NATIONAL TECHNICAL UNIVERSITY OF ATHENS
SCHOOL OF CIVIL ENGINEERING
Department of Water Resources and Environmental Engineering

Investigation of the stochastic behaviour of surface wind speed using K-moments

HS3.3 – Stochastic modelling and real-time control of complex environmental systems

- Faidon Diakomopoulos | Undergraduate student
- Panayiotis Dimitriadis | Post-Doctoral Researcher
- Theano Iliopoulou | PhD Candidate
- Demetris Koutsoyiannis | Professor

Athens, May 2020

Summary of topics

1. Introduction - Dataset
2. Presentation of distributions
3. Methodology of K-moments
4. Results-Conclusions
5. References

Introduction-Dataset

Currently, more and more countries make a shift towards renewable energy to reduce the environmental impact of the use of the fossil fuels. Wind energy has a significant position in this hierarchy, as one of the most efficient to be converted to electric energy, covering the society's needs in the fields of transportation, trade and consumption.

The probabilistic distributions of wind speed are a critical piece of information needed in the assessment of wind energy potential. According to the literature, the distribution that is widely used to imitate the behaviour of wind speed is two parameter Weibull. In this work we use various distributions such as *Weibull, Rayleigh, Gamma, LogNormal, Generalized Pareto, Pareto, Nakagami, Generalized Gamma and Pareto Burr Feller* distribution.

For the comparison of the distributions, we evaluate the fitting performance of these, on the empirical distribution from the data of the MIT station. To achieve this goal, we two goodness of fit parameters in statistics analysis, chi square (χ^2) and root mean square error (RMSE). The expressions of these parameters are illustrated, analytically, in the second chapter of the presentation.

Introduction-Dataset

As dataset it was chosen the observations of MIT station in Cambridge, Massachussets, USA. This station looks credible and includes a high number of observed values (hourly) and a low percentage of zero values (1.03%).

The number of non-zero observations is 584196 with a minimum value of 0.1667 m/s and maximum value of 38.6 m/s. In the following table, the statistical characteristics of the dataset are shown:

MIT WIND DATA STATION					
Period of observations: 71 years (1943-2014)					
MIN	MAX	MEAN	STANDARD DEVIATION	COEFFICIENT OF KURTOSIS	COEFFICIENT OF SKEWNESS
0.1667	38.6	5.4780	2.4307	4.7178	0.8988

Summary of topics

1. Introduction - Dataset
2. Presentation of distributions
3. Methodology of K-moments
4. Results-Conclusions
5. References

Presentation of distributions

- *Weibull distribution*

The most common distribution for the research of the behaviour of surface wind speed in the literature. Akpinar et al.(2009), Xu et al.(2015), Carta et al.(2009), Alavi et al.(2016), Mert et al.(2014), Conradsen et al.(1984), Amaya-Martinez et al.(2014), Lechner et al.(1992), Campisi-Pinto et al.(2020) model wind speed with the two parameter Weibull distribution.

The mathematic expressions of cdf and pdf are given, respectively by Conradsen et al(1984):

cdf: $F(x;a,b) = 1 - \exp\left(-\left(\frac{x}{b}\right)^{a}\right)$

pdf: $f(x;a,b) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^{a}\right)$

where x, a, b are wind speed, shape parameter and scale parameter, respectively.

- *Rayleigh distribution*

The Rayleigh distribution was originally derived by Lord Rayleigh, and can be expressed as a special case of Weibull distribution Ganji et al.(2016) with scale parameter $b = 2$ and shape parameter $a = \sqrt{2}\sigma$

cdf: $F(x;\sigma) = 1 - \mathrm{e}^{-x^2/(2\sigma^2)}$

pdf: $f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}$

Presentation of distributions

- *Gamma distribution*

Another one distribution that has been proposed from Amaya-Martinez et al.(2014) and Alavi et al.(2016) and it seems to fit quite well the observed data values, is the Gamma distribution.

The mathematic expressions of cdf and pdf are the followings:

cdf: $F(x; a, b) = \gamma\,(a, b, x)/\Gamma\,(a)$

pdf: $f(x; a, b) = \frac{b^a x^{a-1} \mathrm{e}^{-bx}}{\Gamma(a)}$

where $\gamma(a, b, x)$ is the lower incomplete gamma function.

- *LogNormal distribution*

This distribution has not been proposed in many articles about wind speed, although is referenced in Carta et al.(2009) and Alavi et al.(2016) and we use, not the hybrid, but the classic type in this work, without the zero values of the dataset.

cdf: $F(x; m, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\log x - m)^2}{2\sigma^2}\right)$

pdf: $f(x; m, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\log x - m)^2}{2\sigma^2}\right)$

Presentation of distributions

➢ *Generalized Pareto distribution*

Generalized Pareto distribution is a distribution, which is used for extreme phenomena like floods and extremes rainfalls, due to is close relationship with extreme values. For wind speed was used by Lechner et al.(1992) and Holmes et al.(1999) and it contained in our work.

