Evaluation of NU-WRF Performance on Air Quality Simulation under Various Model Resolutions – An Investigation within Framework of MICS-Asia Phase III
By Z. Tao et al.

The authors greatly appreciate the reviewers' insightful and constructive comments. We have made substantial revisions in an effort to address the key comments. The following lists our responses (plain text) to the reviewer's comments (bold).

Response to Reviewer #1

1. The simulations conducted over the three nested domains are used in this study for comparisons and evaluations of meteorological and air quality simulations among three different grid-spacings. The way of nesting used in the simulations may have a large impact on the sensitivity of simulation results to grid spacing. Which nesting way is used in the simulations? One-way nesting or two-way nesting? If two-way nesting is selected in the simulations which I doubt according to results presented in the manuscript (e.g., L175-176), how your conclusion or findings can be affected by this selection?

In this study, we chose the one-way nesting approach. If a two-way nesting approach is used, the meteorological fields for the study domain (innermost domain) from 3 horizontal resolutions would be very similar to each other, which would obscure the conclusion of the effect of model resolution on air quality. In the one-way nesting approach, the meteorological fields of the mother domains are independent on those of the respective nested domains, thus representing a cleaner signal of the grid resolution effect on air quality than that of the two-way nesting approach. As illustrated in Figure 6 of the manuscript, surface wind, temperature, PBLH, ground-level short wave radiation, and cloud water path out of each grid display noticeable differences. Due to the resource constrain, this investigation chose the one-way nesting approach, in which the lateral boundary conditions (LBCs) for each grid modeling were not identical. A "cleaner" approach would be to apply the identical LBCs out of the global model to drive NU-WRF under various grid resolutions to evaluate their impacts on air quality. However, we do not anticipate that such "cleaner" approach would remarkably alter the results based on the one-way nesting approach. In the manuscript, we added a statement in section 2 that the one-way nesting was utilized.

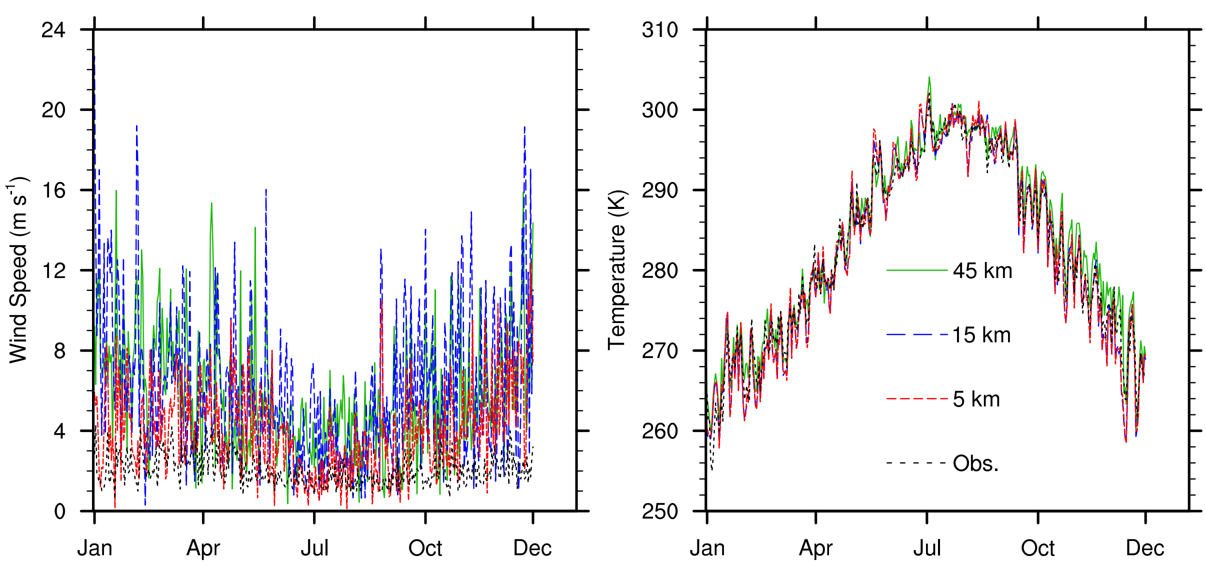
"…A one-way nesting approach was applied so that the values of the mother domains were independent on those of the respective nested domains…"

2. The impact of grid-spacing on met or air quality simulations is highly dependent on terrain complexity of the study region. Given the relatively flat terrain of North China Plain (NCP), the impact of the grid-spacing on the simulations over land cannot be as large as expected. We believe that complex terrain and the situations with a large surface-cover contrast such as coastal regions do require high-resolution simulations. 15-km horizontal resolution that the authors suggest for the MICS-Asia study is definitely not enough to resolve the detailed local wind structures such as land-sea breeze or lake-breeze over the coastal regions which many large cities (e.g., Shanghai and Hong Kong) are located and air pollution is a big concern. It would be very helpful if the authors can shed little bit

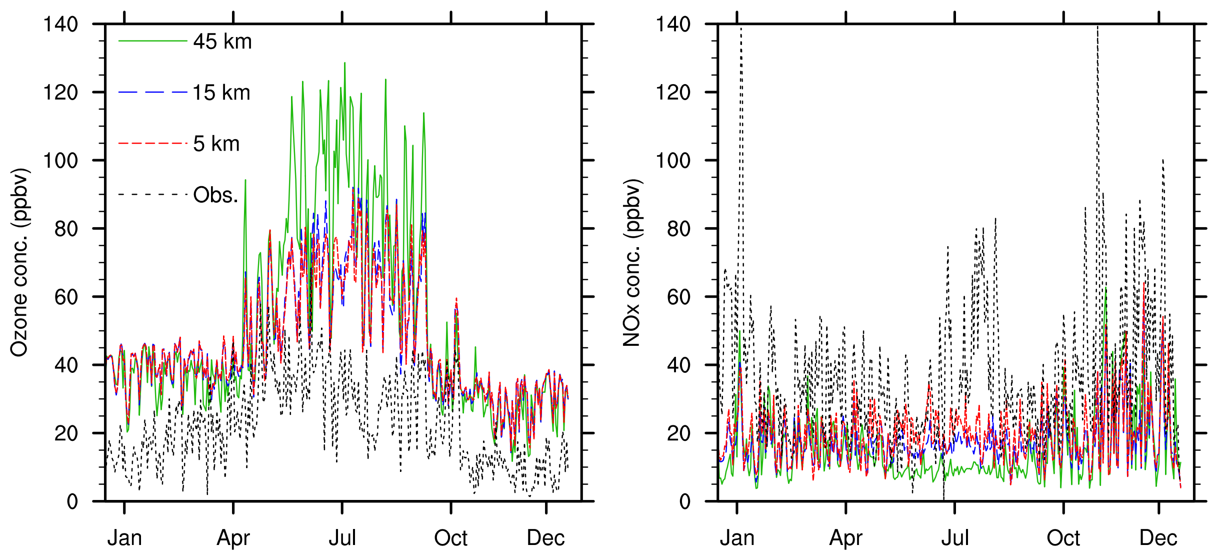
light on discussion of model performance with three grid-spacings at those sites along or near coastal regions.

In this investigation, we have found that the 5-km resolution modeling provided the best results of wind and surface pollutant levels, especially in polluted conditions that were the most relevant to air quality regulation (e.g., compliance of national air quality standards), measured by bias and RMSE. However, the improvement of model performance on air quality was not linear with the resolution increase. For example, the accuracy of modeled surface O_3 out of the 15-km grid was greatly improved over the one from the 45-km grid. Further increase of grid resolution to 5-km, however, showed diminished impact on model performance improvement on O_3 prediction for the study region. In addition, the cost in terms of cpu hours and disk space usage increased dramatically when adopting the 5-km grid, which would be a big hurdle for the inter-model comparison studies such as MICS-Asia that relied on community contributions to model Asia air quality over a relatively long time period. Considering all these factors, we suggest a 15-km resolution grid for future MICS-Asia modeling activity to achieve both accuracy and efficiency.

Of course, the choice of grid resolution also depends on the problems to be solved, such as air quality over coastal areas which show sharp contrasts of surface roughness, albedo, and thermal characteristics. In this investigation, QHD site locates approximately 5 km from the ocean and is subject to sea breeze effects. The temporal profiles of surface wind speed and temperature from the observation and model results out of 3 grids for QHD are shown in the following figure. The results indicated that the choice of grid resolution had large impacts on model simulations at this coastal site. The selection of the 5-km grid reduced biases of both surface temperature and wind speed. The biases of temperature reduced from 1.22 K (45-km) to -0.42 K (15-km), and further down to -0.31 K when the 5-km grid was applied. The biases of surface wind speed for the 45-km, 15-km, and 5-km grids were 3.72, 4.19, and 1.95 m s^{-1}, respectively. Since there were no hourly wind data available to this study, the diurnal changes of sea breeze cannot be evaluated. However, the benefit of finer resolution grid to improving wind simulation was obvious.



The following figure displays the time evolution of surface ozone and NOx concentrations from the observation and model results out of 3 grids for QHD. It can be seen that overall the model, regardless which grid resolution was applied, underestimated ground-level NOx concentrations but overestimated surface ozone levels. The ozone overestimate was especially large during summer months when its photochemical formation was the most efficient. We believe that the inaccurate NOx emissions representations were largely responsible for the model-observation mismatch. On the other hand, the benefit of increasing grid resolution to improving ozone and NOx forecast skills was obvious. The biases of ozone/NOx for the 45-km, 15-km, and 5-km resolution grids were 29.94/-22.46 ppbv, 24.09/-20.29 ppbv, 23.97/-17.95 ppbv, respectively. The respective RMSE were 37.24/28.87 ppbv, 27.28/27.57 ppbv, 27.01/26.38 ppbv. The improvement using the 15-km grid over the 45-km grid was remarkable but that using the 5-km grid over the 15-km grid was marginal.



In summary, the authors agree that, in general, the higher the grid resolution is, the better the simulation results will be. High resolution modeling is especially important to coastal areas and complex terrains where land-surface driving forces are in sharp contrast, such as QHD site. On the other hand, this research also agrees with the findings reported in many other papers that the benefit of higher resolution modeling of air quality starts to diminish at certain point due to the nonlinear nature of the atmospheric system. Balancing the modeling accuracy and computing resource constrain, a 15-km resolution grid has been recommended for future MICS-Asia activities if the investigate domain remains unchanged. We modified the manuscript to make this point explicitly stated in the section 3.1.2.b (Individual site) section 4 (Summary).

In section 3.1.2.b:

"An effort has been put to identify the potential reasons that caused the model-observation discrepancy. First and as discussed previously, the spatial distribution of emissions was one key to determining air quality forecast accuracy. Figure 3s shows the typical time evolutions of surface O_3 and NOx over the rural (XL) and urban (QHD) sites. It can readily be seen that NOx was underestimated at the urban site but overestimated at the rural site. The coarser the grid resolution, the severer the underestimates/overestimates were. This indicated that the 45-km resolution tended to smooth out emissions to make urban (or emissions centers) less polluted but rural more polluted. It in turn led to an overestimate of surface O_3 over the urban sites mainly due to the reduced NOx titration effect, especially at night when there was no photochemical O_3 formation. The statistics showed that the bias of the modeled daytime (7 am ~ 7 pm local time) average surface O_3 was 30% ~ 90% smaller than that of the daily average in the urban sites, no matter which grid resolution was applied. This suggested that in the future the high-resolution emissions, especially proper representation of emission gradients, would be helpful in improving air quality prediction. The effect of emissions gradients associated with the grid resolution would be further discussed in the inter-model comparison section.

Next, the driving meteorology, especially wind, was important to accurately forecast air quality over coastal areas that bore sharp thermal contrasts. QHD site locates approximately 5 km from the ocean and is subject to sea breeze effects. The detailed analysis of meteorology and air quality over QHD was conducted. The results indicated that the choice of grid resolution had large impacts on model simulations at this coastal site. The selection of the 5-km grid reduced biases of both surface temperature and wind speed. The biases of temperature reduced from 1.22 K (45-km) to -0.42 K (15-km), and further down to -0.31 K when the 5-km grid was applied. The biases of surface wind speed for the 45-km, 15-km, and 5-km grids were 3.72, 4.19, and 1.95 m s^{-1}, respectively. The improvement of meteorology forecast helped reducing the biases of air quality modeling. The biases of O_3/NOx for the 45-km, 15-km, and 5-km resolution grids were 29.94/-22.46 ppbv, 24.09/-20.29 ppbv, 23.97/-17.95 ppbv, respectively. The improvement using the 15-km grid over the 45-km grid was remarkable but that using the 5-km grid over the 15-km grid was marginal. The result emphasized the importance of high-resolution modeling to improvements of air quality forecast skills, especially at coastal and complex terrain areas (e.g., QHD and XL)."

In section 4:

"…With regard to MICS-Asia Phase III whose major goal was to examine regional air quality, in general, the finer the grid resolution was, the better the simulation results would be. This was especially true over the coastal areas and complex terrains where a sharp local energy gradient existed. Fine resolution grid was also extremely helpful to reproducing pollutants at higher concentrations that were most relevant to air quality planning and management. However, the benefit of high resolution was not linear with the decrease of grid size. At certain point, the improved modeling accuracy due to an increase in grid resolution was so marginal that it cannot justify the computational cost associated with the fine grid simulation. Based on the balance of modeling accuracy and efficiency, a 15-km horizontal grid appeared to be an appropriate choice to optimize model performance and resource usage if the study domain remained unchanged for future MICS-Asia activities. The study suggested that the high-resolution emissions, especially the proper representation of emission gradients, would be helpful in improving air quality prediction. Moreover, the profile measurements of both meteorology and air quality, in supplement with the ground monitoring networks, would be greatly helpful to identifying model deficiencies and thus improving model forecast skills"

3. Taylor diagram (Figure 2) is a useful way to present model performance, but it is not enough to represent model performance over a large region such as NCP and long- time simulation period such as one year since model performs differently in different sub-regions like urban or rural areas and at time periods (e.g., different reasons). It will be helpful if the authors can provide any model

performance in terms of spatial pattern (e.g., prediction biases) or time series of observation-simulation comparison. The result can be added in an appendix part if pages are limited.

Thanks for the suggestion. We have already had the statistics and discussions of each individual air quality site shown in Figures 3/4 and section 3.1.2.b. We added the time series of observation-simulation comparison averaged over the areas where the monitoring sites were located in the supplement material as shown in the following figures. We also inserted some discussions in section 3.1.1 for meteorological comparisons and in 3.1.2.a for regional average air quality comparisons.

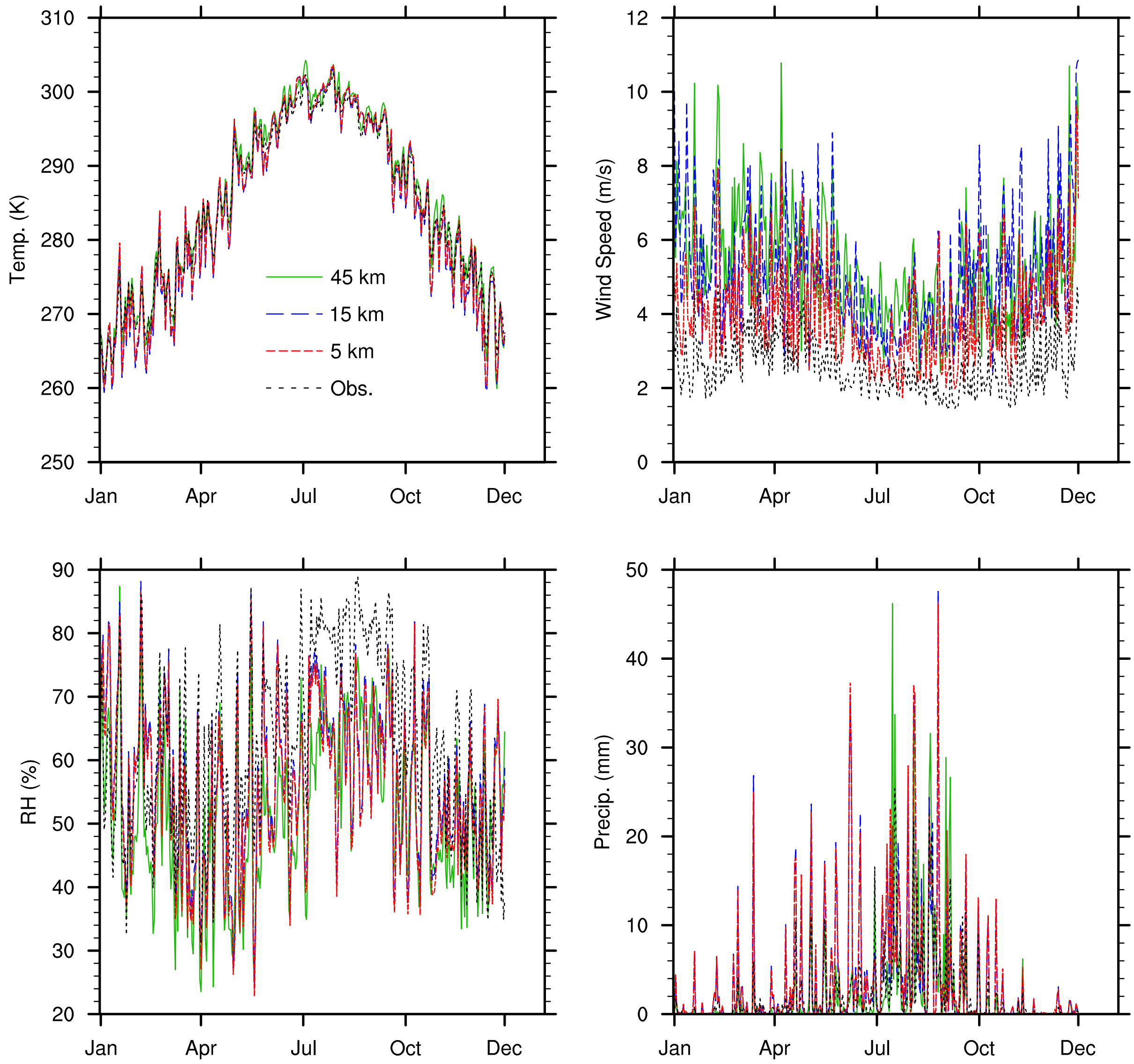


Figure 1s. Time series of surface meteorological parameters of observed vs. modeled values at different grid resolutions in areas where the monitoring sites are located.

