Science of the Total Environment 906 (2024) 167665

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Dynamic evaluation of modeled ozone concentrations in Germany with four chemistry transport models

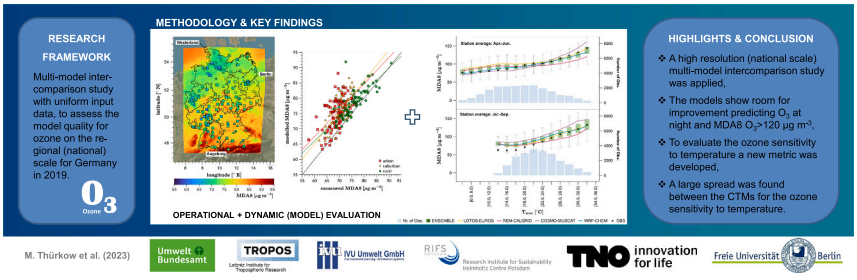
Markus Thürkow [a,*], Martijn Schaap [a,b], Richard Kranenburg [b], Florian Pfäfflin [c], Lina Neunhäuserer [c], Ralf Wolke [d], Bernd Heinold [d], Jens Stoll [d], Aura Lupaşcu [e,1], Stephan Nordmann [f], Andrea Minkos [f], Tim Butler [a,e]

[a] FUB, Institute of Meteorology, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany
[b] TNO, Department Climate, Air and Sustainability, Princetonlaan 6, 3584 CB Utrecht, the Netherlands
[c] IVU Umwelt GmbH, Emmy-Noether-Straße 2, 79110 Freiburg, Germany
[d] TROPOS, Leibniz Institute for Tropospheric Research, Permoserstraße 15, 04318 Leipzig, Germany
[e] RIFS Potsdam, Research Institute for Sustainability, Helmholtz Zentrum Potsdam, Berlinerstraße 130, 14467 Potsdam, Germany
[f] UBA, Umweltbundesamt, Wörlitzer Platz 1, 06844 Dessau-Roßlau, Germany

HIGHLIGHTS

- A high resolution (national scale) multi-model intercomparison study was applied.
- The models show room for improvement predicting O_3 at night and MDA8 O_3 > 120 μg m^3.
- To evaluate the ozone sensitivity to temperature a new metric was developed.
- A large spread was found between the CTMs for the ozone sensitivity to temperature.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Hai Guo

Keywords:
air quality
model evaluation
inter-comparison
ozone
nitrogen oxides

ABSTRACT

Simulating the ozone variability at regional scales using chemistry transport models (CTMs) remains a challenge. We designed a multi-model intercomparison to evaluate, for the first time, four regional CTMs on a national scale for Germany. Simulations were conducted with LOTOS-EUROS, REM-CALGRID, COSMO-MUSCAT and WRF-Chem for January 1st to December 31st, 2019, using prescribed emission information. In general, all models show good performance in the operational evaluation with average temporal correlations of MDA8 O_3 in the range of 0.77–0.87 and RMSE values between 16.3 μg m^{-3} and 20.6 μg m^{-3}. On average, better models' skill has been observed for rural background stations than for the urban background stations as well as for springtime compared to summertime. Our study confirms that the ensemble mean provides a better model-measurement agreement than individual models. All models capture the larger local photochemical production in summer compared to springtime and observed differences between the urban and the rural background. We introduce a

* Corresponding author.
E-mail address: markus.thuerkow@met.fu-berlin.de (M. Thürkow).
[1] Now at ECMWF, Bonn, Germany.

https://doi.org/10.1016/j.scitotenv.2023.167665
Received 28 July 2023; Received in revised form 20 September 2023; Accepted 6 October 2023
Available online 9 October 2023

new indicator to evaluate the dynamic response of ozone to temperature. During summertime a large ensemble spread in the ozone sensitivities to temperature is found with (on average) an underestimation of the ozone sensitivity to temperature, which can be linked to a systematic underestimation of mid-level ozone concentrations. During springtime we observed an ozone episode that is not covered by the models which is likely due to deficiencies in the representation of background ozone in the models. We recommend to focus on a diagnostic evaluation aimed at the model descriptions for biogenic emissions and dry deposition as a follow up and to repeat the operational and dynamic analysis for longer timeframes.

1. Introduction

Ozone (O_3) remains one of the most toxic and ecologically detrimental air pollutants in Europe. Alongside particulate matter and nitrogen oxides, exposure to ozone causes a substantial burden of diseases in Germany (e.g., Krug et al., 2019, 2020). Millions of people are exposed to ozone levels above the WHO guideline values (EEA, 2018). Ozone is not emitted but rather formed through complex chemical reactions in the atmosphere. Processes influencing ozone in the troposphere act over a range of spatial scales, from the global scale (Crutzen, 1973), through the regional scale (Schnell et al., 2015), down to the urban scale (Churkina et al., 2017). Ozone at a given location consists of the transported baseline concentration and the local production (Parrish et al., 2017; Derwent et al., 2018). The oxidation of methane in the remote troposphere and the subsequent long-range transport of remotely produced ozone into Europe contributes to about one third of the annual average baseline ozone concentration over Europe (Butler et al., 2020). The long-range contribution to the background is strongest in spring, while the local production is strongest in summer (HTAP, 2010). High local ozone concentrations associated with ozone threshold exceedance episodes are primarily attributable to regionally emitted ozone precursors (e.g., Reidmiller et al., 2009; Huang et al., 2017; Jonson et al., 2018; Lupaşcu and Butler, 2019; Lupaşcu et al., 2022).

The regional buildup of ozone is highly sensitive to several meteorological parameters such as temperature, moisture, and solar radiation and has been documented in numerous scientific studies based on both measurements and modeling (e.g., Seo et al., 2014; Coates et al., 2016; Kavassalis and Murphy, 2017; Otero et al., 2016, 2018; Luo et al., 2020). The synoptic state, such as high-pressure systems and/or blocking conditions, can boost the formation of ozone over a period of several days (Black et al., 2004). Biogenic volatile organic compound (BVOC) emissions are highly temperature dependent (Jacob and Winner, 2009; Monks et al., 2015) and correlate to clear sky conditions (Guenther et al., 2006, 2012). Their largest impact on ozone formation can be observed on warm and cloud-free days (Tawfik and Steiner, 2013) as photolysis is another crucial factor. Under high nitrogen oxide (NO_X) conditions the photochemical formation of ozone is enhanced with strong solar radiation (Kleinman, 1994), especially during summer when the solar insolation is at its largest (Schaap et al., 2015; Bessagnet et al., 2016). The photolysis of ozone can also act as a sink for ozone, for example in the remote marine boundary layer where NO_X concentrations are low (Oltmans and Levy, 1994). In high NO_X and low NMVOC (non-methane volatile organic compounds) conditions, the same occurs through the production of OH and subsequent reaction with nitrogen dioxide (NO_2). Vegetation is also an important sink of ozone through dry deposition. High temperatures along with low humidity cause plants to close their stomata to conserve water. This reduces the ozone removal (Fowler et al., 2009; Kavassalis and Murphy, 2017). A similar correlation was mentioned by Lin et al. (2020) for low soil moisture content when plants will tend to conserve water. Such a reduction in ozone dry deposition due to vegetative water stress can often be observed during extended episodes of high ozone levels related to heatwaves and droughts (Lin et al., 2020). Churkina et al. (2017) emphasized the role of enhanced precursor emissions such as VOCs and the removal effect of vegetation during heatwaves. Many of the chemical reactions involved in the production of ozone are also faster at higher temperatures (e.g., Sillman, 1995; Atkinson, 2007; Fischer et al., 2014; Coates et al., 2016).

Extensive modeling efforts are required to encompass ozone formation and removal processes with a sufficient high-quality representation notably observed for intense episodes. To hindcast and forecast the ambient air pollution of ozone at all scales, a hierarchy of numerical models is commonly used, consisting of global models (e.g., Young et al., 2018), regional models (e.g., Colette et al., 2017), and depending on the application, urban models (e.g., Maronga et al., 2019). Nowadays, it is common practice to conduct modeling frameworks using a multi-model (ensemble) approach with unified input data (emissions and/or meteorology) as done for example in the Copernicus Atmosphere Monitoring Service (CAMS). Results of an ensemble are often more robust compared to a single simulation and can increase the validity of the model results (Colette et al., 2017; Chen et al., 2019). Numerous model intercomparison studies for ozone have been performed on both global and regional scales (e.g., Rao et al., 2011; Foley et al., 2015a, 2015b; Bessagnet et al., 2016; Colette et al., 2017; Galmarini et al., 2017; Otero et al., 2018; Chen et al., 2019). The response of ozone to a uniform changing emission dataset has been studied in the HTAP framework (e.g., Galmarini et al., 2017; Jonson et al., 2018). Model performance assessments and the evaluation of model processes have been performed within the inter-comparison studies of ACCMIP (e.g., Stevenson et al., 2013; Lamarque et al., 2013) and CCMI (e.g., Morgenstern et al., 2017; Dhomse et al., 2018) for the global scale. The AQMEII (e.g., Rao et al., 2011; Solazzo and Galmarini, 2016; Galmarini et al., 2017), CAMS (e.g., Flemming et al., 2017; Inness et al., 2019) and EURODELTA (e.g., Colette et al., 2017) frameworks can be mentioned as examples for regional model evaluations. Within CAMS the needs for the development of regional air quality modeling aspects have been developed and described in detail. The CAMS_61 service provides information on how to handle daily forecasts and hindcasts using multi-model (ensemble) simulations. Only a few studies for air quality benchmarking on a national level have been reported so far.

The quality of model (ensemble) simulations in the past has predominantly been assessed by determining the model errors with respect to in-situ measurements using the operational evaluation. This can be done using different model quality indicators, such as those provided by the FAIRMODE initiative (Forum for AIR quality MODeling in Europe, https://fairmode.jrc.ec.europa.eu/). To assess the model errors in more detail, the dynamic (model) evaluation can be used (e.g., Dennis et al., 2010; Lecœur and Seigneur, 2013; Henneman et al., 2017). The dynamic evaluation allows a quality assessment of model simulations, based on the analysis of the relationship between air pollutants and different input drivers, such as emissions and/or meteorology. While the operational model evaluation compares the absolute modeled concentration to measurements, the dynamic evaluation is based on the comparison of changes in modeled and observed concentration levels (Dennis et al., 2010; Lecœur and Seigneur, 2013). This assesses whether the models can capture changes in concentrations related to different meteorological conditions or emission changes (Dennis et al., 2010). Modeled and observed sensitivities to different processes can further be diagnosed (Dennis et al., 2010; Henneman et al., 2017). Ozone metrics can be correlated with meteorological parameters as shown in previous studies (e.g., Lecœur and Seigneur, 2013; Otero et al., 2016, 2018). In Lecœur and Seigneur (2013), correlations were estimated for chemistry transport model simulations of particulate matter related to temperature,

precipitation, and wind speed and their 0 to 10 subsequent days. In Otero et al. (2016, 2018), a multiple linear regression analysis for different meteorological conditions was applied to estimate their relevance for ozone. Otero et al. (2016) also indicate the correlation of probabilistic threshold exceedances and estimates the meteorological impact to ozone extreme values.

In this study we (1) perform a model inter-comparison study with uniform input data (notably emissions and meteorology) to (2) assess the model quality on the regional (national) scale for Germany in 2019, by (3) a combined approach of the operational and the dynamic evaluation, aiming (4) to examine model specific behavior as a function of meteorology. In Section 2, we describe the experimental design of the model inter-comparison study, give information on all four models applied and the framework used for the operational and the dynamic evaluation. In Section 3 the results are provided, discussed and put in a broader perspective. In Section 4 the conclusion is presented.

2. Description of models and experimental design

2.1. Participating models

In the context of this study four Eulerian chemistry transport models have been evaluated and compared: LOTOS-EUROS, REM-CALGRID, COSMO-MUSCAT and WRF-Chem (Table 1). All models are designed as regional-scale, limited-area models to perform air quality assessments on short- and long timescales in the lower troposphere and can be used for process studies in scientific research activities as well as for regulatory efforts and policy advice. All models applied have previously participated in several model inter-comparison studies, such as EURODELTA (Colette et al., 2017) and/or AQMEII (Solazzo et al., 2012; Im et al., 2015a, 2015b), in which the performance of the models have been assessed to its peers. The models are also widely used for ozone applications (e.g., Flemming and Stern, 2007; Mar et al., 2016; Escudero et al., 2019). The models differ in their complexity with respect to chemical formation processes, vertical layering and especially in their meteorological driving. We employed two offline (LOTOS-EUROS & REM-CALGRID) and two online models (COSMO-MUSCAT & WRF-Chem). Characteristics relevant for the model's ozone prediction are summarized in Table 1. For a complete description with application examples of the models we refer to the key references given in the tables.

2.2. Experimental design

To guarantee the comparability of the results as much as possible, an intercomparison protocol was developed to harmonize anthropogenic emissions, meteorological input data, boundary conditions, model domains and resolutions as far as possible. The simulations were conducted for the year 2019 over a domain encompassing Germany (Fig. 1, upper panel). Each model simulation has been performed in a nested approach with the outer domain covering Europe. The outer model domain has been employed to encompass the impacts of long-range transport on air quality in Germany. The target domain over Germany is defined by a regular longitude-latitude grid with a spatial resolution of approximately 2×2 km^2 and also covers parts of the neighboring countries, e.g., the Netherlands, Belgium, Luxembourg, and Poland. We allow all models to freely select the number of required intermediate nested grids. By doing so, the expert knowledge for each model system could be used to represent the air quality in the most accurate manner. Note that, the model simulations for the inner nest are performed either with a grid definition matching or with a native grid definition close to the target grid, with all modeled concentrations harmonized to the target 2×2 km^2 grid for the purposes of intercomparison.

The model simulations were driven by meteorological input data provided by the German Weather Service (Deutscher Wetterdienst, DWD, e.g., Reinert et al. (2016)). For the large-scale European simulation, the ICON-EU meteorology with a horizontal resolution of about 7×7 km^2 was selected. Over Germany, the higher-resolved COSMO-D2 model (~2.2 km^2) has been applied. With the operational setup, the DWD provides 60 vertical layers for ICON-EU and 65 for COSMO-D2. The offline models (LOTOS-EUROS & REM-CALGRID) directly make use of these meteorological datasets as 1-hourly input for their chemical transport modeling. WRF-Chem and COSMO-MUSCAT (both online models) use their own methods for further processing the meteorological information according to the needs of their model specifics. For WRF-Chem and COSMO-MUSCAT, the meteorology is simulated as a hindcast itself, driven by initial and boundary conditions of the DWD models. To stay close to the DWD product, the model simulations are re-/initialized every 24–72 h. Aside from the reinitialization of meteorological input fields, WRF-Chem and COSMO-MUSCAT also nudge the 3-D meteorological fields of winds, potential temperature, water vapor mixing ratio, and geopotential every hour to ensure that the fields are not largely diverging from the DWD models.

For Germany, the officially reported anthropogenic emissions were provided by the German Environment Agency (Umweltbundesamt, UBA, www.uba.de) for all participating models individually, gridded using the GRETA system (Gridding Emission Tool for ArcGIS, Schneider et al. (2016)). For the rest of Europe, the regional inventory of CAMS for 2018 (CAMS-REG, Kuenen et al. (2022)) was used. The sector classification for both emission inventories follow the Gridded Nomenclature for Reporting (GNFR). The breakdown of the GNFR sector F (road transport) in GRETA towards the sub-categories F1-F4 is based on a factor split derived from CAMS-REG (Table 2). Area sources from GRETA's GNFR sectors A, B, D, J, and H have been vertically distributed by using the height profile presented in Table 3. All remaining area sources were to be emitted into the lowest model layers. Point sources from GRETA were vertically distributed using the height information that comes with the product. Profiles for the CAMS-recommended height distribution for point and area sources were adjusted by each model group individually on the corresponding grids of the outer nests. For all sectors, the emission time profiles were set to the CAMS temporal profiles to account for the hourly evolution from their emitting activity sectors. For road transport (GNFR sector F) the direct NO_2 emission percentage was set to 20 %. All other NO_X emission sources are distributed with a 97 % to 3 % (NO to NO_2) ratio. The composition of PM and NMVOCs corresponds to CAMS-REG and has been adjusted to the chemical mechanism of the respective models.

The land use classification is based on the EU-wide Corine Land Cover (CLC) dataset for 2018 (EEA, 2021). Over Germany, the dataset of the Federal Agency for Cartography and Geodesy (BKG, https://mis.bkg.bund.de) with a finer resolution of 5 ha was used. Global reanalysis data (ECMWF Atmospheric Composition Reanalysis 4: EAC4) of CAMS were used as boundary conditions around the outer European domain including the model top (Inness et al., 2019). The EAC4 data are based on global chemistry transport calculations in a horizontal resolution of 0.75° and are available in time steps of 3 h.

