A data-driven diagnostic tool for wind turbines under operational variability

†*Simona Bogoevska[1], Eleni Chatzi[2], Elena Dumova-Jovanoska[1], and Rüdiger Höffer[3]

[1]Faculty of Civil Engineering, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia.
[2]Department of Civil, Environmental and Geomatic Engineering, ETH Zürich, Switzerland.
[3]Department of Civil and Environmental Engineering, Ruhr-University Bochum, Germany.

*Presenting author: simona.bogoevska@gf.ukim.edu.mk
†Corresponding author: simona.bogoevska@gf.ukim.edu.mk

Abstract

The need for real-time condition assessment of complex systems relies on implementation of holistic Structural Health Monitoring (SHM) strategies that are capable of tracking structural behavior in a complete operational spectrum of the structure, distinguishing between true system changes and nonthreatening variations.

The proposed data-driven framework utilizes an autonomous bi-component tool able to link monitored structural response with random evolution of Environmental and Operational Parameters (EOP) affecting the monitored system. The approach combines the implementation of a Smoothness Priors Time Varying Autoregressive Moving Average (SP-TARMA) method for modeling the temporal variability in structural response, and a Polynomial Chaos Expansion (PCE) probabilistic model for modeling the propagation of response uncertainty. The computational tool is applied on long-term data, collected from an active sensing system installed for four years on a real operating WT structure located in Dortmund, Germany.

The twenty one-month tracking of the proposed PCE-SPTARMA diagnostic index, further assessed by means of statistic-based analysis, demonstrates that the proposed symbiotic treatment yields a robust model, capable of separating benign EOP fluctuations from potential pattern alterations due to actual structural damage. The obtained data-driven model verifies the future prospective of the strategy for development of an automated SHM diagnostic tool.

Keywords: **Data-driven diagnostics, Operating wind turbine, Structural variability, Uncertainty propagation**

Introduction

Latest technological advancements have fostered extensive application of various sensing techniques and acquisition systems on real engineering structures, thus shifting the focus towards hybrid analysis approaches (data/model) or purely data-based schemes (machine-learning black box approaches). Indeed, owing to various existing sources of uncertainty, complex behavior and variability characterizing the system and environment of actual in-service structures, data-aided assessment of operational engineering structures often remains a more accurate and computationally inexpensive alternative to approaches relying on physical law-based models.

In this context, continuous monitoring strategies facilitate the utilization of more objective and flexible tools pertinent to structural diagnostics and prognosis. Particularly for wind turbines, as systems characterized with time-varying dynamics and alternating operating nature, the adoption of automated identification tools, capable of unprejudiced diagnosis of the structural condition, becomes most valuable.

Semi-data-driven approaches, which rely on fusion of updated Finite Element Model (FEM) of the structure and recorded vibration responses, are reported as promising long-term strategies for monitoring fatigue accumulation, as well as acceleration and strain predictions at unmeasured locations of WT structures [1]-[3]. It is worth mentioning however that Operational Modal Analysis (OMA)-based methods are limited to implementation with time invariant systems [4]-[5], i.e. parked or idling condition of the structure, mode-by-mode or case-by-case investigation [6].

EOP-born variations in structural responses, known to compromise structural performance signatures and mimic real damage states of the structure, have placed data-driven diagnostics as a highly potential approach in tackling the challenge. New emerged strategies rely on eliminating influences of environmental factors from estimated performance indicators with algorithms adopted from the area of statistics, like fitting regression models, or projection methods when influencing variables are not attainable [7]-[9]. As systems are often monitored in an unknown or healthy baseline condition, robust novelty detection strategies and manifestation of detected outliers, related to changing environmental and operational conditions, versus structural damages, are recently gaining popularity in monitored full-scale engineering structures as well [10]-[12]. Comprehensive overviews of further commonly applied statistics based concepts in SHM can be found in [11]-[13].

Instead of filtering out EOPs, an alternative in the data-driven domain lies in integrating both structural response data and influencing agents within probabilistic models [14]-[17]. The workings of a PCE-SPTARMA data-driven tool, previously tested by the authoring team on two operating WT structures [17], are herein further explored by expanding the validation periods of monitored data. Whereas the twenty one-month long implementation on a real operating WT structure confirms the robustness of the strategy, fusion of the proposed strategy with a novelty detection algorithm and probability distribution divergence measure demonstrates the high potential for further automated structural health assessment.

Conceptual model

Conceptualized as a holistic approach, the proposed strategy addresses both behavioral signatures associated to collected WT response data, i.e., (short-term) non-stationarity and long-term temporal variability. This is accomplished through tracking of measured structural responses by an algorithm capable of capturing short and long-term variability of the observed system, thus providing a link between output-only vibration response data and measured EOPs, Fig. 1.

Fluctuations that are typical for the inherent (short-term) system dynamics are modeled by means of a parametric SP-TARMA method. Identified structural performance indicators, corresponding to short-term modeled responses, are then integrated into a PCE tool. The PCE probabilistic modeling approach enables long-term monitoring of structural response variability, further associated to the randomness of measured EOPs.

