Dear anonymous referee 1,

Thank you for your time and effort trying to improve our manuscript.
In response to your review, we updated Figs. 3, 7 and 8 along with new captions and rewrote the caption for Fig. 11. Please find our point-by-point response in red below.
(Our original response in the first round of review remains with "=>". The reviewer's new comments come with "R:" while our new response starts with "A2:".)

Response to anonymous referee 1:

I have used the first review of the present original version of the manuscript as a ground here in my second review of the current study. Additional comments by me are denoted with Italic font style. For some of the issues below I have included the answers from the authors to my comments in the first review, which are marked with blue colors and an arrow at the begging of the text. The comments by me and answers by the authors below are pointing to updated page and line numbers that correspond to the revised version of the manuscript. I have only included issues from my first review of this manuscript that are still relevant to include here.

Major comment

2. There are issues with the language, which need to be improved. In the section Technical corrections below suggestions are given in an attempt to improve the language and clearness of the manuscript. However, my review and corrections of the manuscript concerning the language has only been carried out for the abstract and Introduction to show that the clearness of the text need to be improved. Therefore, I recommend that the full text needs an English proof-check.
=> This manuscript has been internally reviewed twice in our lab and proofread by another native English speaker. As we replied to the reviewer's technical corrections at the very bottom, most of the corrections the reviewer suggested are incorrect in English or distorted our points, if not irrelevant, in the manuscript. The reviewer raised most of the comments or questions regarding MODIS retrievals, which is not the focus of our study but included only for completeness, we thus wanted to stay focused on our goal and highlights of our 10 work. However, we appreciate different views and tried our best to accommodate the reviewer's comments and reflect them in our manuscript unless we have a specific reason not to.

R: It is not ok by the authors to ignore many of my suggestions concerning the language, thus, here they have not replied at all (see technical corrections below). My suggestion of an English proof-check remain. However, when now have reading more of particular Sections 4 and 5 I do not think it is much work needed to improve the text in the Sections 3, 4 and 5.

A2: Please note that, in respect of reviewer's comment, we'd already gone through another round of proof-reading and made necessary corrections throughout the manuscript. But because the reviewer's comments have numerous error in grammar, we do not think they can improve our manuscript. For instance, your last comment was *"conclusions are made in Section 5" is not grammatical correct*.
There, your statement itself is wrong in grammar because you should use "grammatically", not "grammatical". In fact, our sentence "conclusions are made $\sim$" is correct and commonly used in the literature. As we responded in the first round of review, we chose not to take your suggestions when we found your suggestions problematic - either trivial or inappropriate, if not wrong - or when we disagreed with you. But we never ignored any of your comments.

Specific comments

Page 5 Line 17: MODIS is not a satellite and which version is used here, 6.1?
=> We modified "MODIS and GOCI satellites" to "MODIS and GOCI sensors". And yes, version 6.1 was used.

R: Since I do not find it in the revised manuscript, please include text that describe it is Version 6.1 that is used in the present study.
A2: We chose not to specify the versions of retrieval algorithms for both MODIS and GOCI sensors here because it has nothing to do with our observation operators described in the statement.

Line 22. It is not correct to write that AOD measures something. This sentence need to be re-written.

R: By writing “AOD measures the amount. . . ” then the latter word is a verb and it somewhat odd that a parameter measuring something (an active action), although it might be correct in the English language. By using “AOD is a measure of the amount. . . ” Instead then the latter word is a substantive, which is a more suitable phrase. Itt is the same as writing “AOD is an estimate of the amount. . . ..”.
A2: Disagree. We do not see any difference between “AOD measures the amount. . . ” and “AOD is a measure of the amount. . . ”. No changes.

Page 7 Lines 19-21. “Following Remer et al. (2005), observation errors are specified as the retrieval errors: (0.03 + 0.05 * AOD) over ocean and (0.05 + 0.15 * AOD) over land. They do not include the representativeness error and are slightly smaller than those for GOCI AOD, as described below.”

R: The former part of this sentence sounds better now. However, it is very strange that the authors refer to retrieval errors corresponding to the ocean retrievals, estimated by the MODIS aerosol team (including Lorraine Remer), that are not valid anymore. Please update the retrieval error and corresponding references according to my comment in the first review of the original version of the manuscript.

A2: Again, MODIS data was included only for the completeness and we never intended to make the use of MODIS retrievals optimal in this study. For its error specification, we followed previous data assimilation studies like Liu et al. (2011) where it is also referred to Remer et al. (2005). No changes.

Page 9 Lines 7 and 8. The word “validation” can be used when comparing satellite derived AOD against ground-based sunphotometer measurements. However, you cannot validate AOD obtained from passive remote sensing against AOD derived from observations with another satellite sensor used in passive remote sensing.
=> We do not want to argue about how others described their work. The term of “validation” was used in Choi et al. (2018) and we just adopted it here. Also, in a broad sense, the terminology of “validation” is commonly used when one data is evaluated against another independent observation in the data assimilation community, so we do not see it problematic. No changes.
R: You don’t need to argue here and instead of just adopt this term and rely on what is commonly used it is better to go to a dictionary. The word validation is a too strong word to describe an inter-comparison of AOD obtained with passive remote sensing from two different satellite platforms (see also comment below corresponding to Figure 2). The word “evaluation”, used in the text above by the authors or “comparison” are options that are more suitable.
A2: Disagree. In many data assimilation studies, "validation" has long been used interchangeably with "evaluation". No changes.

Lines 18-21. Concerning the sentence “When these different observation errors were applied to GOCI retrievals in the assimilation, the smallest error (ϵ_2) produced slightly better fits to observations specially for the high values (AOD > 2). . . . .” This statement seems not hold, since ϵ_2 is not better than ϵ_1 over land for the situations with lower AOD.
=> We do not understand why our statement doesn’t hold due to the case of lower AOD, which we did not even discuss here. We mentioned that the smallest error (ϵ_2) produced slightly better fits to observations for the high (!) AOD values. We also stated that such a result is not statistically significantly different, so we do not understand why the reviewer is arguing over the statement. No changes.
R: Remove then “specially” from the original sentence in the manuscript, then your argumentation and statement hold.
A2: The smaller the observation error gets, the better the model states fit to the observation. We described the trend here and tried to pull our model values closer to high AOD values, in particular. We believe our statement holds, even with "specially". No changes.

Lines 29-31. Suggestion, change the sentence "This is partly because AOD is not directly associated with surface PM2.5 . . . . " to "This is partly because AOD is not directly related to surface PM2.5 . . . . 2, since "related" is a more correct word to use.

A2: Disagree. The expression of "associated with" is not really different from "related to". No changes.

Section 5 and page 14

Line 26 and first sentence in this section. I think it is too strong positive words are used here when describing the GOCI AOD retrievals. I am sorry about this comment since I refer to the sentence on page 15 instead of the correct page 14: "GOCI AOD retrievals provide reliable and consistent aerosol information, monitoring air pollutants flowing over to the Korean peninsula at high resolution every day." Thus, I think it is too strong positive words used in the first part of the sentence when describing the GOCI AOD satellite retrievals on the accuracy of air quality forecasting This is what the authors write in the abstract "During the Korea-United States Air Quality (KORUS-AQ) period (May 2016), the impact of GOCI AOD on the accuracy of air quality forecasting is examined by comparing with other observations including Moderate Resolution Imaging Spectroradiometer (MODIS) sensors and fine 10 particulate matter (PM2:5) observations at the surface." Concerning the latter parameter the result presented in Figure 5 is not impressive. In addition, you cannot rely the statement on an inter-comparison with AOD derived from observations carried out from another satellite platform. The first and second parts of the sentence is not synchronized and this is an odd phrase "air pollution flowing over. . . ." Therefore, here is a suggestion: "GOCI provide daily AODs with high spatial resolution and frequently detect air pollutants over to the Korean peninsula."

A2: Thanks for your suggestion. We agree that "air pollution flowing over to..." is a bit awkward. We now take out "flowing" and change the part as "air pollutants over the Korean peninsula..."

Figures

Figure 2 It is not correct to write that AOD is retrieved at this time, since it is the observations that is carried out at this time and it is a very long way to come up with an estimate of AOD, for example you have to introduce a model that describe radiation transfer in the atmosphere. Change "retrieved" to "corresponding" in the first sentence of the figure caption to Figure 2. Describe in the figure caption the solid black box introduced.

=> This study is not meant for describing the retrieval process, but how the data is used in the assimilation cycle. The data was processed at the time and that's how they are described and presented in Figure 2. Even in-situ measurements such as radiosonde do not report the values at the exact time (depending on the vertical levels as it goes up), but in the data assimilation context, that's how they are all described. Please note that we already illustrated the temporal distribution of the data in the last paragraph of page 7 ("In terms of temporal distribution,). In response to your last comment, though, we added one paragraph "Domain 2 is marked as a black box in each panel." at the end of the caption.

R: It has nothing to do with reporting the values at the exact time, instead you have to separate "retrieved" and "observed", thus, the TOA radiance was measured/observed at the current time. It is just simply to write it more correctly so you don't hide that AOD derived from satellite measurements is not a direct observation, instead it need to be related to measured radiance scattered only by aerosols. The latter means that the contributions of Rayleigh scattering and surface reflection have to be reduced for. In addition, to relate AOD with TOA radiance associated with purely aerosols you have to introduce radiative transfer calculations. I present here a new suggestion: "Field of GOCI AOD at 550 nm derived/retrieved from observations carried out at 2016-05-01_06:00:00 UTC. . . . . "

A2: Disagree. We already mentioned GOCI data as retrievals. And we've never seen any literature on data assimilation that described the retrieval data in the way you suggested (e.g. "Field of GOCI AOD at 550 nm derived/retrieved from observations carried out at 2016-05-01_06:00:00"). No changes.

Figure 3. You have to include information that the right y-axis goes down to 2000 (the latter value I picked up in the file with the author's response in the review process), since it is otherwise hard for the readers to understand part of the figure. You need to use different symbols for each cycle or at least each pair of cycles. The half-last part of the last sentence in the figure caption need to be re-written.

A2: Figure 3 is now changed with new caption, as suggested.

Figure 4. In the figure caption you have to refer to the body text about the three different types of observation errors. Take the color blind persons in consideration and use the three colors in combination with solid, dashed and dotted lines and separate land and ocean with heavy and normal lines, respectively.

=> This draft uses a lot of colors throughout the figures, and is not meant for color-blinded readers, unfortunately. But based on your comment, we added "The first two errors (ϵ_1 and ϵ_2)) are described in equations (3) - (6) and the third error (ϵ_3) increases by 20% everywhere." in the caption.

R: I agree on "a lot of colors", but why not make it easier, particularly for the colorblind person, by improving those figures for which it possible to do that for? You do not use colors in Figure 3 and it seems to work relatively well (see comments concerning Figure 3 above).

A2: The reason Figure 3 could go with black and gray is because it only needed two different colors. Now we have total of six cases to present here, and they are very close to each other, especially near the low x values. We believe colors can show these errors more clearly. No changes.

Figure 7. It is a lot of space in the figure and therefore write the names of the species in all figures.

=> We decided to put the species name in the main title because the first panel does not have room for it due to the legend. As this figure has to take up the whole page (height-wise) anyway, we decided to keep the main title. No changes.

R: I guess the journal will not accept the presentation of these figure and then it could be worthwhile to wait with the changes. You could in any case already now try to improve the presentation of the figures by using the free spaces on the right and left sides of the figures. Thus, make three rows, 3, 3 and 4, instead and remove "Model levels" and values on the y-axis corresponding to the figures presented at the middle and right positions. This means that you can move the figures closer to each other. The results of it will be that you get somewhat larger figures, thus, it will be possible to remove the titles and put all species names in the figures. In addition, you have to explain the species name in the figure caption or at least write in the figure caption that this information can be found in Section 4.2.

A2: Figure 7 is now replotted. Thanks for your suggestion.

Figure 8. Write "Model levels" connected to the y-axis. => The caption already stated that it 5 is the same as Figure 7. We tried to reserve more x-axis space to zoom in differences between the experiments here, dropping y-axis title intentionally. No changes made.

R: It is not acceptable to refer to the previous figure concerning text belong to the y-axis. You will neither reduce the size of the figures by including this text. Thus, you have free space available on the left side of the figures and you need only to include "Model levels" (once) at this place. You could also exclude the values on the y-axis corresponding to the right figure.

A2: The y-axis label "Model levels" is now added.

Figure 11. Suggestion "Figure 11. Same as Fig. 10, while here the results of forecast accuracy (%) for categorical forecasts are presented, subdivided according to classification of air quality in Tables 2 and 3." Include "Model level" on the y-axis.

=> Figure 11 shows different statistics, not the model level, as shown in the main title. No changes made.

R: The authors are of course right in the comment above, but the figure caption is once again not clear. For example you have to write "with the exception for" instead of just "except" and "Tables 2 and 3" instead of "Table 2 and 3". However, even with these changes suggested the sentence is not clear and complete.

A2: We now changed the entire caption as "Time series of forecast accuracy (%) of the hourly forecasts from the 00 Z initialization for May 4 - 31, 2016 in domain 2 for categorized events based on hourly surface $PM_{2.5}$ concentrations, as defined in Tables 2 and 3."

Technical corrections

It seems that in the creation of the pdf-file, when my review was uploaded into the reviewing system, the space between two comments disappeared. It is pity that I did not notice that. Even so, all the comments are clearly separated in the way that every new comment start with a new line, for example “Line 2” and “Line 3” corresponding to the abstract. My comments are either incomplete. With an example I explain here how to handle the comments. This text “..……. underestimates predicted surface PM2..5……… ” on Line 11 in the abstract means that it is only these words in the sentence that the authors need to put attention on. The following comments presented below remain for the authors to take in consideration, and for some of these I have included a motivation.

Abstract

Line 2. Suggestion “The Korean Geostationary Ocean Color Imager (GOCI) provides, based on daily high temporal and spatial resolution data, unprecedented information on air pollutants over the upstream region of the Korean peninsula for the last decade.” Note that GOCI is not a satellite and the phrase “…..has monitored the East Asian region in high temporal and spatial resolution every day for the last decade…… ” need to be rewritten.
A2: Reviewer seems to be obsessed with terminology. As we already described it as a sensor, we do not make any changes here.

Line 3. “the GOCI aerosol optical depth (AOD)” instead of “the GOCI Aerosol optical depth (AOD).”
A2: Changed, as suggested. Thank you.

Line 6. “…… assimilated with three-dimensional…..technique in the Weather…..”
A2: One can use "in" rather than "with". No changes.

Line 9. “Sensors” are not observations and therefore change to “….(MODIS) AOD……”, but this is a better suggestion “…… (MODIS) AOD and in-situ ground-based fine particle matter (PM2,5).”, since then you don’t need to include “observations” again in the sentence.
A2: We don’t see any differences. No changes.

Line 11.”……. underestimates predicted surface PM2..5……… ” You have to be more clear what you mean with the second time you write “surface PM2..5” in the sentence. Is it “predicted” or “estimated” or something else?
A2: We believe that it is obvious to be the forecast value because we clearly mentioned that the impact lasts for about 6 h at the end of sentence.

Line 13. The last part beginning with “with the most… ” of this sentence is not clear.
=>We clearly demonstrated the most significant contributions to the prediction of heavy pollution events in Figure 11 where the assimilation of GOCI data produced the biggest improvement in b) high pollution accuracy.
R: Make then a new sentence of it “This resulted in the most significant contributions to the prediction of heavy pollution events over South Korea.”
A2: This is basically the same as what we described. No changes.

Introduction
Page 1
Line 21. “Surface concentrations” of what ?
=> of chemical species. We believe this should be clear as the previous paragraph is immediately followed by this one.
R: You have to make the current sentence clear and not rely on the information in the previous paragraph, particularly not in this case when you also describe aerosols in the previous paragraph.
A2: Disagree. We believe it is clear enough for readers. No changes.

Line 23. “The latter is highly dependent on…… ” The word “relies” don’t suit here.
A2: Disagree. "depend on" is not much different from "rely on". No changes.

Page 2

Line 3. "....performs chemical simulations according to 3-km horizontal resolution at present day." Please accept this change or make the original text more clear.
A2: We don't see anything unclear. No changes.

Line 4. What is meant by "these fast-varying complex mechanisms" or what is it pointed to?
=> All the mechanisms described in the previous paragraph, particularly the aerosol-meteorology interaction at short time scales. Again, this sentence is also connected with the paragraph right ahead.
R: To improve the text the best solution is to include the information you touch on in the answer above. Another but not the best solution could be to write it like this : "For such a high-resolution application and for situations with very high aerosol concentrations, the fast-varying complex mechanisms described above might be better represented through online coupling between chemical and meteorological components."
A2: So you suggested that we add "described above" in the middle. The previous paragraph that described the fast-varying complex mechanisms is directly followed by this paragraph. Thus, we see "described above" as redundant. No changes.

The authors have not given any respond to the comments below and again this is not ok.
A2: We did not intend to ignore any of them, but as we made it clear in the first round, we found most of your suggestions below either cosmetic or irrelevant, if not wrong. You basically asked us to rewrite the entire page for pages 1 - 3 for something that does not matter to our main points or the quality of this work. But we reflected a few of them in our manuscript for clarity. Thank you for your time and suggestions.

Lines 9-10. This sentence need to be re-written.
Suggestion: "Chemical modeling are associated with large uncertainties, particularly concerning emission data and to simulate meteorology process. One of the most effective ways of utilizing aerosols is instead to assimilate aerosol observations into the forecast model and improve the initialization of aerosol simulations."
A2: Disagree. No changes.

Line 12. Not clear written: "(usually in the optical properties)"
Suggestion: "(usually the optical properties)"
A2: Incorrect. No changes.

Line 12. Change "observed information" to "information" or "results"
A2: There are no such thing as "observed results". No changes.

Line 15. Change "for" to "of"
A2: It is correct to say "prediction for precipitation,...". No changes.

Line 16. "....conducted in Korea between 1 and 12 June 2016..............."
A2: We already mentioned the time period at the end of the sentence. No changes.

Line 17. Remove "a field campaign"
The original sentence "An international cooperative air quality field study conducted in Korea between 1 and 12 June 2016, named as the Korea-United States Air Quality (KORUS-AQ), was a field campaign jointly developed by air quality researchers in the United States and South Korea to improve our understanding of major contributors to poor air quality in Korea for May 1-June 12, 2016." need to be improved. Note that this sentence include the word "field" twice and this is redundant. I think also that it is possible and more suitable to make the sentence shorter.
Suggestion: The Korea-United States Air Quality (KORUS-AQ) field campaign conducted in Korea between 1 and 12 June 2016 was developed by researchers in the United States and South Korea to improve our understanding of the major contributors to the poor air quality in Korea."

A2: Your suggestion doesn't look any better to us. No changes.

Line 21. "....occurred due to long-range......"
A2: "...occurred by" is not incorrect. Keep it.

Lines 25-27. Based on the Korean Geostationary Ocean Color Imager (GOCI) onboard the Communication, Ocean, and Meteorology Satellite (COMS) retrievals of hourly AOD scenes, for multiple spectral bands, are centred with respect to the Korean peninsula during daytime (Kim et al., 2017). AOD scenes with high spatial and temporal resolutions are available since 2010." The sentence in the original manuscript lose connection at "....spectral bands monitoring....". This is a new suggestion to uses two sentences instead of the original one: "The Korean Geostationary Ocean Color Imager (GOCI) onboard the Communication, Ocean, and Meteorology Satellite (COMS) provides hourly AOD retrievals at multiple spectral bands. GOCI monitoring the East Asian region centered on the Korean peninsula during daytime (Kim et al., 2017). AOD scenes with high spatial and temporal resolutions are available since 2010."
A2: It is just another way of saying. No differences. No changes.

Line 28. "It has been demonstrated"
The original sentence need to be re-written. Here is a new suggestion: "It has been demonstrated that GOCI data are associated with high accuracy compared to the low-orbiting Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite 30 (VIIRS) products (Lee et al. (2010); Wang et al. (2013); Xiao et al. (2016); Choi et al. (2018)"
A2: Again, no differences. No changes.

