In this manuscript, 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 comparison between DA and MOS (see Lines 65-77). 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 model itself. Evaluating the MOS with DA and without DA 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 both MOS and DA methods used in the study. The detailed comments can be found below. In addition, the manuscript is not well 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 can be corrected by using the MOS since no observational data were available during any forecasting periods. It seems ok at some extent to use the first 24-hr corrected forecasts as 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 a 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 use the rest of observational data for verifying the DA’s performance? 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”. Please illustrate clearly how this got 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 12-hr simulations for every day throughout the whole month.
5. It is weird that the horizontal resolutions of the outer domain (75 km×75 km) were even coarser than that of 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 sentence structures. A heavy edit work is needed.

Specific comments

L13: What does “their” represent?

L13-14: Forecast biases caused by uncertainties of met inputs, emissions, and model itself but not the uncertainties themselves can be decreased by using MOS. The statement must be precise. There are many other places having similar issues.

L16: “Clarified” is not an appropriate word here. Please check the similar issue throughout the text.

L20: not “model improvement”, it should be “model forecasting improvement”, something like that. Again, there are too many issues like this. I am not going to list all in the review report.

L22-23: “Such superiority of … could be seen … and extend to …” looks very weird to me. Please rewrite it. Again, check similar issues throughout the manuscript.

L25, L44: It is better to use a separate sentence, rather than use “──”.

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

L34: Numerical models cannot reach the accurate stage forever. Again, be careful for using the words in the writing.

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

L48: I am not sure that emissions and conversion rates of SO2 can be adjusted by using data assimilation. Please clarify this 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 or First, Secondly or Second.

L66-L69: Are you sure that no surface observational data were used in DA for the air quality forecasting research?

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 for the MOS method. You should highlight this work and add it to Section 2.

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

L102: full names of RADM2, MADE, and SORGAM.

L110-112: What do “situation-independence” and “stable effects” mean? Please illustrate them clearly. In addition, the sentence needs to be rewritten although I do not want to point out any more.

L113: “will not be”?

L115-116: $x(t)$ and $y(t)$ did appear once in this manuscript. 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 presentative error $ε\_{r}$ in the result analyses?

L154: What is the main difference between observational error and measurement error?

L164: “.;”?

L166-167: It is better to modify as “Sim\_DM represents the case with or 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 in several other places.

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 O3 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 ?