Reply to reviewer 1:

We thank the reviewers for their positive assessment of the manuscript and for their helpful comments. In the text below, we include the reviewer's original comments in italics, while our responses are listed in regular font.

Reviewer: *The aims of this study are scattered in the introduction and should be clearly presented in the end of the introduction*

Authors: We condensed the goals of the study into one paragraph, now Lines 70-77.

Reviewer: *As to spatial-temporal approach, it seems some recent developments since 2011 have been missed, which should be included. Please see reference 1 for details.*

Authors: Both provided references are included now (Lines 69 and 325)

Reviewer: *Samples size is missing for the three datasets. Please provide.*

Authors: The sample was specified in Lines 100.

Reviewer: *It is not clear what software was used for this study. Please refer your readers to it so that they could apply your method to their study.*

Authors: We specified the software package in Line 77.

Reviewer: *The accuracy measures used, MAE and RMSE, are data unit/scale and variation dependent as detailed in reference 2. Please see the recommendations in this reference for accuracy measure selection*.

Authors: Per reviewer suggestion, we included two unit-independent error measures, RMAE and RRMSE in the revised version of the manuscript., Lines 326-333 and Table 1.

Reviewer: *A statistical test of the cross-validation results in Table 1 may provide more convincing evidence to show the difference between the methods compared*.

Authors: Please see our response to Review 2, to a similar comment.

Reviewer: *The conclusion: it is largely repeating what has been presented in the previous sections. It could be condensed by removing the repetitions.*

Authors: We considerably shortened the second paragraph of the Conclusion section, per reviewer request.

Minor issues:

Reviewer: *Spell out GOSAT, IASI and GOME-2 in the abstract or delete them.*

Authors: Corrected.

Reviewer: *Lines 98-99: this sentence suggests that the method is only applicable for a small region. Please revise and clarify.*

Authors: The confusing sentence has been deleted.

Reviewer: *Lines 144-145: are 'generalized product-sum model' and 'generalized product-sum covariance model' the same? Please keep the name consistent in the paper*.

Authors: Corrected.

Reviewer: *Line 173: delete one 'then'*.

Authors: Corrected.

Reviewer: *For XCO2, only 6 day data were used. Is this too short for ST method? Is it a factor for the poor performance of ST method?*

Authors: In the study by Hammerling et al., 2012 the authors examined optimal temporal aggregation time periods for XCO2 retrievals by analyzing the tradeoff between not having too much temporal variability vs. having sufficient observations in the context of spatial-only interpolation approach. They reported that 4-days temporal resolution gave the best results which points out to the fact that expected decorrelation temporal "length" of CO_2 field is at the order of magnitude of synoptic scales. Based on their analysis, 6d should not be too short.

We believe that the factor affecting "poor" performance of ST (in case of XCO_2) compared to what could be expected are different. We changed the following paragraph to make it more clear (Lines 391-399):

" The difference between the performance of ST and S-approaches obtained through cross-validation becomes most pronounced in processing datasets with measurement errors that are spatially but not temporally correlated. In these cases, an ST approach can use data from adjacent time periods to create the estimate, data that do not have the same regional, spatially-correlated biases. 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." Note that the cross-validation errors and true errors are not identical, the former is just an estimate of the latter.

The direct conclusion from this statement is that ST could perform worse in cross-validation, while in fact it filters regionally correlated measurement errors (not reproduced in time) which brings the focus back on whether the leave-one-out cross validation is the best method for validating this and similar techniques, although it has been used in a series of recent papers (Guo et al., 2013; Zeng et al., 2013, Tadic et al., 2015; Zeng et al., 2016). Please see the response to Reviewer 2. We also checked (not shown in the paper) the timeseries of estimates at selected locations where the difference between S and ST was particularly pronounced. We found that S method produced unrealistically high oscillations in estimates along the temporal axis, while ST kept estimated signal much smoother, which also supports the conclusions.

A hypothetical alternative approach to improve the apparent cross-validation performance would be to explicitly model the retrieval error covariance matrix, instead of assuming the independence of retrieval errors, or, in other words, to isolate measurement clusters having regionally correlated errors. However, such information is usually not available. Interestingly, the very difference in performance between ST vs. S could be used to address this important, but still not fully resolved issue.

Reference: Hammerling, D. M., A. M. Michalak, and S. R. Kawa (2012), Mapping of CO_2 at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2, J. Geophys. Res., 117, D06306, doi:10.1029/2011JD017015.

Reviewer: Lines 345-346: ST method seems not that poor for GOME-2 data. Please revise.

Authors: Corrected.

References:

Guo, L., Lei, L. and Zeng, Z.: Spatiotemporal correlation analysis of satellite-observed CO_2: Case studies in China and USA. Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International, 21-26 July, Melbourne, VIC, 2013.

Zeng, Z., LiPing, L., L. LiJie, G., Li, Z., Bing, Z.,: Incorporating temporal variability to improve geostatistical analysis of satellite-observed CO_2 in China, Chinese Science Bulletin, 58(16), 1948-1954, 2013.

Zeng, Z.; Lei, L.; Hou, S.; Ru, F.; Guan, X.; Zhang, B., A regional gap-filling method based on spatiotemporal variogram model of columns, IEEE Transactions on Geoscience and Remote Sensing, 2014, 52, 3594-3603.

Guo, L., Lei, L., Zeng, Z.C., Zou, P., Liu, D. and Zhang, B., 2015. Evaluation of spatio-temporal variogram models for mapping Xco 2 using satellite observations: A case study in china. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(1), pp.376-385.

Tadić, J., Qiu, X., Yadav, V. and Michalak, A.: Mapping of satellite Earth observations using moving window block kriging, Geosci. Model Dev., 8, 1–9, doi:10.5194/gmd-8-1-2015, 2015.

Guo, L., Lei, L., Zeng, Z.C., Zou, P., Liu, D. and Zhang, B., 2015. Evaluation of spatio-temporal variogram models for mapping Xco 2 using satellite observations: A case study in china. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(1), pp.376-385.

Zeng, Z., Lei, L., Strong, K., Jones, D. B. A., Guo, L., Liu, ., Deng, F., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Henderson, B., Kivi, R., Lindenmaier, R., Morino, I., Notholt, J., Ohyama,H., Petri, C., Sussmann, R., Velazco, V., A., Wennberg, P., O., and Lin, H.: Global land mapping of satellite-observed CO_2 total columns using spatio-temporal geostatistics, International Journal of Digital Earth, DOI: 10.1080/17538947.2016.1156777, 2016.

Reply to reviewer 2:

Reviewer: *I will be upfront and say that 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.

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).

Reference: 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---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.

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---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.*

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 $1/ht^2$ 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 $1/h_t^n$ 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---temporal (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 between 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 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.

Spatio-temporal approach to moving window block kriging of satellite data

Jovan M. Tadić[1], Xuemei Qiu[1], Scot Miller[1] and Anna M. Michalak[1]

[1]{Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA}

Abstract. Numerous existing satellites observe physical or environmental properties of the Earth system. Many of these satellites provide global-scale observations, but these observations are often sparse and noisy. By contrast, contiguous, global maps are often most useful to the scientific community (i.e., level 3 products). We develop a spatiotemporal moving window block kriging method to create contiguous maps from sparse and/or noisy satellite observations. This approach exhibits several advantages over existing methods: 1) it allows for flexibility in setting the spatial resolution of the level 3 map, 2) it is applicable to observations with variable density, 3) it produces a rigorous uncertainty estimate, 4) it exploits both spatial and temporal correlations in the data, and 5) it facilitates estimation in real time. Moreover, this approach only requires ~~a limited number of assumptions –~~the assumption that the observable quantity exhibits spatial and temporal correlations that are inferable from the data. We test this method by creating Level 3 products from satellite observations of CO_2 (XCO_2) from the Greenhouse Gases Observing Satellite (GOSAT~~,~~), CH_4 (XCH_4) from the Infrared Atmospheric Sounding Interferometer (IASI) and solar-induced chlorophyll fluorescence (SIF) from the Global Ozone Monitoring Experiment–2 (GOME-2~~,~~). We evaluate and analyze the difference in performance of spatio-temporal vs. recently developed spatial kriging methods.

