Anonymous Referee #2

We thank Referee #2 for his positive feedback and valuable comments. Please find our point by point responses below. Line numbers in our responses refer to the new manuscript without tracked changes.

The Stuttgart area is characterized by complex topography and known for frequent occurrence of high pollution concentrations. The paper describes a demonstration case study for this area with WRF-Chem at a horizontal resolution of 50 m for a polluted winter day. Overall, the paper addresses an important and interesting topic as air quality studies at such a high resolution are not frequently presented. The paper should definitely be published as an interesting case study with a high-resolution air quality model. However, it may be considered as a bit premature to regard the described setup as an 'air quality forecasting system (AQFS)'.

The results presented in our manuscript are the first steps towards an AQFS. We emphasized that this is still a prototype which, by including additional refinements/enhancements, has the potential to be applied as a forecast system in the future.

Specific comments

Abstract, line 26: 'Together with information about the vertical distribution of PM10 and NO2 from the model, AQFS will serve . . .' What is the difference between the described WRF-Chem simulation and AQFS? Are there any special tools developed for the AQFS? If so, they should also be described in more detail in the paper.

Our AQFS utilizes the WRF-Chem model including the improvements and enhancements described in sections 2.1-2.3.

The introduction is quite long considering the length of the paper and not all points mentioned in the introduction are relevant in the context of the paper (e.g. lines 29-36, remarks offline air quality models or on aerosol-radiation interactions).

We consider it as essential to provide the reader with information about pollutions limits proposed by the World Health Organization (WHO) as this is the basis for the current discussions of protection scenarios. From our point of view, it is also important to mention that there is a longer history of offline models which are presently still widely used around the globe and that the direct interaction between radiation and aerosols is essential. Following your suggestion, we slightly shortened the introduction and the 2nd and 3rd paragraph now read:

"Due to a strong increase of road traffic in major European cities (Thunis et al., 2017), pollution limits are often violated in larger cities. E.g. for particulate matter with particle diameters less than 10 µm (PM10), the critical value is an annual mean concentration of 20 µg m^{-3} or a daily mean value of 50 µg m^{-3} (WHO, 2005). For Nitrogen dioxide (NO_2) the critical values are 200 µg m^{-3} and 40 µg m^{-3} as daily and annual mean values, respectively.

The violation of these pollution limits can lead to health and environmental problems and is currently part of several litigations e.g. at the German Federal Administrative Court dealing with possible driving bans for non low-emission vehicles. The basis for these litigations are mostly few, unevenly distributed local observations. In combination with special meteorological conditions like winter time thermal inversion layers it can be misleading to conclude about the overall air quality in the city only from single observations. According to e.g. the German Federal Immission Control Ordinance it is sufficient that traffic related measurements are representative for a section of 100 m, but this is not representative for the commercial and office districts in the cities that are suffering from traffic control in case of fine dust alerts and residential areas. Namely in residential areas health protection action plans require representative air quality measures."

Line 86: In the meantime, PALM offers also several option to use e.g. COSMO of WRF as drivers (Kadasch et al. and Lin et al. in https://gmd.copernicus.org/articles/special_issue999.html as well as the implementation of WRF interface, which is already included in the PALM system). Besides of this, this remark would better match in the context of lines 96-100.

At the time when the simulation was conducted, these driver options were not available to the community. As our focus is on the WRF model system, we decided to add the work of Lin et al. (2020) to the outlook on page 14, line 488:

"Recently, Lin et al. (2020) developed an interface to use output from high-resolution WRF simulations to force PALM 6.0 in an offline mode which could be another tool in the future to study microscale structures in urban areas."

Line 158: The statement about RADM2 is somewhat odd (63 chemical species including photolysis and more than reactions). Besides of this, the number of 63 includes also water vapor and a passive compound.

Thank you for pointing this out. In this sentence, the number of chemical reactions was missing. To avoid confusion on the exact number of species used in RADM in WRF-chem, we decided to change the sentence on page 5, line 158 to:

"RADM2 features more than 60 chemical species and more than 135 chemical reactions including photolysis"

The authors claim that the model can be applied 'in a forecast and warning mode' (line 173). On the other hand, it is obvious from the domain size and the high resolution the simulations must be quite demanding, which is also mentioned in line 174. How high was the computational effort for the described study (How long did the simulation take on how many nodes)? Besides of this, tools should be supplied to stakeholders for an AQFS.

