Spatial Statistical Data Fusion for Remote Sensing Applications

Hai Nguyen
Science Data Understanding Group
Jet Propulsion Laboratory
California Institute of Technology

May 18, 2010

Outline

Introduction

Review of Spatial Statistical Techniques

Spatial Statistical Data Fusion

Results

Section overview

Introduction to Data Fusion

- Data fusion overview
- Motivating example
- Data fusion problems and issues

Definition

Data fusion is the process of combining information from heterogeneous sources into a single composite picture of the relevant process, such that the composite picture is generally more accurate and complete than that derived from any single source alone (Hall, 2004).

Applicability

- Data collection is often incomplete, sparse, and yields incompatible information.
- Fusion techniques can make optimal use of such data.
- When investment in data collection is high, fusion gives the best return.

Applicability

- Data collection is often incomplete, sparse, and yields incompatible information.
- Fusion techniques can make optimal use of such data.
- When investment in data collection is high, fusion gives the best return.

Aerosols- definition and importance

An aerosol is defined as a suspension of fine, solid particles or liquid droplets in a gas.

- Examples include dust or particle emissions of diesel cars, and sulfates, nitrates and ammonium compounds.
- They contribute to air pollution.
- They affect the climate system in both direct and indirect ways.

Aerosols- definition and importance

An aerosol is defined as a suspension of fine, solid particles or liquid droplets in a gas.

- Examples include dust or particle emissions of diesel cars, and sulfates, nitrates and ammonium compounds.
- They contribute to air pollution.
- They affect the climate system in both direct and indirect ways.

Aerosols- definition and importance

An aerosol is defined as a suspension of fine, solid particles or liquid droplets in a gas.

- Examples include dust or particle emissions of diesel cars, and sulfates, nitrates and ammonium compounds.
- They contribute to air pollution.
- They affect the climate system in both direct and indirect ways.

Aerosols- definition and importance

An aerosol is defined as a suspension of fine, solid particles or liquid droplets in a gas.

- Examples include dust or particle emissions of diesel cars, and sulfates, nitrates and ammonium compounds.
- They contribute to air pollution.
- They affect the climate system in both direct and indirect ways.

Direct aerosol effect

The direct aerosol effects include strong back-scattering and absorption of light.

- Aerosols reflect solar radiation back into space by Mie scattering.
- According to the IPCC, the cooling effect of aerosols is about -.7 watts/m^2, compared to the total global warming effect of 2.5 watts/m^2.
- Light-absorbing aerosols causes absorption instead of scattering or transmission of solar radiation.

Direct aerosol effect

The direct aerosol effects include strong back-scattering and absorption of light.

- Aerosols reflect solar radiation back into space by Mie scattering.
- According to the IPCC, the cooling effect of aerosols is about -.7 watts/m^2, compared to the total global warming effect of 2.5 watts/m^2.
- Light-absorbing aerosols causes absorption instead of scattering or transmission of solar radiation.

Direct aerosol effect

The direct aerosol effects include strong back-scattering and absorption of light.

- Aerosols reflect solar radiation back into space by Mie scattering.
- According to the IPCC, the cooling effect of aerosols is about -.7 watts/m^2, compared to the total global warming effect of 2.5 watts/m^2.
- Light-absorbing aerosols causes absorption instead of scattering or transmission of solar radiation.

Direct aerosol effect

The direct aerosol effects include strong back-scattering and absorption of light.

- Aerosols reflect solar radiation back into space by Mie scattering.
- According to the IPCC, the cooling effect of aerosols is about -.7 watts/m^2, compared to the total global warming effect of 2.5 watts/m^2.
- Light-absorbing aerosols causes absorption instead of scattering or transmission of solar radiation.

Indirect aerosol effect

Aerosol contribute to climate change.

- Aerosol particles can act as nuclei for cloud formation.
- Studies suggest that aerosols shift asian tropical rainfall southward.

Indirect aerosol effect

Aerosol contribute to climate change.

- Aerosol particles can act as nuclei for cloud formation.
- Studies suggest that aerosols shift asian tropical rainfall southward.

Indirect aerosol effect

Aerosol contribute to climate change.

- Aerosol particles can act as nuclei for cloud formation.
- Studies suggest that aerosols shift asian tropical rainfall southward.

Remote sensing illustration

Our study uses data from two satellites:

- Multiangle Imaging SpectroRadiometer (MISR),
- Moderate Resolution Imaging Spectroradiometer (MODIS).

Data from MISR and MODIS

We use two datasets from MISR and MODIS as our motivating example.

- The domain is 30°S latitude to the equator (0° latitude) and the prime meridian (0° longtitude) to 30°E longtitude.
- The time period is from January 1-16, 2001.
- MISR has 9,308 observations, and MODIS has 47,695 observations.

MISR and MODIS aerosol maps

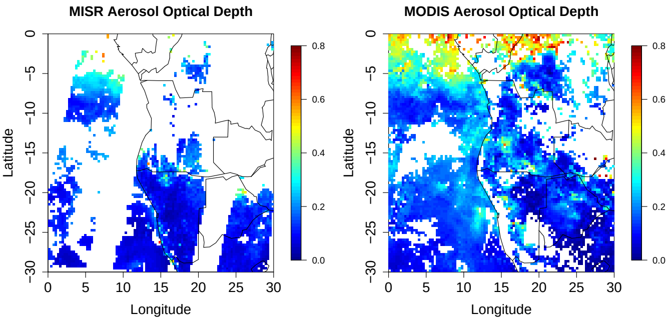

Figure: Maps of AOD.

Massive data size

MISR and MODIS, like many remote sensing instruments, return tens of thousands of data points per day.

- Traditional interpolation methods have computational complexity quadratic or cubic in data size.
- The rapid growth in remote sensing dataset size demands methods that scale well.

Massive data size

MISR and MODIS, like many remote sensing instruments, return tens of thousands of data points per day.

- Traditional interpolation methods have computational complexity quadratic or cubic in data size.
- The rapid growth in remote sensing dataset size demands methods that scale well.

Massive data size

MISR and MODIS, like many remote sensing instruments, return tens of thousands of data points per day.

- Traditional interpolation methods have computational complexity quadratic or cubic in data size.
- The rapid growth in remote sensing dataset size demands methods that scale well.

