T-Return Period Values of Meteorological Design Parameters

M. Le Du and C. Laurent
EDF R&D
1 avenue du General de Gaulle, 92141 Clamart, France, *marc.le-du@edf.fr*

Abstract—Based on a literature review, this paper presents the main methods for estimating T-return-period values of Meteorological values to be used as design parameters : classical Generalized Extreme Values (GEV) distributions - including the famous Gumbel Law - and less classical but useful Peak Over Threshold (POT) methods - including Generalized Pareto Distributions (GPD). The distribution of extremes from mixed parent phenomena is discussed. A particular attention is paid to the statistical conditions the initial parent data must respect for application of Extreme Values Analysis (independence, identical distribution) and the ways to check for these conditions. Tests of nullity of the shape parameter (depending on the shape of tail of parent law) are described. The problem of precision of estimates is evaluated.

I. Introduction

Many codes dealing with the design of overhead lines are derived from reliability based design principles (CEI60826, CENELEC50341, ASCE74, …). Basically, the principle is to design the line for a given reliability level, reasonable regarding security, safety and costs. The reliability level is often defined by the return period of the climatic load that the component must withstand, associated with a specified resistance of components. The reference return period T is generally 50 years, but can be higher depending on the importance of the line, the consequences of a damage, etc.

To reach the reliability level, one must therefore know the climatic parameter X_T (wind velocity, ice load, temperature, …) of return period T. Of course, T is generally much bigger than the period for which we have reliable measured data. So how can we know the 50 or 100 years return period wind velocity when we have 10 to 20 years (or less) experimental data ? The statistical techniques to do this are referring to Extreme Values Analysis.

II. Return period and Probability

The value X_T of a climatic parameter is the value which is exceeded, on average, once every T years, T being the return period. If P(X) is the probability of not exceeding X in one year, T is given by :

$$T(X) = \frac{1}{1-P(X)} \qquad (1)$$

In this expression, P(X) is the cumulative distribution function of the extreme value X. The aim of extreme values analysis is to assess the function P(X) for the considered parameter : once you have P(X), you can get X_T. To do this, one must estimate the parameters defining P(X) with a limited number of extreme observations of X.

III. Conditions for estimation of extreme values distributions

Before describing the main forms of P(X) and the way to estimate their parameters, we must begin with the statistical conditions to be respected so that these estimates are reliable. The classical conditions in statistical estimation are usually expressed as : "the random variables are independent and identically distributed" [1].

A. The "identically distributed" condition

This condition refers to the homogeneity of the data : the selected values must come from the same "parent law" from which the extreme value distribution is derived. If the "regular" values are not driven by the same law, the extreme values law can not be unique. Fig. 1 illustrates a 4-calendar years time-series plot of daily maximum wind values in Marseille (France).

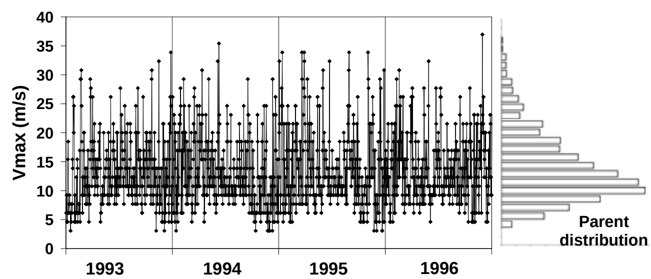

Fig. 1. Example of 4-years time-series data : maximum daily wind in Marseille. The parent distribution is a Weibull law

Checking that the extremes are issued from sequences of identically distributed variables implies an analysis of all the data, not only an analysis of the extremes. The determination of the "parent law" is therefore important. Fig. 1 shows that this data should probably be separated in seasons, the summer and winter wind distributions being obviously very different in this region.

There are numerous reasons why the data could not be homogeneous. The first one is that the measured meteorological parameter can be caused by different phenomena. For example ice loads can be caused by freezing rain, wet snow, rime, … Wind velocity can be caused by extra-tropical winter storms, lighting storms, tornadoes, … It might be necessary to separate the data depending on the origin phenomena before performing an Extreme Values Analysis. An other reason of non homogeneity might be a change of a sensor dedicated to the measure of the meteorological parameter (anemometer, dynamometer, thermometer, …). A faulty instrumentation or changes in the environment of the sensor are other possible causes [2]. A major cause of non homogeneity – a very discussed one – could be a climatic evolution, natural or not, revealing a trend in measured values of the meteorological parameter : increase, decrease, periodicity, etc. When classical Extreme Values Analysis methods are used for the determination of design values, the implicit hypothesis is that statistical properties of *past* measured data are supposed to be kept in the *future*.

