Review of the manuscript “Overview and evaluation of the Community Multiscale Air Quality (CMAQ) model version 5.1

This article describes the changes that have been brought to the CMAQ model from version 5.0.2 (released in April 2014 according to the CMAQ website) and version 5.1, released in Dec. 2015, including comparison of model performance between these two model versions, mostly for ozone and PM2.5 over the continental United States. Five different simulations have been performed for the year 2011 (or two months in this year) to evaluate the changes in model performance due to the global and simultaneous upgrade of WRF and CMAQ version as well as of the emission database, but also to separate the contribution of different changes in the models. The authors show in a convincing way that the improvement between model v. 5.0.2 and v5.1 is substantial, even though it is unclear whether this improvement could be due in part or totally to the change of emission datasets. The sensitivity of versions 5.0.2 and version 5.1 of this model to emission reduction scenarios is also evaluated in terms of RRF (relative response factor) for ozone, and of emission cut simulation / base simulation (for PM25).

This article is definitely within the scope of GMD. The improvements in CMAQ that are presented seem substantial even though for some of them the detailed explanation of what has actually been done and why lacks detail. A considerable work of validation has been performed.

Many aspects of the manuscript need to be improved before final publication can be considered. These aspects include the traceability / reproducibility of results (exact description of model), and providing a real and complete model overview. Also, the possibility that the described improvements in model performance are due totally or partly to the use of a new emission dataset must be ruled out.

Less importantly, a clarification of the links between WRF and CMAQ is needed (all these points are developed below in points GC1 to GC6 of my review which I think should definitely be addressed).

This article has the potential to be read and cited by many researchers or operational modellers that use CMAQ for their studies, operational previsions, and/or contribute to its development. It also reflects a huge amount of work by the CMAQ development team. This is why, **on the one hand, I recommend that this study shall be published in GMD, but I think that some important aspects of the paper definitely need to be improved before publication in GMD, including the detailed description of the model parameterizations that are changed (traceability) and a complete overview of CMAQ v5.1.**

The article is clearly written and the language level is good as far as I can say.

General comments

GC1 Sensitivity to emissions

It is appreciable that sensitivity simulations are performed to evaluate the impact of the various changes between the v5.0.2 setup with WRF 3.4 and NEIv1 emissions and the v5.1 setup with WRF 3.7 and NEIv2 emissions.

However, I have the feeling that, if the idea is to test the sensitivity of the results to the various improvements performed, then a simulation with the v5.1 setup but with NEIv1 emission dataset should be provided. In the present version, all simulations with v5.1 are performed with NEIv2 emissions and all simulations with v5.0.2 are performed with NEIv1 emissions. So the improvement between both model versions, which is shown in a convincing way, could after all be due in part or totally to the improved emission dataset. A sensitivity experiment to emissions is in my opinion needed to rule this hypothesis out and show in a direct way that the improved results are really due to the improvements in WRF and CMAQ and not to better emission input datasets.

I feel that a convincing answer to this caveat needs to be brought before publication. Otherwise, the reader has no proof at all that the described improvements are not just an effect of changing the input emission dataset

Response: A new emissions platform became available after the v5.0.2 simulations were complete. It was felt that in order to obtain the best model results the latest emissions platform should be used and therefore was used for the v5.1 simulations. Sensitivity tests were performed to assess the impact that the changes in the emissions platform had on the model results and the impacts were determined to be small. Obviously it was not made clear in the manuscript that the overall impact from the emission platform change was small. Hopefully this is now made clear in the text. In addition, a figure showing the impact of the emissions platform change on ozone and PM2.5 in January and July has been added to the text to quantify to the reader the impact from the emissions platform change.

GC2: reproducibility/traceability, precise description of parameterizations
In some occasions, statements are done that dome parameterizations have been "improved", or "changed", but failing to described exactly what has been changed in which way, questioning the reproducibility of the results (see below comments C7, C10, C14). Details, references and, if necessary equations need to be brought so that developers of other models are able to test similar changes in their models.
Point 6 in the GMD review criteria states: *"In the case of model description papers, it should in theory be possible for an independent scientist to construct a model that, while not necessarily numerically identical, will produce scientifically equivalent results. Model development papers should be similarly reproducible. For MIP and benchmarking papers, it should be possible for the protocol to be precisely reproduced for an independent model. Descriptions of numerical advances should be precisely reproducible".*
-> *The present manuscript clearly fails to meet this criteria. This is not a reason for rejection because the authors will easily be able to correct this caveat in the review process, but this should definitely be done before final publication.*

Response: The details regarding the parameterizations and options employed in both the WRF and CMAQ simulations has been expanded in Section 3. In addition, the namelists for the WRF simulations have been included in the supplemental material. As for the CMAQ model, the model code for all the versions of the model presented here is available for download through the CMAS Center website. The code is open source and freely available. The input data, including the emission and MCIP (WRF) data are all available upon request from the corresponding author. Even the output from any/all the model simulations performed here are available upon request as well.

