Coherent and Compatible Statistical Models in Structural Analysis

***M. Nogal[1], E. Castillo[2], A. Calviño[2] and A. J. O'Connor[1]**

[1]Department of Civil Engineering, Trinity College Dublin, Dublin 2, Ireland.
[2] Department of Applied Mathematics and Computational Sciences, University of Cantabria, 39005 Santander, Spain.

*Corresponding author: nogalm@tcd.ie

Abstract

Modelling problems in structural analysis requires of a statistical approach that allows us to take into account the random nature of the variables as well as the uncertainties involved in the problem being analysed. However neither all statistical models are valid nor all assumptions are mathematically or physically reasonable. The aim of this paper is twofold: (a) to explain how to build statistical models with mathematical and physical coherence, and (b) to describe the most common mistakes made when building or selecting mathematical and statistical models. We provide some interesting tools to carry out this important task and present some examples that show the inconveniences and consequences derived from an incorrectly established model.

Keywords: Location-scale stable families, Structural analysis statistical models, Specification of multivariate joint distributions, Extreme Values, Probability papers.

Introduction

Before selecting a model to solve a given engineering problem, a very important step consists of dedicating sufficient time to study the problem under consideration in some depth. This means that the engineer must understand the problem, the variables and the implied physical relations, which should be present in the model. For example, an engineer dealing with a breakwater needs to understand that the large waves and winds are the most important agents implied in design. This means that maxima events and then maxima extreme value distributions must be considered. The limited or unlimited range of the random variables involved is also relevant, because this permits excluding either the Weibull or Frechet type of distributions. Ignoring these aspects leads to unconservative or very expensive solutions which are engineeringly regrettable.

It is also convenient to use simple models, that is, as parsimonious as possible and dimensionally consistent. In this line, the Buckingham theorem plays a fundamental role and should be the first step in equation modelling. Apart from reducing the number of variables involved and avoiding us to be concerned about dimensions, it permits us to check if the selected variables are sufficient or need to be completed with additional variables to reproduce a physical problem or phenomena.

Another important decision to be made when building models is the selection of the families of random variables used to reproduce the real ones. In this context, the designer must take into account the variable ranges and be aware that not all distribution families are valid for reproducing all types of variables. In this context, one should know that some distributions are valid only for dimensionless variables (Poisson, beta, binomial, etc.) and that some distributions are not scale (geometric, chi-squared, etc.) or location (gamma, log-normal, etc.) stable. For example, selecting non-scale families means that the resulting models will not be valid for variables when written in terms of different measure units, and then they are inadequate.

Since a statistical analysis requires the joint distribution of all variables involved, the selection of a multivariate model is crucial too. In this line it is important to use feasible models. We point out that in some cases a lack or an excess of simplifying assumptions can lead to undefined or inexistent models, respectively.

As a final example, the designers should be aware of the existence of different probabilistic papers (maximum, minimum, etc.) and that not all data points but only those in the tail of interest must be used when dealing with extremes. Ignoring these facts can be catastrophic.

Since we consider that all these issues are very relevant for engineers, they are discussed with some detail in this paper.

The aim of this paper is twofold. On one hand we introduce some considerations to be taken into account when building statistical models and, on the other hand, we point out some problems we can find when these aspects are not considered. Besides, we provide some tools to facilitate this task together with several examples for a better comprehension of the concepts discussed.

The paper is organized as follows. In Section 2 we present a brief review of some of the statistical models proposed in different Civil Engineering fields. In Section 3 we make some considerations about the units of the random variables and their moments. In Section 4 we emphasize the importance of the Buckingham theorem in order to build parsimonious and dimensionless models. In Section 5 we deal with extreme values and probability papers. In Section 6 we explain and discuss different possibilities to define multivariate models and finally, in Section 7 we give some conclusions.

Some statistical models proposed in the literature

In the Civil engineering literature it is becoming more frequent to find statistical approaches. For example, reliability analysis has reached all engineering fields. Due to the abundant bibliography dealing with this issue, as a sample and for illustration purposes, Table 1 shows a list of some examples of distributions used in the Civil Engineering literature.

Table 1: Some probability distribution families used in the literature together with the corresponding engineering variables.