With this “binocular” visualization of the problem a selected PCE-SPTARMA output feature can serve as a robust diagnostic indicator for separating benign pattern alterations from actual structural damage.

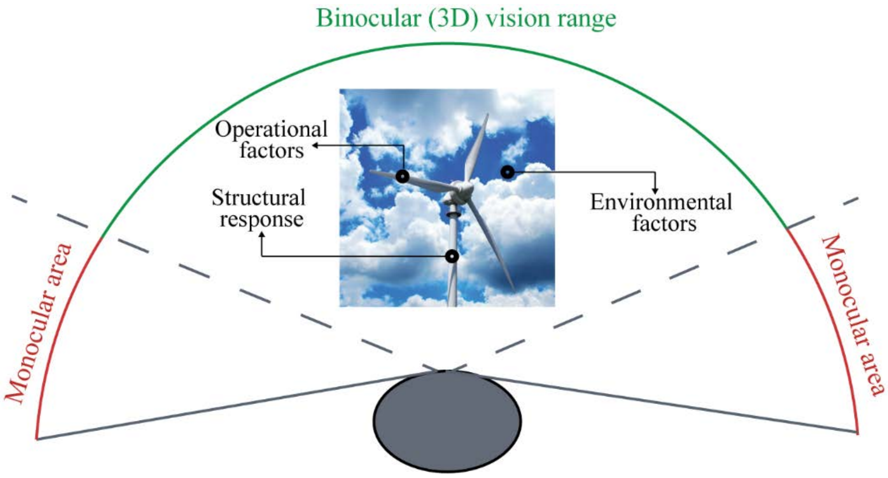


Figure 1. Conceptual model of the SHM strategy as a binocular eye vision metaphor

Theoretical framework

The proposed PCE-SPTARMA tool is a multicomponent algorithm comprising several computational methods commonly applied in a wide range of areas of research and application. The separate methods are herein summarized and presented with a concise theoretical overview. The reader is guided to further appropriate references for more detailed information on the theoretical background.

Modeling non-stationarity

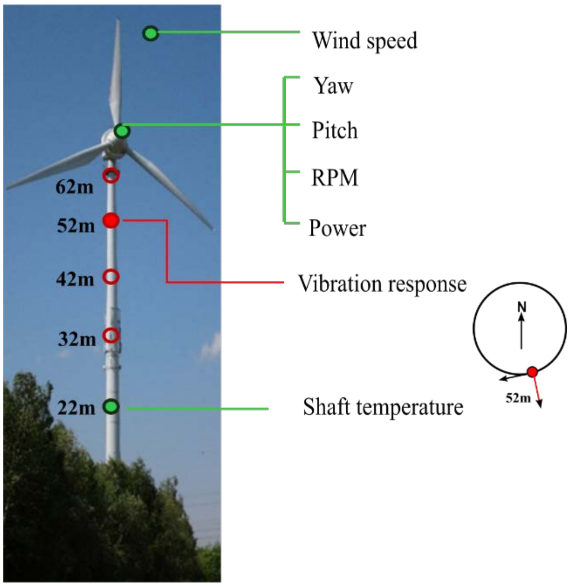
The nonstationary dynamics typical for an operating WT structure can be successfully tracked via the compact parametric formulation provided by the SP-TARMA models [18]. A full SP-TARMA model is completely described by an assemblage of three equations, one representing the modeled signal (system response), and two stochastic difference equations governing the time evolution of the unknown AR and MA parameters of the model. Thus, an adequate modelling of a measured nonstationary signal is ensured by proper selection of three user-defined parameters, i.e. the AR/MA order n, the ratio of the residual variances ν, and the order of the stochastic difference equations κ [19]. Statistical approaches such as minimization of the AIC (Akaike information criterion) or the BIC (Bayesian information criterion) improve the optimal selection of these values without overfitting the modeled signal. Finally, for a selected model $M(n, \nu, \kappa)$ the SP-TARMA model parameters are obtained via the Kalman Filter scheme combined with an Extended Least Squares-like algorithm [18].

Modeling uncertainty

The PCE tool is an uncertainty quantification method, which enables the relationship between outputs (structural response performance indicators) and inputs (environmental and operational loads) to the system. A PCE model can be described by a mathematical expansion of a random system output variable on multivariate polynomial chaos basis functions [17]. Spectral representations, such as the PCE method, rely on several regularity requirements, namely finite variance of the outputs, orthonormality of the polynomial basis, and statistical independence of the input variables [20]. Hence, the polynomial chaos basis functions orthonormal with respect to the probability space of the system's random inputs have to be properly selected to ensure the necessary orthogonality relationship. Furthermore, the statistical independence of input data needs to be properly verified and possibly addressed via computational approaches capable of extracting independent (latent) variables from observed data, such as the Independent Component Analysis (ICA) tool [21]. Then for a selected family of polynomial functions and maximum polynomial order P, the solution of the deterministic unknown parameters of a truncated PCE model are estimated via the least squares approach based on minimization of the sum of the squared residuals between true (observed) and modeled (predicted) system outputs [20].

