The authors would like to thank Anonymous Referee #1 for their constructive comments. Below are our responses.

**Page 3 line 54-55:** I disagree with this statement. Many European WRF-Chem modelling evaluation studies have been published in the last few years.

We have added additional citations of papers that apply WRF-Chem over Europe. However, in our view, studies that focus on evaluation over the whole European domain are still limited to date. If there are particular studies that fulfill this criteria that are not being discussed in the manuscript, the authors would appreciate it if the referee would mention the papers specifically. The sentences in question have been updated as follows to improve clarity: “The use of WRF-Chem over Europe has increased in recent years (e.g., Forkel et al., 2012; Žabkar et al., 2015; Solazzo et al., 2012a, b; Tuccella et al., 2012; Zhang et al., 2013a, b). However, only a limited number of these studies are dedicated to the evaluation of WRF-Chem-simulated meteorology and chemistry over the whole European domain.”

**Page 10 line 293-295:** Please, in order to prove that differences between the two meteorological simulations are negligible provide statistical indexes or a comparison figure in the supplementary material.

A table and figures showing the meteorology from the RADM2 simulation has been added to the supplementary material; see Table S1 and Figures S4-S7. Furthermore, the manuscript has been updated as follows to directly address this question.

“Differences in predicted meteorology between the MOZART and RADM2 simulations are small, with differences in MSLP less than one hundredth of 1%, and differences in T2, WS10, and WD10 generally far below 1%. Since the simulations were run without aerosol-radiative feedbacks, it was expected that the two simulations would show minimal differences in meteorology, and we conclude that differences in O3 and NOx predicted in the MOZART and RADM2 simulations (Section 4.2) are a direct result of differences in the chemistry, rather than chemistry-radiative feedbacks. Statistics for meteorology for the RADM2 simulation can be found in the Supplementary Material, Table S1 and Figures S4-S7.”

**Page 15 line 478:** to be in line with the NOx (NO2 and NO) treatment in MOZART simulation, I suggest to briefly explore NO concentrations in RADM2

A discussion of NO concentrations in RADM2 has been added to the revised manuscript.

“Like for MOZART, NO for RADM2 is underpredicted throughout the domain, with NO concentrations slightly more negatively biased than in MOZART in all seasons except Fall, when NO concentrations are higher for RADM2 than for MOZART and show better agreement with the observations. Temporal correlation for NO2 and NO in RADM2 is also found to show similar behavior to the MOZART simulation.”

**Figure 1:** I suggest to represent temperature using the International System unit (K) here and everywhere else in the text.

The authors prefer to keep temperature in units of Celsius. Although it is not the SI unit, Celsius is widely used in the meteorological community, and is also used in GMD publications (see, e.g.,
Furthermore, when calculating relative bias statistics (MB, NMB, MFB) for temperature as in Table 4, using Kelvin rather than Celsius makes the denominator extremely large and the bias extremely small, making relative bias statistics less meaningful. However, if the editor agrees that the temperature unit should be changed to Kelvin, we will make these changes to our manuscript.
The authors would like to thank Anonymous Referee #2 for their constructive comments. Below are our responses.

1) lines 50-52: The authors give here three examples of air quality models but maybe they could also refer here to the review article of Baklanov et al. (2014) for the online coupled regional meteorology chemistry models in Europe.

The original manuscript did include a citation to Baklanov et al. 2014 (line 53-54). However, we have added in the revised manuscript a more detailed reference to this manuscript: “The application of online coupled regional meteorology chemistry models in Europe, among them WRF-Chem, has been recently reviewed by Baklanov et al. [2014].”

2) lines 62-64: The importance of time variant chemical boundary conditions for simulated near surface ozone over Europe has been also highlighted in other recent regional modelling studies (see e.g. Akritidis et al., 2013).

Following the referee's suggestion, the manuscript has been extended, mentioning a the importance of temporally varying chemical boundary conditions.

“The importance of temporally varying chemical boundary conditions in air quality modeling has also been stressed in other studies (including Akritidis et al., 2013; Andersson et al., 2015).”

3) line 264: Please provide some more information on the selection of the AirBase stations classified as rural background. Do you include stations with class 1–3 according to the Joly-Peuch classification methodology for surface ozone (Joly and Peuch, 2012). This approach has been also applied in a recent study by Katragkou et al. (2015) for the evaluation of MACC reanalysis near-surface ozone over Europe.

We used the classification of stations provided with the metadata in AirBase. This is now indicated in the revised manuscript in Section 3.2.2.

“Because of the relatively coarse horizontal resolution in this model study, model output is only compared against AirBase stations that are classified as "rural background." The station classification was taken from the metadata provided by the EEA for AirBase.”

4) line 283: You may add one sentence with information for the use and value of SOMO35 index.

A brief discussion of the purpose and use of the SOMO35 metric has been added to the manuscript in Section 3.3.

“SOMO35 is an indicator of cumulative annual ozone exposure used in health impact assessments. The accumulated health impact is assumed to be proportional to the sum of concentrations above a cutoff of 35 ppb, chosen because the relationship between O3 and adverse effects is very uncertain below this threshold (WHO, 2013).”
5) Looking the Figures 4 and 9 I am wondering why at the lateral boundaries there are such differences between the two simulations with the different chemical mechanisms (RADM2 and MOZART) even though they are constrained with identical O3 chemical lateral boundary conditions.

The importance of ozone import into the model domain from the lateral boundary conditions depends not only the concentration at the lateral boundary conditions (as the reviewer notes, in the case of MOZART and RADM2 simulations, these concentrations are the same), but also on the dominant wind flows at the edge of the domain. A plot of seasonally averaged wind vectors from ERA-Interim for 2007, which are the fields used to force model meteorology at the edges of our domain, has been added to the Supplementary Material (Figure S2). The dominant flow of air onto the European continent is from the west, and we see that the western (particularly northwestern) edge of the domain is where seasonally-averaged O3 values are most similar between the MOZART and RADM2 simulations. At the northwestern edges of the domain, we see that seasonal average O3 predicted by RADM2 is generally not more than 5% lower than that predicted by MOZART. At the southern and eastern edges of the domain, there is not a strong flow of air into the model domain, which dampens the impact of ozone boundary conditions in this area.

In addition to the addition of Figure S2 to the Supplementary Material, we have made the following addition to the text in Section 4.2.1.

“Absolute O3 concentrations are most similar (i.e., less than 5% different) between the mechanisms near the northwest edges of the domain (see Figures 4 and 9), where the prevailing westerly winds (Supplementary Material, Figure S2) mean that O3 imported from the boundary conditions plays a dominant role.”

6) lines 546-551: Normally with NOx titration we mean the first order removal process of O3 through direct reaction with NO which takes place during nighttime and in the vicinity of large NO emission sources. However the presented results refer to summer daytime and maybe this behaviour is related to the saturated NOx conditions (or VOC sensitive conditions) in these areas (which is a different issue). The split between NOxsaturated or NOx-sensitive regimes is driven by the chemistry of odd hydrogen radicals with HNO3 being the dominant sink in the first case and peroxides the dominant sink in the second case. Maybe the authors could also plot the photochemical regimes in their simulations for the month of July using VOC/NOx or H2O2/NOy ratios (see also the study of Beekmann and Vautard, ACP, 2010).

The reviewer is correct; in this discussion the term "NOx titration behavior" has been replaced with "NOx saturated behavior." Regarding plotting chemical indicators for chemical regime, an additional plot showing the indicator CH2O/NOy has been added to the Supplementary Material (Figure S12); a brief discussion of this figure is now included in Section 4.3. A comparison of our results on NOx vs. VOC sensitivity to the findings of Beekman and Vautard (2010) has also been added to the discussion. The revised discussion is copied below.

“Notably, the U.K., Benelux, northern France and Paris, and northwest Germany show NOx-saturated behavior, in which increased NOx emissions lead to decreased O3 concentrations. NOx-saturated regimes
are also seen around the area of the Mediterranean between Monaco, Genoa and Corsica. An alternate approach to identify areas of NOx-sensitive vs. NOx-saturated regimes is to use indicator ratios (in the base simulation) following Sillman (1995). We have applied this approach with the indicator ratio CH2O/NOy (Figure S12) and find that areas identified as NOx sensitive using the indicator ratio are the same as those identified using the simulation with +30% NOx emissions. These results are also consistent with the areas of Europe found to be NOx saturated in the model study of Beekmann and Vautard (2010). Magnitudes of the observed change in O3 in response to increased NOx emissions are quite similar for both mechanisms, although RADM2 shows slightly stronger NOx saturation (i.e., a stronger decrease in O3 given a 30% increase in NOx emissions) in the area centered around Benelux, and stronger NOx sensitivity over Scandinavia and northwest Russia.”

7) lines 558-559: Mind also that the highest sensitivity for ozone production with regards to VOC is at the regions of high NOx emissions as someone would expect for the regions in the VOC limited regime.

We see that in areas with high NOx emissions such as Benelux, northern France and Germany, and shipping tracks in the Mediterranean, both RADM2 and MOZART predict VOC-sensitive conditions. This point have been added to the discussion in the revised manuscript. However, the increases in O3 with +30% VOC emissions are still relatively small. The text has been updated as follows:

“This areas where MOZART and RADM2 are in agreement in predicting VOC sensitivity (increased O3 concentrations in response to increased VOC emissions) are generally those with high NOx emissions, where one would expect the highest VOC sensitivity based on theory; these areas include Benelux, northern France, northwest Germany, and shipping tracks in the Mediterranean. However, the increase in O3 concentration is modest for both mechanisms; for RADM2 it is generally limited to increases of 2-4% over the base simulation.”

8) lines 565-566: Do you think that the different O3 sensitivity to VOC changes in the two schemes can account for the O3 differences between RADM2 and MOZART (e.g. the lower ozone values in MOZART)? If yes, in which sense?

The results of the +30% VOC sensitivity studies for July (Figure 16) indicate that d[O3]/d[VOC] is higher (more positive) for RADM2 than for MOZART for the chemical regime represented by the models in July 2007. This is an indication that the two mechanisms are simulating different O3 chemical regimes – in the case of RADM2, there is a greater extent of VOC sensitivity, which means that addition of VOC emissions moves the chemistry in the direction of maximum O3 production efficiency, which is not the case for MOZART over much of the domain. A more extensive study would be needed to evaluate whether the conclusion that d[O3]/d[VOC] is higher for RADM2 than for MOZART can be applied more generally. In our simulations, this effect (i.e., more O3 incremental production from VOC in RADM2 than in MOZART) appears to be dominated by other differences between the mechanisms (e.g., the inorganic rate coefficients), given that O3 concentrations predicted by MOZART are always greater than those predicted by RADM2 in our simulations. A discussion of this has been added to Section 4.3:
“The results of the +30% VOC sensitivity studies for July indicate that \( \frac{d[O_3]}{d[VOC]} \) is higher (more positive) for RADM2 than for MOZART for the chemical regime represented by the models in July 2007. This shows that the two mechanisms are simulating different \( O_3 \) chemical regimes – in the case of RADM2, there is greater VOC sensitivity, meaning that addition of VOC emissions moves the chemistry in the direction of maximum \( O_3 \) production efficiency; this is not the case for MOZART over much of the domain. A more extensive study would be needed to evaluate whether the conclusion that \( \frac{d[O_3]}{d[VOC]} \) is higher for RADM2 than for MOZART can be applied more generally.”

9) lines 575-578: This is an interesting result which shows that differences in rate constants can account by 40% for the \( O_3 \) differences between RADM2 and MOZART runs. You may highlight this result a bit more.

This result has been highlighted further in the Abstract and in the Summary and Conclusions. In the Summary and Conclusions section, we further suggest that harmonization of inorganic rate constants could potentially lead to reduced spread in predicted \( O_3 \) among multi-model studies such as AQMEII.