VARIABLE	DISTRIBUTION	VARIABLE	DISTRIBUTION
Geometric and mechanical properties	log-normal and normal	Maximum wave height	reverse Weibull
Material properties	normal, two- and three-parameter Weibull	Two successive wave periods	bivariate Weibull and bivariate Rayleigh
Excedences of wave height or significant wave height	Generalized Pareto	Significant wave heights	Weibull, generalized gamma, generalized beta kind I and beta kind II
Stress range	Raleigh, wide-band, Weibull, beta, log-normal, Rice's and normal distributions	Significant wave height and wave period	Box-Cox + bivariate normal, bivariate log-normal and bivariate Plackett
Loads	Poisson, Gumbel and normal	Small wave heights in large depths	Rayleigh
Wind speed	Frechet, Gumbel, reverse Weibull and log-normal	Joint density of significant wave height, wave period and current and wind speeds	Marginals transformed to normals by Box-Cox transformation plus multivariate normal
Wave period	log-normal	Wave height	Rayleigh distribution and reverse three-parameter Weibull
Fatigue life	Weibull		

It is relevant to say that some of the used models above are theoretically justified and some are used just for convenience or to facilitate calculations or mathematical derivations. For example, the normal model is justified when the random variable being modelled is the sum of a large enough number of other variables. This occurs frequently in strength of materials where in a cross section of volume all the subelements add efforts or collaborate to resistance. Poisson and gamma distributions have been proved to correspond to rare events and the time of occurrence of the *r*-th event, respectively. The Weibull, Gumbel and Frechet extreme value distributions and their reverse versions are justified because they are the limit distributions of maxima or minima, which are very important in Civil Engineering design because in general only maxima (waves, winds, earthquakes, temperatures, etc.) and minima (draughts, fatigue strength, temperatures, etc.) values lead to failure. However, it is not uncommon to see minima models erroneously used for maxima or vice versa. The generalized Pareto distribution is justified because it arises as a limit distribution for exceedances (large waves, winds, etc.) over or shortfalls (rain, temperature, etc.) under a threshold. Rice and Rayleigh distributions are also derived from theoretical models of waves.

Contrary, other distributions, such as the log-normal that arises in order to reproduce asymmetric data, the generalized beta and the models based on the Box-Cox transformation that are used to fit different data histograms, etc. have convenience as motivation.

In the Structures field, for example, [O'Connor and Kenshel (2013)] use the normal distribution to describe concrete material properties, [O'Connor and Enevoldsen (2009)] propose Log-normal distributions for modelling structural parameters and uncertainties associated with modelling, [Simiu et al. (1980)] assume the Fretchet distribution for the wind speed and [Pourzeynali and Datta (2005)] suggests the Raleigh distribution to model the stress range.

In the Material Science field, [Castillo and Fernández-Canteli (2009)] develop a fatigue model using a three-parameter Weibull distribution for a normalizing variable representing the whole S-N field based on a unique distribution function, [Koller et al. (2009)] validate the use of a log-Gumbel fatigue regression model and [Przybilla et al. (2011)] propose a method to obtain the distribution of fracture stress as a three-parameter Weibull cumulative distribution function (cdf) referred to a uniaxially and uniformly tensioned surface element. We can also mention the case of Coast and Ocean Engineering where [Ferreira and Guedes Soares (1999)] assume significant wave heights to follow Beta distributions, [Ferreira and Guedes Soares (1998)] use the Generalized Pareto densities for excedences of wave heights and significant wave heights, or [Ochi (1992)] proposes the Generalized Gamma distribution for significant wave heights.

Another field with a wide variety of stochastic models is Transportation. Some examples are [Lo et al. (1996)] who propose independent Poisson link counts or [Castillo et al. (2012)] who develop a bayesian network considering that the different traffic variables follow a generalized beta distribution. Multinomial models were assumed by [Clark and Watling (2005)] for route flows and shifted-gamma distribution was used by [Castillo et al. (2013)] for modelling the traffic flows.

From the list of publications above we can realize that a large set of distributions has been used. Detected inconsistencies in some of the proposed models motivates the current paper, which presents essential aspects to be considered when building statistical models.

Some considerations on units of the random variable and their moments

One common mistake when building statistical models is to ignore that not all families of distributions are valid for all types of variables. We need to be aware that parameters of statistical families have units. In particular, the mean has the same dimension as the random variable and the variance the squared dimensions.

Example 3.1 (Exponential distribution) *For the exponential distribution Exp(λ) we have:*

$$E[X]=1/\lambda; \qquad Var[X]=1/\lambda^2. \qquad (1)$$

Since the dimension of $1/\lambda^2$ *is the square of the dimension of 1/λ, the inverse of the variable unit, the dimensions are consistent in this case.*

Example 3.2 (Beta distribution) *For the beta distribution, X~Beta(a,b) we have:*

$$E[X]=\frac{a}{a+b}; \qquad Var[X]=\frac{ab}{(a+b)^2(a+b+1)}. \qquad (2)$$

This implies that X must be dimensionless, because in the term a+b+1 a and b must be dimensionless; otherwise they cannot be added to 1 (dimensionless). Once that a and b have been recognized as dimensionless, E[X] and Var[X] are also dimensionless (see (2)).

