Reply to Comments from Reviewer #1

The referee comments are shown in blue.

The responses to the comments are shown in black.

The text included in the revised manuscript are shown in red.

Reviewer #1 comments:

This study uses Global Forecast System with the new Finite Volume Cube-Sphere dynamic core (GFS-FV3) to drive the CMAQ v5.0.2 and evaluates the model results with observational data. The forecast system shows good agreement in meteorological variables and pollutants. This manuscript fits the scope of the journal of Geoscientific Model Development. However, more detailed and in-depth descriptions are expected in several places (see below comments).

Response:

Thank you for your positive comments. Please see below our point-by-point responses.

Specific comments

1. Line27-28: As mentioned here, NAM model is the current meteorological model. Can authors explain more in-depth why chose the GFS-FV3? Please also present some comparison results between these two models.

Response: The necessity and urgency to develop a FV3GFS-CMAQ system is owing to the retirement of the North American Mesoscale forecasting system (NAM) in the NOAA National Weather Service (NWS). The NAM model is no longer being updated (as of 2017). The current regional models for providing high-resolution forecast and guidance in NWS will be eventually replaced by a FV3-based system by 2023. The

FV3GFS-CMAQ system studied in this work is essentially a replacement of the NAQFC's offline coupled to a meteorological driver system by swapping NAM by FV3GFS.

The meteorological performance from the two systems, NAM and FV3GFS, is comparable (see, e.g., https://hmt.noaa.gov/experiments/pdf/WPC-HMT_WWE_2019_Final_Report.pdf). Huang et al. (2019) and Lee et al. (2020, https://www.cmascenter.org/conference/2020/session-presentations10.cfm) provided an air-quality-model-specific performance evaluation when a CMAQ-based Chemical Transport Model was driven by the NAM and FV3GFS meteorology. Although we did not compare performance of NAM versus FV3GFS, the FV3GFS-CMAQ interim NAQFC-β system we analyzed in this paper showed an across-board improvement in terms of the major chemical evaluation performance statistics than that by the NAM driven operational NAQFC.

To address the reviewer's comments, we have added a brief summary on the comparison of NAM and FV3 based on previous publications and reports, see below: the GFS v15 gave lower T2 prediction over CONUS comparing to NAM. While the NAM had slightly biased high T2, the GFS v15 had slightly biased low T2 during Aug 2019. The T2 underprediction was more significant in the Midwest by GFS v15 than its by NAM, especially during daytime. During the 2018-19 winter season, the FV3GFS had similar statistical scores regarding performance in snowfall prediction. While both FV3GFS and NAM gave over-forecasting in accumulated snow under most of precipitation type methods, the FV3GFS had larger over-prediction than NAM. It indicated FV3GFS had a colder forecast. The authors attributed the larger overprediction in snow depth to the consistent cold bias in GFS v15, which was identified by NWS through the intercomparison between GFS v14 and v15.

2. Line 50: "semi- or intermediate-VOCs" is mentioned as the missing sources of $PM_{2.5}$ in abstract. However, none of the S/I VOCs sources are analyzed in the rest of the manuscript. Please double check whether the analysis of this part is omitted.

Response: We have added some discussions on the impact of missing S/IVOCs and related SOA chemistry, see our response #2 to the reviewer #2's comments and also see below the revised text in the manuscript:

In CMAQ v5.0.2, the primary organic aerosol (POA) is processed as non-volatile. The emissions of semivolatile and intermediate volatility organic compounds (S/IVOCs) and their contributions to the secondary organic aerosol (SOA) are not accounted for in the aerosol module. In the recent versions of CMAQ, two approaches linked to POA sources have been implemented. One introduces semi-volatile partitioning and gas-phase oxidation of POA emissions. The other (called pcSOA) accounts for multiple missing sources of anthropogenic SOA formation, including potential missing oxidation pathways and emissions of IVOCs. These two improvements lead to increased organic carbon concentration in summer but decreased level in winter. The changes vary by season as a result of differences in volatility (as dictated by temperature and boundary layer height) and reaction rate between winter and summer. Therefore, the missing S/IVOCs and related SOA chemistry in v5.0.2 are key reasons for the OC overprediction and underprediction during cooler and warmer months, respectively.

