Spatial Statistics for Climate and Weather

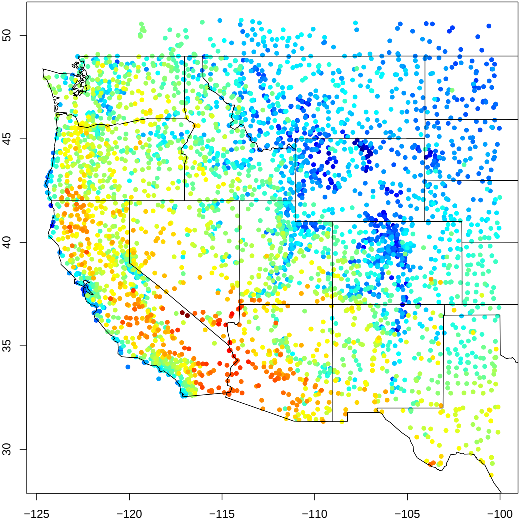
Will Kleiber

Department of Applied Mathematics
University of Colorado
Boulder, CO

Uncertainty Modeling in the Analysis of Weather, Climate and Hydrological Extremes, Banff, Canada
June 17, 2016

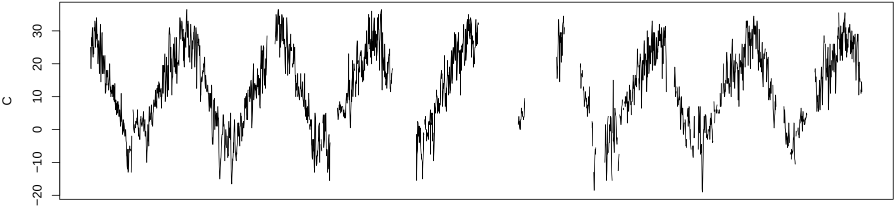
Spatial Statistics

Removing the noise (smoothing)



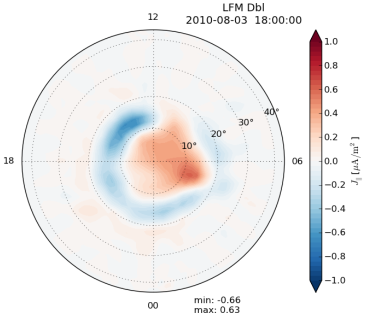
Spatial Statistics

Filling in the gaps (prediction)



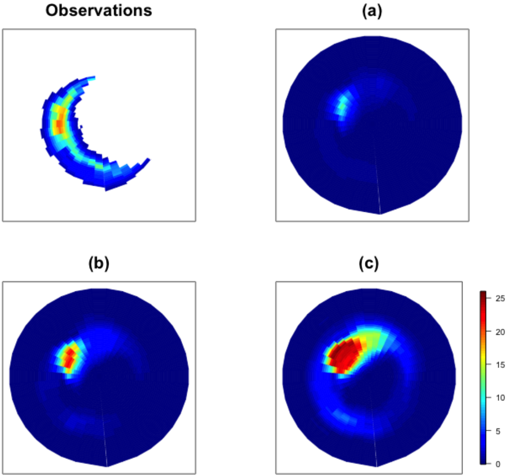
Spatial Statistics

Quantify differences (characterization)



Spatial Statistics

What-if scenarios (simulation)



Spatial Statistics

What is spatial statistics?

Typical goals:

- Removing the noise (smoothing)
- Filling in the gaps (prediction)
- Quantify differences (characterization)
- What-if scenarios (simulation)

Important in all goals is to quantify the uncertainty.

Spatial Statistics

What is spatial statistics?

Typical goals:

- Removing the noise (smoothing)
- Filling in the gaps (prediction)
- Quantify differences (characterization)
- What-if scenarios (simulation)

Important in all goals is to quantify the uncertainty.

Outline:

- Nonstationary processes
- Large datasets
- Multivariate processes

Spatial Statistics

What is spatial statistics?

Typical goals:

- Removing the noise (smoothing)
- Filling in the gaps (prediction)
- Quantify differences (characterization)
- What-if scenarios (simulation)

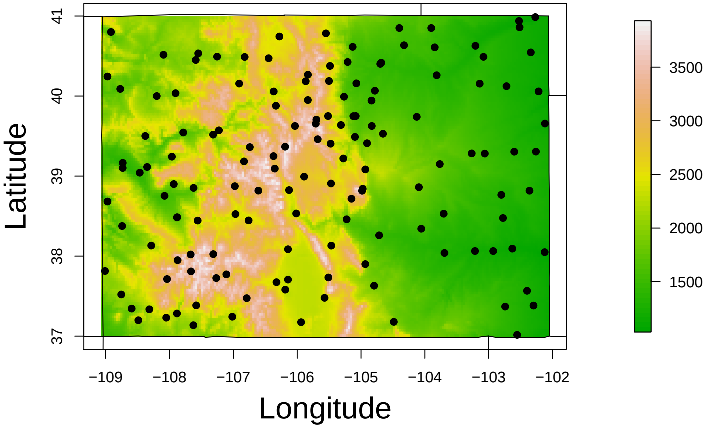
Important in all goals is to quantify the uncertainty.

Outline:

- Extreme(ly nonstationary) processes
- Extreme(ly large) datasets
- Extreme(ly multivariate) processes

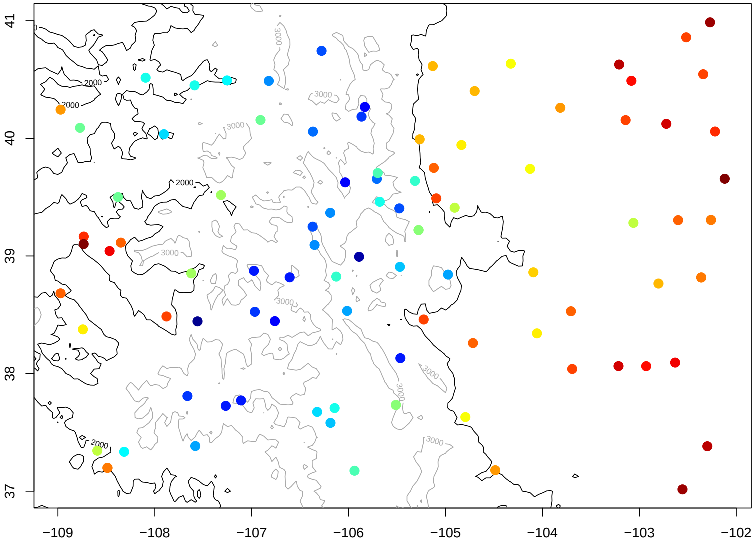
Colorado Data

Data: 145 stations from the Global Historical Climatology Network.
Daily minimum temperature, 1893-2011.



Stevenson Screen and Rain Gauge at Niwot Ridge

Minimum Temperature: June 1, 2010



Notation and Preliminary Ideas

$Z(\mathbf{s})$, indexed by location $\mathbf{s} \in \mathbb{R}^d$, is a Gaussian process if

- For any $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathbb{R}^d$, $(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))^{\mathrm{T}}$ is multivariate normal,

requiring

i) Mean function: $\mathbb{E}\, Z(\mathbf{s}) = \mu(\mathbf{s})$ for all $\mathbf{s} \in \mathbb{R}^d$

ii) Covariance function: $\mathrm{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1, \mathbf{s}_2)$ for all $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^d$.

Why Gaussian? Model is complete with $\mu(\cdot)$ and $C(\cdot, \cdot)$.

Standard Observational Model

Consider an observed process $Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d$,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

where

- $\mu(\mathbf{s})$ fixed mean function
- $Z(\mathbf{s})$ is a mean zero Gaussian process
- $\varepsilon(\mathbf{s})$ is Gaussian white noise ("nugget effect")

Standard Observational Model

Consider an observed process $Y(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d$,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + Z(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

where

- $\mu(\mathbf{s})$ fixed mean function
- $Z(\mathbf{s})$ is a mean zero Gaussian process
- $\varepsilon(\mathbf{s})$ is Gaussian white noise ("nugget effect")

Momentarily use

$$Y(\mathbf{s}) = Z(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

where $\mu(\mathbf{s})$ has already been estimated.

Kriging

Typical goal: Smooth observations $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ to estimate $Z(\mathbf{s}_0)$.

Kriging

Typical goal: Smooth observations $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ to estimate $Z(\mathbf{s}_0)$.
The kriging predictor is

$$\hat{Z}(\mathbf{s}_0) = \frac{1}{n} \sum_{i=1}^{n} w(\mathbf{s}_0, \mathbf{s}_i) Y(\mathbf{s}_i)$$

for weights $w(\mathbf{s}_0, \mathbf{s}_1), \ldots, w(\mathbf{s}_0, \mathbf{s}_n)$ that minimize

$$\mathbb{E}\big(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\big)^2.$$

Kriging

Typical goal: Smooth observations $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$ to estimate $Z(\mathbf{s}_0)$. The kriging predictor is

$$\hat{Z}(\mathbf{s}_0) = \frac{1}{n} \sum_{i=1}^{n} w(\mathbf{s}_0, \mathbf{s}_i) Y(\mathbf{s}_i)$$

for weights $w(\mathbf{s}_0, \mathbf{s}_1), \ldots, w(\mathbf{s}_0, \mathbf{s}_n)$ that minimize

$$\mathbb{E}\big(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\big)^2.$$

If $\mathrm{Cov}\big(Z(\mathbf{s}_1), Z(\mathbf{s}_2)\big) = C(\mathbf{s}_1, \mathbf{s}_2)$ and $\mathrm{Var}\,\varepsilon(\mathbf{s}) = \tau^2$,

$$\hat{Z}(\mathbf{s}_0) = \mathbf{c}^{\mathrm{T}} \left(\Sigma + \tau^2 I\right)^{-1} \mathbf{Y}$$

where $\mathbf{c} = (C(\mathbf{s}_0, \mathbf{s}_i))_i$ and $\Sigma = (C(\mathbf{s}_i, \mathbf{s}_j))_{ij}$.

Kriging Uncertainty

The kriging predictor is

$$\hat{Z}(\mathbf{s}_0) = \mathbf{c}^{\mathrm{T}} \left(\Sigma + \tau^2 I\right)^{-1} \mathbf{Y}$$

with predictive mean squared error

$$\mathbb{E}(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0))^2 = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}^{\mathrm{T}} \left(\Sigma + \tau^2 I\right)^{-1} \mathbf{c}.$$

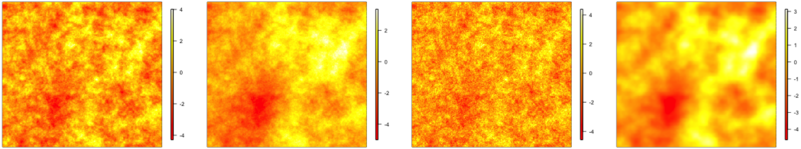
MSE can be approximated via conditional simulations.

