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 Introduction

2 Model

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 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 locations of the N monitors. We will assume the universal kriging setup with only location as a predictor. In otherwords, we assume the $Y(\boldsymbol{s})$ is a geostatistical process with mean function $\mu(\boldsymbol{s})$ and covariance function $C_Y(\boldsymbol{s}, \boldsymbol{t})$ for $\boldsymbol{s}, \boldsymbol{t} \in \mathcal{D}$. Further, we assume that $\mu(\boldsymbol{s}) = \boldsymbol{x}'(\boldsymbol{s})\boldsymbol{\beta}$ where if $\boldsymbol{s} = (u, v)$, then $\boldsymbol{x}'(\boldsymbol{s}) = (1, u, v)$. Finally, 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.

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 $\mathrm{cov}(Z(\boldsymbol{d}_i), Z(\boldsymbol{d}_j)) = C_Y(\boldsymbol{d}_i, \boldsymbol{d}_j) + \sigma^2_\varepsilon 1(\boldsymbol{d}_i = \boldsymbol{d}_j)$, and $\boldsymbol{c}_Y(\boldsymbol{s}_0) = \mathrm{cov}(Y(\boldsymbol{s}_0), \boldsymbol{Z})$ where $\mathrm{cov}(Y(\boldsymbol{s}_0), Z(\boldsymbol{d}_i)) = C_Y(\boldsymbol{s}_0, \boldsymbol{d}_i)$. Then the generalized least squares estimator of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}_{gls} = (\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}_Z^{-1}\boldsymbol{Z}$, the universal kriging predictor of $Y(\boldsymbol{s}_0)$ is $\widehat{Y}(\boldsymbol{s}_0) = \boldsymbol{x}(\boldsymbol{s}_0)'\widehat{\boldsymbol{\beta}}_{gls} + \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{gls})$, and its mean square prediction error is

$$
\begin{aligned}
&\sigma^2_{\widehat{Y}}(\boldsymbol{d}; \boldsymbol{s}_0) = \boldsymbol{c}_Y(\boldsymbol{s}_0)'\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{c}(\boldsymbol{s}_0) + \\
&\quad \left[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}(\boldsymbol{d})\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right]' \left[\boldsymbol{X}(\boldsymbol{d})'\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{X}(\boldsymbol{d})\right]^{-1} \left[\boldsymbol{x}(\boldsymbol{s}_0) - \boldsymbol{X}(\boldsymbol{d})\boldsymbol{C}_Z(\boldsymbol{d})^{-1}\boldsymbol{c}_Y(\boldsymbol{s}_0)\right]
\end{aligned}
$$

where $\boldsymbol{d} = (\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_N)'$ and $\boldsymbol{d}_i = (u_i, v_i)'$ for $i = 1, 2, \ldots, N$. Then the design criterion is

$$
U(\boldsymbol{d}) = \frac{1}{M}\sum_{j=1}^{M} \sigma^2_{\widehat{Y}}(\boldsymbol{d}; \boldsymbol{s}_j)
$$

where $\{\boldsymbol{s}_j\}$ are the M locations we wish to predict and our goal is to minimize U in $\boldsymbol{d}$.

Alternatively if we wish to learn about the entire spatial domain, we can minimize

$$U_C(\boldsymbol{d}) = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} \sigma^2_{\widehat{Y}}(\boldsymbol{d};\boldsymbol{s})d\boldsymbol{s},$$

though this integral is unlikely to be available in closed form and so in practice we would approximate with a criterion with the form of $U(\boldsymbol{d})$. We could also modify $U(\boldsymbol{d})$ by attaching weights to the spatial locations if some locations are more important than others.

We can consider two versions of this optimizatio problem: when $M > N$, i.e. we want to predict at more locations than we can observe, or the opposite case when $M < N$. When $M < N$, it is sensible to restrict ourself to designs where the first M observed locations are exactly the M locations at which we want to predict [CAN WE PROVE THIS?]. When $M > N$, it is no longer necessarily the case that putting a design location at a prediction location is a good idea.

[SHOULD WE CONSIDER OTHER OBJECTIVE FUNCTIONS? SOMETHING DEPENDING ON ENTROPY?]

Also note that $U(\boldsymbol{d})$ depends on the covariance function, $C_Y(\boldsymbol{s},\boldsymbol{t})$, which may depend on unknown parameters. In that case, we can put a prior on those unknown parameters and instead minimize $E_{\boldsymbol{\theta}}[U(\boldsymbol{d};\boldsymbol{\theta})] = \int_\Theta U(\boldsymbol{d};\boldsymbol{\theta})[\boldsymbol{\theta}]d\boldsymbol{\theta}$ where $[\boldsymbol{\theta}]$ is the prior on $\boldsymbol{\theta}$. [CONNECTION TO FACT THAT KRIGING CAN BE DERIVED FROM A BAYESIAN HIERARCHICAL LINEAR MODEL]