1. Introduction

Satellite observations of the Earth's surface and atmosphere provide a valuable window into the functioning of the Earth system. Satellites often provide global observations, but these observations are rarely uniform or contiguous in space/time. The observations can be non-contiguous due to satellite orbit geometries and periods, geophysical limitations (e.g. cloud cover), and temporary instrument malfunctions. Furthermore, satellites may provide a large quantity of data, but individual observations can have a large noise-to-signal ratio. It is often necessary to spatially interpolate the data in order to organize the data onto a regular grid, query the data at a particular location of interest, estimate data at unsampled times and/or locations, and/or map the underlying signal in a noisy dataset. These gridded, interpolated maps are commonly named "Level 3" data (e.g. NASA, 2014) and are often part of the standard suite of satellite data products.

CO_2 column observations (XCO_2) from the Greenhouse Gases Observing Satellite (GOSAT), CH_4 column observations (XCH_4) from the Infrared Atmospheric Sounding Interferometer (IASI) and solar-induced chlorophyll fluorescence (SIF) observations from The Global Ozone Monitoring Experiment–2 (GOME-2) provide prototypical examples of these challenges, and these three satellites are the primary application used throughout this work (see Section 3).

The most commonly-used method for creating Level 3 maps from satellite data is binning. This approach involves taking the mean of all observations within a given grid cell or "bin" (see Kulawik et al., 2010, and Crévoisier et al., 2009 for examples). The binning method, however, has a number of shortfalls that

can lead to inconsistent or inaccurate results. First, different bins contain variable numbers of observations. As a result, some bins will be well-constrained by the data while others may be based upon sparse, noisy observations. Second, binning does not produce uncertainty estimates. Third, this method cannot extrapolate the unknown quantity to bins without any observations.

A broad class of geostatistical methods known as kriging provides an alternative approach to mapping satellite observations. Kriging is a best linear unbiased estimator (for kriging see Chiles and Delfiner, 2012), where covariance functions are used to represent correlations among data. As a result, kriging can account for a variable density of observations and can estimate uncertainties in the resulting maps. Various forms of kriging have recently been used to map satellite Earth observations, particularly for XCO_2 (e.g., Hammerling et al. 2012a,b; Tadić et al., 2015; Zeng et al., 2013; Guo et al., 2013, Zeng et al., 2016). Hammerling et al. (2012a,b) presented an approach to map Orbiting Carbon Observatory-2 (OCO-2) and GOSAT XCO_2 observations, respectively, with non-stationary properties. In our previous study (Tadić et al., 2015) we extended that approach to create XCO_2 maps that can have a different spatial resolution from the resolution or footprint of the original satellite observations. Our previous study and those of Hammerling et al. (2012a,b) accounted for spatial covariances among observations but did not include a temporal component. The present study extends this geostatistical framework from a purely spatial to a spatiotemporal domain.

Spatiotemporal approaches to interpolation can provide a number of advantages relative to purely spatial methods (e.g. Zeng et al., 2016; Guo et al., 2013). A purely spatial approach will usually aggregate observations into temporal blocks; observations within the same block effectively have the same time stamp whether or not those observations are actually synchronous (e.g., Tadić et al., 2015; Hammerling et al., 2012a,b). Any real temporal variability within a block becomes noise. A spatiotemporal approach, by contrast, treats time as an explicit dimension and models covariances among data as a function of time. ~~As a result, the spatiotemporal approach can (1) fill in temporal gaps in the observations, (2) create maps at higher temporal resolutions than purely spatial approach, (3) produce more accurate estimates when observations have variable spatio-temporal coverage, (4) predict future values (i.e. extrapolate temporally).~~

A handful of recent studies have considered temporal relationships when mapping satellite observations of XCO_2. These studies have either used various forms of Kalman smoothing (e.g., Katzfuss and Cressie 2011, Katzfuss and Cressie 2012, Nguyen et al. 2014) or geostatistics (e.g., Guo et al. 2013; Zeng et al. 2013; Zeng et al. 2016). The former group of studies leverages Kalman smoothing to improve the computational tractability of mapping dense or abundant datasets, like OCO-2 and the Atmospheric Infrared Sounder (AIRS). The latter group of studies, by contrast, has applied geostatistics to sparse datasets like those from the GOSAT satellite. A detailed review of spatial and spatio-temporal mapping methods has been published recently (Li and Heap, 2014).

The ~~model developed in~~ goal of this ~~paper also uses geostatistics~~study is to ~~map~~develop a geostatistical spatio-temporal mapping and upscaling method (applicable, but not limited to, satellite observations of XCO_2~~, but we present~~) that exhibits several advances relative to previous ~~efforts.~~methods. It can: (1) fill in temporal gaps in the observations, (2) create maps at higher temporal resolutions than purely spatial approach, (3) produce more accurate estimates when observations have variable spatio-temporal coverage, (4) predict future values (i.e. extrapolate temporally). Among other improvements, we develop an efficient method ~~to subsample~~for subsampling satellite observations and utilize the product-sum covariance model (e.g., De Iaco et al., 2001) that is easy to parameterize, which makes it applicable to both ~~abundant~~dense and sparse datasets. The entire work has been conducted in Matlab 2012a.

Section 2 of this study describes the presented model in detail; it describes an efficient subsampling procedure that can handle very large datasets and a covariance model that can estimate both spatial and

temporal relationships in the data. We then incorporate these components into a spatiotemporal version of moving window block kriging. In sections 3 and 4, we subsequently apply this model to map GOSAT XCO_2, IASI XCH_4 and GOME-2 SIF at multiple time resolutions (including daily).

2. Methods

The spatio-temporal block kriging approach presented in this study proceeds in three steps for each model grid cell and estimation time. First, we subsample the observations within a predetermined spatio-temporal domain (section 2.1). Next, we characterize the local spatio-temporal covariance structure (section 2.2). Finally, we interpolate the satellite observations at the desired spatial resolution (section 2.3).

2.1 Subsampling of observations

The ultimate goal of the proposed subsampling strategy is to reduce the number of observations in the spatio-temporal vicinity of an estimation location to a representative, computationally feasible subset of data. We use a subset of observations (M) to estimate a local set of covariance parameters and use another subset (N) to estimate the desired quantity and associated uncertainty. Note that, for the method presented here, M and N can refer to either the same subset of data or different subsets.

The total number of observations used for covariance parameter estimation (M), is selected to be small enough to make this estimation computationally feasible but large enough to yield a sample representative of both local and regional variability. The optimal subset of N observations used for mapping depends on the actually observed covariance structure which is not known prior to covariance parametrization step. In the example presented in Sect. 3, the optimal observational subset used in a mapping step for each grid cell comprised N points having the highest covariance with the estimation location. In the example below, we set both M and N at 500; larger values of M and N did not have a substantial impact on the estimated parameters and mapped quantity, respectively. ~~Furthermore, M should represent local variability, and larger values of M would encompass more distant, non-local regions.~~

We select subset of observations M for each estimation grid cell by assigning a relative selection probability to each observation based on that observation's spatial and temporal 'separation distances' from the centroid of the grid cell. In the absence of a proper metric for distance in space-time, we model the spatial and temporal components of the overall selection probability separately.

The selection probability (and its components) is described by the following equation:

$$P = P_s \times P_t \propto 1/(A_s h_s)^2 \times e^{-(A_t h_t)^2} \quad (1)$$

where P_s is the spatial component of the relative probability of a given observation being selected, P_t is temporal component, h_s and h_t are distances between estimation location and observations, in space and time, respectively, and A_s and A_t are unit dependent, user defined weighting factors between separation distance in space vs. in time (how deep in space vs. time the sampling should occur). The unit dependent choice of A_s and A_t can be initially based on user expectations of the decorrelation distances in space vs. time and, if necessary, subsequently corrected accounting for actually computed decorrelation lengths in space and time in an iterative fashion. In this way temporal and spatial sampling depths could even be locally optimized and become location-specific. In the examples below (Section 3), A_s and A_t were set to

1 km^{-1}, and 0.5 day^{-1}, respectively, based on the observed average decorrelation distances in space and time (see Fig. 1 and Section 4.1).