Example 3.3 (Weibull distribution) *For the Weibull distribution, X~W(λ,k) we have:*

$$E[X]=\lambda\Gamma\left(1+\frac{1}{k}\right); \quad Var[X]=\lambda^2\left[\Gamma\left(1+\frac{2}{k}\right)-\Gamma^2\left(1+\frac{1}{k}\right)\right], \qquad (3)$$

which implies that k must be dimensionless and λ must have the same dimensions as X, and that the Weibull model can be made consistent for variables of any dimensions.

Example 3.4 (Gamma distribution) *If the random variable X is Gamma X~W(λ,k), the random variable X+a with a≠0 is not gamma any more. This means that the gamma family is not stable with respect to changes in location and has important consequences, because the gamma distribution cannot be used for location variables, such as temperatures. More precisely, if a random temperature is gamma measured in Celsius degrees, it is not gamma when measured in Farenheit or Reamur degrees. Thus, using the gamma family for temperatures is inadequate and misleading.*

Other examples of dimensionless families are the binomial, negative binomial and Poisson. Contrary, normal distributions are examples of statistical families compatible with any dimension.

Parsimonious and dimensionless models: The Buckingham theorem

When a mathematical or statistical model is built, a dimensional analysis of the variables involved must be initially carry out as this allows us to understand some deep relations among these variables and help to avoid dimensional contradictions. Besides, it is recommendable to build a dimensionless model in order to prevent dimensional inconsistencies and in some cases to reduce the problems associated with precision in numerical evaluations. Finally, it is important to work with parsimonious models, that is, the simplest models explaining all the aspects to be considered. To these aims the Rayleigh method of dimensional analysis and its formalization proposed by [Buckingham (1915)] plays a fundamental role. To illustrate, we propose the following example.

Example 4.1 (Corbel Example. Dimensionless variables) *The example deals with a reliability analysis of a corbel by means of the strut-and-tie model represented in Figure 1. In this case we assume two possible failure modes, defined by the limit-state functions* H_1 *and* H_2*:*

$$H_1 \equiv f_s A_s - F_v \tan\theta - F_h = 0, \qquad (4)$$

$$H_2 \equiv f_c A_b - F_v = 0 \qquad (5)$$

where F_v *and* F_h *are the applied vertical and horizontal forces, respectively,* f_s *and* f_c *are the strength of the steel and of the compressed concrete,* θ *is the angle between the compression strut and the tie,* A_s *is the cross sectional area of the passive reinforcement and* A_b *is the area where the action is applied.*

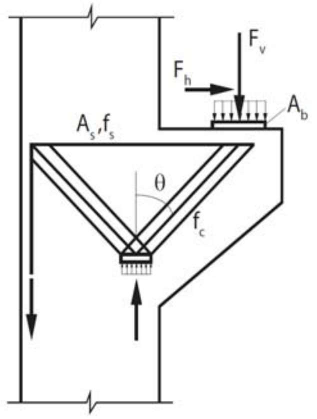

Figure 1: Corbel Example. Strut-and-tie model.

The failure curve can be expressed as the minimum value of the previous limit-state functions, that is,

$$H = min\ \{H_1, H_2\} \tag{6}$$

Based on the Buckingham Π Theorem, we get the dimensional decomposition shown in Table 2, where [F] and [L] denote force and length magnitudes, respectively. We see that the n=7 variables set up a dimensional matrix with rank q=2. Applying the Buckingham Π Theorem, we conclude that the model (6) is equivalent to another with p=n−q=5 dimensionless parameters (ratios).

Table 2: Corbel Example. Dimensional decomposition.

	F_v	F_h	f_s	f_c	θ	A_s	A_b
[F]	1	1	1	1	0	0	0
[L]	0	0	-2	-2	0	2	2

If we use f_s *and* A_s *as reference or normalizing variables, we obtain the following new dimensionless variables:*

$$F_v^* = \frac{F_v}{f_s A_s}; \quad F_h^* = \frac{F_h}{f_s A_s}; \quad f_c^* = \frac{f_c}{f_s}; \quad \theta^* = \theta; \quad A_b^* = \frac{A_b}{A_s}, \tag{7}$$

and the new mathematical expression for the model (6) becomes:

$$H^* = \frac{H}{f_s A_s} = min\left\{1 - F_v^* \tan\theta^* - F_h^*,\ f_c^* A_b^* - F_v^*\right\}, \tag{8}$$

where the asterisks refer to dimensionless variables. The main advantages of using the Buckingham theorem are:

1. The model presents p=5 variables instead of n=7, which implies a reduction in the problem complexity.

2. The variables are independent from any units being considered, avoiding possible dimensional mistakes. Moreover, the normalization modifies the variable ranges and reduces possible numerical precision problems.

