The paper presents WRF-Chem simulations to analyze changes in PM2.5 and O3 concentrations due to urbanization effects in the Chinese city of Chengdu. Two months are simulated, specifically January for PM2.5 and July for O3, and sensitivity experiments are performed for each of the two simulations to evaluate the impact of land-use changes associated with urban expansion and the impact of anthropogenic emissions. The authors also estimate the changes in premature mortality associated with the changes in air pollution concentrations. The study is well designed to differentiate between the meteorological changes associated with urbanization caused by land-use changes and the changes in emissions and is, in my opinion, a nice example of how a modeling study can be used to disentangle different factors. The paper is also overall very well written.

I have, however, some concerns related to the lack of information about the boundary-layer depth and vertical structure despite its supposedly large impact on the results (general comment 1) and the seemingly strong overestimation of PM2.5 in the model (general comment 2).

**General comments**

- The text refers repeatedly to the boundary layer height to explain, e.g., the temporal variations in PM2.5 (e.g., line 287, line 308), which makes of course perfect sense. These statements are, however, rather phrased as assumptions or hypotheses and the structure and development of the boundary layer is never discussed or even evaluated to make sure that the model actually captures this crucial factor correctly. In contrast
to the authors’ statement (line 266), I would say that the model does not reproduce the diurnal cycle of PM2.5 well (Fig. 5). The correlation coefficient is not that high and the model also generally overestimates the observed values during most of the time. This could potentially be connected to a poor representation of the boundary layer.

a.

I would recommend to start by comparing vertical profiles in the model with the radiosounding data to determine how well the model actually captures the vertical structure of the boundary layer. This is also commonly done before comparing the model output to surface observations to evaluate the larger scale first. This may even be helpful in better understanding the model performance for surface wind to see whether the large-scale flow is actually represented correctly.

b.

Can it be shown that the temporal development of the boundary-layer depth and vertical structure can explain the modeled development in PM2.5 and O3 as discussed in the paper? Similarly, the text also hypothesizes about the impact of the urban heat island (line 339: “Urban land use can enhance surface heating leading to an increase in air temperature ... resulting in and increase in the boundary layer height”). Can this actually be shown in the simulations?

c.

Have the authors performed any sensitivity tests with different PBL parameterizations? They can have a large impact on the boundary-layer structure and depth, which could potentially have a large impact on the results.

As mentioned above, the model seems to overestimate PM2.5 relatively strongly, with a mean bias of 23.4 μg m$^{-3}$, which the authors consider to be small (line 264). This is almost the same magnitude as the monthly mean change related to anthropogenic emissions (26.6 μg m$^{-3}$), which the authors call significant (line 373), and twice the monthly mean change related to urban expansion. During individual days, the overestimation reaches even values of up to 200 μg m$^{-3}$ (estimated from Fig. 5). I am somewhat concerned about the value of the results from the sensitivity tests considering the equally large model error and maybe that some processes affecting PM2.5 concentrations are not simulated correctly.

The paper reports very precise numbers for the changes caused by urbanization, e.g., the number of premature deaths resulting from PM2.5 (9386, 2609, 1321, and 1485 from ANAC, CVD, RD, and COPD) and the corresponding decreases of 424, 111, 55, and 56 without anthropogenic emissions. Considering that the equations given in Section 2.3 can likely provide only estimates and that I assume there must be large spatial variations in population density and air pollution throughout the city as well, which will affect the total premature deaths, I find the precision of these numbers somewhat misleading. Similarly, in the introduction a paper is cited by Liao et al.
listing temperature and PM10 concentrations caused by urbanization with a resolution of 0.1ºC and 0.1 μg m⁻³ while providing a range of more than 1ºC and more than 40 μg m⁻³ and no information about the magnitude of the urban expansion.

I found it rather interesting that the model simulations for both PM2.5 and O3 show that very high concentrations can also be observed for large distances downstream of Chengdu. Assuming that there are also large populations in the vicinity of Chengdu, the overall impact would thus be even larger. I understand that an analysis of the effects around the city would be clearly beyond the scope of the paper, but I simply found this another interesting aspect.

Specific comments

Line 54: To which socioeconomic developments is the text referring?

Line 61: “the thermal contrast of the topography” is not entirely clear? Thermal circulations are caused by temperature gradients in the atmosphere.

Line 66: I would suggest to be more precise here and specify in which ways the diffusion conditions are more complicated.

Line 89: “Chengdu has the most complex terrain in the world”. This is a strong statement. How is that defined and determined?

Fig. 1: The figure shows SO2 emissions, but their role is never actually mentioned in the paper. There are also high-emission areas in the southeast corner of the figure. Are these other large cities?

Air quality and meteorological data: How many stations are there in Chengdu, i.e., the
ones that are averaged for the analysis (line 129)? And how strongly do the locations differ?

Line 137: At which height is the lowest model level? Is the boundary layer also properly resolved during the night?

Line 175: The text says that β has units of (μg m⁻³)⁻¹, which makes sense to have a non-dimensional exponent in eq. 5, but Table 3 mentions units of %.

Line 180 and Table 3: The references in the text and in the figure caption are not the same.

