NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

09-13

Local Spatial-Predictor Selection

Jonathan R. Bradley, Noel Cressie and Tao Shi

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email:
anica@uow.edu.au

Local Spatial-Predictor Selection

Jonathan R. Bradley * Noel Cressie * † Tao Shi *

Abstract
Consider the problem of spatial prediction of a random process from a spatial dataset. Global spatial-predictor selection provides a way to choose a single spatial predictor from a number of competing predictors. Instead, we consider local spatial-predictor selection at each spatial location in the domain of interest. This results in a hybrid predictor that could be considered global, since it takes the form of a combination of local predictors; we call this the locally selected spatial predictor. We pursue this idea here using the (empirical) deviance information as our criterion for (global and local) predictor selection. In a small simulation study, the relative performance of this combined predictor, relative to the individual predictors, is assessed.

Key Words: information criteria; generalized degrees of freedom; best linear unbiased predictor; model averaging; model combination

1. Introduction

In this proceedings paper, we consider the use of the empirical deviance information criterion to select from multiple predictors over a lattice $D_s \equiv \{\mathbf{u}_1, ..., \mathbf{u}_N\} \subset \mathbb{R}^d$. Let observed and potential data have the following additive structure:

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \epsilon(\mathbf{s}); \ \mathbf{s} \in D_s, \tag{1}$$

where $Z(\mathbf{s})$ represents a datum at a spatial location $\mathbf{s}$, $Y(\cdot)$ is a *hidden* process, and $\epsilon(\cdot)$ is a measurement-error process independent of $Y(\cdot)$. The process $Y(\cdot)$ is the source of spatial dependence in the data, but its probability distribution is left unspecified at this point. We observe $Z(\cdot)$ at locations $\{\mathbf{s}_1, ..., \mathbf{s}_n\} \subset D_s$, and we define $\mathbf{Z}_O \equiv (Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n))'$ and $\mathbf{Y}_O \equiv (Y(\mathbf{s}_1), ..., Y(\mathbf{s}_n))'$. We assume that $\epsilon(\mathbf{s}_i) \sim \text{Gau}(0, \sigma^2_\epsilon v(\mathbf{s}_i))$, where both $v(\cdot)$ and σ^2_ϵ are known and that for $\mathbf{s}_i \neq \mathbf{s}_j$, $\epsilon(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_j)$ are independent.

There are many methods of spatial prediction, both stochastic and non-stochastic, available in the literature; see, for example, Cressie (1993, Section 5.9). In what follows, we consider K spatial predictors of the form,

$$\hat{Y}^{(k)}(\mathbf{s}; \boldsymbol{\theta}) \equiv \mathbf{H}^{(k)}(\mathbf{s}; \boldsymbol{\theta})' \mathbf{Z}_O; \ \ k = 1, ..., K, \tag{2}$$

where $\mathbf{H}^{(k)}(\mathbf{s}; \boldsymbol{\theta})$ is an n-dimensional vector, and $\boldsymbol{\theta}$ represents unknown parameters that are to be estimated. The predictors given by (2) are linear predictors when $\boldsymbol{\theta}$ is known. For example, the best linear unbiased (i.e., universal kriging) predictor can be written in this form (e.g., Cressie, 1993, p. 152), as can the smoothing-spline predictor (Wahba, 1990). When $\boldsymbol{\theta}$ is estimated from the data $\mathbf{Z}_O$ and substituted into $\mathbf{H}^{(k)}$, the predictor may become nonlinear.

Bradley et al. (2011) considered the case of selecting basis functions for a random-effects model using the *empirical deviance information criterion* (DIC_E), where they argued that DIC_E can be interpreted as an estimate of the mean of the total squared prediction

*Department of Statistics, The Ohio State University, Contact: bradley.324@stat.osu.edu
†Centre for Statistical and Survey Methodology, University of Wollongong

error summed over all observed data locations (see also Vaida and Blanchard, 2005). However, we point out that if a predictor $\hat{Y}^{(k)}(\cdot;\boldsymbol{\theta})$ has minimum total squared prediction error summed over all observed data locations, it may not have the minimum squared prediction error at any individual spatial location in D_s. In this proceedings paper, we consider a *local* empirical deviance information criterion to select predictors at each individual spatial location in D_s.

In Section 2 we review $\mathrm{DIC_E}$, and in Section 3 we define local spatial-predictor selection based on a local $\mathrm{DIC_E}$. In Section 4, we simulate data using the Spatial Random Effects (SRE) model of Cressie and Johannesson (2008) and choose between spatial predictors defined in terms of different classes of basis functions. The empirical evidence suggests that in terms of total squared prediction error, the hybrid predictor consisting of locally selected predictors outperforms each individual predictor. We conclude with a discussion in Section 5.

2. Empirical Deviance Information Criterion

Consider the deviance information criterion (Spiegelhalter et al., 2002) to select between $\{\hat{Y}^{(k)}(\cdot;\boldsymbol{\theta}) : k = 1, ..., K\}$ in the case where $Y(\cdot)$ is Gaussian with known parameters $\boldsymbol{\theta}$,

$$\mathrm{DIC}^{(k)}(\mathbf{Z}_O;\boldsymbol{\theta}) \equiv \sum_{i=1}^{n} \frac{(Z(\mathbf{s}_i) - \hat{Y}^{(k)}(\mathbf{s}_i;\boldsymbol{\theta}))^2}{\sigma_\epsilon^2 v(\mathbf{s}_i)} + 2\sum_{i=1}^{n} \mathbf{H}^{(k)}(\mathbf{s}_i;\boldsymbol{\theta})'\mathbf{e}_i, \tag{3}$$

