Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-351-AC1, 2020

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "Sensitivity of *CHIMERE* to changes in model resolution and chemistry over the northwestern Iberian Peninsula" *by* Swen Brands et al.

Swen Brands et al.

swen.brands@gmail.com

Received and published: 10 January 2020

Response to Anonymous Referee #1

Referee comment: The manuscript of Brands et al. investigates the sensitivity of the CHIMERE model to different horizontal and vertical resolutions and to different chemical mechanisms. The focus is on a region of the northwestern Iberian Peninsula where the CHIMERE model has not been applied so far. The model results are evaluated against observations with a focus on minimum and maximum values. To my opinion the current version of this manuscript does not fit into the scope of ACP(D) as it is mainly a technical analysis of the model system and has no broader scope or gen-

eral implications for atmospheric science (see aims and scope of ACP; "The journal scope is focused on studies with general implications for atmospheric science rather than investigations that are primarily of local or technical interest"). To fit into the scope of ACP(D) a more general geoscientific conclusion/scope of the manuscript would be necessary. However, I think that the technical analysis of the model system is an interesting topic and therefore I propose to revise the manuscript largely (see detailed suggestions below) and hand the manuscript over to the partner journal Geoscientific Model Development (GMD). If the authors wish to publish their manuscript in ACP(D) the scope of the journal should be taken better into account and the manuscript needs major revisions.

Response: We would like to thank you very much for taking the time to review our manuscript and for your helpful and constructive criticism.

During the last months, we have undertaken an extensive review and have adressed all points raised by you not only by mere discussion but primarily by adding additional analyses and results. Most importantly, we have largely extended the number of model experiments (from 8 to 19), including 8 experiments run with the EMEP 2017 inventory, which is why it took a little bit longer to complete this review. To improve the interpretability and transferability of our results and to thrive away from the purely technical aspects, we have included additional figures, table content, text passages and conclusions that are of general interest for atmospheric science. While this should make our manuscript more relevant for the ACP readership, we would also agree on handing it over to GMD, if this is intended by the editor.

General comments

Referee comment: In p4 the authors write '[...]without using downscaling with traffic or population proxies[...]'. Accordingly, also for the finest nest with 0.05 x0.04 resolution, the authors applied emissions with a resolution of 0.1x0.1, right? If this is the case this is an important limitation of the study and needs to be clearly stated.

Response: This is a misinterpretation caused by us, arising from a too short description of the applied emission postprocessing in the first version of the manuscript. We surely did not run the 0.05° x 0.04° (lon x lat) CHIMERE domain with anthropogenic emissions on a 0.1° x 0.1° grid and have never tried to do so since the model would probably crash in this case. Rather, the raw emission from HTAP 2010 (and now also EMEP 2017) have been downscaled to the 0.05° x 0.04° grid with the help of different proxies, as described below. In the revised manuscript, an entire section is dedicated to the applied anthropogenic emission inventories and downscaling techniques (see Section 2.2):

Referee Comment: It is well known that the emission resolution influence the results largely. A too coarse resolution of the emissions can also deteriorate the model results (e.g. Markakis et al., 2015). Therefore, I propose to perform additional model runs with a downscaling of the emissions as this is a general feature of the CHIMERE model. These additional runs can then be used to quantify uncertainties due to missing downscaling of emissions.

Response: Already for the 8 experiments included in the first version of the manuscript, the raw HTAP 2010 emissions had been downscaled with land-use categories, i.e. with the basic downscaling option of the emiSURF program shipped with version 2017r4 of CHIMERE, but no further proxies had been considered. For the revised manuscript, 8 additional experiments have been performed with anthropogenic emissions from EMEP 2017, downscaled with population and traffic proxies, as well as with the location of large point sources. On top of the 16 HTAP and EMEP experiments, 3 additional experiments were run to answer specific questions. For one of these specific experiments, HTAP 2010 was downscaled with land-use categories and population density only, i.e. traffic density and large point sources were not used in this case in order to isolate the effect of the population proxy (see Table 3 in the revised manuscript for the detailed infos). The corresponding results can be found in Section 3.1.2, 3.2.2 and 3.3.

Referee comment: The authors find a poor performance for NO2 of the CHIMERE

model and link this poor performance to deficits of the emission inventory. To my opinion this argument needs a more detailed investigation. Several things should be discussed/considered:

Response: In the first version of the manuscript, the conclusions drawn on the model's performance for NO2 were too strong and have been downweighted in the revised manuscript. The large biases found for this species are due to the fact that the percentage bias is used in a region where the NO2 emissions at background sides are very low, part (see Figures 2 and 6). Consequently, a bias of only a few ug/m^3 in absolute terms translates into a large percentage value at these stations, which then largely increases the width of the boxplots (see e.g. Figure 8a). In the revised manuscript, this has been clarified in lines 376-378.