How can we check for this homogeneity ? As mentioned in [2], "a simple time series plot of the data may be sufficient to reveal the presence of trends or step jumps, and is a necessary precursor to further analysis". More advanced techniques are available to answer questions such as : Do two distributions have the same means or variances? Are two distributions different? One should always wait for numerical differences, and test whether these differences are statistically significant. Most books dealing with statistics describe such tests [1]. The test of Kolmogorov-Smirnov is among the most useful. It gives an answer to the question : Are two datasets drawn from the same distribution function or from different distribution function? This test is described in Appendix. Reference [3] provides numerical tools to implement it. It can be applied to test the similarity of yearly distributions of the meteorological parameter for each couple of years, or for the first and the second half of the time series, etc.

B. The "independence" condition

This condition means that the selected extreme values must come from different events. If you use classical methods where you select only the biggest value each year, there is only a little chance that two values come from the same event, but it is not impossible (one value on the 31st December year Y, an other value 1st January year Y+1). As we will see later, other methods do not ask for the restriction of one value per year, and it is therefore important to pay attention to this independence condition. It is customary to assign a minimum separation time to ensure the independence of extremes. For example, wind speeds must be separated by at least 2 to 3 days [4]. For meteorological parameters, it is useful to check that the time series between two selected extremes reveal this independence: for ice loads, two selected values should be separated with an absence of load.

IV. Shape of the Tail of Parent Law

As mentioned previously, the determination of "parent law" is important to check for the "identically distributed" condition. It is also useful to get an idea of the shape of the tail of the distribution. This shape determines the kind of extreme of the variable of interest. It is customary to distinguish three types of tails : Type I correspond to unlimited extremes with exponentially decreasing probability density (for example Normal or Weibull distribution). Type II corresponds as well to unlimited extremes, but with a thicker tail, meaning a higher probability of occurrence of very big values of parameter X (for example Cauchy distribution). Type III corresponds to finite tails and bounded extremes (for example inverse Weibull or Log-Normal distribution).

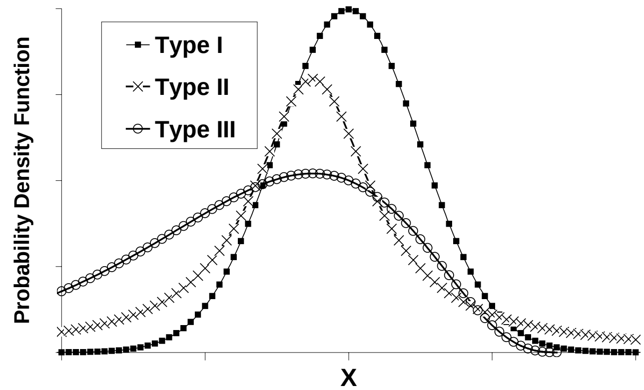

Fig. 2. Example of three types of tails of parent laws : Type I is Normal, Type II is Cauchy and Type III is inverse Gamma.

As we are dealing with extremes, this notion is very important. The use of the traditional "Gumble laws" to assess extremes implies a Type I shape of tail. This assumption might have big consequences on the final value of design parameter X_T. Fig 3 illustrates the variations of X_T with T depending on the type of the shape of tail.

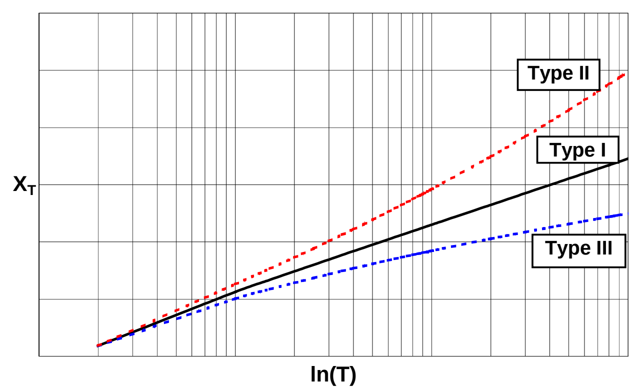

Fig. 3. Variations of X_T with T depending on the type of the shape of tail.

Choosing one type of shape without checking it can lead to overestimate or underestimate X_T depending on the case. We can see that type III, corresponding to parent laws with bounded extremes, will lead to limited variations and almost asymptotic values of X_T for big T. On the contrary, type II, corresponding to unbounded and non-exponentially decreasing extremes, will lead to a very high sensitivity of X_T against T.