Thank you for your comment. As also reviewer #1 asked for more information, we added the following paragraphs to the "experimental set-up" section on page 5, line 173:

"Currently, air pollution modeling with WRF-Chem is a computationally expensive task. Depending on the number of output variables and frequency (5 min in our study), a 24 h simulation currently takes around 36 h wall clock time. For future experiments it is worth to try the I/O quilting option in combination with PNetCDF which should considerably reduce the time spent on I/O.

While the WRF model itself is ready for hybrid parallelism (MPI + OpenMP), the WRF-Chem model can only be used with MPI. If WRF-Chem could be enhanced for additional OpenMP capabilities, this would lead to an increase in computation speed almost linear with the number of OpenMP threads."

We are not sure what is meant by "tools" here? If the AQFS is further refined in the future, this will require a short tutorial for the person or institution operating this system including all pre- and postprocessing steps.

Lines 208-209: The differences should be commented in more quantitative way (E.g. what is the difference between the spatial integral for a) and b)?).

Thank you for your comment. We decided to add the average emission rates from both data sources. The average emission for the CAMS-REG-AP data set is 2 mol km^{-2} h^{-1} and the average emission for the BW-EMISS data set is 7 mol km^{-2} h^{-1}.
The following sentence was added on page 7, line 226:

"The average emissions for this particular time step are 2 mol km^{-2} h^{-1} for the CAMS-REG-AP data set and 7 mol km^{-2} h^{-1} for the BW-EMISS data set."

Section 2.4: Lines 225-232 are not related to observation and should be moved to somewhere else. Chemistry observations from Fig. 5 should also be included here instead.

Thank you for your suggestion. This paragraph has been moved to section 4.2 and has been slightly modified. It now reads (page 10, line 360):

"As the incorporated emissions are from 2014 and are based on annual values, it cannot be expected that the model exactly matches the observed concentrations. For instance, the actual traffic, the sequence of traffic lights and traffic congestions of this particular day cannot be realistically represented. In addition, all diagnosed or prognostic chemical quantities are only available on model levels (with the lowest model half level being at ~15 m above ground) but according to studies of (Glaser, 2003) and (Samad et al., 2020) the concentrations of PM_{10} and NO_2 are often constant up to 150—200 m AGL during daytime."

Following the suggestion of reviewer #1 we also included the simulated NO_2 concentrations in Fig. 5. Figure 5 is also referenced in section 4.2.1 on page 11, line 383.

Glaser, K., Vogt, U., Baumbach, G., Volz-Thomas, A., and Geiss, H. (2003), Vertical profiles of O3, NO2, NOx, VOC, and meteorological parameters during the Berlin Ozone Experiment (BERLIOZ) campaign, J. Geophys. Res., 108, 8253, doi:10.1029/2002JD002475, D4

A. Samad, U. Vogt, A. Panta, D. Uprety,
Vertical distribution of particulate matter, black carbon and ultra-fine particles in Stuttgart, Germany, Atmospheric Pollution Research, Volume 11, Issue 8, 2020, Pages 1441-1450, ISSN 1309-1042, https://doi.org/10.1016/j.apr.2020.05.017.

Lines 291-307: It is not clearly mentioned whether too persistent simulated clouds are really the reason to the too late drop in temperature.

We checked the model output with respect to hydrometeors again. The model simulates a very thin cloud layer at around 1000 m above ground level along the lower left and partly the lower right quadrant of the model domain. This thin cloud layer slowly moves to the SE and starts to diminish around 6 UTC.

We added two sentences on page 9, line 311 for clarification:

"A reason for this delayed temperature drop could be a simulated thin cloud layer around 1000 m AGL which is present in the lower left and partly the lower right quadrant of the model domain. This cloud layer slowly moves in a southeasterly direction and starts to dissolve around 06 UTC."

Line 345: Some comments on the influence of the quite thick lowest layer on the results should be given and (in spite of the objections and restrictions) a small comparison with observed values.

