Spatial Statistical Data Fusion for Remote Sensing Applications

Amy Braverman\textsuperscript{1} Hai Nguyen\textsuperscript{1} Noel Cressie\textsuperscript{2} Matthias Katzfuss\textsuperscript{2} Edward Olsen\textsuperscript{1} Rui Wang\textsuperscript{2} Anna Michalak\textsuperscript{3} Charles Miller\textsuperscript{1}

\textsuperscript{1}Jet Propulsion Laboratory, California Institute of Technology

\textsuperscript{2}Department of Statistics, The Ohio State University

\textsuperscript{3}Department of Civil and Environmental Engineering, University of Michigan

June 22, 2010
The work presented here is based on three related bodies of research:

1. a theoretical foundation developed by Noel Cressie,
2. a method for spatial interpolation of very large data sets called Fixed-Rank Kriging, also developed by Noel Cressie and students,
3. Spatial Statistical Data Fusion, the Ph.D. dissertation of Hai Nguyen, which extends Fixed-Rank Kriging to address the data fusion problem.

This research is supported by the NASA Earth Science Technology Office through its Advanced Information Systems Technology Program.

Some small notational errors in the published version of the paper. If you’d like a corrected copy, please email me at Amy.Braverman@jpl.nasa.gov.
Outline

Introduction

Spatial statistics

Kriging

Fixed-Rank Kringing

Data Fusion

Fusing AIRS and OCO-2 CO2

Conclusions
Data fusion means many things to many different people.

This is true even within the remote sensing community (e.g. "image fusion").

Our definition focuses on Earth science: infer the true value of some quantity of interest from multiple data sources with different statistical characteristics (e.g. resolutions, systematic and random errors, etc.).

The fused data are these estimates (also called the "predictions"). Uncertainties of the estimates (mean squared prediction errors, MSPE's) must accompany the predictions.

Inference from spatial data relies on the theory of spatial statistics, so that is the formalism we use.

First, consider the problem of inferring the true value from a single data source.
Remote sensing data:

(A) is the true field. (B) is discretized into pixels. (C) is noisy (measurement bias and variance added). (D) has missing data.

Given only (D), can we infer (A)?
Let $s_1$ and $s_2$ be the (lat,lon) pairs of two point locations.
Let $Y_1(\cdot)$ and $Y_2(\cdot)$ be random variables representing the values of two quantities (e.g. air temperature and humidity) at the locations of their arguments.

slices of the joint pdf of $Y_1(s_1)$ and $Y_2(s_2)$ at fixed values of $Y_2(s_2)$.

$E[Y_1(s_1)|Y_2(s_2)] = \text{projection of the slice means onto the floor is a line (linear regression).}$
$E[\cdot] = \text{expected value.}$ $[Y_1(s_1)|Y_2(s_2)] = \text{conditional distribution of } Y_1(s_1) \text{ given } Y_2(s_2)$. 
Remote sensing data typically have areal extent (support), not point support.

Slices of the joint pdf of $Y_1(s_1)$ and $Z_2(B(s_2))$
at fixed values of $Z_2(B(s_2))$

\[
Z_2(B(s_2)) = \left[ \frac{1}{|B(s_2)|} \int_{u \in B(s_2)} Y_2(u) du \right] + \epsilon_2(B(s_2)),
\]
where $| \cdot |$ is size, and $\epsilon_2(\cdot)$ is measurement error.

Remote sensing data have measurement error: $E(\epsilon_2(\cdot)) = b\mu$ (bias), $\mu = E(Y(\cdot))$; $Var(\epsilon_2(\cdot)) = \sigma^2$ (variance).
Optimal statistical spatial interpolation (for point support) using covariances to determine weights.

Georges Matheron (1963).

Let \( Y(\cdot) \) be a statistical "process" (random variable) to be estimated at location \( s_0 \) from observations at locations \( s_1, \ldots, s_N \in D \).

