Nonstationary spatial data: think globally act locally

Douglas Nychka, National Center for Atmospheric Research

Summary

- Two applications from climate science
- Nonstationary Gaussian fields
- Unconditional simulation of pattern scaling fields
- Conditional simulation of ocean temperature fields
- Big Data analysis on super computers (Big R)

Challenges:

Building convolution covariance models for large problems and actually computing the beasts!

Credits

- Pattern scaling - simulation, Ashton Weins (CU), Mitchell Crock (CU), Dorit Hammerling (NCAR).
- ARGO floats - conditional simulation, Mikael Kuusela (SAMSI), Michael Stein (UC-Rutgers), Pulong Ma (U Cincinnati)

Kuusela, M. and Stein M. (2017). Locally stationary spatio-temporal interpolation of Argo profiling float data
arXiv:1711.00460v2

PART 1

Spatial problems in climate science

Future Climate

What will the climate be in 60 years?

- Need a scenario of future human activities.

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.

Community Earth System Model (CESM)

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

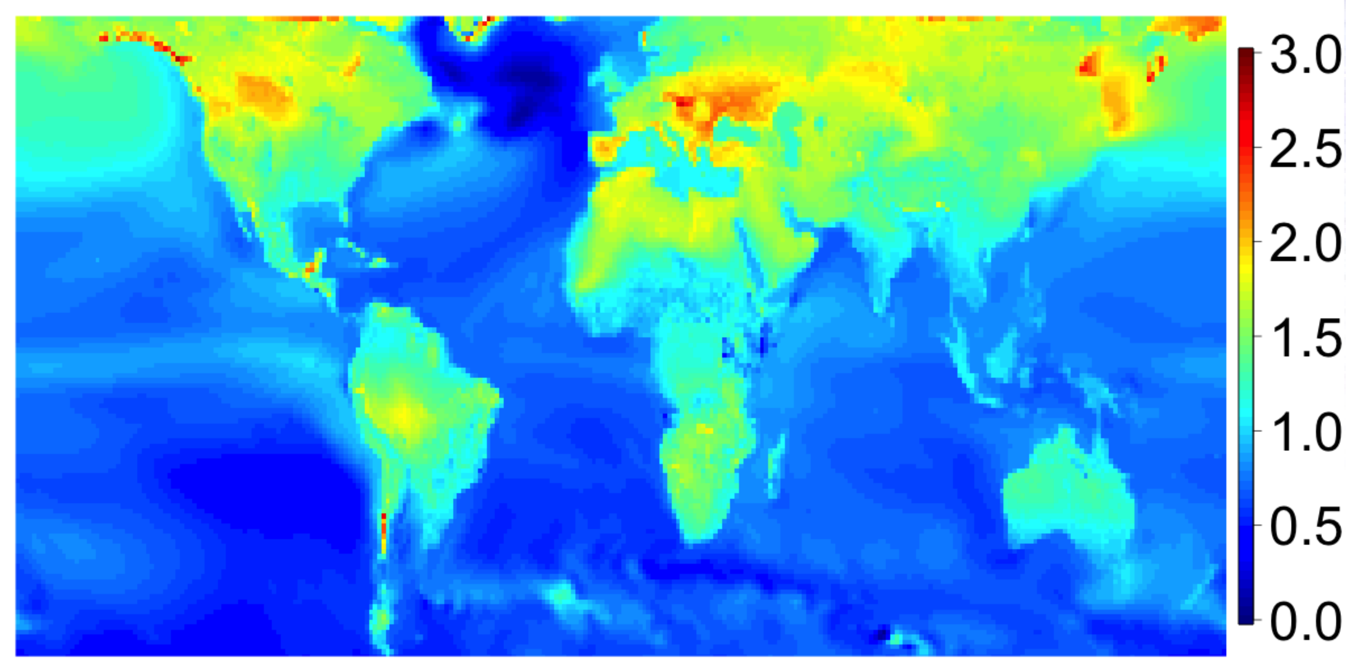
CESM Large Ensemble (CESM-LE)

