August 27-31, 2023
FLORENCE - ITALY
16th International Conference on Wind Engineering

Parent population method for the assessment of return wind speeds: calibration from incomplete datasets

Vincenzo Picozzi, Andac Akbaba, Alberto M. Avossa, Francesco Ricciardelli

University of Campania "Luigi Vanvitelli", Aversa, ITALY
vincenzo.picozzi@unicampania.it

SUMMARY:
This paper deals with the calibration of the Parent Population Method for the assessment of extreme wind speeds starting from incomplete datasets. To this aim, high resolution data from the German Weather Service are used. First, the error on the right tail associated with the calibration of the parent probability density function from the entire available dataset is shown; as an alternative calibration on the 0.1% largest values is proposed. Then the effect of missing data, of false calms and of downsampling on the model parameter λ is investigated. As to downsampling, empirical correction factors are given, on average equal to 3.4 and 6.7 when data are sampled every one hour or every three hours, respectively; it is proved that the use of average values for correction is quite acceptable in terms of accuracy of the predicted return wind speed. Correction for missing data and for the presence of false calms can be carried out considering an equivalent record with duration equal to that of available, non-zero measurements.

Keywords: Extreme value analysis, Process analysis, reliability.

1. INTRODUCTION

The Parent Population Method (PPM) derives the Cumulative Distribution Function (CDF) of the annual maximum wind speed $F_{\hat{V}}(v)$ from the parent Probability Density Function (PDF) $f_V(v)$:

$$F_{\hat{V}}(v) = exp[-\lambda \cdot f_V(v)] \quad (1)$$

where λ is a model parameter, defining the expected number of upcrossings of a sufficiently high threshold *V*. The method was originally developed by Gomes and Vickery (1977) and further investigated and applied over the years (e.g. Lagomarsino et al. 1992, Palutikoff et al. 1999, Freda and Solari 2010, Burlando et al. 2013, Torrielli et al. 2013, Pagnini and Solari 2016).

Advantages with respect to classical Extreme Value (EV) analysis are mainly to be found in the fact that lower quality data are sufficient for its calibration; in fact, shorter and incomplete records suffice, provided that these are properly pre-processed, whereas EV analysis requires long and almost complete records in order to give reliable results. In the following, we will define as *complete record* a series of contiguous values of mean wind speeds, i.e. a series of wind speeds averaged over a period of time coinciding with the sampling time (e.g. 24 daily measurements of 1 hr averaged wind speeds or 144 daily measurements of 10 min averaged wind speeds), without missing values and obtained with an instrument with a virtually zero measurement threshold, so not giving rise to false wind calms. A record not satisfying one or more of the above conditions has to be regarded as an *incomplete record*. Calibration of Eq. (1) means finding estimates of the parameters of $f_V(v)$, and of λ, and incomplete records is what we have in practice to do it.

In Eq. (1), $f_V(v)$ is representative only of the statistical distribution of the parent population, and as such its estimate is independent of whether sampling is contiguous or disjunct (sampling period larger than the averaging time), and it is not affected by missing values; however, it is affected by the presence of false calms. On the other hand, λ depends not only on the statistical distribution of V, but also on the frequency content of the time series $V(t)$, therefore it depends on the sampling time, as well as on the presence of missing values and false calms.

As to $f_V(v)$, it is commonly accepted that a Weibull form is appropriate. However, it is objected that calibration with data containing false calms introduces errors; therefore two alternative solutions are seen, the use of either a hybrid Weibull or a left-censored distribution (Lagomarsino et al. 1992). However, regardless of the presence of false calms it is objected that the parameters well-fitting the body of the distribution do not necessarily fit its right tail with the same accuracy, and this introduces an error when the parent distribution is used to obtain the EV distribution.

As to λ, this is given by the product $\lambda = 2\pi\beta_v\nu_v\sigma_v$, where β_v is the ratio of the average positive rate of change of $V(t)$ to the standard deviation of $\dot{V}$, ν_v is the average cycling rate of $V(t)$, σ_v is the standard deviation of V. The three quantities composing λ can be calculated separately, σ_v depending only on the statistical properties of V, therefore being directly related to $f_V(v)$, and β_v and ν_v being dependent also on the frequency content of $V(t)$.

Within the PPM, starting from the work of Freda and Solari (2010) there has been an effort to correct the model parameter λ. They introduced three correction factors accounting for missing data, sampling period, and wind calms. The correction coefficient for missing data is set to the ratio between the number of data in a complete record and the available data; a correction factor of 4.0 is given for 10 min averages sampled every 3 hours; no correction factor is given for wind calms, as it is assumed that their effect is negligible. Burlando et al. (2013) suggested that correction for downsampling is made directly on the return wind speed rather than on the model parameter λ. The effect of downsampling when applying EV analysis was studied by Picozzi et al. (2022), and a similar approach is used in this paper.