Application case study

The described SHM strategy is implemented and tested for a 0.5MW WT erected in 1997, located in the vicinity of Dortmund, Germany, Fig. 2. A continuous measurement of acceleration response is recorded by triaxial accelerometers (PCB-3713D1FD3G MEMS sensors) mounted at five different height positions on the inner side of the WT shaft. Along with the vibration data, SCADA data are recorded with the same sampling frequency of 100 Hz. Within this paper results are presented for records corresponding to almost two complete years of continuously monitored data (January 2012 to September 2013). The last three months of year 2013, as well as scattered weeks in the previous period, are disregarded from the assessment as a result of missing temperature data from various sensor malfunctions.



Power	500 kW
Height of the rotor center	65 m
Length of blade (3 total)	19.13 m
Rotor speed	18-36 rpm
Blade material	GRP
Tower material	steel
Construction year	1997

Figure 2. Schematic overview of measured data (left), WT structure characteristics (right)

As a first step, acceleration records from a selected sensor location (marked at Fig. 2) were low-pass filtered and down-sampled to 12.5 Hz, with a cut-off frequency at 6 Hz. Subsequently, 10-min long preprocessed data sets were implemented within the short-term framework. The tuning of an appropriate SP-TARMA model to actual 10-min long signals is a crucial point of the short-term modeling phase. Towards this end, plots of the AIC and BIC for model order selection are significant indicative tools that facilitate the fitting process of the user-defined parameters of the SP-TARMA model (i.e. the model order n, the smoothness constraint order κ and the residual variance ratio ν). A detailed inspection of a selection of response data sets in conjunction with their estimated statistical criterion plots revealed an optimal fitting with the parameter values equal to $n=18$, $\kappa=1$, $\nu=0.0001$. For a graphic comparison, Fig. 3 presents a fitted and an over fitted 10-min long data set signal with $\nu=0.0001$ and $\nu=0.001$, respectively. Further discussion and graphical outputs on the SP-TARMA tuning process for the actual WT structure can be found in [17].

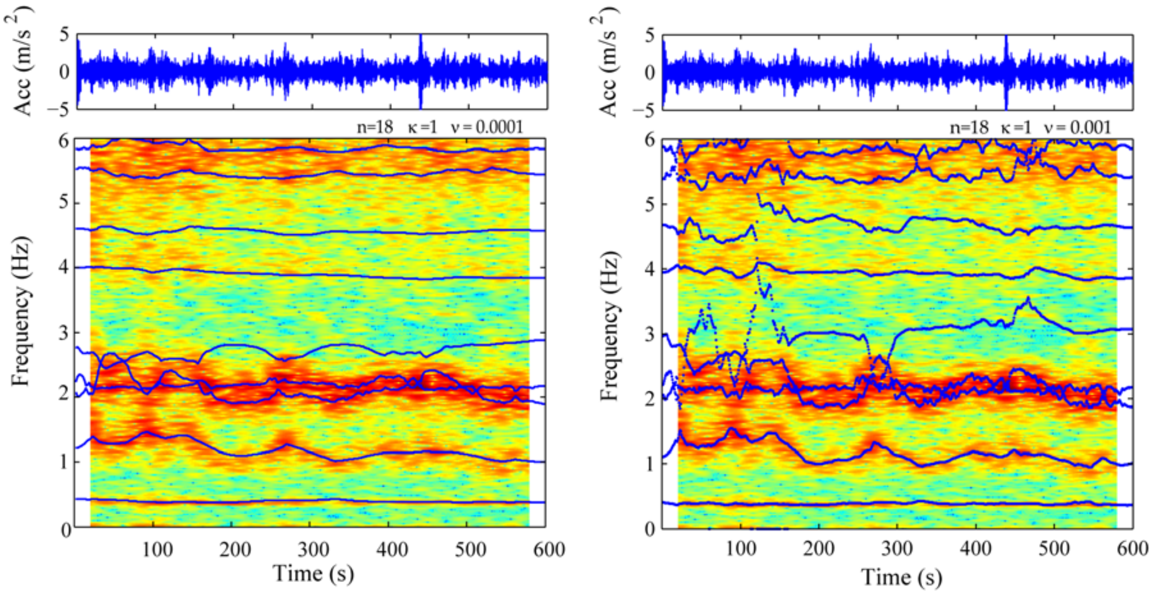


Figure 3. SP-TARMA model tuning, ν =0.0001 (Left), ν =0.001 (Right)

Measured data corresponding to operational and environmental parameters were organized as 10-min averages and further processed to be utilized as input variables into the long-term framework. More precisely, five SCADA parameters (wind velocity, RPM of the rotor, power production, yaw angle, and shaft temperature) were transformed to independent variables via the ICA algorithm. In order to preserve all the existing information the number of ICA latent variables was kept same as the maximum number of available EOP. For the purpose of satisfying the second PCE prerequisite, the ICA estimates are further transformed into uniformly distributed variables via use of the non-parametrically estimated cumulative distribution functions. Hence, in accordance with the uniform PDFs of the input data, the Legendre polynomials are selected as the PC functional basis. The standard deviation (std) of the SP-TARMA residuals for the 10 minute intervals, analyzed as part of the short-term framework, is selected as the PCE output parameter.