Change of support- example

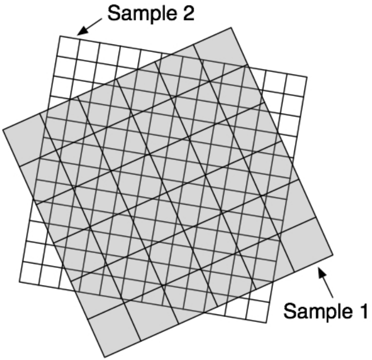

Figure: Example of different footprints. Source: Amy Braverman

Change of support - cont.

Data comes at areal-level support, but we need to estimate the point-level underlying continuous structure.

- Differences include alignments, orientation, shape, and size.
- Ignoring change of support makes inferences susceptable to the ecological fallacy.

Change of support - cont.

Data comes at areal-level support, but we need to estimate the point-level underlying continuous structure.

- Differences include alignments, orientation, shape, and size.
- Ignoring change of support makes inferences susceptable to the ecological fallacy.

Change of support - cont.

Data comes at areal-level support, but we need to estimate the point-level underlying continuous structure.

- Differences include alignments, orientation, shape, and size.
- Ignoring change of support makes inferences susceptable to the ecological fallacy.

Bias correction

Validation exercises (Paradise, 2007) indicate that MISR and MODIS, like other satellite instruments, likely have bias.

- Fusion methodology should include bias correction.
- Estimate of bias coefficients generally require other unbiased data sources.

Bias correction

Validation exercises (Paradise, 2007) indicate that MISR and MODIS, like other satellite instruments, likely have bias.

- Fusion methodology should include bias correction.
- Estimate of bias coefficients generally require other unbiased data sources.

Bias correction

Validation exercises (Paradise, 2007) indicate that MISR and MODIS, like other satellite instruments, likely have bias.

- Fusion methodology should include bias correction.
- Estimate of bias coefficients generally require other unbiased data sources.

Outline

Introduction

Review of Spatial Statistical Techniques

Spatial Statistical Data Fusion

Results

Section overview

Spatial statistics review

- Kriging
- Fixed-ranked kriging

Data assumption

Suppose that we have the following model:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}) \ , \quad \mathbf{s} \in D. \tag{1}$$

- $\mathbf{Z}$ is the data vector.
- $Y(\cdot)$ is the true hidden process.
- $\epsilon(\cdot)$ is a gaussian error process.

Functional form and MSPE

Given a data vector, $\mathbf{Z}$, we define our interpolator at location $\mathbf{s}_0$ as

$$\bar{Y}(\mathbf{s}_0) = \mathbf{a}'\mathbf{Z}.$$

We want to minimize

$$\begin{aligned} \mathrm{E}\left|\mathbf{a}'\mathbf{Z} - \mathbf{Y}(\mathbf{s}_0)\right|^2 &= \mathrm{Var}\left(\mathbf{a}'\mathbf{Z} - Y(\mathbf{s}_0)\right) \\ &= \mathbf{a}'\mathrm{Var}(\mathbf{Z})\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathrm{Cov}(\mathbf{Z}, Y(\mathbf{s}_0)) \\ &= \mathbf{a}'\mathbf{\Sigma}\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathbf{c}_0. \end{aligned}$$

Functional form and MSPE

Given a data vector, $\mathbf{Z}$, we define our interpolator at location $\mathbf{s}_0$ as

$$\bar{Y}(\mathbf{s}_0) = \mathbf{a}'\,\mathbf{Z}.$$

We want to minimize

$$\begin{aligned} \mathrm{E}\,\left|\mathbf{a}'\mathbf{Z} - \mathbf{Y}(\mathbf{s}_0)\right|^2 &= \mathrm{Var}\left(\mathbf{a}'\mathbf{Z} - Y(\mathbf{s}_0)\right) \\ &= \mathbf{a}'\mathrm{Var}(\mathbf{Z})\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathrm{Cov}(\mathbf{Z}, Y(\mathbf{s}_0)) \\ &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathbf{c}_0. \end{aligned}$$

Functional form and MSPE

Given a data vector, $\mathbf{Z}$, we define our interpolator at location $\mathbf{s}_0$ as

$$\bar{Y}(\mathbf{s}_0) = \mathbf{a}'\,\mathbf{Z}.$$

We want to minimize

$$\begin{aligned} \mathrm{E}\,\left|\mathbf{a}'\mathbf{Z} - \mathbf{Y}(\mathbf{s}_0)\right|^2 &= \mathrm{Var}\left(\mathbf{a}'\mathbf{Z} - Y(\mathbf{s}_0)\right) \\ &= \mathbf{a}'\mathrm{Var}(\mathbf{Z})\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathrm{Cov}(\mathbf{Z}, Y(\mathbf{s}_0)) \\ &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathbf{c}_0. \end{aligned}$$

Functional form and MSPE

Given a data vector, $\mathbf{Z}$, we define our interpolator at location $\mathbf{s}_0$ as

$$\bar{Y}(\mathbf{s}_0) = \mathbf{a}'\mathbf{Z}.$$

We want to minimize

$$\begin{aligned}
\mathrm{E}\,|\mathbf{a}'\mathbf{Z} - \mathbf{Y}(\mathbf{s}_0)|^2 &= \mathrm{Var}\left(\mathbf{a}'\mathbf{Z} - Y(\mathbf{s}_0)\right) \\
&= \mathbf{a}'\mathrm{Var}(\mathbf{Z})\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathrm{Cov}(\mathbf{Z}, Y(\mathbf{s}_0)) \\
&= \mathbf{a}'\mathbf{\Sigma}\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathbf{c}_0.
\end{aligned}$$

Solving for the kriging coefficients

We take the derivative of the expected prediction error with respect to **a**:

$$\frac{d}{d\mathbf{a}}(\mathbf{a}'\,\boldsymbol{\Sigma}\,\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\,\mathbf{c}_0) = 2\,\boldsymbol{\Sigma}\,\mathbf{a} - 2\,\mathbf{c}_0$$

Setting this equation to 0 and solving for **a**, we get

$$\begin{aligned} \boldsymbol{\Sigma}\,\mathbf{a} - \mathbf{c}_0 &= 0 \\ \boldsymbol{\Sigma}\,\mathbf{a} &= \mathbf{c}_0 \\ \mathbf{a} &= \boldsymbol{\Sigma}^{-1}\,\mathbf{c}_0. \end{aligned}$$

Solving for the kriging coefficients

We take the derivative of the expected prediction error with respect to **a**:

$$\frac{d}{d\mathbf{a}}(\mathbf{a}'\boldsymbol{\Sigma}\,\mathbf{a} + \mathrm{Var}(Y(\mathbf{s}_0)) - 2\mathbf{a}'\mathbf{c}_0) = 2\,\boldsymbol{\Sigma}\,\mathbf{a} - 2\,\mathbf{c}_0$$

Setting this equation to 0 and solving for **a**, we get

$$\begin{aligned} \boldsymbol{\Sigma}\,\mathbf{a} - \mathbf{c}_0 &= 0 \\ \boldsymbol{\Sigma}\,\mathbf{a} &= \mathbf{c}_0 \\ \mathbf{a} &= \boldsymbol{\Sigma}^{-1}\,\mathbf{c}_0. \end{aligned}$$

Comments on kriging

Kriging is a very popular spatial analysis methodology

- It can handle change of support.
- It can produce estimates of the MSPE.

