Particle Swarm Optimization for Spatial Design

Matthew Simpson[1]

Abstract

KEY WORDS:

[1](to whom correspondence should be addressed) Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100, themattsimpson@gmail.com

1 The Problem

Suppose we are interested in the latent spatial field of some response variable $Y(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{D} \subseteq \Re^2$. Specifically, we are interested in predicting $Y(\boldsymbol{s})$ at a set of target locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M \in \mathcal{D}$. We have the ability to sample N locations anywhere in $\mathcal{D}$, and we wish to place them in order to optimize some design criterion. Let $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N \in \mathcal{D}$ denote the N sampled locations. Suppose that $Y(\boldsymbol{s})$ is a geostatistical process with mean function $\mu(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}$ for some known covariate $\boldsymbol{x}(\boldsymbol{s})$ and covariance function $C(\boldsymbol{s}, \boldsymbol{t})$ for $\boldsymbol{s}, \boldsymbol{t} \in \mathcal{D}$. Once the design points are selected, we observe $Z(\boldsymbol{d}_i)$ for $i = 1, 2, \ldots, N$ where $Z(\boldsymbol{d}) = Y(\boldsymbol{d}) + \varepsilon(\boldsymbol{d})$ and $\varepsilon(\boldsymbol{d})$ is mean zero white noise with variance σ^2_ε, representing measurement error. An intuitive criterion to minimize is the average mean square prediction error (MSPE) from kriging across each of the target locations, each the errrors are weighted appropriately. It turns out that this is intimately related to a more principled design criterion, the expected entropy gain on the predictive distribution: $\mathrm{E}\left\{\log[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] - \log[Y(\boldsymbol{s}_0)]\right\}$. This is equivalent to the mutual information between $Y(\boldsymbol{s}_0)$ and $\boldsymbol{Z}$, which can be expressed as $M[Y(\boldsymbol{s}_0), \boldsymbol{Z}] = H[Y(\boldsymbol{s}_0)] + H[\boldsymbol{Z}] - H[Y(\boldsymbol{s}_0), \boldsymbol{Z}]$ where $H(.)$ is the entropy of the density function of the eclosed random variable.

1.1 Simple Kriging

In simple kriging, $C(.,.)$, $\boldsymbol{\beta}$, and σ^2_ε are all treated as known. Let $\boldsymbol{Z} = [Z(\boldsymbol{d}_1), Z(\boldsymbol{d}_2), \ldots, Z(\boldsymbol{d}_N)]'$, $\boldsymbol{X} = [\boldsymbol{x}(\boldsymbol{d}_1), \boldsymbol{x}(\boldsymbol{d}_2), \ldots, \boldsymbol{x}(\boldsymbol{d}_N)]'$, $\boldsymbol{C}_Z = \mathrm{cov}(\boldsymbol{Z})$ where $\mathrm{cov}[Z(\boldsymbol{d}_i), Z(\boldsymbol{d}_j)] = C(\boldsymbol{d}_i, \boldsymbol{d}_j) + \sigma^2_\varepsilon 1(\boldsymbol{d}_i = \boldsymbol{d}_j)$, and $\boldsymbol{c}_Y(\boldsymbol{s}_0) = \mathrm{cov}[Y(\boldsymbol{s}_0), \boldsymbol{Z}]$ where $\mathrm{cov}[Y(\boldsymbol{s}_0), Z(\boldsymbol{d}_i)] = C(\boldsymbol{s}_0, \boldsymbol{d}_i)$. Then the simple kriging predictor of $Y(\boldsymbol{s}_0)$ is the linear predictor $\boldsymbol{Y}^*(\boldsymbol{s}_0) = \boldsymbol{\lambda}'\boldsymbol{Z} + k$ that MSPE conditional on sampled locations, $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N$. Specifically it minimizes $\mathrm{E}[Y(\boldsymbol{s}_0) - \widehat{Y}_{sk}(\boldsymbol{s}_0)]^2$ over $\boldsymbol{\lambda}$ and k such that $\boldsymbol{\lambda}'\boldsymbol{1} = 1$ where $\boldsymbol{1}$ is a column vector of ones. The simple kriging predictor is easily derived as

$$\widehat{Y}_{sk}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta})$$

with MSPE

$$\sigma^2_{sk}(\boldsymbol{s}_0) = C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0).$$

The simple kriging MSPE a function of the design points, $\boldsymbol{D} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N)$, through $\boldsymbol{c}_Y(\boldsymbol{s}_0)$ and $\boldsymbol{C}_Z^{-1}$.

