

Spatio-Temporal Data Fusion for Very Large Remote Sensing Datasets

Hai NGUYEN

Statistician
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA
(hai.nguyen@jpl.nasa.gov)

Matthias KATZFUSS

Postdoctoral Researcher at Institut für Angewandte Mathematik
Universität Heidelberg
Germany
(katzfuss@gmail.com)

Noel CRESSIE

Distinguished Professor, NIASRA
School of Mathematics and Applied Statistics
University of Wollongong, Wollongong NSW 2522
Australia
(ncressie@uow.edu.au)

Amy BRAVERMAN

Senior Statistician
Jet Propulsion Laboratory
California Institute of Technology
(amy.braverman@jpl.nasa.gov)

Developing global maps of carbon dioxide (CO₂) mole fraction (in units of parts per million) near the Earth's surface can help identify locations where major amounts of CO₂ are entering and exiting the atmosphere, thus providing valuable insights into the carbon cycle and mitigating the greenhouse effect of atmospheric CO₂. Existing satellite remote sensing data do not provide measurements of the CO₂ mole fraction near the surface. Japan's Greenhouse gases Observing SATellite (GOSAT) is sensitive to average CO₂ over the entire column, and NASA's Atmospheric InfraRed Sounder (AIRS) is sensitive to CO₂ in the middle troposphere. One might expect that lower-atmospheric CO₂ could be inferred by differencing GOSAT column-average and AIRS mid-tropospheric data. However, the two instruments have different footprints, measurement-error characteristics, and data coverages. In addition, the spatio-temporal domains are large, and the AIRS dataset is massive. In this article, we describe a spatio-temporal data-fusion (STDF) methodology based on reduced-dimensional Kalman smoothing. Our STDF is able to combine the complementary GOSAT and AIRS datasets to optimally estimate lower-atmospheric CO₂ mole fraction over the whole globe. Further, it is designed for massive remote sensing datasets and accounts for differences in instrument footprint, measurement-error characteristics, and data coverages. This article has supplementary material online.

KEY WORDS: EM algorithm; Fixed rank smoothing; Kalman filter; Multivariate geostatistics; Spatial random effects model.

1. INTRODUCTION

Climate forecasting is an important research topic because of its implications for political, social, and scientific decision making. One area of active research represents the behavior of the atmosphere through general circulation models, which approximate the atmospheric circulation based on equations describing motion of fluids and the input of thermodynamic energy sources such as solar radiation and latent heat (e.g., McGuffie and Henderson-Sellers 1997).

About 8 gigatons (Gt) of carbon dioxide (CO₂) per year enters the atmosphere, and about half of this is anthropogenic. While roughly 4 Gt is absorbed by the ocean and terrestrial processes, there is a yearly increase of about 4 Gt of atmospheric CO₂. Its mole fraction (in units of parts per million – ppm) is one of the most important components for modeling and estimating the climate of the 21st century (Houghton et al. 2001), and a large number of experiments with comprehensive ocean-atmosphere general circulation models (OAGCMs or GCMs) prescribe CO₂-mole-fraction scenarios using relatively simple offline carbon cycle models (Friedlingstein et al. 2006). The

accuracies of the GCM predictions depend on the fidelity of the prescribed CO₂ scenarios with respect to current and future climatic conditions. Here, remote sensing of atmospheric CO₂ from satellites provides a valuable resource for improving understanding and characterization of these CO₂ scenarios.

Carbon dioxide is a naturally occurring chemical compound consisting of one carbon atom and two oxygen atoms, and it cycles in and out of a variety of Earth's "compartments." For example, it is one of the key inputs in photosynthesis, the process used by plants and other organisms to convert light energy from the sun into chemical energy, and it is one of the byproducts in the reverse biological process of respiration. In the atmosphere, CO₂ acts as a greenhouse gas, and its increase since the late 19th century is believed to be playing an important role in global warming (Houghton et al. 2001). In water, CO₂ dissolves to

form carbonic acid, which contributes to ocean acidification and poses a threat to food chains connected to the oceans.

The exchange of CO₂ between the atmosphere and the Earth's surface is a critical part of the global carbon cycle and an important determinant of future climate (Gruber et al. 2009). A carbon sink is a natural or anthropogenic reservoir that sequesters CO₂ from the atmosphere (e.g., photosynthesis by ocean organisms or by terrestrial plants). In contrast, a carbon source releases CO₂ into the atmosphere; an example of a natural source is plant respiration, and examples of anthropogenic sources are land-clearing for agriculture and fossil-fuel burning.

The key to understanding the atmosphere–surface CO₂ exchange is the lower-atmospheric CO₂ mole fraction. Deriving its global distribution over time is important for studying places where high amounts of CO₂ are being generated and removed. However, there is no global, remote-sensing-based map of lower-atmospheric CO₂. Past estimates of CO₂ sources and sinks have relied on ground-based data, whose locations tend to be sparsely distributed around the globe. Consequently, estimates over continental-scale areas such as Siberia, Asia, Africa, South America, and the oceans, which have particularly poor coverage, have large errors. Remote sensing data offer much better coverage, although none of the instruments has sensitivity to CO₂ in the lower atmosphere.

One approach to estimate lower-atmospheric CO₂ is called flux inversion that combines a priori knowledge of sources and sinks, a chemistry and transport model, and satellite CO₂ observations to estimate lower-atmospheric CO₂ (e.g., Chevallier et al. 2005). We take a spatial-statistical approach and combine data from the Greenhouse gases Observing SATellite (GOSAT) and data from the Atmospheric InfraRed Sounder (AIRS) instruments. These two instruments have different sensitivities in the atmospheric column, which make inferences on lower atmospheric CO₂ in principle possible. The purpose of this article is to use spatial and spatio-temporal statistics to predict lower-atmospheric CO₂ by fusing data from the GOSAT and AIRS instruments, accompanied by uncertainty quantification of those predictions.

GOSAT is a polar-orbiting satellite dedicated to the observation of carbon dioxide and methane, both major greenhouse

gases, from space. It flies at approximately 665 kilometers (km) altitude, and it completes an orbit every 100 min. The satellite returns to the same observation location every 3 days (Morino et al. 2011). NASA's Atmospheric CO₂ Observations from Space (ACOS) team uses the raw-radiance data from GOSAT to estimate the column-average CO₂ mole fraction in ppm, extending from the surface to the satellite over a base area corresponding to the instrument's footprint. In this article, we will be using GOSAT retrievals that are processed by the ACOS team to yield Level 2 column-average CO₂ data (see Crisp et al. 2012, for more details), which were available to us through NASA's Goddard Earth Sciences Data and Information Services Center. *Hereafter, we refer to these as ACOS data.*

AIRS is a high-resolution, infrared-spectrometer instrument aboard NASA's Aqua satellite that flies in sun-synchronous orbit at an altitude of approximately 705 km (Aumann et al. 2003). It completes one orbit every 99 min and returns to the same observation location every 16 days. AIRS measures several geophysical quantities such as air and surface temperature, water vapor, and cloud properties, along with greenhouse gases such as ozone, carbon monoxide, carbon dioxide, and methane. In this article, we make use of the AIRS CO₂ product, which is the mole fraction of CO₂ over the mid-tropospheric segment of the atmospheric column (see Chahine et al. 2008, for more details). *Hereafter, we refer to these as AIRS data.*

AIRS observations have circular footprints with 45-km radii, while ACOS observations have circular footprints with 5-km radii. Figure 1 is a schematic diagram of the observational patterns of ACOS and AIRS under ideal conditions; the left and middle panels display the spatial change-of-support issue at hand, namely the vastly different footprint sizes of the two instruments. The right panel displays the instruments' sensitivities to different parts of the atmosphere. The sensitivities differ mostly in the lower atmosphere, and so lower-atmospheric CO₂ can be approximated by (weighted) differencing of the ACOS column-average CO₂ mole fraction and the AIRS mid-tropospheric CO₂ mole fraction (Section 3.1). These two instruments measure their respective physical processes using different technologies, fields-of-view, and retrieval algorithms, which

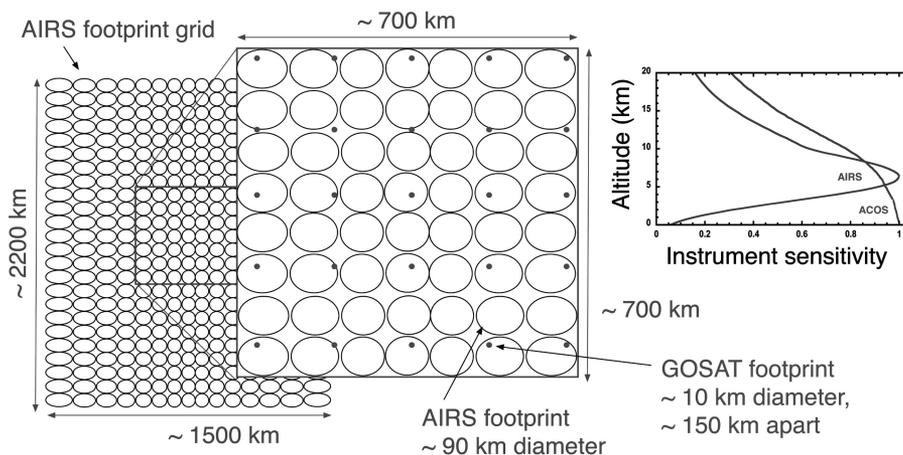


Figure 1. Left and middle panels: GOSAT and AIRS sampling footprints. Right panel: CO₂ sensitivities of GOSAT and AIRS to different parts of the atmosphere.

