Received: 4 July 2020 | Accepted: 13 July 2020
DOI: 10.1002/env.2647

DISCUSSION PAPER

WILEY

Discussion on A high-resolution bilevel skew-*t* stochastic generator for assessing Saudi Arabia's wind energy resources

Adelchi Azzalini

Dipartimento di Scienze Statistiche, Università di Padova, Padova, Italia

Correspondence
Adelchi Azzalini, Dipartimento di Scienze Statistiche, Università di Padova, Padova, Italia.
Email: azzalini@stat.unipd.it

General considerations

It was a pleasure to read this interesting and stimulating paper, and I congratulate the authors on their valuable work.

My perspective is situated in the context of flexible parametric families of probability distributions, where typically new theoretical constructions are presented alongside some small-scale numerical illustration(s) aimed at providing some supporting evidence of the practical usefulness or some interesting feature of the proposed formulation. These illustrations, while possibly useful in the above-indicated role, hardly ever can realistically be regarded as true applications of the probability distribution under consideration, given the limited subject-matter motivation and the lack of implications on the applied side. At the opposite side of the range, in a sense, there are truly applied works which employ these probability distributions in a marginal role, typically for mere numerical data fitting, but without a link with the constructive probabilistic mechanism of these parametric families, so that the adopted parametric family could often be replaced by another one with essentially no difference.

The present contribution does not fit into any of the two roles indicated above. It represents instead one of the few instances where the project is motivated by a large-scale application of potentially major impact in the real world, and the model formulation blends seamlessly subject-matter considerations with formal properties of the stochastic component. This aspect is especially prominent when one looks at the formulation of Section 3.3, which we shall examine shortly.

Another appealing aspect of this work lies in its multidisciplinary organization. While I am totally unable to comment on aspects outside the precincts of statistics and probability, still I am pleased to see that this cross-disciplinary interaction has successfully taken place. Also, one cannot avoid noticing the massive effort which collectively must have gone into this project.

More specific remarks

Section 3.3 provides the framework of the stochastic model; equation (2) and the implied equation (3) represent the core component. Here, it is appropriate to point out the connection with the construction of Zhang and El-Shaarawi (2010), in the sense that the numerator of (2) is built as a suitable nonlinear combination of two independent Gaussian components which is conceptually analogous to equation (3) of Zhang and El-Shaarawi (2010). There is a difference in that here the leading term $\lambda_r\ |U_r|$ varies with the region r and not the specific location s; however, this is not a conceptually crucial aspect. The decisive difference between the two formulations is the introduction here of the random term in the denominator of (2), which leads to a skew-*t* marginal distribution, which substantial increase in flexibility of the model.

A second remark is about parameter estimation, which is a particularly challenging process with massive data sets of this sort, and requires the use various devices to keep the working under control. A specific but relevant instance of this sort is described in the final part of Section 4.1 and more extensively in the Supplementary Material, where the

 © 2020 John Wiley & Sons, Ltd.

authors find it useful to obtain initial estimates by numerical inversion of sample low-order cumulants. In this logic, it is worth mentioning that more reliable estimates are likely to be obtained by employing sample quantiles in place of sample moments, which are known to have instability issues with long-tailed distribution. A procedure which maps quantiles to parameters of the skew-*t* distribution is presented by Azzalini and Salehi (2020).

The final annotation is presented in a somewhat dubitative form, since I am not sure of fully grasping the underlying mechanism. In the simulation work of Section 3.4, the data are sampled from the SKT model described in the first two paragraphs and regarded as "the true data," to which a Gaussian and a SKT model are fitted. Since the SKT sampling mechanism implies a nonzero location of the random term, this can only be interpreted as a systematic component by the Gaussian model, incorporated into the fixed parameters to be estimated. Could it then be that the systematic bias pointed out at the end of the section and illustrated in Figure 4 is due to this effect? In this case, this bias would be linked to the chosen sampling scheme, which this matches the fitted SKT model, while the Gaussian model is inevitably disadvantaged in this respect. If this reading is confirmed, a fairer comparison would be setup by simulating data from a shifted SKT error term, such that its median value is 0, like for the Gaussian distribution.

REFERENCE

Azzalini, A., & Salehi, M. (2020). *Some computational aspects of maximum likelihood estimation of the skew*-t *distribution*. In A. Bekker, D.-G. Chen, & J. T. Ferreira (Eds.), *Computational and methodological statistics and biostatistics*. New York, NY: Springer International Publishing.

How to cite this article: Azzalini A. Discussion on A high-resolution bilevel skew-*t* stochastic generator for assessing Saudi Arabia's wind energy resources. *Environmetrics*. 2020;31:e2647. https://doi.org/10.1002/env.2647