Nonstationary spatial data: think globally act locally

Douglas Nychka,
Colorado School of Mines (CSM)
National Center for Atmospheic Research

Summary

- NCAR Large Ensemble
- Nonstationary Gaussian fields
- Estimating the covariance
- Emulating climate fields

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

Credits

- SAR: Ashton Wiens (CU), Mitchell Krock (CU), Dorit Hammerling (CSM) William Kleiber (CU)

- Parallel computation: Florian Gerber (CSM)

- Climate Model Experiements: Claudia Tebaldi (PNNL) Stacey Alexeeff (Kaiser)

PART 1
Spatial problems in climate science

Future climate

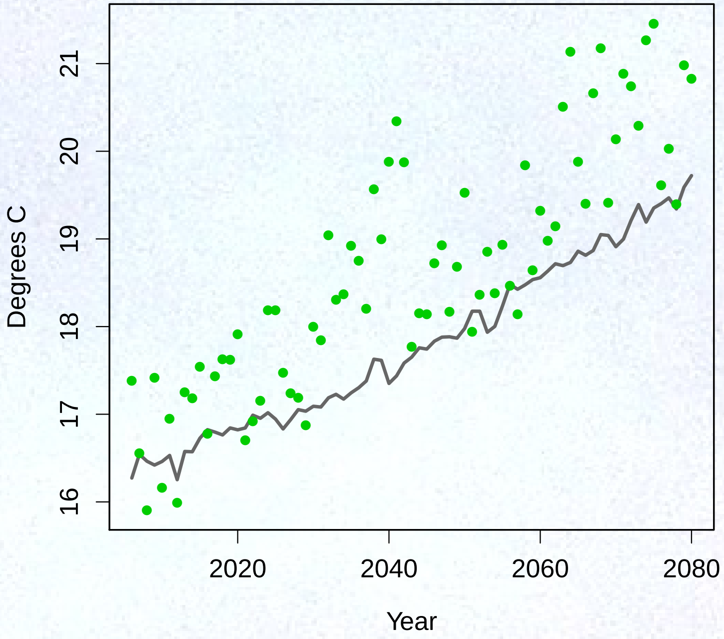
CESM Large Ensemble (CESM-LENS) A 30+ member ensemble of CESM simulations that have been designed to study the local effects of climate change

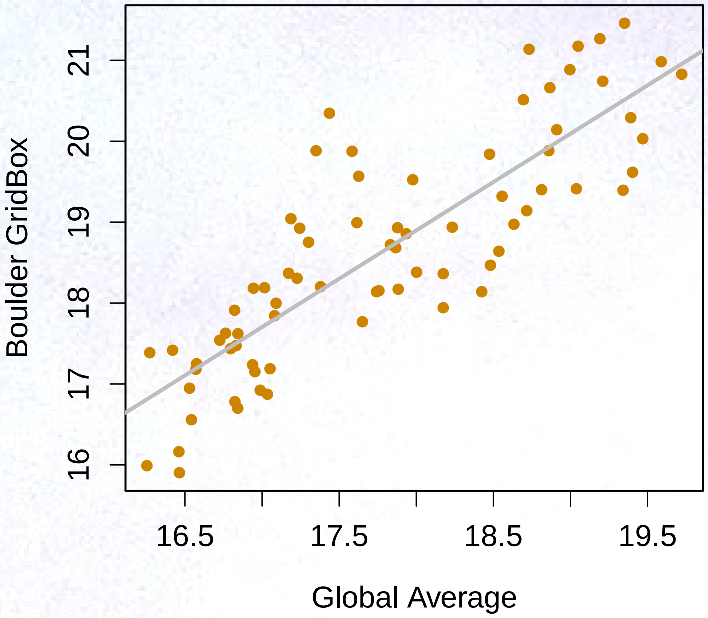
The range of ensemble members characterize some of the uncertainty due to natural variability in the Earth system

