Gaussian Process Regression with Noisy Inputs

Dan Cervone

Harvard Statistics Department

March 3, 2015

A smooth response x over a surface $\mathbb{S} \subset \mathbb{R}^p$.

- For $s_1, \dots, s_n \in \mathbb{S}$,
$$\begin{pmatrix} x(s_1) \\ \vdots \\ x(s_n) \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n))$$
- $[\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)]_{ij} = c(s_i, s_j)$, where c is the *covariance function*.

A smooth response x over a surface $\mathbb{S} \subset \mathbb{R}^p$.

- For $s_1, \ldots, s_n \in \mathbb{S}$,

$$\begin{pmatrix} x(s_1) \\ \vdots \\ x(s_n) \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n))$$

- $[\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)]_{ij} = c(s_i, s_j)$, where c is the *covariance function*.

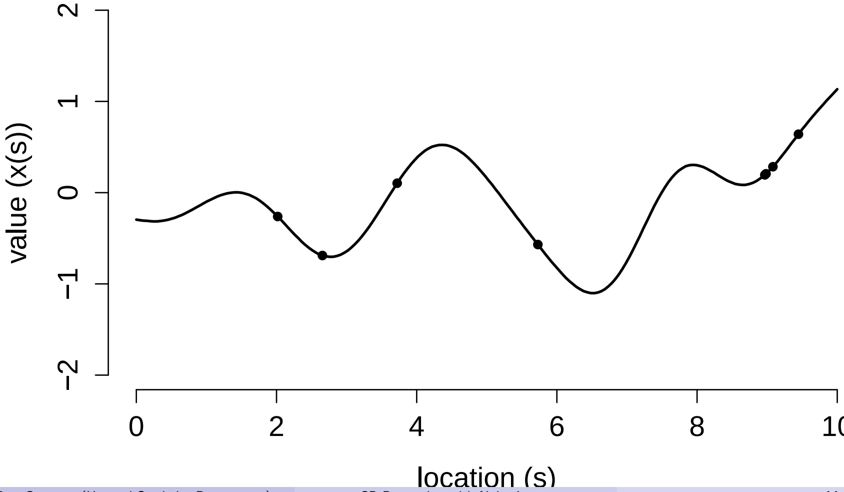
Interpolation/prediction at unobserved locations in input space

- Observe $\mathbf{x}_n = (x(s_1) \ \ldots \ x(s_n))'$.
- Predict $\mathbf{x}_k^* = (x(s_1^*) \ \ldots \ x(s_k^*))'$.

$$\mathbf{x}_k^* | \mathbf{x}_n \sim \mathcal{N}\big(\mathbf{C}(\mathbf{s}_k^*, \mathbf{s}_n)\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{x}_n, \\ \mathbf{C}(\mathbf{s}_k^*, \mathbf{s}_k^*) - \mathbf{C}(\mathbf{s}_k^*, \mathbf{s}_n)\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{C}(\mathbf{s}_n, \mathbf{s}_k^*)\big)$$

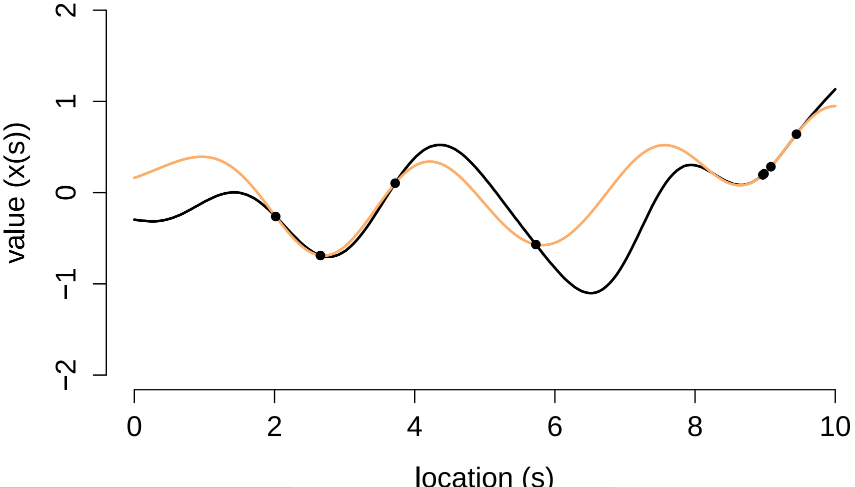
Gaussian process regression

Example



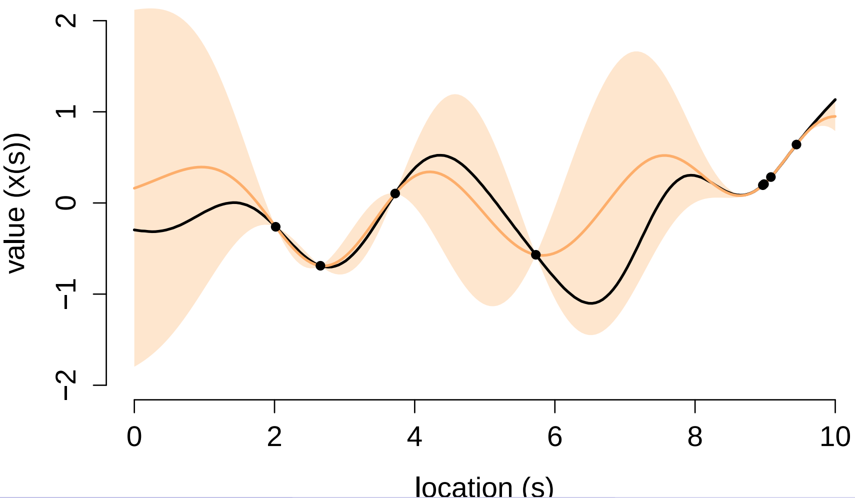
Gaussian process regression

Example



Gaussian process regression

Example



GPs with noisy inputs

Scientific examples

Location error model

Instead of observing x, we observe the process $y(s) = x(s + u)$, where $u \sim g(u)$ are errors in the input space $\mathbb{S}$.

Note:

- We observe $\mathbf{s}_n$, $\mathbf{y}_n$, but wish to predict $x(s^*)$.
- Note: y is never a GP.

Location errors (e.g. geocoding error, map positional error) is a problem in many scientific domains.

- Epidemiology [3, 10, 2].
- Environmental sciences [1, 16].
- Object tracking/computer vision [9, 15].

Measurement error

GP location errors vs errors-in-variables

GP input/location errors:

- $y(s) = x(s + u) + \epsilon.$

Measurement error

GP location errors vs errors-in-variables

GP input/location errors:

- $y(s) = x(s + u) + \epsilon$.

Traditional errors-in-variables model [5]:

- $x^* = f_\theta(x) + \epsilon$.
- $x(s^*) = f_{\theta,\mathbf{s}_n}(\mathbf{x}_n) + \epsilon$ (GP regression).

Measurement error

GP location errors vs errors-in-variables

GP input/location errors:

- $y(s) = x(s + u) + \epsilon$.

Traditional errors-in-variables model [5]:

- $x^* = f_\theta(x) + \epsilon$.
- $x(s^*) = f_{\theta,\mathbf{s}_n}(\mathbf{x}_n) + \epsilon$ (GP regression).
- Observe $y = x + \eta$, ie $\mathbf{y}_n = \mathbf{x}_n + \boldsymbol{\eta}_n$.
- Common to assume $\eta \perp x$ (classical) or $\eta \perp y$ (Berkson).

Measurement error

GP location errors vs errors-in-variables

GP input/location errors:

- $y(s) = x(s + u) + \epsilon$.

Traditional errors-in-variables model [5]:

- $x^* = f_\theta(x) + \epsilon$.
- $x(s^*) = f_{\theta,\mathbf{s}_n}(\mathbf{x}_n) + \epsilon$ (GP regression).
- Observe $y = x + \eta$, ie $\mathbf{y}_n = \mathbf{x}_n + \boldsymbol{\eta}_n$.
- Common to assume $\eta \perp x$ (classical) or $\eta \perp y$ (Berkson).

GP input errors does not yield a traditional errors-in-variables regression problem:

- Errors $y(s) - x(s)$ depend on $x(s)$.
- True regression function is unknown: $x(s^*) = f_{\theta,\mathbf{s}_n+\mathbf{u}_n}(\mathbf{y}_n) + \epsilon$.

Measurement error

Methodology

Methods to properly accounting for noisy inputs are essential for reliable inference in this regime.
We seek:

- Optimal (MSE) point prediction, and interval predictions with correct coverage.
- Consistent/efficient parameter estimation.
- The location-error regime can actually deliver more precise predictions than the error-free regime.

Measurement error

Methodology

Methods to properly accounting for noisy inputs are essential for reliable inference in this regime.

We seek:

- Optimal (MSE) point prediction, and interval predictions with correct coverage.
- Consistent/efficient parameter estimation.
- The location-error regime can actually deliver more precise predictions than the error-free regime.

We discuss three methods:

Measurement error

Methodology

Methods to properly accounting for noisy inputs are essential for reliable inference in this regime.

We seek:

- Optimal (MSE) point prediction, and interval predictions with correct coverage.
- Consistent/efficient parameter estimation.
- The location-error regime can actually deliver more precise predictions than the error-free regime.

We discuss three methods:

- Ignoring location errors.

