

***Response to reviewer #3 of “Analysis of summer O3 in the Madrid air basin with the LOTOS-EUROS chemical transport model” by Miguel Escudero et al.***

The authors would like to acknowledge the comments from reviewer #3. We thank the positive comments. All the suggestions demonstrate a good knowledge of the scientific field and that has resulted in a great improvement of the paper after his/her revision.

Before answering to the reviewer’s comments one by one, the authors want to clarify that the main objective of this paper is to perform a detailed analysis of summer O3 events in a topographically complex basin such as the MAB with the valuable help of LOTOS-EUROS CTM. This model was adapted/configured to provide the best representation of the scientific observations resulting from a comprehensive field campaign carried out in Madrid on July 2006. Our main goal was not to realise a deep analysis of the driving factors of LOTOS-EUROS in O3 simulations (physical and chemical schemes, emission patterns...).

We opted to focus our efforts in finding the best configuration for the particular study area and period. In addition to the horizontal resolution, we knew (from previous experimental studies in the region) that the vertical resolution was a key aspect that had not been taken into account so often in previous simulation studies in the Mediterranean. Moreover, the use of two different meteorological input datasets was planned in order to evaluate if there was a relevant difference in the outputs in the two cases. That is a relevant question for CTMs especially when modelling O3.

What is the state of the art terms of fine restitution of the fields of ozone in the Mediterranean? What has been undertaken at this scale so far? What are the current shortcomings in terms of model performance for ozone, what are the locks for the restitution of fine-scale ozone fields, on the horizontal and on the vertical?

**The following paragraphs have been added to the introduction:**

“Air quality model results vary at different resolutions especially due to the resolution of emissions and the description of the driving meteorology (Fenech et al, 2008). Some authors have found that the impact of higher horizontal resolutions in O3 simulations is more sensitive to the resolution of emissions than to meteorology (Valari and Menut, 2008). Moreover, finer resolution result in less dilution of emissions but also in differences have been found in the O3-NOx interaction (Valari and Menut, 2008, Stock et al., 2014, Schaap et al., 2015).

In the Iberian Peninsula, the use of fine grids (in the order of 1-5 x 1-5 km) has been found beneficial in the context of complex terrains where mesoscale processes acquire importance for interpreting production and transport of O3 (Toll and Baldasano, 2000, Jimenez-Guerrero et al., 2008,). In coastal areas, with complex topography, high resolution simulations have been generally employed with good results (Carvalho et al., 2006, Jimenez et al., 2006, Gonçalves et al., 2009). Moreover fine grids have been recommended for describing O3 variability especially in urban and industrial areas (Jimenez et al, 2006, Baldasano et al., 2011). In general, the use of finer resolution simulations in the Iberian Peninsula generally imply benefits in the O3 description such as improvement in correlation and reduction in bias and errors (Jimenez et al., 2006).

Less importance has been given to the vertical resolution mostly because the vertical O<sub>3</sub> profile evaluation of CTM is difficult due to the lack of experimental vertical O<sub>3</sub> data. In complex domains in the Iberian Peninsula the models may not reproduce O<sub>3</sub> concentrations due to a poor representation of mesoscale flows and layering and accumulation of pollutants (Gonçalves et al., 2009). In general, it has been demonstrated that incrementing vertical resolution would help to resolve meteorological phenomena (Carvalho et al., 2006) and would also offer a more realistic vertical exchange between the boundary layer and the free troposphere (Jimenez et al., 2006)."

*What is required to go further in terms of ozone modelling?*

From the generic point of view, there are a variety of questions that could be named as next steps for the improvement of CTMs regarding O<sub>3</sub> modelling. In general, simulating episodes is challenging due to the impact of multi-driving factors (emissions, meteorology, physical-chemical schemes...). Thus, investigating responses to emission changes (including biogenic emissions that are poorly described) is obviously one of the key aspects but it also relevant to find specific configurations for the CTMs in order to provide the best possible estimates of the formation and transport mechanisms that trigger O<sub>3</sub> concentrations especially in summer. For these and other challenges, it is extremely important to evaluate modelling tools with data obtained in campaigns/measurements using state-of-the-art equipment both on the surface and vertically.

More in particular, on the basis of the results obtained in the present work we would like to mention four particular studies/strategies that could favour CTM development:

- Include vertical measurements of NO<sub>x</sub> in field campaigns.
- Perform measurements of NO<sub>y</sub> for carrying out NO<sub>x</sub>/VOC sensitivity studies (according to Sillman's methodology).
- Improve temporal resolution in emission models.
- Study the vertical exchanges of O<sub>3</sub> in detail.