Line 31. "assimilating AOD derived from MODIS observations (Remer et al., 2005)" Should it be like this: "Liu et al. (2011) were the first to implement assimilation of Aerosol Optical Depth (AOD) retrieved from the MODIS sensors (Remer et al., 2005) into the National Centers for Environmental Prediction (NCEP) Gridpoint Statistical Interpolation (GSI; Wu et al. (2002); Kleist et al. (2009)) system."
A2: Disagree. No changes.

Line 34. "forecasts of a dust storm"
A2: It is correct to state "forecasts in a dust storm event". No changes.

Line 35. "..widely used for air quality forecasting. The system has been extended for....."
Please make two sentences of this relatively long original sentence.
A2: It is only three-line long. No changes.

Page 3
Line 5. "MOdel....."
A2: Changed, as suggested.

Line 8. "....assimilation of AOD improved..."
A2: AOD means AOD retrievals. Why should we take out "retrievals"? No changes.

Line 9. "In the present study, the assimilation............system has been extended to be better used in the GOCI AOD retrievals during the current investigation period..."
A2: Nothing is wrong with our original expression. No changes.

Line 10. Not clear what is meant by this "careful investigation of data characteristics."
R: The change suggested here by me don't hold and I am sorry for that. However, this sentence on line 9, page 3, in the original version of the manuscript need in any case to be improved. Here is a new suggestion: "In the present study, the assimilation

capabilities in the GSI 3DVAR system has been further extended to optimize the use of GOCI AOD retrievals during the KORUS-AQ period. (with careful investigation of data characteristics) I suggest to excluding the last part of this sentence, in the brackets, or improve it.
A2: Disagree. Keep our original statement.

Lines 10-13. The last part of this sentence is not clear: "...... compared to that of other observations."
I suggest creating two sentences of this long original sentence.
A2: Disagree. Keep it.

Line 13. "data and examine"
Sorry, I referred to the wrong line here, thus, it should be Line 12 instead. Here you should remove the redundant word "then" in the original manuscript.
A2: Removed, as suggested.

Line 18. "conclusions are presented"
This is one of several examples showing that the manuscript needs an English proof-check. Thus, the phrase "conclusions are made in Section 5" is not grammatical correct.
A2: Disagree. No changes.

Dear anonymous referee 2,

Thank you for your meticulous review of this draft and great suggestions.

We truly appreciate your help on finding a bug in Fig. 7, which is now replotted with new caption. We also found most of your comments very helpful in clarifications, improving our draft significantly.

Our line-by-line response to your comments can be found in "A2" below. (Our response in the first round remains as "A:".)

Response to anonymous referee 2:

My major comment remains being the evaluation of the modeled/assimilated vertical profiles using data available from the KORUS-AQ campaign. They can address this in different ways. There was an HSRL in Seoul (see sample data in Peterson et al., 2019) that can provide full time series of vertically resolved aerosol extinction that can be compared to the simulations. Also, there was an HSRL in the DC8 aircraft, so the comparison can be done for some days across the Korean peninsula. Also, data from the AMS onboard of the DC-8 can be used to compute PM1 and compare to model estimates, there were 2-3 full vertical profiles over Seoul every day the aircraft flew. I'm not asking the authors to do all of these (although this would be nice), but at least show some effort of trying to assess the skill of the model in representing vertical profiles and if the changes in vertical distribution generated by the assimilation are somewhat reflected in better agreement to the observations. If discrepancies arise this is still useful as one can blame model uncertainties (e.g., computation of optical properties) for them. Louisa Emmons (NCAR ACOM, same institution as authors) is a modeler that was heavily involved in the KORUS-AQ campaign, perhaps consulting with her on this topic might be a good idea.

A2: Thank you for your suggestion. We may need to start with some clarification.

In the first round of your review, you misunderstood our main results as the assimilation of GOCI retrievals had degraded, rather than improved, the forecasts in surface $PM_{2.5}$ concentration when not assimilated with surface $PM_{2.5}$ observations. You made the point as your first and the most important finding, which was completely contradictory to our major conclusion on how useful the assimilation of GOCI retrievals was, particularly for high pollution events.

Thus, in response to your review in the first round, we added three more figures to prove that the GOCI retrievals actually **improved** the forecasts with respect to **independent** $PM_{2.5}$ and AOD observations on the ground, as you suggested.

But, to our surprise, you did not comment on this at all in the 2nd round of review whether our revised manuscript with the objective validation corrected your misunderstanding. As your major comment was all related to that and it is the most critical finding of this study, it is mandatory to agree to each other on this particular point, we believe.

However, skipping the main point, you continue to ask for the model verification on the vertical profiles, now in order to address the issue of model uncertainties, which is off the topic.

When you had thought that our assimilation of AOD degraded forecasts, it was legitimate to request for examining why and how the assimilation of AOD could result in poor forecasts. Along the line, it made sense for you to infer that the vertical profiles in the model could have deviated from the observed ones, thus to ask for evaluating the profiles against other observations. In return, we provided three new figures to support our conclusions and proved that you had misunderstood our results.

Now, if you corrected your misinterpretation of our results, you need to come with another rationale behind your request for the model evaluation.

At this point, we want to remind you that the main goal of this study is to demonstrate the benefit of the GOCI assimilation on the prediction of **surface** $PM_{2.5}$ concentration. And it is our belief that we had provided sufficient evidence for the value of the GOCI assimilation, including independent verification, as you suggested.

Of course, the model uncertainty could play a nontrivial role on the analysis quality, but there are many other factors in the analysis system that can contribute to making the analysis successful to the same extent, such as the estimation of background error covariance, observation error statistics, and the error in the observation operators. In the limitation of the simple GOCART aerosol scheme and the poor quality of emission data, we do not think that accurate representation of vertical profiles in the model is a primary concern of this study on the air pollution at the surface.

We thus disagree with you that we should include the evaluation of the model profiles using additional data from a field campaign. Please note that we already have a total of 16 figures and 4 tables to demonstrate the benefit of the GOCI assimilation.

Comments from initial review

R1: Initial reviewer comments, A: Author response, R2: New reviewer comment

R1: Figure 1. Why show observations for a given time? Why not show maybe an average of the period analized?
A: => Figure 1 simply shows the model domain with the observing network. No changes are made.
R2: But you can also use it to show average concentrations over the period analyzed instead of showing a random time. Also, it would be nice to get a second panel with the 2nd domain to see details on the observation distribution over the Korean peninsula
A2: => Figure 1 is now changed, as suggested.

R1: Figures 8 nd 7. You could model vertical distribution and impact after assimilation using airborne data and surface lidars deployed as part of KORUS-AQ
A: => Not clear on your point here. Figures 7 and 8 show how the model responded to the assimilation of observations in use. This analysis is needed to understand how our assimilation worked in the model space. It has nothing to do with verification.
R2: I apologize for the typo, I meant “You could evaluate model vertical distribution ...” This is related to one of my main comments on using KORUS-AQ observations for evaluating vertical distributions, that the authors have not addressed.

A2: => Now we understood your point. However, we disagree that it is essential to evaluate the vertical distribution of the model variables over South Korea (or in the surrounding region) with two major reasons:
i) The positive impact of the GOCI assimilation is mainly attributed to a wide **horizontal** coverage of the **upstream** region (not the vertical penetration through the troposphere). This point is well illustrated in our figure 9 and summarized in the last section. Unfortunately, this cannot be verified against any other instruments located in Seoul or even with the DC8 aircraft that flew around South Korea (based on the flight tracks depicted in Fig. 1 in Peterson et al. (2019)). As marked in dark red in the GOCI panel in Fig. 9, the biggest impact of the assimilation was found around the Jilin Province in China, northwest of North Korea, where we do not have any measurement even during the KORUS-AQ campaign. That’s why the GOCI AOD retrievals are so valuable, as we stressed throughout the manuscript.
ii) More importantly, the assimilation of AOD is not predominantly dependent on the model performance or behavior on the vertical structure of aerosol species at certain times. A successful data assimilation requires a tremendous effort on understanding and harmonizing various components with disparate characteristics, and the model error is just one of them. Even if we can evaluate the vertical profiles of a few model aerosol variables over a tiny portion of our model domain 1, it wouldn’t give you a synthetic picture of how the vertical structure of the model affected the analysis quality (both quantitatively and qualitatively) because it is not possible to disentangle it from all the complicated behaviors in the model and the analysis systems in a statistically meaningful way.
This is simply not how you can make a real data assimilation system work. Therefore, we chose not to take the reviewer’s suggestion.

Comments by line on revised manuscript

2 1. Provide references for this statement, latest IPCC report should do
A2: Thanks for your suggestion. We checked the latest IPCC report "IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.",
but could not find anything associated with aerosol-meteorology interaction. It reported a great deal of emission levels or effect on the global scale, but not on the local meteorology, which is not directly related to our study. So we chose to not include it.

3 3-8. You can add Park et al (2014) to this paragraph

A2: Again, thanks for the reference. We also recognized several other papers with the assimilation of GOCI using the CMAQ model. But one of the main points of this study is to examine the impact of GOCI assimilation on the online coupled forecasting system, as described in the paragraph. So it is not appropriate to refer to the study using the *offline* coupled system here. Not included.

4 5. You can cite LeGrand e t al. (2019) for the AFWA scheme

A2: Included. Thank you.

8 30-34. If I understood correctly, what you are trying to say here is that there are large values after thinning that were not found in the original data. Please rephrase to make this clearer and to the point.

A2: The statement is now changed in lines 31-32, as below.

"When all the GOCI data thinned in the GSI system were checked for the entire month, there were such extreme values that did not exist in the original dataset for multiple cases." =>"For the month of May 2016, multiple cases with such extreme fake values were found after the thinning process."

9 7-16. I'm assuming subscript 1 and 2 in the error (eqns 3-6) correspond to the different verifying object (AERONET or satellite-based retrievals)? Please specify which is which. AERONET is generally treated as ground truth, so that's probably the one you should be using.

A2: For the clarity, we now added "We assign ϵ_1 following their error specification with respect to AERONET and ϵ_2 based on their expected error against retrieved satellite AOD in GOCI YAER V2.".

But in the data assimilation framework, observation error can be further adjusted based on the model's representativeness. In other words, observation error in the analysis system does not need to be the same as the instrument error. But thanks for your comment.

9 27-29. This sentence is not clear, please rephrase.

A2: It is now changed from "it is not guaranteed that such an analysis better fit to AOD retrievals would actually lead to better forecasts in surface $PM_{2.5}$."

to "it is not guaranteed that the analysis in a good agreement with AOD retrievals would actually lead to better forecasts in surface $PM_{2.5}$."

9 30-31. "This is partly because AOD is not directly associated with surface PM2.5 ..." I think what you are trying to say is that AOD is a column integrated quantity while PM2.5 is measured at the surface?

Might be better to specify it that way, the way it's currently stated is vague and not necessarily true (many approaches exist to compute surface PM2.5 from AOD)

A2: We now added ", a column integrated quantity, " right after AOD.

9 30-31. "... and partly because large uncertainties in the forecast model and the emission forcing can dominate over the analysis error during the model integration" This is only applicable for forecasts, not for the analysis. Might want to split the arguments.

A2: It is applicable for the analysis as it is computed based on the forecast error as well.

But in response to your comment, the paragraph in Page 9, 30-31 is now rewritten as below.

"Even though it would be hard to quantify the model error and the emission uncertainty (and their impact on the forecast quality), it might be worth checking the correlation between GOCI AOD retrievals and surface $PM_{2.5}$ observations before evaluating the impact of GOCI AOD on surface $PM_{2.5}$ forecasts."

=> "Even if the efficiency of assimilating AOD on improving surface $PM_{2.5}$ forecasts can be largely affected by the quality of the forecast model and the emission data in use, the effectiveness of the AOD assimilation is based on the relationship between the column-integrated AOD and $PM_{2.5}$ on the ground. Therefore, it might be worth checking the correlation between GOCI

AOD retrievals and surface $PM_{2.5}$ observations for the cycling period."

9 26 – 10 2. Based on the bad correlation, you might want to mention that this is why a model is needed to “translate” AODs into PM2.5. And that this “translation” might depend on the ability of the model of properly represent the aerosols vertically, and the conversion from aerosol mass to optical properties.
A2: Good point. We now added one more sentence at the end of the paragraph.
"Such an indirect relationship between the two observations makes the analysis challenging because it can induce a large error in the observation operator and heavily depend on the model’s ability of deriving $PM_{2.5}$ from the AODs based on the vertical structure of aerosol variables and the conversion from aerosol mass to optical properties."

Figure 7. How can be sulfate in the background so low (seems to be equal to 0) and different than 0 in the assimilation experiments?
A2: Excellent point. We now fixed a bug in the plotting algorithm which did not properly convert the model variable "sulf" in ppmv to actual "sulfate" in ug/mg in the background fields. Figure 7 is replotted with the correction of the background fields in sulfate. We very much appreciate your careful review on this.

11 5-6 . “When all the observations are assimilated together (in "ALL"), it combines the effect of surface PM2.5 and GOCI retrievals, as expected”. You might want to explain that this combined effect ends up changing the vertical distribution, pulling the surface levels towards PM2.5 and the upper levels to match the AOD columns.
A2: Thanks for your suggestion. The sentence is now modified to replace ", as expected" with "changing the vertical distribution of aerosol species to match with the AOD column values and pulling the surface states towards surface $PM_{2.5}$ concentrations.".

11 32. I think it’s worth mentioning/discussing that although errors are reduced, none of the assimilation experiments are able to reduce the bias compared to the NODA experiment, which is pretty low to start with.
A2: In terms of systematic error, we focused more on the mean absolute error (mae) than the mean error (bias). In that sense, most experiments actually reduced the error from the one in NODA. As we already stated that in the same paragraph (Page 11, lines 25-26), we decided not to add any more discussion related to it.

13 21-29. The data used for evaluation in Fig 14 correspond to 3 urban sites where the 9 km res model with simple aerosol chemistry and emissions that have no hourly variation will have a very hard time representing the observed fluctuations. For instance, see Nault et al., (2018), the model configuration you are running doesn’t even consider secondary organic aerosol formation. You can pick sites that you assimilated that are close-by, plot them in the same way, and I would expect you find similar fluctuations. When you average many sites that are both urban and background, you expect some of these fluctuations to be smoothed out. My point is that I wouldn’t blame observation quality in this case.
A2: Your point is taken. Now we changed
"As raw data, these observations do not seem to be reliable and fluctuate a lot for the entire period," to
"In the assimilation system, raw data are not considered to be reliable,".

14 23-24. I would add “and more detailed aerosol chemistry mechanisms”
A2: Thank you. We now added "and more sophisticated aerosol chemistry mechanisms".

14 26. “The best use . . . ” I would tone down this statement
A2: Now it is changed to "One of the best ways of utilizing such invaluable observations ".

15 2-4. Similar to my previous comment, might be better to just state that one is a column integrated quantity and the other is surface, and their connection depends on the vertical distribution of aerosols and conversion from mass to optical properties.
A2: These aspects are stated in the following points (ii and iii). We only change "surface $PM_{2.5}$" to "$PM_{2.5}$ on the ground".

22-23. Not correct, see Peterson et al (2019). There are periods of stagnation where local contribution generates pollution episodes that are not negligible.
A2: This is correct and consistent with the findings in Peterson et al (2019).
We defined heavy pollution events as the cases with surface $PM_{2.5}$ > 50 μg/m^3, as specified in page 13, line 32.
In Peterson et al. (2019), for the same $PM_{2.5}$ (in the bottom panel of Fig. 3), they highlighted the heavy pollution events (based on the same reference line of 50 μg/m^3) in orange and marked them as "Transport" cases.
Thanks for making us confirm that our finding is correct here. No changes.

Technical Corrections

1 7-10. This sentence reads as you are using MODIS and surface PM to evaluate the assimilation, but are actually assimilating them as well. Please rephrase.
A2: We now added "those of" as "$\sim$ comparing with those of other observations". Thanks for your clarification.

Figure 6. State in the caption which DA experiment is plotted.
A2: We now added a sentence at the end of the caption as below.
Here, "DA" refers to the "ALL" experiment.

10 1. "… coefficient of 0.33, the two observation types …"
A2: "the" is now added in front of "two observation types...", as suggested.

10 17 "… observation types separately, …"
A2: Corrected.

Figure 7. State concentration units in the caption. Also, use full specie name instead of abbreviation or reference location in the text where species are defined, some of the names are not straightforward to to figure out.
A2: Thanks for your suggestion.
Figure 7 is now completely replotted with new caption "Vertical profile of 10 GOCART aerosol variables composed of $PM_{2.5}$ - unspeciated aerosol contributions to $PM_{2.5}$ (P25), sulfate, OC1 and OC2 (BC1 and BC2) as hydrophobic and hydrophilic organic (black) carbon, respectively, DUST1 and DUST2 (SEAS1 and SEAS2) as dust (sea salt) aerosols in the smallest and 2nd smallest size bins. All the variables shown are mixing ratios in the unit of μg/kg. Different experiments are depicted in different colors, as averaged over domain 2 for the period of May 4 - 31, 2016. Analysis ("A") is drawn as solid line while background (e.g. 6-h forecast; "B") as dashed line."

10 29 "… produces the largest PM2.5 throughout the …"
A2: Modified, as suggested.

14 1. "As our best experiment "ALL" analyzed," not clear what this means, please rephrase
A2: It is now changed to "As the analysis of our best experiment "ALL" showed,".

14 9 ."are near sea level"
A2: Now, "the" is omitted, as suggested.

References

LeGrand, S. L., Polashenski, C., Letcher, T. W., Creighton, G. A., Peckham, S. E. and Cetola, J. D.: The AFWA dust emission scheme for the GOCART aerosol model in WRF-Chem v3.8.1, Geosci. Model Dev., doi:10.5194/gmd-12-131-2019, 2019.

Improving air quality forecasting with the assimilation of GOCI AOD retrievals during the KORUS-AQ period

Soyoung Ha[1], Zhiquan Liu[1], Wei Sun[1], Yonghee Lee[2], and Limseok Chang[2]

[1]National Center for Atmospheric Research, Boulder, Colorado, USA
[2]National Institute of Environmental Research, Incheon, South Korea

Correspondence: Soyoung Ha (syha@ucar.edu)

Abstract. The Korean Geostationary Ocean Color Imager (GOCI) satellite has monitored the East Asian region in high temporal and spatial resolution every day for the last decade, providing unprecedented information on air pollutants over the upstream region of the Korean peninsula. In this study, the GOCI aerosol optical depth (AOD), retrieved at 550 nm wavelength, is assimilated to ameliorate the analysis quality, thereby making systematic improvements on air quality forecasting in South Korea. For successful data assimilation, GOCI retrievals are carefully investigated and processed based on data characteristics. The preprocessed data are then assimilated in the three-dimensional variational data assimilation (3DVAR) technique for the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem). During the Korea-United States Air Quality (KORUS-AQ) period (May 2016), the impact of GOCI AOD on the accuracy of air quality forecasting is examined by comparing with those of other observations including Moderate Resolution Imaging Spectroradiometer (MODIS) sensors and fine particulate matter ($PM_{2.5}$) observations at the surface. Consistent with previous studies, the assimilation of surface $PM_{2.5}$ concentrations alone systematically underestimates surface $PM_{2.5}$ and its positive impact lasts mainly for about 6 h. When GOCI AOD retrievals are assimilated with surface $PM_{2.5}$ observations, however, the negative bias is diminished and forecasts are improved up to 24 h, with the most significant contributions to the prediction of heavy pollution events over South Korea.

1 Introduction

With the recent increase of chemical and aerosol observations in the troposphere, chemical data assimilation is expected to play an essential role in improving air quality forecasting, particularly in the real-time environment. Although various data assimilation (or analysis) techniques have been developed for many decades, they were predominantly applied in the context of numerical weather prediction (NWP) (Kalnay, 2003) and have not been extensively exploited for the prediction of air pollution.

