OPTIMAL PREDICTIVE INFERENCE IN LOG-GAUSSIAN RANDOM FIELDS

Victor De Oliveira
Department of Management Science and Statistics
The University of Texas at San Antonio

victor.deoliveira@utsa.edu
http://faculty.business.utsa.edu/vdeolive

Log-Gaussian Random Fields

The random field $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$, with $Z(\mathbf{s}) > 0$ and $D \subset \mathbb{R}^d$, is log-Gaussian if $\{\log(Z(\mathbf{s})) : \mathbf{s} \in D\}$ is Gaussian

Assume for now the mean and covariance functions of $Y(\cdot) = \log(Z(\cdot))$ are

$$\begin{aligned} E\{Y(\mathbf{s})\} &= \mu_Y \\ \operatorname{cov}\{Y(\mathbf{s}), Y(\mathbf{u})\} &= C_Y(\mathbf{s}, \mathbf{u}) \end{aligned}$$

with $\mu_Y \in \mathbb{R}$ unknown and $C_Y(\mathbf{s}, \mathbf{u})$ known covariance function in $\mathbb{R}^d$, and for all $\mathbf{s} \in D$, $C_Y(\mathbf{s}, \mathbf{s}) = \sigma_Y^2$.

The mean and covariance functions of $Z(\cdot)$ are

$$\begin{aligned} E\{Z(\mathbf{s})\} &= \exp\{\mu_Y + \frac{\sigma_Y^2}{2}\} =: \mu_Z \\ \operatorname{cov}\{Z(\mathbf{s}), Z(\mathbf{u})\} &= \mu_Z^2 \big(\exp\{C_Y(\mathbf{s}, \mathbf{u})\} - 1 \big) \end{aligned}$$

Predictive Inference

DATA:

$\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))$ measured at sampling locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$

Let $\mathbf{s}_0 \in D$ unmeasured location in D, and $B \subset D$ a subregion of D

GOALS:

- Obtain predictor of $Z(\mathbf{s}_0)$ [point prediction]
- Obtain predictor of $Z(B) = \frac{1}{|B|} \int_B Z(\mathbf{s}) d\mathbf{s}$ [block prediction]
- Obtain prediction interval for $Z(\mathbf{s}_0)$ [point interval prediction]
- Obtain prediction interval for $Z(B)$ [block interval prediction]

Point Prediction

Let $\hat{Z}_0$ be predictor for Z_0 (within a family) and $L(Z_0, \hat{Z}_0)$ a loss function
The optimal predictor of Z_0 is the predictor that minimizes the risk function

$$r(\hat{Z}_0) = E\{L(Z_0, \hat{Z}_0)\}$$

For squared error loss the risk becomes the mean squared prediction error

$$\text{MSPE}(\hat{Z}_0) = E\{(\hat{Z}_0 - Z_0)^2\}$$

Similar notation is used for prediction of $Z(B)$

Notation: Quantities depending on the prediction location $\mathbf{s}_0$ would be written with subscript '0'

For μ_Y known the optimal predictor and its MSPE are

$$\begin{aligned}
\hat{Z}_0^* &= E\{Z_0 \mid \mathbf{Z}\} = \exp\{\hat{Y}_0^* + \frac{\hat{\sigma}_{0Y}^*}{2}\} \\
\text{MSPE}(\hat{Z}_0^*) &= \text{var}(Z_0) - \text{var}(\hat{Z}_0^*) \\
&= \mu_Z^2\big(\exp\{\sigma_Y^2\} - \exp\{\mathbf{c}_{0Y}'\Sigma_Y^{-1}\mathbf{c}_{0Y}\}\big)
\end{aligned}$$

where

$$\begin{aligned}
\hat{Y}_0^* &= E\{Y_0 \mid \mathbf{Y}\} \\
&= \mu_Y + \mathbf{c}_{0Y}'\Sigma_Y^{-1}(\mathbf{Y} - \mu_Y\mathbf{1}) \\
\hat{\sigma}_{0Y}^* &= \text{var}(Y_0 \mid \mathbf{Y}) \\
&= \sigma_Y^2 - \mathbf{c}_{0Y}'\Sigma_Y^{-1}\mathbf{c}_{0Y}
\end{aligned}$$

$\Sigma_{Y,ij} = C_Y(\mathbf{s}_i, \mathbf{s}_j)$; $\quad \mathbf{c}_{0Y,i} = C_Y(\mathbf{s}_0, \mathbf{s}_i)$

These are known in the geostatistical literature as simple kriging predictors and simple kriging variances of Z_0 and Y_0.

Unbiased Prediction

The most used predictor in practice is the lognormal kriging predictor

$$\hat{Z}_0^{LK} = \exp\{\hat{Y}_0^{OK} + \frac{1}{2}(\sigma_Y^2 - \boldsymbol{\lambda}_{0Y}' \Sigma_Y \boldsymbol{\lambda}_{0Y})\}$$

and

$$\text{MSPE}(\hat{Z}_0^{LK}) = \mu_Z^2 \Big(\exp\{\sigma_Y^2\} + \exp\{\boldsymbol{\lambda}_{0Y}' \Sigma_Y \boldsymbol{\lambda}_{0Y}\} - 2\exp\{\boldsymbol{\lambda}_{0Y}' \Sigma_Y \boldsymbol{\lambda}_{0Y} - m_{0Y}\} \Big)$$

where

$$\begin{aligned} \hat{Y}_0^{OK} &= \boldsymbol{\lambda}_{0Y}' \mathbf{Y} \qquad \text{(the BLUP of } Y_0\text{)} \\ \boldsymbol{\lambda}_{0Y}' &= \Big(\mathbf{c}_{0Y} + \frac{1 - \mathbf{1}' \Sigma_Y^{-1} \mathbf{c}_{0Y}}{\mathbf{1}' \Sigma_Y^{-1} \mathbf{1}} \mathbf{1} \Big)' \Sigma_Y^{-1} \end{aligned}$$

By construction $\hat{Z}_0^{LK}$ satisfies the following optimality property:

Proposition 1 (Cressie, 1993). Predictor $\hat{Z}_0^{LK}$ minimizes $E\{(\log(\hat{Z}_0) - \log(Z_0))^2\}$ over the class of predictors of the form $\hat{Z}_0 = \exp\{\boldsymbol{\lambda}_0' \log(\mathbf{Z}) + k_0\}$, where $\boldsymbol{\lambda}_0 \in \mathbb{R}^n$ and $k_0 \in \mathbb{R}$ are constrained such that $E\{\hat{Z}_0\} = E\{Z_0\}$ for every $\mu_Y \in \mathbb{R}$

Recently, Cox (2004) noted a stronger optimality property:

Proposition 2. Predictor $\hat{Z}_0^{LK}$ minimizes $E\left\{\frac{\left(\hat{Z}_0 - Z_0\right)^2}{\exp\{\mathbf{c}_{0Y}' \Sigma_Y^{-1} \log(\mathbf{Z})\}}\right\}$ over the class of all unbiased predictors of Z_0

Comment. These optimality properties are somewhat unsatisfactory:
◦ The former holds in the transformed log-scale rather than in the original scale of measurement
◦ The latter, although for the original scale, holds with respect to a weighted squared error loss function with little intuitive appeal

Optimal Point Prediction

Consider the family of predictors

$$\mathcal{P}_0 = \left\{ \hat{Z}_0 = \exp\{\mathbf{a}_0'\mathbf{Y} + k_0\} \ : \ k_0 \in \mathbb{R},\ \mathbf{a}_0 \in \mathbb{R}^n,\ \mathbf{a}_0'\mathbf{1} = 1 \right\}$$

which includes many special cases $\hat{Z}_0^{LK}$, $\hat{Z}_0^{N} = \exp\{\hat{Y}_0^{OK}\}$ and others

Theorem 1. The predictor in the $\mathcal{P}_0$ that minimizes $E\{(\hat{Z}_0 - Z_0)^2\}$ is

$$\hat{Z}_0^{ME} = \exp\{\hat{Y}_0^{OK} + \frac{1}{2}(\sigma_Y^2 - \boldsymbol{\lambda}_{0Y}'\Sigma_Y\boldsymbol{\lambda}_{0Y} - 2m_{0Y})\}$$

and

$$\text{MSPE}(\hat{Z}_0^{ME}) = \mu_Z^2\Big(\exp\{\sigma_Y^2\} - \exp\{\boldsymbol{\lambda}_{0Y}'\Sigma_Y\boldsymbol{\lambda}_{0Y} - 2m_{0Y}\}\Big)$$

where $m_{0Y} = \frac{1-\mathbf{1}'\Sigma_Y^{-1}\mathbf{c}_{0Y}}{\mathbf{1}'\Sigma_Y^{-1}\mathbf{1}}$

Block Prediction

A related problem is the prediction, based on (point) data $\mathbf{Z}$, of

$$Z(B) = \frac{1}{|B|} \int_B Z(\mathbf{s}) d\mathbf{s}, \quad B \subset D$$

These variables are clalled 'blocks' in the geostatistical literature

Examples where this arises:

- Environmental Assessment
- Precision Farming

Block Prediction (cont.)