These arguments are now reflected in the paper in the conclusions section:

"The results from this study can be useful to understand the phenomenology of high O<sub>3</sub> episodes in the MAB and to gain knowledge to design appropriate strategies for air-quality management. Further research must be implemented to investigate aspects like the sensitivity to emission reduction scenarios or the role of VOCs with emphasis on the biogenic ones. Moreover, to perform the tasks of validating and optimising CTMs, increasing efforts should be made to conduct more field campaigns in different air basins in the Mediterranean using state-of-the-art equipment to generate data and knowledge about O<sub>3</sub> behaviour both on the surface and vertically. Useful parameters to be included in these campaigns are O<sub>3</sub>, NO<sub>x</sub>, VOC and, when possible, intermediate products like NO<sub>y</sub>, HNO<sub>3</sub> and H<sub>2</sub>O<sub>2</sub> that, according to previous experience (Sillman, 1995), are key parameters for facing model-based NO<sub>x</sub>-VOC sensitivity studies and the assessment of emission inventories. Some of these parameters (especially NO<sub>x</sub>) should also be incorporated in vertical measurements.

In future, similar simulations to the one presented in this study should be performed in the different air basins in the IP where O<sub>3</sub> exceedances have been recorded (Querol et al., 2016). CTMs should be configured specifically for each region or air basin to assure the best performance by capturing the influence of topography and local circulations. For such studies,

we highlight the importance of conducting experimental campaigns that can support the necessary model evaluation.

Finally, it should be noted that when running such fine resolutions for real applications it is also important to work on the emission datasets (out of the scope of this work). Increasing the detail in emission inventory (mainly based on a bottom-up approach) could improve the performance of CTMs when assessing sensitivities or emission scenarios. Moreover, improving time resolution in the emission models can be beneficial for simulating O<sub>3</sub> episodes.”

*In what way does this fine-scale study bring new elements, not only to the evaluation of the model itself, but also to research on the subject?*

The topographical complexity of the MAB along with the large variability of emissions in a conurbation like Madrid required the use of fine scale configurations. In the vertical, it is clear that a considerable number of layers in the model allow a better representation of the vertical variability of O<sub>3</sub> that would have been more difficult to observe and analyse with a coarser resolution (stratospheric intrusion, formation of residual layers, etc). This has been stated by many authors who carried out other fine-scale modelling studies (see discussion about fine resolution studies added to the introduction above). Moreover, the temporal extent of the analysis (1 month) is not common and has offered the opportunity to describe the patterns/scenarios of O<sub>3</sub> episodes in the MAB during summer. Finally, it is innovative the discussion of the role of the BLH because this parameter is highly dependent of the model resolution.

*Also, why do the authors only focus on model resolution?*

In these detailed analysis for regions with complex topography and a great variety of sources, model resolution in the horizontal is an issue. Also vertical resolution was a key aspect on the light of previous studies which studied O<sub>3</sub> phenomenology based on experimental campaigns (Querol et al., 2018 among others). In those studies, vertical fumigation was observed to have a key influence on surface O<sub>3</sub> concentrations. In consequence, if LOTOS-EUROS was able to adequately represent mechanisms such as vertical exchange, formation of O<sub>3</sub> residual layers, high altitude intrusions or fumigation according to what had been observed in the field campaigns, it could be trustfully used for a phenomenological study on O<sub>3</sub> in summer in the MAB. We should bear in mind that LOTOS-EUROS in its standard configuration is a computationally cheap CTM (fast and accurate) and, by this study, we demonstrate that it is also able to reproduce complex events with an obvious resolution adaptation.

*In the conclusions, there is a lack of discussion about the importance of improvements (improved indicators), with regard to the requirements of enhanced configurations (especially CPU time).*

The question of whether it is worthwhile to perform these high resolution simulations (both horizontally and vertically) in terms of CPU time depends on the objectives that you are seeking. In this case, CPU time was increased reasonably when running 70-layer schemes with fine resolution (ECMWF\_70 or WRF\_70) because our objective was to describe O<sub>3</sub> phenomenology in the MAB.