3. Line 173-174: What is the purpose to only extracting the first 24-hour results from each 72-hour forecast? If the first 24-hour results are only needed, why still simulate the next 48 hours?

Response: We mainly focus on the first 24-h forecast for the following 3 reasons:

(1) The experimental GFSv15-CMAQv5.0.2 system is a prototype and still being developed. It is not qualified for operational application yet. Thus, the forecast results for day-2 (forecast for 25-48h) and day-3 (49-72h) were occasionally unavailable due to running issues, system fails, or archive missing, especially during the early stage in the application of this forecast system.

(2) We did the discrete statistics for day-2 performance to compare with the day-1 performance. Since there were a couple days lacking day-2 and day-3 forecast results in Jan, Feb, and Jun, the statistics are not presented for those months in case of keeping the apple-to-apple comparison. We found the day-2 performance are close to the day-1 performance. The difference in MDA8 O_3 predictions are shown in Table R1. The difference in monthly NMBs are up to

3%, mostly within 1%. Similar day-2 $PM_{2.5}$ predictions to day-1 $PM_{2.5}$ prediction could be also found in Table R2.

(3) Many of the studies for previous NAQFC forecasting performance focused on the day-1 performance (e.g., Kang et al., 2010; Lee et al., 2017). The day-1 results presented in the manuscript would be more comparable with other studies.

Table R1. Performance statistics of MDA8 O_3 against AIRNow dataset

	Day1 performance MDA8 O_3, ppb							Day2 performance MDA8 O_3, ppb							
Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr	Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr
Jan	32.1	32.0	-0.1	7.2	-0.4	17.2	0.58	Jan	/		/		/		/
Feb	36.4	35.5	-0.9	7.8	-2.5	16.7	0.58	Feb	/		/		/		/
Mar	44.9	40.4	-4.5	8.7	-10.0	15.8	0.56	Mar	44.9	40.3	-4.6	8.9	-10.2	16.1	0.53
Apr	46.4	43.1	-3.3	7.7	-7.1	13.3	0.62	Apr	46.4	42.9	-3.5	8.1	-7.5	13.8	0.59
May	44.1	42.7	-1.4	7.8	-3.3	13.9	0.67	May	44.1	42.2	-1.9	8.3	-4.4	14.8	0.62
Jun	45.7	43.9	-1.8	10.9	-4.0	18.3	0.59	Jun	/		/		/		/
Jul	44.3	46.6	2.3	9.5	5.2	16.6	0.72	Jul	44.3	46.2	1.9	9.8	4.4	17.1	0.69
Aug	43.7	46.9	3.2	9.4	7.3	16.4	0.74	Aug	43.7	46.6	2.9	9.7	6.7	16.8	0.71
Sept	42.5	45.6	3.1	8.0	7.2	14.4	0.79	Sept	42.5	45.1	2.6	8.1	6.1	14.6	0.77
Oct	37.0	40.4	3.4	7.8	9.3	15.8	0.80	Oct	36.8	40.1	3.3	7.9	9.1	16.0	0.77
Nov	34.2	35.9	1.8	7.6	5.2	16.5	0.72	Nov	34.1	35.1	1.0	7.7	3.0	16.4	0.69
Dec	31.7	33.5	1.8	7.8	5.6	18.6	0.68	Dec	29.8	30.6	0.8	8.0	2.8	20.3	0.60

