INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeib Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
RICE UNIVERSITY

A Comprehensive Approach to

Spatial and Spatiotemporal Dependence

Modeling

by

L. Scott Baggett

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Katherine Bennett Ensor, Professor, Chair
Statistics

James R. Thompson, Professor
Statistics

Andre W. Droxler, Associate Professor
Geology & Geophysics

HOUSTON, TEXAS

APRIL, 2000
ABSTRACT

A Comprehensive Approach to

Spatial and Spatiotemporal Dependence Modeling

by

L. Scott Baggett

One of the most difficult tasks of modeling spatial and
spatiotemporal random fields is that of deriving an accurate representation
of the dependence structure. In practice, the researcher is faced with
selecting the best empirical representation of the data, the proper family of
parametric models, and the most efficient method of parameter estimation
once the model is selected. Each of these decisions has direct
consequence on the prediction accuracy of the modeled random field. In
order to facilitate the process of spatial dependence modeling, a general
class of covariogram estimators is introduced. They are derived by direct
application of Bochner’s theorem on the Fourier-Bessel series
representation of the covariogram. Extensions are derived for one, two
and three dimensions and spatiotemporal extensions for one, two and three
spatial dimensions as well. A spatial application is demonstrated for
prediction of the distribution of sediment contaminants in Galveston Bay
estuary, Texas. Also included is a spatiotemporal application to generate
predictions for sea surface temperatures adjusted for periodic climatic
effects from a long-term study region off southern California.
ACKNOWLEDGMENTS

My first introduction to spatial statistics was through a course offered at Rice University by Dennis Cox. I realized at that time that had these techniques been available to me while working as an oceanographer, the contributions made then would have been greatly enhanced. It is my hope that such contributions are forthcoming. My level of creativity in spatial statistics was greatly enhanced by faculty members working in other disciplines of statistics at Rice as well, particularly James Thompson, David Scott and Katherine Ensor whose courses in simulation, nonparametric function estimation and statistical time series modeling respectively are among the few that will remain with me as a professional statistician.

Even the most dazzling statistical results are useless without interpretation. For this task, Andre Droxler was of exceptional help in untangling an imperceptible confusion of ocean temperatures into a chronological documentary of long-term climatic fluctuation with numerous complex pieces.

Of particular recognition is Katherine Ensor who acted not only as chairman to this thesis committee, but also as mentor in every sense of the definition. Her ability to recognize the interplay of details and the all-inclusive picture simultaneously has a great influence on this work and future endeavors as well.

My sincerest gratitude is due to my wife Tamie, and my three children, Whitney, Kendal and Chandler who tolerated long hours of my absence both physically and in spirit while this work was in progress. I would also like to extend my gratitude to John Kelly, my brother, without whose continued support this thesis might have not been completed.
I dedicate this thesis to my father, whose lessons of the natural world shaped my
deepest curiosity and profound respect of the intricate workings of the universe. I also
dedicate this thesis to my mother who taught me of the rewards for hard work and
diligence. Their influence will be with me and my children always.

Partial support for this research was received from the National Science
Foundation, grant number DMS-9704648. Additional support was provided by the Rice
University Statistical Consulting Laboratory.
# Table of Contents

Chapter 1  
Introduction  

Chapter 2  
Spatial Dependence Modeling  

2.1 Framework  
2.2 Survey of Methods  
2.3 Comments  

Chapter 3  
Comprehensive Spatial Covariance Estimation  

3.1 Background on One-Dimensional Fourier-Bessel Series  
3.2 Spatial Covariance Modeling: The Fourier-Bessel Series Representation  
3.3 Estimation  
3.4 Simulation Results  

Chapter 4  
Spatiotemporal Dependence Modeling  

4.1 Framework  
4.2 Survey of Methods  
4.3 Discussion  

Chapter 5  
Comprehensive Spatiotemporal Covariance Estimation  

5.1 Background on Two-Dimensional Fourier-Bessel Series  
5.2 Spatiotemporal Covariance Modeling: The Fourier-Bessel Series Representation  
5.3 Estimation  

Chapter 6  
Application: Spatial Modeling of Toxic Sediment Load in Galveston Bay Estuary  

6.1 Data Description  
6.2 Covariogram Estimation  
6.3 Spatial Prediction  
6.4 Interpretation
Chapter 7
Application: Spatiotemporal Modeling of Sea Surface Temperatures off Southern California (CalCOFI) 157

7.1 Data Description 158
7.2 Covariogram Estimation 178
7.3 Spatiotemporal Prediction 186
7.4 Interpretation 189

Chapter 8
Conclusions 208

Appendix 213

A. SAS Code - Spatial Covariogram Estimation
B. SAS Code - Spatiotemporal Covariogram Estimation

References 219
PREFACE

In his book, George Christakos (1992) defines modeling as the fundamental process of combining intellectual creativity with physical knowledge and mathematical techniques in order to learn the properties of the mechanisms underlying a physical phenomenon and make predictions. This certainly typifies the endeavor of modern geostatistics. In the process of modeling, we seek to simplify the inherently complex. Our simplifications are not without purpose however in that our goal is to identify the most influential factors of an unknown process. This statement is the foundation that forms the differences between deterministic and stochastic modeling approaches. With deterministic modeling, the goal is to identify all influential factors recognizing that residual random variation is that collection of components we have yet to model. Stochastic modeling alternatively recognizes the residual random variation inherent in nature as that which yields little or no return on investment with respect to modeling effort. In most cases, we are nevertheless required to know fundamental properties of this random component such as low order moments and even distribution where available in order to characterize that which we have chosen not to include in the model itself.

In analyzing Christakos' definition, there are two goals to modeling: 1) to learn the properties of the mechanisms underlying a physical phenomenon and 2) make predictions. If we have accomplished these two objectives, our modeling effort is considered a success. We utilize some comparative set of measures in order to decide whether or not we have met the goals of modeling. In some sense, the two objectives are related in that we are unable to make accurate predictions without sufficient understanding and estimation of the mechanisms underlying the process being studied.
In the case of geostatistical modeling, there are two fundamental mechanisms required to meet the first objective. These are 1) the underlying spatial trend and 2) the dependence structure of the underlying process. Trend modeling has long been one of the major efforts of statistics and much of what has been developed over the last century in the area of trend estimation is directly applicable to geostatistical problems. Alternatively, dependence modeling is particular to geostatistics where typically, parametric spatial covariance models are fit to observed data and subsequently used in the application of spatial interpolation. In most instances, the uncertainty associated with estimation of the covariance model parameters is not reflected in the interpolation step. This has long been a potential source of critique with the common geostatistical paradigm. Few remedies have been suggested and most involve Bayesian estimation.

Extensions of purely spatial models into the spatiotemporal domain are becoming popular, but no single method stands out as superior in most applications. Instead, the techniques developed are largely specific to the problem. These techniques range from that of combining spatial and time series techniques to pure spatial extensions in the time domain. Spatiotemporal modeling however will continue to be an area of debated but contemporary research as this is a rich problem area and there remains much to be gained from its refinement. The practical applications of spatiotemporal modeling are numerous including but certainly not limited to earth sciences, epidemiology, economics and environmental engineering. However, be certain that problems and complexities inherent in pure spatial models are compounded by adding a temporal component. Assumptions valid in the spatial domain often have no meaning in the temporal domain and the converse is true as well.
The work described this thesis seeks to remove some of the objectivity inherent in spatial dependence modeling. The method described here may be used in place of uninformed or uncertain decisions but does not seek to absolve the geostatistical method from important informed determinations by researchers intimately familiar with the problem at hand. Certainly, as with any modeling effort, the conceptual design and relatedness to the problem in many cases sensibly precludes any attempt at automating the entire modeling effort.
1. INTRODUCTION

Covariance estimation is a fundamental part of spatial statistics. Proper estimation of the spatial dependence structure is important not only as an informative investigative tool, but also as a prerequisite to estimation of the underlying process at points where samples are not available. The estimation of spatial covariance however is not a straightforward process. Most of the literature on spatial prediction is not devoted to interpolation methods such as kriging, but instead to the process of dependence estimation that inarguably remains to be the most complex step in spatial modeling.

Consider the crucial process of fitting a theoretical covariogram to observations sampled from an unknown spatial process. First, there is the problem of finding the best empirical representation among the collection of techniques available. Besides the common method-of-moments variogram and covariogram (Cressie, 1993), we have the robust variogram (Cressie and Hawkins, 1980), relative variogram (Isaaks and Srivastava, 1989) and the non-ergodic covariogram (Goovaerts, 1997) to mention but a few examples. Each of these is applicable in specific circumstances and no single method is best in all cases. The decision on which method to apply is usually based on exploratory data analysis combined with conjectured knowledge of the studied process.

Next, consider the problem of optimum bin size determination usually required for computation of the empirical covariance. Although optimal data binning for histograms has been addressed in the literature (e.g. Scott, 1992), the specific problem of optimal binning of so-called spatial covariance clouds has largely been ignored. The researcher commonly applies the generalized rule of thumb binwidth originally suggested by Journel and Huijbregts (1978).
The task of correct parametric model selection is yet another source of uncertainty. The availability of generally adaptive parametric covariance models is limited. Although these models are used to approximate most empirical settings, there are almost always subtleties in the underlying spatial dependence function that known parametric models are unable to account for in part due to their restrictive structure. One example is a spatial dependence structure that exhibits attenuated but irregular periodicity. In such a case, a combination of variograms may be used under the assumption of a separable covariance structure. This technique is often applied, but is accompanied by potential problems. With the separable covariance strategy, it is often difficult to apriori select the best models to apply, leaving the researcher to trial and error model selection. Furthermore, the number of parameters increases with separable models and therefore the risk of parameter identifiability is present. Although parameter selection techniques exist for nonlinear models such as those common to spatial statistics, e.g. (Bunke and Droge, 1985), they are regularly used. Such methods may aid in proper fitting of hybrid separable models where a large number of parameters are involved.

One rarely derives a parametric covariance model that parsimoniously describes the underlying physical relationship between spatially indexed random variables. Christakos (1984) states that existing parametric spatial dependence families, while providing rich sources of models, are not totally satisfactory from a practical viewpoint. The researcher frequently encounters data that are not adequately approximated by any known spatial dependence model. Instead, some other function not included in the known models may offer a better fit. The investigator often uses an experimental model based upon knowledge of the underlying random field, regardless of meeting the
admissibility requirements only to generate meaningless or “embarrassing” (Cressie, 1993) negative prediction variances. In many cases encountered in practice, the functions fitted to experimental data are neither included among the known permissible models nor derived from them such that they are admissible.

Models are typically chosen from a set of known common choices based largely upon perceptible fit. In this context, the adaptation of commonly used models often postulates the true underlying process. For example, the exponential covariogram model describes a spatial first-order Markov process. Another example is the spherical covariogram that measures the volume of intersection between two spheres with Poisson generated random variables as centers. These and other models often found in the typical arsenal of spatial modeling tools unnecessarily place restrictions on our representation of the true model. They are often however the only available choice. Parsimonious dependence modeling becomes important in the context of estimation or prediction and simulation. For this reason, the use of an optimally fit function becomes tantamount to theoretical justification. In many circumstances, the practitioner is required to develop and employ new models that may or may not be valid spatial dependence functions even though they demonstrate a good fit to the observed data.

Parameter estimation routines for parametric covariance models are varied and it is often unclear as to the optimum method to use. McBratney and Webster (1986) compare weighted least squares, generalized least squares, and maximum likelihood methods for fitting sample covariance structures. Maximum likelihood covariance estimators, REML estimators and minimum variance unbiased estimators such as MINQUE and MVUE are described by Mardia and Marshall (1984), Kitanidis (1985),
Stein (1987) and Zimmerman (1989). Cressie (1985) describes a weighted least squares fitting technique that theoretically places more emphasis on model fit close to the origin and simultaneously reduces the importance of small bin sizes. We have found that the weighting criterion tends to minimize the loss function by optimizing for parameters estimates that inflate the model. Diamond and Armstrong (1983) show the prediction stage of the analysis to be reasonably insensitive to the dependence structure chosen. However, a misspecification of the covariance model is readily shown to bias the predictions. We should therefore proceed with fitting the best covariance model possible, free of conscious bias. We have in fact observed that a misspecification of the model type and components combined with improper estimation can have serious effects on prediction. Particular attention is required in the region of influence, specifically close to the origin where prediction neighborhoods are defined.

Researchers are typically faced with the problem of verification of the selected model in the sense of permissibility. Christakos (1984) describes a set of procedures useful in validating spatial dependence models. However, these techniques vary with the particular conjectured model and are often mathematically difficult to employ as the functional form of the model becomes increasingly complex. Often models are postulated that violate admissibility. One example is Carrol et al. (1997) where a spatiotemporal correlation function is employed to study the effects of ozone in Harris County, Texas. Cressie (comment, 1997) proves that the correlation function used is indeed inadmissible. Although Carrol et al. (rejoinder, 1997) support the applied model based on acceptable correlation estimates, Cressie states that this is indeed cause for concern and speaks to a more general issue in statistical modeling - namely, we should
always choose or estimate parameter values from the correct parameter space. We may otherwise never be sure that our inferences are valid.

The form of the model depends on assumptions made on the underlying process. For the assumption of second order stationarity, the covariogram is usually estimated while for the assumption of intrinsic stationarity, the semivariogram is the estimator of choice. Cressie (1993) makes the case in favor of semivariogram estimation, as the class of intrinsically stationary processes is more general. In addition, the semivariogram is estimated more reliably than the covariogram due to unbiasedness of the semivariogram estimator. However Christakos (1984) suggests the covariogram to be simpler to estimate in many cases because the covariogram eventually goes to zero in the limit as distance becomes infinite. For the variogram, the current method of nugget estimation involves the extrapolation of a fitted parametric model to the zero distance. The properties of the extrapolated nugget estimator have not been extensively explored in the literature and are thus questionable. The limiting property of the covariogram will simplify the derivation of our comprehensive model as will be discussed. Although covariogram estimation is described herein, estimation of the semivariogram may similarly be accomplished, but precluded by nugget and sill approximation.

A comprehensive approach to modeling the spatial covariance structure would provide potential benefits beyond that of nullifying the requirement of assuming a parametric model class and functional form such as separability. Variograms that are difficult to model with conventional parametric techniques due to their empirical complexity could easily be accommodated by adaptable counterparts. The anticipated improvement in representation of the underlying spatial dependence is expected to at
least provide more accurate estimates at nonmeasured points. This is critical since the estimated variogram is applied as deterministic in the prediction equations. Perhaps most importantly, an automatic method of generating permissible nonparametric spatial covariance estimates would preclude the requirement of validation with respect to admissibility that accompanies most parametric models. Christakos (1984) describes techniques applicable to the validation of parametric spatial covariance structures. This most often involves taking the Fourier transform of the candidate model and modifying the parameter space such that the spectral density is nonnegative in the case of covariogram estimation and conditionally negative definite in the case of variogram estimation. Further validation must be made with respect to the dimensionality of the sampling space. For example, a model valid in $\mathbb{R}^1$ is not necessarily valid in higher dimensional space. Most often, candidate models are not validated by the researcher with respect to permissibility, but only legitimized by the behavior of the computed prediction variances. The danger in this practice is that the model is only useful for the data and predictions to which it was applied. Further prediction or addition of data followed by prediction or interpolation could potentially result in negative prediction variances. Perhaps most importantly however is the potential for numerical singularities in the estimated covariogram or variogram matrix that could result in unreliable predictions.

The condition of positive semidefiniteness of the covariogram is by definition required for the valid derivation of prediction variances. However, as it is defined, positive semidefiniteness is not readily implementable. The theorem by Bochner (1955) provides a methodological approach for applying this permissibility condition. Work by Armstrong and Diamond (1984) and Christakos (1984) demonstrate the applicability of
Bochner's theorem in checking for permissibility of parametric covariance models.
Results derived within these papers for the parametric case also provide a basis for the
development of permissible nonparametric estimators. Many of the described procedures
in the aforementioned papers are direct applications of Bochner's theorem or results
attributable to the indirect application of Bochner's theorem.

Our motivation herein is therefore to design an adaptable estimator of spatial
covariance such that the researcher is not required to make unnecessary or even invalid
assumptions regarding the underlying spatial process. Such a general estimator should
conform to practical spatial dimensions and automatically satisfy the permissibility
criteria of spatial dependence models while possessing the property of being flexible
enough to model a wide range of processes. We postulate that the true physical model
behind most spatial processes is unknown. Furthermore, there is no assertive proof that
even many common spatial dependence models exist in nature. We therefore most often
are forced to rely on mathematical surrogates that behave in accordance with the
observed process. To this end, an optimal general model is one that exhibits a large
degree of conformity to the observed process. We suggest a spatial covariance estimator
based upon a Fourier-Bessel series expansion of the unknown isotropic and second-order
stationary covariance $c(r)$. We will show that the Fourier-Bessel series provides a
natural basis for the spectral representation of $c(r)$ in terms of the spectrum $C(\omega)$.
Instead of assuming a specific parametric form of $c(r)$, we alternatively propose a large
class of estimators of $C(\omega)$ such that $c(r)$ has the positive-definite property.
2. SPATIAL DEPENDENCE MODELING

A considerable amount of work has been done over the last two decades in the area of spatial dependence modeling. A number of areas however remain to be adequately studied. Increasingly, the science of spatial dependence modeling is seen to draw important results from other disciplines of statistics and mathematics. Certainly as these lateral areas are developed, so will be spatial dependence modeling. The purpose of this chapter is to lay the groundwork for existing techniques. Section 2.1 outlines definitions and describes the fundamental problem in terms of a random sample observed from the unknown spatial random field. The survey contained in Section 2.2 will primarily focus on efforts to move away from classes of parametric models. Work toward development of a comprehensive covariance model for dependent data has focused almost exclusively on kernel regression techniques. Even then, the bulk of work is in parameter-free estimation of autocorrelation functions for time series data. There is a shortage of techniques available specifically for nonparametric or comprehensive spatial covariance models. We describe those few methods available for the purely spatial context below. However, included is a review of available techniques for nonparametric time series autocorrelation function estimation as these methods may be broadened to include the spatial case as well.

The problem of admissibility of the spatial covariance estimator is essentially what makes the problem of model derivation non-trivial. This places a constraint on the sign of the spectral density inclusive of the entire transform domain. Furthermore, the basic definition of admissibility is cumbersome to work with and we instead resort to an important result by Bochner (1955). The problem is similar to that of nonnegativity of
nonparametric density estimators, only the nonnegativity constraint applies to an alternate space, that being the frequency domain. The problem of admissibility is indeed fundamental in that it refers itself to the more important issue of covariance estimation drawn from the class of true covariance models.

2.1 Framework

We begin with a spatially indexed sample $z(s_i), s_i = \{s_{i,1}, \ldots, s_{i,d}\}, i = 1, \ldots, n$ observed in some $d$-dimensional region of interest. We furthermore assume that $z(s_i), i = 1, \ldots, n$ is a single realization of an underlying random process or spatial random field $Z(s), s \in \mathbb{R}^d$. In order to formalize the concept, the following definition is stated.

**Definition 2.1:** Let $(\Omega, F, P)$ be a probability space where $\Omega$ is the sample space, $F$ is the $\sigma$-field of subsets of $\Omega$, and $P$ is a probability measure on the measurable space $(\Omega, F)$. A spatial random field $Z(s), s \in \mathbb{R}^d$ is a collection of random variables $\{z_1(s_1), z_2(s_2), \ldots\}$ such that each random variable in the collection is defined on the probability space $(\Omega, F, P)$ and maps to the measurable space $(\mathbb{R}^d, \mathcal{B}^d)$ where $\mathcal{B}^d$ is a $\sigma$-field of subsets of $\mathbb{R}^d$.

Although Definition 2.1 is consistent throughout most of the literature, other definitions of the spatial random field are available more in the context of the classical definition of a random variable where each sampled value $z(s)$ represents one draw from the random
variable, a function \(Z(s)\) (Christakos, 1992). Definition 1 is more in agreement with Wackernagel's (1995) description of the *regionalized variable* in that a subset of objects are drawn from an infinite collection of objects.

The spatial random field \(Z(s)\) is defined through the finite dimensional distribution function,

\[
F_{s_1,...,s_n}(z(s_1),...,z(s_n)) = P(z(s_1) \leq Z(s_1),...,z(s_n) \leq Z(s_n))
\]  

(2.1)

for any \(n\). Furthermore the distribution in (2.1) is assumed to satisfy the following two conditions, again for any \(n\):

(a) *Symmetry condition:*

\[
F_{s_{(1)},...,s_{(n)}}(z(s_{(1)}),...,z(s_{(n)})) = F_{s_1,...,s_n}(z(s_1),...,z(s_n))
\]

(2.2)

where \((1),..., (n)\) is any permutation of \(1,..., n\).

(b) *Consistency condition:*

\[
F_{s_1,...,s_n,s_{n+1},...,s_{1}}(z(s_1),...,z(s_n),\infty,...,\infty) = F_{s_1,...,s_n}(z(s_1),...,z(s_n))
\]

(2.3)

The most common case is for \(Z(s)\) distributed as multivariate normal, which we denote by \(Z(s) \sim \mathcal{N}_n(\mu, \Sigma)\) where \(\mu\) is \(n\times1\) and \(\Sigma\) is \(n\times n\). The off-diagonal elements of \(\Sigma\) characterize the spatial dependence between pairs of points from different locations, \(s_i\) and \(s_j\) where \(i \neq j\). Given the existence of the probability density function, it is given by

\[
f_{s_1,...,s_n}(z(s_1),...,z(s_n)) = \frac{\partial^n}{\partial z(s_1)\partial z(s_2)\cdots\partial z(s_n)} F_{s_1,...,s_n}(z(s_1),...,z(s_n)).
\]

(2.4)
We are now in a position to define the two most common second-order moment functions used in spatial statistics in terms of expectation operators.

**Definition 2.2:** Given the spatial random field \( Z(s_i), i = 1, \ldots, n \) and

\( s_i = (s_{i,1}, \ldots, s_{i,d}) \in \mathbb{R}^d \) where \( Z(s_i) \) has probability density function

\[ f_{s_1,\ldots,s_n}(z(s_1), \ldots, z(s_n)) \] and \( E(\cdot) \) is the expectation operator and \( \text{var}(\cdot) \) is the variance operator, the conventional spatial covariance function or **covariogram** is defined as

\[
c(s_i, s_j) = E\left(Z(s_i)Z(s_j)\right) - E\left(Z(s_i)\right)E\left(Z(s_j)\right)
\]  \hspace{1cm} (2.5)

and the **variogram** is defined as

\[
\gamma(s_i, s_j) = \frac{1}{2} \text{var}(Z(s_i) - Z(s_j))
\] \hspace{1cm} (2.6)

In the above definition, the covariogram \( c(s_i, s_j) \) and variogram \( \gamma(s_i, s_j) \) are functions of definitive locations. In practice, it is common that the relationships between pairs of observations of the spatial random field are not necessarily location dependent. Some variations on this rule might include location invariance, e.g. the dependence between a pair of observations is a function of the distance \( h = |s_i - s_j| \) between the pair and not necessarily the location. This is defined as **stationarity**. Another aspect is that of rotational invariance. This assumes that a pair of observations collected from the spatial random field depend on radial distance \( r \) where \( r = |h| \) and not on relative direction. A radially invariant spatial dependence function is a function of distance alone. Due to directional invariance, the assumption of **isotropy** allows for a simpler representation and
mathematical tractability of spatial dependence. The validity of the isotropic assumption must be empirically established however as described in Cressie (1993) among others. We operationalize these concepts in the following definitions:

**Definition 2.3:** The random field \( Z(s) \) is defined as *second-order stationary* if the following two conditions are met,

1) \( E(Z(s)) = \mu \) for all \( s \in \mathbb{R}^d \) \hspace{1cm} (2.7)

2) \( c(s_i, s_j) = c(s_i - s_j) \) for all \( s_i, s_j \in \mathbb{R}^d \) \hspace{1cm} (2.8)

Furthermore, the random field \( Z(s) \) is defined as *intrinsically stationary* if the following two conditions are met,

1) \( E(Z(s)) = \mu \) for all \( s \in \mathbb{R}^d \) \hspace{1cm} (2.9)

2) \( \gamma(s_i, s_j) = \gamma(s_i - s_j) \) for all \( s_i, s_j \in \mathbb{R}^d \) \hspace{1cm} (2.10)

**Definition 2.4:** For the random field \( Z(s) \) where \( s \in \mathbb{R}^d \), the covariogram function defined in (2.5) is called *isotropic* if \( c(s_i - s_j) = c(\|s_i - s_j\|) \). Likewise, the variogram defined in (2.6) is *isotropic* if \( \gamma(s_i - s_j) = \gamma(\|s_i - s_j\|) \).

