UNIVERSITÀ DEGLI STUDI DI PALERMO

Meccanica dei Materiali e delle Strutture
Vol. VI (2016), no.1, pp. 235-242
ISSN: 2035-679X
Dipartimento di Ingegneria Civile, Ambientale, Aerospaziale, Dei Materiali
DICAM

BAYES ESTIMATION OF EXTREME WIND LOADS BASED UPON A POISSON-PARETO MODEL

E. Chiodo*, F. De Angelis*, D. Lauria*

* Università di Napoli Federico II, via Claudio, 21, 80125 Napoli, Italy
e-mail: fabio.deangelis@unina.it

Key words: Bayes estimation; Extreme value theory; Pareto distribution; Poisson processes; Reliability of structures; Wind loads.

Abstract. *Bayes estimation of extreme load values in the framework of risk and reliability analysis is investigated. The evaluation of extreme values of wind loadings on structures is performed via a combined employment of a Poisson process model for the "peak–over–threshold" characterization and a Pareto distribution for modelling the so-called "parent distribution" which generates the base load values. The method is applied to the extreme wind speed, both for sake of identification and estimation purpose. This topic has indeed brought about an increasing number of studies in the last years, both for wind energy production assessment and also in risk and reliability analysis. This modeling is difficult due to the uncertainty in wind speed probability distributions. For this purpose, the paper proposes a novel Bayes approach for the estimation of the probability that wind speed is lower than a prefixed extreme value. A large set of numerical simulations are performed in the last part of the paper, in order to illustrate the feasibility and efficiency of the above estimation method, especially when compared to the classical Maximum Likelihood method.*

1 INTRODUCTION

The recent advances in wind engineering motivated a large amount of studies focused on wind speed (WS) statistical distribution. At this purpose, a significant interest has been paid to extreme values (EV) characterization of WS, both for assessing the maximum wind energy production[1-4] and for evaluating risk, safety or reliability[5-8]. It has to be highlighted that the meaning of EV is herein conceived as class of p-quantiles corresponding to high values of p (such as the 0.95 or 0.99-quantiles), or extreme values of WS over an assigned time horizon, as it will be discussed in detail in Sec. III. For sake of clarity, it is recalled that the p-quantile is a value x_p of a random variable (RV) X is such that p100% of the observed value of the RV fall below x_p.

It has to be outlined that extreme quantiles are generally evaluated in the "static" case, i.e. without explicitly taking into account time, as typically performed in most WS studies devoted to wind energy characterization. This means that random variables are used, instead of stochastic processes as would be more realistic. In the following a dynamic approach is pursued in terms of stochastic process.

The forecast of extreme wind speed values, or "wind gusts", plays a crucial role in environmental studies and engineering risk and reliability analysis, to which the present paper mainly refers: it is known, indeed, that wind turbines are designed to be "cut out", in case of

intolerable high wind speeds, as a mean of protection against possible damages. It is not trivial to put in evidence that, with reference to the aspect of energy production, - by keeping in mind the “cubic rule” relationship between the wind power and the wind speed - the extreme upper quantiles of wind power are very sensitive to the correspondent quantiles of wind speed, so that an inaccurate quantiles estimation may involve relatively large errors in the evaluation of the expected wind energy production.
However, it has been emphasized (not only in the specific literature of wind energy investigations, but also in the general statistical literature about extreme values) that the EV estimation is a complex task, since large sample sizes are requested: in the case of not sufficient sample sizes, many models can be employed, from the classical Weibull distribution to the more recent Log-logistic, Lomax or Burr distribution, which generally perform quite similarly in the “central” part of the real WS distribution, i.e. with about the same values of “central parameters” such as the mean and the median values. At the aim of overcoming this "model uncertainty", the paper proposes a Bayes approach for the estimation of the EV probability distribution, which may be suitable under various models, based upon the characterization of extreme WS by means of a proper Poisson process of "exceedances", following a methodology introduced in extreme value theory of stochastic processes, and also applied in[4] within the framework of safety studies. In practice, this methodology can be regarded as strictly related to the so called "peaks-over-threshold" (POT) method[3], which is based upon the stochastic process of the time instants in which the wind-speed exceeds a given threshold.

