

Atmos. Chem. Phys. Discuss., referee comment RC1
<https://doi.org/10.5194/acp-2021-196-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on acp-2021-196

Anonymous Referee #2

Referee comment on "Forecasting and identifying the meteorological and hydrological conditions favoring the occurrence of severe hazes in Beijing and Shanghai using deep learning" by Chien Wang, Atmos. Chem. Phys. Discuss.,
<https://doi.org/10.5194/acp-2021-196-RC1>, 2021

Wang presented the framework of a deep convolutional neural network to forecast haze events in Beijing and Shanghai. He utilized the trained neural network model and cluster analysis to categorize the regional meteorological regimes associated with haze and non-haze events, focusing on the technical aspects, such as the design of the model. The methods on the combination of a convolutional network and cluster analysis in this paper could be of high interest to the community, and it would be useful to better understand what the neural network model learns and why it misses the haze events. However, discussions of new scientific findings and more in-depth interpretations are not sufficient. For instance, generally, there is no significant difference in features between clusters (e.g., the four clusters of TP in Beijing look similar), and the author did not explain the physical meanings of the clusters. Thus, it is hard to see the necessity of cluster analysis without significant differences in clusters and explanations of their physical meanings. Although the methods and ideas are innovative, the manuscript is not well-written, mainly that the structure is a bit challenging for readers without sufficient background to follow. The methods could be helpful to the community once the author clarifies and addresses the important issues and includes more scientific discussions about the findings.

Specific comments:

1. Page 3, line 95-102: This paragraph is confusing. The first part of the paragraph says that U-net can be used when weather patterns associated with the targeted outcome are known or irrelevant to the task, but the second part mentions that when the environmental conditions related to the targeted outcome are yet known, U-net is not applicable. The two paragraphs are conflicting and confusing. Additionally, the sentences are convoluted and difficult to follow. I suggest rewriting the whole paragraph.

2. Page 4, line 130-132: Are the model structure and parameters the same for the two models (Beijing and Shanghai)? If so, shouldn't the models' parameters be optimized separately for the two regions to obtain the best performance for each of them?

3. Page 6, line 187-189: Do you include a testing set (a holdout dataset that has never been used in the training process)? From your results, I think the validation set in this paper is used to tune the hyperparameters of the model and monitor potential overfitting that occurs when training accuracy decreases but validation accuracy increases. An evaluation for a testing set is recommended to assess the performance of the trained model.

4. Page 6, line 197-198: Please include the description of class-weight and batch normalization in the appendix or model description and how they help the imbalance-data issue.

5. Page 7-8, line 246-249: It is worth checking the maps of features for the TP, FP, and FN for April and May versus the TP, FP, and FN for other months to see their differences in the weather and hydrological features.

6. Page 4-9, Section 2 and 3: The subsections of "Kernel size and CNN performance" and "Reducing the number of input features" could be moved to Section 2, as these two subsections are more related to the model architecture and design. Additionally, the author shows the validation results of reducing the number of input features before introducing the methods of reducing the number of features. The two sections together make it difficult for readers to follow. I suggest reorganizing these two sections and the subsections.

7. Page 8, Figure 4: The unit/axis label of top figure?

8. Page 10, line 322: Please specify what VAE stands for.

9. Page 11, Section 4: The feature patterns of the clusters for the TP, FN, and FP are very similar with slight differences in certain features. Generally, I don't see the point of conducting cluster analysis because (1) there is no significant difference between clusters, (2) the author did not explain the physical meanings of each cluster (if based on the slight differences), and (3) the author focused on the differences between TP and FN when explaining the missing haze events, which can be done by simply comparing averaged feature maps of TP and FN without cluster analysis.

10. Page 11, line 349-354: It seems like the feature patterns of the four clusters for the

TP in Beijing cases are very similar (also the clusters for FN and FP). Why did you choose four clusters? Could you justify the purpose of using cluster analysis, given that there is no significant difference between the clusters?

11. Page 12, line 380: If the results of Shanghai are similar to Beijing and there is no regional characteristic for Shanghai, I suggest removing the results of Shanghai from the paper. Or you could add more discussions and highlight the common or different characteristics shown in the two regions.

12. Page 12, Figure 7: (1) Please include the color bar. The title of the features and cluster labels are hard to see; please enlarge them. It will also be helpful to add a contour map or add a point/shape to label Beijing on the map. (2) As mentioned before, there is no significant difference between clusters, and it is hard to read the key messages from 117 plots, especially that they are all super small. I wonder whether it is necessary to demonstrate the results of all the clusters in the main text? Is it possible to only show the results of the major cluster (the cluster with the largest number) and move the results of all clusters to the supplement? A similar issue also is shown in Figures 8 and 9.

13. Page 12, line 386: For the results of unnormalized format, I wonder whether there is a trend shown in the features and the haze events have seasonality or not. If so, how would the trend/seasonality affect the clustered features?

14. Page 11, line 349-354 and Page 13, line 398-399: Do the four clusters represent different regimes/scenarios of haze events based on their differences shown in DTCV, SW1, and SW2?

15. Page 13, line 400 and line 402: It should be Figure S7 and Figure S8?