Stationarity

A Gaussian process $Z(\mathbf{s})$ is stationary if

- $\mathbb{E}Z(\mathbf{s}) = \mu$ is constant across the domain and
- $\mathrm{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1 - \mathbf{s}_2)$ depends only on the lag between locations.

Isotropic if $C(\mathbf{s}_1 - \mathbf{s}_2) = C(\|\mathbf{s}_1 - \mathbf{s}_2\|)$.

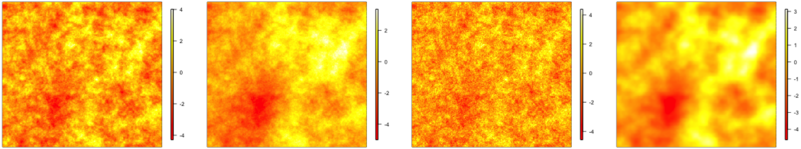


Stationarity

A Gaussian process $Z(\mathbf{s})$ is stationary if

- $\mathbb{E}Z(\mathbf{s}) = \mu$ is constant across the domain and
- $\mathrm{Cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1 - \mathbf{s}_2)$ depends only on the lag between locations.

Isotropic if $C(\mathbf{s}_1 - \mathbf{s}_2) = C(\|\mathbf{s}_1 - \mathbf{s}_2\|)$.

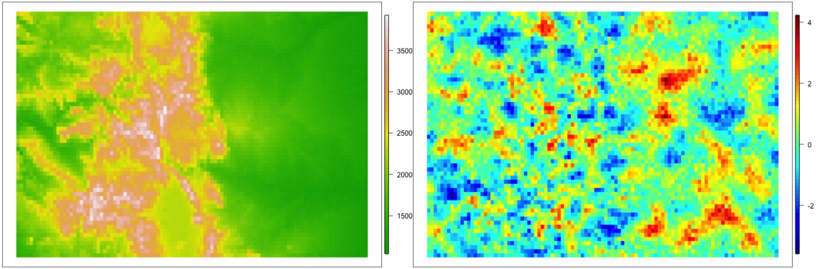


$Z(\mathbf{s})$ is nonstationary if it isn't stationary.

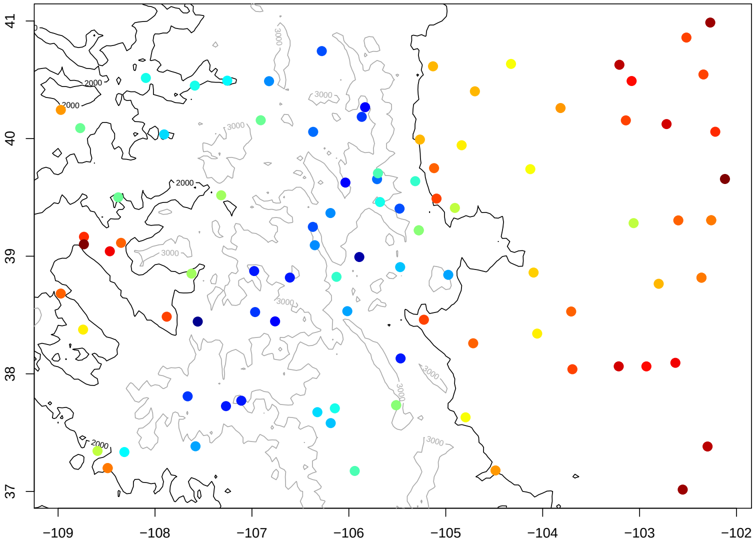
Nonstationary Processes

What might covariance nonstationarity look like?

$$C(\mathbf{s}_1, \mathbf{s}_2) \neq C(\mathbf{s}_1 - \mathbf{s}_2)$$



Minimum Temperature: June 1, 2010



Statistical Model

Model minimum temperature $Y(\mathbf{s}, t)$

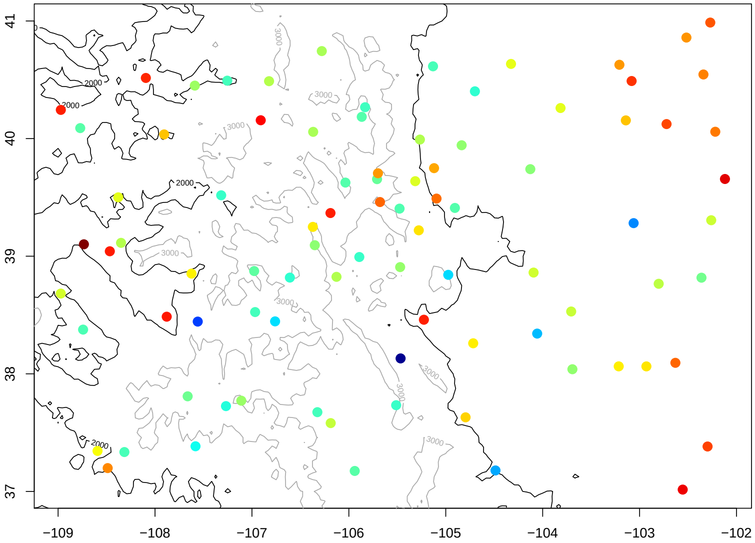
$$
\begin{aligned}
Y(\mathbf{s}, t) &= \boldsymbol{\beta}(\mathbf{s})^{\mathrm{T}} X(\mathbf{s}, t) + Z(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \\
&= \boldsymbol{\beta}(\mathbf{s})^{\mathrm{T}} X(\mathbf{s}, t) + W(\mathbf{s}, t) \\
(\qquad &= \text{Local Climate} + \text{Weather}).
\end{aligned}
$$

$X(\mathbf{s}, t)$ includes seasonal terms and AR(1) behavior.

- Nonstationary mean, estimated locally by least squares
- Is $W(\mathbf{s}, t)$ nonstationary?

(To interpolate local climate, interpolate $\boldsymbol{\beta}(\mathbf{s})$).

Minimum Temperature Residuals: June 1, 2010



How to Model Nonstationarity

- Regularize an empirical covariance matrix (Loader and Switzer 1989; Oehlert 1993)
- Stationary in regions (Haas 1990; Kim et al. 2005)
- Deformation (Sampson and Guttorp 1992)
- Scale mixtures: adaptive spectra (Pintore and Holmes 2007), nonstationary Matérn (Paciorek and Schervish 2006; Stein 2005)
- Process convolution (Higdon 1998; Higdon et al. 1999; Fuentes and Smith 2002)
- Basis-constructed processes (Nychka et al. 2002; Lindgren et al. 2011)

Temperature Example

Temperature model covariance assumptions:

$$\mathrm{Cov}(W(\mathbf{s}, t), W(\mathbf{s}, t+1)) = 0$$

Temperature Example

Temperature model covariance assumptions:

$$\mathrm{Cov}(W(\mathbf{s},t),W(\mathbf{s},t+1))=0$$

$$\mathrm{Cov}(W(\mathbf{s}_1,t),W(\mathbf{s}_2,t))=C(\mathbf{s}_1,\mathbf{s}_2,d(t))+\tau(\mathbf{s}_1,\mathbf{s}_2)^2\mathbb{1}_{[\mathbf{s}_1=\mathbf{s}_2]}$$

Temperature Example

Temperature model covariance assumptions:

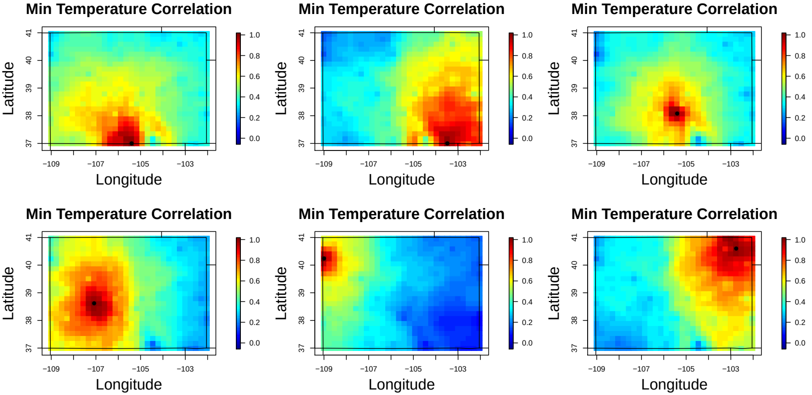
$$\mathrm{Cov}(W(\mathbf{s},t), W(\mathbf{s},t+1)) = 0$$

$$\mathrm{Cov}(W(\mathbf{s}_1,t), W(\mathbf{s}_2,t)) = C(\mathbf{s}_1,\mathbf{s}_2,d(t)) + \tau(\mathbf{s}_1,\mathbf{s}_2)^2 \mathbb{1}_{[\mathbf{s}_1=\mathbf{s}_2]}$$

Estimator for $C(\mathbf{s}_1,\mathbf{s}_2,d(t))$:

$$\frac{\sum_{t=1}^{T}\sum_{k=1}^{n}\sum_{\ell=1}^{n} K_{\lambda_t}\left(\|d(t_0),d(t)\|_d\right) K_\lambda\left(\|\mathbf{s}_1-\mathbf{s}_k\|\right) K_\lambda\left(\|\mathbf{s}_2-\mathbf{s}_\ell\|\right) W(\mathbf{s}_k,t)W(\mathbf{s}_\ell,t)}{\sum_{t=1}^{T}\sum_{k=1}^{n}\sum_{\ell=1}^{n} K_{\lambda_t}(\|d(t_0),d(t)\|_d)K_\lambda(\|\mathbf{s}_1-\mathbf{s}_k\|)K_\lambda(\|\mathbf{s}_2-\mathbf{s}_\ell\|)}$$

Spatial Correlation



Temperature Data

Leave-one-out pseudo-cross-validation comparing kriging under

- Isotropic Matérn model estimated by maximum likelihood
- Nonstationary kernel-smoothed empirical covariances