GC3: Need for a real model overview
The title of the paper is "Overview and evaluation of the Community multiscale Air Quality (CMAQ) model version 5.1" but in my opinion the paper clearly lacks an overview of the model. I think that a Section "Overview of CMAQ 5.1" or similar should be introduced between the Introduction and the

section about the scientific improvements.

This section should include at least the basic information one would expect to find about a chemistry-transport model: since when has this model been developed? What kind of grid does it use? With what type of transport scheme (horizontal and vertical transport), is the transport Eulerian, Lagrangian, mixed? What is the recommended range of use in terms of resolution, domain size, regions of use, vertical extension? What are the inputs that are needed and the variables that are provided as an output? Are the aerosols treated in a sectional or modal way, and which physico-chemical processes are included regarding the aerosols? For example, we read that gravitational sedimentation is now included but we do not know if and how processes such as evaporation, coagulation, dry and wet deposition etc. are treated.

Maybe the authors consider that the focus of this article shall be uniquely the increment from version 5.0.2 to version 5.1 instead of a real model overview, but in this case the title would have to be changed accordingly (and the interest of the paper would be greatly lessened in my opinion, questioning the interest of the publication). In the present state, the paper does not reflect the title, because it does not include an overview of v5.1, only a set of sensitivity studies between v5.0.2 and v5.1.

Response: It's true that this manuscript is intended to provide a description of the changes that were made between CMAQ versions 5.0.2 and 5.1, along with a comparative evaluation of the results from the previous version of the model and the latest version. It is not intended to be a true "overview" of the modeling system. There have been several papers published that present a general overview of the CMAQ model from a physics and functionality perspective (Byun and Schere, 2006 for example, which is referenced in Section 1). The usefulness of presenting results from one model version to another can in some cases be questionable, however in this case it was felt that the scope of the updates made to the modeling system warranted informing the user community of the model updates and providing evaluation results that users of the model can use to determine whether or not that want to use the new model for their applications. As for the title of the manuscript, it was changed to better reflect that this article is a description of the changes in and evaluation of the CMAQv5.1 model, and not an overview of the CMAQ modeling system.

GC4 Need for a general presentation of model outputs
I feel that the reader lacks a spatialized vision of the model outputs and their characteristics compared to the known features of atmospheric composition over north America. It would be very helpful to provide a map of simulated ozone for the months of January and July, as well as simulated NOx, PM2.5 for these months with v5.1, possibly superposed with the measured average where station data is available, or any other way to give a spatialized vision of model outputs in comparison with state-of- the-art knowledge of the atmospheric composition over the continental US.

Response: This is a good suggestion. At the risk of inundating the manuscript with too many figures, we opted to provide seasonal plots of PM2.5 and O3 from CMAQv5.1 in the supplemental material. This should suffice to give the reader the reference that the reviewer is suggesting.

GC5 Better description of modelling setup
The modelling setup (Section 3) should be described more carefully. Very little space is dedicated to describing the CMAQ configuration for the main simulation, and this section mostly describes the changes between NEI emissions v1 and v2. I think that some information is clearly missing: what kind of initial and boundary conditions are used for the main species, what advection schemes, and the main user options that have been chosen for the simulations. This is particularly the case as CMAQ is a very modular model in which many choices are left to the user, as stated on other CMAQ-related documents.

The same applies to the WRF configuration, particularly as WRF seems very intricate with CMAQ (it seems that some of the updates need to be performed simultaneously in WRF and CMAQ). WRF also has many configuration options, and some of them are very critical for chemistry-transport modelling, such as for example the PBL scheme, but also the convection parameterization (if used), the eventual damping options to avoid instability over steep terrain (which is important since the continental US include major mountain rages), etc. Some information is given later in the text but should be given in Section 3 as well. It should also be mentioned what initial and boundary conditions have been used for WRF, and if some nudging has been applied.

It should also be mentioned explicitly if a spinup period has been performed (and discarded) for each simulation. This is particularly the case for the July and January simulations, which last only one month.