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

Mean scaling pattern CESM

Slopes across 30 members for JJA

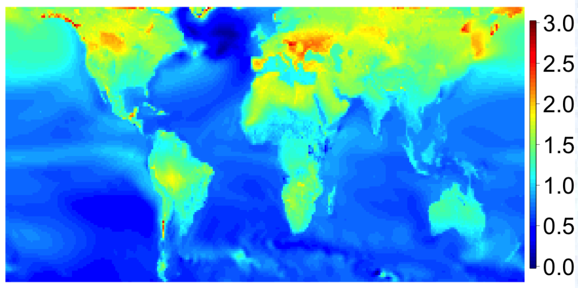


E. g. value of 2.5 means: a 1° global increase implies 2.5° increase locally.

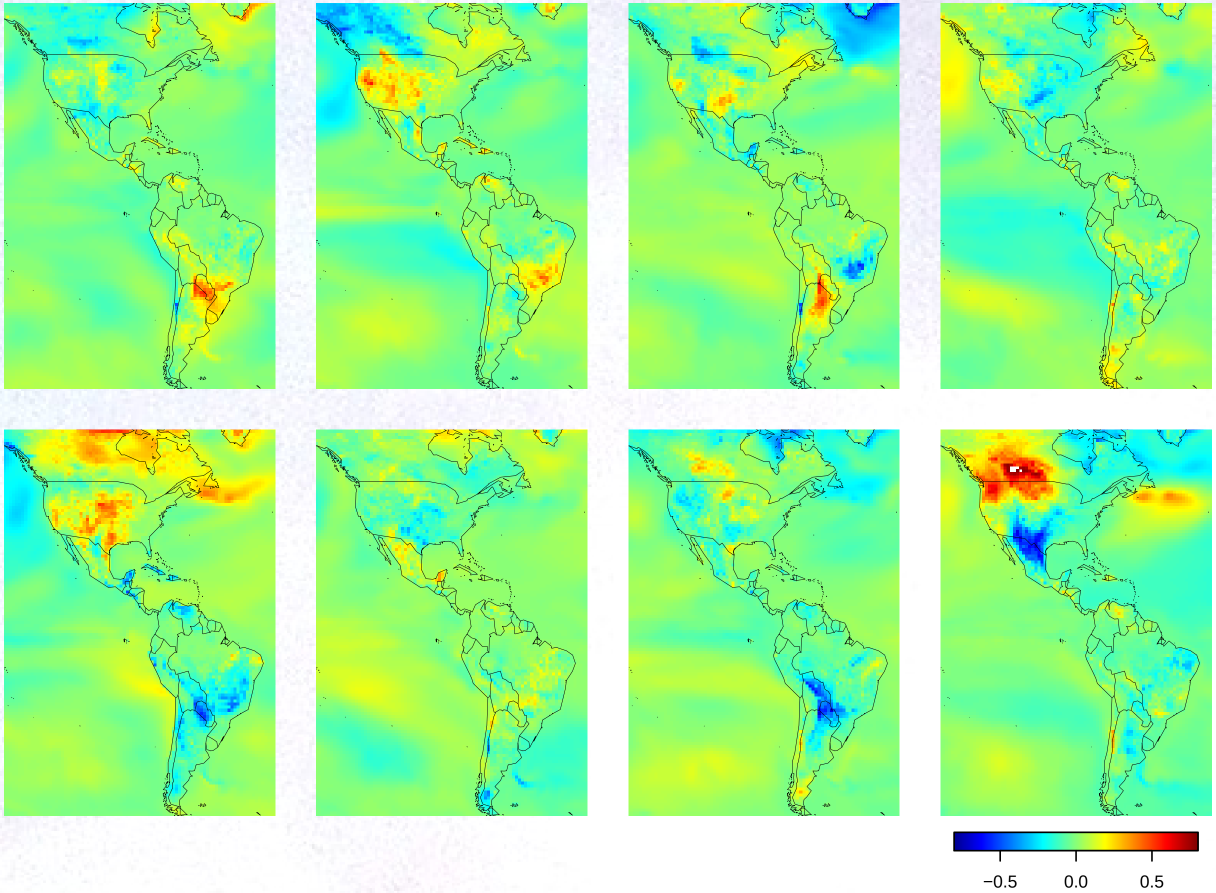
This allows us to determine the local mean temperature change based on a simpler model for the global average temperature