Reference [2] indicates that Type I distribution is usually preferred for extreme wind speeds, explaining that many studies revealed a parent Weibul distribution and "there is no natural upper bound to wind speed anywhere approaching the orders of magnitude at which wind speeds are naturally observed". Type III was chosen in some cases, but Type II is rarely found, generally revealing mixed wind series, the mode and the tail of parent distribution being given by different phenomena.

Type I is also often chosen for extreme precipitations [4], [8], at least for maximum precipitations corresponding to 24 hours or more. For shorter periods, Type II seems to lead to better adjustments [4].

For extreme temperatures, Type I and Type III are usually chosen, depending on the period selection (year, season) or minimal or maximal temperatures [4], [8], [9]. However, one can argue that these discussions are not relying on satisfactory statistical data, as the trend in the evolution of global mean temperatures during last century and specially since the 1960's seems now accepted [10]. Time series for temperatures are probably not identically distributed.

What about accretions ? Among the parameters of interest in the design of overhead lines, ice thickness or ice loads are among the most complex, as they result of combinations of - at least – precipitations, temperatures and winds. As accretions are not permanent and even not frequent depending on the region (no accretion in one or several years), it might be difficult to get enough data to examine the parent distribution, not even talking of the shape of its tail. When the data is available, Type I is often chosen, based on exponentially decreasing type of parent distribution (for example [11]). But Type II has also been preferred, reference [6] explaining that "seems generally true of the extreme ice thickness data". As for temperatures, the trend of accretion loads measured in Studnice in Czech Republic [12] might alert on the risk of non identically distributed data: this data shows a dramatic increase of ice accretion during the last decade of 20th century.

Il might anyway be sterile to argue the type of tails for each parameter in general: many things depend on the local conditions, for icing probably even more than for other parameters. The only way to cope with this is to examine data for each station, and test the shape of tail if possible.

V. Two class of methods for estimation of P(X)

At this stage, we assume the two basic conditions previously described for application of Extreme Values Analysis are fulfilled : the initial data is composed of independent and identically distributed (thus stationary) values.

The two class of classical methods for estimating the distribution of extremes depend on the extreme data sample selection. In the first one, the selected extremes are the maxima of yearly datasets, on which a Generalized Extreme Value (GEV) distribution is fitted. In the second one - known as Peak Over Threshold (POT) method - the selected extremes are all the values exceeding a threshold, on which a Generalized Pareto Distribution (GPD) is fitted. Fig. 4 illustrates the selected data for these two class of methods. Because of the "identically distributed condition", we cut the summer season of Fig. 1 and keep the winter season, from November to March. Another analysis should be made for the summer season.

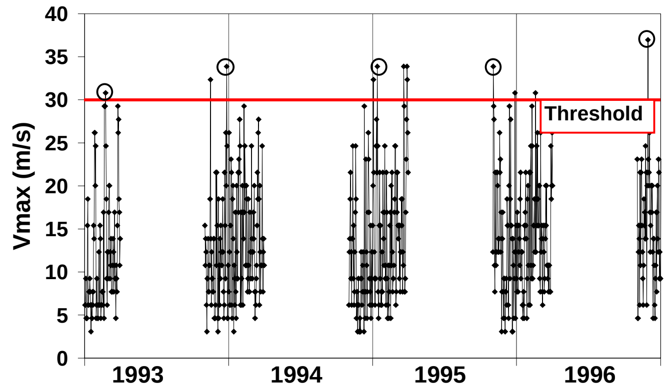

Fig. 4. Example of selection of extremes depending on the method – yearly maxima for GEV (circled) – values exceeding threshold for POT.

GEV methods will use one extreme value per year (5 extremes on Fig. 4). One could notice that for 1994-1995, 3 points reach the maximum value (wind speeds are integer in m/s) but we select only one. The choice of "year" (calendar, seasonal, etc.) might have an influence on the selected data [5]. From Reference [6], "one could argue that it makes more sense to choose maximum ice thickness for the season rather than for the calendar year".

POT methods allow a selection of more values (12 values above 30 m/s on Fig. 4, on which 11 are assumed independent). The selected data depends on the threshold (choice of threshold will be discussed later) but it does not depend on a specific period: several values might be selected in a year, and none in another one.

