Reply to reviewer 2

We would like to thank the reviewer for their time, especially since this is a long paper, and useful & considerate comments. Their comments are repeated below in italic, followed by our answers.

General comments

However, some of the findings have been presented already in the previous papers S16a and S16b.

These papers bring out different aspects of sampling issues, for a variety of observing systems and observables. While there is a thematic overlap, we feel the overlap is small content-wise:

- S16a concerns spatial sampling, in a model evaluation context. It uses regional model data and assumes highly localised observations. It really is a study of sampling errors for continuously measuring (in-situ) ground-sites or incidental flight campaigns. It shows that different observations can lead to very different sampling errors. It includes a lot of sensitivity studies for different strategies in comparing a global model to the observations.
- S16b concerns temporal sampling, in a model evaluation context. It used global model data and real remotely sensed observations. It compared sampling errors to actual model errors and showed them to be of similar magnitude. It also showed that models compared better with real observations after temporal collocation. It showed that the sampling error for VIS remote sensing data would depend on longitude when using daily model data.
- The current study concerns spatio-temporal sampling in a general sense. It uses regional model data and a separate model for the idealised observational spatio-temporal sampling. It allows the study of sampling issues in satellite L3 data (not possible with S16a or S16b) and provides realistic estimates for representation errors after temporal collocation (again not possible with S16a or S16b).

The paragraph describing S16a and S16b (p 2 line 20-28, in the introduction) has been rewritten to clarify this.

Minor/specific comments

Introduction: I would suggest to add a reference related to representation errors in ozone observations, e.g., Sofieva, V. F., Kalakoski, N., Päivärinta, S.-M., Tamminen, J., Laine, M., and Froidevaux, L.: On sampling uncertainty of satellite ozone profile measurements, Atmos. Meas. Tech., 7, 1891-1900, doi:10.5194/amt-7-1891-2014, 2014.

This is an interesting paper that the other reviewer suggested as well. We were not familiar with it but have now added it to the introduction.

Page 3, Section 2.1: Please explain N10/N50 and introduce "BC" as abbreviation for black carbon (used later on in the paper).

Agreed.

Sections 3.2 - 3.5: I would suggest to merge the description of the different figures into one subsection.

We agree the page layout does look a bit awkward, but the benefit (we hope) of the subsections is that readers will be able to quickly look up the description relevant to a particular graph. We suggest to not change this.

Page 8, lines 9/10: Why is clear sky day-light AOT lower than average AOT?

Average day-light AOT is only slightly smaller than night-time AOT (few %) for unknown reasons. However, average clear-sky AOT is decidedly smaller than cloudy AOT (mostly due to increased humidity in the cloudy column). Day-light is mentioned in because it is one of two conditions for valid observations. We have replaced 'clear-sky day-light AOT' with 'observable AOT' and added an explanation

Page 11, Sec. 6.2: Are the numbers the errors due to "purely spatial sampling"?

Indeed. Our assumption is that such in-situ ground-sites measure continuously, at least for the duration of a day. Consequently, daily representation errors are purely due to spatial sampling.

Page 14, Sec. 9.3: Please add a comment here that you find similar results for polar orbiting satellites and geostationary satellites. At least for me this was a bit surprising as I expected lower errors for the geostationary satellite observations due to multiple views per day (instead of one measurement per day for the LEO).

Agreed.

While this result seems counter-intuitive, it is a consequence of 1) temporal variation throughout the day that even the GEO sensor can not observe; 2) cloud masking over 210 by 210 km2 that prevents observation of the entire area by the GEO sensor. These two causes contribute in roughly equal measure to the final representation error for geostationary sensors that can only observe during the day.

Although we do not mention this in the paper, we considered the case of a physically impossible observing system: a geostationary satellite that can observe during both day and night. For areas without (!) clouds, daily representation errors are indeed zero as expected.

Page 15, line 30: Not sure whether I can follow conclusion 3). Could you please add an explanation here. Like referee #1 (her/his comment no. 18) I think that estimates of the monthly mean will improve with increasing number of observations.

While we agree with the reviewer's point, we wanted to study representation errors in the context of realistically achievable number of observations. Reviewer's #1 example is fairly abstract as both cases seldom occur.

The relevant figure is Fig. 8 which shows monthly representation errors for ground-sites as a function of *required* temporal coverage. *Actual* temporal coverage (or the number of observations) will always be higher and is shown by the black dotted line (right axis). The brown line (left axis) represents representation errors when data are *not* collocated (which is what our statement was about). Note that an increased number of observations *may* reduce representation errors, as is shown for Japan. However, for Oklahoma (and most other regions) this error hardly changes with the number of observations. A combination of strong temporal variation throughout the day, and different spatial sampling of the ground-site and represented area prevents an increasing number of observations to reduce representation errors.

Strictly speaking our statement should have read “Using a minimum required number of observations cannot be relied upon to control representation errors.” The text will be changed.

Page 16, lines 7/8: You say that the results were robust across the regions, but what about the selected months? Did you analyze the natural variability of the observables as a function of month? Do you think that the selected months are representative for the whole year / other years? Errors may increase/decrease significantly if natural variability is different for different months.

Errors will increase or decrease with variability but we never saw a significant change (for argument’s sake here defined as a changed by a factor 3, ie. a 30% error becoming a 10% or 90% error). Also, “robust” referred to the second part of the previous sentence (“their behaviour (e.g. impact from sampling or collocation)”). We accept there will be changes in exact error values. We have rephrased to improve clarity.

Fig. 2: I find it difficult to identify the blue line. Is it possible to show only one red line (e.g., mean/median + std.dev. of all observations 2000-2010)?

This figure will be recreated, with the blue line more prominent.

Technical corrections

Page 22, Fig. 5, caption: “210 x 210 km” -> “210 x 210 km^2”

Page 25, Fig. 9, title: “obs: 210 x 10 km^2” -> “obs: 10 x 10 km^2”

Page 31, Fig. 16, title: “obs: 10 x 210 km^2” -> “obs: 210 x 210 km^2”

Page 34, Fig. 20, caption: “PM25” -> “PM2.5” and “km” -> “km^2”

Page 35, Fig. 21, caption: “PM25” -> “PM2.5” and “km” -> “km^2”

Page 36, Fig. 23, caption: “km” -> “km^2”

Page 37, Fig. 24, caption: “km” -> “km^2”

Most of corrections will be implemented in the final paper. Note that in Fig 16, our caption is correct: the LIDAR sweeps out a narrow transect (curtain), represented by a 10 x 210 km2 area.