2. METHODOLOGY

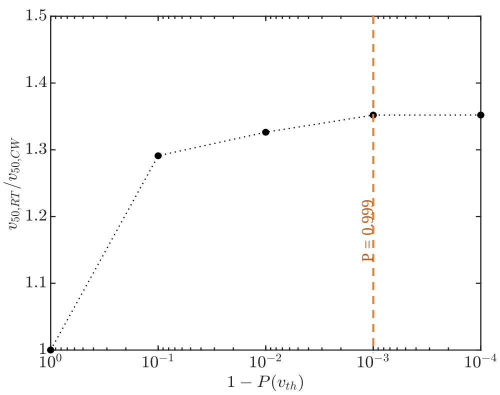
High-resolution data from automated weather stations of the German Weather Service (DWD) were used to investigate the effects of the calibration method and of the incompleteness of the dataset on the design wind speeds obtained through application of the PPM. DWD database contains contiguous values of the 10 min averaged wind speed and direction from about 500 meteorological stations located in Germany. Measurements comply with WMO standards. In this study, data from 114 stations were used, selected such to cover the 24 years between 1995 and 2018 with 95% of available data or more and nearly no calms. High altitude stations were discarded and no correction for roughness and orography was applied, as both do not affect the results.

For the parent probability density function, the classical Weibull model is chosen, calibrated on the right tail values of the available sample. This has the advantage of better reproducing the extreme values, and eliminates potential problems deriving from the presence of false calms, as low values in the sample do not contribute to calibration of the distribution.

The model parameter λ is first calibrated based on the complete datasets, so to obtain target values. Then, deviation from target values due to false calms, missing data and downsampling is investigated by creating artificially incomplete datasets starting from the complete ones. This was done by removing values so to obtain incomplete datasets with prescribed characteristics.

3. RESULTS

When fitting the parent distribution to the right tail of the available data, the question arises of how to select the right tail, i.e. how to select the threshold above which data are retained and below which they are discarded. To this aim a sensitivity study was carried out, as shown in Figure 1; the parent Weibull distribution was calibrated for different values of the threshold probability of exceedance (horizontal axis), and the ratio of the return wind speeds as evaluated with the right tail data and with the entire sample population is plotted on the vertical axis. Figure 1a shows that when the entire sample is used, the estimated return wind speed is about 26% lower than it is when the only the upper 0.1% values are used. Figure 1b shows that the right tail of the parent Weibull distribution is much better approximated when fitted to the right tail data.



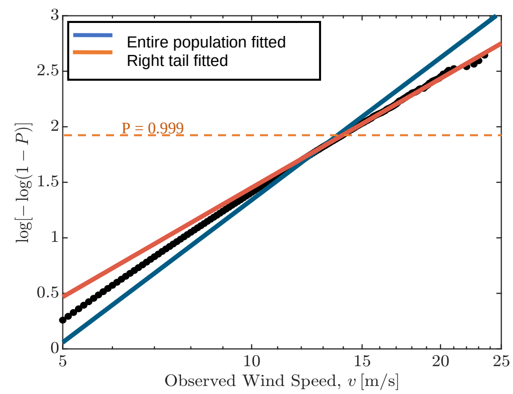


Figure 1. Definition of the right tail threshold and effect on the calibrated Weibull distribution.

To examine the effect of downsampling, a similar procedure to that proposed by Picozzi et al. (2022) was used. First the values of λ were calculated from the complete datasets of the 114 stations as in Gomes and Vickery (1977), see Figure 2a; a mean value of 6875 years^{-1} and a Standard Deviation of 727 years^{-1} (CoV=0.106) were found within the available population.
The complete datasets were then subsampled with different sampling times; for example, setting a sampling time to 20 min gives two artificial datasets from which two values of λ were calculated and averaged. Then the sampling time was set to 30 min, so to obtain three datasets and three values of λ that were again averaged. The procedure was repeated with sampling times up to 6 hrs. Then the ratios between the λ values obtained from the complete datasets and those obtained from the incomplete ones were calculated for different values of the sampling time and plotted in Figure 2b; blue line indicates the median values and dark blue hatch shows the 5% to 95% confidence

interval. These ratios correspond to the empirical downsampling correction factor ρ_λ, which ranges between 2.8 and 3.9 with a mean value of 3.4 and standard deviation of 0.20 (CoV=0.088) for hourly downsampled data, and between 5.2 and 8.9 with a mean of 6.7 and standard deviation of 0.64 (CoV=0.104) for three hourly downsampled dataset.

