Rejoinder on the discussion of 'A high-resolution bilevel skew-t stochastic generator for assessing Saudi Arabia's wind energy resources'

Felipe Tagle[1], Marc G. Genton[2], Andrew Yip[3], Suleiman Mostamandi[3], Georgiy Stenchikov[3] and Stefano Castruccio[1]

August 26, 2020

We would like to thank all four discussants for their discussions on different aspects of our paper. We were particularly pleased to see comments across the entire spectrum of our work, from theoretical properties of the model to geoscience, engineering and wind energy aspects that could provide a pathway to implement a wind energy portfolio in Saudi Arabia, truly reflecting the mission of *Environmetrics*. Throughout this rejoinder, we will denote the discussion of Adelichi Azzalini as A, Sándor Baran as B, Emilio Porcu, Jonas Rysgaard and Valerie Eveloy as PRE and Andew Zammit-Magion as Z.

This rejoinder covers the different points raised by the discussants in different sections. Section 1 discusses theoretical and methodological aspects of our model. Section 2 focuses on inference and also shows new results with a Bayesian implementation with Stan (Carpenter et al. 2017). Section 3 replies to discussion points on validation. Finally, Section 4 covers aspects related to geoscience, engineering and wind energy stemming from this work.

[1]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, United States.
E-mail: scastruc@nd.edu

[2]Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.

[3]Atmospheric and Climate Modeling Group, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2015-CRG4-2640.

1 Modeling aspects

The choice of the model has been motivated by the need to strike a balance between the flexibility of a theoretical construct for a new class of parametric families and practical necessities in the big picture of a large interdisciplinary project. As remarked in the introductory comment by A, it is challenging to propose models in environmental statistics with theoretically justified models, which can also be realistically employed in applications whose focus is beyond marginal distributions. As also noted by A, our model indeed can be seen as conceptually similar to Zhang & El-Shaarawi (2010), with a random denominator allowing for a skew-t marginal distribution.

PRE has proposed a classification of both this model and the alternative formulation of Tagle et al. (2020) in terms of a more general class of additive models, with a common random effect across a partition of the domain. As a result, they highlighted how one of the main shortcomings of the model, namely the presence of discontinuity at the borders of the domain, in this model is shared across this more general class. This limitation, also discussed by Z is practically unavoidable if one is to avoid the latent Gaussian model approach, and has the appealing advantage of allowing to express moments in closed form.

PRE also provided a detailed discussion about the choice of the clusters, advocating for the use of topic-relevant information (when available). To support this point, we would like to offer an example of a multi-resolution model belonging to the proposed class in the context of functional Magnetic Resonance Imaging (fMRI), using a notation similar (not identical due to the presence of a convex combination) to that PRE for the ease of comprehension, and removing any issue of time dependence and trends to simplify the notation. In Castruccio et al. (2018) a mean-zero bi-resolution Gaussian process was proposed

$$X(\mathbf{s}) = (1 - \omega_{r(\mathbf{s})})V_{r(\mathbf{s})} + \omega_{r(\mathbf{s})}\xi_{r(\mathbf{s})}(\mathbf{s}),$$

where the partition $r(\mathbf{s})$ was provided by the Regions of Interest (ROIs) in the brain, and the

fine scale resolution represented the voxel intensity. The spatial process was expressed as a convex combination with parameters $\omega_{r(\mathbf{s})} \in [0, 1]$ (depending on the ROI) between a ROI-specific Gaussian random effect $V_{r(\mathbf{s})} \sim \mathcal{N}_R(\mathbf{0}, \mathbf{\Sigma}_V)$ and a fine-scale Gaussian field $\xi_{r(\mathbf{s})}(\mathbf{s}) \sim \mathcal{N}_{|\mathcal{D}_\ell|}(\mathbf{0}, \mathbf{\Sigma}_\xi)$, where $|\mathcal{D}_\ell|$ is the cardinality of the region $\mathcal{D}_\ell$. Inference was performed in a two-step fashion with graphical LASSO to estimate the connectivity among ROIs. While this model is more limited in its marginal properties than the one we propose here, it is a good example of a bi-resolution model in the general class proposed by PRE where the choice of region is not dictated by a clustering algorithm, but it already provided by the ROIs, regions in the brain localizing different motor and cognitive tasks. Even in this case, however, one may argue that the choice of the partition (*parcellation*) is not unique, and it should be chosen depending on the particular investigation a scientist is interested in.