Response: The model setup for both CMAQ and WRF has been expanded to include many of the options employed in both models. The WRF namelists are now provided in the supplemental material as well. Information regarding the spin-up periods used has also been added to Section 3.

GC6: Precisions about the WRF model and its links with CMAQ

In many parts of the CMAQ and WRF seem very intricate (as soon as the abstract, which states "Version 5.1 of the CMAQ model was released to the public which incorporates a large number of science updates (...) These updates include improvements in the meteorological calculations **in both CMAQ and WRF**". Also on p. 13, l. 8-12 (as well as in the conclusion), we find a sentence that tends to indicate that WRF-CMAQ would even have to be considered as a single "WRF-CMAQ modeling system", that "therefore should be evaluated together". This raises some questions:

* Can CMAQ be used with other meteorological models than WRF (such as reanalysis or outputs from national meteorological centers)? Is it recommended by the CMAQ developers?
* If these models are so intricate together, would it not be relevant to change the title to include this concept of "WRF-CMAQ modeling system"?

Response: This is good point to be made. While CMAQ can be used with other meteorological models (MM5 for example), it does have strong ties to the WRF model through consistent mixing schemes and land-surface treatments. So, while it's not truly correct to call it exclusively the WRF-CMAQ modeling system, what was tested and evaluated here was the WRF-CMAQ modeling system.

Minor general comments

- Some parts of the article are not friendly for a non-US reader. For example, the 2-letter codes for US states are not well-known to the international public. Either the authors should provide a map of these codes, also including the five zones defined p. 14, l. 6-8, or give the complete names of the states.

Response: All state names are now spelled out.

- Many long URLs are given between parenthesis in the text (e.g. p. 6, l. 27-28). They should probably be given as footnotes.

Specific comments:

Section2

C1 p. 5, l. 31: Is this time interval valid for all the domain? Days in July should last much longer than

12 hours at least in the north of the domain, and the daytime interval must be very different from the west to the east of the simulation domain (about 5000 km, which is about 4 hours time lag in the solar time). Using points from 11:45 to 23:45 UTC from west to east would result to using data points from mid-morning to the sunset at the eastern part of the domain, and from dawn to mid-afternoon in the west of the domain, which is critical as cloudiness often has a strong diurnal cycle.
I recommend that all the available daytime data points shall be used for this comparison.

Response: **All available data from the satellite product are used in the average in Figure 1a (Note this Figure has been moved and is now referred to as FigureXa). The figure in the Supplemental Information section S1 shows the number of daytime hours (11:45UTC – 23:45UTC) with available GOES cloud albedo data during July 1 – July 31, 2011 for the modeling domain. Regions in the eastern half of the US have a larger number of available satellite observations (on the order of 390hrs) compared to the western coast which has < 340hrs. Since the reference to the time window of 11:45UTC-23:45UTC caused unnecessary confusion we have removed this from the main text. We now point readers to the Supplemental Information for further description of the hours of available satellite data:**
"The satellite data are available at 15 minutes prior to the top of the hour during daytime hours and were matched to model output at the top of the hour (see section S.1 in the supplemental material for further information)."

C2 p. 5, l. 32: the description of Fig. 1 does not fit that in the caption of Fig. 1 (the latter one seems to be more relevant). The average cloud albedo seems not to be shown.
This should be clarified.

Response: The reviewer is correct. The wrong Figure 1 was included with the original submission of the manuscript. In the revised version of the manuscript this Figure (now called Figure 4) does show the average cloud albedo, consistent with the description in the text. The Figure caption has also been changed to say:
"Figure 4. The average cloud albedo during daytime hours in July 2011 derived from (a) the GOES satellite product (b) WRF3.7 (c) CMAQv5.1 with photolysis/cloud model treatment from v5.0.2 and WRF3.7 inputs (CMAQv5.1_RetroPhot) (d) CMAQv5.1 using WRF3.7 inputs (CMAQv5.1_Base)."

C3 p. 5, l. 36-371: The authors should explain why it is needed to have a convective cloud model within CMAQ and not use cloud fractions and water content provided by WRF, particularly if CMAQ and WRF almost form a single modeling system as stated elsewhere.