In the abstract, discussion of this difference now reads: "Additionally, we found that differences in reaction rate coefficients for inorganic gas phase chemistry in MOZART-4 vs. RADM2 accounted for a difference of 8 \( \mu g \) m\(^{-3}\), or 40% of the summertime difference in \( O_3 \) predicted by the two mechanisms."

In the Summary and Conclusions, the text has been updated as follows. The first sentence was in the original manuscript, the second sentence has been added.

“Although the most fundamental differences between MOZART-4 and RADM2 (and other chemical mechanisms used in regional modeling) is the representation of VOC oxidation chemistry, we find that approximately 40% of the difference seen in predicted \( O_3 \) seen in this study can be explained by differences in inorganic reaction rate coefficients employed by MOZART-4 and RADM2. This result suggests that harmonization of inorganic rate coefficients among chemical mechanisms used for regional air quality modeling might be valuable, and could potentially lead to a smaller spread in model-predicted \( O_3 \) compared to that seen in, e.g., the multi-model studies of AQMEII (Solazzo et al., 2012b; Im et al., 2015).”

10) lines 591-594: Taking into consideration all three (rate constants, deposition and photolysis schemes) it seems that altogether account about 60% for the \( O_3 \) differences between RADM2 and MOZART runs. Is this correct? You may highlight this conclusion.

It is true that if one looks at the average change in \( O_3 \) concentration in these three sensitivity simulations, then a total of 60% of the MOZART-RADM2 difference in \( O_3 \) concentration is explained, assuming that the effects are additive. However, the authors have consciously avoided presenting this as a conclusion in the text; since the effects of inorganic rate constants, photolysis and deposition are highly interconnected, it is reasonable to assume their combined effects may not be simply additive. We consider a quantification of the nonlinearity of this behavior to be outside the scope of this study.

11) Figure 3: I guess here the authors refer to wind direction. Please also provide information on the approach calculating the wind direction difference between obs and model.
The caption for Figure 3 has been fixed and now correctly refers to wind direction rather than wind speed. A more detailed description of how modeled wind direction was compared to observed wind direction has been added to Section 3.3, and reads as follows.

“When applying these statistics to wind direction, wind direction was treated as a scalar quantity, when in fact it is a vector. This simple approach was favored rather than applying a correction (as done by, e.g., Zhang et al. (2013a) in cases where the difference in modeled vs. observed wind direction were greater than 180°). This is not expected to make an important impact on our analysis, especially since northerly winds (i.e., centered around 0°, or equivalently 360°) are not prevalent in Europe (see Figure 3 and Figure S2 in the Supplementary Material).”

12) Figure 16: Maybe it would be better to show the sensitivity result in a percentage scale (from -10 to 10 %).

In Figure 16, the plot has been adjusted to show the percent difference rather than the fractional difference.

Minor comments line 209: delete double "and". line 239: It is "for" instead of "fo". line 305: Maybe "related" instead of "associated" . line 406: It is "configuration" instead of "configuration".

All of the above minor comments have been addressed with corrections in the text.

lines 427-429: The sentence needs rephrasing. It is not clear.

The sentence now reads "Coates et al. (2016) have shown that adding representation of stagnant conditions (which were not represented in Knote et al. (2015)) to a box model increased the sensitivity of predicted O3 to the chemical mechanism, and also improved model agreement with observations."

We believe this has improved the clarity of the original sentence, which read "Coates et al. (2016) have shown that accounting for stagnant conditions in a box model increased the variability in predicted O3 with temperature in a way that better reproduced the variability seen in observational datasets and 3-D model simulations; adding representation of stagnant conditions (which were not represented in Knote et al. (2015)) to the box model also increased the sensitivity of predicted O3 to the chemical mechanism."
Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: Model evaluation and chemical mechanism comparison

Kathleen A. Mar¹, Narendra Ojha², Andrea Pozzer², and Tim M. Butler¹

¹Institute for Advanced Sustainability Studies, Potsdam, Germany
²Atmospheric Chemistry Department, Max Planck Institute for Chemistry, Mainz, Germany

Correspondence to: K. A. Mar (Kathleen.Mar@iass-potsdam.de)

Abstract. We present an evaluation of the online regional model WRF-Chem over Europe with a focus on ground-level ozone (O₃) and nitrogen oxides (NOₓ). The model performance is evaluated for two chemical mechanisms, MOZART-4 and RADM2, for year-long simulations. Model-predicted surface meteorological variables (e.g., temperature, wind speed and direction) compared well overall with surface-based observations, consistent with other WRF studies. WRF-Chem simulations employing MOZART-4 as well as RADM2 chemistry were found to reproduce the observed spatial variability in surface ozone over Europe. However, the absolute O₃ concentrations predicted by the two chemical mechanisms were found to be quite different, with MOZART-4 predicting O₃ concentrations up to 20 µg m⁻³ greater than RADM2 in summer. Compared to observations, MOZART-4 chemistry overpredicted O₃ concentrations for most of Europe in the summer and fall, with a summertime domain-wide mean bias of +10 µg m⁻³ against observations from the AirBase network. In contrast, RADM2 chemistry generally led to an underestimation of O₃ over the European domain in all seasons. We found that the use of the MOZART-4 mechanism, evaluated here for the first time for a European domain, led to lower absolute biases than RADM2 when compared to ground-based observations. The two mechanisms show relatively similar behavior for NOₓ, with both MOZART-4 and RADM2 resulting in a slight underestimation of NOₓ compared to surface observations. Further investigation into the differences between the two mechanisms revealed that the net midday photochemical production rate of O₃ in summer is higher for MOZART-4 than for RADM2 for most of the domain. The largest differences in O₃ production can be seen over Germany, where net O₃ production in MOZART-4 is seen to be higher than in RADM2 by 1.8 ppb hr⁻¹ (3.6 µg m⁻³ hr⁻¹) or more. We also show that, while the two mechanisms exhibit similar NOₓ-sensitivity, RADM2 is approximately twice as sensitive to increases in anthropogenic VOC emissions as MOZART-4. Ad-
ditionally, we found that differences in reaction rate coefficients for inorganic gas phase chemistry
in MOZART-4 vs. RADM2 accounted for a difference of 8 µg m$^{-3}$, or 40% of the summertime dif-
ference in O$_3$ predicted by the two mechanisms. Differences in deposition and photolysis schemes
explained smaller differences in O$_3$. Our results highlight the strong dependence of modeled surface
O$_3$ over Europe on the choice of gas phase chemical mechanism, which we discuss in the context
of overall uncertainties in prediction of ground-level O$_3$ and its associated health impacts (via the
health-related metrics MDA8 and SOMO35).
Tropospheric ozone (O$_3$) is an air pollutant with adverse effects on human and ecosystem health as well as a short-lived climate forcer with a significant warming effect (e.g., Monks et al., 2015; Stevenson et al., 2013; WHO, 2003). In Europe, ozone pollution remains a problem: the European Environmental Agency reports that between 2010 and 2012, 98% of Europe’s urban population was exposed to O$_3$ levels in exceedance of the WHO air quality guideline (EEA, 2014), leading to more than 6000 premature deaths annually (Lelieveld et al., 2015). This is despite the fact that European emissions of ozone precursors, in particular nitrogen oxides (NO$_x$) and volatile organic compounds (VOCs), have decreased significantly since 1990. The persistence of unhealthy levels of ozone in Europe can be attributed to increases in hemispheric background ozone (Wilson et al., 2012) as well as the non-linear relationship between O$_3$ and levels of precursor species NO$_x$ and VOC (EEA, 2014).

Air quality models are employed to understand the drivers of air pollution at a regional scale and to evaluate the roles of and interactions between emissions, meteorology and chemistry. These models fall into two broad categories: offline Chemistry-Transport Models (CTMs), in which meteorology is calculated separately from model chemistry, and "online" models, the category to which WRF-Chem belongs, in which the meteorology and chemistry are coupled, meaning they are solved together in a physically consistent manner (e.g., Zhang, 2008). The meteorology and chemistry components in WRF-Chem use the same horizontal and vertical grids and same timestep, eliminating the need for temporal interpolation (e.g., Grell et al., 2004, 2005).

Air quality modeling studies over the European region have predominantly utilized CTMs, examples of which include EMEP (Simpson et al., 2012), CHIMERE (Terrenoire et al., 2015), and LOTOS-EUROS (Schaap et al., 2008). The application of online coupled regional meteorology-chemistry models in Europe, among them WRF-Chem, has been recently reviewed by Baklanov et al. (2014). The use of WRF-Chem over Europe has increased in recent years (e.g., Forkel et al., 2012; Žabkar et al., 2015; Solazzo et al., 2012a, b; Tuccella et al., 2012; Zhang et al., 2013a, b). However, only a limited number of these studies are dedicated to the evaluation of WRF-Chem-simulated meteorology and chemistry over the whole European domain. The study of Tuccella et al. (2012) evaluated the performance of WRF-Chem using the RADM2 chemical mechanism by comparing domain-wide average values against observations of meteorology and chemistry. However, an evaluation of the spatial distribution of model-simulated meteorology and trace gases is missing. This type of spatial information is extremely pertinent for air quality management applications, where model performance at a national scale can become more relevant than performance metrics applied to the whole of Europe; this information gets lost when only comparing quantities that have been averaged over the entire domain. Additionally, Tuccella et al. (2012) utilized time-invariant chemical boundary conditions, which the authors suggested misrepresented the seasonal changes in the intercontinental transport (Tuccella et al., 2012). The importance of temporally varying chemical
boundary conditions in air quality modeling has also been stressed in other studies (including Akridis et al., 2013; Andersson et al., 2015). In addition to the study of Tuccella et al. (2012), Zhang et al. (2013b) evaluated the performance WRF-Chem-MADRID (Zhang et al., 2010), an unofficial version of WRF-Chem coupled to the Model of Aerosol Dynamics, Reaction, Ionization, and Dissolution (MADRID), over Europe for the month of July 2001, employing the gas-phase mechanism CB05 (Yarwood et al., 2005). This detailed study provides a valuable reference for comparison to the present work, but their simulations are only for one month, rather than the complete seasonal cycle.

Several groups contributed WRF-Chem simulations to the AQMEII project (phase 1 and phase 2) for the European domain (Solazzo et al., 2012b; Im et al., 2015). In AQMEII phase 1, two different WRF-Chem simulations were part of the model ensemble for Europe, but evaluation of model performance for ozone focused on evaluation of the ensemble (Solazzo et al., 2012b), rather than on individual members. In fact, in the analysis of Solazzo et al. (2012b), individual models were anonymized, meaning the performance statistics for the WRF-Chem ensemble members are not explicitly presented. The evaluation of model performance with respect to ozone in AQMEII phase 2 (Im et al., 2015) provides more information on the model performance of the contributing WRF-Chem ensemble members for the European domain. In AQMEII phase 2, seven different WRF-Chem runs were part of the ensemble. Of these seven simulations, four of them used the gas phase chemical mechanism RADM2 (Stockwell et al., 1990), two used the mechanism CBMZ (Zaveri and Peters, 1999), and one used the mechanism RACM (Stockwell et al., 1997; Geiger et al., 2003). All WRF-Chem simulations for Europe in AQMEII phase 2 tended to underestimate ozone concentrations, with annual average normalized mean bias ranging from -1.6 to -15.8 %, depending on the ensemble member.

The purpose of the present study is to perform a detailed evaluation of meteorology and gas phase chemistry simulated by WRF-Chem, including the spatial and seasonal variations over a full year seasonal cycle using time-varying chemical boundary conditions. This evaluation is performed for two different gas phase chemical mechanisms within WRF-Chem, MOZART-4 (Emmons et al., 2010) and RADM2 (Stockwell et al., 1990). As discussed above, the RADM2 mechanism has been popularly used in WRF-Chem for simulation over Europe (Tuccella et al., 2012; Im et al., 2015). The MOZART-4 chemical mechanism has been widely used with WRF-Chem for regional air quality applications outside of Europe (e.g., Pfister et al., 2013; Im et al., 2015). To the authors’ knowledge, however, WRF-Chem with MOZART-4 has not yet been applied and evaluated over a European domain.

