Uncertainty in the pattern scaling of climate models

Douglas Nychka, National Center for Atmospheric Research

National Science Foundation November 2017

Summary

- Pattern scaling of a climate model ensemble (Climate Science)
- Nonstationary Gaussian fields (Applied Stats)
- Spatial autoregressions, multiscale processes (Big Data)
- Results for CESM ensemble
- Data analysis on super computers (Big R)

Challenges:

Building covariance models for large problems, simulating Gaussian fields, and actually computing the beasts!

PART 1

Climate model emulation

Context

What will the climate be in 60 years?

- Need a scenario of future human acitivities.

The representative concentration pathway (RCP) is a synthesis that specifies how greenhouse gases change over time.

- Need a geophysical model to relate the RCP to possible changes in climate.

Earth system models are complex, physical models that integrate the feedbacks in the atmosphere, ocean, land, ice and other components to determine the climate under different conditions.

Community Earth System Model (CESM)

A family of models developed at NCAR and supported by the National Science Foundation.

From global to local

Main benefit of a climate model is in quantifying the local effects of climate change.

CESM Large Ensemble (CESM-LE)

A 30+ member ensemble of CESM simulations that have been designed to study the local effects of climate change
– and the uncertainty due to the natural variability in the earth system.

- $\approx 1^{\circ}$ spatial resolution – about 55K locations
- Simulation period 1920 - 2080
- Using RCP 8.5 after 2005

Classic pattern scaling.

The assumption:

Patterns of temperature change over space are linear functions of the change in global mean temperature.

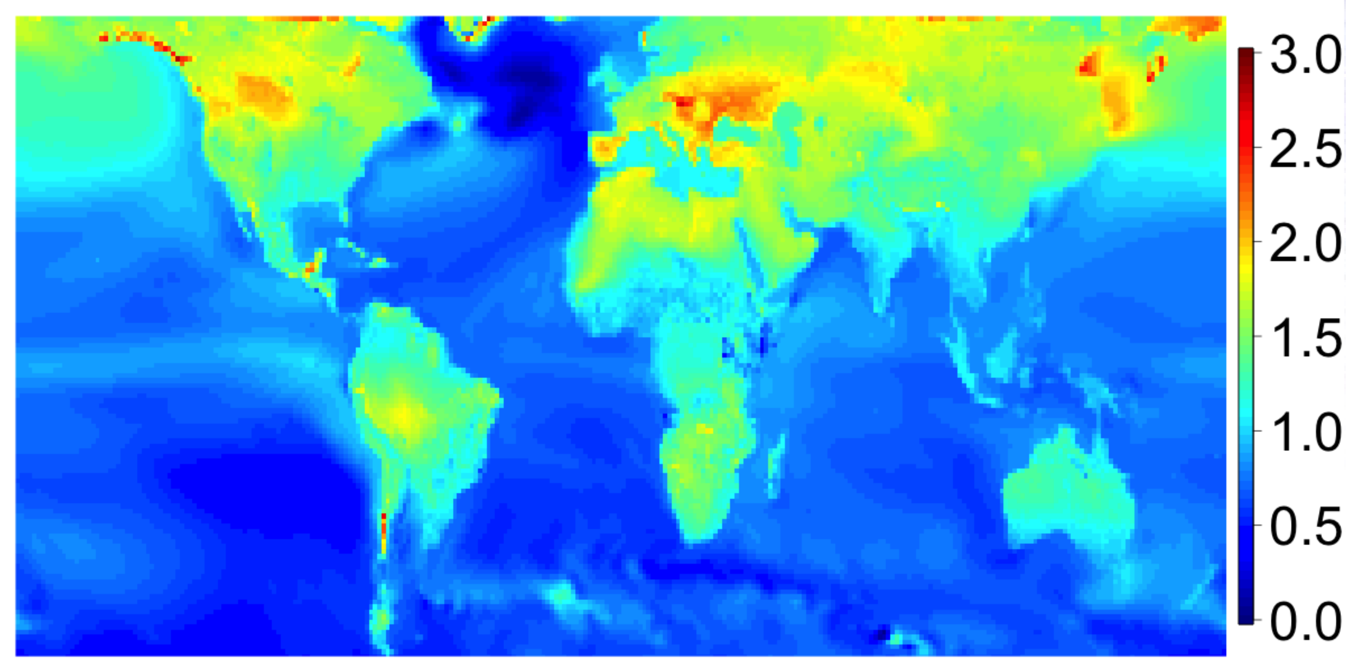
T_t Temperature at time t at a specific location (or climate model grid box)

g_t Global mean temperature at time t.

$$(T_t - T_0) \approx P(g_t - g_0)$$

P is a slope (amplitude) that relates a change in global temperature to one locally.

Mean slopes across 30 members for JJA



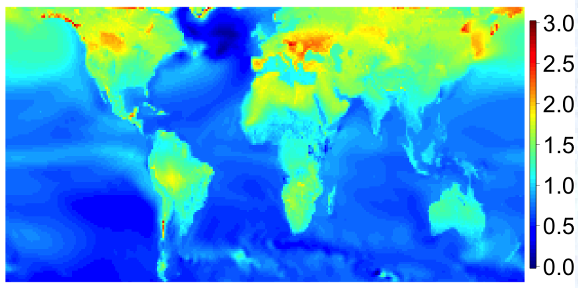
E. g. value of 2.5 means: a 1° global increase implies 2.5° increase locally.