Two predictors have been proposed in the geostatistical literature:

- The lognormal kriging block predictor

$$\hat{Z}(B)^{LK} = \frac{1}{|B|}\int_B \hat{Z}^{LK}(\mathbf{s})d\mathbf{s}$$

- A block predictor motivated by the assumption of "preservation of lognormality"

$$\hat{Z}(B)^{PL} = \exp\left\{\hat{Y}(B)^{OK} + \frac{1}{2}\left(\sigma_Y^2 - \boldsymbol{\lambda}_Y'(B)\Sigma_Y\boldsymbol{\lambda}_Y(B)\right)\right\}$$

where $\hat{Y}(B)^{OK} = \boldsymbol{\lambda}_Y'(B)\mathbf{Y}$ is the BLUP of $Y(B) = \int_B Y(\mathbf{s})d\mathbf{s}/|B|$ based on $\mathbf{Y}$

Both predictior are unbiased for $Z(B)$, but no other properties are given

Optimal Block Prediction

Consider the family of block predictors

$$\mathcal{P}_B = \left\{ \hat{Z}(B) = \frac{1}{|B|} \int_B \exp\{\hat{Y}^{OK}(\mathbf{s}) + k(\mathbf{s})\} d\mathbf{s} \ : \ k(\mathbf{s}) \in \mathcal{C}(B) \right\}$$

where $\mathcal{C}(B)$ is the space of continuous and bounded functions on B

Theorem 2. The predictor in the family of predictors $\mathcal{P}_B$ that minimizes $E\{(\hat{Z}(B) - Z(B))^2\}$ is

$$\hat{Z}(B)^{ME} = \frac{1}{|B|} \int_B \hat{Z}^{ME}(\mathbf{s}) d\mathbf{s}$$

where $\hat{Z}^{ME}(\mathbf{s})$ is the optimal point predictor given before, and

$$\text{MSPE}(\hat{Z}(B)^{ME}) = \frac{\mu_Z^2}{|B|^2} \int_B \int_B \left(\exp\{C_Y(\mathbf{s}, \mathbf{u})\} - \exp\{\boldsymbol{\lambda}_Y'(\mathbf{s}) \Sigma_Y \boldsymbol{\lambda}_Y(\mathbf{u}) - m_Y(\mathbf{s}) - m_Y(\mathbf{u})\} \right) d\mathbf{s} d\mathbf{u}$$

Considering another family of block predictors

$$\tilde{\mathcal{P}}_B = \left\{ \hat{Z}(B) = \exp\{\hat{Y}(B)^{OK} + k_B\} \ : \ k_B \in \mathbb{R} \right\}$$

Theorem 3. The predictor in the family of predictors $\tilde{\mathcal{P}}_B$ that minimizes $E\{(\hat{Z}(B) - Z(B))^2\}$ is given by

$$\hat{Z}(B)^{MP} = \exp\left\{ \hat{Y}(B)^{OK} + \frac{1}{2}\left(\sigma_Y^2 - 3\boldsymbol{\lambda}_Y'(B)\Sigma_Y\boldsymbol{\lambda}_Y(B)\right) + \log\left(\frac{1}{|B|}\int_B e^{\boldsymbol{\lambda}_Y'(B)\mathbf{c}_Y(\mathbf{s})}d\mathbf{s}\right)\right\}$$

and

$$\begin{aligned}\mathrm{MSPE}(\hat{Z}(B)^{MP}) \quad = \quad & \frac{\mu_Z^2}{|B|^2}\left(\int_B\int_B \exp\{C_Y(\mathbf{s},\mathbf{u})\}d\mathbf{s}d\mathbf{u}\right. \\ & \left. - |B|^2\exp\left\{2\log\left(\frac{1}{|B|}\int_B e^{\boldsymbol{\lambda}_Y'(B)}\mathbf{c}_Y(\mathbf{s})d\mathbf{s}\right) - \boldsymbol{\lambda}_Y'(B)\Sigma_Y\boldsymbol{\lambda}_Y(B)\right\}\right)\end{aligned}$$

Comparison of Predictors

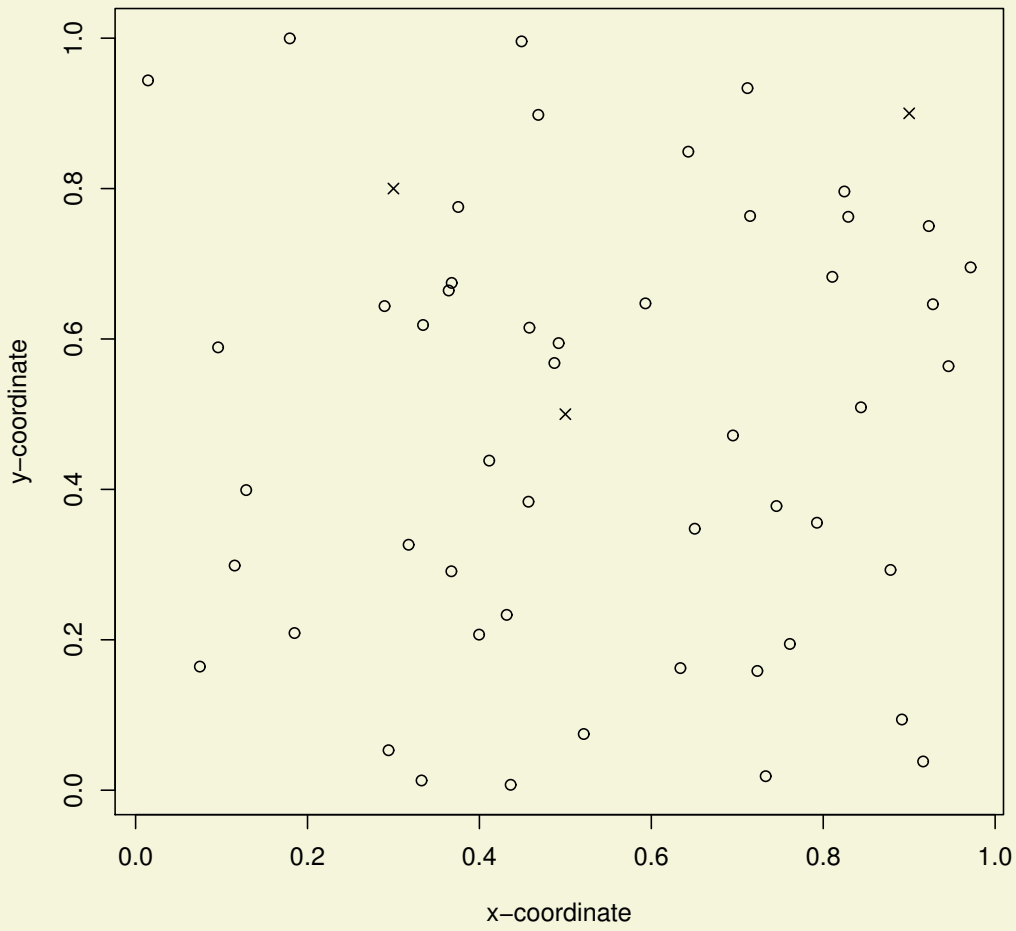
Consider random field $Z(\mathbf{s}) = \exp\{Y(\mathbf{s})\}$, defined on region $D = [0,1] \times [0,1]$, where $\{Y(\mathbf{s}), \mathbf{s} \in D\}$ is Gaussian with

$$E\{Y(\mathbf{s})\} = \mu_Y \quad , \quad C_Y(\mathbf{s}, \mathbf{u}) = \sigma_Y^2 \exp\{-\frac{||\mathbf{s} - \mathbf{u}||}{\theta_Y}\}$$

$\mu_Y \in \mathbb{R}$ and $\sigma_Y^2, \theta_Y > 0$

Data on $Z(\cdot)$ is observed at $n = 50$ sampling locations chosen at random

Sampling Design



○ : sampling location × : prediction location

Comparison of Point Predictors

We compare the values of $\hat{Z}_0^{ME}$ and $\hat{Z}_0^{LK}$ by predicting $Z(\mathbf{s}_0)$ for locations