2.3. Evaluation metrics and measurement data

The modeling results were sampled for locations of the monitoring network from the German Environment Agency and the German federal states, as available in the central database at UBA. Based on the 2×2 km^2 mesh size of the target grid resolution, we only incorporated background sites (urban, suburban, and rural) into the model performance evaluation. Sites located above 900 m were excluded from the analysis. In total, 238/247 (O_3/NO_2) measurement sites were included (Fig. 1). The classification into urban (93/108), suburban (72/71), and rural (73/68) background sites was used to discriminate between more and less polluted areas in Germany. The distinction between rural and urban sites is used to consider any (potential) titration effects and provides information on the modeled ozone production efficiency, which may vary considerably with the location and timing of NO_X emissions.

Table 1
Description of basic structures, schemes and relevant parameters of the four participating models.

	LOTOS-EUROS	REM-CALGRID	COSMO-MUSCAT	WRF-Chem
MODEL OVERVIEW				
Key references	Manders et al. (2017)	Stern (2003)	Wolke et al. (2012)	Grell et al. (2005); Fast et al. (2006)
Version	2.2.002	4.0	5.05–02	3.9.1
Research group	TNO	IVU Umwelt	TROPOS	RIFS
Two-way feedback	Offline	Offline	Online	Online
INTERNAL MODEL GRID STRUCTURE				
Nesting & horizontal resolution	3 domains (28 × 32 km^2; 7 × 8 km^2; 2 × 2 km^2)	3 domains (28 × 32 km^2; 7 × 8 km^2; 2 × 2 km^2)	2 domains (14 × 14 km^2; 2 × 2 km^2)	3 domains (30 × 30 km^2; 10 × 10 km^2; 2 × 2 km^2)
Vertical layers	13	10/8	40/65	38
Vertical extent	8,000/4,000 m (7/10 layers below 1500 m)	3,000 m (8 layers below 1300 m)	8,000 m (MUSCAT: 11/20 layers below 1000 m)	20,000 m (11 layers below 1300 m)
Depth of first layer	20 m	20 m	20 m	25 m
INPUT DATA (HARMONIZED)				
Meteorology	ICON-EU & COSMO-D2 (both DWD)			
Emissions	German Environment Agency (UBA) for Germany & CAMS-REG (Kuenen et al., 2022) for Europe			
Boundary conditions	CAMS-EAC4 (Inness et al., 2019)			
Land use	Corine Land Cover 2018 (EEA, 2021) & the 5-ha dataset of the Federal Agency for Cartography and Geodesy (BKG)			
MODEL PROCESSES				
Advection	Based on Walcek (2000)	Based on Walcek (2000), modified by Yamartino (2003)	Based on Hundsdorfer et al. (1995); Wolke and Knoth (2000); Schlegel et al. (2012)	Third-order Runge–Kutta time-integration (Skamarock et al., 2008)
Vertical diffusion	K_z-theory, with K_z values calculated in the stability parameterization	Based on K-theory	Provided online by COSMO	Vertical turbulent mixing calculated online
Dry deposition	DEPAC (Van Zanten et al., 2010)	DEPAC (Van Zanten et al., 2010)	Based on Seinfeld and Pandis (2006); Schlünzen et al. (2012)	Based on Wesely (1989)
Wet deposition	Banzhaf et al. (2012)	Simple wash-out approach, scavenging rates are calculated from temperature- and species-dependent Henry's Law solubility and rainfall rate	Based on Simpson et al. (2012)	Based on Neu and Prather (2012)
Gas phase chemistry	CBM-IV (Gery et al., 1989) with 38 species and 96 reactions	CBM-IV (Gery et al., 1989) with 36 species and 93 reactions	RACM-MIM2-ext (Stockwell et al., 1997; Karl et al., 2006; Luttkus et al., 2022) with 140 species and 335 reactions	MOZART4 (Emmons et al., 2010) with 140 species and 335 reactions
Cloud chemistry	ph dependent oxidation scheme	No explicit cloud chemistry, simple parameterization of sulfate oxidation in clouds	Based on Schaap et al. (2004)	Double microphysics scheme (Morrison and Gettelman, 2008)
Photolysis	Computed offline with the solar-zenith angle and adjusted online by clouds	Offline computed clear sky rates based on the TUV radiative transfer model, online modified by cloud cover from the meteorological driver	Computed offline with the solar-zenith angle and adjusted online by clouds	Provided online based on clear-sky rates as function of solar-zenith angle and cloud shading factors
Biogenic emissions	Similar to Steinbrecher et al. (2009)	Based on Simpson et al. (1995, 1999)	Based on Steinbrecher et al. (2009)	Based on Guenther et al. (2006)
SOA formation	Not included here	SORGAM (Schell et al., 2001)	Based on Schrödner et al. (2014); Luttkus et al. (2022)	Based on Knote et al. (2014)
Sea salt emissions	Based on Monahan (1986); Mårtensson et al. (2003)	Based on Monahan (1986); Gong et al. (1997a, 1997b)	Based on Long et al. (2011); Sofiev et al. (2011)	Based on Monahan (1986); Gong et al. (1997a, 1997b); O'Dowd et al. (1997)

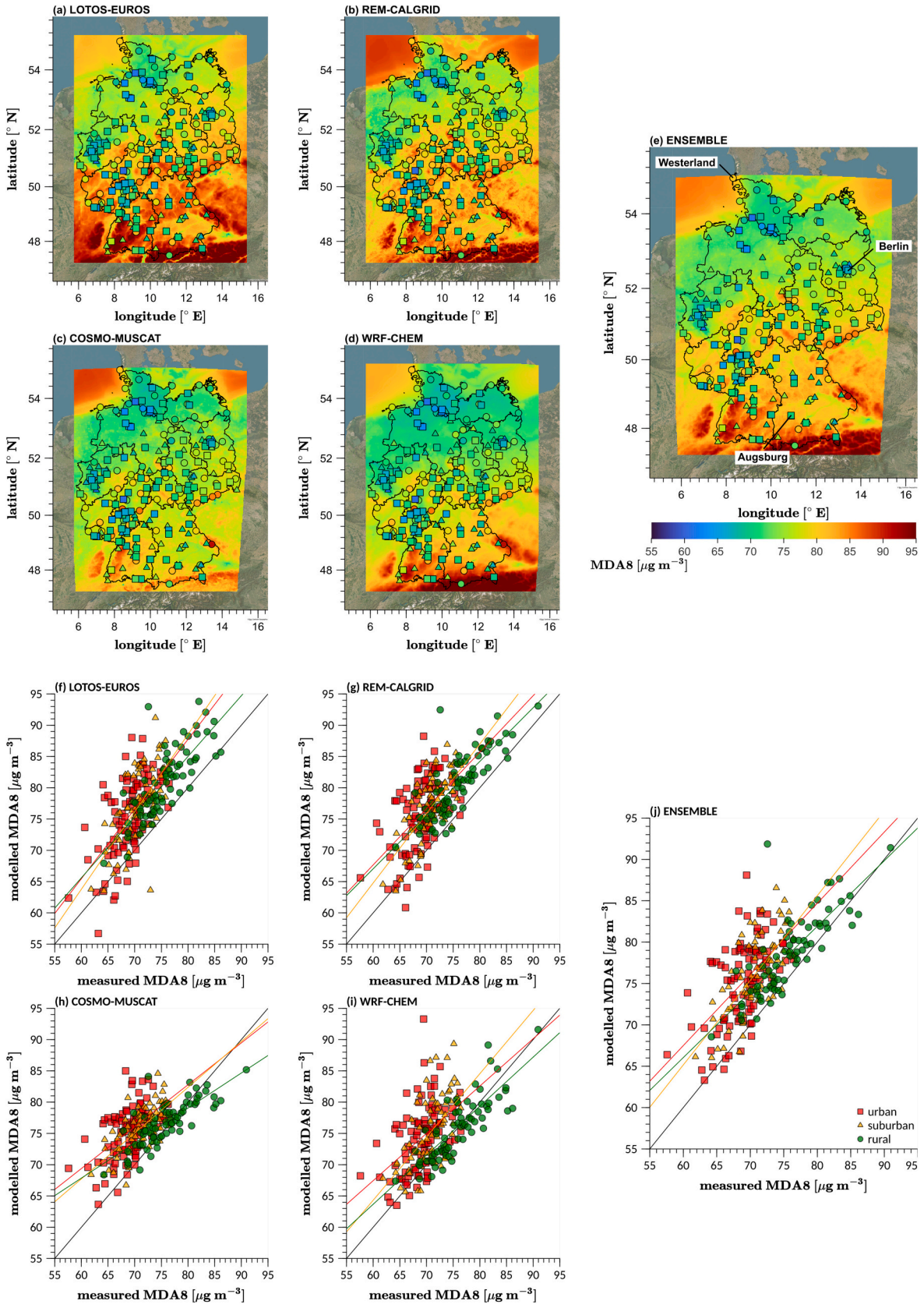


Fig. 1. Annual averaged MDA8 O_3 over Germany for LOTOS-EUROS (a), REM-CALGRID (b), COSMO-MUSCAT (c) and WRF-Chem (d). Measured concentrations of the urban (square), suburban (triangle) and rural (circle) background are shown on top. A regression analysis compares the modeled and the measured concentrations and is shown below the maps for each model (f-i). The ensemble mean of all models is shown in subplots (e) and (j).

Table 2
Breakdown of the GNFR sector F (GRETA) into the transport sub-categories F1 to F4 (CAMS-REG) [fraction of the total].

	F1 (exhaust-gasoline)	F2 (exhaust-diesel)	F3 (exhaust-LPG)	F4 (non-exhaust)
CO	0.75	0.24	0.01	0.00
CO_2	0.33	0.66	0.01	0.00
NH_3	0.74	0.20	0.05	0.00
NMVOC	0.58	0.16	0.03	0.23
NO_X	0.03	0.97	0.00	0.00
PM_{10}	0.05	0.20	0.00	0.75
PM_{25}	0.07	0.31	0.00	0.62
SO_2	0.30	0.70	0.00	0.00

Annual mean concentration maps were created to illustrate differences in the spatial distribution between all models. A regression analysis was performed to quantify the spatial correlation with measurements. Besides the evaluation of the spatial distribution, we generated time series for all stations from which we show selected sites in rural (Westerland: DEUB001), suburban (Augsburg: DEBY099) and urban (Berlin: DEBE034) locations as well as for all monitoring sites averaged. Time series of the daily mean and the mean values of the daily maximum 8-hour average (MDA8) for ozone provide more insight into the synoptic variability. Difference plots address the specific dissimilarities between the individual model simulations. By analyzing average weekly and daily variations, the meteorological impact can be largely removed, and the effect of anthropogenic emissions is emphasized. We therefore calculated average diurnal cycles for all days of the week. For ozone we distinguished between the growing season (April–September) and the winter period (January–March & October–December). We further subdivided the growing season into April–June and July–September, as it was anticipated that during the April–June period ozone concentrations are predominantly affected by long-range transport, while from July to September the local photochemical production plays a more important role (Otero et al., 2018; Butler et al., 2020).

We calculated an ensemble mean from the four models to verify possible benefits when using a simple poor man's ensemble approach. The performance between the individual model members can thus be compared against each other and to this ensemble mean. To evaluate the capability of the ensemble to represent different concentration regimes for ozone, we grouped the MDA8 O_3 into discrete bins based on concentration.

A wide range of statistical indicators were computed to benchmark the model performance. We used mean bias (MB), root mean squared error (RMSE), index of agreement (IOA) and correlation coefficient (R) (see supplementary material). Due to the large number of stations, we calculated average statistics over all sites. Exceedances of the calculated values of the 120 μg m^{-3} EU long-term target value for MDA8 O_3 have been investigated as well. All four models were evaluated on their ability to capture the observed threshold exceedances using time series and statistics. We scored the models on their rate of false alarms (FA), missed alarms (MA), good values below (GA^-), and good values above (GA^+) the threshold for MDA8 O_3 of 120 μg m^{-3}. In an ideal case, the number of false and missed alarms is small compared to the number of good values below and above the threshold. We provide the probability of detection (POD) and the success ratio (SR), that is "comparing the (correct) modeled alerts with the (observed) alerts and the (correct) modeled alerts with (all) alerts issued by the model", respectively (Janssen and Thunis, 2022).

We made use of the air quality modeling benchmarking indicators provided by the FAIRMODE initiative (Janssen and Thunis, 2022). In FAIRMODE, scientific model assessment methods are combined and harmonized using modeling quality indicators (MQIs) and modeling performance indicators (MPIs). All indicators are based on the uncertainty of measurements for each pollutant (RMS_U). Note, that the calculation of the RMS_U differs for time series and annual averaged values. The MQIs are calculated using the ratio between the model error and the RMS_U, scaled by a factor $\beta = 2$, so that the differences between modeled results and observations are allowed to be twice as large as the RMS_U. To emphasize further needs for model improvements we make use of the MPIs that are related to the temporal variation in terms of correlation (MPI_R), bias (MPI_{Bias}), standard deviation (MPI_σ) and high percentile values (MPI_{Perc}). In FAIRMODE the 92.9th percentile of the MDA8 O_3 is used to calculate the MPI_{Perc}. Similar to ozone, the MPI_{Perc} is assessed for nitrogen dioxide that equals the 19th occurrence in 8760 h (99.8th percentile). In FAIRMODE, defined modeling quality objectives (MQO) and modeling performance criteria (MPC) are being used to indicate the limits of applicability of a modeling approach. The MQO and the MQC are fulfilled when the 90th percentile value of corresponding MQI and MPI values are ≤ 1. For a more detailed description of all indicators, the objective and criteria we refer to Janssen and Thunis (2022). The calculation of all indicators applied in this study is given in the supplementary material.

We further assessed the quality of all four models and the ensemble using a dynamic (model) evaluation. The MDA8 for ozone has been used for this evaluation. We classified the prevailing weather conditions by categorizing the meteorological parameters into discrete bins. Similar straightforward methods of clustering meteorological conditions to determine the accumulation of air pollutants have also been applied in previous research activities (e.g., van Pinxteren et al., 2019; Thürkow et al., 2021). The MDA8 for ozone was assigned to each of these discrete classes. As a measure for the meteorological conditions, we selected the daytime maximum temperature (T_{max}) and the corresponding humidity at 2 m altitude ($RH@T_{max}$). We used the DWD COSMO-D2 forecasts as a proxy for observations, as they were applied as input dataset to all four individual models. For statistical robustness we require for each bin that the number count of the observations is larger than the number of stations being used. We further tried to quantify the relationship between temperature and MDA8 O_3 by a very basic calculation. For this, the average mean MDA8 O_3 values of the binned temperature clusters for ($T_{max} = 30$ to 32 °C) and ($T_{max} = 20$ to 22 °C) were subtracted from each other and divided by the 10 °C temperature difference between these bins. We performed the dynamic evaluation for urban and rural sites as well as for different periods of the year separately as described above.

3. Results and discussions

3.1. Spatial distribution

Fig. 1 shows maps of the annual averaged MDA8 O_3 over Germany, for all models and the measured background concentration. A regression analysis compares the modeled and the measured concentrations and is shown below the maps for each model. The ensemble mean is presented aside. In Fig. 2 the same information as in Fig. 1 is given for the nitrogen dioxide concentration.

The spatial ozone distribution calculated by all models shows a very similar pattern (Fig. 1). Largest concentrations in Germany are found in the rural background, including a clear signal of the orography. Here, we

Table 3
Height distribution [m] of area sources for GRETA (A: Public power; B: Industry; D: Fugitive; H: Aviation; J: Waste).