The selection of the second PCE user-defined parameter, the maximum polynomial order, is achieved via supervision of a PCE modeled output parameter Y^{PC} for a selected validation data range that clearly contains new ranges of input data. As presented in Fig. 4, the highest sensitivity to new records (marked red) of measured temperature and RPM values is linked to the maximum order P=5. This results in an evident discrepancy between the original output variable (Y) and the PCE modeled one (Y^{PC}). In addition, it was concluded that further increasing the maximum order does not significantly improve the accuracy of the expansion. More details regarding the PCE model generation, as well as the ICA transformation of the SCADA parameters are further elaborated in [17].

The previously described framework is herein utilized for the twenty one – month period of monitored data of the operating WT. In order to attain acceptable accuracy and alertness, as well as low computational cost, the assessment is performed for one 10-min data set per hour, resulting in total of 14064 analyzed data sets for the stated period. The standard deviation (std) of the PCE residuals is selected as a reliable Diagnostic Index (DI) able to directly demonstrate responsiveness to varying EOP, further verified with outlier analysis of validation input data sets. The proposed SHM strategy as a comprehensive three step tool is summarized in Fig. 5.

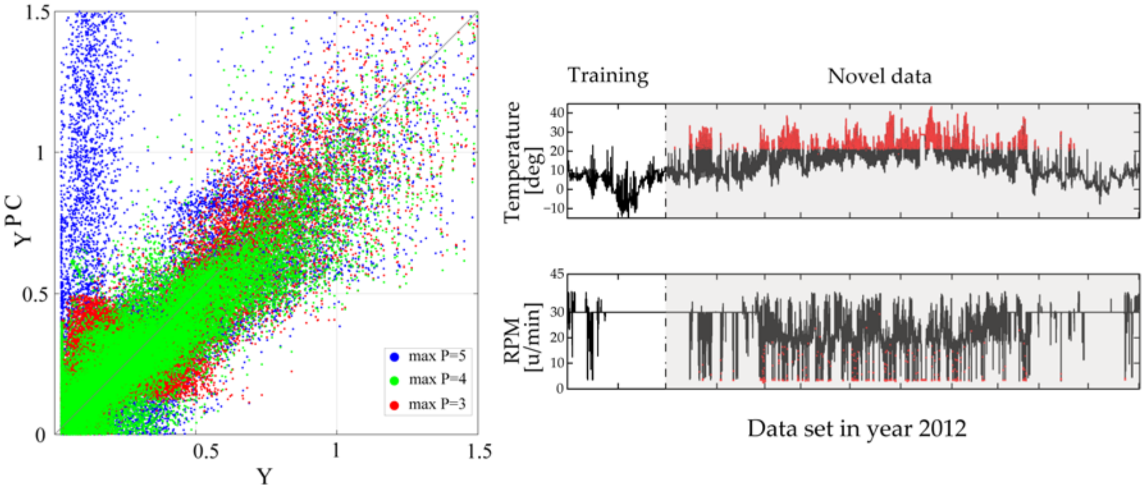


Figure 4. Tuning of PCE maximum order: modeled output (Left), input data (Right)

A preliminary testing of the sensitivity of the obtained DI is performed by means of outlier analysis on the input data time histories with the well-known Mahalanobis Distance (MD) discordancy measure. More precisely, for a p-dimensional multivariate sample $x_1, \dots, x_n$, the MD is defined as, [22]:

$$MD_i = \sqrt{(x_i - t_{tr})^T C_{tr}^{-1} (x_i - t_{tr})} \qquad for\ \ i = 1, \dots, n \tag{1}$$

where t_{tr} is the arithmetic mean and C_{tr} is the sample covariance matrix, estimated for a certain training period of an input data set. The x_i samples from a testing set which have MD beyond a predefined value are interpreted as novelties. Hence, the definition of thresholds is vital part of the process. An adaptive method that takes into account the actual empirical chi-square distribution function of the estimated MD (instead of a fixed quantile) is herein applied [22].

The sensitivity of the index values to unfamiliar EOP fluctuations is tested for two, four and twelve-month training periods. In Fig. 6, for a two-month training period, the validation sets of the estimated DI and statistical outlier analysis (univariate MD plot) of the input data time histories illustrate that index values exceeding the ± 3std thresholds (99.7% confidence intervals calculated for the fitted Gaussian distribution of the PCE estimation set errors) can be linked to novel data ranges of the measured influencing agents, more precisely temperature and RPM values between months March and November year 2012, as well as April and September in year 2013.