We will use the notation \( c(h) \) and \( c(r) \) to denote stationary and isotropic covariograms respectively. Note that by symmetry, \( c(h) = c(-h) \) but for example considering a planar process, \( c(-s_1, s_2) \) may be different from \( c(s_1, s_2) \).
The necessity of Fourier transformation of the covariogram is apparent with Bochner’s theorem. The covariance models are real and even functions as are their \( d \)-fold Fourier transforms when they exist.

**Definition 2.5:** The Fourier transform pair with respect to the covariogram \( c(h) \) is

\[
C(w) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp\{-iw^T h\} c(h) \, dh
\]

\[
c(h) = \int_{\mathbb{R}^d} \exp\{-iw^T h\} C(w) \, dw
\]

where \( d \)-fold integration is implied and \( h = \{h_1, \ldots, h_d\} \in \mathbb{R}^d \) and \( w = \{w_1, \ldots, w_d\} \in \mathbb{R}^d \).

If the spatial random field \( Z(s) \), where \( s \in \mathbb{R}^d \) is isotropic, e.g. \( c(h) = c(r) \), then the corresponding spectrum is isotropic as well where \( |w| = \omega \) and is written as \( C(w) = C(\omega) \). The isotropy assumption allows for a change of support space from cartesian coordinates in equations (2.11) and (2.12) to more convenient spherical coordinates. This transformation reduces the vector functions with domain \( \mathbb{R}^d \) to scalar functions with domain \( \mathbb{R}^1 \) as outlined in the following propositions.

**Proposition 2.1:** (Adler, 1981) Let \( Z(s) \) be an isotropic random field with \( s \in \mathbb{R}^d \).

Then the spatial covariance \( c_d(r) \) may be written as
\[ c_d(r) = (2\pi)^{d/2} \int_0^\infty \frac{J_{(d-2)/2}(\omega r)}{(\omega r)^{(d-2)/2}} C_d(\omega) \omega^{d-1} d\omega \] (2.13)

**Proposition 2.2:** (Adler, 1981) Let \( Z(s) \) be an isotropic random field with \( s \in \mathbb{R}^d \).

Then the spectrum \( C_d(\omega) \) may be written as

\[ C_d(\omega) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \frac{J_{(d-2)/2}(\omega r)}{\omega r} c_d(r) r^{d-1} dr \] (2.14)

The preceding propositions are readily proven by transformation from \( h \) and \( w \) into \( d \)-dimensional spherical coordinates. When isotropy is assumed, we will use the subscript \( d \) in the covariogram and corresponding spectrum to denote support space. In order for \( c_d(r) \) to be a valid covariance function, it is necessary and sufficient that \( c_d(r) \)

admits a representation of the form (2.13) where \( C_d(\omega) \) is a nonnegative bounded function.

When the variation is bounded, the variogram is equivalent to the covariogram function. That is to say, a second order stationary random field is easily shown to be intrinsically stationary. However, the converse is not always true. In the case of a second order stationary random field, it follows from (2.5) and (2.6) that the following holds.

\[ \gamma(h) = c(0) - c(h), \] (2.15)
where \( c(0) = E \left( (Z(s) - \mu_Z)^2 \right) \) and \( h = [h_1, \ldots, h_d] \subseteq \mathbb{R}^d \) is the process variance and 
\( E(Z(s)) = \mu_Z \). Equation (2.15) does not hold in general for an intrinsic random field.

In certain fields, for example hydrology, the researcher is faced with modeling an assumed nonstationary random field. The hypothesis of intrinsic stationarity defined above is then extended to include more sophisticated covariance models, namely generalized covariances. Although they will not be used in the development of our comprehensive covariance estimator, their existence warrants definition in that extensions to these models may be easily made.

**Definition 2.6:** A \( \kappa \)th order intrinsic random function or \( IRF - \kappa \) is a random field \( Z(s) \) where \( s \subseteq \mathbb{R}^d \), that requires a \( \kappa + 1 \)-th order increment to achieve stationarity.

More specifically, the \( \kappa \)-th order increment is the linear combination

\[
Y = \sum_{i=1}^{n} \lambda_i Z(s_i)
\]

with weights \( \lambda_i , i = 1, \ldots, n \) that satisfy

\[
\sum_{i=1}^{n} \lambda_i (s_{i,b}^h, \ldots, s_{i,b_d}^h) = 0
\]

for all integers \( b_1, \ldots, b_n \geq 0 \) such that

\[
\sum_{i=1}^{d} b_i \leq \kappa
\]

and \( s_i = (s_{i,1}, \ldots, s_{i,d}) \subseteq \mathbb{R}^d \) and \( i = 1, \ldots, n \). We then define the \( \kappa \)-th order generalized covariance \( k(s_i - s_j) \) as the stationary part of the nonstationary covariance, \( c(s_i, s_j) \).
The preference of model, $\gamma(h)$, $c(h)$ or $k(h)$ where $h = |s - (s + h)|$ is dependent upon the assumption imposed on the underlying spatial process $Z(s)$. For an intrinsically stationary process, many researchers acknowledge that $\gamma(h)$ is the preferable model over $c(h)$. Namely, the variogram is mean-free. For a second-order stationary process, the covariogram is often the preferred model although the class of covariograms is included in the class of all variograms.

This paper will deal only with processes linked to the intuitive idea of stationarity and hence will not include the class of generalized covariances discussed above. Furthermore, we will deal mostly with isotropic covariance functions since this assumption does away with awkward multiple integrations when computing the Fourier transform. This is an easy assumption to validate and isotropy is most often the case in practice. We will emphasize modeling of the isotropic covariogram, $c(r)$ in the remainder of this thesis although the extension to the case of the isotropic variogram $\gamma(r)$ is straightforward and can be found in Cressie (1993) for example. We proceed by laying ground rules for the admissibility of spatial covariance functions.

**Proposition 2.3:** A continuous covariance $c(h)$ is admissible if and only if $c(h)$ is nonnegative definite, i.e.

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j c(h) \geq 0 \text{ for all } h = |s_i - s_j| \tag{2.16}
$$
In practice, Proposition 2.3 is difficult to implement. Although it serves as a meaningful theoretical definition of the nonnegative definiteness property, it is generally impractical to apply in most operational circumstances. A practical alternative method for verification of admissibility is offered by Bochner’s Theorem stated as follows,

**Theorem 2.1:** (Bochner, 1955) A continuous function, \( f(u) \) where \( u \in \mathbb{R}^d \), is nonnegative definite if and only if it can be expressed as the Fourier transform of a nonnegative symmetric bounded measure \( G(w) \) where \( w \in \mathbb{R}^d \), i.e.

\[
    f(u) = \int_{\mathbb{R}^d} \exp(iw'u) dG(w) \tag{2.17}
\]

Alternatively if \( G(dw) \) can be written as \( g(w) dw \), then (2.17) is identical to writing

\[
    f(u) = \int_{\mathbb{R}^d} \exp(iw'u) g(w) dw, \tag{2.18}
\]

where \( g(w) \geq 0 \) for all \( w \)

\[
i = \sqrt{-1} \quad \text{and} \quad w'u \quad \text{is the inner product} \quad \sum_i w_i u_i.
\]

In words, Bochner’s theorem states that any positive definite function \( f(\cdot) \) corresponds to the covariogram of a second order stationary stochastic process. This statement supported by Bochner’s theorem lays groundwork for the development of techniques to construct valid covariograms. Although other methods not based on (2.16) exist, these are almost always in reference to restricted situations. Alternate procedures to validate a proposed variogram or covariance model for nonnegative definiteness are generally
difficult to apply. Christakos (1984) discusses a set of methods for validation of conjectured covariance models. Many of these involve direct application of Bochner's theorem. In practice, known positive definite functions are fit to the empirical covariance. The parameter space for these known models is well established. Hybrid models can be generated from known models by utilizing the closure properties of admissible covariance functions. These are stated as follows:

**Proposition 2.4:** (Cressie, 1993) Closure Properties of $c(\cdot)$

i) An admissible isotropic model, $c(r)$ where $r = |\mathbf{h}|$ and $\mathbf{h} = \{h_1, \ldots, h_d\} \in \mathbb{R}^n$ is admissible in $\mathbb{R}^m$ where $m \leq n$. The model is not necessarily admissible where $m > n$ however.

ii) If $c_1(\mathbf{h})$ and $c_2(\mathbf{h})$ are valid models for $\mathbf{h} = \{h_1, \ldots, h_d\} \in \mathbb{R}^d$, then

$$c(\mathbf{h}) = c_1(\mathbf{h}) + c_2(\mathbf{h})$$

is a valid model for $\mathbf{h} = \{h_1, \ldots, h_d\} \in \mathbb{R}^d$.

iii) If $c_1(\mathbf{h})$ is a valid model in $\mathbf{h} = \{h_1, \ldots, h_d\} \in \mathbb{R}^d$ and $b > 0$ then $c(\mathbf{h}) = bc_1(\mathbf{h})$

is a valid model for $\mathbf{h} = \{h_1, \ldots, h_d\} \in \mathbb{R}^d$.

iv) If $\{c_i(h_i), \ldots, c_d(h_d)\}$ are admissible covariograms for $h_i \in \mathbb{R}^i$, $i = 1, \ldots, d$ then

$$c(\mathbf{h}) = \prod_{i=1}^{d} c_i(h_i)$$

is an admissible covariogram in $\mathbf{h} = \{h_1, \ldots, h_d\} \in \mathbb{R}^d$. A covariance $c(\mathbf{h})$ written in this form is called **separable**.
Although a variety of empirical covariance representations exist, we will focus on the most common of these, the method of moments estimator. The assumption with this estimator is that the data can be modeled as a second-order stationary process.

**Definition 2.7:** For regularly spaced observations, or observations taken on a regular lattice, the *method of moments covariogram estimator* is defined as

\[
\hat{c}(\mathbf{h}) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - \bar{Z})(Z(s_j) - \bar{Z})
\]

(2.19)

where

\[
\bar{Z} = \frac{1}{n} \sum_{i = 1}^{n} Z(s_i) \text{ and } N(h) = \{(s_i, s_j) : |s_i - s_j| = h; i, j = 1, \ldots, n\}
\]

and \(|N(h)|\) is the number of distinct pairs in \(N(h)\). For irregularly spaced observations, the *method of moments covariogram estimator* is defined as

\[
\hat{c}(h(l)) = \frac{1}{N(h)} \sum_{N(h)} w_{i,j} \left\{ (Z(s_i) - \bar{Z})(Z(s_j) - \bar{Z}) : (i, j) \in N(h); h \in T(h(l)) \right\}
\]

(2.20)

where \(|N(h)|\) is the number of distinct pairs in \(N(h)\) and \(T(h(l))\) is a specified neighborhood of \(h(l) \in \mathbb{R}^d, l = 1, \ldots, k\).

The estimators in (2.19) and (2.20) are well known to have \(O(1/n)\) bias, e.g.

\[E(\hat{c}(h)) = c(h) + O(1/n)\] as described for example in Fuller (1996). Furthermore, the bias could contribute substantially to the mean square error when \(n\) is small (Cressie, 1993). The covariance estimator in (2.20) is also influenced by the selection of
neighborhood width $T(h(l))$. No method of optimizing $T(h(l))$ with respect to minimization of some appropriate loss function has been suggested. The selection of a neighborhood width is chosen in practice such that $|N(h)| \geq 30$ as proposed by Journel and Huijbregts (1978). We will later propose an estimator of $\hat{c}(h)$ that is unbiased and does not require binning as does the common estimator in (2.20) or indeed most of the empirical estimators currently in use. We do not advocate dependence estimation from unbinned empirical covariograms in all situations as binning serves to enhance computational efficiency. With small sample sizes however, binning tends to oversmooth the representative empirical covariogram or may indeed be impractical to perform given the generally accepted bin width of $|N(h)| \geq 30$.

2.2 Survey of Methods

Spatial covariance is classically modeled as the variogram or covariogram function depending upon the hypothesis of stationarity. The models used are typically parametric and most commonly nonlinear in the parameters. In addition, most parametric models assume stationarity of the underlying random field and as such are functions of the distance $h$. Examples of frequently used parametric families of variogram functions are given by Cressie (1993) and generally involve models of the form

$$c(h; \theta) = f(h; \theta) + \varepsilon(h),$$

(2.21)

where $\theta$ is a $p \times 1$ parameter vector and $\theta \in \Theta_p$ for $p$-dimensional parameter space $\Theta$.

In most cases, the parameter space $\Theta_p$ is constrained so that the model (2.21) is
admissible. The error term $\varepsilon(h)$ is assumed to be distributed with zero mean and some variance, itself a function of the distance vector $h$.

Undoubtedly the most common method of covariogram modeling is that of selection and fitting of one of a number of known admissible parametric covariogram and associated variogram model families of the form (2.21). Numerous references are available that describe these model families including Cressie (1993), Goovaerts (1997) and Isaaks and Srivistava (1997) to mention a few. In addition, a number of parametric spatial covariance models are widely available in current geostatistical software packages. As previously noted, this chapter will not include a description of techniques for parametric model selection and fitting as this topic is extensively addressed in the literature. Rather, we will focus on data-driven techniques that utilize nonparametric or otherwise unspecified models. Included are nonparametric estimators of time-series covariances as these are often applicable to the spatial case as well.

Efforts at developing spatial or temporal covariance models that can be generally applied to a wide range of observable random processes have been intended largely for the extension of nonparametric regression methods mostly involving the application of kernel type smoothers, in particularly kernel regression. The method of kernel regression seems applicable in the case of general spatial dependence modeling, however some problems arise with this approach. The construction of the spectrum requires evaluation of the associated spatial dependence function $c(h)$ at all values of the domain $h$ where $h \in \mathbb{R}^d$. Kernel type smoothers only allow evaluation of the function $\hat{c}(h)$ within the domain of the observed data. For variogram estimation in particular, difficulties may arise in extrapolation of the region between the smallest sampled distance and zero.
Therefore, the region within a neighborhood of $h = 0$ is not efficiently estimated. This includes estimation difficulties with the nugget. Attempts at using boundary kernels may provide some remedy, but the situation is largely controlled by the limited ability of kernel smoothers to extrapolate beyond the data range.

Simple interpolation or even smoothing of the empirical variogram is risky in that the empirical representation of the covariance structure is not guaranteed to be admissible. However, the fundamental idea behind nonparametric covariogram or semivariogram estimation is basically the application of a smoothing mechanism to the empirical representation. For this purpose, nonparametric covariance estimation is an important diagnostic tool (Diggle and Verbyla, 1998).

The method described by Shapiro and Botha (1991) is the only one specifically applied to the estimation of an admissible spatial covariance. Their method is based on constrained curve fitting through point estimates of the variogram of an assumed stationary process. The objective function minimizes the fit between the empirical variogram and a discrete representation of the theoretical Fourier transform where the spectral density is defined as a step function. The problem is restricted to processes of second order stationarity or where the variogram is a bounded function. In addition, the method deals only with isotropic spatial covariance functions. This allows the multidimensional $n$-fold Fourier transform to be written as a single integral as will be described below. Although Shapiro and Botha’s method produces a variogram estimator that is conditionally negative definite on a discrete set, it does not necessarily retain that property in the continuum. The spectral distribution function is approximated by a step function although Pilz, Spoeck and Schimek (1997) suggest approximation by piecewise
linear functions as an improvement. The method is relatively straightforward although numerical optimization is a required step. The estimated variogram is often bumpy, perhaps a symptom of the step function approximation to the spectrum. For the one-dimensional isotropic case, from Bochner's theorem, \( c_z(r) \) is continuous and positive definite if representable as

\[
c_z(r) = \int_0^\infty \cos(rt) dF(t),
\]

(2.22)

where \( F(t) \) is a bounded monotonically increasing function on \([0, \infty)\). Assuming second order stationarity and the covariogram representation in (2.22), the weighted least squares fitting problem consists of finding \( c_z(0) \) and \( F(t) \) that minimize

\[
\sum_{i=1}^n w_i \left[ \hat{r}_i - c_0 + \sum_{j=1}^m \cos(rt_j) y_i \right]^2
\]

(2.23)

subject to

\[
y_j \geq 0, \ j = 1, \ldots, m, \ \text{and}
\]

\[
c_0 - \sum_{j=1}^m y_j \geq 0.
\]

The optimization problem in (2.23) is further operationalized numerically by allowing \( F(t) \) to be a step function with a finite number of positive jumps and thus reducing the integration in (2.22) to a summation over the jumps shown in (2.23). The problem is readily extendible to higher dimensions.

Diggle and Verbyla (1998) demonstrate a local polynomial smoother of the variogram as a diagnostic tool. They employ kernel weighted local linear regression as a method of producing a variogram estimate with optimum boundary properties. However,
since the resulting variogram estimate is not used for prediction and subsequently for computation of prediction errors, there is no need to assure the estimator is conditionally negative definite. An optimal bandwidth is derived and the technique suggests the interesting potential use of kernel weighted local polynomial regression estimators of the spatial covariance. Specifically, the authors formulate a nonstationary nonparametric estimate of a 2-dimensional temporal semivariogram. The longitudinal data used are represented as

\[(y_{ij}, t_{ij}), \quad j = 1, ..., n; i = 1, ..., m.\]

where \(y_{ij}\) is the \(j^{th}\) of \(n\) measurements on the \(i^{th}\) of \(m\) subjects and \(t_{ij}\) is the time of measurement of \(y_{ij}\). The authors select a data driven optimal bandwidth by optimizing the total mean square error,

\[
TMSE(h) = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{m} E \left[ \left( \hat{Y}(t_{ij}, t_{ik}) - y(t_{ij}, t_{ik}) \right)^2 \right].
\]

(2.24)

As stated above, an effort at validating the derived semivariogram structure is not undertaken because its use is purely diagnostic and not predictive.

Hyndman and Wand (1997) similarly describe a local polynomial estimator of the autocovariance function of a nonstationary time series. Assume the sample of random pairs, \((X_i, Y_i), ..., (X_n, Y_n)\) to follow the non-stationary correlated errors model,

\[Y_i = m(X_i) + \epsilon_i\]

(2.25)

where \(\text{cov}(\epsilon_i, \epsilon_j) = 0\) for \(|i - j| > \nu\). As vectors, sample and error are \(Y = [Y_1, ..., Y_n]^T\).

\[X = [X_1, ..., X_n]^T\] and \(e = [\epsilon_1, ..., \epsilon_n]^T\) so that the mean function \(m\) is expressed in the
conditional expectation $E[Y | X] = [m(X_1), ..., m(X_n)]^T$ and the $j$ th lag autocorrelation is

$$\gamma_j = [\text{cov}(\epsilon_i, \epsilon_{i-j}), ..., \text{cov}(\epsilon_i, \epsilon_{n-j})].$$

The error covariance matrix is written as

$$V = \text{cov}(\epsilon | X) = \text{diag}(\gamma_0) + \sum_{i=1}^{\infty} \{B_i \text{diag}(\gamma_i) + B_i^T \text{diag}(\gamma_i)\} \quad (2.26)$$

where $B_i$ is the $i$ th order backshift operator, a sub-diagonal matrix with 1's on the $i$ th sub-diagonal and $B_0 = I$, the identity matrix. As an initial step, the authors suggest fitting an arbitrary mean function to the data using any of the general class of estimators of the mean $m$ where $\hat{m} = [\hat{m}(X_1), ..., \hat{m}(X_n)]^T$ of the form

$$\hat{m} = S_1Y. \quad (2.27)$$

$S_1$ is an $n \times n$ data smoother matrix depending on $X$. Specifically, $\hat{m}$ is any linear smoother, e.g. OLS regression, moving average, spline, local polynomial etc. The vector of residuals is then

$$r = (I - S_1)Y. \quad (2.28)$$

It follows then that the $j$ th lag autocorrelation is

$$\gamma_j = E[e \times (B_j e)] \quad (2.29)$$

where $\times$ is taken in context to be the element-wise product of two equally sized matrices.

The proposed estimator of $\gamma_j$ is then

$$\hat{\gamma}_j = S_2 \{r \times (B_j r)\} \quad (2.30)$$
where $S_2$ is the smoother matrix of the lag residuals $r(B\cdot r)$. For the case of $S_2 = \{1/n\}$, the stationary case or usual estimator of the autocorrelation function is

$$
\hat{\gamma}_j = \frac{1}{n} \sum_{i=j+1}^{n} r_i r_{i-j}.
$$

(2.31)

The technique is demonstrated with $S_1$ and $S_2$ as $p$th degree local polynomial smoother matrices where that the $(i, j)$th entries are

$$
\{S_{p,h}\}_{i,j} = e_i^T \left\{ X_p (X_i)^T W_h (X_i) X_p (X_i) \right\}^{-1} X_p (X_i)^T W_h (X_i) e_j
$$

(2.32)

where $e_i$ is the column vector with 1 in the $i$th position and zeros elsewhere,

$$
X_p (x) = \begin{bmatrix}
1 & X_i - x & \cdots & (X_i - x)^p \\
\vdots & \vdots & \ddots & \vdots \\
1 & X_n - x & \cdots & (X_n - x)^p 
\end{bmatrix}
$$

and

$$
W_h (x) = \text{diag} K \left( \frac{X_i - x}{h} \right).
$$

$K(\cdot)$ is the kernel and $h$ is the smoothing parameter. Statistical properties and bias are derived by the authors along with an application to earthquake data. Their method is potentially adaptable to nonstationary spatial observations with binned distances, however in the spatial case the reduction in sample size may increase the variance of the estimator. Furthermore, for the nonstationary case the estimator (2.30) is not necessarily positive definite as this property is sacrificed for the simplicity of the nonparametric estimator. As discussed above, variogram estimation at the nugget is potentially unstable due to the extrapolation limitations of nonparametric estimators.
Similar in nature to Shapiro and Botha (1991), Hall, Fisher and Hoffmann (1994) proposed a technique of modifying a kernel type smoother of the empirical covariance structure such that the spectrum is nonnegative. The method requires deletion of negative lobes or truncation and slope adjustment along with perhaps additional smoothing in the spectral domain. The technique only addresses covariance structures in one dimension, but extension to the estimation of spatial structure functions is suggested. The necessary adjustments made in the spectral domain may be confounded in higher dimensions, particularly for \( d > 2 \). Their method is clearly an operational demonstration of the procedure originally suggested by Masry (1983) described below. The technique of Hall, Fisher and Hoffmann (1994) assumes a stationary process \( Z(t) \). The objective is to estimate \( c_z(\tau) = \text{cov}(Z(t), Z(t+\tau)) \). The kernel estimator is

\[
\hat{c}_z(\tau) = \frac{\sum_i \sum_j \hat{Z}_{ij} K\left\{ (\tau-t) / h \right\}}{\sum_i \sum_j K\left\{ (\tau-t) / h \right\}},
\]

(2.33)

where \( \hat{Z}_{ij} = (Z(t_i) - \bar{Z})(Z(t_j) - \bar{Z}) \), \( \bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z(t_i) \), and \( t_{ij} = t_i - t_j \). By Bochner's theorem, the positive semidefiniteness property is equivalent to nonnegativity of the Fourier transform of \( \rho \), i.e.,

\[
\rho_F(\theta) \geq 0 \quad \forall \theta,
\]

where

\[
\rho_F(\theta) = \int_{-\infty}^{\infty} \rho(t) e^{i\theta t} dt = 2 \int_{0}^{\infty} \rho(t) \cos(\theta t) dt,
\]

(2.34)
and $\rho_F(\cdot)$ is the Fourier transform of $\rho(\cdot)$. In order to produce a covariance function satisfying the positive semidefiniteness property, the authors propose the following:

1. Compute the Fourier transform $\hat{\rho}_F(\cdot)$ of $\hat{\rho}(\cdot)$.

2. Render $\hat{\rho}_F(\cdot)$ nonnegative by additional smoothing or removal of negative lobes. Call the resulting function $\tilde{\rho}_F(\cdot)$.