The mathematic expressions of cdf and pdf, as the are given by Hosking et al.(1987), are the followings:

cdf: $F(x; a, k) = 1 - (1 - k\,x/a)^{1/k}$

pdf: $f(x; a, k) = \frac{1}{a}\left(1 - \frac{kx}{a}\right)^{\frac{1}{k}-1}$

➢ *Pareto distribution*

More precisely, is called pareto II or Lomax and has close relationship with Generalized Pareto distribution concerning the extreme values.

The mathematic expressions of cdf and pdf, as they are given by Koutsoyiannis.(2019a):

cdf: $F(x; \lambda, k) = 1 - \left(1 + \frac{kx}{\lambda}\right)^{-1/k}$

pdf: $f(x; \lambda, k) = \frac{1}{\lambda}\left(1 + \frac{kx}{\lambda}\right)^{-1/k-1}$

Presentation of distributions

- *Generalized Gamma distribution*

The generalized gamma distribution has been proposed by Mert et al.(2014) and Campisi-Pinto et al.(2020) and is a generalization of Weibull distribution with three parameters.

The expression of the pdf is given by Mert et al.(2014):

cdf: $F(x; a, b, k) = \gamma\left((x/b)^k\right)/\Gamma\,(a)$

pdf: $f(x; a, b, k) = kx^{a-1}\mathrm{ex}\rho\left(-(x/b)^k\right)/\left(b^{ka}\Gamma(a)\right)$

- *Nakagami distribution*

This distribution is a special case of generalized gamma distribution with two parameters and also has close relationship with Weibull distribution and its flexibility. Introduced for first time to imitate the behavior of wind speed by Alavi et al.(2016)

cdf: $F(x; m, \Omega) = 1 - \gamma\,(m, m\,x^2/\Omega)/\Gamma\,(m)$

pdf: $f(x; m, \Omega) = \frac{2m^m}{\Gamma(m)\Omega^m}x^{2m-1}\mathrm{e}^{\left(-\frac{m}{\Omega}x^2\right)}$

where $\Omega = E[\underline{x}^2]$ and $m = E\left[\underline{x}^2\right]^2/\left(E[\underline{x}]^2 - E[\underline{x}^2]\right)^2$ $m \geq \frac{1}{2}$.

Fitting of distributions

➢ *Pareto Burr Feller (PBF)*

A three parameter distribution which is also known as Pareto III, Burr XII and Feller. The name "PBF" was introduced by Koutsoyiannis et al.(2018a), and is an excellent distribution to describe the wind speed behavior combining the asymptotic properties of Weibull distribution for low wind speeds and Pareto properties for large ones (Koutsoyiannis et al.(2018a)).

The mathematic expressions of cdf and pdf are the followings:

cdf: $F(x;\lambda,z,\xi) = 1 - \left(1+\left(\frac{x}{\lambda}\right)^{z}\right)^{-\frac{1}{z\xi}}$

pdf: $f(x;\lambda,z,\xi) = \frac{1}{\lambda\xi}\left(\frac{x}{\lambda}\right)^{z-1}\left(1+\left(\frac{x}{\lambda}\right)^{z}\right)^{-\frac{1}{z\xi}-1}$

- *Goodness of fit parameters*

We use two parameters to compare how the predicted values of the previous distributions fit to the observed data from the station. According to literature, two of the most credible and usable parameters are RMSE and chi square error (χ^2).

$RMSE$ is given by Costa Rocha et al.(2011): $RMSE = \frac{1}{N}\left(\sum_{i=1}^{N}(y_i - x_i)^2\right)^{1/2}$ where y_i are the predicted and x_i is the observed values.

χ^2 is given by Thomson et al.(2014): $\chi^2 = \sum_{i=1}^{N}\frac{(f_i - F_i)^2}{F_i}$ where f_i is the ith predicted value and F_i is the observed, respectively.

Fitting of distributions

- *Maximum Likelihood estimation*

The method estimates the parameters to maximize the likelihood function of independent randomly observed sample.

Because of the complexity in the solution of partial differential equations to achieve the previous goal, we choose to maximize the logarithm of the function, Koutsoyiannis.(1996):

$$L(x_1, \dots, x_n, \theta_1, \dots, \theta_p) = \ln f_{x_1,\dots,x_n}(x_1, \dots, x_n, \theta_1, \dots, \theta_n) = \sum_{i=1}^{n} l_n f_x(x_{i_1}\theta_1, \dots, \theta_r)$$

Equation to maximize the L() logarithmic likelihood:

$$\frac{\partial L(x_1,\dots,x_n,\theta_1,\dots,\theta_r)}{\partial \theta_k} = \sum\nolimits_{i=1}^{n} \frac{1}{f_x(x_i,\theta_1,\dots,\theta_r)} \frac{f_x(x_i,\theta_1,\dots,\theta_r)}{\partial \theta_k}$$

Summary of topics

1. Introduction - Dataset
2. Presentation of distributions
3. Methodology of K-moments
4. Results-Conclusions
5. References

Methodology of K-moments

The classical moments, even if useful as theoretical concepts their estimation is not reliable for moment order beyond 2 or three Lombardo et al.(2014). For this reason the classical moments for higher than this order have been termed “unknowable” by Koutsoyiannis et al.(2019b).

