<table>
<thead>
<tr>
<th>Title</th>
<th>A Generalized Measure of Uncertainty in Geostatistical Model Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>SS-2016-0368R1</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a></td>
</tr>
<tr>
<td>DOI</td>
<td>10.5705/ss.202016.0368</td>
</tr>
</tbody>
</table>
| Complete List of Authors | Chun-Shu Chen  
Jun Zhu and  
Tingjin Chu |
| Corresponding Author | Chun-Shu Chen |
| E-mail | cschen@cc.ncue.edu.tw |

Notice: Accepted version subject to English editing.
A Generalized Measure of Uncertainty in Geostatistical Model Selection

Running title: Measure of Uncertainty in Spatial Model Selection

Chun-Shu Chen\textsuperscript{1,*}, Jun Zhu\textsuperscript{2}, and Tingjin Chu\textsuperscript{3}

\textsuperscript{1}Institute of Statistics and Information Science, National Changhua University of Education, Changhua 500, Taiwan
\textsuperscript{2}Department of Statistics and Department of Entomology, University of Wisconsin, Madison, WI 53706, USA
\textsuperscript{3}Center for Applied Statistics and Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China

Abstract: Model selection and model averaging are essential to regression analysis in environmental studies, but determining which of the two approaches is the more appropriate and under what circumstances remains an active research topic. In this paper, we focus on geostatistical regression models for spatially referenced environmental data. For a general information criterion, we develop a new perturbation-based criterion that measures the uncertainty (or, instability) of spatial model selection, as well as an empirical rule for choosing between model selection and model averaging. Statistical inference based on the proposed model selection instability measure is justified both in theory and via a simulation study. The predictive performance of model selection and model averaging can be quite different when the uncertainty in model selection is relatively large, but the performance becomes more comparable as this uncertainty decreases. For illustration, a precipitation data set in the state of Colorado is analyzed.

Key words and phrases: Data perturbation; Generalized degrees of freedom; Geostatistics; Information criterion
mation criterion; Model complexity; Spatial prediction.

1 Introduction

For regression analysis of geostatistical data in many environmental studies, the response variable of interest is often observed along with a set of covariates at spatial sampling locations. Selection of a subset of covariates and prediction of the response at unsampled locations are generally based on fitting spatial linear regression models and choosing a suitable subset of covariates using a model selection criterion such as the Akaike’s information criterion (AIC) (Akaike (1973)). To assess the fitted models, stochasticity of both the parameter estimates and that of the model selection should be considered. In geostatistics, model selection may involve not only selection of covariates but also determination of the spatial error structure. In this paper, we restrict our attention to the selection of a suitable subset of covariates in a geostatistical model. However, the randomness in the selection of models (or, selection uncertainty) is often ignored in the statistical inference post model selection (e.g., Breiman (1996)). To mitigate the effect of selection uncertainty, model averaging that pools multiple fitted models is widely used as it can provide a better prediction than a single best model. Although model selection and model averaging have been well studied for the standard linear regression that assumes independent errors (e.g., Shao (1997), Hoeting, Madigan, Raftery, and Volinsky (1999), Burnham and Anderson (2002), Claeskens and Hjort (2008)), results are far fewer for spatial linear regression in geostatistics. The purpose of this paper is to develop a new approach to geostatistical regression model selection and model averaging for the analysis of spatial data in the environmental sciences.

With model selection (or, covariate selection), generally a best model (or, a best subset of covariates) is selected based on a certain type of criterion and the more important covariates are identified according to the best model. With model averaging, however, several candidate
models are combined based on estimated model weights. Although model averaging tends to give better prediction than model selection, the computational cost is often higher due to the search for suitable weights. It is also more challenging to infer about the relationship between the response and the covariates based on an averaged model. Under the Bayesian framework (e.g., Hoeting, Madigan, Raftery, and Volinsky (1999), Johnson and Hoeting (2011)), the posterior inclusion probability (PIP) of each covariate provides a measure of importance of the covariate in relation to the response, for which a prior specification and Markov chain Monte Carlo (MCMC) for the posterior computation are required. In addition, some Gibbs sampler based methods are also commonly used for Bayesian variable selection such as the stochastic search variable selection (SSVS) algorithm (George and McCulloch (1993)). In this paper, our focus is to investigate the connection between model selection and model averaging in a frequentist framework.

It is in general not clear whether one strategy (i.e., model selection or model averaging) is preferable over the other and under what circumstances. To address this issue, Yuan and Yang (2005) developed a criterion to capture the uncertainty of model selection in standard linear regression with independent errors. Ghosh and Yuan (2009) proposed an $L_1$-norm criterion that measures the instability of model selection for logistic regression with binary data, as well as an empirical rule to suggest whether model selection or model averaging is preferable. In addition, Efron (2014) used a bootstrap based method to discuss the stability of an estimator after model selection for independent observations. The above results are useful for evaluating model selection and model averaging, but the response variables are assumed to be independent.

In geostatistics, Hoeting, Davis, Merton, and Thompson (2006) provided some heuristic arguments in spatial linear regression model selection using AIC. A simulation study indicated that if the spatial dependence is ignored, some important covariates may not be selected and hence
the prediction errors will be high. Huang and Chen (2007) developed an approximately unbiased estimator of the mean squared prediction error (MSPE) for evaluating different spatial predictors based on generalized degrees of freedom and derived asymptotic efficiency results for the proposed method, but the focus was on selection among different predictors obtained by different spatial models for prediction rather than selection of covariates. More recently, Chu, Zhu, and Wang (2011) proposed a penalized maximum likelihood estimation (PMLE) and a one-step sparse approximation to simultaneously select covariates and estimate parameters in spatial linear regression models, but quantification of model selection uncertainty was not considered. The uncertainty of model selection has received much attention under the Bayesian approaches (e.g., Clyde and George (2004), Johnson and Hoeting (2011)) and can be measured via the posterior inference. To the best of our knowledge, however, foundational questions of how to evaluate the uncertainty of model selection and the connection between model selection and model averaging in geostatistical regression settings have not been adequately addressed under the frequentist framework and will be explored in this paper.

Here, we develop new methodology for geostatistical regression model selection and model averaging in the context of two model selection criteria, namely, generalized information criterion (GIC) and conditional generalized information criterion (CGIC). We propose a novel criterion to measure the uncertainty (or, instability) of geostatistical regression model selection based on the selected model and the corresponding predictor. The resulting predictor after model selection and parameter estimation is nonlinear, which makes the problem more challenging to handle than the standard linear regression. Our overall strategy is to develop an index that quantifies the instability in geostatistical regression model selection via a perturbation technique. It simultaneously takes into account the uncertainties of model selection and parameter estimation. By normalizing the
instability index, a generalized instability measure is developed which more accurately reflects the complexity of a model fitting procedure. In addition, we establish a theoretical connection between the proposed index of selection instability and the notion of generalized degrees of freedom for geostatistical regression model selection using GIC and CGIC. For practical applications, we further develop an empirical rule that helps to determine whether model selection or model averaging is preferred under GIC and CGIC.

The remainder of this paper is organized as follows. In Section 2, we describe the geostatistical regression model and the corresponding spatial predictor. In Section 3, we derive the properties of various model selection methods for geostatistical regression and some model averaging methods are also introduced. We then develop an index of model selection instability and a generalized instability measure. Further, a theoretical result associated with the proposed methodology is established. In Section 4, we provide an estimation method of the generalized instability measure and an empirical rule for choosing between model selection and model averaging. The results of two simulation scenarios and a weather data example are given in Sections 5 and 6, respectively. We conclude with a discussion in Section 7, and the technical details are given in the Appendix.

2 Geostatistical Regression Model and Spatial Prediction

2.1 Geostatistical Regression Model

Let $\mathcal{D} \subset \mathbb{R}^2$ be a continuous and bounded study region and let $s$ be an arbitrary location in $\mathcal{D}$. Suppose there are $p$ covariates at location $s$ and are denoted, together with 1 for the intercept, by $x(s) = (1, x_1(s), \ldots, x_p(s))^\prime$. A spatial random field $\{S(s) : s \in \mathcal{D}\}$ of interest is

$$S(s) = \beta_0 + \sum_{j \in M_0} \beta_j x_j(s) + \eta(s),$$

(1)

where $\beta_j$ for $j = 1, \ldots, p$ are regression coefficients, $M_0$ is the index set of the covariates in the true model, and $\eta(\cdot)$ is a spatial random error process that captures the spatial variation of $S(\cdot)$ and
can provide a local adjustment to the mean trend due to unobserved covariates. It is a common practice to assume that $\eta(\cdot)$ follows a Gaussian process with mean zero and covariance function $K(\cdot; \theta)$ parameterized by the vector $\theta$. Readers interested in various covariance functions may refer to Chapter 4 of Schabenberger and Gotway (2005) for more details. Now, the response variable $Z(s)$ at location $s \in D$ is modeled by

$$Z(s) = \beta_0 + \sum_{j \in M_0} \beta_j x_j(s) + \eta(s) + \varepsilon(s), \quad (2)$$

where $\varepsilon(s) \sim N(0, \sigma^2)$ is a measurement error and is independent of the spatial error process $\eta(s)$. We will refer to (2) as the true geostatistical regression model.

