Predictive Gaussian Processes

An alternative to process convolutions, also consisting on a representation of a Gaussian process over a grid, is Predictive Gaussian Processes (Banerjee et al. JRSSB '08).

Consider a set of points $U = (u_1, \ldots, u_p)$. Then we denote the Gaussian process w on U as $\boldsymbol{w}^* = (w(u_1), \ldots, w(u_p)) \sim N_p(0, C^*)$ where $C^*_{ij} = C(u_i, u_j)$, the covariance function that defines w.

Predictive Gaussian Processes

An alternative to process convolutions, also consisting on a representation of a Gaussian process over a grid, is Predictive Gaussian Processes (Banerjee et al. JRSSB '08).

Consider a set of points $U = (u_1, \ldots, u_p)$. Then we denote the Gaussian process w on U as $\boldsymbol{w}^* = (w(u_1), \ldots, w(u_p)) \sim N_p(0, C^*)$ where $C^*_{ij} = C(u_i, u_j)$, the covariance function that defines w.

The spatial kriging interpolator at location s is given as

$$\tilde{w}(s) = E(w(s)|w^*) = \boldsymbol{c}(s)^T (C^*)^{-1} \boldsymbol{w}^*$$

where $\boldsymbol{c}(s) = (C(s, u_1), \ldots, C(s, u_p))$.

Predictive Gaussian Process

The interpolator $\tilde{w}$ defines a new process indexed in s. Its covariance function is given as

$$\tilde{C}(s, s') = \boldsymbol{c}(s)^T (C^*)^{-1} \boldsymbol{c}(s')$$

so $\tilde{w}(s)$ is not a stationary process, even if the original process $w(s)$ is.

Predictive Gaussian Process

The interpolator $\tilde{w}$ defines a new process indexed in s. Its covariance function is given as

$$\tilde{C}(s, s') = \boldsymbol{c}(s)^T (C^*)^{-1} \boldsymbol{c}(s')$$

so $\tilde{w}(s)$ is not a stationary process, even if the original process $w(s)$ is.

If we consider a model

$$Y(s) = \boldsymbol{x}(s)^T \boldsymbol{\beta} + w(s) + \varepsilon(s)$$

we can substitute the Gaussian process $w(s)$ by $\tilde{w}(s)$, which produces an important dimension reduction, as the matrix inversions involved in dealing with $\tilde{w}$ are only $p \times p$ as opposed to $n \times n$.

Linear Representations

We consider representing a Gaussian spatial process as a linear combination of basis functions, not necessarily orthogonal

$$X(\boldsymbol{s}) = \sum_{j=1}^{m} B_j(\boldsymbol{s})\gamma_j = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{\gamma}\ , \quad \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T \sim N_m(0, \boldsymbol{K})$$

Linear Representations

We consider representing a Gaussian spatial process as a linear combination of basis functions, not necessarily orthogonal

$$X(\boldsymbol{s}) = \sum_{j=1}^{m} B_j(\boldsymbol{s})\gamma_j = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{\gamma}\ , \quad \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T \sim N_m(0, \boldsymbol{K})$$

The covariance function of $X(\boldsymbol{s})$ is given by

$$v(\boldsymbol{s}, \boldsymbol{s}') = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{K}\boldsymbol{B}(\boldsymbol{s}')$$

It is non-stationary and reduces the matrix computation burden but allows for very general classes of covariance functions.

Linear Representations

The linear representation allows for fixed rank Kriging (Cressie and Johannesson JRSSB '08).

Linear Representations

The linear representation allows for fixed rank Kriging (Cressie and Johannesson JRSSB '08).

The predictive process approach corresponds to taking

$$\boldsymbol{K} = \boldsymbol{C}^*,$$

and

$$B_j(\boldsymbol{s};\boldsymbol{\psi}) = [\boldsymbol{c}(\boldsymbol{s})^T(\boldsymbol{C}^*)^{-1}]_j$$

Linear Representations

The linear representation allows for fixed rank Kriging (Cressie and Johannesson JRSSB '08).