Measurement error

Methodology

Methods to properly accounting for noisy inputs are essential for reliable inference in this regime.
We seek:

- Optimal (MSE) point prediction, and interval predictions with correct coverage.
- Consistent/efficient parameter estimation.
- The location-error regime can actually deliver more precise predictions than the error-free regime.

We discuss three methods:

- Ignoring location errors.
- Kriging (BLUP), using moment properties of error-induced process y.

Measurement error

Methodology

Methods to properly accounting for noisy inputs are essential for reliable inference in this regime.
We seek:

- Optimal (MSE) point prediction, and interval predictions with correct coverage.
- Consistent/efficient parameter estimation.
- The location-error regime can actually deliver more precise predictions than the error-free regime.

We discuss three methods:

- Ignoring location errors.
- Kriging (BLUP), using moment properties of error-induced process y.
- MCMC on the space $(\mathbf{x}_k^*, \mathbf{u}_n)$.

Ignoring location errors

Sometimes, you can get lucky

Analyst just assumes $\mathbf{y}_n = \mathbf{x}_n$:

- "Kriging Ignoring Location Errors" (KILE) [6]:

$$\hat{x}_{\text{KILE}}(s^*) = \mathbf{C}(s^*, \mathbf{s}_n)\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n.$$

- Parameter inference based on assuming $\mathbf{y}_n = \mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n))$.

Ignoring location errors

Sometimes, you can get lucky

Analyst just assumes $\mathbf{y}_n = \mathbf{x}_n$:

- "Kriging Ignoring Location Errors" (KILE) [6]:

$$\hat{x}_{\text{KILE}}(s^*) = \mathbf{C}(s^*, \mathbf{s}_n)\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n.$$

- Parameter inference based on assuming $\mathbf{y}_n = \mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n))$.

Example: $c(s_1, s_2) = \exp(-(s_1 - s_2)^2)$, and $u \sim \mathcal{N}(0, \sigma_u^2)$.

Ignoring location errors

Sometimes, you can get lucky

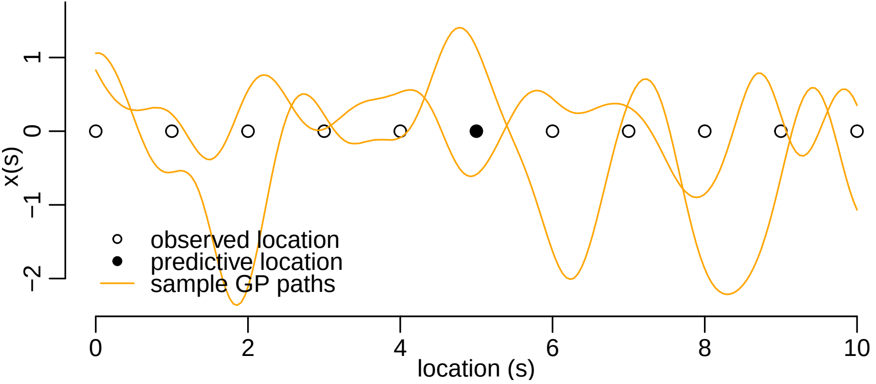
Analyst just assumes $\mathbf{y}_n = \mathbf{x}_n$:

- "Kriging Ignoring Location Errors" (KILE) [6]:

$$\hat{x}_{\mathrm{KILE}}(s^*) = \mathbf{C}(s^*, \mathbf{s}_n)\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n.$$

- Parameter inference based on assuming $\mathbf{y}_n = \mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n))$.

Example: $c(s_1, s_2) = \exp(-(s_1 - s_2)^2)$, and $u \sim \mathcal{N}(0, \sigma_u^2)$.



Ignoring location errors

Sometimes, you can get lucky

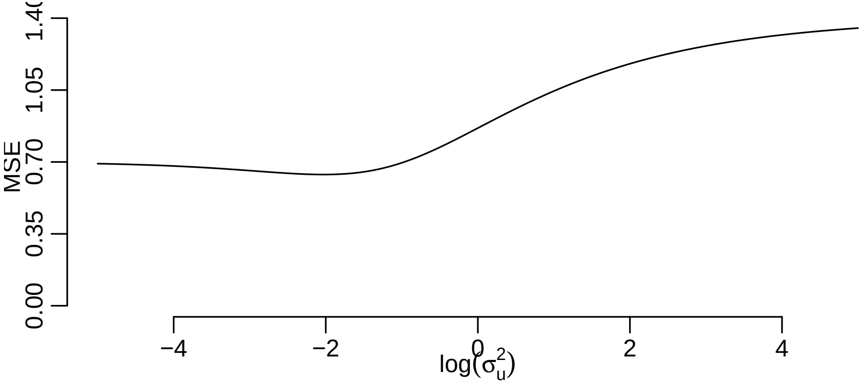
Analyst just assumes $\mathbf{y}_n = \mathbf{x}_n$:

- "Kriging Ignoring Location Errors" (KILE) [6]:

$$\hat{x}_{\text{KILE}}(s^*) = \mathbf{C}(s^*, \mathbf{s}_n)\mathbf{C}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n.$$

- Parameter inference based on assuming $\mathbf{y}_n = \mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{s}_n, \mathbf{s}_n))$.

Example: $c(s_1, s_2) = \exp(-(s_1 - s_2)^2)$, and $u \sim \mathcal{N}(0, \sigma_u^2)$.



Ignoring location errors

Sometimes, disaster strikes

Assuming known covariance funciton, KILE is **not** a self-efficient procedure.

- *Self-efficiency* [12]: estimator cannot be improved by removing/subsampling data.

Ignoring location errors

Sometimes, disaster strikes

Assuming known covariance funciton, KILE is **not** a self-efficient procedure.

- *Self-efficiency* [12]: estimator cannot be improved by removing/subsampling data.

Theorem

Assume covariance function c and error model $u \sim g(u)$ satisfy regularity conditions. Let $\hat{x}^n_{\text{KILE}}(s^*)$ be the KILE estimator for $x(s^*)$ given $\mathbf{x}_n$. Then for any $\mathbf{s}_n$ and s^*, there exists s_{n+1} such that

$$\mathbb{E}[(x(s^*) - \hat{x}^{n+1}_{\text{KILE}}(s^*))^2] \geq \mathbb{E}[(x(s^*) - \hat{x}^n_{\text{KILE}}(s^*))^2].$$

Ignoring location errors

Sometimes, disaster strikes

Assuming known covariance funciton, KILE is **not** a self-efficient procedure.

- *Self-efficiency* [12]: estimator cannot be improved by removing/subsampling data.

Theorem

Assume covariance function c and error model $u \sim g(u)$ satisfy regularity conditions. Let $\hat{x}^n_{\mathrm{KILE}}(s^*)$ be the KILE estimator for $x(s^*)$ given $\mathbf{x}_n$. Then for any $\mathbf{s}_n$ and s^*, there exists s_{n+1} such that

$$\mathbb{E}[(x(s^*) - \hat{x}^{n+1}_{\mathrm{KILE}}(s^*))^2] \geq \mathbb{E}[(x(s^*) - \hat{x}^n_{\mathrm{KILE}}(s^*))^2].$$

Regularity conditions:

- c twice differentiable everywhere.
- $k(s_1, s_2) = \mathbb{E}[c(s_1 + u_1, s_2 + u_2)]$ twice differentiable everywhere except $s_1 = s_2$.

Ignoring location errors

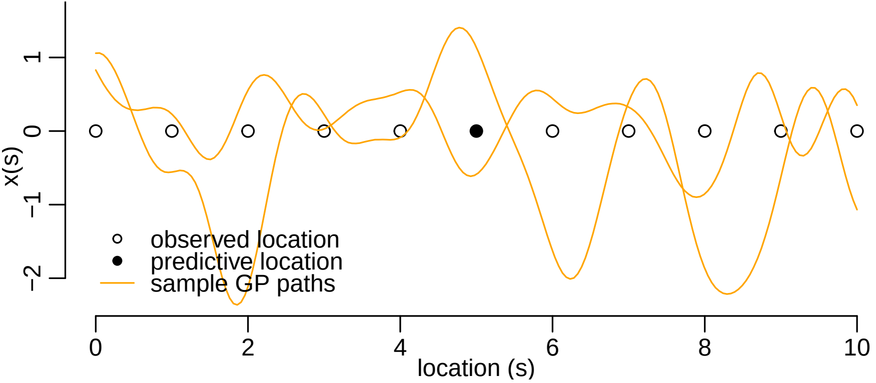
Sometimes, disaster strikes

Example: $c(s_1, s_2) = \exp(-(s_1 - s_2)^2)$, and $u \sim \mathcal{N}(0, 0.04)$.

Ignoring location errors

Sometimes, disaster strikes

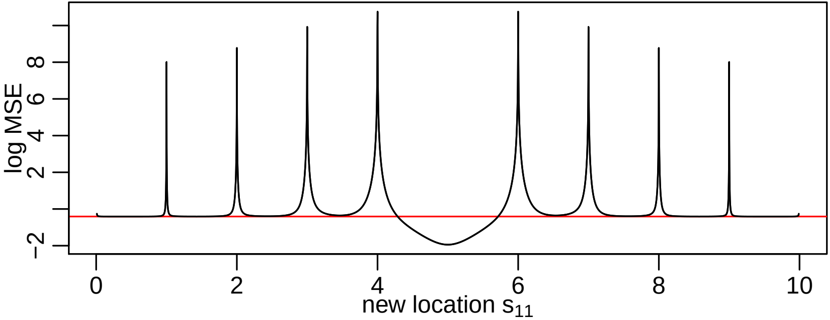
Example: $c(s_1, s_2) = \exp(-(s_1 - s_2)^2)$, and $u \sim \mathcal{N}(0, 0.04)$.