Uncertainties in aerosol chemistry, as well as its multiscale interactions with daily changing weather conditions, make it challenging to predict air pollutants accurately (Grell and Baklanov, 2011; Baklanov et al., 2014; Kong and coauthors, 2015; Baklanov et al., 2017). Surface concentrations are directly affected by transport and dispersion of chemical species through advection, convection, vertical diffusion and surface fluxes. In general, they are strongly driven by external forcing such as anthropogenic and natural emissions. The latter heavily relies on temperature, humidity, and wind speed in the boundary layer as well as solar radiation and soil moisture. Aerosols in turn affect local meteorology via aerosol-meteorology interaction (by directly scattering and absorbing solar radiation and also as sources of cloud condensation nuclei) at short time scales. For the operational air quality forecasting in South Korea, the Korean National Institute of Environmental Research (NIER) performs chemical simulations on 3-km resolution at present (Chang et al., 2016). For such a high-resolution application and for situations with very high aerosol concentrations, these fast-varying complex mechanisms might be better represented through online coupling between chemical and meteorological components. The online coupled forecasting system is particularly suitable for air quality forecasting associated with strong synoptic forcing or long-range transport of air pollutants. Also, finer scale features may require more frequent coupling of the atmospheric system and only the online coupled system can provide the framework for such applications.

With large uncertainties in chemical modeling and emission data, particularly associated with meteorological components, one of the most effective ways of utilizing aerosol observations is to assimilate them into the forecast model and improve the initialization of aerosol simulations. However, due to the scarcity of three-dimensional chemical observations and the complexity of how to project the observed information (usually in the optical properties) onto the parameterized schemes in the chemical model, aerosol/chemical data assimilation in the coupled chemistry and meteorology models has been limited to date (Bocquet et al., 2015). Improving the quality of chemical assimilation will not only improve the prediction of air pollution, but also advance numerical weather prediction (NWP) for precipitation, visibility, and high impact weather.

An international cooperative air quality field study conducted in Korea, named as the Korea-United States Air Quality (KORUS-AQ), was a field campaign jointly developed by air quality researchers in the United States and South Korea to improve our understanding of major contributors to poor air quality in Korea for May 1-June 12, 2016. During this early summer time when it is mostly warm and humid, numerous measurements of pollutants were made at multiple platforms in an effort to identify local and transboundary pollution sources contributing to the formation of ozone and $PM_{2.5}$. Although local emissions played a nontrivial role throughout the period, the highest pollution event occurred by long-range transport from the upwind area on May 25-27, 2016 (Miyazaki et al., 2019). As the transboundary transport cannot be fully measured by surface stations over land, a proper use of satellite data that have a wide spatial coverage would have great potential to improve air quality forecasting for such events.

The Korean Geostationary Ocean Color Imager (GOCI) onboard the Communication, Ocean, and Meteorology Satellite (COMS) provides hourly AOD retrievals at multiple spectral bands monitoring the East Asian region centered on the Korean peninsula during daytime (Kim et al., 2017). Since its launch in 2010, the GOCI satellite has been producing AOD retrievals at high spatial and temporal resolution. It has long been demonstrated that the GOCI data were in high accuracy, comparable to the low-orbiting Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS) products (Lee et al. (2010); Wang et al. (2013); Xiao et al. (2016); Choi et al. (2018)).

Liu et al. (2011) first implemented the capability of assimilating Aerosol Optical Depth (AOD) retrieved from MODIS satellite sensors (Remer et al., 2005) into the National Centers for Environmental Prediction (NCEP) Gridpoint Statistical Interpolation (GSI; Wu et al. (2002); Kleist et al. (2009)) system. Since they confirmed that the AOD assimilation improved aerosol forecasts in a dust storm event that occurred in East Asia, the GSI three-dimensional variational data assimilation (3DVAR) system has been widely used for air quality forecasting and extended for additional aerosol observations such as surface particulate matter - all particles with aerodynamic diameter less than 2.5 μm ($PM_{2.5}$) or up to 10 μm (PM_{10}) (Schwartz et al. (2012) and Jiang et al. (2013), respectively).

GOCI AOD retrievals have been assimilated in several studies to assess their impact on short-term air pollution forecasts in the online coupled forecasting system. Saide et al. (2014) performed the Observing System Experiment (OSE) using the eight bin MOdel for Simulating Aerosol Interactions and Chemistry aerosol model (MOSAIC) (Zaveri et al., 2008) in the WRF-Chem/GSI 3DVAR system. Pang et al. (2018) assimilated AOD retrievals from GOCI and Visible Infrared Imaging Radiometer Suite (VIIRS; Jackson et al. (2013)) to predict surface $PM_{2.5}$ concentrations over Eastern China and found that the assimilation of AOD retrievals improved the forecast accuracy but still underestimated heavy pollution events.

This work further extends the assimilation capabilities in the GSI 3DVAR system to best use GOCI AOD retrievals during the KORUS-AQ period with careful investigation of data characteristics. Aiming at improving the operational air quality forecasting in Korea, which is currently lacking the state-of-the-art analysis system, we are discussing how to effectively assimilate satellite-derived aerosol data and examine its impact on surface $PM_{2.5}$ predictions compared to that of other observations. In the categorical forecasts for different air pollution events, we focus on severe pollution cases describing how air pollutants evolve, coupled with the synoptic weather systems.

A brief overview of the analysis and forecasting systems used in this study is presented in Section 2, followed by cycling experiments with details on observation processing for GOCI retrievals described in Section 3. Results are summarized in Section 4 discussing the observation impact during the cycles and extended forecasts separately. Forecast performances in heavy pollution events are briefly described as well. Finally, conclusions are made in Section 5, along with a discussion on the limitations of this study and suggestions for the future research.

2 The WRF-Chem forecast model and the GSI 3DVAR analysis system

2.1 WRF-Chem forecast model

The model used in this study is an online-coupled meteorology and chemistry model, WRF-Chem version 3.9.1 (Grell et al., 2005). The physics options used in WRF-Chem include the rapid and accurate radiative transfer model for GCM (RRTMG) for long-wave radiation (Iacono et al., 2008), new Goddard shortwave radiation (Chou and Suarez, 1994), the Yonsei University (YSU) planetary boundary layer (PBL) scheme (Hong et al., 2006), the Lin microphysics scheme (Lin et al., 1983), as well as a new Grell 3D cumulus parameterization scheme. These options are chosen based on the operational configuration currently used in the Korean National Institute of Environmental Research (NIER) for their daily air quality forecasting in South Korea. The Goddard Chemistry Aerosol Radiation and Transport (GOCART; Chin et al. (2002)), developed by the National Aeronautics and Space Administration (NASA), is used as an aerosol scheme. Aerosol direct effects are allowed through the interaction between GOCART and the Goddard shortwave radiation scheme (Fast et al. (2006); Barnard et al. (2010)).

The Model for Ozone and Related Chemical Tracers (MOZART) gas phase chemistry (Emmons et al., 2010) is generated with the kinetic preprocessor (KPP) (Damian et al. (2002); Sandu and Sander (2006)), and is used together with the simple GOCART aerosol scheme, known as the MOZCART mechanism (Pfister et al., 2011). The MOZART chemistry in WRF-Chem is designed to run with the Madronich FTUV scheme for photolysis processes (Tie et al., 2003), reading in climatological O3 and O2 overhead columns. It also utilizes the standard WRF-Chem implementation of the Wesley dry deposition scheme (based on Wesely (1989)) allowing for seasonal changes in the dry deposition. The resolved scale wet scavenging is inactivated but convective wet scavenging is applied in the Grell cumulus parameterization. Also, GOCART sea salt emissions and dust emissions with AFWA modifications (LeGrand et al., 2019) are included in this study.

Anthropogenic emissions are estimated offline based on the global EDGAR-Hemispheric Transport of Air Pollutants (HTAP) emission inventory (http://edgar.jrc.ec.europa.eu/htap_v2/) that consisted of 0.1° x 0.1° gridmaps of CH4, CO, SO2, NOx, NMVOC, NH3, PM10, PM2.5, BC and OC from the year of 2010. The emission data mapped to our model grids has a single level with no vertical variations and is generated from the annual mean with no diurnal variations (e.g. time-invariant). In terms of data range, the maximum (average) value of $PM_{2.5}$ in the data, for example, is 3.56 (0.032) μg m^{-2} s^{-1} and 2.84 (0.026) μg m^{-2} s^{-1} in domain 1 and 2, respectively.

Biogenic emissions are built up using the Model of Emission of Gases and Aerosol from Nature (MEGAN; Version 2) (Guenther et al., 2006) and for biomass burning emissions, daily fire estimates provided by Fire Inventory from NCAR (FINN; Wiedinmyer et al. (2011)) are used with tracer transport allowed. All the wrf files including biomass and biomass burning emissions are processed using the MODIS landuse datasets (Friedl et al., 2002).

2.1.1 The GSI 3DVAR analysis system

To assimilate AOD retrievals and surface $PM_{2.5}$ observations in the Weather Research and Forecasting-Chemistry (WRF-Chem) model, the NCEP GSI 3DVAR Version 3.5 system is used. As Liu et al. (2011) and Schwartz et al. (2012) described the details of the system for aerosol data assimilation, only a brief explanation follows. Incorporating observations into the three-dimensional model state space, a 3DVAR system produces the best estimate to the true state by minimizing the differences between observations and background forecasts (e.g. innovations; (o-b)'s), which is called the “analysis". Given the model state vector ($\mathbf{x}$), the penalty function (or cost function) $J(\mathbf{x})$ is defined as

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x_b})^{\mathbf{T}}\mathbf{B^{-1}}(\mathbf{x}-\mathbf{x_b}) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y})^{\mathbf{T}}\mathbf{R^{-1}}(H(\mathbf{x})-\mathbf{y}), \quad (1)$$

where $\mathbf{x_b}$ stands for the background state vector (e.g. forecasts from the previous cycle), $\mathbf{y}$ an observation vector, and H is an observation operator that projects the model states onto the observation space linearly or nonlinearly to compute the model correspondent to each observation. Background and observation error covariance matrices $\mathbf{B}$ and $\mathbf{R}$, respectively, indicate how reliable the background forecast ($\mathbf{B}$ in the first term) and the observed information ($\mathbf{R}$ in the second term) might be to determine how to properly weight the two disparate resources. By minimizing the cost function ($J(\mathbf{x})$) with respect to the model state vector $\mathbf{x}$ at the analysis time, the variational analysis algorithm produces the analysis that fits best to all the observations assimilated within the assimilation time window.

To characterize the forecast error magnitude and its spatial structure, background error covariance $\mathbf{B}$ is estimated for each aerosol species using the National Meteorological Center (NMC) method (Parrish and Derber, 1992) based on the differences between 48- and 24-h WRF-Chem forecasts valid at the same time for 30 samples ending at 0000 UTC in May 2016. The current GSI/3DVAR system does not allow cross-correlation between aerosol species or between aerosol and meteorological variables. As this is a 3DVAR analysis with no time information, B only characterizes the spatial correlations in each analysis variable, which determines how to propagate the observed information across the model grids.

Following Liu et al. (2011) and Schwartz et al. (2012), this study also takes the speciated approach where the analysis vectors are comprised of 15 WRF-Chem/GOCART aerosol variables - sulfate, organic carbon (O) and black carbon (B), mineral dust (D) in five particle-size bins (with effective radii of 0.5, 1.4, 2.4, 4.5, and 8.0 μm), and sea salt (S) in four particle-size bins (with effective radii of 0.3, 1.0, 3.25, and 7.5 μm for dry air), and P as unspeciated aerosol contributions to $PM_{2.5}$ -, as opposed to using total aerosol mass of $PM_{2.5}$ as the analysis variable in Pagowski et al. (2010). For organic and black carbon, hydrophobic and hydrophilic components are considered (e.g. O_1, O_2, B_1, and B_2).

The observation operator $H(\mathbf{x})$ for surface $PM_{2.5}$ requires 10 GOCART aerosol variables as

$$H(\mathbf{x}) = \rho_d[P + D_1 + 0.286D_2 + 1.8(O_1 + O_2) + B_1 + B_2 + S_1 + 0.942S_2 + 1.375U], \qquad (2)$$

where P represents unspeciated aerosol contributions to $PM_{2.5}$; U denotes sulfate; O_1 and O_2 (B_1 and B_2) are hydrophobic and hydrophilic organic (black) carbon, respectively; and D_1 and D_2 (S_1 and S_2) are dust (sea salt) aerosols in the smallest and 2nd smallest size bins. This formula originated from the WRF-Chem diagnostics of $PM_{2.5}$ for the GOCART aerosol scheme. PM observations are mass concentrations in μg/m^3 while all the model variables listed within the bracket in the right-hand side are aerosol mixing ratios (μg/kg), dry density ρ_d is thus required for the unit conversion in equation 2.

In this study, we assimilate AOD retrievals at 550 nm from both MODIS and GOCI sensors using the same observation operator based on the community radiative transfer model (CRTM; Han et al. (2006); Liu and Weng (2006)) as described in Liu et al. (2011). Although the GOCART aerosol scheme is well known to underestimate surface PM concentrations due to the lack of secondary organic aerosol (SOA) formation, nitrate, and ammonium (Liu et al. (2011); Volkamer et al. (2006); McKeen et al. (2009); Pang et al. (2018)), it is widely used in the analysis study because it is the only scheme publicly available for assimilating AOD retrievals from satellite data in the GSI system. Aerosol optical depth (AOD) measures the amount of light extinction by aerosol scattering and absorption in the atmospheric column which depend on the refractive indices and the size distribution of aerosol. In GSI, the CRTM computes the effective radii and the refractive indices of the 14 speciated WRF-Chem/GOCART aerosol species, assuming spherical aerosol particles and lognormal size distributions. Applying single-scattering properties of spheres by Mie theory, the mass extinction coefficient is computed as a function of the effective radius for each aerosol species at a certain wavelength (here, 550 nm) at each model level. The mass extinction coefficient (m^2/g) for each aerosol species multiplied by the aerosol layer mass (g/m^2) produces dimensionless AOD for the species at that level. To represent the entire atmospheric column, model-simulated AOD is then computed as the column integration of AOD for all aerosol species. Using the CRTM as a forward operator, AOD retrievals are assimilated separately or simultaneously with $PM_{2.5}$ observations from the surface network over East Asia, as described in the following section.

3 Cycling Experiments

During the month of May 2016, observations are assimilated in the GSI 3DVAR system to produce the analysis that is used as an initial condition for the following WRF-Chem simulations. WRF-Chem forecasts valid at the next analysis time are then used as first-guess (or background) for the next GSI analysis. In this study, the whole process is repeated every 6 h (called "cycled") for the month-long period. Here we describe the analysis and the forecast systems used in the cycling.

3.1 Model configurations and cycling

All the analyses and the following forecasts are conducted over two one-way nested domains centered on the Korean peninsula, as shown in Fig. 1. Domain 1 uses 175 x 127 horizontal grids at 27-km resolution and domain 2 has 97 x 136 grids at 9-km resolution. Both domains have total of 31 vertical levels up to 50 hPa. The initial and boundary meteorological conditions for domain 1 are provided by the U.K. Met Office Unified Model (UM-MET) global forecasts operated by the Korean Meteorological Administration (KMA) with a horizontal resolution of $\sim$25 km (0.3515°x0.234375°) at 26 isobaric levels every 6 h. This configuration was chosen in the limitation of computational resources, but the use of higher resolutions both in time and space might be desirable to further improve forecast skills in the future. The chemical initial and boundary conditions for domain 1 are taken from the output of the global Model for Ozone and Related Chemical Tracers (MOZART-4) (Emmons et al., 2010) that are converted to WRF-Chem species by using the “mozbc" utility (downloaded from https://www2.acom.ucar.edu/wrf-chem/wrf-chem-tools-community/). Meteorological and chemical fields in domain 1 are reinitialized from global forecasts every cycle while initial and boundary conditions for domain 2 are nested down from domain 1 in a one-way nesting. Aerosol and chemical initial conditions are then overwritten by WRF-Chem forecasts from the previous cycle in each domain. The GSI analysis is consecutively performed in the two domains using the same observations within each domain. During the cycles, 24 h forecasts are initialized from the 00Z analysis every day.

3.2 Observations

3.2.1 Surface $PM_{2.5}$

Hourly surface PM concentrations are provided by the NIER which collects real-time pollutant observations at 361 South Korean stations from AirKorea (http://www.airkorea.or.kr) and those at $\sim$900 Chinese sites from China National Environmental Monitoring Centre (CNEMC; http://www.cnemc.cn). Figure 1 shows the entire surface observing network that was used to assimilate surface $PM_{2.5}$. Observation sites are concentrated in the urban area where many sites are close enough to be overlapped with each other. The Seoul Metropolitan Area (SMA; centered around 37.5°N, 127°E), for example, has hourly reports from total of 41 stations.

As part of data quality control (QC), surface $PM_{2.5}$ concentrations higher than 100 μg/m^3 are not assimilated and observations producing innovations ((o-b)'s) that exceed 100 μg/m^3 were also discarded during the analysis step. To accommodate most measurements in China during heavy pollution events, a much higher threshold of 500 μg/m^3 was once applied as the maximum observed value in our test experiment for the same month-long cycles, but it did not lead to any meaningful changes in the forecast performance over South Korea (not shown). Presumably this is because such high values were observed only over China where air pollutants were already overestimated by the emission data based on the 2010 inventory such that the forecast skills over Korea became insensitive to the assimilation of those additional surface observations in China. Therefore, we applied the original threshold of 100 μg/m^3 to all our experiments presented here.

Observation error is composed of measurement error (ϵ_o) and the representative error (ϵ_r) caused by the discrete model grid spacing (e.g. $\epsilon_{pm_{2.5}} = \sqrt{{\epsilon_o}^2 + {\epsilon_r}^2}$). Following Elbern et al. (2007) and Schwartz et al. (2012), observation error for surface $PM_{2.5}$ increases with the observed value (x_o) as $\epsilon_o = 1.5 + 0.0075 * x_o$. The representative error is formulated as $\epsilon_r = \gamma \epsilon_o \sqrt{\frac{\Delta x}{L}}$ where γ is 0.5, Δx is grid spacing (here, 27 km for domain 1 and 9 km for domain 2) and the scaling factor L is defined as 3 km. Based on this formula, observation error ($\epsilon_{pm_{2.5}}$) ranges from 2.0 to 3.2 μg/m^3 in domain 2, assigining the error of 2.48 μg/m^3 to the $PM_{2.5}$ observation of 50 μg/m^3, for example. In this 3DVAR analysis, observation errors are considered to be uncorrelated so that the observation error covariance matrix $\mathbf{R}$ becomes diagonal. During the 6-h cycling, all the surface observations within $\pm$1 h window at each analysis time were assimilated without further adjustment of observation error.

3.2.2 AOD retrievals and observation preprocessing

Total AOD retrievals at 550 nm from MODIS sensors onboard Terra and Aqua satellites have been widely used in aerosol studies (Zhang and Reid, 2006, 2010; Lee et al., 2011). But the polar-orbiting satellites produce a very limited dataset temporally (mostly around 06 UTC only) and spatially (with a sparse coverage) over Korea during the KORUS-AQ period. The MODIS AOD level 2 products over both land and ocean "dark" area are available at 10 km x 10 km resolution and thinned over 60-km resolution during the GSI analysis in this study. Following Remer et al. (2005), observation errors are specified as the retrieval errors: (0.03 + 0.05 * AOD) over ocean and (0.05 + 0.15 * AOD) over land. They do not include the representativeness error and are slightly smaller than those for GOCI AOD, as described below.

The GOCI satellite monitors the East Asian region centered on the Korean peninsula (36°N, 130°E) covering about 2500 km $\times$ 2500 km. GOCI level II data has eight spectral bands from the visible to near-infrared range (412 to 865 nm) with hourly measurements during daytime from 9:00 (00 UTC) to 17:00 local time (08 UTC) at 6 km resolution. As summarized in Choi et al. (2018), a recently updated GOCI Yonsei aerosol retrieval (YAER) Version 2 algorithm targets cloud- and snow-free pixels over land and cloud- and ice-free pixels over ocean in producing the level II data. By adopting the MODIS and VIIRS aerosol retrieval and cloud-masking algorithms, cloud pixels are filtered to avoid cloud contamination, and high reflectance or highly heterogeneous reflectance pixels are also masked to further increase data accuracy and consistency during the retrieval process.

Unlike MODIS retrievals, GOCI AOD has not been extensively used in the data assimilation community. The GSI system takes most observation types in prepbufr, which has already gone through some processing to be prepared for data assimilation, but the preprocessing algorithms are not publicly available. This means that, when a new dataset is assimilated in GSI, users need to investigate the characteristics of the data (such as temporal and spatial distribution) and thereby make the data suitable for assimilation, which is of crucial importance for the analysis quality.