Temperature Data

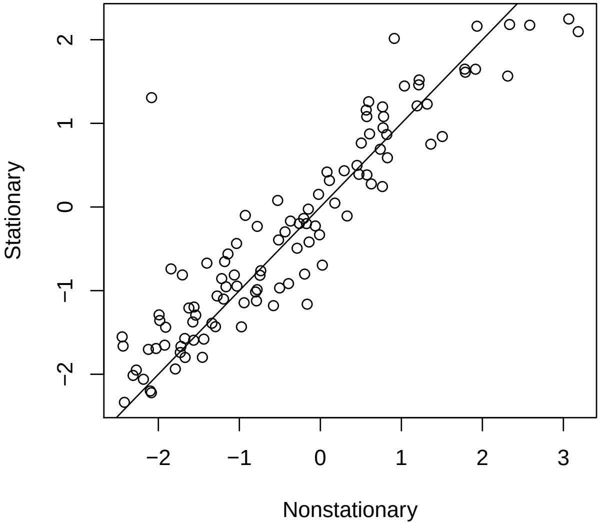
Leave-one-out pseudo-cross-validation comparing kriging under

- Isotropic Matérn model estimated by maximum likelihood
- Nonstationary kernel-smoothed empirical covariances

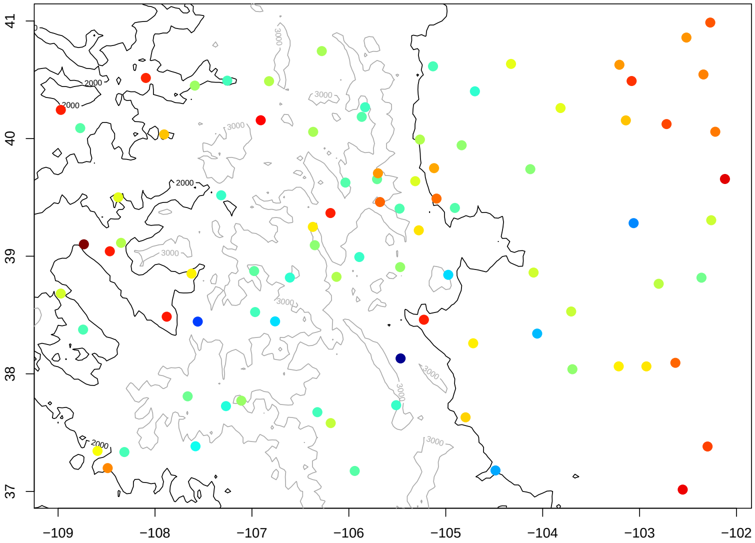
Results:

	RMSE	CRPS
Stationary	1.808	0.983
Nonstationary	1.805	0.983

A Closer Look

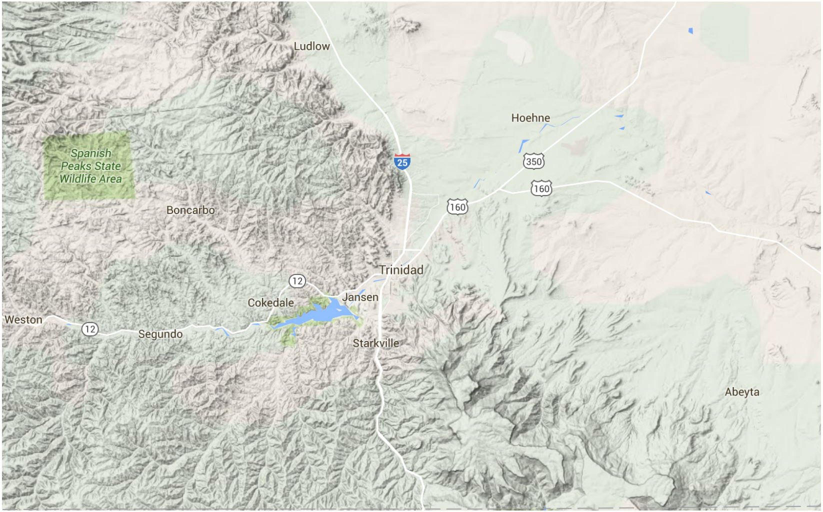


A Closer Look (Trinidad)

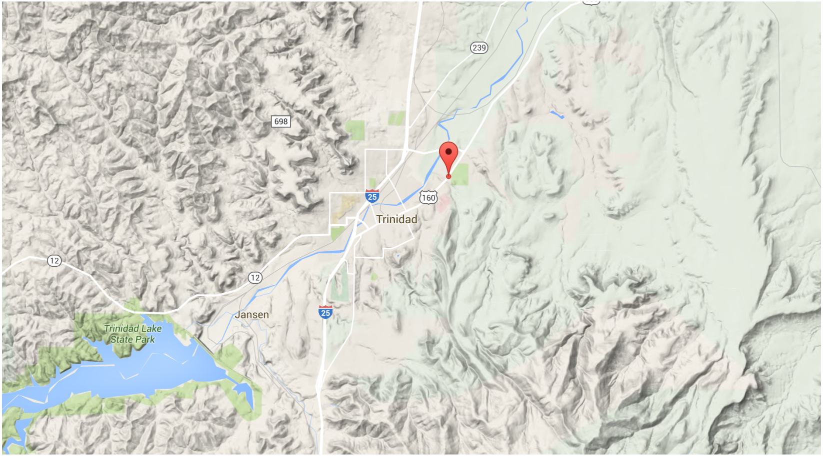


A Closer Look (Trinidad)

A Closer Look (Trinidad)



A Closer Look (Trinidad)



A Closer Look (Trinidad)

Temperature Data Minus Trinidad

Leave-one-out pseudo-cross-validation comparing kriging under

- Isotropic Matérn model estimated by maximum likelihood
- Nonstationary kernel-smoothed empirical covariances

Results:

	RMSE	CRPS	RMSE	CRPS
Stationary	1.808	0.983	1.811	0.984
Nonstationary	1.805	0.983	1.749	0.964

Temperature Data Minus Trinidad

Leave-one-out pseudo-cross-validation comparing kriging under

- Isotropic Matérn model estimated by maximum likelihood
- Nonstationary kernel-smoothed empirical covariances

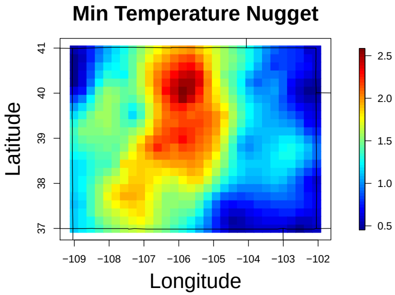
Results:

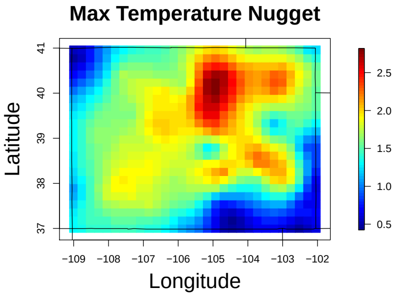
	RMSE	CRPS	RMSE	CRPS
Stationary	1.808	0.983	1.811	0.984
Nonstationary	1.805	0.983	1.749	0.964

A whopping 2-3% improvement.

Nonstationarity: Last Thoughts

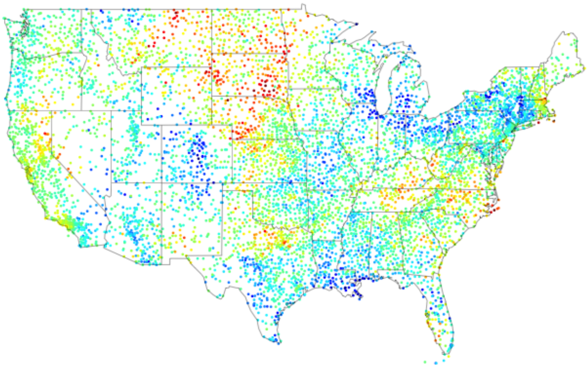
Spatially varying nugget effect seems apparent.





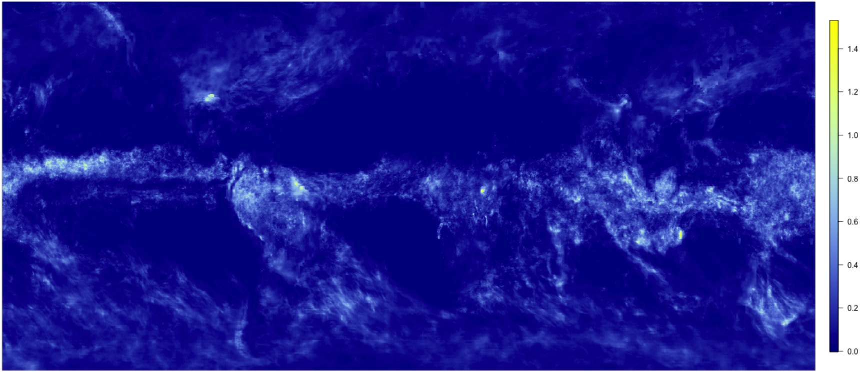
Fuglstad et al. (2014) had a similar experience.

Precipitation anomalies: 7,352 stations



$$\hat{Z}(\mathbf{s}_0) = \mathbf{c}^{\mathrm{T}}(\Sigma + \tau^2 I)^{-1}\mathbf{Y}$$

GPM data: 4,320,000 grid points



Popular approaches

- Fixed rank kriging: low rank representation (Cressie and Johannesson 2008)
- Predictive processes: conditioning leads to a low rank representation (Banerjee et al. 2008)
- Covariance tapering: sparsity via compactly supported covariance (Furrer et al. 2006; Kaufman et al. 2008)
- Full scale approximation: low rank + compactly supported small scale variation (Stein 2008; Sang and Huang 2012)
- Stochastic partial differential equations (Lindgren et al. 2011)
- Multiresolution representations (Nychka et al. 2002; Ferreira and Lee 2007; Nychka et al. 2015; Katzfuss 2016)