This allows us to determine the local mean temperature change based on a simpler model for the global average temperature

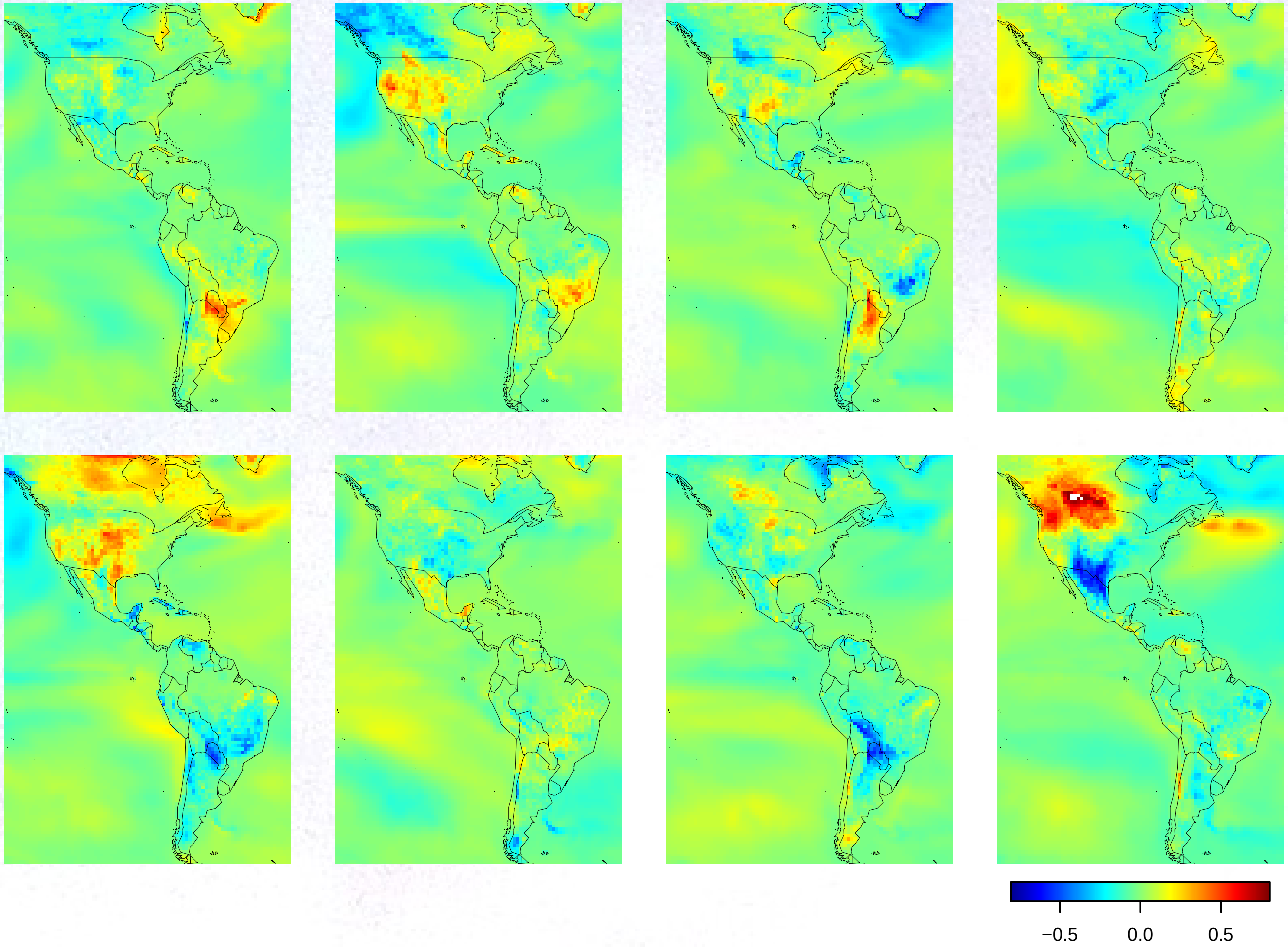
PART 2
A spatial data problem

Individual patterns

Ensemble mean



First 8 *centered* ensemble members



. . . there are 22 more of these!

Goal:

Be able to simulate additional fields efficiently that match the spatial dependence in this 30 member ensemble.

Result:

- Use simple models for other RCP scenarios to generate a large number of ensembles and without the need to do full CESM runs.

- Enable researchers using climate change impact models to drive their models with a larger number of ensembles and without the need to do full CESM runs.

Gaussian process (GP) models

$f(\boldsymbol{x})$ value of the field at location $\boldsymbol{x}$.

$E[f(\boldsymbol{x})] = 0$ and $k(\boldsymbol{x}_1, \boldsymbol{x}_2) = E[f(\boldsymbol{x}_1) f(\boldsymbol{x}_2)]$

- $f(\boldsymbol{x})$ is a GP if any finite collection of $\{f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N)\}$ has a multivariate normal distribution.

- f is mean square continuous (differentiable) if k is continuous (differentiable) in both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

- If the process is stationary and isotropic then

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \varphi(||\boldsymbol{x}_1 - \boldsymbol{x}_2||).$$

– a strong assumption but identifies the direct connection with Radial Basis Functions methods

Estimating covariance parameters by maximum likelihood.

Covariance now depends on some parameters, k_θ.

