This paper designed a spatial-temporal deep learning model to estimate the daily variations and the meteorology-driven $PM_{2.5}$ concentrations in China. The authors applied the GRU and ResNet structures that have been widely used in computer vision. Thus, information on complex interactions between grid points and short-term changes in aerosol concentration is fused to obtain the final aerosol variability. The data pre-processing methods are interesting due to only the synoptic-scale variations (with a removed slow-changed background of $PM_{2.5}$ concentrations) are estimated by the model. Since meteorology is the key factor for synoptic variations in aerosol, this approach could reproduce the short-term variations in $PM_{2.5}$ concentration. In addition, the authors introduced many derived meteorological variables to the model to promote its performance. Hence the results are certainly encouraging even with no AOD and visibility information.

Overall, the paper is succinct and clearly presented. But several issues should be clarified in the manuscript as listed below.

Major comments:
1) Based on the model target described in Section 2.2, the authors treat the synoptic variations as the ratio between daily $PM_{2.5}$ concentration and the "background" concentration. And they defined the background as the 31-day running average for the current year and the previous year. I know the approach aims to remove the variations longer than a one-month timescale, but the definition is somewhat arbitrary, especially the average for the current year and the previous year (rather than the average only for the current year). Here, the author should better give some discussion for the issue.

2) The model described in section 2.3 is slightly simple, which should affect readers' reproduction of the model. At least the authors should provide the channel information of the DL model.

3) In section 3, The author compared the predictive power with the CTM (WRF-Chem) and claimed a better performance. However, the authors only compared the $PM_{2.5}$ concentrations. A comparison of synoptic variability should be added following the paper's topic.

Minor comments:
1) Sections 1 and 2. R2 -> R^2.
2) Sections 1 and 2. PM2.5 -> $PM_{2.5}$.
3) L95 in section 2.2. Quasi-static -> 4) Quasi-static
4) Figure 6. The North China Plain, Northeastern China, Sichuan Basin, etc. are regions in China that should be denoted using boxes on the map.