Assume \( Z(s_i) = Y(s_i) + \epsilon(s_i) \), where \( \epsilon(\cdot) \) is zero-mean white noise with finite variance, \( \sigma^2 \nu(s) \), \( \sigma^2 > 0 \), and \( \nu(\cdot) \) is assumed known.

Assume \( Y \) has linear mean structure:

\[
Y(s) = t(s)'\alpha + \nu(s),
\]

where \( t(\cdot) \) is a vector of known covariates (e.g. latitude and longitude), \( \alpha \) is estimated from the data, and \( \nu(\cdot) \) is small scale variation.

\( \nu(s) \) is assumed to be zero mean with finite, non-zero variance, and (spatial) covariance function,

\[
Cov(\nu(u), \nu(v)) = C(u, v).
\]
Combine all this and get:

\[ Z = T\alpha + \delta, \quad \delta = \nu + \epsilon, \]

where \( \delta, \nu, \) and \( \epsilon \) are vectors of length \( N \), \( T \) is an \( N \times p \) matrix of covariates (\( p = 2 \) for lat/lon).

Note: \( \delta \) is a combination of small scale variation and measurement error. Write:

\[ \text{Cov}(\delta) = \Sigma = C + \sigma^2 V, \]

where \( C \) is an \( N \times N \) matrix of spatial covariances, \( [C]_{ij} = C(s_i, s_j) \), and \( V = \text{diag}(\nu(s_1), \ldots, \nu(s_N)) \). Note: \( V \) allows for non-constant measurement error variance.

The kriging estimator of \( Y(s_0) \) is \( \hat{Y}(s_0) = \mathbf{a}'Z \) where \( \mathbf{a} \) is chosen to minimize \( E\|Y(s_0) - \mathbf{a}'Z\|^2 \) subject to the unbiasedness condition, \( E(\mathbf{a}'Z) = E(Y(s_0)) = \mu \).
Answer:

\[ \hat{Y}(s_0) = t(s_0)\hat{\alpha} + a'(Z - T\hat{\alpha}), \]

\[ RMSPE(\hat{Y}(s_0)) = \left\{C(s_0, s_0) - a'\Sigma a \
+ (t(s_0) - T'a)'(T'\Sigma^{-1}T)^{-1}(t(s_0) - T'a)\right\}^{\frac{1}{2}}, \]

where

\[ \hat{\alpha} = (T'\Sigma^{-1}T)^{-1}T'\Sigma^{-1}Z, \quad a = c(s_0)\Sigma^{-1}, \quad c(s_0) \equiv (C(s_0, s_1), \ldots, C(s_0, s_N))'. \]

If there is bias in the measurement, \( E(\epsilon(s)) = b\mu, \) then

\[ a = \left(\Sigma^{-1} + \Sigma^{-1}1(1 + b)[-1'(1 + b)\Sigma^{-1}1(1 + b)]^{-1}1'(1 + b)\Sigma^{-1}\right)c(s_0) \]

\[ + \Sigma^{-1}1(1 + b)[-1'(1 + b)\Sigma^{-1}1(1 + b)]^{-1}1'(1 + b)\Sigma^{-1}. \]
Unless $C$ is isotropic and stationary, it is hard to invert the $N \times N$ covariance matrix $\Sigma$ because $N$ is very large. Isotropy and stationarity are unrealistic for most geophysical processes, particularly at large scales.

We don’t observe $Z(s)$ (point support), we observe $Z(B(s))$ (footprint or "block" support).

