

Responses to Reviewer 2

The manuscript, "Predicting power ramps from joint distributions of future wind speeds" describes a novel statistical postprocessing method called the multivariate Gaussian regression (MGR) to provide calibrated and sharp hourly ECMWF 100-m wind speed forecasts for lead times +24h to +47h. Observations from three offshore meteorological towers are used for verification. The main advantages to the new method are that it can explicitly model temporal dependencies across multiple lead times and is not limited to the number of members available in the raw ensemble, unlike the various other multivariate postprocessing methods discussed in the paper and compared against in the figures. Skillful modeling of these joint distributions across lead times is essential for reliable power ramp predictions and consequently the balancing of a power grid with wind energy resources. The authors found that the new method outperformed the various other methods commonly used in the literature when assessing the scores and skill scores of wind speed and wind power forecasts. The paper is overall well-written, interesting, and valuable to the wind energy field. Here are my recommendations for further clarity.

Thank you very much for taking the time to read our manuscript, for your well written summary, and the positive feedback! You can find our responses to the comments below.

Scientific and clarification comments

L48: I'm not sure what was meant by the phrase, "and not supported by the....." Can you please reword for clarification?

We have restructured the sentence to clarify our point. It can be found in lines 46 to 48 of the revision: *This assumption – namely, that dependencies between the postprocessed forecasts are the same as between raw NWP forecasts – is not always fulfilled (Ben Bouallègue et al., 2016), for the same reasons that ensemble margins are often miscalibrated and must be postprocessed.*

L59 and throughout: the use of higher resolution may be confusing to readers in this context as higher resolution when referring to NWP forecasts typically means higher temporal or spatial resolution for numerical integration. Here, it seems to mean more ensemble members. Can you replace with another term or distinguish for the reader the meaning in this paper's context?

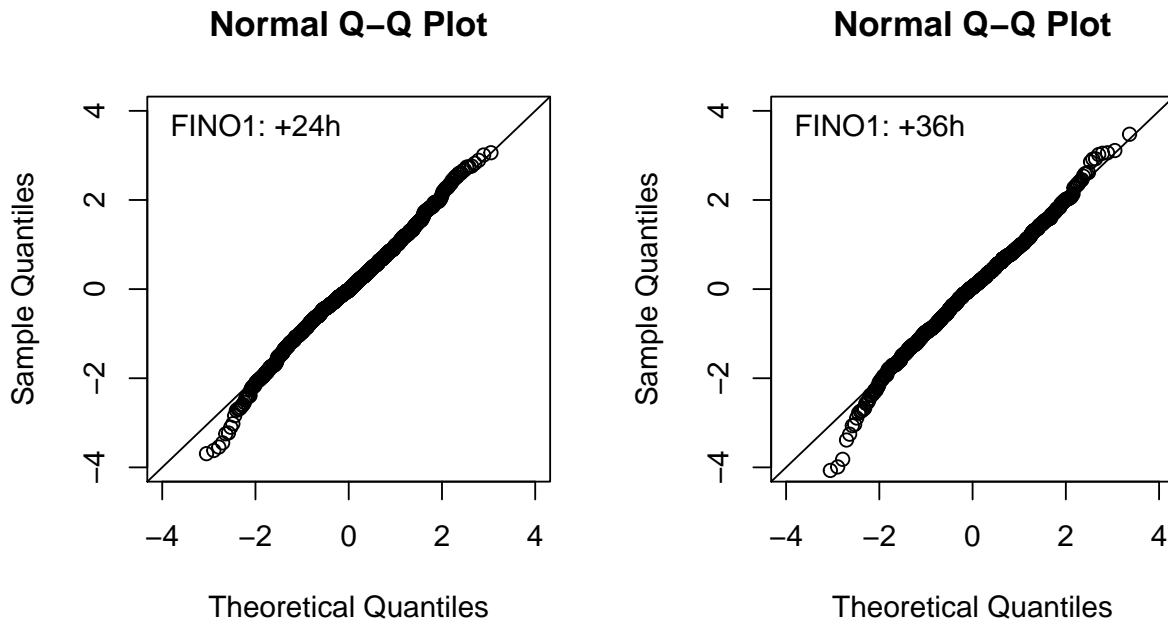
You are right, the multiple meanings of "resolution" could be confusing to the reader. We have added an extra sentence at this point in the text which clarifies the distinction between the resolution of ramp probability forecasts and the resolution of NWP models (in time and/or space). It can be found at lines 60-61 of the revision: *This ramp probability resolution is only a function of the number of ensemble members (i.e., its inverse) and should not be confused with the temporal or spatial resolution of the NWP model.*

L83-87: Please mention somewhere in here that the forecasts are initialized at 00Z. It's mentioned later, but I found myself asking that question in this section.