$$\mathbf{s}_0 = (0.5, 0.5), (0.3, 0.8) \text{ and } (0.9, 0.9)$$

For that note

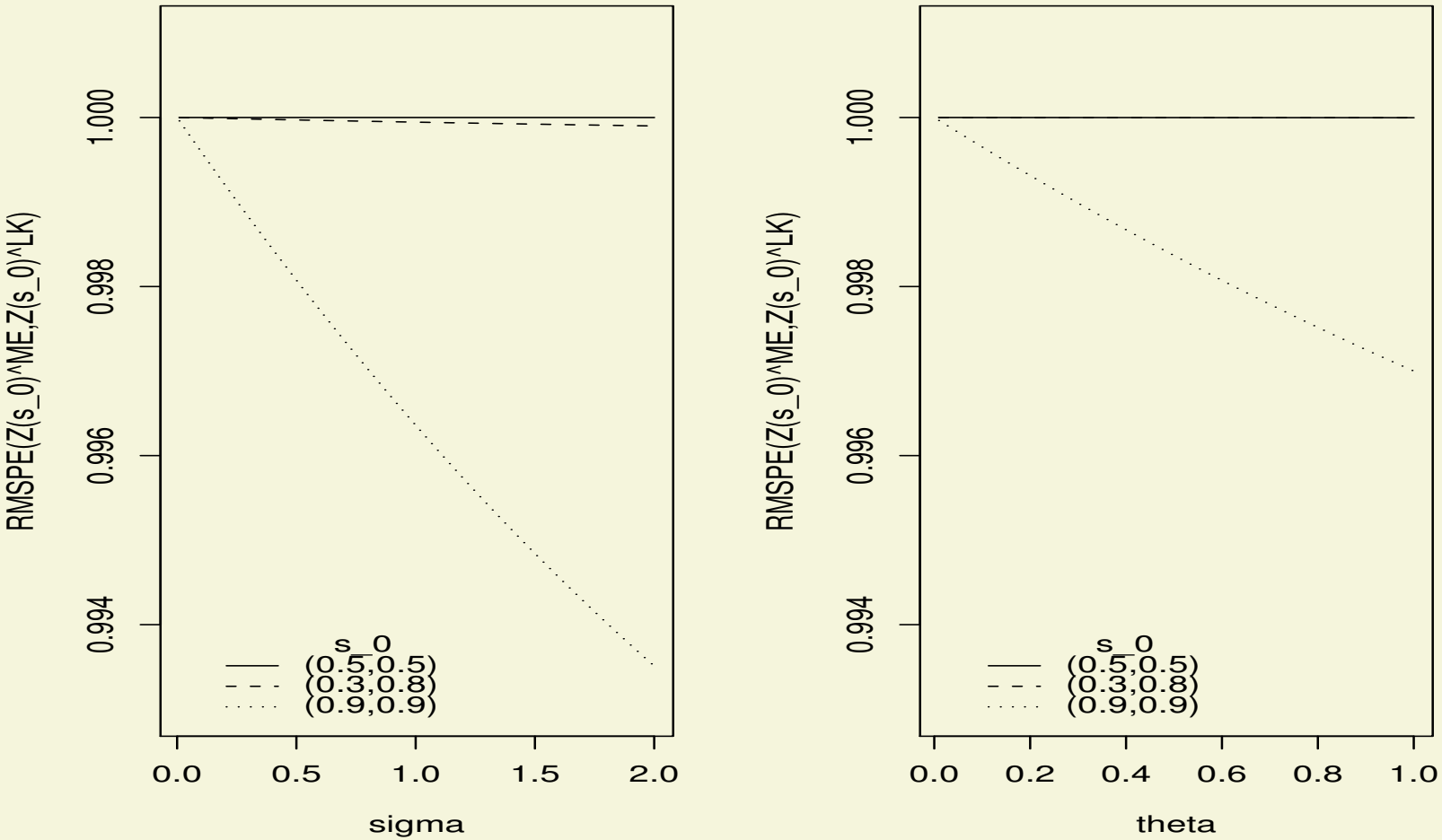
$$\frac{\hat{Z}_0^{ME}}{\hat{Z}_0^{LK}} = \exp\{-m_{0Y}\} = \frac{E\{\hat{Z}_0^{ME}\}}{E\{Z_0\}}$$

does not depend on the observed data

To compare MSPEs use predictive efficiency of $\hat{Z}_0^{ME}$ relative to $\hat{Z}_0^{LK}$

$$\text{RMSPE}(\hat{Z}_0^{ME}, \hat{Z}_0^{LK}) = \frac{\text{MSPE}(\hat{Z}_0^{ME})}{\text{MSPE}(\hat{Z}_0^{LK})}$$

Comparison of Point MSPEs



Comparison of Block Predictors

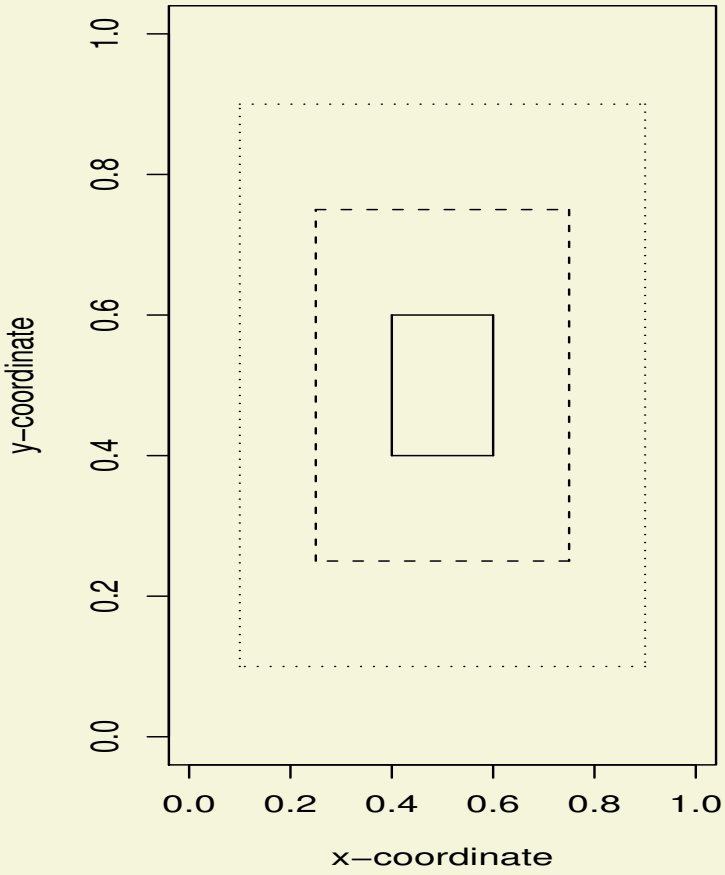
The block predictors are approximated by noting that

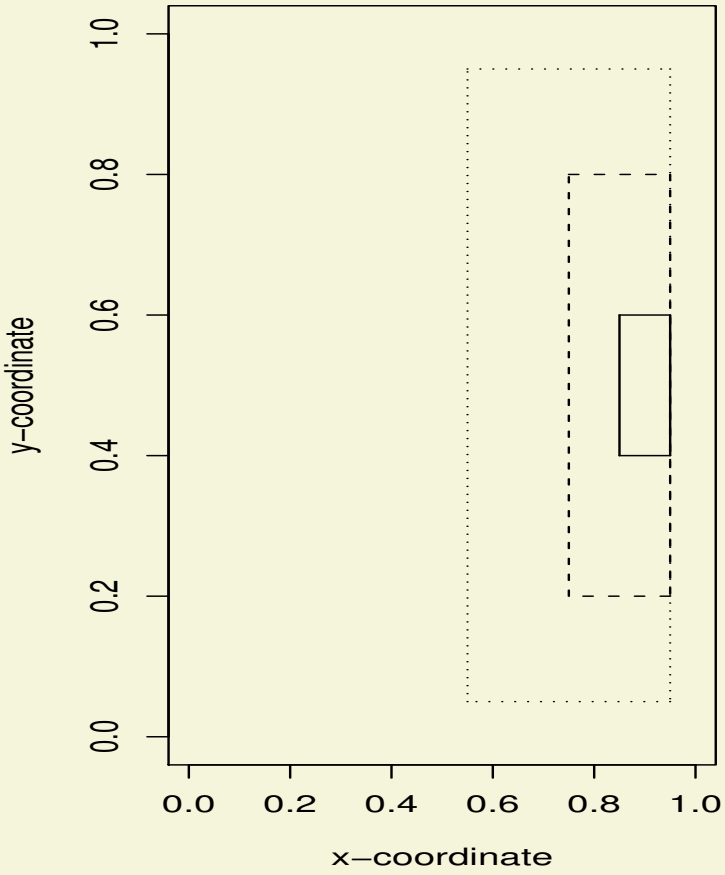
$$\hat{Z}(B)^{LK} = E_{\mathbf{S}}\{\hat{Z}^{LK}(\mathbf{S})\} \quad \text{and} \quad \hat{Z}(B)^{ME} = E_{\mathbf{S}}\{\hat{Z}^{ME}(\mathbf{S})\}$$

where $\hat{Z}^{LK}(\cdot)$ and $\hat{Z}^{ME}(\cdot)$ are the point predictors, and expectation is taken with respect to $\mathbf{S} \sim \text{unif}(B)$

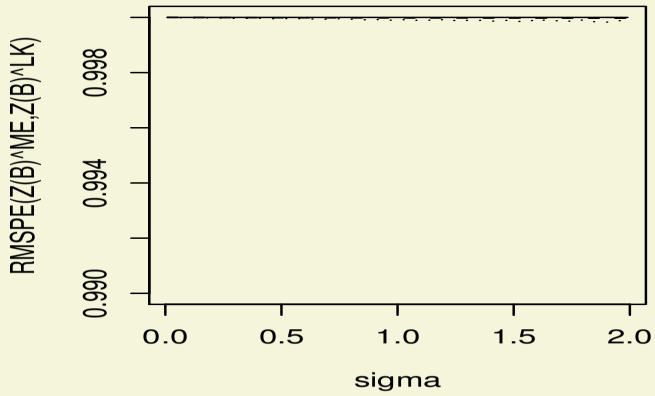
We compare the block predictors of $Z(B)$ for the following sub-regions B