3. The variable H^* becomes more meaningful than the associated dimensional variable H because its value can be compared for different cases when the steel characteristics (area and yield strength) are kept constant.

We point out that the dimensional results can be recovered at the end of the process, undoing the change proposed in (7).

Extreme values and probability papers

In the engineering design of structures we need to deal with extreme values, that is, maxima (for example, loads, moments, etc.) or minima (for example, strength properties). In such cases, a careful selection of extreme value distributions to approximate the distribution of extremes is required. In this paper we deal only with maxima, but the minimum problem is similar. In order to see if a cdf $F(x)$ can be approximated for maxima by a reverse Weibull, Gumbel or Frechet distribution we can use the following theorem by [Castillo (1988)].

Theorem 1 *If F(x) is the cumulative distribution function of a random variable and*

$$\lim_{\varepsilon \to 0} \frac{F^{-1}(1-\varepsilon)-F^{-1}(1-2\varepsilon)}{F^{-1}(1-2\varepsilon)-F^{-1}(1-4\varepsilon)} = 2^{c.} \tag{9}$$

then F(x) can be approximated in its right tail by a Frechet distribution if c>0, a Gumbel distribution if c=0 and a Weibull distribution if c<0.

In particular, if the range of $F(x)$ is limited it cannot be approximated by a Frechet distribution and if it is unlimited, we cannot use a Weibull distribution.

As some interesting examples, Table 3 shows the corresponding approximating distributions of some of the most common distributions for maxima and minima.

Table 3: Corresponding approximating distributions for maxima and minima of the most common distributions.

Distribution	Domain of Attraction	
	Maximal	Minimal
Normal	Gumbel	Gumbel
Exponential	Gumbel	Weibull
Log-normal	Gumbel	Gumbel
Gamma	Gumbel	Weibull
Gumbel	Gumbel	Gumbel
Gumbel	Gumbel	Gumbel
Rayleigh	Gumbel	Weibull
M = maxima m = minima		

Distribution	Domain of Attraction	
	Maximal	Minimal
Uniform	Weibull	Weibull
Weibull	Weibull	Gumbel
Weibull	Gumbel	Weibull
Cauchy	Fréchet	Fréchet
Pareto	Fréchet	Weibull
Fréchet	Fréchet	Gumbel
Fréchet	Gumbel	Fréchet
M = maxima m = minima		

The previous method permits determining the extreme value distributions associated with a given one $F(x)$. However, in practice we do not have this information but only data. In this case we can plot this data on a Maximal Gumbel probability paper, as shown in Figure 2. Then, looking to its right tail and determining whether the data trend is straight or has positive or negative curvature, we can decide about Gumbel, Weibull or Frechet as approximating distributions, respectively.

Building multivariate statistical models

In this section we deal with the problem of defining the joint multivariate density of all the variables which are relevant to the problem under consideration.

There are several ways to define the joint density of a multivariate model. These methods can be classified as underdetermined, overdetermined and uniquely determined methods, depending of the number of imposed conditions.

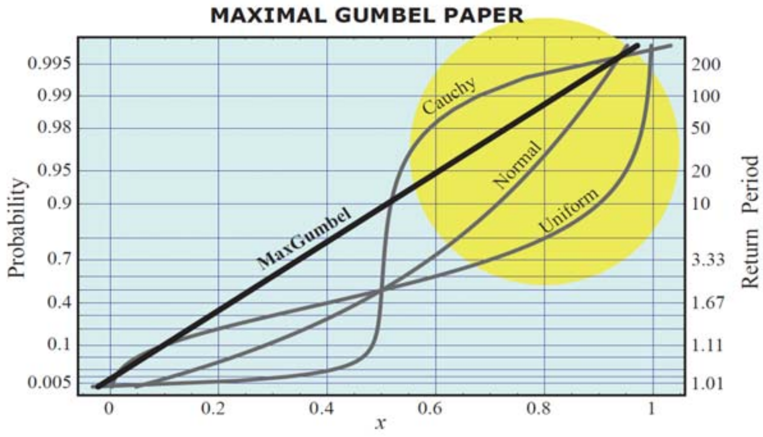

Figure 2: Maximal Gumbel probability paper illustrating the relevant zone.