2 WIND SPEED EXTREME VALUES EVALUATION BY MEANS OF STOCHASTIC PROCESSES

As already remarked in the previous section, in the following the dynamic approach is adopted. Indeed, for guaranteeing the required tower safety margins, being towers built for expected operation lifetimes of many years, the designers have to estimate the *extreme values* of the stochastic process of wind speed, i.e. the maximum "wind gust" amplitude over a prefixed time horizon. Herein and in the sequel, we will denote simply as "gust" the wind gust amplitude over a given time interval, or an extreme value in the sense below explained.
Firstly, let us denote by W=W(t) the stochastic process of WS values over time, where w* is a sufficiently high value (a “threshold” value) of WS such as that every value of WS higher than w* can be considered as a "gust". This value will depend on the tower structure and possibly on given guidelines, and is typically used for defining the "cut-off" value of the WS which depends on machine features, so it is left unspecified here.
Hence, let $N_b(t)$ define the stochastic process of the WS values which cross the "barrier" w*, i.e. the number of the “peaks-over-threshold” in terms of WS[3]. On the basis of mild assumptions, generally satisfied, such as that the mean duration of each gust U_k is much smaller than the mean time between the successive crossings τ_k and that the barrier level is high enough, the $N_b(t)$ process – here defined as “gust counting process” – can be described, as deduced in advanced books on extremal processes, e.g.[9], by the well-known Poisson probability law *p(k,t)* described by:

$$p(k,t) \equiv P\left[N_b(t) = k\right] = e^{-\phi_b t} \cdot \frac{(\phi_b\, t)^k}{k!} \qquad (1)$$
$$k = 0,1,...,+\infty$$

In (1) ϕ_b is the mean number of up-crossings in the unit time. The mean and variance of the process $N_b(t)$ are numerically equal and given by:

$$\mathrm{E}[\mathrm{N_b(t)}] = \mathrm{Var}[\mathrm{N_b(t)}] = \phi_b t \tag{2}$$

Let us focus our attention on the gust amplitude occurring at time T_k: such amplitude is a random variable, here indicated as Z_k. An intuitive safety index for characterizing the *extreme values* of the stochastic process $N_b(t)$ is the maximum gust amplitude over the interest time interval, which is also an obvious index of the damage that the gust process can bring about to the system. This can be accomplished by associating to the stochastic process $N_b(t)$ and the random variables Z_k ($k=1,2,\ldots, N_b(t)$), the following stochastic process:

$$\begin{aligned} &\mathrm{Y(t)} = max[Z_1, Z_2, \ldots, Z_N], \text{ if } N(t) \geq 0 \\ &\mathrm{Y(t)} = 0, \textit{otherwise}. \end{aligned} \tag{3}$$

where (as in the following): $N(t) = N_b(t)$. By assigning a "safety level" z^*, the following safety index (SI) can be consequently defined:

$$\mathrm{S(t)} = \mathrm{P}[\ \mathrm{Y(t)} < z^*\ ] \tag{4}$$

Accordingly, S(t) is the probability that z* is never exceeded over (0,t) since - for every assigned value *n* of N(t) - the following relationship holds:

$$[\max[Z_1, Z_2, .., Z_n] < z^*] \ \textit{if and only if} \ [(Z_1 < z^*)\cap..\cap(Z_N < z^*)] \tag{5}$$

The RV Z_k are assumed to be statistically independent and identically distributed with the common, time-independent, cumulative distribution function:

$$\mathrm{F(x)} = \mathrm{F_Z(x)} = P(Z_k \leq \mathrm{x}),\ \forall\ \mathrm{k} = 1,2,\ldots\ldots\mathrm{n} \tag{6}$$

After trivial manipulations, the following compact expression can be obtained for the above SI under the Poisson hypothesis for $N_b(t)$:

$$\mathrm{S(t)} = \exp[-\varphi_b t(1 - F(z^*))] \tag{7}$$

In the paper, in order to model the parent distribution F(z), a Pareto model[10] is adopted, with positive parameter k, with *cdf:*

$$F(x) = 1 - \left(\frac{1}{x}\right)^k, x > 0 \tag{8}$$

The method of using a combined model of an exceedance stochastic process and a "Pareto distribution" (PD) is similar to a recently adopted model[11] in a different environment, i.e. for modelling extreme temperatures under global warming. As a function of time t, the SI it is an Exponential cdf, as may easily seen by expressing it as:

$$\mathrm{S(t)} = \exp(-\mathrm{q}\varphi t) \tag{9}$$

having defined:

$\phi = \phi_b$ = *mean gust frequency (i.e., expected number of gust occurrence per unit time);*
q=q(z*)= 1 - F(z*) = P(Z_j > z*) = *exceedance probability (EP) of the value z* by any single RV* Z_j.

It has to be highlighted that the EP neither depends on the index j, nor on time. The function q(z*), i.e. the EP which represents the gust cdf, may assume various expressions. In the adopted PD model, inference for the above model can be adequately accomplished by following a Bayesian approach as shown in the following.

