12-2017

Analysis of Data of Different Spatial Support: A Multivariate Process Approach

Kelly-Ann N. Dixon Hamil
Purdue University

ANALYSIS OF DATA OF DIFFERENT SPATIAL SUPPORT: A MULTIVARIATE PROCESS APPROACH

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Kelly-Ann Dixon Hamil

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2017

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Hao Zhang, Chair

Department of Statistics

Dr. Rebecca Doerge

Department of Statistics

Dr. Tonglin Zhang

Department of Statistics

Dr. Songlin Fei

Department of Forestry

Approved by:

Dr. Jun Xie

Head of the Department of Statistics Graduate Program

Dedicated to
my parents and
my husband, Hayden.

ACKNOWLEDGMENTS

This journey has been exciting and filled with learning experiences at every step. As I reflect, I am grateful to all those who have guided and supported me and made this experience memorable.

First, I must acknowledge my God who has not only seen me through all my challenges and accomplishments, but who has blessed me greatly with friends, family and resources.

I thank my advisor, Professor Hao Zhang, whose mentorship and guidance have been invaluable. His willingness to share his passion for applied statistical work has inspired me. I also thank my committee members - Dean Rebecca Doerge, Professor Songlin Fei and Professor Tonglin Zhang - for their continued encouragement, advice and contributions throughout my graduate career.

Outside of my committee, I have also been fortunate to be blessed with other mentors in research and teaching. Thank you Dr Laura Cayon, Dr Zenephia Evans, Professor Edray Goins, Ellen Gundlach, Professor Annmarie Nicely and Professor Mark Ward for you support and advice.

I am grateful to the faculty, staff and graduate students in the Department of Statistics. Specifically, I thank my academic siblings Dr Yong Wang and Dr Juan Hu for their assistance. I also thank Barret Schloerke for his willingness and guidance in implementing parallel computing. Additionally, thank you to Dr Basil Iannone, III for his assistance with the data used in this project.

"Gud fren beta dan pakit moni" and I appreciate this Jamaican proverb even more today than when I began this journey. Thank you to all my friends (near and far) who have supported me through my sujourn in Indiana. Special mention goes to April, Sheridon, Pauline, Tasha, DeLean, Nicole, Monique, Daliha, Sharlene, Natalie, Aleiya, Christelene, Courtney, Ayu, Hilda and Dominique.

Thank you to both my earthly and my spiritual families. Your prayers, care packages, love and other blessings have not gone unnoticed and I truly appreciate each of you. Thank you also to my parents for always being supportive and encouraging me to aim for the stars. Without them, I would not be who I am today.

An extra special thank you goes to my husband, Hayden. For your endless love, encouragement and sacrifice, there are no words powerful enough to say how grateful I am for your presence in my life (Team Hamil to di worl'!).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Dixon Hamil, K. PhD, Purdue University, December 2017. Analysis of Data of Different Spatial Support: A Multivariate Process Approach. Major Professor: Hao Zhang.

Inherent to a spatial variable is the unit of support at which it is measured. In many studies, variables are observed at different support. For example, disease rates might be measured at an aggregated level while temperature is usually measured at specific points. It is still an interesting problem to study the relationship of variables having different support. However, it may be a different problem to statistically model the relationship of variables of different support, particularly when the supports do not have a hierarchical structure.

Currently, cokriging, the use of one or more spatial variables to predict another variable, is applied to variables of the same support. In this work, I extend cokriging for use with variables of different support by constructing a nonparametric cross-covariance matrix. This method is flexible as it applies to any marginal spatial model and is suited to large datasets because it uses latent variables which can assist with dimension reduction.

The proposed nonparametric method is demonstrated with two correlated variables which are measured at different spatial units. In addition, the method is implemented using two algorithms; one which yields an optimized matrix (Wang, 2011) and the other which produces an approximately optimized matrix but is computationally more efficient (Hu 2013). The results show that the method is appropriate for predicting data of different support and that it outperforms some competing methods with respect to predictive performance. Furthermore, as expected, the approximately optimized matrix does not perform as well as the alternative algorithm, but it performs better than the comparative methods.

1. INTRODUCTION

1.1 Motivation

The spatial unit, or support, at which data gathering occurs is typically chosen based on the physical and financial constraints of collection balanced by the information needed for analysis. This therefore means that data collected for different variables and/or by different agencies may not have the same spatial unit despite having been collected in the same spatial field. Additionally, over time, these units may change due to advancement in technology and/or changes in the boundaries of the region being studied. Moreover, due to limited resources, data are now being shared in larger quantities. As a result, data gathered from more than one source may not be compatible. Health outcomes, for example, are known to be related to socio-economic and climatic variables, but these may all be measured at different spatial units. Health outcomes may be measured at the county level, while socio-economic variables may be reported by enumeration districts, and the measurements of climatic variables are taken at specific points. For effective policies and proper planning at the local, regional and global scales, this disparity must be accounted for in any analysis.

In cases where the spatial units are the same but the variables are different, cokriging is used to improve the accuracy of predictions (Stein et al., 1988; Stein & Corsten, 1991; Knotters et al., 1995; Wu et al., 2009). This is especially true in situations where the related variables are measured at more locations than the variable to be predicted (Zhang & Cai, 2015). However, it is not immediately clear how cokriging can be applied to data of different support.

Typically, methods such as Bayesian Hierarchical models (BHMs) (Le & Zidek, 1992) and area-to-point (ATP) kriging methods (Kyriakidis, 2004) have been used to

solve this issue, but they are not always ideal. For example, when using BHMs, the covariance matrix is constructed by conditioning the process with the smaller spatial units (e.g. points) on the process with the larger support (e.g. areas) (Wikle & Berliner, 2005). However, this may not have a scientific interpretation if the process with the smaller spatial unit in not driven by the process with the larger one. This suggests that other methods are needed to handle this and other situations where existing methods may present challenges. As a consequence, I propose an alternative method.

1.2 Proposed Solution

Due to the multivariate nature of the problem, there is the need to model the cross-correlation between the processes. In this dissertation, I advance a method for constructing cross-covariance matrices for data of different support which uses the marginal covariance matrices for each process. Given any two stochastic processes and/or random fields (indexed by i), each of the covariance matrices may be written as $\mathbf{\Sigma}_i = \mathbf{A}_i\mathbf{A}_i'$. It is from this spectral decomposition of the covariance matrices that the cross-covariance matrix can be created.

From Wang (2011), it is known that the cross-covariance between the partial realisation of any 2 spatial processes, $\mathbf{Y}_i$ and $\mathbf{Y}_j$ $i \neq j$ is $Cov(\mathbf{Y}_i, \mathbf{Y}_j) = \mathbf{A}_i\mathbf{J}_{ij}\mathbf{Y}_j$ where $\mathbf{J}_{ij}$ represents the correlation between the set of latent variables from each univariate process and $\mathbf{A}_i$ is from the spectral decomposition outlined above. In order to construct the cross-covariance matrices presented in this dissertation, I assume that the correlation is nonparametric. This method is an alternative to other methods as it is flexible enough to be used with many types of marginal covariance functions. Therefore it is suitable for use with data which are measured at different levels of support and it does not require the variables being used in the analysis to have a hierarchical structure.

The method to create the proposed nonparametric cross-covariance matrices is relatively easy to implement. It is an optimization problem and the compromise is that once the dataset is large, the algorithm may be computationally expensive, especially as it relates to time. One advantage of using latent variables is that they assist with dimension reduction (e.g Cressie and Johannesson (2008)), thus reducing the potential computational issues. Additionally, the algorithm is implemented in parallel.

By employing this method to real data, I show that prediction using the nonparametric method to construct cross-covariance matrices yields improved results when compared to (i) the method using only the marginal model and (ii) the hierarchical model.

1.3 Background

Geostatistical data are spatial data that are measured at a specific point on a continuous physical plane and are usually referenced using a set of coordinates (e.g. longitude and latitude). It is also sometimes referred to as point-level data (e.g. temperature measured at weather stations). Conversely, lattice or areal data are spatial data which represent an area or can be considered as regionally aggregated data (e.g. disease prevalence for counties). The final type of spatial data are called point process data which, unlike the previously mentioned types, occur at locations which are random variables. An example of this is the epicentre of earthquakes. For the purposes of this dissertation, only the first two types of spatial data will be considered and the terms *point* and *area* will be used to refer to point-level data and lattice data respectively. These two terms will also refer to the spatial unit at which the data are collected. Alternatively, these may also be called the level of support for the data, or just support for simplicity. Due to the difference in what each measurement represents, an area or a point, using these data together can be problematic. This is issue of support has two main parts. One is the change of support

problem (CoSP) which focusses on relating the spatial variation of one variable with a given support to that of another variable with different support (Gotway & Young, 2002). The second aspect is how to use data with different support in the same model (e.g. Qu, Li, Zhang, and Wang (2012), Zhu, Carlin, and Gelfand (2003), Smith and Cowles (2007), Zacarias and Andersson (2011)). This last part is the focus of this dissertation.

1.3.1 Covariance Functions for Spatial Data

Typically one of the main goals of the analysis of spatial data is prediction of a value at an unknown location. To accomplish this, researchers use kriging, a technique that yields an unbiased, linear predictor which also minimizes the mean square error. This is known as the best linear unbiased estimator (BLUP) (see Cressie (1990) for further details). Central to using this method of estimation, is the estimation of the marginal covariance function or covariogram, $C(\cdot)$, which measures the covariance between the observations of the process being examined. For $C(\cdot)$ to corresponds to a valid covariance matrix, it needs to satisfy the following properties:

1. $|C(h)| \leq C(0)$ where h is the distance between two points
2. $C(\cdot)$ must be positive definite. That is, for any set of locations $s_1, s_2, \ldots, s_n$ and constants $a_1, a_2, \ldots, a_n$ then $\sum_i \sum_j a_i a_j C(s_i - s_j) \geq 0$
3. $\mathbf{C}(\cdot)$ must be symmetric. That is, $C(Y(0), Y(h)) = C(Y(h), Y(0))$ for any h

In some instances, there is a measurement error associated with infinitesimally small separation distances. This is referred to as the nugget effect (τ^2). In such cases, the covariance function becomes:

$$C(h; \theta, \sigma^2, \tau^2) = \begin{cases} C(h; \theta, \sigma^2), & h > 0 \\ C(0; \theta, \sigma^2) + \tau^2, & h = 0 \end{cases} \tag{1.1}$$

where $C(h; \theta, \sigma^2)$ is continuous at $h = 0$.

Covariance Functions for Geostatistical Data

Suppose $Y(s) = \mu(s) + X(s) + \epsilon(\mathbf{s})$ is a Gaussian random process where $\mu(s)$ is the mean function which may depend on one or more covariates, $X(s)$ measures the spatial dependence and $\epsilon(s)$ is an independent normally distributed error term which does not have any spatial correlation. From this, let the correlation function between two locations s and t be defined as $C(s,t) = Cov(Y(s), Y(t))$. Additionally, let h be equal to the distance between any two locations s and t. The process is considered second-order stationary if the mean is a constant (i.e. $E(Y(s)) = \mu$) and the covariance function only depends on the difference between the two locations (i.e. $Cov(Y(x), (Ys) = Cov(Y(0), Y(s-x))$. Furthermore, the process is called isotropic if the covariance function of the process only depends on the absolute distance between the two points.

One popular and flexible family of covariance functions is called the Matèrn covariance function which is only dependent on the distance between two locations ($\mathbf{h}$). This covariance function is given by

$$C(\mathbf{h}; \sigma^2, \phi, \nu) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\|\mathbf{h}\|}{\phi} \right)^{\nu} K_{\nu} \left(\frac{\|\mathbf{h}\|}{\phi} \right), \qquad \mathbf{h} \in \mathbb{R}^d, \nu > 0 \qquad (1.2)$$

where ν is the parameter which controls the smoothness of the process, ϕ is the range parameter which measures the distance at which the data are no longer correlated, σ^2 is the variance of the process, Γ is the gamma function and K_{ν} is the modified Bessel function of the second kind. Special cases of this family are:

- $\nu = 0.5$ - Exponential - $C(h) = \sigma^2 \exp -(h/\phi)$
- $\nu = 1.5$ - $C(h) = \sigma^2(1+h) \exp -(h/\phi)$
- $\nu \to \infty$ - Gaussian - $C(h) = \sigma^2 \exp -(h^2/\phi)$

Other families of covariance functions include the Powered Exponential family and the Spherical family; the latter being a covariogram family with compact support.

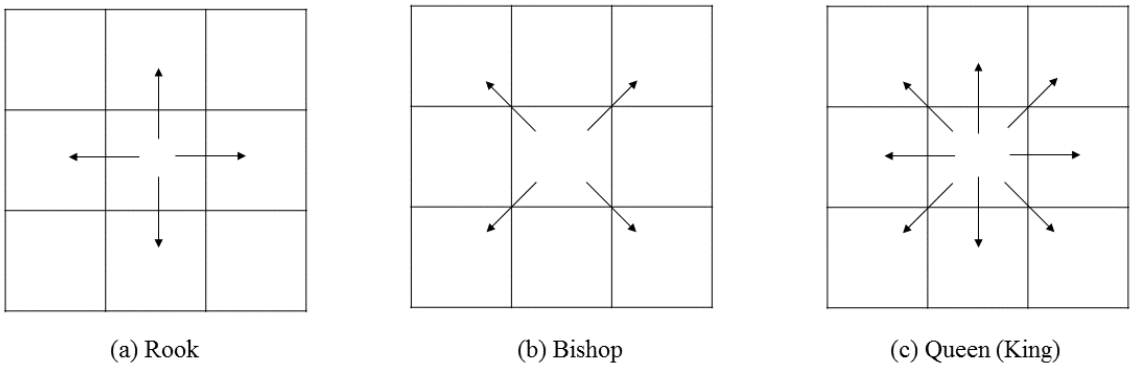
Covariance Functions for Lattice Data

In contrast, when spatial data are measured in aggregate over fixed discrete divisions (areas), models for Gaussian Markov Random Fields (GMRF) are most appropriate (Haran, 2011). Two of the most frequently used models are the conditional autoregressive (CAR) and the simultaneous autoregressive (SAR) models (Held & Rue, 2010).

The SAR model assumes that $Y \sim N(0, \sigma^2(I - \rho W)^{-1}(I - \rho W')^{-1})$ where W is the proximity matrix with w_{ij} denoting the proximity of location i to location j, ρ captures the spatial correlation, σ^2 is the variance of the process and Y is the vector of partial realisations. In this case, it is assumed that the value for any given area is dependent on its immediate neighbours. On the other hand, the CAR model assumes that the value for a given area is conditional on all the other areas in the region being studied. As a result, $Y \sim N(0, \sigma^2(I - \rho W)^{-1}D)$, where all the parameters are defined as above, and D is a diagonal matrix with $d_{ii} = \frac{\sigma^2}{w_{i+}}$ and w_{i+} is the number of neighbours of area i. Researchers tend to use the CAR model more frequently because its dependency structure is applicable to many situations (Banerjee et al., 2015a).

The proximity matrix may be chosen using methods which utilize neighbours or distances. In either case, the matrix is generally binary with $w_{ij} = 1$ if i and j are neighbours (denoted by $i \sim j$) and 0 otherwise and this relationship is symmetric. Typically for a CAR model, the proximity matrix is created based on neighbours, while the SAR model may use either method. There are three common ways of determining whether i is a neighbour of j. The rook, bishop and queen (king) methods are based on whether the neighbouring areas share a non-zero edge, a vertex (point) or both (Figure 1.1). Alternatively, if the distance between the centre of area i and area j is less than some threshold d, then i and j are considered neighbours. An extension of this is the k-nearest neighbour option, which ranks the centroid distances and the

first k are labelled as neighbours. It should be noted that area i is not a neighbour of itself using any of the above-mentioned methods.



