Mapping of satellite Earth observations using moving window block kriging

Jovan M. Tadić[1], Xuemei Qiu[1], Vineet Yadav[1] and Anna M. Michalak[1]

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

Abstract. Global gridded maps (a.k.a. Level 3 products) of Earth system properties observed by satellites are central to understanding the spatiotemporal variability of these properties. They also typically serve either as inputs into biogeochemical models, or as independent data for evaluating such models. Spatial binning is a common method for generating contiguous maps, but this approach results in a loss of information, especially when the measurement noise is low relative to the degree of spatiotemporal variability. Such "binned" fields typically also lack a quantitative measure of uncertainty.

Geostatistical mapping has previously been shown to make higher spatiotemporal resolution maps possible, and also provides a measure uncertainty associated with the gridded products. This study proposes a flexible moving window block kriging method that can be used as a tool for creating high spatiotemporal resolution maps from satellite data. It relies only on the assumption that the observed physical quantity exhibits spatial correlation that can be inferred from the observations. The method has several innovations relative to previously applied methods: 1) it provides flexibility in the spatial resolution of the contiguous maps 2) it is applicable for physical quantities with varying spatiotemporal coverage (i.e., density of measurements) by utilizing a more general and versatile data sampling approach, and 3) it provides rigorous assessments of the uncertainty associated with the gridded products. The method is demonstrated by creating Level 3 products from observations of column-integrated carbon dioxide (XCO_2) from the GOSAT satellite, and solar induced fluorescence (SIF) from the GOME-2 instrument.

1. Introduction

Satellite measurements of a Earth surface and atmospheric quantities have enormous benefits for Earth system science due to their global coverage and near real-time availability. They provide key constraints for developing models representing our understanding of the functioning of the Earth system. However, due to orbit geometries and geophysical limitations, a uniform or contiguous global coverage of these observations in space and/or time is not possible. This necessitates creation of contiguous maps for obtaining measurements at unsampled times and locations for understanding overall patterns, driving biogeochemical or physical models, and/or validating model predictions. Due to their widespread utility, global gridded maps are often part of the standard suite of satellite data products, and are often termed "Level 3" data (e.g. NASA, 2014).

In the case of column-integrated carbon dioxide (XCO_2) and solar induced fluorescence (SIF) observations, the two illustrative applications that will be used in this work, gridded products have been used, for example, to evaluate the representation of water stress in models of photosynthesis (Lee et al. 2013), to assess the performance of a terrestrial biosphere model in representing global CO_2 distributions (Hammerling et al. 2012b), and to constrain a model to assess the relative roles of variations in atmospheric transport and carbon exchange in explaining atmospheric CO_2 variability over the Amazon (Parazoo et al. 2013). The generation of Level 3 products is also often part of the standard processing sequence of observations (e.g. GOSAT Project, 2014; CO_2 DAAD, 2014).

Presently, "binning" is the most widespread method for creating such contiguous maps of satellite data. Such binning typically involves computing the mean of the observations that fall within a grid-cell (aka "bin") of an appropriate geographic size and time window (for applications of binning in the context of satellite retrievals of atmospheric concentration of carbon dioxide see; Kulawik et al., 2010; Crevoisier et al., 2009). However this simplicity comes with some limitations such as: (1) the mean is computed from a different number of measurements across grid-cells, (2) the inability to take into account any redundancy among nearby observations in computing the mean, and (3) the lack of gap filling properties for grid-cells that may contain no observations for a given time window.

The methodological deficiencies of binning can be overcome by using kriging, a geostatistical interpolation approach that takes into account the spatial and/or temporal correlation in the observations. Kriging is a best linear unbiased estimator, with the various implementations of ordinary kriging relying on the assumption of intrinsic stationarity. More typically, a covariance function is used to represent spatial correlation, and second-order stationarity is assumed, i.e. that the mean is constant and the covariance is only a function of the distance between observations (for kriging see; Chiles and Delfiner, 2012). Because the mean and covariance of Earth system observations vary substantially, the kriging tools need to be modified to reflect this nonstationarity. One such method is moving window kriging, in which kriging is performed locally and the covariance parameters are determined locally within pre-specified spatial and/or temporal subdomains (e.g. Haas 1990). The ability of the moving window kriging to reflect local uncertainty has been emphasized to be the most important advantage over kriging methods relying on the global covariance models (e.g. Harris et al., 2010; Walter et al., 2001; Van Tooren and Haas, 1993). Due to this advantage, the moving window kriging has been previously used for creating contiguous maps of satellite remote sensing observations of column-averaged CO_2 (XCO_2) (e.g. Hammerling et al. 2012a and 2012b).

This work proposes a further development of the moving window kriging method for application with satellite observations of Earth system properties. Whereas Hammerling et al. (2012a,b) used ordinary kriging as the basis for obtaining estimates at the spatial support (i.e. resolution or spatial footprint size) of observations, we propose a moving window block kriging method that can yield estimates at any resolution equal to or greater than that of the observations (for discussion on change of support in the context of remote sensing see; Atkinson and Curran, 1995, Collins and Woodcock, 1999; Braverman 2011). The main advantages of the proposed tools are that they make it possible to: (1) select the spatial support/resolution of the mapped quantities, (2) handle large volumes of data by developing subsampling technique that can make moving window block kriging computationally feasible for large number of satellite measurements, and (3) provide rigorous assessments of the uncertainty associated with the contiguous maps.