Next, we consider model selection among the $p$ covariates indexed by $P = \{1, \ldots, p\}$. Let $M$ denote a candidate model as a subset of $P$ and let $M \subseteq 2^P$ denote a class of candidate models. Let $Z = (Z(s_1), \ldots, Z(s_n))' \equiv (Z_1, \ldots, Z_n)'$ be the data observed at $n$ sampling locations $s_1, \ldots, s_n$, and let $X_M = (x_M(s_1), \ldots, x_M(s_n))'$ be an $n \times (|M| + 1)$ design matrix for model $M$ with $|M|$ denoting the number of covariates in $M$. For a given candidate model $M \in \mathcal{M}$, the geostatistical regression model (2) can be rewritten in matrix form

$$Z = X_M \beta_M + \eta + \varepsilon \sim N(X_M \beta_M, \Sigma_Z), \quad (3)$$

where $\beta_M$ is the vector of regression coefficients consisting of $\beta_0$ and $\{\beta_j : j \in M\}$, $\eta = (\eta(s_1), \ldots, \eta(s_n))'$ is the vector of spatial random errors with a covariance matrix $\Sigma_\eta(\theta) = [K(s_i, s_{i'}; \theta)]_{i,i'=1}^n$, $\varepsilon = (\varepsilon(s_1), \ldots, \varepsilon(s_n))' \sim N(0, \sigma^2 I)$ is the vector of independent measurement errors, and $\Sigma_Z = \Sigma_\eta(\theta) + \sigma^2 I$ is the covariance matrix of the data vector $Z$. 
2.2 Spatial Prediction

For predicting the spatial random field \( \{ S(s) : s \in D \} \) based on the data vector \( Z \), we define \( S = (S(s_1), \ldots, S(s_n))' \) at \( n \) sampling locations. By (1), we have

\[
S = X_M \beta_M + \eta \sim N(X_M \beta_M, \Sigma(\theta)). \tag{4}
\]

For a given model \( M \in \mathcal{M} \) with known \( \theta \) and \( \sigma_e^2 \), the best linear unbiased predictor (BLUP) of \( S(s) \) at any location \( s \in D \) is obtained by minimizing the MSPE (see, e.g., Chapter 5 of Schabenberger and Gotway (2005)). The BLUP of \( S(s) \), indexed by a given model \( M \), is given by

\[
\hat{S}_M(s; \theta) = x_M(s) \tilde{\beta}_M + \text{cov}(\eta(s), \eta) \Sigma^{-1}_Z (Z - X_M \tilde{\beta}_M), \tag{5}
\]

where \( \tilde{\beta}_M = (X_M' \Sigma^{-1}_Z X_M)^{-1} X_M' \Sigma^{-1}_Z Z \) is the generalized least squares estimator of \( \beta \) and \( \text{cov}(\eta(s), \eta) = (K(s, s_1; \theta), \ldots, K(s, s_n; \theta)) \). Further, we rewrite the vector of BLUPs \( \hat{S}_M(\theta) = (\hat{S}_M(s_1; \theta), \ldots, \hat{S}_M(s_n; \theta))' \) as

\[
\hat{S}_M(\theta) = H_M(\theta) Z, \tag{6}
\]

where \( H_M(\theta) = \sigma_e^2 \Sigma^{-1}_Z X_M (X_M' \Sigma^{-1}_Z X_M)^{-1} X_M' \Sigma^{-1}_Z + \Sigma(\theta) \Sigma^{-1}_Z \). It follows from (6) that \( \hat{S}_M(\theta) \) is a linear combination of the data vector \( Z \), given \( \theta \) and \( \sigma_e^2 \).

In practice, the parameters \( \theta \) and \( \sigma_e^2 \) are unknown but can be estimated by, for example, the maximum likelihood (ML), the restricted maximum likelihood (REML), or a Bayesian method (e.g., Schabenberger and Gotway (2005)). Here, we use likelihood-based methods to estimate model parameters and in particular, the REML method, because it tends to give less biased estimators than the corresponding ML estimators (McGilchrist (1989), Cressie and Lahiri (1993), Cressie and Lahiri (1996)). Let \( \hat{\theta}_M \) and \( \hat{\sigma}_e^2_M \) denote the REML estimates under the candidate model (3). We have the estimates of the covariance matrix \( \hat{\Sigma}_Z = \hat{\Sigma}_\eta(\hat{\theta}_M) + \hat{\sigma}_e^2_M I \) and the vector of regression
coefficients \( \hat{\beta}_M = (X'_M \hat{\Sigma}_Z^{-1} X_M)^{-1} X'_M \hat{\Sigma}_Z^{-1} Z \). Together with (5), an empirical predictor of \( S(s) \) for any \( s \in D \) is given by

\[
\hat{S}_M(s; \hat{\theta}_M) = x'_M(s) \hat{\beta}_M + \hat{\sigma}_M^2 \hat{\Sigma}_Z^{-1}(Z - X_M \hat{\beta}_M),
\]

where \( \hat{\sigma}_M^2 \) is the REML estimate. Analogous to (6), the predictors of \( S \) at \( n \) sampling locations can be rewritten as

\[
\hat{S}_M(\hat{\theta}_M) = (\hat{S}_M(s_1; \hat{\theta}_M), \ldots, \hat{S}_M(s_n; \hat{\theta}_M))' = \hat{H}_M(\hat{\theta}_M) Z,
\]

where

\[
\hat{H}_M(\hat{\theta}_M) = \hat{\sigma}_M^2 \hat{\Sigma}_Z^{-1} X_M (X'_M \hat{\Sigma}_Z^{-1} X_M)^{-1} X'_M \hat{\Sigma}_Z^{-1} + \hat{\Sigma}_\eta(\hat{\theta}_M) \hat{\Sigma}_Z^{-1}.
\]

Unlike (6), however, the matrix \( \hat{H}_M(\hat{\theta}_M) \) depends on the data vector \( Z \) and hence \( \hat{S}_M(\hat{\theta}_M) \) in (8) is no longer a linear predictor.

3 Index of Selection Instability in Geostatistical Regression

3.1 Model Selection via GIC and CGIC

Under the candidate model (3), we consider a generalized information criterion (GIC) defined as

\[
\text{GIC}_A(M) = -2\ell_M(\hat{\beta}_M, \hat{\sigma}_M^2; \hat{\Sigma}_Z; Z) + \lambda(|M| + |\theta| + 2),
\]

where \( \lambda > 0 \) is a penalty parameter, \( |M| \) is the number of covariates in model \( M \), \( |\theta| \) is the number of unknown parameters in \( \Sigma_\eta(\theta) \), \( \ell_M(\cdot) \) is the log-likelihood function of \( Z \) based on model \( M \), and \( \hat{\beta}_M, \hat{\theta}_M, \) and \( \hat{\sigma}_M^2 \) are the REML estimates. The GIC in (10) includes the Akaike’s information criterion (AIC) with \( \lambda = 2 \) (Akaike (1973)), the Bayesian information criterion (BIC) with \( \lambda = \log(n) \) (Schwarz (1978)), the corrected AIC (AICc) criterion with \( \lambda = 2n/(n - |M| - |\theta| - 2) \) (Hurvich and Tsai (1989)), and the risk inflation criterion (RIC) with \( \lambda = 2\log(p) \) (Foster and...
George (1994)). For a given $\lambda$, the model that has the smallest value of $\text{GIC}_\lambda$ is selected as the best model and is denoted as

$$
\hat{M}(\lambda) = \arg \min_{M \in \mathcal{M}} \text{GIC}_\lambda(M).
$$

(11)

By (8) and (9), the corresponding predictor of $\mathbf{S}$ is

$$
\hat{\mathbf{S}}_{\hat{M}(\lambda)}(\hat{\mathbf{M}}(\lambda)) = \hat{\mathbf{H}}_{\hat{M}(\lambda)}(\hat{\theta}_{\hat{M}(\lambda)})\mathbf{Z}.
$$

(12)

In addition, we consider a conditional generalized information criterion (CGIC) given by

$$
\text{CGIC}_\lambda(M) = n^{-1} \left\{ \| \mathbf{Z} - \hat{\mathbf{S}}_M(\hat{\mathbf{M}}(\lambda)) \|^2 + \lambda \hat{\sigma}_\lambda^2 \text{tr} \left( \hat{\mathbf{H}}_M(\hat{\theta}_M) \right) \right\},
$$

(13)

where $\lambda > 0$ is a penalty parameter and $\hat{\sigma}_\lambda^2$ is an estimate of $\sigma_\lambda^2$ invariant to the model choice obtained by, for example, REML based on the full model. The CGIC in (13) includes the conditional Akaike’s information criterion (CAIC) with $\lambda = 2$ and the conditional BIC (CBIC) criterion with $\lambda = \log(n)$ as special cases (e.g., Vaida and Blanchard (2005), Chen and Huang (2012)). For a given $\lambda$, the selected model based on $\text{CGIC}_\lambda$ is denoted as

$$
\hat{M}_c(\lambda) = \arg \min_{M \in \mathcal{M}} \text{CGIC}_\lambda(M)
$$

(14)

and the corresponding predictor of $\mathbf{S}$ is

$$
\hat{\mathbf{S}}_{\hat{M}_c(\lambda)}(\hat{\theta}_{\hat{M}_c(\lambda)}) = \hat{\mathbf{H}}_{\hat{M}_c(\lambda)}(\hat{\theta}_{\hat{M}_c(\lambda)})\mathbf{Z}.
$$

(15)

Compared with the GIC in (10), the CGIC in (13) not only considers the size of regression term (i.e., $|M|$), but also the complexity of spatial dependence in the model selection procedure. This intuition can be further supported by the following Proposition 1 and its proof is given in the Appendix.
Proposition 1. Under the settings of (3), (8), and (9), let $|M|$ be the number of covariates in model $M$ and define

$$
\hat{Q}_M(\hat{\theta}_M) = X_M(\Sigma_M^{-1} - 1)\Sigma_M^{-1}
$$

with $\Sigma_M = \Sigma_M(\hat{\theta}_M) + \hat{\sigma}_2^2 I$. Then, for any $\lambda > 0$ and $M \in \mathcal{M}$, $CGIC_\lambda(M)$ of (13) can be rewritten as

$$
CGIC_\lambda(M) = n^{-1} \left[ ||Z - \hat{S}_M(\hat{\theta}_M)||^2 + \lambda \hat{\sigma}_2^2 |M| + \lambda \hat{\sigma}_2^2 tr \left( \Sigma_M(\hat{\theta}_M) \Sigma_M^{-1} \{ I - \hat{Q}_M(\hat{\theta}_M) \} \right) \right].
$$

(16)

In Proposition 1, the third term on the right-hand side of (16) reflects the complexity of spatial dependence in the model selection procedure. Hence, the CGIC is expected to be more suitable than the GIC in geostatistical regression model selection, which will be demonstrated by a simulation study in Section 5.