The predictive process approach corresponds to taking

$$\boldsymbol{K} = \boldsymbol{C}^*,$$

and

$$B_j(\boldsymbol{s};\boldsymbol{\psi}) = [\boldsymbol{c}(\boldsymbol{s})^T(\boldsymbol{C}^*)^{-1}]_j$$

The process convolution approach corresponds to letting

$$B_j(\boldsymbol{s};\boldsymbol{\psi}) = b(\boldsymbol{s}-\boldsymbol{u}_j;\boldsymbol{\psi})$$

for some kernel function $b(\cdot;\boldsymbol{\psi})$, and some matrix $\boldsymbol{K}$.

Fixed ranked kriging is thus the most general of the three approaches, allowing for a general basis and a general covariance structure for the coefficients. The predictive process approach uses one specification: the covariance of the process on the grid. It is the most parsimonious approach, but also the least flexible.

Properties of Predictive Processes

- $\tilde{w}$ is an interpolator of w on the grid. Thus $\tilde{w}(u_j) = w(u_j)$.

Properties of Predictive Processes

- $\tilde{w}$ is an interpolator of w on the grid. Thus $\tilde{w}(u_j) = w(u_j)$.
- $\tilde{w}(s)$ is the projection of $w(s)$ onto the Hilbert space generated by all linear combinations of the elements in w^* as well as their mean-square limits. The inner product is given by $E(w(s)w(s'))$.

Properties of Predictive Processes

- $\tilde{w}$ is an interpolator of w on the grid. Thus $\tilde{w}(u_j) = w(u_j)$.
- $\tilde{w}(s)$ is the projection of $w(s)$ onto the Hilbert space generated by all linear combinations of the elements in w^* as well as their mean-square limits. The inner product is given by $E(w(s)w(s'))$.
- As $\tilde{w}(s)$ is a conditional expectation

$$0 \leq \text{var}(\tilde{w}(s)) = \text{var}(w(s)|\boldsymbol{w}^*) = \text{var}(w(s)) - \boldsymbol{c}(s)^T (C^*)^{-1} \boldsymbol{c}(s)$$

implying that $\tilde{w}(s)$ underestimates the variability of $w(s)$. In other words, the variance of $\tilde{w}$ is maximized at the grid points.

Inversion of the Covariance Matrix

With the addition of uncorrelated noise to the linear representation model we have that the covariance matrix of the process can be written as

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{K}\boldsymbol{B}^T + \sigma^2\boldsymbol{V}$$

where $\boldsymbol{V}$ is a diagonal matrix. Using the Sherman-Morrison-Woodbury formulas we obtain the inverse of this matrix as

$$\boldsymbol{\Sigma}^{-1} = (\sigma^2\boldsymbol{V})^{-1} - (\sigma^2\boldsymbol{V})^{-1}\boldsymbol{B}\left(\boldsymbol{K}^{-1} + \boldsymbol{B}^T(\sigma^2\boldsymbol{V})^{-1}\boldsymbol{B}\right)\boldsymbol{B}^T(\sigma^2\boldsymbol{V})^{-1}.$$

This formula requires the inversion of an $m \times m$ covariance matrix, as well as an $n \times n$ diagonal matrix. This implies very strong computational savings.

Modified Predictive Process

In the conditionally linear representation, predictive processes have coefficients γ that correspond to the original process restricted to a grid. This is a convenient property, as it implies that γ has the same physical properties as the original process, which can be useful for modeling. Process convolutions have this property only for kernels with very tight support.

Modified Predictive Process

In the conditionally linear representation, predictive processes have coefficients $\boldsymbol{\gamma}$ that correspond to the original process restricted to a grid. This is a convenient property, as it implies that $\boldsymbol{\gamma}$ has the same physical properties as the original process, which can be useful for modeling. Process convolutions have this property only for kernels with very tight support.