2. Methods

The proposed approach builds on the work of Hammerling et al. (2012a,b), with the goal of increasing the applicability and the flexibility of the nonstationary local kriging approach presented therein. The main innovations are twofold. The first is to allow flexibility in the spatial support of the estimates (i.e. the spatial resolution at which the mapping is conducted). The second is to provide a general approach for subsampling available observations in a manner that (i) captures the local correlation structure in the vicinity of each estimation grid cell and (ii) makes the statistical mapping approach computationally feasible in the case of applications with a very large number of observations.

The mapping proceeds in three steps for each grid cell and each estimation time on a regular grid, in order to create a contiguous map of the satellite observations. These steps are outlined in the subsections below, and include subsampling of the observations, characterization of the local spatial covariance structure, and interpolation at the desired spatial resolution. In Section 3, the new mapping approach is applied to two prototypical examples of satellite observations, namely observations of column-integrated concentration of atmospheric CO_2 concentrations (XCO_2) and observations of surface solar induced fluorescence (SIF), measured by the GOSAT satellite, and by the GOME instrument, respectively.

2.1 Subsampling of observations

The goal of the subsampling strategy is to preferentially sample observations in the vicinity of a given estimation grid cell, such that both the characterization of the local spatial covariance structure, and the ultimate mapped estimate and its associated uncertainty, are representative of local variability. This is accomplished by selecting the total number of observations to be used, N, where N is selected to be large enough to yield a representative sample, but small enough to make mapping computationally feasible on a given computational platform. For the applications presented in Section 3, N=500 and N=1000 for the XCO_2 and SIF mapping, respectively.

N observations are selected for each estimation grid cell by assigning a relative selection probability to each observation based on that observation's separation distance from the centroid of the grid cell. This selection probability could be application-specific, but for the applications presented here we selected:

$$P_s \propto 1/h^2 \tag{1}$$

where P_s is the relative probability of a given observation being selected, and h is the great circle distance between the location x_i of an observation and the centroid x_j of the estimation gridcell:

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

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

The form of P_s in eqn. 1 ensures that a comparable number of observations is selected within any equal-area concentric band around an estimation grid cell, thereby also ensuring that observations that are at close distances to one another are preferentially close to the estimation location. This is a desirable feature because observations that are close to one another define the shape of the variogram at short separation distances (Section 2.2), and the variogram should reflect variability in the vicinity of the estimation grid cell. Different forms of P_s could also be used, for example if more / fewer observations along a given direction were desirable in order to better represent expected correlations along a given direction.

In previous work (Alkhaled et al. 2008; Hammerling et al. 2012a,b), a fixed application-specific window size was instead defined within which all available observations were used, together with a user-defined fraction of observations outside of the window. The window size was based in part on expected scales of variability in the satellite observations. The updated approach presented here reduces the number of user-selected parameters, and explicitly provides a mechanism for ensuring the computational feasibility of mapping in the case of very large datasets, such as the SIF example examined here.

2.2 Characterization of Spatial Covariance

The characterization of the local covariance structure of the observations around each estimation grid cell, based on the subsampled observations, proceeds as described in Hammerling et al. (2012a, Section 2.1), except that (1) all possible pairs of observations are included in the formulation of the raw variogram, and the nugget-effect variance, representative of the retrieval / measurement errors, is not spatially uniform. The reader is referred to that earlier publication for additional details.

Briefly, for each estimation grid cell, a raw variogram is calculated based on the subsampled observations:

$$\gamma(h) = \frac{1}{2}[y(x_i) - y(x_j)]^2 \tag{3}$$

where γ is the raw variogram value for a given pair of observations $y(x_i)$ and $y(x_j)$, and h is the great circle distance between the locations (x_i and x_j) of these observations, as defined in eqn. 2.

A parametric function, the theoretical variogram, is fitted to the raw variogram using non-linear least squares. For the prototypical applications presented here, an exponential variogram function with a nugget effect was used, because it yields a valid covariance function on a sphere (Huang et al., 2011), provided a good match to the known physical characteristics of the observations, and fit the observed variability well:

$$\gamma(h) = \begin{cases} 0, \ for\ h = 0 \\ \sigma^2(1 - \exp\left(-\frac{h}{l}\right) + \sigma_{nug}^2, \ for\ h > 0 \end{cases} \tag{4}$$

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. The nugget component can be either prescribed (as in the XCO_2 example in Section 3) or estimated (as in the SIF example in Section 3), depending on the availability of information about measurement and retrieval errors.

The variogram parameters can be used to define a corresponding local spatial covariance structure for the mapped quantity (XCO_2 or SIF, in the prototypical examples presented here). For the variogram function in eqn. 4 this becomes:

$$q(h) = \sigma^2 \exp\left(-\frac{h}{l}\right) \tag{5}$$

The nugget effect is correspondingly used to define the covariance structure of the measurement and retrieval errors:

$$R(h) = \begin{cases} \sigma_{nug}^2, \ for\ h = 0 \\ 0, \ for\ h > 0 \end{cases} \tag{6}$$

2.3 Mapping using moving window block kriging

Ordinary kriging, a minimum variance linear unbiased mapping method for spatial data, was used in Hammerling et al. (2012a,b) to create contiguous maps of XCO_2. In this approach, the spatial support (i.e. footprint) of the estimates corresponds to that of the observations. Although the mapping can be performed at any spatial interval (e.g. once per $1^{\circ}\times1^{\circ}$ grid cell), the estimates remain representative of the variability at the scale of the observations.