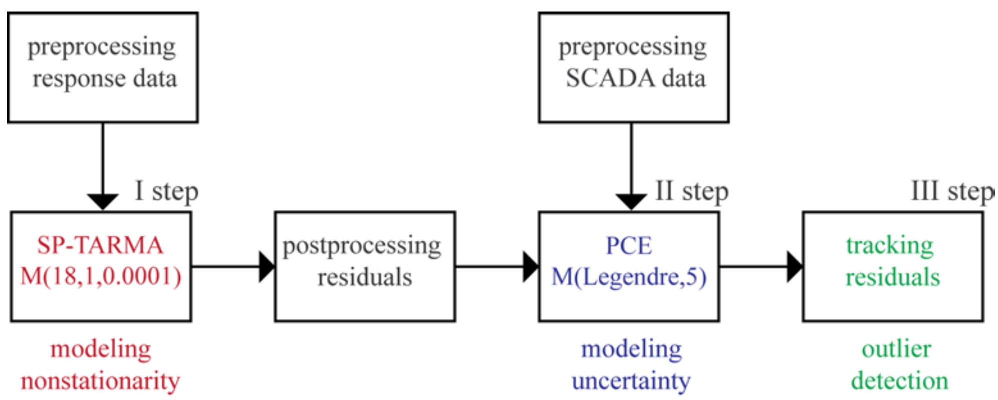


Figure 5. Schematic overview of the proposed SHM framework

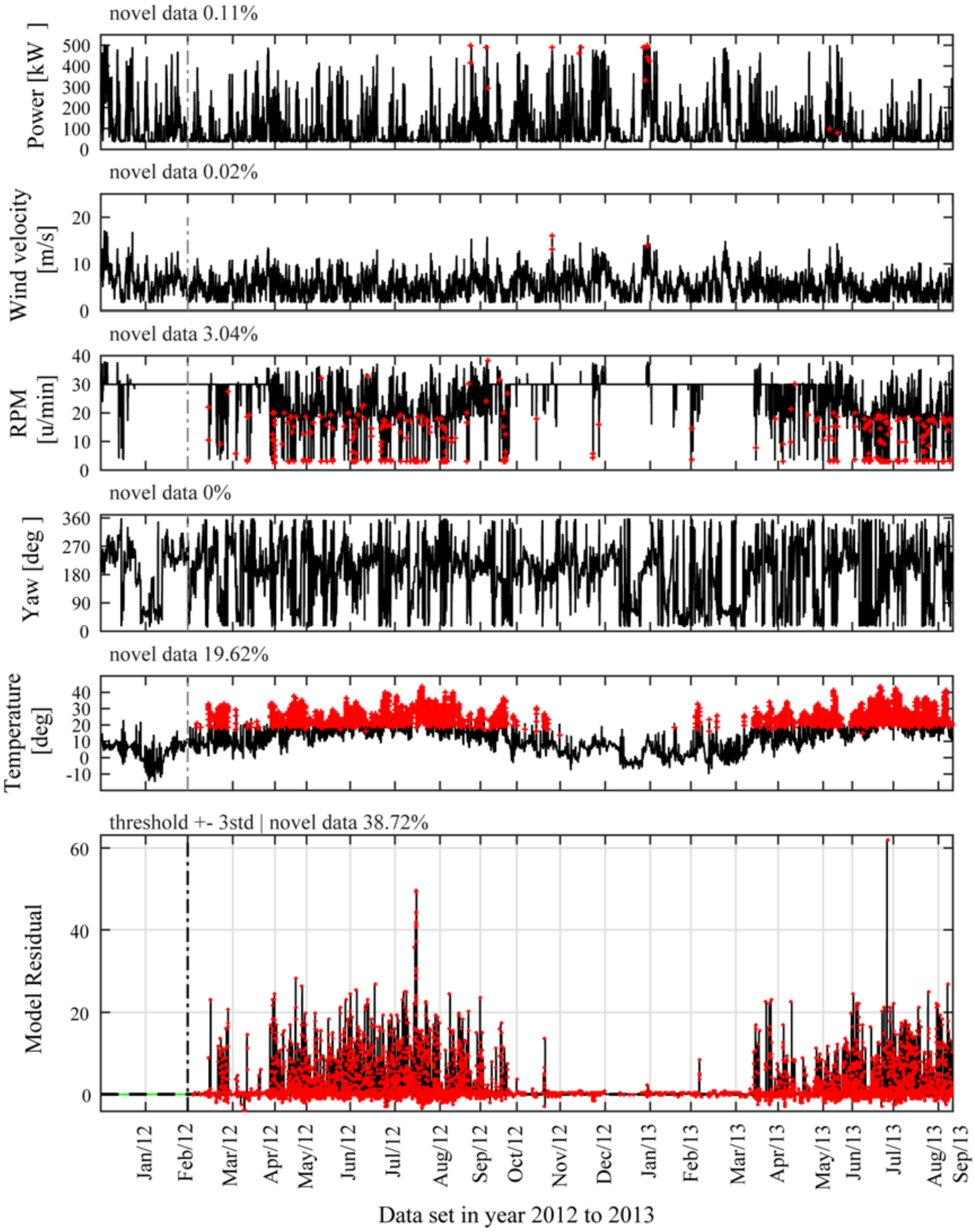


Figure 6. Two-month training set. Identified novel data (red points) within time history of 10-min mean values of measured SCADA and X- chart of the Model residual

With further increase of the training period and redefining the normal condition to include more points on fluctuating EOP (Figures 7-8), the MD outlier percentages decrease (SCADA variables with zero percent are not included in the plots) and correspondingly the DI becomes significantly reduced. In the case of the twelve- month period of training, the MD outlier percentages drop below 0.2% and the DI distribution pattern of the testing set is evidently improved, with substantially less points above the threshold values.

