

Space-time Data Fusion

Amy Braverman, Hai Nguyen, and Edward Olsen
Jet Propulsion Laboratory,
California Institute of Technology
Pasadena, California 91109
Email: Amy.Braverman@jpl.nasa.gov

Noel Cressie
Department of Statistics
The Ohio State University
Columbus, OH 43210
Email: ncressie@stat.osu.edu

Abstract—Space-time Data Fusion (STDF) is a methodology for combining heterogeneous remote sensing data to optimally estimate the true values of a geophysical field of interest, and obtain uncertainties for those estimates. The input data sets may have different observing characteristics including different footprints, spatial resolutions and fields of view, orbit cycles, biases, and noise characteristics. Despite these differences all observed data can be linked to the underlying field, and therefore the each other, by a statistical model. Differences in footprints and other geometric characteristics are accounted for by parameterizing pixel-level remote sensing observations as spatial integrals of true field values lying within pixel boundaries, plus measurement error. Both spatial and temporal correlations in the true field and in the observations are estimated and incorporated through the use of a space-time random effects (STRE) model. Once the models parameters are estimated, we use it to derive expressions for optimal (minimum mean squared error and unbiased) estimates of the true field at any arbitrary location of interest, computed from the observations. Standard errors of these estimates are also produced, allowing confidence intervals to be constructed. The procedure is carried out on a fine spatial grid to approximate a continuous field. We demonstrate STDF by applying it to the problem of estimating CO₂ concentration in the lower-atmosphere using data from the Atmospheric Infrared Sounder (AIRS) and the Japanese Greenhouse Gases Observing Satellite (GOSAT) over one year for the continental US.

I. INTRODUCTION

The motivation for this work is the need to combine data from multiple remote sensing instruments to paint a complete and quantitative picture of the distribution of important geophysical quantities over the Earth and in its atmosphere. We focus on carbon dioxide (CO₂) in the lower part of Earth's atmosphere because this may be a proxy for CO₂ flux, and that flux is an extremely important quantity for understanding the carbon cycle. Moreover, no remote sensing instrument observes flux directly, but NASA's Atmospheric Infrared Sounder (AIRS) and the Japanese Greenhouse Gases Observing Satellite (GOSAT) observe different parts of the atmospheric column which, taken together, can be used to calculate an estimate of lower atmosphere CO₂. In order to produce these estimates and quantify their uncertainties we must account for differences in the input data sets: their resolutions and other spatial and temporal sampling characteristics. We also need to exploit spatial and temporal correlations in both the observations and in the underlying true field and correct for instrument biases.

The fundamental construct that allows us to do these things

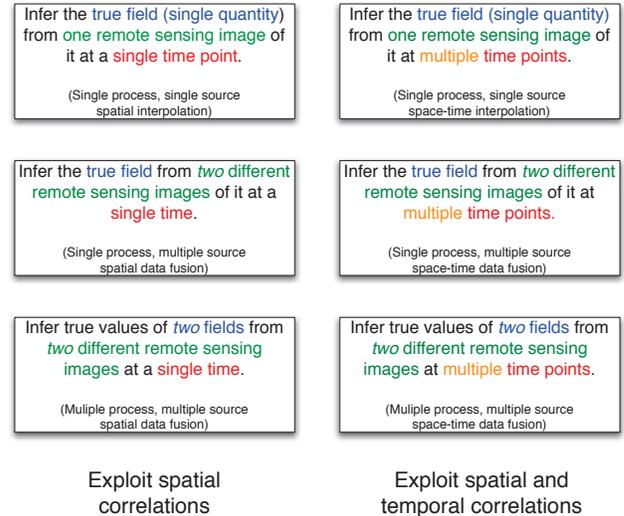


Fig. 1. Problems in remote sensing data analysis. The text boxes describe six related problems. The left column articulates three of these. The top-left box relates to making inferences about a single field of interest (e.g., carbon dioxide) from one remote sensing data set at a single time point. The middle-left box relates to making inferences about a single geophysical field from two data sets, also at a single time. The bottom-left box corresponds to making inferences about two different geophysical fields when each is seen by a different instrument. The right panel shows the same set of scenarios, but for cases where multiple snapshots of the scene are available at a succession of times.

is a statistical model that relates the observed data to the underlying true field of interest at a specified set of locations and times. To introduce the model and how it will be used, we first discuss a basic statistical framework for modeling spatio-temporal data and show how this framework addresses the problem of spatial interpolation using a single data set. Then we show that data fusion is just an extension of the same idea to multiple data sets simultaneously. Finally, we extend the framework to include time so that we take full advantage of both spatial and temporal relationships in the data.

Figure 1 shows the relationships among six different flavors of the space-time interpolation problem, including data fusion. Each box in the Figure states the the problem it represents (e.g., infer the true values of two fields from two different remote sensing images at multiple time points), and gives a specific name (e.g., multiple process, multiple source, space-time data fusion) to that particular version. In this article, we

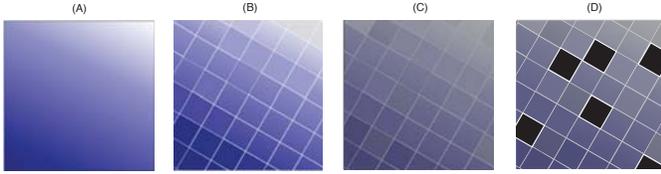


Fig. 2. Remote sensing data and their relationship to the true quantity of interest. The left panel shows the true geophysical field (A). (B) shows the field as it is viewed by a remote sensing instrument. The image is pixelated because each pixel is the average of the true values in (A) belonging to it. The instrument has measurement error, so (C) shows the image corrupted by noise. Noise is assumed to be independent from pixel to pixel. Finally, some pixels may not be observed at all due to instrument observing characteristics (e.g., some instruments cannot see through clouds). (D) shows the image under these conditions.

will build up the story starting with the box in the upper-left and ending with the box in the lower-right: multiple process, multiple source, space-time data fusion, which we generically refer to as space-time data fusion (STDF) since the other five boxes represent special cases of it.