Response: We revised the preceding paragraph to explain why CMAQ cannot not use convective cloud predicted by WRF. The updates in photolysis calculations CMAQ v5.2 related to clouds were intended to ensure internal consistency between cloud mixing, aqueous chemistry and photolysis. The reason cloud treatment in CMAQ is not currently "completely consistent" with WRF is the way that sub-grid convective clouds are handled. The sub-grid convective cloud scheme in CMAQ, which is responsible for convective transport of chemical species, aqueous chemistry, and wet scavenging, is a simple bulk scheme based on the convective cloud model in the Regional Acid Deposition Model (RADM; Chang et al., 1987) but with convective transport based on the Asymmetric Convective Model (Pleim and Chang, 1992). Since the CMAQ cloud scheme uses the convective precipitation rate to diagnose sub-grid mass fluxes, the location and timing of precipitating convective clouds are consistent with WRF. A new convective cloud scheme for CMAQ based on the Kain-Fritsch scheme in WRF is currently being tested to improve consistency across the chemical and meteorological components of the system. This future model update will allow sub-grid cloud fraction and water content information from WRF to be used within the sub-

grid cloud-related processes within the CMAQ system.

C4 Fig. 1: The methodology to produce these maps should be precised. Is a threshold placed on cloud albedo to decide that a particular hour is or is not cloudy? Is this threshold the same in the models and for the GOES data?

Response: The wrong Figure 1 was included with the original submission of the manuscript. In the revised version of the manuscript this Figure (now referred to as Figure 4) shows the average cloud albedo (consistent with the description in the text). There is no threshold used or needed with the new figure.

C5 p. 5, l. 37-38: the statement that the model cloudiness is "more consistent with the WRF parameterization" is possibly correct but not shown by the figure, since WRF3.4 cloudiness is not provided for comparison with CMAQ v5.0.2 cloudiness. Also, it is hardly a surprise, since CMAQ v5.1 clouds are produced from WRF 3.7 cloud data it would be alarming if the results are very different.

Actually, this paragraph seems to me a bit titled towards suggesting that CMAQv5.1 cloudiness is better that CMAQ v5.0.2, but the figure does not allow to make such a statement, since CMAQ 5.1 cloudiness is compared to the cloudiness of WRF 3.7, which is almost the same data, while CMAQ v5.0.2 is compared to the actual satellite data, which is more of a challenge. Actually, in my eyes, visual comparison between Figs 3c and 3d to Fig. 3a shows that, above most of the continental US and the surrounding oceans (except maybe the center-north of the US), CMAQ 5.0.2 cloudiness is in much better agreement than CMAQ 5.1 with the observed cloudiness.

Response: We revised the paragraph in section 2.3 to better explain how WRF and CMAQ differ in the cloud description, specifically the sub-grid or convective clouds. The motivation for the updates in the cloud treatment in CMAQv5.1 is the improved consistency with the WRF parameterizations and this Figure is intended to emphasize this improvement. (Note that the figure in question has been moved to section 4.3 and is now referred to as Figure 4). The cloud albedo in Figure 4c that represents the cloud treatment from CMAQv5.0.2 was based on inputs from WRF3.7, not WRF3.4. This is now made more explicit in the text and the Figure caption to avoid any confusion.

The authors invite to compare Fig. 3c to 3a and 3d to 3b, but I recommend that they also invite the reader to compare Fig. 3d to 3a and explicitly comment the comparison between CMAQ v 5.1 cloudiness to the Goes data and comment why the agreement does not seem as good as with CMAQ v5.0.2 over many areas.

Response: The readers are provided information on the differences between the CMAQv5.1 cloudiness and the GOES data in the following paragraph in section 4.3:
"Two notable issues remain with the v5.1 modeled cloud parametrization. The photolysis cloud parameterization in v5.1 produces more clouds over water compared to the WRF parameterization, which is itself biased high for some parts of the Atlantic Ocean compared to GOES. This issue will be addressed by science updates planned for the CMAQ system and evaluation results are expected to improve in the next CMAQ release (See Section 7.4). A more significant issue, from an air quality perspective, is the under-prediction of clouds over much of the Eastern and West Central US in the WRF predicted clouds, which is now directly passed along to CMAQ. This misclassification of modeled clear sky conditions can contribute to an over prediction of O3 in these regions. Resolving this issue will require changes to the WRF cloud parameterization. Future research will also include changing the sub-grid cloud treatment currently used in the CMAQ system to be consistent with the sub-grid parameterization used in WRF. Section S.1 in the supplemental material provides a table with additional evaluation metrics

of the modeled clouds over oceans versus over land and also describes how cloud albedo was calculated for the three model simulations."

Also, it would be interesting to have the same 4 figures shown and commented for the month of January (eventually as a supplement).

Response: We did attempt to make this figure for January but found that the satellite retrieval method for cloud albedo gave unreliable diagnostic information for locations with snow cover.

