Responses to the comments of anonymous referee #1

We would like to thank for the comments which helped to improve our manuscript. Please find below your comments in blue, our responses in black and modifications in the revised manuscript related to technical or specific comments in italic and inside quotes. Additional text modifications have been done as most figures have been updated following your suggestions. All modifications are highlighted in the revised manuscript.

General comments and overall quality

This paper is mainly focussed on the agreement between modelled and observed surface ozone in Europe in the year 2010 using various sensitivity model runs to identify reasons for mismatches between the modelled and observed levels of ozone. Other species like CO, SO2 and PM2.5 are also included but only to a very small extent. The work is presented in a clear and sound scientific way with no major errors, and overall this is as a robust and well-performed study with interesting findings that certainly deserves to be published. Some questions and comments are given in the following.

A few general comments: Neither the title, abstract or conclusions mention anything about the additional species (SO2, CO and PM2.5) being included in the work. Furthermore, these species constitues a very small part of the paper and apparently with a fairly small implication for surface ozone which is the main focus of the paper. Thus, one could consider to take out these species completely. This is left to the authors (or the editor) to decide.

Thank you for this comment. We would like to keep them to show, even in a small extent, a more general model performance for a better comparison with past and future studies.

A main issue when doing comparisons between observed and modelled ozone is the question of how to treat vertical concentation gradients near the surface. During the summer season the effective dry deposition and uptake in vegetation will lead to signicifant gradients in ozone near the ground. Since the air intake of the ozone monitors mostly are at around 2 m the gas concentrations at that altitude in some way need to be related to the mean concentations in the model's lowest layer, in this case around 20 m. Has this issue been considered and if not – how important could this effect be?

The height of the first layer is indeed very important for the mixing and deposition processes. Menut et al. (2013) showed that setting the top of the model's lowest layer at a lower altitude (8 m compared to 40 m) results in consistent reductions of modeled (for August 2009 in Europe) ozone concentration by 3-12 μg m^{-3} (~1.5-6 ppb) as ozone deposition is enhanced, especially over forested areas. Also, Travis et al. (2017) reported a decrease of ~3 ppb in ozone mixing ratios when accounting for the subgrid vertical gradient between their lowest model level (centered 60 m above ground) and their measurement altitude (10 m). This decrease in ozone was evident for their whole ozone concentration probability

distribution, shifting the distribution to the left (towards the lower ozone concentration values). In most of the applications, the regional models use a surface layer thickness between 20 and 90 m and in models with coarse surface layer (such as 90 m), usually a correction is implemented to represent surface concentrations (Bessagnet et al., 2016). In our application, the top of the lowest layer is at 20 m, but the ozone mixing ratios are calculated at the mid-point of each layer. So in this case the modeled ozone mixing ratios in the lowest layer are at 10 m which is reasonable for regional modeling. We added a sentence to clarify that:

"We used 14 sigma layers going up to 460 hPa with the first layer being approximately 20 m thick. The concentrations are calculated at the mid-point of a given layer, so the modeled values of the first layer correspond to a height of approximately 10 m."

Underestimation of the high peak values is commonly seen in almost every model study. The authors should include some discussion on this general feature with references to a number of relevant modelling papers. Could it be that this artefact is reflecting the unavoidable smoothing (in emissions, meteorology etc) that all CTM relies on?

Several studies (Valari and Menut, 2008; Markakis et al., 2015; Schaap et al., 2015; Kuik et al., 2016) have investigated the impact of the use of a finer model and emissions resolution on the CTM performance and they showed that it can improve the model performance for urban areas but has no significant impact on rural ones (for resolutions higher than the one used in this study (0.250 x 0.125)). Furthermore, it was shown that a finer resolution in the emissions and the model could lead to consistent overall reductions in ozone compared to the coarser resolution due to enhanced ozone titration by NO_x in the urban areas. Therefore, we conclude that the model and emissions resolution of this study cannot be held responsible for the underestimation of high ozone mixing ratios.