Z also remarked that the model does not allow for temporal varying dynamics, and while this is appropriate in the presented work (as shown in the Figure S3), this may be an issue for higher resolution (hourly or sub-hourly) wind. While we ourselves were relatively surprised by the regularity of the temporal dynamics, this seems consistent with the results from other data sets for multi-decadal wind projections, which do not show significant trends across years. It is likely that this could become an issue for hourly data, and several possible generalizations of our model could accommodate for that. Perhaps the simplest approach would be to detect change points in the dynamics and assume a piecewise constant dynamics. While this solution would induce discontinuity, it would preserve a computationally convenient expression of the likelihood.

Z also pointed out that covariances could be used to improve the model's explanatory power. The main modeling challenge compared to a Gaussian (or latent Gaussian) model is that the addition of covariates would impact not just the mean of the model, but also the covariance and other higher-order properties. Therefore, even the simplest approach of modifying equation (2)

so that

$$Y_r(\mathbf{s}) = \sum_{i=1}^{p} \beta_i X_i(\mathbf{s}) + \frac{\lambda_r |U_r| + \eta(\mathbf{s})}{\sqrt{Z_r}},$$

would imply a nontrivial recomputation of the moments. Nevertheless, since height is expected to play a key role in the determination of both mean and variability of wind fields, we agree that the use of static and possibly dynamic covariates could improve the fidelity of the stochastic generator.

2 Inferential aspects

The use of the EM algorithm for inference has made necessary the development of semi-closed form expression for both the expectation and the maximization step, with numerical approximation schemes of some terms, hence resulting in non-negligible computationally demanding resources. Several discussants have suggested alternative methods to ameliorate the computational burden. A proposed the determination of the initial estimates by using sample quantiles instead of sample moments, a more stable choice for long-tailed distribution (Azzalini & Salehi 2020).

Z discussed Laplace approximation and stochastic variational Bayes as alternatives to our EM algorithm. Both choices have relative merits in their ability to achieve approximate Bayesian inference, either with or without a Gaussian approximation. Additionally to Z's proposals we would like to offer an alternative, more automatic albeit less computationally efficient approach with Hamiltonian Monte Carlo inference with the Stan software in R. As part of this new implementation, we also offer a comparison with a Gaussian-log-Gaussian (GLG) model (Palacios & Steel 2006), which models non Gaussianity with a lognormal scale mixing of a Gaussian process.

2.1 A Bayesian bi-resolution skew-t model with Hamiltonian Monte Carlo Inference

Similarly to Tagle et al. (2020), we consider a partition $\mathcal{R}$ of cardinality R, where for each region $r \in \mathcal{R}$, and each $\mathbf{s} \in r$ we have the following model, which we denote as SKT

$$Y(\mathbf{s}) = \sigma_r \frac{\rho_r U_{0,r} + \lambda_r |U_{1,r}| + \eta_{2,r}(\mathbf{s})}{\sqrt{Z_r}}, \tag{1}$$

where $\sigma_r, \rho_r > 0$, $\lambda_r \in \mathbb{R}$. The model assumes that there exist scalar random variables $U_{1,r}$ and Z_r across each element r of the partition, and $\eta_{2,r}$, a stationary Gaussian process, with mean zero and correlation function $C_{\theta_r}(\cdot)$ with the associated covariance matrix $\mathbf{\Sigma}(\boldsymbol{\theta}_r)$. We assume independence of $U_{1,r}$ and Z_r across $\mathcal{R}$, as well as for the Gaussian process $\eta_{2,r}$. The large-scale effect, whose regional component is denoted by $U_{0,r}$ induces the desired spatial dependence, encoded in the correlation matrix $\mathbf{\Sigma}_0$.