(a) Rook (b) Bishop (c) Queen (King)

Figure 1.1. Proximity Options for Contiguous Areas

Multivariate Covariance Functions

In the case of cokriging, the BLUP is given by

$$\hat{Y}_1(s^*) = E(Y_1(s^*)) + \mathbf{k}'\boldsymbol{\Sigma}_{\mathbf{Y}}{}^{-1}(\mathbf{Y} - E(\mathbf{Y})) \tag{1.3}$$

where $E(Y_1(s^*))$ is the mean of Y_1 at location s^*, $\mathbf{k}'$ is the spatial covariance of the multivariate process at the location to be predicted and the locations that are observed. $\boldsymbol{\Sigma}_Y^{-1}$ is the precision matrix, which is the inverse of the multivariate covariance function and $\mathbf{Y}$ is the partial realization of the process.

This covariance matrix, $\boldsymbol{\Sigma}_Y^{-1}$, has two main components - the direct (marginal) covariance functions for each of the processes and the cross-covariance function which describes the covariance structure between different processes. The estimation of this cross-covariance function is one of the main concerns when cokriging is used (Guhaniyogi et al., 2013). Although the cross-covariance function may not necessarily be symmetric, and therefore not positive definite, one property which must be satisfied is that $C_{ii}(0)C_{jj}(0) \geq C_{ij}^2(h)$ for any two processes i and j.

In the case of geostatistical data, there are three main methods used to model the cross-covariance matrix. Given a variance function V, the Proportional Model, $C_{ij} = V_{ij}\rho(h)$ assumes that the direct and cross-covariance functions are all proportional to the same basic spatial correlation function (Wackernagel, 2003). A more flexible model, is the Linear Model of Coregionalization (LMC). In this case, $C(h) = \sum_{k=1}^{K} V_k \rho_k(h)$ (Goulard & Voltz, 1992; Gelfand et al., 2004). The third method is the Multivariate Matèrn (Gneiting et al., 2010) which assumes that the direct and cross-covariance functions are all in the Matèrn family (see 1.2).

The Multivariate CAR model (MCAR) is typically used to model cross-covariance functions for lattice data. As with the univariate model, the full conditional distributions are used to determine the joint distribution. Another method, Smoothed ANOVA (SANOVA) (Zhang et al., 2009) is sometimes used as the complex covariance structures needed for MCAR, may be difficult to determine based on the given dataset (Banerjee et al., 2015b). Despite this, the flexibility of MCAR, including its diverse class of models, is more attractive and is therefore usually the method of choice.

1.3.2 Nonparametric Cross-Covariance Functions

Due to the complexities involved in modelling cross-covariance functions, one approach that has been used in recent times is the semiparametric method (Wang, 2011). The main idea is to create a multivariate covariogram model using a given set of marginal models. The technique first assumes that each individual process, $\mathbf{Y}_i(s)$, is Gaussian having zero mean and a covariogram which belongs to a parametric family. Once the appropriate univariate model is determined and the estimates for each are obtained, a set of latent variables for each process is created. Finally, the nonparametric cross-covariogram is constructed using the correlation between the latent variables from each of the univariate processes. One of the key underlying assumptions of this method is that the multivariate spatial process can be approximated by

a finite weight sum of independent and identically distributed (i.i.d.) $N(0,1)$ random variables and that there is either no correlation or perfect correlation between the random variables related to each process. While this method can be computationally expensive when the dataset is large, the main advantages is that the joint covariance matrix is positive definite and that it can be used with a wide range of marginal covariance functions.

1.3.3 Change of Support (CoSP)

Innate to spatial measurements is the support, which is the spatial unit at which data are measured. This support, may be points or polygons/areas. The choice of these spatial units is solely determined by the researcher or organisation collecting the data. As a result, there are a myriad of options. For example, temperature is typically measured at specific points, while disease incidence is reported by area. There are many choices for the definition of areas and these are usually defined by convenience. As an illustration, administrative units are defined by the main agency which uses them. Therefore units such as postal codes, census tracts, police divisions, electoral divisions, and counties may not have the same boundaries and they may or may not overlap. Additionally, over time, some boundaries change based on population distribution and/or other factors. Furthermore, in some industries (e.g. mining, Matheron (1963)) data can only be measured at specific points, but conclusions are needed based on predictions of a given area. This, coupled with the difference in spatial units, require the use of methods to make predictions at different levels of support.

The CoSP phenomenon occurs across many disciplines and sub disciplines, including economics, ecology, sociology, public health and epidemiology. This has resulted in many names being used to refer to this concept (Gotway & Young, 2002). These include, but are not limited to, multiscale modelling, spatial data transformations, spatially misaligned data, ecological inference, Openshaw and Taylor's (1979) modi-

fiable area unit problem (MAUP), spatial data transformations and multiresolution modelling (Gotway & Young, 2002).

One of the main methods used to handle this type of mismatched data is to use Monte Carlo Markov Chain (MCMC) along with Kriging (Gelfand & Carlin, 2001). In this situation, the spatial process $Y(s)$ is assumed to be normally distributed and MCMC methods are used to obtain the parameter estimates for the covariogram. This technique therefore assumes that the spatial process at the point or area being predicted is conditional normal given $Y(s)$. The method has been applied to point-to-point, point-to-area, area-to-point and area-to-area predictions, even when the area, sometimes referred to as blocks, are not all of the same size and/or of irregular shape. Some of the advantages of this approach include its flexibility for isotropic and geometrically anisotropic forms and it does not require the process to be stationary. Finally, it can easily be extended to spatio-temporal models. Despite this, one constraint is that the technique requires that all locations are measured at the same set of times for spatio-temporal models.

1.4 Structure of the Dissertation

This dissertation is divided into four main parts. Chapter 2 gives a review of the change of support problem, its solutions and the methodology that has been developed to handle models which use data of different support. The third chapter, which is the main contribution of this paper, outlines the methodology used to develop the nonparametric cross-covariance matrices for data of different support, while Chapter 4 looks at the application of the methodology that I developed to forestry data. Here, the performance of the proposed method will also be compared to already existing techniques. The final chapter summarizes the advantages and limitations of the presented approach and discusses the direction of future work in this area.

2. CHANGE OF SUPPORT METHODOLOGY AND APPLICATIONS TO VARIABLES OF DIFFERENT SUPPORT

2.1 Introduction

The term *scale* has various interpretations in different arenas (Turner, 1989; Dungan et al., 2002; Wu, 2004) and is sometimes used ambiguously (Atkinson & Tate, 2000). One key meaning of *scale* is spatial extent, which refers to the area where data are collected (e.g. 1 hectare plots versus 50 hectare plots). This type of differences in scales is directly related to the modifiable area unit problem (MAUP) which often occurs in ecology (Openshaw, 1983; Jelinski & Wu, 1996; Dark & Bram, 2007). Another key meaning, and the focus of this study, is *spatial support* which refers to the shape, size/geometry (eg. volume) and orientation of measurements in a given space (Gotway & Young, 2002; Gelfand, 2010).

Issues of support can be divided into two main areas. The first aspect looks at the estimation of values for one support given that measurements are available at another level of support. An example of this is using point-level temperature data obtained from weather stations within a defined area to estimate the temperature for that area as a whole. This is referred to as the change of support problem (CoSP), which is concerned with making conclusions at one spatial scale given that the observations for the process of interest are made at another spatial scale. Despite this problem being applicable in many disciplines, it originated in mining where there is the need to identify profitable areas (blocks) to mine but due to physical and economic constraints, data is only available at specific points (Matheron, 1963; Cressie, 1990).

The second, but less studied, issue is how different variables of different support may be used to make predictions for one variable. An example is the use of climatic measures (point-level data) to assist in the prediction of health outcomes (area data).

This chapter outlines the solutions and applications for each of these two areas.

2.2 Change of Support Methodology

Change of Support (CoS) methodology looks at solutions to one of three types of predictions for different data levels. The first is aggregation, otherwise called upscaling, which deals with using finer level data to obtain values for a course area. This could include manipulating point-level to predict area values, or using small areas to predict the values for a larger area. In this case, all the small areas are contained in the larger area being predicted. Conversely, disaggregation or downscaling, is concerned with using area data to make predictions at the point level, or for smaller areas. Finally, side scaling, which is also called the overlapping units problem, focusses on obtaining predictions for one area given another in cases where the boundaries for the areas overlap.

The methods used to solve these CoSPs fall into five main categories. Some methods are able to produce results for all three types of data changes, while others are designed to only solve a subset of these issues. When considering solutions to the CoSP, the following 10 criteria should be taken into consideration (Gotway Crawford & Young, 2005). The methods should:

1. Explicitly account for different levels of spatial support;
2. Be able to be utilised for different requirements such as upscaling (aggregation), downscaling (disaggregation) or side-scaling (overlapping units);
3. Result in smooth predicted surfaces across the boundaries for the units being predicted;
4. Produce accurate estimates of the standard error of the predictions;
5. Be able to incorporate covariates to improve predictions;

6. Be able to be used with different types of data - continuous or discrete (e.g. counts, averages, binary data);

7. Produce predictions which lie in the parameter space. That is, the resulting predictions should not be negative if the values should only be positive;

8. Be consistent across scales. Specifically they should satisfy the mean balance property (Huang et al., 2002) or the pycnophylactic property (Tobler, 1979), depending on whether the data is being aggregated or disaggregated. For example, given a series of areas, the predictions for a given area should sum to the observed point data observed within that area (the mean balance property);

9. Be based on the distributional assumptions and the paucity of the model;

10. Be reasonable from a computational perspective.

It should be noted though, that solutions may not be able to achieve all the aforementioned criteria, but they should aim to attain as many as possible.

Additionally, it is also important to note that the shape of the sampling distribution density curve for point data and area data are different (Matheron & Kleingeld, 1987) therefore aggregation, smoothing and/or averaging of data increases spatial homogeneity which can modify the size and/or direction of associations (Cressie, 1996; Wang et al., 2016; Shafran-Nathan et al., 2017).

There have been a wide range of techniques that have been proposed as solutions to the CoSP. Some can be grouped into categories, while others are stand alone methods. The review that follows, examines five of these methods.

2.2.1 Trivial Solutions

According to (Gelfand, 2010), there are two main methods that fall into this category. The first is what has been termed *block averages*, where the average of the values for all of the data at given points, $Y(s_i)$, which are located in a given

block, B_i, are averaged to represent the block value. Although this method is both computationally efficient and easy to implement, it requires that at least there is one point in each block. Additionally, it ignores the spatial process of the observations in and outside of the block of interest.

Another method is to use the point data, $Y(s_i)$, within a given block, B_i, to predict the value at the centre of the block. This is then used to represent the block value. While this method uses spatial correlation to predict the centroid and is still computationally efficient, the value obtained may not only be biased for the block average, but it has a larger variability.

2.2.2 Smoothing

Spatial smoothing, also called surface methods (Fisher & Langford, 1995), are concerned with area-to-point predictions, where given block values, $Y(B_i)$, predictions for points, $Y(s_i)$, can be obtained. This is typically used with socio-economic and population variables, as individual point information is not always available due to privacy laws. Despite some of the constraints outlined below, these methods are generally easy to implement and to obtain measures of uncertainty, while also being constrained to be aggregate consistent. Additionally, they explicitly consider support and some allow for the inclusion of covariates (Brillinger, 1990; Müller et al., 1997). It should be noted that one of the disadvantages of all these methods, is that it is not suitable for upscaling or side-scaling.

Choropleth map

The choropleth map is the simplest method for addressing area-to-point change of support, as it is typically implemented in geographical information systems (GIS). The method is ideal if the the spatial distribution of the point values is unknown. Despite this and the use of spatial correlation at the area level, failure to assume the dependence of point values is problematic (Kyriakidis, 2004).

Support adjusted locally weighted regression

Support adjusted locally weighted regression is a strategy which explicitly accounts for the differences between areal and point support. The main assumption is that there is spatial areal independence but there is spatial correlation between predicted points within a given block (Brillinger, 1990, 1994; Müller et al., 1997), while also assuming that the process varies smoothly across blocks. Due to the areal independence but point dependence assumption, this method may not yield aggregate consistent predictions (Kyriakidis, 2004).

Kernel smoothing

A related technique to support adjusted locally weighted regression is kernel smoothing, which uses intercentroid analysis and is typically applied to the prediction of population estimates (Bracken & Martin, 1989; Martin, 1996). One criticism of this technique is the main assumption that, for a given area, the value can be collapsed into point data, which is typically located at the centre of the area (Kyriakidis, 2004).

Pycnophylactic Interpolation

Laplacian smooth pycnophylactic interpolation (Tobler, 1979) is a method of area-to-point spatial interpolation where the estimated point values are determined by Laplace's partial differential equations and some predetermined boundary conditions. The general method is to superimpose a fine grid of equally spaced points over the area of interest and the values of the density at each point are calculated. Once these are obtained, Dirichlet's integral is used as the smoothing function. This integral imposes the pycnophylactic and non-negativity constraint.

Although this method ensures that the predictions are aggregate consistent, this property makes it difficult to adjust for covariates and to obtain valid measures of

uncertainty. It should be noted though that the Brillinger version of this method may not satisfy the pycnophylactic property. Additionally, the method uses a regular grid of predicted locations but this has been relaxed by (Rase, 2001). Finally, pycnophylactic interpolation is effective when it is reasonable to assume that the variable is smooth over space but is not appropriate when the variable is abruptly discontinuous in spatial distributions, as often occurs with social and economic data (Flowerdew & Green, 1992).

2.2.3 Regression Methods

This technique is typically used when data are collected on one areal unit, called source zones, but predictions are required at another areal unit, referred to as target zones, which are not necessarily aligned with the source units. The boundaries of both sets of zones therefore do not necessarily meet, for example postcodes and wards in Preston, Lancashire, UK (Flowerdew & Green, 1992) and counties and hydrologic basins (Goodchild et al., 1993).

The method assumes a regression model for ancillary variables and the variable of interest, so as to obtain better estimates from areal interpolation, while treating data for the variable that is to be predicted as missing variables (Flowerdew & Green, 1989, 1992). These predicted values are obtained using an iterative process. In order to reasonably estimate the relationship between the ancillary variables and the one to be predicted, an appropriate probability distribution must be selected for the target variable. Additionally, the pycnophylactic constraint must be satisfied. This procedure has been applied to count data (Flowerdew & Green, 1989, 1994), binomial data (Flowerdew et al., 1991; Green, 1990) and continuous variables (Flowerdew & Green, 1992). This last application can require the use of the EM algorithm to obtain the predicted values, as in the case of housing prices (Flowerdew & Green, 1992) and may require the identification of one or more acceptable underlying continuous surfaces (Goodchild et al., 1993).

This method has the advantage of being able to accommodate covariates and different types of variables, whether they be discrete (count or binomial) or continuous. Moreover, the technique is computationally efficient and relatively easy to implement. However, due to the pycnophylactic constraint and the iterative process, accurate measures of uncertainty are not produced (Gotway Crawford & Young, 2005). Additionally, neither the support of the units nor the correlation between the areas within a zone are considered.

2.2.4 Multi-Scale Tree Models

Multi-scale tree models, classed as hierarchical models, views different resolutions as nested, where one resolution, or level of support (a parent) is associated with several children which are at smaller levels of support (Wikle, 2003). Therefore each level of this system, called a tree, corresponds to a different resolution or spatial scale. The algorithm is based on dynamic models defined on a hidden multivariate Gaussian tree-structured process (Chou et al., 1994) which is a suitable alternative to other hierarchical models when the volume of the density of the spatial data and/or the size of the target zone is so large that standard techniques are not suitable (Wikle, 2003). The technique is based on the Kalman filtering algorithm (also called linear quadratic estimation (LQE)) which uses time series observations that have statistical noise and other inaccuracies to obtain estimates of unknown variables. The adaptation of this technique to spatial problems has been termed change-of-resolution Kalman filtering (Chou et al., 1994) and can sometimes be referred to as multi-resolution spatial models (MRSM) (Johannesson et al., 2007).