Log Likelihood:

$$\ell(\boldsymbol{f},\theta) = (1/2)\boldsymbol{f}^T(K_\theta^{-1})\boldsymbol{f} - (1/2)ln|K_\theta| + \text{constants}$$

- $\widehat{\theta}$ = maximizer of $\ell(\boldsymbol{f},\theta)$ – aka the MLE
- Inference (e.g. confidence intervals) based on the curvature of the log likelihood surface at $\widehat{\theta}$

To simulate a GP at locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$

- Form $K_{\theta,i,j} = k_\theta(\boldsymbol{x}_i, \boldsymbol{x}_j)$
- $\boldsymbol{f} = K_\theta^{1/2}\boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0,1)$

The Matérn covariance

$k(\boldsymbol{x}_1, \boldsymbol{x}_2) = \rho \mathcal{C} d^{\nu} \mathcal{K}_{\nu}(d)$, and $d = ||\boldsymbol{x}_1 - \boldsymbol{x}_2||/\alpha$

- $\mathcal{K}_{\nu}$ a modified Bessel function.
- $\mathcal{C}$ a normalizing constant depending on ν.
- Smoothness ν measures number of mean square derivatives and is equivalent to the polynomial tail behavior of the spectral density.
- When $\nu = .5$, Matérn is an exponential covariance, $\nu = \infty$, a Gaussian.

Challenges with large spatial data

- K a dense covariance matrix that has the dimensions of the number of locations. (e.g. 55K for CESM output)

- For statistical computations it is not feasible to evaluate $K^{1/2}$, K^{-1} and $|K|$ when the data set is large.

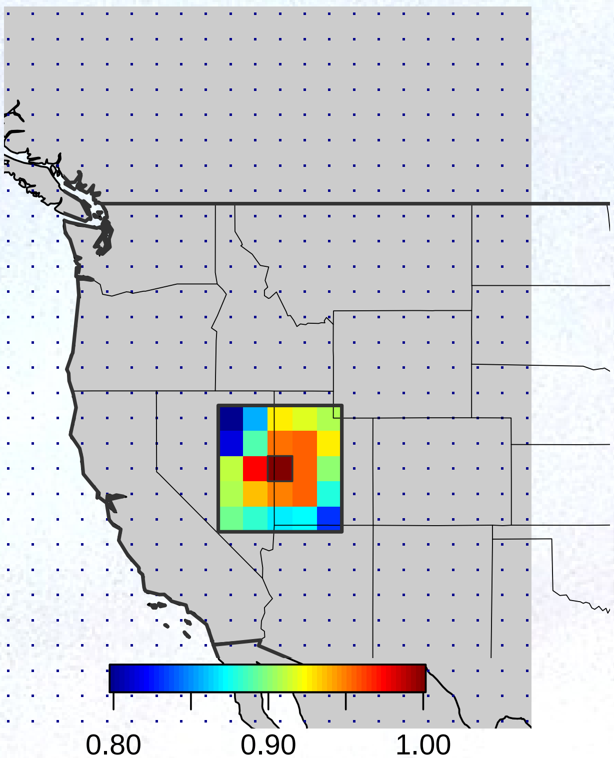
- Typically the GP is not stationary – so a heroic global computation may not solve our problem!

Solution

Fit local GP models to a moving window of the data.

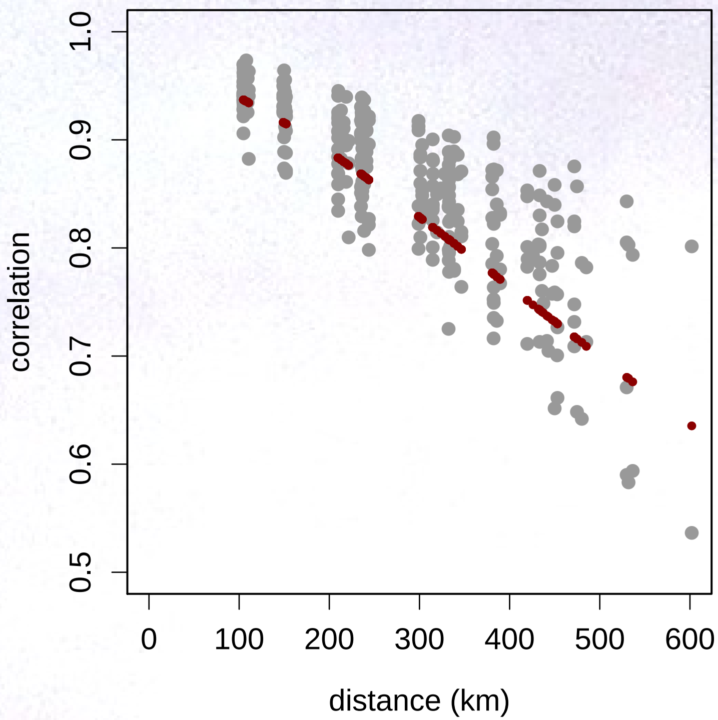
A local spatial model

Correlations of center grid point.



A 5×5 grid of output,
25×30 data points.

