Spatially Correlated Model Selection Method (SCOMS)

Ciro Velasco-Cruz

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Scotland C. Leman, Co-Chair
Eric P. Smith, Co-Chair
Feng Guo
Leanna House

May 04, 2012
Blacksburg, Virginia

Keywords: Spatial statistics; Variable Selection; Non-stationary spatial fields; Ising prior.

Spatially Correlated Model Selection Method (SCOMS)

by

Ciro Velasco-Cruz

ABSTRACT

In this dissertation, a variable selection method for spatial data is developed. It is assumed that the spatial process is non-stationary as a whole but is piece-wise stationary. The pieces where the spatial process is stationary are called regions. The variable selection approach accounts for two sources of correlation: (1) the spatial correlation of the data within the regions, and (2) the correlation of adjacent regions. The variable selection is carried out by including indicator variables that characterize the significance of the regression coefficients. The Ising distribution as prior for the vector of indicator variables, models the dependence of adjacent regions.

We present a case study on brook trout data where the response of interest is the presence/absence of the fish at sites in the eastern United States. We find that the method outperforms the case of the probit regression where the spatial field is assumed stationary and isotropic. Additionally, the method outperformed the case where multiple regions are assumed independent of their neighbors.

Acknowledgments

I thank the Department of Statistics at Virginia Tech for giving me the opportunity to complete my Ph.D. in Statistics.

I want to thank my committee members, but especially Dr. Eric P. Smith and Dr. Scotland C. Leman for their infinite patience and their guidance. I truly believe that without their help I would not complete this work. To Dr. Leman, I honor his optimism and his confidence in me.

To Dr. Jeff Birch who gave me the moral support when I once was in such a need of it, and obviously for his outstanding lectures. I also want to include my friends Bob Hale and his wife, Marlene Hale, that have been always around providing me with the moral support as well, needed throughout the Ph.D.

To my friends and classmates: Dipayan Maiti and Huaiye Zhang, with whom I spent a considerable amount of time discussing statistical issues, and with the same intensity and interest, personal matters, as well.

I thank all the people who have taken part in one way or another on the completion of this thesis research. I hope you, dear and kind friends of mine, do not get upset if your name is not explicitly mentioned above. But be sure that you have a place of honor in my memory and my humble appreciation and love from the bottom of my heart.

Always yours,

Ciro Velasco-Cruz.
(Spring 2012)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In 2000, the Easter Brook Trout Joint Venture (EBTJV) conducted a large study to assess the relationship between the status of the brook trout fish and a set of covariables. A total of 3,337 observational sites were sampled (EBTJV 2006). The sites of observation are scattered along the eastern United States as can be seen in Fig. 1.1, and correspond to streams where the trout exists or used to exist. The information available at the sites includes the status variable that describes whether the brook trout fish is present or absent, and 63 landscape and anthropogenic covariables (also known as stressor metrics; Hudy, Thieling, Gillespie and Smith (2008)). Some analyses have been conducted on this dataset; for example, Thieling (2006) summarized and screened out some stressor metrics, based on their completeness, range, redundancy and responsiveness. In other analysis, Zhang, Thieling, Prins, Smith and Hudy (2008) performed a stepwise variable selection logistic regression, where status was the response variable and five stressor metrics were the regressors (some of them derived as combinations of the original covariables). Zhang et al.'s (2008) goals were to find which of the considered stressor metrics were important to predict the response variable and to develop a model with good classification rates, measured in terms of the Average Fraction Correctly Classified for Fit (AFCCF; Wilkinson (1999)). Zhang et al.'s (2008) approach is to partition the spatial field into regions and to perform a logistic regression at each region. As a result, Zhang et al. found that the case of a single region was outperformed by the case when the spatial field was split into 5 or 6 regions.

Motivated by Zhang et al.'s (2008) goals and approach on the analysis of the brook trout data, in this research we propose a comprehensive methodology with the following charac-

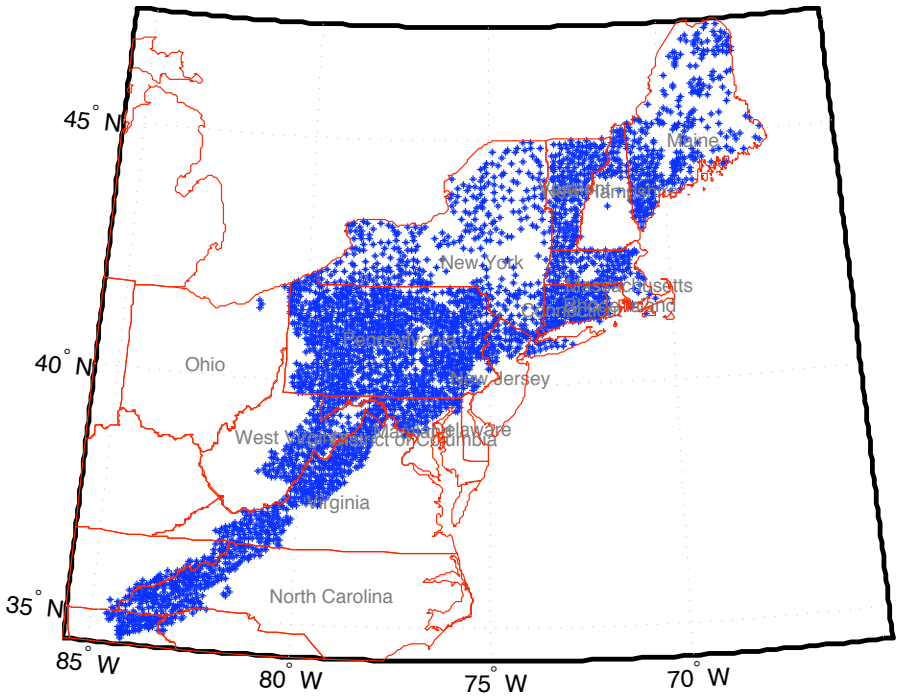


Figure 1.1: Distribution of the brook trout fish dataset. Dots (in blue) are observational sites.

teristics. It assumes that the spatial field that generates the spatial data is non-stationary, but rather piecewise stationary. Therefore, it partitions the spatial field into mutually exclusive regions and fits a regression model at each region. However, in our method, the regression model fit takes into account the spatial correlation of the data available in each region and the correlation between regions, two sources of dependence whose joint effect has never been taken into account before. The between regions dependence is effectively considered by the probability of including a covariable during a variable selection step, thanks to the Ising distribution (Higdon 1998*a*; Møller, Pettitt, Reeves and Berthelsen 2006). The Ising distribution provides a simple form to accommodate the regions dependence. As a result, the models in the regions are correlated with one another. Later, we will see that this methodology is suitable for the analyses of binary data, using a probit regression.

Variable selection has been a recurrent issue in statistical applications. Often, there are situations when one wants parsimonious models to describe relationships between response variables and sets of covariables, simply because based on simpler models the relationships are easier to understand and explain. Simpler models are commonly built with the most significant covariables. In Bayesian variable selection, the uncertainty of whether the co-

variables are significant is quantified by means of indicator variables that characterize the significance of the covariables. The significance of a covariable is equivalently determined by the significance of its regression coefficient, so that the indicator variable is equal to one if the regression coefficient is significant, and zero if the coefficient is non-significant. To model the uncertainty of the indicator variable, sometimes a Bernoulli distribution is assumed. This is the case of the Stochastic Search Variable Selection (SSVS) method proposed by George and McCulloch (1993). This method lets the indicator variable be equal to one with probability p and zero with probability $1-p$, where p is constant and known (p is also called probability of inclusion). Several applications of the SSVS method have taken place since its release. For example, with spatially correlated data, Reich, Fuentes, Herring and Evenson (2010) applied the SSVS method to select significant spatially varying regression coefficients (Gelfand, Kim, Sirmans and Barnerjee 2003). Similarly, Kuo and Mallick (1998) let γ have a Bernoulli distribution, but in their approach the indicator variable is included as another parameter in the regression model, that jointly with the regression coefficient generates the *effect size* parameter. Kuo and Mallick proved that their method is closely related to SSVS. From a more general perspective, to model uncertainty about the indicator variable, γ, in their Spatial Bayesian Variable Selection (SBVS) method for lattice data (see Fig. 1.2 for an example of a lattice), Smith and Fahrmeir (2007) let γ have the Ising distribution. We come back to this method, and describe it thoroughly in Section 2.2.2.

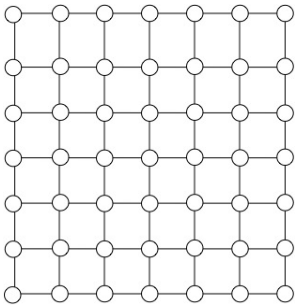


Figure 1.2: A regular lattice example. The circles are observations location. Ideally, all the circles are equally separated. This type of arrangement is commonly found in images, i.e. satellites images.

With respect to partitioning the spatial field into homogeneous regions, Kim, Mallick and Holmes (2005) proposed a method to analyze non-stationary Gaussian processes that divides the spatial field of interest into non-overlapping regions, where the spatial process is assumed stationary and isotropic. The regions are further assumed to be independent of

one another. In Eq. (2.9) we come back to this method.

After pursuing another way to fulfill Zhang et al.'s (2008) goals, we develop a general and comprehensive methodology that formally addresses each of the Zhang et al. goals and provides a more realistic summary of the data at hand. In other words, this thesis introduces an appropriate method for variable selection for spatially correlated data, when the spatial process that generates the data is non-stationary. The spatial field is assumed to be made up of mutually exclusive regions, where the process is stationary and where a set of significant covariables is selected. The significance of the variables is leveraged by the dependence between adjacent regions. The variable selection incorporates the dependence between adjacent regions thanks to the Ising distribution assumed as the probability of inclusion. We invite our interested reader to look at Chapter 3 where details of the method are given. For now, we only give a brief overview that we hope would help to sketch the big picture of the method, known as Spatially Correlated Model Selection (SCOMS).

1.1 An overview on the SCOMS method

Without loss of generality, we present basic concepts of the *Spatially Correlated Model Selection* (SCOMS) method using variables from the brook trout data, such as the status and stressor metrics, as reference. In the dataset, the response of interest is binary, thus either the probit or the logistic regression analyses are appropriate, and also lead to similar conclusions (Weisberg 2005; Myers 1990). For instance, Zhang et al. (2008) analyzed the brook trout dataset by fitting logistic regressions at each region. However, we opt for the probit regression, which is described here. A continuous latent variable, as in Albert and Chib (1993) and Oliveira (2000), is introduced in the analysis. In modeling binary responses, it is convenient to consider the continuous latent variable, typically assumed normally distributed, as the driving variable of the binary outcomes. Then, the two variables, the binary and the continuous, are related through the following mapping function. Let Z be the binary response, and Y be the continuous latent variable, so

$$Z(\mathbf{s}_l) = \begin{cases} 1 & \text{if } Y(\mathbf{s}_l) > 0 \\ 0 & \text{if } Y(\mathbf{s}_l) \leq 0, \end{cases} \tag{1.1}$$

where $\mathbf{s}_l$ is the vector of coordinates of the lth site, $l = 1, \ldots, n$, n is the number of sites, and the status variable, $Z(\mathbf{s}_l)$, is equal to either 1 or 0, depending on whether the trout is present or absent at site $\boldsymbol{s}_l$. In Eq. (1.1) the latent variable Y is assumed normally distributed. Therefore, $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)]'$ is probabilistically specified as

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{1.2}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and variance-covariance matrix of the multivariate normal distribution. We set $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, where $\boldsymbol{X}$ is a matrix of covariates of order $n \times q$, and $\boldsymbol{\beta}$ is a vector of regression coefficients of order q. And, as is typical in spatial applications, the variance-covariance matrix is defined as $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{H}(\cdot) + \tau^2 I$, with $\boldsymbol{H}(\cdot)$ being the spatial correlation matrix, and σ^2 and τ^2 are called the *partial sill* and the *nugget*. However, in this paper we approximate $\boldsymbol{\Sigma}$ by the CAR model (see Eq. (2.15)), for the reasons presented in Sections 2.3.5 and 2.4.

The binary process is completely characterized by the Gaussian latent process, and the mapping in Eq. (1.1) is unique so long as the variance of the spatial field is set to a constant. Usually, $\tau^2 = 1$ allows the identification of the parameters in $\boldsymbol{X}\boldsymbol{\beta}$ (see Section 2.4 for more details). Hence, the rest of this section focuses on the continuous latent variable only, as the response variable, provided that the mapping of Z into Y, or vice-versa, is one to one.

Once the probit model is established as above, the next part of the method is what makes the SCOMS method interesting. To simplify our procedure description, let us temporarily assume that the spatial field of interest is already partitioned into a given number of regions (although later we will see how to construct a partition of the spatial field). For instance, as an illustration let us consider the spatial field in Fig. 1.3. There we have that the field is partitioned into $R = 5$ regions. The assumed model for the continuous latent vector, $\boldsymbol{Y}_i$, is $N(\boldsymbol{X}_i\boldsymbol{\vartheta}_i, \boldsymbol{\Sigma}_i)$, for $i = 1, \ldots, R$, where $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are the response vector and the matrix of covariables corresponding to the ith region, and $\boldsymbol{\vartheta}_i = \boldsymbol{\beta}_i \circ \boldsymbol{\gamma}_i$ is known as the *effect size* (Kuo and Mallick 1998), where $\boldsymbol{\gamma}_i$ is a vector of indicator variables, defined below, and $\boldsymbol{\beta}_i$ is the vector of regression coefficient. The *effect size* parameter is the result of the Hadamard product between vectors $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_i$.

The selection of the most significant covariates that explain variation in the response variable is performed via the variable selection method proposed by (Kuo and Mallick 1998), with a slight modification that permits the accommodation of not only multiple regions but

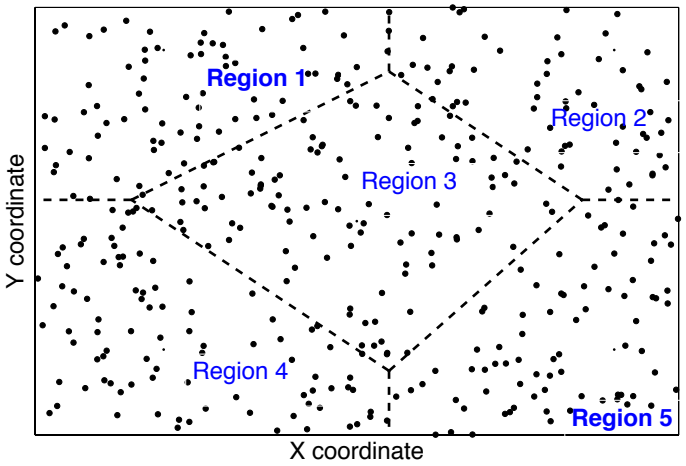


Figure 1.3: An example of a Voronoi tessellated spatial field into 5 regions. Each dot represent a simulated site in the spatial field. The Voronoi tessellation (in dashed lines) generates mutually exclusive partitions.

also the correlation between the regions. We let $\gamma_{i,j}$ be an indicator variable such that $\gamma_{i,j} = 1$ if the stressor metric $X_{i,j}$ is significant, and $\gamma_{i,j} = 0$ if $X_{i,j}$ is non-significant, for $i = 1, \ldots, R$ and $j = 1, \ldots, q$.

Now, let us define a vector of indicator variables to be $\boldsymbol{\gamma}_j = [\gamma_{1,j}, \ldots, \gamma_{R,j}]'$. This vector contains the indicator variables of the jth covariate across all regions. In order to provide cross-correlation between regions, we let

$$\pi(\boldsymbol{\gamma}_j|\theta, R) \propto \exp\left\{\sum_{i\sim i'} \theta\ \omega_{i,i'}\ I(\gamma_{i,j} = \gamma_{i',j})\right\}. \tag{1.3}$$

Eq. (1.3) is the Ising distribution (Higdon 1998*a*; Green and Richardson 2002; Smith and Fahrmeir 2007), where θ and $\omega_{i,i'}$ are intrinsic parameters of this model (both are defined later). The term within the curly brackets in the Ising distribution accounts for the correlation between adjacent regions ($i \sim i'$ is read as i' is adjacent to i). For completely independent regions, i.e. $\theta = 0$, Eq. (1.3) reduces to $\pi(\gamma_{i,j} = 1) = 1/2$. On the other hand, for dependent regions, the Ising distribution increases the probability of coincidences of the indicator variables of the same covariable in two adjacent regions, i.e., the correlation between two regions, i and i', increases the chances of $\gamma_{i,j}$ to be equal to $\gamma_{i',j}$ by an amount proportional to $[\theta\ \omega_{i,i'}]$. Eq. (1.3), therefore, represents a generalization of the probability of $\gamma_{i,j}$ when correlation between adjacent regions is taken into account. As a consequence

of using Eq. (1.3) as the probability of including a regressor in the linear model, $\boldsymbol{X}_i\boldsymbol{\vartheta}_i$, we obtain correlated models, with stronger correlation between models from adjacent regions. A detailed description of the Ising distribution as is utilized in this paper, is given later in Chapter 3.

The scope of this thesis is to present a method for variable selection with spatially correlated data, when the spatial process that generates the data is locally stationary and isotropic. The spatial process is assumed to be made up of mutually exclusive regions, where the process is stationary. The number of regions and their location in the field are unknown, however, the SCOMS method looks for the optimum number of regions. By "optimum" we mean that the number of regions, R, is selected as the one that optimizes one of the following criteria: AIC, BIC, DIC, AFCCF, etc. Furthermore, at each region a selection of the most significant variables occurs. The variable selection takes into account the dependence between adjacent regions; as a result, the models are spatially correlated. The use of the Ising distribution as prior probability of including a covariate in the variable selection step, allows the accommodation of the correlation between adjacent regions, into the fitted models. The metric for the dependence between adjacent regions, ω, is deterministic and based on physical characteristics of the regions, as explained in Section 3.5.2. The search for the partition of the spatial field, given the number of regions, is carried out using the Markov Chain Monte Carlo (MCMC) algorithm, where at each iteration, the spatial field is partitioned via Voronoi tessellations (Okabe, Boots, Sugihara and Chiu 2000).

1.2 The thesis structure

The methodology presented in this thesis is inspired by popular ideas and methods for spatial fields and model selection; Chapter 2 presents an abbreviated literature review that addresses concepts and statistical techniques that in one way or another are influential in building our methodology. In Chapter 3, the proposed methodology is presented in detail. We start from the basic concepts such as the field's partition, the model, the likelihood, etc., when multiple regions are of consideration. Above, in this brief introduction of the method, we presented it assuming that the response was binary, in order to illustrate the method's scope, however, this particular case is a generalization of the method. In Chapter 3 we present the case of the binary response as a generalization of our methodology, which is later utilized to analyze the brook trout data as the motivating example of the proposed methodology. In

that same chapter, we also address in detail the construction of the parameter ω in the Ising distribution, and how to learn about θ. Prior distributions for the unknown parameters are also presented in this chapter.

A simulation study, and the analyses of two subsets of the brook trout dataset as the case studies, along with their results are presented in Chapter 4. We separately explain how to simulate the datasets, and then how to use them for the validation and evaluation of the SCOMS method. The list of stressor metrics for the brook trout dataset, that are considered in the case studies, is given in this chapter as well. Conclusions based on the simulation study, the case studies and some other results, are presented in Chapter 5. The full conditional distributions and simulation results are presented in Appendices A and B.

Chapter 2

Literature review

2.1 Introduction

Let $\boldsymbol{Y}$ be a random vector with elements $[Y_1, \ldots, Y_n]'$, and let $\boldsymbol{X}$ be a matrix with column vectors of covariables of order n, such that $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_q]$. It is said that $\boldsymbol{Y}$ and $\boldsymbol{X}$ are linearly related if

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

is a valid model for $\boldsymbol{Y}$. Here, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_q]'$ is the vector of the regression coefficients, and $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_n]$ is the random vector with the multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance $\boldsymbol{\Sigma}$.

When the response vector, $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)]'$ whose elements are indexed by the coordinates $\boldsymbol{s}_l$ $(l = 1, \ldots, n)$, is spatially correlated, the variance-covariance of $\boldsymbol{\epsilon}$ is typically assumed structured. Commonly, $\boldsymbol{\Sigma}$ is assumed equal to

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{H}(\rho, \|\boldsymbol{s}_l - \boldsymbol{s}_l'\|) + \tau^2 \boldsymbol{I}_{n\times n},$$

where $\sigma^2\boldsymbol{H}(\cdot)$ is an isotropic spatial covariogram, $\boldsymbol{H}(\rho, \|\boldsymbol{s}_l - \boldsymbol{s}_l'\|)$ is the spatial correlation matrix with parameter ρ, and $\|\boldsymbol{s}_l - \boldsymbol{s}_l'\|$ is the distance between sites l and l' (The distance can be either Euclidian or Geodetic (Banerjee 2005) distance). In spatial statistics, the parameters σ^2 and τ^2 are called *partial sill* and *nugget*, and they have an interpretation that is better understood graphically. Fig. 2.1 illustrates a variogram $\gamma(d_{l,l'}) = \tau^2 + \sigma^2(1 - H(\rho, d_{l,l'}))$, where $d_{l,l'} = \|\boldsymbol{s}_l - \boldsymbol{s}_l'\|$; in the illustration, we see that the *sill* is located where the variogram

function flattens, the *nugget* is the variance at $d = 0$ (also known as the variability due to the measurement error), and the difference between the *nugget* and the *sill* is the *partial sill*. In Section 2.3.4, the three most recurrently used isotropic covariograms are presented.

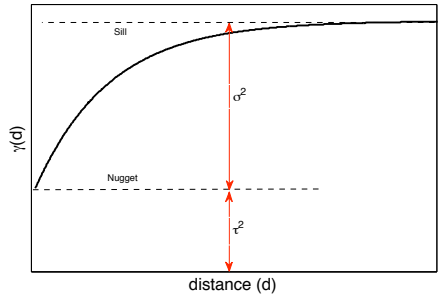


Figure 2.1: An illustration of an isotropic variogram with *partial sill* and *nugget*, σ^2, and τ^2, respectively.

Eq. (2.1) is the model assumed throughout this thesis. The SCOMS methodology modifies the mean and the covariance structure of the model above to obtain proper inferences from the analyses of spatial data whose spatial process is non-stationary. In the forthcoming sections, we address topics that are utilized directly in the SCOMS method and/or that make its application more efficient. For instance, in Section 2.2 some variable selection methods are described. All of them are aimed at determining which covariables should be included in matrix $\boldsymbol{X}$, in Eq. (2.1), in order to obtain parsimonious models. Not surprisingly, some useful features of some variable selection methods mentioned in that section will be made part of the SCOMS method. The approximation of the variance-covariance matrix, $\boldsymbol{\Sigma}$ in Eq. (2.1), by the CAR model is presented in Section 2.3. This approximation is proposed as a way to speed up the multivariate probit analysis in the case study. In that section, we also (a) give the theoretical background that justifies the spatial processes, (b) concretely present the Gaussian random process, highlighting its assumptions of stationarity and isotropy, and (c) present some methods used in practice when the last two assumptions are not supported by the realizations of the spatial processes. Because the response variable in the case study is binary, in Section 2.4 we present a way to make the connection between the Gaussian process and the binary random field.

Fitting and interpretation of the mean $\boldsymbol{X\beta}$ in Eq. (2.1), can represent challenges, especially when the number of regressors, q, is large. To build parsimonious and more understandable and manageable models, there are some variable selection methods designed to reduce the number of regressors in the model. The variable selection methods optimally

search for the regressors whose effects are significant in explaining the response variable. In the following sections we describe some variable selection procedures. This is not by any means an exhaustive enumeration of variable selection methods, but presumably, they are the most commonly used in practical applications.

2.2 Variable Selection methods

The purpose of variable selection procedures is to efficiently search for regressors that are important for explaining the variability of the response variable, and include them as part of the model $\boldsymbol{X\beta}$ in Eq. (2.1). In circumstances when q regressors are available, there are 2^q potential models, and frequently only a few of them are based upon the most significant regressors. The evaluation of all 2^q models represents computational burden even for small q. Hence, systematic procedures, with useful statistical properties have been used to ease and efficiently search for the subset of significant regressors for the response variable.

2.2.1 Likelihood based variable selection methods

Let us start by defining the likelihood function, a concept that will be repeatedly mentioned in what follows of this thesis. Let $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ be a random sample of size n, and let $f(\boldsymbol{y}|\boldsymbol{B}, \boldsymbol{X})$ be the joint distribution of the sample, a function with unknown parameters given in $\boldsymbol{B}$. Then, given that $\boldsymbol{Y} = \boldsymbol{y}$ is observed, the function on $\boldsymbol{B}$ defined by

$$L(\boldsymbol{B}|\boldsymbol{y}, \boldsymbol{X}) = f(\boldsymbol{y}|\boldsymbol{B}, \boldsymbol{X})$$

is called the likelihood function (Cassella and Berger 2002). For instance, the likelihood function for Model (2.1) is

$$L(\boldsymbol{B}|\boldsymbol{y}, \boldsymbol{X}) \propto \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta})\right\},$$

where $\boldsymbol{B} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$.

Criteria such as AIC, BIC and DIC, are based upon the likelihood function. In general, these criteria are typically utilized to discriminate among potentially useful models. Under the assumption that there are at least two competing suitable models for $\boldsymbol{Y}$, one of them is

preferred over the other(s) provided such model yields the smaller value of the AIC, BIC or DIC criterion.

Suppose there are two appropriate models for $\boldsymbol{Y}$, M_1 and M_2, so that $M_1 \subset M_2$; perhaps M_1 is a more parsimonious model. This would occur, for example, if in model M_1 one covariable is left out from $\boldsymbol{X\beta}$, while model M_2 has in $\boldsymbol{X\beta}$ all the available covariables. In this example, M_1 is a particular case of M_2. Each of the three information criteria can be used to discriminate between M_1 and M_2 as follows, recall that $\boldsymbol{B}$ is the parameter that contains all the unknown parameters in the model:

1. Akaike (1974) proposed *An Information Criterion* (AIC), given by:

$$AIC_i = -2\log L(\boldsymbol{B}_i|\boldsymbol{y}, \boldsymbol{X}, M_i) + 2q_i, \text{ for } i = \{1, 2\},$$

   where $L(\boldsymbol{B}_i|\boldsymbol{y}, \boldsymbol{X}, M_i)$ is the likelihood function for the parameters given in model M_i, and q is the dimension of the parameter space or the number of parameters independently adjusted for the maximization of the likelihood (Akaike 1974). To compute AIC, $L(\boldsymbol{B}|\boldsymbol{y}, \boldsymbol{X}, M)$ is evaluated at the Maximum Likelihood Estimator (MLE) of the parameters. For example, if the MLE for $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}$, then the likelihood function $L(\boldsymbol{B}|\boldsymbol{y}, \boldsymbol{X})$ is evaluated at $\widehat{\boldsymbol{\beta}}$. The term $2q$ is known as the adjusting term or penalty. The AIC criterion favors simpler models and penalizes complex ones.

   The AIC helps us to prefer one model over any other. For instance, suppose that $AIC_{M_1} < AIC_{M_2}$. Then, AIC suggests choosing M_1.

2. Schwarz (1978) introduced the *Bayesian Information Criterion* (BIC), given by

$$BIC_i = -2\log L(\boldsymbol{B}_i|\boldsymbol{y}, \boldsymbol{X}, M_i) + q_i \log(n), \text{ for } i = \{1, 2\}.$$

   The BIC and AIC are quite similar. Their obvious difference is that BIC penalizes the fit by the number of parameters in the model combined with the sample size, whereas the AIC only by the number of parameters.

   With BIC, we prefer a model over any other if such model has the smallest BIC. For example suppose that $BIC_{M_1} < BIC_{M_2}$. Therefore, BIC suggests M_1 as the preferred model (over M_2).

   One important feature of the BIC is that it is related to the Bayes Factor (Good 1983; Lee 2004; Kass and Raftery 1995). The Bayes Factor (BF) when comparing

models such as M_1 and M_2 above (albeit BFs compare any kind of models: non-nested and/or nested.), has the following relationship with BIC: $-2\log BF \approx BIC_{M_1} - BIC_{M_2}$, but it holds only when the prior distribution for the regression coefficients, $\boldsymbol{\beta}$, is multivariate normal with mean at the maximum likelihood estimate and variance equal to the expected information matrix for one observation (Raftery 1999; Kass and Raftery 1995).

3. Spiegelhalter, Best, Carlin and van der Linde (2002) proposed the *Deviance Information Criterion* (DIC). The *Deviance* for the regression coefficients is given by

$$D(\boldsymbol{\beta})_i = -2\log L(\boldsymbol{B}_i|\boldsymbol{y}, \boldsymbol{X}, M_i) + 2\log h(\boldsymbol{y}, \boldsymbol{\Sigma}|\boldsymbol{X}), \text{ for } i = \{1, 2\},$$

where $\boldsymbol{\beta}$ is the vector of regression coefficients in the model, and $h(\boldsymbol{y}, \boldsymbol{\Sigma}|\boldsymbol{X})$ is some standardizing factor, a function of the data alone. For example, Spiegelhalter et al. suggest that for distributions that are members of the exponential family, $h(\boldsymbol{y}) = L(\boldsymbol{y}|\boldsymbol{y}, \boldsymbol{\Sigma}, \boldsymbol{X})$, the saturated likelihood, where $\boldsymbol{\Sigma}$ is given as known (although any estimated of this parameter is utilized).

The posterior expectation of the deviance, $\bar{D}_i = E_{\boldsymbol{\beta}|\boldsymbol{y}}(D(\boldsymbol{\beta})_i)$, summarizes the model fit. The penalizing term, $p_{D_i} = \bar{D}_i - D(\bar{\boldsymbol{\beta}})_i$, is understood as the effective number of parameters or as a measure of the *complexity* of the model.

The DIC is defined as follows

$$DIC_i = \bar{D}_i + p_{D_i}.$$

Again, with DIC, we prefer one model over any other if such model has the smallest DIC. For example suppose that $DIC_{M_1} < DIC_{M_2}$. Therefore, DIC suggests M_1.

4. Bayes Factor (BF). Although it is not exactly a likelihood based statistic as the AIC, BIC, or DIC, the BF is also utilized to discriminate models. For Bayesian applications, the BF is the usual statistic to compare any kind of model, either nested or non-nested.

