Approximate conditional simulation for large spatial data

Doug Nychka, Maggie Bailey, Soutir Bandyopadhyay
Colorado School of Mines

Outline

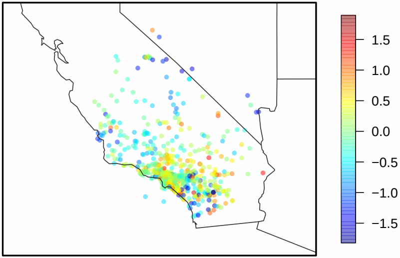
- USGS ShakeMap
- Conditional simulation
- Approximate unconditional simulation
- Approximate fast prediction
- Approximation theory

WaldoTobler:
"everything is related to everything else, but near things are more related than distant things."

... but the distant things become less important given the near things ...

USGS Shake Map data product

Median ground motion for the Ridgecrest, CA earthquake (July 2019),



- Predict motion for locations that are not monitored.
- quantify the uncertainty
- Want to do this fast for a large number of observation locations

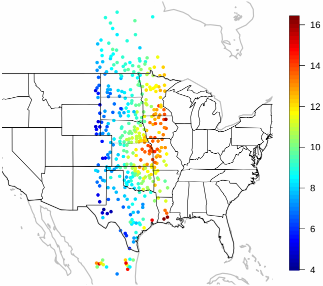
(e.g. $\sim 10^3$) and a large regular grid (e.g. $\sim 512 \times 512$)

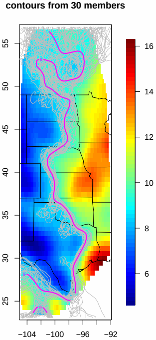
The goal

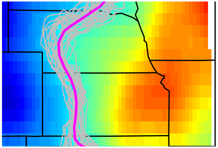
Push the limits of interactive spatial data analysis on a typical laptop.

Conditional Simulation

Summer mean rainfall and the 9in contour







Each grey contour is equally plausible conditional on the data.

Gaussian additive model

$$\mathbf{y}_k = f(\mathbf{s}_k) + \mathbf{e}_k$$

- f a Gaussian process
- $\{e_k\}$ independent. $N(0, \tau^2)$
- Ignore the mean model, i.e. $E(f(\mathbf{s})) = 0$
- Covariance for f, $k(\mathbf{s}, \mathbf{s}')$ assumed *stationary*.

Conditional simulation

- For fixed covariance parameters, conditional distribution

$$[f(s)|\mathbf{y}] \sim MN(\hat{\boldsymbol{\mu}}, \Omega)$$

$\hat{\boldsymbol{\mu}}$ conditional mean and Ω conditional variance.

Draw samples from this distribution for a large regular grid – without explicitly constructing Ω

- Monte Carlo estimate of the prediction error
- Useful for nonlinear features such as level sets, contour lines.

Conditional simulation algorithm

$\mathcal{S}^{\mathcal{G}} = \{s_k^G\}$ Grid locations,
$\mathcal{S}^{\mathcal{O}} = \{s_i^O\}$ Observation locations

1. Find $\hat{\boldsymbol{\mu}} = E[\mathbf{f}^G|\mathbf{y}]$
2. Simulate (unconditional) process on union of $\mathcal{S}^{\mathcal{G}}$ and $\mathcal{S}^{\mathcal{O}}$
   Gives $\mathbf{f}^G$ and $\mathbf{f}^O$ jointly.
3. Form synthetic data $\mathbf{y}^\star = \mathbf{f}^O + \mathbf{e}$
4. Calculate conditional mean, $\hat{f}^G = E[\mathbf{f}^G|\mathbf{y}^\star]$.
   Spatial prediction onto the grid based on the synthetic observations .
5. $\mathbf{f}^\star = \hat{\boldsymbol{\mu}} + (\mathbf{f}^G - \hat{\mathbf{f}}^G)$ is a draw from the conditional distribution.