3. Fourier invert $\tilde{\rho}_F(\cdot)$ to obtain a new positive semidefinite function $\tilde{\rho}(\cdot)$.

In this case, the bandwidth $h$ is selected with respect to asymptotic optimality of the estimator. Other selection criteria may be applied as well. Although not explicitly demonstrated, the authors suggest extension of the method to nonparametric spatial variogram estimation. The method is subjective in the sense that it requires visual deletion of negative parts of the estimated spectrum following bandwidth selection and fitting of the observed autocorrelation. This suggests the resulting spectrum is admissible only within the fitted distance.

Masry (1983) described estimation of a stationary continuous time process from irregularly spaced sampling instants. The method assumes a Poisson sampling rate and utilizes a kernel type smoother demonstrated in the method of Hall, Fisher and Hoffmann (1994) described above. Although the technique is not extended to the estimation of spatial covariance structure, consistent estimation of arbitrary covariance functions is described. The corresponding extension to spectral estimation was developed in Masry (1978) suggesting that the method may be modified such that the corresponding spectrum is nonnegative. Similar work by Scargle (1982, 1989) also deals with irregularly spaced sampling. Although the method is specifically used for spectral estimation from
irregularly spaced sampling in one dimension, Scargle demonstrates nonparametric estimation of the autocorrelation function by inverting the estimated spectral density.

2.3 Comments

The methods discussed in this chapter primarily deal with the problem of covariance fitting once the empirical estimator has been determined. At this stage, the investigator assumes an adequate representation of the dependence structure based upon the observations. The problems associated with binning in the case of unequally spaced observations is not addressed, nor is the problem of biasedness of the estimator in the case of the commonly used method of moments covariogram estimator in (2.19) and (2.20). This additional source of uncertainty in spatial covariance modeling is certain to have implications at the prediction (kriging) stage.

A sound foundation in nonparametric function estimation has been developed over recent years and the performance of associated methods for optimal binwidth selection have certainly been proven in practice. The limitation of any kernel-based method is however seen in the fact that the practical domain consists only of the observed data range. Indeed, even large gaps in the data present problems along with additional complications at the boundaries of the observed data interval. The true function is in most cases expected to extend beyond the data range, as is the case with the covariogram defined for the entire real domain. Taking the Fourier transform of finite domain representations of functions with true infinite domains tends to distort the spectral representation.
3. COMPREHENSIVE COVARIANCE ESTIMATION

The existing collection of spatial covariance models is often limited in practice. While known spatial covariance models are adaptable to a wide variety of problems, their behavior is often seen to be problem specific. For example the functional form of the Gaussian model

\[ c(r) = c(0) \exp\left(-\left(\frac{r}{a_c}\right)^2\right), \quad c(0) \geq 0, a_c > 0 \]

suggests an infinitely differentiable process with a characteristic sigmoidal shape. The development of a more general single covariogram model adaptable to a wide range of cases should involve a fundamental representation of the covariance. Since we can use Bochner's theorem as a basis for development of an admissible covariance, the Fourier transform of the covariance is a logical natural representation with attractive mathematical properties. In the following chapter, we will develop an estimator in the form of a Fourier-Bessel series, of which the familiar Fourier series is a special case.

Watson (1995) describes the historical background and theory of Fourier-Bessel series. Although not as theoretically exhaustive, Bateman (1953) offers a general description with some discussion and references to special cases not mentioned in Watson (1995). Fourier-Bessel series are commonly used in engineering applications. In a spatial context, Christakos (1984) suggests the use of Fourier-Bessel expansions as an excellent approximation of spectral functions of parametric variogram or covariogram models where the Fourier transform is not obtainable directly by integration. A prerequisite for this approximation is that the corresponding spectral function vanishes outside a finite frequency. This is suggested to be the common case.
3.1 Background on One-Dimensional Fourier-Bessel Series

Much of the material contained in this section is explained in Watson (1995). Fourier (1822) described the expansion of an arbitrary function \( f(x) \) as a series of Bessel functions of order zero in order to model heat conduction. Lommel (1868) later extended Fourier's expansion as a series of Bessel functions of arbitrary order. Hankel (1869) rigorously proved existence and uniform convergence of the expansion.

**Definition 3.1:** The *Fourier-Bessel expansion of \( f(x) \) is written as*

\[
f(x) = \sum_{m=1}^{\infty} a_m J_{\nu}(i_m x) \tag{3.1}
\]

where coefficient \( a_m \) is defined as

\[
a_m = \frac{2}{J_{\nu-1}^2(i_m)} \int_0^1 t f(t) J_{\nu}(i_m t) \, dt \tag{3.2}
\]

and the series (3.1) converges to the sum \( f(x) \) for any real \( x \in (0,1) \).

**Definition 3.2:** The function \( J_{\nu}(z) \) is called the *Bessel function of the first kind of order \( \nu \) and is written as*

\[
J_{\nu}(z) = \sum_{k=0}^{\infty} \frac{(-1)^k \left( \frac{1}{2} z \right)^{\nu+2k}}{k! \Gamma(\nu+k+1)} \tag{3.3}
\]

where the constants \( i_1, i_2, \ldots \) are defined as the ordered positive zeros of \( J_{\nu}(z) \).
There are infinitely many zeros of $J_\nu(z)$. Their properties are detailed in Watson (1995). Note specifically that the roots depend upon the order $\nu$ of $J_\nu(z)$ and in fact, $J_\nu(z)$ and $J_{\nu+1}(z)$ have no positive roots in common.

The functional representation in (3.1) is characterized by the basis function $J_\nu(z)$ and is motivated by the orthogonality property of Bessel functions specifically stated as

$$\int_0^t f(t) J_\nu(i_n t) J_\nu(i_m t) \, dt = \begin{cases} 0 & n \neq m \\ \frac{1}{2} J_{\nu+1}^2(i_m) & n = m \end{cases}. $$

This choice of basis function in the series representation of $f(x)$ will also prove to be convenient with respect to the series representation of isotropic and second-order stationary spatial covariance functions $c(r)$. Additional properties of Bessel functions suggest advantages in their use of spatial covariance representation, e.g. Matern (1960). Results regarding boundedness and term-by-term differentiability and integrability of $J_\nu(z)$ with respect to $z$ and $\nu$ are specifically stated in the following theorems.

**Theorem 3.1:** (Watson, 1995) $J_\nu(z)$ is an analytic function of $z$ for all values of $z$ ($z = 0$ possibly being excepted) and is an analytic function of $\nu$ for all values of $\nu$.

**Theorem 3.2:** (Nielsen, 1899) $J_\nu(z)$ is bounded, namely if $J_\nu(z)$ may be expressed as

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(\nu+1)}(1+h)$$

where
\[ |h| < \exp \left( \frac{\frac{1}{2} |z|^2}{|\nu_0 + 1|} \right) - 1, \]

and \[ |\nu_0 + 1| = \min \{ |\nu + 1|, |\nu + 2|, |\nu + 3|, \ldots \}. \]

Derivation of the limit theorems for the series in (3.1) conveniently utilizes the

calculus of residues but is outside the scope of this discussion. The reader is again

referred to Watson (1995) for a full discourse. Three convergence theorems are stated

without proof as follows:

**Theorem 3.3:** (Watson, 1995) Let \( f(t) \) be a function defined arbitrarily on the open

interval \((0,1)\) and let \( \int_0^1 \sqrt{t} f(t) \, dt \) exist and be absolutely convergent. Let

\[ a_m = \frac{2}{J_{\nu+1}(i_m)} \int_0^1 t f(t) J_{\nu}(i_m t) \, dt \text{ where } \nu \geq -\frac{1}{2}. \]

In addition, let \( x \) be any point in \((a,b)\) such that \( 0 < a < b < 1 \) and \( f(t) \) has limited total

fluctuation in \((a,b)\). Then the series \( \sum_{m=1}^{\infty} a_m J_{\nu}(i_m x) \) is convergent with sum

\( \frac{1}{2} \{ f(x+0) + f(x-0) \} \).

If \( f(t) \) is continuous in an open interval \((a,b)\) and satisfies the conditions of Theorem

3.3 then Watson, 1995 shows the Fourier-Bessel expansion associated with \( f(t) \)

converges uniformly to the sum \( f(x) \) throughout the open interval \((a+\varepsilon, b-\varepsilon)\) where
\( \varepsilon \) is a positive number. This is used in the proof to establish the following theorem and corollary;

**Theorem 3.4:** (Watson, 1995) If \( \sqrt{t}f(t) \) has limited total fluctuation in \((a,b)\), where \((a,b)\) is any part or all of the open interval \((0,1)\), then

\[
\int_a^b t f(t) J_v(\lambda t) \, dt = O\left(\frac{1}{\lambda^{\nu+1}}\right) \quad \text{as} \quad \lambda \to \infty.
\]

**Corollary 3.4.1:** (Sheppard, 1889) If \( \sqrt{t}f(t) \) has an integral which is absolutely convergent and \( \sqrt{t}f(t) \) has limited total fluctuation, then for the open interval \( 0 < x \leq 1 \), the following holds.

\[
\frac{2J_v(i_m x)}{J_{\nu+1}(i_m)} \int_0^1 tf(t) J_v(i_m t) \, dt = O\left(\frac{1}{m^{\nu}}\right).
\]

The preceding convergence theorems will be later used in defining the limit properties of the Fourier-Bessel series representation of the isotropic spatial covariance function \( c(r) \). In particular, we will investigate the convergence of the partial sum of the Fourier-Bessel series and the order of magnitude of the terms in the series.

**3.2 Spatial Covariance Modeling: The Fourier-Bessel Series Representation**

The admissible spatial covariance \( c_d(r) \) has a Fourier-Bessel series representation of order \( \nu \) as follows,
**Definition 3.3**: Let $c_d(r) = 0$ for $r > r_c$ such that $c_d(r/r_c)$ is defined arbitrarily for $0 < r/r_c < 1$. The Fourier-Bessel series of order $\nu$ associated with the spatial covariance function $c_d(r)$ is written as

$$
c_d(r) = \sum_{m=1}^{\infty} \frac{2J_\nu(i_mr/r_c)}{J^2_{\nu-1}(i_m)} \int_0^1 tC_d(t) J_\nu(i_mt) dt \quad 0 < r < r_c. \tag{3.4}
$$

The series in (3.4) above is defined for any order $\nu \geq -1/2$. Optimal order of the series for particular $f(x)$ is not discussed in Watson (1995). However, the order parameter $\nu$ may be selected so that the integral in the series representation is conveniently evaluated. For the $d$-dimensional covariogram $c_d(r)$, we show that by selecting the order $\nu$ to be indexed by dimension, the integral in the series representation (3.4) can be manipulated to be the $d$-dimensional spectrum.

**Proposition 3.1**: Let $Z(s)$ be an isotropic random field with $s \in \mathbb{R}^d$ and let $c_d(r) = 0$ for $r > r_c$. The Fourier-Bessel series associated with the corresponding $d$-dimensional covariogram $c_d(r)$ is given by

$$
c_d(r) = \frac{2(2\pi)^{d/2}}{r_c(r/r_c)^{d/2-1}} \sum_{m=1}^{\infty} \left( \frac{J_{(d-2)/2}(i_m r/r_c)}{J^2_{(d-2)/2}(i_m)} \right) (i_m)^{(d-2)/2} C_d(i_m) \tag{3.5}
$$

where $C_d(\cdot)$ is the spectrum of $c_d(r)$. 
Proof: Write \( c_d (r) \) as a Fourier-Bessel series of order \( \nu = (d - 2)/2 \). Recall from proposition 2.2: (Adler, 1981) the following isotropic spectral representation for \( d \) dimensions is given by,

\[
C_d (\omega) = \frac{1}{(2\pi)^{d/2}} \int_0^\infty \frac{J_{(d-2)/2} (\omega r)}{(\omega r)^{(d-2)/2}} c_d (r) r^{d-1} dr.
\]

Multiply the Fourier-Bessel series representation of \( c_d (r) \) by appropriate constants so that the integrand takes the form of the integrand in (3.6) with \( \omega = i_m \) giving,

\[
c_d (r) = \sum_{m=1}^\infty \frac{2 J_{(d-2)/2} (i_m r / r_c) (i_m r / r_c)^{(d-2)/2}}{J_{d/2}^2 (i_m r / r_c) (r / r_c)^{(d-2)/2}} \int_0^1 t \frac{J_{(d-2)/2} (i_m t)}{(i_m r / r_c)^{(d-2)/2}} c_d (t) (r / r_c)^{(d-2)/2} dt.
\]

Next substitute \( t = r / r_c \) in the integral noting that \( c_d (r) = 0 \) when \( r > r_c \) so that the integration limits are defined by \( 0 \leq r \leq r_c \) to give

\[
c_d (r) = \sum_{m=1}^\infty \frac{2 J_{(d-2)/2} (i_m r / r_c)}{J_{d/2}^2 (i_m r / r_c)} \left( \frac{2\pi}{r_c} \right)^{d/2} \left( \frac{i_m r / r_c}{r_c} \right)^{(d-2)/2} \left( \frac{r / r_c}{r_c} \right)^{d-1} \left\{ \frac{1}{(2\pi)^{d/2}} \int_0^{r_c} \frac{J_{(d-2)/2} (i_m r / r_c)}{(i_m r / r_c)^{(d-2)/2}} c (r / r_c) (r / r_c)^{d-1} dr \right\} \times \left( \frac{J_{(d-2)/2} (i_m r / r_c)}{J_{d/2}^2 (i_m r / r_c)} \right) (i_m)^{(d-2)/2} C_d (i_m).
\]

Proposition 3.4: Let \( Z(s) \) be an isotropic random field with \( s \in R^d \) and let \( c_d (r) = 0 \) for \( r > r_c \). Then the spectrum \( C_d (\omega) \) of the Fourier-Bessel series associated with \( c_d (r) \) may be written as
\[ C_d(\omega) = \left(\frac{2}{r_c^2}\right) \sum_{m=1}^{\infty} \frac{C_d(m \pi r_c)}{J_{d/2}(m \pi r_c)} \left(\frac{m \pi r_c}{\omega}\right)^{-(d-2)/2} \int_0^\infty r J_{(d-2)/2}(m \pi r) J_{(d-2)/2}(\omega r) \, dr \] (3.7)

**Proof:** Apply the definition of \( C_d(\omega) \) as the (inverse) Fourier transform of \( c_d(r) \) given in (3.6), and replace \( c_d(r) \) with the Fourier-Bessel series representation of \( c_d(r) \) given in (3.5). For \( d \in \{1, 2, 3\} \), the integral in (3.7) has a closed form solution.

**Remark 3.1:** For a second-order stationary and isotropic random field in \( \mathbb{R}^d \), the Fourier-Bessel series associated with the spatial covariance \( c_1(r) \) is given by

\[ c_1(r) = \frac{2\pi}{r_c} \sum_{m=1}^{\infty} \cos(m \pi r / r_c) C_1(m \pi) \] (3.8)

and the corresponding series associated with the spectrum \( C_1(\omega) \) is given by

\[ C_1(\omega) = 2 \sum_{m=1}^{\infty} \frac{m \pi \cos(r_c \omega) C_1(m \pi)}{(m \pi - r_c \omega)(m \pi + r_c \omega)} \] (3.9)

where the Fourier transform pair \( c_1(r) \Leftrightarrow C_1(\omega) \) is defined as follows,

\[ c_1(r) = \begin{cases} 2 \int_0^\infty \cos(\omega r) C_1(\omega) \, d\omega & r \leq r_c, \\ 0 & r > r_c \end{cases} \]

\[ C_1(\omega) = \frac{1}{\pi} \int_0^\infty \cos(\omega r) c_1(r) \, dr \]

**Proof:** For \( d = 1 \), define the Fourier-Bessel series for \( \nu = (d-1)/2 = -1/2 \) and note that

\[ J_{-1/2}(z) = \sqrt{\frac{2}{z\pi}} \cos(z) \]

and
\[ J_{\nu_2}(z) = \sqrt{\frac{2}{\pi z}} \sin(z). \]

By proposition 3.3 above we get

\[ c_i(r) = \frac{2}{r_c} \sum_{m=1}^{\infty} \frac{\cos \left( \frac{i_m r}{r_c} \right)}{\sin^2 \left( \frac{i_m}{i_m} \right)} \int_0^{r_c} c_i(r/r_c) \cos \left( \frac{i_m r}{r_c} \right) \, dr. \]

Note that \( \pi C_i(\omega) = \int_0^{r_c} \cos(\omega r) c_i(r) \, dr = \int_0^{r_c} \cos(\omega r) c_i(r) \, dr \) so that

\[ c_i(r) = \frac{2\pi}{r_c} \sum_{m=1}^{\infty} \frac{\cos \left( \frac{i_m r}{r_c} \right)}{\sin^2 \left( \frac{i_m}{i_m} \right)} C_i \left( \frac{i_m}{i_m} \right). \]

For the cosine function, the zeros are given at \( m\pi \) for \( m = 1, 2, \ldots \cos(m\pi) = 0 \) for \( m = 1, 2, \ldots \). In addition, the sine function evaluated at the zeros of the cosine function is always 1 or \( \sin \left( \frac{i_m}{i_m} \right) = \sin(m\pi) = 1 \). Therefore, we can write

\[ c_i(r) = \frac{2\pi}{r_c} \sum_{m=1}^{\infty} \cos \left( \frac{m\pi r}{r_c} \right) C_i \left( \frac{m\pi}{i_m} \right). \]

For the spectrum, apply proposition 3.4. Recall that \( C_i(\omega) = \frac{1}{\pi} \int_0^{r_c} \cos(\omega r) c_i(r) \, dr \) so that substituting \( c_i(r) \) by the Fourier-Bessel series gives

\[ C_i(\omega) = \frac{1}{\pi} \int_0^{r_c} \cos(\omega r) \left[ \frac{2\pi}{r_c} \sum_{m=1}^{\infty} \cos \left( \frac{m\pi r}{r_c} \right) C_i \left( \frac{m\pi}{i_m} \right) \right] dr \]

\[ = \frac{2}{r_c} \sum_{m=1}^{\infty} C_i \left( \frac{m\pi}{i_m} \right) \left[ \frac{r_c (r_c \omega \cos(m\pi) \sin(r_c \omega) - m\pi \cos(r_c \omega) \sin(m\pi))}{(r_c \omega - m\pi)(r_c \omega + m\pi)} \right] \]

\[ = 2 \sum_{m=1}^{\infty} \frac{m\pi \cos(r_c \omega) C_i \left( \frac{m\pi}{i_m} \right)}{(m\pi - r_c \omega)(m\pi + r_c \omega)} \]
Remark 3.2: For a second-order stationary and isotropic random field in $R^3$, the Fourier-Bessel series associated with the spatial covariance $c_2(r)$ is given by

$$c_2(r) = \frac{4\pi}{r_c} \sum_{m=1}^{\infty} \frac{J_0(i_m r / r_c)}{J_1^2(i_m)} C_2(i_m)$$

and the corresponding series representation of the spectrum $C_2(\omega)$ is given by

$$C_2(\omega) = 2 \sum_{m=1}^{\infty} \frac{i_m J_0(r_c \omega) C_2(i_m)}{J_1(i_m) (i_m^2 - r_c^2 \omega^2)}$$

where the Fourier transform pair $c_2(r) \leftrightarrow C_2(\omega)$ is defined as follows,

$$c_2(r) = \begin{cases} \frac{2\pi}{r_c} \int_0^r J_0(\omega r) \omega C_2(\omega) \, d\omega & r \leq r_c, \\ 0 & r > r_c \end{cases}$$

$$C_2(\omega) = \frac{1}{2\pi} \int_0^{r_c} J_0(\omega r) r c_2(r) \, dr.$$

Proof: Apply proposition 3.3 for $n = 2$. For the spectrum, apply proposition 3.4 by recalling that $C_2(\omega) = \frac{1}{2\pi} \int_0^{r_c} J_0(\omega r) r c_2(r) \, dr$ so that substituting $c_2(r)$ by the Fourier-Bessel series gives

$$C_2(\omega) = \frac{1}{2\pi} \int_0^{r_c} J_0(\omega r) r \left[ \frac{4\pi}{r_c} \sum_{m=1}^{\infty} \frac{J_0(i_m r / r_c)}{J_1^2(i_m)} C_2(i_m) \right] \, dr$$

$$= \frac{2}{r_c} \sum_{m=1}^{\infty} \frac{C_2(i_m / r_c)}{J_1^2(i_m)} \left[ \frac{r_c^2 (i_m J_0(r_c \omega) J_{-1}^0(i_m) - r_c \omega J_0(i_m) J_{-1}(r_c \omega))}{r_c^2 \omega^2 - i_m^2} \right]$$

$$= \frac{2}{r_c} \sum_{m=1}^{\infty} \frac{C_2(i_m)}{J_1^2(i_m)} \left[ \frac{r_c^2 (i_m J_0(r_c \omega) J_1(i_m) - r_c \omega J_0(i_m) J_1(r_c \omega))}{i_m^2 - r_c^2 \omega^2} \right]$$
\[ = 2 \sum_{m=1}^{\infty} \frac{i_m J_0 (r, \omega) C_2 (i_m)}{J_1 (i_m) (i_m^2 - r_c^2 \omega^2)}. \]

**Remark 3.3:** For a second-order stationary and isotropic random field in \( R^3 \), the Fourier-Bessel series associated with the spatial covariance \( c_s (r) \) is given by

\[ c_s (r) = \frac{4 \pi^2}{rr_c} \sum_{m=1}^{\infty} m \pi \sin (m \pi r / r_c) C_3 (m \pi) \] (3.12)

and the corresponding series representation of the spectrum \( C_3 (\omega) \) is given by

\[ C_3 (\omega) = \frac{2 \pi^2}{\omega} \sum_{m=1}^{\infty} m^2 C_3 (m \pi / r_c) \left[ \frac{2 r_c \omega (\cos (r_c \omega) - 1) - (m^2 \pi^2 - r_c^2 \omega^2) \sin (r_c \omega)}{(m \pi - r_c \omega)^2 (m \pi + r_c \omega)^2} \right] \] (3.13)

where the Fourier transform pair \( c_s (r) \Leftrightarrow C_3 (\omega) \) is defined as follows,

\[ c_s (r) = \begin{cases} 4 \pi \int_0^{r_c} \frac{\sin (\omega r)}{r} \omega C_3 (\omega) d\omega & r \leq r_c, \\ 0 & r > r_c \end{cases} \]

\[ C_3 (\omega) = \frac{1}{2 \omega^2} \int_0^{r_c} \sin (\omega r) r c_s (r) dr. \]