We added a sentence to refer to other studies (Solazzo et al., 2012; Im et al., 2015) that report the model underestimation of high ozone concentrations for a variety of different models and parameterizations in Europe:

"Similar model bias patterns as in this study were also reported by other studies for a variety of different models and parameterizations in Europe, the vast majority of which showed overestimation of low ozone concentrations and significant underestimation of the high ozone levels (Solazzo et al., 2012; Im et al., 2015)."

Specific comments

P3 L8 (and Fig 1). The definition of sub-regions and in particular sub-region 3 seems a bit odd. If the point is to divide Europe into areas with homogenous characteristics with respect to climate and air pollution statistics, then region 3 doesnt seem a very natural choice since it merges clean background sites (West coast of Ireland) with central European sites (e.g. Czech Republic). Apart from perhaps the most northern part, a latitudinally based definition of sub-regions is not very meaningful for Europe. Thus, it would make more sense to split region 3 into two regions or to create another set of sub-

regions better reflecting climatological patterns. (See e.g. the PRUDENCE regions: http://ensemblesrt3.dmi.dk/quicklook/regions.html)

Thank you for this comment. We followed your suggestions and divided our domain in PRUDENCE regions, but with some modifications: i) the separation of the Benelux region for the reasons stated in the text, ii) we kept the Po Valley region instead of the more general "Alps" region in the PRUDENCE regions, iii) we grouped some stations in central France with the stations in central Europe (ME region), and iv) we grouped the few stations in northeastern Europe in the Eastern Europe region (EA) instead of Scandinavia region (SC). We updated all the respective figures and modified some of the text to properly describe and discuss the updated results.

P3 L9. With the model top at 460 hPa, the domain seems shallow compared to the model setup that is normally used for regular modelling in Europe. The authors should inlude some sentences justifying this choice of vertical range.

For the meteorological simulation with WRF we used 31 layers (up to 100 hPa) and for the air quality simulations with CAMx we used a selection of 14 out of 31 layers with higher resolution closer to the surface. We performed an additional simulation with the base case parameterization but this time including all 31 layers for June 2010 (and the last 2 weeks of May that were used as spin-up). Please note that this CAMx test simulation with the 31 layers was performed only for June and not for the whole summer (JJA). The results are shown in Fig. 1 below (we included this figure in the revised supplementary material as Fig. S1). The effect of extending our vertical range to higher altitude and increasing our vertical resolution leads to a small reduction in ozone (by ~1–2 ppb). This slightly improves the model performance in the lower ozone bins but also slightly worsens it for the higher ozone bins. Overall, the impact is quite small which is in line with other studies that have also reported a small sensitivity of the surface ozone to the refinement of the vertical mesh (Menut et al., 2013; Markakis et al., 2015). We added a sentence such that our choice of the vertical range seems more justified:

"Additional tests showed that higher vertical resolution with layers up to 100 hPa would have a negligible effect on surface ozone (see Fig. S1) as also shown by other studies (Menut et al., 2013; Markakis et al., 2015)."

