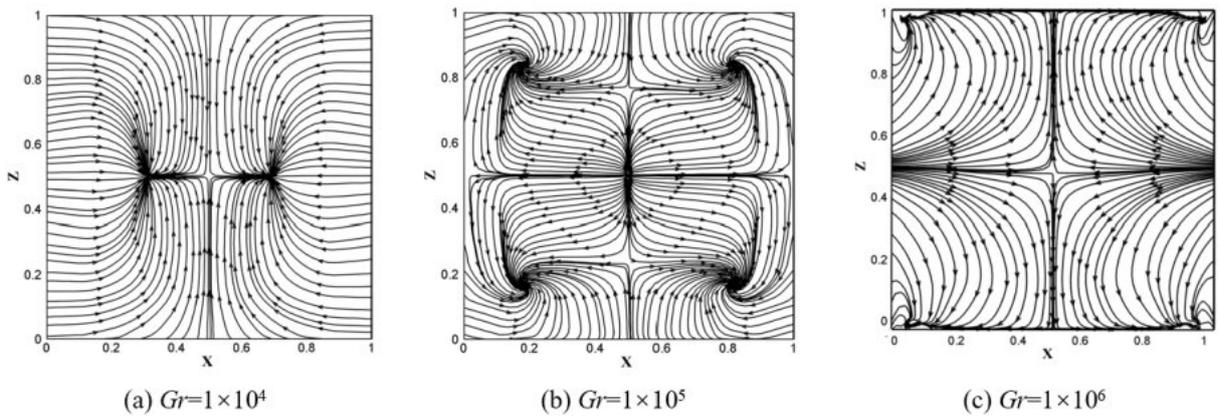
(a) *Gr*=1×10⁴

(b) *Gr*=1×10⁵

(c) *Gr*=1×10⁶

Fig. 8. The distribution of streamline on the plane *Y*=0.5 for (a) *Gr*=1×10⁴，(b) *Gr*=1×10⁵ and (c) *Gr*=1×10⁶.
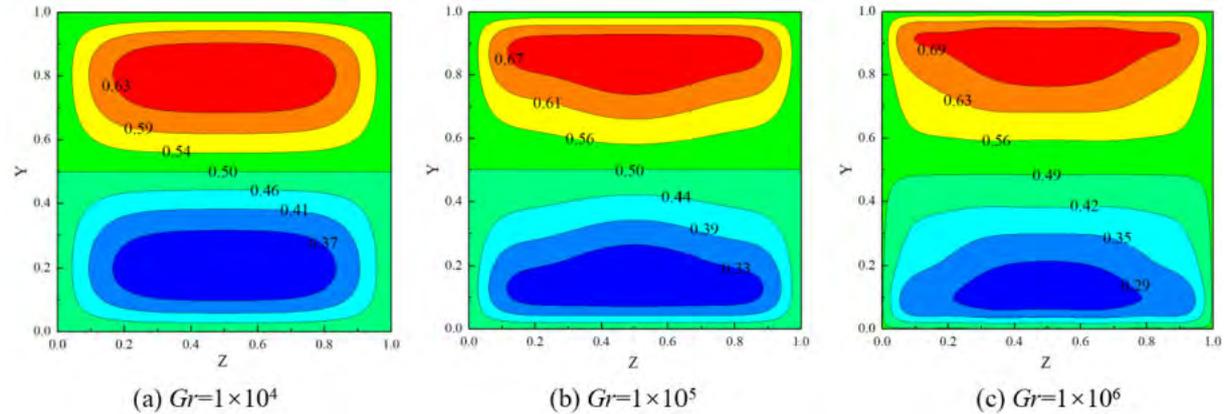
(a) *Gr*=1×10⁴

(b) *Gr*=1×10⁵

(c) *Gr*=1×10⁶

Fig. 9. The distribution of temperature on the plane *X*=0.5 for (a) *Gr*=1×10⁴，(b) *Gr*=1×10⁵ and (c) *Gr*=1×10⁶.

(a) $Gr=1\times10^4$ (b) $Gr=1\times10^5$ (c) $Gr=1\times10^6$
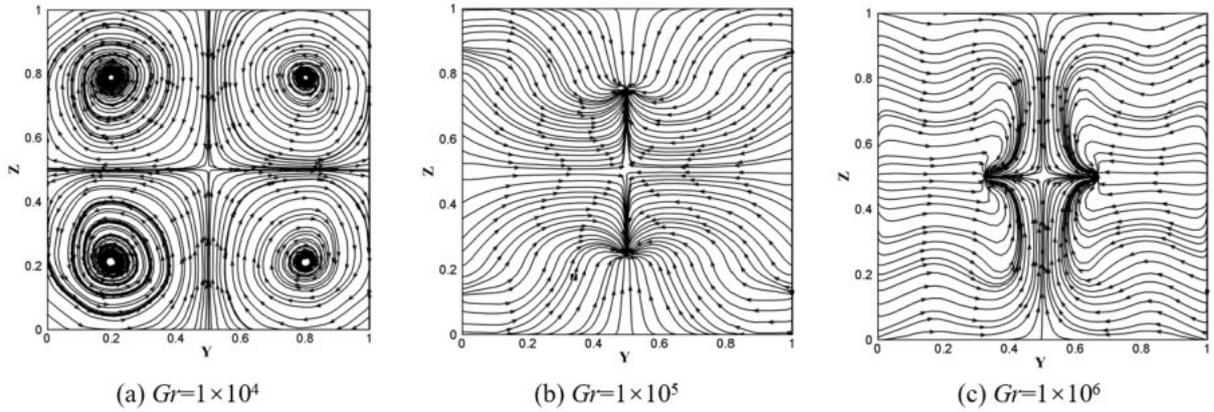
Fig. 10. The distribution of streamline on the plane $X=0.5$ for (a) $Gr=1\times10^4$, (b) $Gr=1\times10^5$ and (c) $Gr=1\times10^6$.
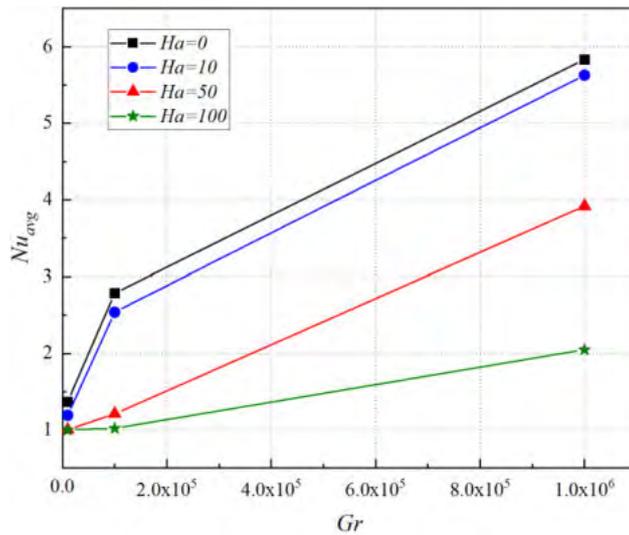
Fig. 11 The distribution of average Nusselt number with Grashof numbers on the hot wall for different Hartman numbers.

(a) $Ha=0$ (b) $Ha=10$
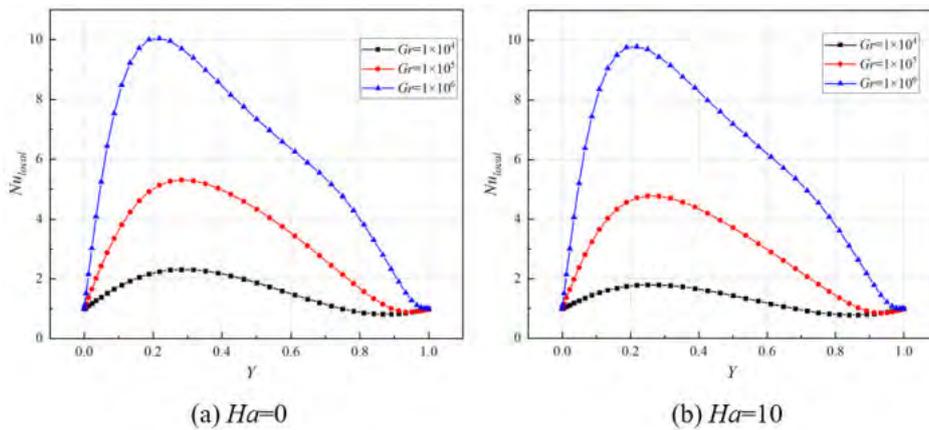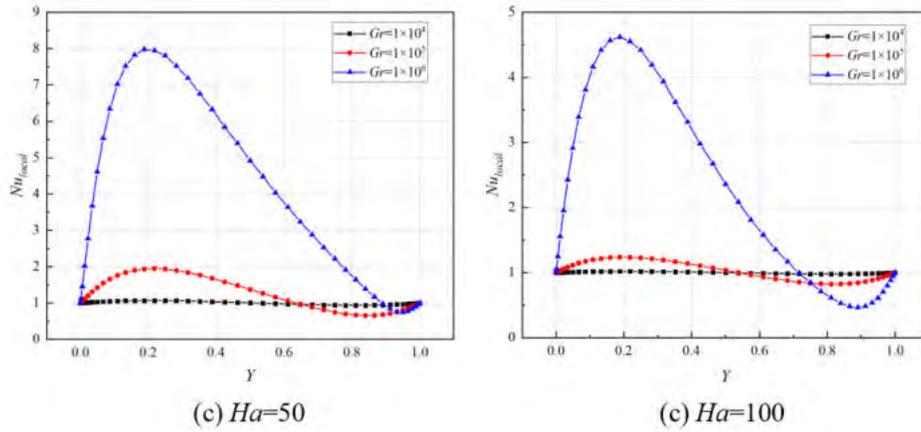
(c) *Ha*=50   (c) *Ha*=100

Fig. 12. The distribution of local Nusselt number alone the line (0, *Y*, 0.5) for (a) *Ha*=0, (b) *Ha*=10, (c) *Ha*=50 and (d) *Ha*=100.

(a) *Ha*=0   (b) *Ha*=10
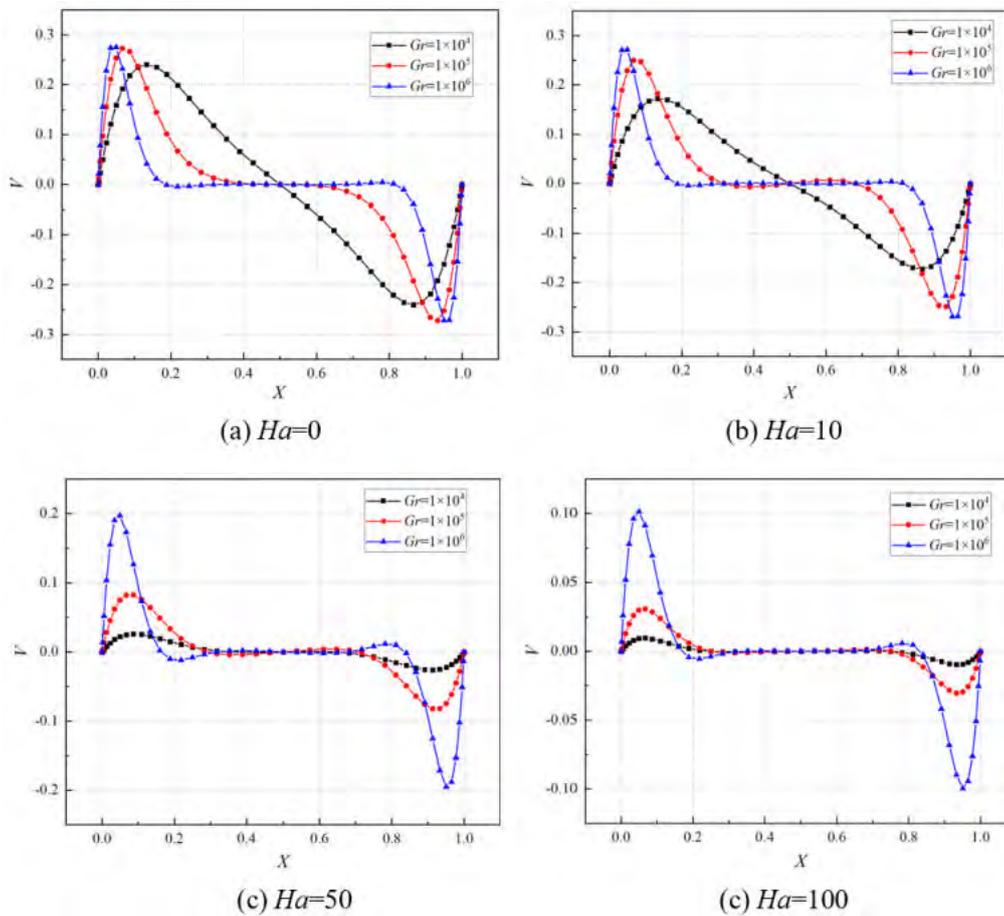
(c) *Ha*=50   (c) *Ha*=100

Fig. 13. The distribution of velocity *V* alone the line (*X*, 0.5, 0.5) for (a) *Ha*=0, (b) *Ha*=10, (c) *Ha*=50, and (d) *Ha*=100.

4.2 Effects of Hartmann number on natural convection

Using the same method as in section 4.1, we next investigate the effects of the magnetic field applied along the *Y* direction on flow and heat transfer at $Gr=1\times10^5$. Fig. 14 illustrates that increasing the Hartmann number results in isothermal surfaces that are approximately parallel to the *Y* direction, particularly for the central part of the cubic cavity. The temperature difference along the *Y* direction decreases.

Fig. 15 demonstrates that at the plane *Z*=0.5, the isotherms become approximately parallel to the isothermal walls as the Hartmann number increases. The magnetic field significantly restricts fluid flow along the *X* direction, and the boundary layers near the horizontal walls gradually disappears, while the boundary layers of the isothermal walls become thicker. These changes indicate that the heat transfer mode changes from convection to conduction, and the flow field structure in Fig. 16 changes to some extent.

Since the direction of the applied magnetic field is parallel to the *Y*-axis, the direction of the Lorentz force is perpendicular to the magnetic field direction, thus affecting the flow fields most significantly at the *Y*=0.5 plane. According to Fig. 17, as the Hartmann number increases, the thermal boundary layers expand towards the center. As shown in Fig. 18 (a) and Fig. 18 (b), when the Hartmann number increases from 10 to 50, the streamlines undergo significant changes. As Hartmann number continues to increase to 100, the changes become less pronounced.
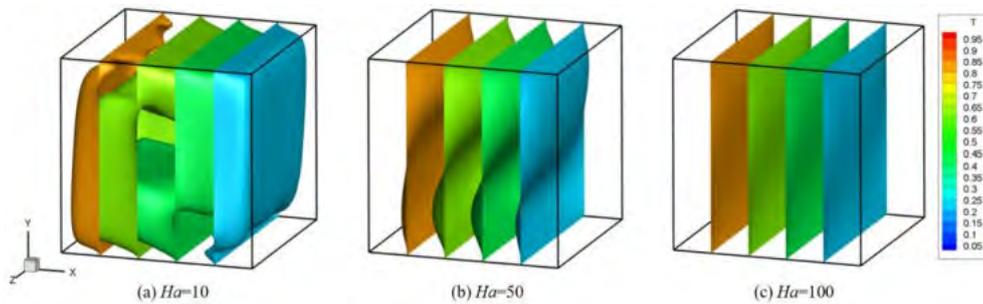
(a) *Ha*=10 (b) *Ha*=50 (c) *Ha*=100

Fig. 14. The distribution of three-dimensional temperature isosurfaces at *Gr*=1×10⁵ for (a) *Ha*=10, (b) *Ha*=50 and (c) *Ha*=100.
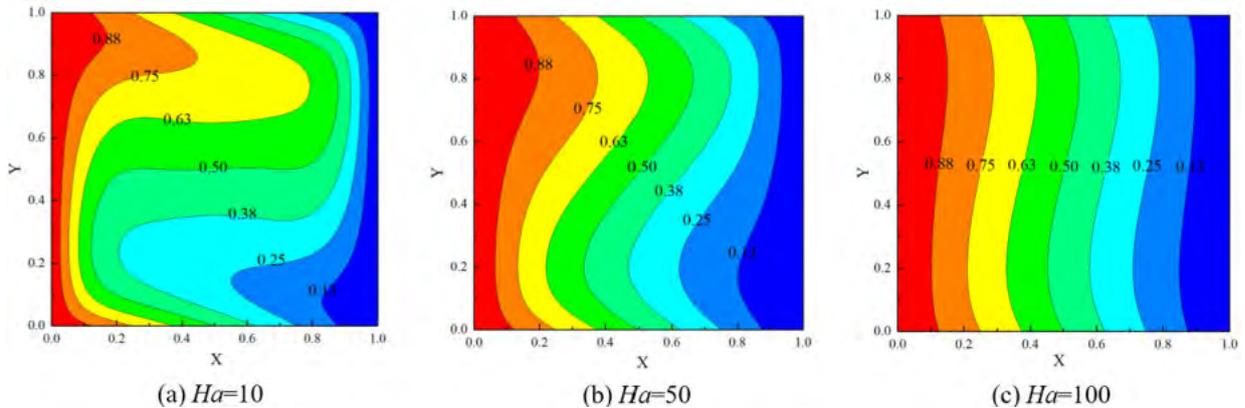
(a) *Ha*=10 (b) *Ha*=50 (c) *Ha*=100

Fig. 15 The distribution of temperature on the plane *Z*=0.5 at *Gr*=1×10⁵ for (a) *Ha*=10, (b) *Ha*=50 and (c) *Ha*=100.
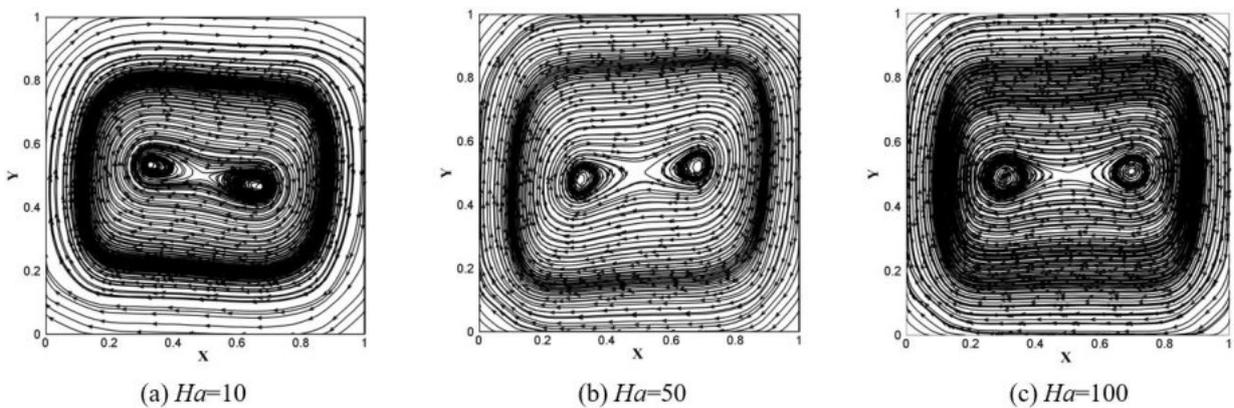
(a) *Ha*=10 (b) *Ha*=50 (c) *Ha*=100

Fig. 16. The distribution of streamline on the plane *Z*=0.5 at *Gr*=1×10⁵ for (a) *Ha*=10, (b) *Ha*=50 and (c) *Ha*=100.
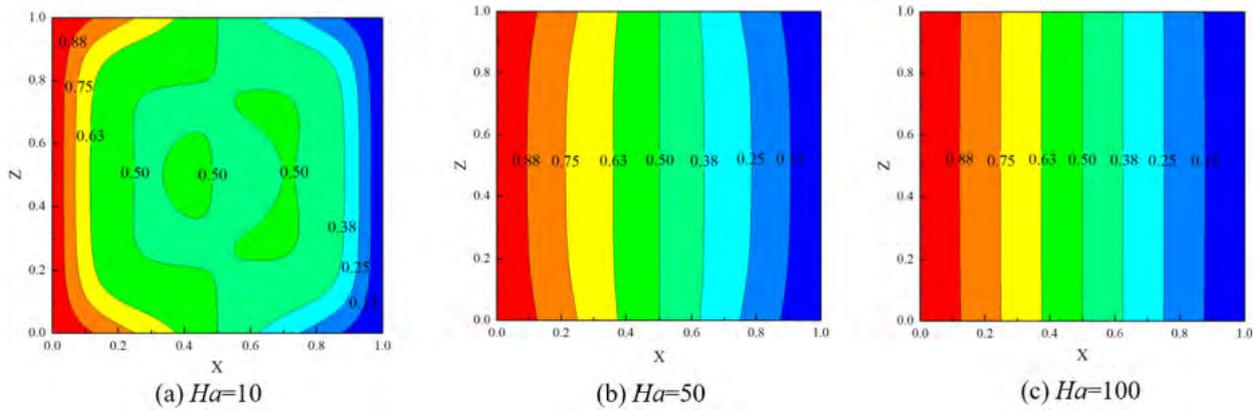
(a) *Ha*=10   (b) *Ha*=50   (c) *Ha*=100

Fig. 17 The distribution of temperature on the plane *Y*=0.5 at *Gr*=1×10⁵ for (a) *Ha*=10, (b) *Ha*=50 and (c) *Ha*=100.
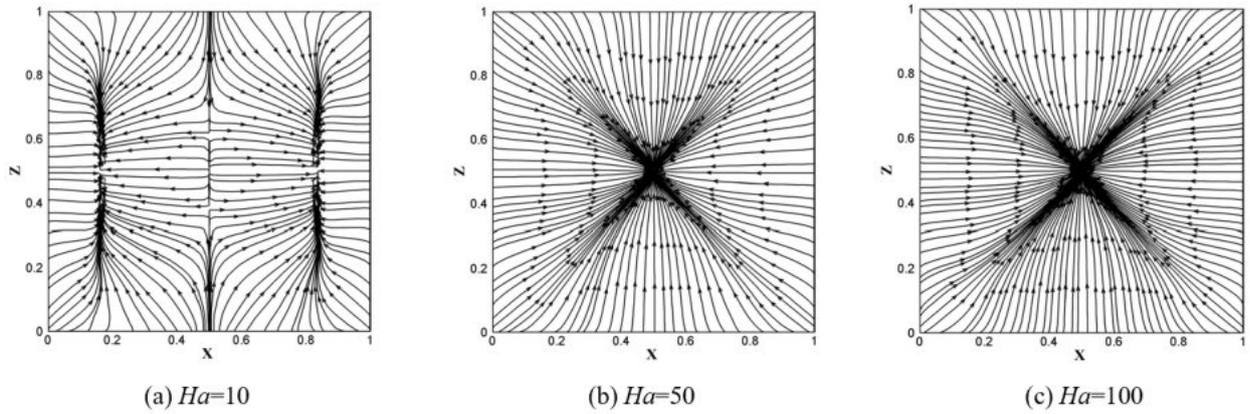
(a) *Ha*=10   (b) *Ha*=50   (c) *Ha*=100

Fig. 18. The distribution of streamline on the plane *Y*=0.5 at *Gr*=1×10⁵ for (a) *Ha*=10, (b) *Ha*=50 and (c) *Ha*=100.

From Fig. 19, it can be observed that at the plane *X*=0.5 the increase of Hartmann number results in a decrease of the maximum temperature difference between the upper and lower parts of the cube. The streamline in Fig. 20 has also undergone significant changes.

Fig. 21 shows that when the Hartmann number increases from 0 to 100, the average Nusselt number of the hot wall decreases by 26.8%, 63.3%, and 64.9% for *Gr*=10⁴, 10⁵, and 10⁶, respectively. Therefore, the magnetic field more obviously affects heat transfer under large Grashof numbers.
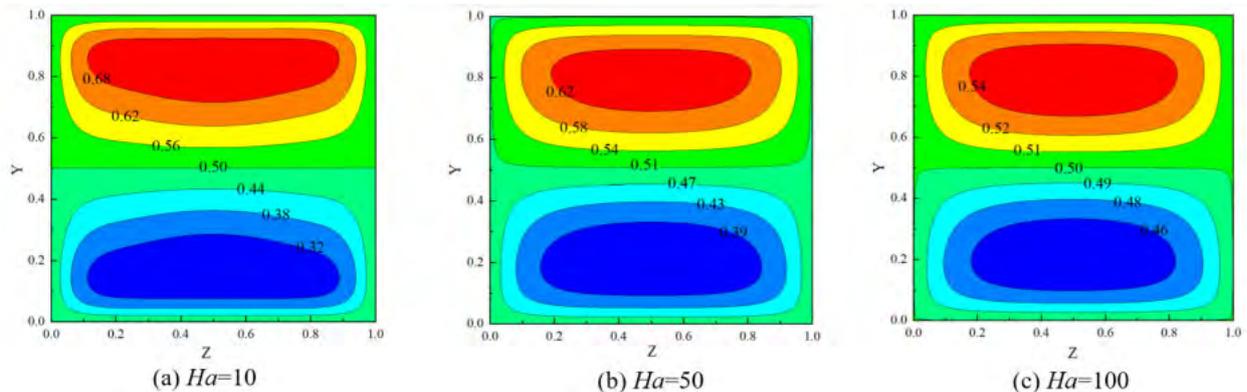
(a) *Ha*=10   (b) *Ha*=50   (c) *Ha*=100

Fig. 19 The distribution of temperature on the plane *X*=0.5 at *Gr*=1×10⁵ for (a) *Ha*=10, (b)
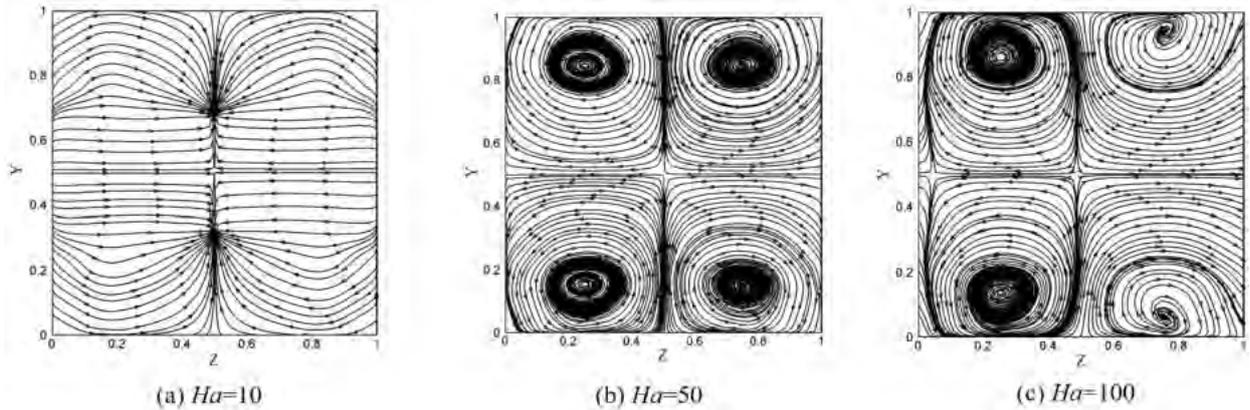
(a) *Ha*=10   (b) *Ha*=50   (c) *Ha*=100

Fig. 20. The distribution of streamline on the plane *X*=0.5 at *Gr*=1×10⁵ for (a) *Ha*=10, (b) *Ha*=50 and (c) *Ha*=100.
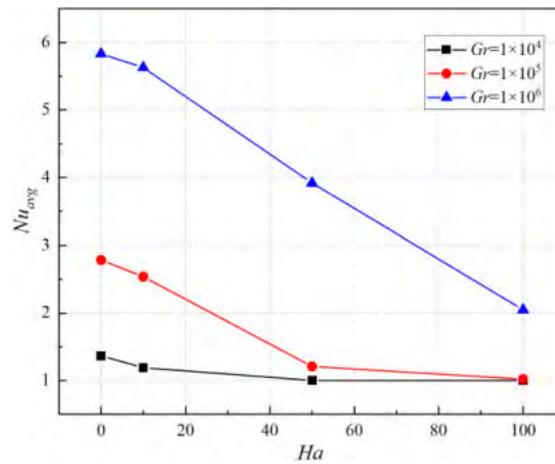
Fig. 21 The distribution of average Nusselt number with Hartmann numbers for different Grashof numbers.

Fig. 22 shows that as Hartmann number increases from 0 to 50, the local Nusselt number decreases significantly. When Grashof increases to 1×10^6, the magnetic field effects are not obvious, and the local Nusselt number is approximately equal to that under no magnetic field (*Ha*=0) at a relatively small Hartmann number (*Ha*=10). The positions corresponding to the maxima of the local Nusselt number remain nearly unchanged. For both smaller Grashof (*Gr*=1 × 10⁴) and larger Hartmann number (*Ha*=100), the local Nusselt number is approximately equal to 1.

Fig. 23 reveals that an increase in the Hartmann number leads to a decrease in the velocity *V* along the *X* direction, and most of the regional flow in the square cavity is suppressed. The velocity and the structure of the flow field are more significantly affected by the Hartmann number for smaller Grashof numbers.

To better understand the influence of the magnetic field on the transverse velocity (*W*), we examine the maximum velocity value in the *Z* direction under *Ha*=0, 10, 30, 40, 50, 60, 80, and 100. Fig. 24 demonstrates that the velocity decreases significantly as the Hartmann number increases from 0 to 50, but the velocity changes little as the Hartmann number increases from 50 to 100.
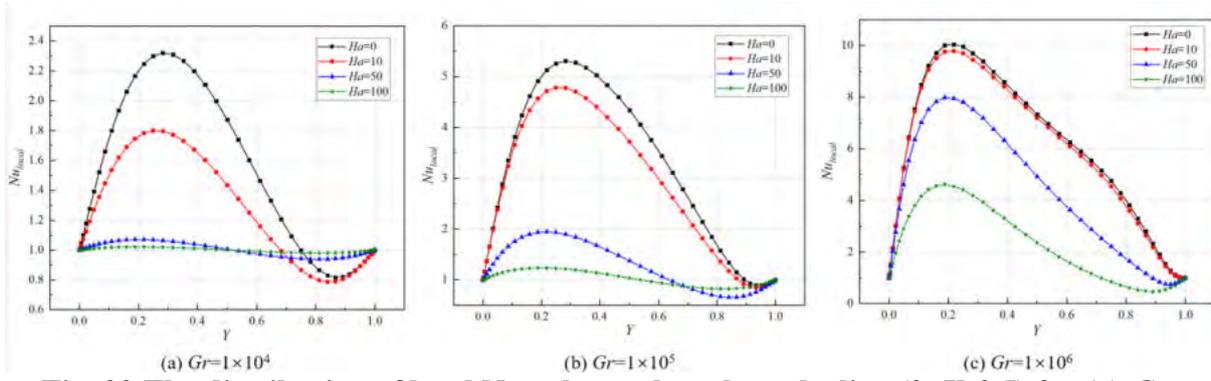
Fig. 22 The distribution of local Nusselt number alone the line (0, Y, 0.5) for (a) Gr= 1×10⁴, (b) Gr=1×10⁵ and (c) Gr=1×10⁶.
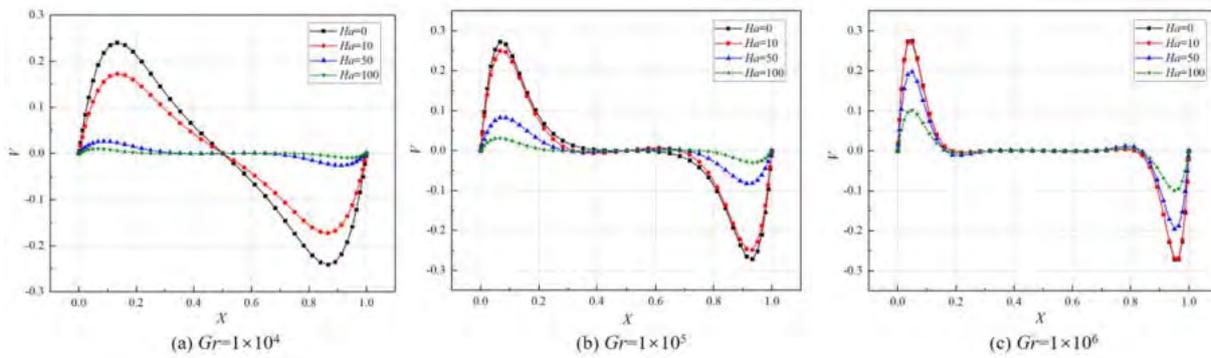
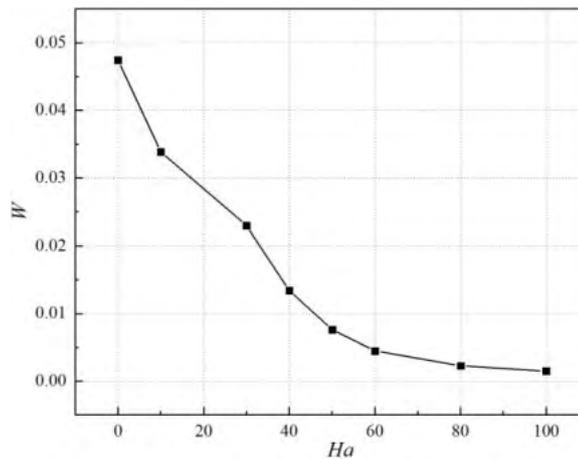**Fig. 23 The distribution of velocity V alone the line (X, 0.5, 0.5) for (a) Gr= 1×10⁴, (b) Gr=1×10⁵ and (c) Gr=1×10⁶.**

Fig. 24 The distribution of the maximum velocity (W) with Hartmann numbers at Gr=1×10⁵.

4.3 The effects of magnetic field directions on natural convection

Later, we analyze the difference in the effects of magnetic field directions (B_X, B_Y) on heat transfer. Fig. 25 shows that the corresponding local Nusselt number of B_X is lower than that of B_Y at the height of 0-0.6, and higher than that of B_Y at 0.6-1.

Table 2 demonstrates that the averaged Nusselt number of the hot wall corresponding to B_Y is higher than that of B_X under any working condition. The difference becomes more apparent as

the Grashof number increases. Thus, we conclude that B_X has greater inhibitory effects on heat transfer than B_Y.
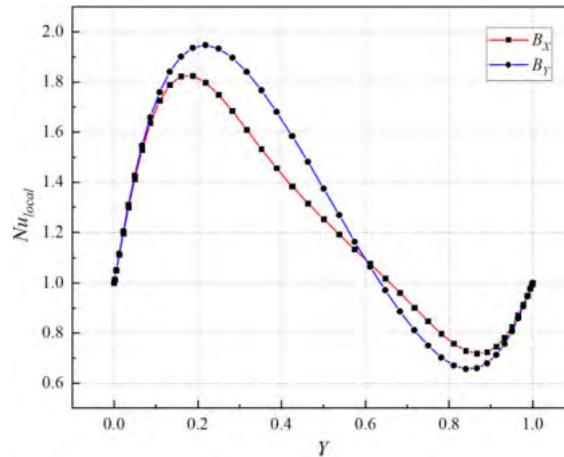
Fig. 25 The distribution of local Nusselt number alone the line (0, *Y*, 0.5) under the field of B_X and B_Y when Gr=1×10⁵.

Table 2 The averaged Nusselt number on the hot wall under different magnetic field directions.

Gr	Ha	Nu_{avg}	
		B_Y	B_X
10^4	0	1.367	1.367
	50	1.003	1.0025
10^5	0	2.785	2.785
	50	1.209	1.175
	100	1.021	1.013
10^6	0	5.831	5.831
	50	3.918	3.258
	100	2.049	1.638

5. Conclusion

In this paper, the MHD convection is investigated in a three-dimensional rectangular cavity filled with a conductive fluid. One pair of walls is maintained at a constant temperature and the others are thermally conducting. The flow and heat transfer equations are solved directly using the high-precision SCM-ACM, and the effects of magnetic field strength, directions, and Grashof numbers on the temperature and flow fields are analyzed. Based on the analysis, the following conclusions are drawn.

(1) For the temperature field, as the Grashof number increases, a stable thermal stratification forms at the center of the cubic cavity, and thermal boundary layers form near the horizontal walls. The increase in Grashof number enhances the heat transfer rate and increases the temperature difference between the upper hot fluid and the lower cold fluid in the cavity. The Hartmann number stabilizes the flow and weakens the heat transfer. The magnetic effects become more significant at higher Grashof numbers.

(2) For the flow field, the increase in Grashof number enhances the convective strength near the wall, resulting in the formation of more vortices inside the square cavity. Under the combined action of centrifugal force and inertial force, these vortices move towards the

corners. When the Hartmann number increases from 0 to 50, significant changes occur in the flow and temperature fields, but further increasing the Hartmann number has less influence. When $Ha > 60$, the velocity in the third direction approaches zero, simplifying the three-dimensional flow into two-dimensional flow.

(3) By comparing the heat transfer rates, we conclude that the magnetic field parallel to the temperature gradient (B_X) has a higher inhibiting effect on heat transfer compared to the magnetic field perpendicular to the temperature gradient (B_Y.)

References

[1] Kabeel, A. E., El-Said, E. M. and Dafea, S. A. (2015) A review of magnetic field effects on flow and heat transfer in liquids: Present status and future potential for studies and applications, *Renewable and Sustainable Energy Reviews* **45**, 830–837.

[2] Davidson, P. A. and Elena, V. B. (2002) An Introduction to Magnetohydrodynamics, *American Journal of Physics* **70**, 781.

[3] Kolsi, L., Oztop, H. F., Alghamdi, A., Abu-Hamdeh, N., Borjini, M. N. and Aissia, H. B. (2016) A computational work on a three dimensional analysis of natural convection and entropy generation in nanofluid filled enclosures with triangular solid insert at the corners, *Journal of Molecular Liquids* **218**, 260–274.

[4] Ismael, M. A., Armaghani, T. and Chamkha, A. J. (2016) Conjugate heat transfer and entropy generation in a cavity filled with a nanofluid-saturated porous media and heated by a triangular solid, *Journal of the Taiwan Institute of Chemical Engineers* **59**, 138–151.

[5] Selimefendigil, F. and Öztop, H. F. (2015) Natural convection and entropy generation of nanofluid filled cavity having different shaped obstacles under the influence of magnetic field and internal heat generation, *Journal of the Taiwan Institute of Chemical Engineers* **56**, 42–56.

[6] Garoosi, F., Hoseininejad, F. and Rashidi, M. M. (2016) Numerical study of natural convection heat transfer in a heat exchanger filled with nanofluids, *Energy* **109**, 664–678.

[7] Spizzichino, A., Zemach, E. and Feldman, Y. (2019) Oscillatory instability of a 3D natural convection flow around a tandem of cold and hot vertically aligned cylinders placed inside a cold cubic enclosure, *International Journal of Heat and Mass Transfer* **141**, 327-345.

[8] Bondareva, N. S. and Sheremet, M. A. (2016) Effect of inclined magnetic field on natural convection melting in a square cavity with a local heat source, *Journal of Magnetism and Magnetic Materials* **419**, 476–484.

[9] Saleh, H. and Hashim, I. (2016) Magnetohydrodynamic natural convection in a rotating enclosure, *Advances in Applied Mathematics and Mechanics* **8**, 279–292.

[10] Rudraiah, N., Barron, R. M., Venkatachalappa, M. and Subbaraya, C. K. (1995) Effect of a magnetic field on free convection in a rectangular enclosure, *International Journal of Engineering Science* **33**, 1075–1084.

[11] Sivasankaran, S. and Ho, C. J. (2008) Effect of temperature dependent properties on MHD convection of water near its density maximum in a square cavity, *International Journal of Thermal Sciences* **47**, 1184–1194.

[12] Sajjadi, H., Delouei, A. A., Atashafrooz, M. and Sheikholeslami, M. (2018) Double MRT Lattice Boltzmann simulation of 3-D MHD natural convection in a cubic cavity with sinusoidal temperature distribution utilizing nanofluid, I*nternational Journal of Heat and Mass Transfer* **126**, 489–503.

[13] Sajjadi, H., Delouei, A. A., Sheikholeslami, M., Atashafrooz, M. and Succi, S. (2019) Simulation of three dimensional MHD natural convection using double MRT Lattice Boltzmann method, *Physica A: Statistical Mechanics and Its Applications* **515**, 474–496.

[14] Zhou, W., Yan, Y., Xie, Y. and Liu, B. (2017) Three dimensional lattice Boltzmann simulation for mixed convection of nanofluids in the presence of magnetic field, *International Communications in Heat and Mass Transfer* **80,** 1–9.

[15] Sheikholeslami, M. and Ellahi, R. (2015) Three dimensional mesoscopic simulation of magnetic field effect on natural convection of nanofluid, *International Journal of Heat and Mass Transfer* **89**, 799–808.

[16] Gelfgat, A. Y. and Zikanov, O. (2018) Computational modeling of magnetoconvection: effects of discretization method, grid refinement and grid stretching, *Computers & Fluids* **175**, 66–82.

[17] Bouchta, S., Feddaoui, M. B. and Dayf, A. (2021) Numerical simulation of free convection in a partially heated three-dimensional enclosure filled with ionanofluid ([C4mim][NTf2]-Cu), *Mathematical Problems in Engineering* **2021**, 1–11.

[18] Okada, K. and Ozoe, H. (1992) Experimental heat transfer rates of natural convection of molten gallium suppressed under an external magnetic Field in either the X, Y, or Z Direction, *Journal of Heat Transfer* **114**, 107–114.

[19] Ozoe, H. and Okada, K. (1989) The effect of the direction of the external magnetic field on the three-dimensional natural convection in a cubical enclosure, *International Journal of Heat and Mass Transfer* **32**,1939–1954.

[20] Zhang, J. K., Cui, M., Li, B. W. and Sun, Y. S. (2020), Performance of combined spectral collocation method and artificial compressibility method for 3D incompressible fluid flow and heat transfer, *International Journal of Numerical Methods for Heat & Fluid Flow* **30**, 5037–5062.

[21] Kajzer, A. and Pozorski, J. (2018) Application of the entropically damped artificial compressibility model to direct numerical simulation of turbulent channel flow, *Computers & Mathematics with Applications* **76**, 997–1013.

[22] Ramachandran, P. and Puri, K. (2019) Entropically damped artificial compressibility for SPH, *Computers & Fluids* **179**, 579–594.

[23] Bogaers, A. E., Kok, S., Reddy, B. D. and Franz, T. (2015) Extending the robustness and efficiency of artificial compressibility for partitioned fluid–structure interactions, *Computer Methods in Applied Mechanics and Engineering* **283**, 1278–1295.

[24] Wang, W., Li, B. W., Varghese, P. L., Leng, X. Y. and Tian, X. Y. (2018) Numerical analysis of three-dimensional MHD natural convection flow in a short horizontal cylindrical annulus, *International Communications in Heat and Mass Transfer* **98,** 273–285.

[25] Zhang, J. K., Dong, H., Zhou, E. Z., Li, B. W. and Tian, X. Y. (2017) A combined method for solving 2D incompressible flow and heat transfer by spectral collocation method and artificial compressibility method, *International Journal of Heat and Mass Transfer* **112**, 289–299.

[26] Chattopadhyay, G., Ranganathan, U. and Millet, S. (2019) Instabilities in viscosity-stratified two-fluid channel flow over an anisotropic-inhomogeneous porous bottom, *Physics of Fluids* **31**, 012103.

[27] Merrill, B. E. and Peet, Y. T. (2019) Moving overlapping grid methodology of spectral accuracy for incompressible flow solutions around rigid bodies in motion, *Journal of Computational Physics* **390**, 121–151.

[28] Raj, R. and Guha, A. (2019) On Bragg resonances and wave triad interactions in two-layered shear flows, *Journal of Fluid Mechanics* **867**, 482–515.

[29] Pourjafar, M., Hejri, A., Bazargan, S. and Sadeghy, K. (2018) On the use of a fluid's elasticity for deliberate rise of Taylor cells in a rotating micro-filter separator, *Physics of Fluids* **30**, 114106.

[30] Fusegi, T., Hyun, J. M., Kuwahara, K. and Farouk, B. (1991) A numerical study of three-dimensional natural convection in a differentially heated cubical enclosure, *International Journal of Heat and Mass Transfer* **34**, 1543-1557.

[31] Tric, E., Labrosse, G. and Betrouni, M. (2000) A first incursion into the 3D structure of natural convection of air in a differentially heated cubic cavity, from accurate numerical solutions, *International Journal of Heat and Mass Transfer* **43**, 4043-4056.

[32] Wakashima, S. and Saitoh, T. S. (2004) Benchmark solutions for natural convection in a cubic cavity using the high-order time–space method, *International Journal of Heat and Mass Transfer* **47**, 853–864.

[33] Luo, X. H., Li, B. W. and Hu, Z. M. (2016) Effects of thermal radiation on MHD flow and heat transfer in a cubic cavity, *International Journal of Heat and Mass Transfer* **92**, 449–466.

[34] Singh, R. J. and Gohil, T. B. (2019) The numerical analysis on the development of Lorentz force and its directional effect on the suppression of buoyancy-driven flow and heat transfer using OpenFOAM. *Computers & Fluids* **179**, 476-489.

Design and load-bearing capacity analysis of bone-inspired lightweight microstructures

*Shuxiang Zhang[1], Jia Li[1], Cong Bi[2], Zehao Wang[2], Dana Tang[1], †Haibin Tang[1]

[1]Nanjing University of Science and Technology, China

[2]Aero Engine Academy of China, China

*Presenting author: shuxiangzhang@njust.edu.cn

†Corresponding author: htang28@njust.edu.cn

Abstract

In the structural design of aero-engines, weight reduction has always been a key point for researchers. Bio-inspired structural design is a potential way to achieve effective weight reduction while ensuring the overall performance of the structure. Inspired by the microstructure of the animal bone, a biomimetic lightweight design is conducted. The load-bearing capacity of bone-inspired lightweight microstructures under different loading conditions are analyzed through numerical simulation. The results show that for the condition where there is weak load along a certain direction, the rectangular-hole bone-inspired lightweight microstructure is the best biomimetic model. Under unknown load conditions, the square-hole bone-inspired lightweight microstructure has the best load-bearing capacity in the weakest direction. The equivalent material parameters of the typical bone-inspired lightweight microstructure are further calculated. With the square-hole bone-inspired lightweight microstructure, a lightweight design of the intermediate casing is performed. The biomimetic lightweight scheme of the intermediate casing reduces the weight by more than 10% compared with the original structure effectively.

Keywords: bone-inspired lightweight microstructure; lightweight; intermediate casing

1. Introduction

Aircrafts operate under extremely complex and harsh service environments, e.g., aerodynamic heating at high Mach numbers, multiple operations in and out of space atmosphere, ultra-long-range cruise flight, ultra-high maneuvering load impact, etc. These pose more stringent requirements on the structural performance. Weight reduction is one of the main ways to improve the performance of the aircrafts. It has been reported that for every 1% reduction in the structural weight of the aircraft, the performance can be improved by 3%-5% [1]. Under that situation, the lightweight design is a development trend in the design of aircrafts [2]-[3].

There are two main design aspects to reduce weight. One method is based on materials [2][4], which means finding or designing new materials with higher specific strength and stiffness. For example, titanium alloy [5] is a light material with low density, good corrosion resistance, high heat resistance, high stiffness and strength. It is an ideal material for aerospace, petrochemical, biomedical and other fields and can be used to improve the ratio of strength to weight. In order to reduce weight, titanium alloy has replaced some aluminum alloy in aircraft parts. Aluminum

magnesium alloy has replaced conventional steel and cast iron for automobiles [6]. In addition to metal alloys, composite materials [7]-[10] are another important material that can be used to reduce weight. They are usually designed to be lighter than single material. The design using composite materials usually has lighter weight than traditional materials. Another method is structural design that is structural improvement or optimization. In this respect the main method is not to change the material itself but to change the material distribution [11], which means optimizing size shape and topology structure. Among them, topology optimization is the most important one, including homogenization method [12], variable density method [13], evolutionary structural optimization [14] and level set method.

Bio-spatial topology is the result of natural selection over billions of years, which reflects the natural evolutionary trend, including optimal spatial configuration, minimal material with the maximum load-bearing capacity. The animal bone [15], the honeycomb [16], and the bamboo [17] are the representative biological structures. In the past several decades, researchers have developed excellent biomimetic structures by mimicking the biological configuration rules. Wang et al. [18] have designed a biomimetic scaffold microstructure based on the principle of structural biomimetics, on the basis of analyzing the configuration characteristics of typical branch structures. Compared with the prototype structure, the mass of the biomimetic microstructure is reduced by 7.35%, and the maximum stress and maximum deformation are significantly reduced. Xing et al. [19] have proposed a biomimetic reinforced frame based on structural optimization by mechanical analysis of shield leaf veins and root structures. The comparison between the biomimetic frame and the traditional frame showed that the total weight is reduced by about 6.0%. If only the weight of the ribs is considered, the weight is reduced by 35.1%. Zhang et al. [20] have evaluated the mechanical properties of the bamboo from a macro perspective, and designed a biomimetic bamboo beam that has lighter mass and higher mechanical performance than traditional beams. Zhang et al. [21] have studied the microstructure and mechanical properties of the ostrich metatarsal bone, and designed a biomimetic tube according to the shape and structural characteristics of the metatarsal bone. The experimental analysis has showed that the biomimetic tube has higher bending resistance and lower mass than ordinary tubular structures. The above-mentioned successful cases of biomimetic microstructure research is meaningful to the design of biomimetic microstructure.

Inspired by the microstructure of the animal bone, the biomimetic lightweight design is performed in the present work. Firstly, the load-bearing capacity of bone-inspired lightweight microstructures under different loading conditions are compared through numerical simulation. Next, the equivalent parameters of the typical bone-inspired lightweight microstructure are calculated. Finally, the optimal bone-inspired lightweight microstructure is applied to the intermediate casing of aero-engine.

2. Bone-inspired design of the lightweight microstructures

The animal bone has good mechanical properties and its microstructure is a typical porous structure. This makes the bone has the characteristics of light weight, high load-bearing capacity and can absorb impact energy well. Inspired by this, the new biomimetic lightweight microstructure design is investigated in the present work. The elliptical-hole, rectangular-hole, and square-hole bone-inspired lightweight microstructures are constructed, as shown in Figure

1(a)-(c). The honeycomb microstructure is exquisite with high strength and light weight, which is composed of many hexagonal cells of the same size. It can bear the load that is much larger than its own weight. Further, the bone-inspired lightweight microstructures that integrate the advantages [22] of the bone and honeycomb are also constructed, i.e., the hexagonal-staggered-hole bone-inspired lightweight microstructure and the polygonal-mixed-hole bone-inspired lightweight microstructure, as shown in Figure 1(d)-(e). The dimensions of these models are set as 20 mm x 20 mm x 20 mm, and the thickness of the outer wall is designed as 1 mm. The filling ratios of the models from Figure 1(a)-(e) are 55.9%, 45.5%, 49.4%, 50.8%, and 53.9%, respectively.
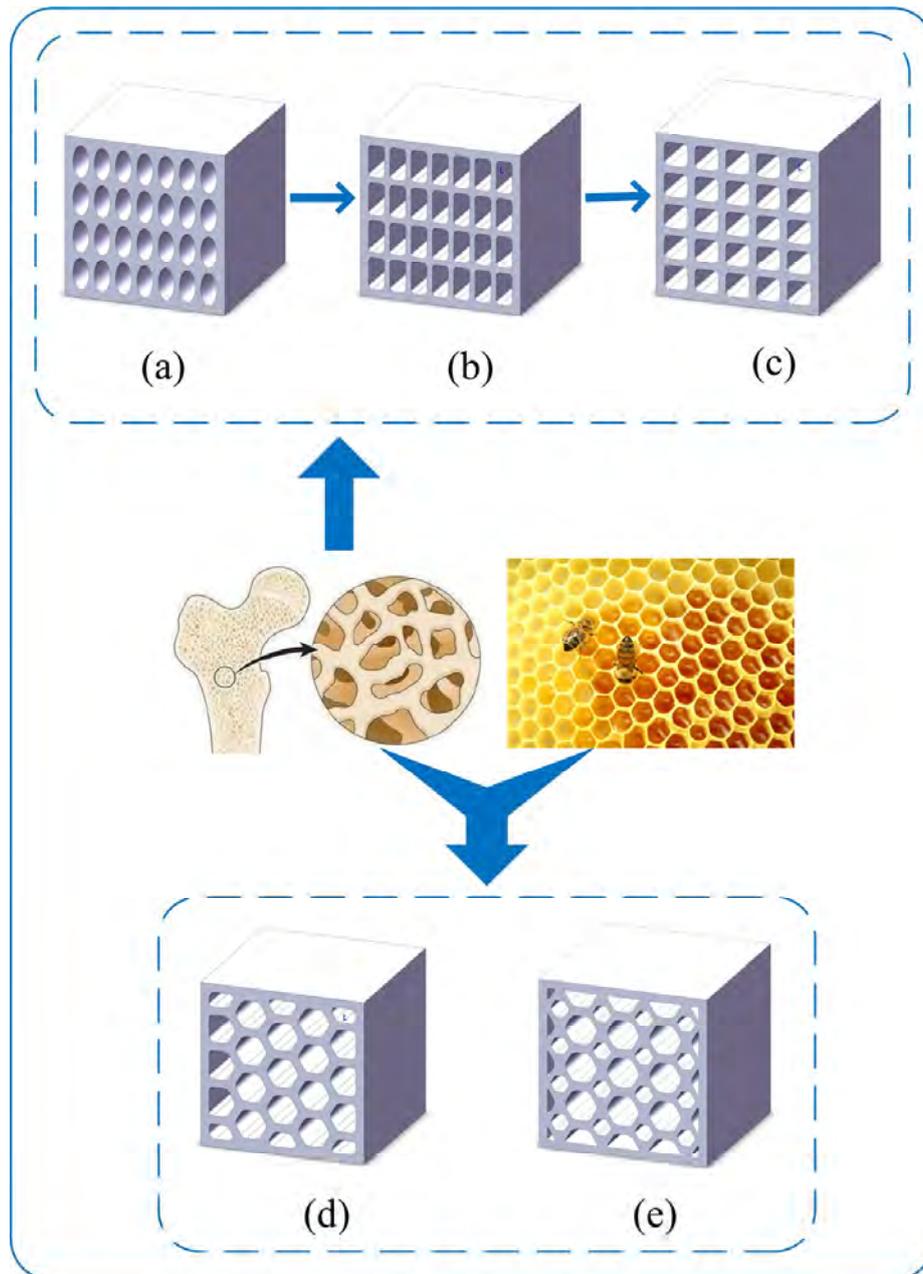
Figure 1. The bone-inspired lightweight microstructure for (a) elliptical-hole, (b) rectangular-hole, (c) square-hole, (d) hexagonal-staggered-hole, and (e) polygonal-mixed-hole

3. Load-bearing capacity analysis of the bone-inspired lightweight microstructures

3.1. Load-bearing capacity calculation of the bone-inspired microstructures

The elliptical-hole, rectangular-hole, square-hole, hexagonal-staggered-hole, and polygonal-mixed-hole bone-inspired lightweight microstructures given in Figure 1 are imported into ABAQUS for mechanical performance analysis, respectively. The material parameters of titanium alloy is adopted for the designed microstructures, and the details have been provided in Ref. [5]. Loads are applied along the Y and Z directions, respectively. The stress contours and stress-strain curves of the models are obtained, as shown in Figure 2(a)-(e) and 3(a)-(e). The reaction force is defined as the maximum load of the biomimetic model, when it undergoes complete plastic deformation. The maximum strength is obtained by dividing the maximum load by the surface area. In order to compare the load-bearing capacity of these models with different filling ratios, the maximum strength need be further normalized. The specific strength is defined by the maximum strength divided by the filling ratio of the model. The load-bearing capacity of these models are finally obtained by the ratio of the specific strength and material strength (complete plastic deformation). For example, the load-bearing capacity of the five bone-inspired lightweight microstructures in Figure 1(a)-(e) along the X direction is 100%.

According to the load-bearing capacity of model defined above, the performance evaluation of biomimetic models introduced in this section is investigated. By extracting the corresponding displacement and reaction force, the load-bearing capacities along the Y direction of elliptical-hole, rectangular-hole, square-hole, hexagonal-staggered-hole and polygonal-mixed-hole bone-inspired lightweight microstructures in Figure 1(a)-(e) are calculated, which are 50.3%, 50.7%, 63.3%, 59.3%, 46.8%, respectively, and those along the Z direction are 64.2%, 73.4%, 63.30%, 49.4%, 52.4%, respectively. The results are listed in Table 1.

Table 1. Load-bearing capacity of the bone-inspired lightweight microstructures

Loading direction	X	Y	Z
elliptical-hole	100%	50.3%	64.2%
rectangular-hole	100%	50.7%	73.4%
square-hole	100%	63.3%	63.3%
hexagonal-staggered-hole	100%	59.3%	49.4%
polygonal-mixed-hole	100%	46.8%	52.4%

In the elliptical-hole bone-inspired lightweight microstructure given in Figure 2(a) and 3(a), there is a significant difference between the load-bearing capacities along the Y and Z directions, which is mainly due to the different cross-sectional areas perpendicular to the Z direction and Y direction of the elliptical-hole bone-inspired lightweight microstructure. From the stress contours of the elliptical-hole model along Y and Z direction in Figure 2(a), it can be seen that the region where the center of two ellipses lies determines the load-bearing capacity of elliptical-hole bone-inspired lightweight microstructure, which is the weakest place. If we want to improve the load-bearing capacity of elliptical-hole bone-inspired lightweight microstructure, we need to make the microstructure as uniform as possible.
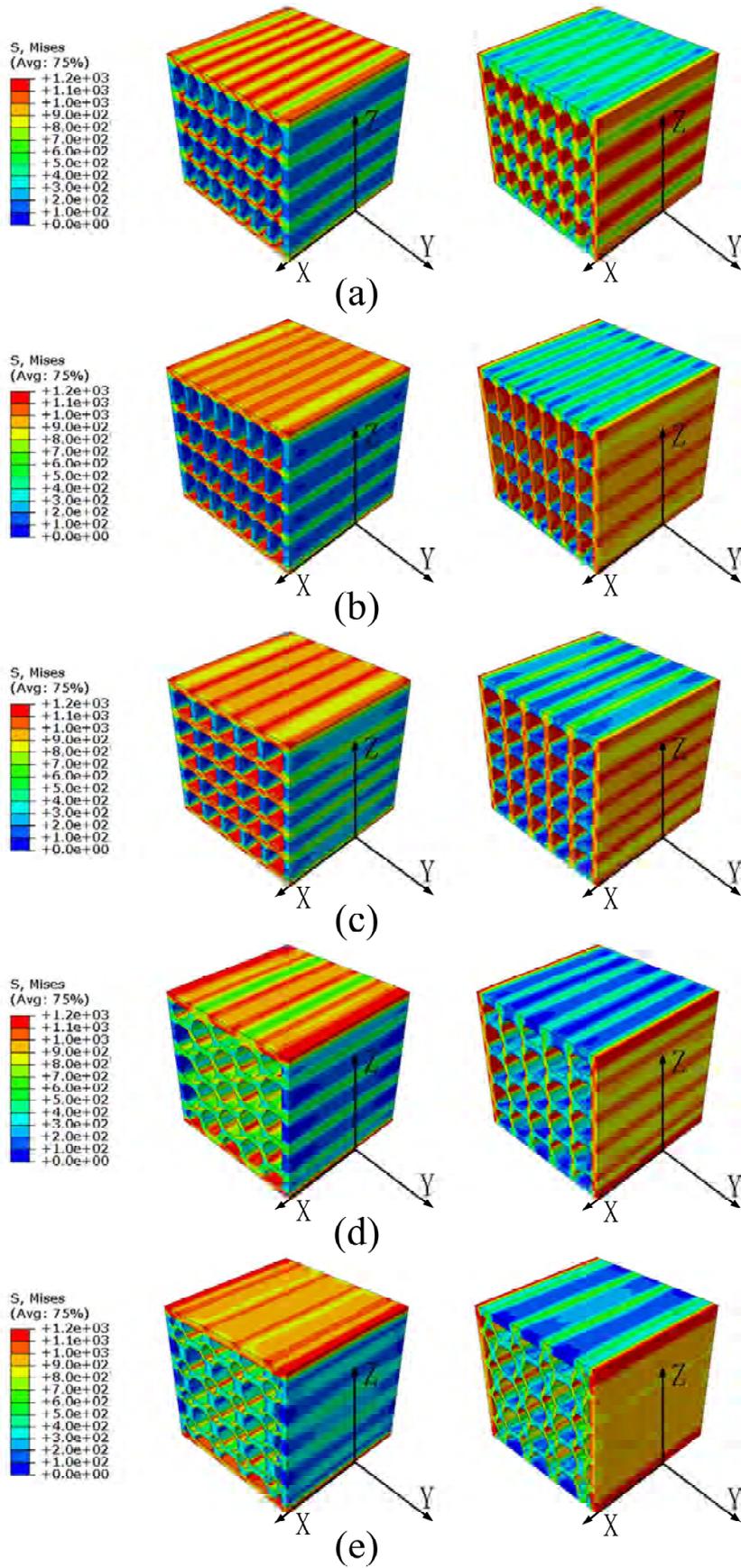
Figure 2. Stress contours of bone-inspired microstructures along the Y and Z directions
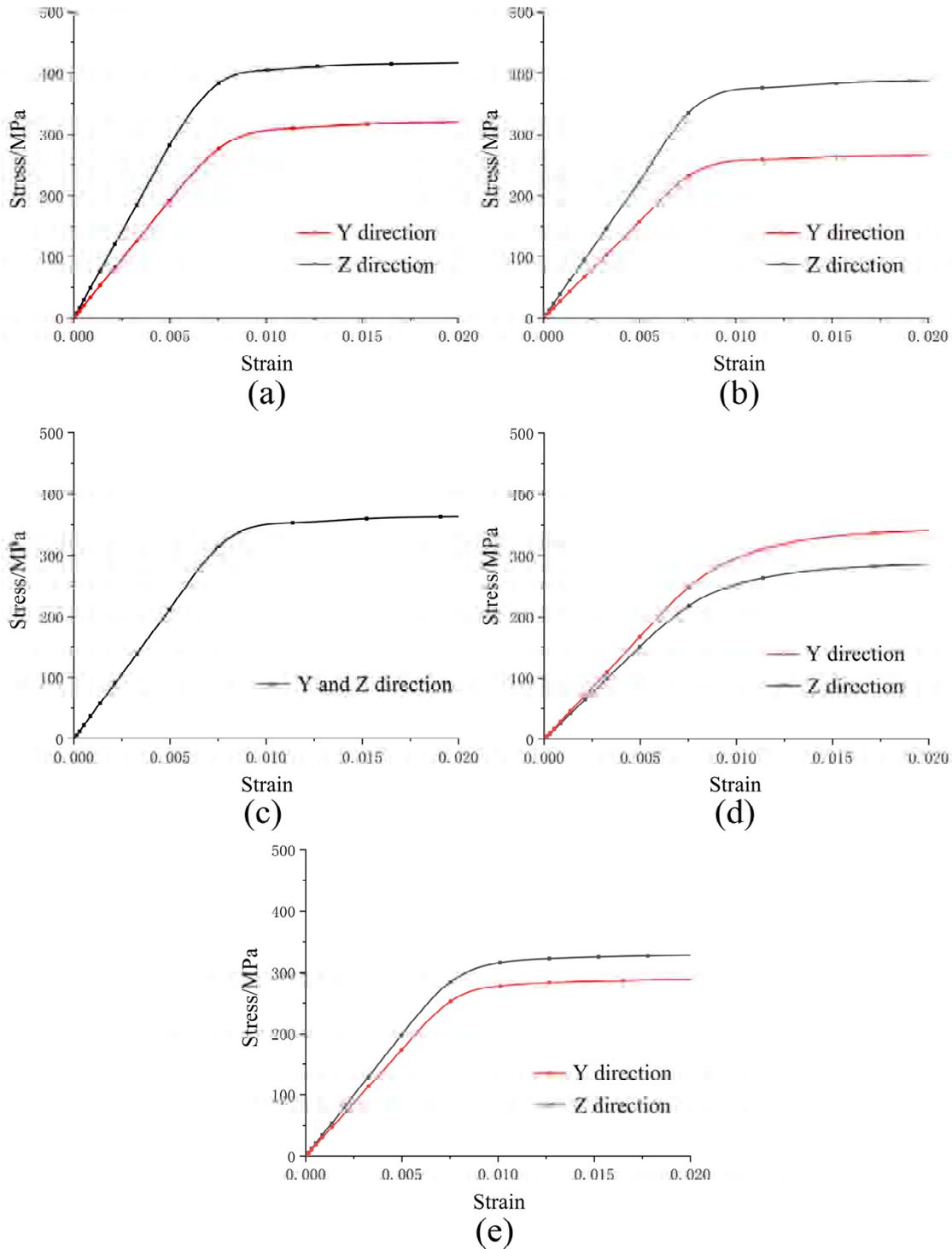
Figure 3. Stress-strain curves of the bone-inspired lightweight microstructures for (a) elliptical-hole, (b) rectangular-hole, (c) square-hole, (d) hexagonal-staggered-hole, and (e) polygonal-mixed-hole

On the basis of the elliptical-hole bone-inspired lightweight microstructure, the rectangular-hole bone-inspired lightweight microstructure is proposed. In Figure 2(b) and Figure 3(b), there is also a significant difference between the load-bearing capacities along the Y and Z directions of the rectangular-hole bone-inspired lightweight microstructure. Similar to the elliptical-hole

bone-inspired lightweight microstructure, the difference between the two results is mainly due to different cross-sectional areas perpendicular to the Z direction and Y direction of the rectangular-hole bone-inspired lightweight microstructure. However, from the calculation results, it can be seen that the rectangular-hole bone-inspired lightweight microstructure given in Figure 2(b) and 3(b) has improved load-bearing capacity compared with the elliptical-hole bone-inspired lightweight microstructure given in Figure 2(a) and 3(a). This is because that the rectangular hole in Figure 1(b) has the same major and minor axes as the elliptical hole in Figure 1(a), and its weak area is the same, but the rectangular hole has a larger area and lower relative filling ratio. According to the calculation method of the load-bearing capacity, under the condition that maximum strength value does not change, the load-bearing capacity of rectangular-hole bone-inspired microstructure in Figure 1(b) is improved.

In order to avoid the anisotropic load-bearing capacities along the Y and Z directions, the square-hole bone-inspired lightweight microstructure is proposed. In Figure 2(c) and 3(c), the load-bearing capacity of the square-hole bone-inspired lightweight microstructure along the Y direction is 63.3%. Since the shape and size of the cross-sections along the Y and Z directions are exactly same for the square-hole bone-inspired lightweight microstructure, the load-bearing capacities along the Y and Z directions are the same. In the biomimetic design, the weak load-bearing capacity along a certain direction should be avoided. Compared with elliptical-hole bone-inspired lightweight microstructure given in Figure 1(a) and rectangular-hole bone-inspired lightweight microstructure given in Figure 1(b), the load-bearing capacity in the weak direction of square-hole bone-inspired lightweight microstructure given in Figure 1(c) is significantly improved.

In Figure 2(d) and 3(d), the load-bearing capacities along the Y and Z directions of hexagonal-staggered-hole bone-inspired lightweight microstructure are 59.3% and 49.4% respectively. It is mainly attributed to the staggered arrangement of the designed holes. There is a weaker area in the Z direction for the hexagonal-staggered-hole bone-inspired lightweight microstructure. Similarly, in Figure 2(e) and 3(e), the load-bearing capacities of polygonal-mixed-hole bone-inspired lightweight microstructure in the Y-axis direction and Z-axis direction are only 46.8% and 52.4%, respectively.

3.2. Equivalent material parameters of the typical bone-inspired microstructure

According to the structural design and analysis in Section 2, under unknown loading conditions, the square-hole bone-inspired lightweight microstructure has optimal load-bearing capacity along its weakest direction. Taking the square-hole bone-inspired lightweight microstructure as an example, its equivalent material property is calculated. The cross-sectional shapes and sizes perpendicular to the Y and Z directions of the square-hole bone-inspired lightweight microstructure are the same, thus elastic moduli along these two directions are also the same. The elastic modulus E is calculated by

$$E = \frac{FL}{A\Delta L} \tag{1}$$

where F is the reaction force during the elastic procedure, ΔL is the displacement, $L = 20$mm is the length of the square, $A = 400$mm^2 is the cross-sectional area of the model. The calculated

elastic moduli E_{11} and E_{22} are 65200MPa and 40109MPa, respectively. The Poisson's ratio is calculated by

$$\varepsilon = \frac{\Delta L}{L}, \quad Nu = \frac{\varepsilon'}{\varepsilon} \tag{2}$$

where ε' is the transverse linear strain, ε is the longitudinal linear strain. The calculated Poisson's ratios Nu_{12} and Nu_{23} are 0.155 and 0.15, respectively. The shear modulus G is calculated by

$$G = \frac{\tau}{\gamma}, \gamma = \tan\theta \tag{3}$$

where τ is the shear stress, γ is the shear strain, θ is the shear angle. The calculated shear moduli G_{12} and G_{23} are 16797Mpa and 24328Mpa, respectively. Table 2 shows the calculated results of the equivalent material parameters for the square-hole bone-inspired lightweight microstructure. The failure of the square-hole bone-inspired lightweight microstructure is determined by the occurrence of the local complete plastic deformation. Thus, the maximum strength is obtained, as shown in Table 3.

Table 2. Equivalent parameters of the square-hole bone-inspired microstructure

Parameter	Value	Parameter	Value	Parameter	Value
E_{11}	65200 MPa	Nu_{12}	0.155	G_{12}	16797 MPa
E_{22}	40109 MPa	Nu_{13}	0.155	G_{13}	16797 MPa
E_{33}	40109 MPa	Nu_{23}	0.150	G_{23}	24328 MPa

Table 3. Maximum strength of the square-hole bone-inspired microstructure

X direction	Y direction	Z direction
583 MPa	369 MPa	369 MPa

4. Biomimetic lightweight design of the intermediate casing

4.1. The principle of the biomimetic lightweight design

In order to achieve effective biomimetic lightweight design of original solid structures, i.e., the intermediate casing of an aero-engine, the failure location and actual performance of the intermediate casing should not be changed in the biomimetic lightweight design. Therefore, during the application of the square-hole bone-inspired lightweight microstructure in the intermediate casing, the original failure region and the stress distribution of the replaced region need be checked simultaneously.

4.2. Failure analysis of the intermediate casing

According to the load analysis of the intermediate casing under the operating condition, it is mainly subjected to the surface pressure and axial load. A designed boundary condition is given to the intermediate casing model. The non-stress concentration regions are selected for the biomimetic lightweight replacement, as plotted in Figure 4. The volume of the replaced regions accounts for 23.5% of the total structure. Since the filling ratio of the used biomimetic microstructure is 49.4%, the weight of the whole casing decreases by more than 10%.

The material parameters of the unreplaced regions remain unchanged, while those of the replaced regions are from the equivalent calculation of the square-hole bone-inspired lightweight microstructure. The coordinates of the material are described by the orthogonal coordinates, and the direction with the strongest load-bearing capacity of the biomimetic microstructure is the same as the axial direction of the intermediate casing. The same boundary condition is applied to the new intermediate casing model, and the calculation result is shown in Figure 5b. It can be found that the failure position before and after replacing material parameters in the intermediate casing model does not change, by the comparison of Figure 5(a) and (b).

In addition to the failure analysis of the whole intermediate casing model, it is necessary to further verify whether there is structural failure at the replaced regions. The stress contour of the replaced regions is shown in Figure 5(c), and the maximum stresses along the X, Y, and Z directions are shown in Table 4. It can be concluded that the stress at the replaced regions with the square-hole bone-inspired lightweight microstructure is not remarkable, which is much lower than the maximum strength.

Table 4. Stress comparison of the replaced regions

Direction	Maximum Stress	Failure stress
X	343 MPa	583 MPa
Y	77 MPa	369 MPa
Z	97 MPa	369 MPa
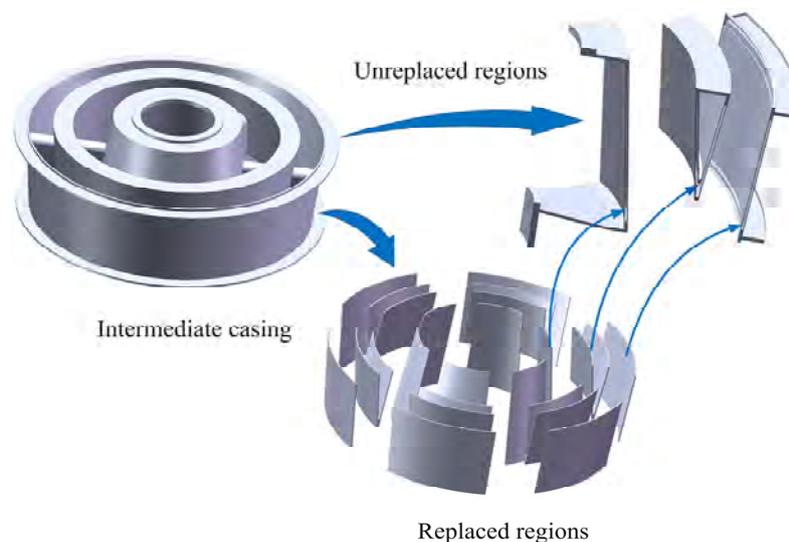
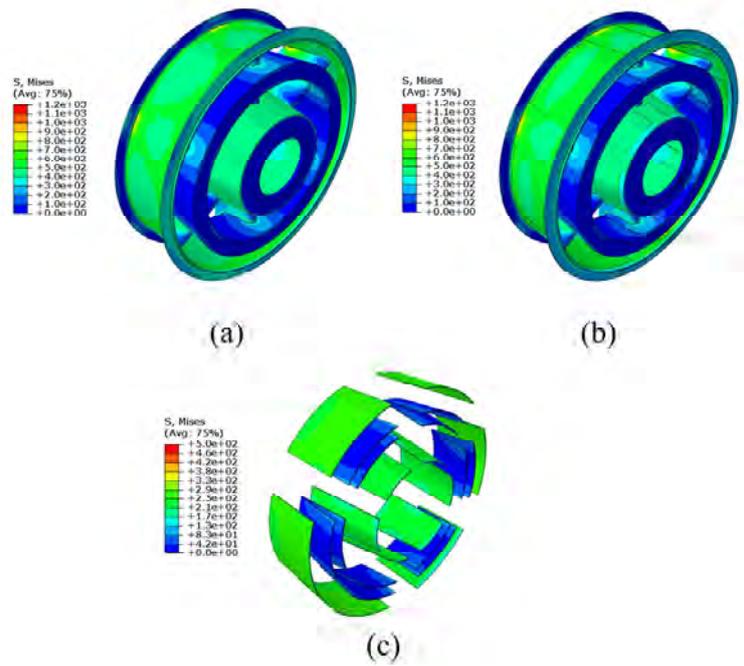**Figure 4. The replaced regions in the intermediate casing model**

Figure 5. Stress contours of the intermediate casing for (a) original intermediate casing, (b) replaced intermediate casing, and (c) replaced regions

5. Conclusion

Based on the biomimetic design theory, the structural design and load-bearing capacity analysis is performed in the present work. The correlation between load-bearing performance and weight reduction ratio is analyzed by numerical simulation. In terms of biomimetic lightweight design, the square-hole bone-inspired lightweight microstructure has optimal load-bearing capacity along its weakest direction. Therefore, under unknown load conditions, the square-hole bone-inspired lightweight microstructure is the best choice. For conditions where load along a certain direction is weak, the rectangular-hole bone-inspired lightweight microstructure is a good choice. During the intermediate casing lightweight design, the square-hole bone-inspired lightweight microstructure is selected for lightweight design. The biomimetic lightweight scheme of the intermediate casing reduces weight by more than 10% compared with the original structure.

Acknowledgement

References

[1] Wang, W., Yuan, L. and Wang, X. (2015) Macrostructure lightweight analysis of aircraft additive manufacturing parts, *Aircraft Design* **35(03)**, 24-28.
[2] Immarigeon, J. P., Holt, R. T., Koul, A. K., Zhao, L., Wallace, W. and Beddoes, J. C. (1995) Lightweight materials for aircraft applications, *Materials Characterization* **35(1)**, 41-67.
[3] Smith, S. and Tlusty, J. (1997) Current trends in high-speed machining, *Journal of Manufacturing Science*

and Engineering **119**, 664–666.

[4] Altenbach, H. (2011). Mechanics of advanced materials for lightweight structures, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* **225(11)**, 2481-2496.

[5] Tang, H., Huang, H., Liu, C., Liu, Z. and Yan, W. (2021) Multi-Scale modelling of structure-property relationship in additively manufactured metallic materials, *International Journal of Mechanical Sciences* **194**, 106185.

[6] Yang, H., Li, L. X., Wang, Q. D. and Guo, L. G. (2010) Research on scientific and technological development in the field of light alloy forming, *Journal of Mechanical Engineering* **46(12)**, 12.

[7] Tang, H., Chen, H., Sun, Q., Chen, Z. and Yan, W. (2021) Experimental and computational analysis of structure-property relationship in carbon fiber reinforced polymer composites fabricated by selective laser sintering, *Composites Part B: Engineering* **204**, 108499.

[8] Tang, H., Sun, Q., Li, Z., Su, X. and Yan, W. (2021) Longitudinal compression failure of 3D printed continuous carbon fiber reinforced composites: an experimental and computational study, *Composites Part A: Applied science and manufacturing* **146**, 106416.

[9] Tang, H., Chen, Z., Xu, H., Liu, Z., Sun, Q., Zhou, G., Yan W., Han, W. and Su, X. (2020) Computational micromechanics model based failure criteria for chopped carbon fiber sheet molding compound composites, *Composites Science and Technology* **200**, 108400.

[10] Tang, H., Zhou, G., Sun, Q., Avinesh, O., Meng, Z., Engler-Pinto, C. and Su, X. (2021) Experimental and computational analysis of bending fatigue failure in chopped carbon fiber chip reinforced composites, *Composite Structures* **275**, 114402.

[11] Tang, D., Yang, K., Gao, T., Liu, T. and Tang, H. (2023) Mechanical-electromagnetic integration design of Al2O3/SiO2 ceramic cellular materials fabricated by digital light processing, *Thin-Walled Structures* **183**, 110437.

[12] Bendsøe, M. P. and Kikuchi, N. (1988) Generating optimal topologies in structural design using a homogenization method, *Computer Methods in Applied Mechanics and Engineering* **71(2)**, 197-224.

[13] Bendsøe, M. P. (1989) Optimal shape design as a material distribution problem, *Structural Optimization* **1**, 193-202.

[14] Xie, Y. M. and Steven, G. P. (1993) A simple evolutionary procedure for structural optimization, *Computers & Structures* **49(5)**, 885-896.

[15] Wu, Y., Bionic lightweight design of aircraft rudder wing structure and its mechanical properties, Master's thesis, Nanjing University of Aeronautics and Astronautics, Nanjing, 2017.

[16] Bao, N., Ma, J., Zhang, X. and Zhong, Z. (2016, July) Structural bionic lightweight design for the stiffened plate of base structure, *In 2016 7th International Conference on Mechanical and Aerospace Engineering* 244-247.

[17] Wang, X. and Song, B. (2022) Application of bionic design inspired by bamboo structures in collapse resistance of thin-walled cylindrical shell steel tower, *Thin-Walled Structures* **171**, 108666.

[18] Wang, Y. H. and Chen, W. Y. (2012) Lightweight design of aircraft support based on structural bionics, *Journal of Mechanical Design* **29**, 77-79.

[19] Xing, D., Chen, W., Xing, D. and Li, Z. (2013) Lightweight design for an aircraft reinforcing frame based on structural bionics. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* **227(8)**, 1338-1346.

[20] Zhang, T., Wang, A., Wang, Q. and Guan, F. (2019) Bending characteristics analysis and lightweight design of a bionic beam inspired by bamboo structures. *Thin-Walled Structures* **142**, 476-498.

[21] Zhang, R., Pang, H., Han, D., Xue, S., Jiang, L., Li, D. and Li, J. (2020) Bionic design in anti-bending and lightweight tube based on the tarsometatarsus of ostrich. *Rendiconti Lincei Scienze Fisiche e Naturali* **31**, 189-201.

[22] Yang, Y., Zhu, Q. X., Wang, W. and Peng, X. (2021) Structure bionic design method oriented to integration of biological advantages. *Structural and Multidisciplinary Optimization* **64(3)**, 1017-1039.

Inverse scattering technique using deep learning for 3-D scalar wave propagation

†*T. Saitoh[1] and S. Hirose[2]

[1]Department of Civil and Environmental Engineering, Gunma University, Japan
[2]Department of Civil and Environmental Engineering, Tokyo Institute of Technology, Tokyo, Japan

*Presenting author: t-saitoh@gunma-u.ac.jp
†Corresponding author: t-saitoh@gunma-u.ac.jp

Abstract

Accurately identifying the position, size, and shape of defects in materials is important for structural monitoring and health diagnostics. The methods for defect shape reconstruction in materials have been developed by several researchers over the years. The inverse scattering technique is known as an effective method for reconstructing a defect and has been applied to many engineering problems. However, in general, the mathematical formulation of the inverse scattering is highly complex, and reconstructing a defect using it is relatively time-consuming. In this study, we propose the utilization of deep learning for inverse scattering analysis to improve computational time and address the problem associated with mathematical treatment. The training data for deep learning is created using the convolution quadrature time-domain boundary element method (CQBEM). The proposed inverse scattering method is validated through several numerical examples.

Keywords: Boundary element method, Time-domain, Defect shape reconstruction, Inverse scattering, 3-D scalar wave propagation.

Introduction

The use of ultrasonic non-destructive testing is widespread in providing safety assurance for structural materials. Accurately identifying the position, size, and shape of defects in materials is important for structural monitoring and health diagnostics. The methods for defect shape reconstruction in materials have been developed by several researchers over the years[1][2][3][4]. The inverse scattering technique[5][6][7] is known as an effective method for reconstructing a defect and has been applied to many engineering problems. However, in general, the inverse scattering formulation is mathematically very complicated and reconstructing a defect using it is relatively time-consuming. The inverse scattering technique is not as commonly used in practical engineering applications compared to the synthetic aperture focusing technique (SAFT)[8][9], which is the most widely used method for defect detection and imaging.

In recent years, the deep learning has attracted attention in many engineering fields. Deep learning is a type of machine learning that involves training artificial neural networks to learn from large amounts of image data. It has proven to be particularly effective for solving image classification problems[10]. The use of deep learning for ultrasonic non-destructive testing has been considered as an effective way to reduce the labor required by inspectors. Meng et al. [11] applied deep learning techniques to develop a defect imaging system for composite materials. Saitoh et al. [12] employed deep learning to solve an inverse scattering problem for a scatterer in an infinite region. This method uses deep learning of B-scope images at receiver points to predict the size and position of a defect. However, it has a limitation as it is only applicable for 2-D scalar wave propagation.
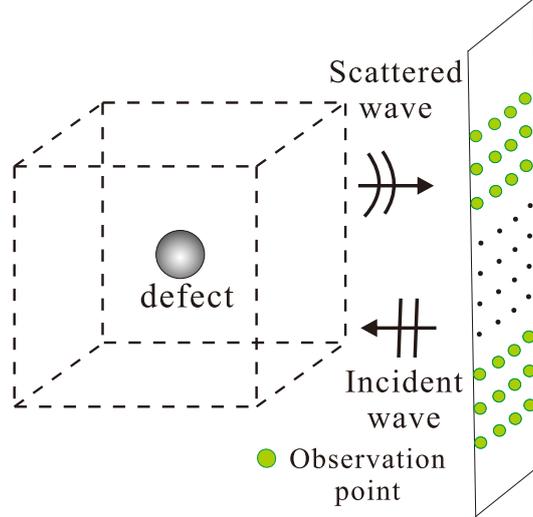
Figure 1: Inverse scattering model.

Therefore, in this research, we extend the application of the inverse scattering technique using deep learning from a 2-D case to a 3-D case. Image data for deep learning is created by using the convolution quadrature time-domain boundary element method (CQBEM). The proposed inverse scattering method is validated through several numerical examples.

Convolution Quadrature Time-Domain Boundary Element Method (CQBEM)

Using waveforms obtained from actual experiments for inverse scattering analysis is highly desirable. However, it is generally the case that waveforms obtained from measurement experiments contain uncertain factors, such as noise. Therefore, in this study, image data for deep learning is created by using the convolution quadrature time-domain boundary element method (CQBEM). The CQBEM is known as a stable time-domain boundary element method (BEM) better than the classical time-domain BEM[13]. In the CQBEM, the time-discretization is carried out using the convolution quadrature method (CQM) proposed by Lubich[14]. In general, the time-domain boundary integral equation at time t for 3-D scalar wave propagation is written as follows:

$$C(\boldsymbol{x})p(\boldsymbol{x},t) = p^{\mathrm{in}}(\boldsymbol{x},t) + \int_S G(\boldsymbol{x},\boldsymbol{y},t) * q(\boldsymbol{y},t)dS_y - \int_S H(\boldsymbol{x},\boldsymbol{y},t) * p(\boldsymbol{y},t)dS_y \quad (1)$$

where $C(\boldsymbol{x})$ is the free-term depends on the boundary surface S, p is the pressure, $p^{\mathrm{in}}(\boldsymbol{x},t)$ is an incident wave. The variable q is defined as the partial derivative of $p(\boldsymbol{y},t)$ with respect to the unit outward normal vector $\boldsymbol{n}$ on the boundary surface S, expressed as $q = \partial p(\boldsymbol{y},t)/\partial n$. In addition, $*$ is the covolution. $G(\boldsymbol{x},\boldsymbol{y},t)$ and $H(\boldsymbol{x},\boldsymbol{y},t)$ are the fundamental solution and its double layer kernel, respectively, for 3-D scalar wave propagation. By discretizing Eq.(1) for time and space using the CQM and the collocation method, respectively, and solving the resulting equation, it is possible to obtain the pressure field $p(\boldsymbol{x},t)$ around a defect. You may refer to the papers discussing CQBEM for scalar wave propagation to obtain more detailed information on the method[15][16][17][18].

Problem Statements and Conditions

This section describes the problem statements and conditions for the inverse scattering analysis. Using the CQBEM described in the previous section, we solve the scattering problem of an in-
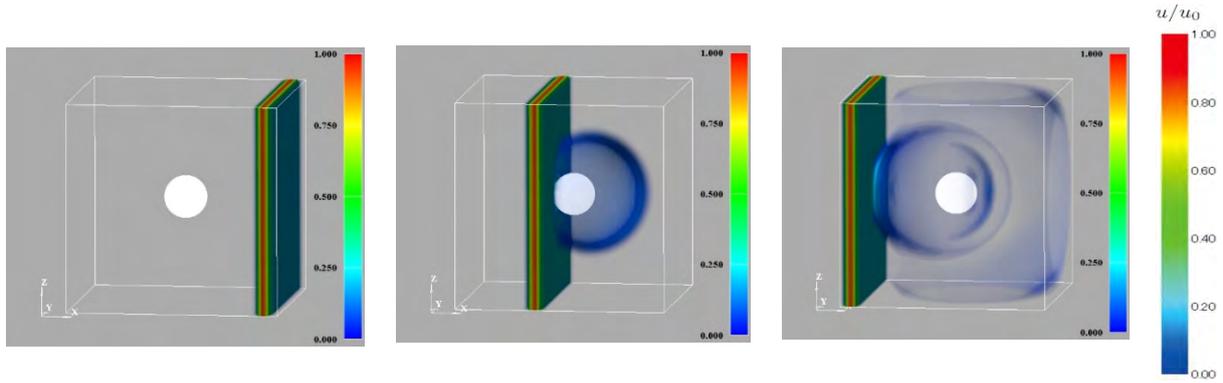
Figure 2: 3-D scalar wave fields around a defect obtaind by the CQBEM.

cident plane wave by a defect with a radius of a in 3-D infinite space, as illustrated in Figure 1. The boundary condition on the surface S of a defect is considered as $q(\boldsymbol{x}, t) = 0$. The scattered waves are received at the observation points, as shown in Fig.1. The total pressure waveforms including scattered pressure waves at each observation point are utilized for this inverse scattering analysis. The observation points are designed to replicate a grid pattern similar to that of a matrix array transducer used in ultrasonic non-destructive testing. It is assumed that there is a single defect present within the analysis domain.

Fig.2 shows an example of the pressure field $p(\boldsymbol{x}, t)$ around the defect with the radius a obtained by using the CQBEM. In Fig.2, the defect is represented as a white sphere. On the left side of Fig. 2, we can observe the incident plane wave before it interacts with the defect. The scattered pressure wave is confirmed in both center and right of Fig.2. The scattered waves generated by the incident plane wave hitting the defect propagate to the far field without producing any other scattered waves. These scattered pressure waves $p(\boldsymbol{x}, t)$ are received at each observation point. The received waveforms are utilized for this inverse scattering analysis, as shown in the following section.

Training data for CNN

To estimate an unknown sphere defect position (x, y, z), which is the center coordinate of the defect in 3-D infinite space, the deep learning[10][19] is implemented in this study. As mentioned in the previous section, the scattered pressure waves are observed at each observation point. For example, at a certain set of observation points, it is possible to obtain scattered pressure waveforms $p(\boldsymbol{x}_i, t)$, as shown in the upper middle of Fig.3. Furthermore, it is straightforward to arrange such received waveforms at all observation points in a row and display them as an image data, as shown in the upper right of Fig.3. Indeed, image data is well-suited for the convolutional neural network (CNN). Such image data can be easily obtained by varying the position and size of the defect in the CQBEM and it is possible to obtain a large amount of such image data. In this study, such image data obtained by the CQBEM are used for training, validation, and testing in the CNN.

As mentioned before, in this problem, we assume an infinite domain for the wave propagation region. If the methods, such as the finite difference method (FDM) or finite element method (FEM), are used, instead of the CQBEM, to solve such wave scattering problems, it is necessary to set up absorbing boundaries to accurately represent the far-field region[20]. However, it is generally challenging to set up a universally perfect absorbing boundary that works effectively

for all types of waves. When using methods such as FDM or FEM, there is a possibility that deep learning might learn the physically unrealistic wave behavior caused by setting up an artificial absorbing boundary condition.

In general, during the CNN training phase, as shown in the lower part of Fig.3, labels such as the horizontal and vertical axes in the images are unnecessary for the deep learning. Therefore, they are removed in advance before performing the training and validation. The CNN architecture used in this analysis can be seen in Fig.4. In the final output, the set of the center coordinate of the defect position (x, y, z) is calculated using the linear function as the activation function.

Results and Discussions

Some numerical examples are shown in this section. 9261 image data are prepared and 80% of them are used for training and the rest 20% for validation. The mini-batch size is 32 and the input image size is 224×224. The hyper parameters, weight decay and learning rate are 1.0e-4. Adam is used in the learning process. GPGPU, Geforce RTX3090 (with 24GB memory), is used for the acceleration of this CNN computation. After conducting training and validation for prepared 9261 image data, the deep learning model is created. The testing for unknown image data, which are not used in the training and validation steps, is performed to determine whether the created deep learning model can accurately estimate the unknown defect position (x, y, z).

Fig.5 shows the results for the proposed inverse scattering. The number of observation points used for this inverse scattering analysis is 64, and they are placed on the $x = 1$ plane as shown in Fig.1. The spacing between the observation points in the y and z directions is $\Delta x = 0.1$, and 8 observation points are arranged in each of the y and z directions. For example, the coordinates of the observation point that serves as the starting point for the grid arrangement are $(x, y, z) = (1, 1, 1)$. In each of Fig.5, the gray and red spheres show the correct and estimated position of the defect. To quantitatively verify the position of the defect, both the predicted and actual coordinates of the defect's center (x, y, z) are also indicated. Note that the results in Fig.5 are normalized by the radius a of the defect. As shown in Fig.5, it can be observed that the estimated position of the defect closely matches the actual position of the defect. Therefore, it can be said that this method has sufficient applicability to estimate the approximate position
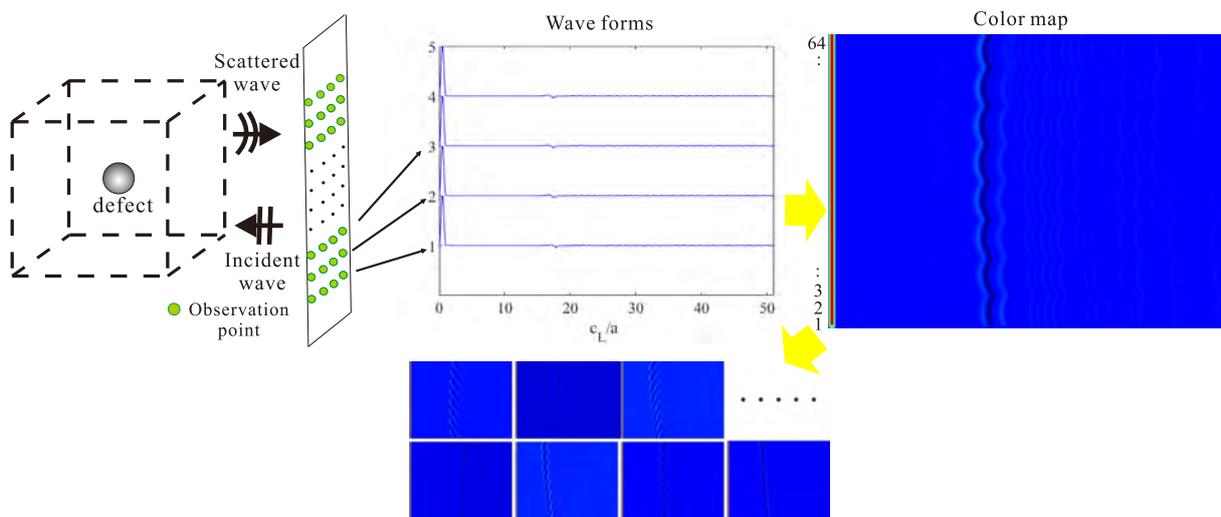

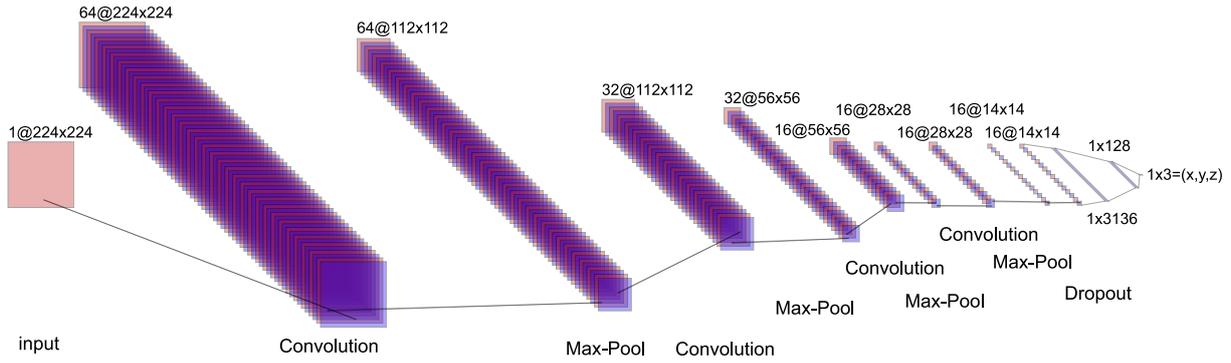**Figure 3: Overview diagram for creating deep learning data.**

Figure 4: CNN architecture used in this problem.

(x, y, z) of a defect.

Conclusions

In this paper, we presented the inverse scattering technique using deep learning for a defect in 3-D infinite space. The 3-D scalar wave propagation and scattering are considered for this formulation. Our results demonstrated that the deep learning-based inverse scattering technique can accurately estimate the defect position in 3-D infinite space. In this method, using deep learning for defect position estimation requires running numerous simulations to generate the necessary training data. This process can indeed demand a significant amount of computation time due to the large number of simulations involved. However, once the deep learning model is created, the time required to estimate the position of an unknown defect is minimal. The trained model can quickly process new input data and provide predictions efficiently, making the estimation process significantly faster compared to the conventional inverse scattering techniques.

In the future, to improve the accuracy of the deep learning model, we will make more big image data sets. In addition, the proposed method for a defect will be extended to that for a defect in 3-D elasic infinite space where P and S waves must be considered. Furthermore, estimating the position of defects based on the results obtained from actual measured waveforms is also our future task.

Acknowledgement

References

[1] Fink, M. (1992) Time reversal of ultrasonic fields - part I: Basic principles, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **39**(5), 555–566.
[2] Dominguez, N., Gibiat, V. and Esquerre, Y. (2005) Time domain topological gradient and time reversal analogy: an inverse method for ultrasonic target detection, *Wave Motion* **42**(1), 31–52.
[3] Bonnet, M. (2006) Topological sensitivity for 3d elastodynamic and acoustic inverse scattering in the time domain, *Computer Methods in Applied Mechanics and Engineering*, **195**, 5239–5254.
[4] Kirsch, A. (2002) The music-algorithm and the factorization method in inverse scattering theory for inhomogeneous media, *Inverse Problems* **18**(4), 1025.
[5] Colton, D., Coyle, J. and Monk, P. (2000) Recent developments in inverse acoustic scattering theory, *SIAM Review* **42**(3), 369–414.
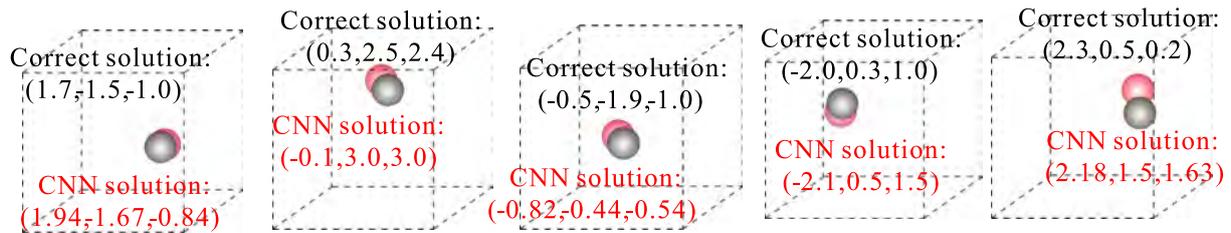
Figure 5: Defect position estimation results using CNN.

[6] Kitahara, M., Nakahata, K. and Hirose, S. (2002) Elastodynamic inversion for shape reconstruction and type classification of flaws, *Wave motion* **36**, 443–455.

[7] Nakahata, K., Saitoh, T. and Hirose, S. (2006) 3-D flaw imaging by inverse scattering analysis using ultrasonic array transducer, *Review of Progress in Quantitative Nondestructive Evaluation* **34**, 717–724.

[8] Langenberg, K. J., Berger, M., Kreutter, T., Mayer, K. and Schmitz, V. (1986) Synthetic aperture focusing technique signal processing, *NDT International* **19**(3), 177–189.

[9] Spies, M., Rieder, H., Dillhöfer, A., Schmitz, V. and Müller, W. (2012) Synthetic aperture focusing and time-of-flight diffraction ultrasonic imaging-past and present, *Journal of Nondestructive Evaluation* **31**(4), 310–323.

[10] Simonyan, K. and Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition, *CoRR*, abs/1409.1556.

[11] Meng, M., Chua, Y. J., Wouterson, E. and Ong, C. P. K. (2017) Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks, *Neurocomputing* **257**, 128–135.

[12] Saitoh, T., Sasaoka, S. and Hirose, S. (2022) Deep learning based 2-D inverse scattering analysis for estimation of position and size of a defect in solid, *Artificial Intelligence and Data Science* **3**(J2), 935–944 (in Japanese).

[13] Mansur, W. J. and Brebbia, C. A. (1983) Transient elastodynamics using a time-stepping technique, *C.A. Brebbia, T. Futagami, M. Tanaka (Eds.), Boundary Elements*, 677-698.

[14] Lubich, C. (1988) Convolution quadrature and discretized operational calculus I, *Numerische Mathematik* **52**(2), 129–145.

[15] Schanz, M. and Antes, H. (1987) Application of operational quadrature methods in time domain boundary element methods, *Meccanica* **32**, 179–186.

[16] Abreu, A., Carrer, J. and Mansur, W. (2003) Scalar wave propagation in 2D: a BEM formulation based on the operational quadrature method, *Engineering Analysis with Boundary Elements* **27**(2), 101–105.

[17] Saitoh, T. and Hirose, S. (2010) Parallelized fast multipole bem based on the convolution quadrature method for 3-D wave propagation problems in time-domain, *IOP Conf. Series: Materials Science and Engineering*, 10:012242.

[18] Saitoh, T., Hirose, S., Fukui, T. and Ishida, T. (2007) Development of a time-domain fast multipole bem based on the operational quadrature method in a wave propagation problem, *Advances in Boundary Element Techniques VIII*, 355–360.

[19] Chollet, F. (2017) *Deep learning with Python*, Manning Publications.

[20] Pled, F. and Desceliers, C. (2022) Review and recent developments on the perfectly matched layer (PML) method for the numerical modeling and simulation of elastic wave propagation in unbounded domains, *Archives of computational methods in engineering*, HAL–03196974.

Construction and verification of composite intensity measures for multi-storey subway station structures based on partial least squares regression

*Wei Yu[1] , †Zhi Yi Chen[1,2]

[1]Department of Geotechnical Engineering, Tongji University, 1239, Siping Road, Shanghai, China
[2]State Key Laboratory of Disaster Reduction in Civil Engineering, Tongji University, 1239 Siping Road, Shanghai, China

*Presenting author: yuwei2020@tongji.edu.cn

†Corresponding author. Email address: zhiyichen@tongji.edu.cn

Abstract

In order to quantitatively evaluate the seismic damage potential of ground motions (GMs) on multi-storey subway station structures and solve the problem of that single intensity measures (IMs) are inadequate in accurately representing the seismic damage potential of GMs, this paper takes the input GMs in the seismic design of multi-storey subway station structures as the research object and constructs composite intensity measures (IMs) that can effectively characterize the damage potential of input GMs on a typical four-storey and three-span subway station structures. Considering the various seismic response of underground structures in different characteristic period type sites, the soil-subway station structure system is divided into four period bands according to the characteristic period of the system. Then, four representative period soil-structure systems in four period bands are selected. Subsequently, 40 GMs were selected for elastoplastic numerical analysis, and the results of numerical analysis were used as sample data for partial least squares regression (PLSR) to construct composite IMs corresponding to each period bands soil-structure systems. Finally, 40 addition GMs are used to verify the correlation between the composite IMs and the seismic damage of multi-storey subway station structures. The results show that the correlation between the composite IMs and the seismic damage of multi-storey subway station structures is better than that of commonly used single IMs.

Keywords: Subway station structures, Damage potential of ground motions, Composite intensity measures, Characteristic periods, Partial least squares regression.

Introduction

In recent decades, several major earthquakes have occurred near large cities, such as the Hanshin earthquake in Japan and the Chichi earthquake in Taiwan, China, have causing huge casualties and economic losses to human society. Among them, there are also damages to underground structures [1, 2]. With the rapid development of the economy and construction, underground structures are gradually developing towards deep and multi-storey. Compared with single or double-storey structures, multi-storey underground structures are more complex, and their seismic damages are more difficult to repair. Therefore, when protecting underground structures from earthquake damage, not only should the seismic isolation technology be fully utilized, but also the most unfavorable input GMs should be considered to ensure that the underground structure is sufficiently safe under earthquake action. Among them, the most unfavorable input GMs refers to the GMs that can make the underground structure in the most unfavorable state [3]. However, how to obtain such input GMs is not clear. Considering that the intensity measures (IMs) are important parameters in characterizing the danger and damage potential of GMs. Therefore, the input GM can be selected through reasonable IMs. The IM should be able to effectively characterize the damage potential of GMs on underground

structures. However, which IM can be used as the characterization parameter is not clear.

In recent years, many scholars have carried out a series of studies on the correlation between the dynamic response of underground structures and IMs. Chen et al. [4] pointed out that the pulse effect plays a dominant role in structural damage in mountain tunnels and that can be characterized by maximum incremental velocity (V_{mi}). Zhong et al. [5] found that peak ground acceleration (A_{pg}) and acceleration spectrum intensity (V_{si}) are the best seismic motion parameters for predicting the dynamic response of single-storey double-span subway station structures under earthquake excitation. Huang et al. [6] contend that A_{pg} is the optimal IM for shallow-buried tunnel structures, while peak ground velocity (V_{pg}) is the optimal IM for medium-depth and deep-buried tunnel structures. Yang et al. [7] found that A_{pg} is more suitable for predicting the vulnerability of shallow-buried underground structures in hard soil layers, while V_{pg} is more suitable for predicting the vulnerability of shallow-buried underground structures in soft soil layers. The above research aims to provide a basis for selecting input GMs for seismic resistance design of related underground structures and has drawn a large number of meaningful conclusions. In fact, due to the strong randomness of seismic motion, using a single IM will result in a large amount of missing information about GMs, which cannot fully reflect the characteristics of GMs and is not conducive to safe seismic resistance design of engineering structures [8].

In view of the problem that the single IMs are inadequate in accurately representing the seismic damage potential of GMs, Liu et al. [9, 10]adopted multiple commonly used single IMs as independent variables and took maximum interlayer displacement angle θ_{max} as dependent variable for ground frame structure, and constructed a linear regression equation for multiple single IMs through PLSR methods. Then, defined the linear combinations of regression as composite IMs, and proved the effectiveness of the constructed composite IMs. In addition, for different characteristic period soil-underground structure system, the structural seismic response and discreteness of IMs are variable, and no IM can represent the damage potential of soil-underground structure systems with different characteristic periods without any invariance. To this end, soil-underground structure systems can be classified according to characteristic periods, and corresponding composite IMs can be constructed for each type of soil-structural system through PLSR method to characterize structural earthquake responses.

This article takes the input GM of the seismic design of a multi-storey subway station structure of four-storey, three-span as the research object. Using the overall damage index of the underground structure as a quantitative indicator of the damage degree of the underground structure. The characteristic period grouping of the soil-underground structure system was studied, and the composite IMs corresponding to each period type of soil-underground structure system was constructed, and the performance of the composite IMs was verified.

1 Muti-storey subway station numerical model and engineering demand parameter

This paper is based on statistical methods to construct composite IMs. Therefore, it is necessary to conduct elastoplastic dynamic time history analysis of subway stations under a certain number of GMs to provide necessary data samples before conducting research on composite IMs. The data samples include representative IMs of input ground motion, as well as seismic dynamic response data of subway stations under earthquake excitation.

1.1 Soil-structure interaction analytical model and parameters

The prototype of the four-storey three-span island subway station is locating in Shanghai. The standard cross-sectional dimensions of the station structure are shown in Figure 1. The thickness of soil cover is 3.2m, and the width and height of the station are 23.6 m and 29.1 m respectively.
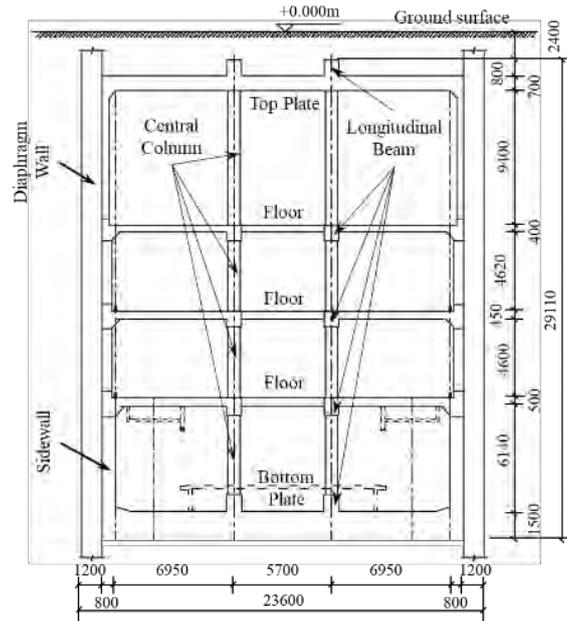
Figure 1: Standard cross-sectional dimensions of subway station structure（unit: mm）

A dynamic soil-structure interaction analysis model was established using the general finite element software ABAQUS. The kinematic hardening Mohr-Coulomb elastoplastic constitutive model was adopted for the soil layer, and the finite element structural model of the soil-underground structure system was shown in Figure 2. The thickness of the model side wall was set as the sum of the thicknesses of the diaphragm wall and the original side wall. According to "Code for Seismic Design of Buildings" in China (GB 50011, 2010) [11], the width of the soil layer on both sides of the structure should be at least three times the width of the structure. In this model, the width of the soil layer on both sides of the structure is 250 meters, and the depth of the soil layer is 85 meters. To facilitate control of the characteristic period of the soil layer, a homogeneous soil layer was proposed. Soil-structure interaction was defined by Coulomb's friction law, with a friction coefficient of 0.4 assumed, corresponding to a friction angle of 22°.

The infinite element is set on both sides of the soil-structure interaction model, which is based on the statistical analysis of Zienkiewicz [12]. The dynamic response analysis of Lysmer and Kuhlemeyer[13]. The infinite element (CINPE4) boundary can effectively reduce the reflection of seismic waves carrying energy at the boundary towards the analysis element area. The central columns were made of C45 grade concrete, and other structural components were made of C35 grade concrete. The reinforcements were insert into the beam element through the *rebar keyword to simulate the mechanical behavior of reinforcement. HRB400 hot-rolled reinforcement was selected for the central column, and HRB335 hot-rolled reinforcement was selected for other parts. The detailed performance parameters of each material were shown in Table 1.

Table1: Material parameters of reinforcement and concrete

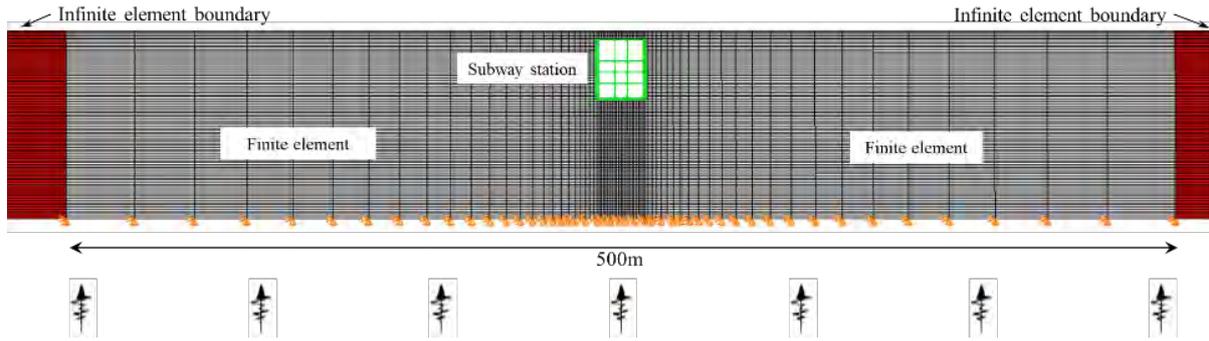| Material		Density (kg/m³)	Module of elasticity （GPa）	Poisson's ratio	Yield strength (MPa)	Axial compressive strength （MPa）	Axial tensile strength （MPa）
Concrete	C45	2500	33.5	0.2	—	29.6	2.51
	C35	2500	31.5	0.2	—	23.4	2.2
Reinforce-ment	HRB400	7850	200	0.3	400	—	—
	HRB335	7850	200	0.3	335	—	—

Figure 2: Finite element analysis model of soil structure interaction

To better simulate the dynamic response of the structure in the elastic-plastic stage, the isotropic bilinear model was used for reinforcement, and the concrete damage plastic model was used for concrete. According to basic theory of Lubliner et al [14] and Lee et al. [15], the damage factor D is used to describe the damage degree and stiffness degradation of concrete:

$$\bar{\sigma} = (1-D) E_0 \bullet \left(\varepsilon - \varepsilon^{pl} \right) \qquad (1)$$

where $\bar{\sigma}$ is the effective stress; E_0 is the initial elastic stiffness; ε is the strain tensor; ε^{pl} is plastic strain tensor; and D is the damage factor, including the tension damage factor D_t and the compression damage factor D_c. Due to different response characteristics in tension and compression, two variables are defined to present damage states respectively as Eq. (2):

$$D = 1 - (1-D_t)(1-D_c) \qquad (2)$$

where, the value of D is within the range of (0, 1); $D=0$ indicates that the structure is in the elastic state; $D>0$ indicates that the structure is in the plastic state.