Ignoring location errors

Sometimes, disaster strikes

Example: $c(s_1, s_2) = \exp(-(s_1 - s_2)^2)$, and $u \sim \mathcal{N}(0, 0.04)$.



Kriging (Best Linear Unbiased Prediction)

KALE [6]

Second moment properties of $y(s)$ and $(x(s^*), y(s))$:

$$\begin{aligned} k(s_1, s_2) &= \mathbb{C}[y(s_1), y(s_2)] = \mathbb{E}[c(s_1 + u_1, s_2 + u_2)] \text{ for } s_1 \neq s_2 \\ k(s, s) &= \mathbb{C}[y(s), y(s)] \quad = \mathbb{E}[c(s + u, s + u)] \\ k^*(s, s^*) &= \mathbb{C}[y(s), x(s^*)] \; = \mathbb{E}[c(s + u, s^*)]. \end{aligned}$$

Kriging (Best Linear Unbiased Prediction)

KALE [6]

Second moment properties of $y(s)$ and $(x(s^*), y(s))$:

$$\begin{aligned} k(s_1, s_2) &= \mathbb{C}[y(s_1), y(s_2)] = \mathbb{E}[c(s_1 + u_1, s_2 + u_2)] \text{ for } s_1 \neq s_2 \\ k(s, s) &= \mathbb{C}[y(s), y(s)] \quad = \mathbb{E}[c(s + u, s + u)] \\ k^*(s, s^*) &= \mathbb{C}[y(s), x(s^*)] \ = \mathbb{E}[c(s + u, s^*)]. \end{aligned}$$

k is the covariance function for y, and we can use it for Kriging adjusting for location error (KALE) [6]:

$$\hat{x}_{\mathrm{KALE}}(s^*) = \mathbf{K}^*(s^*, \mathbf{s}_n)\mathbf{K}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n.$$

Kriging (Best Linear Unbiased Prediction)

KALE [6]

Second moment properties of $y(s)$ and $(x(s^*), y(s))$:

$$\begin{aligned}
k(s_1, s_2) &= \mathbb{C}[y(s_1), y(s_2)] = \mathbb{E}[c(s_1 + u_1, s_2 + u_2)] \text{ for } s_1 \neq s_2 \\
k(s, s) &= \mathbb{C}[y(s), y(s)] \quad = \mathbb{E}[c(s + u, s + u)] \\
k^*(s, s^*) &= \mathbb{C}[y(s), x(s^*)] \; = \mathbb{E}[c(s + u, s^*)].
\end{aligned}$$

k is the covariance function for y, and we can use it for Kriging adjusting for location error (KALE) [6]:

$$\hat{x}_{\mathrm{KALE}}(s^*) = \mathbf{K}^*(s^*, \mathbf{s}_n)\mathbf{K}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n.$$

- For any error structure u, k is a valid covariance function if and only if c is.
- If c is known, then KALE dominates KILE in MSE.

Sometimes, k is available in closed form:

- Example: for $c(s_1, s_2) = \tau^2 \exp(-\beta||s_1 - s_2||^2)$ and $u_i \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_p)$,

$$k(s_1, s_2) = \frac{\tau^2}{(1 + 4\beta\sigma_u^2)^{p/2}} \exp\left(-\frac{\beta}{1 + 4\beta\sigma_u^2}||s_1 - s_2||^2\right).$$

- Not generally true that c and k have same functional form.

Sometimes, k is available in closed form:

- Example: for $c(s_1, s_2) = \tau^2 \exp(-\beta||s_1 - s_2||^2)$ and $u_i \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_p)$,

$$k(s_1, s_2) = \frac{\tau^2}{(1 + 4\beta\sigma_u^2)^{p/2}} \exp\left(-\frac{\beta}{1 + 4\beta\sigma_u^2}||s_1 - s_2||^2\right).$$

- Not generally true that c and k have same functional form.

Most commonly, k is computed by Monte Carlo.

$$k(s_1, s_2) \approx \frac{1}{M} \sum_{i=1}^{M} c(s_1 + u_{1i}, s_2 + u_{2i})$$

- $u_{ji} \stackrel{iid}{\sim} g(u_j)$ for $i = 1, \dots, M$.

Kriging (Best Linear Unbiased Prediction)

Interval estimation

We get interval estimates for KALE by deriving the distribution function of prediction errors:

Proposition

Let

$$W(\mathbf{u}_n) = \mathbb{V}[x(s^*)] + \gamma'\mathbf{C}(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)\gamma - 2\gamma'\mathbf{C}(\mathbf{s}_n + \mathbf{u}_n, s^*)$$

$$\text{where } \gamma = \mathbf{K}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{K}^*(\mathbf{s}_n, s^*).$$

Then

$$\mathbb{P}(x(s^*) - \hat{x}_{\text{KALE}}(s^*) < z) = \mathbb{E}\left[\Phi\left(\frac{z}{\sqrt{W(\mathbf{u}_n)}}\right)\right],$$

where Φ *is the standard normal distribution function.*

We get interval estimates for KALE by deriving the distribution function of prediction errors:

Proposition

Let

$$W(\mathbf{u}_n) = \mathbb{V}[x(s^*)] + \gamma' \mathbf{C}(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)\gamma - 2\gamma' \mathbf{C}(\mathbf{s}_n + \mathbf{u}_n, s^*)$$

$$\textit{where } \gamma = \mathbf{K}(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{K}^*(\mathbf{s}_n, s^*).$$

Then

$$\mathbb{P}(x(s^*) - \hat{x}_{\text{KALE}}(s^*) < z) = \mathbb{E}\left[\Phi\left(\frac{z}{\sqrt{W(\mathbf{u}_n)}}\right)\right],$$

where Φ is the standard normal distribution function.

- These yield *confidence intervals*, not conditional probability intervals.

Inferring parameters of covariance function:

- Likelihood:

$$L(\theta; \mathbf{y}_n) \propto \int |\mathbf{C}_\theta(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)| \exp\left(-\frac{1}{2}\mathbf{y}_n' \mathbf{C}_\theta(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)^{-1}\mathbf{y}_n\right) d\mathbf{u}_n.$$

    - Stochastic EM.

Inferring parameters of covariance function:

- Likelihood:

$$L(\theta; \mathbf{y}_n) \propto \int |\mathbf{C}_\theta(\mathbf{s}_n+\mathbf{u}_n, \mathbf{s}_n+\mathbf{u}_n)| \exp\left(-\frac{1}{2}\mathbf{y}_n' \mathbf{C}_\theta(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)^{-1}\mathbf{y}_n\right) d\mathbf{u}_n.$$

  - Stochastic EM.

- Pseudo-likelihood, based on Gaussian approximation to first two moments [6, 5]:

$$\tilde{L}(\theta; \mathbf{y}_n) \propto |\mathbf{K}_\theta(\mathbf{s}_n, \mathbf{s}_n)|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{y}_n' \mathbf{K}_\theta(\mathbf{s}_n, \mathbf{s}_n)^{-1}\mathbf{y}_n\right).$$

  - Pseudo-score is an unbiased estimating equation.
  - Maximum pseudo-likelihood estimator is asymptotically normal under proper domain conditions.

MCMC

$\mathbf{y}_n$ contains information about location errors $\mathbf{u}_n$:

$$\begin{aligned}\hat{x}(s^*) &= \mathbb{E}[x(s^*)|\mathbf{y}_n] \\ &= \int \left(\mathbf{C}(s^*, \mathbf{s}_n + \mathbf{u}_n)[\mathbf{C}(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)]^{-1}\mathbf{y}_n\right) \pi(\mathbf{u}_n|\mathbf{y}_n) d\mathbf{u}_n.\end{aligned}$$

MCMC

$\mathbf{y}_n$ contains information about location errors $\mathbf{u}_n$:

$$\begin{aligned}\hat{x}(s^*) &= \mathbb{E}[x(s^*)|\mathbf{y}_n] \\ &= \int \left(\mathbf{C}(s^*, \mathbf{s}_n + \mathbf{u}_n)[\mathbf{C}(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)]^{-1}\mathbf{y}_n\right) \pi(\mathbf{u}_n|\mathbf{y}_n) d\mathbf{u}_n.\end{aligned}$$

- Dominates KALE in MSE.
- $x(s^*)|\mathbf{y}_n$ yields conditional probability intervals.
- Naturally incorporates parameter estimation/uncertainty.

MCMC

Distributional assumptions

KALE/KILE does not require Gaussian assumption for x to obtain point predictions, but MCMC procedure does.

MCMC

Distributional assumptions

KALE/KILE does not require Gaussian assumption for x to obtain point predictions, but MCMC procedure does.

- Similar to arguments in favor of conjugate priors when using squared error loss [14], it makes sense to assume normality in x when Kriging.

MCMC

Distributional assumptions

KALE/KILE does not require Gaussian assumption for x to obtain point predictions, but MCMC procedure does.