[Figure 1]

h_s is calculated as the great circle distance between the centroid x_j of the estimation grid cell and the location x_i of an observation:

$$h_s(x_i, x_j) = r\cos^{-1}(\sin\varphi_i \sin\varphi_j + \cos\varphi_i \cos\varphi_j \cos(\lambda_i - \lambda_j)) \quad (2)$$

where φ_i and λ_i are the latitude and longitude of location x_i and r is the radius of the Earth.

The temporal and spatial components of the probability function have different functional forms out of necessity. The measurements often come pre-aggregated in time slices corresponding to hours, days, or longer aggregation time periods, which multiplies the number of observations with the same time stamp. As a result, it is not possible to assign sampling probability along a temporal axis in a manner equivalent to the spatial approach; doing so would result in infinite probabilities assigned to all observations within the time slice of the actual estimation location ($P_t \sim 1/0^2 = \infty$). The same holds for spatially co-located observations. However, since each observation comes with unique spatial coordinates (not pre-binned like in temporal case), we select a simpler form of the spatial component of the sampling function. The defined form of P (Eq. 1) ensures that pairs of observations close to estimation location define the shape of the variogram at short separation distances (the variogram should reflect variability in the spatio-temporal vicinity of the estimation grid cell. See Section 2.2). Different forms of P can be used if directional anisotropy is expected or if more/fewer observations along a given direction are desired to better represent expected correlations.

Previous approaches required the user to choose spatial and temporal windows that determine which neighboring observations to use (see, for comparison, Alkhaled et al. 2008; Hammerling et al. 2012a,b). The approach proposed in this paper, by contrast, requires fewer subjective choices – only the form of sampling function and unit dependent choice of normalizing coefficients A_s and A_t. In addition, our approach is computationally feasible even for very large data sets.

2.2 Characterization of Spatio-temporal Covariance

Existing studies have used a number of models to estimate spatio-temporal covariances for a variety of applications. Models used include the metric model (Dimitrakopoulos and Luo, 1994), linear model (Rouhani and Hall, 1989), product model (De Cesare et al., 1996), non-separable model (Cressie and Huang, 1999), and generalized product-sum covariance model (De Iaco et al., 2001). The approach developed in this paper uses a generalized product-sum covariance model (De Iaco et al., 2001). This model affords a number of advantages relative to other covariance models: (1) a product sum covariance model outperformed other models in terms of prediction accuracy in a recent study using GOSAT satellite data (Guo et al., 2013), (2) it is relatively easy to implement (De Iaco et al., 2001), and (3) it is more flexible than a non-separable covariance model (De Cesare, 2001a).

The product-sum model, as it has been applied in the past, has one important area for improvement. The original procedure (De Iaco et al., 2001) assumed separate modeling of the spatial and temporal covariance (variograms) and their later unification into a spatio-temporal model in the final step. The procedure requires observations approximately in the same location at multiple different times. However, satellite observations are often not perfectly collocated in consequent measurement cycles over the same region. As a result, we would need to assume that each measurement cycle is perfectly co-located with

previous/future cycles, or define an arbitrary tolerance, in order to apply the original approach. This assumption becomes more prone to error if the observations are very sparse, as is often the case with satellites.

Thus, in this study, we cater to specific properties of satellite data and alter the original procedure by estimating all covariance parameters simultaneously, thereby avoiding the aforementioned problem.

We broadly define the covariance as follows:

$$C_{s,t}(h_s,h_t) = \mathrm{Cov}(Z(s_+h_s,t_+h_t), Z(s,t)) \quad (3)$$

The equation shows that covariance between two points (Z) separated in space-time (s,t) depends on their distance in space (h_s) and distance in time (h_t). The following class of valid product–sum covariance models was introduced in De Cesare et al. (2001b) and further developed in De Iaco et al. (2001):

$$C_{s,t}(h_s,h_t) = k_1 C_s(h_s) C_t(h_t) + k_2 C_s(h_s) + k_3 C_t(h_t) \quad (4)$$

where C_t and C_s are valid temporal and spatial covariance models, respectively. De Iaco et al. (2001) proved that for positive definiteness it is sufficient that $k_1 > 0$, $k_2 \geq 0$ and $k_3 \geq 0$. It is interesting to note that from Eq. 4 follows that spatio-temporal covariance models collapses down to purely spatial model in cases where temporal covariance does not exist. Thus, the spatial approach could be viewed as a special case of spatio-temporal modeling.

The model in Eq. 4 corresponds to the spatio-temporal variogram shown in Equation 5. In the original procedure, De Iaco et al., 2001 estimated separate spatial (h_t=0) and temporal (h_s=0) variograms using the data. De Iaco et al., 2001 ~~then~~ then combined these models to obtain the final spatio-temporal variogram model:

$$\gamma_{s,t}(h_s,h_t) = \gamma_{s,t}(h_s,0) + \gamma_{s,t}(0,h_t) - k\gamma_{s,t}(h_s,0)\gamma_{s,t}(0,h_t) \quad (5)$$

where $\gamma_{s,t}(h_s,0)$ and $\gamma_{s,t}(0,h_t)$ are spatio-temporal variograms for h_t=0 and h_s=0, respectively (Figure 2). Parameter k is estimated from the data which makes the model easily applicable:

$$k = \frac{k_s C_s(0) + k_t C_t(0) - C_{s,t}(0,0)}{k_s C_s(0) k_t C_t(0)} \quad (6)$$

where $k_s C_s(0)$ and $k_t C_t(0)$ are spatial and temporal sills (variances) obtained in modeling of separate spatial and temporal variograms. The only condition k has to fulfill in order to create an admissible covariance model is

$$0 < k \leq \frac{1}{max\{\sigma_s^2\left(\gamma_{s,t}(h_s,0)\right); \sigma_t^2\left(\gamma_{s,t}(0,h_t)\right)\}} \quad (7)$$

Due to the specifics of satellite data, we estimate both the covariance parameters and parameter k simultaneously. This approach accounts for constraints that assure a positive definiteness of the model (De Iaco et al., 2001). This simultaneous approach makes the model more applicable to sparse data and data with variable spatial coverage, as is often the case with satellite observations.

We use a Gaussian variogram function with a nugget effect to model temporal covariance in the example presented here (for an overview of variogram models see Chiles and Delfiner, 2012). We use an

exponential model for the spatial variogram. In both cases, we make this choice based upon visual inspection of local variograms at multiple estimation locations:

$$\gamma_t(h_t)(Gaussian) = \begin{cases} 0, \ for\ h_t = 0 \\ {\sigma_t}^2(1 - \exp\left(-\frac{{h_t}^2}{{l_t}^2}\right) + \sigma_{nug}^2, \ for\ h_t > 0 \end{cases} \quad (8)$$

$$\gamma_s(h_s)(exponential) = \begin{cases} 0, \ for\ h_S = 0 \\ {\sigma_s}^2(1 - \exp\left(-\frac{h_S}{l_S}\right) + \sigma_{nug}^2, \ for\ h_S > 0 \end{cases} \quad (9)$$

where σ^2 and l are the variance and correlation length of the quantity being mapped, and σ^2_{nug} is the nugget variance, typically representative of measurement and retrieval errors in the case of satellite observations.

[Figure 2]

Unlike the original procedure in De Iaco et al. (2001), we model the variogram using only two steps. First, we calculate a raw spatio-temporal variogram based on the subsampled observations for each estimation grid cell:

$$\gamma(h_s, h_t) = \frac{1}{2}[y(x_i) - y(x_j)]^2 \quad (10)$$

where γ is the raw spatio-temporal variogram value for a given pair of observations $y(x_i)$ and $y(x_j)$, and h_s and h_t are, respectively, the great circle distance and temporal distance between the spatio-temporal locations (x_i and x_j) of these observations.