Figure 3 shows the numerical calculation results of the distribution of tensile and compressive damage to the subway station structures under a certain seismic excitation.

(a)   (b)
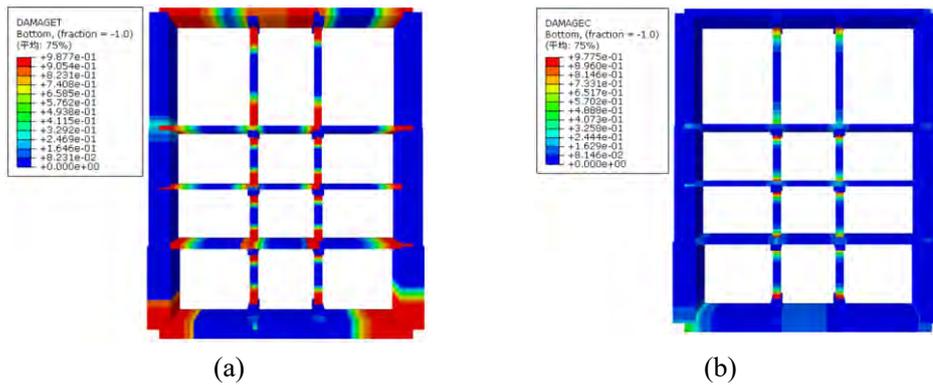
Figure 3: Cloud chart of structural damage: (a) distribution of tensile damage factor D_t; (b) distribution of compression damage factor D_c

1.2 The selection of engineering demand parameter

Engineering demand parameters (EDPs) refers to the index that can quantitatively characterize the degree of structural damage under earthquake excitation and they are crucial in selecting seismic dynamic time-history analyses for underground structures. Currently, commonly used failure indices for underground structures include the sectional bending moment, joint displacement, rotation angle of key components (such as the central column of a subway

station), and maximum layer displacement angle. However, using only the performance indicators of key components to evaluate the seismic performance of underground structures is not comprehensive. The dynamic behaviors of general components, such as increased deformation and stiffness degradation, can also accelerate the damage of key components and lead to the overall destruction of the underground structure. To address this issue, this paper proposes an overall damage index for underground structures that reasonably considers the impact of the size and damage of each component of a subway station on its overall seismic performance based on the P-A dual parameter damage model proposed by Park and Ang [16]. The overall damage index can be expressed by the follow Eqs. (3)-(5):

$$D_{us} = \sum_{i=1}^{n} \lambda_i \cdot D_i \tag{3}$$

$$\lambda_i = \frac{\Phi_i}{\sum_{i=1}^{n} \Phi_i} \tag{4}$$

$$\Phi_i = Den(E_{hi}) = \frac{E_{hi}}{V_i} \tag{5}$$

where, D_{us} is the overall damage index; D_i is the damage index of ith component; V_i is the volume of the ith component; λ_i is the weighting coefficient; and E_{hi} is the total hysteretic energy consumption of ith component. The D_{us} was selected as the EDP of subway station structure. Referring to the describe of the damage levels of mountain tunnels [4], while $D_{us} = 0$, the structure is no damage; $0 < D_{us} \le 0.4$, the structure is slightly damaged; $0.4 < D_{us} \le 0.7$, the structure is severely damaged; $0.7 < D_{us}$, the structure is collapsed.

2 The selection of ground motions and representative intensity measures

2.1 The selection of input ground motion for elastoplastic time-history analysis

In order to obtain better results for the selection of input GMs for elastoplastic dynamic time-history analysis, the composition of data samples should be representative and extensive when the data sample quantity is certain. The GMs used are suggested to be from high seismicity sites associated with earthquakes of moderate-to-large magnitudes, which may impose significant structural demand and are thus meaningful for the estimation of earthquake damage potential [9]. In this study, 40 GMs are downloaded from the PEER strong ground motion database [17] as the input GMs for analysis. The conditions for screening ground motion were that the magnitude was greater than Mw 6.0, and the peak acceleration of ground motion was greater than 0.05g. The brief information of the 40 selected GMs are listed in Table 2. And the intensity level of the selected 40 GMs is visually shows in Figure 4 in the form of A_{si}.
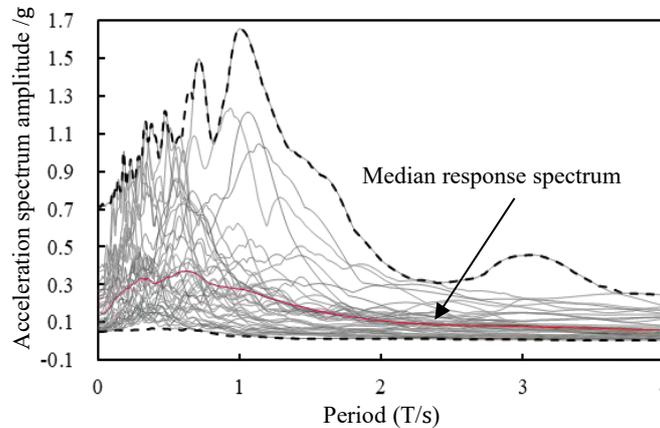

Figure 4: Acceleration intensity of the selected 40 input GMs (damping ratio =0.05)

Table 2: The brief information of the 40 selected ground motion records.

Ground motion number	Earthquake name	Record name	PGA (m/s²)	Ground motion number	Earthquake name	Record name	PGA (m/s²)
1	1999, Chi-Chi, Taiwan, China	RSN1229_CHICHI_CHY078-E	0.880	21		RSN5120_CHUETSU_ISK005EW	0.735
2		RSN1310_CHICHI_ILA004-N	0.640	22	1987, Whittier Narrows, America	RSN608_WHITTIER.A_A-WAT180	1.102
3		RSN1310_CHICHI_ILA004-W	0.732	23		RSN608_WHITTIER.A_A-WAT270	0.983
4		RSN1334_CHICHI_ILA044-N	0.710	24		RSN6206_TOTTORI.1_HRS013EW	0.858
5		RSN1334_CHICHI_ILA044-W	0.818	25		RSN6206_TOTTORI.1_HRS013NS	0.564
6		RSN1357_CHICHI_KAU011-E	0.574	26	2000, Tottori, Japan	RSN6378_TOTTORI.1_TKS001EW	0.641
7		RSN1357_CHICHI_KAU011-N	0.555	27		RSN6959_DARFIELD_REHSN02E	2.629
8	1979, Imperial, Valley, America	RSN178_IMPVALL.H_H-E03140	2.674	28	2010, Darfield, New Zealand	RSN6959_DARFIELD_REHSS88E	2.379
9		RSN178_IMPVALL.H_H-E03230	2.225	29		RSN732_LOMAP_A02043	2.744
10		RSN201_IMPVALL.A_A-E03140	1.380	30	1989, Loma Prieta, America	RSN732_LOMAP_A02133	2.203
11		RSN201_IMPVALL.A_A-E03230	0.970	31		RSN759_LOMAP_A01000	2.576
12	2004, Niigata, Japan	RSN4201_NIIGATA_NIG011EW	0.582	32		RSN759_LOMAP_A01090	2.845
13		RSN4201_NIIGATA_NIG011NS	0.561	33		RSN760_LOMAP_MEN270	1.098
14		RSN4215_NIIGATA_NIG025EW	2.045	34		RSN760_LOMAP_MEN360	1.187
15		RSN4215_NIIGATA_NIG025NS	1.936	35		RSN808_LOMAP_TRI000	1.003
16	1984, Morgan Hill, America	RSN452_MORGAN_A01310	0.648	36		RSN808_LOMAP_TRI090	1.601
17		RSN4989_CHUETSU_FKS020NS	0.521	37	2011, Christchurch, New Zealand	RSN8123_CCHURCH_REHSN02E	3.714
18	2007, Niigataken Chuetsu-oki, Japan	RSN5103_CHUETSU_IBRH07NS	0.592	38		RSN8123_CCHURCH_REHSS88E	7.153
19		RSN5119_CHUETSU_ISK004EW	0.733	39	1994, Northridge-01, America	RSN962_NORTHR_WAT180	0.915
20		RSN5119_CHUETSU_ISK004NS	0.706	40		RSN962_NORTHR_WAT270	0.883

2.2 The selection of representative intensity measures

Reasonable IMs are an important basis for evaluating the EDP of underground structures. The selected representative ground motion IMs should contain as much basic information as possible of the original ground motion. At present, the commonly used IMs include: (1) IMs related to the characteristics of ground motion itself, such as the cumulative absolute velocity (A_{cv}), Arias Strength (I_a), peak ground acceleration (A_{pg}), and peak ground velocity (V_{pg}), etc.; (2) IMs related to the seismic response of underground structures, such as spectral acceleration (S_a), spectral velocity (S_v), spectral displacement (S_d), etc. considering the strong randomness of GMs, this paper covers various ground motion characteristics, including amplitude, energy, and spectral, when selecting representative IMs. According to the frequency of use of these IMs in engineering seismic design or research, 12 types of IMs of 3 categories are selected, as shown in Table 3.

Table 3: The selection of representative IMs

Types	Number	IMs	Definition
Peak Value	1	Peak ground acceleration	$A_{pg}=max\|a(t)\|$
	2	Peak ground velocity	$V_{pg}=max\|v(t)\|$
	3	Peak ground displacement	$D_{pg}=max\|d(t)\|$
	4	Effective peak ground acceleration	$A_{ep}=\dfrac{mean\left(S_a^{0.1\sim0.5}(\xi=0.05)\right)}{2.5}$
	5	Cumulative absolute velocity	$A_{cv}=\displaystyle\int_0^{tot}\|a(t)\|\,dt$
Energy	6	Acceleration spectrum intensity	$A_{si}=\displaystyle\int_{0.1}^{0.5}S_a(T,\xi=0.05)\,dT$
	7	Velocity spectrum intensity	$V_{si}=\displaystyle\int_{0.1}^{0.5}S_v(T,\xi=0.05)\,dT$
	8	Displacement spectrum intensity	$D_{si}=\displaystyle\int_{2.0}^{5.0}S_d(T,\xi=0.05)\,dT$
	9	Arias intensity	$I_a=\dfrac{\pi}{2g}\displaystyle\int_0^{tot}a(t)^2\,dt$
Spectral	10	Acceleration response spectrum	$S_a(T,\xi=0.05)$
	11	Velocity response spectrum	$S_v(T,\xi=0.05)$
	12	Displacement response spectrum	$S_d(T,\xi=0.05)$

Note: $a(t)$ is the acceleration time history of ground motion; $v(t)$ is the velocity time history of ground motion; $d(t)$ is the displacement time history of ground motion; g is the acceleration of gravity; t is a certain moment; $mean(\cdot)$ means the average value of the variable; ξ is the damping ratio; t_{tot} is the total duration of the selected ground motion; T is the natural vibration period of the single degree of freedom structure.

3 Division of characteristic period bands of site

Due to the limitation of soil, the number of stories in underground structures is often small and the difference in natural vibration periods is small. Therefore, the effect of natural vibration periods on the EDP of underground structures is negligible. However, the diversity of characteristic periods at sites where underground structures are located can also cause significant differences in the EDPs of underground structures under earthquake excitation, and it is difficult to study the seismic response of underground structures in each characteristic period site type. Therefore, this section will determine the appropriate limit of underground structure site characteristic period bands through abundant dynamic elastoplastic time-history analysis of a four-storey and three-span subway station structure.

3.1 The selection of representative characteristic period

According to the seismic code of "Code for Seismic Design of Buildings" (GB 50011-2010) [11] and "Code for Seismic Design of Buildings in Shanghai" (DGJ 08-9-2013) [18], the seismic design characteristic period range of the site is 0.2s~1.1s. Therefore, this paper selects 27 representative characteristic period of soil-underground structure from 0.2 s to 1.1 s for analysis. The representative periods are shown in Table 4. The GMs in Table 2 is used as input for each representative characteristic period system model, resulting in a total of 1080 elastoplastic dynamic time-history analyses.

Table 4: The selected representative site periods

Number	Period	Number	Period	Number	Period
1	0.20	10	0.29	19	0.70
2	0.21	11	0.30	20	0.75
3	0.22	12	0.35	21	0.80
4	0.23	13	0.40	22	0.85
5	0.24	14	0.45	23	0.90
6	0.25	15	0.50	24	0.95
7	0.26	16	0.55	25	1.00
8	0.27	17	0.60	26	1.05
9	0.28	18	0.65	27	1.10

3.2 Result of the division of period bands

Based on the numerical analysis results, the EDP of the subway station structure under 40 GMs can be obtained in each representative characteristic period. According to the value of EDP, the rankings of 40 GMs of 27 representative characteristic period are obtained. Then, comparing the degree of ranking change of different characteristic periods. When the degree of ranking change is within an acceptable range, these two characteristic periods can be divided into the same group.

In order to quantitatively describe the degree of ranking change, it is necessary to define a threshold. When the ranking change value does not exceed the given threshold, the greater number of GMs there are, indicating a higher similarity between the compared two representative characteristic periods. In this paper, this threshold is taken as 20% of the total number of 40 GMs. In addition, this paper regulated that the seismic response law of the subway station located in the site of the representative characteristic periods will be considered similar, when the proportion of the number of GMs that not exceeded the given threshold to the total 40 GMs exceed 60%. According to this regulation, the site period can be divided into four types of Type I, Type II, Type III, and Type IV, corresponding to the period band of 0.20s~0.23s, 0.23s~0.26s, 0.26s~0.40s, and 0.40s~1.10s respectively. The results of the proportion of the number of GMs that not exceeded the given threshold to the total 40 GMs are given in Figure 5.
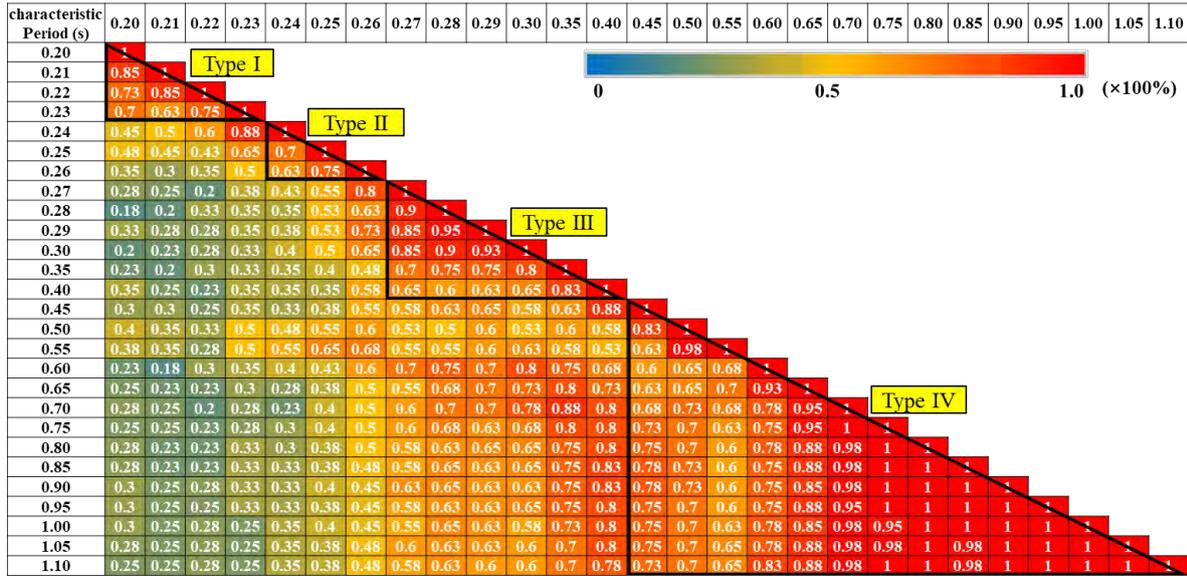
Figure 5: Results of the proportion of the number of GMs that not exceeded the given threshold to the total 40 GMs between different characteristic period

The figure is a lower-triangular heatmap matrix of correlation values between characteristic periods (0.20 s to 1.10 s), with regions labeled Type I, Type II, Type III, and Type IV, and a color scale from 0 to 1.0 (×100%).

characteristic Period (s)	0.20	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00	1.05	1.10
0.20																											
0.21	0.85																										
0.22	0.73	0.85																									
0.23	0.7	0.63	0.75																								
0.24	0.45	0.5	0.6	0.88																							
0.25	0.48	0.45	0.43	0.65	0.7																						
0.26	0.35	0.3	0.35	0.5	0.63	0.75																					
0.27	0.28	0.25	0.2	0.38	0.43	0.55	0.8																				
0.28	0.18	0.2	0.33	0.35	0.35	0.53	0.63	0.9																			
0.29	0.33	0.28	0.28	0.35	0.38	0.53	0.73	0.85	0.95																		
0.30	0.2	0.23	0.28	0.33	0.4	0.5	0.65	0.85	0.9	0.93																	
0.35	0.23	0.2	0.3	0.33	0.35	0.4	0.48	0.7	0.75	0.75	0.8																
0.40	0.35	0.25	0.23	0.35	0.35	0.35	0.58	0.65	0.6	0.63	0.65	0.83															
0.45	0.3	0.3	0.25	0.35	0.33	0.38	0.55	0.58	0.63	0.65	0.58	0.63	0.88														
0.50	0.4	0.35	0.33	0.5	0.48	0.55	0.6	0.53	0.5	0.6	0.53	0.6	0.58	0.83													
0.55	0.38	0.35	0.28	0.5	0.55	0.65	0.68	0.55	0.55	0.6	0.63	0.58	0.63	0.98													
0.60	0.23	0.18	0.3	0.35	0.4	0.43	0.6	0.7	0.75	0.7	0.8	0.75	0.68	0.6	0.65	0.68											
0.65	0.25	0.23	0.23	0.3	0.28	0.38	0.5	0.55	0.68	0.7	0.73	0.8	0.73	0.63	0.65	0.7	0.93										
0.70	0.28	0.25	0.2	0.28	0.23	0.4	0.5	0.6	0.7	0.7	0.78	0.88	0.8	0.68	0.73	0.68	0.78	0.95									
0.75	0.25	0.25	0.23	0.28	0.3	0.4	0.5	0.6	0.68	0.63	0.68	0.8	0.8	0.73	0.7	0.63	0.75	0.95	1								
0.80	0.28	0.23	0.23	0.33	0.3	0.38	0.5	0.58	0.63	0.65	0.65	0.75	0.8	0.75	0.7	0.6	0.78	0.88	0.98	1							
0.85	0.28	0.23	0.23	0.33	0.33	0.38	0.48	0.58	0.65	0.63	0.65	0.75	0.83	0.78	0.73	0.6	0.75	0.88	0.98	1	1						
0.90	0.3	0.25	0.28	0.33	0.33	0.4	0.45	0.63	0.65	0.63	0.63	0.75	0.83	0.78	0.73	0.6	0.75	0.85	0.98	1	1	1					
0.95	0.3	0.25	0.25	0.33	0.33	0.38	0.45	0.58	0.63	0.63	0.65	0.75	0.8	0.75	0.7	0.6	0.75	0.88	0.95	1	1	1	1				
1.00	0.3	0.25	0.28	0.25	0.35	0.4	0.45	0.55	0.65	0.63	0.58	0.73	0.8	0.75	0.7	0.63	0.78	0.85	0.98	0.95	1	1	1	1			
1.05	0.28	0.25	0.28	0.25	0.35	0.38	0.48	0.6	0.63	0.63	0.6	0.7	0.8	0.75	0.7	0.65	0.78	0.88	0.98	0.98	1	0.98	1	1	1		
1.10	0.25	0.25	0.28	0.25	0.35	0.38	0.48	0.58	0.63	0.6	0.6	0.7	0.78	0.73	0.7	0.65	0.83	0.88	0.98	1	1	0.98	1	1	1	1	

4 Construction of composite intensity measures

4.1. Construction process of composite intensity measures

In this study, the dependent variable is the overall damage index D_{us} of the underground structure, and the independent variable is the selected representative IMs. The construction process of composite IMs mainly includes the following steps [19, 20].

Step (1): The dependent variable and independent variable are denoted as $Y=(y)_{n\times1}, y \in R^n$, $X=(x_1, x_2,\cdots, x_p)_{n\times p}, x_i \in R^n$, and the standardized matrix is further denoted as F_0, E_0. Let u_1 be the first principal component of F_0, $u_1=F_0c_1$, $\|c_1\|=1$; t_1 is the first principal component of E_0, $t_1=E_0\omega_1$, $\|\omega_1\|=1$. In PLSR, it is required that the covariance between t_1 and u_1 be maximized, that is

$$Cov(t_1, u_1) = \sqrt{var(t_1)var(u_1)}r(t_1, u_1) \to max \tag{6}$$

where, $r(t_1, u_1)$ is the correlation matrix. Eq. (6) can be transformed into solving the following optimization problem, that is

$$\begin{cases} max\langle E_0\omega_1, F_0c_1\rangle \\ s.t.\ \omega_1^T\omega_1 = 1; \quad c_1^Tc_1 = 1 \end{cases} \tag{7}$$

Step (2): Use the Lagrange algorithm to solve Eq. (7). For univariate partial least squares regression, since there is only one variable in F_0, $u_1=F_0$,

$$t_1 = E_0\omega_1 = \frac{1}{\sqrt{\sum_{j=1}^p r^2(x_j, y)}}\cdot[r(x_1, y)E_{01}, r(x_2, y)E_{02}, \cdots,\ r(x_p, y)E_{0p}] \tag{8}$$

$$\omega_1 = \frac{1}{\sqrt{\sum_{i=1}^{p} r^2(x_i, y)}} \begin{bmatrix} r(x_1, y) \\ r(x_2, y) \\ \vdots \\ r(x_p, y) \end{bmatrix} \tag{9}$$

Step (3): Find the regression of E_0 on t_1 and the regression of F_0 on t_1, that is

$$E_0 = t_1 p_1^{\mathrm{T}} + E_1 \tag{10}$$

$$F_0 = t_1 r_1 + F_1 \tag{11}$$

where, p_1 and r_1 are regression coefficients; E_1 and F_1 are residual matrices. That is,

$$p_1 = \frac{E_0^T t_1}{\|t_1\|^2}, \qquad r_1 = \frac{F_0^T t_1}{\|t_1\|^2} \tag{12}$$

Step (4): Replace E_0 and F_0 with residual matrices E_1 and F_1, and find the second principal components t_2 and u_2.

$$t_2 = E_1 \omega_2; \; u_2 = F_1 c_2 \tag{13}$$

By repeating steps (1)-(3), the second principal components t_2 and u_2 can be obtained.

Step (5): Repeat the previous steps to find t_3, t_4, $\cdots\cdots$, t_h ($h \leq rank(E_0)$), h can be determined by cross-validity analysis.

Further, find the linear regression equation of F_0 with respect to $t_1, t_2, \cdots, t_h$.

$$F_0 = r_1 E_0 \omega_1^* + r_2 E_0 \omega_2^* + \cdots + r_h E_0 \omega_h^* + F_h = E_0 \left(r_1 \omega_1^* + r_2 \omega_2^* + \cdots + r_h \omega_h^* \right) + F_h \tag{14}$$

where,

$$\omega_h^* = \prod_{i}^{h-1} \left(I - \omega_i p_i^{\mathrm{T}} \right) \omega_h \tag{15}$$

Step (6): Let $x_i^* = E_{0i}$, $y^* = F_0$, $\alpha_i = \sum_{h=1}^{m} r_h \omega_{hi}^*$, ω_{hi}^*, where ω_{hi}^* is the ith component of ω_h^*, Eq. (14) can be written as,

$$\hat{y}^* = \beta_1 x_1^* + \beta_2 x_2^* + \cdots + \beta_p x_p^* \tag{16}$$

Restore the standardized variables in Eq. (16) to the original variables:

$$\hat{y} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{17}$$

where, ε is the residual of the regression model.

Step (7): Since the Bootstrap resampling method can generate data samples without requiring other assumptions or adding new samples, and only relies on the given sample information, this paper adopts the regression parameter significance test method based on Bootstrap resampling to test the significance of regression coefficients [10]. To ensure that $\beta_k \neq 0$ ($k = 1, 2, \ldots, p$), a hypothesis test must be performed. The null hypothesis H_0 and alternative hypothesis H_1 are respectively:

$$H_0: \; \beta_k = 0 \tag{18}$$

$$H_1: \; \beta_k \neq 0 \tag{19}$$

Randomly select n_B samples with replacement from the sample sets F_0 and E_0 ($n_B < n$), to

obtain a Bootstrap sample. Repeat *r* times to obtain *r* groups of Bootstrap samples, and find the partial least squares model with *h* principal components for each group of samples, and its partial regression coefficient is:

$$\left\{ \beta_1^b, \beta_2^b, \dots\dots, \beta_p^b \right\} \quad (b=1, 2, \dots\dots, r) \tag{20}$$

Let $\hat{\beta}_k^b = |\beta_k^b - \beta_k|$, ($b=1, 2, \dots\dots, r$; $k = 1, 2, \dots, p$), suppose that the hypothesis test level is α, and the α quantile of $\hat{\beta}_k^b$, $\beta_\alpha(k)$, is taken as the critical value of the rejection region. If $|\beta_k| > \beta_\alpha(k)$, reject H_0, and the independent variable has statistical significance, otherwise remove the variable. Remove variables that have not passed the significance test from the regression equation and continue to repeat the above process until all variables pass the significance test and stop.

4.2 Cross-validity analysis

In step (5) of section 4.1, the number of principal components to be selected can be determined through the following process. First, remove the *i*th sample point from all sample points to form a new set of sample points (containing a total of n-1 sample points), use the new set of sample points, and use *h* principal components to fit a regression equation $y_{h(-i)}$. Then substitute the excluded *i*th sample point into $y_{h(-i)}$, and the fitted value is denoted as $\hat{y}_{h(-i)}$. For $i = 1, 2, \cdots, n$, repeat the above operation and substitute the excluded *i*th sample point into Eq. (17), then the predicted error square sum of $\hat{y}$ can be defined as *PRESS_h*, which is

$$PPRESS_h = \sum_{i=1}^{n} \left[\hat{y} - \hat{y}_{h(-i)} \right]^2 \tag{21}$$

In addition, use all sample points to fit a regression equation with h principal components. At this time, the predicted value of the *i*th sample point is denoted as $\hat{y}_{hi}$, and the error square sum of $\hat{y}$ can be defined as *SS_h*, which is

$$SS_h = \sum_{i=1}^{n} (\hat{y} - \hat{y}_{hi})^2 \tag{22}$$

If the disturbance error of the regression equation with *h* components is smaller than the fitting error of the regression equation with *h*-1 components to a certain extent, it is considered that adding one component t_h can significantly improves the accuracy of the prediction model. Therefore, it is hoped that the ratio of *PRESS_h/SS_h* can be as small as possible. Specify,

$$Q_h^2 = 1 - (PRESS_h/SS_h) \geq 0.0975 \tag{23}$$

When $Q_h^2 \geq 0.0975$, the marginal contribution of the t_h component is significant. And for $k = 1, 2, \cdots, p$, there is at least one *k* such that $Q_h^2 \geq 0.0975$.

4.3 Significance test of regression coefficients

By introducing Python scientific computing libraries of Numpy and Scipy, a Python program code is written to implement PLSR parameter significance analysis based on Bootstrap resampling. Selecting the results of elastoplastic dynamic time-history analysis of 4 soil-underground structure numerical models with natural vibration periods of 0.22s, 0.25s, 0.4s, and 0.8s as data sample. The selected data sample account for four type site condition respectively. According to the process introduced in section 4.1, substitute the data samples to

obtain 4 PLSR models (composite IMs). Figure 6 shows the Q_h^2 values of 12 principal components of four numerical models. It can be seen from Figure 6 that when extracting the third partial least squares component, for 4 numerical models, there are $Q_3^2 < 0.0975$, so according to cross-validity analysis, extracting 2 principal components can achieve the accuracy of prediction.
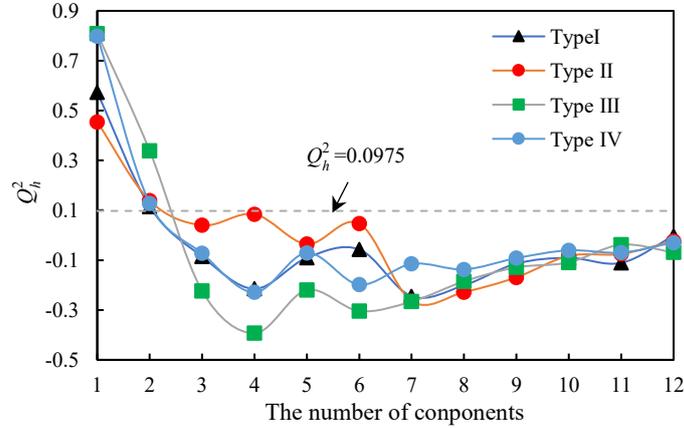
Figure 6: Result of the cross-validation

After obtaining the expression for the composite IMs, a regression parameter significance test method based on bootstrap sampling is used to verify the regression parameters. Taking n_B=30, r=100, a total of 100 bootstrap samples were obtained. Then, the critical value of the rejection domain $\beta_\alpha(k)$ (α is 0.05) is obtained through these samples. If $|\beta_k| > \beta_{0.05}(k)$, then H_0 is rejected, that is, β_k is significantly not equal to 0, otherwise the null hypothesis is accepted, and the significance test is not passed at this time.

Table 5 gives the bootstrap test results for 4 PLSR models. From Table 5, it can be seen that all 12 regression coefficients of Type I, Type III, and Type IV, representative underground structure have passed the significance test. But the regression coefficient of A_{si} of Type II has not passed the significance test. On this basis, the regression coefficient of A_{si} of Type II is removed from the regression equation. And then, the PLSR analysis is performed again according to all the above steps, and the regression coefficient significance test analysis steps based on bootstrap sampling are continued until all regression coefficients of medium and long representative underground structures pass the test.

Table 5: Bootstrap test results of 12 regression coefficients

Types of sites	A_{pg}		V_{pg}		D_{pg}		A_{ep}		A_{cv}		A_{si}	
	$\hat{\beta}_1$	$\beta_{0.05}(1)$	$\hat{\beta}_2$	$\beta_{0.05}(2)$	$\hat{\beta}_3$	$\beta_{0.05}(3)$	$\hat{\beta}_4$	$\beta_{0.05}(4)$	$\hat{\beta}_5$	$\beta_{0.05}(5)$	$\hat{\beta}_6$	$\beta_{0.05}(6)$
Type I	-0.118	0.006	-0.074	0.014	0.084	0.009	0.395	0.028	0.088	0.010	0.032	0.011
Type II	-0.372	0.100	-0.074	0.013	0.219	0.005	0.460	0.050	0.042	0.009	**0.002**	**0.005**
Type III	-0.079	0.007	-0.069	0.013	0.035	0.004	0.380	0.046	0.101	0.003	0.080	0.005
Type IV	-0.059	0.020	-0.089	0.011	-0.014	0.006	0.303	0.015	0.102	0.012	-0.097	0.033

Types of sites	V_{si}		D_{si}		I_a		S_a		S_v		S_d	
	$\hat{\beta}_7$	$\beta_{0.05}(7)$	$\hat{\beta}_8$	$\beta_{0.05}(8)$	$\hat{\beta}_9$	$\beta_{0.05}(9)$	$\hat{\beta}_{10}$	$\beta_{0.05}(10)$	$\hat{\beta}_{11}$	$\beta_{0.05}(11)$	$\hat{\beta}_{12}$	$\beta_{0.05}(12)$
Type I	-0.293	0.066	-0.259	0.058	0.526	0.073	0.181	0.009	0.198	0.004	0.372	0.034
Type II	-0.280	0.026	-0.143	0.016	0.415	0.046	0.140	0.014	0.153	0.016	0.252	0.032
Type III	-0.302	0.071	-0.294	0.080	0.543	0.103	0.189	0.010	0.209	0.008	0.396	0.039
Type IV	-0.228	0.043	-0.321	0.035	0.635	0.057	0.119	0.023	0.142	0.015	0.615	0.090

The composite IMs sought in this paper are finally obtained, and the composite IMs

corresponding to the structures in Type I, Type II, Type III, and Type IV site are denoted as I_{pls1} I_{pls2}, I_{pls3}, and I_{pls4}, respectively. The regression coefficients of the composite IMs were shown in Table 6.

Table 6: The regression coefficients of composite IM

Composite IMs \\ Coefficient	I_{pls1}	I_{pls2}	I_{pls3}	I_{pls4}
β_1	-0.118	-0.407	-0.079	-0.059
β_2	-0.074	-0.157	-0.069	-0.089
β_3	0.084	0.190	0.035	-0.014
β_4	0.395	0.490	0.380	0.303
β_5	0.088	0.015	0.101	0.102
β_6	0.032	—	0.080	-0.097
β_7	-0.293	-0.020	-0.302	-0.228
β_8	-0.259	-0.222	-0.294	-0.321
β_9	0.526	0.447	0.543	0.635
β_{10}	0.181	0.105	0.189	0.119
β_{11}	0.198	0.121	0.209	0.142
β_{12}	0.372	0.239	0.396	0.615
constant	0.822	-0.946	0.899	1.242

5 Correlation analysis of composite IMs

To verify the correlation between the composite IMs and EDP of subway station structures, 100 verified GMs were randomly selected from the PEER strong motion database using the same screening conditions as in section 3.1 to carry out dynamic time-history analysis. The acceleration spectrum intensity of the selected 100 seismic motions is shown in Figure 7.
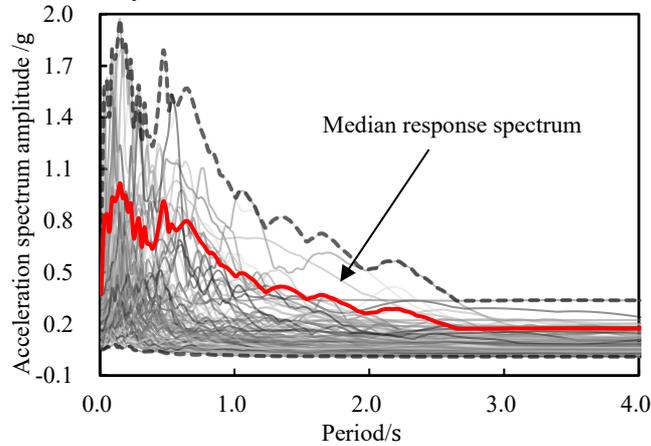
Figure 7: Acceleration intensity of the selected 100 verified GMs (damping ratio =0.05)

In statistics, the Pearson correlation coefficient is commonly used to measure the linear correlation between two variables. The correlation coefficient between two random variables is calculated using Eq. (24):

$$\rho = \frac{Cov(lnIM, lnEDP)}{\sqrt{Var(lnIM)} \cdot \sqrt{Var(lnEDP)}} \tag{24}$$

where, $Cov(\cdot,\cdot)$ represents the covariance between variables, and $Var(\cdot)$ is the variance of the variable. The Pearson correlation coefficient is defined as follows: $0.8 \leq |\rho| \leq 1$, high correlation; $0.5 \leq |\rho| < 0.8$, moderate correlation; $0.3 \leq |\rho| < 0.5$, low correlation; $|\rho| < 0.3$, low or

no correlation.

In this paper, the Pearson correlation coefficient between IMs and EDP of subway station structures is used to measure the performance of IMs in characterizing the damage potential of GMs for subway station structures. Figure 8 shows the calculated results of Pearson correlation coefficients between composite IMs, 12 single IMs and the EDP of subway station in Type I (Figure 8a), Type II (Figure 8b), Type III (Figure 8c), and Type IV (Figure 8d) soil, respectively.

(a) The comparison of Pearson correlation between I_{pls1} and 12 single IMs

(b) The comparison of Pearson correlation between I_{pls2} and 12 single IMs

(c) The comparison of Pearson correlation between I_{pls3} and 12 single IMs

(d) The comparison of Pearson correlation between I_{pls4} and 12 single IMs
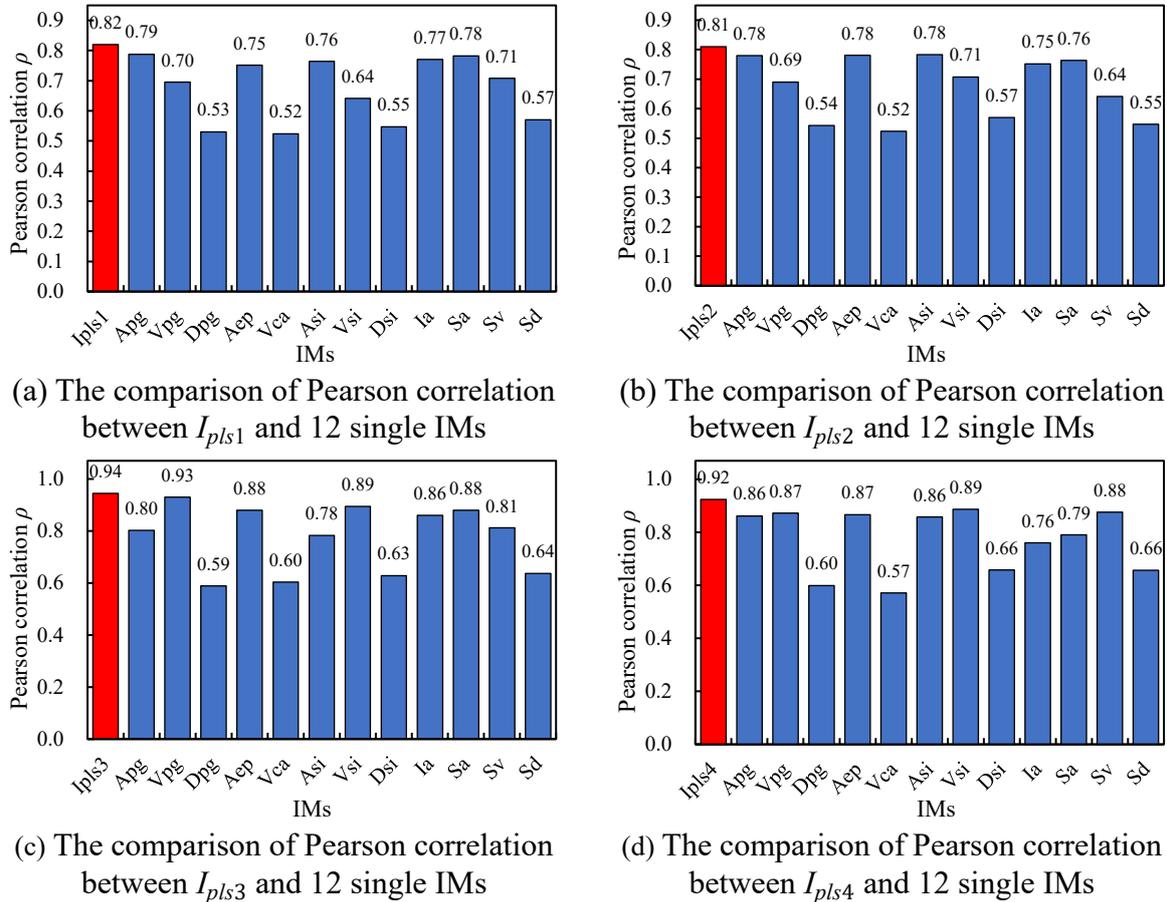
Figure 8: The comparison of pearson correlation of composite IMs and 12 single IMs

From Figure 8, it can be seen that whether in Type I, Type II, Type III or Type IV soil, the correlations between the EDP of subway station structure and composite IMs constructed by PLSR are stronger than that of 21 single IMs. The pearson correlations between EDP of subway station structure and 4 composite IMs are 0.82, 0.81, 0.94, and 0.92, respectively.

6 Conclusions

In order to study the damage potential of different input GMs on muti-storey subway station. This paper conducted 1080 elastoplastic dynamic time-history analysis to divide the type of site condition, and constructed some composite IMs that can effectively characterize the damage potential of input GMs on a typical four-storey and three-span subway station structures in different type of site condition. The research results can provide a reference for the selection of input GMs in the seismic design of relevant muti-storey subway station structures. However, it should be noted that this article is only a study conducted using a four-storey, three-span subway station as an example, and does not have universality for seismic design of different subway station structures. Further research is needed on the impact of the structural forms of subway station on subway stations seismic response in the future to obtain more universal conclusions.

The main conclusions of this paper are as follows:

(1) According to the similarity analysis of the ranking of damage degree of a subway station structure located in 27 sites with different period condition, the characteristic site period can be divided into four types of Type I, Type II, Type III, and Type IV, corresponding to the period bands of 0.20s~0.23s, 0.23s~0.26s, 0.26s~0.40s, and 0.40s~1.10s.

(2) According to Pearson correlation analysis, for underground structure in Type I and Type II site, the acceleration-based IMs, such as A_{pg}, A_{si}, and S_a have higher correlation with EDP, while for underground structure in Type III and Type IV site, velocity-based IMs, such as V_{pg}, V_{si}, and S_v have higher correlation with EDP.

(3) For multi-storey subway station structures in Type I, Type II, Type III, and Type IV site, the correlation between composite IMs and EDP are generally higher than that of 12 single IMs. Thus, the performance of the composite IMs proposed in this paper in characterizing the EDP was preliminarily verified.

References

[1] Parra-Montesinos, G. J., Bobet, A., and Ramirez, J. A. (2006) Evaluation of soil-structure interaction and structural collapse in Daikai subway station during Kobe earthquake. *ACI Structural Journal* **103**, 113.

[2] Wang, W. L., Wang, T. T., Su, J. J., Lin, C. H., Seng, C.R., and Huang, T.H. (2001) Assessment of damage in mountain tunnels due to the Taiwan Chi-Chi Earthquake. *Tunnelling and Underground Space Technology* **16**, 133–150.

[3] Zhai, C. H., and Xie, L. L. (2004) Study on the severest design ground motions. *13th World Conference on Earthquake Engineering Vancouver, BC, Canada*.

[4] Chen, Z. Y., and Wei, J. S. (2013) Correlation between ground motion parameters and lining damage indices for mountain tunnels. *Natural Hazards* **65**, 20.

[5] Zhong, Z. L., Shen, Y. Y., Zhao, M., Li, L. Y., Du, X. L., and Hao, H. (2020) Seismic fragility assessment of the Daikai subway station in layered soil. *Soil Dynamics and Earthquake Engineering* **132**, 106044.

[6] Huang, Z. K., Pitilakis, K., Argyroudis, S., Tsinidis, G., and Zhang, D. M. (2021) Selection of optimal intensity measures for fragility assessment of circular tunnels in soft soil deposits. *Soil Dynamics and Earthquake Engineering* **145**, 106724.

[7] Yang, J., Zhuang, H. Y., Zhang, G. Y., Tang, B. Z., and Xu, C. J. (2023) Seismic performance and fragility of two-story and three-span underground structures using a random forest model and a new damage description method. *Tunnelling and Underground Space Technology* **135**, 104980.

[8] Housner, G. W., Jennings, P. C., and Hall, W. J. (1977) The capacity of extreme earthquake motions to damage structures. *Englewood Cliffs, N.J, 1977. Prentice-Hall, Inc* 1977.

[9] Liu, T. T., Yu, X. H., and Lu, D. G. (2020) An approach to develop compound intensity measures for prediction of damage potential of earthquake records using canonical correlation analysis. *Journal of Earthquake Engineering* **24**, 24.

[10] Liu, T. T., Lu, D. G., and Yu, X. H. (2019) Development of a compound intensity measure using partial least-squares regression and its statistical evaluation based on probabilistic seismic demand analysis. *Soil Dynamics and Earthquake Engineering* **125**, 105725.

[11] GB 50011-2010. (2010) Code for seismic design of buildings. *Beijing: China Architecture & Building Press*. (in Chinese).

[12] Zienkiewicz, O. C. (1986) Mapped infinite elements for exterior wave problems. *International Journal For Numerical Methods In Engineering* **23**, 1388.

[13] Lysmer, J., and Kuhlemeyer, R. (1969) Finite dynamic model for infinite media. *Journal of the Engineering Mechanics Division* **95**, 859–877.

[14] Lubliner, J., Oliver, J., Oller, S., and Onate, E. (1998) A plastic-damage model for concrete. *International Journal of Solids and Structures* **25**, 299–326.

[15] Lee, J., and Fenves, G. L. (1998) Plastic-damage model for cyclic loading of concrete structures. *Journal of Engineering Mechanics* **124**, 892–900.

[16] Park, Y. J., Ang, A. H. S., and Wen, Y. K. (1985) Seismic damage analysis of reinforced concrete buildings. *Journal of Structural Engineering* **111**, 740–757.

[17] Pacific Earthquake Engineering Research Center. PEER NGA-west2 strong motion database [DB/OL]. (2013-3-27) [2022-4-3].

[18] DGJ08-9-2013. (2013) Code for seismic design of buildings. *Shanghai: Tongji University*. (in Chinese)

[19] Draper, N., and Smith, H. (1998) *Applied Regression Analysis*. New York: John Wiley & Sons, 115–131.

[20] Wold, S., Martens, H., and Wold, H. (2006) *The Multivariate Calibration Problem in Chemistry Solved by The PLS Method*. Springer Berlin Heidelberg: Berlin, Heidelberg, 286–293.

Coupling Simulation of Rupture and Seismic Analysis in Cross-Fault Mountain Tunnel

*Liqun Li[1], †Zhiyi Chen[1]

[1]Department of Geotechnical Engineering, Tongji University, Shanghai, 200092, China

*Presenting author: 2011711@tongji.edu.cn
†Corresponding author: zhiyichen@tongji.edu.cn

Abstract

Near-fault mountain tunnels often have severe permanent deformation and structural dislocation during the seismic damage process. In this paper, we establish a cross-fault tunnel-rock model, including the fault zone and the surrounding complete rock mass for a tunnel in western China, and carry out the numerical simulation of the dynamic force by applying the thrust displacement load of the uplift fault and the dynamic seismic load. We analyze the overall slip condition of the rock body in the fracture zone during the uplift process of the upper plate, reveal the development and evolution law of the "funnel zone" of the backlash fault, put forward the influence interval of the fault movement on the tunnel, analyze the dynamic load response of the tunnel caused by decisive seismic action under the condition of the fault's initial dislocation, reveal the coupling effect of the fault's initial dislocation and the seismic dynamic load, and summarize the distribution of internal force and deformation state of the tunnel, the deformation and damage mode of the tunnel under the initial dislocation-strong earthquake coupling is revealed, to provide a theoretical basis for the seismic-resistant and fault-resistant design of tunnels through faults in high-intensity zones.

Keywords: Rupture simulation; Seismic response; Tunnel engineering; Engineering geology; Soft rock

Introduction

Tunnel seismic damage is often more minor than ground structures, and the structure tends to be safer as the depth increases. However, seismic events in recent years have shown that when the surrounding rock conditions are poor, strong ground motion can lead to the overall instability of the surrounding rock-lining structure, in which the disaster effect of tunneling through active faults is particularly significant. Under complex seismic effects such as near-fault ground motion and significant fault dislocations, the damage mechanism of tunnels through active faults differs significantly from that of conventional projects in plains areas. Firstly, under active fault conditions, the upward and downward dislocations of the surrounding rock are constantly developing during tunnel construction. There will be long-term backlash/slip flow during the tunnel operation stage, which makes the force state of the surrounding intact rock-body-fracture-zone-tunnel system constantly changing, and the impact of this quasi-static long-term effect on the seismic performance of cross-fault tunnels is not clear yet. The tunnel constructed in the fracture zone tunnel site area will cause severe damage once the permanent forced displacement, such as fault dislocation occurs in the 2022 Menyuan 6.9 magnitude earthquake, the Qilian-Haiyuan rupture in the north-east edge of the Tibetan Plateau in the western section of the Lenglongling rupture occurred in left-rotation strike-slip dislocation, which caused the Lanzhou-Xinjiang High-Speed Railway Tunnel Daliang Tunnel across the active fault section of the line of the section of the severe offset, sidewall extrusion, crush, second lining collapse, and the side wall is crushed. The sidewalls were extruded and

broken, the second lining was crushed, the roadbed was uplifted, and the electric power and communication equipment were damaged. It is a typical case of high-speed railway cross-fault tunnel damage[1].

Research has been carried out on the disaster-causing effect of long-term creep and slip of faults and related measures. He et al. (2022) established a three-dimensional numerical model of cross-fracture by simulating the creep behavior of surrounding rock during the construction of cross-fault tunnels, analyzed the force and deformation characteristics of the upper and lower disk lining, and got the conclusion that the radius of the fault impact is about double that of the tunnel diameter [2] . Liu et al. [3][4] investigated the influence of positive fault thrust on tunnels with different inclinations of fault-free fracture zones through indoor modeling tests. Johansson et al. [5] determined the structural design criteria based on the fault displacements and failure strains using experiments and numerical simulations. The above studies have analyzed and discussed the mechanical response and damage mechanism of cross-fault tunnels in detail and investigated the catastrophic mechanism of tunnels under different loading methods, fault types, and surrounding rock conditions. However, the existing studies are analyzed mainly by a single loading method, and the coupled thrust-vibration loading method is less; there are limitations in reflecting the geological characteristics of active faults and the viscous-slip dynamic characteristics because the characterization of the fault surface has not been sufficiently considered, and therefore there are limitations in describing the damage mechanism of tunnels in active fault zones; the fault thrust of the loaded tunnels is often calculated by the actual fault thrust of a specific seismic hazard, and the potential fault thrust corresponding to the maximum potential earthquake magnitude of the area is missing. The loaded fault thrust is often calculated according to the actual fault thrust in a particular earthquake hazard, and there is a lack of extended research on the potential maximum thrust loading corresponding to the maximum potential magnitude in the region, thus underestimating the hazards of future potential fault-seismic activities on the tunnel.

Therefore, based on the previous research, this project will carry out fault dislocation-ground vibration coupling simulation for a cross-fault mountain tunnel in the western region. The simulated fault is a right-handed strike-slip dominated active rupture with a thrust component, and the surrounding rock conditions are IV (V in the nucleus of the rupture zone), and the overall slip condition of the rock body of the fracture zone in the process of uplift of the upper plate can be analyzed by inputting the fault dislocation-ground vibration load to reveal the development and evolution law of the "funnel zone" of the retrograde fault. It summarizes and generalizes the internal force distribution and deformation state of the tunnel, reveals the deformation and damage mode of the tunnel under the initial dislocation-strong earthquake coupling, and provides a theoretical basis for the design of seismic-resistant and fault-resistant design of tunnels through faults in high-intensity areas.

Geomechanical model with cross-fault tunnel

Fault parameter

The target tunnel passes through an active backlash fracture, in which the width of the nucleus of the fracture zone is 100m. The total width is 420m, the inclination of the fracture zone is 25°, and the dip angle is 70°, in which the rock bodies on both sides are mainly gneiss. The enclosing rock grade is IV (red part in fig.1), and the fracture zone and the nucleus of the fracture zone and the nucleus are mainly crushed rock, eroded breccia, and tectonic breccia, and the enclosing rock grade is V (green part in fig.1).
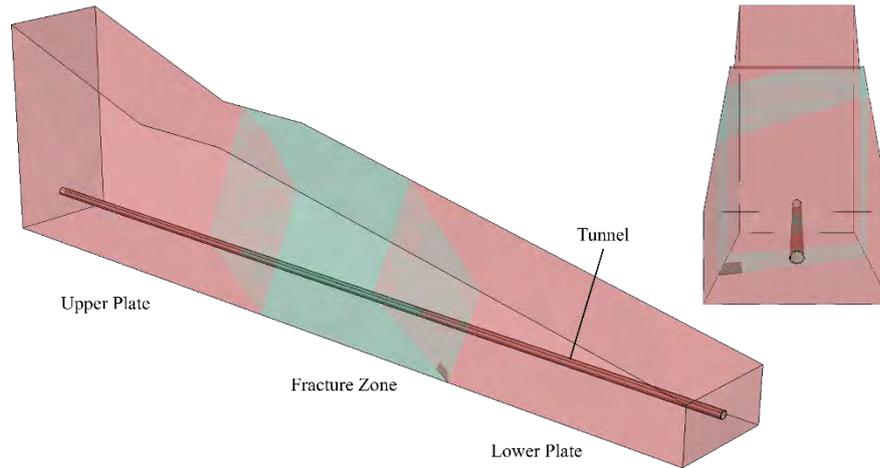
Figure 1. Tunnel-fault geology relation

To accurately simulate the physical response of the fault fracture zone during the uplift of the fault upper disk and to consider the computational efficiency, the intact rock bodies on both sides are simulated by the linear elasticity principal model. In contrast, the soft rock plasticity model simulates the core rock bodies [6][7]. The soft rock plasticity principal model proposed by Crook (2006) is suitable for simulating the physical response of soft rock and consolidated sand materials. It introduces the plastic flow and hardening criterion [7] based on Drucker- Prager principal model, which makes it have the characteristics of smooth yielding plane and finite yielding (yield surface closure), based on Crook. Prager's ontology is based on the introduction of the plastic flow criterion and hardening criterion [7], which makes the characteristics such as smooth yield plane and finite yielding (yield surface closure). Based on the isotropic nonlinear elastic yielding ontology model proposed by Crook [6], the equations for the yield surface are as follows:

$$\sqrt{(e_0 p_c^0 tan\beta)^2 + (g(\sigma)q)^2} - (p - p_t)tan\beta \left(\frac{p - p_c}{p_t - p_c}\right)^{\frac{1}{n_y}} - e_0 p_c^0 tan\beta \qquad (1)$$

Where: $p = -\frac{1}{3}trace\ \sigma$ is the equivalent pressure stress; q is the mises equivalent stress; p_c is the yield stress in hydrostatic compression; p_t is the yield stress in hydrostatic tension; p_c^0 is the initial value of p_c, β is the friction angle, n_y is the material parameter that controls the shape of the yield surface in the $\boldsymbol{p-q}$ plane; and e_0 is the eccentricity parameter.

The materiality parameters of the surrounding rocks and fault fracture zones are shown in Table 1 below:

Table 1. Parameters of rock mass

Area	Density $(kg \cdot m^{-3})$	Young's modulus (GPa)	Poisson's ratio	Cohesive force (MPa)	Friction angle
Upper plate	2600	3.5	0.3	0.6	35
Lower plate	2600	3.5	0.3	0.6	35
Fracture zone	2500	1.2	0.33	0.5	30

Tunnel parameter

Tunnel through the fault section maximum width of 15.8m, maximum height of 14.2m, within the headroom reserved space of 60cm (100 years of wrong amount of 30cm + 30cm reserved space for reinforcement), the arch wall adopts a circular inner profile, and deepen the back arch; at the same time, the use of increasing the amount of reserved deformation, strengthen the initial

support and over-support, curtain grouting of the broken section, second lining to strengthen and other measures. The lining is designed in longitudinal segments (6m) and extends 200m to each side, and deformation joints are set up between the lining segments.
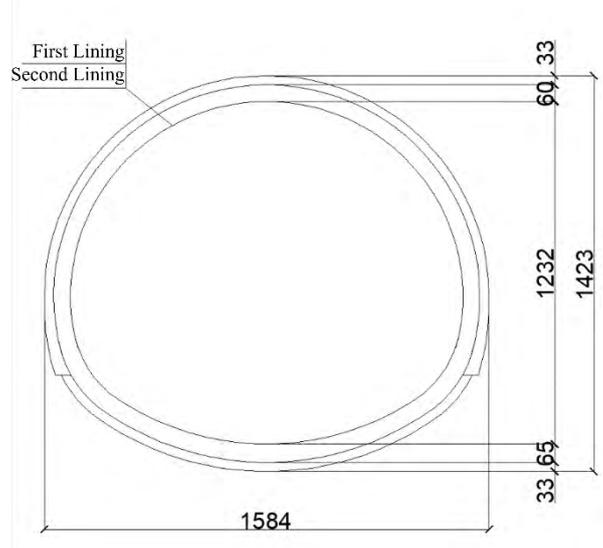
Figure 2. Cross-fault tunnel section

To better simulate the dynamic response of underground structures in the elastoplastic phase, the concrete is modeled by the concrete damage plasticity model. Based on the theoretical foundations of Lubliner et al. [8] and Lee and Fenve [9], the damage factor D is used to characterize the degree of damage and stiffness degradation of concrete.

$$\bar{\sigma} = (1 - D)E_0 \cdot (\varepsilon - \varepsilon_{pl}) \tag{2}$$

Where: $\bar{\sigma}$ is the effective stress, E_0 is the initial elastic stiffness, is the ε strain tensor, and ε_{pl} is the plastic strain tensor.

An isotropic homogeneous solid is used to model the tunnel, in which binding constraints connect the rock mass and two layers of linings. The lining material parameters are shown in Table 2.

Table 2. Parameters of lining

Area	Density $(kg \cdot m^{-3})$	Young's modulus (GPa)	Poisson's ratio	Cohesive force (MPa)	Friction angle
First lining	2300	25	0.2	1.4	65
Second lining	2500	32.5	0.2	1.5	65

Numerical modeling and computation

Finite element model

A general finite element software ABAQUS is used to establish the geomechanical power analysis model of cross-fault tunnel-enclosed rock. Different material properties are assigned to different areas of the surrounding rock to simulate the physical response differences between the upper and lower rock bodies and the fault fracture zone. To avoid the reflection effect of the model boundary in simulating, the model's width is 200m, and the model's height is taken according to the actual geological conditions. The size of the tunnel cross-section of the finite element model is 200m×100m~400m (the height varies along the axial direction of the tunnel), the length is 1510m, the grid encryption is carried out for the area around the tunnel lining, and the cell type adopts C3D10 cells. The schematic diagram of the finite element model is shown in Figure 3:
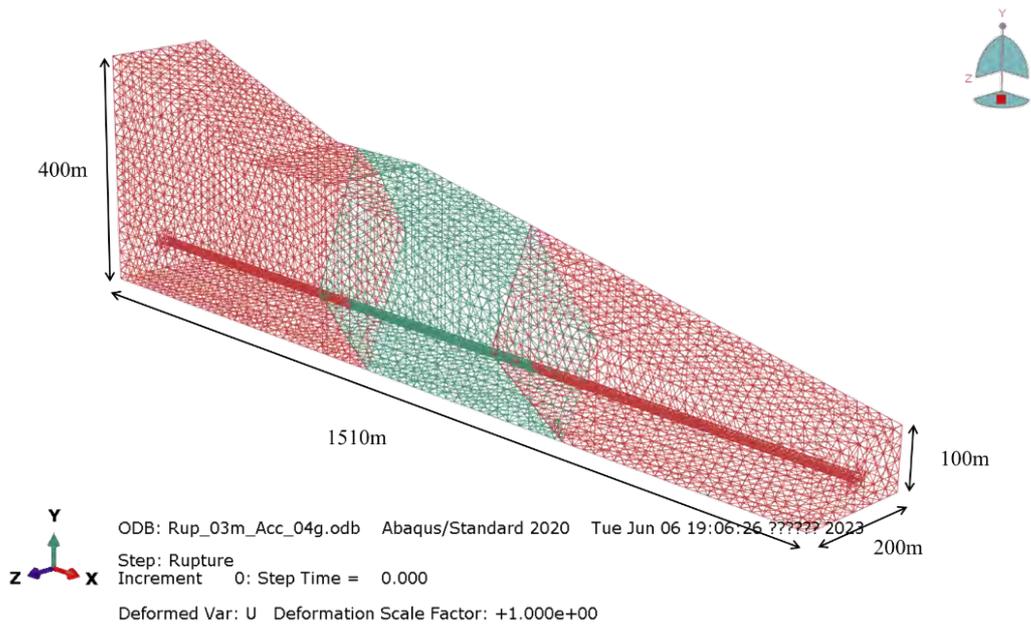
Figure 3. Finite element model

Among them, the fault zone is simulated according to the ground investigation data, and the dip angle of the fault zone is 70°, the inclination is 25°, and the angle with the tunnel axis is 30°. The ground investigation data shows that the fault zone's core is 420 m. Considering the lithological difference between the upper and lower plates of the fault zone and the fault zone, the material parameters of different peripheral rock grades are used to simulate the peripheral rock, and friction contact is set up at the interface of the fault zone. The peripheral rock and the contact conditions are chosen as the penalty function friction contact model with a coefficient of friction of 0.4. To simulate the interaction between the upper plate - the fault zone - the lower plate, the fault zone setup is shown in Fig. 4.
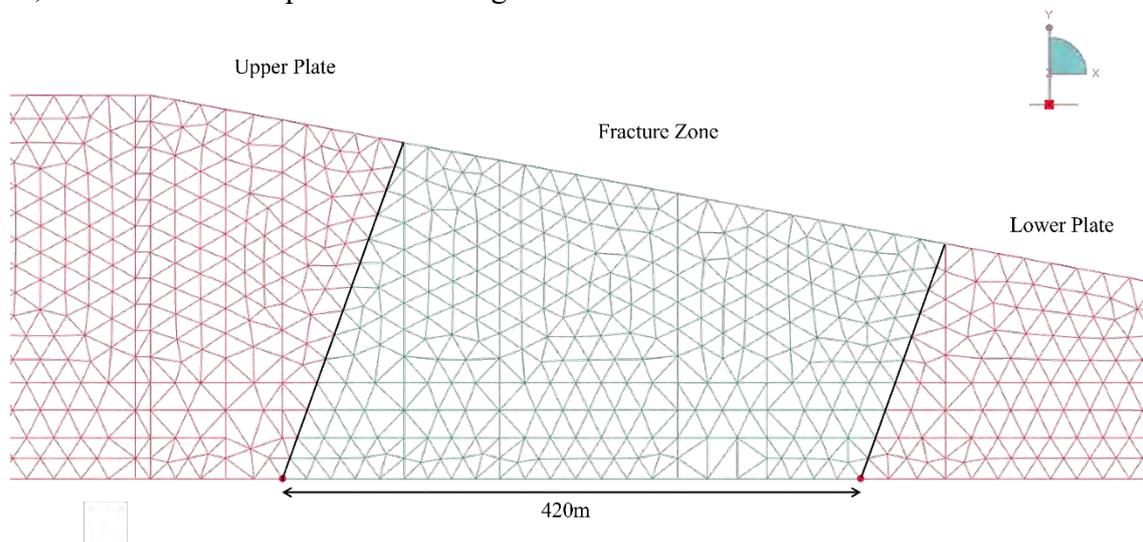
Figure 4. Fracture zone setting

Analysis step and boundary conditions

To study the disaster-causing mechanism of the tunnel under the joint action of faults and strong earthquakes, a two-step loading method is used for the analysis; firstly, the input of thrust displacement loads to the model is carried out in the static analysis step (Rupture step), and after the end of the static analysis step, it enters into the dynamic implicit analysis step (Acc_input step) to carry out the inputs for the ground shaking. In this study, the external load

inputs are backlash thrust input (static) and three-way ground vibration input (dynamic), and different boundary condition settings are used in different sections of the analysis.

For the static thrust analysis, a fixed constraint is applied to the lower plate boundary to constrain the overall lateral boundary displacement of the model. In the upper plate as the active plate, an overall displacement is applied during fault dislocation to simulate the thrust fault dislocation, and the boundary settings are shown in Fig. 5:

Boundary setting: Rupture step
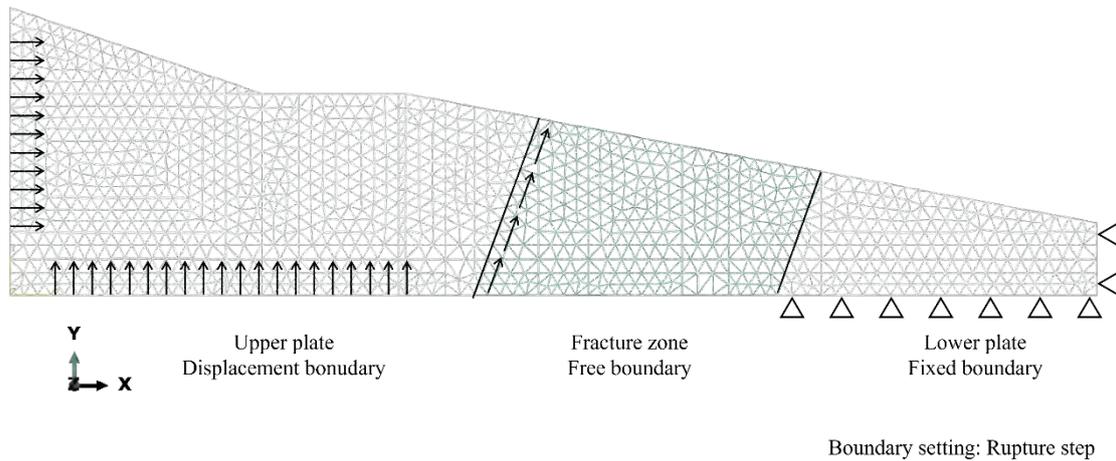
Figure 5. Rupture step boundary setting

In the power step analysis, no displacement constraints are applied at the bottom of the model due to the input of ground vibration loads, the top of the model is a free boundary, and MPC boundaries are used at the side boundaries to prevent reflections of the ground vibration inputs in the analysis domain and the boundary setup is shown in Fig. 6:

Boundary setting: Seismic step
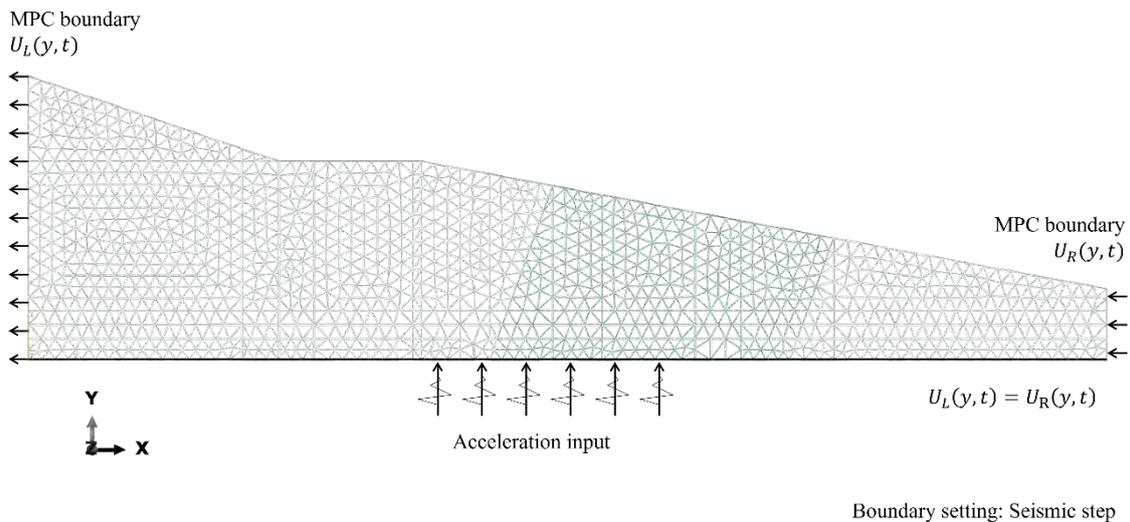
Figure 6. Seismic step boundary setting

Initial geostress

According to the given ground investigation data to model the rock body, using manual iteration for the ground stress balance, to simulate the stress drop during the unloading process of the tunnel excavation peripheral rock, in the process of calculating the ground stress balance of the

excavated part of the rock body (the core) for the stiffness reduction to simulate the stress drop, the core rock body stiffness for the intact rock body of 2/3, the other parameters remain unchanged, after two rounds of iteration of the ground stress balance, the maximum displacement of the soil body to reach 6E-05m to meet the requirements of the calculations, the ground stress balance of the displacement cloud diagram as shown in Fig. 7.
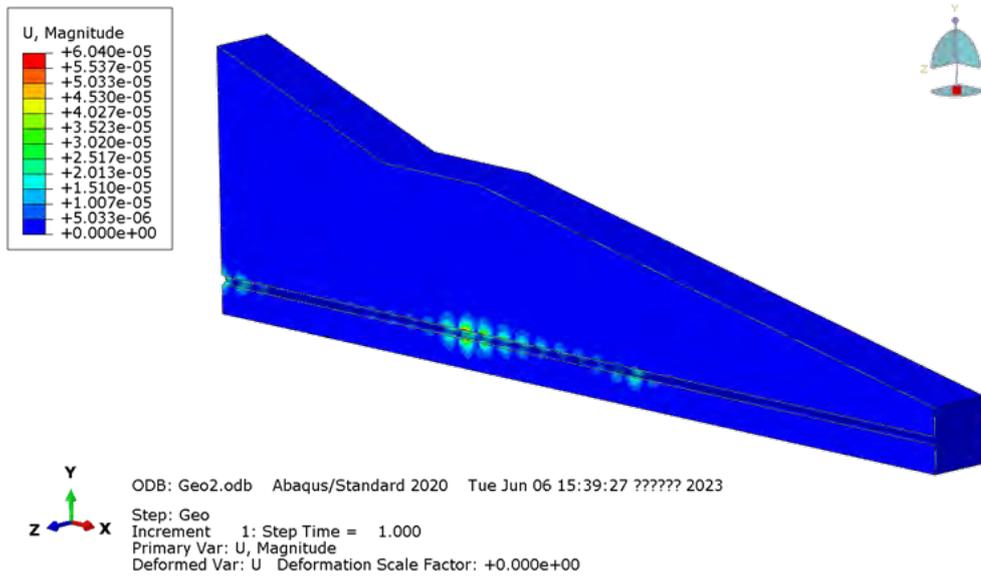
Figure 7. Initial geo-stress

Load input

According to the target tunnel through the fault of the current, the reserved space of the tunnel section is 60cm. This paper intends to simulate the tunnel backlash thrust component, the total thrust amount of 30cm, considering the fault strike angle and the tunnel angle, and the calculation of the tunnel fault thrust input vector component shown in Figure 8.
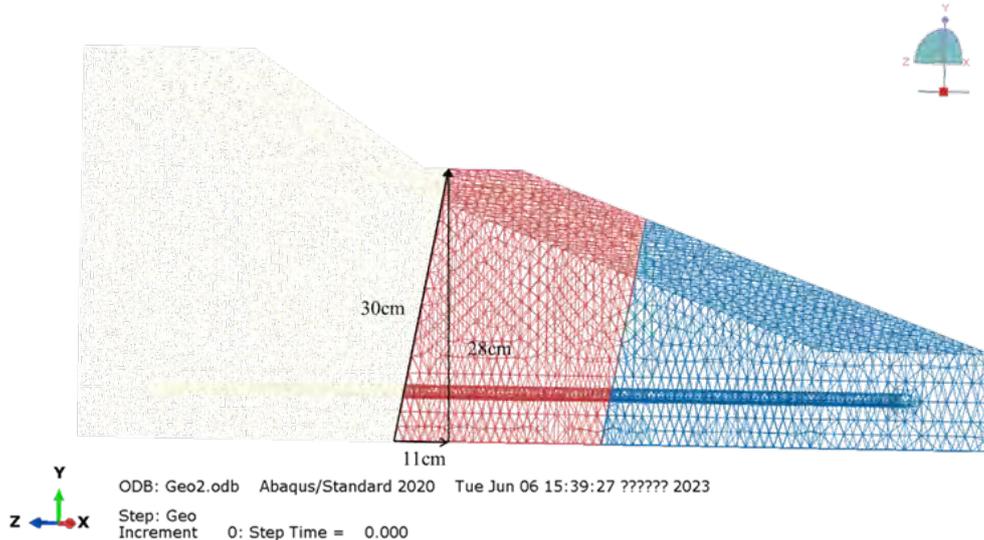
Figure 8. Input of thrust fault

When analyzing the tunnel through the fault under seismic action, selecting seismic waves as the dynamic load is necessary. Due to different acceleration peaks, spectral characteristics, duration, and other characteristics of different seismic waves, there will be differences in the impact on the structure, so the selection of appropriate seismic waves and reasonable processing for the analysis of the seismic law of mountain tunnels under seismic action has a vital role in the analysis of seismic law, which directly determines the analysis of the actual seismic

response of the tunnel structure to reflect more realistically the seismic response. This directly determines whether the analysis can genuinely reflect the actual seismic response of the tunnel structure. According to the site conditions and the maximum peak and spectral characteristics, and other related indicators, it is proposed to use the Wenchuan Wolong three-way seismic wave for the dynamic time-range analysis of the ground shaking input, the ground shaking data after the baseline correction and amplitude modulation (PGA = 0.4g), the time-range curve is shown in Figure 9.
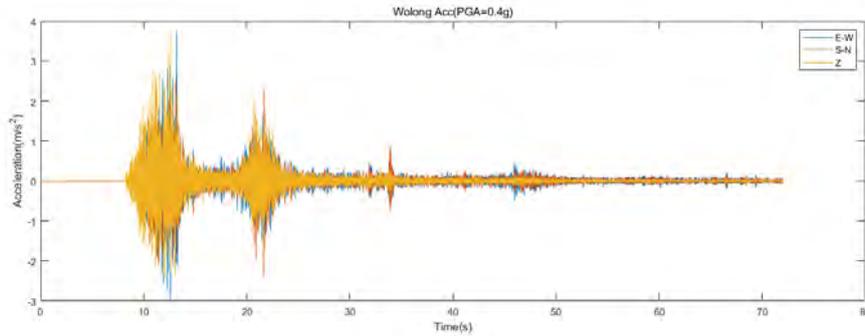
Figure 9. Seismic acceleration input

Dynamic response analysis

Displacement field of thrust fault

Figure 10 shows the rock displacement cloud map of the tunnel uplift process. Comparing the different stages of the backlash uplift of the active plate (upper plate), it can be found that in the early stage of the backlash uplift, there are apparent boundaries in the uplift displacement maps of the rock mass (e.g., d=8,14,17 cm). The displacement boundaries are in the fracture zone. The friction surface of this uplift is generated in the interior of the fracture zone this time. The interaction characteristics of the upper and lower plates' rock mess are dominated by the blocks' friction. With the further uplift of the active disk, in the region of the fracture zone with poor lithology, a triangular shear zone similar to that produced by the rupture of the soil layer under the action of fault thrust in soft soil [10] (e.g., d = 30 cm) appears. In this case, the uplift behavior of the triangular region in the rupture zone is differentiated from that of the upper and lower plates, and the shear zone as a whole is characterized by a widening of its extent and horizontal transport.
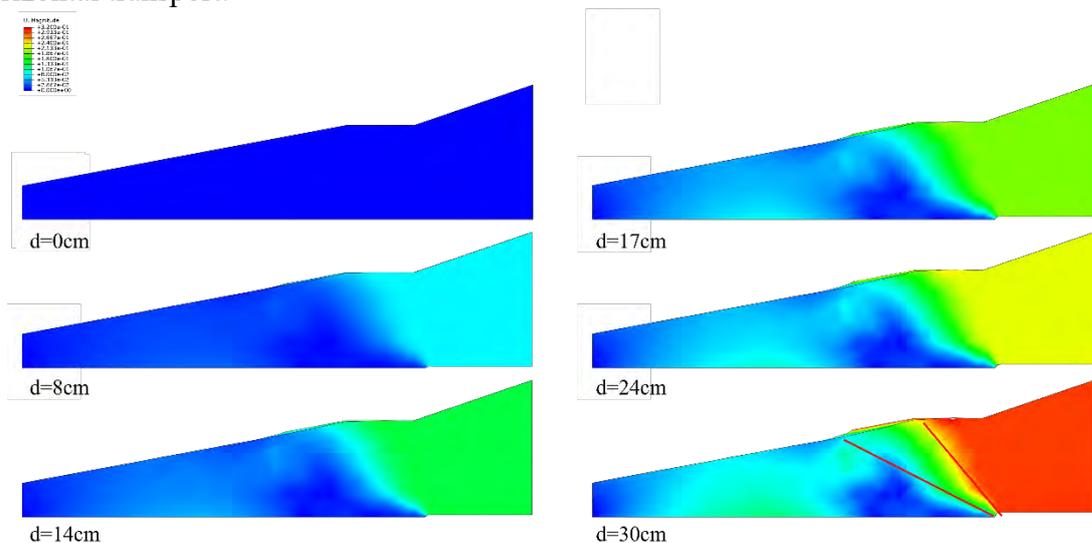
Figure 10. Displacement of the uplift process on reverse faults(d=0~30cm)

Figure 11 shows the displacement vector diagram of the surrounding rock under 30cm uplift, from which it can be seen that, with the gradual uplift of the active disc (red on the right side), the rock body of the central crushing zone shows a more significant large deformation flow tendency, and due to the pulling of the uplift of the upper disc, it makes the part of the crushing zone (green in the central part) appear a more detailed horizontal slip flow, and in the crushing zone - the contact surface of the lower disc shows the downward squeezing The trend is downward compression at the contact surface of the broken zone and the lower disc. After analysis, it is believed that this displacement trend is due to the significant difference in lithology between the upper and lower plate rock mass (red) and the rock mass of the crushed zone (green), which caused the large deformation of the crushed zone by lateral transport under the active lateral uplift. This lateral transport and extrusion trend is expected to make the crushed zone-lower plate junction area subject to additional loads.
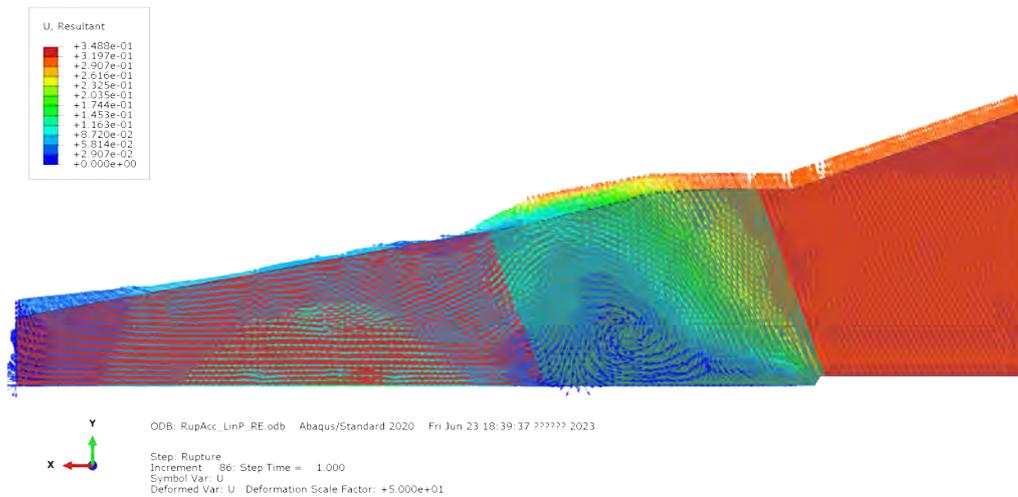
Figure 11. Displacement vectors for rock masses on reverse faults(d=30cm)

Tunnel response during slip

Figure 12 shows the overall deformation of the tunnel at the center-axis position. With the occurrence of the backlash fault thrust, the axial deformation of the tunnel is more apparent, and the maximum deformation reaches 28.6cm, basically the same as the result of the thrust input. And from the figure, the deformation of the tunnel after the location of the interface between the upper disc and the fracture zone has decreased significantly, mainly caused by the decrease in the stiffness of the surrounding rock in the fracture zone section, and the constraints become weaker.
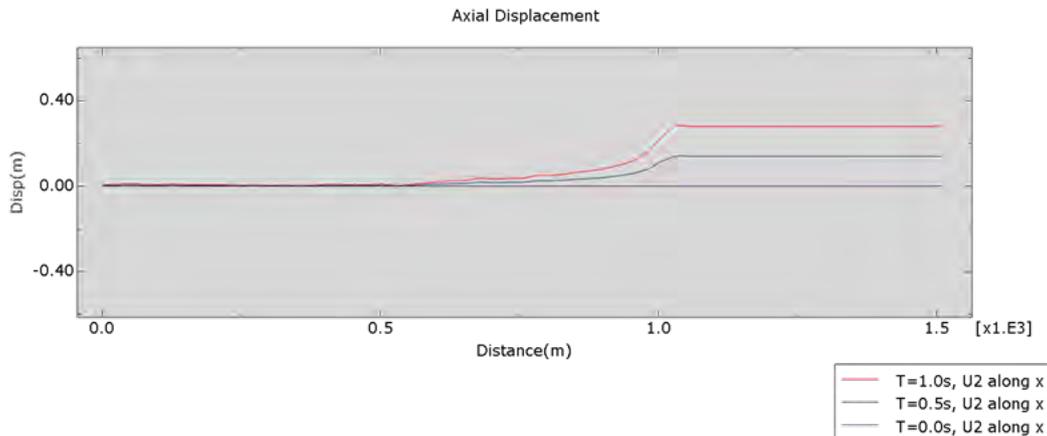
Figure 12. Tunnel axial deformation

The stress along the axial data of the tunnel up-arch location is output. The results are shown in Fig. 13, from which, with the thrust of the retrograde fault, the thrust interface (upper disc - rupture zone) produces a peak stress of 8.49 MPa, and the stress of the weak rock body in the rupture zone produces an apparent elevation. It is worth noting that, with the uplifting process of the fault, the Mises stress at the up-arch position of the rupture zone produces a tendency to decrease and then uplift. This change is caused by the change of the stress state at the up-arch position from compressive to tensile due to the uplifting.
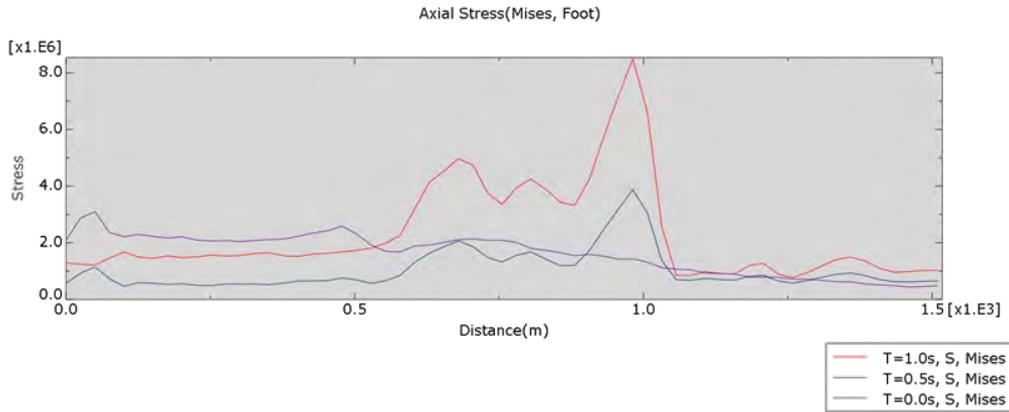
Figure 13. Tunnel axial stress(mises, tunnel invert)

The tunnel's stress and strain cloud diagrams during the fault uplift are shown in Fig. 14. From the figure. It can be seen that with the uplift of the reverse fault, the lining structure appeared to have two more obvious stress concentration areas, respectively, located in the upper plate-crushing zone interface due to the impact of the initial uplift of the upper plate shear effect, the stress rise is more prominent, with the further uplift, the bottom of the tunnel inside the crushing zone close to the side of the lower plate rock body also appeared a more obvious stress elevation, lining the stress state of The conclusion is consistent with the conclusion of 4.1 that the overall lateral transport of the rock mass in the crushing zone, which in turn caused the crushing load at the crushing zone-lower disc interface.
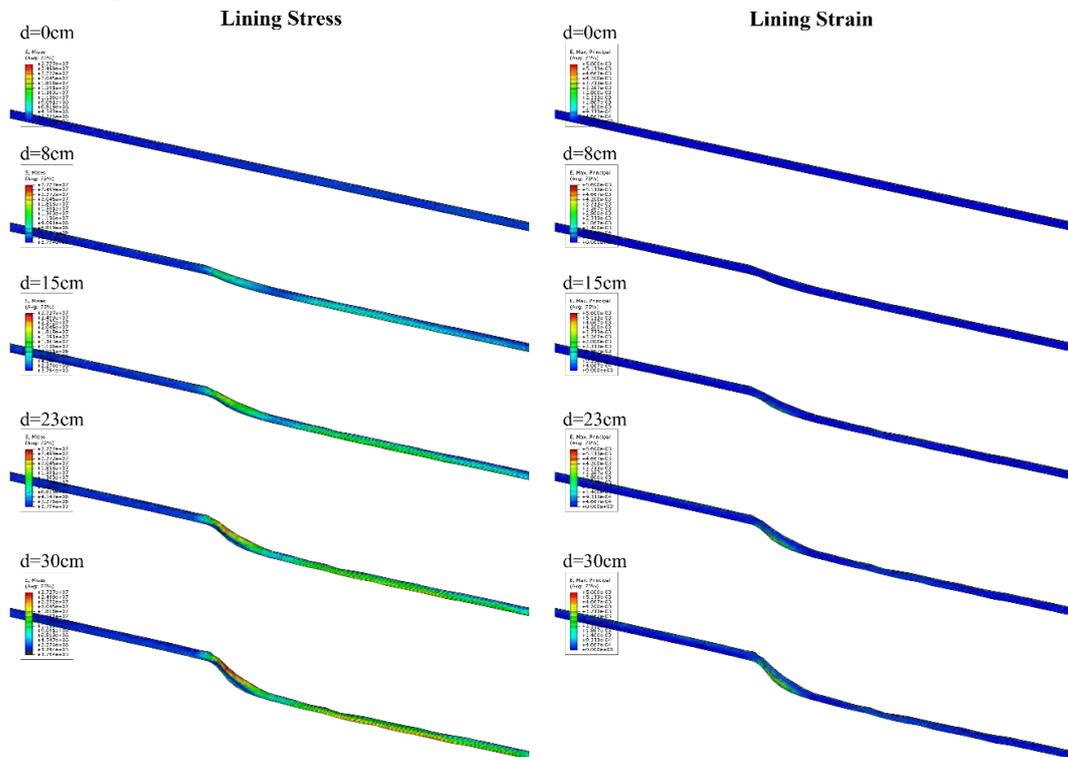
Figure 14. Lining response(rupture)

Seismic response

The results of seismic dynamic time-course calculations of the perimeter rock-lining model are analyzed, and the results of comparing the dynamic response of the perimeter rock at different moments are shown in Fig. 15:

| a. stress response | b. strain response |
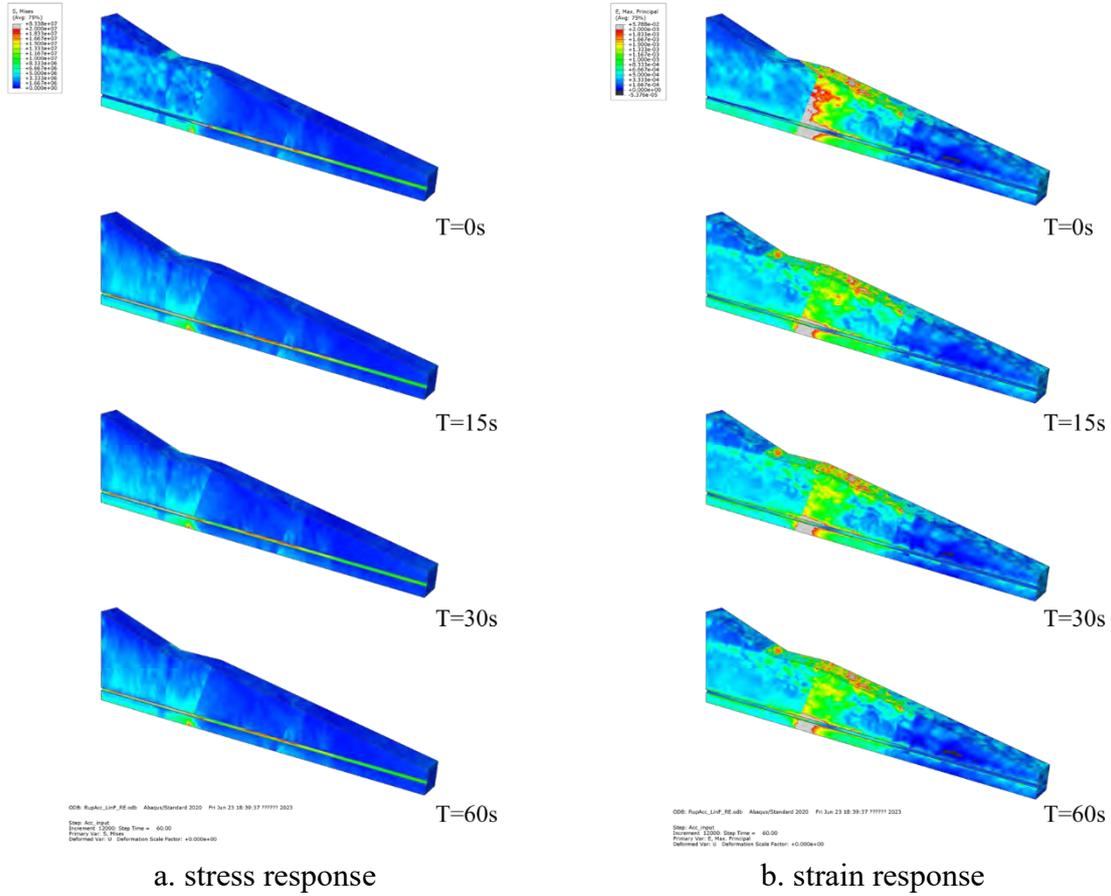
Figure 15. Dynamic response of rock mass

From the figure, it can be seen that the initial stress caused by the reverse-strike fault dislocation occupies the majority of the additional response of the surrounding rock, and the considerable deformation dislocation and stress concentration caused by it play a dominant role in the fault dislocation-strong seismic site response. It is worth noting that with the dynamic input of seismic waves, the surrounding rock stress at the interface of the upper disc-fracture zone shows a decreasing trend, which is initially presumed to be caused by the nonlinear surrounding rock materials with the introduction of the flow law, and to a certain extent the seismic action plays a role in releasing the accumulated strain caused by the long-term creeping and slipping of the fault.

Structural safety analysis

To clearly define the location of the analyzed section, the upper tunnel disc-fracture zone interface is defined as the mileage zero point, which is marked along the axial direction of the tunnel (x-direction), where the positive and negative directions are aligned with the x-axial direction of the model. The coordinates are defined as shown in Figure 16.
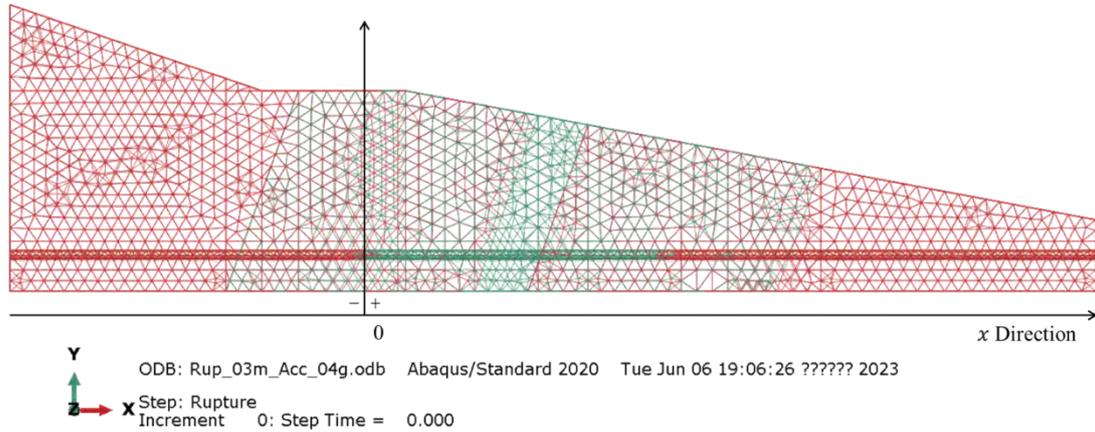
Figure 16. Section definition

Deformation analysis

Figure 17 shows the strain vector diagram of the primary lining structure between the upper disc-fracture zone interface (x=0m) and the fracture zone-lower disc interface (x=377.5m). It can be seen that, under the condition of the backlash dislocation amount of 30cm, the maximum strain value of the tunnel lining structure in this position after the thrust of the x=0m position reached 2.96E-03, which appeared in the position of the right side arch waist. The convergent deformation of the cave body (according to 45°) is from 1.29508e+01m to 1.28984e+01m, and the diameter deformation rate reaches 4.6‰; the maximum strain value of the tunnel lining structure in this position after the thrust of x=377.5m reaches 1.48E-03, and it occurs in the position of the supination arch. The convergent deformation of the cave body (according to 45°) is deformed from 1.29508e+01m to 1.29250e+01m, and the diameter deformation rate reaches 2.3‰.
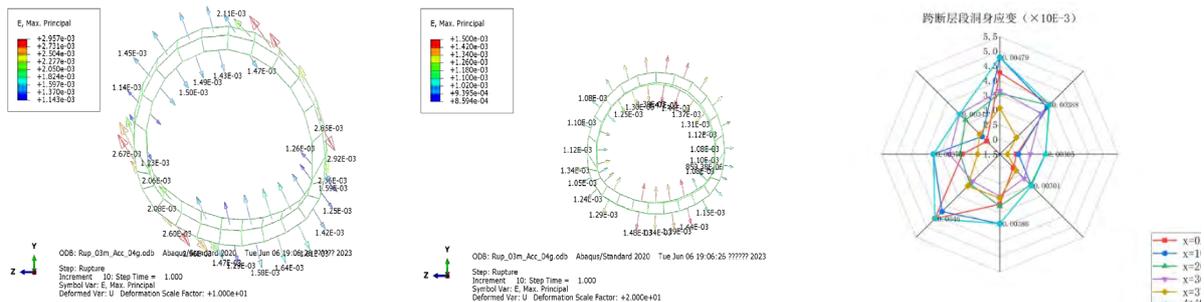
a. interface of the upper plate and fracture zone(x=0m)   a. interface of fracture zone and lower plate(x=0m)   c. strain envelope

Figure 17. Lining strain in fracture zone

Stress analysis

The stress map of the lining at the interfaces on both sides of the extracted fault zone is shown in Fig. 18, from which it can be seen that the lining in section x=0m shows obvious elliptical deformation, with stretching and compression along the 45-degree direction, and the maximum stress of 11.07 MPa occurs at the right arch foot. x=377.5m section has a smaller stress offset than section x=0, and the maximum stress of 11.55 MPa occurs at the left arch waist. Stress. To

113

summarise the stress results after the thrust is applied, the lining arch waist and arch foot position are selected as representative points, and the time course curves of the stresses at the lining arch waist and the elevation arch position at x=0m, 100m, 200m, and 377.5m are extracted. Their envelopes are extracted, and it can be seen in the figure that the peak of the overall strain of the tunnel occurs at the left arch waist (25.3MPa).

| a. interface of the upper plate and fracture zone(x=0m) | a. interface of fracture zone and lower plate(x=0m) | c. stress envelope |
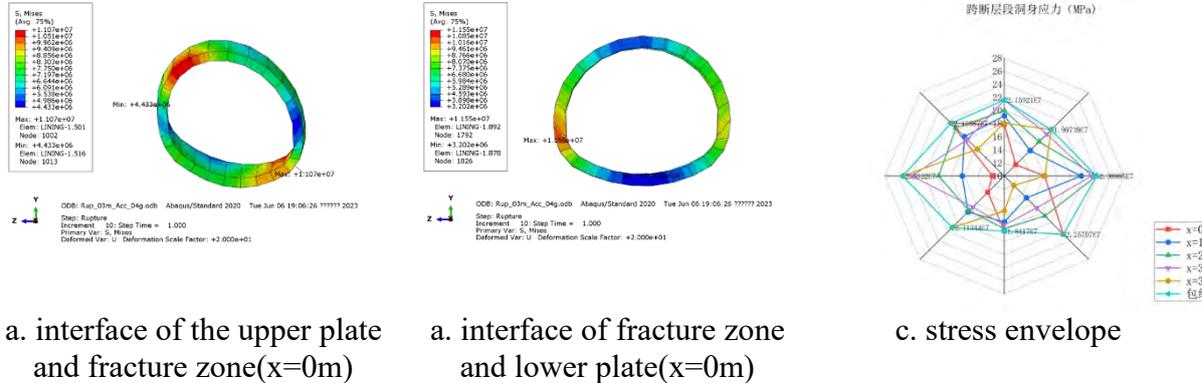
Figure 18. Lining stress in fracture zone

Conclusions

a. With the gradual lifting of the active disc, the rock body of the central crushing zone shows a more significant large deformation flow trend. Due to the pulling of the uplift of the upper disc, the part of the crushing zone shows a more detailed horizontal slip flow, and in the crushing zone - the lower disc contact surface shows a downward extrusion trend. After analysis, it is believed that this displacement trend is due to the significant difference in lithology between the upper and lower plate rock mass and the rock mass of the crushed zone, which makes the crushed zone under the action of active lateral uplift produce a large deformation behavior of transverse transport, and it is expected that this transverse transport, the trend of extrusion will make the crushed zone - lower plate junction area to bear additional loads;

b. The dynamic response of the tunnel is analyzed. Under the existing design conditions, considering the maximum dislocation condition and seismic condition, without considering the section design of the primary support and secondary lining, the maximum dislocation deformation of the tunnel reaches 28.6cm, and the maximum convergent deformation of the tunnel reaches 4.6‰ (diameter deformation rate), of which the maximum stress of 12.18MPa occurs in the location of x=300.00m, and the maximum total strain of 2.824E-03 occurs in the location of x=300.00m. 2.824E-03 appeared at position x=50.0m;

c. With the thrust of the reverse fault, a peak stress of 8.49 MPa was generated at the thrust interface (upper plate-fracture zone), and the stress in the weak rock body within the fracture zone was significantly elevated. The thrust surface and the adjacent area (within 50 m) appeared to have a more typical elliptical cross-section with 45° deformation. It is worth noting that, with the uplift process of the fault, it can be seen that the stress in the superelevation position of the rupture zone tends to decrease and then increase, and this change is caused by the change of the stress state of the superelevation position from compression to tension due to uplift, and the decrease of the stress occurs only in the side of the weakly constrained rupture zone, and does not occur in the side of the integrity of the rock body;

d. The tectonic and joint characteristics of the rock structure are more complicated, especially in the rock body under the influence of fault fracture zones, the tectonic development is more abundant, and it is difficult to obtain the initial stress field accurately, so there is a specific error in the setting of the in-situ stress field in the content of the present study, to further optimize the calculation results in the future using more detailed ground investigation data and the results of the study of the active faults;

e. The coupling between fault dislocation and ground shaking leads to a more complicated disaster stimulation condition, in which there is also the possibility of earthquake-induced stick-slip thrust, so further research can be carried out on the thrust-earthquake multi-catastrophe input condition setting.

References

[1] Zhang, W., Li, M., Ji, Y., Li, G., Zhao, L., & Li, S. (2022). Analysis and enlightenment of typical failure characteristics of tunnels caused by the Menyuan M6. 9 earthquake in Qinghai Province. China Earthquake Engineering Journal, 44(3), 661-669.

[2] He, Y., Sun, X., & Zhang, M. (2022). Investigation on the deformation of segment linings in cross-fault tunnel considering the creep behavior of surrounding rock during construction-operation period. Buildings, 12(10), 1648.

[3] Xuezeng, L., Lianglun, L., & Yunlong, S. (2012). Effect of thrust fault stick-slip rupture on road tunnel. Journal of Tongji University (Natural Science), 40(7), 1008-1014.

[4] Liu, X. Z., Wang, X. L., & Lin, L. L. (2013). Model experiment on effect of normal fault with 75 dip angle stick-slip dislocation on highway tunnel. Chinese Journal of Rock Mechanics and Engineering, 32(8), 1714-1720.

[5] Johansson, J., & Konagai, K. (2007). Fault induced permanent ground deformations: experimental verification of wet and dry soil, numerical findings' relation to field observations of tunnel damage and implications for design. Soil Dynamics and Earthquake Engineering, 27(10), 938-956.

[6] Bigoni, D., & Piccolroaz, A. (2004). Yield criteria for quasibrittle and frictional materials. International journal of solids and structures, 41(11-12), 2855-2878.

[7] Crook, A. J. L., Willson, S. M., Yu, J. G., & Owen, D. R. J. (2006). Predictive modelling of structure evolution in sandbox experiments. Journal of Structural Geology, 28(5), 729-744.

[8] Lubliner, J., Oliver, J., Oller, S., & Onate, E. (1989). A plastic-damage model for concrete. International Journal of solids and structures, 25(3), 299-326.

[9] Lee, J., & Fenves, G. L. (1998). Plastic-damage model for cyclic loading of concrete structures. Journal of engineering mechanics, 124(8), 892-900.

[10] Lin, M. L., Chung, C. F., & Jeng, F. S. (2006). Deformation of overburden soil induced by thrust fault slip. Engineering Geology, 88(1-2), 70-89.

Research on seismic performance indicator of coupled vehicle-track-tunnel system

Shengqi Xia[1], *Zhiyi Chen[1], and Liqun Li[1]

1 Department of Geotechnical Engineering College of Civil Engineering, Tongji University, Shanghai 200092

Abstract

With the continuous construction of high-speed railway tunnels, proper seismic performance indicator for high-speed railway tunnels concerning the train operation safety is urgently needed. To establish such indicator, a coupled vehicle-track-tunnel refined finite element model is developed through the joint simulation of HyperMesh and ABAQUS, in which the interaction between the subsystems and influence of uneven stratum are thoroughly considered. The model is further validated by comparing the numerical results with theoretical solution and measured data. Significant impact of soil layer heterogeneity is also illustrated, indicating the necessity of complex stratum modeling. Based on the refined finite element model, dynamic simulation of the vehicle-track-tunnel coupling system under seismic excitation is performed, and a two-level seismic performance indicator for high-speed railway tunnel is established, which offers a quantitative evaluation approach for seismic design. Though the established seismic performance indicator is proved effective in certain circumstances, further verification and refinement is still required in future research.

Keywords: High-speed railway tunnel; Vehicle-track-tunnel coupling system; Refined FEM model; Seismic performance indicator

1. Introduction

In recent years, the construction of subways and high-speed railways in China has developed at a rapid pace. By December 2021, 51 cities in 31 provinces (autonomous regions and municipalities directly under the Central Government) and the Xinjiang Production and Construction Corps had opened 269 urban railway lines with a total operating length of 8,708 kilometers; from 2020 to 2021, the total operating length of railways will exceed 150,000 kilometers, with more than 4,000 kilometers of new railways put into operation nationwide, including about 2,150 kilometers of high-speed railways. With more trains running at higher speeds, research into the seismic problems of trains has become more urgent. Past earthquakes have shown that earthquakes not only have a negative impact on bridges, road foundations, tunnels, and other structures, but also pose a threat to the safety of train operations: for example, the 1995 magnitude 7.3 earthquake in Kobe, Japan, caused varying degrees of damage to some underground stations and interdistrict tunnels and derailed 15 trains; in 2004, a magnitude 6.8 earthquake in Niigata Prefecture derailed eight out of ten cars on Shinkansen Line 325 near Nagata Station; in 2022, a magnitude 6.9 earthquake in Hualien caused severe twisting of the tracks; and in the same year, the magnitude 6.9 earthquake in Menyuan, Qinghai, caused misalignment and damage to the tunnel lining and bulging of the roadbed[1]. To ensure the safety of high-speed train operation under earthquake, research on coupled train-track-structure

systems is needed.

Regarding the issue of train operation safety in high-speed railway tunnels under earthquake action, some scholars have carried out valuable research work. The issue of train operation safety in high-speed railway tunnels under earthquake action has been studied by some scholars. Yang[2] assumed that the whole ballast track and tunnel structure deformed together, and the safety of train operation and the influence of subgrade isolation components were evaluated from the aspects of deformation and vibration amplitude of track structure under seismic load. Yang[3] established a three-dimensional finite element model of tunnel-train system under earthquake action based on the characteristics of soil-tunnel-train dynamic interaction, and an explicit parallel computing method based on contact load balance was put forward to do the calculation. The dynamic response of tunnel-train system under earthquake action was studied, and the operation safety of subway train was evaluated. The influence of operation speed and earthquake intensity on subway operation safety was also analyzed. Li[4] established a simulation model of train-tunnel-soil system, through which the nonlinear seismic response of train-tunnel-soil system was studied, and the safety operation index of subway vehicle was evaluated. Meanwhile, the lining and subgrade of railway tunnels were also researched by some scholars. Zhang[5] used a simple two-dimensional model to study the dynamic response of tunnel under vehicle-tunnel coupling action based on an example of railway tunnel engineering. Su[6], using finite difference software, established a calculation model of high-speed railway tunnel structure and surrounding rock system. The dynamic response law of tunnel lining structure when the incident angle of seismic wave changes in the horizontal plane was systematically studied. Italian scholar Fabozzi[7] established a process that uses seismic hazard decomposition to define a set of virtual sources that potentially affect railway tunnels. Through this set of sources, the seismic damage probability of tunnel structures and the available time for implementing real-time response measures could be calculated. Zhu[8] used finite difference method to simulate and analyze the displacement and velocity response of a subway tunnel subgrade structure under earthquake, and the action characteristics of earthquakes on lower part structure and subgrade of tunnel were discussed.

It can be seen from the above research that the adverse effects of earthquake on vehicle-track-tunnel coupling system have been paid attention to by existing research, but most related models have simplified the treatment of stratum due to the limitations of modeling means and technical difficulties, which makes it difficult to reflect the dynamic response of train-tunnel system under earthquake in real environment; on the other hand, there is still a lack of verification indicators for train operation safety applicable to high-speed railway tunnels. In order to establish a numerical model for subsequent dynamic response analysis, this study first explored the modeling method for establishing a refined model of vehicle-track-tunnel coupling in complex stratum through HyperMesh and ABAQUS joint simulation based on the relevant theory of vehicle-track dynamics. The establishment of the model included not only geometric model establishment, mesh division technology and method, track structure establishment, wheel-rail contact relationship consideration, etc., but also the use of material constitutive model, boundary condition processing and seismic input method, etc., were discussed. So that the established numerical analysis model could be as close as possible to engineering practice, increasing the reliability of subsequent analysis. On this basis, the complex model established was compared with the simplified model used in most studies to reflect the superiority of the

analysis model. Finally, by carrying out seismic response analysis of vehicle-track-tunnel coupling system, two-level seismic performance indicators for high-speed railway tunnels were established and verified.

2. Three-dimensional refined finite element model

2.1. Tunnel-soil model

Previous research and engineering experience have shown[9–14] that factors such as the existence of stratum heterogeneity and weak interlayer will affect the seismic safety of tunnel engineering to varying degrees. With the development of tunnel construction towards deeper, longer and larger directions, the influence of heterogeneous stratum on tunnels will become more difficult to predict. Therefore, for important projects, it is necessary to consider the complexity of the surrounding stratum when establishing tunnel-stratum models. To meet the needs of subsequent analysis of high-speed train operation safety, this paper selects a section of 1000m length of soil layer transition section of Yangtze River Tunnel along Jiangsu-Zhejiang High-speed Railway for modeling when establishing the analysis model, and considers the real fluctuation situation of the section soil layer when modeling. According to relevant survey data, the position of bedrock surface in the selected section is about 100m below the ground surface, so the soil layer depth is taken as 100m for convenience of seismic wave input. The schematic diagram of the selected section is shown in Figure 2-1.
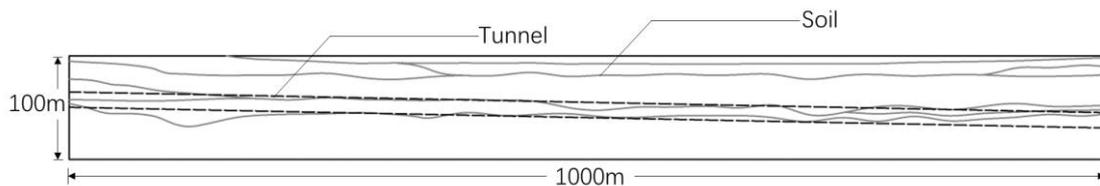
Figure 2-1 Schematic diagram of the longitudinal section of the analysis interval of the soil transition section

According to the conclusion of research[15,16], when using simplified artificial boundary as model dynamic boundary, under the premise of considering soil nonlinearity, the calculation results that meet the engineering accuracy requirements can be obtained when the model width-height ratio B/H=3. When selecting the longitudinal section scale, the soil layer depth is taken as 100m, so according to the above research conclusion, the model transverse width can be set as 300m, and a three-dimensional model can be established by stretching the two-dimensional plane soil layer curve in transverse direction. Thus, the overall size of the model is 100m×300m×1000m. Considering the strong nonlinearity of complex system coupling between vehicle-track-tunnel, this study adopts homogeneous ring method to establish tunnel model when establishing tunnel model, and considers lining and joint action through stiffness reduction coefficient. The three-dimensional finite element model of tunnel-stratum is shown in Figure 2-2.
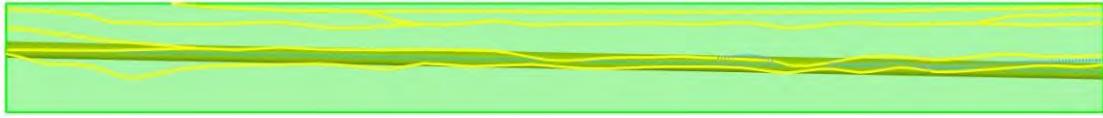
(a)   Longitudinal section diagram
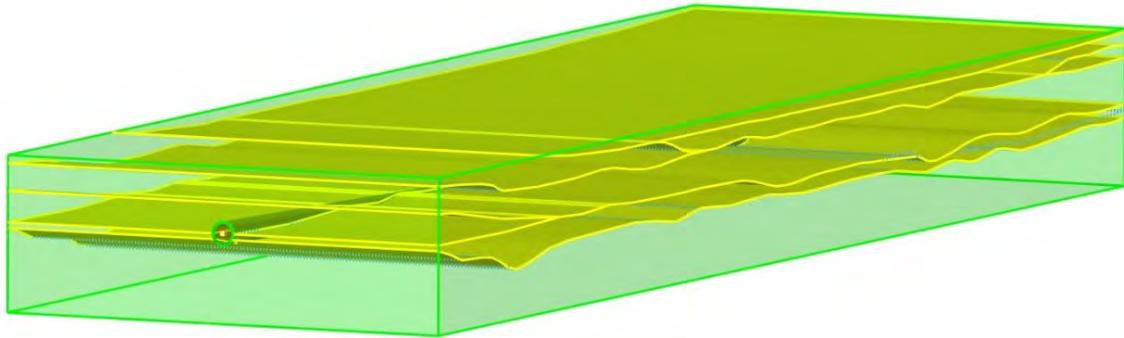
(b)   Schematic diagram of the 3D model

Figure 2-2 Schematic diagram of the finite element model

In the modeling process of this paper, linear elastic model is adopted in both soil and tunnel lining. To restore the dissipation effect of seismic wave in soil layer, Rayleigh damping is adopted for model damping, and the coefficient is determined according to the circular frequency of the first two modes of modal analysis. The stratum parameters involved in soil layer modeling are shown in Table 2-1, and tunnel lining material parameters are shown in Table 2-2.