In Section II, we present a basic model for remote sensing data that articulates quantitatively the relationships between the data we observe and the unobserved true quantity of interest. In Section III we use that model to address the problem of spatial interpolation using a single data set, then extend that to data fusion beginning with one quantity of interest and two data sources but ignoring time. Next, we expand the paradigm to cover the case where there are two quantities of interest, each viewed by one instrument, and finally we incorporate time into our model yielding the STDF methodology. In Section We apply STDF to one year of data from AIRS and GOSAT to estimate lower atmosphere CO₂ on a one-degree spatial grid and two-week time steps for one year, over the continental US. We close with a short discussion of these results.

II. THE STATISTICAL NATURE OF REMOTE SENSING DATA

Remote sensing data are, by their very nature, statistical. Satellite instruments do not directly observe geophysical variables, they observe radiances from which geophysical information is inferred. Observations from space are typically made on spatial units coincident with instruments' pixels, while the true physical process is continuous in space. This discretization of the scene is one source of uncertainty. Another source is that the instrument itself adds measurement error at the pixel level. This error includes both bias ("systematic" error) and variance ("random" error). Finally, there may be additional bias and variability due to the inability of the instrument to observe under certain conditions (e.g. clouds). These relationships are illustrated in Figure 2.

The problem is that we only have access to information like that in panel (D) in Figure 2, and we want to infer the true continuous field in panel (A). We can do this using a spatial statistical model that relates the observations in panel D to the true field in panel (A).

The relationships necessary to make an inference about (A) based on information from (D) can be quantified precisely. Let \mathbf{s} be a point location (e.g. latitude and longitude) in the domain shown in panel (A) of Figure 2. Denote the domain by D , and let $Y(\mathbf{s})$ be a random variable capturing the value of the geophysical variable (equivalently, geophysical "field") of interest at location \mathbf{s} in D . The remote sensing instrument discretizes the scene into pixels (as shown in panel (B)), so we define

$$Y(B) = \frac{1}{|B|} \int_{\mathbf{u} \in B} Y(\mathbf{u}) d\mathbf{u}, \quad (1)$$

as the noiseless value of Y at the resolution of the pixel, denoted by B . $Y(B)$ is simply the average of $Y(\mathbf{s})$ for all \mathbf{s} in the pixel B . The instrument adds random measurement noise to what it sees:

$$Z(B) = Y(B) + \epsilon(B),$$

where $\epsilon(B)$ is an independent Gaussian measurement error term with variance $Var(\epsilon(B)) = \sigma_\epsilon^2$, and possibly some bias, $E(\epsilon(B)) = \mu_\epsilon$. This is shown graphically in panel (C) as a noisy version of panel (B). Finally, some pixels may have missing data as shown in panel (D). We concatenate all the observed values for non-missing pixels in panel (D) to form a column vector

$$\mathbf{Z} = (Z(B_1), \dots, Z(B_N))',$$

which captures the data in the N non-missing pixels.

III. INFERENCE FROM REMOTE SENSING DATA

In this section we use the model presented in Section II to develop the models we will use for data fusion.

A. Single Process Spatial Interpolation and Data Fusion

The simplest inference problem related to remote sensing data is to infer the geophysical field, sometimes called the geophysical "process", at a specified location (or on a grid of locations) from aggregated, noisy, incomplete data like those shown in panel (D) of Figure 2. A simple linear estimate of $Y(\mathbf{s})$ is

$$\hat{Y}(\mathbf{s}) = \mathbf{a}'_s \mathbf{Z},$$

where \mathbf{a}_s is chosen to minimize the mean squared error,

$$MSE(Y(\mathbf{s}), \hat{Y}(\mathbf{s})) = E\|\hat{Y}(\mathbf{s}) - Y(\mathbf{s})\|^2, \quad (2)$$

subject to the condition that the estimate be unbiased:

$$E(\hat{Y}(\mathbf{s})) = E(Y(\mathbf{s})). \quad (3)$$

The problem can be solved using the method of Lagrange multipliers to minimize (2) subject to (3).

Now suppose a second remote sensing data set, possibly with different resolution and statistical characteristics, is available (see Figure 3). As before, denote the geophysical variable's value at location \mathbf{s} by $Y(\mathbf{s})$ but now we have two data vectors,

$$\mathbf{Z}_1 = Z(B_{11}), \dots, Z(B_{1N_1})', \quad \mathbf{Z}_2 = Z(B_{21}), \dots, Z(B_{2N_2})',$$

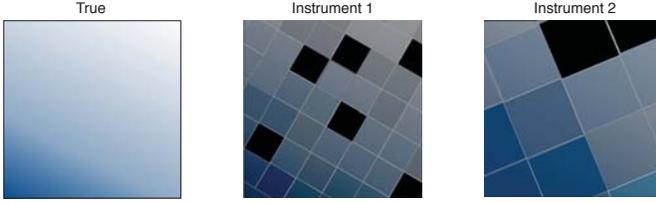


Fig. 3. A true geophysical field (left), and two remote sensing instruments' views of it (center and right).

where B_{km} denotes the m th footprint in the data set for instrument k . For technical reasons it is necessary to make a slight modification to the definition of $Y(B)$ in (1). Instead of defining $Y(B)$ as an integral over an infinite number of locations \mathbf{s} in the domain, we partition the domain into a set of very fine, non-overlapping quantum spatial units called basic areal units (BAU's). Each location \mathbf{s} is associated with one of these BAU's so \mathbf{s} is not continuously varying, but in practice it is nearly so. The domain D is defined to be the union of all the identifying points for the BAU's. The reason for this modification is so that in practice there are a countable number of locations to be considered in any calculation. The modified definitions of $Y(B_{km})$ and $Z_k(B_{km})$ are

$$Y(B_{km}) = \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in D \cap B_{km}} Y(\mathbf{u}), \quad \text{and}$$

$$Z_k(B_{km}) = Y(B_{km}) + \epsilon_k(B_{km}).$$

Note that $Y(B_{1m_1})$ and $Y(B_{2m_2})$ are different from each other because they are averages over different spatial footprints, and that we never actually observe them.