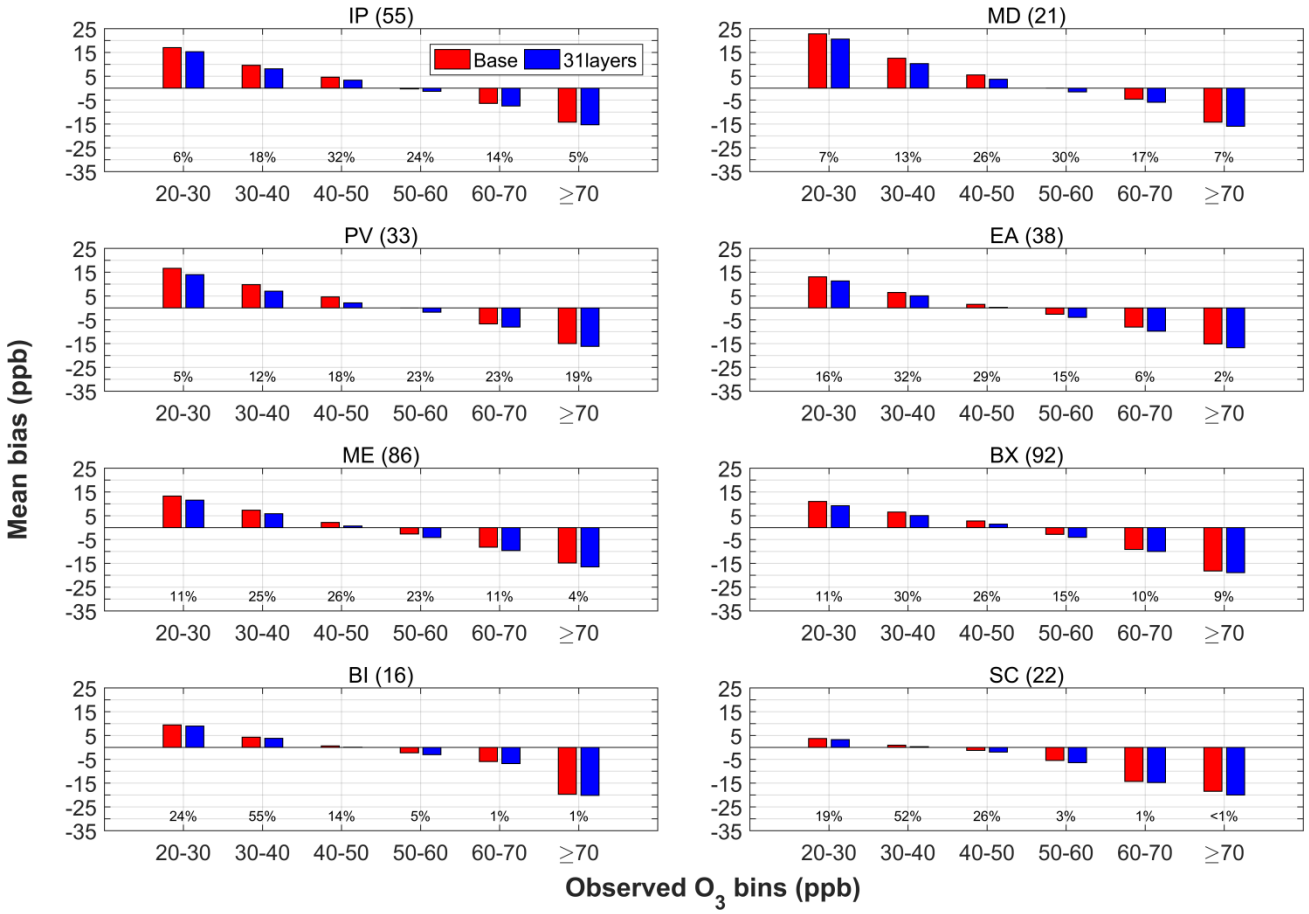

Figure. 1. Effect of increasing vertical resolution (31 layers up to 100 hPa instead of 14 layers up to 460 hPa) on the mean afternoon (12:00–18:00 UTC) bias for surface O_3 mixing ratios in 8 European regions in June 2010. Percentage values below the bars indicate the fraction of the values assigned to each bin for each region. The number of stations available for each region is reported in parentheses at the top of each panel.

P7 L4. For studying ozone peak values the time period 11-16 UT is selected "when the ozone production and mixing ratios often reach their maximum", the author states. This seems as a bit narrow and early to capture the highest ozone peak values. On average, for the entire 6-months summer season, 11-16 UT may be the peak period in some regions (see Fig 3). However, during high ozone episodes the peak values will often occur later in the day, and a period 12-18 UT would seem a more natural choice or even 14-20 UT.

We agree and adapted the 12:00–18:00 UTC time interval, which corresponds to local summertime of 14:00–20:00 for the UTC+1 timezone countries where most of the stations (used in this study) are located.