From a Bayesian perspective, this predictor can be rationalized by a particular Bayesian hierachical model (HM) that has a bit more structure. Specifically, we now assume that $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ and $\{\varepsilon(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ are independent Gaussian processes. Then the posterior predictive distribution for $Y(\boldsymbol{s}_0)$ can be derived as

$$[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] = \int [Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}][\boldsymbol{Y}|\boldsymbol{Z}]d\boldsymbol{Y}$$

where it is easy to derive that

$$\begin{aligned}
Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z} &\sim N\left[\boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right], \\
\boldsymbol{Y}|\boldsymbol{Z} &\sim N\left[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}), \boldsymbol{C}_Y - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y\right]
\end{aligned}$$

so that $Y(\boldsymbol{s}_0)|\boldsymbol{Z}$ is Gaussian with

$$\begin{aligned}
\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \mathrm{E}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} = \boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}), \\
\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} + \mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} \\
&= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0).
\end{aligned}$$

Then a design criterion we could minimize is average simple kriging MSPE across all target locations, $\sum_{i=1}^N \sigma^2_{sk}(\boldsymbol{s}_i)/N$. This is a bit naieve, however, because it does not take into account the covariance structure across the locations.

Let $\boldsymbol{Y}^* = [Y(\boldsymbol{s}_1, Y(\boldsymbol{s}_2, \ldots, Y(\boldsymbol{s}_M)]'$ denote the true vector of response variables at the target locations, $\{\boldsymbol{s}_k \in \mathcal{D} : i = 1, \ldots, M\}$. The covariance between the prediction errors at $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ can be derived as $\mathrm{cov}[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)|\boldsymbol{Z}] = C(\boldsymbol{s}_i, \boldsymbol{s}_j) - \boldsymbol{c}_Y(\boldsymbol{s}_i)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_j)$. Let $\boldsymbol{C}_{Y*} = \mathrm{cov}(\boldsymbol{Y}^*)$ where $\mathrm{cov}[Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)] = C(\boldsymbol{s}_i, \boldsymbol{s}_j)$ and $\boldsymbol{C}_{YY^*} = \mathrm{cov}(\boldsymbol{Y}, \boldsymbol{Y}^*)$ where the i, jth element is given by $\mathrm{cov}[Y(\boldsymbol{d}_i), Y(\boldsymbol{s}_j)] = C(\boldsymbol{d}_i, \boldsymbol{s}_j)$ for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, M$. Then the simple kriging error covariance matrix is given by $\boldsymbol{\Sigma}_{sk} = \boldsymbol{C}_{Y^*} - \boldsymbol{C}'_{YY^*}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{YY^*}$.

Then an intuitive design criterion to minimize is $\log|\boldsymbol{\Sigma}_{sk}|$, which is in some sense the total amount of variation in the distribution of the errors from predicting $\boldsymbol{Y}^*$. If we return to the Bayesian setting and instead maximize the expected entropy gain, we have $M(\boldsymbol{Y}^*, \boldsymbol{Z}) = H(\boldsymbol{Y}^*) + H(\boldsymbol{Z}) - H(\boldsymbol{Y}^*, \boldsymbol{Z})$. To derive these quantities, note that the joint distribution of $(\boldsymbol{Y}^*, \boldsymbol{Z})$ is Gaussian with mean $(\boldsymbol{X}^*\boldsymbol{\beta}, \boldsymbol{X}\boldsymbol{\beta}$ and covariance matrix

$$\mathrm{cov}[(\boldsymbol{Y}^*, \boldsymbol{Z})] = \begin{bmatrix} \boldsymbol{C}_{Y^*} & \boldsymbol{C}'_{YY^*} \\ \boldsymbol{C}_{YY^*} & \boldsymbol{C}_Z \end{bmatrix}$$