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

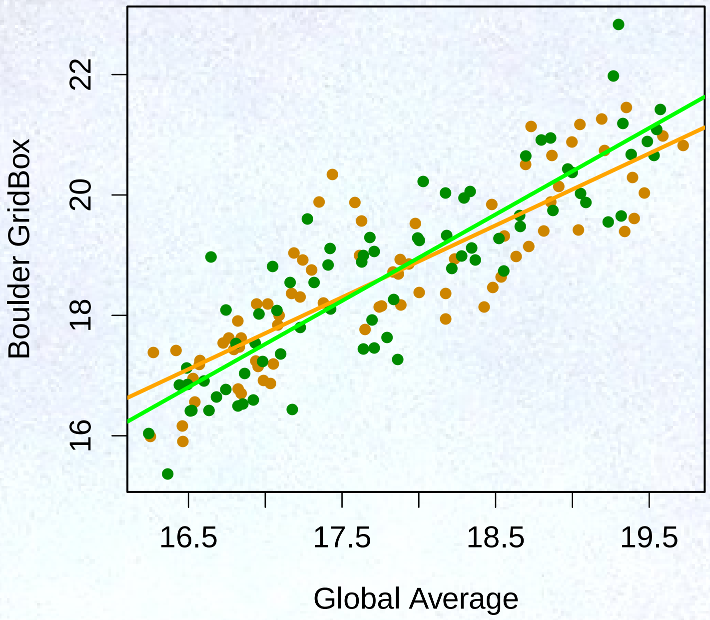
Simulated JJA Temperature

Grid box around Boulder/Denver – average summer temperature.