In order to uniquely determine a multivariate model with an underdetermined method, we have to add some extra conditions. In the case of the overdetermined methods, the solution is not guarantied. For a more detailed description of these methods, see [Arnold et al. (1992,1999,2001)] and [Castillo et al (2014)].

The following example illustrates the cases of overdetermined and underdetermined methods.

Example 6.1 (Normal conditionals model) *[Arnold et al (1999)] demonstrate that there are two families of bivariate distributions with normal conditionals, that is, with conditionals $X|Y{=}y$ and $Y|X{=}x$ which are normals: (a) the normal and (b) a family with regression lines and conditional variances given by:*

$$E(X|Y=y) \;=\; \mu_1(y) = -\frac{m_{12}y^2+m_{11}y+m_{10}}{2(m_{22}y^2+m_{21}y+m_{20})}, \tag{10}$$

$$var(X|Y=y) \;=\; \sigma_1^2(y) = \frac{-1}{2(m_{22}y^2+m_{21}y+m_{20})}, \tag{11}$$

$$E(Y|X=x) \;=\; \mu_2(x) = -\frac{m_{21}x^2+m_{11}x+m_{01}}{2(m_{22}x^2+m_{12}x+m_{02})}, \tag{12}$$

$$var(Y|X=x) \;=\; \sigma_2^2(x) = \frac{-1}{2(m_{22}x^2+m_{12}x+m_{02})}, \tag{13}$$

where the m's are constants.

One example of a normal density is shown in the left plot of Figure 3, where the linear regression lines are shown on the top projection and the normal marginals in the left and right projections. Similarly, the right plot corresponds to a non-normal family, which shows projected non-linear regression lines and non-normal marginals.

If we assume normal conditionals alone, the resulting model is undefined, but if in addition we assume that the $X|Y$ regression line is a proper third degree polynomial, we are in front of an inexistent or impossible model as we can conclude from Equation (10).

The simplest method to define a joint density corresponds to the independent model in which all variables are independent, so that it is sufficient to define the univariate marginals. However, when variables are dependent, the model complicates.

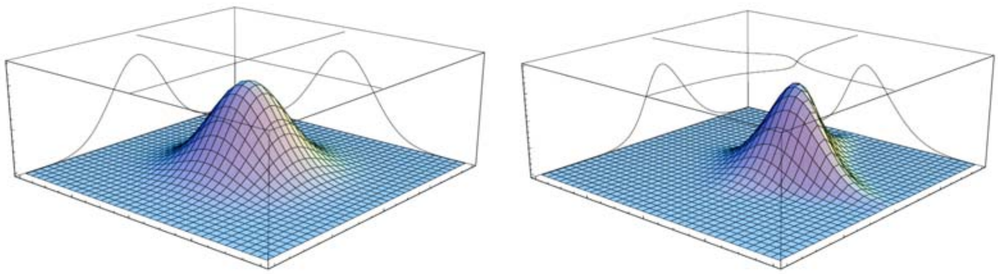

Figure 3: Two illustrative examples of bivariate densities with normal conditionals: normal (left figure) and non-normal (right figure).

Example 6.2 (Corbel Example. Selecting probability distribution families) *In this example we select and discuss the probability distribution families associated with the variables involved in the corbel example introduced previously. To simplify and because the steel manufacture companies are very reliable, we can assume f_s and A_s as deterministic. Thus, we only have the random variables F_v, F_h, f_c, A_b and θ. In addition we can assume that all variables are independent. With the exception of F_v and F_h, this is a reasonable assumption because they involve forces, material strengths, areas and a design angle, whose values are undoubtedly independent. Furthermore, we assume the independence of F_v and F_h. This implies that only marginal distributions are needed in order to build the statistical multidimensional model.*

Table 4 shows the selected marginal distributions, the associated parameters and the corresponding ranges.

Table 4: Corbel Example. Marginal distribution functions for the dimension variables.

