Response to the reviewers

Dear Reviewer:

Thank you for your comments. These are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and responded to them which are described in detail below.

- **Comment No.1:** The authors mention an impressive dataset of 541 meteorological towers in a specific region. Some more details would be relevant in order to understand if the data are comparable – i.e., how does the terrain differ among the various met mast locations, are the measurement heights the same, are the instruments the same (cup anemometers, sonic anemometers, lidars)?

Response 1: Thank you for your comment. The areas of meteorological towers used in this paper are mainly in North China. Most of the meteorological towers are located in the terrain of plains and hills, which are judged as L, M, and H classes in accordance with the terrain in the IEC61400-1:2019 Chapter 11.2 with the approximate proportions of 50%, 30%, and 20%.

The lowest height of these meteorological towers installed with wind speed and direction sensors is 10 m, the highest height is between 70 m and 140 m, and using the wind shear exponent to uniformly extrapolate to a height of 100 m.

The anemometers used in the meteorological towers are cup anemometers, no ultrasonic or LiDAR.

The principle of using the above database is to cover as wide a range of wind parameters as possible, making the model more widely applicable. We add the above description in the manuscript.

- **Comment No.2:** The IEC 61400, ed. 4 standard allows several different approaches to extrapolation, including avoiding the extrapolation altogether by introducing a higher safety factor. It will be useful if the authors could study/compare these different extrapolation approaches in the context of their proposed methodology.

  Response 2: Thank you for your comment. This is a very good suggestion, and we believe that the reason why the IEC standard allows for the existence of different extrapolation methods with a high safety factor is that each method has a rationale and the truth value cannot be verified to a certain extent, and in this case a high safety factor can only be used to ensure safety.
  We have identified the technical route of "fitting before aggregation" through previous research, and in the test case we have only compared the results under this route and called it the IEC method. The aggregation before fitting and the inverse first-order reliability method (IFORM) are not compared in the current manuscript. In fact, this work we are in progress because we also realize the differences in the results of the different methods and the importance of uncertainty analysis in the extrapolation of ultimate loads.

- **Comment No.3:** One significant challenge in the "fitting before aggregation" method is that the distribution fitting on a few values is not very robust, and a few outliers or bad fits can distort the aggregated result. It would be good to check the confidence in the aggregated distribution predictions – for example by doing multiple local distribution fits by bootstrapping the block maxima.

  Response 3: Thank you for your comment. The problem you mentioned is indeed the problem we face in this method. In the process of local distribution fitting, we introduced the Normal, Log-normal, Gumbel and Weibull distributions as candidate distributions, and through the fitting test, we found that none of these distributions can fully satisfy all the samples. As described in Section 4.2, in order to ensure the consistency of the MLP model output, we chose the Weibull distribution, which fits the

most samples better. This also leads to a deviation of the local distribution from reality for some wind speeds, which perturbs the load extrapolated results. In fact, we have seen in the literature that the use of Gaussian mixture model gives better results than distributions such as Weibull to avoid and reduce outliers or bad fits, now we are trying in this way. And we will also try by doing multiple local distribution fits and by bootstrapping the block maxima.

- **Comment No.4:** There is a dependency between the shape and scale parameters in a Weibull distribution fit (if you choose a value of one parameter, it will define what is the value of the other parameter that best represents the data set). Therefore, fitting separate meta models for the scale and shape parameters of the Weibull distribution may limit the accuracy of the results. In the current manuscript, it doesn't get clear if the authors fit one single MLP model with two outputs, or two separate models? Please discuss.

Response 4: Thank you for your comment. Actually we have used one single MLP model with three outputs. But due to the Weibull distribution parameter loc=0 we have used, it can also be considered as two outputs. We will describe it in detail in the manuscript.

Once again, thank you very much for the constructive comments and suggestions which would help us in depth to improve the quality of the manuscript. We will try our best to improve the manuscript. Please feel free to contacts with any questions.

Kind regards,

Pengfei Zhang