At the end of section 3.1.1

"The time series of daily mean wind speed, air temperature, and RH, as well as daily total precipitation averaged over the monitoring sites is illustrated in Figure 1s in the supplement material. It echoed the above findings based on the Taylor diagram. It appeared that NU-WRF constantly overestimated surface wind speed throughout the year with large overestimate occurring in fall and winter, while it severely underestimated RH in summer. Uncertainty in representation of land surface characteristics at least partially explained these biases (Yu, 2014; Gao et al., 2018). High-resolution grid tended to reduce the uncertainty in land surface representation, which would be helpful to improving model performance in meteorology simulation. A more detailed exploration of model-observation mismatch was insightful but beyond the scope of this research."

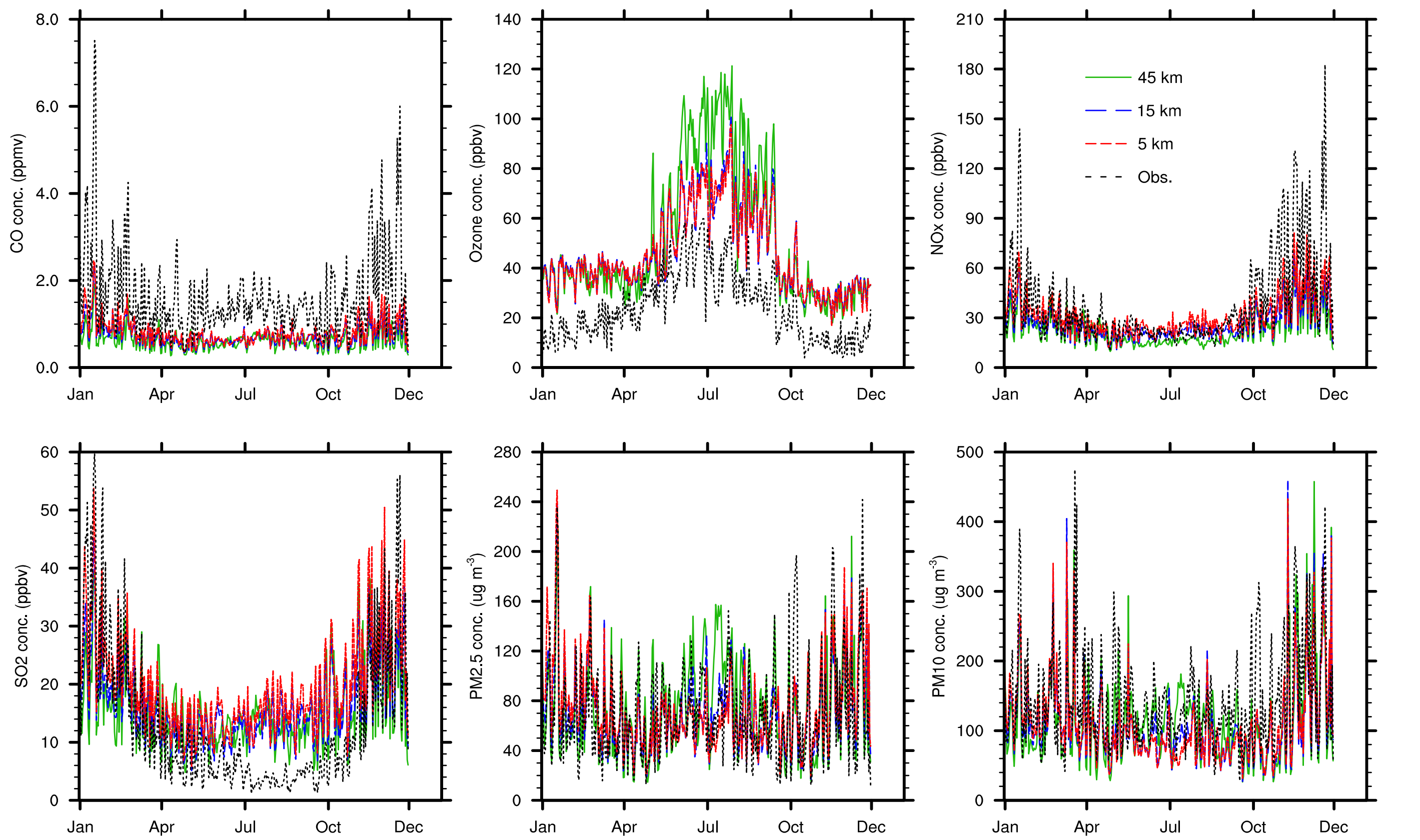


Figure 2s. Time series of surface air quality of observed vs. modeled values at different grid resolutions in areas where the monitoring sites are located.

At the end of section 3.1.2.a

"Figure 2s in the supplement material shows the time series of daily mean air quality averaged over the monitoring sites for the year 2010. The constant underestimate of CO throughout the year, severe underestimate of NOx in fall and winter, and large underestimate of SO_2 in summer all pointed out that the emissions inventory may be incomplete, agreeing with the reports by Kong et al. (2019) and Li et al. (2019). In the future, improvement of the emissions inventory accuracy and more realistic temporal emissions distribution may help improving NU-WRF performance in simulating O_3 photochemistry."

4. Figure 7: It seems that simulated O3 spatial patterns are not matched well with that of its precursors including NOx simulations and isoprene emissions (see Fig.6) at different grid-spacing. For instance, the simulated surface NOx concentrations at the grid-spacings of 15-km and 5-km grids look very similar to those at the grid-spacing of 45-km. However, the simulated O3 concentrations out of the 15-km and 5-km grids are much smaller than those at the grid-spacing of 45-km. More explanations will be helpful to readers for better understanding their relationship and the model performance at varying grid-spacings.

This is a good point. Actually, the other reviewer also raised the similar question. The authors believe, through carefully analysis, that the following two factors play major roles in these results. 1) Ozone photochemistry: ozone is a secondary pollutant formed in the atmosphere in the presence of its precursors such as NOx and VOCs, as well as solar radiation. Except for limited urban areas, ozone formation is typically limited by the availability of NOx in the vast rural areas as illustrated in Figure 7. In this case, the 45-km grid tended to distribute NOx emissions more evenly in the region, effectively decreasing the surface NOx concentration in urban areas but increasing it over rural areas. The larger average wind speeds out of the 45-km grid (Figure 6 and Table 3) in July further smoothed out NOx distributions in NCP. This in turn increased the domain average surface O_3 concentration via photochemistry based on the 45-km resolution results. 2) Vertical lifting effect: fine resolution (e.g., 15-km and 5-km) modeling tended to produce stronger updraft than a coarse resolution modeling (e.g., 45-km) as shown in Figure 4s. This finding is consistent with the work by Lee et al. (2018) who account this partly for the aerosol-cloud interaction induced freezing/evaporation-related invigoration mechanism. The strong uplift would bring more surface pollutants such as NOx into the upper atmosphere, thus further reducing the NOx availability at ground that limits the surface ozone production but increases its formation in the upper atmosphere (see Figure 8 in the manuscript). In future studies, the measured vertical meteorology and pollutant profiles will be extremely helpful in elucidating the reasons.

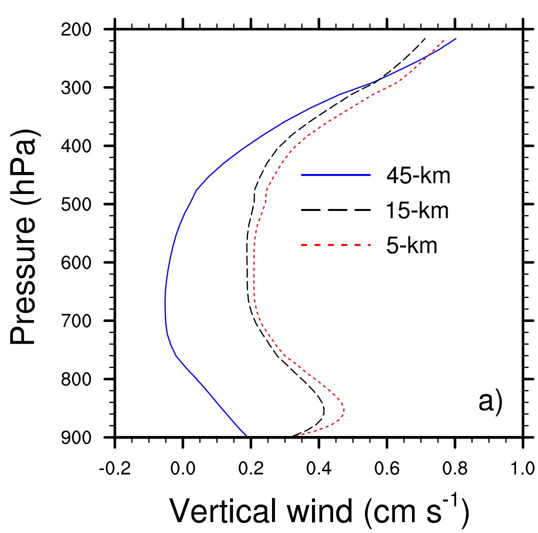


Figure 4s. Domain average vertical wind speed at different grid resolutions along the altitude

A few sentences were added in section 3.2.3:

"…The domain average discussed in this section, however, was the average covering the vast rural area that generally was NOx-limited such that surface O_3 formation was controlled by the availability of NOx – more NOx resulting in more O_3 through photochemical processes. In this case, the 45-km grid tended to distribute NOx emissions more evenly in the region, effectively decreasing the surface NOx concentration in urban areas but increasing it over rural areas. The larger average July wind speed simulated by the 45-km grid (Figure 6 and Table 3) further smoothed out the NOx distribution in NCP. This in turn increased the domain average surface O_3 concentration via photochemistry based on the 45-km resolution results. In addition, vertical lifting played an important role in explaining the maximum regional O_3 in July simulated by the 45-km grid as compared to the results by the other two grid resolutions. As displayed in Figure 4s in the supplement material, a fine resolution modeling (e.g., 5-km) tended to produce a stronger updraft than a coarse resolution modeling (e.g., 45-km), consistent with the findings by Lee et al. (2018). The strong uplift would bring more surface pollutants such as NOx into the upper atmosphere, thus further reducing the NOx availability at ground limiting the surface ozone production but increasing its formation in the upper atmosphere."

5. L430-432: How the maximum PBLH can be observed in Mongolian Plain where surface cover is dominated by grass?

PBL growth is primarily driven by the buoyancy due to surface heating. Thus, PBLH is closely related to the sensible heating at surface. The larger the sensible heating is, the deeper the PBL will be (e.g., Tao et al., 2013). Meanwhile, the high sensible heating is generally associated with a dry soil as reported in Bindlish et al. (2001). Major vegetation coverages over the study domain include grasslands mosaiced with open shrublands (over large portions of the northwest quartile of the domain), croplands (over large portions of eastern part of the domain outside of water), various deciduous forests (areas separate grassland and cropland), and urban. The grassland soil is generally drier than that of other vegetation covers in the domain. This explains why the largest average PBLH is found over the grassland in the northwestern corner of the domain. The text of L432-434 has been modified as:

"…The large average PBLH (more than 1,000 m) was found in the northwestern corner of the domain with a dominant land cover type of grassland mosaiced with open shrubland that appeared to be drier than the other land cover types in the domain. The high sensible heating associated with dry soil tended to produce the deep PBL (Tao et al., 2013)."

6. Table 3: Is it possible to add any available observational data for a comparison? The values presented in Table 3 represent domain average. It is not clear whether the simulations at those grids over ocean were included in the calculations.

The purpose of Table 3 is to facilitate the analysis of inter-resolution model comparison (section 3.2). Therefore, no observational data is listed in this table. The comparisons with the observations have been presented in section 3.1. The regional averages presented in Table 3 were calculated including every grid (land and ocean) within the domain. We changed the Table title to "Domain total emissions and average meteorology and air quality at various resolutions".

7. L60: Is "CHIMERE" defined? Please check similar issue for other abbreviation terms.

CHIMERE is not an abbreviation. It is the name of a Eulerian off-line chemistry-transport model developed in France. We modified the sentence (L61) as "…using the CHIMERE chemistry-transport model at various horizontal resolutions over Paris". We also checked the text and spelled out the abbreviation when it first occurred.

8. L120: Is "off" correct?

We modified the sentence to avoid confusion. The new description is "…new Grell cumulus scheme developed from the ensemble cumulus scheme that allowed subsidence spreading.".

9. L208: "simulated the best precipitation" or "simulated the precipitation best"? I recommend the latter.

We changed the sentence as suggested. We also checked the text to make the recommended changes as appropriate.

Response to Reviewer #2

1. The manuscript concludes that the 15-km resolution model has the overall best performance among the three. This is somewhat surprising as the finest resolution model is often assumed to be better.

In this investigation, we have found that the 5-km resolution modeling provided the best results of wind and surface pollutant levels, especially in polluted conditions that were more relevant to air quality regulation (e.g., compliance of national air quality standards), measured by bias and RMSE. However, the improvement of model performance on air quality was not linear with the resolution increase. For example, the accuracy of modeled surface O_3 out of the 15-km grid was greatly improved over the one from the 45-km grid. Further increase of grid resolution to 5-km, however, showed diminished impact on model performance on O_3 prediction for the study region. In addition, the cost in terms of cpu hours and disk space usage increased dramatically when adopting the 5-km grid resolution, which would be a big hurdle for the inter-model comparison studies such as MICS-Asia that relied on community contributions to model Asian air quality over a relatively long time period. Considering all these factors, we suggest a 15-km resolution grid for future MICS-Asia modeling activity to achieve both accuracy and efficiency. We checked the wording of the manuscript to make it clear that 15-km grid did not provide the best performance but rather was an optimal resolution that balanced the model accuracy and resource usages. For example, we modified the section 4 (Summary) as:

"…With regard to MICS-Asia Phase III whose major goal was to examine regional air quality, in general, the finer the grid resolution was, the better the simulation results would be. This was especially true over the coastal areas and complex terrains where a sharp local energy gradient existed. Fine resolution grid was also extremely helpful to reproducing pollutants at higher concentrations that were most relevant to air quality planning and management. However, the benefit of high resolution was not linear with the decrease of grid size. At certain point, the improved modeling accuracy due to an increase in grid resolution was so marginal that it cannot justify the computational cost associated with the fine grid simulation. Based on the balance of modeling accuracy and efficiency, a 15-km horizontal grid appeared to be an appropriate choice to optimize model performance and resource usage if the study domain remained unchanged for future MICS-Asia activities. The study suggested that the high-resolution emissions, especially the proper representation of emission gradients, would be helpful in improving air quality prediction. Moreover, the profile measurements of both meteorology and air quality, in supplement with the ground monitoring networks, would be greatly helpful to identifying model deficiencies and thus improving model forecast skills."

2. For the most part, the manuscript provides only domain-mean comparison between the three resolutions against observations. Although site-level model evaluation is shown in the figures, they are mere statistics and lack follow-up investigations or discussions that can be linked to certain model processes or input data that can provide insights for model improvement or can be generalized for other regions and time periods. For example, more analysis should be conducted to examine where/when the variations in meteorology and air quality are the largest within the domain that are most challenging for the 5-km model to capture.

This is a very good suggestion. We went back to data and made more analysis. Based on the results, we believe that the following factors account at least partially for the discrepancy between the modeled and observed air quality. 1) Spatial distribution of emissions was one key to determining air quality forecast accuracy. Out of 25 air quality monitoring sites used for model evaluation, 3 were rural sites and the remaining were urban/suburban sites. Figure 3s shows the typical time evolutions of surface ozone and NOx over the rural (XL) and urban (QHD) sites. It can readily be seen that NOx was underestimated at the urban site but overestimated at the rural site. The coarser the grid resolution, the severer the underestimates/overestimates were. This indicated that the 45-km resolution tended to smooth out emissions to make urban (or emissions centers) less polluted but rural more polluted. It in turn led to an overestimate of surface ozone over the urban sites mainly due to the reduced NOx titration effect, especially at night when there was no photochemical ozone formation. The statistics showed that the bias of the modeled daytime (7 am ~ 7 pm local time) average surface O_3 was 30% ~ 90% smaller than that of the daily average in the urban sites, no matter which grid resolution was applied. This suggests that, in the future, the high-resolution emissions, especially proper representation of emission gradients, will be helpful in improving air quality prediction. This point will be revisited in addressing comment 3.

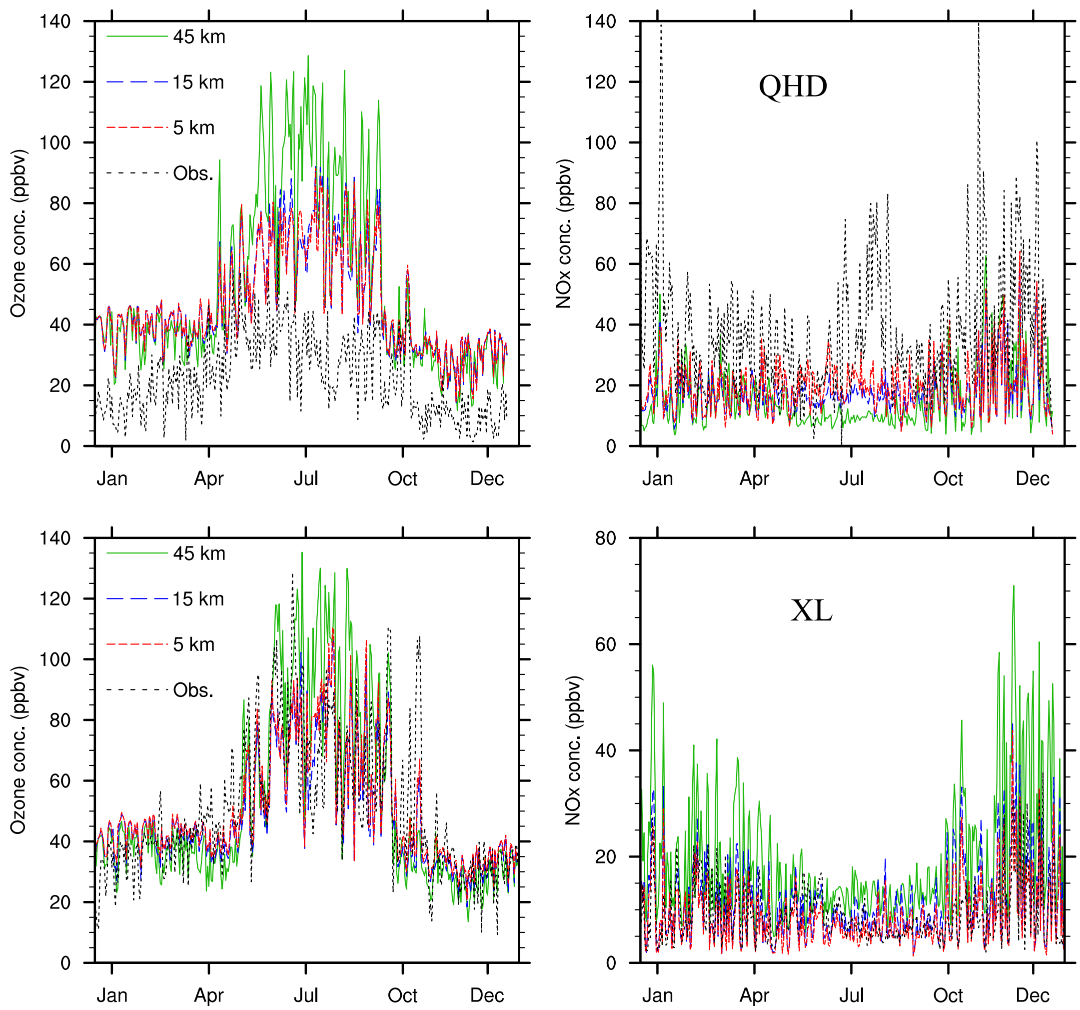
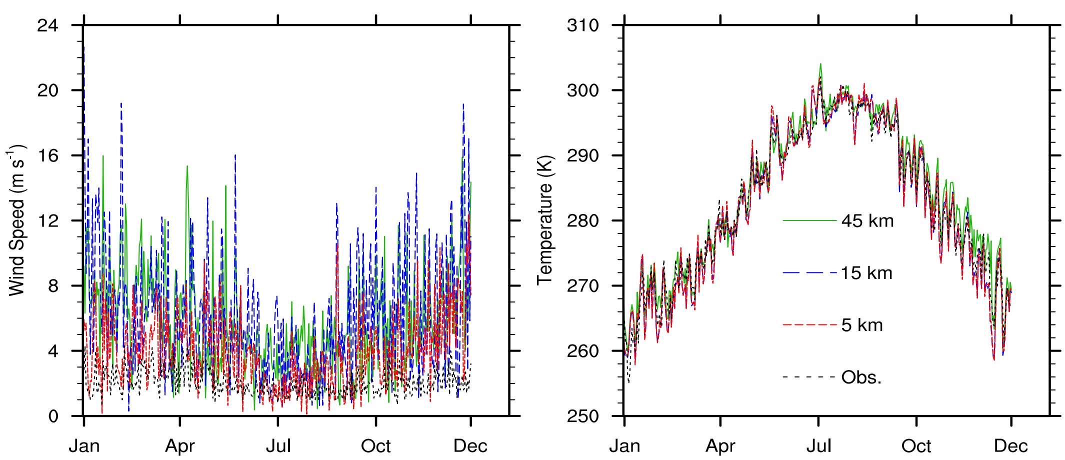


Figure 3s. Time series of surface O_3 and NOx concentrations over QHD (upper) and XL (lower) sites of the observed vs. modeled values at different grid resolutions.