The BF is defined as follows. Recall that the joint distribution of the sample given above, is $f(\boldsymbol{y}|\boldsymbol{B}, \boldsymbol{X})$, and let $P(M_1)$ and $P(M_2)$ be the prior probabilities for models, M_1 and M_2, then the BF that compares model M_1 versus M_2 (Newton and Raftery

1994) is

$$BF = \frac{\int_{\boldsymbol{B}\in\Omega_{M_1}} f(\boldsymbol{y}|\boldsymbol{B},\boldsymbol{X},M_1)P(\boldsymbol{B}|M_1)P(M_1)d\boldsymbol{B}}{\int_{\boldsymbol{B}\in\Omega_{M_2}} f(\boldsymbol{y}|\boldsymbol{B},\boldsymbol{X},M_2)P(\boldsymbol{B}|M_2)P(M_2)d\boldsymbol{B}}, \tag{2.2}$$

where $P(\boldsymbol{B}|M_i)$ is the prior probability distribution for the unknown parameters under model M_i for $i = \{1, 2\}$, and Ω_{M_i} is the parameter space defined by model M_i.

If $BF = 1$ the two compared model are equally preferable, if $BF > 1$ ($BF < 1$) then model M_1 (M_2) should be preferred over M_2 (M_1). For instance, suppose that for our models above, $BF > 1$, then the Bayes factor suggests that model M_1 should be preferred.

For two equiprobable models where $P(M_1) = P(M_2)$, Eq. (2.2) is simplified as

$$BF = \frac{\int_{\boldsymbol{B}\in\Omega_{M_1}} f(\boldsymbol{y}|\boldsymbol{B},\boldsymbol{X},M_1)P(\boldsymbol{B}|M_1)d\boldsymbol{B}}{\int_{\boldsymbol{B}\in\Omega_{M_2}} f(\boldsymbol{y}|\boldsymbol{B},\boldsymbol{X},M_2)P(\boldsymbol{B}|M_2)d\boldsymbol{B}}.$$

The AIC, BIC and DIC criteria evaluate the model fit taking into account the model's complexity through the penalizing term. In statistical analyses, there is not a preferred criterion out of the three, and their utilization depends essentially on the model formulations and the adopted statistical paradigm, namely the frequentist or Bayesian approaches. For instance, for Bayesian hierarchical model specifications, DIC is more appropriate than BIC and AIC (Banerjee, Carlin and Gelfand 2004). Additionally, DIC does not present extra computational issues in Bayesian applications, since it can be computed from the available MCMC samples directly. However, for the AIC and BIC, a maximization over the parameter space is required before their computation, which translates into extra computational time.

The AIC and BIC criteria have been used in an alternative way. Sometimes the mean, $\boldsymbol{X\beta}$ is held fixed, and the difference between two potential models, M_1 and M_2, lies on their variance-covariance structure, say for example, for M_1 a compound symmetry covariance structure is assumed, whilst for M_2 an unstructured one (Littell, Milliken, Stroup, Wolfinger and Schabenberger 2006). Following the given rule, one model is preferred over the other provided its AIC or BIC is the smallest. In some other instances (regression) the variance parameter may be viewed as fix and the regression coefficients variable.

2.2.2 Bayesian variable selection

O'Hara and Sillanpää (2009) reviewed the most common Bayesian variable selection methods. From O'Hara and Sillanpää's (2009) list of methods, the Stochastic Search Variable Selection (SSVS) proposed by George and McCulloch (1993) and Kuo and Mallick's (1998) method are the ones that we present here. Even though it is not in the O'Hara and Sillanpää's (2009) list, we thoroughly describe Smith and Fahrmeir's (2007) method because it introduces the Ising distribution into the variable selection scheme, and the Ising distribution plays a central role in our methodology.

Stochastic Search Variable Selection (SSVS) method

The Stochastic Search Variable Selection (SSVS, George and McCulloch (1993)) is a Bayesian hypothesis test method to determine the significance of the regressor coefficients. Let us start the exposition of the SSVS by letting β_j be the jth regressor coefficient in Eq. (2.1). Now suppose we have the following problem: we want to determine whether β_j belongs to Ω_0 or Ω_A, two subsets of the parameter space Ω, where $\Omega_0 \cup \Omega_A = \Omega$ and $\Omega_0 \cap \Omega_A = \varnothing$. Traditionally, two hypotheses are proposed to equivalently specify the same problem. One of the hypotheses is known as the *null* (H_0) and the other is known as the *alternative* (H_A), which are usually expressed as follows,

$$H_0 : \beta_j \in \Omega_0 \text{ vs } H_A : \beta_j \in \Omega_A. \tag{2.3}$$

For each hypothesis a prior probability of the form $\pi(H_0) = 1 - p_j$ and $\pi(H_A) = p_j$ is assumed (Good 1983). An equivalent specification of the hypotheses above, is as follows. Let γ_j be an indicator variable, such that $\pi(\gamma_j = 0) = 1 - p_j$ and $\pi(\gamma_j = 1) = p_j$. Hence,

$$\beta_j|\gamma_j \sim (1 - \gamma_j)N(0, \tau_j^2) + \gamma_j N(0, c_j^2\tau_j^2). \tag{2.4}$$

Eq. (2.4) can be understood as the prior distribution for β_j in the SSVS method. The first component of Eq. (2.4), that refers to the *null* hypothesis, assumes that $\beta_j \sim N(0, \tau_j^2)$, and the second component referring to the *alternative* hypothesis, assumes that $\beta_j \sim N(0, c_j^2\tau_j^2)$, with $c_j > 0$. When $c_j > 1$ the distribution for β_j is more diffuse. One of the drawbacks of this procedure resides in the best way to specify the components τ and c in Eq. (2.4). There are some solutions to this problem that we will not address here (see George and McCulloch

(1993) for more about this method).

Other method

Kuo and Mallick (1998) proposed a variable selection method for regression models of the form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\vartheta} + \boldsymbol{\epsilon}, \tag{2.5}$$

where $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\vartheta} = [\beta_1\gamma_1, \ldots, \beta_q\gamma_q]'$ is defined as the *effect size* parameter, the regression coefficients $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_q]'$ are the same as those given in Eq. (2.1), and γ_j is the same indicator variable as in Eq. (2.4). This formulation of the variable selection problem

1. allows the quantification of the importance of regressors through the posterior distribution $p(\boldsymbol{\gamma}|\boldsymbol{y})$, and

2. improves the computational efficiency of the algorithm for model fits, because it allows the use of the Gibbs algorithm (Kuo and Mallick 1998).

Spatial Bayesian Variable Selection

We start by defining lattice data. An example is shown in Fig. 2.2. Lattice data is commonly found in images, such as satellite images. In the figure, each empty circle represents a vertex where information is recorded. Typically, for lattice data the vertices are assumed dependent on adjacent vertices. In the illustration, we show one possible arrangement of the dependence that a vertex can have with its adjacent vertices: vertex v_5 is dependent on vertices $v_1, v_2, v_3,$ and v_4.

Smith and Fahrmeir (2007) proposed a variable selection method for lattice data, their Spatial Bayesian Variable Selection (SBVS) method. We give a more than brief description of the SBVS method because the variable selection step in our own method (see Chapter 3) works similarly. Here, a prior probability distribution for p_j in Eq. (2.4), is adopted.

Above in the SSVS method, $\pi(\gamma_j = 1) = p_j$, where p_j is fixed and known. However, in the SBVS method p_j is unknown, hence, a prior distribution for this parameter is formulated. Here, the Ising distribution (Higdon 1994; Higdon 1998*a*; Møller et al. 2006) is assumed as the prior distribution.

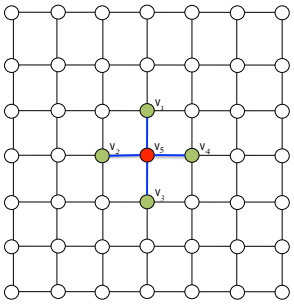


Figure 2.2: The neighborhood structure of dependence in lattice data. Vertex v_5 is connected with locations v_1, v_2, v_3, and v_4. Therefore, the v_5 neighborhood is $NB(v_5) = \{v_1, v_2, v_3, v_4\}$.

Let $\{Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n_i}\}$ be a sequence of observations recorded at the ith vertex of a given lattice (for $i = 1, \ldots, N$, where N is the total number of vertices.). Also, let $\boldsymbol{X}_i$ be a matrix of q covariates available at vertex i. Then,

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i,$$

is the model that describes the relationship between $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$, where $\boldsymbol{\beta}_i = [\beta_{i,1}, \ldots, \beta_{i,q}]'$ is a vector of q regression coefficients. Fig. 2.3 shows an illustration of how the measurements are collected. In a sense, the same lattice is replicated n_i times. At each vertex, measurements of $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ are recorded.

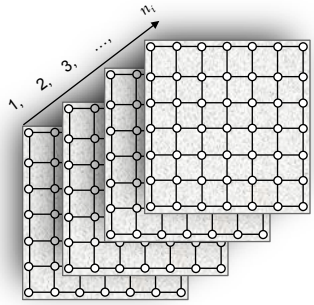


Figure 2.3: Illustration of the data collection procedure for the SBVS method. The same lattice is measured a given number of times.

Let $\boldsymbol{\gamma}_i = [\gamma_{i,1}, \ldots, \gamma_{i,q}]'$ be a vector of indicator variables, where $\gamma_{i,j} = 0$ when $\beta_{i,j} = 0$ and $\gamma_{i,j} = 1$ when $\beta_{i,j} \neq 0$, for $j = 1, \ldots, q$. For variable $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_N']'$, the Ising distribution is assumed as its prior. In particular, if $\boldsymbol{\gamma}_{(j)} = [\gamma_{1,j}, \gamma_{2,j}, \ldots, \gamma_{N,j}]'$ is the vector of indicator

variables corresponding to the jth regressor across vertices $\{1, 2, \ldots, N\}$, the prior density for $\boldsymbol{\gamma}_{(j)}$ is

$$\pi(\boldsymbol{\gamma}_{(j)}) \propto \exp\left\{\sum_{i=1}^{N} \alpha_{i,j}(\gamma_{i,j}) + \sum_{i\sim k} \theta_{i,k,j}\omega_{i,k} I(\gamma_{i,j} = \gamma_{k,j})\right\}, \tag{2.6}$$

i.e. the Ising distribution, where $i \sim k$ means that vertex k is adjacent to vertex i, $\omega_{i,k}$ is the weight of dependence between vertices i and k, and $\sum_{i=1}^{N} \alpha_{i,j}(\gamma_{i,j})$ is the external field. The term $\sum_{i\sim k} \theta_{i,k,j}\omega_{i,k} I(\gamma_{i,j} = \gamma_{k,j})$ is the interaction effect associated with the elements of $\boldsymbol{\gamma}_{(j)}$ for all pairwise neighboring sites.

Eq. (2.6) allows the inclusion of the correlation between neighboring vertices, in the following way: if $\gamma_{k,j} = 0$ at all vertices k that are neighbors of vertex i, then the Ising distribution invites $\gamma_{i,j} = 0$, as well. The indicator variables depend not only on the information within the region, but also on the information from the neighboring regions. Therefore, the linear model in a region is correlated with models from adjacent regions.

When vertices i and k are independent, Eq. (2.6) reduces to a constant, leading back to the SSVS method. It is apparent that this method is a generalization of the SSVS method, and accommodates the spatial correlation of the vertices of the lattice data through the indicator variables.

In sciences such as those related to the environment, the study of the relationships between a set of covariables believed to drive the variations of a feature of interest and the feature of interest (a.k.a. the response variable), is frequently carried out by assuming that their relationship can be explained with a simple model such as the one given in Eq. (2.1). Commonly, the assumed probability distribution in naturally continuous responses, such as air temperature, total rainfall (Stroud, Müller and Sansó 2001), etc., (typically observed at monitoring sites) is the multivariate normal distribution, with parameters $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. The following section presents a review on the theoretical background that supports spatial analyses, when a Gaussian response vector is of interest.

2.3 Gaussian Random Fields

Let us turn to the spatial statistics foundations, specifically to the Gaussian Random Fields in $\mathcal{D}$ ($\mathcal{D} \in \mathbb{R}^2$). It is important to introduce in some detail some of the concepts that we believe help to build a strong understanding of the method we present later in the next

chapter. Let us start with stationarity and isotropy in Gaussian stochastic processes, two concepts that play important roles in spatial statistics.

Let $\{Y(\boldsymbol{s});\ \boldsymbol{s} \in \mathcal{D}\}$ be a stochastic process in $\mathbb{R}^2$, where $\boldsymbol{s}$ is the coordinates vector, i.e. $\boldsymbol{s} = [\text{longitude}, \text{latitude}]$. If $\boldsymbol{Y} = \{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\}$ follows the multivariate Gaussian probability density function, then, the process is said to be a Gaussian Process (GP). In spatial statistics, the components of the random vector $\boldsymbol{Y}$ of the process are assumed spatially dependent, with the degree of association defined by the location of the random variable (Banerjee et al. 2004). Two concepts are important for random processes, *stationarity* and *isotropy*.

2.3.1 Stationarity

A stochastic process is *strictly stationary* if any two vectors, $\{Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)\}$ and $\{Y(\boldsymbol{s}_1 + \boldsymbol{h}), \ldots, Y(\boldsymbol{s}_n + \boldsymbol{h})\}$ with $\boldsymbol{h} \in \mathbb{R}^2$, have the same probability distribution. A less restrictive condition of stationarity for stochastic processes, is *weakly stationary* (for Gaussian processes both concepts are equivalent). A random process is said to be *weakly stationary* (Cressie 1993; Banerjee et al. 2004) if

$$\begin{aligned} &E(Y(\boldsymbol{s})) = \mu \\ &\text{Cov}(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})) = C(\boldsymbol{h}), \quad \text{for } \boldsymbol{s} \text{ and } \boldsymbol{s}+\boldsymbol{h} \in \mathcal{D}, \end{aligned} \tag{2.7}$$

i.e. neither moments depend on the location $\boldsymbol{s}$.

For example, Eq. (2.1) is a GP with mean $\boldsymbol{X\beta}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, both independent of location. A more realistic assumption is one that allows μ and $\boldsymbol{\Sigma}$ be location dependent: the non-stationary case. Such a case is addressed below. For now, we turn to the concept that defines the geometry of the spatial correlation. When the variance-covariance matrix of a stochastic process is only a function of the distance between any two observations of the process at different locations, the spatial field is said to be *isotropic*, otherwise it is said to be *anisotropic*.

2.3.2 Isotropy and anisotropy

When the spatial dependence is not only function of the distance but may also depend on other factors, then a spatial field is said to be *anisotropic*. A particular case of *anisotropy* is the *geometric anisotropy*, which happens when the correlation of the spatial field depends on distance and orientation (Banerjee et al. 2004; Schabenberger and Gotway 2005). The geometry of the correlation of an *isotropic* spatial field depicts circular contours, while the geometry of the correlation of a *geometric anisotropic* spatial field describes elliptical contours. Fig. 2.4 illustrates both, the *isotropic* and the *geometric anisotropic* spatial fields. In both, the strength of the dependence becomes weaker as distance increases. In the left panel, the geometry of an *isotropic* field is depicted, where concentrical circles describe the spatial correlation from the center, c. Two points located at the same distance from c, say a and b, have the same correlation with c. In the right panel, an example of the *geometric anisotropic* field is shown, where ellipses are surrounding a central point c. Two points a and b, equally distant from c, have different correlations with c; the correlation of a with c is weaker than the correlation of b with c. Such difference is due to the positions where the points a and b are with respect to c.

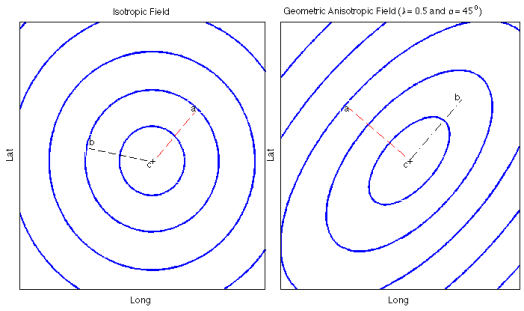


Figure 2.4: An isotropic and a geometric anisotropic random fields. The left panel illustrates the geometry of the spatial correlation of an *isotropic* random field. The right panel illustrates the geometry of the spatial correlation of a *geometric anisotropic* field.

It is known that the *geometric anisotropy* is driven by two parameters, namely the angle of rotation of the main axis of the ellipse (α), and the ratio between the minor and major axes of the ellipse (λ). To accommodate *geometric anisotropy*, a transformation of the coordinates

of the field is carried out as follows

$$\boldsymbol{s}^* = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} \text{longitude} \\ \text{latitude} \end{pmatrix}. \tag{2.8}$$

Where $\boldsymbol{s}^*$ the transformed coordinates vector. The unknowns, α and λ, should be estimated. This transformation turns a *geometric anisotropic* field into an *isotropic* one.

We saw that when the spatial field is *geometric anisotropic*, a simple closed form transformation of the field is enough to handle it. Such transformation orthogonally rotates the spatial field to make it isotropic. Nevertheless, there are situations when the *anisotropicity* of the spatial fields is not only due to the orientation of the sites, as it is in the case of the geometric anisotropy fields, but also due to some other unknown factors, so that simple transformations of the fields are no longer a practical solution. Therefore, other more suitable approaches are used for the analysis of such spatial fields. Some of them are mentioned in the following section.

2.3.3 Non-stationary and non-isotropic spatial fields methods

In particular, a Gaussian process is said to be non-stationary if its parameters are location-dependent as follows

$$\begin{aligned} &E(Y(\boldsymbol{s})) = \mu(\boldsymbol{s}) \\ &\text{Cov}(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})) = C(\boldsymbol{h}, \boldsymbol{s}), \quad \text{for } \boldsymbol{s} \text{ and } \boldsymbol{s}+\boldsymbol{h} \in \mathcal{D}, \end{aligned} \tag{2.9}$$

where the mean and the covariance of the spatial field depend on the location $\boldsymbol{s}$.

When stationarity and isotropy cannot be supported by the realizations of the spatial process, more general approaches for the analyses of such spatial fields are available. Sometimes, although the data do not support it, a spatial process is assumed stationary and as a result, information about the process is lost due to this oversimplification of the problem. Instead, more realistic and richer, but at the same time more complicated approaches should be used for the analyses of non-stationary spatial data. In this section, some methods recurrently applied in practice are presented. These methods allow the analysis of not only non-stationary processes but also non-isotropic ones.

The moving window method

For the analysis of non-stationary spatial fields, Haas (1995) proposed the moving windows approach that consists in constructing a symmetric spatial window around a given site $\boldsymbol{s}_0$, where a prediction of the process is pursued. In this method, the spatial process within the "window" is assumed to be stationary and isotropic. To obtain good predictions, the size of the "window" around $\boldsymbol{s}_0$ is defined as follows. Suppose there are a total of n sites of observation within the spatial field of interest. Hence, there is an $f \in (0, 1)$ such that the number of observations within the window, n_w, is $n_w = n \times f$. The f is known as the fraction of sites that are used to predict site $\boldsymbol{s}_0$. The window's size is determined by the number of observations that the window embraces. An optimal number of observations needed to predict the spatial process at the site $\boldsymbol{s}_0$, or equivalently the optimal fraction of the sample, f, is estimated by cross-validation.

The spatial field deformation method

Sampson and Guttorp (1992) proposed a transformation of the coordinates of the spatial field, similarly as we proceeded above when the spatial fields are geometrically anisotropic (see Section 2.3.2). This method starts by defining the *spatial dispersion* function of the spatial field as $D^2(\boldsymbol{s}_i, \boldsymbol{s}_j) = \text{var}(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))$, where $\{Y(\boldsymbol{s}_i), i = 1, \ldots, n\}$ is a realization of a non-stationary, non-isotropic spatial process. It is assumed that there is a function $f(\cdot)$ such that $\boldsymbol{s}_i^* = f(\boldsymbol{s}_i)$, where $f(\cdot)$ is a transformation function: $\mathbb{R}^2 \to \mathbb{R}^2$. A very particular example is given in Eq. (2.8), but in general, $f(\cdot)$ does not have a closed form. In the transformed space, there is a monotone function $g(\cdot)$ such that $d_{i,j}^2 = g(\|\boldsymbol{s}_i^* - \boldsymbol{s}_j^*\|)$, where $d_{i,j}^2 = \hat{\text{var}}(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))$, an estimate of D^2. An estimate of g, $\hat{g}$, given the $d_{i,j}$ is found via 2D Non-Metric Multidimensional Scaling (NM-MDS) (Izenman 2008). NM-MDS determines a monotone function $\delta \ni \delta(d_{i,j}) = \delta_{i,j} \approx \|\boldsymbol{s}_i^* - \boldsymbol{s}_j^*\|$ implying that $d_{i,j}^2 = (\delta^{-1}(\delta_{i,j}))^2 \approx \hat{g}(\|\boldsymbol{s}_i^* - \boldsymbol{s}_j^*\|)$. An estimate of $f(\cdot)$, $\hat{f}$, is found via thin-plate spline interpolation. The transformed field with coordinates $\boldsymbol{s}^*$, called the *D plane*, is assumed stationary and isotropic. An estimate of the *spatial dispersion* of two sites $\boldsymbol{s}_{i_1}$ and $\boldsymbol{s}_{i_2}$ can be found by first transforming their coordiantes using $\hat{f}$: $\boldsymbol{s}_{i_1}^* = \hat{f}(\boldsymbol{s}_{i_1})$ and $\boldsymbol{s}_{i_2}^* = \hat{f}(\boldsymbol{s}_{i_2})$, then using $\hat{g}$: $d_{i_1,i_2}^2 \approx \hat{g}(\|\boldsymbol{s}_{i_1}^* - \boldsymbol{s}_{i_2}^*\|)$.

Kernel method

Another method for analyzing non-stationary spatial random fields, is based upon *kernel convolution*. A stationary Gaussian process, $\{Y(\boldsymbol{s}_i), i = 1, \ldots, n\}$ can be expressed in terms of the convolution of a white noise process, $\epsilon(\boldsymbol{s})$, and a kernel function, $K(\boldsymbol{s})$, as $Y(\boldsymbol{s}) = \int_{\mathbb{R}^2} K(\boldsymbol{s} - \boldsymbol{u})\epsilon(\boldsymbol{u})d\boldsymbol{u}$. Where $K(\boldsymbol{s})$ is an arbitrary kernel function, for example, $K(\boldsymbol{s}) \propto \exp\{-\frac{1}{2}\|\boldsymbol{s}\|^2\}$ (Higdon 2001). Higdon (1998*b*) modified the kernel function to allocate non-stationarity by just letting K vary with $\boldsymbol{s}$: $K_{\boldsymbol{s}}(\boldsymbol{s})$. This approach lets $K_{\boldsymbol{s}}(\boldsymbol{s})$ smooth the spatial field by convolving different locally stationary (not necessary isotropic) subfields within the spatial field.

Spatial field partition method

Kim et al. (2005) proposed a method to analyze non-stationary Gaussian processes. This method divides the spatial field of interest into non-overlapping pieces called regions, where for each region the process is assumed stationary and isotropic. The regions are further assumed to be independent with one another. Under the assumption that there are a number of piecewise stationary and isotropic, as well as independent regions within the field, a Bayesian algorithm is proposed to search for the regions. The number of regions, R, is assumed to be random. The reversible jump algorithm (Green 1995) is utilized to find the optimal number of regions.

Kim et al.'s (2005) approach becomes an important piece of the SCOMS method we present in the next chapter. We slightly generalize this method by allowing the regions to be dependent of one another, provided the regions are adjacent. Later, in the next chapter we provide details of the SCOMS method. For the time being, it is enough to say that the SCOMS method combines the Ising distribution with the Kuo and Mallick (1998) variable selection method and the Kim et al. (2005) method (for more details see Chapter 3).

Up until now, it should be apparent that approaches that tackle the non-isotropic and non-stationary spatial processes find paths that lead to isotropic fields. In all the methods above, the isotropy assumption is present, either local as in the Kim et al. (2005) and Haas (1995) methods, or global as in the Sampson and Guttorp (1992) method. These methods are evidence of the important role that the concept of isotropy plays in spatial fields, and with the idea of making this section as self-contained as possible, in the following section we

list some of the most commonly used isotropic covariogram functions. For a more exhaustive list of these covariograms, we recommend our interested reader to look at the books by Banerjee et al. (2004) and Schabenberger and Gotway (2005), that we believe are good and complete references for Bayesian and frequentist spatial analyses.

2.3.4 Isotropic Covariance functions

Although non-stationary spatial random fields are more common in practice, stationary fields play an important role in spatial statistics analyses, since most of the procedures listed above that deal with non-stationary processes assume local or, for the deformation method, global stationarity. Whether it is global or local, a stationary spatial random field, defined in $\mathcal{D} \in \mathbb{R}^2$, is a random process with a covariance structure $C(\cdot)$, called covariance function, and it is isotropic if C is a function of the distance between sites only.

There are a number of parametric covariance functions for stationary and isotropic spatial fields. Here, we present the isotropic covariance functions that are commonly used in spatial statistical applications.

1. *Spherical*

$$C(\sigma^2, \phi, d) = \sigma^2 \left(1 - \frac{3}{2} d/\phi + \frac{1}{2}(d/\phi)^3\right), \quad 0 < d \leq \phi. \tag{2.10}$$

   Where $\phi \geq 0$ is the range of the spatial dependence, $\sigma^2 > 0$ is the variance of the field, and d is the distance between pair of locations.

2. *Matérn*

$$C(\sigma^2, \nu, \phi, d) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\phi d)^\nu K_\nu(\phi d), \quad d > 0. \tag{2.11}$$

   Matérn (1986) suggested this covariogram, where $\nu \in (1/2, \infty)$ is a parameter that controls the smoothness of the random field in $\mathcal{D} \in \mathbb{R}^2$, $\phi \geq 0$ is the range of the spatial dependence, σ^2 is the variance of the field, $d \geq 0$ is the separation between two locations, and $K_\nu(\cdot)$ is the modified Bessel function of second kind of order ν.

   There are two particular cases of Eq. (2.11) that results in the two well known covariogram functions,

- *Exponential.* When $\nu = 1/2 \Rightarrow$

$$C(\sigma^2, \phi, d) = \sigma^2 \exp(-\phi d). \tag{2.12}$$

- *Gaussian.* When $\nu \to \infty \Rightarrow$

$$C(\sigma^2, \phi, d) = \sigma^2 \exp(-\phi^2 d^2). \tag{2.13}$$

When the sample size and the extension of the spatial field are large, the analysis of Gaussian Random fields (GRF) carries two complications: (1) the assumptions of stationarity and isotropy are hardly realistic, and (2) large sample sizes carry a computational issue that is known as the big n problem. Alternatives to handle non-stationary and non-isotropic fields were mentioned above. In the following section we present a method that we found really helpful to deal with the big n problem. This approach approximates a GRF by a Markov Random Field (MRF), resulting in a Gaussian Markov Random Fields (GMRF). The approximation is accomplished by the Conditional Autoregressive (CAR) model (Besag 1975; Rue and Tjelmeland 2002; Banerjee, Gelfand, Finley and Sang 2008).

2.3.5 Conditional Autoregressive (CAR) models

CAR models are simple models for spatial data. Before (1975), they were utilized exclusively for lattice data. In 1971, the Hammersley-Clifford theorem (not published by the authors but presented in Besag (1974)) was used to validate their utilization to model Gaussian spatial processes, or somewhat equivalently to the analysis of higher order Markov processes, where the spatial dependence of a site of observation extends to not only on its nearest neighbors (which is the case for the first order Markov process) but to all the sites. Besag (1975) proposed to use CAR models for the analysis of non-lattice data, and he called the new approach the *spatial stochastic interaction for irregularly distributed data points.* The main characteristic of the new approach was that every site becomes a neighbor of every other site. Besag's (1975) paper marked a new way to analyze irregularly distributed spatial data.

Rue and Tjelmeland (2002) showed that the CAR model is the result of the combination of GRFs with MRFs, leading to the Gaussian Markov Random Fields (GMRF) (Banerjee et al. 2008). CAR models have enjoyed a dramatic increase in usage only in the past decade (Banerjee et al. 2004), simply because they provide an easy solution for the big n problem

in Bayesian methods (Banerjee et al. 2008).

Let us assume that a spatial field is observed at fixed locations $l = 1, \ldots, n$. Besag (1975) defined the CAR model as follows.

$$y_l|\boldsymbol{y}_{-l} \sim N\left(\mu_l + \rho \sum_{l'} w_{l,l'}(y_{l'} - \mu_{l'})/w_{l\cdot}, \tau^2/w_{l\cdot}\right). \tag{2.14}$$

Eq. (2.14) is the conditional distribution of the random variable Y_l at site l given the realizations of the random vector $\boldsymbol{y}_{-l}$ at the remaining sites, $\boldsymbol{y}_{-l} = (y_1, \ldots, y_{l-1}, y_{l+1}, \ldots y_n)$. The parameter ρ drives the spatial correlation. The parameter $w_{l,l'}$ is a metric of the spatial association between pair of observations, such that $w_{l,l} = 0$ and $w_{l,l'} > 0$ for $l \neq l'$, $w_{l\cdot} = \sum_{l'=1}^{n} w_{l,l'}$, and $\tau^2/w_{l\cdot}$ is the conditional variance. The mean function of the CAR model includes the dependence of y_l on $\boldsymbol{y}_{-l}$, and the variables become independent when $\rho = 0$.

The joint distribution of $\boldsymbol{Y}$ corresponding to Eq. (2.14), is

$$f(\boldsymbol{y}) \propto \exp\left(-\frac{1}{2\tau^2}(\boldsymbol{y} - \boldsymbol{\mu})'(\boldsymbol{D}_w - \rho\boldsymbol{W})(\boldsymbol{y} - \boldsymbol{\mu})\right), \tag{2.15}$$

the multivariate normal distribution, with mean $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, and variance-covariance matrix $\boldsymbol{\Sigma} = \tau^2(\boldsymbol{D}_w - \rho\boldsymbol{W})^{-1}$, where $\boldsymbol{W} = (w_{l,l'})$ is a symmetric matrix of weights, and $\boldsymbol{D}_w = \text{diag}(w_{i\cdot})$.