where $\boldsymbol{X}^* = [\boldsymbol{x}(\boldsymbol{s}_1), \boldsymbol{x}(\boldsymbol{s}_2), \ldots, \boldsymbol{x}(\boldsymbol{s}_M)]'$. The entropy of the p-variate normal distribution is $\log[(2\pi e)^p|\boldsymbol{\Sigma}|]/2$ where $\boldsymbol{\Sigma}$ is the covariance matrix. This yields

$$\begin{aligned} 2M(\boldsymbol{Y}^*, \boldsymbol{Z}) &= \log|\boldsymbol{C}_{Y^*}| + \log|\boldsymbol{C}_Z| - \log\left|\begin{bmatrix} \boldsymbol{C}_{Y^*} & \boldsymbol{C}'_{YY^*} \\ \boldsymbol{C}_{YY^*} & \boldsymbol{C}_Z \end{bmatrix}\right| \\ &= \log|\boldsymbol{C}_{Y^*}| - \log|\boldsymbol{C}_{Y^*} - \boldsymbol{C}'_{YY^*}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{YY^*}|. \end{aligned}$$

So maximizing the expected entropy gain is equivalent to minimizing $\log|\boldsymbol{C}_{Y^*} - \boldsymbol{C}'_{YY^*}\boldsymbol{C}_Z^{-1}\boldsymbol{C}_{YY^*}|$ since $\boldsymbol{C}_{Y^*}$ does not depend on the sampled points, and further the intuitive criterion coincides with the entropy criterion.

1.2 Universal Kriging

A major limitation of simple kriging is that typically $\boldsymbol{\beta}$, σ_ε^2, and $C(.,.)$ are unknown. Universal kriging attempts to remedy this by allowing $\boldsymbol{\beta}$ to be unknown. The universal kriging predictor is

$$\widehat{Y}_{uk}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{gls})$$

where $\widehat{\boldsymbol{\beta}}_{gls} = (\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{Z}$ is the generalized least squares estimate of $\boldsymbol{\beta}$, and the MSPE of $\widehat{Y}_{uk}(\boldsymbol{s}_0)$ is

$$\begin{aligned}\overline{S}^2_{uk}(\boldsymbol{s}_0) &= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)].\end{aligned}$$

This can also be justified via a Bayesian HM by using the same Gaussian assumptions as before, but further assuming that $\boldsymbol{\beta}$ has the improper uniform prior, i.e. $\boldsymbol{\beta} \sim U(-\infty, \infty)$. Let $\boldsymbol{C}_Y = \text{cov}(\boldsymbol{Y})$ where $\text{cov}(Y(\boldsymbol{d}_i), Y(\boldsymbol{d}_j)) = C(\boldsymbol{d}_i, \boldsymbol{d}_j)$. Now we have

$$[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] = \int\int [Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}][\boldsymbol{Y}|\boldsymbol{Z}]d\boldsymbol{Y}d\boldsymbol{\beta}$$

where

$$\begin{aligned}Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z} &\sim N\left[\boldsymbol{x}(\boldsymbol{s}_0)'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right], \\ \boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{Z} &\sim N\left[\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\boldsymbol{\beta}), \boldsymbol{C}_Y - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y\right], \\ \boldsymbol{\beta}|\boldsymbol{Z} &\sim N\left[\widehat{\boldsymbol{\beta}}_{gls}, (\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\right],\end{aligned}$$

so that $Y(\boldsymbol{s}_0)|\boldsymbol{Z}$ is again Gaussian with

$$\begin{aligned}\text{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \text{E}\left\{\text{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} = \boldsymbol{x}(\boldsymbol{s}_0)'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{gls}), \\ \text{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}] &= \text{E}\left\{\text{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} + \text{var}\left\{\text{E}[Y(\boldsymbol{s}_0)|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{Z}]|\boldsymbol{Z}\right\} \\ &= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) + \text{var}\left\{[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'\boldsymbol{\beta} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{Y}|\boldsymbol{Z}\right\}.\end{aligned}$$

