Modelling of tropospheric NO$_2$ using WRF-Chem with optimized temporal NO$_x$ emission profiles derived from in-situ observations - Comparisons to in-situ, satellite, and MAX-DOAS observations over central Europe

Leon Kuhn[1, 2], Steffen Beirle[2], Vinod Kumar[2], Sergey Osipov[2, 4], Andrea Pozzer[2], Tim Bösch[5], Rajesh Kumar[3], and Thomas Wagner[1, 2]

[1]Institute for Environmental Physics, University of Heidelberg, Germany
[2]Max Planck Institute for Chemistry, Mainz, Germany
[3]National Center for Atmospheric Research, Boulder, United States of America
[4]Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[5]Institute for Environmental Physics, University of Bremen, Germany

Correspondence: Leon Kuhn (l.kuhn@mpic.de)

Abstract. We present a WRF-Chem simulation over central Europe with a high spatial resolution of $3\,\mathrm{km}\times3\,\mathrm{km}$ and a focus on nitrogen dioxide (NO$_2$). A regional emission inventory, issued by the German Environmental Agency, with a spatial resolution of $1\,\mathrm{km}\times1\,\mathrm{km}$ is used. We demonstrate, that by precise temporal modulation of the emission data (use of "temporal profiles"), significant improvement in model accuracy over existing simulations is achieved. Simulated NO$_2$ surface concentrations are

5   compared to measurements from a total of 275 in-situ measurement stations in Germany, where the model was able to reproduce average noontime NO$_2$ concentrations with a bias of $+0.9\,\%$ and $R = 0.76$. A comparison between modelled NO$_2$ vertical column densities (VCDs) and satellite observations from TROPOMI (TROPOspheric Monitoring Instrument) is conducted, where crucial aspects of the observation process, such as altitude-dependent NO$_2$ sensitivity as well as the influence of clouds and a priori assumptions of the retrieval, are taken into account. Simulations and satellite observations are shown to agree with a

10   model bias of $-6.6\,\%$ and $R = 0.84$ for monthly means. Lastly, simulated NO$_2$ concentration profiles are compared to profiles obtained from Multiaxis Differential Optical Absorption Spectroscopy (MAX-DOAS) measurements of five European ground stations using the profile retrieval algorithms from the Mexican MAX-DOAS fit (MMF) and the Mainz Profile Algorithm (MAPA). For stations within Germany, biases of $-5.9\,\%$ to $+50.3\,\%$ were obtained when comparing average noontime NO$_2$ concentrations at different altitudes. Outside of Germany, where lower resolution emission data was used, biases of up to

15    $+78.6\,\%$ were observed. Overall, the study demonstrates that temporal modulation of emission data is crucial for modelling tropospheric NO$_2$ realistically.

1   Introduction

Modelling of regional atmospheric chemistry and transport (RCT) is an important discipline in the field of air quality research. While observational data is often only available at coarse spatial or temporal resolutions, model data can be generated on regular

20   grids of much higher spatio-temporal resolution. Modern RCT models can therefore be used to systematically investigate the processes of transport and (photo)-chemical conversion that trace gases are subject to upon emission into the atmosphere. Most importantly, however, they allow for predictions of trace gas concentrations when observational data is not available and they are also used for operational air quality forecasting. Thus, they give valuable insight into the dynamics of air quality in polluted regions of the earth. RCT models can be "fully coupled", meaning that they include interactions between chemistry

25   and meteorology, or "offline", meaning that they simulate chemistry and meteorology separately. Examples for state of the art RCT models are WRF-Chem (Grell et al., 2005), COSMO/MESSy (Kerkweg and Jöckel, 2012), Lotos-Euros (Manders et al., 2017), CAM-chem (Emmons et al., 2020), and CHIMERE (Menut et al., 2021).

Nitrogen dioxide (NO_2) is one of the most relevant chemical species for air quality in polluted regions. It is toxic to humans and acts as a precursor for ozone (O_3), a key pollutant of urban smog. The hazardous impact of NO_2 on human health has

30   been widely recognized among the scientific community (see e.g. Faustini et al. (2014); Mills et al. (2015); Chowdhury et al. (2021)). Monitoring and predicting realistic distributions of NO_2 in the troposphere is therefore of ongoing political and scientific interest.

Past modelling efforts with focus on tropospheric NO_2 have typically resulted in very similar general outcomes: while researchers found good agreement between modelled and observed meteorological data (such as wind speeds and air temper-

35   ature), systematic disagreements between modelled surface NO_2 concentrations and in-situ observations from ground based measuring stations were found. Visser et al. (2019) report on the results of a WRF-Chem simulation over central Europe in which noontime NO_2 surface concentrations and VCDs are underestimated by 38.5 % and approximately 15 %, respectively. In a comparison of monthly mean NO_2 VCDs, R-values between 0.82 and 0.92 were obtained. The authors identify an underestimation of soil emissions in their emission inventory (TNO-MACC-III, short for "Monitoring Atmospheric Composition

40   and Climate" by the Netherlands Organisation for Applied Scientific Research) as a possible explanation. Kuik et al. (2016) run a WRF-Chem simulation with the same emission inventory as Visser et al. (2019) over the region of Berlin, Germany, and observe even stronger underestimations of surface NO_2 by more than 50 % during daytime and a strong overestimation at night-time in a similar comparison. The study reveals that increasing the spatial resolution (including downscaling of the used emission data) of the model from $15\,\mathrm{km} \times 15\,\mathrm{km}$ to $1\,\mathrm{km} \times 1\,\mathrm{km}$ slightly improves agreement, but not to a satisfying

45   degree. In a subsequent publication, the authors attribute the disagreements to underestimations in their emission data (see Kuik et al., 2018). Mar et al. (2016) report that the choice of the chemical mechanism used by WRF-Chem barely impacts the NO_x underestimation by comparing simulations using the RADM2 and MOZART-4 chemical mechanisms. Kumar et al. (2021) demonstrate in a simulation with the MECO(n) model system over Germany using TNO-MACC-III and regional emission data from the German Environmental Agency (UBA), that agreement between modelled NO_2 concentrations and in-situ observa-

50   tions improves greatly when diurnal and seasonal variability is added to the yearly resolved emission data using reasonably chosen hourly and monthly weighting factors ("temporal profiles"). This temporal upsampling has become common practice among the air quality modelling community and standard values for such temporal profiles have established (see Crippa et al. (2020b); Kumar et al. (2021) and the references within). Kumar et al. (2021) used improved temporal profiles derived from empirical proxies, e.g. car counts on highways at different hours of the day. Altogether the contemporary literature comes to

the clear consensus that RCT simulations tend to underestimate tropospheric NO_2 and that the emission inventories and their preprocessing play a key role in resolving that problem.

A number of publications have theorized that an apparent model bias can be explained because groundbased in-situ measurements systematically overestimate NO_2 concentrations. Conventional in-situ methods often utilize molybdenum converters, which were found to be cross sensitive to other reactive nitrogen species, such as NH_3 or PAN. This issue was discussed e.g. by Dunlea et al. (2007), Steinbacher et al. (2007), Boersma et al. (2009), and Visser et al. (2019). However, a systematic investigation to prove a general bias of the NO_2 in-situ measurements in Europe is still missing. Furthermore, the corresponding European authority claims that their instruments follow strict regulations (defined by the European Norm EN 14221) and that the overestimation of NO_2 was found to be smaller than 7 % in test measurements (German Environmental Agency (a)).

In this paper, we show the results of a WRF-Chem simulation over central Europe for the month of May 2019 with a spatial resolution of $3\,km \times 3\,km$. The goal is to quantify the level of agreement between simulated NO_2 concentrations and VCDs and corresponding results from different observational datasets. In particular, we investigate whether the NO_2 biases of the simulation can be resolved by optimization of the temporal profiles. For this purpose our simulation results are compared to three reference datasets:

1. Surface NO_2 concentrations measured by a network of in-situ instruments over Germany, operated by the German Environmental Agency (UBA, see German Environmental Agency (b))

2. Tropospheric NO_2 vertical column densities (VCDs) measured by TROPOMI (TROPOspheric Monitoring Instrument) on the Sentinel 5 precursor satellite (specifically the NO_2 PAL data product as described in Eskes et al., 2019). This includes recomputing the air mass factors (AMFs) of the retrieval based on our simulation results.

3. NO_2 concentration profiles obtained from five MAX-DOAS instruments that partake in the FRM4DOAS project (see Fayt et al., 2021)

We show that when temporal profiles are chosen such that the simulation shows good agreement with in-situ measurements, the simulation also agrees well with satellite and MAX-DOAS observations. This demonstrates that the total NO_x (= NO_2 + NO) budget assumed in the used emission inventories is realistic. As a direct consequence of this finding, improved simulation strategies like ours could be used for more precise air quality modelling with the focus on NO_2, NO_x or other species in the future.

The paper is structured as follows: Sect. 2 describes the setup of our WRF-Chem simulation and the preprocessing of emission data in detail. Sect. 3 shows the results of our simulation and compares them to a "classical" simulation with empirically derived temporal profiles. Furthermore, the simulated data is compared to the above mentioned observational datasets. In sect. 4 the results are discussed and conclusions are drawn.
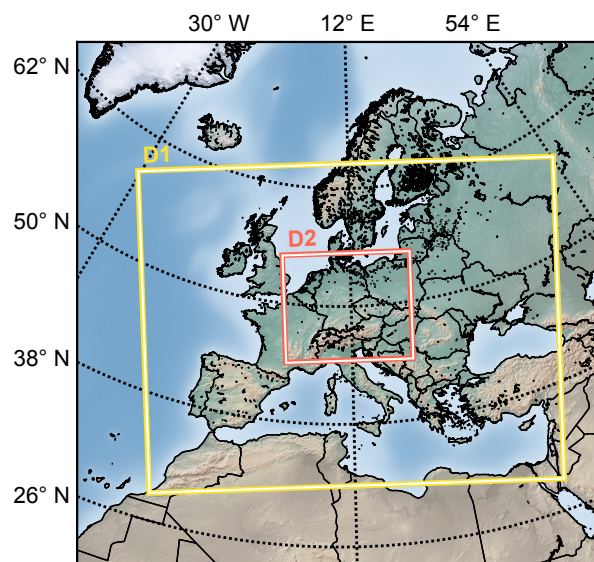
Figure 1. Geographical coverage of the WRF-Chem simulation. The spatial resolutions of the outer domain D1 and the inner domain D2 are $15\,\mathrm{km} \times 15\,\mathrm{km}$, and $3\,\mathrm{km} \times 3\,\mathrm{km}$, respectively.

2 WRF-Chem simulation setup

WRF-Chem (Weather Research and Forecasting model with Chemistry, see Grell et al., 2005) is a mesoscale RCT model. We run WRF-Chem on a twofold nested domain over central Europe for the month of May, 2019, see Fig. 1. The spatial resolutions of the outer and inner domain (called D1, and D2 from hereon) are $15\,\mathrm{km} \times 15\,\mathrm{km}$, and $3\,\mathrm{km} \times 3\,\mathrm{km}$ (with 320×245 pixels and 500×430 pixels, respectively). The temporal resolution of the simulation is 60 seconds on D1 and 12 seconds on D2. Output files are written daily for D1 and hourly for D2. The simulation uses the Thompson microphysics scheme (see Thompson et al., 2008), the RRTMG (Rapid Radiative Transfer Model for General Circulation Models long- and shortwave radiation scheme, see Iacono et al., 2008), the Monin-Obukhov similarity scheme for surface layer modelling (see Monin and Obukhov, 1954), the NOAH Land-Surface Model (see Niu et al., 2011), the YSU boundary layer scheme (see Hong, 2010), and the Grell-Devenyi ensemble scheme for cumulus modelling (see Grell and Dévényi, 2002). For modelling of chemistry, the MOZART chemical mechanism (see Emmons et al., 2010) is coupled to the GOCART aerosol mechanism (see Chin et al., 2000) along with the TUV full photolysis scheme (Madronich (1987); Tie et al. (2003)), which deploys climatological O_3 and O_2 columns. Dry deposition is calculated according to Wesely (1989). Spectral nudging (see e.g. Omrani et al., 2012) to ERA5 reanalysis data (see Hersbach and Dee, 2017) is used. The full list of WRF-Chem simulation options (the "namelist") can be found in the supplement of this article. The simulation uses a total of 43 vertical levels in terrain-following coordinates (see Table A1 for the layer heights up to 6 km). Both domains D1 and D2 receive initial conditions from the CAM-chem model (Emmons

et al., 2020). Additionally, CAM-chem yields the boundary conditions for D1. D2 receives boundary conditions online from the WRF-Chem simulation running on D1.

2.1 Emission preprocessing

The WRF-Chem simulation is driven by different emission inventories for different emission sectors (see also Table 1). Emis-
sions from biomass burning are taken from the Fire Inventory from NCAR (FINN, see Wiedinmyer et al., 2011) with a spatio-
temporal resolution of 1 km × 1 km and 24 hours. Biomass burning emissions are assumed to have a diurnal variation with a
peak in the early afternoon (1 pm local time) and are distributed vertically in the model following the plumerise parametrization
of Freitas et al. (2007). Biogenic emissions are computed online using an implementation of the MEGAN model (see Guenther
et al., 2006). For non-organic anthropogenic emissions a combination of two emission inventories is used: Over Germany,
an inventory of high spatial resolution (1 km × 1 km, resampled to $0.01° × 0.01°$ or ∼ 1.1 km × 0.7 km) is provided by the
German Environmental Agency (Umweltbundesamt, UBA, see Strogies et al., 2020). From hereon, this emission inventory
will be referenced as UBA-E. Using the UBA-E emission data over Germany enables modelling of NO_2 distributions at an
unprecedented high spatial resolution. Outside of Germany the EDGARv5 emission inventory with a moderate spatial resolu-
tion ($0.1° × 0.1°$ or ∼ 11 km × 7 km) is used (see Crippa et al., 2020a). Since UBA-E does not include organic anthropogenic
emissions, EDGARv5 is used on the entirety of D1 and D2 for organic species. Non-methane volatile organic compounds
(NMVOCs) are provided as lumped species and are speciated according to Huang et al. (2017). An alternative to EDGARv5
would have been the TNO-MAC-III emission inventory (Kuenen et al., 2014) which comes at a higher spatial resolution of
$0.0625° × 0.125°$ but was only available for the year 2011 at the time the simulation was run. Since EDGARv5 was available
for 2015 and NO_x emissions have steadily decreased over the past years (see e.g. Anenberg et al., 2022), EDGARv5 was
considered a more reasonable choice.