3 BAYES INFERENCE FOR EXTREME VALUES PROBABILITY DISTRIBUTION

Bayesian Inference is widely recognized as a powerful tool for exploiting both experimental data from the field (which are the only data to which classical statistical inference is referred) and prior knowledge, which in practice can be thought always existing in engineering environment. The integration of these knowledge sources makes possible to tailor an efficient estimation procedure and even to integrate available information, especially (but not only) when there is a certain lack of data .
This approach was adopted in the relevant literature for estimating a generic quantile of the "extremal" process Y(t) of (3). In this section the Bayesian estimation of SI is explained.
In particular, a “power Beta” prior pdf is adopted in the paper for the unique parameter k of the PD, so that a Beta prior pdf for Q is obtained, i.e. K is distributed so that Q= $F(z^*)=1-(1/z^*)^K$ has a a Beta prior pdf . Alternatively, a Beta prior pdf can be assigned directky to Q, if – as reasonable – some prior information are possessed about Q. The “input data” for the estimation is a joint prior pdf, denoted as $g(q,\phi)$, for interest parameters Q and Φ. As well known, the rationale behind the methodology is that the parameters to be estimated are regarded as RV (consequently, capital letter Q and Φ are adopted here for them, except when they are arguments of functions): so, they possess a pdf which can be integrated and updated with field data – denoted by D – by the below reported Bayes’ theorem:

$$g(q,\phi \mid D) = g(q,\phi)L(D \mid q,\phi)/C \tag{10}$$

where:
- $L(D|q,\phi)$ is the Likelihood (pdf or, in the case of discrete observation as in the present case, “probability mass" functions), of the data D conditional to the parameters (q,ϕ);
- C is a constant (with respect to the parameter values):

$$C = \int_0^{+\infty}\int_0^{+\infty} g(q,\phi)L(D \mid q,\phi)dqd\phi \tag{11}$$

As well known, the best Bayes estimate - in the mean square error sense - of a given function τ=τ(q,ϕ) is provided by the posterior mean:

$$\tau^\circ = E[\tau \mid D] = \int_0^{+\infty}\int_0^{+\infty} \tau(q,\phi)g(q,\phi \mid D)dqd\phi \tag{12}$$

where τ° denotes an estimate of the generic parameter τ.
In the interest case, the quantity to be estimated is furnished by the above SI, so that, for an assigned time t=t*:

$$S^{\circ}=\mathrm{E}[\mathrm{S}|D]=\int_{0}^{+\infty}\int_{0}^{+\infty}\exp(-q\phi t^{*})g(q,\phi|D)dqd\phi \tag{13}$$

The Bayes inference herein described employs the well known “conjugate” priors for the RV Q and Φ, i.e. the Beta prior pdf for Q and the Gamma prior pdf for Φ[10]. These two random variables are moreover assumed to satisfy the hypothesis of statistically independence (a reasonable hypothesis, as they are generated by independent physical phenomena).
The *Beta pdf* with positive parameters r and s, is defined, for a RV assuming values on (0,1), as:

$$betapdf(q;r,s)=\frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}q^{r-1}(1-q)^{s-1},\ 0<q<1.$$

Mean value and variance of the Beta pdf are given by:

$$\mu_B=\frac{\mathrm{r}}{\mathrm{r}+\mathrm{s}}\ ;\ \sigma_B^2=\mu_B^2\left\{\frac{\mathrm{s}}{\mathrm{r}(\mathrm{r}+\mathrm{s}+1)}\right\}.$$

The *Gamma pdf* with positive parameters n and δ (shape and scale parameter respectively) is expressed by:

$$gampdf\,(\phi,n,\delta)=\frac{1}{\delta^n\Gamma(n)}x^{n-1}\exp\left(-\frac{x}{\delta}\right),$$

where Γ(α) is the “Euler-Gamma” Function evaluate at α. The mean value and variance of the distribution are:

$$E[\Phi]=n\delta \qquad\qquad Var[X]=n\delta^2$$

By denoting with the suffix “0” the prior pdf parameters and by the suffix “1” the posterior pdf parameters, the prior joint pdf g(q,ϕ) is given by:

$$g(q,\phi)=betapdf(q;r_0,s_0)\cdot gampdf(\phi;n_0,\delta_0) \tag{16}$$

The parameters’ values (r_0,s_0,n_0,δ_0), according to the Bayesian paradigm, are inferred from prior information, i.e. test plant or experts’ judgements. Data collection is based upon recording the number of gusts *N(u)*, occurring in the time interval *u*, and the marking number of gusts *M* that over cross the fixed value z*. Once the time interval *u* has been chosen, the number of gusts *n=N(u)* becomes a constant, and the RV *M* becomes a Binomial RV representing the number of exceedances in *n* independent proofs[10].
By properly combining the prior pdf and the Likelihood Function, according to Bayes' formula, it is trivial to verify that the posterior pdf of Q and ϕ are again Beta and Gamma respectively, with updated values of the parameters which are function of the measured values *n* and *m*; these have the role of “sufficient statistics” for the problem at hand. In particular, the following relation maintains:

$$g(q,\phi|D)=betapdf(q;r_1,s_1)\cdot gampdf(\phi;n_1,\delta_1) \tag{17}$$

where:

$$\mathrm{r}_1=\mathrm{r}_0+\mathrm{m};\quad \mathrm{s}_1=\mathrm{s}_0+n-\mathrm{m};\quad \mathrm{n}_1=\mathrm{n}_0+\mathrm{n};\quad \delta_1=\frac{\delta_0}{1+\mathrm{u}\delta_0} \tag{18}$$

The above relationship involves the posterior conditional independence of Q and ϕ, once assigned the data D. If the following equality holds:

$$n_0 = r_0 + s_0 \tag{19}$$

the prior and posterior pdf can be easily expressed. In this noteworthy case, the product QΦ is Gamma distributed with parameters (r_0, δ_0) according to well known properties of the Beta and Gamma distribution[10]. Hence the product:

$$H = Q\Phi t \tag{20}$$

is still Gamma distributed with parameters (r_0, θ_0), with $\theta_0 = \delta_0 t$ and, since $S(t)=exp(-Q\Phi t)$, the obvious relation can be stated:

$$H = -\ln S \tag{21}$$

This implies that prior pdf for *H* is:

$$g(h) = gampdf(h; r_0, \vartheta_0) \tag{22}$$

and, hence, the prior pdf *p(s)* for the SI is a Negative Log Gamma one:

$$p(s) = \frac{1}{\Gamma(r_0)\vartheta_0^{r_0}} s^{\left(\frac{1}{\vartheta_0}-1\right)} (-\log s)^{r_0-1}, \; 0 \le s \le 1 \tag{23}$$

S is indeed a RV belonging to the interval (0,1).
Following the same way used for updating the prior pdf $g(q,\phi)$ into the posterior pdf $g(q,\phi|D)$, by remembering the analytical relationship between (q,ϕ) and the SI, it is immediate to obtain, for the Safety Index SI=S(t), the following posterior pdf, for the collection data D:

$$p(s \mid D) = nlgpdf(s; r_1, \vartheta_1) \tag{24}$$

with:

$$r_1 = r_0 + m; \theta_1 = \delta_1 t; \delta_1 = \frac{\delta_0}{1 + u\delta_0} \tag{25}$$

This formulation permits a feasible and direct way for obtaining the Bayes estimate of SI, according to the well known properties of Negative Log Gamma function:

$$S^0 = E[S \mid D] = \frac{1}{(1+\theta_1)^{r_1}} \tag{26}$$

If the equality (19) is not satisfied, numerical methods as those reported in[12] have to be adopted, which in any case do not involve critical tasks. It has to be remarked that both p(s|D) and S° depend, of course, on time, since the SI is a function of time. This dependency is obviously embedded in the parameter $\theta_i = \delta_i t$ (i=0,1).

4 EVALUATION OF BAYESIAN ESTIMATION EFFICIENCY BY MEANS OF NUMERICAL SIMULATION

In order to achieve both numerical evidence of the above discussed efficiency of the Bayesian estimator and to demonstrate thoroughly its performances, a large set of numerical experiments have been performed, by means of Monte Carlo simulation[12]. These experiments are in particular focused on:

1. evaluation of the Mean Square Error of the Bayes estimator;
2. comparison of Bayesian estimates with the classical ones, in particular with the most adopted Maximum Likelihood (ML) estimates.

They were conducted for various sample sizes and various input data values. For the sake of brevity, only a significant subset of the results is reported.
Data for the SI were generated from the assumed prior pdf on (Q, Φ), while:
- data on the observed number of gusts over a given time interval u were generated by a Poisson Process of mean frequency Φ (randomly generated previously according to the prior pdf) in the interval $(0,u)$;
- data on the observed exceedance number m were generated by a Binomial RV with parameters (n,Q), being also Q randomly generated previously according to the prior pdf.

The prior data, supposed to be deduced from past observations in this field, and relevant to an extremal WS value w*= 20 m/s, are chosen for sake of illustration as follows (the prior pdf are chosen equal to those in[10] for sake of comparison):

- Φ has a Gamma pdf with $\mu=11.0\ year^{-1}$ and $\sigma = 0.22\ year^{-1}$;
- K has a Power Beta pdf chosen so that Q has a Beta pdf with $\mu=0.02$ and $\sigma = 0.0275$.

