

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 PM_{2.5} 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 PM_{2.5}).

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 predictions, 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

GC2 : reproducibility/traceability, precise description of parameterizations

In some occasions, statements are done that some parameterizations have been “improved”, or “changed”, but failing to describe 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.

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.

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 NO_x, PM_{2.5} 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.

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 intricated 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 ranges), 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.

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

In many parts of the CMAQ and WRF seem very intricated (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 reanalyses or outputs from national meteorological centers) ? Is it recommended by the CMAQ developers ?

* If these models are so intricated together, would it not be relevant to change the title to include this concept of “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.
- 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 :

Section 2

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.

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.

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.

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

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 tilted towards suggesting that CMAQv5.1 cloudiness is better than 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.

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.

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

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

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.

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

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

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 decsribing this algorithm.

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.

C11 p. 7, 25 : please define BELD

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.

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						
....						

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.

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 explanation 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.

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.

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 v5.0.2 is mentioned so much at this point.

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 teh 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.

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 v5.1 . In my opinion, the parts describing differences between v5.1 and v5.0.2 should be moved to Section 4

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.

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.

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).

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.

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.

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 behaviour 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.

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

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 drastical 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.

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”.

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.

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.

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)

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

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 absense of update in the photolysis scheme. Therefore, I find this part of the conclusion (from “The net effect...” to “on average”) a bit questionable.

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.

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)

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.