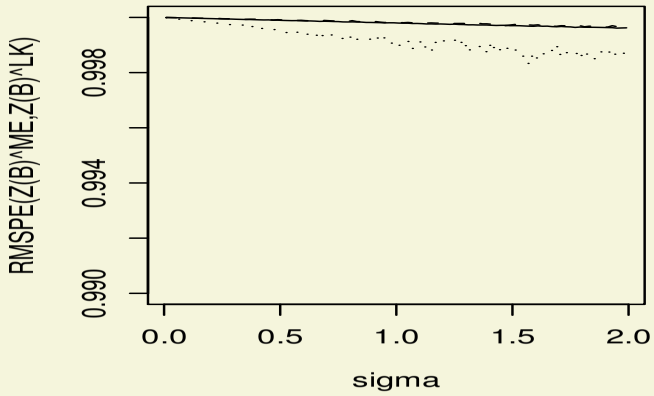
Block Subregions

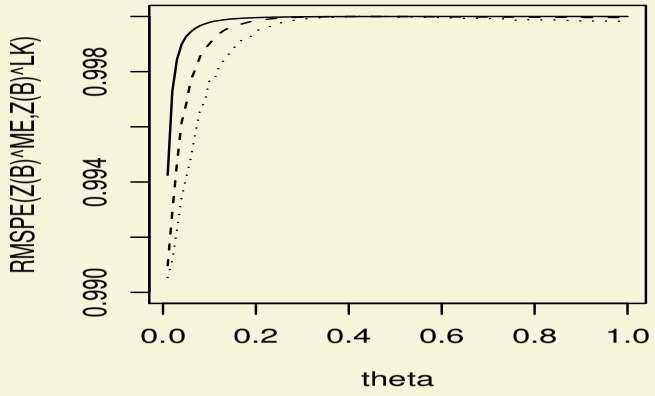


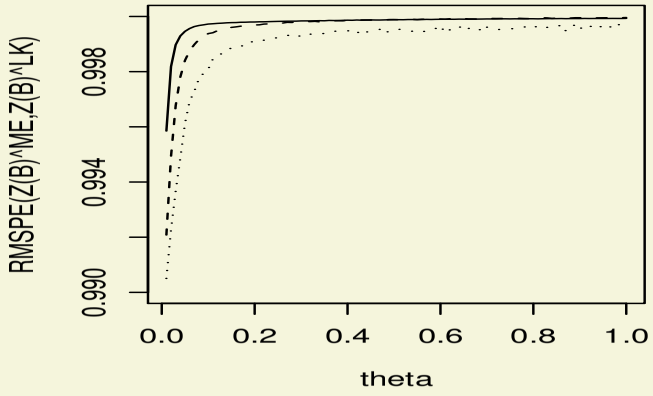


Comparison of Block MSPEs: $Z(B)^{ME}$ and $Z(B)^{LK}$

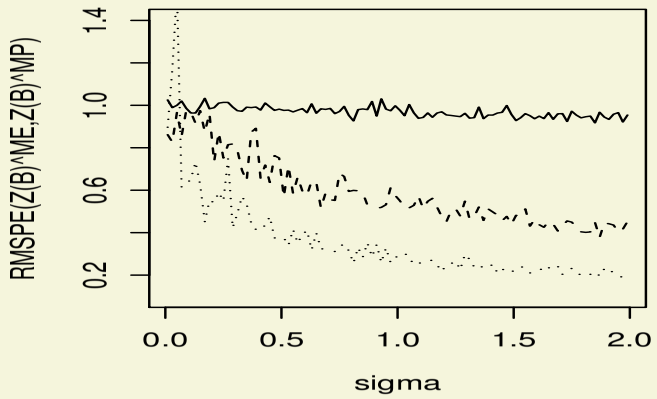


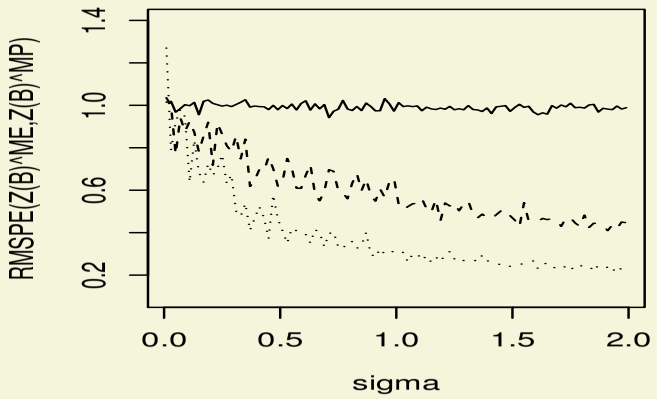


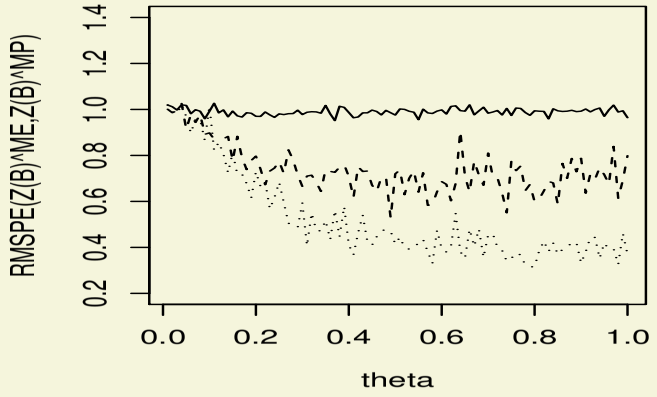


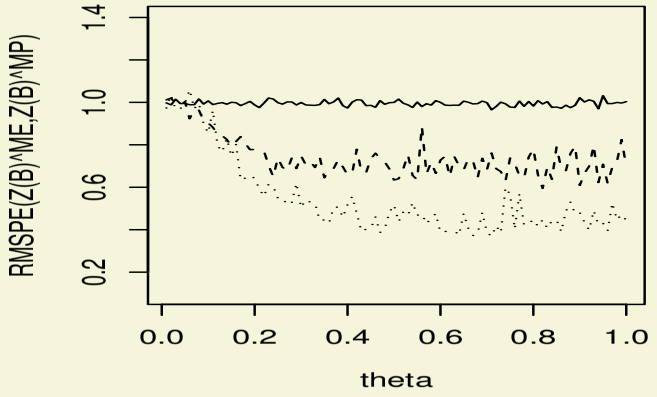


Comparison of Block MSPEs: $Z(B)^{ME}$ and $Z(B)^{LK}$









The Non-constant Mean Case

Suppose now that

$$\mu_Y(\mathbf{s}) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s})$$

$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ unknown regression parameters
$(f_1(\mathbf{s}), \ldots, f_p(\mathbf{s}))'$ known location-dependent covariates
In this case we have:

- The result on optimal point prediction (Theorem 1) can be easily extended
- The results on optimal block prediction (Theorems 2 and 3) *cannot* be extended

First Conclusions

• Point and block predictors for log-Gaussian processes exist that improve upon lognormal kriging, but ...

• Lognormal kriging point and block predictors have (near) optimality properties in the original scale

• The lognormal kriging block predictor is substantially better than the block predictor motivated by "permanence of lognormality"
Also, the best predictor in $\mathcal{P}_B$ is better than the best predictor in $\tilde{\mathcal{P}}_B$

• For random fields with non-constant mean the optimality results also hold for point prediction, but not for block prediction

Interval Prediction

Assume now the mean and covariance functions of $Y(\cdot)$ are

$$\begin{aligned} E\{Y(\mathbf{s})\} &= \sum_{j=1}^{p} \beta_j f_j(\mathbf{s}) \\ \text{cov}\{Y(\mathbf{s}), Y(\mathbf{u})\} &= C(\mathbf{s}, \mathbf{u}) \end{aligned}$$

$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ unknown regression parameters

$f_1(\mathbf{s}), \ldots, f_p(\mathbf{s})$ known location-dependent covariates

$C(\mathbf{s}, \mathbf{u})$ parametric covariance function in $\mathbb{R}^d$, with $C(\mathbf{s}, \mathbf{s}) = \sigma^2 > 0$

Observed Data

We have noisy measurements of the random field $Z(\cdot)$ at known sampling locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$:

$$Z_{i,\text{obs}} = Z(\mathbf{s}_i)\epsilon_i, \qquad i = 1, \ldots, n$$

with $\{\log(\epsilon_i)\} \overset{\text{iid}}{\sim} \text{N}(0, \sigma_\epsilon^2)$ measurement errors distributed independently of $Z(\cdot)$, and $\sigma_\epsilon^2 \geq 0$

GOAL:
Obtain prediction interval for $Z_0 = Z(\mathbf{s}_0)$, the unobserved value of the process at $\mathbf{s}_0 \in D$, based on $\mathbf{Z} = \{Z_{i,\text{obs}}\}_{i=1}^n$

Model parameters are $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^q$, which include $\sigma_\epsilon^2, \sigma^2$ and any other parameters in $C(\mathbf{s}, \mathbf{u})$

Standard Approach

Standard approach to find prediction intervals for $Z(\cdot)$ is to transform prediction intervals for $Y(\cdot)$. Let

$$\mathbf{Y} = \log(\mathbf{Z}) \quad \text{and} \quad Y_0 = \log(Z_0)$$

The BLUP of Y_0 based on $\mathbf{Y}$ and its mean squared prediction error are

$$\hat{Y}_0(\boldsymbol{\vartheta}) = \boldsymbol{\lambda}_0'(\boldsymbol{\vartheta})\mathbf{Y} \quad , \quad \hat{\sigma}_0^2(\boldsymbol{\vartheta}) = \sigma^2 - 2\boldsymbol{\lambda}_0'(\boldsymbol{\vartheta})\mathbf{c}_0(\boldsymbol{\vartheta}) + \boldsymbol{\lambda}_0'(\boldsymbol{\vartheta})\Sigma_{\vartheta}\boldsymbol{\lambda}_0(\boldsymbol{\vartheta})$$

with

$$\boldsymbol{\lambda}_0'(\boldsymbol{\vartheta}) = \left(\mathbf{c}_0(\boldsymbol{\vartheta}) + X(X'\Sigma_{\vartheta}^{-1}X)^{-1}(\mathbf{x}_0 - X'\Sigma_{\vartheta}^{-1}\mathbf{c}_0(\boldsymbol{\vartheta}))\right)'\Sigma_{\vartheta}^{-1}$$