- Similar to arguments in favor of conjugate priors when using squared error loss [14], it makes sense to assume normality in x when Kriging.
- Let $\Pi_{\mathbf{0},\mathbf{C}}$ be the family of joint distributions for $\mathbf{x}_n$ with first two moments $\mathbf{0}, \mathbf{C}$.
- For $\pi_1, \pi_2 \in \Pi_{\mathbf{0},\mathbf{C}}$, let

$$R_{\pi_1}(\pi_2) = \mathbb{E}_{\pi_1}[(\mathbb{E}_{\pi_2}[x(s^*)|\mathbf{x}_n] - x(s^*))^2].$$

MCMC

Distributional assumptions

KALE/KILE does not require Gaussian assumption for x to obtain point predictions, but MCMC procedure does.

- Similar to arguments in favor of conjugate priors when using squared error loss [14], it makes sense to assume normality in x when Kriging.
- Let $\Pi_{\mathbf{0},\mathbf{C}}$ be the family of joint distributions for $\mathbf{x}_n$ with first two moments $\mathbf{0}, \mathbf{C}$.
- For $\pi_1, \pi_2 \in \Pi_{\mathbf{0},\mathbf{C}}$, let

$$R_{\pi_1}(\pi_2) = \mathbb{E}_{\pi_1}[(\mathbb{E}_{\pi_2}[x(s^*)|\mathbf{x}_n] - x(s^*))^2].$$

If π_0 is Gaussian, then for all $\pi \in \Pi_{\mathbf{0},\mathbf{C}}$,

$$R_{\pi}(\pi) \leq R_{\pi}(\pi_0) = R_{\pi_0}(\pi_0) \leq R_{\pi_0}(\pi).$$

KALE/KILE does not require Gaussian assumption for x to obtain point predictions, but MCMC procedure does.

- Similar to arguments in favor of conjugate priors when using squared error loss [14], it makes sense to assume normality in x when Kriging.
- Let $\Pi_{\mathbf{0},\mathbf{C}}$ be the family of joint distributions for $\mathbf{x}_n$ with first two moments $\mathbf{0}, \mathbf{C}$.
- For $\pi_1, \pi_2 \in \Pi_{\mathbf{0},\mathbf{C}}$, let

$$R_{\pi_1}(\pi_2) = \mathbb{E}_{\pi_1}[(\mathbb{E}_{\pi_2}[x(s^*)|\mathbf{x}_n] - x(s^*))^2].$$

If π_0 is Gaussian, then for all $\pi \in \Pi_{\mathbf{0},\mathbf{C}}$,

$$R_{\pi}(\pi) \leq R_{\pi}(\pi_0) = R_{\pi_0}(\pi_0) \leq R_{\pi_0}(\pi).$$

- $R_{\pi_0}(\pi) - R_{\pi_0}(\pi_0)$ is the cost of incorrectly assuming π when x is Gaussian.
- $R_{\pi}(\pi_0) - R_{\pi}(\pi)$ is the *opportunity cost* of a Gaussian assumption.

This problem favors gradient-based MCMC samplers (HMC, MALA):

$$\log(\pi(\theta, \mathbf{u}_n|\mathbf{y}_n)) = -\frac{1}{2}\log(|\mathbf{C}_\theta(\mathbf{u}_n)|) - \frac{1}{2}\mathbf{y}_n'\mathbf{C}_\theta(\mathbf{u}_n)^{-1}\mathbf{y}_n + \text{const.}$$

$$\frac{\partial}{\partial u_i}\log(\pi(\theta, \mathbf{u}_n|\mathbf{y}_n)) =$$

$$\frac{1}{2}\text{Tr}\left(\mathbf{C}_\theta(\mathbf{u}_n)^{-1}\left[\frac{\partial}{\partial u_i}\mathbf{C}_\theta(\mathbf{u}_n)\right]\left(\mathbf{C}_\theta(\mathbf{u}_n)^{-1}\mathbf{y}_n\mathbf{y}_n' - \mathbf{I}_n\right)\right) + \frac{\partial}{\partial u_i}\log(\pi(\mathbf{u}_n))$$

$$\frac{\partial}{\partial \theta_i}\log(\pi(\theta, \mathbf{u}_n|\mathbf{y}_n)) =$$

$$\frac{1}{2}\text{Tr}\left(\mathbf{C}_\theta(\mathbf{u}_n)^{-1}\left[\frac{\partial}{\partial \theta_i}\mathbf{C}_\theta(\mathbf{u}_n)\right]\left(\mathbf{C}_\theta(\mathbf{u}_n)^{-1}\mathbf{y}_n\mathbf{y}_n' - \mathbf{I}_n\right)\right) + \frac{\partial}{\partial \theta_i}\log(\pi(\theta))$$

where $\mathbf{C}_\theta(\mathbf{u}_n) = \mathbf{C}_\theta(\mathbf{s}_n + \mathbf{u}_n, \mathbf{s}_n + \mathbf{u}_n)$.

- Computational complexity of both log-likelihood and gradient dominated by $\mathbf{C}_\theta(\mathbf{u}_n)^{-1}$.

Multimodality is a common problem.

- In error-free regime, likelihood for θ can be multimodal [20].
- In isotropic model with location errors $\mathbf{u}_n$, $\pi(\mathbf{y}_n|\mathbf{u}_n, \theta)$ constant for $\mathbf{u}_n$ across contours preserving pairwise distances.

MCMC
Multimodality

Multimodality is a common problem.

- In error-free regime, likelihood for θ can be multimodal [20].
- In isotropic model with location errors $\mathbf{u}_n$, $\pi(\mathbf{y}_n|\mathbf{u}_n, \theta)$ constant for $\mathbf{u}_n$ across contours preserving pairwise distances.

Example of isolated modes:

- $n = 2$, $p = 1$.
- $c(s_1, s_2) = \exp(-(s_1 - s_2)^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- $u_i \sim \mathcal{N}(0, \sigma_u^2)$.

MCMC

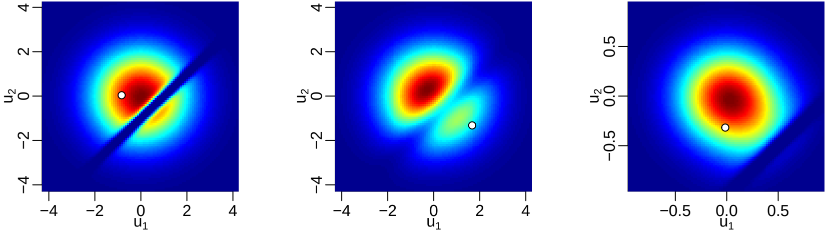
Multimodality

Multimodality is a common problem.

- In error-free regime, likelihood for θ can be multimodal [20].
- In isotropic model with location errors $\mathbf{u}_n$, $\pi(\mathbf{y}_n|\mathbf{u}_n, \theta)$ constant for $\mathbf{u}_n$ across contours preserving pairwise distances.

Example of isolated modes:

- $n = 2$, $p = 1$.
- $c(s_1, s_2) = \exp(-(s_1 - s_2)^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- $u_i \sim \mathcal{N}(0, \sigma_u^2)$.



$\sigma_u^2 = 2$, $\sigma_x^2 = 0.0001$ $\sigma_u^2 = 2$, $\sigma_x^2 = 1$ $\sigma_u^2 = 0.1$, $\sigma_x^2 = 0.0001$

Simulation study

Simulation study compares

- data: $c(s_1, s_2) = \tau^2 \exp(-\beta ||s_1 - s_2||^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- location errors: $u_i \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_2)$.
- methods: KILE, KALE, HMC.
- tasks: parameter inference, point prediction, interval prediction.
- scenarios: parameters assumed known, parameters first estimated.

Simulation study

Simulation study compares

- data: $c(s_1, s_2) = \tau^2 \exp(-\beta||s_1 - s_2||^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- location errors: $u_i \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_2)$.
- methods: KILE, KALE, HMC.
- tasks: parameter inference, point prediction, interval prediction.
- scenarios: parameters assumed known, parameters first estimated.

Parameter	Values used
τ^2	1
β	0.001, 0.01, 0.1, 0.5, 1, 2
σ_x^2	0.0001, 0.01, 0.1, 0.5, 1
σ_u^2	0.0001, 0.01, 0.1, 0.5, 1

Parameter values used in simulation study.

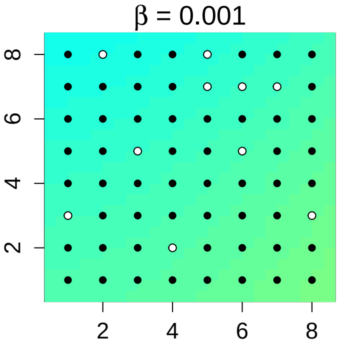
Simulation study

Simulation study compares

- data: $c(s_1, s_2) = \tau^2 \exp(-\beta||s_1 - s_2||^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- location errors: $u_i \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_2)$.
- methods: KILE, KALE, HMC.
- tasks: parameter inference, point prediction, interval prediction.
- scenarios: parameters assumed known, parameters first estimated.

Parameter	Values used
τ^2	1
β	0.001, 0.01, 0.1, 0.5, 1, 2
σ_x^2	0.0001, 0.01, 0.1, 0.5, 1
σ_u^2	0.0001, 0.01, 0.1, 0.5, 1

Parameter values used in simulation study.



- black = observed locations; white = predicted locations.