where $\mathbf{e}_i$ is the i-th column of the $n \times n$ identity matrix. Vaida and Blanchard (2005) show that the deviance information criterion is an unbiased estimator of the mean of the total squared prediction error based on the statistical model given by (1). That is, if $Y(\cdot)$ follows any generic distribution, and $\epsilon(\cdot)$ is a Gaussian white-noise process with mean zero and $\mathrm{var}(\epsilon(\mathbf{s})) = \sigma_\epsilon^2 v(\mathbf{s})$, then

$$E_{\mathbf{Z}_O}(\mathrm{DIC}^{(k)}(\mathbf{Z}_O;\boldsymbol{\theta})|\boldsymbol{\theta}) = E_{\mathbf{Z}_O,\mathbf{Y}}\left(\sum_{i=1}^{n} \frac{(Y(\mathbf{s}_i) - \hat{Y}^{(k)}(\mathbf{s}_i;\boldsymbol{\theta}))^2}{\sigma_\epsilon^2 v(\mathbf{s}_i)}\middle|\boldsymbol{\theta}\right). \tag{4}$$

Bradley et al. (2011) define the empirical deviance information criterion ($\mathrm{DIC_E}$) by substituting estimates of $\boldsymbol{\theta}$ into (3). That is,

$$\mathrm{DIC_E}^{(k)}(\mathbf{Z}_O) \equiv \sum_{i=1}^{n} \frac{(Z(\mathbf{s}_i) - \hat{Y}^{(k)}(\mathbf{s}_i;\hat{\boldsymbol{\theta}}))^2}{\sigma_\epsilon^2 v(\mathbf{s}_i)} + 2\sum_{i=1}^{n} \mathbf{H}^{(k)}(\mathbf{s}_i;\hat{\boldsymbol{\theta}})'\mathbf{e}_i, \tag{5}$$

where $\hat{\boldsymbol{\theta}}$ is a given estimator of $\boldsymbol{\theta}$; the subscript "E" in (5) stands for "empirical." Notice that $\mathrm{DIC_E}$ will depend on the estimator $\hat{\boldsymbol{\theta}}$ chosen.

There has been work in the literature to preserve the unbiasedness result in (4), in the case where parameters $\boldsymbol{\theta}$ are estimated. This can be achieved using a penalty called the generalized degrees of freedom (Ye, 1998; Huang and Chen, 2007; Liang et al., 2008; Greven and Kneib, 2010; Chen and Huang, 2011). From a computational-efficiency point of view, Bradley et al. (2011) provide empirical results that suggest that computing the generalized degrees of freedom is slow, which leads us to consider $\mathrm{DIC_E}$.

The conventional approach for model selection is to choose the k that minimizes $\{\mathrm{DIC_E}^{(k)}(\mathbf{Z}_O) : k = 1, ..., K\}$. This is quantified by the following equation,

$$k^*(\mathbf{Z}_O) \equiv \arg\min \left\{ \mathrm{DIC_E}^{(k)}(\mathbf{Z}_O) : k = 1, ..., K \right\}. \tag{6}$$

Notice that (5) is a "global" criterion since it depends on sums over $i = 1, ..., n$. In the next section, we consider selecting k at each $\mathbf{s}_i \in D_s$; that is, we consider "local" spatial-predictor selection.

3. Local Spatial-Predictor Selection

We begin this section by describing an ideal setting where we have a criterion based on unobserved quantities, that is, an oracle (Donoho and Johnstone, 1994). The normalized squared prediction error at generic spatial location $\mathbf{u}$, namely

$$\mathrm{SPE}^{(k)}(\mathbf{u}, \mathbf{Z}_O) \equiv \frac{(Y(\mathbf{u}) - \hat{Y}^{(k)}(\mathbf{u}; \hat{\boldsymbol{\theta}}))^2}{\sigma_\epsilon^2 v(\mathbf{u})}, \tag{7}$$

is an oracle since it depends on the unobserved quantity $Y(\mathbf{u})$. If we define the following local-selection procedure,

$$\tilde{k}(\mathbf{u}_j, \mathbf{Z}_O) \equiv \arg\min \left\{ \mathrm{SPE}^{(k)}(\mathbf{u}_j, \mathbf{Z}_O) : k = 1, ..., K \right\}, \tag{8}$$

then for any $k = 1, ..., K$, and spatial predictor locations $\mathbf{u}_1, ..., \mathbf{u}_N$, it is easy to see that

$$\sum_{j=1}^{N} \frac{(\, Y(\mathbf{u}_j) - \hat{Y}^{(\tilde{k})}(\mathbf{u}_j; \hat{\boldsymbol{\theta}}) \,)^2}{\sigma_\epsilon^2 v(\mathbf{u}_j)} \leq \sum_{i=1}^{N} \frac{(Y(\mathbf{u}_j) - \hat{Y}^{(k)}(\mathbf{u}_j; \hat{\boldsymbol{\theta}}))^2}{\sigma_\epsilon^2 v(\mathbf{u}_j)}. \tag{9}$$

That is, the hybrid predictor that is a combination of the locally selected predictors, selected according to (8), performs as well or better (in terms of total squared prediction error) than any of the $k = 1, ..., K$ predictors.

