Geosci. Model Dev. Discuss., 2, C591–C600, 2010
www.geosci-model-dev-discuss.net/2/C591/2010/
© Author(s) 2010. This work is distributed under
the Creative Commons Attribute 3.0 License.

Geoscientific Model Development Discussions

Interactive comment on "Assessment of bias-adjusted $PM_{2.5}$ air quality forecasts over the continental United States during 2007" *by* D. Kang et al.

Anonymous Referee #2

Received and published: 3 March 2010

GENERAL COMMENTS

This paper evaluates a post-processing method that combines recent and new hourly PM2.5 predictions from a source-oriented air-quality modeling system (WRF-NMM-CMAQ) with recent hourly PM2.5 measurements from a network of monitoring stations in order to produce improved forecasts at the locations of those monitoring stations (though not elsewhere). A set of model PM2.5 forecasts and measurements from the North American AIRNow meta-network for a one-year period have been used to evaluate this method. The same Kalman Filter Predictor bias-adjustment approach has been used previously for ozone forecasts and was shown to improve forecast skill.

However, due to a number of differences between PM2.5 and ozone, there was no reason to expect the same result a priori for PM2.5, but the results presented in this paper suggest that this technique does improve skill for PM2.5 forecasts as well on average.

This is a well-written and useful paper. It extends previous work on ozone forecasting and shows that the KF bias-adjustment method can add skill for forecasting of PM2.and could be implemented operationally in conjunction with the near-real-time PM2.5 measurements that are available for much of North America from the AIRNow meta-network. I have made a number of specific comments and suggestions below that I would ask the authors to consider, as I believe that addressing them would strengthen the paper further. My recommendation is to accept this paper for publication in Geoscientific Model Development conditional upon minor revisions.

SPECIFIC COMMENTS

1. It is not mentioned in Section 1 whether the U.S. Air Quality Index is defined in terms of hourly values or 8-hourly values or daily maximum values or other quantities. One reason for raising this point is that the analyses described in the manuscript are restricted to observed and forecasted daily mean PM2.5 concentrations even though the measurement data set used had hourly time resolution and both raw and KF bias-adjusted model predictions were also available every hour. It seems that an opportunity to look at diurnal variations in model error, the hour-specific performance of the KF bias-adjustment technique, and any improvement offered by post-processing to the prediction of daily maximum PM2.5 values was missed.

2. There seems to be a disconnect between the Eta model references that are given on p. 1377, l. 8 and the use of the WRF-NMM model noted on p. 1379, l. 14. References for the latter would seem to be more appropriate if that is the meteorological model on which the present study was based.

GMDD

2, C591–C600, 2010

3. On p. 1378, l. 5, it would strengthen this statement if substantiation were provided by referencing several publications as examples of the use of post-processing bias-adjustment techniques with NWP model forecasts.

4. It is not made clear in the manuscript that the approach being discussed is only applicable at locations where near-real-time PM2.5 measurements are available. I have made a few suggestions in the Technical Corrections section of places in the manuscript where some clarification could be added. Neither is there any discussion of the limitations and inconsistencies introduced by this approach. By that I mean (a) that this technique cannot provide any guidance for model predictions away from monitoring station locations and (b) that it introduces implicit inconsistencies between model predictions in between monitoring stations and the bias-adjusted predictions at those station locations that a forecaster would have to be aware of. A few groups have suggested objective-analysis approaches (e.g., Blond et al., 2003, J. Geophys. Res., 108, doi:10.1029/2003JD003679) that it might also be possible to apply to PM2.5 forecasts to address this second problem by modifying the gridded forecast based on the bias-adjusted point-specific forecasts.

5. I acknowledge the ever-present tension between brevity and completeness in describing a study, but there are some details missing from Section 2.1 that I would have liked to have seen:

- What were the horizontal domain, map projection, model top, vertical coordinate, and horizontal and vertical grid spacings used by the WRF-NMM model in this application?
- What was the CMAQ vertical coordinate and model top that was used?
- What was the base year for the U.S. EPA national emissions inventory that was used? Which emissions inventories if any were used to account for Canadian and Mexican emissions?

GMDD

2, C591–C600, 2010

Interactive Comment

- Were biogenic emissions considered, and if so, how?
- Were any intermittent, natural PM emissions sources, such as sea salt from wave breaking, wildfires, and wind-blown dust, considered? If not, this should be noted along with the implication of an expected negative bias for PM concentrations as a consequence.
- What were the chemical lateral boundary conditions used by CMAQ? Did they vary at all in time and in space? Would trans-Pacific transport episodes be represented?
- Which model versions of WRF-NMM and CMAQ (and PREMAQ) were used in this study?
- (p. 1380, l. 14) Was just one 48-hour CMAQ forecast made per day, and if so, why was the 06 UTC time chosen when the meteorological forecasts starting at 00 UTC and 12 UTC are based on more meteorological observations?