Table 2-1 Stratigraphic parameters

Earth	No	Thickness (m)	Top of storey elevation (m)	Density (kg/m³)	Shear wave speed (m/s)	Poisson's ratio
Silt	1	6 to 11	0 to -1.57	1860	103	0.3
Powdered sand	2	4 to 11	-1 to -24	1920	113	0.3
Powdered sand	3	4.1 to 10.2	-6 to -32	1920	132	0.29
Silty powdery clay	4	12.8 to 33.2	2.72 to -16.15	1760	186	0.4
Powdered earth	5	4 to 10.6	-39.3 to -55.8	1780	322	0.3
Fine sand	6	6.9 to 38.2	-62.5 to -100.7	2000	419	0.29

Table 2-2 Tunnel lining material parameters

Materials	Density (kg/m³)	Modulus of elasticity (Pa)	Poisson's ratio
C60 concrete	2600	3.6×10^{10}	0.2

In the three-dimensional model, when tunnel passes through longitudinal heterogeneous soil layer, it will cut and produce complex geometric curves, which brings great difficulty to mesh division. The pre-processing mesh division tool provided by ABAQUS cannot effectively deal with such complex geometric curves. Research shows[17] that using HyperMesh to divide the mesh of complex rock and soil engineering model can greatly simplify the mesh division process, save the pre-processing time while the calculation results remain high accuracy and reliability, thus improving the quality of finite element analysis. Therefore, this paper uses the professional finite element pre-processing software HyperMesh to divide the mesh of this complex three-dimensional model. When dividing the tunnel mesh, because its geometric structure is relatively simple, the two-dimensional surface mesh can be divided at both ends of the tunnel first, and then the Solid Map function in the HyperMesh three-dimensional mesh division module can be used to divide the hexahedral mesh along the longitudinal direction. The reason for choosing hexahedral mesh is that compared with tetrahedral mesh, hexahedral mesh can reduce the number of meshes required for analysis and improve the calculation accuracy, and it is also convenient for post-processing of tunnel structure. When dividing the soil layer mesh, the software cannot automatically generate hexahedral mesh for complex geometric forms, so Tetramesh function with more mature algorithm is used to automatically divide tetrahedral mesh for soil layer. The finite element mesh generated by the above method not only effectively restores the complex three-dimensional geometric shape of tunnel-soil coupling model, but also ensures the accuracy of seismic response analysis of tunnel lining.

2.2. Vehicle-track model

With the continuous increase of train running speed, the interaction problem between railway vehicle and track system becomes more prominent and complex. Academician Zhai Wanming[18] believes that the classical vehicle dynamics and track dynamics theory system treats locomotive vehicle and track as two relatively independent subsystems, which cannot solve the complex dynamic interaction problem of vehicle-track system. In order to solve the above difficulties, Zhai and his research team have formed a mature, widely recognized and authoritative vehicle-track coupling dynamics theory after decades of continuous efforts and exploration. The establishment of vehicle-track dynamic model in this paper mainly refers to the above research results. This paper takes CRHII type EMU as the prototype of train and establishes the dynamic model of vehicle-track system. When establishing the vehicle model, the train is simulated as a multi-rigid-body system, consisting of car body, bogie, wheel, primary suspension and secondary suspension, etc. Each rigid body contains five degrees of freedom: vertical, lateral, roll, yaw and pitch. Each car has 35 degrees of freedom. The rigid bodies are connected by spring-damper units, considering the stiffness and damping in longitudinal, lateral and vertical directions respectively.

(b)   Passenger vehicle-track spatially coupled model (top view)

(a)   Passenger vehicle-track spatially coupled model (end view)
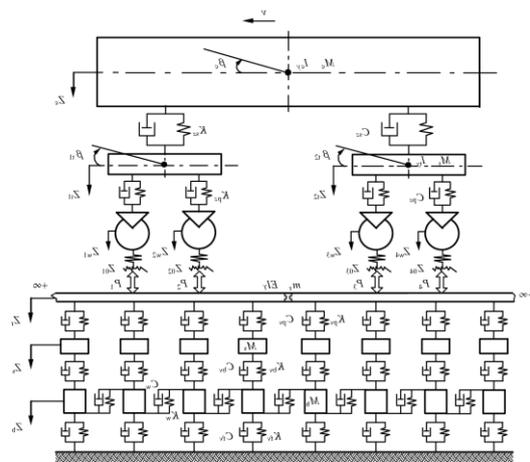
(c)   Passenger vehicle-track spatially coupled model (side view)

Figure 2-3 Schematic diagram of the vehicle-track dynamics model[18]
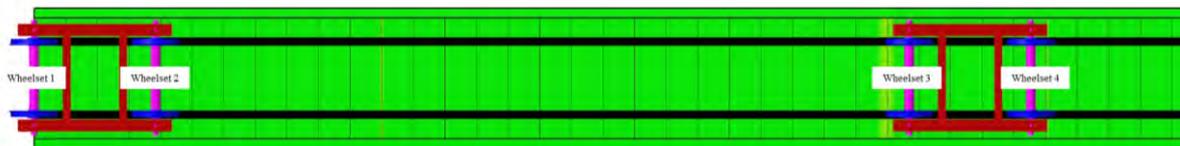
(a)   Corresponding number for each wheelset
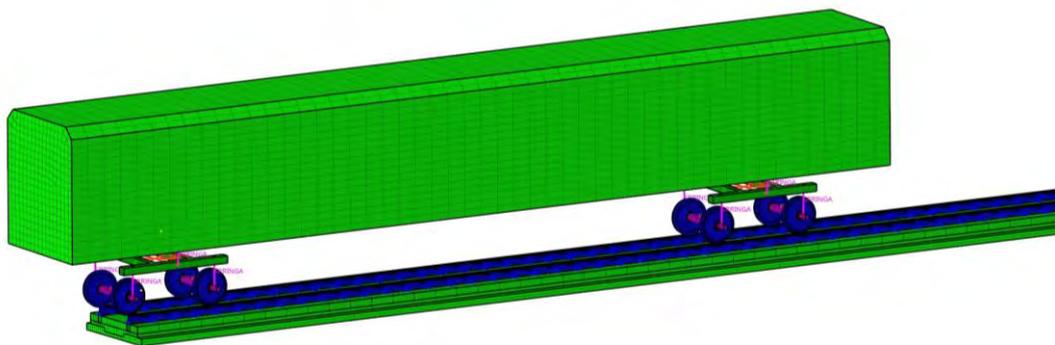
(b)   Refined model of the vehicle-track system

Figure 2-4 Finite element model of the vehicle-track system

The establishment of track model takes CRTSII type slab ballastless track as the prototype, and the model consists of steel rail, fastener, track slab, cement asphalt mortar layer and concrete support layer. Among them, mortar layer and concrete support layer are modeled by solid elements, and fasteners are simplified as a series of spring-damper elements, considering the stiffness and damping in vertical and lateral directions. The main modeling parameters of vehicle-track model are shown in Table 2-3.

Table 2-3 Main parameters for vehicle-track model

Parameters	Value
Fastener vertical stiffness (kN/mm)	60
Fastener damping in vertical direction (kN·s/m)	25
Fastener transverse stiffness (kN/mm)	30
Fastener lateral damping (kN·s/m)	25
One-system suspension vertical stiffness (N/m)	1176000
One-series suspension damping (N·s/m)	19600
Series I suspension lateral stiffness (N/m)	980000
Longitudinal stiffness (N/m) of one series suspension	980000
Second system suspension vertical stiffness (N/m)	40,000
Second series suspension lateral stiffness (N/m)	158760
Second series suspension lateral damping (N·s/m)	58800
Longitudinal stiffness of the second system suspension (N/m)	158760
Body mass (kg)	34000
Wheel pair mass (kg)	1400
Bogie mass (kg)	3200

2.3.   Coupling relationships between subsystems

2.3.1.   Soil-tunnel coupling system

In order to realize the overall dynamic solution of vehicle-track-tunnel-soil coupling system under earthquake action, it is vital to establish accurate coupling relationship between the above subsystems.

There is a complex dynamic contact effect between underground structure and surrounding soil. In the related research of underground structure seismic analysis, there are mainly three methods to define the mechanical behavior of soil-structure contact surface: integral method, contact element method and contact surface pair method[19]. The integral method treats soil and underground structure as a whole for finite element mesh division, and there is no contact interface between them, only difference in material properties. This method is simple to operate and has high solving efficiency, but it often leads to excessive structural response. Contact element method is able to consider the friction and relative sliding between soil and structure

contact surface within small deformation range. On this basis, contact surface pair method can further simulate the dynamic contact nonlinearity problem of large relative displacement and separation between soil and underground structure. However, these two calculation methods will significantly reduce the efficiency of numerical calculation. For complex coupling system dynamics problems with high nonlinearity, appropriate contact relationship should be selected to balance the accuracy and cost of numerical calculation. In ABAQUS, Tie contact is a special form of surface-to-surface contact, and there is no relative movement and separation between the master surface and slave surface of binding constraint, and the slave surface displacement is determined by the master surface displacement. This contact mode balances the accuracy and efficiency of numerical calculation well, and is widely used in tunnel seismic response research. In addition, Zhuang Haiyang[20] believes that the calculation results of underground structure dynamic response without considering contact effect are biased to safety, which is suitable for engineering seismic response analysis of underground structure. Therefore, this paper uses Tie constraint in ABAQUS to realize the coupling of tunnel and surrounding soil.

2.3.2. Train-track coupling system

The rolling contact behavior of wheel and rail is an important content of vehicle system dynamics research, and the wheel-rail action directly affects the motion stability, ride comfort and operation safety of train[21]. Therefore, the accurate description of wheel-rail contact behavior is the premise of studying the train operation safety in coupling system under earthquake action. In actual situation, because both wheel and rail are deformable elastic bodies, they will produce a contact surface when contacting, which is called "contact patch". To study the shape, size and force distribution relationship of contact patch, scholars have established different rolling contact mechanics models. The earliest contact theory was initiated by Hertz in 1882, which first strictly derived the normal contact solution of two contact bodies[22], and proposed the assumption that the contact patch between two elastic bodies is elliptical. Although Hertz contact theory is based on some ideal assumptions, its calculation speed of contact problem is fast, and it meets the engineering accuracy requirements, so it is still widely used in many current research fields. However, for derailment problems, factors such as friction force on wheel-rail contact surface and conformal contact between wheel flange and rail surface will significantly affect vehicle dynamic behavior, and Hertz theory cannot meet vehicle dynamic analysis needs. Non-Hertz contact theory is a theory that arises to solve complex wheel-rail contact problems, among which the most widely recognized one in the industry is Kalker's three-dimensional non-Hertz elastic body rolling contact theory, that is, exact theory or Kalker's variational method[23,24]. The idea of Kalker's exact theory is to transform the basic classical elasticity equations into the residual energy expression of rolling contact problem by displacement variation, so that rolling contact problem is transformed into a mathematical programming problem. The boundary element numerical calculation program CONTACT compiled to solve the above linear programming problem can solve Hertz and non-Hertz normal problems, incremental form slip contact problems and steady-state and non-steady-state rolling contact problems. Because of its accuracy and completeness of calculation results, CONTACT is often used as a verification program for wheel-rail contact related research.

With the improvement of computer performance and the development of calculation methods, finite element method has been widely used in wheel-rail rolling contact problem analysis,

which can better deal with physical nonlinearity, geometric nonlinearity and boundary condition nonlinearity in wheel-rail rolling contact problem. In ABAQUS, a large-scale finite element software, the accurate simulation of wheel-rail contact relationship can be realized by using contact pair algorithm. Because the mesh division of wheel is more refined than that of rail, when defining contact surface, rail surface is selected as master surface and wheel surface as slave surface. The definition of contact property needs to consider normal relationship and tangential relationship between contact surfaces. In principle, rail mesh and wheel mesh should not penetrate each other in normal direction, that is, normal contact relationship is defined as "hard contact". But in dynamic analysis, such strict contact relationship is likely to lead to numerical singularity and make calculation not converge. To solve more effectively and accurately, "soft contact" that allows penetration between contact surfaces needs to be used to define contact relationship. An[25] believes that although there are obvious differences between Hertz-type gap modeling results and non-Hertz-type modeling methods, these differences do not significantly affect the effectiveness of Hertz-type gap modeling in predicting normal force in vehicle-track dynamics. Li[26] also proved that Hertz theory is applicable to standard wheel-rail profile matching contact problem by comparing calculation results of several normal problem solving models. Therefore, this paper calculates normal force-penetration relationship according to Hertz nonlinear contact theory, and imports ABAQUS in table form to define normal contact stiffness. The calculation formula of Hertz nonlinear contact theory is[27]

$$P(t) = \begin{cases} \left[\frac{1}{G}\delta Z(t)\right]^{\frac{3}{2}} & \delta Z(t) > 0 \\ 0 & \delta Z(t) \leq 0 \end{cases} \qquad (1)$$

Where $\delta Z(t)$ denotes the magnitude of the elastic compression displacement occurring between the wheels and rails, G is the wheel-rail contact constant, and for abrasive treads there is $G = 3.68 \times 10^{-8} R^{-0.115} (m/N^{2/3})$, where R is the actual rolling circle radius.

The calculated wheel-track force-penetration relationship is shown in Figure 2-5. Luo and Shi[28] believe that, in simplified cases, the contact stiffness can be approximately taken as $5 \times 10^8 N/m$ for calculation.
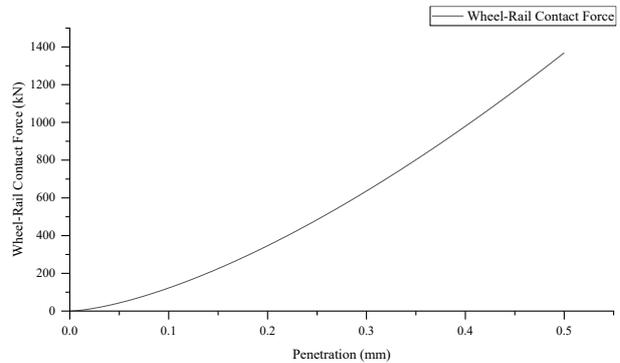
Figure 2-5 Wheel-rail contact force-penetration curve

2.4. *Three-dimensional refined finite element model of the coupled vehicle-track-tunnel system*

Xu and Zhai[29] identified several challenges in vehicle-track-tunnel (VTT) analysis: (1) simplification of large-scale computational model; (2) treatment of VTT as an integrated system, rather than imposing the load induced by vehicle-track interaction as excitation on the supporting structure; (3) enhancement of computational efficiency for solving the dynamic

model of high degree of freedom system; (4) incorporation of stratum boundary conditions. To address these key technical issues, they developed a comprehensive VTT dynamics model, in which the track-tunnel system was modeled by finite element method, and the surrounding soil layer was modeled by infinite element method. This dynamic model not only accounted for the discontinuity of tunnel lining, but also coupled the subsystems within the VTT interaction framework for the first time through vehicle-track coupling dynamics[18]. The matrix equation of the dynamic model is as follows

$$
\begin{bmatrix} \boldsymbol{M}_{vv} & \boldsymbol{M}_{vr} & 0 & 0 \\ \boldsymbol{M}_{rv} & \boldsymbol{M}_{rr} & 0 & 0 \\ 0 & 0 & \boldsymbol{M}_{tt} & 0 \\ 0 & 0 & 0 & \boldsymbol{M}_{ss} \end{bmatrix} \begin{Bmatrix} \ddot{\boldsymbol{X}}_v \\ \ddot{\boldsymbol{X}}_r \\ \ddot{\boldsymbol{X}}_t \\ \ddot{\boldsymbol{X}}_s \end{Bmatrix} + \begin{bmatrix} \boldsymbol{C}_{vy} & \boldsymbol{C}_{vr} & 0 & 0 \\ \boldsymbol{C}_{vv} & \boldsymbol{C}_{\pi} & \boldsymbol{C}_{rt} & 0 \\ 0 & \boldsymbol{C}_{tr} & \boldsymbol{C}_{tt} & \boldsymbol{C}_{ts} \\ 0 & 0 & \boldsymbol{C}_{st} & \boldsymbol{C}_{ss} \end{bmatrix} \begin{Bmatrix} \dot{\boldsymbol{X}}_v \\ \dot{\boldsymbol{X}}_r \\ \dot{\boldsymbol{X}}_t \\ \dot{\boldsymbol{X}}_s \end{Bmatrix}
$$
$$
+ \begin{bmatrix} \boldsymbol{K}_{vv} & \boldsymbol{K}_{vr} & 0 & 0 \\ \boldsymbol{K}_{rv} & \boldsymbol{K}_{rr} & \boldsymbol{K}_{rt} & 0 \\ 0 & \boldsymbol{K}_{tr} & \boldsymbol{K}_{tt} & \boldsymbol{K}_{ts} \\ 0 & 0 & \boldsymbol{K}_{ts} & \boldsymbol{K}_{ss} \end{bmatrix} \begin{Bmatrix} \boldsymbol{X}_v \\ \boldsymbol{X}_r \\ \boldsymbol{X}_t \\ \boldsymbol{X}_s \end{Bmatrix} = \begin{bmatrix} \boldsymbol{F}_Q \\ \boldsymbol{F}_\tau \\ 0 \\ 0 \end{bmatrix} \tag{2}
$$

Where $\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{K}$ denote the mass, damping and stiffness matrices, respectively; $\boldsymbol{X}, \dot{\boldsymbol{X}}$, and $\ddot{\boldsymbol{X}}$ denote displacement, velocity and acceleration, respectively; $\boldsymbol{F}$ denotes the load vector; the subscripts "v", "r", "t" and "s" denote the train, rail, rail and tunnel subsystems; "vr" and "rv" indicate the interaction between the car and rail; "rt" and "tr " denote the interaction between rails and track plates; "ts" and "st" denote the interaction between track plates and tunnels; "vv "rr", "tt" and "ss" represent the matrix of each subsystem itself. This model takes full account of the nonlinearity of train operation and wheel-rail contact relationship, as well as the influence of lining assembly on tunnel structure dynamic response, which provides a powerful analysis tool for studying the dynamic response of vehicle-track-tunnel coupling system under earthquake action.

This paper, based on the aforementioned system coupling concept, develops a three-dimensional refined finite element model of vehicle-track-tunnel coupling system, consisting of 7876564 elements. To achieve a trade-off between solving accuracy and seismic response calculation, the dynamic implicit analysis module of ABAQUS is employed, with a maximum calculation step of 0.02s. The global three-dimensional refined finite element model of vehicle-track-tunnel coupling system is illustrated in Figure 2-6.
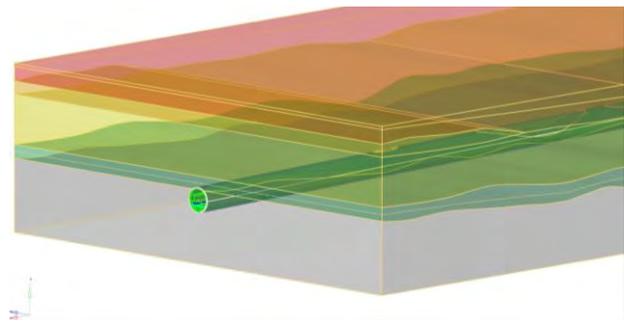
Figure 2-6 3D refined finite element model of the coupled vehicle-track-tunnel system

3. Model validation

3.1. Train-track model validation

In the UIC 519 code, the motion of a free wheel pair on a track can be described by the differential equation:

$$\ddot{y} + \frac{v^2}{er_0} \Delta r = 0 \tag{3}$$

Where y is the wheel pair traverse, $\ddot{y}$ is the acceleration of the wheel pair traverse, e is the contact point span, r_0 is the nominal rolling circle radius, Δr is the wheel diameter difference, and v is the speed of the wheel pair.

Let v be a constant that does not affect the calculation of the equivalent taper, i.e.

$$v = \frac{dx}{dt} \tag{4}$$

Where x is the longitudinal displacement of the wheel pair moving forward along the track. Thus, there is

$$\begin{cases} \dfrac{dy}{dt} = v\dfrac{dy}{dx} \\ \dfrac{d^2y}{dt^2} = v^2\dfrac{d^2y}{dx^2} \end{cases} \tag{5}$$

Bringing equations (4) and (5) into equation (3) gives

$$\frac{d^2y}{dx^2} + \frac{\Delta r}{er_0} = 0 \tag{6}$$

Let the shape of the wheel tread be an angle of γ with a conical shape and

$$\Delta r = 2y\tan\gamma \tag{7}$$

Then the differential equation (3) becomes a constant coefficient second order differential equation:

$$\ddot{y} + \frac{2\tan\gamma}{er_0}y = 0 \tag{8}$$

Take the nominal equivalent taper of 0.038 after matching the tread to the rail profile and make the initial condition of the above equation as follows

$$\begin{cases} y = 0.003 \\ \dfrac{dy}{dx} = 0 \\ x = 0 \end{cases} \tag{9}$$

The variation of wheelset lateral acceleration with respect to wheelset longitudinal displacement can be determined by numerical integration. Figure 3-1 depicts the comparison of wheelset lateral acceleration curves between numerical simulation results and theoretical solutions at the train running speed of 300km/h. Since the theoretical solution neglects the effects of suspension system nonlinearity and wheelset yaw angle, the obtained lateral acceleration curve is a triangular function curve with constant amplitude. In contrast, in the finite element model developed in this paper, the damping of suspension system enhances the lateral stability of train operation, resulting in a lateral acceleration curve that is a quasi-triangular function curve with constant frequency but diminishing amplitude.
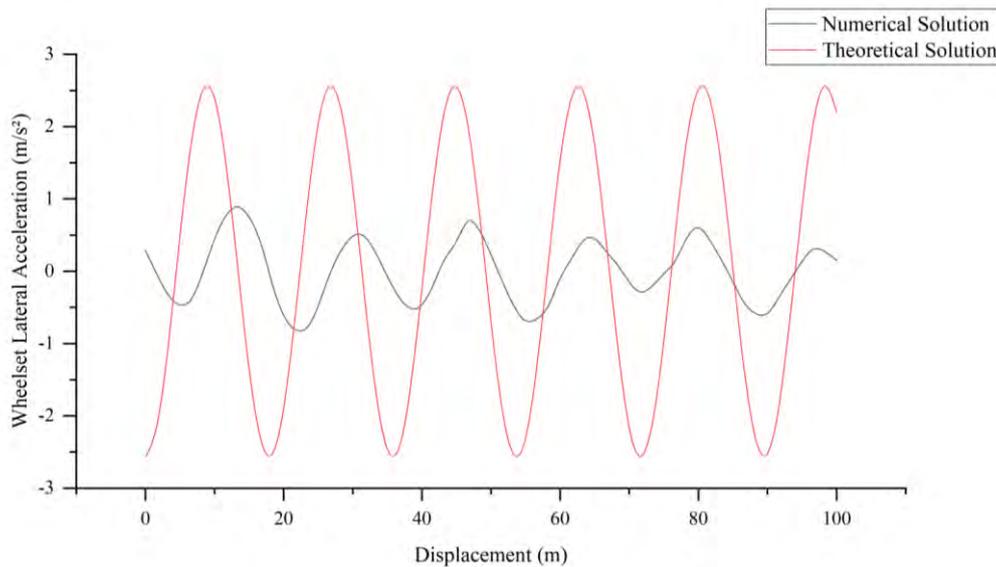
Figure 3-1 Variation of wheel pair lateral acceleration with longitudinal displacement
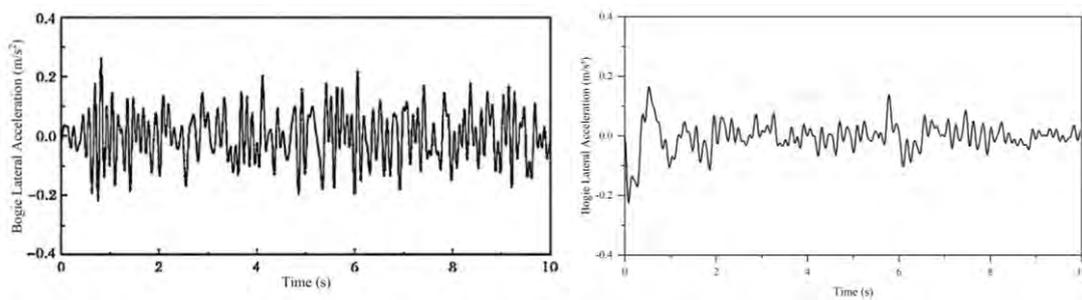
(a) Measured data   (b)Numerical results

Figure 3-2 Comparison of the bogie lateral acceleration between measured data and numerical results

The lateral vibration acceleration of the frame can indicate the lateral motion stability of high-

speed train on the track, hence, the numerical simulation results of this paper are further validated by comparing with the measured data of a type of CRH high-speed EMU[18]. As shown in Figure 3-2, the finite element model developed in this paper can faithfully capture the dynamic features of train, and the train exhibits good motion stability when running at 300km/h, which enhances the reliability of subsequent derailment analysis of high-speed train under seismic excitation.

3.2. Comparison of seismic response of tunnels in complex and homogeneous stratum

3.2.1. Earthquake input

Since the bottom of the model approximately coincides with the bedrock surface, this paper employs the acceleration input method to impose the lateral seismic wave on the bottom of the model, and neglects the effect of non-uniform distribution of seismic wave along longitudinal direction on the dynamic results. The seismic wave is chosen as the Kobe earthquake wave recorded in 1995, which exhibits distinct near-field seismic wave spectrum features. The recording duration of the earthquake wave is 41.99s, the recording time interval is 0.01s, and the peak acceleration is 3.48m/s². To reduce computational cost, only the first 20s waveform is truncated and input during

simulation. The original recorded waveform of Kobe earthquake wave is presented in Figure 3-3.
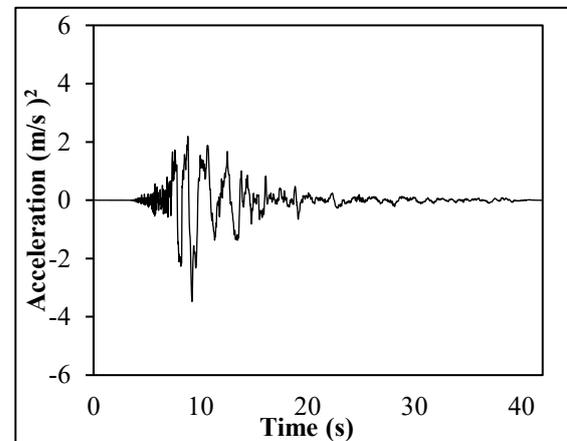
Figure 3-3 Kobe wave acceleration time history diagram

3.2.2. Comparison of seismic response with observation section selection

One of the merits of complex stratum model is that it can acquire observation data from different sections within the same model, and the selection of observation sections can encompass areas with mild soil variation and areas with large longitudinal fluctuation. The final selected observation sections are six in total, and the layout is illustrated in Figure 3-4. Among them, shield tunnel traverses multiple soil layers at S1, S4 and S5 sections, while the soil variation is relatively gentle at S2, S3 and S6 sections.
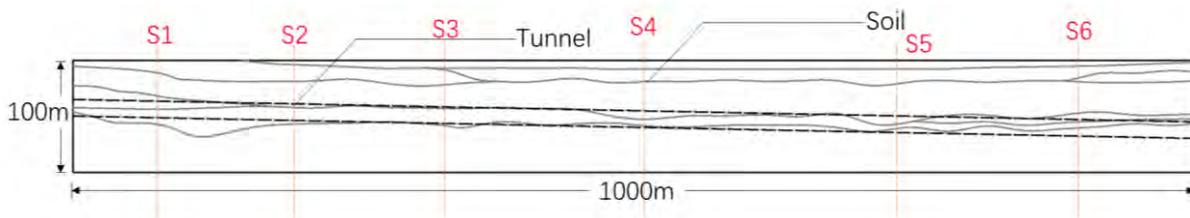
Figure 3-4 Distribution of observation sections

To enhance the persuasiveness of model   comparison, a three-dimensional model

with longitudinal homogeneity is established based on the distribution of surrounding soil at section S1 (Figure 3-5), ensuring the consistency of overall model size, model material and mesh division method with the complex stratum model. The original waveform of Kobe earthquake wave for 20s is imposed from the bottom in acceleration form to obtain the seismic response of the two models.
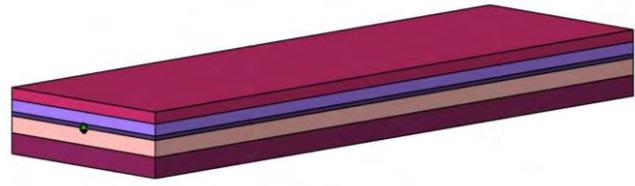
Figure 3-5 Schematic diagram of the longitudinally homogeneous model

According to the numerical simulation results, the soil acceleration response of the two models is relatively large at t=9.5s, thus the lateral acceleration distribution cloud map of the two models at this moment is chosen for comparison. As shown in Figure 3-6, under the premise of identical earthquake wave input, there is a remarkable difference in lateral acceleration distribution between the two models at t=9.5s. In the longitudinally homogeneous model, the lateral acceleration distribution of soil layer exhibits a horizontal stratified pattern, and due to the weak soil properties, there is an evident amplification effect on surface acceleration. In the complex stratum model, owing to the uneven longitudinal distribution of soil layer, the lateral acceleration also displays an uneven pattern related to soil layer geometry at t=9.5s. Meanwhile, the maximum lateral acceleration occurs in the weakest silt soil layer.
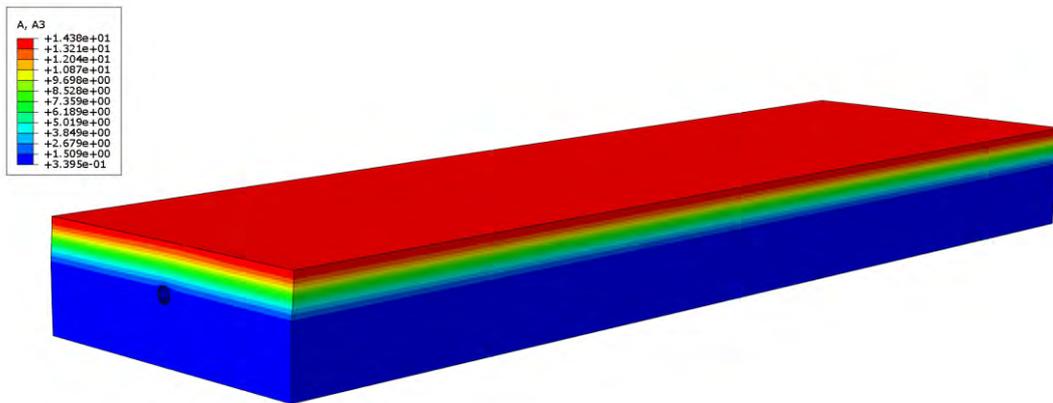
(a)   Cloud map of lateral acceleration distribution for homogeneous stratum model at 9.5s
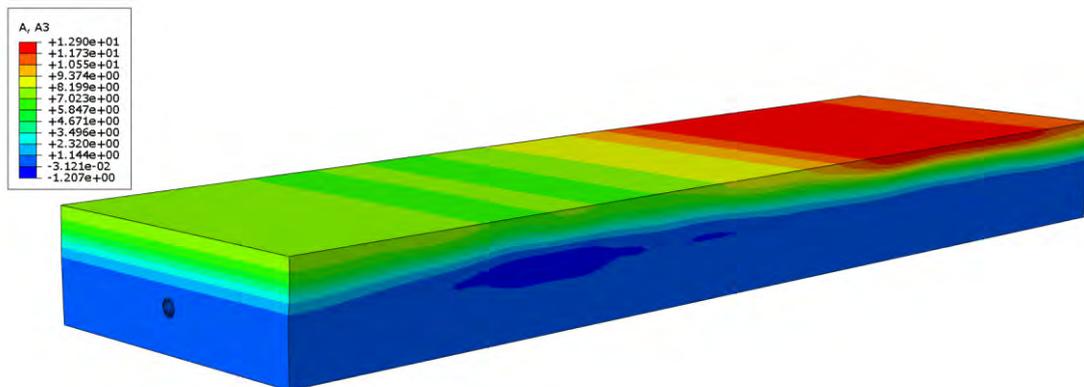
(b)   Cloud map of lateral acceleration distribution for complex stratum model at 9.5s

Figure 3-6 Cloud map of lateral acceleration distribution for complex and homogeneous stratum models

Six sections are selected from the complex stratum model, and the incremental circumferential stress response curve of tunnel lining at its maximum position (located at arch shoulder or arch foot) is plotted as shown in Figure 3-7, and is compared with the longitudinally homogeneous model. On the same section of different models (S1), the peak and fluctuation of incremental stress response curve are basically consistent, but in complex stratum model, the peak stress response value of lining increases by about 8% compared with homogeneous stratum model. Moreover, the stress attenuation is slower in complex stratum model. This indicates that homogeneous stratum model will underestimate tunnel lining seismic response to some extent, and as time goes by, seismic response difference between tunnel in two models gradually enlarges. On different sections of same model (complex stratum model), there is significant difference in stress response. On sections with gentle soil change (S2, S3, S6), circumferential stress peaks of lining are 4.37, 2.61 and 3.80Mpa respectively. On sections with large soil fluctuation (S1, S4, S5), circumferential stress peaks of lining reach 10.40, 9.41 and 9.71Mpa respectively. This indicates that in seismic analysis, influence induced by abrupt change of stratum condition cannot be neglected. Since dynamic response difference of different tunnel sections may affect operation safety of high-speed train passing through, it is necessary to perform complex modeling for surrounding soil layer where vehicle-track-tunnel system is situated in order to accurately reflect dynamic response of vehicle-track-tunnel system under seismic excitation.
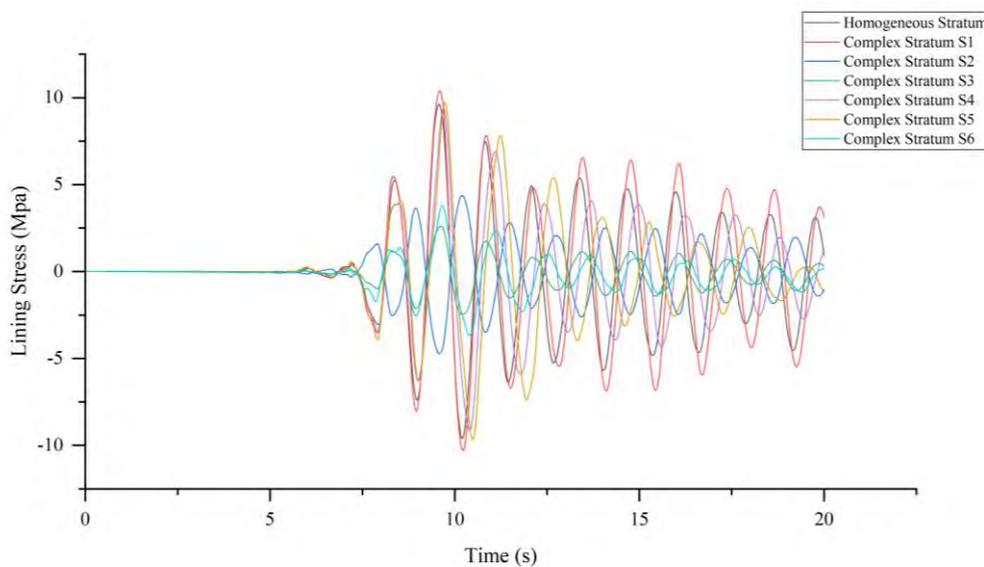
Figure 3-7 Incremental circumferential stress response of tunnel lining in each section

4. Establishment of seismic performance indicator for coupled vehicle-track-tunnel systems

4.1. *Analysis and selection of existing evaluation indicators*

4.1.1. *Train operation safety evaluation indicators*

The assessment of train operation dynamic performance mainly encompasses three aspects: hunting stability, derailment safety and running smoothness, among which, train derailment safety is a vital aspect of train operation safety assessment. The research on derailment problem originated from the renowned derailment coefficient discrimination criterion proposed by French scholar Nadal[30] in 1896, which took the friction force between wheel and rail that can prevent wheel slip as the critical condition of train derailment. He assumed that the lateral force T_y and normal force N between wheel and rail along the transverse direction of contact point satisfy Coulomb friction law, and determined the critical value formula of derailment coefficient:

$$\frac{Q}{P} = \frac{tan\alpha - \mu}{1 + \mu tan\alpha} \tag{10}$$

In the formula Q and P are the wheel-rail contact transverse and vertical forces at the contact point, respectively; α is the contact angle of the wheel edge, and μ is the wheel-rail friction coefficient.

  In addition to derailment coefficient, wheel load reduction rate as a supplementary indicator also gradually draws attention from various countries. After studying and analyzing a large number of freight car derailment accident cases[31]. European, Japanese and North American scholars found that excessive wheel load reduction rate caused by track and vehicle structure parameters, is often the main cause of the above accidents. Japan stipulates in "Railway Structure Design Standard-Explanation of Seismic Design"[32] that the dynamic limit of wheel load reduction rate of new Shinkansen train should not exceed 0.8; European standard EN 14363:2016[33] stipulates that the limit of derailment coefficient is 0.8, and requires that the wheel load reduction rate of train should not exceed 0.6 in curve section. Chinese code "Railway Bridge and Culvert Design Code" (TB 10002-2017)[34] stipulates the limits of derailment coefficient and wheel load vertical reduction rate are 0.8 and 0.6 respectively based on reference to codes of Japan, Europe and other countries. In addition to the above two indicators, there are also limit values for wheel-rail vertical force and wheel-rail lateral force in codes of various countries, but derailment coefficient is still the authorized indicator to measure train derailment safety at present.

With the continuous increasing running speed of passenger train, the assessment system with Nadal derailment criteria as the core also gradually shows drawbacks. Because high-speed train will frequently exhibit jumping track and hunting movement when running, both derailment coefficient and wheel load reduction rate are difficult to accurately describe train's derailment behavior, and their assessment results are often too conservative[35–37]. In view of the defects of the above derailment criteria, many new derailment assessment indicators based on wheel-rail contact relationship and system energy have been proposed successively.

Miyamoto[38] first proposed in 1996 that wheel-rail uplift can be used as an assessment indicator of train derailment, then Zhai[39] proposed a limit value of 27mm for wheel uplift according to the geometric shape of wear-type tread (LM) wheel in China, and pointed out that wheel uplift can be indirectly obtained by measuring action time of derailment coefficient in test. Xiao[36] combined indicators such as derailment coefficient, wheel load reduction rate, wheel uplift and contact point lateral displacement, to construct and analyze derailment safety domain of high-speed train under seismic excitation. Lei[37] proposed a method to assess seismic safety of high-speed train on bridge based on geometric indicators, and divided safety into three levels: safety, danger and derailment according to wheel uplift.

Zeng Qingyuan[40], academician of Central South University and others believe that train derailment is caused by instability of system lateral vibration, first proposed stochastic analysis theory of train derailment energy, solved theoretical problem of train derailment with unknown cause. Guan[41] derived a derailment discrimination criterion determined by peak wheel lateral force, wheel load and collision time according to kinematic relationship and energy conservation law. Sun[42] proposed RMS limit value of wheelset lateral acceleration by analyzing nonlinear dynamic characteristics of high-speed train derailment, which has higher reliability in assessing high-frequency dynamic derailment. Barbosa[43] proposed a three-dimensional derailment discrimination criterion based on wheel-rail force considering wheelset spin and creep effects.

Considering that indicators such as wheel uplift and wheelset lateral displacement have the characteristics of accuracy and easy acquisition, combined with three-dimensional refined finite element model established in this paper, these two indicators are intended to be selected as train derailment indexes. For matching situation between LMA wheel profile and CHN60 rail profile, wheel uplift limit can be calculated from relative geometric position as 28.272mm, corresponding critical wheel lateral displacement limit value is 33.1mm[36].

4.1.2. *Seismic performance indicators for tunnels*

For the operation safety of train in shield tunnel under seismic excitation, Japan proposed[44] in the latest railway standard to adopt tunnel structure displacement and deformation, track surface response and lateral vibration displacement as seismic performance verification indicators for train in tunnel. Although no explicit indicator limit value is set in the book, this idea undoubtedly reflects the continuous advancement of dynamic response research of vehicle-track-tunnel coupling system, and implies a new direction in such field. Although there is no mature high-speed railway tunnel seismic performance indicator at present, seismic performance indicators have been widely applied for general tunnels. Taking displacement type indicator as an example, diameter deformation rate has been adopted as seismic performance indicator of shield tunnel in "Urban Rail Transit Structure Seismic Design Code" (GB50909)[45]. Dong[16] implemented dynamic increment method and obtained limit of diameter deformation rate of shield tunnel in elastic and local elasto-plastic state as 3.40‰ and 6.40‰ respectively. Wang[46] conducted seismic performance analysis of tunnel structure under combined action of load and erosion based on Xiamen Metro Line 2 project, and preliminarily provided diameter deformation rate limit applicable to circular shield tunnel at various performance levels. Japanese scholar Koizumi Atsushi[47] took tunnel inclination

angle as evaluation indicator to measure structure safety, and gave approximate range of tunnel inclination angle at three performance levels.

Displacement type deformation indicator has advantages of clear concept and easy acquisition, but only through deformation indicator can not obtain structure damage evolution process and damage path. Therefore, some scholars carried out research on damage indicator of tunnel seismic resistance. Wang[48] performed wave spectrum analysis on input earthquake wave, and according to energy principle, derived energy response equation under damage state and conducted nonlinear seismic response analysis on tunnel structure based on the equation, discussed weak part and asymptotic failure process of tunnel structure under seismic excitation. Li[49] proposed a reinforced concrete equivalent plastic damage model suitable for damage analysis of tunnels, and gave corresponding damage evaluation indicator. Analysis shows that this constitutive model can accurately reflect the overall mechanical behavior and damage evolution process of reinforced concrete structure. Sun[50] established lining local damage index (LLDI) and lining global damage index (LGDI) according to damage mode of tunnel under Wenchuan earthquake action. The indexes are then used to evaluate influence of single main shock excitation and main shock-aftershock excitation on hydraulic arch tunnel.

Although damage type indicator can more accurately reflect stress state and damage process of tunnel linings, for complex coupling system of vehicle-track-tunnel, extraction and analysis of damage indicator will significantly reduce analysis efficiency. Because the purpose of this study is to establish seismic performance indicator of vehicle-track-tunnel coupling system easy to use in engineering practice, diameter deformation rate as a more mature indicator is intended to be used as deformation analysis indicator of tunnel in coupling model.

4.2. Analysis of the derailment process of high-speed train under seismic excitation

This section takes seismic response of high-speed train running at 300km/h as an example and investigates dynamic derailment process of high-speed train through two derailment safety indicators: wheel uplift and wheelset lateral displacement. The original waveform of 1995 KOBE wave introduced in section 3.2.1 was input to bottom of the model in acceleration form. Figure 4-1 presents the time history curve of wheel uplift (left wheel) and wheelset lateral displacement change of the firstly derailed wheelset (wheelset 2).
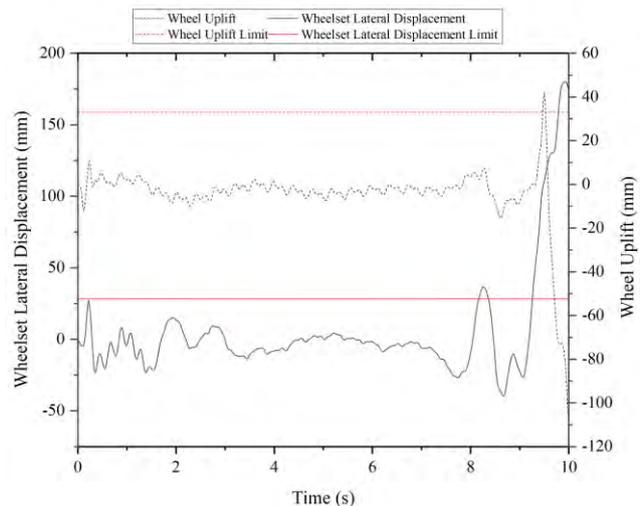
Figure 4-1 Time history diagram of train derailment evaluation indicators
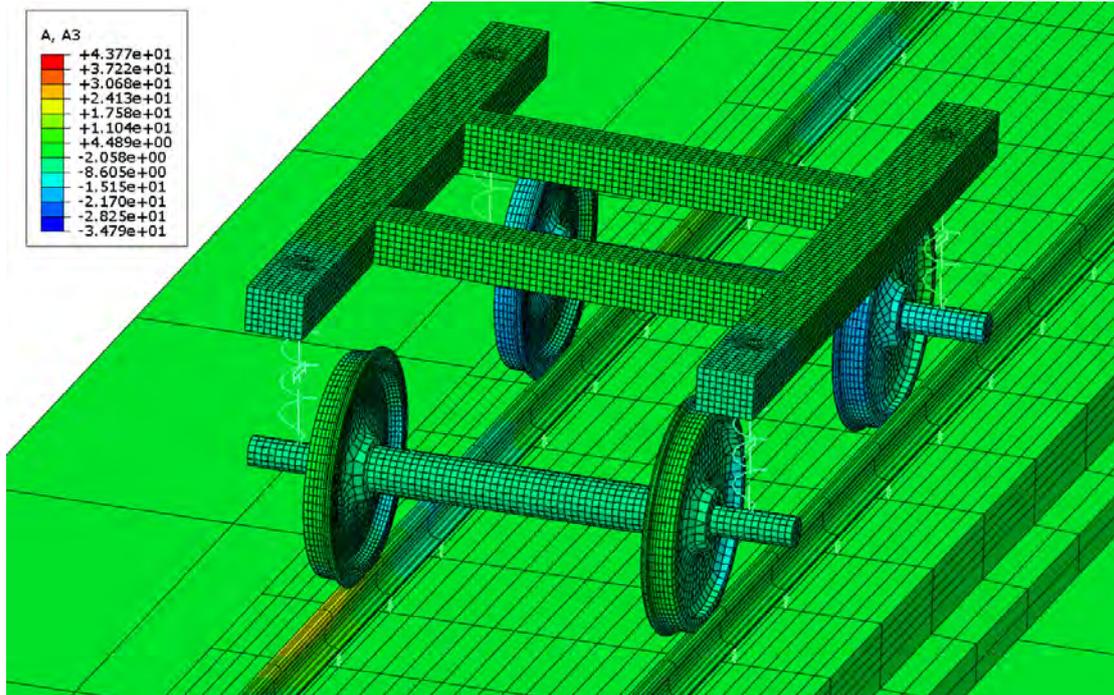
Figure 4-2 Derailment of wheelset 2 at 9.46s

As shown in Figure 4 1, wheelset lateral displacement exceeds limitation for the first time at 8.22s, and to both sides of wheel, the wheelset lateral displacement exceeds 33.1mm, indicating that wheelset 2 exhibits large lateral vibration at this time. Then, wheelset lateral displacement exceeds limitation for the second time and diverges at 9.28s. Wheel uplift only has small fluctuation before 9s, and its amplitude is far lower than limit value level. But after 9s, wheel uplift rapidly increases, finally exceeds the limit value and diverges at 9.46s. It can be revealed from directly observing vehicle operation state (Figure 4-2), that left wheel of wheelset 2 climbs on rail top surface and loses constraint at 9.46s, causing the lateral instability and derailment of the train. By comparing results evaluated by various derailment indicators with the finite element simulation results, it can be known that wheel uplift indicator can more accurately reflect the derailment process of high-speed train under seismic excitation, while wheelset lateral displacement indicator can roughly reflect the lateral instability process of high-speed train. Therefore, subsequent analysis will adopt wheel uplift as dynamic evaluation indicator for the train operation state and adopt wheelset lateral displacement as early warning indicator for derailment of high-speed train.

4.3. Establishment of seismic performance indicators for high-speed railway tunnels

Taking diameter deformation rate as seismic performance evaluation indicator of shield tunnel structure, deformation of six observation sections under seismic excitation is analyzed. By the curve trend in Figure 4-3, it can be known: (1) Tunnel diameter deformation rate response of six sections under seismic excitation is less affected by soil fluctuation, among which S2 section's maximum diameter deformation rate is only 6% higher than that of the S6 section. Diameter deformation rate of all sections has two peaks at 7.26s and 8.2s respectively, and reaches maximum value at 9.32s. (2) At the same time, the lateral displacement of the high-speed train wheelset exceeds the limit for the first time at 8.2s and diverges at 9.32s, which are

consistent with the two peaks of the diameter deformation rate. The wheel lift of the train, which causes derailment, happens slightly later than the maximum value of the diameter deformation rate. Combining above two points, diameter deformation rate indicator of shield tunnel is less affected by soil fluctuation situation, thus has higher applicability for seismic performance evaluation of shield tunnel in soft soil area, and has stronger indication for high-speed train derailment. Therefore, diameter deformation rate can be attempted to be taken as seismic performance indicator for high-speed railway tunnel.
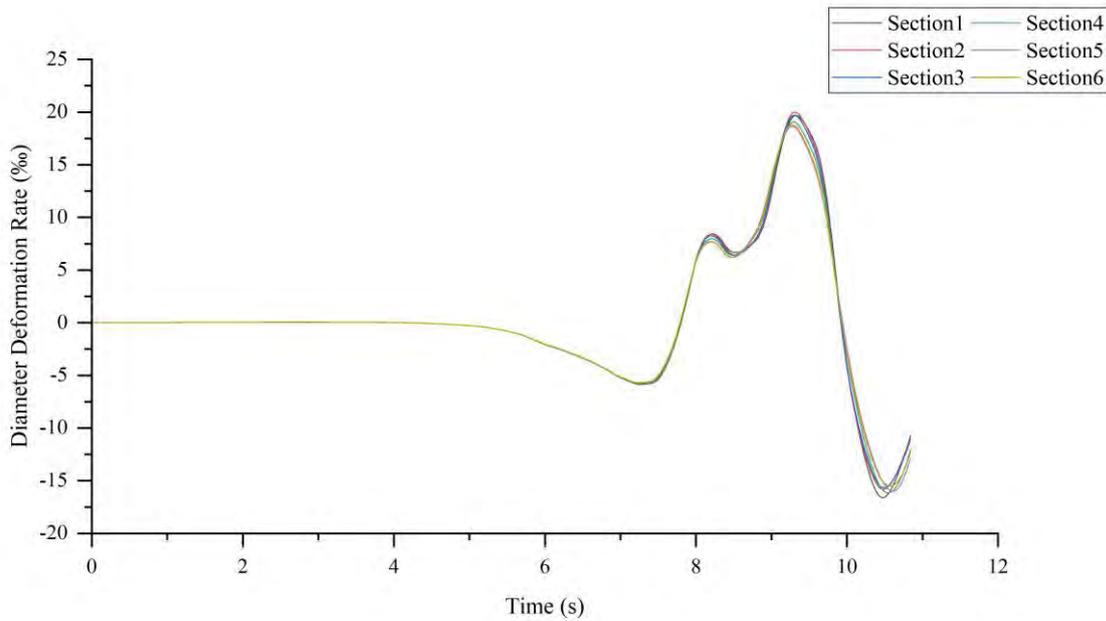
Figure 4-3 Time history diagram of diameter deformation rate for different sections

Based on the aforementioned coupled dynamic analysis of the vehicle-track-tunnel system under seismic excitation, two seismic performance indicators of high-speed rail tunnel with distinct performance levels can be established. Level I is the prognostic indicator of high-speed train derailment. Under this level, the train exhibits considerable lateral oscillation, but remains on track. Level II is the critical indicator of high-speed train derailment. The high-speed train will derail once the tunnel seismic response surpasses this indicator. According to the findings of Dong[16] et al., when the shield tunnel is in the performance state of "potential structural failure", the recommended threshold of its diameter deformation rate is 6.40‰. In this paper, based on this research threshold, the diameter deformation rate of 6.40‰ is adopted as the prognostic indicator of high-speed train derailment in shield tunnel, and the diameter deformation rate of 16‰ is adopted as the critical indicator of high-speed train derailment in accordance with the numerical simulation outcomes of this paper.

Table 4-1 Two-level seismic performance indicator for high-speed railway tunnel

Performance level	Limit values for diameter deformation rate	High-speed train operation state
Level I	6.40‰	Large lateral vibrations on the train, but still not derailed
Level II	16‰	Train about to derail

4.3.1. *Verification of two-level seismic performance indicator for high-speed railway tunnel*

To validate the proposed two-level seismic performance indicator, 1952 Taft wave is input to the established vehicle-track-tunnel coupling FEM model as the ground motion, and the peak ground acceleration is adjusted to 0.2g to acquire more practical results. Since the peak ground acceleration appears at 9.14s, time duration of the numerical simulation is limited to 9.5s to save the calculation cost. The original recorded waveform of Taft earthquake wave is presented in Figure 4-4.
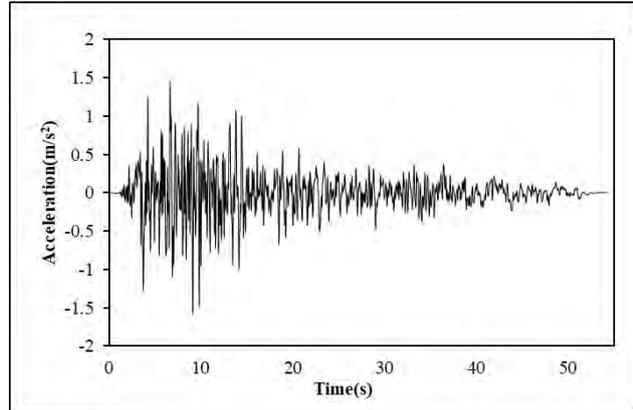
Figure 4-4 Taft wave acceleration time history diagram

As for the verification process, diameter deformation rate on the 6 sections of tunnel is extracted to identify the performance state of tunnel structure, then train high-speed train operation safety is predicted according to the two-level seismic performance indicator. After that, wheel uplift and wheelset lateral displacement of wheelset 2 will be studied to validate the conclusions. Results are shown in Figure 4-6 and Figure 4-5.
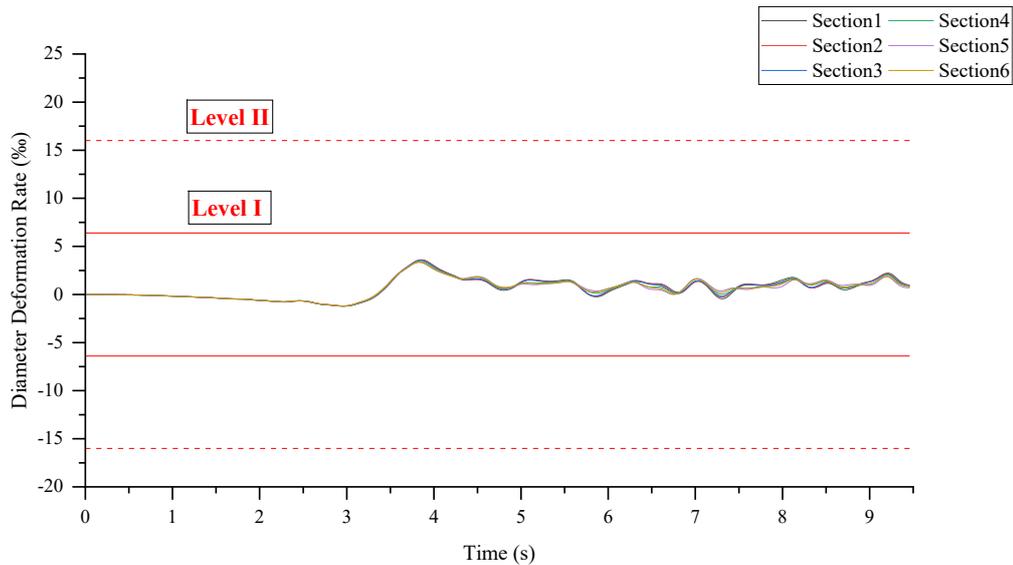
Figure 4-5 Time history diagram of diameter deformation rate for different sections under Taft wave
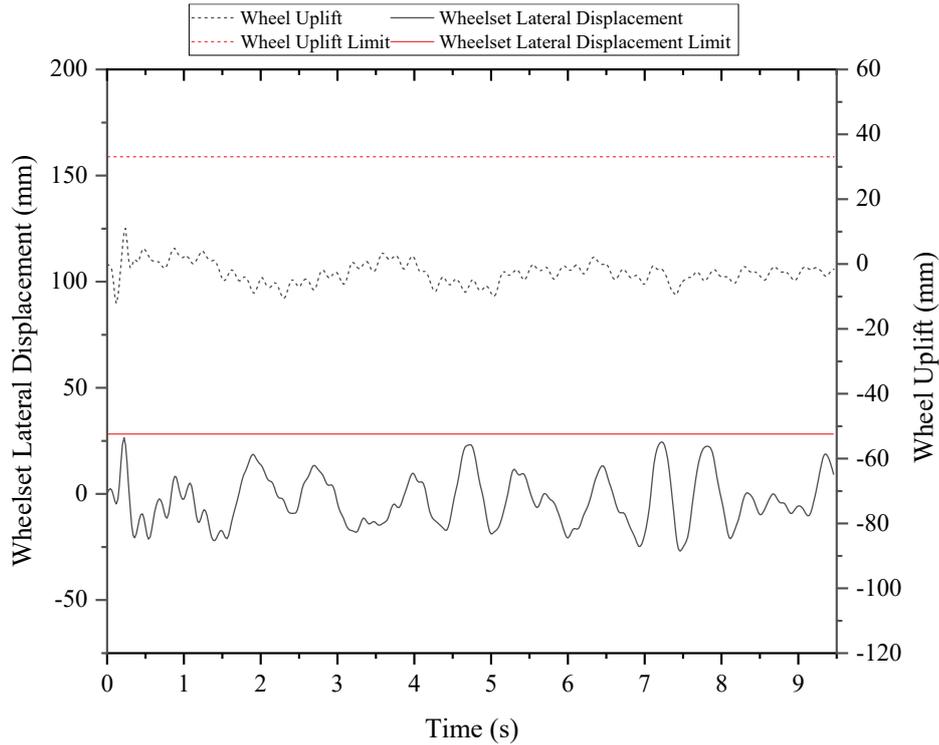
Figure 4-6 Time history diagram of train derailment evaluation indicators under 0.2g Taft wave

From Figure 4-5, it can be shown that the diameter deformation rate of tunnel remains in the range of the Level I limit, indicating the high-speed train should operate under safe condition without the risk of derailment. According to Figure 4-6, both wheelset lateral displacement and wheel uplift haven't exceeded their limit, showing that the high-speed train is running safely in the railway tunnel under the seismic excitation of 0.2g Taft earthquake. The operation state of high-speed train has a high consistency with the established seismic performance indicator, thus well proves the accuracy of the two-level seismic performance indicator for high-speed railway tunnel. However, further refinement should be made through analyses of more engineering cases.

5. Conclusions

This paper develops a three-dimensional refined finite element model of the vehicle-track-tunnel coupled system in complex stratum by employing HyperMesh and ABAQUS joint simulation, investigates the running safety of high-speed train and the dynamic response of shield tunnel under seismic excitation, and eventually proposes two-level seismic performance indicators of high-speed railway tunnel based on the analysis outcomes. The main contributions of this paper are as follows:

(1) This paper develops a three-dimensional refined finite element model of the vehicle-track-tunnel coupled system, considering factors such as vehicle-track dynamics, wheel-rail contact relationship, stratum distribution heterogeneity, etc., which can more precisely simulate the dynamic response of the coupled system under seismic excitation.

(2) This paper validates the reliability of the model by comparing the finite element computation outcomes with theoretical and measured data. Simultaneously, it examines the stress situation of different sections of the tunnel, discloses the significant influence of soil layer heterogeneity on the seismic performance of the tunnel, and indicates the necessity of complex stratum modeling.

(3) Based on the three-dimensional refined finite element model of the vehicle-track-tunnel coupled system, in accordance with the requirements of train operation safety, two-level seismic performance indicator for high-speed railway tunnel is established, which offers a quantitative evaluation approach for seismic design of high-speed railway tunnel, and identifies the most unfavorable working condition by comparative analysis.

This paper provides an efficacious analysis approach and evaluation index for seismic design of high-speed rail tunnel, which is of paramount importance for ensuring high-speed rail operation safety. Meanwhile, there are also some limitations in this paper, such as neglecting the influence of factors such as earthquake wave spectrum characteristics, longitudinal earthquake wave non-uniform distribution on the dynamic response of the coupled. The established seismic performance indicator is also required further experimental and practical verification. These issues need to be further explored and refined in future research.

6. References

[1]   Zhang, W., Li, M., and Ji, Y.P. (2022) Analysis and enlightenment of typical failure characteristics of tunnels caused by the Menyuan M6.9 earthquake in Qinghai Province. *China Earthquake Engineering Journal*. 44 (3), 661–669.

[2]   Yang, N. (2010) Study of Anti-seismic Calculation Methods on Subway Track under Seismic Action, Beijing Jiaotong University, 2010.

[3]   Yang, X., Wang, H.H., and Jin, X.L. (2016) DYNAMIC RESPONSE AND RUNNING SAFETY EVALUATION OF A TUNNEL-TRAIN SYSTEM DURING EARTHQUAKE. *Engineering Mechanics*. 33 (12), 176–185.

[4]   Li, J., Lou, Y., Yang, X., and Jin, X. (2021) Nonlinear seismic analysis of a train-tunnel-soil system and running safety assessment of metro vehicles. *Soil Dynamics and Earthquake Engineering*. 147 106772.

[5]   Zhang, G.B. (2012) The Analysis of the Coupling Vibration of Vehicle and Tunnel, Dalian Jiaotong University, 2012.

[6]   Su, L.J. and Liu, H.Q. (2018) Influence of incident direction of seismic wave on dynamic response of high speed steel tunnel. *Journal of Liaoning Technical University (Natural Science)*. 37 (5), 809–815.

[7]   Fabozzi, S. (2018) Feasibility study of a loss-driven earthquake early warning and rapid response systems for tunnels of the Italian high-speed railway network. *Soil Dynamics and Earthquake Engineering*. 11.

[8]   Zhu, X.Y. and Zhang, Z.Q. (2021) Response Analysis of Urban Rail Transit Tunnel and Ballast Bed Structure under Near Fault Earthquake. *Subgrade Engineering*. (4), 178–181.

[9]   Hu, R.Q. (2014) Analysis on seimic and train vibration of shield tunnel in peat soil, Southwest Jiaotong University, 2014.

[10] Zhang, J., He, C., Geng, P., Wang, W., and Meng, L. (2017) Shaking table tests on longitudinal seismic response of shield tunnel through soft-hard stratum junction. *Chinese Journal of Rock Mechanics and Engineering*. 36 (1), 68–77.

[11] Chen, G.X., Sun, R.R., and Zhao, D.F. (2019) Longitudinal seismic response characteristics of seabed shield

tunnels using submodeling analysis. *Chinese Journal of Geotechnical Engineering*. 41 (11), 1983–1991.

[12] Liang, J.W., Wu, M.T., and Ba, Z.N. (2019) Seismic response analysis of lateral uneven sites with soft-hard connected media. *Chinese Journal of Geotechnical Engineering*. 41 (9), 1599–1608.

[13] Liang, J., Xu, A., Ba, Z., Chen, R., Zhang, W., and Liu, M. (2021) Shaking table test and numerical simulation on ultra-large diameter shield tunnel passing through soft-hard stratum. *Soil Dynamics and Earthquake Engineering*. 147 106790.

[14] Chen, Y.D. (2020) Study on the influence of weak interlayer on tunnel structure under earthquake, Chongqing Jiaotong University, 2020.

[15] Dong, Z.F., Zhu, H.Y., Yan, C., and Zheng, B. (2018) Determination of the width of simplified lateral boundaries in the plane problem of underground structures. *JOURNAL OF VIBRATION AND SHOCK*. 37 (6), 218–224.

[16] Dong, Z.F., Yao, Y.C., and Wang, J.J. (2015) Limiting Value of Diameter Deformation Rate for Shield Tunnels under Earthquake Action. *Urban Mass Transit*. 18 (8), 52–56.

[17] Zhang, P., Zheng, D.J., and Zhang, X.F. (2008) Finite Element Model of Complex Geologic Engineering Based on HyperMesh Software. *Journal of Yangtze River Scientific Research Institute*. (2), 84–86.

[18] Zhai, W. (2020) Vehicle–Track Coupled Dynamics: Theory and Applications. Springer Singapore, Singapore.

[19] Zhao, L. (2016) Study of Dynamic Interaction Between Underground Structure and Soil Subject to Strong Earthquake, Institute of Engineering Mechanics，China EarthquakeAdministration, 2016.

[20] Zhuang, H.Y., Wu, B., and Chen, G.X. (2014) Study on Dynamic Contact Properties of Soil-subway Underground Structure Interaction System. *Journal of Disaster Prevention and Mitigation Engineering*. 34 (6), 678–686.

[21] Luo, J., Wu, C., Liu, X., Mi, D., Zeng, F., and Zeng, Y. (2018) Prediction of soft soil foundation settlement in Guangxi granite area based on fuzzy neural network model. *IOP Conference Series: Earth and Environmental Science*. 108 032034.

[22] Hertz, H. (1882) Ueber die Berührung fester elastischer Körper. *Crll*. 1882 (92), 156–171.

[23] Kalker, J.J. (1979) The computation of three-dimensional rolling contact with dry friction. *International Journal for Numerical Methods in Engineering*. 14 (9), 1293–1307.

[24] Kalker, J.J. (1990) Three-Dimensional Elastic Bodies in Rolling Contact. Springer Netherlands, Dordrecht.

[25] An, B.Y. (2020) NUMERICAL STUDY OF WHEEL-RAIL ROLLING CONTACT, Southwest Jiaotong University, 2020.

[26] Li, Y.Y. (2020) The Research on Wheel Wear Prediction of High-speed Train, Beijing Jiaotong University, 2020.

[27] Iwnicki, S. and Iwnicki, S., Eds. (2006) Handbook of Railway Vehicle Dynamics. 0 ed. CRC Press, .

[28] Luo, R. and Shi, H.L. (2018) Dynamics of Railway Vehicle Systems and Application. Southwest Jiaotong University Press, Chengdu.

[29] Xu, L. and Zhai, W. (2021) Vehicle–track–tunnel dynamic interaction: a finite/infinite element modelling method. *Railway Engineering Science*. 29 (2), 109–126.

[30] Nadal, M.J. (1896) Théorie de la stabilité des locomotives, Part 2: Mouvement de lacet. *Annla Mines*. 10 232–255.

[31] Zhang, H. and Yang, G.Z. (2005) On the derailment and wheel load reduction issues of passenger car bogies. *Rolling Stock*. (6), 10-16+1.

[32] 国土交通省鉄道局 監修；鉄道総合技術研究所 編. (2012) 鉄道構造物等設計標準·同解說: 耐震設計 平成 24 年 9 月. 丸善出版, Tokyo.

[33] (2016) EN 14363 Railway applications - Testing and Simulation for the acceptance of running characteristics of railway vehicles - Running Behaviour and stationary tests. .

[34] (2018) Code for Design on Railway Bridge and Culvert: TB10002-2017.

[35] Shi, H.L., Luo, R., and Zeng, J. (2021) Review on domestic and foreign dynamics evaluation criteria of high-speed train. *Journal of Traffic and Transportation Engineering*. 21 (1), 36–58.

[36] Xiao, X.B. (2013) Study on high-speed train derailment mechanism in severe environment, Southwest Jiaotong University, 2013.

[37] Lei, H.J., Huang, B.K., and Liu, W. (2020) Evaluation method for running safety of train on a high-speed railway bridge based on geometric indexes. *JOURNAL OF VIBRATION AND SHOCK*. 39 (17), 57–63.

[38] Miyamoto, M. (1996) Mechanism of Derailment Phenomena of Railway Vehicles. *QR of RTRI*. 37 (3), 147–155.

[39] Zhai, W.M. and Chen, G. (2001) Method and Criteria for Evaluation of Wheel Derailment Based on Wheel Vertical Rise. *Journal of the China Railway Society*. 23 (2), 17–26.

[40] Zeng, Q.Y., Zeng, J., and Zhou, Z.H. (2006) Analysis Theory of Train Derail ment and Its Application. Central South University Press, Changsha.

[41] Guan, Q.H. and Zeng, J. (2009) Study on derailment induced by lateral impact between wheel and rail. *Journal of Vibration and Shock*. 28 (12), 38-42+201.

[42] Sun, L.X., Yao, J.W., Cheng, L., Li, H.M., and Cui, X. (2020) Critical State EValuation Method for Dynamic Derailment of High Speed Vehicle. *China Railway Science*. 41 (2), 113–122.

[43] Barbosa, R.S. (2009) Safety of a railway wheelset – derailment simulation with increasing lateral force. *Vehicle System Dynamics*. 47 (12), 1493–1510.

[44] 鉄道総合技術研究所 (令和 4 年 5 月) 鉄道構造物等設計標準·同解説（トンネル·シールド編）. 丸善出版, Tokyo.

[45] (2014) GB50909-2014 Code for Seismic Design of Urban Rail Transit Structures.

[46] Wang, S.F. (2017) Research on Seismic Behavior of Deteriorated Shield Tunnel in Chloride Environment, Southwest Jiaotong University, 2017.

[47] 土木学会 (2007) シールドトンネルの耐震検討. 丸善出版, Tokyo.

[48] Wang, Z.Z., Tan, Y.G., and Zhang, Z. (2011) Seismic response analyses of tunnel structures based on damage and energy index. *Journal of Dalian University of Technology*. 51 (6), 861–867.

[49] Li, B.T. (2017) Coupling static-dynamic test method and seismic safety evaluation method of cracked tunnel lining, Southwest Jiaotong University, 2017.

[50] Sun, B., Zhang, S., Deng, M., and Wang, C. (2020) Nonlinear dynamic analysis and damage evaluation of hydraulic arched tunnels under mainshock–aftershock ground motion sequences. *Tunnelling and Underground Space Technology*. 98 103321.

Bilinear constitutive model based theoretical study on interfacial properties of pipeline joints coupling tension and temperature

Hong Yuan[1,2], Jun Han[1], *Yaojie Huang[1], †Lan Zeng[1]

[1] MOE Key Laboratory of Disaster Forecast and Control in Engineering, School of Mechanics and Construction Engineering, Jinan University, China
[2] School of Architectural Engineering, Guangzhou Institute of Science and Technology, Guangzhou 510540, China

*Presenting author: huangyaojie@stu2022.jnu.edu.cn
†Corresponding author: zenglan@jnu.edu.cn

Abstract

In order to better understand and describe the interfacial properties of pipe joints for pipeline structures, this paper theoretically investigates the mechanical properties of adhesively bonded pipeline joint interface under simultaneous action of axial tensile loading and temperature changes based on the bilinear rigid-softening constitutive model. Analytical solutions for interface relative slips, shear stresses, and load-displacement relationship, as well as expressions for ultimate loads and effective bonding lengths, have been derived. Different failure processes including softening or debonding stages caused by the bonding lengths were discussed, demonstrating the full-ranged interfacial failure of pipeline joints and explaining their stress transfer mechanism, interface crack propagation, and ductile behavior. The finite element analysis for some certain cases was also checked to coincide with the solutions. This helps establish a theoretical analysis basis for the application of pipeline joints in actual environments with loading and temperature differences.

Keywords: Pipeline joint; bond-slip model; temperature; interfacial property.

Introduction

Pipe joints are common in pipeline engineering as they are able to provide a connection between components and to extend the length of pipelines. They play an indispensable and important role in piping systems in various fields, such as water conservancy, mechanical, electrical, and aerospace engineering. Traditional joint connection methods include welding, threaded connections, and flanged connections, but they have some drawbacks in terms of extra process requirements, corrosion and leakage issues, and accessory costs. The main disadvantages of welded joints are high labor intensity, high costs, and corrosion at the joint [1]; the most deficiency of threaded connections is uneven stress distribution on the threads, which affects the sealing performance of the connection [2], leading to corrosion and leakage; and those of flanged connections are the large number of process steps and high process requirements [3].

Compared to the above three traditional connection methods, adhesive joints have good performance to avoid the aforementioned drawbacks. Adhesive joints have the characteristics of relatively low weight, ease of manufacturing, and effectively reducing stress concentration [4]. This has high economic benefits in construction, maintenance, and safety. Adhesive joints also have significant improvements such as good water tightness, corrosion resistance, and good fatigue properties [5], making them widely used in ocean engineering equipment such as

platforms and pipelines. Furthermore, the use of adhesive joints can reduce labor intensity because no processing or drilling is required on the original pipelines.

However, the application of adhesive joints still faces many challenges. For example, the mechanical properties of the joint interface are complex, and considering the working environment of the pipeline, moisture and thermal changes can greatly affect the bonding characteristics of the interface, especially the temperature changes [6]. It is generally believed that the fracture energy G_f of the adhesive interface remains unchanged until the glass transition temperature is reached, maintaining its original performance and ultimate failure mode. However, as the temperature increases, fracture energy and brittleness decrease [7]. When the temperature approaches the glass transition temperature of the adhesive, the adhesive will transfer from a glassy state to a highly elastic state, and its fracture strength and elastic modulus will rapidly decrease with elongation increasing [8]-[9]. Plazek and Frund [10] have studied the curing and aging process of epoxy resins and found that the transition temperature of uncured samples increased with curing time, while that of fully cured samples decreased. Buch and Shanahan [11] have conducted modified epoxy resin adhesive to 5,000 hours of thermal aging in a 180-250 degrees Celsius environment and found that the transition temperature decreased sharply. Therefore, the thermal behavior of the adhesive at high temperatures is not only affected by the degree of curing, but also by the temperature environment. Furthermore, when the properties of the adhesive change, the bonding characteristics of the joint interface will be greatly reduced and become more complex. Thus, besides the parametric effect of loading modes, the temperature changes on the bond behavior of pipe joints can not be ignored.

Currently, many researchers have conducted in-depth studies on the performance of pipe joints. They have applied various models and methods to explore the effects of factors such as adhesive thickness and overlap length on the performance of pipe joints. These research results are of great guiding significance for the design and manufacturing of pipe joints optimized for bearing tensile, bending, or the most common torsional loads [12-15], which can help improve their performance and reliability.

In engineering practice, the strength and reliability of bonded joints are crucial, and the bond-slip model is an important tool for designing, evaluating and optimizing the mechanical performance of bonded joints for the whole service life. The failure mode of bonded joints often involves interfacial debonding which shows complex mechanical behaviors such as interfacial fracture and crack propagation. Existing research has shown that the stress-strain relationship of the interface is nonlinear, and cannot be studied via simple linear elastic fracture mechanics methods, which requires a nonlinear adhesive-slip constitutive model to characterize the stress-strain relationship of the interface, that is, the cohesive force model [16]-[18]. The bilinear model is considered one of the most accurate models for expressing the stress-strain relationship of the interface. Therefore, many scholars at home and abroad have studied the bilinear adhesive slip model. Yuan *et al.* [19] have reported an analysis scheme for predicting complete debonding propagation of RC beams reinforced with FRP based on a bilinear adhesive slip model. Wang [20] have proposed a cohesive zone model to analyze the debonding caused by a single crack in the middle. Biscaia *et al.* [21] have analyzed the local adhesive-slip behavior of bonded specimens at high temperatures and proposed a temperature-dependent adhesive-slip model with bilinear behavior based on experiments. Dai *et al.* [22] have developed an adhesive-slip model to accurately estimate the bonding behavior of composite materials bonded to the surface of concrete, which were reinforced with CFRP plates using EBR (external bonding reinforcement) technology. Lu *et*

al. [23] have discussed the problem of the adhesive slip model in FRP reinforced concrete structures, and proposed three new adhesive slip models based on the predictions of mesoscopic scale finite element models, which can accurately predict the bonding strength and strain distribution in the FRP plates. Guo *et al.* [24] have introduced the results of tensile tests on CFRP-steel double-lap bonded joints in the temperature range of -20°C to 60°C, and developed a bilinear local adhesive slip model to describe the relationship between shear stress and interfacial slip at different working temperatures.

However, existing literature has rarely studied the influence of temperature changes on the interface performance of pipe joints, and most studies have not proposed a solution that can predict the entire failure process. Therefore, this paper is based on the mechanical model of the pipe joint subjected to axial tension and the cohesive force model, ensuring its mechanical performance at normal working temperature range, introducing the expansion coefficient of the main pipe and the casing, considering the mechanical performance of the pipe joint interface under the simultaneous action of temperature change and axial tension, and proposing a solution that can predict the entire debonding process. This provides a research idea for studying the effect of temperature load on interfacial damage and also provides a theoretical basis for the application of pipe joints in large temperature difference environments.

Mechanical modeling of pipe joints

2.1 Interface models and assumptions

The main pipe (Pipe 1) and the pipe coupler (Pipe 2) are both thin-walled pipes that are bonded together with a very thin layer of adhesive. Due to symmetry, only the joint on the right side of the pipe is considered. Assuming the distance from the left end of the inner main pipe to the right end of the outer pipe coupler is L, the axial force F conducts on the pipe joint, as shown in Fig. 1.
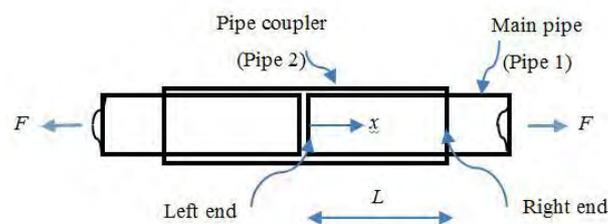
Figure 1. Mechanical model of the pipe joint

In the analysis presented in this paper, the following basic assumptions are made:
(1) The materials of both the main pipe and the pipe coupler are homogeneous and both pipes are in a linear elastic stress-strain state;
(2) The adhesive layer is only subjected to shear stress at its inner and outer surfaces;
(3) The axial external load is shared by the main pipe and the pipe coupler, while the adhesive layer does not bear external loads;
(4) Due to the thin-walled property of both pipes, interfacial layer caused by radial relative displacement is neglected.

2.2 Control Equations

Since the materials of the pipes are in the linear elastic stage, according to the principles of material mechanics, the loads borne by Pipes 1 and 2 are denoted by F_1 and F_2, respectively:

$$F_1 = E_1 A_1 \left(\frac{du_1}{dx} - \alpha_1 \Delta T \right),$$
(1)

$$F_2 = E_2 A_2 \left(\frac{du_2}{dx} - \alpha_2 \Delta T \right).$$
(2)

where E_1、E_2，A_1、A_2，u_1、u_2，α_1、α_2 are the modulus of elasticity, cross-sectional area, axial displacement at the bond interface and coefficient of expansion for Pipes 1 and 2, respectively. The parameter of ΔT is the temperature variation value. As the pipes in the model are thin-walled, the temperature gradient changes in the radial direction are not considered. According to the assumptions, the external loads are borne entirely by the two pipes, so that:

$$F = F_1 + F_2.$$
(3)

In Pipes 1 and 2, shown in Figure 1, the interface displacements are different, so it is necessary to introduce the covariate of axial relative slip at the bonded interface, denoted by δ:

$$\delta = u_1 - u_2.$$
(4)
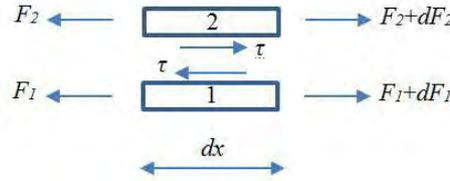
Figure 2. Infinitesimal isolated body

As shown in Fig. 2, a micro-element of length dx is taken for the study and the axial shear stress at the bonded interface is τ. From the equilibrium of the micro-element is obtained that:

$$\tau = \frac{1}{2\pi R} \frac{dF_1}{dx},$$
(5)

where R is the average radius of the pipe joint:

$$R = \frac{1}{2}\left[\left(R_1 + \frac{t_1}{2} \right) + \left(R_2 - \frac{t_2}{2} \right) \right],$$
(6)

where t_1 and t_2 are the thickness of Pipes 1 and 2, respectively, and R_1 and R_2 are the average radii of Pipes 1 and 2, respectively. Substituting Eqs. (1)-(4) into Eq. (5) and introducing the interfacial bond strength and fracture energy, we have:

$$\frac{d^2\delta}{dx^2} - \frac{2G_f}{\tau_f^2} \lambda^2 \tau = 0,$$
(7)

$$F_1 = \frac{\tau_f^2}{2G_f} \frac{2\pi R}{\lambda^2} \left[\frac{d\delta}{dx} + \frac{F}{E_2 A_2} - (\alpha_1 - \alpha_2)\Delta T \right],$$
(8)

where τ_f is the interfacial bond strength, G_f is the interfacial fracture energy and the expression for λ is:

$$\lambda^2 = \frac{\tau_f^2}{2G_f} 2\pi R \left(\frac{1}{E_1 A_1} + \frac{1}{E_2 A_2} \right). \tag{9}$$

2.3 Basic sketch of the bilinear model

In order to gain a better and comprehensive understanding of the interfacial mechanical properties of pipe joints under the combined effect of temperature loading and axial tension, and to explain the mechanism of interfacial crack expansion and peel damage, the bilinear intrinsic model is used. The model in the study only gives the graph and expressions when δ is greater than zero. The complete graph can be obtained by symmetrizing the model according to the origin, as shown in Fig. 3:
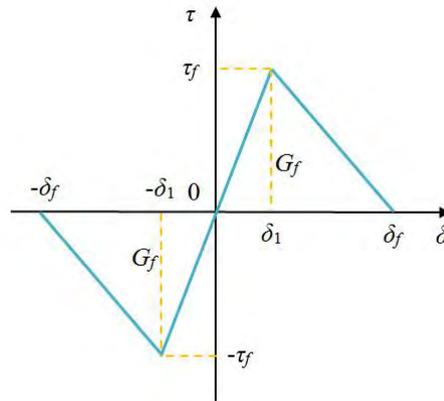
Figure 3. Bilinear constitutive model

As the interfacial slip increases, the interfacial bond shear stress increases linearly up to $|\tau_f|$, corresponding to a linear increase in interfacial slip to $|\delta_1|$, which is the elastic section. As the slip exceeds $|\delta_1|$ and continues to increase, the interface steps into the softening region, and its interfacial shear stress decreases linearly with increasing interfacial slip until the slip increases to $|\delta_f|$ and the shear stress decreases to zero, and the interface enters peeling. For the convenience of the study, it is assumed that the intrinsic model is fully reversible during the interface slip process, i.e. unloading. The mathematical description of the constitutive model is as follows:

$$\tau = f(\delta) = \begin{cases} \dfrac{\tau_f}{\delta_1} \delta & 0 \le |\delta| \le \delta_1 \\[2ex] \dfrac{\tau_f}{\delta_f - \delta_1}(\delta_f - \delta) & \delta_1 < \delta \le \delta_f \\[2ex] \dfrac{\tau_f}{\delta_f - \delta_1}(-\delta_f - \delta) & -\delta_f \le \delta < -\delta_1 \\[2ex] 0 & |\delta| > \delta_f \end{cases} \tag{10}$$

Substituting the mathematical expression (10) for the constitutive model into the control differential Eq. (7) gives the second order differential equation for the amount of slip at the interface:

$$\frac{d^2\delta}{dx^2} - \lambda_1^2\delta = 0 \quad \left(0 \leq |\delta| \leq \delta_1\right), \tag{11a}$$

$$\frac{d^2\delta}{dx^2} + \lambda_2^2\delta = \lambda_2^2\delta_f \quad \left(\delta_1 < \delta \leq \delta_f\right), \tag{11b}$$

$$\frac{d^2\delta}{dx^2} + \lambda_2^2\delta = -\lambda_2^2\delta_f \quad \left(-\delta_f \leq \delta < -\delta_1\right), \tag{11c}$$

$$\frac{d^2\delta}{dx^2} = 0 \quad \left(|\delta| > \delta_f\right), \tag{11d}$$

where:

$$\lambda_2^2 = \frac{\tau_f}{\delta_1}\lambda^2 = \frac{\tau_f}{\delta_1}2\pi R\left(\frac{1}{E_1 A_1} + \frac{1}{E_2 A_2}\right), \tag{12a}$$

$$\lambda_3^2 = \frac{\tau_f}{\delta_f - \delta_1}\lambda^2 = \frac{\tau_f}{\delta_f - \delta_1}2\pi R\left(\frac{1}{E_1 A_1} + \frac{1}{E_2 A_2}\right). \tag{12b}$$

State analysis

3.1 Initial state during temperature loading

Before studying the damage process, the initial state of the interface of the pipe joint when only the temperature influence is applied separately, and then the axial force is superimposed on the initial state. If the pipe joint is only subjected to temperature loading, the interface at both ends of the pipe joint is in the elastic stage, and the relative slip and shear stress distribution at the interface is antisymmetric to the middle of the interface. The governing differential equation for this stage is Eq. (11a), and the boundary conditions when the temperature load is applied are Eqs. (13a) and (13b).

$$F_1(0) = 0, \tag{13a}$$

$$F_1(L) = 0. \tag{13b}$$

Continuous conditions:

$$\delta'(x) \text{ is continuous at } x = a_L, \tag{14a}$$

$$\delta'(x) \text{ is continuous at } x = L - a_R. \tag{14b}$$

If $(\alpha_1-\alpha_2)\Delta T>0$, it means that the expansion effect of the main pipe is more pronounced than the pipe coupler at the same temperature change. Moreover, the axial stress in the main tube is negative due to the restriction of the pipe coupler bonding, i.e. the main pipe is under compression at this time. Based on conditions (13a) and (13b), solving the differential Eq. (11a), the analytical expressions for the relative slip and shear stress distribution at the interface can be obtained as follows:

$$\delta(x) = \frac{1}{\lambda_2}\frac{\cosh(\lambda_2 x) - \cosh\left[\lambda_2(L-x)\right]}{\sinh(\lambda_2 L)}(\alpha_1 - \alpha_2)\Delta T, \tag{15}$$

$$\tau(x) = \frac{\tau_f}{\delta_1} \frac{1}{\lambda_2} \frac{\cosh(\lambda_2 x) - \cosh[\lambda_2(L-x)]}{\sinh(\lambda_2 L)}(\alpha_1 - \alpha_2)\Delta T . \tag{16}$$

If only the effect of the temperature load is considered, i.e. the axial tension F is zero, the softening zone at the left and right ends is antisymmetric along the middle of the bond interface. The relative slip of the interface at the left and right ends can be obtained from Eq. (17):

$$\Delta_L = -\frac{1}{\lambda_2}\tanh\left(\frac{1}{2}\lambda_2 L\right)(\alpha_1 - \alpha_2)\Delta T , \tag{17a}$$

$$\Delta_R = \frac{1}{\lambda_2}\tanh\left(\frac{1}{2}\lambda_2 L\right)(\alpha_1 - \alpha_2)\Delta T . \tag{17b}$$

It can be seen from Eq. (17) that there is no axial tension i.e. when $F=0$, the interface is subject to relative slip, which is different from when the effect of temperature change is not taken into account. If $(\alpha_1-\alpha_2)\Delta T<0$, it means that the expansion effect of the pipe coupler is more pronounced than that of the main pipe at the same temperature change, and that the axial stress in the main tube is positive because the main pipe limits the expansion of the pipe coupler. That is to say, the main pipe is under tension at this point, and Eqs. (15)-(17) are still applicable. Therefore, only the case of $(\alpha_1-\alpha_2)\Delta T>0$ is considered here for the convenience of the study.

As already mentioned, an increase in temperature will lead to an increase in the interfacial bonding properties, at which the glass transition temperature (50°C) is reached, leading to a softening of the bonded interface and a significant change in the properties of the adhesive layer. Consequently, the actual constitutive model will become very complex, which is not the focus of this paper. Therefore, in order to make the study simpler, it is assumed that the bond interface is still in the elastic stage in the initial state caused by the temperature loading here, and it is assumed that the relative slip of the interface caused by the temperature loading is all less than $|\delta_1|$. Letting $\Delta_L=-\delta_1$ and $\Delta_R=\delta_1$, Eq. (17) yields:

$$(\alpha_1 - \alpha_2)\Delta T = \delta_1 \lambda_2 \coth\left(\frac{1}{2}\lambda_2 L\right). \tag{18}$$

At this point, a critical expansion parameter K_3 is defined such that $K_3=\delta_1\lambda_2\coth(\lambda_2 L/2)$. When $0<(\alpha_1-\alpha_2)\Delta T<K_3$, the interface will go through an elastic stage (E stage), an elastic-softening stage (ES stage), an elastic-softening-peeling stage (ESP stage), a softening-elastic-softening-peeling stage (SESP stage) and a softening-peeling stage (SP stage).

3.2 Initial state during temperature loading

3.2.1 Elastic stage
At this point, the superimposed axial tension acting on both ends affects the interface. Moreover, as the load increases, the relative slip at the right end of the pipe joint increases, and the relative slip at the left end of the pipe joint is negative and gradually increases. The interface at this point is still in the elastic stage and the governing equations are the same as in Section 3.1. Due to the superimposed axial tension, the boundary condition (13b) becomes condition (19).

$$F_1(L) = F . \tag{19}$$

Therefore, the analytical expressions for the relative slip and shear stress distribution at the interface are shown as follows,

$$\delta(x) = \left[\frac{F}{E_1 A_1} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2} \frac{\cosh(\lambda_2 x)}{\sinh(\lambda_2 L)}$$
$$- \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2} \frac{\cosh[\lambda_2(L-x)]}{\sinh(\lambda_2 L)}, \tag{20}$$

$$\tau(x) = \frac{\tau_f}{\delta_1} \left\{ \begin{array}{l} \left[\dfrac{F}{E_1 A_1} + (\alpha_1 - \alpha_2)\Delta T \right] \dfrac{1}{\lambda_2} \dfrac{\cosh(\lambda_2 x)}{\sinh(\lambda_2 L)} \\ - \left[-\dfrac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \dfrac{1}{\lambda_2} \dfrac{\cosh[\lambda_2(L-x)]}{\sinh(\lambda_2 L)} \end{array} \right\}. \tag{21}$$

The relative slip of the interface between the left and right ends can be obtained from Eq. (20):

$$\Delta_L = \left[\frac{F}{E_1 A_1} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2} \frac{1}{\sinh(\lambda_2 L)} - \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2} \coth(\lambda_2 L), \tag{22a}$$

$$\Delta_R = \left[\frac{F}{E_1 A_1} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2} \coth(\lambda_2 L) - \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2} \frac{1}{\sinh(\lambda_2 L)}. \tag{22b}$$

Softening of the right end interface occurs when the relative slip at the right end of the pipe joint reaches δ_1 and the shear stress reaches a peak value of τ_f.

3.2.2 Elasticity-softening stage

Define the length of the softening zone at the right end as a_R, which increases as the axial tension increases. The governing equations for this stage are Eqs. (11a) and (11b), with boundary conditions (13a), (19) and:

$$\delta(L - a_R) = \delta_1 \tag{23}$$

The continuity condition is (14b). Based on conditions (13a) and (23), the differential equation (11a) is solved to obtain the analytical expressions for the relative slip and shear stress distribution at the interface in the elastic region of the bonded interface ($0 \leq x \leq L-a_R$) and the expressions are shown as follows:

$$\delta(x) = -\left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2 Y_2} \sinh[\lambda_2(L - a_R - x)] + \frac{\delta_1}{Y_2} \cosh(\lambda_2 x), \tag{24}$$

$$\tau(x) = -\frac{\tau_f}{\delta_1} \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2 Y_2} \sinh[\lambda_2(L - a_R - x)] + \frac{\tau_f}{Y_2} \cosh(\lambda_2 x), \tag{25}$$

where:

$$Y_1 = \sinh[\lambda_2(L - a_R)], \tag{26a}$$

$$Y_2 = \cosh[\lambda_2(L - a_R)]. \tag{26b}$$

Based on conditions (14b) and (23), the differential equation (11b) is solved to obtain the

analytical expressions for the relative slip and shear stress distribution at the interface in the softened region of the bonded interface($L-a_R \leq x \leq L$). The expressions are shown as follow：

$$\delta(x) = \delta_f - \left\{ \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_3} \frac{1}{Y_2} + \delta_1 \frac{\lambda_2}{\lambda_3} \frac{Y_1}{Y_2} \right\} \sin\left[\lambda_3(L-a_R-x)\right]$$
$$-(\delta_f - \delta_1)\cos\left[\lambda_3(L-a_R-x)\right] \tag{27}$$

$$\tau(x) = \frac{\tau_f}{\delta_f - \delta_1} \left\{ \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_3} \frac{1}{Y_2} + \delta_1 \frac{\lambda_2}{\lambda_3} \frac{Y_1}{Y_2} \right\} \sin\left[\lambda_3(L-a_R-x)\right]$$
$$+ \tau_f \cos\left[\lambda_3(L-a_R-x)\right] \tag{28}$$

Based on condition (19), Eq. (27) can be solved to obtain an expression for the external load F by：

$$F = E_2 A_2 \frac{(\alpha_1 - \alpha_2)\Delta T \left[\cos(\lambda_3 a_R) - Y_2\right] + \delta_1 \lambda_2 Y_1 \cos(\lambda_3 a_R) + (\delta_f - \delta_1)\lambda_3 Y_2 \sin(\lambda_3 a_R)}{\rho_2 Y_2 + \cos(\lambda_3 a_R)} \tag{29}$$

The relative slip at the interface between the left and right ends of the pipe joint can be obtained from Eqs. (24) and (27):

$$\Delta_L = -\left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_2 Y_2} \sinh\left[\lambda_2(L-a_R)\right] + \frac{\delta_1}{Y_2}, \tag{30a}$$

$$\Delta_R = \left\{ \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_3} \frac{1}{Y_2} + \delta_1 \frac{\lambda_2}{\lambda_3} \frac{Y_1}{Y_2} \right\} \sin(\lambda_3 a_R) - (\delta_f - \delta_1)\cos(\lambda_3 a_R) + \delta_f \tag{30b}$$

Peeling occurs at the right end when the relative slip of the interface at the right end of the pipe joint reaches δ_f and the shear stress is reduced to zero.

3.2.3 Elastic-softening-peeling stage
Peeling will occur at the interface when the relative slip at the right end of the interface reaches δ_f, defining the length of the peel zone at the right end as d_R. Cracks expand along the bond surface, the peel zone increases and the softening zone moves towards the left end. The governing equations for this stage are Eqs. (11a), (11b) and (11d), with boundary conditions (13a), (19), (31a) and (31b):

$$\delta(L-a_R-d_R) = \delta_1, \tag{31a}$$

$$\delta(L-d_R) = \delta_f. \tag{31b}$$

The continuous condition is (32) as follows:

$$\delta'(x) \text{ is continuous at } x = L-d_R-a_R, \tag{32a}$$

$$\delta'(x) \text{ is continuous at } x = L-d_R. \tag{32b}$$

Based on conditions (32a) and (31a), solving the differential equation (11a) yields the

analytical expressions for the relative slip and shear stress distribution at the interface in the elastic region of the bonded interface ($0 \le x \le L - a_R - d_R$):

$$\delta(x) = -\left(\delta_f - \delta_1\right)\frac{\lambda_3}{\lambda_2}\cot\left(\lambda_3 a_R\right)\sinh\left[\lambda_2\left(L - d_R - a_R - x\right)\right] + \delta_1 \cosh\left[\lambda_2\left(L - d_R - a_R - x\right)\right] \tag{33}$$

$$\tau(x) = -\tau_f \frac{\lambda_2}{\lambda_3}\cot\left(\lambda_3 a_R\right)\sinh\left[\lambda_2\left(L - d_R - a_R - x\right)\right] + \tau_f \cosh\left[\lambda_2\left(L - d_R - a_R - x\right)\right]. \tag{34}$$

Based on conditions (31a) and (31b), solving the differential equation (11b) yields the analytical expressions for the relative slip and shear stress distribution at the interface in the softened region of the bonded interface ($L - a_R - d_R \le x \le L - d_R$):

$$\delta(x) = \delta_f - \left(\delta_f - \delta_1\right)\frac{\sin\left[\lambda_3\left(L - d_R - x\right)\right]}{\sin\left(\lambda_3 a_R\right)}, \tag{35}$$

$$\tau(x) = \tau_f \frac{\sin\left[\lambda_3\left(L - d_R - x\right)\right]}{\sin\left(\lambda_3 a_R\right)}. \tag{36}$$

Based on conditions (31b) and (32b), solving the differential equation (11d) yields an analytical expression for the relative slip of the interface in the stripped region of the bonded interface ($L - d_R \le x \le L$):

$$\delta(x) = \delta_f - \left(\delta_f - \delta_1\right)\lambda_3 \frac{1}{\sin\left(\lambda_3 a_R\right)}\left(L - d_R - x\right). \tag{37}$$

Based on conditions (13a) and (19), Eqs. (33) and (37) can be solved to obtain an expression for the external load F by:

$$F = E_2 A_2\left[-\left(\delta_f - \delta_1\right)\lambda_3 \cot\left(\lambda_3 a_R\right)Y_4 + \delta_1 \lambda_2 Y_3 + \left(\alpha_1 - \alpha_2\right)\Delta T\right], \tag{38a}$$

$$F = E_1 A_1\left[\left(\delta_f - \delta_1\right)\lambda_3 \frac{1}{\sin\left(\lambda_3 a_R\right)} - \left(\alpha_1 - \alpha_2\right)\Delta T\right], \tag{38b}$$

where:

$$Y_3 = \sinh\left[\lambda_2\left(L - d_R - a_R\right)\right], \tag{39a}$$

$$Y_4 = \cosh\left[\lambda_2\left(L - d_R - a_R\right)\right]. \tag{39b}$$

The relation on a_R and d_R can be obtained from Eq. (40):

$$\frac{\lambda_2}{\lambda_3}\left\{\rho_2 \cos\left(\lambda_3 a_R\right)Y_4 + 1\right\} - \rho_2 \sin\left(\lambda_3 a_R\right)Y_3 = \frac{\rho_2 + 1}{\delta_1 \lambda_2}\sin\left(\lambda_3 a_R\right)\left(\alpha_1 - \alpha_2\right)\Delta T \tag{40}$$

The relative slip of the interface at the left and right ends can be obtained from Eqs. (33) and (37), respectively:

$$\Delta_L = -\left(\delta_f - \delta_1\right)\frac{\lambda_3}{\lambda_2}\cot\left(\lambda_3 a_R\right)Y_3 + \delta_1 Y_4 \,, \tag{41a}$$

$$\Delta_R = \delta_f + \left(\delta_f - \delta_1\right)\lambda_3 \frac{1}{\sin\left(\lambda_3 a_R\right)} d_R \,. \tag{41b}$$

When the relative slip of the interface at the left end of the pipe joint reaches δ_1 and the shear stress reaches the peak τ_f, softening of the left end occurs.

3.2.4 Softening-elasticity-softening-peeling stage

When the relative slip of the left end interface reaches δ_1, softening occurs at the left end, and the length of the softening zone at the left end is defined as a_L. The right softening zone and the peeling zone move toward the left end, the elastic zone decreases, and the left softening zone increases. The governing equations for this stage are Eq. (11a), (11b) and (11d). The boundary conditions are (13a), (19), (31a), (31b) and:

$$\delta\left(a_L\right) = \delta_1. \tag{42}$$

The consecutive conditions are Eqs. (14a) and (43):

$$\delta(x) = \delta_f - \delta_f \cos\left[\lambda_1\left(L - d_R - a_R - x\right)\right]. \tag{43}$$

Based on conditions (14a) and (42), solving the differential equation (11b) yields the analytical expressions for the relative slip and shear stress distribution at the interface in the softened region of the bonded interface ($0 \le x \le a_L$):

$$\delta(x) = \delta_f - \left(\delta_f - \delta_1\right)\cos\left[\lambda_3\left(x - a_L\right)\right] - \delta_1 \frac{\lambda_2}{\lambda_3}\frac{Y_6 - 1}{Y_5}\sin\left[\lambda_3\left(x - a_L\right)\right], \tag{44}$$

$$\tau(x) = \tau_f \cos\left[\lambda_3\left(x - a_L\right)\right] + \tau_f \frac{\lambda_3}{\lambda_2}\frac{Y_6 - 1}{Y_5}\sin\left[\lambda_3\left(x - a_L\right)\right], \tag{45}$$

where:

$$Y_5 = \sinh\left[\lambda_2\left(L - d_R - a_L - a_R\right)\right], \tag{46a}$$

$$Y_6 = \cosh\left[\lambda_2\left(L - d_R - a_L - a_R\right)\right]. \tag{46b}$$

Based on conditions (31a) and (42), solving the differential equation (11a) yields the analytical expressions for the relative slip and shear stress distribution at the interface in the elastic region of the bonded interface ($a_L \le x \le L - a_R - d_R$):

$$\delta(x) = \frac{\delta_1}{Y_5}\left\{\sinh\left[\lambda_2\left(L - d_R - a_R - x\right)\right] + \sinh\left[\lambda_2\left(x - a_L\right)\right]\right\}, \tag{47}$$

$$\tau(x) = \frac{\tau_f}{Y_5}\left\{\sinh\left[\lambda_2\left(L - d_R - a_R - x\right)\right] + \sinh\left[\lambda_2\left(x - a_L\right)\right]\right\}. \tag{48}$$

Based on conditions (32a) and (31a), solving the differential equation (11b) yields the analytical expressions for the relative slip and shear stress distribution at the interface in the softened region of the bonded interface ($L - a_R - d_R \le x \le L - d_R$):

$$\delta(x)=\delta_f-\left(\delta_f-\delta_1\right)\cos\left[\lambda_3\left(L-d_R-a_R-x\right)\right]-\delta_1\frac{\lambda_2}{\lambda_3}\frac{Y_6-1}{Y_5}\sin\left[\lambda_3\left(L-d_R-a_R-x\right)\right],\quad(49)$$

$$\tau(x)=\tau_f\cos\left[\lambda_3\left(L-d_R-a_R-x\right)\right]+\tau_f\frac{\lambda_3}{\lambda_2}\frac{Y_6-1}{Y_5}\sin\left[\lambda_3\left(L-d_R-a_R-x\right)\right].\quad(50)$$

Based on conditions (31b) and (32b), solving the differential equation (11d) yields an analytical expression for the relative slip of the interface in the stripped region of the bonded interface ($L\text{-}d_R\leq x\leq L$):

$$\delta(x)=\delta_f-\left\{\left(\delta_f-\delta_1\right)\lambda_3\sin\left(\lambda_2 a_R\right)+\delta_1\lambda_2\frac{Y_6-1}{Y_5}\cos\left(\lambda_3 a_R\right)\right\}\left(L-d_R-x\right).\quad(51)$$

Based on conditions (13a) and (19), Eqs. (44) and (51) can be solved to obtain an expression for the external load F by:

$$F=E_2A_2\frac{\delta_1\lambda_2^2}{\lambda_3}\left[\frac{\lambda_3}{\lambda_2}\frac{Y_6-1}{Y_5}\cos\left(\lambda_3 a_L\right)+\sin\left(\lambda_3 a_L\right)+\frac{\lambda_3}{\delta_1\lambda_2^2}\left(\alpha_1-\alpha_2\right)\Delta T\right],\quad(52a)$$

$$F=E_1A_1\frac{\delta_1\lambda_2^2}{\lambda_3}\left[\frac{\lambda_3}{\lambda_2}\frac{Y_6-1}{Y_5}\cos\left(\lambda_3 a_R\right)+\sin\left(\lambda_3 a_R\right)-\frac{\lambda_3}{\delta_1\lambda_2^2}\left(\alpha_1-\alpha_2\right)\Delta T\right].\quad(52b)$$

The relation on a_L, a_R and d_R can be obtained from the association of Eq. (52):

$$\begin{aligned}&\frac{\lambda_3}{\lambda_2}\frac{Y_6-1}{Y_5}\left[\rho_2\cos\left(\lambda_3 a_L\right)-\cos\left(\lambda_3 a_R\right)\right]+\left[\rho_2\sin\left(\lambda_3 a_L\right)-\sin\left(\lambda_3 a_R\right)\right]\\&=-\left(\rho_2+1\right)\frac{\lambda_3}{\delta_1\lambda_2^2}\left(\alpha_1-\alpha_2\right)\Delta T\end{aligned}.\quad(53)$$

Based on condition (31b), by solving Eq. (49) it is obtained that：

$$\frac{Y_6-1}{Y_5}\sin\left(\lambda_3 a_R\right)=\frac{\lambda_2}{\lambda_3}\cos\left(\lambda_3 a_R\right).\quad(54)$$

The relative slip of the interface at the left and right ends can be obtained from Eqs. (44) and (51), respectively:

$$\Delta_L=\delta_f-\left(\delta_f-\delta_1\right)\cos\left(\lambda_3 a_L\right)+\delta_1\frac{\lambda_2}{\lambda_3}\frac{Y_6-1}{Y_5}\sin\left(\lambda_3 a_L\right),\quad(55a)$$

$$\Delta_R=\delta_f+\left[\left(\delta_f-\delta_1\right)\lambda_3\sin\left(\lambda_3 a_R\right)+\delta_1\lambda_2\frac{Y_6-1}{Y_5}\cos\left(\lambda_3 a_R\right)\right]d_R.\quad(55b)$$

The interface enters the softening-peeling stage when the elastic region of the interface completely disappears and the softening zones on the left and right ends come together. Defining the length as a_u of the entire softening zone of the interface at this point, there are:

$$a_u=\frac{1}{\lambda_3}\arccos\left[-\frac{1}{\rho_2}+\left(\frac{1}{\rho_2}+1\right)\frac{\lambda_3}{\delta_1\lambda_2^2}\left(\alpha_1-\alpha_2\right)\Delta T\right].\quad(56)$$

3.2.5 Softening-peeling stage

When the softening zones on the left and right ends merge and the rigid region of the interface disappears completely, the interface will enter the softening-peeling stage. The controlling differential equations for this stage are Eqs. (11b) and (11d), the boundary conditions are Eqs. (13a), (19) and (31b), the continuity condition is (32b) and the length of the softening section for this stage is defined as a. Based on conditions (13a) and (31b), solving the differential equation (11b) yields the analytical expressions for the relative slip and shear stress distribution at the interface in the softened region of the bonded interface ($0 \leq x \leq L-d_R$):

$$\delta(x) = \delta_f - \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_3} \frac{\sin\left[\lambda_3(L - d_R - x)\right]}{\cos\left[\lambda_3(L - d_R)\right]}, \tag{57}$$

$$\tau(x) = \frac{\tau_f}{\delta_f - \delta_1} \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_3} \frac{\sin\left[\lambda_3(L - d_R - x)\right]}{\cos\left[\lambda_3(L - d_R)\right]}. \tag{58}$$

Since the shear stress in the peeled region is zero, based on conditions (31b) and (32b), solving the differential equation (11d) yields the analytical expression for the relative slip of the interface in the peeled region of the bonded interface ($L-d_R \leq x \leq L$):

$$\delta(x) = \delta_f - \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\cos\left[\lambda_3(L - d_R)\right]}(L - d_R - x). \tag{59}$$

The relative slip of the interface at the left and right ends can be obtained from Eqs. (57) and (59), respectively:

$$\Delta_L = \delta_f - \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\lambda_3} \tan\left[\lambda_3(L - d_R)\right], \tag{60a}$$

$$\Delta_R = \delta_f + \left[-\frac{F}{E_2 A_2} + (\alpha_1 - \alpha_2)\Delta T \right] \frac{1}{\cos\left[\lambda_3(L - d_R)\right]} d_R. \tag{60b}$$

Substituting the boundary condition (23) into Eq. (59), the expression for the length of the softening zone a can be solved for as:

$$a = L - d_R = \frac{1}{\lambda_2} \arccos\left[\frac{-F/(E_2 A_2) + (\alpha_1 - \alpha_2)\Delta T}{F/(E_1 A_1) + (\alpha_1 - \alpha_2)\Delta T} \right]. \tag{61}$$

During the unloading stage, a decreases as the load decreases, indicating that during this stage the peel zone extends towards the left end of the pipe joint and that the interface is completely peeled off when the softening zone decreases to zero. If $\Delta T=0$, i.e. when temperature changes are not taken into account, the length of the softening zone remains constant during the final unloading process.

Numerical simulation

4.1 Parameter selection

The material and geometrical characteristics parameters used in the numerical solution were selected as follows:
$t_1 = 5$ mm, $R_1 = 147.5$ mm, $t_2 = 7$ mm, $R_2 = 154$ mm, $E_1 = 120$ GPa, $E_2 = 200$ GPa.
The expansion coefficients are given as follows:

$\alpha_1 = 2.1 \times 10^{-5}$ ℃, $\alpha_2 = 1.0 \times 10^{-5}$ ℃.

The interface characteristics parameters are given as follows:

$\delta_1 = 0.034$ mm, $\delta_f = 0.16$ mm, $\tau_f = 7.2$ MPa, $G_f = 0.58$ N/mm.

Since the focus of this paper is on the effect of temperature changes on interface damage, the bond length is taken to be a relatively large value:

$L = 600$ mm.

4.2 Simulation results

As the temperature exceeds 50 ℃ , the performance of the bonded layer will change significantly. As with the rigid-softening model, the reference temperature was set at 20 ℃ and a temperature change of 30 ℃ was taken as the study. The load-displacement curve is shown in Fig. 4.
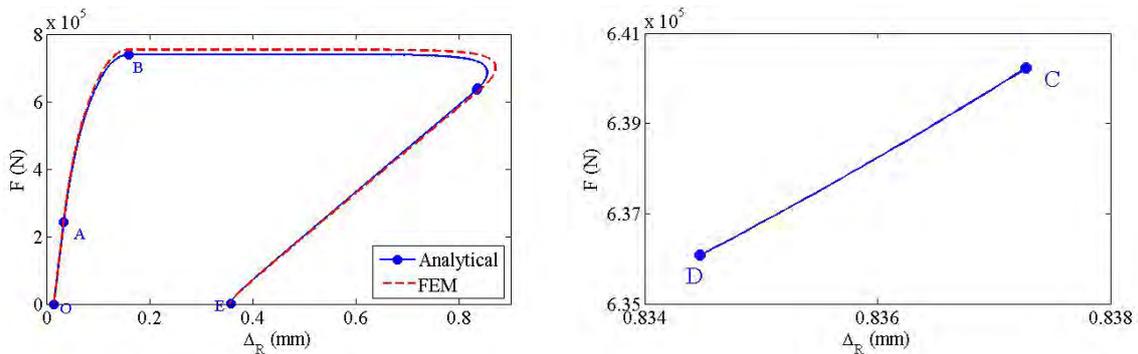
Figure 4. Bilinear model load-displacement curves

Section OA is the elastic stage, AB is the elastic-softening stage, BC is the elastic-softening-peeling stage, CD is the softening-elastic-softening-peeling stage and DE is the softening-peeling stage. When there is only a temperature load and the axial tension is zero, the relative slip of the interface is greater than zero, i.e. the point O is to the right of the origin of the coordinates. When the axial tension is small, the relative slip of the left end interface of the pipe joint is negative, the relative slip of the right end interface of the pipe joint is greater than zero and less than δ_1, and the whole interface is in the elastic stage. With the increase of axial tension, when the relative slip of the right end interface reaches δ_1, the shear stress reaches the peak value τ_f and the right end interface appears to be softened. In the elastic-softening stage, the length of the softening zone at the right end increases, the peak shear stress moves towards the left end, and when the relative slip of the right end interface reaches δ_f, peeling occurs at the right end interface. In the elastic-softening-peeling stage, the length of the peeling zone at the right end increases, and the peak shear stress moves towards the left end. When the relative slip at the left end interface reaches δ_f and the peak shear stress reaches τ_f, the left end interface appears to soften and the bonded interface still has a partially elastic region. In the softening-elastic-softening-peeling stage, the peak shear stresses on both sides approach each other, and when the elastic region disappears completely, the interface enters the softening-peeling stage and unloading begins.