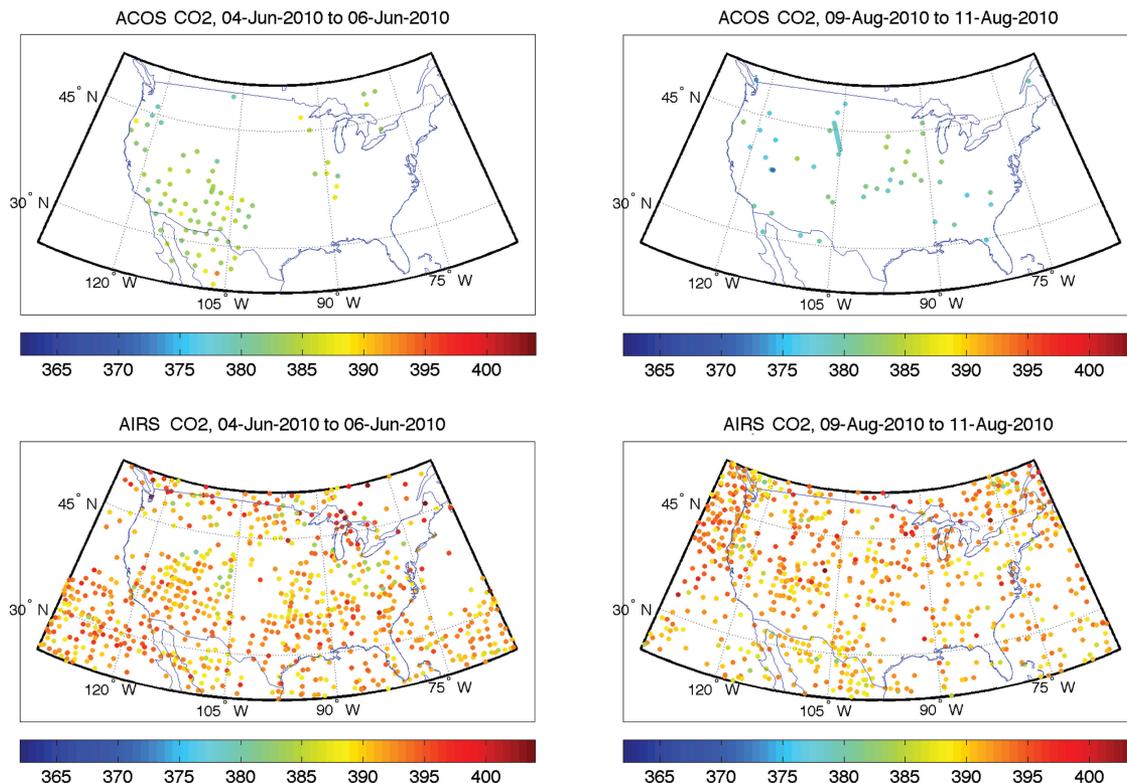


Figure 2. Two 3-day blocks of ACOS data (top row) and AIRS data (bottom row). Shown are data for the three-day block June 4–6, 2010 (left column) and for the three-day block August 9–11, 2010 (right column). The units of measurement are parts per million (ppm). The displayed data-locations represent the centers of the footprint and are *not* scaled according to footprint size.

lead to coverage differences that can both be complementary and reinforcing.

1.1 GOSAT and AIRS Data

In this article, we carry out data fusion on ACOS data and AIRS data over the contiguous United States during the Boreal summer of 2010. The ACOS and AIRS datasets we analyze are located in a region that extends from 25°N latitude to 50°N latitude and from 132°W longitude to 65°W longitude, over the 3-month time period June 1, 2010 to August 30, 2010. We chose this domain because we have ground-based, lower-atmospheric validation data from the National Oceanic and Atmospheric Administration Carbon Cycle Greenhouse Gases (NOAA CCGG) aircraft program within this same region and time period. These aircraft programs collect in situ flask samples of trace gases at different altitudes, from which we can compare our approximations of lower-atmospheric CO₂ mole fraction to these data.

Over any 3-day block, GOSAT takes 56,000 measurements over the globe. However, only 2% to 5% of the data collected are usable since retrievals are limited to clear-sky conditions. The resulting ACOS data are classified into several categories, depending on quality. We only included measurements in the highest-quality category, based on a data-quality filter provided by the ACOS team (G. Osterman, personal communication, February, 2011). Within the spatio-temporal domain described just above, we obtained 3869 ACOS data points and 40,564 AIRS data points. We partitioned the data over these 3 months into 3-day blocks. There are 30 three-day blocks over the three summer

months in 2010, with an average of 128.1 observations per block for ACOS and 1352.1 observations per block for AIRS. Figure 2 shows the ACOS and AIRS data for two such blocks.

The ACOS data are much sparser than the AIRS data, due to different instrument designs and retrieval methodologies; ACOS typically has incomplete coverage of the United States over each 3-day block, while AIRS has reasonably complete coverage. The coverage of the ACOS and AIRS data in Figure 2, when compared to the regular sampling patterns in Figure 1, is uneven due to the presence of clouds and other atmospheric conditions that result in less-than-complete retrievals of CO₂.

1.2 Reviews of Statistical Data Fusion

We are considering two (in principle, many) geophysical processes, whose data we wish to fuse. To exploit the temporal, spatial, and cross-process dependence, any remote sensing data-fusion methodology must overcome two basic difficulties: the potential massiveness of the data and the different footprints of the instruments (i.e., different spatial supports).

Recent spatial and spatio-temporal inferential methodologies that are scalable in data size include those due to Berliner, Wikle, and Milliff (1999; hierarchical Bayesian spatio-temporal model with multiresolution wavelet basis functions and two data sources of different support), Wikle et al. (2001; more general than Berliner, Wikle, and Milliff, 1999, with science-based orthogonal eigenfunctions and multiresolution basis functions to capture residual dependencies), Nychka, Wikle, and Royle (2002; modeling nonstationary covariance functions with

multiresolutional wavelet models), Hooten, Larsen, and Wikle (2003; hierarchical Bayesian model with FFT representation of spatial random effects), Royle and Wikle (2005; spectral parameterization of the spatial Poisson process), Banerjee et al. (2008; approximate optimal prediction with dimension reduction through conditioning on a small set of space-filling locations), Calder (2008; bivariate dynamic process convolution model), Cressie and Johannesson (2008; Fixed Rank Kriging based on the Spatial Random Effects model), Stein and Jun (2008; modeling nonstationary covariance models using the discrete Fourier transform), Lindgren, Rue, and Lindstrom (2011; linking Gaussian fields and Gaussian Markov random fields using stochastic partial differential equations), and Cressie, Shi, and Kang (2010; fixed rank filtering and fixed ranked smoothing based on the Kalman filter and the spatio-temporal random effects model). In this article, we generalize the latter article's approach for a single data source, to data fusion for multiple spatio-temporal data sources.

Cressie, Shi, and Kang (2010) used binned method-of-moments to estimate the spatio-temporal random effects parameters, whereas Katzfuss and Cressie (2011) derived maximum likelihood estimates of such parameters via the expectation-maximization (EM) algorithm. In the spatial setting, Katzfuss and Cressie (2009) demonstrated that their EM estimators are more stable and accurate than the corresponding binned method-of-moments estimators. Consequently, in this article, we pursue spatio-temporal data fusion where the parameters are estimated via the EM algorithm.

In Section 2, we describe a spatio-temporal data-fusion methodology that uses the spatio-temporal random effects model to deal with issues of both big datasets and heterogeneous spatial supports, while incorporating temporal dependence; an EM algorithm is developed to estimate the model's parameters, and we call the final result spatio-temporal data fusion (STDF). In Section 3, we make use of STDF to solve the main problem, namely to estimate lower-atmospheric CO₂ from ACOS and AIRS remote sensing data. We also compare the performance of STDF against a standard NASA methodology. In Section 4, we discuss our findings and possible extensions of STDF, and an Appendix gives the details of our STDF smoothing equations. Additional details on the spatial and spatio-temporal data-fusion methodology are provided in supplementary materials online, along with a zip file of the data we analyze.

2. THE SPATIO-TEMPORAL STATISTICAL MODEL

In this section, we briefly review the spatial statistical framework, give some necessary notation, and present basic derivations for predictions and for maximum likelihood estimation (via the EM algorithm) of our spatio-temporal model's parameters.

2.1 Data Model and Properties

Let $\{Y_t^{(k)}(\mathbf{s}) : \mathbf{s} \in D\}$ be the k th hidden, real-valued spatial process of interest on a *discretized* domain D at time t , where $t = 1, \dots, T$, and without loss of generality we assume $k \in \{1, 2\}$. The spatial domain of interest is written mathematically as $\cup\{S_i \subset \mathbb{R}^d : i = 1, \dots, N_D\}$, which is made up of N_D pre-

specified, fine-scale, nonoverlapping, basic areal units (BAU's) $\{S_i\}$, with respective locations $D \equiv \{\mathbf{p}_i \in S_i : i = 1, \dots, N_D\}$. For example, the set of BAUs could be a set of tiling hexagons, and D could be the hexagons' centroids. One could think of them as the finest possible resolution of scientific interest; while the spatio-temporal random effects model that we shall use below is invariant to their choice, the discretization of the spatial domain is a fundamental step.

Let $\mathbf{Z}_t^{(k)}$ be the vector of noisy observations on $Y_t^{(k)}(\cdot)$ taken by the k th remote sensing instrument at $N_t^{(k)}$ footprints $\{A_{i,t}^{(k)} : i = 1, \dots, N_t^{(k)}\}$ at time t , where a generic footprint A can be expressed as the union of those BAUs whose locations are indexed by $D \cap A$. The observed value over a footprint A by instrument k at time t is modeled as the average of the true process $Y_t^{(k)}(\cdot)$ over the BAUs within that footprint, plus measurement error:

$$Z_t^{(k)}(A) = \frac{1}{|D \cap A|} \left\{ \sum_{\mathbf{s} \in D \cap A} Y_t^{(k)}(\mathbf{s}) \right\} + \epsilon_t^{(k)}(A);$$

$A \subset \mathbb{R}^d, k = 1, 2. \quad (1)$

The measurement-error term, $\epsilon_t^{(k)}(A)$, may have nonzero mean that captures the instrument bias, and it has measurement-error variance $(\sigma_{\epsilon,t}^{(k)})^2 v_t^{(k)}(A) > 0$, where $v_t^{(k)}(\cdot)$ is known and allows for the possibility of nonconstant variance over the domain D . We assume that the measurement-error processes $\epsilon_t^{(1)}(\cdot)$ and $\epsilon_t^{(2)}(\cdot)$ are independent of one another and of $(Y_t^{(1)}(\cdot), Y_t^{(2)}(\cdot))$, and that the measurement-error variances, $\{(\sigma_{\epsilon,t}^{(k)})^2 : k = 1, 2\}$, are known; in practice, these variances are obtained from validation data and/or instrument specification. If unknown, they can be estimated by examining empirical variograms and extrapolating to the origin, such as in Kang, Liu, and Cressie (2009).

Our data model given by (1) can be compared to that presented in Wikle (2003) and Wikle and Berliner (2005). We have possibly nonzero-mean measurement errors, but a very simple independent error structure; Wikle and Berliner's error model has zero mean and a change-of-support effect that exhibits spatial correlation.

