Geosci. Model Dev. Discuss., 7, C724–C735, 2014
www.geosci-model-dev-discuss.net/7/C724/2014/

Geoscientific Model Development Discussions

Interactive comment on "Air quality forecasts at kilometer scale grid over Spanish complex terrains" *by* M. T. Pay et al.

M. T. Pay et al.

maria.pay@bsc.es

Received and published: 3 June 2014

We would like to thanks the Referee #2 for his/her comments which have contributed to increase the quality of the present work. On the one hand, the manuscript has been improved in terms of writing and grammar after a review from a native speaker. On the other hand, all the specific remarks and discussion from Referee #2 have been implemented in the reviewed manuscript.

Please, find in the next paragraphs answers to Referee #2.

Referee #2: While I do agree with the other reviewer's comment regarding the length of the analysis being presented (one month), given the amount of effort required to perform a thorough analysis of the data for multiple domains and grid resolutions, the

short duration does not, in my opinion, significantly harm the analysis presented. However, it does make it impossible to make any general, sweeping conclusions regarding the performance of 4km vs 1km grid resolutions, since model performance can change significantly throughout the year (and from year to year as well). I don't believe the authors make any of these types of generalized conclusions, so that is not an issue. Perhaps in the future the analysis could be extended to a longer time period (perhaps cutting down on the number of domains analyzed).

Authors: As mentioned in the answers to Referee #1, the reason why we selected the present period to study the resolution grid effect is based on the availability, by the time we started the present study, of CALIOPE-AQFS simulations at 1 km resolution during an interesting period in terms of air quality over the three study domains (AND, BCN and MAD).

As the Referee #2 points out, a more comprehensive evaluation could cover for instant a full year. In this sense, an annual evaluation (September 1st 2011-September 1st 2012) for the Barcelona domain has been already discussed in Baldasano et al. (2013) and the results are in accordance with the present work. Anyways, in a future analysis we will expand the period of the analysis to a full year over the three domains. Baldasano, J. M., Arévalo, G., Pay, M.T., and Gassó, S.: Influence of horizontal grid resolution on air quality modelling systems in Barcelona Metropolitan Area (Spain), in: 15th HARMO, Madrid, Spain, 6-9 May 2013, 2013.

Referee #2: And incommensurability between observations and model values will always be an issue, and should probably always be noted, as the comparisons being made are between point observations and grid volume concentrations. But noting whether the measurements are instantaneous values or hourly average values would be useful (same goes for the model values).

Authors: We agree that representativeness challenges continue to be present whenever gridded simulation are compared to observed data at a point in time and space

as modelled concentrations represent a volumetric average over an entire grid cell, meanwhile the stochastic compound embedded in the observations is not accounted for. Measurements have their own uncertainty due to biases and artifacts related to sampling and laboratory analysis methods. The European legislation (2008/50/EC) requires that the uncertainty of measurements meet the air quality objective of 25% for PM10 and PM2.5 and 15% for O3, NO2, and SO2.

As mentioned in the answers to Referee #2 concerning temporal representativeness, in the present comparison both modelled and measured concentrations are hourly averaged. In the case of the CMAQ, the model provides an output file (named ACON*) with hourly averaged concentrations. Concerning observations, which are received in near-real time, the measurements come from automatic monitoring networks, which are hourly averaged by the people who manage those networks.

We have included a comment on that in section 2.4 as follows:

"Representativeness challenges continue to be present whenever gridded simulations are compared to observational data at a point in time and space, as modelled concentrations represent a volumetric average over an entire grid cell, meanwhile the stochastic compound embedded in the observations is not accounted for. Concerning temporal representativeness, in the present comparison both modelled and measured concentrations are hourly averaged. [. . .]"

Referee #2: P2294L4: Define "main pollutants" here.

Authors: The suggestion has been included as follow:

"It provides 48 h forecast of main pollutants (NO2, O3, SO2, PM10, PM2.5, CO, and C6H6) over Spain at 4 km horizontal resolution [. . .]"

Referee #2: P2294L12: Replace "in" with "by". This change applies to the entire article.

Authors: The correction has been amended in the whole manuscript.

Referee #2: P2295L23: Define CAMx and CMAQ here. I don't believe they have been defined yet.