EDGARv5 (in 2015) and UBA-E have time resolutions of 1 month and 1 year, respectively. However, many emissions follow
strong diurnal and seasonal patterns. For example, emissions from car traffic are expected to be lower at nighttime and higher
at daytime and agricultural emissions typically occur in specific months of the year. EDGARv5 can still resolve emission
variations on a monthly time-resolution but is incapable of resolving diurnal patterns. UBA-E provides only annual emissions,
therefore without any temporal pattern. The solution to this problem is to scale the coarsely resolved emission data to an hourly
resolution using presumed hourly, daily, and monthly emission scaling factors, called temporal profiles. The emission rate
$E_{X,k}(m, d, h, \text{lat, lon})$ of a species X from sector k at month m, day d, and hour h at fixed latitude lat and longitude lon is
given as

$$E_{X,k}(m, d, h, \text{lat, lon}) = \hat{E}_{X,k}(\text{lat, lon}) \cdot p_{\text{monthly},k}(m) \cdot p_{\text{daily},k}(d) \cdot p_{\text{hourly},k}(h) \tag{1}$$

where $\hat{E}_{X,k}$ denotes the total emission budget of species X from sector k in the emission inventory and p_{monthly}, p_{daily}, and
p_{hourly} the monthly, daily, and hourly temporal profiles. The individual profiles are normalized to 12 (annual cycle), 7 (weekly
cycle), and 24 (diurnal cycle), respectively. Because different emission sectors follow vastly different temporal patterns, the
temporal profiles are defined for each sector individually. An overview of the emission sectors is given in Table 1. The total

Table 1. Emission sectors in the UBA-E and EDGARv5 emission inventories

Name	Contribution[†] (UBA-E) [%]	Contribution[†] (EDGARv5) [%]
traffic (no resuspension)	43.8	38.7
power industry	18.5	15.0
agricultural soils	10.1	4.5
energy for buildings	7.3	6.6
manufacturing industry	7.2	15.8
non-metallic minerals production	2.5	0.0
production of chemicals	2.4	0.3
shipping	2.1	2.1
iron and steel production	1.8	< 0.1
oil refineries and transformation industry	1.7	1.8
aviation landing and take-off	1.2	1.8
railways, pipelines, and off-road transport	0.9	2.3
production of food, pulp, and paper	0.3	0.3
manure management	0.1	1.1
fuel exploitation	0.1	0.0
solid waste incineration	0.1	0.1
non-ferrous metal production	0	< 0.1
non-energy use of fuels	< 0.1	0.0
agricultural waste burning	N/A	0.3
fossil fuel fires	N/A	< 0.1
aviation climbing and descent	N/A	5.9
aviation cruise	N/A	3.4

N/A = "not available"

[†] Relative contribution of this sector to the overall emission budget (yearly for UBA-E, and for the month of May for EDGARv5) of NO_x

emission rate of species X is obtained by summation over all emission sectors k

$$E_X(m, d, h, \text{lat, lon}) = \sum_k \hat{E}_{X,k}(\text{lat, lon}) \cdot p_{\text{monthly},k}(m) \cdot p_{\text{daily},k}(d) \cdot p_{\text{hourly},k}(h) \qquad (2)$$

Some species are only implicitly contained in the emission inventories. For example, EDGARv5 and UBA-E specify NO_x emissions, but the partitioning into NO and NO_2 must be chosen by the user via a speciation profile $p_{\text{spec}}(k)$. Equation (2) then generalizes to

$$E_X(m, d, h, \text{lat, lon}) = \sum_k \hat{E}_{X_{\text{lump}},k}(\text{lat, lon}) \cdot p_{\text{monthly},k}(m) \cdot p_{\text{daily},k}(d) \cdot p_{\text{hourly},k}(h) \cdot p_{\text{spec},k}(X) \qquad (3)$$

140   where X_{lump} is the lump of species that contains species X (i.e. $X_{\mathrm{lump}} = \mathrm{NO_x}$ when $X = \mathrm{NO_2}$).

In principle, WRF-Chem also supports vertical distribution of trace gas emissions. This is reasonable to consider, given that many strong emissions, like those from combustion stacks, take place at elevated altitudes. In analogy to the mentioned temporal and speciation profiles, this can be modelled using an additional vertical emission profile. Suggestions for vertical emission profiles are given in Bieser et al. (2011) and Pozzer et al. (2009). In our study we omit the use of vertical profiles and

145   inject all emissions into the lowest model layer (0 - 8 m). This simplification was made seeing that vertical profiles would add another dimension of complexity to our main goal, which is to optimize the temporal profiles. Furthermore, precise information on the emission height is often not available. Lastly, assuming a well-mixed troposphere, vertical profiles can be expected to be relevant only in the close vicinity of strong elevated emission sources.

2.2   Optimization of temporal profiles

150   The key difference between previously published RCT simulations and ours is the optimization of the temporal emission profiles. The goal is to find temporal profiles that minimize the model's mean $\mathrm{NO_x}$ bias over the course of the day. The mean relative bias is computed as

$$\mathrm{bias} = \frac{1}{N} \sum_{i=1}^{N} \frac{x_{\mathrm{sim},i} - x_{\mathrm{obs},i}}{x_{\mathrm{obs},i}} \tag{4}$$

where $x_{\mathrm{obs},i}$ denotes the i-th observation and $x_{\mathrm{sim},i}$ the corresponding simulated value. We use the observations from in-situ

155   measurements of background $\mathrm{NO_x}$ surface concentrations as reference. A detailed explanation of the dataset is given in sect. 3.1. Equation (3) hints towards the complexity of the optimization problem: Because the emission inventories include dozens of emission sectors, each with their own emission profiles, this poses an optimization problem with many degrees of freedom. In addition, a single WRF-Chem simulation of just one month takes days to finish even on modern supercomputers. This circumstance makes it nearly impossible to optimize the emission profiles using standard methods like gradient descent. For

160   our simulation we have therefore optimized the emission profiles manually. By "manual optimization" we mean the iterative process of running WRF-Chem, evaluating the simulation results against in-situ observations, and slightly nudging the temporal profiles in a direction in which better agreement between simulation and observations can be expected. Due to the short lifetime of $\mathrm{NO_x}$ the observed concentrations closely follow the temporal profiles. As an initial starting point, the temporal profiles from Kumar et al. (2021) were used. Although the proposed optimization method is rather unconventional, it has a few important

165   benefits. Consider the following:

1. The hourly profiles of many sectors have characteristic shapes, e.g. a peak in the hourly profile of the traffic sector during the morning rush hour. These should be at least coarsely preserved during the optimization process in order to maintain realistic emission behaviour.

2. Because the optimization problem is ill-posed, it is often unclear, of which sector the profiles should be tuned further

170   in order to improve the simulation. Sometimes, the spatial distribution of a specific emission sector matches that of the model error, indicating that this sector should be prioritized.

3. Gradient-based optimization methods depend on hyperparameters, such as the step size. If these are not picked correctly from the start, the optimization may converge slowly or diverge entirely. This becomes unfeasible, when a single forward run takes days to compute.

175 Using a conventional optimization routine, where the gradient of a loss function determines the outcome of a single optimization step, would require to encode aspects 1-3 in the form of mathematical constraints. This makes a rigorous treatment of the problem extremely complex. The manual optimization approach, however, does not require a mathematical formulation and can thus take the discussed aspects into consideration more easily. We run the optimization under the following conditions:

1. Only the most relevant emission sectors are modified during the process. According to Table 1 these are: traffic (no resuspension), power industry, agricultural soils, energy for buildings, and manufacturing industry.

2. Speciation of lumped species follows the recommendations of Huang et al. (2017) and is not further optimized. For NO_x, the partitioning is assumed to be 87.5 % NO and 12.5 % NO_2. This choice reflects the fact that NO_x from combustion processes is mostly emitted as NO, which mostly oxidizes to NO_2 over time. Literature values for typical NO_2/NO_x ratios in anthropogenic emissions range from lower values (e.g. 5.3 %, as reported by Wild et al. (2017) and 7.8 %, as reported by Jimenez et al. (2000)) to much higher values (e.g. 39 %, as reported by Richmond-Bryant et al. (2017) and 36 %, as reported by Costantini et al. (2016)).

3. In order to improve generalization of the temporal profiles, the optimization is performed using simulated and observed data from May 2018. Additionally, we accelerate the optimization by only using two weeks of simulation time for each optimization step. The final temporal profiles are evaluated in a full-month simulation for the year 2019, as described in sect. 2.

Figure 2 gives an overview of the optimization process in a total of 3 steps. Figure 2h reveals that the overall NO_x budget assumed for the simulated month is realistic, meaning that the corresponding monthly weighting factor is reasonable. Similarly, the daily profiles from Kumar et al. (2021) were taken over without changes. What remains are the hourly profiles of each sector. Prior to the optimization, simulated NO_x was too low at daytime and too high at night. Therefore, the hourly profiles were adjusted accordingly. For example, in step 1 (red line), only a single change was made to the traffic sector profile (subfigure a) in order to boost simulated NO_x values in the morning ($\sim$ 6 AM). The principle was extended to the remaining hourly profiles in the next two steps. The full list of emission profiles used in our simulation is attached to the supplement of this article.

2.3 Chemical boundary conditions

As described, our simulation uses CAM-chem model results as chemical boundary conditions (BCs) on the outer domain D1 and initial conditions (ICs) on both domains. In practice, it shows that simulated NO_2 and NO concentrations are mostly insensitive to BCs and ICs. O_3 concentrations, however, are highly sensitive to the O_3 BCs. At the same time, O_3 is known to be a crucial reaction partner of NO_x chemistry, mostly for oxidizing NO in the reaction

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{R1}$$
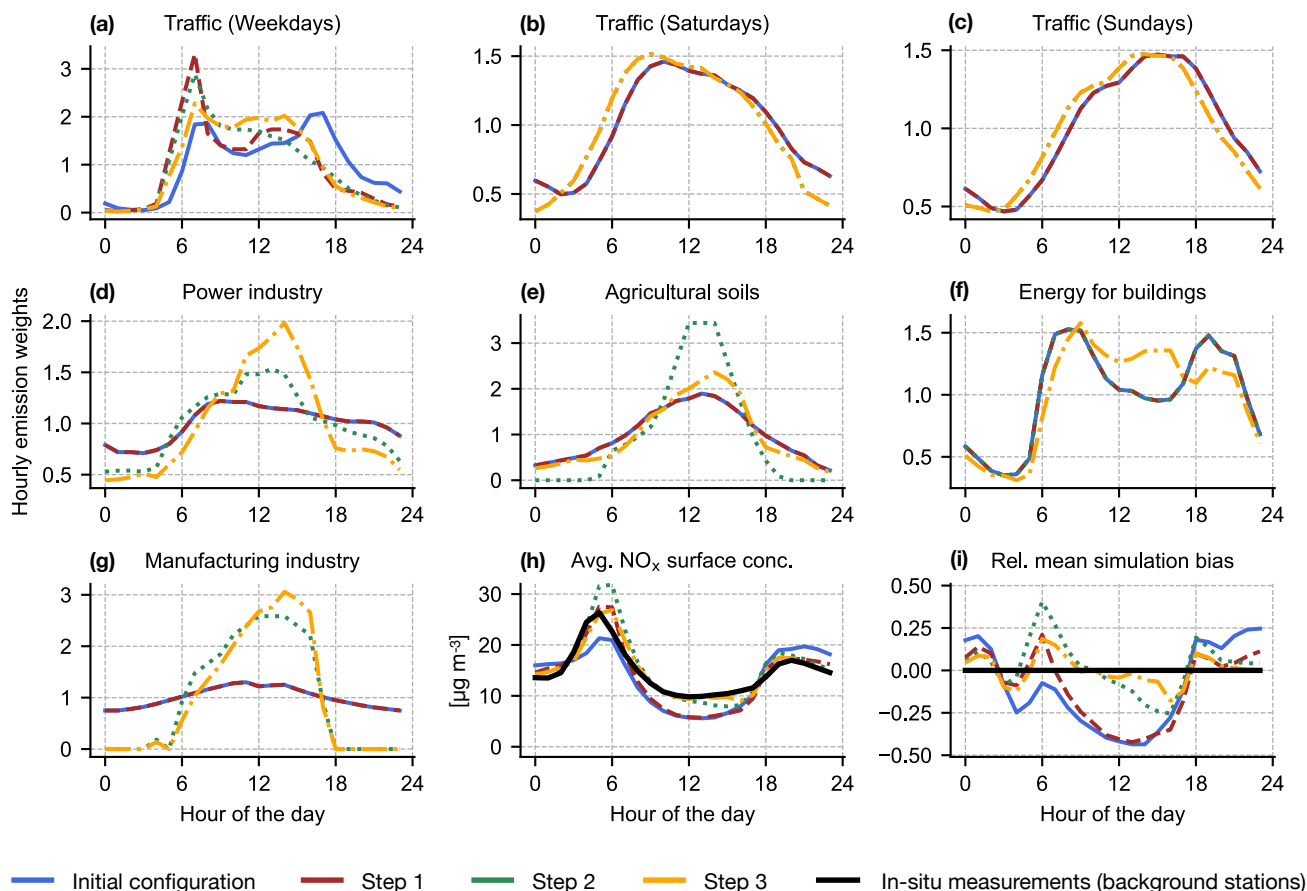
Figure 2. Overview of the optimization process of the hourly emission profiles based on the simulation results and in-situ measurements of May 2018. The profiles used by Kumar et al. (2021) are drawn in blue and our final optimized profiles in orange.

Therefore, in addition to the tuning of temporal emission profiles, we further conduct a tuning of O_3 BCs in the same fashion. As a result, a O_3 BC scaling factor of ~ 0.85 was determined. The results discussed in sect. 3 verify that scaling of the O_3 BCs yields an improvement to the resulting O_3 concentrations.

3 Results

We compare the results of two WRF-Chem simulations for May 2019 (one with the temporal profiles used by Kumar et al. (2021) called "S-old", and one with optimized profiles called "S-new") to the dataset of in-situ surface concentration measurements and show that S-new outperforms S-old in every regard. We then continue to compare S-new to the remaining observational datasets of satellite and MAX-DOAS measurements.

3.1 Comparison of surface concentrations

In-situ measurements of trace gas surface concentrations in Germany are conducted by the UBA. They are available for NO_2, NO and O_3 as hourly mean values. A total of 434 UBA measuring stations are distributed over Germany. 92 % of the stations use a chemiluminescence based measuring method. The remaining 8 % use other methods (cavity enhanced phase shift spectroscopy, diffuse sampling, or photolysis conversion). Of all UBA stations, 63 % are classified as "background", 30 % as "traffic", and 7 % as "industry". In this study we only take "background" stations into consideration. This choice was made seeing that NO_2 concentration can vary strongly near traffic over distances of only 10 - 100 m (see e.g. Beckwith et al., 2019) and timescales of a few minutes. Given that our simulation runs at a spatial resolution of $3\,\mathrm{km} \times 3\,\mathrm{km}$, it is expected to show poor agreement with the traffic stations. Many of the publications referenced in sect. 1 have also restricted their reference data to "background" in-situ measurements.

In order to compare the simulated surface concentrations to the in-situ measurements, they are interpolated from the WRF-Chem simulation grid to the geolocations of the UBA stations. The uncertainty Δx_t of the simulated surface concentrations and the in-situ reference values at time step t is computed via

$$\Delta x_t = \max\left(|x_t - x_{t-1}|, |x_t - x_{t+1}|\right) \tag{5}$$

i.e. the maximal hour-to-hour variation in concentration. Figure 3 shows the average surface concentrations of NO_2, NO, NO_x, and O_3 of both simulations S-old and S-new as well as the corresponding in-situ measurements. The left panel (subfigures a-d) is restricted to the first ten days of the simulated time period (01 May 2019 - 10 May 2019) for easier readability. The right panel (subfigures e-h) shows the average diurnal concentrations obtained from averaging over all days of the simulation period.

Comparing S-old and S-new shows the vast difference in simulation accuracy achieved by using the optimized temporal profiles. S-old consistently overestimates the NO_x concentration at night, but underestimates it at daytime. Particularly around noontime the simulated NO_x concentrations are low-biased by approximately $-20\,\%$. S-new achieves much better agreement with the in-situ measurements: The mean NO_x concentration is reproduced with minimal deviations ($< 2\,\mu\mathrm{g\,m}^{-3}$), regardless of daytime. Given that the used emission inventories treat NO and NO_2 as lumped NO_x, this validates our choice of temporal profiles.

The individual time series of NO_2 and NO also show good agreement with the in-situ observations, with minimal to no deviations at noontime. However, discrepancies during nighttime between simulation and observations remain ($< 4.7\,\mu\mathrm{g\,m}^{-3}$ for NO_2 and $< 4.1\,\mu\mathrm{g\,m}^{-3}$ for NO).

