Response to reviewer 2

Authors appreciate reviewer's thoughtful comments and suggestions, which are greatly helpful for us to improve our manuscript. The manuscript has been revised to accommodate the reviewer's comments.

General comment and response:

General Comments: This paper discussed an application of offline WRF-CMAQ simulations and evaluation both with and without DA during the KORUS-AQ measurement campaign in Korea during May 1 – June 12, 2016. While the paper is well written, and the DA methodology is sound, there are some concerns on missing details and discussion throughout the paper (see Specific comments), rather expected model performance results, and using a such a case study/campaign application as an appropriate demonstration for the readiness of an "operational" air quality forecast system (below). While I recommend accepting the article for publication, these issues should be addressed.

I think it is a bit misleading to call this application of WRF-CMAQ simulations a true "operational" air quality forecasting system, as it is currently only applied for a very short time period (May 1 – June 12) during a detailed measurement campaign. Such detailed observations are likely not available to demonstrate a continuous refresh of satellite/surface observations for DA in an operational system. To support this model system's use as "operational", what confidence is there that the statistical performance will scale to the remainder of the calendar year under different meteorological and chemical (i.e., emissions) environments? I think the "operational" focus of the paper should be dialed back in the paper, in place of a focus on the application of offline WRF-CMAQ chemical DA and evaluation during an intensive measurement campaign in Korea. Towards the end of the paper, it could be discussed of this WRF-CMAQ configuration could be further developed and more comprehensively tested to become an operational air quality forecasting system for Korea.

This issue is compounded by the somewhat expected results in the paper of improved model performance by assimilating data (compared to no DA), the impacts of a higher frequency rate of DA, and lack of testing the important "trade-off" that exists between increased precision and computational cost of DA. In fact, the authors explicitly state that this unsurprising result needs to be further tested, and that the true operational system should be designed under these considerations. These facts further substantiate my concern that it may be premature to consider this paper a description of an "operationally ready" model, but rather an application of testing DA in CMAQ during a relatively intensive KORUS campaign, which is a region that would certainly benefit from further development and testing of an operational air quality forecast model.

Response: Since we agreed that further researches are needed for true "operational" air quality prediction system, the word of "operational" has been removed from the original manuscript, and the title of the paper has also been modified to "Development of Korean Air Quality Prediction System version 1 (KAQPS v1) with focuses on practical issues" in order to avoid any confusion.

Actually, based on this work, the multiple-year tests are being currently conducted with the current sets of WRF-CMAQ-DA system. Once these tests are finished, we will revisit and report the issue of the "operational" air quality prediction system in South Korea, again.

Specific comments and response:

Comment: S1. Lines 88 – 93: Don't like the interchanging of air quality and chemical weather terminology here, as it is too similar to suggest that online chemical weather feedbacks are necessary in air quality models. I suggest revising to avoid any confusion.

Response: All of the terminology "chemical weather" has been replaced with "air quality" in the manuscript to avoid this confusion. Please, check out pp. 4:85 – 90.

Comment: S2. Lines 95 – 96: This is partially being overcome by new high spatial and temporal resolution satellite observations of composition (e.g., TEMPO, TROPOMI, GEMS, etc.). I think although they lack the longer term records, it should be mentioned that strides are being made at tackling these issues for air composition observations.

Response: We totally agree with referee's comment. A paragraph has been added to the revised paper for explaining the efforts to improve spatial and temporal coverage of satellite measurements. Please, refer to the added parts (pp. 5:93 – 101).

Comment: S3. Lines 98-101: What about the issues of coarse model grid scale for CTMs?

Response: Coarse-grid scale for CTM simulations can increase model uncertainties for several cases (Shrestha et al., 2009; Sirithian and Thepanondh, 2016), but not for all. After consideration of the matter, we have decided not to address this issue in detail, because it is beyond the scope of our manuscript.

Comment: S4. Lines 142-155: This section is lacking information.

1) What are the dynamical/physical configurations for WRF (e.g., LSM, land use data, sfc layer, PBL, grid scale microphysics, convective cloud parameterization, etc.) ?

Response: The dynamical and physical configurations for the WRF model simulations were added in the revised paper. Please, see pp. 7:151 – 157.

2) The met processor needs to be defined (i.e., MCIP) and explained for important derived variables from WRF to drive CMAQ. Many physical inconsistencies can arise between WRF-MCIPCMAQ, and this is pivotal information in understanding the physical linkages between the upstream physics in WRF to drive CMAQ.

Response: Information of MCIP has been added to the manuscript. Please, check out pp. 8:169 – 174.