Disadvantages of kriging include

- Kriging requires inversion of the covariance matrix, $\mathbf{\Sigma}$.
- For small datasets, estimates of the covariance function requires simplifying assumptions.

Comments on kriging

Kriging is a very popular spatial analysis methodology

- It can handle change of support.
- It can produce estimates of the MSPE.

Disadvantages of kriging include

- Kriging requires inversion of the covariance matrix, $\mathbf{\Sigma}$.
- For small datasets, estimates of the covariance function requires simplifying assumptions.

Covariance model

Kriging requires the inversion of the $N \times N$ covariance matrix, $\mathbf{\Sigma} = \mathbf{C} + \sigma^2 \mathbf{V}$.

Cressie and Johannesson (2008) model $C(\mathbf{u}, \mathbf{v})$ as

$$C(\mathbf{u}, \mathbf{v}) = \mathbf{S}(\mathbf{u})' \mathbf{K} \mathbf{S}(\mathbf{v}) \quad \mathbf{u}, \mathbf{v} \in D.$$

- $C(\cdot, \cdot)$ is the covariance function.
- $\mathbf{u}, \mathbf{v}$ are locations in the domain, D.
- $\mathbf{S}(\mathbf{v})$ is an basis expansion of location $\mathbf{s}$ into r-dimensions.
- $\mathbf{K}$ is an $r \times r$ matrix.

Covariance model

Kriging requires the inversion of the $N \times N$ covariance matrix, $\mathbf{\Sigma} = \mathbf{C} + \sigma^2 \mathbf{V}$.
Cressie and Johannesson (2008) model $C(\mathbf{u}, \mathbf{v})$ as

$$C(\mathbf{u}, \mathbf{v}) = \mathbf{S}(\mathbf{u})' \mathbf{K} \mathbf{S}(\mathbf{v}) \quad \mathbf{u}, \mathbf{v} \in D.$$

- $C(\cdot, \cdot)$ is the covariance function.
- $\mathbf{u}, \mathbf{v}$ are locations in the domain, D.
- $\mathbf{S}(\mathbf{v})$ is an basis expansion of location $\mathbf{s}$ into r-dimensions.
- $\mathbf{K}$ is an $r \times r$ matrix.

Inverting the FRK covariance matrix

We model $\mathbf{\Sigma} = \sigma^2\mathbf{V} + \mathbf{S}'\mathbf{K}\mathbf{S}$.

- $\mathbf{S}'$ is an $N \times r$ matrix.
- $\mathbf{K}$ is an $r \times r$ matrix.

Using the Sherman-Morrison-Woodbury formula, the exact inversion of $\mathbf{\Sigma}$ is,

$$\mathbf{\Sigma}^{-1} = (\sigma^2\mathbf{V})^{-1} - (\sigma^2\mathbf{V})^{-1}\mathbf{S}'\left(\mathbf{K}^{-1} + \mathbf{S}(\sigma^2\mathbf{V})^{-1}\mathbf{S}'\right)^{-1}\mathbf{S}'(\sigma^2\mathbf{V})^{-1}.$$

Inverting the FRK covariance matrix

We model $\mathbf{\Sigma} = \sigma^2\mathbf{V} + \mathbf{S}'\mathbf{K}\mathbf{S}$.

- $\mathbf{S}'$ is an $N \times r$ matrix.
- $\mathbf{K}$ is an $r \times r$ matrix.

Using the Sherman-Morrison-Woodbury formula, the exact inversion of $\mathbf{\Sigma}$ is,

$$\mathbf{\Sigma}^{-1} = (\sigma^2\mathbf{V})^{-1} - (\sigma^2\mathbf{V})^{-1}\mathbf{S}'\left(\mathbf{K}^{-1} + \mathbf{S}(\sigma^2\mathbf{V})^{-1}\mathbf{S}'\right)^{-1}\mathbf{S}'(\sigma^2\mathbf{V})^{-1}.$$

Notes on FRK

Some notes:

- FRK has order of computation $O(Nr^2)$.
- The parameter $\mathbf{K}$ may be estimated by minimizing with respect to the Frobenius norm.
- $S(\cdot)$ could be any basis expansion.

FRK performed on separately on MISR and MODIS

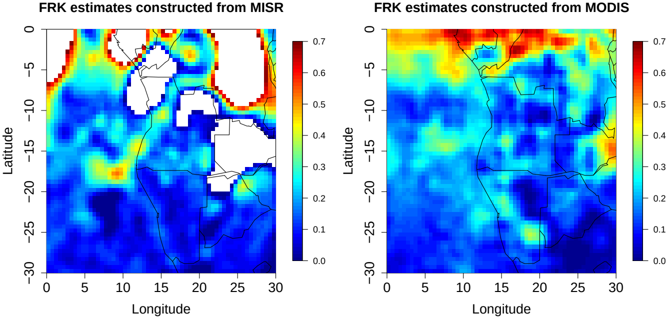

Figure: Single-dataset FRK estimates.

FRK performed on separately on MISR and MODIS

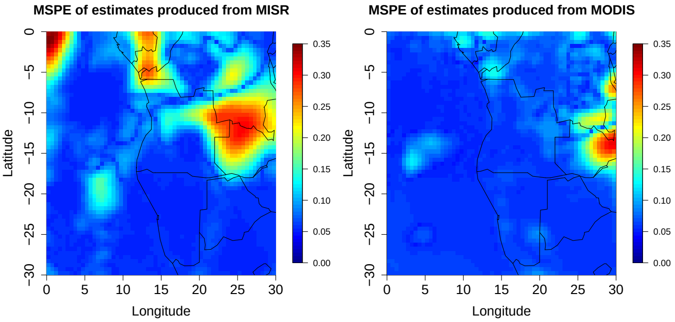

Figure: Prediction errors of single-dataset FRK estimates.