3.2 Model Averaging via GIC and CGIC

The main idea of model averaging is to combine several predictors $\hat{S}_M(\hat{\theta}_M)$ obtained from several candidate models for $M \in \mathcal{M}$ based on estimated model weights. Two approaches to estimate the model weights are frequentist model averaging (FMA) and Bayesian model averaging (BMA). In this paper, we focus on using the FMA. Readers interested in the BMA may refer to Raftery, Madigan, and Hoeting (1997) and Hoeting, Madigan, Raftery, and Volinsky (1999) for details. In order to connect with Section 3.1, we review a commonly used FMA technique based on GIC and CGIC (e.g., Burnham and Anderson (2002), Claeskens and Hjort (2008)). The idea is to consider all the candidate models in $\mathcal{M}$ for averaging, where the weight of each candidate model $M$ is determined by

$$
\hat{w}_\lambda(M) = \frac{\exp \left\{-(1/2)GIC_\lambda(M) \right\}}{\sum_{M^* \in \mathcal{M}} \exp \left\{-(1/2)GIC_\lambda(M^*) \right\}},
$$

(17)

for $GIC_\lambda(M)$ or

$$
\hat{w}_\lambda^c(M) = \frac{\exp \left\{-(1/2)CGIC_\lambda(M) \right\}}{\sum_{M^* \in \mathcal{M}} \exp \left\{-(1/2)CGIC_\lambda(M^*) \right\}},
$$

(18)

for $CGIC_\lambda(M)$. 10
for CGIC$_\lambda (M)$. For a given $\lambda > 0$, the model averaging predictors of $S$ based on GIC$_\lambda (M)$ and CGIC$_\lambda (M)$ are obtained, and are respectively denoted as

$$\hat{S}_\lambda = \sum_{M \in M} \hat{w}_\lambda (M) \hat{S}_M (\hat{\theta}_M)$$

(19)

and

$$\hat{S}_\lambda^c = \sum_{M \in M} \hat{w}_\lambda^c (M) \hat{S}_M (\hat{\theta}_M),$$

(20)

where $\hat{S}_M (\hat{\theta}_M)$ is given in (8). Readers may refer to Burnham and Anderson (2002) and Claeskens and Hjort (2008) for a comprehensive review of model averaging.

3.3 Generalized Instability Measure

In practice, if a model selection procedure is unstable, model averaging strategies may be considered in order to make more accurate predictions. Otherwise, it will be relatively easy to find a best model according to some criterion and then predictions will work well based on the selected model. Therefore, a measure of the instability associated with a model selection procedure is a critical issue when deciding model selection or model averaging.

We now develop new measures of the instability associated with model selection based on GIC and CGIC. Let $e_i$ for $i = 1, \ldots, n$ be the $i$th column of the $n \times n$ identity matrix. For a given penalty parameter $\lambda > 0$ (or, a given model selection criterion), we define an index of selection instability (ISI) as

$$\text{ISI} (\lambda) = E \left( \lim_{\delta \to 0} \delta^{-1} \sum_{i=1}^n \left| \hat{S}_{\hat{\gamma}(\lambda)} (s_i; Z + \delta e_i) - \hat{S}_{\hat{\gamma}(\lambda)} (s_i; Z) \right| \right),$$

(21)

where $\hat{\gamma} (\lambda)$ denotes a selected model which is obtained by GIC$_\lambda$ or CGIC$_\lambda$ according to the data vector $Z$. Further, $\hat{S}_{\hat{\gamma}(\lambda)} (s_i; Z + \delta e_i)$ and $\hat{S}_{\hat{\gamma}(\lambda)} (s_i; Z)$ are the predictors of $S(s_i)$ based on the model $\hat{\gamma} (\lambda)$ applied to $Z + \delta e_i$ and $Z$, respectively. It can be shown that $\hat{S}_{\hat{\gamma}(\lambda)} (s_i; Z) = \hat{S}_{\hat{\gamma}(\lambda)} (s_i; \hat{\theta}_{\hat{\gamma}(\lambda)})$.
and thus, $\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z)$ is an alternative notation of $\hat{S}_{\hat{\gamma}(\lambda)}(s_i; \hat{\theta}_{\hat{\gamma}(\lambda)})$ used here to emphasize that it relies on the data vector $Z$.

The ISI is an $L_1$-norm criterion and can be applied to assess the instability of GIC$_\lambda(M)$ of (10) and CGIC$_\lambda(M)$ of (13). If a model selection procedure is unstable, a minor perturbation (i.e., $\delta e_i$ for $i = 1, \ldots, n$ and $\delta \rightarrow 0$) in the data vector $Z$ would have a high chance to select a very different model, say $\hat{\gamma}^*(\lambda)$ and $\hat{\gamma}^*(\lambda) \neq \hat{\gamma}(\lambda)$. As a result, the difference between $\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z + \delta e_i)$ and $\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z)$ under the same model $\hat{\gamma}(\lambda)$ is expected to be large. In other words, the predictors obtained based on the selected model tend to have large variances and the ISI value of (21) is expected to be larger. Unlike for the standard regression models with independent responses (Ghosh and Yuan (2009)), however, our focus is on geostatistical regression models with spatially dependent responses and the predictors in (21) are nonlinear after model selection and parameter estimation.

The following proposition provides an alternative expression for ISI($\lambda$) of (21) and the proof can be found in the Appendix.

**Proposition 2.** Under the settings of (3), (4), (8), and (9), let $\hat{\gamma}(\lambda)$ be a generic model obtained from GIC$_\lambda(M)$ of (10) or CGIC$_\lambda(M)$ of (13), and the corresponding spatial predictors $\hat{S}_{\hat{\gamma}(\lambda)}(\hat{\theta}_{\hat{\gamma}(\lambda)}) = (\hat{S}_{\hat{\gamma}(\lambda)}(s_1; \hat{\theta}_{\hat{\gamma}(\lambda)}), \ldots, \hat{S}_{\hat{\gamma}(\lambda)}(s_n; \hat{\theta}_{\hat{\gamma}(\lambda)}))'$ of $S$ are given by (12) or (15). Suppose

$$\sum_{i=1}^{n} E\left( |\hat{S}_{\hat{\gamma}(\lambda)}(s_i; \hat{\theta}_{\hat{\gamma}(\lambda)})|/S \right) < \infty \text{ almost surely.}$$

Then, ISI($\lambda$) of (21) is given by

$$ISI(\lambda) = \sum_{i=1}^{n} E\left( \frac{\partial}{\partial S(s_i)} E\left( \hat{S}_{\hat{\gamma}(\lambda)}(s_i; \hat{\theta}_{\hat{\gamma}(\lambda)})/S \right) \right). \quad (22)$$

In Proposition 2, for a given model selection criterion GIC$_\lambda$ or CGIC$_\lambda$ with $\lambda > 0$, the ISI($\lambda$) of (22) can be interpreted as the expected sum of sensitivities of the spatial predictor $\hat{S}_{\hat{\gamma}(\lambda)}(\cdot)$ with respect to the underlying process $S(\cdot)$. That is, if a model fitting procedure is sensitive to the
data vector $Z$, the corresponding ISI value is expected to be large. In other words, a larger ISI value indicates that the corresponding model selection criterion has a higher selection instability. Although it is consistent with the interpretation of the original definition of $\text{ISI}(\lambda)$ in (21), $\text{ISI}(\lambda)$ of (22) has a more intuitive interpretation regarding the instability of a model selection criterion.

The $\text{ISI}(\lambda)$ in (22) is akin to generalized degrees of freedom which can be used to measure the complexity of a model fitting procedure (e.g., Ye (1998), Huang and Chen (2007)). In the absence of spatial dependence of the data vector $Z$, $\text{ISI}(\lambda)$ is reduced to the generalized degrees of freedom (GDF) in Ye (1998). In particular, Ye (1998) developed the GDF to measure the complexity of a model fitting procedure which considers parameter estimation uncertainty and model selection uncertainty, and has been applied by Shen and Ye (2002) in standard linear regression model selection.

Like the usual degrees of freedom in linear regression models, the value of $\text{ISI}(\lambda)$ is closely related to the number of unknown parameters in the model. Thus, for fairer comparison among various model selection criteria, we normalize the $\text{ISI}(\lambda)$ and define a generalized instability measure (GIM) as

$$
\text{GIM}(\lambda) = \frac{\text{ISI}(\lambda)}{|\hat{\gamma}(\lambda)| + |\theta| + 2},
$$

(23)

where $\hat{\gamma}(\lambda)$ is the model selected by $\text{GIC}_\lambda$ or $\text{CGIC}_\lambda$, $|\hat{\gamma}(\lambda)|$ is the number of covariates in model $\hat{\gamma}(\lambda)$, $|\theta|$ is the number of unknown parameters in $\Sigma_\eta(\theta)$, and the additional 2 is for the intercept $\beta_0$ and the noise variance $\sigma^2$. Therefore, our proposal is to use GIM($\lambda$) to measure the instability of a model selection criterion under geostatistical regression settings.
4 Practical Considerations

4.1 Estimation of Generalized Instability Measure

In general, ISI(\(\lambda\)) and GIM(\(\lambda\)) are unknown and need to be estimated. For this purpose, we obtain an approximately unbiased estimator \(\hat{\text{ISI}}(\lambda)\) of ISI(\(\lambda\)) based on a data perturbation technique (Huang and Chen (2007)). Here, we use GIC\(_{\lambda}(M)\) of (10) to illustrate how to estimate ISI(\(\lambda\)) and a similar procedure can be applied to CGIC\(_{\lambda}(M)\) of (13).

Again, we consider model selection among the \(p\) covariates indexed by \(\mathcal{P} = \{1, \ldots, p\}\). Let \(M\) denote a candidate model as a subset of \(\mathcal{P}\) and let \(\mathcal{M} \subseteq 2^\mathcal{P}\) denote a class of candidate models. For a given data vector \(Z\) and a selection criterion GIC\(_{\lambda}(M)\) with \(\lambda > 0\) and \(M \in \mathcal{M}\), the estimation procedure of ISI(\(\lambda\)) comprises the following four steps.