3) How exactly is OBSGRID applied in this model, and how does it relate to the WRF physical configurations?

Response: The OBSGRID does not relate to the WRF physical configurations. This process was conducted to improve the accuracy of initial and boundary conditions for the WRF model simulation via data assimilation technique. A short sentence explaining OBSGRID has been added with more detailed information. Please, check this out at pp. 8:160 – 165.

4) Why was 15x15 km chosen, as opposed to commonly applied forecast models at 12x12 km? While much of this could be provided in supplemental/appendix to preserve brevity in the main text, it still needs to be included somewhere in the manuscript.

Response: Based on Lee et al. (2016)'s work, we chose 15 km by 15 km for CMAQ.

Comment: S5. Lines 157-177: This section is also lacking information.

1) What are the main chemical configurations for CMAQ (e.g., gas-phase chemistry, aerosol mechanism/size, dust/sea-salt, aqueous phase, dry/wet deposition, etc.)?

Response: The main chemical and physical configurations for the CMAQ model simulations were added into the revised manuscript. Please, refer to pp. 8:180 – pp. 9:189.

2) How does the CMAQ configurations interact and physically link with the upstream, driving physical configurations of WRF (comment S4)?

Response: Information on physical configurations of WRF and CMAQ model simulations has been added to the revised manuscript (refer to the responses to comments S4 and S5).

3) Clarification is needed on why MEGAN, rather than in-line BEIS, was used for biogenics in CMAQ. Is this based on literature for the Korean region?

Response: MEGAN was applied to the CMAQ model simulations in this study, because the MEGAN has been utilized in Korean modeling community and has also been widely used in many studies focusing on East Asia including South Korea (Kim et al., 2014; Kim et al., 2017; Lee et al., 2016; Park et al., 2014; Souri et al., 2017).

Comment: S6. Lines 230 – 255: I think this is an important discussion, but please make it clear how much this is a different formulation of AOD compared to the other pre-existing AOD calculations in CMAQv5.1 (e.g., the reconstruction method).

Response: A paragraph has been added in the revised paper. Please, see pp. 13:277 – 286.

Comment: S7. Lines 311 – 316: I have some issue with assuming that $\Delta PM_{2.5}$ exactly scales with ΔPM_{10}, because the inherent differences in some of the sources that make up the PM_{10} bias compared to $PM_{2.5}$. In other words, if most of the ΔPM_{10} is due to missing coarse mode aerosol emissions (e.g., dust etc.), we wouldn't expect this difference to have the same effect on $\Delta PM_{2.5}$.

Response: Yes, it may be a correct point! Our problem has been that the missing sources of the coarse-mode discrepancy (i.e., $\Delta PM_{2.5\text{-}10}$) have not been identified in South Korea. We are thinking that the uncertainty in the fugitive dust emissions from construction sites, road sites, cattle-raising areas, dry mud fields, etc may take significant responsibility for the $\Delta PM_{2.5\text{-}10}$ in South Korea. However, the amounts of such emissions from individual source have been difficult to quantify. In addition to a method we used in this study, we have to add the above amounts in the future study. This may be the reason why Fig. 9 shows larger differences in the PM_{10} predictions than in the $PM_{2.5}$ predictions.

Comment: S8. Lines 391 – 392: Could this also be due to underpredicted NO_2 with the DA run and not enough nighttime ozone titration? This perhaps could be better explored using Ox relationships and looking into different regions of the domain.

Response: A following sentence has been added into the revised manuscript. Please, see pp. 19:424 – 428.

Comment: S9. Lines 397 – 398: It is concerning to draw such a conclusion based only on a single SNU lidar site comparison. Is this truly a widespread issue for nocturnal boundary layers in Korea? While this may indeed be common and well-defined previously using similar WRF physical options as in this study, there needs to be appropriate references here to provide support for your argument.

Response: Although we showed only one site example here, this nocturnal MLH problem has been commonly found in South Korea. Korean modeling community have also been well aware of this problem for a long time (Kang et al., 2016; Nam et al., 2016). A sentence in Lines 397-398 of the original manuscript has been modified, because a more intensive comparison study between lidar-retrieved and model-simulated MLH is necessary. Please, see pp. 20:435 – 437.

Comment: S10. Lines 406 – 408: This methodology is not clear. What set of MLH observations would be used for this effort? It certainly cannot be based on the single SNU lidar site. How would this be done "operationally" in the offline MCIP step between WRF and CMAQ? Bias correcting the MLH may lead to physical inconsistencies with other driving meteorological fields from WRF that were based on a particular set of physical configurations. Overall, this requires more thought likely, raises some concern, and should probably just be removed from the paper.