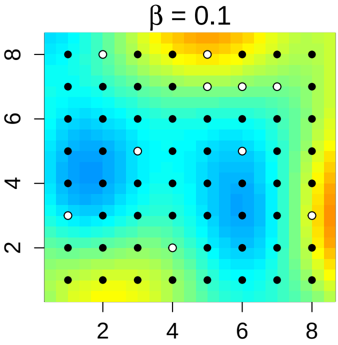
Simulation study

Simulation study compares

- data: $c(s_1, s_2) = \tau^2 \exp(-\beta||s_1 - s_2||^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- location errors: $u_i \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_2)$.
- methods: KILE, KALE, HMC.
- tasks: parameter inference, point prediction, interval prediction.
- scenarios: parameters assumed known, parameters first estimated.

Parameter	Values used
τ^2	1
β	0.001, 0.01, 0.1, 0.5, 1, 2
σ_x^2	0.0001, 0.01, 0.1, 0.5, 1
σ_u^2	0.0001, 0.01, 0.1, 0.5, 1

Parameter values used in simulation study.



- black = observed locations; white = predicted locations.

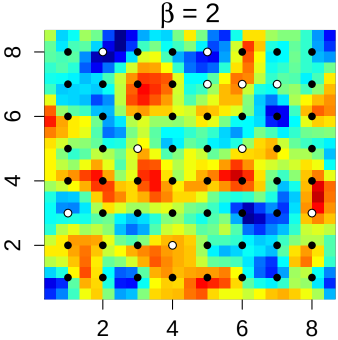
Simulation study

Simulation study compares

- data: $c(s_1, s_2) = \tau^2 \exp(-\beta||s_1 - s_2||^2) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$.
- location errors: $u_i \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_2)$.
- methods: KILE, KALE, HMC.
- tasks: parameter inference, point prediction, interval prediction.
- scenarios: parameters assumed known, parameters first estimated.

Parameter	Values used
τ^2	1
β	0.001, 0.01, 0.1, 0.5, 1, 2
σ_x^2	0.0001, 0.01, 0.1, 0.5, 1
σ_u^2	0.0001, 0.01, 0.1, 0.5, 1

Parameter values used in simulation study.

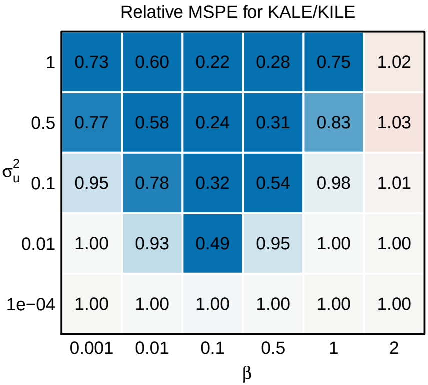


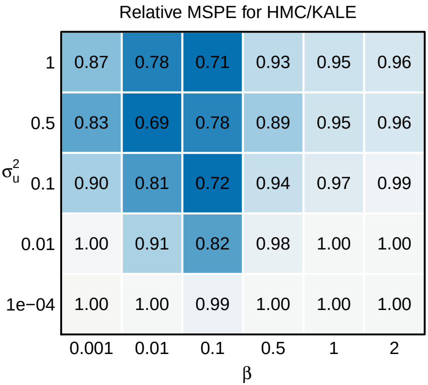
- black = observed locations; white = predicted locations.

Simulation study (known parameters)

MSE ratios

- Parameters assumed known.
- Nugget: $\sigma_x^2 = 0.0001$.

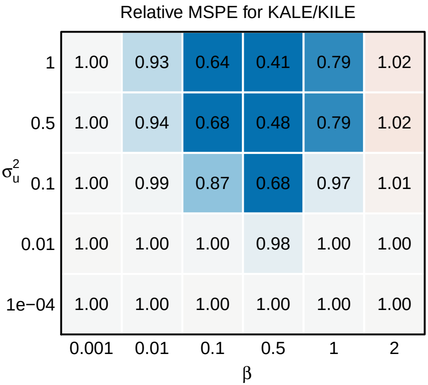


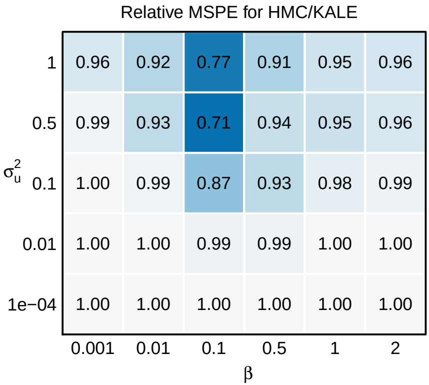


Simulation study (known parameters)

MSE ratios

- Parameters assumed known.
- Nugget: $\sigma_x^2 = 0.01$.





Simulation study (known parameters)

MSE ratios

- Parameters assumed known.
- Nugget: $\sigma_x^2 = 0.1$.

Relative MSPE for KALE/KILE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.00	0.96	0.86	0.67	0.84	1.03
0.5	1.00	0.99	0.94	0.77	0.89	1.01
0.1	1.00	1.00	0.99	0.92	0.99	1.00
0.01	1.00	1.00	1.00	1.00	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

Relative MSPE for HMC/KALE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.00	0.95	0.81	0.94	0.95	0.96
0.5	1.00	0.98	0.89	0.92	0.96	0.97
0.1	1.00	1.00	0.98	0.97	0.98	0.99
0.01	1.00	1.00	1.00	1.00	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

Simulation study (known parameters)

MSE ratios

- Parameters assumed known.
- Nugget: $\sigma_x^2 = 1$.

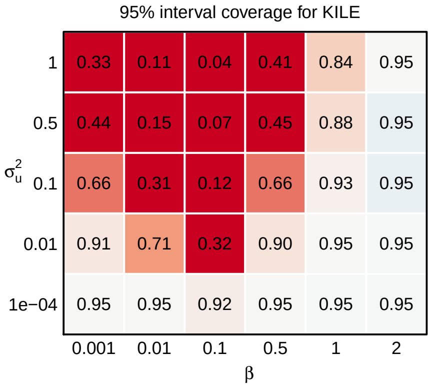
Relative MSPE for KALE/KILE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.00	1.00	0.99	0.92	0.97	1.01
0.5	1.00	1.00	1.00	0.96	0.98	1.01
0.1	1.00	1.00	1.00	1.00	1.00	1.00
0.01	1.00	1.00	1.00	1.00	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

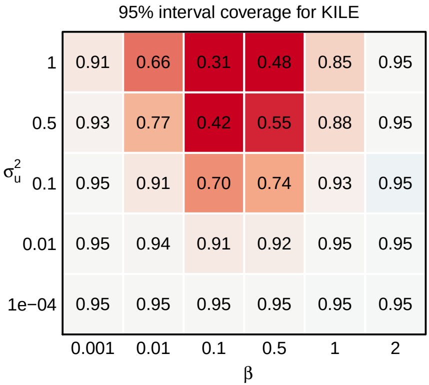
Relative MSPE for HMC/KALE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.00	1.00	0.99	0.98	0.98	0.98
0.5	1.00	1.00	0.98	0.99	0.99	0.98
0.1	1.00	1.00	1.00	0.99	1.00	1.00
0.01	1.00	1.00	1.00	1.00	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

- Parameters assumed known.



- $\sigma_x^2 = 0.0001$

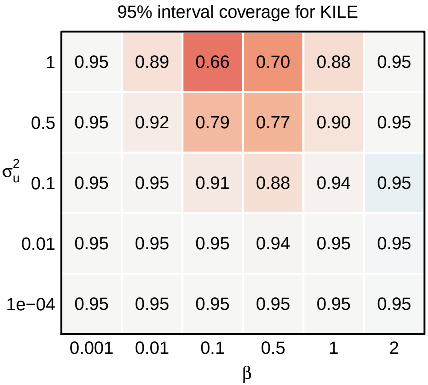


- $\sigma_x^2 = 0.01$

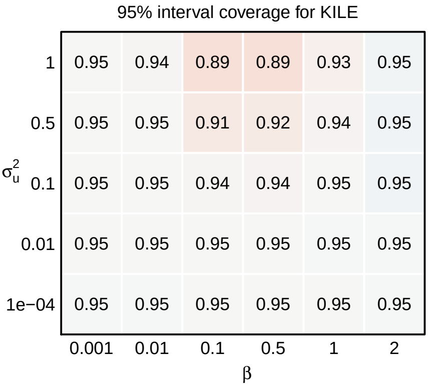
Simulation study (known parameters)

Interval coverage (KILE only)

- Parameters assumed known.



- $\sigma_x^2 = 0.1$



- $\sigma_x^2 = 1$

Simulation study

Estimating parameters

When parameters $\tau^2, \beta, \sigma_x^2$ are unknown:

- KILE: estimated with maximum likelihood.
- KALE: estimated with maximum pseudo-likelihood.
- HMC: given flat priors over reasonable range and sampled jointly with $\mathbf{u}_n$.

When parameters $\tau^2, \beta, \sigma_x^2$ are unknown:

- KILE: estimated with maximum likelihood.
- KALE: estimated with maximum pseudo-likelihood.
- HMC: given flat priors over reasonable range and sampled jointly with $\mathbf{u}_n$.