Here, we instead using block kriging (e.g. Webster, 2000), an approach that yields estimates that represent an average within a specified area. This makes it possible to disassociate the native footprint of the observations from the resolution of the mapped product, thereby making it possible to create contiguous maps at any desired spatial resolution equivalent to or greater than the size of the observation footprints. As with moving window ordinary kriging, block kriging provides an optimal estimate of the quantity being mapped (XCO_2 and SIF, in the prototypical examples presented here) for each estimation location, based on the subsampled observations (Section 2.1) and the local covariance structure (Section 2.2), together with a rigorous assessment of the uncertainty associated with the estimate.

The linear system of equations that is solved to obtain the N weights $\boldsymbol{\lambda}$ assigned to the subsampled observations for a given estimation grid cell is:

$$\begin{bmatrix} \mathbf{Q} + \mathbf{R} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\nu \end{bmatrix} = \begin{bmatrix} \mathbf{q}_A \\ 1 \end{bmatrix} \tag{7}$$

where $\mathbf{Q}$ is a $N\times N$ covariance matrix among the N observations with individual entries as defined in eqn. 5, $\mathbf{R}$ is an $N\times N$ diagonal measurement and retrieval error covariance matrix among the N observations as defined in eqn. 6, $\mathbf{1}$ is an $N\times1$ unity vector, T denotes the vector transpose operation, and $\mathbf{q}_A$ is an $N\times1$ vector of the spatial covariances between the estimation grid cell and the N observation locations, defined as:

$$q_{A,i} = \frac{1}{n}\sum_{j=1}^{n} q\left(h_{i,j}\right) \tag{8}$$

where $q_{A,i}$ is the covariance between the grid cell and observation i, and $q(h_{i,j})$ is defined as in eqn. 5 based on the distance $h_{i,j}$ between observation i and n regularly-spaced locations within the grid cell. In general, the larger the n the better the representation of the area (i.e. grid cell) to observation covariance. For practical purposes, in the applications presented here, n is defined based on the relative footprint of the observations compared to that of the estimation grid cells.

The system in eqn. 7 is solved for $\boldsymbol{\lambda}$ and the Lagrange multiplier ν. These parameters are then used to define the estimate ($\hat{z}$) and estimation uncertainty variance ($\sigma^2_{\hat{z}}$) for the grid cell as:

$$\hat{z} = \boldsymbol{\lambda}^T \mathbf{y} \tag{9}$$

$$\sigma^2_{\hat{z}} = \sigma_{AA} - \boldsymbol{\lambda}^T \mathbf{q}_A + \nu \tag{10}$$

where $\mathbf{y}$ is the $N \times 1$ vector of subsampled observations, and σ_{AA} is the variance of the mapped quantity (XCO_2 or SIF, in the prototypical examples presented here) at the resolution of the estimation grid cell, defined as:

$$\sigma_{AA} = \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} q(h_{j,k}) \tag{11}$$

where $q(h_{j,k})$ is defined as in eqn. 5 based on the distance $h_{j,k}$ between any combination of the n regularly spaced locations within the grid cell defined previously.

3. Example applications

The mapping approach described in Section 2 is demonstrated using two prototypical examples of satellite observations: 1) observations of column-integrated concentrations of atmospheric CO_2 (XCO_2) from the GOSAT satellite, and 2) observations of surface solar induced fluorescence (SIF) from the GOME-2 instrument. These applications differ in the spatial footprint (i.e. support) of the observations (nadir footprint of about 10.5 km diameter at sea level (Kuze et al., 2009), and 40 km × 80 km (Joiner et al, 2013), respectively), the volume of available data (approximately 2×10^3 and 2×10^5 observations per week, respectively), the time scales of variability, and the degree of spatial variability and nonstationary in the observed quantity.

3.1 Global land XCO_2 fields observed by GOSAT

The Japanese Greenhouse Gasses Observing SATellite (GOSAT) (e.g., Kuze et al., 2009) was launched in 2009 and is the first satellite dedicated to global greenhouse gas monitoring, including CO_2 and CH_4. GOSAT flies in a polar, sun-synchronous orbit with a 3-day repeat cycle and an approximately 13:00 local time overpass time. GOSAT XCO_2 data are being used to examine a number of questions in carbon cycle science, including comparing observed and modeled XCO_2 fields (Hammerling et al., 2012b), quantifying sources and sinks of CO_2 (e.g., Deng et al., 2014; Basu et al., 2013, 2014; Chevallier et al., 2014; Takagi et al., 2014), detecting perturbations in the carbon cycle (Guerlet et al., 2013) and interpreting seasonal changes in the carbon balance (Parazoo et al., 2013).

Measurements of XCO_2 (a.k.a. “Level 2” data) are derived using a number of retrieval algorithms, among them NASA’s Atmospheric CO_2 Observations from Space (ACOS) algorithm (e.g., O’Dell et al., 2012; Crisp et al., 2012). Filtered and bias-corrected data from the most up to date version of this algorithm (ACOS v3.4 release 3) are used here to demonstrate the mapping approach presented in Section 2. Approximately 900 successful retrievals are available per three-day repeat cycle, with the majority of observations being over land. These data have substantial retrieval uncertainties (e.g., O’Dell et al., 2012) and include large gaps (e.g., Figure 1). These features prevent the application of simple spatial and temporal binning techniques for generating XCO_2 maps at spatiotemporal scales that are directly useful for addressing existing uncertainties in carbon cycle science.

