

General comments

In this manuscript (AAS-2017-0179), the authors presented a study of comparisons of three-dimensional variation data assimilation (DA) and model output statistics (MOS) on improvement of numerical model forecasts of gaseous and aerosol-phase species. To my understanding, it is not appropriate or not fair to make a direct comparison between these two methods although the authors also pointed out two problems or issues hindering the comparisons between DA and MOS (see Lines 65-77 in the manuscript). This is because the forecast improvement made by DA is only associated with the reduction of uncertainties of initial conditions of chemical species by assimilating observational data whereas the forecast improvement made by MOS is associated with not only the reduction of uncertainties of initial conditions of chemical species, but also the reduction of uncertainties associated with meteorological inputs, emissions, and chemical transport models themselves. Evaluating the MOS with DA and without DA definitely has important implication on improvement of air quality operational forecasts. It is not surprised that DA performance was degrading substantially as the forecast time increased. But it is a little bit surprised that the sensitive run of MOS with DA did not show competitive performance as compared to the MOS only. I am concerned with the MOS and DA methods used in the study. The detailed comments can be found from major comments below. In addition, the manuscript was poorly written. There are a lot of grammar errors and many sentences need significant improvements. Therefore, major submission is recommended to address the following comments.

Major comments

- 1) It is not clear how the first 24-hr forecasts were corrected by using the MOS since no observational data were available during the whole forecasting period. It seems ok at some extent to use the first 24-hr corrected forecasts as the proxy to correct the following 24-hr forecasts when the air quality did not show large change from the first to the second day. However, the MOS correction may cause large forecast bias when the air quality experienced a sudden change from day 1 to day 2 or from day 2 to day 3. How did the MOS handle with those situations?
- 2) Why did the DA method only assimilate part of the observational data and leave the rest of observational data for the verification? This may have a significant impact on the DA results. It is suggested to use all the observational data in the DA application and the DA performance evaluations.
- 3) In section 2.2, the authors mentioned that “the MOS found the statistic relationship from the training samples and applied this relationship to correct the model forecast”. How can the statistic relationships be found from the training and how they were applied to correct the model forecasts. Please illustrate how this can be done in this study.
- 4) On Line 171, “... was integrated 84 h to generate 72 hours’ forecasts of each day with the earliest 12 h being discarded as the spin-up time”. Was the method used by all four sensitivity runs including the cases with DA? How can the verification results be changed if the first 12-hr forecast results were included? In general, it is feasible to use the first 2-3 days as spin-up runs for the one-month simulations but no need to exclude the first 12-hr simulations from each day runs throughout the whole month.

- 5) It is weird that the outer domain was defined with the horizontal resolutions (75 km×75 km) coarser than the NCEP GFS data (0.5°×0.5°) which were used to generate initial and lateral boundary conditions for the simulations in this study.
- 6) The writing has too many issues including grammar errors, inappropriate words, and wrong sentence structures. A heavy edit work is needed.

Specific comments

L20: It is not “model improvement”, it should be “model forecasting improvement” ?

L22-23: “Such superiority of ... could be seen ... and extend to ...” looks very weird. Please rewrite the sentence.

L25, L44: be careful to use “—”.

L33: The statement of “the top priority of local government” needs specific information like the government’s work, or policy or something else.

L34: Numerical models cannot reach to the accurate stage in any case. “A must” is impossible. Please be careful to use any strong words unless you have enough evidence. I saw several places with the same issue.

L38-41: It is difficult to understand the sentence. Please rewrite it.

L48: I do not think that the emissions and conversion rates of SO₂ can be adjusted by using data assimilation. Please double check the original reference and make sure that the citation is correct.

L52: “tried “?

L54: delete “through processing the model outputs”.

L65-65: The statement of “Two problems ... observation” is difficult to understand.

L66 and L72: Firstly → First, Secondly → Second.

L66-L69: Are you sure that no surface observational data were used in DA for any other air quality modeling study and operational forecasting?

L73-75: Please make sure that readers can understand your statements. Be careful to use long sentence.

L82: How did you modify the MOS scheme? This can be considered your contribution to the MOS method in this study. You should highlight this work and add it to Section 2.

L85: “will be” should be consistent with other verbs’ tense here.

L102: Spelling out RADM2, MADE, and SORGAM.

L110-112: What do “situation-independence” and “stable effects” mean? In addition, the sentence needs to be rewritten.

L113: “will not be”?

L115-116: $x(t)$ and $y(t)$ appeared here only. Should you really need them?

L123: What does “low-grade” mean here?

L129: mass should be matter.

L131-133: the sentence needs to be rewritten.

L137: L148: Did you present any results about the representative error ε_r in the result analyses?

L154: How did you define observational error and measurement error? What is their main difference?

L164: “.”?

L166-167: It is better to modify as “Sim_DM represents the case using both DA and MOS methods” something like that.

L169: change to “... between 30 November 2014 and 31 December 2014”.

L177: supplied -> generated. Again the whole sentence needs to be rewritten.

L180: observation data → observational data. Please correct the same problem in other places.

L186: with → and?

L187: who or which? Check the same issue throughout the manuscript.

L198-199: You cannot say that “simulated concentrations seriously overestimate the observed values”. The whole sentence can be written as “The model over-predicted surface O_3 by xxx”.

L202: Need exact definitions of error and bias here.

L204: What about meteorological inputs?

L204-205: What does “24 h forecast” represent? It represents the forecast at the 24th hour or over the period of 1-24 hours?

L211: “concentration are”?

L221: “RMSE were”?

L222-223: delete “what’s more” and “This is to way”?

L224: “could” → “could be”?

L233-234: “will not”?

L241-243: Do you mean “from the -12 forest hour (FHR) to the 10 FHR”? The sentence should be rewritten as “Figures 5 and 6 show ..., respectively”.

L273: “DA produced”?

L317: “forecast 0 h to forecast 11h” should be “the 1st forecast hour (fhr) to the 12th fhr”. Please make correction for other similar issues.

L325: decides → determines.

L324-325: Please provide detailed explanations on why the error’s temporal consistency, rather than its magnitude determines the MOS performance.

L329-330: “... whenever the three changes their setup” needs to be corrected.

L336: Lead to → Leads to. Please rewrite the whole sentence.

L337: delete “a period of only 1 month” → “one month”.

Lxx: It is very tired to list all the writing issues or errors.

P26, Figure 4. All statistical parameters should be put in one table for an easy comparison. All the fonts and axis labels are too small. Figure 4 needs to be replotted.

Figure 5: Please redefine the color bars.

L526: The figure caption needs to be rewritten as “Similar to Figure 5 but for correction coefficients”.

Figures 7 and 8: Please replot two figures by increasing font size and rewrite the captions.

P32/Fig. A.1: 24 Forecast → the 24-h Forecast, 24 Correction → the 24-h Forecast? The font size is too large.

L607: stand for → stands for. Please rewrite the figure caption.