2) The driving meteorology, especially wind, was important to accurately forecast air quality. Take QHD site as an example. QHD site locates approximately 5 km from the ocean and is subject to sea breeze effects. There is a meteorological monitoring station co-locating at QHD. The temporal profiles of surface wind speed and temperature from the observation and model results out of 3 grids for QHD are shown in the following figure. The results indicated that the choice of grid resolution had large impacts on model simulations at this coastal site. The selection of the 5-km grid reduced biases of both surface temperature and wind speed. The biases of temperature reduced from 1.22 K (45-km) to -0.42 K (15-km), and further down to -0.31 K when the 5-km grid was applied. The biases of surface wind speed for the 45-km, 15-km, and 5-km grids were 3.72, 4.19, and 1.95 m s^{-1}, respectively. The improvement of meteorology forecast helped reducing the biases of air quality modeling. The biases of ozone/NOx for the 45-km, 15-km, and 5-km resolution grids were 29.94/-22.46 ppbv, 24.09/-20.29 ppbv, 23.97/-17.95 ppbv, respectively. The improvement using the 15-km grid over the 45-km grid was remarkable but that using the 5-km grid over the 15-km grid was marginal. Vertical wind profile was another important factor to determine surface air quality as shown in the answer to Comment 4. This emphasizes the importance to measure vertical profiles of both meteorology and air quality in the future, which will help improve model skills.



3) Photochemistry mechanism also impacts the model performance. This has been shown in the companion papers by Li et al. (2019) and Kong et al. (2019).

In summary, the authors find out that a high-resolution emissions inventory would greatly help improving the model performances, especially over urban areas and emissions centers. Over the coastal areas (e.g., QHD) and complex terrain areas (e.g., XL), high resolution modeling tends to produce a more realistic wind field that benefits air quality simulation. In the future, the profile measurements of both meteorology and air quality are needed to elucidate the discrepancy between simulation and observation, thus help to improve model skills. We added discussions in section 3.1.2.b (Individual site), section 3.2.3 (see answer to Comment 4), and section 4 (see answer to Comment 1) to reflect the above analysis.

In section 3.1.2.b:

"An effort has been put to identify the potential reasons that caused the model-observation discrepancy. First and as discussed previously, the spatial distribution of emissions was one key to determining air quality forecast accuracy. Figure 3s shows the typical time evolutions of surface O_3 and NOx over the rural (XL) and urban (QHD) sites. It can readily be seen that NOx was underestimated at the urban site but overestimated at the rural site. The coarser the grid resolution, the severer the underestimates/overestimates were. This indicated that the 45-km resolution tended to smooth out emissions to make urban (or emissions centers) less polluted but rural more polluted. It in turn led to an overestimate of surface O_3 over the urban sites mainly due to the reduced NOx titration effect, especially at night when there was no photochemical O_3 formation. The statistics showed that the bias of the modeled daytime (7 am ~ 7 pm local time) average surface O_3 was 30% ~ 90% smaller than that of the daily average in the urban sites, no matter which grid resolution was applied. This suggested that in the future the high-resolution emissions, especially proper representation of emission gradients, would be helpful in improving air quality prediction. The effect of emissions gradients associated with the grid resolution would be further discussed in the inter-model comparison section.

Next, the driving meteorology, especially wind, was important to accurately forecast air quality over coastal areas that bore sharp thermal contrasts. QHD site locates approximately 5 km from the ocean and is subject to sea breeze effects. The detailed analysis of meteorology and air quality over QHD was conducted. The results indicated that the choice of grid resolution had large impacts on model simulations at this coastal site. The selection of the 5-km grid reduced biases of both surface temperature and wind speed. The biases of temperature reduced from 1.22 K (45-km) to -0.42 K (15-km), and further down to -0.31 K when the 5-km grid was applied. The biases of surface wind speed for the 45-km, 15-km, and 5-km grids were 3.72, 4.19, and 1.95 m s^{-1}, respectively. The improvement of meteorology forecast helped reducing the biases of air quality modeling. The biases of O_3/NOx for the 45-km, 15-km, and 5-km resolution grids were 29.94/-22.46 ppbv, 24.09/-20.29 ppbv, 23.97/-17.95 ppbv, respectively. The improvement using the 15-km grid over the 45-km grid was remarkable but that using the 5-km grid over the 15-km grid was marginal. The result emphasized the importance of high-resolution modeling to improvements of air quality forecast skills, especially at coastal and complex terrain areas (e.g., QHD and XL)."

3. It is not clear whether the model input data are resolution aware. Are the underlying emissions inventory data and land surface data (topography, LAI, etc) at a fine resolution of 5 km and then aggregated to the coarser resolutions? If the model is not driven by inputs that can resolve 5-km surface conditions, the 5-km model will not be able to correctly simulate air pollution variations at the 5-km scale.

Thanks for raising a very important point. In addition to the computation constrain, the challenge to employing an ultra-fine resolution modeling is the availability of the input data that are at the same or similar resolution. In this study, the land surface data were derived from the 30s resolution (around 30 m along mid-latitude) MODIS products that were aggregate to the model resolution. However, the MIX anthropogenic and GFEDv3 fire emissions inventories utilized in the study have a resolution of 0.25 by 0.25 degree and 0.5 by 0.5 degree, respectively. As indicated in the answer to Comment 2 above, the uncertainty of emissions may lead to air quality modeling errors. Therefore, the resolution-aware emissions may further improve the model performance using a 5-km grid. We added a caveat in section 4 (Summary) to reflect this.

"… It also was worth noting that the benefit of increasing grid resolution to better surface O_3 and PM2.5 simulations started to diminish when the horizontal resolution reached 15-km, agreeing with the finding by Valari and Menut (2008). There is a caveat, though. The anthropogenic MIX and fire GFEDv3 emissions inventories bore the 0.25º by 0.25º and 0.5º by 0.5º resolution, respectively. These resolutions cannot resolve the 5-km grid. Should a 5-km resolution emissions inventory be available and used, the benefit of high-resolution modeling would likely be more prominent.".

4. Figure 7, top panel: Ozone simulated by the 45-km model is almost 20 ppbv higher than the other two resolutions for July throughout the whole domain, while emissions of ozone precursors and meteorology are not so different. Why? Is this some kind of model error? If the model's oxidant budget is strongly resolution-dependent, one will question whether the model processes are parameterized correctly. A stable model should produce regional-mean concentrations of key species that are more or less consistent between different resolutions; it is the sub-regional variability and extreme concentrations that will differ as the resolution changes. This is reflected in ozone simulated by the 15-km and 5-km grids, but the 45-km model is an outlier.

Thanks for pointing this out. Actually, the first reviewer also raised the similar question. The authors believe, through carefully analysis, that the following two factors play major roles in these results. 1) Ozone photochemistry: ozone is a secondary pollutant formed in the atmospheric in the presence of its precursors such as NOx and VOCs, as well as solar radiation. Except for limited urban areas, ozone formation is typically limited by the availability of NOx in the vast rural areas as illustrated in Figure 7. In this case, the 45-km grid tended to distribute NOx emissions more evenly in the region, effectively decreasing the surface NOx concentration in urban areas but increasing it over rural areas. The larger average wind speeds out of the 45-km grid (Figure 6 and Table 3) in July further smoothed out NOx distributions in NCP. This in turn increased the domain average surface O_3 concentration via photochemistry based on the 45-km resolution results. Actually, the spatial distributions of annual average surface O_3 out of three grids appeared to be less variable. 2) Vertical lifting effect: fine resolution (e.g., 15-km and 5-km) modeling tended to produce a stronger updraft than a coarse resolution modeling (e.g., 45-km) as shown in Figure 4s. This finding is consistent with the work by Lee et al. (2018) who account this partly for the aerosol-cloud interaction induced freezing/evaporation-related invigoration mechanism. The strong uplift would bring more surface pollutants such as NOx into the upper atmosphere, thus further reducing the NOx availability at ground that limits the surface ozone production but increases its formation in the upper atmosphere (see Figure 8 in the manuscript). In future studies, the measured vertical meteorology and pollutant profiles will be extremely helpful in elucidating the reasons.

A few sentences were added in section 3.2.3:
"…The domain average discussed in this section, however, was the average covering the vast rural area that generally was NOx-limited such that surface O_3 formation was controlled by the availability of NOx – more NOx resulting in more O_3 through photochemical processes. In this case, the 45-km grid tended to distribute NOx emissions more evenly in the region, effectively decreasing the surface NOx concentration in urban areas but increasing it over rural areas. The larger average July wind speed simulated by the 45-km grid (Figure 6 and Table 3) further smoothed out the NOx distribution in NCP. This in turn increased the domain average surface O_3 concentration via photochemistry based on the 45-km resolution results. In addition, vertical lifting

played an important role in explaining the maximum regional O_3 in July simulated by the 45-km grid as compared to the results by the other two grid resolutions. As displayed in Figure 4s in the supplement material, a fine resolution modeling (e.g., 5-km) tended to produce a stronger updraft than a coarse resolution modeling (e.g., 45-km), consistent with the findings by Lee et al. (2018). The strong uplift would bring more surface pollutants such as NOx into the upper atmosphere, thus further reducing the NOx availability at ground limiting the surface ozone production but increasing its formation in the upper atmosphere."

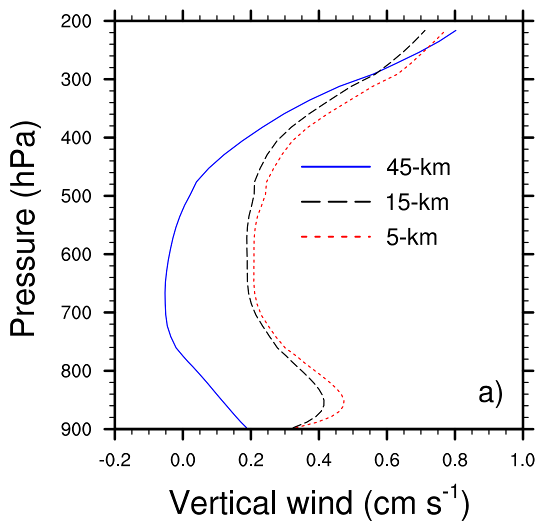


Figure 4s. Domain average vertical wind speed at different grid resolutions along the altitude

5. Table 3: Natural emissions (isoprene, dust, and sea salt) are very different between the three resolutions, varying by almost a factor of two. While these emissions are dependent on meteorology and thus on the model resolution, the standard practice is to implement a scaling factor so that the domain-wide emissions are consistent between different resolutions. Otherwise, it will not be a fair comparison as the emissions are not constant across the three resolutions. As this manuscript is part of a model intercomparison study, these emissions should be consistent with other models participating in the study.

We treated the biogenic, dust, and seasalt emissions that were calculated online as part of the effect of grid resolutions on air quality since the meteorological driving forces of these emissions, such as temperature, solar radiation, and wind, were impacted by the choice of grid resolutions. We think this is a fair justification.

6. Line 215-210: the different conclusion from Gao et al. was due to the difference in observations or in the model setting?

Gao et al. investigated the grid resolution effect on precipitation over the contiguous U.S. Their domain, modeling setup, and observations were all different from the ones used in this study. More importantly, Gao et al. used the processed precipitation data for their model evaluation – their precipitation data were based on the daily rain gauge data that were gridded to the 0.125° resolution using the synergraphic mapping algorithm with topographic adjustment to the monthly precipitation climatology. The processed data promoted the precipitation homogeneity and reduced the chances of model-observation mismatch of a precipitation event. This may be the major reason that two studies draw the opposite conclusions. In the manuscript we emphasized that our conclusion was based on the comparison with the site observation. For example, in section 4 (Summary):

"…The statement on precipitation should be taken with caution since it was based on the comparison with the site observations. Seeing the very heterogeneous nature of precipitation, the penalty of model hitting or missing a rain event was severe. Thus, the coarse grid covering more areas within a grid cell would reduce chances of mistaken precipitation hitting or missing simulations. However, a comparison of modeled precipitations to gridded "observation" that was re-constructed using the synergraphic mapping algorithm with topographic adjustment to the monthly precipitation climatology showed opposite result, where the fine resolution modeling showed superior reproduction of precipitation than the coarse resolution simulation (Gao et al., 2017)."

7. Table 2: I don't understand this table. What are the numbers in each cell and why they are so different?

Table 2 lists the occurrences of exceedances of China's National Ambient Air Quality Standards (NAAQS). Column "Frequency" indicates the time integration of each NAAQS. Column "Class 1" lists the NAAQS for rural sites, and "Class 2" lists the standards for urban-suburban sites. "Obs" lists the occurrences of NAAQS exceedances for each pollutant based on the observations. Columns "45-km", "15-km", and "5-km" list the occurrences of NAAQS exceedances based on the modeling results using "45-km", "15-km", and "5-km" grid resolutions. We added a sentence in section 3.1.2.c:

"Table 2 lists the occurrences of violations of China's national ambient air quality standards (NAAQS) for the six pollutants from both observations and simulations, in which columns "Class 1" and "Class 2" list the standards for rural and urban-suburban sites, respectively, and column "Frequency" indicates the time integration of each NAAQS."

8. Line 32: add "the" before 21st century.

Done.

9. Line 68: remove "however".

Done.

Evaluation of NU-WRF Performance on Air Quality Simulation under Various Model Resolutions – An Investigation within Framework of MICS-Asia Phase III

Zhining Tao[1,2], Mian Chin[2], Meng Gao[3], Tom Kucsera[1,2], Dongchul Kim[1,2], Huisheng Bian[2,4], Jun-ichi Kurokawa[5], Yuesi Wang[6], Zirui Liu[6], Gregory R. Carmichael[7], Zifa Wang[6,8,9], and Hajime Akimoto[10]

1. Universities Space Research Association, Columbia, MD, USA

2. NASA Goddard Space Flight Center, Greenbelt, MD, USA

3. John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

4. University of Maryland at Baltimore County, Baltimore, MD, USA

5. Japan Environmental Sanitation Center, Asia Center for Air Pollution Research, Niigata, 950-2144, Japan

6. State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, 100029, China

7. Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA, USA

8. College of Earth Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

9. Center for Excellence in Urban Atmospheric Environment, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China

10. National Institute for Environmental Studies, Onogawa, Tsukuba, 305-8506, Japan

Formatted: Superscript

Abstract

Horizontal grid resolution has a profound effect on model performances on meteorology and air quality simulations. In contribution to MICS-Asia Phase III, one of whose goals was to identify and reduce model uncertainty in air quality prediction, this study examined the impact of grid resolution on meteorology and air quality over East Asia, focusing on the North China Plain (NCP) region. NASA Unified Weather Research and Forecasting (NU-WRF) model has been applied with the horizontal resolutions at 45-, 15-, and 5-km. The results revealed that, in comparison with ground observations, no single resolution can yield the best model performance for all variables across all stations. From a regional average perspective (i.e., across all monitoring sites), air temperature modeling was not sensitive to the grid resolution but wind and precipitation simulation showed the opposite. NU-WRF with the 5-km grid simulated the wind speed best, while the 45-km grid yielded the most realistic precipitation as compared to the site observations. For air quality simulations, finer resolution generally led to better comparisons with observations for O_3, CO, NOx, and PM2.5. However, the improvement of model performance on air quality was not linear with the resolution increase. The accuracy of modeled surface O_3 out of the 15-km grid was greatly improved over the one from the 45-km grid. Further increase of grid resolution to 5-km, however, showed diminished impact on model performance improvement on O_3 prediction. In addition, 5-km resolution grid showed large advantage to better capture the frequency of high pollution occurrences. This was important for assessment of noncompliance of ambient air quality standards, which was key to air quality planning and management. Balancing the modeling accuracy and resource limitation, a 15-km grid resolution was suggested for future MICS-Asia air quality modeling activity if the research region remained unchanged. This investigation also found out large overestimate of ground-level O_3 and underestimate of surface NOx and CO, likely due to missing emissions of NOx and CO.

Deleted: best

Deleted: finer

Deleted: findings

1. Introduction

Air pollution is a threat to human health/climate and detrimental to ecosystem (Anenberg et al., 2010; https://www.who.int/airpollution/ambient/en/). Lelieveld et al. (2015) estimated that over 3 million premature mortality could be attributable to outdoor air pollution worldwide in 2010 based on their analysis of data and the results from a high-resolution global air quality model. Since the turn of the 21st century, East Asia has undergone remarkable changes in air quality as observed by satellite and ground stations (Jin et al., 2016; Krotkov et al., 2016). In the past decade, haze (fine particle) pollution has become a household name in China and many severe haze events have been reported and their formation mechanisms and associations with global- and meso-scale meteorology have been analyzed (Zhao et al., 2013; Huang et al., 2014; Gao et al., 2016; Cai et al., 2017; Zou et al., 2017). Meanwhile, ground level ozone has been a major air quality concern in China (Wang et al., 2017; Lu et al., 2018), Japan (Akimoto et al., 2015), and South Korea (Seo et al., 2014). In combination with observations from various platforms, chemical transport model (CTM) remains an important tool to understand mechanisms, to investigate spatial-temporal distributions, and to design feasible control strategies of air pollution. However, CTM model uncertainties persist (e.g., Carmichael et al., 2008) and the interpretation of any model results needs caution and exertion of careful analysis.