1. Based on (10) and (11), a model \(\hat{M}(\lambda)\) is selected from \(\mathcal{M}\) and the corresponding number of covariates in model \(\hat{M}(\lambda)\) is denoted as \(|\hat{M}(\lambda)|\).

2. Generate a set of perturbed data vectors \(Z^{*(t)} = (Z^{*(t)}_1, \ldots, Z^{*(t)}_n)' = Z + \tau \xi^{(t)}\) for \(t = 1, \ldots, T\), where \(\xi^{(t)} \sim N(0, \Delta^2 I)\) is independent of \(Z\) and \(\tau \in (0, 1]\) is the size of perturbation.

3. Analogous to Step 1, we select a model among \(\mathcal{M}\) based on GIC\(_{\lambda}(M)\) for each perturbed data vector \(Z^{*(t)}\), where the selected model is denoted as \(\hat{M}^{*(t)}(\lambda)\) and the corresponding estimate of \(\theta\) is \(\hat{\theta}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}\) for \(t = 1, \ldots, T\). Also, the predictor of \(S\) is denoted as \(\hat{S}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}(s_1; \hat{\theta}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}), \ldots, \hat{S}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}(s_n; \hat{\theta}^{*(t)}_{\hat{M}^{*(t)}(\lambda)})\)\('.

4. With \(\bar{S}^*(s_i) = T^{-1} \sum_{t=1}^{T} \hat{S}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}(s_i; \hat{\theta}^{*(t)}_{\hat{M}^{*(t)}(\lambda)})\) and \(\bar{Z}^{*(t)} = T^{-1} \sum_{t=1}^{T} Z^{*(t)}_i\) for \(i = 1, \ldots, n\), approximate the ISI(\(\lambda\)) by

\[
\hat{\text{ISI}}(\lambda) = \frac{1}{(T - 1)\tau^2 \Delta^2} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( \hat{S}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}(s_i; \hat{\theta}^{*(t)}_{\hat{M}^{*(t)}(\lambda)}) - \bar{S}^*(s_i) \right) \left( Z^{*(t)}_i - \bar{Z}^{*(t)} \right).
\]
From $\hat{\text{ISI}}(\lambda)$, an estimator $\hat{\text{GIM}}(\lambda)$ of $\text{GIM}(\lambda)$ for $\text{GIC}_\lambda$ is given by

$$\hat{\text{GIM}}(\lambda) = \frac{\hat{\text{ISI}}(\lambda)}{|\mathcal{M}(\lambda)| + |\mathbf{\theta}| + 2}.$$  \hfill (24)

Henceforth, we use $\hat{\text{GIM}}(\lambda)$ to measure the instability of a model selection criterion when fitting a geostatistical regression model.

4.2 An Empirical Rule

As shown in the simulation study of Section 5, the instability of a model selection criterion is closely related to the size of the underlying true model. Thus, we propose an empirical rule based on $\text{GIM}(\lambda)$ to roughly estimate the size of the underlying true model.

For $j = 1, \ldots, p$, let $\mathcal{P}_j$ be the class of all subsets of size $j$ that are selected from $\mathcal{P} = \{1, \ldots, p\}$. That is, $\mathcal{P}_1 = \{\{1\}, \ldots, \{p\}\}$, $\mathcal{P}_2 = \\{\{a, b\} : a, b = 1, \ldots, p, \text{ and } a \neq b\}$, $\cdots$, $\mathcal{P}_p = \{\{1, \ldots, p\}\}$. Let $\hat{\gamma}_j(\lambda)$ be the best subset among $\mathcal{P}_j$ selected by $\text{GIC}_\lambda$ or $\text{CGIC}_\lambda$. Here, let $\mathcal{P}_0$ be the subset that contains the intercept-only model. Based on (23), we define a generalized instability measure for $\mathcal{P}_j$ as

$$\text{GIM}_j(\lambda) = \frac{\text{ISI}_j(\lambda)}{j + |\mathbf{\theta}| + 2},$$ \hfill (25)

where $\text{ISI}_j(\lambda)$ is akin to the $\text{ISI}(\lambda)$ in (21) but selects models among $\mathcal{P}_j$. The estimate $\hat{\text{GIM}}_j(\lambda)$ of $\text{GIM}_j(\lambda)$ can be obtained by the estimation procedure of $\text{GIM}(\lambda)$ in Section 4.1, where $\mathcal{M}$ is replaced with $\mathcal{P}_j$. Intuitively, if the true model $M_0 \in \mathcal{P}_j$, the value of $\text{GIM}_j(\lambda)$ is expected to decrease as $j$ increases for $0 \leq j < j^*$ and is expected to be stable for $j^* \leq j \leq p$. To see the pattern of change in $\{\hat{\text{GIM}}_j(\lambda) : j = 0, 1, \ldots, p\}$ more clearly, we define a relative difference of generalized instability measures between $\mathcal{P}_j$ and $\mathcal{P}_{j-1}$ as

$$\text{ReGIM}_j(\lambda) = \left|\text{GIM}_j(\lambda) - \text{GIM}_{j-1}(\lambda)\right|; \quad j = 1, \ldots, p.$$ \hfill (26)

15
In practice, we can judge the size of the underlying true model according to the pattern of \( \{ \text{ReGIM}_j(\lambda) : j = 1, \ldots, p \} \) as we will demonstrate in the precipitation data example in Section 6.

5 Simulation Study

We now conduct two simulation scenarios to evaluate the performance of the generalized instability measure \( \widehat{\text{GIM}}(\lambda) \) in (24) for geostatistical regression model selection and model averaging under various situations.

Here, we consider an isotropic and stationary process for the spatial random error process \( \eta(\cdot) \) with a Matérn covariance function \( K(s_A, s_B; \theta) \equiv \sigma^2 \rho(s_A, s_B; a, \nu) \) and \( \theta \equiv (\sigma^2, a, \nu)' \), where \( \rho(s_A, s_B; a, \nu) \) is a Matérn correlation function (Matérn (2013)) defined by

\[
\rho(s_A, s_B; a, \nu) = \frac{\|s_A - s_B\|^{\nu}}{\Gamma(\nu)a^{2\nu-1}} K_\nu(a^{-1}\|s_A - s_B\|). \tag{27}
\]

In (27), \( \|s_A - s_B\| > 0 \) is the Euclidean distance between two locations \( s_A, s_B \in \mathcal{D} \), \( K_\nu(\cdot) \) is a modified Bessel function of the second kind with order \( \nu > 0 \), \( \nu \) is a smoothness parameter that controls the smoothness of the process, \( a > 0 \) is a range parameter that controls the range of spatial dependence, and \( \sigma^2 \) is the variance of \( \eta(\cdot) \). In (27), a larger \( \nu \) value indicates a smoother spatial process, while a larger \( a \) value indicates a stronger spatial dependence.

Next, we consider two simulation scenarios I (strong spatial dependence) and II (weak spatial dependence). The corresponding results are based on \( B = 200 \) simulation replicates.

5.1 Simulation Scenario I

Under the true geostatistical regression model (2), we consider an unit square as the study region \( \mathcal{D} \equiv [0,1]^2 \) in \( \mathbb{R}^2 \) and \( N = 100 \times 100 = 10,000 \) regular grid points, where the coordinate of the \( i \)th grid point is denoted as \( s_i = (s_{i1}, s_{i2})' \) for \( i = 1, \ldots, N \). In our simulation scenario, a total of \( p = 9 \) covariates \( \{x_1(s), \ldots, x_9(s) ; s \in \mathcal{D} \} \) are considered and each covariate is independently generated.
from the standard Gaussian distribution. In addition, the spatial random error process \( \eta(\cdot) \) is generated from a zero-mean Gaussian stationary process, where the covariance matrix is based on the exponential covariance function (i.e., \( \nu = 0.5 \) in (27)) with \( \theta = (1, 1, 0.5)' \). The right-hand side of Figure 1 shows a realization of \( \eta(\cdot) \) which corresponds to a stronger spatial dependence. Further, the measurement error (or, noise) variance \( \sigma_z^2 = 1 \) in (2) is assumed to be known throughout simulation scenario I. We consider five different \( M_0 \) as the underlying true models in the forms of 
\[
\beta_0 + \sum_{j=1}^{k} \beta_j x_j(s) \quad \text{for} \quad k = 1, 3, 5, 7, \text{and} \ 9.
\]
For each \( k \), the regression coefficients are set to be \( \beta_0 = 1, \beta_1 = \cdots = \beta_k = (3/k)^{1/2} \), and \( \beta_{k+1} = \cdots = \beta_p = 0 \) so that the signal-to-noise ratio (SNR) is kept at 4, where the SNR is defined as the ratio of the variance of the signal \( S(\cdot) \) to the noise variance \( \sigma_z^2 \).

In each simulated data set, a sample size of \( n = 50 \) is drawn from the 10,000 grid points based on a simple random sampling scheme. For each true model (i.e., \( k = 1, 3, 5, 7, \text{or} \ 9 \)), we apply six model selection criteria, AIC, BIC, AICc, RIC, CAIC, and CBIC, to select models among \( \mathcal{M} \), where \( \mathcal{M} = 2^{\{1, \ldots, 9\}} \) consists of all the possible combinations of covariates \( x_1, x_2, \ldots, x_9 \), with \( M = \emptyset \in \mathcal{M} \) representing the intercept-only model. Throughout simulation scenario I, the model parameters are estimated by REML.

To assess the instability of model selection based on AIC, BIC, AICc, RIC, CAIC, and CBIC, the corresponding values of \( \hat{G}_\text{IM}(\lambda) \) in (24) are computed. The perturbed sample size is set to \( T = 50 \) and the perturbation size is set to \( \tau = 0.1 \) for computing \( \hat{I}_\text{SI}(\lambda) \).