For the conditional autoregressive models presented in Besag (1975), one has to carefully choose the components of the $\boldsymbol{W}$ matrix. Banerjee et al. (2004) proposed the construction of $\boldsymbol{W}$ as follows. The weight for the spatial dependence of site l on site l', $w_{l,l'} = (\boldsymbol{W})_{l,l'}$, should decay with distance. Intuition says that two observational sites that are close may have stronger dependence on each other than two sites that are far apart. Therefore, a sensible metric for the spatial association between sites is $w_{l,l'} = 1/d_{l,l'}$, where $d_{l,l'}$ is the distance between sites l and l'.

In general, the nature of the response vector $\boldsymbol{Y}$ determines the most appropriate probability distribution to assume. For the linear model given in Eq. (2.1), we assumed that our response variable is continuous and the data supports the normality assumption. It is not always the case. For example, in the brook trout data the response is binary, thus the Bernoulli probability distribution seems to be more appropriate for the response variable in

the data, hence a Binary random field should be assumed. Nevertheless, when the response variable is binary, we can still use the Gaussian process with all its proprieties to analyze the Binary random process. The following section explains the use of a latent Gaussian process that helps to accomplish the task.

2.4 Clipped Gaussian Process

To analyze Binary Random Fields (BRF), Oliveira (2000) proposed the Clipped Gaussian (CG) random fields technique, which is a generalization of the method proposed by Albert and Chib (1993) to the case when the data are spatially correlated. The CG introduces an unobserved latent GRF process underneath a BRF.

The BRF partitions the field into two disjoint subfields. For example, in the brook trout dataset, Z(**s**)=1 at sites where the trout is present, and Z(**s**)=0 where the trout is absent. Therefore, we have a spatial process $\{Z(\boldsymbol{s}_l); l = 1, \ldots, n\}$ made of 0s and 1s. The CG proposes to partition the continuous GRF according to the BRF, in the following way.

Let $\{Z(\boldsymbol{s}_l), \boldsymbol{s}_l \in \mathcal{D}\}_{l=1}^{n}$ be a single realization of the BRF. There exists $\{Y(\boldsymbol{s}_l), \boldsymbol{s}_l \in \mathcal{D}\}_{l=1}^{n}$ such that

$$Z(\boldsymbol{s}_l) = \begin{cases} 1 & \text{if} \quad Y(\boldsymbol{s}_l) \geq 0 \\ 0 & \text{if} \quad Y(\boldsymbol{s}_l) < 0, \end{cases} \tag{2.16}$$

where $\boldsymbol{Y} = [Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)]'$ is a realization of GRF, with the same distribution properties given in Model Eq. (2.1). When mapping the BRF into GRF, it is known that there are an infinite number of possible values of the variance of the GRF that yields the same likelihood without affecting Eq. (2.16) (Oliveira 2000). Therefore, without loss of generality, the main diagonal of the variance-covariance matrix $\boldsymbol{\Sigma}$ for the GRF is set to 1, or in the CAR model $\tau^2 = 1$.

Chapter 3

Spatially Correlated Model Selection

3.1 Introduction

The time to present in detail the foundations of the Spatially Correlated Model Selection (SCOMS) method has come. In Chapter 2 we mentioned the Kim et al.'s (2005) method to analyze non-stationary spatial fields, and we also gave a thorough description of the Smith and Fahrmeir's (2007) method because it introduces the Ising distribution in variable selections. In this chapter we will see that the combination of those methods with the Kuo and Mallick's (1998), will result in the SCOMS method. The SCOMS method is a proposed alternative for variable selection with spatially correlated data whose spatial field is non-stationary. The method searches for mutually exclusive, dependent, stationary and isotropic regions within a non-stationary spatial field; and at each region, it selects the set of significant covariables that explain certain variability on the response variable, taking into account the dependence between regions.

With spatial data, special care should be taken to the structure of the correlation that might best explain the spatial dependence of the data. From an optimistic stand point (although this optimism may fade away because of its intrinsic complexity), spatially correlated data has advantageously more information to exploit for inferences than independent data, in the following sense. Consider a single realization of the spatial process, $Y(\mathbf{s}_l)$. The random variable $Y(\mathbf{s}_l)$ possesses not only information of the process at location $\boldsymbol{s}_l$, where the observation is taken, but also information about its neighboring locations. Whereas a realization of an independent process, regardless its location, contains information only about

its underlying generating process, but never about its neighborhood. Thus, the strength of statistical analyses for independent data relies mainly on the number observations (sample size), whilst the strength of the correlated data analyses relies on the manner in which correlation between realizations of the process is modeled.

Modeling the spatial correlation in non-stationary spatial fields is problematic for spatial data analyses. Typically, in order to simplify their analysis, the spatial fields are assumed stationary and isotropic. However, experience says that in spatial applications, the spatial processes are rarely stationary and isotropic (Fuentes and Smith 2001). Therefore, some important information can be ignored and lost when analyzing data under the stationarity assumption, when its spatial field is actually non-stationary. To provide a more general approach to model the spatial correlation, we propose the SCOMS method to properly analyze data from spatial fields that are neither stationary nor isotropic. The way it works will be explained in the forthcoming sections. For the time being, it is enough to mention that this method models two sources of correlation, one is the spatial correlation of the data within regions and the other is the correlation between regions.

The SCOMS is a simple and novel method to analyze non-stationary spatial fields. Its simplicity and novelty reside in that it combines methodologies already available for applications, such as the Kim et al.'s (2005) method and the Kuo and Mallick's (1998) method, in order to answer sometimes important scientific questions. For instance, suppose that one wants to partition a spatial field of interest into regions where the spatial field is stationary; thus, questions such as:

- Into how many regions should a spatial field of interest be partitioned?
- Where should the boundaries of the regions be located?
- Is there a subset of covariables that are significant in explaining the response variable of interest at the region level?
- How do the covariables in a region depend on those that are in neighboring regions?

can be answered with the SCOMS method.

The regions, as parts of a spatial field, are correlated with each other. The correlation between regions is modeled through the significance of the covariables in such a way that the transitions between regions are smooth if their correlation is high or rough if they are

uncorrelated. For instance, covariables such as elevation, wind speed, slope, anthropogenic activities, soil type, etc., change across the field, and hence their effect on the response variable change as well. Hence, the main task of the SCOMS method is to find regions where the effect of the covariables remains constant. And, if a covariable effect is highly significant in one region, the SCOMS methods invites the covariable to be also significant in regions adjacent to it. In summary, by including the correlation between regions the SCOMS method invites significant covariables in one region, to be significant on adjacent regions. This is only possible thanks to the Ising distribution that models the probability of inclusion in the variable selection. On the other hand, if the spatial field is assumed stationary (i.e. a single region), local information is likely to be lost. For example, mining is one of the human activities that sometimes contaminates water with chemicals, such as acid-generating sulfides and toxic heavy metals, that are toxic for aquatic life in high concentrations, especially for trout (Zhang et al. 2008). Mining is localized at some areas within the brook trout data spatial field. Then, the covariables that quantify the concentration of acid-generating sulfides and toxic heavy metals, will probably not be significant to explain the response variable in large scale analyses; however, their significance will likely be found and explained if the spatial field is partitioned into regions.

In summary, this chapter presents the SCOMS method. This is a variable selection method with spatial data whose spatial field is non-stationary and non-isotropic. The method accomplishes two complementary and fundamental tasks: (1) The identification of locally stationary and dependent regions within a spatial field of interest, similarly as it is done in the Kim et al. (2005). And at each region, (2) variable selection is performed with a slightly modified version of the Kuo and Mallick's (1998) method to accommodate the Ising distribution, in order to model the regional dependence in the models. In the following sections, the details of the method are presented.

3.2 Spatial Field partition

Because the SCOMS method is constructed for piece-wise stationary spatial fields, we need to find the pieces, also called regions, where the field is stationary. Therefore, we need to partition the field into regions. This section describes the way the partition of the field is achieved.

Partitions of a convex spatial field can be obtained by means of the Voronoi tessellation, a tessellating technique that has been used for spatial applications. Okabe et al. (2000) give a comprehensive overview of the role of the Voronoi tessellation in spatial analyses. The Voroin tessellation technique partitions a spatial field into a given number of mutually exclusive regions, as illustrated in Fig. 3.1. In the figure, a square spatial field is tessellated into $R = 10$ non-overlapping regions. The computational efficiency of this technique has made it popular for spatial statistics applications, when the spatial fields are convex. Unfortunately, it does not work with non-convex fields (where probably other more complicated and maybe less computationally efficient methods are required). In order to avoid extra complications in the development of the method, we assume that all the spatial fields in this chapter are convex.

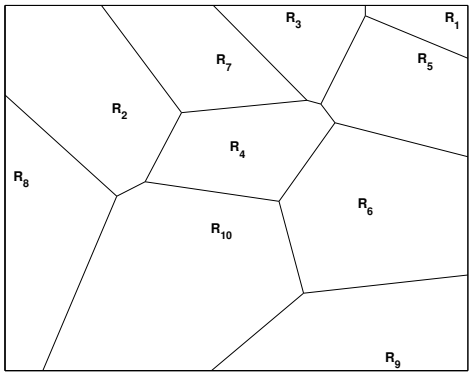


Figure 3.1: A Voronoi tessellation of a spatial random field into 10 non-overlapping regions. The polygons of the regions are defined only by their centroids, $\mathfrak{c}(R)$.

A spatial field in $\mathcal{D} \in \mathbb{R}^2$ is Voronoi tessellated into R mutually exclusive regions by solely giving the polygon that encloses the spatial field and the coordinates of the centroids of the R regions. Hence, let us suppose that there is a polygon that delimits a given spatial field. Let $\mathfrak{c}(R) = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_R)$ be the centroids of the R regions that tessellate the spatial field. The set of regions $\{1, 2, \ldots, R\}$ uniquely determined by $\mathfrak{c}(R)$, defines non-overlapping polygons that encompass sites of observations. The sites are assigned to the regions with respect to the (Euclidian) distance from the centroids of the regions, using the following criterion:

$$\boldsymbol{s}_l \in i\text{th region if } \min_{i'=1,\ldots,R}(\|\boldsymbol{s}_l - \boldsymbol{c}_{i'}\|) = \|\boldsymbol{s}_l - \boldsymbol{c}_i\|, \ \forall\, l = 1, \ldots, n \text{ and } i \in \{1, 2, \ldots, R\}. \quad (3.1)$$

In the SCOMS method, the sites contained in one region are assumed spatially correlated. Therefore, a spatial covariance structure, and also a mean function, of the form $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\mu}_i$ per each region is assumed. By allowing different spatial covariance structures and different means across the spatial field (one pair per region), the spatial field is assumed non-stationary. In what follows, we present the models and the probability distributions assumed in the SCOMS method, for non-stationary spatial fields.

3.3 Local Gaussian distribution and the likelihood function

For the sake of clarity in the description of the model and its distributional assumptions, we introduce some notation. Let $\{Y(\mathbf{s}_l), \boldsymbol{s}_l \in \mathcal{D}\}_{l=1}^{n}$ be a realization of a Gaussian random field observed at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$. Also, let $\boldsymbol{X}_{n\times q}$ be the matrix of q covariables.

Suppose a spatial field is partitioned into a fixed number of regions, R. Let $Y(\boldsymbol{s}_l)_i$ and $\boldsymbol{X}_{i,l}$ be an observation of the spatial field and the q-vector of covariables corresponding to the lth site, with coordinates $\boldsymbol{s}_l$, in region i. Then,

$$\boldsymbol{Y}_i = (Y(\boldsymbol{s}_1)_i, \ldots, Y(\boldsymbol{s}_{n_i})_i)'$$

and

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}'_{i,1} \\ \vdots, \\ \boldsymbol{X}'_{i,n_i} \end{pmatrix}$$

are the observation vector and the design matrix corresponding to region i. Here n_i is the number of sites that region i encompasses, and $i = 1, 2, \ldots, R$, and the total number of sites is $n = \sum_{i=1}^{R} n_i$.

Conditionally independent regions

The model that explains the relationship between the response vector and the set of covariables is as follows. Let us consider a partition of the spatial field of interest. Thus, given the partition, the regions are assumed to be conditional independent of one another, i.e. the

spatial process in one region is independent of the spatial process in other regions. However, the information at the sites within a region is assumed spatially correlated, i.e. elements of the vector $\boldsymbol{Y}_i$ are spatially correlated. Hence, the relationship between $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ can be explained by the following linear model

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{e}_i, \tag{3.2}$$

where $\boldsymbol{\beta}_i$ is the vector of regression coefficients, and $\boldsymbol{e}_i \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$. Here, $\boldsymbol{\Sigma}_i$ is the covariance matrix that models the spatial correlation of the data at region i. Recall that the covariance matrix for the ith region is approximated by $\boldsymbol{\Sigma}_i \approx \tau_i^2(\boldsymbol{D}_{w_i} - \rho_i\boldsymbol{W}_i)^{-1}$, where $\boldsymbol{D}_{w_i}$ and $\boldsymbol{W}_i$ are deterministic, and their definition is explained in Section 2.3.5.

The sampling distribution of the response vector $\boldsymbol{Y}_i$ given the model in Eq. (3.2), is as follows. Let us suppose that there is an observed realization of the process $\boldsymbol{Y}_i = \boldsymbol{y}_i$ at region i, then

$$f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \boldsymbol{X}_i, \tau_i^2, \rho_i) = \left(\frac{1}{2\pi}\right)^{\frac{n_i}{2}} |\boldsymbol{\Sigma}_i|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i)\right\} \tag{3.3}$$

Therefore, the likelihood function for $\boldsymbol{B}_i = (\boldsymbol{\beta}_i, \tau_i^2, \rho_i)$ is

$$L(\boldsymbol{B}_i|\boldsymbol{y}_i, \boldsymbol{X}_i) \propto f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \boldsymbol{X}_i, \tau_i^2, \rho_i). \tag{3.4}$$

The expressions above are supposed to be valid for all the regions in the partition.

Joint distribution for regions

As a consequence of the conditionally independence, as well as the models and the density functions given in Eqs. (3.2) and (3.3), we can specify the density of the vector of realizations of the spatial process, $\boldsymbol{Y} = [\boldsymbol{Y}_i', \ldots, \boldsymbol{Y}_R']'$, as products of the individual marginals. Hence, the joint distribution of $\boldsymbol{Y} = [\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_R]$ given the partition is simply

$$f_{\boldsymbol{Y}}(\boldsymbol{y}|\mathbb{c}(R), \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{\tau}, \boldsymbol{\rho}) \propto \prod_{i=1}^{R} f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i|\boldsymbol{\beta}_i, \boldsymbol{X}_i, \tau_i^2, \rho_i), \tag{3.5}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_R)$, $\boldsymbol{\tau} = (\tau_1^2, \ldots, \tau_R^2)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_R)$. As a result, the likelihood for $\boldsymbol{B} = (\boldsymbol{B}_1, \ldots, \boldsymbol{B}_R)$ is given by

$$L(\boldsymbol{B}, \mathfrak{c}(R)|\boldsymbol{y}, \boldsymbol{X}) \propto f_{\boldsymbol{Y}}(\boldsymbol{y}|\mathfrak{c}(R), \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{\tau}, \boldsymbol{\rho}). \tag{3.6}$$

The SCOMS method assumes a fairly simple model when the response variable is normally distributed and the model given in Eq. (3.2) is capable of explaining the relationship between the dependent and the independent variables. Nevertheless, there are situations where the response variable is binary. In such situations, we can still apply the SCOMS method for the analysis the binary response, but after some suitable adaptations, as explained in the following section.

3.3.1 A generalization of the model

An extension of our model above to a case where the underlying distribution is different to the Gaussian distribution, occurs when the spatial random field is binary. Binary Random Fields occur often in environmental and ecological research. We say that $\{Z(\mathbf{s}_l), \boldsymbol{s}_l \in \mathbb{R}^2\}_{l=1}^n$ is a realization of a binary random process if each $Z(\mathbf{s}_l) \in \{0, 1\}$. For example, $Z(\boldsymbol{s}_l)$ may represent the presence or absence of an instance of interest (Haran 2010), or it may indicate that the concentration of a contaminating substance is above or below a threshold, or that the abundance of a species is low or high. Another example is given by the brook trout dataset, where at each site the response variable indicates the presence or absence of trout.

In general, binary responses occur either as the result of a nominal variable with two levels or through categorizing a continuous variable as a two levels nominal variable (Heagerty and Lele 1998). In modeling a Binary random process, it is sometimes convenient to assume that there is an underlying Gaussian random process governing the binary outcomes. As presented in Section 2.4, the connection between the Binary field and the Gaussian field is uniquely determined if the variance of the Gaussian process is fixed to a constant value, usually equal to 1 (Albert and Chib 1993; Oliveira 2000). This practice is equivalent to the well known probit regression for random processes (Oliveira 2000). The introduction of the latent Gaussian process with mean and variance given in Eq. (3.2) results in a new joint distribution for $\boldsymbol{Y}$ conditional on $\boldsymbol{Z}$, as follows.

For regions $\{1, 2, \ldots, R\} \in \mathcal{D}$ with centroids $\mathfrak{c}(R) = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_R)$, the joint distribution

for the Gaussian process $\boldsymbol{Y} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_R')'$, where $\boldsymbol{Y}_i = (Y_{i,1}, Y_{i,2}, \ldots, Y_{i,n_i})'$, given $\boldsymbol{Z} = (\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_R')$, where $\boldsymbol{Z}_i = (Z_{i,1}, \ldots, Z_{i,n_i})'$ is

$$f_{\boldsymbol{Y}}(\boldsymbol{y}|\mathbb{C}(R), \boldsymbol{Z} = \boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{X}, \boldsymbol{\tau}, \boldsymbol{\rho}) \propto \prod_{i=1}^{R} \{\boldsymbol{I}_{(\boldsymbol{y}_i \geq \boldsymbol{0})} \boldsymbol{I}_{(\boldsymbol{z}_i = \boldsymbol{1})} + \boldsymbol{I}_{(\boldsymbol{y}_i < \boldsymbol{0})} \boldsymbol{I}_{(\boldsymbol{z}_i = \boldsymbol{0})}\} (1/2\pi)^{n_i/2} \times \ldots$$
$$|\boldsymbol{\Sigma}_i|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}_i)\right\}, \quad (3.7)$$

where each element of $\boldsymbol{I}_{(\cdot)}$, $I_{(A)}$ is an indicator function for the event A.

Notice that by transforming the BRF into a GRF, the model given in Eq. (3.2) is preserved, and the BRF only represents an extra level at the top of a hierarchical model specification, where the mapping into the Gaussian random process takes place. It does not represent theoretical complications, but unfortunately it does make the model fitting process computationally more demanding.

The correlation between adjacent regions is taken into account in the model fitting thanks to the Ising distribution embedded in the variable selection method of Kuo and Mallick (1998). The next section presents the Kuo and Mallick's (1998) method and the associated Ising distribution.

3.4 The locally isotropic variable selection method for spatial data

In the SCOMS method, the selection of the set of significant covariables for the response variable is performed using a slightly modified version of the Kuo and Mallick's (1998) method, as follows. First of all, in the SCOMS method we have multiple regions, therefore, the generalization of the Kuo and Mallick's (1998) method to multiple regions leads us to re-state Eq. (2.5) as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\vartheta}_i + \boldsymbol{\epsilon}_i, \text{ for } i = 1, \ldots, R. \quad (3.8)$$

Here the *effect size* parameter is given by $\boldsymbol{\vartheta}_i = [\beta_{i,1}\gamma_{i,1}, \ldots, \beta_{i,q}\gamma_{i,q}]'$; $\boldsymbol{Y}_i$, $\boldsymbol{X}_i$, $\boldsymbol{\beta}_i$ and $\boldsymbol{\epsilon}_i$ are the same as those given in Eq. (3.2).

A *priori* we assume that the indicator variables, $\gamma_{i,j}$, within a region are independent of one another, i.e. for region i, $\gamma_{i,j}$ and $\gamma_{i,j'}$ are assumed *a priori* independent. However, any

two indicator variables related to one covariable but from two regions are dependent, i.e., consider $\gamma_{i,j}$ at region i and $\gamma_{i',j}$ at region i', where $i' \in \mathbb{N}_i$ (where $\mathbb{N}_i$ is the set of regions that are neighbors of region i, and $i \neq i'$). We assume that $\gamma_{i,j}$ and $\gamma_{i',j}$ are correlated. In general, let $\boldsymbol{\gamma}_j = [\gamma_{1,j}, \ldots, \gamma_{R,j}]'$ be the indicator vector made up of indicator variables associated to the jth covariate across regions. We assume the Ising distribution as prior for $\boldsymbol{\gamma}_j$ as follows

$$\pi(\boldsymbol{\gamma}_j|\theta, \mathfrak{C}(R)) = \frac{1}{\Omega(\theta,\omega)_j} \exp\left\{\sum_{i\sim i'} \theta\ \omega_{i,i'}\ I(\gamma_{i,j} = \gamma_{i',j})\right\}, \tag{3.9}$$

where θ is the interaction parameter (i.e. the parameter that determines the overall degree of dependence among the regions) and ω is a metric describing the regions relationship. The constant of integration, $\Omega(\theta,\omega)_j$ does not have a close form. Smith and Fahrmeir (2007) expressed the constant of integration of the Ising distribution as

$$\Omega(\theta,\omega)_j = \sum_{\boldsymbol{\gamma}_j} \exp\left\{\sum_{i\sim i'} \theta\ \omega^*_{i,i'}\ I(\gamma_{i,j} = \gamma_{i',j})\right\}.$$

Given a partition of the spatial field, the parameter ω is fixed and it is modeled as a function of the physical characteristics of the regions as we explain later. The correlation between regions that the Ising distribution accounts for is based directly on adjacent regions. An illustration of Ising distribution is given in the following example.

Example of the Ising distribution

To understand the Ising distribution, Fig. 3.2 presents a simple spatial field partitioned into four regions. For non-symmetric ω, the term within the curly brackets in Eq. (3.9) is expanded as follows,

$$\begin{aligned}\left\{\theta\ \sum_{i\sim i'} \omega_{i,i'}\ I(\gamma_{i,j} = \gamma_{i',j})\right\} = \theta\ \big[&(\omega_{1,2} + \omega_{2,1})\ I(\gamma_{1,j} = \gamma_{2,j}) + \ (\omega_{1,3} + \omega_{3,1})\ I(\gamma_{1,j} = \gamma_{3,j}) + \\ &(\omega_{2,3} + \omega_{3,2})\ I(\gamma_{2,j} = \gamma_{3,j}) + \ (\omega_{2,4} + \omega_{4,2})\ I(\gamma_{2,j} = \gamma_{4,j}) + \\ &(\omega_{3,4} + \omega_{4,3})\ I(\gamma_{3,j} = \gamma_{4,j})\big].\end{aligned}$$

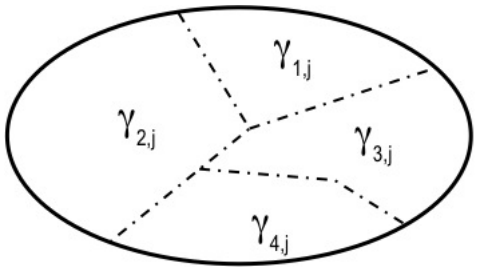


Figure 3.2: An arbitrary spatial field partitioned into four regions. For illustration, we consider the latent indicator variable of the jth covariate.

As shown in the example above, the Ising distribution naturally includes the dependence that a given region has on its adjacent regions. For instance, let us consider regions 1 and 2, if $\gamma_{1,j}$ and $\gamma_{2,j}$ were both equal to either 0 or 1, the Ising distribution would increase the probability of the latent variable to be equal to 0 or 1, by the amount $\theta(\omega_{1,2} + \omega_{2,1})$. On the other hand, if the same two γs were not equal (i.e. one equal to 1 and the other equal to 0), the Ising distribution would not increase or decrease the probability, but would revert to the case where adjacent regions are independent. Therefore, the Ising distribution is a general model for the indicator variables in $\boldsymbol{\gamma}$. A particular case of Eq. (3.9) is apparent when $\theta = 0$ and it also reduces to the independent regions, as in Kim et al.'s (2005) method.

Given the above, the specification of the model is completed. The models at the regions, the likelihoods, and the accommodation of the Ising distribution into the variable selection method, are the most important components of the SCOMS method. The forthcoming sections address the prior specifications for the unknown parameters introduced in the model formulation, as well as the algorithms that will allow one to make inferences.

3.5 The analysis

The linear model given in Eq. (3.8) together with the Ising distribution given in Eq. (3.9) depend on a set of unknown parameters. To learn about them, we rely on Bayesian statistics. In this section, the prior distributions for the unknown parameters are specified. Although, the number of regions, R, and θ in the Ising distribution are unknown, we propose to base their selection on the strategy outlined in Eq. (3.14). For each parameter, whose prior distribution is specified in this section, there is a full conditional posterior distribution that can be found in Appendix A.

3.5.1 Prior specifications

The joint prior distribution for the parameters in the models in Eqs. (3.8) and (3.9) is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\rho}, \mathfrak{c}(R)) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma}|\theta)\pi(\boldsymbol{\tau})\pi(\boldsymbol{\rho})\pi(\mathfrak{c}(R)). \tag{3.10}$$

We assume independent prior distributions for all the unknown parameters, except for θ and for R, as can be seen in the right hand side of Eq. (3.10). In the following, we specify each $\pi(\cdot)$ presented above. Notice that in Section 3.5.3 we present the algorithm that searches for the centroids of the regions, therefore, the prior distribution for the centroids, $\pi(\mathfrak{c}(R))$, is specified in that section as well.

Prior specification for the matrix of the regression coefficients

We start by giving the prior distribution for $\boldsymbol{\beta}$, the matrix of regression coefficients. This parameter contains all the covariables and all regions, as follows.

$$\pi(\boldsymbol{\beta}) = \prod_{i=1}^{R} N_q(\mathbf{0}, \sigma^2_{\boldsymbol{\beta}_i} I_{q\times q}), \tag{3.11}$$

where $I_{q\times q}$ is the identity matrix, and σ^2_β is the variance of the prior distribution. For this hyperparameter, Kuo and Mallick (1998) recommend to set it equal to any value in the range $(0.25, 16)$, so we set $\sigma^2_\beta = 10$.

Prior specification for the matrix of the indicator variables

The joint prior distribution for $\boldsymbol{\gamma}$, the matrix of indicator variables, is

$$\pi(\boldsymbol{\gamma}|\theta) = \prod_{j=1}^{q} \pi(\boldsymbol{\gamma}_j|\theta).$$

Where $\pi(\boldsymbol{\gamma}_j|\theta)$ is given in Eq. (3.9).

An example of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$

To understand the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters, we give the following example. Let us assume that there are $R = 3$ regions and $q = 4$ regressors, hence the elements of each parameter are,

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{1,1} & \beta_{2,1} & \beta_{3,1} \\ \beta_{1,2} & \beta_{2,2} & \beta_{3,2} \\ \beta_{1,3} & \beta_{2,3} & \beta_{3,3} \\ \beta_{1,4} & \beta_{2,4} & \beta_{3,4} \end{bmatrix} \quad \text{and } \boldsymbol{\gamma} = \begin{bmatrix} \gamma_{1,1} & \gamma_{2,1} & \gamma_{3,1} \\ \gamma_{1,2} & \gamma_{2,2} & \gamma_{3,2} \\ \gamma_{1,3} & \gamma_{2,3} & \gamma_{3,3} \\ \gamma_{1,4} & \gamma_{2,4} & \gamma_{3,4} \end{bmatrix}.$$

It is clear from this illustration that for every component in $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ has a corresponding indicator variable. Each column of these two matrices corresponds to a region, and each row corresponds to a covariable.

Prior specification for the variance parameter of the spatial field

The prior distribution for $\boldsymbol{\tau}$ is specified under two different situations that depend on the type of the response variable at hand.

1. When the response of interest is given by $\boldsymbol{Y}$ as in Model Eq. (3.2), $\pi(\boldsymbol{\tau}) \propto \prod_{i=1}^{R} \pi(\tau_i)$, where

$$\pi(\tau_i^2) \propto 1/\tau_i^2, \tag{3.12}$$

   a non-informative prior is a reasonable choice.

2. When the response of interest is binary (see Section 3.3.1), we follow the sampling scheme 1 proposed by Imai and van Dyk (2005), where

$$\tau_i \sim \text{Gamma}(\alpha^*, \beta^*), \text{ for } i = 1, \ldots, R, \tag{3.13}$$

   with $\alpha^* = \beta^* = 1$.

Prior specification for the spatial correlation parameters

Finally, the prior distribution for $\boldsymbol{\rho}$ is specified. Banerjee et al. (2004) simulated data with different values of this parameter and found that, in their words, "a prior for ρ that encour-

ages a consequential amount of spatial association would place most of its mass near $\rho = 1$". Carlin and Banerjee (2002) assumed a Beta as prior distribution for this parameter. Based on these findings on the CAR models, we adopt the following prior for $\boldsymbol{\rho}$,

$$\pi(\boldsymbol{\rho}) \propto \prod_{i=1}^{R} \text{Beta}(\rho_i; a_\rho, b_\rho), \tag{3.14}$$

With respect to the hyper-parameter, Carlin and Banerjee mentioned that some authors assume $a_\rho = 18$ and $b_\rho = 2$ to encourage spatial correlation in CAR models applications, however, such values have been controversial. Nevertheless, we adopt them for the hyper-parameters in Eq. (3.14) for lack of better ones.

Learning about θ and the number of regions

To find out the optimum number of regions to partition the spatial field into and to learn about the parameter θ in the Ising distribution, we propose the following strategy, which becomes especially important for practical applications of the SCOMS method.

1. Firstly, we propose a grid of feasible values for the number of regions, for example $R \in \{1, 2, 3, \ldots\}$, and set $\theta = 0.5$.

2. Set $R = r$ and fit the model considering this number of regions.

3. After convergence, we can summarize the model fit by computing one of the likelihood based criteria, AIC, BIC, DIC, or the *marginal density* of the data $f(\boldsymbol{y}|R = r, \theta = 0.5)$, or by cross-validation (as in the case study Section 4.4).

4. Repeat steps 2 and 3, for all values of $R \in \{1, 2, 3, \ldots\}$.

5. The number of regions is selected as the one that optimizes the chosen model fit evaluation criterion, and set it constant for further analysis.