**Proof:** Define the Fourier-Bessel series for \( \nu = 1/2 \) and note that

\[ J_{\nu_2} (z) = \sqrt{\frac{2}{z} \sin (z)} \]

and

\[ J_{\nu_2} (z) = \sqrt{\frac{2}{z^3} \sin (z) - z \cos (z)}. \]

Apply proposition 3.4 for \( n = 3 \) to get
\[ c_3(r) = \frac{1}{r_c} \sum_{m=1}^{\infty} \sin (i_m r/r_c) \frac{2i_m}{rr_c} \left( \frac{1}{i_m/r_c} \right) \int_0^{r_c} r c_3(r) \sin (i_m r/r_c) \, dr. \]

where we have used the fact that the zeros of the sine function occur at \( i_m = m\pi \) so that \( \sin (i_m) = 0 \). Note that

\[ 2\pi^2 C_3(\omega) = (1/\omega) \int_0^{\infty} \sin (\omega r) r c_3(r) \, dr = (1/\omega) \int_0^{r_c} \sin (\omega r) r c_3(r) \, dr \]

thus substituting

\[ 2\pi^2 C_3(m\pi) \] gives

\[ c_3(r) = \frac{4\pi^3}{rr_c} \sum_{m=1}^{\infty} m \sin (m\pi r/r_c) C_3(m\pi). \]

For the spectrum, apply proposition 3.4 and note that \( C_3(\omega) = \frac{1}{\pi} \int_0^{r_c} \cos (\omega r) c_3(r) \, dr \) so that substituting \( c_3(r) \) by the Fourier-Bessel series gives

\[ C_3(\omega) = \frac{1}{\pi} \int_0^{r_c} \cos (\omega r) \left[ \frac{2\pi}{r_c} \sum_{m=1}^{\infty} \cos (m\pi r/r_c) C_3(m\pi) \right] \, dr \]

\[ = \frac{2\pi^2}{\omega} \sum_{m=1}^{\infty} m^2 C_3(m\pi) \left[ \frac{2r_c \omega (\cos (r_c \omega) - 1) - (m^2 \pi^2 - r_c^2 \omega^2) \sin (r_c \omega)}{(m\pi - r_c \omega)^2 (m\pi + r_c \omega)^2} \right]. \]

Note we have taken into account \( \cos (m\pi) = 1 \) and \( \sin (m\pi) = 0 \) for \( m = 1, 2, ..., \)

### 3.3 Estimation

We propose an estimator based upon the Fourier-Bessel series associated with \( c_d(r) \). If the infinite sum in the representation (3.5) is truncated to \( M \) terms, we can write the covariance as a partial finite sum denoted \( c_{d,M}(r) \).
\[ c_{d,M}(r) = (2\pi)^{d/2} \sum_{m=1}^{M} \frac{2J_{(d-2)/2}(i_m r / r_c)(i_m / r_c)^{(d-2)/2}}{J_{d/2}^2(i_m) r^{(d-2)/2} r_c^2} C_d(i_m / r_c), \quad (3.14) \]

where \( c_{d,M}(r) = 0 \) for \( r \geq r_c \) and true spectrum,

\[ C_d(\omega) = (2\pi)^{-(d/2)} \int_0^{r_c} \frac{r^{(d-2)/2} r}{(\omega)^{(d-2)/2}} c(r / r_c) J_{(d-2)/2}(\omega r) \, dr \quad (3.15) \]

The Fourier transform of (3.14) is the truncated spectrum,

\[ C_{d,M}(\omega) = (2 / r_c^2) \sum_{m=1}^{M} \frac{C_d(i_m / r_c)(i_m / r_c)^{(d-2)/2}}{J_{d/2}^2(i_m) \omega} \int_0^{r_c} r J_{(d-2)/2}(i_m r) J_{(d-2)/2}(\omega r) \, dr \quad (3.16) \]

The true and truncated series spectral representations in (3.15) and (3.16) respectively are related as follows,

\[ \lim_{M \to \infty} C_{d,M}(\omega) = C_d(\omega) \]

The true spectrum \( C_d(\omega) \) in (3.14) is unknown. If \( C_d(\omega) \) is given an arbitrary form dependent upon the \( p \)-dimensional parameter vector \( \theta \in \Theta \), then the spectrum is written as \( C_d(\omega) = C_d(\omega; \theta) \) and the \( M \)-truncated Fourier-Bessel series representation of the spatial covariance is now written as a parameter dependent function,

\[ c_{d,M}(r; \theta) = (2\pi)^{d/2} \sum_{m=1}^{M} \frac{2J_{(d-2)/2}(i_m r / r_c)(i_m / r_c)^{(d-2)/2}}{J_{d/2}^2(i_m) r^{(d-2)/2} r_c^2} C_d(i_m / r_c; \theta), \quad (3.17) \]

where \( d \) is the dimension of the sample space and \( M < \infty \) is the number of terms summed in the truncated Fourier-Bessel series where \( m = 1, \ldots, M \). By Bochner’s theorem, in order for (3.17) to be an admissible covariance, the Fourier transform of \( c_{d,M}(r; \theta) \) is necessarily nonnegative and bounded, e.g.
\[ C_{d,M}(\omega; \theta) \geq 0 \text{ for all } \theta \in \Theta_{p+1} \text{ and } \omega \geq 0. \]

For this to occur, by (3.16) the true spectrum \( C_d(\omega; \theta) \) must also be necessarily nonnegative and bounded, e.g.

\[ C_d(\omega; \theta) \geq 0 \text{ for all } \theta \in \Theta_{p+1} \text{ and } \omega \geq 0. \]

To estimate the Fourier-Bessel series model in (3.5), an adequate model of the spectrum is required. A well-known characteristic of the spectrum of the covariance is that it is a nonnormalized density where the spectrum of the correlation is a true density. Christakos (1984) suggests a method for construction of new correlogram models using known density functions in place of the scaled spectrum \( C_d(\omega)/c_d(0) \) in the spectral representation of the correlation function \( \rho_d(r) \) where the corresponding Fourier transform is given by

\[
\rho_d(r) = \frac{c_d(r)}{c_d(0)} = (2\pi)^{d/2} \int_0^{\infty} \frac{J_{(d-2)/2}(\omega r)}{(\omega r)^{(d-2)/2}} \frac{C_d(\omega)}{c_d(0)} \omega^{d-1} d\omega. \tag{3.18}
\]

For our generalized covariogram model, we require a functional representation of \( C_d(\omega) \) with elasticity adequate to model a variety of true spectral forms. Most of the known distributions can be modeled by systems of distributions summarized in Stuart and Ord (1994). Most spatial spectra are characterized by unimodality. Otherwise a variety of shapes are encountered in practice that exhibit varying degrees of higher moment characteristics. For purposes of general modeling of \( C_d(\omega) \), we utilize a scaled form of the Johnson system of distributions described in Johnson, (1949). The Johnson system of
distributions is based on a transformed normal variate and describes a variety of shapes even greater than that of the Pearson system. Of the three Johnson types, $S_L$, $S_H$, and $S_U$, the Johnson $S_U$ system characterizes a variety of unbounded range distributions. By assuming the Johnson $S_U$ system can adequately model dimension-dependent changes in shape, e.g. $C_d(\omega)$ can be approximated by $C_{S_U}(\omega; \theta)$, the following spectrum model of a scaled Johnson $S_U$ system is proposed,

$$C_{S_U}(\omega; \theta) = \frac{\alpha}{\sqrt{1 + (\omega / \lambda)^2}} \exp \left( -(1/2) \delta^2 \left[ \log \left( (\omega / \lambda) + \sqrt{(\omega / \lambda)^2 + 1} \right) \right]^2 \right)$$

$\theta = \{\alpha, \lambda, \delta \} \in \Theta_3$

where $\omega \in (-\infty, \infty)$ and $\alpha > 0$ in order to assure nonnegativity. The scaling coefficient $\alpha$ takes into account that we are modeling the spectrum of the covariogram where the spectrum is a nonnormalized density. For modeling the spectrum of the spatial correlation or correlogram, the scale parameter $\alpha$ is fixed at unity.

With the substitution of the scaled Johnson $S_U$ system in (3.19) for the spectrum, the form of the covariance representation in (3.17) now becomes

$$c_{d,M}(r; a_0, \ldots, a_p) = (2\pi)^{d/2} \sum_{m=1}^{N} \frac{2J_{(d-2)/2}(i_m r / r_c) (i_m / r_c)^{(d-2)/2}}{J_{d/2}(i_m) r^{(d-2)/2} r_c^{2}} C_{S_U}(i_m; \theta)$$

(3.20)

From (3.20), we naturally propose the following statistical covariogram model,

$$c_d^*(r_k) = (2\pi)^{d/2} \sum_{m=1}^{N} \frac{2J_{(d-2)/2}(i_m r_k / r_c) (i_m / r_c)^{(d-2)/2}}{J_{d/2}(i_m) r_k^{(d-2)/2} r_c^{2}} C_{S_U}(i_m; \theta) + \epsilon_k$$

(3.21)
\[ c_{d,M}(r; \theta) + \epsilon_k \]

where \( \theta = \{\alpha, \lambda, \delta\} \in \Theta \) and \( c_d^*(r_k) \) is the \( k \)-th observed element of the empirical or observed covariogram, \( k = 1, \ldots, K \). The parameter vector \( \theta \) is estimated by minimizing the least squares objective function \( q(\theta) \) where

\[ q(\theta) = \frac{1}{K} \sum_{k=1}^{K} (c_d^*(r_k) - c_{d,M}(r_k; \theta))^2 \]

so that \( \hat{\theta} \) is the maximum likelihood estimator of \( \theta \) given by

\[ \hat{\theta} = \min_{\theta \in \Theta} q(\theta) \text{ and } \hat{\theta} = \{\hat{\alpha}, \hat{\lambda}, \hat{\delta}\} \in \Theta. \]

The \( M \)-truncated Fourier-Bessel series estimator of the covariogram is then given by

\[ c_{d,M}(r; \hat{\theta}_0, \ldots, \hat{\theta}_p) = (2\pi)^{d/2} \sum_{m=1}^{M} \frac{2J_{(d-2)/2}(i_m r / r_c)(i_m / r_c)^{(d-2)/2}}{J_{d/2}(i_m) r^{(d-2)/2} r_c^2} C_{S_{\nu}}(i_m; \hat{\theta}). \]

(3.22)

A matrix representation of the model and estimator is easily constructed. Let \( \mathbf{T}_d(r) \) and \( \mathbf{C}_{S_{\nu}}(\theta) \) be vectors and \( k_d \) is the appropriate \( d \)-specified constant each defined respectively as

\[
\mathbf{T}_d(r) = \begin{bmatrix} t_{d,1}(r) \\ \vdots \\ t_{d,M}(r) \end{bmatrix} \quad \mathbf{C}_{S_{\nu}}(\theta) = \begin{bmatrix} C_{S_{\nu},1}(\theta) \\ \vdots \\ C_{S_{\nu},M}(\theta) \end{bmatrix} \quad k_d = \frac{2(2\pi)^{d/2}}{r_c^2}.
\]

\( \mathbf{T}_d(r) \) is the \( M \times 1 \) vector with elements \( t_{d,m}(r) \) indexed by the \( m^{th} \) positive zero of \( J_{\nu}(z) \), \( m = 1, \ldots, M \) with argument \( r \), written as
\[ t_{d,m}(r) = \frac{J_{(d-2)/2}\left(i_m \rho / r_c\right) \left(i_m / r_c\right)^{(d-2)/2}}{J_{d/2}^2(i_m) r^{(d-2)/2}}. \] (3.23)

\( C_{s_v} (\theta) \) is the \( M \times 1 \) vector of the spectrum elements \( C_{s_v,m} (\theta) \) evaluated at the first \( M \) scaled zeros of \( J_\nu(z) \) and parameterized by \( \theta \in \Theta \) where \( \theta \) is the \( p \times 1 \) vector of parameters, namely

\[ C_{d,m}(\theta) = C_{s_v}(i_m, \theta). \] (3.24)

The \( M \)-truncated series representation of the spatial covariance \( c_{d,M}(r; \theta) \) in (3.17) may then be written as the product of two vectors,

\[ c_{d,M}(r; \theta) = k_d T_d^T(r) C_{s_v}(\theta) \] (3.25)

Similarly, the model in (3.21) may be written as

\[ c^*_d(r_k) = k_d T_d^T(r_k) C_{s_v}(\theta) + \epsilon_k \] (3.26)

and the estimator in (3.22) written as

\[ c_{d,M}(r; \hat{\theta}) = k_d T_d^T(r) C_{s_v}(\hat{\theta}) \] (3.27)

The objective function with respect to \( \theta \) may be written as

\[ q(\theta) = \sum_{k=1}^{K} \left( c^*_d(r_k) - k_d T_d^T(r_k) C_{s_v}(\theta) \right)^2 \]

\[ = \left( c^*_d - k_d T_d C_{s_v}(\theta) \right)^T \left( c^*_d - k_d T_d C_{s_v}(\theta) \right) \] (3.28)

where \( T_d \) is the \( K \times M \) matrix of functions \( t_{d,m}(r_i) \) indexed by the \( m^{th} \) positive zero of \( J_\nu(z) \), \( m = 1, \ldots, M \) with argument \( r_k \), \( k = 1, \ldots, K \).
\[
\mathbf{T}_d = \begin{bmatrix}
    t_{d,1}(r_1) & \cdots & t_{d,M}(r_1) \\
    \vdots & \ddots & \vdots \\
    t_{d,1}(r_K) & \cdots & t_{d,M}(r_K)
\end{bmatrix} = [\mathbf{T}_d(r_1) \cdots \mathbf{T}_d(r_K)]^T
\]

and \( \mathbf{c}_d^* \) is the \( K \times 1 \) vector of observed covariances.

\[
\mathbf{c}_d^* = \begin{bmatrix}
    c_d^*(r_1) \\
    \vdots \\
    c_d^*(r_K)
\end{bmatrix}
\]

The mean squared error is written as

\[
MSE(c_{d,M}(r; \hat{\theta})) = E\left( c_{d,M}(r; \hat{\theta}) - c_d(r) \right)^2
\]

where

\[
E\left( c_{d,M}(r; \hat{\theta}) \right) = k_d \mathbf{T}_d^r(r) E\left( \mathbf{C}_{s_v}(\hat{\theta}) \right)
\]

and

\[
\text{var}\left( c_{d,M}(r; \hat{\theta}) \right) = k_d^2 \mathbf{T}_d^r(r) \hat{\Sigma}_{s_v \theta} \mathbf{T}_d(r)
\]

where \( \hat{\Sigma}_{s_v \theta} \) is the \( M \times M \) variance-covariance matrix with elements

\[
\text{cov}\left( C_{s_{v,i}}(\hat{\theta}), C_{s_{v,j}}(\hat{\theta}) \right) \quad \text{where} \; i, j = 1, \ldots, M .
\]

The \( MSE(c_{d,M}(r; \hat{\theta})) \) may then be written as

\[
MSE(c_{d,M}(r; \hat{\theta})) = \text{var}\left( c_{d,M}(r; \hat{\theta}) \right) + \text{bias}\left( c_{d,M}(r; \hat{\theta}) \right)^2
\]

\[
= k_d^2 \mathbf{T}_d^r(r) \hat{\Sigma}_{s_v \theta} \mathbf{T}_d(r) + \left( c_d(r) - k_d \mathbf{T}_d^r(r) E\left( \mathbf{C}_{s_v}(\hat{\theta}) \right) \right)^2
\]

\[
= k_d^2 \mathbf{T}_d^r(r) \hat{\Sigma}_{s_v \theta} \mathbf{T}_d(r) + c_d(r)^2
\]

\[
-2k_d c_d(r) \mathbf{T}_d^r(r) E\left( \mathbf{C}_{s_v}(\hat{\theta}) \right)
\]
\[ +k_d^2 T_d^T (r) E \left( C_{s_0} \left( \hat{\theta} \right) \right) E \left( C_{s_0} \left( \hat{\theta} \right) \right)^T T_d (r) \] (3.32)

If we can assume that \( \text{cov} \left( C_{s_{i,j}} (\theta) , C_{s_{j,i}} (\theta) \right) = 0 \) for \( i \neq j \), then the variance covariance matrix \( \hat{\Sigma}_{s_{0}, \theta} \) is written as

\[ \hat{\Sigma}_{s_{0}, \theta} = \text{diag} \left( \sigma_{s_{0,1}}^2 (\hat{\theta}) , ..., \sigma_{s_{0,M}}^2 (\hat{\theta}) \right) \] (3.33)

where \( \sigma_{s_{0,m}}^2 (\hat{\theta}) = \text{var} \left( C_{s_{0,m}} (\hat{\theta}) \right) \).

By considering the arithmetic means of the partial sums of the series estimator in (3.22), we can apply Fejer's method described in Lanczos (1966) to compute a more globally accurate approximation. If \( c_{d,m} (r; \hat{a}_0, ..., \hat{a}_p) \) is considered as the \( M^{th} \) partial sum of the infinite series \( c_{d,m} (r; \hat{a}_0, ..., \hat{a}_p) \), then Fejer's method considers the convergence of the series \( \bar{c}_{d,M} (r; \hat{a}_0, ..., \hat{a}_p) \) for \( M = 1, 2, ... \) where

\[ \bar{c}_{d,M} (r; \hat{a}_0, ..., \hat{a}_p) = \frac{1}{M} \sum_{j=0}^{M-1} c_{d,j} (r; \hat{a}_0, ..., \hat{a}_p) \] (3.34)

Note that \( c_{d,0} (r; \hat{a}_0, ..., \hat{a}_p) = 0 \) by definition of the Bessel series in (3.3). By considering the limit,

\[ \bar{c}_{d,\infty} (r; \hat{a}_0, ..., \hat{a}_p) = \lim_{M \to \infty} \bar{c}_{d,M} (r; \hat{a}_0, ..., \hat{a}_p) \] (3.35)

Fejer's method shows that when \( c_{d,\infty} (r; \hat{a}_0, ..., \hat{a}_p) \) exists, it coincides with \( \bar{c}_{d,\infty} (r; \hat{a}_0, ..., \hat{a}_p) \), but in fact \( \bar{c}_{d,\infty} (r; \hat{a}_0, ..., \hat{a}_p) \) may exist in cases where \( c_{d,\infty} (r; \hat{a}_0, ..., \hat{a}_p) \) does not. The usefulness of Fejer's method is illustrated by considering the integrated mean squared error (IMSE) as a measure of average error,
\[ IMSE = \int_{\mathbb{R}} \left[ c_d(r) - \phi_{d,M}(r) \right]^2 dr \]  

(3.36)

where \( \phi_{d,M}(r) \) is some approximation. The partial sum \( c_{d,M}(r) \) given by (3.14) minimizes the \( IMSE \). The local error \( c_d(r) - c_{d,M}(r) \) may still be large at isolated points and we may have to pay the price of a small \( IMSE \) by an increased error at for example discontinuities where Fejer’s method may give better approximations. Indeed we approximate the \( IMSE \) with the squared error loss function \( SSE \),

\[ SSE = \sum_{i=1}^{n} \left[ c_d(r) - c_{d,M}(r; \hat{a}_0, \ldots, \hat{a}_p) \right]^2 \]  

(3.37)

where \( SSE \) is used as the objective function for estimation of the parameter vector \( \hat{\theta} = (\hat{a}_0, \ldots, \hat{a}_p)^T \). Fejer’s method is applied by using a simple weighting scheme. The series estimator given by (3.22) is rewritten as

\[ c_{d,M}(r; \hat{a}_0, \ldots, \hat{a}_p) = (2\pi)^{d/2} \sum_{m=1}^{M} b_m \left[ \frac{2J_{(d-2)/2}(i_m r / r_c)(i_m / r_c)^{(d-2)/2}}{\int_{d/2}^2 (i_m)^{d-2} r_c^2} C_{S_0}(i_m; \hat{\theta}) \right] \]  

(3.38)

where

\[ b_m = (1 - m/M). \]

Lanczos (1966) shows that Gibb’s oscillations in the neighborhood of discontinuities are completely avoided by use of Fejer’s method.

### 3.4 Simulation Results

Random field simulation is a powerful tool for constructing realizations that take into account the structural dependence in space of the studied natural process. In
simulating a specified random field, we make the assumption that the simulated field is the same as an observed natural random field with the same covariance structure. Implication of such an assumption allows us to utilize comparative measures of random field estimation, e.g. the performance of a hypothesized technique in predicting the random field at an unsampled location. We specifically seek to compare the spatial prediction errors between known established methods of estimation and our proposed estimation method.

Numerous measures for evaluating the performance of an estimator exist in the literature. Most of these methods involve numerically intensive cross-validation techniques described by Scott (1992) among others. The basic idea is to delete some of the data and use the remaining data to predict the deleted observations. The prediction error is then inferred from the predicted minus actual values. Adaptation of cross-validation methods to the spatial domain is described by Cressie (1993). The idea of simulating a random field combined with cross validation allows for more control in the ability to measure the comparative performance of the proposed estimation method. For example, we can generate any number and location of observations necessary for adequate comparison.

Numerous techniques for simulation are available and discussed in Christakos (1992). One of the most common methods for generating Gaussian and other random fields is the lower-upper triangular matrix technique or LU technique (Elishakoff, 1983 and Alabert, 1987). The LU technique is currently used for generating random fields in the Splus Spatial Statistics module and is used here for simulation of specified Gaussian random fields. The LU technique consists of three steps:
i) generate the following $m \times m$ covariance matrix

$$
C_z = \begin{bmatrix}
    c_z(s_1, s_1) & \cdots & c_z(s_1, s_m) \\
    \vdots & \ddots & \vdots \\
    c_z(s_m, s_1) & \cdots & c_z(s_m, s_m)
\end{bmatrix}
$$

(3.39)

where $c_z(s_i, s_j)$ is the covariance of a zero-mean random field $Z(s)$.

ii) Because $C_z$ is nonnegative definite, symmetric, Cholesky decomposition yields its representation as the product of a lower triangular matrix $L$ and an upper triangular matrix $U$ or

$$
C_z = LU
$$

(3.40)

iii) We then define the $m \times 1$ vector

$$
Z = LV
$$

(3.41)

where $V$ is a vector of $m$ independent standard normal random variables and $Z$ is the simulated random field.

Note that the random field generated in (3.41) is a linear combination of Gaussian random variables and is therefore itself Gaussian. The expectation is given by

$$
E(Z) = LE(V)
$$

$$
= 0 ,
$$

and covariance

$$
E \left( (LV)(LV)^T \right) = LE \left( VV^T \right) U
$$

$$
= LIU
$$

$$
= C_z .
$$
Consideration of simulation design points is important in assuring the prediction locations lie within valid regions with respect to kriging. For example, we would not consider predicting a point outside a reasonable convex hull defined by the sample locations. For this reason, we propose a pseudo-random location scheme as follows:

i) For a square simulation region, define $q \times q$ equal square subquadrats within the region.

ii) Within each subquadrat, define $n_q$ uniform random locations for a total of $q^2 n_q$ locations within the simulation region.

iii) Simulate the $q^2 n_q$ realizations for the pseudo-random locations using the LU technique described above.

iv) For each subquadrat, exclude $m_q$ observations from the covariogram estimation step. Therefore, a total of $q^2 \left(n_q - m_q\right)$ simulated realizations are used for estimation of the covariance.

v) Cross validate the $q^2 m_q$ realizations using the covariance estimated in step (iv).

The described simulation scheme assures approximate evenly distributed distances including small lag distances within each subquadrat. In addition, the kriging step is performed on locations that have neighboring points. Christakos (1992) refers to the above design scheme as stratified spatial sampling although we have modified the procedure by holding out subquadrat points for estimation at the cross-validation stage.

