EGUsphere, author comment AC1
https://doi.org/10.5194/egusphere-2022-599-AC1, 2022

Reply on RC1

Daan R. Scheepens et al.

Author comment on "Adapting a deep convolutional RNN model with imbalanced regression loss for improved spatio-temporal forecasting of extreme wind speed events in the short to medium range" by Daan R. Scheepens et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-599-AC1, 2022

Dear Reviewer,

Thank you for your review and suggestions for improving the clarity of the manuscript. We proceed to answer your questions to the best of our ability. The changes and clarifications will be added to the final manuscript.

Major comments:

-
  - &
  - We will address the points about model application and motivation better in the introduction. The aim of the paper is to investigate how the spatio-temporal predictions of a deep learning forecasting model may be improved for the extremes through the manipulation of loss function. Indeed, it is likely that any improvements for the extremes go hand-in-hand with predictions deteriorating for non-extremes. We will attempt to show whatever trade-offs there are in this regard more clearly in the results. From an application perspective, such a trade-off would nevertheless be very attractive to improve a model's efficacy as a warning-system: Its ability to distinguish different levels of extremes would be critical while its ability to distinguish different levels of non-extremes would be irrelevant.

    Furthermore, the focus on relative rarity in our definition of an extreme event will be better motivated in the methodology. What we mentioned perhaps too briefly in our manuscript is that the primary reason for making this choice stems from the fact that extreme winds in the absolute sense (e.g. exceeding 25 m/s) occur exclusively off-shore in a very localised region and are thus absent from the majority of coordinates in the ERA5 reanalysis data (where typically only max. wind speeds of 8-10 m/s are present). Defining extreme events rather in terms of their relative rarity at each coordinate allows us to investigate forecasting improvements of extreme events more generally by looking at improvements on the tails of the respective distributions, regardless of what absolute values these tails actually obtain. Demonstrating that forecasting performance on the tails can be improved in this more general context, by adapting the loss function, can be swiftly translated to other cases where the tails of the distributions denote actual hazardous events. This paper serves to help indicate as to how the loss function may be adapted as to best

help improve forecasting performance on the tails. indeed, with the assumption that the tails typically denote extreme events of some form. We will modify the paragraph in question in the methodology accordingly.
  - It is true that probabilistic output is typically preferred by users but here we would argue that in the same way as how deterministic NWP forecasts are commonly aggregated into probabilities by utilising large ensembles, the deterministic forecasts of the ConvLSTM regression model can be aggregated into probabilities e.g. with different ensemble members trained on different subsets of the training set. We will mention this possibility in the discussion.
- We certainly understand the reviewer's concern. However, instead of providing benchmark comparisons with other models, this point has made it apparent that we have had to more clearly state the aim of the paper in the introduction. The aim of the paper is not to investigate our adaptation of the ConvLSTM as compared with other state-of-the-art models. Rather, the aim of the paper is to investigate how the performance of a popular spatio-temporal deep learning model (the ConvLSTM) changes for various thresholds of extreme events by utilising two types of loss functions proposed in the literature on imbalanced regression. For literature on the improvements of the ConvLSTM over other state-of-the-art models, or simpler non-convolutional, or non-recurrent models, the reader could be referred to Shi et al. (2015) or Shi et al. (2017), which we will mention in the introduction.
- We will clarify this point in the methodology. 1000 hPa wind speed is the only variable used. The model takes in 12 consecutive hours of wind speed data over the 64x64 grid, comprising a tensor of size 12x64x64. This tensor is encoded through the encoding network into a hidden state and decoded through the decoding network into an output tensor of size 12x64x64 comprising the forecast of the subsequent 12 hours. The temporal correlations between consecutive hours are taken into account implicitly by the convolutional LSTM layers. For the exact details on how the convolutional LSTM modules achieve this we will refer the reader to Shi et al. (2015).