To compute the last variance term, note that

$$\begin{aligned}\text{cov}(\boldsymbol{Y}|\boldsymbol{Z}) &= \text{E}[\text{cov}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{Z})|\boldsymbol{Z}] + \text{cov}[\text{E}(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{Z})|\boldsymbol{Z}] \\ &= \boldsymbol{C}_Y - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y + (\boldsymbol{I} - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{C}_Z^{-1}\boldsymbol{C}_Y)\end{aligned}$$

and

$$\begin{aligned}\mathrm{cov}(\boldsymbol{Y},\boldsymbol{\beta}|\boldsymbol{X}) &= \mathrm{E}[\mathrm{cov}(\boldsymbol{Y},\boldsymbol{\beta}|\boldsymbol{\beta},\boldsymbol{Z})|\boldsymbol{Z}] + \mathrm{cov}[\mathrm{E}(\boldsymbol{Y}|\boldsymbol{\beta},\boldsymbol{Z}),\boldsymbol{\beta}|\boldsymbol{Z}] \\ &= (\boldsymbol{I} - \boldsymbol{C}_Y\boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}.\end{aligned}$$

Then we have

$$\begin{aligned}&\mathrm{var}\left[Y(\boldsymbol{s}_0)|\boldsymbol{Z}\right] = \\ &\quad C(\boldsymbol{s}_0,\boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)] \\ &\quad + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\left\{\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1} + (\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\right\}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + \boldsymbol{c}_Y(\boldsymbol{s}_0)'(\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)] \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Y^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'(\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{x}(\boldsymbol{s}_0)'(\boldsymbol{C}_Y^{-1} - \boldsymbol{C}_Z^{-1})\boldsymbol{c}_Y(\boldsymbol{s}_0),\end{aligned}$$

which reduces to

$$\begin{aligned}\mathrm{var}\left[Y(\boldsymbol{s}_0)|\boldsymbol{Z}\right] &= C(\boldsymbol{s}_0,\boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\ &\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)].\end{aligned}$$

Then the design criterion in this case is $\overline{S}_{uk}^2 = \sum_{i=1}^N S_{uk}^2(\boldsymbol{s}_i)/N$, and once again $\mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]\right\} = \mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]$, which is constant in $\boldsymbol{Z}$.

1.3 Full Uncertainty Kriging

In both simple and universal kriging the predictors and their MSPEs are the same as the posterior predictive means and posterior predictive variances for a suitably chosen Bayesian HM. In both cases if we wish to choose the sampled locations, $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N$, in order to minimize the MSPE from the kriging estimate, we are effectively minimizing $\sum_{i=1}^N \mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_i)|\boldsymbol{Z}(\boldsymbol{D})]\right\}/N$ in $\boldsymbol{D}$. We can easily extend this to the case where both σ_ε^2

and $C(.,.)$ are unknown, though it may no longer correspond to the optimal linear predictor. Suppose the covariance function is parameterized by $\boldsymbol{\theta} \in \Theta$, denoted by $C_{\boldsymbol{\theta}}(.,.)$, and let $\boldsymbol{\phi} = (\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$. Assume that a priori $\boldsymbol{\phi} \sim [\boldsymbol{\phi}]$ — we relax the restriction that $\boldsymbol{\beta} \sim U(-\infty, \infty)$. Then we can write the expected posterior predictive variance of $Y(\boldsymbol{s}_0)$ as

$$\mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{Z}]\right\} = \mathrm{E}\Bigg(\mathrm{E}\left\{\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}]|\boldsymbol{Z}\right\}\Bigg) + \mathrm{E}\Bigg(\mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}]|\boldsymbol{Z}\right\}\Bigg)$$

where from the universal kriging case we know

$$\begin{aligned}
\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}] &= \left\{\left[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right](\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}' + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\right\}\boldsymbol{C}_Z^{-1}\boldsymbol{Z}, \\
\mathrm{var}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}] &= C(\boldsymbol{s}_0, \boldsymbol{s}_0) - \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0) \\
&\quad + [\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)]'[\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X}][\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)].
\end{aligned}$$

[PROBLEM: $\mathrm{var}\left\{\mathrm{E}[Y(\boldsymbol{s}_0)|\boldsymbol{\theta}, \sigma_\varepsilon^2, \boldsymbol{Z}]|\boldsymbol{Z}\right\}$ IS NOT AVAILABLE IN CLOSED FORM, SO MAYBE WE DO JUST MINIMIZE THE A PRIORI EXPECTED UNIVERSAL KRIGING VARIANCE?]