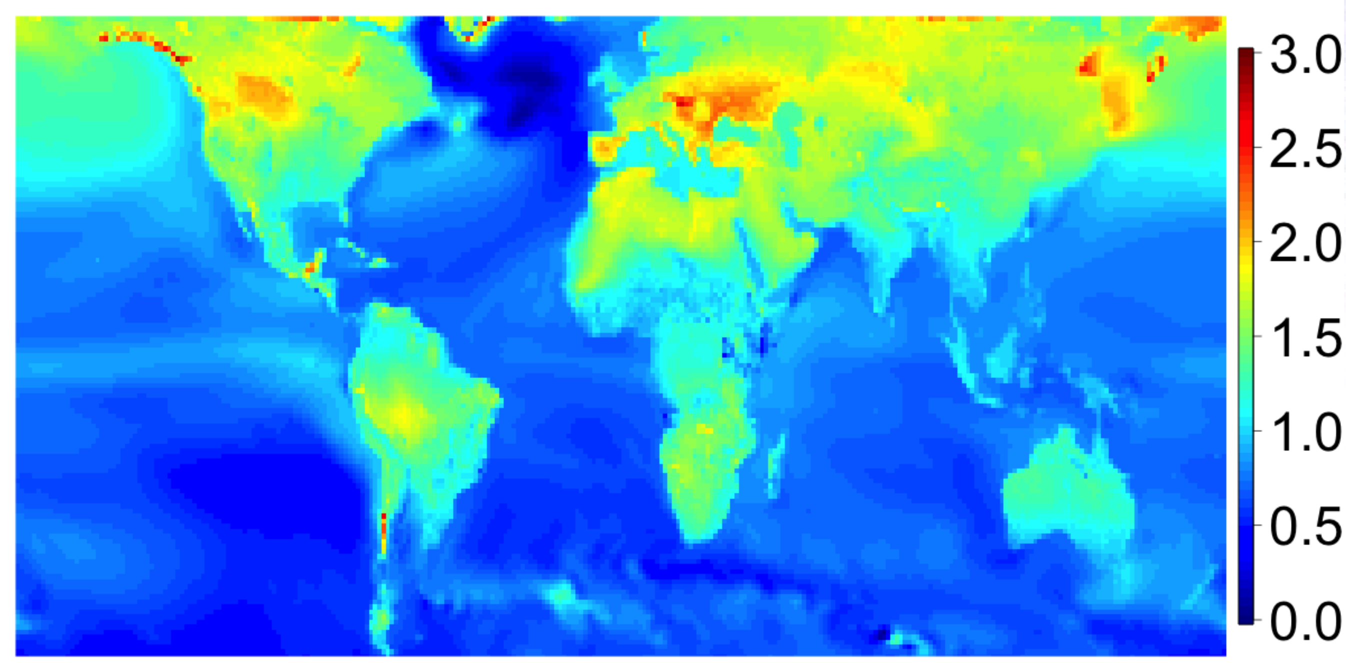
First and second ensemble member



We have 30 of these and at 50K+ gird boxes – and this is just surface temperature!

Mean scaling pattern

OLS slopes across 30 CESM-LENS members for JJA

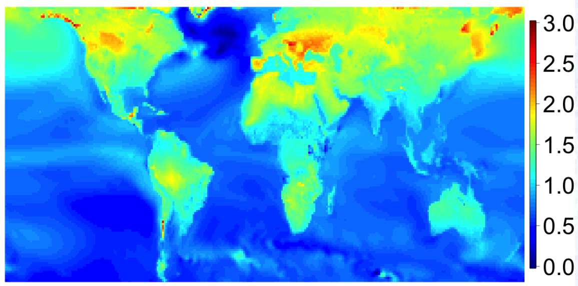


E. g. value of 2.5 means: a 1° global increase implies 2.5° increase locally.

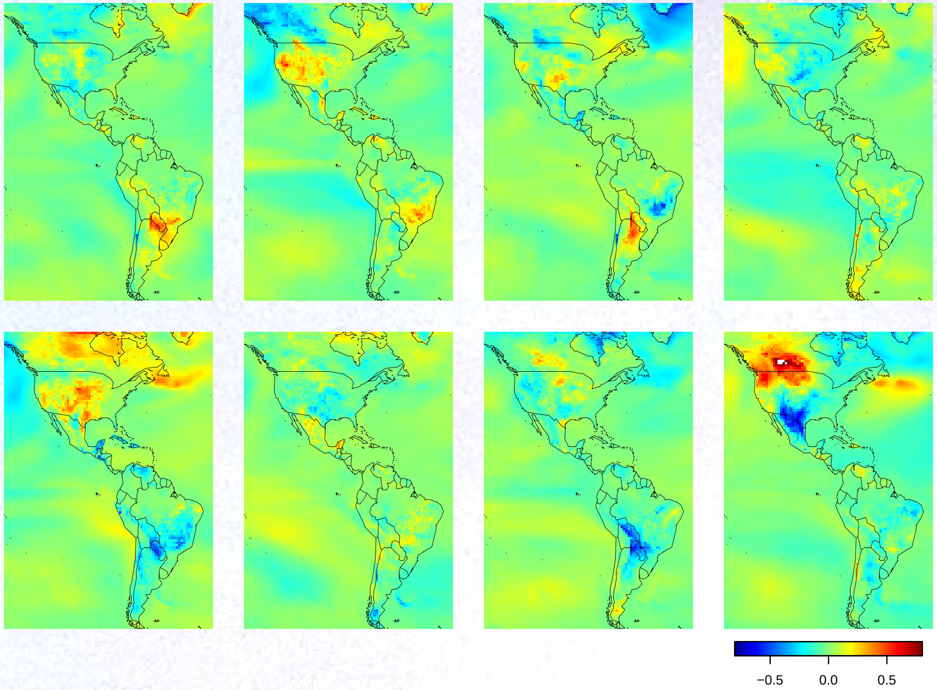
This allows us to determine the local temperature change based on a simpler model for the global average temperature

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.

Build a statistical emulator of this part of the model.

PART 2
Nonstationary Gaussian Processes

Gaussian process models

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

- $E[u(\boldsymbol{s})] = 0$ and $C(\boldsymbol{s}_1, \boldsymbol{s}_2) = E[u(\boldsymbol{s}_1)u(\boldsymbol{s}_2)]$

- A covariance that is stationary and isotropic:

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma^2 \mathcal{M}(\kappa ||\boldsymbol{s}_1 - \boldsymbol{s}_2||).$$

– a strong assumption, note two covariance parameters σ and κ.