Outline

Introduction

Review of Spatial Statistical Techniques

Spatial Statistical Data Fusion

Results

Section overview

Spatial Statistical Data Fusion

- Assumptions about data structure
- SSDF estimates
- Solving the fusion equations
- Scalability
- Change of support
- Estimating SSDF parameters

Model assumptions

We assume the data are generated according to the following model:

$$\begin{aligned} \mathbf{Z}_i &= (Z_i(B_{i1}), Z_i(B_{i2}), \ldots, Z_i(B_{iN_i}))', \\ Z_i(B_{ij}) &= \frac{1}{|B_{ij}|} \int_{\mathbf{u} \in B_{ij}} Y(\mathbf{u}) d\mathbf{u} + \epsilon_i(B_{ij}), \end{aligned} \tag{2}$$

where

- B_{ij} is the jth footprint from data set i,
- $\mathbf{Z}_i$ is the vector of response variable from dataset i ,
- $Y(\cdot)$ is the true process,
- $\epsilon_i(B_{ij})$ is the error process.

Error process assumptions

We assume a multiplicative bias model:

$$\mathrm{E}(\epsilon_i(B_{ij})) = c_i\mu,$$
$$\mathrm{Var}(\epsilon_i(B_{ij})) = \sigma_i^2.$$

Functional form for SSDF

To estimate the underlying process $Y(\cdot)$ at $\mathbf{s}$, we construct a linear combination of $\mathbf{Z}_1$ and $\mathbf{Z}_2$,

$$\hat{Y}(\mathbf{s}) = \mathbf{a}_{1\mathbf{s}}'\mathbf{Z}_1 + \mathbf{a}_{2\mathbf{s}}'\mathbf{Z}_2. \tag{3}$$

We seek values of $\mathbf{a}_{1\mathbf{s}}$ and $\mathbf{a}_{2\mathbf{s}}$ that minimize the mean-squared prediction error for this estimate.

$$\text{minimize MSPE} = \mathrm{E}(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}))^2 \quad \text{subject to} \quad \mathrm{E}(\hat{Y}(\mathbf{s})) = \mu. \tag{4}$$

Functional form for SSDF

To estimate the underlying process $Y(\cdot)$ at $\mathbf{s}$, we construct a linear combination of $\mathbf{Z}_1$ and $\mathbf{Z}_2$,

$$\hat{Y}(\mathbf{s}) = \mathbf{a}_{1\mathbf{s}}' \mathbf{Z}_1 + \mathbf{a}_{2\mathbf{s}}' \mathbf{Z}_2. \tag{3}$$

We seek values of $\mathbf{a}_{1\mathbf{s}}$ and $\mathbf{a}_{2\mathbf{s}}$ that minimize the mean-squared prediction error for this estimate.

$$\text{minimize MSPE} = \mathrm{E}(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}))^2 \quad \text{subject to} \quad \mathrm{E}(\hat{Y}(\mathbf{s})) = \mu. \tag{4}$$

Mean-squared prediction error expansion

The expanded form for the mean-squared prediction error is

$$
\begin{aligned}
E\left(Y(\mathbf{s}) - \hat{Y}(\mathbf{s})\right)^2 &= \mathrm{Var}(\mathbf{a}_{1\mathbf{s}}'\mathbf{Z}_1 + \mathbf{a}_{2\mathbf{s}}'\mathbf{Z}_2 - Y(\mathbf{s})), \\
&= \mathbf{a}_{1\mathbf{s}}'\mathrm{Var}(\mathbf{Z}_1)\mathbf{a}_{1\mathbf{s}} + \mathbf{a}_{2\mathbf{s}}'\mathrm{Var}(\mathbf{Z}_2)\mathbf{a}_{2\mathbf{s}} + \mathrm{Var}(Y(\mathbf{s})) \\
&\quad + 2\mathbf{a}_{1\mathbf{s}}'\mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2)\mathbf{a}_{2\mathbf{s}} - 2\mathbf{a}_{1\mathbf{s}}'\mathrm{Cov}(\mathbf{Z}_1, Y(\mathbf{s})) \\
&\quad - 2\mathbf{a}_{2\mathbf{s}}'\mathrm{Cov}(\mathbf{Z}_2, Y(\mathbf{s})). \qquad (5)
\end{aligned}
$$

Mean-squared prediction error expansion

The expanded form for the mean-squared prediction error is

$$\begin{aligned} E(Y(\mathbf{s}) - \hat{Y}(\mathbf{s}))^2 &= \mathrm{Var}(\mathbf{a}_{1\mathbf{s}}'\mathbf{Z}_1 + \mathbf{a}_{2\mathbf{s}}'\mathbf{Z}_2 - Y(\mathbf{s})), \\ &= \mathbf{a}_{1\mathbf{s}}'\mathrm{Var}(\mathbf{Z}_1)\mathbf{a}_{1\mathbf{s}} + \mathbf{a}_{2\mathbf{s}}'\mathrm{Var}(\mathbf{Z}_2)\mathbf{a}_{2\mathbf{s}} + \mathrm{Var}(Y(\mathbf{s})) \\ &\quad + 2\mathbf{a}_{1\mathbf{s}}'\mathrm{Cov}(\mathbf{Z}_1, \mathbf{Z}_2)\mathbf{a}_{2\mathbf{s}} - 2\mathbf{a}_{1\mathbf{s}}'\mathrm{Cov}(\mathbf{Z}_1, Y(\mathbf{s})) \\ &\quad - 2\mathbf{a}_{2\mathbf{s}}'\mathrm{Cov}(\mathbf{Z}_2, Y(\mathbf{s})). \end{aligned} \tag{5}$$

Unbiasedness constraint

We wish to find the fusion coefficients that minimize the MSPE, subject to the unbiasedness contraint,

$$\begin{aligned} \mathrm{E}(Y_{\mathbf{s}}) = \mu = \mathrm{E}\left(\mathbf{a}'_{1\mathbf{s}}\mathbf{Z}_1 + \mathbf{a}'_{2\mathbf{s}}\mathbf{Z}_2\right) &= E(\hat{Y}_{\mathbf{s}}), \\ \mu &= \mathbf{a}'_{1\mathbf{s}}\mathbf{1}_{N_1}(1+c_1)\mu + \mathbf{a}'_{2\mathbf{s}}\mathbf{1}_{N_2}(1+c_2)\mu, \\ 0 &= \mathbf{a}'_{1\mathbf{s}}\mathbf{1}_{N_1}(1+c_1) + \mathbf{a}'_{2\mathbf{s}}\mathbf{1}_{N_2}(1+c_2) - 1. \end{aligned} \tag{6}$$