Second, we fit the theoretical variogram defined in Eq. 5 to the raw variogram using non-linear least squares. We subsequently calculate the spatiotemporal covariance using the following equation:

$$C_{s,t}(h_s, h_t) = C_{s,t}(0,0) - \gamma_{s,t}(h_s, h_t)) \quad (11)$$

Validity on the sphere. Most covariance models were originally designed for Euclidean space, and their validity in other coordinate systems cannot be assumed *per se*. Huang et al. (2011) examined the validity of several theoretical covariance models in spherical coordinate systems. However, this evaluation has not been done for the spatio-temporal product-sum covariance model. Other studies that use a product-sum covariance model typically assume the validity of this covariance model on a sphere (e.g., Zeng et al., 2013; Zeng et al., 2016). Results from Huang et al. (2011) explicitly validate the exponential covariance model on a sphere, as well as sums of the products of exponential covariance models and constants (provided that the constants are positive). The first term of the product-sum covariance model used in this study (Eq. 4) represents a Hadamard product (Million, 2007) of two positive definite matrices. According to Schur product theorem, a Hadamard product of two positive definite matrices necessarily gives a positive definite matrix (Mathias, 1993). It therefore follows that a generalized product-sum model (Equation 4) is valid on a sphere if its spatial component is valid on a sphere.

2.3 Mapping using spatio-temporal moving window block kriging

This section leverages the sampling function (Sect. 2.1) and the product-sum covariance model (Sect. 2.2) to implement a spatio-temporal version of moving window block kriging. A primary advantage of block kriging is its ability to estimate contiguous maps at any spatial resolution equal to or coarser than the

spatial support (i.e. footprint size) of observations (refer to Sect. 1 and Tadić et. al. 2015). Unlike ordinary kriging method, the spatial support in block kriging corresponds to the average value within each chosen grid cell.

Moving window block kriging requires solving a set of linear equations to obtain a set of weights (λ). These weights must be estimated for each prediction location using N associated observations:

$$\begin{bmatrix} \mathbf{Q}+\mathbf{R} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\nu \end{bmatrix} = \begin{bmatrix} \mathbf{q_A} \\ 1 \end{bmatrix} \quad (12)$$

In this equation, **R** is a diagonal $N \times N$ nugget covariance matrix that describes measurement and retrieval errors, **Q** is a $N \times N$ covariance matrix among the N observations with individual entries as defined in Eqn. 11, **1** is an $N \times 1$ unity vector, ν is a Lagrange multiplier, and $\mathbf{q}_A$ is an $N \times 1$ vector of the spatio-temporal covariances between the N observation locations and the estimation grid cell, defined as:

$$q_{A,i} = \frac{1}{n}\sum_{j=1}^{n} q\left(h_{s_{i,j}}, h_{t_{i,j}}\right) \quad (13)$$

where $q_{A,i}$ is the covariance between the grid cell and observation i. $q(h_{i,j})$ is defined as $C_{s,t}$ in Eqn. 11 based on the distances $h_{s_{i,j}}$ and $h_{t_{i,j}}$ between observation i and n regularly-spaced locations within the grid cell. In the context of satellite measurements, n is a highest number of non-overlapping footprints contained within a grid cell and was calculated based on the relative size of the satellite footprint compared to the size of the estimation grid cells. n varies with latitude, as the size of grid cells decreases with the distance from the equator. The system in Eqn. 12 is solved for the weights (**λ**) and the Lagrange multiplier (ν). We subsequently use these parameters to define the estimate ($\hat{z}$) and estimation uncertainty ($\sigma^2_{\hat{z}}$) for the grid cell:

$$\hat{z} = \boldsymbol{\lambda}^T \mathbf{y} \quad (14)$$

$$\sigma^2_{\hat{z}} = \sigma_{AA} - \boldsymbol{\lambda}^T \mathbf{q}_A + \nu \quad (15)$$

where **y** is the $N \times 1$ vector of subsampled observations, and σ_{AA} is the variance of the observations at the resolution of the estimation grid cell, defined as:

$$\sigma_{AA} = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} q(h_{j,k}) \quad (16)$$

In that equation, $q\left(h_{s_{i,j}}, h_{t_{i,j}}\right)$ is defined as $C_{s,t}$ in Eqn. 11 based on the distances $h_{s_{i,j}}$ and $h_{t_{i,j}}$ between any combination of the n regularly spaced locations within the grid cell defined previously.

3. Example applications

We select three case studies of satellite Level 2 data to demonstrate the properties of the method developed in this paper: column-integrated dry air model fraction of CO_2 (XCO_2) from the Japanese Greenhouse Gas Observing SATellite (GOSAT), CH_4 (XCH_4) from the Infrared Atmospheric Sounding Interferometer (IASI), and solar-induced fluorescence (SIF) the Global Ozone Monitoring Experiment–2 (GOME-2). Level 2 datasets from GOSAT, IASI and GOME-2 have relatively different characteristics.

For example, GOSAT observations are sparse while IASI and GOME-2 are abundant. These diverse datasets are therefore ideal for testing the method developed here.

The method was demonstrated by producing two different sets of maps. First, it was applied at resolutions coarser than native (1 × 1°, 2.5 × 2°, and 1 × 1° for GOSAT, IASI and GOME-2, respectively) to demonstrate block kriging capabilities of the method (Section 3). Second, it was applied at the native resolution of the satellites for cross-validation (method evaluation) purposes (Section 4).

3.1 Total column CO_2 (XCO_2) observed by GOSAT

The Japanese Greenhouse Gas Observing SATellite (GOSAT) (e.g., Kuze et al., 2009), the first satellite dedicated to global greenhouse gas monitoring, was launched in 2009. Basic information about the satellite, its orbit configuration, and the CO_2 column observations are given in our previous study (Tadić et al., 2014). It flies in a polar, sun-synchronous orbit with a 3-day repeat cycle and an approximate 13:00 LT overpass time. GOSAT has a nadir footprint of about 10.5 km diameter at sea level (Kuze et al., 2009) and 2×10^3 observations per week. The XCO_2 observations from GOSAT have large retrieval uncertainties (e.g., O'Dell et al. 2012) and exhibit large spatial and temporal gaps (e.g., Fig. 3a). Although these XCO_2 observations are sparse and noisy, contiguous Level 3 maps are often desirable for environmental and ecological applications~~.~~ (Maksyutov et al., 2013; Liu et al., 2012). To this end, we generate global daily estimates for XCO_2 (August 2-7, 2009) to match the timeframe used in Tadić et al., 2014.

[Figure 3]

We obtain bias-corrected and filtered GOSAT Level 2 observations using NASA's Atmospheric CO_2 Observations from Space (ACOS) algorithm v3.4 release 3 (e.g., O'Dell et al., 2012; Crisp et al., 2012). In this study, we use spatio-temporal moving window block kriging to create a series of contiguous, in-filled global daily maps and associated uncertainties for 2-7 August 2009 (two repeat cycles) (Fig. 3a-c) at 1×1^o resolution. We select the time period to match the time period from our previous study (Tadić et al., 2014). Unlike results from our previous study and other similar studies, which created estimates at 6-day or longer time periods (Hammerling et al., 2012a), we leverage the method developed here to produce maps at the daily scale.

3.2 Total column CH_4 (XCH_4) observed by IASI

The Infrared Atmospheric Sounding Interferometer (IASI) developed by the Centre National d'Etudes Spatiales (CNES) in collaboration with the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) is a Fourier Transform Spectrometer based on a Michelson Interferometer coupled to an integrated imaging system that measures infrared radiation emitted from the Earth. It is carried by MetOp-A, a sun-synchronous polar orbit satellite which flows at an altitude of 817 km. Detailed information about the IASI instrument could be found elsewhere (Crévoisier et al., 2009a,b; Massart et al., 2014). IASI has an instantaneous field of view of 50×50 km, composed of four pixels each 12 km in radius, delivering ~56×10^3 XCH_4 observations per week.

[Figure 4]

Methane Level 2 IASI (0-4 km) data were retrieved at the NOAA/NESDIS using the NUCAPS (NOAA Unique CrIS/ATMS Processing System) algorithm (Gambacorta, 2013; Xiong et al., 2013). For the ice-covered ocean the data for the lower troposphere (0-4 km) are unreliable due to insufficient thermal contrast between the surface and the atmosphere. Filtering parameters have been provided by Xiong (2014, private communication). The data are available at http://www.nsof.class.noaa.gov/. Using the new method, we created a series of contiguous global daily maps and associated uncertainties for the Northern Hemisphere, for February 26-March 4, 2013 (i.e. Figure 4a-c) at $1^{o}\times1^{o}$ resolution. We chose this time period to match the occurrence of the methane "anomaly" North of the coast of Scandinavia.