However, since $Y(\cdot)$ is not observed, we consider the estimation of SPE. In particular, for a location $\mathbf{s}_i$ where there is a datum $Z(\mathbf{s}_i)$, consider the following estimator of $\mathrm{SPE}^{(k)}(\mathbf{s}_i, \mathbf{Z}_O)$:

$$\frac{(Z(\mathbf{s}_i) - \hat{Y}^{(k)}(\mathbf{s}_i; \hat{\boldsymbol{\theta}}))^2}{\sigma_\epsilon^2 v(\mathbf{s}_i)} + 2\mathbf{H}^{(k)}(\mathbf{s}_i; \hat{\boldsymbol{\theta}})' \mathbf{e}_i; \;\; i = 1, ..., n, \tag{10}$$

where the second term is defined by individual components in (2) and (3). From Stein's Lemma (Stein, 1981), it is immediate that

$$\begin{aligned} &E_{\mathbf{Z}_O} \left(\frac{(Z(\mathbf{s}_i) - \hat{Y}^{(k)}(\mathbf{s}_i; \hat{\boldsymbol{\theta}}))^2}{\sigma_\epsilon^2 v(\mathbf{s}_i)} + 2\mathbf{H}^{(k)}(\mathbf{s}_i; \hat{\boldsymbol{\theta}})' \mathbf{e}_i \middle| \boldsymbol{\theta} \right) \\ &= E_{\mathbf{Z}_O, \mathbf{Y}} \left(\mathrm{SPE}^{(k)}(\mathbf{s}_i, \mathbf{Z}_O) | \boldsymbol{\theta} \right); \;\; i = 1, ..., n, \end{aligned} \tag{11}$$

and hence (10) and the oracle criterion (7) share the same mean.

However, $Z(\cdot)$ is typically not observed at all spatial locations in D_s. To address this, we select predictors over a partitioning of $D_s = P_1 \cup \cdots \cup P_A$, where $P_a \cap P_b$ is the empty set for $a \neq b$. It is important that each set of the partition include locations where there

are data. For a given set of the partition, we select a predictor based on all locations where there are data; this then defines the selected predictor at all locations in the set.

Specifically, we define the local empirical deviance information criterion for a set of the partition P_a, as

$$\mathrm{DIC_E}^{(k)}(P_a, \mathbf{Z}_O) \equiv \sum_{\mathbf{s}_i \in P_a} \frac{(Z(\mathbf{s}_i) - \hat{Y}^{(k)}(\mathbf{s}_i; \hat{\boldsymbol{\theta}}))^2}{\sigma_\epsilon^2 v(\mathbf{s}_i)} + 2 \sum_{\mathbf{s}_i \in P_a} \mathbf{H}^{(k)}(\mathbf{s}_i; \hat{\boldsymbol{\theta}})' \mathbf{e}_i; \;\; a = 1, ..., A. \tag{12}$$

Notice that the sums in (12) are over the data locations $\mathbf{s}_i$ contained in P_a.

Now use the spatial predictor that minimizes $\mathrm{DIC_E}^{(k)}(P_a, \mathbf{Z}_O)$ with respect to $k = 1, ..., K$. That is, for all $\mathbf{u} \in P_a$, the optimally selected predictor is

$$k^*(\mathbf{u}, \mathbf{Z}_O) \equiv \arg\min \left\{ \mathrm{DIC_E}^{(k)}(P_a, \mathbf{Z}_O) : k = 1, ..., K \right\}. \tag{13}$$

Now, as a varies from 1,..., A, $k^*(\mathbf{u}, \mathbf{Z}_O)$ is obtained for all $\mathbf{u} \in D_s$. The corresponding locally selected spatial predictor is given by the N-dimensional vector,

$$(\hat{Y}^{(k^*)}(\mathbf{u}_i; \hat{\boldsymbol{\theta}}) : i = 1, ..., N)'. \tag{14}$$

We call (14) the *locally selected spatial predictor*. In the next section, we provide empirical results that indicate that this predictor performs better, in terms of (bias-corrected) total squared prediction error, than any of the individual spatial predictors under consideration.

4. A Small Simulation Study

In this section, we use the SRE model as our simulation model and *Fixed Rank Kriging (FRK)* for spatial prediction (Cressie and Johannesson, 2006, 2008). We simulate data according to (1) with the following model for $Y(\cdot)$:

$$Y(\mathbf{s}) = (\mathbf{S}^{(0)}(\mathbf{s}))' \boldsymbol{\eta} + \xi(\mathbf{s}), \tag{15}$$

where $\mathbf{S}^{(0)}(\mathbf{s}) = (S_1^{(0)}(\mathbf{s}), ..., S_{r^{(0)}}^{(0)}(\mathbf{s}))'$ is an $r^{(0)}$-dimensional vector of basis functions and $\mathbf{s} \in D_s$. The $r^{(0)}$-dimensional random vector $\boldsymbol{\eta}$ is distributed as Gau($\mathbf{0}$, $\mathbf{K}^{(0)}$), where $\mathbf{K}^{(0)}$ is the $r^{(0)} \times r^{(0)}$ covariance matrix of $\boldsymbol{\eta}$, $\xi(\cdot)$ is assumed to be a white-noise Gaussian process with mean zero and variance $(\sigma_\xi^{(0)})^2$, and $\xi(\cdot)$ is assumed to be independent of $\boldsymbol{\eta}$. The fixed but unknown parameters are organized into the set $\boldsymbol{\theta} \equiv \{\mathbf{K}^{(0)}, (\sigma_\xi^{(0)})^2\}$. Recall that $\mathrm{var}(\epsilon(\cdot)) = \sigma_\epsilon^2 v(\cdot)$ is assumed known.

In Section 4.1, we describe the calibration of our simulation model. This includes the specification of the spatial domain D_s, σ_ϵ^2, $v(\cdot)$, $\mathbf{S}^{(0)}(\cdot)$, $\mathbf{K}^{(0)}$, and $(\sigma_\xi^{(0)})^2$.