The simultaneous evaluation of WRF-Chem with two different chemical mechanisms further allows us to evaluate the sensitivity of \( \text{O}_3 \) and \( \text{NO}_x \) to the choice of chemical mechanism in a setup where the differences in model physics and other parameters are minimized. This is in contrast to the study of Im et al. (2015), where the various WRF-Chem ensemble members also used different
schemes for model physics. Coates and Butler (2015) recently investigated the sensitivity of the production of odd oxygen \((O_x, a proxy for production of O_3)\) to the choice of chemical mechanism using a box model, and found that choice of chemical mechanism led to differences in \(O_3\) concentrations on the order of 10 ppb under idealized conditions, although differences between the MOZART-4 and RADM2 chemical mechanisms tended to be closer to 5 ppb. In another box model study, Knote et al. (2015) investigated the sensitivity of \(O_3\), \(NO_x\), and other radicals to the different gas-phase chemical mechanisms used in the models that contributed to the AQMEII phase-2 intercomparison project. Knote et al. (2015) found that the choice of chemical mechanism is responsible for a 5% uncertainty in predicted \(O_3\) concentrations and a 25% uncertainty in predicted \(NO_x\) concentrations.

The present study builds on the work of Coates and Butler (2015) and Knote et al. (2015) by comparing two chemical mechanisms within an online coupled regional air quality model. The use of WRF-Chem provides an advantage in that it is compatible with multiple different chemical mechanisms, allowing us to test the effect of different chemistry with minimal confounding factors due to differences in model physics, etc. Furthermore, the use of an online regional model rather than a box model allows us to examine the sensitivity of model-predicted concentrations to the choice of chemical mechanism under more realistic conditions, in which variations in meteorology and dynamics are fully included. Parameters such as radiation are allowed to vary realistically, and different chemical regimes (\(NO_x\)- vs. VOC-limited) are present (e.g., in different seasons and in different parts of the model domain).

Chemical mechanism comparisons have also been undertaken previously using 3-D regional air quality models, though the majority have focused on comparing the SAPRC-99 mechanism (Carter 1990) with versions of the Carbon Bond mechanism (Gery et al. 1989) over a U.S. domain (Luecken et al. 2008; Faraji et al. 2008; Yarwood et al. 2003; Zhang et al. 2012). Two additional studies have compared versions of the RACM mechanism with RADM2 (Mallet and Sportisse 2006) and CB05 (Kim et al. 2010) using the model Polyphemus (Mallet et al. 2007) for a European domain. Typically, these studies found that simulations using two different chemical mechanisms led to differences in \(O_3\) on the order of 5-10 ppb (Luecken et al. 2008; Zhang et al. 2012; Mallet and Sportisse 2006; Kim et al. 2010), although extreme differences of 30-40 ppb were observed between SAPRC-99 and CB-IV mechanisms when simulating high ozone episodes (Faraji et al. 2008; Yarwood et al. 2003).

In this paper, the model configuration, including emissions and initial and boundary conditions, is described in Section 2. A description of observational datasets for meteorology and chemistry and the evaluation methodology is provided in Section 3. Results for the model evaluation and intercomparison of two chemical are presented in Section 4 followed by a summary and concluding remarks in Section 5.
2 Model Description and Setup

2.1 WRF-Chem

This study utilizes the Weather Research and Forecasting with Chemistry (WRF-Chem) model version 3.5.1. WRF-Chem has been developed collaboratively by NOAA, DOE/PNNL, NCAR and other research institutes. We defined our simulation domain on the Lambert projection. The model domain is centered at 15° E, 52° N, and covers nearly the entire European region. The horizontal resolution is chosen to be 45 km × 45 km. The model domain has 115 and 100 grid points in the west-east and south-north directions respectively.

We have used 35 vertical levels in the model starting from surface to 10 hPa. The lowest model level corresponds to an approximate altitude of 50 m above the surface. Tests have shown that surface layer concentrations in this configuration are effectively the same as when the lowest model level is at a height of 14 m, but with no urban surface physics scheme (the urban physics scheme is incompatible with a 14-m model level). Geographical data including terrain height, soil properties, albedo, etc. are interpolated primarily from USGS (United States Geological Survey data) at 30 sec resolution. The land use classification has been interpolated from the CORINE data at 250 m resolution, which was then mapped to the USGS land use classes used by WRF (see Kuik et al., 2016).

Model simulations are conducted for the period of 23 December 2006 to 31 December 2007. The first week of output was treated as model spin up and has been discarded. The instantaneous model output, stored every hour, has been used for the analysis. The different options used in this study to parametrize the atmospheric processes are listed in Table 1. A namelist is available in the Supplementary Material.

The initial and lateral boundary conditions for the meteorological fields were provided from the ERA-interim reanalysis dataset available from ECMWF. This data is available every 6 hours with a spatial resolution of approximately 80 km (T255 spectral). In order to limit the errors in the WRF simulated meteorology the Four Dimensional Data Assimilation (FDDA) has been applied. In the FDDA, temperature is nudged at all the vertical levels with a nudging coefficient of 0.0003. The horizontal winds are nudged at all the vertical levels, except within the PBL, with the nudging coefficient of 0.0003. Sensitivity studies performed showed that nudging of water vapor highly suppressed the precipitation over Europe in a manner inconsistent with observations. As such, water vapor is not nudged in our simulations. This also follows the approach of, e.g., Miguez-Macho et al. (2004) and Stegehuis et al. (2014). The nudging coefficients for temperature and winds have been chosen following previous studies (Stauffer et al., 1991; Liu et al., 2012). The time step for the simulations has been set at 180 s.
Initial and boundary conditions for chemical fields in WRF-Chem are used from the MOZART-4/GEOS5 simulations (http://www.acd.ucar.edu/wrf-chem/mozart.shtml), with a horizontal resolution of 1.9° × 2.5° and 56 pressure levels. MOZART-4/GEOS-5 simulations use meteorology from the NASA GMAO GEOS-5 model and emissions based on ARCTAS inventory (http://www.cgrer.uiowa.edu/arctas/emission.html).

### 2.2 Emissions

Anthropogenic emissions of CO, NO\textsubscript{x}, SO\textsubscript{2}, NMVOCs, PM\textsubscript{10}, PM\textsubscript{25}, and NH\textsubscript{3} are used from the TNO-MACC II emission inventory for Europe (Kuenen et al., 2014), for the year 2007. These emissions are provided as yearly totals by source sector on a high-resolution (7 km × 7 km) grid. The TNO-MACC II emission inventory is based on emissions reported by member countries to the European Monitoring and Evaluation Program (EMEP), which are then further refined to fill gaps and correct errors and obvious inconsistencies. Emissions are temporally disaggregated based on seasonal, weekly and diurnal cycles provided by Denier van der Gon et al. (2011); Schaap et al. (2005). These temporal profiles vary by source sector according to the SNAP (Selected Nomenclature for Sources of Air Pollution) convention. NMVOC emissions are split into modeled NMVOC species (e.g., ethane, aldehydes) based on von Schneidemesser et al. (2016). NO\textsubscript{x} is emitted as 90% NO and 10% NO\textsubscript{2} by mole. Emissions are distributed into the first seven model vertical layers (the surface and the first 6 model layers above the surface) based on sectoral averages from Bieser et al. (2011), although model runs showed little sensitivity to the distribution of emissions above the surface layer.

The model domain used in this study is larger than the European domain used in the TNO-MACC II inventory (Kuenen et al., 2014). Emissions at our domain edges were filled using the Hemispheric Transport of Air Pollution (HTAP v2.2) emission inventory for the year 2008 (http://edgar.jrc.ec.europa.eu/htap_v2/index.php). The HTAP v2 data, described in detail by Janssens-Maenhout et al. (2015), is harmonized at a spatial resolution of 0.1° x 0.1° and available with monthly time resolution. In our model simulations, no additional weekly or diurnal profiles were applied to the HTAP v2 emissions. Furthermore, all emissions from HTAP were emitted into the surface model layer. Because HTAP emissions were only used at the grid "edge," the differences in temporal and vertical resolution of emissions used for HTAP is not expected to have a significant impact on model results. An example of emissions processed for model input is shown Figure S1 in the Supplementary Material.

Biomass burning emissions are from the Fire Inventory from NCAR (FINN), Version 1 (Wiedinmyer et al., 2011). To avoid the double counting of emissions from agricultural burning (i.e., assuming that the FINN product captures large-scale agricultural burning), emissions of the combustion species CO, NO\textsubscript{x}, and SO\textsubscript{2} from SNAP category 10 (Agriculture) in the TNO-MACC II inventory were not included in model simulations, at the suggestion of H.A.C. van der Gon (personal commu-
namic, 2015). Biogenic Emissions are calculated online based on weather and land use data using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2006).

2.3 Model Chemistry

The two year-long WRF-Chem simulations performed for this study are summarized in Table 2. In the MOZART simulation, gas phase chemistry is represented by the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4) mechanism (Emmons et al., 2010). Tropospheric chemistry is represented by 81 chemical species, which participate in 38 photolysis and 159 gas-phase reactions. The MOZART-4 mechanism includes explicit representation of the NMVOCs ethane, propane, ethene, propene, methanol, isoprene, and α-pinene. Other NMVOC species are represented by lumped species based on the reactive functional groups. In the WRFV3.5.1 code, two bug fixes have been included for the MOZART-4 mechanism: the \( \text{NH}_3+\text{OH} \) rate coefficient has been corrected following Knote et al. (2015), and a correction has been made to treatment of the vertical mixing of MOZART-4 species (A.K. Peterson, personal communication). In the WRF-Chem simulations, we use the version of MOZART-4 coupled to the simple GOCART aerosols mechanism (Ackermann et al., 1998b), known as the MOZCART mechanism. In this paper, we limit our analysis to gas-phase species. Because of this focus, and to simplify the interpretation the mechanism intercomparison (see below), all aerosol radiative feedbacks (i.e., both direct and indirect effects) are turned off in all model simulations in this study.

In the RADM2 simulation, gas phase chemistry is represented by the second generation Regional Acid deposition Model (RADM2) (Stockwell et al., 1990). This mechanism has 63 chemical species which participate in 21 photolysis and 136 gas phase reactions. The NMVOC oxidation in RADM2 is treated in a less-explicit fashion than in MOZART, in which ethane, ethene and isoprene are the only species treated explicitly and all other NMVOCs are assigned to lumped species based on OH reactivity and molecular weight. In WRF-Chem, RADM2 is coupled to the MADE/SORGAM aerosol module, which is based on the Modal Aerosol Dynamics Model for Europe (MADE) (Binkowski and Shankar, 1995; Ackermann et al., 1998a) and Secondary Organic Aerosol Model (SORGAM) (Schell et al., 2001). However, as noted above, in this study we focus our analysis on gas-phase chemistry.

In both the RADM2 and MOZART simulations, the chemical mechanism code was generated with the Kinetic Pre-Processor (KPP) (Damian et al., 2002; Sandu and Sander, 2006), and equations are solved using a Rosenbrock-type solver. Note that when using RADM2 chemistry, there are two different solvers available within WRF-Chem. We chose to use the KPP chemistry and Rosenbrock solver to be consistent with the MOZART runs, and also because the alternative QSSA chemistry solver has been shown to have problems representing \( \text{NO}_x \) titration (Forkel et al., 2015). In particular, the QSSA treatment of RADM2 chemistry was found to result in an under-representation of nocturnal ozone titration for areas with high NO emissions.
Observational datasets

A summary of the observational datasets used for model evaluation can be found in Table 3.