In terms of temporal distribution, most of GOCI level II data are retrieved on 30 min passed each hour in the hourly report. For example, the actual time for most of the data reported at 00 UTC is centralized around 00:30:00 UTC (hh:mm:ss). In the 3DVAR algorithm, there is no time dimension and all observations are considered to be available at the analysis time. To account for temporal distribution, different weights are often given to observations based on the relative distance between the actual report time and the analysis time during the analysis step. However, taking possible latency in data transfer and retrieval processing into consideration, it is not legitimate to assign weights to the retrievals based on their final report time, without further information. Therefore, considering high temporal and spatial variability of aerosols, the assimilation window is set to ± 1 h in order to avoid inconsistent observed information within the window in this study.

Satellite data are known to have a large positive impact on the analysis quality thanks to the high data volume both in time and space, but such high density violates the assumption of uncorrelated observation errors in the analysis algorithm and increases the computation time for the analysis step excessively. Hence, a large volume of satellite retrievals are typically sampled on a regularly spaced grid through the horizontal thinning procedure. In GSI, satellite radiance data can be thinned such that retrievals are randomly sampled at a predefined spacing for each instrument type before getting ingested into the observation operator during the analysis (Rienecker and Coauthors, 2008). This thinning procedure, however, can pick up inconsistent data (near the cloud boundaries, for instance) and is reported as suboptimal (Ochotta et al., 2005; Reale et al., 2018). Therefore, we decided to preprocess GOCI AOD retrievals with superobing where all the data points are averaged within a certain radius. In this study, we superobed GOCI retrievals over each grid box in domain 1 (at 27-km resolution). Figure 2 shows the sample horizontal distribution of GOCI AOD retrievals valid at 06 UTC May 1 2016 before (a) and after (b) preprocessing them, comparing with those thinned over 60 km (c) and 27 km meshes (d) during the GSI analysis, respectively. Some high AOD values in the original dataset (as shown in a), especially on cloud edges, cannot be fully resolved by our 27-km model grids. By averaging all data points over each grid box at 27-km resolution, the superobed data in b) have a better quality control throughout the domain reducing the data volume effectively. A total number of observations marked in the upper right corner of each panel indicates that thinning over the 60 km mesh in c) reduces the number of assimilated observations to 2.5% of that in the original level II data while superobing and thinning over 27-km mesh utilize 8-10 % of the original data representing the whole data coverage fairly well.

It might be noteworthy to make two more points related to data processing here. First, superobing was applied as part of preprocessing before the GSI analysis gets started while the thinning was conducted during the analysis step so that the preprocessing could speed up the GSI analysis up to 25 times (by injecting less than 10 % of the original data and turning off the thinning process). This can facilitate the use of satellite retrievals in the operational air quality forecasting. Next, the thinning algorithm in GSI V3.5 resulted in erroneous values in some places, as indicated by the maximum values in c) and d). For the month of May 2016, multiple cases with such extreme fake values were found after the thinning process. This bug may need to be fixed in the GSI or avoided by bounding the values exceeding the original data.

To examine the effect of data processing on the performance of the analysis and the background during the cycles, we compare two cycling experiments - one with the assimilation of the original level II data thinned over 27-km mesh (named "GOCI_orig" in gray) and the other with the assimilation of GOCI retrievals preprocessed over 27-km grids in domain 1 (called

"GOCI" in black) - in Fig. 3. As GOCI data are reported from 00 to 08 UTC, only 00 and 06 UTC cycles are shown here in consecutive cycle numbers. The time series of (o-a)'s and (o-b)'s in each experiment show that the preprocessed data slightly fit better to the observations than the thinned data, assimilating more retrievals throughout the period. Because the differences between the two experiments are not significant, for the computational efficiency, we decided to preprocess all the GOCI retrievals and assimilate them turning off the thinning process in GSI for the rest of the experiments shown in this study.

Choi et al. (2018) described their improved retrieval algorithm (GOCI YAER V2) with updated cloud-masking and surface reflectance calculations, making a long-term validation against other ground- and satellite-based measurements. In their study, depending on the verifying objects - either ground-based Aerosol Robotic Network (AERONET) or satellite-based retrievals -, they specified uncertainties of GOCI AOD retrievals over land and ocean using two different linear regression formulae. We assign ϵ_1 following their error specification with respect to AERONET and ϵ_2 based on their expected error against retrieved satellite AOD in GOCI YAER V2.

$$\epsilon_1{}^{land} = 0.061 + 0.184\tau_A \tag{3}$$

$$\epsilon_1{}^{ocean} = 0.030 + 0.206\tau_A \tag{4}$$

$$\epsilon_2{}^{land} = 0.073 + 0.137\tau_A \tag{5}$$

$$\epsilon_2{}^{ocean} = 0.037 + 0.185\tau_A \tag{6}$$

where τ_A stands for GOCI AOD values. In an effort to account for representativeness error, we also tried with ϵ_2 increased by 20% everywhere as the third error formula (e.g. $\epsilon_3 = 1.2 \times \epsilon_2$) and compared all three types of errors in Fig. 4. When these different observation errors were applied to GOCI retrievals in the assimilation, the smallest error (ϵ_2) produced slightly better fits to observations specially for the high values (AOD > 2) during the cycles, as expected, but not in a statistically meaningful way (not shown). In fact, it is not straightforward to estimate the representativeness error which is subject to the model resolution (both in horizontal and vertical) and data processing in use. Therefore, in many cases, observation error is specified based on the resulting forecast performance (Ha and Snyder, 2014). But because our forecast skills were not very sensitive to three different error formulae tried here, for the rest of the experiments, ϵ_2 is used as observation error for GOCI retrievals.

The goal of this study is to examine the relative impact of the GOCI assimilation on the prediction of surface $PM_{2.5}$ and ultimately to improve the forecasts for pollution events. Although it is rather easy to render the analysis close to GOCI observations by reducing the observation error, it is not guaranteed that the analysis in a good agreement with AOD retrievals would actually lead to better forecasts in surface $PM_{2.5}$. This is partly because AOD, a column integrated quantity, is not directly associated with surface $PM_{2.5}$ and partly because large uncertainties in the forecast model and the emission forcing can dominate over the analysis error during the model integration. Even if the efficiency of assimilating AOD on improving surface $PM_{2.5}$ forecasts can be largely affected by the quality of the forecast model and the emission data in use, the effectiveness of the AOD assimilation is based on the relationship between the column-integrated AOD and $PM_{2.5}$ on the ground. Therefore, it might be worth checking the correlation between GOCI AOD retrievals and surface $PM_{2.5}$ observations for the cycling period. Figure 5 depicts a scatter diagram of GOCI AOD retrievals at 550 nm and surface $PM_{2.5}$ observations that are co-located in each grid box in domain 1 for the month of May 2016. As shown with the linear regression coefficient of 0.33, the two observation types have low correlations during this period, which is consistent with previous studies (Saide et al., 2014; Pang et al., 2018). Such an indirect relationship between the two observations makes the analysis challenging because it can induce a large error in the observation operator and heavily depend on the model's ability of deriving $PM_{2.5}$ from AOD based on the vertical structure of aerosol variables and the conversion from aerosol mass to optical properties.

4 Results

With a careful design of model configuration and observation processing, the overall impact of assimilating all the available observations ("DA") is illustrated, compared to the baseline experiment without data assimilation ("NODA") in Figure 6. Here, the 0-23 h hourly forecasts from all the 00Z analyses in domain 2 are concatenated for the entire month. Surface $PM_{2.5}$ observations marked as black dots show that the air quality gets distinctively aggravated for the last 7 days, related to long-range transport of air pollutants. With data assimilation ("DA"), the analyses at 00Z and the following forecasts (red) make a better agreement with corresponding observations than those without assimilation (gray), especially from day 15 (e.g. after a full spin-up for two weeks). In particular, on May 25-27, forecast error grows quickly even from the good analysis at 00Z, possibly associated with large uncertainties in lateral boundary conditions and the forecast model in use. However, averaged over the entire period, the mean absolute error (mae) indicates that the performance of 0-23-h forecasts at 9-km resolution gets improved by $\sim$30% through data assimilation.

4.1 Observation impact during the cycles

Given that the aerosol assimilation has a positive impact on air quality forecasting, it might be worth isolating the contribution of each observation type to the improvement of the analysis and the following forecasts. We first assimilate individual observation types separately, naming the experiment following each observation type, then we assimilate them all together (called "ALL"). Figure 7 illustrates the vertical profile of 10 three-dimensional GOCART aerosol variables that are used in diagnosing $PM_{2.5}$ in the GOCART scheme, in the analysis (solid) and background (e.g. 6-h forecast; dashed) averaged over domain 2. Assuming that cycles may need to spin up meteorology and chemistry at least for three days in the regional simulations, all the statistics are computed from day 4 in the rest of the figures. Although the analysis variables only at the lowest model level are used in the observation operator for surface $PM_{2.5}$, the observation impact is detected throughout the atmosphere due to the spatial correlations specified in the background error covariance. Contributions of different observations to each analysis variable vary, with the largest variability in the analysis increments (analysis-minus-background) displayed in sulfate. Interestingly, a large impact of AOD retrievals is noticed in hydrophilic organic carbon (O_2) aloft (e.g. between 12 and 25 levels) and unspeciated aerosol (P) in the boundary layer. The assimilation of all the observations ("ALL") tends to reduce O_2, dust in both size bins (D_1 and D_2) and unspeciated aerosol (P) in the lower atmosphere.

Figure 8 summarizes the effect of different observations on $PM_{2.5}$ in both domains. The assimilation of surface $PM_{2.5}$ observations (green) results in the smallest $PM_{2.5}$ while the GOCI assimilation (blue) produces the largest $PM_{2.5}$ throughout the atmosphere in both domains. When the analysis (solid line) is compared to background (dashed), it is revealed that $PM_{2.5}$ is predominantly increased over domain 1 with the assimilation of GOCI retrievals. Overall, the aerosol assimilation affects the entire profile of $PM_{2.5}$ with the largest impact at the surface.

To understand the observation impact in the horizontal distribution, Fig. 9 shows the analysis increments (analysis-minus-background) averaged over the period of May 4-31, 2016. Generally, the assimilation of surface $PM_{2.5}$ observations ("PM") reduces surface $PM_{2.5}$ over most regions in China while the GOCI assimilation largely increases surface $PM_{2.5}$ almost everywhere, consistent with Fig. 8. As MODIS retrievals have a relatively low coverage of the East Asian region for the entire period, they have the smallest impact among all the observation types. When all the observations are assimilated together (in "ALL"), it combines the effect of surface $PM_{2.5}$ and GOCI retrievals, changing the vertical distribution of aerosol species to match with the AOD column values and pulling the surface states towards surface $PM_{2.5}$ concentrations. While the observing network of surface $PM_{2.5}$ is widely distributed over China, the impact of GOCI data is more centralized over Korea, making unequivocal contributions to air quality forecasting in the Korean peninsula.

Note that we employ the 2010 inventory for our emission data, which does not reflect the emission control started from 2013 in China (Zheng et al., 2018). Given that air pollutants in the emission data constitute the majority of the precursors of $PM_{2.5}$ pollution, surface $PM_{2.5}$ could strongly depend on emissions which might have led to the overestimation in the background (e.g. first-guess). Therefore, the assimilation of surface $PM_{2.5}$ tends to counteract the overestimation driven by the emission data over China. On the other hand, over South Korea, the emission data does not seem to be overestimated and the assimilation of surface $PM_{2.5}$ leads to increasing surface $PM_{2.5}$ most effectively during the cycles.

Different from surface particulate matter, AOD in the background is contingent upon the optical properties described in the observation operator (e.g. CRTM) and the vertical structure of aerosols simulated in the column. The influence of GOCI assimilation may indicate the model deficiencies in the two aspects because the model states are pulled toward the observed information during the analysis step, as depicted in the analysis increment.

4.2 Observation impact on 24-h forecasts

Since the real effect of data assimilation is manifested in the subsequent forecasts, we now examine forecast improvements when initialized by our own analyses. A good analysis is expected to slow down the forecast error growth, leading to better forecasts. In this subsection, forecast errors at the lowest model level are compared between experiments for 24 h with respect to surface observations from various sites in South Korea. As we focus on 9-km simulations over the Korean peninsula, it is hard to anticipate the direct effect of the assimilation beyond 24 h, specially in such a small domain where the weather systems dramatically change from day to day. As shown in Fig. 10, the forecast error is the largest in the baseline experiment ("NODA"), followed by the assimilation of MODIS retrievals alone ("MODIS") in terms of mean absolute error (mae). Note that the analysis in the "PM" experiment is verified against the same surface $PM_{2.5}$ observations used in the assimilation. Therefore, the analysis error is smaller than those in other experiments, but the forecast error grows quickly over the next 24

h. The assimilation of surface $PM_{2.5}$ alone generally underestimates the prediction of surface $PM_{2.5}$ with the fastest growth of forecast error. On the other hand, the assimilation of AOD retrievals (either GOCI or MODIS) alone does not improve the surface analysis and mostly overestimates surface $PM_{2.5}$ for 24 h. This might be ascribed to an imperfection of the forward operator of AOD and the model deficiency in the representation of three-dimensional aerosol species that comprised AOD and $PM_{2.5}$. When assimilated with surface $PM_{2.5}$ observations (in "ALL"), however, AOD retrievals effectively reduce the forecast error and suppress the error growth throughout 24 h forecasts.

Recently, heavy pollution events have often taken place over Korea and considerable attention was drawn to the accuracy of the operational air quality forecasting in the country, particularly in surface $PM_{2.5}$. As it has a great social impact to accurately predict exceedance and non-exceedance events in categorical predictions, it is necessary to evaluate the forecast performance for different categorical events. While Miyazaki et al. (2019) classified the entire KORUS-AQ campaign period into four different phases based on dominant atmospheric circulation patterns, we categorize events for the month of May 2016 based on hourly surface $PM_{2.5}$ concentrations, as summarized in Table 2 and 3. Figure 11 summarizes the evaluation of 24 h forecasts based on the formulae described below.

$$Overall\_Accuracy(\%) = \frac{a1+b2+c3+d4}{N} \times 100 \tag{7}$$

$$High\_Pollution\_Accuracy(\%) = \frac{c3+d4}{III+IV} \times 100 \tag{8}$$

$$Overestimation(\%) = \frac{b1+c1+c2+d1+d2+d3}{N} \times 100 \tag{9}$$

$$Underestimation(\%) = \frac{a2+a3+a4+b3+b4+c4}{N} \times 100 \tag{10}$$

$$False\_Alarm(\%) = \frac{II}{II+IV} \times 100 \tag{11}$$

$$Detection\_Rate(\%) = \frac{IV}{III+IV} \times 100 \tag{12}$$

where $I = a1+a2+b1+b2$, $II = c1+c2+d1+d2$, $III = a3+a4+b3+b4$, and $IV = c3+c4+d3+d4$. The air quality forecasting operated by the Korean NIER is currently evaluated in the same way on a daily basis, except for daily mean values.

In all events, the overall accuracy of 0-24-h forecasts is the highest in "ALL" (~70 %) and the lowest in "NODA" (~60 %), making about 10 % improvement by assimilation during this KORUS-AQ period. It is noted that the forecast error illustrated in Fig. 10 is dominated by days with clear sky or moderate air quality conditions (about two thirds of the month-long period, as shown in Fig. 6) while the forecast accuracy summarized in Fig. 11 is determined by equally weighting different categorical forecasts with different sample sizes. This implies that the categorical forecast evaluation tends to emphasize the forecast accuracy for pollution events (which has a smaller sample size). As such, Fig. 11a highlights the effect of data assimilation on improving air pollution forecasts. Differences between experiments are much larger in high pollution events (Fig. 11b) and the detection rate (Fig. 11f) where AOD retrievals (both GOCI and MODIS) make the biggest positive contributions. While "NODA" produces poor forecasts consistently in most metrics shown in Fig. 11, the forecast accuracy in "PM" (green) drops very quickly for the first 12 h for all events (a) and pollution events (b), indicating that the assimilation of surface $PM_{2.5}$ alone may not be enough to maintain the forecast skills beyond the cycling frequency (e.g. 6 h). It also increasingly underestimates surface $PM_{2.5}$ with time, especially after 20 h, and produces more false alarms even though its overestimation rate is the lowest among all experiments. Overall, the AOD assimilation tends to overestimate the prediction of surface $PM_{2.5}$ with a relatively large false alarm, but clearly helps enhance the forecast accuracy up to 24 h when assimilated with surface $PM_{2.5}$ observations. Even with low correlations with surface $PM_{2.5}$ (as illustrated in Fig. 5), AOD retrievals keep the surface air pollution forecasts from drifting away from the true state, compensating for model deficiencies. This demonstrates that it could be substantially beneficial to monitor a wide range of the surrounding area using the geostationary satellite in the enhancement of air quality forecasts.

In order to verify our forecasts against independent observations, we processed total AOD at 500 nm from the Aerosol Robotic Network (AERONET; https://aeronet.gsfc.nasa.gov/) sites and surface $PM_{2.5}$ concentrations measured at three more stations operated by NIER during the KORUS-AQ field campaign (Fig. 12). The level 2 quality level data are used for AERONET AOD observations as cloud-free and quality-assured data. Figure 13 illustrates the time series of hourly AOD from our experiments compared to hourly averages of AOD observations from 8 AERONET sites (black dots). At all sites, GOCI (blue) produces the largest AOD at most of high peaks while PM (green) and NODA (gray) simulate the smallest AOD throughout the period. Regardless of relative AOD values between the experiments, model forecasts are well matched with observations at low AOD values, but mostly miss high AOD observations, especially during the high pollution events for May 24-27. This leads to the negative mean bias (as (f-o)'s) in all experiments (shown in the legend), implying that our forecasts produce AOD slightly lower than the observed one as a whole. The rms error and mean bias at total of 16 AERONET sites are summarized in Table 4, indicating that GOCI has the smallest forecast error in AOD nation-wide.

Surface $PM_{2.5}$ measurements from three NIER sites were downloaded from https://www-air.larc.nasa.gov/cgi-bin/ArcView/korusaq as raw data with no quality control. They are provided as hourly averages starting from May 9 and compared to our hourly model output for May 9 - 31 (Fig. 14). These observations look somewhat noisy, but our forecasts broadly follow them throughout the period. Similar to the AOD verification shown in Fig. 13, forecasts from GOCI produce the smallest forecast mean bias among all the experiments in Olympic Park (in a) and Daejeon (in b), predicting the high surface $PM_{2.5}$ concentrations between May 24 and 26. But GOCI was worse than other experiments in Ulsan (in c), overestimating surface $PM_{2.5}$, especially during the high pollution days. In the assimilation system, raw data are not considered to be reliable, but this verification is included for the completeness because there was no other instrument that reported surface $PM_{2.5}$ concentrations or all the precursors of $PM_{2.5}$ concentrations to validate $PM_{2.5}$ forecasts on the ground level.

4.3 A heavy pollution case

The effect of assimilating different observations is most distinguishable in high pollution events, as demonstrated in Fig. 11. During the KORUS-AQ period, there were about 5 heavy pollution cases (when surface $PM_{2.5}$ > 50 μg/m^3, as defined in Table 2) over South Korea. The longest and the most severe pollution events have occurred on May 25-26, 2016. Fig. 15 illustrates how air pollutants have been transported from China, associated with the strong synoptic weather systems in the region for a few days. As the analysis of our best experiment "ALL" showed, the Korean peninsula was positioned in the downstream region of the upper-level trough at 500 hPa (in the left panel). In the low troposphere, the center of the North Pacific High was situated in the east of Japan bringing lots of moisture to Korea at 00 UTC 24 May 2016 and blocking the eastward movement of the surface low pressure system located north of Korea (centered around 46°N, 125°E), as shown in Fig. 15d. With the slowly approaching upper-level westerlies, these warm and moist conditions in the low troposphere provided a favorable environment for increasing air pollution in the Korean peninsula for the next few days. At 00 UTC 25 May, the Shangdong area in China (shown as the largest polluted area to the west of Korea) exceeded 150 μg/m^3 in surface $PM_{2.5}$ (Fig. 15b). This area is in high topography with elevations higher than 3.5 km (in height above ground level; AGL) while most regions in South Korea, especially the Seoul Metropolitan Area (SMA), are elevated near sea level. Therefore, when slow and deep baroclinic systems are approaching the Korean peninsula like these events, a deep pool of highly polluted air can be advected from China as a whole to substantially degrade the air quality in South Korea at least for a day or two. This long-range transport case produced an hourly maximum surface $PM_{2.5}$ observation of 117 μg/m^3 over the SMA in Korea at 00 UTC May 26, 2016, as shown in Fig. 16a.