4.1 Calibration of the Simulation Model

We consider the lattice, $D_s \equiv \{\mathbf{s} = (i, j)' : i = -50, ..., 50, j = -50, ..., 50\} = \{\mathbf{u}_1, ..., \mathbf{u}_N\}$, so that $N = 10,201$. We randomly selected $n = 2550$ locations in D_s, resulting in the data locations, $\{\mathbf{s}_1, ..., \mathbf{s}_{2550}\}$. In this small simulation study, $v(\cdot) \equiv 1$, and σ_ϵ^2 is given a value calibrated from (17) below. That is, we simulate $Y(\cdot)$ from (15) and

$$Z(\mathbf{s}_i) = Y(\mathbf{s}_i) + \epsilon(\mathbf{s}_i); \;\; i = 1, ..., 2550, \tag{16}$$

where $\epsilon(\cdot)$ is a Gaussian white-noise process with mean zero and variance σ^2_ϵ.

We simulate $Y(\cdot)$ from (15) by specifying $\mathbf{S}^{(0)}(\cdot)$ to consist of bisquare basis functions (Cressie and Johannesson, 2008) based on a quad-tree of centers (or knots) defined at two resolutions. A generic bisquare function and the two resolutions of the quad-tree centers are displayed in Figures 1 and 2, respectively. The number of basis functions in $\mathbf{S}^{(0)}(\cdot)$ is $r^{(0)} = 562$.

The 562×562 matrix $\mathbf{K}^{(0)}$ is calibrated against the stationary exponential covariance function. We choose $\mathbf{K}^{(0)}$ such that $||\mathbf{S}^{(0)}\mathbf{K}^{(0)}(\mathbf{S}^{(0)})' - \mathbf{\Sigma}^{(0)}||^2_F$ is minimized, where $\mathbf{S}^{(0)} \equiv (\mathbf{S}^{(0)}(\mathbf{u}_1), ..., \mathbf{S}^{(0)}(\mathbf{u}_N))'$, $||\cdot||_F$ is the Frobenius norm, and the (i,j)-th element of $\mathbf{\Sigma}^{(0)}$ is $\exp\{-||\mathbf{u}_i - \mathbf{u}_j||/\kappa\}$, for $\mathbf{u}_i, \mathbf{u}_j \in D_s$. The Frobenius norm is given by $||\mathbf{M}||_F \equiv \{\mathbf{M}'\mathbf{M}\}^{1/2}$ for a square matrix $\mathbf{M}$. The parameter $\kappa > 0$ in the exponential covariance function represents the level of spatial dependence; we set $\kappa = 25$ to obtain medium-strength spatial dependence (Kang et al., 2010; Bradley et al., 2011; Katzfuss and Cressie, 2011).

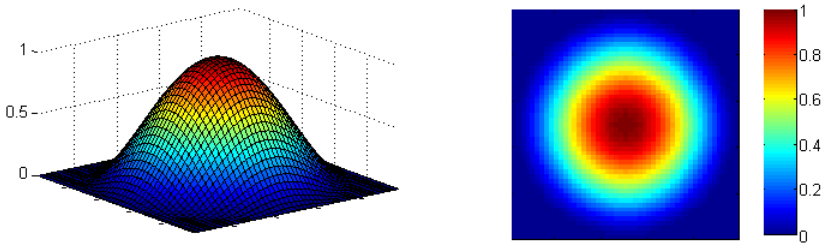

Figure 1: A bisquare function over two-dimensional Euclidean space.

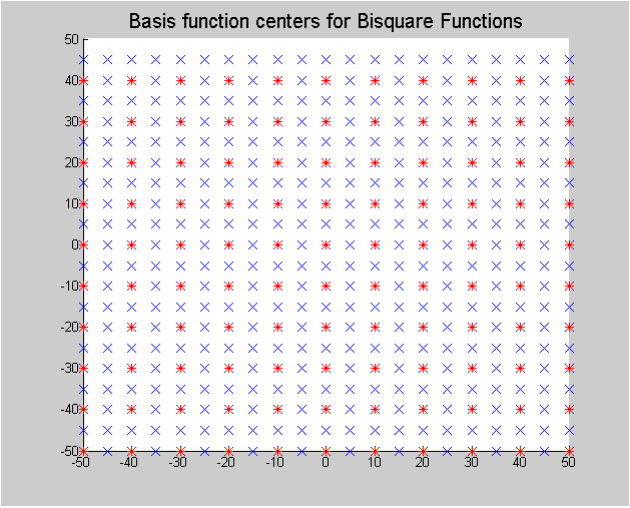

Figure 2: The basis-function centers used for simulation were obtained from a quad-tree. The red '*' corresponds to centers of the bisquare basis functions at the first resolution. The blue 'x' corresponds to centers of the bisquare basis functions at the second resolution.

We calibrate second-moment parameters in the simulation model (15) and (16) by first specifying proportions of variability and then solving for parameters (Kang et al., 2010; Bradley et al., 2011; Katzfuss and Cressie, 2011). Let the fine-scale-variation proportion be denoted as:

$$\text{FVP} \equiv \frac{N(\sigma_\xi^{(0)})^2}{N(\sigma_\xi^{(0)})^2 + \text{trace}(\mathbf{S}^{(0)}\mathbf{K}^{(0)}(\mathbf{S}^{(0)})')}. \tag{17}$$

The values of $\mathbf{K}^{(0)}$ and $\mathbf{S}^{(0)}$ are specified above; consequently, $\text{trace}(\mathbf{S}^{(0)}\mathbf{K}^{(0)}(\mathbf{S}^{(0)})')$ is known. In the simulation, we specify $\text{FVP} = 0.05$ and solve (17) to obtain $(\sigma_\xi^{(0)})^2 = 0.0486$. Let the signal-to-noise ratio be denoted,

$$\text{SNR} \equiv \frac{N(\sigma_\xi^{(0)})^2 + \text{trace}(\mathbf{S}^{(0)}\mathbf{K}^{(0)}(\mathbf{S}^{(0)})')}{N\sigma_\epsilon^2}. \tag{18}$$

In the simulation, we specify $\text{SNR} = 10$, and then we solve(18) to obtain $\sigma_\epsilon^2 = 0.1080$.