The Lagrangian

We wish to find the optimal fusion coefficients such that the MSPE is minimized. Let:
$\mathbf{\Sigma}_{ii} = \mathrm{Var}(\mathbf{Z}_i)$, $\mathbf{\Sigma}_{ik} = \mathrm{Cov}(\mathbf{Z}_i, \mathbf{Z}_k)$, and $\mathbf{c}_{is} = \mathrm{Cov}(\mathbf{Z}_i, \mathbf{s})$.

$$L = \mathbf{a}'_{1s}\mathbf{\Sigma}_{11}\mathbf{a}_{1s} + \mathbf{a}'_{2s}\mathbf{\Sigma}_{22}\mathbf{a}_{2s} + \sigma_{s,s} + 2\mathbf{a}'_{1s}\mathbf{\Sigma}_{12}\mathbf{a}_{2s} - 2\mathbf{a}'_{1s}\mathbf{c}_{1s} - 2\mathbf{a}'_{2s}\mathbf{c}_{2s} + 2m\left[\mathbf{a}'_{1s}\mathbf{1}_{N_1}(1+c_1) + \mathbf{a}'_{2s}\mathbf{1}_{N_2}(1+c_2) - 1\right]. \quad (7)$$

The Lagrangian

We wish to find the optimal fusion coefficients such that the MSPE is minimized. Let:
$\mathbf{\Sigma}_{ii} = \mathrm{Var}(\mathbf{Z}_i)$, $\mathbf{\Sigma}_{ik} = \mathrm{Cov}(\mathbf{Z}_i, \mathbf{Z}_k)$, and $\mathbf{c}_{is} = \mathrm{Cov}(\mathbf{Z}_i, \mathbf{s})$.

$$
\begin{aligned}
L = {} & \mathbf{a}'_{1\mathbf{s}}\mathbf{\Sigma}_{11}\mathbf{a}_{1\mathbf{s}} + \mathbf{a}'_{2\mathbf{s}}\mathbf{\Sigma}_{22}\mathbf{a}_{2\mathbf{s}} + \sigma_{\mathbf{s},\mathbf{s}} + 2\mathbf{a}'_{1\mathbf{s}}\mathbf{\Sigma}_{12}\mathbf{a}_{2\mathbf{s}} - 2\mathbf{a}'_{1\mathbf{s}}\mathbf{c}_{1\mathbf{s}} - 2\mathbf{a}'_{2\mathbf{s}}\mathbf{c}_{2\mathbf{s}} \\
& + 2m\left[\mathbf{a}'_{1\mathbf{s}}\mathbf{1}_{N_1}(1 + c_1) + \mathbf{a}'_{2\mathbf{s}}\mathbf{1}_{N_2}(1 + c_2) - 1\right]. \qquad (7)
\end{aligned}
$$

Derivatives w.r.t. the fusion coefficients

Differentiating L with respect to $\mathbf{a}_{1s}$, $\mathbf{a}_{2s}$, and m, and setting the resulting expressions equal to zero, we have

$$
\begin{aligned}
\frac{\partial L}{\partial \mathbf{a}_1} &= \boldsymbol{\Sigma}_1 \mathbf{a}_{1s} + \boldsymbol{\Sigma}_{12} \mathbf{a}_{2s} - \mathbf{c}_{1s} + m \mathbf{1}_{N_1}(1 + c_1) = 0, \\
\frac{\partial L}{\partial \mathbf{a}_2} &= \boldsymbol{\Sigma}_2 \mathbf{a}_{2s} + \boldsymbol{\Sigma}_{12} \mathbf{a}_{1s} - \mathbf{c}_{2s} + m \mathbf{1}_{N_2}(1 + c_2) = 0, \\
\frac{\partial L}{\partial m} &= \left[\mathbf{a}'_{1s} \mathbf{1}_{N_1}(1 + c_1) + \mathbf{a}'_{2s} \mathbf{1}_{N_2}(1 + c_2) - 1 \right] = 0. \qquad (8)
\end{aligned}
$$

In matrix forms

We can express the equation for the fusion coefficients in matrix form,

$$\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \mathbf{1}_{N_1}(1+c_1) \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \mathbf{1}_{N_2}(1+c_2) \\ \mathbf{1}'_{N_1}(1+c_1) & \mathbf{1}'_{N_2}(1+c_2) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1s} \\ \mathbf{a}_{2s} \\ m \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{1s} \\ \mathbf{c}_{2s} \\ 1 \end{bmatrix}. \quad (9)$$

Solving for the fusion coefficients

The fusion coefficients are

$$\begin{aligned}
\mathbf{a}_{1s} &= -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{a}_{2s} + \boldsymbol{\Sigma}_{11}^{-1}\mathbf{c}_{1s} - \boldsymbol{\Sigma}_{11}^{-1}\mathbf{1}_{N_1}(1+c_1)m, \\
\mathbf{a}_{2s} &= -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{a}_{1s} + \boldsymbol{\Sigma}_{22}^{-1}\mathbf{c}_{2s} - \boldsymbol{\Sigma}_{22}^{-1}\mathbf{1}_{N_2}(1+c_2)m,
\end{aligned}$$

which can be expressed as

$$\begin{aligned}
\mathbf{a}_{1s} &= \mathbf{A}_1^{-1}(\mathbf{B}_1 + \mathbf{C}_1 m), \qquad \text{and} \\
\mathbf{a}_{2s} &= \mathbf{A}_2^{-1}(\mathbf{B}_2 + \mathbf{C}_2 m). \qquad (10)
\end{aligned}$$

Solving for the fusion coefficients

The fusion coefficients are

$$\begin{aligned}
\mathbf{a}_{1s} &= -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{a}_{2s} + \boldsymbol{\Sigma}_{11}^{-1}\mathbf{c}_{1s} - \boldsymbol{\Sigma}_{11}^{-1}\mathbf{1}_{N_1}(1+c_1)m, \\
\mathbf{a}_{2s} &= -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{a}_{1s} + \boldsymbol{\Sigma}_{22}^{-1}\mathbf{c}_{2s} - \boldsymbol{\Sigma}_{22}^{-1}\mathbf{1}_{N_2}(1+c_2)m,
\end{aligned}$$

which can be expressed as

$$\begin{aligned}
\mathbf{a}_{1s} &= \mathbf{A}_1^{-1}(\mathbf{B}_1 + \mathbf{C}_1 m), \qquad \text{and} \\
\mathbf{a}_{2s} &= \mathbf{A}_2^{-1}(\mathbf{B}_2 + \mathbf{C}_2 m).
\end{aligned} \tag{10}$$

Solving for the fusion coefficients- cont.