Table R2. Performance statistics of 24-h avg $PM_{2.5}$ against AIRNow dataset

	Day1 performance 24-h avg $PM_{2.5}$, μg m^{-3}							Day2 performance 24-h avg $PM_{2.5}$, μg m^{-3}							
Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr	Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr
Jan	8.2	13.8	5.5	11.5	66.9	92.3	0.35	Jan	/		/		/		/
Feb	7.9	12.5	4.6	10.0	58.0	81.5	0.53	Feb	/		/		/		/
Mar	7.8	11.0	3.2	9.2	41.2	69.0	0.40	Mar	7.8	11.0	3.2	10.4	41.2	71.1	0.36
Apr	6.3	8.0	1.7	6.3	27.9	61.6	0.33	Apr	6.3	7.5	1.3	5.5	20.1	58.6	0.33
May	6.7	6.9	0.2	4.7	3.3	49.3	0.26	May	6.7	6.5	-0.2	4.6	-2.7	49.0	0.27
Jun	7.1	6.8	-0.3	5.4	-4.2	47.1	0.22	Jun	/		/		/		/
Jul	8.4	8.5	0.1	11.8	1.0	59.8	0.28	Jul	8.4	8.0	-0.4	10.5	-4.7	56.1	0.27
Aug	7.2	6.9	-0.3	4.0	-4.7	40.2	0.33	Aug	7.2	6.8	-0.4	4.1	-5.4	41.0	0.34
Sept	7.0	7.6	0.6	4.7	8.5	44.2	0.48	Sept	7.0	7.0	0.0	4.3	-0.1	43.2	0.51
Oct	6.6	9.6	3.0	9.0	44.7	73.2	0.36	Oct	6.6	8.9	2.2	7.5	33.4	67.4	0.36
Nov	8.9	13.2	4.2	9.8	47.2	72.1	0.48	Nov	8.9	12.8	3.9	9.7	43.3	70.7	0.47
Dec	8.8	13.9	5.1	10.8	57.9	82.5	0.51	Dec	8.8	13.6	4.8	10.9	54.5	82.1	0.49

4. Line 192-193: What are the criteria or references for setting this threshold (120 ppb

and 100 µg m^{-3} for O_3 and $PM_{2.5}$)? How about those abnormal low data?

Response: There are many abnormal records in the raw AIRNow data. We calculate the record numbers above certain thresholds for O_3 and $PM_{2.5}$. For O_3, records above 120, 160, 200, and 300 ppb are 0.31%, 0.17%, 0.08%, and 0.06% of the total records. For $PM_{2.5}$, records above 100, 200, 300, and 500 µg m^{-3} are 0.26%, 0.24%, 0.21%, and 0.20% of the total records. we chose thresholds of 120 ppb for O_3 and 100 µg m^{-3} for $PM_{2.5}$ as they are much higher than most observed values and provide a reasonable representation of outliers, although their selection is more or less arbitrary. We did not exclude abnormally low values.

To address reviewer's comments, we utilize the Quality Assurance/Quality Control information from the AIRNow dataset to filter the invalid records in the revised manuscript. The arbitrary thresholds are no longer used. We redo the statistics and the updated results are shown in Tables R3 and R4 below. As shown, the changes in performance statistics between the two filtering methods are minor, with slightly better results by excluding those outliners and abnormal records. Our major conclusions remain. We update the figures and the relevant sections accordingly in the revised manuscript.

