Referee report for "*On the spatio-temporal representativeness of observations*", submitted by Nick Schutgens et al. to ACP.

Dear authors,

It was with great interest and pleasure that I read your new contribution on the representativeness of observations. As measurement accuracies improve and multi-instrument studies become more and more prevalent, the representativeness of individual and averaged data becomes of increasing importance and this paper therefore addresses a real and acute need. While the general approach (using high-resolution model fields to simulate representativeness errors) is not novel, this paper reports on the first application to estimate the combined spatio-temporal representativeness of aerosol(-related) data and as such clearly deserves publication. Moreover, the manuscript reads very well with a clear structure, few typing errors, and easy-to-read figures. My only truly general remark would be that the work got chopped up into too many (3) papers, leading to some repetition but also requiring the reader to have at hand the other papers, and actually making parts of the previous papers, less than a year old, somewhat obsolete. Of course, this can no longer be changed.

Besides these general thoughts, I have a few **specific concerns and suggestions** for improvements:

1) The title needs to be more specific, clarifying that this paper is about aerosols. The scope of the results presented here does not warrant the current title. For instance, the representativeness of trace gas and other meteorological observations is impacted by many factors not addressed in the current paper (e.g. the photochemical diurnal cycle of NOx, Ozone depletion in the polar vortices, small-scale variability in the water vapour field, the diurnal variability of temperature…). See also the next comment.
2) Even though this paper is about aerosols, the introduction could/should touch more broadly upon the literature that exists in other atmospheric domains, also outside the assimilation context. For instance: the seminal workshop report by Nappo et al. (1982, Workshop on the representativeness of meteorological observations, held in Boulder in June 1981), papers on the sampling uncertainty and representativeness errors in Ozone observations (Sofieva et al., AMT, 2014b; Section 3.1 in Coldewey-Egbers, AMT, 2015) etc.
3) Even though some references are provided in the introduction to empirical estimates of aerosol spatio-temporal variability and some caveats are given in the conclusions, it would be good to have a paragraph providing some quantitative information on the known/expected variability within a model pixel, i.e. variability at scales smaller than 10km and 1hour. This is in particular relevant to assess the completeness of the error estimates for in situ measurements.

Below are some more highly specific or **technical comments**.

1) Section 2.1, 1st sentence seems redundant (basically saying that the simulated fields are those that were simulated).
2) Section 2, more general: are the hourly data hourly averages or hourly snapshots?

3) Page 5, line 10 (about the observational sampling): in reality, the observations don't occur exactly on the x,y,t of the model. Does that matter, and if not, why not?
4) Page 5, line 14: temporal collocation can of course also be used when comparing different measurement (e.g. in situ versus satellite, so not only in observation-model comparisons), so the scope of these results is wider than is portrayed in the paper.
5) Section 3, more general: why only look at temporal collocation and not spatial collocation? Clouds could also be dealt with using spatial masks instead of temporal collocation. For orography, a spatial mask would be the only solution.
6) Page 7, section 3.4, 1st sentence: Fig. 6 is the first box-whisker plot, not Fig.7
7) Page 8, section 4, 1st sentence: Maybe add "only" to the beginning of the sentence: "Only the EMEP…"
8) Section 4 (and subsequent, more general): why this particular choice of 210x210km2? Most current gridded data sets, whether from satellite or model, have better resolution than that.
9) Page 8, line 9: explain why day-light AOT is lower than average AOT, if known.
10) Page 8, line 21: how come? Please explain briefly.
11) Section 5: again, why 210x210?
12) Page 8, line 29-30: is it known why cloudy AOT is larger than clear-sky AOT for these regions?
13) Page 9, line 12. Although you make it explicit later in the paper (in section 5.3), I think it would be good to state earlier on that the strong effect of temporal sampling/the huge gains with temporal collocation, are all about clouds.
14) Page 9, line 14: satellites -> satellite
15) Page 9, line 19: you point out the similar errors between a ground-site and a satellite sounder with a repeat cycle of 1 day. That may be true for the average size of the errors, but the spatio-temporal pattern of those errors should be vastly different, no? The paper contains lots of box-whisker plots summarizing the statistical properties of the representation errors. It would perhaps be nice to see some more maps (like Figs 3 and 4) to be able to judge the spatial patterns of the representation errors. This is to be seen as just a suggestion: if the authors don't see value in that, they can perhaps just include a statement to explain why no further maps are shown.
16) Page 9, line 32: due to -> obtained after temporal
17) Page 12, line 16: please explain somewhere what N10 is.
18) Page 14, line 4-5 (Section 9.2). You state: "The number of observations used in constructing monthly averages cannot be used to control representation errors". I don't understand where this conclusion comes from (which probably indicates I misunderstood something earlier on). I can hardly believe this to be correct: surely a monthly average based on a measurement every day of the month will lead to a better estimate of the monthly mean than an average based on just 1 measurement?

Looking forward to reading the final paper,
A referee