

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2021-430-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-430

Anonymous Referee #1

Referee comment on "Temperature forecasting by deep-learning methods" by Bing Gong et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-430-RC1>, 2022

The present paper presents the use of a stochastic adversarial video prediction model to forecasting the two-meter temperature. While the paper is interesting to read and the conclusions seem valid, I do have several points that I think would have to be addressed before the paper could be considered for publication in GMD. In particular:

1) The present paper is a contribution to an increasing line of research on the application of a deep-learning based methodology to weather prediction, in particular the use of an existing video prediction model applied to weather prediction. This line of work has been of great interest when first showcased through the contributions of Weyn et al. and Dueben et al., just to name a few, but it does not feel that the present paper adds anything substantially new besides using a different architecture for the same problem. One main issue is that meteorological data is fundamentally different from generic video data in that it follows a well-defined system of partial differential equations. In this purely data-driven approach it seems that one has to be willing to throw away more than a hundred years of research on the understanding of these governing equations of hydro-thermodynamics just to be able to use an off-the-shelf video prediction architecture, which does not seem to come close to where traditional numerical methods can go today in terms of relevant forecast metrics. The question to ask is hence whether is indeed the right approach going forward, or whether one should strive to combine data-driven approaches with the inductive bias as provided by the fundamental laws of physics. There is a growing interest in physics-informed machine learning, which allows combining differential equations with data-driven machine learning which in a way seems more appropriate for the present problem at hand. If that was possible for the present model then I think the paper would become much stronger and more suitable for what would actually be required for weather prediction.

2) The selection of features (cloud cover, 850 hPa temperature and two-meter temperature) seems slightly arbitrary. While the authors do provide some justification for the selection of these parameters, there are many more parameters that influence the evolution of the two-meter temperature. The authors then state "A more systematic variable selection process as is typical for data science studies is beyond the scope of this paper.", but I do not believe this is a justifiable statement here, because the authors do carry out a data science study in this paper. Again, if this was the first paper to be written

on using a video prediction model for weather prediction this point could be easily forgiven, but as there are many other papers out there that provide proof of concept that such models can predict the future weather to some degree I think some more work needs to be done here to justify this feature selection, and to show which features have to be selected to get the best possible model results. In the machine learning literature it is customary to carry out ablation studies that showcase the importance of various aspects of the data/model components being used, and I think such a study would be beneficial here as well.

3) The baseline comparison model used is a standard convolutional LSTM model. This is the simplest possible model for video frame prediction and it is well-known to perform rather poorly as it exhibits an excessive amount of diffusion. Thus, beating this baseline is rather straightforward so I wonder if the comparison of the authors' model to this simple model really yields a lot of information about the absolute strength of this model. It would be great to add a somewhat more state-of-the-art comparison model as well to be truly able to assess how good the stochastic adversarial video prediction model is for the present problem. Related to this, it would also be useful to add the performance metrics of traditional weather forecasting models as point of comparison. Right now this information is just provided in the Discussions section but it would be nice to show these metrics in the plots as well.

4) Owing to the interest in data-driven weather forecasting, a standard benchmark "WeatherBench" has been proposed to facilitate comparison with other deep learning based models. The present paper does not use this benchmark but rather investigates the model performance over Europe instead. This makes positioning this work within the wider literature rather challenging so I wonder if it would not be better to provide these results instead (or in addition) for the WeatherBench dataset as well. Again, this would facilitate comparison with other approaches that have been proposed for data-driven weather forecasting.

In summary, while the present paper is interesting to read I do believe there isn't a sufficient amount of novelty yet that warrants publication in GMD in its present form. The main contribution of picking a video prediction model and applying it to weather forecasting has been done several times in the recent literature so this does not feel novel enough anymore unless other open aspects of data-driven weather prediction are investigated in addition. These could be, as indicated above, a combination with differential-equations based models, a more thorough investigation of which parameters are responsible for the success of the proposed model, beating other existing approaches for the exact same problem domain, just to name a few.