Referee comments, answers and changes in the manuscript

General comments

Referee 1

Comparison of the modelled NO2 concentrations in this study with those in Kuik et al (2016) does suggest that the modification to the chemical mixing routines in WRF- Chem in order to set a minimum vertical mixing value helps reduce the overprediction of NO2 at nighttime. I am, however, uncomfortable with the some of the choices made in this routine: using fixed minimum mixing rates, and emissions as a proxy for land use, both seem like clunky fudges to me. CO from shipping emissions would lead to it being activated over the oceans, for example, which I think is counter to the intended influence of this modification. It is good that the authors have noted their use of this modification in this study (and disappointing that if it has been used in previous published studies, that it does not seem to have been documented as being used), as this will add to growing evidence in the literature (c.f. Hu et al, 2013) that this issue exists and needs addressing in a more rigorous and methodical manner.

As changes other than this modification have been made to the model setup between this study and Kuik et al (2016), it would be useful for readers if a figure illustrating the differences in NO, NO2, NOx and O3 concentrations in model simulations with and without this modification could be added to the supplementary material (as this would be illustrative, it would only need to be for a couple of days, not for the entire year).

In addition, recent communication from Georg Grell on the wrf-chem-discussions list indicated that the ACM2 scheme now conducts mixing of chemical species within the PBL physics routines (improving consistency between the treatment of chemical and meteorological tracers). Though this is probably outside the scope of this study, have the authors considered investigating if this scheme improves nighttime pollutant concentrations?

Referee 3

Page 5, lines 1 – 4 and page 10, line 8: This fix of the too weak vertical exchange during nighttime seems quite arbitrary. Is this fix only applied only to pollutant concentrations? Enhanced nocturnal mixing would also affect the thermal stratification which could in turn affect the vertical exchange of pollutants. Therefore, the applied fix should be commented more critically.

Answer

We share the referees' concern on the modification of the chemical mixing routine. This was a very much discussed topic at a workshop of European WRF-Chem users in Germany last year. Several points can be commented on:

- The fix is only applied to pollutant concentrations.
- As the focus of this study is Berlin, which is sufficiently far away from the sea, we do not believe that a potential modification of mixing when shipping emissions are high would impact our results. However, we share the referee's concern. An "easy" (yet still clunky) fix to the shipping emissions problem would be to restrict the increased mixing to grid cells with an urban land use category.
- Illustrative figures have been added to the supplementary material comparing results

from two simulations, one with default TNO diurnal cycles of emissions and no enhanced mixing, and one with diurnal cycles for Berlin and enhanced mixing. In addition, a reference to the supplementary material has been added in the manuscript as indicated below.

- At the time of running the simulations and writing, we were not aware of the new features of the ACM2 scheme, so we did not try it (though we did compare the results from the YSU, MYNN and MYJ schemes). We included the existence of this option in the conclusions as indicated below, note that this could also be an option for a more consistent treatment of meteorology and chemistry over urban areas, and suggest testing it as a follow-up of this study.

Changes in the manuscript:

Page 5, line 4: "Nighttime mixing over urban areas is not accounted for sufficiently by the urban parametrization and the PBL scheme and thus adjusted (dry_dep_driver.F) as described in the Supplementary Material. There, we also show illustrative results (Fig. S1-S3) of two test simulations comparing the impact of changes in this model setup with respect to Kuik et al. (2016), including the modification of nighttime mixing and the modification of the diurnal distribution of emissions (see Sect. 2.4)."

Page 20, line 24: "[…] but is also consistent with deficiencies in other processes varying on the diurnal scale such as the modelled mixing in the planetary boundary layer. The analysis of the model results suggests that the latter is particularly relevant in summer and spring, and that further research is needed in order to better represent urban processes and their coupling with chemistry in WRF-Chem. For example, the changes in the model code applied here to improve nighttime mixing can be critically discussed, and would ideally be replaced by an improved ~~the~~ parameterization of urban processes. The latter would need~~s~~ to better account for urban heat and momentum fluxes for a more realistic representation of mixing both at daytime and at nighttime, particularly in summer. An alternative model configuration to be tested could be the recently extended ACM2 planetary boundary parametrization (Pleim et al., 2007), which now conducts mixing of chemical species within the planetary boundary layer scheme. In addition, measurements of vertical profiles of NOx in urban areas are needed to evaluate and improve models for applications in urban areas."

Referee 1

The locations of measurement stations are not very clear in Figure 2, and it is currently impossible just using this paper to determine the exact location of individual named stations within the study. An extra map (either in the main paper, or in supplementary material), with the locations of named measurements stations made much clearer, would be very useful.

Answer:

Manuscript updated as suggested.

Changes in the manuscript:

To clarify this, a map with station types and airbase codes as well as a table matching the codes with the station names and coordinates has been included in the supplementary material. The material has also been referenced in the main text:

Page 7 lines 12-14: "For the comparison with model results, observations from stations

within Berlin and in the adjacent surroundings in the Federal State of Brandenburg representing “urban background”, “suburban background” and “rural near-city” conditions are used (Fig. 2, also see supplementary Figure S4 and Table S1)”

Referee 1: In Section 6.2 the authors use wind speed and direction data to select data for investigating the impact of changing urban traffic emissions on downwind model NOx predictions. Were there specific reasons for choosing 2 m/s wind speed and 72 data pairs as your cut-off points? How dependent on your data selection criteria are the changes in model bias between simulations - if you adjusted the criteria by +/-10% would your results change greatly?

Answer

The reason for choosing a wind speed threshold was to exclude situations in which the “downwind” station would not be affected by emissions in the city center. Results were only considered for a certain number of data pairs in order to get a sufficiently large number of hourly observations for averaging the results. If the wind speed threshold is changed by +/-10%, the results are not changed greatly, and the ranges of bias reduction indicated in the manuscript are still valid for January, and slightly enlarged for July, with both lower and higher changes (see table included below). Most stations excluded by the threshold of available data pairs have far fewer data pairs available than the threshold, and most stations included have a considerably larger number of data pairs available. An exception are the results for the stations Müggelseedamm and Johanna and Willi Brauer Platz in January, with 64 and 69 available data pairs, both located in the east of the city center. For the latter, the same pattern as reported above is true (higher concentrations simulated in the sensitivity simulation) and is thus in line with the reported results; however the model bias becomes positive in the sensitivity simulation. The model bias for the station Müggelseedamm is unchanged (and positive in both cases), which suggests that the location might not lie within the plume from the city on the selected dates. This is plausible, as the wind direction bin is rather large (90°). Overall, we conclude that the chosen thresholds are suitable for assessing the effect on downwind concentrations.

Table 1: Sensitivity of changes in NO2 model bias depending on the chosen cut-off for wind speed and the number of data pairs to be included. “n” denotes the number of data pairs, “MB” the mean bias averaged over all available data pairs in the respective wind direction bin (see main text). 1.8, 2.0 and 2.2 refer to the cut-off wind speed used for the calculations (in m/s). “diff” indicated the difference in mean bias between the base run and the sensitivity simulation. Highlighted are those stations excluded from the analysis, as the number of available data pairs is below the threshold (72).

name	**month**	**simulation**	**n_1.8**	**MB_1.8**	**n_2.0**	**MB_2.0**	**n_2.2**	**MB_2.2**	**diff_1.8**	**diff_2.0**	**diff_2.2**
Blankenfelde-Mahlow	Jan	base	32	1.48	32	1.48	32	1.48	0.01	0.01	0.01
Blankenfelde-Mahlow	Jan	sensitivity	32	1.47	32	1.47	32	1.47			
Buch	Jan	base	208	-1.68	203	-1.98	199	-2.31	-2.07	-2.11	-2.13

Buch	Jan	sensitivity	209	0.38	204	0.13	200	-0.18			
Frohnau	Jan	base	184	-3.03	182	-3.19	179	-3.22	-1.58	-1.53	-1.55
Frohnau	Jan	sensitivity	185	-1.45	182	-1.66	179	-1.66			
Gross Glienicke	Jan	base	204	-2.55	204	-2.55	204	-2.55	-2.38	-2.38	-2.36
Gross Glienicke	Jan	sensitivity	228	-0.17	228	-0.17	227	-0.19			
Grunewald	Jan	base	188	-5.86	188	-5.86	188	-5.86	-2.88	-2.88	-2.88
Grunewald	Jan	sensitivity	212	-2.98	212	-2.98	212	-2.98			
J. u. W. Brauer Platz	Jan	base	69	-0.19	69	-0.19	67	-0.46	-1.67	-1.67	-1.72
J. u. W. Brauer Platz	Jan	sensitivity	69	1.48	69	1.48	67	1.26			
Mueggelseedamm	Jan	base	64	5.59	64	5.59	64	5.59	0.00	0.00	0.00
Mueggelseedamm	Jan	sensitivity	64	5.59	64	5.59	64	5.59			
Schichauweg	Jan	base	34	2.78	33	2.76	31	2.47	-2.36	-2.47	-2.18
Schichauweg	Jan	sensitivity	34	5.14	33	5.22	30	4.64			
Blankenfelde-Mahlow	Jul	base	157	0.25	151	-0.01	144	0.15	-0.24	-0.42	-0.32
Blankenfelde-Mahlow	Jul	sensitivity	173	0.48	170	0.41	163	0.48			
Buch	Jul	base	47	-2.69	46	-2.56	44	-2.38	-1.11	-1.12	-0.89
Buch	Jul	sensitivity	50	-1.58	49	-1.44	47	-1.49			
Frohnau	Jul	base	41	-2.25	41	-2.25	40	-2.26	-1.96	-1.56	-0.65
Frohnau	Jul	sensitivity	44	-0.28	43	-0.69	41	-1.60			
Gross Glienicke	Jul	base	97	-4.43	91	-5.64	87	-6.35	-0.67	-1.35	-1.24
Gross Glienicke	Jul	sensitivity	96	-3.76	94	-4.30	89	-5.12			
Grunewald	Jul	base	97	-3.70	93	-3.86	86	-4.63	-2.11	-1.63	-2.59
Grunewald	Jul	sensitivity	95	-1.59	92	-2.23	89	-2.03			
J. u. W. Brauer Platz	Jul	base	158	-3.37	149	-3.95	142	-4.50	-1.28	-1.55	-1.02
J. u. W. Brauer Platz	Jul	sensitivity	158	-2.09	152	-2.40	144	-3.48			
Mueggelsee	Jul	base	156	-0.34	148	-0.98	145	-1.32	-0.43	-0.69	-0.24

damm											
Mueggelsee damm	Jul	sensitivity	160	0.09	154	-0.29	148	-1.08			
Schichauweg	Jul	base	143	-3.54	137	-3.64	130	-3.88	-0.65	-0.85	-0.80
Schichauweg	Jul	sensitivity	169	-2.88	158	-2.79	151	-3.08			

However, a small imprecision in the reporting of the results needs to be corrected in the manuscript, see below.

Changes in the manuscript

P. 17, line 20: "With some differences between the stations, the bias of weekday downwind NO2 concentrations was reduced by between ca. 1.5 and ~~3.0~~2.9 µg m-3 January and ca. 0.~~7~~4 and 1.5 µg m-3 in July."

Referee 2

In its current form, it shows a case study for the city of Berlin which maybe should be included in the title. It would be nice if the authors could add a section or a short passage, discussing the questions What do we learn from this study? Are the results transferrable to other urban areas?

P21: Some open questions should be briefly discussed her as mentioned in the beginning: Why is the study important for the field of research? What were the novel findings? What can we learn? Are the results transferrable to other urban areas?

Answer

In agreement with the reviewer, we add to the conclusions as follows:

Changes in the manuscript

Page 21, line 15: "The still negative bias is consistent with the factor being a rather conservative estimate.

The emission inventory used in this study is based on officially reported emissions by the individual countries, and the emissions are spatially distributed by TNO based on proxy data. Assuming the quality and accuracy of the proxy data is similar, at least for larger German cities, and considering that modelling studies for other German cities have also shown an underestimation of simulated NO2 concentrations using the same emission inventory, we would assume that the results found in this study for Berlin may generally be transferrable to at least other German metropolitan areas. The underestimation of NO2 concentrations throughout the day, the consistency of the calculated correction factor with findings from other studies and the improvement of model results applying the correction factor suggest that more research is needed in order to more accurately understand the spatial and temporal variability in real-world NOx emissions from traffic, and apply this understanding to the inventories used in high resolution chemical transport models. Given the above considerations, this not only holds for the urban area of Berlin, but for German and most

likely European metropolitan areas more generally."

Specific Comments

Referee 1

It might be helpful to expand the abbreviations MAM, JJA, SON, and DJF when they first appear, to make the paper more accessible to those who are not familiar with these common abbreviations.

Answer

Manuscript updated as suggested.

Changes in the manuscript

The captions of Tables 3 and 4 as updated to include the explanation of abbreviations: "Data are aggregated as follows: MAM - March, April, May, JJA - June, July, August, SON - September, October, November, DJF - December, January, February."

Referee 2

Page 1: Line 11: It is unclear at this point you have treated the discrepancy between model resolution and resolution of the emission data. What is the purpose of the emission downscaling to 1km and what was the initial resolution of the inventory. These points should be discussed more details in the methodology section, also explaining the downscaling method.

P4: 16: see comment above about the purpose of downscaling

P6: 2: specify chosen categories and reason for the selection

Answer

Since the original resolution of the emission inventory is much coarser (7km x 7km) than the model resolution used in this study (3km x 3km) we find it necessary to downscale the emission categories most relevant for this study based on representative proxy data. Kuik et al. (2016) have shown that this helps to better capture the spatial distribution of pollutant concentrations. Since the downscaling of the emissions was developed for the study presented in Kuik et al. (2016), which also included model simulations at a horizontal resolution of 1kmx1km, the emissions were downscaled to 1kmx1km instead of 3kmx3km.The inventory, reason and process of downscaling are discussed in Section 2.4 (extended as specified below) and in Kuik et al., 2016.

Changes in the manuscript

Page 1, line 11: "The emission data are downscaled from an original resolution of ca. 7 km x 7 km to a resolution of 1 km x 1 km"

Page 6, line 4, "In addition, we only downscale those emission categories (SNAP categories) which are both of main interest for studying NO2 in an urban area and also represented well by the proxy data chosen. This ensures that we are not suggesting a higher precision than achievable with the available proxy data. We thus only downscale emissions from SNAP

categories 2 (residential combustion), 6 (product use) and 71-75 (traffic), as these emissions can be represented well by population density (SNAP 2 and 6) and traffic density (SNAP 71-75)."

Referee 2
P2: 9ff: better EEA
21: better WHO

Referee 3
Page 13, line 9:
Eea and Wmo should be capitalized

Answer
Manuscript updated as suggested.

Changes in the manuscript
EEA and WHO capitalized throughout the manuscript.

Referee 2
Page 2:
10: What is the reason for the downward trend?

Answer
The reference given in the manuscript (EEA 2016) attributes the downward trend to decreasing NOx emissions. We add this information to the manuscript.

Changes in the manuscript
Page, line 10: "While there is a downward trend in NO2 concentrations due to decreasing NOx emissions, […]"

Referee 2
Page 2:
16: specify 'real world driving condition'

Answer
We meant conditions that resemble typical driving situations either because they are simulated accurately in the lab or because the car is actually driving on the road. To clarify, we rephrase as follows:

Changes in the manuscript
"[…] despite increasingly strict emission standards for diesel cars with the introduction of the Euro 5 and Euro 6 norms, ~~in real-world driving conditions~~ under real-world driving conditions, i.e. the pollutants a car produces while being driven on real roads as opposed to being tested in a lab, Euro 5-certified cars exceed the emission limit of […]"

Referee 1
page 3, line 7: it would be more correct to use "and" instead of "with" in this sentence

Answer
Manuscript updated as suggested.

Changes in the manuscript
Page 3, line 7: “However, many modelling studies report discrepancies between modelled and observed NO_2 concentrations […]”

Referee 2
P3: 3,7: Fallmann et al. 2016 presents a modelling study for the city of Stuttgart following similar procedures and gets similar findings

Answer
Thank you for the suggestion; we included this reference in the manuscript.

Changes in the manuscript
Page 3, line 20, “Fallmann et al. (2016) report a negative bias in NO2 concentrations simulated with WRF-Chem at 3kmx3km of ca. 50% on average and up to 60% during daytime.”

Referee 2
P3, 21: unclear: ‘basing’

Answer
We rephrase:

Changes in the manuscript
“Degraeuwe et al. (2016) assess the impact of different diesel NOx emission scenarios on air quality in Antwerp, ~~basing street canyon modelling on~~ ~~urban background concentrations modelled with~~ combining model simulations with LOTOS-EUROS at a horizontal resolution of ca. 7 km x 7 km (urban background) with a street canyon model.”

Referee 2
P4: 1: what is meant by ‘activity data’

Answer
We mean data describing the intensity of the (anthropogenic) activity causing the emissions, e.g. fuel burnt. We include this example to illustrate it:

Changes in the manuscript
Page 4, line 1:“Emissions are typically estimated from a combination of activity data (e.g. fuel burnt) and emission factors.”

Referee 2
P4: 5: where does the large error range come from and how did you consider this in your study?

P15: 16: what are the dominant non-traffic emission sources in the area of Berlin in summer time?

Answer
As we explain in the manuscript the TNO-MACC III emission inventory used here is based on officially reported emissions and thus consists of the reporting countries' best guess of emissions. The aim of this paper is to use these officially reported numbers and assess how much they might underestimate NO2 emissions in the case of Berlin. The emission error range is thus considered as a range in which the actual emissions are expected to fall.

In summer, traffic emissions in Berlin constitute ca. 60% of all NOx emissions according to the TNO-MACC III inventory. The next largest emission source is emissions from (energy) industry. However, it is important to note that, unlike traffic emissions, energy industry emissions are not all directly at the surface, but distributed between the third and seventh model layer (that is, above ca. 95m).

Both points are made clearer in the revised manuscript by adding the following:

Changes in the manuscript
Page 4, line 5, "Emission factors for road transport, for example, may have an error range between 50% and 200% , while emission factors for energy industry emissions, the second largest source of NOx emissions in Berlin, is much better constrained with an error range between 20% and 60% (Kuenen et al., 2014). Emission error ranges for the TNO-MACC III inventory used in this study are determined following the EEA Emission Inventory Guidebook, and depend, for example, on the number of measurements made for deriving the emission factor (EEA, 2013, Kuenen et al., 2014)."

Referee 2
P4:14: better 'model setup'

Answer
Manuscript changed as suggested.

Changes in the manuscript
"The model setup, […]"

Referee 2
P4: 21: State, whether the sensitivity simulations are taken from the output of the 1 year run or whether they are independent experiments.

Answer
To clarify, we change the manuscript as follows:

Changes in the manuscript
Page 4, line 21: "The factor is then tested in two individual one month long ~~sensitivity~~ simulations for January and July 2014, in the following referred to as sensitivity simulations."

Referee 2
P4: General: specify the source and the characteristics of the urban canopy parameters

Answer
Thank you for your comment. The source and characteristics of the parameters are described in detail in Kuik et al. (2016). We would like to refer the reader to this description (e.g. as indicated on page 4, line 30, 31 in the manuscript), in order to avoid making this present manuscript too long.

Changes in the manuscript
None

Referee 2
Page 14, line 5 onwards: What is the model height for the NOx evaluation? How do model and observation height compare?

Referee 3
Page 4, line 29: 'top' should be added after 'model layer'. With a lowest layer depth of 30 m the near surface profiles are not well resolved. How does this affect the simulated near-surface NO_x concentrations?

Answer
Modelled NOx concentrations are grid cell averages, with the first model layer top at 30m. Measurements are done at a height of 3m.

We agree with Reviewer 3 that the surface profile in the lower 30m is not well resolved. We have tested extrapolating the simulated profiles of NOx concentrations to the surface. Extrapolated daytime concentrations typically differed very little from the grid cell averages, while extrapolated nighttime concentrations would generally be much higher than the simulated grid-cell average. This is due to the steep gradient of the simulated NOx profile during nighttime when the planetary boundary layer is shallow. Ideally, the simulated profiles would be compared to measurements, which could then also be used to extrapolate the modelled grid-cell averages to the ground. Unfortunately such measurements are not available.

We discuss challenges in the comparability of grid cell averages with point measurements on page 13, line 7, and the evaluation of modelled vertical profiles with measurements on page 14, line 19/20 and page 20, line 29/30. We changed the manuscript as suggested by Reviewer 3, and add a sentence emphasizing the use of measured vertical profiles when

modelled profiles are not resolved in the lowest layer.

Changes in the manuscript
Page 4, line 29, "[…] with the first model layer top at […]".
Page 14, line 18: "Overall, this discussion shows that the representation of vertical mixing over urban areas might have to be improved to be physically more consistent in regional models, for example by better taking into account urban heat and momentum fluxes and treating the urban parameterization consistently with chemistry. Measurements of vertical profiles of NOx in cities, particularly in the planetary boundary layer, would be helpful in order to evaluate the models and improve the representation of surface NOx concentrations, as the NOx profile in the lowest model layer is not resolved at the model resolution used in this study."

Referee 2
P5: 20ff: how is second part of the simulation initialized, how does it refer to the first model period?

Referee 3
Page 5, line 21: Why is a 4-day spin-up required for the simulation of the last 6 month of the year?