	0–20	20–92	92–184	184–324	324–522	522–781	781–1106
H	0.25	0.25	0.1	0.1	0.1	0.1	0.1
A, B, D, J	0.1	0.8	0.1	0	0	0	0

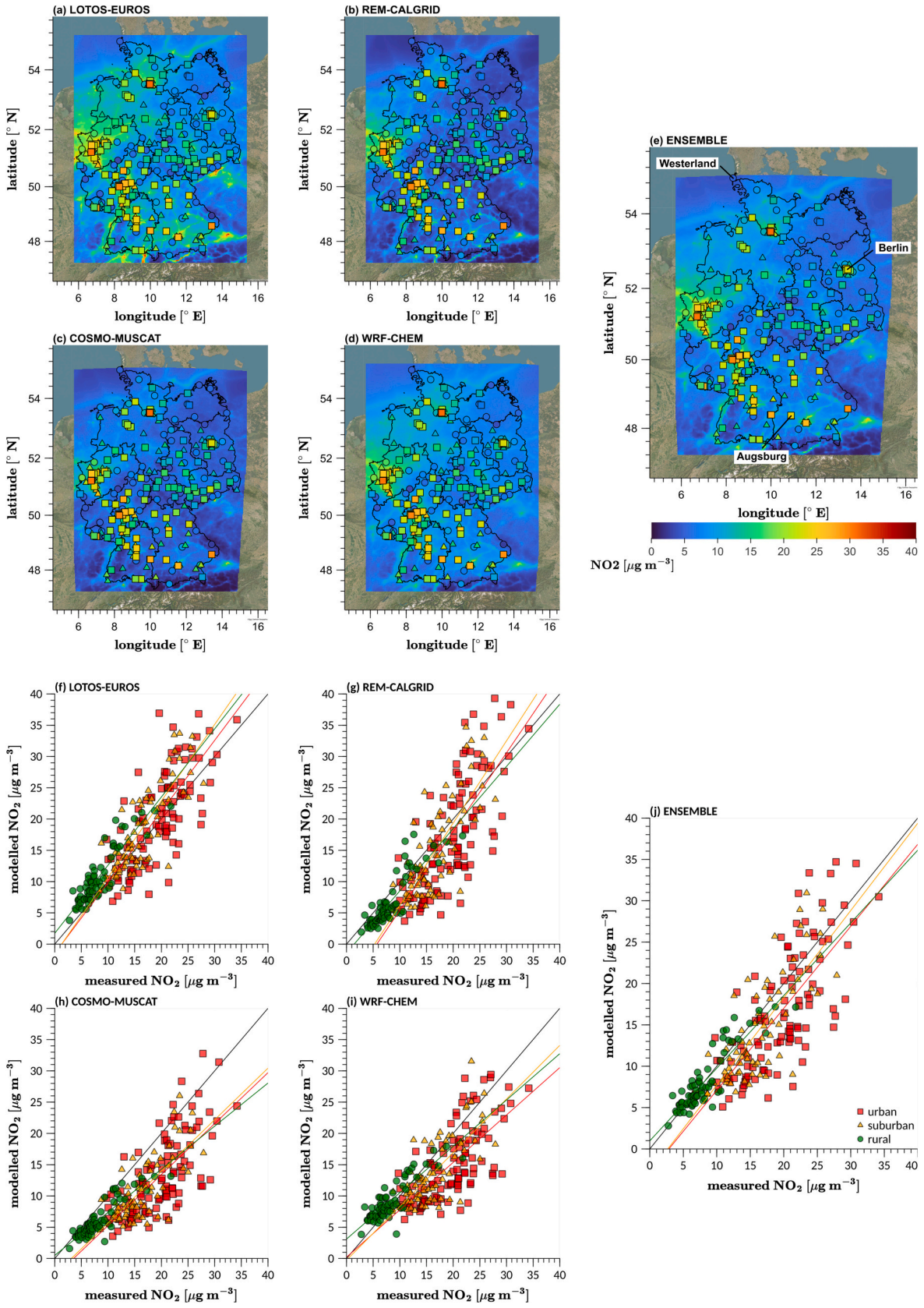


Fig. 2. Same as in Fig. 1 with information given for the nitrogen dioxide concentration.

found for all four models the well-known elevated ozone concentration pattern in the Alps. The lowest values are found in and near NO_X source regions, like the Main-Rhein area, the Ruhr area as well as the larger cities like Berlin, Hamburg, Bremen, and Munich. These minima are due to the well-known titration of ozone by nitrogen monoxide. All models show increased ozone levels above the sea, consistent with a smaller dry deposition sink in comparison to land areas.

The regression analysis for ozone summarizes the average model-measurement comparison indicating different fit slopes for the participating models (Fig. 1). All models overestimate the measured MDA8 O_3 levels on annual average in the German background for rural, suburban and urban areas. Systematic differences among the model participants become obvious in the rural background, over the large cities, the sea, and the Alps. The model members differ mainly in terms of the differentiation between urban and rural regions, with the largest absolute mean bias seen for urban areas. This corresponds with larger fraction errors in the urban background than in the rural background. Rural regions show an overall larger model-measurement agreement. In addition, the spatial pattern varies between the models, e.g., over the sea, at higher elevations, and with respect to the gradient between northern and southern Germany.

Modeled distributions for nitrogen dioxide clearly show the same pattern as the input emissions due to its short lifetime in the atmosphere (Fig. 2). All major cities and highways can be seen in the annual distribution. Most models show an average negative bias for the nitrogen dioxide background concentration. In the urban background larger variations can be recognized for all model participants compared to the rural area.

3.2. Temporal analysis

Fig. 3 shows daily time series of the MDA8 O_3, for the models and measured ozone as average at all background monitoring sites in Germany. The difference for each model against the measured concentration is presented in the lower panel. In Fig. 4 we classified the MDA8 for ozone for observed discrete bins of 20 μg m^{-3} to investigate how well the models can reproduce different ozone concentration regimes. Daily mean time series for three example sites (Westerland, Augsburg and Berlin) are presented in Fig. 5.

The time series (Fig. 3 & Fig. 5) show that the highest ozone levels in 2019 were measured for a few single days in June and two episodes in the second half of July and at the end of August. The dynamic range across the year of the simulated ozone concentration differs between the models, as the ordering of the model systems changes from season to season (Fig. 3 & Fig. 4). Fig. 4 shows that the MDA8 O_3 of about 80–120 μg m^{-3} in summer and about 40–80 μg m^{-3} in winter are captured well by the models. Observed concentration bins of the MDA8 for ozone are overestimated by the models below 80 μg m^{-3} in summer. For winter, the models are biased high for concentration bins of the MDA8 O_3 below 40 μg m^{-3} and biased low for MDA8 O_3 above 80 μg m^{-3}.

We also found that the models capture MDA8 O_3 values above 120 μg m^{-3} during summer and fall to a fairly large extent (Fig. 3 & Fig. 4).

Fig. 3. Time series (a) of the MDA8 O_3 for LOTOS-EUROS (yellow), REM-CALGRID (magenta), COSMO-MUSCAT (purple), WRF-Chem (blue) and the ensemble mean (green). Plotted values reflect the average background concentration of all available monitoring sites in Germany. Measured concentrations are indicated with black circles. The difference for each model and the ensemble mean to the measured concentration is shown in subplot (b). Exceedances above the MDA8 O_3 target value of 120 μg m^{-3} are indicated with red dots.

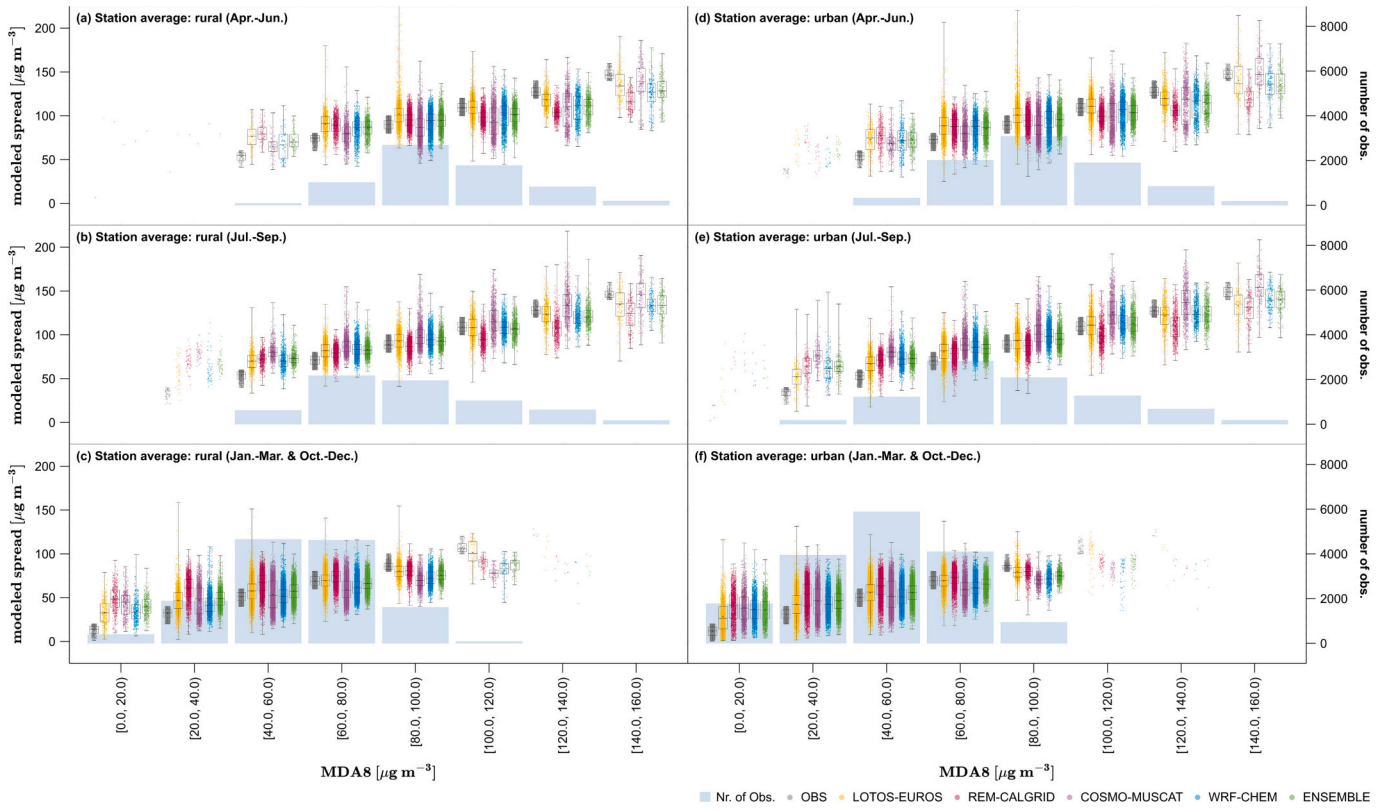


Fig. 4. Modeled ozone concentration (MDA8 O_3) in relation to the measured ozone levels for LOTOS-EUROS (yellow), REM-CALGRID (magenta), COSMO-MUSCAT (purple), WRF-Chem (blue) and the ensemble (green). Observations are shown in grey. Plotted values are binned by the measured background concentration levels of all available monitoring sites in Germany. Each single dot represents one observation (one day of the MDA8 O_3 in the corresponding bin from one station). The box plot shows the mean, median, 25th and 75th percentile as well as the min and max values. Different station classes are separated out for rural (a–c) and urban (d–f) areas. The periods of April–June and July–September are shown in the subplots (a,d) and (b,e) as well as the winter months January–March and October–December in subplots (c,f).

However, on average, the model ensemble underestimates the MDA8 for ozone above the 120 μg m^{-3} air quality target threshold. Especially, the observed high ozone levels during the end of April are not captured by any of the model participants (Fig. 3). Inspection of the individual example sites further reveals model specific features (Fig. 5). We found that all four models can represent the conditions present in coastal areas for example in the German location of Westerland. Ozone background concentrations for large agglomerations like in Berlin and for suburban areas in middle-sized cities represented by the station Augsburg can be captured as well.

As the MDA8 O_3 is thought to be rather insensitive to titration regimes the analysis of the diurnal behavior provides additional information on the ozone concentration itself. In Fig. 6 the day of the week cycles for modeled ozone and the measured levels are shown for rural and urban background sites and in Fig. 7 the same information is presented for nitrogen dioxide. All models reproduce the early morning ozone minimum at the same time and the increase of ozone levels at a similar rate afterwards (Fig. 6). We also found that the models capture the morning and the evening daytime maxima for nitrogen dioxide concentrations (Fig. 7). However, the timing of the daytime ozone maximum between the models can be slightly shifted. We also found that the modeled absolute concentrations largely differ between the model groups. All models hardly capture the low ozone levels at night in summer and winter. Especially in the urban background, the models do not decrease as far and overestimate the ozone levels up to 20 μg m^{-3} during the night. LOTOS-EUROS shows the deepest nighttime dips, which could be related to larger NO_2 concentrations at night.

We also found on average better model-measurement agreement during the growing season than in winter. Especially in July–September the results largely vary between the ensemble members, with the models sometimes showing better agreement and sometimes not. During the middle of the day, the nitrogen dioxide levels are often underestimated by all models. The models slightly tend to overestimate the afternoon maxima of NO_2 in rural areas and in the urban background at weekends. LOTOS-EUROS shows too pronounced morning daytime maxima for NO_2. In winter the treatment of stability and the titration of NO play a key role and still remain difficult. REM-CALGRID converts NO_2 cycles rather well except for rural background areas during the winter season, where the model substantially differs from the model ensemble.

Simulated ground-level ozone has already been shown to be sensitive to the model representation of the local production which can be influenced e.g., by vertical mixing processes in the free troposphere and the planetary boundary layer (Jang et al., 1995; Hogrefe et al., 2018). In addition, the ozone simulated by regional models can be affected by long-range transport and thus can be highly sensitive to the choice of chemical boundary conditions. The concentrations predicted for the outer domain act as boundary conditions and also impact the inflow of ozone through the edges of the inner model domain(s) (Colette et al., 2017; Im et al., 2018). The information on causes for (high) ozone episodes can be obtained from simulations that account for the attribution of different source sectors and source regions (e.g., Pay et al., 2019; Lupaşcu and Butler, 2019; Butler et al., 2020; Lupaşcu et al., 2022; Schaap et al., 2023) and could give further indication on the reasons why the model results largely vary for threshold exceedances of MDA8 O_3 > 120 μg m^{-3} in April compared to the observations.

Ozone exceedance events in the summertime are mainly affected by regional photochemistry (Schaap et al., 2023). Fig. 7 shows (on average) lower nitrogen dioxide concentrations in the observed afternoon rush-

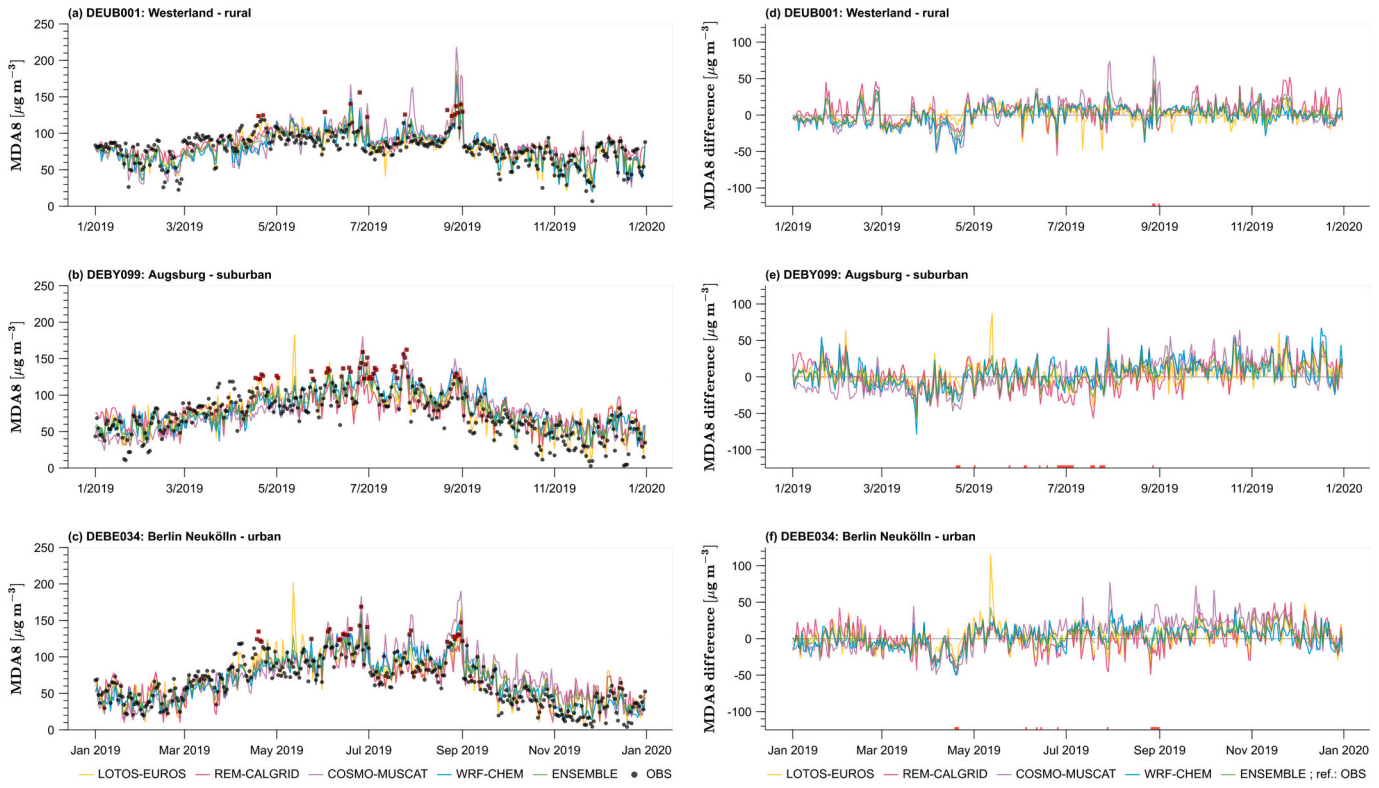


Fig. 5. Same as in Fig. 3 with the information of Westerland (a, d), Augsburg (b, e) and Berlin (c, f) instead of the station average over Germany. Time series of the MDA8 O_3 are presented in subplot (a–c). The differences for modeled and measured levels are subdivided in (d–f).