Authors: CAMx and CMAQ have been defined in the revised version of the manuscript as follows:

"Several studies have evaluated the impact of increasing horizontal resolution on different scales over eastern and southeastern USA using the Community Multiscale Air Quality (CMAQ) and the Comprehensive Air Quality Model with Extensions (CAMx) models ranging from 32 km – 12 km – 4 km [...]"

Referee #2: P2296L28: What is meant by "larger spatial concentration"?

Authors: The original sentence:

"[. . .] bottom-up emission inventories provide better performance and larger spatial concentration than down-scaled inventories."

Has been replaced by:

"[. . .] bottom-up emission inventories provide predicted concentrations and corresponding gradients which are more consistent with observed concentrations than those from down-scaled inventories."

Referee #2: P2297L16: Define OPANA.

Authors: OPANA means OPerational Atmospheric Numerical model for urban and regional Areas. The definition has been included in the revised manuscript as follows:

"The lowest resolution system is OPANA (OPerational Atmospheric Numerical model for urban and regional Areas), running at 27 km x 27 km at the Technical University of Madrid [. . .]"

Referee #2: P2298L23: I assume the numbers provided in parentheses are the length of the mountain ranges and not height. That needs to be made clear in the text.

Authors: The numbers between parentheses are the height of the mountain ranges. I have clarified it the text as follows:

“BCN is a coastal area characterized by several valleys perpendicular to the coastal line and two main mountain ranges, one coastal (500 m height) and one pre-coastal (1000-1700 m height), which induce mesoscale phenomena such as sea-breeze and mountain-valley winds.”

Referee #2: P2298L25: What is meant by “Central System”?

Authors: The Central System is one of the main mountain ranges in the Iberian Peninsula (2.592 m height). It has been defined in the manuscript accordingly as follows:

“MAD is a continental region with a much simpler topography (including the mountain range of the Central System located in the northwestern of MAD, with summits reaching 2500 m height, and the Tajo valley in the southern of MAD), which brings different locally-driven flows.”

Referee #2: Figure 1: In Figure 1 the domains are labeled d1-d5, but here they are named. They should be made consistent.

Authors: In Figure 1 the D-domains make reference to the nested sequence of the simulated domains starting from the mother domain (D1, Europe). However, the acronyms AND, MAD and BCN correspond to the study domains (Andalucia, Madrid and Barcelona, respectively). For instant, the impact of horizontal resolution increase in terms of air quality concentrations over the AND domain is analyzed using the simulations from D2 (IP - 4 km) and D3 (AND - 1 km), which in the rest of the paper is referred as 4 km and 1 km simulations, respectively.

This fact has been clarified in the caption of Figure 1 as follows:

“CALIOPE-AQFS nesting strategy (D-domains) and study domains (Andalucia, AND; Madrid, MAD; and Barcelona, BCN).”

Referee #2: P2299L6: What is meant by “logistic”?

Authors: In this sense, “logistic” means “commercial activities”. This error has been amended.

Referee #2: Section 2.2: What land-use data is used in the WRF simulation? Authors: As indicated in the Section 5 (Conclusion), the WRF uses the land-use data from the U.S. Geological Survey (USGS) which is based on the year 1993. This fact has been mentioned in the Section 2.2 as follows:

“The Noah land-surface model (NoahLSM) based on the U.S. Geological Survey (USGS) land-use data is used by default in the present WRF configuration.”

Referee #2: P2300L25: After collapsing, how many CMAQ layers are in the PBL?

Authors: Six CMAQ sigma layers cover the PBL. It has been implemented in the manuscript as follows:

“[. . .] meanwhile CMAQ vertical levels are obtained by collapsing from the 38 WRF levels to a total of 15 layers steadily increasing from the surface up to 50 hPa, with 6 layers along the PBL and a first layer depth of 39 m.”

Referee #2: P2301L4: Why is the reference here to the previous version of CMAQ stated as v4.5? The previous version of CMAQ before 5.0 is 4.7 (and before that 4.6).

Authors: The comparison between CMAQv4.5 and CMAQv5.0 is based on the CALIOPE-AQFS progress. It has been clarified in the revised manuscript as follows:

“Based on evaluation results with the previous CMAQ version within CALIOPE-AQFS (4.5 vs 5.0) (Pay et al. 2012b), CMAQ has been updated to the version 5.0.1, using the CB05 chemical mechanism (Yarwood et al., 2005), the AERO5 for aerosols modeling, as well as in-line photolysis calculation.”

Referee #2: P2301L8: Why use AERO5 and not AERO6?