Answer
The simulations for the first and second halves of the year were performed in parallel for reasons of computational efficiency. A comparison of the last days of the first simulation with the (overlapping) first days of the second simulation showed no discontinuities.

Changes in the manuscript
Page 5, line 20: "Both simulations are initialized using data from ERA-Interim (meteorology) and MOZART4/GEOS5 (chemistry) and preceded by a spin-up period of 4 days."

Referee 2
P5, 28: Discuss briefly if you can estimate an error from using 2011 emission data for a 2014 simulation?

Answer
From comparing the TNO-MACC III emissions for Germany in the years available, there was generally only a very small (decreasing) trend in reported emissions up to 2011, expected to continue also after 2011. This allows using the latest available year of emissions (2011) also for 2014 simulations (Hugo Denier van der Gon, personal communication). Thus, we expect the error due to the different year of emissions to be much smaller than the uncertainty related to the quantification of traffic NOx emissions. The latter is also supported by other studies reporting similar results using different methods (and different times) to assess the difference between reported and actual NOx emissions.

Changes in the manuscript
Page 5, line 26: "[…] The latest available year is 2011, which we use for simulating the year

2014. From comparing the TNO-MACC III emissions for Germany in the years available, there was generally only a very small (decreasing) trend in reported emissions up to 2011, expected to continue also after 2011. This allows using the latest available year of emissions (2011) also for 2014 simulations (Hugo Denier van der Gon, personal communication)."

Referee 3
Page 6, line 9: Please mention also the heights and not also the layer numbers.

Answer
We include the approximate layer heights in addition. Due to the terrain-following pressure coordinates of WRF, the exact value of the heights varies slightly.

Changes in the manuscript
P. 6, line 11: "[…] are distributed vertically into the first seven layers (see Supplementary Material for further details). For reference, the layer tops are at ca. 30 m (layer 1), 95 m (layer 2), 190 m (layer 3), 310 m (layer 4), 460 m (layer 5), 650 m (layer 6) and 890 m (layer 7)."

Referee 3
Page 6, line 28: Does the percentage refer to NO_2 mass? Please clarify.

Answer
Yes, it does. We include this in the manuscript as follows:

Changes in the manuscript
Page 6, line 28, "[…] NOx is emitted as 10% NO2 and 90% NO (by mass)."

Referee 2
P6, 23: Can you give an estimate of the error in mean values when only considering weekdays?

P13 30/31: earlier it was mentioned that only weekday profiles were applied in the model setup. Please clarify this aspect here.

Referee 3
Page 6, line 23: Why is the same diurnal cycle applied for weekdays and weekends? The traffic counts show certainly a different course for weekdays and weekends.

Answer
Weekly profiles were applied (depending on the day of the week), and daily profiles were applied (depending on the hour of the day). Applied diurnal profiles are not weekday-dependent, but calculated from averaging the profiles of all days of the week (and stations). We clarify this in the manuscript.

The reason for applying the same diurnal cycle on weekends and weekdays is mainly

computational efficiency, as the emission processing applied here is already very complex. Furthermore, this processing is in line with the way emissions are processed in many other modelling studies. As the applied average diurnal cycle is very similar to the weekday diurnal cycle, we expect that this would mainly effect simulated weekend concentrations. Furthermore, a comparison of the root mean square error of the (station-and-weekday-) average diurnal cycle with the (station-) average weekday and weekend diurnal cycles shows that the error is comparably small during daytime, which is the time period of main interest for this study. However, we agree that a distinction between weekend and weekday diurnal cycles might potentially be an improvement of the model setup. This would be part of the improvements in traffic emission diurnal cycles discussed in the conclusions of the manuscript.

Changes in the manuscript

P. 6, line 22: "[…] we apply a uniform diurnal cycle for each day of the week, making no distinction between the diurnal cycle of weekends and weekdays. As mentioned above, we do however apply also a weekly profile, thus the magnitude of the daily emissions on weekends is different from that on weekdays."

Referee 2

P7: 25: Procedure of NO2 retrieval at Air Base station unclear. How do you get the final NO2 concentrations from converted NO?

Answer

We clarify this in the manuscript as follows:

Changes in the manuscript

Page 7, line 11: "[…] as required by EU clean air legislation. The files can directly be downloaded from the AirBase website."

Page 7, line 25: "NO2 concentrations used for this study were measured using chemiluminescence. With this method, NO2 is converted to NO with a molybdenum converter before being detected using chemiluminescence, as NO reacts with O3 to form NO2 and O2 while emitting light (see, e.g. Gerboles et al., 2003, and Steinbacher et al., 2007)."

Referee 2

P8: 20,21: Specify more details about 'using the model setup for [. . .] policy relevant'. Do you mean operational forecast here?

Answer

Rather than operational forecasts, modelling studies aiming at identifying and/or evaluating NO2 reduction measures are meant.

Changes in the manuscript

Page 8, line 20/21: "Modelled NO2 concentrations are evaluated with the aim of using the model setup for policy-relevant analyses of urban NO2 concentrations and NO2 reduction measures with high temporal and spatial resolution, […]"

Referee 3
Page 8, line 25: The equation for MQO should be inserted already here.

Answer
Manuscript updated as suggested.

Changes in the manuscript
Page 8, line 26: Equation 2 is inserted here.

Referee 1
page 8, 2nd paragraph: urRV needs proper formatting twice here

Answer
Manuscript updated as suggested.

Changes in the manuscript
Page 8, lines 7 and 9: Formatting corrected.

Referee 3
Page 10, line 3: The linking between NO, NO_2 and O_3 is also true for offline models.

Answer
We absolutely agree. We rephrase the corresponding sentence to make the wording clearer.

Changes in the manuscript
Page 10, line 1: "[…] we include a brief evaluation of selected key meteorological parameters (temperature, wind speed and direction) as well as further chemical species (O3, NOx), the former because ~~as~~ WRF-Chem is an online-coupled model, and the latter because NO2 is tightly linked to NO and O3."

Referee 3
Page 10, line 25: Please add some details about the 'misreprentation of the diurnal cycles'.

Answer
We assume this comment relates to page 14, line 25. We clarify this sentence by changing the punctuation.

Changes in the manuscript
P. 14, line 25: "This is consistent with an overall underestimation of emission sources active in the morning hours on weekdays and potentially also a misrepresentation of the diurnal cycles of emissions in the model~~.~~: ~~T~~traffic emissions are distributed in the model throughout the day using […]"

Referee 2
P10: 25: state briefly why you changed the model setup compared to Kuik et al (2016) with regard to re-initialization etc. Can you discuss the impact of the modifications on NO, NO2, NOx, O3?

Answer
With regard to re-initialization, we changed the model setup because the length of the simulation presented here is considerably longer. We therefore decided to follow the procedure used to perform the AQMEII simulations (see manuscript for references). Because of the short spin-up times of the meteorology and chemistry in the model, we expect the impact of the reinitialization technique on the analyzed concentrations of NO, NO2, NOx and O3 to be rather small. The main benefit of this technique is an improved comparability of modelled pollutants with observed time series as the simulated meteorology follows the observations more closely. Please also see our answer to comments of referees 1 and 2 (first comment answered above).

Changes in the manuscript
None

Referee 2
P10, line 30: reference height for wind speed and temperature? Model height 30m?

Answer
Clarified in the manuscript as follows:

Changes in the manuscript
Page 10, line 30: "Modelled and observed 2m temperature and 10m wind speed are compared at five stations run by the German Weather Service […]"

Referee 2
P11: 5-21: Indicate briefly the reason for different model performance for different seasons/daytime. Does that relate to general problems with tour model setup (boundary layer scheme, chemical mechanism etc.)

5.1/5.2: how do the biases in meteorology relate to biases in chemical species? Is there a seasonal dependence?

Answer
It is widely known that the performance of meteorological variables depend on the area of interest, time of the day, season, land use data, the weather patterns, the PBL and land surface scheme used to set a model configuration etc. To ensure best model performance, multiple simulations would be required, but this approach is beyond the main purpose of this study. Furthermore, best model performance would still not imply an equally good performance throughout all seasons.

On the comment concerning Sect. 5.1/5.2 – we address this question in detail in Sect. 5.3. As mentioned in Sect. 5.3, we find that the model has some difficulties to capture the variation of NO2 concentrations on time scales below 2.5 days, which might be related to problems in modelled mixing. One seasonally dependent bias that is mentioned consists in higher than observed peaks in the model simulation at rural background stations in autumn, which might be related to a bias in wind direction.

Changes in the manuscript
None

Referee 3
Page 13, line 9: This is the case for all simulations with a grid width of only 3 km.

Answer
Yes. We make this clearer by rephrasing the corresponding sentence as follows:

Changes in the manuscript
P. 13, line 9: "[…] influenced by local sources that cannot be captured by WRF-Chem run at a horizontal resolution of 3 km x 3 km."

Referee 2
P13: 24: How is the link between population density and emission achieved/applied in the model?

Answer
The downscaling is described in Section 2.4 of the manuscript. We make the link clearer in the manuscript as follows:

Changes in the manuscript
Page 13, line 23: "The re-distribution of these emissions based on population density, as described in Sect. 2.4, may also have contributed to a better spatial representation in our study."

Referee 2
P14, 14: better 'too strong' instead of 'too efficient'

Answer
Manuscript changed as suggested.

Changes in the manuscript
Page 14, line 14: "This might be explained by mixing over urban areas during daytime that is too strong ~~efficient.~~"

Referee 2

P14: 5ff: Does a change of the urban canopy model make a change in boundary layer mixing close to the surface? How was the change in model code achieved?
Some boundary layer schemes have been modified recently in order to take into ac- count the mixing of chemical compounds inside the boundary layer. Can you comment on this? What impact might this have to your study?

P14, 15: Did you find a sensitivity to UCM selection?

Answer
In this study we have not tested different UCMs. In the WRF-Chem version used here (v3.8.1), the available urban canopy models are the single-layer UCM, and the more complex multi-layer building effect parametrization (BEP) and building energy model (BEM). The multi-layer parametrizations have not been tested for several reasons:
- They would strongly increase computational costs.
- Other studies have shown that they do not outperform simpler approaches when simulating surface temperature.
- They can only be combined with the MYJ and Boulac PBL schemes. The latter cannot be coupled with chemistry, while the former often leads to strong biases in simulated 2m air temperature.
- None of the urban parametrizations are coupled with chemistry, so the only effect on chemistry would be via improved meteorology.
Please also see Kuik et al., 2016, Section 2.3 for a discussion of these issues and further references, and our response to referees 1 and 2 on the subject of recent boundary layer scheme modifications (first comment).

Changes in the manuscript
None

Referee 2
P14, 25: Is the emission at urban background station generally too low? Maybe you are missing an advection term here?

Answer
Looking at the map of NOx emissions (Fig. 1 in the manuscript), we would assume that the biases are not due to an underestimation only at particular grid cells. The map shows that emissions are generally high in the area of the urban background stations (compare with Fig. 2). Advection of chemical species is solved on the same grid and is consist with the simulated meteorological fields. The relatively good model performance for simulated meteorology suggests that advection in the model is fine.

Changes in the manuscript
None

Referee 1
page 16, line 2: the correction factor should be "f", not "F"?

Answer
Manuscript changed as suggested.

Changes in the manuscript
Page 16, line 2: Nomenclature of correction factor harmonized (f_{NOx}).

Referee 2
P16: 18: Are the 'newer' diesel cars already implemented in the inventory used here?

Answer
The comment relates to the following sentence in the manuscript: "The seasonal differences might also be influenced by the temperature dependence of NOx emissions in newer diesel cars (Hausberger and Matzer, 2017), leading to higher NOx emissions at colder temperatures, which are not captured by the model." (Page 16, line 18).

The emission inventory/emission processing for the modelling does not take into account a potential temperature-dependence of emissions, while Hausberger and Matzer (2017), also cited in the manuscript, present results for a potential temperature-dependence of NOx emissions of Euro 4, Euro 5 and Euro 6 diesel cars. The contribution of Euro 6 cars in Germany was 1.5% in 2014, Euro 5 accounted for 27% and Euro 4 for 46% (Knörr et al., 2016). A temperature-dependence of their NOx-emissions could contribute to seasonal differences in model performance compared to observations, as described in the manuscript.

Changes in the manuscript
None

Referee 2
P17: 6: From your experience: Which part of the model/configuration has to be changed for improving the representation of mixing processes?

Answer
We address this question in the manuscript in Sect. 5.4 Diurnal and weekly variation of the model bias: "Overall, this discussion shows that the representation of vertical mixing over urban areas might have to be improved to be physically more consistent in regional models, for example by better taking into account urban heat and momentum fluxes and treating the urban parameterization consistently with chemistry." (P. 14, line 18). This point is repeated in the conclusions.

Changes in the manuscript
None

Referee 2
P17, line 11: Can the difference between secondary and primary NOx be attributed to wrong NO titration processes? Or other chemical processes?

Answer
As O3/NOx chemistry is non-linear, we do not think that it is straight forward to attribute the differences to one chemical process. Rather, other processes e.g. the evolution of the boundary layer height also have a strong impact on pollutant concentrations and the chemical regime. Thus, an attribution of the differences would need to look at a variety of different processes and factors, which is beyond the scope of this study. We include this in the manuscript as follows:

Changes in the manuscript
Page 17, line 11: “The differences in NO2 and NOx improvements suggest that the impact of the primary NO2 fraction in emitted NOx on ~~observed and~~ modelled NO2 and NOx concentrations, as well as the influence of chemical processes such as NO titration and other relevant physical and chemical processes, might need to be assessed in greater detail.”

Referee 1
page 17, line 19: should it be "wind direction bin", instead of "wind speed bin"?

Answer
Yes, that is true. Changed accordingly.

Changes in the manuscript
Page 17, line 18: “[…] exist in the respective wind ~~speed~~ direction bin […]”

Referee 2
P18: 11: The fact that a higher resolution emission inventory does not improve the model results is an important finding here and has to be explained more detailed. Did you run model experiment with a high-resolution emission inventory as well? What is the resolution here?

Answer
It appears that the reviewer has misunderstood something here. We do not discuss the resolution of the inventory in this part of the text. Conversely, in a previous study (Kuik et al., 2016), we find that local pollution patterns can be represented better when downscaling the emission inventory from a horizontal resolution of 7km to 1km and increasing the model resolution.

We are glad to further respond to this comment in case the reviewer would like to further elaborate on the comment.

Changes in the manuscript
None

Referee 2
P18 15-20: What kind of models do you suggest in terms of higher resolution? Chemical transport models, LES, CFD, dispersion models. . ?

Answer
The paragraph in the manuscript that this comment refers to speaks about the scaling of roadside NOx concentrations with traffic counts, and tests whether a linear relationship exists (as often assumed in the calculation of the diurnal distribution of traffic emissions). Here, the term "model" is used for mathematical relationships between NOx and traffic counts, not in the context of atmospheric chemistry models. To make it clearer, we replace the term "model" in this section with the term "relationship". We are glad to give a more detailed answer to this question if the reviewer would like to clarify what is being referred to here.

Changes in the manuscript
A linear regression ~~model~~ does not explain the variance of observed NOx concentrations at nighttime, as indicated by the R^2 close to 0 in Fig. 9. However, during daytime, traffic counts alone explain up to ca. 40% of observed NOx variance, particularly during the
traffic rush hours. The explained variance is smaller during the afternoon peak. In comparison to a linear ~~model~~ relationship, a quadratic ~~model~~ relationship (NOx / (traffic_count)^2) does not explain more of the observed variance (not shown). An exponential ~~model~~ relationship (NOx/exp(traffic_count)), however, does explain a considerably larger share of the observed variance during daytime and particularly during the traffic rush hours, as depicted in Fig. 9 (up to ca. 60% depending on the station).

Referee 2
P18, 25: meaning: HBEFA

Answer
In this case yes.

Changes in the manuscript
Page 18, line 25, "[…],and emission factors (e.g. from HBEFA) are higher in congested situations compared to free flowing traffic."

Referee 2
P20: 27: Urban processes in WRF-Chem are linked via meteorology only. Do you have a suggestion, how UCMs have to be changed in order to improve AQ simulations?

Answer
Following the line to which the comment refers, we have already included a suggestion on how such an improvement might be achieved: "For example, the parameterization of urban processes needs to better account for urban heat and momentum fluxes for a more realistic representation of mixing both at daytime and at nighttime, particularly in summer."

Changes in the manuscript

None

Referee 1
Page 34, figure 8: your description misses out the colour (blue?) of the sensitivity simulation data

Answer
Manuscript changed as suggested.

Changes in the manuscript
Caption Figure 8: "Time series of hourly observed (black line) and modelled NO2, comparing the base simulation (red) with the sensitivity simulations (blue) […]"

Referee 1
page 38, table 3: in the table caption, the units of MB and RMSE need correcting from ug/m3 to K and m/s (respectively)

Answer
Manuscript changed as suggested.

Changes in the manuscript
Caption Table 3: "Mean bias (MB) and root mean square error (RMSE) are indicated in K (temperature) and m s^{-1} (wind speed), […]"

Additional minor changes
P. 21, line 14: "[…] and also improve~~d~~ modelled downwind concentrations."

References not listed in the manuscript

- Hu, X.-M., P. M. Klein, and M. Xue (2013), Evaluation of the updated YSU planetary boundary layer scheme within WRF for wind resource and air quality assessments, J. Geophys. Res. Atmos., 118, 10,490–10,505, doi:10.1002/jgrd.50823.

- Knörr, W., Heidt, C., Gores, S., Bergk, F.: Aktualisierung „Daten- und Rechenmodell: Energieverbrauch und Schadstoffemissionen des motorisierten Verkehrs in Deutschland 1960-2035" (TREMOD) für die Emissionsberichterstattung 2016 (Berichtsperiode 1990-2014), https://www.ifeu.de/wp-content/uploads/Endbericht_TREMOD_2016_160701.pdf (last access: 11 April 2018)

Top-down quantification of NO_x emissions from traffic in an urban area using a high resolution regional atmospheric chemistry model

Friderike Kuik[1,2], Andreas Kerschbaumer[3], Axel Lauer[4], Aurelia Lupascu[1], Erika von Schneidemesser[1], and Tim M. Butler[1,2]

[1]Institute for Advanced Sustainability Studies, Potsdam, Germany
[2]Freie Universität, Fachbereich Geowissenschaften, Institut für Meteorologie, Berlin, Germany
[3]Senatsverwaltung für Umwelt, Verkehr und Klimaschutz Berlin, Germany
[4]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

Correspondence to: Friderike Kuik (friderike.kuik@gmail.com)

Abstract. With NO_2 limit values being frequently exceeded in European cities, complying with the European air quality regulations still poses a problem for many cities. Traffic is typically a major source of NO_x emissions in urban areas. High resolution chemistry transport modelling can help to assess the impact of high urban NO_x emissions on air quality in and outside of urban areas. However, many modelling studies report an underestimation of modelled NO_x and NO_2 compared with observations. Part of this model bias has been attributed to an underestimation of NO_x emissions, particularly in urban areas. This is consistent with recent measurement studies quantifying underestimations of urban NO_x emissions by current emission inventories, identifying the largest discrepancies when the contribution of traffic NO_x emissions is high. This study applies a high resolution chemistry transport model in combination with ambient measurements in order to assess the potential underestimation of traffic NO_x emissions in a frequently used emission inventory. The emission inventory is based on officially reported values and the Berlin-Brandenburg area in Germany is used as a case study. The WRF-Chem model is used at a 3 km x 3 km horizontal resolution, simulating the whole year 2014. The emission data are downscaled from an original resolution of ca. 7 km x 7 km to a resolution of 1 km x 1 km. An in-depth model evaluation including spectral decomposition of observed and modelled time series and error apportionment suggests that an underestimation in traffic emissions is likely one of the main causes of the bias in modelled NO_2 concentrations in the urban background, where NO_2 concentrations are underestimated by ca. 8 $\mu g\ m^{-3}$ (-30%) on average over the whole year. Furthermore, a diurnal cycle of the bias in modelled NO_2 suggests that a more realistic treatment of the diurnal cycle of traffic emissions might be needed. Model problems in simulating the correct mixing in the urban planetary boundary layer probably play an important role in contributing to the model bias, particularly in summer. Also taking into account this and other possible sources of model bias, a correction factor for traffic NO_x emissions of ca. 3 is estimated for weekday daytime traffic emissions in the core urban area, which corresponds to an overall underestimation of traffic NO_x emissions in the core urban area of ca. 50%. Sensitivity simulations for the months of January and July using the calculated correction factor show that the weekday model bias can be improved from -8.8 $\mu g\ m^{-3}$ (-26%) to -5.4 $\mu g\ m^{-3}$ (-16%) in January on average in the urban background, and -10.3 $\mu g\ m^{-3}$ (-46%) to -7.6 $\mu g\ m^{-3}$ (-34%) in July. In addition, the negative bias of weekday NO_2 concentrations downwind of the city in the rural and suburban background can be reduced from -3.4 $\mu g\ m^{-3}$ (-12%) to -1.2 $\mu g\ m^{-3}$ (-4%) in January and -3.0 $\mu g\ m^{-3}$ (-22%) to -1.9 $\mu g\ m^{-3}$ (-14%) in July. The results and

their consistency with findings from other studies suggest that more research is needed in order to more accurately understand the spatial and temporal variability in real-world NO_x emissions from traffic, and apply this understanding to the inventories used in high resolution chemical transport models.