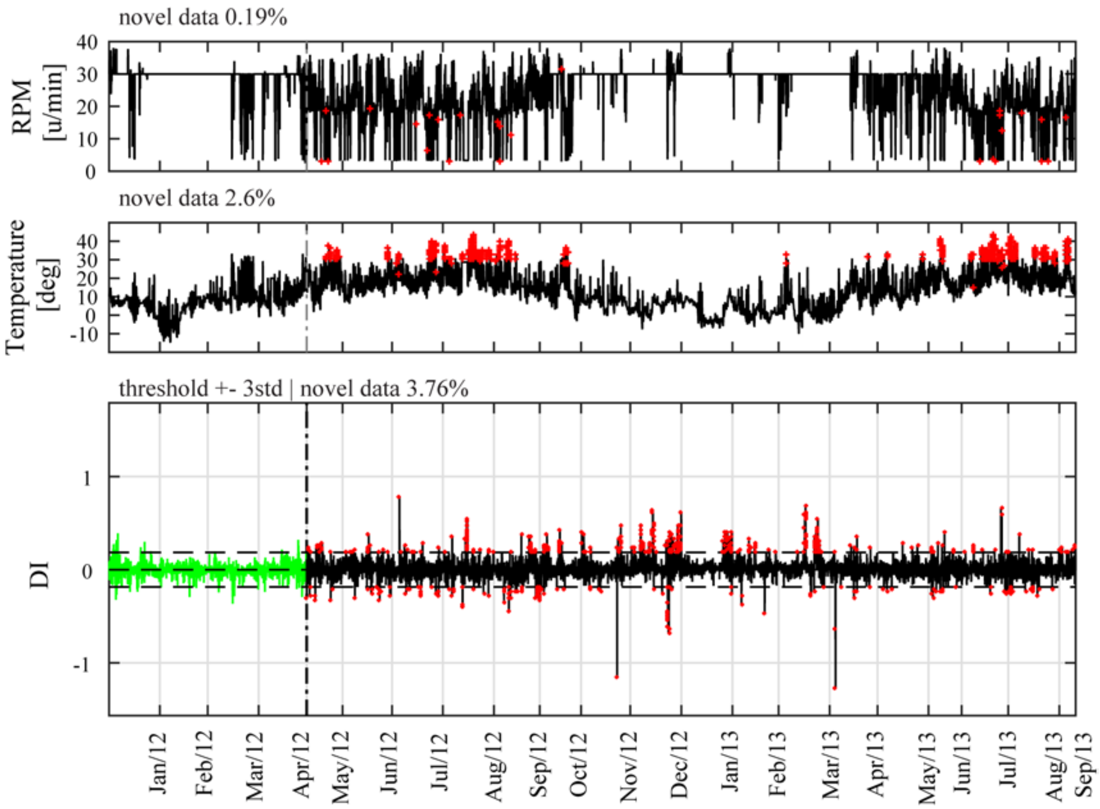


Figure 7. Four-month training set. Identified novel data (red points) within time history of 10-min mean values of measured SCADA and X- chart of the DI