- Anisotropy: A a linear tranformation :

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma^2 \mathcal{M}(||A(\boldsymbol{s}_1 - \boldsymbol{s}_2)||).$$

A is a linear rotation and scaling of the coordinates

Nonstationary covariance functions

- Convolution model (Higdon, Fuentes)

Represent the process first, then figure out the covariance function

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

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

The covariance function:

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \int_{\Re^2} \Psi(\boldsymbol{s}_1, \boldsymbol{s}^*) \Psi(\boldsymbol{s}_2, \boldsymbol{s}^*) d\boldsymbol{s}^*$$

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

2-D exponential kernel example:

$$\Psi(\boldsymbol{s}, \boldsymbol{s}^*) = \sigma(\boldsymbol{s}) e^{-\kappa(\boldsymbol{s})\|\boldsymbol{s}-\boldsymbol{s}^*\|}$$

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma(\boldsymbol{s}_1)\sigma(\boldsymbol{s}_2) \int e^{-\kappa(\boldsymbol{s})||\boldsymbol{s}_1-\boldsymbol{s}^*||} e^{-\kappa(\boldsymbol{s}^*)||\boldsymbol{s}_2-\boldsymbol{s}^*||} d\boldsymbol{s}^*$$

- If $\theta(\boldsymbol{s}) \equiv \theta$ in 2-d this gives a Matérn with smoothness $\nu = 1.0$
- C is positive definite because it is built from the process description.
- For unequal κ no simple closed form for this covariance.
  –direct use of covariance is not feasible for large problems

Ansiotropy:

A = diagonal $\times$ rotation .

$$\Psi(\boldsymbol{s},\boldsymbol{u}) = \sigma e^{-||A(\boldsymbol{s}-\boldsymbol{u})||}$$

Nonstationarity: $A(\boldsymbol{s})$ varies in space.

$$\Psi(\boldsymbol{s},\boldsymbol{s}^*) = \sigma(\boldsymbol{s}) e^{-||A(\boldsymbol{s})(\boldsymbol{s}-\boldsymbol{s}^*)||}$$

Our approach

- Represent the process in terms of a differential operator

- Discretize the problem to give sparse matrices.

Stochastic partial differential equations

Let Ψ can be the Green's function for a stochastic PDE with operator $\mathcal{L}$. I.e.

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

is equivalent to

$$\mathcal{L}(u(\boldsymbol{s})) = dW(\mathrm{s})$$

An SPDE model (Lindgren and Rue) for u ($\alpha = 2$)

$$(\kappa^2 - \nabla \cdot \nabla) u(\boldsymbol{s}) = dW(\mathrm{s})$$

$\Delta u = \frac{\partial^2}{\partial s_1^2} u + \frac{\partial^2}{\partial s_2^2} u \equiv \nabla \cdot \nabla u$

This gives a process that is Matérn, range κ and smoothness 1.

Discretizing

$u_{i,j}$ now the process on a unit lattice

$\frac{\partial^2}{\partial s_1^2}u + \frac{\partial^2}{\partial s_2^2}u$ is approximated by differences

$$\frac{\partial^2}{\partial s_1^2}u \approx (u_{i+1,j} - u_{i,j}) - (u_{i,j} - u_{i-1,j}) = u_{i+1,j} - 2u_{i,j} + u_{i-1,j}$$

The full discretization of the SPDE

$$(\kappa^2 - \Delta)u(\boldsymbol{s}) = dW(\mathrm{s})$$

$$\kappa^2 u(\boldsymbol{s}) - \frac{\partial^2}{\partial s_1^2}u(\boldsymbol{s}) \qquad - \frac{\partial^2}{\partial s_2^2}u(\boldsymbol{s}) \qquad = dW(\mathrm{s})$$

$$\kappa^2 u_{i,j} - (u_{i+1,j} - 2u_{i,j} + u_{i-1,j}) - (u_{i,j+1} - 2u_{i,j} + u_{i,j}) = e_{i,j}$$

$$\kappa^2 u_{i,j} + 4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} = e_{i,j}$$

$$[\kappa^2 + 4]u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} = e_{i,j}$$

A Spatial Autoregression (SAR)

Gridded field of $\boldsymbol{u}$:

.	.	.	.	.
.	.	u_1	.	.
.	u_2	u_*	u_3	.
.	.	u_4	.	.
.	.	.	.	.