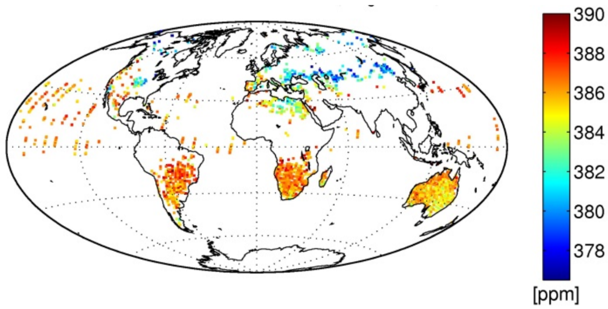


Figure 1. ACOS v3.4 release 3 XCO_2 Level 2 data ("Observations") for August 2-7, 2009.

The approach described in Section 2 is used to create continuous maps, a.k.a. Level 3 data, based on XCO_2 observations obtained over two repeat cycles, namely August 2-7, 2009 (Figure 1). A six-day period is used to balance the competing goals of including as many observations as possible, while avoiding time periods over which the XCO_2 field itself would change substantially (see discussion in Hammerling et al. 2012a). Maps of XCO_2 and associated uncertainties are created at native (Figure 2a,b) and 1° × 1° (Figure 2c,d) resolutions, in order to examine and demonstrate the impact of resolution on mapping uncertainty. Targeting different resolutions is made possible by the use of the moving window block kriging approach presented here. N=500 subsampled observations are used per estimation location. These maps can also be compared to those presented for an equivalent period in Hammerling (2012b, Auxiliary Figures 2 and 3), with methodological differences as described in Section 2, and representative of the estimated XCO_2 at the native resolution of sounding footprints (nadir footprint ~10.5 km diameter) with estimates at 1° × 1.25° intervals.

Results show that, because of the information content of the sparse observations, the estimated fields (Figure 2 a,c) are similar at native and 1° × 1°, but that estimating directly at the coarser 1° × 1° resolution yields lower uncertainties as observations become more informative for spatially-averaged quantities (Figure 3).

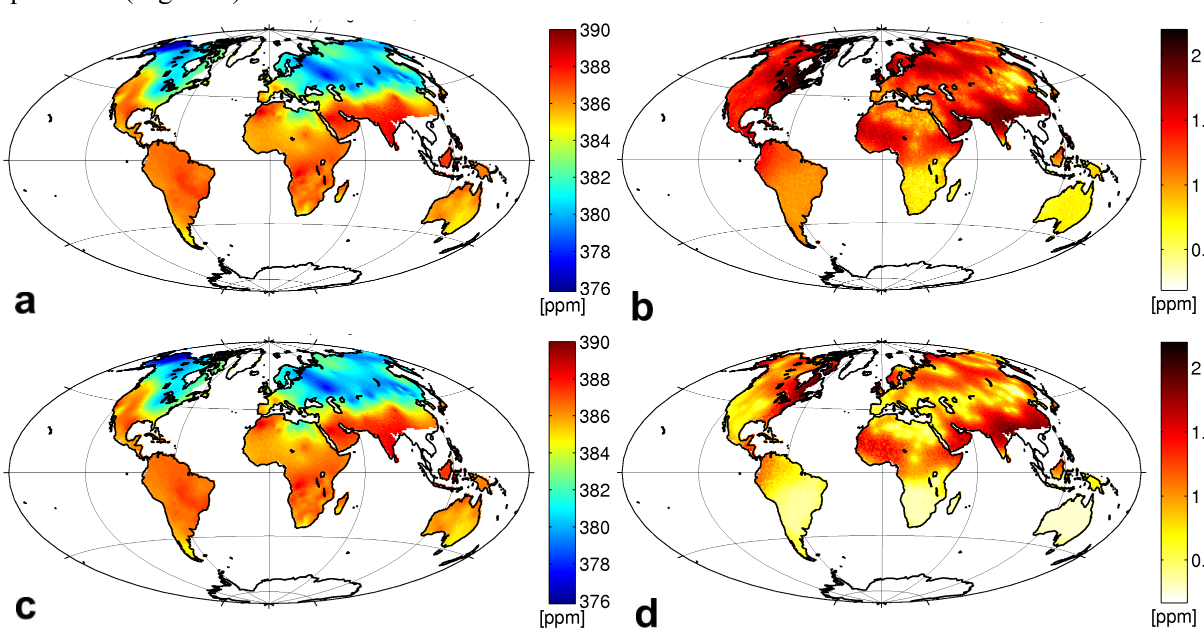


Figure 2. XCO_2 Level 3 maps (a,c) and associated uncertainties (b,d) based on ACOS 3.4 release-3 retrievals ("Estimates") for August 2-7, 2009 at (a,b) native resolution and (c,d) 1° × 1° resolution, obtained using the proposed mapping approach.

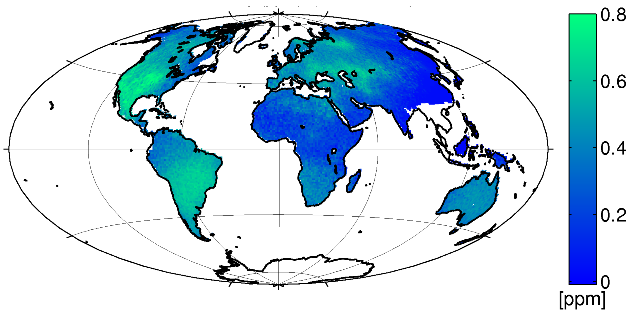


Figure 3. Reduction in estimation uncertainties between the native estimation resolution and the 1°x1° estimation resolution for XCO_2 Level 3 maps based on ACOS 3.4 release-3 retrievals for August 2-7, 2009.