In contrast, K-moments allow estimation for high order statistics p, with significant reliability. Something that justifies and their name “Knowable”. The definitions of K-moments by Koutsoyiannis.(2018b) are the followings:

Noncentral K-moments are defined as:

$$K'_{\rho q} = (\rho - q + 1)E\left[\left(F\left(\underline{x}\right)\right)^{\rho - q}\ \underline{x}^{q}\right]$$

With the same way, central K-moments are defined as:

$$K_{\rho q} = (\rho - q + 1)E\left[F\left(\underline{x}\right)^{\rho - q}\left(\underline{x} - \mu\right)^{q}\right]$$

Finally, the definition of tail-based (noncentral) K-moments is the following:

$$\bar{K}'_{\rho q} = (\rho - q + 1)E\left[\left(\bar{F}(x)\right)^{\rho - q} x^{q}\right]$$

Methodology of K-moments

The creation of unbiased estimators of K-moments is based on the analysis by Landwehr et al.(1979) for the production of unbiased estimators on probability weighted moments and L-moments.

The combination of the definition of K-moments with order statistics, and more specifically the quantity $F(\underline{x})^{p-q}$ with the arrangement of the sample in ascending order we can think the estimators of the form:

$$\hat{\underline{K}}'_{\rho q} = \frac{\rho - q + 1}{n} \sum_{i=1}^{n} \left(F\left(\underline{x}_{(i:n)} \right) \right)^{\rho - q} \underline{x}_{(i:n)}^{q}$$

$$\hat{\underline{K}}_{\rho q} = \frac{\rho - q + 1}{n} \sum_{i=1}^{n} \left(F\left(\underline{x}_{(i:n)} \right) \right)^{\rho - q} \left(\underline{x}_{(i:n)} - \hat{\mu} \right)^{q}$$

$$\hat{\underline{K}}^{+}_{\rho q} = \frac{\rho - q + 1}{n} \sum_{i=1}^{n} \left(2F\left(\underline{x}_{(i:n)} \right) - 1 \right)^{\rho - q} \left(\underline{x}_{(i:n)} - \hat{\mu} \right)^{q}$$

In our work, we use the noncentral estimator of K-moments which does not depend on $\underline{x}_{(i:n)}$ but only in *i* and *n* with the introduced of parameter $b_{i,n,\rho}$ to create a new definition: $i \geq \rho \geq 0$

$$\hat{\underline{K}}'_{\rho q} = \sum_{i=1}^{n} b_{i,n,\rho - q + 1} \underline{x}_{(i:n)}^{q}$$

with $b_{in\rho} = \frac{\rho}{n} \frac{\Gamma(n-\rho+1)}{\Gamma(n)} \frac{\Gamma(\mathrm{i})}{\Gamma(\mathrm{i}-\rho+1)}$ $\mathrm{i} \geq \rho \geq 0$ $q = 1$.

Methodology of K-moments

- Return periods

The final step of the method is to find the empirical and theoretical return periods and to succeed the minimum error of the fitting with each other with the RMSE method.

The expression for the theoretical return period:

$$\frac{T(x)}{D} = \frac{1}{1 - F(x)}$$

where F is the cumulative distribution function.
The empirical return period for the estimated K-moments can be expressed by the relationship:

$$\frac{T\left(K_\rho'\right)}{D} = \rho \cdot \Lambda_\rho$$

where ρ is the moment order.

For the approximation of the quantity Λ_ρ, we can use the following relationship which is given by Koutsoyiannis.(2018b):

$$\Lambda_\rho = \Lambda_\infty + (\Lambda_1 - \Lambda_\infty)\frac{1}{\rho}$$

Methodology of K-moments

In the following table are referenced the value of Λ_1, Λ_∞,β,B for a range of functions which have been used in our work and the are illustrated in the next chapter.

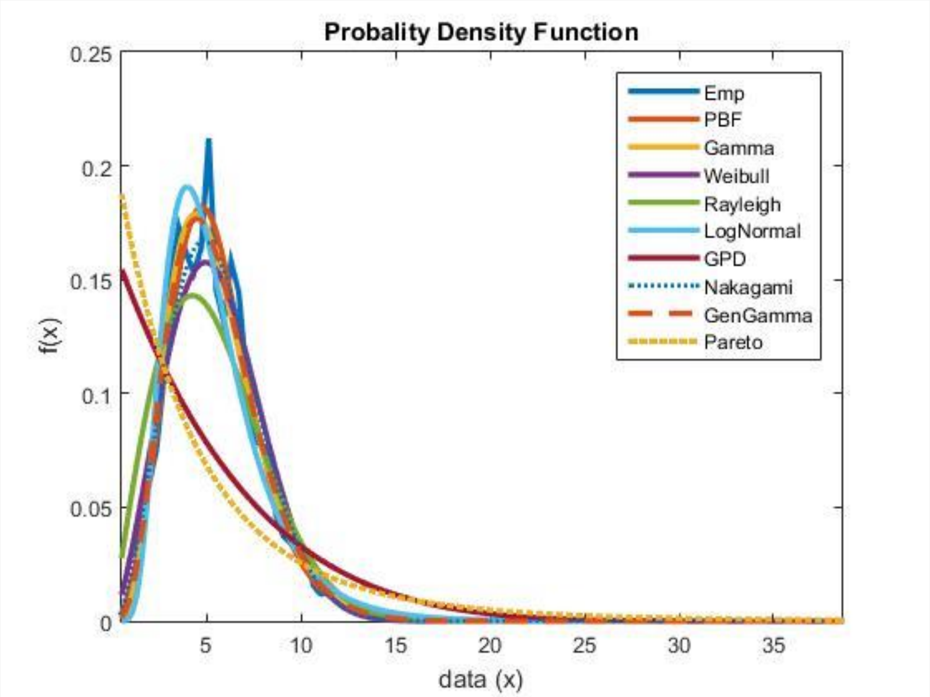
Function	Λ_1	Λ_∞	β	B
PBF	$\left(1+\left(\frac{B\left(\frac{1}{z\xi}-\frac{1}{z},\frac{1}{z}\right)}{z}\right)^{z}\right)^{\frac{1}{z\xi}}$	$\Gamma(1-\xi)^{-\frac{1}{\xi}}$	-	-
GAMMA	$\Gamma\,(a)/\gamma\,(a,a)$	1.781	0	$-0{,}154\ln a$
WEIBULL	$\mathrm{e}^{\left(\Gamma\left(1+\frac{1}{b}\right)\right)^{b}}$	1.781	0	$\frac{1-\mathrm{e}^{\gamma}}{\ln 2}-0{,}07+\frac{0.92}{\zeta}+\frac{0{,}23}{\zeta^{2}}$ $\mathrm{e}^{\gamma}=1{,}781$