3.1 Meteorology

Since WRF-Chem couples the meteorology simulations online with the chemistry, we begin by evaluating the modeled meteorological fields against observations which are driving the simulations of chemical fields. In this study, the WRF-Chem simulated meteorological fields are evaluated against the in situ measurements of mean sea level pressure (MSLP), 2-meter temperature (T2) and 10-meter wind speed and direction (WS10 and WD10, respectively) from the Global Weather Observation dataset provided by the British Atmospheric Data Center (BADC). We chose these meteorological variables for the evaluation as these are expected to have the most significant influence on the gas-phase chemistry, which is the main focus of this study.

3.2 Chemistry

3.2.1 EMEP Network

The EMEP observational dataset provides surface measurements of pollutant concentrations, including tropospheric ozone and its precursors, at stations chosen to be representative of regional background pollution (see, e.g., [Tørseth et al., 2012]). The regional focus is in keeping with the goals of the Convention on Long-range Transboundary Air Pollution (CLRTAP), under which this network is administrated.

3.2.2 AirBase Network

AirBase is the public air quality database of the European Environmental Agency (EEA), and represents a much denser network of monitoring than the EMEP network (http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-7). Because of the relatively coarse horizontal resolution in this model study, model output is only compared against AirBase stations that are classified as "rural background." The station classification was taken from the metadata provided by the EEA for AirBase. Some AirBase stations are also part of the EMEP network; the subset of AirBase stations used in this study exclude any stations that are also part of the EMEP network (since they are already included in the evaluation against EMEP observations).

3.3 Evaluation methodology

Stations were excluded from our season-by-season analysis if the temporal coverage was less than 75%, i.e., if missing or flagged hourly (or 3-hourly) data represented more than 25% of the hourly (or 3-hourly) time series over the entire season. For sensitivity studies that consider the month of
July only, stations were considered that had at least 75% temporal coverage for the month. This criteria was applied for all meteorological and chemistry observations. For comparison of model output to in situ observations, the model gridcell that is closest to the latitude, longitude location of the measurement station was chosen. Statistics calculated include the mean, mean bias (MB), normalized mean bias (NMB), mean fractional bias (MFB) and the temporal correlation coefficient (r). The domain-wide statistics presented in Tables 4 - 9 were calculating by first calculating the statistical quantity hour-by-hour at each station, and then averaging these values over all times (in the season) and all stations. Definitions of calculated statistical quantities can be found in Appendix B.

When applying these statistics to wind direction, wind direction was treated as a scalar quantity, when in fact it is a vector. This simple approach was favored rather than applying a correction (as done by, e.g., Zhang et al. (2013a) in cases where the difference in modeled vs. observed wind direction were greater than 180°). This is not expected to make an important impact on our analysis, especially since northerly winds (i.e., centered around 0°, or equivalently 360°) are not prevalent in Europe (see Figure 3 and Figure S2 in the Supplementary Material).

From hourly concentrations of $O_3$, both observed and modeled, additional ozone metrics for health impacts are calculated. MDA8 is defined as the maximum daily 8-hour mean ozone, in accordance with the European Union’s Air Quality Directive. Note that, for calculation of MDA8, a missing value was assigned if one or more hours of data in the 8-hour average were missing. SOMO35 is an indicator of cumulative annual ozone exposure used in health impact assessments. The accumulated health impact is assumed to be proportional to the sum of concentrations above a cutoff of 35 ppb, chosen because the relationship between $O_3$ and adverse effects is very uncertain below this threshold [WHO, 2013]. Mathematically, SOMO35 is defined as the sum of MDA8 levels over 35 ppb (70 µg m$^{-3}$) over a year, in units of concentration-days, following Amann et al. (2008).

$$SOMO35 = \frac{365}{N_{\text{valid}}} \sum_{\text{iday}} \max(0, C_{\text{iday}} - 70 \, \mu g \, m^{-3})$$

where $N_{\text{valid}}$ is the number of valid (i.e., non-missing) daily values.

4 Results and Discussion

4.1 Evaluation of Meteorology

Table 4 shows a summary of domain-wide statistics evaluating the MOZART model simulation against observations of meteorological variables MSLP, T2, WS10 and WD10; the spatial distribution of these statistics shown in Figures 4-5 for temperature and wind variables. Differences in predicted meteorology between the MOZART and RADM2 simulations are small, with differences in MSLP less than one hundredth of 1%, and differences in T2, WS10, and WD10 generally far
below 1%. Since the simulations were run without aerosol-radiative feedbacks, it was expected that
the two simulations would show minimal differences in meteorology, and we conclude that differ-
ences in O\textsubscript{3} and NO\textsubscript{x} predicted in the MOZART and RADM2 simulations (Section 4.3) are a direct
result of differences in the chemistry, rather than chemistry-radiative feedbacks. Statistics for me-
eteorology for the RADM2 simulation can be found in the Supplementary Material, Table S1 and
Figures S4-S7.

MSLP has been reproduced over the entire European domain with a high degree of skill in every
season for both simulations, with negligible bias (domain-averaged NMB and MFB are zero in all
seasons) and temporal correlation coefficients (r values) of 0.98 or greater (see also Figures S3 and
S7 in the Supplementary Material).

The spatial distribution of seasonal average T2 in the model and observations is shown in Figure 1,
along with the spatial variation in mean bias and temporal (3-hourly) correlation. Overall, the spatial
variability in measured T2 is found to be well-reproduced by WRF-Chem during all the seasons. The
absolute values of mean biases in T2 were generally found to be lower than 1° C. Larger biases in T2
can be found in the Alps, in particular during winter, where T2 is often overpredicted by more than
1° C (Figure 1). This larger bias over mountainous regions, also found in a previous study (Zhang
et al., 2013a), is likely due to the complex mountain terrain and related unresolved local dynamics.
The r values are generally found to be more than 0.9 in all the seasons and show no significant
dependence in near
geographical variation, indicating that the model is able to reproduce the hourly variations in near
surface temperature. Averaged over the entire domain, the mean bias in T2 varies from -0.4 to +
0.3° C depending on the season (Table 4).

The spatial variability in wind speeds, including the seasonality, with strongest winds during the
winter, have been reproduced by the model (Figure 2). However, the model tends to overestimate
winds speeds with larger biases (2 m/s or more) during the winter and fall. The regions showing
greater bias in wind speed include the Alps, coastal regions, and the low-lying areas of northern
Germany and Denmark (Figure 2). The temporal correlation of wind speed is generally above 0.7 in
the northern half of the domain, but is lower (0.4-0.6) in the southern part of the domain, in areas
in the Alps and close to the Mediterranean (Figure 2). Similar behavior for modeled wind speed is
reported by Zhang et al. (2013a), who attributes the overestimation in wind speeds primarily to poor
representation of surface drag exerted by unresolved topographical features, which results in model
limitations in simulating circulation systems such as sea breeze and bay breeze. An overview of the
statistics for wind direction is presented in Table 4, with the spatial distribution shown in Figure 3.

Wind direction over the continent is predominantly from the west and south, and the mean bias in
wind direction is between 20 and 30 degrees depending on the season. Similar to the patterns seen
for wind speed, areas with complex topography (the Alps, the Balkans, the Mediterranean coast)
show the largest biases and the lowest correlations for wind direction.
Overall, we find that WRF-Chem is capable of reproducing the spatial and temporal variations in the European meteorological conditions reasonably well, in a manner consistent with previous studies (e.g., [Zhang et al., 2013a]).

4.2 Evaluation of Chemistry

4.2.1 Ozone

We begin the evaluation of chemistry by examining the seasonal average surface $O_3$ distribution over Europe from the MOZART simulation, as shown in Figure 4. Predicted surface $O_3$ distributions show a clear seasonality, with maximum concentrations during summer. In all seasons, surface $O_3$ concentrations are highest over the Mediterranean region, with values during the spring and summer greater than 110 $\mu g \text{m}^{-3}$. Simulated concentrations reproduce the north-south gradient in $O_3$ seen in the ground-based observations. Figure 5 provides another comparison of seasonal average $O_3$ distributions in the model vs. the observations (from both the AirBase and EMEP networks) and additionally shows the spatial distribution of MB and $r$, the temporal (hourly) correlation coefficient; performance statistics are shown in Table 5 (against observations from the AirBase network) and Table 6 (against observations from the EMEP network). MOZART overpredicts $O_3$ concentrations for most of Europe in the summer and fall. In winter and spring, MOZART tends to underestimate $O_3$ in north-central Europe, but overestimate $O_3$ in southern Europe. Hourly correlation coefficients for $O_3$ are highest (greater than 0.6) in northern Europe (especially France, Germany, and the Benelux region) and in Spain, but are lower (with values of approximately 0.4) throughout Italy and the mountainous regions of the Alps. Notably, Italy and the Alps are the regions within our domain that exhibit the highest biases and lowest correlations with respect to wind direction and speed (Section 4.1), which could explain the poorer temporal correlation for $O_3$ in these areas.

Looking at Tables 5 and 6, we see some differences in the statistical performance of the MOZART simulation when compared to the EMEP vs. the AirBase observational datasets. Considering the EMEP observations over the whole domain (Table 6), MOZART slightly overpredicts $O_3$ in summer, with a summertime mean bias of 4 $\mu g \text{m}^{-3}$, whereas the summertime mean bias when compared the AirBase network is 10 $\mu g \text{m}^{-3}$ (Table 5). In winter and spring, the bias (MB, NMB, and MFB) in MOZART-predicted $O_3$ is more negative when compared to EMEP observations than to AirBase observations. In fall, the sign of the domain-average bias changes if considering the model performance against EMEP vs. AirBase observations. These differences likely reflect differences in the character of the two observational networks. First, we expect that the Airbase rural background sites considered here may be, on average, more influenced by local pollution sources than the EMEP sites, which are selected to be representative of more remote regional background. Secondly, the geographical coverage of AirBase vs. EMEP sites for $O_3$ is slightly different (Figure S8). In particular, coverage of the U.K. and the Nordic countries is almost exclusively via the EMEP network, poten-
tially giving the EMEP observations a northern bias in comparison to the AirBase-only sites. Both features of the measurement networks could explain the lower values of the domain-wide average O$_3$ observed at the EMEP vs. the AirBase stations. In addition to evaluating the model’s ability to simulate hourly O$_3$ concentrations, we also consider MDA8 and SOMO35, two metrics designed to evaluate the impact of ozone on health. The distribution of seasonal average values of MDA8 is shown in Figure 6 for the MOZART simulation. The European Union’s Air Quality Directive states that, as a long-term objective, MDA8 should not exceed the threshold value of 120 µg m$^{-3}$; as a target value this long-term objective should not be exceeded on more than 25 days per year, averaged over 3 years. Figure 6 shows that, at some stations in the Alps and in southern Italy during summer, the average value of MDA8 exceeds 120 µg m$^{-3}$. As seen in Figure 7, the number of days when MDA8 exceeds the 120 µg m$^{-3}$ is greater than 25 in spring alone for much of southern Europe, which is also captured well by the MOZART simulation. MOZART tends to overpredict MDA8 and the days in exceedance of the target value in summer and fall, consistent with the overestimation of hourly average O$_3$ during this season. Since the metric MDA8 is, in effect, a measure of daytime ozone, it is always higher than the straight average of hourly concentrations. As a consequence, MOZART shows greater bias in MDA8 than in average O$_3$ in seasons where average O$_3$ is already overpredicted (Tables 5 and 6). In general, regional and seasonal patterns for MDA8 simulated by MOZART are similar to those for simulated average O$_3$. SOMO35, an indicator for cumulative annual exposure, is shown in Figure 8 for the year 2007. MOZART is able to reproduce the north-south gradient of SOMO35 seen in the observations quite well, while overpredicting the magnitude of SOMO35 by 2 mg m$^{-3}$·days (Table 7). WRF-Chem simulations using the RADM2 chemical mechanism show a spatial and seasonal distribution of surface O$_3$ over Europe (Figures 9 and 10) that is qualitatively similar to that for MOZART. The correlation coefficients for the MOZART and RADM2 simulations are also similar in both magnitude in distribution (Figures 5 and 10). Absolute O$_3$ concentrations are most similar (i.e., less than 5% different) between the mechanisms near the northwest edges of the domain (see Figures 4 and 9), where the prevailing westerly winds (Supplementary Material, Figure S2) mean that O$_3$ imported from the boundary conditions plays a dominant role. However, it is striking to note that the surface O$_3$ concentrations predicted by two different chemical mechanisms are generally quite different, with RADM2 predicting average surface O$_3$ values that are approximately 20 µg m$^{-3}$ lower than those predicted by MOZART in spring and summer (c.f. Figures 4 and 9, Tables 5 and 8 and Tables 6 and 9). In contrast to MOZART, RADM2 underpredicts O$_3$ throughout most of Europe in all seasons. An exception to this is in southern Europe in winter, where RADM2, like MOZART, shows some overprediction of O$_3$ concentrations in southern Europe, particularly near the Mediterranean. RADM2 also overpredicts O$_3$ near the Mediterranean in fall (a season where MOZART overpredicts O$_3$ Europe-wide). The general underprediction of O$_3$ concentrations in RADM2 means that the health metrics MDA8 and SOMO35 are also underpredicted.
Overall, absolute biases (i.e., the absolute value of MB, NMB, and MFB) are smaller for MOZART than for RADM2, indicating that MOZART is more successful overall in reproducing European ground-level O₃.