VI. Basic principle and plotting positions

The basic principle of Extreme Values Analysis is to fit the 2 or 3 parameters defining the cumulative distribution function P(x) to the selected set of data. At this stage, we have selected and ordered N extreme values $\{x_1 \leq \ldots \leq x_N\}$, N being the number of years for GEV, or N depending on the threshold for POT (usually $N_{POT} > N_{GEV}$)

To fit the parameters of P(x), one must estimate $P^*(x_i)$, the value of P(x) for each x_i. We will then get N couples of coordinates $(x_i, P^*(x_i))$ on which we will try to adjust a supposed asymptotic function of P(x). The estimate of $P(x_i)$ is known as the "plotting position". Influenced by the usual way of estimating cumulative density functions - not only their tails - the classical plotting position is often used:

$$P^*(x_i) = i/(N+1) \qquad (2)$$

However, Reference [23] explains that this estimate introduces bias when estimating quantiles of P(x), which is the case for Extreme Values Analysis. The following form of $P^*(x_i)$ is often chosen, either for GEV and POT methods [15], [17], [18], with $-0.5 < a < +0.5$

$$P^*(x_i) = (i-a)/(N+1-2a) \qquad (3)$$

As a simplification, the plotting position $P^*(x_i)=(i-0.35)/N$ is often met in literature.

VII. THE GENERALIZED EXTREME VALUE DISTRIBUTION (GEV)

When the selected extremes x_i are the maxima of yearly datasets, the cumulative distribution function P(x) to be fitted is the Generalized Extreme Value (GEV) distribution, given by [2]:

$$P(x) = \exp\left[-\left(1-k\left(\frac{x-\beta}{\alpha}\right)\right)^{1/k}\right] \qquad k\neq0 \qquad (4)$$

$$P(x) = \exp\left[-\exp\left(-(x-\beta)/\alpha\right)\right] \qquad k=0 \qquad (5)$$

k, β and α are respectively the shape, the location and the scale parameter. The shape of the tail of the parent law will determine the value of the shape parameter k, thus the type of extreme distribution: k=0 for type I, k<0 for type II and k >0 for type III. The aim of the Extreme Value Analysis is to estimate k, α and β, what completely defines the distribution. Once this is done, X_T is given by :

$$X_T = \beta + \frac{\alpha}{k}\left\{1-\left[-\ln\left(1-\frac{1}{T}\right)\right]^k\right\} \qquad k\neq0 \qquad (6)$$

$$X_T = \beta - \alpha \ln\left[-\ln\left(1-\frac{1}{T}\right)\right] \qquad k=0 \qquad (7)$$

A. Graphical estimation of parameters α and β for Gumbel law (k=0)

The large number of parent laws with an exponentially decreasing tail makes the type I or Gumbel law the most famous and most widely used extreme value distribution. Moreover, as you assume the shape parameter is zero, the estimation of the two parameters α and β only is required. Equation (7) can be written as well :

$$x = \alpha\left[-\ln\left(-\ln\left(P(x)\right)\right)\right] + \beta \qquad (8)$$

Thus, a plot of x on the ordinate and –ln[-ln(P(x))] on the abscissa will give a straight line : the slope will give an estimate of α and the intercept will give an estimate of β. The procedure consists in positioning the points (P^*(xi),xi), fit a straight line to these points (for example by using least-squares method), then determine the slope α and the intercept β.

Some authors [13] suggest that a power of the value x can be used instead of the value x itself. The reason for this is that the distribution of the annual extreme is a Gumbel law if the tail of the parent law is exponentially decreasing. If x is supposed to be exponentially decreasing, x^q is even more ! So the convergence to a Gumbel law for x^q is likely to be quicker than for x. Of course, the estimate parameters α and β are not the same for x and x^q. If x^q is used, X_T is not given by equation (7) but by the q^{th} root of the right hand side.

B. Estimation of parameters α, β and k in the general case (k≠0)

As noticed earlier, the use of Gumbel graphical method presuppose k=0. Instead of doing this, we suggest applying the general case, and test afterwards the nullity of shape parameter k. In this general case when k is unknown, the estimation is however less simple, because three parameters must be found. There are several numerical methods for the estimation of these three parameters. The two most commonly used are Probability Weighted Moments (PWM) and Maximum Likelihood (ML) solutions.

1) Probability Weighted Moments (PWM)

The parameters α, β and k can be deduced from the three first statistical moments b_0, b_1 and b_2 of the extreme value X. Unbiased estimates of these three moments for a set of N ordered extreme annual values $\{x_1 \leq x2 \leq \ldots \leq x_N\}$ are given by [15]:

$$b_r = \frac{1}{N}\sum_{i=1}^{N} x_i \times \left[P^*(x_i)\right]^r \text{, r=0,1,2, …} \qquad (9)$$

Where $P^*(x_i)$ is the plotting position (see part VI). From these moments, we calculate an intermediate variable c :

$$c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\ln 2}{\ln 3} \qquad (10)$$

Estimates of parameters α, β and k are then given by :

$$k = 7{,}859\ c + 2{,}9554\ c^2 \qquad (11)$$

$$\alpha = \frac{(2b_1 - b_0)k}{\Gamma(1+k)(1-2^{-k})} \qquad (12)$$

$$\beta = b_0 + \frac{\alpha\left[\Gamma(1+k)-1\right]}{k} \qquad (13)$$

Where Γ is the gamma function, tabulated in statistics books or provided by Microsoft Excel® functions.