Dimension Variable	*Variable Type*	*Lower Bound*	*Upper Bound*	*Assumed Distribution*	*Assumed Parameters*		
					Shape	*Scale*	*Location*
f_s	*Deterministic*	*455124 kN/m²*	*455124 kN/m²*				
A_s	*Deterministic*	*5.92e-4 m²*	*5.92e-4 m²*				
F_v	*Random: Extreme Value*	*positive value*	*positive value*	*3P Max- Weibull*	*0.21*	*36.209 kN*	*68.846 kN*
F_h	*Random: Extreme Value*	*negative value*	*positive value*	*3P Max- Weibull*	*0.236*	*9.776 kN*	*−4.225kN*
f_c	*Random: General*	*positive value*	*positive value*	*Gamma*	*149.50*	*142.5 kN/m²*	
A_b	*Random: General*	*positive value*	*positive value*	*Gamma*	*1.45*	*0.031 m²*	
θ	*Random: General*	$\pi/6$	$\pi/3$	*Generalized Beta*	*α=2 β=15*	$\pi/6$	*−1*

Now, given these selected distributions, we can obtain the distributions for the associated dimensionless problem. With this aim, first we represent A_s and f_s by their expected values, i.e., $A_s \equiv E[A_s] = \mu_{A_s}$ and $f_s \equiv E[f_s] = \mu_{f_s}$, obtaining the following new dimensionless variables:

$$F_v^* = \frac{F_v}{\mu_{f_s}\mu_{A_s}}; \quad F_h^* = \frac{F_h}{\mu_{f_s}\mu_{A_s}}; \quad f_c^* = \frac{f_c}{\mu_{f_s}}; \quad A_b^* = \frac{A_b}{\mu_{A_s}},$$

and the limit-state function:

$$H^{*} = \frac{H}{\mu_{f_s}\mu_{A_s}} = min\left\{1-F_v^{*}\tan\theta-F_h^{*},\quad f_c^{*}A_b^{*}-F_v^{*}\right\}. \tag{14}$$

For the new non-dimensional variables appearing in equation (14), Table 5 shows the corresponding distributions and associated dimensionless parameters. The values have been obtained using $\mu_{f_s}=455124\,kN/m^2$ *and* $\mu_{A_s}=5.92\text{e-}4\,m^2$.

Table 5: Corbel Example. Marginal distribution functions for the dimensionless variables.

Dimensionless Variable	*Assumed Distribution*	*Assumed Parameters*		
		Shape	*Scale*	*Location*
F_v^{*}	*3P Maximum Weibull*	*0.21*	*0.134*	*0.256*
F_h^{*}	*3P Maximum Weibull*	*0.236*	*0.036*	*−0.016*
f_c^{*}	*Gamma*	*149.5*	*3.13e−04*	
f_s	*Gamma*	*1.45*	*52.365*	
θ^{*}	*Generalized Beta*	*α=2, β=15*	$\pi/6$	*−1*

Table 5 shows that only the scale and location parameters are affected by the normalization. Moreover, the parameters of the Generalized Beta distribution remain constant because they are associated with the dimensionless variable $\theta^{*}=\theta$ *. Finally, the statistical families in Table 4 remain in Table 5 because all of them are stable with respect to scale changes.*

However, there exists another way to deal with the dimensionless problem without using scale-stable distributions. The process consists of obtaining the dimensionless sample data before fitting the distribution parameters.

One of the most important methods to define dependent multivariate models is Bayesian networks, which are defined by means of a directed acyclic graph *G* together with the conditional distributions of each of the involved variables given their parents, as follows:

$$f(x_1;x_2;\ldots;x_n)=f_1(x_1)f_2(x_2|x_1)f_3(x_3|x_1;x_2)\ldots f_n(x_n|x_1;x_2;\ldots;x_{n-1})=\prod_{i=1}^{n} f_i(x_i|\pi_i), \tag{15}$$

where π_i are the parents of the variable X_i in the directed acyclic graph *G*. Bayesian networks are the simplest way to reproduce complicated multidimensional families of distributions avoiding incompatibilities.

Example 6.3 (Corbel Example. A multivariate model) *In this example we determine a multivariate model associated to the previously dimensionless corbel example.*

From equations (4) and (5) we know that

$$H_1^{*} = f_1\left(F_h^{*}, F_v^{*}, f_s^{*}, A_s^{*}, \theta\right) \tag{16}$$

$$H_2^{*} = f_2\left(F_v^{*}, f_c^{*}, A_b^{*}\right), \tag{17}$$

where H_1^{*} *and* H_2^{*} *are the dimensionless limit-state functions, using* μ_{A_s} *and* μ_{f_s}.

In order to correctly define the joint distribution of these limit-state functions, we have to carry out an analyses of the dependence relation among the involved variables. In this example, we assume independence among all variables, except between F_v and F_h, because these forces are usually related.