$X = (f_j(\mathbf{s}_i))_{n\times p}$, $\mathbf{x}_0 = (f_1(\mathbf{s}_0), \ldots, f_p(\mathbf{s}_0))'$
$(\Sigma_{\vartheta,ij})_{n\times n} = C(\mathbf{s}_i, \mathbf{s}_j) + \sigma_\epsilon^2\mathbf{1}\{i = j\}$ (positive definite), $(\mathbf{c}_0(\boldsymbol{\vartheta})_i)_{n\times 1} = C(\mathbf{s}_0, \mathbf{s}_i)$

When $\boldsymbol{\vartheta}$ is *known*, it follows

$$\begin{pmatrix} Y_0 \\ \hat{Y}_0(\boldsymbol{\vartheta}) \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} \mathbf{x}_0'\boldsymbol{\beta} \\ \mathbf{x}_0'\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \boldsymbol{\lambda}_0'(\boldsymbol{\vartheta})\mathbf{c}_0(\boldsymbol{\vartheta}) \\ \boldsymbol{\lambda}_0'(\boldsymbol{\vartheta})\mathbf{c}_0(\boldsymbol{\vartheta}) & \boldsymbol{\lambda}_0'(\boldsymbol{\vartheta})\Sigma_{\vartheta}\boldsymbol{\lambda}_0(\boldsymbol{\vartheta}) \end{pmatrix} \right)$$

so $T = Y_0 - \hat{Y}_0(\boldsymbol{\vartheta}) \sim \mathrm{N}(0, \hat{\sigma}_0^2(\boldsymbol{\vartheta}))$ is a pivot for the prediction of Y_0

Then a $1-\alpha$ prediction interval for Y_0 is

$$\hat{Y}_0(\boldsymbol{\vartheta}) \pm \Phi^{-1}(1-\alpha/2)\hat{\sigma}_0(\boldsymbol{\vartheta})$$

and a $1-\alpha$ prediction interval for Z_0 is

$$\exp\{\hat{Y}_0(\boldsymbol{\vartheta}) \pm \Phi^{-1}(1-\alpha/2)\hat{\sigma}_0(\boldsymbol{\vartheta})\}$$

This $I_0^N(\alpha, \boldsymbol{\vartheta})$ is called the *standard* $1-\alpha$ prediction interval for Z_0

Shortest Prediction Intervals: Known Covariance

Consider the family of $1-\alpha$ prediction intervals for Z_0

$$\mathcal{F}_0 = \left\{ \left(\exp\{\hat{Y}_0(\boldsymbol{\vartheta}) - \Phi^{-1}(1-\gamma)\hat{\sigma}_0(\boldsymbol{\vartheta}) \ , \ \exp\{\hat{Y}_0(\boldsymbol{\vartheta}) - \Phi^{-1}(1-\alpha+\gamma)\hat{\sigma}_0(\boldsymbol{\vartheta})\} \right) : \gamma \in [0, \alpha) \right\}$$

which includes the standard prediction interval (obtained for $\gamma = \alpha/2$)

Theorem 4. Let $\alpha \in (0,1)$, $\boldsymbol{\vartheta} \in \Theta$ and $\mathbf{s}_0 \in D$. The shortest prediction interval in $\mathcal{F}_0$ is the one corresponding to the value $\gamma = \gamma_0^{\text{opt}} = \gamma_0^{\text{opt}}(\alpha, \boldsymbol{\vartheta}) \in (0, \alpha/2)$, which is the (unique) solution to the equation

$$\Phi^{-1}(1-\gamma) - \Phi^{-1}(1-\alpha+\gamma) = 2\hat{\sigma}_0(\boldsymbol{\vartheta})$$

Hence the shortest $1-\alpha$ PI for Z_0 in $\mathcal{F}_0$ is

$$I_0^S(\alpha, \boldsymbol{\vartheta} = \left(\exp\{\hat{Y}_0(\boldsymbol{\vartheta}) - \Phi^{-1}(1-\gamma_0^{\text{opt}})\hat{\sigma}_0(\boldsymbol{\vartheta}) \ , \ \exp\{\hat{Y}_0(\boldsymbol{\vartheta}) - \Phi^{-1}(1-\alpha+\gamma_0^{\text{opt}})\hat{\sigma}_0(\boldsymbol{\vartheta})\} \right)$$

Comparison

Let $\mathrm{RL}(I_0^N(\alpha,\boldsymbol{\vartheta}), I_0^S(\alpha,\boldsymbol{\vartheta}))$ be defined as

$$\frac{\mathrm{len}(I_0^S(\alpha,\boldsymbol{\vartheta}))}{\mathrm{len}(I_0^N(\alpha,\boldsymbol{\vartheta}))} = \frac{\exp\{\Phi^{-1}(1-\alpha+\gamma_0^{\mathrm{opt}})\hat{\sigma}_0(\boldsymbol{\vartheta})\} - \exp\{-\Phi^{-1}(1-\gamma_0^{\mathrm{opt}})\hat{\sigma}_0(\boldsymbol{\vartheta})\}}{\exp\{\Phi^{-1}(1-\alpha/2)\hat{\sigma}_0(\boldsymbol{\vartheta})\} - \exp\{-\Phi^{-1}(1-\alpha/2)\hat{\sigma}_0(\boldsymbol{\vartheta})\}}$$

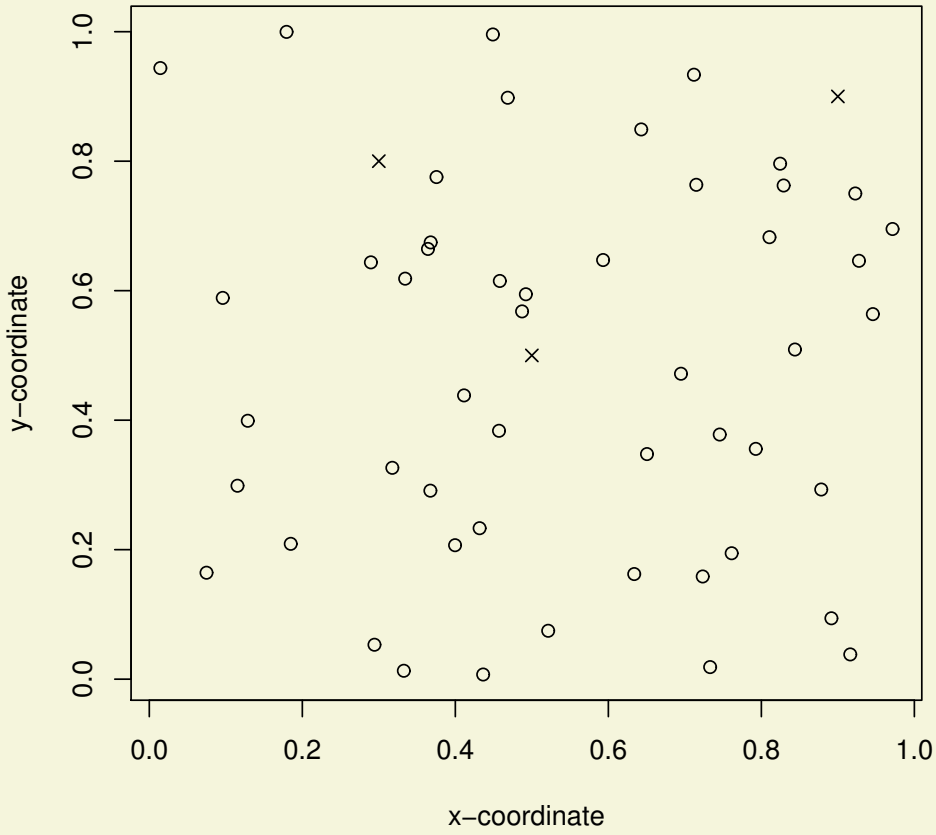
Consider $D = [0,1] \times [0,1]$ and random field $Z(\mathbf{s}) = \exp\{Y(\mathbf{s})\}$, where $\{Y(\mathbf{s}), \mathbf{s} \in D\}$ is Gaussian with constant mean and Matérn covariance function

$$C(\mathbf{s},\mathbf{u}) = \frac{\sigma^2}{2^{\theta_2-1}\Gamma(\theta_2)}\Big(\frac{l}{\theta_1}\Big)^{\theta_2}\mathcal{K}_{\theta_2}\Big(\frac{||\mathbf{s}-\mathbf{u}||}{\theta_1}\Big)$$