We perform the above-described simulation and cross validation for a total of 100 square simulation regions of length 1 per side and $q = 3$ for a total of 9 subquadrats each of area $(1/3) \times (1/3)$. For each subquadrat, $n_q = 10$ random uniform locations and
corresponding simulations are generated and $m_q = 5$ points are held out from the covariance estimation step. This results in $100 \times q^2 \times m_q = 100 \times 9 \times 5 = 4500$ total prediction locations.

The simulation and cross-validation scheme described above is completed for each of three common parametric covariance models, exponential, Gaussian and spherical. Specific forms and parameterizations of each model are described in Table 1. The parameterizations chosen are selected in order to impose a variety of covariogram shapes.

Table 3.1 Parametric covariogram models used for comparison in simulations.

<table>
<thead>
<tr>
<th>Model</th>
<th>Form</th>
<th>$c(0)$</th>
<th>$a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$c(r) = c(0) \exp(-r/a)$, $c(0) &gt; 0$, $a &gt; 0$</td>
<td>0.9</td>
<td>0.3</td>
</tr>
<tr>
<td>Gaussian</td>
<td>$c(r) = c(0) \exp\left(-\left(\frac{r}{a}\right)^{2}\right)$, $c(0) &gt; 0$, $a &gt; 0$</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Spherical</td>
<td>$c(r) = \begin{cases} c(0) \left[1 - \frac{3r}{2a} + \frac{r^3}{2a^3}\right] &amp; \text{if } r \in [0,a] \ 0 &amp; \text{if } r \geq a \end{cases}$, $c(0) &gt; 0$, $a &gt; 0$</td>
<td>0.7</td>
<td>0.2</td>
</tr>
</tbody>
</table>

As an alternative to binning, we fit the model to the covariogram cloud. This eliminates the influence of bin size on either the experimental or true fitted models.

Diggle and Verbyla (1998) model a nonparametric variogram fit to the empirical variogram cloud because the variogram cloud provides an unbiased estimate of the true
variogram. We show conditions under which the covariogram cloud provides an unbiased estimate of the covariogram as well.

**Definition 3.4:** Let \( Z(s) \) be a second-order spatial random field with \( s \in R^d \) observed at locations \( s_i, i = 1, \ldots, n \). The **covariogram cloud** is defined as

\[
\hat{c}(h) = Z(s_i)Z(s_j) \quad \text{for all } i, j = 1, \ldots, n \text{ and } h = s_i - s_j.
\]

Furthermore, if \( Z(s) \) is isotropic, the covariogram cloud is defined as

\[
\hat{c}(r) = Z(s_i)Z(s_j) \quad \text{for all } i, j = 1, \ldots, n \text{ and } r = |h|.
\]

We can define \( r_k \) for \( k = 1, \ldots, K \) where \( r_k \) is the \( k^{th} \) ordered distance of the set of all distances \( \{|s_i - s_j|; i, j = 1, \ldots, n\} \). Then \( c^*(r_k) \) is the \( k^{th} \) observed covariogram cloud element ordered by distance \( r_k \).

**Proposition 3.5:** Let \( Z(s) \) be a second-order zero-mean spatial random field with \( s \in R^d \) observed at locations \( s_i, i = 1, \ldots, n \) with \( h = s_i - s_j \). Then the mean of \( \hat{c}(h) \) is

\[
E(\hat{c}(h)) = c(h).
\]

In addition, if \( Z(s) \) is Gaussian then the variance of \( \hat{c}(h) \) is

\[
\text{var}(\hat{c}(h)) = c(0)^2 + c(h)^2
\]

**Proof:** The mean is straightforward.

\[
E(\hat{c}(h)) = E(Z(s_i)Z(s_j)) = c(h)
\]
for computation of the variance, assume \( Z(s_i) \) and \( Z(s_j) \) are jointly distributed

bivariate normal random variables with

\[
\begin{bmatrix}
Z(s_i) \\
Z(s_j)
\end{bmatrix}
\sim N_2 \begin{bmatrix} 0 & c(h) \\
c(h) & c(0)
\end{bmatrix},
\]

where \( c(h) = E(Z(s_i)Z(s_j)) \) and \( c(0) = E(Z(s_i)^2) \). The density function may be

written as

\[
f(Z(s_i), Z(s_j)) =
\frac{1}{2\pi\sqrt{c(0)^2 - c(h)^2}} \exp \left\{ -\frac{1}{2} \left( \frac{c(0)(Z^2(s_i) - Z^2(s_j)) - 2c(h)Z(s_i)Z(s_j)}{c(0)^2 - c(h)^2} \right) \right\}.
\]

The variance of \( \hat{c}(h) \) is found by direct integration,

\[
\text{var}(\hat{c}(h)) = E(\hat{c}(h)^2) - E(\hat{c}(h))^2
\]

\[
= E(Z(s_i)^2Z(s_j)^2) - c(h)^2
\]

\[
= \int \int Z(s_i)^2Z(s_j)^2 f(Z(s_i), Z(s_j)) dZ(s_i) dZ(s_j) - c(h)^2
\]

\[
= c(0)^2 + 2c(h)^2 - c(h)^2
\]

\[
= c(0)^2 + c(h)^2
\]

**Proposition 3.6:** Let \( Z(s) \) be a second-order zero-mean spatial random field with \( s \in \mathbb{R}^d \) observed at locations \( s_i, i = 1, \ldots, n \) with \( h = s_i - s_j \). Let the sample covariogram
cloud be defined as $\hat{c}(h) = Z(s_i)Z(s_j)$ for all $i, j = 1, \ldots, n$. Then the density of

$$u_h = \hat{c}(h)$$

is given by

$$f_{u_h}(u_h) = \frac{K_0(u_h c(0) / k(h))}{\pi(k(h))^{1/2}} \exp\left(u_h c(h) / k(h)\right), \quad -\infty < u_h < \infty$$

where $K_0(s)$ is the modified Bessel function of the second kind of order zero, and

$$k(h) = c(0)^2 - c(h)^2.$$ 

and the characteristic function is given by

$$\phi(t) = \frac{1}{1 - 2itc(h) + t^2k(h)}.$$ 

The simple kriging or prediction model is described in Cressie (1993). We assume a zero-mean process of the modeled random variable $Z(s_i)$. We seek to optimize the predictions based upon the squared-error loss function,

$$L(Z(s_0), p(Z,s_0)) = (Z(s_0) - p(Z,s_0))^2$$

(3.42)

where $Z(s_0)$ is the prediction at location $s_0$. $Z = \{Z(s_1), \ldots, Z(s_n)\}$ for $n$ locations and $p(Z,s_0)$ is the predictor function. The best linear predictor is given by

$$p(Z,s_0) = \sum_{i=1}^{n} l_i Z(s_i) + k$$

(3.43)

such that $E(Z(s_0) - p(Z,s_0))^2$ is minimized. Such an optimal predictor $p^*$ is given by

$$p^*(Z,s_0) = c^T \Sigma^{-1} Z$$

(3.44)
where $\mathbf{c} = (c(s_0, s_1), \ldots, c(s_0, s_n))^T$, $\Sigma$ is an $n \times n$, whose $(i, j)$th element is $c(s_i, s_j)$, the spatial covariance or covariogram between observations $Z(s_i)$ and $Z(s_j)$. The minimized mean-squared prediction error is given by

$$\sigma_p^2 = C(s_0, s_0) - \mathbf{c}^T \Sigma^{-1} \mathbf{c}.$$ 

For the simulation analysis, the kriging neighborhood is defined as the subquadrat, each containing $n_r = 5$ points used to estimate the covariogram and 5 holdouts for a total of 10 simulated points in each subquadrat. Each simulated field of 9 subquadrats then has a total of $n_q = \sum_{i=1}^{9} n_r = 45$ points used to estimate the covariogram based upon the empirical covariogram cloud.

In order to obtain numerically optimal estimates from the experimental covariogram model in (3.22), we initiate an iterative estimation procedure with reasonable starting values for the parameters and a low order for the series of say $M = 10$. Once estimates are obtained for $M = 10$, those are subsequently used as starting values for a higher order series of $M = 100$. This method of estimation reduces the number of iterations of higher order series estimates with greater precision. Ultimately, prescribed absolute differences between parameter estimates at each iteration are used as stopping criteria.

Performance of the estimated covariograms and of the prediction method in general is assessed using cross-validation as described in Cressie (1993). The closeness of the predicted values to the true values can be characterized by two measures, e.g.

$$CV_A(Z) = (1/n_z) \sum_{j=1}^{n_z} \left[ \left( Z(s_j) - \hat{Z}_{-j}(s_j) \right) / \sigma_{-j}(s_j) \right]$$

(3.45)
or alternatively,

\[
CV_b(Z) = \left( \frac{1}{n_i} \sum_{j=1}^{n_i} \left[ \left( Z(s_j) - \hat{Z}_{-j}(s_j) \right) / \sigma^{-j}(s_j) \right]^2 \right)^{1/2},
\]

where \( \hat{Z}_{-j}(s_j) \) is the predicted simulation point of the hold-outs, \( Z = \{ Z(s_1), ..., Z(s_{n_i}) \} \) and \( \sigma^{-j}(s_j) \) is the prediction error for \( \hat{Z}_{-j}(s_j) \). The cross-validation measure in (3.45) or \( CV_a(Z) \) should be approximately 0 and the cross-validation measure in (3.46) or \( CV_b(Z) \) should be approximately 1 for an adequate predictive model. Goovaerts (1997) discusses a cautious approach when interpreting crossvalidation scores as existing methods fail to provide realistic measures of absolute performance. However, we expect to detect differences in prediction efficiency at least at a comparative level.

Crossvalidation scores are generated for the 1) the estimated true model using the known parameters as starting values, 2) the experimental Fourier-Bessel estimator, and the 3) the unestimated true model using the known parameters. Estimations for the true and experimental models are achieved using the Nelder-Mead simplex algorithm. The Nelder-Mead method precludes the computation of derivatives and performs well in the presence of local minima or noisy functional representations. Results are illustrated graphically. Figs. 3.1 through 3.3 show the distribution of \( CV_a(Z) \) for 100 simulated fields for each of the three models. Figs. 3.4 through 3.6 magnify the distributions in Figs. 3.1 through 3.3 to the regions surrounding the expected crossvalidation score of 0. Figs. 3.7 through 3.9 show the distribution of \( CV_a(Z) \) for 100 simulated fields for each of the three models. Figs. 3.10 through 3.12 magnify the distributions in Figs. 3.7
through 3.9 to the regions surrounding the expected crossvalidation score of 1. In all cases, the true estimated model tended to generate a number of extreme crossvalidation scores for both $CV_A(Z)$ and $CV_B(Z)$ as seen in Figs. 3.1 through 3.3 for $CV_A(Z)$ and Figs. 3.7 through 3.9 for $CV_B(Z)$. This is not true of the experimental Fourier-Bessel estimator which is comparable to the true unestimated model. When the extreme crossvalidation scores are removed as in Figs. 3.4 through 3.6 for $CV_A(Z)$ and Figs. 3.10 through 3.12 for $CV_B(Z)$, the experimental estimator performs as well as the estimated true model. Two exceptions are seen in Figs. 3.10 and 3.11 where the distribution of experimental crossvalidation scores for $CV_B(Z)$ tends to be larger than either the estimated or unestimated true model for the exponential and Gaussian simulated fields. However, the true unestimated model for the Gaussian field tends to display greater variation in the scores and is right-skewed. Overall, the experimental estimator performed as well or better than the estimated true model. One might expect that in practice that a poor crossvalidation score indicative of a poor model fit would encourage the investigator to attempt a new set of starting values or alternative optimization method.

An additional graphical assessment of the comparative performance of each of the three covariogram models is shown in Figs. 3.13 through 3.15 where the predicted vs. the true simulated observations are plotted for each of the exponential, Gaussian and spherical simulated fields. In each case, the experimental Fourier-Bessel series estimator is seen to perform almost identically to the true unestimated and true estimated models. For the exponential and Gaussian random fields, the experimental estimator may exhibit slightly more variation in the predictions, but only a negligible amount.
Figure 3.1 Total distribution of crossvalidation scores $CV_A(Z)$ for 100 exponential simulated random fields.
Figure 3.2 Total distribution of crossvalidation scores $CV_A(Z)$ for 100 Gaussian simulated random fields.
Figure 3.3 Total distribution of crossvalidation scores $CV_a(Z)$ for 100 spherical simulated random fields.
Figure 3.4 Distribution of crossvalidation scores $CV_A(Z)$ for 100 exponential simulated random fields around zero.
Figure 3.5 Distribution of crossvalidation scores $CV_n(Z)$ for 100 Gaussian simulated random fields around zero.
Figure 3.6 Distribution of crossvalidation scores $CV_A(Z)$ for 100 spherical simulated random fields around zero.
Figure 3.7 Total distribution of crossvalidation scores $CV_\beta(Z)$ for 100 exponential simulated random fields.
Figure 3.8 Total distribution of crossvalidation scores $CV_g(Z)$ for 100 Gaussian simulated random fields.
Figure 3.9 Total distribution of crossvalidation scores $CV_{\theta}(Z)$ for 100 spherical simulated random fields.
Figure 3.10 Distribution of crossvalidation scores $CV_b(Z)$ for 100 exponential simulated random fields around zero.
Figure 3.11 Distribution of crossvalidation scores $CV_b(Z)$ for 100 Gaussian simulated random fields around zero.
Figure 3.12 Distribution of crossvalidation scores $CV_{\theta}(Z)$ for 100 spherical simulated random fields around zero.
Figure 3.13 Predicted vs. true simulated observations for 100 simulated exponential random fields of 45 hold-out points in each field.
Figure 3.14 Predicted vs. true simulated observations for 100 simulated Gaussian random fields of 45 hold-out points in each field.
Figure 3.15 Predicted vs. true simulated observations for 100 simulated spherical random fields of 45 hold-out points in each field.
4. SPATIOTEMPORAL DEPENDENCE MODELING

By definition, a spatiotemporal process is one that develops simultaneously in time and space. The precise nature by which time and space dynamically interact to affect the observed process is ultimately what we seek to understand. With this in mind, the most important component of any spatiotemporal modeling exercise is that of characterization of the dependence or covariance structure of the observed process. We hope to use the estimated covariance to relate the behavior of a process at a specific spatial location and time with the behavior of the same process within some predefined local space-time neighborhood. The incorporation of spatial and temporal covariance into a cohesive structure that attempts to explain the process in the spatiotemporal domain has only relatively recently been examined. First and second order restrictions that allow valid stochastic modeling in separate space and time also apply to combined space and time, but additional or modified conditions must often be accounted for in order to adequately predict and interpolate the spatiotemporal process under study.

The optimal spatiotemporal model may be best defined as one that posits the fewest constraints on the observed process. We can only better understand the process through modeling by allowing the proposed model to be as flexible as possible through space and time. For example, spatial variability should be allowed to interact with time. On all temporal scales, the spatial field must be allowed to change on a local or regional spatial scale. This characteristic is ultimately defined by the spatiotemporal covariance structure. Indeed in rare cases, space and time may be assumed separable, but such a covariance model is useful only in those special cases. Ideally, we should define a conceptual covariance model that itself allows the degree of dependence to be defined by
the process itself and not the observer. Such a model is readily adaptable to a broader range of applications.

Modeling assumptions most often depend upon the design of the spatiotemporal experiment itself. For example, one might model a process observed at fixed station locations through time differently from a process observed at random or non-fixed locations through time. In fact, the nature of the problems that spatiotemporal modeling addresses, usually warrants the use of long-term historical datasets that have been collected over time. Therefore, no typical design can be anticipated in order to establish a standard analysis technique.

Important future applications of stochastic spatiotemporal modeling will undoubtedly be efforts directed toward the analysis of massive data sets. One example is spatiotemporal data obtained from geophysical seismic collection or so called 4D seismic in which a resource reservoir is assessed in three spatial dimensions through time in order to determine optimal placement of wells and produce reasonable reservoir forecasts. There currently exists a deterministic technique to model such data, but the computational requirements are significant, typically requiring weeks of computation time on state-of-the-art platforms for even a single realization or 3D image in time. Another example involves the processing of satellite obtained images. Again, this requires spatiotemporal techniques optimized for assimilation of large data sets. Some work is already being done in this area, but only as an interpolatory technique, not taking into account the covariance structure of the process itself, i.e. (Chin and Mariano, 1997) who devise an interpolation method used to model parts of an oceanic front obscured by cloud cover. This technique and others similar to it exemplify the importance of
stochastic spatial analysis as a technique used to enhance high resolution remote data assimilation.

Spatiotemporal modeling is most often attempted with samples that are space abundant and time deficient, or alternatively space deficient and time abundant. Indeed, spatiotemporal datasets both rich in time and space are rare. Given these sampling deficiencies, it is important that not only statistically optimal interpolations and predictions be made, but some measure of confidence must also be included such that an assessment of error may be made for those regions and times that are sparsely sampled. In strictly the temporal domain, this is accomplished most simplistically with confidence bands on the predictions. In the spatial domain, error contours may be easily generated, but there still remains the problem that the prediction variance is underestimated because the estimated variogram is taken to be deterministic when the kriging equations are calculated (Cressie, 1993).

The science of modeling spatial processes as they change through time has long been the domain of complex deterministic models. The most aggressive efforts in deterministic spatiotemporal models have been in the fields of geology, meteorology and oceanography where predictions of complex physical dynamic processes are important in long and short-term prediction. The importance of predicting such large scale processes are apparent for economic reasons, say agricultural production and fisheries harvests, but also for large scale disaster prediction such as regional and global rainfall, temperature extremes and earthquake prediction.

The stochastic approach to spatiotemporal modeling is a relatively recent method and has compared favorably to deterministic models of the same processes (Kaplan et al.,
Perhaps the greatest advantage the stochastic approach has over the deterministic is in the ability of the former to produce prediction errors. Indeed, the stochastic and deterministic approaches can be assumed to be similar up to the level of complexity described in the trend. What remains in the stochastic model is the zero-mean process we actually seek to describe stochastically. The deterministic model alternatively seeks to describe even the microscale structure of the process as non-random. Therefore, while efforts of deterministic modeling are placed on mathematical description of the process at all scales, the efforts of stochastic model building are placed upon describing behavior of the unobserved process by making inferences from the observed realization. The tool constructed for describing such relationships is the spatiotemporal covariance structure.

Section 4.1, includes the background and definitions necessary to understand the concept of the spatiotemporal random field. Naturally, almost unlimited approaches are conceivable and in fact necessary when modeling efforts are attempted. Of primary interest in Section 4.1 is the definition of first and second order moments with respect to the distribution of the underlying field. Section 4.2 focuses on discussion of spatiotemporal dependence models. As expected, spatiotemporal modeling is not as cohesive a science as that of purely spatial or time domain modeling, therefore a variety of techniques are described.

4.1 Framework

Consider a continuous space and time referenced process \( Z(s,t) \) where

\[
s = \{s_1, \ldots, s_d \} \in \mathbb{R}^d \quad \text{and} \quad t \in T \subseteq \mathbb{R}^t_0.
\]

An arbitrarily sampled realization of such a process may be represented by \( Z(s_i, t_i) \), where \( i = 1, \ldots, n \) and thus the sample vector represented
by \( \{ Z(s_i, t_i), ..., Z(s_n, t_n) \} \). If all sites are sampled simultaneously at discrete times \( j = 1, ..., T \), the sample vector may be written as \( \{ Z(s_1, t_1), Z(s_2, t_2), ..., Z(s_T, t_T) \} \). If ergodic, then the sample is from the true underlying continuous space-time process \( Z(s, t) \).

**Definition 4.1:** Let \( \mathcal{S}_{d+t} \) be the set of all \( A \times B \) with \( A \in \mathbb{R}^d \) and \( B \in T \). For \( A \) and \( B \), let the spatiotemporal measure be defined as \( \mu(A \times B) = \phi(A)\psi(B) \). Let \( \Phi \otimes \Psi \) be the product \( \sigma \)-field generated by the set \( \mathcal{S}_{d+t} \). \( \mu \) is then the product measure on \( \Phi \otimes \Psi \).

We can then define the \( \sigma \)-finite measure spaces \((\mathbb{R}^d, \Phi, \phi)\) and \((T, \Psi, \psi)\). A

*spatiotemporal random field* \( Z(s, t) \), \( s \in \mathbb{R}^d \), \( t \in T \) is a measurable function from \( \mathbb{R}^d \times T \) onto \([0, \infty] \). Specifically, \( Z(s, t) \) is measurable for the \( \sigma \)-field \( \Phi \otimes \Psi \) where \( Z(s, t) \in L_1(\mathbb{R}^d \times T, \Phi \otimes \Psi, \mu = \phi \times \psi) \).

The spatiotemporal random field \( Z(s, t) \) can be interpreted as a random field defined in \( d + 1 \) dimensional space or alternatively as a random field in \( d \) dimensional space but indexed by time \( t \). The case considered for modeling mostly depends upon the spatial sampling design in time, but also may depend upon the aggregation properties of the sample.

Similar to the purely spatial random field \( Z(s) \), the spatiotemporal random field \( Z(s, t) \) is defined through the finite dimensional distribution function,
\[ F_{s_i, t_i} (z(s_1, t_1), \ldots, z(s_n, t_n)) = P(z(s_1, t_1) \leq Z(s_1, t_1), \ldots, z(s_n, t_n) \leq Z(s_n, t_n)) \]

(4.1)

for any \( n \). Furthermore the distribution in (2.1) is assumed satisfy the \textit{symmetry} and \textit{consistency} conditions for any \( n \) as defined in Chapter 2.

\textbf{Definition 4.2:} Given the spatiotemporal random field \( Z(s_i, t_i), i = 1, \ldots, n \) with

\[ s_i = (s_{i1}, \ldots, s_{id}) \in \mathbb{R}^d \text{ and } t_i \in T \subset \mathbb{R}^1 \text{ with probability density function} \]

\[ f_{s_i, t_i} (z(s_1, t_1), \ldots, z(s_n, t_n)), \text{ the process } Z(s, t) \text{ admits first and second order moments respectively given by the mean function}, \]

\[ \mu(s, t) = E[Z(s, t)], \]

(4.2)

and the (centered) \textit{spatiotemporal covariance function} or \textit{spatiotemporal covariogram},

\[ c(s_i, t_i; s_j, t_j) = E[(Z(s_i, t_i) - \mu_z(s_i, t_i))(Z(s_j, t_j) - \mu_z(s_j, t_j))]. \]

(4.3)

The \textit{spatiotemporal structure function} or \textit{spatiotemporal variogram} is

\[ \gamma(s_i, t_i; s_j, t_j) = \frac{1}{2} E[(Z(s_i, t_i) - Z(s_j, t_j))^2]. \]

(4.4)

In the case of space homogeneity and time stationarity, the relationship between the spatiotemporal structure function \( \gamma(s_i, t_i; s_j, t_j) \) and spatiotemporal covariance \( c(s_i, t_i; s_j, t_j) \) can be expressed as

\[ c(h, \tau) = c(0, 0) - \gamma(h, \tau). \]

(4.5)
The first and second order moment structure completely characterize the process in the case of a Gaussian field.

Special cases of the underlying process that will be referenced later are defined as follows:

**Definition 4.3:** Given the spatiotemporal random field \( Z(s_i, t_i), i = 1, \ldots, n \) with
\[
s_i = (s_{i,1}, \ldots, s_{i,d}) \in \mathbb{R}^d \quad \text{and} \quad t_i \in T \subset \mathbb{R}^1
\]
with probability density function
\[
f_{s_1, \ldots, s_n, t_1, \ldots, t_n} (z(s_1, t_1), \ldots, z(s_n, t_n)),
\]
the process \( Z(s, t) \) may possess the following class properties:

1. The process \( Z(s, t) \) is called *spatially homogeneous* and *temporally stationary* if the covariance function is space-time invariant with respect to spatial and temporal location or written as
\[
c(s, t) = c(s_i, t_i; s_j, t_j), \quad i \neq j
\]
\[
= c(h, \tau) \tag{4.6}
\]
where \( h = s_i - s_j \) and \( \tau = t_i - t_j \).