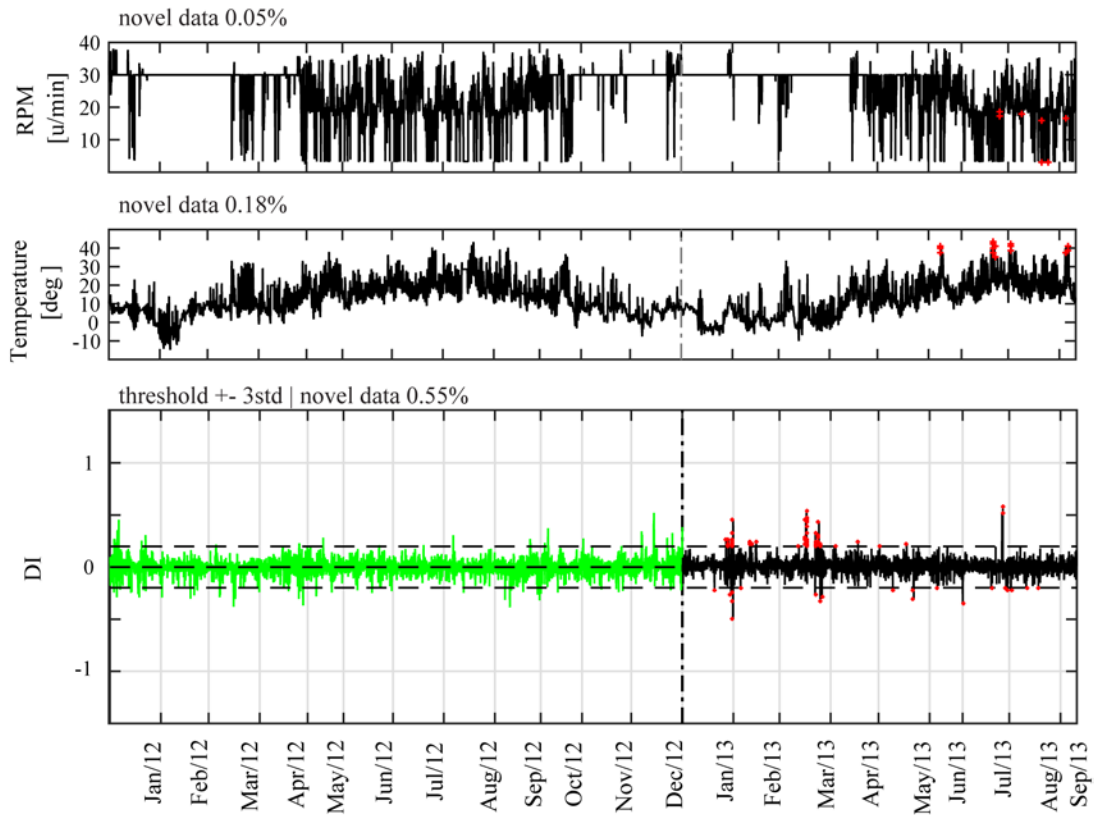


Figure 8. Twelve-month training set. Identified novel data (red points) within time history of 10-min mean values of measured SCADA and X- chart of the DI

Future work

In order to identify connections of specific patterns of structural behavior to relevant operating regimes of the WT system, future research will focus on the long-term tracking of the estimated PCE-SPTARMA diagnostic index. As preliminary presented in Fig. 9, simple statistical measures like the Kullback–Leibler divergence, applied daily on the obtained DI, demonstrate sensitivity to new data ranges of measured EOPs and agree well with the MD-based analysis of SCADA variables (Figs. 6-8).

However, in order to develop a holistic and computationally efficient tool capable of separating benign EOP fluctuations from indicator distortions due to actual structural damage or system malfunction, proper threshold tuning and pattern analysis is crucial. Towards this end, simulated damages will be introduced to the baseline training data of the monitored healthy structure. Finally, autonomous routines, based on robust outlier analysis or similar statistical measures, and an application- ready monitoring mapping for an appropriate timely reaction (model retraining or structural intervention), will be sought as a last step.

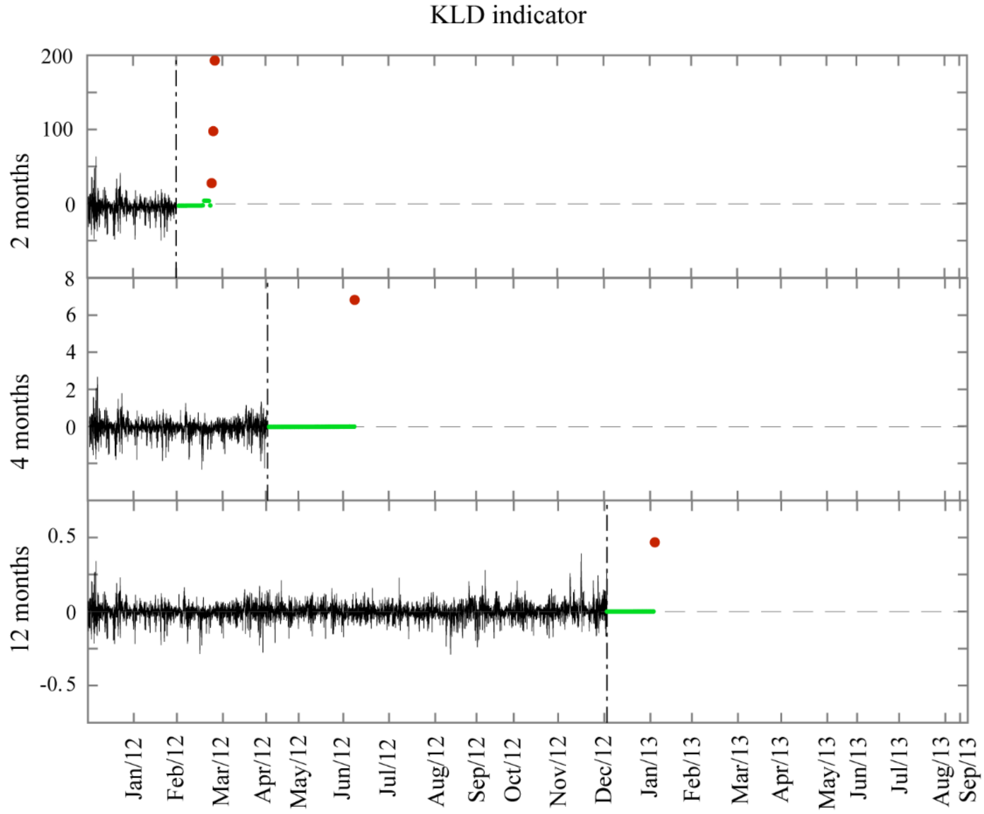


Figure 9. KL-Divergence indicator applied on the DI with cumulative assessment of one DI value per day, and tested for 2, 4, 12 months of training

Conclusions

The proposed strategy delivers a PCE-SPTARMA robust diagnostic index able to capture the non-stationary response and the long-term response variability of an actual operating WT structure for a monitoring period of twenty one months. The potential for further enhancements of the tool, towards real-time computing platform able to guide operators in the management of WT structures, is verified by outlier analysis of recorded SCADA data and preliminary utilization of statistical divergence measures on the obtained index.