Inter-model comparison study provides a valuable way to understand model uncertainties and sheds light on model improvements. With this as one of its major goals, the Model Inter-Comparison Study for Asia (MICS-Asia) was initiated in 1998. Since then MICS-Asia has gone through three phases with emphasis on various aspects of air pollution. Phase I focused on long-range transport and deposition of sulfur over East Asia (Carmichael et al., 2002). Phase II expanded the analysis on more pollutants including nitrogen compounds, particulate matter, and ozone, in addition to sulfur (Carmichael et al., 2008). Fast moving to Phase III, MICS-Asia concentrated on three topics with number one aiming at identifying strengths and weaknesses of current air quality models to provide insights on reducing uncertainties (Gao et al., 2018). There are totally 14 CTMs – 13 regional and 1 global – participating in the coordinated model experiment, which simulated air quality over Asia throughout the year 2010. Due to the constrain of computing resources among participating modeling groups, a 45-km horizontal resolution has been commanded for every team to run the year-long experiment.

This relatively coarse spatial resolution raises the question of how representative the model can resolve key issues relevant to air quality and its planning/regulation, e.g., heterogeneous emissions, inhomogeneous land cover and meteorology. For example, Valari and Menut (2008) explored the issue using the CHIMERE chemistry-transport model at various horizontal resolutions over Paris. They found out that the ozone simulation was especially sensitive to the resolution of emissions. However, the benefit of increasing emissions resolutions to improve ozone forecast skills was not monotonic and at certain point the forecast accuracy decreased upon further resolution increase. Using the Weather Research and Forecasting Chemistry model (WRF-Chem) with various horizontal resolution (3 ~ 24 km) over the Mexico City, Tie et al. (2010) concluded that a 1 to 6 ratio of grid resolution to city size appeared to be a threshold to improve ozone forecasting skill over mega-city areas: the forecast would be improved significantly when model resolution was below this threshold value. On contrary to Valari and Menut (2008), Tie et al. (2010) suggested that the meteorology changes associated with the grid size choice played a more prominent role in contributing to the improvement of ozone forecast skills. More recently, Neal et al. (2017) employed a high-resolution (12 km) air quality model with high-resolution emissions within the Met Office's Unified Model (AQUM) for air quality forecast over the Great Britain.

Deleted: however,

They found out that AQUM significantly improved the forecast accuracy of primary pollutants (e.g., NO_2 and SO_2) but less obviously for secondary pollutants like ozone, as compared with a regional composition-climate model (RCCM, 50 km horizontal resolution). But there was a drawback from their conclusion in that the chemical mechanisms and photolysis rates utilized in AQUM and RCCM were different, complicating the underlying reasons for changes in forecast skills. Lee et al. (2018) examined the importance of aerosol-cloud-radiation interactions to precipitation and the model resolution impact of key meteorological processes that affected precipitation using the Advanced Research WRF model. They found that the coarse model resolution would lower updraft, alter cloud properties (e.g, mass, condensation, evaporation, and deposition), and reduce cloud sensitivity to ambient aerosol changes. They further concluded that the uncertainty associated with resolution was much more than that related to cloud microphysics parameterization. The resultant meteorological condition change would trigger air quality response as well.

Despite the progress, the exploration of impacts of model resolution on local air quality over Asia is rare. Taking advantage of the MICS-Asia platform, we examined the issue over the MICS-Asia domain using the NASA Unified WRF (NU-WRF, Tao et al., 2013, 2016, 2018; Peters-Lidard et al., 2015), focusing on the North China Plain (NCP) that was plagued by frequent heavy air pollution episodes. The investigation would not only assist in gaining insights on how model horizontal resolution affects simulated meteorology and air quality, but also contribute to formulation of uncertainties resulted from model resolutions to the MICS-Asia community. The latter would especially be valuable since most MICS-Asia Phase III model simulations were conducted at a specific horizontal resolution (i.e., 45-km for most participants).

2. NU-WRF model and experiment design

NU-WRF is an integrated regional Earth-system modeling system developed from the advanced research version of WRF-Chem (Grell et al., 2005), which represents atmospheric chemistry, aerosol, cloud, precipitation, and land processes at convection-permitting spatial scales (typically 1-6 km). NU-WRF couples the community WRF-Chem with NASA's Land Information System (LIS), a software framework including a suite of land surface models (LSMs) that are driven by satellite/ground observations and reanalysis data (Kumar et al., 2006; Peters-Lidard et al., 2007). It also couples the Goddard Chemistry Aerosol Radiation and Transport (GOCART) bulk aerosol scheme (Chin et al., 2002, 2007) with the Goddard radiation (Chou and Suares, 1999) and microphysics schemes (Tao et al., 2011; Shi et al., 2014) that allows for fully coupled aerosol-cloud-radiation interaction simulations. In addition, NU-WRF links to the Goddard Satellite Data Simulator Unit (G-SDSU), which converts simulated atmospheric profiles, e.g, clouds, precipitation, and aerosols, into radiance or backscatter signals that can directly be compared with satellite level-1 measurements at a relevant spatial and temporal scale (Matsui et al., 2009, 2013, 2014). In this study, NU-WRF has been employed to carry out the model simulations at various horizontal resolutions using the same set of physical and chemical configurations.

A nested domain setup was configured to this investigation as shown Figure 1. The 45-km resolution mother domain (d01) covered the MICS-Asia Phase III study region. The nested 15-km (d02) and 5-km (d03) domains covered the East Asia and NCP, respectively. A one-way nesting approach was applied so that the values of the mother domains were independent on those of the respective nested domains. This analysis focused on NCP and its adjacent areas with over 1.1 million square kilometers. The key NU-WRF configurations included the updated Goddard cumulus ensemble microphysics scheme (Tao et al., 2011), new Goddard long/shortwave radiation scheme (Chou and Suares, 1999), Monin-Obukhov surface layer scheme, unified Noah land surface model (Ek et al., 2003) with LIS initialization (Peters-Lidard et al., 2015), Yonsei University planetary boundary layer scheme (YSU, Hong et al., 2006), new Grell cumulus scheme developed from the ensemble cumulus scheme (Grell and Devenyi, 2002) that allowed subsidence spreading (Lin et al., 2010), 2nd generation regional acid deposition model (RADM2, Stockwell et al., 1990; Gross and Stockwell, 2003) for trace gases and GOCART for aerosols. In this investigation, the option of fully coupled GOCART-Goddard microphysics and radiation schemes (Shi et al., 2014) has been implemented to account for the aerosol-cloud-radiation interactions.

Deleted: off

Anthropogenic emissions were from the mosaic Asian anthropogenic emissions inventory (MIX, Li et al., 2017) that was developed for the MICS-Asia Phase III. The MIX inventory was at the 0.25º by 0.25º resolution and projected to the study domain under the 45-, 15-, and 5-km horizonal resolutions. Fire emissions were from the 0.5º by 0.5º Global Fire Emissions Database version 3 (GFEDv3, van der Werf et al., 2010; Mu et al., 2011) and also projected to the targeted region. Biogenic emissions were computed online using the Model of Emissions of Gases and Aerosols from Nature version 2 (MEGAN2, Guenther et al., 2006). Dust and sea salt emissions were also calculated online using the dynamic GOCART dust emissions scheme (Kim et al, 2017) and sea salt scheme (Gong, 2003), respectively.

The meteorological Lateral Boundary Conditions (LBCs) were derived from the Modern Era Retrospective-Analysis for Research and Applications (MERRA, Rienecker et al., 2011). The trace gas LBCs were based on the 6-hour results from the Model for OZone And Related chemical Tracers (MOZART, Emmons et al., 2010). The aerosol LBCs were from the global GOCART simulation with a resolution of 1.25 (longitude) by 1 (latitude) degree (Chin et al., 2007). Three horizontal resolutions varied from 45-km to 5-km with 15-km in between. Terrain-following sixty vertical levels stretched from surface to 20 hPa with the 1st layer height of approximately 40 meters from surface. The simulation started on December 20, 2009, and ended on December 31, 2010, with the first 11 days as the spin-up.

3. Results

3.1. Comparisons with observations

The NU-WRF results out of different horizontal resolutions have compared with ground observations using the following statistic measures:

Correlation coefficient: $$r = \frac{\sum_{i=1}^{n}(m_i-\overline{m})(o_i-\overline{o})}{\sqrt{\sum_{i=1}^{n}(m_i-\overline{m})^2}\sqrt{\sum_{i=1}^{n}(o_i-\overline{o})^2}}$$

Mean bias: $$MB = \frac{1}{n}\sum_{i=1}^{n}(m_i - o_i)$$

Normalized mean bias: $$NMB = \frac{\sum_{i=1}^{n}(m_i-o_i)}{\sum_{i=1}^{n}o_i} \times 100\%$$

Root mean square error: $$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(m_i-o_i)^2}{n}}$$

Normalized standard deviation: $$NSD = \frac{\sqrt{\frac{\sum_{i=1}^{n}(m_i-\overline{m})^2}{n-1}}}{\sqrt{\frac{\sum_{i=1}^{n}(o_i-\overline{o})^2}{n-1}}}$$

Where, m_i and o_i denote for the modeled and observed values at time-space pair i; $\overline{m}$ and $\overline{o}$ represent the average modeled and observed values, respectively. r describes the strength and direction of a linear relationship between two variables – a perfect correlation has a value of 1. NMB and MB depict the mean deviation of modeled results from the respective observations. A perfect model simulation yields an *NMB* and a *MB* of 0. $RMSE$ measures the absolute accuracy of a model prediction. The smaller the *RMSE*, the better the model performance is. Similar to *NMB* and *MB*, a *RMSE* of 0 indicates a perfect model prediction. NSD is a measure to check how well the model can reproduce the variations of observations – a value of 1 represents a perfect reproduction of observed variations.

3.1.1. Meteorology

The 2010 meteorological observations were collected at the standard stations operated by China Meteorological Administration (CMA, http://data.cma.cn/en). The locations of each site within our study domain were represented with the black dots in Figure 1. In total there were 77 sites reporting daily average values of wind speed (Wind), air temperature (Temp), and relative humidity (RH), as well as daily total precipitation (Precip). Figure 2 (top row) shows the Taylor diagram summarizing r, NMB, and NSD of the comparison of regional mean (average of observations from 77 sites) daily meteorological variables. Along the azimuthal angle is r. NSD is proportional to the radial distance from the origin. NMB (sign and range) are represented by geometric shapes. The statistical measures under 45-, 15-, and 5-km resolutions are represented by color blue, green, and red, respectively. The closer to the point "Obs" on the Taylor diagram and smaller of NMB, the better the model performance is.

It can be seen that the model horizontal resolution has little impact on surface air temperature simulation. Regardless of resolution selections, the modeled temperature correlated very well with the corresponding observations with r values all approaching 0.99. NU-WRF also reproduced the observed temperature variations well with NSD ranging between 1.05 and 1.10. Meanwhile, NMB was within ±1% for all experimented resolutions. *RMSEs* were 1.13 K, 2.26 K, and 2.02 K for the 45-km, 15-km, and 5-km grids, respectively. The insensitivity of surface air temperature to the choice of model resolutions was also reported by Gao et al. (2017), who used WRF to explore the issue for summer seasons at the 36-, 12-, and 4-km resolutions.

On the other hand, the horizontal resolution has a remarkable effect on surface wind speed as shown in Figure 2 (top row). At 5-km resolution, NU-WRF yielded a r value of 0.75, *NMB* of approximately 54%, and *NSD* of 1.78. NU-WRF simulated a large variation in wind than the observed ones. As comparisons, the values of r, *NMB*, and *NSD* for 15-km and 45-km were 0.54, 95%, 2.14, and 0.71, 103%, 2.01, respectively. The respective *RMSEs* out of the 45-km, 15-km, and 5-km grids were 2.87, 2.82, and 1.67 m s^{-1}. It was apparent that 5-km resolution gave the overall best wind speed simulation compared to the observations, though NU-WRF overestimated the surface wind speed in all cases. The wind speed overestimate, especially under low wind conditions, was a common problem in all MICS-Asia participating models and other weather forecast models (Gao et al., 2018). This overestimate stemmed from many factors, including but not limited to terrain data uncertainty, poor representation of urban surface effect, horizontal and vertical grid resolutions, etc. Dr. Yu (2014) in her doctoral dissertation pointed out that surface wind simulation would be improved upon using more accurate land-use data. This is expected since surface wind is largely dependent on the land surface characteristics, such as albedo and roughness. High-resolution grid tends to have more accurate land-use representation seeing the inhomogeneous nature of land type.

NU-WRF simulations at all three resolutions yielded the similar reproductions of the observed variations in relative humidity (RH) with the *NSD* ranging between 0.87 and 0.88. The modeled RH was less variable than the observed one. While the modeled RH at 45-km resolution (r = 0.84) better correlated with the observations than those at the finer resolutions did (approximately 0.67 for both 15-km and 5-km resolutions), the *NMB* at this resolution was the largest (-17%) among the three cases. The *NMBs* for 15-km and 5-km cases were -10% and -12%, respectively. Overall, NU-WRF underestimated the surface RH. The respective *RMSEs* for 45-km, 15-km, and 5-km resolutions were 13.2%, 12.6%, and 13.3%. The simulation with the 15-km grid appeared to yield the overall best RH in three cases.

It was interesting to find that NU-WRF simulated the precipitation best, as directly compared to the rain gauge data, when using the 45-km grid. At this resolution, NU-WRF gave r of 0.81, *NMB* of 1.7%, *RMSE* of 3.2 mm day^{-1}, and *NSD* of 1.41. As comparisons, the values of r, *NMB*, *RMSE*, and *NSD* for 15-km and 5-km were 0.53, 76%, 5.7 mm day^{-1}, 1.71, and 0.52, 80%, 5.8 mm day^{-1},1.72, respectively. Finer resolutions indeed yielded worse results in precipitation modeling as compared to the site data. This may be because precipitation was a very heterogeneous phenomenon – finer model grid had larger chances to miss a precipitation event or hit an event that was not existent, leading to greater overall bias and poorer correlation. On the contrary, Gao et al. (2017) compared their WRF modeled results to the gridded precipitation based on daily rain gauge data that were gridded to the 0.125° resolution using the synergraphic mapping algorithm with topographic adjustment to the monthly precipitation climatology (Maurer et al., 2004). They reported that the modeled precipitation out of the 4-km resolution was much improved over that out of the coarser 36- or 12-km resolutions.

Deleted: best

The time series of daily mean wind speed, air temperature, and RH, as well as daily total precipitation averaged over the monitoring sites is illustrated in Figure 1s in the supplement material. It echoed the above findings based on the Taylor diagram. It appeared that NU-WRF constantly overestimated surface wind speed throughout the year with large overestimate occurring in fall and winter, while it severely underestimated RH in summer. Uncertainty in representation of land surface characteristics at least partially explained these biases (Yu, 2014; Gao et al., 2018). High-resolution grid tended to reduce the uncertainty in land surface representation, which would be helpful to improving model performance in meteorology simulation. A more detailed exploration of model-observation mismatch was insightful but beyond the scope of this research.

3.1.2. Air quality

The difference seen in the aforementioned meteorology would cause varied performances on air quality simulations at various model horizontal resolutions. In this study, the NU-WRF simulated surface air quality was compared to the corresponding observations. The 2010 ground-level air quality data were obtained from the Chinese Ecosystem Research Network (CERN, http://www.cern.ac.cn) operated by the Institute of Atmospheric Physics of Chinese Academy of Sciences. There were 25 monitoring sites distributed within a 500 km by 500 km area centering around Beijing, China (open diamond in Figure 1). The site locations and characteristics were listed in Table 1. 22 out of 25 sites were either in an urban or a suburban setting, with the balance being in a rural setting. Each site reported hourly concentrations of at least one of the following six pollutants – ozone (O_3), nitrogen oxides (NOx), carbon monoxide (CO), sulfur dioxide (SO_2), and particulate matters with aerodynamic diameters less than 2.5 and 10 μm (PM2.5 and PM10).

a. Regional average

First, the regional mean (averaged across 25 sites) daily surface concentrations from both observations and simulations, paired in space and time, were calculated. The r, *NME*, and *NSD* were then computed and illustrated in a Taylor diagram (Figure 2 (bottom row)).

The six pollutants can be put into two groups – one most relevant to ozone photochemistry including O_3, NOx, and CO, and the other closely tied to aerosols including SO_2, PM2.5, and PM10. It was readily seen that the r values of O_3, NOx, and CO were not very sensitive to the choice of model horizontal resolutions. For O_3, the r values for 45-km, 15-km, and 5-km grids were all around 0.85. The respective r values were 0.84, 0.81, 0.80 for NOx, and 0.80, 0.75, 0.73 for CO. In general, however, NU-WRF reproduced the observed variations in O_3, NOx, and CO better with a fine resolution than with a coarse one. *NSD* of 1.23 for O_3 at 5-km resolution was the closest to 1 among three resolutions (1.24 for 15-km and 2.01 for 45-km). *NSDs* were 0.40, 0.36, 0.46 for NOx, and 0.24, 0.27, 0.31 for CO, under the 45-km, 15-km, and 5-km resolutions, respectively, suggesting that simulations with the finest resolution tended to reproduce the observed variations better than the ones with coarse resolutions for these three trace gases. Meanwhile, NU-WRF yielded the smallest bias when employing the fine resolution grid. *NMBs* for O_3 decreased from 115% to 92% when grid resolutions increased from 45-km to 5-km. *NMBs* were -38%, -30%, -18% for NOx, and -61%, -55%, -51% for CO, under the 45-km, 15-km, and 5-km resolutions, respectively. It was apparent that NU-WRF overestimated surface O_3 but underestimated NOx and CO, consistent with the findings in the companion MICS-Asia III studies that based their results on ensemble model simulations (Li et al., 2019; Kong et al., 2019). The majority of the air quality monitoring sites used in this study were in an urban setting, which typically were in a VOC-limited regime. This meant that the underestimate of NOx would reduce the titration that consumed surface O_3 leading to its overestimate. We further analyzed the model bias for daytime (8-18 local standard time) vs. nighttime. It was found that the nighttime biases for surface O_3 and NOx were approximately 2~4 times higher than those of daytime, consistent with the finding that insufficient NOx titration caused overestimate of modeled surface O_3.