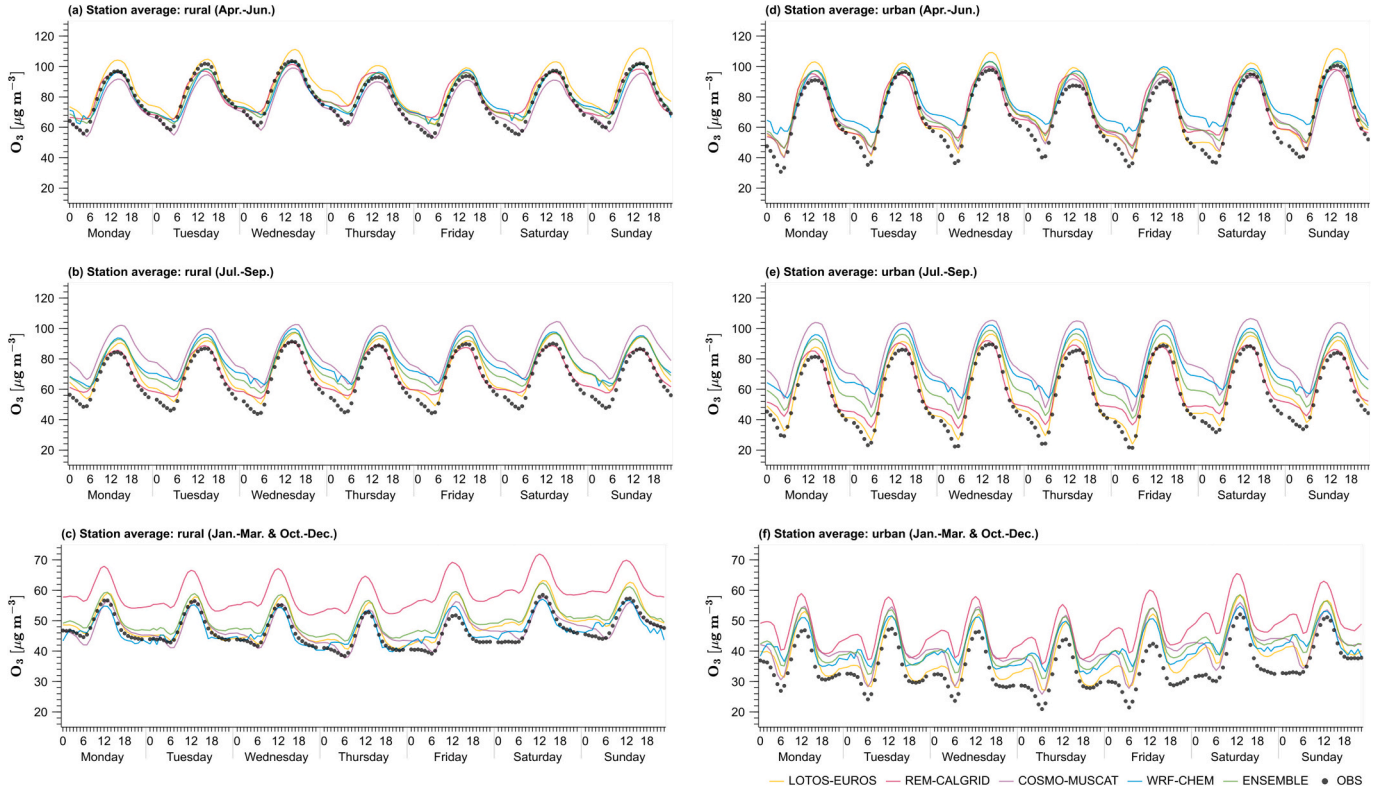


Fig. 6. Day of the week ozone concentration for LOTOS-EUROS (yellow), REM-CALGRID (magenta), COSMO-MUSCAT (purple), WRF-Chem (blue) and the ensemble mean (green). Plotted values reflect the average background concentration of all available monitoring sites in Germany. Measured concentrations are indicated with black circles. Different station classes are separated out for rural (a–c) and urban (d–f) areas. The periods of April–June and July–September are shown in the subplots (a, d) and (b, e) as well as the winter months January–March and October–December in subplots (c, f).

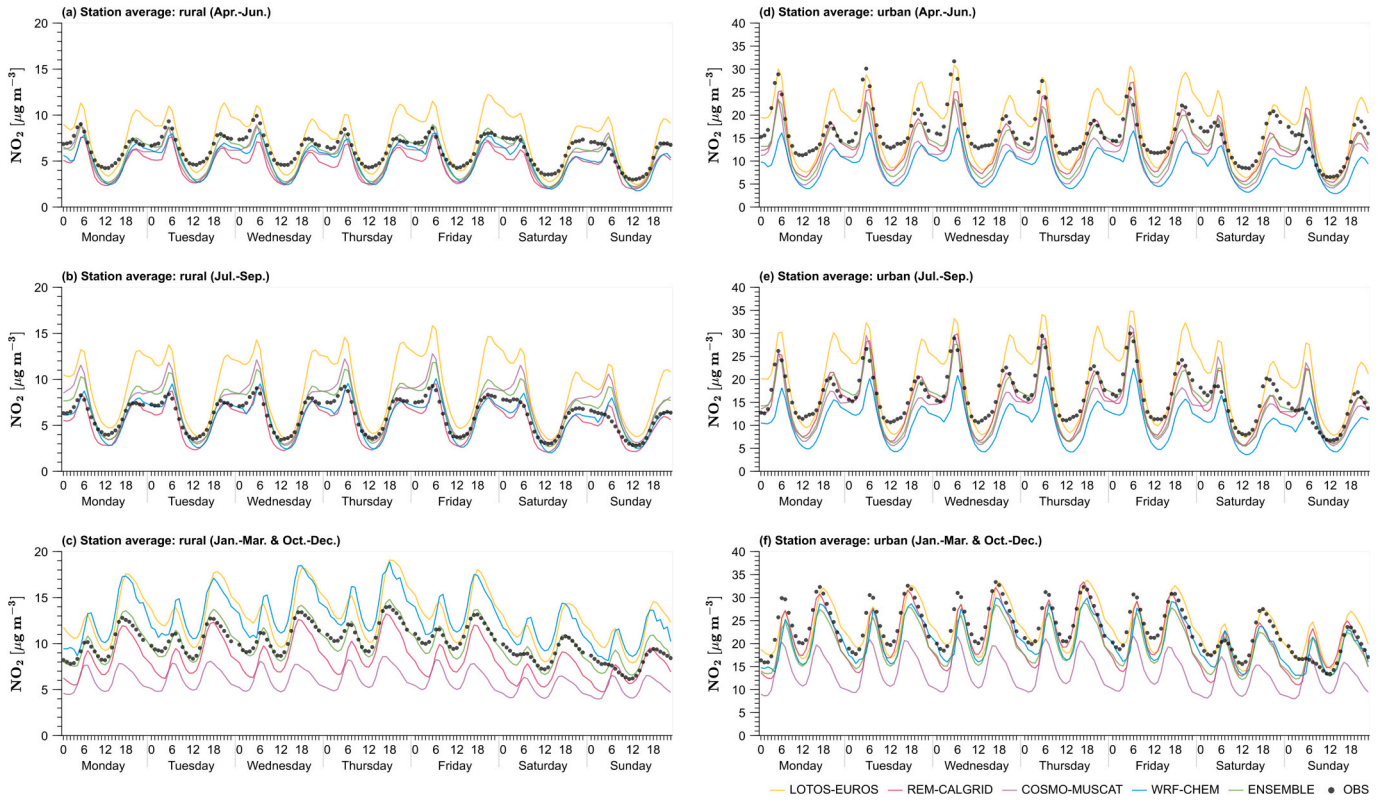


Fig. 7. Same as in Fig. 6 with information given for the nitrogen dioxide concentration.

hour peak (caused by lower emissions in urban areas) on weekends than during the week. This leads to a lowered titration on weekends. Accordingly, lower ozone concentrations can be expected during the rest of the week than on weekends (Koo et al., 2012) in many urban areas. In NO_X limited regions, such as in rural areas or some urban locations, the ozone level can also be lowered with decreased NO_X levels at daytime. This uncertainty to changes in NO_X emissions with respect to O_3 and their response to local photochemistry can be in principle captured by the models. However, addressing the weekend effect requires to analyze several years of data meaning that our single year simulation is not sufficiently long to quantify the effect. In addition, the emission information needs to be refined for this purpose as generic (monthly, weekly and daily) temporal profiles per sector have been used.

3.3. Model performance evaluation by comparison with measurements

In Fig. 8 the mean bias (MB), root mean squared error (RMSE), index of agreement (IOA) and temporal correlation (R) averaged over all German background sites are presented. The statistics for ozone and nitrogen dioxide concentrations were calculated for hourly and daily time series. For ozone, we also provide statistics of the MDA8 and the daytime maximum. Each indicator's value (lowest to largest skill) has been color coded (from dark to light red) to improve the readability and comparability.

The information presented in Fig. 8 illustrates the common behavior that chemistry transport models perform (slightly) better for daily than for hourly time series. Hourly ozone correlation coefficients for all models are between 0.68 and 0.80. Corresponding values of the daily mean time series are between 0.67 and 0.85. Higher correlation coefficients can be observed for the daytime maximum for ozone and the MDA8 O_3 ranging between 0.78–0.86 and 0.77–0.87, respectively. For the hourly time series, the models show a larger root mean squared error (20.4–25.4 μg m^{-3}) than calculated for the daily time series (14.4–21.3 μg m^{-3}) as expected. The root mean squared errors for the daytime maxima and the MDA8 O_3 are comparable to numbers we calculate for the daily time series (both about 16.3–20.6 μg m^{-3}). Modeled mean biases for the daily and hourly time series are comparable to each other (5.3–8.8 μg m^{-3}). The mean bias of the daytime maximum and the MDA8 for ozone (2.9–6.9 μg m^{-3}) are lower compared to the daily and hourly time series. For the index of agreement, we also calculate for hourly and daily time series quite similar numbers (0.78–0.89, each). The index of agreement for the daytime maximum and the MDA8 for ozone is about 0.83–0.91. The overall lowest model skills were calculated for MDA8 O_3 values above the target value of 120 μg m^{-3}. This can be shown for example for the correlation coefficients (0.53–0.61) and the index of agreement (0.56–0.72). We also found the largest spread between the models in terms of the mean error (−14.1 to 8.2 μg m^{-3}) and root mean squared error (12.6–19.6 μg m^{-3}) for MDA8 O_3 values above 120 μg m^{-3}. As shown in the previous section, the models tend to overestimate the observed ozone mass concentration at night. This reduces the model-measurement agreement of the ozone depletion at night and its production afterwards. The challenging representation of the nighttime titration can explain the lower model performance for the hourly ozone assessment compared to daily time series and the daytime maximum or MDA8 for ozone. The low skill of the models to represent MDA8 O_3 values above 120 μg m^{-3} illustrates the challenging nature of chemistry transport models to capture peak ozone concentrations or in fact episodes.

For the nitrogen dioxide concentration, the models show consistent results in the ranking of the model skills. For daily time series we calculate a mean absolute bias of 1.8 to 5.0 μg m^{-3}, a root mean squared error between 7.1 and 9.0 μg m^{-3}, an index of agreement of about 0.57–0.78 and 0.43–0.71 for the correlation coefficient. As expected, the performances for the hourly time series are lower compared to the daily time series (Fig. 8). LOTOS-EUROS is overestimating the measured NO_2 levels on average, whereas the other model members simulate lower nitrogen dioxide concentrations than observed.

The mean model-measurement agreement and skill for the calculated

(a) O_3

		MB	RMSE	IOA	R
Hourly	LOTOS-EUROS	5.37	20.43	0.88	0.80
	REM-CALGRID	8.83	22.90	0.81	0.74
	COSMO-MUSCAT	8.40	25.36	0.80	0.68
	WRF-CHEM	8.54	23.39	0.82	0.73
	ENSEMBLE	7.80	19.49	0.87	0.82
Daily	LOTOS-EUROS	5.26	14.41	0.89	0.85
	REM-CALGRID	8.77	17.93	0.79	0.76
	COSMO-MUSCAT	8.34	21.31	0.78	0.67
	WRF-CHEM	8.53	17.91	0.83	0.78
	ENSEMBLE	7.72	15.35	0.86	0.83
DailyMax	LOTOS-EUROS	6.85	17.74	0.90	0.85
	REM-CALGRID	5.91	19.57	0.83	0.81
	COSMO-MUSCAT	3.86	20.56	0.87	0.78
	WRF-CHEM	2.86	16.42	0.91	0.86
	ENSEMBLE	4.87	15.53	0.92	0.88
MDA8	LOTOS-EUROS	6.05	16.30	0.91	0.87
	REM-CALGRID	5.94	18.73	0.84	0.82
	COSMO-MUSCAT	4.31	20.64	0.87	0.77
	WRF-CHEM	3.50	16.44	0.91	0.85
	ENSEMBLE	4.95	15.13	0.92	0.88
MDA8>120	LOTOS-EUROS	-2.15	13.16	0.72	0.61
	REM-CALGRID	-14.14	19.58	0.56	0.53
	COSMO-MUSCAT	8.18	17.19	0.67	0.58
	WRF-CHEM	-3.84	12.64	0.71	0.60
	ENSEMBLE	0.21	11.33	0.74	0.61

(b) NO_2

		MB	RMSE	IOA	R
Hourly	LOTOS-EUROS	1.79	10.82	0.74	0.61
	REM-CALGRID	-1.84	10.94	0.70	0.59
	COSMO-MUSCAT	-4.94	11.54	0.60	0.44
	WRF-CHEM	-2.45	10.37	0.73	0.60
	ENSEMBLE	-1.86	9.29	0.75	0.66
Daily	LOTOS-EUROS	1.78	7.10	0.78	0.71
	REM-CALGRID	-1.86	7.71	0.72	0.69
	COSMO-MUSCAT	-4.96	8.97	0.57	0.43
	WRF-CHEM	-2.45	7.39	0.77	0.71
	ENSEMBLE	-1.87	6.61	0.77	0.74

Fig. 8. Statistics for ozone (a) and nitrogen dioxide (b). The calculated values of the mean bias (MB), root mean squared error (RMSE), index of agreement (IOA) and correlation coefficient (R) reflect the average over all available background sites. All statistics were calculated for hourly and daily input data. For ozone the MDA8 O_3, MDA8 $O_3 > 120$ µg m^{-3} and daytime maximum is shown. The model skill has been color coded (from dark to light red: lowest to highest skill) to improve the readability and comparability of each statistical index.

statistics is often highest for the model ensemble, calculated as a simple four model mean, for ozone as well as for nitrogen dioxide. Fig. 8 shows that the ensemble mean almost always ranks highest for correlation and index of agreement. Error statistics for mean bias and root mean squared error are also low and often outperform most of the individual ensemble members. Even the performance for MDA8 O_3 values above 120 µg m^{-3} is higher than for the single models.

3.4. Air quality model benchmarking following FAIRMODE

Fig. 9 summarizes the air quality model benchmarking following guidelines within FAIRMODE for ozone and nitrogen dioxide. We found that for ozone all four models fulfill the modeling quality objectives (MQOs) for annually averaged values and for time series of the MDA8, at all German background locations as well as in each subclass (MQI ≤ 1 for 90 % of analyzed stations; indicated with green boxes). Note that the MQIs in rural areas are substantially lower (yearly: 0.24–0.45; MDA8 O_3: 0.42–0.58) than in the urban background (yearly: 0.58–0.66; MDA8 O_3: 0.52–0.62). We found that the MQIs for the annual assessment are often larger than MQI values for time series of the MDA8 in urban areas. This can be simply explained as the measurement uncertainty for the annual assessment is smaller as opposed to time series of the MDA8. In contrast, for rural and suburban areas the MQIs for annual averaged values are lower for most of the model results.

Modeling performance indicators (MPIs) for ozone with respect to bias, correlation, standard deviation and high percentile values were also calculated. MPI values with respect to bias (MPI_{Bias}) for all German background sites are similar between all model systems (0.26–0.32). Values found in the urban background are larger (0.31–0.35) than shown for the rural area (0.12–0.24). Largest differences for MPIs of the ozone assessment between the models were calculated for the correlation, the standard deviation and the representation of high percentile values. We found that the models largely vary in terms of correlation for high ozone concentrations, illustrated by MPI_R values between 0.17 and 0.32. We also identified differences between the models for MPI_σ values of about 0.11–0.34, that reflects the lower season-to-season variation for the models applied.