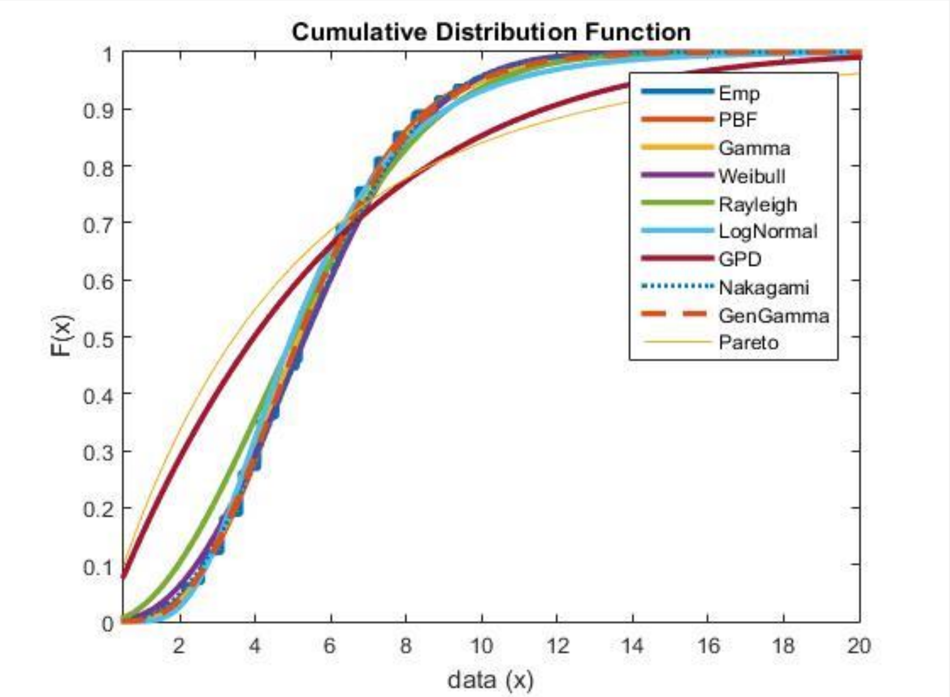
Finally, we estimate the parameters of every function by minimizing the mean square error of the logarithms of the empirical and theoretical return period *focusing on extremes (the tail of the distribution).*

Summary of topics

Results-Conclusions

The left diagram shows the fitting of all the probability density functions with the empirical distribution from the observed data. The right one shows the same functions for the cumulative distribution function, correspondingly. The parameters of the functions have been estimated from maximum likelihood method.