Minor comments:

- Noted.
- Thank you for the comment. Indeed, this is missing in the literature review as the focus on probabilistic was added after the introduction seemed to be finished. We have updated the introduction accordingly and added the following paragraph:

  "Furthermore, a lot of work has been done in recent years on probabilistic weather forecasting and many postprocessing methods have been proposed to improve probabilistic forecasts. Postprocessing is typically applied to ensemble weather- or energy forecasts and attempts to correct biases exhibited by the system and improve overall performance (see e.g. Phipps et al. (2022)) but has been explored to a lesser degree in the context of extreme event prediction. One approach to postprocess ensemble forecasts for extreme events is to utilise extreme-value theory, a review of which can be found in Friederichs et al. (2018). The authors propose separately postprocessing toward the tail distribution and formulate a postprocessing approach for the spatial prediction of wind gusts. Other authors have explored the potential of ML in this context. Ji et al. (2022), for example, investigate two DL-based postprocessing approaches for ensemble precipitation forecasts and compare these against the censored and shifted gamma distribution-based ensemble model output statistics (CSG EMOS) method. The authors report significant improvements of the DL-based approaches over the CSG EMOS and the raw ensemble, particularly for extreme precipitation events. Ashkboos et al. (2022) introduce a 10-ensemble dataset of several atmospheric variables for ML- based postprocessing purposes and compare a set of baselines in their ability to correct forecasts, including extreme events. Alessandrini et al. (2019), on the other hand, demonstrate improved predictions on the right tail of the forecast distribution of analog ensemble (AnEn) wind speed forecasts using a novel bias-

correction method based on linear regression analysis, while Williams et al. (2014) show that flexible bias-correction schemes can be incorporated into standard postprocessing methods, yielding considerable improvements in skill when forecasting extreme events."

- The review of data-driven weather forecasting models will be removed from the introduction as we have come to see that this is, indeed, not the focus of the manuscript.
- For this study initially 5 different pressure levels, including the diagnostic 10 m wind fields, were investigated. With the currently implemented hub heights of wind turbines in Austria of 100 to 135 m a.g.l., the 1000 hPa fields are more appropriate (corresponding to ca. 100-130 m in Eastern Austria (main wind energy region)). Furthermore, reanalysis methods typically interpolate across- and output data at different pressure levels rather than height levels, which also motivates the choice. We have modified the methodology accordingly.
- Thank you for pointing this out. We have decided to repeat the experiments utilising a Yeo-Johnson power transform (Yeo and Johnson, 2000) before the zero-mean, unit-variance normalisation in order to make the distributions more Gaussian-like.
- Noted. We will include another, linear, weighting method in the comparison and will include the SERA loss with three different sets of control-points in order to provide a more complete comparison.
- Thank you - this was indeed overlooked. The final results will be calibrated as recommended in Ferro and Stephenson (2011).
- In principle it would be possible to incorporate the SEDI loss or another loss function into the model. This, however, was out of scope for this work. Machine learning based methods, as well as also statistical methods, tend to smoothen the forecasts and underestimate especially the tails. The idea was to implement a loss function which is able to account for that, sharpen the forecasts and is able to get the intensities in the right order. This is essential for not only wind energy applications (planning of feed-in, curtailment, etc.) but also for e.g. tourism (winter sports), transportation, and forestry. We added the following to the discussion:

  "Another possible extension of this work would be implementing either the SEDI or the FSS as a loss function (see e.g. Lagerquist and Ebert-Uphoff, 2022) or even combine the ConvLSTM with a so-called physics-aware loss function (see e.g. Schweri et al., 2021; Cuomo et al., 2022)."
- Rather than averaged, the results are aggregated over all lead-times. This will be clarified.
- Noted.
- Full fields. We will make sure to clarify this.
- Because Fig. A1 is a comparison of the continuous-valued wind speed fields and as such requires a continuous score like the RMSE to compare. The SEDI is a categorical score that can only be used with discrete data considering events and non-events i.e. after applying a threshold to the continuous wind speed fields. We will make sure to clarify this.