With respect to the observations, please see our comment on section 2.4 above. Unfortunately we do not have 3-dimensional measurements available, but according to a recent study of Samad et al. (2020) and an older study of Glaser et al. (2003), at least the concentrations of PM_{10} and NO_2 are often constant up to an altitude of a few 100 m above ground. We added a paragraph on page 10, line 360:

"As the incorporated emissions are from 2014 and are based on annual values, it cannot be expected that the model exactly matches the observed concentrations. For instance, the actual traffic, the sequence of traffic lights and traffic congestions of this particular day cannot be realistically represented. In addition, all diagnosed or prognostic chemical quantities are only available on model levels (with the lowest model half level being at ~15 m above ground). According to studies of Glaser et al. (2003) and Samad et al. (2020) the concentrations of PM_{10} and NO_2 are often constant up to 150—200 m AGL during daytime."

Line 357 'nocturnal boundary layer height': This is hardly visible in Fig. 11, either use different colors or indicate the PBL height in the figure.

We decided to add the potential temperature gradient to the original Fig. 11. The PBL height can be determined by the strongest gradient (dark blue colors).

The paragraph starting on page 10, line 343 now reads:
"Both locations are characterized by a very stable shallow boundary layer until 09 UTC with a depth of less than 200 m. Between 03 and 09 UTC the temperatures at Schnarrenberg are up to 1.5 K colder near the surface (Fig. 9) resulting in a stronger potential temperature gradient up to 400 m AGL compared to the IPM location. During the day, the boundary layer height increases to 400 m above ground as indicated by the constant potential temperature (e.g. Bauer et al., 2020) which is a typical value for European winter conditions (Seidel et al., 2012; Wang et al., 2020). The PBL heights are also visible by the potential temperature gradients ($\Delta\theta$) shown in Figs. 11c, d. During the morning hours, a very shallow boundary layer was simulated at Schnarrenberg (blue colors in Fig. 11c) while at IPM some fluctuations are present. During daytime, $\Delta\theta$ nicely shows the PBL height evolution up to 400 m AGL, while after sunset the PBL collapses to a very stable layer again (dark blue colors in Figs. 11c, d) with heights between 50—100 m AGL. Calculating the gradient Richardson number (Ri; Chan, 2008) (not shown) and assuming a threshold of 0.25 for a turbulent PBL 0.25 (Seidel et al., 2012; Lee and Wekker, 2016) gives similar results. After sunset around 15:30 UTC, the boundary layer collapses to a night-time stable boundary layer and a temperature inversion occurred again. "

Line 362 'that PM10 is a diagnosed quantity in our model setup.' What does this statement mean? I assume that Figures 12 and 13 show instantaneous values of the concentrations. What is the temporal variability at this high resolution? A presentation of the diurnal course of the pollutants should also be included. What is the effect of the too late drop in temperature on pollutant concentrations?

Our intention was to explain that PM_{10} is a sum of species, which have a diameter of less than 10 µm. All variables/fields shown in the paper are instantaneous values. We deleted this sentence from section 4.2.1 to avoid confusion.

To further illustrate the temporal evolution of NO_2 and PM_{10}, we decided to provide animations of the simulated NO_2 and PM_{10} concentrations as supplementary material.

Regarding the effect of the too late temperature drop on the pollutant concentrations, we unfortunately cannot draw a conclusion here.

Figures 1, 2, 3, 8, 12 and 13: The lat-lon presentation used here is quite unfavorable.

We do not understand what is "unfavorable" here? The lat-lon labeling with geographical coordinates in decimal degrees is reasonable for us.

Figure 4: Prep_chem_sources and MOZBC are separate tools, why are the put together? What is the meaning of the dashed arrow?

We slightly modified the workflow diagram (Fig. 4) . The dashed arrow was introduced unintentionally.

Figure 7: It is hard to distinguish any details. A presentation up to a height of 400 hPa would be sufficient.

We changed the layout of the Skew-T diagrams shown in Fig. 7 by limiting the vertical extent to 400 hPa and reducing the temperature range in addition.

Figure 11: It is hard to distinguish the PBL height from this figure, either change colors or indicate the PBL height.

Please see our answer on your comment to line 357 and Figure 11 above.