Figure 5 shows the shear stress distribution at the interface. As can be seen from the figure, as the relative slip of the interface increases, the interface gradually goes through the elastic stage, elastic-softening stage, elastic-softening-peeling stage, softening-elastic-softening-peeling stage and softening-peeling stage. In the elastic stage, the shear stresses across the bonded interface are all small, making the shear stress at the left end of the pipe joint negative due to the temperature loading. As the axial tension increases, the interface enters the elastic-

softening stage when the shear stress at the right end of the pipe joint reaches a peak value τ_f. The length of the softening zone on the right side increases and the peak shear stress moves to the left. When the shear stress at the right end of the pipe joint decreases to zero and the interface at the right end peels off, the interface enters the elastic-softening-peeling stage. The length of the peeling zone on the right side increases and the peak shear stress moves to the left side. When the shear stress at the left end of the pipe joint reaches a peak τ_f and the left and right sides are separated by an elastic region in the peak, the interface enters the softening-elastic-softening-peeling stage. As the length of the softening zone on the left side increases, the shear stress peaks on the left and right sides are close to each other. When the elastic area of the interface disappears completely, the shear stress peaks on both sides are combined. As a result, the left and right softening zones are merged into one, and the interface enters the softening-peeling stage. The peak shear stress moves towards the left end of the pipe joint and decreases, the softening zone decreases and the peeling zone increases.

(a) Elastic stage

(b) Elastic-softening stage

(b) Elastic-softening-peeling stage

(d) Softening-elasticity-softening-peeling stage

(e) Softening-peeling stage

(f) Partial details
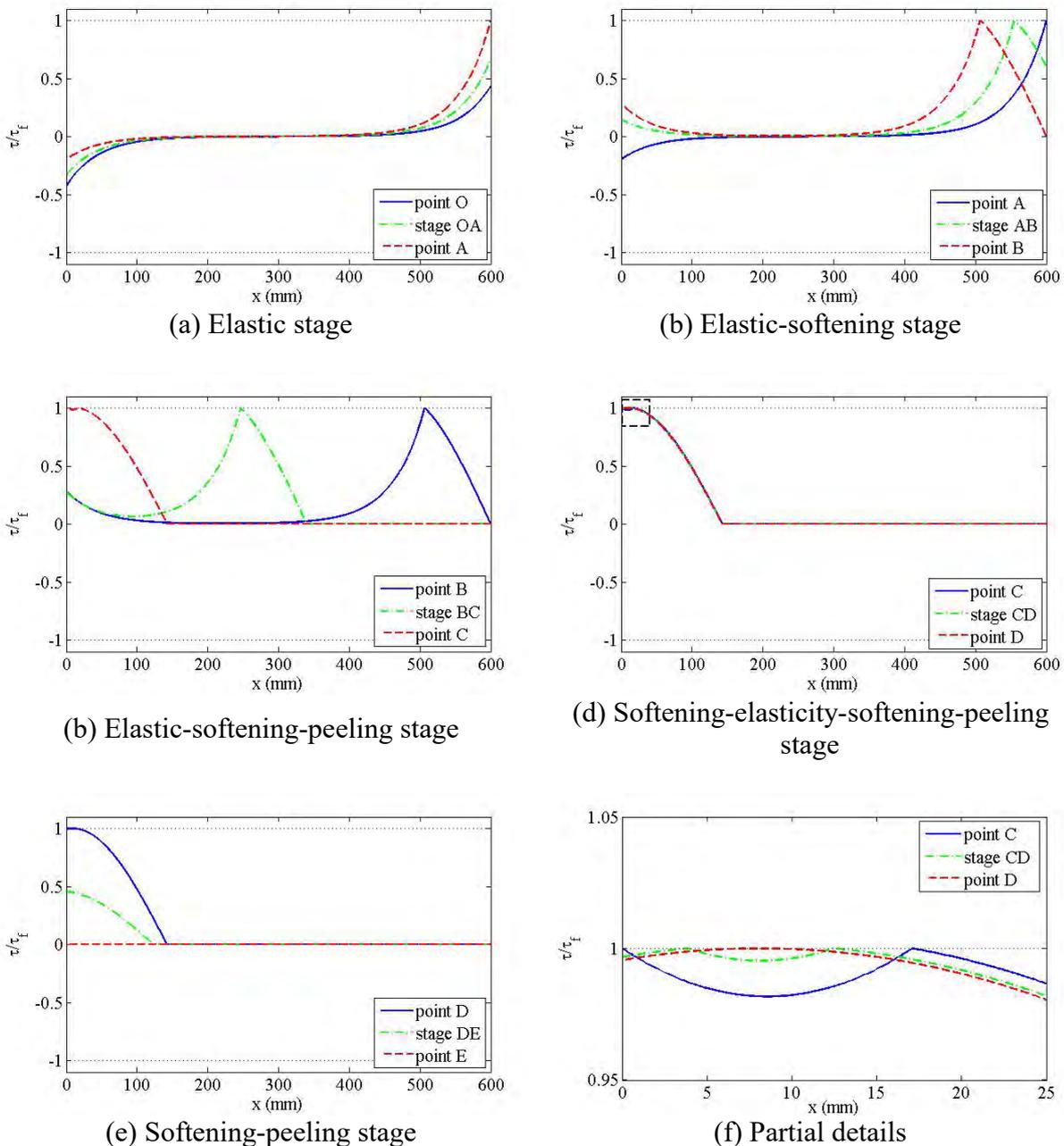
Figure 5. Shear stress distribution at the interface of the bilinear model

As can be seen from Fig. 6, the load-displacement curve can be obtained when the effect of temperature changes is not considered. After considering the effect of temperature changes, the curve starts to the right of the origin, the ultimate load decreases but there is a relative increase in ductility. In the linear unloading stage, the value that eventually returns to the transverse axis is also larger than δ_f. It should be noted that it was emphasized at the beginning that only the case of $(\alpha_1-\alpha_2)\Delta T>0$ is considered, while the opposite is true if $(\alpha_1-\alpha_2)\Delta T<0$. The effect of temperature change also depends on the magnitude of the expansion coefficient of the two pipes and is not determined by the temperature difference alone.
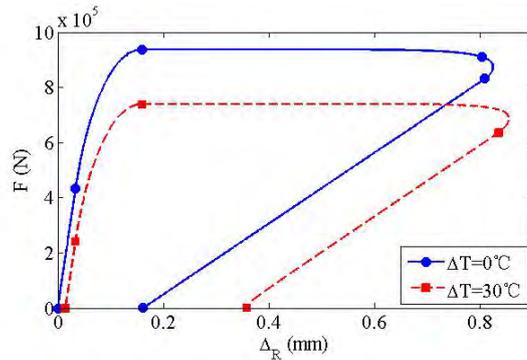
Figure 6. Effect of temperature change on load-displacement curve
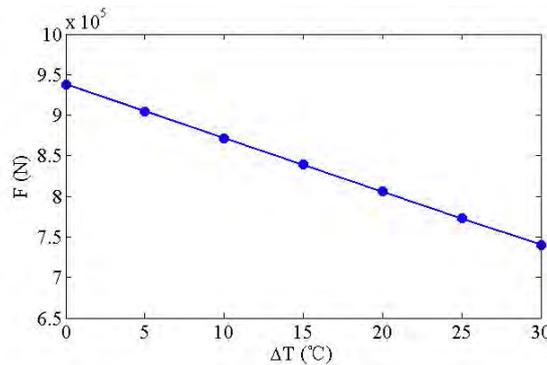
Figure. 7 Effect of temperature change on ultimate load

As can be seen from Fig. 7, the ultimate load is maximum when the effect of temperature change is not taken into account, and decreases linearly with increasing temperature. As with the effect of temperature change on the load-displacement curve, the effect of temperature change depends on the magnitude of the expansion coefficient of the two pipes and is not determined by the temperature difference alone, as it mentioned before.

Summary

Based on the rigid-softening and bilinear instantonal models, the mechanical properties of the interface of the pipe joint are considered when temperature variations and axial tensile forces are simultaneously applied. The analytical solutions for the relative slip, shear stress and load-displacement relationships at the interface, as well as the expressions for the ultimate load and the effective bond length are derived, and the different peel damage processes due to different bond lengths are discussed separately, thus demonstrating the whole peel damage process at the interface of the pipe joint, which is used to explain the stress transfer mechanism, interface crack extension and ductility behavior of the pipe joints. The outcomes can help reasonably describe, design and optimize the interfacial property of adhesively bonded pine joints for pipeline structures in actual engineering during the whole failure process.

Acknowledgements

References

[1] Bordbar, S., Alizadeh, M. and Hashemi, S. H. (2013) Effects of microstructure alteration on corrosion behavior of welded joint in API X70 pipeline steel, *Materials & Design* **45**, 597–604.

[2] Zhou, W., Zhang, R., Ai, S., He, R., Pei, Y. and Fang, D. (2015) Load distribution in threads of porous metal–ceramic functionally graded composite joints subjected to thermomechanical loading, *Composite Structures* **134**, 680–688.

[3] Abid, M. and Nash, D. H. (2006) Structural strength: Gasketed vs non-gasketed flange joint under bolt up and operating condition, *International Journal of Solids and Structures* **43**, 4616–4629.

[4] Xie, Y., Yang, B., Lu, L., Wan, Z. and Liu, X. (2020) Shear strength of bonded joints of carbon fiber reinforced plastic (CFRP) laminates enhanced by a two-step laser surface treatment, *Composite Structures* **232**, 111559.

[5] Zhang, Y., Duan, M., Wang, Y. and Chu, G. (2015) Analytical study of the strength of adhesive joints of riser pipes, *Ships and Offshore Structures* **10**, 545–553.

[6] Viana, G. M. S. O., Costa, M., Banea, M. D. and Da Silva, L. F. M. (2017) A review on the temperature and moisture degradation of adhesive joints, *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications* **231**, 488–501.

[7] Kim, J. K. and Lee, D. G. (2004) Effects of applied pressure and temperature during curing operation on the strength of tubular single-lap adhesive joints, *Journal of Adhesion Science and Technology* **18**, 87–107.

[8] Shimamoto, K., Sekiguchi, Y. and Sato, C. (2016) The critical energy release rate of welded joints between fiber-reinforced thermoplastics and metals when thermal residual stress is considered, *The journal of adhesion* **92**, 306–318.

[9] Marques, E. A. S., Da Silva, L. F., Banea, M. D. and Carbas, R. J. C. (2015) Adhesive joints for low-and high-temperature use: an overview, *The Journal of Adhesion* **91**, 556–585.

[10] Plazek, D. J. and Frund Jr, Z. N. (1990) Epoxy resins (DGEBA): The curing and physical aging process, *Journal of Polymer Science Part B: Polymer Physics* **28**, 431–448.

[11] Buch, X. and Shanahan, M. E. R. (2000) Influence of the gaseous environment on the thermal degradation of a structural epoxy adhesive, *Journal of applied polymer science* **76**, 987–992.

[12] Aimmanee, S., Hongpimolmas, P. and Ruangjirakit, K. (2018) Simplified analytical model for adhesive-bonded tubular joints with isotropic and composite adherends subjected to tension, *International Journal of Adhesion and Adhesives* **86**, 59–72.

[13] Yuan, H., Han, J., Mo, Z. Y. and Zeng, L. (2022) Analysis of full-ranged failure process of bonded pipe Joints under tension, *International Journal of Computational Methods* 2143009.

[14] Yuan, H., Han, J., Zhang, H. L. and Zeng, L. (2022) Debonding analysis of adhesively bonded pipe joints subjected to combined thermal and mechanical loadings, *International Journal of Computational Methods* **19**(6), 2141007.

[15] Pugno, N. and Surace, G. (2001) Tubular bonded joint under torsion: theoretical analysis and optimization for uniform torsional strength, *The Journal of Strain Analysis for Engineering Design* **36**, 17–24.

[16] Campilho, R. D., Banea, M. D., Neto, J. A. B. P. and da Silva, L. F. (2013) Modelling adhesive joints with cohesive zone models: effect of the cohesive law shape of the adhesive layer, *International journal of adhesion and adhesives* **44**, 48–56.

[17] Moghaddas, A., Mostofinejad, D., Saljoughian, A. and Ilia, E. (2021) An empirical FRP-concrete bond-slip model for externally-bonded reinforcement on grooves, *Construction and Building Materials* **281**, 122575.

[18] Zhang, C. (2011) Mechanics and subcritical cracking of FRP-concrete interface, The University of Alabama.

[19] Yuan, H., Teng, J. G., Seracino, R., Wu, Z. S. and Yao, J. (2004) Full-range behavior of FRP-to-concrete bonded joints, *Engineering structures* **26**, 553–565.

[20] Wang, J. (2006) Cohesive zone model of intermediate crack-induced debonding of FRP-plated concrete beam, *International journal of solids and structures* **43**, 6630–6648.

[21] Biscaia, H. C., Chastre, C. and Silva, M. A. (2015). Bond-slip model for FRP-to-concrete bonded joints under external compression. *Composites Part B: Engineering* **80**, 246–259.

[22] Dai, J., Ueda, T. and Sato, Y. (2005) Development of the nonlinear bond stress–slip model of fiber reinforced plastics sheet–concrete interfaces with a simple method, *Journal of composites for construction* **9**, 52–62.

[23] Lu, X. Z., Teng, J. G., Ye, L. P. and Jiang, J. J. (2005) Bond–slip models for FRP sheets/plates bonded to concrete, *Engineering structures* **27**, 920–937.

[24] Guo, D., Liu, Y. L., Gao, W. Y. and Dai, J. G. (2023) Bond behavior of CFRP-to-steel bonded joints at different service temperatures: Experimental study and FE modeling, *Construction and Building Materials* **362**, 129836.

Mathematical model of flow in a one-dimensional open channel with the vertical velocity at the channel bed

†Phuc-Hau Huynh[1], *The Hung Nguyen[2]

[1]The Central College of Transport No V of Danang, Vietnam.
[2]University of Science and Technology, The University of Danang, Vietnam

*Presenting author: profhungthenguyen@gmail.com
†Corresponding author: profhungthenguyen@gmail.com

Abstract

The one-dimensional (1D) mathematical models of open channel flow considered crucial to water resources development and environmental protection have been used popularly in engineering practice. This 1D model is often known as the Saint-Venant equation system and it considers only the main flow velocity along the river with the assumption of a hydrostatic pressure. However, in many cases, there is the effect of river-bed disturbance, and the flow is considered non-hydrostatic pressure. Therefore, in this paper, the authors developed a mathematical model of 1D open channel flow under the influence of the gravitational field, with velocity w* along the gravity acceleration direction at the channel bed. The received governing equations are the more general form of the classic Saint-Venant equation system with the non-hydrostatic pressure.

To solve the derived system of equations, the Taylor-Galerkin finite element method (FEM) with the third order of accuracy was used. The temporal terms were discretized by Taylor expansion series and then the spatial terms were discretized by the Galerkin FEM. In the temporal discretization, the expansion of depth and flow vectors at the time step (n + 1), i.e. $(h, Q)_{n+1}$, is a Taylor series expansion over time t at the right of time t = t_n; next, replacing the spatial terms into this Taylor series was truncated. In the spatial discretization, the second order interpolation function was used. The simulated results based on the proposed numerical scheme showed a good agreement with the experimental data obtained from the physical model built in the National Laboratory for Coastal and River Dynamics in Vietnam.

Keywords: Taylor-Galerkin; finite element method; one-dimensional mathematical models of open channel flow; velocity along gravity acceleration direction at the channel bed, non-hydrostatic pressure; physical model experiment.

Introduction

The 1D modeling of river flow has many applications for water resources development and environmental protection. It has been well established and is commonly known as the Saint-Venant equation based on the simplified assumption that the flow has only velocity along the river axes [2][3][5][8][10]-[12]. However, in reality, there are many problems that this equation is not applied such as flow through the region of water up, obstacles at the river bed [13], etc.
In this paper, the authors take into account which effect of constructing a general mathematical model for the one-dimensional flow under the influence of the gravitational field when the vertical velocity occurs in the river bed.

A general mathematical model of 1D open channel flows

The governing equations used here are the two-dimensional vertical Navier-Stokes equations [4][5][7][8][12][14]. It is integrated over the flow depth and added the boundary conditions in the river bed as vertical velocity $w^*(x,t)$:

$$\frac{\partial u}{\partial t} + u \cdot \frac{\partial u}{\partial x} + w \cdot \frac{\partial u}{\partial z} + \frac{1}{\rho} \cdot \frac{\partial p}{\partial x} - \frac{1}{\rho} \cdot \frac{\partial \tau}{\partial z} = 0 \tag{1a}$$

$$\frac{\partial w}{\partial t} + u \cdot \frac{\partial w}{\partial x} + w \cdot \frac{\partial w}{\partial z} + \frac{1}{\rho} \cdot \frac{\partial p}{\partial z} + g = 0 \tag{1.b}$$

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0 \tag{2}$$
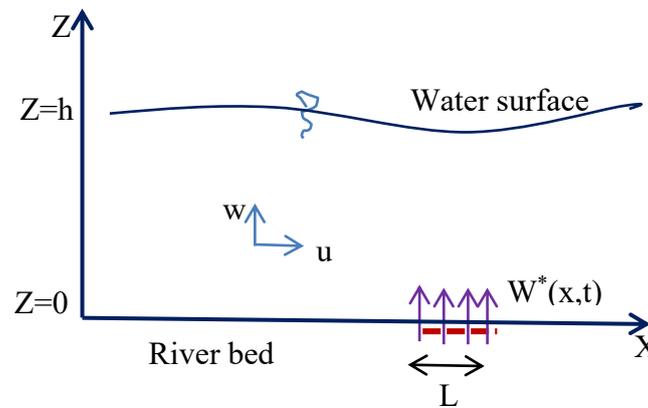
Figure 1. Sketch of the two-dimensional vertical flow problem when the vertical velocity W*(x,t) occurs in the river bed

The vertical axis z is upward, the axis x is horizontal, u is horizontal velocity, w is vertical velocity component, and h is the depth of flow.

Boundary conditions at the water surface: $dh/dt = w_m$ (3)
At the water surface $(z = h)$, the gauge pressure is zero $(p = 0)$ (4)
Boundary conditions at river bed $z = 0$: $w = w^*(x,t)$. (5)
Integrating the continuity equation (2) from the bed to the water surface:

$$\int_o^h \frac{\partial w}{\partial z}.dz = -\int_o^h \frac{\partial u}{\partial x}.dz$$

We receive the vertical velocity at the water surface:

$$w_m = -\int_o^h \frac{\partial u}{\partial x}.dz + w^* \tag{6}$$

Integrating the equation (1.a) from the bed to the water surface, with boundary condition (4) by using Leibnitz's rule [6]:

$$\int_0^h \frac{\partial u}{\partial t}.dz = \frac{\partial}{\partial t}\int_0^h u.dz - u_m \frac{\partial h}{\partial t} \tag{7}$$

   (velocity component in the x direction at the river bed $u = u_0 = 0$; at water surface $u=u_m$)
Applying the integration by part, we have:

$$\int_0^h w.\frac{\partial u}{\partial z}.dz = w_m.u_m - \int_0^h u.\frac{\partial w}{\partial z}.dz$$

Put:
$$J = \int\limits_o^h u.\frac{\partial w}{\partial z}.dz \tag{8}$$

To compute J, based on the continuity equation:
$$\frac{\partial w}{\partial z} = -\frac{\partial u}{\partial x} \tag{9}$$

We have:
$$J = \int\limits_o^h u.\frac{\partial w}{\partial z}.dz = -\int\limits_o^h u.\frac{\partial u}{\partial x}.dz$$

From (6), we have:
$$w_m = -\int\limits_o^h \frac{\partial u}{\partial x}.dz + w^* \tag{10}$$

But, according to Leibnitz's rules [6]:
$$-\int\limits_0^h \frac{\partial u}{\partial x}.dz = -\frac{\partial}{\partial x}\int\limits_0^h u.dz + u_m\frac{\partial h}{\partial x}$$

(velocity component in the x direction at the river bed $u_0 = 0$)
$$w_m = -\frac{\partial}{\partial x}\int\limits_0^h u.dz + u_m\frac{\partial h}{\partial x} + w^*$$

Then, put in (10), we have:
$$\int\limits_0^h w.\frac{\partial u}{\partial z}.dz = w_m.u_m - \int\limits_0^h u.\frac{\partial w}{\partial z}.dz = -u_m\frac{\partial}{\partial x}\int\limits_0^h u.dz +$$
$$+u_m^2.\frac{\partial h}{\partial x} + u_m.w^* + \int\limits_0^h u.\frac{\partial u}{\partial x}.dz$$

As the gauge pressure at water surface is equal to zero; so according to Leibnitz's rule, we receive:
$$\int\limits_0^h \frac{\partial p}{\partial x}.dz = \frac{\partial}{\partial x}\int\limits_0^h p.dz$$

Thus, after integration, equation (1a) is written as follows:
$$\frac{\partial}{\partial t}\int\limits_0^h u.dz - u_m.\frac{\partial h}{\partial t} - u_m\frac{\partial}{\partial x}\int\limits_0^h u.dz + u_m^2.\frac{\partial h}{\partial x} + u_m.w^* + 2\int\limits_0^h u.\frac{\partial u}{\partial x}.dz +$$
$$+\frac{1}{\rho}\frac{\partial}{\partial x}\int\limits_0^h p.dz + \frac{1}{\rho}\tau_b = 0$$

Applying the composite functions and Leibnitz's rule [6], we have:
$$\int\limits_0^h 2u.\frac{\partial u}{\partial x}.dz = \int\limits_0^h .\frac{\partial u^2}{\partial x}.dz = \frac{\partial}{\partial x}\int\limits_0^h u^2 dz - u_m^2.\frac{\partial h}{\partial x}$$

So, equation (1a) is equivalent to:
$$\frac{\partial}{\partial t}\int\limits_0^h u.dz - u_m.\frac{\partial h}{\partial t} - u_m\frac{\partial}{\partial x}\int\limits_0^h u.dz + u_m^2.\frac{\partial h}{\partial x} + u_m.w^* +$$
$$+\frac{\partial}{\partial x}\int\limits_0^h u^2 dz - u_m^2.\frac{\partial h}{\partial x} + \frac{1}{\rho}\frac{\partial}{\partial x}\int\limits_0^h p.dz + \frac{1}{\rho}\tau_b = 0 \tag{11}$$

Notation:
$$< p > = \frac{1}{\rho \cdot h}\int\limits_0^h p.dz \tag{12}$$

Equation (11), after the simplest one, we have:

$$\frac{\partial}{\partial t}\int_0^h u.dz - u_m.\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}\int_0^h u^2 dz - u_m.\frac{\partial}{\partial x}\int_0^h u.dz + u_m.w^* + \frac{\partial}{\partial x}(h.<p>) + \frac{1}{\rho}\tau_b = 0 \qquad (13)$$

Integrating the equation (2) from 0 to z and applying the Leibnitz's rule, we have:

$$\int_0^z \frac{\partial w}{\partial z}.dz = -\int_0^z \frac{\partial u}{\partial x}.dz \quad \rightarrow \quad w = -\frac{\partial}{\partial x}\int_0^z u.dz + u_z.\frac{\partial z}{\partial x} + w^* \qquad (14)$$

Putting (14) into (1.b), we obtain the following equations:

$$\frac{\partial}{\partial t}(-\frac{\partial}{\partial x}\int_0^z u.dz + u_z.\frac{\partial z}{\partial x} + w^*) + u\frac{\partial}{\partial x}(-\frac{\partial}{\partial x}\int_0^z u.dz + u_z.\frac{\partial z}{\partial x} + w^*) +$$

$$(-\frac{\partial}{\partial x}\int_0^z u.dz + u_z.\frac{\partial z}{\partial x} + w^*)*\frac{\partial}{\partial z}(-\frac{\partial}{\partial x}\int_0^z u.dz + u_z.\frac{\partial z}{\partial x} + w^*) + \frac{1}{\rho}\frac{\partial p}{\partial z} + g = 0 \qquad (15)$$

Integrating the equations (15) from z to h, with the boundary condition (4), we receive the expression for the pressure p:

$$\frac{1}{\rho}p = (h-z).(g + \frac{dw^*}{dt}) + \frac{d}{dt}(-\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})) - \frac{d}{dt}(-\frac{\partial}{\partial x}(\overline{U_z}\frac{z^2}{2})) +$$

$$+ \frac{1}{2}(-\frac{\partial}{\partial x}\int_0^h u.dz + u_m\frac{\partial h}{\partial x} + w^*)^2 - \frac{1}{2}(-\frac{\partial}{\partial x}\int_0^z u.dz + u_z\frac{\partial z}{\partial x} + w^*)^2 \qquad (16)$$

and we have :

$$<p> = \frac{1}{\rho \cdot h}\int_0^h p.dz = <p_1> + <p_2> + <p_3> + <p_4> + <p_5> \qquad (17)$$

with:

$$<p_1> = \frac{1}{h}\int_0^h (h-z)\left(g + \frac{dw^*}{dt}\right)dz = \frac{h}{2}(g + \frac{dw^*}{dt})$$

$$<p_2> = \frac{1}{h}\int_0^h \frac{d}{dt}\left(-\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})\right)dz = \frac{1}{h}\frac{d}{dt}\left(-\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})\right)z\Big|_0^h = \frac{d}{dt}[-\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})]$$

$$<p_3> = -\frac{1}{h}\int_0^h \frac{d}{dt}\left(-\frac{\partial}{\partial x}(\overline{U_z}\frac{z^2}{2})\right)dz = -\frac{1}{h}\frac{d}{dt}\left(-\frac{\partial}{\partial x}(\overline{U_z}\frac{z^3}{6})\right)\Big|_0^h = \frac{1}{h}.\frac{d}{dt}[\frac{\partial}{\partial x}(\overline{U_z}\frac{h^3}{6})]$$

$$<p_4> = \frac{1}{2h}[-\frac{\partial}{\partial x}(\overline{U_h}.h) + u_m\frac{\partial h}{\partial x} + w^*]^2 h = \frac{1}{2}[-\frac{\partial}{\partial x}(\overline{U_h}.h) + u_m\frac{\partial h}{\partial x} + w^*]^2$$

$$<p_5> = -\frac{1}{2h}\int_0^h f(z).dz = -\frac{1}{2h}.[f(z)\cdot z]\Big|_0^h + \frac{1}{2h}\int_0^h z.\frac{\partial f(z)}{\partial z}.dz$$

and:

$$f(z) = [-\frac{\partial}{\partial x}\int_0^z u.dz + u_z.\frac{\partial z}{\partial x} + w^*]^2$$

with:

$$\overline{U_z} = \frac{1}{z}\int_0^z u.dz$$

Put:

$$<p_5> = J_1 + J_2$$

$$J_1 = -\frac{1}{2h}.[f(z)\cdot z]\Big|_0^h = -\frac{1}{2}[-\frac{\partial}{\partial x}(\overline{U_z}.h) + u_z.\frac{\partial h}{\partial x} + w^*]^2$$

$$J_2 = \frac{1}{2h}\int_0^h z.\frac{\partial f(z)}{\partial z}.dz = \frac{1}{2h}\int_0^h 2z[-\frac{\partial}{\partial x}(\overline{U_z}\cdot z) + u_z.\frac{\partial z}{\partial x} + w^*].\left[-\frac{\partial^2}{\partial x\partial z}(\overline{U_z}\cdot z) + u_z.\frac{\partial^2 z}{\partial x\partial z}\right]dz$$

$$J_2 = \frac{1}{h}\int_0^h z \frac{\partial}{\partial x}\left(\overline{U_z}\cdot z\right)\frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz - \frac{1}{h}\int_0^h z\cdot u_z \frac{\partial}{\partial x}\left(\overline{U_z}\cdot z\right)\frac{\partial^2 z}{\partial x \partial z}dz - \frac{1}{h}\int_0^h z u_z\cdot \frac{\partial z}{\partial x}\frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz +$$

$$+ \frac{1}{h}\int_0^h z\cdot u_z^2\cdot \frac{\partial z}{\partial x}\cdot \frac{\partial^2 z}{\partial x \partial z}dz - \frac{1}{h}\int_0^h z\cdot w^* \frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz + \frac{1}{h}\int_0^h z\cdot w^* u_z\cdot \frac{\partial^2 z}{\partial x \partial z}dz$$

Put:
$$J_{2A} = -\frac{1}{h}\int_0^h z\cdot w^* \frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz + \frac{1}{h}\int_0^h z\cdot w^* u_z\cdot \frac{\partial^2 z}{\partial x \partial z}dz$$

$$J_{2A} = -\frac{w^*}{h}\frac{\partial}{\partial x}\left(\overline{U_z}\cdot \frac{z^2}{2}\right)\Big|_0^h + \frac{w^* u_z}{h}\frac{\partial\left(z^2/2\right)}{\partial x}\Big|_0^h = -\frac{w^*}{h}\frac{\partial}{\partial x}\left(\overline{U_z}\cdot \frac{h^2}{2}\right) + \frac{w^* u_z}{h}\frac{\partial\left(h^2/2\right)}{\partial x}$$

So:
$$J_2 = \frac{1}{h}\int_0^h z \frac{\partial}{\partial x}\left(\overline{U_z}\cdot z\right)\frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz - \frac{1}{h}\int_0^h z\cdot u_z \frac{\partial}{\partial x}\left(\overline{U_z}\cdot z\right)\frac{\partial^2 z}{\partial x \partial z}dz - \frac{1}{h}\int_0^h z u_z\cdot \frac{\partial z}{\partial x}\frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz +$$

$$\frac{1}{h}\int_0^h z\cdot u_z^2\cdot \frac{\partial z}{\partial x}\cdot \frac{\partial^2 z}{\partial x \partial z}dz - \frac{w^*}{h}\frac{\partial}{\partial x}\left(\overline{U_z}\cdot \frac{h^2}{2}\right) + \frac{w^* u_z}{h}\frac{\partial\left(h^2/2\right)}{\partial x}$$

$$<p_5> = J_1 + J_2 = -\frac{1}{2}\left[-\frac{\partial}{\partial x}\left(\overline{U_z}.h\right) + u_z\frac{\partial h}{\partial x} + w^*\right]^2 + \frac{1}{h}\int_0^h z \frac{\partial}{\partial x}\left(\overline{U_z}\cdot z\right)\frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz -$$

$$\frac{1}{h}\int_0^h z\cdot u_z \frac{\partial}{\partial x}\left(\overline{U_z}\cdot z\right)\frac{\partial^2 z}{\partial x \partial z}dz - \frac{1}{h}\int_0^h z u_z\cdot \frac{\partial z}{\partial x}\frac{\partial^2}{\partial x \partial z}\left(\overline{U_z}\cdot z\right)dz + \frac{1}{h}\int_0^h z\cdot u_z^2\cdot \frac{\partial z}{\partial x}\cdot \frac{\partial^2 z}{\partial x \partial z}dz -$$

$$\frac{w^*}{h}\frac{\partial}{\partial x}\left(\overline{U_z}\cdot \frac{h^2}{2}\right) + \frac{w^* u_z}{h}\frac{\partial\left(h^2/2\right)}{\partial x}$$

Replacing $<p_i>$ into (17), we get:

$$<p> = \frac{h}{2}\left(g + \frac{dw^*}{dt}\right) + \frac{d}{dt}\left[-\frac{\partial}{\partial x}\left(\overline{U_z}\frac{h^2}{2}\right)\right] + \frac{1}{h}\cdot\frac{d}{dt}\left[\frac{\partial}{\partial x}\left(\overline{U_z}\frac{h^3}{6}\right)\right] +$$

$$+ \frac{1}{2}\left[-\frac{\partial}{\partial x}\left(\overline{U_h}.h\right) + u_m\frac{\partial h}{\partial x} + w^*\right]^2 - \frac{1}{2}\left[-\frac{\partial}{\partial x}\left(\overline{U_z}.h\right) + u_z\frac{\partial h}{\partial x} + w^*\right]^2 +$$

$$+ \frac{1}{h}\int_0^h\left[z \frac{\partial}{\partial x}\left(\overline{U_z}.z\right).\frac{\partial^2}{\partial x.\partial z}\left(\overline{U_z}.z\right)\right]dz - \frac{1}{h}\int_0^h\left[z.u_z\cdot \frac{\partial}{\partial x}\left(\overline{U_z}.z\right).\frac{\partial^2 z}{\partial x.\partial z}\right]dz -$$

$$- \frac{1}{h}\int_0^h\left[z.u_z\left(\frac{\partial z}{\partial x}\right).\left(\frac{\partial^2}{\partial x.\partial z}\left(\overline{U_z}.z\right)\right)\right]dz + \frac{1}{h}\int_0^h\left[z.u_z^2\cdot\left(\frac{\partial z}{\partial x}\right).\left(\frac{\partial^2 z}{\partial x.\partial z}\right)\right]dz -$$

$$- \frac{w^*}{h}\cdot\frac{\partial}{\partial x}\left(\overline{U_z}\frac{h^2}{2}\right) + \frac{w^*.u_z}{h}\cdot\frac{\partial(h^2/2)}{\partial x} \qquad (18)$$

Replacing (18) with (13), we have:

$$\frac{\partial}{\partial t}\int_0^h u.dz + \frac{\partial}{\partial x}\int_0^h u^2 dz - u_m\cdot\frac{\partial}{\partial x}\int_0^h u.dz - u_m\cdot\frac{\partial h}{\partial t} + u_m.w^* +$$

$$+ \frac{\partial}{\partial x}\left(\frac{h^2}{2}\left(g + \frac{dw^*}{dt}\right)\right) + \frac{\partial}{\partial x}\left(h\frac{d}{dt}\left[-\frac{\partial}{\partial x}\left(\overline{U_z}\frac{h^2}{2}\right)\right]\right) + \frac{\partial}{\partial x}\left(\frac{d}{dt}\left[\frac{\partial}{\partial x}\left(\overline{U_z}\frac{h^3}{6}\right)\right]\right) +$$

$$+ \frac{1}{2}\frac{\partial}{\partial x}\left(h\left[-\frac{\partial}{\partial x}\left(\overline{U_h}.h\right) + u_m\frac{\partial h}{\partial x} + w^*\right]^2\right) - \frac{1}{2}\frac{\partial}{\partial x}\left(h\left[-\frac{\partial}{\partial x}\left(\overline{U_z}.h\right) + U_z\frac{\partial h}{\partial x} + w^*\right]^2\right) +$$

$$+ \frac{\partial}{\partial x}\left(\int_0^h\left[z\frac{\partial}{\partial x}\left(\overline{U_z}.z\right).\frac{\partial^2}{\partial x.\partial z}\left(\overline{U_z}.z\right]\right) - \frac{\partial}{\partial x}\left(\int_0^h\left[z.u_z\cdot\frac{\partial}{\partial x}\left(\overline{U_z}.z\right).\frac{\partial^2 z}{\partial x.\partial z}\right]dz\right) -$$

$$-\frac{\partial}{\partial x}\left(\int_0^h [z.u_z(\frac{\partial z}{\partial x}).(\frac{\partial^2}{\partial x.\partial z}(\overline{U_z}.z))]dz\right)+\frac{\partial}{\partial x}\left(\int_0^h [z.u_z^2.(\frac{\partial z}{\partial x}).(\frac{\partial^2 z}{\partial x.\partial z})]dz\right)-$$

$$-\frac{\partial}{\partial x}\left(w^*.\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})\right)+\frac{\partial}{\partial x}\left(w^*.u_z.\frac{\partial(h^2/2)}{\partial x}\right)+\frac{1}{\rho}\tau_b = 0 \qquad (19)$$

Assuming the dw/dz >> dw/dx, equation (19) is equivalent to:

$$\frac{\partial}{\partial t}\int_0^h u.dz +\frac{\partial}{\partial x}\int_0^h u^2 dz - u_m.\frac{\partial}{\partial x}\int_0^h u.dz - u_m.\frac{\partial h}{\partial t}+u_m.w^* +\frac{\partial}{\partial x}\frac{h^2}{2}g +\frac{\partial}{\partial x}\left(\frac{h^2}{2}\frac{\partial w^*}{\partial t}\right)+$$

$$+\frac{\partial}{\partial x}(\frac{h^2}{2}u\frac{\partial w^*}{\partial x})+\frac{\partial}{\partial x}\left(h\frac{\partial}{\partial t}[-\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})]\right)+\frac{\partial}{\partial x}\left(hu\frac{\partial}{\partial x}[-\frac{\partial}{\partial x}(\overline{U_z}\frac{h^2}{2})]\right)+$$

$$+\frac{\partial}{\partial x}\left(\frac{\partial}{\partial t}[\frac{\partial}{\partial x}(\overline{U_z}\frac{h^3}{6})]\right)+\frac{\partial}{\partial x}\left(u\frac{\partial}{\partial x}[\frac{\partial}{\partial x}(\overline{U_z}\frac{h^3}{6})]\right)+\frac{\partial}{\partial x}\left(\frac{h}{2}[\frac{\partial}{\partial x}(\overline{U_h}.h)]^2\right)+\frac{\partial}{\partial x}\left(\frac{h}{2}[u_m\frac{\partial h}{\partial x}]^2\right)$$

$$+\frac{\partial}{\partial x}\left(\frac{h}{2}w^{*2}\right)+\frac{\partial}{\partial x}\left(h[-\frac{\partial}{\partial x}(\overline{U_h}.h)][u_m\frac{\partial h}{\partial x}]\right)+\frac{\partial}{\partial x}\left(hw^*[-\frac{\partial}{\partial x}(\overline{U_h}.h)]\right)+\frac{\partial}{\partial x}\left(hw^*[u_m\frac{\partial h}{\partial x}]\right)$$

$$-\frac{\partial}{\partial x}\left(\frac{h}{2}[\frac{\partial}{\partial x}(\overline{U_z}.h)]^2\right)-\frac{\partial}{\partial x}\left(\frac{h}{2}[u_z\frac{\partial h}{\partial x}]^2\right)-\frac{\partial}{\partial x}\left(\frac{h}{2}[w^*]^2\right)-\frac{\partial}{\partial x}\left(h[-\frac{\partial}{\partial x}(\overline{U_z}.h)][u_z\frac{\partial h}{\partial x}]\right)$$

$$-\frac{\partial}{\partial x}\left(hw^*[-\frac{\partial}{\partial x}(\overline{U_z}.h)]\right)-\frac{\partial}{\partial x}\left(hw^*[u_z\frac{\partial h}{\partial x}]\right)+\frac{\partial}{\partial x}\left(\int_0^h [z\frac{\partial}{\partial x}(\overline{U_z}.z).\frac{\partial^2}{\partial x.\partial z}(\overline{U_z}.z]dz\right)-$$

$$-\frac{\partial}{\partial x}\left(\int_0^h [z.u_z.\frac{\partial}{\partial x}(\overline{U_z}.z).\frac{\partial^2 z}{\partial x.\partial z}]dz\right)-\frac{\partial}{\partial x}\left(\int_0^h [z.u_z(\frac{\partial z}{\partial x}).(\frac{\partial^2}{\partial x.\partial z}(\overline{U_z}.z))]dz\right)+$$

$$+\frac{\partial}{\partial x}\left(\int_0^h [z.u_z^2.(\frac{\partial z}{\partial x}).(\frac{\partial^2 z}{\partial x.\partial z})]dz\right)-\frac{\partial}{\partial x}\left(hw^*.\frac{\partial}{\partial x}(\overline{U_z}h)\right)+\frac{\partial}{\partial x}\left(hw^*.u_z.\frac{\partial h}{\partial x}\right)+\frac{1}{\rho}\tau_b = 0 \qquad (20)$$

Assuming that the w* and h change slowly; based on order analysis of terms of equation (20), simpliying the equation (20) and bypassing the high order infinitesimal, we obtain the first equation:

$$\frac{\partial}{\partial t}\int_0^h u.dz -\frac{1}{3}.\frac{\partial}{\partial x}\left\{h^3\left(\frac{\partial^2 u}{\partial x\partial t}\right)\right\}+\frac{\partial}{\partial x}\int_0^h u^2.dz - u_m.\frac{\partial h}{\partial t}-u_m\frac{\partial}{\partial x}\int_0^h u.dz +u_m.w^*$$

$$+g\frac{\partial}{\partial x}.(\frac{h^2}{2})+\frac{1}{\rho}\tau_b = 0 \qquad (21)$$

Substituting (14) into (3), and putting z = h, we receive the second equation:

$$\frac{dh}{dt}=\frac{\partial h}{\partial t}+u_m\frac{\partial h}{\partial x}=-\frac{\partial}{\partial x}\int_0^h u.dz +u_m\frac{\partial h}{\partial x}+w^* \qquad (22)$$

Thus, we obtain the system of two partial differential equation as follows:

$$\frac{\partial}{\partial t}\int_0^h u.dz -\frac{1}{3}.\frac{\partial}{\partial x}\left\{h^3\left(\frac{\partial^2 u}{\partial x\partial t}\right)\right\}+\frac{\partial}{\partial x}\int_0^h u^2.dz - u_m.\frac{\partial h}{\partial t}-u_m\frac{\partial}{\partial x}\int_0^h u.dz +u_m.w^* +g\frac{\partial}{\partial x}.(\frac{h^2}{2})+\frac{1}{\rho}\tau_b = 0$$

$$\frac{\partial h}{\partial t}+\frac{\partial}{\partial x}(\overline{U}_h.h)=w^* \qquad (23)$$

Now, we modify the system of equations (23) into the following forms:

$$\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial x} = \frac{A}{h} w^* + q$$

$$\frac{\partial Q}{\partial t} - \frac{1}{3} \cdot \frac{\partial}{\partial x} \left\{ h^2 \left(\frac{\partial^2 Q}{\partial x \partial t} \right) \right\} + \frac{\partial \left(Q^2 / A \right)}{\partial x} + (g + a) A \frac{\partial h}{\partial x} =$$

$$gAi - gn^2 \frac{Q|Q|}{A} R^{-4/3} + \beta q v \tag{24}$$

where:
A is the cross-section area (m^2), R is the hydraulic radius (m), Q is the discharge (m^3/s), q is the lateral inflow (m^3/s/m), β is the coefficient factor of direction of lateral inflow, w* is the vertical velocity at river bed (m/s), $a = \frac{\partial w *}{\partial t}$ is the acceleration of vertical velocity, h is the water depth, v is the average velocity at a cross-section, i is the slope of the river bed, g is the acceleration of gravity (m/s^2), n is the roughness coefficient, t is the time (s), x is the coordinate along the flow direction (m), $\frac{\partial Q}{\partial t}$ is the inertial term, $\frac{\partial (Q^2 / A)}{\partial x}$ is the convective term, $(g + a) A \frac{\partial h}{\partial x}$ is the force produced by the pressure of fluid which is influenced by the acceleration of vertical velocity at the river bed $a = \frac{\partial w *}{\partial t}$.

Comment:

The system of equations (24) differs from the classical Saint-Venant equation system by these three terms: (i) w* in the continuity equation is the vertical velocity at the river bed (ii); $a = \frac{\partial w*}{\partial t}$ is the acceleration due to the vertical velocity w* at the river bed and (iii) gAi in the equation of motion; qualitatively, the system of equations (24) is an extension of the 1D Saint-Venant equation system. The system of equations can describe the flow problem in one-dimension in which there is an occurrence of large vertical velocity, a barrier at the river bed, etc.

Numerical Approximation

The Taylor-Galerkin finite element method, the third order accurate in time [1][14], is applied to solve the equation (24). With the Taylor-Galerkin solution, the discretization is first conducted in time by Taylor's expansion and then in space by the Galerkin procedure [1][14]. With the temporal discretization, let us perform the quantity $(h,Q)^{n+1}$ in a Taylor series of Δt around t=t_n up to the third order. In the spatial discretization, we use the Standard Galerkin finite element method with using the second order interpolation function.

(a)
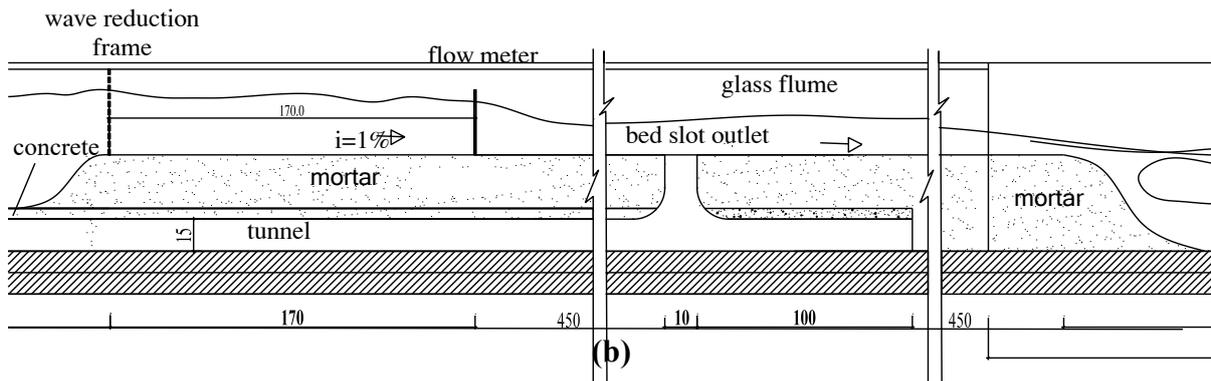
(b)

**Figure 2. (a) Water surface in a flume, case experiment with discharge
Q=0.045(m³/s); (b) Sketch of the lab experiment at National
Laboratory for Coastal and River Dynamics in Vietnam**

Lab Experiment

This paper presents a laboratory experiment to verify the numerical solution of one-dimensional open channel flow having the vertical velocity W^* at the river bed (Eq.24). The lab experiment was conducted at the National Laboratory for Coastal and River Dynamics in Vietnam. The experimental model is the flow in a glass flume having rectangular cross-section; for creating the vertical velocity at the bottom of flume, the flume was divided into two parts, the upper flow was the main stream and the lower flow was sprayed up for generating the vertical velocity W^*.
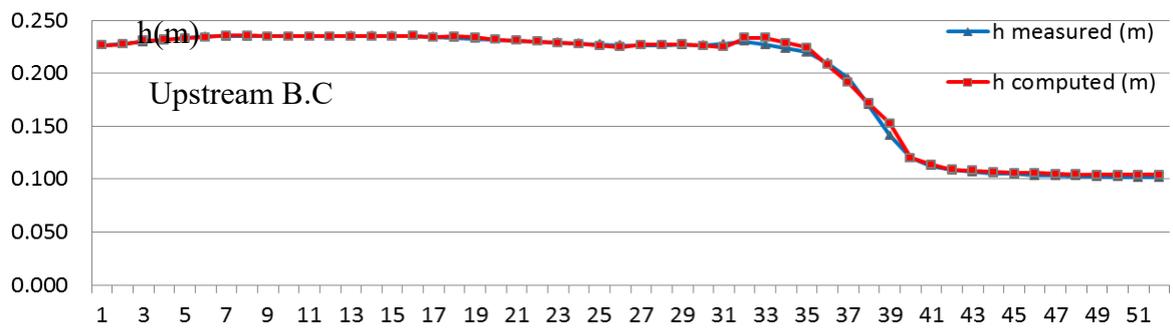
**Figure 3. The water surface in a flume in the case of the experiment with discharge
Q=0.045(m³/s) at upstream boundary condition (B.C)**

Comments

In general, the solution of the system of equation (24) according to the Taylor-Galerkin finite element method is quite consistent with the experimental data (Fig. 3); but because of disturbance, the water surface has a greater difference at the cross-section where having the vertical velocity W^* at the bottom of the flume, the max error is 5.5%.

Conclusions

Qualitatively, the system of equations (24) is the extension of the classical 1D Saint-Venant equation system [2][3][5][8][9][11][12]. This equation system is solved by the Taylor-Galerkin finite element method with the third order of accuracy in time.

The received governing equation is shown to be able to capture the modeling of 1D river flow better, such as in the case of having the vertical velocity at the river bed in condition of one-dimensional flow problem.

The algorithm and programs were verified by comparing the numerical results with the laboratory experiment data which were shown successfully validated.

References

[1] D. Ambrosi and L. Quartapelley (1998), A Taylor Galerkin Method for Simulating Nonlinear Dispersive Water Waves, *Journal of computational physics* **146**.

[2] J.A.Cunge, F.M.Holly, Jr, A.Verwey (1980), *Pratical aspects of computational river hydraulics,* Pitman Publishing INC.

[3] Environmental Modeling Research Laboratory of Brigham Young University (2002), *Surface water modeling system,* Brigham young university.

[4] Le Van Nghi, Tran Dinh Hoi, Nguyen The Hung (2004), Finite element method with high-accuracy, two-stage, of the two vertical Reynolds flow model, *Proceeding of The National Conference on Fluid Mechanics* **2003**, 353-362.

[5] M. Hanif Chaudhry (1993), *Open Chanel Flow. Prentice Hall*, Englewood Cliffs, New Jersey.

[6] Nguyen Dinh Tri (2004), *Higher mathematics*, Tome 3, Construction Pub. House.

[7] Nguyen The Hung (1990), Mathematical model of the two dimensional vertical flow, *Journal of Vietnam National Science & Technology* **7+8**, Hanoi.

[8] Nguyen The Hung (2004), *Finite Element Method in Flow Problems*, Construction Pub. House.

[9] Rainer Ansorge (2003), *Mathematical models of fluid dynamics,* Wiley-VCH GmbH &Co. KgaA.

[10] S.Vedula, P.P. and Mujumdar (2005), *Water Resources Systems : Modelling Techniques and Analysis,* Tata-McGraw Hill.

[11] Ven-te-chow, David R. Maidment, Larry W.Mays (1998), *Applied Hydrology,* McGraw-Hill, Inc.

[12] Weiming Wu (2007), *Computational River Dynamics,* Taylor and Francis e-Library.

[13] Xingwei Chen, and Yee-Meng Chiew, M.ASCE (2004), Velocity Distribution of Turbulent Open-Channel Flow with Bed Suction, *Journal of Hydraulic Engineering* **Vol. 130, No. 2**.

[14] O.C. Zienkiewicz, R.L. Taylor and J.Z. Zhu (2005), *The Finite Element Method for Fluid Dynamics,* Published by Butterworth-Heinemann.

Damage identification in space frame structures using convolutional neural networks and modal strain energy

*Duy-Vu Dinh[1,2]**, †Khac-Duy Nguyen**[3]**, Tan-Phu Vo**[1,2] **and †Duc-Duy Ho**[1,2]

[1]Faculty of Civil Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam
[2]Vietnam National University Ho Chi Minh City, Ho Chi Minh City, Vietnam
[3]School of Civil and Environmental Engineering, Faculty of Engineering, Queensland University of Technology, Australia

*Presenting author: ddvu.sdh212@hcmut.edu.vn
†Corresponding author: k25.nguyen@qut.edu.au
†Corresponding author: hoducduy@hcmut.edu.vn

Abstract

Structural health monitoring (SHM) is a growing field of monitoring and evaluating the health state of structures. 3-D frame structures are one of the most common structures these days. Over their life spans, they could experience damage such as corrosion and cracks due to material degradation or overloading. This study aims to develop a reliable real-time damage identification method for space frame structures. The aim of this study can be achieved by using deep learning techniques and vibration parameters. A new damage identification method is proposed by utilizing convolutional neural networks (CNNs) to detect damage from modal strain energy. Modal strain energy is well-known for its high sensitivity to damage or change of stiffness, and therefore, it can be utilized as a damage index in SHM. CNNs have the ability to self-learn and immediately make predictions, making them suitable for real-time monitoring of structural health. To validate the effectiveness of the proposed method, a 3-D steel frame is used as a case study. The results demonstrate that the proposed method accurately detects, localizes, and quantifies damage in the frame. This study also paves the way for further research on applying deep learning techniques for SHM and other areas of structural engineering.

Keywords: Damage identification, Modal strain energy, Convolutional neural network, Space frame.

1 Introduction

During the life cycle of structures, damage can be induced by various objective and subjective factors. Damage can result in reduced stiffness, instability, and potential implications for the overall durability of the structure. Therefore, damage identification has become a critical aspect of structural integrity. The damage identification process involves four key levels: detection of damage occurrence, prediction of damage locations, assessment of damage severity, and estimation of residual life [1]. Two main approaches are commonly employed in this field: destructive testing and vibration-based methods. Structural health monitoring (SHM) has gained considerable popularity due to its inherent advantages, such as flexibility in measurements, relatively low costs, and non-destructive. The presence of structural defects leads to changes in dynamic characteristics, such as natural frequency, mode shape, and modal strain energy.

The finite element method (FEM) has played a pivotal role in providing computer models of structures, enabling the analysis of their behavior, frequency response, and mode shapes based on parameters such as mass and stiffness. When damage occurs, leading to alterations in stiffness, the dynamic behavior of the computer model is also changed. This technique provides

an opportunity to employ inverse problem-solving techniques to determine the location and extent of the damage. SHM damage detection method comprises two essential components: a vibration-based index and an algorithm. Modal strain energy index (MSE) is a widely used vibration-based index based on mode shapes and exhibits sensitivity to changes in structural stiffness [2]. The effectiveness of MSE in assessing the location and extent of structural damage in three-dimensional structures has often been demonstrated, surpassing other mode shape-based approaches. The modal strain energy method and its variants, such as modal strain energy change index (MSEC) and modal strain energy change ratio index (MSECR), have been successfully applied to detect multiple damages in structures, including trusses and plain frames. Numerous studies have focused on damage detection in such structures using modal information and heuristic optimization techniques, including particle swarm optimization [3], hybrid multi-objective genetic algorithm [4], isogeometric analysis [5]. Although these methods have shown high accuracy in identifying the locations and extent of damage in space frames, the optimization process may limit their real-time monitoring capabilities.

The utilization of convolutional neural networks (CNNs) has gained significant popularity due to their capacity for self-learning and information storage, resulting in significantly reduced prediction time and improved real-time monitoring. CNNs improve the ability to detect damage in structures with a large number of elements by taking advantage of the feature extraction capabilities of this algorithm [6][7][8].

This study presents a novel approach to identifying damage in spaceframe structures by exploring the feasibility of using modal strain energy for deep learning-assisted damage identification. The proposed methodology aims to detect, locate, and quantify the damage effectively. A comprehensive data set is initially generated through finite element analysis (FEA), capturing the damage index based on modal stress energy for various randomly generated scenarios. Subsequently, a CNN model is developed, trained, and validated using distinct training and validation sets derived from the data set. Finally, the trained CNN model is tested for its capacity to identify the position and severity of damage in structures. This integrated approach has great potential to improve real-time damage assessment in complex 3-D frame structures.

In this study, the following outline is presented to elucidate the critical components of the research study. Section 2 provides a comprehensive discussion of the proposed approach's theoretical foundations. It encompasses exploring structural vibration principles, the modal strain energy index, and the use of CNNs in damage identification. Section 3 delves into the detailed methodology of the proposed approach, describing the steps involved in integrating the modal strain energy index and CNNs for real-time damage detection in 3-D frame structures. Moving on to section 4, a case study focuses on a 4-story steel space frame with dimensions comparable to a laboratory-scale structure. This case study aims to demonstrate the effectiveness and applicability of the proposed approach in a practical scenario. Finally, section 5 concludes the article, summarizing the findings, discussing their implications, and providing insights into future research directions. The paper aims to provide a comprehensive understanding of the theoretical foundations, methodology, and practical application of the proposed approach for real-time damage identification in 3-D frame structure through this outline.

2 Theory

2.1 Stiffness reduction

Damage in a structure could be characterized by a reduction in stiffness [9]-[12]. Characterizing damage in structures solely based on stiffness reduction may not encompass all types of

damage that can occur. This particular approach has its limitations within the scope of this study. However, it remains applicable for detecting potential linear forms of damage, such as stiffness reduction resulting from bolt loosening, corrosion, or fatigue cracking induced by cyclic loading. Despite these considerations, for computational analysis, it is suitable to express degradation in terms of stiffness reduction within the context of this study.

In the context of computational analysis, it is suitable to represent the deterioration or degradation of structural elements by quantifying it as a reduction in stiffness. Therefore, to simulate damages occurring at any specific element within the structure, the original value of Young's modulus for each element is adjusted using a specific formulation. This methodology ensures a precise representation of structural deterioration within the computational framework employed in this research.

$$(EI)_j^d = (1 - \alpha_j)(EI)_j^h \tag{1}$$

In the above equations, $(EI)_j^h$ and $(EI)_j^d$ represent the stiffness of elements in undamaged and damaged states, respectively, while denoting the percentage reduction in stiffness for element α_j. Consequently, the stiffness reduction corresponding to elements prone to damage can be expressed using a vector of percentage reduction as follows:

$$\alpha = (\alpha_1, \alpha_2, ..., \alpha_k) \tag{2}$$

Here, α represents a vector that indicates the extent of damage in percentage for the elements requiring identification.

2.2 Damage sensitivity features

2.2.1 Natural frequency change

The equation of motion for a dynamic system n-degrees-of-freedom (n-DoF) can be expressed as:

$$\mathbf{K} \times \mathbf{\Phi} = \mathbf{M} \times \mathbf{\Phi} \times \omega^2 \tag{3}$$

Where $\mathbf{K}$ and $\mathbf{M}$ are the $n \times n$ system stiffness and mass matrices, $\mathbf{\Phi}$ is the mode shape matrix, and ω^2 is the square of angular frequency vector. The eigenvalue problem is solved to obtain eigenvalues (square of angular frequency) and eigenvectors (mode shape) using the stiffness and mass matrices of the frame.

$$\mathbf{\Phi}_i^T \times \mathbf{M} \times \mathbf{\Phi}_i = 1 \tag{4}$$

$$\mathbf{\Phi}_i^T \times \mathbf{K} \times \mathbf{\Phi}_i = \omega_i^2 \tag{5}$$

The frequency of the i-th vibration mode is calculated from the angular frequency of the i-th vibration mode.

$$f_i = \frac{\omega}{2\pi} \tag{6}$$

Natural frequencies are relatively straightforward to measure, and implementing methods for measuring them is easily achievable in practical applications [13]. On the other hand, measuring

natural frequencies to identify structural damage has limitations. Slight frequency changes in larger structures may not indicate significant damage, and variations in mass or temperature can create uncertainties [11]. The natural frequency change has been utilized in various studies as an effective method for damage detection [10][13].

The frequency change index is determined as the difference in frequency in the i-th mode of vibration between after and before damage, given by:

$$fC_i = f_i^d - f_i^h \tag{7}$$

Where f_i^h and f_i^d are frequencies of the structures in the i-th mode of vibration corresponding to the undamaged state and the state requiring damage identification, respectively.

2.3   Modal strain energy

The modal strain energy index MSE [2], which is composed of the element stiffness matrix and the mode shape, is a highly sensitive index compared to other indices constructed from structural vibration analysis, such as natural frequencies and mode shapes, for the purpose of detecting structural damage [14][15]. In the FEM, the displacement of an element is approximated by a shape function, which is a function of the nodal displacements. In this case, the expression for the modal strain energy of the j-th element in the i-th mode of vibration is given as follows:

$$MSE_{ij} = \mathbf{\Phi}_i^T \mathbf{K}_j \mathbf{\Phi}_i \tag{8}$$

Where $\mathbf{\Phi}_i$ is the mode shape vector in the i-th mode of vibration, and $\mathbf{K}_j$ is the stiffness matrix of the j-th element in the global coordinate system.

The modal strain energy of the j-th element in the i-th vibration mode, corresponding to the undamaged state MSE_{ij}^h and the damaged state MSE_{ij}^d, is expressed as follows:

$$MSE_{ij}^h = \mathbf{\Phi}_i^{hT} \mathbf{K}_j \mathbf{\Phi}_i^h \tag{9}$$

$$MSE_{ij}^d = \mathbf{\Phi}_i^{dT} \mathbf{K}_j \mathbf{\Phi}_i^d \tag{10}$$

Where $\mathbf{\Phi}_i^h$ and $\mathbf{\Phi}_i^d$ are the mode shape vectors for the undamaged and damaged states in the i-th vibration mode, respectively; is the stiffness matrix of the j-th element in the global coordinate system. In both Eq. (9) and Eq. (10), the undamaged elemental stiffness matrix is utilized. Although the stiffness matrix of a damaged element differs from that of an undamaged element, the undamaged stiffness matrix can be used as a substitute since the specific matrix for the damaged state is unknown [14]. This approximation may introduce some differences, but it is considered acceptable, as the stiffness matrix is a weighting factor in the modal strain energy index, affecting the vibration modes.

The modal strain energy change index (MSEC) [14] is developed based on the difference in modal strain energy values between an element's damaged and undamaged states. The absolute value of MSEC of a damaged element is greater than that of any undamaged element. Undamaged elements connected to the damaged element also have smaller MSEC. If an undamaged element is farther away from the damaged element, its MSEC value is even lower. This holds for the cases where the structure has multiple damaged elements [14]. The MSEC index is determined as the difference in the modal strain energy of the j-th element in the i-th mode of

vibration, given by:

$$MSEC_{ij} = MSE_{ij}^d - MSE_{ij}^h \tag{11}$$

Where MSE_{ij}^h and MSE_{ij}^d are the modal strain energies of the j-th element in the i-th mode of vibration corresponding to the undamaged state and the state requiring identification of damage, respectively, calculated using Eq. (9) and Eq. (10).

2.4 Deep learning and convolutional neural networks

Deep learning (DL), a subset of machine learning in the field of artificial intelligence, is a potent approach to extracting complex, high-level abstractions and data representations from large datasets [16]. DL learns to extract features directly from raw data autonomously, eliminating the need for human intervention in feature extraction. This characteristic enables DL to establish a direct mapping from input data to desired outputs without the requirement of pre-extracting features. Furthermore, DL exhibits proficiency in capturing complex and abstract features by hierarchically building them on simpler, low-level learned features [17].

CNNs exhibit a distinct structure consisting of input layers, multiple hidden layers, and an output layer. Each layer within the CNN framework consists of interconnected processing units that collectively contribute to the network's functioning. Non-linear transformations are applied to each layer's inputs, generating representations as outputs. Specifically, the input layer receives and transmits the input signals to the hidden layers. In the context of solving multi-objective optimization problems in engineering structures, the design variables of the optimization problem serve as inputs to the CNN model. The hidden layers receive these design variables and learn the relationships between the design variables and the objective and constraint values. Ultimately, the output layer provides predictions for the objective and constraint values of the optimization problem.

3 Proposed CNN Model and Methodology

3.1 Proposed CNN model

Preparation of Vibration Parameters

Using the proposed method, the elements of a CNN model to detect damage in the space frame involve several components shown in Fig. 1. The FEM computes the corresponding vibration parameters for different random damage scenarios. The steps in calculating these parameters are: First, an FEA model is simulated using Python programming language, incorporating material parameters and geometric dimensions of the structural element and the whole frame. Random damage scenarios are generated, where each element is assigned a stiffness reduction value within the $[0, 1]$ range, representing each element's stiffness reduction in the FEA model. Next, the stiffness and mass matrices for the elements are determined in the local coordinate system and the global coordinate system using the FEM and the given damage scenarios. Subsequently, the stiffness and mass matrices of the frame are determined. The eigenvalue problem is solved to obtain eigenvalues (square of angular frequency) and eigenvectors (mode shape) using the stiffness and mass matrices of the frame. These eigenvectors are normalized by using the mass and stiffness matrices. Finally, the modal strain energy and frequency-related indices are computed. These steps in preparing the training data involve the construction of the FEA model, determining stiffness and mass matrices, solving the eigenvalue problem, and computation of relevant indices related to modal strain energy MSEC and frequency change.
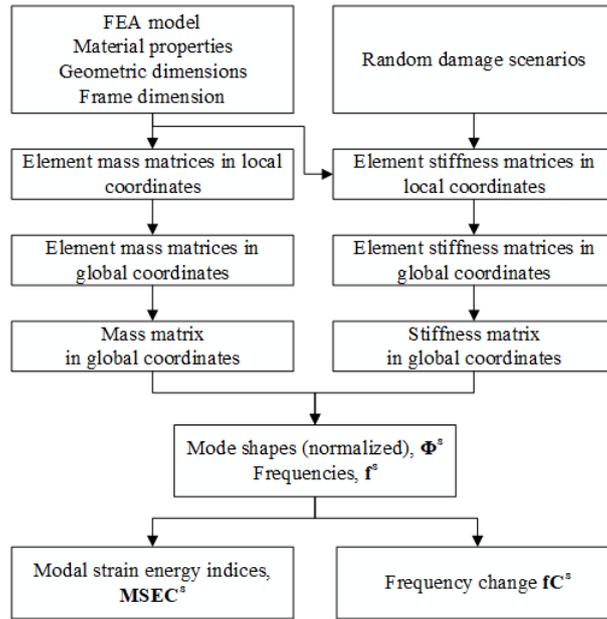
Figure 1: Vibration parameter preparation

Input Data

The above data are then combined according to the input data set requirements. The proposed model analyzes two input data sets for modal strain energy, MSEC with and without frequency change.

These various data sets are used to explore and evaluate the effectiveness of different strain energy indices in identifying and characterizing damage. By considering or omitting frequency information, the proposed method aims to assess the impact of frequency on the accuracy and reliability of damage detection. The different combinations of modal strain energy and frequency-related parameters enable a comprehensive investigation of their interplay in identifying structural damage. The corresponding results and analyzes provide valuable information on the performance and potential applications of the proposed method.

Output Data

The output data are presented in Table 1.

Table 1: Output data

Output data			
α_1	α_2	...	α_n

CNN Model Architecture

The proposed method uses the CNN model to leverage its self-learning and predictive capabilities to identify damage in the space frame based on the input data representing the modal strain energy. Additionally, the CNN model includes a convolutional layer capable of extracting features from the input data corresponding to the considered vibration modes.

The CNN model consists of a convolutional layer in the convolution phase and multiple hidden layers in the fully connected phase, as shown in Table 2. The convolutional layer has a size equal to the number of input data rows (representing the initial vibration modes). The convolutional layer filters aid in separating the vibration mode features of each mode with the corresponding

filter. The convolution phase includes fully connected layers, and the i-th FC layer has nodes. These parameters depend on the structure, number of nodes, beam elements, and input data.

Table 2: Proposed CNN model

Phase	Layer	Component	Activation function
-	Input	$X \times Y \times 1$	-
Convolutional	Convolutional	$X \times 1 \times C_1$	ReLU
Fully Connected	Flatten	$X \times Y \times 1$	-
Fully Connected	Fully Connected	FC_1	ReLU
Fully Connected	Fully Connected	FC_2	ReLU
Fully Connected	Fully Connected	...	...
Fully Connected	Fully Connected	FC_k	ReLU
Fully Connected	Output	n	-

Other Components of the CNN Model

The Rectified Linear Units activation function (ReLU) is widely used in DL due to its non-linear characteristics. It selectively activates neurons, enabling them to be deactivated if the output of the linear transformation is.

3.2   Methodology

To determine the location and extent of damage in space frame structures, a methodology combining the modal strain energy method and CNNs is used in training and detection. The process begins with generating random damage scenarios, then determining the vibration parameters associated with these scenarios using the modal strain energy method based on FEM. Then, a CNN model is created to predict structural damage based on the obtained parameters. The CNN model is trained using random scenarios until the desired number of training iterations is achieved, with updates being made to the detection CNN model. Finally, the vibration data of the structure to be detected is processed according to the input data requirements and fed into the detection model, which produces an identification outcome.

This approach combines the capabilities of the modal strain energy method and CNNs to enhance the accuracy and effectiveness of damage detection in 3-D frame structures. The random generation of damage scenarios enables comprehensive training and evaluation of the CNN model, ensuring its robustness and adaptability. By incorporating the modal strain energy parameters obtained through the FEM, the CNN model is able to capture and analyze the unique patterns and characteristics of structural damage. The detection flowchart provides a systematic and reliable framework for identifying and assessing damage in 3-D frame structures, contributing to the advancement of SHM and maintenance practices.

4   Numerical case study

4.1   Structure's information and FEA model

A case study involving a 3-D modular steel frame structure has been undertaken to assess the effectiveness of the proposed damage detection method. The structure comprises simple column and frame (beam) elements, and an asymmetric configurations four-story structure has been developed to serve as the basis for analysis. A steel space frame with four floors along the Z-axis, two spans along the Y-axis, and one span along the X-axis consists of 52 bar elements. Fig. 2 shows the case-study structure.

The column section dimensions measure 0.00635 m by 0.0508 m, with a length of 0.6096 m. Similarly, the frame section dimensions are 0.00635 m by 0.0508 m, with a depth of 0.6096 m
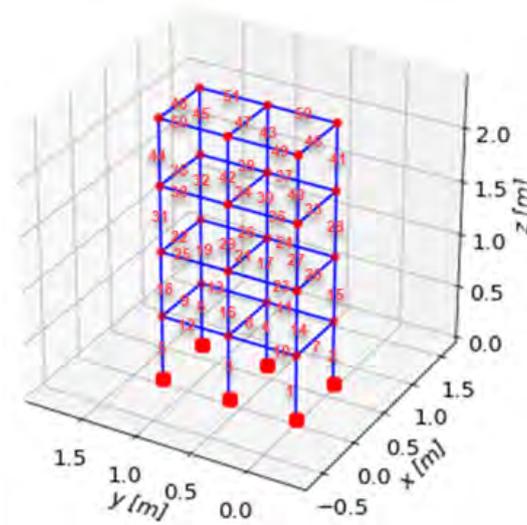
Figure 2: structure of case-study frame

for each bay. These specifications form the basis for constructing the modular steel frame structure, which enables the examination of different structural types and configurations for damage detection evaluation. The material properties, geometric dimensions, and span dimensions are presented respectively in Table 3, Table 4, Table 5.

Table 3: Material properties

Bar element	Young's Modulus, E (N/m^2)	Density, ρ (kg/m^3)	Poisson coefficient, ν
Beam	1.96×10^8	7.8	0.3
Column	1.96×10^8	7.8	0.3

Table 4: Geometry properties

Bar element	Width, b (m)	Height/depth, h (m)
Beam	0.00635	0.0508
Column	0.00635	0.0508

Table 5: Span dimensions

Axis	X (m)	Y (m)	Z (m)
Dimension	0.6096	0.6096	0.6096

The structure under investigation has been subjected to numerical modeling using the FEM. In this model, each node is equipped with six degrees-of-freedom (DoFs), encompassing three translational DoFs and three rotational DoFs. This configuration allows for a comprehensive analysis of the structural behavior. To accurately capture the stiffness properties of the structural elements, (12×12) stiffness matrices and consistent mass matrices have been employed. These matrices facilitate a detailed investigation into the structure's response under various loading conditions. Overall, the numerical model incorporates 144 DoFs, providing a meticulous representation of the structure's dynamic characteristics.

4.2 Damage Scenario

In Fig. 6, the structure under consideration exhibits multiple damages across five elements, each with varying proportions of damage and different types of components. These bars are arranged from the base to the top of the structure.

Table 6: Damage Scenario

Element	1	10	17	33	52
Extent of damage (%)	20	30	20	30	30

4.3 CNN model

The first eight mode shapes are used. Based on the proposed method in section 3, the CNNs model with different components is presented in Table 7.

Table 7: Component of the architecture of CNN model

n	m	C_1	k	FC_1	FC_2	FC_3	FC_4
52	8	200	4	1500	1500	1500	1500

Table 7 presents the components of the CNNs model, where the values for each component are specified. These values include the number of elements in the structure, the number of initial vibration modes considered based on the original structure, the number of filters in each fully connected layer (excluding the flattening layer and the output layer) and the number of nodes in the respective fully connected layer. During the training phase, the damage scenarios are generated randomly with severity ranging from 0% to 50%. A total of 100,000 randomly generated damage scenarios are created.

4.4 Damage Detection Results

Input data: MSEC
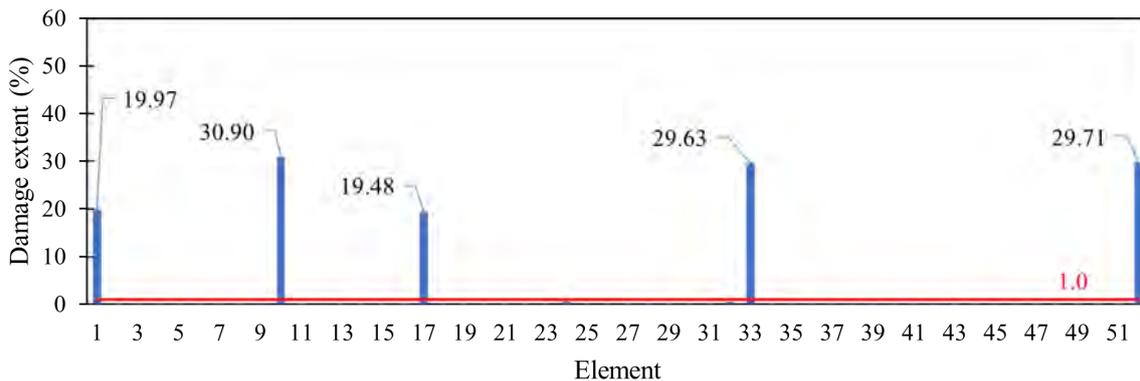
Figure 3: Results of the proposed method with input data: MSEC

The proposed method excelled in accurately identifying the positions of damaged beams by utilizing the MSEC index of the first eight vibration modes as input data. This method showcased its capability to detect damage locations within spatial frame structures, surpassing the threshold of 1.0%, as described in the prescribed scenarios (Fig. 3).

Regarding assessing the severity of the damage, Fig. 3 provides insight into the average error achieved by the proposed method when employing the MSEC index of the first eight vibration

modes. In particular, the average error was found to be a remarkable 1.59%, with a maximum error of 3.01%.

Input data: MSEC and frequency change
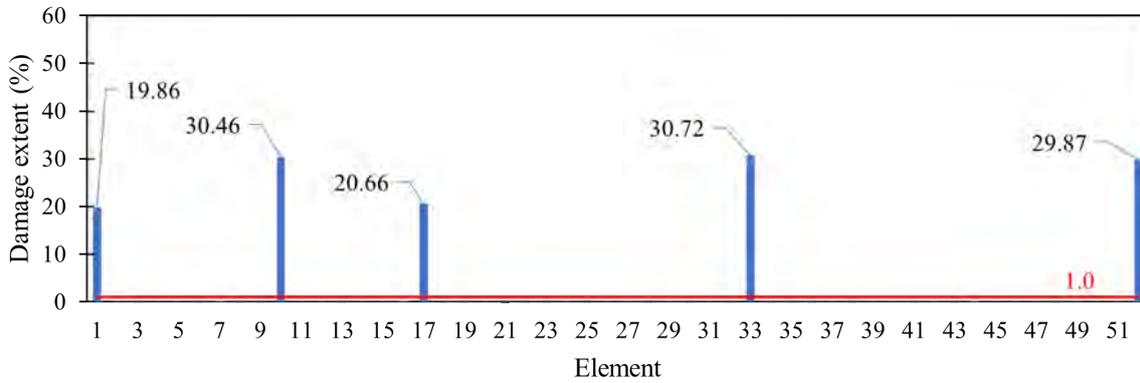
Figure 4: Results of the proposed method with input data: fC and MSEC

The proposed method effectively identified the precise locations of the damaged beams according to the specified scenarios by incorporating the frequency change and the MSEC index of the first eight vibration modes as input data. This method demonstrated its remarkable ability to accurately identify the positions of structural damage with a severity exceeding the threshold of 1.0% (Fig. 4).

Regarding the assessment of damage severity, the proposed method that uses frequency change and MSEC of the first eight vibration modes as indicators achieved an average error of 1.67% and a maximum error of 3.30% (Fig. 4).

Comparison

The comparison of the results obtained from the proposed method reveals that the damage detection capability of the method varies when different input data sets are used. All the input data sets employed in the proposed method enable the detection of the location and severity of the damage. Both data sets include or exclude the natural frequency, with corresponding damage thresholds of 1.0%.
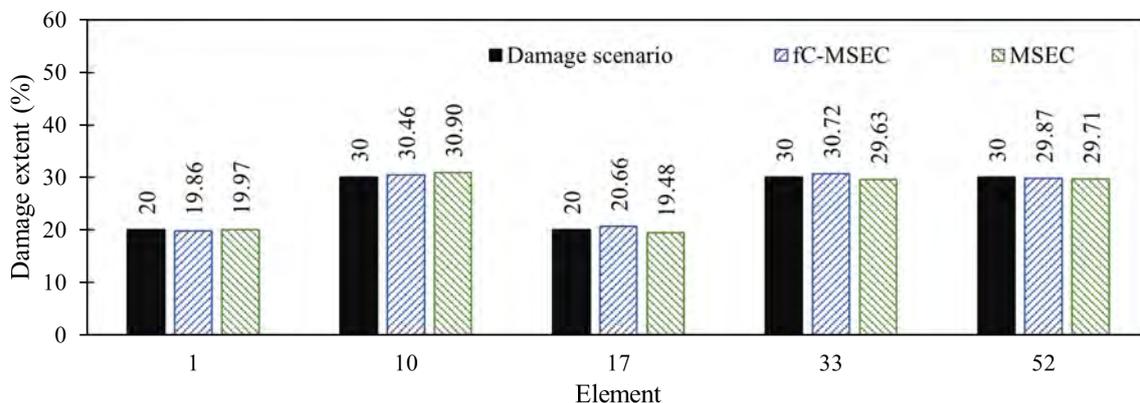
Figure 5: Detection damage extent of damage scenario using the proposed method with various input data

Regarding assessing the severity of the damage, most of the data sets provided exhibit average errors below 1.0%. On the other hand, the (MSE) data set exhibits the highest maximum error,

reaching 4.77%.
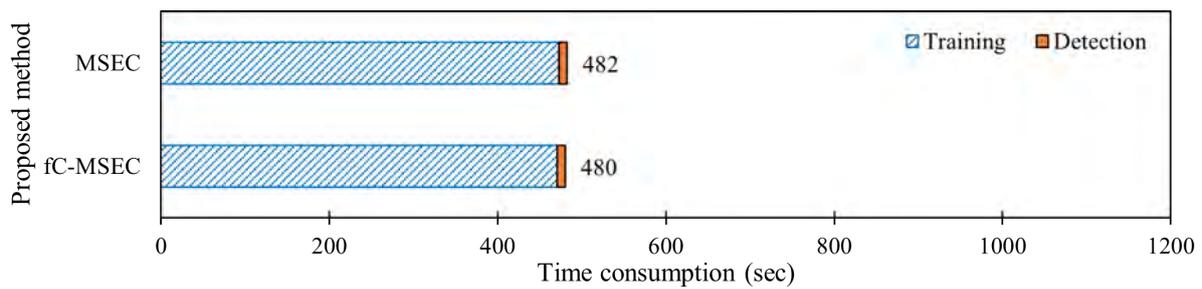
Figure 6: Damage extent detection of damage scenario using the proposed method with various input data

Regarding computational time, which includes training and damage detection time, the different input data sets show similar detection times of around 500 seconds (Fig. 6). The proposed method predominantly consumes time during the training phase of CNNs, while the detection time is nearly instantaneous. Therefore, the method can be considered capable of real-time detection.

The proposed method utilizes the modal strain energy-based approach combined with CNNs to demonstrate its remarkable capability in damage identification.

5 Conclusion

The proposed method demonstrates a remarkable ability to swiftly and accurately identify the presence, location, and extent of damage. The detection time is nearly instantaneous, offering significant advantages in practical applications. Furthermore, both input data sets, comprising the MSEC with and without frequency change index, exhibit exceptional quality, further bolstering the reliability and robustness of our approach. The effectiveness of our method is underscored by its ability to provide detailed and precise information, enabling informed decision-making in structural health monitoring. These findings contribute valuable insights to the field and highlight the potential for broader adoption of our approach in academic and practical contexts.

Acknowledgment

References

[1] Doebling, S. W., Farrar, C. R. and Prime, M. B. (1998) A summary review of vibration-based damage identification methods. *Shock and Vibration Digest* **30**, 91–105.

[2] Stubbs, N., Kim, J. T. and Farrar, C. R. (1995) Field verification of a nondestructive damage localization and severity estimation algorithm. *Proceedings of the Thirteenth International Modal Analysis Conference, IMAC XIII* **2**, 210–218, Nashville, USA.

[3] Seyedpoor, S. M. (2012) A two stage method for structural damage detection using a modal strain energy based index and particle swarm optimization. *International Journal of Non-Linear Mechanics* **47**, 1–8.

[4] Cha, Y. and Buyukozturk, O. (2015) Structural damage detection using modal strain energy and hybrid multi-objective optimization. *Computer-Aided Civil and Infrastructure Engineering* **30**, 347–358.

[5] Khatir, S., Wahab, M. A., Boutchicha, D. and Khatir, T. (2019) Structural health monitoring using

modal strain energy damage indicator coupled with teaching-learning-based optimization algorithm and isogeometric analysis. *Journal of Sound and Vibration* **448**, 230–246.

[6] Khodabandehlou, H., Pekcan, G. and Fadali, M. S. (2018) Vibration-based structural condition assessment using convolution neural networks. *Structural Control and Health Monitoring* **26**, e2308.

[7] Yu, Y., Wang, C., Gu, X. and Li, J. (2019) A novel deep learning-based method for damage identification of smart building structures. *Structural Health Monitoring* **18**, 143–163.

[8] Teng, S., Chen, G., Liu, G., Lv, J., and Fangsen, C. (2019) Modal strain energy-based structural damage detection using convolutional neural networks. *Applied Sciences* **9**, 3376.

[9] Van Houten, E. E., Paulsen, K. D., Miga, M. I., Kennedy, F. E. and Weaver, J. B. (1999) An overlapping subzone technique for MR-based elastic property reconstruction. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **42**, 779–786.

[10] Lee, Y. S. and Chung, M. J. (2000) A study on crack detection using eigenfrequency test data. *Computers & Structures* **77**, 327-342.

[11] Kim, J. T., Ryu, Y. S., Cho, H. M. and Stubbs, N. (2003) Damage identification in beam-type structures: frequency-based method vs mode-shape-based method. *Engineering Structures* **25**, 57–67.

[12] Barroso, L. R. and Rodriguez, R. (2004) Damage detection utilizing the damage index method to a benchmark structure. *Journal of Engineering Mechanics* **130**, 142–151.

[13] Cawley, P. and Adams, R. D. (1979) The location of defects in structures from measurements of natural frequencies. *The Journal of Strain Analysis for Engineering Design* **14**, 49–57.

[14] Shi, Z. Y., Law, S. S. and Zhang, L. M. (1998) Structural damage localization from modal strain energy change. *Journal of Sound and Vibration* **218**, 825–844.

[15] Sazonov, E. and Klinkhachorn, P. (2005) Optimal spatial sampling interval for damage detection by curvature or strain energy mode shapes. *Journal of Sound and Vibration* **285**, 783–801.

[16] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *nature* **521**, 436–444.

[17] Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M. and Inman, D. J. (2018) Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks. *Journal of Sound and Vibration* **424**, 158–172.

Effect of concentration of ion in the solution on the deformation behavior of double network gel under simple tension

*†I. Riku and K. Mimura

Department of Mechanical Engineering, Osaka Metropolitan University, Japan

*†riku@omu.ac.jp

Abstract

Double network (DN) gels are soft-elastic materials consisting of a three-dimensional cross-linked polymer network and liquid filling the space among such network. To take advantage of the swelling-deformation behavior of DN gel and explore its possibility for applications in engineering as a structural member, the effect of concentration of ions in the solution on the deformation behavior of DN gel should be clarified, especially when the DN gel is loaded inside the solution. In this study, we at first proposed a nonaffine model for the DN gel to account for the change of the entangling structure of polymer chains during the swelling-deformation process, in which the change of the number of polymer chains per unit volume is depending on the first invariant of right Cauchy-Green deformation tensor. And then, the swelling ratio and the simple tensional mechanical response of DN gel is calculated when the DN gel is loaded inside the solution with different concentration of ions. The results show that the nonaffine model can reproduce the experimental results of the swelling ratio very well. Moreover, the high concentration of ions leads to the high deformation resistance.

Keywords: Double network gel, Concentration of ion, Nonaffine model, Swelling ratio, Simple tension.

Introduction

Double network (DN) hydrogel is synthesized from monomers in the presence of a cross-linker via a two-step network formation: the first step is forming a tightly cross-linked network gel of polyelectrolyte. Then, the first gel is immersed in an aqueous solution of a second monomer with a low ratio of cross-linking agent and carrying out a second polymerization in the first network. Owing to the polyelectrolyte nature, the first polyelectrolyte gel highly swells in the second monomer solution. As a result, the first gel network is highly extended in the final product and the amount of the second network in the final gel is in large excess to that of the first network[1].

The DN hydrogels have drawn much attention as an innovative material having both high water content and high mechanical strength and toughness. The phenomenon such as yielding and neck propagation can be observed in some tough DN gels. For example, on tensile tests of DN gels that made from relatively sparse first networks, narrow zones appear in the sample and grow up with further stretching. During the neck propagation, a plateau region appears in the loading curve. The plateau value of the tensile stress hardly depends on the stretching rate. After the neck propagation, the gel becomes fairly soft, showing an elastic modulus *ca.* 1/10 of the virgin sample, and sustains large elongation, up to an elongation strain of around 20. The observations on the softened gels after the tensile test demonstrate that irreversible structural change takes place inside the gels, although their appearance is almost unchanged.

To account for such irreversible structural change in the DN gel, in one of our former studies[2, 3, 4], we investigated the development of the microstructure of polymer chains network of the

gel based on a nonaffine polymer chains network model[5], which was originally developed for the orientation hardening of amorphous polymers and may account for the change in the entanglement situation for the physical linkages during the deformation processes. It was found that the nonaffine polymer chains network model together with a stablization algorithm for localized transformation of strain energy can be employed to reproduce the phenomenon of neck propagation in DN gels very well. In this study, the effect of concentration of ions in the solution on the swelling ratio and the simple tensional mechnical response of DN gel are to be investigated.

Conditions of Equilibrium

The basic idea of the derivation of the conditions of equilibrium for the dry polymer chains network bearing acidic groups (pH-sensitive gel) and the aqueous solution is from the work done by Marcombe et al.[6]. It is convenient to consider that, for a pH-sensitive gel, the variational statement includes the following fields: the displacement of the network, the concentations of the solvent and ions, and the degree of acidic dissociation. The variations are subject to auxiliary conditions of several types, including the conservation of various species, incompressibility of molecules, and electroneutrality in the gel and in the external solution.

When the network absorbs C number of solvent molecules, the deformation gradient of the network is expressed by $\mathbf{F}$, the ratio of volume of the swollen gel to that of the dry network is determined as $J = \det \mathbf{F}$. As an idealization, it is assumed that the volume of the gel is a function of the concentration of the solvent:

$$J = 1 + vC. \tag{1}$$

That is, all molecules in a gel are incompressible, and the volume of the gel is the sum of the volume of the dry network and the volume of the pure solvent molecules, where v is the volume per solvent molecule. Eq. (1) determines the concentration of solvent C, once the deformation gradient $\mathbf{F}$ is known.

The Helmholtz free energy of the gel in the current state, W, can be assumed to be a sum of several contributions[6]:

$$W = W_{net} + W_{sol} + W_{ion} + W_{dis}, \tag{2}$$

where W_{net} is due to stretching the network, W_{sol} mixing the solvent with the network, W_{ion} mixing ions with the solvent, and W_{dis} dissociating the acidic groups. For the convenience of formulation, it is preferred to introduce another free energy function $\hat{W}$ by using a Legendre transformation:

$$\hat{W} = W - \mu C, \tag{3}$$

where μ is the chemical potential of the solvent molecules and its value is dependent on the concentrations of the ions. Eq. (1), Eq. (2) and Eq. (3) form the bases for the model of ideal gels. In equilibrium, the change of the Helmholtz free energy of the composite vanishes and one can obtain that

$$s_{ki} = \frac{\partial \hat{W}}{\partial F_{ik}}, \quad C = -\frac{\partial \hat{W}}{\partial \mu}, \tag{4}$$

where s_{ki} is the nominal stress. Empoloying Eq. (4), the swelling-deformation responses of the gel under different mechanical and chemical constraints can be investigated directly.

Free Energy Functions

In the original Flory-Rehner model[7], specific functions are adopted for W_{net} and W_{sol}. In this study, we employ the best known formulation[8] as:

$$W = \frac{1}{2}Nk_BT(F_{ik}F_{ik} - 3 - 2\log J) - \frac{k_BT}{v}\left[vC\log\left(1 + \frac{1}{vC}\right) + \frac{\chi}{1 + vC}\right], \qquad (5)$$

where N is the number of polymer chains per unit volume, i.e. the density of crosslinks of the polymer chains, k_B is Boltzmann constant, T is the absolute temperature and χ is a dimensionless measure of the enthalpy of mixing. A combination of Eq. (1), Eq. (3) and Eq. (5) gives the desired free energy function:

$$\hat{W} = \frac{1}{2}Nk_BT(F_{ik}F_{ik} - 3 - 2\log J) - \frac{k_BT}{v}\left[(J-1)\log\left(\frac{J}{J-1}\right) + \frac{\chi}{J}\right] - \frac{\mu}{v}(J-1). \quad (6)$$

Usually, the affine movement of the polymer chains is assumed and the value of N is fixed, and the affine model of the gel is constructed. However, the observations on the DN gel after the tensile test demonstrate that irreversible structural change takes place inside the gel, although their appearance is almost unchanged[1]. When the DN gel is in tensile, the 1st polymer chains network is quite brittle and breaks into small pieces at small extensions. Subsequently, the 1st polymer chains network fragments into small clusters and the clusters play a role of crosslinkers of the 2nd polymer chains network. This irreversible structural change during the tension of the DN gel can be considered as one kind of the nonaffine movement of the polymer chains, i.e. the density of crosslinks of the polymer chains N vary from its initial value at the dry state, N_0, to its current value which is depend on the deformation gradient of the network $\mathbf{F}$. In this study, the nonaffine model of the DN gel is constructed and the value of N is proposed to vary inverse proportionally with the first invariant of right Cauchy-Green deformation tensor, I_1:

$$\frac{N}{N_0} = \frac{1}{f} + \left(1 - \frac{1}{f}\right)\cdot\frac{3}{I_1}, \quad I_1 = F_{ik}F_{ik}, \qquad (7)$$

where f is a scaling factor and can be specified based on the value of the density of crosslinks of the 1st network and that of the 2nd network at the free swelling state.

Stress-Stretch Relations

Inserting Eq. (6) into Eq. (4), we obtain that

$$\begin{aligned}\frac{s_{ki}}{k_BT/v} &= Nv(F_{ik} - H_{ik}) + \frac{1}{2}v(I_1 - 3 - 2\log J)\frac{\partial N}{\partial F_{ik}} \\ &\quad + \left[J\log\left(1 - \frac{1}{J}\right) + 1 + \frac{\chi}{J} - \frac{\mu}{k_BT}J\right]H_{ik}.\end{aligned} \qquad (8)$$

Recall an algebraic identity, $\partial J/\partial F_{ik} = JH_{ik}$, where H_{ik} is the transpose of the inverse of the deformation gradient $\mathbf{F}$. For simplicity, we describe the deformation of the gel in the coordinates of principal stretches. Let λ_1, λ_2, λ_3 be the principal stretches of the gel, so that $\mathbf{F} = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$, $J = \lambda_1\lambda_2\lambda_3$ and $I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$.

Free Swelling State

Submerged in the solvent-containing environment but subject to no applied forces, the gel attains a state of equilibrium, the free swelling state, characterized by an isotropic swelling ratio, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_0$. Based on Eq. (8), the relation between the principal stretch λ_0 and the chemical potential of the solvent molecules μ can be simplified as:

$$N_0 v \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_0^3} \right) + \log \left(1 - \frac{1}{\lambda_0^3} \right) + \frac{1}{\lambda_0^3} + \frac{\chi}{\lambda_0^6} = \frac{\mu}{k_B T}. \tag{9}$$

Simple Tension State

The gel is equilibrated in a solvent of chemical potential μ, and is subject to a uniaxial stress s_1 along the longitudinal direction. The state of deformation can be characterized by the longitudinal stretch λ_1 and two transverse stretch $\lambda_2 = \lambda_3$. The stresses in the transverse directions vanish, so that Eq. (8) gives

$$Nv \left(\lambda_2 - \frac{1}{\lambda_2} \right) - 3N_0 \lambda_0^2 \left(1 - \frac{1}{f} \right) \left[\lambda_1^2 + 2\lambda_2^2 - 3 - 2\log \left(\lambda_1 \lambda_2^2 \right) \right] \frac{\lambda_2}{(\lambda_1^2 + 2\lambda_2^2)^2}$$
$$+ \left[\lambda_1 \lambda_2^2 \log \left(1 - \frac{1}{\lambda_1 \lambda_2^2} \right) + 1 + \frac{\chi}{\lambda_1 \lambda_2^2} - \frac{\mu}{k_B T} \lambda_1 \lambda_2^2 \right] \frac{1}{\lambda_2} = 0. \tag{10}$$

Eq.(10) determine the transverse stretch λ_2 for a given longitudinal stretch λ_1. Eq. (8) also relates the longitudinal stress s_1 to the stretches λ_1 and λ_2 as:

$$\frac{s_1}{k_B T/v} = Nv \left(\lambda_1 - \frac{1}{\lambda_1} \right) - 3N_0 \lambda_0^2 \left(1 - \frac{1}{f} \right) \left[\lambda_1^2 + 2\lambda_2^2 - 3 - 2\log \left(\lambda_1 \lambda_2^2 \right) \right] \frac{\lambda_1}{(\lambda_1^2 + 2\lambda_2^2)^2}$$
$$+ \left[\lambda_1 \lambda_2^2 \log \left(1 - \frac{1}{\lambda_1 \lambda_2^2} \right) + 1 + \frac{\chi}{\lambda_1 \lambda_2^2} - \frac{\mu}{k_B T} \lambda_1 \lambda_2^2 \right] \frac{1}{\lambda_1}. \tag{11}$$

Parameters

In this study, we have normalized the chemical potential of the solvent molecules μ by $k_B T$, and normalized the stress s by $k_B T/v$ as shown in Eq.(8). A representative value of the volume per solvent molecule is $v = 10^{-28}\text{m}^3$. At room temperature, $kT = 4 \times 10^{-21}\text{J}$ and $kT/v = 4 \times 10^7\text{Pa}$. The elastic modulus of the dry network is $N_0 kT$. For $N_0 v = 10^{-3}$, the elastic modulus is $N_0 kT = 4 \times 10^4\text{Pa}$. On the other hand, the parameter introduced in the nonaffine model of the DN gel is specified as $f = 15$, which is same as the value presumed by our former study[2].

Results

Fig. 1 shows some experimental results[9] of the relationships between the swelling ratio of the gel with different crosslink density and the concentration of ions in the solution under free swelling condition. When the value of pH is smaller than 4 or is large than 6, the effect of the concentration of ions on the swelling ratio of the gel can be negligible. However, when the value of pH is larger than 4 and smaller than 6, the swelling ratio of the gel increases dramatically if the the concentration of ions decreases. Furthermore, the rate of such change is remarkable when the crosslink density decreases. On the other hand, the theoretical predictions of the nonaffine model are shown by red curves and a good agreement with the experimental results can be confirmed.

Fig. 2 shows some theoretical results of the simple tensional mechanical responses of the gel in the solution with different concentration of ions. When the concentration of ions is large (pH=2), yield point appears at the early deformation stage. When the concentration of ions is small, the mechanical responses of the gel is quite similar to that of rubber-like materials.
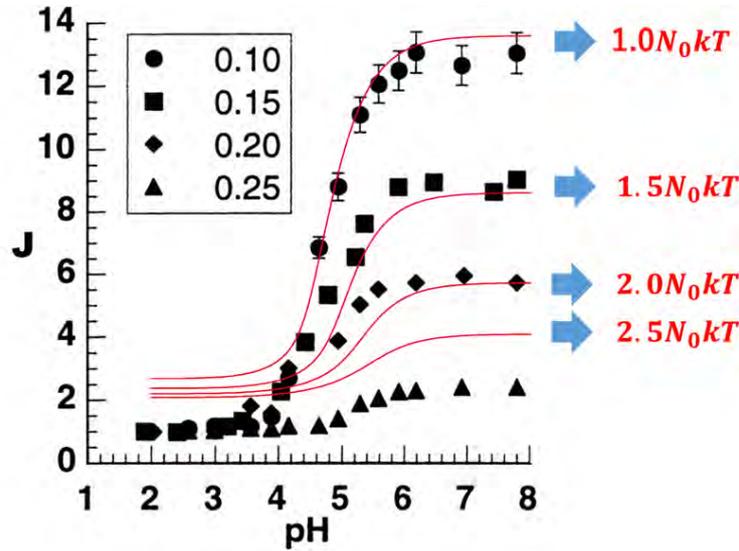
Figure 1: Comparison between theoretical predictions and experimental results of the gel inside the solution under free swelling condition
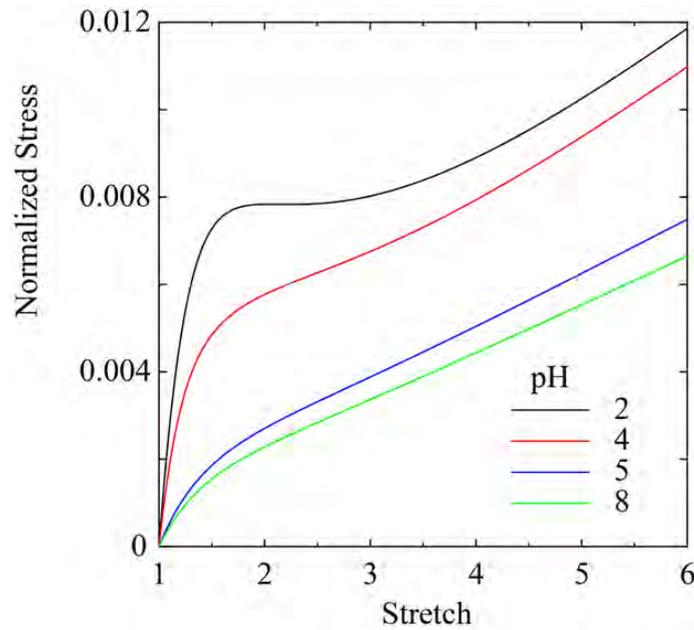
Figure 2: Theoretical predictions of the gel inside the solution under simple tension condition

Conclusions

In this study, the effect of concentration of ions in the solution on the swelling ratio and the simple tensional mechnical response of DN gel are to be investigated with the nonaffine model. The results show that the nonaffine model can be employed to reproduce the experimental results of the swelling ratio of the gel in the solution with different concentration of ions very well.

References

[1] Gong, J.P. (2010) Why are double network hydrogels so tough? *Soft Matter* **6**, 2583–2590.

[2] Riku, I. and Mimura, K. (2015) Study on mechanical model of double network hydrogel with extremely high strength. *Kobunshi Ronbunshu* **72**, 765–772. (in Japanese)

[3] Riku, I. and Mimura, K. (2019) Study on the change of entangling structure of molecular chains during the tensional and swelling process of elastomeric gel. *Proceedings of the 10th International Conference on Computational Methods*, 391–396.

[4] Riku, I. and Mimura, K. (2022) Numberical simulation of neck propagation in double network hydrogel. *Proceedings of the 15th World Congress on Computational Mechanics*, DOI：10.23967/ wccm-apcom.2022.085.

[5] Riku, I. and Mimura, K. (2010) Computational characterization on mechanical behavior of polymer electrolyte membrane based on nonaffine molecular chain network model. *The International Journal of Mechanical Sciences* **52**, 287–294.

[6] Marcombe, R., Cai, S. Q., Hong, W., Zhao, X. H., Lapusta Y. and Suo, Z. G. (2010) A theory of constrained swelling of pH-sensitive hydorgel. *Soft Matter* **6**, 784–793.

[7] Flory, P. J. and Rehner, J. (1943) Statistical mechanics of cross-linked polymer networks II. Swelling. *Journal of Applied Physics* **11**, 521–526.

[8] Hong, W., Liu, Z. S. and Suo, Z. G. (2009) Inhomogeneous swelling of a gel in equilibrium with a solvent and mechanical load. *The International Journal of Solids and Structures* **46**, 3282-3289.

[9] Eichenbaum, G. M., Kiser, P. F., Dobrynin, A. V., Simon S. A. and Needman D. (1999) Investigation of the swelling response and loading of ionic microgels with drugs and proteins: the dependence on cross-link density. *Macromolecules* **32**, 4867-4878.

Weight initialization in physics-informed neural networks to enhance consistency of mass-loss predictions of plant cells undergoing drying

*C.P. Batuwatta-Gamage¹, C.M. Rathnayaka¹,², H.C.P. Karunasena¹,³, M.A. Karim¹, †YuanTong Gu¹

¹ School of Mechanical, Medical and Process Engineering, Faculty of Engineering, Queensland University of Technology (QUT), Australia.
² School of Science, Technology and Engineering, University of the Sunshine Coast (UniSC), Australia.
³ Dept. of Mechanical and Manufacturing Engineering, Faculty of Engineering, University of Ruhuna, Sri Lanka.

*Presenting author: chanaka.gamage@hdr.qut.edu.au
†Corresponding author: yuantong.gu@qut.edu.au

Abstract

This study aimed to highlight the significance of weight initialization towards the consistency of Physics-Informed-Neural-Network-based (PINN-based) predictions for spatiotemporal problems in engineering and science. Accordingly, here a PINN was developed for mass transfer analysis (i.e., PINN-MT) for a single plant cell undergoing drying. While solving Fick's laws of diffusion for a cell domain and predicting mass loss and moisture concentration based on convective mass transfer at the cell wall boundary, PINN-MT utilizes moisture concentration at fresh state of the cell as an initial condition. The governing equations, boundary conditions, and initial conditions were incorporated to the corresponding PINN through the loss function. Residuals of these equations and initial-and-boundary conditions were minimized during the training process of PINN, which can predict moisture concentration variations in time and space domains. However, spatiotemporal problems typically involve many tunable hyperparameters that can make the training process more complicated, leading to inconsistent predictions and loss-convergence problems. This is uncommon in the context of traditional computational approaches. To address this complexity associated with PINNs, pre-trained weight initialization can be adopted, enhancing the ability of PINN-MT to provide consistent solutions via automatic differentiation. In this context, this study assessed the effectiveness and efficiency of PINN-MT coupled with weight initialization to address training complexities and to provide consistent solutions for spatiotemporal problems in engineering and science.

Keywords: Physics informed neural networks, Weight initialization, Mass transfer, Plant cells, Food drying

Introduction

Computational modeling is a powerful tool for studying and predicting plant-based food drying behaviors. By considering the complex physical and chemical processes involved in drying, computational models can optimize drying conditions, improve product quality, and reduce energy consumption [1, 2]. The application of computational modeling spans from the microscale to the industrial scale, including macro and bulk scales. Multiscale modeling, another computational modeling concept, merges models at different scales of resolution, enabling a comprehensive and accurate understanding of complex systems such as plant-based food structures [3, 4].

Understanding the microstructural behaviors of complex plant-based foods, given their heterogeneous, hygroscopic, and porous nature, is essential in optimizing the drying process. However, these investigations are not easily achievable only through experimental attempts

due to method-specific limitations and higher cost for infrastructure. To approximate associated characteristic variations, various constitutive models based on partial differential equations (PDEs) have been developed. Moreover, numerical differentiation using both mesh-based and mesh-free computational tools have been employed under comparatively complex physical conditions [1, 2, 5]. Nevertheless, due to method-specific challenges, solving PDEs using these traditional computational approaches has proven difficult for investigating non-linear and complex microstructural changes [6, 7]. Recently, neural network-based automatic differentiation (AD) has emerged as a new computational technique paradigm [8]. In this context, a novel concept known as the Physics-Informed Neural Network (PINN) has shown promising results in various applications [9-13].

PINN is an innovative scientific machine learning concept that is capable of approximating solutions with no involvement of mesh or particles. Governing equations (Eqn), boundary conditions (BC), and initial conditions (IC) are incorporated into the loss function as residuals, which guide the neural network (NN) to adhere to these equations and conditions. The meshless nature of PINN, along with its ability to solve multi-domain and inverse problems with minimal effort and modification, has catalyzed the development of the concept, particularly in areas where traditional computational approaches face difficulties due to ill-posed, high-dimensional conditions and the demand for expensive computational resources [9]. Despite the achieved success, PINNs have not yet surpassed finite element analysis in terms of computational time and still face challenges when training the neural network for multi-loss functions within multi-task learning. Specifically, PINN often generates inconsistent predictions and encounters loss-convergence problems due to its approximation-based iterative approach. These inconsistencies pose a challenge to the repeatability of the solution approach with PINN, compared to traditional methods. The main reason for this is the initial random weights of the neural network, which significantly impact the final solution. PINNs are trained using a stochastic optimization algorithm, rendering the training process non-deterministic [14]. As a result, the final optimized weights and biases of the neural network can vary each time the model is trained. This becomes particularly noticeable when solving highly non-linear and dynamic problems, such as plant-based food drying [9]. Consequently, this poses a significant challenge when applying PINN to solve real-world applications such as soft matters and biomechanics. In data-science applications, various initialization approaches apart from random initialization such as constant initialization, data-driven initialization and hybrid initialization can be involved with their slight variations [15]. However, most of these methods have been developed for training NNs with labeled data, which are not entirely suitable for the computational analysis required by PINNs. In this context, this study aimed to investigate the issues associated with random initialization and highlight the importance of weight initialization with nearly optimized weights and biases. Accordingly, a series of simulations have been conducted to investigate the consistency of time-domain mass loss and related spatial domain moisture concentration attributes in a single plant cell undergoing convective drying.

Methodology

Fick's laws of diffusion are fundamental to understand and model mass transfer effects. The laws describe how particles move from areas of high concentration to areas of lower concentration in spatiotemporal domain. By incorporating Fick's laws of diffusion into models and simulations, scientists and engineers can analyze and optimize processes involving mass transfer. It provides a solid foundation for developing physics-based models along with computational tools to investigate practical applications, including drug delivery systems, environmental pollution, material science, and biological tissue engineering.

2-D representation of single cell

In this study, the 2-dimensional (2-D) cell geometry has been approximated by considering a circular surface representing a cylindrical shape in 3D as illustrated in Figure 1 [16].

(a) 3D cell geometry   (b) 2D cell domain
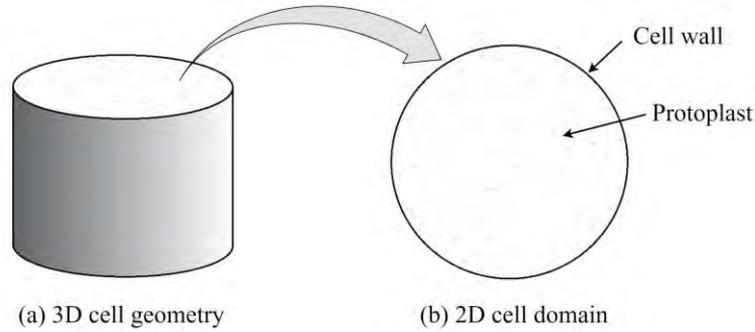
Figure 1: 2D cell domain approximation form 3D cylindrical shape cell

If the top surface of the cylinder is the domain for 2D analysis, the spatio-temporal domain for the single plant cell in the 2D domain can be defined using x and y coordinates in meters, along with time t in seconds. Fick's laws of diffusion can be expressed as follows to predict the time-varying mass loss and spatial variability of moisture concentration, considering the IC and BC [9].

$$\text{Eqn:} \quad \frac{\partial c}{\partial t} = D.\left(\frac{\partial^2 c}{\partial x^2}, \frac{\partial^2 c}{\partial y^2} \right)$$

$$\text{BC:} \quad D.\left(\frac{\partial c}{\partial x}, \frac{\partial c}{\partial y} \right)\bigg|_{un} - \left(h_m (c - c_{Out}) \right) \tag{1}$$

$$\text{IC:} \quad \left(c_0\big|_{t=0} \right) - \left(\frac{M_{wb,T,i} \cdot \rho_i}{M_w} \times P_{cell,i} \right)$$

The moisture concentration inside the cell and outside the cell are defined as c and c_{out} respectively. The spatial coordinates are defined as x and y, and the mass transfer coefficient is defined as h_m. The mass flux variations normal to the boundary are represented by un in the boundary condition equation. In the initial condition equation, the initial tissue wet basis moisture content is defined as $M_{wb,T,i}$, the initial material density is defined as ρ_i, the molar mass of water is defined as M_w and the initial percentage of intercellular water content is defined as $P_{cell,i}$.

Residuals of the above-mentioned equations and conditions have been integrated to a feed forward neural network to develop a PINN for mass transfer analysis in drying plant cells (PINN-MT), as shown below.
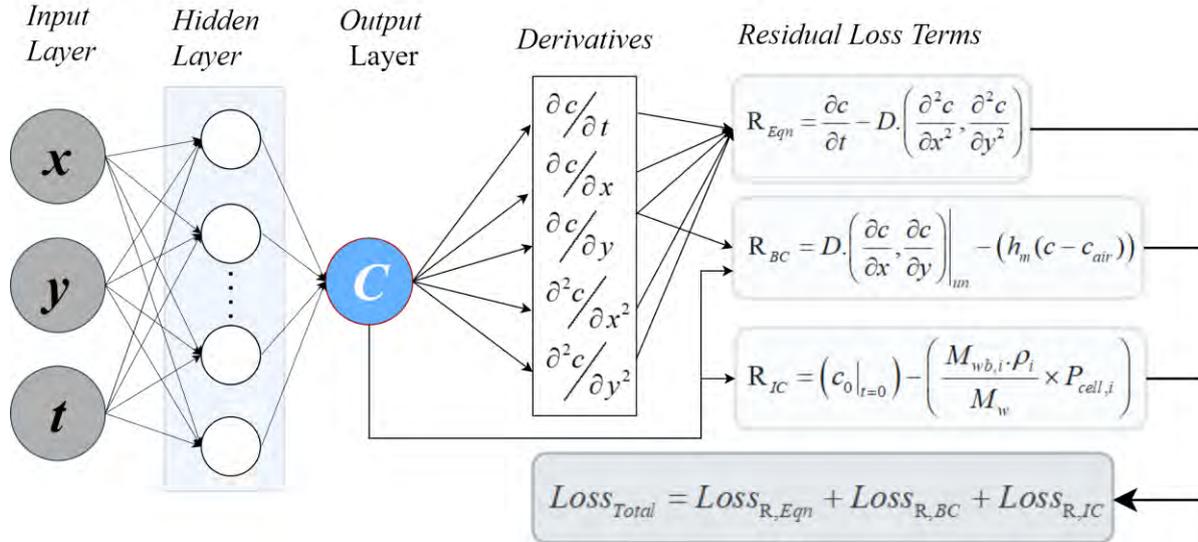
Figure 2: PINN-MT architecture with x, y, and t as independent inputs and c as the dependent output

The PINN-MT for this analysis has been implemented using a radial basis activation function, consisting of a single hidden layer with 750 nodes as in Figure 2. The mass transfer characteristics have been investigated for an apple parenchyma cell and the mass transfer coefficient, denoted as h_m, has been calculated to ensure significant mass loss within a simulation drying time of 10 seconds. To solve the mass transfer model, the PINN-MT requires training to minimize the residual loss terms by optimizing the neural network parameters. The specific parameters utilized in this study are presented in Table 1. For readers interested in more detailed information, further elaboration can be found in [9].