Recall the form of k for this simulation:

$$k(s_1, s_2) = \frac{\tau^2}{(1 + 4\beta\sigma_u^2)^{p/2}} \exp\left(-\frac{\beta}{1 + 4\beta\sigma_u^2}||s_1 - s_2||^2\right).$$

- $\tau^2, \beta, \sigma_u^2$ not identifiable.
- MLE invariance yields same estimated covariance function for KALE/KILE, though Kriging equations will be different.

Simulation study (unknown parameters)

MSE ratios

- Parameters unknown and first estimated.
- Nugget: $\sigma_x^2 = 0.0001$.

Relative MSPE for KALE/KILE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.98	1.02	0.88	1.00	0.98	1.00
0.5	0.97	0.97	0.95	0.97	0.99	1.00
0.1	1.00	1.00	0.97	0.98	1.00	1.00
0.01	1.00	1.00	1.00	0.99	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

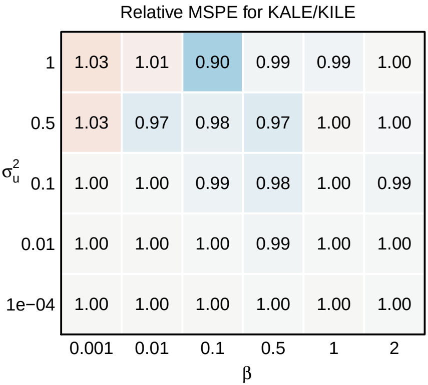
Relative MSPE for HMC/KALE

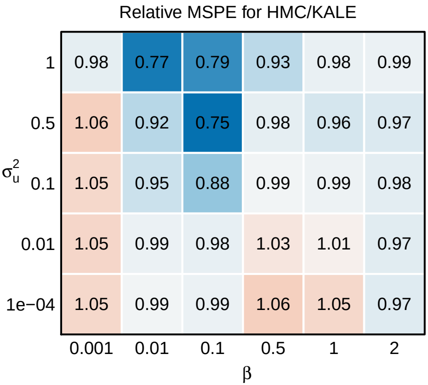
σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.92	0.77	0.82	1.00	0.98	0.98
0.5	0.97	0.74	0.80	0.94	0.96	0.98
0.1	0.96	0.83	0.77	1.00	0.98	0.97
0.01	0.99	0.92	0.85	1.09	1.02	0.98
1e−04	0.99	1.00	0.98	1.14	1.03	0.97

Simulation study (unknown parameters)

MSE ratios

- Parameters unknown and first estimated.
- Nugget: $\sigma_x^2 = 0.01$.





Simulation study (unknown parameters)

MSE ratios

- Parameters unknown and first estimated.
- Nugget: $\sigma_x^2 = 0.1$.

Relative MSPE for KALE/KILE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.05	1.02	0.94	0.97	0.99	0.99
0.5	1.02	0.86	0.98	0.99	1.00	1.00
0.1	1.01	1.01	1.01	0.98	0.99	0.99
0.01	1.00	1.00	1.00	1.00	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

Relative MSPE for HMC/KALE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.31	0.85	0.82	0.99	0.97	0.98
0.5	1.14	0.97	0.87	0.97	0.95	0.95
0.1	1.16	0.98	0.97	0.98	0.97	0.97
0.01	1.18	0.99	0.99	0.98	1.00	0.98
1e−04	1.15	0.99	0.98	0.98	1.00	0.96

Simulation study (unknown parameters)

MSE ratios

- Parameters unknown and first estimated.
- Nugget: $\sigma_x^2 = 1$.

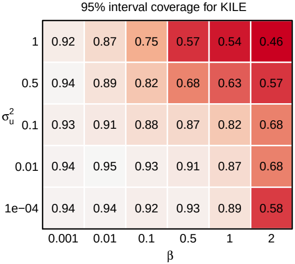
Relative MSPE for KALE/KILE

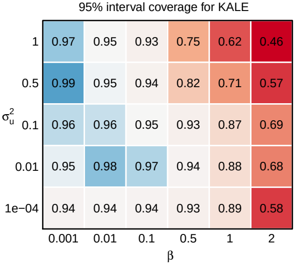
σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.03	1.08	1.01	0.99	0.99	0.99
0.5	0.90	1.05	1.02	1.00	1.00	0.99
0.1	0.92	1.00	1.01	1.01	1.00	0.99
0.01	1.00	1.00	1.00	1.00	1.00	1.00
1e−04	1.00	1.00	1.00	1.00	1.00	1.00

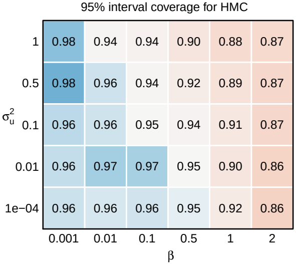
Relative MSPE for HMC/KALE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.58	1.17	0.91	0.96	0.95	0.94
0.5	1.35	1.10	0.95	0.94	0.94	0.95
0.1	1.35	1.28	1.05	0.96	0.94	0.96
0.01	1.21	1.11	1.06	0.96	0.94	0.95
1e−04	1.56	1.13	1.01	0.98	0.95	0.96

- Nugget: $\sigma_x^2 = 0.0001$.







Simulation study (unknown parameters)

Interval coverage

- Nugget: $\sigma_x^2 = 0.01$.

95% interval coverage for KILE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.92	0.88	0.76	0.55	0.56	0.48
0.5	0.93	0.91	0.85	0.71	0.59	0.57
0.1	0.92	0.92	0.91	0.86	0.80	0.67
0.01	0.93	0.93	0.92	0.89	0.85	0.70
1e−04	0.93	0.93	0.93	0.90	0.89	0.73

95% interval coverage for KALE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.94	0.94	0.92	0.70	0.58	0.43
0.5	0.94	0.95	0.94	0.84	0.64	0.57
0.1	0.93	0.94	0.96	0.92	0.85	0.68
0.01	0.93	0.94	0.94	0.92	0.86	0.70
1e−04	0.93	0.93	0.93	0.90	0.89	0.73

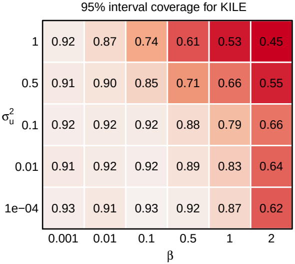
95% interval coverage for HMC

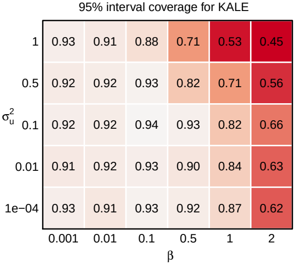
σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.97	0.96	0.94	0.91	0.88	0.87
0.5	0.97	0.96	0.95	0.93	0.89	0.87
0.1	0.97	0.96	0.96	0.93	0.91	0.87
0.01	0.97	0.96	0.95	0.94	0.91	0.87
1e−04	0.97	0.96	0.95	0.94	0.90	0.86

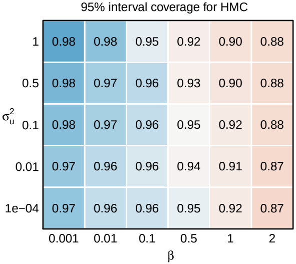
Simulation study (unknown parameters)

Interval coverage

- Nugget: $\sigma_x^2 = 0.1$.







Simulation study (unknown parameters)

Interval coverage

- Nugget: $\sigma_x^2 = 1$.

95% interval coverage for KILE

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.88	0.88	0.77	0.62	0.59	0.54
0.5	0.89	0.85	0.83	0.70	0.62	0.60
0.1	0.89	0.91	0.88	0.85	0.72	0.57
0.01	0.86	0.87	0.87	0.88	0.79	0.57
1e−04	0.86	0.88	0.90	0.88	0.80	0.56

95% interval coverage for KALE

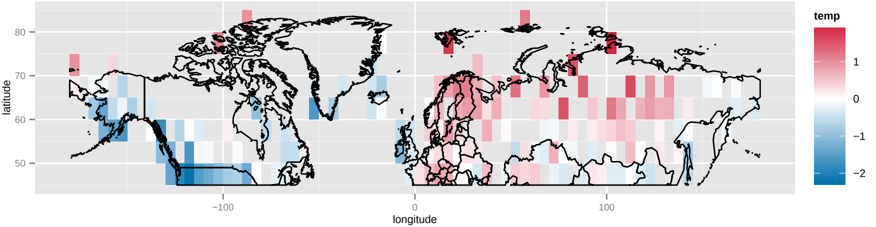
σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	0.87	0.89	0.82	0.65	0.53	0.43
0.5	0.87	0.86	0.86	0.70	0.58	0.52
0.1	0.89	0.89	0.89	0.86	0.72	0.56
0.01	0.86	0.87	0.87	0.88	0.79	0.57
1e−04	0.86	0.88	0.90	0.88	0.80	0.56

95% interval coverage for HMC

σ_u^2 \ β	0.001	0.01	0.1	0.5	1	2
1	1.00	0.99	0.99	0.98	0.97	0.96
0.5	0.99	0.99	0.99	0.98	0.96	0.96
0.1	0.99	0.99	0.99	0.98	0.97	0.94
0.01	0.99	0.99	0.98	0.98	0.97	0.95
1e−04	0.99	0.99	0.98	0.98	0.96	0.95