Cressie and Johannesson (2008) introduced Fixed-Rank Kriging (FRK) as a way to deal with these problems. Model the covariance function as

$$C(u, v) = S(u)'KS(v), \quad u, v \in D,$$

for some $r \times r$ positive-definite matrix $K$, $r \ll N$. $S(\cdot)$ is the basis expansion of a point location into a fixed set of (not necessarily orthogonal) basis functions, $S_j(\cdot)$: $S(u) = (S_1(u), \ldots, S_r(u))'$. 
Multi-resolution spatial basis functions. The spatial domain is subdivided into three levels of resolution, each a factor of two finer than its parent. At resolution \( l \), each cell center \((m_1, \ldots, m_4 \text{ in the left panel, but not shown in the others})\) is the center of a circle of diameter \( r_l \). Point location \( u \) belongs to three circles, and is distances \( d_1, d_2, \) and \( d_3 \) from the cell centers, respectively. These distances determine the basis function values at resolution \( l \), as shown in the right panel.
Then $\Sigma = \sigma^2 V + S'KS$, and

$$\Sigma^{-1} = (\sigma^2 V)^{-1} - (\sigma^2 V)^{-1}S'(K^{-1} + S(\sigma^2 V)^{-1}S')^{-1}S'(\sigma^2 V)^{-1},$$

by the Sherman-Morrison-Woodbury formula (Henderson and Searle, 1981). This only requires the inversion of $K$ and $(K^{-1} + S'(\sigma^2 V)^{-1}S)$, both of which are $r \times r$ matrices. Order of computation is $O(Nr^2)$, not $O(N^3)$.

The FRK kriging predictors and their uncertainties are

$$\hat{Y}(s_0) = t(s_0)\hat{\alpha} + S(s_0)'\hat{K}S\hat{\Sigma}^{-1}(Z - T\hat{\alpha}),$$

$$RMSPE(\hat{Y}(s_0)) = \{S(s_0)'\hat{K}S - S(s_0)'\hat{K}S\hat{\Sigma}^{-1}S'\hat{K}S(s_0) + (t(s_0) - T'\hat{\Sigma}^{-1}S'\hat{K}S(s_0))(T'\hat{\Sigma}^{-1}T)^{-1}(t(s_0) - T'\hat{\Sigma}^{-1}S'\hat{K}S(s_0))\}^{\frac{1}{2}}.$$

$\hat{\Sigma} = \sigma^2 V + S'\hat{K}S$ and $\hat{K}$ is estimated from the data. (Assume $\sigma^2$ is given and no measurement bias.)
What about estimating $K$? We can use the footprint-level data:

$$
\text{Cov}(Z(B(s_k)), Z(B(s_l))) = \\
\text{Cov} \left[ \frac{1}{|B(s_k)|} \int_{u \in B(s_k)} Y(u) du + \epsilon(B(s_k)), \frac{1}{|B(s_l)|} \int_{v \in B(s_l)} Y(v) dv + \epsilon(B(s_l)) \right],
$$

$$
= \frac{1}{|B(s_k)|} \frac{1}{|B(s_l)|} \int_{u \in B(s_k)} \int_{v \in B(s_l)} \text{Cov}(Y(u), Y(v)) du dv
$$

$$
= \frac{1}{|B(s_k)|} \int_{u \in B(s_k)} S(u)' du K \frac{1}{|B(s_l)|} \int_{v \in B(s_l)} S(v)' dv,
$$

$$
= \tilde{S}(B(s_k))' K \tilde{S}(B(s_l)),
$$

where $\tilde{S}(B(s)) = \left( \tilde{S}_1(B(s)), \ldots, \tilde{S}_r(B(s)) \right)$, $\tilde{S}_j(B(s)) = \frac{1}{|B(s)|} \int_{u \in B(s)} S_j(u) du$. 
Estimate $\mathbf{K}$ by:

1. Subdividing the domain into coarse bins (e.g. resolution of a coarse level of $\mathbf{S}$) and calculating $\hat{\mathbf{\Sigma}}$, an empirical estimate of the spatial covariance matrix (details omitted in the interest of time).