S-old shows a consistent overestimation of O_3 of approximately $20\,\mu\mathrm{g\,m}^{-3}$ regardless of daytime, which is reduced to approximately $10\,\mu\mathrm{g\,m}^{-3}$ in S-new. The improvements are mostly attributed to the adjustments to the O_3 boundary conditions described in sect. 2.3. A version of Fig. 3 showing traffic stations instead of background stations can be found in appendix A (Fig. A1). The NO_2 concentration timeline of the traffic observations has a fundamentally different shape compared to the background observations, with practically no noontime low and no evening peak. As expected, the modelled and observed concentration timelines do not agree well for the traffic stations. Interestingly, although this was not considered during the optimization process, the temporal profiles we have obtained for the traffic sector (see Fig. 2a) have a shape similar to the NO_x
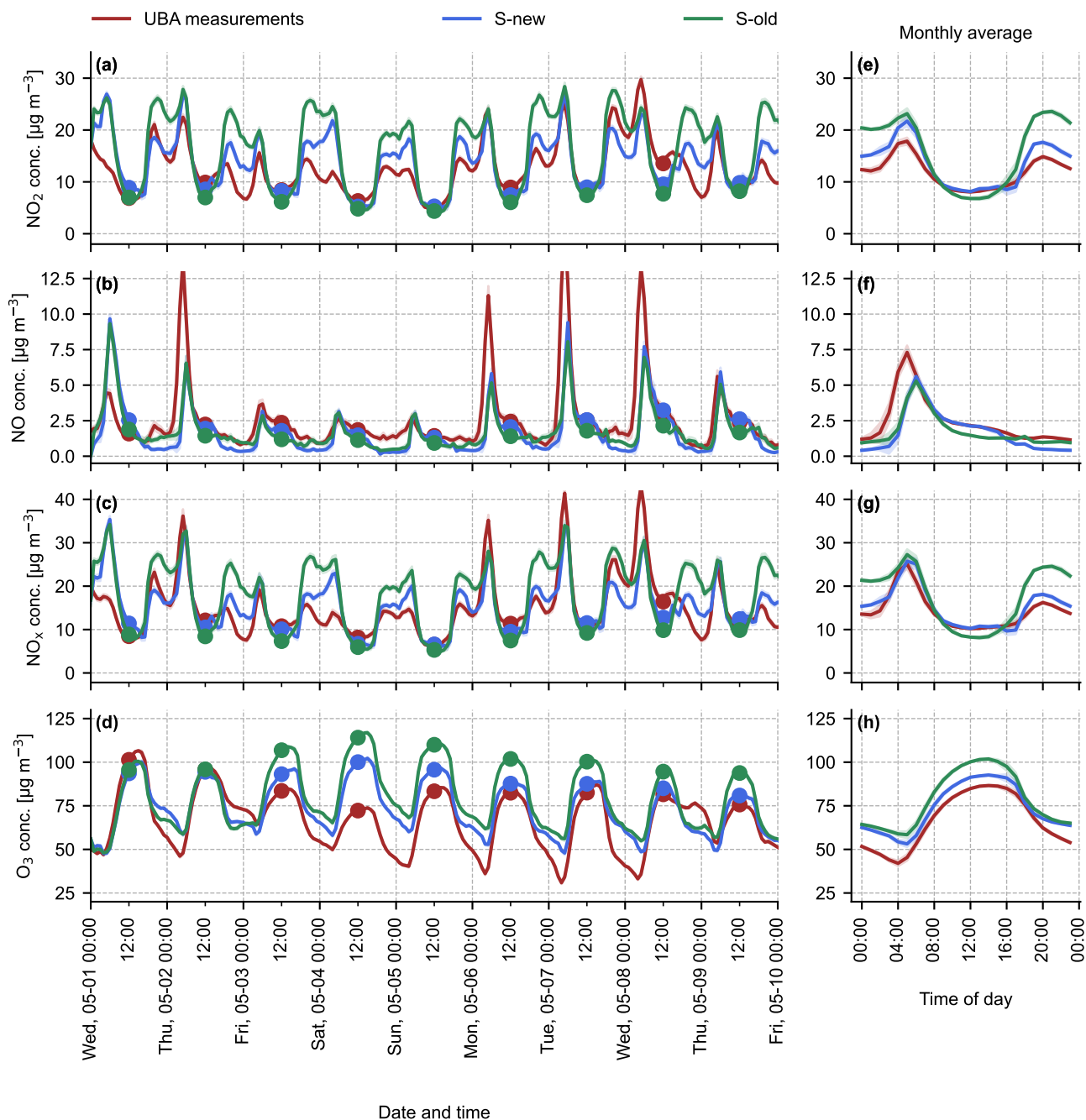
Figure 3. Time series of NO_2, NO, NO_x, and O_3 surface concentrations. The red lines show reference values obtained from the UBA background in-situ stations. The green and blue lines show the simulation results of S-old and S-new, respectively. **(a)** - **(d)** display a time series spanning multiple days (01 May 2019 - 10 May 2019) and **(e)** - **(f)** display the corresponding average concentration values over the entire month of May 2019.

Table 2. Statistical summary for the surface concentration comparison (all hours of the day, "background" stations only)

Trace gas	Mean [μg m^{-3}]			RMSE [μg m^{-3}]		Bias		R	
	UBA	S-new	S-old	S-new	S-old	S-new	S-old	S-new	S-old
NO$_2$	11.9	13.6	15.9	3.7	6.4	+14.3 %	+34.0 %	0.77	0.72
NO	2.5	1.7	1.8	1.9	1.8	-30.0 %	-27.8 %	0.63	0.65
NO$_x$	14.3	15.3	17.7	4.2	6.6	+6.7 %	+23.4 %	0.76	0.67
O$_3$	66.2	73.4	78.5	12.2	16.8	+10.9 %	+18.7 %	0.85	0.82

Table 3. Statistical summary for the surface concentration comparison (at noon, "background" stations only)

Trace gas	Mean [μg m^{-3}]			RMSE [μg m^{-3}]		Bias		R	
	UBA	S-new	S-old	S-new	S-old	S-new	S-old	S-new	S-old
NO$_2$	8.1	8.1	6.8	1.7	2.0	+0.9 %	-15.7 %	0.76	0.75
NO	2.1	2.1	1.5	0.5	0.8	-1.5 %	-31.8 %	0.72	0.66
NO$_x$	10.2	10.2	8.3	2.0	2.7	+0.4 %	-19.1 %	0.77	0.75
O$_3$	84.9	91.4	100.2	11.1	19.0	+7.6 %	+18.0 %	0.72	0.58

time series of the traffic stations (see Fig. A1g), whereas the corresponding initial profile deviates quite strongly in shape. This can be seen as further validation of our results.

Figure 4 emphasizes the improvement in modelling accuracy via scatter plots for NO$_2$, NO, NO$_x$, and O$_3$ surface concentrations at noontime. The results of S-new are consistently closer to the 1:1 line and linear regressions through the point clouds

250 of each plot yield slopes much closer to 1 with similar (small) intercepts compared to S-old. For example, comparing Fig. 4a and Fig. 4b shows slopes of 0.84 (S-new) and 0.73 (S-old) for NO$_2$. A version of Fig. 4 showing values for all hours of the day can be found in appendix A (see Fig. A2). A summary of corresponding statistical diagnostics is given in Tables 2 and 3 for all hours of the day and noontime only, respectively. The Root Mean Square Error (RMSE) is computed according to

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(x_{\text{sim},i} - x_{\text{obs},i}\right)^2}{N}} \qquad (6)$$

255 where $x_{\text{obs},i}$ denotes the i-th observation and $x_{\text{sim},i}$ the corresponding simulated value. Here the RMSE is computed from the averages over all stations used to generate Fig. 3 and 4.

Comparing the entries of Tables 2 and 3 shows that S-new consistently outperforms S-old by showing lower RMSE, lower bias, and higher R-values. Most importantly, the biases of total and noontime NO$_x$ are reduced from +23.4 % to +6.7 % and from −19.1 % to +0.4 %, respectively. Similarly, the biases of total and noontime NO$_2$ are reduced from +34.0 % to +14.3 %

260 and from −15.7 % to +0.9 %. There appears to be only one minor exception to the overall improvement, namely total NO (see
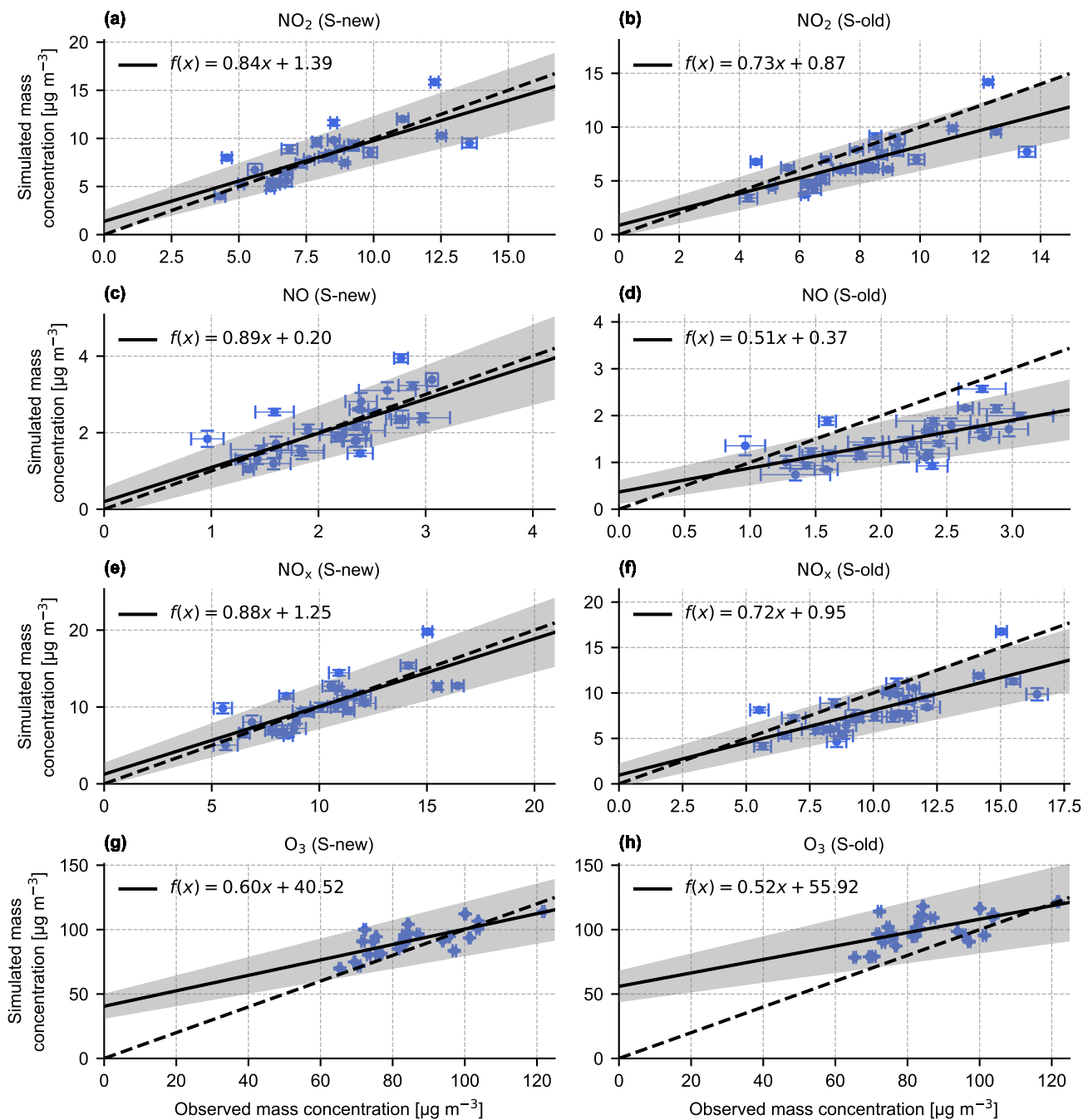
Figure 4. Scatter plots of simulated vs. observed surface concentrations of NO_2, NO, NO_x and O_3. One scatter point resembles the mean over all "background" stations at noontime of one day. The left panel **(a)** - **(g)** shows the results of the new simulation (S-new) and the right panel **(b)** - **(h)** shows the results of the simulation with the original temporal profiles (S-old). The dashed line in each subplot marks the 1:1 line.

second line of Table 2). However, the difference in RMSE between S-new and S-old is on the scale of $\sim 0.1 \ \mu\mathrm{g} \ \mathrm{m}^{-3}$ for NO,
which is minuscule.

3.2 Comparison of NO_2 vertical column densities

As a second diagnostic, we compare simulated NO_2 vertical column densities (VCDs) to observations of the TROPOMI
265 (TROPOspheric Monitoring Instrument, see Veefkind et al., 2012) satellite instrument. The satellite overpass occurs daily,
typically at 13:30 local time in central Europe, with a pixel size of $3.5 \ \mathrm{km} \times 5.5 \ \mathrm{km}$ in NADIR viewing geometry. TROPOMI
measures spectra of backscattered sunlight from earth's surface, from which tropospheric slant column densities (SCDs) are
computed using Differential Optical Absorption Spectroscopy (DOAS, see Platt and Stutz, 2008). Tropospheric VCDs are
obtained from tropospheric SCDs via

$$270 \quad \mathrm{VCD_{trop}} = \frac{\mathrm{SCD_{trop}}}{\mathrm{AMF_{trop}}} \qquad (7)$$

where $\mathrm{AMF_{trop}}$ stands for the tropospheric air mass factor. The tropospheric AMF depends on a number of atmospheric and
surface conditions and assumptions, e.g. cloud properties, the stratospheric column, surface albedo, and the relative a priori
NO_2 profile. In the TROPOMI retrieval, the NO_2 a priori profiles are taken from the TM5 global CT model (see Krol et al.
(2005); Williams et al. (2017)), with a spatial resolution of $\sim 1° \times 1° \approx 110 \ \mathrm{km} \times 70 \ \mathrm{km}$ where our simulation domain D2 is
275 located. A detailed description of the retrieval algorithm can be found in van Geffen et al. (2022) and Eskes et al. (2021). From
hereon we exclusively refer to tropospheric VCDs and SCDs, respectively.

Simulated VCDs are obtained by vertical integration of box VCDs (= partial columns) from WRF-Chem. The WRF-Chem
output is interpolated to the vertical grid of the TM5 model. The box VCDs are computed by multiplying the NO_2 concentration
within each model grid cell with its vertical extent. The partial columns are then summed up:

$$280 \quad \mathrm{VCD_{sim}} = \sum_{l<13} c_l \cdot \Delta h_l \qquad (8)$$

where $\mathrm{VCD_{sim}}$ denotes the simulated VCD, l the TM5 model layer index, c_l the NO_2 concentration in layer l, and Δh_l the
vertical extent of layer l. Usually the summation would be conducted up until layer l_{tp}, the tropopause layer of the TM5 model.
Typically, $l_{\mathrm{tp}} \approx 16$. WRF-Chem, however, is designed to be a tropospheric model, meaning that model results within the
tropopause region should be used with caution. We observe an increase in modelled NO_2 concentrations above 10 km altitude
285 (which equates approximately to layer 13 in the TM5 model, see Fig. A3), which is likely due to stratosphere-troposphere
exchange. We therefore limit the summation of box VCDs to the lowest 13 layers. Then, pairs of simulated and observed
VCDs are obtained by interpolating the WRF-Chem data to the horizontal TROPOMI grid in space (linear) and time (nearest
neighbour).

In order to make a representative comparison between simulated and observed VCDs, the following further aspects are taken
290 into consideration:

1. TROPOMI is not equally sensitive to all layers of the troposphere. This circumstance is described by the averaging
   kernels (AKs, see van Geffen et al., 2022). Thus the AMF depends on relative NO_2 a priori profiles provided by the

TM5 model. It has been demonstrated that replacing TM5 a priori profiles with profiles from high-resolution RCT models improves the VCD retrieval significantly (see e.g. Ialongo et al. (2020); Tack et al. (2021); Liu et al. (2021)).

295 We incorporate the AKs and our simulated high-resolution NO_2 a priori profiles by computing corrected observed VCDs following Eskes et al. (2019):

$$\text{VCD}_{\text{obs, corr}} = \text{VCD}_{\text{obs}} \cdot \frac{\text{AMF}_{\text{trop}}}{\text{AMF}} \cdot \frac{\sum_{l<13} c_l \cdot \Delta h_l}{\sum_{l<13} c_l \cdot \Delta h_l \cdot A_l} \qquad (9)$$

where VCD_{obs} denotes the observed NO_2 VCD, AMF_{trop} the tropospheric AMF, and A the averaging kernel vector.

2. Each observed VCD has an associated q-value that describes the retrieval quality on a range from 0 (bad) to 1 (good)
300   (see van Geffen et al., 2022). q-values of ≥ 0.75 also require the cloud radiance fraction (crf) to be smaller than 0.5, which effectively acts as a cloud filter. The effect of clouds on the measurement process depends on the height of the cloud layer: High clouds can shield off lower layers of the atmosphere, decreasing the NO_2 sensitivity to almost zero below the cloud layer. Vice versa, cloud layers directly below NO_2 layers can enhance the observed NO_2 absorption due to the increased cloud albedo. If cloud layers and NO_2 layers are at the same altitude, NO_2 absorption can be increased
305   by an enhanced light path due to multiple scattering. The influence of clouds on satellite measurements of NO_2 has been discussed e.g. in Martin et al. (2002), Kokhanovsky and Rozanov (2008), and Liu et al. (2021). We apply a q-filter that removes all observations with $q < 0.75$, as recommended by Eskes et al. (2019).