Kriging weight function

Recall model

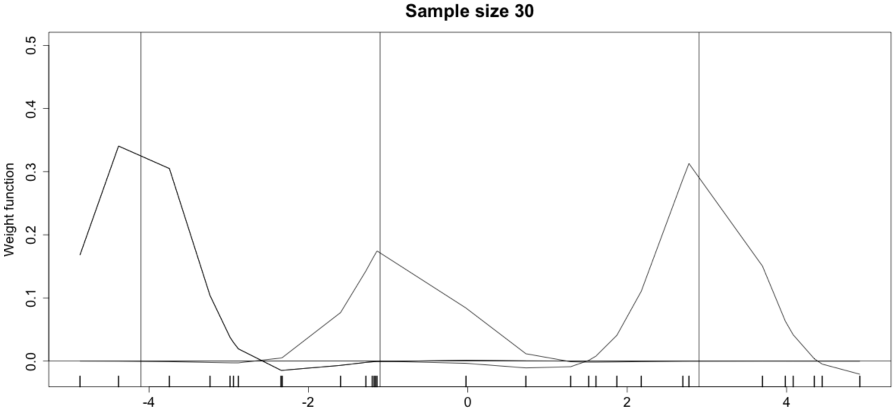
$$Y(\mathbf{s}) = Z(\mathbf{s}) + \varepsilon(\mathbf{s})$$

and the kriging predictor

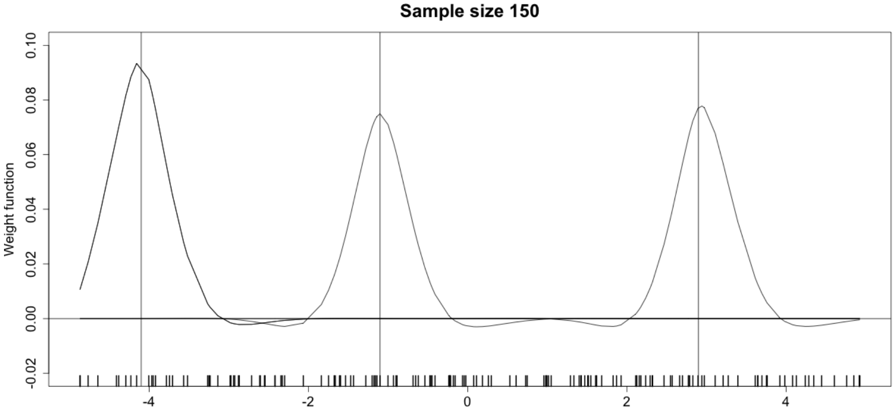
$$\begin{aligned}\hat{Z}(\mathbf{s}_0) &= \mathbf{c}^{\mathrm{T}} \left(\Sigma + \tau^2 I\right)^{-1} \mathbf{Y} \\ &= \frac{1}{n} \sum_{i=1}^{n} w(\mathbf{s}_0, \mathbf{s}_i) Y(\mathbf{s}_i)\end{aligned}$$

How does $w(\cdot, \cdot)$ behave as a function of $\mathbf{s}_1, \ldots, \mathbf{s}_n$?

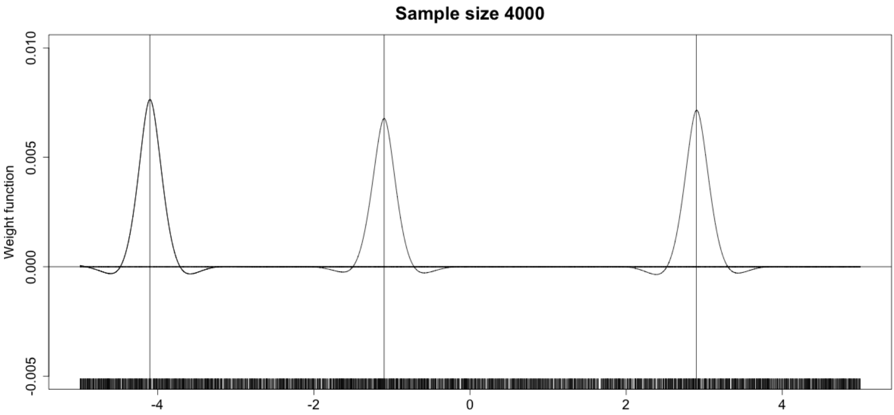
Kriging weight function



Kriging weight function



Kriging weight function



Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

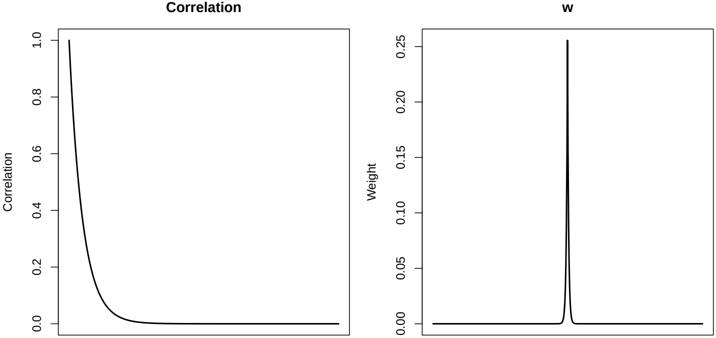
where G is an idealized kernel called the equivalent kernel.

Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

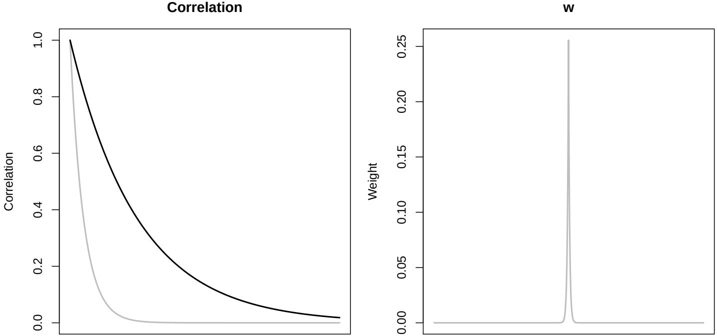


Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

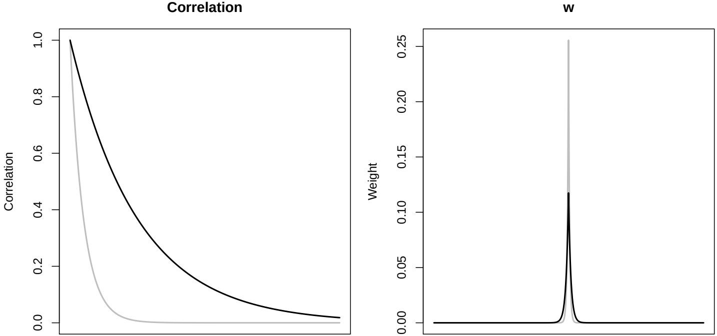


Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

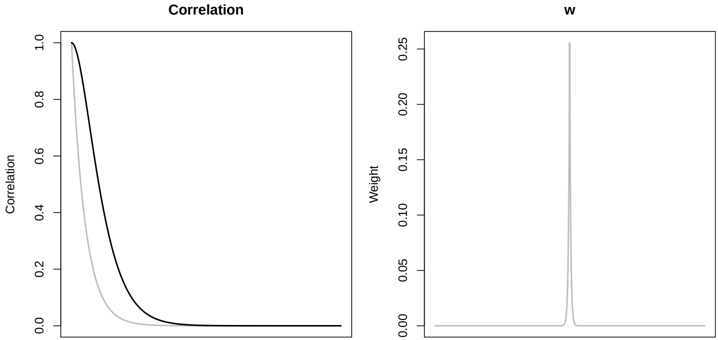


Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

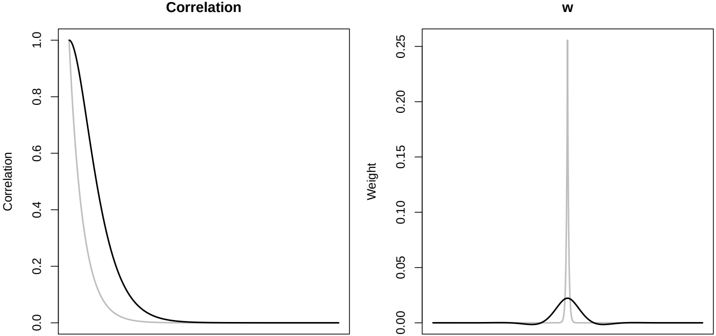


Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

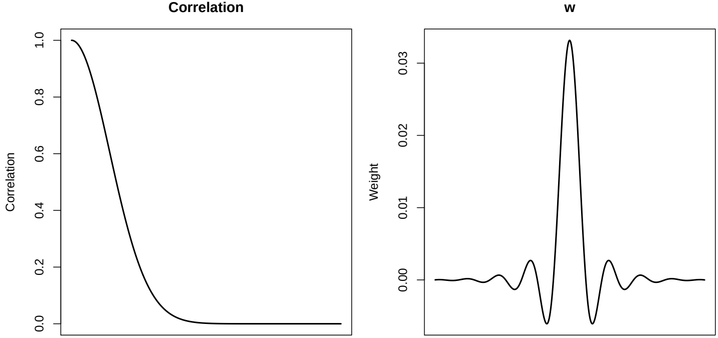


Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.



Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

- For basis representation models G is known analytically
- For "run of the mill" isotropic covariances, G is defined as an integral ($\to$ numerical approximation required).

Approximating w

As $n \to \infty$ it can be shown that

$$w(\mathbf{s}_1, \mathbf{s}_2) \to G(\mathbf{s}_1, \mathbf{s}_2)$$

where G is an idealized kernel called the equivalent kernel.

- For basis representation models G is known analytically
- For "run of the mill" isotropic covariances, G is defined as an integral ($\to$ numerical approximation required).

What if we try

$$\begin{aligned}\hat{Z}_{EK}(\mathbf{s}_0) &= \frac{1}{n}\sum_{i=1}^{n} G(\mathbf{s}_0, \mathbf{s}_i) Y(\mathbf{s}_i) \\ &\approx \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{s}_0, \mathbf{s}_i) Y(\mathbf{s}_i)\end{aligned}$$

(equivalent kriging)?