Individual patterns

Ensemble mean



First 8 out of 30 *centered* ensemble members

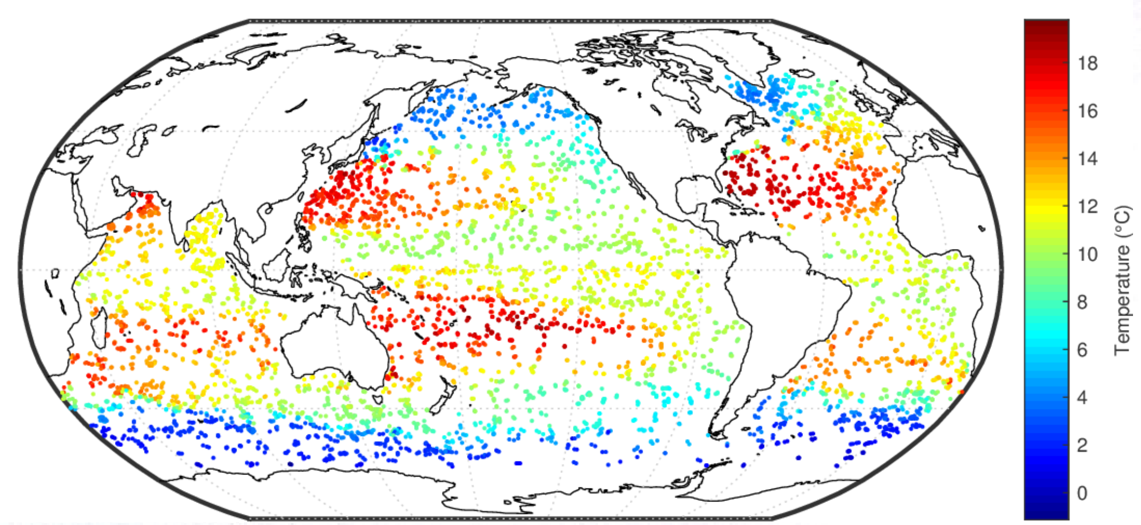


Goal: Simulate additional fields efficiently that match the spatial dependence in this 30 member ensemble.

Ocean heat content

- The ARGO observation network provides profiles of ocean temperature and salinity
- Measurements are irregular in space and time, $\approx$ 4000 floats taking profiles every 10 days

Temperature at 300 db for February, 2012.



Goal: Estimate the temperature field at different depths and times with measures of uncertainty

PART 2
Nonstationary Gaussian Processes

Gaussian process models

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

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

• $f(\boldsymbol{s})$ is a Gaussian process if any finite collection of $\{f(\boldsymbol{s}_1), \ldots, f(\boldsymbol{s}_N)\}$ has a multivariate normal distribution.

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

• An example of exponential covariance for a process that is stationary and isotropic:

$$k(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma^2 e^{(\frac{||\boldsymbol{s}_1 - \boldsymbol{s}_2||}{\theta})}.$$

– a strong assumption, note two covariance parameters σ and θ

Matérn covariance function

$k(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma^2 \text{Matern}_\nu(d) = \sigma^2 \mathcal{C} d^\nu \mathcal{K}_\nu(d)$,
and $d = ||\boldsymbol{s}_1 - \boldsymbol{s}_2||/\theta$

- $\mathcal{K}_\nu$ 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.
- σ^2 the process marginal variance
- When $\nu = .5$, Matérn is an exponential covariance, $\nu = \infty$, a Gaussian.

Nonstationary covariance functions

- Convolution model (Higdon, Fuentes)

Represent the process first, then figure out the covariance function

$$g(\boldsymbol{s}) = \int_{\Re^2} H(\boldsymbol{s}, \boldsymbol{u}) dW(\boldsymbol{u})$$

$dW(\boldsymbol{u})$ a two dimensional standard, white noise process.

The covariance function:

$$k(\boldsymbol{s}_1, \boldsymbol{s}_2) = \int_{\Re^2} H(\boldsymbol{s}_1, \boldsymbol{u}) H(\boldsymbol{s}_2, \boldsymbol{u}) d\boldsymbol{u}$$

- H can be the Green's function for a stochastic PDE (– a connection to INLA)

2-D exponential kernel example:

$$H(\boldsymbol{s},\boldsymbol{u}) = \frac{\sigma(\boldsymbol{s})}{\theta(\boldsymbol{s})} e^{-||\boldsymbol{s}-\boldsymbol{u}||/\theta(\boldsymbol{s})}$$

$$k_\theta(\boldsymbol{s}_1,\boldsymbol{s}_2) = \sigma(\boldsymbol{s}_1)\sigma(\boldsymbol{s}_2)\int \frac{1}{\theta(\boldsymbol{s}_1)\theta(\boldsymbol{s}_2)} e^{-||\boldsymbol{s}_1-\boldsymbol{u}||/\theta(\boldsymbol{s}_1)} e^{-||\boldsymbol{s}_2-\boldsymbol{u}||/\theta(\boldsymbol{s}_2)} d\boldsymbol{u}$$

- If $\theta(\boldsymbol{s}) \equiv \theta$ in 2-d this gives a Matérn with smoothness $\nu = 1.0$
- For unequal θ no simple closed form for this covariance.

Scale mixture (Paciorek, Stein)

$\nu = 1.0$

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

where

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

These are different models.

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

Open question how to figure out which model is more appropriate

Joint distribution

Observations

$$\boldsymbol{y}(\boldsymbol{s}_i) = \boldsymbol{f}(\boldsymbol{s}_i) + \boldsymbol{e}_i$$

$K_{i,j} = k(\boldsymbol{s}_i, \boldsymbol{s}_j)$ and $\boldsymbol{f}$ is $MN(0, K)$

log Likelihood

$$\ell(\boldsymbol{y}, [\sigma^2, \theta, \tau]) = -(1/2)\boldsymbol{y}^T (K + \tau^2 I)^{-1} \boldsymbol{y} - (1/2) log|K + \tau^2| + \mathcal{C}$$

- Maximize to find parameters
- For large data sets K is also large.