Table R3. Performance statistics of MDA8 O_3 against AIRNow dataset

	MDA8 O_3, ppb in GMDD submission								MDA8 O_3, ppb with updated QC						
Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr	Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr
Jan	32.4	32.0	-0.3	7.9	-1.1	18.0	0.52	Jan	32.1	32.0	-0.1	7.2	-0.4	17.2	0.58
Feb	36.7	35.7	-1.1	8.4	-2.9	17.4	0.53	Feb	36.4	35.5	-0.9	7.8	-2.5	16.7	0.58
Mar	45.1	40.4	-4.7	8.9	-10.4	16.0	0.55	Mar	44.9	40.4	-4.5	8.7	-10.0	15.8	0.56
Apr	46.6	43.1	-3.5	8.0	-7.5	13.5	0.61	Apr	46.4	43.1	-3.3	7.7	-7.1	13.3	0.62
May	44.3	42.7	-1.6	7.9	-3.7	14.0	0.66	May	44.1	42.7	-1.4	7.8	-3.3	13.9	0.67
Jun	45.9	43.9	-2.0	11.2	-4.4	18.5	0.58	Jun	45.7	43.9	-1.8	10.9	-4.0	18.3	0.59
Jul	44.5	46.6	2.1	9.7	4.7	16.7	0.70	Jul	44.3	46.6	2.3	9.5	5.2	16.6	0.72
Aug	43.9	46.9	3.0	9.5	6.8	16.3	0.73	Aug	43.7	46.9	3.2	9.4	7.3	16.4	0.74
Sept	42.7	45.6	2.9	8.1	6.8	14.5	0.78	Sept	42.5	45.6	3.1	8.0	7.2	14.4	0.79
Oct	37.2	40.2	3.1	8.0	8.3	15.8	0.77	Oct	37.0	40.4	3.4	7.8	9.3	15.8	0.80
Nov	34.3	34.8	0.5	8.4	1.6	16.9	0.64	Nov	34.2	35.9	1.8	7.6	5.2	16.5	0.72
Dec	30.7	31.2	0.5	9.0	1.6	20.5	0.49	Dec	31.7	33.5	1.8	7.8	5.6	18.6	0.68
O_3-season	44.3	45.1	0.9	9.4	2.0	16.0	0.67	O_3-season	44.1	45.1	1.0	9.2	2.5	16.0	0.69
Non O_3-season	38.2	37.4	-0.9	8.4	-2.3	16.4	0.68	Non O_3-season	37.7	37.5	-0.2	7.8	-0.4	16.0	0.72
Annual	41.1	41.0	-0.1	8.9	-0.1	16.2	0.70	Annual	40.5	40.9	0.4	8.5	1.0	16.0	0.73

Table R4. Performance statistics of 24-h avg $PM_{2.5}$ against AIRNow dataset

24-h avg $PM_{2.5}$, µg m^{-3} in GMDD submission								24-h avg $PM_{2.5}$, µg m^{-3} with updated QC							
Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr	Period	Mean Obs.	Mean Sim.	MB	RMSE	NMB,%	NME,%	Corr
Jan	8.3	13.8	5.5	11.4	66.4	92.4	0.34	Jan	8.2	13.8	5.5	11.5	66.9	92.3	0.35
Feb	8.0	12.5	4.5	10.0	55.9	81.0	0.51	Feb	7.9	12.5	4.6	10.0	58.0	81.5	0.53
Mar	7.9	11.0	3.1	9.4	39.6	68.9	0.38	Mar	7.8	11.0	3.2	9.2	41.2	69.0	0.40
Apr	6.3	8.0	1.7	6.6	26.5	62.0	0.30	Apr	6.3	8.0	1.7	6.3	27.9	61.6	0.33
May	6.8	6.9	0.2	5.0	2.3	49.8	0.23	May	6.7	6.9	0.2	4.7	3.3	49.3	0.26
Jun	7.2	6.8	-0.4	5.6	-5.1	47.4	0.20	Jun	7.1	6.8	-0.3	5.4	-4.2	47.1	0.22
Jul	8.3	8.5	0.1	11.7	1.7	59.9	0.30	Jul	8.4	8.5	0.1	11.8	1.0	59.8	0.28
Aug	7.3	6.9	-0.4	4.1	-5.2	40.4	0.33	Aug	7.2	6.9	-0.3	4.0	-4.7	40.2	0.33
Sept	7.0	7.6	0.5	4.7	7.6	44.4	0.48	Sept	7.0	7.6	0.6	4.7	8.5	44.2	0.48
Oct	6.7	9.5	2.8	8.6	41.7	71.9	0.35	Oct	6.6	9.6	3.0	9.0	44.7	73.2	0.36
Nov	9.0	13.2	4.2	9.8	46.7	72.0	0.48	Nov	8.9	13.2	4.2	9.8	47.2	72.1	0.48
Dec	8.8	13.8	5.0	11.0	56.6	82.9	0.49	Dec	8.8	13.9	5.1	10.8	57.9	82.5	0.51
DJF	8.4	13.4	5.0	10.8	59.7	85.6	0.45	DJF	8.3	13.4	5.1	10.8	61.0	85.5	0.46
MAM	7.0	8.6	1.6	7.2	23.5	60.6	0.33	MAM	6.9	8.6	1.7	7.0	24.8	60.4	0.36
JJA	7.6	7.4	-0.2	7.9	-2.6	49.7	0.26	JJA	7.6	7.4	-0.2	7.8	-2.5	49.5	0.27
SON	7.5	10.0	2.5	8.0	33.0	63.4	0.45	SON	7.5	10.1	2.6	8.1	34.4	63.8	0.46
Annual	7.6	9.8	2.2	8.6	29.0	65.3	0.40	Annual	7.6	9.9	2.3	8.5	30.0	65.2	0.41

5. Line 259-263: What is this “artificial temporal allocation algorithm”? Please introduce more details about this algorithm.

