Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-192-AC2, 2017

Geoscientific Model Development Discussions

Interactive comment on "Spatio-temporal approach to moving window block kriging of satellite data" *by* Jovan M. Tadić et al.

Jovan M. Tadić et al.

jotadic@lycos.com

Received and published: 5 January 2017

Reviewer: I will be upfront and say that as reviewer, I am not well-Â■âĂŘversed 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- âĂŘ2) 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-Â■âĂŘ3 maps vs. direct data assimilation approaches

would be worthwhile, perhaps pointing to scientific results using this method that would have been missed otherwise.

Authors: We would like to emphasize that the methodological advances we presented go beyond the application space defined by three chosen examples. Also, considering the presented method purely as a "mapping" method represents an over-simplification. The method can be, of course, used to produce maps, but it is also capable of upscaling the observations providing estimates at support larger than the support of observations, with associated uncertainties. Example: Imagine that we intend to compare XCO2 derived from OCO-2 and GOSAT retrievals. The direct comparison is not possible because of at least three reasons: (a) the measurements are not collocated (and thus mapping is required), (b) the averaging kernels are different, and (c) the measurements have considerably different spatial statistical properties - support (and thus upscaling of the OCO-2 observations is required). The differences in support can cause substantial differences in reported values (see Tadic and Michalak, 2016). The example shows that even a simple comparison of the same physical quantity measured by two satellites requires a relatively complicated mapping and upscaling methods. The similar conceptual problem remains when model outputs, usually given at regular grids and standardized support, are compared to observational datasets, and when satellite products have to be compared to in situ observations (for example Aircore or aircraft profiles) which are not collocated. Interpolated products could be useful for providing background concentration estimates or initial condition estimates, for example in inverse modeling studies. NOAA has recognized the problem stemming out from the inconsistency in spatio-temporal coverage, and provided justification for mapping: http://www.esrl.noaa.gov/gmd/ccgg/globalview/index.html. The data assimilation systems indeed ingest observations rather than mapped products, but mapping and upscaling method presented here is not limited to greenhouse gas measurements. While transitions of the type Level 2(obs.) - > Level 4(flux patterns), and later eventually Level 4 - > Level 3(maps) are possible, not all the physical quantities have Level 4 data. Actually the solar induced fluorescence (SIF) is a good example. Level 3 data

have been used or generated in a number of recent studies, at the same time providing the insight into their value and scope of application: Liu et al., 2012; Basu et al., 2014; Maksyutov et al., 2013, etc. There are at least few studies we are aware of that currently use mapped and upscaled products: 1) Shiga at al. (Carnegie institution for science) currently use spatio-temporally (ST) mapped SIF as ancillary data in inversion studies, and preliminary results show that ST product is more consistent with atmospheric CO2 observations, then purely spatial product. The publication will follow soon (private communication). 2) Zheng et al. (Yale University) currently use mapped SIF product to study the impact of extreme drought on photosynthesis. The publication will follow soon, too (private communication).

References:

Tadić, J. M., & Michalak, A. M. (2016). On the effect of spatial variability and support on validation of remote sensing observations of CO2. Atmospheric Environment, 132, 309–316.

Liu, J., I. Fung, E. Kalnay, J.-S. Kang, E. T. Olsen, and L. Chen (2012), Simultaneous assimilation of AIRS Xco2 and meteorological observations in a carbon climate model with an ensemble Kalman filter, J. Geophys. Res., 117, D05309,doi:10.1029/2011JD016642.

Basu, S., M. Krol, A. Butz, C. Clerbaux, Y. Sawa, T. Machida, H. Matsueda, C. Frankenberg, O. P. Hasekamp, and I. Aben (2014), The seasonal variation of the CO2 flux over Tropical Asia estimated from GOSAT, CONTRAIL, and IASI, Geophys. Res. Lett., 41, 1809–1815, doi:10.1002/2013GL059105.

Maksyutov, S., Takagi, H., Valsala, V. K., Saito, M., Oda, T., Saeki, T., Belikov, D. A., Saito, R., Ito, A., Yoshida, Y., Morino, I., Uchino, O., Andres, R. J., and Yokota, T.: Regional CO2 flux estimates for 2009–2010 based on GOSAT and ground-based CO2 observations, Atmos. Chem. Phys., 13, 9351–9373, doi:10.5194/acp-13-9351-2013, 2013.

Reviewer: Beyond that, the few basic statistics on the quality of the spatio-Â■âĂŘtemporal (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-Â■âĂŘ3 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 $\sim$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-Â■âĂŘbased 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-Â■âĂŘday 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.