To address the variance underestimation issue we consider

$$\hat{w}(s) = \tilde{w}(s) + \tilde{\varepsilon}(s), \quad \tilde{\varepsilon}(s) \sim N(0, \text{var}(w(s)) - \boldsymbol{c}(s)^T (C^*)^{-1} \boldsymbol{c}(s))$$

where the $\tilde{\varepsilon}(s)$ are uncorrelated. The former implies that $\hat{w}(s)$ has the same expectation as $\tilde{w}(s)$ and var$(\hat{w}(s))$ =var$(w(s))$.

Modified Predictive Process

The full model for the observables is then

$$Y(s) = \boldsymbol{x}(s)^T \boldsymbol{\beta} + \tilde{w}(s) + \tilde{\varepsilon}(s) + \varepsilon(s)$$

Thus, an additional source of non-homogeneous micro-scale variability is included to account for the variance underestimation.

The fact that only one covariance function is needed to obtain this model is a big plus. But the flip side is that only one input is used to determine the correlation between the grid points, the basis function and the nugget effect.

Uncorrelated Expansions

Given a conditionally linear representation

$$X(s) = \sum_{j=1}^{P} B_j(s)\gamma_j$$

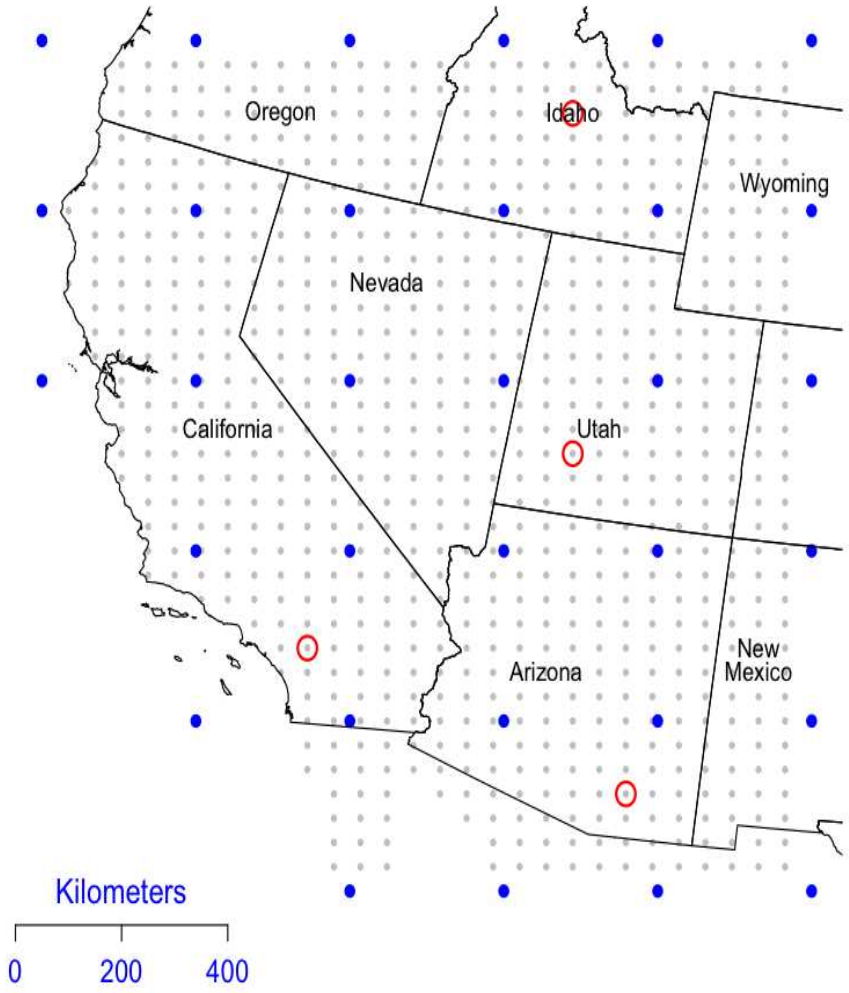
it is possible to obtain a representation that is not orthogonal, but has uncorrelated coefficients.