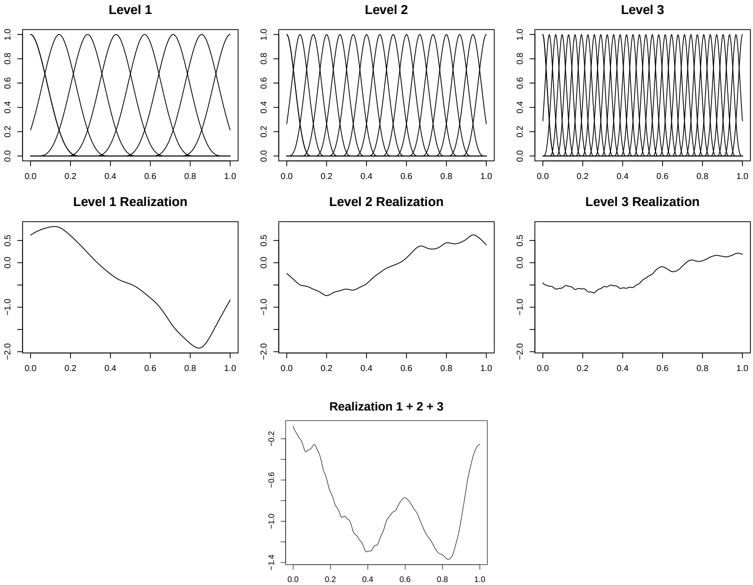
Generic Basis Model

Suppose

$$Z(\mathbf{s}) = \sum_{i=1}^{\infty} c_i \phi_i(\mathbf{s})$$

- c_i are stochastic
- $\phi_i(\mathbf{s})$ are some fixed, useful basis functions

Multiresolution Process



Generic Basis Model Equivalent Kernel

Suppose

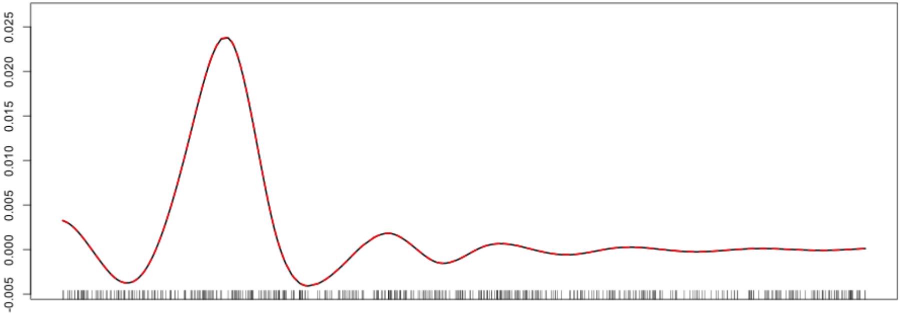
$$Z(\mathbf{s}) = \sum_{i=1}^{\infty} c_i \phi_i(\mathbf{s}),$$

then the equivalent kernel is

$$G(\mathbf{s}_1, \mathbf{s}_2) = \Phi(\mathbf{s}_1)^{\mathrm{T}} (P + \lambda Q)^{-1} \Phi(\mathbf{s}_2)$$

where $\lambda = \tau^2/n$.

Approximation of w (With Corrections)



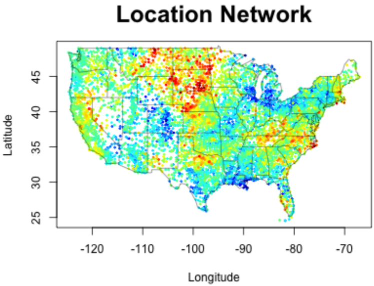
Statistical Models: Timing

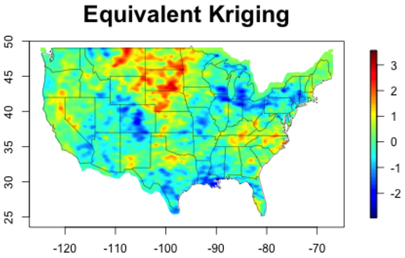
- US data: multiresolution covariance with 52674 basis functions
- GPM data: exponential covariance

Parameters estimated by cross-validation.

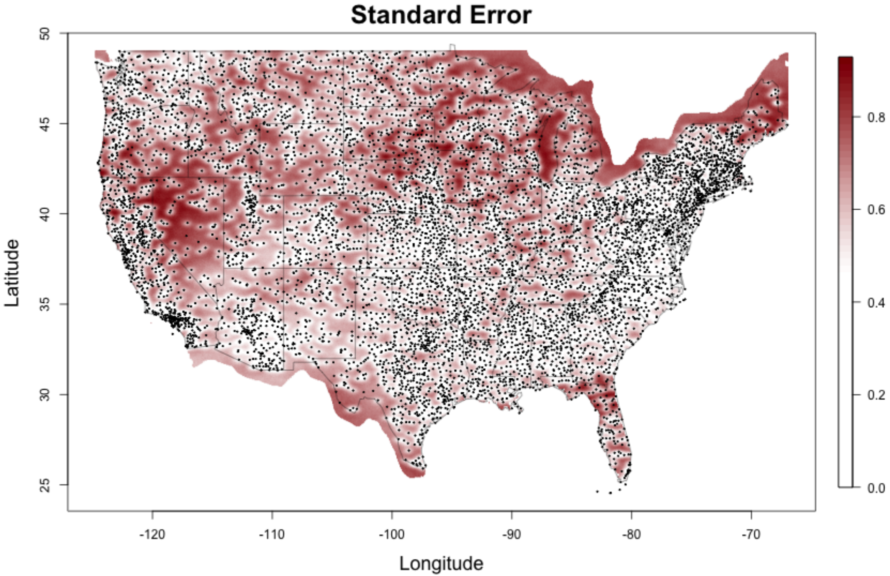
- US data: Kriging to 524888 locations (with remainders): 2.6 seconds
- GPM data: Kriging to 4320000 locations: 81 seconds

Precipitation Results

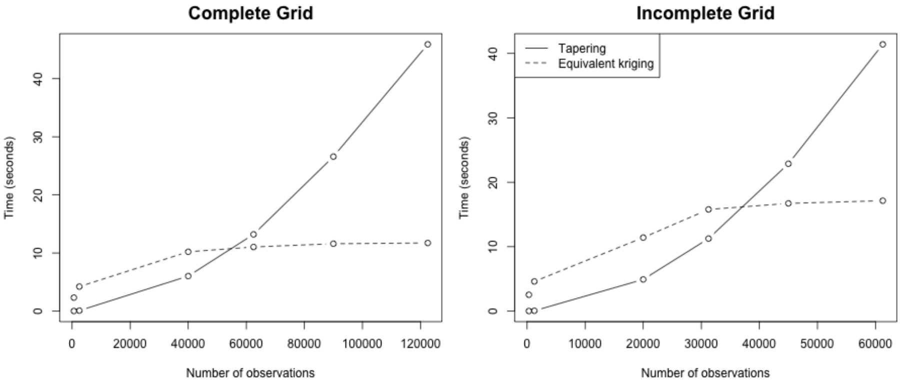




Precipitation Results



Timing Results: Covariance Tapering



NOAA Global Ensemble Forecast System Reforecast

GEFS reforecast project version 2:

- 2012 version of NCEP's GEFS
- 11-member ensemble, daily from 00 UTC initial conditions
- T254 ($\sim$ 50 km) to 8 days, T190 ($\sim$ 70 km) to 16 days

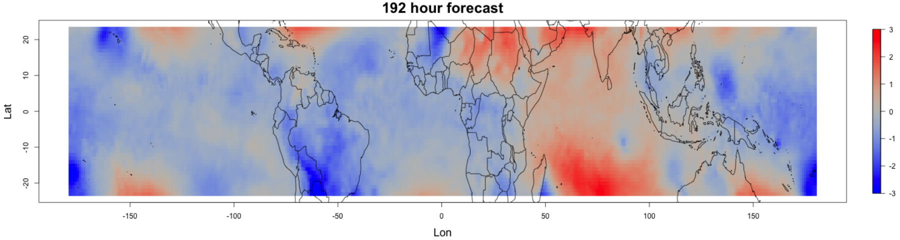
Sea level pressure at forecast horizons:

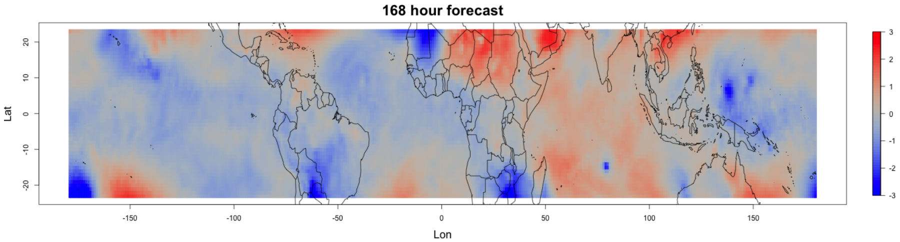
- 0 hours
- 24 hours, 48 hours, . . . , 192 hours (8 days)

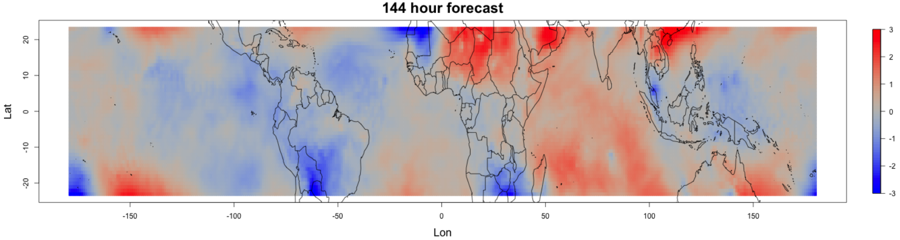
over first 90 days of 2014.

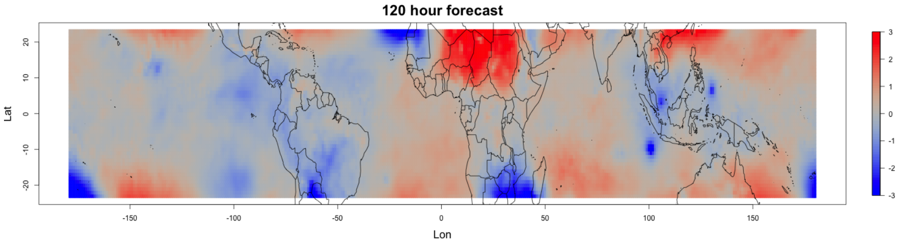
Statistical goal:

- Quantify the improvement and similarity between forecasts and realizing surfaces

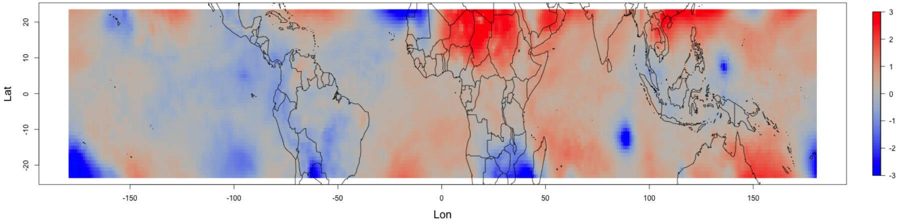

192 hour forecast
Lat
Lon
20
10
0
-10
-20
-150
-100
-50
0
50
100
150
3
2
1
0
-1
-2
-3