Similar to ozone, we found for NO_2 that the annual averaged values of the MQI are larger in the urban background than in rural areas compared to input data for time series (for NO_2, hourly values). All four models as well fulfill the MQI for nitrogen dioxide of the hourly assessment for all German background locations and their split into subcategories for urban to rural sites. Except for LOTOS-EUROS, the MQI for annual averaged values of nitrogen dioxide is not fulfilled by the models in the urban background (MQI > 1; indicated with red boxes). However, the parameters for the calculation of the MQI on annual basis are still under discussion in FAIRMODE. Nevertheless, the MPIs for nitrogen dioxide can all be fulfilled for most of the applied models. COSMO-MUSCAT in addition misses the criteria for high percentile values for all background stations and for the urban and suburban regions.

In agreement to the performance evaluation presented in the previous section, the result of the ensemble mean is often more reliable than the single ensemble members and ranks lowest in terms of numbers for the MQIs and the MPIs for ozone as well as for nitrogen dioxide. All MQI

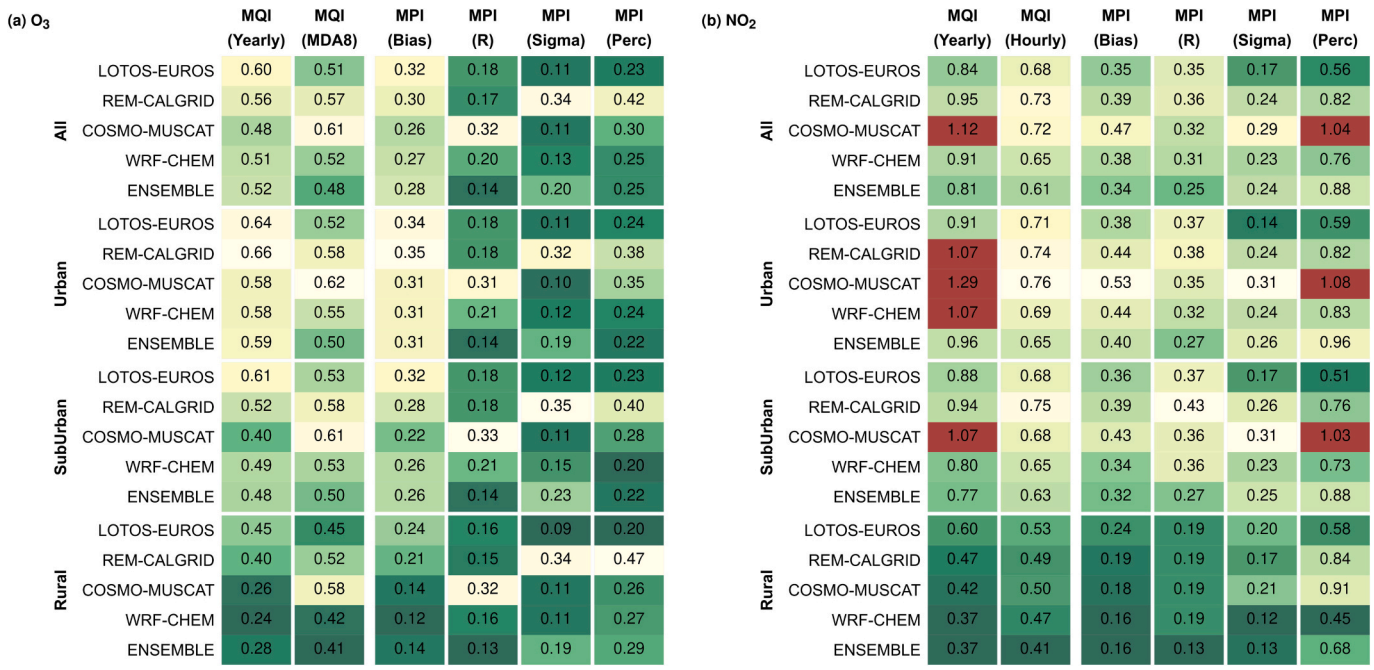


Fig. 9. Summary of indicators for air quality model benchmarking following the Air Quality Directive 2008/50/EC (AQD) from FAIRMODE. The modeling quality indicator (MQI) and the modeling performance indicators (MPIs) were calculated for ozone (a) and nitrogen dioxide (b). Statistics for the MQI were calculated separately, for yearly averaged model results (Yearly) as well as for hourly (Hourly, nitrogen dioxide) and the MDA8 O_3 (MDA8, ozone) input data. The MPIs are calculated in relation to the temporal variability and show modeled discrepancies for the bias (Bias), correlation (R), standard deviation (Sigma) and high percentile values (Perc). The indicators are calculated for all available background sites in Germany (All) or their selection of site classes (Urban, SubUrban, Rural). The model skill has been color coded (from light to dark green: lowest to highest skill) to improve the readability and comparability of each statistical index. Red color-coded boxes are used to indicate that the modeling quality objective (MQO) or modeling performance criteria (MPC) are not fulfilled.

Table 4
Threshold exceedance events of MDA8 $O_3 > 120$ µg m^{-3} as total of all background stations used, per model and station type (good values below the threshold: GA^-; good values above the threshold: GA^+; missed alarms: MA; false alarms: FA; probability of detection: POD; success ratio: SR).

Period	Site class	GA^-	GA^+	MA	FA	POD	SR
LOTOS-EUROS	All	77,992	3,438	2,266	2,608	0.60	0.57
	Rural	23,927	1,135	798	710	0.59	0.62
	Suburban	23,550	1,074	669	785	0.62	0.58
	Urban	30,515	1,229	799	1,113	0.61	0.52
REM-CALGRID	All	80,352	1,431	4,273	248	0.25	0.85
	Rural	24,580	396	1,537	57	0.20	0.87
	Suburban	24,284	445	1,298	51	0.26	0.90
	Urban	31,488	590	1,438	140	0.29	0.81
COSMO-MUSCAT	All	77,412	3,776	1,928	3,188	0.66	0.54
	Rural	23,856	1,201	732	781	0.62	0.61
	Suburban	23,354	1,156	587	981	0.66	0.54
	Urban	30,202	1,419	609	1,426	0.70	0.50
WRF-Chem	All	78,862	3,237	2,467	1,738	0.57	0.65
	Rural	24,314	962	971	323	0.50	0.75
	Suburban	23,773	1,050	693	562	0.60	0.65
	Urban	30,775	1,225	803	853	0.60	0.59
ENSEMBLE	All	79,551	2,906	2,798	1,049	0.51	0.73
	Rural	24,427	873	1,060	210	0.45	0.81
	Suburban	24,033	901	842	302	0.52	0.75
	Urban	31,091	1,132	896	537	0.56	0.68

and MPI values as calculated and presented for the ensemble mean are fulfilling the MQO and MPC, respectively.

Table 4 provides statistical information on exceedances for MDA8 O_3 above 120 µg m^{-3} using false alarms (FA), missed alarms (MA), good values above (GA^+) and below (GA^-) the threshold as well as the probability of detection (POD) and the success ratio (SR) for German background sites differentiated per station type and model. We found that on average all models lack performance capturing exceedances of MDA8 O_3 above 120 µg m^{-3}. The number of false (248–3,188) and missed alarms (1,928–4,273) largely varies between the models and are large compared to the number of good values above the threshold (1,431–3,776) for all German background sites. The averaged model's performance is expressed by both, the probability of detection (POD) and the success ratio (SR). For all German sites, we calculate numbers of the probability of detection ranging between 25 and 66 % and a success ratio of about 54 to 85 %. REM-CALGRID shows by far the largest SR (85 %), but also the lowest POD (25 %). The other three ensemble members show a very consistent picture in the ranking of the SR (54–65 %) and the POD (57–66 %). Statistics calculated for the ensemble mean are biased to the lowest model performance and thus show high values for

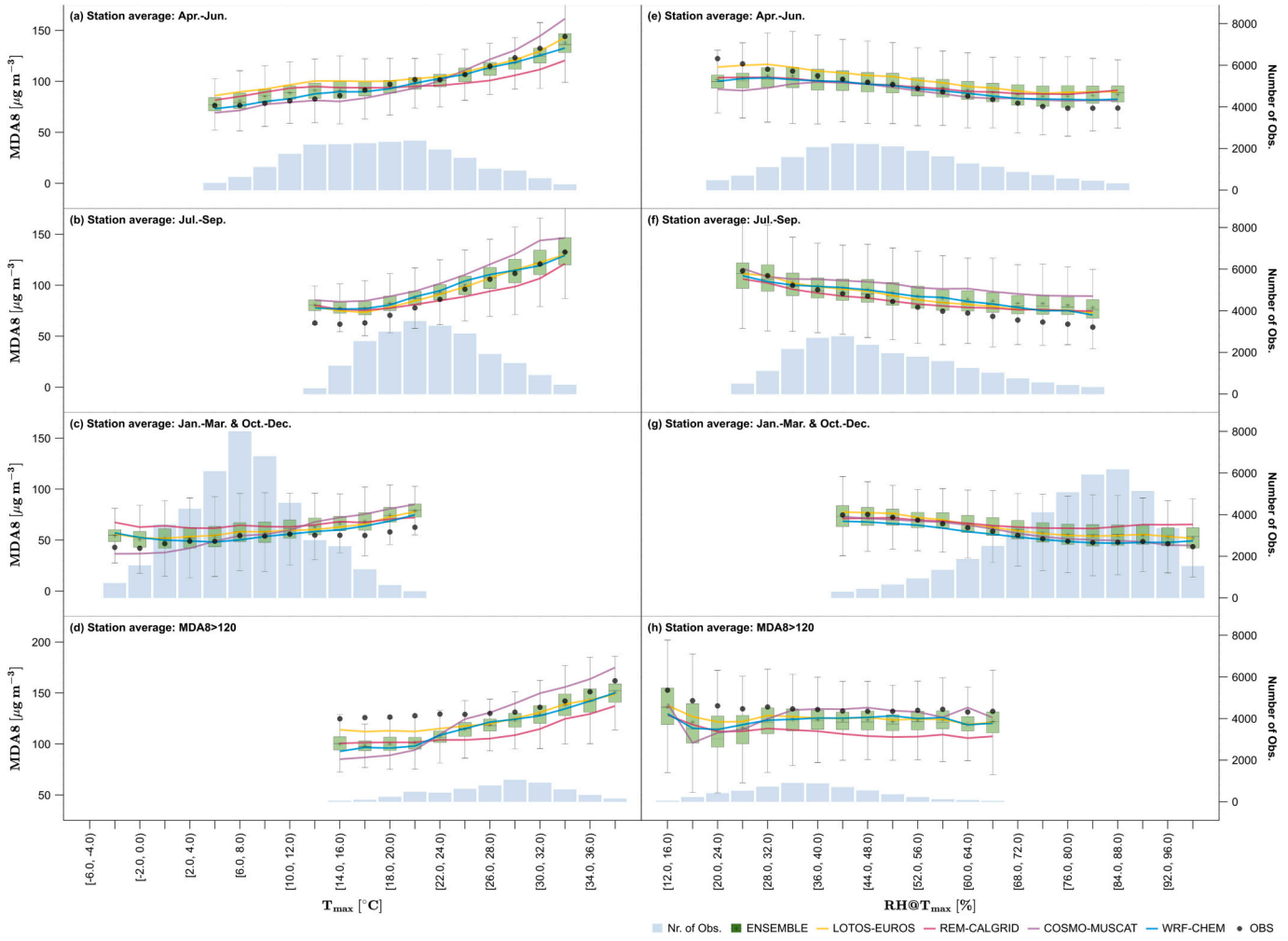


Fig. 10. Ozone concentration (MDA8 O_3) in relation to temperature (a–d) and humidity (e–h) for LOTOS-EUROS (yellow), REM-CALGRID (magenta), COSMO-MUSCAT (purple), WRF-Chem (blue) and the ensemble (boxplots). Plotted values reflect the average background concentration of all available monitoring sites in Germany. Measured concentrations are indicated with black circles. Box plots of the ensemble show the mean, median, 25th and 75th percentile as well as the two times standard deviation as min and max. Mean values are used for the ensemble members and the measurements. The periods of April–June and July–September are shown in the subplots (a, e) and (b, f) as well as the winter month January–March & October–December in subplots (c, g). The relation to temperature and humidity for exceedances above the MDA8 O_3 target value of 120 µg m^{-3} are shown in subplots (d) and (h).

the success ratio (73 %) but a low probability of detection (51 %) for all German background sites. We also noticed that the values of the POD are larger in the urban background than in rural areas. Vice versa, the SR is larger in the rural background.

In FAIRMODE it has already been stated that large inconsistencies may occur for different indicators depending on whether time series (daily/hourly) or annual averaged values are included in the analysis (Monteiro et al., 2018). Following Monteiro et al. (2018), the MQO for hourly or daily time series is often attained, whereas it is not the case for annual values. We could show that all models fulfill the quality standards set by the FAIRMODE guideline values for annual averages as well as for time series in the background locations, whereas for specific regions (e.g., rural vs. urban areas) the model performance can differ. The capability to reproduce extreme events in model intercomparison and evaluation studies is often ignored as many times only averaged values were obtained or only high percentile values were investigated (Monteiro et al., 2018). We therefore also focused on threshold exceedances as recommended by FAIRMODE and which are typically highly coherent in time and space over scales of hundreds of kilometers (Schnell et al., 2015; Carro-Calvo et al., 2017). Regional models are most suited to simulate human exposure to ozone events when the grid resolution is considering the spatial extent of the urban background in the region of interest (Kuik et al., 2016), while for simulation of large-scale ozone exceedance events in the rural background in Europe, grid resolutions between 10 and 20 km have been shown to be optimal (Schaap et al., 2015). However, in our study we could show that the representation of ozone threshold exceedance events in urban areas for MDA8 $O_3 > 120$ µg m^{-3} is limited even on a 2 × 2 km^2 scale.

3.5. Dynamic (ozone) model evaluation

As introduced, ozone is highly correlated to temperature which is an important driver especially during ozone episodes. In Fig. 10 we show the common relation that higher ozone concentrations occur more often during warmer weather conditions, while at colder temperatures the ozone concentrations more likely remain low. Due to the physical relation between the daytime maxima of temperature and their corresponding humidity, lower levels of the relative humidity are more often connected to higher ozone concentrations and vice versa.

Fig. 10 shows that the ensemble mean reports a good agreement of the measured concentration in relation to temperature for April–June. Largest differences to the observed ozone-temperature dependence were found for July–September. In July–September, the measured ozone levels for temperatures lower than ~18 °C are biased high for the

Table 5
Dependence of ozone concentration on temperature, for ~21 °C to ~31 °C [μg m^{-3} $°C^{-1}$].

	Site class	Obs.	LOTOS-EUROS	REM-CALGRID	COSMO-MUSCAT	WRF-Chem	ENSEMBLE
April–June	All	3.06	2.71	1.65	5.10	2.72	2.86
	Rural	3.60	3.08	1.82	5.14	2.80	3.06
	Urban	2.88	2.47	1.43	5.05	2.62	2.68
July–September	All	4.29	3.77	2.56	4.98	3.06	3.45
	Rural	4.13	3.38	2.44	4.78	2.80	3.17
	Urban	4.47	4.06	2.66	5.22	3.34	3.68
MDA8 O_3 > 120 μg m^{-3}	All	0.82	1.87	1.30	5.57	2.97	2.82
	Rural	1.08	1.97	1.32	4.77	2.61	2.66
	Urban	0.69	1.77	1.25	6.03	3.23	2.90

ensemble mean and for all models. In winter, all models and the ensemble mean are biased high for temperatures higher than ~14 °C. Differences between urban and rural sites are on average smaller compared to seasonal deviations (see supplementary material).

However, we also observed a large spread for all models in their ozone sensitivity to different temperature regimes. The ozone temperature dependence for temperatures higher than ~26 °C in April–September has been well captured by WRF-Chem and LOTOS-EUROS. For COSMO-MUSCAT and WRF-Chem we observed good agreement capturing the ozone sensitivity to temperature for temperatures lower than ~18 °C in April–June. In winter, COSMO-MUSCAT is in good agreement with observed concentrations for temperatures lower than ~14 °C. On the other hand, COSMO-MUSCAT shows on average larger ozone sensitivity to temperature for warmer conditions in spring- and summertime. Vice versa, the lowest sensitivity of ozone to temperature was found for REM-CALGRID. This too high and too low sensitivity to temperature observed for COSMO-MUSCAT and REM-CALGRID causes an overestimation (COSMO-MUSCAT) and underestimation (REM-CALGRID) of the ozone levels for temperatures higher than ~26 °C in spring- and summertime. For temperatures lower than ~14 °C and ~18 °C in April–June, we observed slightly biased high ozone levels for REM-CALGRID and for LOTOS-EUROS. For colder conditions in winter, we also observed high biased ozone concentrations for LOTOS-EUROS, WRF-Chem and REM-CALGRID.