Referee comment: The emission inventory is for the year 2010. What was the emission changes in the last 8 years? What trend do ground-level measurements of NO2 show? Are there more up to date emission inventories available (for example from TNO,EDGAR etc.)? If so, why don't use them?

Response: To fix the problem of outdated emissions, the aforementioned 8 experiments based on EMEP 2017 were additionally conducted (see Section 2.3 and Table 3) and the respective results added to the revised manuscript (see Section 3). EMEP 2017 is the most up-to-date emission inventory provided by the European Union and the one year difference to our study period (summer 2018) is assumed to be negligible.

Referee comment: What is the performance of NO2 at the different stations types? How well does the model perform at the 'background' stations? How well at the "traffic" and "industry" stations?

Response: The performance of the four considered chemical species (including NO2) at background, traffic and industry sides was already shortly assessed in the first version of the manuscript (see Figure 6 therein). For the revised manuscript, this part of the study has been extended (see Figure 10 and Section 3.3 for the respective results).

Referee comment: Where are the “traffic” stations located? Does it make sense to evaluate a model with resolutions of 4 to 5 km and emissions at around 10 km resolution with measurements at the street scale? I guess it makes more sense to evaluate against the measurements at (urban) background stations or average all values of the ’traffic’ stations of one particular city.

Response: This is a good point which not only applies to traffic stations but also to industrial sides or to any other sides affected by localized emissions. However, to our understanding, the downscaling techniques applied here have precisely been developed with the aim to improve CHIMERE’s forecast skill on the local scale (Mailler et al. 2017) and, from this point of view, traffic or industrial stations should not be removed from the validation. To solve this dilemma, we do show the model’s performance at traffic (and also industry) sides, which often are outliers if compared to the results at the remaining locations, but only in case they do not hamper the interpretation of the results. Consequently, outliers are shown in the overlay maps (Section 3.1.1 and 3.2.1) but ignored in the boxplots (Section 3.1.2 and 3.2.2) since their inclusion in the latter would blow up the scale of the x-axes.

Furthermore, we use outlier-resistant statistics such as the median instead of the arithmetic mean (see vertical orange line in the centre of the boxplots). The median model performance at traffic sides is shown in panels j, k and l of Figure 10. A description of the corresponding results can be found in Section 3.3. City-specific averages are an interesting alternative and will be considered in future studies.

Referee comment: How well does WRF reproduce the observed meteorology? Too efficient mixing of the boundary layer might cause problems in reproducing the measurements. Please provide at least a basic meteorological evaluation of the used meteorological data.

Response: For now more than a decade, the WRF configuration used to drive CHIMERE has been the backbone of the real-time forecasting system used by the

Galician governmental weather service MeteoGalicia. During the course of time, its performance has been steadily supervised by a team of operational forecasters as part of their day-to-day business and any large errors have been eliminated, if possible. In the revised manuscript, WRF’s performance for a typical summertime heat day is illustrated in Supplementary Figure 1, which demonstrates that orographic and costal effects on the local weather are fairly well reproduced by the model (see also lines 95-99 of the revised manuscript).

Referee comment: The authors focus only on daily minimum and maximum values. I agree that especially the maximum values are very important with respect to air quality issues. However, to my opinion it would be very important to investigate also the general ability of the model to represent the hourly variability of the measurements. Therefore, I propose to further perform statistical analysis of the whole time series for each station and not only for minimum and maximum values.

Response: As suggested by you, a verification of the hourly time series has been included in Figure 10 of the revised manuscripts and the respective results are described in Section 3.3. We found that CHIMRE’s performance for hourly data in many aspects is similar to the performance obtained for daily maxima.

Referee comment: Further, the analysis does not take into account that the model concentrations could be shifted geographically (e.g. minima and maxima are misplaced due to coarse resolution of the emissions). Therefore, I propose to provide additionally overlay plots (maybe only in the supplement) combining the geographical distribution of the modelled concentration and the measured concentrations as an example see Fig. 5 of Knote et al, 2011).

Response: To assess the possibility of geographical shifts in the modelled concentrations, and to present our results in a more illustrative manner, we have introduced a total of 60 overlay maps, distributed over 4 Figures (see Figure 2, 3, 6 and 7 and Section 3.1.1 and 3.2.1 for a description of the respective results). For the statistics assessed in the present study (i.e. the temporal mean and standard deviation of the daily maximum and minimum time series), we have little indications of abrupt spatial shifts in the modelled values. Instead, the modelled values change rather smoothly from one grid-box to another and a shift of the modelled values by a few grid-boxes in any direction would lead to similar verification results.