2. Find $\mathbf{K}$ that minimize the distance between $\tilde{\mathbf{\Sigma}}$ and $\hat{\mathbf{\Sigma}}$:

$$
\| \hat{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}(\mathbf{K}) \|_F = tr \left( (\hat{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}(\mathbf{K}))'(\hat{\mathbf{\Sigma}} - \tilde{\mathbf{\Sigma}}(\mathbf{K})) \right).
$$

This yields a method-of-moments estimate, $\hat{\mathbf{K}}$. 
Example: FRK AIRS CO2, May 1-3, 2003

Prediction

RMSPE

Raw Data

Simple average, one degree grid
Locations south of $60^\circ$S screened out.

Results with $RMSPE > .5$ screened out.

Prediction grid is rectangular, $1^\circ \times 1^\circ$.

396 basis functions at three coarsest levels of resolution on hexagonal discrete global grid (DGG; Sahr and White, 1998). Intercell distances at level 1 $\approx 4,400$ km; at level 2 $\approx 2,500$ km; at level 3 $\approx 1,400$ km.

Binning for method-of-moments estimation of $K$ uses level 5 DGG hexagons (intercell distance $\approx 400$ km).

Computation time: 15 seconds of a 3 GHz MacBook Pro.
Surely having a second data set must help, especially if it’s measuring the same thing. If not, it will still help if the second measurement is correlated with the first.

Single process multiple source (SPMS), $Y_1(\cdot) = Y_2(\cdot) = Y(\cdot)$ or multiple process multiple source (MPMS), $Y_1(\cdot) \neq Y_2(\cdot)$. 
\begin{align*}
E(Y_i(s)) &= \mu_i, \quad E(\epsilon_j(B_{jk})) = b_j \mu_j, \quad \text{Var}(\epsilon_j(B_{jk})) = \sigma_j^2, \quad B_{jk} = B_j(s_k), \\
\text{where } i \text{ indexes process (variable), } j \text{ indexes instrument, and } k \text{ indexes footprint.}
\end{align*}

\begin{align*}
Z_j &= (Z_j(B_{j1}), \ldots, Z_j(B_{jN_j}))', \\
Z_j(B_{jk}) &= \frac{1}{|B_{jk}|} \int_{u \in B_{jk}} Y_j(u) \, du + \epsilon_j(B_{jk}),
\end{align*}

\begin{equation}
\hat{Y}(s) = \begin{bmatrix}
\hat{Y}_1(s) \\
\hat{Y}_2(s)
\end{bmatrix} = \begin{bmatrix}
a'_{11s}Z_1 + a'_{12s}Z_2 \\
a'_{21s}Z_1 + a'_{22s}Z_2
\end{bmatrix}.
\end{equation}

Minimize \( E \left\{ \left( \hat{Y}_1(s) - Y_1(s) \right)^2 + \left( \hat{Y}_2(s) - Y_2(s) \right)^2 \right\} \) subject to

\begin{align*}
E(\hat{Y}_1(s)) &= b_1 a'_{11s} 1_{N_1} \mu_1 + b_2 a'_{12s} 1_{N_2} \mu_2 = \mu_1 \quad \text{and} \\
E(\hat{Y}_2(s)) &= b_1 a'_{21s} 1_{N_1} \mu_1 + b_2 a'_{22s} 1_{N_2} \mu_2 = \mu_2.
\end{align*}
\[ a_{i1s} = A_1^{-1}(B_{i1} + C_1 m_i), \quad \text{and} \quad a_{i2s} = A_2^{-1}(B_{i2} + C_2 m_i). \]

where

\[ m_i = \frac{(1 - 1'N_1 A_1^{-1} B_{i1}(1 + b_1) + 1'N_2 A_2^{-1} B_{i2}(1 + b_2))}{(1'N_1 A_1^{-1} C_1(1 + b_1) + 1'N_2 A_2^{-1} C_2(1 + b_2))}, \quad \text{and} \]