Figure 5 shows the result of our comparison in the form of monthly mean values (i.e. the total average over May 2019). Subfigures a-e show the comparison between simulated NO_2 VCDs and TROPOMI observations using the TM5 NO_2 a priori
310 profiles. Strong enhancements occur in west Germany (Ruhr region) and along the Rhine river. Moderate enhancements occur over logistical hotpots (e.g. the port of Hamburg with a nearby coal power plant), and larger cities (e.g. Berlin and Munich; see Fig. 8 for an overview of the geographical regions mentioned here). Subfigure c shows the difference between simulated and observed VCDs and reveals the model's tendency to overestimate the observed VCDs. The overestimation is close to zero in rural regions but reaches up to 10^{16} molec. cm^{-2} over hotspot regions. The simulated and observed VCDs correlate with
315 an R-value of 0.83, an RMSE of $1.23 \cdot 10^{15}$ molec. cm^{-2}, and a bias of $+17.6$ %. Subfigures f-j show the same comparison, but with recomputed airmass factors according to eq. (9). The comparison improves significantly with an R-value of 0.84, an RMSE of $1.11 \cdot 10^{15}$ molec. cm^{-2}, and a bias of only -6.6 %.

Subfigures k-o show the spatial distribution of simulated and observed surface NO_2 concentrations. Model overestimations of NO_2 VCD and surface concentration occur in similar geographic regions, particularly in the Ruhr region, along the Rhine
320 river, and the port of Hamburg. This hints towards a possible overestimation of NO_x emissions in these regions.

Figure 6 presents further comparisons of simulated and observed NO_2 VCDs in the style of Fig. 5. Here, we only show comparisons using reprocessed observations according to eq. (9). Subfigures a-e show the comparison restricted to Germany, where the high-resolution UBA-E emission inventory was used. Compared to the full domain, an increase in RMSE ($1.46 \cdot 10^{15}$ molec. cm^{-2}) and decrease in bias ($+3.3$%) is observed. The increase in RMSE can be attributed to the Ruhr region in
325 west Germany, which is one of the most polluted regions within our simulation domain. Subfigures f-j show the simulation
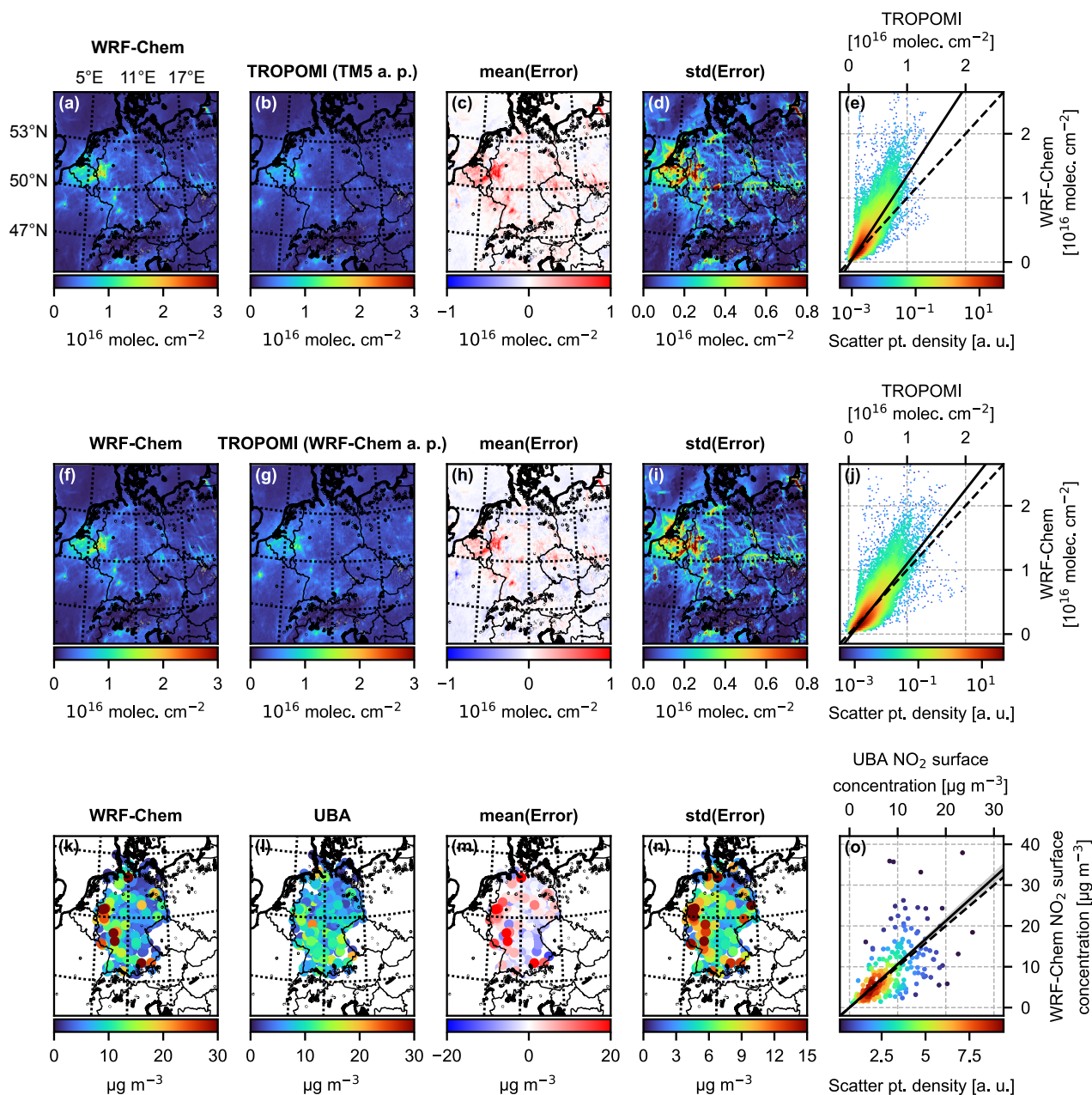
Figure 5. Comparison of simulated and observed NO_2 VCDs **(a)**-**(j)** and surface NO_2 concentrations **(k)**-**(o)**. The first row **(a)**-**(e)** shows the comparison to the original observations using low-resolution TM5 a priori NO_2 profiles. The second row **(f)**-**(j)** shows the comparison with reprocessed observations using high-resolution WRF-Chem a priori profiles and averaging kernels. All satellite observations were restricted to cases with $q \geq 0.75$. The third row **(k)**-**(o)** shows a comparison between simulated and observed surface NO_2 concentrations from "background" stations.
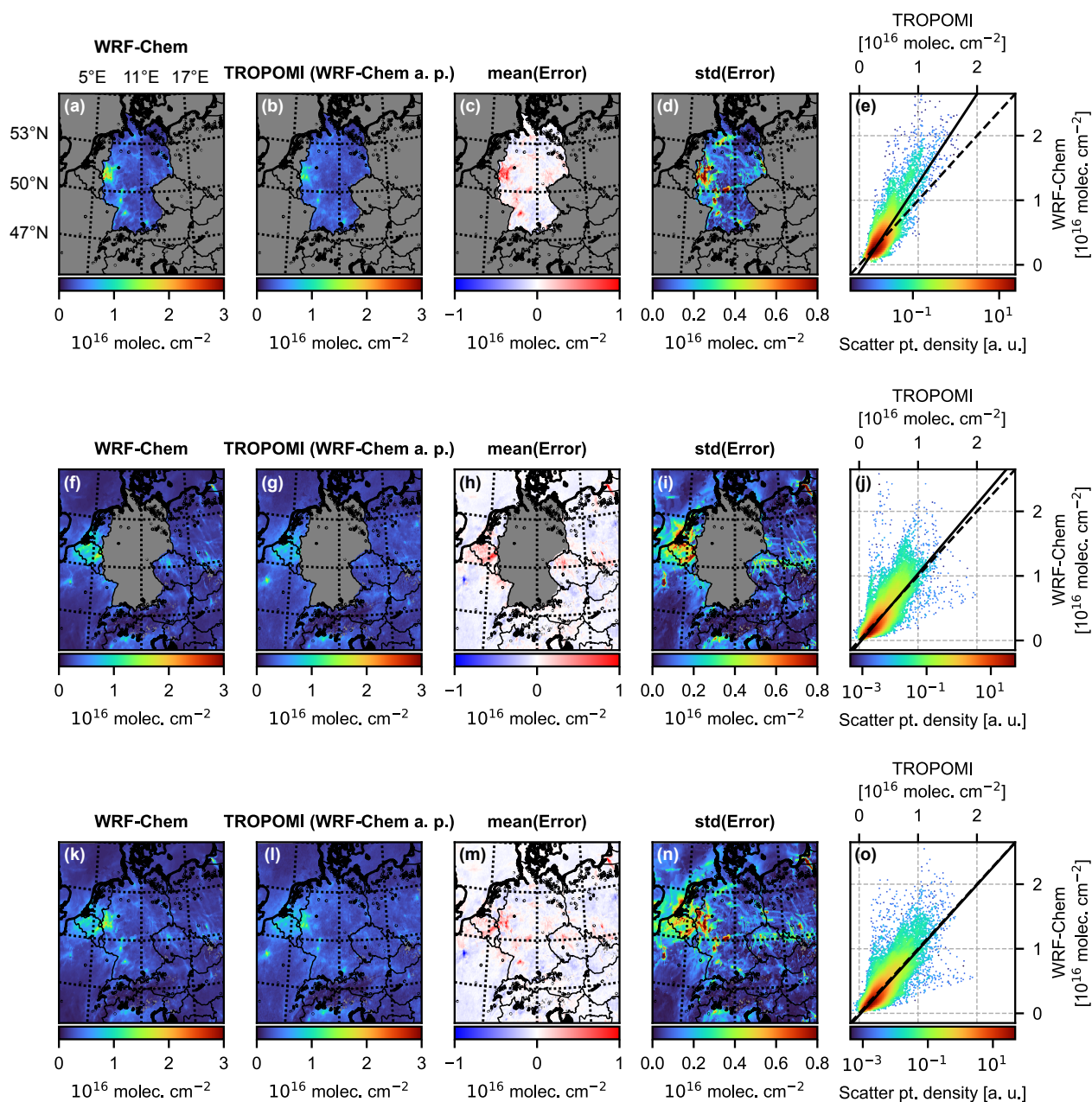
Figure 6. Further comparisons of simulated and observed NO$_2$ VCDs. The first row **(a)**-**(e)** shows the comparison only for Germany. The second row **(f)**-**(j)** shows the comparison for the entire simulation domain except for Germany. The third row **(k)**-**(o)** shows a comparison where instead of the newly optimized temporal profiles, the original ones by Kumar et al. (2021) were used.
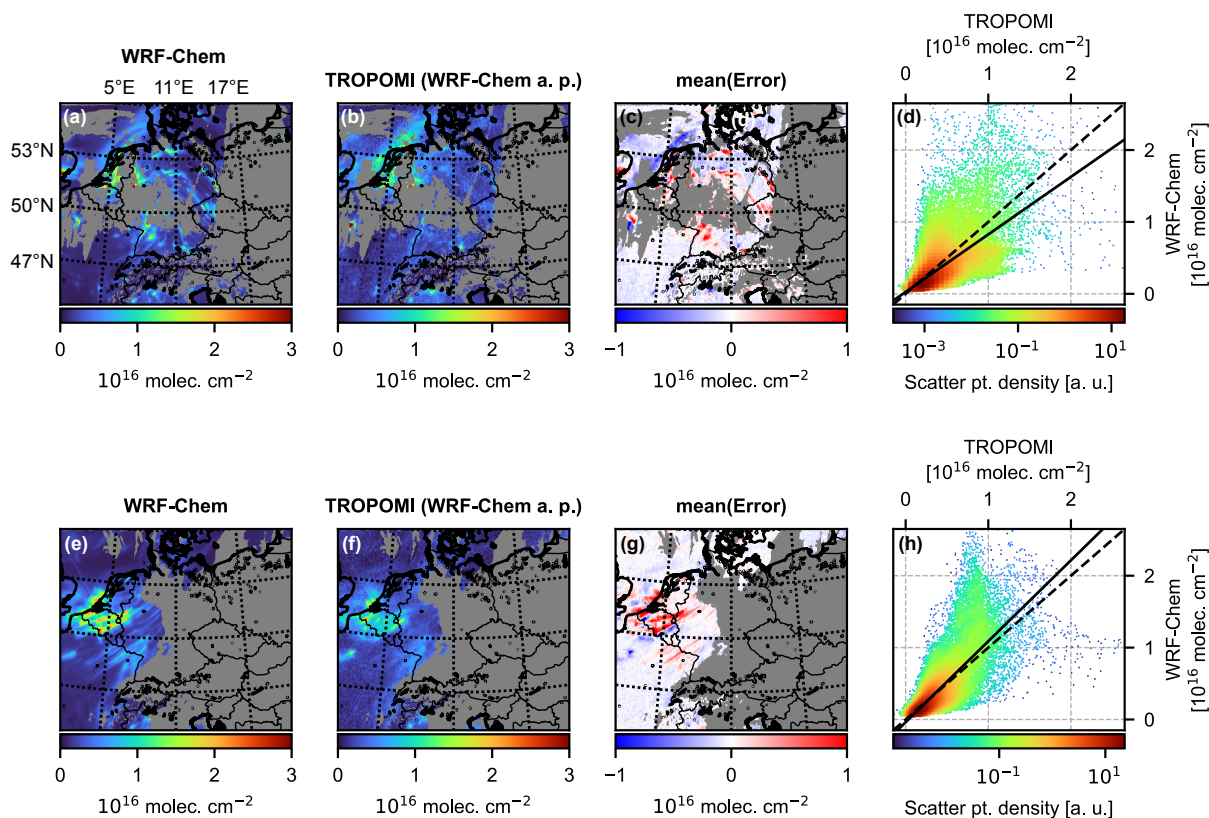
Figure 7. Comparison of simulated and observed NO_2 VCDs for two exemplary individual days. **(a)**-**(d)** shows 23 May 2019, **(e)**-**(h)** shows 15 May 2019. Gray regions are filtered due to the strict masking of clouded pixels.

domain with Germany excluded. Although there are still some strongly polluted regions where WRF-Chem underestimates the NO_2 VCD (e.g. the Netherlands) there are also large regions with model underestimation (e.g. Paris, France, Poland, Austria, and Hungary). This results in an overall mean bias of -9.6% on the displayed domain.

330 Figure 6k-o shows the comparison on the full simulation domain using the original temporal profiles by Kumar et al. (2021), referred to as S-old previously. Although the linear fit in subfigure o is remarkably good with a slope of 1.00, the overall mean bias is -10.2%. This is in agreement with sect. 3.1, where it was shown that S-old suffers from a low bias of -15.7% for NO_2 surface concentrations around noon time. However, in stark contrast to the surface concentrations, which were 16.6 % lower in S-old, the simulated NO_2 VCDs are reduced by only 3.6 % in comparison to S-new. This shows that surface NO_2 concentrations are highly sensitive to the hourly profiles (i.e. the diurnal distribution of emissions), while the integrated NO_2

335 column mostly depends on the total emission budget.