Model biases for O₃ in both the MOZART and RADM2 simulations are in line with biases found in other regional modeling studies for Europe. For instance, values for the NMB in European summertime O₃ ranged from less than -20% to greater than +20% depending on the ensemble member in AQMEII (Solazzo et al., 2012b; Im et al., 2015), compared to values of -18% and +14% for the RADM2 and MOZART simulations, respectively, in the present study. Zhang et al. (2013b) found domain-wide values for NMB for O₃ ranging from +4.2% to +19.1% for the month of July 2001, depending on their model configuration. Tuccella et al. (2012) report a domain-average mean bias in O₃ of -1.4 µg m⁻³ averaged over the whole year. Although the work of Tuccella et al. (2012) uses the RADM2 chemical mechanism and simulates the year 2007, similar to the RADM2 simulation in the present study, there are several differences in model configuration that could explain the observed differences in predicted O₃, including the use of time-invariant chemical boundary conditions, the use of the QSSA rather than the Rosenbrock chemical solver (which has been shown to make a difference Forkel et al. (see 2015)), and the use of an alternate emissions inventory (from EMEP).

The temporal correlation with hourly measurements for O₃ in this study are also in line with other regional modeling studies of O₃ for Europe. Simulations with both chemical mechanisms lead to reasonable correlations between the model-predicted and observed O₃ concentrations over the entire domain, with r values generally in the range of 0.6-0.8 (Figures 5 and 10, Tables 5 and 8). This is consistent with the hourly correlation coefficient for O₃ of 0.62 reported by Tuccella et al. (2012), where their r value represents an average over the entire year of 2007. Zhang et al. (2013b) also report correlation coefficients of 0.6-0.7 for hourly O₃ over the European domain (horizontal resolution 0.5°) using the CB05 gas-phase chemical mechanism in WRF-Chem.

In addition to evaluating the performance of the MOZART and RADM2 simulations on their ability to reproduce ground-level ozone concentrations, we compare the observed sensitivity of modeled O₃ to the choice of chemical mechanism to other studies that have investigated the uncertainty in 3-D model predictions associated with the choice of chemical mechanism. Knote et al. (2015) used box model simulations based on AQMEII phase 2, and concluded that the uncertainty in predicted O₃ in a 3-D model solely due to the choice of gas phase chemical mechanism should be of the order of 5%, or 4 ppbv (8 µg m⁻³). This is quite a bit smaller than the sensitivity to chemical mechanism found in this study, where we see differences in summertime average O₃ of 20 µg m⁻³, corresponding to a relative difference of approximately 40%. Coates et al. (2016) have shown that adding representation of stagnant conditions (which were not represented in Knote et al. (2015)) to a box model increased the sensitivity of predicted O₃ to the chemical mechanism, and also improved model agreement with observations. This result suggests that day-to-day variability in meteorological conditions and
Another interesting basis for comparison is the study of Mallet and Sportisse (2006), who investigate uncertainty in the CTM Polyphemus due to various physical parameterizations, including chemical mechanism (comparing RACM and RADM2), using an ensemble approach. They estimated an overall uncertainty in O$_3$ concentrations of 17% based on choices for physical parameterizations in general, but identified the choice of chemical mechanism along with the turbulent closure parameterization as the two most important drivers of this uncertainty. Simulations using the RACM vs. RADM2 mechanisms yielded differences in average O$_3$ concentrations of 7-13 µg m$^{-3}$, depending on the other parameterizations used. It is clear that the sensitivity of O$_3$ to the use of the MOZART vs. RADM2 chemical mechanism in this study is large compared to other studies of mechanism comparisons in 3-D models (see also Luecken et al., 2008; Kim et al., 2010), though even larger absolute differences in hourly O$_3$ concentrations (up to 40 ppb, or 80 µg m$^{-3}$) have been found in studies of episodic ozone (Faraji et al., 2008; Yarwood et al., 2003). It is possible that MOZART and RADM2 as implemented in this study are examples of chemical mechanisms that are extremely different from one another on a spectrum of other commonly-used mechanisms; the differences between the two mechanisms will be further explored in Section 4.3.

4.2.2 Nitrogen oxides

Seasonal average surface-level NO$_x$ for the MOZART simulation are shown in Figure 11. Several hotspots in the spatial distribution of NO$_x$ mixing ratios are apparent, as expected based on the intensity of emissions in these areas. NO$_x$ hotspots with concentrations of more than 30 µg m$^{-3}$ are visible over parts of France, Belgium, Germany and Russia. Similar high concentrations are also seen over the marine regions close to Barcelona, Monaco, and southern France. As shown in Table 5 the MOZART simulation slightly underpredicts domain-average NO$_x$ concentrations for all seasons when comparing to AirBase observations. In Figures 12 and 13 we examine the spatial distribution of NO$_x$ broken down into its components, NO$_2$ and NO, together with the spatial distribution of MB and r. The MOZART simulation overestimates NO$_2$ in the U.K., northern France, Belgium, and central Germany, all of which are regions known for having high NO$_x$ emissions and concentrations. However this does not hold true for the Netherlands, a neighboring region with high emissions where MOZART tends to underpredict rather than overpredict NO$_2$ concentrations. NO, on the other hand, is significantly underpredicted compared to surface measurements throughout the domain. This may be partially due to the relatively coarse horizontal resolution of the model, in which fresh NO emissions are immediately diluted over a large area, and could also be a consequence of model deficiencies in representing NO$_x$ chemical cycles. Artifacts related to reporting of low NO concentrations approaching measurement detection limits could also play a role (observed time
series for NO typically show a baseline of 1-2 µg m\(^{-3}\), whereas modeled concentrations reach a baseline of zero).

Domain average temporal correlation coefficients (r) against hourly measurements of NO\(_x\), NO\(_2\), and NO (Tables 5 and 6) range from approximately 0.2 to 0.5, which is lower than correlations for O\(_3\) but consistent with other studies, discussed further below. In all seasons, the domain-averaged temporal correlation coefficient is higher when compared to EMEP vs. AirBase observations. This is attributed to lesser local influences and therefore better regional representativeness of the EMEP stations. No exceptional patterns are seen in the spatial distribution of r for NO\(_2\) or NO, although correlation appears slightly better in the northern part of the domain. The MOZART simulation shows the highest domain-average correlation coefficients (r) for NO\(_x\), NO\(_2\), and NO in winter and fall, and the lowest domain-average r values in summer.

NO\(_x\) predicted by the RADM2 simulation shows fairly similar behavior to NO\(_x\) predicted by the MOZART simulation (cf. Figures 12 and 14 and additional Figures S10-S11 in the Supplementary Material). In general, simulated NO\(_x\) concentrations are slightly higher for MOZART than for RADM2. Domain-wide average NO\(_x\) concentrations predicted by MOZART are approximately 2 µg m\(^{-3}\) higher than for RADM2 in all seasons except winter, where the difference is approximately 3 µg m\(^{-3}\) (cf. Tables 5 and 8). The spatial distribution of MB for NO\(_2\) for the RADM2 simulation generally shows the same patterns as observed for the MOZART simulation, namely a slight overestimation in the U.K., northern France, Belgium, and central Germany. Like for MOZART, NO for RADM2 is underpredicted throughout the domain, with NO concentrations slightly more negatively biased than in MOZART in all seasons except Fall, when NO concentrations are higher for RADM2 than for MOZART and show better agreement with the observations. Temporal correlation for NO\(_2\) and NO in RADM2 is also found to show similar behavior to the MOZART simulation.

An exception to the similarity observed between the mechanisms for NO\(_x\) can be seen over central Germany in winter, where MB values for NO\(_2\) are 6-10 µg m\(^{-3}\) for MOZART (Figure 12), but in the range of 0-6 µg m\(^{-3}\) for RADM2 (Figure 14). Differences in NO\(_x\) concentrations predicted by the MOZART vs. RADM2 simulations are generally less than 20%, consistent with Knote et al. (2015), who conclude that uncertainty due to choice in chemical mechanism leads to an uncertainty of up to 25% in 3-D model simulations.

Performance of the present simulations with respect to NO\(_2\) can also be compared to previous published studies (note that none of the above-cited studies perform a validation for NO or NO\(_x\)). Zhang et al. (2013b) reports NMB values of approximately -15% for NO\(_2\) for WRF-Chem simulations against hourly AirBase measurements for July 2001, in line with values of -12% and -19% for the MOZART and RADM2 simulations in this study, respectively. Tuccella et al. (2012) report a MB for NO\(_2\) of -0.9 µg m\(^{-3}\) averaged over the whole year; for comparison the RADM2 simulation in this study shows a MB in the range of -2.5 to -1 µg m\(^{-3}\) for fall, spring and summer, but a MB of
+0.67 µg m\(^{-3}\) in summer. Evaluation of NO\(_2\) was not treated in detail in the AQMEII studies, but Im et al. (2015) report that the models for the European domain underestimate NO\(_2\) by 9% to 45%.

4.3 Characterization of MOZART vs. RADM2 differences

In this section, we explore the differences in surface O\(_3\) between the MOZART and RADM2 simulations by examining net O\(_3\), NO\(_2\), and NO production rates as well as the NO\(_x\)- and VOC-sensitivity of the two mechanisms. We further conducted sensitivity simulations to investigate the relative contributions of different sources to the observed differences in surface O\(_3\) predicted by MOZART and RADM2. The month of July was chosen for the sensitivity simulations since O\(_3\) concentrations over Europe are highest during summer, and thus summer is most the most important season when considering air quality exceedances and health impacts of O\(_3\). Additionally, MOZART and RADM2 show the largest differences in predicted O\(_3\) during this season (see Tables 5 and 8).

To gain insight into model behavior for O\(_3\), we added terms to the model output representing hourly accumulated tendencies, i.e., the change in concentration of a species due to photochemistry only, for July simulations using MOZART and RADM2. The hourly net photochemical production rate was calculated as the difference in the accumulated tendency from one timestep to another. Figure 15 shows the average of the midday (11:00-14:00 CEST, or 9:00-12:00 UTC) photochemical production rate of O\(_3\) and NO\(_x\) components for both the MOZART and RADM2 simulations. (Note that the net photochemical production rate is shown here in ppb hr\(^{-1}\) for more intuitive comparison of production and loss of the different species on a mole basis; µg m\(^{-3}\) was used in Section 4.2 because this is the unit in which limit and target values in the EU Air Quality Directive are expressed.)