If we denote by $\mathbf{Y}_r = \{Y(\mathbf{s}), \mathbf{s} \in r\}$, by $\mathbf{U}_0 = (U_{0,1}, \ldots, U_{0,R})^\top$, and by n_r the number of sampled sites in the region r, an equivalent hierarchical representation of (1) is given by,

$$\begin{aligned} \mathbf{Y}_r | U_{0,r} = u_{0,r}, U_{1,r} = u_{1,r}, Z_r = z_r &\sim \mathcal{N}_{n_r}\left(\sigma_r \tfrac{\rho_r u_{0,r} + \lambda_r |u_{1,r}|}{\sqrt{z_r}}, \tfrac{\sigma_r^2}{z_r}\mathbf{\Sigma}(\boldsymbol{\theta}_r)\right), \\ \mathbf{U}_0 &\sim \mathcal{N}_R(\mathbf{0}, \mathbf{\Sigma}_0), \\ U_{1,r} &\overset{iid}{\sim} \mathcal{N}(0,1), \\ Z_r &\overset{iid}{\sim} \text{Gamma}(\nu_r/2, \nu_r/2). \end{aligned} \tag{2}$$

The vector of parameters for each region is $\boldsymbol{\gamma}_r = (\sigma_r, \rho_r, \lambda_r, \nu_r, \boldsymbol{\theta}_r^\top)^\top$, with $\boldsymbol{\theta}_r$ collecting the parameters of the region-specific correlation function, and the total set of parameter consists of $\boldsymbol{\theta} = (\{\boldsymbol{\gamma}_r, r \in \mathcal{R}\}, \mathbf{\Sigma}_0)$. The non-negativity constraints on σ_r, ρ_r and ν_r motivate the use of prior distributions with support on $\mathbb{R}^+$. We use non-informative priors for all parameters, following weak prior suggestions in the Stan documentation. In particular, we choose normal random variables centered at the parameter true values with large variances (Stan automatically adds a truncation when a lower bound is specified). We run 3 chains for 2,000 iterations each, discarding the first 1000 as warm-up. Inference is performed on simulated data considering $N = 15, 50$, and 200 spatial replicates.

Computations were performed on a cluster with 2.3Ghz AMD EPYC 7451 24-Core Processors (48 threads), of which only 3 were used for any of the scripts, one for each of the noted 3 chains of the Stan inference procedure. Running the 2,000 iterations took between 90 to 120 minutes.

2.2 Simulation study

We firstly explore various parameter and regional configurations, with varying degrees of marginal skewness and kurtosis, as well as spatial dependence, to offer some guidance regarding the number of spatial replicates and the spatial design that leads to posterior distributions that are reasonably well-concentrated. Secondly, we investigate the degradation in predictive performance of a model which only allows for flexibility in capturing heavy-tailed behavior. In particular, we consider the GLG model of Palacios & Steel (2006)

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta} + \frac{\eta(\mathbf{s})}{\sqrt{\xi(\mathbf{s})}} + \varepsilon(\mathbf{s}),$$

where the first term represents a spatial trend, $\eta(\mathbf{s})$ is a stationary Gaussian process with mean 0, variance σ^2 and isotropic correlation function $C_\theta(h)$, $h = \|\mathbf{s}-\mathbf{s}'\| > 0$, $\mathbf{s}, \mathbf{s}' \in \mathcal{D}$, and the last term is an independent and uncorrelated Gaussian process with mean 0 and variance τ^2 representing the nugget effect. Here, $\xi(\mathbf{s})$ is such that $\log\{\xi(\mathbf{s})\}$ follows an independent Gaussian process with constant mean surface at $-\nu/2$ and covariance function $\nu C_\theta(h)$, with $\nu > 0$. Marginally, $\xi(\mathbf{s})$ follows a lognormal distribution with mean 1 and variance $\exp(\nu) - 1$. The marginal kurtosis of $Y(\mathbf{s})$ is given by $3\exp(\nu)$; thus large values of ν correspond to heavy-tailed behavior, while Gaussian tails are recovered in the limit as $\nu \to 0$. The covariance function $C(h)$, without a nugget effect, for the GLG model is given by

$$C(h) = \sigma^2 C_\theta(h)\, e^{\nu\left(1+\frac{1}{4}[C_\theta(h)-1]\right)}.$$