To compare with six model selection criteria, predictions based on model averaging methods as illustrated in Section 3.2 are also conducted. For each true model (i.e., \( k = 1, 3, 5, 7, \text{or} \ 9 \)), the six model averaging methods based on (17)–(20) are referred to as AIC-MA, BIC-MA, AICc-MA, RIC-MA, CAIC-MA, and CBIC-MA.
5.2 Simulation Result I

We compare the prediction performance of various model selection and model averaging methods by examining an average squared prediction error (ASPE):

\[
\text{ASPE} = B^{-1} \sum_{b=1}^{B} \left[ n^{-1} \sum_{i=1}^{n} \left\{ \hat{S}^{(b)}(s_i) - S^{(b)}(s_i) \right\}^2 \right],
\]

where \( \hat{S}^{(b)}(s) \) is a generic predictor of \( S^{(b)}(s) \) corresponding to (12), (15), (19), or (20) for the \( b \)th simulated data set and \( S^{(b)}(s) \) is an underlying random variable of interest at the location \( s \) for the \( b \)th simulation replicate. In addition, we examine the performance of variable selection for GIC\(_\lambda\) and CGIC\(_\lambda\) under the five true models. Table 1 shows the ASPE values for the six model selection criteria (i.e., AIC, BIC, AICc, RIC, CAIC, and CBIC) and the six model averaging methods (i.e., AIC-MA, BIC-MA, AICc-MA, RIC-MA, CAIC-MA, and CBIC-MA) under the five true models (i.e., \( k = 1, 3, 5, 7, \) and 9). The corresponding variable selection results for the six model selection criteria are given in Table 2. For ease of comparison, the ASPE values in Table 1 are also plotted in Figure 2(a), where we omit the results of AICc, AICc-MA, RIC, and RIC-MA because the patterns are similar.

Table 1 and Figure 2(a) indicate that the CGIC\(_\lambda\) performs well in most cases, as it is closely related to the MSPE. For example, the CAIC is an unbiased estimator of the MSPE plus the noise variance when the model parameters are known (Vaida and Blanchard (2005)). Comparing model selection and model averaging (e.g., AIC versus AIC-MA, BIC versus BIC-MA, etc.) in terms of the ASPE values, we find that with the smaller models (e.g., \( k = 1, 3, \) or 5) with fewer covariates in our simulation, the ASPE values of model selection and model averaging give quite different prediction results. For example, AIC with an average ASPE value 0.5357 and a standard error 0.0193 is significantly different to AIC-MA which has an average ASPE value 0.4598 with a standard error 0.0184 (see, e.g., \( k = 1 \) in Table 1). In general, model averaging outperforms model...
selection. However, the prediction results of model selection and model averaging become more comparable for the true models with more covariates (e.g., $k = 7, 9$) under this simulation scenario.

In terms of the rate of selecting the true model (Table 2), a model selection criterion with a larger penalty parameter (e.g., BIC, RIC, and CBIC) penalizes more for a model with more covariates and hence tends to have a higher rate of selecting the true model when the underlying true model has fewer covariates, and vice versa. Further, by Proposition 1, $\text{CGIC}_\lambda$ considers the complexity of a spatial dependence in its selection procedure. As is expected, $\text{CGIC}_\lambda$ outperforms the corresponding $\text{GIC}_\lambda$. Somewhat surprisingly, the RIC criterion tends to have a relatively high rate of selecting the underlying true model. In other words, besides the conditional information criteria and our proposed empirical rule in Section 4.2, RIC is also a suitable criterion for determining the size of the underlying true model.

The left panel of Table 3 shows the generalized instability measure under the AIC, BIC, AICc, RIC, CAIC, and CBIC criteria by computing $\widehat{\text{GIM}}(\lambda)$ of (24), and the corresponding results are also plotted in Figure 2(b). CAIC and CBIC tend to have smaller GIM values than AIC, BIC, AICc, and RIC under geostatistical regression model selection; this agrees with Proposition 1. It also indicates that the conditional information criteria are more stable and thus are more suitable than unconditional information criteria in geostatistical regression model selection, as shown in Tables 1 and 2. In addition, from a prediction point of view and for a given model selection criterion (see, e.g., Figure 2(a) and Figure 2(b)), we find that model selection and model averaging can be quite different when the model selection uncertainty (i.e., value of $\widehat{\text{GIM}}(\lambda)$) is relatively large, but tends to be more similar as model selection uncertainty decreases.

Next, we perform a sensitivity analysis to evaluate the effect of the perturbation size $\tau$ on the computation of $\widehat{\text{GIM}}(\lambda)$. The average values of $\widehat{\text{GIM}}(\lambda)$ based on $k = 5$ and 200 simulation
replicates are shown in the right panel of Table 3, along with the standard errors. The numerical results indicate that the computation of $\hat{\text{GIM}}(\lambda)$ is not sensitive to the choice of $\tau$. In the past, Huang and Chen (2007) and Chen, Yang, and Li (2014) also showed that model selection and weight selection of model averaging using the data perturbation approach are not sensitive to the choice of $\tau$ under the frameworks of geostatistical models. Therefore, a nonadaptive $\tau = 0.1$ throughout the simulation scenario I is acceptable. Shen and Huang (2006) further developed a methodology about the adaptive choice of $\tau$ or the optimal choice of $\tau$ under different data sets, although the computation is more time-consuming.

5.3 Simulation Scenario II

To evaluate the performance of the generalized instability measure $\hat{\text{GIM}}(\lambda)$ in (24) for geostatistical regression model selection more realistically, the second simulation scenario is designed based on a real data example.

We consider a weather data set in the state of Colorado (Reich and Davis (2008), Chu, Zhu, and Wang (2011)), where the response variable is the March mean precipitation (inches per 24-hour period) on the log scale from 261 weather stations. For each weather station, $p = 10$ covariates are available. In addition to elevation ($x_1$), slope ($x_2$), and aspect ($x_3$), seven spectral bands from a MODIS satellite imagery (i.e., B1M–B7M) are available and are successively denoted as $x_4$, $x_5$, $x_6$, $x_7$, $x_8$, $x_9$, and $x_{10}$. Under the geostatistical regression model (2) with the covariance matrix of $\eta(\cdot)$ being defined in (27), we consider three different $M_0$ as the underlying true models in the forms of $\beta_0 + \sum_{j=1}^{k} \beta_j x_j(s)$ for $k = 1, 5,$ and 9. For each $k$, the regression coefficients, $\theta = (\sigma^2, a, \nu)'$, and $\sigma^2$ are estimated by REML based on the precipitation data set. These estimated parameter values are displayed in Tables 4 and 5, and are used to be the true parameter values of the underlying geostatistical regression model (2) when simulating new responses for the three
cases (i.e., \( k = 1, 5, \) and 9). In the simulation scenario II, the spatial random error process \( \eta(\cdot) \) is generated from a zero-mean Gaussian stationary process, where the covariance matrix is based on the exponential covariance function (i.e., \( \nu = 0.5 \) in (27)) with the range parameter \( a \) being around \( 0.1 \sim 0.14 \) as shown in Table 5. The left-hand side of Figure 1 shows a realization of \( \eta(\cdot) \) with \((\sigma^2_{\eta}, a, \nu) = (1, 0.1, 0.5)\). Compared with the simulation scenario I, simulation scenario II corresponds to a weaker spatial dependence. In addition, the SNR values for the three cases (i.e., \( k = 1, 5, \) and 9) are also shown in Table 5. Compared with the SNR = 4 in the simulation scenario I, the SNR value in the simulation scenario II is smaller and hence it indicates that the noise is stronger.

In each simulated data set, a sample size of \( n = 100 \) is drawn from 261 weather stations based on a simple random sampling scheme. For each true model (i.e., \( k = 1, 5, \) or 9), we apply six model selection criteria, AIC, BIC, AICc, RIC, CAIC, and CBIC, to select models among all the possible combinations of covariates. For CAIC and CBIC which are special cases of the CGIC in (13), an estimate \( \hat{\sigma}^2 \) of \( \sigma^2 \) is obtained by REML based on the full model so that it is invariant to the model choice. In addition, the estimate \( \hat{\sigma}^2 \) is used in the data perturbation procedure when computing \( \hat{GIM}(\lambda) \) of (24). Throughout the simulation scenario II, the model parameters are estimated by REML.

### 5.4 Simulation Result II

To assess the instability of model selection based on AIC, BIC, AICc, RIC, CAIC, and CBIC in geostatistical regression, the corresponding values of \( \hat{GIM}(\lambda) \) in (24) are computed based on a data perturbation approach as illustrated in Section 4.1, where the perturbation size is set to \( \tau = 0.1 \) for computing \( \hat{ISI}(\lambda) \). The left panel of Table 6 shows the results of \( \hat{GIM}(\lambda) \) under the AIC, BIC, AICc, RIC, CAIC, and CBIC criteria for the three cases (i.e., \( k = 1, 5, \) and 9) based on
200 simulation replicates. Analogous to simulation scenario I, the conditional information criteria (e.g., CAIC and CBIC) tend to have smaller GIM values than unconditional information criteria (e.g., AIC, BIC, AICc, and RIC) in the simulation scenario II. It indicates that the conditional information criteria are more stable and thus are more suitable than unconditional information criteria in geostatistical regression model selection. We again conclude that model selection and model averaging will give more comparable results as the uncertainty of model selection decreases. This is as expected, because the tendency of GIM in the simulation scenario II is similar to that in the simulation scenario I for a given model selection criterion.