Matérn fit with MLEs and $\nu = 1.0$



Data model:

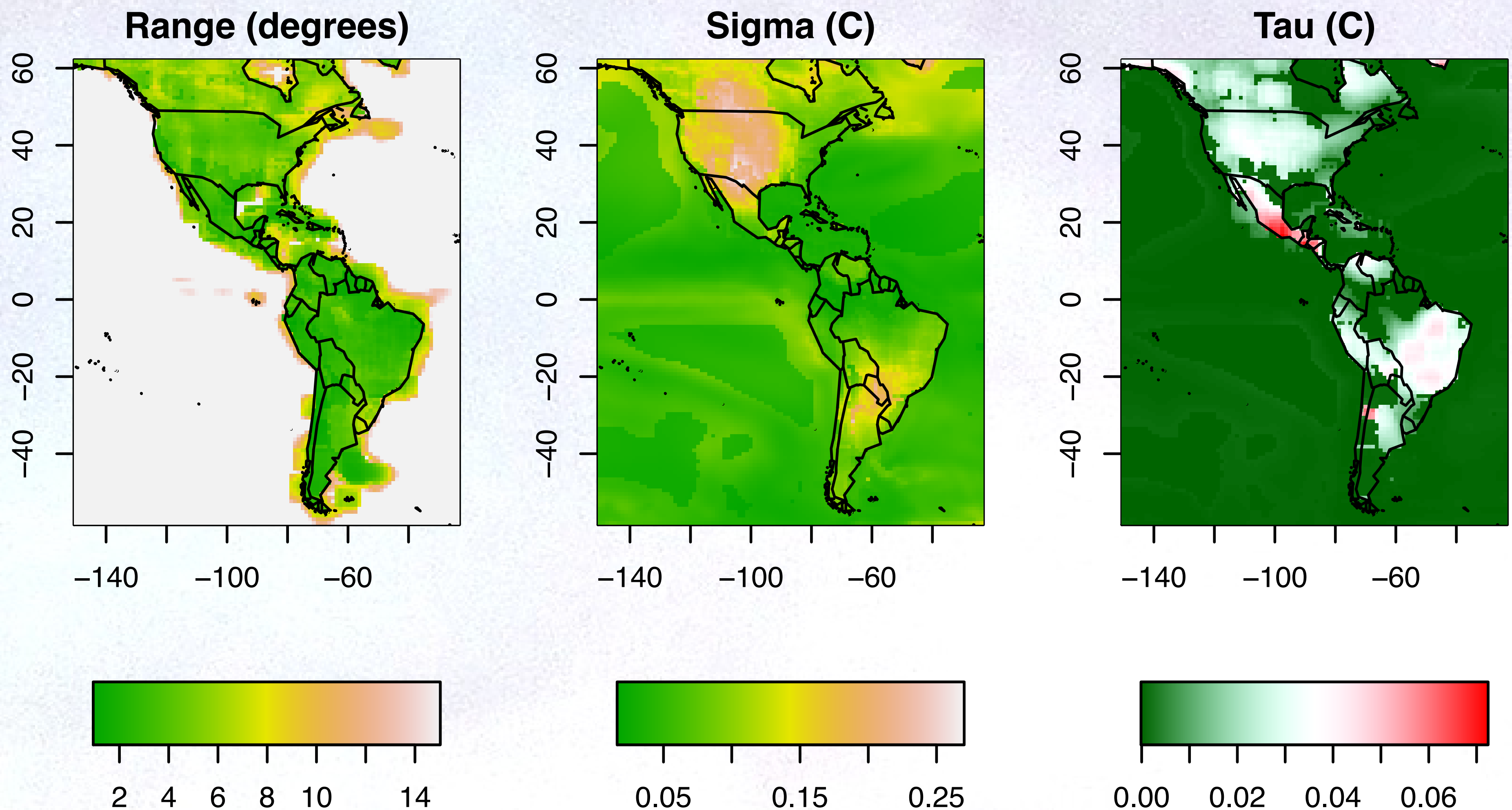
$P(\boldsymbol{x}_i) = f(\boldsymbol{x}_i) + \epsilon_i$

f – Matérn GP

ϵ_i – uncorrelated noise

Fitting all grid boxes

Matérn MLEs



- $11\times$ 11 windows using coordinates in degrees
- About 13K grid boxes in this subregion

More on nonstationary covariances

- Convolution model (Higdon, Fuentes)

with exponential kernels;

$$k_\theta(\boldsymbol{x}_1, \boldsymbol{x}_2) \sim \int e^{-||\boldsymbol{x}_1 - \boldsymbol{u}||/\theta(\boldsymbol{x}_1)} e^{-||\boldsymbol{u} - \boldsymbol{x}_2||/\theta(\boldsymbol{x}_2)} d\boldsymbol{u}$$

If $\theta(\boldsymbol{x}) \equiv \theta$ in 2-d this gives a Matern with smoothness $\nu = 1.0$

- Scale mixture (Paciorek, Stein)

$\nu = 1.0$

$$k(\boldsymbol{x}_1, \boldsymbol{x}_2) \sim d\mathcal{K}_1(d),$$

where

$$d = \frac{||\boldsymbol{x}_1 - \boldsymbol{x}_2||}{\sqrt{\theta(\boldsymbol{x}_1)^2 + \theta(\boldsymbol{x}_2)^2}}$$

These are different models.

Conjecture: as $\theta(\boldsymbol{x}_2) \to 0$ give different smoothness at $\boldsymbol{x}_2$

Some drawbacks

- Approximate simulation does not address long range dependence beyond the local neighborhoods.

- This is not a global model and the covariance is implicit.

- Simulation is slow because each grid point requires a separate eigen-decomposition.