2.2.1 Parameter estimation

We compare the proposed model's posterior performance against the competing GLG model based on simulations drawn from the former model with varying degrees of skewness. Of course, we expect to see a degradation in performance for the GLG model but it is instructive to assess along which aspects such failings occur. We generate synthetic spatial replicates from the SKT model assuming a 12 region setup, each region a unit square, arranged in a 4 by 3 lattice configuration, and choose 15 points randomly from each unit square, for a total of 170 points. We assume a fixed value for the large-scale effect parameter, $\rho_r = 0.3$, as well as its range parameter, ϕ_0, set at 2. We consider a control simulation with no skewness, i.e., $\lambda_r = 0$, and test simulations with values $\lambda_r = 1.7, 1, 0.28$ for the first 3 regions, implying marginal skewness values of 0.15, 0.5 and 0.75, respectively, and 0 for the remaining regions. Because of the parameterization of the skew-t distribution, ν_r needs to be modified accordingly to keep the excess kurtosis fixed at 2 across all regions. In particular $\nu_r = 9.2, 7.8, 7$ for the first 3 regions and $\nu_r = 7$ for the remaining 9 regions. Lastly, we consider different values of the regional range parameter $\phi_r = 0.1, 0.5, 1$, to partially offset the increase in intra-regional spatial dependence that accompanies the increases in λ_r. From the control simulation we derive two fitted models, GLG-control and SKT-control, and similarly, we denote those from test simulations as GLG-skew and SKT-skew.

Figure 1 displays the 95% and 80% credible intervals for the fitted GLG models for simulations with $\phi_r = 0.1$. The posterior marginal distributions of ϕ_r for the first 3 regions of the GLG-skew model do not show clear departures in their median with the other regions nor with those of GLG-control, and a similar assessment can be made of their variability. Slight differences emerge in ν, as the presence of skewness results in reduced posterior medians for the affected regions. The first region, in particular, achieves the lowest regional value, falling below 0.1, implying a reduction in the implied kurtosis and spatial dependence. In contrast, the posterior medians of ρ_r for the skewness-affected regions show marked differences, with values exceeding 2 and 1 in

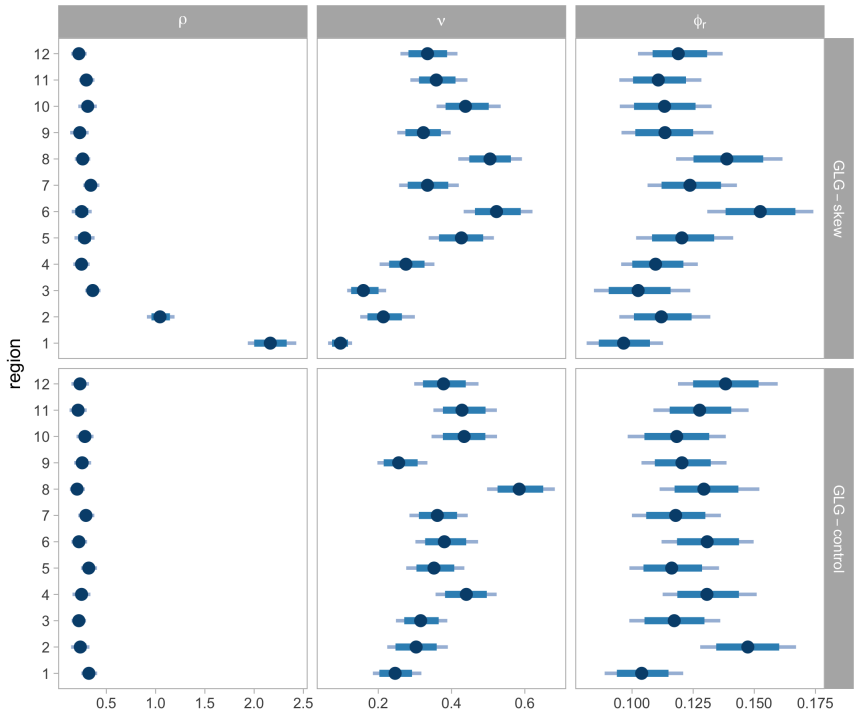

Figure 1: The posterior median, and the 95% (light blue) and 80% (dark blue) credible intervals for ρ, ν and ϕ for the each of the 12 regions, for the GLG-control and GLG-skew models.

the first and second region, respectively, while the rest are closely concentrated in the vicinity of 0.4. This represents a significant increase in the strength of the regional spatial dependence.