P7 L14. Some details (geographical location and alitude) of the 8 stations with data on both T and O3 should be given, e.g. in a map.

We included a table in the supplement (Table S4) which provides the geographical location and elevation of these 8 stations.

P10 L17 (and Fig 4). How representative are the mean diurnal cycle of NO2 for this very large region? The header states that only 8 sites are included, and presumably (with some knowledge of the Airbase data) most of these sites are from the Northern UK?

After the implementation of the different regions separation, there are now only 3 stations in Scandinavia (SC region). Since this a small number of stations for such large region, the NO_2 results should not be interpreted as a robust representation of the whole SC region. We added a sentence to clarify that:

"However, for BI and SC regions the NO_2 results should not be interpreted as a robust representation of the whole region due to the small number of sites (4 and 3 respectively) that are included."

Technical corrections

P2 L8-9. Consider rewriting this sentence: "Apart from the ozone precursor emissions, the other key driver of the surface ozone concentrations, as well as its chemistry, is the meteorology; from local to global scale". To state that meteorology is a "key driver" of surface ozone concentrations is somewhat meaningless without a few words explaing how met could affect ozone.

We added a few words to explain how meteorology can affect ozone:

"For example, on the local scale changes in shortwave solar radiation and temperature can directly influence the ozone photochemistry, and changes in wind speed or vertical mixing can lead to accumulation or dilution of the ozone precursor concentrations as well as ozone itself. On the global scale, changes in atmospheric circulation patterns can influence the continental transport of ozone concentrations and its precursors, the stratosphere–troposphere ozone exchange and the local meteorology."

P2 L13-14. This sentence is imprecise. Although T is peaking in the afternoon, incoming solar radiation is not. Rewrite.

We corrected the sentence:

"The peak values of surface ozone concentrations usually occur in the summer afternoon hours when the temperature reaches its diurnal maximum and the incoming solar radiation is still ample."

P2 L18. are -> is

Corrected.

P2 L19. This phrase should be reformulated and clearified: "The evaluation of modeled ozone production from the ozone concentrations may not be a safe option."

The sentence was reformulated:

"The evaluation of modeled ozone production by just comparing modeled ozone concentrations with measurements may be misleading, as an agreement between modeled and observed ozone concentrations might just be the result of compensating errors."

P5 L4. This phrase should be reformulated: "… large discrepancies have been obvious …"

We corrected that phrase:

"… large discrepancies have been observed …"

P7 L14: Rephrase this: "… surface stations which contain both temperature and ozone …". (The station doesn't "contain" temperature and ozone.

We rephrased that part of the sentence:

"… surface stations (see Table S4 for details), which have measurements of both temperature and ozone, …"

P8 L23. The text and the caption of Table 5 should explain for which time period (summer season?) these statistics were based on and for what type of data (hourly, afternoon means or something else?).

We updated the caption of Table 5:

"Table 5. Model performance evaluation for the daily mean concentrations of the chemical species in summer (JJA) 2010. The units for MB, MGE and RMSE are in ppb for the gas species and in $\mu g\ m^{-3}$ for the $PM_{2.5}$."

We updated the text:

"The overall model performance for the daily mean concentrations of the air pollutants in summer (JJA) 2010 (Table 5) was reasonably good."

P10 L13. Typo: "overestimation" should be changed to "underestimation"

Corrected.

P10 L17. Rewrite. The word "now" doesn't seem appropriate. Change e.g. to "in this region" or something else.

We reformulated that part of the sentence:

"…appears to be more pronounced…"

Fig 5 and Fig 6. The time period (11-16 UT?) which the afternoon average is based on should be given in the Figure captions.

We updated all Figure captions to include the time interval that refers to the afternoon average (12:00–18:00 UTC).

P12 L15. Rephrase this: "… seems to be the most effective scenario …" (The point is presumingly that the model scenario with increased emissions of both NOx and VOC is the scenario that gives the best fit with the observations).