2. The process \( Z(s, t) \) is called *isotropic* if the covariance is radially invariant or a function of spatial distance alone
\[
c(s_i, t_i; s_j, t_j) = c(r, \tau) \tag{4.7}
\]
where \( r = |h| \).

3. The process \( Z(s, t) \) is said to have *constant mean* if
\( \mu(s,t) = \mu. \) \hfill (4.8)

4. The process \( Z(s,t) \) is said to have *stationary increments* if the spatiotemporal semivariogram is a function of the spatial and temporal increments only,

\[ \gamma(s_i, t_i; s_j, t_j) = \gamma(h, \tau). \] \hfill (4.9)

**Definition 4.4:** A process \( Z(s,t) \) that satisfies both (4.6) and (4.8) is said to be *second order stationary* or *weakly stationary*. A process \( Z(s,t) \) that satisfies both (4.8) and (4.9) is said to be *intrinsically stationary* or to satisfy the *intrinsic hypothesis*.

The notion of isotropy presents the first complication with space-time modeling in that conditions that apply to time do not in general apply in a spatial context. For example, there is no real meaning of isotropy in the temporal sense. Likewise, time recursivity has no meaning in the spatial sense. Due to complications such as these, space-time covariance structures require rigorous definition such that assumptions made upon either the time or space domain are not extended through both as meaningless constructs.

Typically, the underlying spatiotemporal process may be written as the sum of a deterministic trend component \( \mu(s,t) \) and a single pure stochastic component \( \eta(s,t) \).

\[ Z(s,t) = \mu(s,t) + \eta(s,t). \] \hfill (4.10)
Most modeling approaches separate the spatiotemporal process into a sum of gross or expected trend \( \mu(s,t) \), unknown trend that includes the distinct parts of the process we wish to learn more about \( \nu(s,t) \), and a white noise component \( \epsilon(s,t) \) as

\[
Z(s,t) = \mu(s,t) + \nu(s,t) + \epsilon(s,t).
\]  \hspace{1cm} (4.11)

The additive separation in (4.11) is arbitrary and indeed non-unique. From an operational point of view, one seeks to discover new information regarding the underlying continuous process \( Z(s,t) \). From this standpoint, one may approach the additive separation in (4.11) as modeling the sum of a known deterministic part and an unknown part with a stochastic component. For example, rainfall amounts in space and time may be empirically decomposed as

\[
Z(s,t) = \delta(t) + Y(s,t)
\]  \hspace{1cm} (4.12)

where \( \delta(t) \) is the known large scale seasonal components and the process \( Y(s,t) \) contains information on smaller scale local variation in space and time. The investigator then concentrates on modeling \( Y(s,t) \) in order to elucidate unknown spatial and temporal variation in rainfall amounts.

The easiest approach involves modeling the known trend \( \mu(s,t) = \sum_{i=1}^{p} \delta_i(s,t) \) as a \( p^{th} \) degree polynomial in \( s \) and/or \( t \) where typically \( p \leq 2 \). The form of the trend function employed is largely objective and is selected with regard to the data and particular features to be modeled. There currently exist a number of methods for modeling and accounting for the known or gross spatiotemporal trend. Of these, two methods are most commonly used, e.g. median polish (Cressie, 1986) and polynomial.
regression (Ripley, 1981). Undoubtedly, nonparametric regression techniques are also used to detrend spatiotemporal data but we are not aware of examples in the literature.

Various theoretical forms of the population spatiotemporal semivariogram and covariogram are seen in the literature. Some of these are developed based upon characteristics of the process being modeled while others are used simply out of mathematical convenience. Still other spatiotemporal semivariogram and covariogram models are selected due to their ability to be fit to a wider variety of empirical representations. Regardless of the theoretical model, in order to compute meaningful prediction estimates, the spatiotemporal covariance function must meet specific admissibility criteria. As in the spatial case, we state the following necessary and sufficient conditions in order for a continuous function to be a valid spatiotemporal covariogram or semivariogram.

**Proposition 4.1:** Admissible conditions for spatiotemporal covariance structures.

1. The spatiotemporal covariogram \( c(s_i, t_i; s_j, t_j) \) must be nonnegative definite,

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j c(s_i, t_i; s_j, t_j) \geq 0 \quad \forall n \tag{4.13}
\]

for any complex number \( \lambda_i \).

2. The spatiotemporal semivariogram \( \gamma(s_i, t_i; s_j, t_j) \) must be conditionally negative definite,

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma(s_i, t_i; s_j, t_j) \leq 0 \quad \forall n \tag{4.14}
\]

when
\[ \sum_{i=1}^{n} \lambda_i = 0. \]

These conditions assure that the variance of the prediction estimates will be nonnegative. Potential theoretical covariance models may still provide meaningful prediction estimates and even valid prediction errors for the data at hand, but may prove meaningless when fit to an empirical covariance model with different data. It is often difficult and sometimes impossible to identify a valid covariance structure by directly applying (4.13) or (4.14). Fortunately, alternative verification methods based upon Bochner (1955) are simpler to apply.

In practice, theoretical spatial parametric forms known to be admissible are often combined to form spatiotemporal parametric composite covariograms and semivariograms. These models are generally classified as follows:

1. **space-time factorable or separable**,
   \[
   c(h, \tau) = c_1(h)c_2(\tau) \tag{4.15}
   \]
   \[
   \gamma(h, \tau) = \gamma_1(h)\gamma_2(\tau). \tag{4.16}
   \]

2. **space-time zonal anisotropic**,
   \[
   c(h, \tau) = c_1(h) + c_2(\tau) \tag{4.17}
   \]
   \[
   \gamma(h, \tau) = \gamma_1(h) + \gamma_2(\tau). \tag{4.18}
   \]

3. **combined space-time factorable and zonal anisotropic**,
   \[
   c(h, \tau) = c_1(h) + c_2(\tau) + c_1(h)c_2(\tau) \tag{4.19}
   \]
   \[
   \gamma(h, \tau) = \gamma_1(h) + \gamma_2(\tau) + \gamma_1(h)\gamma_2(\tau). \tag{4.20}
   \]
Myers and Journel (1990) point out that models incorporating zonal anisotropies are at risk of being non-admissible. They show by example that a non-invertible kriging system may be easily constructed due to the positive semidefiniteness of the covariogram when strict positive definiteness is required. Nevertheless, these types of models will continue to be employed as they are readily available, easily fitted in most cases and provide

### 4.2 Survey of Methods

For spatiotemporal applications, most works precondition the observed field by transforming the assumed spatially nonhomogeneous and temporally nonstationary random process to one of spatial homogeneity and temporal stationarity. Most approaches use simple polynomial detrending of the observed process in order to yield a convenient detrended process upon which estimation and prediction are done. The modeling step itself is as varied as the application. Numerous creative techniques are described, but may not necessarily be amenable to a wide range of applications. As an exercise, spatiotemporal modeling seeks to provide information to specific problems. The methods of aggregation, detrending, covariance modeling and prediction are all subject to constraints posed by those distinct problems. Although not all-inclusive, we will describe some of the unique approaches to spatiotemporal covariance modeling in order to provide the reader with an idea of the variety of techniques available. We will begin with earlier approaches to spatiotemporal covariance modeling and proceed to more recent efforts.

Rodriguez-Iturbe and Mejia (1974) describe a rainfall process in terms of its correlation structure in time and space for purposes of optimal design of precipitation
monitoring networks. The factorable spatiotemporal covariogram may be expressed as follows

\[ c(h, \tau) = \sigma_z^2 \rho(h) \phi(\tau), \]

where \( \sigma_z^2 \) is the process variance, \( \rho(h) \) is the purely spatial component and \( \phi(\tau) \) is the purely temporal component. Under second-order space-time stationarity, the spatiotemporal variogram may be written as

\[ \gamma(h, \tau) = c(0, 0) - c(h, \tau) \]

so that

\[ \gamma(h, \tau) = \sigma_z^2 (1 - \rho(h) \phi(\tau)). \]

The forms of \( \rho(h) \) and \( \phi(\tau) \) for this study are selected based primarily on known spatial and temporal behavior of the rainfall process. The temporal correlation is written specifically as

\[ \phi(\tau) = \theta^{k - |\tau|}, \]

and the spatial correlation written as

\[ \rho(r) = \beta r K_1(\beta r). \]

where \( r = |h| \), \( \beta \) is a constant, and \( K_1(\cdot) \) is the modified Bessel function of the first kind.

The spatiotemporal aspects of storm cell rainfall generation are modeled by Bras and Rodriguez-Iturbe (1976) with particular emphasis on utilizing the known behavior of storm cells. Although the model is not adaptable to situations other than that for which it was developed, it demonstrates the incorporation of established behavior with the
stochastic aspect of the data and therefore, the physics of the process under investigation may be utilized in describing the covariance structure. For the model, the authors incorporate a meteorological assumption known as Taylor's hypothesis (1937) which states correlation in time is equivalent to that in space if time is transformed to space in the mean direction of storm movement. In short, the hypothesis allows the conversion of time to space and therefore reduces the $\mathbb{R}^s \times T$ spatiotemporal problem to one of purely space in $\mathbb{R}^{s+1}$. The general form of the spatiotemporal covariance for the spatiotemporal process $Z(s_i, t_i)$ where $s_i = (x_i, y_i) \in \mathbb{R}^2$ is written as

$$c(x_i, y_i, t_i; x_j, y_j, t_j).$$

(4.26)

If it is assumed for simplicity that the storm moves parallel to the $x$ axis only, then under Taylor's hypothesis, (4.26) may be written as

$$c(x_i, y_i, t_i; x_j, y_j, t_j) = c(x_i + Ut_i, y_i, x_j + Ut_j, y_j)$$

(4.27)

where $U$ is the storm average velocity in the $x$ direction. The extension to angular storm movement readily follows.

Bastin et al. (1984) models average areal rainfall within two studied basins, time-variant estimates of the simple variogram

$$\gamma(h) = \alpha h^\beta$$

(4.28)

were obtained in order to incorporate the resulting spatiotemporal semivariogram into a regional spatiotemporal model. Initially, a variogram was fit as a function of month $m$ where

$$\gamma(h) = \alpha(m) h^{\beta(m)}.$$

(4.29)
Experimental fitting of (4.29) showed \( \beta(m) \) to be relatively time independent, e.g.

\[ \beta(m) = \beta, \]  
while \( \alpha(m) \) demonstrated clear seasonality and were independently estimated to give the final model,

\[ \gamma(h) = \alpha(m)h^\beta. \]  

(4.30)

The data are indexed by month and the corresponding monthly indexed version of (4.30)

For a study of sulfate deposition, Bilonick (1985) develops a combined space-time zonal anisotropic hybrid model where the spatiotemporal covariance is expressed as the sum of purely spatial and purely temporal semivariograms. The composite spatiotemporal semivariogram model is written as

\[ \gamma(r, \tau) = \gamma_0 + \gamma_p(\tau) + \gamma_s(\tau) + \gamma_L(r), \]  

(4.31)

where \( \gamma_0 \) is the nugget effect, \( \gamma_p(\tau) \) is the overall seasonal component, \( \gamma_s(\tau) \) is the short range temporal component and \( \gamma_L(r) \) is an isotropic spatial component where \( r = |h| \), each specifically written as

\[ \gamma_0 = c_0 \]

\[ \gamma_p(\tau) = c_p\left(1 - 0.5\cos(2\pi\tau)\right) \]

\[ \gamma_s(\tau) = \begin{cases} 0, & \tau = 0 \\ c_s\left(3\tau(2a_s)^{-1} - 2\tau^3a_s^{-3}\right), & 0 < \tau < a_s \\ c_s, & \tau \geq a_s \end{cases} \]

\[ \gamma_L(r) = c_Lr. \]

Models such as this are often plagued by overparameterization and therefore difficult to fit. It is important to obtain reasonable starting values to the parameters and exercise the
fit with adequate sensitivity analysis. If the data demonstrate enough variability, it is not uncommon that the empirical covariance function has non-unique local minima. In this case, the empirical spatiotemporal variogram was sufficiently smooth that estimates to the theoretical model in (4.31) were readily obtained. Furthermore, as discussed above, Myers and Journel (1990) show the zonal anisotropic model to be possibly non-admissible as a valid dependence model. Nevertheless, numerous examples exist in the literature of composite zonal anisotropic models. Erskoll (1997), Buxton and Pate (1994), Huevelink, Musters and Pebesma (1997), De Cesare, Myers and Posa (1997), Haas (1995), and Rouhani and Hall (1989) are some examples.

A general model of atmospheric chemical deposition is developed by Egbert and Lettenmaier (1986) where the spatial covariance structure is estimated based primarily on the method of moments. Instead of seasonal detrending, the weekly incremented data are analyzed by season, the goal being to understand small scale (weekly) and large scale (yearly) variation. The spatial covariance to be estimated is of the form

$$c(s_i, s_j, t) = E(Z(s_i, t)Z(s_j, t))$$  \hspace{1cm} (4.32)

where \( t = |t_i - t_j| \) and the method of moments estimator is given by

$$\hat{c}(s_i, s_j; t) = \frac{1}{N} \sum_{y=1}^{Y} \sum_{w=1}^{W} (Z(s_i, w_i, y) - \bar{Z}(s_i, y))(Z(s_j, w_i + t, y) - \bar{Z}(s_j, y))$$  \hspace{1cm} (4.33)

where \( y = 1, ..., Y \) is the year increment, \( w = 1, ..., W \) is the week increment, \( N \) is the total number of terms in the crossproduct and

$$\bar{Z}(s_k, y) = \frac{1}{W} \sum_{w=1}^{W} Z(s_k, w_i, y).$$  \hspace{1cm} (4.34)
The resulting covariogram matrix is then smoothed by moving average with smoothing parameter \( r_0 \) such that all spatial locations defined by
\[
    r - r_0 \leq |s_i - s_j| \leq r + r_0
\]
are included in the final estimator. For (4.33), it is shown that \( E(\hat{c}(s_i, s_j, t)) = c(s_i, s_j, t) \) prior to the smoothing step. Although the authors did not address the problem of admissibility of the smoothed estimator, a similar estimator was presented in a regression context by Glassbey (1988) who shows that the locally averaged method of moments covariance estimator is in fact positive semidefinite. The estimator is derived as spatially inhomogeneous and temporally stationary, however the extension to the spatially homogeneous and isotropic case may be easily made. In addition, the authors discuss extensions to general kernel type estimators as originally proposed by Toyama and Veneziano (1980).

Rouhani and Wackernagel (1990) develop an approach to the analysis of time-rich and space-poor data entirely within a geostatistical framework. The authors assume intrinsic stationarity with respect to time increments. The empirical semivariogram for the \( k^{th} \) increment is written as
\[
    \hat{\gamma}(s_i, s_j, \tau_k) = \frac{1}{2T_k} \sum_{\alpha=1}^{T_k} \{Z(s_i, t_\alpha) - Z(s_i, t_\alpha + \tau)\}\{Z(s_j, t_\alpha) - Z(s_j, t_\alpha + \tau)\}. \quad (4.36)
\]
where \( \tau \) is the time lag belonging to lag class \( \tau_k \) and \( T_k \) is the cardinality of lag class \( \tau_k \).

The technique used to concentrate on the temporal aspect of the data is that of the nested variogram model introduced by Journel and Huijbregts (1978) where the experimental variograms are modeled as sums of elementary variograms. This model presupposes that
the phenomenon of interest may be modeled by a sum of simple random processes, each contributing to the observed process on a different specific temporal scale. The theoretical variogram is written as

$$\gamma(s_i, s_j, \tau) = \sum_{u=1}^{S} \gamma_u(s_i, s_j, \tau)$$

$$= \sum_{u=1}^{S} b_u(s_i, s_j, \tau) g_u(\tau), \quad (4.37)$$

where $g_u(\tau)$ are elementary variogram functions, known to be conditionally negative definite. Additionally, the coefficient matrices $B_u$ composed of the individual coefficients for the $u^{th}$ elementary variogram, $b_u(s_i, s_j, \tau)$ are themselves conditionally negative definite. The classical variance-covariance matrix $\Gamma$ is related to the coefficient matrices as follows;

$$\Gamma = \sum_{u=1}^{S} B_u \quad (4.38)$$

that suggests $\Gamma$ is composed of a mixture of variogram structures at different spatial scales.

The authors used wave or hole-effect variograms as their elementary variogram functions in order to account for periodicity in the data. Thus each elementary variogram is of the form

$$g_u(\tau) = 1 - \exp(-\tau/r_u) \cos(2\pi\tau/l_u), \quad (4.39)$$

where $r_u$ is the range or extent of oscillation of the hole function and $l_u$ is the period of the cyclic variogram. Once again, the model is recognized as zonal anisotropic and suffers from non-invertibility of the kriging system as described by Myers and Journel.
(1990). However, the model is included here because of its popularity and unique approach to scaled decomposition of the field.

The goal of the method presented by Sampson and Guttrop (1992) is to provide a general smooth admissible nonparametric spatial correlation model of a specified spatial dispersion measure \( \text{var} \left( Z(x_i,t) - Z(x_j,t) \right) \). By using multidimensional scaling, the spatial dispersion measures are transformed from the geographic plane \( G \) onto a Euclidean dispersion plane \( D \) where the distances represent stationary and isotropic spatial dispersions. The variogram model is thus generally represented as

\[
\gamma(h) = D^2(x_a,x_b) = g \left( \| f(x_a) - f(x_b) \| \right),
\]

(4.40)

where \( f \) traditionally adjusts for anisotropy and \( g \) serves the role of the standard variogram. By allowing \( f \) to be a 1-1 nonlinear mapping, e.g. MDS where \( f : G \to D \), a smoothed function \( \hat{g} \) may be estimated that computes an effective isotropic and homogeneous structure function useful for estimating the dispersion structure, in this case the variogram for both observed and unobserved locations.

The method is computationally intensive, and dimensionally handicapped as smoothing splines are used to estimate \( g \). In addition, the selection of smoothing parameters is critical in defining the shape if the dispersion region, but at the same time somewhat arbitrary. The method of Sampson and Guttrop is perhaps the only described technique on nonparametric spatiotemporal covariance modeling. The spatiotemporal extension is discussed in Guttrop, Meiring and Sampson (1994) with application to ozone modeling.
Mardia and Goodall (1993) utilize a factorable covariance structure in the modeling of multivariate observations taken at a number of sparse sites over rich time. The process is modeled as the sum of a selected trend surface and a mean-zero, second-order stationary Gaussian process. The first approach is to temporally detrend the observed $k$-variate process $\mathbf{Z}(s, t) = (Z_1(s, t), \ldots, Z_k(s, t))$ and consider the detrended process as a purely spatial repeated measures experiment for the model

$$\mathbf{Z}(s) = \mathbf{m}(s) + \mathbf{e}(s). \quad (4.41)$$

The second comparative approach is to include the temporal component for the model

$$\mathbf{Z}(s, t) = \mathbf{m}(s, t) + \mathbf{e}(s, t). \quad (4.42)$$

For both cases, the overall covariance $\Sigma$ decomposable as

$$\Sigma = \Sigma_T \otimes \Sigma_S \otimes \Sigma_V, \quad (4.43)$$

where $\Sigma_T$, $\Sigma_S$, $\Sigma_V$ are the covariance matrices of the temporal, spatial and multivariate components respectively. The authors utilize the spatial deformation method of Sampson and Guttorp (1992) in the case space nonhomogeneity.

Posa (1993) describes a time-varying spatiotemporal structure function as a spatial covariance extension in which the process variance or sill is a function of time, i.e.,

$$\gamma(h, t) = A_0(t) S(h), \quad t = 1, \ldots, T \quad (4.44)$$

where $S(h)$ is simply the pure theoretical spatial structure function with a sill equal to one. The representation is somewhat misleading since there is no attempt to explicitly model $A_0(t)$. Instead, discrete empirical determinations of $A_0(t)$ are used to seasonally
index the model in (4.44), hence there is no true modeled temporal dependence structure in space.

Carroll et al., (1997) use a factorable spatiotemporal correlation model to characterize urban ozone concentrations and measure human exposure as a function of population. The eleven monitoring stations are sparsely located in Harris County, Texas but provide incomplete hourly measurements of ozone concentration from 1980 to 1993. The general form of the model is

$$Z(s, t) = g(t) + \epsilon(s, t)$$ (4.45)

where $g(t)$ is a polynomial trend function of hour, month, temperature and squared temperature. The covariance function of the residual process $\epsilon(s, t)$ is then given as

$$\text{cov}(\epsilon(s_i, t_i), \epsilon(s_j, t_j)) = \sigma^2 \rho(d, \nu), \ i \neq j$$ (4.46)

where

$$\rho(d, \nu) = \begin{cases} 1 & \text{if } d = \nu = 0 \\ \phi_{\nu} \psi_{\nu} & \text{otherwise} \end{cases}$$ (4.47)

and $d = \|s_i - s_j\|$, and $\nu = |t_2 - t_1|$. The functions $\phi_{\nu}$ and $\psi_{\nu}$ are defined as

$$\phi_{\nu} = e^{a_0 + a_1 \nu + a_2 \nu^2}$$

$$\psi_{\nu} = e^{b_0 + b_1 \nu + b_2 \nu^2}$$

The full form of the model is then written as

$$\text{cov}(\epsilon(s_i, t_i), \epsilon(s_j, t_j)) = \begin{cases} \sigma^2 & \text{if } d = \nu = 0 \\ \sigma^2 e^{d(a_0 + a_1 \nu + a_2 \nu^2 + b_0 + b_1 \nu + b_2 \nu^2)} & \text{otherwise}, \end{cases}$$ (4.48)
where the first term in the exponent of (4.48) depends on both the time and spatial
difference and the second term depends on only the time difference. The authors do not
analytically prove that (4.48) is positive semi-definite and indeed although it is
empirically useful in that reasonable interpolations were calculated, the theoretical form
is in fact inadmissible as a valid covariogram as indicated by Cressie (1997) in a
rejoinder.