NU-WRF simulated less variations in 3 aerosol related pollutants than those of observations under all applied horizontal resolutions. The *NSDs* ranged from 0.56 (for SO_2 at 15-km resolution) to 0.96 (for PM2.5 at 45-km resolution). Though it reproduced the observed SO_2 variations the best (*NSD* = 0.68) with 5-km resolution, NU-WRF yielded the best *NSD* for PM2.5 (0.96) and PM10 (0.92) when 45-km resolution was employed. Similar to 3 trace gases relevant to surface O_3 formation, the choice of model resolution had a limited effect on r statistics. The r values varied from 0.70 (45-km resolution) to 0.76 (both 15- and 5-km) for surface SO_2, and from 0.68 (45-km resolution) to 0.63 (5-km) for PM2.5. The r values for PM10 were all around 0.58 under the selected resolutions. The impact of model resolution on *NMBs* showed mixed information – while the smallest *NMBs* for SO_2 (20%) and PM10 (-19%) were achieved using the 45-km resolution, the smallest *NMB* for PM2.5 (1.5%) was observed at the 15-km resolution. The model underestimate of PM10 was consistent with the findings of the companion investigation using the multi-model ensemble analysis (Chen et al., 2019).

Figure 2s in the supplement material shows the time series of daily mean air quality averaged over the monitoring sites for the year 2010. The constant underestimate of CO throughout the year, severe underestimate of NOx in fall and winter, and large underestimate of SO_2 in summer all pointed out that the emissions inventory may be incomplete, agreeing with the reports by Kong et al. (2019) and Li et al. (2019). In the future, improvement of the emissions inventory accuracy and more realistic temporal emissions distribution may help improving NU-WRF performance in simulating O_3 photochemistry.

Moved down [1]: In the future, improvement of the emissions inventory accuracy and more realistic temporal emissions distribution may help improving NU-WRF performance in simulating O_3 photochemistry.

Formatted: Subscript

Moved (insertion) [1]

b. Individual site

The daily average concentrations of each pollutants were calculated and paired in space and time at each air quality monitoring site. Then the statistics at each individual site was computed.

Figure 3 illustrates the comparisons of *MB*, *RMSE*, and correlation coefficient (r) of surface O_3 from different horizontal resolutions at each site. It can be found that there was no single resolution that yielded the best correlation across all sites. For example, the simulation with the 45-km horizontal resolution gave the best correlation over sites BD, CFD, CZ, HJ, SJZ, SQL, TG, TJ, TS, XH, XL, YF, YJ, and ZJK. On the other end of spectrum, BJT, DT, and LTH achieved the best correlation when the 5-km grid was applied. QHD saw the best correlation out of the simulation with the 15-km resolution. In any cases, however, the variations of r values from different horizontal resolutions at each site were small (less than 0.04). On the other hand, NU-WRF yielded the worst *MB* and *RMSE* when employing the 45-km resolution grid, while *MB* and *RMSE* were similar between simulations with 15-km and 5-km resolutions. Typically, at sites with urban/suburban settings, *MB (RMSE)* based on the 45-km grid was approximately 15~30% (20~40%) higher than that out of the 15-km or 5-km grids. It appeared that NU-WRF tended to have a better performance on ground-level O_3 simulation when increasing the horizontal resolution from 45-km to 15-km, but further finer resolution had diminished impact on improving surface O_3 modeling. This was consistent with the finding by Valari and Menut (2008) who concluded that the benefit of finer horizontal resolution grid to improving surface O_3 forecast skill would diminish at certain point.

Figure 4 shows the PM2.5 case of comparisons of *MB*, *RMSE*, and r. Only 10 sites reported PM2.5 measurements over year 2010. In general, the NU-WRF simulation with the 45-km grid correlated better to the respective observations than the other 2 resolutions. The only exception was site BD that saw the best correlation for the 5-km resolution. *MB* and *RMSE* results were mixed with no single resolution giving superior results across all sites. Over 2 rural sites (LS and XL), the simulations with the 15-km or 5-km grids yielded remarkably smaller *MB* but correlated less to the corresponding observations than the one with the 45-km grid. Over 8 urban/suburban sites, BD, SQL, and TG experienced the smallest *MB* when employing the 5-km resolution grid, while TG, TJ, and XH saw the least bias at the 45-km resolution. The smallest *MB* at BJT and LTH occurred using the 15-km grid.

At the individual site level, the impact of grid resolution on surface NOx and CO (figures not shown) modeling was similar to that at the regional average. Finer resolution simulation generally reduced *MB* and *RMSE*. The results out of the 45-km grid always had the largest bias. The underestimates of NOx at least partially explained the overestimate of surface O_3 at each site due to a less efficient NO-titration of O_3. This suggested that a higher resolution modeling with more accurate spatial representation of NOx emissions would help improving its performance on surface O_3 simulations.

The signals for SO_2 and PM10 (figures not shown) simulations were mixed as well. For example, the largest bias for SO_2 simulation over sites BD, CZ, GA, HS, LS, QA, QHD, XH, XL, YF, and YJ occurred when applying the 45-km grid, while the maximum bias over BJT, DT, HJ, LF, LTH, SJZ, SQL, TG, TJ, TS, ZJK, and ZZ happened at the 5-km resolution. Sites CD and CFD saw the largest bias at the 15-km resolution. Unlike PM10 that was almost always underestimated at each site regardless of grid resolutions, SO_2 was overestimated at 18 out of 25 sites and underestimated at the remaining 7 sites.

An effort has been put to identify the potential reasons that caused the model-observation discrepancy. First and as discussed previously, the spatial distribution of emissions was one key to determining air quality forecast accuracy. Figure 3s shows the typical time evolutions of surface

O_3 and NOx over the rural (XL) and urban (QHD) sites. It can readily be seen that NOx was underestimated at the urban site but overestimated at the rural site. The coarser the grid resolution, the severer the underestimates/overestimates were. This indicated that the 45-km resolution tended to smooth out emissions to make urban (or emissions centers) less polluted but rural more polluted. It in turn led to an overestimate of surface O_3 over the urban sites mainly due to the reduced NOx titration effect, especially at night when there was no photochemical O_3 formation. The statistics showed that the bias of the modeled daytime (7 am ~ 7 pm local time) average surface O_3 was 30% ~ 90% smaller than that of the daily average in the urban sites, no matter which grid resolution was applied. This suggested that in the future the high-resolution emissions, especially proper representation of emission gradients, would be helpful in improving air quality prediction. The effect of emissions gradients associated with the grid resolution would be further discussed in the inter-model comparison section.

Next, the driving meteorology, especially wind, was important to accurately forecast air quality over coastal areas that bore sharp thermal contrasts. QHD site locates approximately 5 km from the ocean and is subject to sea breeze effects. The detailed analysis of meteorology and air quality over QHD was conducted. The results indicated that the choice of grid resolution had large impacts on model simulations at this coastal site. The selection of the 5-km grid reduced biases of both surface temperature and wind speed. The biases of temperature reduced from 1.22 K (45-km) to -0.42 K (15-km), and further down to -0.31 K when the 5-km grid was applied. The biases of surface wind speed for the 45-km, 15-km, and 5-km grids were 3.72, 4.19, and 1.95 m s^{-1}, respectively. The improvement of meteorology forecast helped reducing the biases of air quality modeling. The biases of O_3/NOx for the 45-km, 15-km, and 5-km resolution grids were 29.94/-22.46 ppbv, 24.09/-20.29 ppbv, 23.97/-17.95 ppbv, respectively. The improvement using the 15-km grid over the 45-km grid was remarkable but that using the 5-km grid over the 15-km grid was marginal. The result emphasized the importance of high-resolution modeling to improvements of air quality forecast skills, especially at coastal and complex terrain areas (e.g., QHD and XL).

Formatted: Subscript
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Subscript
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Justified
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Subscript
Formatted: Font: (Default) Times New Roman, 12 pt
Formatted: Font: (Default) Times New Roman, 12 pt

c. Extreme values

High concentrations of air pollutants are of more concerns because of their adverse health effects on both human beings and ecosystem. High pollutant concentrations also pose a greater risk for non-compliance of the ambient air quality standards. Therefore, evaluations of impacts of grid resolution on extreme concentrations of air pollutants are desirable.

Figure 5 displays the probability density function distributions of six pollutants based on hourly surface concentrations across the monitoring sites. This analysis was focused on high pollutant concentrations with the cutoff values for CO, O_3, NOx, SO_2, PM2.5, and PM10 being 1.1 ppmv, 60 ppbv, 25 ppbv, 5.5 ppbv, 15 μg m^{-3}, and 30 μg m^{-3}, respectively. It appeared that NU-WRF, regardless of the grid resolutions, failed to simulate surface CO with concentrations more than 4 ppmv, likely due to the underestimate of CO emissions (Kong et al., 2019). The grid resolution appeared to have limited impacts on surface PM10 simulations when its concentrations were more than 200 μg m^{-3}. On the other hand, the grid resolution showed large impacts on NU-WRF's capability in simulating high surface concentrations of O_3, NOx, SO_2, and PM2.5. For surface O_3 with concentrations more than 100 ppbv, the NU-WRF results with the 45-km grid appeared to better agree with the probability distribution of observations. For surface NOx with concentrations more than 70 ppbv, the NU-WRF results with the 5-km resolution grid better mimicked the observed distribution. Modeling with the 5-km grid also yielded the best results of distributions, in comparisons to the respective observations, of SO_2 with concentrations more than 45 ppbv, and of PM2.5 with concentrations greater than 120 μg m^{-3}.

Table 2 lists the occurrences of violations of China's national ambient air quality standards (NAAQS) for the six pollutants from both observations and simulations, in which columns "Class 1" and "Class 2" list the standards for rural and urban-suburban sites, respectively, and column "Frequency" indicates the time integration of each NAAQS. It was apparent that NU-WRF failed to report CO violations at any grid resolutions. No CO NAAQS violation was simulated but the observation showed that surface CO exceeded the national standard by more than 1000 times. NU-WRF underestimated the NAAQS exceedances of NOx and SO_2. A higher-resolution grid appeared to be able to catch more violations although the modeled results at the 5-km resolution only captured 33% and 10% observed exceedances of NOx and SO_2, respectively. NU-WRF overestimated surface O_3 and PM2.5 when their concentrations were more than the corresponding NAAQS. The fine grid resolution (i.e., 5-km) appeared to reduce the overestimation of surface O_3 exceedances largely as compared to the 45-km grid but only marginally compared with the 15-km grid. Compared to the observed occurrences of surface O_3 standard violation (3,684), the simulated exceedances were 6.7, 2.8, and 2.7 times higher when employing the 45-km, 15-km, and 5-km resolution grid, respectively. The observations showed 1,343 occurrences of surface PM2.5 exceedances, while the modeled exceedances were 377, 267, and 231 more for the 45-km, 15-km, and 5-km grids, respectively. As for surface PM10, the modeled exceedances were approximately 27%, 43%, and 41% less than the observed one for the 45-km, 15-km, and 5-km grids, respectively.

Deleted: .

3.2. Inter-resolution comparisons

It is informative to compare the NU-WRF results out of different horizontal resolutions. This, in addition to the discussion in section 3.1.2.b, can help understand the reasons why model resolution matters.

3.2.1. Emissions

There were two types of emissions applied in this study. One was the prescribed emissions out of the anthropogenic and wild fire sources, and the other was emissions computed online using the real-time meteorology (or dynamic emissions) including emissions from biogenic sources, dust sources, and sea spray. Amounts and temporal variations of dynamic emissions depended on surrounding environmental conditions. For example, air temperature and solar radiation regulates biogenic emissions (Guenther et al., 2006). Surface wind speed plays a major role in both dust (Ginoux et al., 2001; Chin et al., 2002) and sea salt emissions (Gong, 2003).

For the prescribed emissions, the differences of domain total masses out of each grid were small (less than 5%). However, the emission gradient around sources of a fine resolution grid appeared to be sharper than that of a coarse resolution grid. This meant that a coarse grid tended to distribute the prescribed emissions more evenly into the domain, while a fine grid tended to produce more extreme concentrations of primary pollutants (emitted directly from a source) such as NOx and SO_2, as shown in Table 2.

Online calculated emissions, on the other hand, displayed large differences in both gradient and total mass. Similar to the case of prescribed emissions, a fine resolution grid tended to give a sharper gradient of dynamic emissions than a coarse resolution grid did, as highlighted in Figure 6 (1st row) that illustrated the biogenic isoprene emissions (mol km^{-2} hr^{-1}) on a typical summer day. It was apparent that much more details were simulated using a fine resolution grid - the flow of Yellow River can even be seen on the 5-km resolution map that was otherwise invisible from the coarser resolution maps. Meanwhile, the total masses of dynamic emissions showed large difference out of different resolution grids as listed in Table 3. On an annual basis, the domain total isoprene emissions were 740,562 tons when estimated using the 45-km grid, approximately 85% and 86% of those with the 15-km and 5-km grids, respectively. The total dust emissions out of the 45-km grid were 2,431 tons, only 54% and 62% of those based on the respective 15-km and 5-km grids. The percentage contrasts for sea salt emissions were even larger with emissions out of the 15-km and 5-km grids being 1.3 and 1.6 times more than those of the 45-km grid, respectively. It should be noted that although they differed greatly between out of the 45-km and 15-km grids, the dynamic emissions out of the 5-km grid were much closer to those out of the 15-km grid, partially explaining why the impact of model resolution on surface air quality was less remarkable by increasing the resolution from 15-km to 5-km than from 45-km to 15-km.

The spatial (gradient) and mass variations in emissions out of different resolution grids would result in difference in air quality simulations.

3.2.2. Meteorology

It's been reported that simulated meteorology varies in response to selections of model grid resolutions (e.g., Tie et al., 2010; Lee et al., 2018). Meteorology plays an important role in regulating regional air quality – it affects emissions amount originating from biogenic, dust, and sea sources; it impacts atmospheric chemical and photochemical transformation; and it directs air flows and the associated transport of trace gases and aerosols. In this investigation, a few meteorological parameters key to air pollutant generation and accumulation were analyzed, including surface wind, air temperature, downward shortwave flux at surface (SWDOWN), planetary boundary layer height (PBLH), and cloud water (liquid + ice) path (CWP). We focused on months that were prone to deteriorated PM2.5 (January) and O_3 (July) air quality as shown in Figure 6 and Table 3.

NU-WRF simulated a similar direction of surface wind in July 2010 over the eastern portion of the domain (2nd row of Figure 6). In general, average wind speed was larger over Bohai Sea and Yellow Sea than over the surrounding land areas with dominating wind direction being south and southeast. Based on the results from the 15-km and 5-km grids, the peak average wind speeds over 4 m s^{-1} were found in Bohai Bay blowing to Tianjin and Beijing. However, such a peak was absent from the 45-km grid simulation. In the west portion of the domain, the wind direction changed from southeast in the south to southwest in the north in general. Compared to the more organized wind directions out of the 45-km grid, wind directions out of the 15- and 5-km grids were more chaotic. Averaged over the domain, the January mean wind speed out of the 45-km grid was 2.92 m s^{-1}, which were 7% and 16% larger than those of the 15-km and 5-km grids, respectively. The largest July mean wind speed was again simulated with the 45-km grid, 10% and 12% larger than the corresponding wind speed out of the 15-km and 5-km grids, respectively.

Overall, NU-WRF simulated very similar magnitudes and spatial patterns of surface air temperature in July (3rd row of Figure 6), regardless of the selections of grid resolutions. Large portions of the NCP experienced more than 300 K of July average air temperature. The minimum average temperature of approximately 290 K was found in the central north part of the domain, which was part of the Mongolian Plateau with the elevation being over 1,500 m above the sea level. The domain average January and July surface air temperature were around 268 K and 300 K, respectively, for simulations out of all three grids.

As expected, the modeling results from all three grids (4th row of Figure 6) showed that July average PBLH over sea was much smaller than that over land. The large average PBLH (more

Deleted: maximum

than 1,000 m) was found in the northwestern corner of the domain with a dominant land cover type of grassland mosaiced with open shrubland that appeared to be drier than the other land cover types in the domain. The high sensible heating associated with dry soil tended to produce the deep PBL (Tao et al., 2013). The largest domain-average PBLHs in January and July were found from the simulations out of the 15-km and 45-km grids, respectively. In January, the differences of the domain-average PBLHs from different grid resolutions were small and within 2%. In July, however, such difference can be over 9%.

Regardless of the grid resolutions, NU-WRF simulated a generally southeast-northwest gradient of SWDOWN in July with the highest flux (over 300 W m^{-2}) occurring in the northwestern domain (5th row of Figure 6). The differences between the maximum and minimum domain average SWDOWN out of 3 grids were 5.6% and 3.3% in January and July, respectively.

CWP represented the vertical integration of cloud water (including both liquid and ice phases) contents and can be regarded as a proxy of cloud amount and coverage. Opposite to the SWDOWN case, NU-WRF modeled a generally northwest-southeast gradient of CWP in July with the high values found in the southeastern domain (6th row of Figure 6). This was understandable since cloud reflects and scatters the incoming solar radiation and thus affect SWDOWN. Large cloud existence tended to reduce the solar flux reaching the underneath Earth surface. The CWP differences among the model results out of different grid resolutions appeared to be larger than SWDOWN differences. In July, the domain average CWPs out of the 15-km and 5-km grids were 37% and 33% larger than that of the 45-km grid, respectively. The gaps were even larger in January, during which the domain average CWPs from the 15-km and 5-km grids were approximately 1.6 times larger than that from the 45-km grid.

3.2.3. Air Quality

In response to the aforementioned emissions and meteorological variations resulted from the selections of model grid resolutions, changes in regional air quality ensued as illustrated in Figure 7 and Table 3. This figure shows the July average concentrations of ground-level O_3 and its precursors of NOx and CO, as well as the January mean concentrations of surface SO_2, PM2.5, and PM10, during which month the respective pollutants tended to reach high concentrations.

O_3 is a secondary pollutant that is formed in the atmosphere through complex photochemical processes upon existences of its precursors such as NOx and volatile organic compounds (VOC). Figure 7 (row 1) shows that the spatial distributions of surface O_3 are similar to each other but the concentrations out of the 15-km and 5-km grids are smaller than those from the 45-km grid. The domain average surface O_3 concentration in July was approximately 87 ppbv based on the results from the 45-km grid, 26% and 25% higher than those out of the 15-km and 5-km grid, respectively. In January, however, the highest domain average concentration occurred when the 5-km grid was used, which was 5.3% higher than that out of the 45-km grid.