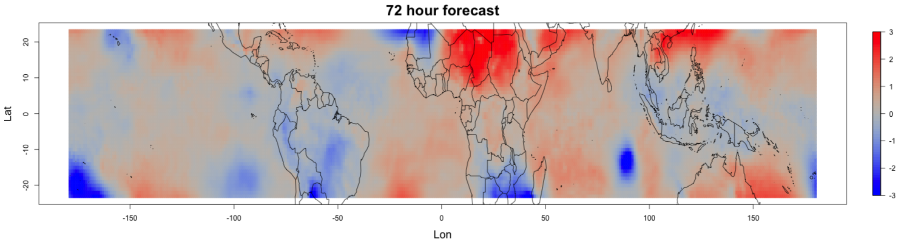

168 hour forecast
Lat
Lon
20
10
0
-10
-20
-150
-100
-50
0
50
100
150
3
2
1
0
-1
-2
-3

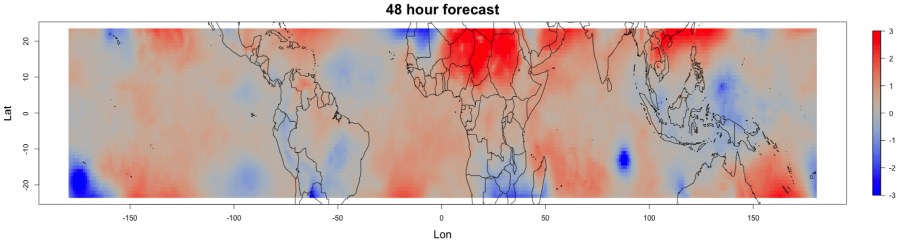

144 hour forecast
Lat
Lon
20
10
0
-10
-20
-150
-100
-50
0
50
100
150
3
2
1
0
-1
-2
-3


120 hour forecast
Lat
Lon
-150
-100
-50
0
50
100
150
20
10
0
-10
-20
3
2
1
0
-1
-2
-3

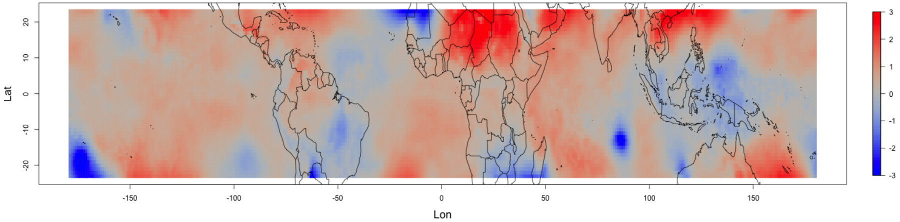
96 hour forecast

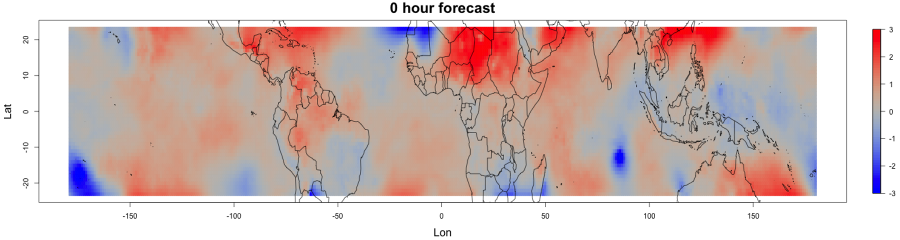



72 hour forecast
Lat
Lon
20
10
0
-10
-20
-150
-100
-50
0
50
100
150
3
2
1
0
-1
-2
-3


48 hour forecast
Lat
Lon
20
10
0
-10
-20
-150
-100
-50
0
50
100
150
3
2
1
0
-1
-2
-3

24 hour forecast





0 hour forecast
Lat
Lon
20
10
0
-10
-20
-150
-100
-50
0
50
100
150
3
2
1
0
-1
-2
-3

Introduction to Multivariate Spatial Modeling

A typical model for p observed spatial processes is

$$\begin{pmatrix} Y_1(\mathbf{s}) \\ Y_2(\mathbf{s}) \\ \vdots \\ Y_p(\mathbf{s}) \end{pmatrix} = \mathbf{Y} = \boldsymbol{\mu} + \mathbf{Z} + \boldsymbol{\varepsilon} = \begin{pmatrix} \mu_1(\mathbf{s}) \\ \mu_2(\mathbf{s}) \\ \vdots \\ \mu_p(\mathbf{s}) \end{pmatrix} + \begin{pmatrix} Z_1(\mathbf{s}) \\ Z_2(\mathbf{s}) \\ \vdots \\ Z_p(\mathbf{s}) \end{pmatrix} + \begin{pmatrix} \varepsilon_1(\mathbf{s}) \\ \varepsilon_2(\mathbf{s}) \\ \vdots \\ \varepsilon_p(\mathbf{s}) \end{pmatrix}$$

where

- $\boldsymbol{\mu}(\mathbf{s})$ is a fixed unknown vector of functions
- $\mathbf{Z}(\mathbf{s})$ is a mean zero p-variate correlated stochastic process
- $\boldsymbol{\varepsilon}(\mathbf{s})$ is a mean zero p-variate white noise process

Cross-Covariance Functions

Dependence is usually specified by choosing

- (Direct)-Covariance functions $C_{ii}(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_i(\mathbf{s}_2))$
- Cross-covariance functions $C_{ij}(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)), i \neq j$.

Cross-Covariance Functions

Dependence is usually specified by choosing

- (Direct)-Covariance functions $C_{ii}(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_i(\mathbf{s}_2))$
- Cross-covariance functions $C_{ij}(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)), i \neq j$.

We require these to be nonnegative definite in that

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{ik} a_{j\ell} C_{ij}(\mathbf{s}_k - \mathbf{s}_\ell) \geq 0.$$

Cross-Covariance Functions

Dependence is usually specified by choosing

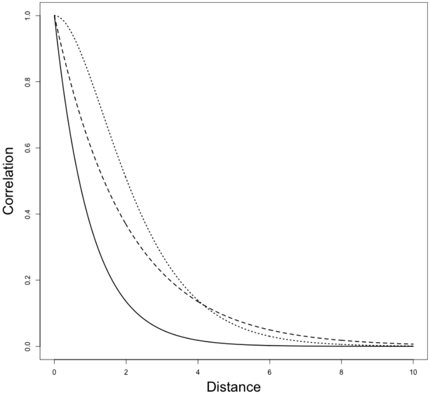
- (Direct)-Covariance functions $C_{ii}(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_i(\mathbf{s}_2))$
- Cross-covariance functions $C_{ij}(\mathbf{s}_1 - \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2)), i \neq j$.

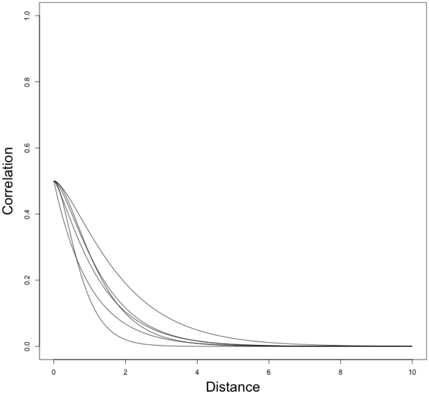
We require these to be nonnegative definite in that

$$\sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{n} \sum_{\ell=1}^{n} a_{ik} a_{j\ell} C_{ij}(\mathbf{s}_k - \mathbf{s}_\ell) \geq 0.$$

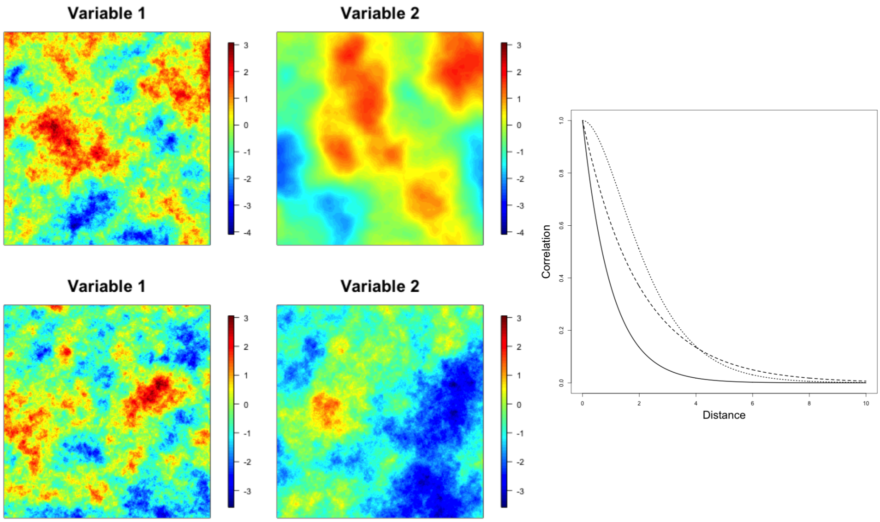
This is a very difficult condition to ensure for some arbitrary proposed model, so most models are constructed to satisfy it.