The k th true process at time t , namely $Y_t^{(k)}(\cdot)$, is assumed to have a linear mean structure and two components of spatio-temporal statistical dependence:

$$Y_t^{(k)}(\mathbf{s}) = \mathbf{x}_t^{(k)}(\mathbf{s})' \boldsymbol{\alpha}_t^{(k)} + \mathbf{S}_t^{(k)}(\mathbf{s})' \boldsymbol{\eta}_t^{(k)} + \xi_t^{(k)}(\mathbf{s}); \quad \mathbf{s} \in D, \quad (2)$$

where we now describe each component of the right-hand side of (2). The first term is not random and assumes a linear model in $p_t^{(k)}$ covariates, $\mathbf{x}_t^{(k)}(\cdot) \equiv (x_{a,t}^{(k)}(\cdot) : a = 1, \dots, p_t^{(k)})'$, where the regression-coefficient vector $\boldsymbol{\alpha}_t^{(k)}$ is to be estimated. The middle term, $\mathbf{S}_t^{(k)}(\cdot)' \boldsymbol{\eta}_t^{(k)}$, captures the smooth spatial dependence and is expressed as the inner product of an $r_t^{(k)}$ -dimensional vector of known spatial basis functions, $\mathbf{S}_t^{(k)}(\cdot)$, and an $r_t^{(k)}$ -dimensional Gaussian random variable, $\boldsymbol{\eta}_t^{(k)} \sim N(\mathbf{0}, \mathbf{K}_t^{(k)})$; see Cressie and Johannesson (2008). We also assume that the random effects, $\boldsymbol{\eta}_t^{(1)}$ and $\boldsymbol{\eta}_t^{(2)}$, are jointly normal and that $\text{cov}(\boldsymbol{\eta}_t^{(k)}, \boldsymbol{\eta}_t^{(l)}) \equiv \mathbf{K}_t^{(k,l)}$, where we write $\mathbf{K}_t^{(k)} \equiv \mathbf{K}_t^{(k,k)}$. The last term in (2), $\xi_t^{(k)}(\cdot)$, is made up of spatially and temporally independent Gaussian variables with mean zero and variance $(\sigma_{\xi,t}^{(k)})^2$. We assume that $\xi_t^{(1)}(\cdot)$ is

independent of $\xi_t^{(2)}(\cdot)$ and of $\eta_t^{(2)}$; similarly, we assume that $\xi_t^{(2)}(\cdot)$ is independent of $\eta_t^{(1)}$.

To allow for the possibility of instrument bias, we assume that the measurement-error process, $\epsilon_t^{(k)}(\cdot)$, satisfies $E(\epsilon_t^{(k)}(A)) = c^{(k)}E(Y_t^{(k)}(A))$, where zero bias is captured by $c^{(k)} = 0$. The multiplicative bias coefficients $\{c^{(k)} : k = 1, 2\}$ are assumed known, typically from validation experiments or from comparison with independent, unbiased data sources.

Combining Equations (1) and (2), we can assemble the scalars into column vectors and the row vectors into matrices to form a spatial linear mixed effects model:

$$\mathbf{Z}_t^{(k)} = \mathbf{X}_t^{(k)} \boldsymbol{\alpha}_t^{(k)} + \mathbf{S}_t^{(k)} \boldsymbol{\eta}_t^{(k)} + \boldsymbol{\xi}_t^{(k)} + \boldsymbol{\epsilon}_t^{(k)}; \quad k = 1, 2,$$

where $\mathbf{S}_t^{(k)}$ is an $N_t^{(k)} \times r_t^{(k)}$ matrix with m th row given by the $r_t^{(k)}$ -dimensional vector $\mathbf{S}_t^{(k)}(A_{m,t}^{(k)})'$, $\boldsymbol{\xi}_t^{(k)} \equiv (\xi_t^{(k)}(A_{i,t}^{(k)})) : i = 1, \dots, N_t^{(k)}$, and the other terms are defined analogously. It is important to note that while the functions $\mathbf{x}_t^{(k)}(\cdot)$, $\mathbf{S}_t^{(k)}(\cdot)$, and $\xi_t^{(k)}(\cdot)$ were originally defined at the BAU level, their definitions over any footprint A are given by

$$\begin{aligned} \mathbf{x}_t^{(k)}(A) &\equiv \frac{1}{|D \cap A|} \sum_{\mathbf{s} \in D \cap A} \mathbf{x}_t^{(k)}(\mathbf{s}) \\ \mathbf{S}_t^{(k)}(A) &\equiv \frac{1}{|D \cap A|} \sum_{\mathbf{s} \in D \cap A} \mathbf{S}_t^{(k)}(\mathbf{s}) \\ \xi_t^{(k)}(A) &\equiv \frac{1}{|D \cap A|} \sum_{\mathbf{s} \in D \cap A} \xi_t^{(k)}(\mathbf{s}), \end{aligned}$$

and we can similarly define the process $Y(\cdot)$ over a footprint as

$$Y_t^{(k)}(A) \equiv \frac{1}{|D \cap A|} \sum_{\mathbf{s} \in D \cap A} Y_t^{(k)}(\mathbf{s}).$$

At time t , we can stack datasets $\mathbf{Z}_t^{(1)}$ and $\mathbf{Z}_t^{(2)}$ to form a vector of dimension $N_t \equiv N_t^{(1)} + N_t^{(2)}$:

$$\begin{aligned} \begin{pmatrix} \mathbf{Z}_t^{(1)} \\ \mathbf{Z}_t^{(2)} \end{pmatrix} &= \begin{pmatrix} \mathbf{X}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_t^{(2)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_t^{(1)} \\ \boldsymbol{\alpha}_t^{(2)} \end{pmatrix} + \begin{pmatrix} \mathbf{S}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t^{(2)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t^{(1)} \\ \boldsymbol{\eta}_t^{(2)} \end{pmatrix} \\ &+ \begin{pmatrix} \boldsymbol{\xi}_t^{(1)} \\ \boldsymbol{\xi}_t^{(2)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_t^{(1)} \\ \boldsymbol{\epsilon}_t^{(2)} \end{pmatrix}, \end{aligned}$$

or equivalently,

$$\mathbf{Z}_t = \mathbf{X}_t \boldsymbol{\alpha}_t + \mathbf{S}_t \boldsymbol{\eta}_t + \boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t, \quad (3)$$

where the dimension of the fixed but unknown vector $\boldsymbol{\alpha}_t$ is $p_t \equiv p_t^{(1)} + p_t^{(2)}$, the stacked random vectors $\boldsymbol{\eta}_t$, $\boldsymbol{\xi}_t$, and $\boldsymbol{\epsilon}_t$ are assumed to be independent of one another, and the dimension of $\boldsymbol{\eta}_t$ is $r_t \equiv r_t^{(1)} + r_t^{(2)}$.

The all-important temporal dependence is established by assuming that the mean-zero vectors $\{\boldsymbol{\eta}_t : t = 0, \dots, T\}$ follow a first-order vector-autoregressive process:

$$\boldsymbol{\eta}_t | \boldsymbol{\eta}_{t-1}, \dots, \boldsymbol{\eta}_0 \sim N_t(\mathbf{H}_t \boldsymbol{\eta}_{t-1}, \mathbf{U}_t); \quad t = 1, 2, \dots, \quad (4)$$

with initial state $\boldsymbol{\eta}_0 \sim N_{r_0}(\mathbf{0}, \mathbf{K}_0)$. The $r_t \times r_{t-1}$ matrix \mathbf{H}_t is called the propagator matrix, and the $r_t \times r_t$ covariance matrix \mathbf{U}_t is called the innovation matrix.

From Cressie, Shi, and Kang (2010), the spatio-temporal random effects (STRE) model used in this article relies on the

following bivariate mean-zero process:

$$\begin{pmatrix} \mathbf{S}_t^{(1)}(\cdot)' & \mathbf{0}' \\ \mathbf{0}' & \mathbf{S}_t^{(2)}(\cdot)' \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_t^{(1)} \\ \boldsymbol{\eta}_t^{(2)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_t^{(1)}(\cdot) \\ \boldsymbol{\xi}_t^{(2)}(\cdot) \end{pmatrix} \equiv \mathbf{S}_t(\cdot)' \boldsymbol{\eta}_t + \boldsymbol{\xi}_t(\cdot),$$

where $\mathbf{S}_t(\cdot)$ is $2 \times r_t$, $\{\boldsymbol{\eta}_t : t = 0, 1, 2, \dots\}$ evolves according to (4), $\boldsymbol{\xi}_t(\cdot)$ is independent of $\boldsymbol{\eta}_t$, and $\boldsymbol{\xi}_t^{(1)}(\cdot)$ and $\boldsymbol{\xi}_t^{(2)}(\cdot)$ are independent.

The STRE model has a remarkable change-of-support property (see the supplementary material, Section A.1) that allows the covariance matrix of the data vector \mathbf{Z}_t to be written in terms of the BAU-level parameters defined below (2):

$$\boldsymbol{\Sigma}_t \equiv \text{var}(\mathbf{Z}_t) = \mathbf{S}_t \mathbf{K}_t \mathbf{S}_t' + \mathbf{C}_t \mathbf{E}_t + \mathbf{V}_t,$$

where $\mathbf{K}_t \equiv \text{var}(\boldsymbol{\eta}_t)$, $\mathbf{V}_t \equiv \text{var}(\boldsymbol{\epsilon}_t)$, and

$$\mathbf{C}_t \mathbf{E}_t \equiv \text{var}(\boldsymbol{\xi}_t) = \begin{pmatrix} \mathbf{C}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_t^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{E}_t^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_t^{(2)} \end{pmatrix}, \quad (5)$$

for $N_t^{(k)} \times N_t^{(k)}$ matrices $\mathbf{C}_t^{(k)} \equiv (\sigma_{\xi,t}^{(k)2} \mathbf{I}_{N_t^{(k)}})$ and $\mathbf{E}_t^{(k)} \equiv [\frac{|D \cap A_{i,t}^{(k)} \cap A_{j,t}^{(k)}|}{|D \cap A_{i,t}^{(k)}| |D \cap A_{j,t}^{(k)}|} : i, j = 1, \dots, N_t^{(k)}]$. From Cressie and Johannesson (2008), the inverse of the covariance matrix $\boldsymbol{\Sigma}_t$ can be computed rapidly via the Sherman-Morrison-Woodbury formula (e.g., Henderson and Searle 1981):

$$\boldsymbol{\Sigma}_t^{-1} = \mathbf{D}_t^{-1} - \mathbf{D}_t^{-1} \mathbf{S}_t [\mathbf{K}_t^{-1} + \mathbf{S}_t' \mathbf{D}_t^{-1} \mathbf{S}_t]^{-1} \mathbf{S}_t' \mathbf{D}_t^{-1}, \quad (6)$$

where $\mathbf{D}_t \equiv \mathbf{C}_t \mathbf{E}_t + \mathbf{V}_t$.