C6 p. 6, l. 11-12: "the rate constant (...) low-pressure limit": please clarify, and define what are N and Fc in the subsequent parenthesis.

Response: We opted to remove the statement containing the values for N and Fc, as they are the same as those reported by Bridier et al. (1991) which is referenced in the section. Including the updated values seemed unnecessary given the that they are available in both the cited paper and the CMAQv5.1 technical documentation, and would actually be quite laborious to define here.

C7 p. 7, l. 11-19
The modification to the sea-salt emission schemes is described in a rather vague way. I recommend that the following information is added:
- Give the equations that control the sea-salt emission processes in CMAQ in open ocean and in the surf-zone and the reference from which they were inferred. It would also be helpful to the reader to also show the size distribution of sea-salts in the former and in the new version ("the size distribution (...) was expanded to better reflect ..." seems a bit to vague to me, not permitting reproducibility.

Response: Numerous references were included in this section that contain the information requested above. In addition, the CMAQv5.1 technical documentation also contains detailed information regarding this update to the model. It seems excessive to rehash here the specific equations and values implemented in the new model when they are easily obtained through the citations provided. The combination of the citations provided and CMAQ technical documentation should be provide information needed for reproducibility. In addition, the CMAQ code itself containing the updates is available as well.

C8 subsection 2.1, p. 3, seems to address issues about vertical mixing and air-surface exchanges rather than explicit transport. I think it would fit better in the 2.5 subsection.

Response: This section has been moved to the end of Section 2.5.

C9 p. 7, l. 23: Niinements to Niinemets?

Response: The reference name has been corrected and added to the list of references.

C10 p. 7, l. 23-24: "A leaf temperature algorithm was implemented that replaced the 2m-temperature...". I think there is not enough detail to guarantee reproducibility, or the possible use of this algorithm by other modellers. I recommend that the idea of this algorithm and its equations are given, and if possible that the interested reader shall be oriented to a publication describing this algorithm.

Response: A detailed description of the updated algorithm is available in the CMAQv5.1 technical documentation, which is now provided as a link in the text and the reader is now encouraged to reference for additional details.

C11 p. 7, l. 24-25: it is unclear to me how 2m-temperature can be consistent with emission factor measurements (2m temperature should be consistent with 2m-temperature measurements...). Please reformulate or clarify this sentence.

Response: This statement has been expanded to make it more clear what exactly was changed to make the calculation in the CMAQv5.1 more consistent regarding the 2-meter temperature and emission factor in BEIS.

C11 p. 7, 25: please define BELD

Response: BELD has already defined in the text as "Biogenic Emission Land-use Data".

Section 3

C12, p. 8, l. 5: it is interesting that the model is run up to 50 hPa, well into the stratosphere. I think it would be interesting to have a glimpse of the model outputs in the stratosphere (or throughout the whole atmospheric column) and their validity, particularly regarding ozone. For upper troposphere/lower stratosphere, comparison with either satellite sata or aircraft data such as MOZAIC (http://www.iagos.fr/web/) would, I think bring something useful to this study. Also, it would be interesting to know whether additional reactions are needed in CMAQ to take into account the lower stratospheric chemistry, which is different from tropospheric chemistry.

Response: It's really beyond of the scope of what this paper is trying to accomplish to get into the details of the treatment of the stratosphere in CMAQ. There are papers however, both published and in development, that discuss stratospheric treatment in CMAQ in terms of hemispheric CMAQ model simulations. Those papers do take advantage of the some of the measurement data referred to by the reviewer. In short however, there are checks within the CMAQ model to deal with stratospheric ozone, utilizing climatological ozone profiles and comparing those against model simulated upper-level ozone values for consistency. If the values are found to be inconsistent, then the user is warned and in some cases the model simulation is stopped.

C13, Table 2. This table is useful but not easy to read. I suggest that it is converted into a table with several columns:

Simulation name	CMAQ version	WRF version	NEI version	Photolysis scheme	Chemical scheme	Simulation period
CMAQ_5.0.2_Ba se						
....						

Response: The layout of Table 2 has been updated accordingly.

C14, p. 9, l. 4: There are many ways to simulate plume-rise. The scheme that is being used and the underlying data (chimney height, flow speed and temperature if relevant etc.) should be described, or the reader should be referred to a previous publication describing the plume-rise strategy in CMAQ, to permit reproducibility.

Response: The CMAQ plume rise has been described in previous publications (e.g. Foley et al.) and follows the same implementation as SMOKE. A link was added that provides details on the

plume rise calculation used in CMAQ.