As was the case earlier, the optimal (minimum mean squared error, unbiased) estimate of $Y(\mathbf{s})$ is of the form

$$\hat{Y}(\mathbf{s}) = \mathbf{a}'_{1\mathbf{s}} \mathbf{Z}_1 + \mathbf{a}'_{2\mathbf{s}} \mathbf{Z}_2, \quad (4)$$

where $\mathbf{a}_{1\mathbf{s}}$ and $\mathbf{a}_{2\mathbf{s}}$ are the solutions to the constrained minimization of (2) subject to the unbiasedness condition, (3). We call $\mathbf{a}_{1\mathbf{s}}$ and $\mathbf{a}_{2\mathbf{s}}$ the data fusion coefficients and (4) the data fusion estimator of $Y(\mathbf{s})$. While it is intuitively clear that the data fusion problem can be solved this way, there is a more general formulation that will be required in order to introduce time into the framework.

The formulation in (4) focuses on solving for the data fusion coefficients, $(\mathbf{a}_{1\mathbf{s}}, \mathbf{a}_{2\mathbf{s}})$, at every location \mathbf{s} for which an inference will be made. However, our interest is not so much in the coefficients themselves as it is in the estimates of $Y(\mathbf{s})$. An alternative model for $Y(\mathbf{s})$ is

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \nu(\mathbf{s}) + \xi(\mathbf{s}), \quad (5)$$

$$\mu(\mathbf{s}) = \mathbf{t}(\mathbf{s})' \boldsymbol{\alpha}, \quad \nu(\mathbf{s}) = \mathbf{S}(\mathbf{s})' \boldsymbol{\eta}.$$

The model in Equation (5) is called the Spatial Random Effects model (SRE), and it expresses $Y(\mathbf{s})$ as a sum of three components. The first component, $\mu(\mathbf{s})$, is called the trend, and reflects the large scale behavior of Y . The trend is the behavior of $Y(\mathbf{s})$ that is easily explained by a simple statistical model

such as the regression, $\mu(\mathbf{s}) = \mathbf{t}(\mathbf{s})' \boldsymbol{\alpha}$. $\mathbf{t}(\mathbf{s})$ is (say) a vector of the latitude and longitude of location \mathbf{s} , and $\boldsymbol{\alpha}$ is the least squares regression coefficient.

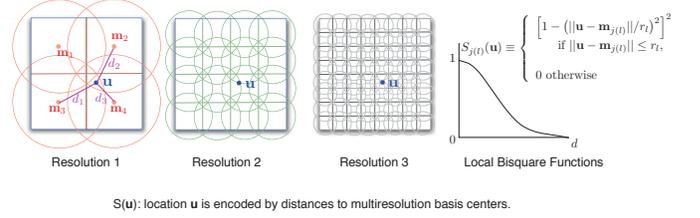


Fig. 4. Multiresolution basis for encoding location \mathbf{u} .

The component $\nu(\mathbf{s})$ is called the spatial covariance term, and it captures the spatial covariance structure of $Y(\mathbf{s})$; we assume it that it is Gaussian with mean zero and $Cov(\nu(\mathbf{u}), \nu(\mathbf{v}))$ (the covariance between two locations) that can be modeled as a linear combination of the elements of a hidden state vector, $\boldsymbol{\eta}$. $\mathbf{S}(\mathbf{s})$ is the weight vector for location \mathbf{s} that tells us how to weight the elements of $\boldsymbol{\eta}$ in order to arrive at an approximation for that portion of the structure of $Y(\mathbf{s})$ that is not captured by the trend. The elements of $\boldsymbol{\eta}$ correspond to a set of locations defined at several specified levels of spatial resolution. For example, in Figure 4 the domain is divided into four subregions with centers $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$, and \mathbf{m}_4 at the coarsest level of spatial resolution, into 16 subregions at the second level of resolution, and into 64 subregions at the third level of resolution. The centers for the latter two levels are not shown. A spatial location \mathbf{u} is encoded by the local bisquare function value of its distances from these 84 centers. The local bisquare function decays as shown in the right panel of Figure 4. These 84 distances are the components of the weight vector $\mathbf{S}(\mathbf{u})$, and $\boldsymbol{\eta}$ is also of dimension $r = 84$ in the example provided by the figure. The encoded values of the locations of interest are said to form a multiresolution basis set, and, since the encoding scheme is chosen by us, they are known and fixed. $\boldsymbol{\eta}$, on the other hand, is not known and must be estimated from the data. It turns out that $\boldsymbol{\eta}$ can be estimated from footprint level data as we shall see below.

The final component of the model (5) is the fine-scale variation term, $\xi(\mathbf{s})$. This accounts for variability not captured by the trend or spatial covariance terms. One can also think of it as sub-BAU variation. It is assumed to be Gaussian with mean zero and variance σ_ξ^2 for all locations.

The estimate of $Y(\mathbf{s})$ implied by the model (5) is

$$\hat{Y}(\mathbf{s}) = \hat{\mu}(\mathbf{s}) + \hat{\nu}(\mathbf{s}) + \hat{\xi}(\mathbf{s}),$$

$$= \mathbf{t}(\mathbf{s})' \hat{\boldsymbol{\alpha}} + \mathbf{S}(\mathbf{s})' \hat{\boldsymbol{\eta}} + \hat{\xi}(\mathbf{s}). \quad (6)$$

To proceed, we need to estimate $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$, and $\xi(\mathbf{s})$. It will be helpful to define the following quantities for the k th data set,

$k = 1, 2$:

$$\begin{aligned}
\mathbf{Y}_k &= (Y(B_{k1}), \dots, Y(B_{kN_k}))', \\
Y(B_{km}) &= \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} Y(\mathbf{u}), \\
\mathbf{T}_k &= (\mathbf{t}(B_{k1}), \dots, \mathbf{t}(B_{kN_k}))', \\
\mathbf{t}(B_{km}) &= \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \mathbf{t}(\mathbf{u}), \\
\boldsymbol{\mu}_k &= (\mu(B_{k1}), \dots, \mu(B_{kN_k}))', \\
\mu(B_{km}) &= \left[\frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \mathbf{t}(\mathbf{u})' \boldsymbol{\alpha} \right] \\
&= \mathbf{t}(B_{km})' \boldsymbol{\alpha}, \\
\boldsymbol{\nu}_k &= (\nu(B_{k1}), \dots, \nu(B_{kN_k}))', \\
\nu(B_{km}) &= \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \nu(\mathbf{u}), \\
\boldsymbol{\xi}_k &= (\xi(B_{k1}), \dots, \xi(B_{kN_k}))', \\
\xi(B_{km}) &= \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \xi(\mathbf{u}), \\
\mathbf{S}_k &= (\mathbf{S}(B_{k1}), \dots, \mathbf{S}(B_{kN_k}))', \\
\mathbf{S}(B_{km}) &= \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \mathbf{S}(\mathbf{u}), \\
\boldsymbol{\epsilon}_k &= (\epsilon_k(B_{k1}), \dots, \epsilon_k(B_{kN_k}))'.
\end{aligned}$$