So far only monthly means were considered in the evaluation. Figure 7 shows the comparison for two exemplary individual days: 23 May and 15 May 2019. When looking at single days, large regions of the maps are filtered due to the strict masking of

Table 4. Statistical summary of the results shown in Fig. 5, 6, and 7

Referring to	NO$_2$ a priori profiles	Mean (sim.)	Mean (obs.)	Slope	Intercept	RMSE	Bias	R
Fig. 5a-e	TM5	2.52	2.14	1.39	-0.04	1.23	+17.6 %	0.83
Fig. 5f-j	WRF-Chem	2.52	2.70	1.15	-0.06	1.11	-6.6 %	0.84
Germany only, Fig. 6a-e	WRF-Chem	3.48	3.37	1.38	-0.12	1.46	+3.3 %	0.83
without Germany, Fig. 6f-j	WRF-Chem	2.30	2.54	1.08	-0.04	1.01	-9.6 %	0.83
old temporal profiles, Fig. 6k-o	WRF-Chem	2.41	2.68	1.00	-0.03	0.97	-10.2 %	0.83
single day (23 May 2019, Fig. 7a-d)	WRF-Chem	2.78	3.15	0.80	0.03	2.32	-11.8 %	0.66
single day (15 May 2019, Fig. 7e-h)	WRF-Chem	2.89	2.89	1.13	-0.04	1.92	-0.3 %	0.80

Means and RMSE are given in units of $[10^{15} \text{ molec. cm}^{-2}]$

clouded pixels. This is the case for all days of May 2019. The two displayed days were chosen because they cover the most polluted regions of the simulation domain well. In comparison to monthly means, the modelled NO$_2$ VCDs are smeared out. This

340 is most likely an artefact of the simulated wind directions and speeds. In a monthly mean, such effects are expected to cancel out statistically. Furthermore, a significant decrease in R-value (0.66 and 0.80) and an increase in RMSE ($2.32 \cdot 10^{15}$ molec cm^{-2} and $1.92 \cdot 10^{15}$ molec cm^{-2}) is observed in comparison to monthly means ($R = 0.84$, RMSE $= 1.11 \cdot 10^{15}$ molec cm^{-2}). This is expected, seeing that averaging reduces the statistical noise of both datasets. Noise sources that contribute statistically include the DOAS fit error (from the SCD retrieval), uncertainties of the AMF (e.g. due to clouds) and meteorological model

345 variables. Table 4 summarizes the results of this section.

3.3 Comparison of NO$_2$ concentration profiles

As a third diagnostic we compare modelled NO$_2$ concentration profiles to profiles obtained from MAX-DOAS (Multi-Axis Differential Optical Absorption Spectroscopy) measurements. In contrast to the data shown in sect. 3.1 and 3.2, profile comparison allows for assessment of the model's capability to capture vertical distributions of NO$_2$. MAX-DOAS measurements

350 use the DOAS principle (see Platt and Stutz, 2008) to obtain trace gas differential slant column densities (dSCDs) at different elevation angles. By application of an inversion algorithm a discretized concentration vector c is obtained, whose entries denote the target gas concentration in different atmospheric layers. An overview of different inversion algorithms can be found in Frieß et al. (2019). We use data from the FRM4DOAS network (see Fayt et al., 2021) which applies the Mexican MAX-DOAS fit (MMF, see Friedrich et al., 2019) and the Mainz Profile Algorithm (MAPA, see Beirle et al., 2019). While MMF is based on

355 optimal estimation (see Rodgers, 2000), MAPA uses Monte Carlo simulation in order to determine profile shape parameters which combine (possibly lifted) box profiles and exponential profiles. Purely exponential concentration profiles, however, can not be obtained from MAPA.

Figure 8. Part of central Europe covering Germany, Belgium and Netherlands with important geographical regions marked in red. The blue stars show the locations of the MAX-DOAS measurements reported in sect. 3.3.

Five MAX-DOAS instruments are operated within our simulation domain D2: Mainz (Germany), Bremen (Germany), Heidelberg (Germany), De Bilt (Netherlands), and Uccle (Belgium). Figure 8 shows the locations of these stations. A single station typically yields 2-4 NO_2 profile measurements per hour. In order to compare simulated and observed profiles, the WRF-Chem dataset is interpolated to the geolocations and measurement times of the MAX-DOAS instruments. We use nearest-neighbour interpolation in space and time. The uncertainty of the simulated NO_2 profiles is obtained as the standard deviation of the surrounding eight WRF-Chem grid cells.

The MMF inversion algorithm provides averaging kernels (AKs), represented by a $(h \times h)$ matrix, where h is the number of atmospheric layers considered. Here, $h = 20$, comprising retrieval altitudes up to 4 km. Using the AKs, the MAX-DOAS measuring sensitivity to different altitudes can be applied to the simulated profiles. For this purpose the simulated NO_2 profiles are vertically interpolated to the layers of the MAX-DOAS retrieval grid (linear interpolation). Then, the AKs are applied via

$$c_{\text{sim, corr}} = \mathbf{A}c_{\text{sim}} + (\mathbb{1} - \mathbf{A})c_{\text{ap}} \tag{10}$$

where $c_{\text{sim, corr}}$ is the corrected simulated profile, $\mathbf{A}$ the averaging kernel matrix, c_{sim} the original simulated profile, $\mathbb{1}$ the unity matrix, and c_{ap} the a priori profile see (see Rodgers and Connor, 2003). MAPA neither provides AKs, nor depends on an a priori profile. All profiles flagged as "erroneous" by MAPA were dismissed from the evaluation.

Figure 9 shows averaged NO_2 profiles from the MAX-DOAS station Mainz, Germany, in the time window from 11 AM to 02 PM for a selection of individual days. The aim is to give an overview of the variety of observed and modelled profile shapes.
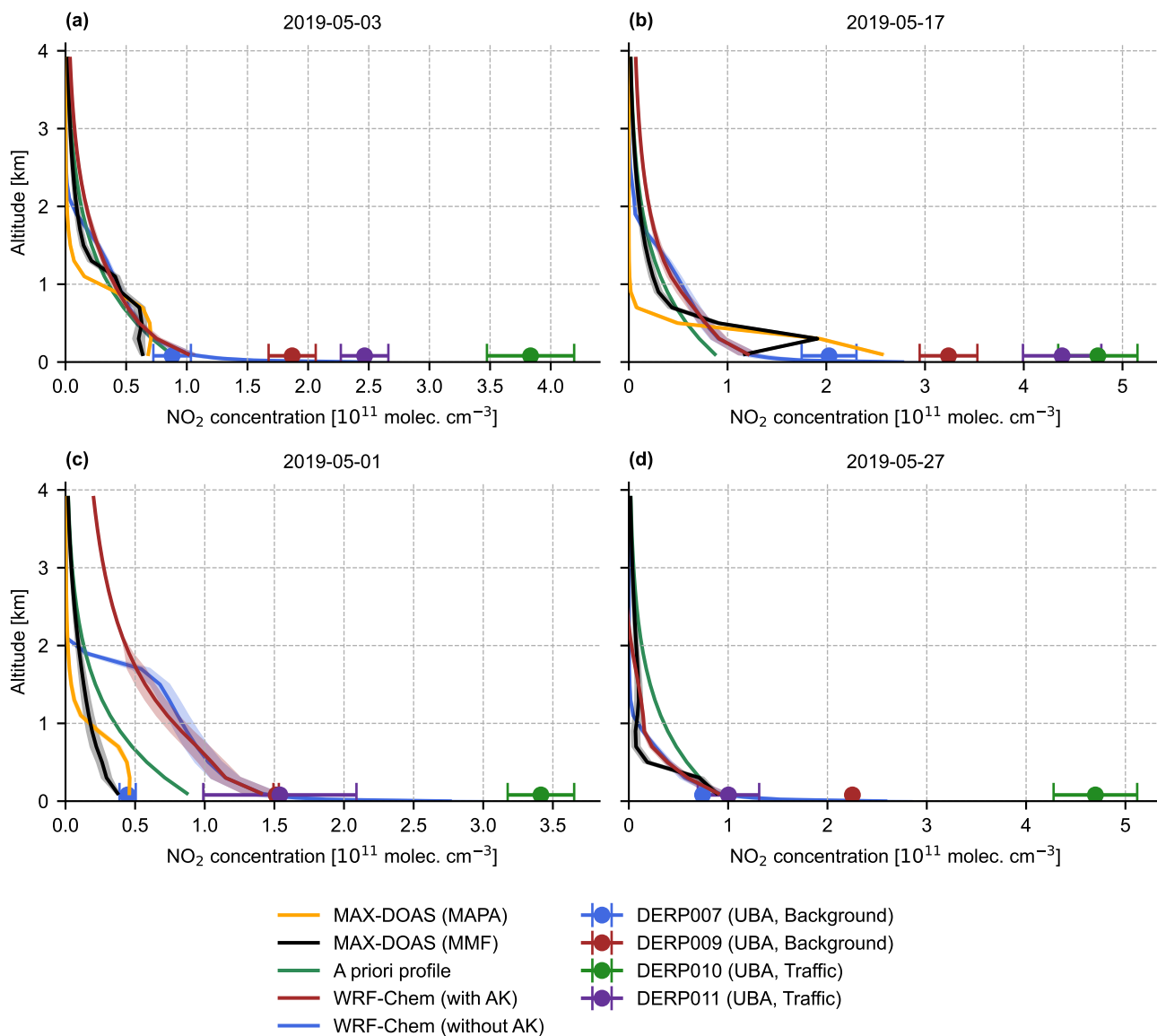
Figure 9. Comparison of exemplary profiles for Mainz around noontime (11 AM - 02 PM) for four different days: **(a)** 03 May 2019, **(b)** 17 May 2019, **(c)** 01 May 2019, **(d)** 27 May 2019. Additionally, average NO_2 surface concentrations measured by UBA measuring stations within a radius of 5 km of the MAX-DOAS instrument are drawn as colored scatter markers at 0 m altitude.

Additionally, the average NO_2 surface concentrations measured by UBA stations within a radius of 5 km of the MAX-DOAS

375   instrument are drawn as colored scatter markers at 0 m altitude.

Figure 9a shows a typical scenario: On this day (03 May 2019) observation and simulation show good general agreement. All three datasets (MMF, MAPA, and WRF-Chem) yield profiles of similar shape. Above an altitude of ~ 2 km the profiles quickly approach zero, which is characteristic for the transitioning regime between the planetary boundary layer and the free troposphere. At low altitudes of 0-150 m the simulated profile has a strong exponential tail, typical for the surface layer into which most NO_x emissions are injected (the lowest ~ 30 m). Furthermore, the simulated profiles show good agreement with the measurements of the nearby UBA in-situ stations. High vertical resolution turns out to be crucial for the comparison to the observed surface concentrations: In the FRM4DOAS dataset, MAPA and MMF were operated with a vertical resolution of 200 m. In a direct comparison, the concentrations obtained from MAPA or MMF in the lowest layer (0-200 m) underestimate the in-situ observations at the surface because of the limited vertical resolution of the MAX-DOAS measurements and the rather coarse retrieval grid used by FRM4DOAS. Meanwhile, the lowest layer of our WRF-Chem simulation only spans 0 - 8 m, which allows for a much more representative comparison.

Subfigure 9b shows a day on which an elevated NO_2 layer was detected by MMF. This occurs regularly at all MAX-DOAS stations in this study and is characterized by strongly enhanced NO_2 concentrations at higher altitudes (here at ~ 500 m). Elevated layers are typically caused by elevated emissions, e.g. from a power plant stack at a few hundred meters height. Additionally, transport events which advect NO_2 from the surface layer could be the cause. However, no corresponding enhancements can be seen in the simulated NO_2 profile. As described in sect. 2.1, we do not use vertical emission profiles in the WRF-Chem simulation. Furthermore, it is possible that the overall spatial resolution of our simulation limits its ability to model comparable elevated trace gas abundances. MAX-DOAS instruments typically have a field of view of $\sim 1°$, which is small enough to resolve e.g. plumes from power plant stacks. Our WRF-Chem simulation has a spatial resolution of $3\,\mathrm{km} \times 3\,\mathrm{km} \times 100\,\mathrm{m}$ at ~ 500 m altitude and therefore dilutes trace gas concentrations into comparably large grid cell volumes. Lastly, horizontal concentration gradients can be expected near strong emission sources. These can lead to apparently elevated profiles, because the MAX-DOAS profile inversion makes the simplifying assumption of horizontally homogeneous distributions. Comparison of simulated and observed elevated urban emissions is therefore an advanced problem and not further adressed here.

Subfigure 9c shows a day on which the agreement between model and observations was poor, with deviations of > 100 %. Such days are outliers and do not represent the overall quality of the model or the measurements, however, they should not remain unmentioned. Possible reasons for such deviations include e.g. falsely modelled wind directions (affecting trace gas transport) or cloud cover (affecting photolysis).

Subfigure 9d shows a day on which the modelled profile has an extremely steep exponential gradient in combination with a relatively thin boundary layer. At the transition from the boundary layer to the surface layer (~ 30 - 50 m above ground), the simulated profiles shows an increase of almost 400 %. Interestingly, on this day (27 May 2019), one of the background stations (DERP009) measured more than twice the surface concentration compared to a nearby traffic station (DERP011). This phenomenon was observed on only one other day (10 May 2019) of the simulated month, on which very similar profile shapes were observed. Furthermore, no MAPA profile is available on this day due to the filtering by error flags. Also, there is good agreement between modelled and observed profiles above the surface layer. It is therefore plausible to assume that the steep

Table 5. Statistical summary of the results shown in Fig. 10 and 11

Location	Mean (sim.)	Mean (obs.)	Slope	Intercept	RMSE	Bias
Mainz	3.05	2.03	1.32	0.38	1.35	+50.3 %
Bremen	1.83	1.94	1.15	-0.41	1.42	-5.9 %
Heidelberg	2.57	1.98	0.97	0.66	0.93	+30.1 %
De Bilt	2.40	2.05	1.16	0.02	0.54	+16.9 %
Uccle	4.11	2.30	1.58	0.48	2.44	+78.6 %

Means, intercept, and RMSE are given in units of $[10^{10} \text{ molec. cm}^{-3}]$.

NO_2 gradient towards the surface is not due to a faulty model, but, on the contrary, indicates the model's ability to capture scenarios in which e.g. vertical mixing is mostly suppressed and large portions of the emitted NO_2 remain in the surface layer.

In order to condense the remaining evaluation, we will focus on monthly noontime averages from hereon, i.e. plots of the same structure as shown in Fig. 9, but averaged over the entire simulation period of May 2019 (Fig. 10 and 11). In both figures the right-side panel shows scatter plots of averaged NO_2 concentrations (WRF-Chem vs. MMF, due to availability of AKs) at different altitudes.

Figure 10 shows the results of the three German MAX-DOAS locations Mainz, Bremen, and Heidelberg. Overall, the three stations show qualitatively similar profiles. At Bremen (see subfigures c-d), a consistent elevated layer at ~ 1000 m altitude was detected, which could be explained by multiple power plants in the near vicinity of the instrument. This has been discussed in detail by Bösch (2018). The scatter plots (subfigures b, d, f) show a strong correlation between simulation and measurements and yield linear fits with slopes of 1.32 for Mainz, 1.15 for Bremen, and 0.97 for Heidelberg. The results shown in Fig. 11 show similar agreement in De Bilt but larger deviations in Uccle. As a result, the slope of 1.58 for the linear fit in Fig. 11d is significantly larger than for the other stations. Table 5 gives a statistical summary of the results shown in Fig. 10 and Fig. 11.

As an additional diagnostic, a correlation analysis between observed (MMF) and simulated (with AKs applied) NO_2 concentrations in the lowest 0-200 m above the surface is shown in Appendix A (see Fig. A4).