Again, we perform a sensitivity analysis to evaluate the effect of the perturbation size $\tau$ on the computation of $\hat{\text{GIM}}(\lambda)$. The average values of $\hat{\text{GIM}}(\lambda)$ and the corresponding standard errors based on $k = 5$ and 200 simulation replicates are given in the right panel of Table 6. For a given model selection criterion, the numerical results indicate that the computation of $\hat{\text{GIM}}(\lambda)$ is not sensitive to the different choices of $\tau$. As illustrated in Section 4.1, an approximately unbiased estimator $\hat{\text{ISI}}(\lambda)$ of $\text{ISI}(\lambda)$ can be obtained based on a data perturbation technique as $\tau \to 0^+$ (Huang and Chen (2007)), but it may produce numerical instability if $\tau$ is too small (e.g., $\tau = 10^{-10}$). Therefore, we use $\tau = 0.1$ throughout the simulation scenarios I and II that results in a somewhat larger bias but a smaller variance (e.g., Shen and Huang (2006), Huang and Chen (2007), Chen, Yang, and Li (2014)).

The results from simulation scenarios I and II suggest that the size of the underlying true model is an important factor that impacts the instability of geostatistical regression model selection. In practice, we may simultaneously perform conditional information criteria (e.g., CAIC and CBIC), the RIC criterion, and the empirical rule in Section 4.2 to jointly infer the size of the underlying true model. If the number of covariates in the underlying true model is relatively small, a model
selection procedure will be relatively unstable and thus a model averaging method is preferred for more accurate spatial prediction, while the conditional information criterion or the RIC criterion is recommended for variable selection. On the other hand, when selection uncertainty decreases (see, e.g., the left panel of Table 3 and Figure 2(b)), the model selection procedure tends to be stable and thus model selection and model averaging will give quite comparable prediction results (see, e.g., Table 1 and Figure 2(a)) under a given model selection criterion and the corresponding model averaging method. Thus, prediction and variable selection can be made based on a model selection criterion, where the conditional information criteria are still preferable.

6 Precipitation Data Example

To demonstrate the application of our proposed methodology, a precipitation data set in the state of Colorado (Reich and Davis (2008), Chu, Zhu, and Wang (2011)) as illustrated in Section 5.3 is analyzed. We consider all the possible combinations of covariates (i.e., $2^{10}$ candidate models) to investigate the size of the underlying geostatistical regression model for this data set.

Guided by the simulation study, we apply RIC, CAIC, CBIC, and the empirical rule developed in Section 4.2 to jointly estimate the size of the underlying geostatistical regression model. Table 7 provides three best models which are respectively selected by RIC, CAIC, and CBIC among $2^{10}$ candidate models. In addition, the estimated results of $\text{ReGIM}_j(\lambda)$ of (26) for $j = 1, \ldots, 10$ are shown in Figure 3. For estimating $\text{ReGIM}_j(\lambda)$, an estimate of $\text{GIM}_j(\lambda)$ of (25) is computed based on the estimation procedure of $\text{GIM}(\lambda)$ in Section 4.1 with $\mathcal{M}$ being replaced with $\mathcal{P}_j$, where $\mathcal{P}_j$ is the class of all subsets of size $j$ that are selected from $\mathcal{P} = \{1, \ldots, 10\}$. For example, $\mathcal{P}_1 = \{\{1\}, \ldots, \{10\}\}$, $\mathcal{P}_2 = \{\{a, b\} : a, b = 1, \ldots, 10, \text{ and } a \neq b\}$, $\cdots$, $\mathcal{P}_{10} = \{\{1, \ldots, 10\}\}$ as illustrated in Section 4.2. The results in Table 7 indicate that there should be 2 to 3 important covariates in the geostatistical regression model for precipitation. Two covariates (elevation and
B4M) are selected by RIC, which agrees with a previous analysis (Chu, Zhu, and Wang (2011)), whereas three covariates (elevation, slope and B1M) are selected by CAIC and two covariates (slope and B7M) are selected by CBIC. It is also quite clear that ReGIM$_j$(λ) becomes more stable when $j$ is more than 2, as the elbow pattern indicates in Figure 3. These results suggest that the size of the underlying geostatistical regression model for this data set is relatively small and thus model averaging would be more suitable for predicting precipitation at unsampled locations. Further, RIC, CAIC, and CBIC are the recommended model selection criteria for choosing the important covariates.

7 Conclusions and Discussion

In this paper, a novel GIM criterion is proposed to measure the instability of geostatistical regression model selection. The proposed criterion takes into account the uncertainties of model selection and parameter estimation and thus more accurately reflects the complexity of a given model fitting procedure. A theoretical connection between the GIM criterion and the notion of generalized degrees of freedom in geostatistical regression models is also established.

From the results of two simulation scenarios and a data example, we recommend that the conditional information criteria (e.g., CAIC and CBIC), the RIC criterion, and the empirical rule developed in Section 4.2 can be used in practice together to estimate the size of the underlying geostatistical regression model. When model selection uncertainty is relatively large (e.g., when the model size is relatively small in our simulation scenarios and the data example), the model averaging method is preferred when spatial prediction is the main interest, and the conditional information criterion or the RIC criterion is recommended for identifying important covariates. However, when model selection uncertainty decreases, the predictive performance by model selection and model averaging tends to be similar. Thus, prediction and selection of covariates can both be obtained.
from a model selection criterion, where the conditional information criterion is preferable.

Although our proposed GIM criterion is developed under the settings of geostatistical regression models, it could also be extended to explore model selection or model averaging under other modeling frameworks such as linear mixed-effects models. For parameter estimation, it is known that parameters in the Matérn correlation function cannot be estimated well, even when increasing amounts of data are collected densely in a fixed domain (i.e., the infill asymptotic framework). In addition, Irvine, Gitelman, and Hoeting (2007) have compared the performance of ML and REML estimates of spatial covariance parameters under various sampling designs via simulation studies and showed that these estimates still have room for improvement. While our focus here is on the selection of covariates, the instability of the estimation of spatial covariance parameters is an important issue to address and could be explored based on the notion of our proposed GIM criterion. Further theoretical and numerical investigations are of interest, which we leave for future research.

A limitation of the proposed method is its computational intensity when dealing with massive data sets or a large number of covariates, because the estimation procedure of GIM(\(\lambda\)) in Section 4.1 is time-consuming that involves inverting an \(n \times n\) covariance matrix and selecting a best model from a large class of candidate models. To overcome such challenges, computationally more efficient methods can be considered such as the least angle regression (e.g., Efron, Hastie, Johnstone, and Tibshirani (2004)), covariance tapering (e.g., Furrer, Genton, and Nychka (2006)), and the low-rank methods (e.g., Cressie and Johannesson (2008)). We leave the theoretical and numerical investigations for future research.

**Acknowledgments**

We would like to thank the editor (Professor Ruey S. Tsay), the associate editor, and the two referees for their helpful comments and suggestions. In particular, we would like to thank Professor
Appendix:

**Proof of Proposition 1.** Since \( Q_M(\theta_M) = X_M(X_M', \Sigma_Z^{-1} X_M)^{-1} X_M', \Sigma_Z^{-1} \), we rewrite \( \hat{H}_M(\theta_M) \) of (9) as

\[
\hat{H}_M(\theta_M) = \hat{\Sigma}_Z^{-1} X_M(X_M', \Sigma_Z^{-1} X_M)^{-1} X_M', \hat{\Sigma}_Z^{-1} + \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1}
\]

\[
= \hat{\Sigma}_Z^{-1} Q_M(\theta_M) + \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1}
\]

\[
= \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1} - \Sigma_{\eta}(\theta_M) \Sigma_{Z}^{-1} Q_M(\theta_M) + \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1} Q_M(\theta_M) + \hat{\Sigma}_{\eta, M} \hat{\Sigma}_Z^{-1} Q_M(\theta_M)
\]

\[
= \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1} (I - Q_M(\theta_M)) + \left( \Sigma_{\eta}(\theta_M) + \hat{\Sigma}_{\eta, M} I \right) \left( \Sigma_{\eta}(\theta_M) + \hat{\Sigma}_{\eta, M} I \right)^{-1} Q_M(\theta_M)
\]

It follows that

\[
\text{tr} \left( \hat{H}_M(\theta_M) \right) = \text{tr} \left( \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1} (I - Q_M(\theta_M)) \right) + \text{tr} \left( Q_M(\theta_M) \right)
\]

\[
= \text{tr} \left( \Sigma_{\eta}(\theta_M) \hat{\Sigma}_Z^{-1} (I - Q_M(\theta_M)) \right) + |M|.
\]

Combining the above equation with (13), we obtain the desired result. This completes the proof.

**Proof of Proposition 2.** Following (12), (15), and the definition of ISI(\( \lambda \)) in (21), we have

\[
\delta^{-1} \left\{ \hat{S}_{\gamma}(\lambda)(s_i; Z + \delta e_i) - \hat{S}_{\gamma}(\lambda)(s_i; Z) \right\}
\]

\[
= \delta^{-1} \left\{ \left[ \hat{H}_{\gamma}(\lambda)(0; Z + \delta e_i) \right] - \left[ \hat{H}_{\gamma}(\lambda)(\hat{\theta}_{\gamma}(\lambda); Z) \right] \right\} = \left[ \hat{H}_{\gamma}(\lambda)(\hat{\theta}_{\gamma}(\lambda)) \right]_{ii}
\]

\[
= \delta^{-1} \left[ \hat{\Sigma}_Z^{-1} X_{\gamma}(\lambda) \left\{ X_{\gamma}(\lambda)' \hat{\Sigma}_Z^{-1} X_{\gamma}(\lambda) \right\}^{-1} X_{\gamma}(\lambda)' \hat{\Sigma}_Z^{-1} \right]_{ii} + \left[ \hat{\Sigma}_{\eta}(\hat{\theta}_{\gamma}(\lambda)) \hat{\Sigma}_Z^{-1} \right]_{ii},
\]