These are vectors or matrices of footprint-level quantities defined as the spatial averages of the BAU-level quantities associated with locations within those footprints. Note in particular that $\mu(B_{km})$ can be expressed as a linear combination of $\mathbf{t}(B_{km})$ and $\boldsymbol{\alpha}$, which is independent of level. The basis functions $S(B_{km})$ also obey this linear aggregation relationship, and this is crucial because it allows us to write,

$$\begin{aligned}
\nu(B_{km}) &= \frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \nu(\mathbf{u}), \\
&= \left[\frac{1}{|D \cap B_{km}|} \sum_{\mathbf{u} \in (D \cap B_{km})} \mathbf{S}(\mathbf{u})' \right] \boldsymbol{\eta}, \\
&= \mathbf{S}(B_{km})' \boldsymbol{\eta}, \tag{7}
\end{aligned}$$

and hence $\boldsymbol{\eta}$ is also independent of level of resolution. Both $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ can therefore be estimated from footprint-level data. The model relating *all* the data to the unknown quantities is

$$\begin{aligned}
\begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} &= \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}, \\
&= \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \end{pmatrix} \boldsymbol{\alpha} + \begin{pmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{pmatrix} \boldsymbol{\eta} + \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix},
\end{aligned}$$

or more compactly,

$$\mathbf{Z} = \mathbf{T}\boldsymbol{\alpha} + \mathbf{S}\boldsymbol{\eta} + \boldsymbol{\xi} + \boldsymbol{\epsilon}, \tag{8}$$

where $\mathbf{Z} = (\mathbf{Z}_1', \mathbf{Z}_2')'$, $\mathbf{T} = (\mathbf{T}_1', \mathbf{T}_2')'$, $\mathbf{S} = (\mathbf{S}_1', \mathbf{S}_2')'$, $\boldsymbol{\xi} = (\boldsymbol{\xi}_1', \boldsymbol{\xi}_2')'$, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2')'$.

To estimate $\boldsymbol{\alpha}$, we regress \mathbf{Z} on \mathbf{T} if there is no bias in the instrument measurements. No bias means that the expectations of $\epsilon_1(B_{1m})$ and $\epsilon_2(B_{2m})$ are both zero. It is, of course, possible that the instruments do have biases which we would then model as $E(\epsilon_1(B_{1m})) = c_1\mu(B_{1m})$ and $E(\epsilon_2(B_{2m})) = c_2\mu(B_{2m})$ where c_k are multiplicative bias coefficients. The choice of a multiplicative bias model rather than an additive one is somewhat a matter of convenience at this point. If the bias coefficients are non-zero, then we define

$$\mathbf{C} = \begin{bmatrix} (1 + c_1)\mathbf{I}_{N_1} & \mathbf{0} \\ \mathbf{0} & (1 + c_2)\mathbf{I}_{N_2} \end{bmatrix},$$

where \mathbf{I}_{N_k} is the $N_k \times N_k$ identity matrix, and regress \mathbf{Z} on \mathbf{CT} to account for the biases. The solution is,

$$\hat{\boldsymbol{\alpha}} = [(\mathbf{CT})'(\mathbf{CT})]^{-1}(\mathbf{CT})'\mathbf{Z}.$$

While $\hat{\boldsymbol{\alpha}}$ is an estimate and therefore has a variance, we treat it as fixed from this point forward.

To estimate $\boldsymbol{\eta}$, we use a Bayesian formalism, and assume that a priori $\boldsymbol{\eta}$ is (r -dimensional) multivariate Gaussian with mean zero and ($r \times r$) covariance matrix \mathbf{K} :

$$\boldsymbol{\eta} \sim N_r(\mathbf{0}, \mathbf{K}).$$

The optimal a posteriori estimate of $\boldsymbol{\eta}$ is

$$\hat{\boldsymbol{\eta}} = E(\boldsymbol{\eta}|\mathbf{Z}) = \mathbf{G}(\mathbf{Z} - \boldsymbol{\mu}), \quad \mathbf{G} = \mathbf{KS}'\boldsymbol{\Sigma}_Z^{-1}. \tag{9}$$

The covariance matrix of \mathbf{Z} is $\boldsymbol{\Sigma}_Z$ and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$. Using an Empirical Bayes approach, we estimate \mathbf{K} off-line and consider it fixed for the remainder of this analysis. Currently we use the method-of-moments for this. Details can be found in [1].

Finally, to estimate $\xi(\mathbf{s})$ we assume a priori that

$$\xi(\mathbf{s}) \sim N(0, \sigma_\xi^2),$$

with σ_ξ^2 estimated off-line as was done for \mathbf{K} above (See [1]). The optimal a posteriori estimate of $\xi(\mathbf{s})$ is

$$\hat{\xi}(\mathbf{s}) = E(\xi(\mathbf{s})|\mathbf{Z}) = \mathbf{b}(\mathbf{s})'\boldsymbol{\Sigma}_Z^{-1}(\mathbf{Z} - \boldsymbol{\mu}), \tag{10}$$

where $\mathbf{b}(\mathbf{s}) = (\mathbf{b}_1(\mathbf{s})', \mathbf{b}_2(\mathbf{s})')'$, and

$$\mathbf{b}_k(\mathbf{s}) = \sigma_\xi^2 \left(\frac{I(\mathbf{s} \in B_{k1})}{|D \cap B_{k1}|}, \dots, \frac{I(\mathbf{s} \in B_{kN_k})}{|D \cap B_{kN_k}|} \right)'.$$

Although $\boldsymbol{\eta}$ and $\xi(\mathbf{s})$ are independent a priori, they are not independent a posteriori because they both depend on the same data, \mathbf{Z} . The formulas for $\hat{\boldsymbol{\eta}}$ and $\hat{\xi}(\mathbf{s})$ given in (9) and (10) are correct, but when we introduce time, this dependence will become important.