6. The discussion in Section 2.2 of the PM2.5 observations misses a number of issues. First, it is not mentioned but the TEOM measurement bias is known to vary with ambient temperature and hence with season, so that wintertime biases are considerably larger than summertime ones. Second, were the TEOM measurements that were reported to AIRNow bias-corrected or not before transmission to AIRNow (this may also vary by network)? This is a fundamental issue because if the measurements themselves are biased, then the bias-adjustment technique will adjust model predictions towards the measurement bias. Third, some beta attenuation measurement (BAM) instruments are also employed in North America to measure real-time hourly PM2.5 concentration, and these instruments have different error characteristics than TEOMs; were any measurements from these instruments considered? And fourth, it might be worth noting that although other PM2.5 concentration measurements are made

GMDD
2, C591–C600, 2010

CC BY

and are available retrospectively (e.g., IMPROVE network), they are not near-real-time measurements nor are they available through AIRNow, which means that they are not useful for bias adjustment of operational forecasts.

7. Figure 1 has some problems. First, the caption suggests that it shows the forecast domain whereas it only shows that the domain includes 48 states of the U.S.A. plus the District of Columbia. It does not show where the lateral boundaries of the domain are actually located, which would be valuable information. Second, the figure (and Figure 7) shows that PM2.5 measurements are also available from southern Canada, but nowhere is it mentioned (e.g., Section 2.4) whether the Canadian measurements are used or not in the calculation of the continental and subregional statistics (i.e., Tables 1 and 2, Figures 2, 4, 5a, 6, 8, 9, 10, 11). MB was calculated for these stations for Figure 7.

8. The second paragraph of Section 2.4 describes a subregional division of the continental U.S. into seven regions based on analysis of an O3 climatology, but the authors do not indicate whether they believe that this division is similarly relevant for PM2.5, given its different characteristics from O3. Figure 1 also shows only six regions, not seven?

9. I think the sentence at the end of the third paragraph of Section 2.4 sows confusion. First, no explanation is given as to why the cool season was divided into two parts (perhaps in recognition of the significant change to CMAQ noted in Section 2.1 that was made in mid-September 2007, perhaps not). And second, none of the analyses presented in the figures and tables thereafter mention these two subseasons; instead, they all seem to be based on measurements and predictions from Jan. 1 to mid-April and from September to December 2007 (see Table 1 and Figures 5, 7, 8, 10, and 11). There is also only one subsequent reference to a first or second cool season thereafter in the text that I noticed (p. 1383, l. 4). If the authors wish to retain the terms “first cool season” and “second cool season”, perhaps they should also refer to the “*full* cool season” in the appropriate table and figure captions.

GMDD

2, C591–C600, 2010

10. Based on the formulas in Kang et al. (2005), the calculation of the two categorical statistics in Section 3.4 of this paper appears to have been done correctly. However, different practitioners use different names for these two statistics, and, as noted recently by Barnes et al. (2009, Wea. Forecasting, 24, 1452-1454), there has been much confusion in the literature between false alarm rate (F) and false alarm ratio (FAR). As a service to the reader, it would be helpful if alternate names for these two statistics could also be mentioned and the Barnes et al. (2009) note referenced. For example, perhaps p. 1388, l. 22-23 of the manuscript could be reworded as follows:

"Figure 11 displays the false alarm ratio (FAR; also known as probability of false alarm) and hit rate (H; also known as probability of detection)(see Kang et al., 2005; Barnes et al., 2009) for the raw model and ..."].

It would also help a non-specialist reader if the possible range and the interpretation of extreme values were given for FAR and H in this section.

11. The discussion of Figure 11 near the end of Section 3.4 states that the KF forecasts "increased the H values for all the sub-regions except for the LM and RM". Would it be more correct to append the phrase "in the warm season and the UM in the cool season"?

12. Two references cited on p. 1378 (McKeen et al. 2007 and Appel et al., 2008) are missing from the References section.

13. Following the example of Table 3 of Eder et al. (2009, Atmos. Environ., 43, 2312-2320), it would be useful to add three columns to both Tables 1 and 2 to provide sample size N, observed mean O, and modeled mean M for each subregion.

14. For Figure 10, would it be possible to indicate the number of samples for each bin somehow either in the text or on the figure itself?