Figure 2 distills the impact of these parameter value differences on the implied spatial dependence, by displaying the difference between the correlation functions evaluated at the posterior samples and the true correlation functions. Due to space considerations, we only show the differences pertaining to the SKT simulation with $\phi_r = 1$ for the first 3 regions, and choose the 4th as representative of the remaining 9 regions. Significant biases can be seen for the correlation functions of the GLG-skew model in the first two regions, while the differences are indistinguishable from those of the fitted SKT-skew model for the third and beyond. These biases at distances larger than 0.2 approximately vary between 0.2 and 0.3 in the first region, and 0.1 and 0.25 for the second, across the 3,000 posterior samples.

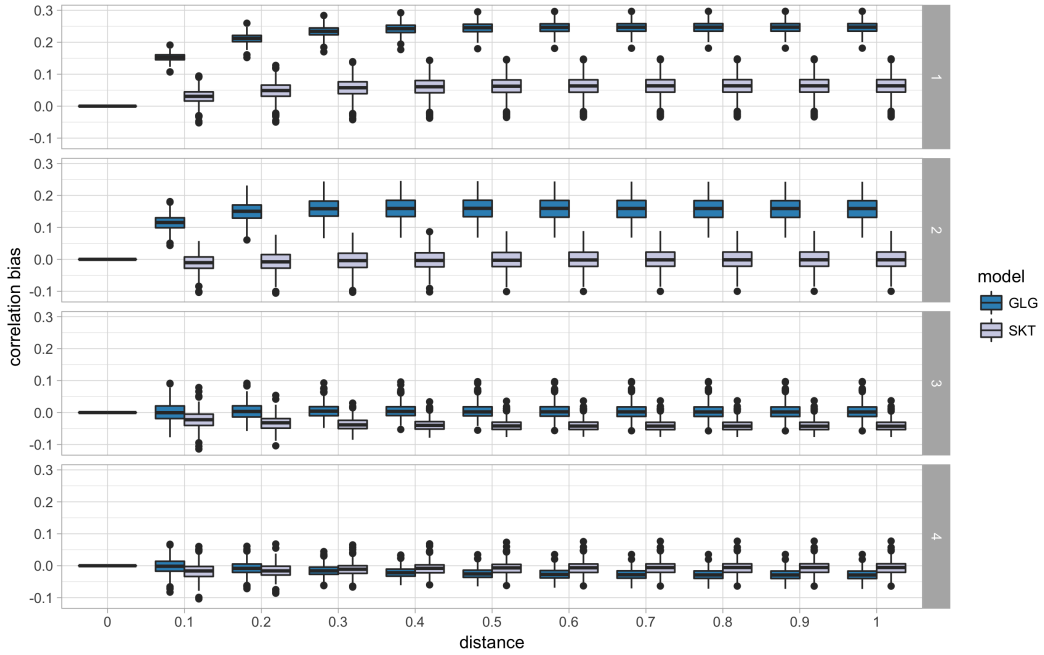

Figure 2: Difference between the correlation functions for the GLG-skew and SKT-skew models evaluated at the posterior samples of the SKT simulation with $\phi_r = 0.1$, and the true correlation functions underlying said simulation (displayed in Fig. S.1) for the first 4 regions.

2.2.2 Predictive performance

We compare posterior predictive performance using standard metrics, the continuous ranked probability score (CRPS; Gneiting et al. (2007)) and the mean absolute error (MAE). Table 1 shows the out-of-sample posterior predictive performance based on the CRPS and MAE for the GLG and SKT models for the various values of ϕ_r, for the first 3 regions and the range of values of the remaining 9. The effect of diminishing ϕ_r is clearly reflected in the predictive performance across statistics and fitted models: the greater the strength of spatial dependence, the better the predictive ability of the model. Comparing the performance of the models, the statistics clearly favor the SKT model, across all values of ϕ_r. Interestingly, a comparison using in-sample values did not show any degradation in predictive performance in neither of the two statistics, despite the GLG's inherent inability to match the marginal structure. An inspection of the posterior samples of the large scale effect U_0 revealed that in contrast with the model assumption of a centered normal distribution, the means were consistently positive across the replicates used for

the fitting, with an average of 0.72, underscoring the model misspecification.

Table 1: Out-of-sample continuous ranked probability score (CRPS) and mean-absolute error (MAE) values for the first three regions and the minimum and maximum of the other 9 regions.