1 Introduction

Limit values for ambient NO_2 concentrations (Ambient Air Quality Directive 2008/50/EC) as well as NO_x exhaust emission standards are set by European legislation, but ambient measurements show that NO_2 concentrations still frequently exceed the European annual mean limit value of 40 µg m^{-3} (EEA, 2016; Minkos et al., 2017). For example, 12% of all measurement sites in Europe registered exceedances of the annual mean limit value in 2014, most of them located at the roadside. Within Europe, Germany had the highest median NO_2 concentrations in 2014 (EEA, 2016), where it was estimated that the limit value was exceeded at 57% of all traffic sites (Minkos et al., 2017). While there is a downward trend in NO_2 concentrations due to decreasing NO_x emissions, extrapolating the current trend to 2020, exceedances are still expected at 7% of the stations in 2020, requiring additional measures in order for the European air quality goals to be met (EEA, 2016).

In general, traffic is the most important source of NO_x emissions in Europe contributing 46% in 2014 in the EU-28, with considerably higher contributions to ambient NO_2 concentrations in urban areas (EEA, 2016). NO_x emissions from diesel vehicles, the main traffic NO_x source, have recently been a strong focus of international media attention: despite increasingly strict emission standards for diesel cars with the introduction of the Euro 5 and Euro 6 norms, ~~in~~ under real-world driving conditions, i.e. the pollutants a car produces while being driven on real roads as opposed to being tested in a lab, Euro 5-certified cars exceed the emission limit of 0.18 g km^{-1} by an average factor of 4-5 (e.g. EEA, 2016; Hausberger and Matzer, 2017) and the newer Euro 6-cars exceed the emission limit of 0.08 g km^{-1} by an average factor of 6-7 (e.g. EEA, 2016; Briefing, 2016).

NO_x impacts human health, ecosystems and climate directly and indirectly as precursor of tropospheric ozone (O_3) and particulate matter (PM). Health impacts of NO_2 include adverse respiratory effects (WHO, 2013), and the effect of road traffic NO_2 on premature mortality might be more than ten times larger than the effect of road traffic PM2.5 (Harrison and Beddows, 2017).

In order to support policy makers in identifying suitable measures to reduce roadside and urban background NO_2 concentrations to levels well below the limit value, as well as for assessing the health impact of current and future NO_2 concentrations, air pollution modelling is a valuable tool (e.g. Von Schneidemesser et al., 2017). Chemistry transport models can be used to assess the impact of local emissions on air chemistry and air quality in the surroundings and downwind of the emission sources. Online-coupled models, such as the chemistry version of the Weather Research and Forecasting model (WRF-Chem, Grell et al., 2005), have several advantages compared with offline approaches. These include, for example, a numerically more consistent treatment and a more realistic representation of the atmosphere, particularly in case of high model resolution (Grell and Baklanov, 2011).

Due to its short lifetime in the atmosphere, NO_2 is more spatially variable than for example O_3, particularly in urban areas with locally high NO_x emissions. This is one of the reasons why models with higher spatial resolutions of a few km are capable of representing observed NO_2 concentrations better than coarser models, with better performance if emission input and meteorological data are also available at the high resolution (e.g. Schaap et al., 2015). In terms of model evaluation, comparing NO_2 concentrations averaged over a coarse model grid cell with point measurements can lead to mismatches (Solazzo et al., 2017), with a better comparability achieved through high model resolutions of only a few km or less, depending on the size of the city. Simulating air quality in Mexico City, Tie et al. (2010) showed that reasonable model results can be achieved at a ratio of city size to model resolution of ca. 6:1.

However, many modelling studies report discrepancies between modelled ~~with~~ and observed NO_2 concentrations, which are in parts attributed to an underestimation of traffic NO_x emissions. All but one model simulating the European domain during the model intercomparison project AQMEII phase 2 underestimate annual mean NO_2 concentrations by 9%-45% on average. Some of them overestimate NO_2 concentrations at nighttime (Im et al., 2015), meaning that daytime concentrations are underestimated even more than the average model bias would indicate. Similarly, the European models contributing to the more recent AQMEII phase 3 intercomparison show an under-prediction of NO_2 concentrations throughout the whole year, with the sole exception of one model (Solazzo et al., 2017). In the Eurodelta model intercomparison study (Bessagnet et al., 2016), the participating models simulate NO_2 concentrations reasonably well on average compared with observations in the rural background, but most models show an underestimation of daytime NO_2 on average, particularly in summer (Fig. 9 from Bessagnet et al., 2016). Few studies focus particularly on NO_2 in urban areas: Terrenoire et al. (2015) simulated air quality over Europe at a horizontal resolution of 0.125° x 0.0625° with the CHIMERE model for 2009 and found that NO_2 concentrations are underestimated by more than 50% in urban areas. Schaap et al. (2015) show that the bias in modelled NO_2 concentrations in urban areas is reduced with increasing model resolutions, but still report negative biases for a model resolution of 7 km x 7 km, between 6 and 10 µg m^{-3} for different offline-coupled chemistry transport models. Fallmann et al. (2016) report a negative bias in NO_2 concentrations simulated with WRF-Chem at 3kmx3km of ca. 50% on average and up to 60% during daytime. Degraeuwe et al. (2016) assess the impact of different diesel NO_x emission scenarios on air quality in Antwerp, ~~basing street canyon modelling on urban background concentrations modelled~~ combining model simulations with LOTOS-EUROS at a horizontal resolution of ca. 7 km x 7 km (urban background) with a street canyon model. They report a low bias in modelled urban background NO_2 concentrations of ca. 20%, requiring bias correction for the further analysis of the emission scenarios. Kuik et al. (2016) evaluated air quality simulated with WRF-Chem over the Berlin-Brandenburg region and found underestimations of NO_2 concentrations at daytime, and overestimations at nighttime.

Many studies attribute an underestimation of observed NO_2 concentrations to an underestimation of emissions (e.g. Solazzo and Galmarini, 2016; Degraeuwe et al., 2016; Giordano et al., 2015) and particularly traffic emissions in urban areas (Terrenoire et al., 2015). Further reported causes of the disagreement include problems with simulating the correct PBL height and mixing in the model (e.g. Solazzo et al., 2017; Kuik et al., 2016).

Modelling studies for North America report lower negative or even positive biases in modelled NO_2 concentrations (e.g Solazzo et al., 2017). While total NO_x emissions reported for Europe are on average already larger than for North America by

a factor of more than 2 (Im et al., 2015), these differences might indicate an even larger contribution of diesel car emissions to measured NO_2 concentrations, as the share of diesel cars is a main difference in emission sources between Europe and North America. Thus, large differences in the model bias between Europe and North America would be consistent with an underestimation of diesel traffic emissions in Europe.

Emissions are typically estimated from a combination of activity data (e.g. fuel burnt) and emission factors. Emission factors for road transport emissions depend on the fuel type and the car type (heavy duty or light duty, exhaust treatment) as well as on the driving conditions including road type and speed (e.g. Hausberger and Matzer, 2017). While activity data are only assumed to have an uncertainty of ca. 5%-10%, the emission factor is more difficult to quantify in many cases (Kuenen et al., 2014, and references therein). Emission factors for road transport, for example, may have an error range between 50% and 200%~~(Kuenen et al., 2014) .~~ , while emission factors for energy industry emissions, the second largest source of NO_x emissions in Berlin, is much better constrained with an error range between 20% and 60% (Kuenen et al., 2014) . Emission error ranges for the TNO-MACC III inventory used in this study are determined following the EEA Emission Inventory Guidebook, and depend, for example, on the number of measurements made for deriving the emission factor (EEA, 2013; Kuenen et al., 2014) . Recent studies for London show that NO_x emissions from flux measurements are up to 80% (Lee et al., 2015), or a factor of 1.5-2 (Vaughan et al., 2016) higher than NO_x emissions from the UK National Atmospheric Emissions Inventory, with largest discrepancies found in cases where traffic is the dominant source of NO_x concentrations. Karl et al. (2017) conclude from eddy covariance measurements in Austria that traffic related NO_x emissions in emission inventories frequently used by air quality models can be underestimated by up to a factor of 4 for countries where diesel cars represent a major fraction of the vehicle fleet and have a significant contribution to reported biases in modelled NO_2 concentrations.

In this study the aim is to quantify the underestimation of traffic emissions in a widely used state-of-the art emission inventory based on officially reported emissions, for simulating NO_2 concentrations in an urban area with high resolution. We use the Berlin-Brandenburg area as a case study, and use the WRF-Chem model to simulate NO_2 concentrations. The model setup, model simulations and input data are described in Sect. 2, and observational data used are described in Sect. 3. The emission inventory used here is the TNO-MACC III inventory (Kuenen et al., 2014), downscaled to ca. 1 km x 1 km for the Berlin-Brandenburg area (Kuik et al., 2016), also described in Sect. 2. The analysis builds on advanced model evaluation techniques including an operational and a diagnostic evaluation (outlined in Sect. 4) of the modelled NO_2 concentrations (Dennis et al., 2010), with the aim of assessing the contribution of different sources of model error (Sect. 5). Based on this analysis, a correction factor for traffic emissions is calculated, and additional sources of the model bias are discussed (Sect. 6). The factor is then tested in two ~~sensitivity~~ individual one month long simulations for January and July ~~2014.~~ 2014, in the following referred to as sensitivity simulations. In addition, we analyse observational data of NO_2 concentrations and traffic counts, assessing the linear scaling assumed between emissions and traffic counts for the temporal distribution of emissions in chemistry transport models. Section 7 closes with a summary and conclusions from the results.

2 Model simulations

2.1 Model setup

We use the Weather Research and Forecasting model (WRF) version 3.8.1 (Skamarock et al., 2008), with chemistry and aerosols (WRF-Chem, Grell et al., 2005; Fast et al., 2006). The setup includes two model domains centred over Berlin, at horizontal resolutions of 15 km x 15 km and 3 km x 3 km, using one-way nesting. The model top is at 50 hPa, using 35 vertical levels with the first model layer top at approximately 30m above the surface. There are 12 levels in the lowest 3km. Urban processes (meteorology) are parametrized with the single layer urban canopy model, with input parameters specified for Berlin as described in Kuik et al. (2016) and three urban land use categories. The setup further includes the RADM2 chemical mechanism with the Kinetic Pre-Processor (KPP) and the MADE/SORGAM aerosol scheme. The MOZART chemical mechanism is used in a sensitivity test. All physics and chemistry schemes used in this study are listed in Table 1.

Small changes in the code have been made. The initialization of the dry deposition (module_dep_simple.F) has been adapted in order to account for three urban land use categories as described in Kuik et al. (2016) and references therein. Nighttime mixing over urban areas is not accounted for sufficiently by the urban parametrization and the PBL scheme and thus adjusted (dry_dep_driver.F) as described in the Supplementary Material. There, we also show illustrative results (Fig. S1-S3) of two test simulations comparing the impact of changes in this model setup with respect to Kuik et al. (2016) , including the modification of nighttime mixing and the modification of the diurnal distribution of traffic emissions (see Sect. 2.4).

2.2 Model input data

We use the European Centre for Medium-Range Weather Forecasting (ECMWF) Interim reanalysis (ERA-Interim, Dee et al., 2011) with a horizontal resolution of 0.75°x0.75°, and a temporal resolution of 6h, interpolated to 37 pressure levels (with 29 levels below 50 hPa) as meteorological initial and lateral boundary conditions. The sea surface temperature is updated every 6 hours. The data are interpolated to the model grid using the standard WRF pre-processing system (WPS). Chemical boundary conditions for trace gases and particulate matter are created from simulations with the global chemistry transport Model for Ozone and Related chemical Tracers (MOZART-4/GEOS-5 Emmons et al., 2010). Instead of the standard USGS land use data we use CORINE data (EEA, 2014), remapped to the USGS classes, using three categories characterizing the urban area. This provides are more realistic characterization of the land use in the Berlin-Brandenburg area (Kuik et al., 2016; Churkina et al., 2017). The emission input data and its pre-processing are described in Sect. 2.4.

2.3 Simulation procedure

Following the work flow used in AQMEII phase 2 (Brunner et al., 2015), we re-initialize the simulation every two days, with a one day spin-up of the model meteorology. To ensure consistency in the chemical fields, we start each new two-day simulation from the chemistry fields of the previous simulation. For the base run using the RADM2 chemistry scheme we do a full year (2014) simulation. The results of this simulation are used to derive a correction factor for road traffic emissions, as explained in

Section 6.1. For computational reasons, the simulation is divided into two parts covering the first six and the last six months of the year. Both simulations are initialized using data from ERA-Interim (meteorology) and MOZART4/GEOS5 (chemistry) and preceded by a spin-up period of 4 days. We do a one-month sensitivity simulation with the MOZART chemistry scheme (July 2014), and two sensitivity simulations with increased traffic emissions for January and July 2014, all with the same simulation procedure. All model simulations are listed in Table 2.

2.4 Emissions

2.4.1 General description

The emission data used in this study are from the TNO-MACC III inventory (Kuenen et al., 2014). The latest available year is 2011, which we use for simulating the year 2014. From comparing the TNO-MACC III emissions for Germany in the years available, there was generally only a very small (decreasing) trend in reported emissions up to 2011, expected to continue also after 2011. This allows using the latest available year of emissions (2011) also for 2014 simulations (Hugo Denier van der Gon, personal communication). Details on the emission inventory and the way these emissions are used in the present WRF-Chem setup can be found in Kuik et al. (2016) and references therein, and are briefly summarised here. The data are originally at a horizontal resolution of ca. 7 km x 7 km, which we downscale for the Berlin-Brandenburg region based on proxy data (Fig. 1). As it has been shown that downscaling the emission data to the resolution of the model grid helps to better capture the spatial distribution of air pollutant concentrations, we updated the downscaling procedure (see Supplementary Material). The updates include an extension of the region for which the emissions are downscaled from Berlin to the whole Berlin-Brandenburg region. In addition, we only downscale those emission categories (SNAP categories) which are both of main interest for studying NO_2 in an urban area and also represented well by the proxy data chosen. This ensures that we are not suggesting a higher precision than achievable with the available proxy data. We thus only downscale emissions from SNAP categories 2 (residential combustion), 6 (product use) and 71-75 (traffic), as these emissions can be represented well by population density (SNAP 2 and 6) and traffic density (SNAP 71-75).

2.4.2 Emission processing

Kuik et al. (2016) concluded that when simulating urban air quality with high resolution and using emission input data at high resolution, a more detailed treatment of the vertical distribution of point source emissions might further improve the model results. For this reason in this study, the emissions are distributed vertically based on profiles adapted from Bieser et al. (2011), i.e. emissions from the energy industry are distributed between the third and seventh model layer, emissions from other industrial sources as well as from the extraction and distribution of fossil fuels are distributed between the first four model layers, waste treatment emissions are distributed in the first five model layers and airport emissions (LTO cycle) are distributed vertically into the first seven layers (see Supplementary Material for further details). For reference, the layer tops are at ca. 30 m (layer 1), 95 m (layer 2), 190 m (layer 3), 310 m (layer 4), 460 m (layer 5), 650 m (layer 6) and 890 m (layer 7).

TNO-MACC III emissions are provided as annual totals. For each emission (SNAP) category separately, we apply factors distributing the emissions for each month, day of the week (weekend vs. weekday), and hour of the day (diurnal cycle) based on Builtjes et al. (2002), with the exception of the diurnal cycle of traffic emissions. Previous studies highlighted the importance of using locally available information when specifying temporal profiles of emissions (e.g. Mues et al., 2014). Here we apply a diurnal cycle of traffic emissions (fraction of total daily emissions per hour of the day) calculated based on traffic counts provided by the Berlin Senate Department for the Environment, Transport and Climate (data used from 2007-2016) and by the German Federal Highway Research Institute BASt (Bundesanstalt für Straßenwesen, 2017, data used from 2003-2016). The diurnal cycle applied here is obtained by calculating the fraction of average daily traffic counts in Berlin at each hour of the day, thus assuming a linear scaling of traffic emissions with traffic counts as also assumed by Builtjes et al. (2002). Following Builtjes et al. (2002), we apply a uniform diurnal cycle for each day of the week, making no distinction between the diurnal cycle of weekends and weekdays. As mentioned above, we do however apply also a weekly profile, thus the magnitude of the daily emissions on weekends is different from that on weekdays. The main differences between the profiles calculated based on locally available information and the hourly emission factors from Builtjes et al. (2002) include an earlier increase of traffic emissions in the morning by ca. one hour and more evenly distributed high traffic emissions during the day with less pronounced morning and afternoon peaks.

NO_x is emitted mainly as NO, but also includes a fraction directly emitted as NO_2 (the primary NO_2 fraction, f-NO_2) by combustion engines. Here, NO_x is emitted as NO for all SNAP categories except "road transport" and "non-road transport". For non-road transport and all road transport emissions except diesel, NO_x is emitted as 10% NO_2 and 90% NO (by mass). Road transport diesel emissions include both light duty vehicle (LDV) and heavy duty vehicle (HDV) emissions. For the latter, we also assume a f-NO_2 of 10%, while for light duty vehicles we assume a f-NO_2 of 26% (Carslaw, 2005). Combining this with the TNO-MACC share of diesel emissions attributable to LDV (43%) and HDV (57%), we obtain a combined f-NO_2 for road transport diesel NO_x emissions (SNAP 72) of 17%. Test simulations varying the f-NO_2 for diesel LDV between 10% and 55% have shown that the simulated NO, NO_2 and NO_x concentrations have very little sensitivity towards the f-NO_2 of LDV diesel emissions, while small differences in the simulated ozone concentrations were seen. As further sensitivity simulations on this topic are beyond the scope of this study and differences were small, we chose to use a f-NO_2 that was around the mid-point of those values documented for LDV diesel NO_x emissions (26%).

2.4.3 Comparison of the downscaled TNO-MACC III emissions with a local inventory

Local NO_x emissions from road transport are available for 2009 (Berlin Senate Department for the Environment, Climate and Transport, online). In comparison with the downscaled TNO-MACC III emissions for the Berlin grid cells (2011), traffic NO_x emissions from the local inventory are 6% higher. The geographical distribution of the emissions in the local inventory is very similar to the downscaled version of TNO-MACC III used in this study (Fig. 1).

3 Observational data

3.1 AirBase observations and NO_2 uncertainty

NO_2, NO_x and O_3 measurements are taken from AirBase (EEA, 2017), a database compiling air quality observations from the EU Member States and associated countries, performed as required by EU clean air legislation. The files can directly be downloaded from the AirBase website. In the case of Germany the measurements are performed by the federal states. For the comparison with model results, observations from stations within Berlin and in the adjacent surroundings in the Federal State of Brandenburg representing "urban background", "suburban background" and "rural near-city" conditions are used (Fig. 2, also see supplementary Figure S4 and Table S1). For our analysis, we re-classify the AirBase station "DEBE066" in Berlin-Karlshorst from "urban background" to "suburban background", as the station is not located in the core area of the city and pollutant concentrations measured there are similar to concentrations measured at other suburban background stations. As a result, four stations for each classification type are used in this study: Amrumer Straße (DEBE010), Brückenstraße (DEBE068), Belziger Straße (DEBE018) and Nansenstraße (DEBE034) in the urban background, Blankenfelde-Mahlow (DEBB086), Buch (DEBE051), Groß Glienicke (DEBB075) and Johanna und Willy Brauer Platz (DEBE066) in the suburban background, and Frohnau (DEBE062), Grunewald (DEBE032), Müggelseedamm (DEBE056) and Schichauweg (DEBE027) in the rural near-city background.

In addition, five measurement stations representing "traffic" conditions within Berlin, which are located next to major roads within the core area of the city, and assumed to be primarily influenced by traffic emissions, are used for the observation-based analysis (Sect. 6.3).

NO_2 concentrations used for this study were measured using chemiluminescence. With this method, NO_2 is converted to NO with a molybdenum converter before being detected using chemiluminescence, as NO reacts with O3 to form NO2 and O2 while emitting light (see, e.g. Gerboles et al., 2003; Steinbacher et al., 2007) . A limitation of this method is that other nitrogen-containing species (PAN, HNO_3) are also converted to NO in this process. In a comparison study, Steinbacher et al. (2007) found that only 73%-82% of the measured NO_2 with this method is "real" NO_2, at a rural background site in Switzerland. However, they state that reasonable results are obtained with this type of converter at urban background sites. Villena et al. (2012) compared NO_2 concentrations in urban smog conditions in Santiago de Chile using chemiluminescence detection with a molybdenum converter and differential optical absorption spectroscopy and found large differences between measured concentrations during daytime. Further sources of uncertainty are introduced in the detection itself, for which NO reacts with O_3, producing the luminescence signal to be detected. Gerboles et al. (2003) assess the uncertainty of NO_2 measurements, and Pernigotti et al. (2013) derive a simplified procedure in order to calculate the NO_2 measurement uncertainty, which we apply in order to obtain a rough estimate of the uncertainty range of NO_2. Accordingly, the uncertainty (u) of the observed NO_2 concentrations x at time i is quantified as follows:

$$u(x_i) = u_{r_{RV}} \cdot \sqrt{(1-\alpha)x_i^2 + \alpha \cdot RV^2} \quad (1)$$

Here, ~~urRV~~ $u_{r_{RV}}$ is an estimate of the relative uncertainty around a reference value RV, and α is the fraction of uncertainty not proportional to the reference value. We use the coefficients corresponding to the mean uncertainties of the individual parameters, i.e.~~urRV~~$u_{r_{RV}}$=0.09, α=0.06 and the reference value RV=200 µg m^{-3} (Pernigotti et al., 2013).