TECHNICAL CORRECTIONS

GMDD

2, C591–C600, 2010

p. 1376, l. 2. Change “particular” to “particulate”.

p. 1376, l. 15. Capitalize “pacific coast”.

p. 1376, l. 18. Change “systematical” to “systematic”.

p. 1377, l. 12. “integral(?) data set”

p. 1377, l. 20. Perhaps “emissions *and* physical, chemical, and ...”.

p. 1377, l. 22. Insert comma before “resulting”.

p. 1377, l. 23. Change “poses” to “pose”.

p. 1377, l. 29. (Also p. 1378, l. 6) Insert hyphen between “bias” and “adjustment”.

p. 1378, l. 5. Change “model” to “models”.

p. 1378, l. 7. Would suggest inserting the phrase “at locations with PM2.5 monitors” before “is warranted” [see Specific Comment 4].

p. 1378, l. 11. Change “refer” to “referred”.

p. 1378, l. 16. Perhaps “It was not clear whether they would be readily applicable for PM forecasts ...”.

p. 1379, l. 9. This is where it might be expected, but there is no “bridge” here to the rest of the manuscript via a brief description of the structure of the rest of the manuscript.

p. 1379, l. 15. This is the first use of AQF but this acronym is not defined (same comment for “PDFs” on p. 1385, l. 3).

p. 1379, l. 16. Perhaps “... model, which simulates the transport ...”.

p. 1379, l. 17. Change “substance” to “substances”.

p. 1379, l. 21. Change “spacings” to “spacing”.

p. 1380, l. 4. Perhaps “... can improve *point-specific* forecast results over the raw

GMDD

2, C591–C600, 2010

model ...” [cf. Specific Comment 4].

p. 1381, l. 24. Perhaps “from then on” in place of “further on”.

p. 1382, l. 4. Change “Normalize Mean Error” to “Normalized Mean Error”.

p. 1382, l. 5. IOA is not included in this list but is used in Figure 8.

p. 1382, l. 7. Perhaps “Two categorical metrics, False Alarm Ratio (FAR) and Hit Rate (H), are used ...”.

p. 1382, l. 16. “... time series *are* ...”.

p. 1382, l. 18. “... during *the* cool season ...”, “... during *the* warm season ...” (see also p. 1385, l. 2 and l. 12, p. 1386, l. 17 and l. 20, etc.).

p. 1382, l. 20. “... the time series *are* divided into ...”.

p. 1382, l. 22. Change “Further more” to “Furthermore”.

p. 1383, l. 3. Perhaps “... overestimated PM2.5 concentrations *on average* during the ...”. p. 1383, l. 20. “... Coast presents a completely different story, ...”.

p. 1383, l. 26 Change “attributed” to “attributable”.

p. 1384, l. 23. Don’t need hyphen here in “bias adjustment”, since it is not being used as a compound adjective. Same comment on p. 1385, l. 21, p. 1386, l. 10, etc.).

p. 1386, l. 23. Perhaps “reduce” instead of “rectify”.

p. 1388, l. 25. Perhaps “An *exceedance* threshold value of ...”.

p. 1389, l. 13. Suggest “... during the year of 2007 *for locations with hourly PM2.5 monitors*” [see Specific Comment 4].

p. 1389, l. 21. Perhaps “the warm season; *in contrast* the opposite is true ...”.

p. 1389, l. 26. Suggest “... significantly improved the PM2.5 forecasts *for locations*

GMDD

2, C591–C600, 2010

with hourly PM2.5 monitors as revealed by reductions ...” [see Specific Comment 4].

p. 1390, l. 3. Perhaps “... transition of seasons or model changes”.

p. 1390, l. 29. Perhaps “... in issuance of air-quality-degradation-related health advisories”.

p. 1392, l. 21. Change “Mckeen” to “McKeen”.

Fig. 1. In caption change “AIRNOW” to “AIRNow”.

Fig. 2. Add to caption that time series are daily means *over all stations*.

Fig. 3. Add “subregions” at end of caption.

Fig. 4. Revise caption to indicate that scatterplots are based on forecasts and measurements for all stations and all days.

Fig. 10. Third line of figure caption seems out of place; e.g., why is Figure 9 referred to?

Fig. 11. Why not continue to use the abbreviation “MOD” instead of “MD” (cf. Tables 1 and 2 and Figures 2, 3, 7, and 8)?

One general comment is that GMD is an international journal and it should be clear from the article text that a air quality modeling system for the U.S. is being discussed. There are several places in the text where this clarification could be made:

p. 1376, l. 2. Insert “for the US” after “forecasts”.

p. 1376, l. 25. Insert “in the US” before “to compute the Air Quality Index”.

p. 1377, l. 3. Insert “in the US” after “publicly available” (otherwise, reference list is parochial: e.g., see http://www.esrl.noaa.gov/csd/events/iwaqfr/ for other groups and agencies involved in real-time ozone and PM forecasting).

p. 1388, l. 25. Perhaps “... based on the U.S. National Ambient ...”.

GMDD

2, C591–C600, 2010

p. 1389, l. 16. “the entire domain” or “the continental U.S. portion of the domain”?

Interactive comment on Geosci. Model Dev. Discuss., 2, 1375, 2009.