The observed timeseries show values of MDA8 O_3 just over 120 μg m^{-3} already in the temperature range from ~15 to ~24 °C which are systematically underestimated by all models (Fig. 10). LOTOS-EUROS comes closest to covering MDA8 O_3 > 120 μg m^{-3} for temperatures lower than ~24 °C. REM-CALGRID captures the ozone-temperature dependence for MDA8 O_3 > 120 μg m^{-3} to a fairly large extent but is on average biased low as well. The models show a similar ozone temperature-dependence for temperatures higher than ~28 °C. However, temperatures at which the MDA8 O_3 typically starts to rise largely differ between the models (~22–26 °C) and the observations (~30 °C). COSMO-MUSCAT shows the most pronounced ozone temperature sensitivity and therefore the largest underestimation for temperatures lower than ~24 °C and high biased MDA8 O_3 > 120 μg m^{-3} for temperatures larger than ~30 °C.

In Table 5 we present the temperature increase per Celsius between ~21 °C and ~31 °C. The ensemble mean and nearly all models underestimate the observed ozone dependence to temperature in April–June and July–September for all locations. Nevertheless, the higher temperature sensitivity in July–September (OBS: 4.3 μg m^{-3} $°C^{-1}$; ENS_{ALL}: 3.5 μg m^{-3} $°C^{-1}$) compared with April to June (OBS: 3.1 μg m^{-3} $°C^{-1}$; ENS_{ALL}: 2.9 μg m^{-3} $°C^{-1}$) is captured by most of the models. All models also capture the observed lower ozone sensitivity to temperature for urban areas compared to rural stations in April–June and vice versa in July–September (see Table 5). For MDA8 O_3 > 120 μg m^{-3}, all models show a too large ozone dependence on temperature (OBS_{ALL}: 0.8 μg m^{-3} $°C^{-1}$; ENS_{ALL}: 2.8 μg m^{-3} $°C^{-1}$), where REM-CALGRID comes closest covering the temperature increase per 10 °C for MDA8 O_3 > 120 μg m^{-3}.

In Fig. 10 we show the sensitivity to the relative humidity and observed consistent results with an overall good model measurement agreement for all models. The models slightly overestimate the ozone concentrations for humid conditions equal to a relative humidity larger than ~68 % in April–June and larger than ~64 % in July–September. Largest high biased concentrations for humid conditions have been found for COSMO-MUSCAT in July–September and for REM-CALGRID in winter. Except for LOTOS-EUROS, in April–June the models are also biased low for a relative humidity lower than ~28 %. MDA8 O_3 values above 120 μg m^{-3} are biased low as well for a relative humidity lower than ~28 % for all models. For REM-CALGRID a good agreement to the ozone-humidity dependence is observed. However, similar to results shown before for the ozone sensitivity to temperature, a too low modeled concentration for MDA8 O_3 > 120 μg m^{-3} is observed.

In Otero et al. (2018) the influence of meteorology on ozone has been calculated for a model ensemble and the observations using a multiple linear regression analysis on a multi-decadal time scale. The study showed that the models can reproduce the temperature (T_{max}) response to a fairly large extent. We precise this finding and argue that the observed ozone-temperature dependence shows highest agreement for warmer conditions, but the models can also largely vary in the modeled ozone sensitivity to temperature. In contrast to our results, Otero et al. (2018) identified a better model-measurement agreement for summer compared to spring. We found large season-to-season differences and high biased ozone levels for July–September that seem to be connected with moderate temperatures and thus a higher relative humidity during the summertime, especially at night. Following Otero et al. (2018) the relation between humidity and ozone is more difficult to capture. We found an overall good agreement of the modeled ozone-humidity dependence. However, the models largely underestimate high ozone concentrations (MDA8 O_3 > 120 μg m^{-3}) for dry conditions during springtime. We speculate that this could be connected with an insufficient representation of modeled deposition velocities covering vegetative water stress and low sensitivity of NO_X and BVOC emissions (e.g., Churkina et al., 2017; Lin et al., 2020).

4. Summary & conclusion

We have successfully compared and evaluated the performance of four regional chemistry transport models and their ensemble. Our work shares similar features with previous multi-model intercomparison studies for ozone but focuses on a higher grid resolution and performs a national scale quality assessment, which might be more relevant with a future revision of the European air quality directive.

The results demonstrate that all models satisfy the modeling quality objectives and criteria (MQO/MQC) set by the FAIRMODE initiative for every modeling quality and performance indicator (MQI/MPI) for annual averaged ozone concentrations as well as for time series of the MDA8 O_3. All models showed good performance of the operational evaluation using standard statistical indicators such as bias, correlation or for high percentile values. The analysis of the model skills demonstrates better performance in rural areas than in the urban background and for springtime compared to summertime. We also showed an improved performance for the ensemble over the individual model participants. Despite the high model-measurement agreement, the

individual models and their ensemble show room for improvement at simulating threshold exceedances for MDA8 O_3 above 120 μg m^{-3}. Here, we compute high values for missed alarms and false alarms. We also found rather low model measurement agreement for observed lower ozone concentrations at night, which is likely related to difficulties in simulating the stable nocturnal boundary layer.

A novel feature of this study is the dynamic evaluation of modeled ozone with respect to temperature. In general, the models correctly reproduced the observed higher sensitivity of ozone with respect to temperature in summer than in springtime, which is linked to overall higher local photochemical production of ozone in summer. The models also correctly captured the differences in the temperature sensitivity of ozone between rural and urban areas in both spring and summer seasons, with a lower temperature sensitivity in urban areas in springtime and a higher sensitivity in summertime, showing that the 2 × 2 km^2 model resolution used in this study is adequate for simulating the dynamic response of ozone to temperature in urban background areas. Despite capturing the observed spring to summer and urban to rural differences in the temperature sensitivity of ozone, there was still a large spread in the modeled temperature sensitivities. The sensitivity of ozone to temperature also depends on the ratio of NO_X and BVOC. For example, in NO_X rich urbanized regions the temperature-sensitivity to ozone is higher than in NO_X limited regions (Otero et al., 2021). Future work should focus on the representation of temperature sensitive processes in models, such as BVOC and nitrogen oxide emissions.

In general, the models underestimated the observed temperature sensitivity of ozone in summer, despite showing high bias in simulated ozone concentrations compared with the measurements. Both, the high bias and the underestimated temperature sensitivity can be linked to a systematic overestimation of ozone concentrations in the 60–80 μg m^{-3} range, which make up most of the observed ozone concentrations in summer and tend to occur at moderate temperatures. The models also show a high bias with respect to observations in the same concentration range in spring. However, this is not reflected in the overall springtime model bias due to a systematic underestimation of higher values (over 100 μg m^{-3}) in spring, which tend to occur at relatively low temperatures (below 25 °C) compared with similar exceedance events in summer. The missed springtime exceedance events and the general overestimation of concentrations in the 60–80 μg m^{-3} range could be linked to deficiencies in the representation of background ozone, including long-range transport, which can be further addressed with source apportionment techniques and validation at the European scale. An important caveat in this context is that these results are based only on observations and model simulations for the year of 2019. Future work should focus on longer periods to determine whether these results can be generalized to longer timeframes.

In this study we focused on an operational and dynamic evaluation of four chemistry transport models for ozone in Germany. The dynamic evaluation would benefit from an extension to a decadal time scale as shown e.g., by Colette et al. (2017) and Banzhaf et al. (2015). We used a poor man's ensemble to assess the added value of the combined information in the ensemble. The evaluation of larger ensembles of chemistry transport models is currently taking place within the Copernicus Atmosphere Monitoring Service, enabling it to address additional ensemble properties which are not covered here. An important next step for the work presented here is to diagnose the reasons for the mismatches with observations and differences between the models. Such diagnostic evaluation requires investigating the process descriptions in more detail. As the uncertainties related to the biogenic emissions and the dry deposition parameterizations are large (e.g., Im et al., 2015b), we recommend to focus on the representation of these process descriptions by evaluation of dedicated simulations for research campaigns providing the relevant process information.

CRediT authorship contribution statement

M. Thürkow: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization **T. Butler:** Conceptualization, Methodology, Writing - Original Draft, Supervision, Project administration, Funding acquisition **M. Schaap:** Conceptualization, Methodology, Writing - Original Draft, Supervision **All Authors:** Resources, Writing - Review & Editing. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was funded by the German Environment Agency (Umweltbundesamt - UBA), Dessau-Roßlau within the framework of the "OzonEval" research initiative (project number: FKZ 3720 51 201 0). Thanks to Hanna Thürkow for proofreading and the final assistance when submitting the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.167665.

References

Atkinson, R., 2007. Rate constants for the atmospheric reactions of alkoxy radicals: an updated estimation method. Atmos. Environ. 41 https://doi.org/10.1016/j.atmosenv.2007.07.002.

Banzhaf, S., Schaap, M., Kerschbaumer, A., Reimer, E., Stern, R., van der Swaluw, E., Builtjes, P., 2012. Implementation and evaluation of pH-dependent cloud chemistry and wet deposition in the chemical transport model REM-Calgrid. Atmos. Environ. 49, 378–390. https://doi.org/10.1016/j.atmosenv.2011.10.069.

Banzhaf, S., Schaap, M., Kranenburg, R., Manders, A.M.M., Segers, A.J., Visschedijk, A.J.H., Denier van der Gon, H.A.C., Kuenen, J.J.P., van Meijgaard, E., van Ulft, L.H., Cofala, J., Builtjes, P.J.H., 2015. Dynamic model evaluation for secondary inorganic aerosol and its precursors over Europe between 1990 and 2009. Geosci. Model Dev. 8, 1047–1070. https://doi.org/10.5194/gmd-8-1047-2015.

Bessagnet, B., Pirovano, G., Mircea, M., Cuvelier, C., Aulinger, A., Calori, G., Ciarelli, G., Manders, A., Stern, R., Tsyro, S., García Vivanco, M., Thunis, P., Pay, M.-T., Colette, A., Couvidat, F., Meleux, F., Rouïl, L., Ung, A., Aksoyoglu, S., Baldasano, J. M., Bieser, J., Briganti, G., Cappelletti, A., D'Isidoro, M., Finardi, S., Kranenburg, R., Silibello, C., Carnevale, C., Aas, W., Dupont, J.-C., Fagerli, H., Gonzalez, L., Menut, L., Prévôt, A.S.H., Roberts, P., White, L., 2016. Presentation of the EURODELTA III intercomparison exercise – evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology. Atmos. Chem. Phys. 16 https://doi.org/10.5194/acp-16-12667-2016.

Black, E., Blackburn, M., Harrison, G., Hoskins, B., Methven, J., 2004. Factors contributing to the summer 2003 European heatwave. Weather 59. https://doi.org/10.1256/wea.74.04.

Butler, T., Lupascu, A., Nalam, A., 2020. Attribution of ground-level ozone to anthropogenic and natural sources of nitrogen oxides and reactive carbon in a global chemical transport model. Atmos. Chem. Phys. 20, 10707–10731. https://doi.org/10.5194/acp-20-10707-2020.

Carro-Calvo, L., Ordóñez, C., García-Herrera, R., Schnell, J.L., 2017. Spatial clustering and meteorological drivers of summer ozone in Europe. Atmos. Environ. 167, 496–510. https://doi.org/10.1016/j.atmosenv.2017.08.050.

Chen, L., Gao, Y., Zhang, M., Fu, J.S., Zhu, J., Liao, H., Li, J., Huang, K., Ge, B., Wang, X., Lam, Y.F., Lin, C.-Y., Itahashi, S., Nagashima, T., Kajino, M., Yamaji, K., Wang, Z., Kurokawa, J., 2019. MICS-Asia III: multi-model comparison and evaluation of aerosol over East Asia. Atmos. Chem. Phys. 19, 11911–11937. https://doi.org/10.5194/acp-19-11911-2019.

Churkina, G., Kuik, F., Bonn, B., Lauer, A., Grote, R., Tomiak, K., Butler, T.M., 2017. Effect of VOC emissions from vegetation on air quality in Berlin during a heatwave. Environ. Sci. Technol. 51 https://doi.org/10.1021/acs.est.6b06514.

Coates, J., Mar, K.A., Ojha, N., Butler, T.M., 2016. The influence of temperature on ozone production under varying NO_x conditions – a modelling study. Atmos. Chem. Phys. 16 https://doi.org/10.5194/acp-16-11601-2016.

Colette, A., Andersson, C., Manders, A., Mar, K., Mircea, M., Pay, M.-T., Raffort, V., Tsyro, S., Cuvelier, C., Adani, M., Bessagnet, B., Bergström, R., Briganti, G., Butler, T., Cappelletti, A., Couvidat, F., D'Isidoro, M., Doumbia, T., Fagerli, H., Granier, C., Heyes, C., Klimont, Z., Ojha, N., Otero, N., Schaap, M., Sindelarova, K., Stegehuis, A.I., Roustan, Y., Vautard, R., van Meijgaard, E., Vivanco, M.G., Wind, P., 2017. EURODELTA-trends, a multi-model experiment of air quality hindcast in Europe over 1990–2010. Geosci. Model Dev. 10, 3255–3276. https://doi.org/10.5194/gmd-10-3255-2017.

Crutzen, P., 1973. A discussion of the chemistry of some minor constituents in the stratosphere and troposphere. Pure Appl. Geophys. 106–108, 1385–1399. https://doi.org/10.1007/BF00881092.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S.T., Scheffe, R., Schere, K., Steyn, D., Venkatram, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. Environ. Fluid Mech. 10 https://doi.org/10.1007/s10652-009-9163-2.

Derwent, R.G., Manning, A.J., Simmonds, P.G., Spain, T.G., O'Doherty, S., 2018. Long-term trends in ozone in baseline and European regionally-polluted air at Mace Head, Ireland over a 30-year period. Atmos. Environ. 179, 279–287. https://doi.org/10.1016/j.atmosenv.2018.02.024.

Dhomse, S.S., Kinnison, D., Chipperfield, M.P., Salawitch, R.J., Cionni, I., Hegglin, M.I., Abraham, N.L., Akiyoshi, H., Archibald, A.T., Bednarz, E.M., Bekki, S., Braesicke, P., Butchart, N., Dameris, M., Deushi, M., Frith, S., Hardiman, S.C., Hassler, B., Horowitz, L.W., Hu, R.M., Jöckel, P., Josse, B., Kirner, O., Kremser, S., Langematz, U., Lewis, J., Marchand, M., Lin, M., Mancini, E., Marécal, V., Michou, M., Morgenstern, O., O'Connor, F.M., Oman, L., Pitari, G., Plummer, D.A., Pyle, J.A., Revell, L.E., Rozanov, E., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tilmes, S., Visioni, D., Yamashita, Y., Zeng, G., 2018. Estimates of ozone return dates from chemistry-climate model initiative simulations. Atmos. Chem. Phys. 18 https://doi.org/10.5194/acp-18-8409-2018.

EEA, 2018. Air Quality in Europe - 2018 Report. Eur. Environ. Agency Rep. No 12/2018.

EEA, 2021. CORINE Land Cover - User Manual. Copernicus L. Monit. Serv. 1.0.

Emmons, L.K., Walters, S., Hess, P.G., Lamarque, J.-F., Pfister, G.G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S.L., Kloster, S., 2010. Description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). Geosci. Model Dev. 3, 43–67. https://doi.org/10.5194/gmd-3-43-2010.

Escudero, M., Segers, A., Kranenburg, R., Querol, X., Alastuey, A., Borge, R., de la Paz, D., Gangoiti, G., Schaap, M., 2019. Analysis of summer O_3 in the Madrid air basin with the LOTOS-EUROS chemical transport model. Atmos. Chem. Phys. 19, 14211–14232. https://doi.org/10.5194/acp-19-14211-2019.

Fast, J.D., Gustafson Jr., W.I., Easter, R.C., Zaveri, R.A., Barnard, J.C., Chapman, E.G., Grell, G.A., Peckham, S.E., 2006. Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model. J. Geophys. Res. Atmos. 111 https://doi.org/10.1029/2005JD006721.

Fischer, E.V., Jacob, D.J., Yantosca, R.M., Sulprizio, M.P., Millet, D.B., Mao, J., Paulot, F., Singh, H.B., Roiger, A., Ries, L., Talbot, R.W., Dzepina, K., Pandey Deolal, S., 2014. Atmospheric peroxyacetyl nitrate (PAN): a global budget and source attribution. Atmos. Chem. Phys. 14, 2679–2698. https://doi.org/10.5194/acp-14-2679-2014.