A single realization of $Y(\cdot)$ based on the calibrations above is shown in Figure 3 for illustration. In Section 4.2, we give two spatial predictors (i.e., $K = 2$) from which our local spatial-predictor selection will be illustrated.

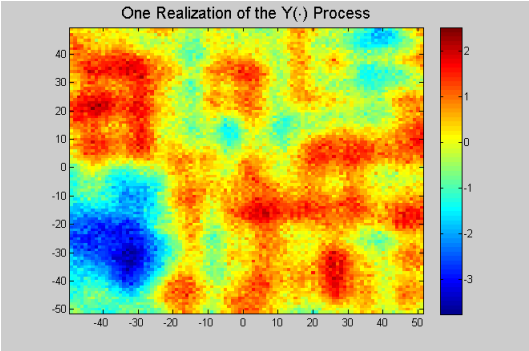

Figure 3: One realization from $Y(\cdot)$ (Section 4.1), which was obtained from an SRE model with bisquare basis functions.

4.2 Spatial mapping with Fixed Rank Kriging (FRK)

In this proceedings paper, we consider the simple case of selecting from two FRK predictors. For a given set of $r^{(k)}$ basis functions, $\mathbf{S}^{(k)}(\cdot)$, the k-th FRK predictor is based on the SRE model:

$$\mathbf{S}^{(k)}(\cdot)'\boldsymbol{\eta}^{(k)} + \xi^{(k)}(\cdot), \tag{19}$$

where $\text{var}(\boldsymbol{\eta}^{(k)}) \equiv \mathbf{K}^{(k)}$ and $\text{var}(\xi^{(k)}(\cdot)) \equiv (\sigma_\xi^{(k)})^2$; $k = 1, 2$. Cressie and Johannesson (2008) derive the FRK predictor as,

$$\begin{aligned} &E(Y(\mathbf{u})|\mathbf{Z}_O, \hat{\boldsymbol{\theta}}^{EM}, \mathbf{S}^{(k)}(\cdot)) \\ &\equiv \text{cov}_{\mathbf{Z}_O}(\mathbf{Z}_O, Y(\mathbf{u})|\hat{\mathbf{K}}^{EM}, (\hat{\sigma}_\xi^{EM})^2, \mathbf{S}^{(k)}(\cdot))'\text{cov}_{\mathbf{Z}_O}(\mathbf{Z}_O|\hat{\mathbf{K}}^{EM}, (\hat{\sigma}_\xi^{EM})^2, \mathbf{S}^{(k)}(\cdot))^{-1}\mathbf{Z}_O, \end{aligned} \tag{20}$$

where $\hat{\boldsymbol{\theta}}^{EM} \equiv \{\hat{\mathbf{K}}^{EM}, (\hat{\sigma}_\xi^{EM})^2\}$ is an EM estimate of $\boldsymbol{\theta} = \{\mathbf{K}^{(k)}, (\sigma_\xi^{(k)})^2\}$ (Katzfuss and Cressie, 2011) based on the EM algorithm. Also in (20), $\mathbf{S}^{(k)}(\cdot)$ is an $r^{(k)}$-dimensional vector of basis functions that are not necessarily the bisquare functions used to generate the data.

In (20), the n-dimensional vector,

$$\begin{aligned}
&\text{cov}(\mathbf{Z}_O, Y(\mathbf{u})|\hat{\mathbf{K}}^{EM}, (\hat{\sigma}_\xi^{EM})^2, \mathbf{S}^{(k)}) \\
&= \mathbf{S}^{(k)}\hat{\mathbf{K}}^{EM}\,\mathbf{S}^{(k)}(\mathbf{u}) + \text{cov}_{\mathbf{Z}_O}(\mathbf{Z}_O, \xi(\mathbf{u})|\hat{\mathbf{K}}^{EM}, (\hat{\sigma}_\xi^{EM})^2, \mathbf{S}^{(k)}) \\
&= \mathbf{S}^{(k)}\hat{\mathbf{K}}^{EM}\,\mathbf{S}^{(k)}(\mathbf{u}) + (\hat{\sigma}_\xi^{EM})^2(I(\mathbf{u} = \mathbf{s}_1), ..., I(\mathbf{u} = \mathbf{s}_n))', \qquad (21)
\end{aligned}$$

where $\mathbf{S}^{(k)} \equiv (\mathbf{S}^{(k)}(\mathbf{s}_1), ..., \mathbf{S}^{(k)}(\mathbf{s}_n))'$ is the $n \times r^{(k)}$ matrix of basis functions formed from $\mathbf{S}^{(k)}(\cdot)$, and $I(\cdot)$ denotes the indicator function. Also in (20), the $n \times n$ matrix,

$$\text{cov}(\mathbf{Z}_O|\hat{\mathbf{K}}^{EM}, (\hat{\sigma}_\xi^{EM})^2, \mathbf{S}^{(k)}) = \mathbf{S}^{(k)}\hat{\mathbf{K}}^{EM}\,(\mathbf{S}^{(k)})' + (\hat{\sigma}_\xi^{EM})^2\mathbf{I} + \sigma_\epsilon^2\mathbf{I},$$

where $\mathbf{I}$ denotes the $n \times n$ identity matrix.