Simulating a Gaussian Process

at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_M$

- Form $K_{i,j} = k(\boldsymbol{s}_i, \boldsymbol{s}_j)$ covariance matrix at locations
- $\boldsymbol{f} = \Omega \boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0, 1)$

Ω is the matrix square root of K

Conditional simulation of a Gaussian Process

at locations $\boldsymbol{s}^g_1, \ldots, \boldsymbol{s}^g_M$

conditional on observations $y_1, \ldots, y_N$ at $\boldsymbol{s}^o_1, \ldots, \boldsymbol{s}^o_N$

- Form $K_{o,o}$ Covariance matrix at observations locations

$K_{g,g}$ Covariance matrix at grid locations

$K_{g,o}$ Cross-covariance matrix between grid and observation locations

- $\boldsymbol{f} = \hat{\boldsymbol{f}} + \Omega \boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0, 1)$

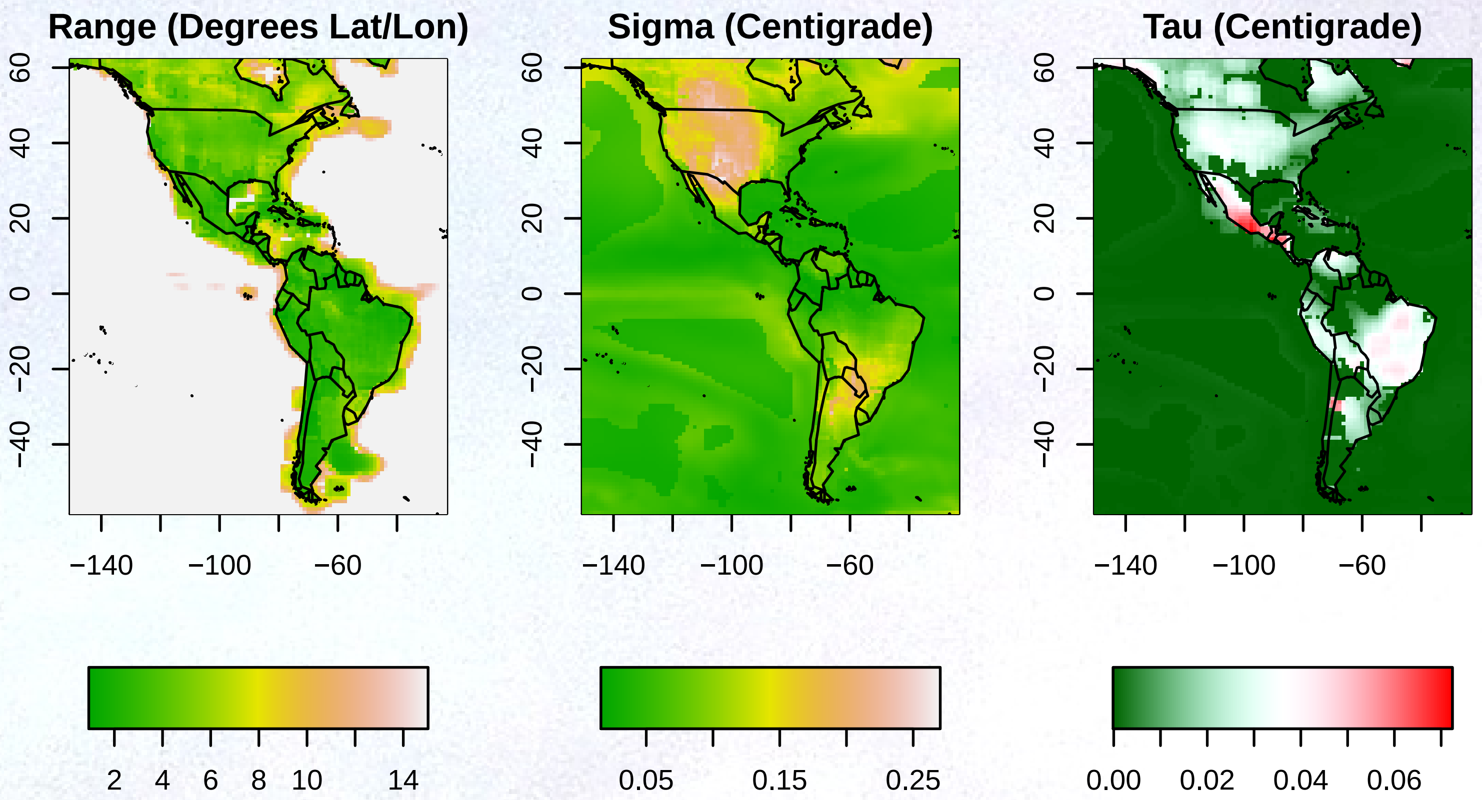
$\hat{\boldsymbol{f}}$ the conditional expectation for bbf (aka Kriging)

Ω is the matrix square root of $K_{g,g} - K_{g,o}(K_{o,o} + \tau^2 I)^{-1} K^T_{g,o}$

PART 3 Unconditional simulation

Climate model patterns

Local Matérn MLEs for the 30 member ensemble patterns



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

What should we do with these?