We have added to the conclusions section a discussion about this:

**“The main objective of the paper is to provide a phenomenological interpretation of O3 events in the area after performing a detailed evaluation of the best configuration of the model for the specific area and period. Regarding the specific question of the reasonable number of vertical levels in the model configuration, it is dependent on the objective of the study. In this study the environmental analysis was the main objective and it was logical and feasible from the perspective of CPU time to employ a considerable number of vertical levels because it allowed a better representation of the vertical variability of O3. In other studies such as air quality forecasting or long term analyses in which CPU time may be large, the reasonable number of levels can be less.”**

*It would also be important to give a strategy to choose a future configuration: is it reasonable to choose so many vertical levels (70)?*

**This answer is linked with the previous one. The use of 70 layers allows the model to represent a vertical complexity that would not be possible with the usual number of layers in standard CTM configurations. This was important for this work because we aimed to perform a correct phenomenological study of O3 in a complex region. Is it reasonable to make operational analyses or forecasts with that high number of layers? Probably we cannot afford that but the objective is different in those cases. Just to highlight the relevance of the vertical configuration of the model for O3 simulations, this study has triggered a discussion among the developers of LOTOS-EUROS on incrementing the number of vertical layers in the standard version of the model without putting in risk the computational efficiency of the model. As commented in the previous question, a paragraph discussing about this has been added to the conclusions section.**

*Also, still for a contextualization (this time of the results and not of the stakes), the interpretations of the episodes (which are precise), lack some context. There has been indeed a lot of studies of ozone formation episodes in the literature. The formation of ozone downwind sources, and its dependence upon wind speed and vertical dilution are known: what exactly are the new knowledge at the end of the study, concerning the phenomenology of episodes and their properties, or concerning the ability of a CTM to simulate them? How can these new knowledge improve air quality management strategies?*

**This study provides new knowledge on the occurrence of O3 episodes in the MAB. Among these, we could highlight three. Firstly, it is the first time that summer O3 events in the MAB have been classified in detail, taking into account transport features and the relevance of vertical exchange. This is then a necessary step to investigate emission scenarios and control strategies to be applied. Moreover, this study is innovative because it demonstrates the transfer of O3 produced at the MAB towards other air basins under certain meteorological scenarios (see, for example, section 3.2.3). Finally, regarding the ability of CTM to simulate complex O3 events in Southern Europe, this study demonstrate the relevance of horizontal and, especially, vertical resolution in CTM configuration.**

*Are we just talking about improved scores from one version to another, or do some configurations allow to depict specific features that do not appear in the others?*

**Apart from the relevant improvement of scores associated with increments in the horizontal resolution, it is clear that, for example, the 5-layer schemes (ECMWF\_5 and WRF\_5) cannot describe vertical exchange of O<sub>3</sub> since they do not have enough resolution. This means that, with low number of layers, it is impossible, for example, to observe the formation of residual layers or the influence of stratospheric intrusions.**

*Specific comments:*

*Page 10 – line 5 - “In the plots corresponding to ECMWF\_5 and ECMWF\_70 runs we observe systematic positive bias especially in the period 14–20 UTC when the formation is strong although it only spiked with low wind speed. This feature was not so marked in the three remaining configurations and, in particular, in the two WRF runs the bias values were randomly distributed around zero.” How do the authors explain the mid-day biases of ECMWF compared to WRF? Is it just a problem of resolution of the meteorological calculations, or does it depend on the meteorological model itself?*

**We suspect that it can be attributed mainly to the resolution given that the ECMWF run performed with higher resolution (ECMWF\_HR\_70) responded more closely to the WRF runs.**

*Line 15 page 15: the best overall performance is analyzed on which criteria? Which parameters are used to affirm that the simulation is better? Should it be the restitution of the diurnal peak, the phasing of the morning increase, the total amount on the vertical, or just the indicator average...? In particular, the WRF70 has a strong underestimation of diurnal ozone in Figure 4 at El Pardo and this is still considered as the “best run”. What about this feature at other stations?*