Response: Following reviewer's suggestion, the related sentences of Lines 406-408 have been removed from the manuscript since we agree that careful consideration of the MLH bias correction is needed. We are thinking that the bias correction will not lead to physical inconsistency in the “off-line” mode modeling, but it can create a problem in the “on-line (two-way)” modeling.

Comment: S11. Lines 412 - 413: CMAQ already has the “capability” to predict aerosol composition. Thus, it should be restated to say a “: : :a strong capability of our DA system is to improve predictions of CMAQ aerosol composition”.

Response: We modified the corresponding sentence into “a strong capability of our DA system is to improve predictions of CMAQ aerosol composition”. Please, check this out at pp. 20:441 – 442.

Comment: S12. Lines 416 – 424: These changes in model performance would be elucidated if an addition column showing the absolute bias difference plot (colored in Red Blue shading) for the two runs compared to surface observations in Figure 10. This can be achieved by interpolating the closest model to the observations points.

Response: Yes, it is a good idea! Following reviewer’s suggestion, bias difference plot has been added into Fig. 10, and the caption of the figure has also been changed. Please, check out the modified Fig. 10 at pp. 45 in the revised manuscript.

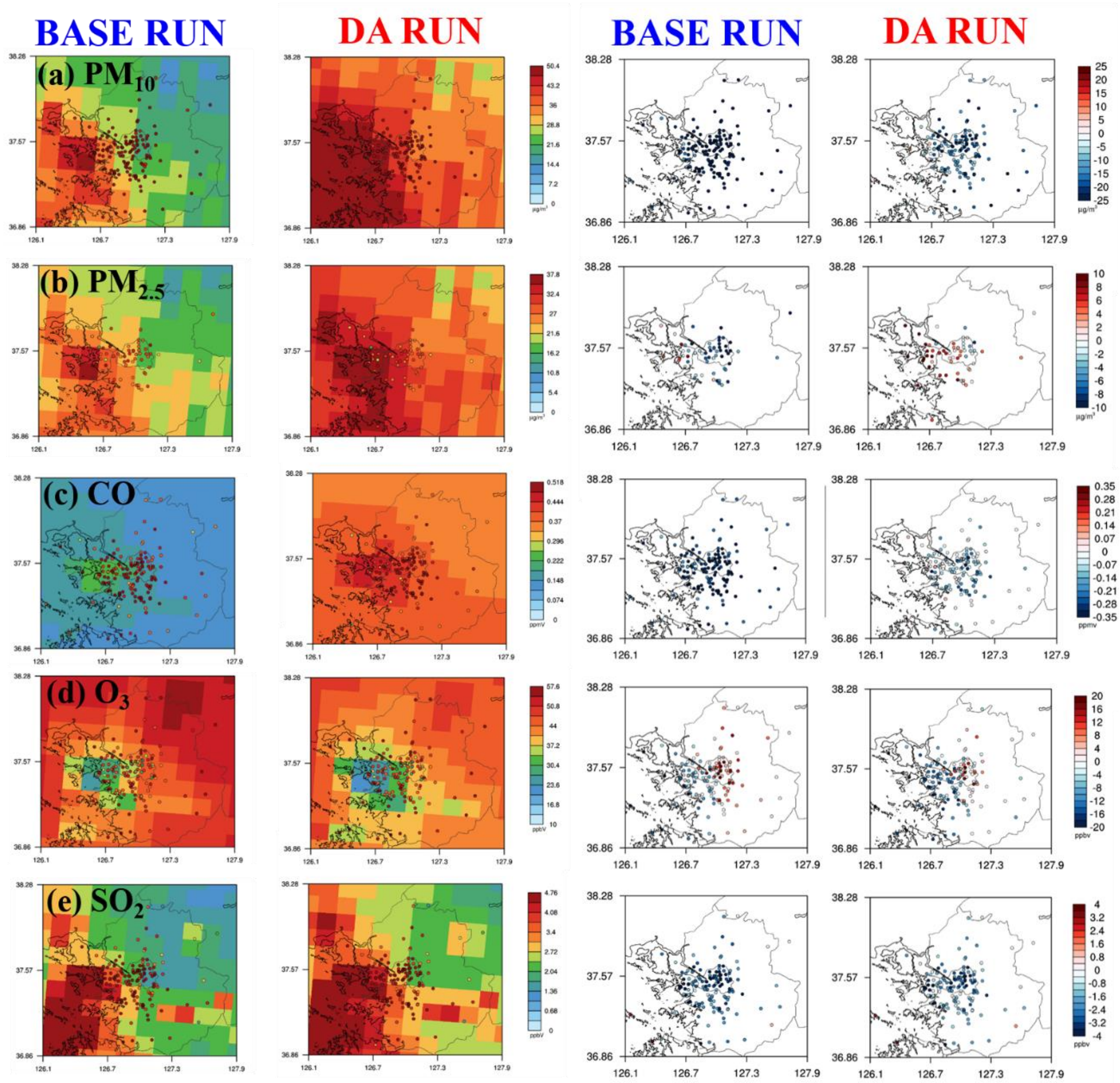

Figure 10. Spatial distributions (first and second columns) and bias (third and fourth columns) of (a) PM_{10}, (b) $PM_{2.5}$, (c) CO, (d) SO_2, and (e) O_3 over Seoul Metropolitan Area (SMA) for the entire period of the KORUS-AQ campaign. Colored circles of first and second columns represent the concentrations of the air pollutants observed at the Air Korea stations in the SMA.

Comment: S13. Lines 428 – 429: This is confusing, as it appears you are talking about an additional model simulation to the DA run. I thought that the adjusted NOx observations are used in the DA run, as discussed for the results in Lines 363-373 and comparing results from Figures 7a-b.

Response: As addressed in Lines 428-429 of the original manuscript, Air Korea observations for NO_2 were not applied in a 100 % accurate way in the current version of Korean air quality prediction system. It is because NO_2 mixing ratios measured at the Air Korea sites are contaminated by other nitrogen gases due to "NO_2 measurement artifacts" as discussed in Lines 374-387 (original manuscript). In Fig. 7(a), the results of a DA RUN by assimilating CMAQ outputs with Air Korea-observed O_3 mixing ratios are shown. Fig. 7(b) depicts the results of the test run by assimilating CMAQ outputs with Air Korea-observed both O_3 and NO_2 mixing ratios. But, here we used "$0.8\times NO_2$ mixing ratios", following