For the primary pollutants, i.e., NOx, CO, and SO_2 (rows 2-4 of Figure 7, respectively), which were emitted directly by their sources, the spatial distributions of their concentrations mimicked closely with their emission distributions. High concentrations centered around emission sources with a reducing gradient outward. The domain average concentrations of these 3 pollutants out of the 45-km grid results were always the largest in both January and July. The average surface NOx concentrations from the simulations out of the 15-km and 5-km grids were around 24% lower than their counterparts out of the 45-km grid in January. In July, the differences were reduced to 7.9% and 11.8% for the 15-km and 5-km grids, respectively. On the other hand, the larger percentage differences, as compared to the results out of the 45-km grid, occurred in July than in

Deleted: portion

Deleted: , also in the Mongolian Plateau

January for both CO and SO_2. For example, the surface CO concentrations out of the 5-km grid were 12.3% and 30.6% lower than those based on the 45-km grid in January and July, respectively. The respective ground-level SO_2 concentrations from the 5-km grid were 20.5% and 38.9% lower than those from the 45-km grid in January and July.

It was interesting to note that among the 3 cases, the domain average July surface O_3 and NOx concentrations were both the highest out of the 45-km grid, contrary to the results discussed in section 3.1.2a where the highest O_3 concentration occurred out of the simulation using the 45-km grid while the highest NOx concentration happened with the 5-km grid. This seemingly contradicting result was internally consistent. Section 3.1.2a actually depicted the average surface concentrations in an urban environment (23 of 25 monitoring sites were in an urban/suburban setting), where surface O_3 formation was typically VOC controlled such that NO tended to consume O_3 through titrations. As discussed in section 3.2.1, a 5-km grid gave a much sharper emissions gradient with anthropogenic emissions concentrating in urban/suburban areas. This led to higher NOx concentrations around urban/suburban areas out of the simulation with the 5-km grid, which effectively resulted in lower O_3 concentrations there through the NO titration effect. The domain average discussed in this section, however, was the average covering the vast rural area that generally was NOx-limited such that surface O_3 formation was controlled by the availability of NOx – more NOx resulting in more O_3 through photochemical processes. In this case, the 45-km grid tended to distribute NOx emissions more evenly in the region, effectively decreasing the surface NOx concentration in urban areas but increasing it over rural areas. The larger average July wind speed simulated by the 45-km grid (Figure 6 and Table 3) further smoothed out the NOx distribution in NCP. This in turn increased the domain average surface O_3 concentration via photochemistry based on the 45-km resolution results. In addition, vertical lifting played an important role in explaining the maximum regional O_3 in July simulated by the 45-km grid as compared to the results by the other two grid resolutions. As displayed in Figure 4s in the supplement material, a fine resolution modeling (e.g., 5-km) tended to produce a stronger updraft than a coarse resolution modeling (e.g., 45-km), consistent with the findings by Lee et al. (2018). The strong uplift would bring more surface pollutants such as NOx into the upper atmosphere, thus further reducing the NOx availability at ground limiting the surface ozone production but increasing its formation in the upper atmosphere.

Vertical distributions of O_3 also tend to have a sizable impact on next day's surface O_3 levels (e.g., Kuang et al., 2011; Caputi et al., 2019). Figure 8 illustrates the domain average profiles of vertical wind, NOx, O_3 (panels a~c), and the average diurnal distribution of surface O_3 (panel d) over July. Here we limited our discussion on the results from the 15- and 5-km grids since 45-km grid artificially allowed more NOx emissions spreading to rural areas to produce much more O_3 as shown in the previous paragraph. Lee et al. (2018) claimed that a coarse resolution model appeared to lower updraft as compared with a fine resolution modeling. This study agreed with their finding as illustrated in Figure 8 (panel a). The domain average July vertical wind out of the simulation with the 5-km grid ranged from 0.25 to 0.45 cm s^{-1} (upward) between 800 hPa and 400 hPa, stronger than the corresponding one out of the 15-km grid. The reason was complex and the aerosol-cloud interaction induced freezing/evaporation-related invigoration mechanism played a role (Lee et al., 2018). The stronger upward wind tended to lift more gaseous pollutants up to the free troposphere as shown in Figure 8 (panel b (NOx) and c (O_3)). The pollutants there would have visible impacts on the following-day surface air quality, especially on O_3 levels at night and in the morning when sun breaks out the nocturnal planetary boundary layer, as evidenced in Figure 8 (panel d). At night with no photochemical formation, surface O_3 concentration was largely

Formatted: Subscript

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Formatted: Font: (Default) Times New Roman, 12 pt

Deleted: In addition, the higher air temperature and stronger SWDOWN in July out of the 45-km grid as compared to other two resolutions favored more surface O_3 generations.

controlled by upper-level O_3 mixing down, NO titration and O_3 dry deposition. With the virtually same average surface NO concentrations out of the 15- and 5-km grids, the upper-level O_3 mixing down appeared to control the relative magnitudes of surface O_3 concentrations simulated using the 15- and 5-km grids. This partially explained why, at night and early morning, the ground level O_3 concentrations were higher out of the 5-km grid than from the 15-km grid. During daytime when the photochemical formation of O_3 takes control, the regional average surface O_3 concentrations is largely determined by the availability of O_3 precursors (i.e., NOx and VOC) and ambient environmental conditions. In this case, more spreading NOx emissions out of the 15-km grid appeared to generate more surface O_3 than the 5-km grid did.

PM2.5 and PM10 were mixed pollutants that not only were emitted by various sources but also were generated in the atmosphere through physical and chemical processes. Figure 7 shows that high surface concentrations of PM2.5 (more than 120 $\mu g\ m^{-3}$, row 5) and of PM10 (more than 170 $\mu g\ m^{-3}$, row 6) were still found around the source areas based on the modeling results out of the 15-km and 5-km grids. However, high PM2.5 and PM10 concentrations spread out to larger areas based on the results from the 45-km grid as compared to the ones from the finer grid resolutions. Similar to the primary pollutants, the largest domain average surface concentrations occurred when a 45-km grid was used for the NU-WRF simulation. The domain average PM2.5 concentrations out of the 15-km and 5-km grids in January were 15.7% and 14% lower than those from the 45-km grid, respectively. The surface PM2.5 concentration differences among results out of different grid resolutions grew larger in July, reaching 48% when comparing the result from the 5-km grid to that from the 45-km grid. The domain average surface PM10 concentrations showed similar pattern to that of PM2.5 with the results out of the 5-km grid being 12.2% and 44.2% smaller than that from the 45-km grid.

It is worth noting that the magnitudes and spatial distributions of ground-level pollutants were close to each other between the results out of the 15-km and 5-km grids. This again indicates that the improvement of fine grid resolution modeling reduces at a certain point. In future MICS-Asia efforts, a 15-km grid appears to offer the optimized results balanced with performance and resources.

4. Summary

Contributing to MICS-Asia Phase III whose goals included identifying and reducing air quality modeling uncertainty over the region, this investigation examined the impact of model grid resolutions on the performances of meteorology and air quality simulation. To achieve this, NU-WRF was employed to simulate 2010 air quality over the NCP region with three grid resolutions of 45-km, 15-km, and 5-km. The modeling results were compared to the observations of surface meteorology archived by CMA, and of ground-level air quality collected in CERN. The inter-model comparison among the simulation results out of three grids were also conducted to understand the reasons why model resolution mattered.

The analysis showed that there was no single resolution which would yield the best reproduction of meteorology and air quality across all monitoring sites. From a regional average prospective (i.e., across all monitoring sites in this study), the choice of grid resolution appeared to have a minimum influence on air temperature modeling but affected wind, RH, and precipitation simulation profoundly. A 5-km grid appeared to give the best wind simulation as compared to the observations quantified by bias, RMSE, standard deviation, and correlation. Compared to the one using the 45-km grid, the simulated wind speed from a 5-km grid reduced the positive bias by 46.8%. While a 15-km grid yielded the best overall performance on RH modeling, the result out of the 45-km grid gave the most realistic reproduction of precipitation. The statement on precipitation should be taken with caution since it was based on the comparison with the site observations. Seeing the very heterogeneous nature of precipitation, the penalty of model hitting or missing a rain event was severe. Thus, the coarse grid covering more areas within a grid cell would reduce chances of mistaken precipitation hitting or missing simulations. However, a comparison of modeled precipitations to gridded "observation" that was re-constructed using the synergraphic mapping algorithm with topographic adjustment to the monthly precipitation climatology showed opposite result, where the fine resolution modeling showed superior reproduction of precipitation than the coarse resolution simulation (Gao et al., 2017).

The simulated meteorology differences due to the selection of grid resolution would consequently lead to differences in air quality simulation. Air pollutant concentrations were basically determined by their emissions and underlying meteorology that directed their formation (e.g., O_3 and aerosols), transport, and removal processes. For the prescribed emissions originated from anthropogenic and wild fire sources, the grid resolution had limited influence on emission amount – less than 5% difference with each other under the different resolution grids – but large impact on emission spatial distribution with sharper emission gradient around sources out of a fine resolution grid than from a coarse resolution one. For the dynamic emissions driven by meteorology, not only was an emission gradient around a source larger out of a higher resolution grid, but also the total emission amount varied greatly. For example, the domain total annual biogenic isoprene emissions from a 5-km grid was about 16% larger than those out of a 45-km grid due to the underlying differences in land cover and meteorology.

Though the impact of grid resolution on air quality varied from location to location, finer grid yielded better results for daily mean surface O_3, NOx, CO, and PM2.5 simulations from a regional average perspective. For example, after reducing the grid resolution from 45-km to 15-km, the positive bias of daily mean surface O_3 and PM2.5 decreased by 15% and 75%, respectively. Fine resolution modeling was especially beneficial to high pollutant concentration forecast. This was important to air quality management. Taking China's NAAQS as cutoff values for each pollutant, the frequencies of noncompliance occurrences of O_3, NOx, SO_2, and PM2.5 out of the 5-km grid simulation were much closer to the observations than those out of the 45-km modeling were. For example, the simulation with the 5-km grid produced 168% and 17% more exceedances in NAAQS of O_3 and PM2.5, respectively, whereas the respective exceedances were 573% and 28% more with modeling using the 45-km grid, as compared to the observed exceedances. It also was worth noting that the benefit of increasing grid resolution to better surface O_3 and PM2.5 simulations started to diminish when the horizontal resolution reached 15-km, agreeing with the finding by Valari and Menut (2008). There was a caveat, though. The anthropogenic MIX and fire GFEDv3 emissions inventories bore the 0.25º by 0.25º and 0.5º by 0.5º resolution, respectively. These resolutions cannot resolve the 5-km grid. Should a 5-km resolution emissions inventory be available and used, the benefit of high-resolution modeling would likely have been more prominent.

It should be pointed out that NU-WRF significantly overestimated surface O_3 concentration but underestimated ground-level CO and NOx concentrations regardless of grid resolutions. This was true not only on the regional averages but also at majority of the monitoring sites. The missing emissions was believed to be largely responsible for this result (Kong et al., 2019). Underestimate of surface NOx tended to increase ground-level O_3 due to the reduced titration effect, especially at night over urban areas that were typically NOx abundant.

Deleted: s

Deleted: -

Formatted: Subscript

Deleted:

Deleted: .

In conclusion, grid resolution had a profound effect on NU-WRF performance on meteorology and air quality over the East Asia. Fine resolution grid did not always generate the best modeling results and the proper selection of horizontal resolution hinged on investigation topics for a given set of physics and chemistry choices in a model. With regard to MICS-Asia Phase III whose major goal was to examine regional air quality, in general, the finer the grid resolution was, the better the simulation results would be. This was especially true over the coastal areas and complex terrains where a sharp local energy gradient existed. Fine resolution grid was also extremely helpful to reproducing pollutants at higher concentrations that were most relevant to air quality planning and management. However, the benefit of high resolution was not linear with the decrease of grid size. At certain point, the improved modeling accuracy due to an increase in grid resolution was so marginal that it cannot justify the computational cost associated with the fine grid simulation. Based on the balance of modeling accuracy and efficiency, a 15-km horizontal grid appeared to be an appropriate choice to optimize model performance and resource usage if the study domain remained unchanged for future MICS-Asia activities. The study suggested that the high-resolution emissions, especially the proper representation of emission gradients, would be helpful in improving air quality prediction. Moreover, the profile measurements of both meteorology and air quality, in supplement with the ground monitoring networks, would be greatly helpful to identifying model deficiencies and thus improving model forecast skills.

Deleted: .

Competing interests

The authors declare that they have no conflict of interest.

Author contribution

ZT and MC designed the experiments. ZT, MG, TK, DK, and HB carried out the experiments working on various modeling components. YW and ZL collected, organized, and archived the ground air quality measurement data. All authors contributed to model result analysis and interpretation. ZT prepared the manuscript with contributions from all co-authors.

Acknowledgement

This work was supported by the NASA's Atmospheric Composition: Modeling and Analysis Program (ACMAP) and Modeling, Analysis, and Prediction (MAP) program. The authors thank MICS-Asia for its organized platform of discussion and data sharing. This work is not possible without the supercomputing and mass storage support by NASA Center for Climate Simulation (NCCS). All data collected and generated for this research are archived and stored on NCCS servers. Due to the sheer size of data, it is impractical to upload data to a public domain repository. However, the authors will be happy to share data on an individual request basis.

References

Akimoto, H., Mori, Y., Sasaki, K., Nakanishi, H., Ohizumi, T., and Itano, Y.: Analysis of monitoring data of ground-level ozone in Japan for long-term trend during 1990-2010: Causes of temporal and spatial variation, Atmos. Environ., 102, 302-310, 2015.

Anenberg, S. C., Horowitz, L. W., Tong, D. Q., and West J. J.: An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling, Environ. Health Perspect., 118(9), 1189-1195, doi:10.1289/ehp.0901220, 2010.

Cai, W., Li, K., Liao, H., Wang, H., and Wu, L.: Weather conditions conductive to Beijing severe haze more frequent under climate change, Nature Climate Change, 7, 257-263, doi:10.1038/NCLIMAE3249, 2017.

Caputi, D. J., Faloona, I., Trousdell, J., Smoot, J., Falk, N., and Conley, S.: Residual layer ozone, mixing, and the nocturnal jet in California's San Joaquin Valley, Atmos. Chem. Phys., 19, 4721-4740, doi:10.5194/acp-19-4721-2019, 2019.

Carmichael, G., Calori, G., Hayami, H., Uno, I., Cho, S. Y., Engardt, M., Kim, S. B., Ichikawa, Y., Ikeda, Y., Woo, J. H., Ueda, H., and Amann, M.: The MICS-Asia study: model intercomparison of long-range transport and sulfur deposition in East Asia, Atmos. Environ., 36, 175-199, 10.1016/s1352-2310(01)00448-4, 2002.

Carmichael, G., Sakurai, T., Streets, D., Hozumi, Y., Ueda, H., Park, S., Fung, C., Han, Z., Kajino, M., and Engardt, M.: MICS-Asia II: The model intercomparison study for Asia Phase II: methodology and overview of findings, Atmos. Environ., 42, 3468-3490, 10.1016/j.atmosenv.2007.04.007, 2008.

Chen, L., Gao, Y., Zhang, M., Fu, J. S., Zhu, J., Liao, H., Li, J., Huang, K., Ge, B., Wang, X., LAM, Y. F., Lin, C.-Y., Itahashi, S., Nagashima, T., Kajino, M., Yamaji, K., Wang, Z., and Kurokawa, J.-I.: MICS-Asia III: Multi-model comparison and evaluation of aerosol over East Asia, Atmos. Chem. Phys. Discuss., http://doi.org/10.5194/acp-2018-1346, in review, 2019.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, J. Atmos. Sci., 59, 461–483, 2002.

Chin, M., Diehl, T., Ginoux, P., and Malm, W.: Intercontinental transport of pollution and dust aerosols: implications for regional air quality, Atmos. Chem. Phys., 7, 5501-5517, doi:10.5194/acp-7-5501-2007, 2007.

Chou, M.-D. and Suarez, M. J.: A solar radiation parameterization (CLIRAD-SW) for atmospheric studies, NASA Tech. Rep. NASA/TM-1999-10460, vol. 15, 38 pp, 1999.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model, J. Geophys. Res., 108, 8851, doi:10.1029/2002JD003296, 2003.

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), Geosci. Model Dev., 3, 43-67, 2010.

Gao, M., Carmichael, G. R., Wang, Y., Saide, P. E., Yu, M., Xin, J., Liu, Z., and Wang, Z.: Modeling study of the 2010 regional haze event in the North China Plain, Atmos. Chem. Phys., 16, 1673-1691, doi:10.5194/acp-16-1673-2016, 2016.

Gao, M., Han, Z., Liu, Z., Li, M., Xin, J., Tao, Z., et al.: Air quality and climate change, Topic 3 of the Model Inter-Comparison study for Asia Phase III (MICS-Asia III) – Part 1: Overview and model evaluation, Atmos. Chem. Phys., 18 (7), 4859-4884, doi:10.5194/acp-18-4859-2018, 2018.

Gao, Y., Leung, L. R., Zhao, C., and Hagos, S.: Sensitivity of U.S. summer precipitation to model resolution and convective parameterizations across gray zone resolution, J. Geophys. Res. Atmos., 122, 2,714-2,733, doi:10.1002/2016JD025896, 2017.

Ginoux, P., Chin, M., Tegen, I., Prospero, J., Holben, B., Dubovik, O., and Lin, S.-J.: Sources and global distributions of dust aerosols simulated with the GOCART model, J. Geophys. Res., 106, 20,255-20,273, 2001.

Gong, S. L.: A parameterization of sea-salt aerosol source function for sub- and super-micron particles, Glob. Biogeochemical Cycles, 17, No. 4, 1097, doi: 10.1029/2003GB002079, 2003.

Grell, G. A. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett., 29, 1693, doi:10.1029/2002GL015311, 2002.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock W. C., and Eder, B.: Fully coupled ''online'' chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, 2005.

Gross, A. and Stockwell, W. R.: Comparison of the EMEP, RADM2 and RACM Mechanisms, J. Atmos. Chem., 44, 151-170, 2003.

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), Atmos. Chem. Phys., 6, 3181-3210, 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, 2006.

Huang, R.-J., Zhang, Y., Bozzetti, C., Ho, K.-F., Cao, J.-J., Han, Y., Daellenbach, K. R., Slowik, J. G., Platt, S. M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S. M., Bruns, E. A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., An, Z., Szidat, S., Baltensperger, U., Haddad, I. E., and Prevot, A. S. H.: High secondary aerosol contribution to particulate pollution during haze events in China, Nature, 514, 218-222, doi:10.1038/nature13774, 2014.

