Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2020-237-RC2, 2020

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Recommendations on benchmarks for photochemical grid model applications in China: Part I – $PM_{2.5}$ and chemical species" *by* Ling Huang et al.

Anonymous Referee #3

Received and published: 6 August 2020

Severe air quality problem in China has recently attracted attention from the public as well as the scientific community. Photochemical grid models (PGMs) are frequently used to investigate the phenomenon and develop emission control strategies. The number of PGMs-based research articles for scientific and regulatory applications has surged. This study aimed to apply model performance evaluation (MPE) methodologies developed in U.S. to evaluate PGMs used in China for total PM2.5 and speciated PM components. A total of 128 recent peer reviewed articles based on one of four most popular PGMs were compiled and different model configurations as well as statistical metrics were evaluated. The benchmarks were developed and recommended for two tiers: "goals" and "criteria" for evaluating PGM applications in China. Although

the methodologies and metrics used in this study are not novel, or adopted from several studies conducted in U.S., the derived results/recommendations/conclusions are scientifically sound and its logic or context is reasonable. This study is expected to provide guidance for future PGM evaluations in China.

The manuscript is well written and the logic and context are well presented and easy to follow. Several comments are provided below in hope that these will assist the authors strengthen the manuscript.

Technical comments:

The methodologies and metrics adopted in this study are well established and published in several literature, and 128 relevant modeling studies conducted in China were compiled in this study for the model performance evaluation. Although the information provided in this study is useful for the modeling community, the analysis was relatively straightforward so I would consider this study a critical literature review, instead of a novel study. to strengthen the scope of this study, one would expect that the authors go beyond what was accomplished in the U.S. studies and consider additional analyses such as the following.

Although the manuscript describes the reasons why China specific modeling performance evaluation are needed, there are many commonalities across the air quality modeling community worldwide so comparison can be made among studies conducted in China or elsewhere. Emery et al (2017) indicates that "While we primarily address U.S. modeling and regulatory settings, these recommendations are relevant to any such applications of state-of-the-science photochemical models." The comparison of benchmarks from this study with Emery et al (2017) shows similarities. Thus, it seems that the benchmarks developed in China in this study confirm their worldwide applicability for other super-regional, regional, or local modeling domains. It would be valuable if the authors discuss the broader implication of these findings.

A total of 128 peer-reviewed articles were compiled for this study. Are there articles or

studies that were excluded from this study but could be potentially included by reapplying the metrics used in this study? Some of the studies may not report any MPE results but could be recalculated to get MPE results if needed. Please add some discussion on those studies, especially on those with speciated PM components since the number of these studies is very limited. If applicable, please include any additional studies so the dataset is larger or more meaningful. In addition to peer-reviewed articles, there may be non-peer-reviewed reports which deal with PGM applications (e.g., US EPA's PGM reports). I wonder if there are such reports published by Chinese central or provincial government or NGOs that can be included in this study.

On the other hand, are there cases (excluded in this study) that the authors can reapply the benchmarks recommended in this study to demonstrate the improved model robustness or validity? For example, there may be PGM studies that did not use the metrics adopted in this study but the evaluation may be improved after these metrics or benchmarks are applied.

Some benchmarks for speciated PM components are questionable due to the number of available studies, which may lead to biased or inconclusive results. Although caution is warranted, I wonder if the dataset can be enriched by including some studies elsewhere (e.g., U.S. studies) since benchmarks for speciated PM components were not studied in Emery et al (2017). I understand the focus of this study is in China, but it seems that benchmarks developed in China and U.S. are valid, comparable in both countries.

In-depth discussion on statistical metrics in "Impact of temporal and spatial resolution" (page 7, lines 35-37) is needed. This is counter-intuitive that the wider ranges are associated with the larger number of data points. What data is needed to improve the confidence on the benchmarks developed for speciated PM components?

Minor comments:

Page 5, line 29: Table 2 should be Table 1.

Page 6, line 40: please check the number. It seems one single study is in spring, not summer. There are 5 studies in summer.

Page 7, line 5-14: "Figure 6" is missing in this section and should appear somewhere.

Page 7, line 4 and 15: "Impact" is misspelled.

Page 7, line 33-34: it seems the R values correctly correspond to the coarsest resolution (80km) but off to the finest resolution (3km).

Page 8, line 5-13: it seems that the R values in the text correspond to different percentile. For instance, the 33rd percentile value should be 0.64 for hourly to 0.91 for monthly results while the 67th percentile should be 0.5 for hourly to 0.70 for monthly. Please check the remaining values in the text against Figure 9.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2020-237, 2020.