Sample Size, n	MSEB	MSEL	REFF
5	0.0113	0.0394	3.4867
15	0.0083	0.0208	2. 5060
30	0.0065	0.0110	1.6923
50	0.0045	0.0067	1.4889

Table 1: Values of sample MSEB, MSEL and REFF versus sample size, for a set of $N=10^4$ Monte Carlo random samples.

The choice of the two parameters (n_0,δ_0) of the prior Gamma pdf, and those (r_0,s_0) of the prior Beta pdf are easily obtained by inverting the relations between the mean and variance of these priors and the relevant parameters.
For each sample size n, a number of $N=10^4$ replications has been performed in which the above RV Φ and Q were generated according to the above pdf, and the Bayes estimate of S was deduced. In particular, the results for various sample-sizes ($n=5$, $n=15$, $n=30$, $n=50$) are reported in Table 1, in terms of the classical indexes:
MSEB: Mean Square Error of the Bayes estimator;
MSEL: Mean Square Error of the ML estimator;
REFF=MSEL /MSEB.
The "REFF" index is the well known "relative efficiency" of the Bayes estimator with respect to the ML estimator. The above "Mean Square Errors" have been obtained at the end of each simulation as the averages over the N sampled estimator's square errors.
It is remarked, by observing the MSEB index, that the efficiency of Bayesian estimation increases – as always occurs – when number of data is exiguous. Moreover, it has to be highlighted that it is very

well performing, much more than classic one (as the REFF values clearly put in evidence), also when many data are available.

5 CONCLUSIONS

The paper proposes an interesting approach for the estimation of the probability that wind speed is lower than a prefixed extreme value which might be dangerous in terms of safety. From a probabilistic point of view, the method is based on the POT method for describing the stochastic processes of WS extremes in time and a PD for the parent distribution, exploiting the Bayes estimation method for inference on the above probability, which allows to define a proper "safety index" for a wind tower. A large set of numerical simulations, performed in the last part of the paper, shows the absolute and relative efficiency of the proposed method of estimation. As a further step to be taken for future studies, it would be worthwhile to assess the robustness of the proposed estimation method. At this purpose, it will be suitably verified that this approach provides satisfactory estimates also when the true prior models are different from the ones -the conjugate Beta and Gamma pdf- assumed in the present work.

REFERENCES

[1] Y.M. Kantar, I. Usta, "Analysis of Wind Speed Distribution", *Energy Conversion and Management*, **49** (2), 962-973 (2008).
[2] D. Villanueva, A. Feijoo, "Wind Power Distributions: A review of their Applications", *Renewable and Sustainable Energy Reviews*, **14**, 1490-1495 (2010).
[3] Y. An, M.D. Pandey, "A comparison of methods of extreme wind speed estimation", *J. Wind Eng. Ind. Aerodyn.*, **93**, 535–545 (2005).
[4] E. Chiodo, D. Lauria, C. Pisani, "Bayes estimation of wind speed extreme values", Proceedings of 3rd Renewable Power Generation Conference (RPG 2014), Napoli 24-26 September, (2014).
[5] P. Tabbuso, S.M.J. Spence, L. Palizzolo, A. Pirrotta, A. Kareem, "An efficient framework for the elasto-plastic reliability assessment of uncertain wind excited systems", *Structural Safety*, **58**, 69-78 (2016).
[6] L. Carassale, G. Solari, "Wind modes for structural dynamics: a continuous approach", *Probabilistic Engineering Mechanics*, **17** (2), 157-166 (2002).
[7] L.C. Pagnini, G. Solari, "Joint modeling of the parent population and extreme value distributions of the mean wind velocity", *Journal of Structural Engineering*, **142** (2), Article n. 04015138 (2016).
[8] M. Grigoriu, G. Samorodnitsky, "Reliability of dynamic systems in random environment by extreme value theory", *Probabilistic Engineering Mechanics*, **38**, 54-69 (2014).
[9] M. R. Leadbetter, G. Lindgren, H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, N.Y., (1983).
[10] N.L. Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, J. Wiley, Vol. 1 and 2, (1995).
[11] Y.Z. Osman, R. Fealy, J.C. Sweenwy, "Modelling extreme temperatures in Ireland under global warming using a hybrid peak-over-threshold and generalized Pareto distribution approach", *International Journal of Global Warming*, **7** (1), 21-47 (2015).
[12] S.J. Press, *Subjective and objective Bayesian statistics: principles, models and applications*, New York, John Wiley&Sons, (2003).