3.2 Meteorological data

In order to complement the analysis and to investigate potential influences of the modelled meteorology on modelled NO_2 concentrations, we include a comparison of modelled meteorology with observations. This includes observations of 2m temperature, and 10m wind speed and direction, all provided by the German Weather Service and available online (Kaspar et al., 2013). In addition, mixing layer height derived from ceilometer measurements at Nansentraße during the BAERLIN2014 campaign (Geiß et al., 2017) are used for a qualitative comparison with the modelled mixing layer height (see Kuik et al., 2016, for a discussion of this type of comparison). The data are generally available between 20 June and 27 August 2014, but include a number of gaps.

4 Analysis and evaluation metrics

4.1 Analysis of model results

Modelled NO_2 concentrations are evaluated with the aim of using the model setup for policy-relevant analyses of urban NO_2 concentrations and NO_2 reduction measures with high temporal and spatial resolution, and in order to identify the main sources of the errors in modelled NO_2 concentrations. For this, we use both operational and diagnostic evaluation metrics, which are explained in the following.

Operational evaluation metrics applied here are based on Thunis et al. (2012) and Pernigotti et al. (2013). They include an analysis of the mean bias (MB) and normalized mean bias (NMB), the correlation coefficient (R), and the root mean square error (RMSE, as defined in the Supplementary Material). The model error is compared with the model quality objective (MQO) and performance criteria calculated from NO_2 observations and their uncertainty. The MQO is defined as follows:

$$\mathrm{MQO} = \frac{1}{2}\frac{\mathrm{RMSE}}{\mathrm{RMS_U}} \tag{2}$$

Following Thunis et al. (2012) and Pernigotti et al. (2013), a MQO lower than 0.5 indicates that the model results are on average within the range of the measurement uncertainty, and further efforts to improve model performance are not meaningful. A MQO between 0.5 and 1 indicates that the uncertainties of model and observations overlap, and that the model might still be a better predictor of the true value than the observations. A MQO greater than 1, on the other hand, indicates significant differences between the model and the observations. ~~The MQO is defined as follows:~~

~~$$\mathrm{MQO} = \frac{1}{2}\frac{\mathrm{RMSE}}{\mathrm{RMS_U}}$$~~

With RMS_U being the root mean square of the measurement uncertainty. The performance criteria for mean bias, normalized mean bias and correlation coefficient as defined in Pernigotti et al. (2013) are listed in the Supplementary Material. As the uncertainty of NO_2 measurements is partly concentration-dependent, the MQO and the other performance criteria differ between station classes and seasons.

The operational evaluation and model quality objectives are intended to support an assessment of the extent to which a model can be used for policy-relevant analyses, but do not point to the underlying processes that might lead to a disagreement between model results and observations. Furthermore, the calculation of the NO_2 measurement uncertainty underlying the calculation of the MQO and performance criteria is also based on a number of uncertain parameters.

We thus complement the analysis with a diagnostic evaluation, comparing the individual spectral components of the modelled and observed time series. This is done following Solazzo and Galmarini (2016) and Solazzo et al. (2017): we use a Kolmogorov-Zurbenko filter (Zurbenko, 1986), a widely used filter in the analysis of air quality data based on calculating the iterative moving average of a time series, in order to decompose the modelled and observed time series into contributions from different time scales. The Kolmogorov-Zurbenko filter is a low pass filter, with the length of the moving average window and the number of iterations determining the spectral component to be filtered. Taking the difference between two filtered time series (band-pass filter) makes it possible to decompose the observed and measured time series into an intra-diurnal component (ID, < 0.5 days), a diurnal component (DU, 0.5-2.5 days), a synoptic component (SY, 2.5-21 days) and a long-term component (LT, >21 days) with the property

$$TS(x) = LT(x) + SY(x) + DU(x) + ID(x). \quad (3)$$

Here, TS describes the full time series of the species x. This is described in detail in Solazzo et al. (2017) and Solazzo and Galmarini (2016) and references therein. Further detail is also given in the Supplementary Material.

By assessing the error of each component individually it is then easier to relate the error to the model process(es) characteristic at the respective time scale. The error analysis of the different spectral components is done by "error apportionment" (Solazzo et al., 2017), breaking down the mean square error (MSE) into bias, variance (σ) error and minimum achievable mean square error (mMSE) as follows:

$$\mathrm{MSE} = (\mathrm{mod} - \mathrm{obs})^2 + (\sigma_{\mathrm{mod}} - \mathrm{r}\sigma_{\mathrm{obs}})^2 + \mathrm{mMSE} \quad (4)$$

As described by Solazzo and Galmarini (2016), the minimum achievable mean square error is determined by the observed variability that is not reproduced by the model. While this approach helps investigating the sources of model errors, it does not allow for clearly identifying or quantifying them as several processes take place on similar time scales, and because this filtering method does not allow for a complete separation of the different spectral components (see Solazzo et al., 2017, for a discussion of this issue).

In addition to this operational and diagnostic analysis of simulated NO_2 concentrations, we include a brief evaluation of selected key meteorological parameters (temperature, wind speed and direction) as well as further chemical species (O_3, NO_x)~~as~~, the former because WRF-Chem is an online-coupled model~~and~~, and the latter because NO_2 is tightly linked to NO and O_3.

4.2 Observation-based analysis

As traffic emissions are the focus of this study, the analysis of the model results is complemented with an analysis based on observations of roadside and urban background NO_2 concentrations and traffic counts. Like in many chemistry transport modelling studies, we assume a linear scaling of traffic emissions with traffic counts, which are used as a proxy for calculating time profiles of traffic emissions for each month, day of the week and hour of the day. While it has been shown that model results can be improved by taking into account country-specific driving patterns as well as by applying separate diurnal cycles for heavy and light duty vehicles (Mues et al., 2014), local traffic conditions (e.g. congestion) are currently not taken into account in the calculation of the diurnal cycles.

Using observations of traffic counts and roadside NO_x concentrations in Berlin obtained at the same locations and times (data described in Sect. 2.4.2 and 3.1), we assess how much of the observed variance in NO_x concentrations can be explained with traffic counts in a linear model. In addition to a linear fit, other types of relationships (e.g. quadratic, exponential) are also explored. We neglect other influences on observed NO_x concentrations such as other emission sources and large-scale and local meteorological conditions. In order to account for different conditions at different hours of the day, we fit the data separately for each hour of the day. The intention of this analysis is not to build a statistical model for roadside NO_x concentrations, but rather to give insight into the type of relationship between roadside NO_x concentrations and traffic counts, complementing the model simulations done in this study.

5 Model evaluation

5.1 Meteorology

An in-depth evaluation of modelled meteorology obtained with a similar model setup is presented in Kuik et al. (2016) for the summer (JJA) of 2014. Here, model results for the whole year of 2014 are presented and discussed. Changes in the model setup compared with the setup presented in Kuik et al. (2016) are the planetary boundary layer scheme (MYNN, Nakanishi and Niino (2006) instead of YSU, Hong et al. (2006)) and re-initialization of the model meteorology every 2 days as described in Sect. 2. Tests showed that though the change in planetary boundary layer scheme did not introduce considerable improvements, it did seem to lead to a slightly better match of model results with observations in the timing of the decrease of the boundary layer in the evening. Here an additional brief model evaluation is done in order to ensure that the modelled meteorology still reproduces observations reasonably well.

Modelled and observed ~~temperature and~~ 2m temperature and 10m wind speed are compared at five stations run by the German Weather Service, including Schönefeld, Tegel and Tempelhof in Berlin and Lindenberg and Potsdam outside of Berlin (Table 3). Across the stations, annual mean temperature is simulated well, with mean biases smaller than -1°C outside of Berlin and just above -1°C within Berlin. Modelled and observed hourly temperatures correlate well with R=0.96 at all five stations. Small seasonal differences exist, with somewhat higher biases in winter (as large as -1.7°C in Tegel) and somewhat lower biases in spring (e.g. -0.1°C in Schönefeld). Annual mean wind speed is somewhat overestimated within Berlin (between 0.02 m/s

and 0.45 m/s, or up to 13%), with correlations of the hourly values between 0.74 and 0.78 within Berlin. In winter, wind speed is slightly underestimated at two out of the three stations within Berlin (-2% and -7% at Tegel and Schönefeld, respectively), while it is overestimated somewhat more in spring and summer (up to 0.58 m/s, or ca. 20% in Tegel). In spring and summer, the main wind directions are captured relatively well by the model (see Fig. ~~S2 and S3~~ S6 and S7 in the Supplementary Material). In autumn, wind from the east, the main wind direction, is modelled less frequently than observed, but wind from the south-east is modelled too frequently compared with observations. In winter, modelled wind comes from south and south-west too frequently compared with observations, at the expense of south-easterly wind directions, as depicted in Fig. ~~S2 and S3~~S6 and S7. Compared with Kuik et al. (2016), an improvement in summer mean bias in wind speed is seen; with the JJA mean bias between 0.3 and 0.4 m/s smaller than that of the comparable simulation in Kuik et al. (2016) at all Berlin stations, and JJA correlation coefficients improved by ca. 0.1. This can probably be attributed to the continuous re-initialization of modelled meteorology in this simulation.

In addition, modelled and ceilometer-derived mixing layer heights (MLH) are compared (Fig. ~~S4~~ S8 in the Supplementary Material). Even though a quantitative comparison between the modelled MLH and the MLH height derived from optical measurements is difficult to interpret (see Kuik et al., 2016), a qualitative comparison of mean diurnal cycles gives insight into the timing of the deepening of the MLH. The comparison shows that the modelled increase of the summer MLH in the morning is too early, already starting at ca. 4 am in the model. Though the precise time of the observed MLH increase cannot be determined from the available data, it takes place between 5am and 7am (Fig. ~~S4~~ S8 in the Supplementary Material). An early modelled deepening of the mixing layer might lead to a too early and thus too strong mixing of chemical species in the model.

5.2 Operational evaluation of simulated chemical species

Seasonally and station-class averaged performance metrics are listed in Table 4 for NO_2, NO_x and O_3. NO_2 and total NO_x are biased low throughout the seasons and station classes, with the highest (absolute and relative) mean biases for urban background stations both annually and seasonally. The model bias is relatively low at rural and suburban background stations, with annual mean biases of only up to -2.8 µg m^{-3} (-19%). Correlation coefficients of modelled with observed hourly concentrations are R=0.50 and R=0.55 in the rural and suburban background, respectively.

NO_2 at urban background sites is biased by -7.8 µg m^{-3} (-29%) on average, with a higher negative bias in spring (-10.2 µg m^{-3}, -38%) and summer (-9.3 µg m^{-3}, -41%) and smaller negative biases in autumn (-4.9 µg m^{-3}, -17%) and winter (-6.8 µg m^{-3}, -22%). Modelled hourly concentrations correlate reasonably well with observations in autumn, spring and winter (R between 0.51 and 0.55), but worse in summer (0.36).

Modelled hourly ozone concentrations correlate reasonably well with observations at all station classes throughout the whole year (R between 0.70 and 0.73), but with lower correlations for individual seasons. This shows that intra-seasonal differences are represented well by WRF-Chem, with slightly worse representations of inter-seasonal variations. Modelled ozone concentrations are biased high at most stations and in most seasons, with the exception of a low bias in summer in the urban background.

For NO_2, the MQO (Eq. 2) is greater than 0.5, but smaller than 1, both annually averaged and in all seasons at rural near-city background and suburban background stations. For urban background sites the MQO is larger than 1 both on annual average and in spring and summer, and just below 1 in autumn and winter, emphasizing that the model performs reasonably well in the rural and suburban background, but the disagreement between model results and observations is larger in the urban background. This suggests that processes or emissions typical for urban areas are an important source of model error.

In order to test the sensitivity of the results to the selected chemical mechanism, we compare modelled NO_2 and total NO_x concentrations for July with two different chemical mechanisms: RADM2 (the base configuration in this study) and MOZART. For all station classes in and around Berlin, the modelled NO_x and NO_2 concentrations only show very small mean differences of -0.04 to -0.4 µg m^3 (NO_x) and -0.4 to -0.5 µg m^3 (NO_2, RADM2 - MOZART). This suggests that the model bias in NO_2 and total NO_x concentrations of the base configuration is not strongly influenced by the choice of chemical mechanism, but rather results from other sources of error.

5.3 Diagnostic evaluation of simulated NO_2 concentrations

In order to further assess the model performance and identify main sources of the model bias, a diagnostic evaluation is done, by spectrally decomposing the modelled and observed time series of NO_2 and analysing the type of error of each component.

Averaging the decomposed time series over each station class, the modelled long term (LT) and synoptic (SY) components as defined in Sect. 4.1 correlate well with the observations: the correlation coefficient for the LT component is 0.83, 0.81 and 0.72 for rural near-city, suburban and urban background, respectively, and 0.60, 0.63 and 0.65 for the SY component (Fig. 3). This suggests that changes on time scales of ca. 2.5 days to a few weeks are captured relatively well by WRF-Chem, which includes for example the modelled synoptic (meteorological) situation and is consistent with the good model performance in simulating observed meteorology. The correlation coefficients for the diurnal (DU) component are smaller, with 0.45, 0.52 and 0.48 for rural near-city, suburban and urban background, respectively. This suggests that the model has more difficulties in capturing variations at time scales of a few hours to 2.5 days than on longer time scales. This might be related to the diurnal variations in modelled mixing, but also to the diurnal cycle of emissions. Particularly the latter is strongly influenced by traffic emissions in the urban area and might also point to deviations of the model-prescribed diurnal cycle in emissions from the real-world diurnal cycle.

With the procedure used for spectrally decomposing the NO_2 time series, the LT component is the only systematically biased component, with the other components fluctuating around zero. Decomposing the model error shows that the bias of the LT component has the largest contribution to the error for urban background stations (ca. 30%, Fig. 4). NO_2 has a short life time and is mainly influenced by local and regional sources. This means that the boundary conditions are not likely to be a strong source of error. The negative bias in the LT component is consistent with both problems in daytime vertical mixing and an underestimation of emissions. As discussed in Sect. 3.1, NO_2 concentrations detected with chemiluminescence using a molybdenum converter might be biased high due to interferences with other nitrogen-containing species (e.g. PAN, HNO_3) and could further contribute to discrepancies between modelled and observed NO_2 concentrations.

The second largest error at urban background stations and the largest error at rural near-city and suburban background stations is the mMSE of the diurnal component. This means that part of the observed variability is not reproduced by the model and is consistent with the comparably lower correlation coefficients of the diurnal component compared with the synoptic and long term components. Solazzo et al. (2017) relate this error to problems in comparing single point measurements with model grid cell values (incommensurability) and a disagreement in timing of modelled and observed concentrations, amongst others. The incommensurability can, in the case of NO_2, come from NO_2 observations being influenced by local sources that cannot be captured by WRF-Chem run at a horizontal resolution of 3 km x 3 km. The temporal variation of modelled NO_2 concentrations, in case of the diurnal component, can be influenced by the temporal profiles prescribed to the emission input data. Thus, the error is consistent with problems in the prescribed diurnal cycles of emissions including traffic emissions, but might also be related to a diurnally varying bias in emissions.

At rural near-city background stations, there is a relatively large contribution of the variance error of the diurnal component. This is probably caused by an overestimation of the standard deviation of observed diurnal components in autumn (Fig. ~~S5~~ S9 in the Supplementary Material), particularly pronounced at the site Frohnau in the north or Berlin, slightly west of the main emission sources. This might be explained by the disagreement in modelled and observed wind direction in autumn, leading to higher than observed NO_2 peaks in the model.

Solazzo et al. (2017) present a diagnostic model evaluation of the AQMEII phase 3 model simulations for the year 2010 and report the largest error of modelled NO_2 in winter, both for the European and North American domains simulated in AQMEII. Our results show the opposite for urban background stations (Fig. ~~S5~~ S9 in the Supplementary Material): the model error, and particularly the bias, is smallest in autumn and winter. While Solazzo et al. (2017) attribute the winter bias to a potential underestimation in residential combustion emissions, these seem to be captured comparably well by the TNO-MACC III inventory in the case of Berlin. The re-distribution of these emissions based on population density, as described in Sect. 2.4.2, may also have contributed to a better spatial representation in our study.

5.4 Diurnal and weekly variation of the model bias

The results from the operational and diagnostic evaluation of modelled NO_2 concentrations suggest that emissions within the urban area are a main source of model error, both contributing to the model bias and the lower correlation with observations. Traffic emissions have the largest contribution to urban NO_x emissions. As traffic emissions have a distinct weekly and diurnal cycle, we additionally assess mean diurnal cycles of modelled and observed NO_2 concentrations as well as the differences between weekdays and weekends. This also helps to further assess the contribution of problems in modelled mixing to the model error. In addition, we analyse the MQO and performance criteria separately for weekends (Saturday and Sunday) and weekdays (Monday through Friday). Public holidays that fall on a weekday are excluded from this analysis, as they were not treated separately from regular weekdays in the emission processing.

The comparison of mean modelled and observed NO_2 diurnal cycles shows distinct differences between station classes and weekend and weekday diurnal cycles (Fig. 5). The diurnal cycle of observed NO_2 concentrations is modelled reasonably well for rural and suburban background stations. In particular, nighttime concentrations are simulated well for rural and suburban

background stations, and mostly underestimated in the urban background. Other WRF-Chem modelling studies often report too little mixing at nighttime over urban areas leading to a strong overestimation of observed concentrations. In this study as in other modelling studies using WRF-Chem (Ravan Ahmadov, pers. comm.), a modification of the model code was applied in order to increase nighttime mixing. This, in combination with a more realistic vertical distribution of point source emissions (as described in Section 2.4.2), seems to improve model performance for NO_2 during nighttime. In addition, tests revealed that this change to the model code does not impact modelled daytime concentrations.

During weekdays, there is an underestimation of the observed morning peak in all seasons and at all station classes. Weekend diurnal cycles are modelled well at rural and suburban background stations. At urban background stations there is a larger disagreement between modelled and observed concentrations throughout the whole day on both weekends and weekdays. The underestimation of daytime urban background NO_2 concentrations is particularly strong in summer and spring. This might be explained by mixing over urban areas during daytime that is too ~~efficient~~strong, caused for example by a turbulent diffusion coefficient that is too large during daytime over urban areas in the lowest model layer. Other modelling studies have reported similar problems, reducing the coefficient over urban areas (e.g. of the CHIMERE model setup used in Schaap et al., 2015). An onset of the deepening of the boundary layer that is too early (Sect. 5.1) might further contribute to the disagreement in the modelled and observed morning peaks. Overall, this discussion shows that the representation of vertical mixing over urban areas might have to be improved to be physically more consistent in regional models, for example by better taking into account urban heat and momentum fluxes and treating the urban parameterization consistently with chemistry. Measurements of vertical profiles of NO_x in cities, particularly in the planetary boundary layer, would be helpful in order to evaluate the models and improve the representation of surface NO_x concentrations, as the NO_x profile in the lowest model layer is not resolved at the model resolution used in this study.

The model underestimation of observed daytime NO_2 concentrations at urban background stations is stronger on weekdays than on weekends, and is particularly noticeable during the morning hours. This is consistent with an overall underestimation of emission sources active in the morning hours on weekdays and potentially also a misrepresentation of the diurnal cycles of emissions in the model~~. Traffic~~ : traffic emissions are distributed in the model throughout the day using a linear scaling with traffic counts (Sect. 2.4.2), which might fall short of accounting for relatively higher emissions during situations with high traffic and associated congestion. This issue is further assessed in Sect. 6.3.

Generally, throughout all seasons, the NO_2 MQO is not met on weekdays for urban background stations, but is smaller than 1 on weekends (Fig. 6). The pattern of the model-observation disagreement, and particularly the weekend-weekday differences, are consistent with traffic emissions as a main source of the bias, having a particularly large contribution to observed urban background concentrations.

6 Top-down quantification of NO_x emissions from traffic

6.1 Calculation of a correction factor

The results from the operational and diagnostic evaluation of modelled NO_2 concentrations suggest that traffic emissions are a main source of model error in the urban background: the bias and the mMSE of the diurnal component have the largest contribution to the model error in the urban background throughout all seasons, which is consistent with both an underestimation of the magnitude of traffic emissions, and a problem with their temporal distribution. This is further supported by the smaller (absolute and relative) daytime bias of modelled NO_2 concentrations on weekends, where there is less traffic. In the following, we derive a correction factor based on this model bias, which represents the degree to which traffic emissions are underestimated in Berlin, but also takes into account that other sources of model error are likely to also contribute to this bias.

Besides biases in traffic emissions, problems in modelled mixing, which is particularly relevant in summer and spring when the mixed layer is deeper than in other seasons, might contribute to the model bias. Other contributions to the NO_2 bias might come from deviations of modelled from observed wind speed in certain periods, and a potential overestimation of NO_2 in the observations by detection of other nitrogen containing compounds as discussed above. These sources of error are likely to impact the model results equally on both weekends and weekdays, whereas an underestimation of traffic emissions will have the largest impact on the results on weekdays. For the quantification of the underestimation of traffic emissions we assume that the weekend bias is entirely caused by non-traffic-emission-related sources of error and thus use the difference between weekday and weekend bias as an estimate for the traffic related bias. We use the weekday-weekend difference of the relative biases (Fig. 7), thus assuming that the model error due to other sources than traffic emissions roughly scales with the magnitude of modelled concentrations. These are both conservative assumptions, as the correction factor would be much larger if the whole weekday bias was regarded as caused by traffic emissions, and it would also be larger if the absolute weekday/weekend difference was used.