Jin, Y., Andersson, H., and Zhang, S.: Air pollution control policies in China: A retrospective and prospects, Int. J. Environ. Res. Public Health, 13(12), 1219, doi:10.3390/ijerph13121219, 2016.

Kim, D., Chin, M., Kemp, E. M., Tao, Z., Peters-Lidard, C. D., and Ginoux, P.: Development of high-resolution dynamic dust source function – A case study with a stong dust storm n a regional model, Atmos. Environ., 159, 11-25, doi:10.1016/j.atmosenv.2017.03.045, 2017.

Kong, L., Tang, X., Zhu, J., Wang, Z., Fu, J. S., Wang, X., Itahashi, S., Yamaji, K., Nagashima, T., Lee, H.-J., Kim, C.-H., Lin, C.-Y., Chen, L., Zhang, M., Tao, Z., Li, J., Kajino, M., Liao, H., Sudo, K., Wang, Y., Pan, Y., Tang, G., Li, M., Wu, Q., Ge, B., and Carmichael, G. R.: Evaluation and uncertainty investigation of the NO_2, C and NH_3 modeling over China under the framework of MICS-Asia III, Atmos. Chem. Phys. Discuss., http://doi.org/10.5194/acp-2018-1158, accepted, 2019.

Deleted: in review

Krotkov, N. A., McLinden, C. A., Li, C., Lamsal, L. N., Celarier, E. A., Marchenko, S. V., Swartz, W. H., Bucsela, E. J., Joiner, J., Duncan, B. N., Boersma, K. F., Veefkind, J. P., Levelt, P. F., Fioletov, V. E., Dickerson, R. R., He, H., Lu, Z., and Streets, D. G.: Aura OMI observations of regional SO_2 and NO_2 pollution changes from 2005 to 2015, Atmos. Chem. Phys., 16(7), 4605–4629, doi:10.5194/acp-16-4605-2016, 2016.

Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J. Olden, S., Lighty, L., Eastman, J. L., Doty, B., Dirmeyer P., Adams, J., Mitchell K., Wood, E. F., and Sheffield, J.: Land Information System – An Interoperable Framework for High Resolution Land Surface Modeling, Environ. Modelling & Software, 21, 1,402-1,415, 2006.

Kuang, S., Newchurch, M. J., Burris, J., Wang, L., Buckley, P. I., Johnson, S., Knupp, K., Huang, G., Phillips, D., and Canrell, W.: Nocturnal ozone enhancement in the lower troposphere observed by lidar, Atmos. Environ., 45, 6078-6084, 2011.

Lee, S. S. Li, Z., Zhang, Y. Yoo, H., Kim, S., Kim, B.-G., Choi, Y.-S., Mok, J., Um, J., Choi, K. O., and Dong, D.: Effects of model resolution and parameterizations on the simulations of clouds, precipitation, and their interactions with aerosols, Atmos. Chem. Phys., 18, 13–29, doi:10.5194/acp-18-13-2018, 2018.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality on a global scale, Nature, 525, 367-371, 2015.

Li, J., Nagashima, T., Kong, L., Ge., B., Yamaji, K., Fu, J. S., Wang, X., Fan, Q., Itahashi, S., Lee, H.-J., Kim, C.-H., Lin, C.-Y., Zhang, M., Tao, Z., Kajino, M., Liao, H., Li, M., Woo, J.-H., Kurokawa, J., Wang, Z., Wu, Q., Akimoto, H., Carmichael, G. R., and Wang, Z.: Model evaluation and intercomparison of surface-level ozone and relevant species in East Asia in the context of MICS-Asia phase III, Part I: overview, Atmos. Chem. Phys., 19, 12993-13015, http://doi.org/10.5194/acp-19-12993-2019, 2019.

Li, M., Zhang, Q., Kurokawa, J. I., Woo, J. H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17, 935-963, doi:10.5194/acp-17-935-2017, 2017.

Lin, M., Holloway, T., Carmichael, G. R., and Fiore, A. M.: Quantifying pollution inflow and outflow over East Asia in spring with regional and global models, Atmos. Chem. Phys., 10, 4221-4239, doi: 10.5194/acp-10-4221-2010, 2010.

Lu, X., Hong, J., Zhang, L., Cooper, O. R., Schultz, M. G., Xu, X., Wang, T., Gao, M., Zhao, Y., and Zhang, Y.: Severe surface ozone pollution in China: A global perspective, Environ. Sci. Technol. Lett., 5, 487-494, 2018.

Matsui, T., Tao, W.-K., Masunaga, H., Kummerow, C. D., Olson, W. S., Teruyuki, N., Sekiguchi, M., Chou, M., Nakajima, T. Y., Li, X., Chern, J., Shi, J. J., Zeng, X., Posselt, D. J., and Suzuki, K.: Goddard Satellite Data Simulation Unit: Multi-Sensor Satellite Simulators to Support Aerosol- Cloud-Precipitation Satellite Missions, Eos Trans., 90(52), Fall Meet. Suppl., Abstract A21D-0268, 2009.

Matsui, T., Iguchi, T., Li, X., Han, M., Tao, W.-K., Petersen, W., L'Ecuyer, T., Meneghini, R., Olson, W., Kummerow, C. D., Hou, A. Y., Schwaller, M. R., Stocker, E. F., and Kwiatkowski, J.: GPM satellite simulator over ground validation sites, Bull. Am. Meteor. Soc., 94, 1653-1660, doi:10.1175/BAMS-D-12-00160.1, 2013.

Matui, T., Santanello, J., Shi, J. J., Tao, W.-K., Wu, D., Peters-Lidard, C., Kemp, E., Chin, M., Starr, D., Sekiguchi, M., and Aires, F.: Introducing multi-sensor satellite radiance-based evaluation for regional earth system modeling, J. Geophys. Res., 119, 8450-8475, doi:10.1002/2013JD021424, 2014.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B.: A long-term

Deleted: -I.

Deleted: -

Deleted: . Discuss.

Field Code Changed

Deleted: 2018-1283, in review,

Formatted: Default Paragraph Font, Font: (Default) +Body (Calibri)

hydrologically based dataset of land surface fluxes and states for the continuous United State, J. Clim., 15(22), 3,237-3,251, 2002.

Mu, M., Randerson, J. T., van der Werf, G. R., Giglio, L., Kasibhatla, P., Morton, D., Collatz, G. J., DeFries, R. S., Hyer, E. J., Prins, E. M., Griffith, D. W. T., Wunch, D., Toon, G. C., Sherlock, V., and Wennberg, P. O.: Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model prediction of carbon monoxide, J. Geophys. Res., 116, D24303, doi:10.1029/2011JD016245, 2011.

Neal, L. S., Dalvi, M., Folberth, G., McInnes, R. N., Agnew, P., O'Connor, F. M., Savage, N. H., and Tilbee, M.: A description and evaluation of an air quality model nested within global and regional composition-climate models using MetUM, Geosci. Model Dev., 10, 3941-3962, doi:10.5194/gmd-10-3941-2017, 2017.

Peters-Lidard, C. D., Houser, P. R., Tian, Y., Kumar, S. V., Geiger, J., Olden, S., Lighty, L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E. F., and Sheffield, J.: High-performance Earth system modeling with NASA/GSFC's Land Information System, Innovations in Systems and Software Engineering, 3, 157–165, 2007.

Peters-Lidard, C. D., Kemp, E. M., Matsui, T., Santanello, J. A., Kumar, S. V., Jacob, J. P., Clune, T., Tao, W.-K., Chin, M., Hou, A., Case, J. L., Kim, D., Kim, K.-M., Lau, W., Liu, Y., Shi, J., Starr, D., Tan, Q., Tao, Z., Zaitchik, B. F., Zavodsky, B., Zhang, S. Q., and Zupanski, M.: Integrated modeling of aerosol, cloud, precipitation and land processes at satellite-resolved scales. Environmental Modeling & Software 67, 149-159, 2015.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA - NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24, 3624–3648, 2011.

Seo, J., Youn, D., Kim, J. Y., and Lee, H.: Extensive spatiotemporal analyses of surface ozone and related meteorological variables in South Korea for the period 1999-2010, Atmos. Chem. Phys., 14, 6395-6415, doi:10.5194/acp-14-6395-2014, 2014.

Shi, J. J., Matsui, T., Tao, W.-K., Tan, Q., Peters-Lidard, C. D., Chin, M., Pickering, K., Guy, N., Lang, S., and Kemp, E. M.: Implementation of an aerosol-cloud microphysics-radiation coupling into the NASA Unified WRF: Simulation results for the 6-7 August 2006 AMMA Special Observing Period, Q. J. R. Meteorol. Soc., doi:10.1002/qj.2286, 2014.

Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X.: The Second Generation Regional Acid Deposition Model Chemical Mechanism for Regional Air Quality Modeling, J. Geophys. Res., 95, 16343-16367, 1990.

Tao, W.-K., Shi, J. J., Chen, S. S., Lang, S., Lin, P.-L., Hong, S.-Y., Peters-Lidard, C., and Hou, A.: The impact of microphysical schemes on hurricane intensity and track, Asia-Pacific J. Atmos. Sci., 47, 1–16, 2011.

Tao, Z., Santanello, J. A., Chin, M., Zhou, S., Tan, Q., Kemp, E. M., and Peters-Lidard, C. D.: Effect of land cover on atmospheric processes and air quality over the continental United States – A NASA Unified WRF (NU-WRF) model study, Atmos. Chem. Phys., 13, 6207-6226, doi: 10.5194/acp-13-6207-2013, 2013.

Tao, Z., Yu, H., and Chin, M.: Impact of transpacific aerosol on air quality over the United States: A perspective from aerosol-cloud-radiation interactions, Atmos. Environ., 125, 48-60, doi:10.1016/j.atmosenv.2015.10.083, 2016.

Tao, Z., Braun, S. A., Shi, J. J., Chin, M., Kim, D., Matsui, T., and Peters-Lidard, C. D.: Microphysics and radiation effect of dust on Saharan air layer: An HS3 case study, Mon. Wea. Rev., 146, 1813-1835, doi:10.1175/MWR-D-17-0279.1, 2018.

Tie, X., Brasseur, G., and Ying, Z.: Impact of model resolution on chemical ozone formation in Mexico City: application of the WRF-Chem model, Atmos. Chem. Phys., 10, 8983-8995, doi:10.5194/acp-10-8983-2010, 2010.

Valari, M. and Menut, L.: Does an increase in air qulaity model's resolution bring surface ozone concentrations closer to reality?, J. Atmos. Oceanic Tech., 25, 1955-1968, doi:10.1175/2008JTECHA1123.1, 2008.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11707-11735, doi:10.5194/acp-10-11707-2010, 2010.

Wang, W.-N., Cheng, T.-H., Gu, X.-F., Chen, H., Guo, H., Wang, Y., Bao, F.-W., Shi, S.-Y., Xu, B.-R., Zuo, X., Meng, C., and Zhang, X.-C.: Assessing spatial and temporal patterns of observed ground-level ozone in China, Sci. Rep., 7, 3651, doi:10.1038/s41598-017-03929-w, 2017.

Yu, Man: An assessment of urbanization impact on China by using WRF-Chem and configuration optimization, PhD (Doctor of Philosophy) thesis, University of Iowa, https://doi.org/10.17077/etd.lzfu2tj8, 2014.

Zhao, X. J., Zhao, P. S., Xu, J., Meng, W., Pu, W. W., Dong, F., He, D., and Shi, Q. F.: Analysis of a winter regional haze event and its formation mechanism in the North China Plain, Atmos. Chem. Phys., 13, 5685-5696, doi:10.5194/acp-13-5685-2013, 2013.

Zou, Y., Wang, Y., Zhang, Y., and Koo, J.-H.: Arctic sea ice, Eurasia snow, and extreme winter haze in China, Sci. Adv., 3, e1602751, 2017.

Table 1. Information of Air Quality Observation Sites

Site Name	Symbol	Longitude	Latitude	Altitude (m)	Setting
Baoding	BD	115.441	38.824	4	Urban
Beijing Tower	BJT	116.372	39.974	44	Urban
Chengde	CD	117.925	40.973	395	Urban
Caofeidian	CFD	118.442	39.270	11	Urban
Cangzhou	CZ	116.779	38.286	12	Urban
Datong	DT	113.389	40.089	1058	Urban
Gu An	GA	115.734	39.149	21	Rural
Hejian	HJ	116.079	38.423	66	Urban
Hengshui	HS	115.656	37.742	77	Urban
Langfang	LF	116.689	39.549	19	Urban
Lingshan	LS	115.431	39.968	116	Rural
Longtan Lake	LTH	116.430	39.870	31	Urban
Qian An	QA	118.800	40.100	54	Urban
Qinhuangdao	QHD	119.570	39.950	2.4	Urban
Shijiazhuang	SJZ	114.529	38.028	70	Urban
Shuangqing Road	SQL	116.338	40.007	58	Urban
Tanggu	TG	117.717	39.044	13	Urban
Tianjin	TJ	117.206	39.075	2	Urban
Tangshan	TS	118.156	39.624	14	Urban
Xianghe	XH	116.962	39.754	9	Suburban
Xinglong	XL	117.576	40.394	879	Rural
Yangfang	YF	116.126	40.147	78	Suburban
Yanjiao	YJ	116.824	39.961	26	Suburban
Zhangjiakou	ZJK	114.918	40.771	777	Urban
Zhuozhou	ZZ	115.988	39.460	48	Suburban

Table 2. Comparisons of occurrences of exceedances of China's National Ambient Air Quality Standards between observations and simulations*

	Frequency	Class 1	Class 2	Obs.	45-km	15-km	5-km
CO	Hourly	10	10	1,150	0	0	0
O_3	Hourly	160	200	3,684	24,807	10,283	9,880
NOx	Hourly	250	250	9,009	14	520	3,003
SO_2	Hourly	150	500	393	0	2	39
PM2.5	24-hours	35	75	1,343	1,720	1,610	1,574
PM10	24-hours	50	150	2,834	2,067	1,617	1,676

* Class 1/2 standards are for rural/suburban-urban, respectively. Units are ppbm for CO, ppbv for O_3, NOx, and SO_2, and μg m^{-3} for PM2.5 and PM10

Formatted: Subscript

Formatted: Subscript

Deleted: .

Table 3. Domain total emissions and average meteorology and air quality at various resolutions

Deleted: Regional

Variables	Period	45-km	15-km	5-km
Biogenic Isoprene (tons)	Annual	740,562	869,317	862,199
Dust (tons)	Annual	2,431	4,485	3,910
Sea salt (tons)	Annual	548	1,287	1,417
Surface air temperature (K)	January	268	267	268
	July	300	299	299
Surface wind speed ($m\ s^{-1}$)	January	2.92	2.73	2.51
	July	1.70	1.54	1.52
SWDOWN ($W\ m^{-2}$)	January	124	117	117
	July	249	242	250
PBLH (m)	January	333	338	331
	July	627	595	574
CWP ($g\ m^{-2}$)	January	4.34	11.3	11.1
	July	41.4	56.8	55.2
Surface O_3 (ppbv)	January	37.5	39.4	39.5
	July	86.8	68.8	69.2
Surface NOx (ppbv)	January	19.8	14.9	15.0
	July	9.03	8.32	7.96
Surface CO (ppmv)	January	0.600	0.521	0.526
	July	0.444	0.336	0.308
Surface SO_2 (ppbv)	January	16.6	12.9	13.2
	July	10.2	6.55	6.23
Surface PM2.5 ($\mu g\ m^{-3}$)	January	70.9	59.8	61.0
	July	89.3	58.0	46.2
Surface PM10 ($\mu g\ m^{-3}$)	January	102	88.1	89.6
	July	108	74.9	60.3

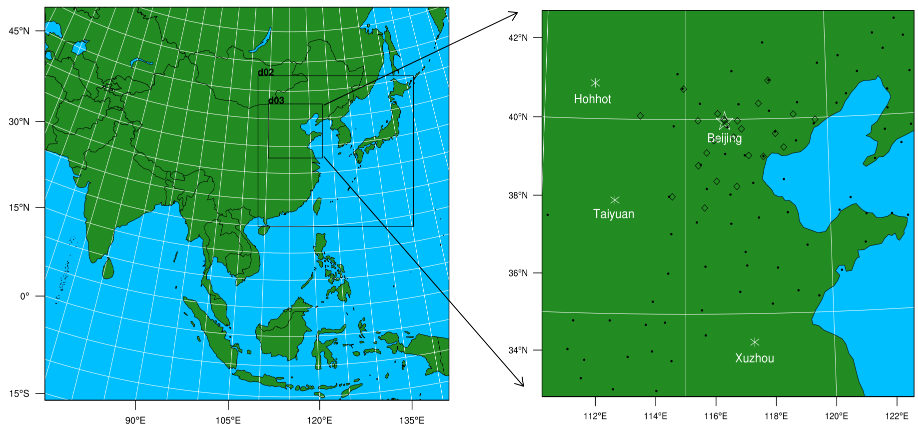


Figure 1. NU-WRF domain set-up. Left panel is the nested MICS-Asia domains; right panel is the enlarged NCP domain (d03) with diamonds representing the air quality monitoring sites and black dots denoting for the meteorological stations. Locations of four cities are marked for position reference.