2) Maximum likelihood (ML) method

This method is quite complex, as it is necessary to solve a set of three non linear equations of k, α and β. This requires an iterative procedure and a “first guess” of the three parameters for the beginning of iteration. Statisticians usually accept this complexity because this method is likely to provide the best estimates of the parameters α, β and k. We will not develop further this method. Details can be found for example in [16].

3) Test of nullity of parameter k

As mentioned earlier, the correct “choice” of k is very important, as the value of X_T is very sensitive to this parameter. We saw that the PWM and ML methods allows to estimate the three parameters while avoiding to make such a choice : the parameter k will not be chosen but calculated. In any case where the assumption k=0 is done, it might be useful to check it. Statistically, it means testing the hypothesis k=0 versus the hypothesis k≠0. Reference [17] gives different tests, depending on the method chosen for the estimation of the three parameters. The simplest is the “median test” : If x* is the median of the set of N ordered extreme annual values $\{x_1 \leq x2 \leq \ldots \leq x_N\}$, under the hypothesis that the x_i obey the Gumbel distribution, the following value d_{med} is asymptotically Normally distributed with mean m_{med} and standard deviation σ_{med} :

$$d_{med} = \ln\left[(x_N - x^*)/(x^* - x_1)\right] \qquad (14)$$

$$m_{med} = \ln\left[-\frac{\ln(N)}{\ln\left(-\ln 2/\ln\left(1-0.5^{1/N}\right)\right)}\right] \qquad (15)$$

$$\sigma_{med} = 1/\left(0.861\ln(N) - 0.490\right) \qquad (16)$$

Therefore, at the 5% significance level, the hypothesis k=0 can be rejected if the standardized value $(d_{med}-m_{med})/\sigma_{med}$ lies outside the interval [-1.96, +1.96].

When k is estimated using the PWM method, the estimate of k under a Gumble hypothesis is asymptotically distributed with zero mean and standard deviation $\sigma_{PWM} = \sqrt{0{,}5633/N}$. Therefore, at the 5% significance level, the hypothesis k=0 is rejected if the standardized value k/σ_{PWM} lies outside the interval [-1.96, +1.96].

The brackets can be adjusted for other confidence levels, following standardized Normal fractiles.

C. *Advantages and drawbacks of GEV*

The main advantage of the GEV methods presented above is their simplicity at the application stage. This is specially true for Gumble laws, allowing for a quite simple estimation of the parameters. Due to the selection of extremes, the “independence” condition is moreover easily reached.

The main drawback of these methods is that only one value per period is selected, which means that you need long time series to get enough data for a reliable estimation of the parameters of law. A rule often used for Gumbel laws says that, with a data set issued from n years, you can estimate with a “reasonable” confidence interval a 2*n return period value (confidence intervals will be discussed later). Thus, for a 50 years return period value, you need 25 years long time series. As you select one value in each period, the implicit hypothesis is that each period is subjected to identically distributed events. Obviously, it might not be the case for icing in places where no events occur during long periods. Reference [6] also points out that you select only one value in each period, even though several big events can occur in the same period. In a way, you ignore some information.

VIII. Peak Over Threshold (POT) methods

A. *Generalized Pareto Distribution (GPD)*

When the selected extremes x_i are all the values exceeding a threshold, the cumulative distribution function P(x) to be fitted is the Generalized Pareto Distribution (GPD) given by :

$$P(x) = 1 - \left[1 - \frac{k}{\alpha}(x - S)\right]^{1/k} \qquad k\neq 0 \qquad (17)$$

Like the GEV distribution, the GPD has a shape parameter k and a scale parameter α. The shape parameter k is the same than for GEV distributions. The parameter S is the selected threshold. For k=0, the GPD is a simple exponential distribution :

$$P(x) = 1 - \exp\left[-\frac{(x-S)}{\alpha}\right] \qquad k=0 \qquad (18)$$

If the threshold is selected high enough, the number of observations above the threshold per year is distributed following a Poisson law with rate λ. An estimate of λ is given by n/M, where M is the number of years of observation and n is the total number of values exceeding the threshold S. X_T is then given by one of the following expression, depending on the value of k :

$$X_T = S + \frac{\alpha}{k}\left[1 - (\lambda T)^{-k}\right] \qquad k\neq 0 \qquad (19)$$

$$X_T = S + \alpha\ln(\lambda T) \qquad k=0 \qquad (20)$$

The estimation of α and k might use the PWM method described previously. As two parameters are unknown, only the two first moment b_0 and b_1 given by (9) are required. The parameters α and k are estimated by [6]:

$$k = (4b_1 - 3b_0 + S)/(b_0 - 2b_1) \qquad (21)$$

$$\alpha = (b_0 - S)(1 + k) \qquad (22)$$

For the classical case k=0, $\alpha=b_0-S$, which lead to a quite simple solution, either for the estimation of required parameters and for the calculation of X_T.