We changed that to:

"… has the most effective improvement in the model performance for the BX region …"

One additional remark:

Fig S10 in the Supplements shows the mean vertical O3 profile as measured by sondes and modelled for six European sites for summer 2010. All these profiles indicate a fairly marked underestimation of ozone by the model at the top model layer (about 4 km or a bit higher). Perhaps the authors could comment on this and on what possible consequenses this could have for the model performance at the surface. Does this reflect a systematic bias (underestimation) in the boundary condition of ozone at the upper boundary? This question should be seen together with my previous comment that the vertical extent of the model domain seems somewhat shallow compared to standard CTMs that typically extend up to the tropopause. This question also leads to the question whether part of the systematic underestimation of ozone at the surface could be explained by the combination of too low concentration used as upper boundary conditions, a shallow vertical model range and uncertainties in the vertical exchange processes of ozone (the vertical gradients from the sondes differ markedly from the modelled ones).

Our model top layer is at about 4.5-5 km a.g.l. Regarding the impact of a higher vertical range and resolution, we discussed it in detail in your specific comment (2) above where we showed that this impact was small. Regarding the ozone vertical profiles, the modeled vertical gradients seem weaker which might be an indication of increased vertical mixing. Similar issues were encountered and addressed by Travis et al. (2017), which were attributed to increased top-down mixing. When they applied some corrections to reduce the top-down mixing, the modeled and observed gradients were in much better agreement and the modeled surface ozone concentrations consistently decreased by 2–3 ppb as less ozone was transported from the upper layers to the surface. Therefore, this uncertainty could partially be responsible for the overestimation of the low ozone mixing ratios but a potential correction would also lead to a higher model underestimation in the high ozone bins. This issue of a potential increased top-down vertical mixing does not seem to be strongly related to our vertical structure choices, as the modeled ozone vertical gradients did not change significantly when we included all 31 layers as shown in Fig. 2 below. Regarding the large model underestimation of ozone in the upper layers, this could be related to a systematic bias in the boundary conditions, as has been also reported by other modeling studies (Giordano et al., 2015; Im et al., 2015; Solazzo et al., 2017). However, as we have shown in Fig. S12, the impact of increasing both the lateral and top boundary conditions

of ozone diminishes closer to the surface and into the interior of the domain. Furthermore, an increase in the ozone boundary conditions will increase the modeled ozone in all ozone bins and so the reduction of the model's underestimation in the high ozone bins will be accompanied by an overestimation in the rest of the ozone bins. In addition, with a reduced top-down mixing the effect of the higher altitude ozone concentrations on the surface will be smaller. Therefore, we conclude that the aforementioned uncertainties seem to be more related to the surface ozone overestimation than underestimation.