One notable difference between observations (a) and all the model simulations (b-f) in Fig. 16 is that 9-km forecasts driven by 0.1° x 0.1° anthropogenic emissions cannot simulate such a high spatial variability across stations. During this heavy pollution event, there were dozens of missing observations to have a less number of stations in a) than all the experiments (b-f). With only 145 stations reporting high concentrations (e.g. surface $PM_{2.5}$ > 50 μg/m^3), the observed distribution still shows a sharp gradient between the stations, specially in SMA. Consistent with all the previous figures, the assimilation of surface $PM_{2.5}$ alone (in "PM") underpredicts surface $PM_{2.5}$ (even more than "NODA") while GOCI overpredicts surface $PM_{2.5}$ most among all observation types almost everywhere except for SMA. MODIS retrievals slightly increase the concentrations from NODA (by ∼10 μg/m^3), with the spatial distribution almost the same as that of NODA. In the concurrent assimilation of all the observations (in "ALL"), a moderate overestimation is presented everywhere, but higher levels of pollution in SMA are not simulated either. To resolve such a large variability between urban and rural area and to increase the sharpness of the forecast accuracy, the use of higher grid resolutions (such as 3 km), more accurate emission data and more sophisticated aerosol chemistry mechanisms might be indispensable.

5 Conclusions and discussion

GOCI AOD retrievals provide reliable and consistent aerosol information, monitoring air pollutants over the Korean peninsula at high resolution every day. One of the best ways of utilizing such invaluable observations is to inject them into the forecast system through data assimilation and better initialize numerical forecasts. For the successful assimilation of real observations, specially retrievals from satellites, extra attention should be paid to processing the data properly, based on the characteristics. The spatial and temporal representativeness of GOCI retrievals was carefully examined and the corresponding data processing was conducted before assimilation in this study. We averaged all the pixels over each grid box at 27-km resolution (e.g. superobing) instead of thinning them randomly, for instance.

It is worth noting several challenges in the assimilation of AOD retrievals for improving the prediction of surface $PM_{2.5}$ concentrations: i) AOD is not directly associated with $PM_{2.5}$ concentration on the ground. Although the two datasets can be highly correlated in specific conditions such as cloud-free, low boundary layer heights and low relative humidity, the overall correlation is low (~0.3) in the present study and it is hard to expect the direct impact on each other. ii) an observation operator for AOD has errors due to the simplification and the limited aerosol specifications in the community radiative transfer model (CRTM). iii) significant model error, which is presumably one of the most critical issues. In the 3DVAR assimilation, in particular, the model estimates of AOD, a column-integrated quantity, are strongly constrained by the model error structure of each aerosol species both horizontally and vertically.

Even with these challenges, however, satellite-based AOD, especially from geostationary satellites like GOCI, can be extremely useful for improving the prediction of air pollution on a daily basis. In the situation where the air quality can be largely affected by long-range transport of air pollutants, such consistent information on the wide upstream area is essential but hard to be obtained otherwise.

Using the GSI 3DVAR system coupled with the WRF-Chem forecast model, we assimilated the satellite AOD retrievals as well as surface $PM_{2.5}$ observations for the month of May 2016 during the KORUS-AQ period. Compared to the baseline experiment (“NODA"), the simultaneous assimilation of various observations consistently improved the prediction of ground $PM_{2.5}$ for 24-h forecasts, reducing systematic error and false alarms. The assimilation of ground $PM_{2.5}$ alone improved the analysis during the cycles, reducing the analysis error to almost half the size compared to the experiment without assimilation. However, the forecast error grew very quickly over the next 12 hours, underestimating $PM_{2.5}$ at the surface, especially in the heavy pollution events where the forecast accuracy dropped from over 70% to ~30% only in four hours. Meanwhile, the GOCI AOD retrievals alone tended to overestimate surface $PM_{2.5}$ but significantly contributed to improving air quality forecasts up to 24 h when assimilated with surface $PM_{2.5}$ observations. The effect of data assimilation is most distinguishable and remarkable for high pollution events. During the month of May 2016, most heavy pollution events were associated with long-range transport from China. In such cases, it was particularly beneficial to monitor the wide upstream region using geostationary instruments such as GOCI.

To assess the effect of data assimilation with respect to independent observations, 0-23 h forecasts from different experiments are verified against AOD from AERONET sites and ground $PM_{2.5}$ measurements from the sites operated during the KORUS-

AQ field campaign. In this verification, the assimilation of GOCI retrievals is the most effective in improving the forecast performance at most sites, especially for high pollution events.

Even with the successful data assimilation, there are several limitations in this study. First, the simple GOCART aerosol scheme is well known for the underestimation of air pollutants due to the lack of the aerosol size distribution and the secondary organic aerosol (SOA) formation. We had to use the scheme for the assimilation of AOD retrievals since the observation operator for AOD was only built for the GOCART scheme in the GSI system. Next, as there is no cross-covariance between aerosol and meteorological variables considered in the background error covariance estimates, the influence of aerosols on meteorological variables was not fully simulated in this study. Without the assimilation of meteorological observations, it was not possible to make an optimal estimate that is fully coupled between chemistry and meteorology although the meteorological information was provided through the first guess and lateral boundary conditions. Finally, the emission inventory used in this study was based on the annual mean of 2010, which did not reflect the actual emissions for the year of 2016, especially over China. The large bias and uncertainties in the emission data was particularly detrimental to the assimilation of surface $PM_{2.5}$ alone.

To overcome the systematic underestimation of the GOCART aerosol scheme in the assimilation context, there is an ongoing effort for a new development of an interface for more sophisticated aerosol schemes such as MOSAIC and/or the Modal for Aerosol Dynamics in Europe and the Volatility Basis Set (MADE/VBS; Ackermann et al. (1998), Ahmadov et al. (2012)) in the WRFDA system (Barker et al., 2012). This would be advantageous for more realistic forecast behavior in high resolution applications.

The positive impact of data assimilation is generally limited to 24-h forecast because of three major reasons: First, most air pollutants have a short lifetime due to dry and wet deposition and transformations through interactions with solar radiation and clouds. Secondly, pollutant transport and transformations in chemical transport models are strongly driven by external forcing, such as emissions, boundary conditions, and meteorological fields. Lastly, there are large uncertainties in aerosol and gas-phase chemistry parameterized in chemical transport models. Therefore, to extend the period of forecast improvements, emission data needs to be improved and large uncertainties in chemical and meteorological boundary conditions should be minimized. It has been shown that the estimation of emission inventories as part of the DA procedure can help extend the impact of data assimilation in longer forecasts (Elbern et al., 2007; Kumar et al., 2019). Also, more sophisticated aerosol and chemical mechanisms might be able to improve air quality forecasting by reducing model deficiencies (Chen et al., 2019). A simultaneous assimilation of meteorological observations and measurements of individual chemical species as well as particulate matter would be certainly beneficial in both NWP and air quality forecasting. To better account for high nonlinearities and uncertainties of aerosol forecasting on small scales, more advanced analysis techniques such as ensemble or hybrid data assimilation would be more desirable.

Code and data availability. The WRF-Chem v3.9.1 and the GSI v3.5 codes used in this paper will be available upon request. Input observations and boundary conditions for a sample test period can be also provided upon request.

Author contributions. ZL helped formulating the study and WS performed initial test runs. YL and LC provided input datasets, partially funding this study. SH designed and ran the experiments, analyzed the results and wrote the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. All the experiments presented here were performed on the Cheyenne supercomputer at the National Center for Atmospheric Research (NCAR). This work was jointly supported by the National Science Foundation under Grant No. M0856145 and the grant from the National Institute of Environment Research (NIER), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIER-SP2018-252). We acknowledge use of the WRF-Chem preprocessor tool (mozbc, fire_emiss, megan_bio_emiss, and anthro_emiss) provided by the Atmospheric Chemistry Observations and Modeling Lab (ACOM) of NCAR. Authors are also thankful for Seunghee Lee, Ganghan Kim and Myong-In Lee at UNIST in South Korea for their help transferring the input data for our experiments. Dave Gill and Wei Wang at MMM/NCAR helped us processing data in WPS and tuning the WRF configuration, respectively. Finally, special thanks should go to Gabriele Pfister at ACOM/NCAR and Dan Chen at China Meteorological Administration for their internal review, which greatly improved the manuscript.

References

Ackermann, I. J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F. S., and Shankar, U.: Modal aerosol dynamics model for Europe: development and first applications, Atmospheric Environment, 32, 2981 – 2999, https://doi.org/https://doi.org/10.1016/S1352-2310(98)00006-5, 1998.

Ahmadov, R., McKeen, S. A., Robinson, A. L., Bahreini, R., Middlebrook, A. M., de Gouw, J. A., Meagher, J., Hsie, E.-Y., Edgerton, E., Shaw, S., and Trainer, M.: A volatility basis set model for summertime secondary organic aerosols over the eastern United States in 2006, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD016831, 2012.

Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Moussiopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., and Zhang, Y.: Online coupled regional meteorology chemistry models in Europe: current status and prospects, Atmospheric Chemistry and Physics, 14, 317–398, https://doi.org/10.5194/acp-14-317-2014, https://www.atmos-chem-phys.net/14/317/2014/, 2014.

Baklanov, A., Brunner, D., Carmichael, G., Flemming, J., Freitas, S., Gauss, M., Hov, o., Mathur, R., Schlünzen, K. H., Seigneur, C., and Vogel, B.: Key Issues for Seamless Integrated Chemistry–Meteorology Modeling, Bulletin of the American Meteorological Society, 98, 2285–2292, https://doi.org/10.1175/BAMS-D-15-00166.1, 2017.

Barker, D., Huang, X.-Y., Liu, Z., AulignÃ©, T., Zhang, X., Rugg, S., Ajjaji, R., Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y.-R., Henderson, T., Huang, W., Lin, H.-C., Michalakes, J., Rizvi, S., and Zhang, X.: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA, Bulletin of the American Meteorological Society, 93, 831–843, https://doi.org/10.1175/BAMS-D-11-00167.1, 2012.

Barnard, J. C., Fast, J. D., Paredes-Miranda, G., Arnott, W. P., and Laskin, A.: Technical Note: Evaluation of the WRF-Chem "Aerosol Chemical to Aerosol Optical Properties" Module using data from the MILAGRO campaign, Atmospheric Chemistry and Physics, 10, 7325–7340, https://doi.org/10.5194/acp-10-7325-2010, 2010.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, Atmospheric Chemistry and Physics, 15, 5325–5358, https://doi.org/10.5194/acp-15-5325-2015, https://www.atmos-chem-phys.net/15/5325/2015/, 2015.

Chang, L.-S., Cho, A., Park, H., Nam, K., Kim, D., Hong, J.-H., and Song, C.-K.: Human-model hybrid Korean air quality forecasting system, Journal of the Air & Waste Management Association, 66, 896–911, https://doi.org/10.1080/10962247.2016.1206995, 2016.

Chen, D., Liu, Z., Ban, J., Zhao, P., and Chen, M.: Retrospective analysis of 2015–2017 wintertime $\mathrm{PM}_{2.5}$ in China: response to emission regulations and the role of meteorology, Atmospheric Chemistry and Physics, 19, 7409–7427, https://doi.org/10.5194/acp-19-7409-2019, 2019.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B., Duncan, B., Martin, R., Logan, J., Higurashi, A., and Nakajima, T.: Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements, Journal of the Atmospheric Sciences, 59, 461–483, https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2, 2002.

Choi, M., Kim, J., Lee, J., Kim, M., Park, Y.-J., Holben, B., Eck, T. F., Li, Z., and Song, C. H.: GOCI Yonsei aerosol retrieval version 2 products: an improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia, Atmospheric Measurement Techniques, 11, 385–408, https://doi.org/10.5194/amt-11-385-2018, 2018.

Chou, M.-D. and Suarez, M. J.: An efficient thermal infrared radiation parameterization for use in general circulation models., Tech. Memo 104606 [NTIS N95-15745], NASA, 1994.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G. R.: The kinetic preprocessor KPP-a software environment for solving chemical kinetics, Computers and Chemical Engineering, 26, 1567 – 1579, https://doi.org/https://doi.org/10.1016/S0098-1354(02)00128-X, 2002.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmospheric Chemistry and Physics, 7, 3749–3769, https://doi.org/10.5194/acp-7-3749-2007, 2007.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geoscientific Model Development, 3, 43–67, https://doi.org/10.5194/gmd-3-43-2010, 2010.

Fast, J. D., Gustafson Jr., W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., and Peckham, S. E.: Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006721, 2006.

Friedl, M., McIver, D., Hodges, J., Zhang, X., Muchoney, D., Strahler, A., Woodcock, C., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global land cover mapping from MODIS: algorithms and early results, Remote Sensing of Environment, 83, 287 – 302, https://doi.org/https://doi.org/10.1016/S0034-4257(02)00078-0, 2002.

Grell, G. and Baklanov, A.: Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches, Atmospheric Environment, 45, 6845 – 6851, https://doi.org/https://doi.org/10.1016/j.atmosenv.2011.01.017, http://www.sciencedirect.com/science/article/pii/S1352231011000240, modeling of Air Quality Impacts, Forecasting and Interactions with Climate., 2011.

Grell, G., Peckham, S., Schmitz, R., McKeen, S., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled “online” chemistry within the WRF model., Atmos. Environ., 39, 6957–6975, https://doi.org/http://dx.doi.org/10.1016/j.atmosenv.2005.04.027, 2005.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmospheric Chemistry and Physics, 6, 3181–3210, https://hal.archives-ouvertes.fr/hal-00295995, 2006.

Ha, S.-Y. and Snyder, C.: Influence of Surface Observations in Mesoscale Data Assimilation Using an Ensemble Kalman Filter, Mon. Wea. Rev., 142, 1489–1508, https://doi.org/10.1175/MWR-D-13-00108.1, 2014.

Han, Y., van Delst, P., Liu, Q., Weng, F., Yan, B., Treadon, R., and Derber, J.: JCSDA Community Radiative Transfer Model (CRTM) - Version 1, NOAA tech. rep. NESDIS 122, 2006.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes., Mon. Wea. Rev., 134, 2318–2341, https://doi.org/10.1175/MWR3199.1, 2006.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models., J. Geophys. Res., 113, D13 103, https://doi.org/10.1029/2008JD009944, 2008.

Jackson, J. M., Liu, H., Laszlo, I., Kondragunta, S., Remer, L. A., Huang, J., and Huang, H.-C.: Suomi-NPP VIIRS aerosol algorithms and data products, Journal of Geophysical Research: Atmospheres, 118, 12,673–12,689, https://doi.org/10.1002/2013JD020449, 2013.

Jiang, Z., Liu, Z., Wang, T., Schwartz, C. S., Lin, H.-C., and Jiang, F.: Probing into the impact of 3DVAR assimilation of surface PM10 observations over China using process analysis, Journal of Geophysical Research: Atmospheres, 118, 6738–6749, https://doi.org/10.1002/jgrd.50495, 2013.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge Univ. Press, Cambridge, http://cds.cern.ch/record/992314, 2003.

Kim, J., Kim, M., and Choi, M.: Monitoring Aerosol Properties in East Asia from Geostationary Orbit: GOCI, MI and GEMS, pp. 323–333, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-59489-7_15, 2017.

Kleist, D. T., D. F. Parrish, J. C. D., Treadon, R., Wu, W.-S., and Lord, S.: Introduction of the GSI into the NCEP Global Data Assimilation System, Wea. Forecasting, 24, 1691–1705, 2009.

Kong, X. and coauthors: Analysis of meteorology-chemistry interactions during air pollution episodes using online coupled models within AQMEII phase-2, Atmospheric Environment, 115, 527 – 540, https://doi.org/https://doi.org/10.1016/j.atmosenv.2014.09.020, 2015.

Kumar, R., Delle Monache, L., Bresch, J., Saide, P. E., Tang, Y., Liu, Z., da Silva, A. M., Alessandrini, S., Pfister, G., Edwards, D., Lee, P., and Djalalova, I.: Toward Improving Short-Term Predictions of Fine Particulate Matter Over the United States Via Assimilation of Satellite Aerosol Optical Depth Retrievals, Journal of Geophysical Research: Atmospheres, 124, 2753–2773, https://doi.org/10.1029/2018JD029009, 2019.

Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., and Koutrakis, P.: A novel calibration approach of MODIS AOD data to predict $PM_{2.5}$ concentrations, Atmospheric Chemistry and Physics, 11, 7991–8002, https://doi.org/10.5194/acp-11-7991-2011, 2011.

Lee, J., Kim, J., Song, C. H., Ryu, J.-H., Ahn, Y.-H., and Song, C.: Algorithm for retrieval of aerosol optical properties over the ocean from the Geostationary Ocean Color Imager, Remote Sensing of Environment, 114, 1077 – 1088, https://doi.org/https://doi.org/10.1016/j.rse.2009.12.021, 2010.

LeGrand, S. L., Polashenski, C., Letcher, T. W., Creighton, G. A., Peckham, S. E., and Cetola, J. D.: The AFWA dust emission scheme for the GOCART aerosol model in WRF-Chem v3.8.1, Geoscientific Model Development, 12, 131–166, https://doi.org/10.5194/gmd-12-131-2019, https://www.geosci-model-dev.net/12/131/2019/, 2019.

Lin, Y.-L., Farley, R. D., and Orville, H. D.: Bulk parameterization of the snow field in a cloud model, Journal of Climate and Applied Meteorology, 22, 1065–1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, 1983.

Liu, Q. and Weng, F.: Advanced Doubling-Adding Method for Radiative Transfer in Planetary Atmospheres, Journal of the Atmospheric Sciences, 63, 3459–3465, https://doi.org/10.1175/JAS3808.1, 2006.

Liu, Z., Liu, Q., Lin, H.-C., Schwartz, C. S., Lee, Y.-H., and Wang, T.: Three-dimensional variational assimilation of MODIS aerosol optical depth: Implementation and application to a dust storm over East Asia, J. Geophys. Res., 116, https://doi.org/10.1029/2011JD016159, 2011.

McKeen, S., Grell, G., Peckham, S., Wilczak, J., Djalalova, I., Hsie, E.-Y., Frost, G., Peischl, J., Schwarz, J., Spackman, R., Holloway, J., de Gouw, J., Warneke, C., Gong, W., Bouchet, V., Gaudreault, S., Racine, J., McHenry, J., McQueen, J., Lee, P., Tang, Y., Carmichael, G. R., and Mathur, R.: An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field study, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011697, 2009.

Miyazaki, K., Sekiya, T., Fu, D., Bowman, K. W., Kulawik, S. S., Sudo, K., Walker, T., Kanaya, Y., Takigawa, M., Ogochi, K., Eskes, H., Boersma, K. F., Thompson, A. M., Gaubert, B., Barre, J., and Emmons, L. K.: Balance of Emission and Dynamical Controls on Ozone During the Korea-United States Air Quality Campaign From Multiconstituent Satellite Data Assimilation, Journal of Geophysical Research: Atmospheres, 124, 387–413, https://doi.org/10.1029/2018JD028912, 2019.

Ochotta, T., Gebhardt, C., Saupe, D., and Wergen, W.: Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods, Quarterly Journal of the Royal Meteorological Society, 131, 3427–3437, https://doi.org/10.1256/qj.05.94, 2005.

Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D.: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting?Chemistry model and Grid-point Statistical Interpolation, Quarterly Journal of the Royal Meteorological Society, 136, 2013–2024, https://doi.org/10.1002/qj.700, 2010.