3.3 Global land solar-induced fluorescence fields observed by GOME-2

The GOME-2 (The Global Ozone Monitoring Experiment–2) instrument on board METOP-A (e.g., Joiner et al., 2013) observes solar-induced fluorescence (SIF). The GOME-2 spatial footprint (i.e. support) of the observations is 40 km × 80 km (Joiner et al, 2013), and the volume of available data is approximately 2×10^{5} SIF observations per week.

[Figure 5]

Multiple recent studies have demonstrated the potential use of satellite observations of solar-induced fluorescence (SIF) for understanding the photosynthetic CO_2 uptake at large scales (Joiner et al., 2011; Joiner et al., 2012; Joiner et al., 2013; Frankenberg et al., 2011; Frankenberg et al., 2012; Guanter et al., 2012, Lee et al., 2013; Frankenberg et al., 2014). Satellite SIF measurements can be used with land surface models to understand GPP response to environmental stress (e.g., Lee et al., 2013) and to improve the representation of GPP. GOME-2 provides the highest spatial and temporal density of data, among all available datasets.

In the example presented here we use SIF GOME-2 v.14 data (Joiner et al., 2013) with the approach described in Section 2 to create contiguous maps of SIF at a single spatial resolution ($1^{o} \times 1^{o}$) and daily temporal resolutions. Maps of SIF and associated uncertainties are created at daily temporal resolutions covering 5-14 May, 2012, some of which are shown on Figures 5a-c.

4. Method evaluation: accuracy, precision and bias

4.1 Accuracy, precision and bias

We use a leave-one-out cross validation technique to assess the performance of spatio-temporal (ST) versus spatial moving window block kriging. We produce these estimates at the native resolution of GOSAT, IASI and GOME-2 satellites/instruments, which allowed a direct comparison to measured values. For IASI and GOME-2, for each day in February 26-March 4, 2013, and May 5-14, 2012, respectively, 10% of available observational data were randomly selected for use in leave-one-out cross-validation and their coordinates extracted. For XCO_2, all GOSAT XCO_2 observations for each day in August 2-7, 2009, were used. We assess the accuracy (the difference between estimates and withheld observations) of both methods using two common measures: (1) Mean Absolute Error (MAE), and (2) Root Mean Squared Error (RMSE). ~~We also~~We also use two more recently proposed measures (Li and

Heap, 2011; Li, 2016) that remove the effect of unit/scal. The first is relative mean absolute error (RMAE) that is given as:

$$\mathrm{RMAE} = \frac{1}{n}\sum_{i=1}^{n}|(\hat{z}_i - y_i)/o_i| \times 100 \quad (17)$$

and the second is relative root mean square error (RRMSE), as follows:

$$\mathrm{RRMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{z}_i|/y_i)^2\right]^{1/2} \times 100 \quad (18)$$

where n is the number of observations or samples, o is observed value, and p is predicted or estimated value.

We assess the performance of each method using two additional measures: (3) the accuracy of the uncertainty bounds (the degree to which the reported uncertainties capture the difference between estimates and withheld observations) and (4) bias (the mean difference between estimates and withheld observations).

We parameterize the temporal component of the spatio-temporal sampling function in such a way that observations located +/- 3 days from the actual date had 10% probability of being sampled compared to observations from the actual day (see Fig 1a). We compare the results to spatial kriging estimates obtained in two different ways, based on observations only from the actual day (1d) and based on observations from +/-3 days from the actual day (7d). This latter case is analogous to the +/- 3-day window that we use for the ST approach. In this 7d case, we obtain these spatial kriging results by assuming the entire observational dataset collected within the selected time period (actual day +/- 3 days) is perfectly temporally correlated. In other words, we use all observations as though they were collected at the same time. We then produce estimates at locations of observations collected within the selected timeframe and compare the performance of the two methods. We repeat procedure described in Section 2 for every observation selected for cross-validation, and we average the statistics, displayed in Table 1.

[Table 1]

According to the results, the spatio-temporal approach performs better than the spatial (7d) approach in all three cases and in all performance measures (for example, spatial (7d) MAE was 6-10% larger). The comparison clearly shows that proper characterization of the temporal covariance between two points residing in different time periods (days), embedded into spatio-temporal approach, improves kriging performance. In IASI ~~case~~and GOME-2 cases, the spatio-temporal method also performed better than spatial (1d). However, in case of GOSAT~~ and GOME-2~~ data, spatio-temporal approach slightly underperformed the spatial (1d) approach having 12% higher MAE (please see Section 4.2 for discussion).

We observed that RMAE and RRMSE error measures should be used with caution in cases when observations can take real zero values, like in the GOME-2 case. In such cases the division by close-to-

zero values result in extremely high RMAE and RRMSE values, which overall limits the applicability of these error measures.

We evaluate the accuracy of the uncertainty bounds by examining how often those bounds encapsulate withheld observations. The percentage of observations that fall outside the uncertainty bounds in spatio-temporal approach is comparable to that of the spatial method, confirming the accuracy of the estimated uncertainty bounds (for normally-distributed data the percentage of observations that fall outside of the one, two, and three ~~estimation~~estimations standard deviation ($\sigma_{\hat{z}}$) uncertainty bounds should be 32%, 5% and 0.3%, respectively). The fraction of observations that fall outside the uncertainty bound is generally lower than would be expected for normally-distributed data, and our results may indicate non-normal features in the data.

4.2 When is spatio-temporal modeling recommended?

A ST approach can afford advantages over purely spatial methods when temporal data correlations and data coverage are strong. Indeed, in many cases, the ST approach is more accurate than a purely spatial method (Table 1). This result is consistent with existing literature which uniformly reports that ST approaches are more accurate than spatial approaches (Zeng et al., 2013; Guo et al., 2013; Zeng et al., 2016).

However, although considering information from days preceding and following the target estimation day should in principle always provide a further constraint on the estimate, this does not guarantee that an ST method will always outperform a spatial-only method in practice. The prime reasons for this are two-fold. First, because computational limitations cap the number of observations that can be considered, considering observations across multiple days necessarily leads to a reduction in the spatial density of observations being considered. This first factor can be partially alleviated by carefully designing the selection probability function (Eqn. 1). The second reason is that implementing a ST approach involves the estimation of a larger number of covariance parameters (Eqn. 4-9) relative to a spatial-only approach, which can introduce additional uncertainty. Indeed, we observe that the purely spatial approach performs better than the ST method in some cases (e.g., the GOSAT ~~and GOME-2 1d cases).~~case).

Overall, a ST approach is likely to outperform a spatial-only approach when the data exhibit one (or more) of three characteristics. First, a ST approach is likely better when the data are sparse or unequally distributed. In these cases, a ST approach can intelligently leverage data in adjacent time periods to compensate for the sparsity of data in the time period of interest. Second, an ST approach works well for datasets with temporal gaps (e.g., due to cloud cover or instrument malfunction). An ST approach can fill these gaps while a spatial-only approach cannot be used for temporal gap-filling. Third, an ST-approach is well-suited to datasets with regional biases that manifest in one time slice but that do not repeat in adjacent time slices. ~~Phrased differently, an ST approach is well suited to datasets with~~The difference between the performance of ST and S-approaches obtained through cross-validation becomes most pronounced in processing datasets with measurement errors that are spatially but not temporally correlated. In these cases, an ST approach can use data from adjacent time periods to ~~create the~~obtain an estimate, data that do not have the same regional, spatially-correlated biases. 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. A spatial-only approach, by contrast, will reproduce these regional biases because it does not use data from adjacent times when creating the estimate. As a result, a spatial-only approach will appear to perform better in cross validation, but the ST approach will more accurately reflect the true, underlying process.

5. Conclusions

In this study, we develop a method to create high spatio-temporal resolution maps from satellite data using spatio-temporal moving window block kriging based on product-sum covariance model. The method relies on a limited number of assumptions: that the observed physical quantity is spatio-temporally auto-correlated, and that its nature can be inferred from the observations.