Authors: The science devoted to the speciation of PM aerosols has significantly improved in AERO6 compared to AERO5 including, for instance, speciation of PM fraction (including Fe, K, Mg, Ca and Ti), primary organic aerosol aging, and some updates in the sulfur chemistry. However, we have decided not to use the AERO6 module for two reasons:

1- The computation cost and the size of the input files (in the emissions) and output files significantly increase in AERO6 because of the increase of the number of species. These are critical issues when working with high resolutions in a forecast mode.

2- There are no specific emission profiles to speciate PM fine emission to new ion species in Spain.

Referee #2: P2301L24: What is meant by "300 min"?

Authors: This 300 min is the computational time used to simulate 48 h of meteorology, emission and air quality.

Referee #2: P2301L27: What is meant by "soft reservation"? Section 2.3: I'm not sure how much value this section adds to the manuscript. Every group uses a different computer configuration for their modeling efforts, so these numbers are really unique to your modeling exercise. While some readers may find the information interesting, I think most readers will not find the information overly useful. If a strong argument can be made for keeping the section, then fine. Perhaps it could be consolidated into a single paragraph however.

Authors: In this case, a soft reservation means a special book over the whole supercomputational resources to be sure that the forecast will run with a sufficient number of CPUs.

I agree with the Referee #2 that this computational setup is unique according to our resources and objective. However, the increase of computational resources and horizontal resolution in forecast issues requires of this kind of setups. In this sense, I have

kept this section, but it has been synthetized as follows:

"Running CALIOPE-AQFS at 4 and 1 km is a technical challenge. The simulations are run on MareNostrum supercomputer (Intel Xeon E5-2670, 16 CPUs and 64 GB RAM memory per node) at BSC-CNS. Table 1 depicts the computational requirements to forecast air quality at 48 h in the domains at hands. The numbers of CPUs are chosen to maximize CPU efficiency. Thanks to the parallelization of meteorological and air quality models, MareNostrum uses up to 256 CPUs. Due to the variable nature and complex dependencies, the computational time to forecast 48 h of air quality fields for the 4 domains is 8-9 hours. The most computational demanding domain is the AND at 1 km resolution (366x358 cells, 256 CPU max., and 300 min). For the April 2013 simulation, times add up to 2880 CPU hours/day, or 86400 CPU hours in one CPU (9.86 years). The storage used for the April 2013 output files was 6.13 TB ($\sim$200 GB/day)."

Referee #2: P2302L12: I assume the 1 ug/m3 is a MINIMUM cutoff. That should be made clear.

Authors: Yes, 1ug/m3 is a minimum cutoff. It has been clarified in the manuscript.

Referee #2: P2302L16: Define METAR.

Authors: METAR means METeorological Aerodrome Report. It has been included in the revised manuscript.

Referee #2: P2303L14: What is meant by "considering the 75% of the values"? It's not really clear here.

Authors: It has been clarified as follows:

"Note that mean and maximum concentrations are calculated considering at least the 75% of the data in the corresponding time base".

Referee #2: P2303L25: What is meant by "maps are conserved"?

Authors: The sentence means that when the resolution increases from 4 km to 1 km, the O3 concentration pattern are similar. However, slightly differences appear along areas with high NOx emission where titration processes are very active. This senesce has been rewritten in the revised manuscript as follows:

“Consequently, when the resolution increases the monthly mean O3 concentration maps are almost identical, although the NOx titration effect on O3 is significant along highways and major point sources”

Referee #2: P2304L8: Not sure the language “significantly better textured” is appropriate here. I think the authors are just trying to indicate that the roadways are more easily identified and better defined in the 1km simulation than the 4km, but that doesn’t necessarily mean they are “better”.

Authors: We agree with the Referee #2, we cannot say they are “better” because they have not been compared with 2D observations yet. We are trying to say that 1 km simulation allows to more easily identified roadways and even mountains. In this sense, we have replaced “better textured”/“significantly better textured” by “more easily identified”/“more textured”.

Referee #2: P2305L25: The authors need to be consistent with their language when describing the results. Here, the authors state “monthly r slightly decreases when resolution increases from 0.67 to 0.58”. That’s a difference of 0.09. However, just above the authors state “slopes significantly improve with resolution increase from 0.72 to 0.77 for NO2 and from 0.50 to 0.54”. Both of those increases are much less than the decrease for PM10. The authors need to be fair here and use consistent language instead of highlighting the improvement as “significant ” and the degradation as “slightly decreases”.