26
where $\hat{\Sigma}_Z = \hat{\Sigma}_n(\hat{\theta}_{\hat{\gamma}(\lambda)}) + \hat{\sigma}^2 \hat{\varepsilon}_{\hat{\gamma}(\lambda)} I$, $[A]_{ii}$ denotes the $i$th element of vector $A$, and $[B]_{ii}$ denotes the $i$th diagonal element of matrix $B$. Let $P$ and $G$ be two $n \times n$ positive definite matrices and let $R$ be an $n \times m$ matrix with $\text{rank}(R) = m$. The following four basic properties of positive definite matrices (e.g., Harville (1997)) are used in proving Proposition 2: (i) $P^{-1}$ and $G^{-1}$ are positive definite; (ii) $R'PR$ and $R'GR$ are positive definite; (iii) $PG$ and $GPG$ are positive definite; (iv) If $PG = GP$, then $PG$ is also positive definite. Because $\hat{\Sigma}_n(\hat{\theta}_{\hat{\gamma}(\lambda)})$ and $\hat{\Sigma}_Z$ are $n \times n$ positive definite matrices and $X_{\hat{\gamma}(\lambda)}$ is an $n \times (|\hat{\gamma}(\lambda)| + 1)$ design matrix with $\text{rank}(X_{\hat{\gamma}(\lambda)}) = |\hat{\gamma}(\lambda)| + 1$, $\hat{\Sigma}_Z^{-1} X_{\hat{\gamma}(\lambda)} (X_{\hat{\gamma}(\lambda)}' \hat{\Sigma}_Z^{-1} X_{\hat{\gamma}(\lambda)})^{-1} X_{\hat{\gamma}(\lambda)}' \hat{\Sigma}_Z^{-1}$ is positive definite by (i)-(iii) and $\hat{\Sigma}_n(\hat{\theta}_{\hat{\gamma}(\lambda)}) \hat{\Sigma}_Z^{-1}$ is also positive definite by (i) and (iv). As a consequence, $[\hat{\Sigma}_Z^{-1} X_{\hat{\gamma}(\lambda)} (X_{\hat{\gamma}(\lambda)}' \hat{\Sigma}_Z^{-1} X_{\hat{\gamma}(\lambda)})^{-1} X_{\hat{\gamma}(\lambda)}' \hat{\Sigma}_Z^{-1}]_{ii} > 0$ and $[\hat{\Sigma}_n(\hat{\theta}_{\hat{\gamma}(\lambda)}) \hat{\Sigma}_Z^{-1}]_{ii} > 0$ for $i = 1, \ldots, n$. Thus,

$$\text{ISI}(\lambda) = E \left[ \lim_{\delta \to 0} \delta^{-1} \sum_{i=1}^{n} \left| \hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z + \delta e_i) - \hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z) \right| \right]$$

$$= \sum_{i=1}^{n} \lim_{\delta \to 0} \delta^{-1} E \left[ \hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z + \delta e_i) - \hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z) \right]$$

$$= \sum_{i=1}^{n} \lim_{\delta \to 0} \delta^{-1} \left[ EE(\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z + \delta e_i) | S) - EE(\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z) | S) \right]$$

$$= \sum_{i=1}^{n} E \left[ \lim_{\delta \to 0} \delta^{-1} \left( E(\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z + \delta e_i) | S) - E(\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z) | S) \right) \right]$$

$$= \sum_{i=1}^{n} E \left( \frac{\partial}{\partial S(s_i)} E(\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z) | S) \right),$$

where $\hat{S}_{\hat{\gamma}(\lambda)}(s_i; Z)$ and $\hat{S}_{\hat{\gamma}(\lambda)}(s_i; \hat{\theta}_{\hat{\gamma}(\lambda)})$ are the same as illustrated in (21). This completes the proof.

References


Table 1: Average squared prediction errors of six model selection criteria (AIC, BIC, AICc, RIC, CAIC, CBIC) and six model averaging methods (AIC-MA, BIC-MA, AICc-MA, RIC-MA, CAIC-MA, CBIC-MA) under five true models ($k$) based on 200 simulation replicates for simulation scenario I. The values in parentheses are the corresponding standard errors.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>0.5357</td>
<td>0.5318</td>
<td>0.5264</td>
<td>0.5201</td>
<td>0.5172</td>
</tr>
<tr>
<td></td>
<td>(0.0193)</td>
<td>(0.0189)</td>
<td>(0.0188)</td>
<td>(0.0184)</td>
<td>(0.0184)</td>
</tr>
<tr>
<td>AIC-MA</td>
<td>0.4598</td>
<td>0.4739</td>
<td>0.4814</td>
<td>0.4944</td>
<td>0.5130</td>
</tr>
<tr>
<td></td>
<td>(0.0184)</td>
<td>(0.0180)</td>
<td>(0.0174)</td>
<td>(0.0161)</td>
<td>(0.0150)</td>
</tr>
<tr>
<td>BIC</td>
<td>0.5044</td>
<td>0.5055</td>
<td>0.5029</td>
<td>0.5082</td>
<td>0.5379</td>
</tr>
<tr>
<td></td>
<td>(0.0216)</td>
<td>(0.0205)</td>
<td>(0.0200)</td>
<td>(0.0188)</td>
<td>(0.0187)</td>
</tr>
<tr>
<td>BIC-MA</td>
<td>0.4352</td>
<td>0.4568</td>
<td>0.4693</td>
<td>0.4924</td>
<td>0.5267</td>
</tr>
<tr>
<td></td>
<td>(0.0190)</td>
<td>(0.0184)</td>
<td>(0.0175)</td>
<td>(0.0151)</td>
<td>(0.0134)</td>
</tr>
<tr>
<td>AICc</td>
<td>0.5284</td>
<td>0.5241</td>
<td>0.5057</td>
<td>0.5123</td>
<td>0.5325</td>
</tr>
<tr>
<td></td>
<td>(0.0204)</td>
<td>(0.0203)</td>
<td>(0.0197)</td>
<td>(0.0190)</td>
<td>(0.0186)</td>
</tr>
<tr>
<td>AICc-MA</td>
<td>0.4461</td>
<td>0.4627</td>
<td>0.4722</td>
<td>0.4912</td>
<td>0.5205</td>
</tr>
<tr>
<td></td>
<td>(0.0187)</td>
<td>(0.0182)</td>
<td>(0.0175)</td>
<td>(0.0154)</td>
<td>(0.0137)</td>
</tr>
<tr>
<td>RIC</td>
<td>0.4876</td>
<td>0.4993</td>
<td>0.4842</td>
<td>0.5079</td>
<td>0.5488</td>
</tr>
<tr>
<td></td>
<td>(0.0217)</td>
<td>(0.0207)</td>
<td>(0.0191)</td>
<td>(0.0188)</td>
<td>(0.0182)</td>
</tr>
<tr>
<td>RIC-MA</td>
<td>0.4303</td>
<td>0.4540</td>
<td>0.4677</td>
<td>0.4945</td>
<td>0.5359</td>
</tr>
<tr>
<td></td>
<td>(0.0192)</td>
<td>(0.0185)</td>
<td>(0.0175)</td>
<td>(0.0148)</td>
<td>(0.0131)</td>
</tr>
<tr>
<td>CAIC</td>
<td>0.3866</td>
<td>0.4086</td>
<td>0.4263</td>
<td>0.4300</td>
<td>0.4580</td>
</tr>
<tr>
<td></td>
<td>(0.0155)</td>
<td>(0.0158)</td>
<td>(0.0141)</td>
<td>(0.0111)</td>
<td>(0.0114)</td>
</tr>
<tr>
<td>CAIC-MA</td>
<td>0.3650</td>
<td>0.3749</td>
<td>0.3974</td>
<td>0.4198</td>
<td>0.4660</td>
</tr>
<tr>
<td></td>
<td>(0.0159)</td>
<td>(0.0126)</td>
<td>(0.0118)</td>
<td>(0.0094)</td>
<td>(0.0101)</td>
</tr>
<tr>
<td>CBIC</td>
<td>0.3720</td>
<td>0.4416</td>
<td>0.4482</td>
<td>0.4699</td>
<td>0.5508</td>
</tr>
<tr>
<td></td>
<td>(0.0164)</td>
<td>(0.0199)</td>
<td>(0.0169)</td>
<td>(0.0141)</td>
<td>(0.0145)</td>
</tr>
<tr>
<td>CBIC-MA</td>
<td>0.3538</td>
<td>0.4092</td>
<td>0.4190</td>
<td>0.4597</td>
<td>0.5303</td>
</tr>
<tr>
<td></td>
<td>(0.0162)</td>
<td>(0.0175)</td>
<td>(0.0136)</td>
<td>(0.0116)</td>
<td>(0.0111)</td>
</tr>
</tbody>
</table>

32
Table 2: Frequencies of the number of selected covariates for six information criteria (AIC, BIC, AICC, RIC, CAIC, CBIC) under five true models (k) based on 200 simulation replicates for simulation scenario I. The symbol “∗” indicates that the true model is selected.