4   Discussion and conclusion

We have presented new WRF-Chem simulation results with specific focus on NO_2. The simulation was run on a domain over central Europe with a spatial resolution of $3 \text{ km} \times 3 \text{ km}$ for the month of May 2019. Over Germany, a new emission inventory from the German Environmental Agency with a resolution of $1 \text{ km} \times 1 \text{ km}$ was used. Outside of Germany, the EDGARv5 emission inventory with a resolution of $11 \text{ km} \times 7 \text{ km}$ was used. A new set of temporal emission profiles for the most important emission sectors (traffic, power industry, agricultural soils, energy for buildings, and manufacturing industry) was determined by comparing simulated NO_x surface concentrations to corresponding in-situ measurements in Germany. Once good agreement was found, the model results were also compared to satellite measurements of the NO_2 VCD and NO_2 concentration profiles
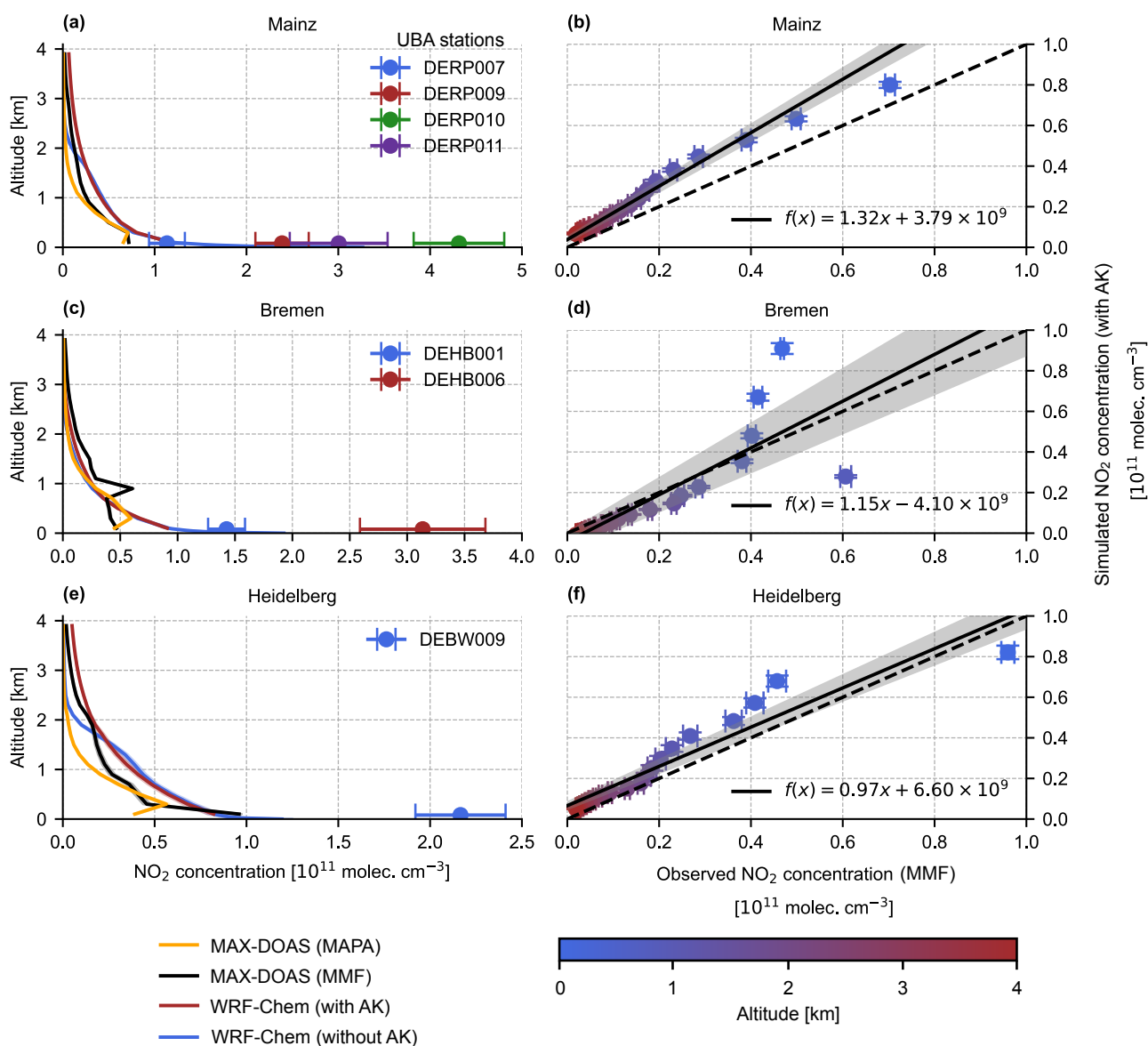
Figure 10. Comparison of averaged NO$_2$ profiles for Mainz, Bremen, and Heidelberg around noontime (11 AM - 02 PM). The left column (a), (c), (e) shows the NO$_2$ profiles as obtained by MMF, MAPA, and WRF-Chem. The right column (b), (d), (f) shows the corresponding scatter plots of averaged NO$_2$ concentrations at different altitudes (here: WRF-Chem vs. MMF).
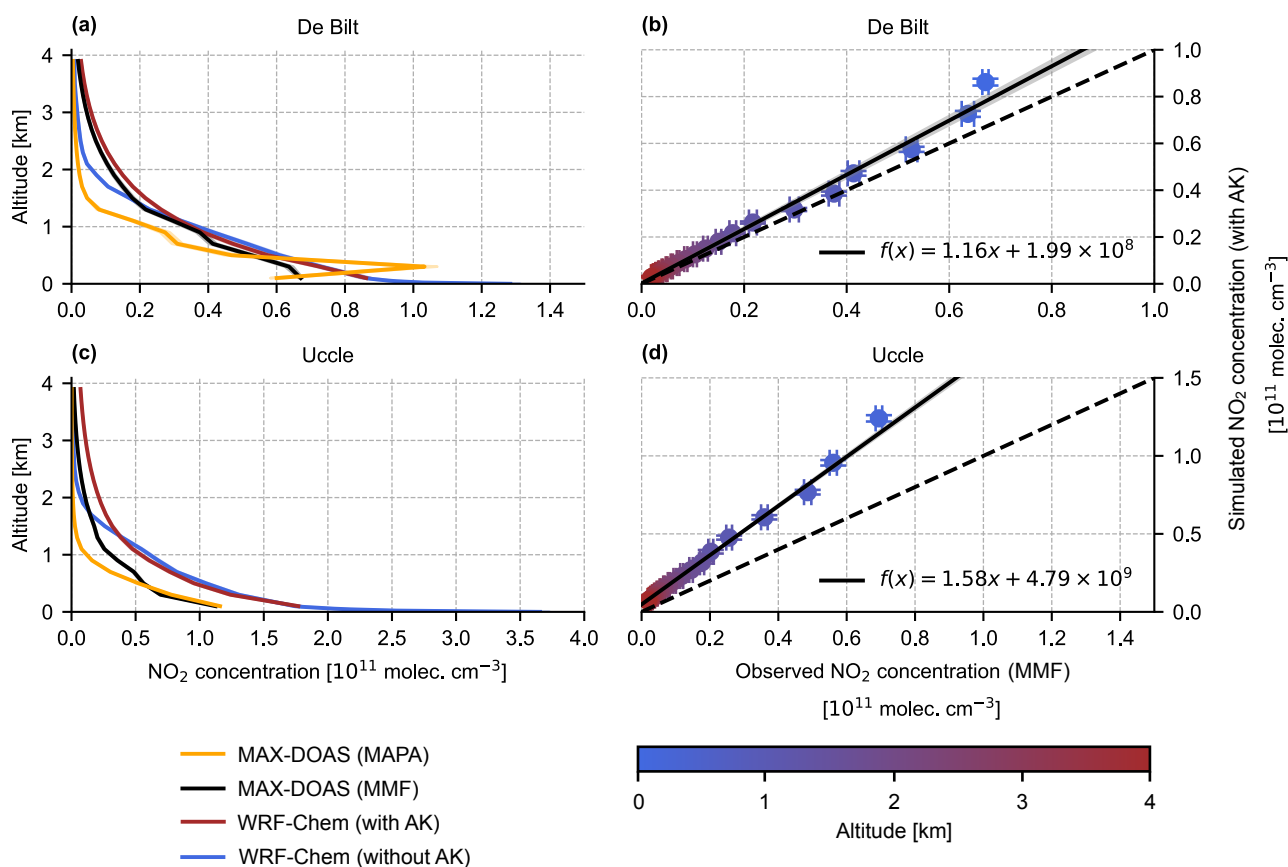
Figure 11. Like Fig. 10, but for the stations De Bilt and Uccle.

obtained from MAX-DOAS measurements. The main focus of the study was to investigate, whether RCT simulations can be improved by fine-tuning of the temporal emission profiles and whether good agreement with all three observational datasets

435   can be achieved.

In sect. 3.1 modelled surface concentrations of NO_2, NO, NO_x and O_3 were compared to in-situ measurements. Here, we used simulation results of two WRF-Chem runs: Firstly, a run with conventional temporal profiles used in previously published literature, and secondly, a run with our newly optimized temporal profiles. The optimized simulation showed significant improvements over the conventional simulation. In particular, the diurnal variation of trace gas concentration was much more

440   accurately reproduced using the new temporal profiles. NO_x biases were reduced from +23.4 % to +6.7 % (all hours of the day considered) and from -19.1 % to +0.4 % (at noontime). NO_2 biases were reduced from +34.0 % to +14.3 % (all hours of the day considered) and from -15.7 % to +0.9 % (at noontime). The corresponding correlation coefficients also slightly improved. This way, the noon-time underestimations of NO_x and NO_2 could be resolved without any changes to the total emission budget.

Underestimations like these were identified in recent literature, but have remained an open issue so far. Despite the significant improvements in NO_x and NO_2 modelling accuracy, some discrepancies remain: While NO_x concentrations appear to be accurate throughout the day, the NO_2/NO ratio is too large before and after noontime. It was found that this problem can not be resolved by changing the speciation ratio of NO_x, which can be explained by the rapid conversion of NO to NO_2 directly after emission. It is therefore more likely that this issue relates to chemistry: Both O_3 and VOCs oxidize NO to NO_2, and a slight model overestimation of O_3 (+10.9 %) was found. Additionally, O_3 concentrations are influenced by VOCs. However, the reference data includes no measurements of VOCs, which means no direct validation of the modelled VOC concentrations is possible. If such measurements become available in the future, the methods described in this article can be extended to further improve the simulation results.

In sect. 3.2 modelled tropospheric NO_2 VCDs were compared to TROPOMI measurements. In order to make a representative comparison, the air mass factors of the TROPOMI retrieval were recomputed using the high-resolution NO_2 profiles from our simulation. The distribution of modelled and observed NO_2 VCDs was found to be similar and agrees with previously reported simulation results. Monthly means of simulated and observed VCDs were shown to correlate strongly ($R = 0.84$) with a model bias of -6.6%. Recomputing the air mass factors using the NO_2 profiles from the regional model was found to be crucial for good agreement: With the original air mass factors the model bias was much larger (+17.6 %). Comparison between the spatial distribution of NO_2 VCDs and surface concentration revealed that in regions where the simulation overestimated the VCDs, it also tended to overestimate the surface concentrations. This was observed in the strongly polluted regions of west Germany and the Netherlands. On the other hand, a general tendency for model underestimation in less polluted regions was found. This observation hints towards faulty NO_x emissions in the corresponding regions. While surface concentrations were shown to be highly sensitive to the choice of temporal profiles, only a weak dependence was found for the NO_2 VCDs. In a comparison to the simulation run with conventional temporal profiles, NO_2 VCDs were only +3.6 % larger on average. The surface concentrations, on the other hand, were 16.6 % larger.

In sect. 3.3 modelled noontime NO_2 concentration profiles were compared to profiles from MAX-DOAS measurements at 5 locations in central Europe from the FRM4DOAS network. The inversion algorithms MMF and MAPA were used. General agreement in the overall shape and magnitude of the profiles was found, with per-station concentration biases of +50.3 % (Mainz), −5.9 % (Bremen), +30.1 % (Heidelberg), +16.9 % (De Bilt) and +78.6 % (Uccle). Two qualitative differences between observed and simulated profiles (as produced by WRF-Chem, i.e. without application of averaging kernels) were identified: Firstly, the simulated profiles show steep NO_2 gradients close to the surface and, furthermore, agree well with the collocated surface concentration measurements. The observed profiles do not resolve the surface layer, most likely due to the coarse retrieval grid with 200 m layer height. As a consequence, the observed profiles tend to underestimate the surface concentration measurements. Secondly, elevated NO_2 layers were identified in some of the observed profiles, but not in the corresponding simulated profiles. Here, the lack of vertical emission profiles, the overall spatial resolution of our WRF-Chem simulation, and the influence of horizontal concentration gradients on the MAX-DOAS inversion procedure were identified as possible explanations.

We draw two main conclusions: Firstly, our study shows that the model accuracy for NO_x and NO_2 in RCT simulations can be strongly improved by optimizing the corresponding hourly emission profiles. In particular, the typical underestimations of

480 noontime NO_x and NO_2 can be resolved. We have described a flexible manual approach for temporal profile tuning, which only requires an iterative chain of model forward runs and comparison to corresponding in-situ observations. Secondly, the intercomparison of the observational datasets used for model validation (here: in-situ measurements, satellite observations, and MAX-DOAS measurements) are overall consistent and agree with our simulation results.

It is important to emphazise the scope of our study. We have shown exemplary temporal profiles and simulation results

485 which *only* represent central Europe in mid-summer. Influences such as regulations for industrial processes, changes in the prevailing energy sources, or reduction of cars as a primary means of personal transport may impact the hourly profiles in the future. Furthermore, it is likely that the hourly profiles inhibit a seasonal dependence. Therefore, the temporal profiles we have presented here should not be seen as generally valid, but rather as a specific result obtained in the framework of our case study for the purpose of proving how sensitive simulation results are to them. However, our general procedure of finding temporal

490 profiles can be applied wherever suitable observational data is available, making the method viable for large parts of the world and other seasons. In the future, our method of deriving temporal profiles could be made more rigorous by experimenting with modern optimization algorithms. This will take time and require the synergy of both, modelling and inversion expertise, but has the potential to make our method applicable to larger domains and simulation periods with less effort.

Data availability. All data are available from the authors upon request.

495 *Author contributions.* LK, TW, SB, and VK developed the question of research. LK and SO carried out the optimization process for the temporal emission profiles. VK, LK, SO, and RK conducted the WRF-Chem simulations. LK, VK, TW, SO, SB, AP, and TB made the comparison between simulation and observational datasets. LK wrote the paper, with all authors contributing by revising it interactively.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the Umweltbundesamt and Deutscher Wetterdienst for providing in-situ measurement data and the

500 UBA emission inventory. Furthermore, we acknowledge Udo Frieß and François Hendrick for providing access to and guidance through the FRM4DOAS datasets. Data analysis and visualization were performed in Python 3.9 including standard libraries such as numpy, scipy, pandas, netCDF4, matplotlib, and basemap. For processing of emission data the HERMES software was used. One of the co-authors (R. Kumar) is from the National Center for Atmospheric Research, which is sponsored by the National Science Foundation of the United States.

References

505   Anenberg, S., Mohegh, A., Goldberg, D., Kerr, G., Brauer, M., Burkart, K., Hystad, P., Larkin, A., Wozniak, S., and Lamsal, L.: Long-term trends in urban NO_2 concentrations and associated paediatric asthma incidence: estimates from global datasets, The Lancet Planetary Health, 6, https://doi.org/10.1016/s2542-5196(21)00255-2, 2022.

Beckwith, M., Bates, E., Gillah, A., and Carslaw, N.: NO_2 hotspots: Are we measuring in the right places?, Atmospheric Environment: X, 2, 100 025, https://doi.org/10.1016/j.aeaoa.2019.100025, 2019.

510   Beirle, S., Dörner, S., Donner, S., Remmers, J., Wang, Y., and Wagner, T.: The Mainz profile algorithm (MAPA), Atmospheric Measurement Techniques, 12, 1785–1806, https://doi.org/10.5194/amt-12-1785-2019, 2019.

Bieser, J., Aulinger, A., Matthias, V., Quante, M., and van der Gon, H.: Vertical emission profiles for Europe based on plume rise calculations, Environmental Pollution, 159, 2935–2946, https://doi.org/10.1016/j.envpol.2011.04.030, 2011.

Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J.: Validation of urban NO_2 concentrations
515   and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities, Atmospheric Chemistry and Physics, 9, 3867–3879, https://doi.org/10.5194/acp-9-3867-2009, 2009.

Bösch, T.: Detailed analysis of MAX-DOAS measurements in Bremen: Spatial and temporal distribution of aerosols, formaldehyde and nitrogen dioxide, Ph.D. thesis, Universität Bremen, 2018.

Chin, M., Rood, R. B., Lin, S.-J., Müller, J.-F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global
520   model GOCART: Model description and global properties, Journal of Geophysical Research: Atmospheres, 105, 24 671–24 687, https://doi.org/10.1029/2000jd900384, 2000.

Chowdhury, S., Haines, A., Klingmüller, K., Kumar, V., Pozzer, A., Venkataraman, C., Witt, C., and Lelieveld, J.: Global and national assessment of the incidence of asthma in children and adolescents from major sources of ambient NO_2, Environmental Research Letters, 16, 035 020, https://doi.org/10.1088/1748-9326/abe909, 2021.