Response: The “artificial temporal allocation algorithm” means the calculation in preprocessing from accumulated precipitation, which is recorded originally in GFS v15 outputs, to hourly precipitation, which will be used by CMAQ model. To address reviewer’s comments, we provide a more detailed description in the revised manuscript as follows:

The precipitation from the original FV3 outputs are recorded as 6-h accumulated precipitations. Artificial errors were introduced to the forecast by an issue in precipitation preprocessing during the early stage development of the GFSv15-CMAQv5.0.2 system. The precipitation at first hour of the 6-h cycle would be dropped occasionally. We corrected this issue and the hourly precipitation still shows large underprediction against surface monitoring networks. It indicates the difficulty for the forecast system in capturing the temporal precipitation, especially during summer (Figure S4). During the summer season, the discrepancy in capturing the short-term heavy rainfall worsens the model performance in predicting hourly precipitation. Besides, we use the threshold of 0.1 mm hr^{-1} to filter the valid records. If the model predicts precipitation that did not occur, the record will be excluded into the statistics calculation. However, all the predicted precipitation is counted in the spatial evaluation against the ensemble datasets of GPCP and CCPA. Therefore, the spatial performance of monthly accumulated precipitation shows better agreement than its of hourly statistics.

In general, the relatively poor performance of the forecast system in capturing the precipitation at the same hours with the observation is the major cause for the large underprediction in hourly statistics.

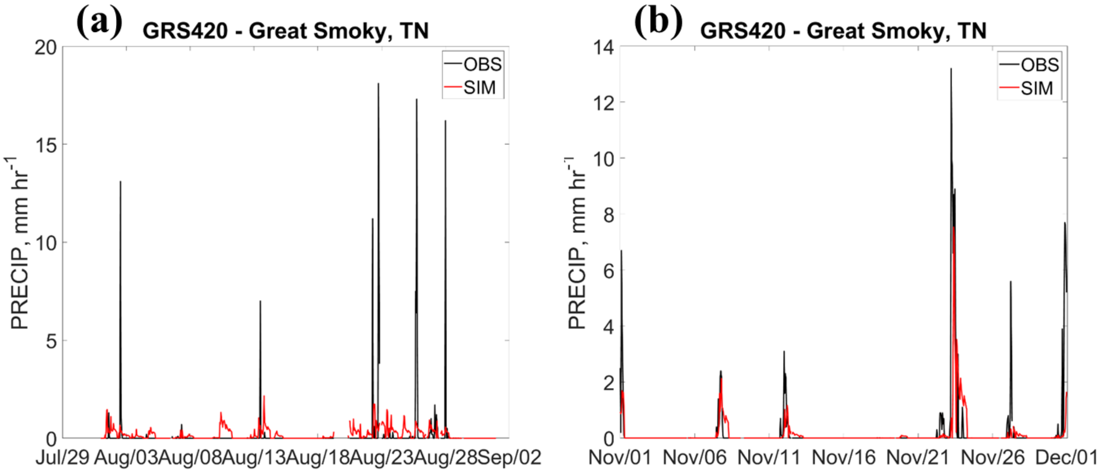

Figure S4. Time series of observed and predicted hourly precipitation at two CASTNET sites at site GRS420 in Tennessee during (a) Aug, and (b) Nov.