Pang, J., Liu, Z., Wang, X., Bresch, J., Ban, J., Chen, D., and Kim, J.: Assimilating AOD retrievals from GOCI and VIIRS to forecast surface PM2.5 episodes over Eastern China, Atmospheric Environment, 179, 288 – 304, https://doi.org/https://doi.org/10.1016/j.atmosenv.2018.02.011, 2018.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center?s spectral statistical-interpolation analysis system, Mon. Wea. Rev., 120, 1747?1763, 1992.

Pfister, G. G., Avise, J., Wiedinmyer, C., Edwards, D. P., Emmons, L. K., Diskin, G. D., Podolske, J., and Wisthaler, A.: CO source contribution analysis for California during ARCTAS-CARB, Atmospheric Chemistry and Physics, 11, 7515–7532, https://doi.org/10.5194/acp-11-7515-2011, 2011.

Reale, O., McGrath-Spangler, E. L., McCarty, W., Holdaway, D., and Gelaro, R.: Impact of Adaptively Thinned AIRS Cloud-Cleared Radiances on Tropical Cyclone Representation in a Global Data Assimilation and Forecast System, Weather and Forecasting, 33, 909–931, https://doi.org/10.1175/WAF-D-17-0175.1, 2018.

Remer, L. A., Kaufman, Y. J., TanrÃ©, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, Journal of the Atmospheric Sciences, 62, 947–973, https://doi.org/10.1175/JAS3385.1, 2005.

Rienecker, M. and Coauthors: The GEOS-5 Data Assimilation System: Documentation of versions 5.0.1, 5.1.0, and 5.2.0. NASA/TM-2008-104606, NASA Tech. Rep. Series on Global Modeling and Data Assimilation, Tech. Note NASA/TM-2008-104606, NASA, https://ntrs.nasa.gov/search.jsp?R=20120011955, 2008.

Saide, P. E., Kim, J., Song, C. H., Choi, M., Cheng, Y., and Carmichael, G. R.: Assimilation of next generation geostationary aerosol optical depth retrievals to improve air quality simulations, Geophysical Research Letters, 41, 9188–9196, https://doi.org/10.1002/2014GL062089, 2014.

Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, Atmospheric Chemistry and Physics, 6, 187–195, https://doi.org/10.5194/acp-6-187-2006, https://www.atmos-chem-phys.net/6/187/2006/, 2006.

Schwartz, C. S., Liu, Z., Lin, H.-C., and McKeen, S. A.: Simultaneous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2011JD017383, 2012.

Tie, X., Madronich, S., Walters, S., Zhang, R., Rasch, P., and Collins, W.: Effect of clouds on photolysis and oxidants in the troposphere, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2003JD003659, 2003.

Volkamer, R., Jimenez, J. L., San Martini, F., Dzepina, K., Zhang, Q., Salcedo, D., Molina, L. T., Worsnop, D. R., and Molina, M. J.: Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL026899, 2006.

Wang, M., Ahn, J.-H., Jiang, L., Shi, W., Son, S., Park, Y.-J., and Ryu, J.-H.: Ocean color products from the Korean Geostationary Ocean Color Imager (GOCI), Optics Express, 21, 3835–3849, https://doi.org/10.1364/OE.21.003835, 2013.

Wesely, M.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmospheric Environment, 23, 1293 – 1304, https://doi.org/https://doi.org/10.1016/0004-6981(89)90153-4, 1989.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625–641, https://doi.org/10.5194/gmd-4-625-2011, 2011.

Wu, W.-S., Purser, R. J., and Parrish, D. F.: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances, Monthly Weather Review, 130, 2905–2916, https://doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2, 2002.

Xiao, Q., Zhang, H., Choi, M., Li, S., Kondragunta, S., Kim, J., Holben, B., Levy, R. C., and Liu, Y.: Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia, Atmospheric Chemistry and Physics, 16, 1255–1269, https://doi.org/10.5194/acp-16-1255-2016, 2016.

Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD008782, 2008.

Zhang, J. and Reid, J. S.: MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006898, 2006.

Zhang, J. and Reid, J. S.: A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products, Atmospheric Chemistry and Physics, 10, 10 949–10 963, https://doi.org/10.5194/acp-10-10949-2010, 2010.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmospheric Chemistry and Physics, 18, 14 095–14 111, https://doi.org/10.5194/acp-18-14095-2018, 2018.

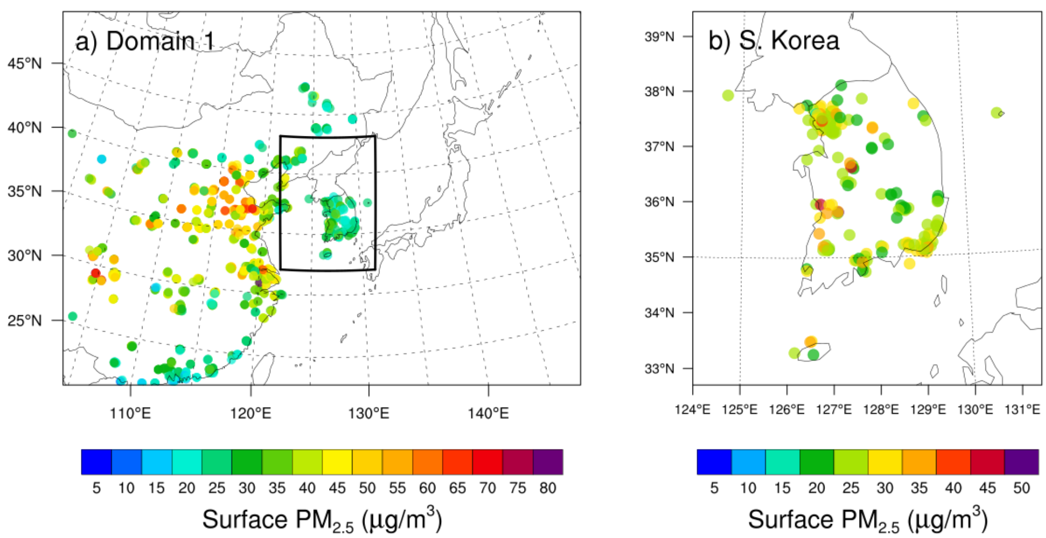

Figure 1. Surface observation network with 960 Chinese stations and 361 Korean stations in domain 1 (a) and zoomed in over South Korea in (b). A black box in a) indicates domain 2 over the Korean peninsula. Dots indicate surface $PM_{2.5}$ observations averaged over the month of May 2016.

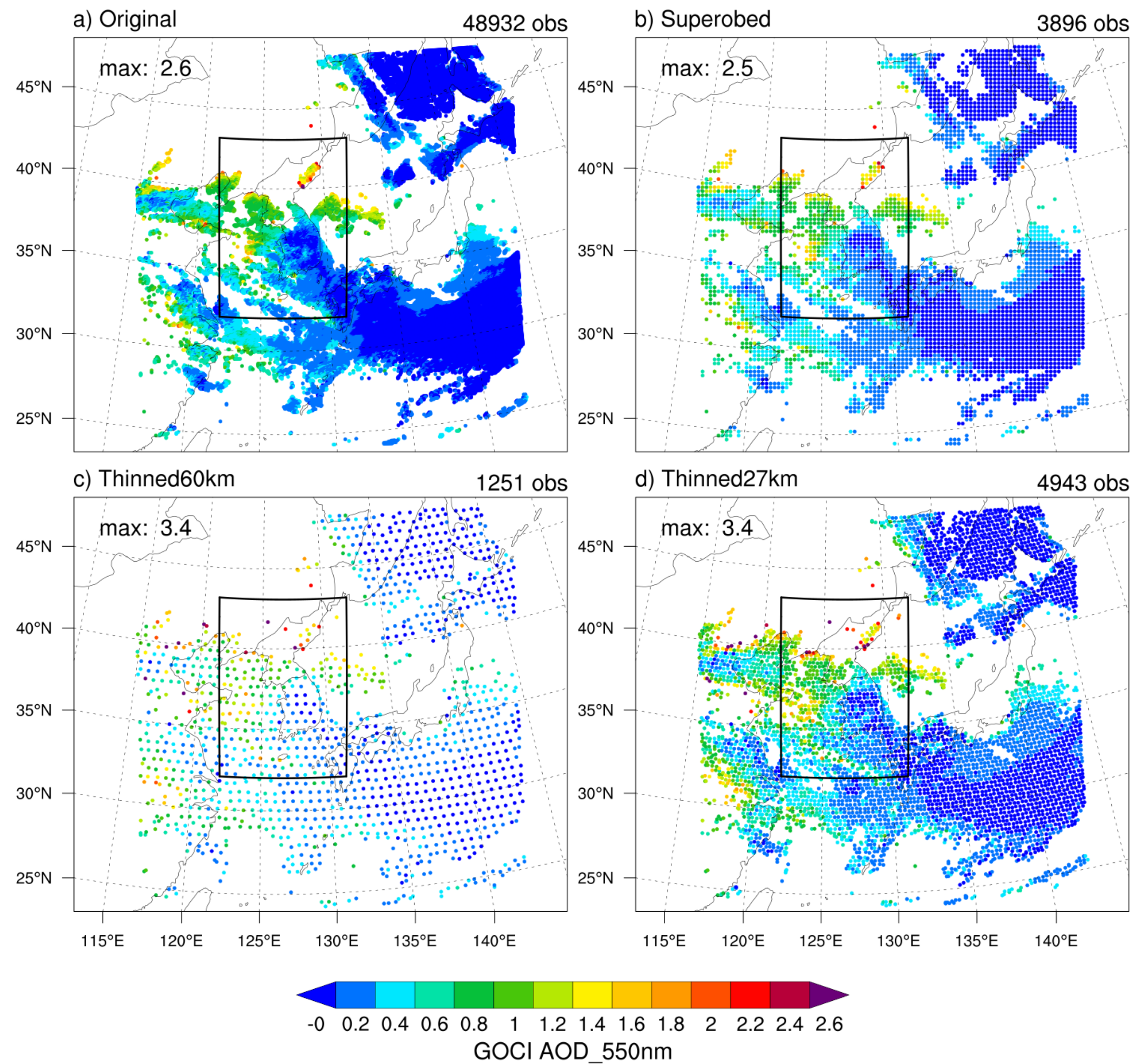

Figure 2. Horizontal distribution of GOCI AOD at 550 nm retrieved at 2016-05-01_06:00:00 UTC in a) the original Level II data at 6 km resolution b) the preprocessed at 27 km resolution before GSI, and the data thinned over c) 60 km and d) 27 km resolution during the GSI analysis, respectively. A total number of observations available for the GSI analysis is shown in the upper right corner of each panel and the maximum value in the upper left corner of each map. Domain 2 is marked as a black box in each panel.

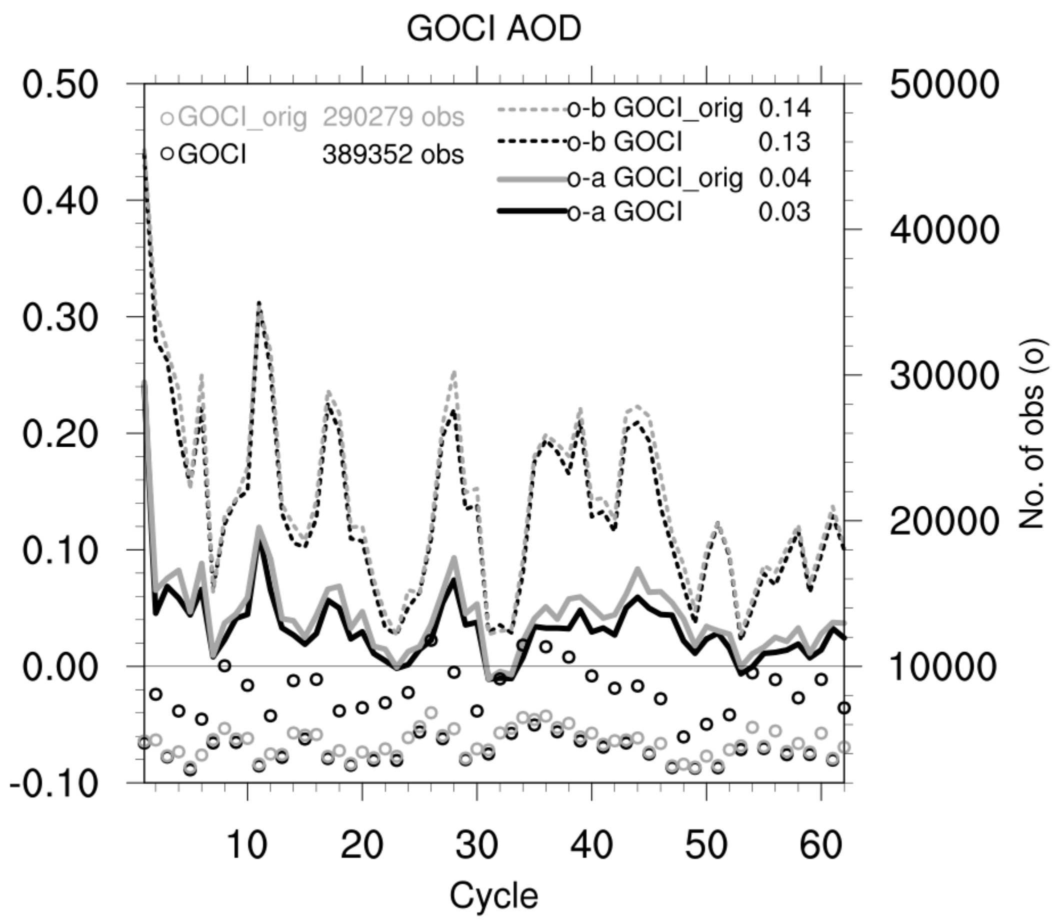

Figure 3. Time series of observation-minus-analysis (o-a; solid lines) and observation-minus-background (o-b; dotted) with respect to GOCI AOD retrievals at 550 nm for two cycling experiments over domain 1. The "GOCI_orig" experiment assimilates the original data thinned over 27-km mesh (in gray) while the "GOCI" experiment assimilates GOCI retrievals averaged over 27-km grids in domain 1 (black). Cycle-mean values are displayed next to each component. Total number of observations assimilated in each experiment at each cycle is also plotted as “o” sign on the right y-axis ranging from 2,000 to 12,000.

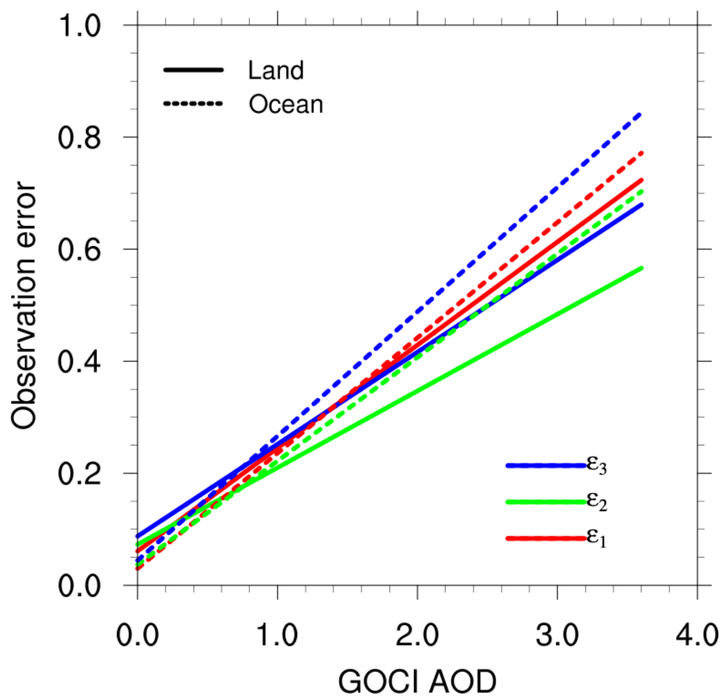

Figure 4. Three different types of observation errors (ϵ) applied to GOCI AOD retrievals over land (solid line) and ocean (dashed line), respectively. The first two errors (ϵ_1 and ϵ_2) are described in equations (3) - (6) and the third error (ϵ_3) increases ϵ_2 by 20% everywhere.

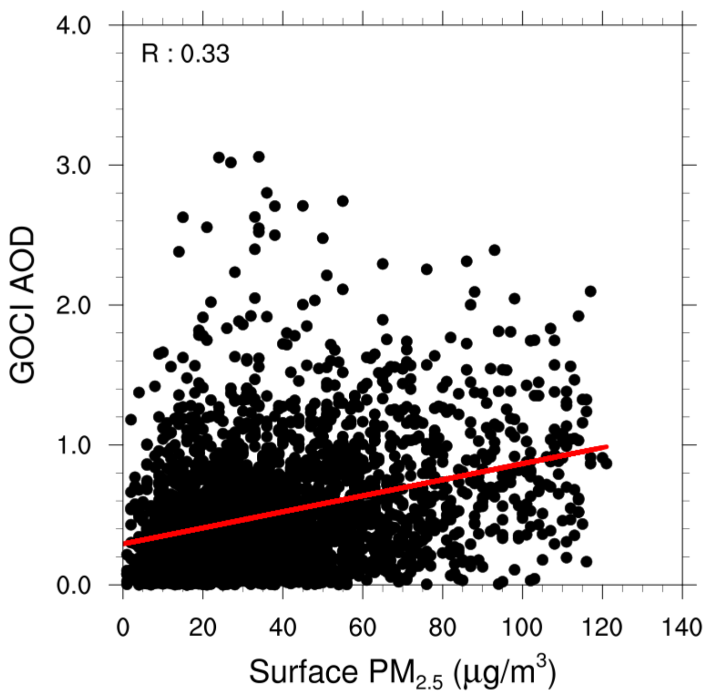

Figure 5. Scatter plots of GOCI AOD retrievals versus ground $PM_{2.5}$ observations collocated in domain 1 for the month of May 2016. The value of R is the correlation coefficient between the two observation types based on the linear regression shown as the red line.

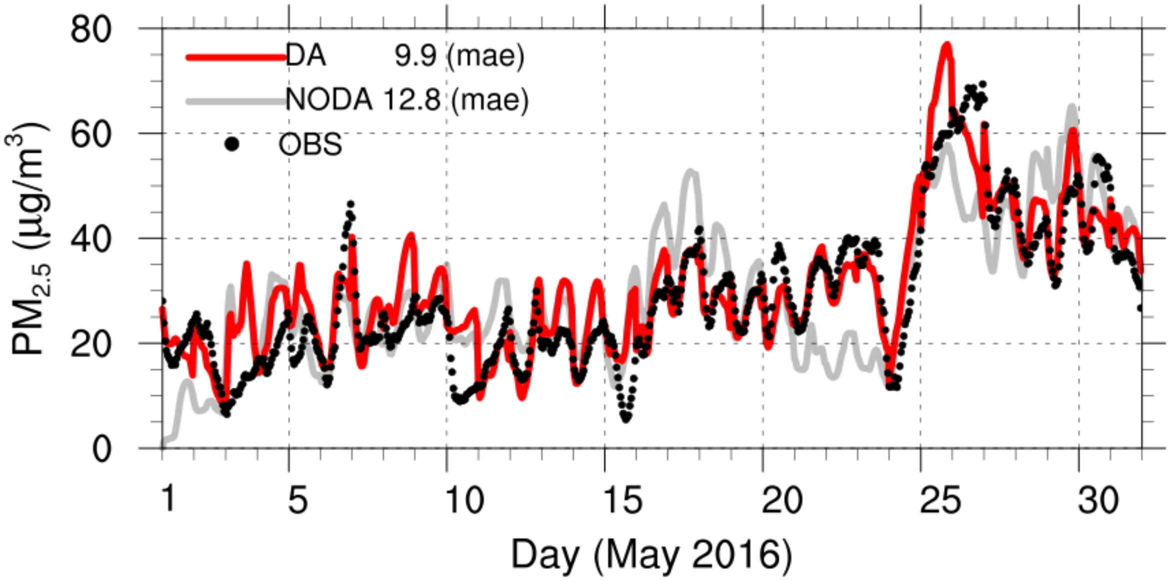

Figure 6. Time series of surface $PM_{2.5}$ simulated with (DA; red) and without assimilation (NODA; gray) in domain 2, representing hourly 0-23-h forecasts from 00Z every day, as averages over 361 stations over South Korea. Corresponding observations are marked as black dots. The mean absolute error (mae; $|o - f|$) averaged over the entire period is shown for each experiment. Here, "DA" refers to the "ALL" experiment.