Overall, the spatial variability as well as the magnitudes of net O\(_3\) production rates are found to be similar for MOZART-4 and RADM2 chemistry (Figure 15). For both mechanisms, the greatest midday net O\(_3\) production rates are found in southern Europe, particularly over the Mediterranean and Atlantic coasts. The difference in net O\(_3\) production rate between the two mechanisms is also shown in Figure 15. MOZART exhibits greater net O\(_3\) photochemical production rates than RADM2 for most of Europe, with the exception of the southeast corner of the domain (Greece, Turkey, and the nearby Mediterranean), where net O\(_3\) production rates are greater for RADM2. The difference in net O\(_3\) production rate (MOZART-RADM2) shows a large maximum over central Europe, centering over Germany and extending west and east into France and Poland. Over Germany, net O\(_3\) production in MOZART is seen to be higher than in RADM2 by 1.8 ppb hr\(^{-1}\) or more.

As expected, regions of high NO\(_2\) production in both MOZART and RADM2 simulations are seen over the high NO\(_x\)-emission regions including Benelux, southern England, western Germany, the Po Valley, and major cities including Paris and Moscow. The difference in net NO\(_2\) production rate between the two mechanisms is also highest where the absolute NO\(_2\) production rates are highest; in these areas the net NO\(_2\) production rate is lower for MOZART than for RADM2 by greater than 0.25 ppb hr\(^{-1}\). Furthermore, areas where the two mechanisms show the greatest differences in net
NO$_2$ production rate tend to be the areas where the net O$_3$ production rate is most different between the two mechanisms, including the large maximum over the Netherlands and northwest Germany.

To further investigate the differences between ozone chemistry in MOZART vs. RADM2, we performed two additional sensitivity studies with each mechanism: one in which all anthropogenic NO$_x$ emissions were increased by 30%, and one in which all anthropogenic VOC emissions were increased by 30%. We then examined the change in O$_3$ concentrations due to these emission perturbations to diagnose whether the chemical mechanisms were operating in a NO$_x$-sensitive or a VOC-sensitive regime. Results are shown in Figure 16. For the simulations where NO$_x$ emissions were increased by 30%, MOZART and RADM2 show very similar behavior. Most of the domain is NO$_x$ sensitive, with increased NO$_x$ emissions resulting in increased modeled O$_3$. Notably, the U.K., Benelux, northern France and Paris, and northwest Germany show NO$_x$-saturated behavior, in which increased NO$_x$ emissions lead to decreased O$_3$ concentrations. NO$_x$-saturated regimes are also seen around the area of the Mediterranean between Monaco, Genoa and Corsica. An alternate approach to identify areas of NO$_x$-sensitive vs. NO$_x$-saturated regimes is to use indicator ratios (in the base simulation) following Sillman (1995). We have applied this approach with the indicator ratio CH$_2$O/NO$_y$ (Figure S12) and find that areas identified as NO$_x$ sensitive using the indicator ratio are the same as those identified using the simulation with +30% NO$_x$ emissions. These results are also consistent with the areas of Europe found to be NOx saturated in the model study of Beekmann and Vautard (2010). Magnitudes of the observed change in O$_3$ in response to increased NO$_x$ emissions are quite similar for both mechanisms, although RADM2 shows slightly stronger NO$_x$ saturation (i.e., a stronger decrease in O$_3$ given a 30% increase in NO$_x$ emissions) in the area centered around Benelux, and stronger NO$_x$ sensitivity over Scandinavia and northwest Russia.

In contrast to the similar behavior seen for NO$_x$ sensitivity, the VOC sensitivity exhibited by the two mechanisms is quite different (Figure 16, lower panel). For both MOZART and RADM2, the effect of increased anthropogenic VOC emissions on O$_3$ is smaller than the effect of increased NO$_x$ emissions. The MOZART simulation shows very little impact of increased VOC emissions on O$_3$, with differences in average O$_3$ concentration generally confined to ± 2% of the base simulation. In contrast, increasing VOC emissions in the RADM2 simulations leads to increased O$_3$ concentrations throughout nearly the entire domain. Areas where MOZART and RADM2 are in agreement in predicting VOC sensitivity (increased O$_3$ concentrations in response to increased VOC emissions) are generally those with high NO$_x$ emissions, where one would expect the highest VOC sensitivity based on theory; these areas include Benelux, northern France, northwest Germany, and shipping tracks in the Mediterranean. However, the increase in O$_3$ concentration is modest for both mechanisms; for RADM2 it is generally limited to increases of 2-4% over the base simulation. The results of the +30% VOC sensitivity studies for July indicate that d[O$_3$]/d[VOC] is higher (more positive) for RADM2 than for MOZART for the chemical regime represented by the models in July 2007. This shows that the two mechanisms are simulating different O$_3$ chemical regimes – in the case of
RADM2, there is greater VOC sensitivity, meaning that addition of VOC emissions moves the chemistry in the direction of maximum $O_3$ production efficiency; this is not the case for MOZART over much of the domain. A more extensive study would be needed to evaluate whether the conclusion that $d[O_3]/d[VOC]$ is higher for RADM2 than for MOZART can be applied more generally.

Taken as a whole, Figure [16] shows that MOZART behaves in a classically $NO_x$-sensitive manner for most of domain, with $O_3$ responding to changes in $NO_x$ but showing little response to changes in anthropogenic VOC. $NO_x$-saturated behavior is also observed, particularly around the area of U.K., Benelux, and northern France and Germany. RADM2, on the other hand, exhibits more of a mixed $NO_x$-VOC-sensitivity for much of the domain. The $NO_x$ sensitivity seen in RADM2 is very similar to that seen in MOZART, but the response of RADM2 to changes in VOC is much stronger (by about a factor of two) than observed in MOZART. With the exception of some small areas in the North and Baltic Sea south of Norway and Sweden, RADM2 predicts $O_3$ increases with VOC increases throughout the entire domain. This difference in VOC sensitivity seen between the mechanisms has implications for policy decisions, as it indicates uncertainty in the European response of $O_3$ to policies designed to reduce anthropogenic VOC emissions.

In addition to characterizing mechanism behavior with respect to net photochemical $O_3$ production and $NO_x$- and VOC-sensitivity, we evaluate the contribution of other sources that could explain the large differences in predicted $O_3$ between the MOZART and RADM2 simulations. First, MOZART and RADM2 use different rate coefficients for several inorganic gas phase chemical reactions. To test the effect of these differences, all RADM2 inorganic reaction rate coefficients were changed so that they matched those used in MOZART simulations in the cases where the reactions are the same in both mechanisms (Section S3 in the Supplementary Material). The differences in inorganic rate coefficients between the two mechanisms explain a significant difference in predicted $O_3$ concentrations: when RADM2 is run with inorganic rate coefficients from MOZART, the resulting domain-mean $O_3$ is higher by more than 8 µg m$^{-3}$ for the month of July, approximately 40% of the difference in predicted $O_3$.

Besides the gas-phase chemistry itself, there are some differences in the implementation of MOZART-4 vs. RADM2 in WRF-Chem that could also contribute to the observed differences in modeled $O_3$: in particular, in the treatment of dry deposition and photolysis (described in the Supplementary Material, Section S2). To test the effect of differences in treatment of dry deposition, we conducted an additional sensitivity in which we modified the RADM2 simulation to treat dry deposition in the same way as it is treated in MOZART. However, this led to only a small difference in average ozone (an increase of 1 µg m$^{-3}$), indicating that modeled surface $O_3$ concentrations are relatively insensitive to these differences in the treatment of dry deposition, at least in the summer. In a sensitivity test where we modified the model code so that the MOZART simulation ran with the same photolysis scheme as used in our RADM2 simulation (i.e., with the Madronich TUV scheme and without reading in climatological $O_3$ and $O_2$ columns), we found that average $O_3$ for July decreases...
by 3 µg m$^{-3}$. This indicates that modeled O$_3$ is also somewhat sensitive to differences in the treatment of photolysis in MOZART and RADM2. However, taken together, our sensitivity simulations suggest that the differences in the inorganic reaction rate coefficients are more important than the differing treatments of dry deposition and photolysis in explaining the differences in predicted O$_3$ between the RADM2 and MOZART simulations.

5 Summary and Conclusions

In this paper, we present a detailed description of a WRF-Chem setup over the European domain and provide an evaluation of the simulated meteorological and chemical fields with an emphasis on model’s ability to reproduce the spatial and temporal distribution of ground-level O$_3$ and NO$_x$. Within WRF-Chem we compare the performance of two different chemical mechanisms: MOZART-4, for which we present the first model evaluation for a European domain, and RADM2. Overall, we found that our WRF-Chem setup reproduced the spatial and seasonal variations in the meteorological parameters over Europe, with biases and correlations consistent with previous studies. Simulations using the MOZART-4 as well as RADM2 chemical mechanisms were found to reproduce the spatial and temporal distributions in ground-level O$_3$ over Europe, based on observations from the EMEP and Airbase networks. However, we find significant differences in O$_3$ concentrations predicted by the two chemical mechanisms, with RADM2 predicting as much as 20 µg m$^{-3}$ less O$_3$ than MOZART during the spring and summer seasons. In general, MOZART-4 chemistry overpredicts O$_3$ concentrations for most of Europe in the summer and fall, whereas RADM2 leads to an underestimation of O$_3$ over the European domain in all seasons. Taken as a whole, use of MOZART-4 chemistry performs better, leading to lower absolute model biases in O$_3$. This is the case when considering hourly O$_3$ concentrations as well as metrics relevant for human health, such as MDA8 and SOMO35. Despite the large differences in predicted O$_3$, the two mechanisms show relatively similar behavior for NO$_x$, with both MOZART and RADM2 simulations resulting in a slight underestimation of NO$_x$ compared to surface observations.

The net midday photochemical production rate of O$_3$ in summer is found to be higher for MOZART than for RADM2 for most of the domain, with the largest differences between the mechanisms seen over Germany, where the net O$_3$ photochemical production for MOZART is higher than for RADM2 by greater than 1.8 ppb hr$^{-1}$ (3.6 µg m$^{-3}$ hr$^{-1}$). However, we have shown that RADM2 is approximately twice as sensitive to increases in anthropogenic VOC emissions as MOZART, suggesting that, under local VOC-limited conditions not seen at the regional scale of our simulations, RADM2 is likely to produce O$_3$ at a greater rate than MOZART. Despite the differences in sensitivity to changes in VOC emissions exhibited by the two mechanisms, sensitivity to changes in NO$_x$ emissions in MOZART and RADM2 are found to be similar.
Our results indicate that modeled surface $O_3$ over Europe is sensitive to the choice of gas phase chemical mechanism, with observed differences in $O_3$ between mechanisms that are larger than those seen in many past studies. Although the most fundamental differences between MOZART-4 and RADM2 (and other chemical mechanisms used in regional modeling) is the representation of VOC oxidation chemistry, we find that approximately 40% of the difference seen in predicted $O_3$ seen in this study can be explained by differences in inorganic reaction rate coefficients employed by MOZART-4 and RADM2. This result suggests that harmonization of inorganic rate coefficients among chemical mechanisms used for regional air quality modeling might be valuable, and could potentially lead to a smaller spread in model-predicted $O_3$ compared to that seen in, e.g., the multi-model studies of AQMEII (Solazzo et al., 2012b; Im et al., 2015). Further investigation of chemical mechanism behavior within 3-D models in general would be helpful to constrain uncertainties in regional air quality modeling.

6 Code availability

The WRF-Chem model is an open-source, publicly available software. The code is being continually improved, with new releases approximately twice per year. WRF-Chem code can be downloaded at [http://www2.mmm.ucar.edu/wrf/users/download/get_source.html](http://www2.mmm.ucar.edu/wrf/users/download/get_source.html). The corresponding author will provide the bug fixes to version 3.5.1 used in this study, described in Section 2.3, upon request.