We have added the forecast initialization time as you suggested. It can be found at lines 86-87 of the revision: *The 50 perturbed ECMWF ensemble members have a spatial resolution of 18 km, with 91 model levels and a temporal resolution of one hour and are always initialized at 00 UTC.*

Sect 3: Wind speeds can only take on positive values. Why not use the truncated Gaussian distribution truncated at zero rather than the untruncated Gaussian distribution?

In our application we consider wind speeds at 100m AGL over homogeneous (ocean) terrain, where the distributional means are typically large relative to the forecast uncertainty. Subsequently, the large majority of the distribution's mass is positive and taking into account truncation at zero is not so important. We have included Q-Q plots for +24h and +36h forecasts at FINO1 postprocessed with NGR below.



Furthermore, our aim is to forecast turbine power – which is not produced below the cut-in wind speed of 3 m/s – so the exact details of the wind speed distribution around zero are not that important. In the end, a slightly negative wind speed forecast can be treated the same as one which is slightly positive since they both produce no power. This point was already made in Sec. 2.1, but we have now added a quick explanation to the beginning of Sec. 3. It can be found at lines 113-114 of the revision: *At the FINO stations, hourly wind speed observations obs_i are generally large compared to the corresponding spread of the ECMWF ensemble forecast, so that truncation at zero can be neglected.*

Equations and equation explanations in the text: *Some of the equations and explanations of variables in the text reuse symbols which I found difficult to keep track of while reading. For example, i and j referred to lead time and next lead time, respectively in some earlier places in the paper while i referred to ensemble member in other parts (e.g., equation 5) and a new variable, m , became the new lead time variable. The definition of N also changed as well. For better readability, please make the variable symbols unique and consistent throughout the paper.*

We agree that the notation was confusing and have modified the variable names to ensure consistency.

L253: *Can you please provide more details about what the DSS score tells the reader about the forecast? Does it inform about the performance of the calibration and sharpness, just one, or some other aspect of the forecast?*

The DSS is the negative log-likelihood of the observations given the forecast distribution. It is a proper score and thus takes into account both forecast calibration and sharpness. We have added this information in line 260 of the revision: *The DSS is a proper score which measures both sharpness and calibration simultaneously.*

L295: *Can you please state why the COP(Err) was selected as the reference method for wind speed as opposed to the other methods?*

We have chosen COP(Err) as the reference method for joint distributional forecasts because it is the existing method which performs best at our stations. According to the DSS, it is only outperformed by (some of the) novel MGR models.

Figure 1: *How were the 50 members out of 1000 selected? Would the gray lines show comparable spread or much more spread if the full 1000 members were shown?*

The 50 members shown in the figure are the first 50 vectors (out of 1000 total) simulated from the postprocessed multivariate Gaussian distribution. An ensemble with 1000 members would be expected to have the same variance (or spread) as one with 50 members since both are simulated from the same distribution. On the other hand, one

would naturally expect the difference between the weakest and strongest wind speeds (i.e., the range of the ensemble forecast) to increase with the number of simulated members.

Figure 2: The caption says that the data were aggregated by month and year. Is it just showing December data like Figure 1 or all months? If all months, do you expect there to be differences in the scores, and thus optimal method, if all months were not aggregated?

The boxplots visualize average monthly forecast skill across all months in the dataset (e.g., $12 \cdot 5 = 60$ data points per boxplot for 5 years of forecasts). These values are obtained as follows: First, the DSS for each joint distribution forecast is computed and the difference taken to the DSS of the reference method $COP(Err)$. Subsequently, these differences are averaged for each month of each year (e.g, distinct values for December 2019, January 2020, etc.) and visualized. Alternatively, one could avoid aggregating the scores (i.e., averaging over each month) and instead plot differences in DSS for all forecast days. This does not significantly change the results except that the boxplots span a larger range of values (there is more variance in daily DSS differences than in monthly-averaged DSS differences) and the interpretation of the quartiles changes.

L323: Can you please state why the ECC method was selected as the reference method for power ramps as opposed to the other methods?

ECC was chosen as the reference method for the power ramp forecasts because it is the standard multivariate NWP postprocessing method. This choice also highlights the improvements in skill that can result from increasing the resolution of ramp predictions by simulating larger ensembles (something which is not possible with ECC.)

Technical comments

L148: Spelling of seasonally.

We've fixed the spelling mistake.

L160: Spelling of postprocessed individual.

We've fixed the spelling mistake.

Eq(5): Please define symbol.

I'm not sure which symbol in Eq(5) you are referring to. If you mean the Θ in Eq(6), we have rewritten the equation in such a way that it becomes unnecessary. The variable names in Eq(5) have been changed to ensure consistency following your previous comment.

L208: I think you can remove on before become.

Thanks, we have removed the extra *on*.

Figures 1 & 4: Please put the corresponding forecast hour along the horizontal axis or state the forecast hour range of the figure within the caption.

We have added the forecast range to the captions of both figures.