Spatial Prediction: An Exhibition

Dale Zimmerman

Department of Statistics and Actuarial Science
University of Iowa

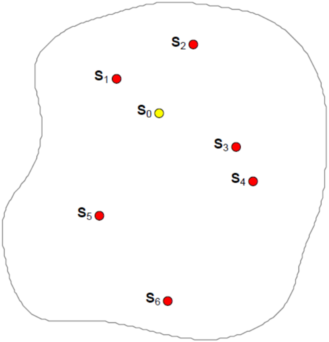
May 6, 2018

Overview

1. Classical spatial prediction (kriging)
2. Some unsurprising features of kriging, with exhibits
3. Some surprising(?) features of kriging, with exhibits
4. Optimal designs for spatial prediction, with exhibits

Classical spatial prediction (kriging)

Available: n observations $Y_1, \ldots, Y_n$ of a variable, taken at distinct sites $\mathbf{s}_i \in D$ $(i = 1, \ldots, n)$, written alternatively as $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$.



Objectives:

1. To predict a value of the same variable at site $\mathbf{s}_0$, i.e., $Y(\mathbf{s}_0)$
2. To characterize the uncertainty of the prediction

The problem is also called *spatial interpolation* by some.

Spatial prediction methods

Many different spatial prediction methods have been proposed:

- Deterministic
  - Inverse distance weighting
  - Interpolating polynomials
  - Splines and other forms of nonparametric smoothing
  - Others (e.g., radial basis function neural networks, . . .)
- Stochastic
  - Least squares regression on spatial coordinates or functions thereof with iid errors (trend surface analysis)
  - Kriging (ordinary, universal, and other types)

Kriging methods are based on a random field model for the observations and predictand(s).

Best unbiased prediction under a random field model

Best unbiased prediction:

- $\mathbf{Y} = [Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)]'$ comprises an observed sample of one realization of a continuously-indexed random process, or *random field*, $Y(\cdot) \equiv \{Y(\mathbf{s}) : \mathbf{s} \in D\}$
- $Y(\mathbf{s}_0)$ is an unobserved member of the same realization
- Assume $Y(\cdot)$ has finite second moments
- Among all predictors $p(\mathbf{Y})$ with finite second moments, the conditional mean

$$E[Y(\mathbf{s}_0)|\mathbf{Y}]$$

minimizes the MSPE $E\{[p(\mathbf{Y}) - Y(\mathbf{s}_0)]^2\}$.

BLUP under a random field model

Best linear unbiased prediction (BLUP) (Goldberger, 1962; Matheron, 1962):

Reduce required knowledge to that of the

- Mean function $m(\mathbf{s}) = E[Y(\mathbf{s})]$, and
- Covariance function $C(\mathbf{s}, \mathbf{u}) = \text{Cov}[Y(\mathbf{s}, \mathbf{u})]$,

and further assume that $m(\cdot)$ is a linear function of unknown parameters, i.e.,

$$m(\mathbf{s}; \boldsymbol{\beta}) = \sum_{j=1}^{p} \beta_j f_j(\mathbf{s})$$

where $f_1(\mathbf{s}) \equiv 1$.

BLUP under a random field model

Then the best linear unbiased predictor (BLUP) of $Y(\mathbf{s}_0)$ is

$$\hat{Y}_K(\mathbf{s}_0) = [\mathbf{c} + \mathbf{X}(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})]'\mathbf{C}^{-1}\mathbf{Y} = \sum_{i=1}^{n} \lambda_i Y(\mathbf{s}_i),$$

say, where

$$\mathbf{X} = [f_j(\mathbf{s}_i)], \qquad \mathbf{x} = [f_j(\mathbf{s}_0)],$$
$$\mathbf{C} = [C(\mathbf{s}_i, \mathbf{s}_j)], \qquad \mathbf{c} = [C(\mathbf{s}_i, \mathbf{s}_0)].$$

The minimized MSPE associated with $\hat{Y}_K(\mathbf{s}_0)$ is given by

$$\begin{aligned} \sigma_K^2(\mathbf{s}_0) &= \mathrm{var}(\hat{Y}_K(\mathbf{s}_0) - Y(\mathbf{s}_0)) \\ &= C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}'\mathbf{C}^{-1}\mathbf{c} \\ &\quad + (\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c})'(\mathbf{X}'\mathbf{C}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\mathbf{C}^{-1}\mathbf{c}). \end{aligned}$$

If $Y(\cdot)$ is Gaussian, $\hat{Y}_K(\mathbf{s}_0)$ and $E[Y(\mathbf{s}_0)|\mathbf{Y}]$ coincide.

