This work aims to evaluate TROPOMI NO2 retrievals against airborne (APEX) observations over Antwerp with the help of the WRF-chem online CTM. A large part of the manuscript is devoted to the evaluation of WRF-chem during the period of interest, both as regards meteorology and atmospheric composition using a variety of in situ observations. The performance of the model is also tested using an extensive number of different parameterizations for the PBL height and the most suitable for the case in question is identified. Further analysis is performed to evaluate the emissions that were used and a new weekly profile for NOx emissions is estimated and used in the sections that follow. Finally, TROPOMI NO2 retrievals are compared against APEX and the bias of TROPOMI is estimated, with the aid of WRF-chem taking into account the averaging kernels.

The manuscript is well written and structured. The authors follow a long investigative line, aiming to identify and control several of the uncertainties that meddle with such evaluation efforts. It also provides a wide range of insights in the performance of a model (WRF-chem) and on methods to evaluate satellite and airborne column observations. However, there are some elements of the paper that should be improved before publication:

**General comments:**

The manuscript uses averaging of in situ measurements quite extensively for purposes of evaluation of the model but also as a basis for further calculations. This is understandably convenient in many ways, as e.g. the consolidation of measurements makes presentation more concise and easy to follow, but presents certain challenges as observations are often inhomogeneous and stations not always representative of the entire domain in question. The authors should discuss this challenges and attempt to justify their averaging strategy and consider adding per station plots in the supplement.
section 2.1.1: Not entirely clear here why this was in two separate runs/periods. Could you please explain in the text?

section 2.2.8: The description of the methodology for the calculation of the emissions is quite standard, please consider shortening it.

sections 4.1.2 and 4.1.3: The authors should try to shortly present their motives for comparing various meteorological variables with the model running with different PBL schemes.

section 4.2.1 Is testing which parameterization(s) work best in a specific case by performing a sensitivity study such as this the indicated way of working with a model like WRF-chem? Or are there other reasons to do this in this work?

section 4.2.2 The purpose of this section is not sufficiently explained/supported. The way it is presented, it hardly adds to the analysis. Running the model with temporally constant emissions is not really a sensible option to choose from. The paragraph could be removed, unless the authors make more it transparent how it integrates to the rest of the manuscript.

section 4.3.2: Considering the fact that these bias corrections (formulas 7 & 8) are based on a comparison for two days, over Antwerp and for a certain range of NO2 column values, one is left wondering if/how they could be used in some way outside the frame of this study. The formulas are also mentioned in the discussion later on, but the comparison with other studies there is done by means of of the regression relation between TROPOMI and an independent observation (APEX) and not formulas 7 & 8.

section 4.3.3 Similarly, the authors could comment on the general applicability of the emission adjustments introduced here. Can this crude inversion be proposed as a method that can be used outside this study?

**Specific comments:**

I. 14: read "generally good performance". Also, please provide figures to support this qualitative comment.

I. 28-29: Provide some numbers to support the qualitative remarks.

I. 6-12x10^15 molec/cm2. Is this range of values low, high?

I. 71: The official validation reports can also be cited here, found in: https://mpc-vdaf.tropomi.eu/index.php/nitrogen-dioxide?start=7

I. 75-76 Some references could be added here, e.g.: 


l. 87: "recent" is a stretch, it is probably not meant in absolute terms but in terms of proximity to the period of the study.

l.92-93: Please rephrase, it's probably not a promise anymore.

l. 136-139: Not clear why the 15-day period was not run as one continuous hindcast and had to be split in partially overlapping smaller runs.

l. 164: Which CAMS model would that be, the global or the regional? Please specify.

l.165: What is CAM-chem? Please provide reference.

l. 212: "the entire model domain": is it d01 or d02?

l. 274: "over the two domains": over both domains? What would that mean exactly?

l.370 please use standard syntax: "unknown reasons".

Figure 12: NO2* has been defined already so no real need to explicitly refer to the correction of measurement interference in the caption. The same for various places in the text.

l. 569: Please provide reference for the conversion rate.
I. 571-573: How relevant could that be? The comparison for NO and NO2 in the afternoon of the 29th appears to be quite good (figure 13).

I. 607: "unclear reasons"

I.617: "data" should be specified.

I.617: Homogenize "NO2" throughout the manuscript by using subscript.

I. 804: "a few percent": please provide a figure