Other estimation methods are possible, including ML methods described for GEV distributions. Because these methods can reach their limit for particular values of k (outside the range [-0.5;+0.5]), Castillo and Hadi proposed the Elemental Percentile Method (EPM). This method consists in getting different values of k and α corresponding to several percentile values of x, then take the median of these k and α [18].

B. *Renewal method*

Instead of using a GPD, one can also fit the following and closed form of P(x) to the selected data above threshold :

$$P(x) = \sum_{m=0}^{\infty} Q(m/S) \times \left[F(x/S)\right]^m \qquad (23)$$

Q(m/S) is the probability of observing m annual number of values above the threshold S. Q is classically a Poisson law.

F(x/S) is the probability that the annual maximum is lower than x, knowing it is greater than S. The function F is typically modelled with an exponential law or a Weibull law.

The aim is then to estimate the parameters of the chosen laws for the functions Q and F, while adjusting them to the data. The relevance of the chosen laws can be tested by using statistical tests (Kolmogorov-Smirnof). We will not fully develop this method, that can be a bit complex at the application stage : one must “play” with the threshold S so that you find a good fitness of both laws Q and F to the data above the threshold. If x is “big enough”, which must be the case for extremes, P(x) can be simplified :

$$P_{x-big}(x) \approx 1 - \lambda\left(1 - F(x/S)\right) \qquad (24)$$

Where λ is the “excess rate” defined previously for GPD. If F is modelled with a Weibull law, the final form will be :

$$P_{x-big}(x) \approx 1 - \lambda\exp\left[-\frac{x-S}{b}\right]^p \qquad (25)$$

The selection of threshold S defines the excess rate λ. The parameters b and p can then be adjusted to the data above the threshold S, with a PWM or ML method. This late formula has been used to estimate the T-return-period values of floods [19]. This method was also eventually chosen to model extreme wind speeds in France following the storms of 1999.

C. Choice of Threshold

The main difficulty with POT methods is the choice of the threshold [2] : "the threshold be set high enough so that only true peaks […] are selected. The threshold must be set low enough to ensure that enough data are selected for satisfactory determination of the distribution parameters". The threshold depends on the data and might differ from a site to an other depending on their climatic exposure.

Coles [22] describes two methods for the threshold selection. The first one consists in plotting the mean of the excesses (x_i-S) as a function of the threshold S ("mean residual life plot"). For a GPD, this graph should be a straight line. An appropriate threshold can be chosen by selecting the lowest value above which the graph is a straight line, but this plot can however be difficult to interpret. The second procedure consist in estimating the parameters of the distribution for a range of thresholds, and look for stability of these parameters. The appropriate threshold is the value above which the estimates of k and (α-k) are assumed constant.

More straightforward methods consist in choosing a threshold corresponding to a high percentile of the data (the x% biggest) or for a given value of excess rate λ [6].

IX. PRECISION OF ESTIMATES

The precision of estimates of X_T is given by its standard error $\sigma[X_T]$. The following expressions [2] give the standard errors for GEV and GPD and k=0.

$$\sigma_{GEV}[X_T] = \left[\frac{\alpha^2}{N}\left(0.608\,y^2 + 0.514\,y + 1.109\right)\right]^{1/2} \quad (26)$$

$$\text{with } y = \ln\left[-\ln\left(1 - 1/T\right)\right]$$

$$\sigma_{GPD}[X_T] = \left[\frac{\alpha^2}{N}\left\{1 + \left[\ln(\lambda T)\right]^2\right\}\right]^{\frac{1}{2}} \quad (27)$$

These values are the theoretical minima (known as Cramer-Rao bounds). Reference [20] studied different estimation methods of GEV distribution parameters and showed the sampling errors could overtake this bound of 30% for classical least-square Gumble plots, PWM and ML estimates giving more precise values. Assuming that the sampling error is normally distributed, the confidence interval of X_T can be assessed by $X_T \pm w^*\sigma[X_T]$, w depending on the required confidence level. Typical values of confidence limits are between one and two standard errors each side of X_T.