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

A series of recent studies has demonstrated the potential use of satellite observations of solar-induced fluorescence (SIF) for understanding and quantifying photosynthetic CO_2 uptake at large scales, using data from the GOSAT satellite (e.g., Joiner et al., 2011; Frankenberg et al., 2011; Guanter et al., 2012, Joiner et al., 2012; Lee et al., 2013; Frankenberg et al., 2012), the SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY) instrument on board ENVISAT (e.g., Joiner et al., 2012), the GOME-2 (The Global Ozone Monitoring Experiment–2) instrument on board METOP-A (e.g., Joiner et al., 2013), and the Orbiting Carbon Observatory (OCO-2) (e.g., Frankenberg et al., 2014). Satellite measurements of fluorescence can be used with land surface models to improve the representation of GPP and to understand GPP response to environmental stress (e.g., Lee et al., 2013). Among available datasets, GOME-2 provides the highest spatial and temporal density of data.

Until now, studies of SIF have relied on spatially and temporally binned average observations at monthly or coarser timescales and 1° or coarser spatial scales (e.g., Figure 4). The coarse spatial and temporal scales were used to overcome, through the use of simple averaging, spatial gaps in observations and the relatively high uncertainties associated with individual retrievals. One of the limitations of such an approach is that it inherently discards information about SIF variability at fine spatial and temporal scales, which is important for understanding the impact of transient effects such as changes in phenology and water availability (Lee et al., 2013), and developing biospheric models that can represent these effects correctly. A second limitation is the lack of a direct and robust quantification of the uncertainty associated with the mapped products, complicating uncertainty analysis in subsequent applications using the data.

As a second demonstration of the mapping approach proposed here, we therefore 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° × 1°), but at multiple temporal resolutions. The examination of shorter time periods was selected in order to more directly respond to scientific opportunities in the use of SIF data, and to complement the spatial-resolution-focused demonstration of Section 3.1. Maps of SIF and associated uncertainties are created at one, six, and 31 day temporal resolutions in August, 2009 (Figure 5), where August 2009 was chosen for convenience to correspond with the XCO_2 application presented in Section 3.1. N=1000 subsampled observations are used per estimation location. The monthly map can also be compared to the monthly binned map presented in Figure 4.

Results show that the proposed approach can leverage nearby observations to create realistic contiguous maps even at one-day resolution (Figure 5a,b), although, as expected, uncertainties are reduced (Figure 5d) at coarse temporal resolutions, just as was seen for coarser spatial resolutions in the XCO_2 application. In fact, the regions with little or no SIF data for the one-day application are clearly visible as high-uncertainty bands in Figure 5b, and a user could explicitly decide whether such uncertainties are acceptable, or too high for a given scientific application. When maps are intended to be used to drive and/or validate biogeochemical models, therefore, having the ability to choose a desirable balance between temporal resolution and mapping uncertainty presents a considerable advantage.

Ideally, the temporal resolution at which maps are obtained is as fine as possible so as to capture the dynamics of the observed physical quantity, in this case SIF. The choice of optimal temporal resolution thus, in general, defines a trade-off between having sufficient observations for adequate spatial coverage, while minimizing the impact of temporal variability in the quantity being examined (Hammerling et al., 2012a). From Figure 6 it is apparent that the presented approach makes it possible to obtain maps at temporal resolutions much higher than the monthly (or coarser) resolution of current binned products. As expected, the more abundant observations available at 6-day temporal resolution (Figure 6d) lead to decreased estimation uncertainty compared to 1-day resolution (Figure 6b). However, at monthly temporal resolutions (Figure 6e,f) the temporal variability in SIF over a 31-day period increases the discrepancy among (spatially) nearby observations, leading to increased uncertainties at coarse time scales. This effect is apparent in comparing Figures 6d and 6f, as uncertainty increases over, for example, eastern South America. A similar trade-off was also noted in selecting mapping time-scales for XCO_2 (Hammerling et al. 2012a), and further speaks to the advantage of being able to select a mapping timescale based on scientific need and uncertainty tolerance, as is possible with the approach presented here.