Model	CRPS_1	CRPS_2	CRPS_3	CRPS_{min}	CRPS_{max}	MAE_1	MAE_2	MAE_3	MAE_{min}	MAE_{max}
GLG-skew ($\phi_r = 0.1$)	1.36	0.76	0.66	0.57	0.71	3.39	2.89	1.39	1.59	2.50
GLG-skew ($\phi_r = 0.5$)	0.45	0.36	0.40	0.33	0.47	4.67	3.52	1.91	1.84	4.40
GLG-skew ($\phi_r = 1.0$)	0.31	0.26	0.29	0.24	0.34	4.91	3.70	1.95	1.91	5.00
SKT-skew ($\phi_r = 0.1$)	0.95	0.66	0.64	0.57	0.72	2.77	2.40	1.37	1.53	2.42
SKT-skew ($\phi_r = 0.5$)	0.40	0.34	0.39	0.32	0.44	4.57	3.41	1.86	1.79	4.17
SKT-skew ($\phi_r = 1.0$)	0.28	0.24	0.29	0.23	0.33	4.87	3.64	1.92	1.89	4.88

Lastly, we compare the marginal posterior predictive distributions with the out-of sample marginal distributions of the simulated data. Figure 3 compares the marginal posterior predictive distribution for a point chosen at random from the first four regions of our spatial design, of the GLG-skew model, and the out-of-sample marginal distribution of the associated SKT simulation. Here the anticipated discrepancies emerge, as the GLG-skew model with its symmetry constraint is incapable of reproducing the skewness of the SKT simulation, particularly for the first region, and gradually vanishing through the subsequent regions.

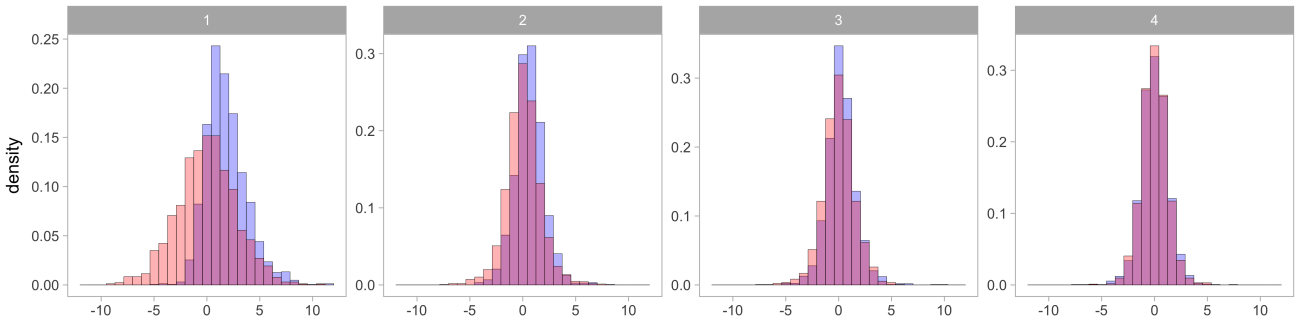

Figure 3: The marginal posterior predictive distribution for the (red) GLG-skew model for a point chosen at random for regions 1 through 4, and the associated out-of-sample marginal distribution (blue) of the SKT simulation.

3 Validation aspects

The use of statistical models as stochastic approximations for expensive, deterministic models (or parts thereof) has a long history in statistics and environmental science. However, despite the sheer number of applications ranging from emulation, data assimilation, stochastic weather

generation and more recently data compression, a fundamental point raised by A is: how do we decide to which extent the approximation is good for our purpose? In other words, would statistical surrogates from our model pass a 'Turing test' and be indistinguishable from the climate/weather model output? While this question is at the center of the development of validation methods for stochastic weather generators, it has been mostly addressed in the context of time series. The push for new quantitative methods to compress data, led by the National Center for Atmospheric Research (NCAR), as well as the wide availability of spatio-temporal models, is prompting a renewed discussion on which validation methods are appropriate as Turing tests in space and space/time.

On this topic, we would like to offer an image/video processing perspective. As argued by Castruccio et al. (2019), determination of similarity between maps of physical quantities from the weather model and its approximation can be regarded as a visual problem. While image processing has developed quantitative metrics of similarity, empirical evidence suggests that the perception of similarity is strongly determined by the level of expertise of the subject performing the test. In other words, a subject expert is expected to be considerably more skilled in determining differences between the original output and the statistical model than a casual subject.