Referee comment: I guess that the model runtime (and needed resources) of the different experiments differ heavily. It would therefore be very beneficial to confront the performance of the model with of the different configurations with the model runtime and to give recommendations about the trade-off between model runtime and model performance. This could have important implications for other people using CHIMERE. Further, if the authors (or others) plan to use the CHIMERE model for operational forecasts this trade-off would be very important information.

Response: Thanks for this valuable comment. In the last column of the revised Table 3, the runtime (in seconds) of the 8 experiments run with EMEP are listed for a typical summertime heat day. The respective values for the experiments conducted with HTAP are virtually identical, but cannot be stated since the exact values unfortunately have not been saved (see lines 192-194 of the revised manuscript). On the basis of this figure, the reader should be able to make a trade-off between model runtime and performance considering her/his own resources and requirements.

Referee comment: The description of the performed sensitivity studies (Sect. 2.2) is much too short. Readers familiar with the CHIMERE system might be able to follow the description of the authors; readers from outside the 'CHIMERE world' are lost. Please provide more details about the two different vertical grids (e.g. by a figure showing the different levels). Further, please describe the differences of the two chemical mechanisms in more detail. How do they differ? I know that Mailler et al., 2017 provides some details, but details which are very important for this study should be repeated in the manuscript. Further, Menut et al., 2013 already provide a short comparison of the MELCHIOR2 and SAPRC07 mechanisms. How do the findings from the authors

compare to the findings of Menut et al., 2013?

Response: As suggested by you, the description of the applied sensitivity tests has been extended and the orography and model layers for the two horizontal and vertical resolution set-ups have been added to Figure 1 of the revised manuscript. Also, the main conceptual differences between full Melchior and SAPRC have been stated and so are the main differences between the simulations run with reduced Melchior (or Melchior2) and SAPRC found in Menut et al. 2013. However, we here apply the full Melchior mechanism instead of the reduced one, meaning that a comparison with Menut et al. 2013 on equal terms is unfortunately not possible. After stating all this in lines 172-183 of the revised manuscript, we found larger average O_3 maxima for SAPRC than for full Melchior, leading to larger positive biases for the former (see lines 312-313 and Figure 4e).

Specific comments

Referee comment: The line numbers in the manuscript seem to be wrong, at least on page 8. I here refer to the line numbers given in the manuscript

Response: In the revised manuscript, the line numbering has been fixed.

Referee comment: Abstract: Please provide information about the period consider for this investigation in the abstract (e.g. 20.7.2018-31.8.2018).

Response: In the revised manuscript, the study period is now stated right at the beginning of the abstract, as requested by you.

Referee comment: Section 2.1: Studies show that ozone on the Iberian Peninsula is heavily influenced by long range transport (e.g. Pay et al., 2019). Hence, boundary conditions are very important in this region. Therefore, please provide more information about the temporal update frequency of the boundary conditions in Sect. 2.1.

Response: In the revised manuscript, the influence of long range transport on the ozone conditions in the Iberian Peninsula is stated in lines 74-79 and the mentioned study is cited. The boundary conditions data from C-IFS used in the present study are 3-hourly and this is stated in lines 115-116 of the revised manuscript.

Referee comment: Further, the authors mention that chemical boundary conditions stem from different systems (C-IFS, MACC). Further, the meteorological boundary conditions for WRF stem from GFS. Please provide short discussions about the influence of inconsistent chemical (and meteorological) boundary conditions.

Response: The fact that the boundary conditions are from different sources and that our WRF simulations are not forced with IFS but GFS data is something we have to live with since 1) C-IFS does not provide all chemical species necessary to run CHIMERE, which is why we also use MACC data and 2) we unfortunately do not have access to the meteorological forecasts from IFS, which is why we use GFS to force WRF instead. The impact these unavoidable inconsistencies have on CHIMERE's performance is expected to grow with increasing leadtime. Since the leadtime of the analysed time series is relatively short in our study (27 hours from initialization at the utmost), we assume it to be of minor importance here. This is pointed out in lines 124-126 of the revised manuscript.

Referee comment: Further, the authors mention that dust information from C-IFS needed to be scaled. What about the other components? Please provide short information about the quality of the chemical boundary data for the investigated period.

Response: In lines 115-126 of the revised manuscript, the chemical boundary data is now described with more detail. The scaling factor of 0.6 is only applied for desert dust. All other species taken from C-IFS are scaled by a factor of 1 (i.e. no scaling). The forecast skill of the C-IFS model is described in Flemming et al. 2015, which is also cited along these lines.

Referee comment: P6l10 The authors mention that for dust there is only a minor benefit when the model top height is increased. For ozone, however transport from the stratosphere is a very important feature which is missing in the applied set-up. Please

comment on this issue.