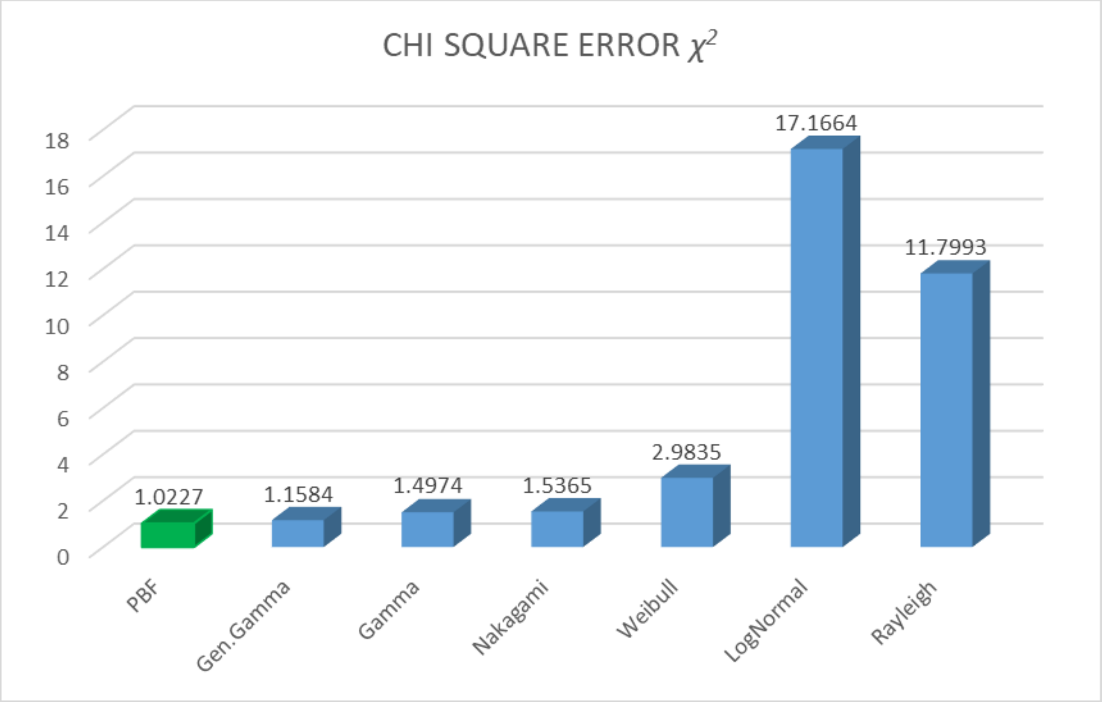


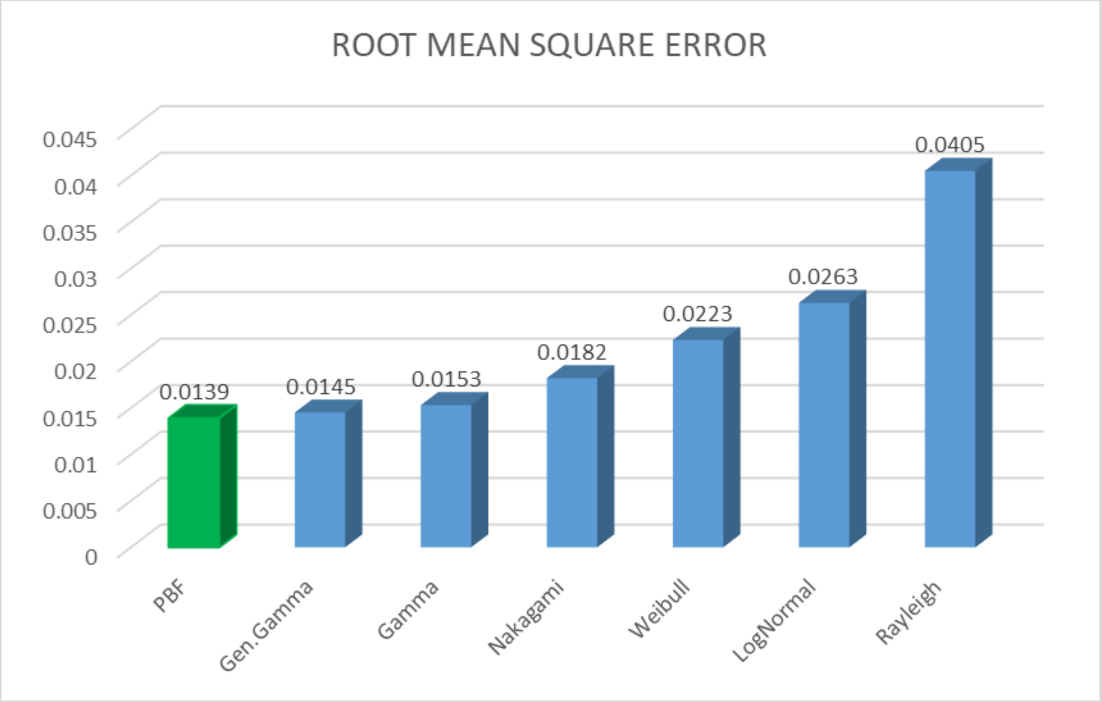


As a general remark, the GPD and Pareto distributions are not suitable to express the behaviour of wind speed for the dataset and secondarily the Rayleigh distribution seems to fit not as well as the other 5 functions. The calculation of the error is the tool that could help to find the most effective distribution.