6. Turning to θ, we proceed similarly as in step 1. A grid of feasible values for this parameter is proposed and evaluated using the same criterion selected in step 3. The grid for θ can be, for example $\theta \in \{0.5, 1.0, \ldots, 3.5, 4.0\}$.

7. Select θ that optimizes the chosen criterion.

Once R and θ are selected, they are fixed at these values, and the analysis proceeds.

Later, this rather general strategy will be particularized according to the pursued evaluation. For instance, in Chapter 4 we specify the statistic that summarizes the simulation study, and in Section 4.4 we revisite this strategy and accommodate cross-validation evaluation in it, and define the statistic that summarizes the cross-validation results.

3.5.2 Metrics for ω

The parameter $\omega_{i,i'}$ in Eq. (3.9), accounts for the level of interdependence between regions i and i'. We base this parameter on physical characteristics of the regions, hence, two different ways are proposed to compute ω.

1. *ω as a function of the length of the shared edge between two adjacent regions.* Let L_i be the sum of all lengths that region i shares with all the regions in $\mathbb{N}_i$, where $\mathbb{N}_i$ is the set of neighbors of region i, and let $l_{i,i'}$ be the length shared with region i' ($i' \in \mathbb{N}_i$), then

$$\omega_{i,i'} = \frac{l_{i,i'}}{L_i} \tag{3.15}$$

   An example of ω as a function of the shared edge, is presented in Fig. 3.3(a). In the figure, there are four different regions within the spatial field. The Mixed region has Forest and Agriculture as neighbors. The edge shared by Mixed and Forest region (in red) is $l_{1,2}$ (where the index 1 is for Mixed and the index 2 is for Forest). And $L_1 = l_{1,2}+l_{1,3}$ (where the index 3 is for Agriculture). Hence $\omega_{1,2}$ follows from Eq. (3.15).

2. *ω as a function of the area of adjacent regions.* Suppose that the area of the polygon defined by region i is a_i, $\forall\ i = 1, \ldots, R$. Then, let $A_i = \sum_{i'=1}^{|\mathbb{N}_i|} a_{i'} + a_i$, be the total area covered by region R_i and its adjacent regions. Then

$$\omega_{i,i'} = \frac{a_{i'}}{A_i}, \tag{3.16}$$

   for $i' \in \mathbb{N}_i$.

   Fig. 3.3(b) shows an example of ω when it is a function of the area of adjacent regions. For Mixed, $A_1 = a_1 + a_2 + a_3$ (where the indexes 1 is for Mixed, 2 is for Forest, and 3

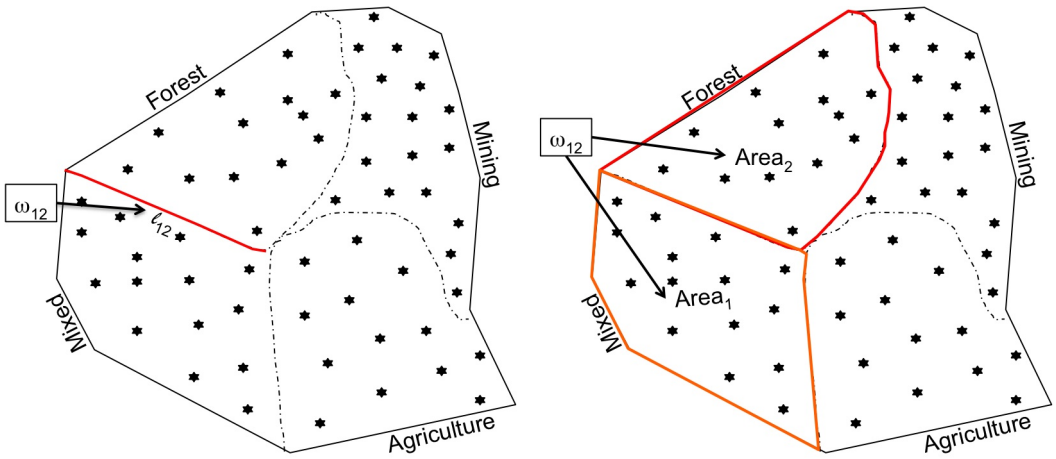


Figure 3.3: Illustration of the metric that measures of the relationship between regions. Regions 1 and 2 (Mixed and Forest), is presented. In panel (a), $\omega_{1,2}$ as a function of the length of the edge shared by these two regions (in red). In panel (b), $\omega_{1,2}$ as a function of the area of the polygons defined by these two regions.

is for Agriculture.). Then, $\omega_{1,2}$ between Mixed (in orange) and Forest (in red) follows from Eq. (3.16).

3.5.3 Algorithm for the search for the optimum partition of the spatial field

The Voronoi tessellation procedure provides a simple way to partition the spatial field. Its computational simplicity makes it attractive for Bayesian spatial analyses (Okabe et al. 2000). The polygons of the regions constructed by the Voronoi technique have linear boundaries; this sometimes can be restrictive. Additionally, the Voronoi tessellation works when the polygon that encloses the spatial field is convex. For non-convex cases, the spatial field tessellation becomes computationally complex. Nevertheless, in this research we use Voronoi tessellation to partition the spatial field, for its simplicity.

The search for the right partition of the spatial field, which is one of the goals of this methodology, is performed iteratively via the Metropolis-Hastings algorithm (Hastings 1970). At each iteration, given the centroids of the R regions, $\mathfrak{c}(R) \in \mathcal{D}$, the Voronoi tessellation

partitions the spatial field into R non-overlapping regions, see for example Fig. 3.1. Once the regions are determined, their physical characteristics are the inputs for the computation of ω and for the identification of the neighborhood of the regions.

In order to the search for the right partition of the spatial field into R regions, we proceed as follows

1. For iteration $t = 1$, the Markov chain of the centriods starts by randomly selecting the R centroids, $\mathfrak{c}(R) \in \mathcal{D}$. We call the starting centroids $\mathfrak{c}^1$. Given $\mathfrak{c}^1$ a set of regions are defined via Voronoi tessellation.

2. At a given iteration, $t > 1$, and taking one region at the time, a new centroid is proposed using the Gaussian distribution as the proposal distribution. For instance, we propose a new centroid, $\boldsymbol{c}_i^*$, for region i as $\boldsymbol{c}_i^* \sim N_2(\boldsymbol{c}_i^t, \psi_i^2 I)$, where ψ_i^2 is the variance of the proposal distribution, and I is an identity matrix of order 2. Then, $\boldsymbol{c}_i^{t+1} = \boldsymbol{c}_i^*$ with probability

$$\alpha_i = \min\left\{\frac{L(\boldsymbol{B}, (\boldsymbol{c}_1^t, \ldots, \boldsymbol{c}_{i-1}^t, \boldsymbol{c}_i^*, \boldsymbol{c}_{i+1}^t, \ldots, \boldsymbol{c}_R^t), \cdot) \times \pi(\boldsymbol{c}_i^*)}{L(\boldsymbol{B}, (\boldsymbol{c}_1^t, \ldots, \boldsymbol{c}_{i-1}^t, \boldsymbol{c}_i^t, \boldsymbol{c}_{i+1}^t, \ldots, \boldsymbol{c}_R^t), \cdot) \times \pi(\boldsymbol{c}_i^t)}, 1\right\}, \tag{3.17}$$

   for $i = 1, \ldots, R$. $L(\cdot)$ is the likelihood function, given in Eq. (3.6) or Eq. (3.7), depending on the type of the response variable.

The prior distribution for the ith centroid, $\boldsymbol{c}_i$, is

$$\pi(\boldsymbol{c}_i) = \text{Uniform}(\mathbb{P}),\ \forall\ i = 1, \ldots, R, \tag{3.18}$$

where $\boldsymbol{c}_i = [\text{longitude}_i, \text{latitude}_i]$, $\mathbb{P}$ is the polygon that encloses the spatial field of interest. Therefore, $\pi(\boldsymbol{c}_i) \geq 0$ if $\boldsymbol{c}_i \in \mathbb{P}$.

In the next chapter we evaluate the method and also apply it to a case study. A complete description on how the data are simulated is also provided.

Chapter 4

Evaluation of the method by simulation

The evaluation of the *Spatially Correlated Model Selection* (SCOMS) method is presented in this chapter. This evaluation involves two sequential steps: (1) The simulation of the datasets and (2) the application of the SCOMS methods to the simulated datasets. In order to accomplish the first step, some parameters are required. We start by defining the spatial field where the data are simulated, and its partition into regions. The spatial field and the physical characteristics of its regions are inputs to compute the metric ω in the Ising distribution. The parameter $\theta > 0$ is specified as well. These two parameters and the partition of the spatial field allow the simulation of the parameter $\boldsymbol{\gamma}$, the matrix of indicator variables. The parameter $\boldsymbol{\gamma}$ defines the models in each region, and includes the correlation between regions. In Section 4.1.2 we present the steps followed to simulate this parameter to guarantee that its realizations indeed include the correlation between the adjacent regions.

The datasets are simulated under three different conditions, that as of now will be called *experiments*. These *experiments* are illustrated in Fig. 4.1. The three experimental conditions result from letting ω in Eq. (3.9) to be function of

Experiment 1. the shared edge, $\omega(L)$,

Experiment 2. the area, $\omega(A)$, and,

Experiment 3. $\omega = 0$, independent regions.

When $\omega = 0$ the adjacent regions are independent of one another (the case studied by Kim

et al. (2005)). Notice that the case of independent regions is also achieved when θ in Eq. (3.9) is set to zero.

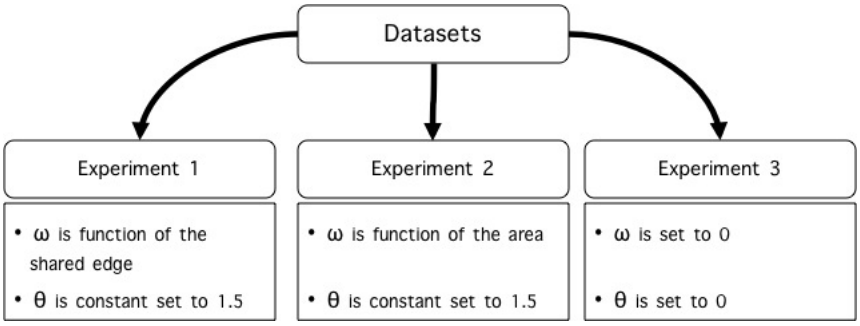


Figure 4.1: The three experimental conditions under the datasets are simulated.

The simulation study is carried out by simulating datasets where the response variable is given by the random vector $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\vartheta}$ and $\boldsymbol{\Sigma}$ are the mean function and the variance-covariance of the Gaussian process $\boldsymbol{Y}$. The parameter $\boldsymbol{\vartheta} = \boldsymbol{\beta} \circ \boldsymbol{\gamma}$ is defined as the *effect size* (Kuo and Mallick 1998), which is the Hadamard product between parameters $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_i$. Each dataset is simulated considering the following steps, where each is explained later in this chapter.

- Generation of a grid of locations within a given spatial field.
- Partition of the field into R regions, and the computation of ω metrics.
- Simulation of the covariates matrix, $\boldsymbol{X}$.
- Simulation of the matrix of indicator variables, $\boldsymbol{\gamma}$.
- Simulation of the matrix of regression coefficients, $\boldsymbol{\beta}$.
- Simulation of the vector of spatial correlation, $\boldsymbol{\rho}$.
- Computation of $\boldsymbol{\Sigma}_i$ as in the CAR model given in Eq. (2.15), for $i = 1, \ldots, R$.
- Simulation of the response vector, $\boldsymbol{Y}_i$, with mean $\boldsymbol{X}_i\boldsymbol{\vartheta}_i$ (Eq. (2.5)) and covariance matrix $\boldsymbol{\Sigma}_i$.

Step (2) of the SCOMS method's evaluation starts once the datasets are simulated. This part of the study consists of the application of the SCOMS method to the simulated data. In

each of the *experimental* conditions, three different analyses, also called *cases*, are performed on the datasets. They depend on how the metric ω is selected when the SCOMS method is applied. *Case* 1 occurs when the analysis is performed assuming $\omega(L)$, *case* 2 comes up assuming $\omega(A)$, and *case* 3 happens when $\omega = 0$. These cases are thoroughly explain in Table 4.1.

The following section describes in detail how the synthetic datasets are obtained. Once the data are simulated, the application of the SCOMS method to the analysis of these data is explained in Section 4.2. The statistics utilized to summarize the simulation results, and the evaluation of the performance of the SCOMS methods when applied to different cases are presented in Section 4.2.1. At the end of this chapter, in Section 4.4, the case studies and their results are presented.

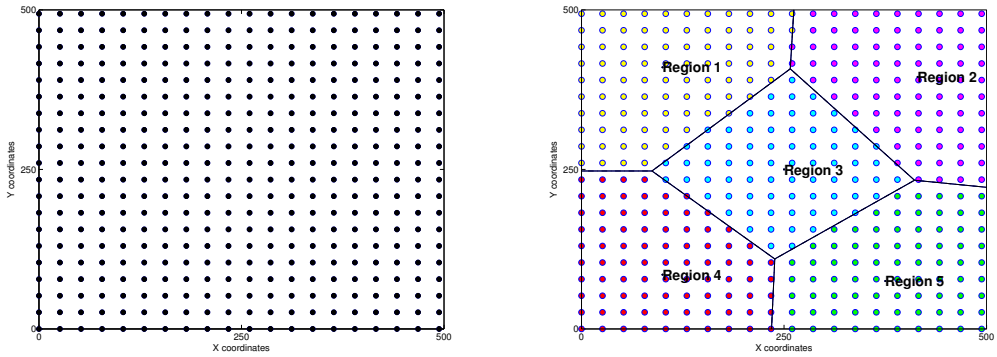
4.1 Data and correlated model simulation

In this section, we start with the definition of the spatial field where the data are going to be simulated, and the identification of the sites where supposedly the information is recorded. As a next step, we generate the matrix of covariates, $\boldsymbol{X}$, the matrix of indicator variables, $\boldsymbol{\gamma}$, the matrix of regression coefficients, $\boldsymbol{\beta}$, the spatial correlation parameter, $\boldsymbol{\rho}$, for the CAR model, etc.

4.1.1 Definition of the spatial field and its partition

A rectangular spatial field is assumed, with vertices at (0,0), (0,500), (500,500) and (500,0) of the Cartesian plane. A grid of 400 sites is defined, each site with coordinate $\boldsymbol{s}_l = (x_l, y_l)$, for $l = 1, \ldots, 400$. Fig. 4.2(a) shows the simulated spatial field and the 400 sites evenly distributed across the field.

A partition of the field is considered. The partition has $R = 5$ mutually exclusive regions. The assignation of the $n = 400$ sites to the regions is performed by the application of the simple rule explained in Section 3.2. Notice that the partition of the spatial field and the grid of sites do not change in the simulation study. Fig. 4.2(b) displays the spatial field and its partition, where colored sites are presented. Sites or dots with the same color belongs to the one region.



(a) (b)

Figure 4.2: The simulated spatial field with the sites of observations and its 5 regions. In panel (a) the simulated spatial field with the grid of the 400 sites. In panel (b) the partition of the simulated spatial field into $R = 5$ regions. The sites are colored to help in the identification of the region they belong to.

The following sections describe the simulation of the design matrix and other parameters, whose values do change over the simulations.

4.1.2 Design matrix and *effect size* matrix simulation

The design matrix, $\boldsymbol{X}$, is simulated by drawing independent samples from a normal distribution with mean 0 and standard deviation 1. Each column in $\boldsymbol{X}$ is a random vector of order 400 from a $N(0, 1)$. A total of $q = 10$ vectors are simulated that are assumed to be the regressors. Hence, $\boldsymbol{X}$ is matrix of covariates of order 400×10.

The simulation of $\boldsymbol{\gamma}$ is carried out in a way that realizations of this parameter incorporates the regions dependence. For this, we need the polygon of the spatial field and its partition as inputs to simulate $\boldsymbol{\gamma}$. The followings steps are required to guarantee that realizations of $\boldsymbol{\gamma}$ indeed embeds the correlation of the regions.

Step 1. Given the spatial field and its partition (see Fig. 4.2(b)), we compute ω such that $\omega_{i,i'} > 0$ if region $i' \in \mathbb{N}_i$, and $\omega_{i,i'} = 0$ if region $i' \notin \mathbb{N}_i$, where $\mathbb{N}_i$ is the set of adjacent regions of region i. The parameter ω is a function of physical characteristics of the regions, i.e. it is a function of the shared edge ($\omega(L)$) or of the area ($\omega(A)$) of adjacent regions. The computation of ω is explained in Section 3.5.2.

Step 2. The parameter that measures the overall interaction of the regions in the Ising distribution is fixed to a given value, usually $\theta = 1.5$. Although it changes depending on to the purpose of the dataset, as seen below.

Step 3. The matrix of indicator variables, $\boldsymbol{\gamma}$, is simulated according to Eq. (3.9). Every element of $\boldsymbol{\gamma}$ is simulated with its conditional distribution, $\pi(\gamma_{i,j}|\boldsymbol{\gamma}_{-(i,j)})$, given in Eq. (A.10) in Appendix A, with $l^* = 0$. The simulation starts assuming that $\pi(\gamma_{i,j} = 1|\boldsymbol{\gamma}_{-(i,j)}) = 0.5,\ \forall\ (i,j)$. The indicator variable, $\gamma_{i,j}$, is defined by a Bernoulli draw with success probability equal to 0.5, for $i = 1, \ldots, R$ and $j = 1, \ldots, q$. This is the initial value for $\boldsymbol{\gamma}$ matrix, called $\boldsymbol{\gamma}^0$.

Step 4. Starting with $\boldsymbol{\gamma}^0$, the conditional probability $\pi(\gamma_{i,j}|\boldsymbol{\gamma}_{-(i,j)})$ is updated, and this step and Step Item 3 are repeated for 20,000 times in order to approach to an stationary state for $\boldsymbol{\gamma}$.

Step 5. Finally, $\boldsymbol{\gamma}$ is chosen to be equal to the next simulated value after the stationary state has been reached, in other words $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{20001}$. Or equivalently, in order to highlight that this parameter is obtained for a given ω, we can set $\boldsymbol{\gamma}(\omega) = \boldsymbol{\gamma}^{20001}$.

Step 6. The regression coefficients matrix, $\boldsymbol{\beta}$, is simulated by drawing random numbers from the normal distribution, conditional on $\boldsymbol{\gamma}$, such that if $\gamma_{i,j} = 1$ then $\beta_{i,j} \sim N(0, 10)$, and $\beta_{i,j} = 0$ otherwise, for $i = 1, \ldots, R,$ and $j = 1, \ldots, q$.

Step 7. The parameter $\boldsymbol{\rho}$ is simulated from its prior given in Eq. (3.14).

With $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{\rho}$ and $\boldsymbol{\gamma}(\omega)$ simulated, we can simulate the response variable $\boldsymbol{Y}$ using Eq. (3.8). But before that, the variance-covariance matrix, $\boldsymbol{\Sigma}$, should be generated. The following explains the simulation of these quantities: $\boldsymbol{\Sigma}$ and $\boldsymbol{Y}$.

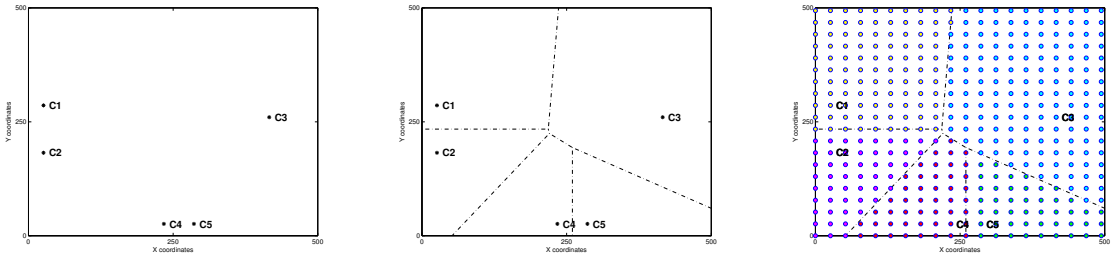
4.1.3 Continuous response simulation

The parameter $\boldsymbol{\Sigma}$, is the covariance matrix of the CAR model. As we have seen before, it is $\boldsymbol{\Sigma}_i = \tau^2(\boldsymbol{D}_w^i - \rho_i \boldsymbol{W}_i)^{-1}$, with $\boldsymbol{D}$ and $\boldsymbol{W}$ computed as explained in Section 2.3.5. To facilitate the simulation study, and without loss of generality, we set $\tau = 1$ in this expression. The simulation of $\boldsymbol{\Sigma}$ is straightforward, once the deterministic components, $\boldsymbol{D}$ and $\boldsymbol{W}$, are available.

With $\boldsymbol{X}_i$ and $\boldsymbol{\vartheta}_i$ $(= \boldsymbol{\beta}_i \circ \boldsymbol{\gamma}_i)$[a] for region i $(i = 1, \ldots, R)$, $\boldsymbol{Y}$ is generated by sampling from the multivariate normal distribution, with mean $[(\boldsymbol{X}_1\boldsymbol{\vartheta}_1)', \ldots, (\boldsymbol{X}_R\boldsymbol{\vartheta}_R)']'$ and covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_R)$.

4.2 Model evaluation scheme

The SCOMS method evaluation starts with the random selection of $R = 5$ centroids within the spatial field. With these centroids, the spatial field is tessellated into five regions, using Voronoi tessellation. Then, the assignation of the sites to the regions follows, using the rule explained in Section 3.2. This process is illustrated in Fig. 4.3. The randomly selected centroids are the black stars in Fig. 4.3(a). With the selected centroids, the five mutually exclusive regions are created by the Voronoi tessellation (an illustration of the regions is presented in Fig. 4.3(b)). The allocation of the sites that belong to each region is shown in Fig. 4.3(c).



(a) (b) (c)

Figure 4.3: Example of the process followed to identify the regions in the spatial field. Panel (a) illustrates the selection of the R centroids, five centroids are randomly selected. Panel (b) shows tessellated field with respect to the selected centroids, this creates the regions. Panel (c) shows the result of grouping the sites into the regions.

Once the spatial field is tessellated, a neighborhood or adjacency matrix of the regions is created. From our illustration in Fig. 4.3, the adjacency matrix is shown in Table 4.1. For adjacent regions, the matrix assigns 1 and 0 otherwise. This matrix makes easier the identification of the adjacency set, $\mathbb{N}$, of each region. For example, the adjacency set for region 1 is $\mathbb{N}_1 = \{\text{Region2}, \text{Region3}\}$ (which corresponds to the first row of the adjacency

[a]The Hadamard product

matrix), set that agrees of what is shown in Fig. 4.3(b). Moreover, areas and perimeters of each region are computed and used as inputs to compute ω. Both, the adjacency matrix and ω are inputs to compute $\pi(\boldsymbol{\gamma}_j)$ with the Ising distribution (Eq. (3.9)), as explained in Step 3 of Section 4.1.2.

Table 4.1: The adjacency matrix, also referred to as the neighborhood system. For adjacent regions, the adjacency is 1, and 0 otherwise.

	Region 1	Region 2	Region 3	Region 4	Region 5
Region 1	0	1	1	0	0
Region 2	1	0	1	1	0
Region 3	1	1	0	1	1
Region 4	0	1	1	0	1
Region 5	0	0	1	1	0

The initial partition of the spatial field (Fig. 4.3) marks the starting state of the chain for the partitions or equivalently the starting state of the chain of the centroids. The search for the best partition is performed by the Monte Carlo Markov Chain (MCMC) algorithm, as explained in Section 3.5.3. The updates for the partition is performed element-wise, where a centroid to be updated is selected at random. For the selected centroid, a new centroid is proposed from the proposal distribution. The new centroid is evaluated, if accepted, then the new centroid is kept, but if rejected the old centroid is kept. The proposal distribution is normal. For instance, the proposed centroid for the ith region is $\boldsymbol{c}_i^{new} \sim N_2(\boldsymbol{c}_i^{old}, \psi^2\, I)$ (where $\boldsymbol{c}_i = (\text{longitude}_i, \text{latitude}_i)$), where ψ is the standard deviation of the proposal distribution. The element-wise updating of the centroids is found to favor fast convergence of the Markov chains, compared with the block updating. At the end, we have a chain of centroids of the form

$$\mathbb{c}^0, \mathbb{c}^1, \mathbb{c}^2, \ldots, \mathbb{c}^{T-1}, \mathbb{c}^T,$$

where T is the number of iterations.

After all the centroids are updated, the adjacency matrix and ω are re-computed. And the matrix of indicator variables, $\boldsymbol{\gamma}$, is updated by sampling from its full conditional $\pi(\boldsymbol{\gamma}|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\rho})$ (this full conditional is given after $\boldsymbol{\beta}$ is integrated out), so is $\boldsymbol{\beta}$ by sampling from its full conditional $\pi(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\gamma}, \boldsymbol{\rho})$. Recall that the full conditionals are given in Appendix A. The parameter $\boldsymbol{\rho}$ is also updated through Metropolis-Hastings steps, since its full conditional distribution is an unknown distribution. After several iterations, we end with a chain for $\boldsymbol{\gamma}$,

$\boldsymbol{\rho}$ and $\boldsymbol{\beta}$ similar to

$$\{\boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\rho}^0\}, \{\boldsymbol{\gamma}^1, \boldsymbol{\beta}^1, \boldsymbol{\rho}^1\}, \{\boldsymbol{\gamma}^2, \boldsymbol{\beta}^2, \boldsymbol{\rho}^2\}, \ldots, \{\boldsymbol{\gamma}^T, \boldsymbol{\beta}^T, \boldsymbol{\rho}^T\},$$

that combined with the centroids chain leads to

$$\{\mathbb{c}^0, \boldsymbol{\gamma}^0, \boldsymbol{\beta}^0, \boldsymbol{\rho}^0\}, \{\mathbb{c}^1, \boldsymbol{\gamma}^1, \boldsymbol{\beta}^1, \boldsymbol{\rho}^1\}, \{\mathbb{c}^2, \boldsymbol{\gamma}^2, \boldsymbol{\beta}^2, \boldsymbol{\rho}^2\}, \ldots, \{\mathbb{c}^T, \boldsymbol{\gamma}^T, \boldsymbol{\beta}^T, \boldsymbol{\rho}^T\}.$$

Experiments

The simulation study is done under the different experimental conditions mentioned above and graphically illustrated in Fig. 4.1. For all the simulated datasets, the number of regions is the same, i.e. $R = 5$, as mentioned in Section 4.1. The data simulation and inference on the simulated data create the following experimental situations: We define an *experiment* to be the conditions under which the data are simulated, and a *case* to describe the conditions under which the inference is performed on the simulated data.

Experiment 1. In this experiment, the datasets are simulated under correlated models. We base the metric, ω, on the shared edge between neighboring regions, $\omega(L)$, and set $\theta = 1.5$. (See Fig. 4.4 for a graphical illustration of the inference performed on the datasets simulated under *experiment 1.*)

- *Case* 1. Inference is carried out based on ω assumed to be a function of the shared length, $\omega(L)$, between adjacent regions, and $\theta = 1.5$.
- *Case* 2. Inference is carried out based on ω assumed to be a function of the area, $\omega(A)$, between adjacent regions, and $\theta = 1.5$.
- *Case* 3. Inference is carried out under the independent models, where $\theta = 0$ in Eq. (3.9).

Experiment 2. In this other experiment, the datasets are simulated under correlated models as well, but with ω based upon the area of adjacent regions, $\omega(A)$, and $\theta = 1.5$. (See Fig. 4.5 for a graphical illustration of the inference performed on the datasets simulated under *experiment 2.*)

- *Case* 1. Inference is carried out based on ω as a function of the shared edge, $\omega(L)$, and $\theta = 1.5$.

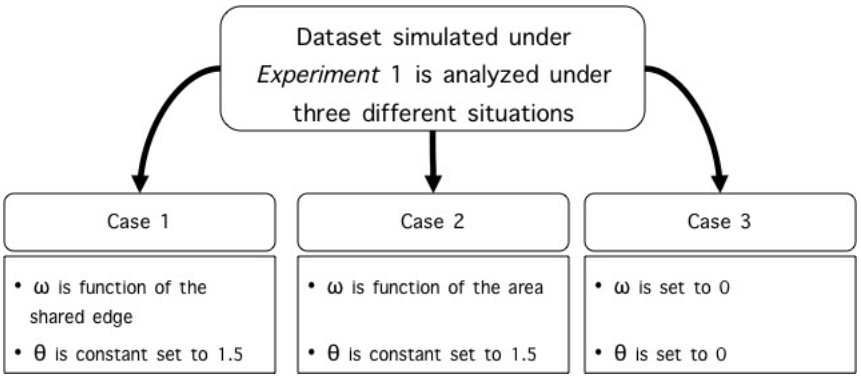


Figure 4.4: The analysis of the dataset simulated under *experiment 1* is carried out under three different conditions, also known as *Cases*.

Case 2. Inference is carried out based on ω as a function of the area, $\omega(A)$, and $\theta = 1.5$.

Case 3. Inference is carried out based on independent models, $\omega = 0$.

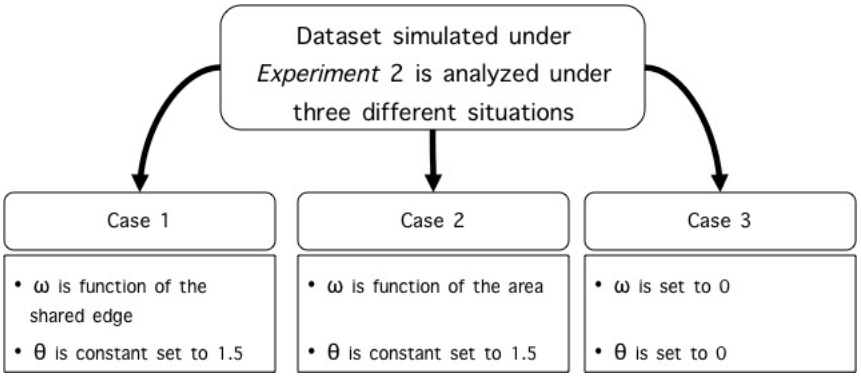


Figure 4.5: The analysis of the dataset simulated under *experiment 2* is carried out under three different conditions or *cases*.

Experiment 3. In this last experiment, the datasets are simulated under independent models, i.e. $\theta = 0$. (See Fig. 4.6 for a graphical illustration of the inference performed on the datasets simulated under *experiment 3*.)