The method has several advantages over previously applied methods~~, as alluded to in Sect. 1: 1) it allows for the creation of contiguous maps at varying spatio-temporal resolution, 2) it can create maps at temporal resolutions shorter than achievable by other binning or kriging methods, 3) it can be applied for creating contiguous maps for physical quantities with varying spatio-temporal coverage (i.e., density of measurements), 4) it provides assessments of the uncertainty of interpolated values, 5) it utilizes all spatio-temporally available information to generate estimates, 6~~. Apart from the advances alluded to in Sect. 1: 1) it improves covariance parameters estimation procedure because it does not model spatial and temporal covariance separately, ~~7~~2) it allows for great flexibility in the choice of sampling function and ~~8~~3) it provides estimates even for the time periods when measurements are not available. It can exploit correlations with both past and future periods of the observed time spot to provide the most accurate estimates.

We demonstrate the applicability of this method by creating Level 3 products from the GOSAT XCO_2, IASI CH_4 and GOME-2 SIF data. Sparse XCO_2 observations from GOSAT and dense XCH_4 and SIF observations from IASI and GOME-2 make a perfect test ground for the method. We show that the proposed method can even map XCO_2 on daily time scales. 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. The factors which could affect the performance of the ST method are discussed in Section 4.2.

This approach could be used in the future to produce real-time estimates not only of XCO_2, XCH_4 or SIF, but of other environmental data observed by satellites which exhibit spatio-temporal autocorrelations. Especially important could be satellite datasets that have spatially, but not temporally, correlated errors. In such cases, sampling across several time periods could perhaps help isolate and remove them, which should be a subject of further studies.

The method could be applied in a standalone mode or as part of a broader satellite data processing package. Maps produced by the spatio-temporal approach could then be incorporated into physical and biogeochemical models of the Earth system.

6. Code availability

The documented Matlab source code is available at the Researchgate website (https://www.researchgate.net/publication/311595272_Spatio-temporal_approach_to_moving_window_block_kriging_of_satellite_data_v10_code; DOI: 10.13140/RG.2.2.21411.04643).

Acknowledgement

This work was supported by the National Aeronautics and Space Administration (NASA) through grant no. NNX12AB90G and NNX13AC48G, and the National Science Foundation (NSF) through grant no. 1342076. Satellite CH_4 IASI v5 data are supplied by the NOAA National Environmental Satellite, Data,

and Information Service (NESDIS): http://www.nsof.class.noaa.gov/. We would also like to thank Leonid Yurganov (JCET) and Nathaniel Lebedda (University of Maryland) for helpful information and discussions.

References

Chiles, J.-P. and Delfiner, P.: Geostatistics, second edition, Wiley, 2012.

Cressie, N. and Huang, H. C.: Classes of nonseperable, spatio-temporal stationary covariance functions, Journal of the American Statistical Association, 94, 1 – 53, 1999.

Crévoisier, C., Nobileau, D., Fiore, A. M., Armante, R., Chédin, A., and Scott, N. A.: Tropospheric methane in the tropics – first year from IASI hyperspectral infrared observations, Atmos. Chem. Phys., 9, 6337–6350, doi:10.5194/acp-9-6337-2009, 2009a.

Crévoisier, C., Chédin, A., Matsueda, H., Machida, T., Armante, R.,and Scott, N. A.: First year of upper tropospheric integrated content of CO_2 from IASI hyperspectral infrared observations, Atmos. Chem. Phys., 9, 4797–4810, doi:10.5194/acp-9-4797-2009, 2009b.

Crisp, D., Fisher, B. M., O'Dell, C., Frankenberg, C., Basilio, R., Bösch, H., Brown, L. R., Castano, R., Connor, B., Deutscher, N. M., Eldering, A., Griffith, D., Gunson, M., Kuze, A., Man10 drake, L., McDuffie, J., Messerschmidt, J., Miller, C. E., Morino, I., Natraj, V., Notholt, J., O'Brien, D. M., Oyafuso, F., Polonsky, I., Robinson, J., Salawitch, R., Sherlock, V., Smyth, M., Suto, H., Taylor, T. E., Thompson, D. R., Wennberg, P. O., Wunch, D., and Yung, Y. L.: The ACOS CO_2 retrieval algorithm – Part II: Global XCO_2 data characterization, Atmos. Meas. Tech., 5, 687–707, doi:10.5194/amt-5-687-2012, 2012.

De Cesare, L., Myers, D. E., and Posa, D.: Spatio-temporal modelling of SO_2 in Milan district. In: Baafi, E. Y., Schofield, N. A. (Eds.), Geostatistics Wollongong. Kluwer Academic Publishing, Dordrecht, pp. 1031 – 1042, 1996.

De Cesare, L., Myers, D. E., and Posa, D.: Product–sum covariance for space–time modeling: an environmental application. Environmetrics 12, 11 – 23, 2001a.

De Cesare, L., Myers, D., and Posa, D.: Estimating and modeling space–time correlation structures: Stat. Prob. Lett., v. 51, no. 1, 9–14, 2001b.

Dimitrakopoulos, R., and Luo, X.: Spatiotemporal Modeling: Covariances and Ordinary Kriging Systems, Geostatistics for the Next Century, Kluwer Academic Publishers, Dordrecht, 88-93, 1994.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., Toon, G. C., Butz, A., Jung, M., Kuze, A., and Yokota, T.: New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, Geophys. Res. Lett., 38, L17706, doi:10.1029/2011GL048738, 2011.

Frankenberg C, O'Dell C, Guanter L, McDue J.: Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: implications for its retrieval and interferences with atmospheric CO2 retrievals. Meas. Tech. 5, 2081–2094., doi:10.5194/amt-5-2081-2012, 2012.

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., et al.: Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2, Remote Sensing of Environment, 147, 1-12, doi: 10.1016/j.rse.2014.02.007, 2014.

Gambacorta A.: The NOAA Unique CrIS/ATMS Processing System (NUCAPS): Algorithm Theoretical Basis Documentation, NOAA Center for Weather and Climate Prediction (NCWCP), Version 1.0, August 21, 2013, http://www.star.nesdis.noaa.gov/jpss/documents/ATBD/NUCAPS_ATBD_20130821.pdf

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P. E., Gomez-Dans, J., Kuze, A., Suto, H., and Grainger, R. G.: Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements, Remote Sens. Environ., 121, 236– 251, doi:10.1016/j.rse.2012.02.006, 2012.

Guo, L., Lei, L. and Zeng, Z.: Spatiotemporal correlation analysis of satellite-observed CO_2: Case studies in China and USA. Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International, 21-26 July, Melbourne, VIC, 2013.

Hammerling, D. M., Michalak, A. M., and Kawa, S. R.: Mapping of CO_2 at high spatiotemporal resolution using satellite observations: Global distributions from OCO-2, J. Geophys. Res., 117, D06306, doi:10.1029/2011JD017015, 2012a.

Hammerling, D. M., Michalak, A. M. O'Dell, C., and Kawa, S. R.: Global CO_2 distributions over land from the Greenhouse Gases Observing Satellite (GOSAT), Geophysical Research Letters, 39, L08804, doi:10.1029/2012GL051203, 2012b.

Huang, C., Zhang, H., and Robeson, S. M.: On the validity of commonly used covariance and variogram functions on the sphere, Math. Geosci., 43, MR2824128, 721–733, doi: 10.1007/s11004-011-9344-7, 2011.

De Iaco, S., Myers, D., and Posa, D.: Space-time analysis using a general product–sum model: Stat. Probab. Lett., v. 52, no. 1, 21–28, 2001.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Yoshida, Y., Corp, L. A., and Middleton, E. M.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space, Biogeosciences, 8, 637-651, doi:10.5194/bg-8-637-2011, 2011.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Middleton, E. M., Campbell, P. K. E., Yoshida, Y., Kuze, A., and Corp, L. A.: Filling-in of near-infrared solar lines by terrestrial fluorescence and other geophysical effects: simulations and space-based observations from SCIAMACHY and GOSAT, Atmos. Meas. Tech., 5, 809–829, doi:10.5194/amt-5-809-2012, 2012.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, Atmos. Meas. Tech., 6, 2803-2823, doi:10.5194/amt-6-2803-2013, 2013.

Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., and Chen, L.: 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, 2012.

Matthias, K., and Cressie, N.: Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets, Journal of Time Series Analysis 32.4, 430-446, doi: 10.1111/j.1467-9892.2011.00732.x, 2011.