In Section 4.3, we consider selecting predictors based on thin-plate splines and on facet-like linear functions. Let $\mathbf{S}^{(1)}(\cdot)$ consist of thin-plate splines based on the radial basis function, $h^2\log(h)$ (e.g., Wahba, 1990, p. 31), and let $\mathbf{S}^{(2)}(\cdot)$ consists of facet-like linear functions from Obled and Creutin (1986). A plot of these basis functions in two-dimensional Euclidean space is given in Figures 4 and 5, respectively. In Figure 6, we display the centers that define $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$.

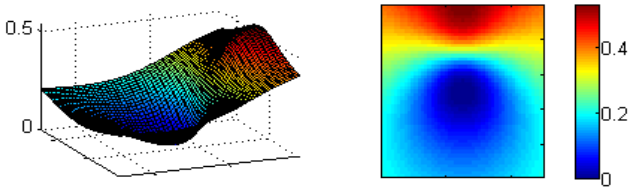

Figure 4: Example of a thin-plate spline from Wahba (1990, p. 31) over two-dimensional Euclidean space.

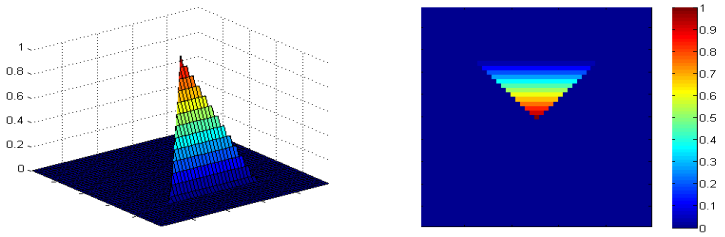

Figure 5: Example of a facet-like linear function over two-dimensional Euclidean space.

In our simulation experiment, we generated data based on an SRE model with $r^{(0)} = 562$ bisquare basis functions. Then we used two FRK predictors based on basis functions given by (misspecified) thin-plate splines ($r^{(1)} = 441$) and facet-like linear functions

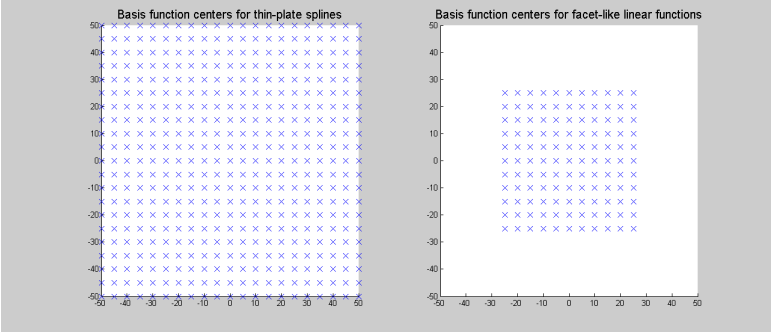

Figure 6: Basis-function centers for thin-plate splines (left-panel) and facet-like linear functions (right-panel). For both predictors, only one resolution is used, and $r^{(1)}$ = 441 and $r^{(2)} = 121$.

($r^{(2)} = 121$), respectively. In Figure 6, it is evident that the facet-like linear functions $\mathbf{S}^{(2)}(\cdot)$ might perform worse in some regions as there are no basis-function centers defined outside the region $\{(i,j) : i,j = -25, ..., 25\}$. In those regions near the boundary of D_s, $\mathbf{S}^{(2)}(\cdot)$ is identically zero. In future research, this simulation will be expanded to include different configurations of centers of $\mathbf{S}^{(k)}(\cdot)$ and hence of $r^{(k)}(\cdot)$, for $k = 1, ..., K > 2$.

4.3 Results for a Single Realization

In this section, we provide results for a single realization, to illustrate our proposed local spatial-predictor selection. The simulated observations are displayed in Figure 7; note that there are regions of missing data throughout D_s.

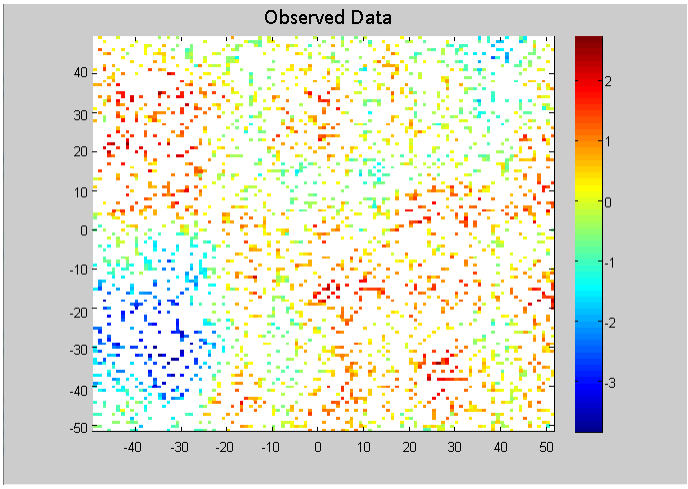

Figure 7: The data locations where color indicates a datum's value. White indicates the spatial locations where no data were observed.

