Firstly, I truly appreciate the constructive comments and suggestions from the two reviewers. A statement reflecting this appreciation has been added in the Acknowledgement section of the revised manuscript. The following are the point-to-point responses to the reviewers' comments (marked with Italic font).

Reviewer #1

Specific comments
1. The selected threshold for defining a severe haze event in the 2-class training is set to days whose surface visibility decreases below the 25 While the value of that percentile varies in time and space, I suggest elaborating more on selecting that exact percentile threshold.

The point is well taken, the corresponding text has been revised to "Although p25 values vary interannually, their long-term means actually represent…".

2. Please elaborate on the input data and its possible effects on the results. How many ground stations are used for the analysis in each city? What are the potential limitations that are resulted from the ERA5 spatial resolution of 0.25 degrees?

I assume this question matters both label and input. For labeling data, this has been made more clearly by adding "…observations in corresponding airports of these two cities during during the time from 1979 to 2019…". The input data obtained from much larger domains to reflect regional weather and hydrological conditions used for forecasting the occurrence of haze at locations of interest. The grid numbers have been provided in the original manuscript already, as in Line 176-177 "…Beijing (64x96 grids) and Shanghai (64x64 grids)…". On the resolution, the discussions of kernel size (please note that this has been moved to the Section 2 in the revised manuscript) alongside highlights in both Abstract and Conclusion have provided insights on that, i.e., the machine actually prefers feature patterns in a larger scale (5-6 degree) than a smaller one in performing the forecast, therefore, the 0.25 degree resolution of ERA5 is adequate for the purpose.

3. Please include, possibly as a supplemental, some technical details regarding the CNN analysis. What were the data and computational volumes and costs? What kind of computation platform was used, how long each training session take? What were optimization and approximation procedures implemented? Such information could assist future researchers when planning their analyses and also provide the scientific community with a technological benchmark for comparison with future projects.

A brief description has been added to the end of Section 2 as "Entire trainings have been conducted using a NVIDIA Tesla V100-SXM2 GPU cluster, costing 25s and 17s per epoch for the machine of Beijing and Shanghai, respectively". The paper also highlights the optimization of an important hyperparameter for meteorological applications, *i.e.*, the kernel size of first two CNN layers. In addition, text also added to explain algorithms such as class-weight and batch normalization. Discussions on other "routine" procedures are omitted or could be referred to Wang (2020, arXiv) to limit the paper size within a reasonable length.

Technical corrections
1. Please keep consistency in number representation along the manuscript (e.g. "11,376" in P. 4 Line 132 vs. "14975" in P. 11 Line 338.

Done.

2. Please keep consistency in technical terminology along the manuscript (e.g., "class-1" vs. "class 1", etc.).

Done.

3. Please specify the unit following physical quantities (e.g. "...heights at 500 (Z500) and 850 (Z850) hPa" in P. 6 Line 188).

Good point, the unit of geopotential height is in meter, here hPa is the unit for pressure levels. It seems to me that the meaning should be quite clear. In a later part of the cluster analysis using non-normalized quantities marked by color bars (e.g., Fig. 8 caption), all the units are indeed provided.

4. Please follow a consistent terminology for classes 0 and 1 in the 2-class analysis (e.g., "non-haze events" and "severe haze events").

The point is well taken. I have checked the text rather thoroughly to make sure class 1 and class 0 are only used when classification is concerned while the other for event-based discussions.

5. 2 Line 38: Please change "event" to "events".

Done.

6. 2 Line 39: Please change "has" to "has".

I believe the reviewer meant change "has" to "have", if so, done.

7. 4 Line 119: Please change "Introduction" to "introduction".

Done.

8. 5 Line 158: Please add punctuation marks where necessary.

Done.

9. 5 Line 176: I suggest replacing or dropping the words "longitude-latitude" that are already self-embedded in surface map objects.

Done.

10. 6 Line 214: Please change "metrics" to "metric".

Metrics is adequate here for the purpose.

11. 6 Lines 219-220: Please clarify that sentence.

The sentence has been revised to "…a validation accuracy of 80% (frequency of non-haze events or no-skill forecasting accuracy) in both…".

12. 7 Line 225: Please correct a typo "class0weight"

Done.

13. 8 Line 267: Please modify to "there are many hyper-parameters in HazeNet that need …".

I believe the reviewer was referring to Line 257. A sentence has been added there as "As in the cases of other CNNs, there are many hyperparameters need to be determined or optimized. These have been done through numerous testing trainings. In practice, it occurs that,…".

14. 12 Line 398: Please change "soli" to "soil".

Done.

15. 8 – caption (Lines 405-408): Please specify the explicit variable descriptions for better readability.

Done.