Appendix A: Abbreviations and Acronyms

DJF: December-January-February (winter)

EDGAR: Emission Database for Global Atmospheric Research

EEA: European Environmental Agency

EOS: Earth Observing System

GEOS5: Goddard Earth Observing System Model, Version 5

GOCART: Goddard Chemistry Aerosol Radiation and Transport

HTAP: Hemispheric Transport of Air Pollution

JJA: June-July-August (summer)

MADE: Modal Aerosol Dynamics Model for Europe

MAM: March-April-May (spring)

MERRA: Modern Era-Retrospective Analysis for Research and Applications

NCEP: National Centers for Environmental Prediction

NCAR: National Center for Atmospheric Research

SON: September-October-November (fall)

SORGAM: Secondary Organic Aerosol Model

WRF-Chem: Weather Research and Forecasting with Chemistry
Appendix B: Definitions of statistical quantities

The statistical quantities used for model evaluation are defined below. Let $Obs^j_i$ and $Mod^j_i$ be the observed and modeled quantities at time $i$ and station $j$, respectively. $N_{obs}^j$ represents the number of temporal data points evaluated at station $j$, and $N_{obs}$ represents the total number of data points (each representing a time $i$ and a station $j$) evaluated in the domain.

The Mean Bias (MB) at a specific station (e.g., Figure 5) is calculated as

$$MB^j = \frac{1}{N_{obs}^j} \sum_{i=1}^{N_{obs}^j} (Mod^j_i - Obs^j_i)$$

and the domain-wide Mean Bias (e.g., Table 5) as

$$MB = \frac{1}{N_{obs}} \sum_{i,j=1}^{N_{obs}} (Mod^j_i - Obs^j_i)$$

Domain-wide values for Normalized Mean Bias (NMB) and Mean Fractional Bias (MFB) are calculated analogously.

$$NMB = \frac{\sum_{i=1}^{N_{obs}} (Mod^j_i - Obs^j_i)}{\sum_{i=1}^{N_{obs}} Obs^j_i}$$

$$MFB = \frac{1}{N_{obs}} \sum_{i,j=1}^{N_{obs}} \frac{(Mod^j_i - Obs^j_i)}{\frac{Mod^j_i + Obs^j_i}{2}}$$

Temporal correlation between model results and observation is evaluated using the Pearson correlation coefficient ($r$). The value of $r$ is calculated at each station using

$$r^j = \frac{\sum_{i=1}^{N_{obs}} (Mod^j_i - \overline{Mod^j})(Obs^j_i - \overline{Obs^j})}{\sigma_{mod} \times \sigma_{obs}}$$

Here, the numerator represents the covariance between the model and observations, $\overline{Mod^j}$ and $\overline{Obs^j}$ represent the mean of the model and observations, respectively, and $\sigma$ is the standard deviation.

The domain-wide correlation coefficients (e.g., Table 5) is then calculated as

$$r = \frac{1}{N_j} \sum_{j} r^j$$

where $N_j$ is the total number of stations.
Acknowledgements. The authors would like to thank Renate Forkel for valuable discussions regarding the setup of our WRF-Chem simulation, and two anonymous reviewers for their feedback. The authors also thank Jane Coates for sharing her technique for VOC speciation and valuable discussions regarding chemical mechanisms. We thank TNO for access to the TNO-MACC II emissions inventory, and Hugo Denier van der Gon for helpful discussions regarding emissions. The HTAP v2.2 anthropogenic emissions were obtained from http://edgar.jrc.ec.europa.eu/htap_v2/index.php. The authors thank Christophe Knote and Anna Katinka Petersen for sharing bug fixes for the WRF-Chem MOZART code. WRF-Chem tools for preprocessing boundary conditions as well as biogenic, fire, and anthropogenic emissions were provided by NCAR (http://www.acom.ucar.edu/wrf-chem/download.shtml). Initial and boundary conditions for meteorological fields were obtained from ECMWF, http://www.ecmwf.int/en/research/climate-reanalysis/era-interim. Initial and boundary conditions for chemical fields were from MOZART-4/GEOS5, provided by NCAR at http://www.acd.ucar.edu/wrf-chem/mozart.shtml. Corine land cover data was obtained from http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-2. We acknowledge the UK Met Office for providing the Global Weather Observation dataset via the British Atmospheric Data Centre. We acknowledge EMEP and the Norwegian Institute for Air Research (NILU) for providing the EMEP chemical observation data via the EBAS database (ebas.nilu.no). AirBase is the public air quality database of the EEA; data were obtained at http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-7. The WRF-Chem simulations have been performed on the supercomputer HYDRA (http://www.rzg.mpg.de/).
References


Table 1. WRF-Chem options used in model simulations.

<table>
<thead>
<tr>
<th>Atmospheric Process</th>
<th>Option used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud microphysics</td>
<td>Lin et al. scheme (Lin et al., 1983)</td>
</tr>
<tr>
<td>Longwave radiation</td>
<td>RRTMG (Iacono et al., 2008)</td>
</tr>
<tr>
<td>Shortwave radiation</td>
<td>Goddard shortwave scheme (Chou and Suarez, 1994)</td>
</tr>
<tr>
<td>Surface Layer</td>
<td>MM5 Similarity based on Monin-Obukhov scheme (Beljaars, 1995)</td>
</tr>
<tr>
<td>Land-surface Physics</td>
<td>Noah Land Surface Model (Chen and Dudhia, 2001)</td>
</tr>
<tr>
<td>Urban surface physics</td>
<td>Urban Canopy Model (Kusaka and Kimura, 2004)</td>
</tr>
<tr>
<td>Planetary boundary layer</td>
<td>Yonsei University scheme (Hong et al., 2006)</td>
</tr>
<tr>
<td>Cumulus parametrization</td>
<td>Grell 3D scheme (Grell and Dévényi, 2002)</td>
</tr>
</tbody>
</table>

Table 2. Description of WRF-Chem simulations performed for this study.

<table>
<thead>
<tr>
<th>Simulation Name</th>
<th>Model Chemistry</th>
<th>Photolysis Scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) MOZART</td>
<td>MOZART- 4 chemistry with goetz aerosols, KPP solver</td>
<td>Madronich F-TUV photolysis</td>
</tr>
<tr>
<td>(2) RADM2</td>
<td>RADM2 chemistry with MADE/SORGAM aerosols, KPP solver</td>
<td>Madronich photolysis (TUV)</td>
</tr>
</tbody>
</table>

Table 3. Observational datasets used for model evaluation.

<table>
<thead>
<tr>
<th>Database</th>
<th>Parameter</th>
<th>Temporal Resolution</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BADC Global Weather Observation Data</td>
<td>MSLP, T2, WS10, WD10</td>
<td>3-hourly</td>
<td><a href="http://badc.nerc.ac.uk/home/">http://badc.nerc.ac.uk/home/</a></td>
</tr>
<tr>
<td>EMEP</td>
<td>NO₂, NO, NOₓ</td>
<td>hourly</td>
<td><a href="http://ebas.nilu.no/">http://ebas.nilu.no/</a></td>
</tr>
</tbody>
</table>
Table 4. Domain-wide statistical performance of WRF-Chem against 3-hourly meteorological observations from BADC. Modeled quantities are from the MOZART simulation.

<table>
<thead>
<tr>
<th></th>
<th>Winter (DJF)</th>
<th>Spring (MAM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
</tr>
<tr>
<td>Mean-Obs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSLP (hPa)</td>
<td>1015.41</td>
<td>1014.79</td>
</tr>
<tr>
<td>T2 (°C)</td>
<td>2.51</td>
<td>2.99</td>
</tr>
<tr>
<td>WS10 (m/s)</td>
<td>4.31</td>
<td>5.60</td>
</tr>
<tr>
<td>WD10 (deg)</td>
<td>175.53</td>
<td>203.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean-Obs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer (JJA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSLP (hPa)</td>
<td>1012.12</td>
<td>1012.11</td>
</tr>
<tr>
<td>T2 (°C)</td>
<td>17.82</td>
<td>17.70</td>
</tr>
<tr>
<td>WS10 (m/s)</td>
<td>3.45</td>
<td>3.90</td>
</tr>
<tr>
<td>WD10 (deg)</td>
<td>173.88</td>
<td>196.92</td>
</tr>
</tbody>
</table>
Table 5. Statistics for MOZART simulation against hourly observations from the AirBase network. Means and MB are expressed in µg m^{-3}; NMB, MFB, and r are unitless. r is the hourly temporal correlation coefficient for all quantities except MDA8, for which it represents the daily temporal correlation coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Winter (DJF)</th>
<th></th>
<th>Spring (MAM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
</tr>
<tr>
<td>O_3</td>
<td>53.82</td>
<td>48.34</td>
<td>-5.44</td>
</tr>
<tr>
<td>MDA8</td>
<td>67.50</td>
<td>64.20</td>
<td>-3.30</td>
</tr>
<tr>
<td>NO_x</td>
<td>20.22</td>
<td>16.99</td>
<td>-3.20</td>
</tr>
<tr>
<td>NO_2</td>
<td>14.40</td>
<td>14.83</td>
<td>0.48</td>
</tr>
<tr>
<td>NO</td>
<td>4.27</td>
<td>1.18</td>
<td>-3.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Summer (JJA)</th>
<th></th>
<th>Fall (SON)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
</tr>
<tr>
<td>O_3</td>
<td>70.84</td>
<td>80.72</td>
<td>9.92</td>
</tr>
<tr>
<td>MDA8</td>
<td>94.51</td>
<td>110.37</td>
<td>15.86</td>
</tr>
<tr>
<td>NO_x</td>
<td>10.63</td>
<td>10.57</td>
<td>-0.10</td>
</tr>
<tr>
<td>NO_2</td>
<td>8.30</td>
<td>9.66</td>
<td>1.37</td>
</tr>
<tr>
<td>NO</td>
<td>2.01</td>
<td>0.48</td>
<td>-1.53</td>
</tr>
</tbody>
</table>
Table 6. Statistics for MOZART simulation against hourly observations from the EMEP network. Means and MB are expressed in µg m\(^{-3}\); NMB, MFB, and r are unitless. r is the hourly temporal correlation coefficient for all quantities except MDA8, for which it represents the daily temporal correlation coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Winter (DJF)</th>
<th>Spring (MAM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Winter (DJF)</td>
<td>Spring (MAM)</td>
</tr>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O(_3)</td>
<td>54.54</td>
<td>43.82</td>
</tr>
<tr>
<td></td>
<td>78.99</td>
<td>68.62</td>
</tr>
<tr>
<td>MDA8</td>
<td>64.66</td>
<td>55.09</td>
</tr>
<tr>
<td></td>
<td>95.64</td>
<td>90.15</td>
</tr>
<tr>
<td>NO(_x)</td>
<td>11.36</td>
<td>12.39</td>
</tr>
<tr>
<td></td>
<td>10.21</td>
<td>10.44</td>
</tr>
<tr>
<td>NO(_2)</td>
<td>10.19</td>
<td>13.24</td>
</tr>
<tr>
<td></td>
<td>8.07</td>
<td>10.72</td>
</tr>
<tr>
<td>NO2</td>
<td>2.10</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>1.34</td>
<td>0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Summer (JJA)</th>
<th>Fall (SON)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summer (JJA)</td>
<td>Fall (SON)</td>
</tr>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O(_3)</td>
<td>72.08</td>
<td>76.39</td>
</tr>
<tr>
<td></td>
<td>53.24</td>
<td>52.05</td>
</tr>
<tr>
<td>MDA8</td>
<td>91.24</td>
<td>101.48</td>
</tr>
<tr>
<td></td>
<td>66.99</td>
<td>70.37</td>
</tr>
<tr>
<td>NO(_x)</td>
<td>7.62</td>
<td>8.44</td>
</tr>
<tr>
<td></td>
<td>11.83</td>
<td>12.14</td>
</tr>
<tr>
<td>NO(_2)</td>
<td>6.07</td>
<td>9.10</td>
</tr>
<tr>
<td></td>
<td>8.88</td>
<td>13.81</td>
</tr>
<tr>
<td>NO2</td>
<td>1.23</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>1.42</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Table 7. Statistics for yearly SOMO35 in mg m\(^{-3}\) · days.