The spatiotemporal covariance function discussed in Christakos (1996, 1997) is
represented as the decomposition of a spatially nonhomogeneous and temporally
nonstationary ordinary covariance, \( c(s_i, t_i; s_j, t_j) \) into a spatially homogeneous and
temporally stationary generalized covariance function (GCF), \( k(h, \tau) \) and a space/time
polynomial of orders \( v \) and \( \mu \) respectively, denoted \( p(v, \mu) \), e.g.

\[
c(s_i, t_i; s_j, t_j) = k(h, \tau) + p(v, \mu), \tag{4.49}
\]

where \( |s_i - s_j| = h \) and \( |t_i - t_j| = \tau \). The GCF can then be expressed in terms of the
corresponding ordinary covariance of a space homogenous and time stationary random
field \( Y(s, t) \) such that \( Y(s, t) = Q[X(s, t)] \) with space homogeneous and time stationary
GCF \( k(h, \tau) \). The space-time operator \( Q \) is a simple detrending mechanism composed
of any applicable functional form. Estimation and detrended prediction are then carried
out on the mathematically convenient process \( Y(s, t) \). Another example of usage of
generalized covariance is Cox and Ensor (1995) where spatiotemporal ozone levels are
modeled.
Numerous techniques exist for developing isotropic spatial covariance structures. Most of these involve the extension of valid covariogram or semivariogram models. Matern (1960) describes among other techniques, one that utilizes the theorem that if \( \mu(a) \) is a measure in a space \( U \) and \( c(r; a) \) is a covariance integrable over the subspace \( V \subset U \) for every \( r \), then the function
\[
c(r) = \int_V c(r; a) \, d\mu(a)
\]  
(4.50)
is also an admissible covariance. The widely used Matern class of covariance models is derived in this manner. Matheron (1973) presents a method of extending admissible covariance structures in \( R^1 \) to \( R^n \) where \( n > 1 \). Briefly stated, if \( c_1(r) \) is a valid covariogram for a stationary process, then by the turning bands operator, the following is a valid covariogram in \( R^n \),
\[
c_n(r) = 2G \left( \frac{n}{2} \right) \pi^{-\frac{n}{2}} \left[ G \left( \frac{n-1}{2} \right) \right]^{-1} \int_0^1 c_1(ur) \left[ 1 - u^2 \right]^{(n-3)/2} \, du,
\]  
(4.51)
where \( G \) is the gamma function. Closed forms of (4.50) and (4.51) both involve Bessel functions and in addition, (4.51) involve Struve functions (Gradshteyn and Ryzhik, 1965). Other approaches involve direct application of difference or differential equations as described by Whittle (1954). By using the Fourier-Bessel approximations of the spectral functions \( \Gamma_n(\omega) \) of a spatial variogram model \( \gamma(r) \), Christakos (1984) presents a method of constructing valid covariograms in \( R^n \). Consider the approximate spectral decomposition,
\[
\rho_n(r) = \frac{c_n(r)}{c(0)} = \frac{(2\pi)^{n/2}}{r^{(n-2)/2}} \int_{R^n} J_{(n-2)/2}(\omega r) \omega^{n/2} \frac{C_n(\omega)}{c(0)} \, d\omega,
\]  
(4.52)
where

\[ \frac{C_n(\omega)}{c(0)} = P_\nu(v), \]  

(4.53)

a probability function of an \( n \) dimensional random vector. Under the premise that it is easier to generate a pdf than a valid spectral density, the generation of a valid spatial covariance in \( \mathbb{R}^n \) is as follows:

1. construct a pdf \( P_\nu(v) \) for the random vector in \( \mathbb{R}^n \).

2. substitute \( P_\nu(v) \) into (4.52) by using the relation (4.53).

Hence, the only apparent obstacle to the method is choice of an appropriate pdf \( P_\nu(v) \).

Whittle (1954) introduced spatial covariance structures based on stochastic differential equations. This work was extended by Vecchia (1985, 1988, 1992), Jones (1989) and Jones and Vecchia (1993) and extended to the spatiotemporal context by Jones and Zhang (1997). The linear regression model for unequally spaced spatiotemporal data is considered,

\[ z(x, y, t) = \sum_{i=1}^{p} \beta_i f_i(x, y, t) + \xi(x, y, t) \]  

(4.54)

where \( f_i(x, y, t) \) are the components of the overall trend function and \( \xi(x, y, t) \) is a homogenous Gaussian spatiotemporal random field with zero mean. Jones and Zhang (1997) present the derivation of a separable covariance model of \( \epsilon(x, y, t) \) based upon the continuous AR(1) process represented by the following stochastic differential equation,

\[ \left( \frac{d}{dt} + \alpha \right) \xi(t) = \epsilon(t), \]  

(4.55)
and continuous fixed-time spatial process represented by the differential equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial x^2} - \phi^2\right)\xi(x, y) = \varepsilon(x, y)$$  \hspace{1cm} (4.56)

where \(\varepsilon(x, y)\) and \(\varepsilon(t)\) are zero-mean Gaussian white noise processes with variance \(\sigma^2\). The resulting correlation function is

$$\Gamma(r, t) = e^{-a\phi r}K_1(\phi r).$$  \hspace{1cm} (4.57)

where \(K_1(\cdot)\) is the modified Bessel function of the second kind, order one with stochastic differential equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial x^2} - \phi^2\right)\left(\frac{d}{dt} + \alpha\right)\xi(x, y, t) = \varepsilon(x, y, t).$$  \hspace{1cm} (4.58)

By considering the stochastic differential equation for the Matern class,

$$\left(\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} - \phi^2\right)^\prime_\prime \xi(x_1, \ldots, x_d) = \varepsilon(x_1, \ldots, x_d),$$  \hspace{1cm} (4.59)

a non-separable model is naturally derived by considering the stochastic differential equation

$$\left(\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} - \phi^2\right)^\prime_\prime - c\frac{\partial}{\partial t}\xi(x_1, \ldots, x_d, t) = \varepsilon(x_1, \ldots, x_d, t).$$  \hspace{1cm} (4.60)

The associated covariance function for \(d = 2\) is

$$\Gamma(r, t) = 2\phi^2\sigma^2\int_0^r \frac{se^{-(s^2 + \phi^2)^\prime_\prime t/c}}{(s^2 + \phi^2)^\prime_\prime} J_0(sr)\,ds$$  \hspace{1cm} (4.61)

and requires numerical integration to evaluate. The described technique requires basic assumptions regarding the physical underlying process, however the stochastic
differential equations proposed above are elementary diffusion models applicable to various random field processes. The process is easily modified but may complicate the derived non-separable covariance.

More recently, an alternative approaches to modeling the spatiotemporal covariance structure followed by prediction are the utilization of hierarchical models as described in Gelfand et al. (1997), Waller, Carlin and Xia (1997) and Berliner, Wikle and Cressie (1999). Another approach worth noting is that described by Huang and Cressie (1996) where a Kalman filter is used to generate predictions of snow water runoff in space and time.

4.3 Discussion

Perhaps the most evident conclusion drawn upon past work is the lack of a cohesive strategy for modeling observations gathered through space and time. This is in contrast to the purely spatial case. The reason for such a lack of uniformity is most likely due to the complexity in sampling design introduced by the time component. For example, observations may be collected at regular or irregular discrete time intervals across all locations, or regular or irregular time intervals by station where each station is sampled at different times. Often both cases but more often the latter case leads to the problem of continuous modeling as discussed in Christakos (1992) or modeling of for example temporally aggregated data as in Haslett and Raftery (1989) or Carroll et al., (1997).

Also to be considered is the added complexity of spatiotemporal trend modeling. For the purely spatial case, there is no clear definition of spatial trend. Over-fitting of a
trend model results in apparent removal of empirical covariance thus making the kriging or prediction stage less optimal in the sense that it merely mimics linear interpolation. For the spatiotemporal case, one must consider the simultaneous trend in both space and time where time is not treated simply as an added dimension. For gross trend removal in the spatiotemporal case, methods commonly used in the purely spatial case are adaptable, however one must consider the periodic nature of temporal data common to most spatiotemporal models, e.g. seasonality or diurnal effects. This requires techniques such as proper aggregation of the data prior to trend modeling.

In practice, one commonly applies *universal spatiotemporal kriging* to a properly estimated spatiotemporal covariance structure. As discussed, modeling of the covariance structure is the most critical stage of spatiotemporal modeling. The assumption of a separable covariance structure simplifies the modeling effort but is inarguably a restriction to the more general case. Furthermore, covariance separability is difficult support in most cases. Therefore, a covariance model that freely allows space-time interaction and considers separability as a special case would be anticipated to provide not only more accurate estimates at the prediction stage, but also reveal more information in terms of the underlying physical mechanism driving the observed process.
5. COMPREHENSIVE SPATIOTEMPORAL COVARIANCE ESTIMATION

The complexities encountered in spatial covariance model selection and estimation are compounded in the space-time domain. As discussed in Chapter 4, assumptions such as space-time separability are necessary in order to apply known admissible covariance functions. The class of models that assumes space-time separability fall under the more general heading of zonal anisotropies. Myers and Journel (1990) show by example that zonal anisotropic covariance matrices may be non-invertible because such models are only conditionally positive semidefinite instead of (strictly) positive definite. This is an important result considering the number of spatiotemporal representations in the literature modeled under the separability of covariances hypothesis.

Extension of the Fourier-Bessel series spatial covariance estimator is easily accomplished to include the additional dimension of time. The advantages of such an estimator is shown to be primarily attributed to the flexibility allowed in modeling the spectrum. The Fourier-Bessel series spatiotemporal covariance representation can be developed as either a separable or non-separable model. Indeed, (non) separability of the spectrum implies (non) separability of the covariance function (Christakos, 1992). For spatiotemporal modeling, the ability to forego the separability assumption strengthens the effort. Non-separability implies space-time interaction to be expressed in a model more representative of the behavior of the true random field. We expect space and time to interact in a more complex manner than with simple additive and multiplicative effects. The same can be said for representation in the spectral domain as well.
5.1 Background on Two-Dimensional Fourier-Bessel Series

Two-dimensional Fourier series representations, a special case of the two-dimensional Fourier-Bessel series expansion, are common in the literature with widespread applications. The extension from the general one-dimensional Fourier-Bessel series representation is easily made to two dimensions. This representation is necessary in order to extend the purely spatial estimator described in chapter 3 to that of an adequate spatiotemporal estimator.

**Definition 5.1:** The two-dimensional Fourier-Bessel expansion of the function \( f(x, y) \) is written as

\[
f(x, y) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} c_{mn} J_{\nu_x}(i_m x) J_{\nu_y}(j_n y)
\]

where the coefficient \( b_{mn} \) is defined as

\[
c_{mn} = \left( \frac{2}{J_{\nu_x+1}(i_m)} \right) \left( \frac{2}{J_{\nu_y+1}(j_n)} \right) \int_0^1 \int_0^1 uv f(u, v) J_{\nu_x}(i_m u) J_{\nu_y}(j_n v) \, du \, dv.
\]

and the series in (5.1) converges to the sum \( f(x, y) \) for any real \( x \in (0, 1) \) and \( y \in (0, 1) \).

\( J_{\nu_x}(w) \) and \( J_{\nu_y}(z) \) are Bessel functions of the first kind of order \( \nu_x \) and \( \nu_y \) respectively and are described in Definition 3.2. Note that the orders \( \nu_x \) and \( \nu_y \) are not required to be the same indicating that the Bessel functions \( J_{\nu_x}(x) \) and \( J_{\nu_y}(y) \) do not necessarily share
common roots $i_m$ and $j_n$, where $i_m, m = 1, \ldots$ are the ordered positive zeros of $J_{\nu_x}(x)$ and $j_n, n = 1, \ldots$ are the ordered positive zeros of $J_{\nu_y}(y)$.

Limit theorems for the two-dimensional Fourier-Bessel series are derived by using results for the one-dimensional case presented in Chapter 3. The following extension theorems for the two-dimensional case are then stated as follows with proof.

**Theorem 5.1**: Let $f(s, t)$ be a function defined arbitrarily on the open product interval $(0,1) \times (0,1)$ and let $\int_0^1 \sqrt{s} f(s, t_0) ds$ and $\int_0^1 \sqrt{t} f(s_0, t) dt$ exist for fixed $s = s_0$ and $t = t_0$ and be absolutely convergent. Let

$$c_{mn} = \left( \frac{2}{J_{\nu_x+1}^2(i_m)} \right) \left( \frac{2}{J_{\nu_y+1}^2(j_n)} \right) \int_0^1 \int_0^1 uvf(u, v) J_{\nu_x}(i_m u) J_{\nu_y}(j_n v) dudv$$

where $\nu_x \geq -\frac{1}{2}$ and $\nu_y \geq -\frac{1}{2}$.

In addition, let $(x, y)$ be any point in the product space $(a, b) \times (c, d)$ such that

$0 < a < b < 1$ and $0 < c < d < 1$ and $f(s, t)$ has limited total fluctuation in $(a, b) \times (c, d)$.

Then the series $\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} c_{mn} J_{\nu_x}(i_m x) J_{\nu_y}(j_n y)$ is convergent with sum

$$\frac{1}{4} \left\{ f(x + 0, y + 0) + f(x + 0, y - 0) + f(x - 0, y + 0) + f(x - 0, y - 0) \right\}.$$

**Proof**: Define the $(M, N)$th partial sum of the iterated series in (5.1) as

$$S_{MN} = \sum_{m=1}^{M} \sum_{n=1}^{N} s_{mn}$$

where
\[ s_{mn} = c_{mn} J_{\nu_1}(i_m x) J_{\nu_2}(j_n y) \]

by Theorem 3.2 and Corollary 3.4.1, the set

\[ \{ s_{mn} \leq \| s_{mn} \| : m, n \in \mathbb{N} \} \quad (5.3) \]

is bounded as a product of bounded elements and \( \mathbb{N} \) is the set of positive integers. Then by hypothesis, there exists a positive real number \( A \) which is an upper bound for the set in (5.3). Specifically, if \( M \) is fixed, we observe that

\[ \sum_{n=1}^{N} s_{mn} \leq \sum_{n=1}^{N} \| s_{mn} \| \leq \sum_{m=1}^{M} \sum_{n=1}^{N} \| s_{mn} \| \leq A \]

for each \( N \in \mathbb{N} \). It thus follows that for each \( M \in \mathbb{N} \), the single series \( \sum_{n=1}^{N} s_{mn} \) is absolutely convergent to an element \( t_M \) in \( \mathbb{R} \). Translating notation back to the original Fourier-Bessel series in (5.1),

\[ S_{MN} = \sum_{m=1}^{M} \sum_{n=1}^{N} c_{mn} J_{\nu_1}(i_m x) J_{\nu_2}(j_n y) \]

\[ = \sum_{n=1}^{N} c_{1n} J_{\nu_1}(i_1 x) J_{\nu_2}(j_n y) + \sum_{n=1}^{N} c_{2n} J_{\nu_1}(i_2 x) J_{\nu_2}(j_n y) + \ldots \]

\[ + \sum_{n=1}^{N} c_{Mn} J_{\nu_1}(i_M x) J_{\nu_2}(j_n y) \]

we infer that

\[ \lim_{N \to \infty} S_{MN} = \sum_{n=1}^{\infty} c_{1n} J_{\nu_1}(i_1 x) J_{\nu_2}(j_n y) + \sum_{n=1}^{\infty} c_{2n} J_{\nu_1}(i_2 x) J_{\nu_2}(j_n y) + \ldots \]

\[ + \sum_{n=1}^{\infty} c_{Mn} J_{\nu_1}(i_M x) J_{\nu_2}(j_n y) \]

and assuming interchangeability of the iterated integrals.
\[
\lim_{N} S_{MN} = \sum_{m=1}^{M} \left\{ \left( \frac{2J_{v_{e}}(i_{m}x)}{J_{v_{e+1}}^{2}(i_{m})} \right) \int_{0}^{1} uJ_{v_{e}}(i_{m}u) \sum_{n=1}^{N} \left( \frac{2J_{v_{e}}(j_{n}y)}{J_{v_{e+1}}^{2}(j_{n})} \right) \int_{0}^{1} v f(u,v) J_{v_{e}}(j_{n}v) \, dv \, du \right\}.
\]

where the inner summation may be written as

\[
\sum_{n=1}^{N} \left( \frac{2J_{v_{e}}(j_{n}y)}{J_{v_{e+1}}^{2}(j_{n})} \right) \int_{0}^{1} v f(u,v) J_{v_{e}}(j_{n}v) \, dv = \sum_{n=1}^{N} b_{n}(x) J_{v_{e}}(j_{n}y)
\]

where

\[
b_{n}(x) = \frac{2}{J_{v_{e+1}}(j_{n})} \int_{0}^{1} v f(u,v) J_{v_{e}}(j_{n}v) \, dv.
\]

so that the series is written as

\[
\sum_{m=1}^{M} \left\{ \left( \frac{2J_{v_{e}}(i_{m}x)}{J_{v_{e+1}}^{2}(i_{m})} \right) \int_{0}^{1} uJ_{v_{e}}(i_{m}u) \sum_{n=1}^{N} b_{n}(x) J_{v_{e}}(j_{n}y) \, du \right\}
\]

Then by Theorem 3.3 and results established above, the series \( \sum_{n=1}^{N} b_{n}(x) J_{v_{e}}(j_{n}y) \) is convergent with sum \( \frac{1}{2} \{ f(x, y+0) + f(x, y-0) \} \) so that the full series is convergent in \( m \) and \( n \) to

\[
\lim_{M, N} S_{MN} = \sum_{m=1}^{M} \left\{ \left( \frac{2J_{v_{e}}(i_{m}x)}{J_{v_{e+1}}^{2}(i_{m})} \right) \int_{0}^{1} uJ_{v_{e}}(i_{m}u) \frac{1}{2} \{ f(u, y+0) + f(u, y-0) \} \, du \right\}
\]

\[
= \sum_{m=1}^{M} a_{m}(y) J_{v_{e}}(i_{m}x)
\]

where

\[
a_{m}(y) = \frac{2}{J_{v_{e+1}}^{2}(i_{m})} \int_{0}^{1} u \frac{1}{2} \{ f(u, y+0) + f(u, y-0) \} J_{v_{e}}(i_{m}u) \, du
\]
and again by Theorem 3.3 and established results above, \( \sum_{n=1}^{\infty} a_n (y) J_{\nu} (i_n x) \) is convergent to
\[
\frac{1}{4} \left\{ f(x+0, y+0) + f(x+0, y-0) + f(x-0, y+0) + f(x-0, y-0) \right\}.
\]

5.2 Spatiotemporal Covariance Modeling: The Fourier-Bessel Series Representation

In definition 3.3, we defined the Fourier-Bessel series representation of the spatial covariance function \( c_d (r) \) for an isotropic random field \( Z(s) \) where \( s \in \mathbb{R}^d \). For the spatiotemporal case, we can now define the spatiotemporal covariance function \( c_d (r, \tau) \) for the space-isotropic and time-stationary random field \( Z(s, t) \) where \( (s, t) \in \mathbb{R}^d \times T \).

**Proposition 5.1:** (Christakos, 1992) The following space-isotropic spectral representation of \( c_d (r, \tau) \) for \( d \) spatial dimensions is given by

\[
C(\omega, \lambda) = (2\pi)^{-d/2} \int_{0}^{\infty} \int_{0}^{\infty} \frac{J_{(d-1)/2} (\omega r)}{(\omega r)^{(d-2)/2}} \exp(i\lambda \tau) r^{d-1} c(r, \tau) \, dr \, d\tau
\]

\[
= (2\pi)^{-d/2} \int_{0}^{\infty} \int_{0}^{\infty} \frac{J_{(d-1)/2} (\omega r)}{(\omega r)^{(d-2)/2}} \left(\{\cos(\lambda \tau) + i \sin(\lambda \tau)\}\right) r^{d-1} c(r, \tau) \, dr \, d\tau
\]

(5.4)

where \( i \) with no subscript is taken in context to be \( i^2 = -1 \).

**Proposition 5.2:** Let \( Z(s, t) \) be a space-isotropic and time-stationary random field where \( (s, t) \in \mathbb{R}^d \times T \) and \( T \subseteq \mathbb{R}_{t,0} = \{ t \in \mathbb{R}, t \geq 0 \} \). Let \( c_d (r, \tau) = 0 \) for \( r > r_c \) and \( \tau > \tau_c \). The Fourier-Bessel series associated with the corresponding covariogram \( c_d (r, \tau) \) is given by
\[ c(r, \tau) = \frac{2(2\pi)^{d/2}}{r \tau c (r / r_c)} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \left( \frac{J_{(d-2)/2} (i_m r / r_c)}{J_{d/2}^2 (i_m)} \right) \left( \frac{J_{-1/2} (j_n \tau / \tau_c)}{J_{1/2}^2 (j_n)} \right) (i_m)^{(d-2)/2} C_d (i_m, j_n) \]

(5.5)

where \( C_d (i_m, j_n) \) is the bivariate spectrum of \( c_d (r, \tau) \).

**Proof:** Write \( c_d (r, \tau) \) as a Fourier-Bessel series of order \( \nu_r = (d - 2)/2 \) and \( \nu_\tau = -1/2 \).

The isotropic spectral representation of \( c_d (r, \tau) \) is rewritten from Proposition 5.1 as

\[
C(\omega, \lambda) = (2\pi)^{-d/2} \int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{J_{(d-2)/2} (\omega r)}{(\omega r)^{(d-2)/2}} \left( \left\{ \cos (\lambda \tau) + i \sin (\lambda \tau) \right\} \right) r^{d-1} c(r, \tau) dr d\tau
\]

(5.6)

Assuming \( c_d (r, \tau) \) to be even in time \( \tau \), e.g. \( c_d (r, \tau) = c_d (r, -\tau) \) and using the properties \( \sin (-\lambda \tau) = -\sin (\lambda \tau) \) and \( \cos (-\lambda \tau) = \cos (\lambda \tau) \), equation (5.6) can be written as

\[
C(\omega, \lambda) = 2(2\pi)^{-d/2} \int_{0}^{\infty} \int_{0}^{\infty} \frac{J_{(d-2)/2} (\omega r)}{(\omega r)^{(d-2)/2}} \cos (\lambda \tau) r^{d-1} c(r, \tau) dr d\tau .
\]

(5.7)

By writing the Fourier-Bessel series representation of \( c_d (r, \tau) \) in explicit form and performing the substitutions \( u = r / r_c \) and \( v = \tau / \tau_c \), the series becomes

\[
c_d (r, \tau) = \frac{(2\pi)^{d/2}}{2} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \left( \frac{2 J_{(d-2)/2} (i_m r / r_c)}{J_{d/2}^2 (i_m)} \right) \left( \frac{2 J_{-1/2} (j_n \tau / \tau_c)}{J_{1/2}^2 (j_n)} \right) \frac{(\tau / \tau_c) (i_m r / r_c)^{(d-2)/2}}{(r / r_c)^{(d-1)/2}} \times
\]

\[
\frac{2}{(2\pi)^{d/2}} \int_{0}^{\infty} \int_{0}^{\infty} \frac{(r / r_c)^{d-2}}{(\tau / \tau_c) (i_m r / r_c) (j_n r / r_c)} u \nu (u, v) J_{(d-2)/2} (i_m u) J_{-1/2} (j_n v) dudv
\]
By substituting (5.7) into (5.8), we obtain
\[
c(r, \tau) = \frac{2(2\pi)^{d/2}}{r_r \tau (r / r_c)^{d/2}} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \left( \frac{J_{(d-2)/2}(i_m r / r_c)}{J_{d/2}(i_m)} \right) \left( \frac{J_{-(d-2)/2}(j_n \tau / \tau_c)}{J_{d/2}(j_n)} \right) (i_m)^{(d-2)/2} C_d(i_m, j_n).
\]

**Proposition 5.3:** Let \( Z(s, t) \) be a space-isotropic and time-stationary random field where \( (s, t) \in R^d \times T \). Let the spatiotemporal covariogram \( c_d(r, \tau) = 0 \) for \( r > r_c \) and \( \tau > \tau_c \). The spectrum of the Fourier-Bessel series associated with \( c(r, \tau) \) is
\[
C_d(\omega, \lambda) = \sum_{m=1}^\infty \sum_{n=1}^\infty C_d(i_m, j_n) \cdot F_2(a, b, z) \cdot k_{mn} \int_0^r r J_{(d-2)/2}(i_m r / r_c) J_{(d-2)/2}(r \omega) \, dr
\]
\[
+ \sum_{m=1}^\infty \sum_{n=1}^\infty C_d(i_m, j_n) \cdot F_2(a, b, z) \cdot k_{mn} \int_0^r r J_{(d-2)/2}(i_m r / r_c) J_{(d-2)/2}(r \omega) \, dr
\]

where
\[
k_{mn} = \sqrt{\frac{i_m \pi}{2 \alpha^2}} \left( \frac{i_m r_c}{\omega} \right)^{d/2} \left( \frac{\csc(j_n)}{J_{d/2}(i_m)} \right)^2
\]

and \( F_2(a, b, z) \) is the generalized hypergeometric function defined as
\[
F_2(a, b, z) = \sum_{k=0}^\infty \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} z^k / k!
\]

where for (5.9), \( a = \{1/4\} \), \( b = \{1/2, 5/4\} \), \( z_2 = -1/4(\lambda + j_n / \tau_c) \tau_c^2 \) and \( z_3 = -1/4(\lambda - j_n / \tau_c) \tau_c^2 \).
proof: Apply the definition of \( C_d(\omega, \lambda) \) as the inverse Fourier transform of the Fourier-Bessel series representation of \( c(r, \tau) \). For \( d \in \{1, 2, 3\} \) the integrals in (5.9) have closed form solutions.