534 Katzfuss, M. and Cressie, N.: Bayesian hierarchical spatio-temporal smoothing for very large datasets.
535 Environmetrics, 23: 94–107. doi: 10.1002/env.1147, 2012.

536 Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon
537 observation Fourier-transform spectrometer on the Greenhouse Gases 5 Observing Satellite for
538 greenhouse gases monitoring, Appl. Optics, 48, 6716–6733, doi:10.1364/AO.48.006716, 2009.

539 Lee, J.-E., Frankenberg, C., van der Tol, C., Berry, J. A., Guanter, L., Boyce, C. K., Fisher, J. B.,
540 Morrow, E., Worden, J. R., Asefi, S., Badgley, G., and Saatchi S.: Forest productivity and water stress in
541 Amazonia: observations from GOSAT chlorophyll fluorescence Proc. R. Soc. B, 280(1762), doi:
542 10.1098/rspb.2013.0171, 2013.

543 Li, J., and Heap, A.D.: Spatial interpolation methods applied in the environmental sciences: a review.
544 Environ. Model. Softw. 53, 173–189, doi: http://dx.doi.org/10.1016/j.envsoft.2013.12.008, 2014.

545 Li, J. and Heap, A.: A review of comparative studies of spatial interpolation methods in environmental
546 sciences: performance and impact factors. Ecol. Inf. 6, 228e241, 2011.

547 Li, J.: Assessing spatial predictive models in the environmental sciences: Accuracy measures, data
548 variation and variance explained, Environmental Modelling & Software, Volume 80, Pages 1-8, ISSN
549 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2016.02.004, 2016.

550 Maksyutov, S., Takagi, H., Valsala, V. K., Saito, M., Oda, T., Saeki, T., Belikov, D. A., Saito, R.,
551 Ito, A., Yoshida, Y., Morino, I., Uchino, O., Andres, R. J., and Yokota, T.: Regional CO_2 flux
552 estimates for 2009–2010 based on GOSAT and ground-based CO2 observations, Atmos. Chem.
553 Phys., 13, 9351–9373, doi:10.5194/acp-13-9351-2013, 2013.

554 Mathias, R.: Matrix completions, norms and Hadamard products, Proc. Amer. Math. Soc. 117 (4), 905–
555 918, 1993.

556 Million, E.: The Hadamard Product. http://buzzard.ups.edu/courses/2007spring/projects/million-
557 paper.pdf, 2007. Accessed on 15 January 2015.

558 NASA Earth Science: http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-
559 eosdis-data-products/, last access: 23 July, 2014.

560 Nguyen, H., Katzfuss, M., Cressie, N. and Braverman, A.: Spatio-Temporal Data Fusion for Very Large
561 Remote Sensing Datasets, Technometrics, Vol. 56, Iss. 2, doi: 10.1080/00401706.2013.831774, 2014.

562 O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi, M., Eldering,
563 D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polon15 sky, I., Smyth,
564 M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO_2 retrieval algorithm – Part
565 1: Description and validation against synthetic observations, Atmos. Meas. Tech., 5, 99–121,
566 doi:10.5194/amt-5-99-2012, 2012.

567 Rouhani, S., and Hall, T.J.: Space-Time Kriging of Groundwater Data. In: Armstrong, M. (Ed.),
568 Geostatistics. Kluwer Academic Publishers, Dordrecht, Vol. 2, 639-651, 1989.

569 Tadić, J., Qiu, X., Yadav, V. and Michalak, A.: Mapping of satellite Earth observations using moving
570 window block kriging, Geosci. Model Dev., 8, 1–9, doi:10.5194/gmd-8-1-2015, 2015.

Xiong X., Barnet C., Maddy E., Gambacorta A., King T., and Wofsy S.: Mid-upper tropospheric methane retrieval from IASI and its validation. Atmos Meas Tech. 6: 2255-2265, 2013.

Zeng, Z., LiPing, L., L. LiJie, G., Li, Z., Bing, Z.,: Incorporating temporal variability to improve geostatistical analysis of satellite-observed CO_2 in China, Chinese Science Bulletin, 58(16), 1948-1954, 2013.

Zeng, Z., Lei, L., Strong, K., Jones, D. B. A., Guo, L., Liu, ., Deng, F., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Henderson, B., Kivi, R., Lindenmaier, R., Morino, I., Notholt, J., Ohyama,H., Petri, C., Sussmann, R., Velazco, V., A., Wennberg, P., O., and Lin, H.: Global land mapping of satellite-observed CO_2 total columns using spatio-temporal geostatistics, International Journal of Digital Earth, DOI: 10.1080/17538947.2016.1156777, 2016.

Table 1. Cross-validation results of GOSAT XCO_2, IASI XCH_4 and GOME-2 SIF datasets using spatio-temporal and spatial methods, including mean absolute error (MAE), root mean squared error (RMSE), relative mean absolute error (RMAE), relative root mean square error (RRMSE), percent of observations lying outside of one, two, and three standard deviations ($\sigma_{\hat{z}}$) of the mapping uncertainty, and mean difference. MAE, RMSE and bias units for GOSAT, IASI and GOME-2 are ppm, ppb and mW/m^2/sr/nm, respectively. RMAE and RRMSE are unitless, and due to the reasons explained in Section 4.1 given only for GOSAT and IASI. Shaded fields represent best estimate in each category for every satellite.

		GOSAT XCO_2			IASI XCH_4			GOME-2 SIF		
		ST	1d	7d	ST	1d	7d	ST	1d	7d
Estimates	Mean absolute error (MAE)	0.83	0.74	0.~~88~~89	19.19	20.23	21.~~03~~04	0.52	0.~~51~~54	0.~~66~~54
	Root mean squared error (RMSE)	1.12	0.98	1.21	25.25	27.10	27.77	0.~~67~~68	0.~~65~~69	0.~~87~~69
	Relative mean absolute error (RMAE)	0.22	0.19	0.23	1.04	1.09	1.14	-	-	-

	Relative root mean square error (RRMSE)	0.29	0.25	0.31	1.37	1.46	1.50	-	-	-
Uncertainties	% observations falling outside $1\sigma_{\hat{z}}$ uncertainty	9.13	15.03	10.70	11.02	9.06	13.84	14.60	12.14	24.80
	% observations falling outside $2\sigma_{\hat{z}}$ uncertainty	1.12	3.01	1.39	0.48	0.51	0.86	1.20	0.64	4.33
	% observations falling outside $3\sigma_{\hat{z}}$ uncertainty	0.067	0.52	0.13	0.04	0.046	0.022	0.11	0.05	0.83
Bias	Mean difference	-0.012	0.0066	-0.034	0.28	-0.14	0.58	0.016	0.0013	0.032

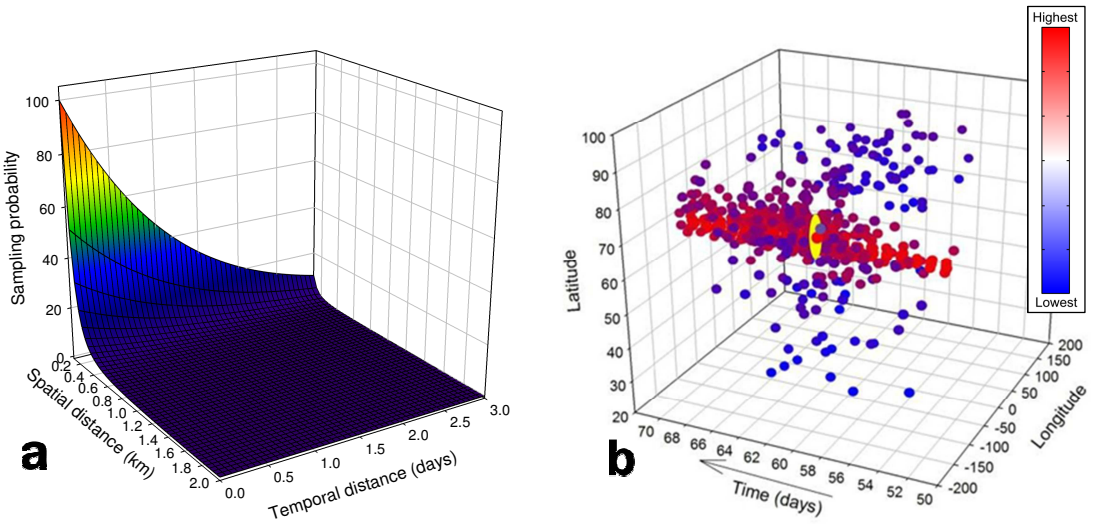
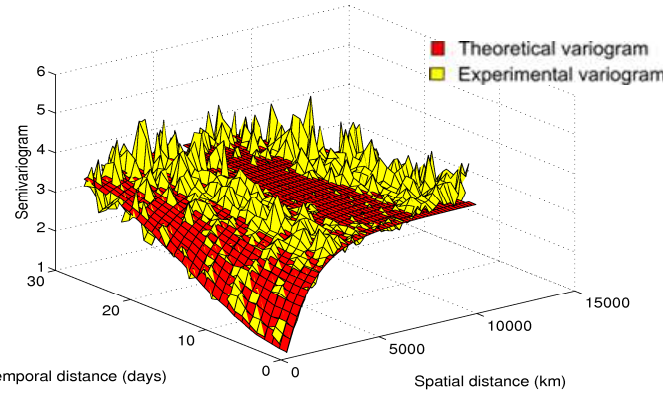