Flemming, J., Stern, R., 2007. Testing model accuracy measures according to the EU directives—examples using the chemical transport model REM-CALGRID. Atmos. Environ. 41, 9206–9216. https://doi.org/10.1016/j.atmosenv.2007.07.050.

Flemming, J., Benedetti, A., Inness, A., Engelen, J.R., Jones, L., Huijnen, V., Remy, S., Parrington, M., Suttie, M., Bozzo, A., Peuch, V.H., Akritidis, D., Katragkou, E., 2017. The CAMS interim reanalysis of carbon monoxide, ozone and aerosol for 2003-2015. Atmos. Chem. Phys. 17 https://doi.org/10.5194/acp-17-1945-2017.

Foley, Kristen M., Dolwick, P., Hogrefe, C., Simon, H., Timin, B., Possiel, N., 2015a. Dynamic evaluation of CMAQ part II: evaluation of relative response factor metrics for ozone attainment demonstrations. Atmos. Environ. 103 https://doi.org/10.1016/j.atmosenv.2014.12.039.

Foley, Kristen M., Hogrefe, C., Pouliot, G., Possiel, N., Roselle, S.J., Simon, H., Timin, B., 2015b. Dynamic evaluation of CMAQ part I: separating the effects of changing emissions and changing meteorology on ozone levels between 2002 and 2005 in the eastern US. Atmos. Environ. 103, 247–255. https://doi.org/10.1016/j.atmosenv.2014.12.038.

Fowler, D., Pilegaard, K., Sutton, M.A., Ambus, P., Raivonen, M., Duyzer, J., Simpson, D., Fagerli, H., Fuzzi, S., Schjoerring, J.K., Granier, C., Neftel, A., Isaksen, I.S.A., Laj, P., Maione, M., Monks, P.S., Burkhardt, J., Daemmgen, U., Neirynck, J., Personne, E., Wichink-Kruit, R., Butterbach-Bahl, K., Flechard, C., Tuovinen, J.P., Coyle, M., Gerosa, G., Loubet, B., Altimir, N., Gruenhage, L., Ammann, C., Cieslik, S., Paoletti, E., Mikkelsen, T.N., Ro-Poulsen, H., Cellier, P., Cape, J.N., Horváth, L., Loreto, F., Niinemets, Ü., Palmer, P.I., Rinne, J., Misztal, P., Nemitz, E., Nilsson, D., Pryor, S., Gallagher, M.W., Vesala, T., Skiba, U., Brüggemann, N., Zechmeister-Boltenstern, S., Williams, J., O'Dowd, C., Facchini, M.C., de Leeuw, G., Flossman, A., Chaumerliac, N., Erisman, J.W., 2009. Atmospheric composition change: ecosystems–atmosphere interactions. Atmos. Environ. 43, 5193–5267. https://doi.org/10.1016/j.atmosenv.2009.07.068.

Galmarini, S., Koffi, B., Solazzo, E., Keating, T., Hogrefe, C., Schulz, M., Benedictow, A., Griesfeller, J.J., Janssens-Maenhout, G., Carmichael, G., Fu, J., Dentener, F., 2017. Technical note: coordination and harmonization of the multi-scale, multi-model activities HTAP2, AQMEII3, and MICS-Asia3: simulations, emission inventories, boundary conditions, and model~output~formats. Atmos. Chem. Phys. 17, 1543–1555. https://doi.org/10.5194/acp-17-1543-2017.

Gery, M.W., Whitten, G.Z., Killus, J.P., Dodge, M.C., 1989. A photochemical kinetics mechanism for urban and regional scale computer modeling. J. Geophys. Res. Atmos. 94, 12925–12956. https://doi.org/10.1029/JD094iD10p12925.

Gong, S.L., Barrie, L.A., Blanchet, J.-P., 1997a. Modeling sea-salt aerosols in the atmosphere: 1. Model development. J. Geophys. Res. Atmos. 102, 3805–3818. https://doi.org/10.1029/96JD02953.

Gong, S.L., Barrie, L.A., Prospero, J.M., Savoie, D.L., Ayers, G.P., Blanchet, J.-P., Spacek, L., 1997b. Modeling Sea-salt aerosols in the atmosphere: 2. Atmospheric concentrations and fluxes. J. Geophys. Res. Atmos. 102, 3819–3830. https://doi.org/10.1029/96JD03401.

Grell, G.A., Peckham, S.E., Schmitz, R., McKeen, S.A., Frost, G., Skamarock, W.C., Eder, B., 2005. Fully coupled "online" chemistry within the WRF model. Atmos. Environ. 39, 6957–6975. https://doi.org/10.1016/j.atmosenv.2005.04.027.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P.I., Geron, C., 2006. Estimates of global terrestrial isoprene emissions using MEGAN (model of emissions of gases and aerosols from nature). Atmos. Chem. Phys. 6 https://doi.org/10.5194/acp-6-3181-2006.

Guenther, A.B., Jiang, X., Heald, C.L., Sakulyanontvittaya, T., Duhl, T., Emmons, L.K., Wang, X., 2012. The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. Geosci. Model Dev. 5, 1471–1492. https://doi.org/10.5194/gmd-5-1471-2012.

Henneman, L.R.F., Liu, C., Hu, Y., Mulholland, J.A., Russell, A.G., 2017. Air quality modeling for accountability research: operational, dynamic, and diagnostic evaluation. Atmos. Environ. 166 https://doi.org/10.1016/j.atmosenv.2017.07.049.

Hogrefe, C., Liu, P., Pouliot, G., Mathur, R., Roselle, S., Flemming, J., Lin, M., Park, R.J., 2018. Impacts of different characterizations of large-scale background on simulated regional-scale ozone over the continental United States. Atmos. Chem. Phys. 18, 3839–3864. https://doi.org/10.5194/acp-18-3839-2018.

HTAP, 2010. Hemispheric Transport of Air Pollution 2010. United Nations.

Huang, M., Carmichael, G.R., Pierce, R.B., Jo, D.S., Park, R.J., Flemming, J., Emmons, L. K., Bowman, K.W., Henze, D.K., Davila, Y., Sudo, K., Jonson, J.E., Tronstad Lund, M., Janssens-Maenhout, G., Dentener, F.J., Keating, T.J., Oetjen, H., Payne, V.H., 2017. Impact of intercontinental pollution transport on North American ozone airpollution: an HTAP phase 2 multi-model study. Atmos. Chem. Phys. 17, 5721–5750. https://doi.org/10.5194/acp-17-5721-2017.

Hundsdorfer, W., Koren, B., vanLoon, M., Verwer, J.G., 1995. A positive finite-difference advection scheme. J. Comput. Phys. 117, 35–46. https://doi.org/10.1006/jcph.1995.1042.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baró, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Denier van der Gon, H., Flemming, J., Forkel, R., Giordano, L., Jiménez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Makar, P.A., Manders-Groot, A., Neal, L., Pérez, J.L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R.S., Syrakov, D., Torian, A., Tuccella, P., Wang, K., Werhahn, J., Wolke, R., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., Galmarini, S., 2015a. Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part II: Particulate matter. Atmos. Environ. 115, 421–441. https://doi.org/10.1016/j.atmosenv.2014.08.072.

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

Im, U., Christensen, J.H., Geels, C., Hansen, K.M., Brandt, J., Solazzo, E., Alyuz, U., Balzarini, A., Baro, R., Bellasio, R., Bianconi, R., Bieser, J., Colette, A., Curci, G., Farrow, A., Flemming, J., Fraser, A., Jimenez-Guerrero, P., Kitwiroon, N., Liu, P., Nopmongcol, U., Palacios-Peña, L., Pirovano, G., Pozzoli, L., Prank, M., Rose, R., Sokhi, R., Tuccella, P., Unal, A., Vivanco, M.G., Yarwood, G., Hogrefe, C., Galmarini, S., 2018. Influence of anthropogenic emissions and boundary conditions on multi-modelsimulations of major air pollutants over Europe and North America in theframework of AQMEII3. Atmos. Chem. Phys. 18, 8929–8952. https://doi.org/10.5194/acp-18-8929-2018.

Inness, A., Ades, M., Agust\'{\i}-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J.J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., Suttie, M., 2019. The CAMS reanalysis of atmospheric composition. Atmos. Chem. Phys. 19, 3515–3556. https://doi.org/10.5194/acp-19-3515-2019.

Jacob, D.J., Winner, D.A., 2009. Effect of climate change on air quality. Atmos. Environ. 43 https://doi.org/10.1016/j.atmosenv.2008.09.051.

Jang, J.-C.C., Jeffries, H.E., Tonnesen, S., 1995. Sensitivity of ozone to model grid resolution — II. Detailed process analysis for ozone chemistry. Atmos. Environ. 29, 3101–3114. https://doi.org/10.1016/1352-2310(95)00119-J.

Janssen, S., Thunis, P., 2022. FAIRMODE Guidance Document on Modelling Quality Objectives and Benchmarking: Version 3.3. Publications Office of the European Union (doi:doi/10.2760/41988).

Jonson, J.E., Schulz, M., Emmons, L., Flemming, J., Henze, D., Sudo, K., Tronstad Lund, M., Lin, M., Benedictow, A., Koffi, B., Dentener, F., Keating, T., Kivi, R., Davila, Y., 2018. The effects of intercontinental emission sources on European air

pollution levels. Atmos. Chem. Phys. 18, 13655–13672. https://doi.org/10.5194/acp-18-13655-2018.
Karl, M., Dorn, H.-P., Holland, F., Koppmann, R., Poppe, D., Rupp, L., Schaub, A., Wahner, A., 2006. Product study of the reaction of OH radicals with isoprene in the atmosphere simulation chamber SAPHIR. J. Atmos. Chem. 55, 167–187. https://doi.org/10.1007/s10874-006-9034-x.
Kavassalis, S.C., Murphy, J.G., 2017. Understanding ozone-meteorology correlations: a role for dry deposition. Geophys. Res. Lett. 44 https://doi.org/10.1002/2016GL071791.
Kleinman, L.I., 1994. Low and high NOx tropospheric photochemistry. J. Geophys. Res. 99 https://doi.org/10.1029/94jd01028.
Knote, C., Hodzic, A., Jimenez, J.L., Volkamer, R., Orlando, J.J., Baidar, S., Brioude, J., Fast, J., Gentner, D.R., Goldstein, A.H., Hayes, P.L., Knighton, W.B., Oetjen, H., Setyan, A., Stark, H., Thalman, R., Tyndall, G., Washenfelder, R., Waxman, E., Zhang, Q., 2014. Simulation of semi-explicit mechanisms of SOA formation from glyoxal in aerosol in a 3-D model. Atmos. Chem. Phys. 14, 6213–6239. https://doi.org/10.5194/acp-14-6213-2014.
Koo, B., Jung, J., Pollack, A.K., Lindhjem, C., Jimenez, M., Yarwood, G., 2012. Impact of meteorology and anthropogenic emissions on the local and regional ozone weekend effect in Midwestern US. Atmos. Environ. 57 https://doi.org/10.1016/j.atmosenv.2012.04.043.
Krug, A., Fenner, D., Holtmann, A., Scherer, D., 2019. Occurrence and coupling of heat and ozone events and their relation to mortality rates in Berlin, Germany, between 2000 and 2014. Atmosphere (Basel) 10, 348. https://doi.org/10.3390/atmos10060348.
Krug, A., Fenner, D., Mücke, H.-G., Scherer, D., 2020. The contribution of air temperature and ozone to mortality rates during hot weather episodes in eight German cities during the years 2000 and 2017. Nat. Hazards Earth Syst. Sci. 20, 3083–3097. https://doi.org/10.5194/nhess-20-3083-2020.
Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I., van der Gon, H., 2022. CAMS-REG-v4: a state-of-the-art high-resolution European emission inventoryfor air quality modelling. Earth Syst. Sci. Data 14, 491–515. https://doi.org/10.5194/essd-14-491-2022.
Kuik, F., Lauer, A., Churkina, G., van der Gon, H.A.C., Fenner, D., Mar, K.A., Butler, T.M., 2016. Air quality modelling in the Berlin–Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data. Geosci. Model Dev. 9, 4339–4363. https://doi.org/10.5194/gmd-9-4339-2016.
Lamarque, J.-F., Shindell, D.T., Josse, B., Young, P.J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W.J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S.J., Horowitz, L.W., Lee, Y.H., MacKenzie, I.A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S.T., Schulz, M., Skeie, R. B., Stevenson, D.S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., Zeng, G., 2013. The atmospheric chemistry and climate model Intercomparison project (ACCMIP): overview and description of models, simulations and climate diagnostics. Geosci. Model Dev. 6, 179–206. https://doi.org/10.5194/gmd-6-179-2013.
Lecœur, È., Seigneur, C., 2013. Dynamic evaluation of a multi-year model simulation of particulate matter concentrations over Europe. Atmos. Chem. Phys. 13 https://doi.org/10.5194/acp-13-4319-2013.
Lin, M., Horowitz, L.W., Xie, Y., Paulot, F., Malyshev, S., Shevliakova, E., Finco, A., Gerosa, G., Kubistin, D., Pilegaard, K., 2020. Vegetation feedbacks during drought exacerbate ozone air pollution extremes in Europe. Nat. Clim. Chang. 10 https://doi.org/10.1038/s41558-020-0743-y.
Long, M.S., Keene, W.C., Kieber, D.J., Erickson, D.J., Maring, H., 2011. A sea-state based source function for size- and composition-resolved marine aerosol production. Atmos. Chem. Phys. 11, 1203–1216. https://doi.org/10.5194/acp-11-1203-2011.
Luo, H., Yang, L., Yuan, Z., Zhao, K., Zhang, S., Duan, Y., Huang, R., Fu, Q., 2020. Synoptic condition-driven summertime ozone formation regime in Shanghai and the implication for dynamic ozone control strategies. Sci. Total Environ. 745 https://doi.org/10.1016/j.scitotenv.2020.141130.
Lupaşcu, A., Butler, T., 2019. Source attribution of European surface O3 using a tagged O3 mechanism. Atmos. Chem. Phys. 19, 14535–14558. https://doi.org/10.5194/acp-19-14535-2019.
Lupaşcu, A., Otero, N., Minkos, A., Butler, T., 2022. Attribution of surface ozone to NO_x and volatile organic compound sources during two different high ozone events. Atmos. Chem. Phys. 22, 11675–11699. https://doi.org/10.5194/acp-22-11675-2022.
Luttkus, M.L., Hoffmann, E.H., Poulain, L., Tilgner, A., Wolke, R., 2022. The effect of land use classification on the gas-phase and particle composition of the troposphere: tree species versus forest type information. J. Geophys. Res. Atmos. 127, e2021JD035305 https://doi.org/10.1029/2021JD035305.
Manders, A.M.M., Builtjes, P.J.H., Curier, L., van der Gon, H.A.C., Hendriks, C., Jonkers, S., Kranenburg, R., Kuenen, J.J.P., Segers, A.J., Timmermans, R.M.A., Visschedijk, A.J.H., Wichink Kruit, R.J., van Pul, W.A.J., Sauter, F.J., van der Swaluw, E., Swart, D.P.J., Douros, J., Eskes, H., van Meijgaard, E., van Ulft, B., van Velthoven, P., Banzhaf, S., Mues, A.C., Stern, R., Fu, G., Lu, S., Heemink, A., van Velzen, N., Schaap, M., 2017. Curriculum vitae of the LOTOS–EUROS (v2.0) chemistry transport model. Geosci. Model Dev. 10, 4145–4173. https://doi.org/10.5194/gmd-10-4145-2017.
Mar, K., Ojha, N., Pozzer, A., Butler, T., 2016. Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: model evaluation and chemical mechanism comparison. Geosci. Model Dev. 9, 3699. https://doi.org/10.5194/gmd-9-3699-2016.
Maronga, B., Gross, G., Raasch, S., Banzhaf, S., Forkel, R., Heldens, W., Kanani-Sühring, F., Matzarakis, A., Mauder, M., Pavlik, D., Pfafferott, J., Schubert, S., Seckmeyer, G., Sieker, H., Winderlich, K., 2019. Development of a new urban climate model based on the model PALM - project overview, planned work, and first achievements. Meteorol. Z. 28, 105–119. https://doi.org/10.1127/metz/2019/0909.
Mårtensson, E.M., Nilsson, E.D., de Leeuw, G., Cohen, L.H., Hansson, H.-C., 2003. Laboratory simulations and parameterization of the primary marine aerosol production. J. Geophys. Res. Atmos. 108 https://doi.org/10.1029/2002JD002263.
Monahan, E.C., 1986. The ocean as a source for atmospheric particles. In: Buat-Ménard, P. (Ed.), The Role of Air-Sea Exchange in Geochemical Cycling. Springer, Netherlands, Dordrecht, pp. 129–163. https://doi.org/10.1007/978-94-009-4738-2_6.
Monks, P.S., Archibald, A.T., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K.S., Mills, G.E., Stevenson, D.S., Tarasova, O., Thouret, V., Von Schneidemesser, E., Sommariva, R., Wild, O., Williams, M.L., 2015. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmos. Chem. Phys. https://doi.org/10.5194/acp-15-8889-2015.
Monteiro, A., Durka, P., Flandorfer, C., Georgieva, E., Guerreiro, C., Kushta, J., Malherbe, L., Maiheu, B., Miranda, A.I., Santos, G., Stocker, J., Trimpeneers, E., Tognet, F., Stortini, M., Wesseling, J., Janssen, S., Thunis, P., 2018. Strengths and weaknesses of the FAIRMODE benchmarking methodology for the evaluation of air quality models. Air Qual. Atmos. Health 11. https://doi.org/10.1007/s11869-018-0554-8.
Morgenstern, O., Hegglin, M., Rozanov, E., O'Connor, F., Luke Abraham, N., Akiyoshi, H., Archibald, A., Bekki, S., Butchart, N., Chipperfield, M., Deushi, M., Dhomse, S., Garcia, R., Hardiman, S., Horowitz, L., Jöckel, P., Josse, B., Kinnison, D., Lin, M., Mancini, E., Manyin, M., Marchand, M., Marécal, V., Michou, M., Oman, L., Pitari, G., Plummer, D., Revell, L., Saint-Martin, D., Schofield, R., Stenke, A., Stone, K., Sudo, K., Tanaka, T., Tilmes, S., Yamashita, Y., Yoshida, K., Zeng, G., 2017. Review of the global models used within phase 1 of the Chemistry-Climate Model Initiative (CCMI). Geosci. Model Dev. https://doi.org/10.5194/gmd-10-639-2017.
Morrison, H., Gettelman, A., 2008. A new two-moment bulk Stratiform cloud microphysics scheme in the community atmosphere model, version 3 (CAM3). Part I: description and numerical tests. J. Climate 21, 3642–3659. https://doi.org/10.1175/2008JCLI2105.1.
Neu, J.L., Prather, M.J., 2012. Toward a more physical representation of precipitation scavenging in global chemistry models: cloud overlap and ice physics and their impact on tropospheric ozone. Atmos. Chem. Phys. 12, 3289–3310. https://doi.org/10.5194/acp-12-3289-2012.
O'Dowd, C.D., Smith, M.H., Consterdine, I.E., Lowe, J.A., 1997. Marine aerosol, sea-salt, and the marine sulphur cycle: a short review. Atmos. Environ. 31, 73–80. https://doi.org/10.1016/S1352-2310(96)00106-9.
Oltmans, S.J., Levy, H., 1994. Surface ozone measurements from a global network. Atmos. Environ. 28, 9–24. https://doi.org/10.1016/1352-2310(94)90019-1.
Otero, N., Sillmann, J., Schnell, J.L., Rust, H.W., Butler, T., 2016. Synoptic and meteorological drivers of extreme ozone concentrations over Europe. Environ. Res. Lett. 11 https://doi.org/10.1088/1748-9326/11/2/024005.
Otero, N., Sillmann, J., Mar, K.A., Rust, H.W., Solberg, S., Andersson, C., Engardt, M., Bergström, R., Bessagnet, B., Colette, A., Couvidat, F., Cuvelier, C., Tsyro, S., Fagerli, H., Schaap, M., Manders, A., Mircea, M., Briganti, G., Cappelletti, A., Adani, M., D'Isidoro, M., Pay, M.-T., Theobald, M., Vivanco, M.G., Wind, P., Ojha, N., Raffort, V., Butler, T., 2018. A multi-model comparison of meteorological drivers of surface ozone over Europe. Atmos. Chem. Phys. 18 https://doi.org/10.5194/acp-18-12269-2018.
Otero, N., Rust, H.W., Butler, T., 2021. Temperature dependence of tropospheric ozone under NOx reductions over Germany. Atmos. Environ. 253, 118334 https://doi.org/10.1016/j.atmosenv.2021.118334.
Parrish, D.D., Petropavlovskikh, I., Oltmans, S.J., 2017. Reversal of long-term trend in baseline ozone concentrations at the North American west coast. Geophys. Res. Lett. 44 https://doi.org/10.1002/2017GL074960.
Pay, M.T., Gangoiti, G., Guevara, M., Napelenok, S., Querol, X., Jorba, O., Pérez Garc \'{\i}a-Pando, C., 2019. Ozone source apportionment during peak summer events over southwesternEurope. Atmos. Chem. Phys. 19, 5467–5494. https://doi.org/10.5194/acp-19-5467-2019.
Rao, S.T., Galmarini, S., Puckett, K., 2011. Air quality model evaluation international initiative (AQMEII): advancing the state of the science in regional photochemical modeling and its applications. Bull. Am. Meteorol. Soc. 92, 23–30. https://doi.org/10.1175/2010BAMS3069.1.
Reidmiller, D.R., Fiore, A.M., Jaffe, D.A., Bergmann, D., Cuvelier, C., Dentener, F.J., Duncan, B.N., Folberth, G., Gauss, M., Gong, S., Hess, P., Jonson, J.E., Keating, T., Lupu, A., Marmer, E., Park, R., Schultz, M.G., Shindell, D.T., Szopa, S., Vivanco, M. G., Wild, O., Zuber, A., 2009. The influence of foreign vs. North American emissions on surface ozone in the US. Atmos. Chem. Phys. 9, 5027–5042. https://doi.org/10.5194/acp-9-5027-2009.
Reinert, D., Prill, F., Frank, H., Denhard, M., Baldauf, M., Schraff, C., Gebhardt, C., Marsigli, C., Zängl, G., 2016. DWD Database Reference for the Global and Regional ICON and ICON-EPS Forecasting System - Version 2.2.2. Dtsch. Wetterd. (DWD), Res. Dev.
Schaap, M., van Loon, M., ten Brink, H.M., Dentener, F.J., Builtjes, P.J.H., 2004. Secondary inorganic aerosol simulations for Europe with special attention to nitrate. Atmos. Chem. Phys. 4, 857–874. https://doi.org/10.5194/acp-4-857-2004.
Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J.M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M.T., Rouïl, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., Wind, P., 2015. Performance of European chemistry transport models as function of horizontal resolution. Atmos. Environ. 112, 90–105. https://doi.org/10.1016/j.atmosenv.2015.04.003.