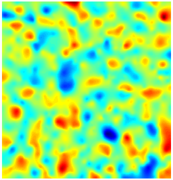
- Initial work was to approximate Step 2. (PART1)
- Once Step 2 was sped up Step 4 became the bottleneck. (PART2)

PART1

Approximate unconditional simulation

Approximate unconditional simulation

A GP process realization



- Simulate on $\mathcal{S}^{\mathcal{G}}$ by circulant embedding $\rightarrow \mathbf{f}^G$

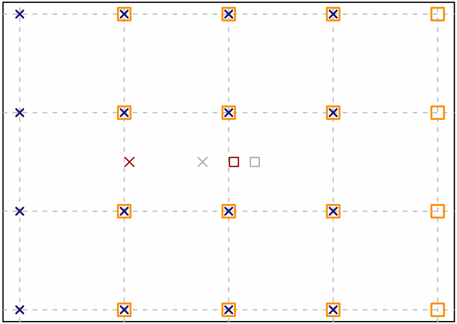
Requires that the circulant embedding gives a positive definite covariance function.

- Approximate conditional simulation of $\mathbf{f}^O$ given $\mathbf{f}^G$

Relies on local prediction and exploits the **screening** effect.

The details

Prediction using a 4×4 neighborhoods for some (four) observation locations.



- X and X generated conditional on X
- □ and □ generated conditional on □

- *Observations in separate boxes are made conditionally independent.*

Making this work

- Circulant embedding is so efficient one can generate very fine grids that separate most observation locations.

- With a stationary covariance the covariance matrix of the nearest grid locations is the same for all neighborhoods – Find its inverse once.

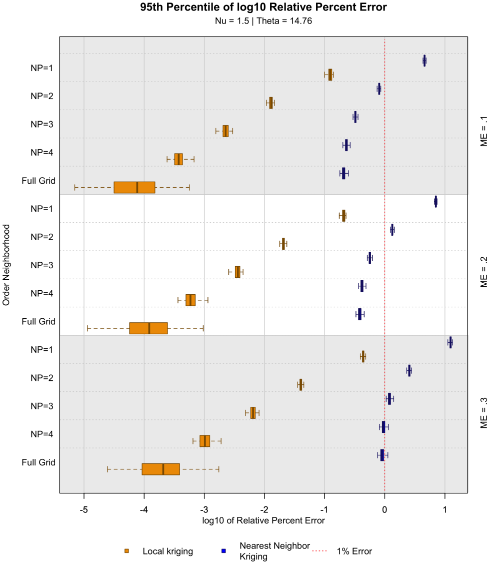
- Computing the (approximate) conditional mean for all observation locations is efficiently coded as a sparse matrix multiplication.

- Adding the extra variation to the conditional mean is on the order of the number of observations.

- Using a "model misspecification" calculation we show the approximate prediction SEs from this scheme are close to the exact ones.

Accuracy

Relative errors between exact and approximate prediction SEs.



PART2

Approximate fast spatial prediction

Prediction step

It is suprisingly simple
Prediction at a grid point.

$$\hat{f}(\mathbf{s}_k^G) = \sum_{i=1}^{n} k(\mathbf{s}_k^G, \mathbf{s}_i^O)\mathbf{c}_i$$

- Coefficient vector **c** found by Cholesky solves (Kriging) based on synthetic data.
- When there are many grid points this is time consuming.

Another (strange) Prediction Problem

Predict the off grid covariance from only on-grid covariances.

$$k(\mathbf{s}_k^G, \mathbf{s}_i^O) \approx \sum_{j=1}^{M} k(\mathbf{s}_k^G, \mathbf{s}_j^G) A_{j,i}$$

- Use only neighborhoods around $\mathbf{s}_i^O$ not the full grid, where A is a sparse matrix
- $A_{j,i}$ are already found from PART 1 for a completely different reason.

Amazing!