SAR weights:

.	.	.	.	.
.	.	-1	.	.
.	-1	$\kappa^2 + 4$	-1	.
.	.	-1	.	.
.	.	.	.	.

The SAR:

$$(\kappa^2 + 4)u_* - (u_1 + u_2 + u_3 + u_4) = \text{white noise}$$

- κ is not exactly equal to the continuous range.
- $B\boldsymbol{u} = \text{i.i.d.}N(0,1)$ where B is a sparse matrix
- Covariance matrix for $\boldsymbol{u}$ is $(B^{-1}B^{-T})$

- A SAR is a CAR with precision $B^T B$

Key ideas

$Bu = \text{i.i.d.} N(0, 1)$

- Sparsity in B facilitates computation for large problems

- Coefficients of B can vary row by row (i.e. lattice point by lattice point) to represent nonstationarity

- $\boldsymbol{u} = B^{-1}\boldsymbol{e}$, a discretized version of a convolution process.

B^{-1} plays the role of Ψ in $u(\boldsymbol{s}) = \int_{\Re^2} \Psi(\boldsymbol{s}, \boldsymbol{s}^*) dW(\boldsymbol{s}^*)$

- $B\boldsymbol{u}$, transforms the field (possibly observations) to white noise

B plays the role of $\mathcal{L}$ in $\mathcal{L}u = dW$

The Anisotropic Stencil

For a single lattice point (Flugstadt, et al.)

.	.	.	.	.
.	$2H_{12}$	$-H_{22}$	$-2H_{12}$	.
.	$-H_{11}$	$\kappa^2 + 2H_{11} + 2H_{22}$	$-H_{11}$	.
.	$-2H_{12}$	$-H_{22}$	$2H_{12}$	.
.	.	.	.	.

- $H = AA^T$
- Key idea is to allow these weights to vary for each lattice point.

PART 3 Fitting the model

Local Likelihoods

- Estimate A in local windows for a Matérn smoothness 1.
- Translate this covariance model into the SAR form.

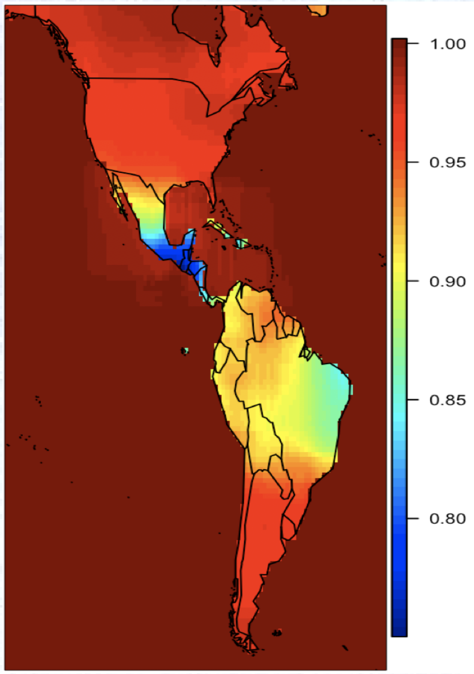
What is new?

- Quantified the accuracy of *local* Matérn estimates based on replicates.
- Established the translation from local covariance estimates into a global SAR model.