Table 1: Parameters used for mass transfer and moisture concentration investigations [9]

Parameter	Value
Cell diameter (d)	150 μm
Diffusion coefficient (D)	4.19 x 10^{-10} (m^2 s^{-1})
Convective mass transfer coefficient (h_m)	3 x 10^{-6} (m s^{-1})
Initial tissue dry basis mass ($M_{wb,T,i}$)	6.80 (kg kg^{-1})
Molar mass of water (M_w)	0.018016 (kg mol^{-1})
Universal gas constant (R)	8.314 (J $mol^{-1}K^{-1}$))
Initial density of apple (ρ_i)	837 (kg m^{-3})
Initial moisture concentration (c_0)	39829 (mol m^{-3})
Moisture concentration in air (c_{out})	0.654 (mol m^{-3})

Results and Discussions

The detailed development and performance evaluation of the PINN-MT model for solving mass transfer attributes have been extensively examined and compared against well-established finite element approaches [9]. Building upon this previous work, this study specifically explored the impact of weight initialization on the training convergence and prediction consistency within the PINN-MT framework. Initial investigations primarily concentrated on the challenges associated with random weight initialization. Subsequently,

the effectiveness of weight initialization from pre-trained models through transfer learning is presented in conjunction with the PINN-MT investigations.

Case Study 01: PINN-MT with Xavier random initialization

Xavier Initialization is a widely used method for initializing the weights of neural networks. It involves randomly assigning values to the weights based on the input and output dimensions of the weight matrices. The random nature of Xavier initialization ensures that each weight in the network starts with a different value. This randomness is beneficial most of the time as it allows the network to explore a broader range of parameter configurations during the training process. It helps to prevent the network from getting stuck in symmetric states and encourages the learning of diverse representations [15].

Here, the PINN-MT model was trained independently for three times using Xavier random initialization. The aim was to investigate the problem of consistency and convergence in the predictions of time-domain dry-basis moisture mass loss. The results of these three training runs are depicted in Figure 3, which display the predicted mass loss over time according to the PINN-MT model. By examining the graph, the consistency and convergence of predictions of the model can be analyzed and evaluated.
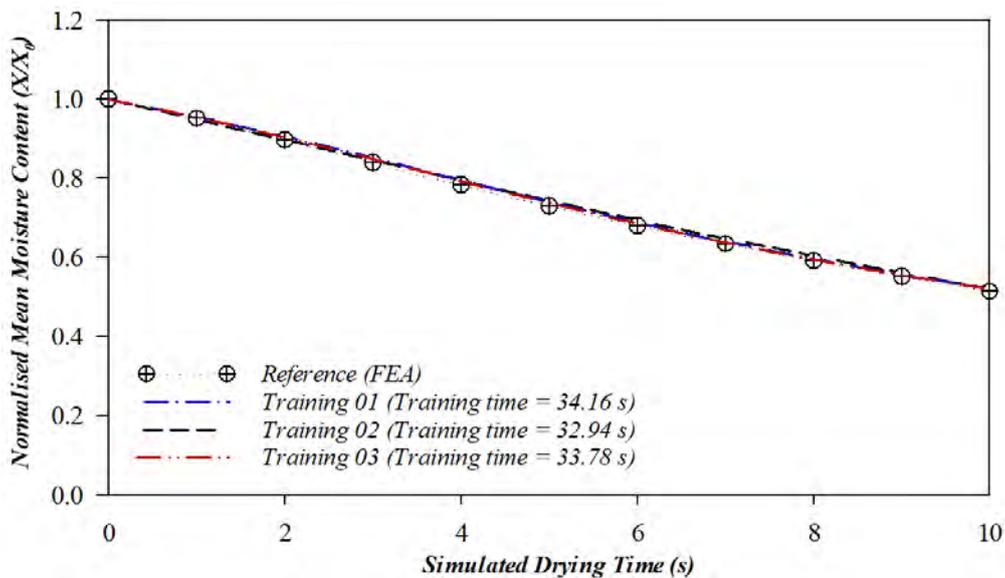
Figure 3: Dry-basis moisture mass variation predictions against simulated drying time

If all the parameters in the neural network, such as the activation function, learning rate, and number of nodes per hidden layer, have been appropriately set, the PINN-MT model can provide qualitative predictions of variations in mass loss, as depicted in Figure 3. However, to meet the requirements of scientific and engineering solutions, it is essential to assess the consistency and accuracy of the model's predictions. This can be evaluated by comparing the dry-basis moisture mass values predicted by the PINN-MT model for each training and further extending the comparison to benchmark results obtained from finite element analysis (FEA). By quantitatively evaluating the performance of the model against the FEA benchmarks (refer to Table 2), a more comprehensive understanding of its capabilities and limitations can be obtained. This analysis facilitates a more reliable assessment of the ability of the model to provide consistent and accurate predictions in practical applications.

Table 2: The normalized mean moisture content calculated from three independent trainings of PINN-MT, compared to FEA reference values.

Simulation Type	Iteration	Normalized mean moisture content predictions in each timestep									
		1 s	2 s	3 s	4 s	5 s	6 s	7 s	8 s	9 s	10 s
Reference:											
FEA	-	0.952	0.896	0.838	0.782	0.729	0.680	0.634	0.590	0.550	0.513
PINN-MT:											
Training 01	5000	0.949	0.898	0.846	0.794	0.744	0.695	0.648	0.603	0.560	0.520
Training 02	5000	0.956	0.906	0.851	0.795	0.741	0.689	0.640	0.596	0.556	0.520
Training 03	5000	0.955	0.903	0.848	0.793	0.738	0.686	0.637	0.593	0.554	0.520

The study also includes an evaluation and qualitative comparison of the spatiotemporal-dependent moisture concentration variations within the cell. The predicted moisture concentration profiles inside the cell are analyzed and compared, providing insights into how the concentration changes over time and space. These qualitative comparisons are depicted in Figure 4. By examining these figures, a visual understanding of how the moisture concentration evolves within the cell can be obtained.
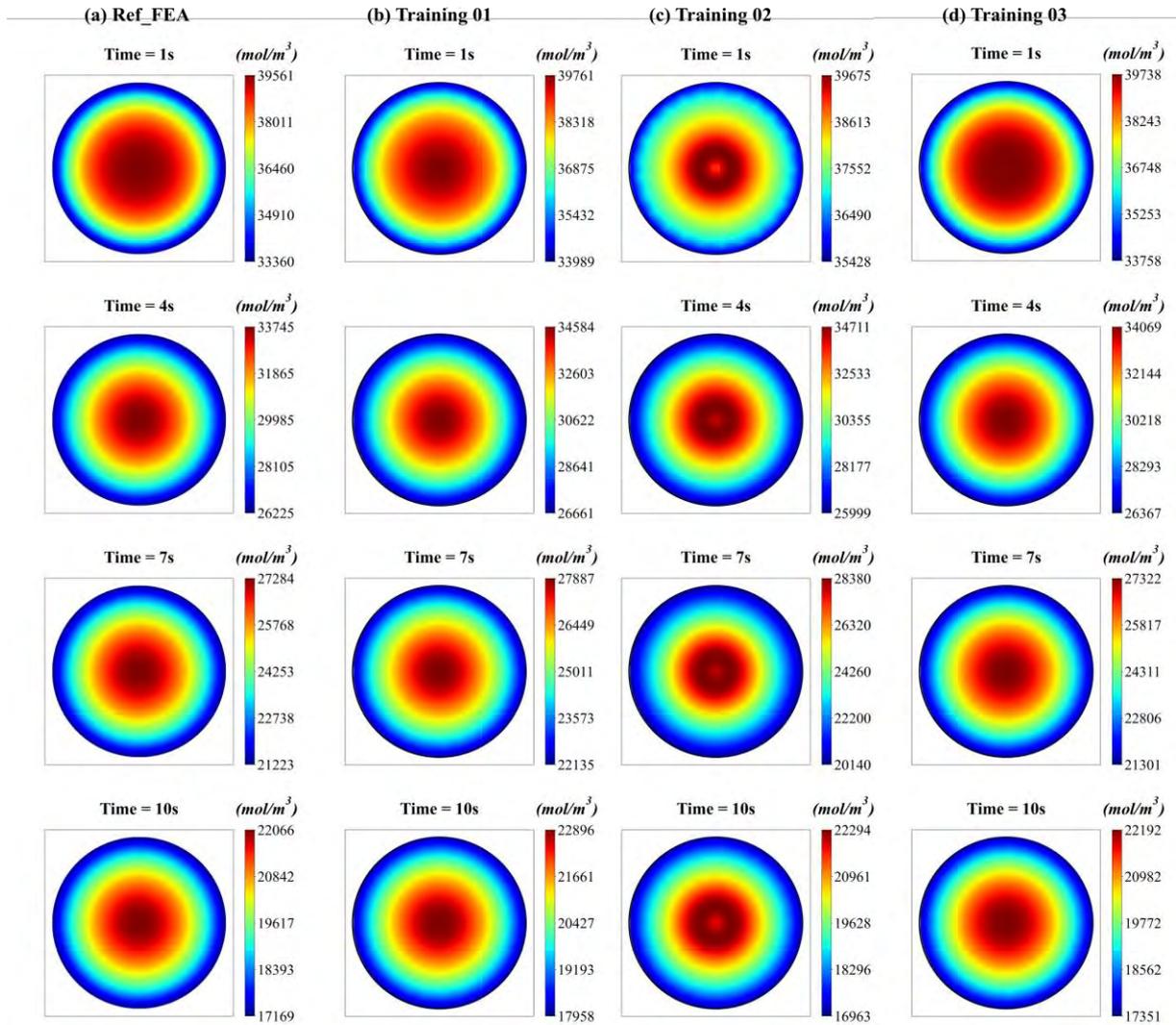
Figure 4: Qualitative variations of moisture concentration predictions inside the cell for selected drying times: (a) Benchmark reference results from FEA, (b) Training 01 results, (c) Training 02 results, and (d) Training 03 results.

For a clearer understanding of the prediction variations quantitatively, the maximum and minimum moisture concentration values inside the cell at each time step have been compared in Table 3. This comparison allows for a quick assessment of the range and extremities of moisture concentration within the cell over time, providing insights into the temporal variations and trends during the drying process.

Table 3: The maximum and minimum moisture concentration values from three independent trainings of PINN-MT, compared to FEA reference values.

Simulation Type		Max and min moisture concentration predictions in each timestep									
		1 s	2 s	3 s	4 s	5 s	6 s	7 s	8 s	9 s	10 s
Reference: FEA	Max	39561	38175	36082	33745	31447	29289	27284	25418	23683	22066
	Min	33360	30477	28194	26225	24436	22774	21223	19776	18426	17169
PINN-MT:											
Training 01	Max	39761	38585	36774	34584	32260	29990	27887	26003	23345	22896
	Min	33989	30830	28492	26661	25080	23597	22135	20690	19286	17958
Training 02	Max	39675	38273	36599	34711	32669	30538	28380	26256	24215	22294
	Min	35428	31943	28772	25999	23654	21718	20140	18885	17811	16963
Training 03	Max	39738	38384	36372	34069	31707	29432	27322	25407	23697	22192
	Min	33758	30771	28400	26367	24545	22863	21301	19858	18542	17351

As observed from the results presented in Table 2 (normalized mean moisture content) and Table 3 (maximum and minimum moisture concentration), it is clear that the quantitative values obtained from three independent trainings of the PINN-MT model differ from each other. Furthermore, certain random initializations have resulted in convergence problems, leading to notable discrepancies, as evidenced by the results of Training 02 (refer to Figure 4-c). Specifically, the moisture concentration at the center of the cell exhibits significant variations compared to the reference FEA results. These prediction discrepancies primarily stem from the random initialization of the neural network at the beginning of training. While constant initialization may seem like a potential solution to address these issues and avoid the randomness associated with weight initialization, determining the optimal constant starting weight and bias values for a given PINN is not a straightforward task.

Case Study 02: PINN-MT with nearly optimized weight initialisation

To address the issue mentioned, the most suitable approach is to initialize the weights from a pre-trained same PINN model, where the weights and biases are nearly optimized for the specific loss terms based on the integrated physics equations and conditions. This can be easily accomplished in PyTorch by saving and loading the state dictionary, which maps each parameter of the model to its corresponding PyTorch tensor. By adopting this technique, the PINN-MT model can benefit from the knowledge and patterns acquired during the pre-training phase, resulting in improved consistency and convergence of predictions. In particular, NN parameters (weights and biases) from the Training attempt 01 from the above study have been saved and initialized to the PINN-MT to train the PINN-MT with nearly optimized NN parameters. The relevant qualitative values of normalized mean moisture content and the maximum and minimum moisture concentration values are tabulated in Table 01 and Table 06 respectively.

To this end, the parameters (weights and biases) of the neural network from Training attempt 01 were saved and re-used to initialize the PINN-MT model when training for the same problem the second time. This approach allowed the PINN-MT model to be trained with nearly optimized neural network parameters. The consistency and the convergence of resulting predictions were evaluated, focusing on the quantitative values of the normalized mean moisture content as in Table 4 and both the maximum and minimum moisture concentration values in

Table 5.

Table 4: The normalized mean moisture content calculated from three trainings of PINN-MT with nearly optimized weight initialization, compared to FEA reference values.

Simulation Type	Iteration	Normalized mean moisture content predictions in each timestep									
		1 s	2 s	3 s	4 s	5 s	6 s	7 s	8 s	9 s	10 s
Reference:											
FEA	-	0.952	0.896	0.838	0.782	0.729	0.680	0.634	0.590	0.550	0.513
PINN-MT:											
Training 01	1000	0.955	0.901	0.842	0.786	0.733	0.682	0.635	0.591	0.551	0.514
Training 02	1000	0.955	0.901	0.842	0.786	0.733	0.682	0.635	0.591	0.551	0.514
Training 03	1000	0.955	0.901	0.842	0.786	0.733	0.682	0.635	0.591	0.551	0.514

Table 5: The maximum and minimum moisture concentration values from three trainings of PINN-MT with nearly optimized weight initialization, compared to FEA reference values.

Simulation Type		Max and min moisture concentration predictions in each timestep									
		1 s	2 s	3 s	4 s	5 s	6 s	7 s	8 s	9 s	10 s
Reference: FEA	Max	39561	38175	36082	33745	31447	29289	27284	25418	23683	22066
	Min	33360	30477	28194	26225	24436	22774	21223	19776	18426	17169
PINN-MT:											
Training 01	Max	39557	38074	36079	33785	31481	29324	27339	25494	23753	22094
	Min	33291	30491	28190	26187	24373	22700	21151	19720	18406	17206
Training 02	Max	39557	38074	36079	33785	31481	29324	27339	25494	23753	22094
	Min	33291	30491	28190	26187	24373	22700	21151	19720	18406	17206
Training 03	Max	39557	38074	36079	33785	31481	29324	27339	25494	23753	22094
	Min	33291	30491	28190	26187	24373	22700	21151	19720	18406	17206

The results obtained from initializing the training with nearly optimized NN parameters demonstrate a very higher consistency in the predicted normalized mean moisture content and the maximum and minimum moisture concentration values. Therefore, it is evident that this approach ensures reliable convergence during training and provides consistent predictions. Moreover, this approach ensures reliable fast convergence during training and provides consistent predictions with at least 80% reduced computational time.

In addition to that, this nearly optimized weight initialization approach proves to be highly beneficial when applying the PINN model to solve physics-based models under different conditions or subjected to varying parameters. By transferring the nearly optimized weights and biases from a model developed for one plant-based food cell (such as apple) to other plant-based cells (such as potato and carrot), the convergence of the results can be significantly improved. This leads to substantial reductions in training time, with the potential for up to 80%-time savings. Further details and investigations on this approach can be found in [9].

Conclusion and Outlook

In conclusion, this study has highlighted the challenges of inconsistency and loss-convergence problems in Physics Informed Neural Network (PINN) predictions when solving physics-based models without label data. The random weight initialization of feed forward neural networks contributes to these problems, leading to varying predictions with each training attempt. In contrast, FEA does not exhibit such inconsistencies.

To address these problems, the effectiveness of weight initialization from a pre-trained model has been investigated using a Fick's laws diffusion-based mass transfer model coupled with a deep neural network (PINN-MT). While Xavier random initialization can yield qualitative predictions close to benchmark FEA results, different training attempts produce varying quantitative values. To overcome these challenges, the effectiveness of weight initialization from a pre-trained model has been investigated. By utilizing nearly optimized weights and biases, the PINN training process becomes more consistent and converges reliably. This approach also serves as an effective regularization method for PINN-based models, enabling the application of a developed PINN model with known physics to different materials or conditions. The key advantage of this approach is the reduction in computational costs and inconsistencies.

In summary, the weight initialization from pre-trained model through transfer learning steps can be summarized as below.

1. **Pre-train the developed PINN model**: Start by training a PINN model on a related physics problem with random initialization. This pre-training step allows the network to learn general features and representations of the underlying physics.

2. **Save weights and biases matrix of the pre-trained model:** Once the model is trained initially, extract the learned weights and biases of the network. These weights represent the knowledge and patterns learned from the pre-training phase. However, make sure at least the model results closely aligned with expected results.

3. **For the target PINN:**

   a. For the same application: Use transfer learning to initialize the saved weights and biases matrix when training PINN for the second time.
   b. For different similar applications: Develop the PINN with same size neural network with the similar integrated physics and conditions. Use transfer learning to initialize the saved weights and biases matrix in the previous steps when training PINN for this new similar application.

   This step ensures that the model starts with knowledge already acquired during pre-training.

4. **Fine-tune the target PINN:** Train the target PINN by adjusting the learning rate and other hyperparameters as needed for the target problem until convergence, probably with less epochs.

5. **Evaluate and iterate:** Evaluate the performance of the fine-tuned PINN on the targeted problem's validation. If the results are satisfactory, the model is ready for predictions. Otherwise, repeat steps 3 to 6, adjusting the architecture or hyperparameters as needed, until achieve the desired performance.

Acknowledgement

References

1. Rathnayaka Mudiyanselage, C.M., et al., *Novel trends in numerical modelling of plant food tissues and their morphological changes during drying – A review.* Journal of Food Engineering, 2017. **194**: p. 24-39.
2. Datta, A.K., *Toward computer-aided food engineering: Mechanistic frameworks for evolution of product, quality and safety during processing.* Journal of Food Engineering, 2016. **176**: p. 9-27.
3. Welsh, Z., et al., *Multiscale Modeling for Food Drying: State of the Art.* Compr Rev Food Sci Food Saf, 2018. **17**(5): p. 1293-1308.
4. Wijerathne, W., et al., *A coarse-grained multiscale model to simulate morphological changes of food-plant tissues undergoing drying.* Soft Matter, 2019. **15**(5): p. 901-916.

5.   Khan, M.I.H., et al., *Fundamental Understanding of Heat and Mass Transfer Processes for Physics-Informed Machine Learning-Based Drying Modelling.* Energies, 2022. **15**(24).

6.   Batuwatta-Gamage, C.P., et al., *A physics-informed neural network-based surrogate framework to predict moisture concentration and shrinkage of a plant cell during drying.* Journal of Food Engineering, 2022. **332**.

7.   Cuomo, S., et al., *Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next.* Journal of Scientific Computing, 2022. **92**(3): p. 88.

8.   Raissi, M., P. Perdikaris, and G.E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations.* Journal of Computational Physics, 2019. **378**: p. 686-707.

9.   Batuwatta-Gamage, C.P., et al., *A novel physics-informed neural networks approach (PINN-MT) to solve mass transfer in plant cells during drying.* Biosystems Engineering, 2023. **230**: p. 219-241.

10.   Jeong, H., et al., *A Physics-Informed Neural Network-based Topology Optimization (PINNTO) framework for structural optimization.* Engineering Structures, 2023. **278**: p. 115484.

11.   Bai, J., et al., *An Introduction to Programming Physics-Informed Neural Network-Based Computational Solid Mechanics.* International Journal of Computational Methods. **0**(0): p. 2350013.

12.   Raissi, M., A. Yazdani, and G.E. Karniadakis, *Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations.* Science, 2020. **367**(6481): p. 1026-1030.

13.   Karniadakis, G.E., et al., *Physics-informed machine learning.* Nature Reviews Physics, 2021. **3**(6): p. 422-440.

14.   Markidis, S., *The Old and the New: Can Physics-Informed Deep-Learning Replace Traditional Linear Solvers?* Front Big Data, 2021. **4**: p. 669097.

15.   Narkhede, M.V., P.P. Bartakke, and M.S. Sutaone, *A review on weight initialization strategies for neural networks.* Artificial Intelligence Review, 2022. **55**(1): p. 291-322.

16.   Karunasena, H.C.P., et al., *Simulation of plant cell shrinkage during drying – A SPH–DEM approach.* Engineering Analysis with Boundary Elements, 2014. **44**: p. 1-18.

The static behaviors study of 2D structures with cell-based smoothed isogeometric analysis

*Liming Zhou[1,2], Jinwen Geng[1,2], Pengxu Chen [1,2], He Zhu [1,2], and Zhiqiang Gao [1,2]

[1] Key Laboratory of CNC Equipment Reliability, Ministry of Education, Jilin University, No. 5988 Renmin Street, Changchun, 130025, P. R. China

[2]School of Mechanical and Aerospace Engineering, Jilin University, No. 5988 Renmin Street, Changchun, 130025, P. R. China

*Corresponding author: lmzhou@jlu.edu.cn

Abstract

In order to further improve the precision and accuracy of engineering structures analysis, Cell-based smoothed isogeometric analysis (CS-IGA) is suggested as a solution to the problems of mathematical modeling and displacement solving. This method is based on strain smoothing technology used to isogeometric analysis (IGA). Therefore, the value of the integral points is only needed to obtain without corresponding derivative when calculating the strain matrix. It greatly improves the accuracy and efficiency of CS-IGA. In addition, the NURBS curves are used in CS-IGA to build mathematical models, which can accurately describe complex geometric models and improve the precision of design analysis. This modeling method makes it possible for seamlessly integrating CAD and CAE, which avoids the grid processing work in engineering problems and reduces the cost of the design. Compared with other traditional numerical methods, CS-IGA has higher accuracy, convergence and computational efficiency. It will be verified by the examples in this paper. Overall, this paper can give a hand for analyzing complex 2D structures.

Keywords: Cell-based smoothed isogeometric analysis, Strain smoothing technology, NURBS curves

1.Introduction

With the continuous deepening of the fourth industrial revolution, the number of engineering projects are increasing and engineering problems have become the focus of contemporary scholars. As the foundation of engineering analysis, structural design and analysis has been paid more and more attention by scholars. At present, the mechanical characteristics of the structure are analyzed by three methods: analytical method, experimental method and numerical analysis. The experimental method and the analytical method have some limitations in accuracy and efficiency when analyzing the structure with complex boundary. However, with the rapid development of computer technology, numerical analysis has greatly improved the calculation accuracy and efficiency and has become the first choice for scholars to analyze complex structural problems.

UP to now, finite element method (FEM), as the most mature mathematical analysis method, which is widely used in structural analysis problems. Pinnola et al. [1] proposed a new two-node FEM to approximate bending problems of elastic beams driven by different stresses. Experiments and numerical examples are used to confirm precision of the method. Smoljanovic et al. [2] proposed a mathematical model of a beam structure with a mixture of finite-discrete

elements and verified the accuracy of solving the beam-plate structure through numerical examples. Wei et al. [3] used Hermite curve to approach the solution of a nonlinear Euler beam. This method is verified the correctness of the solution through numerical examples and gives the corresponding error. Hosseinian et al. [4] proposed a rectangular finite element formula for solving thin plates. The shape function of the formula is obtained by multiplying the shape function of the element and the Euler beam, which can accurately analyze the dynamics and static mechanics of the beam and plate structure. Patuelli et al. [5] established a two-node beam element model to calculate and analyze the dynamics of beam with flexural and torsion combination. The feasibility of the method is verified by experiments and data simulation. Al-Zahrani et al. [6] solved the free vibration problem of axial and bidirectional functionally gradient beams based on the mixing law of Vogit model. The method also analyzes the effect of material ratio on natural frequency under axial, bidirectional and boundary conditions.

Meshless method is a numerical analysis method widely used in computer simulation. Compared with FEM, meshless method can be more flexible in constructing the grid according to the needs of physical problems. In engineering problems, this method can effectively avoid the calculation error caused by mesh distortion and improve the precision of calculating complex structures. Common meshless methods include element-free Galerkin method [7]-[11], material point method [12]-[16], Radial basis function [17]-[21], smoothed finite element method [22]-[26]. Jarak et al. [27] derived the meshless Petrov-Galerkin method and solved the linear high-order differential equation, which provided the formula basis for analyzing plate structures. Gao et al. [28] derived the grid-free manifold method by combining least square method with finite covering technique. This method can solve dynamic problems and the accuracy of the method is verified. Chien et al. [29] derived the cell-based smoothed finite element method to reduce the influence of mesh distortion and analyze the vibration problem of shell structures. Li et al. [30] proposed an edge-based smoothed tetrahedral finite element method to solve the thermodynamic problems of three-dimensional structures and verified the precision through numerical examples.

For further improving the accuracy and efficiency of engineering structure analysis, Hughes et al. [31]-[32] proposed a new numerical method - isogeometric analysis (IGA). This method is based on NURBS curve to build the geometric model of the object. This enables CAD and CAE techniques to be seamlessly integrated with high precision and efficiency for analyzing complex geometric models. Due to the property of NURBS basis function, the special treatment of the boundary conditions is required [33]-[35]. Lee et al. [36] used IGA to establish rod elements and solve free vibration problems of elastic rods. Milic et al. [37] proposed an analytical method combining the Kirchhoff-Love shell formula with NURBS basis function to obtain the solution of the shell structures. Fahrendorf et al. [38]] proposed an isogeometric method applied to mixed stress-displacement to analyze the structure of elastoplastic materials. The precision of the method is verified by numerical examples. Pekovic et al. [39] presented an isogeometric plate finite element formula based on Reddy third-order shear deformation theory to solve the dynamic problems of composite structures. Lee et al. [40] proposed an isogeometric shell element based on the Reissner-Mindlin shell theory and analyzed the vibration problem of shell structures. However, the traditional Gauss integration is applied to IGA method, it will produce

some errors for the analysis of complex structures. Therefore, the strain smoothing technology in the meshless method is extended to IGA in this paper. The cell-based smoothed isogeometric analysis (CS-IGA) is proposed to solve structural problems in engineering. It can calculate the strain matrix by the basis function values of the midpoint on the subdomain boundary. It can greatly improve the calculation accuracy and efficiency.

In this paper, CS-IGA is mainly used to analyze 2D beam and plate structure problems. In chapter 2, the basic equations and boundary conditions are derived. The basic theory of CS-IGA and the strain smoothing technology are given in this chapter. In chapter 3, the superiority of CS-IGA is verified by numerical results. In chapter 4, the conclusion of this paper is given

2. Basic theory

2.1. Basic equations

In order to solve the problem of deformation of two-dimensional isotropic materials under different loads, it mathematically boils down to the calculation of mechanical parameters that satisfy the basic equations.

The equilibrium equation can be expressed:

$$\sigma_{ij,j} + f_i = 0 \tag{1}$$

where σ is Cauchy stress; f_i is the component of body force along the axis. The linear constitutive equations can be presented as:

$$\sigma_{ij} = C_{ijkl} S_{kl} \tag{2}$$

$$[\mathbf{C}] = \frac{E(1-v)}{(1+v)(1-2v)} \begin{bmatrix} 1 & \dfrac{v}{1-v} & \dfrac{v}{1-v} & 0 & 0 & 0 \\[2ex] \dfrac{v}{1-v} & 1 & \dfrac{v}{1-v} & 0 & 0 & 0 \\[2ex] \dfrac{v}{1-v} & \dfrac{v}{1-v} & 1 & 0 & 0 & 0 \\[2ex] 0 & 0 & 0 & \dfrac{1-2v}{2(2-v)} & 0 & 0 \\[2ex] 0 & 0 & 0 & 0 & \dfrac{1-2v}{2(2-v)} & 0 \\[2ex] 0 & 0 & 0 & 0 & 0 & \dfrac{1-2v}{2(2-v)} \end{bmatrix} \tag{3}$$

where C_{ijkl} is elastic constant; S_{kl} is strain constant; E is elasticity modulus; v is poisson ratio. The geometric equations have the following form:

$$S_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right), (i, j=1, 2, 3) \tag{4}$$

where u_i is the component of displacement.

2.2. Boundary conditions
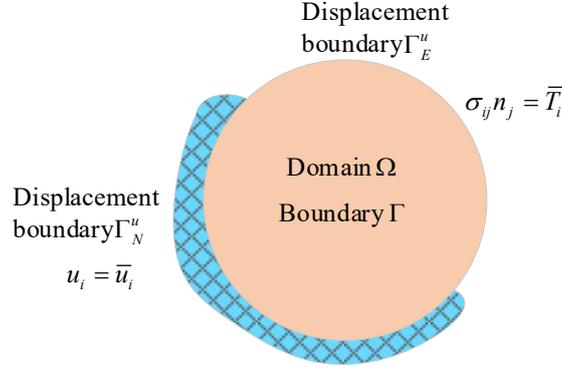
Figure 1. Boundary conditions of elastic structures

For boundary conditions, the global boundary of the 2D structure is Γ. The natural and essential boundary conditions can be respectively expressed as:

$$u_i = \bar{u}_i \text{, at the boundary } \Gamma_N^u \tag{5}$$

$$\sigma_{ij} n_j = \bar{T}_i \text{, at the boundary } \Gamma_E^u \tag{6}$$

where $\bar{u}_i$ represents the displacement; Γ_N^u represents the nature boundary; Γ_E^u represents the essential boundary; n_j is the component of the outer normal vector. It is noted that elastic structures satisfy the condition $\Gamma = \Gamma_N^u \cup \Gamma_E^u$, $\Gamma_N^u \cap \Gamma_E^u = \varnothing$. The specific boundary situation is shown in Figure 1.

2.3. Cell-based smoothed isogeometric analysis

This method is based on the construction of NURBS. The basis function of NURBS can be obtained by dealing with the B-spline basis function. The knot vector of a B-spline of degree m is defined as $\xi_i = [\xi_1, \cdots, \xi_{N+m+1}] \in \Re$, the number of repetitions for both the beginning and ending knots is m. For any control point A with degree m, the basis function of B-spline can be represented as:

$$N_{A,0}(\xi) = \begin{cases} 1, \text{if } \xi_A \le \xi < \xi_{A+1} \\ 0, \text{otherwise} \end{cases} \tag{7}$$

$$N_{A,m}(\xi) = \frac{\xi - \xi_A}{\xi_{A+m} - \xi_A} N_{A,m-1}(\xi) + \frac{\xi_{A+m+1} - \xi}{\xi_{A+m+1} - \xi_{A+1}} N_{A+1,m-1}(\xi) \tag{8}$$

The NURBS basis function of degree m can be expressed as:

$$R_{A,m}(\xi) = \frac{\omega_A N_{A,m}(\xi)}{\sum_B \omega_B N_{B,m}(\xi)} \tag{9}$$

where ω_A represents the weight of control point A. The NURBS basis function satisfies the following equation:

$$\sum_{A=1}^{N} R_{A,m}(\xi) = 1 \tag{10}$$

In one-dimensional space, The NURBS curve can be expressed as:

$$C(\xi) = \sum_{A=1}^{N} R_{A,m}(\xi) P_A \tag{11}$$

where P_A is the of control point. Similarly, the two-dimensional NURBS surface can be expressed as:

$$S(\xi,\eta) = \sum_{A=1}^{N} \sum_{B=1}^{M} R_{AB,\text{mn}}(\xi,\eta) P_{AB} \tag{12}$$

Although it is obvious that NURBS is not an interpolating function, the displacement $u(\mathbf{x})$ and geometric position $x(\xi)$ can be obtained by the following formula.

$$\begin{cases} u(\mathbf{x}) = \sum_{A=1}^{NP} R_A(\xi) \mathrm{c}_A \\ x(\xi) = \sum_{A=1}^{NP} R_A(\xi) P_A \end{cases} \quad \mathbf{x} = \{x,y\}, \xi = \{\xi,\eta\} \tag{13}$$

where NP reprensents the number of control points; c_A reprensents the relevant parameters of the control point P_A.

Then the strain smoothing technology is extended to the method, and any element can be divided into several subdomains of the same size. As can be seen from Figure 2, The subdomains of parameter space and physical space are one-to-one corresponding and can be transformed by Jacobian matrix. In the Figure 2, $\Omega_\xi^{a_i}$ is the subdomain in the parametric space; $\Omega_x^{a_i}$ is the subdomain in the physical space; $x_c^{a_i}$ and $\xi_c^{a_i}$ are the central point of the subdomain in the corresponding space; $\Gamma_\xi^{a_i}$ is the boundary of subdomain $\Omega_\xi^{a_i}$. The equation can be expressed from the subdomain as:

$$\begin{cases} \sum_{a=1}^{nel} \sum_{i=1}^{ns} \Omega_\xi^{a_i} = \Omega_\xi \\ \Omega_\xi^{a_i} \cap \Omega_\xi^{a_j} = \varnothing \end{cases} \tag{14}$$
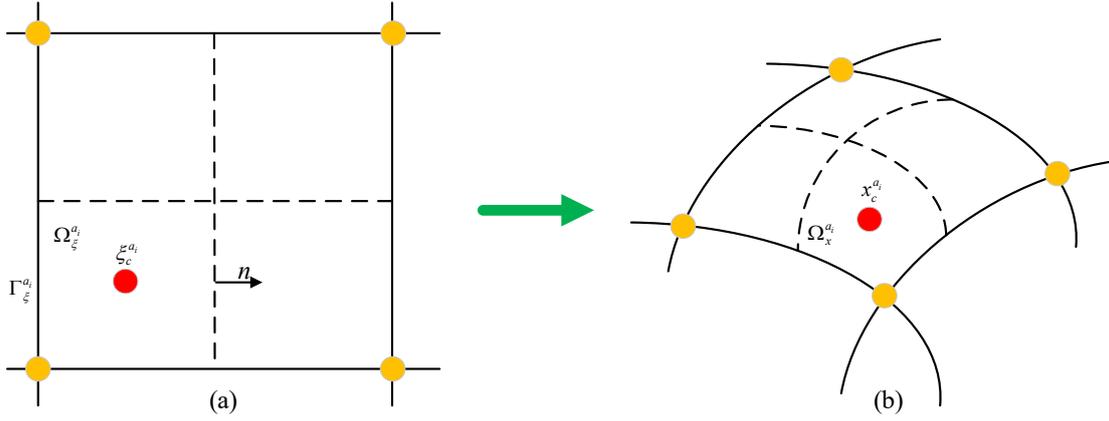
Figure 2. The corresponding relation of each space subdomain: (a) Parametric subdomain; (b) Physical subdomain
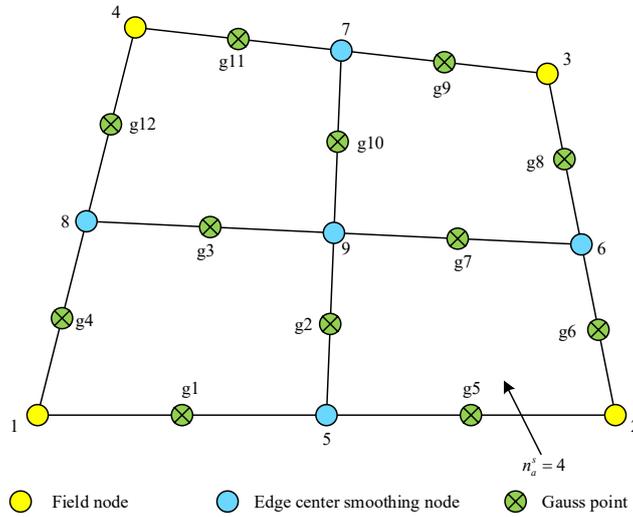
Figure 3. Positions of Gauss points in the 4 smoothing domains

The numerical integral calculation can be carried out through the midpoint of the boundary in the corresponding subdomain in the parameter space, as shown in the Figure 3. It only requires the value of the basis function at the midpoint. It completely avoids the calculation of the derivative. The strain smoothing technology can be expressed as:

$$\tilde{\boldsymbol{\xi}}^h\left(\mathbf{x}_c^{a_i}\right) = \frac{1}{A_x^{a_i}} \int_{\Omega_x^{a_i}} \boldsymbol{\xi}^h\left(\mathbf{x}\right) d\Omega$$

$$= \frac{1}{A_x^{a_i}} \int_{\Omega_x^{a_i}} \sum_{A=1}^{NC} \left\{ \begin{array}{c} R_{A,x} u_{Ax} \\ R_{A,y} u_{Ay} \\ R_{A,y} u_{Ax} + R_{A,x} u_{Ay} \end{array} \right\} d\Omega \qquad (15)$$

$$= \sum_{A=1}^{NC} \tilde{\mathbf{B}}_A(\mathbf{x}_c^{a_i}) \mathbf{u}_A$$

where

$$\tilde{\mathbf{B}}_A\left(\mathbf{x}_c^{a_i}\right) = \begin{bmatrix} \tilde{\nabla}_x R_A\left(\xi_c^{a_i}\right) & 0 \\ 0 & \tilde{\nabla}_y R_A\left(\xi_c^{a_i}\right) \\ \tilde{\nabla}_y R_A\left(\xi_c^{a_i}\right) & \tilde{\nabla}_x R_A\left(\xi_c^{a_i}\right) \end{bmatrix} \tag{16}$$

$$\tilde{\nabla}_x R_A\left(\mathbf{x}_c^{a_i}\right) = \frac{1}{A_x^{a_i}} \int_{\Omega_x^{a_i}} \nabla_x R_A(\mathbf{x})\mathrm{d}\Omega \tag{17}$$

where $\tilde{\xi}^h\left(\mathbf{x}_c^{a_i}\right)$ represents the smoothed strain; $\xi^h(\mathbf{x})$ represents the standard strain; $A_x^{a_i}$ represents the area of the subdomain $\Omega_x^{a_i}$; NC represents the total number of control points; $\tilde{\mathbf{B}}_A\left(\mathbf{x}_c^{a_i}\right)$ represents smoothed strain matrix; $\nabla_x R_A(\mathbf{x})$ represents the standard gradient; $\tilde{\nabla}_x R_A\left(\mathbf{x}_c^{a_i}\right)$ represents the smoothed gradient in the sub-domain $\Omega_x^{a_i}$.

The smoothed gradient $\tilde{\nabla}_x R_A\left(\mathbf{x}_c^{a_i}\right)$ is further presented as:

$$\begin{aligned} \tilde{\nabla}_x R_A\left(\mathbf{x}_c^{a_i}\right) &= \frac{1}{A_x^{a_i}} \int_{\Omega_x^{a_i}} \left[\mathbf{J}(\xi)\right]^{-\mathrm{T}}\left[\nabla_\xi R_A(\xi)\right]\mathrm{d}\Omega \\ &\approx \left[\tilde{\mathbf{J}}\left(\xi_c^{a_i}\right)\right]^{-\mathrm{T}} \frac{1}{\tilde{J}\left(\xi_c^{a_i}\right)A_\xi^{a_i}} \int_{\Omega_x^{a_i}} \nabla_\xi R_A(\xi)\tilde{J}\left(\xi_c^{a_i}\right)\mathrm{d}\Omega \\ &= \left[\tilde{\mathbf{J}}\left(\xi_c^{a_i}\right)\right]^{-\mathrm{T}} \frac{1}{A_\xi^{a_i}} \int_{\Omega_x^{a_i}} \nabla_\xi R_A(\xi)\mathrm{d}\Omega \\ &= \left[\tilde{\mathbf{J}}\left(\xi_c^{a_i}\right)\right]^{-\mathrm{T}} \tilde{\nabla}_\xi R_A(\xi) \end{aligned} \tag{18}$$

where

$$\begin{aligned} \tilde{\nabla}_\xi R_A(\xi) &= \frac{1}{A_\xi^{a_i}} \int_{\Omega_\xi^{a_i}} \nabla_\xi R_A(\xi)\mathrm{d}\Omega \\ &= \frac{1}{A_\xi^{a_i}} \int_{\Omega_\xi^{a_i}} R_A(\xi)\mathbf{n}(\xi)\mathrm{d}\Omega \end{aligned} \tag{19}$$

where $\mathbf{n}(\xi)$ represents the outer normal vector of $\Gamma_\xi^{a_i}$; $\tilde{J}\left(\xi_c^{a_i}\right)$ represents the Jacobian matrix of the central point $\xi_c^{a_i}$.

$$\tilde{\mathbf{J}}\left(\xi_c^{a_i}\right) = \sum_{A=1}^{NC} \begin{bmatrix} x_A\tilde{\nabla}_\xi R_A\left(\xi_c^{a_i}\right) & x_A\tilde{\nabla}_\eta R_A\left(\xi_c^{a_i}\right) \\ y_A\tilde{\nabla}_\xi R_A\left(\xi_c^{a_i}\right) & y_A\tilde{\nabla}_\eta R_A\left(\xi_c^{a_i}\right) \end{bmatrix} \tag{20}$$

$$\tilde{J}\left(\xi_c^{a_i}\right) = \det\left[\tilde{\mathbf{J}}\left(\xi_c^{a_i}\right)\right] \tag{21}$$

According to the principle of virtual work, the equation can be expressed as:

$$\int_{\Omega} (\sigma_{ij,j} + b_i)\delta u_i d\Omega - \int_{\Gamma_u^2} (\sigma_{ij} n_j - \bar{T}_i)\delta u_i d\Gamma = 0 \tag{22}$$

The above equilibrium equation can be further simplified as:

$$\int_{\Omega} -\sigma_{ij}\delta u_i d\Omega + \int_{\Omega} b_i\delta u_i d\Omega + \int_{\Gamma_u^2} \bar{T}_i\delta u_i d\Gamma = 0 \tag{23}$$

The expression of the functional for total potential energy Π in the analysis domain can be presented as:

$$\Pi = \int_{\Omega} -\sigma_{ij} u_i d\Omega + \int_{\Omega} b_i u_i d\Omega + \int_{\Gamma_u^2} \bar{T}_i u_i d\Gamma \tag{24}$$

Let $\delta\Pi = 0$, then the strain smoothing technique is applied to the formula to simplify the corresponding system equation. Finally, the equilibrium equation of the system is obtained.

$$\mathbf{Ku} = \mathbf{P} \tag{25}$$

where

$$\mathbf{K} = \sum_{a=1}^{nel} \sum_{i=1}^{ns} [\tilde{\mathbf{B}}(\boldsymbol{\xi}_c^{a_i})]^{\mathrm{T}} \mathbf{C}\tilde{\mathbf{B}}(\boldsymbol{\xi}_c^{a_i}) \mathbf{J}(\boldsymbol{\xi}_c^{a_i}) A_{\xi}^{a_i} \tag{26}$$

$$\begin{aligned}
\mathbf{P} &= \sum_{a=1}^{nel} \sum_{i=1}^{ns} [\mathbf{R}_A(\xi_c^{a_i})]^{\mathrm{T}} \mathbf{b}(\xi_c^{a_i}) \mathbf{R}_A(\xi_c^{a_i}) \mathbf{J}(\xi_c^{a_i}) A_{\xi}^{a_i} \\
&+ \sum_{a=1}^{nel} \sum_{j=1}^{nb} [\mathbf{R}_A(\xi_b^{a_j})]^{\mathrm{T}} \bar{\mathbf{T}}(\xi_b^{a_j}) \mathbf{R}_A(\xi_b^{a_j}) \mathbf{J}^b(\xi_b^{a_j}) L_{\xi}^{a_j}
\end{aligned} \tag{27}$$

where nb is the number of sides that the external force is applied to the boundary; $\xi_b^{a_j}$ is the midpoint of the side; a_j is the corresponding sub-domain; J^b is the Jacobian; $L_{\xi}^{a_j}$ is the length of the side.

3. Numerical examples

3.1. Validation
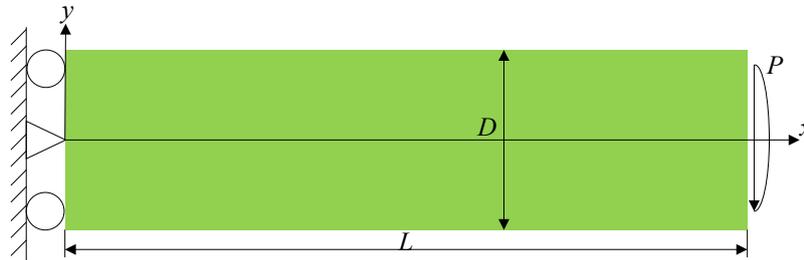
Figure 4. Geometric model of the cantilever beam

In order to verify the high precision and high convergence of CS-IGA in solving statics problems, a cantilever beam with an analytical solution under parabolic load is used for analysis. As shown in Figure 4, the geometric and material parameters are taken as $L = 48$m, $D = 12$m, $P = 1000$N, $E = 3\times10^7$Pa, $v = 0.3$. The free end of the cantilever beam is subjected to parabolic surface force and the left end is fixed. The problem can be thought of as a plane stress problem

and the thickness is selected as unit thickness.

The expression for the parabolic surface force applied at the free end is given by

$$\tau_{xy}\big|_{x=L} = \frac{P}{2I}\left(\frac{h^2}{4} - y^2\right) \tag{28}$$

The analytical solution to the cantilever beam can be expressed as [41]:

$$u(x,y) = -\frac{Py}{6EI}\left[(6L - 3x)x + (2+v)\left(y^2 - \frac{h^2}{4}\right)\right] \tag{29}$$

$$v(x,y) = \frac{P}{6EI}\left[3vy^2(L-x) + (4+5v)\frac{h^2 x}{4} + (3L-x)x^2\right] \tag{30}$$

$$I = \frac{h^3}{12} \tag{31}$$

where *I* represents the inertia moment per unit thickness; *E* represents elasticity modulus; *v* represents poisson ratio.
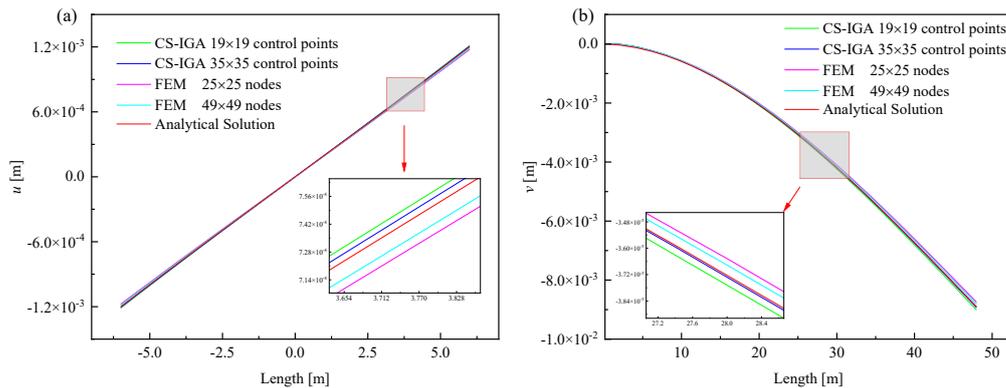
Figure 5. Displacement at upper side of cantilever beam: (a) *x*-direction displacement component *u*; (b) *y*-direction displacement component *v*

As can be seen from Figure 5, the displacement *u* on the central line and the displacement *v* on the central axis are calculated by different discrete schemes. By comparing with the analytical solution, the results of CS-IGA are basically consistent with the analytical solution. It verifies the high accuracy of CS-IGA. At the same time, the calculation results of CS-IGA under different discrete schemes are basically consistent with the analytical solutions, which verifies that CS-IGA has good convergence and effectiveness. It is easy to find that compared with FEM, CS-IGA uses fewer nodes but achieves higher computational accuracy.

In addition, the displacement contour of the cantilever beam structure is given in Figure 6. The results calculated by CS-IGA are approximate to reference solution. It further shows that CS-IGA has good computational accuracy.
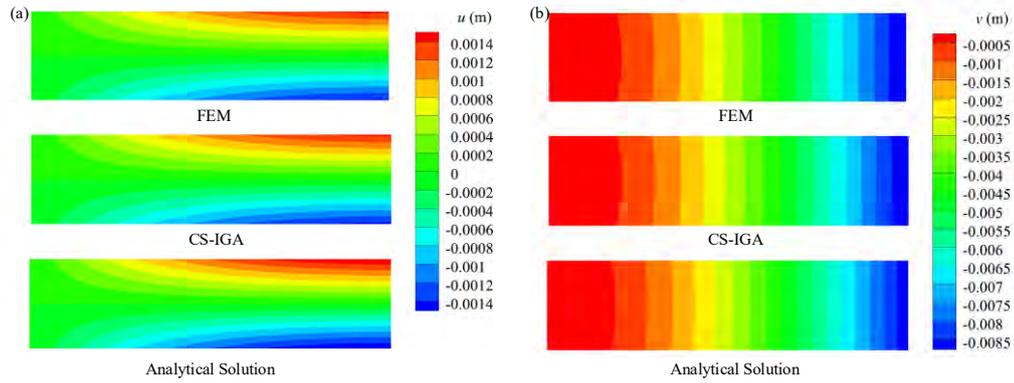
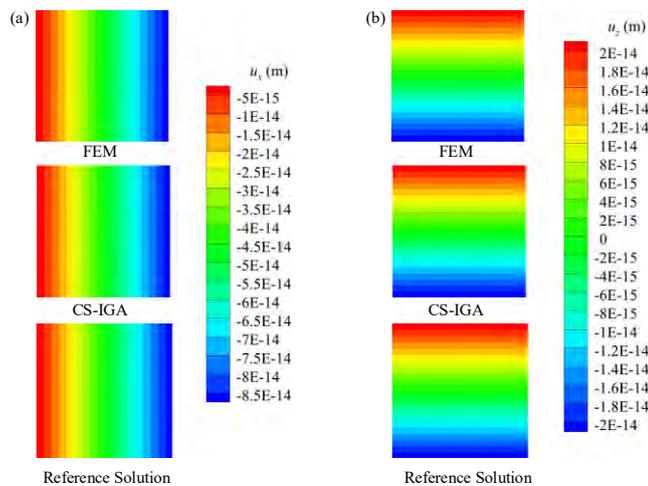**Figure 6. The displacement contour plots: (a) *x*-direction displacement component *u*; (b) *y*-direction displacement component *v***

3.2. Trapezoidal cantilever beam

In this section, CS-IGA and FEM are applied to solve the displacement of the upper surface of a trapezoidal cantilever beam respectively. In the CS-IGA, the order of the NURBS basis function is $m = n = 3$. The node vectors are $\mathbf{E}$ = [0 0 0 0 1 2 2 2 2] and $\mathbf{H}$ = [0 0 0 0 1 2 2 2 2]. The integration point is the midpoint of the edge of the subdomain. The geometry of the structure (length L = 2m, width H = 0.3m and h = 0.1m) is shown in Figure 7. The material parameters are taken as E=7.6×10^{11}Pa, v=0.33. The left end is fixed with u_x = 0 (x=0). The structure is subjected to a concentrated force load F = 100N above the free end.
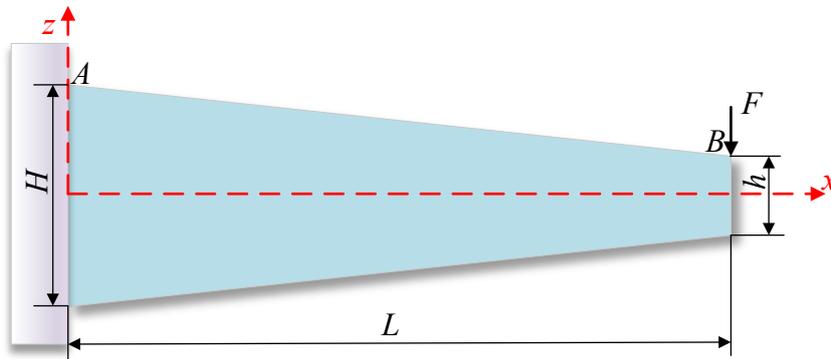
Figure 7. Geometric model of the trapezoidal cantilever beam
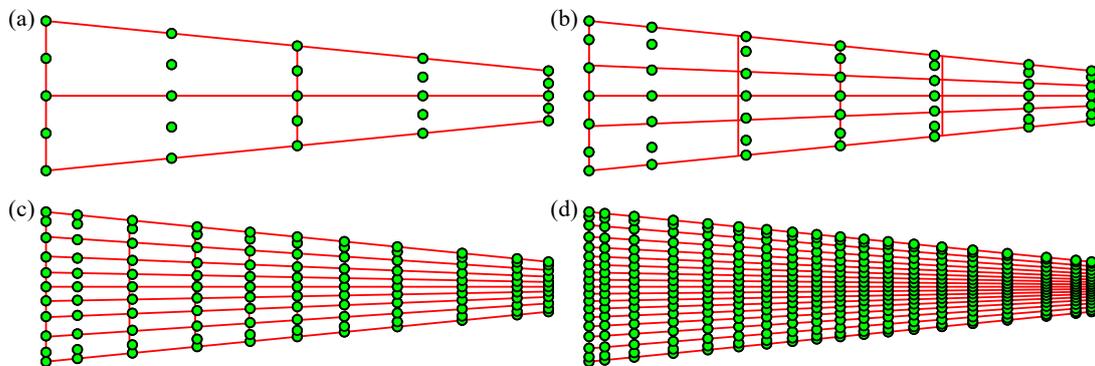
Figure 8. Different mesh divisions for the trapezoidal beam: (a) Primary mesh; (b)Refine once; (c) Refine twice; (d) Refine third

In this example, the plane stress condition and the third-degree NURBS basis function are used

206

to provide the asymptotic fine uniform mesh for geometric discretization of the piezoelectric cantilever beam. The different discretized meshes are shown in Figure 8.
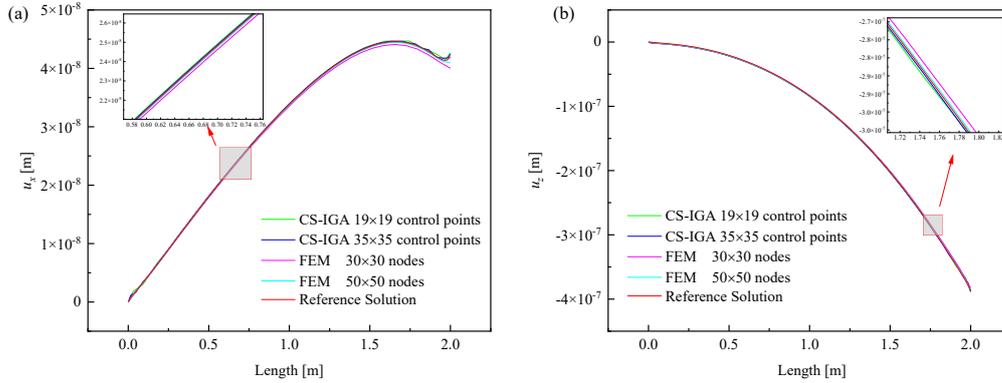
Figure 9. Displacement at edge *AB* of trapezoidal cantilever beam: (a) *x*-direction displacement component *u_x*; (b) *z*-direction displacement component *u_z*

In this numerical example, the displacement of the structure is calculated by the FEM using 900 nodes and 2500 nodes. In CS-IGA, four subdomains are spilt for an element. The structure is calculated by using 361 control points and 1225 control points. The reference solution is obtained by using FEM with 10000 nodes. The results of the trapezoidal cantilever beam are obtained, including the displacement u_x in the x direction and u_z along the z direction. The calculated results are shown in the Figure 9(a) to Figure 9(b).
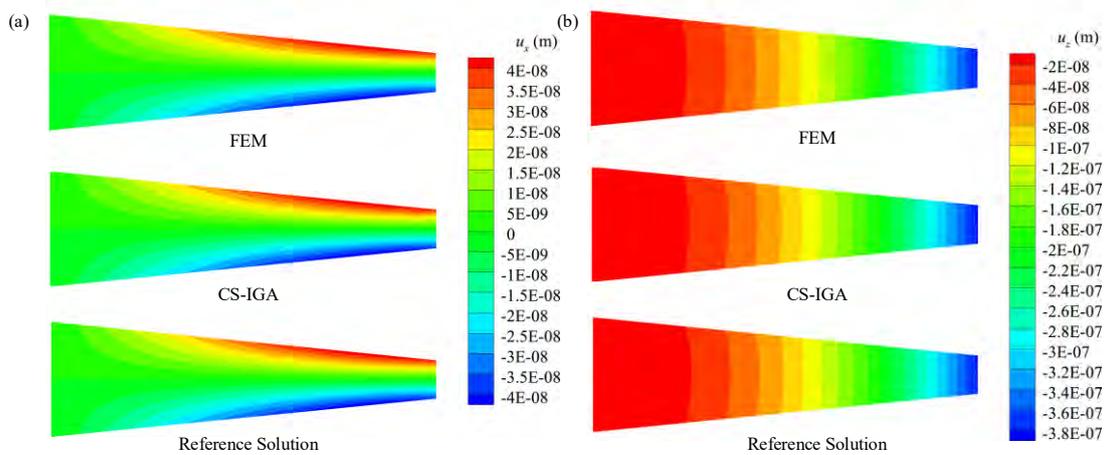
Figure 10. The displacement contour plots: (a) *x*-direction displacement component *u_x*; (b) *z*-direction displacement component *u_z*

Additionally, the displacement contour of trapezoidal cantilever beam structure is calculated by using CS-IGA and FEM. It can be seen from Figure 10 that the solution obtained by CS-IGA is similar to the reference solution. It also indicates that CS-IGA has a high accuracy.

3.3. Superposition cantilever beam formed by two materials

The geometric model of the superposition cantilever beam composed of nickel and copper is shown in Figure 11. The geometric and material parameters are taken as $L = 3$m, $h = 0.1$m, $P = 1000$N/m, $E_N = 2.1 \times 10^{11}$Pa, $v_N = 0.31$, $E_C = 1.1 \times 10^{11}$Pa, $v_C = 0.33$. The side *AB* of the cantilever beam is under uniform force and the left side is the fixed end. The model is constructed by

NURBS basis function. The node vectors are **E** = [0 0 0 0 1 2 2 2 2] and **H** = [0 0 0 0 1 2 2 2 2].
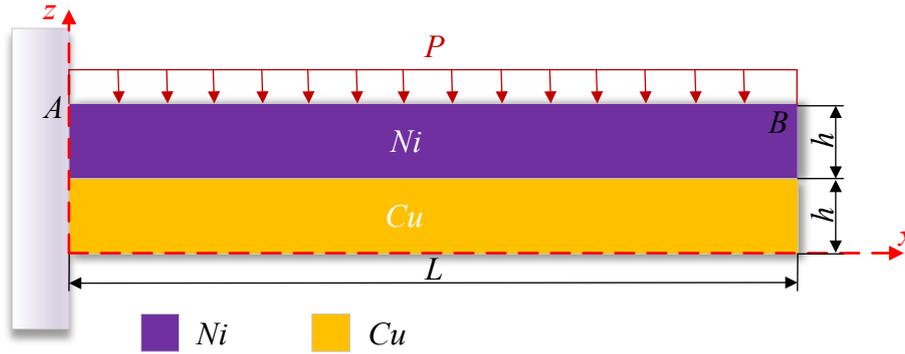
Figure 11. Geometric model of the superposition cantilever composed of nickel and copper

In the Figure 12, the displacement of the structure is calculated by the FEM using 19×19 nodes and 35×35 nodes. In CS-IGA, the structure is calculated by using the same number of control points. The reference solution is obtained by using FEM with 99×99 nodes. The results of the calculation of edge *AB* were obtained, including the displacement u_x along the *x* direction and u_z along the *z* direction. The results are shown in the Fig. 12(a) to Fig. 12(b). It will prove that this method has certain advantages in analyzing structural problems.
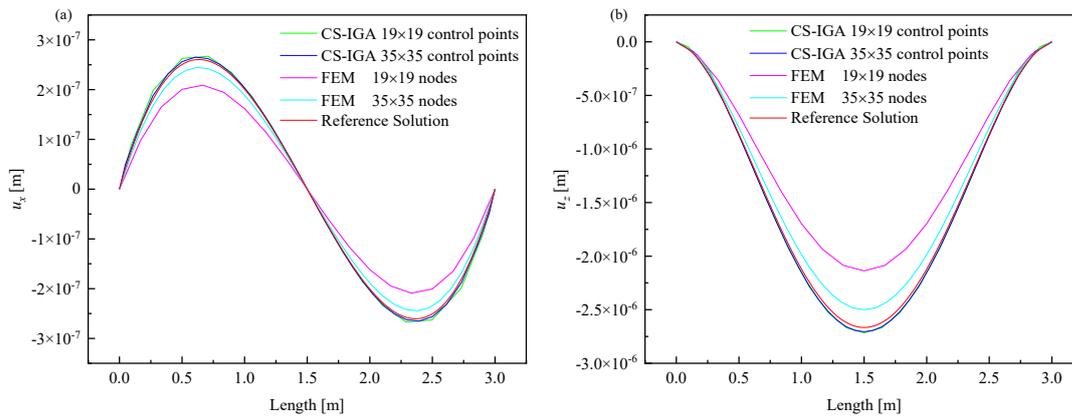
Figure 12. Displacement at edge *AB* of superposition cantilever beam: (a) *x*-direction displacement component u_x; (b) *z*-direction displacement component u_z
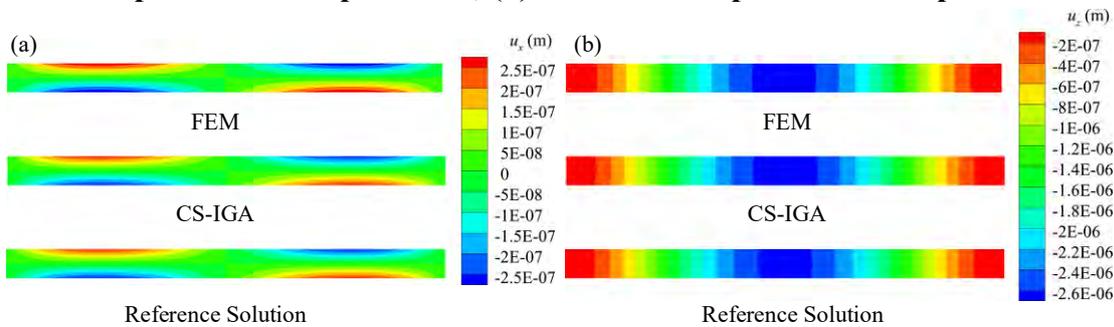
Figure 13. The displacement contour plots: (a) *x*-direction displacement component u_x; (b) *z*-direction displacement component u_z.

In the Figure 13, it can be seen that the displacement of the whole beam obtained by using CS-

IGA is consistent with the reference displacement. The high precision of CS-IGA is shown in analyzing the composite beam structure of multiple materials.

3.4. A square plate under uniform load
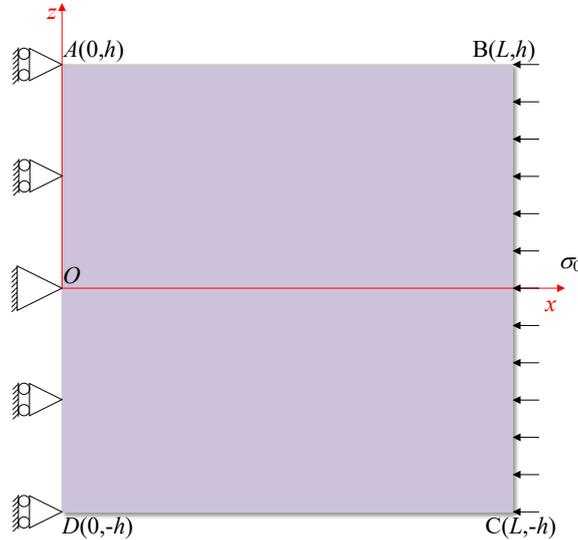
Figure 14. Geometric model of the square plate

Considering a square plate under uniform load, CS-IGA is used to calculate the displacement of the structure. The geometric model is shown in Figure 14. The geometric and material parameters are taken as $L = 1$mm, $h = 0.5$mm, $P = 1000$N, $E = 1 \times 10^{11}$Pa, $v = 0.33$. The boundary conditions are given by

$$\sigma_z(x, z = \pm h) = 0, \quad \tau_{xz}(x, z = \pm h) = 0 \tag{32}$$

$$\sigma_x(x = L, z) = \sigma_0, \quad \tau_{xz}(x = L, z) = 0 \tag{33}$$

$$u_x(x = 0, z) = 0, \quad u_z(x = 0, z = 0) = 0 \tag{34}$$

Figure 15. The displacement contour plots: (a) x-direction displacement component u_x; (b) z-direction displacement component u_z

The displacement contour of square plate structure is shown in Figure 15. The displacement

results calculated by CS-IGA are basically consistent with the reference solution, which shows that the method has good precision.

4. Conclusion

The statics of two-dimensional structures is analyzed by CS-IGA method. Several classical numerical examples are used to verify the precision of the displacement results under external forces and different boundary conditions. Compared the results with FEM, the efficiency and convergence of CS-IGA are further verified. The following are the conclusions:

(1) It is proved that CS-IGA has good accuracy, high efficiency and convergence in two-dimensional structural analysis;

(2) When calculating the stiffness matrix, CS-IGA only needs to calculate the shape function value of the integral point, which avoids the calculation of the derivative;

(3) Compared to FEM, CS-IGA can obtain more accurate results using fewer control points;

(4) CS-IGA has a good application prospect for solving multi-physics coupling problems.

CRediT authorship contribution statement

Liming Zhou: Methodology, Writing - original draft, Writing - review & editing. **Jinwen Geng:** Validation, Software, Writing - review & editing. **Pengxu Chen:** Writing - review & editing. Peng Liu: Writing - review & editing. **He Zhu:** Writing - review & editing. **Zhiqiang Gao:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

[1] Pinnola, F. P., Vaccaro, M. S., Barretta, R., and de Sciarra, F. M. (2021) Finite element method for stress-driven nonlocal beams, *Engineering Analysis with Boundary Elements* **134**, 22-34.

[2] Smoljanovic, H., Uzelac, I., Trogrlic, B., Zivaljic, N., and Munjiza, A. (2018) A computationally efficient numerical model for a dynamic analysis of beam type structures based on the combined finite-discrete element method, *Materialwissenschaft Und Werkstofftechnik* **49**(5), 651-665.

[3] Wei, D. M., and Liu, Y. (2012) Analytic and finite element solutions of the power-law Euler-Bernoulli beams, *Finite Elements in Analysis and Design* **52**, 31-40.

[4] Hosseinian, N., and Attarnejad, R. (2023) A New Rectangular Finite Element for Static and Dynamic Analysis of Arbitrarily Tapered Plates, *International Journal of Applied Mechanics* **15**(04), 2250099.

[5] Patuelli, C., Cestino, E., and Frulla, G. (2023) A Beam Finite Element for Static and Dynamic Analysis of Composite and Stiffened Structures with Bending-Torsion Coupling, *Aerospace* **10**(2), 142.

[6] Al-Zahrani, M. A., Asiri, S. A., Ahmed, K. I., and Eltaher, M. A. (2022) Free Vibration Analysis of 2D Functionally Graded Strip Beam using Finite Element Method, *Journal of Applied and Computational Mechanics* **8**(4), 1422-1430.

[7] Amirani, M. C., Khalili, S. M. R., and Nemati, N. (2009) Free vibration analysis of sandwich beam with FG core using the element free Galerkin method, *Composite Structures* **90**(3), 373-379.

[8] Wu, C., Xiang, H., and Du, X. P. (2016) Improved three-variable element-free Galerkin method for vibration analysis of beam-column models, *Journal of Mechanical Science and Technology* **30**(9), 4121-4131.

[9] Chen, C. L., and Hsu, C. H. (2019) An element free Galerkin method for static behavior of a magneto-electro-elastic beam in thermal environments, *Smart Materials and Structures* **28**(11), 115034.

[10] Xie, D., Jian, K. L., and Wen, W. B. (2017) An element-free Galerkin approach for rigid-flexible coupling dynamics in 2D state, *Applied Mathematics and Computation* **310**, 149-168.

[11] Yan, X. S. (2022) Free vibration analysis of a rotating nanobeam using integral form of Eringen's nonlocal theory and element-free Galerkin method, *Applied Physics a-Materials Science & Processing* **128**(8), 641.

[12] Liu, Y., Si, X. N., Liu, P., and Zhang, X. (2018) Mesoscopic modeling and simulation of 3D orthogonal woven composites using material point method, *Composite Structures* **203**, 425-435.

[13] Htike, H., Chen, W., Gu, Y., and Yang, J. J. (2011) Material Point Method with RBF Interpolation, *Cmes-Computer Modeling in Engineering & Sciences* **72**(4), 247-272.

[14] Andersen, S., and Andersen, L. (2010) Analysis of spatial interpolation in the material-point method, *Computers & Structures* **88**(7-8), 506-518.

[15] Beuth, L., Wieckowski, Z., and Vermeer, P. A. (2011) Solution of quasi-static large-strain problems by the material point method, *International Journal for Numerical and Analytical Methods in Geomechanics* **35**(13), 1451-1465.

[16] Kang, J. G., Homel, M. A., and Herbold, E. B. (2022) Beam elements with frictional contact in the material point method, *International Journal for Numerical Methods in Engineering* **123**(4), 1013-1035.

[17] Hsu, M. H., and Tan, P. J. B. (2015) Dynamic analysis of grouped turbo blades using radial basis function, *Advances in Mechanical Engineering* **7**(6), 1-9.

[18] Xiang, S., Jin, Y. X., Jiang, S. X., Bi, Z. Y., and Wang, K. M. (2011) Meshless global radial point collocation method for three-dimensional partial differential equations, *Engineering Analysis with Boundary Elements* **35**(3), 289-297.

[19] Panchore, V. (2022) Selection of Radial Basis Functions for the Accuracy of Meshfree Galerkin Method in Rotating Euler-Bernoulli Beam Problem, *International journal of applied and computational mathematics* **8**(3), 121.

[20] Xiang, S., and Kang, G. W. (2018) Meshless solution of the problem on the static behavior of thin and thick laminated composite beams, *Mechanics of Composite Materials* **54**(1), 89-98.

[21] Chu, F. Y., Wang, L. H., Zhong, Z., and He, J. Z. (2014) Hermite radial basis collocation method for vibration of functionally graded plates with in-plane material inhomogeneity, *Computers & Structures* **142**, 79-89.

[22] Liu, G. R., Nguyen, T. T., Dai, K. Y., and Lam, K. Y. (2007) Theoretical aspects of the smoothed finite

element method (SFEM), *International Journal for Numerical Methods in Engineering* **71**(8), 902-930.

[23] Zhang, Z. Q., and Liu, G. R. (2010) Upper and lower bounds for natural frequencies: A property of the smoothed finite element methods, *International Journal for Numerical Methods in Engineering* **84**(2), 149-178.

[24] Liu, G. R., Nguyen-Thoi, T., Nguyen-Xuan, H., and Lam, K. Y. (2009) A node-based smoothed finite element method (NS-FEM) for upper bound solutions to solid mechanics problems, *Computers & Structures* **87**(1-2), 14-26.

[25] Nguyen-Thoi, T., Liu, G. R., and Nguyen-Xuan, H. (2009) Additional properties of the node-based smoothed finite element method (NS-FEM) for solid mechanics problems, *International Journal of Computational Methods* **6**(4), 633-666.

[26] Liu, G. R., Nguyen-Thoi, T., and Lam, K. Y. (2009) An edge-based smoothed finite element method (ES-FEM) for static, free and forced vibration analyses of solids, *Journal of Sound and Vibration* **320**(4-5), 1100-1130.

[27] Jarak, T., Jalusic, B., and Soric, J. (2020) Mixed meshless local petrov-galerkin methods for solving linear fourth-order differential equations, *Transactions of Famena* **44**(1), 1-12.

[28] Gao, H. F., and Wei, G. F. (2016) Complex Variable Meshless Manifold Method for Elastic Dynamic Problems, *Mathematical Problems in Engineering* **2016**, 5803457.

[29] Chien, T. H., Nhon, N. T., Hung, N. X., Rabczuk, T., and Bordas, S. (2011) A cell - based smoothed finite element method for free vibration and buckling analysis of shells, *Ksce Journal of Civil Engineering* **15**(2), 347-361.

[30] Li, E., He, Z. C., and Xu, X. (2013) An edge-based smoothed tetrahedron finite element method (ES-T-FEM) for thermomechanical problems, *International Journal of Heat and Mass Transfer* **66**, 723-732.

[31] Hughes, T. J. R., Reali, A., and Sangalli, G. (2010) Efficient quadrature for NURBS-based isogeometric analysis, *Computer Methods in Applied Mechanics and Engineering* **199**(5-8), 301-313.

[32] Auricchio, F., Calabro, F., Hughes, T. J. R., Reali, A., and Sangalli, G. (2012) A simple algorithm for obtaining nearly optimal quadrature rules for NURBS-based isogeometric analysis, *Computer Methods in Applied Mechanics and Engineering* **249**, 15-27.

[33] Mitchell, T. J., Govindjee, S., and Taylor, R. L. (2011). A Method for Enforcement of Dirichlet Boundary Conditions in Isogeometric Analysis. In D. Mueller-Hoeppe, S. Loehnert & S. Reese (Eds.), *Recent Developments and Innovative Applications in Computational Mechanics* (pp. 283-293).

[34] Wang, D. D., and Xuan, J. C. (2010) An improved NURBS-based isogeometric analysis with enhanced treatment of essential boundary conditions, *Computer Methods in Applied Mechanics and Engineering* **199**(37-40), 2425-2436.

[35] Kobis, M. A., and Arnold, M. (2014) Numerical solution of penalty formulations for constrained mechanical systems using heterogeneous multiscale methods, *Journal of Computational and Applied Mathematics* **262**, 193-204.

[36] Lee, S. J., and Sub, P. K. (2011) Free Vibration Analysis of Elastic Bars using Isogeometric Approach, *Architectural Research* **13**(3), 41-47.

[37] Milic, P., and Marinkovic, D. (2015) Isogeometric FE analysis of complex thin-walled structures, *Transactions of Famena* **39**(1), 15-26.

[38] Fahrendorf, F., Morganti, S., Reali, A., Hughes, T. J. R., and De Lorenzis, L. (2020) Mixed stress-displacement isogeometric collocation for nearly incompressible elasticity and elastoplasticity, *Computer Methods in Applied Mechanics and Engineering* **369**, 113112.

[39] Pekovic, O., Stupar, S., Simonovic, A., Svorcan, J., and Komarov, D. (2014) Isogeometric bending analysis

of composite plates based on a higher-order shear deformation theory, *Journal of Mechanical Science and Technology* **28**(8), 3153-3162.

[40] Lee, S. J. (2016) Free Vibrations of Plates and Shells with an Isogeometric RM Shell Element, *Architectural Research* **18**(2), 65-74.

[41] Wang, D. D., Zhang, H. J., and Xuan, J. C. (2012) A strain smoothing formulation for NURBS-based isogeometric finite element analysis, *Science China-Physics Mechanics & Astronomy* **55**(1), 132-140.

CFD Modeling of Wind Driven Rain Impacts in Singapore

†*George XU[1], Kendrick TAN[1], Zhengwei GE[1], Venugopalan RAGHAVAN[1], Harish GOPALAN[1], Hee Joo POH[1], Yong ENG[2], Hwee Sien TAN[2] and Irene LEE[2]

[1]Institute of High Performance Computing (IHPC), Agency for Science Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore
[2]Building & Research Institute, Housing Development Board, HDB Hub, 480 Lorong 6 Toa Payoh, Singapore 310480, Republic of Singapore

*Presenting author: xuxg@ihpc.a-star.edu.sg
†Corresponding author: xuxg@ihpc.a-star.edu.sg

Abstract

Wind driven rain (WDR) often has adverse impacts on built environment, in terms of damages to building structure and the durability of building façade. In tropical countries such as Singapore, safety concerns related to the floor wetting by rain infiltrated into occupational or transit areas have received increasing attention in the building design and operation. Under the green and sustainable building design in Singapore, WDR control for minimizing the safety risks becomes one of the critical performance-based design features.

Among the methods used to study WDR, CFD models, including both the Eulerian-Lagrangian approach and the Eulerian-Eulerian approach, have been widely adopted to address the impacts of WDR by researchers and designers in urban planning and green building design in Singapore. In-depth illustration of our recent progress in formulating and enhancing both CFD-based approaches which were implemented in the OpenFOAM framework is presented. Questions related to modelling assumptions, drag prediction, computational accuracy and efficiency are addressed. The CFD-based approaches have been applied to simulate the solutions for four different environmental test cases. These case studies demonstrate well that CFD-based approaches could be used as design tools to accurately and efficiently address the impacts of WDR on built environment design.

Keywords: Wind Driven Rain, Eulerian-Lagrangian, Eulerian-Eulerian, Raindrop Trajectory, CFD

Introduction

When a raindrop falls due to gravity, its trajectory will be diverted towards the wind direction due to the aerodynamic force exerted by the wind flow. When such wind driven rain (WDR) strikes on a building, it may have adverse impacts on the building structures, particularly under strong wind conditions.

As widely reported in literature [1], WDR may cause the damage of external façade and structures of a building and badly affect the durability of building surface finishing. If the building envelopment is not air-tight well, the rainwater may ingress through the gaps or cracks, which may subsequently induce mould growth and increased moisture contamination in the indoor environment. In addition, WDR may lead to the unexpected wetting of floors in transit areas and even occupied regions, which can impose severe safety inconvenience and threat to commuters or occupants. The latter impact due to WDR has been receiving greater attention by building designers and building authorities in Singapore [2].

As a small country with a tropical climate, Singapore always experiences high precipitation conditions. While there is no distinct wet or dry season in Singapore, monthly variations in rainfall do exist. Higher rainfall occurs from November to January during the wet phase of north-east Monsoon season. While the driest month falls in February. In terms of spatial distribution, rainfall is higher over the northern and western parts of Singapore and decreases towards the eastern part of the island. The typical rainfall intensities fall in the range from 10mm/hr to 120mm/hr, based on the past precipitation records from 1981 to 2010 [3]. Generally, rain may occur for almost half duration of each month and the rainfall intensity is relatively higher than that in countries with four seasons. Thus, special concerns about WDR impacts on occupied and transit areas have been increasingly addressed during the building design and operation stages.

In the Green Mark (GM) certification guidelines for residential and non-residential buildings [2], which are advocated by the Building and Construction Authority (BCA) of Singapore, computational fluid dynamics (CFD) based approaches are proposed for assessing the building performance impacted by WDR. The Eulerian-Lagrangian approach has been widely and primarily used for WDR studies in early days, in which the discrete phase model is adopted to model the trajectories of raindrops. Recently, Eulerian-Eulerian approach developed by Kubilay *et al* [4], in which the raindrops are modelled as continuous phases, gives an alternative option to simulate the WDR impacts. In addition to the two CFD-based simulation approaches, there are two more approaches for WDR impact studies: on-site measurement and semi-empirical methods [5]. The former is a post-study approach, which is not suitable for design stage where the building is not completed. While the latter suffers in accuracy, as surrounding buildings impose significant changes to the wind profiles approaching the building of interest. Owing to rapid progress in computing technologies, CFD based approaches have been highly preferred to evaluate the impacts of WDR and the design changes to effectively mitigate these impacts.

Both CFD-based simulation approaches have been developed and specifically customized for the efficient and effective study of WDR under Singapore climate conditions. Special attentions are paid to the development of Eulerian-Lagrangian approach based on the OpenFOAM framework [6], and its applicability in comparison to the Eulerian-Eulerian approach.

In this paper, the modelling assumptions for the CFD simulations will be briefly introduced, followed by the introduction of the two CFD-based approaches. Some innovative and user-friendly features, which are developed for enhanced accuracy and efficiency of the CFD solution procedure, are highlighted for brevity. WDR studies for four selected use cases, based on the proposed CFD approaches, will be addressed in the Results & Discussion section. Finally, a brief concluding remark for the current study is summarized.

CFD-based Modelling Approaches for WDR Simulation

- *Modelling assumptions*

In both CFD-based modelling approaches, it is reasonable to assume that the raindrop flow may not affect the airflow due to the wind, because of the relatively low volume fraction of raindrops [7]. Thus, the one-way coupling sequential solution procedure is adopted in the CFD approaches. The airflow subjected to the log-law wind profile corresponding to the neutral-state of atmospheric boundary layer (ABL) flow [8], is first simulated. Next, the trajectories or spatial distributions of the rain phase are simulated by the Lagrangian approach and the Eulerian approach, respectively.

The raindrops are assumed to be in a constant spherical shape and travels at the terminal velocities for the given diameters. Evaporation, deformation, breakup and the interactions among raindrops are ignored in the simulations.

In addition, when the raindrops strike on the buildings' structures, they are assumed to be trapped on the surfaces. The run-off along building surfaces, splashing, break-up and coalescence phenomena are ignored for the purpose of computational efficiency.

Steady-state CFD simulations for both incompressible turbulent airflow and the transport of the rain phases are carried out. Post-processing of wind driven rain impacts, in terms of specific catch ratio (SCR) [1], raindrop trajectories and wetting depth, can be performed with the help of python-based scripts that are specifically developed as user friendly features in the project.

- *Eulerian approach for airflow solutions*

  - Continuity Equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0 \tag{1}$$

  - Momentum Equation:

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \boldsymbol{u} + \boldsymbol{g} \tag{2}$$

In the above equations, ρ, $\boldsymbol{u}$, p, ν, $\boldsymbol{g}$ and t represent the air density, air velocity, static pressure, kinematic viscosity of the air, gravitational acceleration, and time, respectively.

- *Lagrangian approach for raindrop simulation*

In the Eulerian-Lagrangian approach, each raindrop is assumed be to a discrete phase and its tempo-spatial trajectories are traced with the following two equations:

  - Velocity of the raindrop:

$$\frac{d\boldsymbol{x_d}}{dt} = \boldsymbol{u_d} \tag{3}$$

  - Forces balance on the raindrop:

$$m_d \frac{d\boldsymbol{u_d}}{dt} = \Sigma \boldsymbol{F} = \boldsymbol{F_g} + \boldsymbol{F_D} \tag{4}$$

where the force due to gravitational acceleration is calculated as $\boldsymbol{F_g} = \rho_d \frac{\pi D_d^3}{6} \boldsymbol{g}$, the wind load exerted on the raindrop is estimated as $\boldsymbol{F_D} = C_D \frac{\pi D_d^2}{8} \rho_d (\boldsymbol{u} - \boldsymbol{u_d}) |\boldsymbol{u} - \boldsymbol{u_d}|$. The drag coefficient is estimated by curve fitting to the experimental data obtained by Gunn and Kinzer [9] and Stokes law, in the following form [1]

$$C_D = \begin{cases} \frac{24}{Re}, Re \leq 0.1 \\ 10^Y, Re > 0.1 \end{cases} \tag{5}$$

where Y is the 8th-order curve-fitting polynomial function, i.e. $Y = \sum_{i=0}^{8} A_i X^i$, $X = \log 10(Re)$ and the coefficients $A_0 = 1.4114$, $A_1 = -0.9058$, $A_2 = 0.0847$, $A_3 = 0.0233$, $A_4 = -0.0034$, $A_5 = -0.0021$, $A_6 = 8.3387 \times 10^{-5}$, $A_7 = 1.1488 \times 10^{-4}$, $A_8 = 1.1490 \times 10^{-5}$.

In the above formulae, $\boldsymbol{x_d}$, $\boldsymbol{u_d}$, m_d, ρ_d and D_d denote the location, velocity, mass, density and diameter, respectively, of the raindrop of interest.

- *Eulerian approach for raindrop simulations*

In the Eulerian approach for raindrop simulations, the rain is assumed to be a continuous phase dispersed in the air flow field. A set of species transport equations, as shown in Equations (6)-(7) below, are solved to obtain the spatial distribution of the raindrops of concern:

- Mass conservation of multi species:

$$\frac{\partial \alpha_d}{\partial t} + \frac{\partial \alpha_d \overline{u_{d,j}}}{\partial x_j} = 0 \qquad (6)$$

- Momentum conservation for individual specie:

$$\frac{\partial \alpha_d \overline{u_{d,i}}}{\partial t} + \frac{\partial \alpha_d \overline{u_{d,i}}\,\overline{u_{d,j}}}{\partial x_j} + \frac{\partial \alpha_d \overline{u'_{d,i}u'_{d,j}}}{\partial x_j} = \alpha_d g_i + \alpha_d \frac{3\mu_a}{\rho_w d^2}\frac{C_d Re_R}{4}\left(\overline{u_i} - \overline{u_{d,i}}\right) \qquad (7)$$

where

d	is the raindrop diameter;
α_d	is the volume fraction of the rain phase d;
$u_{d,j}$	is the velocity component of rain phase d;
u_i	is the velocity component of the wind in the i-th direction;
ρ_w	is the density of water;
μ_α	is the dynamic viscosity of air;
g	is the gravitational acceleration;
C_d	is the drag coefficient for the raindrop;
Re_R	denotes the relative Reynolds number calculated using the relative velocity between the air and rain phases.

- *Innovative features implemented in the Eulerian-Lagrangian approach*

It should be noted that the Eulerian-Eulerian approach, which was originally developed by Kubilay *et al* [4], is adopted in this study. We have put in efforts in customizing the application to be compliant with the tropical climate conditions and design concerns. Major efforts of the researchers have been spent on the development of Eulerian-Lagrangian approach for WDR simulations, based on the OpenFOAM v2012 framework [6]. In addition to the raindrop solvers based on Equations (3)-(4), some innovative features, which are developed primarily for efficient and user-friendly solutions, are summarized below:
- Smart selection of release plane [10]
- Catch ratio calculation based on mass conservation of stream tubes bounded raindrops [1]
- Wetting marker based on erosion rate [11]
- Particle trajectory calculation based on decomposed solution datasets in parallel computing

For brevity, technical details about the innovative features mentioned above will not be illustrated in the conference paper.

Results & Discussion

The Eulerian-Lagrangian approach has been actively applied to study the impacts of WDR in various test cases to address its accuracy, efficiency and its applicability, in comparison to the Eulerian-Eulerian approach. In this paper, WDR studies for four different building environments are addressed in depth in this section.

- *Unobstructed urban flow*

This test case corresponds to a fully developed neutral-state ABL flow throughout the computational domain [10]. The analytical solutions for airflow and raindrop trajectories can be derived, subject to the parameters for ABL flow and raindrop diameters of concern.

The proposed Eulerian-Lagrangian approach has been adopted to simulate the trajectories for raindrops with diameters of 0.5mm, 1mm, 2mm and 5mm, which are primarily advocated in the BCA GM certification code [2]. The Eulerian-Lagrangian simulation results for the various raindrop diameters are shown in Figure 1 in comparison to the analytical trajectories.

Figure 1(b) illustrates the changes to the downward velocity of the raindrops when raindrops are transported within the computational domain. It is observed that the simulated velocity components are in good agreement with their theoretical terminal velocities which are 2m/s, 4m/s, 6.5m/s and 9m/s, respectively. This is thanks to the correct implementation of raindrop drag coefficients in the Eulerian-Lagrangian solver. It should be noted that the correct way for predicting the drag coefficient for raindrops has also been implemented in the Eulerian-Eulerian solver that is use in current study.

(a) Computational domain for unobstructed urban flow

(b) Invariant velocity component along gravitational direction

(c) Trajectories of raindrops
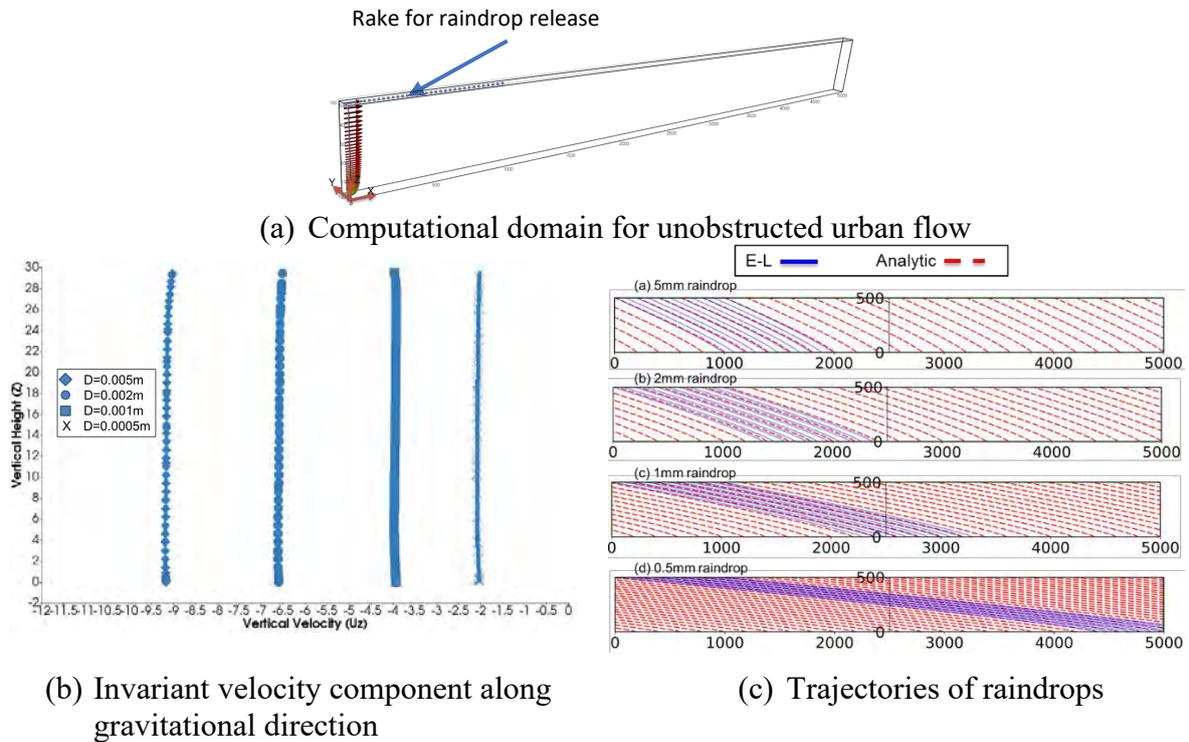
Figure 1. Trajectories of raindrops under unobstructed urban flow domain.

It is observed that the Eulerian-Lagrangian simulation results agree well with the analytical solutions. Under the same incoming wind conditions, smaller raindrops travel further in the horizontal direction before landing on the ground. Based on the wetting-related safety concerns due to WDR in Singapore, the smaller raindrops may be more problematic due to a deeper wetting depth, when subjected to stronger incoming wind conditions. This principle may be deviated if there are obstructions due to surrounding buildings.

This verification process also implies that the analytical solution derived for neutral-state log-law ABL wind profile is accurate. Thus, it is subsequently adopted for the automatic selection of release plane of raindrops in the Eulerian-Lagrangian solution procedure [10]. In comparison to the conventional trial-and-error manner for determining an effective release plane for raindrops in Eulerian-Lagrangian solution procedure, this innovative feature, as detailed in [10], can remarkably reduce the simulation efforts by up to 75%.

- *Two parallel wide buildings*

The two parallel wide buildings model [12] has been adopted for the verification of the developed Eulerian-Lagrangian and Eulerian-Eulerian approaches. Figure 2 shows the computational domain where the ABL flow normal to the longitudinal direction of the wide buildings is selected for study. The specific catch ratios on the two windward facades, obtained with Eulerian-Lagrangian approach for 1mm raindrops, are shown in Figure 2(b). In

general, the SCR on both facades increases along the vertical height. This may be attributed to the increment of approaching wind speed along the vertical direction. Also, the catch ratio on the taller building is lower than that of the low-height building at the same height. This is due to the blockage effect from the upstream low-height building, which effective lowers the local wind speed approaching the taller building.

(a) Computational domain

(b) Contour plots of specific catch ratios on windward facades

(c) Specific catch ratio profile along the central vertical line on the taller building windward façade for 1mm droplets

(d) Specific catch ratio profile along the central vertical line on the taller building windward façade for 5mm raindrops
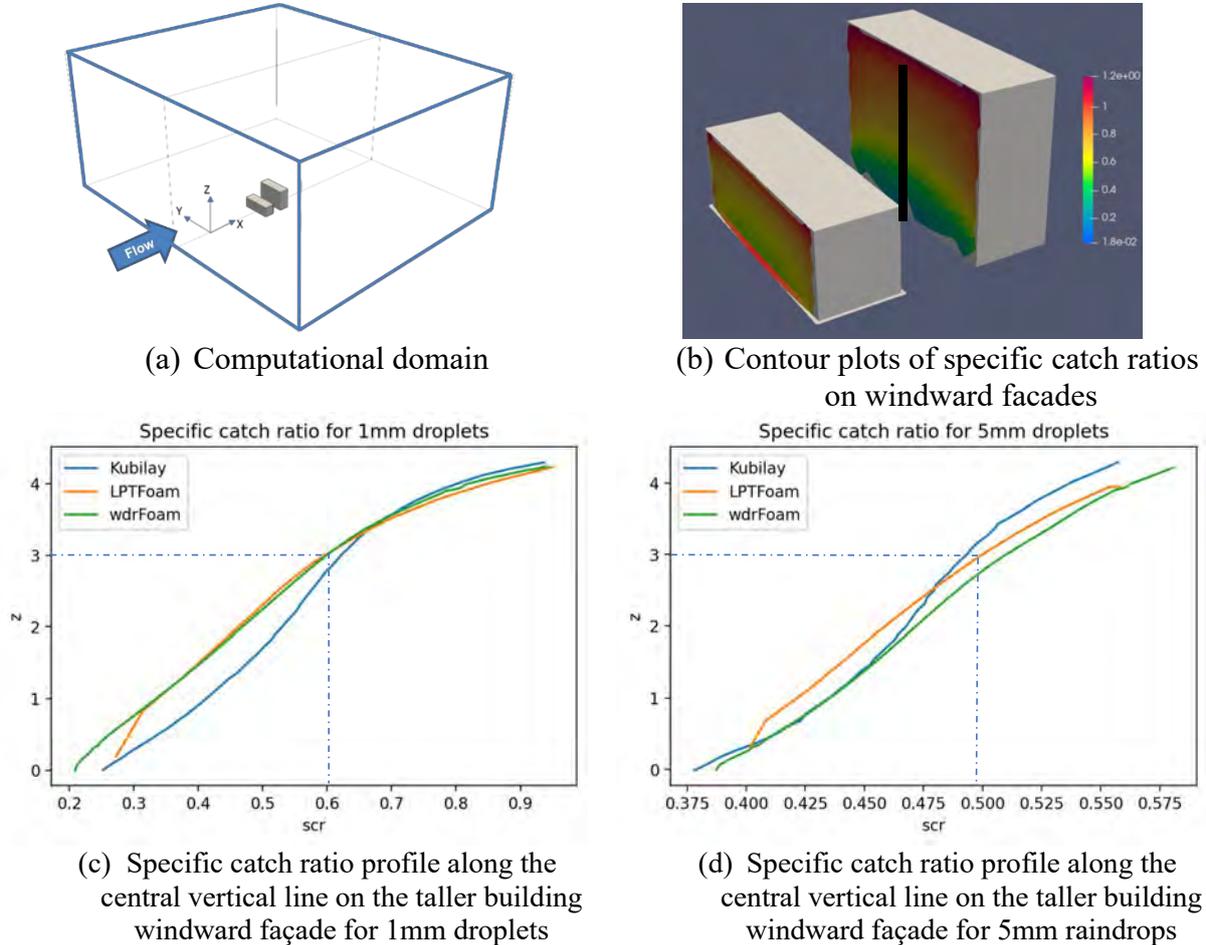
Figure 2: Simulated specific catch ratios on windward facades for the two parallel wide buildings model.

The SCR profiles, for 1mm and 5mm raindrops, along the central vertical line of the windward façade of the taller building are shown in Figures 2(c)-(d). The Eulerian-Lagrangian and Eulerian-Eulerian solutions are plotted with the literature data [12]. Our simulation results are observed to be agreeable with the literature data in terms of their trends and magnitudes.

In addition, as highlighted by dot-dash lines in Figures 2(c)-(d), at the vertical height of 3m, the SCR values for 1mm and 5mm raindrops are approximately 0.6 and 0.49, respectively. This implies that the smaller raindrops may impose relatively much stronger WDR impacts on building structures.

- *City model*

A simple city model [8], which is created by revising a tutorial model for wind around surrounding buildings in OpenFOAM v2012 package [6], are simulated using the developed Eulerian-Lagrangian approach.

Figures 3(a)-(b) show the geometrical model for the computational domain and the close-up view of the area of interest for WDR studies, respectively. The top surface, in particular, the

area under the dome shelter, of the brick-shaped base building is the primary focus in this study.

The wetted areas for 1mm raindrops are analysed under two ABL wind conditions corresponding to the wind speed of 3m/s (low-speed wind scenario) and of 17m/s (high-speed wind scenario), respectively. The selected winds are upstream of the building of interest and interact with a set of high-rise buildings before reaching it.

The simulated wetting areas for the base building and the ground surface under the two wind scenarios are shown in Figures 3(c)-(d), respectively, from a top-view angle. The wetted areas are denoted with the marker in red based on the erosion rate predicted in Eulerian-Lagrangian solution procedure. On the other hand, the blue region indicates the areas free from rain impact. Yellow arrows indicate the tunnel effects when incoming wind go through the gaps between adjacent buildings, which can carry the raindrops to travel towards the areas of interest. As shown in Figure 3(c), the air wake zone behind the high-rise buildings does create the shelter effect over the top surface of the base building, resulting in the blue portion free of WDR impact. In addition, the area under the dome is affected by the ingress of 1mm raindrops taking place at the front entrance edges only.

As the wind speed increases, Figure 3(d) reveals the large area free of WDR impact, due to the enlarged shelter effect from the high-rise buildings in front. In other words, the stronger wind flow deflected by the high-rise buildings effectively blow the falling 1mm raindrops away from the wake zone where the base building exists. Meanwhile, the augmented wind flow through the tunnels between high-rise buildings also carry the raindrops to travel longer distances, leading to deep penetration of 1mm raindrops into the area under the dome.

This test cases indicate that the impacts of WDR are highly dependent on the ABL wind conditions, as well as the blockages from surrounding buildings in the typical city environment.

(a) Computational domain

(b) Area of interest

(c) Wetting area under low-speed wind scenario

(d) Wetting area under high-speed wind scenario
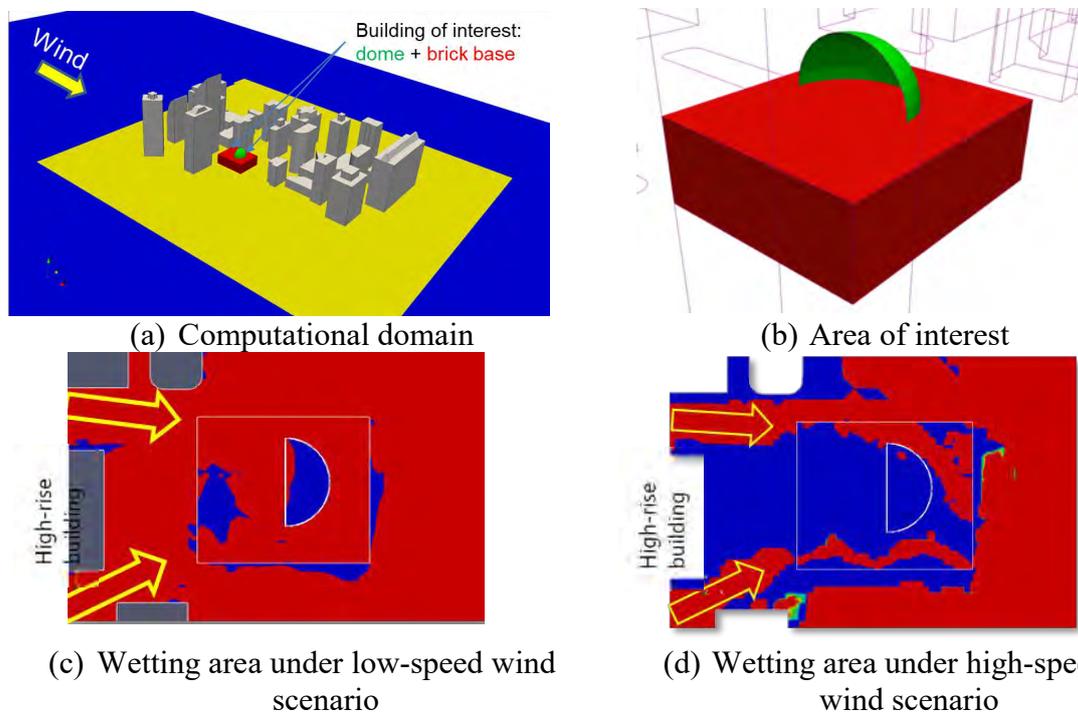
Figure 3. Wetting predictions by Eulerian-Lagrangian approach for a city model.

Parametric studies about the computational efficiency of Eulerian-Lagrangian approach has been carried out, based on the city model under the high-speed wind scenario. Relevant

simulations are performed on a high-end Ubuntu-based workstation. The details about the workstation specification are summarized in Table 1 below.

Table 1: Specifications of computing workstation used for studies.

Components	Specs
Operating System	Ubuntu 20.04, x86_64
No. of CPUs	24
Model of CPU	Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz
RAM	384G (available)

The efficiency test is limited to the dependence of computational time for the Lagrangian solution process on a number of 1mm raindrops released from the same area. As shown in Figure 4, the CPU time for Lagrangian solution is linearly proportional to the number of raindrops to be simulated. The linear curve-fit equation is also presented in Figure 4. The efficiency test would be useful for modelers to estimate the simulation time when the number of raindrops is used in their simulations.
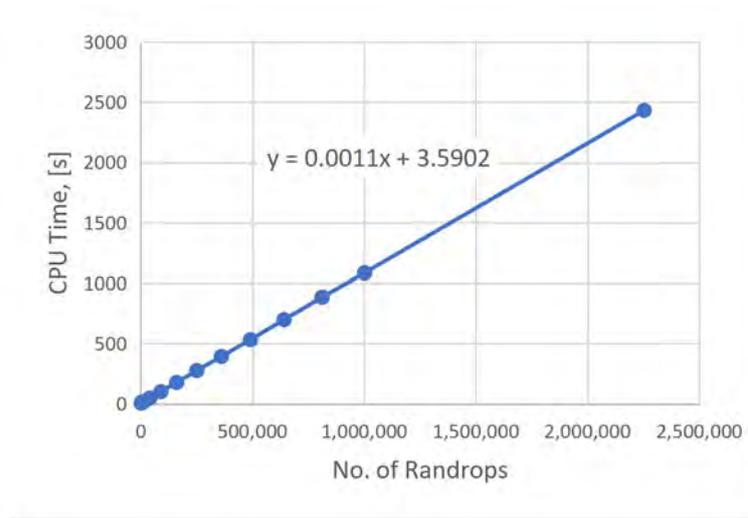
Figure 4. Computational efficiency of Eulerian-Lagrangian approach for the city model under the high-speed ABL flow conditions.

- *Neighborhood buildings*

The last test case applies the developed simulation-based approaches to assess the effectiveness of a mitigation measure for the design of neighbourhood buildings, as shown in Figure 5(a). The design is focused on the pick-up and drop-off (PUDO) shelter area, as depicted in Figure 5(b), within a cluster of neighbourhood buildings. Since WDR is expected to affect the PUDO area, vertical drop panel extended downwards from the perimeter of the shelter is proposed by the designers to mitigate WDR impact.
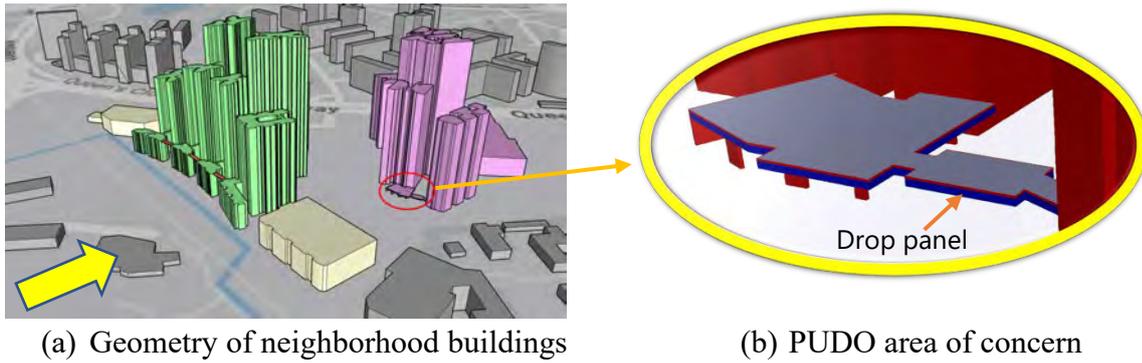
(a) Geometry of neighborhood buildings   (b) PUDO area of concern
Figure 5. Geometrical model for neighborhood buildings and the PUDO area of concern in WDR study.

In this study, both CFD-based WDR approaches have been adopted to study the 2mm raindrops impacts under the ABL wind flow normal to the PUDO region. The ABL flow is in accordance with the wind speed of 6.6m/s at the reference height of 15m above the ground.

The wetted areas by 2mm raindrops obtained from Eulerian-Lagrangian and Eulerian-Eulerian approaches are shown in Figure 6 and Figure 7, respectively. As shown in Figure 6, the drop panel reduces the wetting depth in the region as highlighted by the oval shape, but not as effectively in the other regions under the shelter. Similar results are also observed in the solutions from Eulerian-Eulerian approach, as shown in Figure 7. Both approaches give rise to agreeable prediction of wetting depths for cases without and with the drop panel proposed.

It is worth nothing that the wetted areas from Eulerian-Eulerian approach are based on the value of volume fraction of respective raindrops, while in Eulerian-Lagrangian approach, the marker based on erosion rate is in use. The procedure to select a suitable threshold values for the two different variables for the determination of the wetting depth is still unclear. Relevant measured data will be required to justify the appropriate choice of the threshold values.
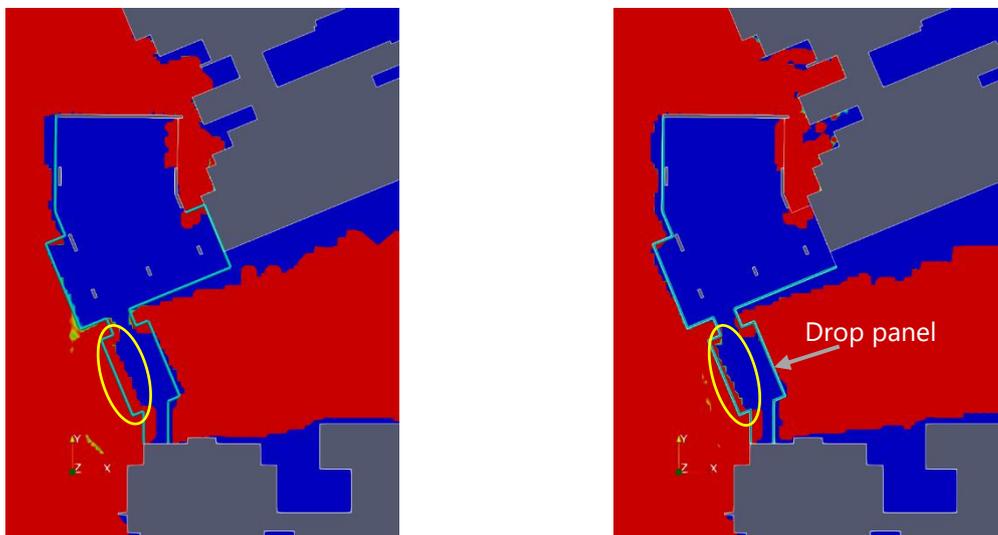
(a) Without vertical drop panel   (b) With vertical drop panel
Figure 6. Wetting areas obtained from Eulerian-Lagrangian approach.

(a) Without vertical drop panel   (b) With vertical drop panel
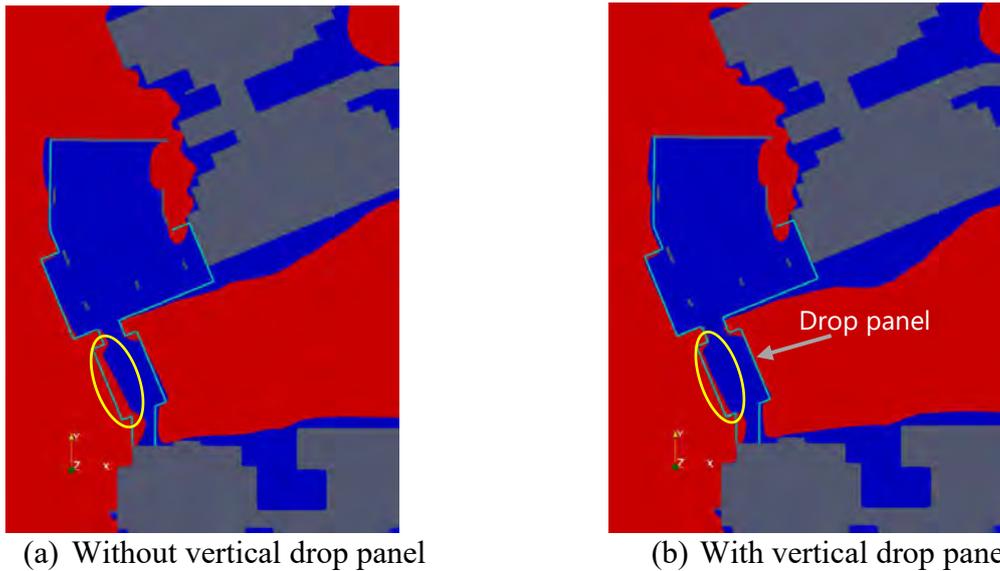
Figure 7. Wetting areas obtained from Eulerian-Eulerian approach.

Conclusions

In this paper, the two different CFD-based solution procedures for WDR studies, which have been developed recently under OpenFOAM v2012 framework [6], are presented with the solutions for four various environmental test cases. The accuracy, efficiency, and applicability of both CFD approaches, in particular about the Eulerian-Lagrangian approach, are well addressed throughout the paper, from the design concerns in Singapore about WDR studies.