525   Costantini, M. G., Khalek, I., McDonald, J. D., and van Erp, A. M.: The Advanced Collaborative Emissions Study (ACES) of 2007- and 2010-Emissions Compliant Heavy-Duty Diesel Engines: Characterization of Emissions and Health Effects, Emission Control Science and Technology, 2, 215–227, https://doi.org/10.1007/s40825-016-0046-y, 2016.

Crippa, M., Guizzardi, D., Oreggioni, G., Muntean, M., and Schaaf, E.: EDGARv5.0 Air Pollutant Emissions, https://doi.org/10.1594/PANGAEA.921922, 2020a.

530   Crippa, M., Solazzo, E., Huang, G., Guizzardi, D., Koffi, E., Muntean, M., Schieberle, C., Friedrich, R., and Janssens-Maenhout, G.: High resolution temporal profiles in the Emissions Database for Global Atmospheric Research, Scientific Data, 7, https://doi.org/10.1038/s41597-020-0462-2, 2020b.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., Martini, F. S., Sheehy, P. M., Zahniser, M. S., Shorter, J. H., Wormhoudt, J. C., Lamb, B. K., Allwine, E. J., Gaffney, J. S., Marley, N. A., Grutter, M., Marquez, C., Blanco, S., Cardenas, B., Retama, A., Villegas,
535   C. R. R., Kolb, C. E., Molina, L. T., and Molina, M. J.: Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment, Atmospheric Chemistry and Physics, 7, 2691–2704, https://doi.org/10.5194/acp-7-2691-2007, 2007.

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., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geoscientific Model Development, 3, 43–67, https://doi.org/10.5194/gmd-3-43-2010,
540   2010.

Emmons, L. K., Schwantes, R. H., Orlando, J. J., Tyndall, G., Kinnison, D., Lamarque, J.-F., Marsh, D., Mills, M. J., Tilmes, S., Bardeen, C., Buchholz, R. R., Conley, A., Gettelman, A., Garcia, R., Simpson, I., Blake, D. R., Meinardi, S., and Pétron, G.: The Chemistry Mechanism in the Community Earth System Model Version 2 (CESM2), Journal of Advances in Modeling Earth Systems, 12, https://doi.org/10.1029/2019ms001882, 2020.

545 Eskes, H., van Geffen, J., Sneep, M., Apituley, A., and Veefkind, J.: Sentinel-5 precursor/TROPOMI Level 2 Product User Manual Nitrogen-dioxide, Royal Netherlands Meteorological Institute, 2019.

Eskes, H., van Geffen, J., Sneep, M., Veefkind, P., Niemeijer, S., and Zehner, C.: S5P Nitrogen Dioxide v02.03.01 intermediate reprocessing on the S5P-PAL system: Readme file, 2021.

Faustini, A., Rapp, R., and Forastiere, F.: Nitrogen dioxide and mortality: review and meta-analysis of long-term studies, European Respiratory Journal, 44, 744–753, https://doi.org/10.1183/09031936.00114713, 2014.

Fayt, C., Friedrich, M., and Hendrick, F.: Fiducial Reference Measurements for Ground-Based DOAS Air-Quality Observations, Royal Belgian Institute for Space Aeronomy, https://frm4doas.aeronomie.be/ProjectDir/FRM4DOAS_CCN02_D20_MAXDOAS_Network_Operational_Processing_System_Architecture_Design_Document__v2.0_20210903.pdf, 2021.

Freitas, S. R., Longo, K. M., Chatfield, R., Latham, D., Dias, M. A. F. S., Andreae, M. O., Prins, E., Santos, J. C., Gielow, R., and Carvalho, J. A.: Including the sub-grid scale plume rise of vegetation fires in low resolution atmospheric transport models, Atmospheric Chemistry and Physics, 7, 3385–3398, https://doi.org/10.5194/acp-7-3385-2007, 2007.

Friedrich, M. M., Rivera, C., Stremme, W., Ojeda, Z., Arellano, J., Bezanilla, A., García-Reynoso, J. A., and Grutter, M.: NO_2 vertical profiles and column densities from MAX-DOAS measurements in Mexico City, Atmospheric Measurement Techniques, 12, 2545–2565, https://doi.org/10.5194/amt-12-2545-2019, 2019.

560 Frieß, U., Beirle, S., Bonilla, L. A., Bösch, T., Friedrich, M. M., Hendrick, F., Piters, A., Richter, A., van Roozendael, M., Rozanov, V. V., Spinei, E., Tirpitz, J.-L., Vlemmix, T., Wagner, T., and Wang, Y.: Intercomparison of MAX-DOAS vertical profile retrieval algorithms: studies using synthetic data, Atmospheric Measurement Techniques, 12, 2155–2181, https://doi.org/10.5194/amt-12-2155-2019, 2019.

German Environmental Agency (a): Private communication, 7th April 2022.

German Environmental Agency (b): Current air data, http://umweltbundesamt.de/en/data/air/air-data/stations, last access: 26th August 2022.

565 Grell, G. A. and Dévényi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophysical Research Letters, 29, 38–1–38–4, https://doi.org/10.1029/2002gl015311, 2002.

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

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmospheric Chemistry and Physics, 6, 3181–3210, https://doi.org/10.5194/acp-6-3181-2006, 2006.

Hersbach, H. and Dee, D.: ERA5 reanalysis in production, ECMWF Newsletter 147, 2017.

Hong, S.-Y.: A new stable boundary-layer mixing scheme and its impact on the simulated East Asian summer monsoon, Quarterly Journal of the Royal Meteorological Society, 136, 1481–1496, https://doi.org/10.1002/qj.665, 2010.

575 Huang, G., Brook, R., Crippa, M., Janssens-Maenhout, G., Schieberle, C., Dore, C., Guizzardi, D., Muntean, M., Schaaf, E., and Friedrich, R.: Speciation of anthropogenic emissions of non-methane volatile organic compounds: a global gridded data set for 1970–2012, Atmospheric Chemistry and Physics, 17, 7683–7701, https://doi.org/10.5194/acp-17-7683-2017, 2017.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research, 113, https://doi.org/10.1029/2008jd009944, 2008.

Ialongo, I., Virta, H., Eskes, H., Hovila, J., and Douros, J.: Comparison of TROPOMI/Sentinel-5 Precursor NO_2 observations with ground-based measurements in Helsinki, Atmospheric Measurement Techniques, 13, 205–218, https://doi.org/10.5194/amt-13-205-2020, 2020.

Jimenez, J. L., Mcrae, G. J., Nelson, D. D., Zahniser, M. S., and Kolb, C. E.: Remote Sensing of NO and NO_2 Emissions from Heavy-Duty Diesel Trucks Using Tunable Diode Lasers, Environmental Science & Technology, 34, 2380–2387, https://doi.org/10.1021/es9911622, 2000.

Kerkweg, A. and Jöckel, P.: The 1-way on-line coupled atmospheric chemistry model system MECO(n) – Part 1: Description of the limited-area atmospheric chemistry model COSMO/MESSy, Geoscientific Model Development, 5, 87–110, https://doi.org/10.5194/gmd-5-87-2012, 2012.

Kokhanovsky, A. and Rozanov, V.: The uncertainties of satellite DOAS total ozone retrieval for a cloudy sky, Atmospheric Research, 87, 27–36, https://doi.org/https://doi.org/10.1016/j.atmosres.2007.04.006, 2008.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmospheric Chemistry and Physics, 5, 417–432, https://doi.org/10.5194/acp-5-417-2005, 2005.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and van der Gon, H. A. C. D.: TNO-MACC_II emission inventory: a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmospheric Chemistry and Physics, 14, 10 963–10 976, https://doi.org/10.5194/acp-14-10963-2014, 2014.

Kuik, F., Lauer, A., Churkina, G., van der Gon, H. A. C. D., Fenner, D., Mar, K. A., and Butler, T. M.: Air quality modelling in the Berlin–Brandenburg region using WRF-Chem v3.7.1: sensitivity to resolution of model grid and input data, Geoscientific Model Development, 9, 4339–4363, https://doi.org/10.5194/gmd-9-4339-2016, 2016.

Kuik, F., Kerschbaumer, A., Lauer, A., Lupascu, A., von Schneidemesser, E., and Butler, T. M.: Top-down quantification of NO_x emissions from traffic in an urban area using a high-resolution regional atmospheric chemistry model, Atmospheric Chemistry and Physics, 18, 8203–8225, https://doi.org/10.5194/acp-18-8203-2018, 2018.

Kumar, V., Remmers, J., Beirle, S., Fallmann, J., Kerkweg, A., Lelieveld, J., Mertens, M., Pozzer, A., Steil, B., Barra, M., Tost, H., and Wagner, T.: Evaluation of the coupled high-resolution atmospheric chemistry model system MECO(n) using in situ and MAX-DOAS NO_2 measurements, Atmospheric Measurement Techniques, 14, 5241–5269, https://doi.org/10.5194/amt-14-5241-2021, 2021.

Liu, S., Valks, P., Pinardi, G., Xu, J., Chan, K. L., Argyrouli, A., Lutz, R., Beirle, S., Khorsandi, E., Baier, F., Huijnen, V., Bais, A., Donner, S., Dörner, S., Gratsea, M., Hendrick, F., Karagkiozidis, D., Lange, K., Piters, A. J. M., Remmers, J., Richter, A., Van Roozendael, M., Wagner, T., Wenig, M., and Loyola, D. G.: An improved TROPOMI tropospheric NO_2 research product over Europe, Atmospheric Measurement Techniques, 14, 7297–7327, https://doi.org/10.5194/amt-14-7297-2021, 2021.

Madronich, S.: Photodissociation in the atmosphere: 1. Actinic flux and the effects of ground reflections and clouds, Journal of Geophysical Research, 92, 9740, https://doi.org/10.1029/jd092id08p09740, 1987.

Manders, A. M. M., Builtjes, P. J. H., Curier, L., van der Gon, H. A. C. D., Hendriks, C., Jonkers, S., Kranenburg, R., Kuenen, J. J. P., Segers, A. J., Timmermans, R. M. A., Visschedijk, A. J. H., Kruit, R. J. W., 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,

615   A., van Velzen, N., and Schaap, M.: Curriculum vitae of the LOTOS–EUROS (v2.0) chemistry transport model, Geoscientific Model Development, 10, 4145–4173, https://doi.org/10.5194/gmd-10-4145-2017, 2017.

Mar, K. A., Ojha, N., Pozzer, A., and Butler, T. M.: Ozone air quality simulations with WRF-Chem (v3.5.1) over Europe: model evaluation and chemical mechanism comparison, Geoscientific Model Development, 9, 3699–3728, https://doi.org/10.5194/gmd-9-3699-2016, 2016.

Martin, R. V., Chance, K., Jacob, D. J., Kurosu, T. P., Spurr, R. J. D., Bucsela, E., Gleason, J. F., Palmer, P. I., Bey, I., Fiore, A. M., Li, Q.,
620   Yantosca, R. M., and Koelemeijer, R. B. A.: An improved retrieval of tropospheric nitrogen dioxide from GOME, Journal of Geophysical Research, 107, https://doi.org/10.1029/2001jd001027, 2002.

Menut, L., Bessagnet, B., Briant, R., Cholakian, A., Couvidat, F., Mailler, S., Pennel, R., Siour, G., Tuccella, P., Turquety, S., and Valari, M.: The CHIMERE v2020r1 online chemistry-transport model, Geoscientific Model Development, 14, 6781–6811, https://doi.org/10.5194/gmd-14-6781-2021, 2021.

625   Mills, I. C., Atkinson, R. W., Kang, S., Walton, H., and Anderson, H. R.: Quantitative systematic review of the associations between short-term exposure to nitrogen dioxide and mortality and hospital admissions, BMJ Open, 5, e006 946–e006 946, https://doi.org/10.1136/bmjopen-2014-006946, 2015.

Monin, A. and Obukhov, S.: Basic laws of turbulent mixing in the surface layer of the atmosphere, Tr. Akad. Nauk. SSSR Geophiz. Inst., 24, 163–187, 1954.

630   Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, https://doi.org/10.1029/2010jd015139, 2011.

Omrani, H., Drobinski, P., and Dubos, T.: Spectral nudging in regional climate modelling: how strongly should we nudge?, Quarterly Journal of the Royal Meteorological Society, 138, 1808–1813, https://doi.org/10.1002/qj.1894, 2012.

635   Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy, Springer Berlin Heidelberg, https://doi.org/10.1007/978-3-540-75776-4, 2008.

Pozzer, A., Jöckel, P., and Aardenne, J. V.: The influence of the vertical distribution of emissions on tropospheric chemistry, Atmospheric Chemistry and Physics, https://doi.org/https://doi.org/10.5194/acp-9-9417-2009, 2009.

Richmond-Bryant, J., Owen, R. C., Graham, S., Snyder, M., McDow, S., Oakes, M., and Kimbrough, S.: Estimation of on-road NO_2 con-
640   centrations, NO_2/NO_x ratios, and related roadway gradients from near-road monitoring data, Air Quality, Atmosphere & Health, 10, 611–625, https://doi.org/10.1007/s11869-016-0455-7, 2017.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding, WORLD SCIENTIFIC, https://doi.org/10.1142/3171, 2000.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002jd002299, 2003.

645   Steinbacher, M., Zellweger, C., Schwarzenbach, B., Bugmann, S., Buchmann, B., Ordóñez, C., Prevot, A. S. H., and Hueglin, C.: Nitrogen oxide measurements at rural sites in Switzerland: Bias of conventional measurement techniques, Journal of Geophysical Research, 112, https://doi.org/10.1029/2006jd007971, 2007.

Strogies, M., Gniffke, P., and Hausmann, K.: German Informative Inventory Report 2020, http://iir-de-2020.wikidot.com/summary (last access: 13 July 2021), 2020.

650   Tack, F., Merlaud, A., Iordache, M.-D., Pinardi, G., Dimitropoulou, E., Eskes, H., Bomans, B., Veefkind, P., and Van Roozendael, M.: Assessment of the TROPOMI tropospheric NO_2 product based on airborne APEX observations, Atmospheric Measurement Techniques, 14, 615–646, https://doi.org/10.5194/amt-14-615-2021, 2021.

Thompson, G., Field, P. R., Rasmussen, R. M., and Hall, W. D.: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization, Monthly Weather Review, 136, 5095–5115, https://doi.org/10.1175/2008mwr2387.1, 2008.

Tie, X., Madronich, S., Walters, S., Zhang, R., Rasch, P., and Collins, W.: Effect of clouds on photolysis and oxidants in the troposphere, Journal of Geophysical Research, 108, https://doi.org/10.1029/2003jd003659, 2003.

van Geffen, J., Eskes, H. J., Boersma, K., and Veefkind, J.: TROPOMI ATBD of the total and tropospheric NO_2 data products, Royal Netherlands Meteorological Institute, 2022.

Veefkind, J., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H., de Haan, J., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, https://doi.org/10.1016/j.rse.2011.09.027, 2012.

Visser, A. J., Boersma, K. F., Ganzeveld, L. N., and Krol, M. C.: European NO_x emissions in WRF-Chem derived from OMI: impacts on summertime surface ozone, Atmospheric Chemistry and Physics, 19, 11 821–11 841, https://doi.org/10.5194/acp-19-11821-2019, 2019.

Wesely, M.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmospheric Environment (1967), 23, 1293–1304, https://doi.org/10.1016/0004-6981(89)90153-4, 1989.

Wiedinmyer, C., Akagi, S. K., Yokelson, R. J., Emmons, L. K., Al-Saadi, J. A., Orlando, J. J., and Soja, A. J.: The Fire INventory from NCAR (FINN): a high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625–641, https://doi.org/10.5194/gmd-4-625-2011, 2011.

Wild, R. J., Dubé, W. P., Aikin, K. C., Eilerman, S. J., Neuman, J. A., Peischl, J., Ryerson, T. B., and Brown, S. S.: On-road measurements of vehicle NO_2/NO_x emission ratios in Denver, Colorado, USA, Atmospheric Environment, 148, 182–189, https://doi.org/10.1016/j.atmosenv.2016.10.039, 2017.

Williams, J. E., Boersma, K. F., Le Sager, P., and Verstraeten, W. W.: The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation, Geoscientific Model Development, 10, 721–750, https://doi.org/10.5194/gmd-10-721-2017, 2017.