Lines 192 and 439: How are “pollution episodes” defined? Are these continuous periods, during which the air quality standard is exceeded, i.e., they can have different lengths? Does “pollution” always mean exceeding the standards (e.g., line 232) in this paper?

Line 213: What does “administrative division adjustment” mean?

Line 222: I am not sure how meaningful it is to calculate and discuss an increase or decrease between the first and last year if there is no real trend and the inter-annual variability is larger than the overall trend.

Table 4: Are these numbers calculated for individual days and then summed over the whole year? Section 2.3 does not say for which time period the given equations are applicable. It is mentioned that equation 5 is applied to MDA8 O3, but it does not say either over which time period the PM2.5 values have to be averaged.

Line 235: “cold westerly winds from the north” seems to be a contradiction. Are the winds from the west or from the north?
Line 239: Why do the humidity differences cause the inversion? Actually, it seems that the temperature decreases with height, i.e., there is no inversion, but only a stable layer.

Fig. 3: The 850-hPa level should be lower than the Tibetan Plateau. So what do the plotted fields over the Tibetan Plateau actually show? The areas above 850 hPa should best be masked in the plots. Also, are these monthly averages?

Fig. 4: (I) What do the black dots to the right of the subfigures (where one usually finds the wind arrows) mean? (ii) Are the profiles from the radiosoundings or also from the NCEP analysis as Fig. 3? Are they monthly means? (iii) Why are both 00 and 12 UTC shown if the differences are never discussed?

Line 269: A comparison of RH may not actually say much about the model performance in simulating humidity correctly because it combines the effects of humidity and temperature. It might be better to use, e.g., mixing ratio or specific humidity if the observations are available to convert the units.

Fig. 5

Are the measurements averages over all stations in the city as suggested in the previous section? If so is the model output equally averaged over all urban grid cells or over all grid cells closest to the stations?

The figure and figure caption do not say which of the shown datasets (black dots or colored lines) is the model and which the observations. Based on a statement in the text that the model underestimates RH, I assume that the observations are the dots.

Wind speed, RH, and temperature seem to have a resolution of only 1 m s\(^{-1}\), 5 or 10\%, and 1\(^\circ\)C, respectively, which makes it really difficult to compare the diurnal cycles because, e.g., wind speed remains basically constant for hours. This will also affect the quantitative comparison (correlation coefficient). Is this the actual resolution of the data?
Why is wind direction not shown?

Line 292: “mountain-valley breezes” – Where can they be seen in the figure? Chengdu seems to be located in a northeasterly flow directed towards the mountains all day and night. How does that agree with diurnal circulations?

Line 294: It is very difficult to see in the horizontal cross sections whether there are actually converging flows, because the figures show the flow at the first model level (I assume so because the figure caption does not say), which means very large elevation differences. The shown wind field is about 2-3 km higher in the leftmost part of the figure than in the right part. Based on the figures, one could also argue that the high concentrations are simply located at the foot of the mountains, i.e., they are trapped by the terrain.

Line 298: Is this really a secondary circulation that is forced by the terrain? Could it not be the upper-level westerly flow that is shown in Fig. 3? This would also agree with the absence of this “secondary circulation” in July, when a weak-gradient situation prevails at upper levels.

Technical comments, typos, etc.

Several of the figure axis labels are very small and thus almost impossible to read, e.g., axes labels in Fig. 2 (in particular the days of the week, but also the months), Fig. 4., Figs. 6-11.

For past events, the past tense should be used: e.g., “increased by 24.2%” (line 214). Similarly, lines 217, 232 and 438. There may be others.

Lines 25 and 30: It would be good to add that the mentioned changes in PM2.5
concentrations are based on monthly means.

Line 26: MDA8 is not defined

Line 55: “increases in anthropogenic emissions”

Line 60: The reference list contains only Qian et al. (2022).

Line 67: “notable pollution episodes”

Line 67: Chow et al. (2013) is not included in the reference list.

Line 71: I don’t understand the meaning of the sentence “Although the principles ...”

Line 91: It would be really helpful to label these terrain features in Fig. 1a?

Fig. 1: It would be helpful to indicate the location of Chengdu in subfigure (b). I assume it corresponds to the red location in the center with high emission values?

Line 115: “Data and methods” might be a more common section title.

Line 200: “The high temperature and strong sunlight contribute to the elevated ...”
Line 189: “two crucial air pollutants that account for air pollution”. This is a somewhat awkward sentence. “that account for air pollution” can easily be removed.

Line 237: “to the west of Chengdu”

Line 240: “cold air”

Line 314: “and mixing” (remove “well”)

Line 360: “can increase” suggests that this is a maximum value. Or is the text also referring to a monthly average as for PM2.5? It would actually help to always add “monthly mean” when talking about monthly mean values (e.g., line 404) to avoid misunderstandings.

Fig. 10: Is there a reason why the color scale is not adjusted to show the whole range? Based on the color bar it looks like the highest values are around 10 μg m⁻³, whereas the text actually mentions a monthly average of 26.6 μg m⁻³.

Line 407: “assess” instead of “access”

Line 545: This reference is not cited in the manuscript.