Approximate Prediction step

$$
\begin{aligned}
\hat{f}(\mathbf{s}_k^G) &= \sum_{i=1}^{n} k(\mathbf{s}_k^G, \mathbf{s}_i^O)\mathbf{c}_i \\
&\approx \sum_{i=1}^{n} \left[\sum_{j=1}^{M} k(\mathbf{s}_k^G, \mathbf{s}_j^G) A_{j,i} \right] \mathbf{c}_i = \sum_{j=1}^{M} k(\mathbf{s}_k^G, \mathbf{s}_j^G) \left[\sum_{i=1}^{n} A_{j,i} \mathbf{c}_i \right] \\
&= \sum_{j=1}^{M} k(\mathbf{s}_k^G, \mathbf{s}_j^G) \mathbf{c}_j^{\star}
\end{aligned}
$$

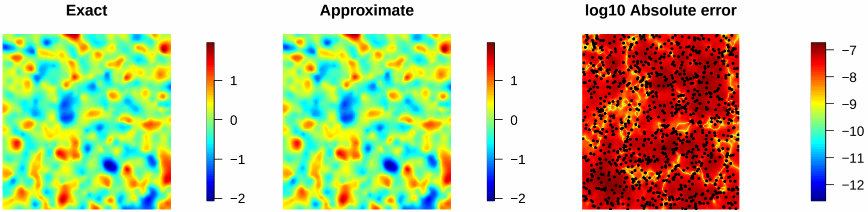
- $\mathbf{c}_j^{\star}$ these are now coefficients on the regular grid
- k is stationary – last sum is just a convolution.

We have approximated a huge and dense matrix multiplication by a sparse matrix multiplication and a convolution.

Does it work?

Matern: $\alpha = .08$ $\nu = 1.0$
$n = 500$, 300×300 grid, 8×8 neighborhood

Predicted fields



A factor of 25 speedup!

Approximation theory

Recall

$$k(\mathbf{s}_k^G, \mathbf{s}_i^O) \approx \sum_{j=1}^{M} k(\mathbf{s}_k^G, \mathbf{s}_j^G) A_{j,i}$$

or as functions.

$$k(., \mathbf{s}_i^O) \approx \sum_{j=1}^{M} k(., \mathbf{s}_j^G) A_{j,i}$$

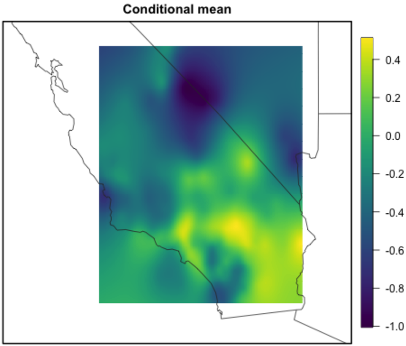
Let $\mathcal{H}$ be the reproducing kernel space for $k(.,.)$

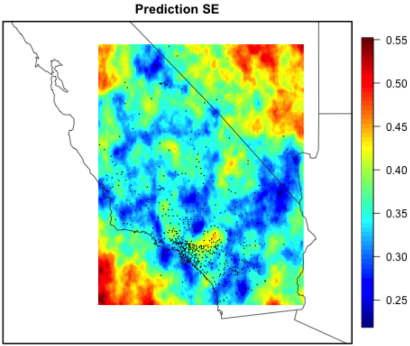
$$\| k(., \mathbf{s}_i^O) - \sum_{j=1}^{M} k(., \mathbf{s}_j^G) A_{j,i} \|_{\mathcal{H}}^2$$

- this measure of the error in the approximation is known as the *power function* in interpolation theory
- Key term in deriving general bounds on the interpolation error of any $f \in \mathcal{H}$ by points $\{\mathbf{s}^G\}$.

USGS ShakeMap

Approximation 50 member ensemble 300 $\times$ 300 grid $\rightarrow$ 30 seconds
Exact and single member 100 $\times$ 100 grid $\rightarrow$ 30 seconds





Thank you