Appendix A: Additional figures and tables

Table A1. Layer extent of the lowest 24 layers in our WRF-Chem simulation

Layer number	Layer bottom [m]	Layer top [m]	Layer number	Layer bottom [m]	Layer top [m]
1	0	8	13	1370	1546
2	8	33	14	1546	1697
3	33	66	15	1697	1841
4	66	125	16	1841	1937
5	125	209	17	1937	2035
6	209	310	18	2035	2183
7	310	429	19	2183	2374
8	429	575	20	2374	2661
9	575	741	21	2661	3142
10	741	935	22	3142	3907
11	935	1178	23	3907	4762
12	1178	1370	24	4762	5643

Exact layer bottoms and tops depend on location and time. The values given here are averages.

Table A2. Statistical summary of the results shown in Fig. A4

Location	Mean (sim.)	Mean (obs.)	Slope	Intercept	RMSE	Bias	R
Mainz	11.69	7.18	0.07	11.15	10.42	+62.8 %	0.13
Bremen	9.09	4.68	0.64	6.12	8.63	+94.3 %	0.29
Heidelberg	8.21	9.60	0.15	6.78	17.06	-14.5 %	0.38
De Bilt	8.62	6.71	0.69	3.97	6.31	+28.5 %	0.56
Uccle	17.69	11.24	0.57	11.32	13.35	+57.3 %	0.45

Means, intercept, and RMSE are given in units of $[10^{10} \ \mathrm{molec. \ cm^{-3}}]$.
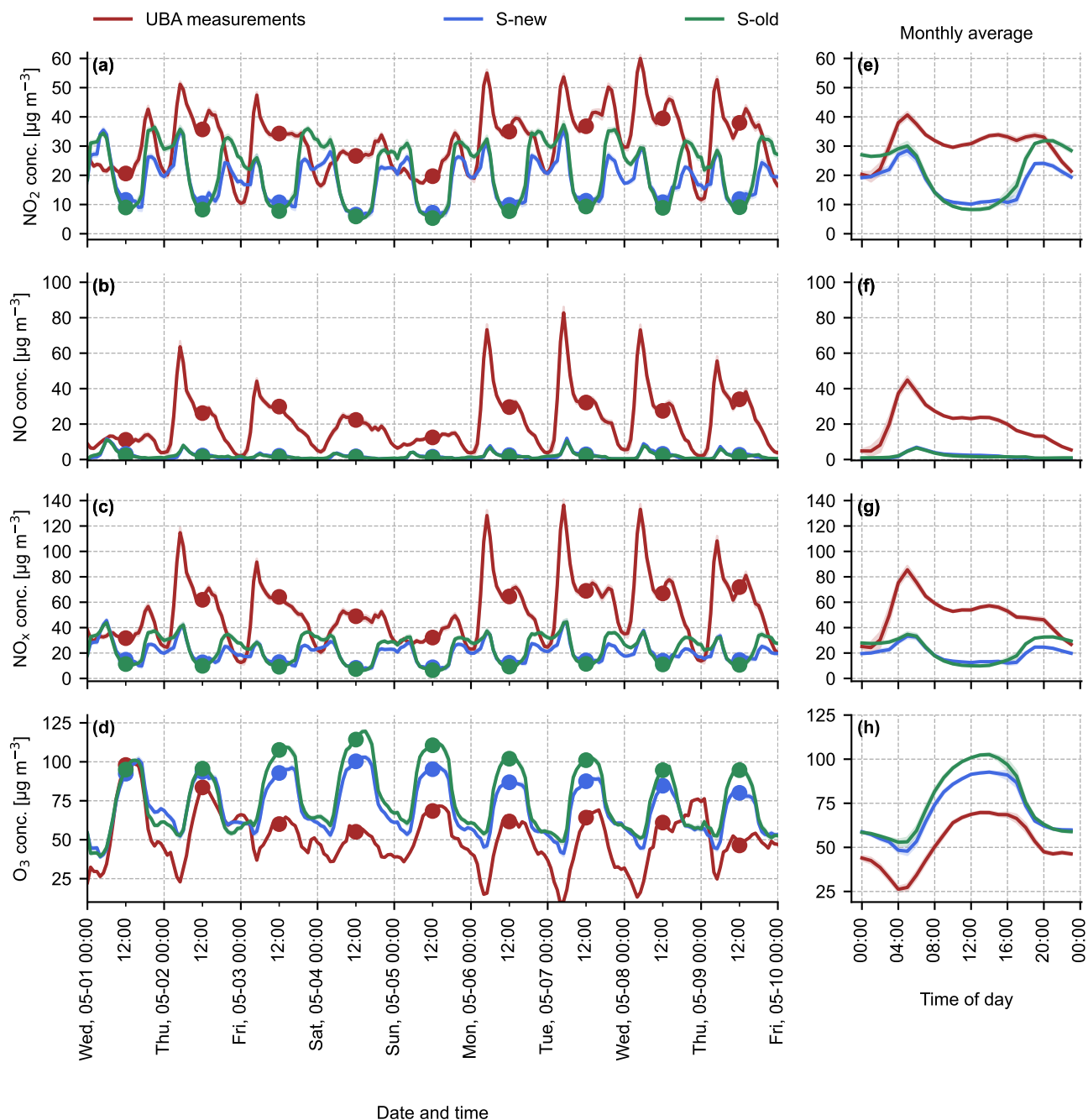
Figure A1. Like Fig. 3, but for traffic stations instead of background stations.
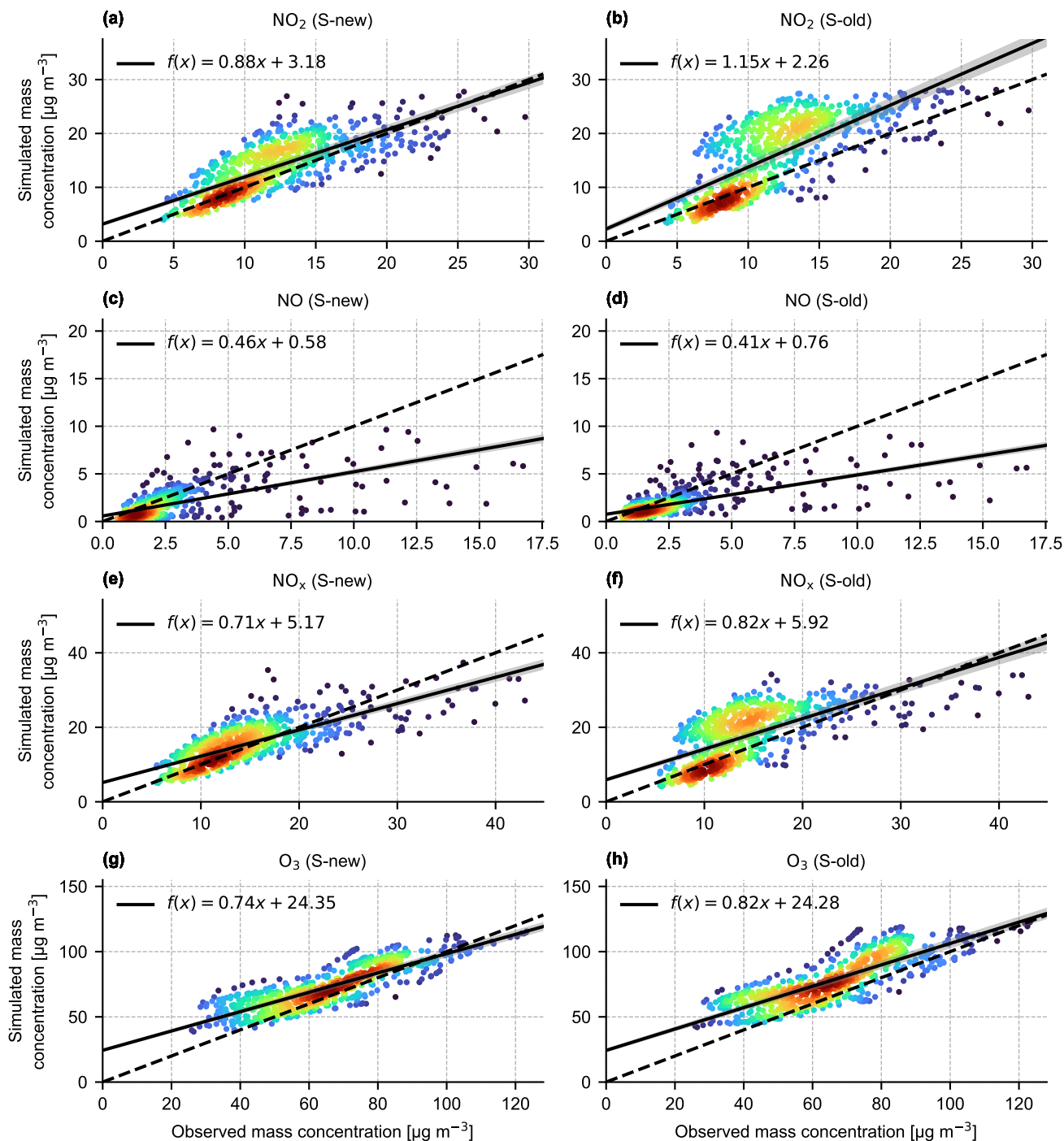
Figure A2. Like Fig. 4, but for all hours of the day instead for noontime only. Error bars were omitted and a color coding (without physical meaning) applied for easier readability.
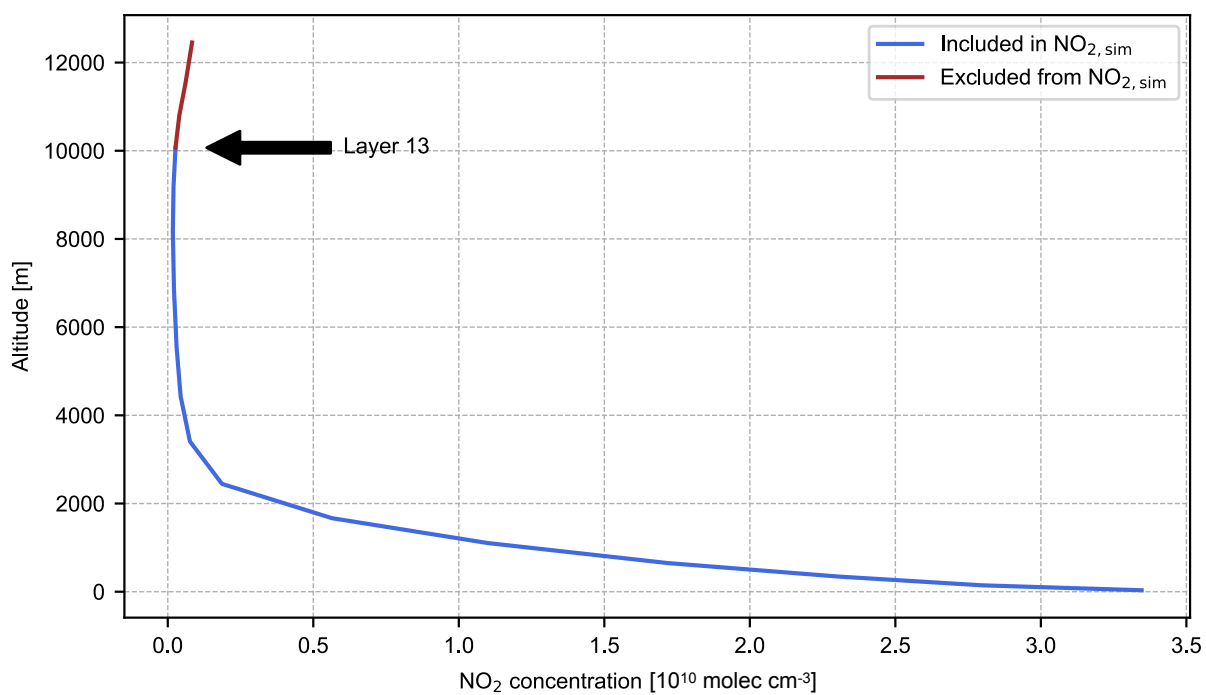
Figure A3. Average simulated NO_2 profile, vertically interpolated to the TM5 grid.
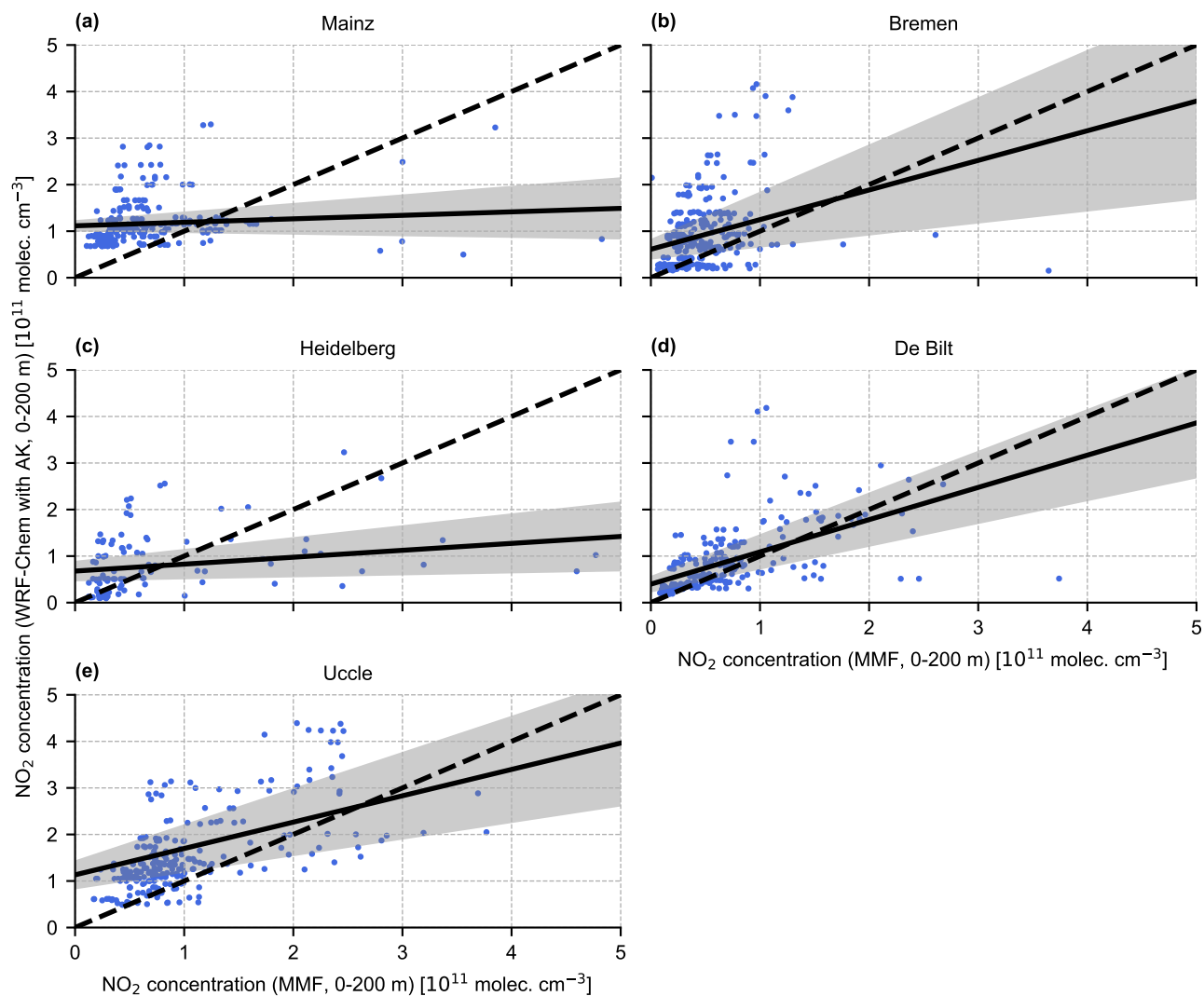
Figure A4. Correlation analysis between observed (MMF) and simulated (averaging kernels applied) NO_2 concentrations in the lowest 0-200 m above the surface. In each subplot, the dashed line marks the 1:1 line and the solid line is a linear fit through the point cloud. The R-values and further statistical diagnostics of each plot are found in Table A2.