$\mathbf{A}_i, \mathbf{B}_i$ and $\mathbf{C}_i$ are defined as

$$
\begin{aligned}
\mathbf{A}_1 &\equiv (\mathbf{I}_{N_1} - \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}), \quad \text{an } N_1 \times N_1 \text{ matrix,} \\
\mathbf{A}_2 &\equiv (\mathbf{I}_{N_2} - \mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}), \quad \text{an } N_2 \times N_2 \text{ matrix,} \\
\mathbf{B}_1 &\equiv \mathbf{\Sigma}_{11}^{-1}(\mathbf{c}_{1s} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{c}_{2s}), \quad \text{an } N_1\text{-dimensional vector,} \\
\mathbf{B}_2 &\equiv \mathbf{\Sigma}_{22}^{-1}(\mathbf{c}_{2s} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{c}_{1s}), \quad \text{an } N_2\text{-dimensional vector,} \\
\mathbf{C}_1 &= -\mathbf{\Sigma}_{11}^{-1}(\mathbf{1}_{N_1}(1+c_1) - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{1}_{N_2}(1+c_2)), \quad \text{an } N_1\text{-dimensional vector,} \\
\mathbf{C}_2 &= -\mathbf{\Sigma}_{22}^{-1}(\mathbf{1}_{N_2}(1+c_2) - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{1}_{N_1}(1+c_1)), \quad \text{an } N_2\text{-dimensional vector}\,,
\end{aligned}
\tag{11}
$$

Using FRK covariance structure

We need to invert $\mathbf{A}_i$. If we use the FRK covariance structure, the inversions are

$$
\begin{aligned}
\boldsymbol{\Sigma}_{ii}^{-1} &= (\sigma_i^2 \mathbf{V}_i + \mathbf{S}_i' \mathbf{K} \mathbf{S}_i)^{-1}, \\
&= (\sigma_i^2 \mathbf{V}_i)^{-1} - (\sigma_i^2 \mathbf{V}_i)^{-1} \mathbf{S}_i' \left(\mathbf{K}^{-1} + \mathbf{S}_i (\sigma_i^2 \mathbf{V}_i)^{-1} \mathbf{S}_i' \right)^{-1} \mathbf{S}_i (\sigma_i^2 \mathbf{V}_i)^{-1}. \\
\mathbf{A}_1^{-1} &= (\mathbf{I}_{N_1} - \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21})^{-1}, \\
&= (\mathbf{I}_{N_1} - \boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_1' \mathbf{K} \mathbf{S}_2 \, \boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_2' \mathbf{K} \mathbf{S}_1)^{-1} \quad \text{(by the Woodbury identity)}, \\
&= \mathbf{I}_{N_1} - \boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_1' \left(\mathbf{K}^{-1} + \mathbf{S}_2 \boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_2' \mathbf{K} \mathbf{S}_1 \boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_1' \right)^{-1} \mathbf{S}_2 \boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_2' \mathbf{K} \mathbf{S}_1, \\
\mathbf{A}_2^{-1} &= \mathbf{I}_{N_2} - \boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_2' \left(\mathbf{K}^{-1} + \mathbf{S}_1 \boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_1' \mathbf{K} \mathbf{S}_2 \boldsymbol{\Sigma}_{22}^{-1} \mathbf{S}_2' \right)^{-1} \mathbf{S}_1 \boldsymbol{\Sigma}_{11}^{-1} \mathbf{S}_1' \mathbf{K} \mathbf{S}_2.
\end{aligned}
$$

Scalability- notes

- Given the FRK structure, the inversion is exact.
- Only $\mathbf{K}$, $\left(\mathbf{K}^{-1} + \mathbf{S}_i(\sigma_i^2\mathbf{V}_i)^{-1}\mathbf{S}_i'\right)$, $\left(\mathbf{K}^{-1} + \mathbf{S}_2\,\mathbf{\Sigma}_{22}^{-1}\,\mathbf{S}_2'\mathbf{K}\mathbf{S}_1\mathbf{\Sigma}_{11}^{-1}\,\mathbf{S}_1'\right)$, and $\left(\mathbf{K}^{-1} + \mathbf{S}_1\,\mathbf{\Sigma}_{11}^{-1}\,\mathbf{S}_1'\mathbf{K}\mathbf{S}_2\mathbf{\Sigma}_{22}^{-1}\,\mathbf{S}_2'\right)$ need inversion.
- We incur constant cost of $O(r^3)$ for these inversions.
- The overall computational complexity of the method is $O(Nr^2)$.

Different spatial support

So far, we assumed the following covariance model:

$$C(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{S}(\mathbf{s}_1)'\mathbf{K}\mathbf{S}(\mathbf{s}_2). \tag{12}$$

Our data, however, are observed at areal-level.

- We need to estimate $\mathbf{K}$ from areal-level data.

Data generation model review

Recall that we assume the following model for the data

$$
\begin{aligned}
\mathbf{Z} &= (Z(B_1), Z(B_2), \ldots, Z(B_n))', && (13)\\
Z(B_i) &= \frac{1}{|B_i|} \int_{\mathbf{u} \in B_i} Y(\mathbf{u}) d\mathbf{u} + \epsilon(B_i). && (14)
\end{aligned}
$$