\[ A_1 \equiv (I_{N_1} - \hat{\Sigma}^{-1} \Sigma_{11} \hat{\Sigma}_{12} \hat{\Sigma}_{22} \hat{\Sigma}_{21}), \]
\[ A_2 \equiv (I_{N_2} - \hat{\Sigma}_{22} \hat{\Sigma}_{21} \hat{\Sigma}_{11} \hat{\Sigma}_{12}), \]
\[ B_{i1} \equiv \hat{\Sigma}_{11}^{-1} (c_{i1s} - \hat{\Sigma}_{12} \hat{\Sigma}_{22} c_{i2s}), \]
\[ B_{i2} \equiv \hat{\Sigma}_{22}^{-1} (c_{i2s} - \hat{\Sigma}_{21} \hat{\Sigma}_{11} c_{i1s}), \]
\[ C_1 = -\hat{\Sigma}_{11}^{-1} \left[ 1_{N_1}(1 + b_1) - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} 1_{N_2}(1 + b_2) \right], \]
\[ C_2 = -\hat{\Sigma}_{22}^{-1} \left[ 1_{N_2}(1 + b_2) - \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} 1_{N_1}(1 + b_1) \right]. \]

\( K_{ij} \) estimated using method of moments as described earlier.
OCO-2 will measure total column CO2 on $1.1 \times 2.25$ km footprints.

AIRS measures mid-tropospheric (and above) CO2 at 90 km resolution.

$Y_1(\cdot) =$ total column CO2, $Y_2(\cdot) =$ CO2 in the mid-troposphere and above.

$Y_1(s) - Y_2(s) =$ planetary boundary layer CO2 (PBL CO2).
PBL is that portion of the lower atmosphere that is dragged along by the rotation of the Earth.

CO2 fluxes from the surface enter the atmosphere at the PBL.

Changes in PBL CO2 at any time should generally be correlated with flux.

Monitoring PBL CO2 may allow monitoring of sources and sinks.
Fusing AIRS and OCO-2 CO2

- OCO-like (OCOL) synthetic data downscaled from Parallel Climate Transport Model (PCTM) at $1^\circ \times 1.25^\circ$ resolution to $1 \times 2$ km resolution.

Estimate $Y_{PBL}(s) = Y_1(s) - Y_2(s) = (w_1, w_2) \cdot (\hat{Y}_1(s), (\hat{Y}_2(s)))'$, with $(w_1, w_2) = (1, -1)$.

$MSPE = (w_1, w_2) \cdot Cov \left[ \begin{pmatrix} \hat{Y}_1(s) \\ \hat{Y}_2(s) \end{pmatrix} - \begin{pmatrix} Y_1(s) \\ Y_2(s) \end{pmatrix} \right] \cdot (w_1, w_2)'$. 
Prediction grid: $1^\circ \times 1^\circ$ rectangular; 396 basis functions at three levels of DGG etc. SE’s shown are truncated at .5 ppm, some higher.

Computation time on 3 GHz MacBook Pro $\approx 5$ minutes, half for computation of MSPE’s.

Performed for 89 overlapping three day blocks May 1 through July 31, 2003.
Conclusions

- Unlike more ad-hoc approaches, this methodology is *inferential*. It yields formal statistical estimates and their uncertainties relative to truth.
- We’ve demonstrated computational feasibility, but need to study trade-off’s related to number and kind of basis functions, binning and estimation of $K$’s etc.
- Difficult to judge results because we have combined AIRS observations with synthetic OCO-2. Need to create synthetic "truth", derive synthetic AIRS and OCO-2, fuse and judge results against "truth".
- Results depend crucially on biases and variances of measurement error terms. Instrument team validation results are the only sources of this information.
- Methodology is potentially applicable to many other situations.
- Next: Spatio-temporal Data Fusion (STDF) based on Fixed-Rank Filtering (Kang, Cressie, and Shi, 2009).

*Copyright 2010 California Institute of Technology, all rights reserved. Government sponsorship acknowledged.*