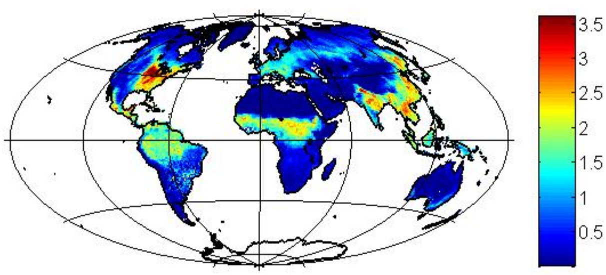
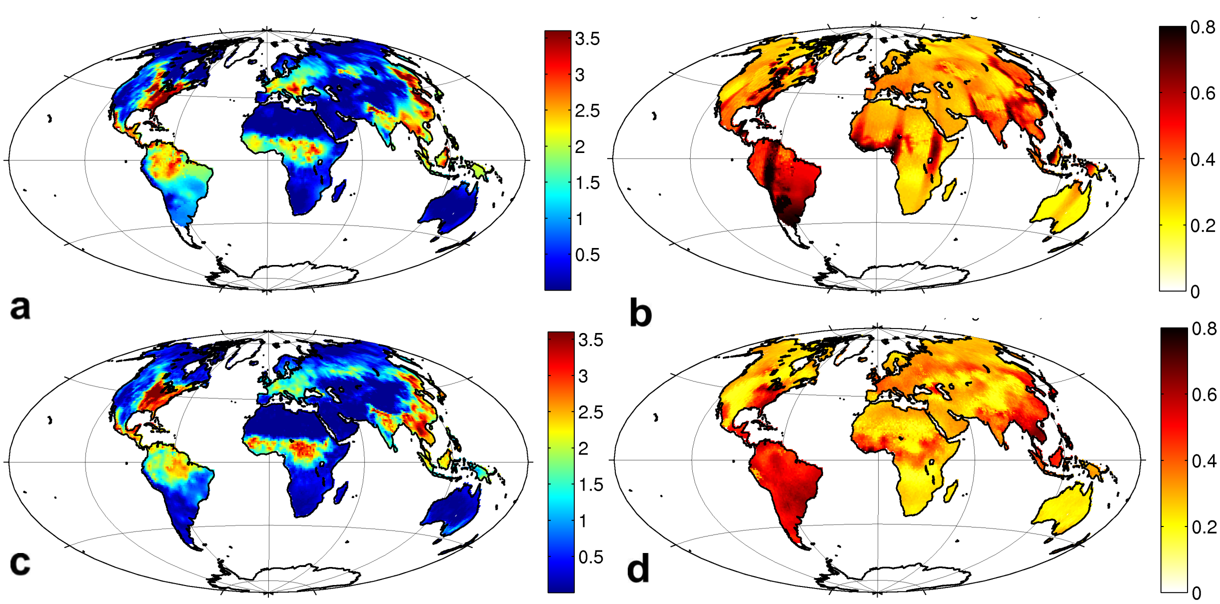


Figure 4. Monthly-averaged binned map of GOME-2 SIF data for August 1-31, 2009 ($mW/m^2/sr/nm$).



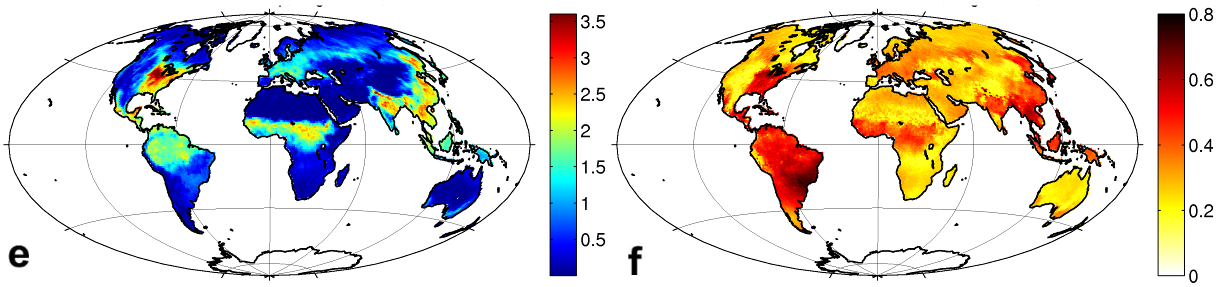


Figure 5. Maps of global SIF (mW/m^2/sr/nm) (a,c,e) and associated estimation uncertainties expressed as standard deviations (b,d,f), for August 1, 2009 (a,b), August 2-7, 2009 (c,d) and August 1-31, 2009 (e,f) obtained using GOME-2 observations and the presented mapping approach at 1° × 1° spatial resolution.

4. Method evaluation

Leave-one-out cross-validation is used to evaluate the performance of the proposed method. In doing so, the goal is for the predicted values to be as directly comparable as possible to the observation being held back. With that goal in mind, the cross-validation analysis is performed for maps generated at 1-day temporal resolution, and at the native spatial resolution of the sounding footprints.

We apply this strategy for both SIF and XCO_2 test cases. For SIF, for each day in August 1-7, 2009, 10% of available GOME-2 SIF 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 in leave-one-out cross validation. All three mapping steps (see Sections 2.1-2.3) are repeated *ab initio* during cross-validation. The performance of the mapping method is tested in terms of the accuracy of the best estimates (the difference between estimates and withheld observations), and the accuracy of the uncertainty bounds (the degree to which the reported uncertainties capture the difference between estimates and withheld observations) and bias (the mean difference between estimates and withheld observations).

The accuracy of the maps at daily temporal resolution and native spatial resolution is evaluated using the mean absolute difference (MAD) and the root mean squared difference (RMSD) between the mapped estimates and observations held back in leave-one-out cross-validation (Table 1). Although an absolute target value for these accuracy metrics is not available, it is interesting to note that the MAD and RMSD are comparable to the reported measurement uncertainty in both satellite datasets (0.77 ppm for GOSAT XCO_2, 0.55 mW/m^2/sr/nm for GOME-2 SIF). We also compare the GOSAT XCO_2 values to those obtained from by applying the method developed in Hammerling et al. (2012a), which yielded a MAD of 0.86 ppm and a RMSD of 1.20 ppm, demonstrating comparable performance, but with the additional benefits provided by the new method as described in Section 2.

Table 1. Cross-validation results of GOSAT XCO_2 and GOME-2 SIF datasets, including mean absolute difference, root mean squared difference, percent of observations lying outside of one, two, and three standard deviations ($\sigma_{\hat{z}}$) of the mapping uncertainty, and mean difference.