Point-level and areal-level covariance

We can find expression for the areal-covariance in terms of the point-level covariance,

$$
\begin{aligned}
\mathrm{Cov}(Z(B_k), Z(B_l)) &= \mathrm{Cov}\left(\frac{1}{|B_k|} \int_{\mathbf{u} \in B_k} Y(\mathbf{u}) d\mathbf{u} \; + \; \epsilon(B_k), \right. \\
&\qquad \left. \frac{1}{|B_l|} \int_{\mathbf{v} \in B_l} Y(\mathbf{v}) d\mathbf{v} \; + \; \epsilon(B_l) \right), \\
&= \mathrm{Cov}\left(\frac{1}{|B_k|} \int_{\mathbf{u} \in B_k} Y(\mathbf{u}) d\mathbf{u} \, , \; \frac{1}{|B_l|} \int_{\mathbf{v} \in B_l} Y(\mathbf{v}) d\mathbf{v} \right) \\
&\quad + \mathrm{Cov}\left(\frac{1}{|B_k|} \int_{\mathbf{u} \in B_k} Y(\mathbf{u}) d\mathbf{u}, \epsilon(B_l) \right) \\
&\quad + \mathrm{Cov}\left(\frac{1}{|B_l|} \int_{\mathbf{v} \in B_l} Y(\mathbf{v}) d\mathbf{v}, \epsilon(B_k) \right) \\
&\quad + \mathrm{Cov}\left(\epsilon(B_k), \epsilon(B_l) \right)) \, , \qquad (15)
\end{aligned}
$$

Point-level and areal-level covariance

$$
\begin{aligned}
\mathrm{Cov}(Z(B_k), Z(B_l)) &= \frac{1}{|B_k|}\frac{1}{|B_l|}\int_{\mathbf{u}\in B_k}\int_{\mathbf{v}\in B_l}\mathrm{Cov}\left(Y(\mathbf{u}), Y(\mathbf{v})\right) \quad d\mathbf{u}\; d\mathbf{v}, \\
&= \frac{1}{|B_k|}\int_{\mathbf{u}\in B_k}\mathbf{S}(\mathbf{u})'d\mathbf{u} \quad \mathbf{K} \quad \frac{1}{|B_l|}\int_{\mathbf{v}\in B_l}\mathbf{S}(\mathbf{v}) \quad d\mathbf{v}, \\
&= \widetilde{\mathbf{S}}(B_k)' \; \mathbf{K} \; \widetilde{\mathbf{S}}(B_l), \qquad (16)
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{\mathbf{S}}(B_i) &= \left(\widetilde{S}_1(B_i), \widetilde{S}_2(B_i), \ldots, \widetilde{S}_r(B_i)\right), \\
\text{with} \qquad \widetilde{S}_j(B_i) &= \frac{1}{|B_i|}\int_{\mathbf{u}\in B_i} S_j(\mathbf{u}) \quad d\mathbf{u}.
\end{aligned}
$$

Point-level and areal-level covariance

$$\begin{aligned}
\mathrm{Cov}(Z(B_k), Z(B_l)) &= \frac{1}{|B_k|}\frac{1}{|B_l|}\int_{\mathbf{u}\in B_k}\int_{\mathbf{v}\in B_l} \mathrm{Cov}\left(Y(\mathbf{u}), Y(\mathbf{v})\right) \; d\mathbf{u}\; d\mathbf{v}, \\
&= \frac{1}{|B_k|}\int_{\mathbf{u}\in B_k} \mathbf{S}(\mathbf{u})' d\mathbf{u} \quad \mathbf{K} \quad \frac{1}{|B_l|}\int_{\mathbf{v}\in B_l} \mathbf{S}(\mathbf{v}) \; d\mathbf{v}, \\
&= \widetilde{\mathbf{S}}(B_k)' \; \mathbf{K} \; \widetilde{\mathbf{S}}(B_l),
\end{aligned} \tag{16}$$

where

$$\begin{aligned}
\widetilde{\mathbf{S}}(B_i) &= \left(\widetilde{S}_1(B_i), \widetilde{S}_2(B_i), \ldots, \widetilde{S}_r(B_i)\right), \\
\text{with} \qquad \widetilde{S}_j(B_i) &= \frac{1}{|B_i|}\int_{\mathbf{u}\in B_i} S_j(\mathbf{u}) \; d\mathbf{u}.
\end{aligned}$$

Point-level and areal-level covariance

$$
\begin{aligned}
\mathrm{Cov}(Z(B_k), Z(B_l)) &= \frac{1}{|B_k|}\frac{1}{|B_l|}\int_{\mathbf{u}\in B_k}\int_{\mathbf{v}\in B_l} \mathrm{Cov}\left(Y(\mathbf{u}), Y(\mathbf{v})\right) \;\; d\mathbf{u}\; d\mathbf{v}, \\
&= \frac{1}{|B_k|}\int_{\mathbf{u}\in B_k} \mathbf{S}(\mathbf{u})' d\mathbf{u} \;\; \mathbf{K} \;\; \frac{1}{|B_l|}\int_{\mathbf{v}\in B_l} \mathbf{S}(\mathbf{v}) \;\; d\mathbf{v}, \\
&= \widetilde{\mathbf{S}}(B_k)' \; \mathbf{K} \; \widetilde{\mathbf{S}}(B_l), \qquad (16)
\end{aligned}
$$

where

$$
\begin{aligned}
\widetilde{\mathbf{S}}(B_i) &= \left(\widetilde{S}_1(B_i), \widetilde{S}_2(B_i), \ldots, \widetilde{S}_r(B_i)\right), \\
\text{with} \quad \widetilde{S}_j(B_i) &= \frac{1}{|B_i|}\int_{\mathbf{u}\in B_i} S_j(\mathbf{u}) \;\; d\mathbf{u}.
\end{aligned}
$$

Estimating $\mathbf{K}$

Given the empirical covariance matrices, $\boldsymbol{\Sigma}_{ij}$, and $\mathbf{S}$, we model the covariance as

$$
\begin{aligned}
\boldsymbol{\Sigma}_{11} &= \tilde{\mathbf{S}}_1' \mathbf{K} \tilde{\mathbf{S}}_1 + \sigma_1^2 \mathbf{V}_1', \\
\boldsymbol{\Sigma}_{12} &= \tilde{\mathbf{S}}_1' \mathbf{K} \tilde{\mathbf{S}}_2, \\
\boldsymbol{\Sigma}_{21} &= \tilde{\mathbf{S}}_2' \mathbf{K} \tilde{\mathbf{S}}_1, \\
\boldsymbol{\Sigma}_{22} &= \tilde{\mathbf{S}}_2' \mathbf{K} \tilde{\mathbf{S}}_2 + \sigma_2^2 \mathbf{V}_2'.
\end{aligned}
$$

Estimating **K**- cont.