We consider the selection of a spatial predictor from among $\{\hat{Y}^{(1)}(\mathbf{u}_i; \hat{\boldsymbol{\theta}}^{EM}), \hat{Y}^{(2)}(\mathbf{u}_i; \hat{\boldsymbol{\theta}}^{EM})\}$, for each $i = 1, ..., N = 10,201$. We choose a simple partition of size $A = 20 \times 20 = 400$, which is displayed in Figure 8 (i.e., $P_1, ..., P_{400}$). In Figure 9, we show the optimal local

selection of $k^*(\cdot, \mathbf{Z}_O)$ from among $\{1, 2\}$.

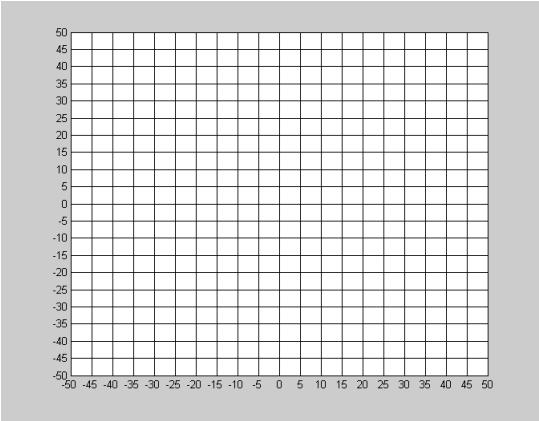

Figure 8: The grid-boxes represent $\{P_a : a = 1, ..., A\}$, a partition of D_s. Here $A = 400$.

Clearly, the predictor based on the facet-like linear functions is selected in the center of the spatial domain D_s. For this single realization, thin-plate splines tend to be chosen more outside the center of the spatial domain, which is consistent with the specification of the basis function centers displayed in Figure 6. Figure 10 shows the locally selected spatial predictor defined by (14), which should be compared to Figure 3.

There are obvious discontinuties in the locally selected predictor, which we shall now discuss. When comparing the center of Figure 10 to the center of Figure 3 we see that the locally selected spatial predictor (consisting mostly of predictions based on $k = 1$) appears to be recovering the values of $Y(\cdot)$ within this region. Outside the center of the spatial domain, we see a very large discrepency between Figures 10 and 3 indicating that neither the thin-plate splines nor the facet-like linear functions are performing well. This plot might lead one to include basis-function centers outside the center of the spatial domain when using facet-like linear functions for FRK prediction.

The stark differences between spatial predictors chosen at different grid cells might also occur because the partition in Figure 8 is too coarse (i.e., A might be too small). Instead of using a coarse regular grid, one could use a fine irregular tessellation (e.g., a Dirchlet tesselation, Cressie, 1993, pp. 373 - 375). We are currently investigating different ways to specify $\{P_1, ..., P_A\}$. However, in this small simulation study our focus is primarily on the performance of the locally selected spatial predictor in terms of squared prediction error using the partition in Figure 8.

Predictor	Total SPE
k = 1	2.7773e+003
k = 2	3.2599e+003
k^*	2.7534e+003

Table 1: The table shows total SPE given by (21). Recall that $k = 1$ corresponds to the FRK predictor using facet-like linear functions, and $k = 2$ corresponds to the FRK predictor using thin-plate splines. Here, k^* corresponds to the locally selected spatial predictor.

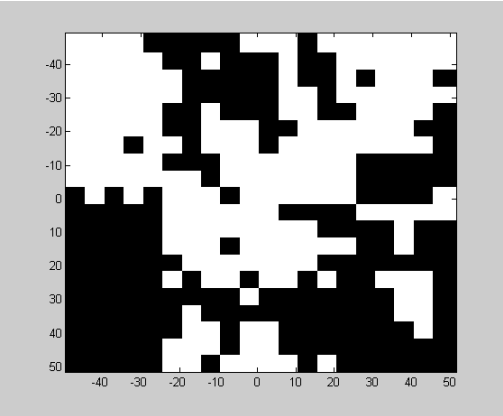

Figure 9: A map of $\{k^*(\mathbf{u}_i, \mathbf{Z}_O) : i = 1, ..., N\}$ for the data displayed in Figure 7. White indicates that the FRK predictor based on facet-like linear functions was chosen. Black indicates that the FRK predictor based on thin-plate splines was chosen.

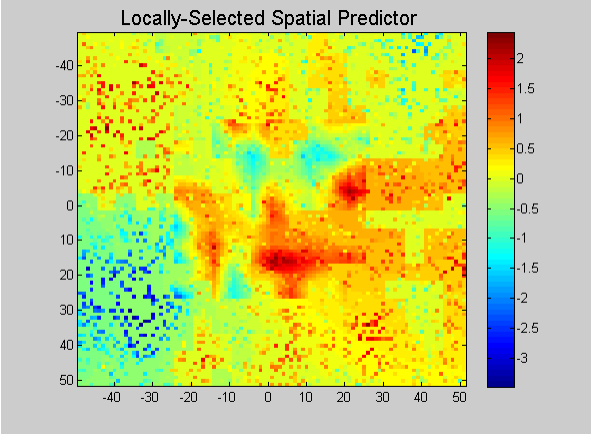

Figure 10: The locally selected spatial predictor given by (14). This should be compared to $Y(\cdot)$ in Figure 3.

Define the *total SPE* as,

$$\mathrm{TSPE}^{(k)}(\mathbf{Z}_O) \equiv \sum_{i=1}^{N} \mathrm{SPE}^{(k)}(\mathbf{u}_i, \mathbf{Z}_O), \tag{22}$$

where $\mathrm{SPE}^{(k)}$ is defined by (7). Notice that total SPE depends on the true value of $Y(\cdot)$, something we can know from the simulation. In Table 1, we report these values. For this particular realization, the locally selected spatial predictor outperforms either of the individual spatial predictors under consideration.