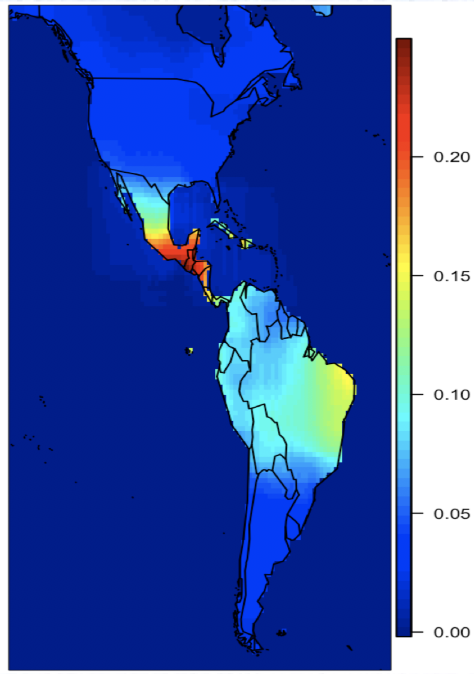
Climate model patterns

Local Matérn MLEs for the 30 member ensemble patterns

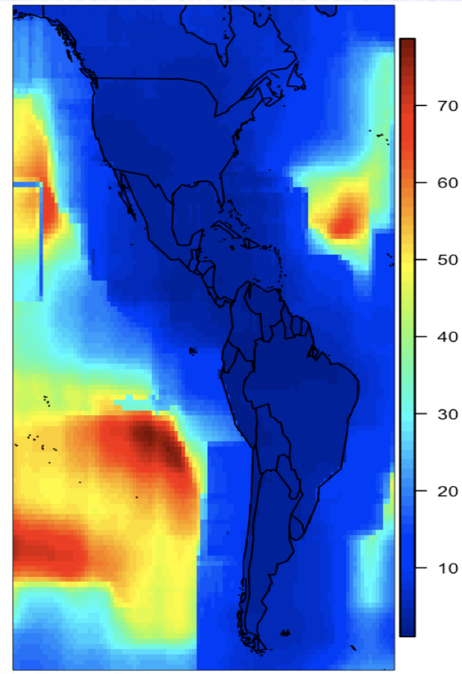
Sill variance



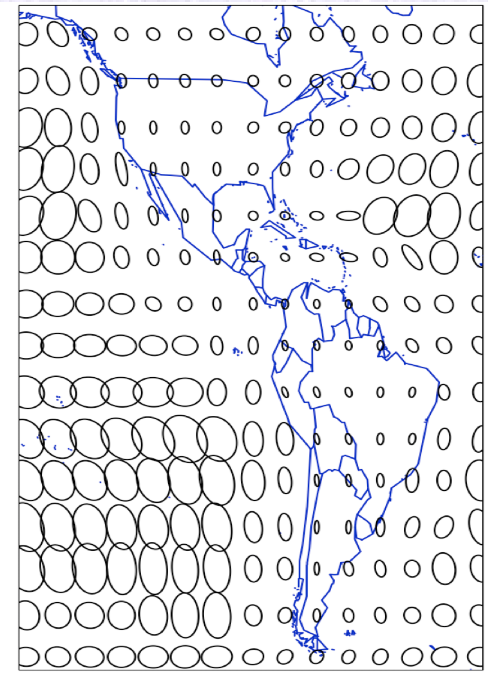
Nugget variance



$1/\sqrt{\kappa_1 \kappa_2}$



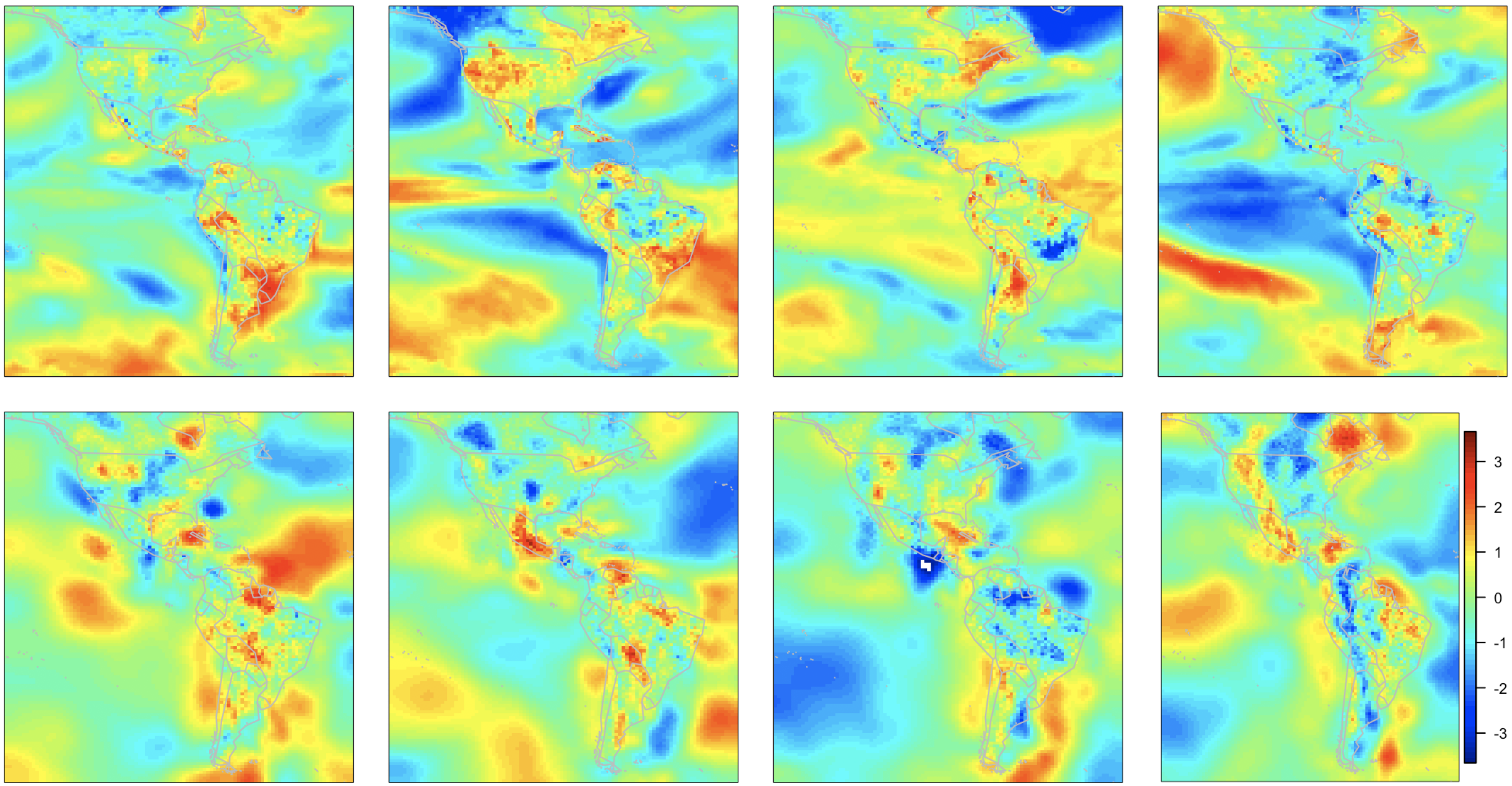
Orientation



PART 4
Simulation and checking

