I believe this paper addresses an important issue relating to the appropriateness of comparisons between ground-based in situ and satellite remote sensing observations and/or model output. In the abstract, the authors highlight the importance of being able to identify or categorize sites to identify those which will be more or less suitable for comparison with satellite data or use in data assimilation. However, no further mention is made in the manuscript of how the analysis carried out would help in this task. I feel the paper would be strengthened by some more concrete examples of how this approach is helpful. For example, would one restrict comparisons to sites whose 12 hour catchment area is smaller than the satellite footprint or model grid box?

In general, I was a little confused with some of the terminology in the manuscript. The terms 'catchment area' and 'footprint' are used before they are defined. P20025, L22-24 is the first time in the text that the term 'catchment area' is used. How is differentiated from the area of representativeness? You state that the catchment area is independent of the pollutant - does it only take into account advection and not the impact of emission/deposition? Later in Section 2.2.2 it is stated that the catchment area is defined to include areas from which fluxes will make a significant impact on the receptor site, so how is it that the catchment area is independent of the emission sources? Does the catchment area calculation assume hypothetical evenly distributed fluxes on constant magnitude?

P20024, L20-23 When you say emission and deposition data are proxies for concentrations, do you really mean that the mixing ratios will scale with the rates, or that the variability in the mixing ratios will have similar spatial and temporal scales to the rates of emission and deposition? As written, it's a little unclear.

P20029 L17-22 This sentence is too long and difficult to understand.

P20033 L13, supplementary, not supplementing (also P20044, L3)

Section 3.1 and Figure 1 – this is the first time that the term footprint is used in the analysis. What does it represent? What are the units for the colour scale?

In Section 3.6, a less computationally expensive approach is presented (with results in the Supplementary material). Could the authors be more explicit about the type of application for which this approach could be sufficient?

The standard deviation of mixing ratios at each site as calculated from a full year of data, whereas the deposition velocity used for ozone was specifically for summertime conditions. Do you think your agreement with observations might change is you had allowed the deposition velocity to vary seasonally?