Rewrite the previous equation as

$$\begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{11} - \sigma_1^2 \mathbf{V}_1 & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22} - \sigma_2^2 \mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{S}}_1' \\ \tilde{\mathbf{S}}_2' \end{bmatrix} \mathbf{K} \begin{bmatrix} \tilde{\mathbf{S}}_1 & \tilde{\mathbf{S}}_2 \end{bmatrix}. \tag{17}$$

Estimating **K**- cont.

Minimize the differences with respect to the Frobenius norm,

$$\mathbf{K} = \mathbf{Q}(\mathbf{R}')^{-1} \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{11} - \sigma_1^2 \mathbf{V}_1 & \hat{\boldsymbol{\Sigma}}_{12} \\ \hat{\boldsymbol{\Sigma}}_{21} & \hat{\boldsymbol{\Sigma}}_{22} - \sigma_2^2 \mathbf{V}_2 \end{bmatrix} \mathbf{R}'\mathbf{Q}'. \tag{18}$$

- **Q** and **R** are derived from QR decomposition of **S**.

Outline

Introduction

Review of Spatial Statistical Techniques

Spatial Statistical Data Fusion

Results

Section overview

Spatial Statistical Data Fusion

- SSDF on MISR and MODIS
- SSDF on synthetic example

Fusing MISR and MODIS

We apply SSDF to MISR and MODIS data.

- We use a 30×30 grid for estimating the empirical covariance matrices.
- We use 342 bisquare basis functions for $\mathbf{S}$.

Maps of basis functions and binning grid

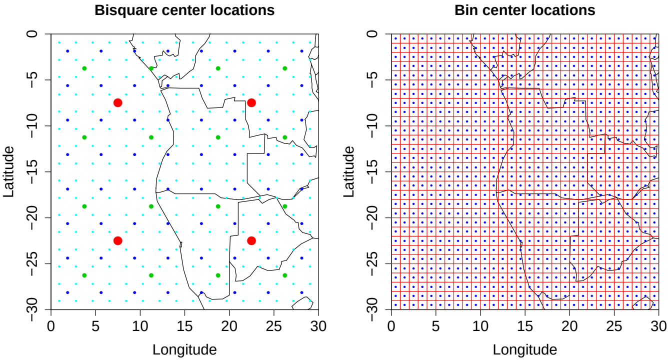

Figure: Map of bin centers and basis centers.

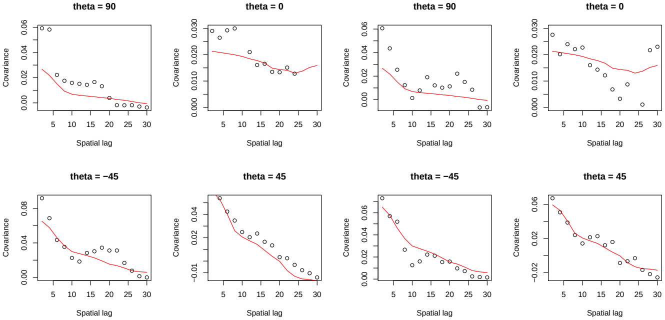

Figure: Empirical vs. estimated covariances for MISR (left four) and MODIS (right four).

Review of the raw data

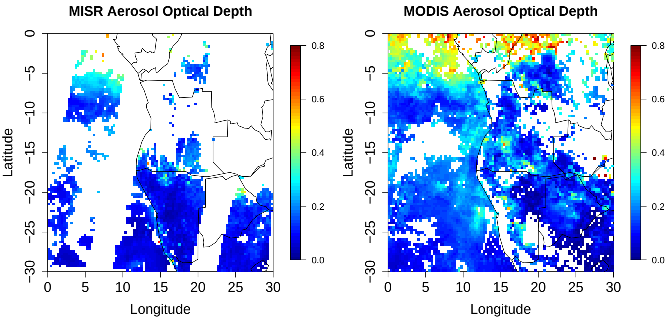

Figure: Maps of AOD.

SSDF and single-dataset estimates

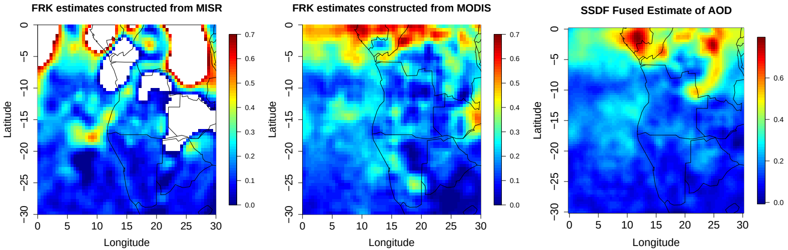

Figure: Estimates of AOD using single-dataset FRK and SSDF.

MSPE for estimates

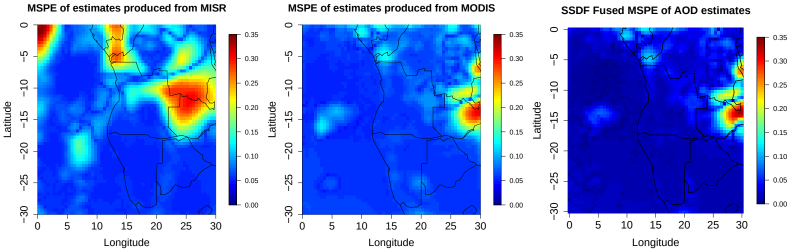

Figure: MSPE of estimates using single-dataset FRK and SSDF.

AERONET locations

Over the same time and spatial region, we have data for 3 AErosol ROBotics NETwork (AERONET) Stations.

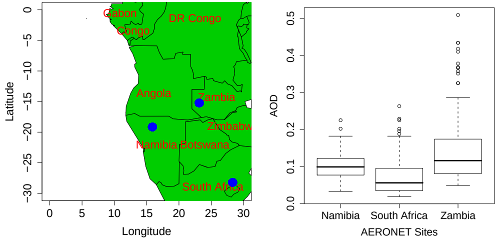

Figure: Left: Locations of AERONET stations. Right: Boxplots of observations at AERONET sites.

AERONET comparisons to model estimates

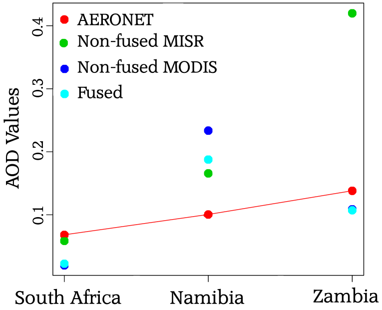

Figure: Individual FRK predictions vs SSDF predictions for AERONET sites.