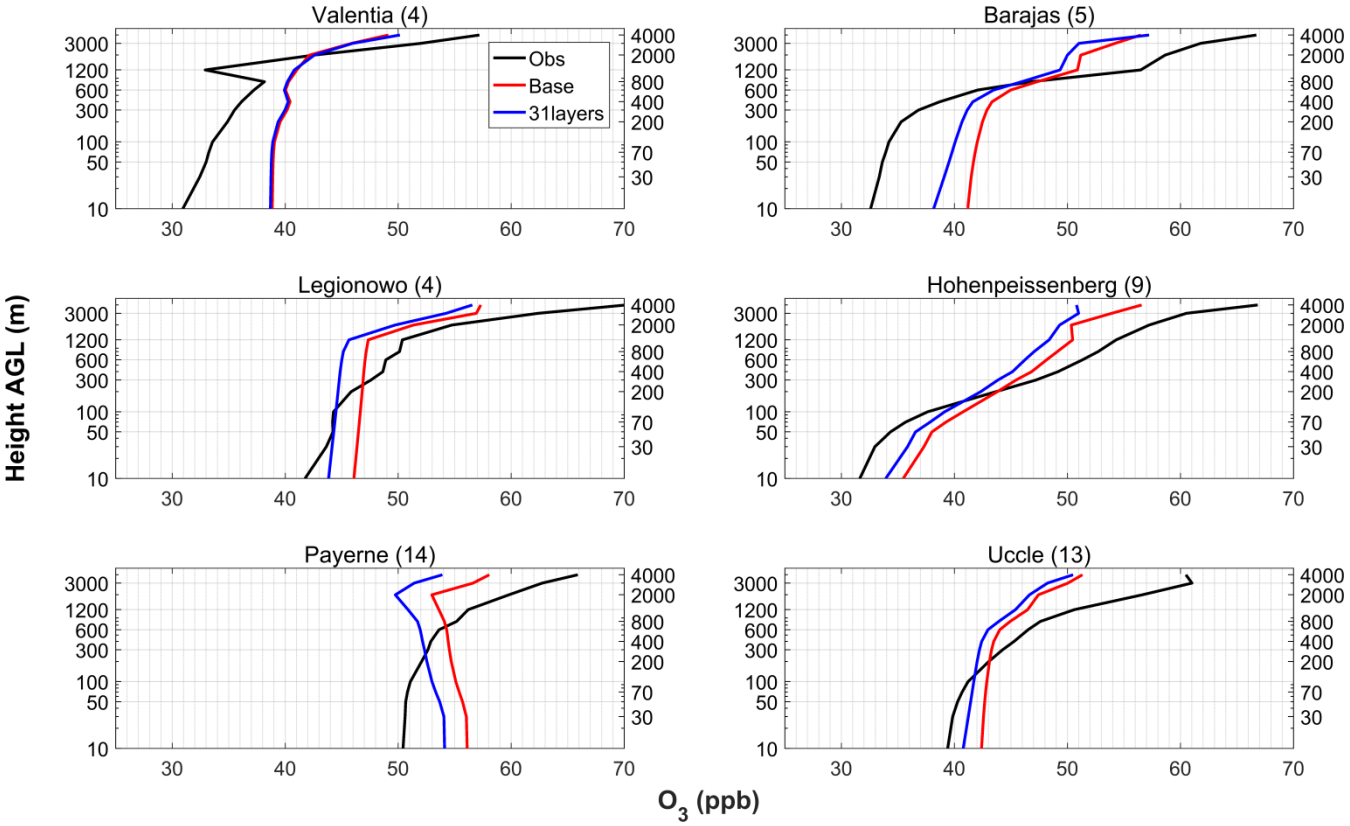

Figure 2. Effect of increasing vertical resolution (31 layers up to 100 hPa instead of 14 layers up to 460 hPa) on the ozone vertical profiles for 6 stations in June 2010. The number of ozonesondes available for each station is reported in parentheses at the top of each panel. Heights of 14 model layers are shown on both y-axes which are in logarithmic scale.

References

Bessagnet, B., Pirovano, G., Mircea, M., Cuvelier, C., Aulinger, A., Calori, G., Ciarelli, G., Manders, A., Stern, R., Tsyro, S., García Vivanco, M., Thunis, P., Pay, M. T., Colette, A., Couvidat, F., Meleux, F., Rouïl, L., Ung, A., Aksoyoglu, S., Baldasano, J. M., Bieser, J., Briganti, G., Cappelletti, A., D'Isidoro, M., Finardi, S., Kranenburg, R., Silibello, C., Carnevale, C., Aas, W., Dupont, J. C., Fagerli, H., Gonzalez, L., Menut, L., Prévôt, A. S. H., Roberts, P., and White, L.: Presentation of

the EURODELTA III intercomparison exercise – evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology, Atmos. Chem. Phys., 16, 12667-12701, doi:10.5194/acp-16-12667-2016, 2016.