Authors: Two comments listed above are related to each other and will be handled together. First, the statistics differs in three test cases so the general conclusions would be pretentious. We provided potential explanations for a poorer performance of the ST approach in GOSAT cross-validation (Lines 384-399). We would like to point out to our reply to Reviewer 1 about errors that are spatially but not temporally correlated, and its effect on the apparent poorer performance of the method, in one satellite case and based on the specific metrics used here. The poorer performance could actually result from ST method providing more accurate, unbiased estimates, yet this has to be further studied. While leave-one-out cross validation might not be the best method for providing the accurate error statistics (as we pointed out both in our reply to reviewer 1 and in the manuscript (Lines 394-396: "Although the resulting estimate may appear inferior during cross-validation, this is because that estimate will not reproduce regional biases in data from the time slice of interest.") it has a long tradition as tool used to assess the performance of similar methods, and we decided to present its results, but pointing

out to potential problems in using it. The synthetic study suggested by Reviewer only for GOSAT case could be usable, but there are at least two entailed problems: (1) we would like to keep consistent error statistics tools across all examples and, (2) synthetic experiment like the suggested one would require a realistic individual retrieval uncertainty estimate. Making assumption about the individual retrieval uncertainty would just mean pushing the problem down the line. There is a long list of studies (see Reference in response to reviewer 1) which all relied upon leave-one-out cross validation done in the manner similar to the one from this study, and to assure comparability between the results we followed the same pattern.

Reviewer: 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.

Authors: Corrected.

Reviewer: 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?

Authors: The subsampling is always necessary in moving window approach to preferentially focus on variability near (in spatio-temporal sense) an estimation location, independently of the available number of observations. In addition, in case of GOME-2 and IASI the number of available observations significantly grows if multiple time slices are included. For example, the covariance matrix covering two weeks of IASI data would have 3 billion entries. It is clear that some kind of subsampling has to be done in order to keep the problem at computationally feasible levels. The estimates do not degrade gradually when subsampling fewer and fewer data points, they rather stay fairly

constant over a certain range of subsampled dataset sizes, and then start to degrade at certain level. To determine such a level one has to produce a series of estimates for the same location while subsampling fewer and fewer measurements, until estimates start to differ above the acceptable threshold. We implemented similar approach (Lines 100-101).

Reviewer: 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 hs' = max(hs, hmin) where hmin is some minimum resolution distance. (And for Co2 and CH4 I would argue making this at least 10-Â■âĂŘ20 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-Â■âĂŘ134 could be shortened or eliminated.

Authors: The choice of the form of the subsampling function is one of the subjective choices the modeler has to make. Instead of arguing why we did select Ps and Pt forms as in the paper, we would like to explain why the 1/ht2 was not used. The satellite data (Level 2) come with continuous spatial and discretized temporal coordinates. Phrased differently, data are temporally pre-aggregated (day 1, day 2, etc.). Any form of the temporal component of the sampling function from 1/htn family would lead to sampling only from the time slice of the estimation location because 1/0 would result in an infinite sampling probability for such observations, unlike observations in other time slices. So the selection of exponential form for the temporal component partially came out of necessity. We do not quite understand the argument about 0.5/1km distance from the center given the spatial resolution 10km (GOSAT). While sampling probability is indeed 4 times higher for 0.5km distant observations, the number of available observations in

combination with selected number of points to be subsampled leads to sampling of all of those points regardless a relative sampling probability difference between them. It is more important that the sampling probability between points 10 and 100 km away differs by factor of 100. We absolutely accept the idea that sampling probability function form can take different shapes, and that it actually can account for anisotropy, and the choice presented in the paper represents just one example. There is indeed no physical justification for the forms selected, like reviewer commented, and in principle it could be replaced with exponential form. However, we do not see that it makes the conceptual presentation of the approach stronger.

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

Authors: We included two references in the Line 269 per reviewer suggestion.

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

Authors: Corrected.

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

Authors: Corrected. Thank you for the suggestion.

Reviewer: 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?

Authors: It is questionable if a comparison between 1d spatial and ST is more fair. In the case of a comparison between ST and 7d spatial we actually produce estimates using the same data, it is just that in the ST approach the temporal covariance be-

tween them is properly characterized, and in the 7d spatial it is not the case. But one might argue that it is more fair, given that we use the same observational data to produce estimates. The 1d spatial case can produce apparently better statistics because of the biases that are spatially correlated but do not reproduce in time, like we mentioned before, and thus it comes back to the question of the selection of the best error metrics, because leave-one-out cross validation yields the numbers which show the degree of regional consistency between the data, not its true accuracy. This discussion is provided in the paper at Lines 391-399. Yet unpublished results show that, based on BIC score, ST method yields SIF estimates that are more consistent with atmospheric observations of CO2 (private communication, Shiga et al., Carnegie Institution for Science).

Reviewer: 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.

Authors: We cite the commented sentence: "The method generally yields more precise and accurate (and unbiased) estimates compared to spatial method which used the same observations but assumed perfect temporal correlation between data." We believe it is clear that this sentence was meant to express that ST yield better results than S 7d (". . .compared to spatial method which used the same observations but assumed perfect temporal correlation between data. . ."). It did not mean to address 1d spatial vs ST comparison. In case of GOSAT, IASI and GOME-2 ST yielded 6, 9 and 4% lower MAE. Those values are consistent with other studies that evaluated ST vs spatial (Guo et al., 2013, Zeng et al., 2013 and 2016).

At the end, the reported statistics for GOME-2 is now slightly changed in the Table 1, as we found a small glitch in the code we used to process GOME-2 dataset. Now, the S method was found to produce better estimates than ST approach only in GOSAT 1d case, for the reason we discussed above. In all other cases, ST method was found to yield best error statistics.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-192, 2016.