Responses to Reviewer 1

The paper describes an improved methodology for the probabilistic forecast of hourly wind speeds on a forecast horizon of one day. Focus of the work is the prediction of power ramps, i.e, the strong increase or decrease of wind power over a time window of one or more hours. Basis of such predictions is typically an ensemble of physics-based numerical wheather predictions (NWP), which is the transformed into wind power predictions using idealized wind turbine power curves.

The authors address the temporal multi-point correlation structure of wind speeds and their forecasts as one crucial problem to improve probabilistic forecasts. Together with existing methods they introduce their own approach to explicitly model the joint multivariate distributions of hourly wind speeds with respect to their mutual temporal dependencies. This new approach of Multivariate Gaussian Regression (MGR) has previously been published in a journal on statistics and econometry, and is consequently described only shortly.

The main part of the paper performs a detailed comparison of the various methods at the example of one power ramp event measured at the German FINO 1 platform in 2019. The proposed method of MGR outperforms the other appraoches in the comparison.

The paper addresses a highly relevant problem in wind power forecasting, namely the so-called power ramps. Moreover, with the temporal multi-point correlation structure of wind speeds the authors address one of the central and most demanding challenges of the field and of atmospheric flows in general. Their approach is promising and the given example is convincing.

Technically, the paper is well written and well readable. The structure is clear and comprehensive. English language style is fluent, precise, and, as far as I can say, correct. Results are presented clearly and with very appropriate graphics. References are given whereever necessary.

The reviewer is an expert neither in NWP nor in the advanced mathematical approaches of the paper. However, in my eyes this paper makes an important contribution, and it is almost ready for publication.

Thank you very much for taking the time to read and review our manuscript and for the positive feedback! We are happy to hear you think our paper makes an important contribution. Please find a response to your comments below.

General remarks

The demonstration of the proposed method using just one single example of a power ramp is quite limited. However, given the length of the paper of already 20 pages, more examples do not seem to make sense. Could the authors comment on the performance of the method for more examples, or, elaborate on possibilities of a wider evaluation?

It is true that we have only visualized wind speed and power ramp forecasts for a single day. On the other hand, the scores we compute and use to quantify forecast performance – of both multivariate wind speed forecasts (via the DSS) and power ramp forecasts (via the area under ROC curve and Brier score) – are based on all forecast days within the dataset and not just the single day visualized.

This has been clarified in lines 67-70 of the revision: *In Sect. 4, the postprocessing methods are first illustrated using an example case where a power ramp occured. Subsequently, the out-of-sample performance of multivariate wind speed forecasts is evaluted across all cases (i.e., days) using scoring rules (Gneiting et al., 2007) and the predictive skill of the ramp probabilities derived from these forecasts quantified as well.*

Specific remarks

P. 3 L. 85: The bi-linear interpolation between grid points is assumably widely used and probably also accepted. However, it is knwon to reduce fluctuation amplitudes. It would be helpful to have any estimate to what extent that effect is present for the given case.

This is an interesting point! We use a bilinear interpolation because the neighboring grid points of FINO stations are all quite homogeneous (i.e., offshore, over the ocean). Our manuscript is also more focused on investigating

the multivariate postprocessing methods which can be used to predict joint distributions of wind speeds from NWP ensemble forecasts at specific locations, rather than the details of how these ensembles forecasts are obtained in the first place. Still, this is definitely a topic which should be systematically investigated and may play a significant role in certain situations (e.g., perhaps in more complex terrain).

We have touched upon this topic in lines 347-349 of the Discussion: *Although this work has focused on the different methods which can be used to postprocess NWPs, the preprocessing methods initially used to interpolate these NWPs to a specific location also play a role. This could be an interesting topic for future work.*

Technical remarks

P. 3 L. 77: The phrase "but observations generally far from zero" does not seem to make sense. Please double-check.

We have rewritten this sentence for improved clarity. It can be found in lines 79-81 of the revision: *The distributions of wind speeds observed at the three sites are skewed, but since the NWP ensemble (Sect. 2.2) performs well for next-day lead times and wind speeds are generally high, prediction errors can be approximated by Gaussian distributions.*

P. 17 L. 355: Inserting a "that" after "ensure" would be helpful, although (to my understanding) not strictly necessary.

We have added a "that" for improved readability.