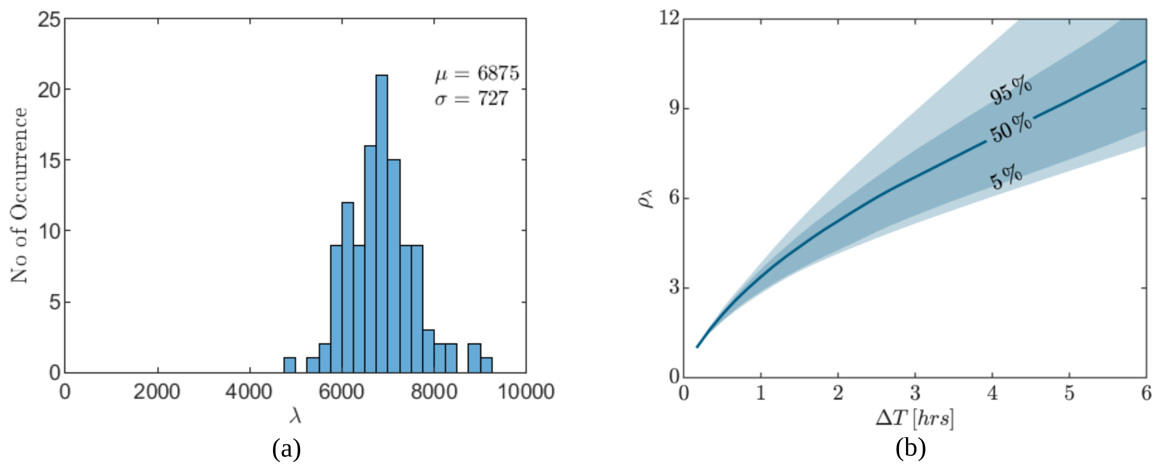


Figure 2. Model parameters for the PPM from the complete datasets (a) and correction factors for downsampling.

Correction for missing data can be simply performed by considering an equivalent duration of the record, equal to the original one to which the total duration of missing data has been subtracted. This is a quite reasonable approach in case there is a limited number of long duration periods of missing data; it is less appropriate when there is a large number of short duration periods of missing data, to the limit single data missing here and there in the record.

Finally, the effect of false wind calms can be assimilated to that of missing data, by considering all zero readings as they were missing ones.

REFERENCES

Burlando, M., De Gaetano, P., Pizzo, M., Repetto, M. P., Solari, G., and Tizzi, M., 2013. Wind climate analysis in complex terrain. J. Wind Eng. Ind. Aerodyn., 123, 349–362.

DWD Climate Data Center (CDC): Historical 10-minute station observations of mean wind speed and wind direction for Germany, version V1, 2019.

Freda, A., and Solari, G., 2010. A pilot study of the wind speed along the Rome-Naples HS/HC railway line. Part 2—Probabilistic analyses and methodology assessment. J. Wind Eng. Ind. Aerodyn., 98(8–9), 404–416.

Gomes, L., and Vickery, B. J., 1977. On the prediction of extreme wind speed from the parent distribution. J. Ind. Aerodyn., 2(1), 21–36.

Lagomarsino, S., Piccardo, G., and Solari, G., 1992. Statistical analysis of high return period wind speeds. J. Wind Eng. Ind. Aerodyn., 41(1–3), 485–496.

Pagnini, L. C., Solari, G., 2016. Joint Modeling of the parent population and Extreme Value Distributions of the mean wind velocity. ASCE, Vol. 142, Issue 2.

Palutikoff, J. P., Brabson, B. B., Lister, D. H., Adcock, S. T., 1999. A review of methods to calculate extreme wind speeds, Meteorological Applications n. 6, pp. 119–132.

Picozzi, V., Akbaba, A., Avossa, A.M., Ricciardelli, F., 2022. Correction of historical records to improve the reliability of design wind speeds. Engineering Structures, Volume 265,114473.

Torrielli, A., Repetto, M. P., and Solari, G. (2013). Extreme wind speeds from long-term synthetic records. J. Wind Eng. Ind. Aerodyn., 115, 22–38

World Meteorological Organization - Guide to Meteorological Instruments and Methods of Observation - Preliminary seventh edition - WMO-No. 8 - Secretariat of the World Meteorological Organization – Geneva – Switzerland - 2006.