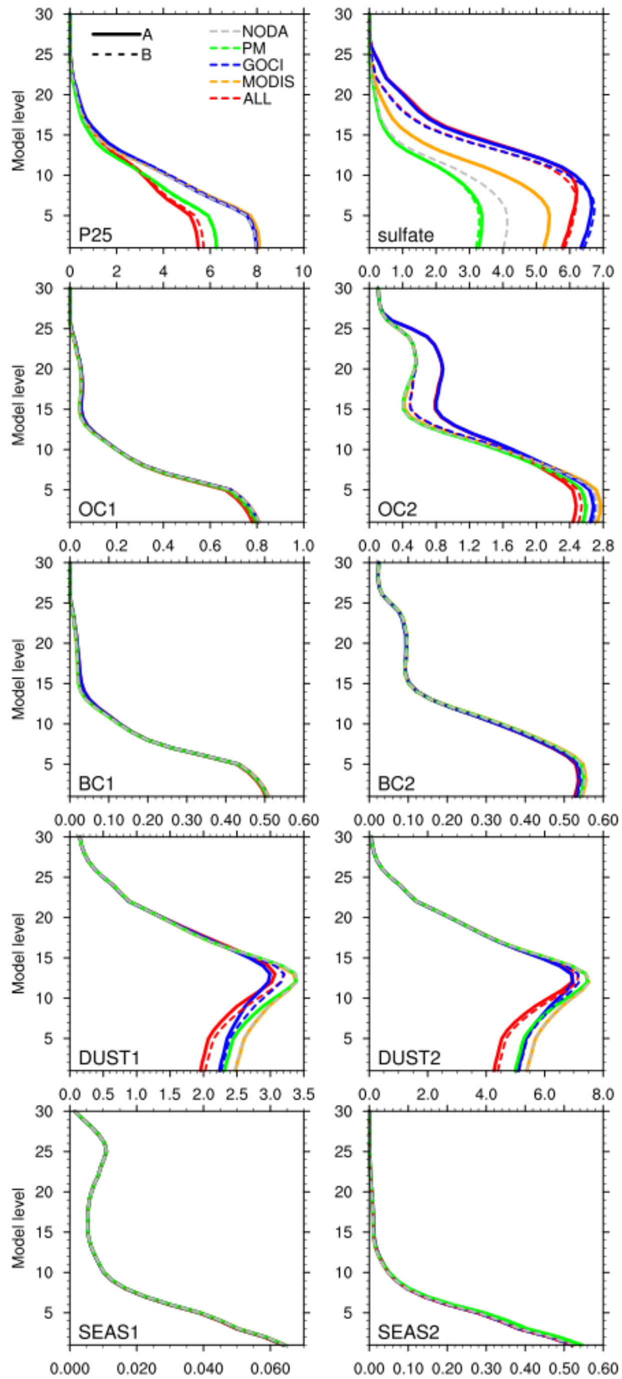

Figure 7. Vertical profile of 10 GOCART aerosol variables composed of $PM_{2.5}$ - unspeciated aerosol contributions to $PM_{2.5}$ (P25), sulfate, OC1 and OC2 (BC1 and BC2) as hydrophobic and hydrophilic organic (black) carbon, respectively, DUST1 and DUST2 (SEAS1 and SEAS2) as dust (sea salt) aerosols in the smallest and 2nd smallest size bins. All the variables shown are mixing ratios in the unit of μg/kg. Different experiments are depicted in different colors, as averaged over domain 2 for the period of May 4 - 31, 2016. Analysis ("A") is drawn as solid line while background (e.g. 6-h forecast; "B") as dashed line.

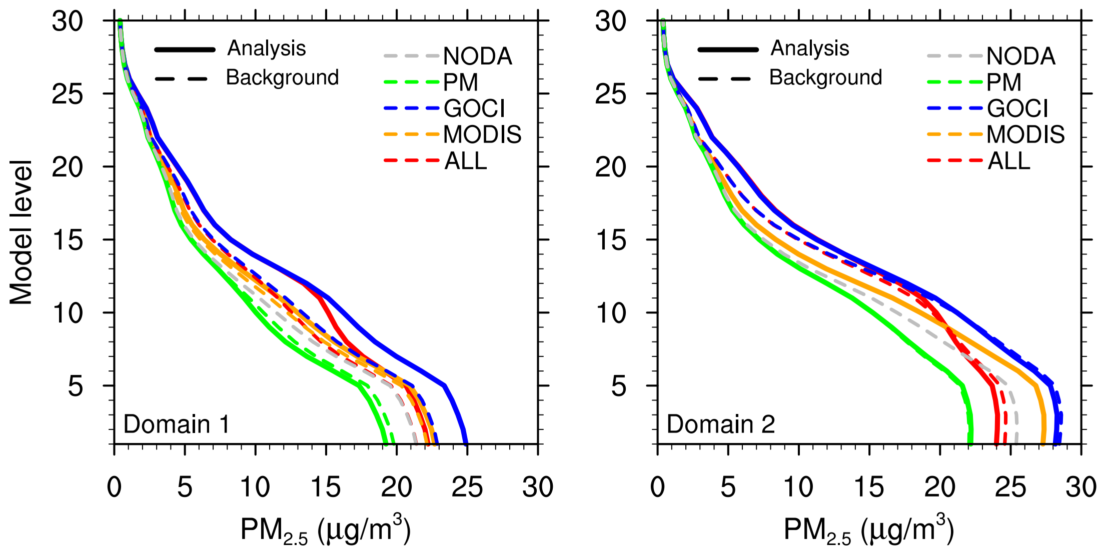

Figure 8. Same as Figure 7, except for $PM_{2.5}$ in both domains.

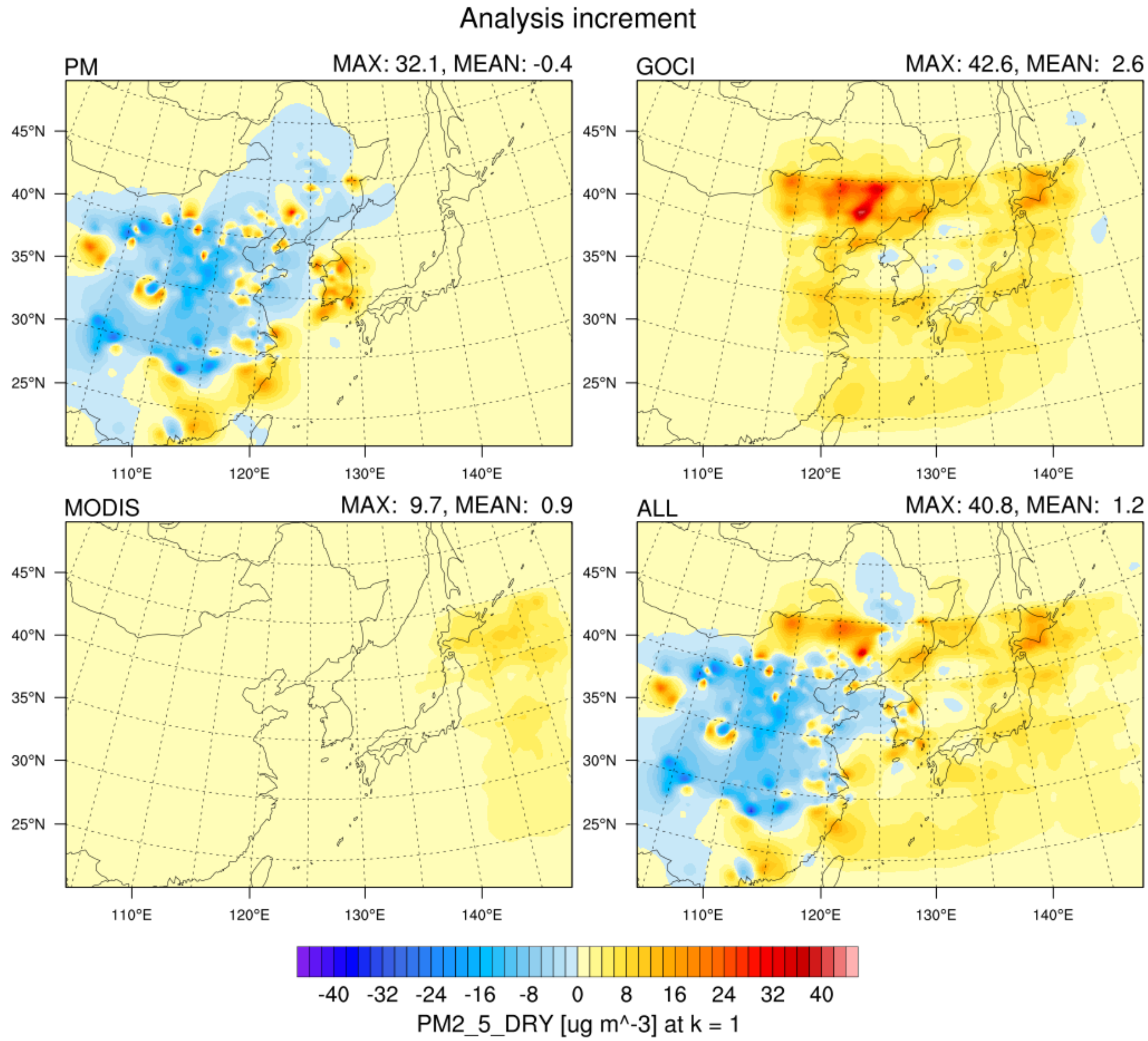

Figure 9. Horizontal distribution of analysis increments (analysis-minus-background) in PM2_5_DRY, the model variable corresponding to $PM_{2.5}$, at the lowest level in domain 1, averaged over the period of May 4 - 31, 2016. Maximum and mean values of the domain in each experiment are shown in the upper right corner of each panel.

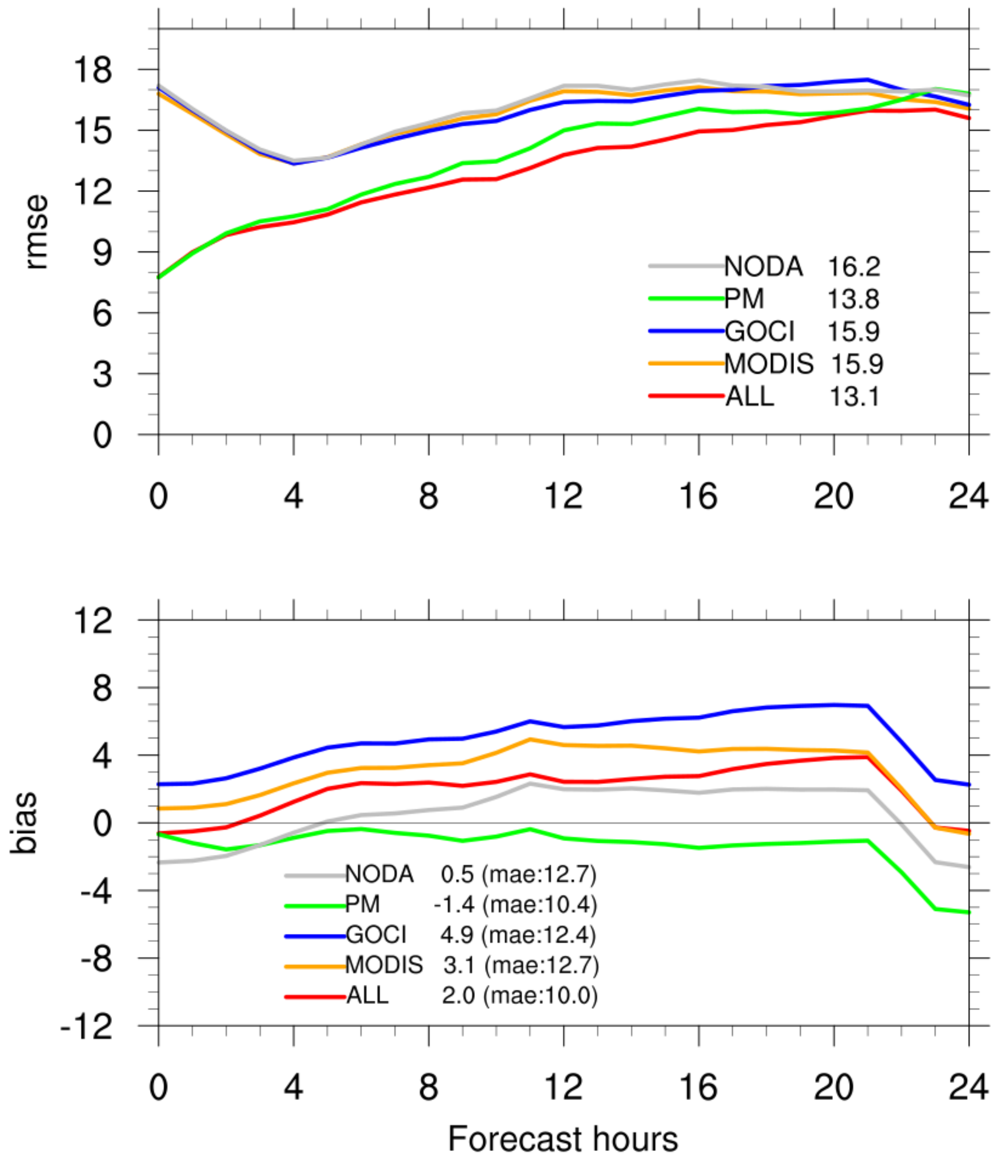

Figure 10. Time series of root-mean-square-error (rmse; upper panel) and bias (lower panel) of the hourly forecasts from the 00 Z initialization for May 4 - 31, 2016. Different experiments in domain 2 are verified against surface $PM_{2.5}$ observations from 361 stations in South Korea. An average of 0-24 h forecast errors is shown next to each experiment name. The mean absolute error (mae) over the 24-h forecasts is also shown in the lower panel.

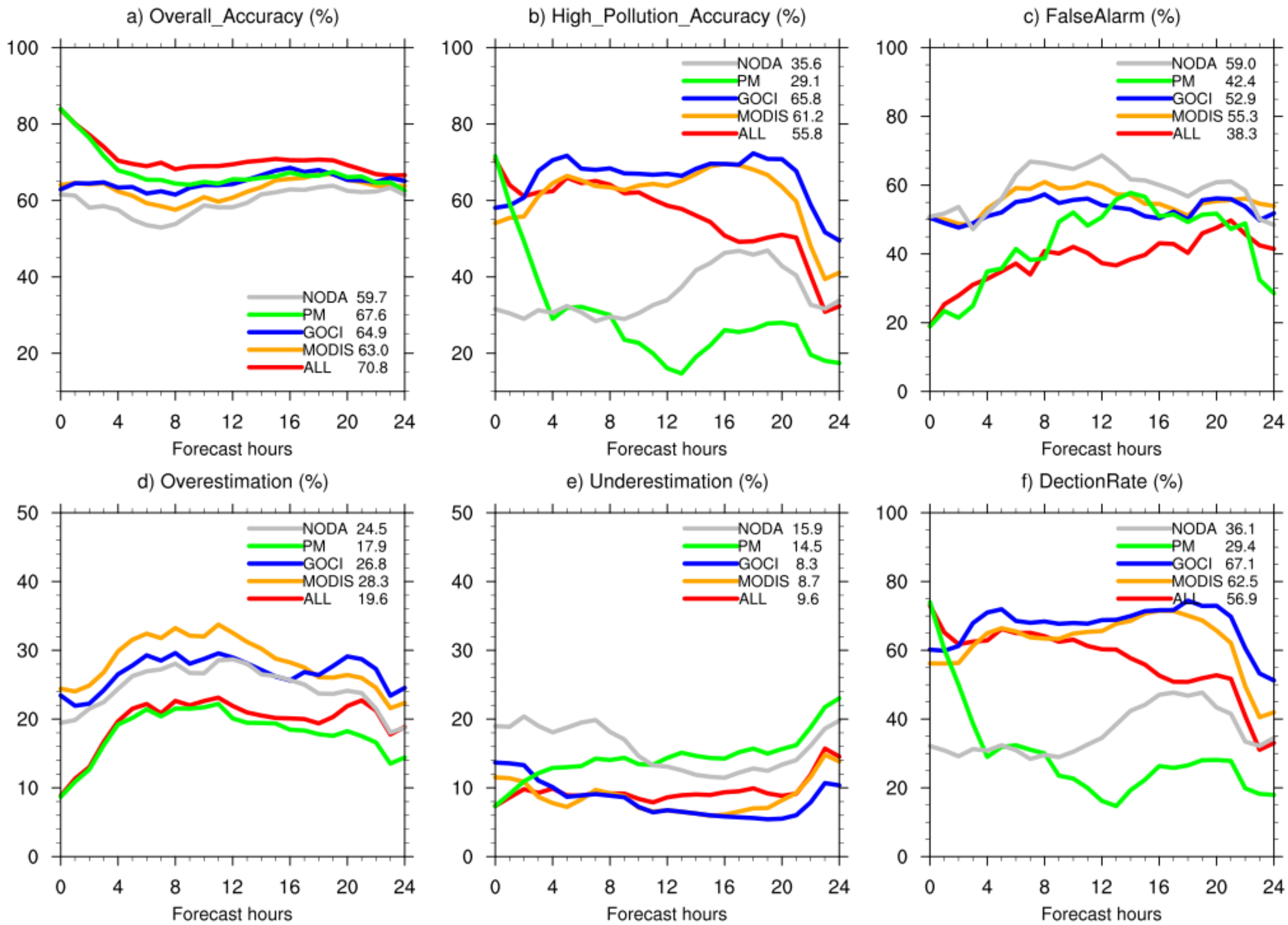

Figure 11. Time series of forecast accuracy (%) of the hourly forecasts from the 00 Z initialization for May 4 - 31, 2016 in domain 2 for categorized events based on hourly surface $PM_{2.5}$ concentrations, as defined in Tables 2 and 3.

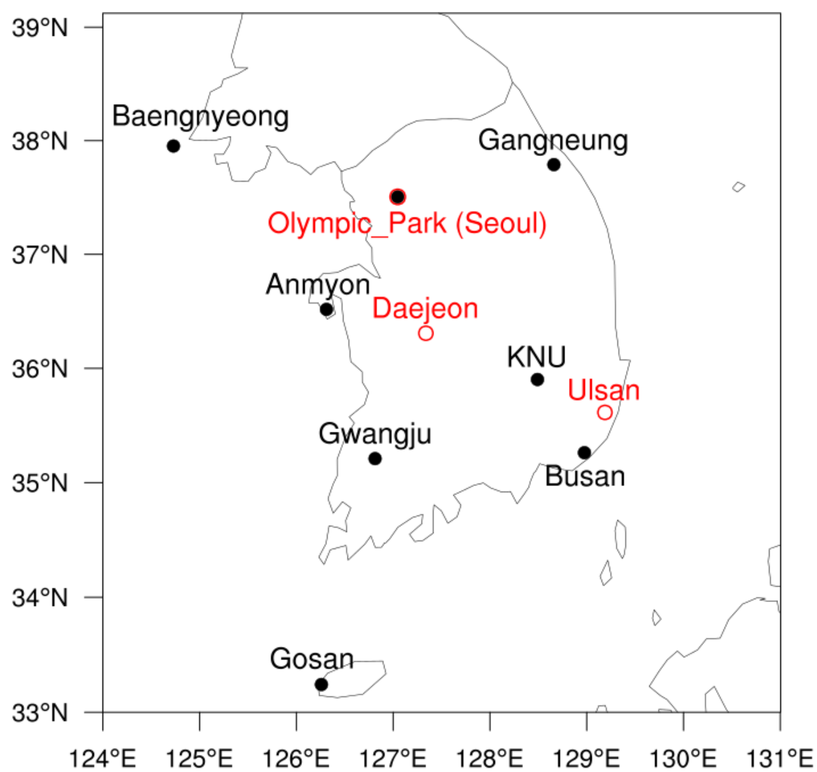

Figure 12. Map of AERONET sites (black dots) used for verification shown in Fig. 13. Three red open dots are the stations operated by NIER to measure surface $PM_{2.5}$ concentrations during the KORUS-AQ field campaign, which are used in the verification illustrated in Fig. 14.