The tree structure (Figure 2.1) is defined based on the centroid of a resolution where each parent node has children. If a node has no children, then it is called a leaf, while if no children spawned a node, then it is called the root. Based on these definitions, the algorithm has two steps (Chou et al., 1994). The first is the uptree (leaves to root) filtering step where the optimal predictor of a specific node is

computed recursively based on the data observed both at that node and any of its descendants. It therefore results in the root being estimated based on all available data. The second step is the downtree (root to leaves) smoothing. This part of the method recursively computes the optimal predictor for each node from the root to leaves based on all of the data.

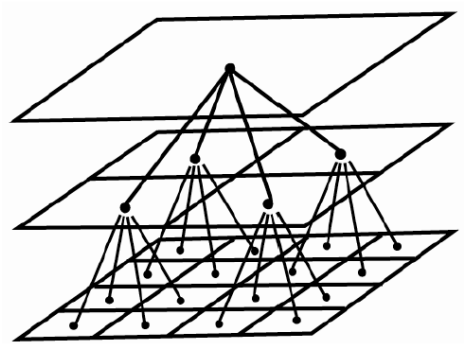


Figure 2.1. Spatial Tree Structure (Source: Gotway & Young, 2002)

The main idea is to assume that the observed data, $\mathbf{Z}(s)$, are related to the vector to be predicted, $\mathbf{X}(s)$, also called the state vector, by (2.1).

$$\mathbf{Z}(s) = K(s)\mathbf{X}(s) + \epsilon(s) \tag{2.1}$$

where $K(s)$ is a $n \times m$ deterministic selection matrix that relates the observed measurements to those to be predicted, while $\epsilon(s)$ is white noise independent of $\mathbf{X}(s)$. The state vector, $\mathbf{X}(s)$, is then assumed to be related to its parent through another deterministic function $\Phi(s)$ (2.2)

$$\mathbf{X}(s) = \Phi(s)\mathbf{X}(p_s) + \eta(s) \tag{2.2}$$

where $\mathbf{X}(p_s)$ is the parent node of $\mathbf{X}(s)$ and $\eta(s)$ is white noise that is independent of $\mathbf{X}(p_s)$.

The method has been extended to satisfy the mass balance property and accommodate nonstationary spatial dependence (Huang et al., 2002) and to include temporal components at a course resolution(Johannesson et al., 2007). Areas of application

include modelling of ocean surface height (Fieguth et al., 1995), image restoration (Banham & Katsaggelos, 1996) and remote sensing (Huang et al., 2002; Simone et al., 2002).

The main appeal of spatial tree models is its computational efficiency due to the recursive nature of the Kalman filter which means that the order of computation is proportional to the number of nodes at the finest level (Gotway & Young, 2002; Huang et al., 2002). Furthermore it incorporates spatial dependence while also providing measures of uncertainty associated with the predictions. Despite its attractiveness, it does not explicitly account for changes in support that may occur with changes in resolutions (Gotway & Young, 2002) and statistical parameter estimation may be difficult (Gotway Crawford & Young, 2005). Finally, there has been no extension to cases where the spatial scales may overlap.

2.2.5 Geostatistical Methods

Overall, Geostatistical methods have proven to be quite helpful especially when profitability is of key importance (Cressie, 1991) (e.g. the mining industry). These methods are so named because they focus on modelling the covariance function for the support of interest.

Block Kriging

Block kriging is one of the most used methods for point-to-area predictions (Burgess & Webster, 1980; Carroll et al., 1995; Kern & Coyle, 2000; Cressie, 2006; Young et al., 2008; Keshavarzi et al., 2011; Ross et al., 2013; Wang et al., 2014; Kang et al., 2017). The procedure uses data which have been observed at specific points to predict the average of the process at the area level while accounting for shape, size and the orientation of the blocks (Journel & Huijbregts, 2003; Chilès & Delfiner, 2012).

Assume that data are available for a specific variable $Y(\mathbf{s})$ at a set of known locations, where $\mathbf{s}$ varies continuously over a given spatial domain $\mathcal{D}$. Also assume

that $Y(\mathbf{s})$ has mean $\mu(s)$ and covariance $Cov(Y(s_i), Y(s_j)) = C(s_i, s_j)$ and that there is an area $\mathbf{B}$ which has a given volume which is also a subset of the domain $\mathcal{D}$. $Y(\mathbf{B})$ can then be written in terms of point data such that

$$Y(\mathbf{B}) = \frac{1}{|B|} \int_B Y(\mathbf{s}) ds \tag{2.3}$$

and

$$Cov(Y(B_i), Y(B_j)) = \frac{1}{|B_i|} \frac{1}{|B_j|} \int_{B_i} \int_{B_j} Cov(u, v) du dv \tag{2.4}$$

The point-to-area covariance is therefore given by

$$Cov(Y(B_i), Y(s_j)) = \frac{1}{|B_i|} \int_{B_i} Cov(u, v) du \tag{2.5}$$

with universal block kriging predictor equal to

$$\hat{Y}(B_i) = \sum_j^n \lambda_j Y(s_j) \tag{2.6}$$

where λ_j are the optimal weights which minimise the prediction mean square error and $s_j \in B_i$ for $j = 1, \ldots, n$.

Overall this method is flexible as it can accommodate different covariance models and different methods of parameter estimation. Furthermore, it can be extended to area-to-area kriging. Despite these advantages, it can become computationally expensive as the technique involves the inversion of large matrices which is of $O(n^3)$. It is also is only applicable for upscaling.

Disjunctive Kriging (DK)

Although other change of support models exist (eg. affine correction and log-normal models (Journel & Huijbregts, 1978; Matheron, 1978; Journel, 1980; Isaaks & Srivastava, 1989 in Machuca-Mory, Babak, and Deutsch (2008))), the disjunctive kriging (DK) model is the most frequently used. Disjunctive kriging is a technique

which uses a bivariate probability model to estimate a given function (Ortiz et al., 2005). If a specific bivariate assumption is used, this can be achieved through the use of isofactorial models which are constructed using factorial decompositions of transformed distributions into polynomials with an orthonormal base (Machuca-Mory et al., 2008). Due to the fact that the resulting random variable are uncorrelated, the optimal estimate is obtained by simple kriging each component (Ortiz et al., 2005). This is a distinct advantage of this technique. This method is typically used to solve the point-to-block and the block-to-block change of support scenarios and can be used for different types of variables (discrete and continuous).

Isofactorial models, as presented in (Wackernagel, 2003) model the bivariate distribution of any two points which lie in the domain of a stationary random function. Although they were developed in Quantum Mechanics, they have been used by statisticians for Markov processes and specifically in the field of Geostatistics since 1973 (Matheron, 1984). The model is given by (2.7)

$$F(dy_x, dy_{x+h}) = F(dy_x)F(dy_{x+h}) \sum_{k=0}^{\infty} T_k(\mathbf{h})\chi_k(Y_x)\chi_k(Y_{x+h}) \tag{2.7}$$

where $T_k(\mathbf{h})$ is the correlation function between a pair of points; χ_k are the orthonormal polynomials; $F(du, dv)$ is the symmetric bivariate distribution; and $F(du)$ is the marginal distribution given by $\int_v F(du, dv)$.

There are three types of isofactorial models - two extreme models and one intermediate model. The extreme models are the diffusion type, which are used for processes which gradually change from one location to another, and the mosaic type, which are used in situations where within a given block the process is relatively constant but exhibits sudden changes between blocks. There are also two intermediate models which allow for both diffusion and mosaic type conditions in one model. These are the Barycentric and Beta models. The Barycentric model includes a measure of dissemination ($\beta \in (0, 1)$), where 0 is the diffusion model and 1 is the mosaic model.

The general method has five main steps (Machuca-Mory et al., 2008; Ortiz et al., 2005):

1. Transform the original data using the appropriate score transformation (e.g. if transforming to Gaussian, use the normal score transformation);

2. Fit the variogram model for the original data and calculate the variance of the block support, $\bar{\gamma}(v, v)$, which is the average of the variogram within the block;

3. Fit the distribution of the transformed data using the relevant polynomials (Table 2.1 (Wackernagel, 2003));

4. Calculate the new distribution for the block support in terms of the relevant polynomials;

5. Transform the polynomial-expanded distribution for the block support back to the original units.

In the case of the Barycentric model, the estimate of β, the effective measure of dissemination (β_{eff}), is calculated between Steps 1 and 2.

Table 2.1.
Isofactorial Models and Related Polynomials

Type of Model	**Transformation**	**Related Polynomials**
Diffusion	Gaussian	Hermite
	Gamma	Laguerre
	Beta	Jacobi
Mosaic (infinite # of states)	Poisson	Charlier
	Negative Binomial	Meixner
Mosaic (finite # of states)	Binomial	Krawtchouk
	Jacobi	Discrete Jacobi
	Anti-Jacobi	Anti-Jacobi

Unfortunately, due to the complexity of the technique and difficulty of the related literature (Ortiz et al., 2005; Verly, 1983), disjunctive kriging is not used as often as

other methods. Additionally, these methods assume a fixed-size unit of prediction, which is not practical. Also, using the orthonormal expansion is known to have related convergence issues (Verly, 1983). Conversely, this method allows for the prediction of non-linear functions and reduces the computational requirement because predictions of each polynomial are executed using separate kriging systems. Moreover, measures of uncertainty are also generated by this technique (Gotway & Young, 2002).

Other Non-Linear Geostatistical Methods

Along with disjunctive kriging, there are other non-linear methods which assist in solving the change of the support problem. Although opting to look for the best estimator (minimum variance estimator) within both the linear and non-linear families will result in a gain in accuracy, it should be noted that the use of non-linear estimators is accompanied with more complex execution and stronger requirements (Journel & Huijbregts, 2003). Most of the methods in this category focus on obtaining the required estimates by modelling the conditional cumulative distribution function (ccdf) using either a parametric or nonparametric approach (Goovaerts, 2001).

The first of these methods is the multi-Gaussian approach where point data is transformed into Gaussian data and then the attractive properties of the conditional normal distribution are used to obtain estimates. In this procedure, the blocks to be predicted are discretized into points, u'_j, and the block value is approximated using $Z(B) \approx \frac{1}{N}\sum_{j=1}^{N} Z(u'_j)$. The point data values are then transformed into Gaussian variables using the relationship $Z(s) = \phi Y(s)$, where ϕ is based on the normal score transformation. Then the ccdf is estimated using simulation and the fact that

$$F_B(z|\mathbf{Z}) \approx P(\frac{1}{N}\sum_{j=1}^{N} Z(u'_j) < z)|Z(s_1), \ldots, Z(s_n) \tag{2.8}$$

$$\approx P(\sum_{j=1}^{N} \phi(Y(u'_j)) < Nz|Y(s_1), \ldots, Y(s_n) \tag{2.9}$$

where $F_B(z|\mathbf{Z})$ is estimated as the proportion of vectors satisfying $\sum_{j=1}^{N} \phi(Y(u'_j)) < Nz$. Once these are obtained, the ccdf is back-transformed to obtain the ccdf of the original variable (Goovaerts, 2001; Gotway & Young, 2002). The multi-Gaussian approach is appealing because it is easier to execute than disjunctive kriging while providing consistent solutions. Furthermore, in a simulation of ore deposits, it provided slightly better precision than other non-linear methods (Verly, 1983). Alternatively, despite the non-linear nature of estimation, the values are averaged linearly in space, which may not be a reasonable transformation for some variables (e.g. hydraulic conductivity). Additionally the Gaussian assumption may not be valid for all variables (e.g. variables in soil science). Finally, there may be the need to allow for spatial dependence at very small or very large values (e.g. soil pollution (Goovaerts, 1999) which is not accommodated by this technique. All these limitations are discussed in further detail by (Goovaerts, 2001).

Another non-linear approach has been called indicator kriging, which builds on the methodology from the multi-Gaussian method but instead focusses on characterising the spatial variability of indicator functions. First, the block is discretized and $Z(u'_j)$ is simulated at each node. The simulated block values are then approximated using the same aggregation technique utilised in the multi-Gaussian approach. Each observation is then transformed into a set of K indicator variables which correspond to K threshold values such that $I(Z(B_1) \leq z), \ldots, I(Z(B_n) \leq z)$ where

$$I(Z(B) \leq z) = \begin{cases} 1 & \text{if } Z(B) \leq z \\ 0 & \text{otherwise} \end{cases} \tag{2.10}$$

The K ccdf values are then estimated by kriging the indicator data and the final result is calculated by interpolating or extrapolating the estimated probabilities (Journel, 1983; Goovaerts, 2001; Gotway & Young, 2002). This technique has been used in estimating the spatiotemporal distribution of hydrogen-ion deposits (Bilonick, 1988) and developing map classification schemes (Solow, 1986).

One advantage of this method is that there is no assumption about the underlying distribution of the data. However, the technique tends to yield worse approximations to the conditional expectation than its disjunctive kringing counterpart and the system of kriging equations can be large (Cressie, 1991).

Area-to-Point (AtP) Kriging

Kyriakidis (2004) developed a statistical smoothing method called Area-to-Point (AtP) kriging. It has been applied to soil science (Goovaerts, 2010; Schirrmann et al., 2012; Kerry et al., 2012), medical geography (Goovaerts, 2010), crime data (Kerry et al., 2010) and hedonic pricing models (Yoo & Kyriakidis, 2009).

AtP kriging is the counterpart of block kriging where area data are assumed to be a linear combination of point measurements within the areas. In AtP kriging, the system of normal equations uses the average between-block semivariances and the block-to-point semivariances which are calculated from the discretized points. To do this, the point semivariogram needs to be estimated using an iterative semivariogram deconvolution (Journel & Huijbregts, 2003; Goovaerts, 2008) since only the variogram for the area data is known. The deconvoluted point semivariogram is then regularized and used to obtain the AtP kriging. This method maintains pycnophylatic property while also allowing for the inclusion of covariates. Despite these properties, model accuracy decreases when the subsamples are not regularly distributed over the sampling areas (Schirrmann et al., 2012).

Bayesian Hierarchical Models (BHMs)

A Bayesian alternative to kriging was first introduced by Le and Zidek (1992) where analysis was restricted to linear interpolation because, in this case, the approach depends on the prior distributions only through the first and second moments. The proposed method followed the basic structure of Bayesian analysis, where in the first level of the hierarchy, the spatial covariance function is left unspecified and the

uncertainty about this function is accounted for in the second level through the priors. If necessary, additional levels can be incorporated into the analysis. This approach though, did not account for differences in support and focussed on the use of Gaussian variables, therefore resulting in a Gaussian joint distribution.

Due to the proliferation of ecological research and GIS information being recorded using different regional boundaries, there is a demand for methods which allow researchers to use information from misaligned sources with non-Gaussian data. The Bayesian hierarchical approach lends to such situations due to its flexibility especially as it relates to the choice of likelihoods and priors. The technique has therefore been applied to population interpolation over misaligned regions (Mugglin & Carlin, 1998) and the prediction of surface wind fields using simulated data and geographic models (Berliner et al., 2003). It has also been extended to include a temporal component (Wikle et al., 2001; Gelfand & Carlin, 2001) which can also incorporate point referenced and areal data types of mismatches (Wikle & Berliner, 2005).