Now using (6) our estimate of $Y(\mathbf{s})$ is

$$\begin{aligned}
\hat{Y}(\mathbf{s}) &= \mu(\mathbf{s}) + \hat{\nu}(\mathbf{s}) + \hat{\xi}(\mathbf{s}), \\
&= \mathbf{t}(\mathbf{s})' \boldsymbol{\alpha} + \mathbf{S}(\mathbf{s})' \hat{\boldsymbol{\eta}} + \hat{\xi}(\mathbf{s}), \\
&= \mathbf{t}(\mathbf{s})' \boldsymbol{\alpha} + \mathbf{S}(\mathbf{s})' \mathbf{KS}' \boldsymbol{\Sigma}_Z^{-1} (\mathbf{Z} - \mathbf{T}\boldsymbol{\alpha}) \\
&\quad + \mathbf{b}(\mathbf{s})' \boldsymbol{\Sigma}_Z^{-1} (\mathbf{Z} - \mathbf{T}\boldsymbol{\alpha}), \\
&= \mathbf{t}(\mathbf{s})' \boldsymbol{\alpha} + (\mathbf{S}(\mathbf{s})' \mathbf{KS}' + \mathbf{b}(\mathbf{s})') \\
&\quad \times \boldsymbol{\Sigma}_Z^{-1} (\mathbf{Z} - \mathbf{T}\boldsymbol{\alpha}). \tag{11}
\end{aligned}$$

We treat α , and therefore μ also, as known, fixed quantities here. The variance of the estimate is

$$\begin{aligned} \text{Var}(\hat{Y}(\mathbf{s})) &= [(\mathbf{S}(\mathbf{s})' \mathbf{K} \mathbf{S}' + \mathbf{b}(\mathbf{s})') \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}] \boldsymbol{\Sigma}_{\mathbf{Z}} \\ &\times [(\mathbf{S}(\mathbf{s})' \mathbf{K} \mathbf{S}' + \mathbf{b}(\mathbf{s})') \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}]'. \end{aligned} \quad (12)$$

We can see from the form of (11) that the fusion coefficients from Equation (4), $\mathbf{a}_{\mathbf{s}} = (\mathbf{a}_{1\mathbf{s}}', \mathbf{a}_{2\mathbf{s}}')'$, are given by

$$\mathbf{a}_{\mathbf{s}} = (\mathbf{S}(\mathbf{s})' \mathbf{K} \mathbf{S}' + \mathbf{b}(\mathbf{s})') \boldsymbol{\Sigma}_{\mathbf{Z}}^{-1},$$

and are applied to the detrended data, $(\mathbf{Z} - \mathbf{T}\alpha)$, prior to forming the final estimate of $Y(\mathbf{s})$.

The reader may notice that the formulas in (11) and (12) depend on being able to calculate the inverse of the very large $(N_1 + N_2) \times (N_1 + N_2)$ matrix, $\boldsymbol{\Sigma}_{\mathbf{Z}}$. This is the variance-covariance matrix of all the data from both instruments at all locations at which they observe. The form of $\boldsymbol{\Sigma}_{\mathbf{Z}}$ lends itself to the application of the Sherman-Morrison-Woodbury matrix inversion formula [2] which allows for computationally efficient inversion. Details are omitted here in the interest of space, but can be found in [1].

B. The Multiple Process, Multiple Source Case

The methodology of the previous section generalizes readily to the case of two (or more) quantities of interest, or “processes”. For example, in this project, the two processes of interest are two different features of the CO2 column in the atmosphere: total column CO2, which is observed by Japan’s Greenhouse gases Observing Satellite (GOSAT), and mid-tropospheric CO2, which is measured by NASA’s Atmospheric Infrared Sounder (AIRS) instrument. We let $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}))'$ be the vector of process values at location \mathbf{s} , where $Y_1(\mathbf{s})$ denotes total column CO2, and $Y_2(\mathbf{s})$ denotes mid-tropospheric CO2. All the formulas developed earlier apply here, but there are now two α ’s, two η ’s, and two $\xi(\cdot)$ ’s, one for each process.

Define

$$\boldsymbol{\mu}^C = \begin{pmatrix} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \mathbf{T}^C \boldsymbol{\alpha}^C,$$

$$\boldsymbol{\nu}^C = \begin{pmatrix} \mathbf{S}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \mathbf{S}^C \boldsymbol{\eta}^C, \quad \text{and}$$

$$\boldsymbol{\xi}^C = \begin{pmatrix} \xi_{11} \\ \xi_{22} \end{pmatrix},$$

where $\boldsymbol{\xi}_{kk} = (\xi_k(B_{k1}), \dots, \xi_k(B_{kN_k}))'$. The formula in (11) holds in the multiple process case with \mathbf{Y} for Y , $\boldsymbol{\mu}^C$ for μ , $\boldsymbol{\nu}^C$ for ν , and $\boldsymbol{\xi}^C$ for ξ . The formulas for the components of these terms generalize straightforward ways to the vector case, as does the variance in (12). Estimation of the required parameters becomes more complex, but still tractable. Details can be found in [3].

C. Multiple Process, Multiple Source Space-time Data Fusion

In this section we present space-time data fusion for the multiple process, multiple source problem. In general, let $\hat{\theta}_t$ be the estimate of the quantity θ at time t . Let $\hat{\theta}_{t|t-1}$ be the estimate of θ at time t based on data up to and including time $t-1$, and let $\hat{\theta}_{t|t}$ be the estimate of θ at time t based on data up to and including time t . The main idea behind STDF is to use the Kalman filter to update the joint estimate of $(\boldsymbol{\eta}_t, \boldsymbol{\xi}_t)$ at each time step given the previous estimates and new data acquired at that time point. At time step t we “forecast” the new value of $\boldsymbol{\eta}_t$ from the relation (13) (before seeing the data) to produce $\hat{\boldsymbol{\eta}}_{t|t-1}$. Then, after the data for time step t arrive, we “update” the forecast in light of the data. $\boldsymbol{\xi}_t$ is not forecasted at each time step, but is updated jointly with $\boldsymbol{\eta}_t$ because the two quantities are statistically dependent.