information given by Jung et al. (2017). We believe this work is not 100 % perfect! That's why we call the DA RUN a preliminary DA RUN. In the future, we attempt to correct these artifacts of NO_2 mixing ratios. Then, we will revisit the impacts of the NO_2 assimilation on O_3 mixing ratios. To avoid confusion regarding this issue, the corresponding sentence was modified in the revised manuscript (please, see pp. 18:394 – 396).

Comment: S14. Lines 525 – 529: Is the AI referring to its application to rapid refresh of emissions, chemical reaction/mechanism replacements, or something else? Is there a large body of research that shows AI can even "improve" air quality forecasts? I think the body of work shows that AI can be used speed up the gas chemistry in regional CTMs, while not suffering model performance degradation. Also, if the provided citation to Kim et al. (2019) is of no help, because it is not included in the reference list (see T6 correction below).

Response: Kim et al. (2019) recently published a paper in ACP that employed a deep recurrent neural network system based on long short-term memory (LSTM) model for daily PM_{10} and $PM_{2.5}$ predictions in South Korea. The prediction system was optimized by iterative model trainings with the inputs of ground-based observations for PM_{10}, $PM_{2.5}$, and the observed meteorological variables including wind speed, wind direction, relative humidity, precipitation, etc. Their AI-based prediction system showed better performances than the CMAQ model simulations. However, the current AI system works only for points where ground-based observations are made. Therefore, we expect that a combination of the AI system with the currently developed air quality prediction system can produce a more accurate air quality forecast over South Korea. Regarding this issue, please refer to pp. 25:550 – 559.

Technical corrections and response:

Comment: T1. Line 120: Tang et al. (2017) is not found in the reference list.

Response: "Tang et al., 2017" has been added into the references. Please, check this out at pp. 34:830 – 834.

Comment: T2. Line 145: Replace "dynamic" with "dynamical".

Response: "dynamic" has been replaced by "dynamical". Please, check this out at pp. 7:149.

Comment: T3. Line 147: Replace "National Centers for Environmental Prediction Final Analysis data (NCEP FNL)" with "National Centers for Environmental Prediction (NCEP) Final (FNL) Operational Global Analysis data on 1° × 1° grids". Is 1° × 1° correct?

Response: Yes, 1° × 1° is correct. "National Centers for Environmental Prediction Final Analysis data (NCEP FNL)" has been replaced with "National Centers for Environmental Prediction (NCEP) Final (FNL) Operational Global Analysis data on 1° x 1° grids". Please, see this out pp. 7:157 – pp. 8:159.