Authors: We totally agree with the reviewer comment. We have commented the results from an objective point of view.

Referee #2: P2305L26: Surely the value here should not be 0.4 (I assume it should be 0.04).

Authors: The reviewer is right. It has been amended in the revised manuscript.

Referee #2: P2306L14: A lot of these values lack units. Units need to be added for all values where appropriate.

Authors: The units have been added accordingly.

Referee #2: P2307L11: I authors say “bias” but the values are in percent, so it must actually be some kind of normalized bias being presented.

Authors: The bias (B) for categorical evaluation is not exactly a real normalized bias. But it is expressed in %. Following the reviewer 1 suggestion a description of categorical statistics has been added in the revised supplementary material.

Referee #2: P2307L16: A number of times the incorrect abbreviation CIS is used instead of CSI.

Authors: It has been corrected.

Referee #2: P2310L3: A reference should be included here regarding the model performance for morning and evening transitions.

Authors: We agree with the reviewer, and some references supporting this issue have been provided. The sentence has been rewritten as follows:

“Simulations by photochemical modeling systems are known not to reproduce faithfully the morning hours after sunrise and the evening hours after sunset when the mixing height experiences rapid changes.”

by

“Several works indicates that WRF does not faithfully reproduce the morning and evening transition over urban environment maybe because it does not model the heat

retention in cities (Makar et al., 2006; Appel et al., 2013)." Makar, P. A., Gravel, S., Chirkov, V., Strawbridge, K. B., Froude, F., Arnold, J., and Brook, J.: Heat flux, urban properties, and regional weather, Atmos. Environ., 40, 2750–2766, 2006."

Referee #2: P2314L10: Change "better captured" to "more evident". Also, the NO2 measurements are likely not made right on the roadway, so it's probably not possible to attribute the improvement in NO2 performance at finer resolution to only the roadways. If the NO2 measurements are made right at the roadways, it would be good to state this earlier in the text regarding the proximity of the NO2 measurements to the major roadways.

Authors: The change "better captured" to "more evident" has been implemented. Concerning the NO2 measurements, as it shown in Figure 2, most of the stations in the urban domains of BCN and MAD are classified according to Garber et al. (2002) as traffic stations, which means they are located at building up areas under the direct influence of traffic emissions. These stations can be located either at the roads or nearby the road. In the case of BCN and MAD where more than 60% of NOx emissions come from the on-road traffic, the improvement of the NO2 performance at finer resolution could be attributed to a more comprehensive modeling of the emission and chemistry near traffic emissions.

Referee #2: P2316L20: This detail should be included earlier in the text. Also, why was such an old land-use data set employed? Using a more up-to-date land-use data set could improve the model results significantly.

Authors: We agree with the Referee #2. A description of the land-use data implemented in WRF has been included in the section 2.2.

We used the USGS land-used data because the WRF works by default with this kind of categories. For next improvement of CALIOPE-AQFS, we have implemented the land-use data from a high resolution and updated data base called CORINE land cover following the methodology of Pineda et al. (2004) to do the assignation between cate-

gories in USGS and CORINE.

Referee #2: P2316L30: What is the CORINE data set? A very brief description of the data would be nice here.

Authors: The CORINE (Coordination of Information on the Environment) Land Cover (CLC) is the database about the coverage and use of the land in the European Union managed by the European Environmental Agency. The CLC has a resolution of 100 m and includes 44 land cover classes. The first inventory was based on 1990 (CLC1990), it has been updated to the year 2000 (CLC2000), and recently to 2006 (CLC2006).

We have included a comment on that in the revised manuscript as follows:

"Furthermore, to get benefits of the resolution increase, the meteorological modeling should include an improved description of the land used instead of the USGS based on the year 1993. In this sense, the Coordination of Information on the Environment (CORINE) provides a high resolution (100 m) land use database developed by the European Environmental Agency updated to the year 2006 (CLC2006) (EEA, 2007) which can be implemented in the WRF model following the methodology described in Pineda et al. (2004)." "EEA: CLC2006 technical guidelines. EEA Technical Report 17/2007. ISBN 978-90-9167-968-3. doi 10.2800/12134, 2007."

Interactive comment on Geosci. Model Dev. Discuss., 7, 2293, 2014.