Review of the manuscript by Cifuentes et al.: "Uncertainty assessment of TROPOMI NO2 over Europe using ground-based remote sensing observations"

The manuscript presents an assessment of key sources of uncertainty in tropospheric NO2 columns from the TROPOMI satellite instrument by comparing them with Pandora and MAX-DOAS ground-based observations and by using model simulations over Europe. The authors provide a detailed description of the uncertainties associated with satellite observations and of the different factors affecting the results of the comparison with ground-based data. The results can be quite useful as a reference for future validation studies aimed at interpreting the various causes of differences between satellite and ground-based data. The paper is scientifically sound and the method appropriate. I recommend publication after addressing the following minor revisions:

1) One of the main conclusions of the paper is that the uncertainty assessment is too optimistic and that the differences between ground-based and satellite observations exceed the total expected uncertainties. In line 562 the authors state: "These results suggest that the uncertainty estimates for the individual instruments or the representation errors derived in this study may be somewhat optimistic. Alternatively, the discrepancies could indicate the presence of additional sources of uncertainty that have not yet been accounted for in the current analysis."

   Could the authors further discuss, possibly with examples (e.g. in the conclusion section), what aspects of the uncertainty estimates may have been underestimated, and how? In addition, could they elaborate on potential unaccounted sources of uncertainty not included in this study, and provide recommendations for future studies

2) Concerning the different temporal sampling, in line 352 the authors state: "For temporal alignment, MAX-DOAS and Pandora observations are averaged within a 1-hour window (¬±30 minutes) centered on the TROPOMI satellite overpass time."

   Doesn't this choice play a role in the calculated differences? Can the authors quantify this effect for a set of reasonable temporal averaging windows and/or better justify their choice?

3) Regarding spatial representation errors: you assess these in the Netherlands to be up to 6%. How applicable is this estimate to other sites? Can the authors elaborate on the factors that may influence this estimate, particularly in regions with different geographical or emission characteristics? Do you have an idea, based on the literature, of how much higher this variability could be?

4) Figure 7: Please include a description of all lines and colors shown in the different panels. What does the gray line represent? What is shown in the middle panels?

5) Can the authors comment on the overall scalability of their findings beyond Europe? What are the expected issues outside this domain, for example in much more polluted regions or at high-altitude sites?