2.2 Spatio-Temporal Data Fusion (STDF)

Suppose that we are interested in predicting a stacked vector of the two processes $Y_t^{(1)}(\cdot)$ and $Y_t^{(2)}(\cdot)$ at a set of locations P (which may consist of areal and/or BAU prediction locations) at time $t \in \{1, \dots, T\}$, based on data $\mathbf{Z}_1, \dots, \mathbf{Z}_T$. Notice that we could allow P to depend on t , but here we choose not to for simplicity of exposition. Let \mathbf{S}_t^P , \mathbf{X}_t^P , and $\boldsymbol{\xi}_t^P$ represent the stacked vectors and stacked matrices derived by evaluating the corresponding terms in (3) at the set of n_P prediction locations P ; $t = 1, \dots, T$ (and let \mathbf{Y}_t^P be defined similarly). For example, if $\mathbf{Y}_t^{(k)P}$ and $\mathbf{S}_t^{(k)P}$ are vectors corresponding to the set of prediction locations, P , then \mathbf{Y}_t^P and \mathbf{S}_t^P are defined as

$$\mathbf{Y}_t^P \equiv \begin{pmatrix} \mathbf{Y}_t^{(1)P} \\ \mathbf{Y}_t^{(2)P} \end{pmatrix}, \quad \text{and} \quad \mathbf{S}_t^P \equiv \begin{pmatrix} \mathbf{S}_t^{(1)P} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t^{(2)P} \end{pmatrix}.$$

Consequently,

$$\mathbf{Y}_t^P \equiv \mathbf{X}_t^P \boldsymbol{\alpha}_t + \mathbf{S}_t^P \boldsymbol{\eta}_t + \boldsymbol{\xi}_t^P; \quad t = 1, \dots, T.$$

Let $\boldsymbol{\theta}$ denote the parameter values consisting of $\{\boldsymbol{\alpha}_t, \mathbf{K}_t, (\sigma_{\xi,t}^{(1)})^2, (\sigma_{\xi,t}^{(2)})^2, \mathbf{H}_t, \mathbf{U}_t : t = 1, \dots, T\}$. Assuming that $\boldsymbol{\theta}$ is known, optimal prediction of \mathbf{Y}_t^P is a result of optimal prediction of $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t^P$, jointly. Cressie, Shi, and Kang (2010) described a computationally efficient procedure to obtain the posterior expectations and covariances for $\{\boldsymbol{\eta}_t\}$ and $\{\boldsymbol{\xi}_t^P\}$. Their methodology, called Fixed Rank Smoothing (FRS), is an extension of the Kalman smoother and consists of two parts: forward-filtering

and backward-smoothing. A description of that methodology, with some modifications to account for the two *different* fine-scale variance parameters $\{(\sigma_{\xi,t}^{(k)})^2 : k = 1, 2\}$, may be found in the Appendix.

Having obtained the joint posterior distribution of $\boldsymbol{\eta}_t$ and $\boldsymbol{\xi}_t^P$, given $\mathbf{Z}_{1:T} \equiv \mathbf{Z}_1, \dots, \mathbf{Z}_T$, which is multivariate normal, then the posterior distribution of \mathbf{Y}_t^P is also multivariate normal. The posterior mean is a $2n_P$ -dimensional vector

$$\mathbf{Y}_{t|T}^P = \begin{pmatrix} \mathbf{Y}_{t|T}^{(1)P} \\ \mathbf{Y}_{t|T}^{(2)P} \end{pmatrix} = \mathbf{X}_t^P \boldsymbol{\alpha}_t + \mathbf{S}_t^P \boldsymbol{\eta}_{t|T} + \boldsymbol{\xi}_{t|T}^P, \quad (7)$$

where $\boldsymbol{\eta}_{t|T} \equiv E(\boldsymbol{\eta}_t | \mathbf{Z}_{1:T})$ and $\boldsymbol{\xi}_{t|T}^P \equiv E(\boldsymbol{\xi}_t^P | \mathbf{Z}_{1:T})$. The $2n_P \times 2n_P$ mean-squared-prediction-error matrix (which can be shown to be equal to the posterior covariance matrix) is

$$\mathbf{M}_{t|T}^P \equiv E\left([\mathbf{Y}_t^P - \mathbf{Y}_{t|T}^P][\mathbf{Y}_t^P - \mathbf{Y}_{t|T}^P]'\right) \equiv \begin{pmatrix} \mathbf{M}_{t|T}^{(1,1)P} & \mathbf{M}_{t|T}^{(1,2)P} \\ \mathbf{M}_{t|T}^{(2,1)P} & \mathbf{M}_{t|T}^{(2,2)P} \end{pmatrix}.$$

Then,

$$\mathbf{M}_{t|T}^P = \mathbf{S}_t^P \mathbf{P}_{t|T} (\mathbf{S}_t^P)' + \mathbf{R}_{t|T}^P + 2\mathbf{S}_t^P \mathbf{W}_{t|T}^P, \quad (8)$$

where $\mathbf{P}_{t|T} \equiv \text{var}(\boldsymbol{\eta}_t | \mathbf{Z}_{1:T})$, $\mathbf{R}_{t|T}^P \equiv \text{var}(\boldsymbol{\xi}_t^P | \mathbf{Z}_{1:T})$, and $\mathbf{W}_{t|T}^P \equiv \text{cov}(\boldsymbol{\eta}_t, \boldsymbol{\xi}_t^P | \mathbf{Z}_{1:T})$. We call (7) and (8) the spatio-temporal data fusion (STDF) equations.

Computation of the various smoothing quantities in this section requires inverting the large $N_t \times N_t$ matrix, $\mathbf{S}_t \mathbf{P}_{t|T-1} \mathbf{S}_t' + \mathbf{D}_t$, where $\mathbf{P}_{t|T-1} \equiv \text{var}(\boldsymbol{\eta}_t | \mathbf{Z}_{1:(t-1)})$ (see the Appendix for details), but from (6) the computational complexity of the inversion is $O(r_t^2 N_t)$. Therefore, STDF has computational complexity that is linear with respect to data size, making it well suited for remote sensing applications, where spatio-temporal datasets tend to be massive.

When there is only a single spatio-temporal dataset on a single process, then STDF in (7) and (8) reduces to FRS given by Cressie, Shi, and Kang (2010). A special case often encountered in practice consists of data that were observed during a single time period; that is, $T = 1$. The data in this special case may be considered to be spatial-only, for which a discussion is included in Section A of the supplementary material.

2.3 EM Algorithm for Parameter Estimation

In Section 2.2, we assumed that $\boldsymbol{\theta}$ the vector of parameters was known. In practice, it needs to be estimated from the data. In this section, we use the EM algorithm to obtain maximum-likelihood parameter estimates of $\boldsymbol{\theta}$ from data $\{\mathbf{Z}_t : t = 1, \dots, T\}$; here $\{\boldsymbol{\eta}_t\}$ and $\{\boldsymbol{\xi}_t\}$ are considered to be ‘‘missing data’’ (see Xu and Wikle 2007; Katzfuss and Cressie 2011, for the case of a single dataset). These estimates may then be substituted into the STDF Equations (7) and (8) in Section 3.2.

Let $\boldsymbol{\theta}^{[b]}$ be the parameter vector at the b th EM iteration. The conditional expectations and covariance matrices for the ‘‘missing data’’ are defined as

$$\boldsymbol{\eta}_{t|T}^{[b]} \equiv E_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t | \mathbf{Z}_{1:T}) \quad (9)$$

$$\boldsymbol{\xi}_{t|T}^{[b]} \equiv E_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}) \quad (10)$$

$$\mathbf{P}_{t,t-1|T}^{[b]} \equiv \text{var}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1} | \mathbf{Z}_{1:T}) \quad (11)$$

$$\mathbf{R}_{t|T}^{[b]} \equiv \text{var}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}) \quad (12)$$

$$\mathbf{W}_{t|T}^{[b]} \equiv \text{cov}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t, \boldsymbol{\xi}_t | \mathbf{Z}_{1:T}) \quad (13)$$

$$\mathbf{P}_{t,t-1|T}^{[b]} \equiv \text{cov}_{\boldsymbol{\theta}^{[b]}}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1} | \mathbf{Z}_{1:T}). \quad (14)$$

The quantities above may be obtained using the smoothing equations in the Appendix by (temporarily) setting P to be the set of observed locations at time t . There are identifiability issues when both \mathbf{H}_t and \mathbf{U}_t are allowed to vary freely with $t \in \{1, \dots, T\}$, which we address by letting $r_1 = r_2 = \dots = r_T$, $\mathbf{H} \equiv \mathbf{H}_1 = \dots = \mathbf{H}_T$, and $\mathbf{U} \equiv \mathbf{U}_1 = \dots = \mathbf{U}_T$. For some problems (e.g., see Section 3), this assumption might be modified to constant \mathbf{H} and \mathbf{U} within sequences of successive time point that partition $\{1, 2, \dots, T\}$. We define $\mathbf{K}_t^{[b+1]} \equiv \mathbf{P}_{t|T}^{[b]} + \boldsymbol{\eta}_{t|T}^{[b]} \boldsymbol{\eta}_{t|T}^{[b]}'$ and $\mathbf{L}_t^{[b+1]} \equiv \mathbf{P}_{t,t-1|T}^{[b]} + \boldsymbol{\eta}_{t|T}^{[b]} \boldsymbol{\eta}_{t-1|T}^{[b]}'$. Then, following Katzfuss and Cressie (2011), the EM updates for $\boldsymbol{\theta}^{[b+1]}$ are

$$\boldsymbol{\alpha}_t^{[b+1]} = (\mathbf{X}_t' \mathbf{Q} \mathbf{V}_t^{-1} \mathbf{Q} \mathbf{X}_t)^{-1} \mathbf{X}_t' \mathbf{Q} \mathbf{V}_t^{-1} [\mathbf{Z}_t - \mathbf{S}_t \boldsymbol{\eta}_{t|T}^{[b]} - \boldsymbol{\xi}_{t|T}^{[b]}], \quad (15)$$

$$\mathbf{K}_0^{[b+1]} = \mathbf{P}_{0|T}^{[b]} + \boldsymbol{\eta}_{0|T}^{[b]} \boldsymbol{\eta}_{0|T}^{[b]}' \quad (16)$$

$$(\sigma_{\xi,t}^{(1)})^2 [b+1] = \frac{1}{N_t^{(1)}} \text{trace}((\mathbf{E}_t^{-1} [\mathbf{R}_{t|T}^{[b]} + \boldsymbol{\xi}_{t|T}^{[b]} \boldsymbol{\xi}_{t|T}^{[b]'}])_{[1, N_t^{(1)})} \quad (17)$$

$$(\sigma_{\xi,t}^{(2)})^2 [b+1] = \frac{1}{N_t^{(2)}} \text{trace}((\mathbf{E}_t^{-1} [\mathbf{R}_{t|T}^{[b]} + \boldsymbol{\xi}_{t|T}^{[b]} \boldsymbol{\xi}_{t|T}^{[b]'}])_{[N_t^{(1)}+1, N_t]} \quad (18)$$

$$\mathbf{H}^{[b+1]} = \left(\sum_{t=1}^T \mathbf{L}_t^{[b+1]} \right) \left(\sum_{t=0}^{T-1} \mathbf{K}_t^{[b+1]} \right)^{-1} \quad (19)$$

$$\mathbf{U}^{[b+1]} = \left(\sum_{t=1}^T \mathbf{K}_t^{[b+1]} - \mathbf{H}^{[b+1]} \sum_{t=1}^T \mathbf{L}_t^{[b+1]} \right) / T, \quad (20)$$

where $(\mathbf{A})_{[i,j]}$; $j \geq i$ is the sub-block of the square matrix \mathbf{A} consisting of all elements of \mathbf{A} whose row and column indices both belong to the set given by the sequence of successive integers $\{i, i+1, \dots, j\}$. The EM estimator is $\hat{\boldsymbol{\theta}}_{\text{EM}} \equiv \lim_{b \rightarrow \infty} \boldsymbol{\theta}^{[b]}$, which is a solution to the likelihood equations under certain regularity conditions (e.g., Katzfuss and Cressie 2011). Some recommendations about EM convergence criteria and parameter starting values are included in Section B of the supplementary material.