Section 4

C15, p. 9, l. 24-30: I think the effect of the changes in the treatment of clouds should also be considered at that point. In July, the increase in O3 between both versions of WRF is strong in the SW United States (from Louisiana to Virginia). If one goes back to Fig. 1, we can see that this area was represented as cloudy by the CMAQv5.0.2 version with WRF 3.4, but almost cloud-free by WRF 3.7. It appears to me that this reduction of cloudiness between WRF 3.4 and WRF 3.7 is also a very plausible explication for O3 increase in summertime over this area, particularly as the same is observed in western Mexico and the states of Arizona, Colorado, Utah and New-Mexico.

Response: Agreed that the effect of the change in clouds between the two versions of the model is an important driver in the difference in ozone between the two simulations. The effect of the change in clouds is specifically addressed in section 4.3 that compares the different cloud treatments used in CMAQv5.0.2 and CMAQv5.1. Inherent in this discussion in the effect that the reduced cloudiness in v5.1 has on the ozone mixing ratios.

Section 5

C15, p.13, l. 1-5: I think the word “stochastic” is not appropriate. I very much prefer “subgrid variations”, which is also used. It is very much a modeller vision to consider that everything below grid resolution is essentially stochastic, by the fact that a station close to a highway measures higher contamination levels than a station 3 miles away in a forest is perfectly deterministic.

Response: The term stochastic has been widely used and applied to these subgrid variations, and is the terminology used in the referenced article. We’re just being consistent with the terminology used in that article.

C16, p. 13, l. 6-20: This part describes the increments between both model versions, and would probably be more at its place in Section 4 than in Section 5. section 5 is about the validation of v5.1 so it is confusing that v.5.0.2 is mentioned so much at this point.

Response: Section 4 is intended to only focus on a few of the major updates and the impact those specific updates had on the model performance. Section 5 is intended to show the overall performance of the model for ozone and PM2.5. While it would be possible to only show the results of the v5.1 simulation here, it seems useful to also include the v5.0.2 results alongside the v5.1 results so that the reader can quickly identify how they might expect ozone and PM2.5 to change between a v5.0.2 and v5.1 simulation. For that reason, we opted to keep results from both model simulations in this section. This is a fairly common way to present results of a model update such as this.

C17, p. 13, l. 15-20: This is a significant caveat that should be mentioned in the model description much earlier than that. The fact that windblown dust treatment was not available for the article simulations and neither for the public release is not anecdotic in my opinion. Also, it would be appreciated that the statement that dust contributions are “small and episodic” is made more quantitative, for example by providing a map of average dust concentrations (and variability) in v5.0.2.

Response: Granted the effect of windblown dust can be large in isolated areas, generally a short amount of time. In the United States, windblown dust concentrations are maximized in the springtime in the desert southwest. We’ve added quantitative values of the springtime seasonal average concentration of soil from CMAQv5.0.2. Overall the seasonal average concentration is

very small compared to the overall PM2.5 seasonal average concentration.

C18, Section 5.1:
This section essentially describes the differences between v5.1 and v5.0.2 . While this is pertinent for the study, it does not seem to fit within Section 5, which is about evaluation of v5.1 . Only the parts referring to Fig. 9 and to Figs. S2-S5, such as p. 14, l. 3-14, l. 26,27, etc. partly treat of the evaluation of v.5.1. In my opinion, the parts describing differences between v5.1 and v5.0.2 should be moved to Section 4.

Response: This section was intended to serve two purposes. First, it is intended to show the performance of CMAQv5.1 versus observations, which of course is useful to users. Secondly, it is also intended to show the change in performance of ozone and PM2.5 between CMAQv5.0.2 and CMAQv5.1 so that users can assess for themselves whether the change in model performance is significant for their application or not. To accomplish this, we present the evaluation results against observations for both versions of the model. Section 4 was intended to only show the impact of the individual model updates, while Section 5 shows the performance of the entire CMAQv5.0.2 and CMAQv5.1 modeling systems. Hopefully the reviewer can agree that this is an acceptable way to accomplish these goals.

C19: figs S2 to S5 should be moved into the main manuscript (because **they do present material permitting an objective evaluation towards the observations in absolute terms, which is lacking in most of the manuscript**). These figures (S2 to S5) also show a rather spectacular improvement from v5.0.2 to v5.1, except maybe for springtime. I think it would be fair that the authors insist more on this very strong improvement of their model results. There are some formatting problems in these Figs. S2-S5 : the simulation name in the caption do not fit exactly the ones given in Tab. 2 , legend of the vertical axis of the top middle panels fail to state that it is the bias which is plotted. If it takes too much space to bring all figures S2-S5 into the manuscript, the authors should maybe consider providing a table with the average observed and modelled values for PM25 as well as the RMSE and correlations for both model versions, and for the four seasons.