According to (Gelfand & Carlin, 2001), if $Y_S' = (Y(s_1), \ldots, Y(s_n))$ in a continuous spatial process, $S \in \mathcal{D}$, observed at locations s_i, $i = 1, \ldots, n$ then $Y_S'|\beta, \theta \sim N(\mu_s(\beta), H_s(\theta))$, where β is the mean trend and θ is the parameters for the stationary covariance function. Given that predictions, $Y(B)$, are needed at the areal level $B \in \mathcal{D}$ but not necessarily a subset of S, then $Y_B|Y_S, \beta, \theta \sim N\left(\mu_B^*(\beta), H_B^*(\theta)\right)$ where

$$\mu_B^*(\beta) = \mu_B(\beta) + H_{s,B}'(\theta)H_s^{-1}(\theta)\left(Y_s - \mu_s(\beta)\right) \tag{2.11}$$

$$H_B^*(\theta) = H_B(\theta) - H_{s,B}'(\theta)H_s^{-1}(\theta)H_{s,B}(\theta) \tag{2.12}$$

A set of locations, $l = 1, \ldots, L_k$ are then selected uniformly and independently for each block B_k to predicted. Once priors are selected for β and θ, $\hat{f}(Y_B|Y_s, \beta, \theta)$ is calculated using Monte Carlo integration and samples taken from this distribution. Finally, $\hat{Y}(B_k) = \dfrac{1}{L_k}\sum_l Y(s_{kl})$. This can also be extended to the spatiotemporal case.

This method is attractive because it uses solid statistical theory while being able to accommodate more complex models which account for different levels of support. It is also easily extended to spatio-temporal models. Additionally, it is flexible for

isotropic and geometrically anisotropic forms and it does not require the process to be stationary. Furthermore, it has the advantage of the resulting posterior predictive distribution being able to produce a complete description of uncertainty. Alternatively, the BHM technique can be computationally intensive and most applications examine solving one change of support problem at any given time (exception - Wikle and Berliner (2005)). Additionally, the priors for the parameters need to be carefully chosen. It should also be noted that this technique requires that all locations are measured at the same set of times for spatio-temporal models.

2.3 Combining Data of Different Support

Although variables are correlated, they may not be measured using the same spatial units due either to the type of data or the organization collecting the data. In such cases, it is necessary to use methods which take account of these differences so as to obtain more accurate predictions.

The AtP krging methodology outlined in Section 2.2.5 has also been used to obtain predictions when the support of the variable to be predicted and the covariates differ. One example is the prediction of soil nitrogen (point-level data) using land use (area data) (Qu, Li, Zhang, Wang, M, et al., 2012).

BHMs (Section 2.2.5) and Bayesian hierarchical regression have been used to obtain predictions from models which have data of different support. The BHMs follow the same method outlined in Section 2.2.7 and has been used in the prediction of uranium measurements at the point level using areal uranium measurements and point-level radon measurements (Smith & Cowles, 2007).

Bayesian hierarchical regression has especially been used in studies where health outcomes (usually area data) are correlated with climatic or environmental measures (usually point data). Examples of this include the incidence of malaria in Mozambique (Zacarias & Andersson, 2011), pædiatric emergency room visits for asthma (Zhu et al., 2003) and mortality (Fuentes et al., 2006). Using the example from Zhu et al.,

the ozone data, measured at points, are realigned to the zip code blocks (the ER visits data) using BHMs and then Bayesian hierarchical regression is used to obtain predictions for ER visits with ozone values as a covariate.

2.4 Summary

In summary, based on the criteria suggested by (Gotway Crawford & Young, 2005), not all methods posses all the requirements. In reality, this may not be possible and so the trade-off is to use the best method for the data given, as each method is suitable once certain conditions are met and therefore are still of merit. For example, BHMs and disjunctive kriging are elegant solutions to the CoSP but their complexity and dependence on many precise assumptions may reduce their attractiveness once those assumption cannot be justified (Gotway Crawford & Young, 2005).

Additionally, BMHs and Bayesian hierarchical regression have been used to solve the problem covariates with different spatial units from the one to be predicted. Despite the attractiveness of these methods, the careful choice of priors is required and the models may also suffer from convergence issues.

It is believed that while adding to the methodology available for combining data of different support, the proposed semiparametric solution could be an attractive Geostatistical method as it could reduce the complexity of implementation, produce measures of uncertainty, while still taking account of the spatial correlation and the levels of support between two or more related variables.

3. NONPARAMETRIC CROSS-COVARIANCE FUNCTIONS

Cross-covariance functions, which measure the relationship between two or more spatial processes, are generally hard to construct due to the fact that the corresponding covariance matrix needs to be symmetric and positive definite. To guarantee these properties, researchers gravitate towards using parametric covariance functions to model the cross-covariance matrices. Examples of these models include linear model of coregionalization (Goulard & Voltz, 1992; Gelfand et al., 2004), separable cross-covariance function (Mardia & Goodall, 1993), kernel convolution (Ver Hoef & Barry, 1998) and covariance convolution (Gaspari & Cohn, 1999; Majumdar & Gelfand, 2007). Typically, these models have been applied to data with the same support. For data of different support, hierarchical models have been used to create cross-covariance matrices for data of different support (Le & Zidek, 1992; Mugglin & Carlin, 1998; Gelfand & Carlin, 2001; Wikle et al., 2001; Zhu et al., 2003; Zacarias & Andersson, 2011). Although these models are useful, they make restrictive assumptions and are therefore not suitable in all cases.

As an alternative, I propose building a nonparametric cross-covariance function which is potentially more attractive as it does not depend on the hierarchical structure of the data. More specifically, the boundaries for one process do not need to be aligned with the boundaries for the other process being examined. Additionally, with the proposed model, the continuous process does not have to be conditioned on the areal process as required by hierarchical models. This is especially useful in cases where the scientific interpretation is based on the areal process being conditioned on the continuous process. The value of the proposed method also increases when the relative flexibility and computational ease is taken into consideration since only the marginal covariance functions are used in the construction of the cross-covariance matrices.

In this chapter, the method for constructing nonparametric cross-covariance matrices is outlined, along with an example. The formulas for prediction and cross-validation are also discussed.

3.1 Construction of Nonparametric Cross-Covariance Matrices for Spatial Models

To facilitate the construction of cross-covariance matrices for spatial models, it is necessary to first obtain the marginal covariance matrix for each process being studied. From these matrices, the cross-covariance matrix is constructed.

3.1.1 Marginal Models

Once marginal covariance matrices are created, cross-covariance functions may be constructed for a combination of two or more stochastic processes and/or random fields. Irrespective of the type of process, the covariance matrix of the marginal model is always symmetric and positive definite. Given these properties, any covariance matrix, $\mathbf{\Sigma}$, can be expressed as (3.1).

$$\mathbf{\Sigma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}' = \mathbf{Q}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{Q}' \tag{3.1}$$

where $\mathbf{Q}$ is an orthonormal matrix and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues. The covariance matrix can therefore be expressed as (3.2).

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{A}' \tag{3.2}$$

where $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}^{\frac{1}{2}}$. From this, the partial realisation of a spatial process $\mathbf{Y}$ with zero mean and covariance $\mathbf{\Sigma}$ may be written in terms of $\mathbf{A}$ and vector of latent $N(0,1)$ variables $\mathbf{Z}$. The spatial process can therefore be written as (3.3)

$$\mathbf{Y} = \mathbf{A}\mathbf{Z} \tag{3.3}$$

where $\mathbf{A}$ is a $n \times n$ matrix, $\mathbf{Z}$ is a $n \times 1$ matrix and n is the number of observations for the given process. One of the advantages of expressing the covariance of the process

in this form is that if n is very large, then the first q latent variables and their columns of $\mathbf{A}$ can be used for computational ease, where q is chosen to be much less than n. The first q are chosen since they relate to the first q ordered eigenvalues. In that case, (3.3) is an approximation which improves with larger values of q.

The Truncated Karhunen-Loéve Expansion

For continuous processes,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \sum_{j=1}^{\infty} a_j(s) Z_j + e(s) \tag{3.4}$$

and given n observations,

$$\mathbf{Y} = \sum_{i=j}^{q} a_j Z_j + e. \tag{3.5}$$

The $a_j(s)$ and Z_j components in (3.4) may be constructed using different methods including the convolution method (Hidgon, 1998), fixed rank kriging (Cressie & Johannesson, 2008), and Predictive Processes (Banerjee et al., 2008). However, the Karhunen-Loéve expansion (Karhunen, 1946; Loève, 1955), which is based on latent dimensions, produces optimal results when compared to the aforementioned methods (Hu, 2013). The method is sometimes referred to using different names such as Principal Component Analysis (Hotelling, 1933), Empirical Component Analysis (Lorenz, 1956), Empirical Eigenfunction Decomposition (Sirovich, 1987) and Singular Value Decomposition (Golub & Loan, 1996).

Using the Karhunen-Loéve expansion, the process $Y(\mathbf{s})$ is

$$Y(\mathbf{s}) = \mathbf{A}(s)\mathbf{Z} \tag{3.6}$$

where $\mathbf{Z}_i$ is a vector of q i.i.d. $N(0,1)$ random variables and $\mathbf{A}_i$ is an $n \times q$ matrix with the terms in the i^{th} row and j^{th} column equation to $\sqrt{\lambda_j} f_j(\mathbf{s}_i)$. More specifically, the $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} \sqrt{\lambda_1} f_1(\mathbf{s}_1) & \sqrt{\lambda_2} f_2(\mathbf{s}_1) & \cdots & \sqrt{\lambda_q} f_q(\mathbf{s}_1) \\ \sqrt{\lambda_1} f_1(\mathbf{s}_2) & \sqrt{\lambda_2} f_2(\mathbf{s}_2) & \cdots & \sqrt{\lambda_q} f_q(\mathbf{s}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\lambda_1} f_1(\mathbf{s}_n) & \sqrt{\lambda_2} f_2(\mathbf{s}_n) & \cdots & \sqrt{\lambda_q} f_q(\mathbf{s}_n) \end{bmatrix} \tag{3.7}$$

As mentioned previously, this ability to represent the spatial process using a set of q latent variables, where $q \ll n$, is advantageous as it allows for the analysis of large datasets because if inversion is required, the order of operations and memory are reduced from $O(n^3)$ and $O(n^2)$ respectively.

3.1.2 Nonparametric Cross-Covariance Functions

Nonparametric cross-covariance functions have been successfully developed and used to improve prediction for data with the same support by (Wang, 2011; Hu, 2013). The idea is that once the covariance functions for each univariate process are obtained and the process expressed in the form of (3.3), the cross-covariance function is constructed using the correlation between the latent variables from each of the univariate processes.

Suppose we have an observation vector $\mathbf{Y} = (\mathbf{Y}_1', \mathbf{Y}_2', \ldots, \mathbf{Y}_p')$ where $\mathbf{Y}_i$ is the vector of observations for the i^{th} variable and

$$\mathbf{Y}_i = \mathbf{A}_i \mathbf{Z}_i \tag{3.8}$$

where $\mathbf{A}$ and $\mathbf{Z}$ are determined as outlined in Section 3.1.2. Of note, if $\mathbf{Y_i}$ is a stochastic process, then a nugget effect needs to be added to avoid singularity of the overall cross-covariance matrix. Therefore for a continuous process, $\mathbf{Y_i}$ in (3.8) becomes

$$\mathbf{Y}_i = \mathbf{A}_i \mathbf{Z}_i + \tau_i \mathbb{1}. \tag{3.9}$$

It can be shown that the cross-covariance between the partial realisation of any two spatial processes, $\mathbf{Y}_i$ and $\mathbf{Y}_j$ $i \neq j$, is

$$\begin{aligned} Cov(\mathbf{Y}_i, \mathbf{Y}_j) &= \mathbf{A}_i Cov(\mathbf{Z}_i, \mathbf{Z}_j)\mathbf{A}_j' \\ &= \mathbf{A}_i \mathbf{J}_{ij} \mathbf{A}_j' \end{aligned} \tag{3.10}$$

Consequently, there are ${}_pC_2$ $\mathbf{J}_{ij}$ matrices. In implementations by Wang (2011) and Hu (2013), it is assumed that $\mathbf{J}_{ij}$ is a $q \times q$ matrix, where q is the length of the Z_i vector, which is the same length for each process. It is also assumed that the relationship between any two pairs of latent variables from the two different processes is perfectly correlated or not correlated at all. As a result, $\mathbf{J}_{ij}$ is a matrix of 1's and 0's. The log likelihood function for the joint process is used to determine whether which pairs are perfectly correlated. That is, find $\mathbf{J}_{ij}$ such that

$$\mathbf{J}_{ij} = \arg\max_{\mathbf{J}_{ij}} l(\mathbf{J}) \tag{3.11}$$

where l is the log likelihood function for the joint process. However, for greater flexibility, in the method proposed here, $\mathbf{J}_{ij}$ is no longer assumed to be square, but a $q_i \times q_j$, where q_i is assumed to be the length of $\mathbf{Z}_i$ and q_j is assumed to be the length of $\mathbf{Z}_j$.

The progressive search algorithm was found to give the best predictive performance when used to identify which pairs are perfectly correlated (Wang, 2011). The procedure begins with a $q_1 \times q_2$ zero matrix and searches the entire matrix for the pair that maximises the log likelihood function in (3.11). Once the pair which satisfies this condition is identified, a 1 is placed in the cell and the corresponding column and row are removed from the search space. The algorithm then repeats the process for each iteration until all the correlated pairs are found. It should be noted that there can only be one 1 in any given row or column and the maximum number of 1s is the $min(q_i, q_j)$. Therefore,the total number of iterations is $c = min(q_i, q_j)$. In this case, a total of $\sum_{k=0}^{c-1}(q_i \times q_j) - (q_i + q_j)(k-1)$ cells are searched and k is the number of 1's already allocated. In order to improve this, the search for each correlated pair

which maximises the likelihood within each iteration is computed in parallel. An alternate algorithm searches each row successively for the pair which maximises (3.11). This greatly reduces the search space since in any given iteration, only one row is searched and not all the possible positions for a 1 in the entire matrix (Hu, 2013). This alternate algorithm yields an approximately maximised $\mathbf{J}_{ij}$.

3.1.3 Example with Data of Different Support

Without loss of generality, assume that there are two processes that are known to be correlated, but each is measured at a different level of support. Also assume that both processes have zero mean. Let $\mathbf{Y}_1 = (Y_{11}, Y_{12}, \ldots, Y_{1m})'$ be the realisation of a areal/block Gaussian Markov Random Field (GMRF) at m different areas. If the marginal covariance matrix, Σ_{11} is written as in (3.2), then

$$\mathbf{Y}_1 = \mathbf{A}_1\mathbf{Z}_1 \tag{3.12}$$

Also, let $\mathbf{Y}_2 = (Y_{21}, Y_{22}, \ldots, Y_{2n})'$ $n \neq m$ be the realisation of a zero mean point-level Gaussian process. The model is therefore written as in (3.9) to give

$$\mathbf{Y}_2 = \mathbf{A}_2\mathbf{Z}_2 + \tau_2\mathbb{1} \tag{3.13}$$

Resulting from the fact that each process is Gaussian, the joint model $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)'$ is bivariate normal with zero mean and variance-covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{12}' & \mathbf{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1\mathbf{A}_1' & Cov(\mathbf{Y}_1, \mathbf{Y}_2) \\ Cov(\mathbf{Y}_1, \mathbf{Y}_2)' & \mathbf{A}_2\mathbf{A}_2' + \tau_2^2 I_{q_2} \end{pmatrix} \tag{3.14}$$

Given (3.10) the covariance between the areal and point-level process is

$$Cov(\mathbf{Y_1}, \mathbf{Y_2}) = \mathbf{A}_1 Cov(\mathbf{Z}_1, \mathbf{Z}_2)\mathbf{A}_2' \tag{3.15}$$

where $\mathbf{J}$ is still a $q_1 \times q_2$ correlation matrix between $Cov(Z_{1k}, Z_{2l})$ with $k = 1, \ldots, q_1$ and $l = 1, \ldots, q_2$. Therefore, the variance-covariance matrix in (3.14) becomes

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{A}_1\mathbf{A}_1' & \mathbf{A}_1\mathbf{J}\mathbf{A}_2' \\ \mathbf{A}_2\mathbf{J}'\mathbf{A}_1' & \mathbf{A}_2\mathbf{A}_2' + \tau_2^2 I_{q_2} \end{pmatrix} \tag{3.16}$$

where $Cov(Z_{1k}, Z_{2,l}) = 0$ or $Cov(Z_{1k}, Z_{2,l}) = 1$ and

$$l(\mathbf{J}) = -\frac{1}{2}log2\pi - \frac{1}{2}|\mathbf{\Sigma}| - \frac{1}{2}\mathbf{Y}'\mathbf{\Sigma}^{-1}\mathbf{Y}. \tag{3.17}$$

Once this is established, the construction of the $\mathbf{J}$ matrix is as outlined above in Section 3.1.2 with the maximum number of 1's being $c = min(m, n)$.