These points motivate a different approach to representing the GP

PART 3:

Markov random fields, multi-scale processes

A Spatial Autoregression (SAR)

Gridded field:

$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & c_1 & . & . \\ . & c_2 & c_* & c_3 & . \\ . & . & c_4 & . & . \\ . & . & . & . & . \end{array}$$

SAR weights:

$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & \text{-1} & . & . \\ . & \text{-1} & a & -1 & . \\ . & . & \text{-1} & . & . \\ . & . & . & . & . \end{array}$$

The filter:

$ac_* - (c_1 + c_2 + c_3 + c_4) =$ white noise

- a needs to be greater than 4.
- $B\boldsymbol{c} =$ i.i.d.$N(0,1)$ where B is a sparse matrix
- Covariance for $\boldsymbol{c}$ is $(B^T B)^{-1} = Q^{-1}$

Q known as the precision or information matrix.

Generalization ...

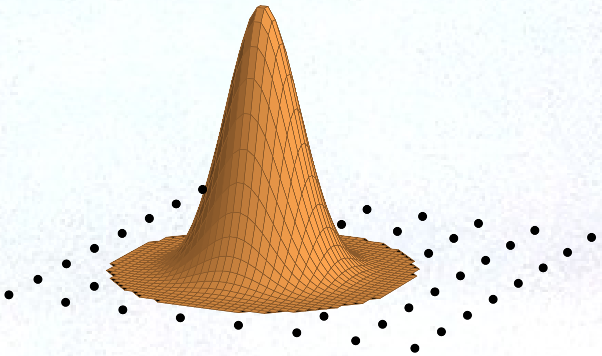
Representing a random surface:

$$g(x) = \sum_j \phi_j(x) c_j$$

- $\boldsymbol{c}$ is the random field from the SAR.
- $\{\phi_j(x)\}$ are radial basis functions :

$$\phi_j(x) = \psi(||\boldsymbol{x} - \boldsymbol{u}_j||/\delta)$$

A member of the Wendland basis functions



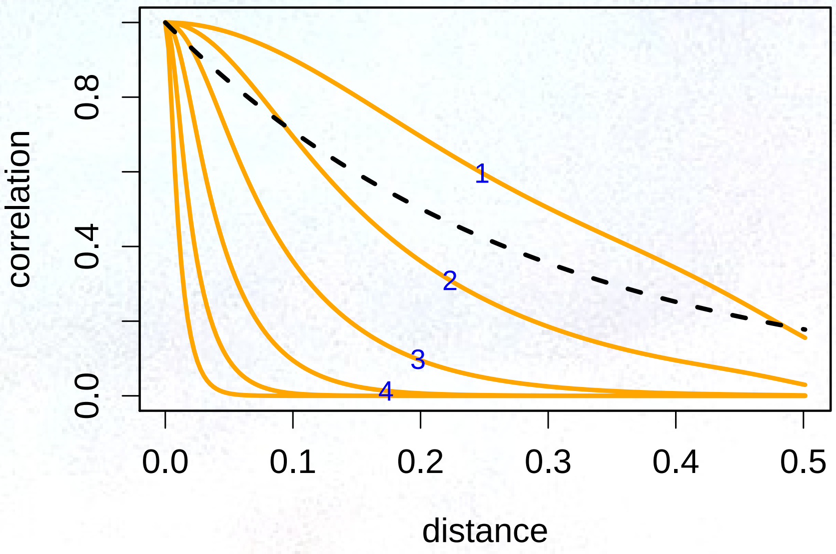
Multi-resolution extension

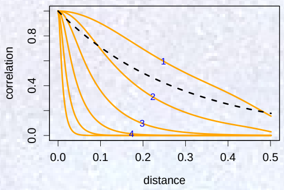
Add together different resolution levels to approximate the Matérn GP.

$$g(x) = \sum_j \phi_{j1}(x)c_{j1} + \sum_j \phi_{j2}(x)c_{j2} + \ldots + \sum_j \phi_{jL}(x)c_{jL}$$