In order to estimate the correction factor for traffic NO_x emissions, we combine the weekday increment of the model bias as defined above with the average fraction of NO_x emissions from traffic to total NO_x emissions in Berlin. The nighttime model bias on weekends and weekdays at urban background stations is of similar magnitude on weekends and weekdays (Fig. 5). A t-test shows that the differences between weekday and weekend bias are not statistically significant at a 95% confidence interval after ca. 17:00 UTC and before ca. 5:00 UTC (depending on the season). Furthermore, traffic emissions used in the model contribute only little to the total NO_x emissions before 6:00 UTC. This suggests that an underestimation of traffic emissions is only likely to have a significant contribution to the bias in modelled NO_2 concentrations between ca. 6:00 and 17:00 UTC. Within the core area of the city where traffic is high (all areas within the "S-Bahn ring"/main core of the city), the average contribution of traffic NO_x to total NO_x between 6:00 and 17:00 UTC is between ca. 30% and 55%, depending on the month and hour of the day. Seasonal average values over the indicated time period are used for the calculation of the correction factor, with 37% in winter, 47% in summer and 42% in autumn and spring.

With the above assumptions, we quantify the underestimation of traffic NO_x emissions in the core urban area on weekdays between 6:00 and 17:00 UTC as follows, calculating a correction factor ~~F~~f_{NOx}:

$$f_{NOx} = \frac{1}{1+NMB} \cdot \frac{1}{s_t}, \quad (5)$$

With the (negative) NMB = $\frac{mod-obs}{obs}$, and s_t denoting the traffic share of NO_x emissions. Averaged over all urban background stations, and all seasons, as well as the time period between 6:00 and 17:00 this results in a correction factor of ca. 3. When averaged over all hours of the day, this factor corresponds to an overall underestimation of NO_x traffic emissions in the urban centre by a factor of ca. 2, and an underestimation of all-source NO_x emissions in the urban centre by a factor of ca. 1.5.

In order to gain more insight into the underestimation of the NO_x emissions, we calculate a separate correction factor for each hour and season based on hourly mean seasonal biases and traffic NO_x emission shares (Fig. ~~S6~~ S10 in the Supplementary Material). The seasonal correction factors show a small increase between 6:00 and 8:00 with a subsequent decrease, and then remain relatively constant from 11:00 to 17:00. The diurnal variations of the factors for the different seasons are qualitatively similar, and the factors vary in magnitude within a range of ca. 1 between the seasons, with the factors being larger in winter than in summer. The diurnal cycle of the correction factor could be due to a diurnally varying importance of other sources of the modelled NO_2 bias than the traffic emissions, such as mixing, but might also be due to a disagreement in the prescribed diurnal cycle of traffic emissions with the real-world diurnal cycle of traffic emissions. The seasonal differences can at least partly be explained with the seasonally varying relevance of other sources of model error, such as mixing, which has a bigger impact in summer and thus also leads to a bigger bias on the weekends, reducing the weekday increment. The seasonal differences might also be influenced by the temperature dependence of NO_x emissions in newer diesel cars (Hausberger and Matzer, 2017), leading to higher NO_x emissions at colder temperatures, which are not captured by the model.

Overall, the assumptions in these calculations are rather conservative: assuming the weekend bias is not caused by an underestimation of traffic emissions at all is likely to underestimate the effect of any traffic bias. As mentioned above, using the absolute weekday increment of the bias would also lead to higher correction factors. A further discussion of the model bias and correction factor looking into potential reasons contributing to an underestimation of traffic NO_x emissions is presented in Sect. 6.4.

6.2 Sensitivity simulation with increased emissions

The weekday correction factor was applied to NO_x traffic emissions for the core urban area of Berlin (within the "S-Bahn Ring") and tested in two sensitivity simulations for January and July 2014. The results (Table 5 and Fig. 8) show that the bias of modelled NO_2 concentrations at urban background stations decreases on average by 2.6 µg m^{-3} (NMB decreases from -24% to -16%) in January, and by 2.0 µg m^{-3} (from -43% to -34%) in July when applying the correction factors for NO_x emissions from traffic. The decrease is larger when only considering weekdays, with a mean bias lower by 3.4 µg m^{-3} (from -26% to -16%) in January and by 2.7 µg m^{-3} (from -46% to -34%) in July. NO_2 concentrations on weekends are still represented reasonably well by the model in January (Fig. 8).The weekend bias is only changed (decreased) by lower than 0.4 µg m^{-3} in both cases. Only a minor change would be expected, since emissions on the weekend are not changed in the sensitivity

simulations, compared to the base simulation. In January, the correlation of modelled with observed NO_2 concentrations in the urban background is improved by between 0.03 and 0.06 for urban background stations in the sensitivity simulation, but this is not the case in July (Table 5). The lack of improvement in the July correlation coefficient could be related to nighttime concentrations in July that seem to be very sensitive to the increase in emissions during daytime (Fig. 8, lower panel). Despite an improved representation of nighttime concentrations compared to a previous study (Kuik et al., 2016), this sensitivity suggests the need for further attention to mixing processes in urban areas in high resolution chemistry transport models.

Bigger improvements are seen when comparing total NO_x: the mean bias for urban background stations is reduced from -16.4 to -10.3 µg m^{-3} (NMB decreased from -35% to -22%) in January and from -11.1 to -8.1 µg m^{-3} (from -45% to -33%) in July. Only considering weekday concentrations, these are improved by 8.1 and 3.9 µg m^{-3} (from -37% to -12% and -48% to -33%) in January and July, respectively. The differences in NO_2 and NO_x improvements suggest that the impact of the primary NO_2 fraction in emitted NO_x on ~~observed and modelled concentrations~~ modelled NO_2 and NO_x concentrations, as well as the influence of chemical processes such as NO titration and other relevant physical and chemical processes, might need to be assessed in greater detail.

While on average the normalised mean bias in modelled rural and suburban background is only reduced by 1-2% in both January and July, the simulation of NO_2 and NO_x concentrations downwind of the city centre is improved considerably in the sensitivity simulation. For analysing the change in modelled downwind concentrations, the results are broadly divided based on four main wind directions (N, W, S, E). For each wind direction bin, the results of two stations outside the core urban area are analysed, with Frohnau and Buch in the North, Johanna und Willi Brauer Platz and Mueggelseedamm in the East, Schichauweg and Blankenfelde-Mahlow in the South and Gross Glienicke and Grunewald in the West. Only situations with wind speeds above 2 m/s are considered. The statistics are calculated for stations where at least 72 hourly model-observation pairs exist in the respective wind ~~speed~~ direction bin, leaving 4 stations in January and 6 stations in July for the analysis, with between 91 and 228 model-observation pairs. With some differences between the stations, the bias of weekday downwind NO_2 concentrations was reduced by between ca. 1.5 and ~~3~~ 2.9 µg m^{-3} January and ca. ~~0.7~~ 0.4 and 1.5 µg m^{-3} in July. Thus, downwind NO_2 concentrations in the sensitivity simulation are only biased by ca. -4% (January) and -14% (July) on average (as compared to -12% and -22% in the base run). This shows that the increase in traffic emissions also helps improve modelled downwind concentrations.

Overall, in both January and July, the bias in modelled urban background NO_2 and NO_x is improved but still negative. Modelled downwind NO_2 concentrations are improved considerably, but with low negative biases remaining also in this case. The improvements are consistent with an underestimation of traffic emissions being a main source of error. However, the results also suggest that on the one side traffic emissions might still be too low, which is consistent with the correction factor being a rather conservative estimate. On the other side, a still negative bias is also consistent with other sources of error contributing considerably to the model-observation differences as discussed previously. A relatively large bias in July remains, consistent with the mixing being an additional main source of error particularly in summer.

Modelled O_3 concentrations are not very sensitive to the changes in NO_x concentrations. On average, modelled O_3 is reduced at urban background stations in January by 1.5 µg m^{-3} (NMB decreases from 29% to 22%). In July, the increased NO_x leads to

a reduction in the already negatively biased O_3 from the model, with the mean bias changing from -7.3 to -8.6 µg m^{-3} (-11% to -13%). Similarly, simulated O_3 concentrations downwind of the city (in analogue to the downwind NO_2 concentrations described above) are biased negatively in both base run and sensitivity study in July. The bias of downwind concentrations changes from -5.4 µg m^{-3} (-7%) in the base run to -6.8 µg m^{-3} (-9%) in the sensitivity run. The negative bias in both NO_x and O_3 in the base run is consistent with the model simulating insufficient NO_x emissions in a NO_x-limited ozone production regime. The reduction of O_3 concentrations in response to increased NO_x emissions is however consistent with the model actually being in a NO_x-saturated (VOC limited) ozone production regime. The representation of VOC emissions in the model could play a role in explaining this discrepancy, as for example biogenic VOC emissions in the Berlin-Brandenburg urban area are underestimated when using WRF-Chem and MEGAN (Churkina et al., 2017). A comprehensive analysis of the simulated ozone production regime is beyond the scope of this work.

6.3 Analysis based on traffic counts

The model bias and the calculated correction factors show a diurnal cycle, with a larger model bias/correction factor in the morning hours. As explained in Sect. 5.3 and 5.4, one reason for this might be differences between prescribed and real-world diurnal cycles of the emissions. The diurnal cycle of traffic emissions in the model is calculated based on traffic counts for Berlin, assuming a linear scaling of traffic emissions with traffic counts, as done in many modelling studies.

Here, we use three years of hourly observations of roadside NO_x concentrations and traffic counts measured at the same stations in order to get insights into the relationship between NO_x concentrations and traffic counts. A linear regression model does not explain the variance of observed NO_x concentrations at nighttime, as indicated by the R^2 close to 0 in Fig. 9. However, during daytime, traffic counts alone explain up to ca. 40% of observed NO_x variance, particularly during the traffic rush hours. The explained variance is smaller during the afternoon peak. In comparison to a linear ~~model~~relationship, a quadratic ~~model~~ relationship ($NO_x \propto (\text{traffic\_count})^2$) does not explain more of the observed variance (not shown). An exponential ~~model~~ relationship ($NO_x \propto \exp(\text{traffic\_count})$), however, does explain a considerably larger share of the observed variance during daytime and particularly during the traffic rush hours, as depicted in Fig. 9 (up to ca. 60% depending on the station).

This simple comparison suggests that roadside NO_x concentrations, and thus most likely also road transport NO_x emissions, scale more than linearly with traffic counts at times when the traffic intensity is high and underline that the assumption of a linear scaling of traffic emissions with traffic counts does not reflect the diurnal variation of traffic emissions sufficiently. More highly congested roads are typical in the morning, and emission factors (e.g. from HBEFA) are higher in congested situations compared to free flowing traffic. Differences in congestion could contribute to explaining the non-linear scaling of NO_x concentrations with traffic intensity. While the impact might not be large when simulating air quality with coarser models, it might play a more important role for high resolution air quality modelling, and the temporal distribution of emissions could potentially be improved when taking these differences into account.

6.4 Discussion of traffic emissions

Based on a comparison of modelled with observed NO_2 concentrations, we estimate that traffic emissions in the urban core of Berlin are underestimated by a factor of ca. 3 on weekdays between ca. 6:00 and 17:00 UTC. This corresponds to an overall underestimation of NO_x traffic emissions (all day average) in the urban centre by a factor of ca. 2, and an underestimation of total NO_x emissions (all day average) in the urban centre by a factor of ca. 1.5. Reasons for the underestimation of emissions used in this study can include limitations in the applicability of the emission inventory used here for high resolution urban air quality modelling, problems in the temporal distribution of emissions, but also a general underestimation of traffic NO_x emissions in the inventories. These three points are discussed further in the following.

First, while a reasonably good model performance can be achieved using the downscaled version of the TNO-MACC III inventory outside of the urban areas, the deviations of modelled from observed NO_2 in the urban background might point to limitations in the applicability of these types of emission inventories for high resolution modelling of NO_2 in urban areas. The horizontal resolution of the original TNO-MACC III emission data is ca. 7 km x 7 km and national totals are disaggregated on the grid based on traffic intensities. Spatial differences in congestion, with emissions greatly varying between the different driving conditions and with car speed (e.g. Hausberger and Matzer, 2017), are probably not well resolved. A comparison of the downscaled version of the TNO-MACC III inventory for Berlin with a local inventory has, however, not revealed major differences in road transport emissions (see Sect. 2.4.3), suggesting that a static highly resolved local inventory based on detailed local information is not likely to improve the model results by much.

Second, in addition to spatially unresolved differences in driving conditions and related emission factors locally increasing the underestimation of emissions, the diurnal cycle of the bias in all seasons suggest that the diurnal cycle of traffic emissions also does not sufficiently account for temporal differences in driving conditions. This is consistent with the observation-based analysis, suggesting that observed NO_x concentrations do not scale linearly with traffic counts. While these assumptions might be valid for coarser model resolutions, they may need to be revisited when going to higher resolutions with a focus on urban areas. However, modelled NO_2 concentrations are broadly underestimated throughout the day, which means that deviations of the model diurnal cycle from the real-world diurnal cycle alone cannot explain the underestimation of modelled NO_2 and NO_x concentrations.

Third, traffic NO_x emissions may be underestimated generally by emission inventories. The correction factor calculated here is in line with the results from other studies quantifying traffic emission underestimations in Europe, reporting traffic NO_x underestimations of around 80% (Lee et al., 2015), a factor of 1.5-2 (Lee et al., 2015) and up to a factor of 4 (Karl et al., 2017). A potential reason for the underestimation in NO_x emissions from traffic can be discrepancies between real-world emission factors and those used in emission inventories. Even though HBEFA emission factors, which are often used for calculating emissions, are based on real-world driving conditions, the latest update of the handbook reports higher emission factors than previously assumed for Euro 6 and Euro 4 diesel cars (Hausberger and Matzer, 2017), e.g. an increase by ca. 50% in case of Euro 6 vehicles (Fig. 14 in Hausberger and Matzer, 2017). In addition, the update assesses the temperature dependence of emission factors and concludes that it may lead to increases in NO_x emissions of more than 30%, compared with standard test

conditions. NO_x emissions from diesel cars increase with decreasing temperatures (Hausberger and Matzer, 2017). This may also contribute to the larger correction factor calculated for the winter months. Finally, while some amount of congestion is included/assumed in the emission inventories, this might be an aspect that is underestimated in terms of severity and extent.

The first and second point of this discussion suggest that improvements might be achieved by combining high resolution chemical transport models with more detailed approaches of calculating emissions. Coupling with a traffic model, for example, might allow for not only being able to take local difference in traffic conditions into account, but also prescribe a more realistic diurnal cycle of traffic emissions. Dispersion modelling and street canyon modelling (e.g. OSPM, Berkowicz, 2000) often already take a more detailed calculation of traffic emissions into account, and different emission modelling approaches exist (e.g. traffic models such as MATSim, Horni et al., 2016). The benefit of high resolution chemistry transport modelling, e.g. their ability to assess the impact of different emission sources on air quality on larger scales and downwind of the main emission sources, could be further exploited if coupled with existing, more detailed approaches in calculating traffic emissions or general improvements in the accuracy and resolution of emission inventories.

The consistent findings that inventories of European traffic emissions may be underestimated, coming from studies using very different methodologies, suggests that further research is necessary in order to understand real-world traffic emissions and represent them in the inventories accordingly. Alternative measurement approaches could help verify the assumptions underlying the calculation of emissions, and help identify potential systematic problems.

7 Summary and conclusions

Several modelling studies, particularly for Europe, have reported an underestimation of modelled NO_2 concentrations compared with observations. Measurement studies also suggest that there might be considerable differences between measured urban NO_x emissions and emissions provided by emission inventories based on official reporting, particularly when the contribution of traffic is large. This study quantifies the underestimation of traffic NO_x emissions using WRF-Chem in a top-down approach, with the Berlin-Brandenburg area in Germany as a case study. The emission inventory used here is TNO-MACC III, downscaled to 1 km x 1 km over the Berlin-Brandenburg area based on local proxy data. The downscaled traffic emissions averaged over Berlin only differ by 6% from a local bottom-up traffic emission inventory.

A diagnostic evaluation of the model results shows that particularly in the rural and suburban background, the long term and synoptic components representing processes at time scales of the order of 2.5 to 21 (synoptic) and longer than 21 days (long term) are simulated well by the model. This suggests that the modelled impact of meteorology on concentrations is represented well overall. The largest contribution to the model error comes from the (negative) bias in the urban background, and from deviations of modelled from observed variability of the diurnal component (0.2-2.5 days). This suggests a possible underestimation of urban emissions, of which traffic is the single most important contributor to NO_x emissions, but is also consistent with deficiencies in other processes varying on the diurnal scale such as the modelled mixing in the planetary boundary layer. The analysis of the model results suggests that the latter is particularly relevant in summer and spring, and that further research is needed in order to better represent urban processes their coupling with chemistry in WRF-Chem. For

example, the changes in the model code applied here to improve nighttime mixing can be critically discussed, and would ideally be replaced by an improved parameterization of urban processesneeds . The latter would need to better account for urban heat and momentum fluxes for a more realistic representation of mixing both at daytime and at nighttime, particularly in summer. An alternative model configuration to be tested could be the recently extended ACM2 planetary boundary parametrization (Pleim, 2007) , which now conducts mixing of chemical species within the planetary boundary layer scheme. In addition, measurements of vertical profiles of NO_x in urban areas are needed to evaluate and improve models for applications in urban areas.

The analysis of the diurnal cycle of the model bias as well as a simple observation-based calculation showing that roadside NO_x concentrations scale non-linearly with traffic counts suggest that a further source of error is likely the prescribed diurnal cycle used for traffic emissions. In this study as well as in many other modelling studies, the diurnal cycle of traffic emissions is calculated assuming a linear scaling of traffic emissions with traffic counts. While this might be sufficient for coarser model resolutions, high resolution urban air quality modelling with chemistry transport models might benefit from a more detailed temporal distribution not only taking into account traffic intensity via a scaling with traffic counts, but also diurnal differences in congestion.

We quantify the underestimation of traffic emissions based on the finding that the weekday bias in modelled NO_2 is larger than on weekends and that the contribution of traffic NO_x to total NO_x emissions in the urban area is typically higher on weekdays. The results suggest that traffic emissions are underestimated by ca. a factor of 3 in the core urban area on weekdays when traffic is highest (6:00 to 17:00 UTC). The underlying assumption is that other sources of model errors influence the model bias equally on weekdays and weekends, with the underestimation of traffic emissions having the largest effect on modelled NO_2 concentrations on weekdays. This underestimation corresponds to an underestimation of weekly mean traffic NO_x emissions in the core urban area of ca. a factor of ca. 2 and an underestimation of total NO_x emissions in the city centre by a factor of ca. 1.5. Two sensitivity simulations for January and July 2014 with NO_x emissions from traffic scaled with the estimated correction factor show that increased traffic emissions improve the model bias in NO_2 and NO_x concentrations in both seasons in the urban background, and also improved improve modelled downwind concentrations. The still negative bias is consistent with the factor being a rather conservative estimate.

The emission inventory used in this study is based on officially reported emissions by the individual countries, and the emissions are spatially distributed by TNO based on proxy data. Assuming the quality and accuracy of the proxy data is similar at least for larger German cities, and considering that modelling studies for other German cities have also shown an underestimation of simulated NO_2 concentrations using the same emission inventory, we would assume that the results found in this study for Berlin may generally be transferrable to at least other German metropolitan areas. The underestimation of NO_2 concentrations throughout the day, the consistency of the calculated correction factor with findings from other studies and the improvement of model results applying the correction factor suggest that more research is needed in order to more accurately understand the spatial and temporal variability in real-world NO_x emissions from traffic, and apply this understanding to the inventories used in high resolution chemical transport models. Given the above considerations, this not only holds for the urban area of Berlin, but for German and most likely European metropolitan areas more generally.

8 Code availability

WRF-Chem is an open-source, publicly available community model. A new, improved version is released approximately twice a year. The WRF-Chem code is available at http://www2.mmm.ucar.edu/wrf/users/download/get_source.html. The corresponding author will provide the modifications introduced and described in Sect. 2.1 upon request.

9 Data availability

The observational and model input data used in this study are publicly available (references indicated in the manuscript) or available upon request.

Acknowledgements. This work was hosted by IASS Potsdam, with financial support provided by the Federal Ministry of Education and Research of Germany (fBMBF) and the Ministry for Science, Research and Culture of the State of Brandenburg (MWFK). The authors would like to thank Hugo Denier van der Gon and Jeroen Kuenen (TNO) for providing data of and information pertaining to the TNO-MACC III inventory and cooperation on the downscaling of the data to a higher resolution; Klaus Schäfer (KIT) for ceilometer data; and Mark Lawrence (IASS), Martijn Schaap (TNO), Ravan Ahmadov (NOAA), Joana Leitao (IASS), Stefano Galmarini (JRC) and Betsy Weatherhead (NCAR) for discussions that helped shape the manuscript.