Some realizations of the process

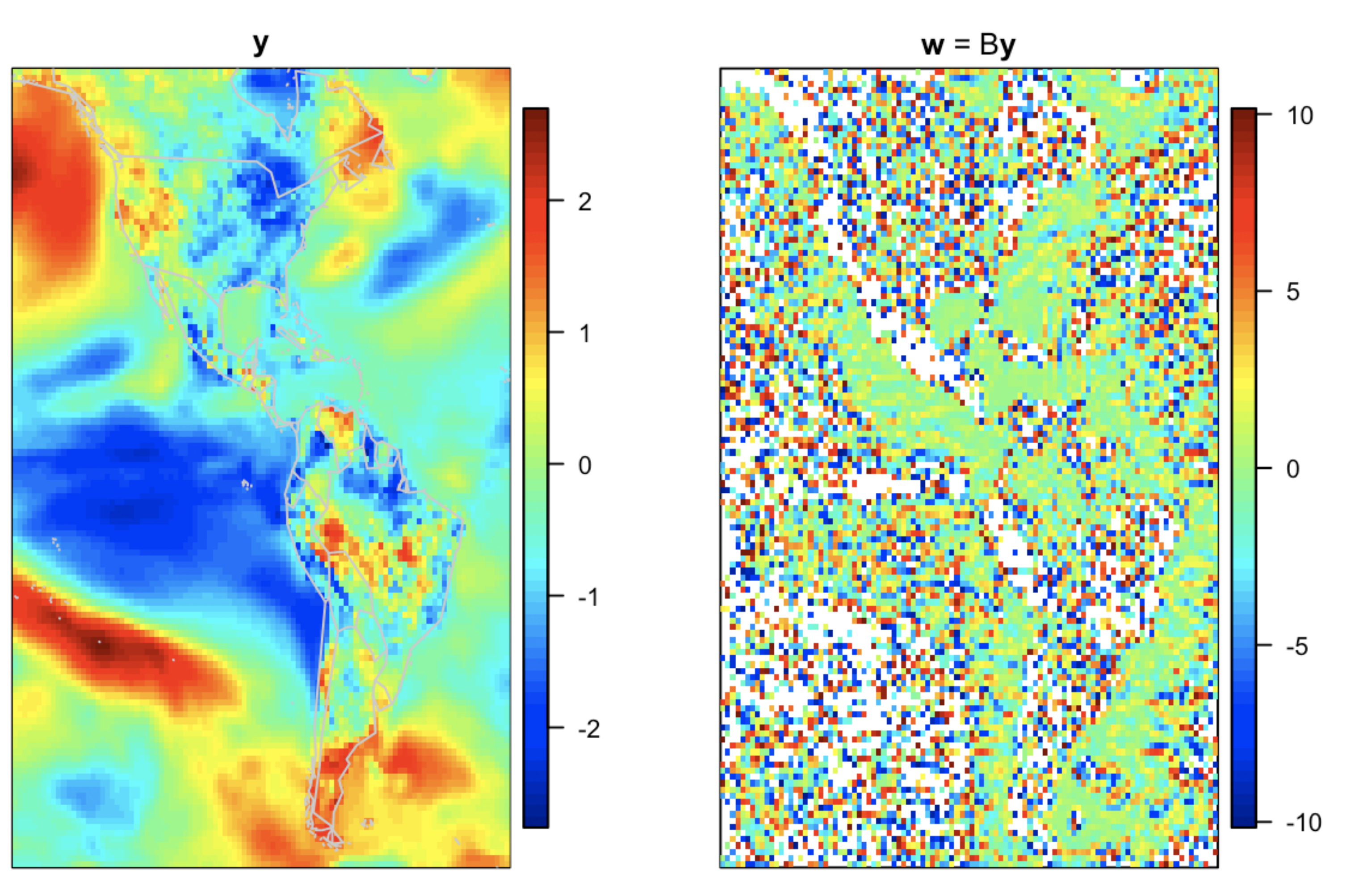
Top four ensemble members from LENS climate model experiment.



Bottom four realizations from SAR model.

Transforming to white noise

Applying the B SAR operator to the first ensemble member



What is next?

- Combine SAR approximation with basis functions for unequally spaced observations

- Improve approximation adding processes at different scale levels.

- Improve whitening transformation.

- Land verses Ocean ?

.	.	.	.	.
.	$2H_{12}$	$-H_{22}$	$-2H_{12}$	.
.	$-H_{11}$	$\kappa^2 + 2H_{11} + 2H_{22}$	$-H_{11}$	.
.	$-2H_{12}$	$-H_{22}$	$2H_{12}$	.
.	.	.	.	.

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!