Response: By design of our experiments, stratospheric ozone intrusions are inherited from the C-IFS model in which the stratospheric O3 chemistry is parametrized, at least in the L60 model version used here (Flemming et al. 2015). Such intrusions, if present at all, were very weak during our study period (summer 2018) and the associated effects are thus assumed to be negligible.

Referee comment: P8l5 How did the authors sample the model data? The authors took the results at the lowest level in the corresponding grid box, right? But did the authors chose instantaneous model results or temporal (e.g. hourly averaged) model output?

Response: The standard hourly output from the CHIMERE simulations has been compared with hourly average observations. These hourly data were then used to compute the daily minima and maxima. In the revised manuscript, this is stated in lines 223-225.

Referee comment: P9l21 Especially in complex terrain the height of the lowest model layer and the height of the station might not fit together. Therefore, please provide a comparison of the station height and the height of the model at the lowest layer and check how large the differences are.

Response: We found little indications that the model errors at the higher stations, where height differences between the model and the real world are expected to be largest, significantly differ from those at the remaining stations of the same type (background). A more detailed assessment on this issue is interesting but out of the scope of the present paper. It will be undertaken in future studies.

Technical corrections

Referee comment: Figure 2: Please fix the legend ('hores')

Response: This issue has been fixed, thanks for careful reading.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-351,

2019.

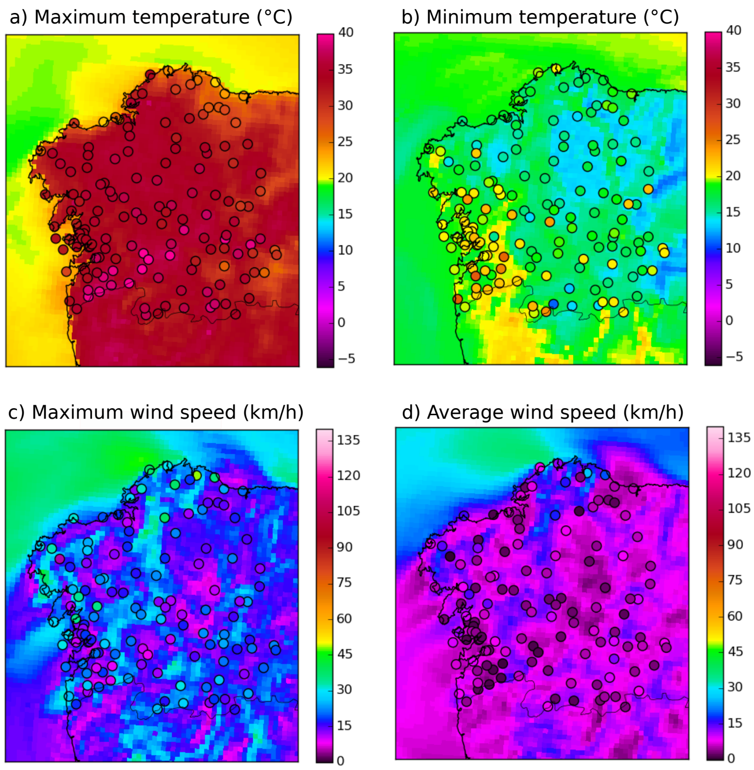

Fig. 1. Suppl. Fig. 1) Daily observations vs. WRF forecasts for a August 5, 2018, (a) max. and (b) min. temperature, (c) max. and (d) average wind speed

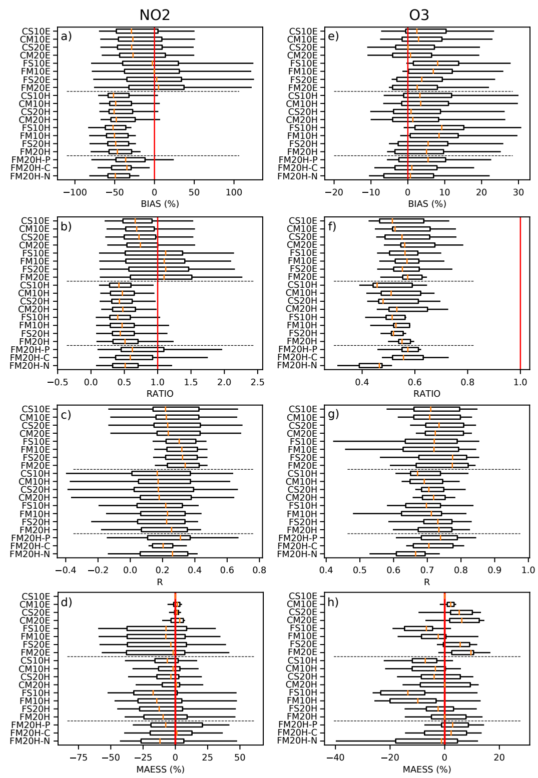

Fig. 2. Suppl. Fig. 2) As Figure 4, but for background sides only