Comment: T4. Line 413. Replace "matters" with "matter"

Response: "matters" has been deleted in the manuscript, following the referee's advice (S11) . Thank you!

Comment: T5. Line 501: Replace “ground” with “near-surface”

Response: “ground” has been replaced with “near-surface”. Please, check this out at pp. 24:526 – 527.

Comment: T6. Line 527: Kim et al. (2019) is not found in the reference list.

Response: “Kim et al. (2019)” has been added to the references. Please, check this out at pp. 32:752 – 755.

References

Kang, M., Lim, Y.-K., Cho, C., Kim, K. R., Park, J. S. and Kim, B.-J.: Accuracy Assessment of Planetary Boundary Layer Height for the WRF Model Using Temporal High Resolution Radio-sonde Observations, , 26, doi:10.14191/ATMOS.2016.26.4.673, 2016.

Kim, H. C., Kim, S., Kim, B.-U., Jin, C.-S., Hong, S., Park, R., Son, S.-W., Bae, C., Bae, M., Song, C.-K. and Stein, A.: Recent increase of surface particulate matter concentrations in the Seoul Metropolitan Area, Korea, Scientific Reports, 7(1), 4710, doi:10.1038/s41598-017-05092-8, 2017.

Kim, H. S., Park, I., Song, C. H., Lee, K., Yun, J. W., Kim, H. K., Jeon, M., Lee, J. and Han, K. M.: Development of a daily PM10 and PM2.5 prediction system using a deep long short-term memory neural network model, Atmos. Chem. Phys., 19(20), 12935–12951, doi:10.5194/acp-19-12935-2019, 2019.

Kim, H.-K., Woo, J.-H., Park, R. S., Song, C. H., Kim, J.-H., Ban, S.-J. and Park, J.-H.: Impacts of different plant functional types on ambient ozone predictions in the Seoul Metropolitan Areas (SMAs), Korea, Atmospheric Chemistry and Physics, 14(14), 7461–7484, doi:10.5194/acp-14-7461-2014, 2014.

Lee, S., Song, C. H., Park, R. S., Park, M. E., Han, K. M., Kim, J., Choi, M., Ghim, Y. S. and Woo, J.-H.: GIST-PM-Asia v1: development of a numerical system to improve particulate matter forecasts in South Korea using geostationary satellite-retrieved aerosol optical data over Northeast Asia, Geosci. Model Dev., 9(1), 17–39, doi:10.5194/gmd-9-17-2016, 2016.

Nam, H.-G., Choi, W., Kim, Y.-J., Shim, J.-K., Cho, B.-C. and Kim, B.-G.: Estimate and Analysis of Planetary Boundary Layer Height (PBLH) using a Mobile Lidar Vehicle system, , 32, doi:10.7780/KJRS.2016.32.3.9, 2016.

Park, M. E., Song, C. H., Park, R. S., Lee, J., Kim, J., Lee, S., Woo, J.-H., Carmichael, G. R., Eck, T. F., Holben, B. N., Lee, S.-S., Song, C. K. and Hong, Y. D.: New approach to monitor transboundary particulate pollution over Northeast Asia, Atmos. Chem. Phys., 14(2), 659–674, doi:10.5194/acp-14-659-2014, 2014.

Shrestha, K. L., Kondo, A., Kaga, A. and Inoue, Y.: High-resolution modeling and evaluation of ozone air quality of Osaka using MM5-CMAQ system, Journal of Environmental Sciences, 21(6), 782–789, doi:https://doi.org/10.1016/S1001-0742(08)62341-4, 2009.

Sirithian, D. and Thepanondh, S.: Influence of Grid Resolution in Modeling of Air Pollution from Open Burning, Atmosphere, 7(7), doi:10.3390/atmos7070093, 2016.

Souri, A. H., Choi, Y., Jeon, W., Woo, J.-H., Zhang, Q. and Kurokawa, J.: Remote sensing evidence of decadal changes in major tropospheric ozone precursors over East Asia, Journal of Geophysical Research: Atmospheres, 122(4), 2474–2492, doi:10.1002/2016JD025663, 2017.

Tang, Y., Pagowski, M., Chai, T., Pan, L., Lee, P., Baker, B., Kumar, R., Delle Monache, L., Tong, D. and Kim, H.-C.: A case study of aerosol data assimilation with the Community Multi-scale Air Quality Model over the contiguous United States using 3D-Var and optimal interpolation methods, Geosci. Model Dev., 10(12), 4743–4758, doi:10.5194/gmd-10-4743-2017, 2017.