Results-Conclusions

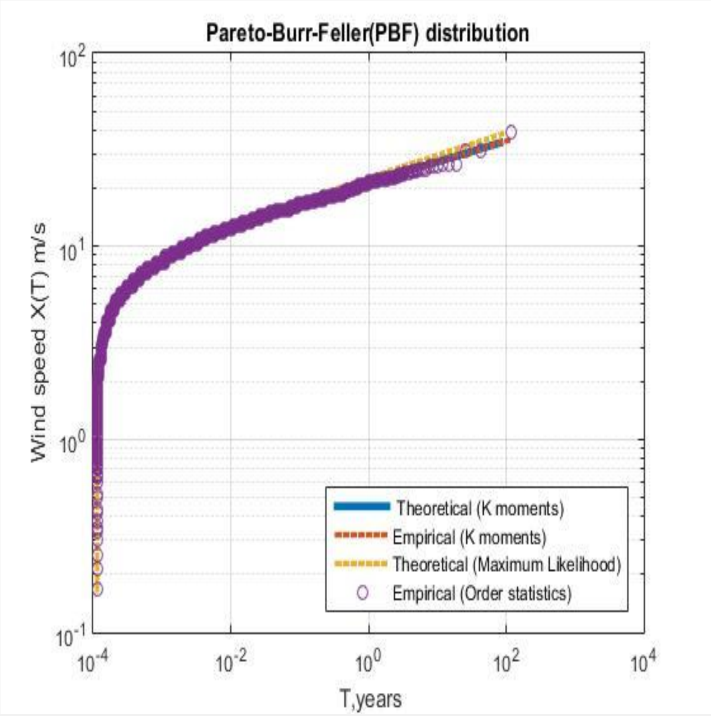
The two diagrams show the results of the two goodness of fit parameters. It is clear that the function with the best fit on the observed data is the Pareto Burr Feller (PBF), which is highlighted with green colour in the two diagrams. The distribution that gives the maximum error is Rayleigh, as it had been forecasted from the previous distribution diagrams.

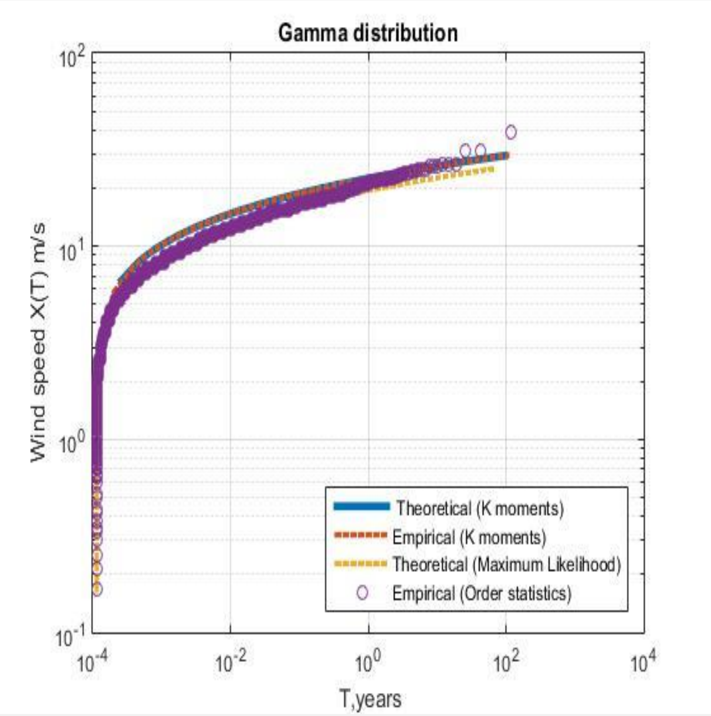


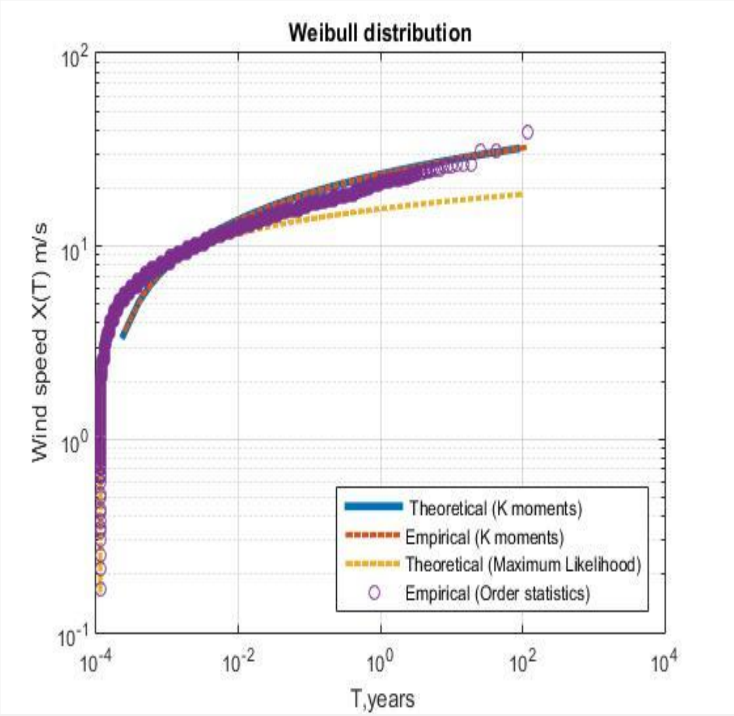


Results-Conclusions

In the next three diagrams the K-moments are presented for three distributions (PBF, Gamma and Weibull) to evaluate the behaviour of these in comparison with the curve of distributions with parameters from maximum likelihood method.