4 Geoscience, engineering and wind energy aspects

While this work indeed does represent a cross-disciplinary effort intended to go beyond the statistical community, as noticed by A, several important points were raised regarding extension of the present methodology to provide a practical pathway towards the implementation of this plan, from both a geoscience and engineering perspective. An important point raised by PRE and also by Z relates to the role of a stochastic approximation for a single run as a representative reproduction of the uncertainty. We are now in a position to have much more robust evidence from an ensemble and a more articulated use of wind extrapolation at hub height and different

turbines (see the points below), but this is an important question to address in the case of a large cross-disciplinary study such as this where multiple simulations controlling for all the desired sources of uncertainty can only be provided in the span of years, if ever.

Is then one run sufficient to capture some of the most important sources of uncertainty? In this particular case, we believe that it is case, as we did not see evidence of inter-annual trends in the wind patterns from the (short) WRF simulations, but also for multi-decadal reanalysis (MERRA-2, Gelaro et al. (2017)) and present and future simulations (MENA CORDEX, Jones et al. (2011)). Therefore, given the regional nature of the problem, one may regard the year-to-year variability as a surrogate for the internal variability of the model, hence lending support for the use of *stochastic generators* originally used for multi-decadal global climate model ensembles (Jeong et al. 2018, 2019, Tagle et al. 2019, Fuglstad & Castruccio 2020).

Further points strictly related to wind energy production have been raised by PRE, and given the importance of these remarks, we address them on a point-by-point basis:

(i) In collaboration with the Atmospheric Chemistry and Aerosol Modeling Group at the University of Notre Dame, we have now generated a new WRF ensemble comprising of four runs between 2013-2016, with two different planetary boundary layer (PBL) schemes and two different surface layer schemes. The simulation is nested and comprises of resolutions from 9km to 3km in Saudi Arabia, and the output is hourly data for most of the domain and half-hourly for the western part of the country. The simulations have 40 vertical levels with approximately 20 meters spacing in the lower levels up to 200 meters to have high detail in the region of interest for wind turbines, see Giani et al. (2020) for a full description. This also addresses the point raised by Z and B on the use of a single WRF run and the need of uncertainty quantification: the ensemble represents the first attempt to control for variability of PBL and surface layer schemes, which were identified by our collaborators as two of the most important sources of uncertainty to quantify. The sheer amount of

computational power required for just four runs (240,000 core hours), along with the cost of storage of the considerable amount of data generated (111 Tb) has prevented us further sensitivity analysis against other sources of interest, most noticeably boundary conditions. An alternative, useful approach would be to consider off-the-shelves forecast data such as ECMWF-HRES, as pointed out by B, which could be used for additional validation of the results.

(ii) We agree that the functional form of extrapolation, i.e. the power law, has been subject to increased scrutiny and alternative modeling strategies have been currently sought to decrease the extrapolation error. One of the most recent work on the topic (Vassallo et al. 2020) successfully applied a feedforward neural network, providing up to 52% improvement in extrapolation over the power law. The use of such approach is however here problematic as 1) no observational data at high altitude were available at the time of this work and 2) even assuming such data were available, more than 80,000 neural networks should be trained for each location, unless another a full domain convolutional neural network approach would be deployed.

(iii) In our siting work (Giani et al. 2020) we constrained the search of sites to locations without a complex terrain precisely to avoid these effects. One of our group member is also currently developing simulations with WRF runs acting as boundary conditions for Large Eddy Simulations (LES) models in order to bring the simulation resolutions below 1km, resolve at least some of the turbulence and hence be able to capture the effects of complex terrain and high altitude.

(iv) One of our group members is currently finalizing a work specifically targeted at finding the best generalization of the power law for hourly data for the ensemble mentioned in part (i). Our new results indicate that a model with a heteroscedastic (due to atmospheric stability), location-specific, hourly-varying power law is indeed considerably more appropriate. Such

a study was not possible in the work object of this discussion, as the WRF run did not have sufficiently many pressure levels near the surface to test the power law.

(v) We again refer to our siting work (Giani et al. 2020), where we have used a database comprising of a wide array of wind turbines with different power curves. The same work also addresses the issue of the use of different optimal turbines for different sites.

(vi) In the siting work (Giani et al. 2020) we have considered error propagation generated by wind, as well as uncertainty on the building and maintenance costs, as well as connection to the electric grid when providing the final levelized cost of energy.