- Assume that the parameter estimates at the center of the window are good estimates for the parameter "fields" $\sigma(\boldsymbol{s})$, $\theta(\boldsymbol{s})$, and $\tau(\boldsymbol{s})$.

- RECALL Form $K_{i,j} = k(\boldsymbol{s}_i, \boldsymbol{s}_j)$ covariance matrix at all observation locations
- $\boldsymbol{f} = \Omega \boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0,1)$

Ω is the matrix square root of K

PROBLEM:

K is too big for computation.

SOLUTION:

Reexpress model in more computable form.

- Approximate the nonstationary model with a spatial autoregressive model (SAR)

(Parameters of the local Matérn models encoded as parameter fields in the LatticeKrig model.)

- Exploit sparse matrix methods to implement

$\boldsymbol{f} = \Omega \boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0,1)$

Ω^T is the sparse Cholesky decoposition of K

A Spatial Autoregression (SAR)

Gridded field:

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

SAR weights:

$$\begin{array}{ccccc} . & . & . & . & . \\ . & . & -1 & . & . \\ . & -1 & a(\boldsymbol{s}) & -1 & . \\ . & . & -1 & . & . \\ . & . & . & . & . \end{array}$$

The filter:

$a(\boldsymbol{s})c_* - (c_1 + c_2 + c_3 + c_4) =$ white noise

- $a(\boldsymbol{s})$ needs to be greater than 4.
  $1/\sqrt{a(\boldsymbol{s}) - 4}$ – an approximate range parameter
- $B\boldsymbol{c} =$ i.i.d.$N(0, 1)$ where B is a sparse matrix
- Covariance for $\boldsymbol{c}$ is $(B^T B)^{-1} = Q^{-1} = K$

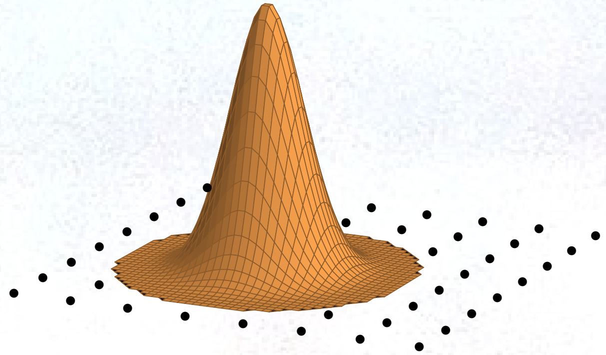
Representing a random surface

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

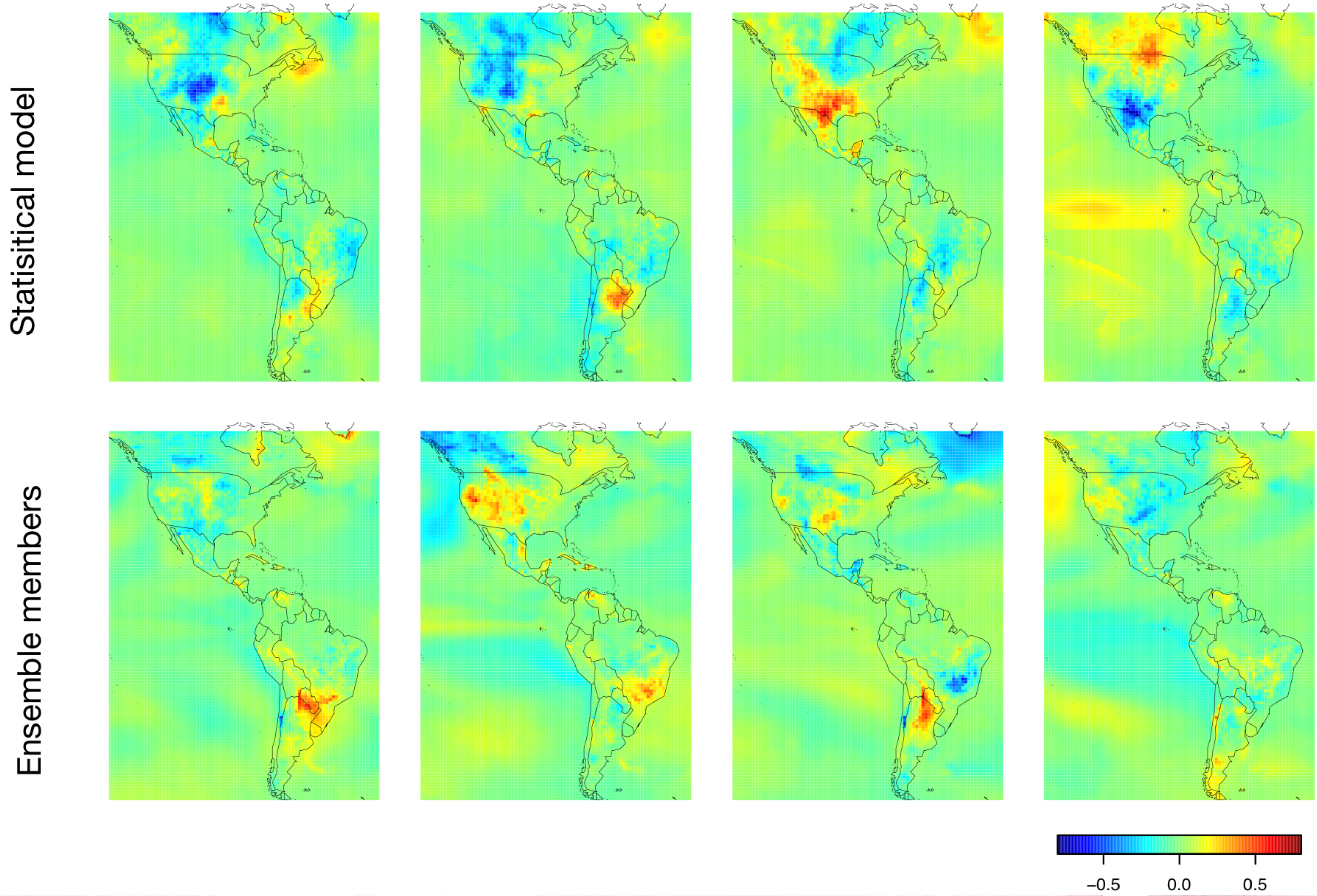
- $\boldsymbol{c}$ is the random field from the SAR.
- $\{\phi_j(x)\}$ are compact, radial basis functions :