BLUP under a random field model

$\hat{Y}_K(\mathbf{s}_0)$ and $\sigma^2_K(\mathbf{s}_0)$ are called the *kriging predictor* and *kriging variance* (*ordinary* if mean is constant, *universal* otherwise)

For pragmatic reasons, the following additional assumptions are often made about the covariance function:

1. It's not completely known; rather, it is known up to the value of a vector of unknown parameters $\boldsymbol{\theta}$; so we write it as $C(\mathbf{s}, \mathbf{u}; \boldsymbol{\theta})$.
2. It's *second-order stationary*, i.e., $C(\mathbf{s}, \mathbf{u}; \boldsymbol{\theta})$ depends on $\mathbf{s}$ and $\mathbf{u}$ only through $\mathbf{h} = \mathbf{s} - \mathbf{u}$, and we write $C(\mathbf{s}, \mathbf{u}; \boldsymbol{\theta}) = C(\mathbf{h}; \boldsymbol{\theta})$.
3. It's *isotropic*, i.e., it's second-order stationary plus $C(\mathbf{h}; \boldsymbol{\theta})$ depends on $\mathbf{h}$ only through $h = \|\mathbf{h}\| = (\mathbf{h}'\mathbf{h})^{1/2}$, and we write $C(\mathbf{h}; \boldsymbol{\theta}) = C(h; \boldsymbol{\theta})$.

Second-order stationary covariance functions

A parametric second-order stationary covariance function must satisfy two mathematical requirements, for all $\boldsymbol{\theta}$ in a parameter space $\boldsymbol{\Theta}$:

1. Evenness, i.e.

$$C(\mathbf{h};\boldsymbol{\theta}) = C(-\mathbf{h};\boldsymbol{\theta}) \text{ for all } \mathbf{h}.$$

2. Positive definiteness, i.e.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(\mathbf{s}_i - \mathbf{s}_j;\boldsymbol{\theta}) \geq 0$$

for all n, all sequences $\{a_i\colon\ i = 1, \ldots, n\}$, and all sequences of spatial locations $\{\mathbf{s}_i \in D\colon\ i = 1, \ldots, n\}$.

Second-order stationary covariance functions

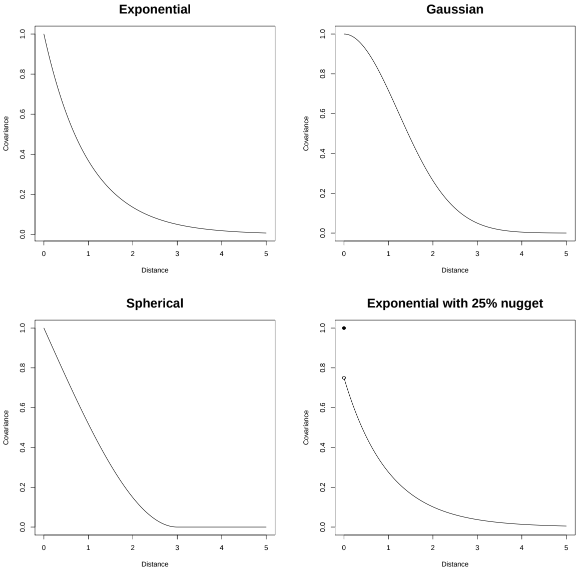
Where do we find such functions?

Bochner's Theorem from analysis tells us, in effect, that any real-valued characteristic function of a d-dimensional random vector is even and nonnegative definite, and thus any real-valued d-dimensional characteristic function could serve as a *valid* covariance function in R^d.

Examples:

- Exponential, $C(h;\boldsymbol{\theta}) = \theta_1 e^{-h/\theta_2}$
- Gaussian, $C(h;\boldsymbol{\theta}) = \theta_1 e^{-h^2/\theta_2}$
- Spherical, $C(h;\boldsymbol{\theta}) = \theta_1 \left(1 - \frac{3h}{2\theta_2} + \frac{h^3}{2\theta_2^3}\right) I(h \leq \theta_2)$