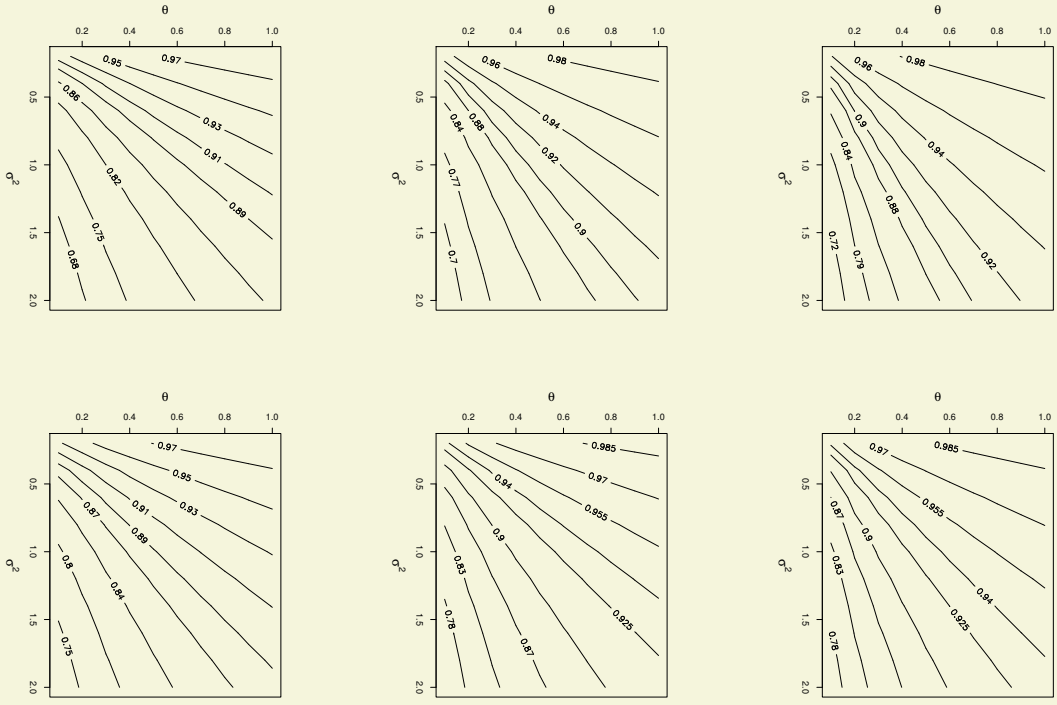
$\boldsymbol{\vartheta} = (\sigma^2, \theta_1, \theta_2)$ are covariance parameters

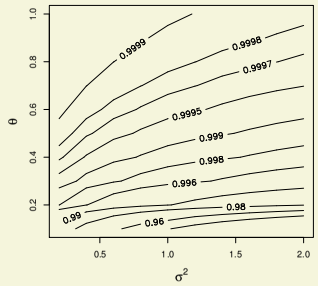
We consider the cases $\theta_2 = 0.5$ and $\theta_2 = 1.5$, and $\sigma_\epsilon^2 = 0$ (no measurement error)

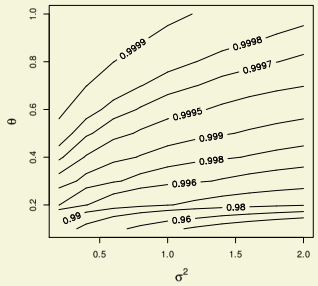
Sampling Design

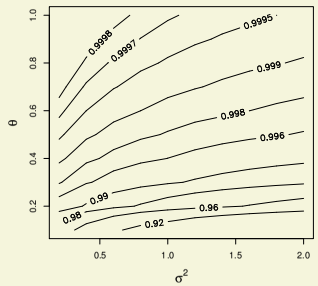


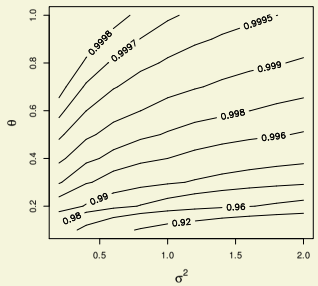
$\circ$: sampling location $\times$: prediction location

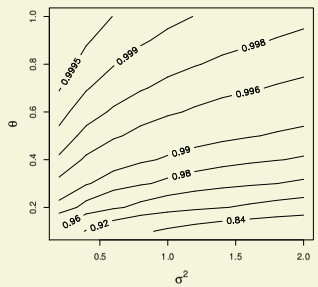

θ
σ²
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0
0.95
0.97
0.86
0.93
0.91
0.82
0.89
0.75
0.88
0.96
0.98
0.94
0.88
0.84
0.92
0.77
0.9
0.7
0.96
0.98
0.9
0.94
0.84
0.88
0.72
0.79
0.92
0.97
0.95
0.87
0.93
0.91
0.89
0.8
0.84
0.75
0.985
0.97
0.94
0.955
0.9
0.925
0.83
0.78
0.87
0.97
0.985
0.9
0.955
0.83
0.84
0.78
0.925

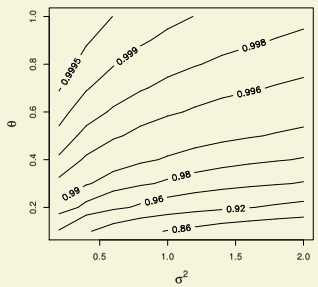

θ
σ²
0.9999
0.9998
0.9997
0.9995
0.999
0.998
0.996
0.99
0.98
0.96


θ
σ²
0.9999
0.9998
0.9997
0.9995
0.999
0.998
0.996
0.99
0.98
0.96


θ
σ²
0.9998
0.9997
0.9995
0.999
0.998
0.996
0.99
0.98
0.96
0.92


θ
σ²
0.9998
0.9997
0.9995
0.999
0.998
0.996
0.99
0.98
0.96
0.92


θ
σ²
0.9995
0.999
0.998
0.996
0.99
0.98
0.96
0.92
0.84


θ
σ²
0.9995
0.999
0.998
0.996
0.99
0.98
0.96
0.92
0.86

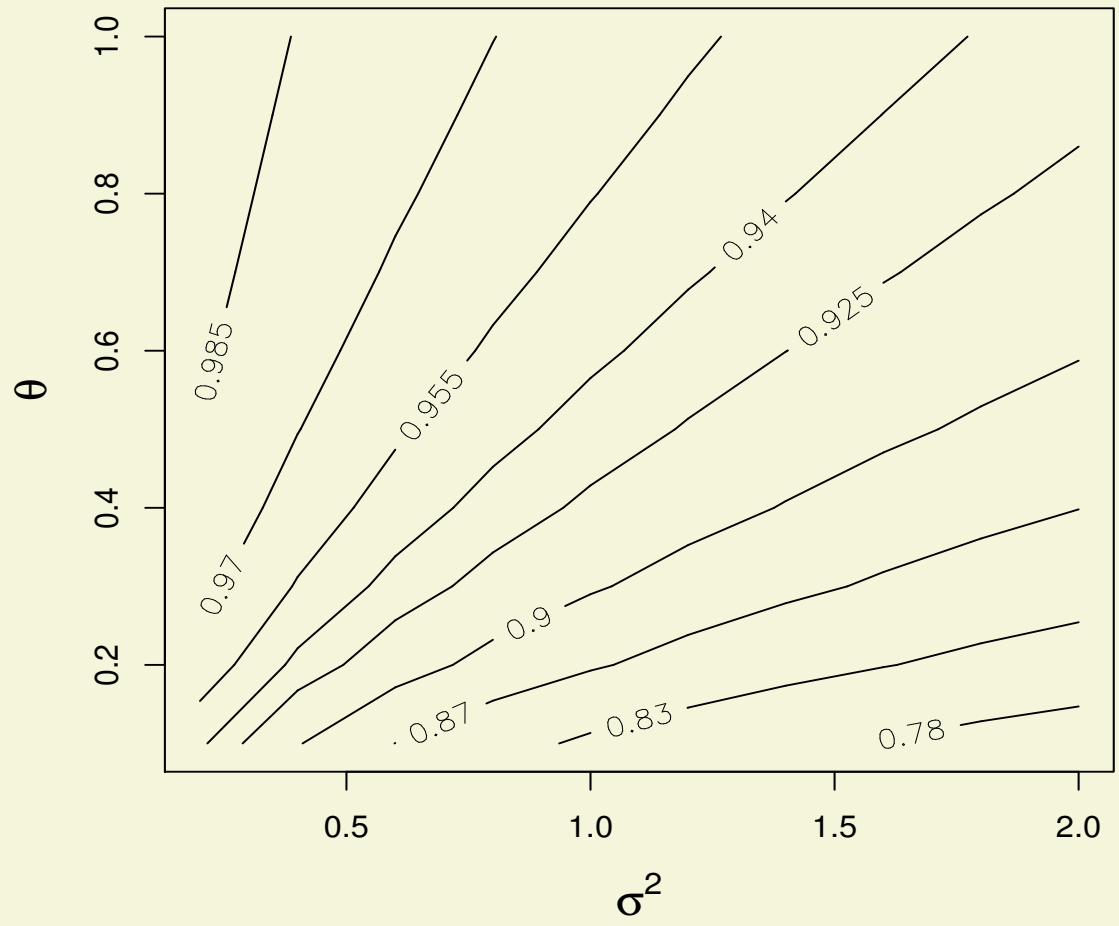


Figure 1: $\theta_2 = 0.5$

Findings

- Length reductions in the range 1–35%
- Length reductions decrease when confidence level increases
- Length reductions decrease when smoothness of the process increases
- The largest length reductions are obtained in models with
    - highly asymmetric marginals (σ^2 large)
    - moderate to weak dependence (θ small)