Let $K = P\Lambda P^T$, where P is orthogonal and Λ is diagonal with elements λ_j. Let $\alpha \sim N_p(0, \Lambda)$, then we can write $\gamma = P\alpha$. Thus

$$X(s) = B(s)^T\gamma = (B(s)^T P)\alpha = \psi(s)^T\alpha = \sum_{j=1}^{p} \psi_j(s)\alpha_j$$

The coefficients α_j are independent. We can order them according to λ_j and obtain the factors $\psi_j(s)$ in order of variability.

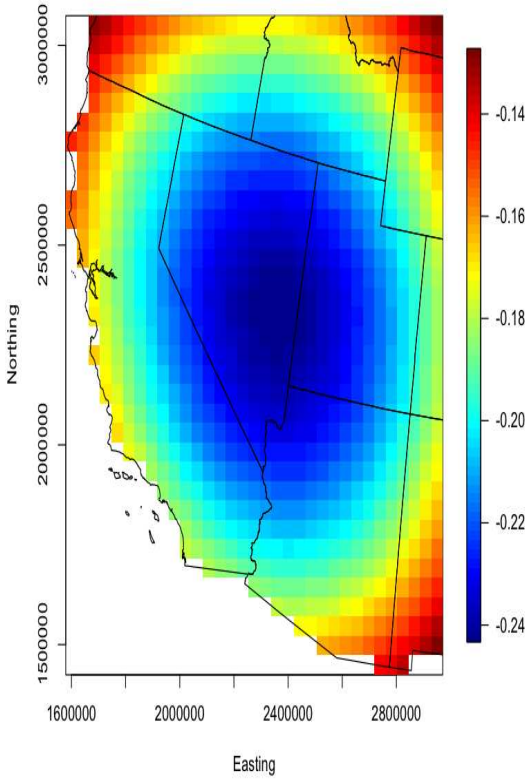
Predictive Processes for RCM Output



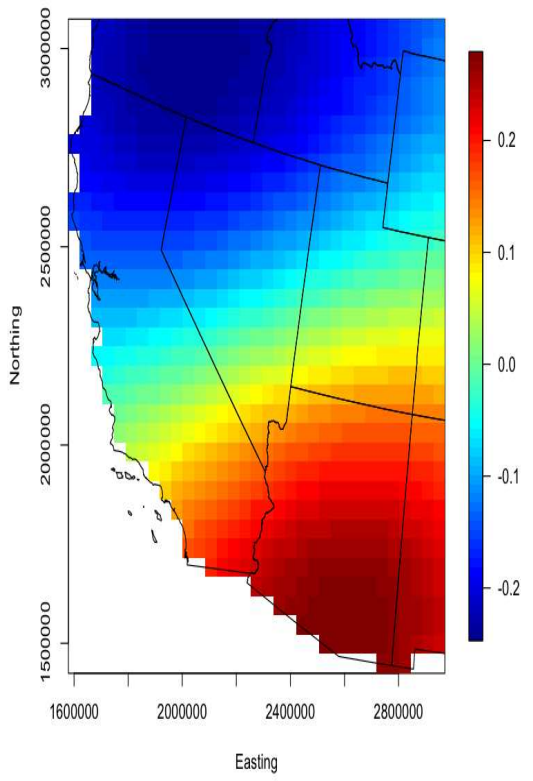
We consider output from Regional Climate Models consisting on a grid of 50 Km resolution. These simulations are taken from the North American Regional Climate Change Assessment Program (NARCCAP). In this case, it is useful to estimate the main modes of spatial variability. We Use a Predictive process with exponential correlation.

First Three Factors

$\Psi_{(1)}$



$\Psi_{(2)}$



$\Psi_{(3)}$