Due to the fact that variables involved in H_2^* are independent we can compute the joint probability by means of the set of all marginal, that is,

$$f(H_2^*) = f(F_v^*, f_c^*, A_b^*) = f(F_v^*) f(f_c^*) f(A_b^*), \tag{18}$$

However, in the case of the , H_1^* applying equation (15) to determine this joint probability, we get:

$$\begin{aligned} f(H_1^*) &= f(F_v^*, F_h^*, f_s^*, A_s^*, \theta) = \\ &= f(F_h^*) f(F_v^*|F_h^*) f(f_s^*|F_h^*, F_v^*) f(A_s^*|F_h^*, F_v^*, f_s^*) f(\theta|F_v^*, F_h^*, f_s^*, A_s^*) = \\ &= f(F_h^*) f(F_v^*|F_h^*) f(f_s^*) f(A_s^*) f(\theta), \end{aligned} \tag{19}$$

which requires to know the conditional distribution of F_v^* given F_h^* .

With this aim, we represent the data (F_h^*, F_v^*) (see left Figure 4) and observe that they exhibit the following linear regression:

$$F_v^* = 3.99\, F_h^* + 0.31\,. \tag{20}$$

Next, we find that the residuals follow a maximal Weibull model (see right Figure 4):

$$F_R(r) = \exp\left\{-\left[1-k\left(\frac{r-\lambda}{\sigma}\right)\right]^{1/k}\right\}; \quad 1-k\left(\frac{r-\lambda}{\sigma}\right) \geqslant 0. \tag{21}$$

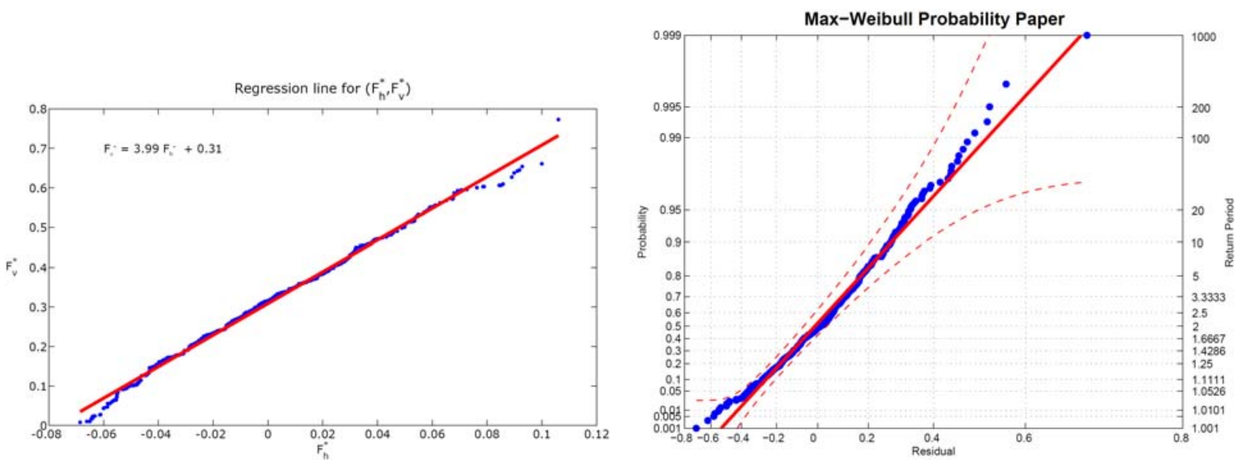

Figure 4: Data and regression line for the Corbel example and residuals given on a Normal probability plot.

Combining this expression with the regression equation (20) leads to the final model for $F_v^*|F_h^*$

$$F_{F_v^*|F_h^*}(f_v^*|f_h^*) = \exp\left\{-\left[1-k\left(\frac{F_v^* - 3.99\, F_h^* - 0.31 - \lambda}{\sigma}\right)\right]^{1/k}\right\} \tag{22}$$

only valid for

$$1-k\left(\frac{F_v^* - 3.99\, F_h^* - 0.31 - \lambda}{\sigma}\right) \geqslant 0 \tag{23}$$

Then, the estimation of the Weibull parameters using the maximum likelihood method leads to

$$k=0.262; \qquad \sigma=0.218; \qquad \mu=-0.0789.$$

In this way, the joint probability of this multivariate model becomes defined and we can evaluate the failure probabilities.

Conclusions

The following conclusions can be drawn from this paper:

1. Random variables and the parameters of statistical distributions are dimensional. These must be taken into consideration when statistical models are selected, otherwise, inadequate models can be obtained leading to important dimensional problems.

2. A previous dimensional analysis of the variables involved must be performed before building a model. This leads to a deep understanding of the relations among the involved variables, avoids dimensional inconsistencies and reduces numerical precision problems. In this direction, the Π Buckingham theorem is the most convenient and recommendable tool to be used.