References

Berkowicz, R.: OSPM—a parameterised street pollution model, in: Urban Air Quality: Measurement, Modelling and Management, pp. 323–331, Springer, doi:10.1007/978-94-010-0932-4_35, 2000.

Berlin Senate Department for Urban Development and Housing: Environment Atlas Berlin, http://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/ed312_01.htm, (last access: December 2015), 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., and White, L.: Presentation of the EURODELTA III intercomparison exercise – evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology, Atmospheric Chemistry and Physics, 16, 12 667–12 701, doi:10.5194/acp-16-12667-2016, https://www.atmos-chem-phys.net/16/12667/2016/, 2016.

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

Briefing, I.: NO_x emissions from heavy-duty and light-duty diesel vehicles in the EU: Comparison of real-world performance and current type-approval requirements, http://www.theicct.org/sites/default/files/publications/Euro-VI-versus-6_ICCT_briefing_06012017.pdf, 2016.

Brunner, D., Savage, N., Jorba, O., Eder, B., Giordano, L., Badia, A., Balzarini, A., Baró, R., Bianconi, R., Chemel, C., Curci, G., Forkel, R., Jiménez-guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Im, U., Knote, C., Makar, P., Manders-groot, A., Meijgaard, E. V., Neal, L., Pérez, J. L., Pirovano, G., Jose, R. S., Schröder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Hogrefe, C., and Galmarini, S.: Comparative analysis of meteorological performance of coupled chemistry-meteorology models in the context of AQMEII phase 2, Atmos Environ, 115, 470–498, doi:10.1016/j.atmosenv.2014.12.032, 2015.

Builtjes, P., Loon, M. V., Schaap, M., Teeuwisse, S., Visschedijk, A., and Bloos, J.: The development of an emission data base over Europe and further contributions of TNO-MEP, https://www.umweltbundesamt.de/en/publikationen/development-of-an-emission-data-over-europe-further, 2002.

Bundesanstalt für Straßenwesen: http://www.bast.de/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/Stundenwerte.html?nn=626916, 2017.

Carslaw, D. C.: Evidence of an increasing NO 2/NOX emissions ratio from road traffic emissions, Atmospheric Environment, 39, 4793–4802, doi:10.1016/j.atmosenv.2005.06.023, 2005.

Churkina, G., Kuik, F., Bonn, B., Lauer, A., Grote, R., Tomiak, K., and Butler, T. M.: Effect of VOC Emissions from Vegetation on Air Quality in Berlin during a Heatwave, Environmental Science & Technology, doi:10.1021/acs.est.6b06514, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A. P., Monge-sanz, B. M., Morcrette, J.-j., Park, B.-k., Peubey, C., De Rosnay, P., Tavolato, C., Thépaut, J.-n., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q J Roy Meteor Soc, 137, 553–597, doi:10.1002/qj.828, 2011.

Degraeuwe, B., Thunis, P., Clappier, A., Weiss, M., Lefebvre, W., Janssen, S., and Vranckx, S.: Impact of passenger car NOx emissions and NO2 fractions on urban NO2 pollution–Scenario analysis for the city of Antwerp, Belgium, Atmospheric Environment, 126, 218–224, doi:10.1016/j.atmosenv.2015.11.042, 2016.

Dennis, R., Fox, T., Fuentes, M., Gilliland, A., Hanna, S., Hogrefe, C., Irwin, J., Rao, S. T., Scheffe, R., Schere, K., and Others: A framework for evaluating regional-scale numerical photochemical modeling systems, Environmental Fluid Mechanics, 10, 471–489, doi:10.1007/s10652-009-9163-2, 2010.

EEA: EMEP/EEA air pollutant emission inventory guidebook 2013 - Technical guidance to prepare national emission inventories, https://www.eea.europa.eu/publications/emep-eea-guidebook-2013, European Environment Agency, Copenhagen, Denmark, 2013.

EEA: CORINE land cover data 2006, updated, http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-3, European Environment Agency, Copenhagen, Denmark, 2014.

EEA: Air quality in Europe - 2016 report, doi:10.2800/80982, https://www.eea.europa.eu/publications/air-quality-in-europe-2016, 2016.

EEA: AirBase -The European air quality database, https://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8, 2017.

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, doi:10.5194/gmd-3-43-2010, 2010.

Fallmann, J., Forkel, R., and Emeis, S.: Secondary effects of urban heat island mitigation measures on air quality, Atmos Environ, 125, Part A, 199–211, doi:10.1016/j.atmosenv.2015.10.094, 2016.

Fast, J. D., Gustafson, W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., and Peckham, S. E.: 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, doi:10.1029/2005JD006721, d21305, 2006.

Geiß, A., Wiegner, M., Bonn, B., Schäfer, K., Forkel, R., Von Schneidemesser, E., Münkel, C., Chan, K. L., and Nothard, R.: Mixing layer height as an indicator for urban air quality?, Atmospheric Measurement Techniques, 10, 2969, doi:10.5194/amt-2017-53, 2017.

Gerboles, M., Lagler, F., Rembges, D., and Brun, C.: Assessment of uncertainty of NO 2 measurements by the chemiluminescence method and discussion of the quality objective of the NO 2 European Directive, Journal of Environmental Monitoring, 5, 529–540, doi:10.1039/b302358c, 2003.

Giordano, L., Brunner, D., Flemming, J., Hogrefe, C., Im, U., Bianconi, R., Badia, A., Balzarini, A., Baró, R., Chemel, C., Curci, G., Forkel, R., Jiménez-guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J., Makar, P., Manders-groot, A., Neal, L., Pérez, J., Pirovano, G., Pouliot, G., José, R. S., Savage, N., Schröder, W., Sokhi, R., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Žabkar, R., Zhang, Y., and Galmarini, S.: Assessment of the MACC reanalysis and its influence as chemical boundary conditions for regional air quality modeling in AQMEII-2, Atmospheric Environment, 115, 371–388, doi:10.1016/j.atmosenv.2015.02.034, http://www.sciencedirect.com/science/article/pii/S1352231015001533, 2015.

Grell, G. and Baklanov, A.: Integrated modeling for forecasting weather and air quality: A call for fully coupled approaches, Atmospheric Environment, 45, 6845–6851, doi:10.1016/j.atmosenv.2011.01.017, 2011.

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, Atmos. Environ., 39, 6957–6975, doi:10.1016/j.atmosenv.2005.04.027, 2005.

Harrison, R. M. and Beddows, D. C.: Efficacy of Recent Emissions Controls on Road Vehicles in Europe and Implications for Public Health, Scientific Reports, 7, 1152, doi:10.1038/s41598-017-01135-2, 2017.

Hausberger, S. and Matzer, C.: Update of Emission Factors for EURO 4, EURO 5 and EURO 6 Diesel Passenger Cars for the HBEFA Version 3.3, http://www.hbefa.net/e/documents/HBEFA3-3_TUG_finalreport_01062016.pdf, 2017.

Hong, S.-y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly weather review, 134, 2318–2341, 2006.

Horni, A., Nagel, K., and Axhausen, K. W.: The Multi-Agent Transport Simulation MATSim, doi:10.5334/baw, 2016.

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., Makar, P. A., Manders-groot, A., Neal, L., Pérez, J. L., Pirovano, G., Pouliot, G., Jose, R. S., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, J., Wolke, R., Yahya, K., Zabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., and Galmarini, S.: Evaluation of operational on-line-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2. Part I: Ozone, Atmos. Environ., 115, 404–420, doi:10.1016/j.atmosenv.2014.09.042, 2015.

Karl, T., Graus, M., Striednig, M., Lamprecht, C., Hammerle, A., Wohlfahrt, G., Held, A., Heyden, L. V. D., Deventer, M. J., Krismer, A., Haun, C., Feichter, R., and Lee, J.: Urban eddy covariance measurements reveal significant missing NOx emissions in Central Europe, Scientific Reports, pp. 1–9, doi:10.1038/s41598-017-02699-9, http://eprints.whiterose.ac.uk/117076/, 2017.

Kaspar, F., Müller-westermeier, G., Penda, E., Mächel, H., Zimmermann, K., Kaiser-weiss, A., and Deutschländer, T.: Monitoring of climate change in Germany – data, products and services of Germany's National Climate Data Centre, Advances in Science and Research, 10, 99–106, doi:10.5194/asr-10-99-2013, 2013.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier Van Der Gon, H. A. C.: TNO-MACC_ II emission inventory; a multi-year (2003-2009) consistent high-resolution European emission inventory for air quality modelling, Atmos Chem Phys, 14, 10 963–10 976, doi:10.5194/acp-14-10963-2014, 2014.

Kuik, F., Lauer, A., Churkina, G., Denier Van Der Gon, H. A. C., 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, doi:10.5194/gmd-9-4339-2016, https://www.geosci-model-dev.net/9/4339/2016/, 2016.

Lee, J. D., Helfter, C., Purvis, R. M., Beevers, S. D., Carslaw, D. C., Lewis, A. C., Møller, S. J., Tremper, A., Vaughan, A., and Nemitz, E. G.: Measurement of NOx Fluxes from a Tall Tower in Central London, UK and Comparison with Emissions Inventories, Environmental Science & Technology, 49, 1025–1034, doi:10.1021/es5049072, pMID: 25494849, 2015.

Minkos, A., Dauert, U., Feigenspan, S., and Kessinger, S.: Luftqualität 2016, https://www.umweltbundesamt.de/publikationen/luftqualitaet-2016, 2017.

Mues, A., Kuenen, J., Hendriks, C., Manders, A., Segers, A., Scholz, Y., Hueglin, C., Builtjes, P., and Schaap, M.: Sensitivity of air pollution simulations with LOTOS-EUROS to the temporal distribution of anthropogenic emissions, Atmospheric Chemistry and Physics, 14, 939–955, doi:10.5194/acp-14-939-2014, https://www.atmos-chem-phys.net/14/939/2014/, 2014.

Nakanishi, M. and Niino, H.: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog, Boundary-Layer Meteorology, 119, 397–407, 2006.

Pernigotti, D., Gerboles, M., Belis, C., and Thunis, P.: Model quality objectives based on measurement uncertainty. Part II: NO 2 and PM 10, Atmospheric environment, 79, 869–878, doi:10.1016/j.atmosenv.2013.07.045, 2013.

Pleim, J. E.: A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing, Journal of Applied Meteorology and Climatology, 46, 1383–1395, doi:10.1175/JAM2539.1, https://doi.org/10.1175/JAM2539.1, 2007.

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

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X.-y., Wang, W., and Powers, J.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note/TN-475+STR, doi:10.5065/D68S4MVH, 2008.

Solazzo, E. and Galmarini, S.: Error apportionment for atmospheric chemistry-transport models–a new approach to model evaluation, Atmospheric Chemistry and Physics, 16, 6263–6283, doi:10.5194/acp-16-6263-2016, 2016.

Solazzo, E., Bianconi, R., Hogrefe, C., Curci, G., Tuccella, P., Alyuz, U., Balzarini, A., Baró, R., Bellasio, R., Bieser, J., and Others: Evaluation and error apportionment of an ensemble of atmospheric chemistry transport modeling systems: multivariable temporal and spatial breakdown, Atmospheric Chemistry and Physics, 17, 3001–3054, doi:10.5194/acp-17-3001-2017, 2017.

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

Terrenoire, E., Bessagnet, B., Rouïl, L., Tognet, F., Pirovano, G., Létinois, L., Beauchamp, M., Colette, A., Thunis, P., Amann, M., and Menut, L.: High-resolution air quality simulation over Europe with the chemistry transport model CHIMERE, Geoscientific Model Development, 8, 21–42, doi:10.5194/gmd-8-21-2015, https://www.geosci-model-dev.net/8/21/2015/, 2015.

Thunis, P., Pederzoli, A., and Pernigotti, D.: Performance criteria to evaluate air quality modeling applications, Atmospheric Environment, 59, 476–482, doi:10.1016/j.atmosenv.2012.05.043, 2012.

Tie, X., Brasseur, G., and Ying, Z.: Impact of model resolution on chemical ozone formation in Mexico City: application of the WRF-Chem model, Atmos Chem Phys, 10, 8983–8995, doi:10.5194/acp-10-8983-2010, 2010.

Vaughan, A. R., Lee, J. D., Misztal, P. K., Metzger, S., Shaw, M. D., Lewis, A. C., Purvis, R. M., Carslaw, D. C., Goldstein, A. H., Hewitt, C. N., and Others: Spatially resolved flux measurements of NOx from London suggest significantly higher emissions than predicted by inventories, Faraday discussions, 189, 455–472, doi:10.1039/c5fd00170f, 2016.

Villena, G., Bejan, I., Kurtenbach, R., Wiesen, P., and Kleffmann, J.: Interferences of commercial NO2 instruments in the urban atmosphere and in a smog chamber, Atmospheric Measurement Techniques, 5, 149, doi:10.5194/amt-5-149-2012, 2012.

Von Schneidemesser, E., Kuik, F., Mar, K. A., and Butler, T.: Potential reductions in ambient NO2 concentrations from meeting diesel vehicle emissions standards, Environmental Research Letters, doi:10.1088/1748-9326/aa8c84, 2017.

WHO: Review of evidence on health aspects of air pollution – REVIHAAP, http://www.euro.who.int/__data/assets/pdf_file/0020/182432/e96762-final.pdf, 2013.

Zurbenko, I.: The spectral analysis of time series, Elsevier North-Holland, Inc., doi:10.2307/2348176, 1986.

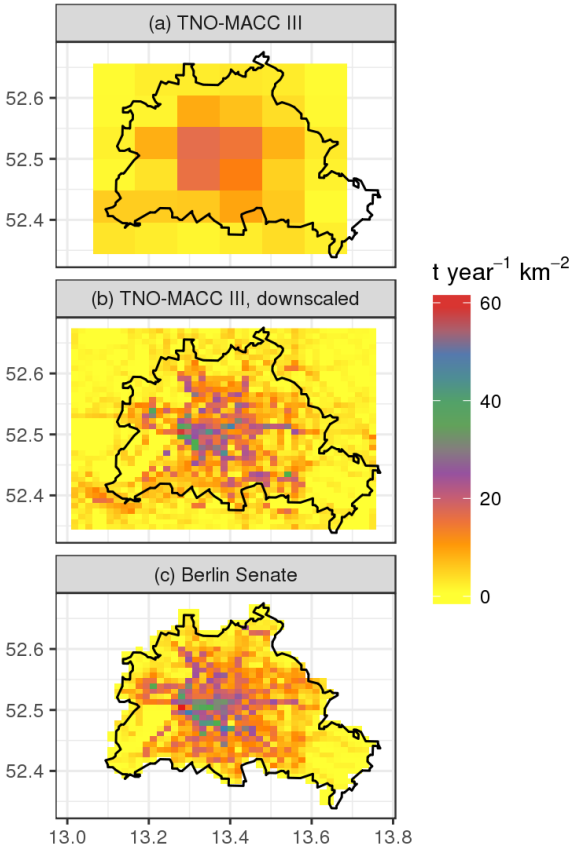

Figure 1. Total NO_x emissions from traffic in Berlin; (a) from the TNO-MACC III inventory, 2011; (b) from the TNO-MACC III inventory, downscaled to a horizontal resolution of ca. 1 km, 2011; (c) from the Berlin Senate Department for the Environment, 2009.

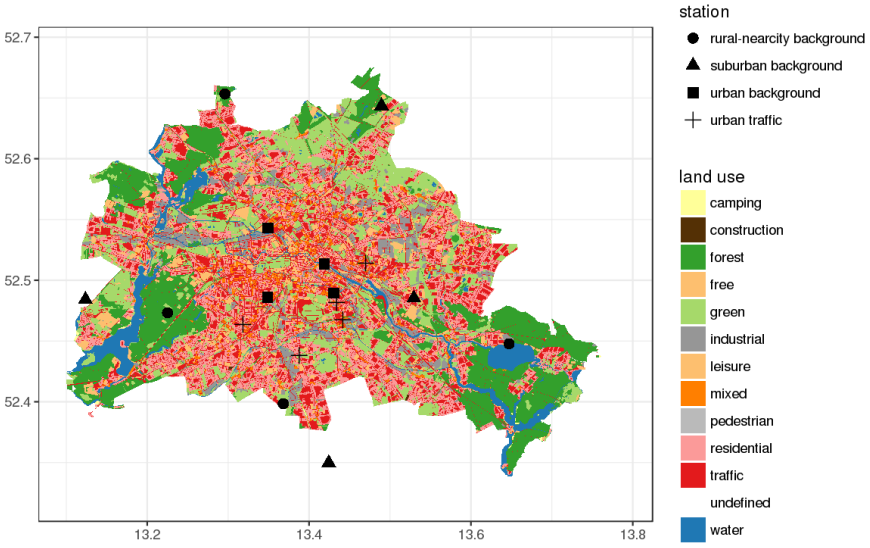

Figure 2. Locations of measurement stations in and close to Berlin, including their AirBase station area classification and type and the land use classes in Berlin according to Berlin Senate Department for Urban Development and Housing (the 2015).

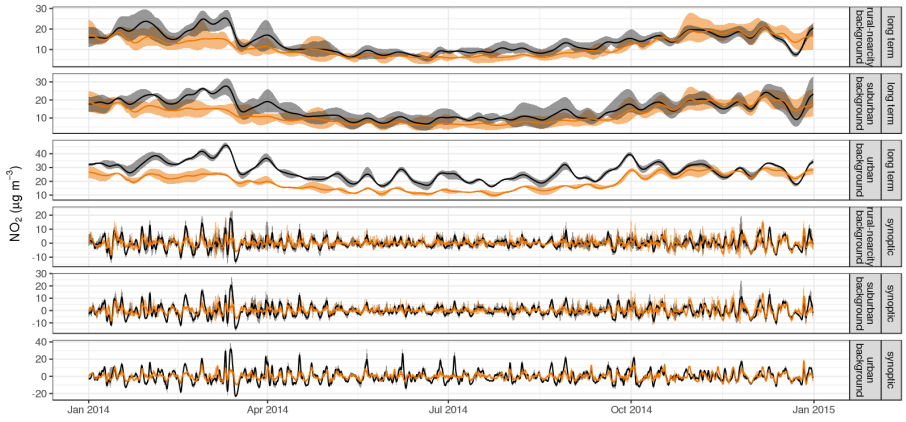

Figure 3. Long term and synoptic components of modeled (orange) and observed (black) time series, averaged over all stations of each station class. The shaded areas show the variability (25th and 75th percentiles) between the different stations within each class. Note the variable y-axis.

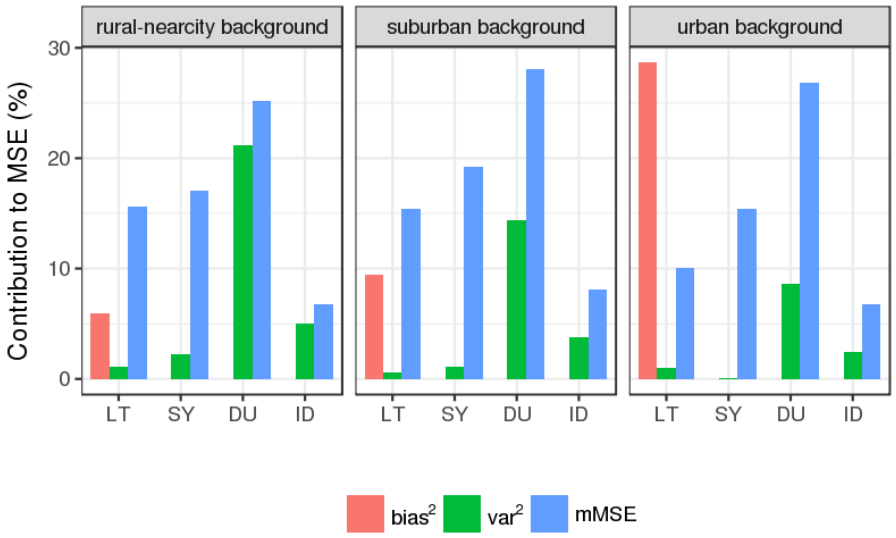

Figure 4. Contribution of different types of error to the mean square error of the model, per station class. The mean square error is divided into squared bias ($bias^2$), variance error (var^2) and minimum mean square error (mMSE) of the long term (LT), synoptic (SY), diurnal (DU) and intra-diurnal (ID) components (see Sect. 4.1 for further details).