6. Line 335: What is "Higher predicted $PM_{2.5}$, typically SOA, in California is expected in the future using GFS-FV3-CMAQv5.3." means? Does it mean that GFS-FV3-CMAQv5.3 would predict higher concentrations than GFSv15-CMAQv5.0.2 for $PM_{2.5}$? If so, what leads to these higher concentrations in GFS-FV3-CMAQv5.3? An updated mechanism or some updated PM sources? Which one is more important for the $PM_{2.5}$ prediction?

Response: As discussed in the comment #2, the GFS-CMAQv5.3 system is expected to give higher predicted SOA in California during summer compared to the current GFSv15-CMAQv5.0.2 system. The primary PM emissions generally decrease from previous NEI to the more recent NEI. Some previews and intercomparison could be seen at http://views.cira.colostate.edu/wiki/wiki/10202/inventory-collaborative-2016v1-emissions-modeling-platform. However, the updated chemical mechanism also includes enhanced SOA formation from anthropogenic and biogenic sources, and is one of the key factors in improving the underestimation of organic aerosol in CA during summer 2010. Therefore, the updated mechanism would be more important for the $PM_{2.5}$ underprediction in those areas.

7. Line 347-359: It's better to move the method introduction to the section 2.

Response: The method for categorical evaluation has been moved to section 2 now.

8. Line 383-385: As mentioned above, GFS-FV3-CMAQv5.3 will have higher $PM_{2.5}$ concentrations. Since the significant overprediction of $PM_{2.5}$ leads the poor performance in capturing the category of "Unhealthy for Sensitive Groups" in cooler months mentioned here, whether the updated system GFS-FV3-CMAQv5.3 would have worse prediction? Can authors provide any suggestion to avoid this?

Response: The combined effect of semivolatile POA and pcSOA tends to decrease organic aerosol in winter. In addition to the semivolatile POA and pcSOA mentioned above, monoterpene SOA was also updated in CMAQv5.3. The impact of updated monoterpene SOA chemistry is more significant during summer because the BVOC emissions are much more reactive in summer than other months in southeastern US (Pye et al., 2018, 2019). Therefore, the POA and SOA updates in v5.3 are likely to lead to improvements at all times of year. The revised discussion is added in the main text. Please refer to the text in red in the response #2 in the reviewer #2 comments.

9. Section 3.3: What is the difference of the meteorological prediction among regions? Please introduce it. It would be helpful to explain the pollutant prediction bias in different region.

Response: The regional performance of meteorological prediction and its relationship with the chemical prediction are added in the revised text:

We further quantify the meteorology-chemistry relationships by conducting the region-specific evaluation of the meteorological variables. The regional performance for the major variables is shown in Figure S9. The regional biases in T2 predictions show high consistency with the regional biases in MDA8 O_3. It indicates that the cold biases in the Midwest (including region 5) and the warm biases near the Gulf coast (including regions of 4 and 6) are important factors for the O_3 underprediction and overprediction in those regions, respectively. The ozone temperature relationship was found (S. Sillman and Samson, 1995; Sillman, 1999). O_3 is expected to increase with increasing temperature within specific range of temperature (Bloomer et al., 2009; Shen et al., 2016). The surface MDA8 O_3-temperature relationship was found at approximately 3-6 ppb K^{-1} in the eastern US (Rasmussen et al., 2012). According to such relationships, the biases in T2 predictions could explain large portion of the O_3 biases. Heavy convective precipitation and tropical cyclones occur more often in the southeastern US, where are mainly regions of 4 and 6. Therefore, the performance in precipitation predictions are lower in those two regions comparing to other regions as we have shown the model has relatively poor performance in capturing short-term heavy rains during summer seasons in section 3.1. Meanwhile, the performance in wind predictions in regions 4 and 6 is relatively poor. Such performance in the meteorological predictions is consistent with the mixed performance in $PM_{2.5}$ prediction in regions 4 and 6. The discrepancy in meteorological inputs, mainly in precipitations and wind, can be attributes to the low temporal agreement shown as correlations of predicted $PM_{2.5}$ in those two regions.

Technical corrections

1. Line 1: “GFSv15-FV3-CMAQv5.0.2” should be “GFSv15-CMAQv5.0.2” to be consistent with the expression in other part of the manuscript.