As the number N of selected values x_i is bigger for POT methods, the sampling error is likely to be lower than for GEV methods, giving better precision on X_T (depending however on the value of the excess rate λ).

One should notice that this error is only the statistical sampling one. It assumes the data is precise (what can be discussed, specially for the extremes), and the method is correct (conditions respected, k actually null, …).

X. MIXED DISTRIBUTIONS

As mentioned in the conditions for extreme values distributions, the data used must be homogeneous ("identically distributed" condition). One should not mix data coming from different phenomena with different parent laws. The procedure consists in isolating the data for each distribution/phenomena, and estimate the distributions of extremes separately. For example, if two phenomena (glaze and rime) are involved, the parameters of each distribution $P_{GLAZE}(x)$ and $P_{RIME}(x)$ are estimated separately. The final cumulative distribution function of X is then given by :

$$P(x) = 1 - \frac{1}{T(x)} = 1 - \left[1 - P_{GLAZE}(x)\right] - \left[1 - P_{RIME}(x)\right] \quad (28)$$

This expression will eventually give the value of X corresponding to a given return period T, assuming the two phenomena. As P_{GLAZE} and P_{RIME} are different distributions, there is not straightforward expression of the final X_T in the general case.

It may however not be necessary to separate the data if the extremes are obviously dominated by a single mechanism. Separation will moreover lower the number of data available for the estimation of the parameters of each distribution.

XI. WHEN CONDITIONS ARE NOT MET

At the beginning of this paper, we insisted on the statistical conditions required so that Extreme Value Analysis can be performed : independence and identical distribution.

Meeting the independence condition is a matter of selection of data. This might lead to lower the number of available points to fit the parameters of the extreme value distribution. If this number of points is low, the notion of "superstation" might be useful to extent it [2], [6]. The idea is to consider all the data measured in several stations, belonging to a region consistent from the meteorological phenomena of interest. The value of X_T estimated with this grouped data shall then be representative for the region. The selection of data must be made with care, as the independence of data is still a condition, and as one must be sure of the homogeneity of the climatic conditions in the region.

Problems are more critical if the "identically condition" is not respected. As addressed in another paper of this conference, accretion conditions in France are mainly due to wet snow events, likely to happen once every several years, sometimes with high severity. In that case, we do not have the data required to check for this condition, and each single data is an extreme value by itself. This lack of data leads us to fit distributions to the ice loads modelled from conditions, and not only from observed values.

The most complex situation probably concerns data revealing a trend. Usual statistical methods do not allow to take such a trend into account, although a linear dependence of the estimated parameters of extreme distribution with time can be considered, which means that X_T is supposed to evolve with time. Addressing such potential trends leads to fundamental questions, beyond statistical tools: How can we detect trends in extremes and not only in mean values? What is this trend supposed to be in the future ? etc.

XII. Conclusion

The key word in Extreme values Analysis is data. The final choice of a method will highly depend on the amount of available values of the climatic / design parameter of interest. One must remind that the GEV or POT distributions presented in this paper are asymptotic functions, likely to be reached for big number of samples of independent and identically distributed values. They will eventually converge to the same values of X_T.

When such large datasets are available, GEV methods have the important advantage of simplicity at the application stage. However, as they require one extreme value per year, they are probably limited to frequently exposed areas.

POT methods give an interesting alternative for lower series of data, as they allow for the selection of several values each year. Moreover, they seem better adapted to irregularly exposed areas. From these two aspects, POT methods are very interesting for ice loads data, specially in regions where accretions might happen from time to time, but sometimes severely. They are however requiring more decisions, specially in the selection of the threshold.

Whatever the method, the analysis of the distribution of the parent laws should not be spared, and specially checking for the "independence and identically distributed" conditions of the selected data. In particular, this analysis may lead to the detection of different phenomena at the origin of the value of interest, requiring a separation of data to avoid mixed distribution. It may also reveal a non stationary data, for which these classical Extreme Value Analysis methods are not valid.

In that case, the application of these methods will give a conventional but still necessary design value.