Data example

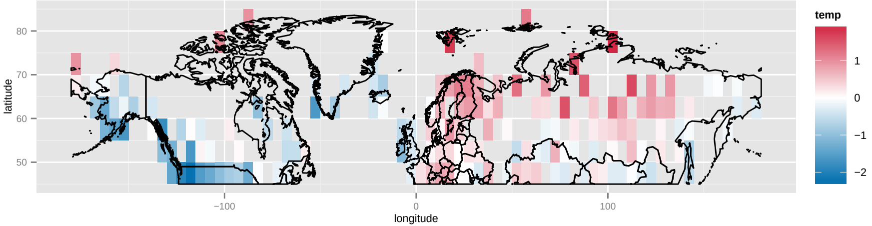
Interpolating northern hemisphere temperature anomolies for summer 2011[1]



[1] Data available: http://www.cru.uea.ac.uk/cru/data/temperature/

Data example

Interpolating northern hemisphere temperature anomolies for summer 2011[1]

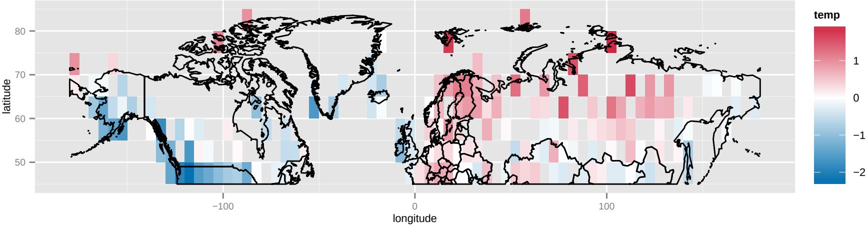


- Temperatures are averaged over April–September time window and $5^\circ \times 5^\circ$ long–lat grid cell.
- Values expressed as anomolies relative to 1860-2010 average [18].
- We further subtract the 2011 mean.
- Numerous pre-processing steps and adjustments to data [4, 13, 7].

[1] Data available: http://www.cru.uea.ac.uk/cru/data/temperature/

Data example

Interpolating northern hemisphere temperature anomolies for summer 2011[1]



- Temperatures are averaged over April–September time window and $5^\circ \times 5^\circ$ long–lat grid cell.
- Values expressed as anomolies relative to 1860-2010 average [18].
- We further subtract the 2011 mean.
- Numerous pre-processing steps and adjustments to data [4, 13, 7].

Geo-referencing by grid cells is a location error problem.

[1] Data available: http://www.cru.uea.ac.uk/cru/data/temperature/

Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

Covariance function is based on distance along the Earth's surface [19]:

$$c(s_1, s_2) = \tau^2 \exp(-\beta\Delta) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$$

$$\Delta = 2r \arcsin \sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\psi_2 - \psi_1}{2}\right)},$$

- $s = (\psi, \phi)$ are longitude, latitude pairs.
- $r = 6371$ is the Earth's radius in km.

Covariance function is based on distance along the Earth's surface [19]:

$$c(s_1, s_2) = \tau^2 \exp(-\beta\Delta) + \sigma_x^2 \mathbf{1}[s_1 = s_2]$$

$$\Delta = 2r \arcsin \sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\psi_2 - \psi_1}{2}\right)},$$

- $s = (\psi, \phi)$ are longitude, latitude pairs.
- $r = 6371$ is the Earth's radius in km.

We assume location errors are i.i.d. in terms of distance on the Earth's surface:

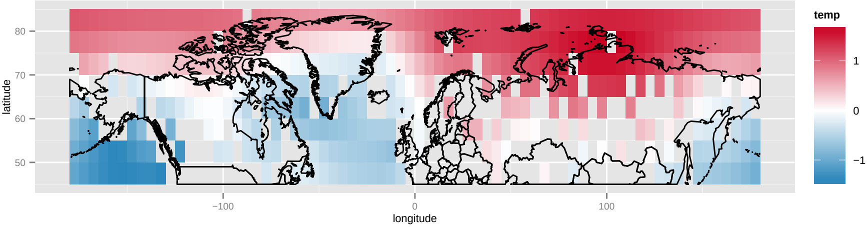
$$u_i \sim \mathcal{N}\left(\mathbf{0}, \sigma_u^2 \left(\frac{180}{\pi r}\right)^2 \begin{pmatrix} \frac{1}{\cos^2(\phi_i)} & 0 \\ 0 & 1 \end{pmatrix}\right).$$

- $\sigma_u^2 = 500$ yields a 28km expected distance between $s + u$ and s.

Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

KALE/KILE approach:



Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

KALE/KILE approach:

	$\hat{\tau}^2$	$\hat{\beta}$	$\hat{\sigma}_x^2$
KILE	1.167	1.428×10^{-4}	0.075
KALE	1.167	1.430×10^{-4}	0.074

Parameter estimates

Data example

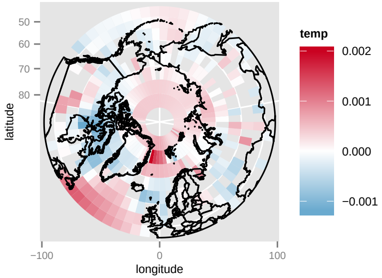
Interpolating northern hemisphere temperature anomolies for summer 2011

KALE/KILE approach:

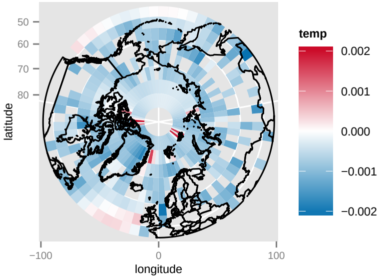
	$\hat{\tau}^2$	$\hat{\beta}$	$\hat{\sigma}_x^2$
KILE	1.167	1.428×10^{-4}	0.075
KALE	1.167	1.430×10^{-4}	0.074

Parameter estimates

KALE - KILE for point predictions:



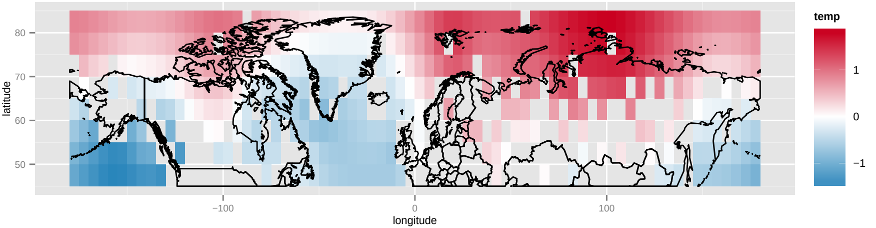
KALE - KILE for interval length:



Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

HMC approach



Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

HMC approach

This looks different from Kriging estimates

- HMC also averages over posterior parameter uncertainty.
- More meaningful comparison is against $\sigma_u^2 = 0$ model using HMC.

Data example

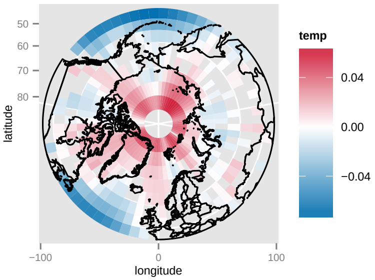
Interpolating northern hemisphere temperature anomolies for summer 2011

HMC approach

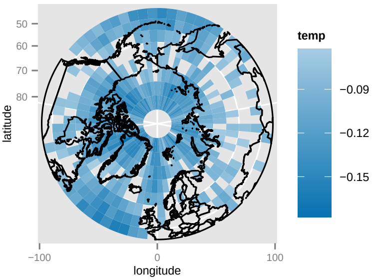
This looks different from Kriging estimates

- HMC also averages over posterior parameter uncertainty.
- More meaningful comparison is against $\sigma_u^2 = 0$ model using HMC.

$\{\sigma_u^2 = 500\} - \{\sigma_u^2 = 0\}$ point predictions:



$\{\sigma_u^2 = 500\} - \{\sigma_u^2 = 0\}$ interval lengths:



Data example

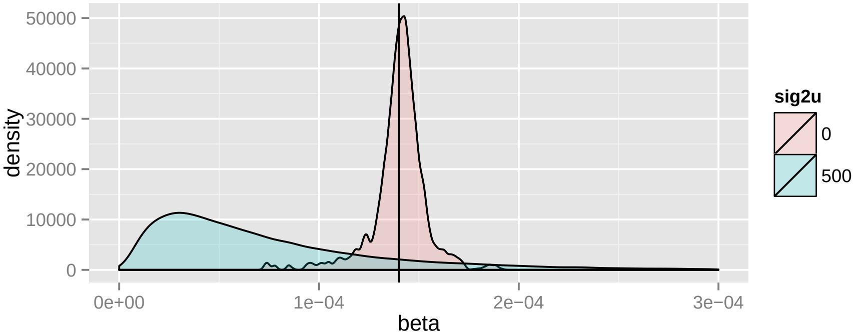
Interpolating northern hemisphere temperature anomolies for summer 2011

HMC differs in parameter inference for $\{\sigma_u^2 = 500\}$ and $\{\sigma_u^2 = 0\}$ models:

Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

HMC differs in parameter inference for $\{\sigma_u^2 = 500\}$ and $\{\sigma_u^2 = 0\}$ models:

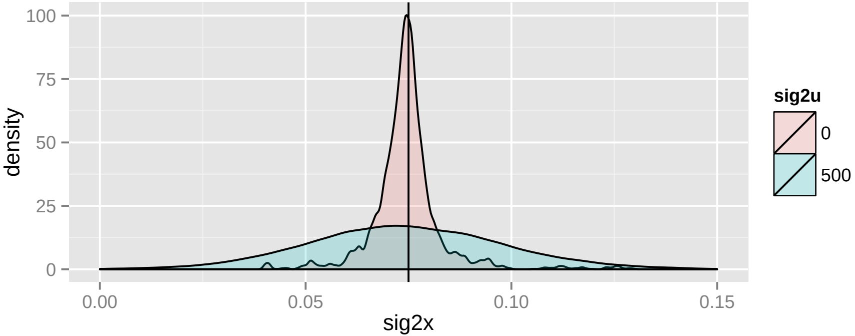


- HMC with $\{\sigma_u^2 = 0\}$ agrees with parameter inference from KALE/KILE.