3.2 Prediction

The main aim of obtaining estimates for the covariances matrices for each process is to make reasonable predictions for at least one of the processes being studied. Given the two process model with realisation vector $\mathbf{Y}$, it can be shown that the simple cokriging predictor given the representation as in (3.3) (Wang, 2011; Hu, 2013) is

$$\hat{\mathbf{Y}}_i = E(\hat{Y}_i) + \mathbf{K}'\mathbf{\Sigma}^{-1}\left(\mathbf{Y} - E(\mathbf{Y})\right) \tag{3.18}$$

where $\hat{\mathbf{Y}}_i$ is a set of locations related to process i, whether areas or points, where predictions are required. The corresponding variance is

$$\sigma^2_{sck} = Var(\hat{Y}_i) - \mathbf{K}\mathbf{K}'\mathbf{\Sigma}^{-1}\mathbf{K} \tag{3.19}$$

where

$$\mathbf{K} = Cov(\mathbf{Y}, \hat{\mathbf{Y}}_i) = \begin{pmatrix} Cov(\mathbf{Y}_1, \hat{\mathbf{Y}}_i)' \\ Cov(\mathbf{Y}_2, \hat{\mathbf{Y}}_i)' \end{pmatrix} \tag{3.20}$$

and $\mathbf{\Sigma}^{-1}$ is the inverse of the joint covariance matrix. More specifically, the elements in (3.20) are defined as

$$Cov(\mathbf{Y}_j, \hat{\mathbf{Y}}_i) = \mathbf{A_j}\mathbf{J}a_i \qquad \text{if } i \neq j \tag{3.21}$$

$$Cov(\mathbf{Y}_j, \hat{\mathbf{Y}}_i) = \mathbf{A_j}a_i \qquad \text{if } i = j \tag{3.22}$$

where $\mathbf{A_j}$ is the $\mathbf{A}$ matrix from (3.2) related to process j and a_i represents the subset of rows in the $\mathbf{A_i}$ matrix corresponding to the unobserved locations. Therefore

$$\mathbf{K} = (a_i'A_1', a_i'A_2')' \tag{3.23}$$

3.3 Cross-Validation

Criteria such as goodness of fit and predictive performance are used to choose between the results of competing statistical methods, since the main aim of using nonparametric cross-covariance matrices as outlined above is to make predictions at specific locations irrespective of whether they are for areas, Cartesian points or both. As a result, determining the best model should be based on predictive performance. Cross-validation, the method used for evaluating this type of performance, is where the resulting model which was built using a training dataset is used to do predictions on a test dataset. One option is to use the drop-one prediction method. This technique which is sometimes referred to as leave-out-one cross-validation, calculates the best linear prediction at a specific location using all the other observations in the dataset. It therefore means that information from the location being predicted is not being used to build the model for prediction at that point. Although this is a more efficient use of the available data when compared to other cross-validation methods, the technique can be time-consuming. Notwithstanding, (Zhang & Wang, 2010) developed a computationally efficient algorithm for this procedure and its related measures.

Assuming a zero-mean process and $\mathbf{Y}$ is a realisation of that process, then the drop-one prediction is (3.24)

$$\hat{Y}_{-i} = -\sum_{j \neq i} \frac{q_{ij} Y_j}{q_{ii}} \tag{3.24}$$

where q_{ij} is the $(i,j)^{th}$ element of the precision matrix, $\mathbf{V}$, which is the inverse of the variance-covariance matrix $\mathbf{V} = \mathbf{\Sigma}^{-1}$. The corresponding drop-one variance is

$$\hat{\sigma}^2_{-i} = \mathbf{\Lambda}^{-1} = E(Y_i - \hat{Y}_{-i})^2$$

$$= \frac{1}{q_{ii}} \tag{3.25}$$

3.3.1 Prediction Scores

For the purpose of this dissertation, four predictive scores will be used to evaluate the performance of the proposed method. In each case, smaller values indicate better predictive performance.

Mean Squared Error and Root Mean Squared Error

The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are frequently used measures for determining predictive performance. The MSE is the average of the squared difference between the observed value and the predicted value at a specific location. The MSE is calculated as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_{-i}\right)^2$$

$$= \frac{1}{n}||\mathbf{\Lambda}^{-1}\mathbf{V}\mathbf{Y}||^2 \tag{3.26}$$

where n is the total number of observations and the RMSE is simply the square root of the MSE.

Logarithmic Score

One shortcoming of the MSE and RMSE is that they do not consider the predictive distribution of the process being studied. Conversely, the Logarithmic Score (LogS) (Gneiting et al., 2007) takes account of both the drop-one predicted value and the drop-one predicted variance. The LogS, assuming a Gaussian distribution, is calculating using (3.27).

$$\text{LogS} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2}\log(2\pi\hat{\sigma}^2_{-i} + \frac{1}{2}z_i^2)\right] \tag{3.27}$$

where

$$z_i = \frac{Y_i - \hat{Y}_{-i}}{\hat{\sigma}_{-i}}$$

$$= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}\mathbf{Y} \tag{3.28}$$

Continuous Ranked Probability Score

LogS, despite having many attractive properties (Roulston & Smith, 2002), lacks robustness (Gneiting & Raftery, 2007). The Continuous Ranked Probability Score (CPRS) accounts for this limitation while still using the predictive distribution (Gneiting et al., 2007; Gneiting & Raftery, 2007). Let $F_i(y)$ be the predictive cumulative distribution function of Y where

$$F_i(y) = P\left(Y(r_i) \leq y | Y(r_j), j \neq i\right) \tag{3.29}$$

and r represents the location, whether it be an area (B) or a point (s). Furthermore, an indicator function is defined such that

$$\mathbf{1}_{\{Y(r_i)\leq y\}} = \begin{cases} 1 & \text{if } Y(r_i) \leq y \\ 0 & \text{otherwise} \end{cases} \tag{3.30}$$

Then the CPRS is

$$\text{CPRS} = \frac{1}{n}\sum_{-\infty}^{\infty}\left(F_i(y) - \mathbf{1}_{\{Y(r_i)\leq y\}}\right)^2 \tag{3.31}$$

where $F_i(y)$ and $\mathbf{1}_{\{Y(r_i)\leq y\}}$ are defined as in (3.29) and (3.30) respectively. If y is Gaussian, then the CPRS can be written in terms of the z_is defined in (3.28)

$$\text{CPRS} = \frac{1}{n}\sum_{1}^{n}\hat{\sigma}_{-i}\left(z_i\left(2\Phi(z_i) - 1\right) + 2\Phi(z_i) - \frac{1}{\sqrt{\pi}}\right) \tag{3.32}$$

4. MODELLING DATA OF DIFFERENT SUPPORT

Related variables are often used to help improve the accuracy of predicted data. In many cases, due to agencies using different measurements and boundaries, these correlated variables are not necessarily of the same support, nor cover the same areas. Many of the methods outlined in Chapter 2 do not allow for the use of different variables measured at different levels of support to be combined for use in spatial modelling. Furthermore, hierarchical methods applied to data of different support are not appropriate if the supports of the data lack a hierarchical structure where the boundaries of one scale overlap with, or are not subsumed within the boundaries of another scale. The proposed nonparametric cross-covariance matrix is a suitable alternative, as it is not only applicable in situations where measurement areas do not overlap, but also particularly useful in cases where the variables of interest are measured at different levels of support and/or the implementation of extant methods is intricate, because of the absence of alternative methods.

The usefulness and flexibility of the proposed method is demonstrated using two correlated variables which are of different support. Once the marginal models for each variable are obtained, the nonparametric method will be used to construct the cross-covariance matrix and its predictive performance evaluated. The motivating dataset, Native Biomass and its correlated variable, Temperature, illustrates that the method is appropriate for predicting data of different support and outperforms an alternative method.

4.1 The Data and Marginal Models

4.1.1 Native Biomass

Biomass is organic material derived from living or recently living organisms. Native Biomass is therefore just organic material obtained from flora that are native to the forests and other areas in which they are located. This organic matter is of importance because it has implications for major industries, such as transportation where approximately 10% of Biomass energy is used for transportation fuels, and rural areas as Biomass helps to bolster these economies through the lumber, paper and pulp industries. Biomass also has important implications for the environment as it can assist in mitigating against climate change while reducing the risk of fires (Bartuska, 2006; Schoene & Killmann, 2007). Biomass and its prediction therefore have implications for local, regional and global policies. For ease of reference, the term Biomass will be used to refer to Native Biomass for the remainder of this dissertation.

The United States is divided into two main ecological regions called Domains. The area of interest is the Eastern Humid Domain, which is just over $3,500,000\ km^2$ and can be further divided into 91 units, called Sections (Cleland et al., 2007). Biomass was collected from Forest Inventory and Analysis Program (FIA) plots (Oswalt et al., 2015) and the average Biomass is available for each Section. Biomass is spatially clustered (Figure 4.1), which is supported by the Moran's I (Moran, 1950) statistic being -0.654 ($p < 2.2e$-16) when the queen weighting method is used.

The distribution of Biomass is fairly normally distributed as seen from the histogram (Figure 4.2). Therefore the Gaussian conditional autoregressive (CAR) model is used as the joint probability model for Biomass. Let $\mathbf{Y}_1$ denote the Biomass observed at $n_1 = 91$ Sections. Then $\mathbf{Y}_1$ has a multivariate normal distribution with a constant mean and a precision matrix (i.e., the inverse of the covariance matrix)

$$\mathbf{\Sigma}_{11}^{-1} = \frac{1}{\tau^*}(M_w - \rho W) \tag{4.1}$$

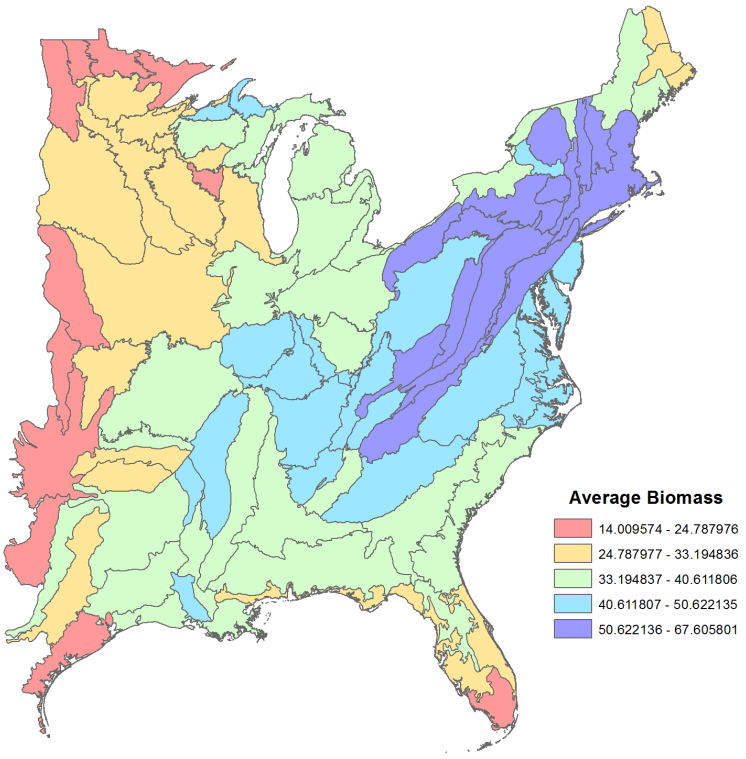


Figure 4.1. Ecological Sections - Average Biomass

where τ^* is the variance parameter, ρ is the parameter which controls spatial dependency, M_w is a diagonal matrix with the number of neighbours for each Section and W is the proximity matrix with $w_{ij} = 1$ if j is a neighbour of i and 0 otherwise.

The parameters were estimated by maximizing the Gaussian likelihood function. The resulting estimates are $\hat{\rho} = 0.991$ and $\hat{\tau^*} = 193.2632$.

4.1.2 Temperature

Temperature is known to be correlated with Biomass (Bartuska, 2006; Schoene & Killmann, 2007). The 30-year Normal Annual Temperature data was collected from the National Oceanic and Atmospheric Administration (NOAA) who hosts the National Climatic Data Center's (NCDC) three-decade (1981-2010) averages for Temperature (^oF), among other variables. 30-year Normals are "period averages com-

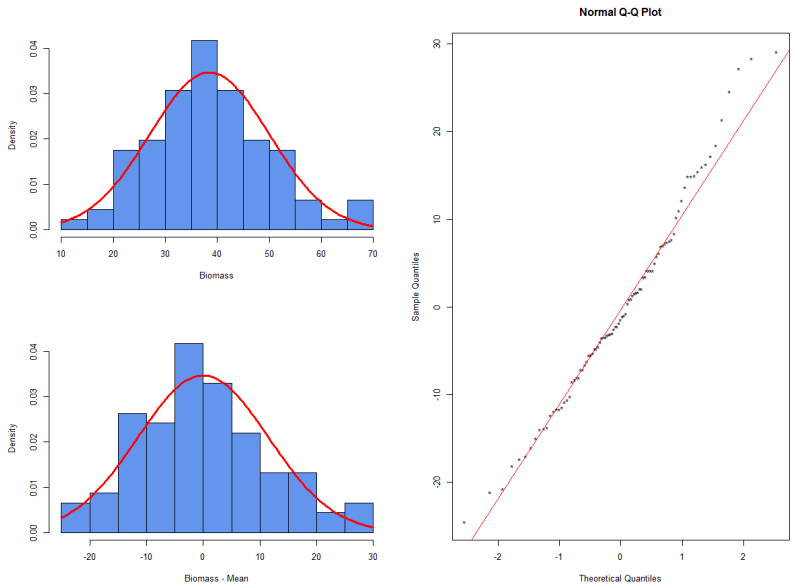


Figure 4.2. Average Biomass

puted for a uniform and relatively long period comprising at least three consecutive ten-year periods" (World Meteorological Organization, 1989). These calculations are required by all members of the World Meteorological Organization (WMO). These 30-year normals are sometimes referred to as climatological data and is the average of the Annual Maximum Temperature Normal and the Annual Minimum Temperature Normal (World Meteorological Organization, 1989; Arguez, Applequist, et al., 2012; Arguez, Durre, et al., 2012).

The following variables were observed at $n_2 = 3,617$ weather stations which fall within the study area: Temperature, latitude, longitude and altitude at each weather station. It should be noted that the number of weather stations within each Section (Figure 4.3) is not related to the size of the Section, as the placement of stations is independently decided by NOAA.

