Review of Tadic et al. "Spatio-temporal approach to moving window block kriging of satellite data"

This paper introduces a spatio-temporal model of environmental data that can be used to produced maps of more uniform coverage, than are capable from the raw satellite observations alone. It applies this approach to three environmental variables from the carbon cycle science community: column co2 (XCO2), column methane (XCH4), and solar-induced chlorophyll fluorescence (SIF). This work follows on a long series of papers by a number of authors with similar models.

General Comments

I will be upfront and say that as as reviewer, I am not well-versed in the geostatistical estimation literature, and am rather an expert on these carbon cycle variables themselves. So my review will focus less on the details of this particular approach, and rather some bigger picture questions.

My main complaint on this work, which honestly is more a complaint about the entire field who does this, and is not particular to this paper, is that it fails to really explain the utility of kriged satellite data beyond simply "pretty pictures". Most data users who attempt to extract scientific results from the data do not use 3D maps. The reason is the data assimilation systems typically ingest the sounding (level-2) data directly (e.g., Houweling et al., 2016; Massart et al., 2016). Therefore, some commentary (like a paragraph in the introduction section) on the use of level-3 maps vs. direct data assimilation approaches would be worthwhile, perhaps pointing to scientific results using this method that would have been missed otherwise.

Beyond that, the few basic statistics on the quality of the spatio-temporal (ST) method over and above pure spatial methods do not really argue that the ST approach buys you much. The actual statistics given in Table 1 are really rather similar between pure spatial vs. the ST method. So the paper seems to argue that this is really useful, but the data really don't back it up. My read is that 1-3 day spatial approaches are really quite adequate for this purpose.

Finally, the validation approach is probably not valid for the GOSAT case. This is because there are only ~14 orbits per day, and huge swaths of the globe are missing even if all the data are used. Therefore, you don't really learn the error statistics unless you perform a simulation-based test where you start with a "true" map, sample it like the satellite would, along with realistic observation errors, and then run it through the kriging algorithm to reconstruct the 1-day map. This paper would be much enhanced if such a realistic validation test were performed. I realize the authors can easily say "beyond the scope of this paper" because what I am suggesting is not easy, but it is really the only way I can see to get at the true errors in the proposed algorithm.

Specific Comments

Abstract: Makes that statement that this approach only requires a limited number of assumptions – that "the observable quantity exhibits spatial and temporal correlations that are inferable from the data." But this seems like a single assumption? Are there more assumptions? Please reword as necessary.

Section 2.1 I don't get why subsampling is necessary. The data volumes don't seem that large. Is it really just because using ALL the data to define the correlations is computationally infeasible? Please expand on this point a bit in this section. Or it just doesn't buy you anything? If the latter, then how do you determine how much subsampling is justified before you start to introduce errors?
Equation 1: I just don't get the difference between the Ps and Pt terms. Pt I get. Ps I don't. For instance, in this method, soundings that are 0.5 km from the center of the grid box are 4 times more likely to be selected than soundings 1 km from the center. Even when the spatial resolution of the soundings themselves is 10 km!, and typically decorrelation lengths of CO2 and CH4 in the atmosphere are more like 100+ km! It seems like an exponential structure for Ps makes a lot more sense. Or at least something like $h_s' = \max(h_s, h_{min})$ where h_{min} is some minimum resolution distance. (And for Co2 and CH4 I would argue making this at least 10-20 km). There is no physical justification actually cited for these functional forms. If the functional form for Ps is changed to exponential, then obviously the entire discussion from likes 122-134 could be shortened or eliminated.

Line 268: …ecological applications. Please provide some references here.

Technical Comments
Line 229: "is a Lagrange multiplier" is missing the actual variable.

Line 316 (and later): ST is never defined. Suggestion you modify the sentence here to say …performance of spatio-temporal (ST) versus…

Page 10, top: I disagree with the conclusions stated here. The MAE and RMSE even for the 7d results seem really only marginally better for ST. And 1d pure spatial, which seems like a more fair comparison as the ST is also done at the daily scale, seems to do as well or better than ST! Also the % lying outside the different uncertainy bounds doesn't seem useful, especially considering that the numbers are significantly less than that expected from pure Gaussian errors. Could the authors explain why they are so much less?

Conclusions near line 404: Again I just don't the ST approach being better. It is only marginally better than 7d, and is slightly worse than 1d. At best this is a wash. Please reword.

Citations

Houweling, S., et al, 2015. "An intercomparison of inverse models for estimating sources and sinks of CO_2 using GOSAT measurements." *Journal of Geophysical Research: Atmospheres* 120.10, 5253-5266.

Massart, Sébastien, et al., 2016."Ability of the 4-D-Var analysis of the GOSAT BESD XCO_2 retrievals to characterize atmospheric CO_2 at large and synoptic scales." *Atmospheric Chemistry and Physics* 16.3, 1653-1671.