STDF assumes a state-space model that employs an order-one, vector auto-regression as the state equation,

$$\boldsymbol{\eta}_{t+1}^C = \mathbf{H}_{t+1} \boldsymbol{\eta}_t^C + \boldsymbol{\zeta}_{t+1}^C, \quad (13)$$

where $\boldsymbol{\eta}_t^C$ is an $(r_1 + r_2)$ -dimensional state vector, and \mathbf{H}_t is the $(r_1 + r_2) \times (r_1 + r_2)$ state transition matrix at time t . $\boldsymbol{\zeta}_{t+1}^C$ is an $(r_1 + r_2)$ -dimensional zero-mean Gaussian vector with covariance matrix \mathbf{U}_{t+1} . \mathbf{U}_{t+1} is assumed to be independent of $\boldsymbol{\eta}_t^C$. The data observed by instrument k over the footprint A at time t is generated according to the following model,

$$Z_t^{(k)}(A) = \mu_t^{(k)}(A) + \mathbf{S}_t^{(k)'}(A) \boldsymbol{\eta}_t^{(k)} + \xi_t^{(k)}(A) + \epsilon_t^{(k)}(A),$$

$k = 1, 2$. We stack the scalars into column vectors and the row vectors into matrices to form the following model for all the observed data at time t for instrument k :

$$\mathbf{Z}_t^{(k)} = \boldsymbol{\mu}_t^{(k)} + \mathbf{S}_t^{(k)'} \boldsymbol{\eta}_t^{(k)} + \boldsymbol{\xi}_t^{(k)} + \boldsymbol{\epsilon}_t^{(k)}, \quad (14)$$

We again stack the elements of (14) to form a meta-data set,

$$\mathbf{Z}_t^C = \boldsymbol{\mu}_t^C + \mathbf{S}_t^{C'} \boldsymbol{\eta}_t^C + \boldsymbol{\xi}_t^C + \boldsymbol{\epsilon}_t^C,$$

where $\mathbf{Z}_t^C = (\mathbf{Z}_t^{(1)'}, \mathbf{Z}_t^{(2)'})'$, $\boldsymbol{\mu}_t^C = (\boldsymbol{\mu}_t^{(1)'}, \boldsymbol{\mu}_t^{(2)'})'$, $\boldsymbol{\eta}_t^C = (\boldsymbol{\eta}_t^{(1)'}, \boldsymbol{\eta}_t^{(2)'})'$, $\boldsymbol{\xi}_t^C = (\boldsymbol{\xi}_t^{(1)'}, \boldsymbol{\xi}_t^{(2)'})'$, $\boldsymbol{\epsilon}_t^C = (\boldsymbol{\epsilon}_t^{(1)'}, \boldsymbol{\epsilon}_t^{(2)'})'$, and

$$\mathbf{S}_t^C = \begin{pmatrix} \mathbf{S}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t^{(2)} \end{pmatrix}.$$

Given all the data from both instruments at time $t-1$, \mathbf{Z}_{t-1}^C , and at time t , \mathbf{Z}_t^C , the STDF algorithm is:

- 1) Estimate $\mathbf{K}_{t,t}$, $\mathbf{K}_{t-1,t-1}$, and $\mathbf{K}_{t,t-1}$ where

$$\mathbf{K}_{t_1,t_2} = \text{Cov}(\boldsymbol{\eta}_{t_1}^C, \boldsymbol{\eta}_{t_2}^C).$$

- 2) Compute the state transition matrix,

$$\mathbf{H}_t = \mathbf{K}_{t,t-1}' \mathbf{K}_{t,t}^{-1}.$$

- 3) Estimate $\mathbf{U}_t = \mathbf{K}_{t,t} - \mathbf{H}_t \mathbf{L}_t$, where $\mathbf{L}_t = \mathbf{K}_{t+1,t}$.
- 4) The one-step-ahead forecast is

$$\boldsymbol{\eta}_{t|t-1}^C \equiv E(\boldsymbol{\eta}_t^C | \mathbf{Z}_{t-1}^C) = \mathbf{H}_t \boldsymbol{\eta}_{t-1|t-1}^C.$$

5) Estimate the forecasted prediction error matrix,

$$\begin{aligned} \mathbf{P}_{t|t-1} &= E \left[\left(\boldsymbol{\eta}_{t|t-1}^C - \boldsymbol{\eta}_t^C \right) \left(\boldsymbol{\eta}_{t|t-1}^C - \boldsymbol{\eta}_t^C \right)' \right] \\ &= \mathbf{H}_t \mathbf{P}_{t-1|t-1} \mathbf{H}_t' + \mathbf{U}_t. \end{aligned}$$

6) The Kalman gain matrix is

$$\mathbf{G}_t = \mathbf{P}_{t|t-1} (\mathbf{S}_t^C)' \left[\mathbf{S}_t^C \mathbf{P}_{t|t-1} (\mathbf{S}_t^C)' + \mathbf{D}_t \right]^{-1},$$

where $\mathbf{D}_t = \text{Var} \left(\boldsymbol{\xi}_t^C + \boldsymbol{\epsilon}_t^C \right)$.

7) The updated forecast is

$$\boldsymbol{\eta}_{t|t}^C = \boldsymbol{\eta}_{t|t-1}^C + \mathbf{G}_t \left(\mathbf{Z}_t^C - \boldsymbol{\mu}_t^C - \mathbf{S}_t^C \boldsymbol{\eta}_{t|t-1}^C \right).$$

8) The updated forecast prediction error matrix is,

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{G}_t \mathbf{S}_t^C \mathbf{P}_{t|t-1}.$$

9) Compute the updated estimate of $\boldsymbol{\xi}_{t|t}^C$ at the target location \mathbf{s} :

$$\begin{aligned} \boldsymbol{\xi}_{t|t}^C(\mathbf{s}) &= \mathbf{c}_t^C(\mathbf{s})' \left[(\mathbf{S}_t^C)' \mathbf{P}_{t|t-1} \mathbf{S}_t^C - \mathbf{D}_t \right]^{-1} \\ &\quad \times \left(\mathbf{Z}_t^C - \boldsymbol{\mu}_t^C - \mathbf{S}_t^C \boldsymbol{\eta}_{t|t-1}^C \right), \end{aligned}$$

where $\mathbf{c}_t^C(\mathbf{s}) = \text{Cov} \left(\mathbf{Z}_t^C, \boldsymbol{\xi}_{t|t}^C(\mathbf{s}) \right)$.