Shortest Prediction Intervals: Unknown Covariance

The previous prediction intervals depend on $\boldsymbol{\vartheta}$, which is unknown
The most immediate fix to this problem is to use $I_0^N(\alpha, \hat{\boldsymbol{\vartheta}})$ and $I_0^S(\alpha, \hat{\boldsymbol{\vartheta}})$, where $\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}(\mathbf{z})$ is an estimate of $\boldsymbol{\vartheta}$. These are called plug-in prediction intervals

Drawback: Plug-in PIs have coverage properties that differ from the nominal coverage properties, usually having smaller coverage than the desired coverage since these PIs intervals do not take into account the sampling variability of the parameter estimates

A solution: Calibrate these plug-in PIs [Cox (1975) and Beran (1990)]

Calibrated Prediction Intervals

The *coverage probability* function of $I_0^S(\alpha, \hat{\boldsymbol{\vartheta}})$ is

$$\pi_0(\alpha, \boldsymbol{\vartheta}) = P_{\vartheta}\{Z_0 \in I_0^S(\alpha, \hat{\boldsymbol{\vartheta}}(\mathbf{Z}))\}$$

We start by estimating $\pi_0(\cdot, \boldsymbol{\vartheta})$ with $\pi_0(\cdot, \hat{\boldsymbol{\vartheta}})$
The basic idea of calibrating plug-in prediction intervals is to find $\alpha_c \in (0, 1)$ for which it holds, exactly or approximately, that

$$\pi_0(\alpha_c, \hat{\boldsymbol{\vartheta}}) = 1 - \alpha$$

The calibrated prediction interval is $I_0^S(\alpha_c, \hat{\boldsymbol{\vartheta}})$, which by construction has coverage probability close to $1 - \alpha$

$\pi_0(\cdot, \hat{\boldsymbol{\vartheta}})$ is usually not available in closed form so it needs to be approximated

Bootstrap Calibration

Let $\pi_0^*(\cdot, \hat{\boldsymbol{\vartheta}})$ be a Monte Carlo estimate of $\pi_0(\cdot, \hat{\boldsymbol{\vartheta}})$

One way is simulate $(\mathbf{Z}_j^*, Z_{0j}^*)$, say B times, from the lognormal model with parameters $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\mathbf{z})$ and $\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}(\mathbf{z})$, and estimate $\pi_0(x, \hat{\boldsymbol{\vartheta}})$ with

$$\frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\{\hat{Y}_0(\boldsymbol{\vartheta}_j^*) - \Phi^{-1}(1-\gamma_{0j}^{\mathrm{opt}*})\hat{\sigma}_0(\boldsymbol{\vartheta}_j^*) < Y_{0j}^* < \hat{Y}_0(\boldsymbol{\vartheta}_j^*) + \Phi^{-1}(1-x+\gamma_{0j}^{\mathrm{opt}*})\hat{\sigma}_0(\boldsymbol{\vartheta}_j^*)\}$$

$Y_{0j}^* = \log(Z_{0j}^*)$, $\boldsymbol{\vartheta}_j^* = \hat{\boldsymbol{\vartheta}}(\mathbf{Z}_j^*)$ and $\gamma_{0j}^{\mathrm{opt}*} = \gamma_0^{\mathrm{opt}}(x, \boldsymbol{\vartheta}_j^*)$

A better way is to use 'Rao-Backwellization' based on the identity

$$\pi_0(\alpha, \boldsymbol{\vartheta}) = E_{\vartheta}\left\{\Phi\left(\frac{U_0(\alpha, \hat{\boldsymbol{\vartheta}}, \mathbf{Y}) - \eta_0(\mathbf{0}, \boldsymbol{\vartheta}, \mathbf{Y})}{\tau_0(\boldsymbol{\vartheta})}\right) - \Phi\left(\frac{L_0(\alpha, \hat{\boldsymbol{\vartheta}}, \mathbf{Y}) - \eta_0(\mathbf{0}, \boldsymbol{\vartheta}, \mathbf{Y})}{\tau_0(\boldsymbol{\vartheta})}\right)\right\}$$

$L_0(x, \boldsymbol{\vartheta}, \mathbf{Y}) = \hat{Y}_0(\boldsymbol{\vartheta}) - \Phi^{-1}(1-\gamma_0^{\mathrm{opt}})\hat{\sigma}_0(\boldsymbol{\vartheta})$

$U_0(x, \boldsymbol{\vartheta}, \mathbf{Y}) = \hat{Y}_0(\boldsymbol{\vartheta}) + \Phi^{-1}(1-x+\gamma_0^{\mathrm{opt}})\hat{\sigma}_0(\boldsymbol{\vartheta})$

Algorithm

Step 1. Compute the ML (or REML) estimate $\hat{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\vartheta}}(\mathbf{z})$ from the observed data $\mathbf{z}$

Step 2. Simulate B independent and identically distributed bootstrap samples $\{\mathbf{Y}_j^* : 1 \leq j \leq B\}$ from the Gaussian random field with $\boldsymbol{\beta} = \mathbf{0}$ and $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$

Step 3. For each $j = 1, \ldots, B$, compute the estimate $\boldsymbol{\vartheta}_j^* = \hat{\boldsymbol{\vartheta}}(\exp(\mathbf{Y}_j^*))$ based on the bootstrap sample $\mathbf{Y}_j^*$

Step 4. For each $\mathbf{s}_0 \in D$ where a PI is sought, compute $L_{0j}^* = L_0(x, \boldsymbol{\vartheta}_j^*, \mathbf{Y}_j^*)$ and $U_{0j}^* = U_0(x, \boldsymbol{\vartheta}_j^*, \mathbf{Y}_j^*)$. Then for $x \in (0, 1)$ estimate $\pi_0(x, \hat{\boldsymbol{\vartheta}})$ by

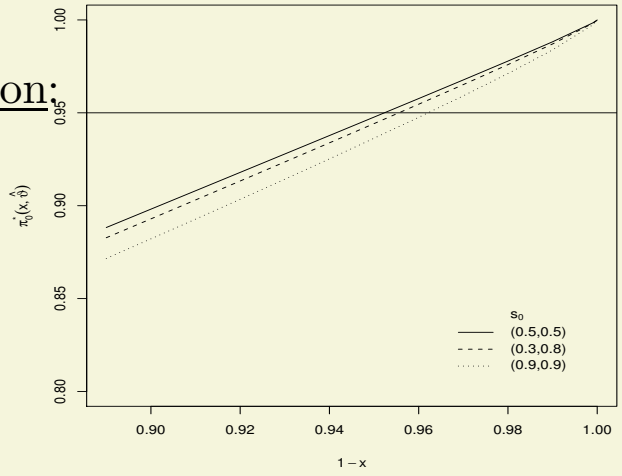
$$\pi_0^*(x, \hat{\boldsymbol{\vartheta}}) = \frac{1}{B} \sum_{j=1}^{B} \left[\Phi\left(\frac{U_{0j}^* - \hat{\eta}_{0j}^*}{\hat{\tau}_0} \right) - \Phi\left(\frac{L_{0j}^* - \hat{\eta}_{0j}^*}{\hat{\tau}_0} \right) \right]$$

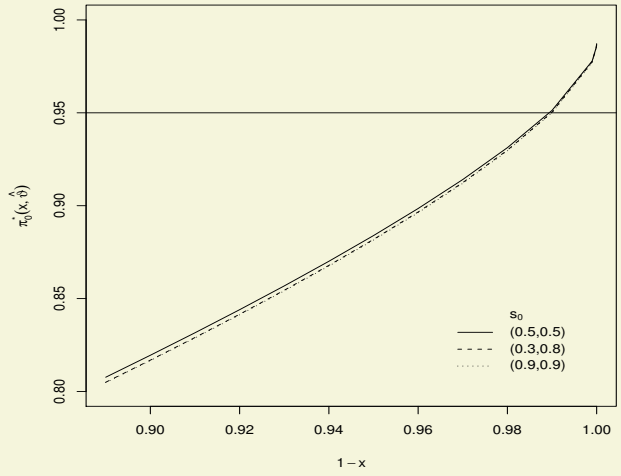
where $\hat{\eta}_{0j}^* = \eta_0(\mathbf{0}, \hat{\boldsymbol{\vartheta}}, \mathbf{Y}_j^*)$ and $\hat{\tau}_0 = \tau_0(\hat{\boldsymbol{\vartheta}})$

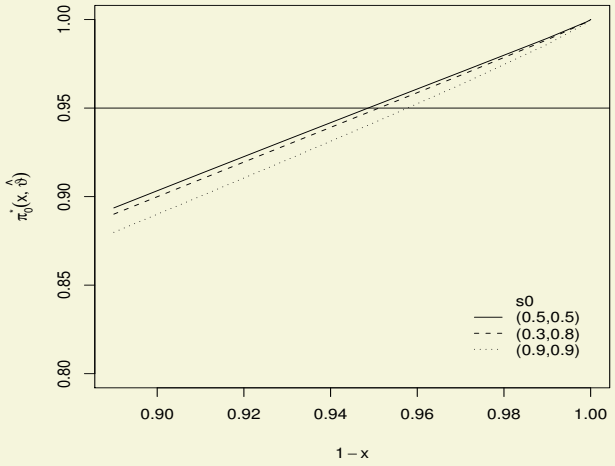
Finally, α_c is found as the solution (in x) of