The temperature exhibits a clear spatial trend (Figure 4.4), which is assumed to depend on longitude, latitude and elevation. Let $Y_2(s)$ denote the temperature at location s. $Y_2(s)$ is modelled by a Gaussian process with mean $\mu(s)$ and an exponential

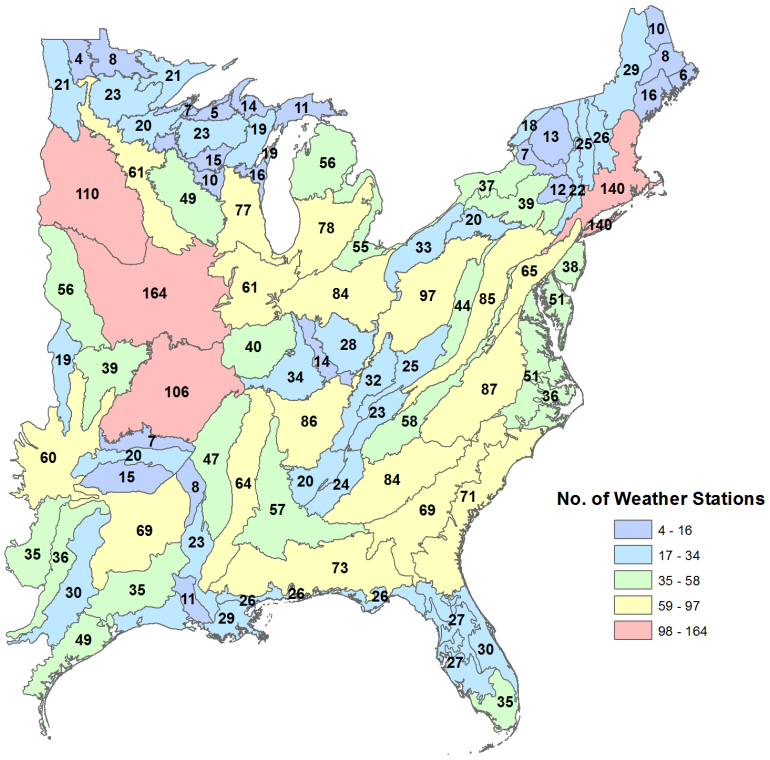


Figure 4.3. Weather Stations in the Eastern Domain

covariance function with partial variance σ^2, a nugget effect τ^2 and a range parameter ϕ:

$$C(h) = \begin{cases} \sigma^2 \exp(-h/\phi) & h > 0 \\ \sigma^2 \exp(-h/\phi) + \tau^2 & h = 0 \end{cases} \tag{4.2}$$

where h denotes the distance between two spatial locations. The mean function is assumed to have the following parametric form (4.3)

$$\begin{aligned} \mu(s) = \beta_0 + Lat(s)\beta_1 + Long(s)\beta_2 + Ele(s)\beta_3 + Lat(s)^2\beta_4 \\ + Long(s)^2\beta_5 + Lat(s) \times Long(s)\beta_6 \end{aligned} \tag{4.3}$$

where $Lat(s), Long(s)$ and $Ele(s)$ are latitude, longitude, and elevation at location s, $\beta_0, \ldots, \beta_6$ are unknown parameters to be estimated. The least squares estimate for these parameters are $\beta_0 = 186.3, \beta_1 = -1.808, \beta_2 = 1.563, \beta_3 = -0.00916, \beta_4 = -0.01307, \beta_5 = 0.006254, \beta_6 = -0.01426$. The detrended observed data $\mathbf{Y}_2$ therefore

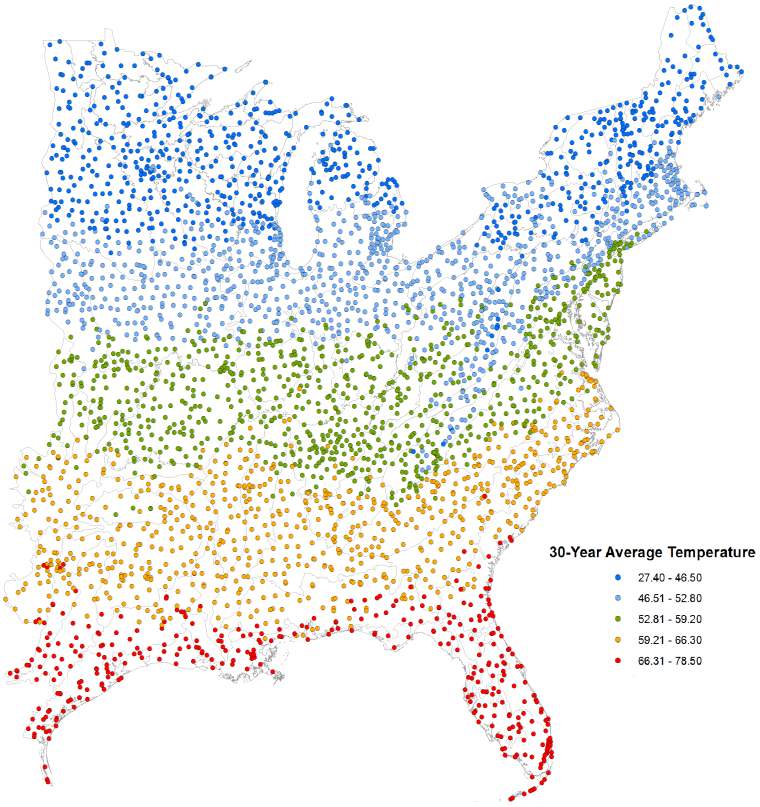


Figure 4.4. Weather Stations in the Eastern Domain

has a multivariate normal distribution with the mean $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}_{22}$, where $\mathbf{m} = (\mu(s_1), \ldots, \mu(s_n))'$ and $\mathbf{\Sigma_{22}}$ has the (i,j)th element being $C(||s_i - s_j||)$ that depends on parameters σ^2, ϕ and τ^2. The parameters are estimated by maximizing the following log-likelihood

$$-(1/2)\log(|\mathbf{\Sigma}_{22}|) - (1/2)(\mathbf{Y}_2 - \mathbf{m})'\mathbf{\Sigma}_{22}^{-1}(\mathbf{Y}_2 - \mathbf{m}). \tag{4.4}$$

The maximum likelihood estimates for the parameters were: $\hat{\sigma}^2 = 0.4188$, $\hat{\phi} = 143.8055\ km$ and $\hat{\tau}^2 = 0.7399$.

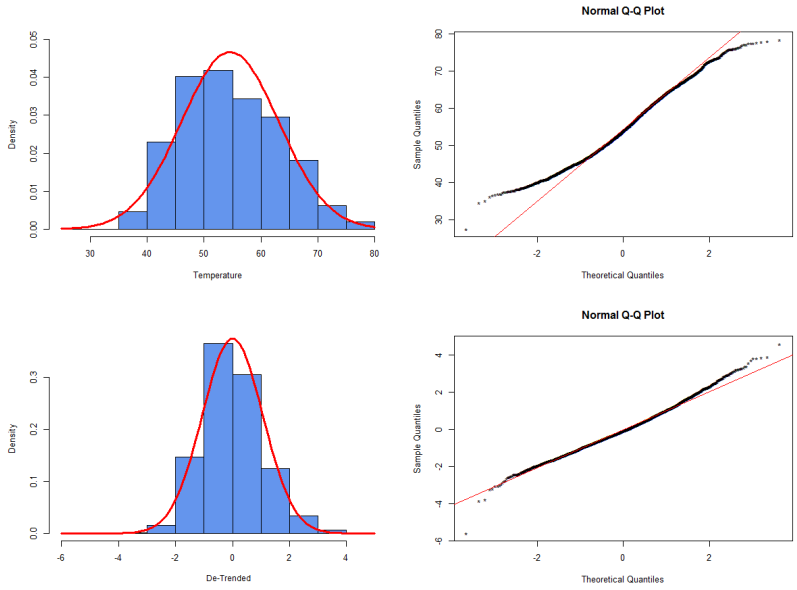


Figure 4.5. Upper Panel: Original Temperature Data and Lower Panel: De-Trended Temperature Data

4.2 Joint Models

Once the marginal models are fitted, the joint model can be constructed using the non-parametric cross-covariance matrices, as described in the previous chapter. For the purpose of comparison, a hierarchical model is also considered. The following two subsections describe the two models, respectively.

4.2.1 Semiparametric Model

In this subsection, for simplicity let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ denote the detrended Biomass and Temperature, respectively. Since $\mathbf{Y}_1$ has a normal distribution, it can be written

$$\mathbf{Y}_1 = \mathbf{A}_1\mathbf{Z}_1. \tag{4.5}$$

where $\mathbf{Z}_1$ has i.i.d. standard normal elements.

Similarly, $\mathbf{Y}_2$ can be written as

$$\mathbf{Y}_2 = \mathbf{A}_2\mathbf{Z}_2 + \tau\mathbf{Z}_0 \tag{4.6}$$

where $\mathbf{Z}_0$ and $\mathbf{Z}_2$ are independent of each other, and each have i.i.d. standard normal elements.

Then the joint distribution of $\mathbf{Y} = (\mathbf{Y_1}, \mathbf{Y_2})'$ is multivariate normal with covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{A}_1\mathbf{A}_1' & \mathbf{A}_1\mathbf{J}\mathbf{A}_2' \\ \mathbf{A}_2\mathbf{J}'\mathbf{A}_1' & \mathbf{A}_2\mathbf{A}_2' + \tau^2\mathbf{I} \end{pmatrix} \tag{4.7}$$

where $\mathbf{A}_1$ is a $n_1 \times n_1$ matrix, $\mathbf{A}_2$ is $n_2 \times n_2$ and $\mathbf{J}$ is $n_1 \times n_2$, with $Cov(Z_{1k}, Z_{2,l}) = 0$ or $Cov(Z_{1k}, Z_{2,l}) = 1$. As outlined in Chapter 3, the $\mathbf{J}$ matrix can only have at most $c = min(q_1, q_2)$ perfectly correlated pairs. In this case $c = n_1 = 91$. Two latent variables, Z_{1k} and Z_{2l} are determined to have a correlation of 1 if it maximises (4.8).

$$l(\mathbf{J}) = -\frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\mathbf{Y}'\mathbf{\Sigma}^{-1}\mathbf{Y} \tag{4.8}$$

Since the dimension of $\mathbf{Z}_1$ is large ($n_2 = 3617$), we made an assumption to simplify the computation to maximize (4.8). We assume that $\mathbf{Z}_1$ is possibly dependent with only the first 100 elements of $\mathbf{Z}_2$, and is independent with all other elements of $\mathbf{Z}_2$. This assumption effectively reduces the dimension of $\mathbf{J}$ to $n_1 \times 100$. The corresponding eigenvalues for the first 100 elements in $\mathbf{Z}_2$ explain approximately 64% of the variation in temperature.

4.2.2 Hierarchical Model

Since the locations of Temperature are nested within the sections of Biomass, the hierarchical model can be easily formulated through the conditional distribution of $\mathbf{Y}_2$ on $\mathbf{Y}_1$. Again, both $\mathbf{Y}_2$ and $\mathbf{Y}_1$ are detrended and therefore have mean 0.

More specifically,

$$\mathbf{Y}_2|\mathbf{Y}_1 \sim MVN(\mu_T, \mathbf{V}_T) \tag{4.9}$$

where $\mathbf{V}_T$ is assumed to be the exponential covariance function with a nugget effect.

In (4.9), $\mu_T = \mathbf{X}\beta$ where the design matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{pmatrix} 1 & Y_{11} \\ 1 & Y_{11} \\ \vdots & \vdots \\ 1 & Y_{11} \\ 1 & Y_{12} \\ \vdots & \vdots \\ 1 & Y_{12} \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & Y_{1\ 91} \end{pmatrix} \tag{4.10}$$

and Y_{1i} is the Biomass in Section i and is constant for all weather stations within a given section. Also

$$\beta = \begin{pmatrix} a \\ b \end{pmatrix} \tag{4.11}$$

The following estimates were obtained after maximising the conditional log likelihood (4.12)

$$\log(l) \propto -\frac{1}{2}\ \log|\mathbf{V}_T| - \frac{1}{2}\left(\mathbf{Y}_2 - \mathbf{X}\beta\right)' \mathbf{V}_T^{-1}\left(\mathbf{Y}_2 - \mathbf{X}\beta\right) \tag{4.12}$$

$\hat{\phi} = 140.2045$, $\hat{\tau} = 0.7406$, $\hat{\sigma}^2 = 0.4049$, $a = 0.045366$, $b = 0.000498$.

Using the laws of covariances, expectation and total variance, the joint distribution is

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{M} \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{11}' & \mathbf{\Sigma}_{22} \end{pmatrix}\right) \tag{4.13}$$

where $\mathbf{\Sigma}_{11}$ is the inverse of the precision matrix in (4.1) and $\mathbf{\Sigma}_{12} = \mathbf{B}\mathbf{\Sigma}_{11}$ with $\mathbf{B}$ being a $n_2 \times n_1$ matrix

$$\mathbf{B} = \begin{pmatrix} b\mathbf{1}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & b\mathbf{1}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & b\mathbf{1}_{91} \end{pmatrix} \tag{4.14}$$

where the length of $\mathbf{1}_i$ is the number of weather stations in Section i. Also, $\mathbf{\Sigma}_{22} = \mathbf{V}_T + \mathbf{B}\mathbf{\Sigma}_{11}\mathbf{B}'$. For the mean vector, $\mathbf{0}$ has length $= n_1$ and $\mathbf{A}$ has length $= n_2$ with $\mathbf{M} = a\mathbf{1}$ where a is from (4.11).

4.3 Predictive Performance and Cross-Validation

The predictive performance of the marginal, hierarchical and proposed semiparametric models were evaluated using four predictive scores, namely, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Logarithmic Score (LogS) and Continuous Ranked Probability Score (CPRS). Additionally, the drop-one predicted values were also obtained. The predictive performance of the semi-parametric methodology using both the fully optimised **J** (Joint - SP Long) and the approximately optimized **J** (Joint - SP Short), the baseline marginal model (Marginal) and the hierarchical model (Joint - HM) was assessed for Biomass and Temperature separately.

4.3.1 Algorithm Performance

As outlined in Chapter 3, there are two algorithms used to construct the **J** matrix and both begin with a $n_1 \times 100$ matrix of zeros. The first algorithm (Joint - SP Long) searches the entire **J** matrix to find the pair which maximises the likelihood in (4.8). Once this pair is found, a 1 is placed in the related cell and the corresponding row and column are removed from the search area. This process is repeated 90 times until all correlated pairs are found.

The second algorithm (Joint - SP Short) begins by searching the first row for the perfectly correlated latent variables in **J** that maximises the likelihood. Once that

pair is identified, the column is removed from any subsequent searches. Irrespective of whether a perfectly correlated pair was found, the row is never searched again.

Note that it is possible for no pair to maximise the likelihood in a given iteration. Therefore, in the case of the Joint - SP Short algorithm, no column is removed from the search area, while for the Joint - SP Long, no column or row is removed. This means that there may be less than n_1 perfectly correlated pairs. Additionally, to reduce computational time, the search within each iteration is run in parallel.

The function to construct the **J** matrix was executed using R-3.4.1 on a server running Ubuntu 16.04 OS with 1 TB RAM and 12 CPU cores. The computational results are presented in Table 4.1.

Table 4.1.
Overall Computational Performance

Algorithm	**# of Pairs**	**Time**	
		Elapsed	**Relative**
Short	85	2022.86	1.000
Long	86	90526.33	44.752

Of the pairs identified, only 12 were the same in both algorithms. Additionally, only Row 81 was not selected to have any perfectly correlated pairs in both algorithms.

4.3.2 Predictive Scores

The predictive scores for Biomass show that both proposed Joint-SP models outperform the Marginal and the Joint-HM models (Table 4.2). This is shown by the markedly lower values for both Joint-SP models. This noticeable improvement in the predictive performance for Biomass may be as a result of the dense distribution of the Temperature weather stations (3617) compared with only 91 measurements on Biomass. Additionally, in this case, the Joint-HM model does slightly worse than the

Marginal model. Of note, although the Joint-SP Short model is a marked improvement over the Joint-HM and Marginal models, it performs slightly worse than the Joint-SP Long model.