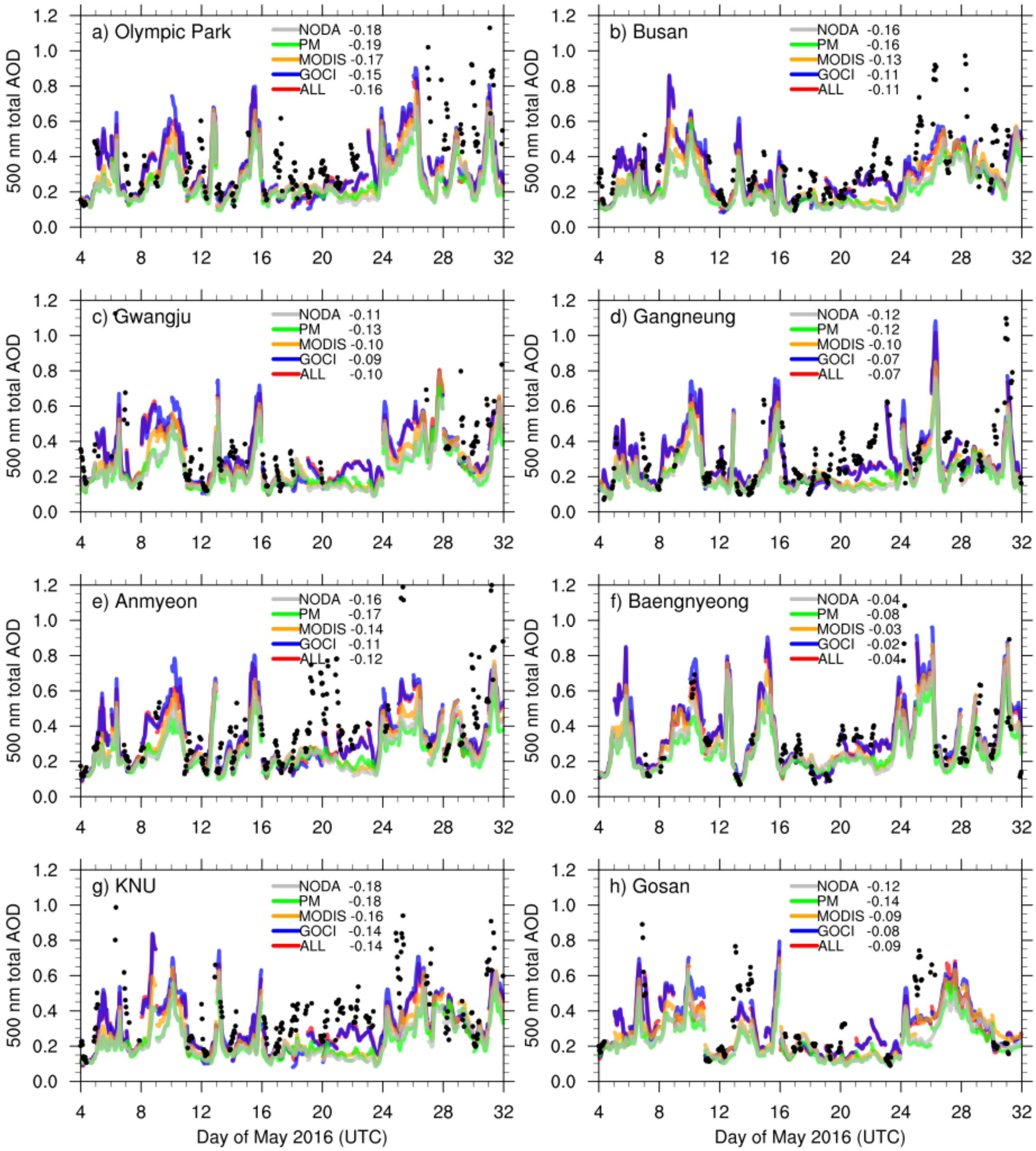

Figure 13. Hourly time series of total AOD at 500 nm from 0000 UTC 04 May to 2300 UTC 31 May at 8 different AERONET sites. Model values in different colors represent output every hour beginning at the initial time and ending at the 23rd hour of integration patched together for each 0000 UTC forecast. The bias (as (f-o)'s) averaged over the entire period is shown next to each experiment name. AERONET observations represent hourly averages as black dots.

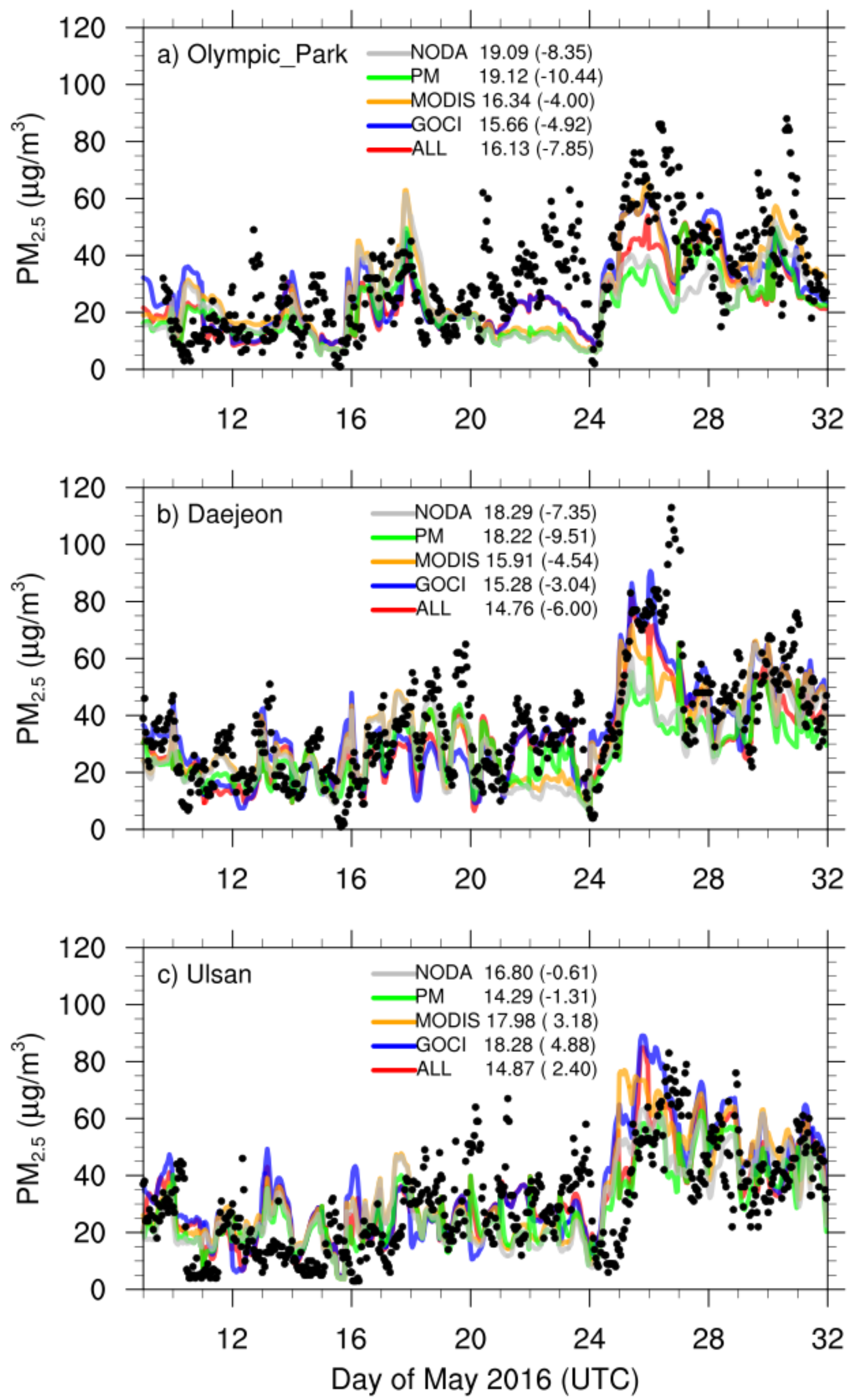

Figure 14. Same as Fig. 13, but for surface $PM_{2.5}$ concentrations from 0000 UTC 09 May to 2300 UTC 31 May at a) Olympic Park in Seoul b) Daejeon and c) Ulsan. The sites are marked as red open dots in Fig. 12. The rmse over the whole period is written next to each experiment name, along with the mean bias (as (f-o)'s) in the parenthesis.

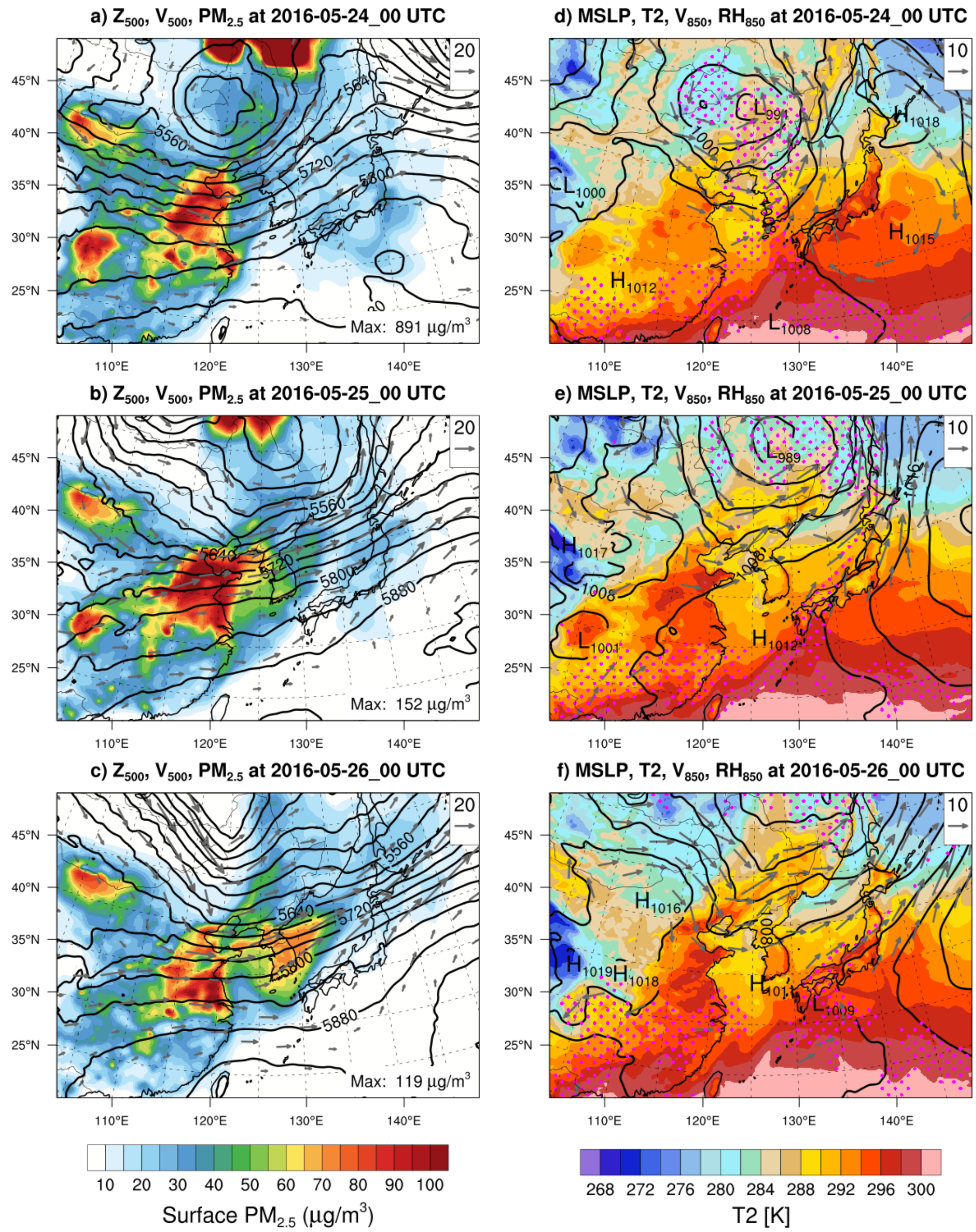

Figure 15. The GSI 3DVAR analyses at 27-km resolution in domain 1 in the "ALL" experiment for three days from 24 to 26 May 2016 at 00 UTC (top to bottom). In the left panel, the horizontal distribution of surface $PM_{2.5}$ ([μg/m^3], filled), geopotential height (contours every 40 m) and horizontal winds ([m/s] in gray vectors) at 500 hPa illustrates that the long-range transport of air pollution from China causes the heavy pollution over South Korea. In the right panel, mean sea level pressure (contours every 4 hPa), 2-m temperature ([K], filled), relative humidity (> 90% in pink dots) and horizontal winds ([m/s] in gray vectors) at 850 hPa represent the weather system in the low troposphere at the same time.

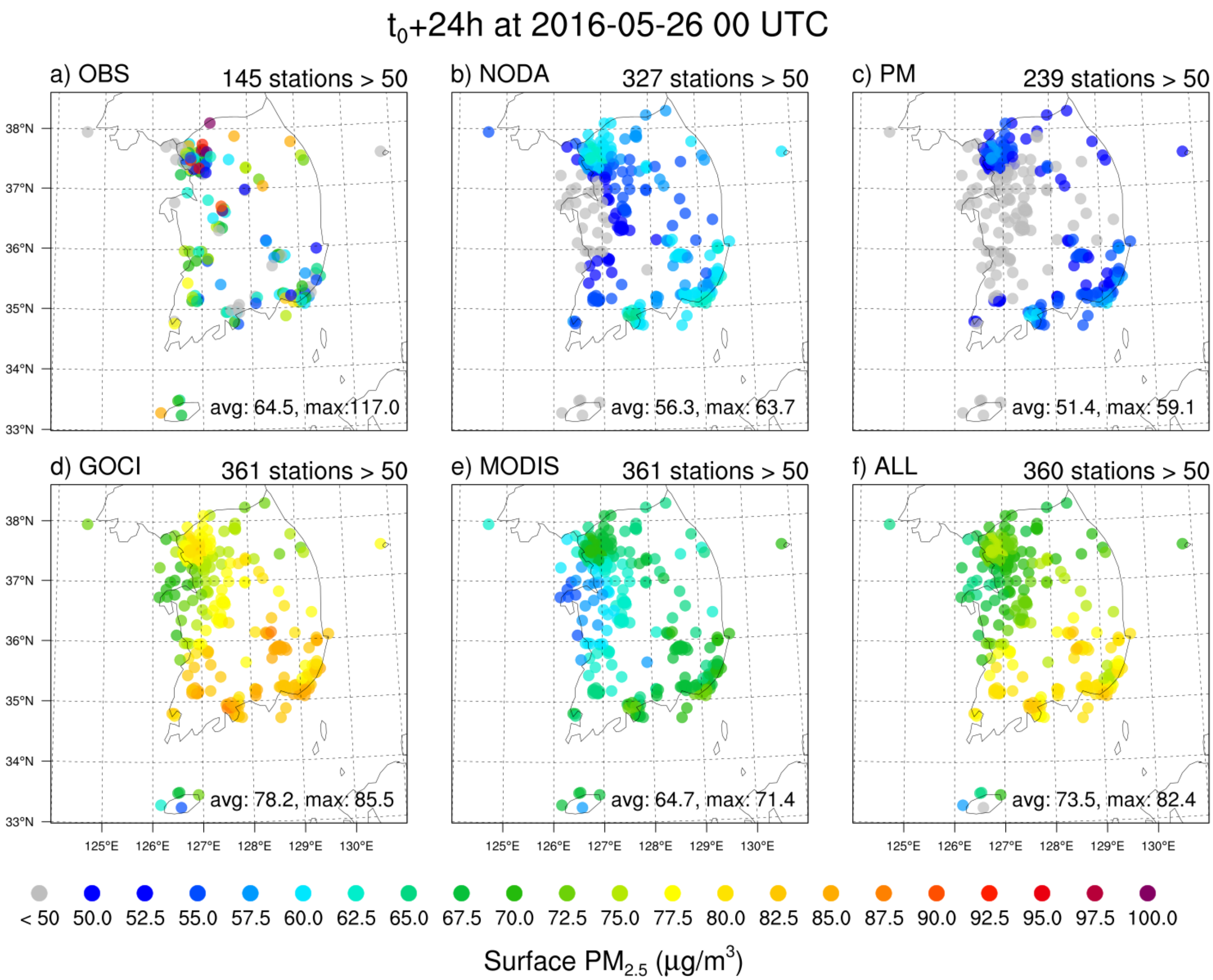

Figure 16. Horizontal distribution of 24-h forecast in 9-km simulations in $PM_{2.5}$ at the lowest level in each experiment compared to a) observations from 361 stations in South Korea valid at 00 UTC 26 May 2016.

Table 1. Physical and chemical parameterizations used in the experiments.

Physical processes	Parameterization schemes
Aerosol chemistry	GOCART
Gas-phase chemistry	MOZART-4
Photolysis	Fast-TUV
Cloud microphysics	Lin
Cumulus	Grell 3D ensemble
Longwave radiation	RRTMG
Shortwave radiation	Goddard
PBL	YSU
Surface layer	Monin-Obukhov
Land surface	Noah

Table 2. Air quality index values

Concentration (μg/m^3, hourly)	Good	Moderate	Unhealthy	Very Unhealthy
$PM_{2.5}$	0-15	16-50	51-100	> 100

Table 3. Categorical forecasts for different air pollution events

Category		Forecast			
		Good	Moderate	Unhealthy	Very Unhealthy
Observation	Good	a1	b1	c1	d1
	Moderate	a2	b2	c2	d2
	Unhealthy	a3	b3	c3	d3
	Very Unhealthy	a4	b4	c4	d4

Table 4. Forecast error in total AOD at 500 nm verified against AERONET sites, computed over 0-23 h forecasts from 00Z analysis for May 4 - 31.

	rmse					bias				
	NODA	PM	MODIS	GOCI	ALL	NODA	PM	MODIS	GOCI	All
OlympicPark	0.26	0.27	0.25	0.23	0.24	-0.18	-0.19	-0.17	-0.15	-0.16
Busan	0.22	0.22	0.19	0.17	0.18	-0.16	-0.16	-0.13	-0.11	-0.11
Gwangju	0.18	0.19	0.17	0.16	0.16	-0.11	-0.13	-0.1	-0.09	-0.1
Gangneung	0.18	0.18	0.17	0.13	0.13	-0.12	-0.12	-0.1	-0.07	-0.07
Anmyeon	0.26	0.27	0.24	0.22	0.22	-0.16	-0.17	-0.14	-0.11	-0.12
Baengnyeong	0.15	0.15	0.14	0.13	0.13	-0.04	-0.08	-0.03	-0.02	-0.04
KNU	0.24	0.25	0.22	0.2	0.21	-0.18	-0.18	-0.16	-0.14	-0.14
Gosan	0.21	0.21	0.18	0.15	0.15	-0.12	-0.14	-0.09	-0.08	-0.09
Seoul_SNU	0.22	0.23	0.21	0.2	0.2	-0.14	-0.16	-0.13	-0.12	-0.12
NIER	0.21	0.21	0.2	0.19	0.19	-0.13	-0.15	-0.12	-0.11	-0.12
YSU	0.22	0.23	0.21	0.2	0.2	-0.15	-0.17	-0.14	-0.13	-0.13
Daegwallyeong	0.13	0.12	0.12	0.09	0.09	-0.08	-0.07	-0.06	-0.03	-0.03
Iksan	0.35	0.35	0.33	0.28	0.29	-0.26	-0.26	-0.24	-0.2	-0.2
Ulsan	0.21	0.22	0.19	0.17	0.17	-0.17	-0.17	-0.14	-0.12	-0.13
Mokpo	0.21	0.22	0.19	0.18	0.18	-0.13	-0.14	-0.11	-0.1	-0.1
Taehwa	0.27	0.28	0.25	0.23	0.24	-0.17	-0.19	-0.16	-0.14	-0.15