		GOSAT XCO_2	GOME-2 SIF
Estimates	Mean absolute difference	0.85 ppm	0.47 $mW/m^2/sr/nm$
	Root mean squared difference	1.15 ppm	0.61 $mW/m^2/sr/nm$
Uncertainties	% observations falling outside $1\sigma_{\hat{z}}$ uncertainty	10.06%	11.23%
	% observations falling outside $2\sigma_{\hat{z}}$ uncertainty	0.96%	0.60%
	% observations falling outside $3\sigma_{\hat{z}}$ uncertainty	0.18%	0.03%
Bias	Mean difference	-0.007 ppm	0.002 $mW/m^2/sr/nm$

Estimation uncertainties reflect the locations and number of observations surrounding the estimation location, the degree of spatial variability in the mapped field in the vicinity of the estimation location, and the spatiotemporal support of the estimates. The accuracy of the uncertainties obtained from the mapping method is evaluated by quantifying the reliability with which the uncertainty bounds associated with the estimates capture the values of the withheld observations. Specifically, we calculate the percentage of estimation locations where the withheld observations fall outside of the one, two, and three estimation standard deviation ($\sigma_{\hat{z}}$) uncertainty bounds. For independent, normally-distributed data, these percentages should be approximately 32%, 5% and 0.3%, respectively. Although these assumptions do not hold here, these values still provide a general indication of expected performance.

For both applications, the percentage of observations falling outside of the uncertainty bounds is lower than would be expected for normally-distributed data (Table 1), showing good mapping accuracy. These percentages are very similar when the analysis is repeated using the method developed by Hammerling et al. (2012a). The lower percentages are due to the fact that observations are not normally distributed.

Finally, the bias of the developed method is quantified using the mean difference between estimates and the withheld observations in the leave-one-out cross-validation. Theoretically, mean difference should approach zero as the number of cross-validation points increases if the method provides perfectly unbiased estimates. The mean difference for both applications (Table 1) was several orders of magnitude lower than the observed spatial gradients in the mapped quantities (e.g., Figure 1, Figure 4), and was not statistically significant (p>>0.05: p=0.86 for GOSAT XCO_2; p=0.63 for GOME-2 SIF). The approach therefore yields unbiased estimates.

5. Conclusions

In this study we propose a flexible moving window block kriging method that can be used as a tool for creating high spatiotemporal resolution maps from satellite data. The method can be applied in a standalone mode, or as a part of broader satellite data processing package. The resulting maps can also be incorporated into biogeochemical and physical models of the Earth system. The approach relies only on the assumption that the observed physical quantity exhibits spatial correlation that can be inferred from the observations. The method has several advantages over previously applied methods: 1) it allows for the creation of contiguous maps at varying spatio-temporal resolution, 2) it can be applied for creating contiguous maps for physical quantities with varying spatio-temporal coverage (aka density of measurements), 3) it provides assessments of the uncertainty of interpolated values. The approach emphasizes the use of local covariance structures in predictions by an arbitrary selection of the sampling function, limiting the radius around estimation locations and adjusting the number of sampled points to a fraction of available measurements. The approach also limits the number of partially-subjective ancillary parameters required, making it applicable across a variety of applications.

The method was demonstrated by creating Level 3 products from two datasets with considerably different spatio-temporal properties. While the GOSAT XCO_2 observations were relatively sparse, the GOME-2 SIF data had a much higher spatio-temporal density. In the case of GOSAT XCO_2, the effects of making predictions at different spatial supports (i.e. resolutions) were analyzed, showing that a decrease in the resolution slightly affects estimates ('smoothing' effect) and more significantly estimation uncertainties (reduced uncertainties at coarser resolution). In the case of GOME-2 SIF, the focus was kept on the effect of different aggregation time periods by creating maps at higher temporal resolutions. This example demonstrated the importance of being able to select a mapping timescale based on scientific need and uncertainty tolerance as optimal temporal resolution results from a trade-off between having sufficient observations for adequate spatial coverage, while minimizing the impact of temporal variability in the quantity being examined. In this it was shown that even daily Level 3 maps could be successfully created by the proposed method. For both datasets, the method was shown to yield precise, accurate, and unbiased estimates. The results clearly indicate that contiguous maps can be created at different spatial resolutions for time periods shorter than achievable by binning/averaging.

The resulting maps can be used to support the development of improved models of the Earth system, both by serving as driver data and validation data for such models.

Acknowledgement

This work was supported by the National Aeronautics and Space Administration (NASA) through grant no. NNX08AJ92G, and the National Science Foundation (NSF) through grant no. 1342076. The XCO_2 data were produced by the ACOS/OCO-2 project at the Jet Propulsion Laboratory, California Institute of Technology, and obtained from Christopher O'Dell (CSU, ACOS/OCO-2 Algorithms Team; http://reef.atmos.colostate.edu/~odell/odell.html). We thank NASA and the ACOS/OCO-2 project for providing the data, and the three Japanese parties (NIES, JAXA, MOE) for making the GOSAT spectra available to the scientific community. We also thank Joanna Joiner (NASA/GSFC) for providing GOME-2 SIF data, and NASA Carbon Cycle Science program (NNH10DA001N) for funding the SIF research.

References

Alkhaled, A. A., Michalak, A. M. Olsen, S., Kawa, S. R., and Wang, J.-W.: A global evaluation of the regional spatial variability of column integrated CO_2 distributions, Journal of Geophysical Research – Atmospheres, 113, D20303, doi:10.1029/2007JD009693, 2008.