Case 1. Inference is carried out based on ω as function of the shared edge, $\omega(L)$, with $\theta = 0.5$.

Case 2. Inference is carried out based on ω as function of the shared edge, $\omega(L)$, with $\theta = 1.5$.

Case 3. Inference is carried out with independent models, where $\theta = 0.5$.

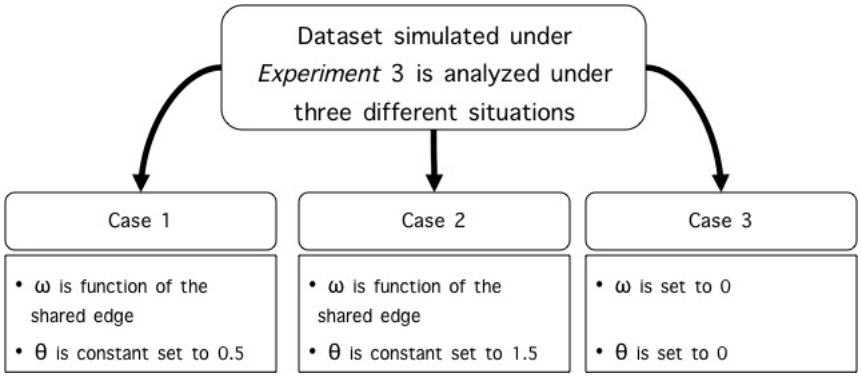


Figure 4.6: The analysis of the dataset simulated under *experiment 3* is carried out under three different conditions or *cases*.

Since in practice it is difficult to know what creates the dependence between regions, we consider *experiment 1 case 2*, and *experiment 2 case 1*, to evaluate how sensitive the SCOMS method is to the choice of the metric of dependence, ω. *Case 1* in *experiment 1* and *case 2* in *experiment 2* are considered to validate the algorithm and to compare results of the other *cases* in the *experiments*. *Case* 3 in both *experiments*, 1 and 2, is included to evaluate the performance of the SCOMS method in situations where inference is carried out assuming independence between regions, when truly the datasets were generated with dependent regions.

Experiment 3 is included to evaluate the performance of the SCOMS method in situations where the datasets are simulated under independent regions. Both *cases*, 1 and 2, evaluate the performance of the method with mild ($\theta = 0.5$) and strong ($\theta = 1.5$) correlations between adjacent regions. And, similarly as in the other experiments, *case 3* is presented to compare the results obtained from the other two cases, in *experiment* 3.

4.2.1 Goodness of fit and Models comparison

The model that relates the response variable with the set of covariables depends on the vector of indicator variables ($\boldsymbol{\gamma}_i$) and the regression coefficients ($\boldsymbol{\beta}_i$) through the *effect size* parameter, $\boldsymbol{\vartheta}_i = \boldsymbol{\gamma}_i \circ \boldsymbol{\beta}_i$, as follows

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\vartheta}_i + \boldsymbol{e}_i.$$

The parameter $\boldsymbol{\gamma}_i$ determines the model and also brings the correlation between regions into the model. Observations are simulated from this model if the parameters $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_i$ are at hand. Hence, one sensible way to evaluate the performance of the SCOMS method in the simulation study is by comparing the $\boldsymbol{\gamma}$ that simulates the data, given the model above, with its samples from its stationary distribution obtained after convergence of the Markov chains. Another way is to evaluate the SCOMS goodness of fit, through the marginal density of the data. The two SCOMS performance evaluations are presented in this section.

AFCCF

To evaluate the performance of the SCOMS method in the selection of the simulated model, say $\boldsymbol{\gamma}^{(real)}$, we compare the samples of this parameter after convergence of its chain, namely $\boldsymbol{\gamma}_{AC}$, with $\boldsymbol{\gamma}^{(real)}$. The statistic that summarizes the comparison is the *Average Fraction Correctly Classified for Fit*, also known as AFCCF (Wilkinson 1999; Zhang et al. 2008), computed in the following way.

With N samples of $\boldsymbol{\gamma}$ after convergence, say $\{\boldsymbol{\gamma}_{AC}^t : t = 1 \ldots, N\}$ (AC stands for After Convergence), it is always possible to estimate the posterior probability of $\gamma_{i,j} = 1$, $\hat{p}_{i,j}$, as the arithmetic mean of the $\boldsymbol{\gamma}_{AC}$ samples, as follows

$$\hat{p}_{i,j} = \frac{\sum_{t=1}^{N} I_{(\gamma_{AC,i,j}^t = 1)}}{N},$$

where $I_{(\cdot)}$ is an indicator variable. Then, the $AFCCF_\gamma$ is given by

$$AFCCF_\gamma = \frac{\sum_{i=1}^{R}\sum_{j=1}^{q}\left\{\gamma_{i,j}^{(real)}\hat{p}_{i,j} + (1-\gamma_{i,j}^{(real)})(1-\hat{p}_{i,j})\right\}}{R\ q}, \tag{4.1}$$

where $\gamma_{i,j}^{(real)} \in \{0,1\}$ is the (i,j)th component of $\boldsymbol{\gamma}^{(real)}$. The $AFCCF_\gamma \in [0,1]$, with values closer to one as desirable in good model fits.

Marginal Density

A statistic utilized to judge model fits is the *marginal density* of the data, namely $f(\boldsymbol{y}|M)$. To estimate the marginal density, we compute the harmonic means of the likelihood, evaluated

at samples of the parameters after convergence. The estimation procedure was proposed by Newton and Raftery (1994), and has the following representation,

$$f(\boldsymbol{y}|M) = \frac{N}{\sum_{t=1}^{N} \frac{1}{L^*(\boldsymbol{\gamma}^t, \boldsymbol{\rho}^t|\boldsymbol{Y}, \boldsymbol{X})}}, \tag{4.2}$$

where M is the model under a particular condition, and $L^*(\cdot)$ is the likelihood given in Eq. (A.7) in Appendix A, $\boldsymbol{\gamma}^t$ and $\boldsymbol{\rho}^t$ are samples from the posterior distribution obtained through Metropolis-Hastings steps.

The marginal density of the data turns into a particularly important tool in the simulation study that helps to identify the optimum number of regions and the paremeter θ. Later we will see that this statistic accurately selects the correct number of regions, however, it is partially reliable when used as a tool to select the optimum value for θ in the Ising distribution.

What follows are the results of the model fit study when the SCOMS method is utilized for inference on the datasets simulated under the three *experimental* conditions presented above.

4.3 Simulation results

The simulation results under the different experimental conditions and cases listed above, are based on chains of $T = 40,000$ iterations. The first 30,000 are considered burn-in, and the last 10,000 are taken as samples from the stationary distributions of the parameters involved. Therefore, the number of samples after convergence that are considered to compute the $AFCCF_\gamma$ or *marginal density* of the data is $N = 10,000$.

The correlation between neighboring models is paramount in this research. Therefore, in the summarization of the simulation results we focus mainly in the $\boldsymbol{\gamma}$ matrix, which is the parameter that allows the accommodation of the correlation between models of adjacent regions and summarizes the importance of covariables. We compute the $AFCCF_\gamma$ to summarize the results of each *case* in the *experiments*. Note that in Appendix B, we are including some tables with summarizations of the simulation results, that are not explicitly addressed in this section.

As an empirical result obtained from the simulation study below, we found that the harmonic mean of the likelihood, used to estimate the *marginal density* of the data, results in not completely reliable estimates. Weak conclusive results were obtained when it was utilized in the selection of the optimum value for the parameter θ in the Ising distribution.

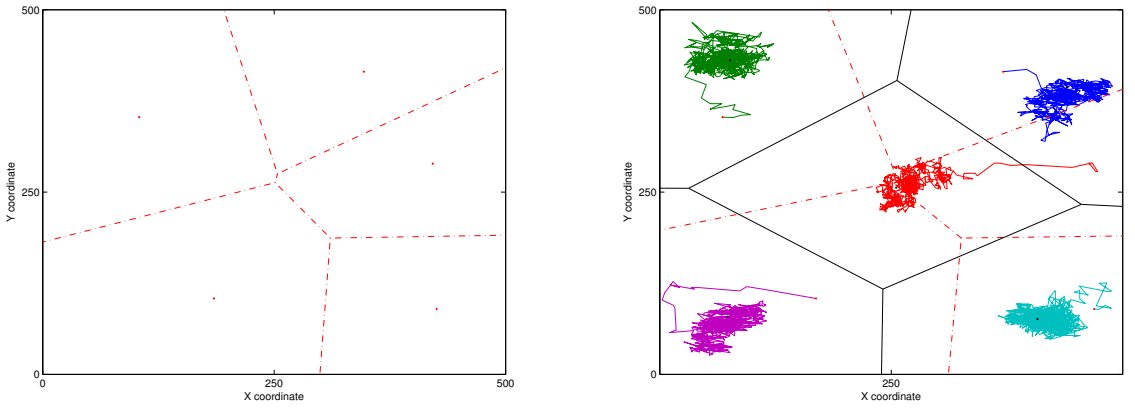
4.3.1 Results from simulated data

Simulation study: *Experiment 1*

One hundred datasets were simulated under experimental condition one. Each dataset was simulated independently of one another. The SCOMS method was applied to those one hundred datasets. As a first step in the inferential analysis, we verified convergence of the Markov chains. As examples, Figs. 4.7 and 4.8 display the initial partition of the spatial field and the trace plots of the centroids, obtained from the analysis of a dataset under *cases* 1 and 2. For *experiment 1 case* 1, Fig. 4.7(a) shows the starting centroids of the regions and their corresponding partition of the field, while in Fig. 4.7(b) the trace plots of the centroids for the 40,000 MCMC iterations is shown (the trace plot plots every 50^{th} sampled centroids). The solid lines in Fig. 4.7(b) represent the partition corresponding to the last sampled centroids. It is obvious how similar the partition of the field shown in Fig. 4.7(b) is to the one that was used to simulate the datasets, shown in Fig. 4.2(b); this indicates that the centroids chains have converged. Similar plots are shown for *case* 2 in Figs. 4.8(a) and 4.8(b). Analogous plots have also been observed from the analysis of the datasets under *case 3* (not shown), where convergence of the sampled centroids to the "real" centroids of the regions have been verified.

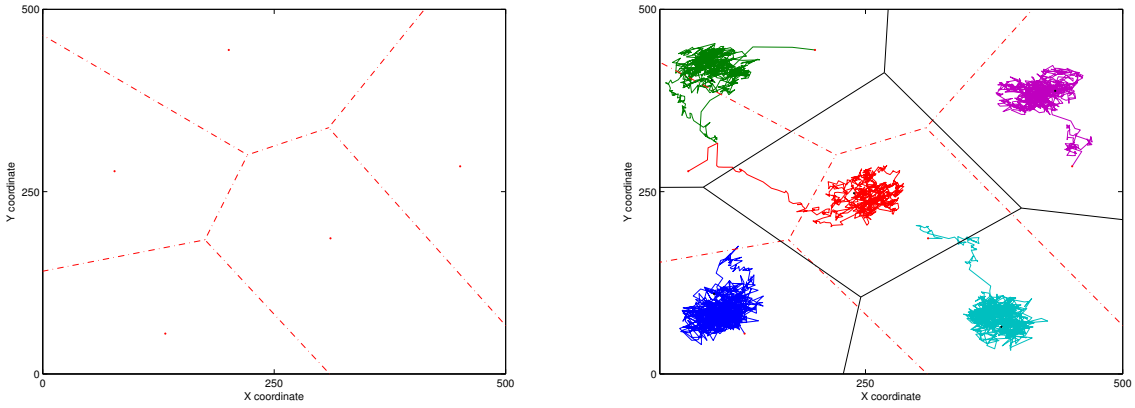
The analysis of the one hundred datasets is summarized by computing the $AFCCF_\gamma$ as in Eq. (4.1), for each of the *cases* in *experiment* 1. Fig. 4.9 shows the box-plots of the $AFCCF_\gamma$s.

In Fig. 4.9, we see that on average *cases* 1 and 2 choose the "real" model (i.e. the $\boldsymbol{\gamma}^{(real)}$) around 95% of the times, while *case* 3 chooses the real model about 91% of the times. Thus, the boxplots reveal that *case 3* is giving less desirable results, as expected. The poor performance of *case* 3 is because it does the inference assuming independent adjacent regions, when in reality they are not. To assess how different the *cases* are of one another, a pairwise comparison is performed on the $AFCCF_\gamma$ means of each *case* in *experiment 1.* The results



(a) (b)

Figure 4.7: Starting centroids and their trace plot under *case 1* in *experiment 1*. Panel (a) shows the initial centroids and their corresponding partition of the field. Panel (b) shows the final partition of the field after 40,000 MCMC iterations (solid black), and the trace plots for the 5 centroids (every 50^{th} iteration is plotted).



(a) (b)

Figure 4.8: Starting centroids and their trace plot under *case 2* in *experiment 1*. Panel (a) shows the initial centroids and their corresponding partition of the field. Panel (b) shows the final partition of the field after 40,000 iterations (solid black), and the trace plot for the 5 centroids (every 50^{th} iteration is plotted).

of the comparisons are shown in Table 4.2.

The boxplots of *cases* 1 and 2 (Fig. 4.9) look similar to each other, indicating that their $AFCCF_\gamma$s are not different, and this is validated with the result shown in the first row

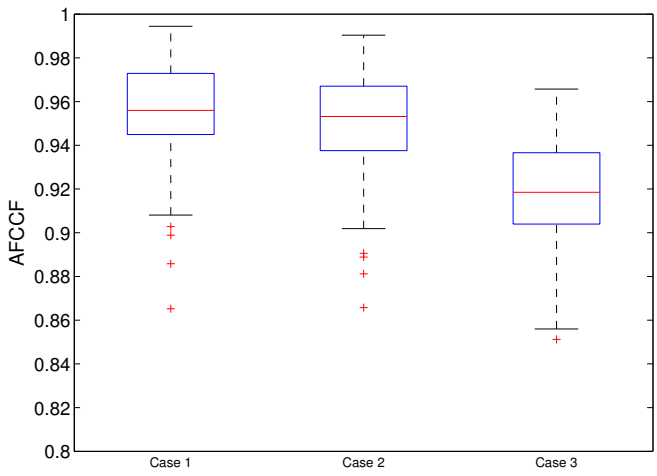


Figure 4.9: The boxplots of the $AFCCF_\gamma$ for *experiment 1 cases* 1, 2 and 3.

Table 4.2: Summary results of the pairwise comparison of the $AFCCF_\gamma$ means from the simulation study in *experiment 1*. The means of the $AFCCF_\gamma$ of the *cases* are compared with one another and a 95% confidence interval is computed. The last column, Reference, identifies which means result statistically different.

$AFCCF_\gamma$ **means**	$AFCCF_\gamma$ **means difference**			
comparison of	L.Limit	Means diff.	U. Limit	Ref.
Case 1 vs Case 2	-0.0034	0.0045	0.0125	
Case 1 vs Case 3	0.0289	0.0369	0.0448	*
Case 2 vs Case 3	0.0243	0.0323	0.0403	*

of Table 4.2 where the $AFCCF_\gamma$ means of the two cases are compared. The difference of these two means is 0.0045, and a 95% confidence interval on the difference includes zero. Therefore, we can say that: even though the ωs in both *cases* are not based on the same physical characteristics of the regions, *cases* 1 and 2 lead to similar inferential conclusions. To some extent, it seems that as long as the correlation between adjacent regions is included in the analysis, when the models are correlated, the SCOMS method is estimating the right models more often, regardless how the metric ω is modeled.

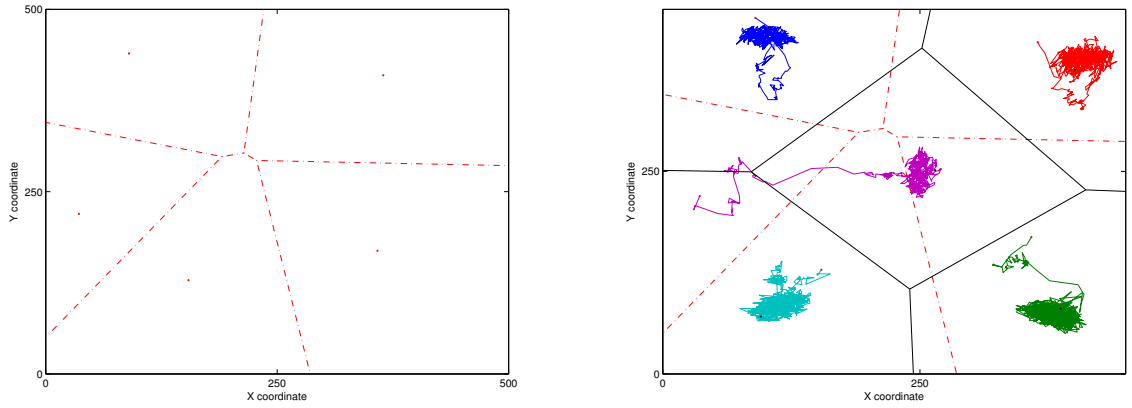
The results of the comparisons between *case* 3 with *cases* 1 and 2 are shown in the second and third rows of Table 4.2. In both comparisons, their corresponding 95% confidence intervals on the $AFCCF_\gamma$s mean differences do not include zero. This proves what we have

said before: *case* 3 is given the less desirable results and its performance is statistically different from *cases* 1 and 2.

Additionally, we computed the empirical coverage rates for the effect size $\boldsymbol{\vartheta}$ parameter, even though they are not of primary concern in this research. The empirical coverage rates are presented in Table B.1 in Appendix B. In that table, we can see that the coverage rates are close to their nominal levels, which is 95%.

Simulation study: *Experiment 2*

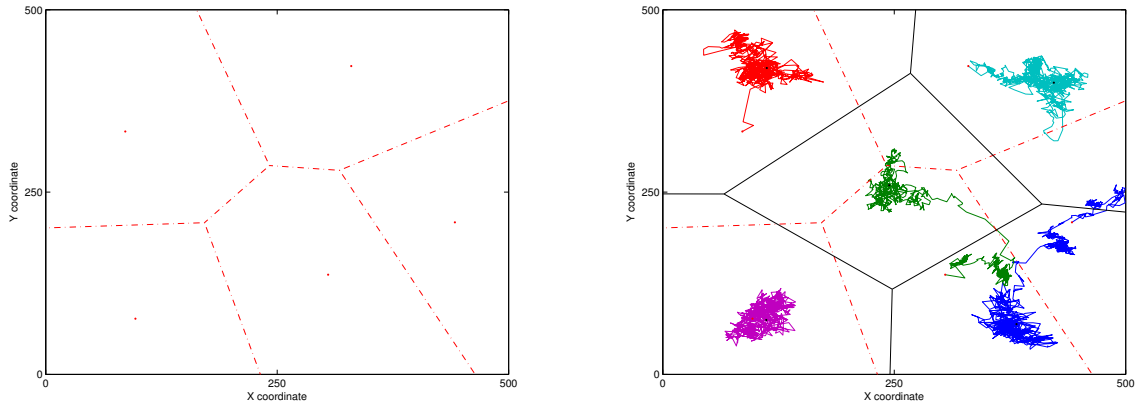
One hundred datasets were simulated under experimental condition two, and analyzed using the SCOMS method. Similarly as in *experiment 1*, we start the summarization of the simulation results by verifying convergence of the chains of the centroids. Examples of the convergence are presented in Figs. 4.10 and 4.11. After a burn-in period, the MCMC samples of the centroids come from where the "real" centroids are located (see Fig. 4.2(b)). We verified all the trace plots, in all the *cases*, to guarantee fair comparisons between *cases* in *experiment 2*.



(a) (b)

Figure 4.10: Starting centroids and their trace plot under *case 1* in *experiment 2*. Panel (a) shows the initial centroids and their corresponding partition of the field. Panel (b) shows the final partition of the field after 40,000 MCMC iterations (solid black line), and the trace plot for the 5 centroids (every 50^{th} iteration is plotted).

The analysis of the one hundred datasets is summarized by the computation of the



(a) (b)

Figure 4.11: Starting centroids and their trace plot under *case 2* in *experiment 2*. Panel (a) shows the initial centroids and their corresponding partition of the field. Panel (b) shows the final partition of the field after 40,000 MCMC iterations (solid black line), and the trace plot for the 5 centroids (every 50^{th} iteration is plotted).

$AFCCF_\gamma$ on the $\boldsymbol{\gamma}$ parameter. Fig. 4.12 presents the box-plots of the $AFCCF_\gamma$ for each *case* in *experiment 2*. It is clear that *case 3*, which assumes independence between models, gives the less desirable results, while *cases* 1 and 2 are both giving similar $AFCCF_\gamma$ results. Between the first two *cases*, their boxplots looks similar to each other, they both are centered at a common value roughly equal to 0.94. Thus, their $AFCCF_\gamma$s can be assumed similar. On the other hand, the boxplot of *case* 3 looks different from those of *cases* 1 and 2, just as in *experiment* 1. This boxplot is centered around 0.92, off by roughly 0.02 from the other two cases (we will see below that this is a significant difference).

To support our last statement, we compared the mean of the $AFCCF_\gamma$s of each *case* with any other, and the results are shown in Table 4.3. This table shows that *case 3* is indeed different from *cases* 1 and 2 (the second and third rows), but it also says that *cases* 1 and 2 are not different of each other (the first row in the table).

Simulation study: *Experiment 3*

The analysis of the one hundred datasets generated under *experiment 3*, is summarized by the $AFCCF_\gamma$ for the parameter $\boldsymbol{\gamma}$, in each *case*. As a common practice, before summarizing

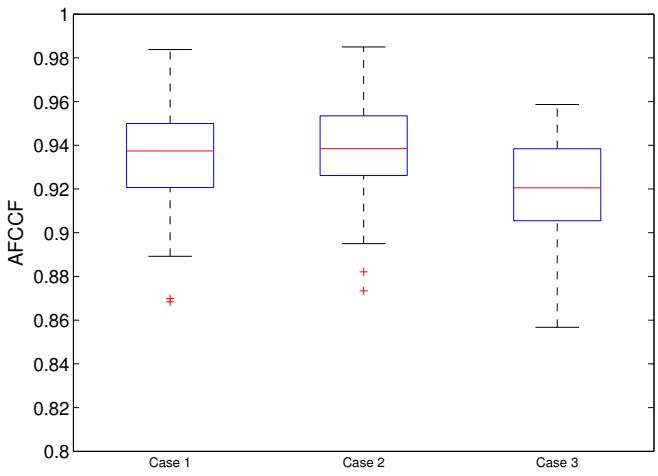


Figure 4.12: The boxplots of the $AFCCF_\gamma$ *experiment 2 cases* 1, 2 and 3.

Table 4.3: Summary results of the pairwise comparison of the $AFCCF_\gamma$ means from the simulation study in *experiment 2*. The means of the $AFCCF_\gamma$ of the *cases* are compared with one another. The last column, Reference, identifies which means result statistically different.

$AFCCF_\gamma$ **means**	$AFCCF_\gamma$ **means difference**			
comparison of	L.Limit	Means diff.	U. Limit	Ref.
Case 1 vs Case 2	-0.0108	-0.0034	0.0039	
Case 1 vs Case 3	0.0100	0.0173	0.0247	*
Case 2 vs Case 3	0.0134	0.0208	0.0281	*

the simulation results, we verified convergence of the chains by looking at the trace plots of each unknown in the model (not shown) to guarantee fair comparisons.

Fig. 4.13 presents the box-plots of the $AFCCF_\gamma$s computed for each of the *cases* in *experiment 3*. The *case 1* and *case 3* boxplots are very similar of each other, and both are centered at 0.92. Thus, as a preliminary conclusion based on the boxplots, we found that *case 1* and *case 3* are equally precise on selecting the real model ($\boldsymbol{\gamma}^{(real)}$). Recall that *case 3* does inference based on the independent model assumption (that matches the way the datasets are simulated, in this experiment), while *cases* 1 and 2 do the inference under the assumption that the models are correlated, with the strength of the correlation given by $\theta = 0.5$ and $\theta = 1.5$, respectively. This result empirically proves that for datasets with

uncorrelated models, our method is leading to a similar conclusion as the one obtained when the inference is done under the independent model assumption, provided θ is set to a relative small value, such as $\theta = 0.5$.

When $\theta = 1.5$ in *experiment* 3, the *case 2* boxplot (Fig. 4.13) is clearly located below of the other two boxplots. This indicates that *case* 2 gives the less desirable results. In this case, the SCOMS method is forcing the models of the regions to be strongly dependent of one another, but the data do not support it.

Table 4.4 presents the pairwise comparison of the $AFCCF_\gamma$ means of the *cases*. In agreement with the preliminary conclusion above, the $AFCCF_\gamma$ mean corresponding to *case 3* is not statistically different from the one corresponding to *case 1* (comparison shown in the second row of the table), but the $AFCCF_\gamma$ means of *cases* 1 and 3 are statistically different from the mean of the $AFCCF_\gamma$ corresponding to *case 2* (comparisons shown in the first and third rows of the table).

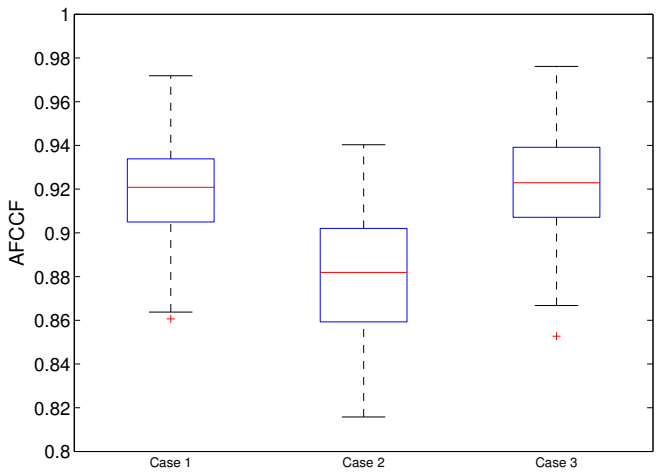


Figure 4.13: The boxplots of the $AFCCF_\gamma$ for *experiment 3 cases* 1, 2 and 3.

Table 4.4: Summary results of the pairwise comparison of the $AFCCF_\gamma$ means from the simulation study in *experiment 3*. The means of the $AFCCF_\gamma$ of the *cases* are compared with one another. The last column, Reference, identifies which means result statistically different.

$AFCCF_\gamma$ **means**	$AFCCF_\gamma$ **means difference**			
comparison of	L.Limit	Means diff.	U. Limit	Ref.
Case 1 vs Case 2	0.0292	0.0377	0.0461	*
Case 1 vs Case 3	-0.0122	-0.0037	0.0048	
Case 2 vs Case 3	-0.0498	-0.0414	-0.0329	*

4.3.2 SCOMS performance on the number of regions and θ identification.

A dataset was simulated under the following conditions. In a spatial field partitioned into $R = 5$ regions, whose geometry is equal to the one of the spatial field given in Fig. 4.2(b), a dataset with 400 sites was generated assuming ω be function of the shared edge and $\theta = 2.0$. The number of covariates is the same as before, $q = 10$.

With this dataset, the SCOMS method is evaluated to learn about how it does when the number of regions and the parameter θ are varied in a range of feasible values. The simulated dataset was analyzed by the SCOMS method on a grid of regions and θ values. In order to judge the model fit, we computed the logarithm of the *marginal density* of the data estimated as in Eq. (4.2). This simulation study is carried out to see whether the SCOMS method can identify the R and θ assumed in the data simulation.

We allow the number of regions, R, to take values on the set $\{4, 5, 6\}$ and θ on the set $\{0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$. At each R-θ combination, the $\log(f(\boldsymbol{y}|R, \theta))$ is computed. The result of this simulation study is presented in Table 4.5.

Table 4.5: Summary of the evaluation of the method under different regions and θs. The logarithm of the marginal density of the data, $\log(f(\boldsymbol{y}|R, \theta))$, was estimated as the goodness of fit statistic. The dataset is simulated with $\theta = 2.0$.

Region	$\theta = 0.5$	$\theta = 1.0$	$\theta = 1.5$	$\theta = 2.0$	$\theta = 2.5$	$\theta = 3.0$
$R = 4$	-1722.9	-1384.8	-1815.7	-1990.1	-1769.1	-1820.2
$R = 5$	-704.5	-704.8	-700.7	-699.9	-701.6	-707.7
$R = 6$	-716.2	-720.5	-716.5	-715.8	-732.9	-715.7

From Table 4.5, we see the logarithm of the marginal density of the data is maximized at $R = 5$ for all θ in its grid of feasible values. This results supports the following conclusion: the SCOMS method efficiently identifies the appropriate number of regions for the data.

When $R = 5$ in the table, the logarithm of the marginal density is maximized at $\theta = 2.0$, as expected. However, a closer inspection of the logarithm of the marginal density of the data when $R = 5$, reveals a rather weak evidence in favor of the real θ, i.e. the $\log f(\boldsymbol{y}|R = 5, \theta = 1.5)$ and $\log f(\boldsymbol{y}|R = 5, \theta = 2.0)$ are only slightly different. Hence, we could conclude that the model fit, based on the marginal density of the data, is as good with $\theta = 1.5$ as it is with $\theta = 2.0$. Therefore, the $\log f(\boldsymbol{y}|R = 5, \theta)$ does not provides strong support to choose either of these two values.

Provoked by the results just presented above, we repeat the evaluation of the SCOMS method utilizing another synthetic dataset. This data are simulated similarly as the one before, in the same spatial field, with the same ω, etc., the only difference is that we set $\theta = 1.5$. The purpose of this evaluation is to see how the $\log f(\boldsymbol{y}|R, \theta)$, as the statistic utilized to select the optimum number of regions and value for θ, performs on cases when the parameter θ that generates the data, is smaller.

Table 4.6 shows the results of the analysis of this other dataset. Again, the number of regions is well identified by the $\log f(\boldsymbol{y}|R, \theta)$, and it also shows that $\log f(\boldsymbol{y}|R = 5, \theta)$ failed on the identification of the right value of θ. When $R = 5$, the marginal density of the data is maximized at $\theta = 1.0$, and not at $\theta = 1.5$ as it must be.

Table 4.6: Summary of the evaluation of the method under different regions and θs. The logarithm of the marginal density of the data, $\log(f(\boldsymbol{y}|R, \theta))$, was estimated as goodness of fit statistic. The dataset is simulated with $\theta = 1.5$.