$$g(x) = g_1(x) + g_2(x) + ... + g_L(x)$$

Correlation functions for $g_1, \ldots, g_6$ and a target exponential

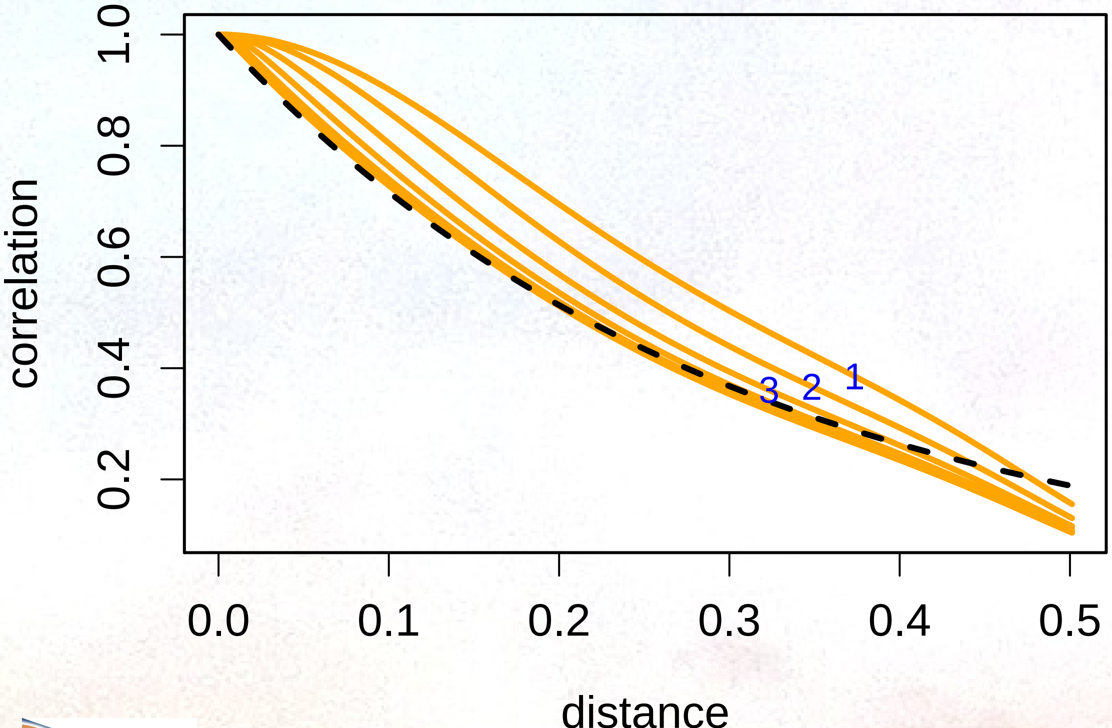




Weighting by asymptotic theory for exponential

$VAR(g_{\text{level}}) = 2^{-\text{level}/2}$
that is $\approx$ 0.707, 0.50, 0.354, 0.25, 0.177, 0.125 .

Correlation functions adding levels and the target exponential



Computing

To simulate the GP

Basis function matrix: $\Phi_{i,j} = \phi_j(\boldsymbol{x}_i)$

- Sparse solve of $V^T \boldsymbol{c} = \boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0,1)$

V the Choleski factor of Q, i.e. $Q = V^T V$

- $\boldsymbol{f} = \Phi \boldsymbol{c}$.
- This a global simulation for a well-defined GP.
- Includes possible long-range correlations.
- At least two orders of magnitude faster than the local approach.

PART 4:

Nonstationary Analysis

A global model:

Can encode all rows of B from local fitting giving the sparse precision matrix $Q = BB^T$

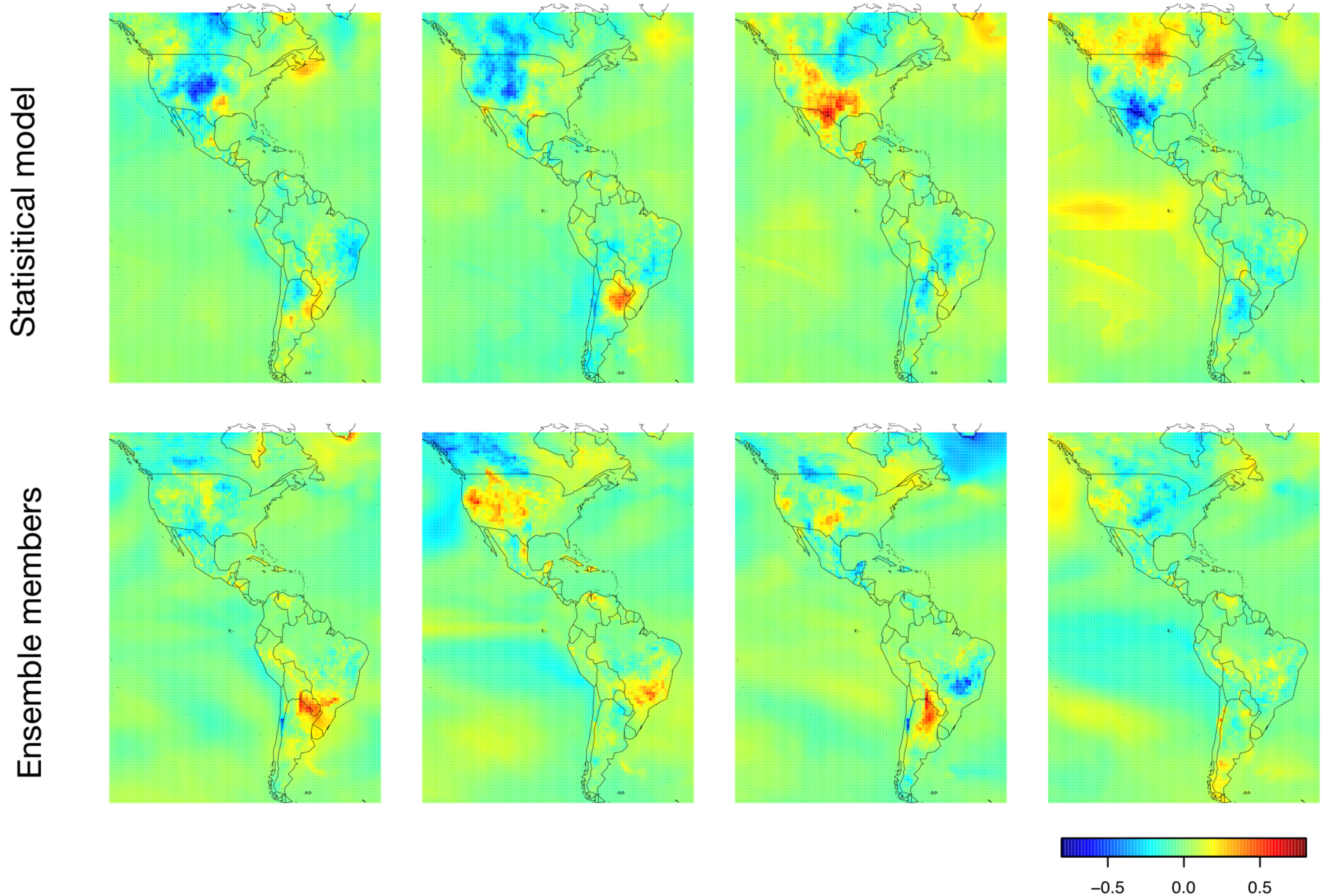
A lazy (but faster) route

Fit the local *Matérn* model and approximate using the LatticeKrig process model. From $\hat{\theta}, \hat{\nu}$ translate to weights and a.