Atkinson, P. M. and Curran, P. J.: Defining an optimal size of support for remote sensing investigations, IEEE Transactions on Geoscience and Remote Sensing, 33, 768-776, 1995.

Baker, I. T., Berry, J. A., Lee, J., Frankenberg, C., and Denning, S.: Constraining Simulated Photosynthesis with Fluorescence Observations, American Geophysical Union, Fall Meeting 2012, abstract #B12B-02.

Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13, 8695-8717, doi:10.5194/acp-13-8695-2013, 2013.

Basu, S., Krol, M., Butz, A., Clerbaux, C., Sawa, Y., Machida, T., Matsueda, H., Frankenberg, C., HAsekamp, O. P., and Aben, I.: The seasonal variation of the CO_2 flux over Tropical Asia estimated from GOSAT, CONTRAIL, and IASI, Geophys. Res. Lett., 41, 1809–1815, doi:10.1002/2013GL059105, 2014.

Braverman, A., Nguyen, H., Olsen, E., Miller, C., Cressie, N., Kratzfuss, M., Wang, R., and Michalak, A.: Geostatistical data fusion for remote sensing applications, NASA, Greenbelt, MD, USA, NASA Annu. Rep., Apr 2011, 2011 Report from the ESTO Advanced Information Systems Technology (AIST) Program, NASA, Greenbelt, MD, USA, 2011.

Chevallier, F., Palemer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional CO_2 flux estimates from in situ and spaceborne measurements of atmospheric CO_2, Geophys. Res. Lett., 41, 1065–1070, doi:10.1002/2013GL058772, 2014.

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

CO_2 DAAD: http://dge.stanford.edu/labs/michalaklab/CO2DAAD/XCO2maps.html, last access: 23 July 2014.

Collins, J. B. and Woodcock, C. E.: Geostatistical estimation of resolution-dependent variance in remotely sensed images, Photogrammetric Engineering and Remote Sensing, 65, 41-50, 1999.

Crevoisier, 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, 2009.

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., Mandrake, 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 XCO2 retrieval algorithm, Part 2: Global XCO_2 data characterization, Atmos. Meas. Tech, 5, 687-707, doi:10.5194/amt-5-687-2012, 2012.

Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of atmospheric CO_2 from GOSAT XCO_2 data, Atmos. Chem. Phys., 14, 3703-3727, doi:10.5194/acp-14-3703-2014, 2014.

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 CO_2 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., Pollock, R., and Taylor, T. E.: 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.

GOSAT Project: http://www.gosat.nies.go.jp/eng/gosat/page5.htm, last access: 23 July 2014.

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.

Guerlet, S., Basu, S., Butz, A., Krol, M., Hahne, P., Houweling, S., Hasekamp, O. P., and Aben, I.: Reduced carbon uptake during the 2010 Northern Hemisphere summer from GOSAT, Geophys. Res. Lett., 40, 2378–2383, doi:10.1002/grl.50402, 2013.

Haas, T. C.: Lognormal and moving window methods of estimating acid deposition. J. Am. Stat. Assoc., 85:950–963, 1990.

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.

Harris, P., Charlton, M., and Fotheringham, A.S.: Moving window kriging with geographically weighted variograms. SERRA 24:1193–1209, 2010.

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, 2011.

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.

Kulawik, S. S., Jones, D. B. A., Nassar, R., Irion, F. W., Worden, J. R., Bowman, K. W., Machida, T., Matsueda, H., Sawa, Y., Biraud, S. C., Fischer, M. L., and Jacobson, A. R.: Characterization of Tropospheric Emission Spectrometer (TES) CO_2 for carbon cycle science, Atmos. Chem. Phys., 10, 5601-5623, doi:10.5194/acp-10-5601-2010, 2010.

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

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

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

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

Parazoo, N. C., Bowman, K., Frankenberg, C., Lee, J. E., Fisher, J. B., Worden, J., Jones, D. B. A., Berry, J., Collatz, G. J., Baker, I. T., Jung, M., Liu, J., and Osterman, G.: Interpreting seasonal changes in the carbon balance of southern Amazonia using measurements of XCO_2 and chlorophyll fluorescence from GOSAT. Geophys. Res. Lett., 40, 2829–2833, doi:10.1002/grl.50452, 2013.

Takagi, H., Houweling, S., Andres, R. J., Belikov, D., Bril, A., Boesch, H., Butz, A., Guerlet, S., Hasekamp, O., Maksyutov, S., Morino, I., Oda, T., O'Dell, C. W., Oshchepkov, S., Parker, R., Saito, M., Uchino, O., Yokota, T., Yoshida, Y., and Valsala, V.: Influence of differences in current GOSAT XCO_2 retrievals on surface flux estimation, Geophys. Res. Lett., 41, 2598–2605, doi:10.1002/2013GL059174, 2014.

Van Tooren, C.F., and Haas, T.C.: A site investigation strategy using moving window kriging and automated semivariogram modelling. In: Contaminated soil '93. Kluwer Academic Press, Dordrecht, 609–622, 1993.

Walter, C., McBratney, A.B., Douaoui, A., and Minasny, B.: Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. Aust. J. Soil. Res., 39, 259–272, 2001.

Webster, R.: Geostatistics for Engineers and Earth Scientists. European Journal of Soil Science, 51, 541–549, doi:10.1046/j.1365-2389.2000.00334-9.x, 2000.