In summary, the key innovative features integrated into the CFD-based solution procedures are listed as follows:

- Two-step sequential solution procedure for WDR simulation: Steady-state airflow simulations, followed by one-way coupling WDR simulations;
- Lagrangian and Eulerian solvers have been developed and customized for simulating WDR impacts, subject to Singapore's special concern about the wetting-related safety;
- Additional user-friendly features which are developed for Eulerian-Lagrangian approach, including
  - ✓ Smart selection of raindrop release plane
  - ✓ Catch ratio calculation based on mass conservation for stream tubes bounded by raindrops
  - ✓ Erosion rate as wetting marker
  - ✓ Particle trajectory calculation base on decomposed solution data for fast processing

Case studies here demonstrate that the developed CFD-based approaches can be used as the design toolkit for accurately assessing the WDR impacts, and subsequently evaluating and improving the mitigation measures for WDR control.

Acknowledgments

References

[1] David Segersson. Numerical quantification of driving rain on buildings, RMK No. 103, December 2003.

[2] BCA GREEN MARK: Technical Guide and Requirements, The Building and Construction Authority of Singapore, 2021.

[3] Lakshmi Sutha Kumar, Yee Hui Lee and Jin Teong Ong. Truncated Gamma Drop Size Distribution Models for Rain Attenuation in Singapore. *IEEE Transactions on antennas and propagation*, Vol. 58, No. 4 (2010).

[4] Kubilay, A., Derome, D., Blocken, B. & Carmeliet, J. CFD simulation and validation of wind-driven rain on a building facade with an Eulerian multiphase model. Building and environment 61, 69-81 (2013).

[5] Blocken, B. & Carmeliet, J. A review of wind-driven rain research in building science. *Journal of wind engineering and industrial aerodynamics*, Vol. 92, 1079-1130 (2004).

[6] User Guide, OpenFOAM v2012, https://www.openfoam.com/news/main-news/openfoam-v20-12.

[7] Elghobashi, S. On predicting particle-laden turbulent flows, *Applied Scientific Research*, Vol. 52, 309-329 (1994).

[8] Richards, P. and Hoxey, R. Appropriate boundary conditions for computational wind engineering models using the k-ϵ turbulence model. *Journal of wind engineering and industrial aerodynamics*, Vol. 46, 145-153 (1993).

[9] Gunn, R. and Kinzer, G. D. The terminal velocity of fall for water droplets in stagnant air. *Journal of Meteorology*, Vol. 6, 243-248 (1949).

[10] George Xu, Kendrick Tan, Zhengwei Ge, Hee Joo Poh, Chin Chun OOI, Yong ENG, Automatic selection of release plane for Lagrangian-based wind-driven rain studies, *International Journal of Wind Engineering and Industrial Aerodynamics*, 232 (2023) 105242. https://doi.org/10.1016/j.jweia.2022.105242

[11] Tang, W. and Davidson, C. I. Erosion of limestone building surfaces caused by wind-driven rain: 2. Numerical modeling. *Atmospheric Environment*, Vol. 38, 5601-5609 (2004).

[12] Kubilay, A., Derome, D., Blocken, B., Carmeliet, J., Wind-driven rain on two parallel wide buildings: field measurements and cfd simulations. *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 146, 11–28 (2015).

Dynamic mode decomposition analysis of turbulent flow around a surface-piercing finite circular cylinder

*Songtao Chen, Weiwen Zhao, †Decheng Wan

Computational Marine Hydrodynamics Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

*Presenting author: songtao.chen@sjtu.edu.cn
†Corresponding author: dcwan@sjtu.edu.cn

Abstract

The turbulent flow past a surface-piercing finite cylinder at $Re = 2.7 \times 10^5$ and $Fr = 1.1$ is analyzed in depth using the dynamic mode decomposition (DMD) method, where the velocity fields at three representative depths serve as two-dimensional snapshots. According to the frequency spectra, the dominant DMD modes can be divided into low-frequency ($St < 0.1$) and vortex shedding ($0.1 < St < 0.3$) regions. In the low-frequency region, the two clusters of DMD modes are due to free shear layers ($\sim St = 0.01$) and the free end ($\sim St = 0.06$), respectively. Furthermore, the disappearances of vortex shedding modes at $z/D = -3.5$ and free end modes at $z/D = -0.75$ indicate the inhibition of Karman vortex shedding near the free end and the negligible free-end effect near the free surface.

Keywords: Dynamic mode decomposition, turbulent flow, finite circular cylinder, free surface, free shear layers, vortex shedding

Introduction

With the rapid development of high-performance computers and experimental techniques, computational fluid dynamics (CFD) and particle image velocimetry (PIV) can obtain high-resolution flow field data for analysis. Nevertheless, how to further utilize massive data effectively has become a troublesome issue. In this regard, the dynamic mode decomposition (DMD) method [1] has grown in popularity in recent years. Among the various modal analysis methods [2][3], the DMD method can capture dynamic modes with associated frequencies and growth rates using linear approximation. In contrast to the spatial orthogonality of the proper orthogonal decomposition (POD) method [4], this characteristic makes the DMD method suitable for flow mechanism analysis and short-term forecasting.

Nowadays, the DMD method has been widely used in high-fidelity numerical simulations of moderate and high Reynolds number flows. Gadalla et al. used large eddy simulation and Ffowcs Williams and Hawkings (FWH) analogy to calculate the hydroacoustic noise of the flow around a sphere at $Re = 5000$ [5]. In their study, they focused on constructing reduced-order models using the DMD and POD methods and evaluating their ability to predict noise. Yin and Ong used the Spalart-Allmaras delayed detached eddy simulation (SA-DDES) to simulate the flow around a wall-mounted square structure at $Re = 1.19 \times 10^5$ [6]. For the wake analysis, they applied the DMD method to the two-dimensional snapshots at the mid-span point. By combining the spectra of velocity components at several points and the distribution of the dominant DMD modes in the frequency domain, they successfully identified four primary modes: low-frequency, subharmonic KH, KH, and secondary harmonic modes. In addition to the turbulent flow around a bluff body, the DMD method has also been applied to

the turbulent wall-bounded flows. Sayadi et al. used direct numerical simulation (DNS) and LES with a dynamic subgrid-scale model to investigate the H-type transition in a flat-plate boundary layer [7]. They performed the DMD analysis on the two-dimensional skin friction data and three-dimensional Q-criterion data. By comparing LES and DNS results, the under-prediction of turbulent skin friction can be attributed to the low-frequency modes, which account for most of the Reynolds shear stress gradient in the boundary layer.

In the present study, the DMD method is applied to the turbulent flow past a surface-piercing finite cylinder at $Re = 2.7 \times 10^5$ and $Fr = 1.1$. The primary objective of this paper is to gain new insights into the flow characteristics using the DMD method. The remainder of this paper is organized as follows. First, the numerical algorithm of the DMD is introduced. Then, the dominant DMD modes on three representative planes are shown to reveal the local flow behaviors. Finally, the main conclusions are drawn.

Numerical method

In the present study, the dynamic mode decomposition (DMD) method is used to analyze the wake field. The basic idea behind the DMD is to extract dynamic modes based on the linear approximation of the system, each of which has a specific frequency and growth rate. In this section, the algorithm of the DMD is introduced in detail.

For a target system, the DMD uses a linear operator $\mathbf{A}$ to approximate the evolution of the non-linear system:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k, \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the snapshot representing the discrete state of the system at time k, $k \in [1, ..., m]$ is the snapshot number, and n is the dimension of each individual snapshot. On this basis, arrange a sequence of snapshots into the following two matrices by columns:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{m-1}], \ \mathbf{X}' = [\mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_m], \tag{2}$$

where the time interval between adjacent snapshots is Δt. Then, the linear operator $\mathbf{A}$ can be approximated as:

$$\mathbf{A} \approx \mathbf{X}'\mathbf{X}^\dagger \in \mathbb{R}^{n \times n}, \tag{3}$$

where $\mathbf{X}^\dagger$ is the Moore-Penrose pseudoinverse of $\mathbf{X}$. In fluid mechanics problems, the dimension of an individual snapshot n is usually much larger than the total number of snapshots m. Therefore, the high-dimensional operator $\mathbf{A}$ is usually not obtained by directly solving Eq. (3). Instead, a low-dimensional operator $\tilde{\mathbf{A}}$ that has the same eigenvalues as $\mathbf{A}$ is considered. First, the singular value decomposition (SVD) is applied to the matrix $\mathbf{X}$ to compute the truncated low-rank approximation:

$$\mathbf{X} \approx \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^*, \tag{4}$$

where $r \leq m-1$ is the SVD truncation rank and $*$ denotes the conjugate transpose. Among them, the columns of the matrix $\mathbf{U}_r \in \mathbb{C}^{n \times r}$ are the proper orthogonal decomposition (POD) modes, which are orthogonal to each other. Then, the low-dimensional operator $\tilde{\mathbf{A}}$ can be obtained using $\mathbf{U}_r$:

$$\tilde{\mathbf{A}} = \mathbf{U}_r^* \mathbf{X}' \mathbf{V}_r \boldsymbol{\Sigma}_r^{-1} \in \mathbb{C}^{r \times r}. \tag{5}$$

Since $\tilde{\mathbf{A}}$ has the same eigenvalues as $\mathbf{A}$, the eigenvalues of $\mathbf{A}$ can be computed as follows:

$$\tilde{\mathbf{A}}\mathbf{W} = \mathbf{W}\boldsymbol{\Lambda}, \tag{6}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_r] \in \mathbb{C}^{r \times r}$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_r) \in \mathbb{C}^{r \times r}$. The corresponding eigenvectors of $\mathbf{A}$ can be obtained:

$$\boldsymbol{\Phi} = \mathbf{U}_r \mathbf{W}, \tag{7}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, ..., \boldsymbol{\varphi}_r] \in \mathbb{C}^{r \times r}$ are also known as the DMD modes. In addition, the frequency f and growth rate σ of the DMD mode can be described as follows:

$$f = \text{Im}(\log(\lambda_i) / \Delta t), \tag{8}$$

$$\sigma = \text{Re}(\log(\lambda_i) / \Delta t), \tag{9}$$

where Im and Re represent the imaginary and real parts, respectively. For the implementation of the above algorithm, the pyDMD Python package [8] is used to perform the following DMD analysis.

Results and discussion

In the present study, the turbulent flow past a surface-piercing finite cylinder at $Re = 2.7 \times 10^5$ and $Fr = 1.1$ is analyzed using the DMD method. The laboratory experiments were performed in the towing tank at the Iowa Institute of Hydraulics Research (IIHR) [9]. The diameter of the cylinder is $D = 0.2$ m, and the immersed length h is $4D$. The delayed detached-eddy simulation (DDES) method based on the shear stress transport (SST) model [10] is used to simulate this large-scale separation flow, and the piecewise-linear interface calculation (PLIC) technique is used to capture the interface. For a high-fidelity numerical simulation, the total number of the structured computational mesh is 14.66×10^6. The detailed computational setup and validation can refer to our previous work [11].

To perform the DMD analysis, two-dimensional snapshots of the velocity field on three representative horizontal planes are collected, as shown in Fig. 1. Each snapshot has about 7.2×10^4 elements. The time span of the snapshots is 9.9 s after the turbulent flow has fully developed, and the uniform time interval Δt between each snapshot is 0.01 s. Figs. 2-4 show the eigenvalue distributions of the dominant DMD modes on the complex plane at $z/D = -0.75$, -2, and -3.5, respectively. To extract the dominant modes, the truncation rank of the reduced SVD approximation depends on the number of singular values reaching the total energy of 0.99. In addition, the modes are ordered by their frequency. As shown in the enlargements of Figs. 2-4, Mode 1 corresponds to the static mode, i.e., the time-averaged flow field, whose imaginary part is zero. Other complex-conjugate modes are all located near the unit circle, indicating that their oscillatory states are stable. Moreover, the number of the dominant DMD modes decreases as the depth approaches the free surface and free end, suggesting the decreasing complexity of the local flow behavior.
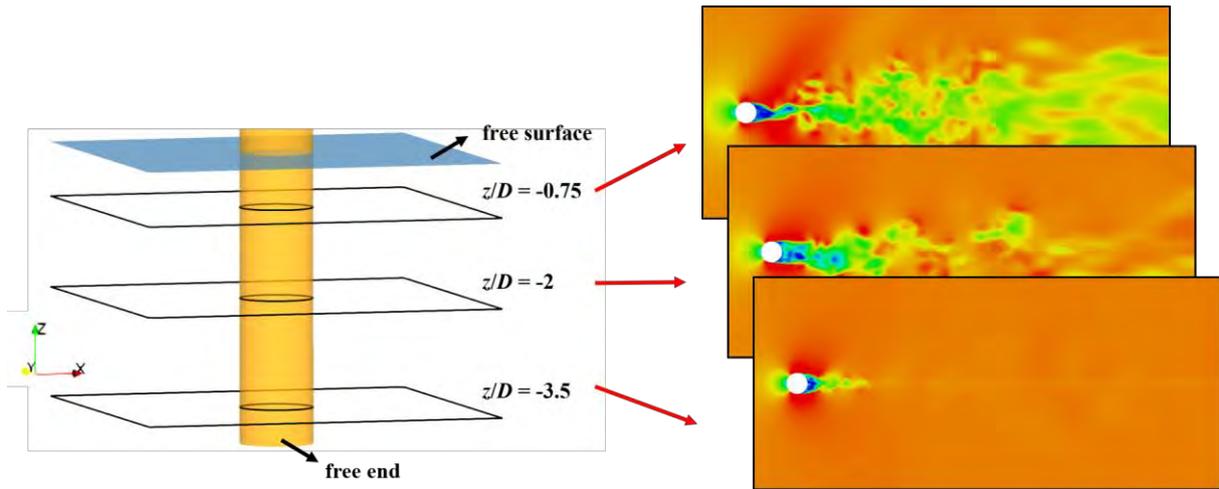
Figure 1. DMD analysis of the velocity field on three representative horizontal planes.
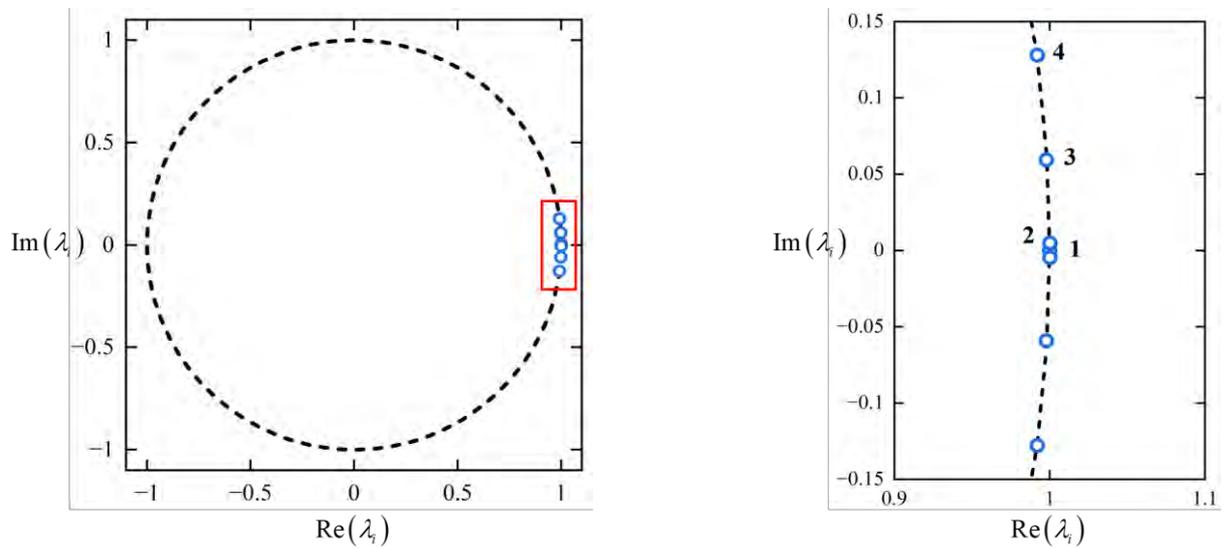
Figure 2. Eigenvalue distribution of dominant DMD modes on the complex plane at *z/D* = −0.75. (The left is the overall view, and the right is the enlarged view.)
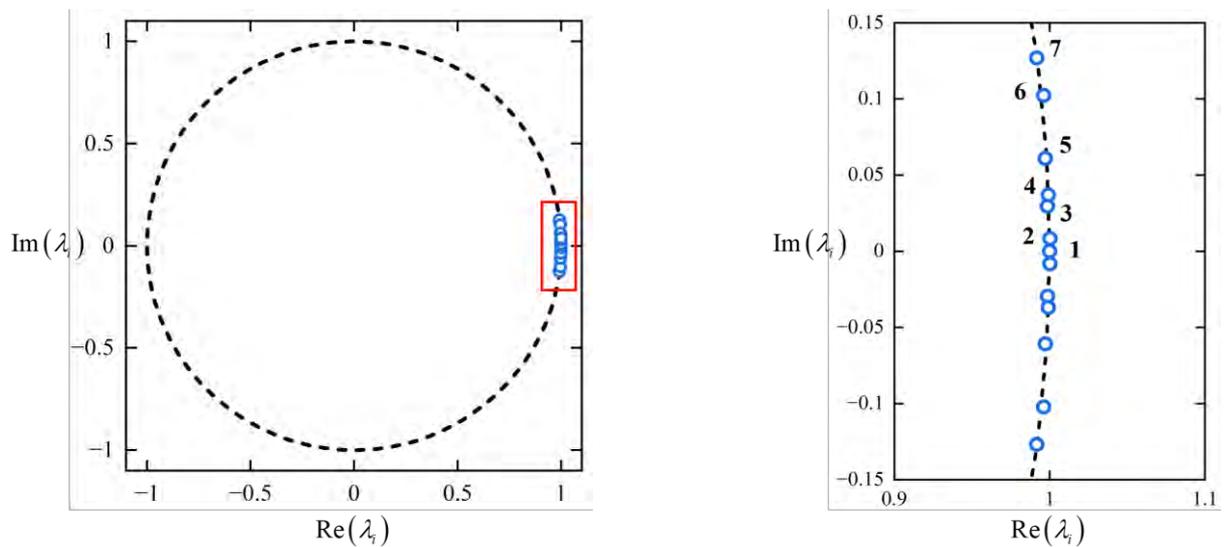
Figure 3. Eigenvalue distribution of dominant DMD modes on the complex plane at *z/D* = −2. (The left is the overall view, and the right is the enlarged view.)
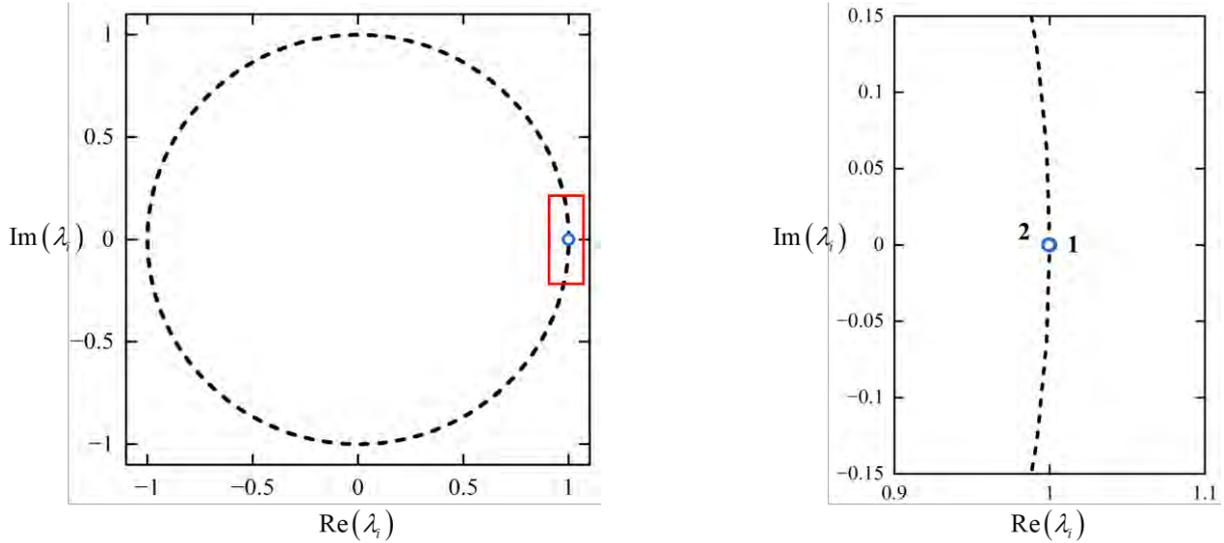
Figure 4. Eigenvalue distribution of dominant DMD modes on the complex plane at $z/D = -3.5$. (The left is the overall view, and the right is the enlarged view.)

To further investigate the flow characteristic of each mode, the frequency spectra of the dominant DMD modes and the lift coefficient are shown in Fig. 5. The lift coefficient of the cylinder is defined as:

$$C_L = \frac{F_L}{0.5\rho U_\infty^2 A} \tag{10}$$

where F_L is the lift force, ρ is the water density, U_∞ is the inflow velocity, and $A = Dh$ is the projected area at still water. On the other hand, the frequency f takes the non-dimensional form of the Strouhal number:

$$St = \frac{fD}{U_\infty} \tag{11}$$

Since the frequency of the static mode is zero, only the oscillatory modes are shown here. However, the oscillatory mode at $z/D = -3.5$ is not shown because its frequency is close to zero.

According to previous experimental measurements of flow around a circular cylinder [12], the Strouhal number in the critical regime ($2 \times 10^5 < Re < 3.5 \times 10^5$) varies between 0.2 and 0.3. Therefore, multiple peaks of the lift coefficient spectrum can be observed in this region, defined as the vortex shedding region. On the other hand, based on our previous fast Fourier transform (FFT) analysis of pressure and velocity [11], the significant peak around $St = 0.06$ in the low-frequency region can be attributed to the free-end effect. At the medium depth $z/D = -2$, the oscillatory DMD modes are mainly in the low-frequency and vortex shedding regions. Furthermore, each mode in the vortex shedding region exactly corresponds to a peak of the lift coefficient spectrum. As for the low-frequency region, the modes around $St = 0.01$ can be attributed to free shear layers, which will be visualized later. When approaching the free surface, the oscillatory modes near the significant peak disappear, indicating the free end effect is negligible at $z/D = -0.75$.
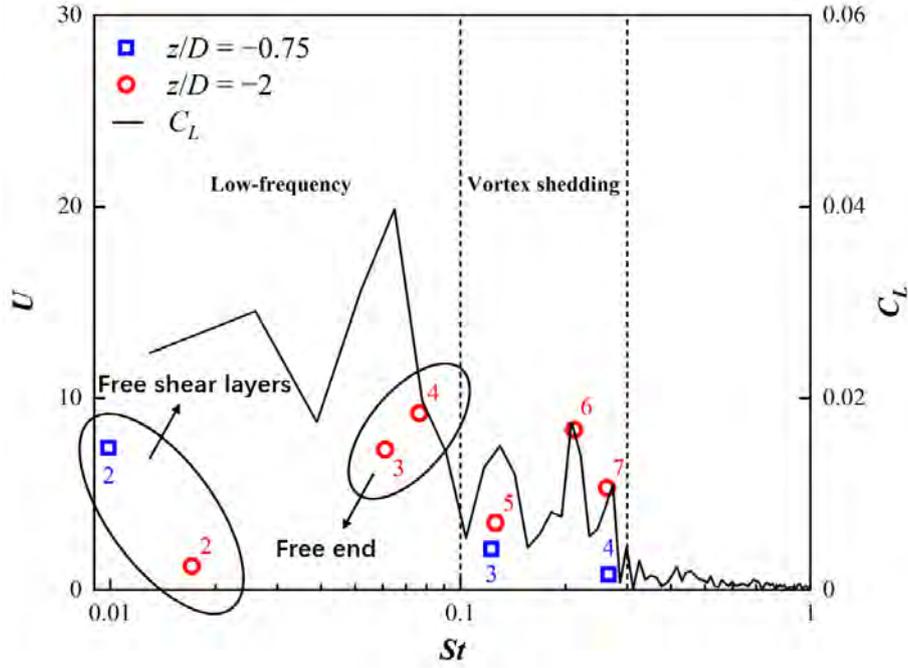
Figure 5. Frequency spectra of dominant DMD modes and lift coefficient.

Figs. 6-8 further show the contours of each velocity component of the dominant DMD modes. It can be observed intuitively that the flow pattern of Mode 1 (static mode) at each depth is the same as the time-averaged flow field. As for Mode 2, large-scale structures can be observed, especially in the streamwise contours. Correspondingly, Fig. 9 shows the contours of variance of streamwise velocity fluctuation, where the high magnitude regions indicate the presence of symmetric free shear layers. By comparison, it can be found that the pattern of Mode 2 is similar to the shape of free shear layers. Therefore, we can conclude that the low-frequency modes around $St = 0.01$ are due to large-scale symmetric free shear layers. For Modes 3 and 4 at $z/D = -2$, Figs. 7(c) and (d) show that the length scales of the primary structures are also large, which can be attributed to the inclined large-scale arch vortices caused by the free end [11]. In contrast, the wavelength of the structures in vortex shedding modes becomes smaller, and the pattern becomes alternatively regular, especially as shown in Figs. 6(d) and 7(g). This characteristic is due to the alternative Karman vortex shedding, causing strong fluctuations even in the far wake. However, there is no vortex shedding mode at $z/D = -3.5$, implying that the Karman vortex shedding is suppressed by the nearby free end.

(a)

(b)
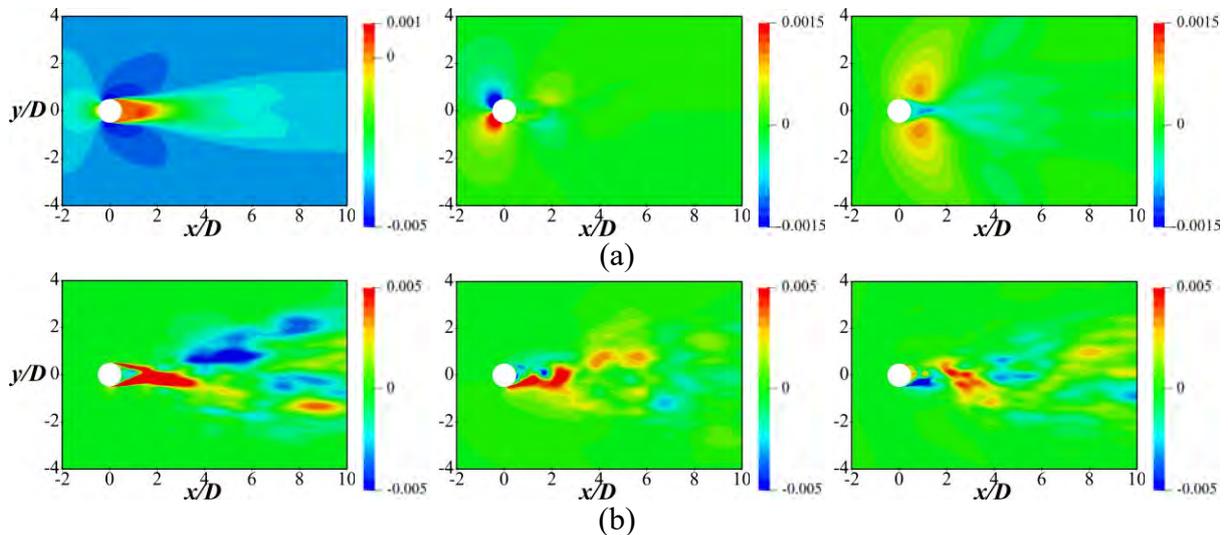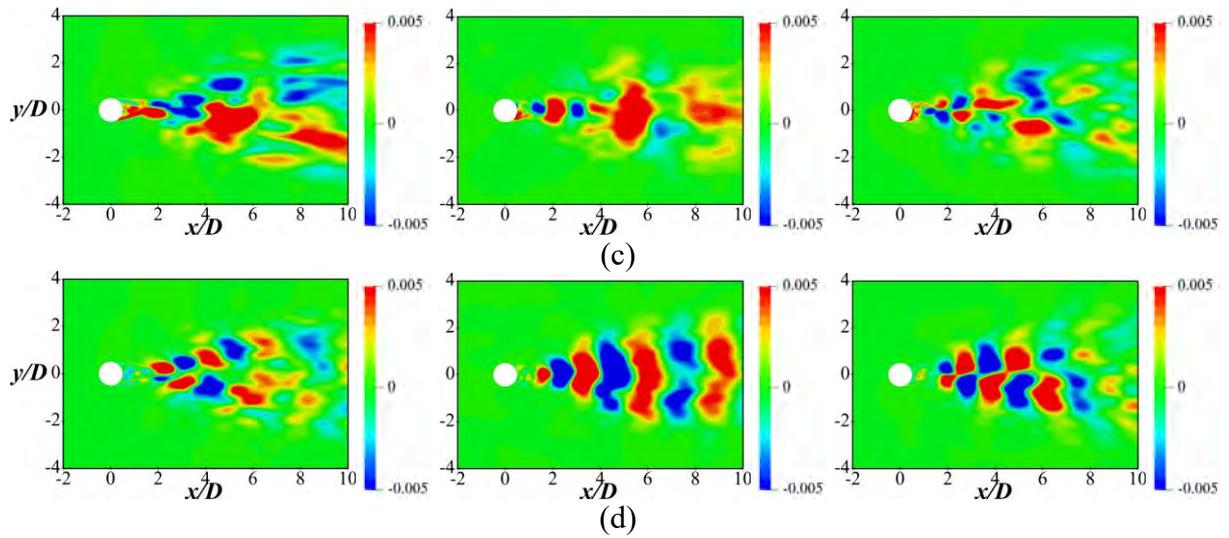
Figure 6. Contours of each velocity component of the dominant DMD modes at *z/D* = −0.75: (a) Mode 1, (b) Mode 2, (c) Mode 3, and (d) Mode4. (The left column is the streamwise velocity, the middle column is the cross-stream velocity, and the right column is the spanwise velocity.)
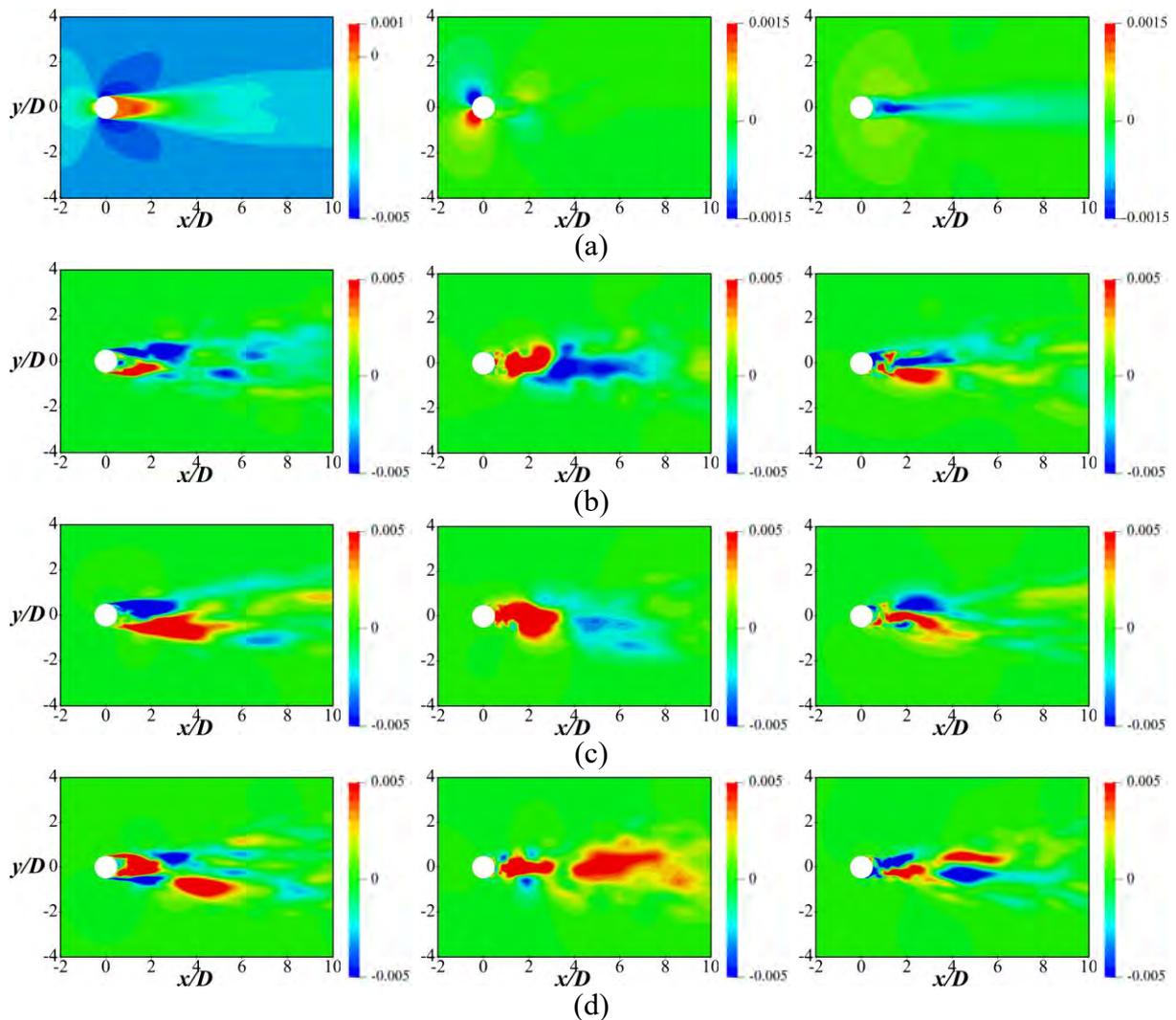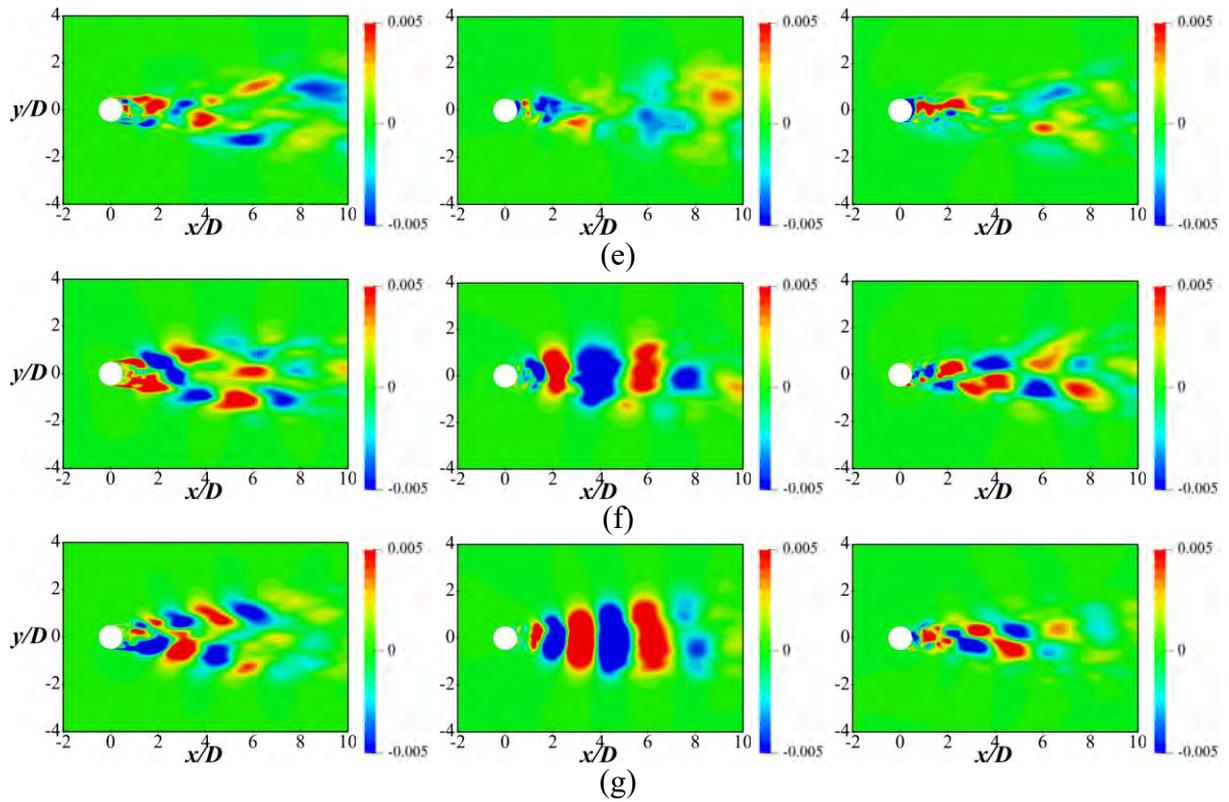
Figure 7. Contours of each velocity component of the dominant DMD modes at *z/D* = −2: (a) Mode 1, (b) Mode 2, (c) Mode 3, (d) Mode 4, (e) Mode 5, (f) Mode 6, and (g) Mode 7. (The left column is the streamwise velocity, the middle column is the cross-stream velocity, and the right column is the spanwise velocity.)
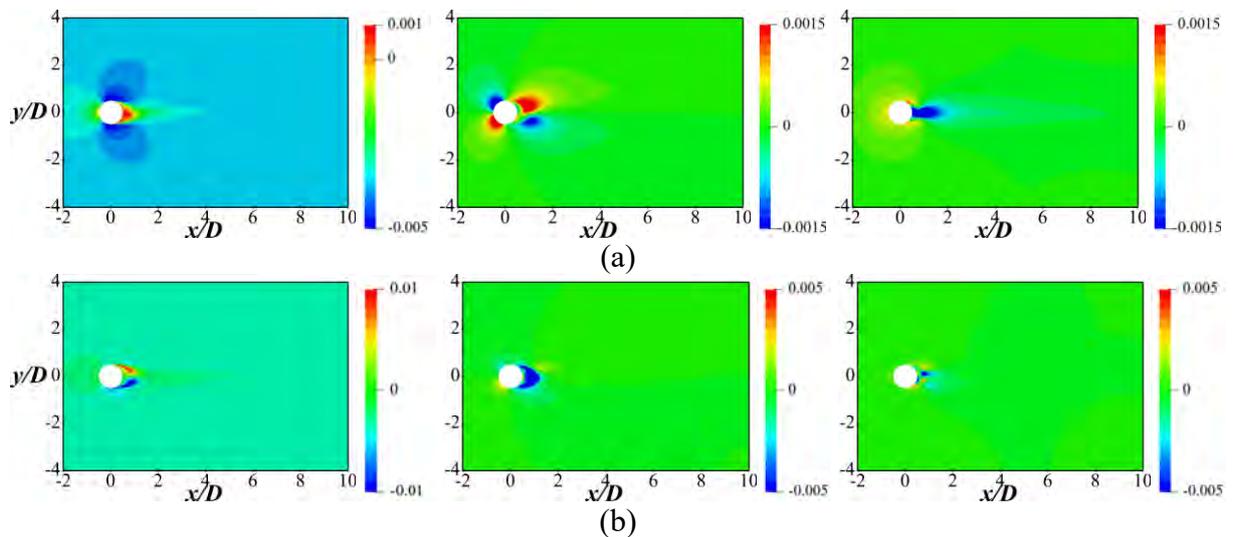
Figure 8. Contours of each velocity component of the dominant DMD modes at *z/D* = −3.5: (a) Mode 1 and (b) Mode 2. (The left column is the streamwise velocity, the middle column is the cross-stream velocity, and the right column is the spanwise velocity.)
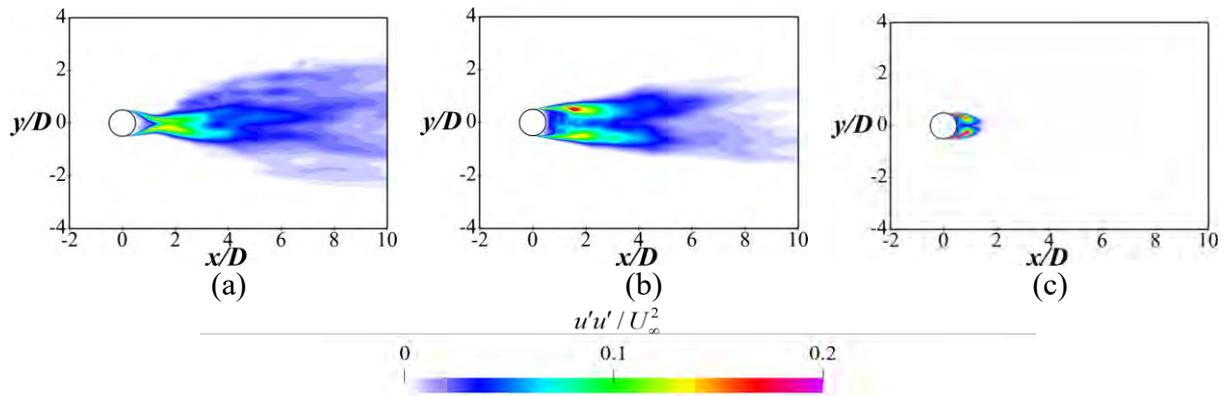
Figure 9. Contours of variance of streamwise velocity fluctuation at (a) $z/D = -0.75$**, (b)** $z/D = -2$**, and (c)** $z/D = -3.5$**.**

Conclusions

In this paper, the DMD method is used to provide an in-depth analysis of the turbulent flow past a surface-piercing finite cylinder at $Re = 2.7 \times 10^5$ and $Fr = 1.1$. The velocity fields are sampled on three representative horizontal planes ($z/D = -0.75$, -2, and -3.5) as two-dimensional snapshots. The main conclusions are as follows.

The frequency spectra show that the dominant DMD modes can be roughly divided into low-frequency and vortex shedding regions. In the low-frequency region, the modes around $St = 0.01$ are due to the large-scale symmetric free shear layers, while the others around $St = 0.06$ are due to the free end effect. For the vortex shedding region, the modes cluster in the range of $0.1 < St < 0.3$. Furthermore, the wavelength of the structures in each vortex shedding mode becomes smaller, and the pattern exhibits an alternative regular distribution due to the Karman vortex shedding. In addition, the vortex shedding modes disappear at $z/D = -3.5$, indicating the inhibition of the Karman vortex shedding near the free end. On the other hand, the low-frequency modes around $St = 0.01$ disappear at $z/D = -0.75$, suggesting the negligible free-end effect.

In the near future, we plan to use three-dimensional snapshots instead to perform the DMD analysis. For such flows with significant three-dimensional effects, this improvement may help to better characterize the flow pattern of each mode.

Acknowledgements

References

[1] Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics* **656**, 5-28.
[2] Taira, K., Brunton, S. L., Dawson, S. T., Rowley, C. W., Colonius, T., McKeon, B. J., ... and Ukeiley, L. S. (2017). Modal analysis of fluid flows: An overview. *AIAA JOURNAL* **55**(12), 4013-4041.

[3] Taira, K., Hemati, M. S., Brunton, S. L., Sun, Y., Duraisamy, K., Bagheri, S., ... and Yeh, C. A. (2020). Modal analysis of fluid flows: Applications and outlook. *AIAA JOURNAL* **58**(3), 998-1022.

[4] Lumley, J. L. The structure of inhomogeneous turbulent flows. *Proceedings of the International Colloquium on the Fine Scale Structure of the Atmosphere and Its Influence on Radio Wave Propagation*, 1967, 166–178.

[5] Gadalla, M., Cianferra, M., Tezzele, M., Stabile, G., Mola, A., and Rozza, G. (2021). On the comparison of LES data-driven reduced order approaches for hydroacoustic analysis. *Computers and Fluids* **216**, 104819.

[6] Yin, G., and Ong, M. C. (2021). Numerical analysis on flow around a wall-mounted square structure using Dynamic Mode Decomposition. *Ocean Engineering* **223**, 108647.

[7] Sayadi, T., Nichols, J. W., Schmid, P. J., and Jovanovic, M. R. Dynamic mode decomposition of H-type transition to turbulence. *Proceedings of the 65th Annual Meeting of the APS Division of Fluid Dynamics*, San Diego, California, November 18–20, 2012, 5-14.

[8] Demo, N., Tezzele, M., and Rozza, G. (2018) PyDMD: Python dynamic mode decomposition. *Journal of Open Source Software* **3**(22), 530.

[9] Stern, F. (2016) Integrated High-Fidelity Validation Experiments and LES for a Surface-Piercing Truncated Cylinder for Sub and Critical Reynolds and Froude Numbers. In *NATO Specialist Meeting*.

[10] Gritskevich, M. S., Garbaruk, A. V., Schütze, J., and Menter, F. R. (2012) Development of DDES and IDDES formulations for the k-ω shear stress transport model. *Flow, turbulence and combustion* **88**, 431-449.

[11] Chen, S., Zhao, W., and Wan, D. (2022) Turbulent structures and characteristics of flows past a vertical surface-piercing finite circular cylinder. *Physics of Fluids* **34**(1), 015115.

[12] Tropea, C., Yarin, A. L., and Foss, J. F. (2007) *Springer handbook of experimental fluid mechanics*. Springer Berlin, Heidelberg, Germany.

Physics informed neural learning of wavefields using Gabor basis functions

*Tariq Alkhalifah[1] and Xinquan Huang[1]

[1]Physical sciences and engineering (PSE), King Abdullah University of science and technology, Saudi Arabia

Presenting and corresponding author: tariq.alkhalifah@kaust.edu.sa

Abstract

Physics informed neural networks (PINNs) functional solutions of partial differential equations (PDEs) have recently garnered a lot of attention due to their flexibility and interpolation abilities. They, however, are expensive to train especially for complex functions like wavefields. We propose to improve the efficiency and accuracy for such wavefield solutions by formulating them as linear combinations of Gabor basis functions that satisfy the wave equation. For the Helmholtz equation, this can be accomplished by outfitting the fully-connected neural network model with a learnable Gabor layer prior to the last hidden layer with linear connections to the output. So, the neurons with nonlinear activations prior to the Gabor layer are tasked with predicting the amplitude of the Gabor function. Tests on realistic examples demonstrate its effectiveness compared to the Vanilla PINN implementation.

Keywords: PINNs, wave equation, Gabor transforms.

Introduction

Recently, machine learning (ML) and specifically deep neural networks (DNNs) have made tremendous inroads in helping us solve physical problems. Considering that DNNs can act as universal function approximaters [1], and considering that automatic differentiation (AD), which is available in almost all ML packages [2], provide a mix of numerical and symbolic based differentiation of the function with respect to the input, [3] used these features to demonstrate the network's capability in learning functional solutions to nonlinear partial differential equations, and referred to this process as physics-informed neural networks (PINNs). The functional representation has shown considerable flexibility in handling irregular domains and providing compact solutions. However, PINNs are still trying to carve out their territory as a credible alternative to numerical methods, considering the high training cost. In our domain, PINNs were utilized to solve the eikonal equation [5], as well as the Helmholtz equation [6]. However, in all these applications, though the NN solution demonstrated flexibility, it was not that efficient to obtain.

We aim here to improve the efficiency and accuracy for such wavefield solutions by formulating them as linear summations of Gabor functions that satisfy the wave equation. This is a natural extension as Gabor functions have proven to be optimal basis functions for wavefields [4]. So, we train the network to predict wavefield solutions as a linear summation of learnable Gabor functions that satisfy the Helmholtz equation. From the examples, we aim to demonstrate that this implementation alone provides considerable accuracy and efficiency improvements over the conventional implementation.

Methodology

A Gabor function represents a plane wave with a specific wavenumber (k), phase (ϕ)and direction (θ) modulated by a Gaussian weight function at a particular location (μ) in the domain

of interest specified by a variance (α) to localize the plane wave. Such a function with the proper wavelength satisfy the wave equation and have been effective basis functions for complex wavefields [4]. For 2D media, specified by x laterally and z depth, a Gabor function is given by:

$$G(k, x, z) = e^{-\frac{(\tilde{x}-\mu_1)^2 + (\tilde{z}-\mu_2)^2}{\alpha}} \cos(k\tilde{x} + \phi), \tag{1}$$

where

$$
\begin{aligned}
\tilde{x} &= x\cos\theta + z\sin\theta, \\
\tilde{z} &= z\cos\theta - x\sin\theta.
\end{aligned}
\tag{2}
$$

We incorporate the Gabor function 1 into the $L-1$ layer of a fully connected neural network, where L represents the number of hidden layers. The output of the $L-2$ layer (as well as previous layers) is given by

$$Z_{i+1} = \sigma(W_i Z_i + b_i), \tag{3}$$

where i is the layer index, which is $L-2$, σ is an activation function, W_i are the weights, and b_i are the biases for that layer. This output is used as an amplitude weight for the Gabor functions in which the output of the $L-1$ layer is given

$$Z_L = G Z_{L-1} + b_{L-1}, \tag{4}$$

where the Gabor functions will learn a unique θ, σ, μ, and ϕ for each neuron within the $L-1$ layer. Finally, the last hidden layer is linearly connected to the output, the real and imaginary parts of the wavefield. This implies that the output wavefield is linear summation of Gabor functions. The wavenumber of the Gabor function is given by the

$$k = \frac{\omega}{v(x, z)}, \tag{5}$$

where ω represents the frequency we are solving for, and v is the velocity corresponding to the input coordinates in 2D $\{x, z\}$. Thus, the input coordinates to the NN function also contributes to the Gabor layer through the velocity and through equation 2. Including the frequency in the Gabor function is important for the speedy convergence of PINNs [7].

Using the Helmholtz equation in 2D:

$$\nabla^2 u(x, z) + k^2 u(x, z) = f(x, z), \tag{6}$$

as the loss term to train PINNs has admitted wavefield solutions given by neural networks with input the coordinate values and outputs the wavefield's real and imaginary values corresponding to that location and source, acting as a function. We train such PINNs using random inputs from the domain of interest (regardless of the shape, which promotes flexibility) and after training, we usually evaluate the NN at a regular grid for a single source for ease of visualization. Nevertheless, the wavefield is stored in an NN.

The PINN network is given by fully connected hidden layers with activation functions in every layer other than the last. So, to adapt PINNs to wavefields, we multiply the layer before last with learnable Gabor functions in which the frequency is the only parameter fixed, and thus, these Gabor functions satisfy the corresponding Helmholtz equation for that frequency. These Gabor functions, with amplitudes given by the neurons output from the previous layers, are linearly connected to last hidden layer, which is connected to the output wavefield value, also

linearly (Figure 1). The additional last hidden layer allows for summation of Gabor functions to provide additional varieties for the output.
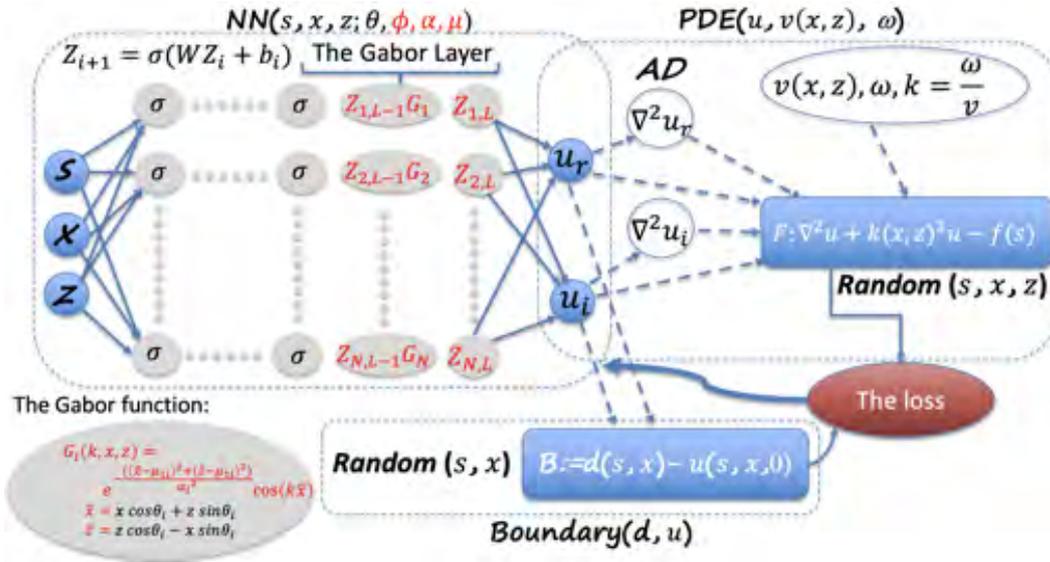
Figure 1: A schematic diagram of the new PINN architecture with Gabor functions. The framework includes three parts, the network NN, and the constraint PDE, with its parameters, and the boundary condition. The network takes (for training) random input source location at the surface and lateral and depth coordinates, and outputs the real and imaginary parts of the wavefield that satisfies, after training, the Helmholtz equation F, and the boundary condition B. The neurons of every layer are like the conventional PINN ones with activation functions (shown on the top left), other than the last two layers with Gabor functions multiplied by the output of these neurons, with neural network weights W and biases b. Also learned are the Gabor parameters other than the frequency. The Gabor function G is shared lower left. AD stands for automatic differentiation.

Results

We test the approach initially to solve the scattered Helmholtz equation for the velocity model in Figure 2a, with a background 1.5 km/s velocity, for a 4 Hz wavefield. The real part of the numerical solution (as reference) is shown in Figure 2b. We compare the performance of this implementation to the vanilla implementation. For 4 Hz, we use a simple network given by 3 hidden layers of 256 neurons in each layer, and we also consider 512 neurons for the vanilla implementation, and use the sine activation function. We train all three implementations (ours-256/134.4k, MLP-256/133.12k, MLP-512/528.39 trainable parameters) using 40000 random samples of (x, z, s), an Adam optimizer, and a learning rate of 0.001.

Figure 2c shows the loss function, which is given mainly by the Helmholtz equation for the three implementations. We can observe the fast convergence of our proposed method. This is also reflected in the predictions shown on Figures d-I after 1500, 3500, and 20000 epochs (left to right) for MLP-256, MLP-512, and our method (top to bottom). The vanilla PINN (MLP-256) struggled in fitting the wavefield even after 20000 epochs. Meanwhile, for the same size network, our proposed method converged fast and provided a reasonable solution to this 3D function (x, z, s). The vanilla PINN with double the neurons (MLP-512), and much higher cost, managed to converge, but at a slower pace. Considering this fast convergence of the Gabor based PINN, we will share results corresponding to large Gabor networks to solve for higher frequency wavefields at the presentation of this work.
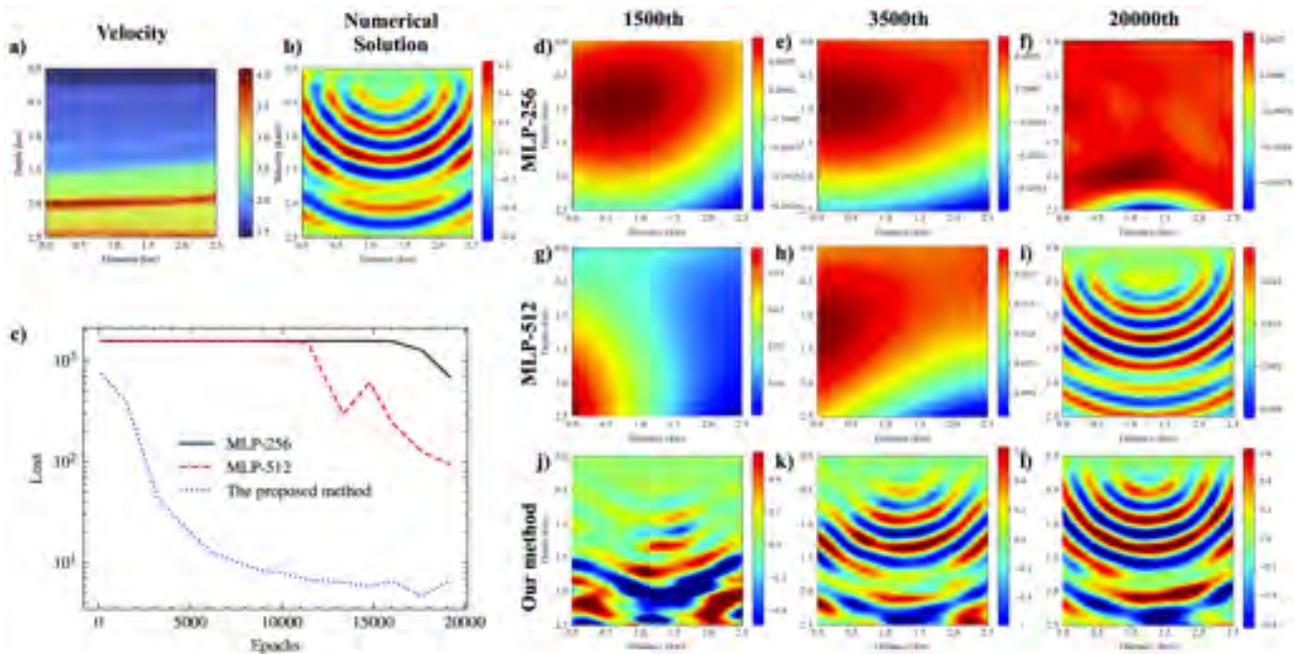
Figure 2: The considered velocity model, and b) the numerical solution. c) The training loss function (satisfying the PDE) for the vanilla PINN with 256 neurons (MLP-256), the vanilla PINN with 512 neurons (MLP-512), the proposed PINN with 256 neurons. The results of the predictions on a regular grid for a source at 1.25 km on the surface are shown in b-j after training for 1500 (first column), 3500 (second column), and 20000 (third column) epochs.

Conclusions

We equipped physics informed neural network with learnable Gabor functions that satisfy the Helmholtz equation. The Gabor functions are placed in the hidden layer before last, and we use linear connections to the output so they maintain their solution property. This implementation allowed for fast convergence of PINN for acoustic wavefield solutions.

References

[1] Pinkus, A., 1999, Approximation theory of the mlp model in neural networks: Acta Numerica, **8**, 143–195.

[2] Baydin, A. G., B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, 2015, Automatic differentiation in machine learning: a survey: arXiv.

[3] Raissi, M., P. Perdikaris, and G. Karniadakis, 2019, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations: Journal of Computational Physics, **378**, 686 – 707.

[4] Arnold, J., 1995, Phase-space localization and discrete representations of wave fields: J. Opt. Soc. Am., **12**, 111–123.

[5] bin Waheed, U., E. Haghighat, T. Alkhalifah, C. Song, and Q. Hao, 2021, Pinneik: Eikonal solution using physics-informed neural networks: Computers & Geosciences, **155**, 104833.

[6] Alkhalifah, T., C. Song, U. bin Waheed, and Q. Hao, 2021, Wavefield solutions from machine learned functions constrained by the helmholtz equation: Artificial Intelligence in Geosciences, **2**, 11–19.

[7] ———, 2021b, Single reference frequency loss for multi-frequency wavefield representation using physics-informed neural networks: Presented at the NeurIPS 2021 AI for Science Workshop.

Cell Mechanic based on a centroidal void cylindrical Tensegrity Model to evaluate the Vibration of a Cellular Cytoskeleton

Eiji Nouchi[1], Tomoteru Oka[2], Noriyuki Kataoka[3], Yoshihisa Kawano[4],

*†Buntara S. Gan[1]

[1]Department of Architecture, College of Engineering, Nihon University, Japan
[2]Department of Mechanical Engineering, Graduate School of Engineering, Nihon University, Japan
[3]Department of Mechanical Engineering, College of Engineering, Nihon University, Japan
[4]Department of Health Science, Faculty of Physical Education, International Pacific University, Japan

*Presenting author: gan.buntarasthenly@nihon-u.ac.jp
†Corresponding author: gan.buntarasthenly@nihon-u.ac.jp

Abstract

Within the cytoplasm of a cell, there is a complex network of protein fibers that help maintain the cell's shape, secure some cells in specific positions, and allow cytoplasm and vesicles to move within the cell. These protein fibers enable cells within multicellular organisms to move. Collectively, this skeleton-like complex network of protein fibers is known as the Cellular Cytoskeleton in biology. There are many structural model hypotheses that scholars have proposed. However, there were constrained to the efforts of abstraction and conception to identify the mechanical behavior of the Cytoskeleton structure. Understanding the mechanical behavior of the Cytoskeleton is highly important to learn the important function of biological processes in a cell to heal humans from diseases by developing the appropriate medicines to cure the broken/infected cell. Moreover, many viruses' structures can be observed and identified from the cellular mechanical point of view. Sophisticated treatment methods to tame the viruses' activities could be discovered in the near future.

There are three types of filamentous proteins: filamentous actin (F-actin), intermediate filaments, and microtubules. The dynamic mechanism of a cellular cytoskeleton is essential for its role as a cell, and its accurate characterization has been a long-standing problem for cell scientists. A cytoskeleton's vibrations are highly influenced by interactions of filamentous proteins mediated by axial vibration of the stiff microtubules (compressive member) and lateral vibration of F-actin (tensile member). Among various structures in a cell, the cytoplasmic contractile bundles, Lamellipodia, and filipodia cells can be modeled by a symmetrical cylinder-shape self-equilibrium Tensegrity with different radii at the top and bottom of the cylinder. The truncated cone-like cylinder model is made to be small in height compared to the radii.

The tensegrity self-vibrational behavior of the Cytoskeleton is investigated to calculate the Cytoskeleton's natural frequencies, which are composed of the individual vibration of microtubules and F-actins experimental data. The Spectral Element Method based on the Wittrick-Williams procedure is adopted to solve the vibrational of the cellular Cytoskeleton. Various n-polygon cylindrical truncated cone-shaped skeletons to mimic the cytoskeletons are presented to demonstrate the robustness of the present study.

Keywords: Cell Mechanic, Tensegrity Structure, Vibration, Cytoskeleton, Spectral Element Method

Introduction

Cells that makeup living organisms are exposed to various mechanical stimuli. In multicellular organisms, a network of cells forms a tissue. A system of tissues forms an organ [3,6,10,25]. Mechanical stimulation is involved in the background of cell morphology, motility, proliferation, and segregation and plays an important role in cell development and homeostasis. Inside a cell, there are organelles. Among the organelles, a cytoskeleton is a membrane that encloses the other organelles. The cellular cytoskeleton determines the cell's shape and triggers cell movement and form changes [1, 5, 19, 31-32, 35]. When receiving an external force, stress is generated in the intracellular cytoskeleton and cell adhesive apparatus. It has become clear that proteins constituting the cytoskeleton and cell adhesive apparatus, such as actin filaments (actin filaments), perceive mechanical stimuli. The actin scaffold undergoes mechanical stimulation in the cytoskeleton and is rapidly remodeled [47]. The role of this structure in mechanical response and its molecular mechanism has attracted attention [10,22,33,42,50-51]. Intermediate filaments have also been suggested to be involved in converting mechanical stimuli into chemical signals. However, although intermediate filaments are thought to play an important role in converting mechanical stimuli into chemical signals, the molecular mechanism remains unclear [1,8,14,21,34,44-45].

In recent years, there has been an idea that the structure of a cytoskeleton is in the form of a tensegrity structure [26-30]. Although tensegrity was originally used in the self-equilibrium concept in the field of structural mechanics, its structure is believed to be applied to the shape of a cell. The cytoskeleton comprises actin filaments, intermediate filaments, and microtubules, forming a three-dimensional filamentous structure inside the cell [11,23]. Microtubules in the cytoskeleton are rigid, and actin and intermediate filaments are elastic, indicating a specific composition of a tensegrity structure.

A tensegrity structure is a statically indeterminate structure composed of continuous cables, tensile members, and discontinuous struts, compressive members, composite structures [12-13,46]. Pins join these members, and the structure is in a self-equilibrium state without being supported. Many research results on tensegrity have been published in mechanics and architecture [15,38]. Various methods have been proposed in architecture for tensegrity structures' morphogenesis and their vibration characteristics [16,18,38-41,48].

In this study, as an initial attempt, we imitate the vibrational behaviors of a cytoskeleton by using a twisted and truncated n-plex cylindrical tensegrity structure. Fig. 1 illustrates the tensegrity model used to mimic the cytoskeleton structure. Sensitivity analyses are conducted to investigate the parameters that rule the vibrational behaviors of the cytoskeleton. In a real application, the parameters can be calibrated by using a result obtained from measurements of the real cytoskeleton.
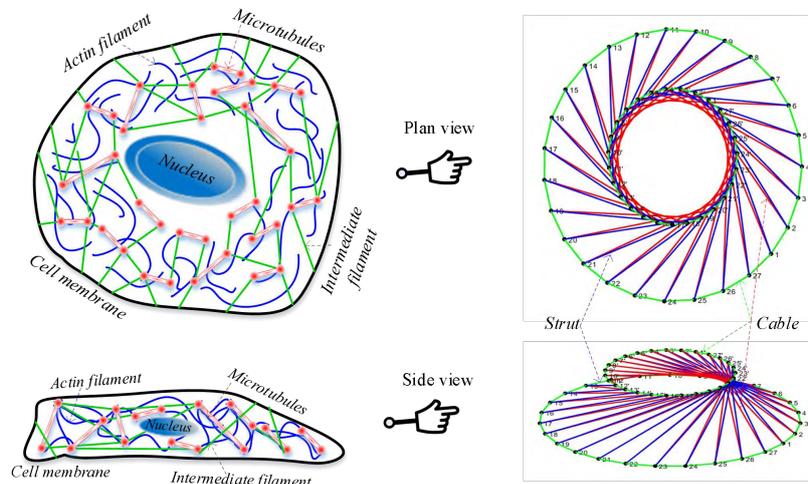
Figure 1. A twisted cylindrical tensegrity model to mimic a cytoskeleton structure

Formulation of Tensegrity in Vibration

Fig. 2 shows a strut and a cable, two basic types of Tensegrity members. The compressive member strut and tensile member cable represent the microtubule and actin filament, respectively. The vibration of the strut is modeled as an axially vibrating microtubule of a cytoskeleton. At the same time, the cable's vibration is modeled to be a laterally vibrating actin filament of a cytoskeleton. The microtubule is presumed to have an axial rigidity represented by prestress force, P, and Young's modulus of elasticity, E. In contrast, the actin filament is presumed to have only an axial rigidity represented by prestress force, P.
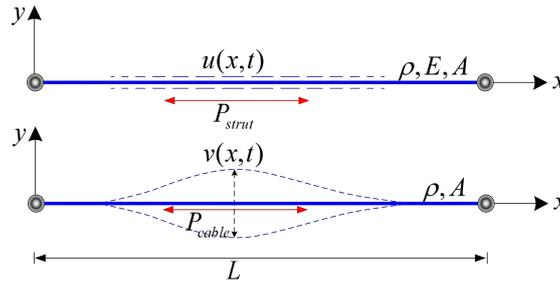
Figure 2. Vibrational models of strut and cable members

The equations of vibrational motion [2] are given as follows:

$$\rho A \ddot{u} - (EA + P_{strut}) u'' = 0 \quad \textit{for strut}$$
$$\rho A \ddot{v} - P_{cable} v'' = 0 \quad \textit{for cable}$$

(1)

where, $u(x,t)$ and $v(x,t)$ are the axial and lateral displacements which are the function of position and time, respectively. The dot and prime superscripts denote the derivatives with respect to time t and spatial coordinate x, respectively.

The internal forces and boundary conditions are given as,

$$N(0,t) = -N_1(t), N(L,t) = N_2(t) \textit{ or } u(0,t) = u_1(t), u(L,t) = u_2(t) \quad \textit{for strut}$$
$$Q(0,t) = -Q_1(t), Q(L,t) = Q_2(t) \textit{ or } v(0,t) = v_1(t), v(L,t) = v_2(t) \quad \textit{for cable}$$

(2)

where $N(x,t), Q(x,t)$ are the axial and shear forces defined by,

$$N(x,t) = (EA + P_{strut}) u'(x,t)$$
$$Q(x,t) = P_{strut} v'(x,t)$$

(3)

Spectral Element Formulation

Analysis was performed using the Spectral Element Method (SEM) [1-2], a combination of the Finite Element Method (FEM), the Dynamic Stiffness Method (DSM), and the Spectral Analysis Method (SAM). The vibration of a tensegrity structure varies with the individual vibrational frequency and the wavelength of the members.

The FEM is a famous computational method used in many fields of engineering and science. However, it is difficult for the conventional FEM to analyze a structure such as a tensegrity

with no supporting boundaries, no external force, and a self-equilibrium state of the members. The DSM uses an exact dynamic stiffness matrix and implicitly considers the mass in the stiffness equation. The SAM uses the Fast Fourier Transform and has the property that the error converges exponentially if the solution is a smooth function ("exponential convergence"), converging much faster than the FEM. In spectral analysis, the dominant differential equation can be solved by infinitely adding waves with different frequencies and the time history of the solution that can be obtained from the Inverse Fourier Transform in the frequency-domain spectral components.

Fig. 3 depicts the concept of SEM as a combination of FEM, SAM, and SAM techniques in computing the vibration of Tensegrity structures.
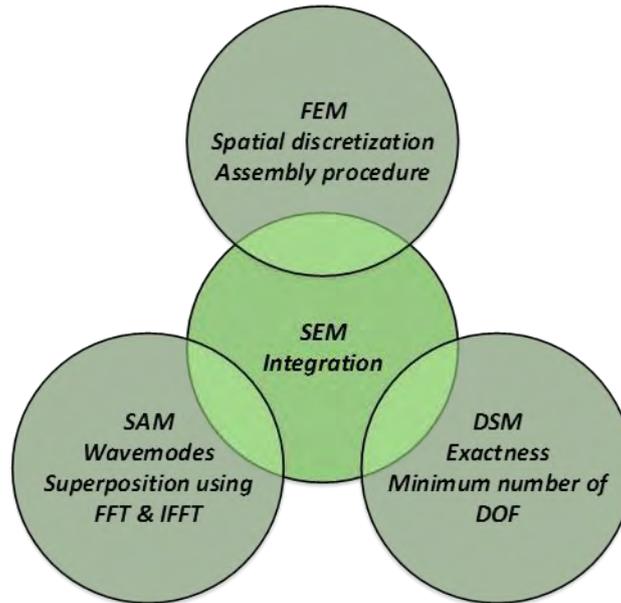
Figure 3. Spectral Element Method (SEM) Outline

Governing equations in the frequency domain

Assume the dynamic response of the Tensegrity in generalized coordinate $w(x,t)$, whereas in the spectral form given by,

$$w(x,t) = \frac{1}{M} \sum_{m=0}^{M-1} W_m(x) e^{i\omega_m t} \tag{4}$$

where $W_m(x) e^{i\omega_m t}$ is the generalized spectral components of the $u(x,t)$ or $v(x,t)$, M is the sampling number, and ω_m is the mth natural frequency.

In spectral forms, the generalized boundary forces $F_1(t)$ and $F_2(t)$, the generalized boundary displacements $w_1(t)$ and $w_2(t)$ are also assumed to be as follows,

$$\{F_1(t), F_2(t)\} = \frac{1}{M} \sum_{m=0}^{M-1} \{F_{1m}, F_{2m}\} e^{i\omega_m t} \tag{5}$$

$$\{w_1(t), w_2(t)\} = \frac{1}{M}\sum_{m=0}^{M-1}\{W_{1m}, W_{2m}\}e^{i\omega_m t} \tag{6}$$

Spectral nodal Degree of Freedom (DOF) and Forces

The spectral nodal generalized displacements $W(U,V)$ and forces $F(N,Q)$ are depicted in Fig. 4. Detail derivation of formulas can be found in [2]. Only relevant formulas are presented in the following.
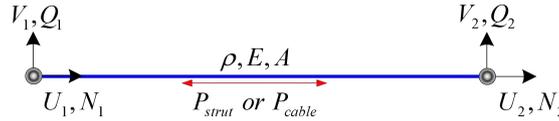
Figure 4. Spectral nodal of generalized DOFs and Forces

The spectral nodal DOFs vector is given by

$$\mathbf{d} = \{W_1 \quad W_2\}^T = \{W(0) \quad W(L)\}^T \tag{7}$$

and the spectral nodal forces vector is given by

$$\mathbf{f} = \{F_1 \quad F_2\}^T = \{-F(0) \quad +F(L)\}^T \tag{8}$$

Dynamic shape function

The solution of (4) can be given by,

$$W(x) = \sum_{r=1}^{2} a_r e^{-ik_r x} = \mathbf{e}(x;\omega)\boldsymbol{a} \tag{9}$$

where, the two wavenumbers can be determined by,

$$k_1 = +\omega\sqrt{\frac{\rho A}{P}} \quad and \quad k_2 = -\omega\sqrt{\frac{\rho A}{P}} \tag{10}$$

with $\mathbf{e}(x;\omega) = \begin{bmatrix} e^{-ik_1 x} & e^{-ik_2 x} \end{bmatrix}$ and $\boldsymbol{a} = \{a_1 \quad a_2\}^T$.

The dynamic shape function can be obtained from,

$$W(x) = \mathbf{N}(x;\omega)\boldsymbol{a} \tag{11}$$

where $\mathbf{N}(x;\omega)$ is given by,

$$\mathbf{N}(x;\omega) = \mathbf{e}(x;\omega)\mathbf{H}_S^{-1}(\omega) = \frac{1}{e^{-ik_2 L} - e^{-ik_1 L}}\left[\left(e^{-i(k_2 L + k_1 x)} - e^{-i(k_1 L + k_2 x)}\right) \quad \left(-e^{-ik_1 x} + e^{-ik_2 x}\right)\right] \tag{12}$$

The spectral nodal DOFs vector of (7) can be formulated by substitution of (11) into (12),

which gives,

$$\mathbf{d} = \mathbf{H}_S(\omega)\boldsymbol{a} \tag{13}$$

where,

$$\mathbf{H}_S(\omega) = \begin{bmatrix} 1 & 1 \\ e^{-ik_1 L} & e^{-ik_2 L} \end{bmatrix}$$

Weak form of governing equation

The spectral element equation of Tensegrity structures can be given as,

$$\mathbf{S}(\omega)\mathbf{d} = \mathbf{0} \tag{14}$$

where,

$$\mathbf{S}(\omega) = \left(\mathbf{H}_S^{-1}(\omega)\right)^T \mathbf{D}(\omega)\mathbf{H}_S^{-1}(\omega) \tag{15}$$

with,

$$\mathbf{D}(\omega) = -P\mathbf{K}^T \mathbf{E}_S \mathbf{K} - \rho A \omega^2 \mathbf{E}_S \tag{16}$$

The derivation of the dynamic shape function (12) can be written as,

$$\mathbf{N}'(x;\omega) = -i\mathbf{e}(x;\omega)\mathbf{K}\mathbf{H}^{-1}(\omega) \tag{17}$$

where,

$$\mathbf{K} = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$$

$$\mathbf{E}_S(\omega) = \int_0^L \left[\mathbf{e}^T(x;\omega)\mathbf{e}(x;\omega)\right] dx = L$$

The Wittrick-Williams procedure is adopted to solve the natural vibrations of the cellular cytoskeleton based on the dynamic Tensegrity formulations above.

Vibrational characteristic of Tensegrity

Analysis was performed using the Spectral Element Method (SEM), which is a combination of the Finite Element Method (FEM), the Dynamic Stiffness Method (DSM), and the Spectral Analysis Method (SAM).

Natural Vibrations of components of Tensegrity

The vibration of a tensegrity structure is highly determined by the combination of vibrations of its components (cables and struts) behaviors. Physical data of each cytoskeleton component is obtained from the following literature.

- Length: Actin [3-6] = 3~20 μm (60 nm, unit length); Microtubule [4-6] = 10 μm

- Diameter: Actin [3-6] = 7~10 nm; Microtubule [3-6] = 25 nm

- Density: Actin [7] = 1.38~1.40 gram/cm^3; Microtubule [8] = 1.41 gram/cm^3

- Elastic modulus: Microtubule [8] = ~ 1 GPa

The tensions per unit length are obtained from the self-equilibrium equation in [1]. Table 1 shows the physical properties of 9-plex cytoskeleton components based on the literature.

Table 1. Physical properties of 9-plex cytoskeleton components

Component	Microtubule (diagonal strut)	Actin-filament (top and bottom cables)	Actin-filament (diagonal cable)
Length (nm)	1912.8	500.0	1711.1
Diameter (nm)	25	7	7
Density (µg/nm^3)	1.41e-15	1.38e-15	1.38e-15
Elastic Modulus (Pa)	3.0e-5	-	-
Tension per unit length (aN/nm)	-2.736e-6	4.0e-8	2.736e-6

Figure 5 shows the model of a 9-plex cytoskeleton tensegrity structure and the results of natural frequencies that satisfy (14), as can be seen as drops in the logarithmic values of the determinant of dynamic stiffness. The first drop is called the structure's first mode of natural frequency.
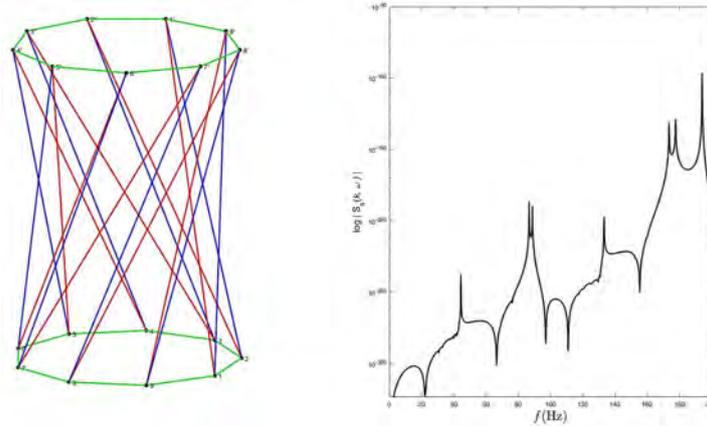
Figure 5. Vibration of a 9-plex cytoskeleton tensegrity structure (1st mode: 21.963 Hz)

Figure 6 shows natural frequency graphs of each component of the 9-plex cytoskeleton tensegrity structure.

Top/bottom actin filament (9.704 Hz)

Microtubule (22.196 Hz)

Diagonal actin filament (43.381 Hz)
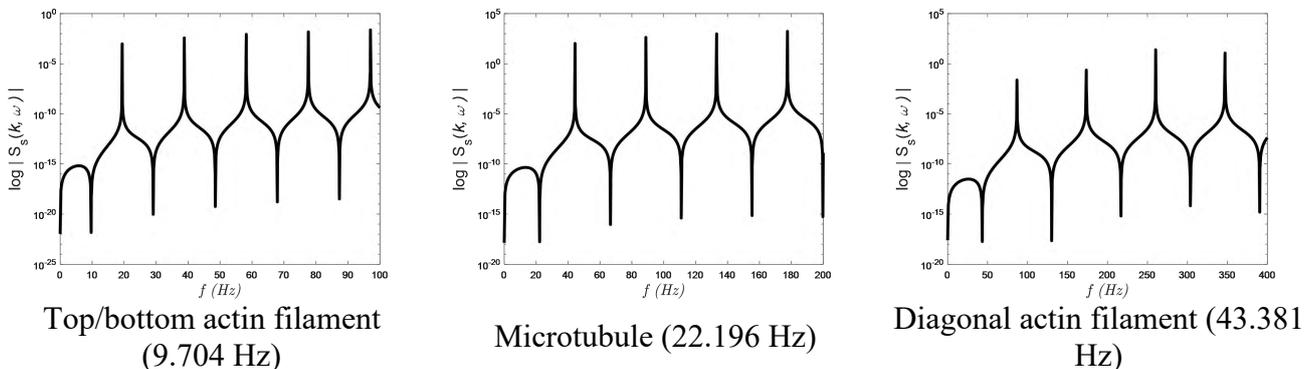
Figure 6. The first five modes vibration of struts and cables in a 9-plex tensegrity cell

Height variation

The variation of heights of the cytoskeleton structure is investigated in this section. Table 2 shows the height variation configurations of the tensegrity structures with their physical properties, which are required to compute the natural frequencies of the structures.

Table 2. Physical properties of 9-plex tensegrities

Height (nm)		1193.2 (0.2H_0)	894.94 (0.6H_0)	1193.2 (0.8H_0)	1491.6 (1.0H_0)
Length (nm)	Microtubule	1234.1	1495.0	1690.5	1912.8
	Actin top&bottom		500.0		
	Actin diagonal	890.0	1226.4	1458.4	1711.1
Diameter (nm)	Microtubule		25		
	Actin		7		
Density[3] (μg/nm^3)	Microtubule		1.41e-15		
	Actin		1.38e-15		
Elastic Modulus[6]	Microtubule		3.0e-5		
Tensile Coefficient (aN/nm)	Actin top (as reference)		4.0e-6		

Figures 7-9 show the model of a 9-plex cytoskeleton tensegrity structure with height variations of 0.8H, 0.6H, and 0.2H, respectively. The first mode of natural frequencies of the various height cytoskeleton tensegrity structures that satisfy (14) can be seen at the first drops in the logarithmic values of the determinant of dynamic stiffness.
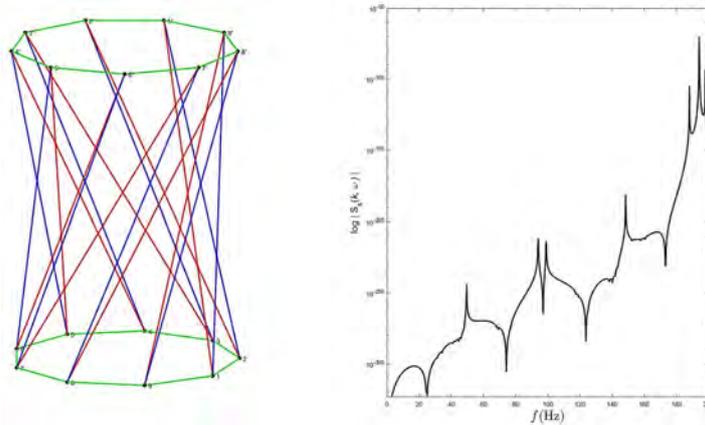
Figure 7. Vibration of a 9-plex tensegrity cell (0.8H = 25.000 Hz)
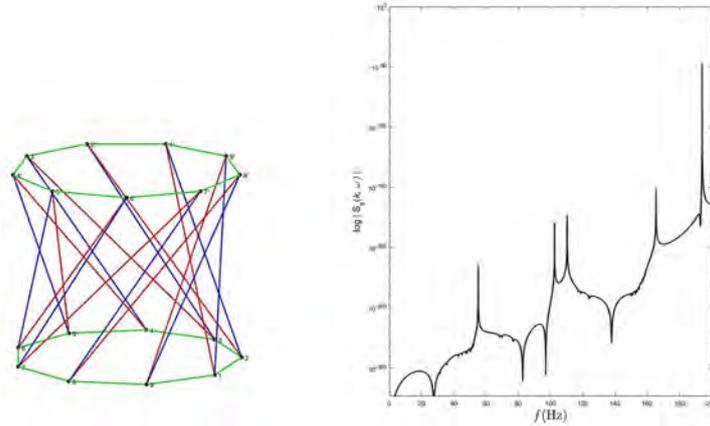
Figure 8. Vibration of a 9-plex tensegrity cell (0.6H = 27.399 Hz)
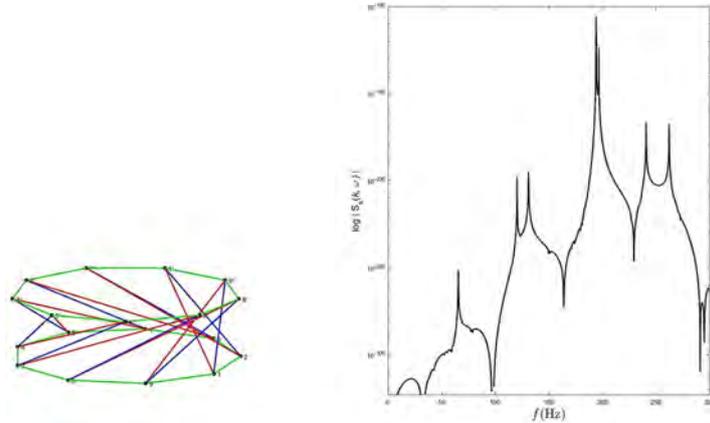
Figure 9. Vibration of a 9-plex tensegrity cell (0.2H = 30.240 Hz)

n-plex variation

This section investigates the variation of 0.2H height of n-plex cytoskeleton twisted cylindrical tensegrity structures. Table 3 shows the physical properties of 9, 12, and 15-plex cytoskeleton tensegrity structures, respectively.

Table 3. Physical properties of *n*-plex tensegrities

n-Plex		9	12	15
Height (nm)			20.24 (*1.4% H$_0$*)	
Length (nm)	Microtubule	1847.3	1950.2	2482.7
	Actin top&bottom		500&1000	
	Actin diagonal	1396.7	2394.8	2927.2
Diameter (nm)	Microtubule		25	
	Actin		7	
Density[3] (μg/nm^3)	Microtubule		1.41e-15	
	Actin		1.38e-15	
Elastic Modulus[6] (Pa)	Microtubule		3.0e+9	
Tensile Coefficient (aN/nm)	Actin top (as referen		4.0e-6	

Figures 10-12 show the model of an *n*-plex cytoskeleton tensegrity structure with 0.2H of height. The first mode of natural frequencies of the *n*-plex cytoskeleton tensegrity structures that satisfy (14) can be seen at the first drops in the logarithmic values of the determinant of dynamic stiffness.
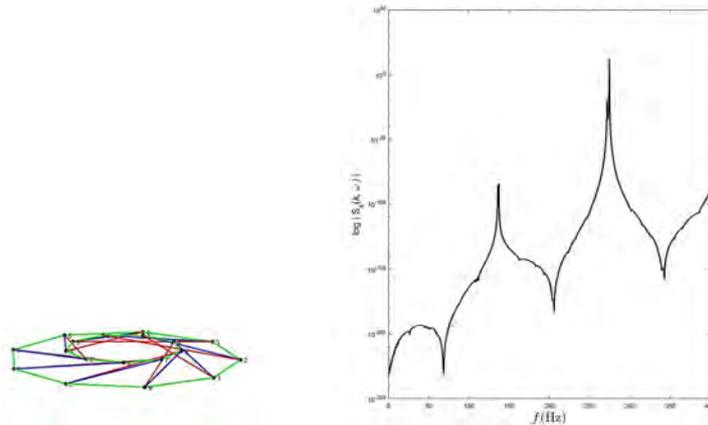
Figure 10. Vibration of a truncated 9-plex conic tensegrity cell (f_1 = 68.61 Hz)
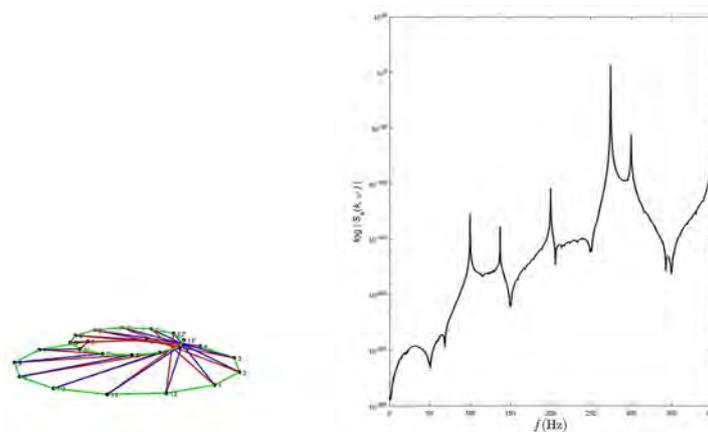
Figure 11. Vibration of a truncated 12-plex conic tensegrity cell (f_1 = 50.61 Hz)
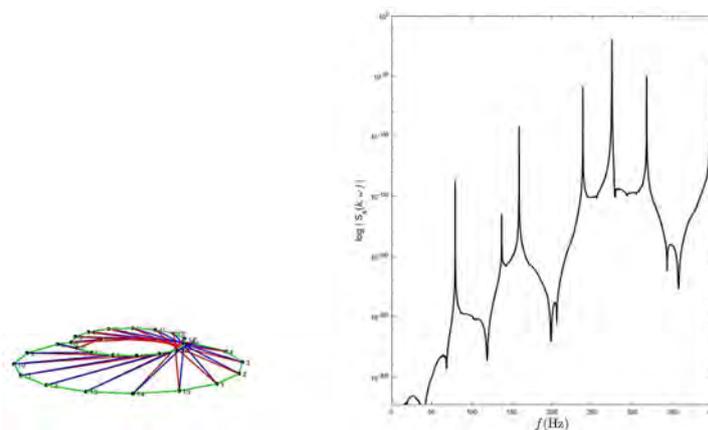
Figure 12. Vibration of a truncated 15-plex conic tensegrity cell (f_1 = 35.49 Hz)

Summary and conclusion

In Figs. 5-6, we can observe that the cellular 9-plex tensegrity has the 1st mode natural frequency, which is close to the 1st mode natural frequency of its strut (Microtubule) because the actin filaments are assumed to vibrate in their lateral direction. In contrast, the microtubule is assumed to vibrate in its axial direction, controlling the vibrational behaviors of the tensegrity.

The height of 9-plex tensegrity is varied by 0.8, 0.6, and 0.2 times the height H. Figs. 7-9 show that the 1st mode of their natural frequency was increasing by reducing the height.

Fixing the height of the 9-plex tensegrity with the 0.2H height, the 1st modes of the 12- and 15-plex tensegrity show the reduction of natural frequency since they become stiffer than the 9-plex tensegrity.

By using tensegrity, we obtained physical properties to reproduce the fluctuations of cells. In the future, we will investigate the significance of fluctuations in the cellular structure composed of actin filaments and microtubules subjected to traction force by observing them under different temperatures and mechanical conditions and by approximating the shape of actual cells. Also, this time, we reproduced the cell shape using a cylindrical model, but since the actual cell has a complicated shape, it is necessary to use random numbers and precise modeling of the cytoskeleton structure for verification. Further elucidation of the molecular mechanism of cytoskeletal control by mechanical stimulation received by cells is expected by developing techniques that can measure the force and shape of the cytoskeleton.

All the cells that make up our body receive mechanical stimuli and maintain homeostasis through appropriate haptic responses. Its failure is expected to be the cause of many diseases. By advancing the elucidation of the process by which cells receive mechanical stimuli and transform them into chemical signals, many discoveries have been made about problems that could not be explained only by chemical signals. It is expected that it will be useful in investigating mechanical cell behaviors.

Conflicts of Interest

The authors declare no conflict of interest.

Declaration of generative AI

The authors declare the the present work has no AI generated products.

References

[1] G. Albrecht-Buehler, Role of cortical tension in fibroblast shape and movement., Cell Motil. Cytoskeleton. 7 (1987) 54–67. https://doi.org/10.1002/cm.970070108.

[2] J.R. Banerjee, Dynamic stiffness formulation for structural elements: A general approach, Comput. Struct. 63 (1997) 101–103. https://doi.org/10.1016/S0045-7949(96)00326-4.

[3] T. Barata, V. Vieira, R. Rodrigues, R.P. das Neves, M. Rocha, Reconstruction of tissue-specific genome-scale metabolic models for human cancer stem cells, Comput. Biol. Med. 142 (2022). https://doi.org/10.1016/j.compbiomed.2021.105177.

[4] D.E. Beskos, G. V. Narayanan, Dynamic response of frameworks by numerical laplace transform, Comput. Methods Appl. Mech. Eng. 37 (1983) 289–307. https://doi.org/10.1016/0045-7825(83)90080-4.

[5] M.E. Chicurel, C.S. Chen, D.E. Ingber, Cellular control lies in the balance of forces, Curr. Opin. Cell Biol. 10 (1998) 232–239. https://doi.org/10.1016/S0955-0674(98)80145-2.

[6] D.W. Deamer, The origin of cellular life, Planets Life Emerg. Sci. Astrobiol. 22 (2007) 187–210. https://doi.org/10.1017/cbo9780511812958.011.

[7] B. Demé, D. Hess, M. Tristl, L.T. Lee, E. Sackmann, Binding of actin filaments to charged lipid monolayers: Film balance experiments combined with neutron reflectivity, Eur. Phys. J. E. 2 (2000) 125–136. https://doi.org/10.1007/s101890050046.

[8] C. Dong, R. Skalak, K.L.P. Sung, Cytoplasmic rheology of passive neutrophils, Biorheology. 28 (1991) 557–567. https://doi.org/10.3233/BIR-1991-28607.

[9] J.F. Doyle, Wave Propagation in Structures, Wave Propag. Struct. (1989) 126–156. https://doi.org/10.1007/978-1-4684-0344-2_6.

[10]   J. Feher, Cell Structure, Quant. Hum. Physiol. (2017) 101–119. https://doi.org/10.1016/b978-0-12-800883-6.00010-0.

[11]   E.G. Fey, K.M. Wan, S. Penman, Epithelial cytoskeletal framework and nuclear matrix-intermediate filament scaffold: three-dimensional organization and protein composition, J. Cell Biol. 98 (1984) 1973–1984. https://doi.org/10.1083/jcb.98.6.1973.

[12]   Fuller, Tensegrity, Portf. Artnews Annu. (1961) 112–127.

[13]   R.B. Fuller, Conceptuality of Fundamental Structures, 1965. https://books.google.co.jp/books?id=HEL6NwAACAAJ.

[14]   Y.C. Fung, S.Q. Liu, Elementary mechanics of the endothelium of blood vessels, J. Biomech. Eng. 115 (1993) 1–12. https://doi.org/10.1115/1.2895465.

[15]   H. Furuya, Concept of Deployable Tensegrity Structures in Space Application, Int. J. Sp. Struct. 7 (1992) 143–151. https://doi.org/10.1177/026635119200700207.

[16]   B.S. Gan, Self-vibrational Analysis of a Tensegrity, in: N. Tien Khiem, T. Van Lien, N. Xuan Hung (Eds.), Lect. Notes Mech. Eng., Springer Singapore, Singapore, 2022: pp. 87–99. https://doi.org/10.1007/978-981-16-3239-6_7.

[17]   B.S. Gan, Computational Modeling of Tensegrity Structures: Art, Nature, Mechanical and Biological Systems, Springer Cham, 2019. https://doi.org/10.1007/978-3-030-17836-9.

[18]   B.S. Gan, J. Zhang, D.K. Nguyen, E. Nouchi, Node-based genetic form-finding of irregular tensegrity structures, Comput. Struct. 159 (2015) 61–73. https://doi.org/10.1016/j.compstruc.2015.07.003.

[19]   M.L. Gardel, K.E. Kasza, C.P. Brangwynne, J. Liu, D.A. Weitz, Chapter 19 Mechanical Response of Cytoskeletal Networks, Methods Cell Biol. 89 (2008) 487–519. https://doi.org/10.1016/S0091-679X(08)00619-5.

[20]   E. Ghavanloo, F. Daneshmand, M. Amabili, Vibration analysis of a single microtubule surrounded by cytoplasm, Phys. E Low-Dimensional Syst. Nanostructures. 43 (2010) 192–198. https://doi.org/10.1016/j.physe.2010.07.016.

[21]   F. Gittes, B. Mickey, J. Nettleton, J. Howard, Flexural rigidity of microtubules and actin filaments measured from thermal fluctuations in shape, J. Cell Biol. 120 (1993) 923–934. https://doi.org/10.1083/jcb.120.4.923.

[22]   D.C. Haspinger, S. Klinge, G.A. Holzapfel, Numerical analysis of the impact of cytoskeletal actin filament density alterations onto the diffusive vesicle-mediated cell transport, PLoS Comput. Biol. 17 (2021) 1–26. https://doi.org/10.1371/journal.pcbi.1008784.

[23]   J.E. Heuser, M.W. Kirschner, Filament organization revealed in platinum replicas of freeze-dried cytoskeletons, J. Cell Biol. 86 (1980) 212–234. https://doi.org/10.1083/jcb.86.1.212.

[24]   M. Igaev, H. Grubmüller, Microtubule instability driven by longitudinal and lateral strain propagation, PLoS Comput. Biol. 16 (2020) 1–21. https://doi.org/10.1371/journal.pcbi.1008132.

[25]   D.E. Ingber, Cellular tensegrity: Defining new rules of biological design that govern the cytoskeleton, J. Cell Sci. 104 (1993) 613–627. https://doi.org/10.1242/jcs.104.3.613.

[26]   D.E. Ingber, J.D. Jamieson, Tumor formation and malignant invasion: role of basal lamina, in: L.A. Liotta, I.R. Hart (Eds.), Tumor Invasion and Metastasis, Springer Netherlands, Dordrecht, 1982: pp. 335–357. https://doi.org/10.1007/978-94-009-7511-8_20.

[27]   D.E. Ingber, J.A. Madri, J.D. Jamieson, Role of basal lamina in neoplastic disorganization of tissue architecture, Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 3901–3905. https://doi.org/10.1073/pnas.78.6.3901.

[28]   D.E. Ingber, Tensegrity I. Cell structure and hierarchical systems biology, J. Cell Sci. 116 (2003) 1157–1173. https://doi.org/10.1242/jcs.00359.

[29]   D.E. Ingber, The riddle of morphogenesis: A question of solution chemistry or molecular cell engineering?, Cell. 75 (1993) 1249–1252. https://doi.org/10.1016/0092-8674(93)90612-T.

[30]   D.E. Ingber, Tensegrity II. How structural networks influence cellular information processing networks, J. Cell Sci. 116 (2003) 1397–1408. https://doi.org/10.1242/jcs.00360.

[31]   D.E. Ingber, L. Dike, L. Hansen, S. Karp, H. Liley, A. Maniotis, H. McNamee, D. Mooney, G. Plopper, J. Sims, N. Wang, Cellular Tensegrity: Exploring How Mechanical Changes in the Cytoskeleton Regulate Cell Growth, Migration, and Tissue Pattern during Morphogenesis, Int. Rev. Cytol. 150 (1994) 173–224. https://doi.org/10.1016/S0074-7696(08)61542-9.

[32]   D.E. Ingber, D. Prusty, Z. Sun, H. Betensky, N. Wang, Cell shape, cytoskeletal mechanics, and cell cycle control in angiogenesis, J. Biomech. 28 (1995) 1471–1484. https://doi.org/10.1016/0021-9290(95)00095-X.

[33]   V.V.S.V. Jakka, J. Bursa, Impact of physiological loads of arterial wall on nucleus deformation in endothelial cells: A computational study, Comput. Biol. Med. 143 (2022) 105266. https://doi.org/10.1016/j.compbiomed.2022.105266.

[34] P.A. Janmey, U. Euteneuer, P. Traub, M. Schliwa, Viscoelastic properties of vimentin compared with other filamentous biopolymer networks, J. Cell Biol. 113 (1991) 155–160. https://doi.org/10.1083/jcb.113.1.155.

[35] P.A. Janmey, The cytoskeleton and cell signaling: Component localization and mechanical coupling, Physiol. Rev. 78 (1998) 763–781. https://doi.org/10.1152/physrev.1998.78.3.763.

[36] U. Lee, Spectral Element Method in Structural Dynamics, John Wiley & Sons (Asia) Pte Ltd, 2009. https://doi.org/10.1002/9780470823767.

[37] A.Y.T. Leung, Dynamic Stiffness and Substructures, Springer-Verlag, 1993. https://doi.org/10.1007/978-1-4471-2026-1.

[38] P. Motro, R., Najari, S., Jouanna, Tensegrity systems. From design to realization., in: Proc. First Int. Conf. Light. Struct. Archit., 1986.

[39] B. Moussa, N. Ben Kahla, J.C. Pons, Evolution of natural frequencies in tensegrity systems: A case study, Int. J. Sp. Struct. 16 (2001) 57–73. https://doi.org/10.1260/0266351011495322.

[40] H. Murakami, Static and dynamic analyses of tensegrity structures. Part 1. Nonlinear equations of motion, Int. J. Solids Struct. 38 (2001) 3599–3613. https://doi.org/10.1016/S0020-7683(00)00232-8.

[41] I.J. Oppenheim, W.O. Williams, Vibration of an elastic tensegrity structure, Eur. J. Mech. A/Solids. 20 (2001) 1023–1031. https://doi.org/10.1016/S0997-7538(01)01181-0.

[42] A.F. Pegoraro, P. Janmey, D.A. Weitz, Mechanical properties of the cytoskeleton and cells, Cold Spring Harb. Perspect. Biol. 9 (2017). https://doi.org/10.1101/cshperspect.a022038.

[43] J. Pokorný, F. Jelínek, V. Trkal, I. Lamprecht, R. Hölzel, Vibrations in Microtubules, J. Biol. Phys. 23 (1997) 171–179. https://doi.org/10.1023/A:1005092601078.

[44] G.W. Schmid-Schönbein, T. Kosawada, R. Skalak, S. Chien, Membrane model of endothelial cells and leukocytes. A proposal for the origin of a cortical stress, J. Biomech. Eng. 117 (1995) 171–178. https://doi.org/10.1115/1.2795999.

[45] G.C. Sieck, Controversies in physiology, J. Appl. Physiol. 89 (2000) 407. https://doi.org/10.1152/jappl.2000.89.2.407.

[46] K. Snelson, Snelson on the tensegrity invention, Int. J. Sp. Struct. 11 (1996) 43–48. https://doi.org/10.1177/026635119601-207.

[47] D. Stamenović, D.E. Ingber, Models of cytoskeletal mechanics of adherent cells, Biomech. Model. Mechanobiol. 1 (2002) 95–108. https://doi.org/10.1007/s10237-002-0009-9.

[48] C. Sultan, M. Corless, R.E. Skelton, Linear dynamics of tensegrity structures, Eng. Struct. 24 (2002) 671–685. https://doi.org/10.1016/S0141-0296(01)00130-4.

[49] A. Tounsi, H. Heireche, H. Benhassaini, M. Missouri, Vibration and length-dependent flexural rigidity of protein microtubules using higher order shear deformation theory, J. Theor. Biol. 266 (2010) 250–255. https://doi.org/10.1016/j.jtbi.2010.06.037.

[50] Y. Ujihara, M. Nakamura, M. Soga, K. Koshiyama, H. Miyazaki, S. Wada, Computational studies on strain transmission from a collagen gel construct to a cell and its internal cytoskeletal filaments, Comput. Biol. Med. 56 (2015) 20–29. https://doi.org/10.1016/j.compbiomed.2014.10.015.

[51] P. Urdeitx, S.J. Mousavi, S. Avril, M.H. Doweidar, Computational modeling of multiple myeloma interactions with resident bone marrow cells, Comput. Biol. Med. 153 (2023) 106458. https://doi.org/10.1016/j.compbiomed.2022.106458.

[52] W.H. Wittrick, F.W. Williams, A general algorithm for computing natural frequencies of elastic structures, Q. J. Mech. Appl. Math. 24 (1971) 263–284. https://doi.org/10.1093/qjmam/24.3.263.

Effect of generic submarine appendages on free surface hydrophysical signatures in the stratified fluid

Fenglai Huang, *Liushuai Cao, †Decheng Wan

Computational Marine Hydrodynamic Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China.

*Presenting author: liushuaicao@sjtu.edu.cn
†Corresponding author: dcwan@sjtu.edu.cn

Abstract

Stratification will contribute to the anisotropy of the submarine wave height and a longer and wider influencing zone on the free surface. However, the contribution ratio of appendages to the waveform characteristics on the free surface is unclear. This greatly hindered the research on the optimization design of the submarine shape with the goal of reducing the hydrophysical signatures on the free surface. In this study, taking the Joubert BB2 submarine model as an example, three different types of submarines in the linearly stratified fluid are simulated: a bare hull, a full-appendage Joubert BB2 submarine, and the full-appendage Joubert BB2 submarine with a ducted propeller. The Reynolds number is about 10^6, and the internal Froude number is about 3 in these three cases. After quantality analyzed the density fluctuations, internal waves, and complex hydrophysical signatures on the free surface of these three cases. The results show that the existence of the appendages will cause an obvious phenomenon on the free surface, and will result in the waveform formed by the submarine on the free surface being "farther", "wider", "faster" and longer duration.

Keywords: Stratified fluid; Joubert BB2 submarine; hydrophysical signatures

Introduction

An essential characteristic of oceans is density stratification. Stratification can significantly affect the mixing process and dynamics of the fluid since it can induce internal waves and alteration in the level and spatial distribution of wake turbulence. It is known that the wake dynamics depend strongly on the Reynolds number, which is the ratio of the inertia force to the viscous forces, and has a significant impact on bluff body wakes.

$$Re = \frac{UD}{\upsilon} \tag{1}$$

here, U is the body speed, D is the diameter of the sphere, and υ is the kinematic viscosity of the water.

Unlike homogeneous fluid, stratified wakes also need to take into account the buoyancy effect in addition to the conflicting impacts of inertial and viscous forces. The ratio of the inertia force to the buoyancy force called internal Froude number Fr_D, is used to describe the various stratification levels.

$$Fr_D = \frac{U}{ND} \tag{2}$$

$$N = [-(g/\rho_0)(\partial \rho/\partial z)]^{1/2} \tag{3}$$

where N is the Brunt-Vaisala frequency, ρ_0 is the undisturbed density profile, and g is the gravitational acceleration. These two critical parameters, the internal Froude number and the Reynolds number, typically have an additive effect on stratified wakes.

There are majority of earlier studies generally concentrated on conventional simple geometries, such as the sphere[1]-[4], cylinder[5], disk[6], and prolate spheroid[7], even if there has been tremendous advancement in the stratified fluid. For simplicity, most basic studies on the flow field surrounding marine vessels typically ignore the effect of stratification.

The movement of a ship and submerged bodies in a stratified fluid has been modeled in a number of publications[8]-[14]. These studies included Carrica[10], which provided CFD simulations of self-propulsion close to the surface in waves and calm water; Xue[11], which looked into the connection between the motion state of moving submerged bodies and wakes; and Ma[12]-[13], which attempted to quantitatively analyze the effects of submarine speed, depth, and density gradient on the free surface wave pattern in linearly stratified and a strongly stratified fluid. As Huang[14] stated, stratification will contribute to the anisotropy of the submarine wave height and a longer and wider influencing zone on the free surface. Then, what is the contribution ratio of appendages to the waveform characteristics on the free surface? In other words, we hope to know the contribution ratio of sails, rudders, and propellers to the waveform characteristics on the free surface, so that being more careful when we design appendages.

In this study, take the Joubert BB2 submarine model as an example, three different types of submarines in the linearly stratified fluid are simulated: a bare hull, a full-appendage Joubert BB2 submarine, and the full-appendage Joubert BB2 submarine with a ducted propeller. The Reynolds number is about 10^6, and the internal Froude number is about 3 in these three cases. We will explore the following issues: (1) the visualization of hydrophysical signatures of submarines, (2) quantitatively analyze how stratification influences the hydrophysical signatures of submarines including the propagation distance in the *xOy* plane, the wave height, and the Kelvin wedge angles; (3) reveal appendages affection on the hydrophysical signatures of submarines.

Numerical methods

Governing equations

Based on Huang[14], we solved the Navier-Stokes equations under Boussinesq approximation and proposed a thermocline model to achieve the linearly stratified environment. The governing equations are shown as follows:

$$\nabla \cdot \boldsymbol{u} = 0 \tag{4}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\frac{1}{\rho}\nabla P + \upsilon\nabla^2\boldsymbol{u} + \beta\boldsymbol{g}T' + \boldsymbol{f} \tag{5}$$

$$\frac{\partial T'}{\partial t} + (\boldsymbol{u} \cdot \nabla)T' + \gamma u_z = \alpha\nabla^2 T' \tag{6}$$

where $\boldsymbol{u}$, ρ, P, υ, α, β, $\boldsymbol{g}$, T', $\boldsymbol{f}$ denote the velocity, density, modified pressure, kinematic viscosity, thermal diffusivity, volumetric expansion coefficient, gravitational acceleration, temperature fluctuation, and large-scale force required to maintain the turbulence, respectively.

The thermocline model

As Huang[14] stated, we established a relationship between the density ρ and temperature T so that we can achieve the linear stratification in the STAR-CCM+ software. the density ρ and temperature T can be written as follows:

$$\begin{cases} T = f(z) = T_0 + c_1 \cdot z \\ \rho = \rho(z) = g(z) = \rho_0 - c_2 \cdot z \end{cases} \tag{6}$$

where T_0, ρ_0 are the reference temperature and density, c_1 and c_2 are the temperature and density gradient, respectively.

Turbulence modeling

The whole problem is simulated with Reynolds averaged Navier-Stokes (RANS) equation, and the Shear-Stress Transport (SST) k-ω turbulence model is adopted to close the equation. The governing equations of k and ω can be written as follows:

$$\frac{\partial}{\partial t}(\rho k) + \nabla \cdot (\rho \boldsymbol{v} k) = \nabla \cdot \left(\mu_{eff,k} \nabla k\right) + \underbrace{P_k - \beta^* \rho k \omega}_{S^k} \tag{8}$$

$$\frac{\partial}{\partial t}(\rho \omega) + \nabla \cdot (\rho \boldsymbol{v} \omega)$$
$$= \nabla \cdot \left(\mu_{eff,\omega} \nabla \omega\right) + \underbrace{\widetilde{C_\alpha} \frac{\omega}{k} P_k - \widetilde{C_\beta} \rho \omega^2 + 2(1 - F_1) \sigma_{\omega 2} \frac{\rho}{\omega} \nabla k \cdot \nabla \omega}_{S^\omega} \tag{9}$$

where all their coefficients depend on the Menter and Kuntz[15].

The volume of fluid method

The volume of fluid (VOF) multiphase model is adopted in this paper Hirt[16]. Depending on the different values of the volume fraction q we can distinguish the position of the interface:

- $q = 0$ the cell is completely void of phase air;
- $q = 1$ the cell is completely filled with phase water;   (10)
- $0 < q < 1$ indicate the presence of an interface between air and water.

Computational setup and validation

The geometric model

The mother geometric model under study is the Joubert BB2 submarine model designed by Joubert[17], which is now an international benchmark for underwater vehicles. In this paper, we simulated three different types of Joubert BB2: a bare hull, a full-appendage Joubert BB2 submarine, and the full-appendage Joubert BB2 submarine with a ducted propeller to analyze the complex hydrophysical signature on the free surface when they operating near the free surface, respectively. The schematic diagram for each of the three types of Joubert BB2 is shown in Figure 1. The Joubert BB2 geometric key parameters are displayed in Table 1 for both full and model scales, and the scaling ratio λ is equal to 18.348 in accordance with accepted international conventions. The propeller is INSEAN E779A, which is a four-blade righthand propeller, the duct is designed to match the propeller, and their main parameters are shown in Table 2, the schematic diagram is illustrated in Figure 1 (c).