Response: The FV3 dynamical core was firstly implemented in the operational GFS starting at v15. To include the complete information of the model versions in the manuscript title per the requirement of submission on Geoscientific Model Development, we incorporate the abbreviation of “GFSv15-FV3-CMAQv5.0.2” for the air quality forecasting system: GFS v15 with FV3 dynamical core offline coupling with CMAQ v5.0.2.

2. Line 214:215: the term “ozone season” should be rewrite as “O_3-season” to be consistent with the expression in other part of the manuscript.

Response: The sentence is reworded as “O_3-season”.

3. Line 419: the term “overpredicted” should be “underpredicted”.

Response: The sentence is corrected.

4. Line 539: “nemsio” should be “NEMSIO” to be consistent with the expression in other part of the manuscript.

Response: The “nemsio” is reworded as “NEMSIO”.

5. Figure 2, Figure 8b and 8d: Some labels and lines are overlap. Please modify these pictures and make it clearer.

Response: The labels in these figures are adjusted to be shown more clearly.

6. Figure 8: The serial number of the figure ((a), (b), (c), (d)) should be in front of the title.

Response: The serial numbers are adjusted to be in front of the title.

7. Figure 8a, 8c: The term “CONUS” should be “Overall”.

Response: The labels for the term “CONUS” in these two figures are reworded as “Overall”.

8. Figure 9b, 9d: Part of the labels of the figure is missing. Please modify it.

Response: Figure 9 is adjusted to show the labels completely.

Reference

Kang, D., Mathur, R., & Trivikrama Rao, S. (2010). Real-time bias-adjusted O3 and PM2.5 air quality index forecasts and their performance evaluations over the continental United States. *Atmospheric Environment*, *44*(18), 2203–2212. https://doi.org/10.1016/j.atmosenv.2010.03.017

Lee, P., McQueen, J., Stajner, I., Huang, J., Pan, L., Tong, D., et al. (2017). NAQFC Developmental Forecast Guidance for Fine Particulate Matter (PM 2.5) . *Weather and Forecasting*, *32*(1), 343–360. https://doi.org/10.1175/waf-d-15-0163.1

Murphy, B. N., Woody, M. C., Jimenez, J. L., Carlton, A. M. G., Hayes, P. L., Liu, S., et al. (2017). Semivolatile POA and parameterized total combustion SOA in CMAQv5.2: Impacts on source strength and partitioning. *Atmospheric Chemistry and Physics*, *17*(18), 11107–11133. https://doi.org/10.5194/acp-17-11107-2017

Pye, H. O. T., Zuend, A., Fry, J. L., Isaacman-VanWertz, G., Capps, S. L., Appel, K. W., et al. (2018). Coupling of organic and inorganic aerosol systems and the effect on gas&ndash;particle partitioning in the southeastern US. *Atmospheric Chemistry and Physics*, *18*(1), 357–370. https://doi.org/10.5194/acp-18-357-2018

Pye, H. O. T., D’Ambro, E. L., Lee, B. H., Schobesberger, S., Takeuchi, M., Zhao, Y., et al. (2019). Anthropogenic enhancements to production of highly oxygenated molecules from autoxidation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(14), 6641–6646. https://doi.org/10.1073/pnas.1810774116

Rasmussen, D. J., Fiore, A. M., Naik, V., Horowitz, L. W., McGinnis, S. J., & Schultz, M. G. (2012). Surface ozone-temperature relationships in the eastern US: A monthly climatology for evaluating chemistry-climate models. *Atmospheric Environment*, *47*, 142–153. https://doi.org/10.1016/j.atmosenv.2011.11.021

Huang, J., McQueen, J., Yang, B., Shafran, P., Pan, L., Huang, H., Bhattacharjee, P., Tang, Y., Campbell, P., Tong, D., Lee, P., Stajner, I., Kain, J., Tirado-Delgado, J. and Koch, D.: Impact of global scale FV3 versus regional scale NAM meteorological driver model predictions on regional air quality forecasting. The 100th AGU Fall Meeting, San Francisco, CA, 9-13 December 2019, 2019.