Region	$\theta = 0.5$	$\theta = 1.0$	$\theta = 1.5$	$\theta = 2.0$	$\theta = 2.5$
$R = 4$	-1541.5	-2918.2	-2035.2	-1959.0	-2201.3
$R = 5$	-645.7	-643.1	-647.2	-647.4	-649.7
$R = 6$	-662.2	-661.1	-725.4	-664.7	-663.7

As mentioned before, the *marginal density* of the data is an useful statistic for the selection of the optimum number of regions as seen above. Unfortunately, based on the results of these two simulations, it turned out that the *marginal density* is not equally useful for the selection of the optimum value of the parameter θ.

4.4 Case study

4.4.1 Data

This presents details about the analysis of the brook trout dataset utilizing the Spatially Correlated Model Selection (SCOMS) method. The dataset has 3,337 sampled sites with the following information. The status of the brook trout fish and a set of landscape and anthropogenic variables. The location of observations are scattered along the eastern part of the United States, as shown in Fig. 1.1. The status variable, which is the response variable in the forthcoming analysis, describes whether the trout is present or absent at the sites of observation, while the landscape and anthropogenic variables (a.k.a. stressor metrics) describe physical characteristics of the sites, such as types of cover crops, type of human activities nearby, water quality where the fish lives, etc. The brook trout data has been analyzed by Thieling (2006), and later by Zhang et al. (2008). Thieling screened the stressor metrics based on criteria such as: completeness, range, redundancy, and responsiveness (see Thieling (2006) and Hudy et al. (2008) for details on the screening criteria). Zhang et al. fit logistic regressions to the data, where five stressor metrics where included as regressors. Among the five stressor metrics that Zhang et al. included in their analysis, some were derived as combination of the original stressor metrics.

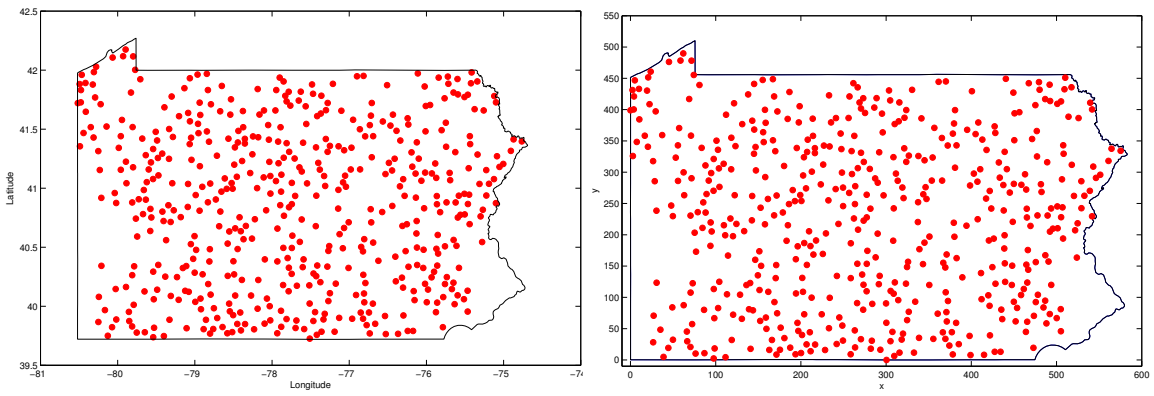
In this analysis, three regressors additional to the Zhang et al.'s (2008) five stressor metrics are included. Table 4.7 lists all the variables and their descriptions, that take part in the current analysis. The STATUS is the binary response variable. The LONGITUDE and LATITUDE are the coordinates of the observational sites, used later as inputs to compute the distance between sites. The INDUST_TRANS, TRANSITIONAL and MIXED_FOREST are the three extra regressors included in the current analysis, and the last five are the metrics that Zhang et al. (2008) considered in their analysis. In summary, the last eight regressors in Table 4.7 are the ones that may have a significant effect on explaining some variation in the response variable, but it will be the results of the SCOMS method that will tell which are the regressor that are indeed the real driving factors of the STATUS variable.

Table 4.7: Set of candidate metrics.

Metric	Description
STATUS	Brook trout population status in a subwatershed. This binary variable is equal to 1 if the trout is present and 0 otherwise
LONGITUDE	Longitude measured in decimal degrees
LATITUDE	Latitude measured in decimal degrees
INDUST_TRANS	Percentage commercial/industrial/transportation in the subwatershed
TRANSITIONAL	Percentage transitional -areas of sparse vegetation in the subwatershed
MIXED_FOREST	Percentage mixed forested lands in the subwatershed
ELEVATION	Mean elevation in meters of the subwatershed
LROAD_DNS	Subwatershed road density (km of road per km^2 of land), in logarithm scale
PERCENT_AG	Percentage of subwatershed area in agricultural use
LCHEM	An environmental factor that incorporated depositional NO_3 and SO_4, in logarithm scale
PERCENT_FOREST	Percentage forest lands in the water corridor of the subwatershed

Two different subsets of the brook trout data are analyzed. Both are based on a sample size of 500 sites. The first subset is a random selection of sites belonging to the state of Pennsylvania (hereafter called the Penn data). The second is a random selection of sites from the brook trout data (hereafter denoted WBT for Whole Brook Trout) after discarding the information belonging to the states of Vermont, New Hampshire and Maine, because in those states almost uniformly the STATUS=1 to all the sites (Zhang et al. 2008).

A scatter plot of the selected sites in the Penn data are given in Fig. 4.14(a). The geometrical similarity of this spatial field with the one assumed in the simulation study (Fig. 4.2(a)) is apparent. To make these two spatial fields even more similar, we transform the coordinates of the Penn data. The transformation is



(a) (b)

Figure 4.14: Distribution of sites in Pennsylvania. Panel (a) shows the distribution of the 500 sampled sites in the Penn data. The points (in red) represent sites of observations. Panel (b) shows the transformed Penn data spatial field.

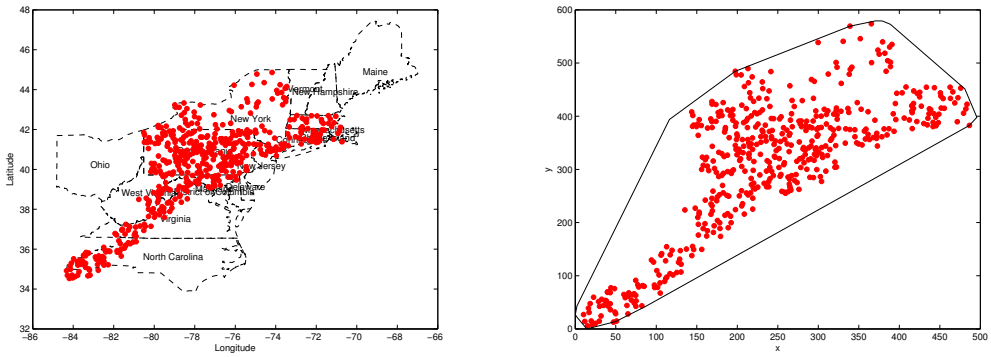
$$
\begin{aligned}
& y = 200 * \text{LATITUDE} \\
& \Rightarrow \text{Latitude}^* = y - (\min(y) - 0.01) \\
& x = 100 * \text{LONGITUDE} \\
& \Rightarrow \text{Longitude}^* = x - (\min(x) - 0.01),
\end{aligned}
$$

that yields a new spatial field shown in Fig. 4.14(b). After the transformation, the new coordinates range approximately between $[0, 600] \times [0, 560]$. Given the similarity between the Euclidian and the Geodesic (Banerjee 2005) distances, the former computed using the transformed coordinates and the latter using the original coordinates, we work with the transformed coordinates, because the Euclidian distances are computed more efficiently, using functions already available in common software (MATLAB software, for example).

Likewise, the sites associated with the WBT data are presented in Fig. 4.15(a). To have a clear idea of the coverage of this data set, we added a map in the background to the scatter plot.

Fig. 4.15(b) shows the WBT data after transforming the coordinates using the following equations

$$
\begin{aligned}
& y = 55 * \text{LATITUDE} \\
& \Rightarrow \text{Latitude}^* = y - (\min(y) - 0.01) \\
& x = 35 * \text{LONGITUDE} \\
& \Rightarrow \text{Longitude}^* = x - (\min(x) - 0.01).
\end{aligned}
$$



(a) (b)

Figure 4.15: Map of the WBT spatial field. Panel (a) shows the map of the WBT spatial field. Panel (b) displays the transformed WBT spatial field with the closest convex hull as the polygon that surrounds the field.

The scatter plot of the WBT also plots the closest convex polygon containing the data. This polygon is the convex hull (Okabe et al. 2000) of the original dataset.

4.4.2 The Model for the Binary Response Variable

To model the binary response, STATUS, of the brook trout dataset, we apply the generalization of the SCOMS method, presented in Section 3.3.1. The hierarchical specification of the model has at its top rung the mapping of the binary random field into the Gaussian random field as follows.

$$Z_{i,l} = \begin{cases} 1 & Y_{i,l} \geq 0, \\ 0 & Y_{i,l} < 0. \end{cases}$$

Where $Z_{i,l}$ is the binary response at the lth site of observation in the ith region, $i = 1, \ldots, R$ and $l = 1, \ldots, n_i$. The binary response is 1 if brook trout is present and 0 otherwise. The next level of the hierarchy has an underlying Gaussian process specification as follows.

$$\boldsymbol{Y}_i \sim N(\boldsymbol{X}_i \boldsymbol{\vartheta}_i, \boldsymbol{\Sigma}_i).$$

This model is the same as the one presented in Eq. (3.8), where $\boldsymbol{\vartheta}_i$ is the effect size, and $\boldsymbol{\Sigma}_i$ is approximated by the CAR model, as before. The design matrix, $\boldsymbol{X}_i$ is a matrix of centered and scaled covariables.

Given the spatial correlation of the Gaussian random field, to accomplish the mapping above, the conditional distribution of every Y_l is needed. To efficiently carry out the mapping, the conditional distribution is directly given by the CAR model in Eq. (2.14).

Further levels in the hierarchical model, where the prior distributions for the unknowns: namely $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$, are specified, are left unchanged; these priors were given in Section 3.5.1, and for $\boldsymbol{\tau}$ we follow the sampling scheme 1 proposed by Imai and van Dyk (2005) (mentioned in Section 3.5.1).

4.4.3 Selection of regions based on model fit and model prediction

In applications of the SCOMS methods to spatially correlated data, including the brook trout data and subsets of it such as the Penn data or the WBT data, a fairly computationally efficient strategy for the selection of the optimum number of regions to partition the spatial field (R) and the parameter θ in the Ising distribution is presented in this section. This strategy is based upon cross-validation, that evaluates the predictive ability of the model of a given method, more specifically of the SCOMS method. We noticed that a strict and careful application of the strategy guarantees the successful selection of the optima for both unknown parameters, R and θ. The strategy includes the following steps. Note that although we use the Penn data to illustrate the steps, the strategy is general and can be applied to the analysis of any kind of data.

Step 1. A set of feasible regions is defined. The definition of the set of feasible values depends on the size of the spatial field, and also on the expert's opinions. For example, for the Penn data, we believe that the optimum number of regions is contained in the set $R \in \{1, 2, 3, 4\}$. At this step, the parameter θ is set to a constant, for example $\theta = 0.5$. Likewise, ω is set to be a function of a physical characteristic of the region. In the simulation study, we did not find any difference between making ω a function of the shared edge ($\omega(L)$) or a function of the area ($\omega(A)$). Therefore, the choice of ω is left to the modeler. In this application, we choose $\omega(L)$.

Step 2. The number of sites in the dataset is split into two disjoint subsets, one is the training dataset, $[\boldsymbol{Z}^1\boldsymbol{X}^1]$, and the other is the validation dataset, $[\boldsymbol{Z}^0\boldsymbol{X}^0]$. The validation dataset is a random selection of usually the 10% of the sites (For the Penn data, this is equal to 50 sites that are selected uniformly across the spatial field).

Step 3. With the information from the steps above, the MCMC algorithm is initialized utilizing only the training dataset. The algorithm is stopped once it completes a pre-specified number of iterations, T.

Step 4. After convergence of the Markov chains, the $AFCCF_{CV}$ (given by Eq. (4.6) below) is computed using the validation dataset only. In this application, the number of iterations is $T = 100,000$, where the last 15,000 samples are taken as samples from the stationary distributions of the unknown parameters.

Step 5. For each R, steps 2, 3 and Item 4 are repeated a given number of times. Each time with different subsets of training and validation data. For this case study, we repeat it five times.

Step 6. Select $R = r$ as the one that maximizes the mean or the median of the $AFCCF_{CV}$s.

Step 7. The number of regions is fixed at $R = r$. Proceed with the cross-validation but now allowing the regional dependence parameter, θ, to vary on a set of feasible values. In the Penn data, we assume that $\theta \in \{0.5, 1.0, 1.5, 2.0\}$. For every θ in the set of values, repeat the cross-validation a given number of times. We repeat it for four times.

Step 8. Select θ as the one that maximizes the mean or the median of the $AFCCF_{CV}$s.

Step 9. Summarize the values of the effect size $\boldsymbol{\vartheta} = \boldsymbol{\beta} \circ \boldsymbol{\gamma}$ and other parameters in the model.

Step 10. Conclude.

4.4.4 The computation of the AFCCF for cross-validation

In datasets where the response variable is binary, as it is in the brook trout dataset and its subset: the Penn and WBT data, we can still apply the SCOMS method with appropriate modifications of the underlying model and the sampling distribution. Above, in Section 4.4.2, we specified the model for binary responses, which represents a generalization of the model assumed for the SCOMS method, as detailed in Section 3.3. In the SCOMS method, the optima for the number of regions and the parameter θ in the Ising distribution need to be found. Both parameters are proposed to be selected based on cross-validation. The $AFCCF_{CV}$ is the statistic that summarizes the cross-validation results. The AFCCF was

utilized in the simulation study to evaluate how accurate the real models were found by the SCOMS method when applied under different *cases* and *experiments*. In this section, the computation of the AFCCF as the summarizing statistic for the cross-validation evaluation is carried out differently. What follows are the details on the computation of the $AFCCF_{CV}$ for cross-validation is addressed.

To compute the $AFCCF_{CV}$ for cross-validation, it is necessary to calculate the posterior predictive probability $P(\boldsymbol{Z}^0|\boldsymbol{Z}^1, \boldsymbol{X})$ for the validation data $[\boldsymbol{Z}^0, \boldsymbol{X}^0]$ given the training data $[\boldsymbol{Z}^1, \boldsymbol{X}^1]$, where $\boldsymbol{X} = [\boldsymbol{X}^0, \boldsymbol{X}^1]$. For instance, suppose that we have R regions and we know which sites in the validation data are in the ith region, $\boldsymbol{Z}_i^0$. Since the binary random field is uniquely determined by the Gaussian random field, we can estimate the posterior predictive distributions in terms of the latent Gaussian process as follows. We need to have realizations of the latent process $\boldsymbol{Y}_i^0$ given realizations of $\boldsymbol{Y}_i^1$. Hence, in terms of the latent variables, we have that the posterior predictive distribution for $\boldsymbol{Y}_i^0$ is given by

$$P(\boldsymbol{Y}_i^0|\boldsymbol{Y}_i^1, \boldsymbol{X}_i) = \int P(\boldsymbol{Y}_i^0|\boldsymbol{Y}_i^1, \boldsymbol{X}_i, \boldsymbol{B}_i)P(\boldsymbol{B}_i|\boldsymbol{Y}_i^1, \boldsymbol{X}_i^1)d\boldsymbol{B}_i. \tag{4.3}$$

Here $\boldsymbol{B}_i = [\boldsymbol{\beta}_i, \boldsymbol{\gamma}_i, \rho_i]$, and $P(\boldsymbol{B}_i|\boldsymbol{Y}_i^1, \boldsymbol{X}_i^1))$ is the posterior distribution of $\boldsymbol{B}_i$ given the training data (assuming that $\boldsymbol{Y}_i^1$ and $\boldsymbol{Z}_i^1$ are in agreement, whenever $Z_{i,l} = 1 \Rightarrow Y_{i,l} \geq 0$, and $Z_{i,l} = 0 \Rightarrow Y_{i,l} < 0$ otherwise). From our model specification above, the conditional density $P(\boldsymbol{Y}_i^0|\boldsymbol{Y}_i^1, \boldsymbol{X}_i, \boldsymbol{B}_i)$ has the following closed form

$$P(\boldsymbol{Y}_i^0|\boldsymbol{Y}_i^1, \boldsymbol{X}_i, \boldsymbol{B}_i) = N(\boldsymbol{\mu}_{0|1}^{(i)}, \boldsymbol{\Sigma}_{0|1}^{(i)}), \text{ for } i = 1, \ldots, R, \tag{4.4}$$

where (we ignore the region subindex for clarity's sake)

$$\boldsymbol{\mu}_{0|1} = \boldsymbol{X}^0\boldsymbol{\vartheta} + \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}^1 - \boldsymbol{X}^1\boldsymbol{\vartheta}),$$

and

$$\boldsymbol{\Sigma}_{0|1} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{01}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{10}.$$

Here

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{01} \\ \boldsymbol{\Sigma}_{10} & \boldsymbol{\Sigma}_{11} \end{bmatrix}$$

is a partition of the variance-covariance matrix, corresponding to the partitioned latent vector

$$\begin{bmatrix} \boldsymbol{Y}^0 \\ \boldsymbol{Y}^1 \end{bmatrix}.$$

Given N samples of $\boldsymbol{B}$ and $\boldsymbol{Y}^1$, say $\{\boldsymbol{B}^{(1)}, \boldsymbol{Y}^{1(1)}, \ldots, \boldsymbol{B}^{(N)}, \boldsymbol{Y}^{1(N)}\}$ after convergence, we draw N samples from Eq. (4.4), $\{\boldsymbol{Y}^{0(1)}, \ldots, \boldsymbol{Y}^{0(N)}\}$ and we estimate $P(\boldsymbol{Z}^0|\boldsymbol{Z}^1, \boldsymbol{X})$ as follows

$$\hat{p}(\boldsymbol{z}^0|\boldsymbol{z}^1) = P(\boldsymbol{Z}^0|\boldsymbol{Z}^1, \boldsymbol{X}) = \frac{\sum_{t=1}^{N} \boldsymbol{I}_{(\boldsymbol{Y}^{0(t)} > \boldsymbol{0})}}{N}, \tag{4.5}$$

where $\boldsymbol{I}_{(.)}$ is an indicator vector, and $\hat{p}(\boldsymbol{z}^0|\boldsymbol{z}^1)$ is a vector of estimated probabilities.

Once $\hat{p}(\boldsymbol{z}^0|\boldsymbol{z}^1)$ is computed, the $AFCCF_{CV}$ for cross-validation is as follows

$$AFCCF_{CV} = \frac{\sum_{l=1}^{n_0} \left[z_l^0 \ \hat{p}(\boldsymbol{z}^0|\boldsymbol{z}^1)_l + (1 - z_l^0)(1 - \hat{p}(\boldsymbol{z}^0|\boldsymbol{z}^1)_l) \right]}{n_0}, \tag{4.6}$$

where $z_l^0 \in \{0, 1\}$ is the observation of Z_l^0 and n_0 is the number of sites in the validation dataset.

4.5 Results for the Penn and the WBT data

A probit regression analysis is carried out for both datasets, the Penn and the WBT data. In these cases, the STATUS is the variable for the binary process (the Z random variable in our model). In accordance with the model specification, we assume that there is a latent process (the Y random variable) hidden underneath the binary process, whose specification is given in Eq. (3.8). These two processes, the binary and the Gaussian, are connected via the formula given in Eq. (2.16). As mentioned earlier, the current analysis represents an extension or a generalization of the SCOMS method, exactly as it is explained in Section 3.3.1.

There are five different validation datasets of size $n_0 = 50$. For each validation set, we have a training dataset with 450 observations. As said before, after convergence, we computed the $AFCCF_{CV}$ for each of the five replications, to define the optimum number of regions, while $\theta = 0.5$. The results of this analysis follows.

4.5.1 Results of the analysis of the Penn data

The analysis the Penn data starts with the application of the steps 1 to 6 of the strategy explained in Section 4.4.3, on the five training-validation subsets of the data. As a result, boxplots of the $AFCCF_{CV}$ for each region considered in the grid of regions, $R \in \{1, 2, 3, 4\}$, are computed and shown in Fig. 4.16.

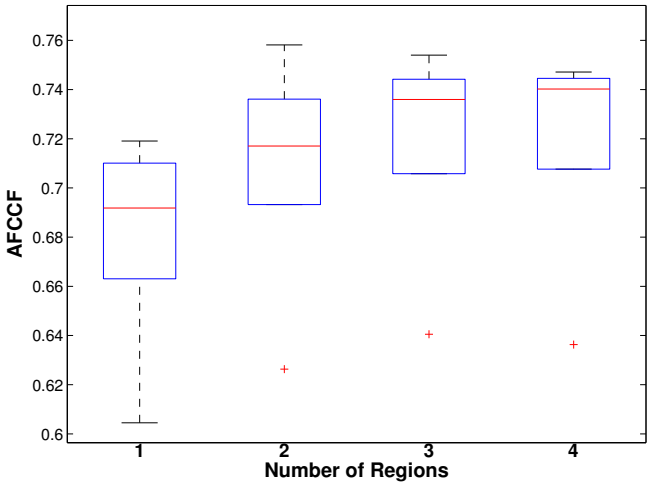


Figure 4.16: The $AFCCF_{CV}$ boxplots for the five cross-validation evaluations for the Penn data ($\theta = 0.5$).

According to Fig. 4.16, the $AFCCF_{CV}$ plateaus when the spatial field is partitioned into $R = 3$; therefore we take $R = 3$ as the optimum number of regions to partition the Penn spatial field. A comment about an outlier value of the $AFCCF_{CV}$ in Fig. 4.16. This value occurred completely by chance, from a sample selected at random. But even the $AFCCF_{CV}$ values for that sample increases and plateaus when $R = 3$ and $R = 4$.

It is worth noticing, from Fig. 4.16, that in terms of prediction, the case when the field of the Penn data is partitioned into R = 3 regions dominates the case when the field is assumed stationary (R=1), resulting in a relative improvement of the $AFCCF_{CV}$ of about 7%.

The next part of the analysis involves the selection of θ for the Ising distribution. This analysis consists in the application of the steps 7 and 8 of the strategy of Section 4.4.3, on four training-validation subsets of the data. The number of regions is fixed to $R = 3$, and the parameter θ is allowed to take values on the grid $\{0.5, 1.0, 1.5, 2.0\}$. As a result, Fig. 4.17

shows the boxplots of the $AFCCF_{CV}$ for each θ in the grid. The set of boxplots, suggest $\theta = 1.5$ as the more suicandidate for the Penn data, since it gives the highest median of the $AFCCF_{CV}$.

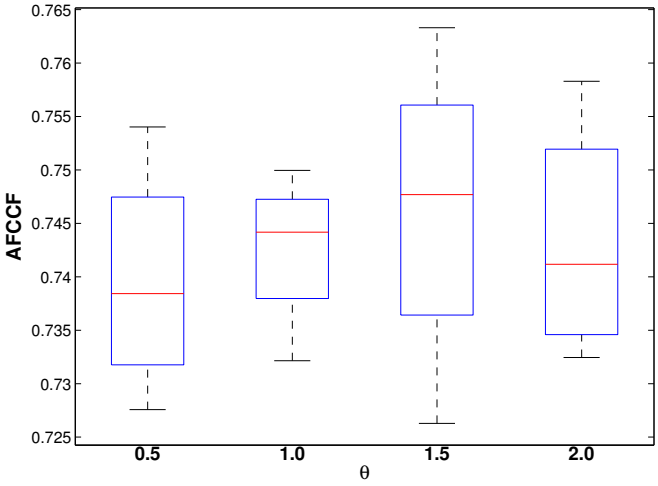


Figure 4.17: The $AFCCF_{CV}$ boxplots for the four cross-validation evaluations for the Penn data, where $R = 3$.

Therefore, based on the cross-validation criterion the optima number of regions and the parameter θ are $R = 3$ and $\theta = 1.5$ for the Penn data.

Finally, in order to estimate the unknowns, namely $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\rho}$, in the model given in Eq. (3.8), we re-ran the algorithm for $T = 100,000$ iterations, with $R = 3$ and $\theta = 1.5$, with the complete dataset, i.e. the 500 sites. After discarding the first 85,000 interations, we computed the effect size estimates (see Eq. (3.8) for the definition of effect size parameter) and their 90% credible interval, as well as the credible interval for ρ in the CAR model, based on the last 15,000 iterations, where convergence of the Markov chains is reached.

Tables 4.8 to 4.10 and 4.11 summarize the results of the analysis of the Penn data. In Table 4.8 we present the posterior probability of the indicator variable $\gamma_{i,j}$, associated with the jth stressor metric in region i, i.e. $P(\gamma_{i,j} = 1|Data)$. It is clear that the stressor metrics with high posterior probability are ELEVATION and PERCENT_FOREST, in the three regions. In region 1, PERCENT_FOREST has posterior probability slightly less than 0.5, suggesting that the variable is of low importance. Some other covariables are of low importance, as well. For instance, in region 3, LCHEM and PERCENT_AG are

the next two most important stressor metrics, with $P(\gamma_{3,LCHEM} = 1|Data) = 1.85\%$ and $P(\gamma_{3,PERCENT\_AG} = 1|Data) = 1.81\%$, but are not important.

Table 4.8: The posterior relative importance of the stressor metrics measured in term of the posterior probability $\gamma_{i,j}$ as $P(\gamma_{i,j} = 1|Data)$ (Penn data).

	Region		
Metric	**I**	**II**	**III**
INDUST_TRANS	0.0021	0.0136	0.0111
TRANSITIONAL	0.0015	0.0023	0.0132
MIXED_FOREST	0.0041	0.0023	0.0081
ELEVATION	**0.9942**	**0.9087**	**1.0000**
LCHEM	0.0025	0.0040	0.0185
LRDKM	0.0028	0.0025	0.0144
PERCENT_AG	0.0051	0.0056	0.0181
PERCENT_FOREST	0.4566	**1.0000**	**0.9979**

Tables 4.9, 4.10 and 4.11 contain the point estimate and the 90% credible interval of the effect size, $\boldsymbol{\vartheta}_i$, and the parameter ρ_i for the CAR model for $i = 1, 2, 3$.

Table 4.9: The 90% credible intervals and point estimate of the effect sizes ($\boldsymbol{\vartheta}_1$), and ρ in the CAR model, in region 1 (Penn data).

	90% Credible Interval		
Metric	**L. Lim.**	**Median**	**U.Lim.**
INDUST_TRANS	-0.0065	-0.0001	0.0063
TRANSITIONAL	-0.0038	0.0000	0.0038
MIXED_FOREST	-0.0228	0.0007	0.0242
ELEVATION	**0.1841**	**0.3238**	**0.4635**
LCHEM	-0.0101	-0.0002	0.0097
LRDKM	-0.0106	-0.0001	0.0103
PERCENT_AG	-0.0249	-0.0007	0.0236
PERCENT_FOREST	-0.0851	0.0327	0.1506
ρ (CAR model)	0.7627	0.8997	0.9713

It is apparent that ELEVATION and PERCENT_FOREST are the most important stressor metrics out of the eight variables considered in the analysis to explain the STATUS variable. In the three regions, the effect sizes of these two stressor metrics are positive, meaning that it is more likely to find trout at subwatersheds located at high elevations and with high

Table 4.10: The 90% Credible intervals and point estimate of the effect sizes ($\boldsymbol{\vartheta}_2$), and ρ in the CAR model, in region 2 (Penn data).

	90% Credible Interval		
Metric	**L. Lim.**	**Median**	**U.Lim.**
INDUST_TRANS	-0.0558	0.0038	0.0633
TRANSITIONAL	-0.0100	0.0000	0.0100
MIXED_FOREST	-0.0077	-0.0001	0.0075
ELEVATION	**0.0185**	**0.3404**	**0.6623**
LCHEM	-0.0190	0.0005	0.0201
LRDKM	-0.0144	0.0003	0.0150
PERCENT_AG	-0.0332	0.0003	0.0337
PERCENT_FOREST	**0.4894**	**0.7673**	**1.0452**
ρ (CAR model)	0.8217	0.9316	0.9816

Table 4.11: The 90% credible intervals and point estimate of the effect sizes ($\boldsymbol{\vartheta}_3$) and ρ in the CAR model, in region 3 (Penn data).

	90% Credible Interval		
Metric	**L. Lim.**	**Median**	**U.Lim.**
INDUST_TRANS	-0.0600	0.0022	0.0644
TRANSITIONAL	-0.0735	0.0044	0.0822
MIXED_FOREST	-0.0483	0.0004	0.0491
ELEVATION	**1.0012**	**1.8228**	**2.6445**
LCHEM	-0.1337	-0.0088	0.1162
LRDKM	-0.1489	-0.0079	0.1331
PERCENT_AG	-0.1750	-0.0086	0.1577
PERCENT_FOREST	**1.1454**	**2.0266**	**2.9077**
ρ (CAR model)	0.7533	0.8969	0.9732

presence of forest. Strictly speaking, in region 1, ELEVATION is the only driving stressor metric, but the relative importance of PERCENT_FOREST given in Table 4.8 suggests a slightly different conclusion. Because the $P(\gamma_{1,PERCENT\_FOREST} = 1|Data) = 45.66\%$, one can say that PERCENT_FOREST has a scant significant effect.

