Comments on: Comparing and selecting spatial predictors using local criteria

Martin P. Tingley · Benjamin A. Shaby

Published online: 29 November 2014
© Sociedad de Estadística e Investigación Operativa 2014

We applaud the authors on a thought provoking manuscript. In what follows, we discuss the use of mean squared error as a model selection criterion, comment on the interpretation and utility of maps of the selected predictor, and then present a simulation study that both demonstrates a new application, and reveals some limitations, of the Locally Selected Predictor concept.

1 Model selection on the basis of squared prediction error

Bradley et al. (2014) select between competing spatial predictors, including traditional stationary kriging and several non-stationary and reduced rank models, based on expected local squared prediction error. The inclusion of reduced rank models as candidate predictors warrants further discussion, as the performance of these predictors has recently been questioned by Stein (2014), largely on the basis of the Kullback–Leibler divergence between the low rank approximation and the true data generating measure. Applications in Stein (2014) further show that reduced rank methods can result in inefficient spatial interpolation as measured by mean squared prediction error.

In their simulation study (Bradley et al. 2014), focus on evaluating the mean squared prediction errors of the Locally Selected Predictor as a number of parameters of the simulation, including the signal to noise ratio and the neighborhood used to select

This comment refers to the invited paper available at doi:[10.1007/s11749-014-0415-1](https://doi.org/10.1007/s11749-014-0415-1).

M. P. Tingley (✉)

Department of Meteorology and Department of Statistics, Pennsylvania State University,
University Park, PA 16802, USA
e-mail: mpt14@psu.edu

B. A. Shaby

Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

between the predictors, are changed. They do not, however, provide the prediction errors associated with each of the candidate models. Given the questions raised by [Stein \(2014\)](#), it would be worthwhile to compare the mean squared prediction error of the various low rank candidate predictors to that of the Locally Selected Predictor. In addition, providing information on the gains in predictive performance that accompany the considerably larger effort of fitting and then selecting between seven spatial models would help illustrate the practical utility of the Locally Selected Predictor. Note that in their case study of global, satellite-derived CO₂ data, [Bradley et al. \(2014\)](#) find that the percentage improvement in the root squared testing error of the Locally Selected Predictor over the best global model is only 0.67 %.

2 Stability and interpretation of maps of the selected predictor

[Bradley et al. \(2014\)](#) mention in their discussion that, even if data are simulated under one of the candidate prediction models, the Locally Selected Predictor will not necessarily select that predictor at all locations. More should be made of this point: as we demonstrate below, the Locally Selected Predictor will not always, or even usually, select the model used to generate the data. As Kriging under the true model, provided it is Gaussian, is the best linear unbiased predictor, the inability of the Locally Selected Predictor scheme to recover that true model is troubling. More generally, additional commentary and results concerning the properties of maps of the selected predictor would aid interpretation. In the context of their simulation experiment, [Bradley et al. \(2014\)](#) do not report or discuss the fractional area for which each candidate predictor is selected, the stability of that fraction across the simulations, or how the properties of maps of the selected predictor may relate to the properties—stationarity, for example—of the data generating process or the candidate predictors themselves.

Regions of disparate spatial behavior are potentially indicative of different physical mechanisms giving rise to the observations. The agreement between the selected predictor map and spatial variability in the data generating mechanism is therefore indicative of the utility of the Locally Selected Predictor in diagnosing spatial variability in the underlying physical processes. However, [Bradley et al. \(2014\)](#) do not show that they can recover with reasonable fidelity the correct map of the selected predictor if the target spatial field is generated by stitching together two (or more) spatial fields, generated from distinct spatial models that are then included as candidate predictors. A complicating issue in this regard is that the candidate spatial predictors are fit *globally* but selected *locally*. For example, if the spatial field is a mixture of two mean-zero Gaussian processes, with exponential covariances featuring different range parameters, then even if an exponential covariance model is included as a candidate predictor, it may never be selected as global parameter estimation results in an estimated range parameter that is everywhere incorrect. That is, even if there are two distinct processes at work, over distinct spatial domains, Locally Selected Prediction does not necessarily afford a rational framework for disentangling them, as each candidate model is itself optimized globally. We return to this point in our own simulation study below, where we demonstrate how Locally Selected Prediction can be used to model non-stationary fields from a candidate pool of stationary models with parameters that are not globally optimal.

Bradley et al. (2014) provide only brief commentary on maps of the selected spatial predictor for their case study of global, satellite-derived CO₂ data. These maps are ragged, composed of small and irregular areas where each predictor is deemed optimal (Bradley et al. (2014), Fig. 8). They are difficult to interpret on scientific grounds and do not reflect the underlying geography—for example, there is no correspondence with coastlines, or the boundaries of climatic zones. Two factors may be limiting the scientific utility of maps of the selected predictor in this application. First, as discussed above, Bradley et al. (2014) fit the candidate pool of predictors globally, and then select between them locally. Second, three of the four spatial predictors used in the CO₂ example are non-stationary, so are arguably too flexible to be useful candidates, if the goal (different from optimizing prediction) is to use maps of the selected predictor to diagnose regions of disparate spatial behavior that may be indicative of variability in physical mechanisms. This is not to say that the non-stationary models are not individually useful for such a diagnosis (for example, by examining the spatial pattern of weights associated with the Fixed Rank Kriging basis functions at each resolution), but rather that the use of non-stationary candidate predictors limits the interpretability of maps of the selected predictor.