Table 4.2.
Predictive Performance - Biomass

Model	MSE	RMSE	LogS	CPRS
Marginal	35.625	5.969	3.204	5.753
Joint-HM	35.696	5.974	3.203	5.758
Joint-SP Short	1.701	1.304	1.784	1.386
Joint-SP Long	**0.452**	**0.672**	**1.511**	**0.949**

Conversely for Temperature, the difference between the predictive scores for the models is not as stark. Despite that, the results showed that the Joint-SP models perform slightly better than the Marginal and Joint-HM models (Table 4.3). This is not surprising since the Temperature data is quite dense over the region being studied and the inclusion of Biomass does not add much more information to the models. Also, the performance of the Marginal and Joint-HM models were approximately the same for all criteria. Again, the difference between the Joint-SP Long and the Joint-SP Short is very small, even though the Joint-SP Long still gives better results.

Table 4.3.
Predictive Performance - Temperature

Model	MSE	RMSE	LogS	CPRS
Marginal	0.868	0.932	1.349	0.903
Joint-HM	0.868	0.932	1.385	0.902
Joint-SP Short	0.831	0.911	1.327	0.897
Joint-SP Long	**0.826**	**0.909**	**1.325**	**0.886**

4.3.3 Prediction Error

The drop-one prediction scores were obtained and the prediction errors, that is the difference between the observed and the drop-one predicted values, for each model calculated. The prediction error for the Joint-SP Long model is superimposed on the graphs with errors for the comparison models.

The plots for both Biomass and Temperature both corroborate the results obtained in Tables 4.2 and 4.3. The prediction errors for the Joint-SP models are relatively small (between ± 4) when compared to the other two models where the errors lie between -15 and +20 (Figure 4.6).

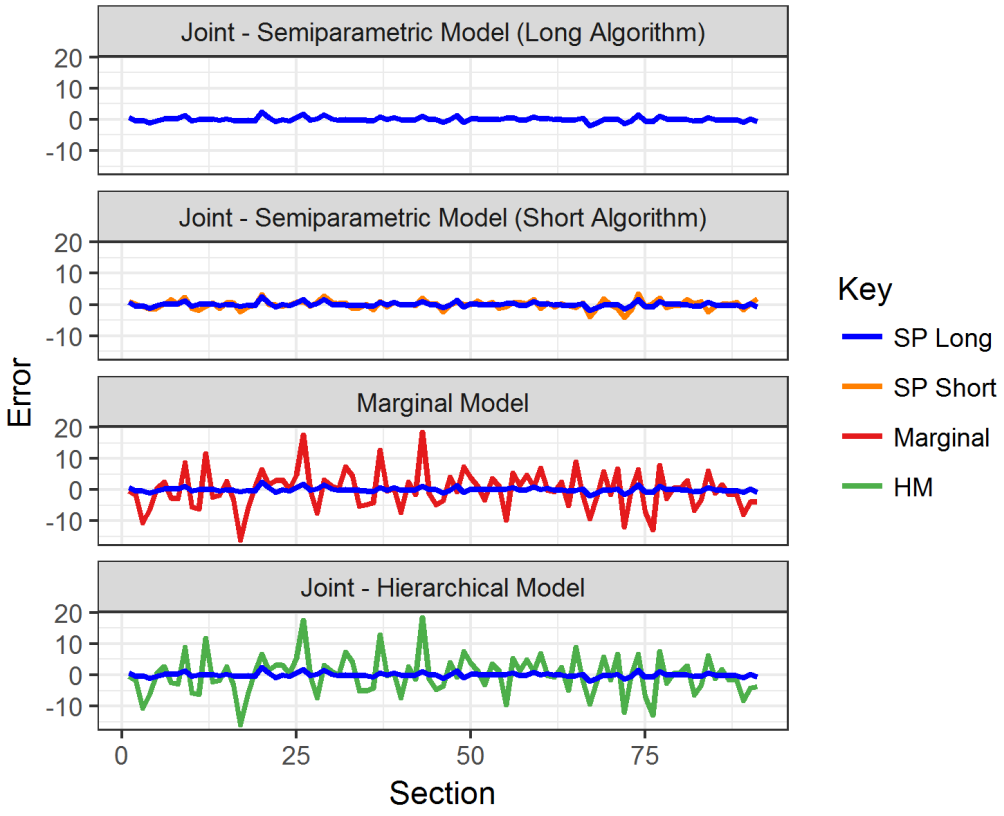


Figure 4.6. Prediction Error for Biomass - Joint Semiparametric Models Compared with Joint Hierarchical and Marginal Models

Finally, in contrast to the results for Biomass, but not unexpectedly, the four plots for Temperature are all similar, ranging from -6 to +4.5. (Figure 4.7).

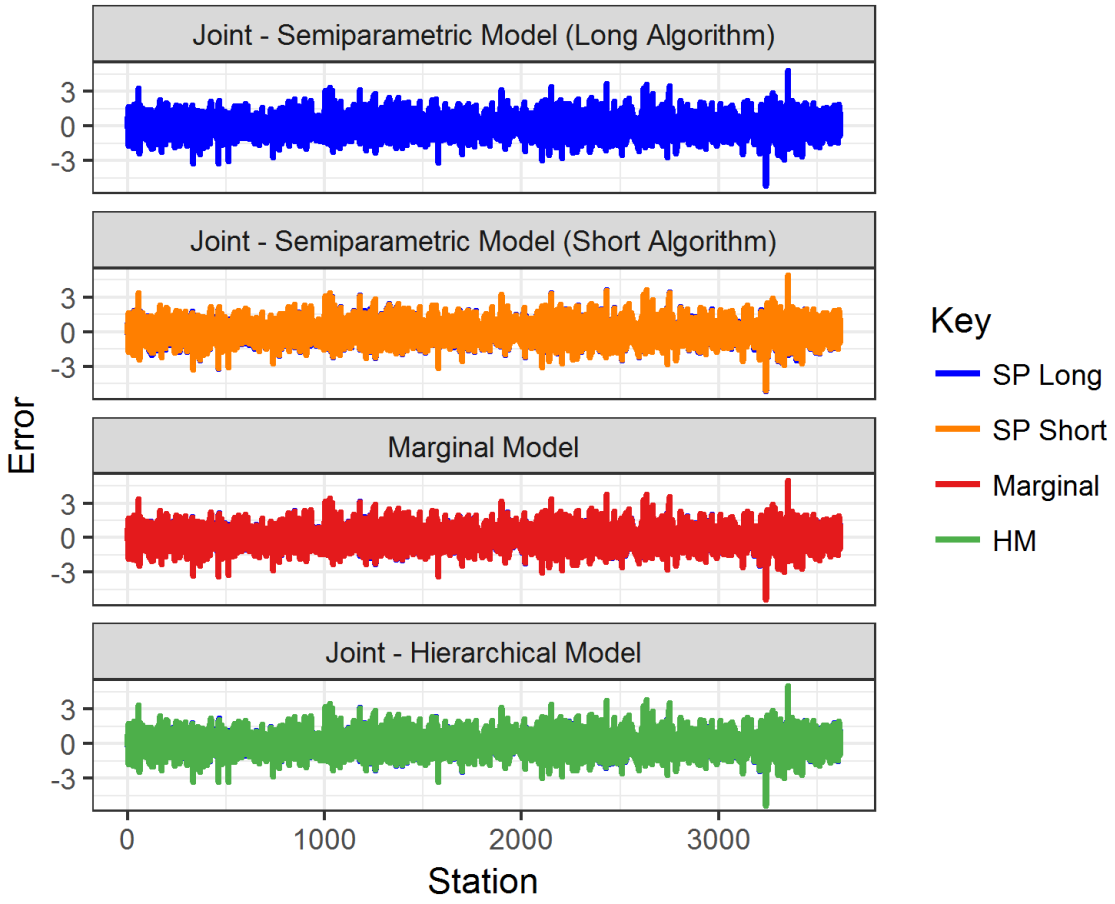


Figure 4.7. Prediction Error for Temperature - Joint Semiparametric Models Compared with Joint Hierarchical and Marginal Models

5. SUMMARY AND CONCLUSION

The aim of this dissertation was to build a nonparametric cross-covariance matrix to improve the accuracy of predicted data when related variables of different support are available. The proposed semiparametric covariance method, which uses the marginal covariance matrices to create the cross-covariance matrix, yielded more accurate predictions than the comparative marginal and hierarchical models.

Since the marginal covariance functions are used in the construction of the nonparametric cross-covariance matrix, the method is attractive because of its flexibility for use in many different circumstances. For example, there is no need to make adjustments for the difference in the support at which the data is measured, or to consider the effect of misaligned boundaries.

For practitioners this method is appealing because the estimation of the marginal covariance function does not require advanced statistical knowledge since this estimation has been implemented for many spatial covariance functions in menu driven applications (eg ARGIS (ESRI, 2011)) and R (R Development Core Team, 2006) packages such as geoR (Ribeiro Jr. & Diggle, 2001) and fields (Nychka et al., 2015). Therefore the method is attractive since it does not require hierarchical analysis which requires more in depth statistical knowledge. Also, the calculation of the cross-covariance matrix is really an optimization problem which can also be executed in already established programmes such as R (R Development Core Team, 2006).

One limitation of the proposed method, is that the computational cost increases in $O(n^2)$ especially when a large number of latent variables are needed to reasonable explain the variability within the process. The joint semiparametric method using the short algorithm performed better than the marginal and hierarchical models, but marginally worse than the algorithm using the fully optimised likelihood. Notwithstanding, the long version of the algorithm took approximately 45 times longer than

the short algorithm. In practice, it is therefore recommended that the short algorithm be used when computational resources are limited and datasets are large.

In light of the fact that the proposed method does not require any adjustments for data that do not use the same levels of support or administrative boundaries, the contribution of this work enhances decision and policy-making abilities in an age where vast amounts of data are collected and there is increased data-sharing among different agencies.

5.1 Future Work

In this work, nonparametric cross-covariance matrices are applied to data that is assumed to have a normal distribution. By extension, this method needs to be applied to data from other distributions, such as count and binary data, especially because non-Gaussian data play an important part in research (eg the prediction disease incidence which typically has a Poisson distribution, is often related to climatic variables). Additionally, the next logical movement is to extend the method to prediction in the spatiotemporal case, as time indexed variables are an increasingly important area in spatial statistics. Finally, there is scope for further development of an algorithm which is more computationally efficient.

REFERENCES

REFERENCES

Arguez, A., Applequist, S., Vose, R. S., Durre, I., Squires, M. F., & Yin, X. (2012). *NOAA's 19812010 Climate Normals: Methodology of Temperature-related Normals* (Tech. Rep.).

Arguez, A., Durre, I., Applequist, S., Vose, R. S., Squires, M. F., Yin, X., ... Owen, T. W. (2012). NOAA's 1981-2010 U.S. Climate Normals: An Overview. *Bulletin of the American Meteorological Society*, *93*(11), 1687–1697.

Atkinson, P. M., & Tate, N. J. (2000). Spatial Scale Problems and Geostatical Solutions: A Review. *Professional Geographer*, *52*(4), 607.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2015a). Basics of Areal Data. In *Hierarchical modeling and analysis for spatial data* (2nd ed., pp. 73–96). CRC Press/Taylor & Francis Group.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2015b). Models for Multivariate Areal Data. In *Hierarchical modeling and analysis for spatial data* (2nd ed., pp. 305–328). CRC Press/Taylor & Francis Group.

Banerjee, S., Gelfand, A. E., & Finley, A. O. (2008). Gaussian predictive process models for large spatial. *Society*, 825–848.

Banham, M. R., & Katsaggelos, A. K. (1996). Spatially adaptive wavelet-based multiscale image restoration. *Image Processing, IEEE Transactions on*, *5*(4), 619–634.

Bartuska, A. (2006). *Why biomass is important–The role of the USDA Forest Service in managing and using biomass for energy and other uses* (Tech. Rep.). Research & Development, USDA Forest Service.

Berliner, L. M., Milliff, R. F., & Wikle, C. K. (2003). Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research*, *108*(C4).

Bilonick, R. A. (1988). Monthly Hydrogen Ion Deposition Maps for the Northeastern U.S. from July 1982 to September 1984. *Atmospheric Environment*, *22*(9), 1909–1924.

Bracken, I., & Martin, D. (1989). The generation of spatial population distributions from census centroid data. *Environment and Planning A*, *21*, 537–543.

Brillinger, D. R. (1990). Spatial Temporal Modelling of Spatially Aggregate Birth Data. *Survey Methodology*, *16*(2), 255–269.

Brillinger, D. R. (1994). Examples of Scientific Problems and Data Analyses in Demography, Neurophysiology and Seismology. *Journal of Computational and Graphical Statistics*, *3*(1), 1–22.

Burgess, T. M., & Webster, R. (1980). Optimal Interpolation and Isarithmic Mapping of Soil Properties. *Journal of Soil Science*, *31*(2), 333–341.

Carroll, S. S., Day, G. N., Cressie, N., & Carroll, T. R. (1995). Spatial Modeling of Snow Water Equivalent Using Airborne and Ground-Based Snow Data. *Environmetrics*, *6*(2), 127–139.

Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

Chou, K. C., Willsky, A. S., & Benveniste, A. (1994). Multiscale recursive estimation, data fusion, and regularization. *IEEE Trans. Automat. Contr.*, *39*(3), 464–478.

Cleland, D., Freeouf, J., Keys, J., Nowacki, G., Carpenter, C., & McNab, W. H. (2007). *Ecological Subregions: Sections and Subsections for the conterminous United States* (Tech. Rep.). General Technical Report WO-76. U.S. Department of Agriculture, Forest Service, Washington, DC.

Cressie, N. (1990). The Origins of Kriging. *Mathematical Geology*, *22*(3), 239–252.

Cressie, N. (1991). *Statistica for Spatial Data*. John Wiley & Sons Inc.

Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, *3*(2-3), 159–180.

Cressie, N. (2006). Block Kriging for Lognormal Spatial Processes. *Mathematical Geology*, *38*(4), 413–443.

Cressie, N., & Johannesson, G. (2008). Fixed Rank Kriging for Very Large Spatial Data Sets. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *70*(1), 209–226.

Dark, S., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., ... Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*(5), 626–640.

ESRI. (2011). *ArcGIS Desktop.* Redlands, CA: Environmental Systems Research Institute.

Fieguth, P. W., Karl, W. C., & Willsky, A. S. (1995). Multiresolution Optimal Interpolation and Statistical Analysis of TOPEXPOSEIDON Satellite Altimetry. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 280–292.

Fisher, P. F., & Langford, M. (1995). Modelling the Errors in Areal Interpolation between Zonal Systems by Monte Carlo Simulation. *Environment and Planning A*, *27*(2), 211–224.

Flowerdew, R., & Green, M. (1989). Statistical methods for inference between incompatible zonal systems. In *The accuracy of spatial databases* (pp. 239–247).

Flowerdew, R., & Green, M. (1992). Developments in areal interpolation methods and GIS. *The Annals of Regional Science*, *26*, 67–78.

Flowerdew, R., & Green, M. (1994). Data Integration: Statistical Methods of Transferring Data Between Zonal Systems. In I. Masser & M. Blakemore (Eds.), *Handling geographic information: Methodology and potential applications* (pp. 38–54). London: Longman.

Flowerdew, R., Green, M., & Kehris, E. (1991). Using Areal Interpolation Methods in Geographic Information Systems. *Papers in Regional Science*, *70*(3), 303–315.

Fuentes, M., Song, H.-R., Ghosh, S. K., Holland, D. M., & Davis, J. M. (2006). Spatial Association Between Speciated Fine Particles and Mortality. *Biometrics*, *62*(3), 855–863.

Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757.

Gelfand, A. E. (2010). Misaligned Spatial Data: The Change of Support Problem. In P. Gelfand, Alan E., Diggle, Peter J., Fuentes, Montserrat, Guttorp (Ed.), *Handbook of spatial statistics* (pp. 517–538). CRC Press.

Gelfand, A. E., & Carlin, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics (Oxford, England)*, *2*(1), 31–45.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., & Sirmans, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, *13*(2), 263–312.

Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *69*(2), 243–268.