4.4 Results for Multiple Realizations from the Simulation

We simulated 500 independent realizations according to (15) and (16). For each of the 500 replications the same 2250 observed spatial locations, as seen in Figure 7, were used. The partition $\{P_1, ..., P_{400}\}$ displayed in Figure 8 was used for local spatial-predictor selection for each of the 500 replicates. We found that the locally selected spatial predictor in (14) had smaller total SPE than either of the individual predictors in 99.8% of the 500 simulation runs.

Although the locally selected spatial predictor consistently had smaller total SPE, the amount of improvement is also of interest. In Figure 11, we display kernel-smoothed histograms of the relative improvement,

$$\frac{\mathrm{TSPE}^{(k)}(\mathbf{Z}_O) - \mathrm{TSPE}^{k^*}(\mathbf{Z}_O)}{\mathrm{TSPE}^{(k)}(\mathbf{Z}_O)}, \tag{23}$$

for $k = 1, 2$. If (23) is "close" to 0, then the amount of improvement that the locally selected predictor provides is small. In Figure 11, we display $K = 2$ kernel-smoothed histograms of (23) based on the 500 independent realizations of $\mathbf{Z}_O$. The corresponding modes are approximately 0.3 ($k = 1$) and 0.45 ($k = 2$), indicating an improvement for the locally selected spatial predictor. From Figure 11, the value of the ratio in (23) is much more variable over independent replicates when $k = 1$ than when $k = 2$.

5. Discussion

In this proceedings paper, we have introduced the notion of local spatial-predictor selection from a finite set of spatial predictors over a lattice D_s, using the empirical deviance information criterion. Simulations were based on an SRE model using bisquare spatial basis functions, but FRK predictors were based on thin-plate splines and on facet-like linear functions. Using the total squared prediction error as our criterion, we saw that the locally selected spatial predictor outperformed each of the individual predictors, although there are discontinuities that occur.

6. Acknowledgments

This research was partially supported by NASA's Earth Science Technology Office through its Advanced Information Systems Technology program.

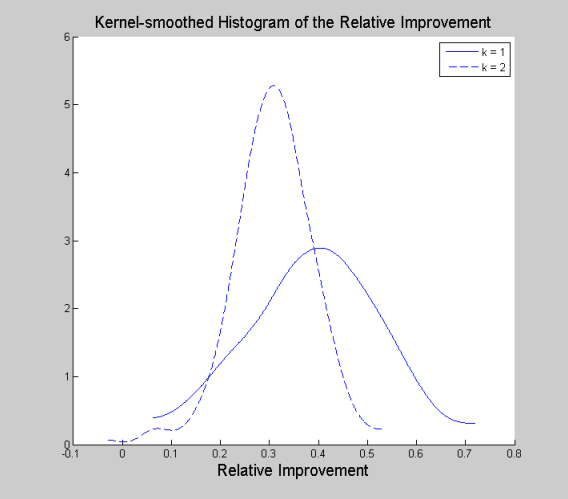

Figure 11: Kernel-smoothed histograms of the ratio in (23) over the 500 independent replicates. The solid line corresponds to $k = 1$ and the dashed line corresponds to $k = 2$.

References

Bradley, J. R., Cressie, N., and Shi, T. (2011). "Selection of rank and basis functions in the Spatial Random Effects Model." In *2011 Proceedings of the Joint Statistical Meetings*, 3393–3406. Alexandria, VA: American Statistical Association.

Chen, C. S. and Huang, H. C. (2011). "Geostatistical model averaging based on conditional information criteria." *Environmental and Ecological Statistics*, 19, 23–35.

Cressie, N. (1993). *Statistics for Spatial Data, rev. ed*. New York, NY: Wiley.

Cressie, N. and Johannesson, G. (2006). "Spatial prediction for massive data sets." In *Australian Academy of Science Elizabeth and Frederick White Conference*, 1–11. Australian Academy of Science, Canberra.

— (2008). "Fixed rank kriging for very large spatial data sets." *Journal of the Royal Statistical Society, Series B*, 70, 209–226.

Donoho, D. and Johnstone, I. (1994). "Ideal spatial adaptation by wavelet shrinkage." *Biometrika*, 81, 425–455.

Greven, S. and Kneib, T. (2010). " On the behaviour of marginal and conditional AIC in linear mixed models." *Biometrika*, 97, 773–789.

Huang, H. C. and Chen, C. S. (2007). " Optimal geostatistical model selection." *Journal of the American Statistical Association*, 102, 1009–1024.

Kang, E. L., Cressie, N., and Shi, T. (2010). " Using temporal variability to improve spatial mapping with application to satellite data." *Canadian Journal of Statistics*, 38, 271–289.

Katzfuss, M. and Cressie, N. (2011). "Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets." *Journal of Time Series Analysis*, 32, 430–446.

Liang, H., Wu, H., and Zou, G. (2008). "A note on conditional AIC for linear mixed-effects models." *Biometrika*, 95, 773–778.

Obled, C. and Creutin, J. (1986). "Some developments in the use of empirical orthogonal functions for mapping meteorological fields." *Journal of Applied Meteorology*, 25, 1189–1204.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society, Series B*, 64, 583–616.

Stein, C. (1981). "Estimation of the mean of the multivariate normal distribution." *Annals of Statistics*, 9, 1135–1151.

Vaida, F. and Blanchard, S. (2005). "Conditional Akaike information for mixed-effects models." *Biometrika*, 92, 351 – 370.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Ye, J. (1998). "On measuring and correcting the effects of data mining and model selection." *Journal of the American Statistical Association*, 93, 120 – 131.