10) The optimal estimate of $\mathbf{Y}(\mathbf{s}, t) = (Y_1(\mathbf{s}, t), Y_2(\mathbf{s}, t))'$ is

$$\hat{\mathbf{Y}}(\mathbf{s}, t) = \left(\boldsymbol{\mu}_t^C(\mathbf{s}) + \mathbf{S}_t^C(\mathbf{s}) \boldsymbol{\eta}_{t|t}^C + \boldsymbol{\xi}_{t|t}^C(\mathbf{s}) \right). \quad (15)$$

11) The mean squared error matrix for $\hat{\mathbf{Y}}(\mathbf{s}, t)$ is

$$\begin{aligned} \text{Var}(\hat{\mathbf{Y}}(\mathbf{s}, t)) &= \mathbf{S}_t^C(\mathbf{s})' \mathbf{P}_{t|t} \mathbf{S}_t^C(\mathbf{s}) + \boldsymbol{\sigma}_\xi^2 \\ &\quad - \mathbf{c}_t^C(\mathbf{s})' \left[\mathbf{S}_t^C \mathbf{P}_{t|t-1} \mathbf{S}_t^C - \mathbf{D}_t \right]^{-1} \mathbf{c}_t^C(\mathbf{s}) \\ &\quad - 2 \mathbf{S}_t^C(\mathbf{s})' \mathbf{K}_{t,t} \mathbf{S}_t^C \boldsymbol{\Sigma}_{t,t}^{-1} \mathbf{c}_t^C(\mathbf{s}), \end{aligned} \quad (16)$$

where $\boldsymbol{\Sigma}_{t,t} = \text{Var}(\mathbf{Z}_t^C)$, and $\boldsymbol{\sigma}_\xi^2$ is a 2×2 diagonal matrix with $\text{Var}(\xi^{(k)})$ on the diagonal.

IV. ESTIMATING LOWER ATMOSPHERE CO₂

We performed multiple process, multiple source, space-time data fusion to estimate the vector of process values of total column and mid-tropospheric CO₂ on a half-degree grid over the continental US from February 2010 through December 2010 in two-week blocks. $\mathbf{Y}(\mathbf{s}, t) = (Y_1(\mathbf{s}, t), Y_2(\mathbf{s}, t))'$, here $Y_1(\mathbf{s}, t)$ is total column CO₂ at location \mathbf{s} and time t , and $Y_2(\mathbf{s}, t)$ is mid-tropospheric CO₂ at location \mathbf{s} and time t . The estimate of lower atmosphere CO₂ is a simple linear function of this vector:

$$\hat{\Delta}(\mathbf{s}, t) = (1 \quad -1) \hat{\mathbf{Y}}(\mathbf{s}, t). \quad (17)$$

The uncertainty of the estimate $\hat{\Delta}(\mathbf{s}, t)$ is its standard error,

$$\sigma(\hat{\Delta}(\mathbf{s}, t)) = \sqrt{(1 \quad -1) \text{Var}(\hat{\mathbf{Y}}(\mathbf{s}, t)) (1 \quad -1)'}. \quad (18)$$

Our total column CO₂ data come from the GOSAT instrument, which provides observations on footprints 10 km

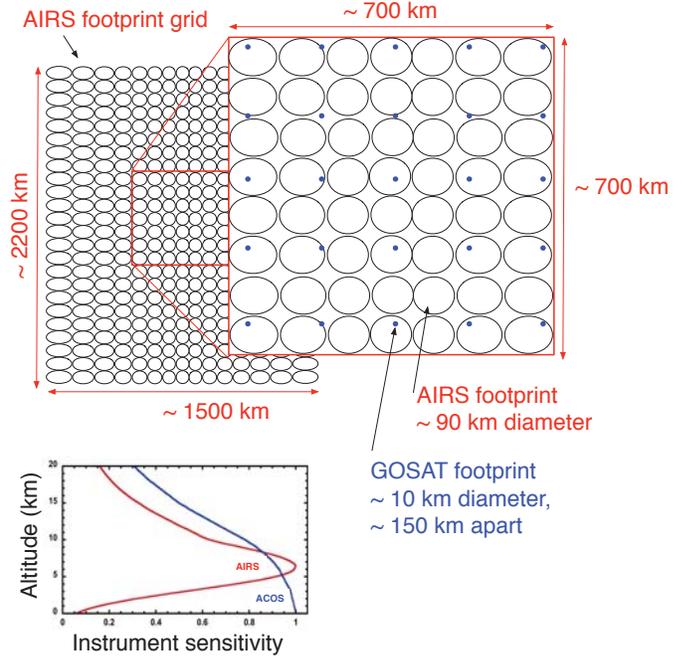


Fig. 5. AIRS and GOSAT horizontal footprints and vertical measurement characteristics. In the instrument sensitivity graph, GOSAT is labeled “ACOS” since JPL’s Atmospheric Carbon Dioxide Observations from Space team actually produced the data used here.

in diameter and spaced roughly 150 km apart. GOSAT observes only over land. Mid-tropospheric CO₂ data come from NASA’s AIRS instrument, which observes on 90 km diameter footprints that are spaced roughly 90 km apart over both land and ocean. Figure 5 is a schematic diagram of the footprint geometries of the two instruments. GOSAT was designed with a very small footprint in order to maximize the number of cloud-free footprints. AIRS has a very large footprint by comparison, and it achieves near-global coverage every two days. These two geometries complement each other, as do the sensitivities of the two instruments to different parts of the column shown at the bottom of Figure 5. The primary difference in sensitivities is in the lower part of the atmosphere, hence the logic that suggests differencing them to estimate lower atmosphere CO₂.