Correlations vs. Cross-Correlations

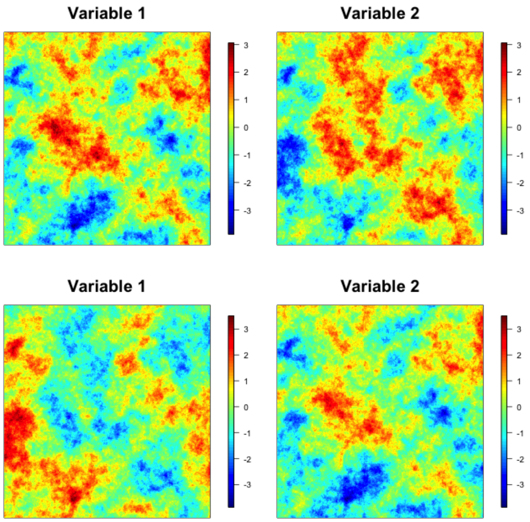


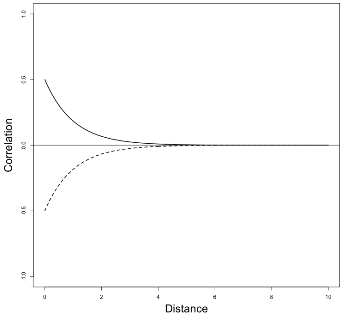


Marginal Range, Smoothness

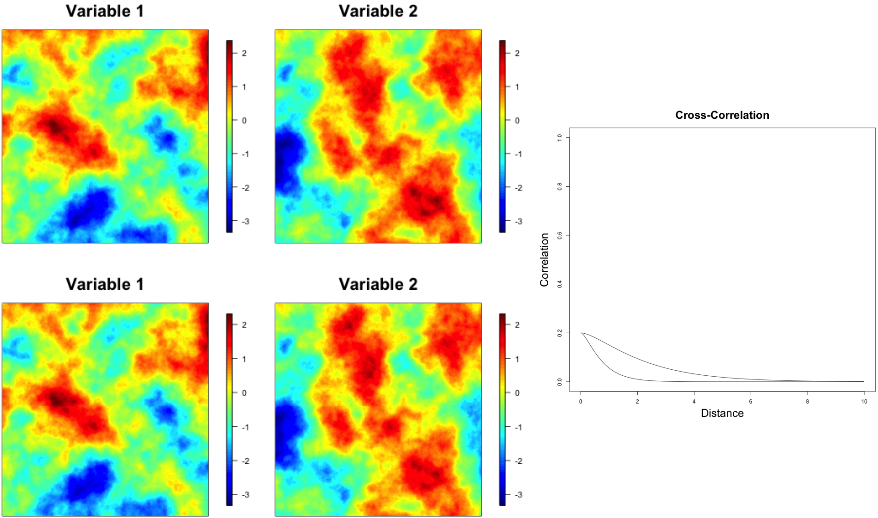


Correlation Coefficient



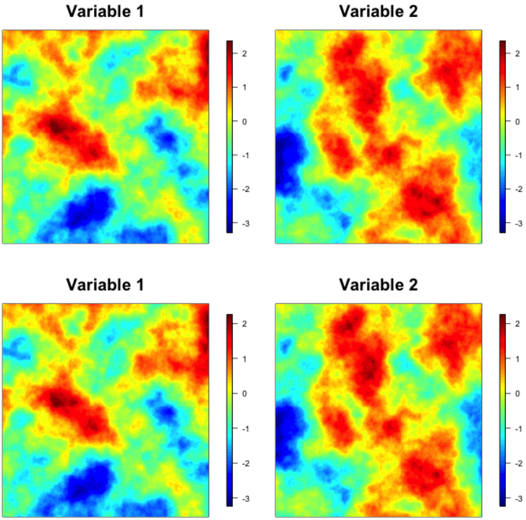


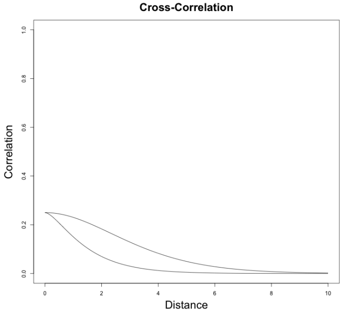
Cross-Range



Variable 1
Variable 2
Variable 1
Variable 2
2
1
0
-1
-2
-3
Cross-Correlation
Correlation
Distance
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10

Cross-Smoothness





Spectra for Multivariate Random Fields

Consider

$$f_{ij}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} C_{ij}(\mathbf{h}) \exp(-i\boldsymbol{\omega}^{\mathrm{T}}\mathbf{h}) \mathrm{d}\mathbf{h}.$$

- $f_{ii}(\boldsymbol{\omega})$ is the spectral density for $C_{ii}(\mathbf{h})$
- $f_{ij}(\boldsymbol{\omega})$ is the cross-spectral density for $C_{ij}(\mathbf{h})$

Spectra for Multivariate Random Fields

Consider

$$f_{ij}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} C_{ij}(\mathbf{h}) \exp(-i\boldsymbol{\omega}^{\mathrm{T}}\mathbf{h}) \mathrm{d}\mathbf{h}.$$

- $f_{ii}(\boldsymbol{\omega})$ is the spectral density for $C_{ii}(\mathbf{h})$
- $f_{ij}(\boldsymbol{\omega})$ is the cross-spectral density for $C_{ij}(\mathbf{h})$

- $f_{ii}(\boldsymbol{\omega})$ is the amount of variability of $Z_i(\mathbf{s})$ that can be attributed to frequency ω.
- What about $f_{ij}(\boldsymbol{\omega})$?

Coherence

Define the coherence at frequency ω between $Z_1(\mathbf{s})$ and $Z_2(\mathbf{s})$ as

$$\gamma(\boldsymbol{\omega}) = \frac{|f_{12}(\boldsymbol{\omega})|}{\sqrt{f_{11}(\boldsymbol{\omega})f_{22}(\boldsymbol{\omega})}} \in [0, 1].$$

Coherence is the amount of variability that can be attributed to a linear relationship between two processes at frequency ω.

Coherence

Define the coherence at frequency ω between $Z_1(\mathbf{s})$ and $Z_2(\mathbf{s})$ as

$$\gamma(\omega) = \frac{|f_{12}(\omega)|}{\sqrt{f_{11}(\omega) f_{22}(\omega)}} \in [0, 1].$$

Coherence is the amount of variability that can be attributed to a linear relationship between two processes at frequency ω.

Moreover, the $K(\mathbf{u})$ that minimizes

$$\mathbb{E} \left| Z_1(\mathbf{s}_0) - \int_{\mathbb{R}^d} K(\mathbf{u} - \mathbf{s}_0) Z_2(\mathbf{u}) \mathrm{d}\mathbf{u} \right|^2$$

is

$$K(\mathbf{u}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \sqrt{\frac{f_{11}(\omega)}{f_{22}(\omega)}} \gamma(\omega) \exp(-i\omega^{\mathrm{T}}\mathbf{u}) \mathrm{d}\omega.$$

Simple Coherence Example

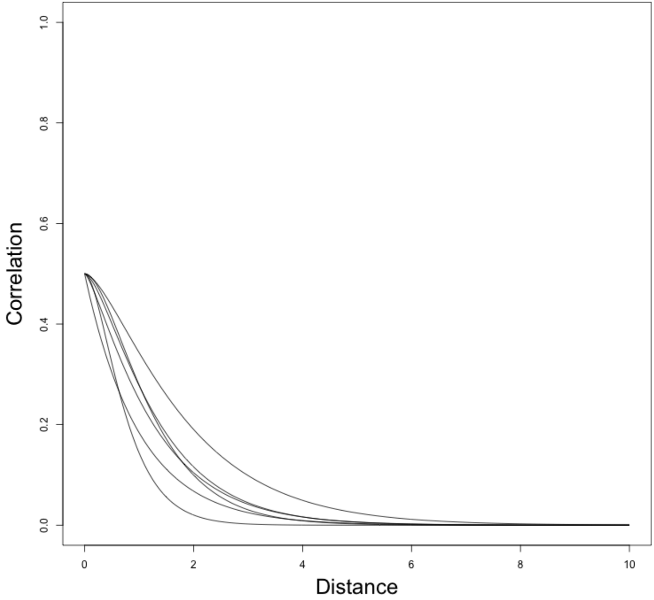
Suppose

$$Z_1(s) = U_1 \cos(\omega_0 s)$$
$$Z_2(s) = U_1 \cos(\omega_0 s) + U_2 \cos(\omega_1 s)$$

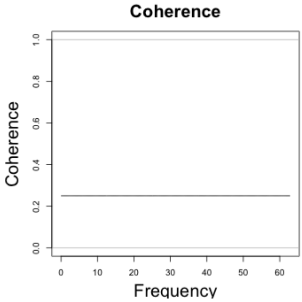
for $\omega_0 \neq \omega_1$ and U_1 and U_2 uncorrelated. Then

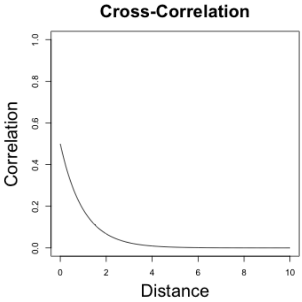
$$\gamma(\omega) = \begin{cases} 1 & \omega = \omega_0 \\ 0 & \text{otherwise} \end{cases}$$