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

A member of the Wendland basis functions



Emulating pattern scaling fields



PART 4
Conditional simulation

Ocean temperatures

Predicted surface temperature field from ARGO float observations (Kuusela and Stein (2017))

- Covariance parameters are from Matérn family
- local windows of 20×20 degrees and 1 month
- student-T distribution used to account for heavy tailed observations.

What should we do with these?

- Assume that the parameter estimates at the center of the window are good estimates for the parameter "fields" $\sigma(\boldsymbol{s})$, $\theta(\boldsymbol{s})$, and $\tau(\boldsymbol{s})$.

- RECALL
- $\boldsymbol{f} = \hat{\boldsymbol{f}} + \Omega \boldsymbol{e}$ where $\boldsymbol{e}$ are iid $N(0,1)$

$\hat{\boldsymbol{f}}$ the conditional expectation for bbf (aka Kriging)

Ω is the matrix square root of

$$K_{g,g} - K_{g,o}(K_{o,o})^{-1} K_{g,o}^T$$

PROBLEM: all the Ks are too big for computation.

SOLUTION: Simulate conditional field by moving local neighborhoods

- Generate a realization of $\boldsymbol{e}$ on the grid.

LOOP OVER GRID LOCATIONS

- For each grid location evaluate Ω in a local neighbor centered at this point, Ω_{local}

- Find the symmetric square root of Ω_{local}

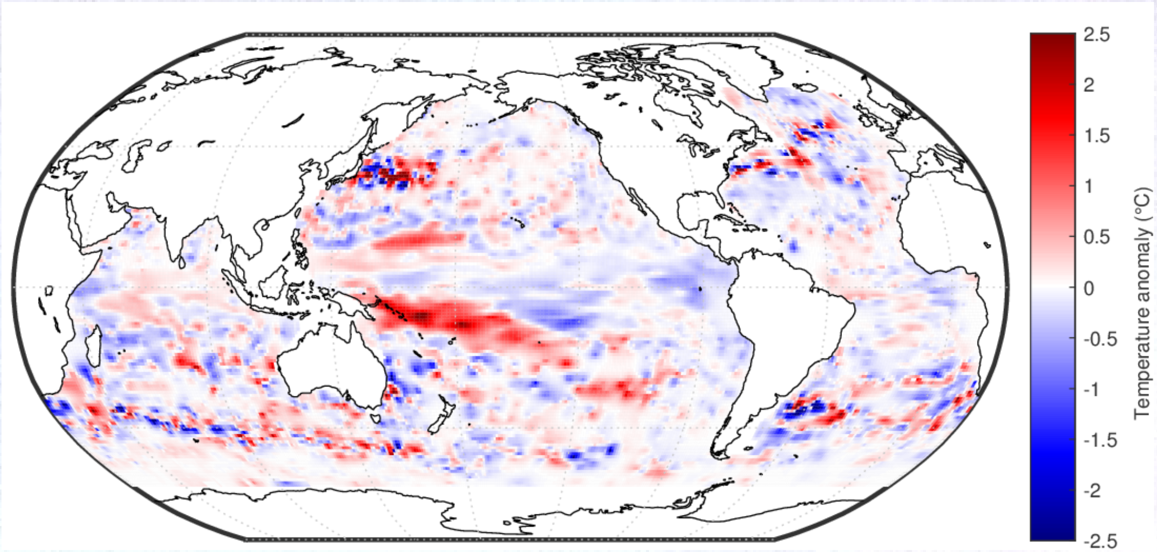
- Apply the center row of square root matrix to the right subset of $\boldsymbol{e}$. (throw the other rows away!)