We organized the data into 21, 15-day periods indexed by $t = 1, 2, \dots, 21$ and aggregated all data for period t for GOSAT into $\mathbf{Z}_t^{(1)}$ and all data for AIRS into $\mathbf{Z}_t^{(2)}$. The bias for GOSAT was set to 20 parts per million (ppm) after some consultation and experimentation (discussed below). The bias for AIRS was determined to be zero on the basis of our own comparison of AIRS CO₂ retrievals to validation data provided by the AIRS team. Measurement error variances for the two data sets were estimated along with the other parameters of the underlying space-time model discussed earlier. The multiresolution spatial basis set, \mathbf{S} , has three levels of resolution and 386 multiresolution centers. We produced estimates of $\mathbf{Y}(\mathbf{s}, t)$ per Equation (15) for point locations \mathbf{s} spaced every half-degree of latitude and longitude over the continental US

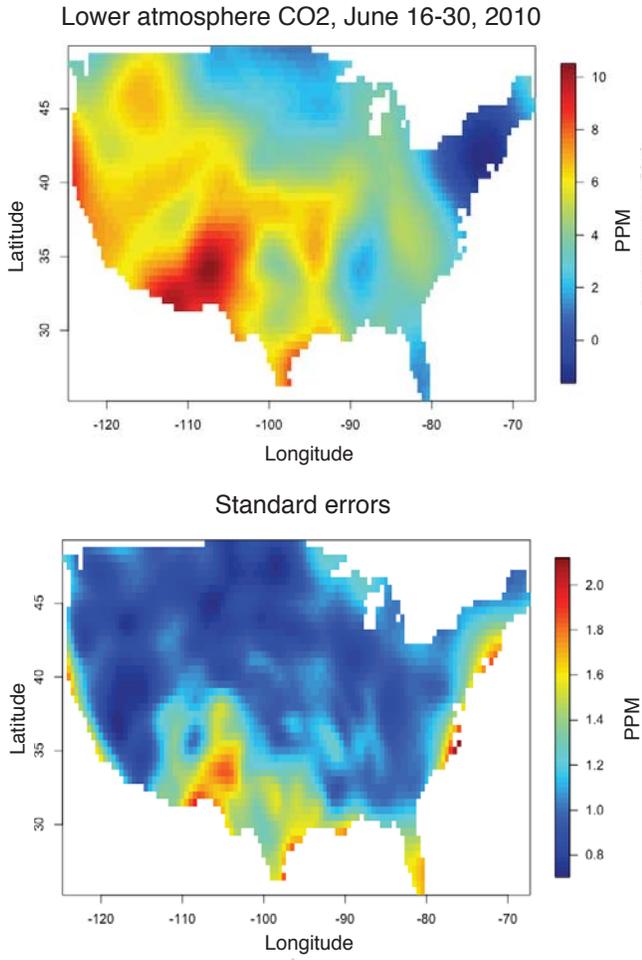


Fig. 6. Results of space-time data fusion to estimate lower atmosphere CO₂ for one time period. The top panel are the estimates, and the bottom panel are the associated uncertainties.

for each of the 21 time points. The error covariance matrix, Equation (16), is also computed for each estimate. All that remains is to compute $(\hat{Y}_1(\mathbf{s}, t) - \hat{Y}_2(\mathbf{s}, t))$ from Equation (17) and its standard error from Equation (18).

Figure 6 shows the results for one representative time period, with the estimates, $\hat{\Delta}(\mathbf{s}, t)$, in the top panel and the corresponding uncertainties, $\sigma(\hat{\Delta}(\mathbf{s}, t))$ in the bottom panel. Several features are obvious. First, is the prominent hotspot over New Mexico, but notice also that these estimates have high uncertainties. The estimate at the center of the hotspot region is about 10 parts per million (PPM), with a standard error of 2 PPM. A 95 percent confidence interval for lower atmosphere CO₂ is then $[10 - 1.96(2), 10 + 1.96(2)]$. The lower bound of this interval is near 6 PPM and so would not appear to be quite so outstanding. Second, there is a strong southwest-northeast gradient that is consistent with what is known about the effect of net ecosystem exchange: CO₂ is emitted as green plants reach the peak of their photosynthetic activity in the summer. Finally, note that the standard errors tend to be higher along the coasts since, with land data only,

there is less neighboring data with which to work.

V. CONCLUSION

We have demonstrated that STDF can be used to leverage both spatial and temporal dependence to estimate a function of two spatially continuous geophysical fields from noisy observations with different statistical characteristics. The maps in Figure 6 look like they may provide reasonable estimates, but these have yet to be validated against independent in-situ observations. It is also worth emphasizing that the validity of both the estimates and uncertainties depends on the means and standard deviations of the measurement error distributions and on other modeling choices discussed earlier. In this exercise, we used measurement-error statistics based on the judgment and experience of members of the instrument teams. A more rigorous analysis will ultimately be required as will a careful evaluation of the sensitivities of our results to the other modeling assumptions.

Near-term methodological improvements center on reducing the duration of a time step in the STDF analysis. Currently, our method aggregates data over 15 days because the GOSAT data are sparse, and estimates of statistical model parameters are unstable with fewer observations. However, CO₂ transport occurs on shorter time scales, and the science community would prefer time steps on the order of three days. We have used the method of moments to estimate model parameters here, but we are investigating expectation maximization (EM) as a more stable alternative. We are also beginning the process of validating our lower atmosphere CO₂ estimates by comparing them to in-situ observations with the help of the instrument validation teams.

ACKNOWLEDGMENT

We would like to thank NASA's Earth Science Technology Office for supporting this work under its Advanced Information Systems Technology Program.

The research described in this article was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Copyright 2011. All rights reserved. Government sponsorship acknowledged.

REFERENCES

- [1] H. Nguyen, N. Cressie, and A. Braverman (2010), Spatial-statistical data fusion for remote sensing applications, Department of Statistics Technical Report No. 849, The Ohio State University. Submitted to the Journal of the American Statistical Association.
- [2] H.V. Henderson and S.R. Searle (1981). On deriving the inverse of a sum of matrices, *SIAM Review*, Vol. 23, No. 1, pp. 53-60.
- [3] H. Nguyen (2009). Spatial Statistical Data Fusion for Remote Sensing Applications, Ph.D. Dissertation, Department of Statistics, UCLA. Available at http://directory.stat.ucla.edu/alumni_profile.php?directory=365.