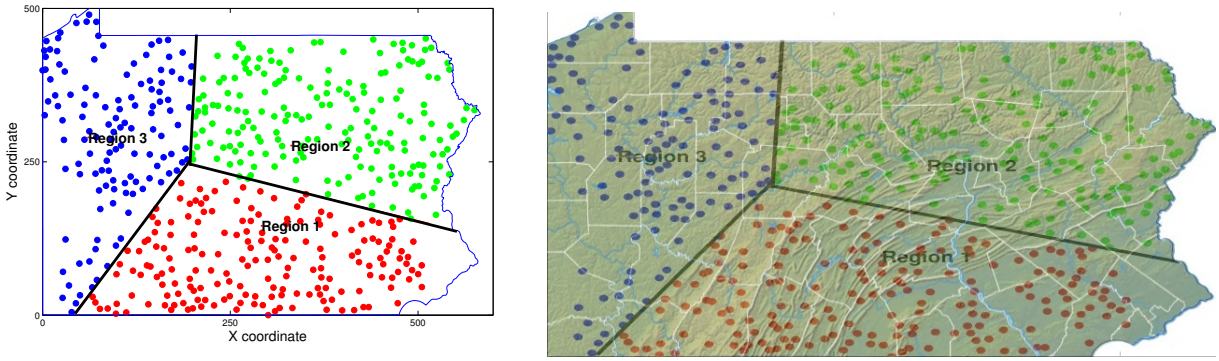
As a result of the application of the SCOMS method to the analysis of the Penn data, the relationship that the STATUS variable has with the two significant stressor metrics changes across the regions. The summary results of region 1, 2 and 3 presented in the Tables 4.9, 4.10 and 4.11, show that the effect sizes of the significant stressor metrics are different among the

regions. For instance, in region 1 the effect sizes of ELEVATION and PERCENT_FOREST are 0.324 and 0.033, respectively. In region 2 the effect sizes of the same metrics are 0.340 and 0.767. And in region 3, the effect sizes are 1.823 and 2.026. We can see that the effect of the ELEVATION and PERCENT_FOREST stressor metrics on the STATUS variable is stronger in region 3, weaker in region 1, and somewhat in the middle in region 2. In summary, the three regions, each one with its own model, prove that the spatial field is not stationary. Moreover, although all have the same significant stressor metrics, the models are different from one another because the effect sizes of the stressor metrics change from model to model. In conclusion, the SCOMS method is efficiently finding three dependent regions where the spatial field is stationary, and also identifies the most significant regressors in each region.

The effect sizes of LCHEM and PERCENT_AG in region 3 come out to be non-significant (not suprising since their relative importance shown in Table 4.8 are small). Nevertheless, their point estimates given in Table 4.11 are negative. This suggests that even a negligible contribution, both LCHEM and PERCENT_AG metrics reduce the chances of finding brook trout fish at sites where these two metrics have large values. This is somewhat in agreement with intuition, since both metrics are derived from human related activities, their presence results in higher perturbations of the fish's habitat.

The final partition of the spatial field is shown in Fig. 4.18(a). Fig. 4.18(b) shows a map of the topography of the state of Pennsylvania together with the partition suggested by the analysis. This lets us see that some of the boundaries that separates the regions follow topographical features. For example, the boundary that separates region 3 from region 1 runs over the ridge of the mountains. On the other hand, the upper part of regions 1 and lower part of region 2 share similar topographical conditions, but the topography present in the upper part of region 2 is not present in region 1, and the left-bottom's topography of region 1 is different from region 2. These characteristics are somehow determining the regions.

Finally, if the regions were independent of one another, the parameter θ in the Ising distribution would be any value close to $\theta = 0$, hence $\theta = 0.5$ would give higher values of the $AFCCF_{CV}$ compare with the $AFCCF_{CV}$ corresponding to $\theta > 0.5$. Hence, having found that $\theta = 1.5$ is interpreted as that the regions are rather dependent of one another, and their correlation is quite strong. Therefore, the SCOMS method is efficiently uncovering the regions' correlation, and using it to model the posterior distribution of $\boldsymbol{\gamma}$, the matrix of indicator variables.



(a) (b)

Figure 4.18: Partition of the Pennsylvania spatial field into $R = 3$ regions. Panel (a) the partition of the Penn spatial field into $R = 3$ regions. This partition is the mean of the after burn in (the last 15,000 iterations) samples of partitions. Panel (b) the partition of the Penn spatial field overlaid on a topography map. The solid dots are sites of observations, and the solid lines are the region boundaries.

In the next part we present the results of the analysis of the Penn data, but this time the binary spatial field is assumed stationary and isotropic. Above, we found that this approach is the worst for prediction, evaluated by the $AFCCF_{CV}$ statistic (see Fig. 4.16, when $R = 1$). However, this is likely the first analysis one might perform on the data set, and probably use it as a reference to compare any other analyses that answer the same type questions.

The Penn data analysis: a stationary and isotropic spatial field

Even though the assumption of stationarity in the Penn data resulted in a model with the lowest predictive ability, evaluated by the $AFCCF_{CV}$ shown in Fig. 4.16 when $R = 1$, this section presents the result of the analysis of the stationary and isotropic Penn data.

Table 4.12 shows the posterior probability of the indicator variable, γ_j, associated to the jth stressor metrics. It also shows the 90% credible interval for the unknown parameters. A single region was considered because the stationarity assumption implies that a single model or region in the spatial field is enough to explain the relationship of the binary response variable with the covariables.

From Table 4.12 we see that ELEVATION is again significant and its effect on the binary response is positive, as before. A somewhat different result to the one obtained above when

the spatial field is partitioned into three regions, we found in this analysis that the other significant metric is the PERCENT_AG, and its effect is negative. The negative effect of this stressor metric in the response variable is sensible. Agriculture, as any other human activity, tends to perturb natural habitats. And its effects is such that it may cause some species (the brook trout specifically) to reduce their presences at places affected by this human activity.

In an attempt to explain this rather inconsistent result, namely: In one analysis PERCENT_FOREST metric is found significant but not PERCENT_AG, and in other analysis the latter is found significant but not the former, we went deeper only to find that both metrics, PERCENT_AG and PERCENT_FOREST, are highly linearly correlated. Their sample correlation is -0.8816. This result explains why in this analysis PERCENT_AG was the other most significant stressor metric, rather than PERCENT_FOREST as found in the analysis of the data when three regions where assumed. And due to the negative correlation between these two stressor metrics, their effect sizes signs differ, one negative and the other positive. Before, in the analysis above, the effect size of the PERCENT_FOREST is positive in all the three regions, in this analysis the effect size of PERCENT_AG is negative. Therefore, the signs on the effect size of each metric are in agreement with their correlation.

Clearly, both stressor metrics bear important information to explain variations on the binary response variable, but because of their high correlation only one of them is allowed to be considered in the model to avoid redundancy. Therefore, the variable selection procedure as proposed by Kuo and Mallick (1998) adapted for the SCOMS method is efficiently handling the multicollinearity problem.

4.6 Analysis results for the WBT data

To analyze the WBT (whole brook trout) data, we follow the process applied before for the analysis of the Penn data. First, we apply steps 1 to 6 of the strategy presented in Section 4.4.3 to find the optimum number of regions. The set of feasible regions is $R \in \{1, 2, 3, 4, 5, 6, 7, 8\}$, and $\theta = 0.5$. The result of this part of the analysis is shown in Fig. 4.19. The boxplots correspond to the Rs in the set of feasible number of regions. Fig. 4.19 suggests that the median of the $AFCCF_{CV}$ is maximized when $R = 6$, whose boxplot seems to be placed slightly above all others.

From Fig. 4.19, one can notice that in prediction terms, the case when the field of the

Table 4.12: The summary of the analysis of the Pennsylvania brook trout data, when its spatial field is assumed stationary and isotropic. The posterior probability of the indicator variable $\mathbf{P}(\gamma_\mathbf{j} = \mathbf{1}|\ldots)$, the 90% credible intervals and the point estimate of the effect sizes (ϑ) and ρ.

		90% Credible Interval		
Metric	$\mathbf{P}(\gamma_\mathbf{j} = \mathbf{1}\|\ldots)$	**L. Lim.**	**Median**	**U.Lim.**
INDUST_TRANS	0.0253	-0.0186	-0.0013	0.0161
TRANSITIONAL	0.0127	-0.0091	-0.0003	0.0086
MIXED_FOREST	0.0447	-0.0264	0.0032	0.0329
ELEVATION	**1.0000**	**0.2361**	**0.3158**	**0.3955**
LCHEM	0.0621	-0.0400	-0.0047	0.0305
LRDKM	0.0759	-0.0523	-0.0070	0.0384
PERCENT_AG	**0.9945**	**-0.3710**	**-0.2838**	**-0.1967**
PERCENT_FOREST	0.0385	-0.0437	0.0017	0.0470
ρ (CAR model)	—	0.8376	0.9346	0.9829

WBT data is partitioned into R = 6 regions dominates the case when the field is assumed stationary (R=1), resulting in a relative improvement of the $AFCCF_{CV}$ of about 7.4%.

The second part of the analysis implies the application of steps 7 and 8 of the strategy, to select the optimum value for θ in the Ising distribution. The number of regions is set to $R = 6$, and we allow θ to take values on the set of feasible values $\{0.0, 0.5, 1.0, 1.5, 2.0, 2.5\}$. The results of this part of the analysis is shown in Fig. 4.20. There we can see that the θ that maximizes the median of the $AFCCF_{CV}$ is $\theta = 0.5$.

The final part of this analysis consisted in re-running the algorithm again but now with $R = 6$ and $\theta = 0.5$ considering the 500 sites in the WBT dataset. The algorithm was run for $T = 500,000$, the first 450,000 iterations were discarded and the last 50,000 were taken as samples from the stationary distribution of the parameters. The results are presented in Table 4.13. In that table, the point estimate of the *effect size* (Est.) and besides it, in parenthesis, the posterior probability of the indicator variable corresponding to each stressor metric are presented. Below these two values, the 90% credible interval of the *effect size* is given. Few stressor metrics were found to be important. In regions 1, 4 and 6 only PERCENT_FOREST is significant; in regions 2 and 5 only ELEVATION is significant; in region 3, MIXED_FOREST and PERCENT_FOREST are significant. Some comments based on the results in Table 4.13 are brought to the reader's attention.

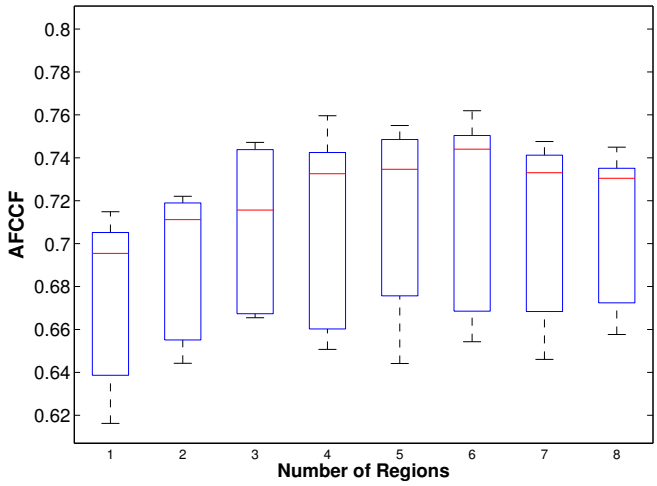


Figure 4.19: The $AFCCF_{CV}$ boxplots for the five cross-validation evaluations on the WBT data to select R. Here the number of regions, R, varies in a set of feasible values (here $\theta = 0.5$).

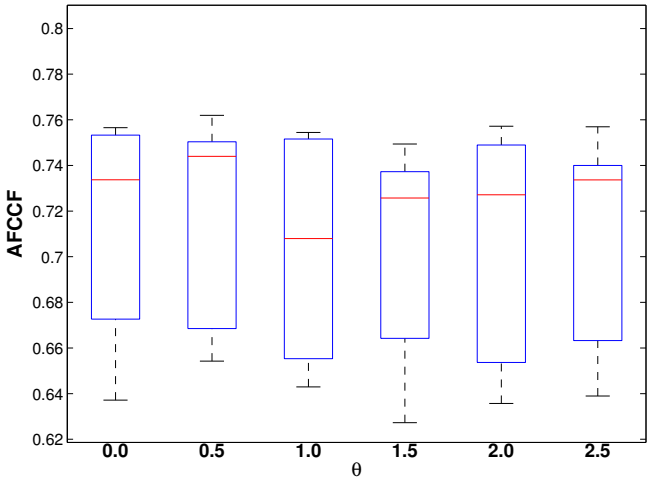


Figure 4.20: The $AFCCF_{CV}$ boxplots for the five cross-validation evaluations on the WBT data to select θ. The parameter θ varies in a set of feasible values (here $R = 6$).

- The *effect sizes* of the significant stressor metrics are different from region to region. It means that the relationship between the STATUS variables and the significant stressor metrics is different across the spatial field, a consistent result related to the assumption of non-stationarity of the field.

- All the *effect sizes* are positive. It means that the higher the value of the stressor metric, the higher the chances of the brook trout fish to be found in the regions.

Table 4.13: The point estimation (Est.), the posterior probability of the indicator variable ($P(\boldsymbol{\gamma}_{ij}) = P(\gamma_{i,j} = 1|\ldots)$), and the 90% Credible Interval for the *effect sizes*, for WTB data.

	Regions					
	1	**2**	**3**	**4**	**5**	**6**
Metric	**Est.**($P(\boldsymbol{\gamma}_{1j})$) **(L.L, U.L)**	**Est.**($P(\boldsymbol{\gamma}_{2j})$) **(L.L, U.L)**	**Est.**($P(\boldsymbol{\gamma}_{3j})$) **(L.L, U.L)**	**Est.**($P(\boldsymbol{\gamma}_{4j})$) **(L.L, U.L)**	**Est.**($P(\boldsymbol{\gamma}_{5j})$) **(L.L, U.L)**	**Est.**($P(\boldsymbol{\gamma}_{6j})$) **(L.L, U.L)**[b]
INDUST_TRANS	0.00 (0.32) (-4.55, 0.00)	0.00 (0.05) (0.00, 0.00)	0.00 (0.16) (-0.71, 0.13)	0.00 (0.27) (-0.09, 0.75)	0.00 (0.14) (-0.87, 0.00)	0.00 (0.28) (-3.49, 0.28)
TRANSITIONAL	0.00 (0.17) (-0.07, 1.71)	0.00 (0.01) (0.00, 0.00)	0.00 (0.17) (-0.87, 0.42)	0.00 (0.25) (-1.84, 0.49)	0.00 (0.06) (0.00, 0.00)	0.00 (0.16) (-2.21, 0.17)
MIXED_FOREST	0.00 (0.35) (-2.64, 0.78)	0.00 (0.03) (0.00, 0.00)	**2.11 (1.00) (1.13, 3.55)**	0.00 (0.26) (0.00, 0.67)	0.00 (0.10) (0.00, 0.38)	0.00 (0.16) (0.00, 1.35)
ELEVATION	0.00 (0.28) (-0.10, 4.49)	**0.39 (0.96) (0.15, 0.58)**	0.00 (0.36) (-2.51, 0.49)	0.00 (0.16) (-0.58, 0.00)	**1.88 (1.00) (1.10, 2.96)**	0.00 (0.22) (-1.34, 0.00)
LCHEM	0.00 (0.17) (0.00, 1.63)	0.00 (0.04) (0.00, 0.00)	0.00 (0.09) (0.00, 0.00)	0.00 (0.43) (-2.18, 0.00)	0.00 (0.05) (0.00, 0.00)	0.00 (0.13) (-0.25, 0.00)
LRDKM	0.00 (0.15) (-1.86, 0.00)	0.00 (0.14) (-0.22, 0.00)	0.00 (0.13) (-0.44, 0.07)	0.00 (0.12) (-0.25, 0.00)	0.00 (0.21) (-1.05, 0.00)	0.00 (0.08) (0.00, 0.00)
PERCENT_AG	0.00 (0.14) (0.00, 0.60)	0.00 (0.03) (0.00, 0.00)	0.00 (0.16) (0.00, 0.97)	0.00 (0.32) (0.00, 1.35)	0.00 (0.07) (-0.10, 0.00)	0.00 (0.44) (0.00, 3.72)
PERCENT_FOREST	**0.51 (0.72) (-0.35, 1.73)**	0.00 (0.09) (0.00, 0.24)	**3.76 (1.00) (1.89, 6.76)**	**0.77 (0.83) (0.00, 2.52)**	0.00 (0.11) (0.00, 0.06)	**0.90 (0.97) (0.43, 5.57)**
ρ (CAR model)	0.92 (0.79, 0.98)	0.94 (0.83, 0.98)	0.91 (0.77, 0.98)	0.92 (0.78, 0.98)	0.89 (0.75, 0.97)	0.91 (0.77, 0.98)

The locations of the regions in the spatial field are shown in Fig. 4.21.

The results of the analysis of the WBT data when the spatial field is assumed stationary are presented in the following section.

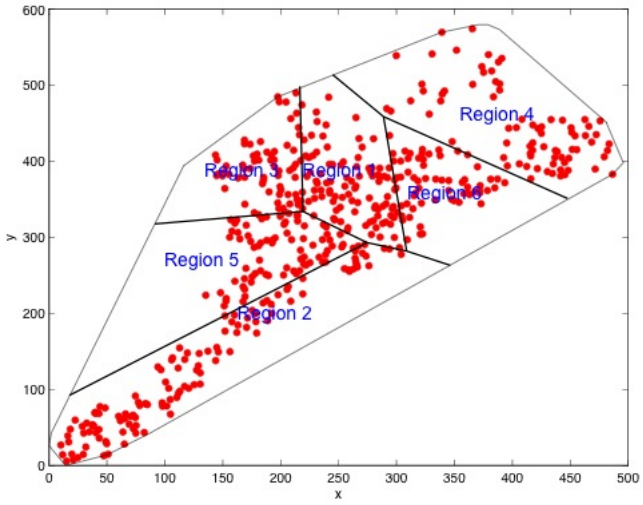


Figure 4.21: Partition of the WBT spatial field into R = 6 regions. The partition is the mean of the after burn in (the last 50,000 iterations) samples of partition.

The WBT data analysis: a stationary and isotropic spatial field

The results of the analysis of the WBT data when the field is assumed stationary and isotropic, are given in Table 4.14. This shows the posterior probability of the indicator variable, the point estimate and the 90% credible interval of the stressor metrics effect sizes. There is only one significant stressor metric, PERCENT_FOREST, with a positive effect on the STATUS variable. Based on this result, one can say that if the spatial field of the WBT data are assumed stationary, crucial information is lost. Some stressor metrics whose significance is local are not discovered. To recover the locally significant information, one must assume that the WBT data's spatial field is non-stationary, as before where the spatial field was partitioned into $R = 6$ regions, and apply the SCOMS method.

The results from the stationary and non-stationary analyses of the WBT data, empirically prove the potential of the SCOMS method. The SCOMS method gives results that might help to widen perspectives and to unveil in more detail the relationships between the response variable and the covariables.

Table 4.14: Summary of the analysis of the WBT data when stationarity is assumed. The posterior probability of the indicator variable $\mathbf{P}(\gamma_\mathbf{j} = \mathbf{1}|\ldots)$, the 90% credible intervals and the point estimate of the effect sizes (ϑ) for the stressor metrics, and the ρ.

		90% Credible Interval		
Metric	$\mathbf{P}(\gamma_\mathbf{j} = \mathbf{1}\|\ldots)$	**L. Lim.**	**Median**	**U.Lim.**
INDUST_TRANS	0.0579	0.0000	0.0000	0.0299
TRANSITIONAL	0.0119	0.0000	0.0000	0.0000
MIXED_FOREST	0.0147	0.0000	0.0000	0.0000
ELEVATION	0.0133	0.0000	0.0000	0.0000
LCHEM	0.0161	0.0000	0.0000	0.0000
LRDKM	0.1111	-0.1115	0.0000	0.0000
PERCENT_AG	0.0119	0.0000	0.0000	0.0000
PERCENT_FOREST	**1.0000**	**0.3362**	**0.4096**	**0.4754**
ρ (CAR model)	——	0.8678	0.9473	0.9858

Chapter 5

Discussion and Conclusions

This thesis proposed and evaluated the *Spatially Correlated Model Selection* (SCOMS) method, as an alternative to the analysis of non-stationary and non-isotropic spatial fields, when variable selection is of concern. The SCOMS method approach consists in partitioning the spatial field into a number of regions, R. At each region the following is assumed,

- the spatial process is stationary,
- a linear model of the form $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\vartheta}_i + \boldsymbol{\epsilon}_i$ is fit, with $\boldsymbol{\epsilon}_i \sim CAR$ (the Conditional Autoregressive model given in Eq. (2.15)), and $i = 1, \ldots, R$; and also
- it is dependent on adjacent regions.

The dependence between regions is quantified and included in the analysis through the Ising distribution assumed as prior for the indicator variables $\gamma_{i,j}$ in the variable selection procedure proposed by Kuo and Mallick (1998). In agreement with the Kuo and Mallick's (1998) method, the *effect size* parameter is defined as $\boldsymbol{\vartheta}_i = [\beta_{i,1}\gamma_{i,1}, \ldots, \beta_{i,q}\gamma_{i,q}]'$, where $\beta_{i,j}$ are the regression coefficients of the jth covariate at the ith region. The Ising distribution (Smith and Fahrmeir 2007) depends on two parameters. The parameter ω that measures the relationship between adjacent regions, and the parameter θ which quantifies the overall interaction or dependence among regions, so that if $\theta = 0$ the regions are independent of each other. In this research we proposed two different ways to measure ω, both based on physical characteristics of the regions (see Section 3.5.2). The two alternatives for ω were evaluated through a simulation study presented in Chapter 4. The results of the simulation study were given in Section 4.3.1, and the conclusions based on the simulation results are

presented later in this chapter. For the parameter θ and the optimum number of regions, R, we proposed a procedure based on cross-validation (described in Eq. (3.14)) to identify their optimal values. The cross-validation procedure was applied in the analysis of the brook trout data (see Section 4.4), which is the data utilized to apply the SCOMS method as case studies. Some conclusion from the results of the analysis of the case study, are presented below.

In summary and based on the results obtained from the simulation study and from the case studies, we believe that the SCOMS method is a valid alternative for the analysis of non-stationary spatial fields, and provides simple answers to questions such as

- In how many regions should a spatial field of interest be partitioned?
- Where should the boundaries of the regions be located?
- Is there a subset of covariates that are significant in explaining the response variable of interest at the regions level?

The following sections address in more details the results of both analyses: the simulation and the case studies. At the end of this chapter, we present general aspects of the methods as well as some alternatives to overcome computational issues found during the evaluation of the method.

5.1 The simulation study

As common practice in evaluations of statistical methods, the simulation study was designed to show that the SCOMS method works under its underlying modeling assumptions. What is important to highlight is that this method showed robustness on the choice of the metric ω, that measures the dependence of adjacent regions. When comparing the results of *cases 1* and *2* in *experiments 1* and *2*, we did not find any significant difference between the cases in either experiments, i.e. the performance of the SCOMS method was as good with $\omega(L)$ (as a function of the shared edge of adjacent regions) as with $\omega(A)$ (as a function of the area of adjacent regions), regardless the ω that generated the data, when the comparisons are based on the $AFCCF_\gamma$ calculated on the $\boldsymbol{\gamma}$ parameter, as the summarizing statistic explained in Section 4.2.1. On the other hand, when inference was carried out under the

assumption that the regions were independent (accomplished by making $\theta = 0$, in the Ising distribution), on datasets generated with dependent regions, we obtained poorer results than when the dependence between regions were considered. These are the results from *case* 3 in *experiments 1* and *2*, presented in Figs. 4.9 and 4.12.

When data are simulated on independent regions, as long as θ is assumed small (*case 1* in *experiment 3*) the SCOMS method gives as good results as when $\theta = 0$ (the case of independent regions, *case 3* in *experiment 3*). But as θ departs from zero (*case 2* in *experiment 3*), the method gives poorer results for inference, as observed in Fig. 4.13.

In the simulation study, the *marginal density* of the data was utilized in the selection of the optimum number of regions and in the determination of θ. In Tables 4.5 and 4.6 we see that the *marginal density* of the data successfully identified the correct number of regions. However, this provides weak evidence or non-conclusive results useful for the selection of the optimum value of θ, since when θ was equal to 1.5, the *marginal density* of the data was maximized at $\theta = 1.0$ (see Table 4.6).

We believe that the harmonic mean of the likelihood function as a way to estimate the *marginal density* (Newton and Raftery 1994) is not appropriate for the model formulations such as the one specified in the SCOMS method. Nevertheless, more investigation is needed to completely understand the reasons of the non-conclusiveness of the estimates of the *marginal density* of the data, especially when it is utilized as an statistic for the selection of θ in the Ising distribution.

The use of the *marginal density* is, therefore, restricted to the selection the number of regions, R. Nonetheless, if it is carefully applied, it could also be utilized for the identification of a value for θ (bearing in mind that the suggested θ could be in the vicinity of the optimum value), only when the response of interest is continuous. For cases where the response is binary, as in the case study, we definitely suggest to base the selection of the R and θ on cross-validation.

The results from the simulation study showed that the SCOMS method is a general procedure to analyze non-stationary spatial fields. With this method, one can investigate about the optimum number of regions to partition a spatial field, and whether these regions are dependent or independent of one another through the parameter θ that determines the degree of the correlation of the regions. As illustrations of what was just said, we have the simulation results presented in Tables 4.5 and 4.6, and the analyses of the data in the

case studies. In the case studies, we saw how to proceed to learn about these parameters in practice. What makes the SCOMS a general method is the Ising distribution assumed as prior for the parameter of indicator variables, $\boldsymbol{\gamma}$. The Ising distribution models the dependence of the regions. Two particular cases of the SCOMS method are worth mentioning: 1) When the optimum number of regions $R = 1$, the SCOMS method reduces to the case when the spatial field is stationary. And 2) when the optimum value for $\theta \simeq 0$, the SCOMS method reduces to the case of independent regions (the case studied by Kim et al. (2005)).

5.2 Case Study

5.2.1 The Penn data

The cross-validation performed to a subset of 500 sites located within the Pennsylvania state boundaries, helped select the optimum number of regions, $R = 3$, and the right value of $\theta = 1.5$, that results in good prediction rates, judge by the $AFCCF_{CV}$ (where values of the $AFCCF_{CV}$ closer to one are preferred) as the statistic that summarizes the cross-validation results.

The case of a single model across the spatial field of interest which directly assumes stationarity of the field, gave smaller values of the $AFCCF_{CV}$ statistic (see Fig. 4.16). According the results of the analysis, it is more appropriate to partition the spatial field of the Penn data into $R = 3$ regions and fit different models on each region. This results proves that the binary spatial field is non-stationary.

The Pennsylvania brook trout spatial field is partitioned into three regions

Given the model that establishes the relationship between the binary response variable (the status of the brook trout fish) with the set of covariates, we found that in the three regions ELEVATION and PERCENT_FOREST were the most significant stressor metrics for the STATUS variable, although in region 1 PERCENT_FOREST was "almost" significant (see Tables 4.9, 4.10 and 4.11).

The point estimate of the *effect sizes* are positive for both metrics: ELEVATION and PERCENT_FOREST. This is in agreement with the nature of the trout. Since brook trout

is adapted to cold waters, intuitively, the temperature of the streams water decreases as the elevation increases. The metric PERCENT_FOREST plays an important role in terms of habitat alteration, since intuitively the more forest present in the sub-watershed the fish's habitat is less perturb. Therefore the probability of the presence of the fish in sub-watersheds with high elevations and/or high percent of forest increases.

As a result of the analysis of the brook trout data set, we found that $\theta = 1.5$. This suggests that the three regions that split the Pennsylvania spatial field are correlated. If they were independent, the most suitable θ would have been the one close to zero, therefore, the $AFCCF_{CV}$ associate with $\theta = 0.5$ would have been greater than the $AFCCF_{CV}$ associated with any other $\theta > 0.5$. But instead, the cross-validation analysis suggested $\theta = 1.5$. Hence, the application of the SCOMS method helps to undercover the correlation between regions and, through the Ising distribution, to include this correlation to leverage the significance of stressor metrics in the models.

It is also important to highlight that the metrics LCHEM and PERCENT_AG in region 3 might potentially have negative impact on the trout in the long run, particularly if mining and agricultural activities increase in the region.

The Pennsylvania brook trout spatial field is assumed stationary

The analysis of the Penn data assuming that its spatial field is stationary gives the following results. There were two significant stressor metrics: ELEVATION and PERCENT_AG. The PERCENT_FOREST metric was not found important in this analysis, but it was found significant when the spatial field was assumed non-stationary. However, in order to find a reason for this supposed inconsistency between the two approaches: the non-stationary and the stationary, we found that the metrics PERCENT_AG and PERCENT_FOREST are negatively correlated (-0.8816). As a result, the SCOMS method chose one out of two highly correlated stressor metrics as significant, showing that it is efficiently handling the multicollinearity problem.

5.2.2 The WBT data

The cross-validation analysis of the WBT data results in the following: The optimum number of regions of the spatial field and the optimum value for θ in the Ising distribution are $R = 6$

and $\theta = 0.5$. Three stressor metrics are significant to explain the STATUS variable with the following allocation. In regions 1, 4 and 6 only PERCENT_FOREST is significant; in regions 2 and 5 only ELEVATION is significant; in region 3, MIXED_FOREST and PERCENT_FOREST are significant (see Table 4.13). All the *effect sizes* of the significant stressor metrics are positive in all the regions. The final partition of the field is presented in Fig. 4.21.

The WBT data analysis under stationary assumption

Even though the case of one region resulted in a model fit with the poorest predictive performance measured in terms of the $AFCCF_{CV}$ (see Fig. 4.19), the WBT data are analyzed under the assumption that its spatial field is stationary. As results, we found that the PERCENT_FOREST is the only significant stressor metric for the STATUS variable. This proves that this analysis was unable to uncover stressor metrics whose significance is high but localized. However, using the SCOMS method on the analysis of non-stationary spatial field can help to find those locally significant metrics.

5.3 Miscellaneous

The estimation of the parameters of the variance-covariance matrix in a Gaussian Random Field (GRF) brings computational problems particularly with large sample sizes, this issue is even more markedly if the GRF is used as a latent process underneath the Binary Random Field (BRF) as in the case study. Calculation of determinants and inverses of large matrices are computationally demanding. The numerical complexity of the GRF grows at $O(n^3)$ rate (Banerjee et al. 2008). Hence, for the generalization of the SCOMS method to binary responses, the variance-covariance of the Gaussian process is approximated by CAR model as a way to speed up the mapping between the BRF and the GRF. This approximation was also used in the simulation study. There, we assumed the CAR model as a simpler model to save us the computation burden that matrix inversions represents. Based on the simulation result, the CAR model has shown good potential for making the algorithm's processing faster. But we are aware about the potential inappropriateness of the CAR model with non-lattice data. A detailed evaluation to address how much is lost or/and gained with the CAR model is needed, but at least now we have found it advantageous for the SCOMS method,

particularly for probit regressions of binary random fields.