<table>
<thead>
<tr>
<th>k</th>
<th>Criterion</th>
<th>Number of covariates</th>
<th>Average number of selected covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>AIC</td>
<td>0 0∗ 0 0 3 12 12 24 34 115</td>
<td>8.10</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>0 48∗ 79 55 13 5 0 0 0 0</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>AICC</td>
<td>0 1∗ 19 57 77 43 3 0 0 0</td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td>RIC</td>
<td>0 76∗ 85 32 6 1 0 0 0 0</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>CAIC</td>
<td>0 28∗ 55 60 31 17 4 4 1 0</td>
<td>2.94</td>
</tr>
<tr>
<td></td>
<td>CBIC</td>
<td>0 78∗ 63 24 15 11 3 4 2 0</td>
<td>2.27</td>
</tr>
<tr>
<td>3</td>
<td>AIC</td>
<td>0 0 0 0 0 2 13 22 35 128</td>
<td>8.37</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>0 0 0 70∗ 84 43 2 1 0 0</td>
<td>3.90</td>
</tr>
<tr>
<td></td>
<td>AICC</td>
<td>0 0 0 10∗ 72 82 31 5 0 0</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>RIC</td>
<td>0 0 0 106∗ 67 26 1 0 0 0</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>CAIC</td>
<td>0 0 2 45∗ 73 52 21 6 1 0</td>
<td>4.34</td>
</tr>
<tr>
<td></td>
<td>CBIC</td>
<td>0 0 12 98∗ 61 18 5 4 2 0</td>
<td>3.63</td>
</tr>
<tr>
<td>5</td>
<td>AIC</td>
<td>0 0 0 0 0 1∗ 4 11 44 140</td>
<td>8.59</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>0 0 0 0 0 103∗ 77 19 1 0</td>
<td>5.59</td>
</tr>
<tr>
<td></td>
<td>AICC</td>
<td>0 0 0 0 58∗ 101 36 5 0 0</td>
<td>5.94</td>
</tr>
<tr>
<td></td>
<td>RIC</td>
<td>0 0 0 0 1 129∗ 60 10 0 0</td>
<td>5.40</td>
</tr>
<tr>
<td></td>
<td>CAIC</td>
<td>0 0 1 0 15 79∗ 72 29 4 0</td>
<td>5.62</td>
</tr>
<tr>
<td></td>
<td>CBIC</td>
<td>0 0 0 3 31 113∗ 43 8 2 0</td>
<td>5.14</td>
</tr>
<tr>
<td>7</td>
<td>AIC</td>
<td>0 0 0 0 0 0 0 4∗ 36 160</td>
<td>8.78</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>0 0 0 0 0 1 6 140∗ 51 2</td>
<td>7.24</td>
</tr>
<tr>
<td></td>
<td>AICC</td>
<td>0 0 0 0 0 2 118∗ 75 5 0</td>
<td>7.42</td>
</tr>
<tr>
<td></td>
<td>RIC</td>
<td>0 0 0 0 0 1 11 151∗ 35 2</td>
<td>7.13</td>
</tr>
<tr>
<td></td>
<td>CAIC</td>
<td>0 0 0 0 1 4 32 106∗ 52 5</td>
<td>7.10</td>
</tr>
<tr>
<td></td>
<td>CBIC</td>
<td>0 0 0 1 2 12 50 110∗ 24 1</td>
<td>6.71</td>
</tr>
<tr>
<td>9</td>
<td>AIC</td>
<td>0 0 0 0 0 0 0 0 2 198∗</td>
<td>8.99</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>0 0 0 0 0 0 3 5 30 162∗</td>
<td>8.76</td>
</tr>
<tr>
<td></td>
<td>AICC</td>
<td>0 0 0 0 0 0 2 31 167∗</td>
<td>8.83</td>
</tr>
<tr>
<td></td>
<td>RIC</td>
<td>0 0 0 0 0 4 9 46 141∗</td>
<td>8.62</td>
</tr>
<tr>
<td></td>
<td>CAIC</td>
<td>0 0 0 0 0 1 33 6 16 121∗</td>
<td>8.45</td>
</tr>
<tr>
<td></td>
<td>CBIC</td>
<td>0 0 0 0 1 8 17 31 72 71∗</td>
<td>7.89</td>
</tr>
</tbody>
</table>
Table 3: Average values of generalized instability measure (GIM) of six information criteria (AIC, BIC, AICc, RIC, CAIC, CBIC) under five true models \((k)\) (left panel) for simulation scenario I and sensitivity analysis of the proposed generalized instability measure (GIM) with respect to various perturbation sizes \(\tau\) for \(k = 5\) (right panel) based on 200 simulation replicates under simulation scenario I. The values in parentheses are the corresponding standard errors.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>(1)</th>
<th>(3)</th>
<th>(5)</th>
<th>(7)</th>
<th>(9)</th>
<th>(0.1)</th>
<th>(0.3)</th>
<th>(0.5)</th>
<th>(0.7)</th>
<th>(0.9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>1.92</td>
<td>1.83</td>
<td>1.74</td>
<td>1.67</td>
<td>1.62</td>
<td>1.74</td>
<td>1.69</td>
<td>1.68</td>
<td>1.70</td>
<td>1.70</td>
</tr>
<tr>
<td>BIC</td>
<td>3.76</td>
<td>2.86</td>
<td>2.23</td>
<td>1.86</td>
<td>1.68</td>
<td>2.23</td>
<td>2.23</td>
<td>2.22</td>
<td>2.26</td>
<td>2.29</td>
</tr>
<tr>
<td>AICc</td>
<td>3.16</td>
<td>2.64</td>
<td>2.17</td>
<td>1.85</td>
<td>1.68</td>
<td>2.17</td>
<td>2.20</td>
<td>2.19</td>
<td>2.22</td>
<td>2.25</td>
</tr>
<tr>
<td>RIC</td>
<td>3.91</td>
<td>2.94</td>
<td>2.19</td>
<td>1.87</td>
<td>1.73</td>
<td>2.19</td>
<td>2.23</td>
<td>2.22</td>
<td>2.26</td>
<td>2.30</td>
</tr>
<tr>
<td>CAIC</td>
<td>2.10</td>
<td>1.63</td>
<td>1.41</td>
<td>1.31</td>
<td>1.28</td>
<td>1.41</td>
<td>1.38</td>
<td>1.35</td>
<td>1.35</td>
<td>1.32</td>
</tr>
<tr>
<td>CBIC</td>
<td>1.87</td>
<td>1.36</td>
<td>1.16</td>
<td>1.15</td>
<td>1.25</td>
<td>1.16</td>
<td>1.17</td>
<td>1.15</td>
<td>1.15</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Table 4: Regression coefficients of three true models \((k)\) of simulation scenario II based on the precipitation data set.

<table>
<thead>
<tr>
<th>Regression coefficients</th>
<th>(k)</th>
<th>(\beta_0)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
<th>(\beta_3)</th>
<th>(\beta_4)</th>
<th>(\beta_5)</th>
<th>(\beta_6)</th>
<th>(\beta_7)</th>
<th>(\beta_9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2950</td>
<td>0.5716</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1.2948</td>
<td>0.4554</td>
<td>0.0344</td>
<td>-0.0027</td>
<td>-0.1395</td>
<td>0.0139</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1.3105</td>
<td>0.4425</td>
<td>0.0317</td>
<td>-0.0085</td>
<td>0.1902</td>
<td>-0.0108</td>
<td>-0.0265</td>
<td>-0.2487</td>
<td>0.1269</td>
<td>-0.1769</td>
</tr>
</tbody>
</table>

Table 5: Parameter values in the covariance matrix of the data vector and the SNR values for three true models \((k)\) of simulation scenario II based on the precipitation data set.

<table>
<thead>
<tr>
<th>Parameter values</th>
<th>(k)</th>
<th>(a)</th>
<th>(\nu)</th>
<th>(\sigma^2_I)</th>
<th>(\sigma^2_I)</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0985</td>
<td>0.5</td>
<td>0.2153</td>
<td>0.2119</td>
<td></td>
<td>2.56</td>
</tr>
<tr>
<td>5</td>
<td>0.1220</td>
<td>0.5</td>
<td>0.1497</td>
<td>0.2281</td>
<td></td>
<td>1.66</td>
</tr>
<tr>
<td>9</td>
<td>0.1435</td>
<td>0.5</td>
<td>0.1745</td>
<td>0.2281</td>
<td></td>
<td>2.27</td>
</tr>
</tbody>
</table>
Table 6: Average values of generalized instability measure (GIM) of six information criteria (AIC, BIC, AICc, RIC, CAIC, CBIC) under three true models ($k$) (left panel) for simulation scenario II and sensitivity analysis of the proposed generalized instability measure (GIM) with respect to various perturbation sizes $\tau$ for $k = 5$ (right panel) based on 200 simulation replicates and the precipitation data set under simulation scenario II. The values in parentheses are the corresponding standard errors.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>$k$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>AIC</td>
<td>4.12</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.35)</td>
</tr>
<tr>
<td>BIC</td>
<td>4.85</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>AICc</td>
<td>4.19</td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
<td>(0.36)</td>
</tr>
<tr>
<td>RIC</td>
<td>4.81</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td>(0.44)</td>
</tr>
<tr>
<td>CAIC</td>
<td>2.49</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>CBIC</td>
<td>1.41</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.18)</td>
</tr>
<tr>
<td></td>
<td>3.88</td>
<td>3.64</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.24)</td>
</tr>
<tr>
<td></td>
<td>4.67</td>
<td>4.16</td>
</tr>
<tr>
<td></td>
<td>(0.45)</td>
<td>(0.32)</td>
</tr>
<tr>
<td></td>
<td>3.98</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.25)</td>
</tr>
<tr>
<td></td>
<td>4.63</td>
<td>4.07</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.31)</td>
</tr>
<tr>
<td></td>
<td>2.22</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.14)</td>
</tr>
<tr>
<td></td>
<td>1.59</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(0.13)</td>
</tr>
</tbody>
</table>

Table 7: Selected models with the corresponding covariates (“Yes”) among $2^{10}$ candidate models for RIC, CAIC, and CBIC applied to the precipitation data example.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Covariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIC</td>
<td>Elevation Yes Slope - Aspect - B1M - B2M - B3M - B4M - B5M - B6M - B7M -</td>
</tr>
<tr>
<td>CAIC</td>
<td>Yes - Yes - - - - - - - - - - - -</td>
</tr>
<tr>
<td>CBIC</td>
<td>- Yes - - - - - - - - Yes</td>
</tr>
</tbody>
</table>

35
Figure 1: Realizations of $\eta(\cdot)$ for $(\sigma_{\eta}^2, a, \nu) = (1, 0.1, 0.5)$ (left) and $(\sigma_{\eta}^2, a, \nu) = (1, 1, 0.5)$ (right).

Figure 2: (a) Average squared prediction error (ASPE) versus five true models ($k$) for four information criteria (AIC, BIC, CAIC, CBIC) and four model averaging methods (AIC-MA, BIC-MA, CAIC-MA, CBIC-MA) based on 200 simulation replicates under simulation scenario I; (b) Average value of generalized instability measure (GIM) versus five true models ($k$) for six information criteria (AIC, BIC, AICc, RIC, CAIC, CBIC) based on 200 simulation replicates under simulation scenario I.
Figure 3: Estimates of relative difference of generalized instability measures (ReGIM) for RIC, CAIC, and CBIC applied to the precipitation data example.