3. Identification of the adequate extreme value distribution is very important in real practice. There are theorems that allow us to decide which of the Weibull, Gumbel or Frechet distributions or their reverse versions corresponds to a given cdf $F(x)$.

4. We must be aware of the fact that different probability papers exists. With respect to extreme value analysis there are two Gumbel probability papers, one for maxima and one for minima. It is important to realize that only the tail of interest must be plotted and fitted.

5. Care must be taken in selecting the adequate multivariate joint density functions. In this line, we must be aware that an excess of assumptions leads to impossible models, and a lack of them, to undefined models. Finally, Bayesian networks is the most adequate method to define the joint distributions, based on a directed acyclic graph and the conditional distributions of each of the random variable given their parents.

Acknowledgements

The authors are indebted to the DURATINET Project (Project No. 2008-1/049) and the University of Cantabria (Puente Project) for partial support of this work.

References

O'Connor, A. J. and Kenshel, O. Experimental evaluation of the scale of fluctuation for spatial variability modeling of chloride-induced reinforced concrete corrosion, Journal Of Bridge Engineering 18(1) (2013) 3–14.

O'Connor, A. J. andEnevoldsen, I. Probability-based assessment of highway bridges according to the new danish guideline, Structure and Infrastructure Engineering 5(2) (2009) 157–168.

Simiu, E., Changery, M. and Filliben, J. J. Extreme wind speeds at 129 stations in the contiguous united states, Journal of the structural division-ASCE 106(4) (1980) 809–817.

Pourzeynali, S. and Datta, T. Reliability analysis of suspension bridges against fatigue failure from the gusting of wind, Journal of Bridge Engineering 10(3) (2005) 262–271.

Castillo E. and Fernández-Canteli, A. A Unified Statistical Methodology for Modelling Fatigue Damage, Springer, Netherlands, 2009.

Koller, R., Ruiz-Ripoll M. L., García A., Fernández-Canteli A. and Castillo, E., Experimental validation of a statistical model for the wöhler field corresponding to any stress level and amplitude, International Journal of Fatigue 31(1) (2009) 231–241.

Przybilla, C., Fernández-Canteli, A. and Castillo, E. Deriving the primary cumulative distribution function of fracture stress for brittle materials from 3- and 4-point bending tests, Journal of the European Ceramic Society 31 (2011) 451–60.

Ferreira, J. A., Guedes Soares, C. Modelling the long-term distribution of significant wave height with the Beta and Gamma models, Ocean Engineering 26(8) (1999) 713–725.

Ferreira, J., and Guedes Soares, C. An application of the peaks over threshold method to predict extremes of significant wave height, Journal of Offshore Mechanics and Arctic Engineering 120 (3) (1998) 165–176.
Ochi, M. New approach in estimating the severest sea state from statistical data, in: Proceedings on the 23rd Coastal Engineering Conference, ASCE, New York, 1992, pp. 512—523.
Lo, H. P., Zhang, N. and Lam, W.Estimation of an origin-destination matrix with random link choice proportions: a statistical approach., Transportation Research, Part B 30:4 (1996) 309–324.
Castillo, E., Nogal, M., Menéndez, J., Sánchez-Cambronero, S. and Jiménez, P. Stochastic demand dynamic traffic models using generalized beta-gaussian Bayesian networks, IEEE Transactions on Intelligent Transportation Systems 13 (2) (2012) 565–581.
Clark, S. andWatling, D. Modelling network travel time reliability under stochastic demand, European Journal of Operational Research 39 (2005) 119–140.
Castillo, E., Jiménez, P., Menéndez, J. and Nogal, M. A Bayesian method for estimating traffic flows based on plate scanning, Transportation 40 (2013) 173–201.
Buckingham, E. The principle of similitude, Nature 96 (1915) 396–397.
Castillo, E. Extreme Value Theory in Engineering, Academic Press, New York, 1988.
Arnold, B. C., Castillo, E. and Sarabia, J. M. Conditionally specified distributions, Springer-Verlag, 1992.
Arnold, B. C., Castillo, E. and Sarabia, J. M. Conditional specification of statistical models, Springer, 1999.
Arnold, B. C., Castillo, E. and Sarabia, J. M. Conditionally specified distributions: an introduction (with comments and a rejoinder by the authors), Statistical Science 16 (3) (2001) 249–274.
Castillo, E., O'Connor, A. J., Nogal, M. and Calviño, A. On the physical and probabilistic consistency of civil engineering random models, Structural Safety (2014) Submitted.