Future Analysis

Estimate the weights and a directly for the LatticeKrig model.

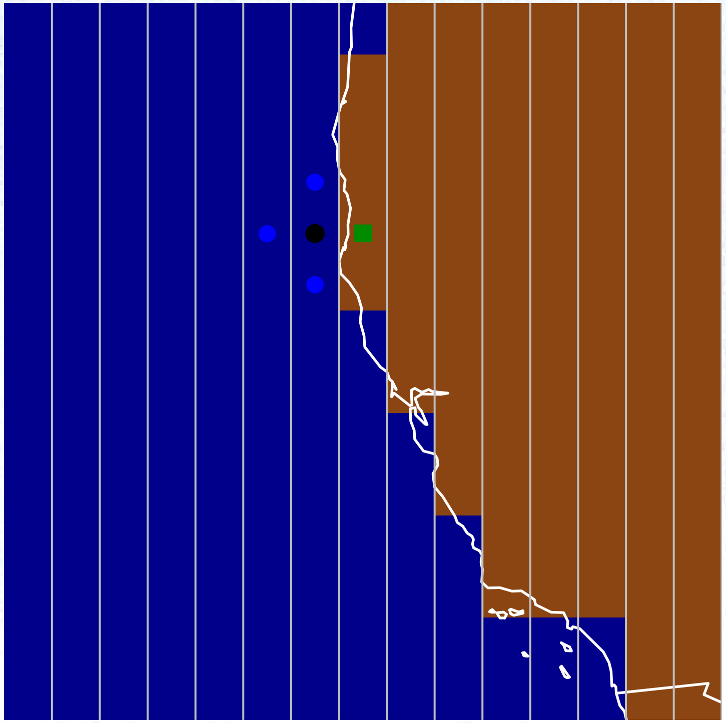
Emulating model fields



The land/ocean boundary

Now work directly with the LatticeKrig process pararmeters

Land/ Ocean mask over Califorina coast



SAR weights:

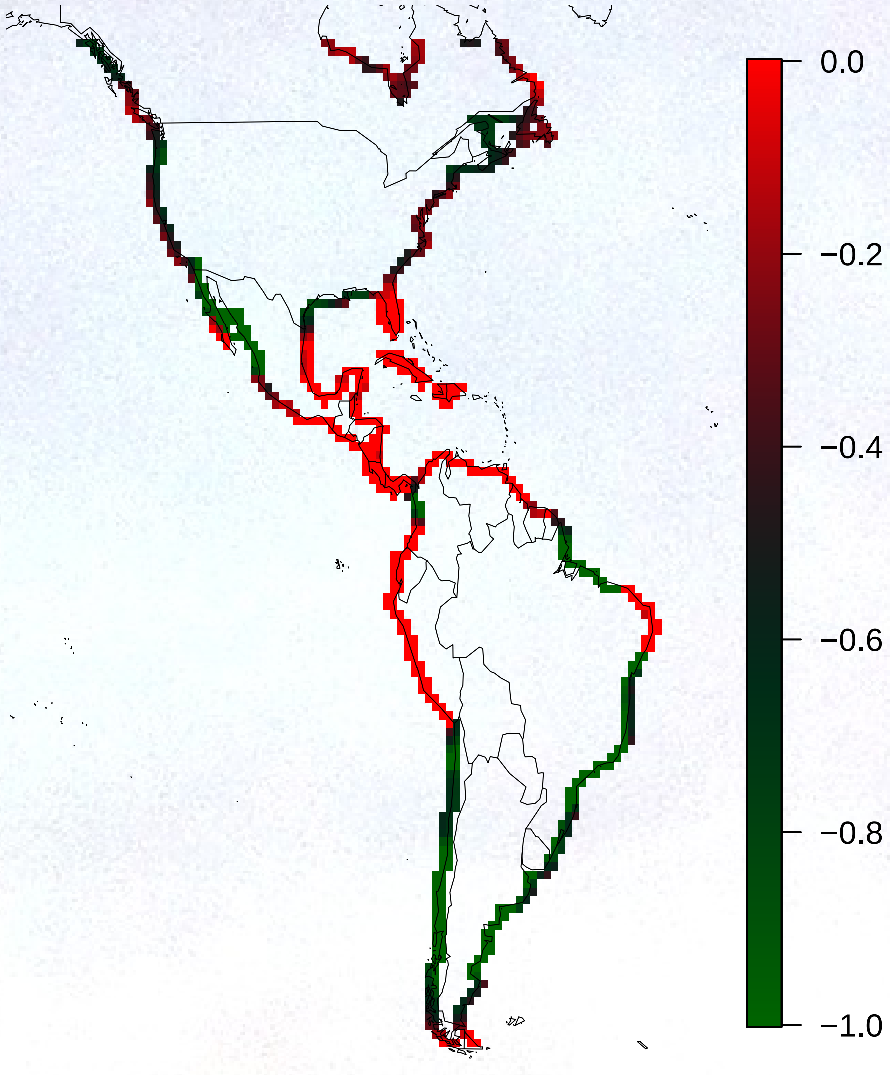
$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & -1 & . & . \\ . & -1 & a & \alpha & . \\ . & . & -1 & . & . \\ . & . & . & . & . \end{array}$$