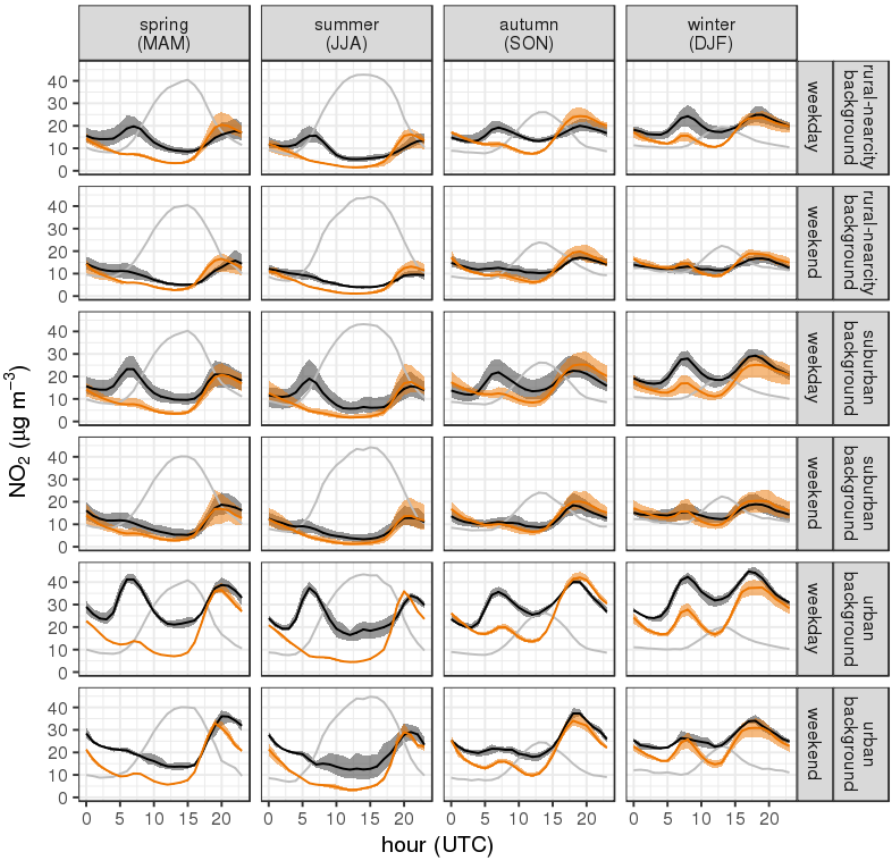

Figure 5. Mean diurnal cycles of modelled (orange) and observed (black) NO_2 concentrations, by station class and weekday/weekend. Shaded areas show the variability between the different stations' mean diurnal cycles (25th and 75th percentiles). Grey lines show the mean modeled planetary boundary layer heights at the respective grid points (scaled, but the relative changes between different hours and seasons are maintained).

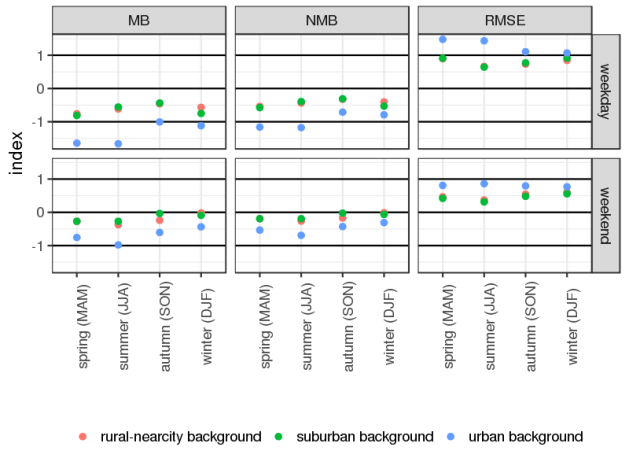

Figure 6. Skill of WRF-Chem in simulating daytime (6-17 UTC) observed NO_2 concentrations. The index represents the the model quality objective for the root mean square error (Sect. 4.1) and the performance criteria for mean and normalized mean bias (described in the Supplementary material), for weekend and weekday days and each month/season.

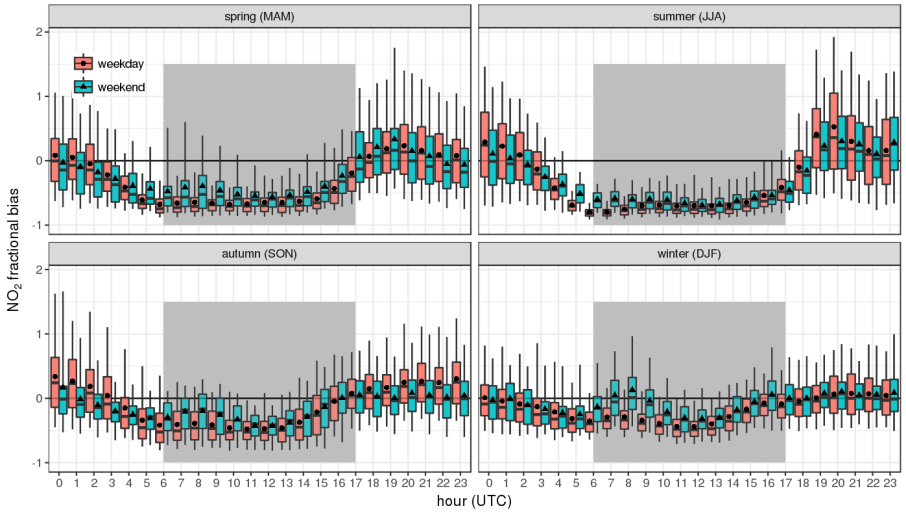

Figure 7. Relative bias in modeled NO_2 concentrations at urban background sites in Berlin, averaged over each season, hour and weekend/weekday. The boxplot show median (line), 25th and 75th (box), 5th and 95th (whiskers) percentiles of the hourly bias. Points show the mean. The grey shaded area shows the time period considered for quantifying the underestimation of daytime traffic emissions.

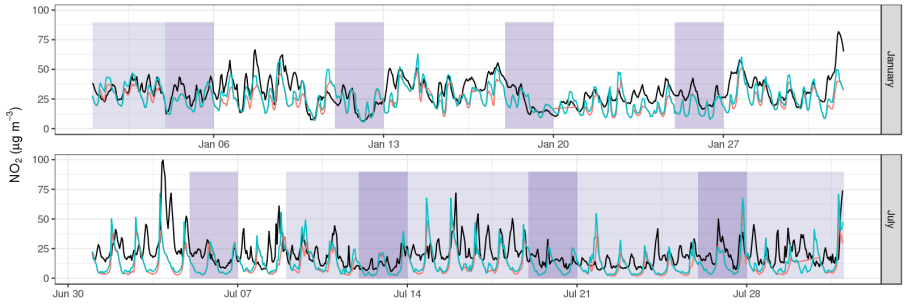

Figure 8. Time series of hourly observed (black line) and modelled NO_2, comparing the base simulation (red) with the sensitivity simulations (blue) using increased traffic emissions by a factor of 3 between 6:00 UTC and 17:00 UTC on weekdays. The time series are averaged over all 4 urban background stations. Weekends are highlighted in dark purple, and holidays are highlighted in light purple.

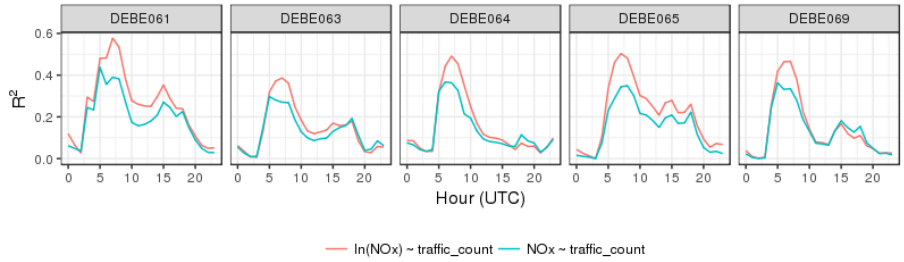

Figure 9. Comparison of R^2 for linear and exponential fit of roadside NO_x concentrations with traffic counts.

Table 1. Model configuration and input data.

Process	Option/dataset	Remarks
Land surface model	Noah LSM	CORINE land use data
Urban processes	single layer UCM	3 categories: roofs, walls, trees
Boundary layer	MYNN	
Cumulus convection	Grell-Freitas	switched on for both domains
Cloud microphysics	Morrison double-moment	
Radiation (sw+lw)	RRTMG	
Aerosols	MADE/SORGAM	chem_opt=106
Chemistry	RADM2	with KPP
Photolysis	Madronich F-TUV	
Anthropogenic emissions	TNO-MACC III	see Sect. 2.4 for details
Biogenic emissions	online	MEGAN
Dust and sea salt emissions	online	dust_opt=3, seas_opt=2
Meteorological boundary conditions	ERA-Interim	sst_update=1
Chemical boundary conditions	MOZART4-GEOS5	

Table 2. Model simulations presented in this paper.

Simulation	Chemistry	Period	Emissions
2014_ref	RADM2	1.1.2014 - 31.12.2014	TNO-MACC III
2014_moz	MOZART	July 2014	TNO-MACC III
2014_emis	RADM2	January 2014	TNO-MACC III
		July 2014	Traffic NO_x increased (Sect. 6.1)

Table 3. Modeled meteorology compared to observations, annual and seasonal performance indicators. Mean bias (MB) and root mean square error (RMSE) are indicated in ~~µg~~ K (temperature) and m $^{-3}$s^{-1} (wind speed), the normalized mean bias (NMB) and correlation coefficient (R) are unitless. Data are aggregated as follows: MAM - March, April, May, JJA - June, July, August, SON - September, October, November, DJF - December, January, February.

		Temperature				Wind speed			
		MB	NMB	RMSE	R	MB	NMB	RMSE	R
Lindenberg	2014	-0.66	-0.06	2.17	0.96	0.81	0.25	1.63	0.69
	spring (MAM)	-0.48	-0.04	2.26	0.91	0.87	0.27	1.8	0.68
	summer (JJA)	-0.89	-0.05	2.4	0.89	0.87	0.31	1.68	0.61
	autumn (SON)	-0.55	-0.05	1.99	0.95	0.92	0.32	1.58	0.63
	winter (DJF)	-0.7	-0.3	2	0.93	0.57	0.14	1.43	0.76
Potsdam	2014	-0.71	-0.06	2.25	0.96	-0.47	-0.12	1.4	0.69
	spring (MAM)	-0.49	-0.04	2.16	0.93	-0.34	-0.08	1.42	0.73
	summer (JJA)	-0.45	-0.02	2.1	0.91	-0.14	-0.04	1.41	0.55
	autumn (SON)	-0.76	-0.07	2.34	0.94	-0.46	-0.12	1.27	0.62
	winter (DJF)	-1.15	-0.44	2.42	0.9	-0.99	-0.2	1.47	0.79
Schoenefeld	2014	-0.61	-0.06	2.16	0.96	0.02	0	1.3	0.78
	spring (MAM)	-0.12	-0.01	1.97	0.94	0.07	0.02	1.34	0.81
	summer (JJA)	-0.63	-0.03	2.1	0.92	0.16	0.05	1.39	0.66
	autumn (SON)	-0.73	-0.06	2.27	0.94	0.12	0.04	1.21	0.7
	winter (DJF)	-0.98	-0.39	2.3	0.91	-0.31	-0.07	1.25	0.85
Tegel	2014	-1.25	-0.11	2.48	0.96	0.4	0.12	1.33	0.75
	spring (MAM)	-0.83	-0.07	2.26	0.93	0.58	0.18	1.43	0.77
	summer (JJA)	-1.02	-0.05	2.23	0.92	0.58	0.2	1.44	0.66
	autumn (SON)	-1.44	-0.12	2.65	0.94	0.47	0.16	1.24	0.69
	winter (DJF)	-1.72	-0.54	2.76	0.9	-0.08	-0.02	1.17	0.84
Tempelhof	2014	-1.21	-0.11	2.51	0.96	0.45	0.13	1.32	0.74
	spring (MAM)	-0.67	-0.06	2.22	0.93	0.47	0.13	1.38	0.77
	summer (JJA)	-1.17	-0.06	2.35	0.91	0.5	0.16	1.51	0.63
	autumn (SON)	-1.38	-0.11	2.65	0.93	0.43	0.14	1.25	0.7
	winter (DJF)	-1.63	-0.54	2.76	0.9	0.38	0.1	1.12	0.81

Table 4. Modeled chemistry, seasonal performance indicators (averaged for each station class, each class includes 4 stations) and the model quality objective for NO_2. Mean bias (MB) and root mean square error (RMSE) are indicated in µg m^{-3}, the normalized mean bias (NMB) and correlation coefficient (R) are unitless. Data are aggregated as follows: MAM - March, April, May, JJA - June, July, August, SON - September, October, November, DJF - December, January, February.

		NO_2				NO_x				O_3				NO_2
		MB	NMB	RMSE	R	MB	NMB	RMSE	R	MB	NMB	RMSE	R	MQO
rural-	2014	-2.12	-0.16	10.2	0.5	-3.77	-0.23	15	0.48	5.02	0.11	22.49	0.7	0.78
nearcity	autumn (SON)	-0.97	-0.06	9.97	0.48	-3.9	-0.19	16.2	0.45	11.96	0.41	22.71	0.66	0.76
backgr.	spring (MAM)	-2.91	-0.23	11.69	0.42	-4.26	-0.29	16.39	0.37	3.88	0.07	23.42	0.62	0.89
	summer (JJA)	-2.38	-0.26	8.23	0.37	-2.88	-0.28	9.57	0.32	1.41	0.02	25.49	0.61	0.64
	winter (DJF)	-2.2	-0.12	10.66	0.47	-4.08	-0.18	16.83	0.46	2.85	0.09	17.18	0.55	0.79
suburban	2014	-2.8	-0.19	10.67	0.55	-7.2	-0.35	20.13	0.48	4.88	0.11	22.45	0.7	0.8
backgr.	autumn (SON)	-0.76	-0.05	10.32	0.52	-7.92	-0.32	23.12	0.44	12.22	0.42	22.39	0.67	0.78
	spring (MAM)	-4.41	-0.31	12.2	0.49	-8.25	-0.44	21.71	0.39	4.15	0.07	24.06	0.61	0.92
	summer (JJA)	-2.88	-0.29	9.01	0.44	-5.12	-0.4	13.14	0.34	1.16	0.02	25.49	0.64	0.7
	winter (DJF)	-3.14	-0.16	10.96	0.53	-7.57	-0.28	21.24	0.49	2.02	0.06	16.53	0.57	0.8
urban	2014	-7.83	-0.29	16.69	0.51	-15.84	-0.4	35.57	0.47	3.25	0.08	21.01	0.73	1.13
backgr.	autumn (SON)	-4.89	-0.17	13.9	0.55	-16.9	-0.36	37.3	0.48	9.09	0.37	19.69	0.71	0.95
	spring (MAM)	-10.23	-0.38	19.71	0.51	-17.09	-0.47	40.68	0.4	3.07	0.06	22.62	0.62	1.32
	summer (JJA)	-9.26	-0.41	18.16	0.36	-13.3	-0.47	28.92	0.24	-1.94	-0.03	24.85	0.6	1.28
	winter (DJF)	-6.84	-0.22	14.05	0.53	-16.16	-0.34	34.41	0.5	2.93	0.12	15.35	0.58	0.93

Table 5. Statistics of modelled NO_2 and NO_x concentrations for January and July, for the base simulation and for the sensitivity simulation with increased traffic emissions, at the urban background stations in Berlin. Mean bias (MB) and root mean square error (RMSE) are indicated in µg m^{-3}, the normalized mean bias (NMB) and correlation coefficient (R) are unitless.

		NO_2				NO_x			
		MB	NMB	RMSE	R	MB	NMB	RMSE	R
Amrumer Str.									
Jan	2014_ref	-6.77	-0.21	11.93	0.65	-14.63	-0.31	28.85	0.65
	2014_emis	-4.16	-0.13	11.09	0.68	-8.01	-0.17	26.18	0.64
July	2014_ref	-10.12	-0.47	17.88	0.34	-12.25	-0.49	22.88	0.27
	2014_emis	-8.9	-0.42	18	0.33	-10.51	-0.42	23.21	0.25
Belziger Str.									
Jan	2014_ref	-10.64	-0.32	15.39	0.51	-20.88	-0.41	34.23	0.51
	2014_emis	-7.97	-0.24	14.04	0.53	-14.57	-0.29	31.6	0.51
July	2014_ref	-7.32	-0.37	15.32	0.31	-8.38	-0.37	17.81	0.23
	2014_emis	-4.02	-0.2	16.73	0.22	-3.67	-0.16	20.43	0.12
Nansenstr.									
Jan	2014_ref	-6.09	-0.21	11.12	0.61	-12.81	-0.3	25.05	0.56
	2014_emis	-3.72	-0.13	10.28	0.64	-7.44	-0.18	23.28	0.55
July	2014_ref	-8.73	-0.43	15.33	0.42	-10.89	-0.45	19.18	0.35
	2014_emis	-6.88	-0.34	15.43	0.36	-8.25	-0.34	19.37	0.28
Brückenstr.									
Jan	2014_ref	-7.1	-0.23	12.51	0.57	-17.27	-0.36	36.4	0.51
	2014_emis	-4.5	-0.15	11.13	0.63	-11.16	-0.24	32.58	0.54
July	2014_ref	-9.92	-0.46	17.24	0.25	-12.85	-0.49	22.7	0.17
	2014_emis	-8.05	-0.37	16.87	0.26	-10.16	-0.39	22.13	0.18

S1 Changes to the model code and impact on results

S1.1 Code changes

The WRF-Chem code file dry_dep_driver.F (v.3.8.1) was changed in order to also allow for increased nighttime mixing in grid cells with high emissions in case an urban physics scheme is used (starting from line 685):

```
      if (p_e_co >= param_first_scalar )then
! if (sf_urban_physics .eq. 0 ) then
            if (emis_ant(i,kts,j,p_e_co) .gt. 0) then
             ekmfull(kts:kts+10) = max(ekmfull(kts:kts+10),1.)
            endif
            if (emis_ant(i,kts,j,p_e_co) .gt. 200) then
             ekmfull(kts:kte/2) = max(ekmfull(kts:kte/2),2.)
            endif
            if (p_e_pm25i > param_first_scalar )then
             if (emis_ant(i,kts,j,p_e_pm25i)+ emis_ant(i,kts,j,p_e_pm25j)
              .GT. 8.19e-4*200) then
              ekmfull(kts:kte/2) = max(ekmfull(kts:kte/2),2.)
             endif
            endif
            if (p_e_pm_25 > param_first_scalar )then
             if (emis_ant(i,kts,j,p_e_pm_25) .GT. 8.19e-4*200) then
              ekmfull(kts:kte/2) = max(ekmfull(kts:kte/2),2.)
             endif
! endif
            endif
```

S1.2 Impact of code changes

Two test simulations illustrate the impact of some of the modifications made to this model setup with respect to Kuik et al., 2016. The differences between the two model simulations shown in this section are the application of the above-described modification or not, and the use of default (TNO) diurnal cycles of traffic emissions and a diurnal cycle calculated based on traffic counts in Berlin. The latter is expected to mainly impact the results during daytime for urban background stations and, to a smaller extent, suburban background stations. The figures show results for simulated and observed NO_2 concentrations.

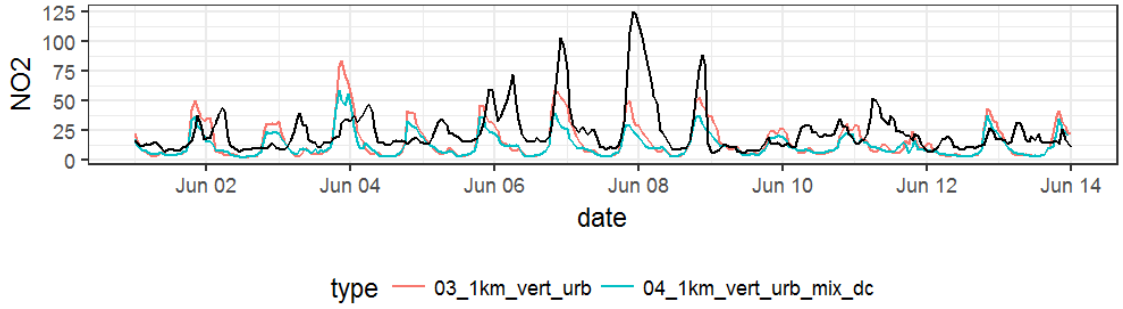

Figure S1. Comparison of modelled NO_2 concentrations of simulations without modified mixing routine and default diurnal cycle of traffic emissions (03_1km_vert_urb) and with modified mixing routine and a diurnal cycle of traffic emissions calculated based on traffic counts in Berlin (04_km_vert_urb_mix_dc). The black line shows observations. Results are averaged over all urban background stations used in this study.

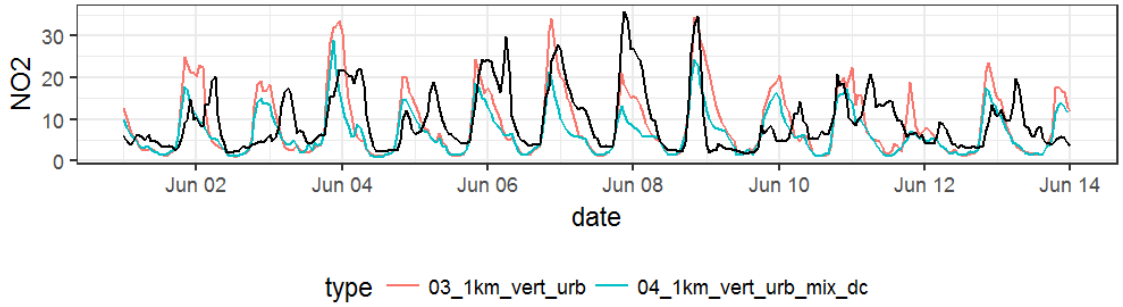

Figure S2. As Fig. S1, but for suburban background stations.