The Voronoi tessellation was utilized to partition the spatial field into regions. This tessellation technique is computationally efficient, and it works only when the polygons of the spatial fields and regions are convex; for spatial fields with non convex polygons this tessellation technique is not longer a solution. One way to work around the non-strictly convex polygons is by working with a convex hull (Okabe et al. 2000) of the data, which gives the closest convex polygon of the spatial field of the data, as we did in the analysis of the WBT data (see Fig. 4.15(b)). In general, the tessellations of non-convex polygons represent computational challenges that we did not investigate in this research, but undoubtably it will require further and closer examination.

Appendix A

Full conditionals derivation

A.1 The linear model

For binary random fields, we have that $Z_{i,l}$, the binary response at site l located at region i, can be mapped into a Gaussian random field as follows,

$$Z_{i,l} = \begin{cases} 1 & \text{if } Y_{i,l} \geq 0 \\ 0 & \text{if } Y_{i,l} < 0. \end{cases} \tag{A.1}$$

The response $\boldsymbol{Y}$ at region i is linearly related with the matrix of covariables as follows

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\vartheta}_i + \boldsymbol{\epsilon}_i, \text{ for } i = 1, \ldots, R, \tag{A.2}$$

where $\boldsymbol{\vartheta}_i = [\beta_{i,0}\gamma_{i,0}, \ldots, \beta_{i,q}\gamma_{i,p}]'$, $\boldsymbol{X}_i = [\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,q}]$ is the design matrix, and

$$\gamma_{i,j} = \begin{cases} 1, & \boldsymbol{x}_{i,j} \text{ is included} \\ 0, & \boldsymbol{x}_{i,j} \text{ is omitted.} \end{cases} \tag{A.3}$$

The random term is

$$\boldsymbol{\epsilon}_i \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i), \tag{A.4}$$

where $\boldsymbol{\Sigma}_i \approx \tau_i^2(\boldsymbol{D}_{\omega_i} - \rho_i \boldsymbol{W}_i)^{-1}$ given in Eq. (2.15). To simplify computation, and to favor the identification of the parameters in the mean model when the response is binary, $\tau_i = 1$.

A.2 Prior specifications

For the unknown parameters, we specify independent prior distributions. For $\boldsymbol{\beta}$, we assume the prior

$$\pi(\boldsymbol{\beta}) = \prod_{i=1}^{R} N_q(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_{q\times q}), \tag{A.5}$$

where $\sigma_\beta^2 = 10$.

For $\boldsymbol{\gamma}$, we assume the Ising prior as follows,

$$\pi(\boldsymbol{\gamma}|\theta) \propto \prod_{j=1}^{q} \exp\left\{\theta \sum_{i\sim i'} \omega_{i,i'}\ I(\gamma_{i,j} = \gamma_{i',j})\right\}. \tag{A.6}$$

Although unknown, the parameter θ is treated as fixed. The parameter ω is deterministic and based upon physical characteristics of the regions.

A.3 Full conditionals

Let $\boldsymbol{Y} = \boldsymbol{y}$ is observed, the likelihood function is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X}) = \prod_{i=1}^{R} (2\pi)^{-n_i/2} |\Sigma_i|^{-1/2} \exp\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\vartheta}_i)'\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\vartheta}_i)\},$$

but since this methodology is concerned in how the models are correlated, we are interested in the likelihood function after integrating out the parameter $\boldsymbol{\beta}$, given its prior distribution in Eq. (A.5). We leave $L(\cdot)$ in terms of $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$, because $\boldsymbol{\gamma}$ is the parameter that conveys the model correlation. This marginal likelihood is given by

$$\begin{aligned} L^*(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X}) = \ & \prod_{i=1}^{R} (2\pi)^{-n_i/2} \frac{|\boldsymbol{\Sigma}_i|^{-1/2}|\boldsymbol{\Sigma}_\beta|^{-1/2}}{|\boldsymbol{X}_i^{*\prime}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i^* + \boldsymbol{\Sigma}_\beta^{-1}|^{1/2}} \times \\ & \exp\{-\tfrac{1}{2}(\boldsymbol{y}_i'(\boldsymbol{\Sigma}_i^{-1} - \boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i^*(\boldsymbol{X}_i^{*\prime}\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i^* + \boldsymbol{\Sigma}_\beta^{-1})^{-1}\boldsymbol{X}_i^{*\prime}\boldsymbol{\Sigma}_i^{-1})\boldsymbol{y}_i)\}, \end{aligned} \tag{A.7}$$

where $\boldsymbol{c}(R)$ is the vector of centroids of the regions, whose prior distribution is given in Eq. (3.18).

The posterior distribution for all the unknown parameters is

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X}) \propto L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\rho})\pi(\boldsymbol{c}(R)).$$

The full conditional for $\boldsymbol{\beta}_i$ is

$$\pi(\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{\gamma}_i, \rho_i, \boldsymbol{c}(R)_i) = N_q\left(\boldsymbol{D}_i^* \boldsymbol{X}_i^{*'} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{y}_i, \boldsymbol{D}_i^*\right), \tag{A.8}$$

where $\boldsymbol{D}_i^* = \left[\boldsymbol{X}_i^{*'} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i^* + \boldsymbol{\Sigma}_\beta^{-1}\right]^{-1}$, $\boldsymbol{\Sigma}_\beta^{-1} = \frac{1}{\sigma_\beta^2}\boldsymbol{I}$, and $\boldsymbol{X}_i^* = [\boldsymbol{X}_{i,1}\gamma_{i,1}, \ldots, \boldsymbol{X}_{i,q}\gamma_{i,q}]$. And the full conditional for $\boldsymbol{\gamma}$ with $\boldsymbol{\beta}$ integrated out is

$$\pi(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\rho}, \boldsymbol{c}(R)) \propto \exp\left\{\theta \sum_{j=1}^{q} \sum_{i \sim i'} \omega_{i,i'} I(\gamma_{i,j} = \gamma_{i'j}) + \log L^*(\boldsymbol{\gamma}, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{X})\right\}. \tag{A.9}$$

A computational efficient expression derived from Eq. (A.9) is the conditional for $\gamma_{i,j}$, $\pi(\gamma_{i,j}|\boldsymbol{\gamma}_{-(i,j)}, \ldots)$, where $\boldsymbol{\gamma}_{-(i,j)}$ is the $\boldsymbol{\gamma}$ matrix with its (i,j) element omitted. Such conditional is expressed as follows,

$$\pi(\gamma_{i,j} = 1|\boldsymbol{\gamma}_{-(i,j)}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{c}(R)) = \frac{1}{1 + \exp\left\{\theta \sum_{i' \in \mathbb{N}_i} \omega_{i,i'}[I(\gamma_{i'j} = 0) - I(\gamma_{i'j} = 1)] + l^*\right\}} \tag{A.10}$$

where $l^* = \left[\log L^*(\gamma_{i,j} = 0, \boldsymbol{\gamma}_{-(i,j)}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X}) - \log L^*(\gamma_{i,j} = 1, \boldsymbol{\gamma}_{-(i,j)}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{Y}, \boldsymbol{X})\right]$, and $\mathbb{N}_i$ is the neighborhood set for region i. And

$$\pi(\gamma_{i,j} = 0|\boldsymbol{\gamma}_{-(i,j)}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\rho}, \boldsymbol{c}(R)) = 1 - \pi(\gamma_{i,j} = 1|\boldsymbol{\gamma}_{-(i,j)}, \boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\rho}, \boldsymbol{c}(R))$$

The full conditional for $\boldsymbol{c}(R)$ is given by

$$\pi(\boldsymbol{c}(R)|\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{X}, \boldsymbol{y}) \propto L^*(\boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X})\pi(\boldsymbol{c}(R)).$$

Finally, the full conditional for $\boldsymbol{\rho}$ is given by

$$\pi(\boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto L(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{c}(R)|\boldsymbol{y}, \boldsymbol{X})\pi(\boldsymbol{\rho}),$$

where $\pi(\boldsymbol{\rho})$ is found in Eq. (3.14).

Appendix B

Simulation Results

B.1 *Experiment 1:* simulation results

Table B.1: Empirical coverage rates for the elements in $\boldsymbol{\vartheta}$ for *Experiment 1*, computed as a percentage of times that the real parameter is in the 95% credible intervals, based on 100 replications.

		Regions				
Case	Parameter	1	2	3	4	5
Case 1	ϑ_1	97	95	99	97	94
	ϑ_2	97	98	98	94	98
	ϑ_3	96	98	98	98	98
	ϑ_4	97	97	98	98	95
	ϑ_5	99	97	96	95	94
	ϑ_6	94	97	95	99	97
	ϑ_7	97	97	99	100	100
	ϑ_8	98	96	96	98	96
	ϑ_9	97	98	97	99	97
	ϑ_{10}	97	94	96	96	99
Case 2	ϑ_1	96	95	99	96	94
	ϑ_2	97	98	98	94	97
	ϑ_3	96	98	99	97	97
	ϑ_4	97	97	98	98	94
	ϑ_5	97	97	95	95	95
	ϑ_6	94	96	95	100	97
	ϑ_7	97	97	99	100	100
	ϑ_8	98	96	95	98	97
	ϑ_9	96	98	97	99	97
	ϑ_{10}	97	95	96	95	99
Case 3	ϑ_1	94	94	98	97	93
	ϑ_2	97	97	98	95	97
	ϑ_3	96	98	98	97	95
	ϑ_4	97	97	98	95	94
	ϑ_5	98	97	95	93	93
	ϑ_6	93	93	92	100	96
	ϑ_7	98	97	98	97	100
	ϑ_8	98	95	94	97	94
	ϑ_9	95	96	94	96	96
	ϑ_10	95	93	92	94	95

Table B.2: AFCCFs from *Cases* 1, 2 and 3, in *Experiment 1.*

	AFCCF				AFCCF		
Rep	Case 1	Case 2	Case 3	Rep	Case 1	Case 2	Case 3
1	0.952	0.952	0.941	51	0.962	0.958	0.944
2	0.953	0.948	0.932	52	0.914	0.920	0.892
3	0.973	0.962	0.925	53	0.991	0.987	0.948
4	0.945	0.934	0.871	54	0.939	0.948	0.941
5	0.968	0.972	0.934	55	0.963	0.959	0.921
6	0.908	0.902	0.863	56	0.977	0.975	0.947
7	0.955	0.948	0.903	57	0.985	0.981	0.944
8	0.938	0.934	0.913	58	0.929	0.935	0.939
9	0.922	0.925	0.924	59	0.986	0.979	0.934
10	0.914	0.926	0.909	60	0.963	0.955	0.898
11	0.951	0.944	0.923	61	0.950	0.951	0.937
12	0.947	0.946	0.918	62	0.942	0.945	0.908
13	0.946	0.943	0.899	63	0.959	0.955	0.931
14	0.947	0.940	0.910	64	0.918	0.908	0.908
15	0.946	0.941	0.882	65	0.946	0.929	0.889
16	0.964	0.961	0.919	66	0.993	0.990	0.966
17	0.975	0.972	0.944	67	0.953	0.955	0.924
18	0.975	0.967	0.919	68	0.930	0.924	0.909
19	0.865	0.866	0.860	69	0.963	0.956	0.908
20	0.903	0.910	0.892	70	0.985	0.975	0.933
21	0.973	0.971	0.928	71	0.922	0.917	0.910
22	0.954	0.952	0.938	72	0.964	0.953	0.903
23	0.886	0.881	0.856	73	0.956	0.949	0.917
24	0.993	0.988	0.951	74	0.964	0.961	0.939
25	0.953	0.950	0.926	75	0.916	0.908	0.900
26	0.971	0.970	0.925	76	0.935	0.934	0.927
27	0.984	0.978	0.940	77	0.950	0.936	0.881
28	0.947	0.953	0.940	78	0.945	0.942	0.911
29	0.977	0.975	0.949	79	0.947	0.939	0.897
30	0.975	0.966	0.911	80	0.955	0.957	0.917
31	0.931	0.934	0.918	81	0.933	0.925	0.884
32	0.938	0.940	0.920	82	0.899	0.889	0.851
33	0.947	0.891	0.925	83	0.926	0.909	0.875
34	0.982	0.982	0.958	84	0.969	0.955	0.904
35	0.981	0.974	0.926	85	0.985	0.984	0.945
36	0.962	0.956	0.908	86	0.940	0.928	0.888
37	0.946	0.945	0.914	87	0.974	0.968	0.904
38	0.979	0.960	0.924	88	0.960	0.941	0.942
39	0.992	0.987	0.954	89	0.954	0.948	0.917
40	0.968	0.963	0.917	90	0.975	0.971	0.937
41	0.959	0.951	0.894	91	0.937	0.933	0.883
42	0.959	0.960	0.936	92	0.994	0.990	0.935
43	0.968	0.973	0.942	93	0.928	0.929	0.903
44	0.987	0.973	0.915	94	0.947	0.953	0.951
45	0.948	0.945	0.914	95	0.956	0.954	0.921
46	0.983	0.980	0.935	96	0.975	0.962	0.911
47	0.966	0.964	0.930	97	0.965	0.960	0.909
48	0.973	0.978	0.953	98	0.959	0.964	0.939
49	0.954	0.948	0.896	99	0.962	0.951	0.897
50	0.972	0.962	0.917	100	0.971	0.969	0.915

B.2 *Experiment 2:* simulation results

Table B.3: Empirical coverage rates for the elements in $\boldsymbol{\vartheta}$ for *Experiment 2*, computed as a percentage of times that the real parameter is in the 95% credible intervals, based on 100 replications.

		Regions				
Case	Parameter	1	2	3	4	5
Case 1	ϑ_1	94	96	98	96	97
	ϑ_2	97	96	96	94	93
	ϑ_3	96	96	97	95	97
	ϑ_4	98	99	92	96	97
	ϑ_5	97	97	99	96	95
	ϑ_6	98	97	98	98	98
	ϑ_7	97	97	94	99	96
	ϑ_8	96	96	95	96	93
	ϑ_9	100	96	90	97	97
	ϑ_{10}	93	99	97	97	96
Case 2	ϑ_1	94	95	98	97	97
	ϑ_2	96	96	95	92	94
	ϑ_3	96	96	96	95	97
	ϑ_4	98	99	93	96	97
	ϑ_5	97	97	98	96	95
	ϑ_6	97	96	99	98	98
	ϑ_7	97	97	94	99	96
	ϑ_8	96	95	94	96	93
	ϑ_9	99	97	90	97	97
	ϑ_{10}	93	100	97	96	95
Case 3	ϑ_1	93	95	98	94	94
	ϑ_2	95	95	97	94	96
	ϑ_3	94	95	95	96	95
	ϑ_4	96	97	92	95	97
	ϑ_5	95	96	98	95	90
	ϑ_6	94	95	98	97	99
	ϑ_7	97	97	91	97	97
	ϑ_8	97	95	93	94	92
	ϑ_9	100	95	90	96	97
	ϑ_{10}	91	97	96	94	93

Table B.4: AFCCFs from *Cases* 1, 2 and 3, in *Experiment 2*.

	AFCCF				AFCCF		
Rep	Case 1	Case 2	Case 3	Rep	Case 1	Case 2	Case 3
1	0.951	0.942	0.909	51	0.870	0.882	0.876
2	0.933	0.925	0.903	52	0.962	0.959	0.939
3	0.960	0.966	0.955	53	0.943	0.946	0.940
4	0.908	0.912	0.913	54	0.967	0.974	0.945
5	0.958	0.952	0.924	55	0.910	0.920	0.913
6	0.947	0.943	0.893	56	0.938	0.938	0.912
7	0.942	0.938	0.919	57	0.936	0.934	0.918
8	0.902	0.903	0.896	58	0.931	0.931	0.898
9	0.956	0.963	0.935	59	0.903	0.897	0.873
10	0.962	0.971	0.952	60	0.946	0.940	0.876
11	0.925	0.934	0.920	61	0.965	0.964	0.953
12	0.929	0.934	0.911	62	0.938	0.932	0.892
13	0.966	0.967	0.946	63	0.921	0.933	0.934
14	0.925	0.932	0.942	64	0.957	0.965	0.959
15	0.908	0.911	0.905	65	0.917	0.932	0.938
16	0.973	0.975	0.933	66	0.934	0.938	0.921
17	0.950	0.954	0.932	67	0.969	0.963	0.943
18	0.948	0.952	0.930	68	0.944	0.957	0.956
19	0.934	0.936	0.908	69	0.868	0.873	0.857
20	0.911	0.921	0.907	70	0.953	0.949	0.922
21	0.889	0.906	0.919	71	0.924	0.925	0.902
22	0.924	0.925	0.901	72	0.905	0.900	0.871
23	0.957	0.962	0.953	73	0.953	0.953	0.910
24	0.931	0.933	0.920	74	0.902	0.898	0.864
25	0.929	0.927	0.885	75	0.947	0.958	0.943
26	0.982	0.978	0.923	76	0.916	0.924	0.921
27	0.909	0.911	0.886	77	0.949	0.952	0.915
28	0.934	0.948	0.939	78	0.956	0.953	0.924
29	0.937	0.933	0.910	79	0.925	0.936	0.928
30	0.964	0.955	0.904	80	0.920	0.923	0.909
31	0.934	0.931	0.891	81	0.945	0.953	0.923
32	0.915	0.919	0.909	82	0.940	0.945	0.940
33	0.932	0.934	0.926	83	0.934	0.940	0.922
34	0.918	0.942	0.945	84	0.959	0.950	0.911
35	0.969	0.970	0.929	85	0.914	0.907	0.866
36	0.942	0.939	0.907	86	0.919	0.927	0.909
37	0.942	0.959	0.947	87	0.945	0.948	0.920
38	0.919	0.924	0.912	88	0.950	0.936	0.888
39	0.942	0.949	0.947	89	0.940	0.951	0.940
40	0.918	0.920	0.906	90	0.900	0.914	0.909
41	0.950	0.955	0.923	91	0.943	0.935	0.904
42	0.911	0.932	0.942	92	0.890	0.895	0.887
43	0.961	0.963	0.932	93	0.961	0.960	0.938
44	0.921	0.938	0.921	94	0.932	0.936	0.906
45	0.934	0.935	0.898	95	0.938	0.960	0.945
46	0.946	0.957	0.940	96	0.942	0.943	0.921
47	0.932	0.942	0.925	97	0.936	0.948	0.934
48	0.904	0.903	0.892	98	0.965	0.973	0.946
49	0.984	0.985	0.944	99	0.949	0.948	0.928
50	0.923	0.926	0.897	100	0.947	0.953	0.929

B.3 *Experiment 3:* simulation results

Table B.5: Empirical coverage rates for the elements in $\boldsymbol{\vartheta}$ for *Experiment 3*, computed as a percentage of times that the real parameter is in the 95% credible intervals, based on 100 replications.

		Regions				
Case	Parameter	1	2	3	4	5
Case 1	ϑ_1	99	98	96	98	96
	ϑ_2	100	95	91	95	92
	ϑ_3	97	98	93	95	92
	ϑ_4	95	99	97	96	92
	ϑ_5	97	93	93	92	96
	ϑ_6	100	98	95	96	94
	ϑ_7	97	97	94	96	98
	ϑ_8	96	96	95	100	92
	ϑ_9	97	94	95	97	94
	ϑ_{10}	93	97	94	98	97
Case 2	ϑ_1	98	97	94	99	96
	ϑ_2	100	93	94	95	92
	ϑ_3	97	97	93	96	93
	ϑ_4	96	100	96	96	93
	ϑ_5	97	93	93	92	96
	ϑ_6	100	97	94	95	95
	ϑ_7	96	96	96	96	98
	ϑ_8	96	96	97	100	93
	ϑ_9	97	94	95	97	94
	ϑ_{10}	94	97	94	97	97
Case 3	ϑ_1	97	98	96	99	96
	ϑ_2	99	94	91	95	92
	ϑ_3	97	99	95	95	94
	ϑ_4	95	99	96	97	91
	ϑ_5	98	94	94	92	96
	ϑ_6	99	98	95	96	95
	ϑ_7	95	98	94	96	98
	ϑ_8	96	95	94	100	92
	ϑ_9	97	92	95	97	94
	ϑ_{10}	93	97	93	98	95

Table B.6: AFCCFs from *Cases* 1, 2 and 3, in *Experiment 3*.

	AFCCF				AFCCF		
Rep	Case 1	Case 2	Case 3	Rep	Case 1	Case 2	Case 3
1	0.934	0.940	0.925	51	0.911	0.910	0.904
2	0.922	0.872	0.936	52	0.920	0.873	0.923
3	0.881	0.841	0.873	53	0.919	0.900	0.920
4	0.952	0.893	0.962	54	0.904	0.880	0.907
5	0.932	0.841	0.947	55	0.915	0.892	0.913
6	0.921	0.865	0.933	56	0.944	0.901	0.952
7	0.935	0.882	0.946	57	0.966	0.934	0.967
8	0.926	0.859	0.940	58	0.961	0.907	0.973
9	0.904	0.849	0.913	59	0.965	0.905	0.976
10	0.906	0.872	0.907	60	0.861	0.849	0.853
11	0.940	0.868	0.952	61	0.923	0.894	0.931
12	0.923	0.897	0.922	62	0.911	0.884	0.909
13	0.914	0.856	0.925	63	0.896	0.872	0.895
14	0.935	0.900	0.938	64	0.930	0.875	0.936
15	0.871	0.850	0.871	65	0.886	0.852	0.893
16	0.947	0.897	0.958	66	0.872	0.851	0.871
17	0.972	0.940	0.970	67	0.923	0.885	0.922
18	0.925	0.918	0.912	68	0.934	0.881	0.946
19	0.923	0.860	0.930	69	0.931	0.928	0.920
20	0.919	0.873	0.928	70	0.909	0.886	0.909
21	0.880	0.850	0.883	71	0.927	0.891	0.930
22	0.906	0.878	0.905	72	0.889	0.840	0.898
23	0.918	0.882	0.919	73	0.942	0.905	0.946
24	0.901	0.891	0.902	74	0.898	0.865	0.901
25	0.929	0.855	0.946	75	0.930	0.887	0.933
26	0.912	0.917	0.898	76	0.892	0.840	0.901
27	0.947	0.914	0.947	77	0.913	0.872	0.916
28	0.908	0.855	0.921	78	0.910	0.886	0.912
29	0.921	0.900	0.915	79	0.949	0.925	0.945
30	0.911	0.840	0.930	80	0.896	0.830	0.913
31	0.937	0.905	0.946	81	0.904	0.883	0.904
32	0.897	0.858	0.900	82	0.879	0.858	0.879
33	0.900	0.838	0.905	83	0.925	0.884	0.929
34	0.900	0.864	0.900	84	0.928	0.899	0.928
35	0.916	0.880	0.920	85	0.914	0.860	0.923
36	0.931	0.886	0.937	86	0.864	0.816	0.867
37	0.916	0.905	0.915	87	0.924	0.914	0.922
38	0.942	0.903	0.943	88	0.902	0.846	0.911
39	0.940	0.924	0.932	89	0.948	0.940	0.936
40	0.952	0.910	0.956	90	0.926	0.915	0.922
41	0.953	0.919	0.955	91	0.940	0.929	0.938
42	0.868	0.821	0.879	92	0.890	0.877	0.887
43	0.912	0.867	0.920	93	0.929	0.899	0.935
44	0.957	0.914	0.958	94	0.932	0.889	0.946
45	0.951	0.900	0.956	95	0.915	0.844	0.928
46	0.883	0.829	0.891	96	0.929	0.889	0.929
47	0.922	0.881	0.928	97	0.906	0.874	0.912
48	0.932	0.869	0.943	98	0.954	0.910	0.957
49	0.919	0.858	0.930	99	0.934	0.910	0.932
50	0.919	0.877	0.922	100	0.898	0.864	0.902

Bibliography

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19(6), 716–723.

Albert, J. H., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88(422), 669–679.

Banerjee, S. (2005), "On Geodetic Distance Computations in Spatial Modeling," *Biometrics*, 61(2), 617–625.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, 1st. edn Chapman and Hall/CRC.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian Predictive Process Models for Large Spatial Data Sets," *Journal of the Royal Statistical Society. Series B*, 70(4), 825–848.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society. Series B*, 36(2), 192–236.

Besag, J. (1975), "Statistical Analysis of Non-Lattice Data," *Journal of the Royal Statistical Society. Series D*, 24(3), 179–195.

Carlin, B. P., and Banerjee, S. (2002), "Hierarchical Multivariate CAR Models for Spatio-Temporal Correlated Survival Data,", Bayesian Statistics 7. University Press.

Cassella, G., and Berger, R. L. (2002), *Statistical Inference*, 2nd. edn Duxbury.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, Revised edn New York: Wiley.

EBTJV (2006), "Eastern Brook Trout Joint Venture,", `http://www.easternbrooktrout.org`.

Fuentes, M., and Smith, R. L. (2001), "A New Class of Nonstationary Spatial Models,", Technical report, Nort Caroline State University.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Barnerjee, S. (2003), "Spatial Modeling With Spatially Varying Coefficient Processes," *Journal of the American Statistical Association*, 98(462), 387–396.

George, E. I., and McCulloch, R. E. (1993), "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), 881–889.

Good, I. J. (1983), *Good Thinking. The Foundation of the Probability and Its Applications* University of Minnesota Press.

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Determination," *Biometrika*, 82(4), 711–732.

Green, P. J., and Richardson, S. (2002), "Hidden Markov Models and Disease Mapping," *Journal of the American Statistical Association*, 97(460), 1055–1070.

Haas, T. C. (1995), "Local Prediction of a Spatio-Temporal Process with an Application to Wet Sulfate Deposition," *Journal of the American Statistical Association*, 90(432), 1189–1199.

Haran, M. (2010), "Gaussian Random Field Models for Spatial Data,", No published.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57(1), 97–109.

Heagerty, P. J., and Lele, S. R. (1998), "A Composite Likelihood Approach to Binary Spatial Data," *Journal of the American Statistical Association*, 93(443), 1099–1111.

Higdon, D. (1994), Spatial Application of Markov Chain Monte Carlo for Bayesian Inference (unpublished), PhD thesis, The University of Washington.

Higdon, D. (1998*a*), "Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications," *Journal of the American Statistical Association*, 93(442), 585–595.

Higdon, D. (1998*b*), "A Process-Convolution Approach to Modeling Temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics*, 5(2), 173–190.

Higdon, D. (2001), "Space and Space–Time Modeling Using Process Convolutions,", Technical Report. Institute of Statistics and Decision Sciences, Duke University.

Hudy, M., Thieling, T. M., Gillespie, N., and Smith, E. P. (2008), "Distribution, Status, and Land Use Characteristics of Subwatersheds within the Native Range of Brook Trout in the Eastern United States," *North American Journal of Fisheries Management*, 28, 1069–1085.

Imai, K., and van Dyk, D. A. (2005), "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation," *Journal of Econometrics*, 124, 311–334.

Izenman, A. J. (2008), *Modern Multivariate Statistics Techniques*, 1st. edn Springer.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factor," *Journal of the American Statistical Association*, 90(430), 773–795.

Kim, H.-M., Mallick, B. K., and Holmes, C. C. (2005), "Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes," *Journal of the American Statistical Association*, 100(470), 653–668.

Kuo, L., and Mallick, B. (1998), "Variable Selection for Regression Models," *The Indian Journal of Statistics, Series B*, 60(1), 65–81.

Lee, P. M. (2004), *Bayesian Statistics: an Introduction*, 3rd. edn Hodder Arnold.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006), *SAS for Mixed Models*, 2nd. edn SAS Institute Inc.

Matérn, B. (1986), *Spatial Variation*, 2nd. edn Springer-Verlag.

Møller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. (2006), "An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants," *Biometrika*, 93(2), 451–458.

Myers, R. H. (1990), *Classical and Modern Regression With Applications*, 2nd. edn Duxbury Thomson Learning.

Newton, M. A., and Raftery, A. E. (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society. Series B*, 56(1), 3–48.

O'Hara, R. B., and Sillanpää, M. J. (2009), "A Review of Bayesian Variable Selection Methods: What, How and Which," *Bayesian Analysis*, 4(1), 85–118.

Okabe, A., Boots, B., Sugihara, K., and Chiu, S. N. (2000), *Spatial Tessellation. Concepts and Applications of Voronoi Diagrams*, 2nd. edn John Wiley & Sons.

Oliveira, V. D. (2000), "Bayesian Prediction of Clipped Gaussian Random Field," *Computational Statistical and Data Analysis*, 34(3), 299–314.

Raftery, A. E. (1999), "Bayes Factor and BIC. Comment on 'A Critique of the Bayesian Information Criterion for Model Selection'," *Sociological Methods and Research*, 27(3), 411–427.

Reich, B. J., Fuentes, M., Herring, A. H., and Evenson, K. R. (2010), "Bayesian Variable Selection for Multivariate Spatially Varying Coefficient Regression," *Biometrics*, 66(3), 772–782.

Rue, H., and Tjelmeland, H. (2002), "Fitting Gaussian Markov Random Fields to Gaussian Fields," *The Scandinavian Journal of Statistics*, 29(1), 31–49.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87(417), 108–119.

Schabenberger, O., and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, 1st. edn Chapman and Hall/CRC.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6(2), 461–464.

Smith, M., and Fahrmeir, L. (2007), "Spatial Bayesian Variable Selection With Application to Functional Magnetic Resonance Imaging," *Journal of the American Statistical Association*, 102(478), 417–431.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Jornal of the Royal Statistical Society. Series B*, 64(4), 583–639.

Stroud, J. R., Müller, P., and Sansó, B. (2001), "Dynamic Models for Spatiotemporal Data," *Journal of the Royal Statistical Society. Series B*, 63(4), 673–689.

Thieling, T. M. (2006), Assessment and Predictive Model for Brook Trout (*Salvelinus fontinalis*) Population Status in the Eastern United States, PhD thesis, James Madison University.

Weisberg, S. (2005), *Applied Linear Regression*, Probability and Statistics, 3rd. edn Wiley.

Wilkinson, L. (1999), *SYSTAT*, SPSS, Chicago.

Zhang, H., Thieling, T., Prins, S. C. B., Smith, E. P., and Hudy, M. (2008), "Model-Based Clustering in a Brook Trout Classification Study within the Eastern United States," *Transactions of the American Fisheries Society*, 137(3), 841–851.