Giordano, L., Brunner, D., Flemming, J., Hogrefe, C., Im, U., Bianconi, R., Badia, A., Balzarini, A., Baró, R., Chemel, C., Curci, G., Forkel, R., Jiménez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J. J. P., Makar, P. A., Manders-Groot, A., Neal, L., Pérez, J. L., Pirovano, G., Pouliot, G., San José, R., Savage, N., Schröder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Žabkar, R., Zhang, Y., and Galmarini, S.: Assessment of the MACC reanalysis and its influence as chemical boundary conditions for regional air quality modeling in AQMEII-2, Atmos. Environ., 115, 371-388, doi:10.1016/j.atmosenv.2015.02.034, 2015.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Flemming, J., Forkel, R., Giordano, L., Jiménez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J. J. P., Makar, P. A., Manders-Groot, A., Neal, L., Pérez, J. L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., and Galmarini, S.: Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone, Atmos. Environ., 115, 404-420, doi:10.1016/j.atmosenv.2014.09.042, 2015.

Kuik, F., Lauer, A., Churkina, G., van der Gon, H. A. D., Fenner, D., Mar, K. A., and Butler, T. M.: Air quality modelling in the Berlin-Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data, Geosci. Model Dev., 9, 4339, doi:10.5194/gmd-2016-190, 2016.

Markakis, K., Valari, M., Perrussel, O., Sanchez, O., and Honore, C.: Climate-forced air-quality modeling at the urban scale: sensitivity to model resolution, emissions and meteorology, Atmos. Chem. Phys., 15, 7703-7723, doi:10.5194/acp-15-7703-2015, 2015.

Menut, L., Bessagnet, B., Colette, A., and Khvorostiyanov, D.: On the impact of the vertical resolution on chemistry-transport modelling, Atmos. Environ., 67, 370-384, doi:10.1016/j.atmosenv.2012.11.026, 2013.

Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J. M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M. T., Rouïl, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., and Wind, P.: Performance of European chemistry transport models as function of horizontal resolution, Atmos. Environ., 112, 90-105, doi:10.1016/j.atmosenv.2015.04.003, 2015.

Solazzo, E., Bianconi, R., Vautard, R., Appel, K. W., Moran, M. D., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J. H., Chemel, C., Coll, I., Denier van der Gon, H., Ferreira, J., Forkel, R., Francis, X. V., Grell, G., Grossi, P., Hansen, A. B., Jeričević, A., Kraljević, L., Miranda, A. I., Nopmongcol, U., Pirovano, G., Prank, M., Riccio, A., Sartelet, K. N., Schaap, M., Silver, J. D., Sokhi, R. S., Vira, J., Werhahn, J., Wolke, R., Yarwood, G., Zhang, J., Rao, S. T., and Galmarini, S.: Model

evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII, Atmos. Environ., 53, 60-74, doi:10.1016/j.atmosenv.2012.01.003, 2012.

Solazzo, E., Bianconi, R., Hogrefe, C., Curci, G., Tuccella, P., Alyuz, U., Balzarini, A., Baró, R., Bellasio, R., Bieser, J., Brandt, J., Christensen, J. H., Colette, A., Francis, X., Fraser, A., Vivanco, M. G., Jiménez-Guerrero, P., Im, U., Manders, A., Nopmongcol, U., Kitwiroon, N., Pirovano, G., Pozzoli, L., Prank, M., Sokhi, R. S., Unal, A., Yarwood, G., and Galmarini, S.: Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown, Atmos. Chem. Phys., 17, 3001-3054, doi:10.5194/acp-17-3001-2017, 2017.

Travis, K. R., Jacob, D. J., Keller, C. A., Kuang, S., Lin, J., Newchurch, M. J., and Thompson, A. M.: Resolving ozone vertical gradients in air quality models, Atmos. Chem. Phys. Discuss., 2017, 1-18, doi:10.5194/acp-2017-596, 2017.

Valari, M., and Menut, L.: Does an Increase in Air Quality Models' Resolution Bring Surface Ozone Concentrations Closer to Reality?, Journal of Atmospheric and Oceanic Technology, 25, 1955-1968, doi:10.1175/2008jtecha1123.1, 2008.