(vii) In yet another work in progress we have focused on some areas of interest in Saudi Arabia and performed simulations at 1 km with different wind farm configurations. Preliminary results indicate configurations between 5 to 7 rotor diameters imply a loss in the range of 10% to 20% due to the wake effect.

An additional point raised by Z was the change of wind patterns over relatively short time spans observed over many regions of the world. A WRF simulation over decades would provide strong evidence, but it is presently impossible due to computational and most importantly storage constraints. We have however preliminary evidence from an analysis of the Middle East North Africa - Coordinated Regional Climate Downscaling Experiment (MENA-CORDEX, Jones et al. (2011)) that there is no significant trend in the vast majority of the country across a past (1980-2005) and a future (2025-2050) 26-years period. This analysis was performed at a considerably coarser resolution ($0.44^\circ \times 0.44^\circ$), but still lends some support on the claim that the siting provided in this work or in Giani et al. (2020) could be still more or less unchanged in the coming decades.

Finally, while the issue of extreme wind was not raised by the discussants, we nevertheless like to point out that there is a non-negligible risk of turbine disruption operation due to strong wind. Our group has addressed this topic in a recent work (Chen et al. 2020).

References

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', Springer International Publishing.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017), 'Stan: A probabilistic programming language', *Journal of Statistical Software* **76**(1).

Castruccio, S., Genton, M. G. & Sun, Y. (2019), 'Visualizing spatiotemporal models with virtual reality: from fully immersive environments to applications in stereoscopic view (with discussion)', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **182**(2), 379–387.

Castruccio, S., Ombao, H. & Genton, M. G. (2018), 'A scalable multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data', *Biometrics* **74**(3), 823–833.

Chen, W., Castruccio, S. & Genton, M. (2020), 'Assessing the risk of disruption of wind turbine operations in Saudi Arabia using Bayesian spatial extremes', *Extremes* . in press.

Fuglstad, G.-A. & Castruccio, S. (2020), 'Compression of climate simulations with a nonstationary global spatiotemporal spde model', *Annals of Applied Statistics* **14**(2), 542–559.

Gelaro, R., McCarty, W., Surez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M. & Zhao, B. (2017), 'The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)', *Journal of Climate* **30**(14), 5419–5454.

Giani, P., Tagle, F., Genton, M. G., Castruccio, S. & Crippa, P. (2020), 'Closing the gap between wind energy targets and implementation for emerging countries', *Applied Energy* **269**, 115085.

Gneiting, T., Balabdaoui, F. & Raftery, A. E. (2007), 'Probabilistic forecasts, calibration

and sharpness', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(2), 243–268.

Jeong, J., Castruccio, S., Crippa, P. & Genton, M. G. (2018), 'Reducing storage of global wind ensembles with stochastic generators', *Annals of Applied Statistics* **12**(1), 490–509.

Jeong, J., Yan, Y., Castruccio, S. & Genton, M. G. (2019), 'A stochastic generator of global monthly wind energy with tukey g-and-h autoregressive processes', *Statistica Sinica* **29**, 1105–1126.

Jones, C., Giorgi, F. & Asrar, G. (2011), 'The coordinated regional downscaling experiment: CORDEX - an international downscaling link to CMIP5', *CLIVAR exchanges* **16**(2), 34–40.

Palacios, M. B. & Steel, M. F. J. (2006), 'Non-Gaussian Bayesian geostatistical modeling', *Journal of the American Statistical Association* **101**(474), 604–618.

Tagle, F., Castruccio, S., Crippa, P. & Genton, M. G. (2019), 'A non-gaussian spatio-temporal model for daily wind speeds based on a multi-variate skew-t distribution', *Journal of Time Series Analysis* **40**(3), 312–326.

Tagle, F., Castruccio, S. & Genton, M. G. (2020), 'A hierarchical bi-resolution spatial skew-t model', *Spatial Statistics* **35**, 100398.

Vassallo, D., Krishnamurthy, R. & Fernando, H. J. S. (2020), 'Decreasing wind speed extrapolation error via domain-specific feature extraction and selection', *Wind Energy Science* **5**(3), 959–975.

Zhang, H. & El-Shaarawi, A. (2010), 'On spatial skew-gaussian processes and applications', *Environmetrics* **21**(1), 33–47.