Data example

Interpolating northern hemisphere temperature anomolies for summer 2011

HMC differs in parameter inference for $\{\sigma_u^2 = 500\}$ and $\{\sigma_u^2 = 0\}$ models:



- HMC with $\{\sigma_u^2 = 0\}$ agrees with parameter inference from KALE/KILE.

Conclusions

Location errors and noisy inputs are a common (but often ignored) problem in GP regression.

Conclusions

Location errors and noisy inputs are a common (but often ignored) problem in GP regression.

Analyst can get away with ignoring location errors when:

- Spatial correlations are either very strong or very weak and location errors are sufficiently small.
- Nugget variance σ_x^2 is large.

Conclusions

Location errors and noisy inputs are a common (but often ignored) problem in GP regression.

Analyst can get away with ignoring location errors when:

- Spatial correlations are either very strong or very weak and location errors are sufficiently small.
- Nugget variance σ_x^2 is large.

Kriging using moment properties of y (KALE) is an acceptable solution in some situations:

- Dominates KILE in MSE when covariance parameters are known.
- Provides correct confidence intervals when covariance parameters are known.

Conclusions

Location errors and noisy inputs are a common (but often ignored) problem in GP regression.

Analyst can get away with ignoring location errors when:

- Spatial correlations are either very strong or very weak and location errors are sufficiently small.
- Nugget variance σ_x^2 is large.

Kriging using moment properties of y (KALE) is an acceptable solution in some situations:

- Dominates KILE in MSE when covariance parameters are known.
- Provides correct confidence intervals when covariance parameters are known.

Using MCMC, we gain several additional advantages:

- Dominates KALE in MSE when covariance parameters are known.
- Provides correct confidence intervals when covariance parameters are known.
- Naturally incorporates uncertainty from estimated parameters.

Conclusions

Location errors and noisy inputs are a common (but often ignored) problem in GP regression.

Analyst can get away with ignoring location errors when:

- Spatial correlations are either very strong or very weak and location errors are sufficiently small.
- Nugget variance σ_x^2 is large.

Kriging using moment properties of y (KALE) is an acceptable solution in some situations:

- Dominates KILE in MSE when covariance parameters are known.
- Provides correct confidence intervals when covariance parameters are known.

Using MCMC, we gain several additional advantages:

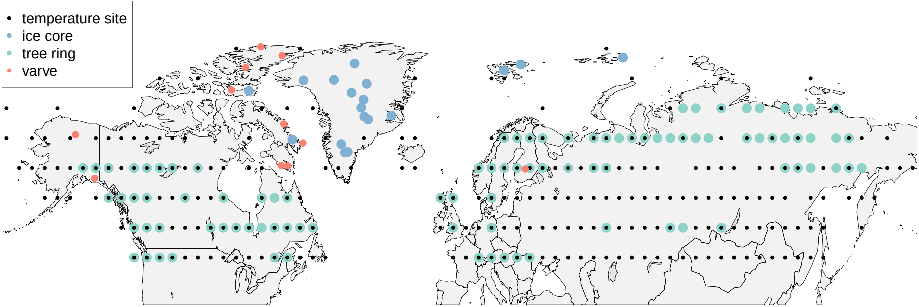
- Dominates KALE in MSE when covariance parameters are known.
- Provides correct confidence intervals when covariance parameters are known.
- Naturally incorporates uncertainty from estimated parameters.

Difficulties that remain:

- Prior sensitivity is an issue, particularly for spatial problems.
- MCMC covergence issues due to multiple (isolated) modes.
- Coverage guarantees when parameters are estimated.

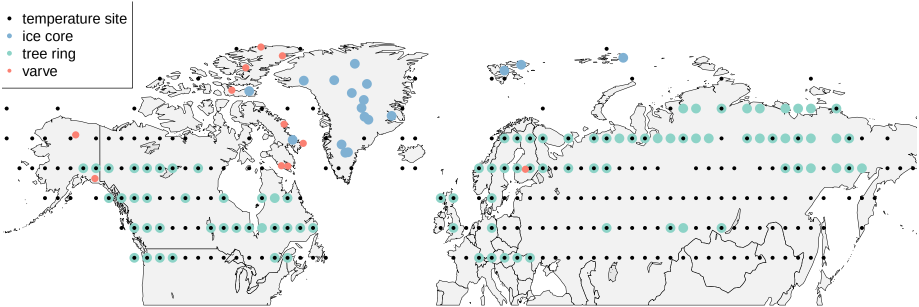
Future work

Climate reconstruction



Future work

Climate reconstruction



- Incorporating proxy data, with location uncertainties [11].
- Spatiotemporal heteroskedasticity in location errors.
- Nonstationary covariance behavior [17, 8].

Thanks to

This work:

- Natesh Pillai
- Peter Huybers
- Luke Bornn

My dissertation committee:

- Natesh Pillai
- Carl Morris
- Luke Bornn

Faculty, classmates, and friends in the Statistics Department.

References I

[1] J. J. Barber, A. E. Gelfand, and J. A. Silander.
Modelling map positional error to infer true feature location.
Canadian Journal of Statistics, 34(4):659–676, 2006.

[2] L. Beale, J. J. Abellan, S. Hodgson, and L. Jarup.
Methodologic issues and approaches to spatial epidemiology.
Environ Health Perspect, 116(8):1105–1110, 2008.

[3] M. R. Bonner, D. Han, J. Nie, P. Rogerson, J. E. Vena, and J. L. Freudenheim.
Positional accuracy of geocoded addresses in epidemiologic research.
Epidemiology, 14(4):408–412, 2003.

[4] P. Brohan, J. J. Kennedy, I. Harris, S. F. Tett, and P. D. Jones.
Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850.
Journal of Geophysical Research: Atmospheres (1984–2012), 111(D12), 2006.

[5] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu.
Measurement error in nonlinear models: a modern perspective.
CRC press, 2006.

[6] N. Cressie and J. Kornak.
Spatial statistics in the presence of location error with an application to remote sensing of the environment.
Statistical science, 18(4):436–456, 2003.

[7] P. Jones, D. Lister, T. Osborn, C. Harpham, M. Salmon, and C. Morice.
Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010.
Journal of Geophysical Research: Atmospheres (1984–2012), 117(D5), 2012.

[8] M. Jun and M. L. Stein.
Nonstationary covariance models for global data.
The Annals of Applied Statistics, pages 1271–1289, 2008.

[9] D. Koller, K. Daniilidis, T. Thorhallson, and H.-H. Nagel.
Model-based object tracking in traffic scenes.
In *Computer VisionECCV'92*, pages 437–452. Springer, 1992.

References II

[10] N. Krieger, J. T. Chen, P. D. Waterman, M.-J. Soobader, S. Subramanian, and R. Carson.
Geocoding and monitoring of us socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter? the public health disparities geocoding project.
American journal of epidemiology, 156(5):471–482, 2002.

[11] M. E. Mann, Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni.
Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia.
Proceedings of the National Academy of Sciences, 105(36):13252–13257, 2008.

[12] X.-L. Meng.
Multiple-imputation inferences with uncongenial sources of input.
Statistical Science, pages 538–558, 1994.

[13] C. P. Morice, J. J. Kennedy, N. A. Rayner, and P. D. Jones.
Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set.
Journal of Geophysical Research: Atmospheres (1984–2012), 117(D8), 2012.

[14] C. N. Morris.
Natural exponential families with quadratic variance functions: statistical theory.
The Annals of Statistics, pages 515–529, 1983.

[15] D. Pfoser and C. S. Jensen.
Capturing the uncertainty of moving-object representations.
In *Advances in Spatial Databases*, pages 111–131. Springer, 1999.

[16] D. Rocchini, J. Hortal, S. Lengyel, J. M. Lobo, A. Jimenez-Valverde, C. Ricotta, G. Bacaro, and A. Chiarucci.
Accounting for uncertainty when mapping species distributions: the need for maps of ignorance.
Progress in Physical Geography, 35(2):211–226, 2011.

[17] M. L. Stein.
Space–time covariance functions.
Journal of the American Statistical Association, 100(469):310–321, 2005.

[18] M. P. Tingley.
A bayesian anova scheme for calculating climate anomalies, with applications to the instrumental temperature record.
Journal of Climate, 25(2):777–791, 2012.

References III

[19] M. P. Tingley and P. Huybers.
A bayesian algorithm for reconstructing climate anomalies in space and time. part i: Development and applications to paleoclimate reconstruction problems.
Journal of Climate, 23(10):2759–2781, 2010.

[20] J. Warnes and B. Ripley.
Problems with likelihood estimation of covariance functions of spatial gaussian processes.
Biometrika, 74(3):640–642, 1987.