Some isotropic covariance functions



Empirical BLUP, or Kriging in practice

- Because we don't know $\boldsymbol{\theta}$ in practice, $\mathbf{Y}$ is used to estimate it, and the BLUP and MSPE formulas are used with this estimate substituted for $\boldsymbol{\theta}$. This is called *empirical BLUP*, or *E-BLUP*.
- This approach assumes *ergodicity*, i.e., the ability to consistently estimate $C(h)$ at some lags h from observed pairs $\{Y(\mathbf{s}), Y(\mathbf{t}) : \|\mathbf{s} - \mathbf{t}\| = h\}$ within the sampled realization.
- If $Y(\cdot)$ is Gaussian, a sufficient condition for ergodicity is $\lim_{h\to\infty} C(h; \boldsymbol{\theta}) \to 0$.
- If $Y(\cdot)$ is Gaussian, REML is widely used to estimate $\boldsymbol{\theta}$.
- If $Y(\cdot)$ is Gaussian, the MSPE of the E-BLUP is at least as large as that of the BLUP (Harville & Jeske, 1992), and we may more accurately approximate the E-BLUP's MSPE by

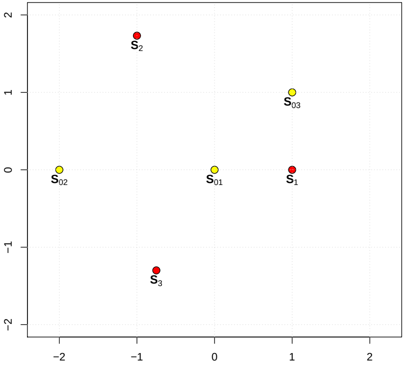
$$\sigma_K^2(\mathbf{s}_0) + tr[\mathbf{A}(\mathbf{s}_0; \boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})].$$

Some unsurprising features of kriging

Recall that the kriging predictor (BLUP) is a linear combination of the available data, $Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)$; thus each observation has a "weight" in the predictor. Now we exhibit effects that various characteristics of a particular kriging situation have on these weights and on the kriging variance.

We do this using a series of toy examples, in which the mean function is taken to be constant and the reference covariance function is taken to be isotropic exponential with unit variance and range parameter 1.0 ($\Rightarrow$ "effective range" $= 3.0$), i.e., $C(h) = e^{-h}$, $h \geq 0$.

Exhibit #1: Effect of distance



$\mathbf{s}_0$	kriging weights λ_1	λ_2	λ_3	σ^2_{OK}
(0, 0)	0.462	0.233	0.305	0.876
(-2, 0)	0.247	0.366	0.386	1.141
(1, 1)	0.533	0.278	0.188	0.967

Exhibit #2: Effect of distance on kriging variance

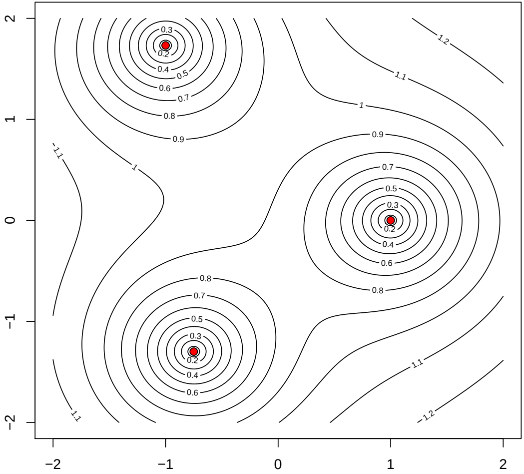
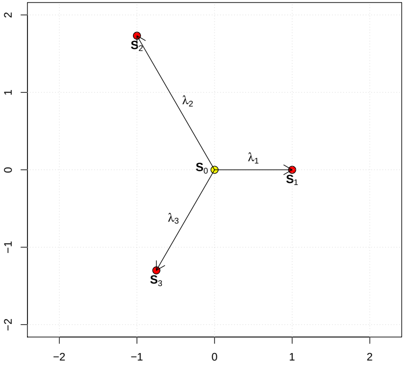


Exhibit #3: Effect of type of covariance function



	kriging weights			
Covariance	λ_1	λ_2	λ_3	σ^2_{OK}
exponential	0.462	0.233	0.305	0.876
Gaussian	0.583	0.132	0.285	0.351
spherical	0.524	0.177	0.299	0.648

Exhibit #4: Effect of type of covariance function on prediction of entire realization

Prediction over $[0, 1]$ from 10 observations. Top — using spherical covariance function; bottom — using Gaussian covariance function; both with same variance and correlation range:

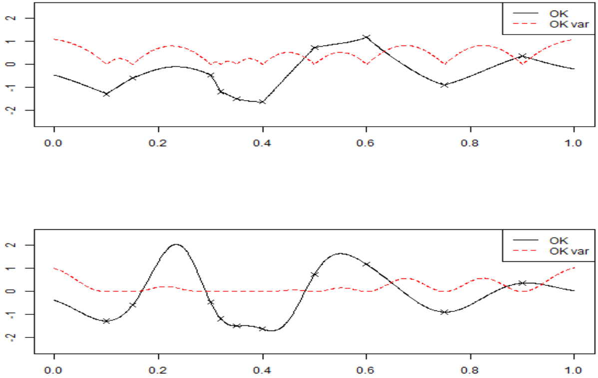
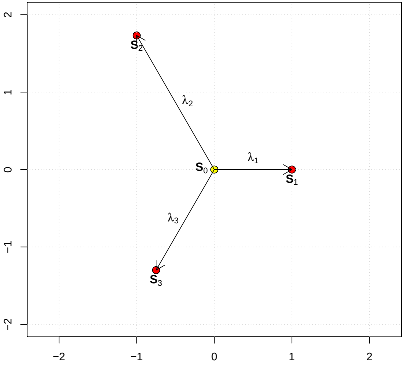


Exhibit #5: Effect of nugget effect



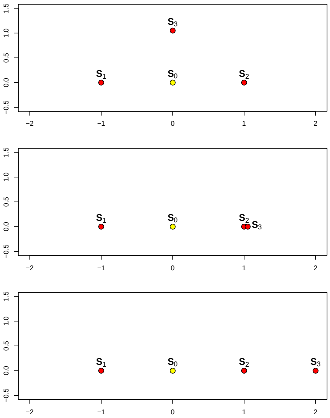
Nugget	kriging weights λ_1	λ_2	λ_3	σ^2_{OK}
0%	0.462	0.233	0.305	0.876
25%	0.427	0.259	0.314	0.995
50%	0.394	0.285	0.321	1.111

Some surprising(?) features of kriging

So far, everything about spatial BLUP makes perfect sense. Next, we consider two features that are perhaps somewhat surprising. They are:

1. The screening effect
2. Perfect interpolation

Exhibit #6: The screening effect



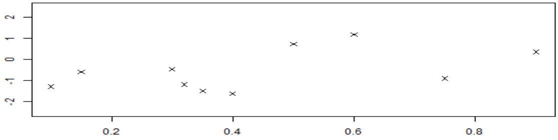
$\mathbf{s}_3$	kriging weights λ_1	λ_2	λ_3	σ^2_{OK}
(0, 1.05)	0.357	0.357	0.287	0.741
(1.05, 0)	0.498	0.402	0.101	0.831
(2, 0)	0.463	0.421	0.116	0.817

The screening effect

- The ith kriging weight is closely related to the partial correlation between $Y(\mathbf{s}_0)$ and $Y(\mathbf{s}_i)$, adjusted for the remaining observations (conditional correlation if $Y(\cdot)$ is Gaussian)
- These partial correlations depend on the spatial configuration in ways that are not easy to characterize
- The screening effect has been used to advantage to (a) reduce computational burden by using a moving kriging neighborhood, and (b) mitigate nonstationarity

Exhibit #7: "Perfect" interpolation

The BLUP at any data location is merely the observation at that location, and the MSPE there is zero. Example with 10 observations on unit interval in 1-D, constant mean, exponential covariance function with range parameter 0.05:



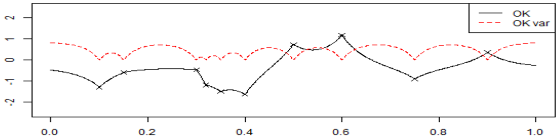
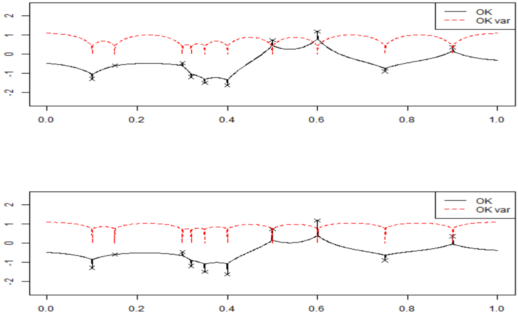


Exhibit #8: "Perfect" interpolation when there's a nugget effect

The phenomenon persists even if the covariance function has a nugget effect (top panel 25% nugget; bottom panel 50% nugget):