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Observation network</th>
<th>Obs</th>
<th>Model</th>
<th>MB</th>
<th>NMB</th>
<th>MFB</th>
<th>no. stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOZART</td>
<td>AirBase</td>
<td>6.23</td>
<td>8.22</td>
<td>1.98</td>
<td>0.32</td>
<td>0.30</td>
<td>375</td>
</tr>
<tr>
<td>MOZART</td>
<td>EMEP</td>
<td>5.73</td>
<td>6.27</td>
<td>0.51</td>
<td>0.09</td>
<td>0.11</td>
<td>122</td>
</tr>
<tr>
<td>RADM2</td>
<td>AirBase</td>
<td>6.23</td>
<td>2.55</td>
<td>-3.68</td>
<td>-0.59</td>
<td>-0.87</td>
<td>375</td>
</tr>
<tr>
<td>RADM2</td>
<td>EMEP</td>
<td>5.73</td>
<td>1.84</td>
<td>-3.91</td>
<td>-0.68</td>
<td>-1.13</td>
<td>122</td>
</tr>
</tbody>
</table>
Table 8. Statistics for RADM2 simulation against hourly observations from the AirBase network. Means and MB are expressed in $\mu g m^{-3}$; NMB, MFB, and r are unitless. r is the hourly temporal correlation coefficient for all quantities except MDA8, for which it represents the daily temporal correlation coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Winter (DJF)</th>
<th></th>
<th>Spring (MAM)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
<td>NMB</td>
</tr>
<tr>
<td>O$_3$</td>
<td>53.82</td>
<td>41.57</td>
<td>-12.18</td>
<td>-0.23</td>
</tr>
<tr>
<td>MDA8</td>
<td>67.50</td>
<td>56.04</td>
<td>-11.46</td>
<td>-0.17</td>
</tr>
<tr>
<td>NO$_x$</td>
<td>20.22</td>
<td>13.75</td>
<td>-6.45</td>
<td>-0.32</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>14.40</td>
<td>11.90</td>
<td>-2.47</td>
<td>-0.17</td>
</tr>
<tr>
<td>NO</td>
<td>4.27</td>
<td>0.97</td>
<td>-3.31</td>
<td>-0.77</td>
</tr>
<tr>
<td></td>
<td>Summer (JJA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O$_3$</td>
<td>70.84</td>
<td>57.79</td>
<td>-13.01</td>
<td>-0.18</td>
</tr>
<tr>
<td>MDA8</td>
<td>94.51</td>
<td>80.59</td>
<td>-13.92</td>
<td>-0.15</td>
</tr>
<tr>
<td>NO$_x$</td>
<td>10.63</td>
<td>9.79</td>
<td>-0.87</td>
<td>-0.08</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>8.30</td>
<td>8.95</td>
<td>0.67</td>
<td>0.08</td>
</tr>
<tr>
<td>NO</td>
<td>2.01</td>
<td>0.46</td>
<td>-1.55</td>
<td>-0.77</td>
</tr>
</tbody>
</table>
Table 9. Statistics for RADM2 simulation against hourly observations from the EMEP network. Means and MB are expressed in µg m⁻³; NMB, MFB, and r are unitless. r is the hourly temporal correlation coefficient for all quantities except MDA8, for which it represents the daily temporal correlation coefficient.

<table>
<thead>
<tr>
<th></th>
<th>Winter (DJF)</th>
<th></th>
<th>Winter (DJF)</th>
<th></th>
<th>Winter (DJF)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
<td>NMB</td>
<td>MFB</td>
<td>r</td>
</tr>
<tr>
<td>O₃</td>
<td>54.54</td>
<td>38.67</td>
<td>-15.62</td>
<td>-0.29</td>
<td>-0.36</td>
<td>0.54</td>
</tr>
<tr>
<td>MDA8</td>
<td>64.66</td>
<td>49.40</td>
<td>-15.26</td>
<td>-0.24</td>
<td>-0.27</td>
<td>0.56</td>
</tr>
<tr>
<td>NOₓ</td>
<td>11.36</td>
<td>10.31</td>
<td>-0.99</td>
<td>-0.09</td>
<td>-0.02</td>
<td>0.38</td>
</tr>
<tr>
<td>NO₂</td>
<td>10.19</td>
<td>10.72</td>
<td>0.56</td>
<td>0.06</td>
<td>0.03</td>
<td>0.51</td>
</tr>
<tr>
<td>NO</td>
<td>2.10</td>
<td>1.16</td>
<td>-0.93</td>
<td>-0.44</td>
<td>-0.67</td>
<td>0.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Spring (MAM)</th>
<th></th>
<th>Spring (MAM)</th>
<th></th>
<th>Spring (MAM)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
<td>NMB</td>
<td>MFB</td>
<td>r</td>
</tr>
<tr>
<td>O₃</td>
<td>78.99</td>
<td>53.24</td>
<td>-25.83</td>
<td>-0.33</td>
<td>-0.40</td>
<td>0.49</td>
</tr>
<tr>
<td>MDA8</td>
<td>95.64</td>
<td>71.04</td>
<td>-24.60</td>
<td>-0.26</td>
<td>-0.29</td>
<td>0.55</td>
</tr>
<tr>
<td>NOₓ</td>
<td>10.21</td>
<td>8.76</td>
<td>-1.31</td>
<td>-0.13</td>
<td>-0.24</td>
<td>0.30</td>
</tr>
<tr>
<td>NO₂</td>
<td>8.07</td>
<td>9.11</td>
<td>0.95</td>
<td>0.12</td>
<td>-0.19</td>
<td>0.34</td>
</tr>
<tr>
<td>NO</td>
<td>1.34</td>
<td>0.68</td>
<td>-0.67</td>
<td>-0.50</td>
<td>-0.59</td>
<td>0.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Summer (JJA)</th>
<th></th>
<th>Summer (JJA)</th>
<th></th>
<th>Summer (JJA)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
<td>NMB</td>
<td>MFB</td>
<td>r</td>
</tr>
<tr>
<td>O₃</td>
<td>72.08</td>
<td>55.65</td>
<td>-16.65</td>
<td>-0.23</td>
<td>-0.24</td>
<td>0.58</td>
</tr>
<tr>
<td>MDA8</td>
<td>91.24</td>
<td>74.75</td>
<td>-16.49</td>
<td>-0.18</td>
<td>-0.19</td>
<td>0.69</td>
</tr>
<tr>
<td>NOₓ</td>
<td>7.62</td>
<td>7.61</td>
<td>0.10</td>
<td>0.01</td>
<td>-0.24</td>
<td>0.28</td>
</tr>
<tr>
<td>NO₂</td>
<td>6.07</td>
<td>8.33</td>
<td>2.20</td>
<td>0.36</td>
<td>-0.02</td>
<td>0.29</td>
</tr>
<tr>
<td>NO</td>
<td>1.23</td>
<td>0.52</td>
<td>-0.73</td>
<td>-0.59</td>
<td>-0.58</td>
<td>0.25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Fall (SON)</th>
<th></th>
<th>Fall (SON)</th>
<th></th>
<th>Fall (SON)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean-Obs</td>
<td>Mean-Mod</td>
<td>MB</td>
<td>NMB</td>
<td>MFB</td>
<td>r</td>
</tr>
<tr>
<td>O₃</td>
<td>53.24</td>
<td>39.89</td>
<td>-13.21</td>
<td>-0.25</td>
<td>-0.29</td>
<td>0.57</td>
</tr>
<tr>
<td>MDA8</td>
<td>66.99</td>
<td>54.31</td>
<td>-12.68</td>
<td>-0.19</td>
<td>-0.21</td>
<td>0.63</td>
</tr>
<tr>
<td>NOₓ</td>
<td>11.83</td>
<td>10.59</td>
<td>-0.82</td>
<td>-0.07</td>
<td>-0.13</td>
<td>0.32</td>
</tr>
<tr>
<td>NO₂</td>
<td>8.88</td>
<td>11.48</td>
<td>2.71</td>
<td>0.31</td>
<td>0.04</td>
<td>0.39</td>
</tr>
<tr>
<td>NO</td>
<td>1.42</td>
<td>1.43</td>
<td>0.07</td>
<td>0.05</td>
<td>-0.31</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Figure 1. Seasonal average values of 2-meter temperature (T2) in degrees Celsius. Model results and statistics are shown for the MOZART simulation at the locations of the observations.
Figure 2. Seasonal average values of 10-meter wind speed (WS10) in m/s. Model results and statistics are shown for the MOZART simulation at the locations of the observations.
Figure 3. Seasonal average values of 10-meter wind direction (WD10) in degrees. Model results and statistics are shown for the MOZART simulation at the locations of the observations.
Figure 4. Seasonal average values of surface O$_3$ in µg m$^{-3}$. Contours are model output from the MOZART simulation. Filled dots represent hourly measurements at AirBase rural background stations, filled squares represent measurements at EMEP stations.
Figure 5. Seasonal average values of surface O₃ in µg m⁻³ from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values from MOZART for corresponding locations. The Mean Bias (MB, in µg m⁻³) and temporal correlation coefficient (r) for hourly values are also shown at the location of station observations.
Figure 6. Seasonal average values of MDA8 in µg m$^{-3}$ calculated from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values from MOZART for corresponding locations. The Mean Bias (MB, in µg m$^{-3}$) and temporal correlation coefficient (r) for daily values are also shown at the location of station observations.
Figure 7. Number of days of exceedances of the EU long-term objective value for MDA8 (120 µg m$^{-3}$) at AirBase (circles) and EMEP (squares) station locations. Shown are totals by season for observations and the MOZART and RADM2 simulations. For simplicity of viewing the data, stations with no exceedances are not plotted.
Figure 8. Yearly values of SOMO35 in mg m$^{-3}$·days calculated from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values for corresponding locations.
Figure 9. Seasonal average values of surface $O_3$ in $\mu g m^{-3}$. Contours are model output from the RADM2 simulation. Filled dots represent hourly measurements at AirBase rural background stations, filled squares represent measurements at EMEP stations.
Figure 10. Seasonal average values of surface $O_3$ in $\mu g m^{-3}$ from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values from RADM2 for corresponding locations. The Mean Bias (MB, in $\mu g m^{-3}$) and temporal correlation coefficient ($r$) for hourly values are also shown at the location of station observations.
Figure 11. Seasonal average values of surface NO$_x$ in µg m$^{-3}$. Contours are model output from the MOZART simulation. Filled dots represent hourly measurements at AirBase rural background stations, filled squares represent measurements at EMEP stations.
Figure 12. Seasonal average values of surface NO$_2$ in $\mu$g m$^{-3}$ from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values from MOZART for corresponding locations. The Mean Bias (MB) and temporal correlation coefficient (r) for hourly values are also shown at the location of station observations.
Figure 13. Seasonal average values of surface NO in µg m\(^{-3}\) from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values from MOZART for corresponding locations. The Mean Bias (MB) and temporal correlation coefficient (r) for hourly values are also shown at the location of station observations.
Figure 14. Seasonal average values of surface NO$_2$ in $\mu$g m$^{-3}$ from hourly measurements at AirBase (circles) and EMEP (squares) stations, and modeled values from RADM2 for corresponding locations. The Mean Bias (MB) and temporal correlation coefficient ($r$) for hourly values are also shown at the location of station observations.
Figure 15. Net midday (11:00 - 14:00 CEST) photochemical production rate in ppb hr$^{-1}$ for O$_3$, NO$_2$, and NO shown for MOZART and RADM2 for July 2007. The last row shows the difference in net production rate in ppb hr$^{-1}$ (RADM2 subtracted from MOZART).
Figure 16. Sensitivity of average $O_3$ for July 2007 to a 30% increase in emissions of $NO_x$ (upper row) or VOC (lower row), shown for the MOZART and RADM2 chemical mechanisms. Shown here is the percent change in $O_3$ concentration, i.e., $100 \times \frac{[O_3]_{+\text{30\% emissions}} - [O_3]_{\text{base}}}{[O_3]_{\text{base}}}$. 