END LOOP

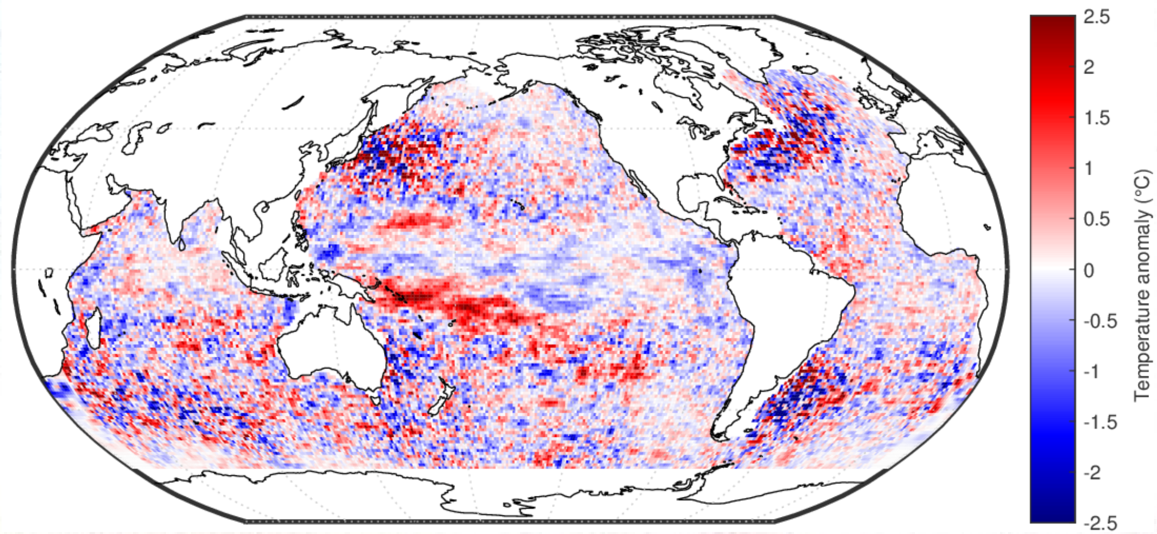
This is an embarrassingly parallel computation.