Response: Ideally we would like to include all the figures in the main text, as they all present relevant material. However, it would create an extremely lengthy article with all those figures from the supplemental material included. As a comprise, three new figures have been added that include the observed and modeled seasonal diurnal profiles for PM2.5, O3 and NOx. These figures do not include the MB, RMSE and correlation, and the reader is referred to the supplemental figures for that information.

C20, Section 5.2:
I would make essentially the same general recommendations than for Section 5.1: that the parts treating of increments from v5.0.2 to v5.1 be moved into Section 4, and that more focus is put on figures S7-S14, bringing some of them into the manuscript, and/or providing a table with the most relevant statistical parameters for both ozone and Nox. The title of the section should probably include Nox as well as ozone.

Response: See responses to C18 and C19.

C21, p. 16, l. 6-12: This difference in summertime ozone concentrations over the eastern US is rather significant and in my opinion can be attributed to the change in meteorology between WRF 3.4 to 3.7 (Fig. 2b): the similarity between Fig. 2b and 10c is striking and the numbers and patterns correspond quite well. I think the authors should comment that, and also the fact that the model bias for ozone in summertime over these regions is increased in v5.1, corresponding to the fact that cloud cover is underestimated in these regions in v5.1 (Fig. 1).

Response: Added to the text the strong correlation between the overall change in ozone in v5.1 and the change in ozone due to the WRF/CMAQ meteorological updates, along with the increase due to increased photolysis in v5.1.

Section 5.2 ("comparison to aircraft measurements")

C22-1: A general comment about this part is that comparing with a single vertical profile is not enough to state an improvement or a deterioration in a model's performance. The analysis of other profiles should be included, either from the same campaign, or from routine MOZAIC measurements, which are abundant above the continental US.

Response: The authors agree that titling this section as a comparison to aircraft measurements is a bit misleading even though it does include a comparison to a single day of aircraft measurements. Since the objective of this section was to evaluate the change in model performance for NOY, AN and PNs, the section has been retitled to reflect its purpose. In addition, several statistical metrics of NOY performance have also now been included in the section to help expand the analysis provided and highlight the greatly improved performance of NOY in CMAQv5.1. While it would be nice to be able to show additional profiles from other days and include measurements from other networks, the point of the section was to simply inform the reader of the large improvement in NOY performance and an example of the change in ANs and PNs mixing ratios that can be expected in the new model.

C22 Note, there is a problem in numbering, because previous section is numbered 5.2 as well. I find it very interesting to give some comparison with aircraft measurements, even though it would be great to have it at upper altitude as well, either from this measurement campaign, or from the routine MOZAIC (or equivalent) measurements.

Response: The section number has been corrected.

C23 Please provide the coordinates and altitude of "Edgewood, MD", as well as the hour (and duration, if relevant) of the considered flight, because PBL structure and the behavior of the real and modelled atmosphere depends a lot on the time of day, and if possible more meteorological context (was it a clear-sky, cloudy, rainy day at that place?). This would help a lot the reader to analyze the figures.

Response: The location and elevation of the Edgewood site have been added to the text. The profiles themselves represent an average of several vertical spirals that took place over Edgewood that day, roughly taking place in the morning, early afternoon and late afternoon. This has now been stated in the text.

C24 Fig. 13: please increase the size of fonts in the panels, it is hard to read in printed version.

Response: The font size has been increased in the figure.

C25 Do the authors have an idea why the Nox (and Noy) values in Fig. 13 are reduced so drastically between both model versions (about 50% for Nox)? Model simulations describe a rather young air mass with Nox/Noy ratio around 60%, while the Nox to Noy ratio in the measurement is about 30%, typical of a much more aged (and clean) air mass, suggesting different trajectories in the model than in reality. Nox level being very dependent on anthropogenic emissions, is it possible that this drastic reduction is due at least in part to the emission update? These changes between model versions seem more dramatic than the smooth statistical changes that appear in the statistical scores between v. 5.0.2. and v. 5.1.