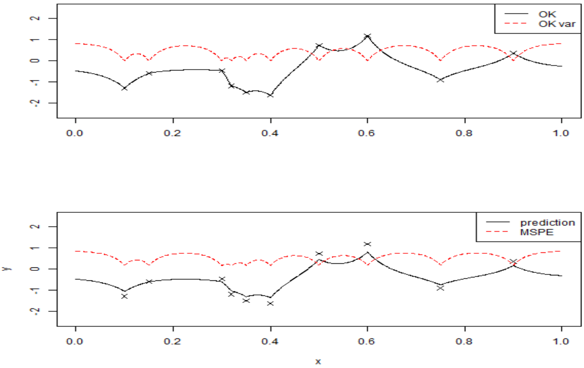
Geologists and many other kriging practitioners view perfect interpolation favorably, as it fully "honors the data."

Exhibit #9: Noiseless prediction

If we prefer to predict a noiseless version of $Y(\mathbf{s}_0)$, i.e., $\boldsymbol{\beta}_1 + W(\mathbf{s}_0)$ in the decomposition

$$Y(\mathbf{s}_0) = \beta_1 + W(\mathbf{s}_0) + \epsilon(\mathbf{s}_0)$$

where $W(\cdot)$ is mean-square continuous and $\epsilon(\cdot)$ is white noise, we obtain intuitively reasonable results (bottom panel is noiseless prediction):

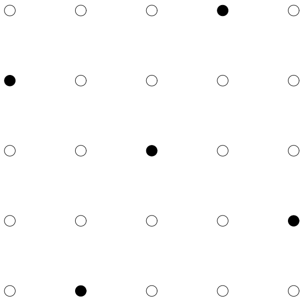


Optimal design for spatial prediction

- The quality of kriging is affected substantially by the spatial configuration of observation sites
- The "Equivalence Theorem" relating optimal design for prediction to optimal design for regression parameter estimation does not apply when the observations are correlated
- Compared to independence, spatial correlation improves prediction but worsens regression parameter estimation
- Work on optimal design for kriging has focused on choosing sites to minimize the maximum or average kriging variance over the region of interest (doesn't depend on data, but does depend on covariance function)
- Best designs using these criteria are "regular" (or nearly so) if the mean is constant, with relatively more points near the periphery of the region of interest if the mean is planar

Exhibit #10: Optimal design for BLUP with known covariance function

A member of the equivalence class of 5-point designs that minimize the maximum ordinary kriging variance over a 5×5 grid, when $C(h) = 0.5^h$ (in units of grid spacing):

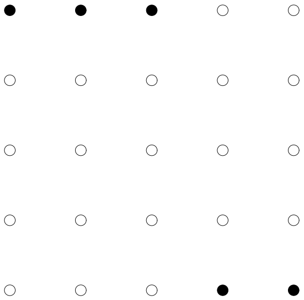


Optimal design for estimating the covariance function

- Since $C(h; \boldsymbol{\theta})$ must be estimated in practice, a total emphasis on prediction under an assumption of a known covariance function seems misplaced
- Zhu & Stein (2005) and Zimmerman (2006) proposed choosing a design to minimize $|\mathbf{B}(\boldsymbol{\theta})|$, the determinant of the inverse of the information matrix associated with the REML estimator of $\boldsymbol{\theta}$
- Best designs using this criterion are very different; they typically consist of several clusters or linear strands, dispersed rather uniformly throughout the study region

Exhibit #11: Optimal design for estimating the covariance function

A member of the equivalence class of 5-point designs that minimize $|\mathbf{B}(\boldsymbol{\theta})|$ when $C(h) = 0.5^h$ (in units of grid spacing):



Optimal design for E-BLUP

- Neither of the previous two design approaches address the problem of most practical interest — the quality of E-BLUP
- Zimmerman (2006) proposed choosing a design to minimize the Harville & Jeske (1992) approximation to the E-BLUP's MSPE,

$$\sigma_K^2(\mathbf{s}_o; \boldsymbol{\theta}) + tr[\mathbf{A}(\mathbf{s}_0; \boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})]$$

  where $\mathbf{A}(\mathbf{s}_0; \boldsymbol{\theta}) = \text{Var}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \hat{Y}_K(\mathbf{s}_0)\right]$
- Best designs using this criterion tend to be hybrids of the previous two: mostly regularly spaced with a few clusters

Exhibit #12: Optimal design for E-BLUP

A member of the equivalence class of 5-point designs that minimize the maximum approximate E-BLUP ordinary kriging variance over a 5×5 grid, when $C(h) = 0.5^h$ (in units of grid spacing):