

Geosci. Model Dev. Discuss., referee comment RC2
<https://doi.org/10.5194/gmd-2021-276-RC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-276

Anonymous Referee #2

Referee comment on "EuLerian Identification of ascending AirStreams (ELIAS 2.0) in numerical weather prediction and climate models – Part 1: Development of deep learning model" by Julian F. Quinting and Christian M. Grams, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-276-RC2>, 2021

This study offers a well thought out alternative to a previous simple statistical model (logistic regression), and represents a significant improvement over the previous implementation. The Convolutional Neural Network (CNN) allows the consideration of non-local spatially dependent predictor/predictand variables, and is readily more interpretable than the logistic regression model. Generally, this is a strong paper, and subject to the minor comments below, the manuscript should be published. My main concern is with the with the generation of figure 4, which I detail in the minor comments below. Additionally, please add a section on limitations of the method, caveats, and future improvements that could be applied to this work.

Minor Points

L55-L64 This is a good justification for a computer vision based machine learning approach.

L66-67 this feels like a bit of a misrepresentation, I would change to say "CNNs identifying salient features in the input space which influence the desired prediction."

L72. I would say that it is "originally designed" as a semantic-segmentation model, as it's applications are now much further reaching.

L106 How much model degradation occurs without this 5th predictor? Figure 2 seems to indicate that the seasonality is not a big factor, as much as including time-lagged ascent

information. Figure 7. Confirms it is not a factor. This seems like something that needs to be explored or commented on further. Is this due to the normalization around the date of interest, and the selection of data around the forecast date. I think it is worth testing whether this variable affects the final skill of the model when you are not selecting data in a 30-day randomized window. Or de-emphasize this line in the introduction in general, as you immediately remove this variable as a predictor.

L157. The non-linearity is not necessarily required. Has it been tested to use linear activations? This would give you an idea of the linearity of the actually predictor/predictand relationship. You have two competing predictor improvements in this model (compared to local logistic regression) 1) the addition of a spatial component via convolution 2) the nonlinear predictor/predictand relationship. It would be good to test what is a bigger factor for model improvement, my inkling is the spatial information is more valuable.

L165. The debate over the efficacy of dropout is distracting to this paper. I would take it out.

L208. Please specify what dataset the MCC threshold tuning (0.05-0.95) tuning was done on.

L245. Readers would benefit from a quick summary of Quinting and Gram's (2021) logistic regression model.

L255 The authors do not define why $\pm 10\%$ is considered perfectly reliable (nor do they test via any subsampling), either justify this more clearly, or I would suggest adopting the Bröcker and Smith reliability diagram framework (Bröcker, J., & Smith, L. A. (2007). Increasing the Reliability of Reliability Diagrams, *Weather and Forecasting*, 22(3), 651-661.)

L270-280 Can you justify why this process should be performed on the testing dataset and not the validation dataset? It appears as if this is tuning a hyperparameter, and you are increasing your model bias skill on the testing data. It seems like the thresholds should be determined on the validation data as you don't plan on running the expensive Lagrangian framework model when implementing this CNN in the future. This seems concerning for this figure.

Grammar edits.

L70 missing space "intrusions (Silverman..)"

L370 remove "aims to" --> "UNet CNN that identifies"