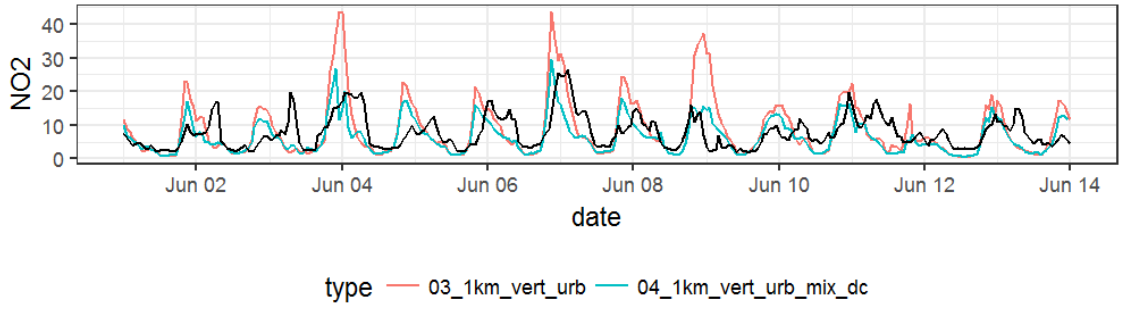

Figure S3. As Fig. S1, but for rural background stations.

S2 Emission processing

S2.1 Downscaling

We used TNO-MACC III emission data and in cooperation with TNO downscaled the data from a horizontal resolution of ca. 7kmx7km to a ca. 1kmx1km. We based the downscaling on proxy data, including population density (Environment Database of the Berlin Senate Department for the Environment, Transport and Climate Protection, Landscan 2010 data), traffic density

for the area of Berlin (Environment Database of the Berlin Senate Department for the Environment, Transport and Climate Protection) and the road network of Brandenburg (OpenStreetMap). Population data is used to downscale emissions from residential combustion (SNAP2) and product use (SNAP6), traffic data is used to downscale emissions from traffic (SNAP 71-75). The 1kmx1km emission grid is defined so that each coarse grid cell of 7kmx7km is divided into 7x7 parts. From each of the proxy datasets a factor is then calculated indicating the proportion of each proxy data type in one high resolution grid cell within one coarse grid cell. These factors are used in order to downscale the respective emissions in the respective area.

S2.2 Modification of airport emissions for Berlin

Airport emissions in Berlin, designated by point sources within non-road transport emissions in the TNO-MACC III inventory, are split into airport emissions into emissions on the ground and emissions from the LTO-cycle. We attribute 60% of the emissions to emissions on the ground, and the remaining emissions to emissions from the LTO cycle, where we distribute the emissions equally into all layers below 900m. The LTO-cycle includes emissions from takeoff, landing and aircraft cruise up to ca. 900m. Furthermore, the TNO-MACC III inventory still includes emissions from the Berlin-Tempelhof airport, which has been closed to air traffic in 2008. In addition, emissions from Tegel airport seemed unrealistically larger than emissions from Schönefeld airport. Thus we summed the emissions from all three major airports in the Berlin-Brandenburg region included in the TNO-MACC III inventory and re-distributed them onto the two airports active in 2014, based on activity data, attributing 75% of the emissions to Tegel and 25% to Schönefeld.

S3 Model evaluation

S3.1 Observations

Table S1. Coordinates, names, codes (airbase) and station types of measurement stations used in this paper. Different from what is indicated in the airbase metadata, the station DEBE066 has been defined as suburban background station in this paper, as its characteristics and location are more in line with suburban background conditions in Berlin than with urban background conditions.

Station code	Station name	Station type	Longitude	Latitude
DEBE027	Schichauweg	rural-nearcity background	13.368	52.398
DEBE032	Grunewald	rural-nearcity background	13.225	52.473
DEBE056	Mueggelseedamm	rural-nearcity background	13.647	52.448
DEBE062	Frohnau	rural-nearcity background	13.296	52.653
DEBB075	Gross Glienicke	suburban background	13.124	52.484
DEBB086	Blankenfelde-Mahlow	suburban background	13.424	52.35
DEBE051	Buch	suburban background	13.49	52.644
DEBE066	J. u. W. Brauer Platz	suburban background	13.53	52.485
DEBE010	Amrumer Str.	urban background	13.349	52.543
DEBE018	Belziger Str.	urban background	13.349	52.486
DEBE034	Nansenstr.	urban background	13.431	52.489
DEBE068	Brueckenstr.	urban background	13.419	52.514
DEBE061	Schildhornstr.	urban traffic	13.318	52.464
DEBE063	Silbersteinstr	urban traffic	13.442	52.468
DEBE064	Karl Marx Str.	urban traffic	13.434	52.482
DEBE065	Frankfurter Allee	urban traffic	13.47	52.514
DEBE069	Mariendorfer Damm	urban traffic	13.388	52.438

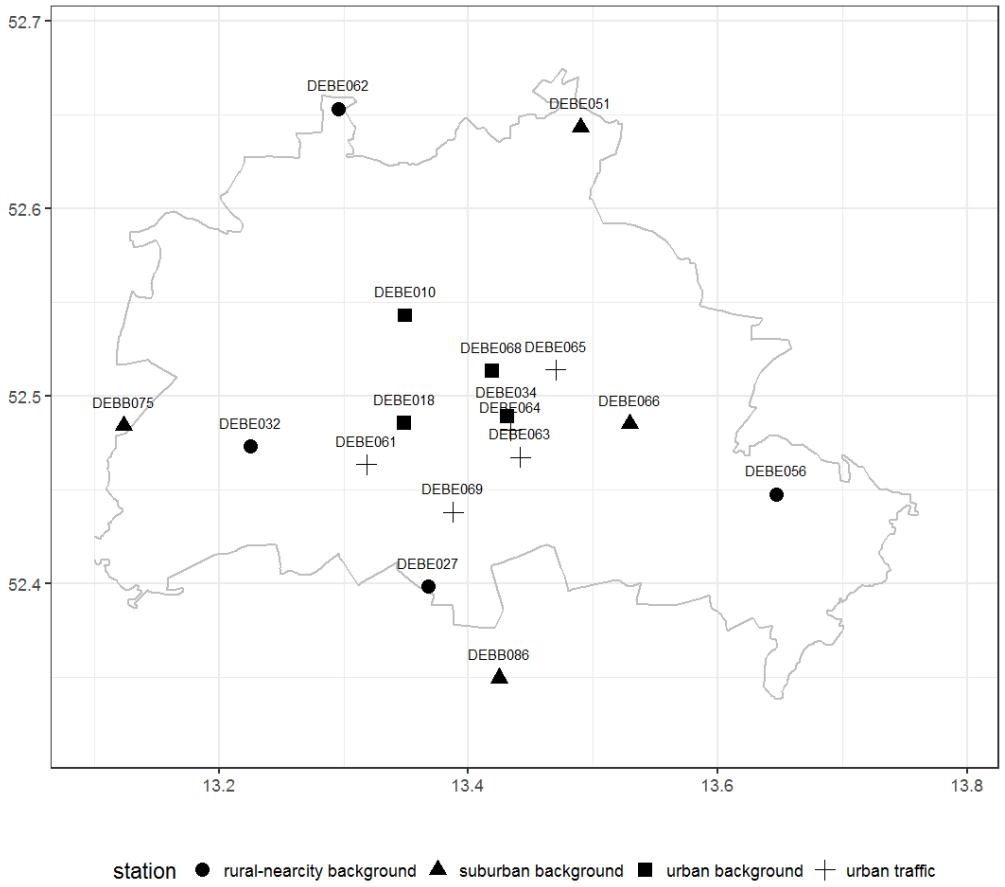

Figure S4. Locations of measurement stations used in this paper including their station type and airbase code.

S3.2 Statistical indicators

The statistical indicators used in this study include the mean bias (MB), normalized mean bias (NMB), root mean square error (RMSE) and Pearson correlation coefficient (R) and are defined as follows, with the model results M, observations O, number of model-observations pairs N and standard deviation σ:

$$MB = \sum_{i=1}^{N}(M_i - O_i) \quad (1)$$

$$NMB = \frac{\sum_{i=1}^{N}(M_i - O_i)}{\sum_{i=1}^{N} O_i} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)} \quad (3)$$

$$R = \frac{\frac{1}{N}\sum_{i=1}^{N}(M_i - \overline{M})(O_i - \overline{O})}{\sigma_M \sigma_O} \quad (4)$$

S3.3 Additional model performance indicators

In addition to the model quality objective (MQO), performance indicators for the mean bias and normalized mean bias are indicated in the manuscript and defined as follows, following Pernigotti et al. (2013):

$$|NMB| < \frac{2RMS_U}{\overline{O}} \quad (5)$$

$$|MB| < 2U(\overline{O}) \quad (6)$$

With the root mean square of the measurement uncertainty

$$RMS_U = \sqrt{\frac{1}{N}\sum_{i=1}^{N} U_{O_i}^2} \quad (7)$$

and $U(\overline{O})$ the uncertainty of the mean of the observed time series.

S4 Spectral decomposition

Observed and modelled time series are spectrally decomposed into a long term (LT), synoptic (SY), diurnal (DU) and intra-diurnal (ID) component, following Hogrefe et al. (2000), and Galmarini et al. (2013). A Kolmogorov-Zurbenko filter $kz_{m,k}$ was used with the time windows m and smoothing parameter k, time series x and time t:

$$ID(t) = x(t) - kz_{3,3}(x(t)) \quad (8)$$

$$DU(t) = kz_{3,3}(x(t)) - kz_{13,5}(x(t)) \quad (9)$$

$$SY(t) = kz_{13,5}(x(t)) - kz_{103,5}(x(t)) \quad (10)$$

$$LT(t) = kz_{103,5}(x(t)) \quad (11)$$

The decomposition was done in R, using the library *kza*.

S5 Supplementary figures

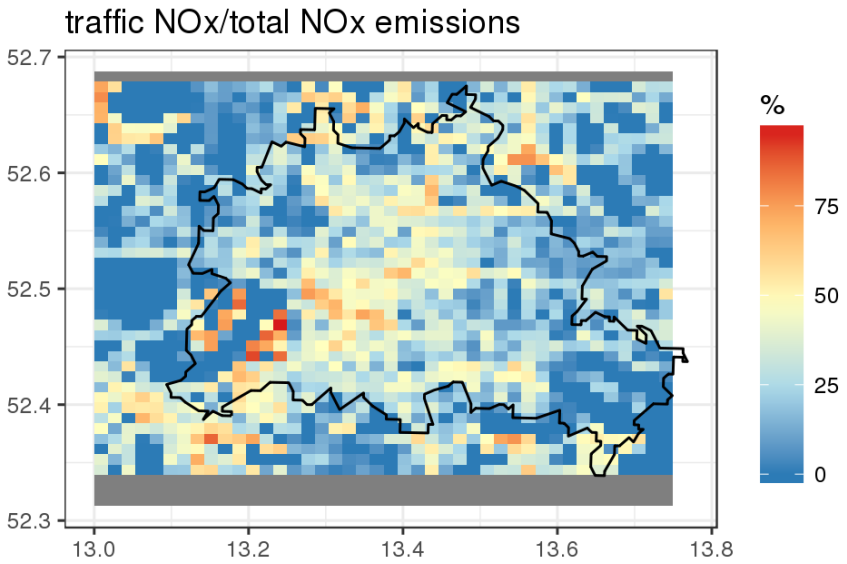

Figure S5. Contribution of traffic NOx emissions to total annual surface NOx emissions in the Berlin-Brandenburg area, based on the downscaled version of TNO-MACC III.

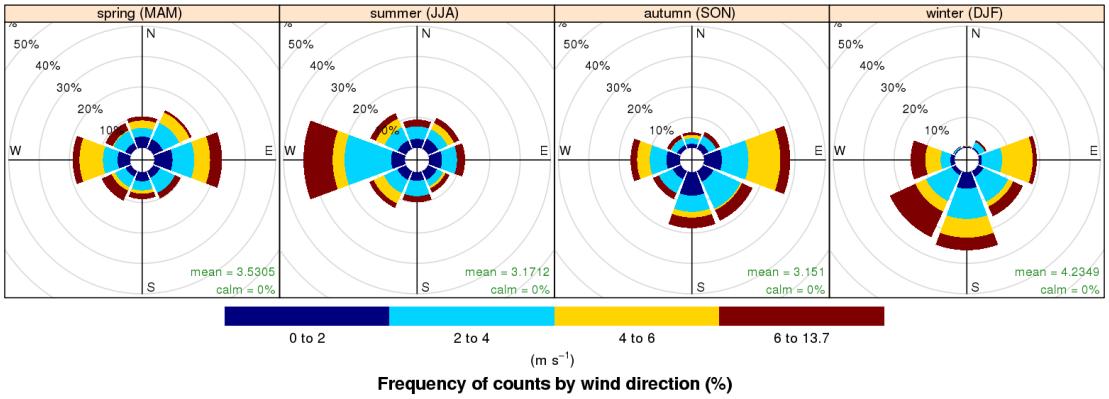

Figure S6. Wind rose showing the frequency distribution of wind speed and direction for the Berlin DWD stations, observations.

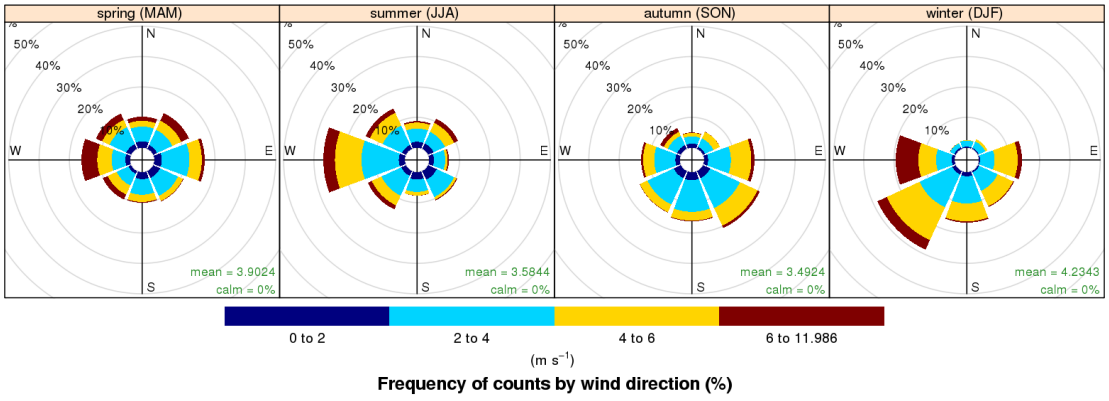

Figure S7. Wind rose showing the frequency distribution of wind speed and direction for the Berlin DWD stations, model results.

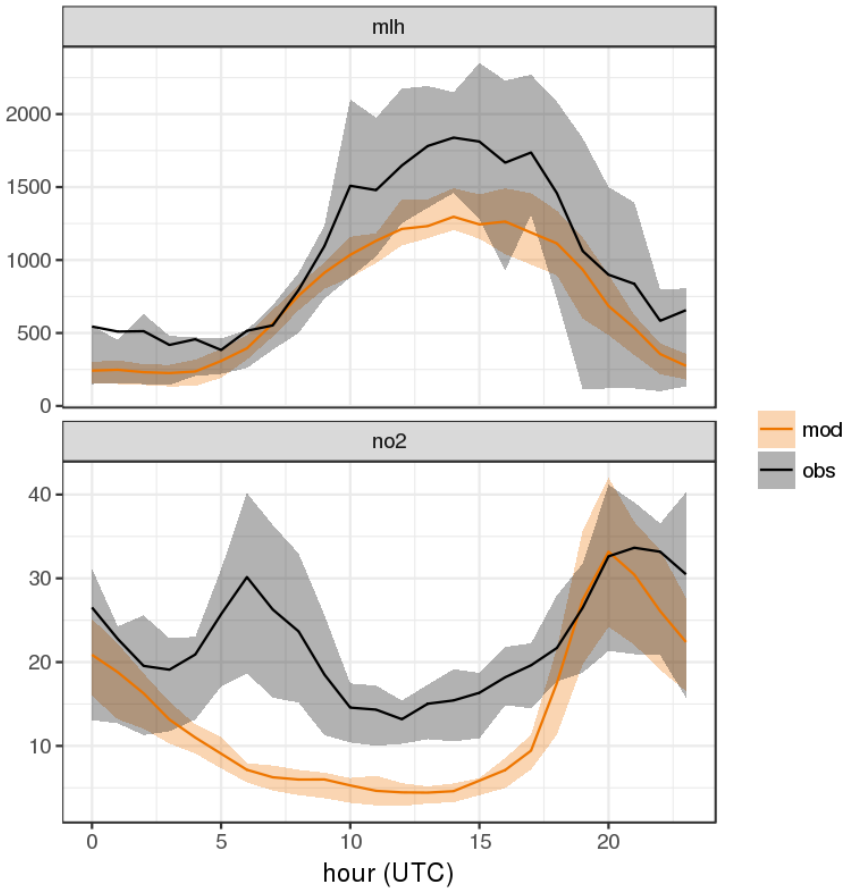

Figure S8. Mean diurnal cycles of observed (obs) and modelled (mod) mixing layer height (MLH) and NO_2 concentrations at Nansenstraße. All data are only averaged over times when MLH observed with a ceilometer is available. This includes between 24-57 hourly values between 20 June and 27 August 2014. The shaded areas show the 25th and 75th percentiles of the data. MLH is given in m, NO_2 concentrations are given in μg m^{-3}.

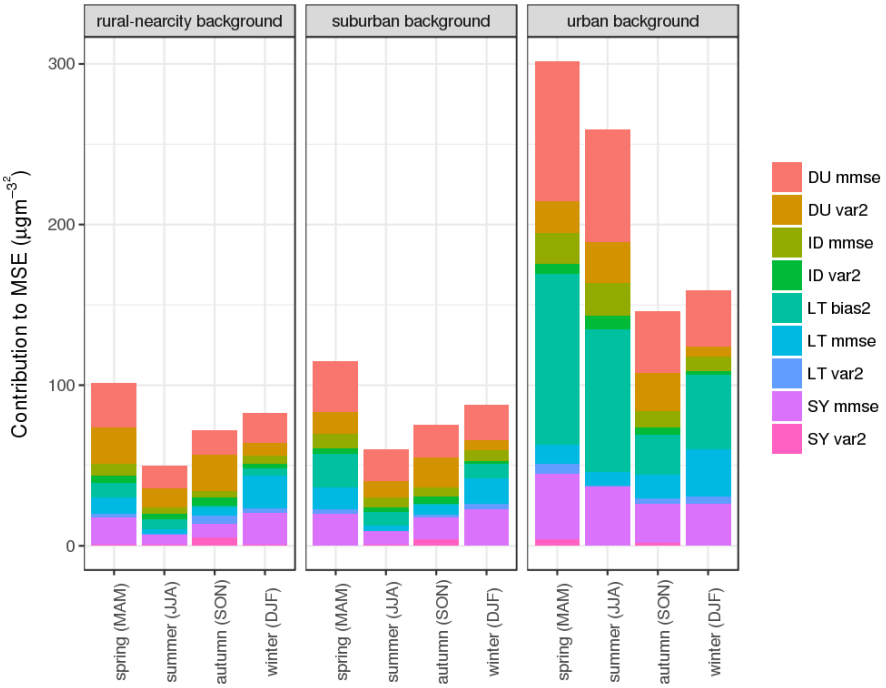

Figure S9. Contribution to mean square error of model results per season and station class.

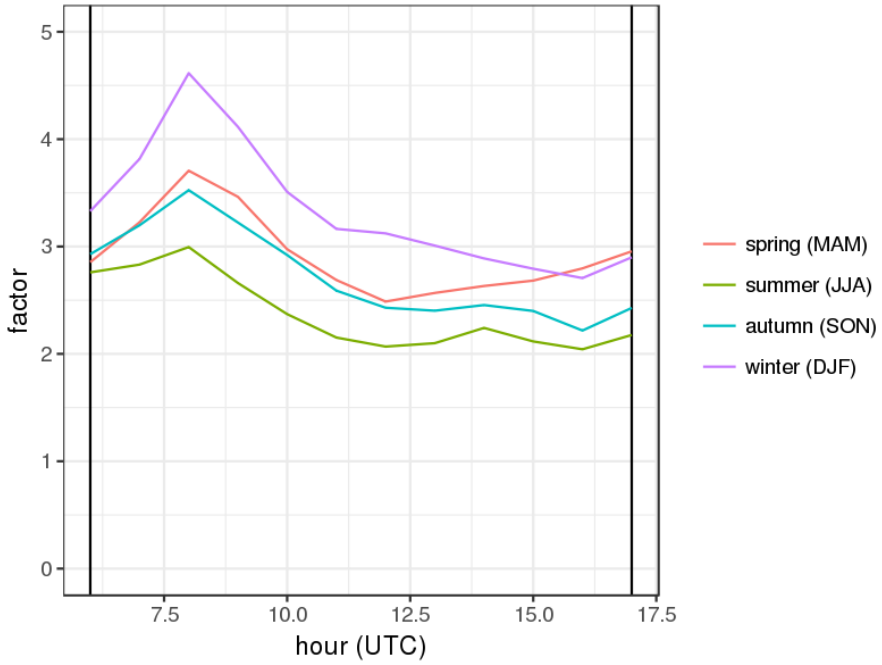

Figure S10. Time- and season-dependent NO_x-emission correction factor.