$\alpha = -1$ isotropic same as the ocean links

$\alpha = 0$ no correlation of grid box with land

Boundary parameters

Using a 9X9 moving window and maximum likelihood



PART 5:

Parallel computation with R

The Cheyenne supercomputer.

$\approx$145K cores = 4032 nodes $\times$ 36 cores
and each core with 2Gb memory
52Pb parallel file system

- Core-hours are available to the NSF research community.
- Simple application process for graduate student allocations.
- Implementation of R on batch and interactive nodes.

Are zillions of R workers feasible?

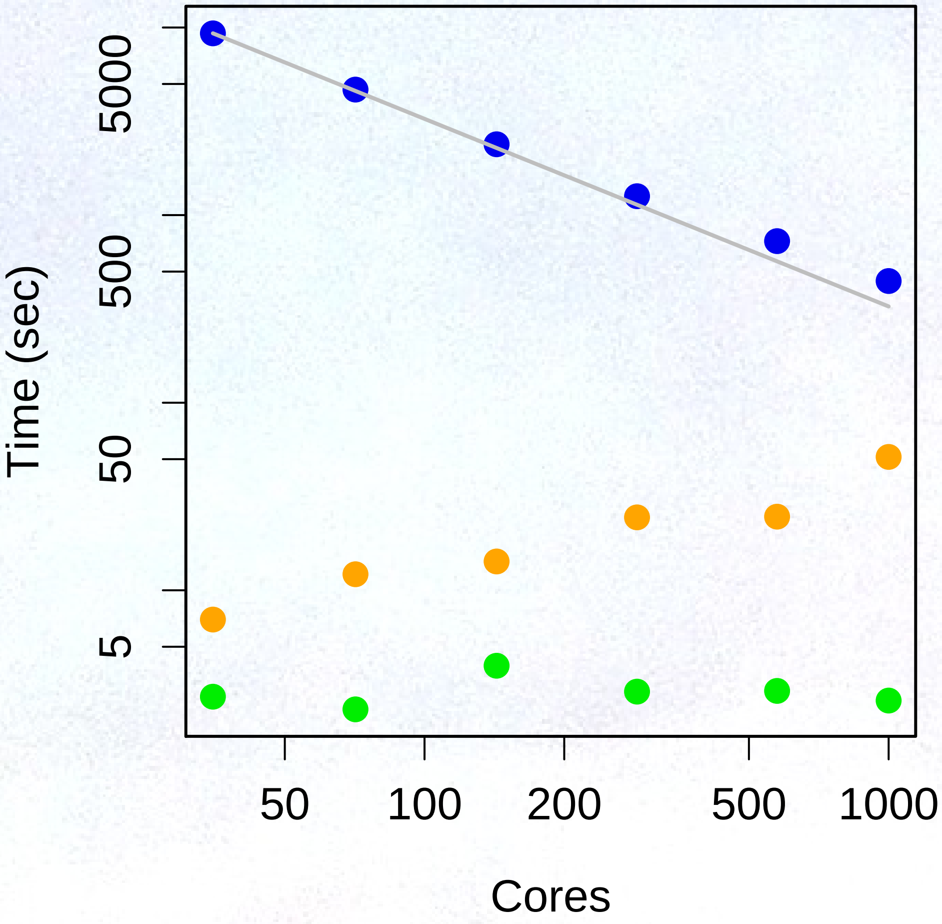
Yes for embarassingly parallel data analysis.

- `Rmpi` used to initiate many parallel R sessions from within a supervisor R session.
- Time to initiate 1000 workers takes about 1 minute.
- Little time lost in broadcasting the data object (12Mb) – about 3 seconds.

Approximate linear scaling using Rmpi

Individual times for:

spawn broadcast apply



Wall clock time in seconds to fit 1000 9×9 blocks with the LatticeKrig model.

Summary

- Emulation of climate model experiments for interpolation and uncertainty quantification is a fruitful area for data science.

- Local covariance fitting can capture variation in complex model output.

- Markov random field based models are suited for large data sets.

- There is an emerging role for supercomputers to support data analysis.

Software

- `fields` R package, Nychka et al. (2000 - present)
- `LatticeKrig` R package, Nychka et al. (2014- present)
- `HPC4Stats` SAMSI short course August 2017, Nychka, Hammerling and Lenssen.

Background reading

Nychka,D., Hammerling, D. , Krock, M. Wiens, A. (2017). Modeling and emulation of nonstationary Gaussian fields.
arXiv:1711.08077

Alexeeff, S. E., Nychka, D., Sain, S. R., & Tebaldi, C. (2016). Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments.
Climatic Change, 1-15.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., & Sain, S. (2015). A multi-resolution Gaussian process model for the analysis of large spatial datasets.
Journal of Computational and Graphical Statistics, 24(2), 579-599.

Thank you!