ARGO analysis

Conditional Mean



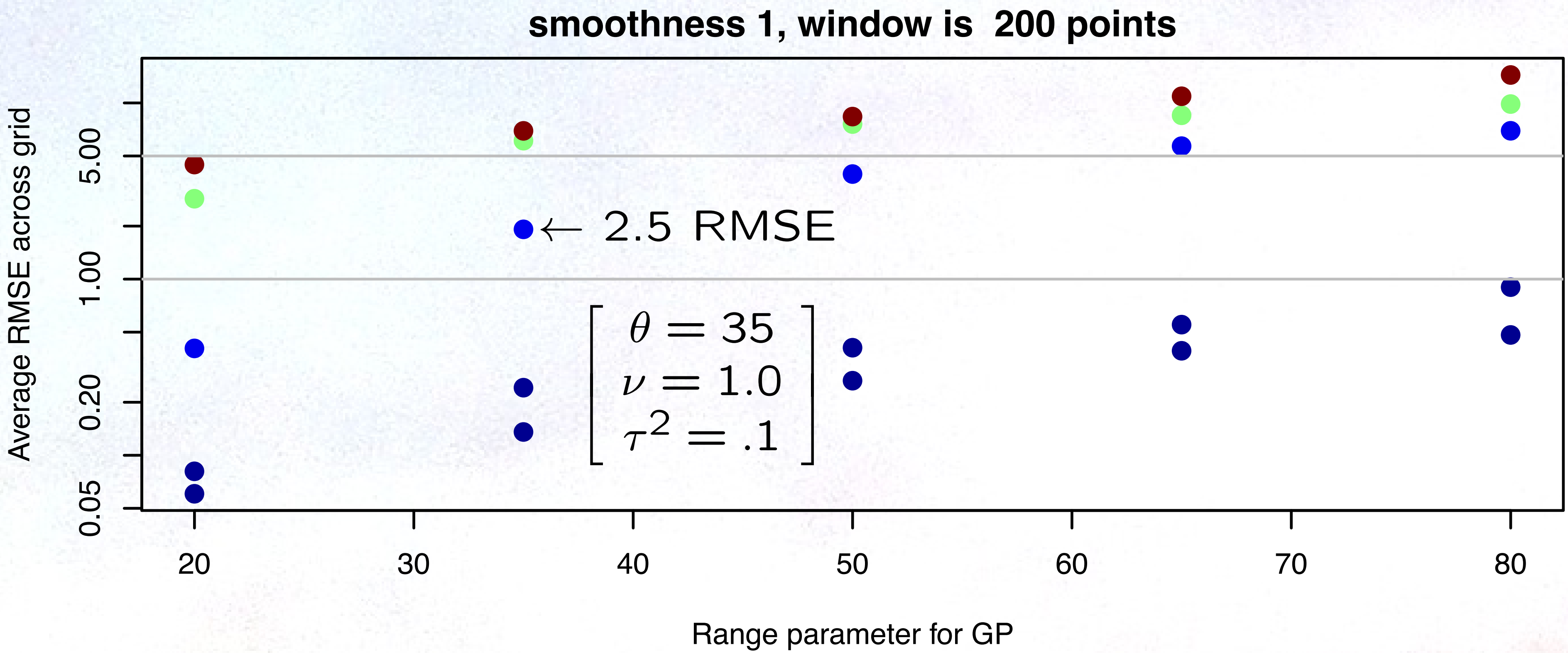
Draw from Conditional Distribution



Why this works

- The "screening effect" for spatial prediction suggests that the Ω matrix will largely depend on a local neighborhood of the observations.

- Can compute explicitly how well the center row of $\Omega^{1/2}_{\text{local}}$ approximates a much larger domain/neighborhood.



Values of τ^2 **.005**, **.01**, **.1**, **.5**, **1.0**

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.

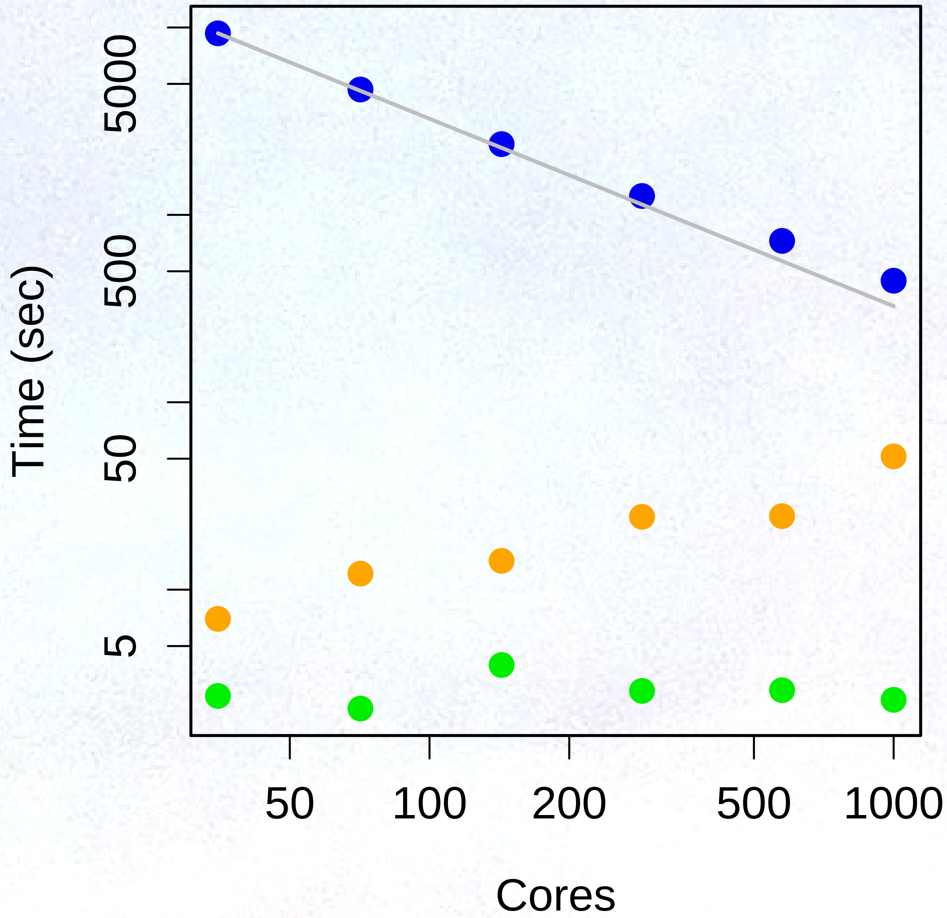
• Time to initiate 1000 workers takes about 1 minute.

• 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 and in geophysical fields.

- 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

Kuusela , M and Stein M. (2017). Locally stationary spatio-temporal interpolation of Argo profiling float data
arXiv:1711.00460v2

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!