Gneiting, T., Kleiber, W., & Schlather, M. (2010). Matérn Cross-Covariance Functions for Multivariate Random Fields. *Journal of the American Statistical Association*, *105*(491), 1167–1177.

Gneiting, T., & Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, *102*(477), 359–378.

Golub, G. H., & Loan, C. F. V. (1996). *Matrix Computations*. Baltimore, MD, USA: John Hopkins Unviersity Press.

Goodchild, M. F., Anselin, L., & Deichmann, U. (1993). A Framework for the Areal Interpolation of Socioeconomic Data. *Environment and Planning A*, *25*(3), 383–397.

Goovaerts, P. (1999). Geostatistics in soil science: state-of-art and perspectives. *Geoderma*, *89*, 1–45.

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma*, *103*(1-2), 3–26.

Goovaerts, P. (2008). Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences*, *40*(1), 101–128.

Goovaerts, P. (2010). Applications to Soil Science and Medical Geography. *Mathematical Geosciences*, *42*(5), 535–554.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*(458), 632–648.

Gotway Crawford, C. A., & Young, L. J. (2005). Change of support : an interdisciplinary challenge. In *Geostatistics for environmental applications* (pp. 1–13).

Goulard, M., & Voltz, M. (1992). Linear Coregionalization Model: Tools for Estimation and Choice of Cross-Variogram Matrix. *Mathematical Geologyeology*, *24*(3), 269–286.

Green, M. (1990). Statistical Models for Areal Interpolation. In J. Harts, H. F. L. Ottens, & H. J. Scholten (Eds.), *Egis '90 proceedings vol. 1* (pp. 392–399). Utrecht, Netherlands: EGIS Foundation.

Guhaniyogi, R., Finley, A. O., Banerjee, S., & Kobe, R. K. (2013). Modeling complex spatial dependencies: low-rank spatially varying cross-covariances with application to soil nutrient data. *Journal of Agricultural, Biological, and Environmental Statistics*, *18*(3), 274–298.

Haran, M. (2011). Gaussian random field models for spatial data. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *Handbook of markov chain monte carlo* (pp. 449–478). CRC Press/Taylor & Francis Group.

Held, L., & Rue, H. (2010). Conditional and Intrinsic Autoregressions. In A. E. Gelfand, P. J. Diggle, M. Fuentes, & P. Guttrop (Eds.), *Handbook of spatial statistics* (pp. 201–216). CRC Press/Taylor & Francis Group.

Hidgon, D. (1998). A process-convolution approach to modeling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, *5*(2), 173–190.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441.

Hu, J. (2013). *Optimal Low Rank Model for Multivariate Spatial Data* (Unpublished doctoral dissertation). Purdue University.

Huang, H.-C., Cressie, N., & Gabrosek, J. (2002). Fast, Resolution-Consistent Spatial Prediction of Global Processes From Satellite Data. *Journal of Computational and Graphical Statistics*, *11*(1), 63–88.

Jelinski, D., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecol.*, *11*(3), 129–140.

Johannesson, G., Cressie, N., & Huang, H. C. (2007). Dynamic multi-resolution spatial models. *Environmental and Ecological Statistics*, *14*(1), 5–25.

Journel, A. G. (1983). Nonparametric Estimation of Spatiai Distributions. *Mathematical Geology*, *15*(3), 445–468.

Journel, A. G., & Huijbregts, C. J. (2003). *Mining Geostatistics*. New Jersey: The Blackburn Press.

Kang, J., Jin, R., Li, X., Member, S., & Zhang, Y. (2017). Block Kriging With Measurement Errors : A Case Study of the Spatial Prediction of Soil Moisture in the Middle Reaches of Heihe River Basin. *IEEE Geoscience and Remote Sensing Letters*, *14*(1), 87–91.

Karhunen, K. (1946). Zur spektraltheorie stochastischer prozesse. *Annales Academiæ Scientiarum Fennicæ Series A. I.*, *37*.

Kern, J. W., & Coyle, K. O. (2000). Global block kriging to estimate biomass from acoustic surveys for zooplankton in the western Aleutian Islands. *Canadian Journal of Fisheries and Aquatic Sciences*, *57*, 2112–2121.

Kerry, R., Goovaerts, P., Haining, R. P., & Ceccato, V. A. (2010). Applying Geostatistical Analysis to Crime Data: Car-Related Thefts in the Baltic States. *Geografiska Annaler*, *42*(1), 53–77.

Kerry, R., Goovaerts, P., Rawlins, B. G., & Marchant, B. P. (2012). Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, *170*, 347–358.

Keshavarzi, A., Sarmadian, F., & Ahmadi, A. (2011). Spatially-based model of land suitability analysis using Block Kriging. *Australian Journal of Crop Science*, *5*(12), 1533–1541.

Knotters, M., Brus, D. J., & Oude Voshaar, J. H. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, *67*(3-4), 227–246.

Kyriakidis, P. C. (2004). A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geographical Analysis*, *36*(3), 259–289.

Le, N. D., & Zidek, J. V. (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to Kriging. *Journal of Multivariate Analysis*, *43*(2), 351–374.

Loève, M. M. (1955). *Probability Theory: Foundations, Random Series*. Princeton, New Jersey: van Nostrand.

Lorenz, E. N. (1956). *Empirical Orthogonal Functions and Statistical Weather Prediction* (Tech. Rep.). Scientific Report No. 1. Statistical Forecasting Project.

Machuca-Mory, D., Babak, O., & Deutsch, C. V. (2008). Flexible Change of Support Model Suitable for a Wide Range of Mineralization Styles. *Mining Engineering*, *60*(2), 63–72.

Majumdar, A., & Gelfand, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology*, *39*(2), 225–245.

Mardia, K. V., & Goodall, C. R. (1993). Spatial-Temporal Analysis of Multivariate Environmental Monitoring Data. In G. P. Patil & C. R. Rao (Eds.), *Multivariate environmental statistics* (pp. 347–386). Elsevier Science Publishers.

Martin, D. (1996). An assessment of surface and zonal models of population. *International Journal of Geographic Information Systems*, *10*(8), 973–989.

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology*, *58*, 1246–1266.

Matheron, G. (1984). Isofactorial models and change of support. In *Geostatistics for natural resources characterization* (pp. 449–467). Springer Netherlands.

Matheron, G., & Kleingeld, W. (1987). The Evolution of Geostatistics. *APCOM 87 - Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries*, *3*, 9–12.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*(1), 17–23.

Mugglin, A. S., & Carlin, B. p. (1998). Hierarchical Modeling in Geographic Information Systems : Population Interpolation over Incompatible Zones. *Journal of Agricultural, Biological, and Environmental Statistics*, *3*(2), 111–130.

Müller, H.-G., Stadtmüller, U., & Tabnak, F. (1997). Spatial Smoothing of Geographically Aggregated Data , with Application to the Construction of Incidence Maps. *Journal of the American Statistical Association*, *92*(437), 61–71.

Nychka, D., Furrer, R., Paige, J., & Sain, S. (2015). *fields: Tools for spatial data.* Boulder, CO, USA: University Corporation for Atmospheric Research. Retrieved from `www.image.ucar.edu/fields`

Openshaw, S. (1983). *The Modifiable Areal Unit Problem.* Norwick, Norfolk: Geo Books.

Openshaw, S., & Taylor, P. J. (1979). A millions or so correlation coefficients: three experiments on the modifiable area unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences* (pp. 127–144). Pion.

Ortiz, J. M., Oz, B., & Deutsch, C. V. (2005). A Step by Step Guide to Bi-Gaussian Disjunctive Kriging. In O. Leuangthong & C. V. Deutsch (Eds.), *Geostatistics banff 2004* (Volume 14 ed., pp. 1097–1102). Springer Science & Business Media.

Oswalt, C. M., Fei, S., Guo, Q., Iannone III, B. V., Oswalt, S. N., Pijanowski, B. C., & Potter, K. M. (2015). A subcontinental view of forest plant invasions. *NeoBiota*, *24*, 49–54.

Qu, M., Li, W., Zhang, C., & Wang, S. (2012). Effect of Land Use Types on the Spatial Prediction of Soil Nitrogen. *GIScience and Remote Sensing*, *49*(3), 397–411.

Qu, M., Li, W., Zhang, C., Wang, S., M, Q., W, L., ... S, W. (2012). Effect of Land Use Types on the Spatial Prediction of Soil Nitrogen. *GIScience and Remote Sensing*, *49*(3), 397–411.

R Development Core Team. (2006). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from `http://www.r-project.org`

Rase, W.-D. (2001). Volume-preserving interpolation of a smooth surface from polygon-related data. *Journal of Geographical Systems*, *3*(2), 199–213.

Ribeiro Jr., P., & Diggle, P. (2001). geoR: a package for geostatistical analysis. *R-NEWS*, *1*(2), 15–18. Retrieved from `http://cran.r-project.org/doc/Rnews`

Ross, C. W., Grunwald, S., & Myers, D. B. (2013). Spatiotemporal modeling of soil organic carbon stocks across a subtropical region. *Science of the Total Environment*, *461-462*, 149–157.

Roulston, M. S., & Smith, L. A. (2002). Evaluating Probabilistic Forecasts Using Information Theory. *Monthly Weather Review*, *130*(6), 1653–1660.

Schirrmann, M., Herbst, R., Wagner, P., & Gebbers, R. (2012). Area-to-Point Kriging of Soil Phosphorus Composite Samples. *Communications in Soil Science and Plant Analysis*, *43*(7), 1024–1041.

Schoene, D., & Killmann, W. (2007). *Definitional issues related to reducing emissions from deforestation in developing countries*. Rome: Food and Agriculture Organization of the United Nations.

Shafran-Nathan, R., Levy, I., Levin, N., & Broday, D. M. (2017). Ecological bias in environmental health studies: the problem of aggregation of multiple data sources. *Air Quality, Atmosphere and Health*, *10*(4), 411–420.

Simone, G., Farina, A., Morabito, F., Serpico, S., & Bruzzone, L. (2002). Image fusion techniques for remote sensing applications. *Information Fusion*, *3*(1), 3–15.

Sirovich, L. (1987). Turbulence and the Dynamics of Coherent Structures Part 1: Coherent Structures. *Quarterly of Applied Mathematics*, *45*(3), 561–571.

Smith, B. J., & Cowles, M. K. (2007). Correlating point-referenced radon and areal uranium data arising from a common spatial process. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *56*(3), 313–326.

Solow, A. R. (1986). Mapping by Simple Indicator Kriging. *Mathematical Geology*, *18*(3), 335–352.

Stein, A., & Corsten, L. C. A. (1991). Universal Kriging and Cokriging as a Regression Procedure. *Biometrics*, *47*(2), 575–587.

Stein, A., Hoogerwerf, M., & Bouma, J. (1988). Use of Soil-Map Delineations to Improve (Co-) Kriging of Point Data on Moisture Deficits. *Geoderma*, *43*, 163–177.

Tobler, W. R. (1979). Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, *74*(367), 519–530.

Turner, M. G. (1989). Landscape Ecology: The Effect of Pattern on Process. *Annual Review of Ecology and Systematics*, *20*(1), 171–197.

Ver Hoef, J. M., & Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, *69*(907), 275–294.

Verly, G. (1983). The Multigaussian Approach and its Applications to the Estimation of Local Reserves. *Mathematical Geology*, *15*(2), 259–286.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications* (3rd ed.). Berlin: Springer-Verlag.

Wang, J., Ge, Y., Song, Y., & Li, X. (2014). A Geostatistical Approach to Upscale Soil Moisture with Unequal Precision Observations. *IEEE Geoscience and Remote Sensing Letters*, *11*(12), 2125–2129.

Wang, J., Wang, Y., & Zeng, H. (2016). A geostatistical approach to the change-of-support problem and variable-support data fusion in spatial analysis. *Journal of Geographical Systems*, *18*(1), 45–66.

Wang, Y. (2011). *A Multivariate Spatial Modeling Approach Using Semi-parametric Covariance Functions* (Unpublished doctoral dissertation). Purdue University.

Wikle, C. K. (2003). Hierarchical Models in Environmental Science. *International Statistical Review*, *71*(2), 181–199.

Wikle, C. K., & Berliner, L. M. (2005). Combining information across spatial scales. *Technometrics*, *47*(1), 80–91.

Wikle, C. K., Milliff, R. F., Nychka, D., & Berliner, L. M. (2001). Spatiotemporal Hierarchical Bayesian Modeling Tropical Ocean Surface Winds. *Journal of the American Statistical Association*, *96*(454), 382–397.

World Meteorological Organization. (1989). *Calculation of monthly and annual 30-year standard normals* (Tech. Rep. No. WMO-TD/No. 341). Geneva.

Wu, C., Wu, J., Luo, Y., Zhang, L., & DeGloria, S. D. (2009). Spatial Prediction of Soil Organic Matter Content Using Cokriging with Remotely Sensed Data. *Soil Science Society of America Journal*, *73*(4), 1202.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis. *Landscape Ecol.*, *19*(2), 125–138.

Yoo, E. H., & Kyriakidis, P. C. (2009). Area-to-Point Kriging in Spatial Hedonic Pricing Models. *Journal of Geographical Systems*, *11*(4), 381–406.

Young, L. J., Gotway, C. A., Yang, J., Kearney, G., & DuClos, C. (2008). Assessing the association between environmental impacts and health outcomes: A case study from Florida. *Statistics in Medicine*, *27*(20), 3998–4015.

Zacarias, O. P., & Andersson, M. (2011). Spatial and temporal patterns of malaria incidence in Mozambique. *Malaria Journal*, *10*(1), 189.

Zhang, H., & Cai, W. (2015). When Doesn't Cokriging Outperform Kriging? *Statistical Science*, *30*(2), 176–180.

Zhang, H., & Wang, Y. (2010). Kriging and cross-validation for massive spatial data. *Environmetrics*, *21*(3-4), 290–304.

Zhang, Y., Hodges, J. S., & Banerjee, S. (2009). Smoothed ANOVA with Spatial Effects as a Competitor to MCAR in Multivariate Spatial Smoothing. *Annals of Applied Statistics*, *3*(4), 1805–1830.

Zhu, L., Carlin, B. P., & Gelfand, A. E. (2003). Hierarchical regression with misaligned spatial data: Relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics*, *14*(5), 537–557.

VITA

VITA

Kelly-Ann Dixon Hamil graduated from the University of the West Indies, Mona Campus with a BSc. (Hons.) in Statistics and Economics (double major) in November 1999. She then worked as a statistician with the Ministry of National Security and Justice, Jamaica, W.I. for 2 years before receiving a Commonwealth Scholarship to read for a MSc. in Statistics at the University of Kent at Canterbury, UK. On completion, Dixon Hamil returned to her undergraduate Alma Mater where she lectured undergraduate and graduate courses in Statistics, along with directing the Statistics degree programme and collaborating with other researchers in Economics, Education and Social Work.

In 2010 she received a Fulbright Scholarship to read for a PhD in Statistics at Purdue University, Indiana, USA. During her tenure at Purdue, she began working with Professor Hao Zhang which led to interdisciplinary research with colleagues in the Forestry Department. Dixon Hamil specialises in scaling issues related to spatial data. In her dissertation, Dixon Hamil proposed a nonparametric cross-covariance matrix for modelling data of different support, which resulted in better predictions when compared to traditional models.

While at Purdue, Dixon Hamil also obtained a M.S. in Mathematical Statistics and worked with the Statistical Consulting Service (SCS) where she assisted over 35 clients from the Purdue community. She was also a Teaching Assistant and Instructor for Elementary Statistics (STAT301) and has won many teaching awards including the Graduate Teaching Award (Purdue University Teaching Academy) and the Excellence in Teaching Award (Purdue Graduate School), the highest level of recognition given to a Graduate Teaching Assistant. She is an Associate Fellow of the Purdue Teaching Academy and has an Advanced Graduate Teaching Certificate (Center Instructional Excellence, Purdue University).