3. STDF TO OBTAIN LOWER-ATMOSPHERIC CO₂ MOLE FRACTION

In this section, we apply the STDF methodology presented in Section 2 to ACOS data and AIRS data to derive lower-atmospheric CO₂ mole fraction over the contiguous United States, and we compare our approach to a standard NASA methodology.

3.1 Inferring Lower-Atmospheric CO₂ Mole Fraction

To construct the Basic Areal Units, or BAUs (Section 2.1), we discretized the $25^\circ \times 67^\circ$ spatial domain that covers the contiguous U.S. into a fine-scale grid of regular, (approximately) equal-area hexagons using Discrete Global Grid software (Carr et al. 1998; Sahr 2001). Specifically, we used resolution 16 of

the ISEA Aperture 3 Hexagon (ISEA3H) global grid, with an inter-cell distance of 1.170 km and a cell area of 1.185 km². We define these hexagons as the BAUs $\{A_i : i = 1, \dots, N_D\}$, whose centers comprise the index set D ; here, $N_D = 13,755,692$. We constructed the three-dimensional covariate $\mathbf{x}_i^{(k)}(\cdot)$ using the constant 1, latitude, and longitude. For the elements of the vector of basis functions, $\mathbf{S}_i^{(k)}(\cdot)$, we used local bisquare functions:

$$f_{a(b)}(\mathbf{u}) = \begin{cases} \left(1 - \frac{\|\mathbf{u} - \mathbf{m}_{a(b)}\|^2}{w(b)^2}\right)^2, & \text{for } \|\mathbf{u} - \mathbf{m}_{a(b)}\| \leq w(b), \\ 0, & \text{otherwise.} \end{cases}$$

Here \mathbf{u} and $\mathbf{m}_{a(b)} \in \mathbb{S}^2$, the 2-sphere; $\mathbf{m}_{a(b)}$ is the a th center point of the b th resolution, for $b = 1, 2, \dots, b_0$; $\|\cdot\|$ denotes great-arc distance; and $w(b)$ is taken to be 1.5 times the shortest distance between any two center points at resolution b . Following Cressie and Johannesson (2008), we computed a diagnostic summary of the SRE parameter estimates by comparing theoretical semivariograms to empirical semivariograms as functions of spatial lag. Based on the diagnostics, we used $b_0 = 2$ resolutions, namely levels 3 and 4 of the ISEA Aperture 3 Hexagon (ISEA3H) global grid: Level 3 provided 14 evenly spaced basis-function centers on a hexagonal grid (inter-cell distance of 1,476 km) over the contiguous United States; and level 4 provided 51 evenly spaced basis-function centers on a finer hexagonal grid (inter-cell distance of 852 km) over the same region. Smaller-scale spatial variation is modeled with the random process $\xi_i(\cdot)$ in (2). For this analysis, we assumed that the covariate-vector function, $\mathbf{x}_i^{(k)}(\cdot)$, and the basis functions, $\mathbf{S}_i^{(k)}(\cdot)$, do not depend on process k nor on time t . We also assumed that all observations within the k th dataset have the same measurement-error variability, so we let $v_i^{(k)}(\cdot) = 1$ for both $k = 1$ and $k = 2$.

Studies comparing monthly seasonal variations of AIRS retrievals to Matsueda airborne measurements show that AIRS measurements have an additive bias of 1.0 ppm and a measurement-error standard deviation of 3.1 ppm (Matsueda, Inoue, and Ishii 2002; Chahine et al. 2008). Validation studies comparing ACOS retrievals against Total Carbon Column Observing Network (TCCON) data indicate that ACOS measurements have a multiplicative bias of -2% and a measurement-error standard deviation of 5.1 ppm (Crisp et al. 2010; Wunch et al. 2011) and (G. Osterman, personal communication, February, 2011). We removed the additive bias from AIRS data by subtracting 1.0 ppm from all AIRS observations prior to applying STDF. Consequently, the multiplicative bias coefficients are $c^{(1)} = 0$ and $c^{(2)} = -.02$ for AIRS and ACOS, respectively. The standard deviations of the measurement errors reported by NASA were used as the measurement-error parameters in our model (i.e., $\sigma_{\epsilon,t}^{(1)} = 3.1$ and $\sigma_{\epsilon,t}^{(2)} = 5.1$).

We ran STDF for *each* of the three summer months (i.e., $T = 10$ for each analysis). While the parameters \mathbf{H} and \mathbf{U} do not vary with time within each month, they are permitted to change between months with each application of STDF to allow for large-scale temporal variability. We chose starting values $\theta^{[0]}$ for the first summer month (June, 2010) as discussed in Section B of the supplementary material. To reduce the time taken for EM estimation for July and August, we made use of STDF parameter estimates from the previous month (i.e., we initialized STDF parameters for July using the converged STDF

parameters for June). We made optimal (smoothing) predictions and derived corresponding standard errors jointly for column-average CO₂ and mid-tropospheric CO₂ for circular footprints with radius 45 km around the center points of a $1^\circ \times 1^\circ$ latitude-longitude grid over the contiguous United States for each of the 30 time periods covered by the data.

To combine column-average CO₂ and mid-tropospheric CO₂, we need to account for their vertical extent. In remote sensing, air pressure is used as a proxy for altitude, for physical reasons (Crisp et al. 2010). We made the simplifying assumption that the air pressure at the surface of the Earth is 1000 hectopascals (hPa), and the air pressure at the satellite instrument is 0 hPa. The middle troposphere is often defined to be the portion of the atmosphere between 500 hPa and 300 hPa (Moore, Remedios, and Waterfall 2010). We made an additional simplifying assumption that the CO₂ component above 300 hPa can be ignored, because the number of CO₂ molecules at the corresponding altitudes is comparatively small.

From column-average CO₂ mole fraction, $Y_{\text{ACOS}}(\mathbf{s})$, and mid-tropospheric CO₂ mole fraction, $Y_{\text{AIRS}}(\mathbf{s})$, at location \mathbf{s} , we approximated lower-atmospheric (i.e., 0 hPa to 300 hPa) CO₂ mole fraction, $Y_{\text{LA}}(\mathbf{s})$, as,

$$\begin{aligned} Y_{\text{LA}}(\mathbf{s}) &= \frac{(1000 - 300)Y_{\text{ACOS}}(\mathbf{s}) - (500 - 300)Y_{\text{AIRS}}(\mathbf{s})}{1000 - 500} \\ &= \frac{7}{5}Y_{\text{ACOS}}(\mathbf{s}) - \frac{2}{5}Y_{\text{AIRS}}(\mathbf{s}). \end{aligned} \quad (21)$$

From the weighted difference (21), it is straightforward to obtain the prediction standard error at location \mathbf{s} as,

$$\sigma_{\text{LA}}^2(\mathbf{s}) \equiv (7/5, -2/5)\mathbf{M}_{t|T}(\mathbf{s})(7/5, -2/5)',$$

where $\mathbf{M}_{t|T}(\mathbf{s})$ is the 2×2 mean-squared-prediction-error matrix for the optimal bivariate predictor, $\hat{\mathbf{Y}}_{t|T}(\mathbf{s}) \equiv (\hat{Y}_{t|T,\text{ACOS}}(\mathbf{s}), \hat{Y}_{t|T,\text{AIRS}}(\mathbf{s}))'$. The lower-atmospheric CO₂ spatial field given by (21) is a first-order approximation, since the instruments' sensitivities to different parts of the atmosphere are assumed to be indicator functions instead of the continuous functions shown in Figure 1.

We obtained smoothed values of lower-atmospheric CO₂ mole fraction and the corresponding prediction standard errors over the contiguous United States between June 1 and August 30, 2010 (i.e., for 3×10 three-day blocks). In Figure 3, we show prediction maps in ppm for the two blocks centered on June 5 and August 10, so that the reader may compare them with the raw data for the same three-day blocks (see Figure 2). The prediction map for early June (Figure 3, left panel) indicates that the lower-atmospheric CO₂ values are high in the West and the North East. The western plume may be related to the topography of the Rocky Mountains, while the northeastern plume is likely related to anthropogenic sources arising from dense urban environments. In the second prediction map (Figure 3, right panel) for early August, there is a marked decrease in the overall CO₂ values compared to those of early June. The declining CO₂ mole fraction over this timeframe is consistent with our understanding of the seasonal CO₂ cycle in the northern hemisphere. In the summer, growing plants and other photosynthesizing organisms absorb CO₂ and convert it into organic matter.