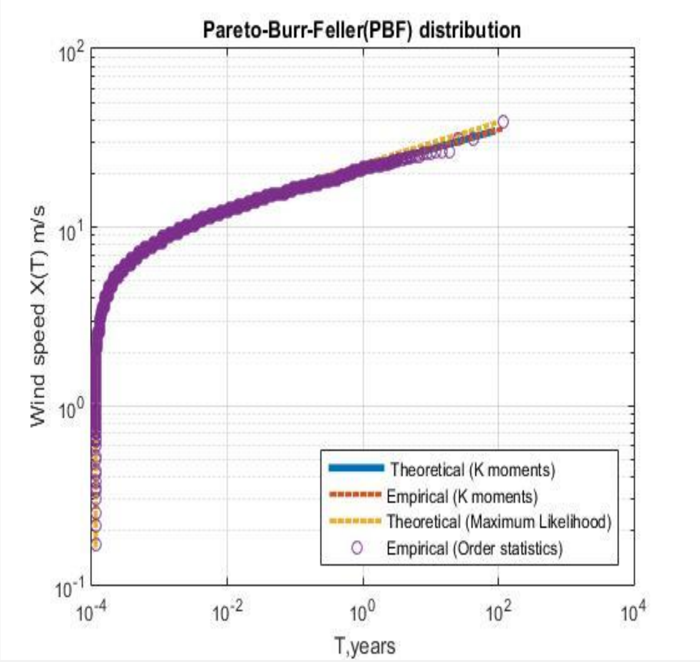


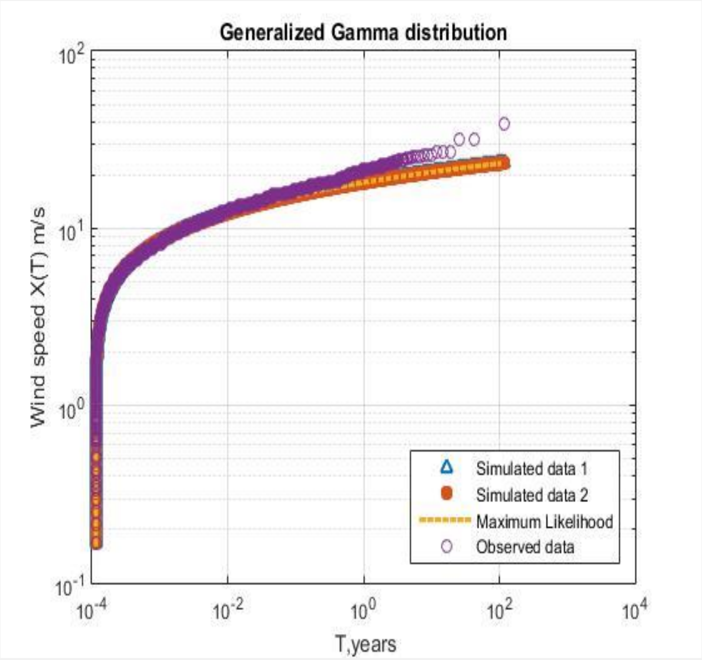


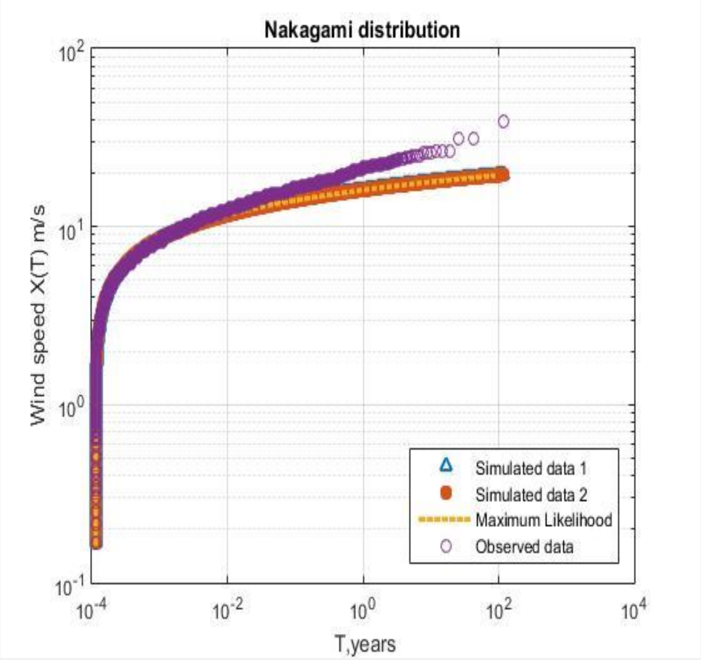
As general remark, it is obvious that for the three different distributions the curves with parameters which are estimated by K-moments concern the distribution tail, in contrast with the curves from maximum likelihood method which have a quite satisfactory behaviour for the body but are failed to fit in the distribution tail. In comparison of the three distributions the *PBF seems to fit perfect in the entire set of values, especialy the PBF curves from K-moments present an almost perfect fit and for large return periods, but both the curves from different methods are close to each other, showing that PBF is the appropriate function to imitate the wind speed behaviour.*

Results-Conclusions

The left diagram for the PBF distribution is the same as the previous page, but now it is compared with the diagrams of the other two distributions with the curves that has been produced from the estimated, of maximum likelihood method, parameters.







The order statistics of the left diagram are the same as the observed data in the other two. Using middle and right diagram's data, it is shown that the simulated values curves, in which the y-axis presents random values and x-axis the values of return period $\frac{T}{D} = \frac{1}{1-F}$, respectively, with the maximum likelihood's parameters, fit, exactly, in the curves of our dataset with parameters of maximum likelihood (orange split line). These curves, even if have quite good adapt in the body of distribution, in any case they cannot approach the extreme values, like the PBF.