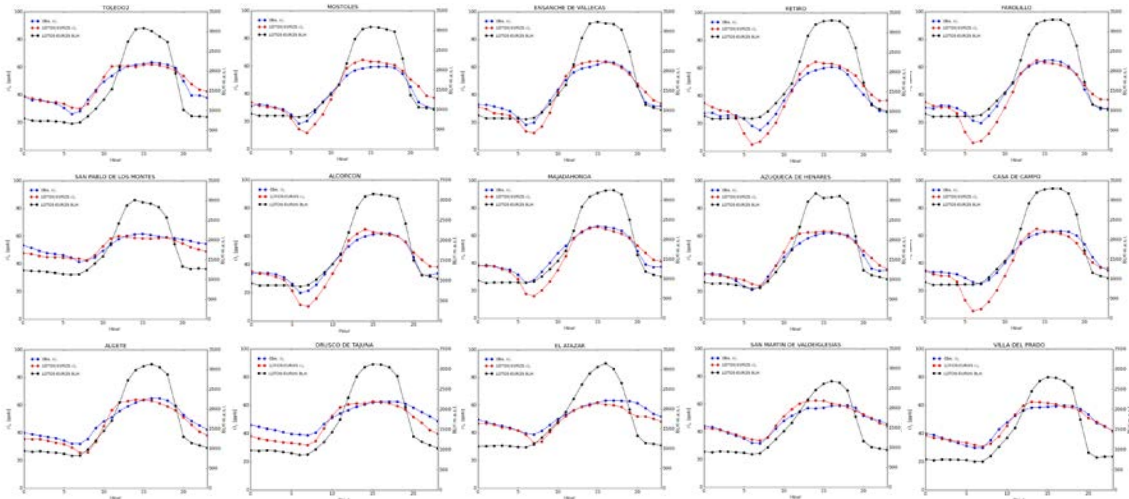
**The best overall performance was selected from all parameters obtained in the validation assessment:**

- i. From the inspection of r values we concluded that the configurations with higher vertical resolution perform better than the 5-layer schemes.
- ii. From the analysis of the model bias (Figures 2, 3 and S3), we concluded that clear improvements were observed using finer spatial resolution so either WRF simulations (WRF\_5 and WRF\_70) or ECMWF\_HR\_70 were the best option on this regard.
- iii. The assessment of NMSE (Figure 2) revealed that WRF\_70 showed slightly lower errors than the other configurations although differences were not drastic.
- iv. After the evaluation of mean daily cycles (Figure 4) we found that the five-level configurations (ECMWF\_5 and WRF\_5) presented a not realistic sharp increment in O<sub>3</sub> concentrations from 6 to 7 UTC. Although the timing of the increase was better reproduced in these runs than in the 70-level schemes, the latter were more realistic in the daily evolution of O<sub>3</sub> concentrations (without steep increases in the morning).
- v. From the comparison of vertical profiles, we concluded that obviously high resolution in the vertical was needed to capture the layering of pollutants reflected in the O<sub>3</sub> soundings.

**Summarising, the best overall configuration for the analysis of O<sub>3</sub> in this region/period were setups with high resolution both in the vertical and in the horizontal, namely ECMWF\_HR\_70**

and WRF\_70. In particular, we used WRF\_70 for vertical cuts because spatial resolution over the study area was better.

The underestimation of diurnal O3 in the mean daily cycle at El Pardo was not present in other stations on a routine basis. The following pictures present cases in which the mentioned underestimation is not observed.



The difference between what is observed and simulated is not always specified, even if we can guess it. Example: in "Figure 7. Longitudinal and latitudinal vertical crosssections of NO2 and O3 for 16 July 2016". Same for figure 9 and figure 11. Also, for page 17 lines 1 to 10, it is important to specify that it is a vision drawn by the model and not the result of observations.

We have made this clear in those Figures' captions. In page 17 we have added a comment making clear that the interpretation is made by means of the analysis of simulation results:

"From the simulation results, we can interpret the evolution of O3. At 18 UTC of 16 July, we can see how O3 levels increased drastically, and the boundary layer depth grew up to 3500 m a.s.l. aided by convection at 18 UTC. Normally, REC events show higher planetary boundary layer (PBL) heights in the evening. Figure 7 shows how the strong convection during REC events injected ground-level pollutants at high altitudes during the late afternoon and the evening reaching up to 3500 m a.s.l. as illustrated in the NO2 plots. When the night-time stable boundary layer forms after sunset, air masses with high O3 that originated near the surface during the previous day were decoupled and remained in the residual layer at altitudes ranging between 2000 and 4000 m a.s.l. forming reservoir layers (00 and 06 UTC cross-sections in Figure 7) which can fumigate the following day. These reservoir layers can also be observed as a relatively thin band at an altitude of 2000–4000 m a.s.l. during every night of the REC period (Figures 6 and S5)."

Figures 6, 8, 10: The location/typology of the station groups should be mentioned.

This information is presented in Figure 1. Thus, in order to clarify we have added in the figure caption of Figures 6, 8, 10 the reference to Figure 1.

*Figure 7: it would be more convenient for the reader to visualize on a map the latitudinal and longitudinal cuts.*

**Exactly as in the previous comment, the map with the latitudinal and longitudinal cuts is presented in Figure 1. We refer again in the figures' captions to Figure 1.**