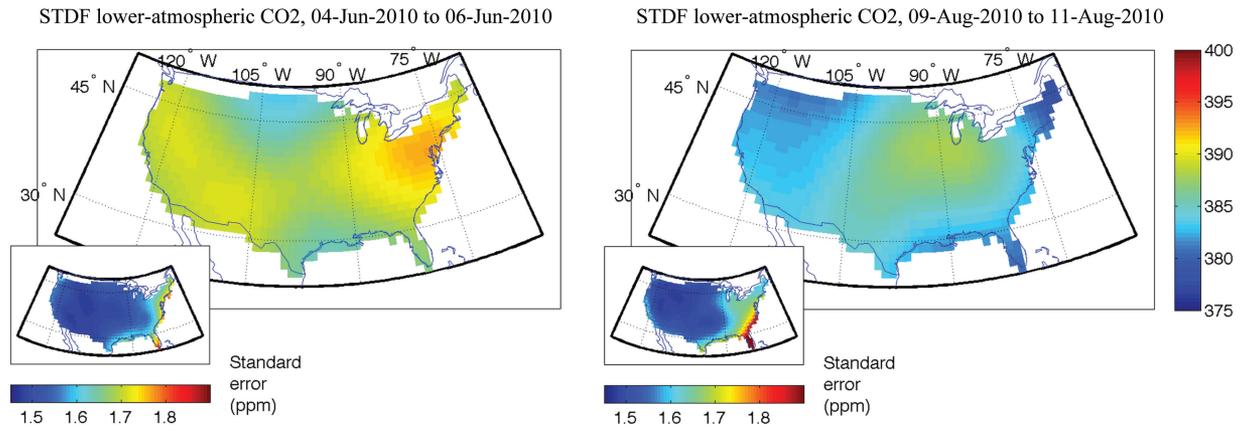


Figure 3. STDF lower-atmospheric CO₂ prediction maps and corresponding prediction-standard-error maps (inset) for the periods June 4–6, 2010 (left) and August 9–11, 2010 (right). Units are ppm.

Notice that the STDF prediction-standard-error maps tend to reflect the observational pattern of the ACOS data; the prediction standard errors are lower in the western part of United States, where we have good ACOS coverage, and they are higher in the eastern part, where we have sparse ACOS coverage. The ACOS data that we used in this study are only available over land, and thus estimates made near the transition between land and ocean in the 30 three-day blocks (e.g., California, East Coast) tend to have higher prediction standard errors.

3.2 Comparison to NOAA Flight Data

The National Oceanic and Atmospheric Administration (NOAA) has been sampling lower-atmospheric CO₂ from aircraft flights over Lamont, Oklahoma and over Homer, Illinois, among others. The program's mission is to capture the seasonal and interannual trends of trace-gas mixing ratios. The aircrafts typically collect flask samples of air at different altitudes throughout the boundary layer and free troposphere (up to 8 km). These are then analyzed by NOAA's Earth System Research Laboratory for important trace gases such as CO₂. Due

to logistical and operational challenges, these aircraft measurements of lower-atmospheric CO₂ can be sparse relative to the large spatio-temporal domain in our study.

In Figure 4, we display the NOAA aircraft data at these two locations and time periods against the corresponding 95% prediction intervals for STDF lower-atmospheric CO₂. The aircrafts fly in an ascending spiral from the surface up to about 8 km, with a majority of the measurements being collected between 1 km and 6 km. The pressure boundary at the lower part of the troposphere is 500 hPa, which corresponds roughly to an altitude of 5.5 km, so these aircraft measurements mostly reflect lower-atmospheric CO₂. Aircraft data at the beginning of a flight tend to be more unstable due to calibration and atmospheric issues, and consequently there are a moderate number of outliers in Figure 4, all identifiable by their low altitude. It is important to note that NOAA aircraft observations are instantaneous CO₂ mole fraction. On the other hand, our STDF predictions represent the average CO₂ mole fraction over the *entire* lower atmosphere within a 3-day period in the column of interest.

Despite the mismatch in spatial and temporal support, the STDF lower-atmospheric CO₂ predictions compare quite well

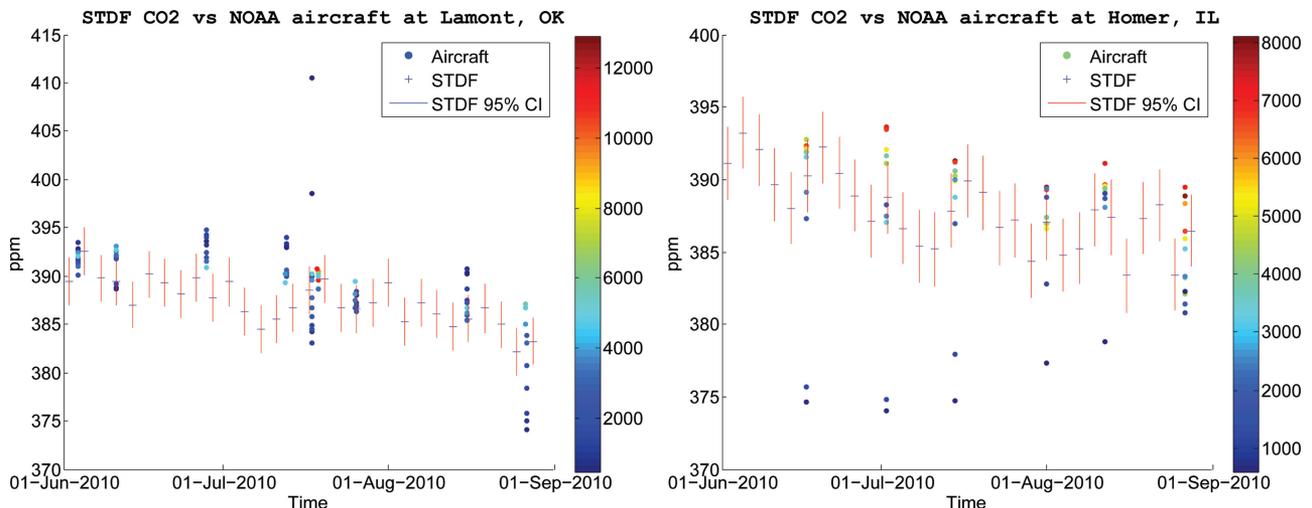


Figure 4. Shown are 95% prediction intervals for STDF lower-atmospheric CO₂ (in red) and NOAA aircraft CO₂ (colored circles) at Lamont, Oklahoma (left panel), and Homer, Illinois (right panel). The altitudes of aircraft observations (in meters) are indicated by the color bars.

to the NOAA data. Discounting the low-altitude outliers, the NOAA data at Homer, Illinois (Figure 4, right panel) mostly fall within the STDF prediction intervals. The aircraft data at Lamont also correlate well with the STDF confidence intervals. The interday small-scale fluctuations in the lower-atmospheric CO₂ mole fraction may represent atmospheric transport and surface exchange, while the large-scale declining trend in mole fraction represents seasonality. Both the NOAA data and the STDF lower-atmospheric CO₂ values at the two locations appear to capture the well known seasonal CO₂ drawdown, which features declining CO₂ mole fraction in the boreal summer as plants use atmospheric CO₂ in photosynthesis. The STDF-predicted drawdown of about 7 ppm over the three summer months is consistent with the seasonal carbon cycle (Russell and Wallace 2004).

3.3 A Comparison to Kriging

We compare our STDF methodology against spatial-only, single-dataset kriging, which was chosen because of its simplicity and the fact that it is widely used in remote sensing (see, e.g., Rossi, Dungan, and Beck 1994; Atkinson and Lewis 2000; Chatterjee et al. 2010). Classical kriging as developed by Matheron (1962) is popular in the remote sensing community because it explicitly models the spatial dependence and produces estimates of uncertainty. However, its reliance on a stationary variogram and its avoidance here of temporal dependence are features that we expect STDF will improve upon.

For six randomly chosen time blocks (centered on the dates June 14, June 17, July 11, July 26, August 1, and August 16), we withheld all ACOS and AIRS data in a reserved region inside the contiguous United States from 36°N latitude to 43°N latitude and from 105°E longitude to 95°E longitude as test data. The remaining data were used as “training data” for STDF and kriging. We then used the fitted STDF and kriging models to make predictions of column-average and mid-tropospheric CO₂ at the test locations within the reserved region in the six time blocks.

The STDF procedure was applied to the training data as described earlier in this section (with the exception that we did not correct for biases in AIRS and ACOS, to make the predictions match the held-out validation data). For kriging, we made predictions for each CO₂ product within each reserved region using only training data from the same time block. Semivariogram parameters for AIRS data were estimated for each time period from the nonwithheld data using empirical robust semivariogram estimates (Cressie 1993, Section 2.4); since ACOS data are very sparse, we combined all ACOS data over the 3 months to estimate the semivariogram parameters. We chose to use a spherical semivariogram model based on examination of the empirical semivariograms.

Since CO₂ is known to have zonal and meridional variability, we assumed a geometrically anisotropic semivariogram model. In general, column-average CO₂ tends to have longer meridional (i.e., longitudinal) correlation length, and mid-tropospheric CO₂ tends to be less anisotropic because the air above the lower atmosphere is well mixed. Examination of the empirical semivariograms from the ACOS data indicates that column-average CO₂ is highly anisotropic with a range of 8.4° in the latitudinal

Table 1. Continuous rank probability score for kriging and STDF when applied to withheld data from ACOS and AIRS. A smaller value represents a better prediction performance

	Kriging	STDF
ACOS	1.75	1.70
AIRS	1.66	1.55

direction and 14.8° in the longitudinal direction. On the other hand, mid-tropospheric CO₂ is roughly isotropic with a range of 14.2° in all directions.

Having obtained predictive distributions at the withheld locations using both STDF and kriging, we evaluated their performance using the continuous rank probability score (CRPS), a strictly proper scoring rule that generalizes the absolute error and assigns a numerical score based on a predictive distribution and the corresponding realized observation (Gneiting and Raftery 2007).

Since our test data consist of observations, $\{Z_t^{(k)}(A)\}$, instead of the true values, $\{Y_t^{(k)}(A)\}$, we obtained approximate predictive distributions for $\{Z_t^{(k)}(A)\}$ by adding the measurement-error variances of the instruments to the STDF and kriging variances, and then we calculated the CRPS as described in Gneiting and Raftery (2007).

We display the CRPS for both methods in Table 1, averaged over all test data. In both instances, STDF has smaller CRPS, indicating better forecasting performance. For column-average CO₂, STDF’s CRPS is a couple of percent smaller than kriging’s; for mid-tropospheric CO₂, STDF’s CRPS is about 7% smaller. These results indicate that taking into account temporal and inter-dataset correlations can improve bivariate predictions.

In addition to taking advantage of spatial, temporal, and interdataset dependence, STDF is computationally efficient. The STDF predictions on June data, where we initialized \mathbf{K}_0 , \mathbf{H} , and \mathbf{U} as described in Section B of the supplementary material, took 4 min on a 3.06 GHz machine with an Intel Duo Core processor, and most of the time was devoted to iterating the EM algorithm until convergence. In subsequent runs on July data and August data, we initialized the code with much-improved starting values, and the EM algorithm converged within 1 min on both runs. This speed makes STDF particularly well suited for use in analyzing remote sensing data, where an important component of the usability of methodology is whether the associated algorithm can process 1 day’s worth of satellite data in (much) less than a day. Our algorithm processes 3 months of satellite data in approximately 6 min.