Summary of topics

1. Introduction - Dataset
2. Presentation of distributions
3. Methodology of K-moments
4. Results-Conclusions
5. References

References

- Akpinar, S. and Akpinar, E.K., 2009. Estimation of wind energy potential using finite mixture distribution models, Elsevier, Energy Conversion and Management, Volume 50, Issue 4, Pages 877-884.
- Alavi, O., Mohammadi, K. and Mostafaeipour, A., 2016. Evaluating the suitability of wind speed probability distribution models: A case of study of east and southeast parts of Iran, Elsevier, Energy Conversion and Management, Volume 119, Pages 101-108 .
- Amaya-Martínez, P. A., Saavedra-Montes, A. J. and Arango-Zuluaga, E. I., 2014. A statistical analysis of wind speed distribution models in the Aburrá Valley, Colombia. CT&F - Ciencia, Tecnología y Futuro, 5(5), 121-136.
- Campisi-Pinto, S., Gianchandani, K. and Ashkenazy, Y., 2020. Statistical tests for the distribution of surface wind and current speeds across the globe, Elsevier, Renewable Energy, Volume 149, Pages 861-876
- Carta, J.A., Ramirez, P. and Velazquez, S.,2009. A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands, Elsevier, Renewable and Sustainable Energy Reviews, Volume 13, Issue 5, Pages 933-955.
- Conradsen, K., Nielsen, L.B. and Prahm, L.P., 1984. Review of Weibull Statistics for Estimation of Wind Speed Distributions, American Meteorological Society, Journals Online, https://doi.org/10.1175/1520-0450(1984)023<1173:ROWSFE>2.0.CO;2.
- Costa-Rocha, P.A., Coelho de Sousa, R., Freitas de Andrade, C. and Vieira da Silva, M.E., 2012. Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil, Elsevier, Applied Energy Volume 89, Issue 1, Pages 395-400
- Holmes, J.D. and Moriarty, W.W., 1999, Application of the generalized Pareto distribution to extreme value analysis in wind engineering, Elsevier, Journal of Wind Engineering and Industrial Aerodynamics,Volume 83, Issues 1–3, Pages 1-10.
- Hosking, J.R.M. and Wallis, J.R., 1987. Parameter and Quantile Estimation for the Generalized Pareto Distribution, Technometrics, Vol. 29, No. 3, pp. 339-349.
- Ganji, M., Bevrani, H., Hami-Golzar, N. and Zabihi, S., 2016. The Weibull–Rayleigh Distribution, Some Properties, and Applications, J Math Sci 218, 269–277, https://doi.org/10.1007/s10958-016-3028-2 .
- Koutsoyiannis, D., 1996. Statistical Hydrology, Edition 3, National Technical University of Athens, pp 76.
- Koutsoyiannis, D., Dimitriadis, P., Lombardo, F. and Stevens, S., 2018a. From Fractals to Stochastics: Seeking Theoretical Consistency in Analysis of Geophysical Data, Springer International Publishing, DOI 10.1007/978-3-319-58895-7_14.
- Koutsoyiannis, D.,2018b. Modelling extreme rainfall in the era of climate change concerns: Towards a consistent stochastic methodology, School for Young Scientists “Modelling and forecasting of river flows and managing hydrological risks: Towards a new generation of methods , Russian Academy of Sciences, Lomonosov Moscow State University, Moscow, doi:10.13140/RG.2.2.22015.25766.
- Koutsoyiannis, D., 2019a. Extreme-oriented selection and fitting of probability distributions, Presentation for EGU.
- Koutsoyiannis, D., 2019b. Knowable moments for high-order stochastic characterization and modelling of hydrological processes, Hydrological Sciences Journal, 64 (1), 19–33, doi:10.1080/02626667.2018.1556794.

References

- Landwehr, J.M., Matalas, N.C., and Wallis, J.R., 1979. Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resour. Res., 15(5), 1055-1064.
- Lechner, J.A., Leigh, S.D. and Simiu, E., 1992. Recent Approaches to Extreme Value Estimation with Application to Wind Speeds,. Part I: the Pickands Method, Elsevier, Journal of Wind Engineering and Industrial Aerodynamics, 41-44, 509-519.
- Lombardo, F., Volpi, E., Koutsoyiannis, D., and Papalexiou, S.M., 2014. Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. Hydrology and Earth System Sciences, 18, 243–255, doi: 10.5194/hess-18-243-2014.
- Mert, I. and Karakus, C., 2015. A statistical analysis of wind speed data using Burr, generalized gamma, and Weibull distributions in Antakya, Turkey, Tubitak, doi:10.3906/elk-1402-66.
- Thompson, R. and Emery, W., 2014. Goodness-of-Fit Test, Elsevier, Data Analysis Methods in Physical Oceanography (Third Edition), Pages 219-311.
- Xu, X., Yan, Z. and Xu, S., 2015. Estimating wind speed probability distribution by diffusion-based kernel density method, Elsevier, Electric Power Systems Research, Volume 121, Pages 28-37.

THANK YOU