Table 1. Main parameters of Joubert BB2 submarine (model scale 1:18.348)

Description	Symbol	Scale	
		Full	Model
Length	L (m)	70.2	3.8260
Beam	B (m)	9.6	0.5232
Draft to Duck	D (m)	10.6	0.5777
Draft to Sail top	D_{sail} (m)	16.2	0.8829
Wetted surface area	S_{wa} (m^2)	2142.3	6.3635

Table 2. The parameters of the ducted propeller

Propeller		INSEAN E779A
Number of blades	Z	4
Propeller diameter	D_p (mm)	227.27
Hub diameter	D_H (mm)	45.53
Pitch ratio (nominal)	P/D_p	1.1
Duct length	L_{duct}	$0.35D_p$
Tip clearance	t (mm)	5

(a) Bare hull model

(b) Joubert BB2 submarine model

(c) Joubert BB2 submarine model with a ducked propeller
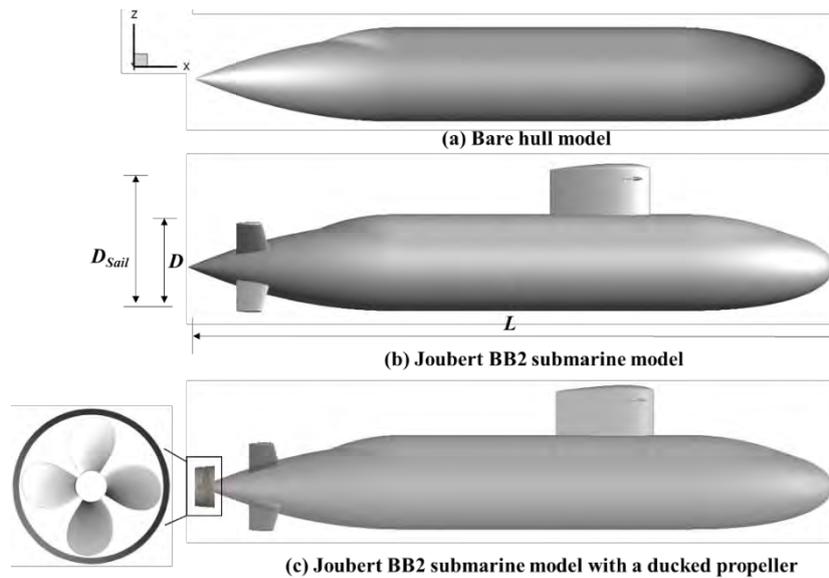
Figure 1. Schematic diagram of the three different types of submarines

Computational domain and discretized mesh

The computational domain size used in this study is substantially bigger than the needed domain size stated in prior studies for flow past a submarine. With the submarine located at the front of the domain, the entire domain extends to 4L in front of the submarine, 20L behind the submarine, and 7L in each broadside to reveal the main characteristics of the waveform on the free surface in each case. To prevent wall effects, the vertical dimension is set to 3L. A sponge layer with a length of 1L is inserted in the inlet and broadside to prevent wave reflection. This larger

computational domain provides convenience for studying the evolution of wake characteristics on the free surface.

In addition, four different boundary condition types are used in this paper. The Joubert BB2 submarine has a non-slip wall boundary condition, velocity inflow and outflow boundary conditions are used for the inlet and outlet, respectively. Symmetry boundary conditions are used on the remaining portion of the domain.

Figure 2 shows the whole calculation grids for the three cases. In cases of bare hull and the full-appendage Joubert BB2 submarine the original mesh is a half mesh with symmetry plane $y = 0$, then they are mirrored. In the case of the full-appendage Joubert BB2 submarine, to ensure the proper resolution of the shear layer and lessen force and moment mistakes, more cells are added to the port, starboard, rudder, and stern regions as displayed in Figure 2 (b). The total number of cells is about 14 million and 16 million for the bare hull case and the full-appendage Joubert BB2 submarine case, respectively. In the case of the full-appendage Joubert BB2 submarine with a ducted propeller, since the propeller rotation is not symmetrical, the mesh has to be whole, and the final total grid is about 33 million. Both in the three cases, the boundary layers are set to 12, the minimum cell thickness from the hull is 1×10^{-4}m, and the $y+$ is continuously maintained between $30 \sim 100$.

(a) Bare hull model

(b) Joubert BB2 submarine model

(c) Joubert BB2 submarine model with a ducked propeller
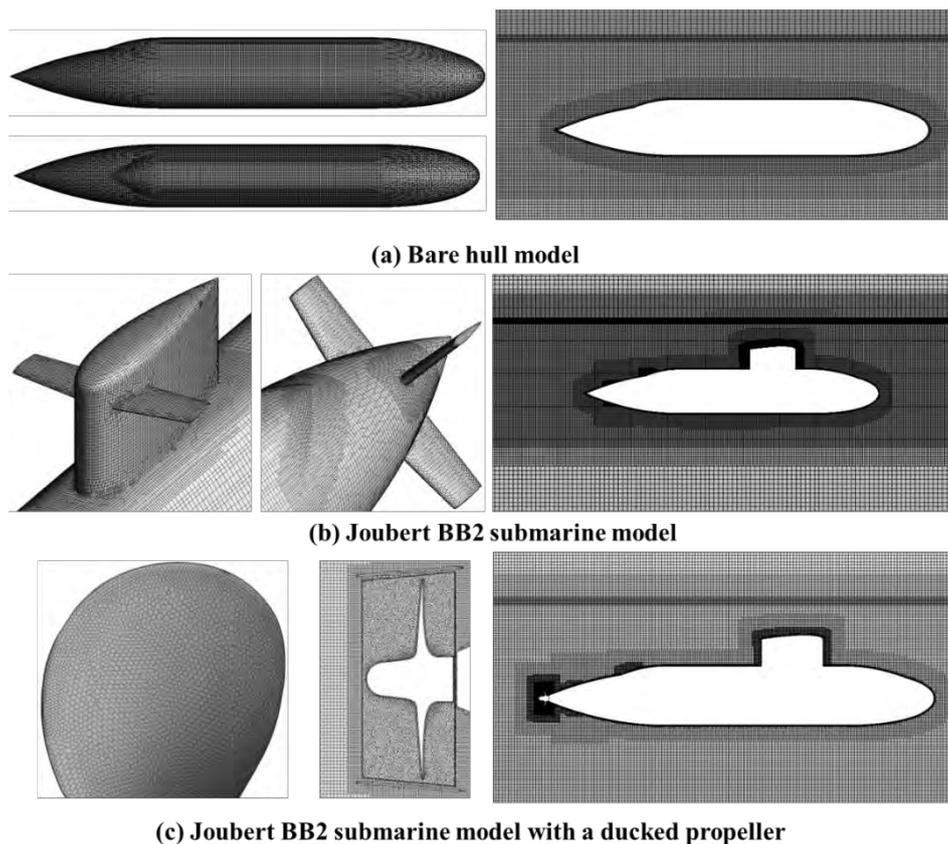
Figure 2. The grid of the three different types of submarines

Initial conditions for the stratification simulation

In this paper, all cases are performed at model scale submarines with the submerged depth equals $z = 0.25L$ and an onset velocity of $U = 1.201$m/s. Regarding the stratification, we model a weakly stratified circumstance where internal Froude number $Fr_D \approx 3$. The polynomial

expression shown in Eqn. (6) can be used to specify the linear density stratification as a function of temperature. The distance from the submarine to the free surface, or the submerged depth z, determines the temperature. As shown in Figure 3, $T_0 = 287.1$ K and varies from 288.15 K to 286.05 K at the top and bottom, and $\rho_0 = 1010$ kg/m^3 and varies from 1000 kg/m^3 to 1020 kg/m^3. Figure 3 helps visualize the initial conditions set for the stratified case.
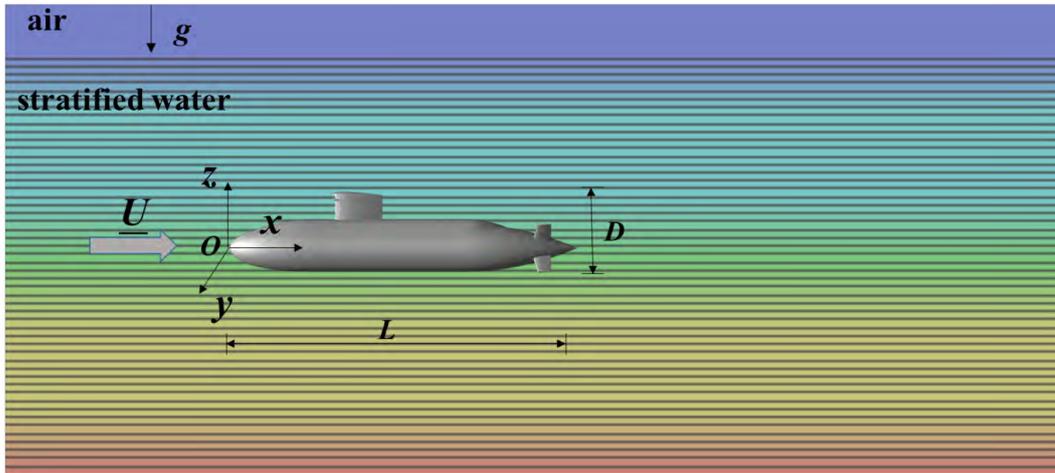
Figure 3. The initial condition for the stratification simulation

Results and discussion

In this section, by quantitatively comparing the internal waves and free surface wake characteristics on the free surface of a bare hull, a full-appendage Joubert BB2 submarine, and the full-appendage Joubert BB2 submarine with a ducted propeller to reveal appendages affection on the hydrophysical signatures on the free surface. The following section has two parts, the first part is density fluctuations and internal waves which help to figure out how appendages influence the wakes in the stratified environment, and the second part is to analyze appendages affection on the hydrophysical signatures on the free surface. In this session, used case 1 represents the bare hull case, case 2 is the full-appendage Joubert BB2 submarine case, and case 3 stands for the full-appendage Joubert BB2 submarine with a ducted propeller case for simplify.

Density fluctuations and Internal waves

Due to the existence of the submarine, internal waves and turbulent wakes are both generated in the stratified fluid, resulting in a narrowing area near the tail of the submarine in the density profile shown in Figures 4 (a), 5 (a), and 6 (a). When analyzing the density profile near the submarine, seven *y-z* slices are created. These slices are placed at $x/L = 0$, $x/L = 0.5$, $x/L = 1.0$, $x/L = 1.5$, $x/L = 2.0$, $x/L = 2.5$, $x/L = 3.0$, where $x/L = 0$ corresponds to the tail of the submarine both in the three cases, and L is the value of the length of the Joubert BB2 submarine. The internal waves affect the original stable density field, which illustrates in the Figures 4, 5, and 6, contours for density stratification become distorted. With the increase of appendages, the density profiles changed more violently.

For example, when referring to the slice $x/L = 0$, in the bare hull case (case 1), the variation of the vertical direction z of the density layer is about [-0.3B, 0.6B], and the variation range of the lateral direction y is about [-0.2B, 0.2B], as shown in Figure 4 (a), and B is the beam of the Joubert BB2 submarine. In the full-appendage Joubert BB2 submarine case (case 2), as shown

in Figure 5 (a), the change of the vertical direction *z* of the density layer is about [-0.3*B*, 1.15*B*], the variation range of the lateral direction is about [-0.3*B*, 0.3*B*], which increases by about 91.67% in the *z* direction, and 50% in the *y* direction compared with case 1. Figure 6 (a) illustrates the full-appendage Joubert BB2 submarine with a ducted propeller case (case 3), at this time, the variation of the vertical and lateral directions are roughly the same as case 2. The difference between Figure 5 (a) and Figure 6 (a) is that due to the rotation of the ducted propeller, the contours of the density layers are greatly destroyed and intersect in Figure 6 (a), on the contrary, although the density layers have a large change, they do not intersect in Figure 5 (a).

(a) *x /L* = 0   (b) *x /L* = 0.5

(c) *x /L* = 1.0   (d) *x /L* = 1.5

(e) *x /L* = 2.0   (f) *x /L* = 2.5

(g) *x /L* = 3.0

Density of water: 1000 1004 1008 1012 1016 1020
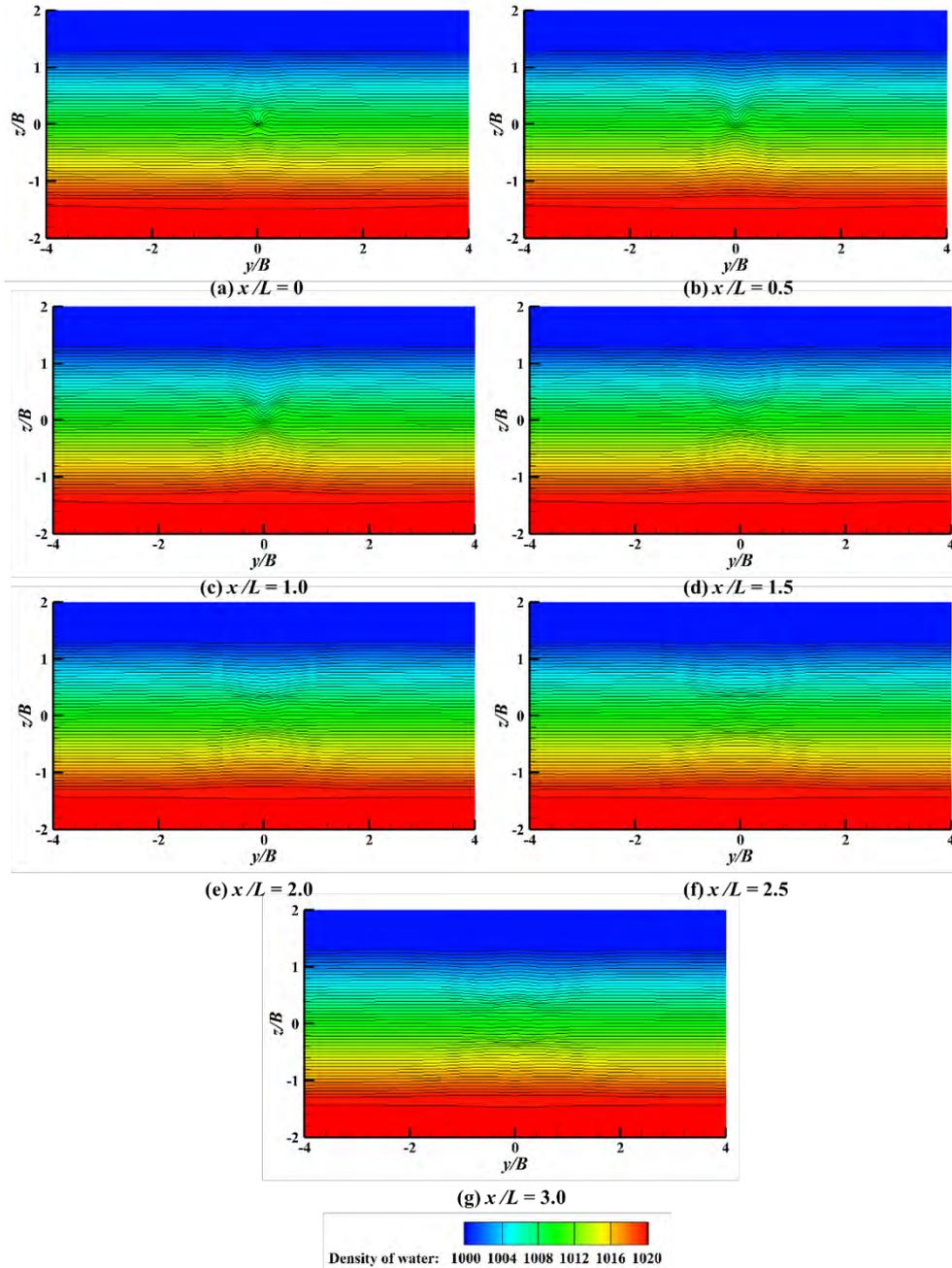
Figure 4. The distribution of density in seven slices along the *x* direction in the near field of the bare hull model

When the slice is *x/L* = 1.0, in case 2, as shown in Figure 5 (c), the change degree of the density layers in the vertical direction is lessened compared with the *x/L* = 0 slice, but the lateral variation range begins to expand, increasing to [-0.6*B*, 0.6*B*]. The vertical (*z*) direction change

of the density profile caused by the propeller's agitation is still severe and the contour of the density layers still has a huge distortion as illustrated in Figure 6(c), when referring to case 3. At this time, the variation range of the lateral direction continues to expand to about [-0.7B, 0.7B].

When the slice is x/L= 3.0, it can be seen that the change degree of the density profile of the three cases is basically similar, as shown in Figures 4 (g), 5 (g), and 6 (g). At this time, the influence of the existence of the submarine on the density profile begins to decrease, the vertical change of the density stratification begins to decrease, and the influence range gradually develops to the lateral direction.
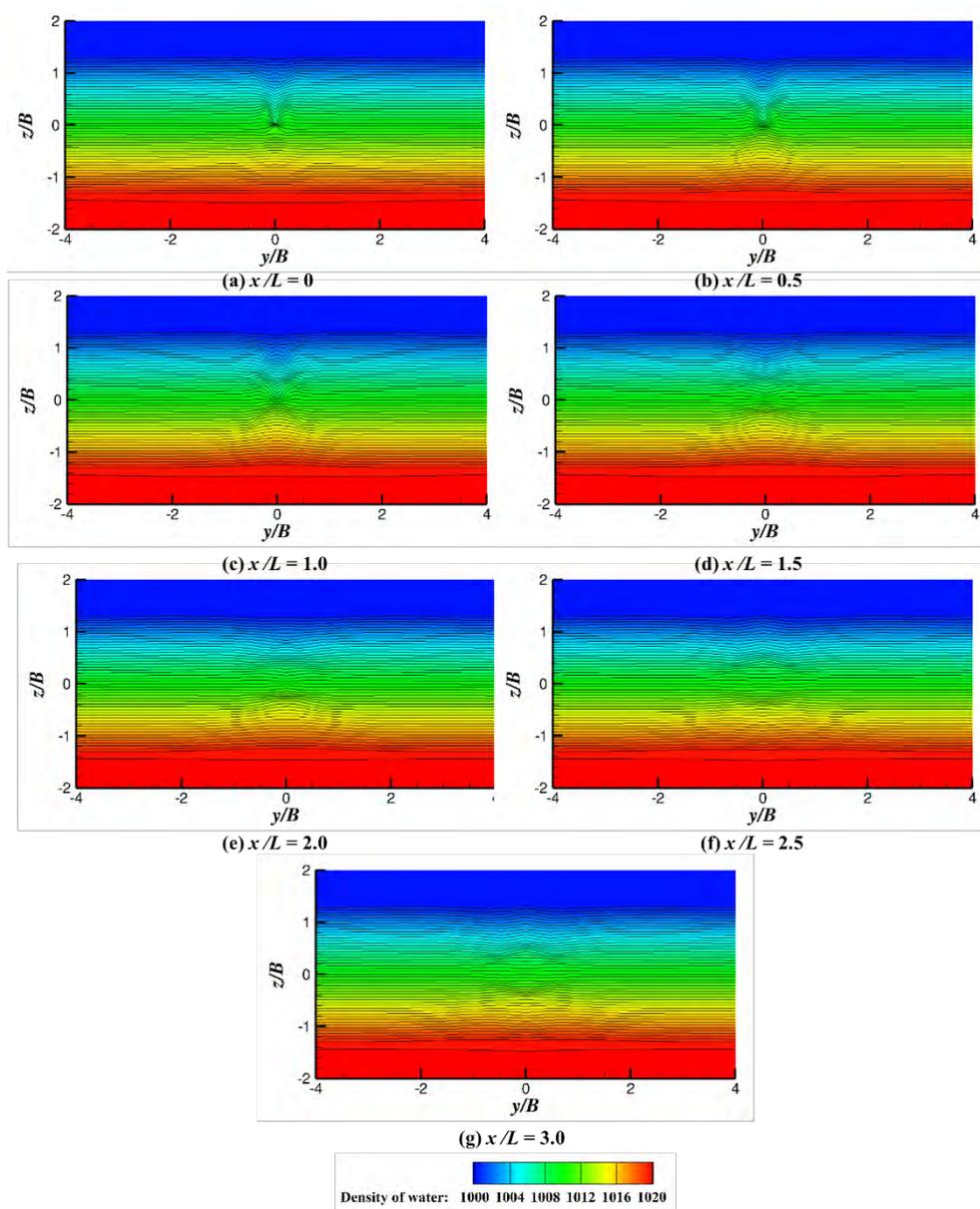
Figure 5. The distribution of density in seven slices along the x direction in the near field of the Joubert BB2 submarine model

According to Spedding[18], projection of motion onto horizontal plane $\{x, y\}$ yields an estimate of the two horizontal velocity components, $\boldsymbol{q} = \{u, v\}$, as well as their gradient values ω_z and Δ_z for internal waves. These two factors are derived to show the flow characteristics. The magnitudes of the divergence field Δ_z, can be utilized as an indicator of internal waves since

they are roughly proportional to the amplitudes of vortex and wave motions in a plane of constant z.

$$\Delta_z = \nabla \cdot \boldsymbol{q} \tag{11}$$
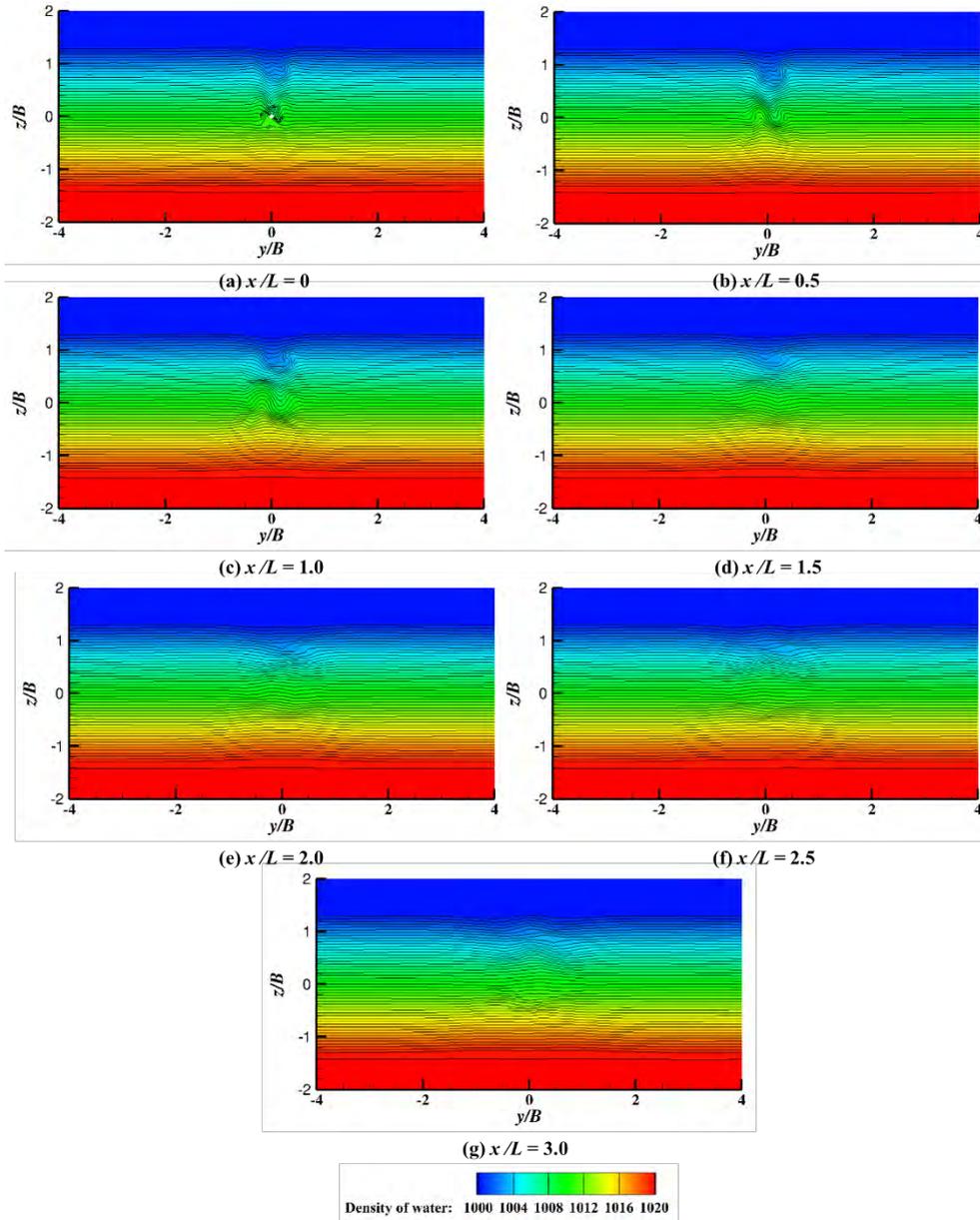
Figure 6. **The distribution of density in seven slices along the x direction in the near field of the Joubert BB2 submarine model with ducted propeller**

As Figure 7 shows, a plane is picked with constant $z = 0$ and colored it using the divergence field Δ_z. The submarine is emitting internal waves after time passed, and internal waves can be seen far behind the submarine due to coherent structures shed from the body and downstream wake turbulence. In Figure 7, it can be seen that shapes of the internal wave emitted by submarines in the three cases are roughly the same, indicating that the shape of internal wave is independent of the geometry, which is also consistent with the conclusion obtained by Spedding[18]. By comparing Figure 7 (a) and (b), we can find that the increase of appendages obviously enhances the strength of the internal waves in the near field. In the far field, by comparing Figure 7 (b) and (c), since the propeller itself acts as a power source, the internal wave intensity and the propagation range both in the x and y directions are increased. Moreover,

as Huang[14] pointed out that internal waves will alter the wave patterns on the free surface, which will discuss in the following part.

(a) Bare hull   (b) Joubert BB2   (c) Joubert BB2 with ducked propeller

internal_wave: -0.005 -0.003 -0.001 0.001 0.003 0.005
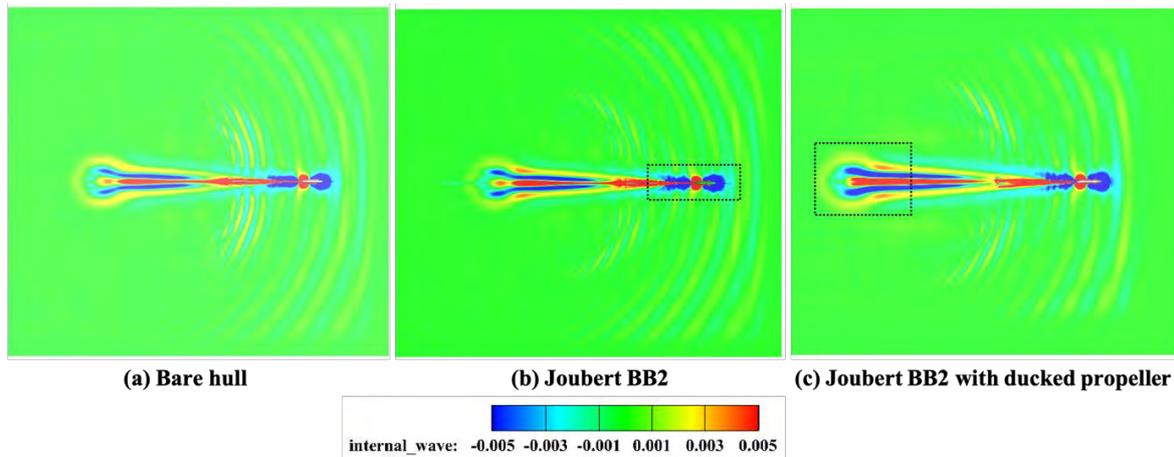
Figure 7. Comparison of the internal wave of the different three cases

Free surface Waves

The displacement of a submerged body in the fluid causes pressure differences around it, which generate waves, these phenomena are depicted in Figures 8, 9, and 10, which show the waveform in different three cases on the free surface over time, respectively. The wave system on the free surface consists of transverse waves and divergent waves. The Kelvin wedge on the free surface, which typically occurs in ship wave systems, can also be thought of as a synonym for it.

Table 3. Comparison of the wave height on the free surface of the different three cases

Case	The maximum wave crest value H_1 (m)	Difference	The maximum wave trough value H_2 (m)	Difference
1	0.0131	-	-0.0124	-
2	0.0206	+57.25%	-0.0174	+40.32%
3	0.0201	+53.44%	-0.0174	+40.32%

Table 4. Comparison of the wake propagation characteristics on the free surface of the different three cases

Case	The longitudinal propagation distances	Difference	The transverse propagation distances	Difference
1	65D	-	39D	-
2	68D	+4.62%	43D	+10.26%
3	69D	+6.15%	44D	+12.82%

When *t*=40s, there are about 40 complete waveforms on the free surface in the bare hull case (case 1), as shown in Figure 8 (h). At this time, the longitudinal propagation distance of the waveform is about 65*D*, and the transverse propagation distance is about 39*D*, where the maximum wave crest value is 0.0131m, and the maximum wave trough value is -0.0124m. From Figures 9 (h) and 10 (h), there are roughly 50 and 53 full waveforms on the free surface for the case of the full-appendage Joubert BB2 submarine (case 2) and the full-appendage

Joubert BB2 submarine with a ducted propeller case (case 3) at $t = 40$s, respectively. Moreover, the longitudinal propagation distances for these two cases are roughly $68D$ and $69D$, and the transverse propagation distances are about $43D$ and $44D$, respectively.

The contribution ratio of appendages affects the free surface characteristics will be mainly analyzed from three aspects: the maximum wave height, the angle of the Kelvin wedge and the wave velocity on the free surface. Table 3 compares the maximum wave height on the free surface with the three cases. In the full-appendage Joubert BB2 submarine case, the maximum wave crest value H_1 is 0.0206m, and the maximum wave trough value H_2 is -0.0174m. The absolute values of the maximum wave crest and trough increased by 57.25% and 40.32%, respectively, when compared to the bare hull case. That means after adding appendages such as sail and rudders, the value of the wave height on the free surface will be significantly increased.

After comparing cases 2 and 3, we found a very interesting phenomenon, the maximum wave crest value H_1 in case 3 is a little bit smaller than in case 2. This may be because of the role of the ducted propeller, it helps reduce the influence of the vertical stratification, and the rotation of the propeller will stir the heavier water to the upper layer Huang[14] so that the value of wave crest and trough will be suppressed.
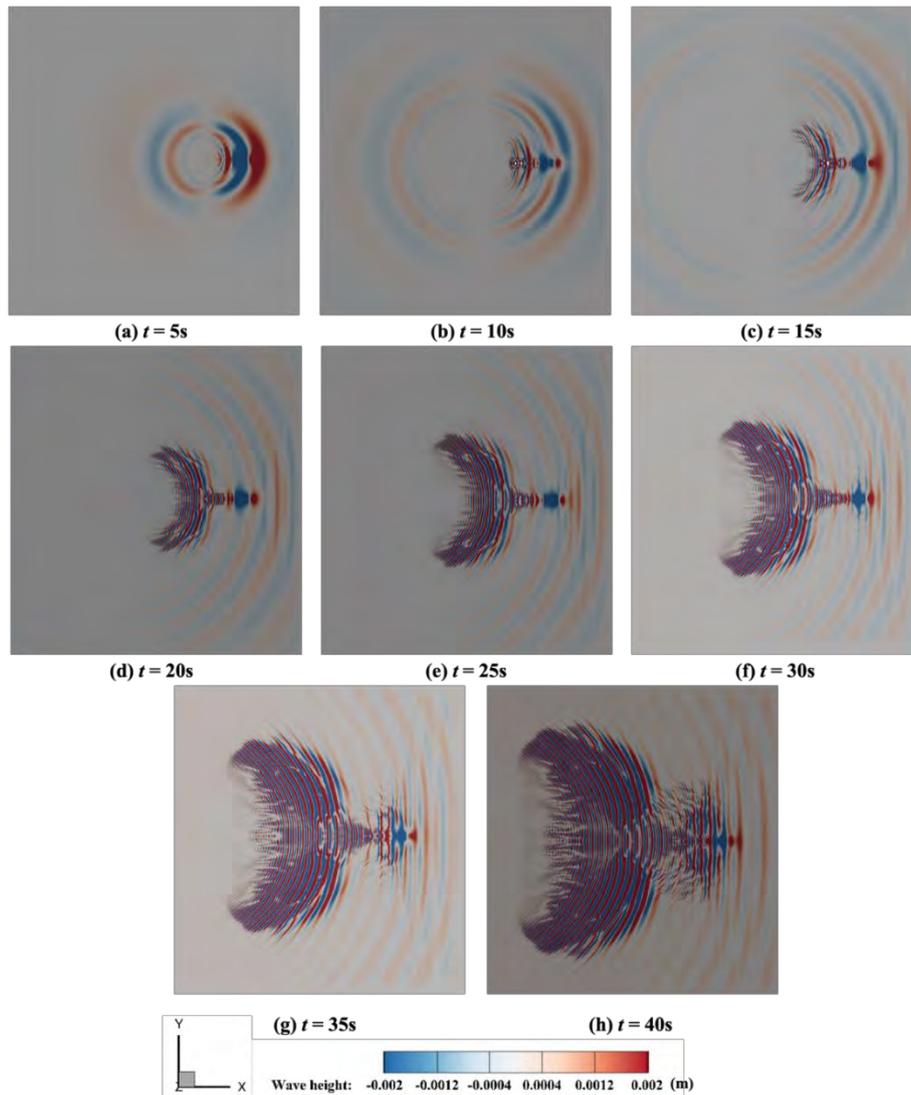
Figure 8. The wave generation process on the free surface of the bare hull model
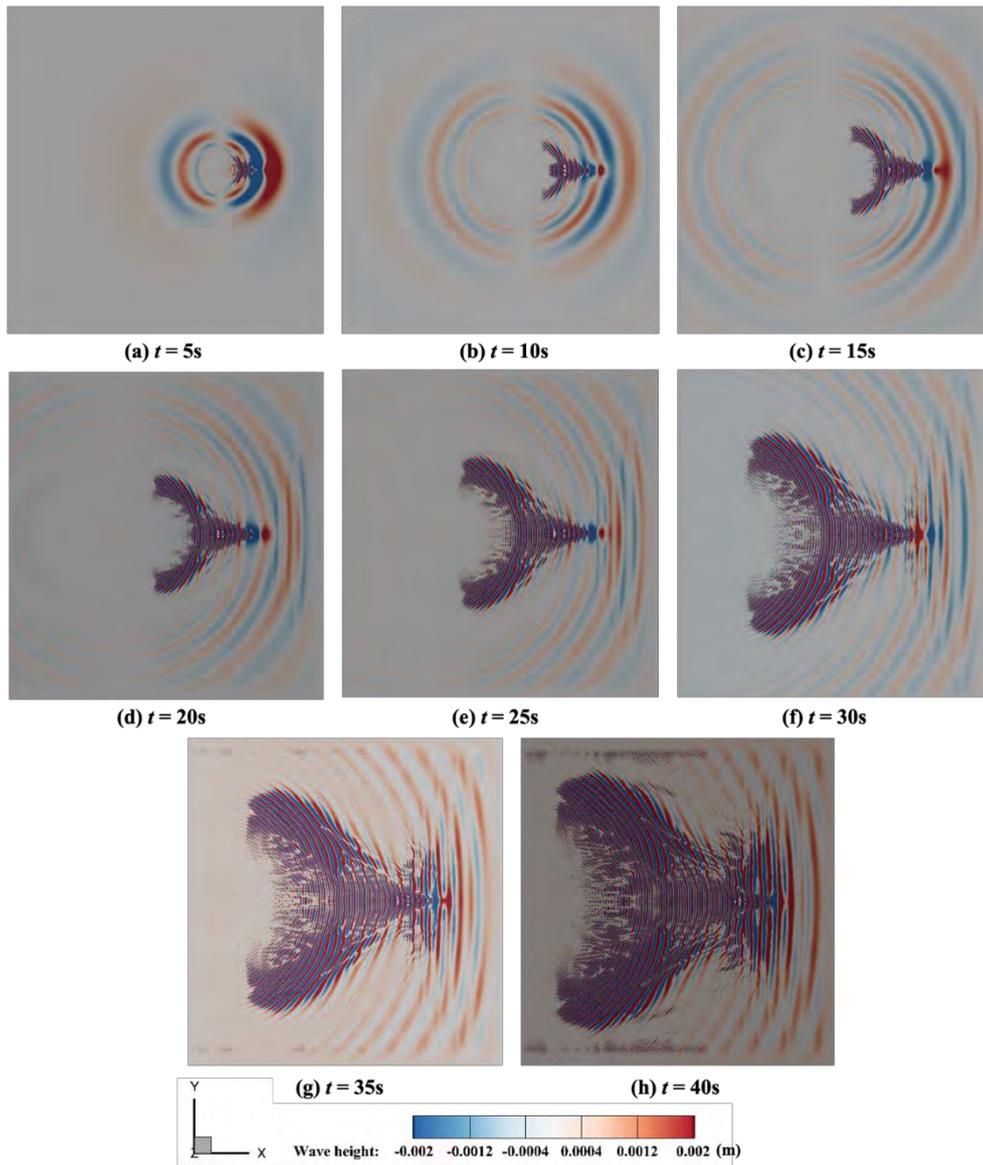
Figure 9. The wave generation process on the free surface of the Joubert BB2 submarine model

Table 4 is a comparison of the wake propagation characteristics on the free surface of the three cases. In case 1, the longitudinal propagation distance on the free surface of the waveform is about $65D$, and the transverse propagation is about $39D$, where D is the draft to duck of the Joubert BB2 submarine. After adding appendages, in case 2, the waveform on the free surface becomes "longer" (in x direction) and "wider" (in y direction), where the longitudinal propagation distance increases by 4.62%, and the transverse propagation distance increase by 10.26% compared case 1. It can also be observed in Figures 8 (h) and 9 (h), the Kelvin wedge angles increase gradually after adding appendages. The longitudinal and transverse propagation distances of case 3 are the longest and widest, which are $69D$ and $44D$, respectively. Although the ducted propeller helps reduce the influence of the vertical stratification compared to a normal propeller, it does not affect itself as a power source to make wakes on the free surface spread farther and wider. Compared to case 1, the longitudinal propagation distance increases by 6.15%, and the transverse propagation distance increases by about 12.82%.
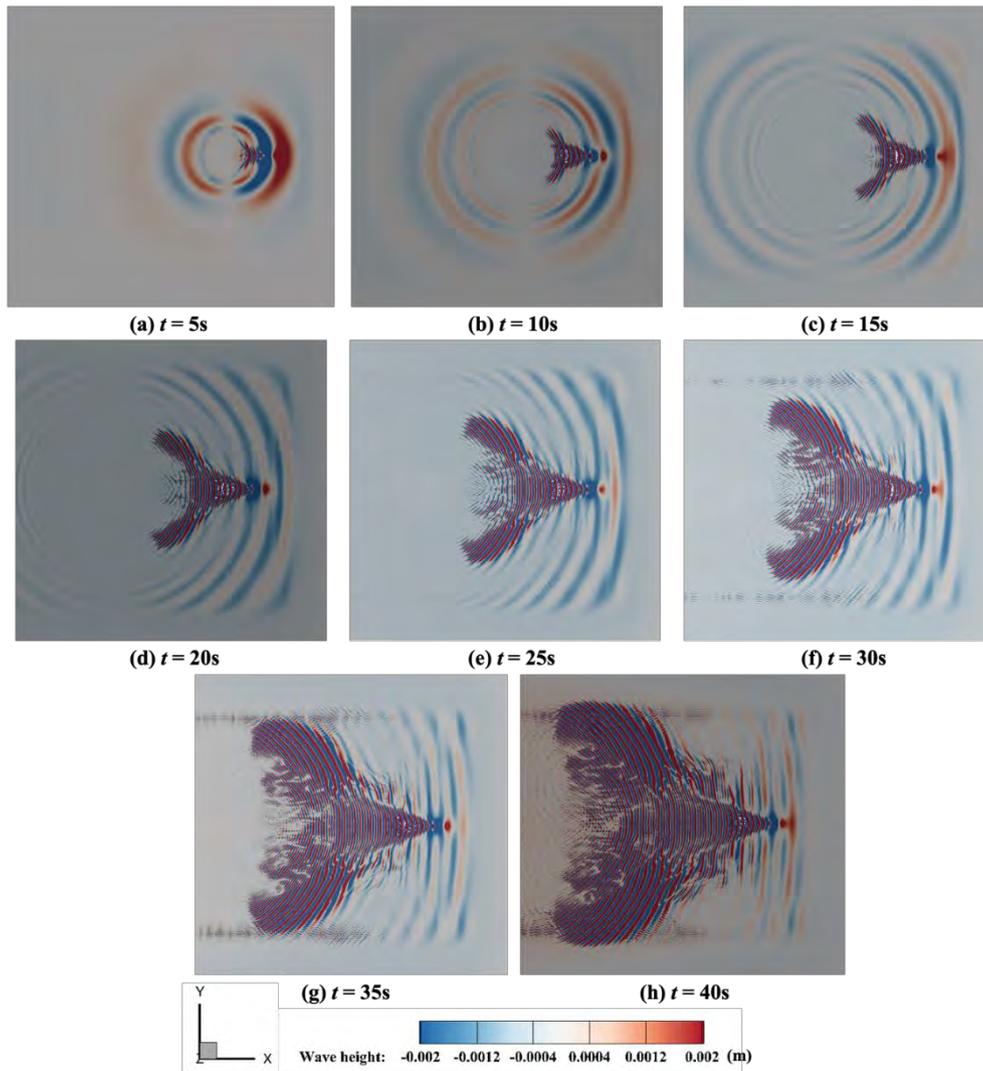
Figure 10. The wave generation process on the free surface of the Joubert BB2 submarine model with a ducted propeller

It is crucial to quantitatively assess how the steady-state free surface profile differs with three cases when examining the waveform on the free surface. Five *x-z* slices are produced in each of these three cases at the free surface as shown in Figure 11. These slices are positioned at *y/B* = 0, 0.25, 0.5, 0.75, and 1.0, where *y* = 0 corresponds to the mid-plane of the submarine and *y* = *B* represents to slices placed a half width beyond the port and starboard edge of the submarine, *B* is the beam.

The steady-state free surface profiles along a submarine's length are displayed on each slice of Figure 11. In these pictures, the location of the submarine is [0, 6.7*D*], and considered about 50*D* downstream evolution on the free surface. The black line in Figure 11 represents the bare hull case, which also is called case 1, the red line is the case of the full-appendage Joubert BB2 submarine, also named case 2 for simplicity, the blue line indicates the full-appendage Joubert BB2 submarine with a ducted propeller case, also called it case 3. One feature is clearly shown in Figure 11 (a) ~ (f) is that the free surface wave heights display a typical oscillatory behavior both in three cases. Moreover, the anisotropy of the wave height in the free surface is more pronounces due to stratification, where the values of wave crest are slightly larger than the wave trough in three cases. This phenomenon is in coincidence with Huang[14].

Secondly, adding appendages will significantly increase the crest and trough values on the free surface, and by comparing the curves under different slices, it is found that the transverse propagation is significantly increased after adding sail and rudders, and within the range of 1 time the width of the submarine, the value of the wave crest and trough will not attenuate.

(a) Location of the chosen 5 specific planes

(b) y = 0B

(c) y = 0.25B

(d) y = 0.5B

(e) y = 0.75B

(f) y = 1.0B
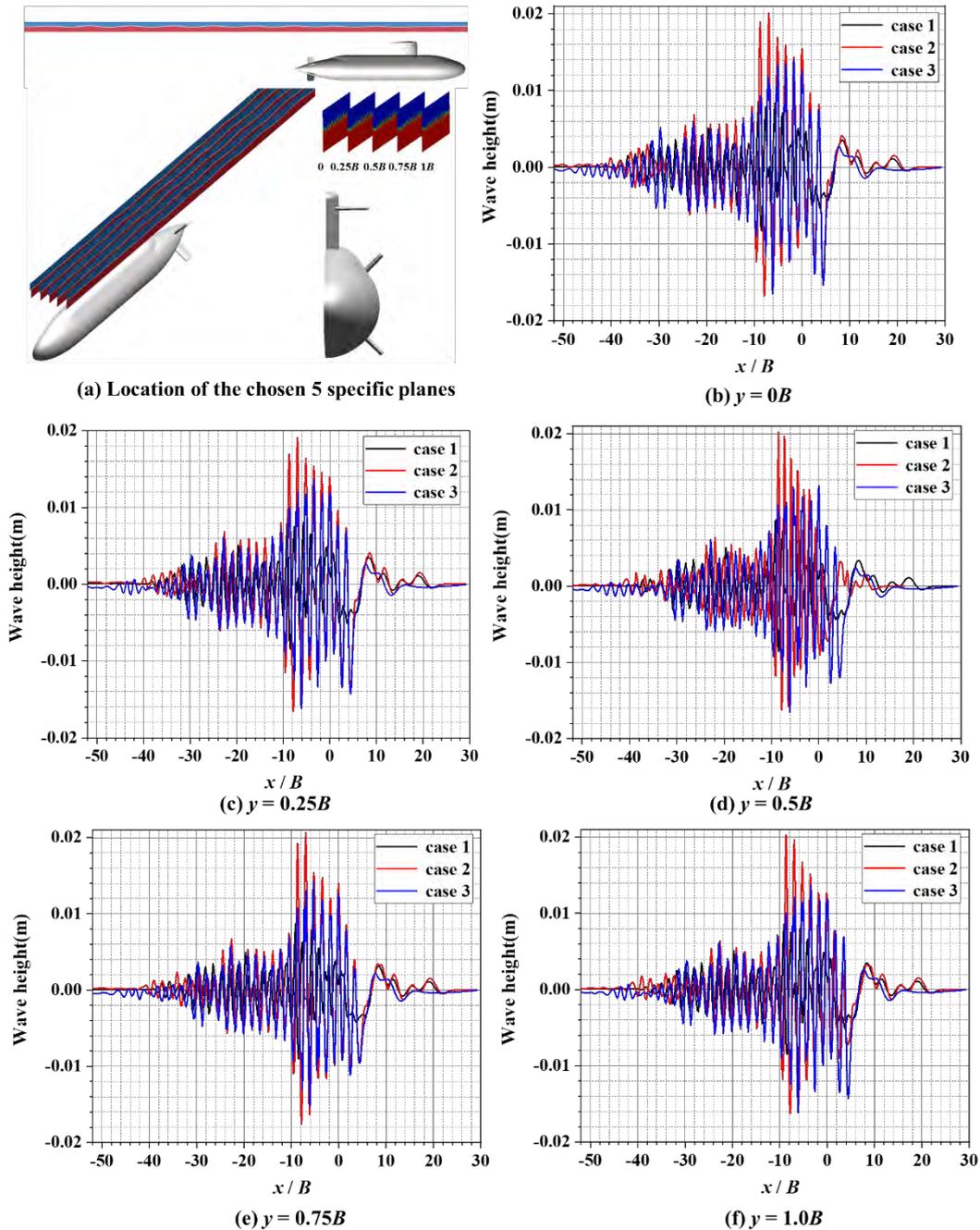
Figure 11. Comparison of the wave height on the free surface of the different three cases (chosen specific 5 planes）

Finally, due to the rectification effect of the ducted propeller, there is little difference between the crest and trough values of the free surface between case 2 and 3, but the propagation distance in the x direction of case 3 is 10D longer than case 2, and 15D longer than case 1. That means

the ducted propeller will make the wave trace left by the submarine on the free surface exist farther, wider, and last for a longer time.

Additionally, the axial velocity contour on the free surface are examined in three cases shown in Figure 12, the axial velocity contour has similar shapes and trend with the wave heights. The axial velocity of case 3 is higher than the other two cases, in short, the existence of the ducted propeller will cause the waveform formed by the submarine on the free surface to be "farther", "wider", "faster" and longer duration.
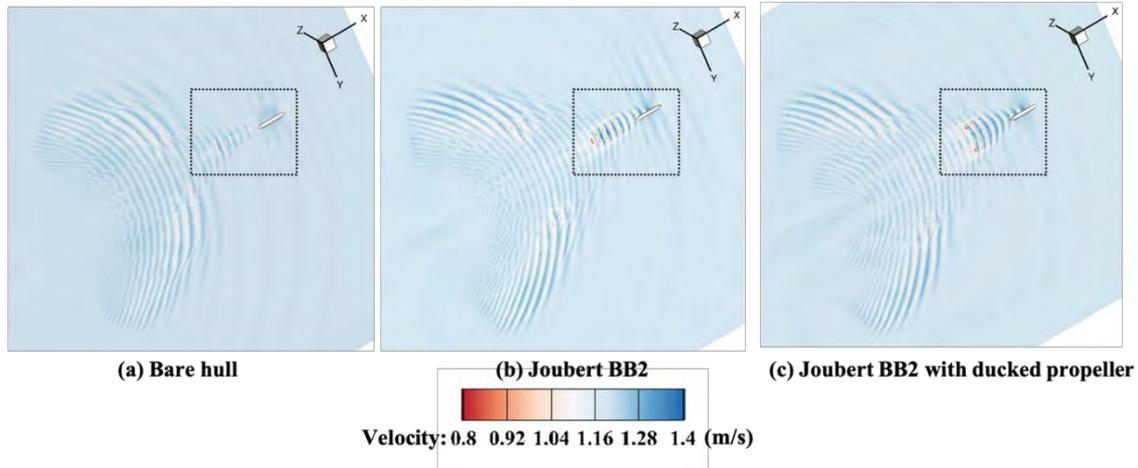
Figure 12. Comparison of the axial velocity on the free surface of the different three cases

Conclusions

Take the Joubert BB2 submarine model as an example, three different types of submarines are simulated: a bare hull, a full-appendage Joubert BB2 submarine, and the full-appendage Joubert BB2 submarine with a ducted propeller. All the simulations are placed in the linearly stratified fluid background with a constant buoyancy frequency. We analyzed the density fluctuations, internal waves, and complex hydrophysical signatures on the free surface when submarines operate near the free surface, respectively. The proportional contribution of appendages and ducted propeller to waveform characteristics on the free surface were quantitatively assessed. The conclusions are as follows:

(1) After adding sails and rudders, compared with the bare hull model, the maximum value of the wave crest on the free surface increases by about 57.25%, and the maximum value of the wave trough increases by 40.32%. The longitudinal and transverse propagation distances increase by 10.26% and 4.62%, respectively. It means that the phenomenon on the free surface will be more obvious, which raises the possibility of being detected.

(2) The ducted propeller has little effect on the value of the wave crest and trough on the free surface, but in terms of propagation distance, the longitudinal and transverse propagation distances increase by 1.53% and 2.56% compared full-appendage Joubert BB2 submarine case. In short, the existence of the ducted propeller will cause the waveform formed by the submarine on the free surface to be "farther", "wider", "faster" and longer duration.

Acknowledgements

References

[1] Diamessis P J, Domaradzki J, Hesthaven J. A spectral multidomain penalty method model for the simulation of high Reynolds number localized incompressible stratified turbulence[J]. Journal of Computational Physics, 2005, 202(1): 298-322.

[2] Diamessis P J, Spedding G R, Domaradzki J A. Similarity scaling and vorticity structure in high-Reynolds-number stably stratified turbulent wakes[J]. Journal of Fluid Mechanics, 2011, 671: 52-95.

[3] Zhou Q, Diamessis P J. Large-scale characteristics of stratified wake turbulence at varying Reynolds number[J]. Physical Review Fluids, 2019, 4(8): 084802.

[4] Cao L s, Huang F l, Liu C, et al. Vortical structures and wakes of a sphere in homogeneous and density stratified fluid[J]. Journal of Hydrodynamics, 2021, 33(2): 207-215.

[5] Ortiz-Tarin J L, Chongsiripinyo K C, Sarkar S. Stratified flow past a prolate spheroid[J]. Phys. Rev. Fluids, 2019, 4: 094803.

[6] Mercier M, Wang S, Péméja J, et al. Settling disks in a linearly stratified fluid[J]. Journal of Fluid Mechanics, 2020, 885: A2.

[7] Chashechkin Y D, Zagumennyi I V. Formation of waves, vortices and ligaments in 2D stratified flows around obstacles[J]. Physica Scripta, 2019, 94(5): 054003.

[8] Esmaeilpour, M., 2017. A ship advancing in a stratified fluid: the dead water effect. Doctoral Thesis, The university of Iowa.

[9] Li, J., Martin, J.E., Carrica, P.M., 2019. Large-scale simulation of ship bubbly wake during a maneuver in stratified flow. Ocean Eng. 173, 643–658.

[10] Carrica, P.M., Kim, Y., Martin, J.E., 2019. Near-surface self propulsion of a generic submarine in calm water and waves. Ocean Eng. 183, 87–105.

[11] Xue, F., Jin, W., Qiu, S., Yang, J., 2020. Wake Features of Moving Submerged Bodies and Motion State Inversion of Submarines. IEEE Access 8, 12713–12724.

[12] Ma, W., Li, Y., Ding, Y., Duan, F., Hu, K., 2020. Numerical investigation of internal wave and free surface wave induced by the DARPA Suboff moving in a strongly stratified fluid. Ships Offshore Struct. 15, 587–604.

[13] Ma, W., Li, Y., Ding, Y., Hu, K., Lan, L., 2018. Numerical simulations of linearly stratified flow past submerged bodies. Pol. Marit. Res. S 3.

[14] Huang, F., Meng, Q., Cao, L., Wan, D., 2022. Wakes and free surface signatures of a generic submarine in the homogeneous and linearly stratified fluid. Ocean Engineering 250, 111062.

[15] Joubert, P.P.N., 2006. Some Aspects of Submarine Design - Part 2. Shape of a Submarine 2026.

[16] Menter, F.R., Kuntz, M., 2004. Adaptation of Eddy-Viscosity Turbulence Models to Unsteady Separated Flow Behind Vehicles, in: McCallen, R., Browand, F., Ross, J. (Eds.), The Aerodynamics of Heavy Vehicles: Trucks, Buses, and Trains, Lecture Notes in Applied and Computational Mechanics. Springer, Berlin, Heidelberg, pp. 339–352.

[17] Hirt, C.W., Nichols, B.D., 1981. Volume of fluid (VOF) method for the dynamics of free boundaries. J. Comput. Phys. 39, 201–225.

[18] Spedding, G.R., 2014. Wake signature detection. Annu. Rev. Fluid Mech. 46, 273–302

Numerical simulation of damped free liquid sloshing inside a spherical tank

*Congyi Huang, Weiwen Zhao, †Decheng Wan

Computational Marine Hydrodynamic Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering,
Shanghai Jiao Tong University, Shanghai 200240, China

*Presenting author: cyhuang@sjtu.edu.cn
†Corresponding author: dcwan@sjtu.edu.cn

Abstract

Liquid sloshing is a very complex phenomenon with large free surface deformation and strong nonlinear characteristics, which may bring danger to the safety of the structure. On one hand, the violently sloshing will exert large slamming pressure on the bulkhead, leading to the deformation even the damage of the structure. On the other hand, the liquid sloshing will stimulate the motion of the tank causing dynamic instability. Therefore, it is of great significance to study the liquid sloshing. In this paper, the inner-house solver based on MPS method MLParticle-SJTU is used to simulate the liquid sloshing phenomenon inside a spherical tank according to the design requirements of aerospace vehicle. Considering h/R=0.2~1.8 six groups of different liquid filling rates, apply initial excitation to the tank and allow it to decay freely. The natural frequencies and damping of different filling rates are calculated respectively. The results show that with the increase of water depth, the natural frequency of liquid sloshing increases, and the damping coefficient decreases.

Keywords: Moving Particle Semi-implicit (MPS), liquid sloshing, Spherical tank, natural frequency, free decay motion.

Introduction

Liquid sloshing is a phenomenon happens inside a partially loaded tank under external excitation. This phenomenon is produced by the combined effect of surface tension and gravity, and it is a very complex fluid motion phenomenon There is a free surface, which may have large deformation such as curling and breaking when the sloshing phenomenon is intense, showing a strong nonlinearity. Liquid sloshing is very common in daily life, for example, when a partially loaded liquid cargo ship (such as LNG, LPG or crude oil ship) encountered wave loads, the movement of the tank will stimulate the internal liquid sloshing. When an aircraft is taking-off, taxiing, landing, or adjusting attitude in the air, the liquid sloshing will also be triggered inside the oil storage tanks.

The occurrence of liquid sloshing is highly randomized and its potential damage to the structure cannot be ignored. On the one hand, the violently sloshing will generate a great pressure on the bulkhead, leading to deformation or damage of the structure. On the other hand, the liquid sloshing may also stimulate the motion response of the structure, which will cause the dynamic instability. Therefore, it is of great significance to accurately simulate and forecast the liquid sloshing phenomena that may occur inside the tank.

There are a lot of researchers who has been working on this liquid sloshing problems. Yang and Peugeot (2014) used CFD to simulate the phenomenon of tank sloshing in a spherical liquid tank and gave the equivalent sloshing mass and the equivalent center of mass position through the equations [1]. Gabriele et al. (2018) simulated the dynamics of a ship with internal liquid tank sloshing in the time domain by a joint simulation approach [2]. Cao et al.(2019) simulated the liquid sloshing phenomenon in a broken ship using the SPH method

[3]. Álvaro et al.(2021) have studied the equilibrium meniscus and axisymmetric oscillations of a ferrofluid solution in a cylindrical tank subjected to a static nonuniform magnetic field under microgravity conditions for the first time [4]. Mashy et al. (2021) used the SPH method to accurately simulate long periods of vigorous oscillatory flow in an arbitrarily shaped partially filled water tank [5]. Chen et al. (2022) performed numerical simulations using the fluid-volume method to study the liquid-gas interface on the rotor surface under microgravity conditions and proposed a dimensionless theoretical solution for the profile [6]. Wang et al. (2022) performed computational fluid dynamics (CFD) simulations using a volume-of-fluid (VOF) model to analyze the effects of propellant volume fill ratio and acceleration conditions on fluid flow in a blade-type surface tension fuel tank [7]. Yang et al. (2022), simulated the intense sloshing flow by Smoothed Particle Hydrodynamics (SPH) method [8].

When the external excitation frequency is close to the system natural frequency, the system will reach the resonance state. At this time, the liquid sloshing is the most violent, the sloshing amplitude is the largest, and the load applied to the structure is also the highest. Therefore, we need to predict the natural frequency with different water depths. In this paper, the freely damped sloshing phenomena inside a tank with different water depth are simulated respectively. The natural frequency and the damping coefficient of the system is calculated.

Numerical method

In this paper, a meshless particle method, the Moving Particle Semi-implicit Method (MPS) is used to simulate the liquid sloshing inside a spherical liquid cabin. The MPS method is a CFD method based on the Lagrangian representation. Different from the traditional mesh-based methods, the MPS method discretizes the computational domain into a series of particles, which do not have a fixed topological relationship with each other but carry their own physical information, such as displacements, velocities, accelerations, etc., respectively. Compared with the traditional mesh-based methods, the particle methods has the advantage that, since there is no mesh in its computation, there will be no mesh distortion or brokenness when simulating the violent flows with free surface large deformation. And the meshless particle method has high quality in tracking the free surface, dealing with the moving boundary and other problems. Therefore, the meshless particle method is very suitable for simulating problems with free surface large deformation characteristics, such as dam-break, water entry and liquid sloshing.

A lot of works based on the MPS method has been carried out by our group, including algorithm improvement, solver development, and applications For solving the pressure oscillation problems that often occur in the traditional MPS method when calculating the pressure by the Poisson pressure equations, the group has proposed a set of improved MPS algorithms by improving the pressure solving method, the pressure gradient model, and the method of judging the free surface [9]. Based on this improved MPS method, our group has developed a meshless particle method solver MLParticle-SJTU, which can be run on the CPU, and the multi-CPU parallel computing can also be realized [10]. In order to make full use of the advantages of the MPS method in simulating violent flows, the group also developed algorithms for coupling the MPS method with other methods, including the MPS-FEM coupling algorithm for simulating fluid-structure coupling phenomena and independently developed the MPSFEM-SJTU solver [11]. And the MPSDEM coupling algorithm for simulating particle flows and independently developed the MPSDEM-SJTU solver [12]. However, due to the low efficiency, the MPS method cannot be used in the simulation of large-scale large scale computational applications. Therefore, a high-performance meshless particle solver MPSGPU-SJTU which can run on GPU is developed by our group, and it can improve the computational speed greatly while guaranteeing the precision and accuracy of the

original particle solver [13]. In previous published papers, the numerical method of the MPS method has been introduced in detail [14-15]. Therefore, in this paper, the MPS method will be introduced just simply in the following part.

Governing equations

The governing equations for MPS method is shown as Eqs. (1), (2).

$$\frac{1}{\rho}\frac{D\rho}{Dt} = -\nabla \cdot V = 0 \tag{1}$$

$$\frac{DV}{Dt} = -\frac{1}{\rho}\nabla p + \nu\nabla^2 V + g \tag{2}$$

where ρ is the fluid density, V is the velocity vector, p is the pressure, ν is kinematic viscosity, g is gravitational acceleration vector and t indicates time.

Particle interaction models

In MPS, particles interact with each other by a kernel function. The kernel function used in this paper is given in Eqs. (3a), 3(b).

$$W(r) = \frac{r_e}{0.85r + 0.15r_e} - 1, \ 0 \le r < r_e \tag{3a}$$

$$W(r) = 0, \ r_e \le r \tag{3a}$$

where r is the particle distance, r_e is the influence radius

The following equations are the particle interaction models used to discretize the terms in the governing equations.

$$\langle \nabla P \rangle_i = \frac{D}{n^0}\sum_{j \ne i}\frac{P_j + P_i}{\left|r_j - r_i\right|^2}(r_j - r_i)W\left(\left|r_j - r_i\right|\right) \tag{4}$$

$$\langle \nabla \cdot V \rangle_i = \frac{D}{n^0}\sum_{j \ne i}\frac{(V_j - V_i) \cdot (r_j - r_i)}{\left|r_j - r_i\right|^2}W\left(\left|r_j - r_i\right|\right) \tag{5}$$

$$\langle \nabla^2 \phi \rangle_i = \frac{2D}{n^0\lambda}\sum_{j \ne i}(\phi_j - \phi_i)W\left(\left|r_j - r_i\right|\right) \tag{6}$$

$$\lambda = \frac{\displaystyle\sum_{j \ne i}W\left(\left|r_j - r_i\right|\right)\left|r_j - r_i\right|^2}{\displaystyle\sum_{j \ne i}W\left(\left|r_j - r_i\right|\right)} \tag{7}$$

$$\langle n \rangle_i = \sum_{j \ne i}W\left(\left|r_j - r_i\right|\right) \tag{8}$$

where D is the number of dimensions, r is the position vector and n^0 is the particle number density at the initial time, and the particle number density is defined as Eq. (8).

Pressure Poisson equation

In this paper, the pressure is calculated by solving the mixed source term Pressure Poisson equation (PPE) proposed by Tanaka and Masunaga[16], Lee et al.[17], which can be written as the following equation.

$$\left\langle \nabla^2 p^{k+1} \right\rangle_i = (1-\gamma)\frac{\rho}{\Delta t}\nabla \cdot V_i^* - \gamma\frac{\rho}{\Delta t^2}\frac{\left\langle n^k \right\rangle_i - n^0}{n^0} \tag{9}$$

where p^{k+1}, Δt and V_i^* are the pressure of the step $k+1$, time step and temporal velocity and γ is the weight of the particle number density term between 0 to 1.

Free surface detection

The particle number density is used to determine whether a particle is a free surface particle. If $\langle n \rangle_i / n^0 < 0.80$, the particle will be determined as a free surface particle. If $\langle n \rangle_i / n^0 > 0.97$, the particle will be regarded as an internal fluid particle. If $0.80 < \langle n \rangle_i / n^0 < 0.97$, an additional function needs to be introduced.

$$\langle F \rangle_i = \frac{D}{n^0}\frac{(r_i - r_j)}{|r_i - r_j|}\sum_{j \neq i} W(|r_i - r_j|) \tag{10}$$

where $\langle F \rangle_i$ can be regarded as the asymmetric arrangements of neighbor particles. $|F|^0$ is the initial value of $|F|$ for surface particles. If $\langle |F| \rangle_i > 0.9|F|^0$, the particle will be regarded as a free surface particle.

Boundary condition

The multilayer particles boundary condition is applied to fulfil the support region of the particles near the boundary, as shown in Fig. 1. The pressure of the wall particle is calculated by solving the PPE equation, and ghost particle pressure is obtained by interpolation.
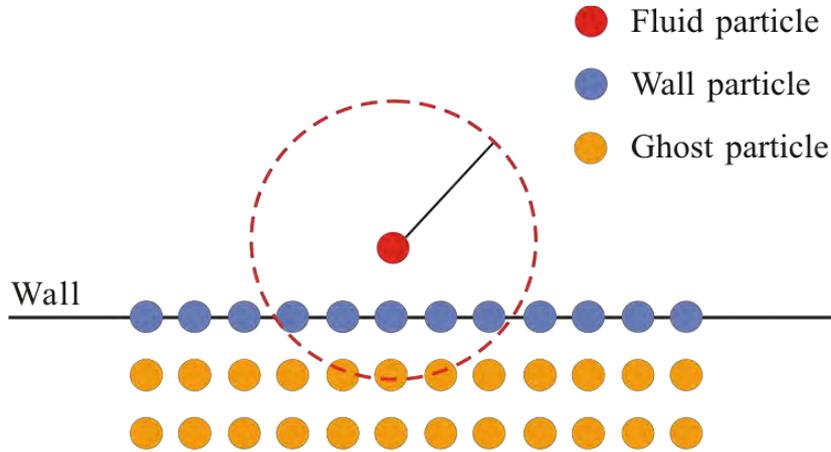
Figure 1. Schematic of boundary particles

Free surface boundary condition

Virtual particles are arranged over the free surface to fulfill the support domain of the free surface particle, as shown in Fig. 2. And the particle number density and the pressure Laplacian model is modified as:

$$\langle n \rangle_i = \sum_{j \neq i} W_{ji} + \sum_{\text{virtual}} W_{\text{virtual},i} = + \sum_{\text{virtual}} W_{\text{virtual},i} \tag{11}$$

$$\left\langle \nabla^2 p^{k+1} \right\rangle_i = \frac{2D}{n^0 \lambda} \sum_{j \neq i} (p_j - p_i) W_{ji} + \frac{2D}{n^0 \lambda} \sum_{\text{virtual}} (p_{\text{virtual}} - p_i) W_{\text{vritual},i} \tag{12}$$

where n_i^* and $W_{\text{vritual},i}$ denote the particle number density of actual neighboring particles and virtual neighboring particles separately.
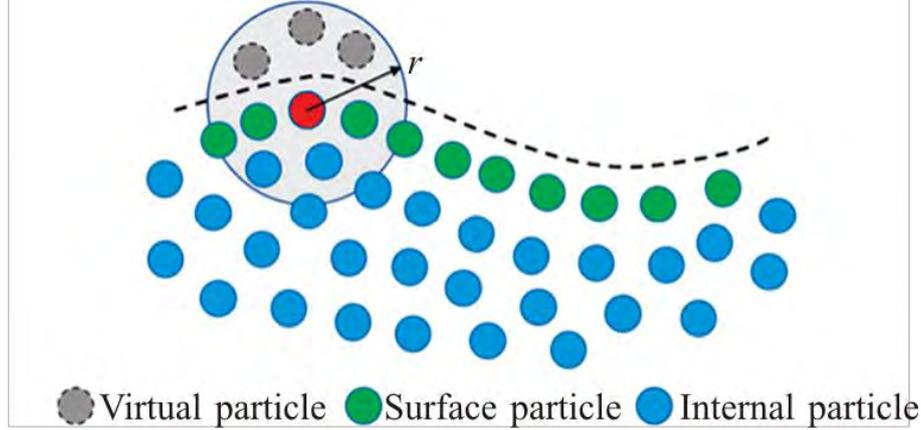
Fig. 2 Schematic of free surface boundary condition

Surface tension model

The surface tension is imposed by adding a source term to the governing equations, which is proposed by Brackbill et al.[18], as the following equations show.

$$\rho \frac{\mathrm{D}\boldsymbol{V}}{\mathrm{D}t} = -\nabla p + \mu \nabla^2 \boldsymbol{V} + \boldsymbol{F}^V + \boldsymbol{F}^S \tag{13}$$

$$\boldsymbol{F}^S = \sigma \kappa \nabla C \tag{14}$$

where σ is surface tension coefficient, κ is inter- face curvature and C is function to mark the free surface particles. For free surface particles $C = 1$, and for internal particles $C = 0$.

Numerical Simulation

In this paper, the free damping liquid sloshing phenomena inside a spherical tank with different water depth is simulated. The diameter of the tank is R=2 m, and six liquid depths are set up with h/R=0.2, 0.4, 0.8, 1.2, 1.6, 1.8 respectively. The initial particle spacing is Δx=0.008 m=1/250R, and the time step is Δt=0.0008 s. The total number of particles corresponding to the six water depths are 181877, 254541, 516971, 665064, 919709, and 1040385, respectively.

The liquid inside the tank is stationary at the starting moment. So the transverse oscillation motion is applied to the tank as initial excitation. The transverse oscillation frequency is 3.14 rad/s, and the duration is 2s. After the initial excitation, the liquid sloshing inside the tank decays freely. During the simulation, the center-of-mass coordinates of the liquid in the x, y, and z directions are recorded. And we pay attention to the deformation of the free surface. The frequency of the sloshing can be calculated from the time history curve. And the damping coefficient is calculated by Eq. (15).

$$\xi = \frac{\ln\left(x_n / x_m\right)}{2\pi m} \tag{15}$$

The numerical simulation in this paper is done by our in-house meshless particle method solver solver MLParticle-SJTU. This solver has been used to simulate the liquid sloshing problems inside different shapes of liquid tanks many times. For example, Chen et al. (2019) simulated the sloshing phenomena inside a rectangular tank and a liquefied natural gas (LNG) tank [19]. Zhang et al. simulated Faraday waves inside a rectangular cabin with heave motion [20]. Huang et al. (2023) simulated Faraday waves inside a cylindrical cabin and a hexagonal prism cabin with heave motion [21]. In these previously papers, the simulation results are in good agreement with the published results obtained by other numerical methods in other groups, which can verify the correctness and effectiveness of the solver used in this paper. Therefore, the correctness of the solver will not be verified again in this paper.

Fig. 3 shows the time-history curves of the coordinates of the mass center when the water depth is h/R = 0.2. In the picture, the black line indicates the coordinates along the sway direction, the blue line indicates the coordinates in the direction of gravity, and the green line indicates the coordinates in the direction perpendicular to the paper surface. As can be seen from the curves that the time-history curve of the coordinate perpendicular to the paper is always approximately zero, indicating that the liquid center of mass nearly has no displacement in this direction and shows a good symmetry of the sloshing. In the gravity direction, the coordinate curve also shows only small fluctuations. The sloshing phenomenon of the liquid is mainly shown in the tank movement direction. From Fig. 3, it can be seen that after the initial excitation lasting for 1s, the liquid sloshing shows a free decay state, with the peak value gradually decreasing, which agrees well with the characteristics of the free decay motion. Through the calculation of this curve, it can be obtained that, , the intrinsic frequency of the liquid oscillation is about 4.1688 rad/s at this water depth, and the damping ratio is about 0.3997. Fig. 4 shows the distribution of the liquid and the pressure distribution. It can be seen from the figure that the fluid pressure distribution is relatively uniform, which verifies the stability of the solver in solving the problem. And it can be seen from the figure that at this water depth, the sloshing phenomenon in the liquid tank is not violent, and the shape of the free surface is kept well with no surface curling and breaking.

Fig. 3 The mass center coordinates with h/R=0.2

(a) t=1.04s (b) t=1.36s

(c) t=1.68s (d) t=2.08s
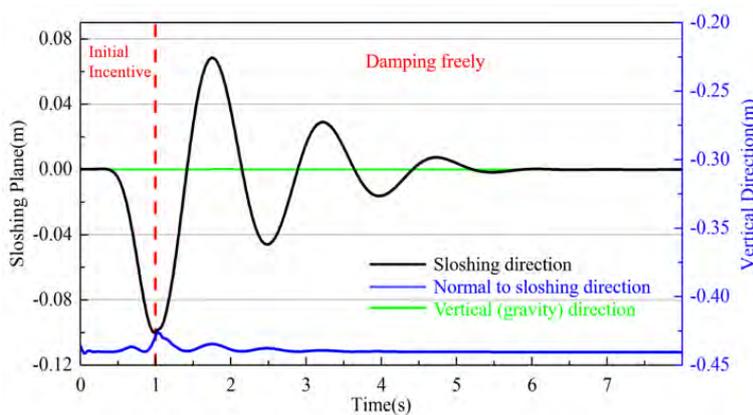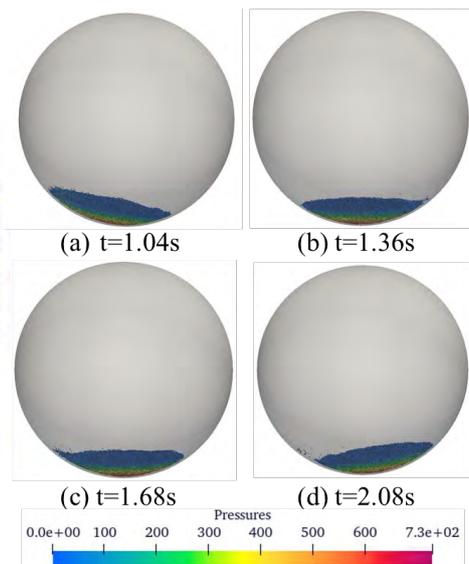
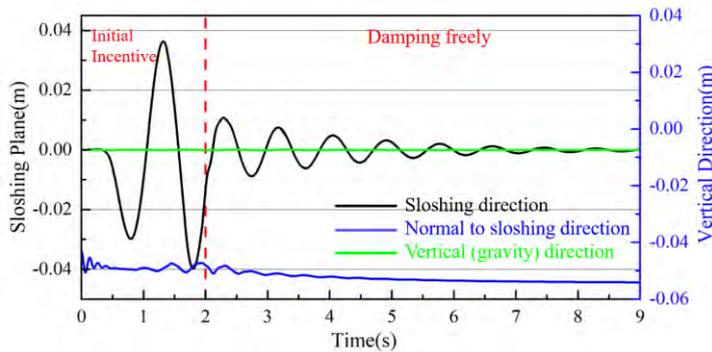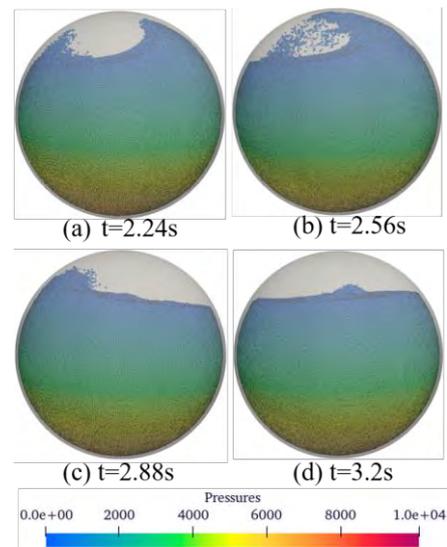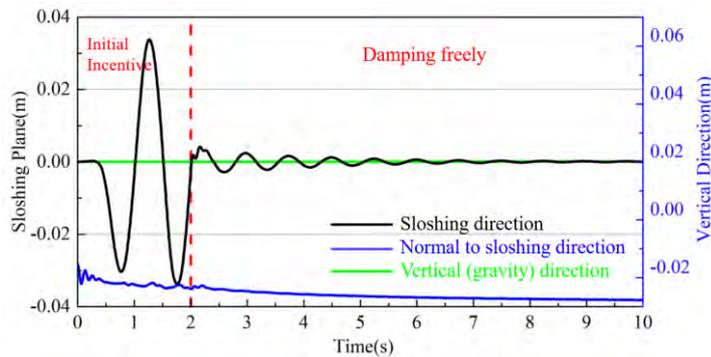**Fig. 4 Liquid distribution of the liquid inside the tank at different times with h/R=0.2**
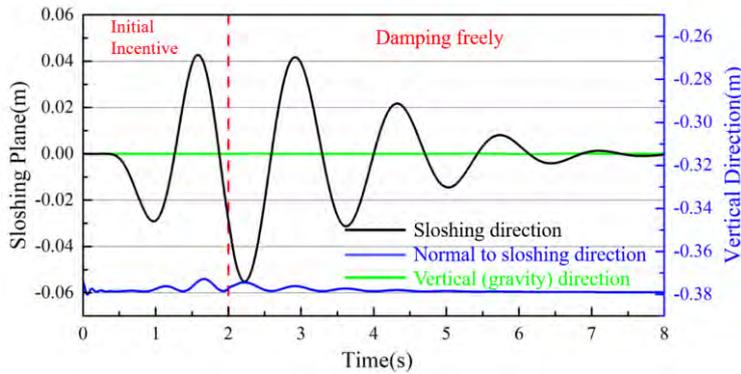
Fig. 5 The mass center coordinates with h/R=0.4
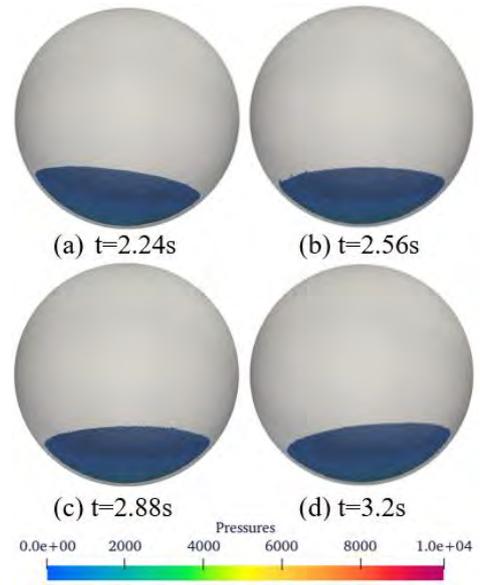
Fig. 6 Liquid distribution of the liquid inside the tank at different times with h/R=0.4

Fig. 5 shows the time history curve of the coordinates of the mass center when the water depth is h/R=0.4. It can be seen from this graph that the characteristics of the curve are basically the same as the case with water depth h/R=0.2. The liquid sloshing is mainly manifested in the direction along the tank movement, and the movement along the other two directions is very small. Through the calculation of the curve, it can be obtained that at this depth, the frequency of free attenuation is 4.3975 rad/s, and the damping ratio is 0.2267. Fig. 6 represents the pressure distribution of the liquid at this depth. It can be seen from the figure that the liquid sloshing phenomenon also presents a relatively smooth state at this water depth, and there is no surface curling or breaking.
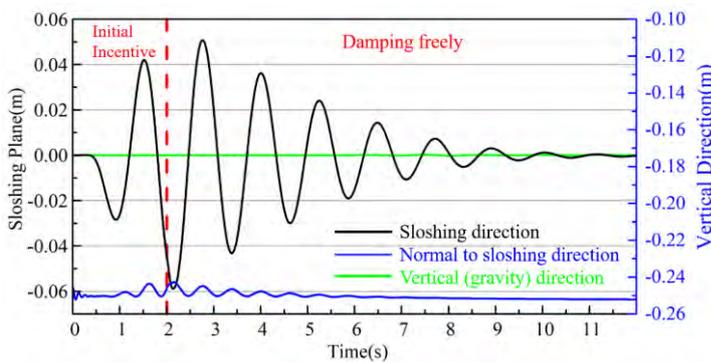
Fig. 7 The mass center coordinates with h/R=0.8
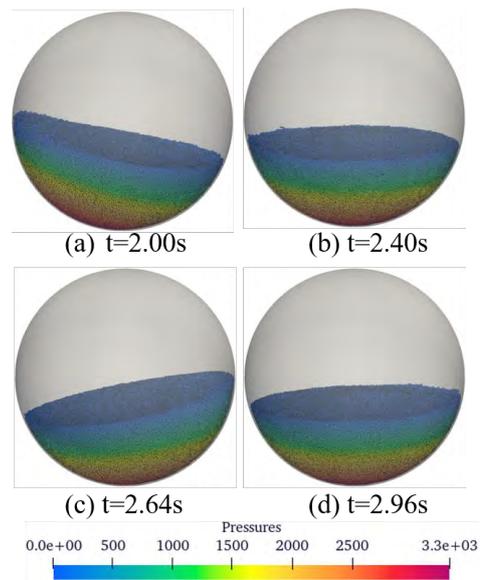
Fig. 8 Liquid distribution of the liquid inside the tank at different times with h/R=0.8
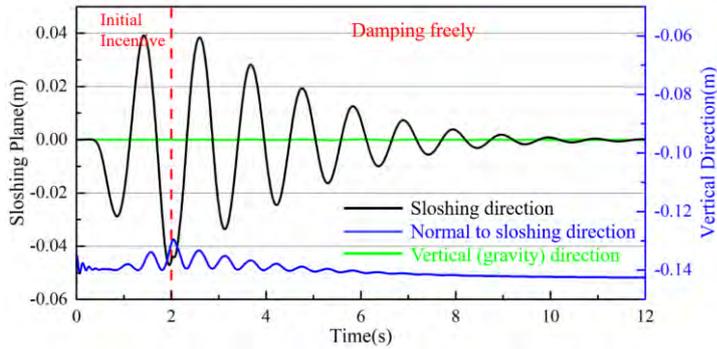
Fig. 9 The mass center coordinates with h/R=1.2
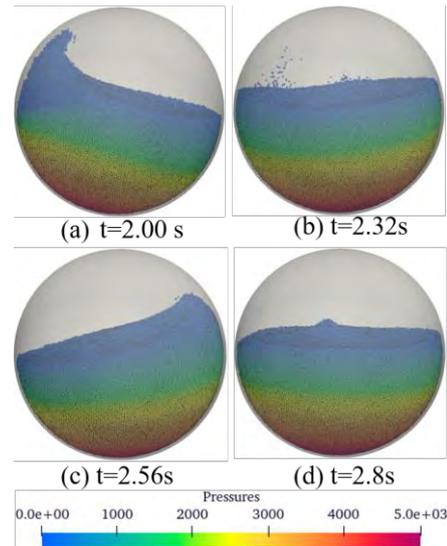
Fig. 10 Liquid distribution of the liquid inside the tank at different times with h/R=1.2

Fig. 7 shows the time history curve of the coordinate when the water depth is h/R=0.8. It can be calculated that the frequency of sloshing at this depth is 5.1206 rad/s, and the damping ratio is 0.0986. Fig. 8 represents the pressure distribution of the liquid at this depth. It can be seen that the liquid sloshing phenomenon also presents a relatively stable state at this water depth. Fig. 9 shows the time history curve of the coordinates when the water depth is h/R=1.2. It can be obtained that the frequency of free attenuation at this depth is 5.776 rad/s, and the damping ratio is 0.0799. Fig. 10 represents the pressure distribution of the liquid at this depth. It can be seen that at this water depth, the sloshing phenomenon of the liquid becomes more intense than before. At this depth, the liquid in the tank climbs upward along the spherical bulkhead during the sloshing, which firstly forms a curling phenomenon, and then the wave surface breaks up, and some droplet splashing phenomena appear in the liquid chamber.

Fig. 11 The mass center coordinates with h/R=1.6

Fig. 12 Liquid distribution of the liquid inside the tank at different times with h/R=1.6

Fig. 13 The mass center coordinates with h/R=1.8

(a) t=2.24s (b) t=2.56s

(c) t=2.88s (d) t=3.2s
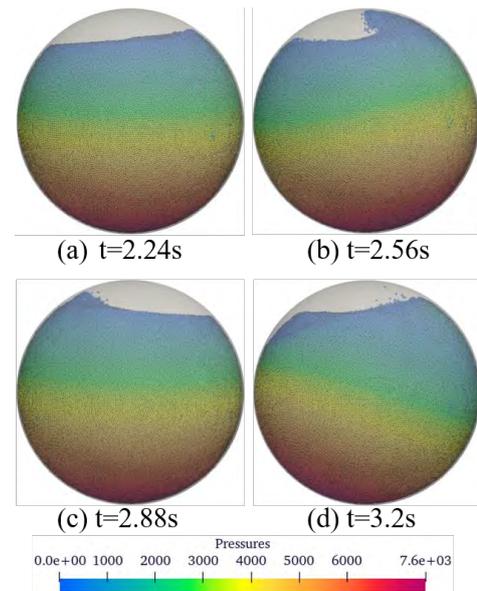
Fig. 14 Liquid distribution of the liquid inside the tank at different times with h/R=1.8

Fig. 11 and Fig. 13 represent the time history curves of the liquid center when the water depth is h/R=1.6 and h/R=1.8. As can be seen from the figures, the sloshing amplitudes of these two cases are much smaller compared to the previous cases. In the first 2s, the large amplitude indicates that the liquid inside the tank is moving together with the tank. When the initial excitation stops, the amplitude of the liquid center of mass curve decreases significantly, indicating that the motion of the liquid center of mass with respect to the tank has been relatively small. It can be seen that the higher filling rate of the liquid chamber can well inhibit the sloshing phenomenon. For the case with h/R=1.6, the sloshing frequency is 7.3513 rad/s, and the damping ratio is 0.0753. For the case with h/R=1.8, the sloshing frequency is 8.3331 rad/s, and the damping ratio is 0.0541. Figures 12 and 14 represent the pressure distribution of the liquid for the two cases h/R=1.6 and h/R=1.8, respectively. It can be seen from the figures that at both sets of depths, an obvious liquid surface curling and breaking occurs, which may because that the angle between the free surface and the bulkhead at this depth is too small, hindering the movement of the liquid climbing upwards along the bulkhead.

In order to further analyze the effect of the water depth on the frequency and the damping coefficient, the frequency and the damping coefficient of the sloshing at different water depth are plotted in a graph for comparison, as shown in Fig. 15. It can be seen that with water depth increasing, the frequency of the liquid sloshing increases, while the damping ratio of the decreases gradually. It can be assumed that increasing water depth brings increasing mass and increasing intrinsic frequency, which are in line with the basic theory in vibration mechanics. The increase in water depth and the decrease in damping coefficient may be due to the fact that the relative motion between the water and the bulkhead is reduced after the water depth is increased, which slows down the damping process of the liquid sloshing.
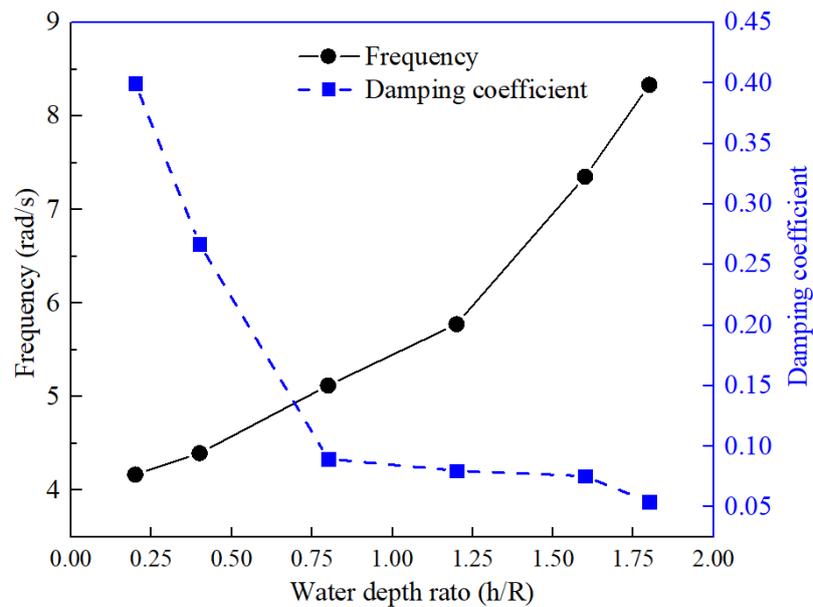
Fig. 15 Comparisons of frequency and damping coefficient at different depths

Conclusions

In this paper, the freely damping sloshing inside a spherical tank is simulated at six different water depths. The nature frequency and the damping coefficients are calculated by the mass center coordinates curves. By comparing the obtained results, it can be found that with the increase of the water depth, the natural frequency of the system gradually increase, while the damping coefficient gradually decreases. Under the same initial excitation, different water depths correspond to different intensities of liquid sloshing. In this paper, for the three cases with h/R>1, the free surface is curled and broken. In contrast, for the three cases with h/R<1, the shape of the free liquid surface remains relatively smooth and stable. This may be due to the fact that when h/R>1, the angle between the free liquid surface and the bulkhead is less than 90°, and the liquid surface curls and breaks under the combined effect of bulkhead pressure and gravity. In the future, the liquid sloshing phenomenon inside the spherical liquid chamber under forced vibration will be further simulated, and the pressure on the bulkhead will be calculated.

Acknowledgements

References

[1] Yang, H. Q., Peupeot, J. (2014) Propellant Sloshing Parameter Extraction from Computational-Fluid-Dynamics Analysis, *Journal of Spacecraft and Rockets*, **51**(3): 908-916.
[2] Gabriele, B., Jose L. (2018) Co-simulation of ship motions and sloshing in tanks, Ocean Engineering, 152(2):353-376.
[3] Cao, X., Tao, L., Zhang, A., Ming, F. (2019) Smoothed particle hydrodynamics (SPH) model for coupled analysis of a damaged ship with internal sloshing in beam seas, *Physics of Fluids*, **31**(3):032103.

[4] Álvaro, R., Ma, H., Tim, H. J. et al (2021) Axisymmetric Ferrofluid Oscillations in a Cylindrical Tank in Microgravity, *Microgravity Science and Technology*, **33**(4).

[5] Mashy, G., Zhou, Y., José, M., et al. (2021) Smooth particle hydrodynamics simulations of long-duration violent three-dimensional sloshing in tanks, *Ocean Engineering*, **229**:108925.

[6] Chen, S., Wu, D., Li, Y., et al. (2022) Profiles of Liquid on the Surface of Revolution with Varying Cross-section under Microgravity, Microgravity Science and Technology, 34(6).

[7] Wang, L., Zhang, X., Yun, Y., Liu, J., Li, W., Huang, B. (2022) Numerical Simulation of Reorientation Process Under Different Conditions in a Vane-type Surface Tension Propellant Tank, *Microgravity Science and Technology*, **34**: 37.

[8] Yang,X., Zhang, Z., Zhang, G., et al. (2022) Simulating multi-phase sloshing flows with the SPH method, *Applied Ocean Research*, **118**:102989.

[9] Zhang Y., Wan, D. (2014) Comparative study of MPS method and level-set method for sloshing flows, *Journal of Hydrodynamics*, **26**(4): 577-585.

[10] Zhang, Y., *Development and Application of 3D Parallel Improved Meshless MPS Method*, PhD Thesis, Shanghai Jiao Tong University, China, 2014.

[11] Zhang, G., Zha, R., Wan, D. (2022) MPS–FEM coupled method for 3D dam-break flows with elastic gate structures, *European Journal of Mechanics / B Fluids*, **94**: 171–189.

[12] Xie, F., Zhao, W., Wan, D. (2021) MPS-DEM coupling method for interaction between fluid and thin elastic structures, *Ocean Engineering*, **236**: 109449.

[13] Chen, X., Wan, D. (2019) GPU accelerated MPS method for large-scale 3-D violent free surface flows, *Ocean Engineering*, **171**: 677–694.

[14] Tang, Z., Zhang, Y., Wan, D. (2016) Multi-resolution MPS method for free surface flows, *International Journal of Computational Methods*, **13**(4): 1641018.

[15] Huang, C., Zhang, G., Zhao, W., Wan, D. (2022) Hydroelastic responses of an elastic cylinder impacting on the free surface by MPS-FEM coupled method, *Acta Mechanica Sinica*, **38**(11): 322057.

[16] Tanaka, M., Masunaga, T. (2010) Stabilization and smoothing of pressure in MPS method by quasi-compressibility, *Journal of Computational Physics*, **229**(11): 4279- 4290.

[17] Lee, B., Park, J., Kim, M., Hwang, S. (2011) Step-by-step improvement of MPS method in simulating violent free-surface motions and impact-loads, *Computer Methods in Applied Mechanics and Engineering*, **200**(9-12): 1113-1125.

[18] Brackbill, J. U., Kothe, D. B., Zemach, C. (1992) A continuum method for modeling surface tension, *Journal of Computational Physics*, **100**(2): 335-354.

[19] Chen, X., Wan, D. (2019) Numerical simulation of liquid sloshing in LNG tank using GPU-accelerated MPS method, *Chinese Journal of Theoretical and Applied Mechanics*, **51**(3): 714-729.

[20] Zhang, G., Zhao, W., Wan, D. (2022) Numerical simulations of sloshing waves in vertically excited square tank by Improved MPS method, *Journal of Hydrodynamics*, **34**(1): 76-84.

[21] Huang, C., Zhao, W., Wan, D. (2023) Numerical simulations of faraday waves in cylindrical and hexagonal tanks based on MPS method, *Journal of Hydrodynamics*, **35**(2): 93-103.

Flow transition and noise analysis of axisymmetric body based on LES model

*Zhiqiang Liu, Yabo Wei, Jianhua Wang, †Decheng Wan

Computational Marine Hydrodynamics Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

*Presenting author: zqliu012345@ sjtu.edu.cn
†Corresponding author: dcwan@sjtu.edu.cn

Abstract

Transition flow is an important phenomenon in hydrodynamics, it affects the flow development and the flow noise. In this paper, an axisymmetric body is selected as a base model. the Wall Model Large Eddy Simulation (WMLES) is used to capture the transition flow and the pressure pulsation on the surface of the axisymmetric body under the high Reynolds number $Re=4.8*10^7$. The vortex structures, pressure pulsation, and other flow parameters are analyzed to study the turbulence transition. According to the results, some mesh generation measures can be summarized to capture the transition flow and pressure pulsation. And in the transition region, the amplitude of pressure pulsation is larger than in the laminar region and turbulent region. In the laminar-turbulent transition region, the vortex structure is complex, which includes laminar vortex and turbulent vortex structures. These studies can provide some references in the future.

Keywords: transition flow; axisymmetric body; high Reynolds number; Large Eddy Simulation; pressure pulsation

Introduction

Transitional flow is an important phenomenon in fluid flow. It involves the laminar flow, transitional flow and turbulent flow. It encompasses flow mechanisms and noise generation Many researches have been done study on the aspect. Yang et al. (1992)use the large eddy simulation to calculate the flow transition flat plate and compared with the experiment, and the agreement with the available experimental results is surprisingly good. It demonstrates that laminar-turbulence flow in boundary layer can be simulated by numerical simulation. Arnal et al. (2000) provide a method to estimate the location of the transition onset and it restricted to subsonic flows (incompressible and transonic). Xiao et al. (2019) provided a description of flat plate transitional flow based on a structural ensemble dynamics (SED) theory of wall turbulence. Manchester et al. (2020) used the LES method calculate transitional flow of the Food and Drug Administration's (FDA) benchmark nozzle model, and comparison of predicted turbulence quantities with experimental measurements at a Reynolds number of 2000. The results show that LES model can accurately capture laminar quantities throughout the model. Bose et al. (2018) reviewed the development of WMLES method in recent years, and assess the validity of the assumption in equation derivation. Ren et al. (2022) use the WMLES, shear stress transport k-w DES (SST-DES), and Spalart–Allmaras DES (SA-DES) methods to simulate the flow over a cylinder and verify the WMLES method. Yalçın (2021) used the in-house high-order solver to calculate the boundary layer transition and cavity flow noise and considering the overset grid technique. Yang (2014) applied high-order spatial and temporal schemes for an unstructured, pressure-based N-S solver to simulate transition flow past airfoil. Shaikh (1997)

used sequences of computer-generated deterministic white noise to excite a laminar boundary layer via a loudspeaker embedded in a flat-plate model. Ye et al. (2019) investigate the surface roughness influence for the boundary layer transition and noise effect.

In this paper, we used the Wall-modeled Large Eddy Simulation (WMLES) to calculate the symmetry-body transitional flow based on OpenFOAM at a high Reynold number $4.8*10^7$. We focus on the turbulence flow and at the same time, we also talk about the transitional flow. Considering the noise effect when the symmetry body sail under the water, we utilize the spectral analysis methods to analyze the noise.

Methodology

For solving incompressible viscous flow in single-phase, here is the large eddy simulation, Eq. (1) and Eq. (2):

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}\left(\overline{u}_i\overline{u}_j\right) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} \tag{1}$$

$$\frac{\partial \overline{u}_j}{\partial x_j} = 0 \tag{2}$$

Where, $\overline{u}_i$ is the velocity and $\overline{p}$ is the pressure, these are obtained by spatial filtering. T_{ij} is the stress tensor and it can be described as follows Eq. (3):

$$\tau_{ij} = 2v\overline{s}_{ij} + \tau_{ij}^{sgs} \tag{3}$$

In the formula, $\overline{S}_{ij}$ is the component of the strain tensor after spatial filtering. τ_{ij}^{sgs} is sub-grid stress tensor, see Eq. (4). Based on Boussinesq assumptions, we can get the sub-grid kinematic viscosity.

$$\tau_{ij}^{sgs} = 2v_{sgs}\overline{s}_{ij} + \frac{1}{3}\tau_{kk}^{sgs}\delta_{ij} \tag{4}$$

In formula, the second term in the right of the equation is usually calculated in pressure term, thus, it doesn't need consideration. The control equation of single-phase large eddy simulation can be described as Eq. (5):

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}\left(\overline{u}_i\overline{u}_j\right) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \frac{\partial\left(2v\overline{s}_{ij} + 2v_{sgs}\overline{S}_{ij}\right)}{\partial x_j} \tag{5}$$

By filtering the turbulent field, WRLES directly solves the turbulent motion larger than a certain scale (filter width), and the motion smaller than this scale is modeled. Filtering in physical space is achieved through integral operations, expressed by the following formula Eq. (6):

$$\iiint_{\Delta/2}^{\Delta/2} u_i(\xi,t)G(x-\xi)d\xi_1\xi_2\xi_3 \tag{6}$$

Where $G(x-\xi)$ is the filter function. Filtering the N-S and continuity equations at the grid scale, we can get the control equation, see Eq. (7):

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial \overline{u_i u_j}}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + v\frac{\partial^2 \overline{u}_i}{\partial x_j \partial x_j} \tag{7}$$

In this formula, the new unknown parameter appear : $\overline{u_i u_j}$. For closing the control equation, we introduce the sublattice model, Wall-adapting local eddy-viscosity model (WALE). Here, $\tau_{ij}^{sgs} = \overline{u}_i\overline{u}_j - \overline{u_i u_j}$, $\tau_{ij} = 2\mu s_{ij} + \tau_{ij}^{sgs}$, τ_{ij}^{sgs} represent the stress tensor. The advantage of WALE is that the value is zero automatically in pure shear flow region show as Eq. (8).

$$\boldsymbol{v_{sgs}} = (C_W \cdot \Delta)^2 \frac{\left(s_{ij}^d s_{ij}^d\right)^{3/2}}{\left(\overline{s}_{ij}\overline{s}_{ij}\right)^{5/2} + \left(s_{ij}^d s_{ij}^d\right)^{5/4}} \tag{8}$$

We use the Wall-modeled LES method to calculate the sub-grid stress. The base logical of Wall-modeled LES is solving the boundary inner layer by model and directly solving the flow by N-S equation for outer boundary layer flow. Compared to WRLES (Wall-resolved Large Eddy Simulation, Wall-resolved LES), WMLES greatly reduces the amount of computation, and improves the computational efficiency.

For obtaining the wall boundary stress, we solving the thin boundary layer equations based on ODE-based model, as Eq. (9), (10), (11).

$$\frac{\partial \tilde{u}_i}{\partial t} + \frac{\partial \tilde{u}_i \tilde{u}_j}{\partial x_j} + \frac{1}{\rho}\frac{\partial \tilde{p}}{\partial x_i} = \frac{\partial}{\partial y}\left[(\nu + \tilde{\nu}_t)\frac{\partial \tilde{u}_i}{\partial y}\right], i = 1,3,5 \tag{9}$$

$$\tilde{u}_i(y_w) = 0, \tilde{u}_i(y^*) = U(y^*) \tag{10}$$

$$\frac{\partial}{\partial x_2}\left[(\nu + \nu_{sgs})\frac{\partial \langle u_i \rangle}{\partial x_2}\right] = F_i \tag{11}$$

$\tilde{u}_i$ is the mean velocity component parallel to the wall (i=1 is the streamwise component, i=3 is the spreading component). $\tilde{\nu}_t$ represents the vortex viscosity of the modeled Reynolds stress in the bottom layer of the boundary layer. y_w represents the wall and y^* represents a certain distance perpendicular to the wall. After obtaining the velocity at the bottom of the near-wall boundary layer, the desired wall stress can be obtained from $\tau_{w,i} = \mu \frac{\partial \tilde{u}_i}{\partial y}$.

The basic idea of WMLES is to directly compute and analyze the flow field in the outer layer of the boundary layer, while the flow field in the inner layer of the boundary layer is modelled by taking the idea of building a model. Compared with WRLES (Wall-resolved Large Eddy Simulation, Wall-resolved LES), WMLES greatly reduces the amount of computation, and it will have a great advantage in computational efficiency when it is applied to the simulation method of high-Reynolds-number gas-liquid two-phase boundary-layer flow based on the modeling method.

Application

In this section, we will introduce the application of WMLES method. We select a symmetry body to calculate the turbulence flow and noise effect in turbulence region. The model is shown as the Figure 1. The main particular parameter is described in Table 1.
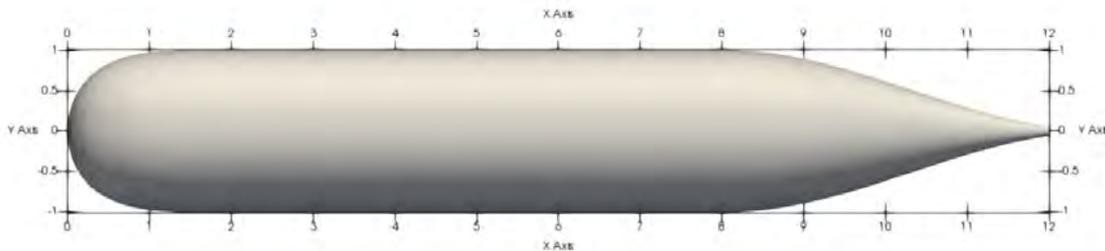
Figure 1. Symmetry-body model

Based on OpenFOAM grid generation tools called snappyHexMeshDict, we get the computational grid. The number of boundary layer is 30, the first boundary layer thickness is $8.04*e^{-4}$. For WMLES, because it solves the boundary layer flow by solving boundary layer

equation by wall model. The expansion rate is 1.06. the mesh is 32.60 M. The y-Plus is 105. The mesh is shown in Figure 2.

Table 1. Main particular parameters of symmetry-body model

parameters	Value/m
Length	12
diameter	2
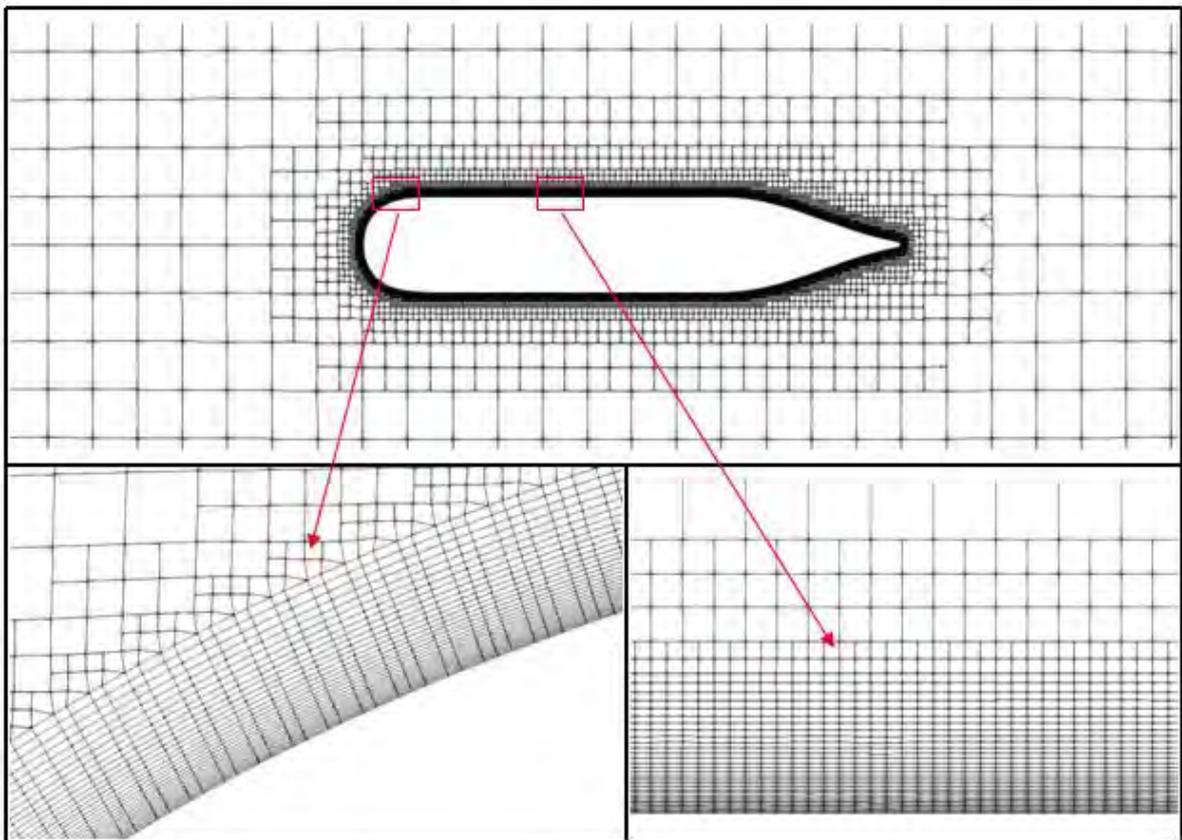
Figure 2. Mesh generation and boundary layer mesh

For get precise solution, the RANS model is used to estimate the boundary layer thickness firstly. According to misread the boundary layer thickness, we ensure the more than 30 layers in the boundary layer for calculation precise.

For analysis the high frequency of noise, the delta T is set as $1*10^{-4}$ and control the Courant number <1. Figure 3 and Figure 4 is the mean velocity field and mean pressure field. From the velocity and pressure field, no significant difference from the actual. Figure 5 is the vortex field, the evolution of the vortex structure represents a turning point in the flow.
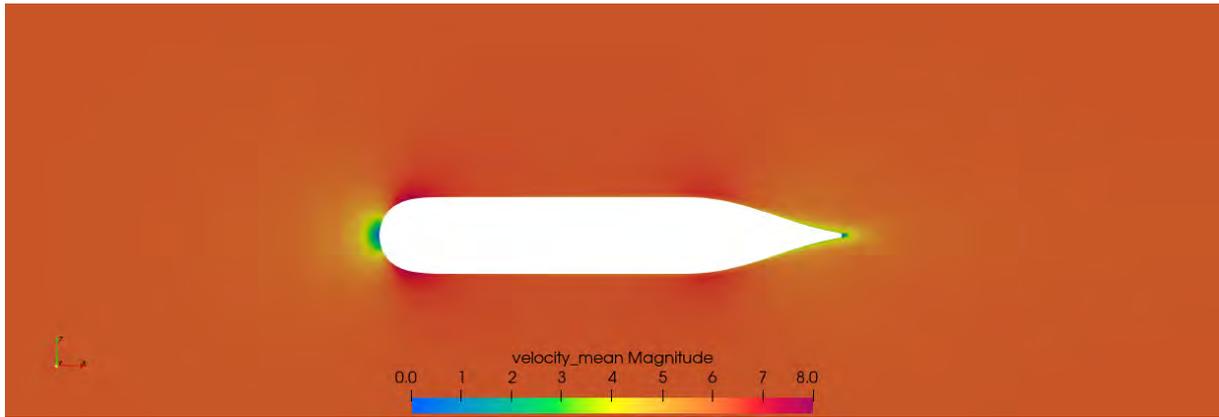
Figure 3. Velocity mean value
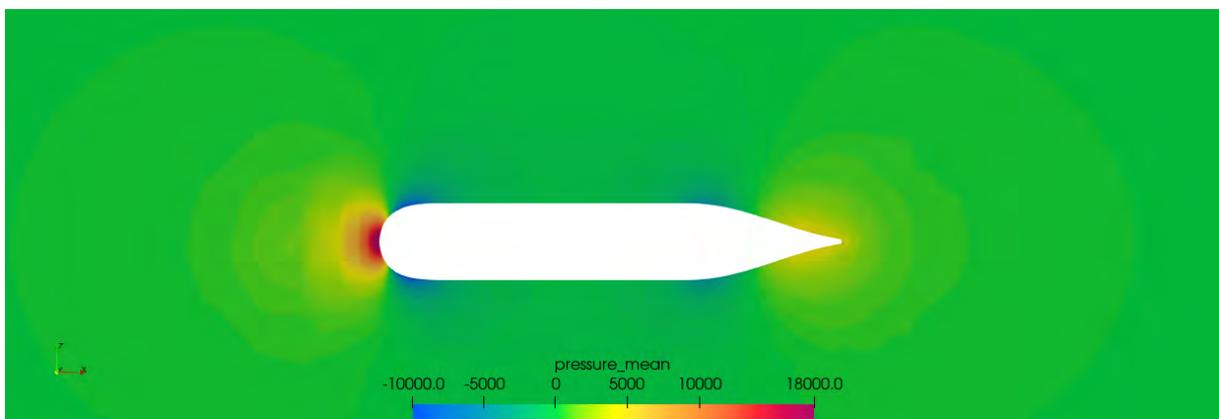
Figure 4. Pressure mean value

Figure 5. the vortex structure (Q=3)

For study the transition flow and noise quantitatively, we set a series of probes to monitor the pressure data in turbulence flow. The probes distribution is shown in Figure 5. 200 measurement points were placed within a distance of 8 meters from the head to the parallel midbody.
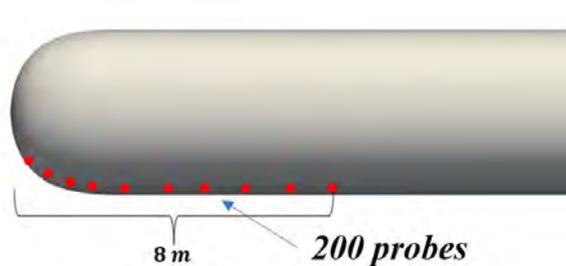
Figure 6. Probes distribution on the symmetry-body surface

Figure 7 is the mean pressure along the flow direction in 8 meters. Responding to Figure 4, the pressure changes from high to low along the flow direction in the bow. The Figure 8 is the pressure root-mean-square change along the symmetry-body. The pressure variance maps out the sharpness of the pressure fluctuations. The occurrence of transition flow and the length of the transition zone can be seen in the change in pressure variance.
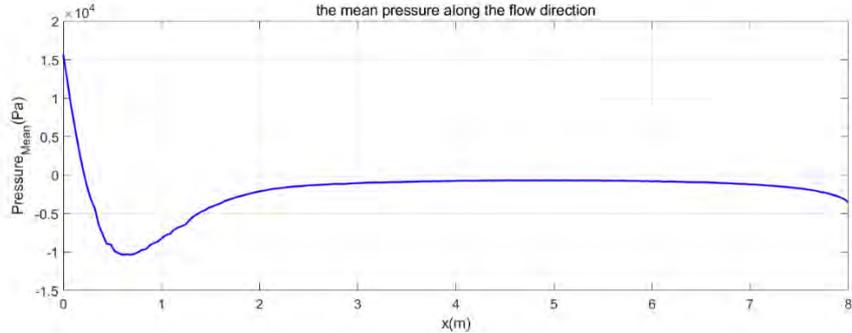
Figure 7. The mean pressure along the flow direction
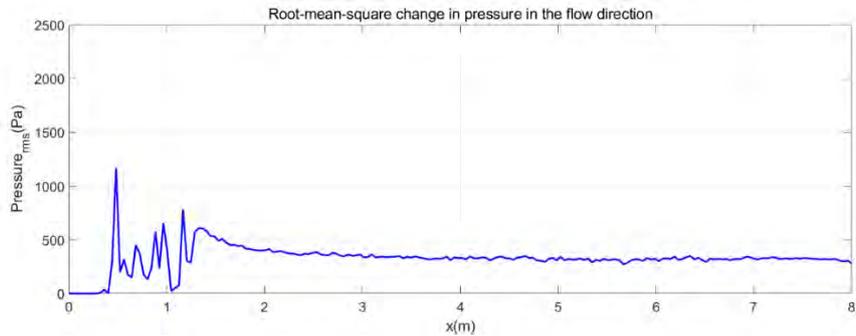
Figure 8. Root-mean-square change in pressure along the flow direction

According to the root-mean-square change, the transition position is 0.4m in x-axis. And the length of transition region is about 0.9m. the result indicate the pressure pulsation in transition region is larger than turbulence region, which is agreement with other studies (Abarbanel et al., 1994; Arakeri, 1975).
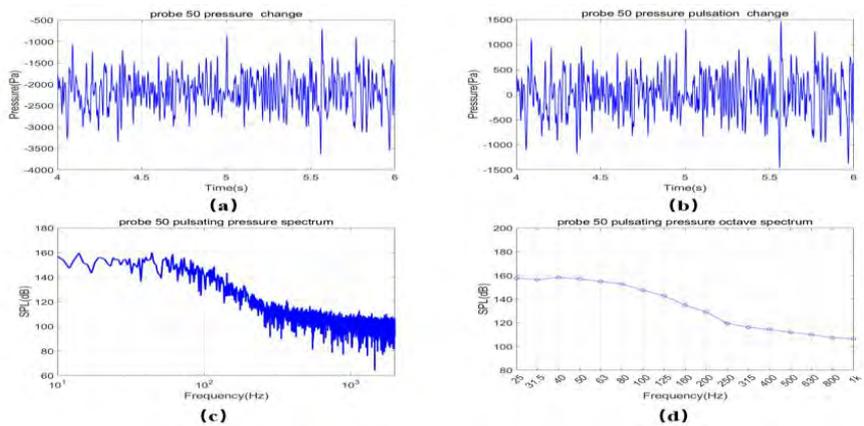
Figure 9. Probe 50 (x=2.04 m) data

((a) pressure change vary time; (b) pressure pulsation change vary time; (c) pulsating pressure spectrum; (d) pulsating pressure octave spectrum;))
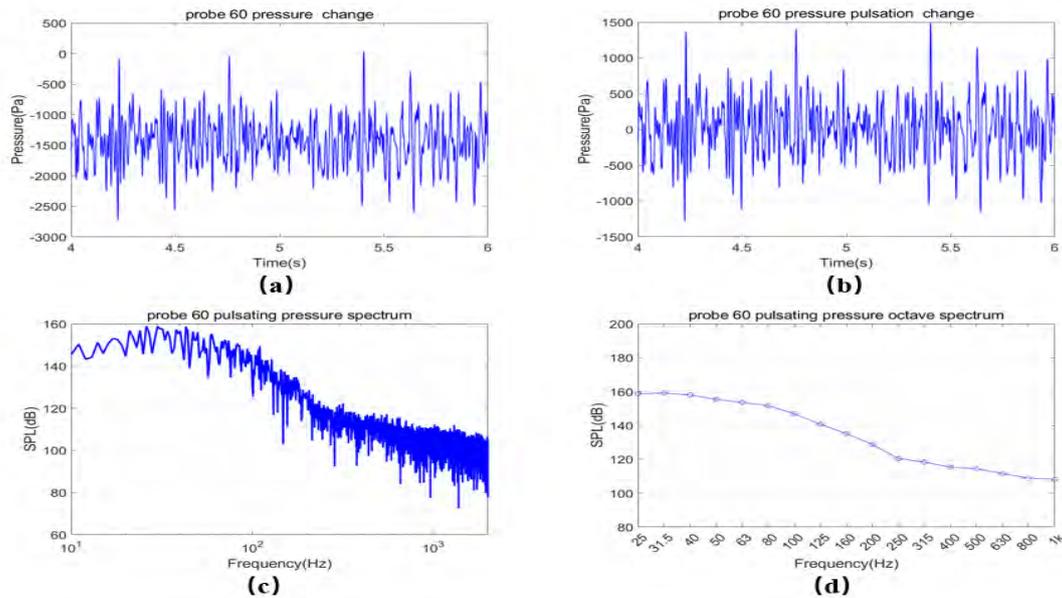
Figure 10. Probe 60 (x=2.04 m) data

((a) pressure change vary time; (b) pressure pulsation change vary time; (c) pulsating pressure spectrum; (d) pulsating pressure octave spectrum;))

Figure 9 and Figure 10 is pressure data and noise analysis on the probe 50 and probe 60. From the pressure pulsation, we can observer the pressure high frequency vibration in turbulent flow. The pressure pulsation data were post-processed using spectral analysis, and the power spectral density function was obtained using the Fourier transform. Based on Figure 9 and Figure 10, the noise is above 140 dB in low frequency. These results agree with other studies. Based on the noise analysis, we can design the low-noise axis-symmetric body form to avoid the noise influence.