Schaap, M., Kranenburg, R., Thürkow, M., Lupaşcu, A., Butler, T., 2023. Ozone source apportionment with a tagging approach in the LOTOS-EUROS model. In: EGU Gen. Assem. 2023, Vienna, Austria, 24–28 Apr 2023, EGU23-12861. https://doi.org/10.5194/egusphere-egu23-12861.

Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A., 2001. Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. J. Geophys. Res. Atmos. 106, 28275–28293. https://doi.org/10.1029/2001JD000384.

Schlegel, M., Knoth, O., Arnold, M., Wolke, R., 2012. Implementation of multirate time integration methods for air pollution modelling. Geosci. Model Dev. 5, 1395–1405. https://doi.org/10.5194/gmd-5-1395-2012.

Schlünzen, K.H., Bungert, U., Flagg, D.D., Fock, B.H., Gierisch, A., 2012. Technical Documentation of the Multiscale Model System M-SYS (METRAS, MITRAS, MECTM, MICTM, MESIM). MEMI Technical Report 3. Meteorol. Institut, KlimaCampus, Univ. Hamburg.

Schneider, C., Pelzer, M., Toenges-Schuller, N., Nacken, M., Niederau, A., 2016. ArcGIS basierte Lösung zur detaillierten, deutschlandweiten Verteilung (Gridding) nationaler Emissionsjahreswerte auf Basis des Inventars zur Emissionsberichterstattung - Kurzfassung; UBA TEXTE 71/2016. Für Mensch Umwelt.

Schnell, J.L., Prather, M.J., Josse, B., Naik, V., Horowitz, L.W., Cameron-Smith, P., Bergmann, D., Zeng, G., Plummer, D.A., Sudo, K., Nagashima, T., Shindell, D.T., Faluvegi, G., Strode, S.A., 2015. Use of North American and European air quality networks to evaluate global chemistry–climate modeling of surface ozone. Atmos. Chem. Phys. 15, 10581–10596. https://doi.org/10.5194/acp-15-10581-2015.

Schrödner, R., Tilgner, A., Wolke, R., Herrmann, H., 2014. Modeling the multiphase processing of an urban and a rural air mass with COSMO–MUSCAT. Urban Clim. 10, 720–731. https://doi.org/10.1016/j.uclim.2014.02.001.

Seinfeld, J.H., Pandis, S.N., 2006. Atmospheric Chemistry and Physics; From Air Pollution toClimate Change, second edition. John Wiley Sons, United States Am.

Seo, J., Youn, D., Kim, J.Y., Lee, H., 2014. Extensive spatiotemporal analyses of surface ozone and related meteorological variables in South Korea for the period 1999–2010. Atmos. Chem. Phys. 14, 6395–6415. https://doi.org/10.5194/acp-14-6395-2014.

Sillman, S., 1995. The use of NOy, H2O2, and HNO3 as indicators for ozone-NOx-hydrocarbon sensitivity in urban locations. J. Geophys. Res. 100 https://doi.org/10.1029/94jd02953.

Simpson, D., Guenther, A., Hewitt, C.N., Steinbrecher, R., 1995. Biogenic emissions in Europe: 1. Estimates and uncertainties. J. Geophys. Res. Atmos. 100, 22875–22890. https://doi.org/10.1029/95JD02368.

Simpson, D., Winiwarter, W., Börjesson, G., Cinderby, S., Ferreiro, A., Guenther, A., Hewitt, C.N., Janson, R., Khalil, M.A.K., Owen, S., Pierce, T.E., Puxbaum, H., Shearer, M., Skiba, U., Steinbrecher, R., Tarrasón, L., Öquist, M.G., 1999. Inventorying emissions from nature in Europe. J. Geophys. Res. Atmos. 104, 8113–8152. https://doi.org/10.1029/98JD02747.

Simpson, D., Benedictow, A., Berge, H., Bergström, R., Emberson, L.D., Fagerli, H., Flechard, C.R., Hayman, G.D., Gauss, M., Jonson, J.E., Jenkin, M.E., Nyiri, A., Richter, C., Semeena, V.S., Tsyro, S., Tuovinen, J.-P., Valdebenito, Á., Wind, P., 2012. The EMEP MSC-W chemical transport model - technical description. Atmos. Chem. Phys. 12, 7825–7865. https://doi.org/10.5194/acp-12-7825-2012.

Skamarock, W.C., Klemp, J., Dudhia, J., Gill, D.O., Barker, D., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3 27, pp. 3–27.

Sofiev, M., Soares, J., Prank, M., de Leeuw, G., Kukkonen, J., 2011. A regional-to-global model of emission and transport of sea salt particles in the atmosphere. J. Geophys. Res. Atmos. 116 https://doi.org/10.1029/2010JD014713.

Solazzo, E., Galmarini, S., 2016. Error apportionment for atmospheric chemistry-transport models – a newapproach to model evaluation. Atmos. Chem. Phys. 16, 6263–6283. https://doi.org/10.5194/acp-16-6263-2016.

Solazzo, E., Bianconi, R., Vautard, R., Appel, W., Moran, M., Hogrefe, C., Bessagnet, B., Brandt, J., Christensen, J., Chemel, C., Coll, I., van der Gon, H., Ferreira, J., Forkel, R., Francis, X., Grell, G., Grossi, P., Hansen, A., Jeričević, A., Galmarini, S., 2012. Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. Atmos. Environ. 53, 60–74. https://doi.org/10.1016/j.atmosenv.2012.01.003.

Steinbrecher, R., Smiatek, G., Köble, R., Seufert, G., Theloke, J., Hauff, K., Ciccioli, P., Vautard, R., Curci, G., 2009. Intra- and inter-annual variability of VOC emissions from natural and semi-natural vegetation in Europe and neighbouring countries. Atmos. Environ. 43, 1380–1391. https://doi.org/10.1016/j.atmosenv.2008.09.072.

Stern, R., 2003. Development and Application of the Chemical Transport Model REM/CALGRID - UFOPLAN-Ref. No. 298 41 252. Umweltbundesamt.

Stevenson, D.S., Young, P.J., Naik, V., Lamarque, J.-F., Shindell, D.T., Voulgarakis, A., Skeie, R.B., Dalsoren, S.B., Myhre, G., Berntsen, T.K., Folberth, G.A., Rumbold, S.T., Collins, W.J., MacKenzie, I.A., Doherty, R.M., Zeng, G., van Noije, T.P.C., Strunk, A., Bergmann, D., Cameron-Smith, P., Plummer, D.A., Strode, S.A., Horowitz, L., Lee, Y. H., Szopa, S., Sudo, K., Nagashima, T., Josse, B., Cionni, I., Righi, M., Eyring, V., Conley, A., Bowman, K.W., Wild, O., Archibald, A., 2013. Tropospheric ozone changes, radiative forcing and attribution to emissions in the atmospheric chemistry and climate model Intercomparison project (ACCMIP). Atmos. Chem. Phys. 13, 3063–3085. https://doi.org/10.5194/acp-13-3063-2013.

Stockwell, W.R., Kirchner, F., Kuhn, M., Seefeld, S., 1997. A new mechanism for regional atmospheric chemistry modeling. J. Geophys. Res. Atmos. 102, 25847–25879. https://doi.org/10.1029/97JD00849.

Tawfik, A.B., Steiner, A.L., 2013. A proposed physical mechanism for ozone-meteorology correlations using land-atmosphere coupling regimes. Atmos. Environ. 72 https://doi.org/10.1016/j.atmosenv.2013.03.002.

Thürkow, M., Kirchner, I., Kranenburg, R., Timmermans, R.M.A., Schaap, M., 2021. A multi-meteorological comparison for episodes of PM10 concentrations in the Berlin agglomeration area in Germany with the LOTOS-EUROS CTM. Atmos. Environ. 244, 117946 https://doi.org/10.1016/j.atmosenv.2020.117946.

van Pinxteren, D., Mothes, F., Spindler, G., Fomba, K.W., Herrmann, H., 2019. Trans-boundary PM10: quantifying impact and sources during winter 2016/17 in eastern Germany. Atmos. Environ. 200, 119–130. https://doi.org/10.1016/j.atmosenv.2018.11.061.

Van Zanten, M.C., Sauter, F.J., Wichink Kruit, R.J., Van Jaarsveld, J.A., Van Pul, W.A.J., 2010. Description of the DEPAC Module: Dry Deposition Modelling With DEPAC_GCN2010; Report 680180001/2010. RIVM, P.O. Box 1, 3720 BA Bilthoven, the Netherlands Tel +31 30 274 91 11 www.rivm.nl.

Walcek, C.J., 2000. Minor flux adjustment near mixing ratio extremes for simplified yet highly accurate monotonic calculation of tracer advection. J. Geophys. Res. Atmos. 105, 9335–9348. https://doi.org/10.1029/1999JD901142.

Wesely, M.L., 1989. Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models. Atmos. Environ. 23, 1293–1304. https://doi.org/10.1016/0004-6981(89)90153-4.

Wolke, R., Knoth, O., 2000. Implicit–explicit Runge–Kutta methods applied to atmospheric chemistry-transport modelling. Environ. Model. Software 15, 711–719. https://doi.org/10.1016/S1364-8152(00)00034-7.

Wolke, R., Schröder, W., Schrödner, R., Renner, E., 2012. Influence of grid resolution and meteorological forcing on simulated European air quality: a sensitivity study with the modeling system COSMO–MUSCAT. Atmos. Environ. 53, 110–130. https://doi.org/10.1016/j.atmosenv.2012.02.085.

Yamartino, R.J., 2003. Refined 3-d Transport and Horizontal Diffusion for the REM/CALGRID Air Quality Model. Freie Univ. Berlin, Inst. für Meteorol.

Young, P.J., Naik, V., Fiore, A.M., Gaudel, A., Guo, J., Lin, M.Y., Neu, J.L., Parrish, D.D., Rieder, H.E., Schnell, J.L., Tilmes, S., Wild, O., Zhang, L., Ziemke, J., Brandt, J., Delcloo, A., Doherty, R.M., Geels, C., Hegglin, M.I., Hu, L., Im, U., Kumar, R., Luhar, A., Murray, L., Plummer, D., Rodriguez, J., Saiz-Lopez, A., Schultz, M.G., Woodhouse, M.T., Zeng, G., 2018. Tropospheric ozone assessment report: assessment of global-scale model performance for global and regional ozone distributions, variability, and trends. Elem. Sci. Anthr. 6 https://doi.org/10.1525/elementa.265.