It is also worth noting that the candidate spatial predictors used in the analysis of the CO₂ data are of different characters. For example, the SPD predictor (Bradley et al. (2014), Fig. 7, third plot) is smoother than all other candidate predictors, and is indeed too smooth to be viewed as reasonable, entirely missing the area of low CO₂ concentrations in Northern Eurasia that is visually present in the training, validation, and testing data. By any common-sense assessment, this looks to be a poor fit to the data—yet it is selected as optimal for a significant fraction of the globe, presumably in regions where the field does not feature sharp spatial gradients.

The disparate properties of the candidate predictors suggest that a modification to the Locally Selected Predictor framework may result in maps of the selected predictor that are more informative with respect to physical mechanisms. Consider a spatial field composed of a smoothly and slowly varying background field, interspersed with regions of more local-scale variability caused by a distinct physical process. As an example, air pollution data may feature regions of fine-scale variability caused by localized sources of the pollutant (for example, Fuentes (2002)). If two candidate predictors are separately optimized for prediction in regions dominated by the slowly varying background state, and by the more local-scale behavior, respectively, maps of the selected predictor may then be useful for diagnosing regions dominated by one of two distinct physical mechanisms, and may therefore be more amenable to scientific interpretation. These observations motivate our simulation study.

3 Constructing non-stationary spatial processes

A simple modification to the Locally Selected Predictor procedure may allow for improved and efficient predictions of non-stationary spatial fields using a set of simple, stationary models as candidate predictors. In addition, the use of stationary candidate predictors that are not globally optimized may results in maps of the selected predictor that are more scientifically interpretable.

Fig. 1 Upper row the spatial weights, $\mathbf{X}_1$ and $\mathbf{X}_2$. Lower row examples of the two spatial fields, β_1 and β_2 , that are then mixed according to the spatial weights

We illustrate these concepts via a simulation study. The spatial field is formed as a spatially varying weighted average of two stationary and isotropic spatial processes, with different range parameters (c.f. [Gelfand et al. 2003](#)). For each location $\mathbf{s}$,

$$Y(\mathbf{s}) = X_1(\mathbf{s}) \cdot \beta_1(\mathbf{s}) + X_2(\mathbf{s}) \cdot \beta_2(\mathbf{s}), \quad (1)$$

where the weights, $\mathbf{X}_1$ and $\mathbf{X}_2$, vary in the x direction only (Fig. 1) and are specified such that the variance of $\mathbf{Y}$ is constant across space. β_1 and β_2 are mean-zero Gaussian spatial processes, with powered exponential covariances with common partial sills of unity and common powers in the exponential of 1.5. The range parameters differ between the two processes; β_1 has a range of 2, and β_2 a range of 0.1. Examples of β_1 and β_2 are shown in Fig. 1, and the corresponding $\mathbf{Y}$ in Fig. 2.

We remove a grid of 5×5 blocks of observations from $\mathbf{Y}$, and define within each a quincunx composed of four validation locations at the corners and one central prediction-and-testing location (white squares and black symbols in Fig. 2). Following [Bradley et al. \(2014\)](#), we select between competing spatial predictors at each of the

Fig. 2 Left panel an example of the final spatial field. White spaces indicate regions where observations are withheld when predicting and when fitting via MLE. Each white square contains four validation locations (circles) and a prediction-and-testing location (plus sign). Right panel analytical MSE for predictions at the validation locations according to Model 1 (long-range parameter) less the MSE for predictions according to Model 2 (short-range parameter)

prediction-and-testing locations on the basis of the mean squared error (MSE) over the four validation points within the corresponding block of withheld observations. Given the experimental design, this selection strategy is equivalent to either moving window or nearest-neighbor predictor selection [see Bradley et al. (2014), Eqs. 14 and 15].

The two candidate predictors, referred to as Models 1 and 2, employ traditional stationary Kriging (following the terminology of Bradley et al. 2014) and use the actual covariance parameters used to generate β_1 and β_2 , respectively. Our simulation study therefore takes the form of an Oracle experiment, as all parameters are treated as known (c.f., Li et al. 2010). Analytical calculations of the expected squared error at each validation location for both Models 1 and 2 indicate that in expectation, Model 1 is preferred at the 15 prediction-and-testing locations with largest x -coordinates (Fig. 2).