The linear scalability of the STDF computations makes it especially relevant in the face of rapidly improving remote sensing technologies, where advances in design and manufacturing have vastly improved the data yield of remote sensing instruments. Modern instruments, such as the Orbiting Carbon Observatory 2 (OCO-2), scheduled to be launched on July 1, 2014, may be able to collect up to 75 times the daily yield of the instruments used in this study. The traditional kriging methodology used as a comparison in this section, has computational complexity $O(N^3)$, so a 75-fold increase in data size would equate to a 421,875-fold increase in computational time. However, for STDF, a 75-fold increase in data size simply leads to a 75-fold

increase in computational time. This means that processing 3 months of OCO-2 data (instead of GOSAT data) using STDF would take 450 minutes, or seven and a half hours. Our STDF methodology clearly still passes the usability test in the case of OCO-2's expected larger datasets.

4. DISCUSSION AND CONCLUSIONS

This article is concerned with spatio-temporal prediction of lower-atmospheric CO₂ over the contiguous United States from two remote sensing instruments—GOSAT/ACOS and AIRS. We introduce spatio-temporal data fusion (STDF) as a solution to this problem, which makes optimal predictions of a weighted difference of column-average CO₂ and mid-tropospheric CO₂ from noisy and incomplete spatio-temporal datasets. The Spatio-Temporal Random Effects (STRE) model underlying STDF is especially attractive in that it allows for seamless change-of-support and scalability to massive data sizes.

In our comparison of STDF outputs and available aircraft validation data, we show that STDF is able to reproduce the seasonal feature of the annual CO₂ drawdown and approximate the CO₂ trend at Lamont, OK and Homer, IL. Current estimates of CO₂ sources and sinks in general circulation models tend to rely on ground-based data, whose sparse locations around the globe lead to large uncertainties that can propagate “downstream” into climate-model predictions. High-coverage STDF-derived CO₂ from ACOS data and AIRS data can help inform, validate, and improve characterization of these CO₂ scenarios, leading to improved climate forecasts from general circulation models.

In Section 3, we used STDF to predict a measure of lower-atmospheric CO₂ mole fraction via data from two instruments. The methodology can be readily applied to other remote sensing datasets, especially those where the data sizes make interpolation via traditional methodologies (e.g., splines, loess, simple kriging, etc.) infeasible. Indeed, STDF can be applied not only to closely related geophysical processes like column-average CO₂ and mid-tropospheric CO₂, but to more disparate data (e.g., carbon dioxide and air temperature) to make joint predictions of the two underlying correlated processes. STDF can capitalize on the between-process correlation and produce more accurate predictions than one would obtain from either of the two datasets alone. In general, STDF is effective when there is strong temporal dependence between consecutive time blocks, and the corresponding datasets complement each other in terms of data coverage.

In this article, we have chosen to estimate the spatio-temporal model's parameters using the EM algorithm. Because we then proceed with inference on the latent processes by substituting the parameter estimates into the STDF equations of Section 3.2, our prediction standard errors do not include variability due to uncertainty in the parameter estimates. We could rectify this by putting a prior distribution on the parameter vector θ and carry out Bayesian inference to produce optimal predictions and associated posterior uncertainties (e.g., Katzfuss and Cressie 2012, in the single-instrument case). While the dimension-reduction feature of the STRE model would reduce the computational burden of the Monte Carlo algorithm, computational time would be much longer overall (hours/days instead of minutes) than that

of EM-based STDF. In remote sensing applications, it is very important to develop algorithms with reasonable run-times, so that they can accommodate the constant influx of new data. Estimating parameters instead of putting priors on them is a very effective compromise.

In summary, we have developed STDF and applied it to spatio-temporal prediction of lower-atmospheric CO₂. We can obtain high global coverage with known uncertainties, and hence, we can construct CO₂-mole-fraction scenarios for use in general circulation models. The scalability of STDF makes it especially appropriate for the massive datasets often found in remote sensing of the environment.

APPENDIX

STDF SMOOTHING EQUATIONS

Let $\mathbf{Z}_{1:\tilde{t}} \equiv (\mathbf{Z}'_1, \dots, \mathbf{Z}'_{\tilde{t}})'$, for $\tilde{t} = 1, \dots, T$, and we define $\eta_{t|\tilde{t}} \equiv E(\eta_t | \mathbf{Z}_{1:\tilde{t}})$ and $\xi_{t|\tilde{t}}^P \equiv E(\xi_t^P | \mathbf{Z}_{1:\tilde{t}})$ as the conditional expectations of the respective quantities given data $\mathbf{Z}_{1:\tilde{t}}$. Similarly, we denote the conditional covariance matrix of η_t by $\mathbf{P}_{t|\tilde{t}} \equiv \text{var}(\eta_t | \mathbf{Z}_{1:\tilde{t}})$, the conditional covariance matrix of ξ_t^P by $\mathbf{R}_{t|\tilde{t}}^P \equiv \text{var}(\xi_t^P | \mathbf{Z}_{1:\tilde{t}})$, and the conditional covariance matrix between η_t and ξ_t^P by $\mathbf{W}_{t|\tilde{t}}^P \equiv \text{cov}(\eta_t, \xi_t^P | \mathbf{Z}_{1:\tilde{t}})$.

We first initialize $\eta_{0|0} = \mathbf{0}$ and $\mathbf{P}_{0|0} = \mathbf{K}_0$. The *filtering quantities* for $t = 1, \dots, T$ are given by the recursive relationships:

$$\eta_{t|t} = \eta_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{S}'_t [\mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{S}'_t + \mathbf{D}_t]^{-1} \times (\mathbf{Z}_t - \mathbf{Q} \mathbf{X}_t \alpha_t - \mathbf{S}_t \eta_{t|t-1}) \quad (\text{A.1})$$

$$\xi_{t|t}^P = \mathbf{C}_t^{PZ} \mathbf{E}_t^{PZ} [\mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{S}'_t + \mathbf{D}_t]^{-1} \times (\mathbf{Z}_t - \mathbf{Q} \mathbf{X}_t \alpha_t - \mathbf{S}_t \eta_{t|t-1}) \quad (\text{A.2})$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{S}'_t [\mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{S}'_t + \mathbf{D}_t]^{-1} \mathbf{S}_t \mathbf{P}_{t|t-1} \quad (\text{A.3})$$

$$\mathbf{R}_{t|t}^P = \mathbf{C}_t^P \mathbf{E}_t^P - \mathbf{C}_t^{PZ} \mathbf{E}_t^{PZ} [\mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{S}'_t + \mathbf{D}_t]^{-1} (\mathbf{E}_t^{PZ})' (\mathbf{C}_t^{PZ})', \quad (\text{A.4})$$

$$\mathbf{W}_{t|t}^P = -\mathbf{P}_{t|t-1} \mathbf{S}'_t [\mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{S}'_t + \mathbf{D}_t]^{-1} (\mathbf{E}_t^{PZ})' (\mathbf{C}_t^{PZ})', \quad (\text{A.5})$$

where

$$\mathbf{Q} \equiv \begin{pmatrix} (1 + c^{(1)}) \mathbf{I}_{N_t^{(1)}} & \mathbf{0} \\ \mathbf{0} & (1 + c^{(2)}) \mathbf{I}_{N_t^{(2)}} \end{pmatrix},$$

$\text{var}(\xi_t^P) = \mathbf{C}_t^P \mathbf{E}_t^P$, $\text{cov}(\xi_t^P, \xi_t) = \mathbf{C}_t^{PZ} \mathbf{E}_t^{PZ}$, and \mathbf{C}_t^P , \mathbf{C}_t^{PZ} , \mathbf{E}_t^P , and \mathbf{E}_t^{PZ} are defined analogously to the terms \mathbf{C}_t and \mathbf{E}_t under (5). The one-step-ahead forecasts are

$$\begin{aligned} \eta_{t|t-1} &= \mathbf{H}_t \eta_{t-1|t-1} \\ \mathbf{P}_{t|t-1} &= \mathbf{H}_t \mathbf{P}_{t-1|t-1} \mathbf{H}'_t + \mathbf{U}_t. \end{aligned}$$

Having calculated the conditional expectations and covariances for $t = 1, \dots, T$ from (A.1)–(A.5), we obtain the *smoothing quantities* by updating “backward” in time (i.e., for $t = T - 1, T - 2, \dots, 0$):

$$\eta_{t|T} = \eta_{t|t} + \mathbf{J}_t (\eta_{t+1|T} - \eta_{t+1|t}) \quad (\text{A.6})$$

$$\xi_{t|T}^P = \xi_{t|t}^P + \mathbf{B}_t (\eta_{t+1|T} - \eta_{t+1|t}) \quad (\text{A.7})$$

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{J}'_t \quad (\text{A.8})$$

$$\mathbf{R}_{t|T}^P = \mathbf{R}_{t|t}^P + \mathbf{B}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{B}'_t \quad (\text{A.9})$$

$$\mathbf{W}_{t|T}^P = \mathbf{W}_{t|t}^P + \mathbf{J}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{B}'_t, \quad (\text{A.10})$$

where

$$\mathbf{J}_t \equiv \mathbf{P}_{t|t} \mathbf{H}'_{t+1} \mathbf{P}_{t+1|t}^{-1}$$

$$\mathbf{B}_t \equiv -\mathbf{C}_t^{PZ} \mathbf{E}_t^{PZ} [\mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{S}'_t + \mathbf{D}_t]^{-1} \mathbf{S}_t \mathbf{P}_{t|t-1} \mathbf{H}'_{t+1} \mathbf{P}_{t+1|t}^{-1}.$$

The cross-covariance term, $\mathbf{P}_{t,t-1|T} \equiv \text{cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_{t-1} | \mathbf{Z}_{1:T})$, is given by

$$\mathbf{P}_{T,T-1|T} = (\mathbf{I}_r - \mathbf{P}_{T|T-1} \mathbf{S}'_T [\mathbf{S}_T \mathbf{P}_{T|T-1} \mathbf{S}'_T + \mathbf{D}_T]^{-1} \mathbf{S}_T) \times \mathbf{H}_T \mathbf{P}_{T-1|T-1}$$

$$\mathbf{P}_{t,t-1|T} = \mathbf{P}_{t|t} \mathbf{J}'_{t-1} + \mathbf{J}_t (\mathbf{P}_{t+1,t|T} - \mathbf{H}_{t+1} \mathbf{P}_{t|t}) \mathbf{J}'_{t-1};$$

$$t = 1, \dots, T - 1,$$

where \mathbf{I}_r is the $r \times r$ identity matrix.

In applications where real-time processing is important, the smoothing approach in this section can be modified to take a filtering perspective. In this case, we would carry out only the filtering steps in (A.1)–(A.5), and the conditional expectations and covariance matrices for the “missing data” in (9)–(14) would be conditioned on $\mathbf{Z}_{1:t}$ instead of $\mathbf{Z}_{1:T}$.

