General Comments

This manuscript aims to develop circulation-based indices to predict different levels of PM2.5 concentrations over three main regions in China - Beijing–Tianjin–Hebei (BTH), the Yangtze River Delta (YRD), and the Pearl River Delta (PRD). The manuscript is well written for the most part. The statistical analyses used to determine the influence of meteorology on PM2.5 are robust, with the assumptions and references clearly indicated. The proposed indices improve on the capability of circulation-based indices to distinguish PM2.5 pollution levels in BTH and provide the first daily circulation-based indices specifically for YRD and PRD.

I have some comments and questions, indicated below.

Specific Comments

(1) Page 2 Line 50: Temperature is also an important factor contributing to the variability in air quality and should be included in the paragraph.
(2) Page 3 Line 72: What differentiates the analyses presented in the current study from Leung et al., 2018 and Hou at al., 2019?
(3) Page 3 Line 91: Please follow the ERA5 guidelines to cite their datasets (https://confluence.ecmwf.int/pages/viewpage.action?pageId=197704114)
(4) Section 2: The datasets used in the study (ERA5, CAQRA, GPCP) are all at different spatial resolutions. How is this accounted for in the analysis?
(5) Page 5 Line 155-160 and Figure 2: Why is there a non-significant correlation between $PM_{2.5}$ and RH in YRD?
(6) Page 6 Line 165: Formatting ($PM_{2.5}$ does not appear in-line with the other text)
(7) Page 6 Line 169: Remind the reader what SLP stands for
(8) Page 6: Line 186: The author mentions –

"Considering the moderate correlations found for YRD and PRD, we further investigate the influence of large-scale circulation on daily PM2.5 variability through its direct effect on the most important regional meteorological variables identified separately for the three regions."

However, considering that the absolute correlation coefficients do not exceed 0.12 for YRD and PRD (line 182 Page 6), please re-word this sentence to reflect that the correlations are "low".

(9) Section 4.3: Why are the circulation variables used here (PRD) different from the other two regions?

(10) My main question is – if the most relevant meteorological fields explain more variance than the circulation indices, what is the value in using these indices? The authors briefly answer this in the last paragraph: **"Although the circulation indices explain less variance than the most relevant regional meteorological fields for YRD and PRD, we expect weather**

prediction and climate models to better represent these features of the large-scale circulation than regional meteorological fields such as surface wind speed and RH."

I would encourage the authors to elaborate on their argument and present more evidence for this – why would weather prediction and climate models better represent these indices over the meteorological fields?

(11) Finally, I found the last section "Discussion and Conclusions" to be too long. I would recommend cutting it down to a paragraph or two, to provide a succinct summary of the main findings from their analysis.