References

[1] Iliopoulos, A., Devriendt, C., Guillaume, P. and Van Hemelrijck, D., Continuous fatigue assessment of an offshore wind turbine using a limited number of vibration sensors, *Proceedings of the 7th European Workshop on Structural Health Monitoring,* Nantes, France, 2014.

[2] Maes, K., Iliopoulos, A., Weijtjens, W., Devriendt, C. and Lombaert, G. (2016) Dynamic strain estimation for fatigue assessment of an offshore monopile wind turbine using filtering and modal expansion algorithms, *Mechanical Systems and Signal Processing* ***76–77***, 592–611.

[3] Van der Male, P. and Lourens, E., Operational vibration-based response estimation for offshore wind lattice structures, *Proceedings of the 33rd International Modal Analysis Conference*, Niezrecki, C., Ed., Orlando, Florida USA, 2015, Volume 7, 83–96.

[4] Tcherniak, D., Chauhan, S. and Hansen, M. H., Applicability limits of operational modal analysis to operational wind turbines, *Proceedings of 28th International Modal Analysis Conference*, Jacksonville, Florida USA, 2010.

[5] Ozbek, M., Meng, F. and Rixen, D. J. (2013) Challenges in testing and monitoring the in-operation vibration characteristics of wind turbines, *Mechanical Systems and Signal Processing* **41**, 649–666.

[6] Limongelli, M.P., Chatzi, E., Döhler, M., Lombaert, G. and Reynders, E., Towards extraction of vibration-based damage indicators, *Proceedings of 8th European Workshop On Structural Health Monitoring*, Bilbao, Spain, 2016.

[7] Weijtjens, W., Verbelen, T., De Sitter, G. and Devriendt, C. (2015) Foundation structural health monitoring of an offshore wind turbine- a full-scale case study, *Structural Health Monitoring* **15**, 389-402.

[8] Oliveira, G., *Vibration-based structural health monitoring of wind turbines*, PhD Thesis, University of Porto, Portugal, 2016.

[9] Hu, W. H., Thöns, S., Rohrmann, R. G., Said, S. and Rücker, W. (2015) Vibration-based structural health monitoring of a wind turbine system Part II: Environmental/operational effects on dynamic properties, *Engineering Structures* **89**, 273–290.

[10] Dervilis, N., Worden, K. and Cross, E. J. (2015) On robust regression analysis as a means of exploring environmental and operational conditions for SHM data, *Journal of Sound and Vibration* **347**, 279–296.

[11] Worden, K., Baldacchino, T., Rowson J. and Cross E., Some recent developments in SHM based on nonstationary time series analysis, *Proceedings of the IEEE* **104**, 1589-1603, 2016.

[12] Deraemaeker, A. and Worden, K. (2018) A comparison of linear approaches to filter out environmental effects in structural health monitoring, *Mechanical Systems and Signal Processing* **105**, 1–15.

[13] Martinez-Luengo, M., Kolios, A. and Wang, L. (2016) Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm, *Renewable and Sustainable Energy Reviews* **64**, 91–105.

[14] Spiridonakos, M. and Chatzi, E., Polynomial chaos expansion models for SHM under environmental variability, *Proceedings of 9th International Conference on Structural Dynamics*, Porto, Portugal, 2014.

[15] Spiridonakos, M., Chatzi, E. and Sudret, B. (2016) Polynomial chaos expansion models for the monitoring of structures under operational variability, *ASCE- ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* **2**.

[16] Avendaño-Valencia, L.D. and Chatzi, E., (2017) Sensitivity driven robust vibration-based damage diagnosis under uncertainty through hierarchical Bayes time-series representations, *Procedia Engineering* **199**, 1852-1857.

[17] Bogoevska, S., Spiridonakos, M., Chatzi, E., Dumova-Jovanoska, E. and Höffer, R. (2017) A Data-Driven Diagnostic Framework for Wind Turbine Structures: A Holistic Approach, *Sensors* **17**, 720.

[18] Spiridonakos, M. D., Poulimenos, A. G. and Fassois, S. D. (2009) Output-only identification and dynamic analysis of time-varying mechanical structures under random excitation: A comparative assessment of parametric methods, *Journal of Sound and Vibration* **329**, 768-785.

[19] Akaike, H. and Kitagawa, G., Eds. (1999) *The practice of time series analysis,* Chapter 11, Springer-Verlag, New York.

[20] Le Maître, O. P. and Knio, O. M. (2010) *Spectral Methods for Uncertainty Quantification*, Springer, Netherlands.

[21] Hyvärinen, A. and Oja, E. (2000) Independent component analysis: algorithms and applications, *Neural Networks* **13**, 411–430.

[22] Filzmoser, P., Garrett, R. G. and Reimann, C., *Multivariate outlier detection in exploration geochemistry*, Technical report TS 03-5, Department of Statistics, Vienna University of Technology, Austria, 2003.