Conclusions

In this paper, we use the WMLES to simulate the flow over a axis-symmetry body. We focus on the transition flow and noise analysis. Compare with other LES models, WMLES model does not have to strictly satisfy the requirement that y-plus is less than 1. According to the pressure root-mean-square value change in bow, compared with turbulence development region fully, the pressure pulsating is larger in transition region, which is agreement with other studies. It indicates the WMLES has the capacity to calculate the boundary layer flow and transition flow. Utilizing the Fourier transform and spectral analysis, we get the sound pressure level (SPL) and one-third octave spectrum. Based on the WMLES and spectral analysis, we can design the low-noise symmetry body form in the future.

Acknowledgements

References

Abarbanel, H. D. I., Katz, R. A., Galib, T., Cembrola, J., & Frison, T. W. (1994). Nonlinear analysis of high-Reynolds-number flows over a buoyant axisymmetric body. *Physical Review E, 49*(5), 4003-4018. doi:10.1103/PhysRevE.49.4003

Arakeri, V. H. (1975). A Note on the Transition Observations on an Axisymmetric Body and Some Related Fluctuating Wall Pressure Measurements. *Journal of Fluids Engineering, 97*(1), 82-86. doi:10.1115/1.3447222

Arnal, D., & Casalis, G. (2000). Laminar-turbulent transition prediction in three-dimensional flows. *Progress in Aerospace Sciences, 36*(2), 173-191. doi:Doi 10.1016/S0376-0421(00)00002-6

Bose, S. T., & Park, G. I. (2018). Wall-modeled large-eddy simulation for complex turbulent flows. *Annual review of fluid mechanics, 50*, 535-561.

Manchester, E. L., & Xu, X. Y. (2020). The effect of turbulence on transitional flow in the FDA's benchmark nozzle model using large-eddy simulation. *International Journal for Numerical Methods in Biomedical Engineering, 36*(10), e3389.

Ren, X., Su, H., Yu, H.-H., & Yan, Z. (2022). Wall-modeled large eddy simulation and detached eddy simulation of wall-mounted separated flow via OpenFOAM. *Aerospace, 9*(12), 759.

Shaikh, F. N. (1997). Investigation of transition to turbulence using white-noise excitation and local analysis techniques. *Journal of Fluid Mechanics, 348*, 29-83. doi:10.1017/S0022112097006629

Xiao, M.-J., & She, Z.-S. (2019). Symmetry-based description of laminar-turbulent transition. *Science China Physics, Mechanics & Astronomy, 62*(9), 994711. doi:10.1007/s11433-018-9377-0

Yalçın, Ö. (2021). DELAYED DETACHED-EDDY SIMULATION BASED PREDICTIONS OF BOUNDARY LAYER TRANSITION AND CAVITY FLOW NOISE.

Yang, H. Q. (2014). *Simulation of Tonal Noise Generated by Transition Flow Past Airfoil Using High-Order Schemes on Unstructured Grids.* Paper presented at the 20th AIAA/CEAS Aeroacoustics Conference.

Yang, Z., & Voke, P. R. (1992). Numerical simulation of boundary layer transition in the presence of free stream turbulence. In (pp. 398-402).

Ye, Q., Avallone, F., Ragni, D., Choudhari, M. M., & Casalino, D. (2019). *Effect of surface roughness on boundary layer transition and far field noise.* Paper presented at the 25th AIAA/CEAS Aeroacoustics Conference.

Numerical study of pressure fluctuation and DMD analysis for SUBOFF model

*Lianjie Yu, Yuan Zhuang, †Decheng Wan

Computational Marine Hydrodynamic Lab (CMHL), School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

*Presenting author: leo_yates@163.com
†Corresponding author: dcwan@sjtu.edu.cn

Abstract

The prediction of underwater noise requires accurate simulation of flow field fluctuations. The paper focuses on the SUBOFF submarine model and analyzes fluctuating pressure. Dynamic Mode Decomposition (DMD) is used to study the correlation of pressure fluctuation and flow structures. Large eddy simulation (LES) method is employed to improve accuracy. Probes and arrays are set on the submarine's surface, allowing for analysis of linear spectrum, octave spectrum, and wavenumber-frequency spectrum. The results show that the resistance force can be predicted accurately with LES method, exhibiting an error of less than 1% compared to experiments. Fluctuating pressure within the submarine varies significantly across different zones. Convective velocity varies across different regions, decreasing from upstream to downstream. DMD results reveal that low-mode vortex structures dominate the flow field, characterized by larger scales. Conversely, high-mode structures exhibit smaller scales and lower energy ratios.

Keywords: SUBOFF, LES, pressure fluctuation, DMD.

Introduction

With the rapid reduction of the range for passive sonar, active sonar has been widely used, making the impact of self noise performance prominent. Under the same conditions, the sonar detection distance increases exponentially once the self noise level decreases by 10dB[1]. Hydrodynamic noise is found the main component of the self noise of underwater vehicles. It is proportional to the 5-6th power of speed approximately. At speeds above 10kn, hydrodynamic noise becomes the main component of sonar self noise[2].

The pressure fluctuation in the turbulent boundary layer (TBL) is the most important component of hydrodynamic noise. Early measurements found that very strong fluctuating pressure can be received when the hydrophone is placed in the turbulent area. However, once the sensor leaves the turbulent area, the fluctuating pressure decays quickly, and the decay speed is greater than r^{-1} that is satisfied by the spherical sound field. The spectrum of fluctuating pressure inside and outside the turbulent zone are different significantly. The low-frequency component is strong inside the turbulent zone, while the high-frequency component is difficult to detect. Both the low and high frequency components outside the turbulent zone can be detected. When a dome is added outside the sensor, the fluctuating pressure is reduced greatly.

The TBL fluctuating pressure is of importance in the underwater noise problem. It is one of the main sources of self noise[3], such as torpedo homing array, towed linear array sonar, side array sonar, etc. Here, the TBL fluctuating pressure produces flow noise directly. It excites the walls of the carrier to vibrate. On the one hand, it radiates noise to the surrounding medium and

generates cabin noise inside. In this case, the pressure fluctuation acts as a random excitation for the vibration of plates and shells, which radiate noise in return.

Experiments show that different physical quantities play a major role inside and outside the turbulent zone. The fluctuating pressure measured in the turbulent zone is mainly acting on the sensor directly, with a large amplitude and low frequency. It is a local quantity that does not propagate outward and decays quickly when leaving the turbulent zone. What is measured outside the turbulent zone is the sound pressure radiated and propagated from within the TBL. The pressure fluctuation that only exist within the turbulent zone and cannot propagate outside is known as "pseudo sound"[4].

The difference between the two is as follows: (1) In the phenomenal perspective, radiated noise exists both in the turbulent and non-turbulent regions, while fluctuating pressure only exists in the turbulent zone. There must be a transition region between the two, where pressure fluctuation is of the same order as sound pressure[5]. As the hydrophone moves away from the turbulent region, the sound pressure decays according to the usual r^{-1} law, while fluctuating pressure decreases to zero at a faster rate r^{-n} (n>1). (2) In the physical mechanism perspective, the fluctuating pressure is caused by the turbulent velocity fluctuations and is an incompressible component of the pressure field. While the sound pressure is caused by density fluctuations generated by Reynolds stress, which is a compressible component of the pressure field. (3) In the mathematical perspective, the sound pressure obeys the wave equation, while the fluctuating pressure obeys the Poisson equation.

In general, wavenumber-frequency spectrum ($k - \omega$ spectrum) is used to describe the spatial-temporal correlation characteristics of fluctuating pressure in TBL. In experiments, linear arrays are used to collect time history data, and wavenumber-frequency spectrum is obtained based on this matrix transformation. It can be regarded as the spatial Fourier transform of cross spectral density. The frequency result obtained by one hydrophone is equivalent to spatial filtering. At present, there are some semi-theoretical $k - \omega$ spectrum formula obtained by measurements. For example, Corcos model is applicable to the rigid plate TBL. Chase model can reflect the convective characteristics better. Smol'yakov Tkachenko model performs better in the low-wavenumber region. Although these models have been validated in the given conditions such as flat plates, there is no model for the fluctuating pressure in the transition zone, especially for submarine bodies.

Using DMD to analyze flow structures is a popular post processing method at present. Yen [6] studied the flow around spheres with different radii in the pipe, and used DMD to explore the flow structures under different Reynolds numbers. According to the corresponding peak frequencies of three St values, the DMD snapshot sampling frequency is set to 5 times the maximum St frequency to ensure that the dominant mode can be captured. The velocity iso-surface is displayed, and it is found that the structure scale of the low-frequency peak is the largest, and the scale of the K-H instability is the smallest.

The structure of this paper is arranged as follows: the calculation methods involved is introduced briefly in the second chapter. The numerical settings and the SUBOFF resistance experiment verification in introduced in the third chapter. In the fourth chapter, the pressure fluctuation is analyzed in linear spectrum, octave spectrum and wavenumber-frequency spectrum, and the correlation between vorticity field and fluctuating pressure is studied using DMD method. Finally, the conclusion is given in chapter 5.

Mathematical Foundation

Large Eddy Simulation

In order to solve the incompressible unsteady viscous fluid problem, the single-phase flow large eddy simulation (LES) equation is introduced:

$$\frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}\left(\overline{u}_i \overline{u}_j\right) = -\frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \frac{\partial \tau_{ij}}{\partial x_j} \tag{1}$$

Among them, $\overline{u}_i$ is the velocity after spatial filtering, and $\overline{p}$ is the pressure after spatial filtering. τ_{ij} is a stress tensor, which can be further expressed in the following form:

$$\tau_{ij} = 2\nu \overline{s}_{ij} + \tau_{ij}^{sgs} \tag{2}$$

Among them, $\overline{S}_{ij}$ is the strain tensor component after spatial filtering. τ_{ij}^{sgs} is the subgrid stress tensor, which is based on the Boussinesq assumption to establish a subgrid model and provide subgrid kinematic viscosity. By imitating the form of a viscous stress tensor, it is associated with the strain tensor:

$$\tau_{ij}^{sgs} = 2\nu_{sgs}\overline{s}_{ij} + \frac{1}{3}\tau_{kk}^{sgs}\delta_{ij} \tag{3}$$

Pressure Spectrum analysis methods

The pressure fluctuation dealt with in the turbulent boundary layer are generally random signals that cannot be expressed in clear mathematical terms and can only be analyzed by statistical methods based on the theory of stochastic processes. The steady random signal itself is still uncertain, but its correlation function is deterministic. The Fourier transform of the correlation function can be expressed as the power spectral density (PSD) function of the random signal at zero mean, and the PSD feature characterizes the magnitude of the energy value of the random signal at each frequency point. The PSD can only be estimated for a random process with a finite number of sample points, and different spectral estimation methods are used to obtain different PSD characteristics.

In this paper, the Welch modified periodogram method is used, which applies the window function $w(n)$ to the signal segments before calculating their respective periodograms. If the signal is divided into L segments, the estimation of the Welch method for each segment is

$$I_M^l(f) = \frac{1}{UM}\left|\sum_{n=0}^{M-1} x_M^l(n)w(n)e^{-j\omega T}\right|, 1 \leq l \leq L, L \leq \frac{N}{M} \tag{4}$$

Welch method reduces the interference of different spectral peaks in the real PSD by adding windows (non-rectangular windows) to each segment of data, and truncates the data so that the variance of the estimation results becomes $1/L$ of the original one. When there is some overlap between segments, the variance of the estimation results can be further reduced by increasing the number of segments as long as the overlap between different segments is not too large.

Wavenumber-frequency spectrum

The wavenumber-frequency spectrum of turbulent fluctuation pressure is defined as the two-dimensional Fourier transform of the correlation function of the fluctuation pressure in time and space. The mathematical expression is:

$$\phi(\vec{k}, \omega) = \int_{-\infty}^{+\infty} R(\vec{\xi}, \tau) e^{-(\vec{k} \cdot \vec{\xi} + \omega \cdot \tau)} d\vec{\xi} d\tau \tag{5}$$

In experiments, the hydrophone array with linear arrangement is usually used to collect time history data, and the wavenumber frequency spectrum is obtained based on this matrix transformation. Its discrete form expression is:

$$\phi_{pp}(k, \omega) = \frac{\left| \sum_{n=1}^{N} \sum_{m=1}^{M} W(x_m, t_n) p_m(x_m, t_n) e^{-j(k_x x_m + \omega t_n)} \Delta x \Delta t \right|}{(2\pi)^2 NMC_w} \tag{6}$$

Among them, p_m represents the pulsating pressure measured at the m-th measuring point, N is the number of time samplings, M is the number of measuring points, Δx is the spacing between measuring points, and Δt is the time step size.

DMD analysis

The basic idea of DMD analysis is linear transformation. For example, a signal U_{tx} is defined, x is the value of each point on a one-dimensional signal, and t is the change of the whole signal under change with time. The time sampling points are N and the space sampling points are m. The purpose of DMD is to decompose the system for linear transformation. The kth time step signal U_{xk} can be decomposed by DMD into

$$U_{x,k} = \sum_{i=1}^{m} b_i \cdot \lambda_i^k \cdot \phi_i \tag{7}$$

where b is the initial value, λ is the characteristic root of matrix A, and ϕ is the corresponding mode.
In the DMD method, Singular Value Decomposition (SVD) is used to implement the linear transformation. It is not expanded here.

Numerical setup

Geometric model

The calculation is carried out using a published SUBOFF submarine model with a length of 4.365 m, a maximum diameter of 0.508 m, an incoming velocity of 3.051 m/s and a Reynolds number of 1.584×10^7 in terms of boat length. OpenFOAM, an open source software, was used for the calculation, and the noise calculation module was developed independently based on this software. A cylindrical computational domain was chosen, with the circumferential cylinder using a boundary type of symmetry, for modelling infinitely deep waters. The submarine model was placed far enough away from the inlet and outlet surfaces to avoid the boundary having an effect on the flow.
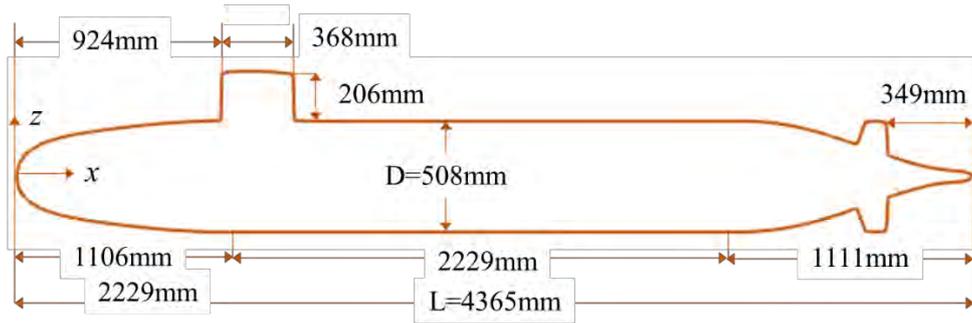
Figure 1. The geometry model of SUBOFF submarine

Grids and probe settings

An unstructured mesh with local refinement near the fin is used. The total number of grids is approximately 12.64 million and the minimum value of y+ is 0.89, which satisfies the boundary layer capture requirement.
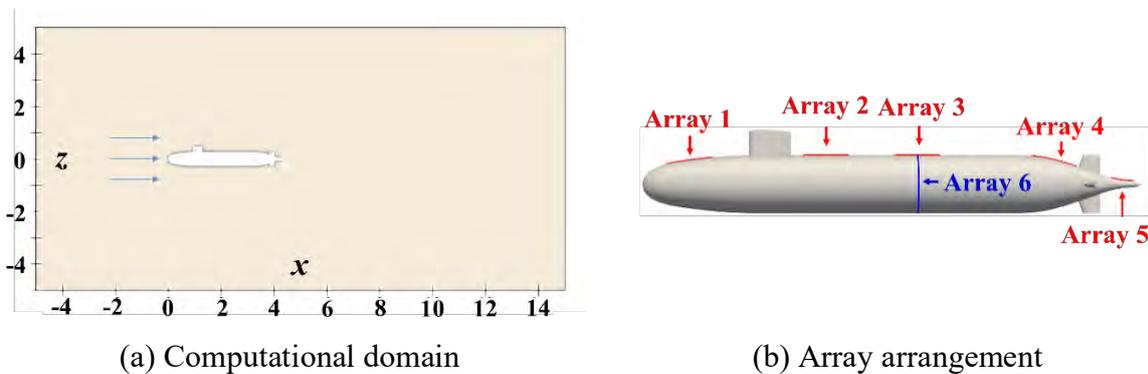
(a) Computational domain   (b) Array arrangement

Figure 2. Computation domain and probe settings

To investigate the pulsating pressure characteristics on the surface of the submarine, a series of arrays consisting of pressure fluctuation measurement points and wave number frequency spectra have been established. Figure 2 illustrates the arrangement, with linear arrays numbered 1-5 positioned evenly along the submarine's backbone in the direction of flow. Array 1 spans an x-direction range of 0.2-0.6m, array 2 covers 1.4-1.8m, array 3 spans 2.2-2.6m, array 4 ranges from 3.4-3.8m, and array 5 spans 4.1-4.3m. The spacing in each linear arrays is 3mm. Additionally, a circular array labeled as 'Array 6' comprises 185 measurement points distributed evenly on the starboard half of the submarine at x=2.5m. Moreover, 1510 pressure fluctuation measurement points positioned at 3mm intervals along the submarine's backbone were utilized for linear and octave spectral analysis. These measurement points cover various zones including laminar, transition, turbulent, and separation areas.

Computational validation

After the calculation is stable, the total resistance of the submarine is obtained as 103.1N, which is close to the experiment result of 102.3N[7], with an error of less than 1%, as Figure 3 shows.
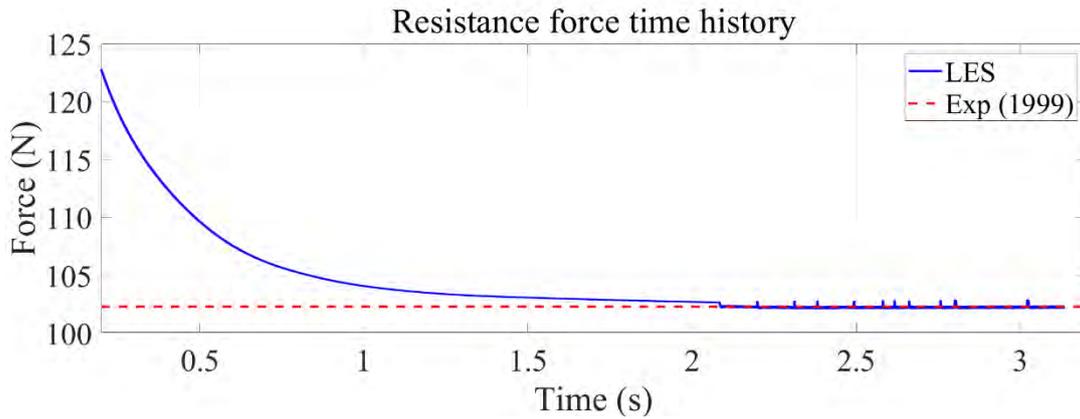
Figure 3. The time history of resistance compared with experiment results[7]

Results and discussion

Pressure fluctuation

In addition to the time-averaged quantities, the pressure fluctuation quantities are also validated in this paper. This is because the accuracy of the hydrodynamic noise forecasts depends to a large extent on the accuracy of the pressure fluctuation forecasts.

Here, the wall pressure measurement points are set on the upper edge line of the submarine, at different distance positions from the head, for the calculation of the power spectral density (PSD). The number value of the probe multiplied by 0.003m is the distance from the head to the measurement point. PSD is calculated using the welch correction periodogram method with an overlap value of 50%, which ensures that the variance of the results is minimized. The linear spectrum is divided into bands according to the 1/3 power of 2 and the octave spectrum density is obtained by integration in each bands, as Figure 4 shows.
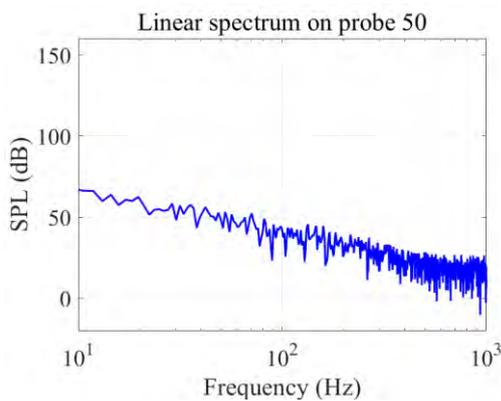
(a) Linear spectrum of laminar zone   (b) Octave spectrum of laminar zone
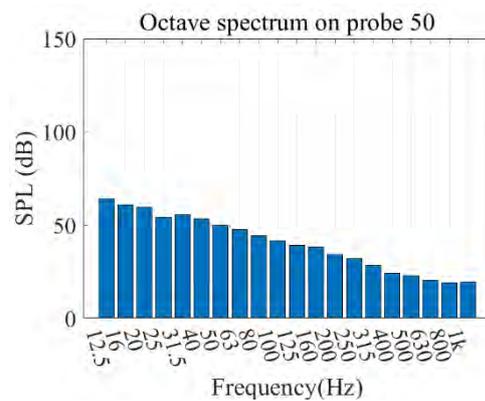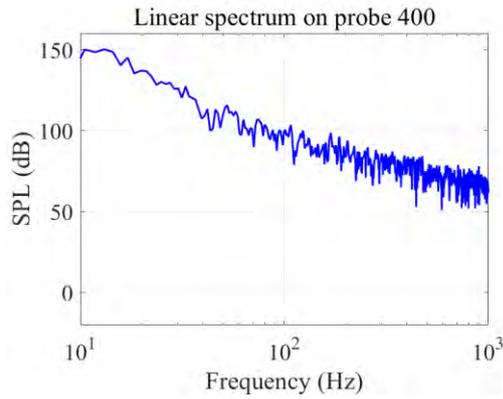
(c) Linear spectrum of sail zone
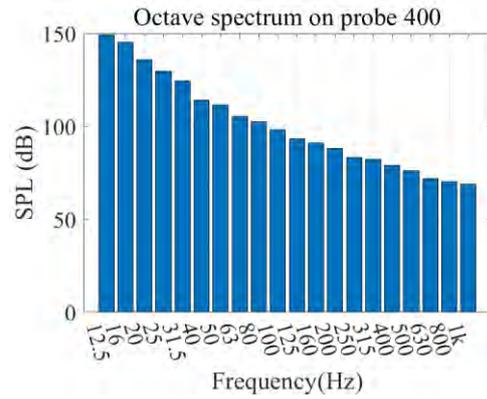
(d) Octave spectrum of sail zone
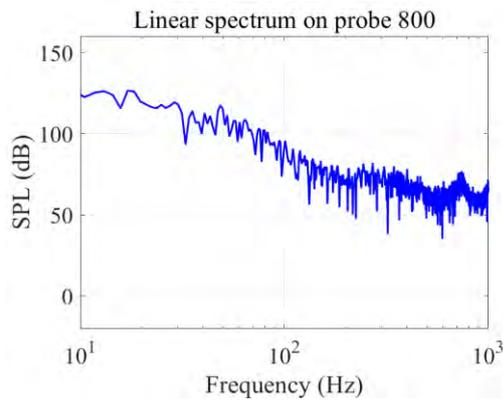
(e) Linear spectrum of turbulence zone
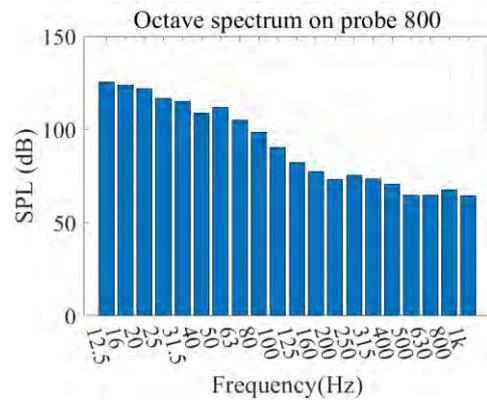
(f) Octave spectrum of turbulence zone
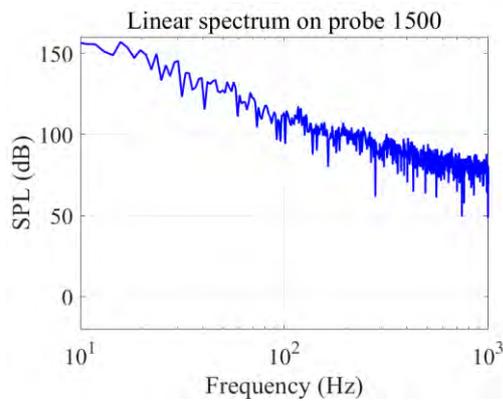
(g) Linear spectrum of shedding zone
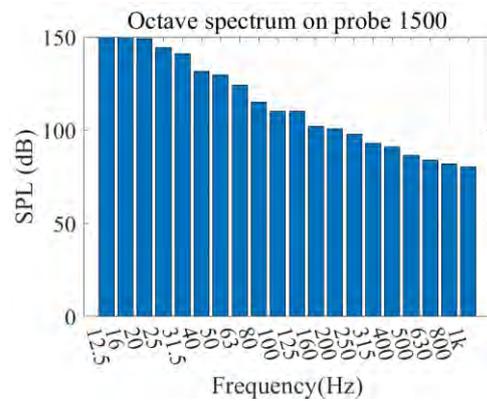
(h) Octave spectrum of shedding zone

Figure 4. The pressure fluctuation in different zones

Observing the pressure fluctuation results in the four different zones, it can be found that the smallest values are found in the laminar zone, the larger values in the turbulence zone and the largest values in the sail zone and vortex shedding zone. This is in line with our experience. The laminar zone fluctuates weakly. Therefore, the pressure fluctuation is small. While in the sail and vortex shedding regions, there is flow separation and the pressure fluctuation is large. In addition, it can be seen that there is a "plateau area" for the fluctuated pressure in the turbulence region, but not in the other regions, which may be due to the unique randomness of the turbulent boundary layer.

Wavenumber-frequency spectrum

According to the array arrangement in Figure 2(b), the sampling process is performed from 2s to 2.5s period with a sampling frequency of 2 kHz. The results are converted into wavenumber-frequency spectrum as shown in Figure 5.
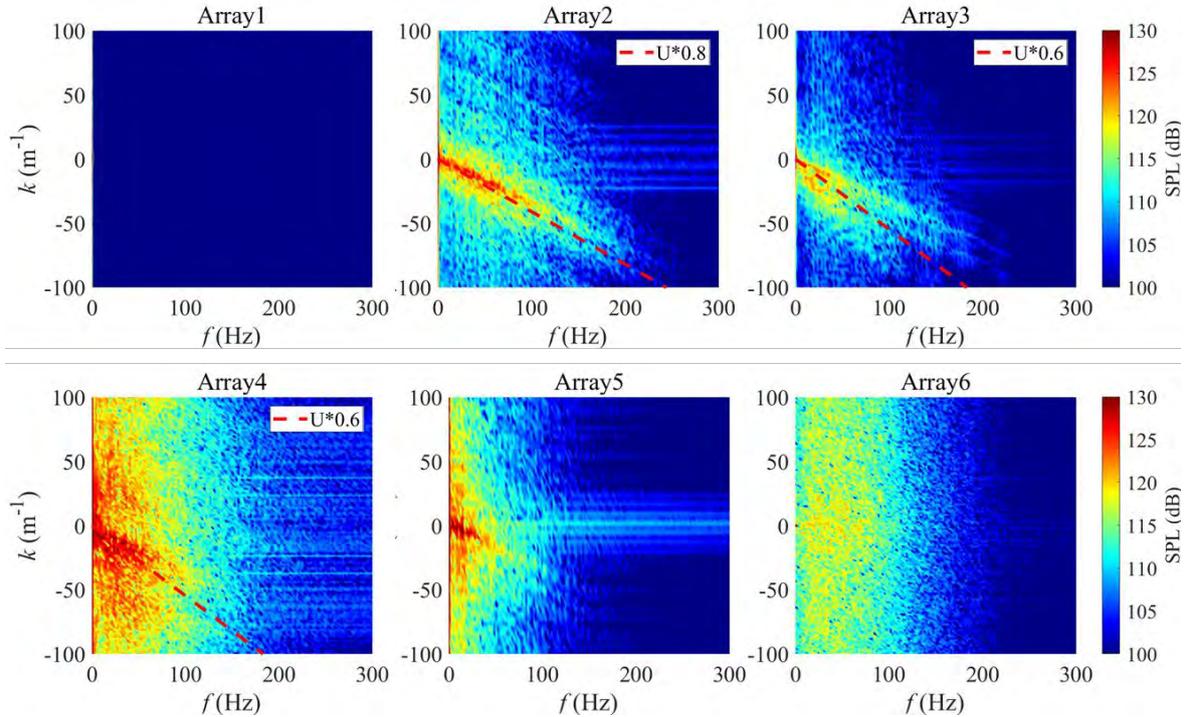
Figure 5. The wavenumber-frequency spectrum in different zones

Since there is no significant convection in the laminar flow region, Array1 has no significant convective ridges. In contrast, the later arrays Array 2-4 all exhibit significant convective ridges. According to the experimental results, the convective velocity is about 0.6-0.8 times of the incoming flow velocity. It can be seen from Figure 5 that from Array2-4, the convective velocity decreases gradually (from 0.8*U to 0.6*U) as the location moves downstream. Array 5 is in the vortex shedding zone, so the convetive phenomenon is not prominent. Array 6 is a circumferential array with no obvious pattern, indicating that the flow is fluctuating randomly in the circumferential direction.

DMD analysis

It is well known that the submarine has two main vortex systems, which are the sail vortex system and the wake vortex system, as is shown in Figure 6. Here, we choose the sail vortex shedding system on *y*=0 slice for DMD analysis.

After the flow becomes stable, 100 moments are selected at equal intervals for vortex snapshot sampling in the 0.1s time range. This ensures full coverage of the frequency range from 10 Hz to 1 kHz for DMD analysis. The vorticity is related to the acoustic pressure, and the acoustic pressure is higher in energy at low frequencies. Therefore, the focus is on the flow structure in low-frequency modes.
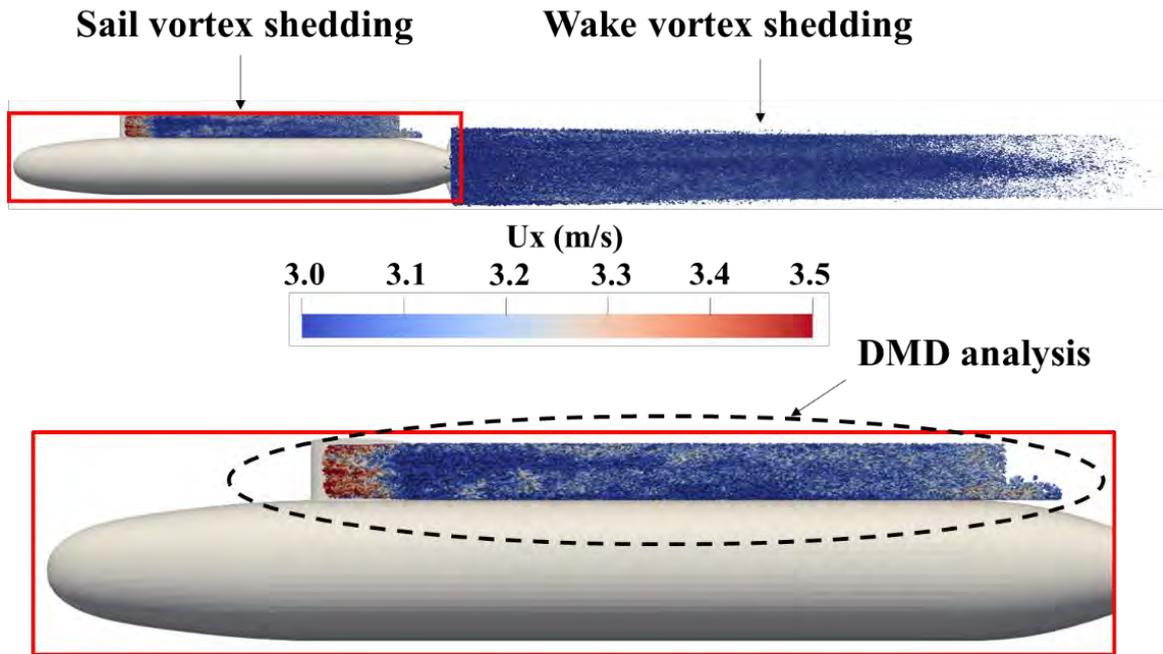
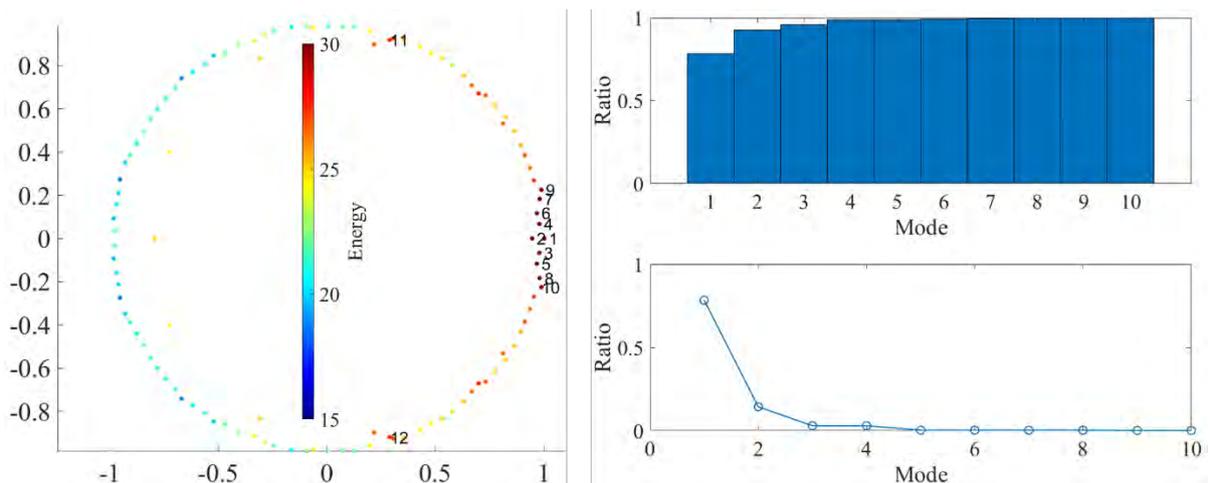**Figure 6. The vortex shedding systems around SUBOFF submarine**

Figure 7. The characteristic root distribution (left) and the energy ratio of the first 10 modes (right)

Figure 7 (left) shows the distribution of each eigenvalue. The colors represent the energy magnitude, and the first 12 order eigenvalues are marked. It is easy to see that they have the highest energy. Figure 7 (right) also illustrates that the first 10 modes occupy almost 100% of the total energy of the vorticity.

From Figure 8, it can be seen that the vortex structure scale is longer in low modes. The scale becomes smaller gradually as the mode number increases. In addition, it can be found that the Q values are larger in low modes and smaller in high modes. It indicates that the low-mode energy dominates the flow and determines the frequency distribution of the pressure fluctuation and hydrodynamic noise.
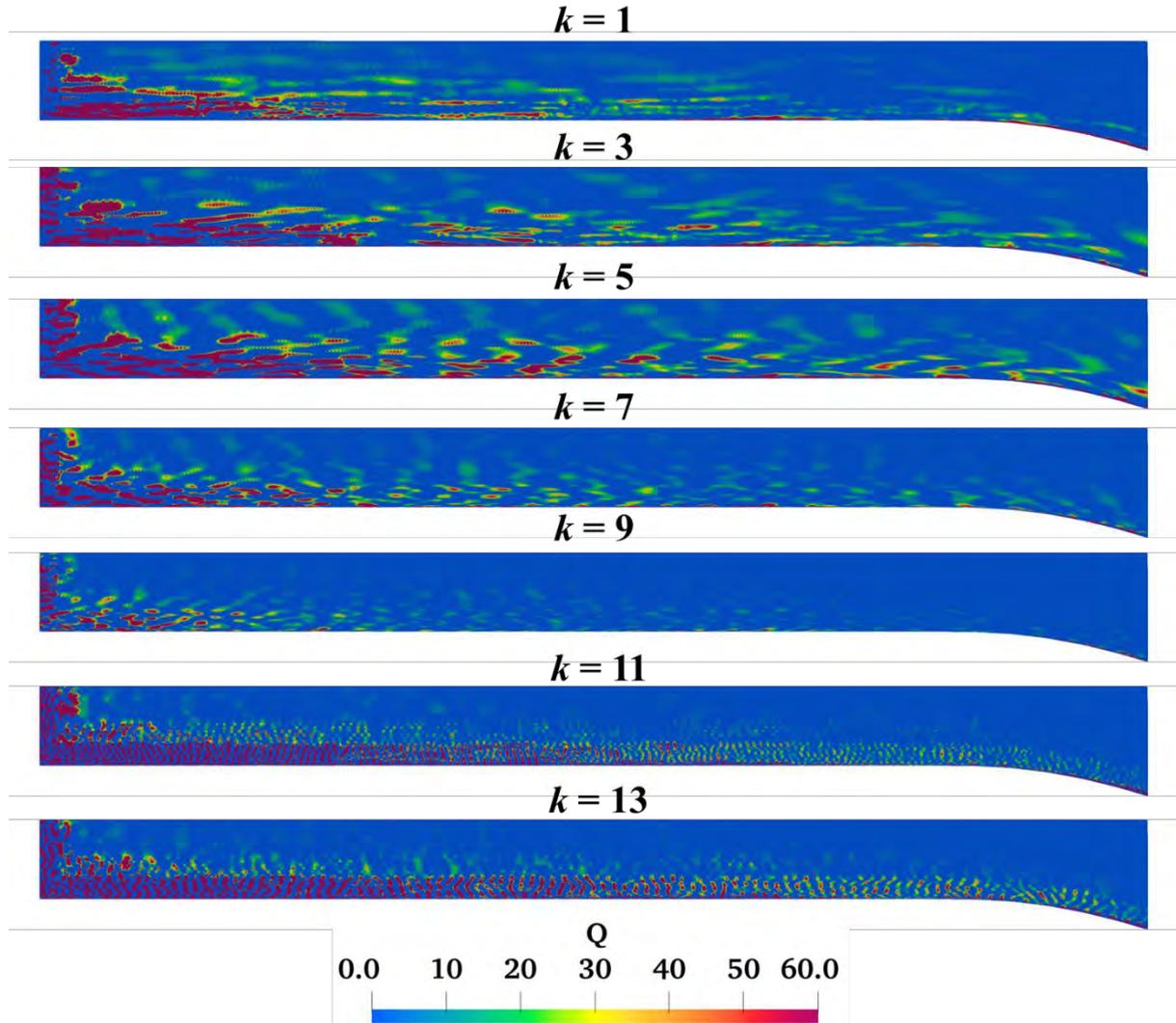
Figure 8. The Q cloud map in different modes

Conclusions

For SUBOFF submarine model, the refined simulation of the flow field and the analysis of the fluctuating pressure are carried out in this paper, serving the hydrodynamic noise prediction. The open source software OpenFOAM is used to perform the flow field simulation, and LES method is used to improve the accuracy. Several probes and arrays are set on the surface of the submarine, and the linear spectrum, octave spectrum and wavenumber-frequency spectrum are analyzed. Finally, DMD method is used to analyze the relationship between the vortex structures downstream of sail and the pressure fluctuation in frequency. The conclusions are as follows:

(1) The resistance force of SOBOFF can be predicted accurately by LES method, and the error is less than 1% compared with experiment, which lays the foundation for fluctuating pressure analysis.

(2) The frequency distribution of fluctuating pressure of the submarine is very different in different zones. The fluctuating value in the laminar flow area can be ignored, and the pressure fluctuation in the vortex shedding zone is larger obviously than that in the turbulent flow area.

There are also differences in the convective velocity in different regions, which decreases from upstream to downstream.

(3) According to the DMD analysis results, the low-mode vortex structures dominate the flow field, and their scales are larger. However, the scale of the high-mode structure is small, and the energy ratio is low.

Acknowledgements

References

[1] Yu, L., Zhao, W., Wan, D. and Wu, J. (2022). Nonlinear noise of hydrofoil cavitation considering sound velocity variation and phase transitions. *Ocean Engineering*, **264**, 112506.

[2] Yu, L. J., Wu, J. W. and Wan, D. C. (2022). Correlation analysis between underwater noise and Liutex for DTMB4119 propeller. *Journal of Hydrodynamics*, **34(4)**, 585-595.

[3] Shi, B., Yang, X., Jin, G., He, G. and Wang, S. (2019). Wall-modeling for large-eddy simulation of flows around an axisymmetric body using the diffuse-interface immersed boundary method. *Applied Mathematics and Mechanics*, **40(3)**, 305-320.

[4] Yu, L., Wan, D. and Chen, L. (2021). Numerical study of flow-induced noise around cylinder and Rod-NACA0012 hydrofoil. *In The 31st International Ocean and Polar Engineering Conference*. OnePetro.

[5] Mukha, T., Bensow, R. E. and Liefvendahl, M. (2021). Predictive accuracy of wall-modelled large-eddy simulation on unstructured grids. *Computers & Fluids*, **221**, 104885.

[6] Yu, L., Zhao, W., & Wan, D. (2023). Dual Mesh Technique for Hydroacoustic Prediction of Cavitating Flows around a Three-Dimensional Bullet. *In Journal of Physics: Conference Series* (Vol. **2458**, No. 1, p. 012021). IOP Publishing.

[7] Yin, G. and Ong, M. C. (2020). On the wake flow behind a sphere in a pipe flow at low Reynolds numbers. *Physics of Fluids*, **32(10)**, 103605.

[8] Liu, H. L. and Huang, T. T. (1998). Summary of DARPA SUBOFF experimental program data. *Naval Surface Warfare Center Carderock Div Bethesda MD Hydromechanics Directorate*.

[9] Yu, L., Zhao, W., & Wan, D. (2023, March). Propeller Wake and Noise Analysis Based on the Third-Generation Vortex. *In Liutex and Third Generation of Vortex Identification: Workshop from Aerospace and Aeronautics World Forum 2021* (pp. 257-270). Singapore: Springer Nature Singapore.

A new method based on perturbation for vibration characteristics of risers

Zhen Wang†* [1]， Jiaqi Han[2],

[1]School of Mathematical Sciences, Beihang University, China
[2]School of Mathematical Sciences, Dalian University of Technology, China.

*Presenting author: wangzmath@163.com
†Corresponding author: wangzmath@163.com

Abstract

By using a new method based on perturbation method and Fredholm theory, the eigenvalues and eigenfunctions of Euler-Bernoulli beam vibration are solved analytically. The eigenvalues and eigenfunctions under boundary conditions are given for the tension beam problem which considers the variation of tension in vertical marine riser from top to bottom caused by gravity effect. The asymptotic analytical results are compared with classical results to verify the effectiveness of the perturbation method. The influence of the change of the bottom tension on the perturbation method is analyzed. The results show that: for each order of natural frequency of riser, the natural frequency will gradually increase with the increase of bottom tension, and the obtained natural frequency will become more and more accurate; when the bottom tension is constant, the perturbation method in this paper can give good accuracy for higher order natural frequency rather than lower order.

Keywords: Euler-Bernoulli beam; perturbation method; Fredholm theory; eigenvalues; eigenfunctions

1. Introduction

The ocean is rich in oil, natural gas and other energy resources, developing deep sea oil and gas resources will alleviate the tension of energy supply and demand effectively and guarantee national energy security effectively. In recent years, with the development of deep-sea oil and gas fields, oil and gas mixed transport riser has attracted more and more attention[1]. The dynamic response of deep-sea risers under severe sea conditions such as typhoon is related to platform safety. The accurate analysis of dynamic response and safety prediction of marine riser is very important for platform decision. When the natural frequency of the deep-sea riser is close to the frequency of the external force acting on the riser, the riser amplitude will increase significantly, and complex coupling resonance phenomenon will occur, leading to the fatigue damage and even failure of the riser[2], and then cause huge economic losses and serious environmental pollution.

The study of vibration characteristics of marine riser is very important to understand the safety of its dynamic behavior, and determining the natural frequency and mode shape of the riser is the key to analyze its dynamic behavior. Païdoussis and Issid (1974)[3] dealt with the dynamics and stability of flexible pipes containing flowing fluid by using physical experimental models and mathematical models, and the influence of internal fluid on the vibration characteristics of the pipeline is further determined. Considering the gravity effect of the vertical ocean riser, the tension in the riser changes from top to bottom. The governing equation is transformed into Euler Bernoulli beam equation with varying coefficient, which is hard to obtain analytic solution. Many domestic and foreign scholars have studied it with

numerical methods and semi-analytic methods. Dareing and Huang(1976)[4] gave the mathematical basis and numerical data for determining natural frequency, and an approximate solution is also given with power series method, Kirk et al.(1979)[5] used the classical Rayleigh-Ritz method to calculate the natural frequency, whose precision is higher than power series method, Kim(1988)[6] tried to use the Wentzel-Kramers-Brillouin (WKB) method to obtain natural frequencies and critical buckling loads of risers simply supported at both ends, the solution is agree well with the result under power series method, Soltanahmadi(1992)[7] obtained the natural frequency of single and multiple flexible riser system based on Fourier method, Zhang et al.(2000)[8] used finite element method to analysis the natural frequency, Cheng et al.(2002)[9] combined the dynamic stiffness method with the WKB theory, which assumes that the coefficients in the differential equation of motion are slowly varying, Graves and Dareing(2004)[10] used segmentation method to analysis the vibration of riser which is particularly useful for higher order natural frequency, Chen et al.(2009)[11] used differential transformation method to calculate the natural frequency and mode shapes under eight different boundary conditions, furthermore, Chen et al.(2014)[12] adopted a new method named variational iteration method to discuss and analyze the natural frequency of the riser. Duan et al. (2021)[13] investigated vortex-induced vibration(VIV) dynamics of a fluid-conveying flexible riser subjected to external shear current, Wang and Zou(2022)[14] analyzed the effects of varying tension along pipe and buffer station on natural frequencies and mode shapes.

In engineering, many beams have non-uniform forms of force, material and stiffness. The corresponding vibration equation is a fourth order partial differential equation with variable coefficients whose solution is the key to analysis the natural vibration characteristics of marine riser. In this paper, we propose a new alternative method to calculate the natural frequency and mode shapes of the risers. The most significant advantage of this method is that it provides the analytical expression of the relationship between the inhomogeneity of tension variation along space and natural frequency.

In Section 2, the lateral vibration equation of riser considered the varying tension are presented. Section 3 outlines the perturbation method and Fredholm alternative theorem which are used to obtain the eigenvalues and eigenfunctions of the vibration of nonuniform Euler-Bernoulli beams. In Section 4, the natural frequencies and mode shapes of varying tension riser are given with the above method under the concrete numerical examples, the effectiveness of the perturbation method is verified by comparing with the classical methods simultaneously. The variation of natural frequency with parameter is also analyzed. Finally, Section 5 concludes with the main findings.

2. Governing equation

Production riser usually can be simplified as a long, continuous tube that is straight and vertical in the whole length direction, which is used to connect offshore platform to the sea floor. The structure diagram is shown in Fig.1. The coordinate system is established, taking the seabed as the origin of coordinates, the X-axis direction is the vertical upward direction of the riser, and the Y-direction is the horizontal direction, which is also the direction of the riser vibration displacement.
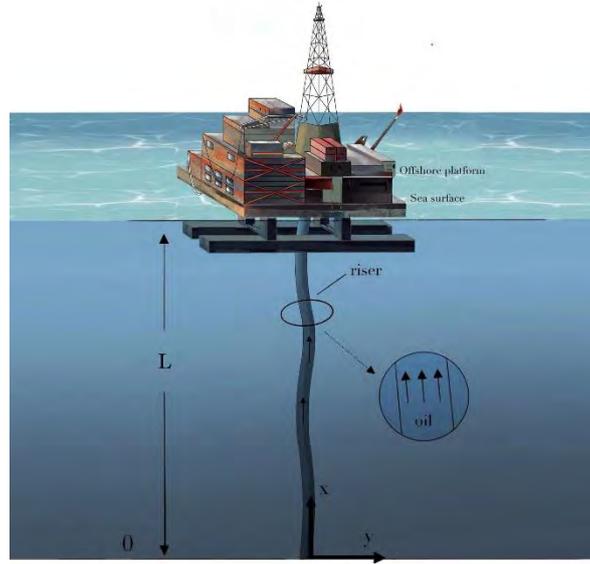
Fig. 1. Schematic diagram of riser structure

It is assumed that the riser is an Euler Bernoulli beam equation, and the influence of flexural rigidity and internal and external medium are considered. Using the form of following partial differential equation proposed in paper[4], the equation of transverse motion of riser can be written as：

$$\frac{\partial^2}{\partial \bar{x}^2}(E(\bar{x})I(\bar{x})\frac{\partial^2 \bar{y}}{\partial \bar{x}^2}) - \frac{\partial}{\partial \bar{x}}(T_e(\bar{x})\frac{\partial \bar{y}}{\partial \bar{x}}) + m_e(\bar{x})\frac{\partial^2 \bar{y}}{\partial t^2} = 0 \ , \tag{1}$$

where $T_e(\bar{x})$ is effective axial force and can be estimated as[15]:

$$T_e(\bar{x}) = T_B + \int_0^{\bar{x}} w(\bar{x})d\bar{x} + A_0(\bar{x})P_0(\bar{x}) - A_i(\bar{x})P_i(\bar{x}). \tag{2}$$

The physical meaning of the symbols used in Eq. (1), Eq. (2) are showed in Table 1:

Table 1. The symbols of riser

Serial	Parameter	Symbol
1	Lateral vibration displacement of the riser	$\bar{y}(\bar{x},t)$
2	Flexural rigidity of the riser	$E(\bar{x})I(\bar{x})$
3	Effective mass (per unit length) participating in the vibration of	$m_e(\bar{x})$
4	Effective axial force	$T_e(\bar{x})$
5	Tension at bottom of the riser	T_B
6	Inside cross-sectional area of the riser	$A_i(\bar{x})$
7	Outside cross-sectional area of the riser	$A_0(\bar{x})$
8	Internal pressure acted on the riser's wall	$P_i(\bar{x})$
9	External pressure acted on the riser's wall	$P_0(\bar{x})$
10	Net weight of the riser (per unit length)	$w(\bar{x})$
11	Length of entire riser	L
12	Weight density of inner media	γ_0
13	Weight density of outer media	γ_i

In this paper, the riser is assumed to be hinged at both ends of the riser ($x = 0$ and $x = L$). The boundary conditions can be expressed as follows:

$$\overline{y}(0,t) = 0, \overline{y}(1,t) = 0, \frac{\partial^2 \overline{y}(\overline{x},t)}{\partial \overline{x}^2}\bigg|_{\overline{x}=0} = 0, \frac{\partial^2 \overline{y}(\overline{x},t)}{\partial \overline{x}^2}\bigg|_{\overline{x}=1} = 0 . \tag{3}$$

Using variable separation method, the lateral vibration displacement of the riser $\overline{y}(\overline{x},t)$ can be expressed as the product of a spatial function $y^*(\overline{x})$ and a time function $T^*(t)$:

$$\overline{y}(\overline{x},t) = y^*(\overline{x})T^*(t) . \tag{4}$$

For simplifying the governing equation, the dimensionless variables are provided as follows[11]:

$$x = \frac{\overline{x}}{L}, y = \frac{y^*}{L}, b(x) = \frac{E(\overline{x})I(\overline{x})}{E(0)I(0)}, n(x) = \frac{T_e(\overline{x})L^2}{E(0)I(0)}, q(x) = \frac{m_e(\overline{x})}{m_e(0)}, \omega^2 = \frac{m_e(0)\lambda^2 L^4}{E(0)I(0)} , \tag{5}$$

where λ is the natural frequency of the riser. By using the spectral representation of Doyle[16], Eq. (1) can be written as:

$$\frac{\partial^2}{\partial x^2}(b(x)\frac{\partial^2 y}{\partial x^2}) - \frac{\partial}{\partial x}(n(x)\frac{\partial y}{\partial x}) - q(x)\omega^2 y = 0 . \tag{6}$$

In this paper, the flexural rigidity, axial force and effective mass (per unit length) participating in the vibration of the riser are considered uniform, Eq. (6) can be simplified as[11]:

$$y'''' - ((T + vx)y')' - \omega^2 y = 0 , \tag{7}$$

where $T = \frac{[T_B + L(A_0\gamma_0 - A_i\gamma_i)]L^2}{EI}, v = \frac{[w - A_0\gamma_0 + A_i\gamma_i]L^3}{EI}$, γ_0 is the weight density of inner media, γ_i is the weight density of outer media,

with non-dimensional boundary conditions:

$$y(0) = 0, y(1) = 0, \frac{\partial^2 y(x)}{\partial x^2}\bigg|_{x=0} = 0, \frac{\partial^2 y(x)}{\partial x^2}\bigg|_{x=1} = 0 . \tag{8}$$

3. Perturbation method

Defining $\varepsilon = \frac{v}{T}$, Eq. (7) of free vibration of the marine riser can be written as:

$$y'''' - T(1 + \varepsilon x)y'' - T\varepsilon y' - \omega^2 y = 0 . \tag{9}$$

To solve Eq. (9) with the boundary conditions Eq. (8) by perturbation method, we suppose that the natural frequency ω and the solution of Eq. (9) y are in a series of ε:

$$\omega = \sum_{i=0}^{\infty}\omega_i\varepsilon^i , \quad y = \sum_{i=0}^{\infty}y_i\varepsilon^i . \tag{10}$$

Then substituting Eq. (10) into Eq. (9), we can obtain:

$$\sum_{i=0}^{\infty}y_i''''\varepsilon^i - T(1+\varepsilon)\sum_{i=0}^{\infty}y_i''\varepsilon^i - T\varepsilon\sum_{i=0}^{\infty}y_i'\varepsilon^i - (\sum_{i=0}^{\infty}\omega_i\varepsilon^i)(\sum_{i=0}^{\infty}\omega_i\varepsilon^i)(\sum_{i=0}^{\infty}y_i\varepsilon^i) = 0 , \tag{11}$$

with the boundary conditions:

$$y_i(0) = 0, y_i(1) = 0, y_i''(0) = 0, y_i''(1) = 0 . \tag{12}$$

By collecting the coefficients of ε^i, and setting to zero one by one, we can obtain an equation system about y_j, $\omega_j, (j = 0,1,...i)$. As a result, one can solve the equations corresponding to each $\varepsilon^i, (i = 0,1,2,...)$ to obtain y_i to solve the equation system step by step.

The governing equation is self-adjoint, which has many perfect properties. It can be found in classical books[17]. All of eigenvalues are real, positive and simple, the eigenfunctions are orthogonality to each other in considered finite interval[14]. In fact, the equation corresponding to each ε^i will have an infinite number of eigenvalues and corresponding eigenfunctions if the

boundary conditions are satisfied. As a result, we write the series of eigenvalues and eigenfunctions in vector form, the Eq. (10) is written as:

$$\vec{\omega} = \sum_{i=0}^{\infty} \vec{\omega_i}\varepsilon^i = \begin{pmatrix} \omega_{01} \\ \vdots \\ \omega_{0N} \end{pmatrix} + \varepsilon \begin{pmatrix} \omega_{11} \\ \vdots \\ \omega_{1N} \end{pmatrix} + \varepsilon^2 \begin{pmatrix} \omega_{21} \\ \vdots \\ \omega_{2N} \end{pmatrix} + \dots$$

$$\vec{Y} = \sum_{i=0}^{\infty} \vec{Y_i}\varepsilon^i = \begin{pmatrix} Y_{01} \\ \vdots \\ Y_{0N} \end{pmatrix} + \varepsilon \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1N} \end{pmatrix} + \varepsilon^2 \begin{pmatrix} Y_{21} \\ \vdots \\ Y_{2N} \end{pmatrix} + \dots , \tag{13}$$

where $\vec{\omega_i} = \begin{pmatrix} \omega_{i1} \\ \vdots \\ \omega_{iN} \end{pmatrix}$, $\vec{Y_0} = \begin{pmatrix} Y_{01} \\ \vdots \\ Y_{0N} \end{pmatrix}$ ($N \to \infty$) is the eigenvalues and the eigenfunctions of the

equation corresponding to ε^i .The eigenvalues are also called the natural frequencies, and the corresponding eigenfunctions are called the mode functions.

Then we will solve $\vec{\omega_i}$,$\vec{Y_i}$ one by one. First, considering the coefficients of ε^0 ,we can get:

$$\varepsilon^0 : \quad y_0'''' - Ty_0'' - \omega_0^2 y_0 = 0 . \tag{14}$$

The character equation is:

$$\lambda^4 - T\lambda^2 - \omega_0^2 = 0 , \tag{15}$$

then we can get eigenvalues $\vec{\omega_0}$ and eigenfunctions $\vec{Y_0}$ of the Eq. (14) with the boundary conditions Eq. (12) as follows:

$$\vec{\omega_0} = \begin{pmatrix} \omega_{01} \\ \omega_{02} \\ \vdots \\ \omega_{0N} \end{pmatrix}, \quad \omega_{0j} = j\pi\sqrt{(j\pi)^2 + T}, j = 1,2,\dots$$

$$\vec{Y_0} = \begin{pmatrix} Y_{01} \\ Y_{02} \\ \vdots \\ Y_{0N} \end{pmatrix}, \quad Y_{0j} = \sin(j\pi x), j = 1,2,\dots . \tag{16}$$

Then, considering the coefficients of ε^1 ,we can get:

$$\varepsilon^1 : \quad y_1'''' - Ty_1'' - Txy_0'' - Ty_0' - \omega_0^2 y_1 - 2\omega_0\omega_1 y_0 = 0 . \tag{17}$$

Substituting the solution of Eq. (17) $\vec{\omega_1}$,$\vec{Y_1}$ and $\vec{\omega_0}$ into the Eq. (17), we can obtain:

$$\vec{Y_1}'''' - T\vec{Y_1}'' - \vec{\omega_0}^2 \vec{Y_1} = Tx\vec{Y_0}'' - T\vec{Y_0}' + 2\vec{\omega_0}\vec{\omega_1}\vec{Y_0} . \tag{18}$$

The right side of the equation Eq. (18) is marked as function $f_0(x,\vec{\omega_1})$. Expanding Eq. (18):

$$Y_{11}'''' - TY_{11}'' - \omega_{01}^2 Y_{11} = TxY_{01}'' + TY_{01}' + 2\omega_{01}\omega_{11}Y_{01} \tag{19-1}$$

$$\vdots$$

$$Y_{1i}'''' - TY_{1i}'' - \omega_{0i}^2 Y_{1i} = TxY_{0i}'' + TY_{0i}' + 2\omega_{0i}\omega_{1i}Y_{0i} \tag{19-i}$$

$$\vdots$$

$$Y_{1N}'''' - TY_{1N}'' - \omega_{0N}^2 Y_{1N} = TxY_{0N}'' + TY_{0N}' + 2\omega_{0N}\omega_{1N}Y_{0N} . \tag{19-N}$$

According to the Fredholm alternative theorem, we can obtain each ω_{1i} needs to satisfy the following relationship:

$$\int_0^1 f_0(x,\omega_{1i})Y_{0i}dx = 0 . \tag{20}$$

then we get the expression of ω_{1i} and $\vec{\omega_1}$:

$$\vec{\omega_1} = \begin{pmatrix} \omega_{11} \\ \omega_{12} \\ \vdots \\ \omega_{1N} \end{pmatrix}, \quad \omega_{1i} = -\frac{1}{4}\frac{iT\pi}{\sqrt{(i\pi)^2 + T}}, i = 1, 2, \ldots. \tag{21}$$

For each ω_{1i}, by substituting it into $f_0(x, \omega_{1i})$, we can obtain corresponding Y_{1i} through solving the Eq.(19-i), we can suppose that Y_{1i} is in a series of Y_{0j}, then expand and rearrange Eq.(19-i), we can obtain:

$$Y_{1j} = \sum_{j=1}^{\infty} c_{ij} Y_{0j}, i = 1, 2, \ldots \tag{22}$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{\infty} c_{ij}(\omega_{0j}^2 - \omega_{0i}^2) Y_{0j} = f_0(x, \omega_{1i}). \tag{23}$$

Integrating both sides of Eq. (23) after multiplying Y_{0j}, we can obtain the expression of the undetermined coefficient c_{ij}:

$$c_{ij} = \begin{cases} 0 & i = j \\ \dfrac{1}{\omega_{0j}^2 - \omega_{0i}^2} \cdot \dfrac{\int_0^1 f_0(x, \omega_{1i}) Y_{0j} dx}{\int_0^1 Y_{0j} Y_{0j} dx} & i \neq j \end{cases}, \tag{24}$$

and the solution of Eq. (17) $\vec{Y_1}$:

$$\vec{Y_1} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1N} \end{pmatrix} = \begin{pmatrix} 0 & c_{12} & \cdots & c_{1N} \\ c_{21} & 0 & \cdots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \cdots & 0 \end{pmatrix} \begin{pmatrix} Y_{01} \\ Y_{02} \\ \vdots \\ Y_{0N} \end{pmatrix}. \tag{25}$$

Similar to the above calculation method, we can get $\vec{\omega_i}, \vec{Y_i}$ through $\vec{\omega_0}, \vec{\omega_1}, \ldots, \vec{\omega_{i-1}}$, $\vec{Y_0}, \vec{Y_1}, \ldots, \vec{Y_{i-1}}$ one by one, and eventually we can get the natural frequency $\vec{\omega}$ and modal functions of the entire tube $\vec{Y}$. We firmly believe that this method can also be applied to many other nonlinear problems with small amount of calculation and good accuracy.

In order to facilitate modal analysis, this paper adopt the following modal normalization method[19]:

$$\hat{Y}(x) = \frac{Y(x)}{\int_0^1 |Y(x)| dx}. \tag{26}$$

4. Results

4.1 Effectiveness of method

This section compares and analyzes the effectiveness of the method through concrete examples. The numerical example data are obtained from Dareing and Huang[4], which is a typical marine drilling riser in the northern North Sea. The whole riser is assumed uniform, which is means the flexural rigidity $E(\bar{x})I(\bar{x})$, the effective mass (per unit length) $m_e(\bar{x})$, the weight of the riser (per unit length) $w(\bar{x})$ are constant. The numerical example data is as follows: the riser length $L = 152.4m$, $EI = 2.7007 \cdot 10^8\, Nm^2$, $m_e = 995.9092\, kg/m$, $w = 3123.0952\, N/m$, the inside and outside cross-sectional area of the riser $A_i = 0.2778m^2$, $A_0 = 0.2917m^2$, the weight

density of inner and outer media $\gamma_i = 13352.6993\,N/m^3$, $\gamma_0 = 10179.2674\,N/m^3$ .We can obtain $T = 100, v = 50$.

By the perturbation method described above, we get the natural frequency and modal function by truncating to the second power term ε^2 , which need to calculate the first two terms, i.e. $\overrightarrow{\omega_0}, \overrightarrow{Y_0}$, $\overrightarrow{\omega_1}, \overrightarrow{Y_1}$. The first five natural frequencies are showed in Table 2. The same data is also adopted and analyzed by Kim (1988) [6] used WKB method, Zhang et al.(2000) [8] used finite element method, Cheng et al.(2002) [9] used WKB-DSSM method, Graves and Dareing(2004)[10] used segmentation method, Chen(2009) [11] used differential transformation method, Chen(2014) [12] used variational iteration method.

Table 2. The first five natural frequency under different methods

Natural frequency	ω_1	ω_2	ω_3	ω_4	ω_5
Power series (Dareing and Huang 1976)	0.8150	1.8036	3.0876	4.7375	6.7890
WKB (Kim 1988)	0.8148	1.8039	3.0833	4.7333	6.7893
WKB-DSSM (Cheng et.al.2002)	0.8150	1.8037	3.0878	4.7377	6.7896
Segmentation (Graves and Dareing 2004)	0.8225	1.8033	3.0819	4.7281	6.7796
DTM,VIM (Chen et.al.2009,2014)	0.8150	1.8038	3.0879	4.7377	6.7899
Perturbation method	**0.8223**	**1.8129**	**3.0960**	**4.7442**	**6.7945**
Error= (Perturbation method- Power series)/ Power series	*0.008957*	*0.0051560*	*0.0027210*	*0.001414*	*0.000810*

Table 2 depicts not only the first five order natural frequency are obtained by the classical calculation method and the perturbation method in this paper but also the error between power series method of Dareing and perturbation method. The results show: the natural frequencies obtained by the perturbation method are close to those found using other methods, the maximum relative error is of order $O(10^{-3})$; at the same time, the relative error decreases with the increase of natural frequency order.

One of the remarkable advantages of the perturbation method in this paper is the low computational cost. The data in Table 2 indicates the relative error no more than order $O(10^{-3})$ that it just need to truncate to the first two terms, i.e. we only need to solve the Eq.(14), Eq.(16) and simultaneously the relative error of natural frequencies $\omega_i (i \geq 5)$ can attain order $O(10^{-4})$. In other words, the perturbation method in this paper is easier to obtain accurate higher order frequencies. The determination of high order natural frequencies for risers sometimes becomes important in engineering practice[12].
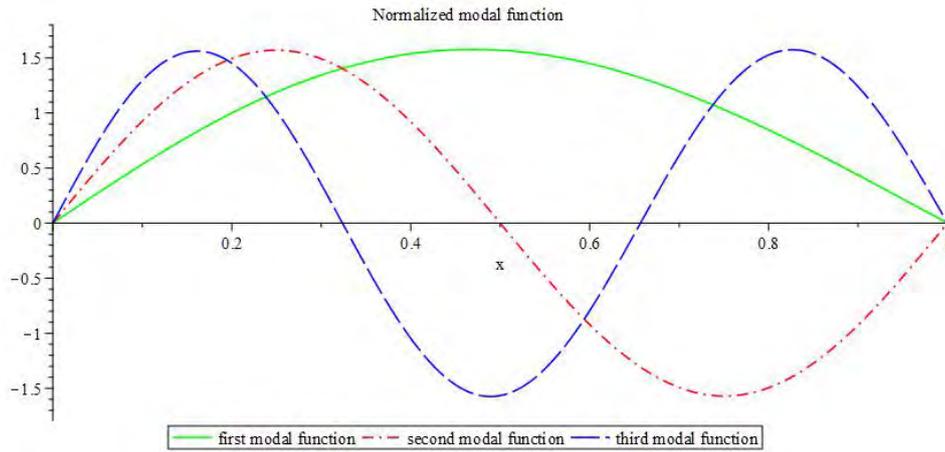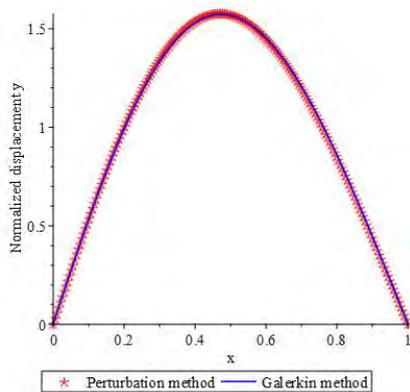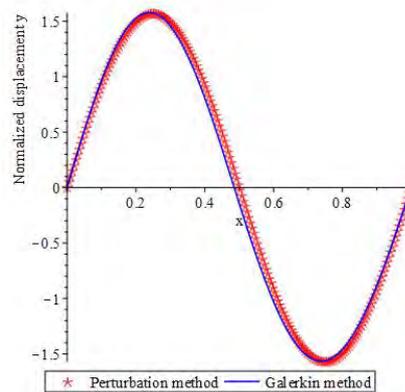
(a)
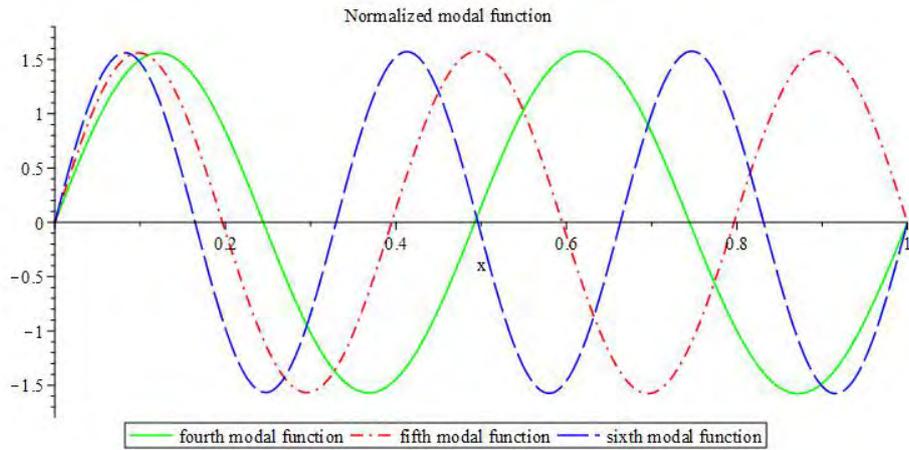
(b)

Fig.2. Normalized mode functions at the first six natural frequencies. (a) the first three mode, (b) the fourth, fifth, sixth mode

The mode shapes can be plotted after obtaining the natural frequency. Fig. 2 shows the first six normalized modal functions which normalized with Eq. (26).
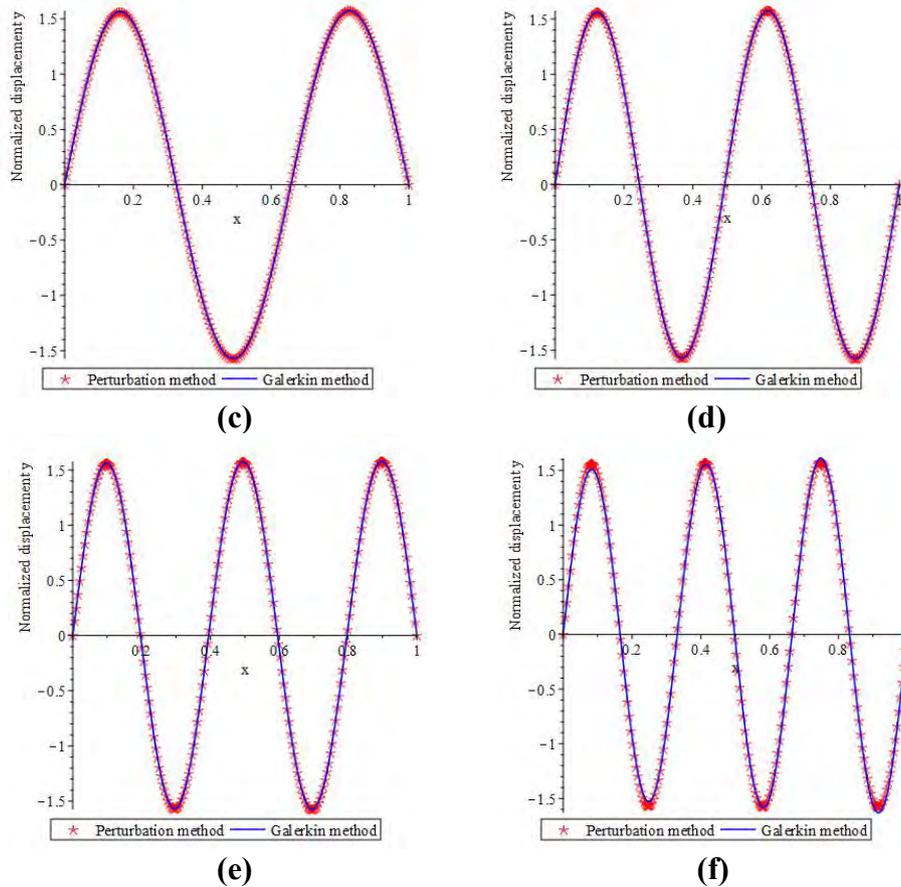

(a)

(b)

Fig.3. Comparison of normalized mode functions between Galerkin method and perturbation method. (a) first mode, (b) second mode, (c) third mode, (d) fourth mode, (e) fifth mode, (f) sixth mode

In this paper, we take Galerkin method as the standard solution which can be found in the book of Païdoussis[18] to compare and analyze with the perturbation method. The first six normalized mode functions with Galerkin method and perturbation method can be obtained in Fig.3. It can be seen that the mode shapes with perturbation method are agree well with that with Galerkin method.

4.2 Influence of parameter ε

Factors such as the change of the bottom tension, the length of riser and the weight density of inner and outer media will affect the natural frequency and mode shape of the riser. To facilitate the study of the effect of the bottom tension on the natural frequency, we can directly equivalently consider the change in ε . We only change the bottom tension T_B , and the other parameters are the same as in the above numerical example. In this paper we consider the case is that T change from 100 to 550 by varying T_B when $v = 50$. Fig.4 shows the corresponding first six natural frequency varies with ε which is range from 1/11 to 1/2 under two different methods. Fig.4 indicate: (i) the two lines almost coincide shows that the perturbation method has little deviation with the classical method, and also confirms the effectiveness of the perturbation method; (ii) for each frequency order, as the ε decreases i.e. bottom tension increases, the natural frequency increases gradually.
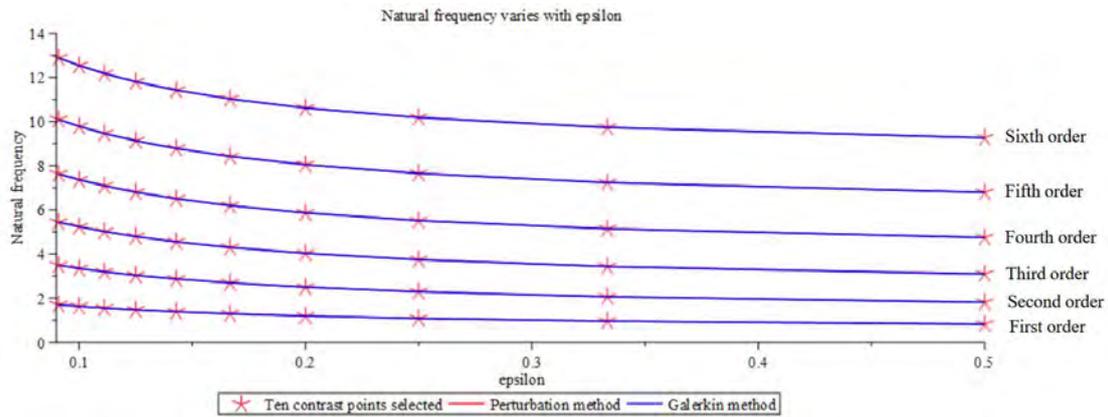
Fig.4. Comparison of first six natural frequency between Galerkin method and perturbation method

For a clearer comparison, we select the 5 points at ε =1/2,1/4,1/6,1/8,1/11, the natural frequency was calculated by Galerkin method and perturbation method respectively, which is shown in Table 3. Simultaneously, the calculation method of relative error in Table 3 is as follows:

Relative error = (Perturbation method – Galerkin method)/ Galerkin method

Table 3. The first five natural frequencies under Galerkin method and perturbation method

		$\varepsilon=1/11$	$\varepsilon=1/8$	$\varepsilon=1/6$	$\varepsilon=1/4$	$\varepsilon=1/2$
First order	Galerkin method	1.7031	1.4683	1.2881	1.0781	0.8150
	Perturbation method	1.7039	1.4695	1.2899	1.0812	0.8223
	Relative error	**0.00047**	**0.000817**	**0.001397**	**0.002875**	**0.008957**
Second order	Galerkin method	3.4915	3.0352	2.6883	2.2892	1.8037
	Perturbation method	3.4929	3.0373	2.6912	2.2939	1.8129
	Relative error	**0.000401**	**0.000692**	**0.001079**	**0.002053**	**0.005101**
Third order	Galerkin method	5.4438	4.7891	4.2975	3.7419	3.0878
	Perturbation method	5.4455	4.7915	4.3008	3.7467	3.0960
	Relative error	**0.000312**	**0.000501**	**0.000768**	**0.001283**	**0.002656**
Fourth order	Galerkin method	7.6270	6.8017	6.1906	5.5121	4.7375
	Perturbation method	7.6289	6.8042	6.1938	5.5165	4.7442
	Relative error	**0.000249**	**0.000368**	**0.000517**	**0.000798**	**0.001414**
Fifth order	Galerkin method	10.0953	9.1273	8.4204	7.6485	6.7894
	Perturbation method	10.0971	9.1297	8.4234	7.6523	6.7946
	Relative error	**0.000178**	**0.000263**	**0.000356**	**0.000497**	**0.000766**
Sixth order	Galerkin method	12.8900	11.8051	11.0226	10.1802	9.2615
	Perturbation method	12.8918	11.8073	11.0252	10.1834	9.2656
	Relative error	**0.00014**	**0.000186**	**0.000236**	**0.000314**	**0.000443**

From Table 3,we can find: (i) for each order of frequency, the relative errors decreases as ε decreases, it means the accurate natural frequencies will easier to get as the ε decreases; (ii) the relative error of the first four natural frequencies no more than order $O(10^{-3})$, however it can achieve order $O(10^{-4})$ at most from the fifth order natural frequency. It indicates the accuracy of the perturbation method increases as the order of natural frequency increases.
The above conclusions obtained by perturbation method are in line with previous cognition which also confirms the validity of the method in this paper.

Conclusions

In this paper, aiming to solve the problem of changed tension riser, we provide an alternative calculation method named perturbation method for getting natural frequency and mode shapes of vibration for deep sea oil production system. This method gives us a clearer understanding of the influence of riser tension inhomogeneity on the natural frequency and mode functions.

Firstly, the analytic expressions of eigenvalues and eigenfunctions of vibration of Euler-Bernoulli beams equation are obtained by using the perturbation method and Fredholm alternative theorem.

Secondly, the effectiveness of the perturbation method is verified by comparative analysis. Using the same numerical example, the natural frequency obtained through perturbation method agree well with that obtained by most classical method in the literature, and it can achieve the same relative error with the low computational cost. The mode shapes under perturbation method and Galerkin method almost coincide.

Thirdly, we analyze the influence of the choice of parameter ε on the natural frequency. The results indicate that with the decrease of ε ,the natural frequency will increase gradually, and the relative error will decrease gradually. At the same time, the accuracy of the perturbation method increases as the order of natural frequency increases.

References

[1] Sparks CP, (2006) Fundamentals of Marine Riser Mechanics: Basic Principles and Simplified Analysis. The United States of America: Library of Congress Cataloging-in-Publication Data.

[2] Kong L B, Zhang H M, Chen Y B, et al. (2018) The analysis of influence factors on natural frequency of deep-sea risers. Ship Science and Technology,40(1): 70-74.

[3] Païdoussis MP, Issid NT, (1974) Dynamic stability of pipes conveying fluid. Journal of Sound and Vibration,33(3): 267-294.

[4] Dareing, D.W., Huang, T., (1976) Natural frequency of marine drilling riser. Journal of Petroleum Technology. Tech28,813–818.

[5] Kirk, C.L., Etok, E.U., Cooper, M.T., (1979) Dynamic and static analysis of a marine riser. Appl. Ocean Res. 1 (3),125–135.

[6] Kim YC, (1988) Natural frequencies and critical buckling loads of marine risers. Journal of Offshore Mechanics and Arctic Engineering,110(1): 2-8.

[7] A. Soltanahmadi, (1992) Determination of flexible riser natural frequencies using Fourier analysis. Marine Structures,5(2-3): 193-203.

[8] Zhang YL, Gorman DG, Reese JM, (2001) A finite element method for modelling the vibration of initially tensioned thin-walled orthotropic cylindrical tubes conveying fluid. Journal of Sound & Vibration,245(1): 93-112.

[9] Cheng Y, Vandiver JK, Moe G, (2002) The linear vibration analysis of marine risers using the WKB-based dynamic stiffness method. Journal of Sound and Vibration,251(4): 750-760.

[10] Graves, J.R., Dareing, D.W., (2004) Direct method for determining natural frequencies of marine risers in Deep Water. ASME, J. Energy Reso. Tech 126,47–53.

[11] Chen, Y.F., Chai, Y.H., Li, X., Zhou, J., (2009) An extraction of the natural frequencies and mode shapes of marine risers by the method of differential transformation. Computers & Structures, 87 (21-22),1384–1393.

[12] Chen Y, Zhang J, Zhang H, et al. (2015) Re-examination of natural frequencies of marine risers by variational iteration method. Ocean Engineering,94(jan.15):132-139.

[13] Duan, J., Zhou, J., You, Y., et al., (2021) Effect of internal flow on vortex-induced vibration dynamics of a flexible mining riser in external shear current. Mar. Struct.80, 103094.

[14] Zhen Wang, Li Zou, (2022) Natural frequency analysis of deep-sea mining riser considering varying tension and buffer station, Ocean Engineering, Volume 264,112372.

[15] Sparks CP, (1984) The influence of tension, pressure and weight on pipes and risers deformation and stresses. ASME J Energy Resour Tech,106:46–54.

[16] Doyle JF, Wave propagation in structures. 175 5th Avenue, New York: Springer-Verlag; 1989.

[17] Courant, R., Hilbert, D., Methods of Mathematical Physics. I. Interscience Publishers, Inc, New York, 1953.

[18] Michael P. Païdoussis, Fluid-Structure Interactions: Slender Structures and Axial Flow, Volume1, Academic Press, London, 1998.

[19] Chai YH, Wang CM, (2006) An application of differential transformation to stability analysis of heavy columns. International Journal of Structural Stability and Dynamics,6(3): 317-33.

Machine Learning Based Approaches for Ultimate Compression Capacity Prediction of Concrete Filled Double Skin Steel Tube Columns

*Piyawat Boonlertnirun[1], Rut Su[1], Arnut Sutha[1] and †Sawekchai Tangaramvong[1]

[1]Center of Excellence in Applied Mechanics and Structures, Department of Civil Engineering, Chulalongkorn University, 10330 Bangkok, Thailand

*Presenting author: 6472051421@student.chula.ac.th
†Corresponding author: sawekchai.t@chula.ac.th

Abstract

This paper presents the machine learning-based methods that construct the accurate model prediction for the maximum compression capacity responses of concrete filled double skin steel tube (CFDST) columns under uniaxial compression forces. The so-called surrogate-assisted model is generated from the set of training data collected from available experimental results. The dataset is classified into two classes, namely the behaviors of short (section failure) and long (member failure) CFDST columns. Two machine learning, including gaussian process regression (GPR) and extreme gradient boosting (XGBoost), methods are encoded in this study. The input data considers geometry (i.e., external and internal diameters/thicknesses of steel tubes, and column length) and material properties (concrete compressive strength, and yield strengths of external and internal steel tubes) of the columns. The output data is the maximum compression capacity of the CFDST columns. The total training datasets comprise of 122 data from the short column tests and 146 data from the long column tests. The surrogate-assisted models determine the accurate uniaxial compression strengths for both short and long CFDST columns, where the good comparisons with relevant standard design specifications are evidenced. The long column responses given by the GPR model are more accurate than those performed by the XGBoost approach, but the short column responses given by the XGBoost approach are more accurate than those performed by the GPR.

Keywords: Concrete filled columns; Double skin steel tubes; Machine learning; GPR; XGBoost.

Introduction

In recent years, the realm of computer science has experienced notable progress, particularly in the domains of artificial intelligence (AI) and machine learning (ML) systems. These intelligent systems can seamlessly integrate with various engineering functionalities, offering structural engineers the ability to predict structural responses based on reliable input data. However, the AI and ML applications in structural engineering remain relatively limited, necessitating further exploration and study.

Concrete filled double steel tube (CFDST) columns depicted in Figure 1 possess numerous advantages owing to their exceptional strength and capacity to absorb substantial energy, making them highly suitable for seismic resistance in the construction of tall buildings, bridges and large structures. Alongside their remarkable load-bearing capabilities, the CFDST columns offer the additional benefits of rapid construction and reduce labor requirements, resulting in cost savings. Ongoing researches focus on the load capacity responses and their

predictive equations that assist engineers to design the CFDST columns accurately and the modification of relevant standard specifications, e.g., EUROCODE 4 and AISC [1, 2].

This study presents the two ML-based, called gaussian process regression (GPR) and extreme gradient boosting (XGBoost), methods to construct the surrogate-assisted model that can accurately approximate the maximum uniaxial compression capacities of CFDST columns. Two classes of (namely short and long) column responses are considered. The surrogate-assisted models are trained from the available dataset, including 122 short column tests in one case and 146 long column tests in the other. The accuracy of the generated predictive models are investigated through the comparisons with design standard specifications. The most efficient ML method suiting the CFDST column applications is summarized.

Methodology

Basic principle

The prediction models developed to estimate the compression capacity of CFDST columns are categorized into two distinct models based on the columns' weight-bearing behaviors: one for short columns and another for long columns. The machine learning algorithms employed in this research encompasses GPR and XGBoost. The dataset adopts in training the algorithms included both experimental test data and data generated through the ABAQUS simulation program.

The evaluation of model performance relies on the R^2 coefficient of determination, as defined by the following equation:

$$R^2 = 100 \times \left(1 - \frac{\sum_{i=1}^{N}(y_i - y'_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}\right), \tag{1}$$

where y_i represents the experimental data, y'_i represents the predicted values, and $\bar{y}$ denotes the mean of the experimental data. The obtained accuracy of the models is subsequently compared with that of established design codes such as AISC and EUROCODE 4. By employing this evaluation metric, the research aims to assess the efficacy of the machine learning models in predicting the compression capacity of CFDST columns and gauge their alignment with industry-standard design practices.
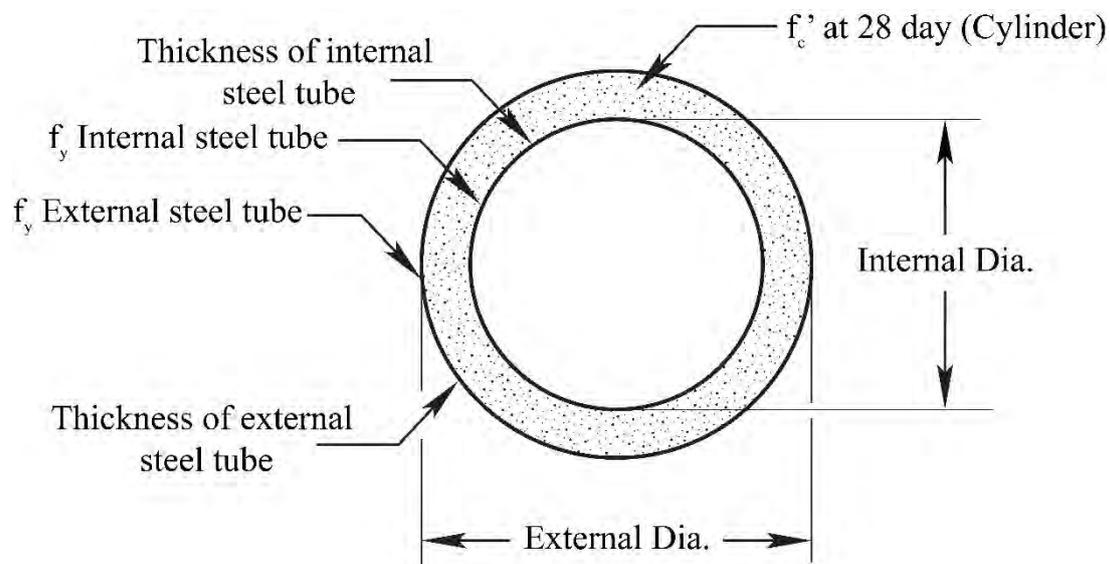

Figure 1. Cross section of CFDST.

AISC specifications

The original AISC specification [2] provides an equation specifically designed for steel reinforced concrete columns. To employ with Concrete Filled Double Steel Tube (CFDST) columns, there are two sets of equations to consider:

Case 1: $P_e < 0.44P_o$

$$(P_u)_{AISC} = 0.877P_e,$$
(2)

Case 2: $P_e \geq 0.44P_o$

$$(P_u)_{AISC} = P_o\left[0.658^{\frac{P_o}{P_e}}\right],$$
(3)

where P_o and P_e can be calculated from equations (4) and (5), f_{syo} and f_{syi} are the yield strength of the external and internal steel tubes, A_{so} and A_{si} are the section area of the internal and external steel tubes, and f'_c and A_c are the compressive strength of the concrete and section area of the concrete, respectively. K is for long columns using pinned support which is equal to 1 and L is the length of an unbracing column.

The parameter $(EI)_{eff}$ can be calculated from equation (6) where E_{so} and E_{si} are the Elastic Modulus of the external and internal steel tubes and E_{cm} can be calculated from equation (7). K_c is calculated from equation (8) with a maximum value of 0.9. W_c is the weight unit of concrete and has a value in the range 2300-2500 kg/m³.

$$P_o = f_{syo}A_{so} + 0.95f'_cA_c + f_{syi}A_{si},$$
(4)

$$P_e = \frac{\pi^2(EI)_{eff}}{(KL)^2},$$
(5)

$$EI_{eff} = E_{so}I_{so} + K_cE_{cm}I_c + E_{si}I_{si},$$
(6)

$$E_{cm} = w_c^{1.5}0.043\sqrt{f_c'},$$
(7)

$$K_c = 0.6 + 2\left(\frac{A_{so}}{A_c+A_{so}}\right) \leq 0.9.$$
(8)

EUROCODE4 specifications

The original EUROCODE4 equation is for Concrete Filled Steel Tubular (CFST), so for CFDST it was modified to add the factor of having an internal steel tube. There are thus 2 cases of equations:

Case 1: $\bar{\lambda} \leq 0.5$

$$(P_u)_{EC4} = n_af_{syo}A_{so} + f'_cA_c\left[1 + n_c\left(\frac{t_o}{D_o}\right)\left(\frac{f_{syo}}{f'_c}\right)\right] + n_af_{syi}A_{si},$$
(9)

Case 2: $\bar{\lambda} > 0.5$

$$(P_u)_{EC4} = f_{syo}A_{so} + f'_cA_c + f_{syi}A_{si},$$
(10)

Where n_a and n_c are reduction factor due to slenderness of column, and can be calculated as follows:

$$n_a = 0.25(3 + 2\bar{\lambda}) \leq 1.0,$$
(11)

$$n_c = 4.9 - 18.5\bar{\lambda} + 17\bar{\lambda}^2 \geq 0.$$
(12)

Next, $\bar{\lambda}$ can be calculated from equation (13), P_{cr} can be calculated from equation (15) and EI_{eff} can be calculated from equation (16) where $K_c = 0.6$.

$$\bar{\lambda} = \sqrt{\frac{P_{pl,Rd}}{P_{cr}}}, \tag{13}$$

$$P_{pl,Rd} = f_{syo}A_{so} + 0.85f_c'A_c + f_{syi}A_{si}, \tag{14}$$

$$P_{cr} = \frac{\pi^2(EI)_{eff}}{(KL)^2}, \tag{15}$$

$$EI_{eff} = E_{so}I_{so} + K_cE_{cm}I_c + E_{si}I_{si}. \tag{16}$$

Machine learning refers to the capability of a computing system to autonomously learn and analyze the relationships between input and output data by examining training data. Numerous algorithms exist for machine learning, each with its own strengths and weaknesses depending on the nature of the data. In the present study, the primary algorithms employed are GPR (Gaussian Process Regression) and XGboost. Previous research has demonstrated the efficacy of these algorithms in similar references [3,4].

Training data

For separating between short and long columns, the slenderness ratio [5] is used as in equation (17) where L_e is the effective column length, I is area moment of inertia and A is cross section of composite column.

$$\text{Slenderness ratio} = \frac{L_e}{\sqrt{I/A}} \tag{17}$$

Stub columns are those with slenderness ratio of less than 22 [5]. For stub columns, there are 122 data sets from actual tests [9-13,15-17] for training the machine learning algorithms, and the means came out as in Table 1.

Table 1. Training data for short column model.

Input	Maximum	Minimum	Average
External Diameter (mm)	603.40	114.00	212.33
External Thickness (mm)	6.77	0.90	3.97
F_{yo} (MPa)	535.00	220.00	362.65
Inner Diameter (mm)	402.10	33.50	107.06
Inner Thickness (mm)	5.77	0.90	3.06
F_{yi} (MPa)	425.00	220.00	358.86
F_c (MPa)	141.00	18.70	50.78
High (mm)	2502.00	342.00	617.31
Slenderness Ratio	21.66	5.55	2283.79
N_u EXP (kN)	5383.50	540.00	2283.79

Long columns, defined as those with a slenderness ratio exceeding 22 [5], play a significant role in this study. A total of 28 data sets derived from actual tests [9,14,21], along with 132 data sets obtained from ABAQUS simulations, are employed for training purposes. Evaluation criteria are established based on prior investigations [6, 7]. It should be noted that comparing the predictions generated by the simulations with those derived from real-world tests, the achieved R-Square value reaches an impressive 92.47%, as depicted in Figure 3.
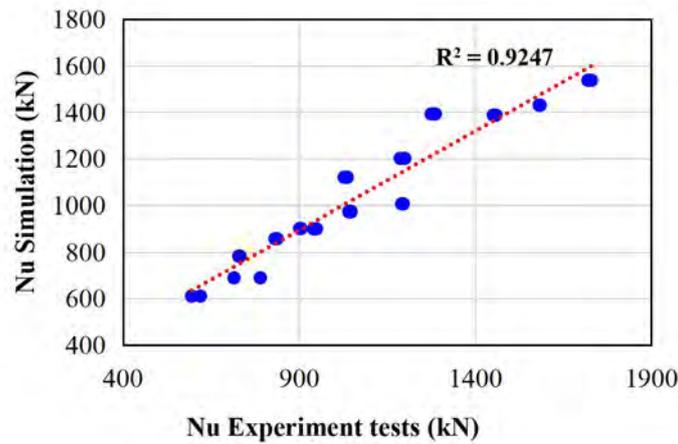
Figure .. R-Square values from simulations versus from actual tests

The evaluation of concrete's compressive capacity involves conducting tests using a 150 × 300 mm cylinder mold, following the established ASTM standard. In order to ensure consistency, data obtained from tests conducted with different types of molds are all converted to the 150 × 300 mm cylindrical mold format. This conversion is achieved by multiplying the results by a specific coefficient, as detailed in previous research [6].

Table 2. Training data for long column model.

Input	Maximum	Minimum	Average
External Diameter (mm)	550.00	114.00	216.67
External Thickness (mm)	10.00	1.96	3.42
F_{yo} (MPa)	549.00	259.00	388.60
Inner Diameter (mm)	250.00	45.00	95.99
Inner Thickness (mm)	10.00	1.96	3.16
F_{yi} (MPa)	549.00	298.00	395.49
F_c (MPa)	53.12	24.50	31.79
High (mm)	6500.00	1200.00	2757.98
Slenderness Ratio	95.93	23.67	48.24
N_u EXP (kN)	9882.58	497.19	2003.23

Result and conclusion

The GPR and XGBoost machine learning algorithms are utilized to create predictive models for the compression capacity of Concrete Filled Double Steel Tubes (CFDST). The predictive models specifically designed for stub columns are trained using data from 122 actual tests, yielding impressive results with an R-Square of 94.02% for GPR and 95.53% for XGBoost. In comparison, the R-Squares for the standard codes, AISC and EC4, are considerably lower at 82.97% and 85.90% respectively, demonstrating the superior accuracy of the machine learning models over the standard code.

On the other hand, for long columns, the algorithms are trained using a combination of data from 16 actual tests and 130 simulations conducted with the ABAQUS program. The resulting predictive models achieve an R-Square of 92.12% for GPR and 89.72% for XGBoost. Although these models exhibit slightly lower accuracy compared to the standard

codes (AISC with an R-Square of 93.43% and EC4 with an R-Square of 94.46%), they still provide valuable insights into the compressive capacity of long CFDST columns.

Overall, the machine learning algorithms showcase their potential by outperforming the standard codes in predicting the compression capacity of stub columns. While the accuracy of the predictive models for long columns is slightly lower, they still offer significant improvements over the standard codes, underscoring the value of incorporating machine learning techniques in the design and analysis of CFDST structures.

Table 3. R-square from each predictive model (stub columns).

Predictive Model	R^2	Average (N_u Exp/N_u Predict)
GPR	94.02	0.94
XGBoost	95.53	1.00
AISC	82.97	1.08
EC4	85.90	0.97
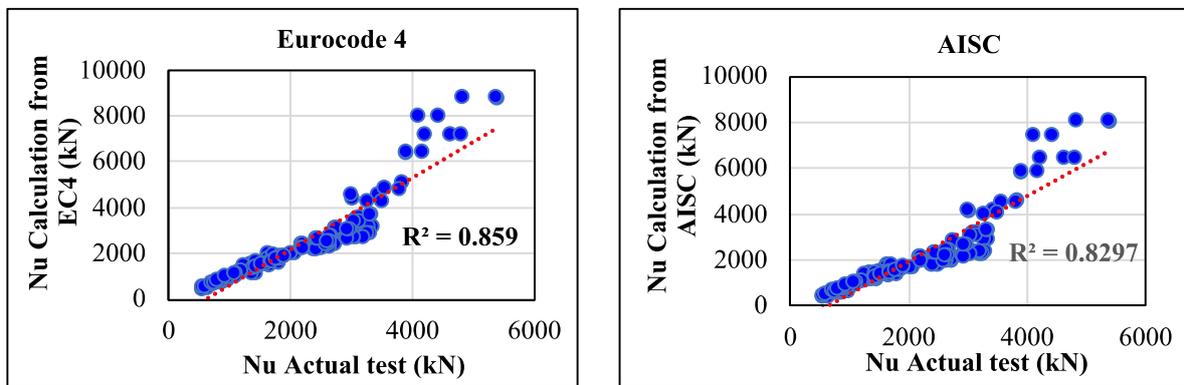
Figure 4. R-Square compared between actual test results and calculations based on the standard design code (Stub column).
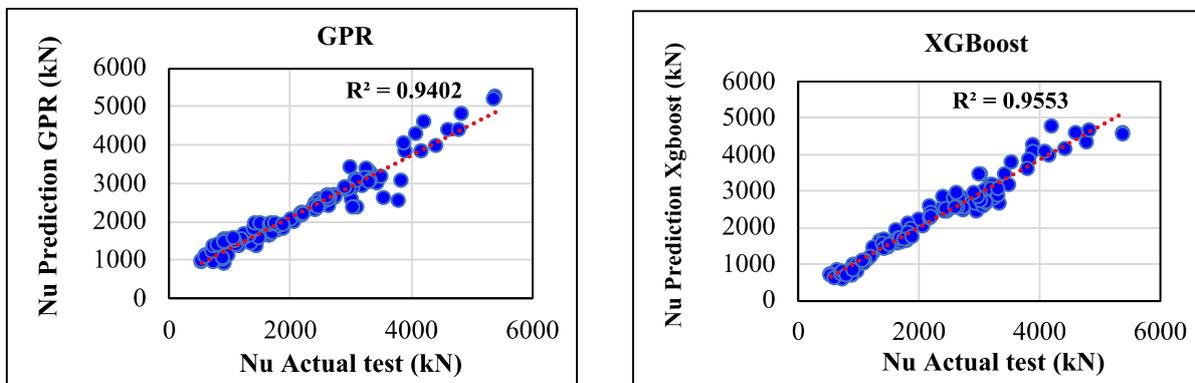
Figure 5. R-Square compared between actual test results and predictions made by machine learning algorithm (Stub column).

Table 4. R-square from each predictive model (long columns).

Predictive Model	R^2	Average (N_u Exp/N_u Predict)
GPR	92.12	1.00
XGBoost	89.72	1.03
AISC	94.46	1.00
EC4	93.43	0.85
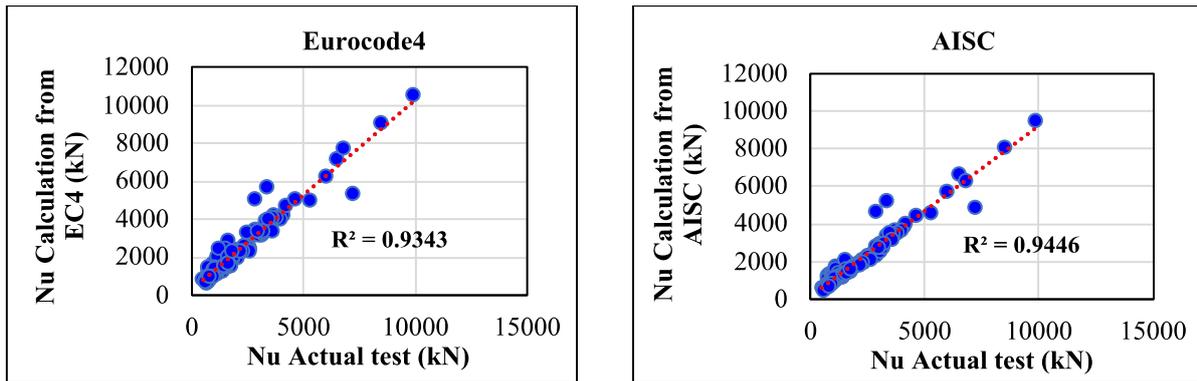
Figure 6. R-Square compared between actual test results and calculations based on the standard design code (long column).
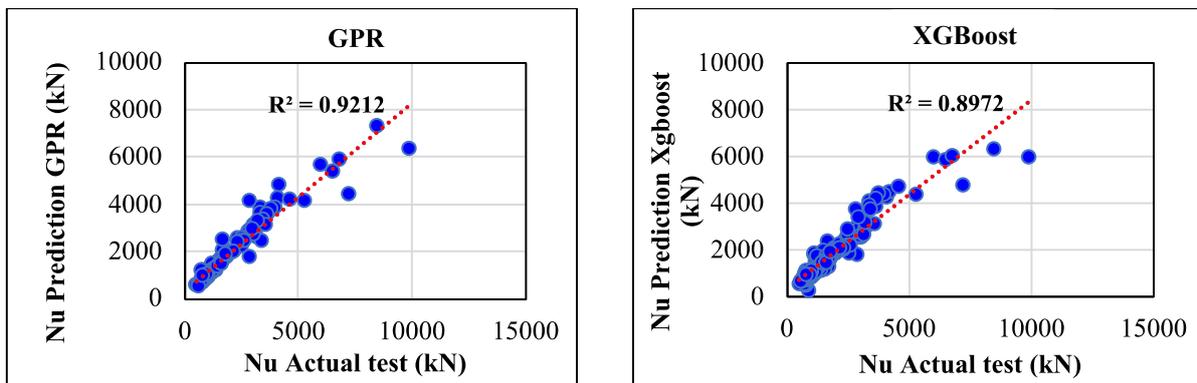
Figure 7. R-Square compared between actual test results and predictions made by machine learning algorithm (long column).

Acknowledgements

References

[1] Eurocode 4. (2004) Design of composite steel and concrete structures Part1.1: general rukes and rules for building. London: British Standard Institution, ENV 1994-1-1.

[2] AISC. (2010) Load and resistance factor design specification, for structural steel building. Chicago: American Institute of Steel Construction.

[3] Rasmussen, C.E. and Williams, C.K.I. (2006) Gaussian process for machine learning, The MIT press.

[4] Chen,T. and Guestrin,C. (2016) Xgboost: A Scalable tree boosting system, KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[5] Hassanein,M.F. and Kharoob,O.F. (2014) Analysis of circular concrete-filled double skin tubular slender columns with external stainless-steel tubes, *Thin-Walled Struct.* **79**,23-37.

[6] Elwell,D.J. and Fu,G. (1995) Compression testing of concrete cylinder vs cube, *Transportation Research Publications.* **119**,11-12.

[7] Ipek,S. and Güneyis,E.M. (2020) Nonlinear finite element analysis of double skin composite columns subjected to axial loading, *Archives of Civil and Mechanical Engineering.* **20**,9.

[8] Yan,X.F. and Zhao,Y.G. (2020) Compressive strength of axially loaded circular concrete-filled double-skin steel tubular short columns, *J.Constr. Steel Res.***170**,106-114.

[9] Tao,Z. and Han,L.H. (2004) Behavior of concrete-filled double skin (CHS inner and CHS outer) steel tubular stub columns and beam-columns, *J.Constr. Steel Res*. **60**,1129-1158.

[10] Zhao,X.L, Tong,L.W. and Wang,X.Y. (2010) CFDST stub columns subjected to large deformation axial loading, *Eng. Struct*. **32**,692-703.

[11] Elchalakani,M, Zhao,X.L. and Grzebieta,R. (2002) Tests on concrete filled double-skin (CHS outer and SHS inner) composite short columns under axial compression, *Thin-Walled Struct*. **40**,415-441.

[12]   Chen,J. and Hai,Y. (2011) Research on bearing capacity of short concrete filled double skin steel tubes columns under axial compression, *Adv. Mater. Res*. **168-170**,2154-2157.

[13]   Chen,J, Ni,Y.Y. and Jin,W.L. (2015) Column tests of dodecagonal section double skin concrete-filled steel tubes, *Thin-Walled Struct*. **88**,28-40.

[14]   Essopjee,Y. and Dundu,M. (2015) Performance of concrete-filled double-skin circular tubes in compression, *Compos. Struct*. **133**,1276-1283.

[15]   Yan,X.F, Zhao,Y.G. and Lin,.S. (2021) Compressive behaviour of circular CFDST short columns with high- and ultrahigh-strength concrete, *Thin-Walled Struct*. **164**,107-898.

[16]   Ekmekyapar,T. and Hasan,H.G. (2019) the influence of the inner steel tube on the compression behaviour of the concrete filled double skin steel tube (CFDST) columns, *Mar. Struct*. **66**,197-212.

[17]   Ekmekyapar,T, Alwan,O.H. and Hasan,H.G. (2019) B.A. Shehab, B.J.M. AL-Eliwi, Comparison of classical, double skin and double section CFST stub columns: Experiments and design formulations, *Constr. Steel Res*.**155**,192-204.

[18]   Pagoulatou,M, Sheehan,T, Dai,X.H. acircular concrete-filled double-skin steel tubular (CFDST) stub columnd Lam,D. (2014) Finite element analysis on the capacity of ns, *Eng. Struct*. **72**,102-112.

[19]   Uenaka,K, Kitoh,H. and Sonoda,K. (2010) Concrete filled double skin circular stub columns under compression, *Thin Walled Struct*. **48**,19-24.

[20]   Yan,X.F. and Zhao,Y.G. (2021) Experimental and numerical studies of circular sandwiched concrete axially loaded CFDST short columns, *Eng. Struct*. **230**,111167.

[21]   Han,L.H, Li,Y.J. and Liao,F.Y. (2011) Concrete-filled double skin steel tubular (CFDST) columns subjected to long-term sustained loading, *Thin Walled Struct*. **49**,1534-1543.

AUTHOR INDEX