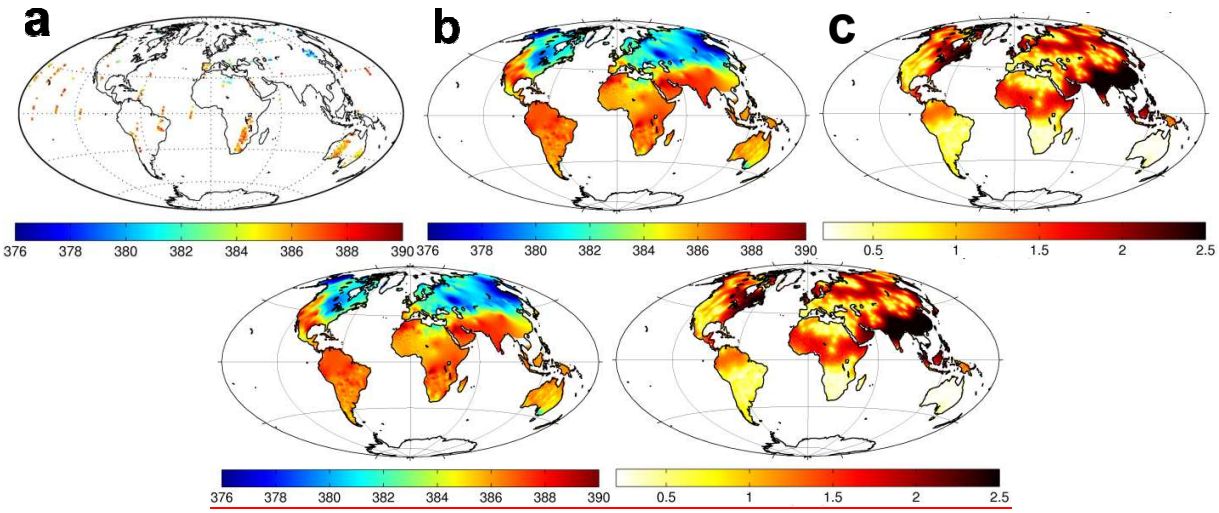
Figure 1. (a) Sampling probability as a decreasing function of spatial and temporal distance as used in this study, (b) The typical example of subsampled IASI Level 2 XCH_4 (altitude below 4 km) data for a selected estimation location (yellow circle). Color of observations shows semivariance between observation and estimation location (blue-lowest, red-highest). Due to stronger temporal covariance, the relative decrease of the sampling probability along temporal axis is smaller than with spatial distance.



607

608 **Figure 2.** Illustration of experimental and fitted theoretical spatio-temporal variogram for GOSAT XCO_2
609 data.

610



611

612

613 **Figure 3.** (a) GOSAT/ACOS v3.4 XCO_2 retrievals (Level 2 data) (ppm) for August 3, 2009 (b)
614 Contiguous global GOSAT/ACOS v3.4 maps (Level 3 data) (ppm) for the same day obtained using
615 Spatio-temporal Moving Window Block Kriging at 1 × 1° spatial resolution, (c) associated uncertainties,
616 given as 1-sigma ($\sigma_{\hat{z}}$) (ppm).

617

618

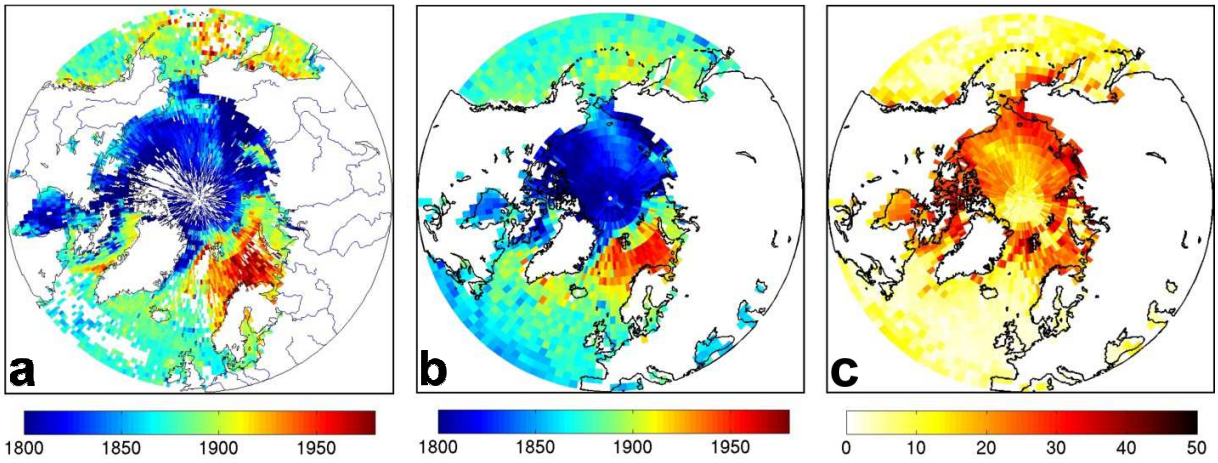


Figure 4. (a) IASI XCH_4 (0-4 km) retrievals (ppb) for March 2, 2013 (sea only), (b) Contiguous IASI maps for Northern Hemisphere for the same day obtained using Spatio-temporal Moving Window Block Kriging at 2.5 × 2° spatial resolution and (c) associated uncertainties, given as 1-sigma ($\sigma_{\hat{z}}$) (ppb).

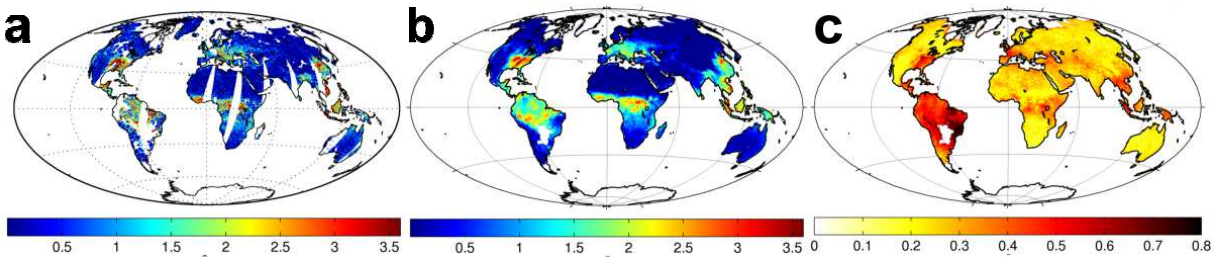


Figure 5. (a) GOME-2 SIF v14 retrievals (Level 2 data) (mW/m^2/sr/nm) for May 5, 2012, (b) Contiguous global GOME-2/SIF v14 maps (Level 3 data) (mW/m^2/sr/nm) for the same day obtained using Spatio-temporal Moving Window Block Kriging at 1 × 1° spatial resolution, (c) associated uncertainties, given as 1-sigma ($\sigma_{\hat{z}}$) (mW/m^2/sr/nm).