Response: The NOx and NOy mixing ratios decrease due to the changes made to the atmospheric chemistry in CMAQv5.1 (Section 2.4.1), while the NOx mixing ratios would be expected to decrease due to greater photolysis in v5.1 (Section 2.3). On average, the NOy mixing ratios in CMAQv5.1 decreased 21% in July and 13% in January. So, in the plots shown in Figure 13, the NOy mixing ratio decrease of about 30% seems reasonable for an isolated case. The emission platform update had little to no impact on the NOX concentrations, so the emission change does not contribute to the decrease in NOx seen in the figure.

Section 6

C26 p. 17, l. 21 : these notions are not necessarily familiar to the reader. I think it should be precised that these notions apply to the United States of America, and possibly add a reference that explains what are the SIPs and "Federal rules".

Response: Added text to explain that SIPs and Federal rules aim to reduce emissions through regulations in order to meet mandated air quality standards.

C27 p. 18, l. 2 : Text and figure caption of Fig. 14 announce that a ratio (emission cut simulation / base simulation) will be shown, but the panels show that what is shown is a "RRF", a notion which is not defined. If RRF is to be actually used, then it should be defined, and possibly, some clues shall be given about the use of this indicator, which is not known to the entire modelling community.

Response: Since the value presented is indeed a ratio and not a true RRF calculation, the term RRF has been removed from the plots and caption, and replaced with the explanation that a ratio of concentration has been used.

C28 Figure 14: it is not clear to me what kind of samples populates the box plots. Are the samples made from model grid cells, model time series at given locations? Also, the sample size for of each bin should be precised (for example, the appearance of the rightmost box plot for the case of January suggests that the sample size may be very small. A bit more methodological precisions for this plot (as well as Fig. 15) would be welcome. I would also suggest that the format of Fig. 15 is applied to Fig. 14 as well, which avoids introducing the RRF, and permits the reader to evaluate the reduction obtained in v5.0.2, the reduction obtained in v5.1, and visualize and evaluate the difference between the responsiveness in both versions. I think Fig. 14 does not allow as to know if the difference in responsiveness between both model versions amounts to 2% or 50% of the expected model response, while Fig. 15 does.

Response: Figure 14 has been updated to remove the term RRF to the correct description as a daily ratio (as is presented in Figure 15). The bins are populated from model grid cells, which is now stated in the caption. In addition, the number of model grid cells in each bin is now included above the x-axis.

Section 7

C29 p. 18, l. 35: I do not agree that the model has been "evaluated in terms of operational performance" since the evaluation has been performed for year 2011, so more like a reanalysis than an operational forecast model. I suggest to replace by "evaluated by comparison of a simulation of year 2011 to routine measurements of ozone, Nox and PM25 from xxx ground stations" (or something equivalent)

Response: Removed the word "operational" as it is a source of confusion and requires additional context. Here we use the term operational to refer to evaluation against observations, not an

evaluation of the model in an operational (e.g. forecast) mode. However, it seemed easier to remove the word operational and let the evaluation results speak for themselves.

C30 p. 19, l. 10: "to decrease the amount of sub-grid in the photolysis calculation" : please clarify, come words seem to be missing here.

Response: Added the word "clouds" after sub-grid.

C31 p. 19, l. 10-13: it also seems that switching from WRF3.4 to WRF3.7 had a strong effect in reducing model cloudiness over the continental US (Fig. 1), in turn increasing summertime ozone levels over the concerned areas (Fig. 2b), even in the absence of update in the photolysis scheme. Therefore, I find this part of the conclusion (from "The net effect..." to "on average") a bit questionable.

Response: Hopefully with the addition of the word "clouds" from the above comment it's clear that we are referring to the decrease in cloudiness in the model as the driving factor in increasing summertime ozone. It's actually the updates to the cloud treatment in CMAQ (to make them consistent with WRF clouds) and not the transition from WRF3.4 to WRF3.7 that drive the difference in the clouds in CMAQ.

C32 p. 19, l. 13-14: if I am not wrong, these options are not really been described in the main development, neither which one of these options was chosen to obtain the results described here.

Response: The availability of these options was mentioned briefly in Section 2.3 and is mentioned again briefly here as new option available in CMAQv5.1.

C33 p. 20, l. 2: I think the authors should state explicitly which known issues they are referring to (because this may be of interest to model users)

Response: Reworded to remove the words "known issues". The known issues being referred to was the windblown dust treatment in v5.1, which is referred to in the next sentence anyway.

C34 p. 20, l. 10-11: I think the WRF website should be referred to as well since extensive use of WRF has been made and it seems critical that users use CMAQ with a recent WRF version.

Response: The WRF website was added to the Data availability section.