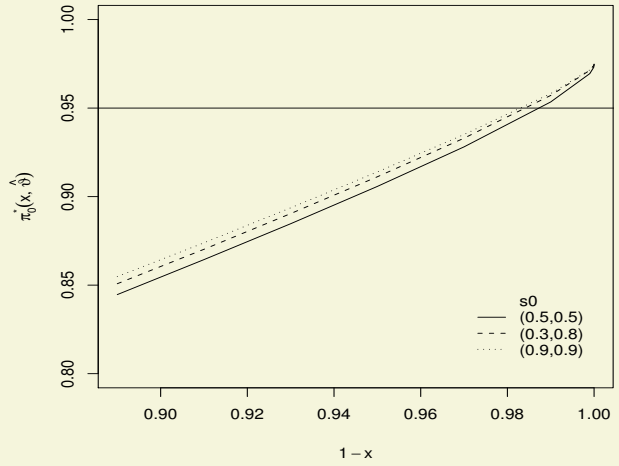
$$\pi_0^*(x, \hat{\boldsymbol{\vartheta}}) = (1 - \alpha)$$

Illustration:









When the data have no nugget effect of calibration tends to be minor

But when data contain measurement error effect of calibration tends to be substantial

σ^2	0.1		1.0	
θ	0.1	0.5	0.1	0.5
		$\sigma_\epsilon^2 = 0$		
plug-in standard	1.068 [.939]	0.551 [.951]	5.620 [.940]	3.004 [.948]
plug-in shortest	1.033 [.939]	0.545 [.950]	4.450 [.942]	2.779 [.948]
calibrated standard	1.124 [.949]	0.557 [.953]	6.111 [.951]	3.050 [.951]
calibrated shortest	1.068 [.946]	0.548 [.951]	4.694 [.949]	2.795 [.949]
		$\sigma_\epsilon^2 = \frac{\sigma^2}{4}$		
plug-in standard	1.011 [.899]	0.522 [.895]	6.131 [.899]	2.897 [.896]
plug-in shortest	0.981 [.900]	0.518 [.895]	4.908 [.909]	2.686 [.902]
calibrated standard	1.164 [.934]	0.544 [.908]	8.597 [.936]	3.041 [.910]
calibrated shortest	1.098 [.928]	0.531 [.904]	6.750 [.937]	2.775 [.911]
		$\sigma_\epsilon^2 = 0$		
plug-in standard	1.225 [.933]	0.727 [.936]	6.507 [.934]	4.025 [.933]
plug-in shortest	1.174 [.934]	0.716 [.937]	4.945 [.937]	3.537 [.936]
calibrated standard	1.313 [.948]	0.779 [.952]	7.367 [.949]	4.422 [.949]
calibrated shortest	1.232 [.944]	0.750 [.948]	5.400 [.947]	3.776 [.947]
		$\sigma_\epsilon^2 = \frac{\sigma^2}{4}$		
plug-in standard	1.175 [.891]	0.697 [.896]	6.960 [.890]	3.907 [.896]
plug-in shortest	1.129 [.894]	0.686 [.897]	5.345 [.903]	3.445 [.901]
calibrated standard	1.377 [.932]	0.761 [.923]	10.168 [.933]	4.390 [.923]
calibrated shortest	1.285 [.927]	0.731 [.917]	7.636 [.935]	3.755 [.921]

Findings

- Length reductions in the range 2–25%
- The largest length reductions are obtained in models with
  - highly asymmetric marginals (σ^2 large)
  - moderate to weak dependence (θ small)
- The effect of calibration is largest when data contain measurement error
- Calibration improves coverage, but sometimes falls a bit short

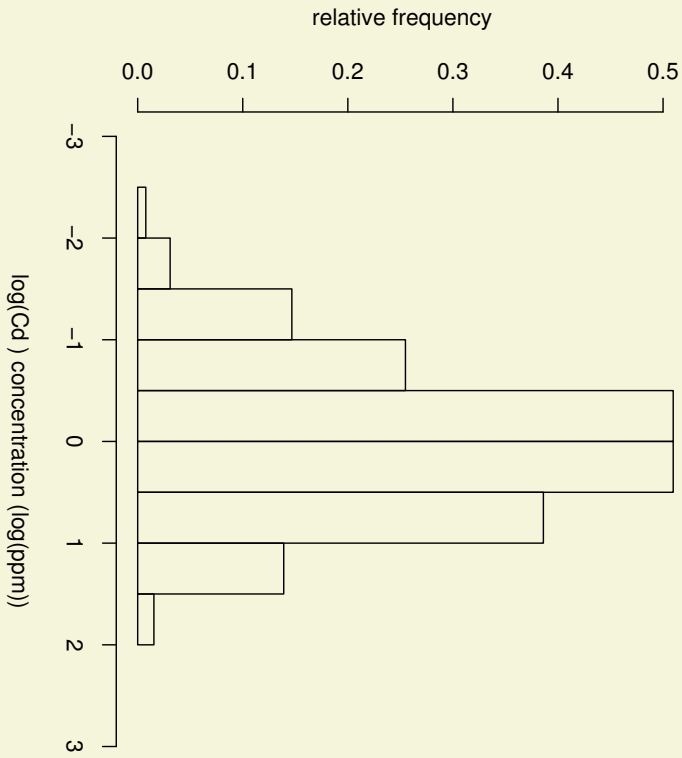
Data Example

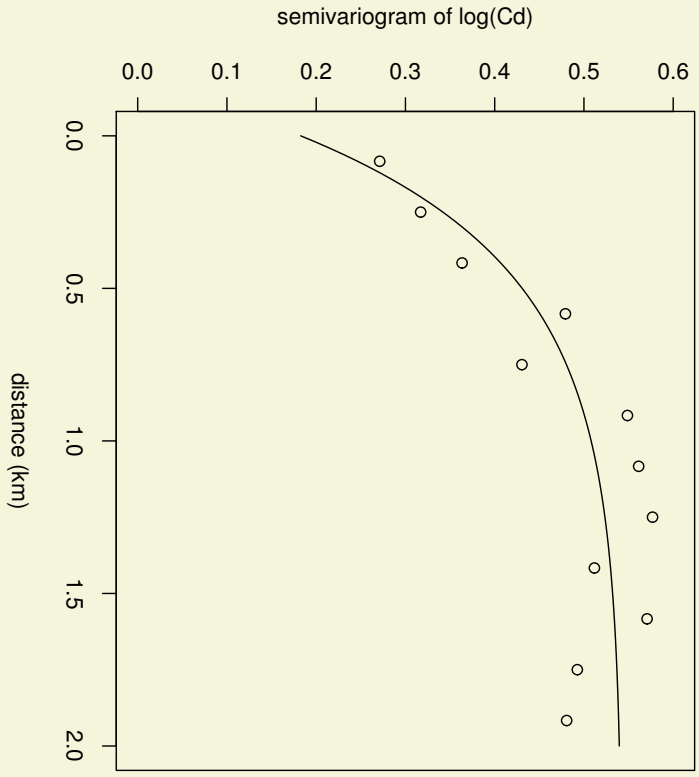
Data on cadmium (Cd) concentrations (in ppm) measured at 259 locations in a region of about 15 km^2 in the Switzerland, collected in 1992

Measurements at 100 locations are used for validation

Exploratory analysis suggests the "best" model is the log-Gaussian random field associated with constant mean, nugget and exponential covariance function:

$$\hat{\beta}_1 = 0.084, \quad \hat{\sigma}^2 = 0.4, \quad \hat{\theta}_1 = 0.177 \quad \text{and} \quad \hat{\sigma}^2_\epsilon = 0.073$$


relative frequency
0.0
0.1
0.2
0.3
0.4
0.5
log(Cd) concentration (log(ppm))
−3
−2
−1
0
1
2
3


semivariogram of log(Cd)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
distance (km)
0.0
0.5
1.0
1.5
2.0

	plug-in			calibrated		
	standard	shortest	RL	standard	shortest	RL
1	1.169	1.053	0.900	1.176	1.053	0.890
2	4.763	4.201	0.882	4.853	4.230	0.871
3	4.527	3.908	0.863	4.663	3.999	0.858
4	3.550	3.094	0.871	3.636	3.130	0.861
5	3.337	2.881	0.863	3.433	2.940	0.856
6	2.999	2.615	0.872	3.100	2.680	0.864
7	3.984	3.494	0.877	4.075	3.568	0.875
8	2.738	2.363	0.863	2.832	2.401	0.847
9	2.813	2.479	0.881	2.873	2.502	0.871
10	3.817	3.326	0.871	3.911	3.393	0.867
⋮	⋮	⋮	⋮	⋮	⋮	⋮

- Relative lengths of shortest prediction intervals w.r.t. standard prediction intervals vary from 0.84 to 0.89 (average of 14% length reduction)
- Proportion of 95% plug-in standard, plug-in shortest, calibrated standard and calibrated shortest prediction intervals covering the corresponding Cd observed value at 100 validation locations were, respectively, $0.93, 0.93, 0.93$ and 0.95
- Calibrated shortest prediction intervals seem to have coverage close to nominal