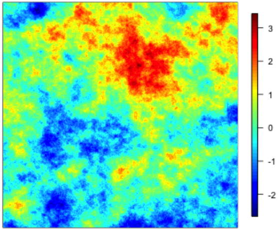
Cross-Correlations

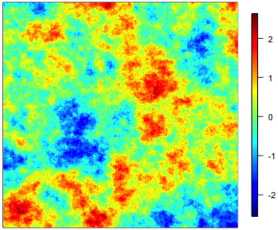


Coherence vs. Cross-Correlation

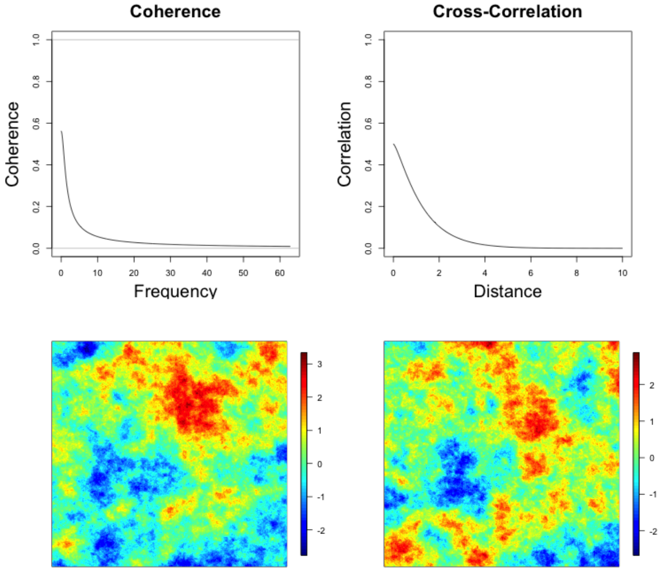




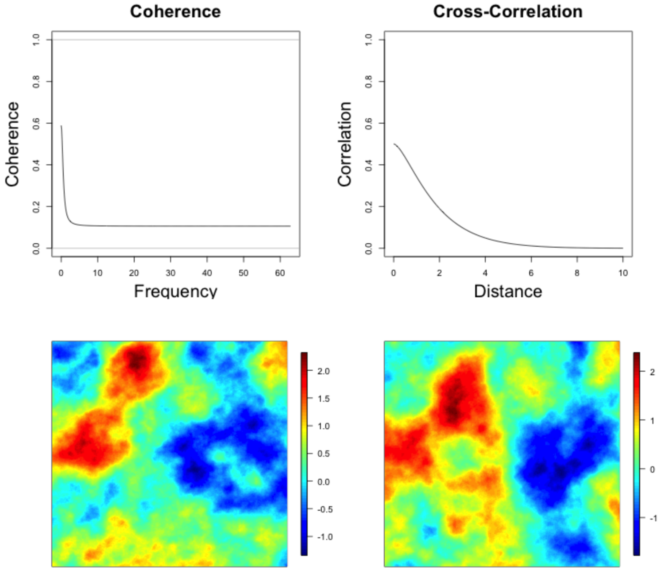




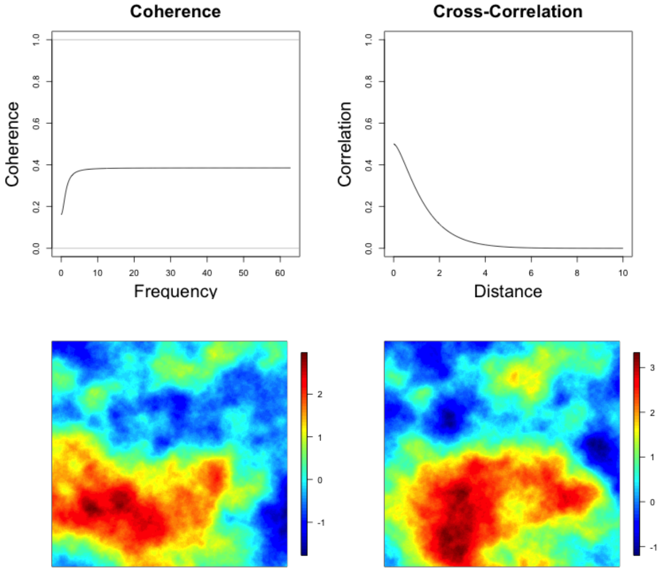
Coherence vs. Cross-Correlation



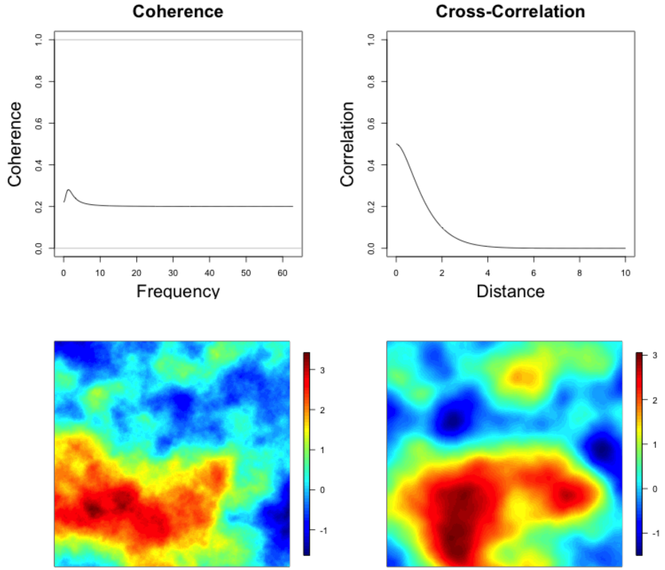
Coherence vs. Cross-Correlation



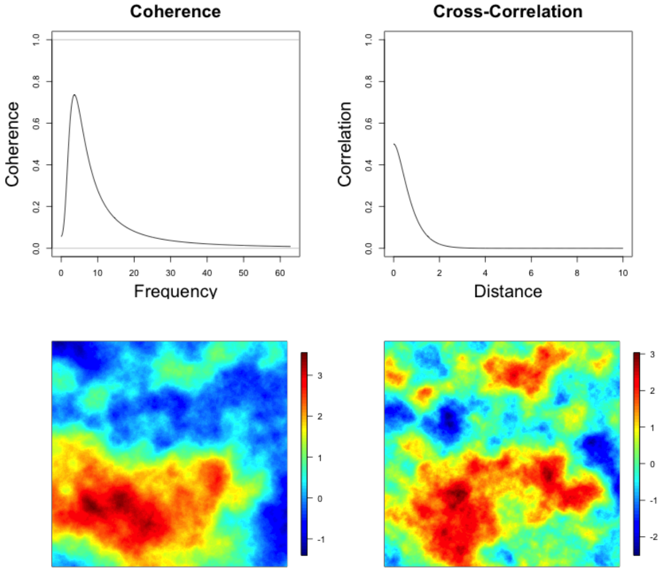
Coherence vs. Cross-Correlation



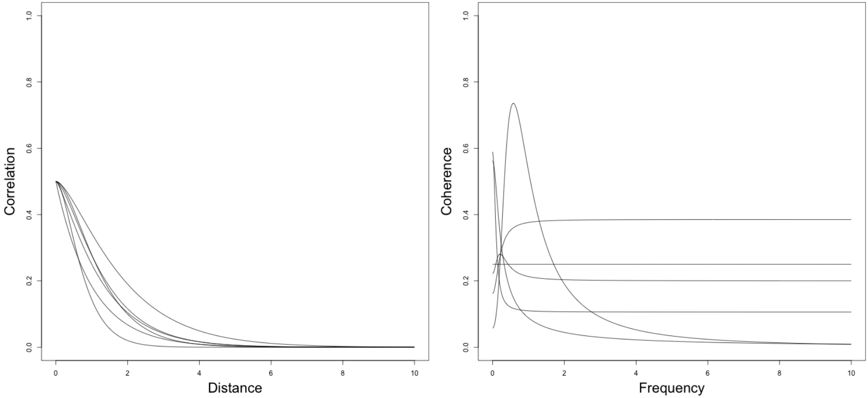
Coherence vs. Cross-Correlation



Coherence vs. Cross-Correlation



Cross-Correlations vs. Coherences



Multivariate Matérn Implications

A bivariate Matérn model has

$$\gamma(\boldsymbol{\omega})^2 = \rho^2 \frac{\Gamma(\nu_{12} + d/2)^2 \Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_1 + d/2)\Gamma(\nu_2 + d/2)\Gamma(\nu_{12})^2} \frac{a_{12}^{4\nu_{12}}}{a_1^{2\nu_1} a_2^{2\nu_2}} \times \frac{(a_1^2 + \|\boldsymbol{\omega}\|^2)^{\nu_1 + d/2}(a_2^2 + \|\boldsymbol{\omega}\|^2)^{\nu_2 + d/2}}{(a_{12}^2 + \|\boldsymbol{\omega}\|^2)^{2\nu_{12} + d}}.$$

Multivariate Matérn Implications

A bivariate Matérn model has

$$\gamma(\boldsymbol{\omega})^2 = \rho^2 \frac{\Gamma(\nu_{12}+d/2)^2\Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_1+d/2)\Gamma(\nu_2+d/2)\Gamma(\nu_{12})^2} \frac{a_{12}^{4\nu_{12}}}{a_1^{2\nu_1}a_2^{2\nu_2}} \times \frac{(a_1^2+\|\boldsymbol{\omega}\|^2)^{\nu_1+d/2}(a_2^2+\|\boldsymbol{\omega}\|^2)^{\nu_2+d/2}}{(a_{12}^2+\|\boldsymbol{\omega}\|^2)^{2\nu_{12}+d}}.$$

Results:

- Force $\nu_{12} > (\nu_1+\nu_2)/2$, else coherence does not decay at arbitrarily high frequencies
- a_{12} controls location of peak of coherence

Estimation: Periodogram

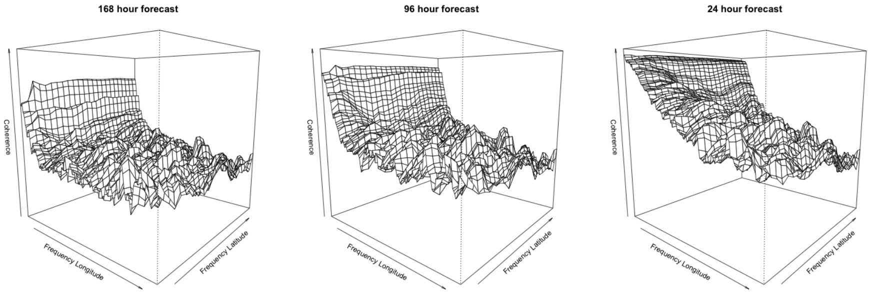
The spatial periodogram matrix is $\mathbf{I}(\omega) = (I_{k\ell}(\omega))_{k,\ell=1}^{p}$ where

$$I_{k\ell}(\omega) = \frac{\delta}{(2\pi)^p N} \left(\sum_{k=1}^{N} Z_k(\mathbf{s}_k) \exp(-i\mathbf{s}_k^{\mathrm{T}}\omega) \right) \overline{\left(\sum_{k=1}^{N} Z_\ell(\mathbf{s}_k) \exp(-i\mathbf{s}_k^{\mathrm{T}}\omega) \right)}$$

and is available at Fourier frequencies.

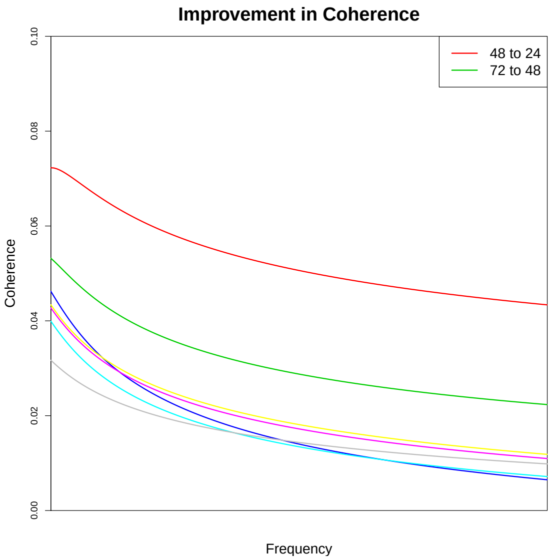
- Need to smooth periodograms for consistency
- GEFS example: average empirical coherences over 90 days in dataset

GEFS SLP Coherences



Estimated absolute coherence functions for the GEFS pressure data between (a) 0h and 168h (7 days), (b) 0h and 96h (4 days) and (c) 0h and 24h (1 day).

GEFS Pressure Example



Discussion

- Nonstationarity: what is the goal?
- Estimation for large datasets: which scales do we care about?
- Multivariate processes: what are we modeling?

Unfair reference list:

Kleiber, W., Katz, R.W. and Rajagopalan, B. (2013). “Daily Minimum and Maximum Temperature Simulation over Complex Terrain”, *Annals of Applied Statistics*, **7**, 588–612,

Kleiber, W. and Nychka, D. (2015). “Equivalent Kriging”, *Spatial Statistics*, **12**, 31–49.

Kleiber, W. (2016). “Coherence for Multivariate Random Fields”, *Statistica Sinica*, minor revision.