We also find the maximum likelihood estimate of the range parameter for each realization of $\mathbf{Y}$, and use it to predict at the prediction-and-testing locations via traditional stationary kriging. Deviations from stationarity in actual applications are likely to be more subtle than those simulated here (a stationary model is clearly inappropriate for $\mathbf{Y}$; Fig. 1), and our motivation is to understand the improvements in the MSE of predictions, averaged over the prediction-and-testing locations, of the Locally Selected Predictor as compared with traditional stationary Kriging using the globally derived MLE range parameter.

Local selection between the two stationary models, with short- and long-range parameters, respectively, reduces MSE as compared with predictions using the globally derived MLE—even though each individually is a poor domain-wide fit to the data. The average MSE over the 25 testing locations and across 1,000 simulations is 0.377 for the MLE, 0.379 for Model 1, 0.391 for Model 2, and 0.323 for the Locally Selected Predictor using Models 1 and 2 as candidates. Although Models 1 and 2 each result in higher

Fig. 3 The fraction of simulations for which Model 1 (long-range parameter) is selected at each prediction-and-testing location

domain-averaged MSE than predictions from the MLE, local selection between them results in an approximately 14 % improvement in MSE as compared with the MLE.

We are primarily interested in determining if the Locally Selected Predictor can correctly identify spatial regions dominated by the long- versus short-range parameters: with what fidelity can we recover the different data generating mechanisms? Inasmuch as we are able to correctly identify the regions dominated by Models 1 versus Model 2, and vice versa, results point to the possibility of using Locally Selected Predictor to identify regions dominated by different physical processes. This goal is substantially different from that of [Bradley et al. \(2014\)](#), where the discussion focusses on improving MSE of predictions by selecting between a number of reasonable spatial predictors, each fit globally. Here, we consider two stationary models, *neither one of which is globally optimal by any criteria*, select between them on the basis of locally estimated MSE, and assess agreement with the data generating mechanism.

There is a great deal of variability across simulations with respect to maps of the selection between Model 1 and 2 (Fig. 3; compare with [Bradley et al. \(2014\)](#) Fig. 8), despite the radically different range parameters. For the right-most column of testing locations, where $X_1(\mathbf{s}) = 1$ and $X_2(\mathbf{s}) = 0$, Model 1 is selected in about 86 % of simulations, while for the left-most column of testing locations, where $X_1(\mathbf{s}) = 0$ and $X_2(\mathbf{s}) = 1$, Model 1 is still selected in about 43 % of simulations (Fig. 3). The testing locations in the central (according to the x coordinate) portion of the domain feature intermediate values, and are not as easy to interpret as both $X_1(\mathbf{s})$ and $X_2(\mathbf{s})$ are non-zero. Where the process $\mathbf{Y}$ is influenced solely by the long-range parameter, the Locally Selected Predictor does an adequate job of identifying the correct model. In regions where the short-range parameter is dominant, the Locally Selected Predictor

selects the incorrect model in almost half of the simulations. The tendency of the Locally Selected Predictor to more frequently select a model with a larger, as opposed to smaller, range parameter than was used to generate the data is in accord with results from Kaufman and Shaby (2013) that predictive performance, as measured by local MSE, is only weakly impacted if the range parameter is larger than optimal.

Results of this simulation study point to both a potential new use of Locally Selected Predictors, and to the possible limitations of using the resulting maps of the selected predictor as a way to diagnose spatial variability in the mechanism(s) producing the observations. Given a process that is clearly non-stationary, predictive MSE is improved when forming the Locally Selected Predictor from two stationary candidate predictors, neither of which is globally optimized. However, the fidelity of the Locally Selected Predictor to the data generating mechanism in our simulations is low, particularly in regions dominated by the short-range process: even in this idealized setting, maps of the selected predictor across simulations are sufficiently variable to preclude their use as a means of identifying regions of disparate spatial behavior. We hope that further refinements of the novel, Locally Selected Predictor idea of Bradley et al. (2014) will permit for analyses that extend beyond optimized predictions and seek to identify or understand the scientific phenomena producing the data.

References

- Bradley JR, Cressie N, Shi T (2014) Comparing and selecting spatial predictors using local criteria. *TEST*. doi:[10.1007/s11749-014-0415-1](https://doi.org/10.1007/s11749-014-0415-1)
- Fuentes M (2002) Spectral methods for nonstationary spatial processes. *Biometrika* 89(1):197–210
- Gelfand AE, Kim H-J, Sirmans CF, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. *J Am Stat Assoc* 98(462):387–396
- Kaufman CG, Shaby BA (2013) The role of the range parameter for estimation and prediction in geostatistics. *Biometrika* 100(2):473–484
- Li B, Nychka DW, Ammann CM (2010) The value of multi-proxy reconstruction of past climate. *J Am Stat Assoc* 105(491):883–911 With discussions and rejoinder
- Stein ML (2014) Limitations on low rank approximations for covariance matrices of spatial data. *Spat. Stat.* 8:1–19