SUPPLEMENTARY MATERIALS

The following supplementary materials can be obtained via a single download.

Spatial-only data fusion document: The file “STDF spatial-only supplement.pdf” describes data model, prediction equations, EM starting values, EM convergence criteria, and EM parameter estimation for the case of spatial-only datasets.

ACOS and AIRS datasets: The zip file “STDF data.zip” contains a folder with the ACOS data and the AIRS data along with a README.txt file describing the data format.

ACKNOWLEDGMENTS

The research described in this article was carried out in part by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. It is supported by NASA’s Earth Science Technology Office through its Advanced Information Systems Technology program. Katzfuss’ research was partially supported by the Mathematics Center Heidelberg. Cressie’s research was partially supported by the Naval Surface Warfare Center, Dahlgren Division.

[Received December 2011. Revised May 2013.]

REFERENCES

Atkinson, P. M., and Lewis, P. (2000), “Geostatistical Classification for Remote Sensing: An Introduction,” *Computers and Geosciences*, 26, 361–371. [9]

Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. L., and Susskind, J. (2003), “AIRS/AMSU/HSB on the Aqua Mission: Design, Science Objectives, Data Products, and Processing Systems,” *IEEE Transactions on Geoscience and Remote Sensing*, 41, 253–264. [2]

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), “Gaussian Prediction Process Models for Large Spatial Data Sets,” *Journal of the Royal Statistical Society, Series B*, 70, 825–848. [4]

Berliner, L., Wikle, C., and Milliff, R. (1999), “Multiresolution Wavelet Analyses in Hierarchical Bayesian Turbulence Models,” in *Bayesian Inference in Wavelet-Based Models*, eds. P. Miller and B. Vidakovic. Springer Lecture Notes in Statistics, No. 141. New York: Springer-Verlag. [3]

Calder, C. (2008), “A Bayesian Dynamic Process Convolution Approach to Modelling the Point Distribution of PM_{2.5} and PM₁₀,” *Environmetrics*, 19, 39–48. [4]

Carr, D., Kahn, R., Sahr, K., and Olsen, T. (1998), “ISEA Discrete Global Grids,” *Statistical Computing and Statistical Graphics Newsletter*, 8, 31–39. [6]

Chahine, M. T., Chen, L., Dimotakis, P., Jiang, X., Li, Q., Olsen, E. T., Pagano, T., Randerson, J., and Yung, Y. L. (2008), “Satellite Remote Sounding of Mid-Tropospheric CO₂,” *Geophysical Research Letters*, 35, L17807, doi: 10.1029/2008GL035022. [2,7]

Chatterjee, A., Michalak, A. M., Kahn, R. A., Paradise, S. R., Braverman, A. J., and Miller, C. E. (2010), “A Geostatistical Data Fusion Technique for Merging Remote Sensing and Ground-Based Observations of Aerosol Optical Thickness,” *Journal of Geophysical Research*, 115, D20207, doi:10.1029/2009JD013765. [9]

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P. (2005), “Inferring CO₂ Sources and Sinks From Satellite Observations: Method and Application to TOVS Data,” *Journal of Geophysical Research: Atmospheres*, 110, doi:10.1029/2005JD006390. [2]

Cressie, N. (1993), *Statistics for Spatial Data* (revised ed.), New York: Wiley-Interscience. [9]

Cressie, N., and Johannesson, G. (2008), “Fixed Rank Kriging for Very Large Spatial Data Sets,” *Journal of the Royal Statistical Society, Series B*, 70, 209–226. [4,5,7]

Cressie, N., Shi, T., and Kang, E. L. (2010), “Fixed Rank Filtering for Spatio-Temporal Data,” *Journal of Computational and Graphical Statistics*, 19, 724–745. [4,5,6]

Crisp, D., Boesch, H., Brown, L., Castano, R., Christi, M., Conner, B., Frankenberg, C., McDuffie, J., Miller, C., Natraj, V., O’Dell, C., O’Brien, D., Polonski, I., Oyafuso, F., Thompson, D., Toon, G., and Spurr, R. (2010), “OCO (Orbiting Carbon Observatory): Level 2 Full Physics Retrieval Algorithm Theoretical Basis,” Version 1.0 Rev. 4, November 10, 2010. JPL, NASA, Pasadena, CA. [7]

Crisp, D., Fisher, B. M., O’Dell, C., Frankenberg, C., Basilio, R., Bsch, 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. (2012), “The ACOS CO₂ Retrieval Algorithm—Part II: Global XCO₂ Data Characterization,” *Atmospheric Measurement Techniques*, 5, 687–707. [2]

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N. (2006), “Climate-Carbon Cycle Feedback Analysis: Results From the C4MIP Model Intercomparison,” *Journal of Climate*, 19, 3337–3353. [1]

Gneiting, T., and Raftery, A. E. (2007), “Strictly Proper Scoring Rules, Prediction, and Estimation,” *Journal of the American Statistical Association*, 102, 359–378. [9]

Gruber, N., Gloor, M., Fletcher, S. E. M., Dutkiewicz, S., Follows, M., Doney, S. C., Gerber, M., Jacobson, A. R., Lindsay, K., Menemenlis, D., Mouchet, A., Mueller, S. A., Sarmiento, J. L., and Takahashi, T. (2009), “Oceanic Sources, Sinks, and Transport of Atmospheric CO₂,” *Global Biogeochemical Cycles*, 23, GB1005, doi:10.1029/2008GB003349. [2]

Henderson, H., and Searle, S. (1981), “On Deriving the Inverse of a Sum of Matrices,” *SIAM Review*, 23, 53–60. [5]

Hooten, M. B., Larsen, D., and Wikle, C. (2003), “Predicting the Spatial Distribution of Ground Flora on Large Domains Using a Hierarchical Bayesian Model,” *Landscape Ecology*, 18, 487–502. [4]

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A. (2001), *Climate Change 2001: The Scientific Basis*, Cambridge, UK: Cambridge University Press. [1]

Kang, E. L., Liu, D., and Cressie, N. (2009), “Statistical Analysis of Small-Area Data Based on Independence, Spatial, Non-Hierarchical, and Hierarchical Models,” *Computational Statistics and Data Analysis*, 53, 3016–3022. [4]

Katzfuss, M., and Cressie, N. (2009), “Maximum Likelihood Estimation of Covariance Parameters in the Spatial Random Effects Model,” in *Proceedings of the 2009 Joint Statistical Meetings*, Alexandria, VA: American Statistical Association, 3378–3390. [4]

——— (2011), “Spatio-Temporal Smoothing and EM Estimation for Massive Remote-Sensing Data Sets,” *Journal of Time Series Analysis*, 32, 430–446. [4,6]

——— (2012), “Bayesian Hierarchical Spatio-Temporal Smoothing for Very Large Datasets,” *Environmetrics*, 23, 94–107. [10]

Lindgren, F., Rue, H., and Lindstrom, J. (2011), “An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic

- Partial Different Equation Approach,” *Journal of Royal Statistical Society, Series B*, 73, 423–498. [4]
- Matheron, G. (1962), “Treatise on Applied Geostatistics, Vol 1,” *Memoires du Bureau de Recherches Géologiques et Minières*, No. 14. Editions Technip, Paris. [9]
- Matsueda, H., Inoue, H. Y., and Ishii, M. (2002), “Aircraft Observation of Carbon Dioxide at 8–13 km Altitude Over the Western Pacific From 1993 to 1999,” *Tellus, Series B*, 54, 1–21. [7]
- McGuffie, K., and Henderson-Sellers, A. (1997), *A Climate Modelling Primer*, New York: Wiley. [1]
- Moore, D. P., Remedios, J. J., and Waterfall, A. M. (2010), “Global Distributions of Acetone in the Upper Troposphere From MIPAS-E Spectra,” *Atmospheric Chemistry and Physics Discussions*, 10, 23539–23557. [7]
- Morino, I., Uchino, O., Inoue, M., Yoshida, Y., Yokota, T., Wennberg, P. O., Toon, G. C., Wunch, D., Roehl, C. M., Notholt, J., Warneke, T., Messerschmidt, J., Griffith, D. W. T., Deutscher, N. M., Sherlock, V., Connor, B., Robinson, J., Sussmann, R., and Rettinger, M. (2011), “Preliminary Validation of Column-Averaged Volume Mixing Ratios of Carbon Dioxide and Methane Retrieved From GOSAT Short-Wavelength Infrared Spectra,” *Atmospheric Measurement Techniques*, 4, 1061–1076. [2]
- Nychka, D., Wikle, C. K., and Royle, J. (2002), “Multiresolution Models for Nonstationary Spatial Covariance Functions,” *Statistical Modelling*, 2, 315–331. [3]
- Rossi, R. E., Dungan, J. L., and Beck, L. R. (1994), “Kriging in the Shadows: Geostatistical Interpolation for Remote Sensing,” *Remote Sensing of Environment*, 49, 32–40. [9]
- Royle, J., and Wikle, C. (2005), “Efficient Statistical Mapping of Avian Count Data,” *Ecological and Environmental Statistics*, 12, 225–243. [4]
- Russell, J., and Wallace, J. (2004), “Annual Carbon Dioxide Drawdown and the Northern Annular Mode,” *Global Biogeochemical Cycles*, 18, 1033, doi:10.1029/2000GB001317. [9]
- Sahr, K. (2001), “DGGRID Version 3.1b. User Documentation for Discrete Global Grid Generation Software,” available at <http://www.sou.edu/cs/sahr/dgg/>. [6]
- Stein, M. L., and Jun, M. (2008), “Nonstationary Covariance Models for Global Data,” *Annals of Applied Statistics*, 2, 1271–1289. [4]
- Wikle, C. K. (2003), “Hierarchical Models in Environmental Science,” *International Statistical Review*, 71, 181–199. [4]
- Wikle, C. K., and Berliner, L. M. (2005), “Combining Information Across Spatial Scales,” *Technometrics*, 47, 80–91. [4]
- Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001), “Spatio-Temporal Hierarchical Bayesian Modeling: Tropical Ocean Surface Winds,” *Journal of the American Statistical Association*, 96, 382–397. [3]
- Wunch, D., Toon, G., Blavier, J., Washenfelder, R., Notholt, J., Connor, B., Griffith, D., Sherlock, V., and Wennberg, P. (2011), “The Total Carbon Column Observing Network,” *Philosophical Transactions of the Royal Society A*, 369, 2087–2112. [7]
- Xu, K., and Wikle, C. K. (2007), “Estimation of Parameterized Spatio-Temporal Dynamic Models,” *Journal of Statistical Planning and Inference*, 137, 567–588. [6]