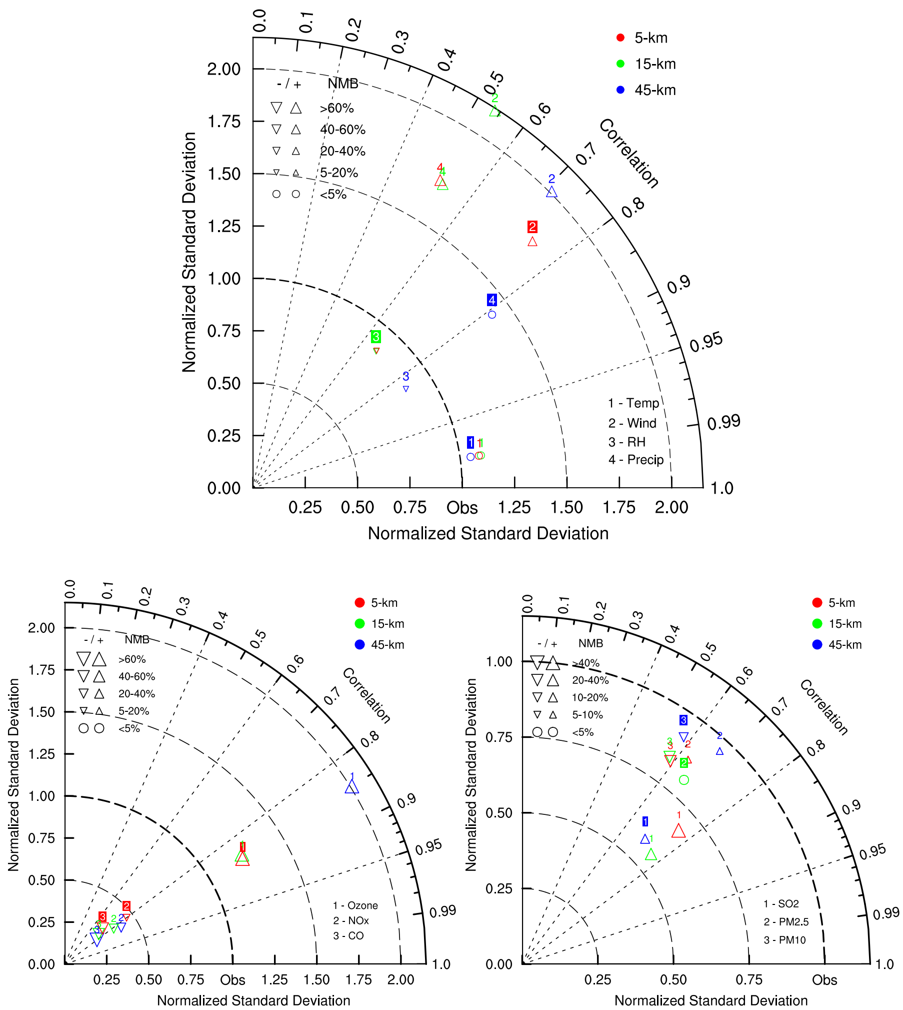


Figure 2. Taylor diagram for evaluations of NU-WRF performances on meteorology (top row) and air quality (bottom row) simulations at three resolutions

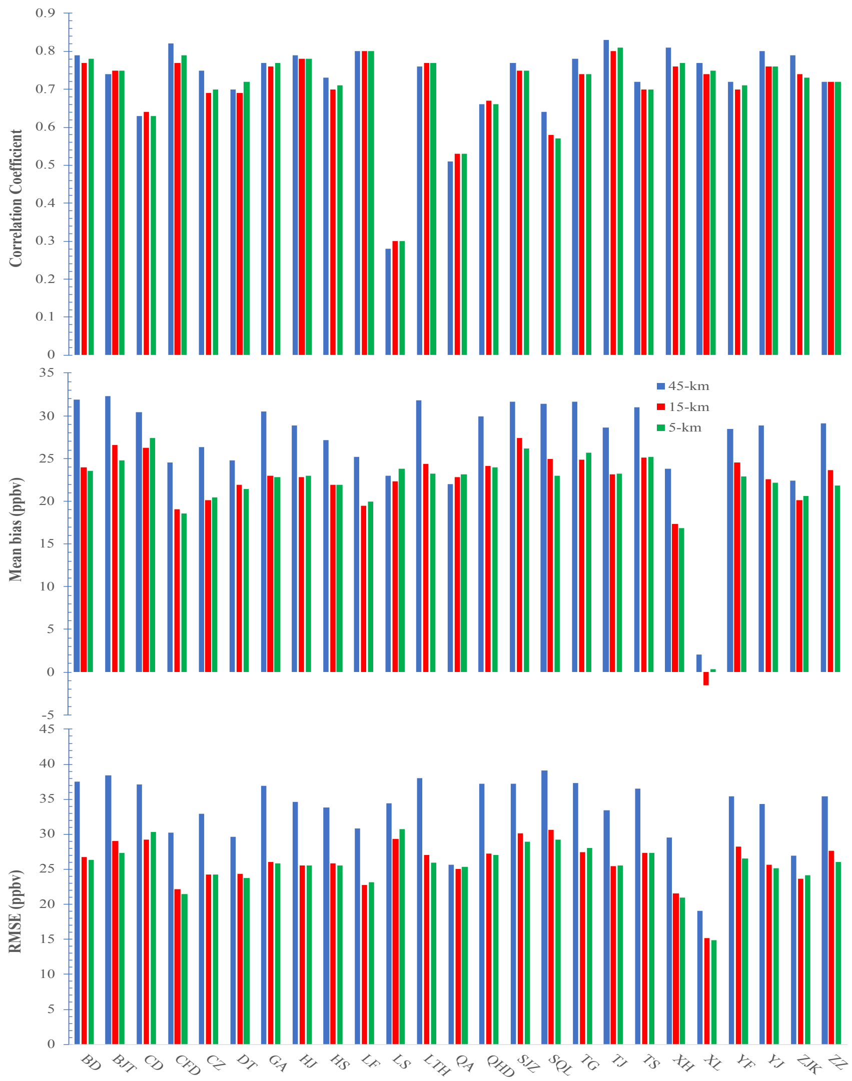


Figure 3. Comparisons of *MB*, *RMSE*, and correlation coefficient (r) of surface O_3 from different horizontal resolutions at each air quality monitoring site

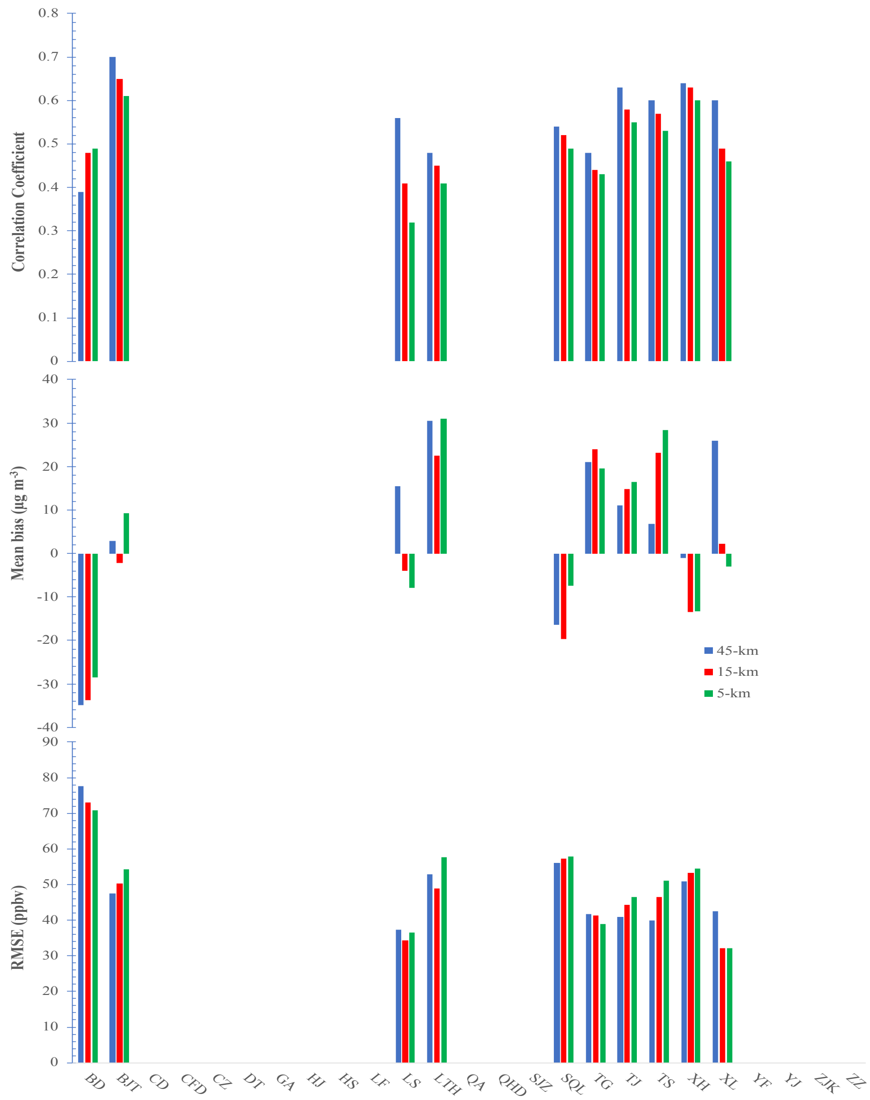


Figure 4. Comparisons of *MB*, *RMSE*, and correlation coefficient (r) of surface PM2.5 from different horizontal resolutions at each air quality monitoring site (blank space indicates no data are available)

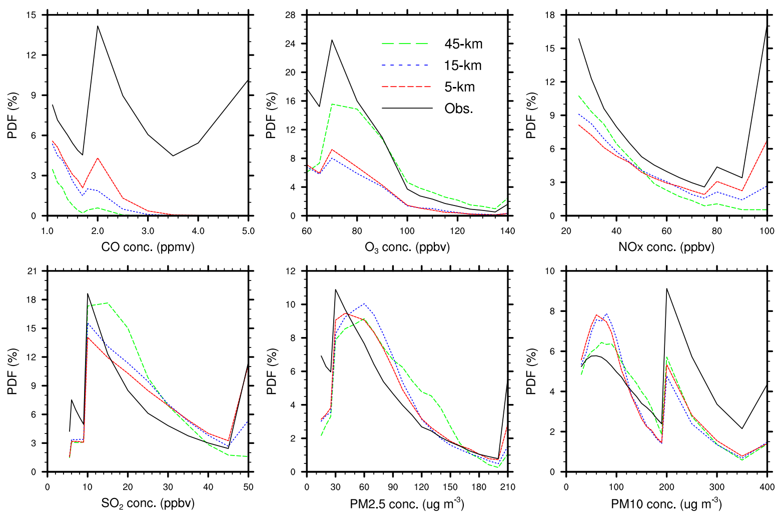


Figure 5. Probability density function (PDF) plots for hourly concentrations of surface air quality

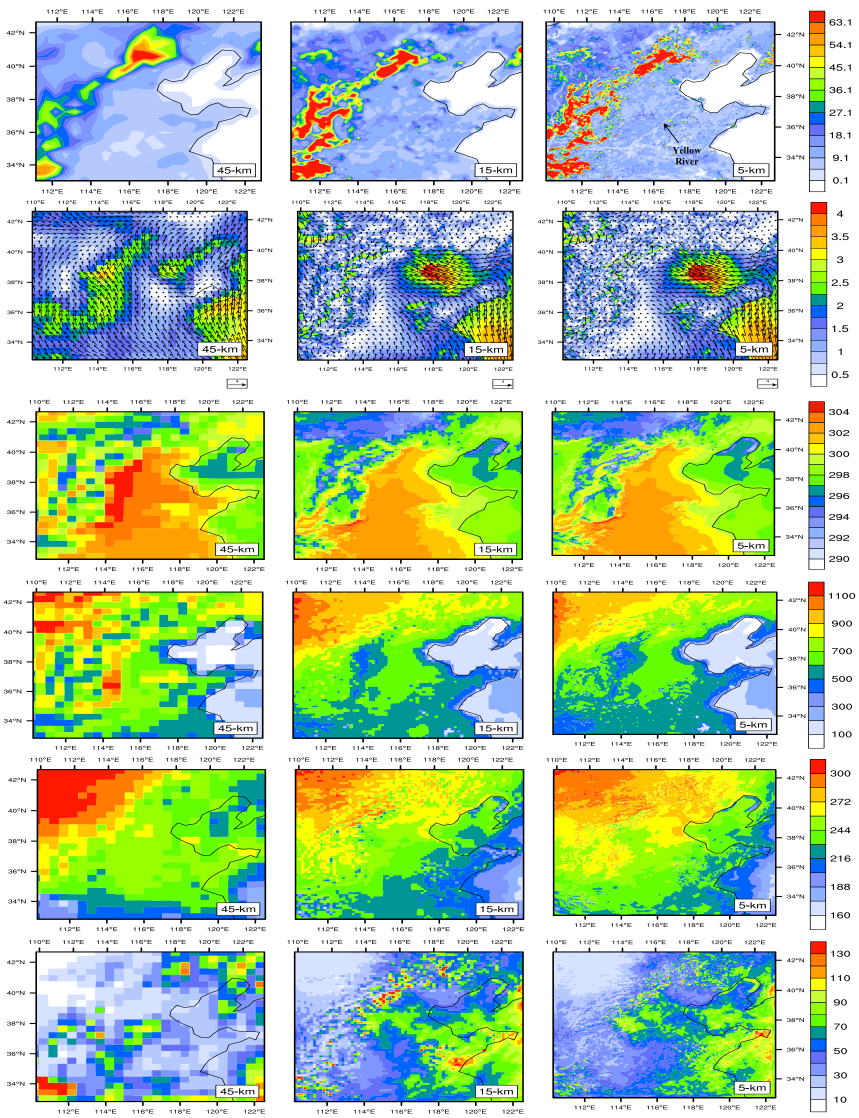


Figure 6. Simulated emissions and July average meteorology from 3 grids: 1st row = isoprene emissions (mol km^{-2} hr^{-1}) from biogenic sources on a typical summer day; 2nd row = surface wind vector with the shade representing wind speed (m s^{-1}); 3rd row = surface air temperature (K); 4th row = PBLH (m); 5th row = SWDOWN (W m^{-2}); 6th row = CWP (g m^{-2}).

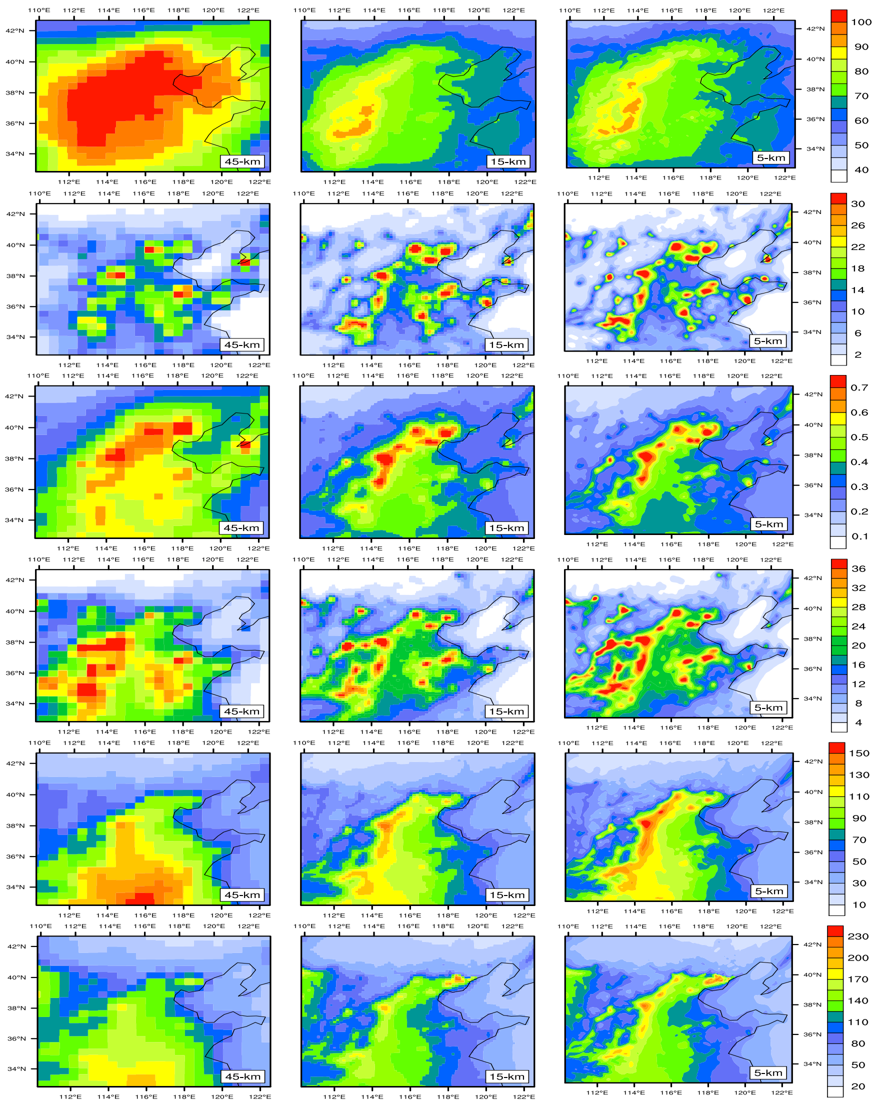


Figure 7. Simulated January (SO_2, PM2.5, and PM10) and July (O_3, NOx, and CO) surface average air quality from 3 grids: 1st row = O_3 (ppbv); 2nd row = NOx (ppbv) 3rd row = CO (ppmv); 4th row = SO_2 (ppbv); 5th row = PM2.5 ($\mu g\ m^{-3}$); 6th row = PM10 (($\mu g\ m^{-3}$).

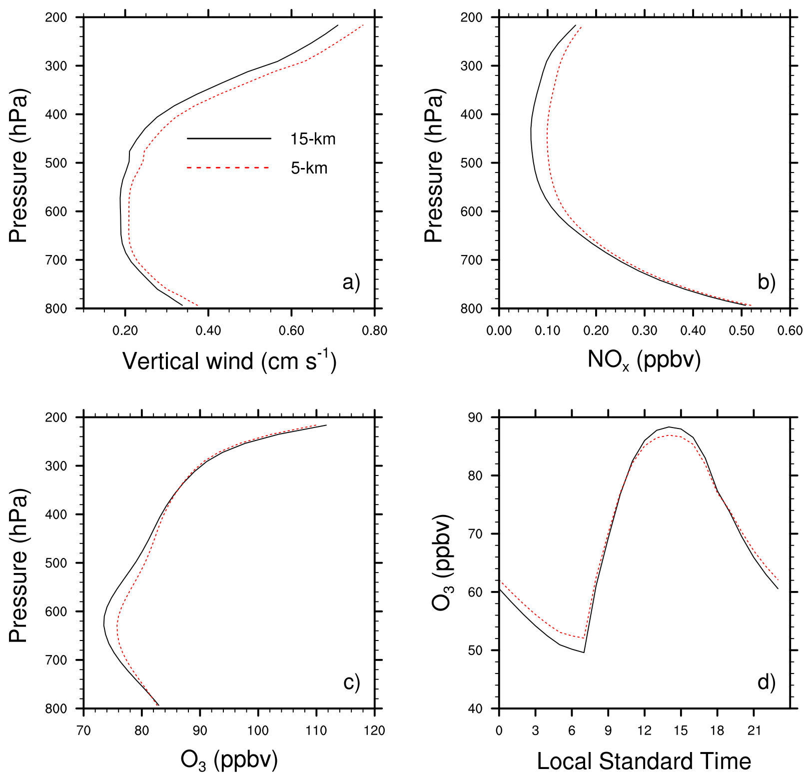


Figure 8. Domain average profiles of vertical wind, NOx, and O3 concentrations (Panels a~c) and domain average diurnal variations of surface O_3 over July.