XIII. Appendix - Test of Kolmogorov-Smirnov

Let $F_{n1}(x)$ and $G_{n2}(x)$ be the empirical cumulative distribution function of two samples of respective sizes n1 and n2. The statistic D is the maximum value of the difference between $F_{n1}(x)$ and $G_{n2}(x)$. The test consists in rejecting the hypothesis that the two data sets are issued from the same distribution if the distance D is "too big". The acceptance depends on the required significance level. For a sufficiently large datasets (n1,n2 > 80) the hypothesis is rejected if :

$D > 1.358\sqrt{n_1+n_2}/\sqrt{n_1 n_2}$ for a risk level of 5%

$D > 1.628\sqrt{n_1+n_2}/\sqrt{n_1 n_2}$ for a risk level of 1%

For other risk levels or smaller datasets, please consult statistic tables [1]. This test can be used as well to check for the adjustment of an empirical cumulative distribution function $F_n(x)$ to a theoretical one, by replacing D by the maximum value of the difference between the empirical and the theoretical cumulative distribution functions, and by replacing the value $\sqrt{n_1+n_2}/\sqrt{n_1 n_2}$ by n.

XIV. References

[1] G.Saporta, « Probabilites, Analyse des donnees et statistiques », Paris : Technip, 1990, p. 263. in French

[2] J.P. Plautikof, B.B. Brabson, D. H. Lister and S.T. Adcock, "A review of methods to calculate extreme wind speeds", Meteorol. Appl. 6 pp 119-132, 1999

[3] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, "Numerical Recipes – the art of scientific computing", Cambridge University Press, New York 1988, Available : http://www.nr.com/

[4] AIEA, « Phenomenes meteorologiques extremes et choix des sites de centrale nucleaire », collection securite 50-SG-S11 A, Vienne, 1983, in French, exist in english.

[5] J. Laflamme, "Spatial variation of extreme values in the case of freezing rain icing", Proccedings of the Sixth IWAIS, pp 19-24, 1993

[6] K.F. Jones, R. Morris, "Investigation of the differences in the application of the Canadian and U.S. methodology for estimates of ice loads for a 50-yr return period", CEATI-WISMIG group report T003700-3306, 2002

[7] L. Gomes and B.J. Vickery, "Extreme wind speeds in mixed wind climates", J. Ind. Aerodyn. 2 pp 331-334, 1978

[8] C. Merlier, "Essai d'apllication de la loi generalisee des valeurs extremes", Rapport Meteo-France, 2001, in French

[9] David, "Etude climatologique des conditions de vents forts et basses temperatures sur le territoire français", Rapport Meteo-France, 1988, in French

[10] WMO, "WMO statement on the status of the global climate in 2003", 2003, available on http://www.wmo.ch.

[11] K. Savadjiev and M. Farzaneh, "Modeling Asymptotic Distributions of Extreme Icing from Various Initial Distributions," 10 th International Workshop on Atmospheric Icing of Structures, Brno, Czech Republic, June, 2002.

[12] F. Popolansky, J. Kruzik,. P. Lehky, J. Hrabanek and J. Lago, "Ice Monitoring at Stand Studnice. Tuned Vibration Control of Overhead Line Conductors", CIGRE Paper 22-105, Session –1998, Paris.

[13] R.I. Harris "Gumble revisited – a new look at extreme value statistics applied to wind speeds", J. Wind. Eng. Ind. Aerodyn., 1998

[14] H. Klajnmic "Estimation et comparaison des niveaux de retour des vitesses extrêmes des vents", Societe Francaise de Statistique, XXXVIèmes journees de statistiques, Montpellier 2004, Available : http://www.agro-montpellier.fr/sfds/ , in French

[15] J.R. Hosking, J.R. Wallis, E.F. Wood "Estimation of the Generalized Extreme Value distribution by the method of probability weighted moments", Technometrics, vol. 27, no 3, pp 251-26, 1985

[16] J.R. Hosking, "Maximum Likelihood Estimation of the parameters of the GEV distribution", Royal Statistic. Soc. 34, pp 301-310, 1985

[17] T. Farago, R.K. Katz „Extremes and Design Values in Climatology", WMO, World Climate Programme, WCP-14, TD-No 386, 1990

[18] E. Castillo, A.S. Hadi, "Fitting the Generalized Pareto Distribution to Data", J. Of the American Statistical Association, 1997, vol. 92, n0. 440

[19] L. Miquel, "Guide pratique d'estimation des probabilites de crues", Collection de la Direction des Etudes et Recherches EDF, Ed. Eyrolles, Paris, 1984

[20] E. Simiu, M. Asce, J. Bietry, J.F. Filliben, "Sampling errors in estimation of extreme wind », J. of the Structural Division, pp 491-501, 1978

[21] C. Dreveton, "Etude des evenements tempetes", rapport Meteo-France, Note SCEM No 23, 1997, in French

[22] S. Coles, "An introduction to statistical modelling of extreme values", Springer Verlag, London, 2001

[23] N.W. Arnell, M. Beran and J.M. Hosking, "unbiased plotting positions for the general extreme value distribution", J. of Hydrology No 86, 1986, pp 59-69.