Further simplification for \( d \in \{1, 2, 3\} \) of the spatiotemporal spectrum of the Fourier-Bessel series representation given in (5.9) still contains forms of the generalized hypergeometric function \( _pF_q(a, b; \tau) \). These are easily evaluated with a symbolic computer package such as \textit{Mathematica} and will not be discussed further.

Specific forms of (5.5) for \( d \in \{1, 2, 3\} \) contain simplifications of the Bessel functions of order \( (d-2)/2 \). The derivations follow directly those of the purely spatial case in chapter 3 with a temporal component included. Each spatiotemporal case for \( d \in \{1, 2, 3\} \) includes evaluation of \( J_{-1/2}(2z) = \sqrt{2\pi/z} \cos(z) \).

\[ \textbf{5.3 Estimation} \]

As with the spatial case in Chapter 3, the proposed estimator is based upon the Fourier-Bessel series representation of the spatiotemporal covariogram \( c(r, \tau) \). The infinite sum in the representation (5.5) is truncated in time and space to \( M \) and \( N \) terms accordingly as follows,

\[
c_{d,M,N}(r, \tau) = \frac{2(2\pi)^{d/2}}{r_{t_c}(r/r_c)^{(d-1)/2}} \sum_{m=1}^{M} \sum_{n=1}^{N} \left( \frac{J_{(d-2)/2}(i_m r / r_c)}{J_{d/2}^2(i_m)} \right) \left( \frac{J_{-1/2}(j_n \tau / \tau_c)}{J_{1/2}^2(j_n)} \right) (i_m)^{(d-2)/2} C_d(i_m, j_n)
\]

(5.10)
where \( c(r, \tau) = 0 \) for \( r \geq r_c \) and \( \tau \geq \tau_c \). If \( C_d(\omega, \lambda) \) is represented as an arbitrary parameterized function, \( C_d(\omega, \lambda; \theta) \) dependent upon \( p \)-dimensional parameter vector \( \theta \in \Theta_p \subset \mathbb{R}^p \), then the series representation in (5.10) becomes

\[
c_{d,M,N}(r, \tau; \theta) = \frac{2(2\pi)^{d/2}}{r_c \tau_c (r / r_c)^{d/2 - 1}} \sum_{m=1}^{M} \sum_{n=1}^{N} \left( \frac{J_{(d-2)/2}(i_m r / r_c)}{J_{-1/2}(i_m)} \right) \left( \frac{J_{-1/2}(j_n \tau / \tau_c)}{J_{1/2}(j_n)} \right) (i_m)^{(d-2)/2} \ C_d(i_m, j_n; \theta)
\]

(5.11)

where \( M, N < \infty \). By Bochner's theorem, in order for (5.11) to be an admissible spatiotemporal covariance, the Fourier transform of \( c_d(r, \tau; \theta) \) is necessarily nonnegative and bounded, e.g.

\[
C_{d,M,N}(\omega, \lambda; \theta) \geq 0 \text{ for all } \theta \in \Theta_p \text{ and } \omega, \lambda \geq 0.
\]

The true spectrum, \( C_d(\omega, \lambda) \) must by necessity also be nonnegative and bounded, e.g.

\[
C_{d,\infty}(\omega, \lambda; \theta) \geq 0 \text{ for all } \theta \in \Theta_p \text{ and } \omega, \lambda \geq 0.
\]

Following the spatial case, we propose the two-dimensional scaled Johnson \( S_U \) distribution as follows (Mardia, 1970):

\[
C_{S_U}(\omega, \lambda; \theta) = \frac{\alpha}{(1 - \rho^2)^{1/2}} \frac{\delta_1 \delta_2}{\beta_1 \beta_2 \sqrt{1 + (\omega / \beta_1)^2} \sqrt{1 + (\lambda / \beta_2)^2}} \times \exp \left( \frac{-1}{2(1 - \rho^2)} \left[ F(\omega; \beta_1, \delta_1) - 2\rho G(\omega; \beta_1, \beta_2, \delta_1, \delta_2) + H(\lambda; \beta_2, \delta_2) \right] \right)
\]

(5.12)

where

\[
F(\omega; \beta_1, \delta_1) = \delta_1 \log \left( \frac{(\omega / \beta_1) + \sqrt{(\omega / \beta_1)^2 + 1}}{1} \right)
\]
\[ G(\omega, \lambda; \beta_1, \delta_2, \beta_2, \delta_2) = \]
\[ \delta_1 \log \left( (\omega / \beta_1) + \sqrt{(\omega / \beta_1)^2 + 1} \right) \times \]
\[ \delta_2 \log \left( (\lambda / \beta_2) + \sqrt{(\lambda / \beta_2)^2 + 1} \right) \]
\[ H(\lambda; \beta_2, \delta_2) = \delta_2 \log \left( (\lambda / \beta_2) + \sqrt{(\lambda / \beta_2)^2 + 1} \right)^2 \]

where \((\omega, \lambda) \in (-\infty, \infty)\) and \(\alpha > 0\) in order to assure nonnegativity. The scaling coefficient \(\alpha\) takes into account that we are modeling the spectrum of a bivariate covariogram where the spectrum is a nonnormalized bivariate density. For modeling the spectrum of the spatiotemporal correlation or correlogram, the scale parameter is fixed at unity.

Construction of a spatiotemporal model of the bivariate covariogram follows from that for the one-dimensional case presented in Chapter 3. Assuming the bivariate Johnson \(S_u\) distribution takes into account dimension-dependent forms, e.g.

\[ C_d(\omega, \lambda; \theta) = C_{S_u}(\omega, \lambda; \theta), \]

the series representation in (5.11)

\[ c_{d,M,N}(r, \tau; \theta) = \frac{2(2\pi)^{d/2}}{r_t^d (r / r_e)^{d/2-1}} \sum_{m=1}^{M} \sum_{n=1}^{N} \left( \frac{J_{(d-2)/2}(i_m r / r_e)}{J_{(d/2)}(i_m)} \right) \left( \frac{J_{-1/2}(j_n \tau / \tau_e)}{J_{1/2}(j_n)} \right) (i_m)^{(d-2)/2} C_{S_u}(i_m, j_n; \theta) \]

(5.13)

We then propose the following model of the spatiotemporal covariogram,

\[ c_d^*(r_k, \tau_i) = \]
\[ \frac{2(2\pi)^{d/2}}{r_t^d (r / r_e)^{d/2-1}} \sum_{m=1}^{M} \sum_{n=1}^{N} \left( \frac{J_{(d-2)/2}(i_m r_k / r_e)}{J_{(d/2)}(i_m)} \right) \left( \frac{J_{-1/2}(j_n \tau_i / \tau_e)}{J_{1/2}(j_n)} \right) (i_m)^{(d-2)/2} C_{S_u}(i_n, j_n; \theta) + \epsilon_{kl} \]

\[ = c_{d,M,N}(r_k, \tau_i; \theta) + \epsilon_{kl} \]

(5.14)
where \( \theta = \{ \alpha, \rho, \beta_1, \beta_2, \delta_1, \delta_2 \} \in \Theta_6 \) and \( c_d^*(r_k, \tau_l) \) is the \((k,l)\)th observed element of the empirical covariogram, \( k = 1, \ldots, K \) and \( l = 1, \ldots, L \). The parameter vector \( \theta \) is estimated by minimizing the least squares objective function \( q(\theta) \)

\[
q(\theta) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \left( c_d^*(r_k, \tau_l) - c_{d,M,N}(r_k, \tau_l; \theta) \right)^2
\]

so that \( \hat{\theta} \) is the maximum likelihood estimator of \( \theta \) given by

\[
\hat{\theta} = \min_{\theta \in \Theta} q(\theta) \quad \text{and} \quad \hat{\theta} = \{ \hat{\alpha}, \hat{\rho}, \hat{\beta}_1, \hat{\beta}_2, \hat{\delta}_1, \hat{\delta}_2 \} \in \Theta_6.
\]

The \( M, N \)-truncated Fourier-Bessel series estimator of the covariogram is then given by

\[
c_{d,M,N}(r, \tau; \hat{\theta}) = \frac{2(2\pi)^{d/2}}{r_c^{d/2-1}} \sum_{m=1}^{M} \sum_{n=1}^{N} \left( \frac{J_{(d-2)/2}(i_m r / r_c)}{J_{d/2}(i_m r)} \right) \left( \frac{J_{(d-2)/2}(j_n \tau / \tau_c)}{J_{d/2}(j_n \tau)} \right) (i_m)^{(d-2)/2} C_{s_c}(j_n, j_n; \hat{\theta})
\]

Thus far, we have not discussed selection of the space and time limit parameters \( r_c \) and \( \tau_c \). In practice, one could set the limit parameters based upon knowledge of the process combined with observation of the empirical spatiotemporal covariogram.

Another alternative selection method consists of including \( r_c \) and \( \tau_c \) in the parameter set and estimation using the objective function in (5.15) where the parameter vector is modified to include the limit parameters, e.g. \( \theta = \{ \alpha, \rho, \beta_1, \beta_2, \delta_1, \delta_2, r_c, \tau_c \} \in \Theta_8 \). This increases the estimation time and places additional burden on the particular estimation technique used. The Nelder-Mead simplex algorithm has demonstrated proficiency in the estimation of comparatively large parameter vectors. This method is used in the spatiotemporal application of sea-surface temperatures in Chapter 7. Alternatively, one
could simply select \( r_e \) and \( \tau_e \) at distances where spatiotemporal dependence is known to be zero.

Sampling in space and time are generally characterized in practice by a large number of sample points. Although fitting (5.16) to the space-time covariogram cloud is an option, it may be more practical to consider the common binned representation of the empirical spatiotemporal covariogram. This is particularly applicable when one considers the covariogram cloud to consist of \( n(n-1)/2 \) individual observations for \( n \) sampled locations in space and time. Binning is commonly performed on a trial-and-error basis until a reasonable representation of the observed covariogram is computed. This method of subjective binning could be called into question but considering the number of sampled locations in space and time combined with the added computational burden of the estimator in (5.16), the current choice of empirical covariogram is the binned version. Certainly as computational performance improves, the preferred choice will be the spatiotemporal covariogram cloud.

The space homogeneous / time stationary method of moments estimator of the population spatiotemporal covariogram is given by

\[
\hat{c}(r, \tau) = \frac{1}{\bar{N}(r, \tau)} \sum_{i \neq j \in \mathcal{N}(r, \tau)} (Z(s_i, t_i) - \bar{Z})(Z(s_j, t_j) - \bar{Z}),
\]

(5.17)

where

\[
\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z(s_i, t_i)
\]

and

\[
\bar{N}(r, \tau) = \left\{(s_i, s_j) : \|s_i - s_j\| = r, t_i - t_j = \tau; t_i - t_j \geq 0; i, j = 1, \ldots, n\right\}.
\]
and \( r \) and \( \tau \) are taken in context to be space and time increments. Often \( r \) and \( \tau \) are binned such that \( r \in B(\{r(l)\}) \) where \( l = 1, \ldots, B_r \) and \( \tau \in B(\{\tau(k)\}) \) where \( k = 1, \ldots, B_\tau \). This is usually done with irregularly sampled locations in space and/or time. Journel and Huijbregts (1978) recommend the size of the bins be minimized to contain no less than 30 distances. For purposes of estimation using the empirical binned spatiotemporal covariogram in (5.17), we replace \( c_d^*(r_k, t_l) \) with \( \hat{c}(r_k, \tau_i) \) for the \( k \) th spatial and \( l \) th temporal bin. The resulting objective function is then

\[
q(\theta) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \left( \hat{c}(r_k, t_l) - c_{d,M,N}(r_k, t_l; \theta) \right)^2. 
\] (5.18)
6. APPLICATION: SPATIAL MODELING OF TOXIC SEDIMENT LOAD IN GALVESTON BAY ESTUARY

The Environmental Monitoring and Assessment Program (EMAP) is a research program to develop the tools necessary to monitor and assess the status and trends of national ecological resources. EMAP objectives are to advance the science of ecological monitoring and ecological risk assessment, guide national monitoring with improved scientific understanding of ecosystem integrity and dynamics, and demonstrate the Committee on Environmental and Natural Resources (CENR) framework through large regional projects. Galveston Bay estuary is a targeted site for EMAP reporting. The proximity of Galveston Bay to major industrial centers makes it particularly vulnerable to anthropogenic effects. The bay is a source of commercial shellfish harvesting and extensive recreational angling. One objective of the EMAP survey is to determine the levels of heavy metal contaminants in the biotic component of the estuary and the level of contamination in the sediment as well. Sources of contamination include runoff, vessel waste and industrial point discharge.

The distribution of sediment contaminants is complicated by regular dredging activities in the ship channel that mix and redistribute contaminants dependent on location, e.g. depth and flow conditions. However, for the most part dredging occurs only in the ship channel where a proper depth must be maintained in order to allow navigable passage. Contaminant distribution is also dependent upon sediment type with finer sediments expected to retain the contaminant more efficiently than coarse sediment or sand with a greater degree of interstitial flushing.
The organic and inorganic compound concentrations measured included: 15 major and trace elements, 27 individual Poly-Aromatic Hydrocarbon (PAH) compounds, 30 pesticides (including DDT's), 25 individual Poly-Chlorinated Biphenyl (PCB) congeners, mono-, di- and tri-butyltin (MBT, DBT, TBT), 1 organophosphate, Total Organic Carbon (TOC) and acid volatile sulfides (AVS). Values include concentrations summed for several major groups: total PAH's, Low and High Molecular Weight PAH's, PCB's, DDT's, BHC’s and Chlordanes. As an applied exercise, the spatial Fourier-Bessel series estimator of the covariogram \( c_2(r) \) is estimated and kriging or prediction plots are generated for summed or total PCB's, total DDT's, total PAH's and an important heavy metal, Arsenic.

All data used in the analysis are publicly available at http://www.epa.gov/emap/.

6.1 Data Description

Sampling stations were randomly selected within the open bay as locations to conduct a specific suite of monitoring activities and sample gathering. More intensive non-random sample locations were also surveyed in tributaries that feed the bay. A map of station locations is shown in Figure 6.1. The Galveston Bay system is composed of Galveston Bay proper, Trinity Bay and East Bay. West Bay contains a single station. The primary tributaries that feed the bay system are the Trinity River with mouth at the northeastern corner feeding Trinity Bay, the natural drainage system from Buffalo Bayou and the Houston ship channel located in the northwestern corner of the bay and the ship channel opening into East Bay.
Figure 6.1 Station locations for the EMAP Galveston Bay study.
Samples were obtained by lowering a grab sampler through the water column such that travel through the last 5 meters was no faster than 1 m/sec. The grab penetrated the sediment by gravity releasing a trigger allowing the jaws to close. When the grab was pulled from the sediment using the winch, the jaws closed, encapsulating the sediment sample. The chance of sampling the exact same location twice was minimized. After three grabs were taken, the boat was moved five meters downstream by letting out the appropriate length of anchor line.

Stainless steel utensils were used to remove the top two cm of sediment from a grab. The sediment was removed to a stainless steel bowl and placed in a cooler of ice to remain cold, but unfrozen. The grab sampler was rinsed and re-deployed. This procedure was repeated until approximately 3,000 cc of sediment was collected. The sediment was mixed by hand until thoroughly homogenized, and aliquots were placed immediately into pre-cleaned glass jars (for organics) or plastic containers (for inorganics and AVS). The samples were immediately stored on ice following collection. The remainder of the sediment was split between grain size and sediment toxicity samples. Sampling was conducted between 24 September 1993 and 10 October 1993.

A successful grab had relatively level, intact sediment over the entire area of the grab and a sediment depth of 7-10 centimeters. Unacceptable grabs included those: containing no sediments, which were partially filled or had shelly substrates or grossly slumped surfaces. Grabs completely filled to the top, where the sediment was oozing out of the hinged top, were also unacceptable.

Upon receipt at the laboratory, the samples were frozen pending analysis. The frozen sediment samples were thawed and thoroughly homogenized prior to analysis.
Separate aliquots of the homogenized sediment were removed. The aliquots were processed for several types of chemical analyses. These included: inorganic analyses (major and trace elements); butyltins (MBT, TBT and DBT)); organic analyses (PAH's, PCB's, aliphatic hydrocarbons, and pesticides); total organic carbon (TOC); and acid volatile sulfides (AVS). Table 6.1 indicates summary statistics for each of the four chemical contaminants.

**Table 6.1** Summary statistics for the modeled contaminants (EMAP study).

<table>
<thead>
<tr>
<th>Contaminant</th>
<th>n</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>44</td>
<td>5.98</td>
<td>2.04</td>
<td>1.62</td>
<td>11.090</td>
</tr>
<tr>
<td>PCB</td>
<td>43</td>
<td>2.75</td>
<td>2.36</td>
<td>0.10</td>
<td>9.00</td>
</tr>
<tr>
<td>DDT</td>
<td>40</td>
<td>0.50</td>
<td>0.45</td>
<td>0.10</td>
<td>2.00</td>
</tr>
<tr>
<td>PAH</td>
<td>42</td>
<td>177.13</td>
<td>234.56</td>
<td>5.70</td>
<td>1375.40</td>
</tr>
</tbody>
</table>

### 6.2 Covariogram Estimation

We formulate the objective of the development of spatial maps of the contaminants as that to identify potential point sources. Although raw contamination level of the sediment is important to commercial and recreational fisheries of particularly shellfishes such as oysters, we adjust the observed contamination level for sediment type and total organic carbon because both are important factors in retention of the contaminant and thereby confuse the identification of a potential source. Adjustment for sediment type and total organic carbon may also better represent the level of contaminant in the water column. Sediment type is coded as percent silt-clay or 1-percent sand. Total
organic carbon is measured analytically. Raw contaminant level is adjusted by first estimating the following model and then using the residuals to model the covariograms.

\[ z_i = \mu + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \]  \hspace{1cm} (6.1)

where

- \( z_i \) = observed concentration at station \( i \), \( i = 1, \ldots, n_r \), and \( n_r \) total stations
- \( \mu \) = overall mean
- \( \beta_1 \) = regression coefficient for sediment type
- \( \beta_2 \) = regression coefficient for total organic carbon
- \( x_{i1} \) = measured sediment type at station \( i \)
- \( x_{i2} \) = measured total organic carbon at station \( i \)
- \( u_i \) = model error

Therefore, the modeled random variable is

\[ \hat{u}_i = z_i - \left( \hat{\mu} + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} \right) \]  \hspace{1cm} (6.2)

In practice, the most common empirical covariogram is a binned estimate at a specified lag distance. The selection of bin widths and total lag distance generally follows two widely accepted rules:

i) each bin must have a cardinality of at least 30, and

ii) the empirical covariogram is computed for \( 1/2 \) the maximum distance of the study region.

With small sample sizes, rule i) above is often violated resulting covariograms that fail to adequately represent the observed field. It is often preferable to adjust the bin widths such that each bin contains an acceptable number of observed covariogram cloud points. Empirical covariograms tend also to show increased variability with large distances, hence attention to rule ii) is important. Furthermore, we are primarily concerned with
predicting points using neighborhood of observed points of relatively small distances from the predicted in order to increase the reliability of the estimate.

The procedure of covariogram estimation is a two-stage process. The empirical representation of the binned covariogram is still largely arbitrary and often subjectively predicated by the experimenter. Knowledge of the nature of dependence of the unknown random field is often taken into account in deriving the empirical covariogram and such knowledge often drives the binning itself.

For the EMAP covariograms, the study region is defined by an $80km \times 80km$ region. We therefore compute the empirical covariogram for a maximum distance of $40km$. A total of six empirical binned covariogram estimates are computed for each of the four contaminants. The first bin distance is zero. This point is an estimate of the process variance, $\sigma^2$. The second bin is computed for the distance $(0 < r < 5km)$. The remaining four bins are relatively equally spaced with width approximately $10km$. All bins are chosen such that the cardinality is at least 30. For estimation of the covariogram, the mean binned distance is used.

Plots of directional empirical variograms for the four contaminants were generated in order to visually assess the assumption of isotropy. These are shown in Figs. 6.2a - 6.2d. The empirical N-S and E-W variograms are similar in structure such that there is no reason to question the isotropy assumption. In addition, plots of detrended contaminant level by direction were generated in order to assess stationarity. These are shown in Figs. 6.3a - 6.3d. No detectable directional trends are observed for any of the four contaminants. We therefore fit the same radially invariant covariogram model within contaminant for all directions without preliminary spatial detrending.
The covariogram model used is the two-dimensional experimental Fourier-Bessel series form given by

\[ c_{d,M} (r; \hat{a}_0, \ldots \hat{a}_p) = (2\pi)^{d/2} \sum_{m=1}^{M} \frac{2J_{(d-2)/2}(i_m r / r_c)(i_m / r_c)^{(d-2)/2}}{J_{d/2}^2(i_m)(i_m / r_c)^{(d-2)/2} r_c^2} \exp \left( \sum_{k=0}^p \hat{a}_k (i_m / r_c)^k \right). \]

Starting values for all parameters and all four contaminants are the same with the exception of \( r_c \) upon which acceptable convergence relied. For all contaminants, the series was computed for first \( M = 10 \) and then for \( M = 100 \) using estimates obtained from the lower order series as starting values as described in Chapter 3. Final estimates are shown in Table 6.2. Covariogram estimation was performed using the Nelder-Mead simplex optimization algorithm. The empirical and fitted experimental covariogram model are shown for each of the four contaminants in Figs. 6.4a - 6.4d.

**Table 6.2** Final parameter estimates for spatial covariograms (EMAP study).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>As</th>
<th>DDT</th>
<th>PAH</th>
<th>PCB</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta )</td>
<td>-2.75</td>
<td>2.97</td>
<td>-4.68</td>
<td>-4.01</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>-57.98</td>
<td>-13.93</td>
<td>-28.83</td>
<td>-79.28</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>52.72</td>
<td>13.44</td>
<td>285.5</td>
<td>80.60</td>
</tr>
<tr>
<td>( r_c )</td>
<td>40029</td>
<td>9993</td>
<td>13119</td>
<td>39989</td>
</tr>
<tr>
<td>MSE</td>
<td>0.068</td>
<td>0.000</td>
<td>0.116</td>
<td>0.117</td>
</tr>
</tbody>
</table>

All of the covariograms are similar in shape and suggestive of strong local dependence on the order of \(< 5000m\). The Fourier-Bessel model tends to fit well close to the origin and in most cases closely interpolates the empirical covariogram within \( 0 \) to \( 5000m \). This is a desirable feature in that spatial prediction will place more weight on closer within-neighborhood observations.
Figure 6.2a Directional empirical covariograms for As.
Figure 6.2b  Directional empirical covariograms for DDT.
**Figure 6.2c** Directional empirical covariograms for PAH.
Figure 6.2d Directional empirical covariograms for PCB.
Figure 6.3a  Directional trend plots for As.
Figure 6.3b Directional trend plots for DDT.
Figure 6.3c  Directional trend plots for PAH.
Figure 6.3d Directional trend plots for PCB.
Figure 6.4a  Empirical and fitted experimental covariogram for As.
Figure 6.4b  Empirical and fitted experimental covariogram for DDT.
EMAP / total PAH
observed and fitted covariograms

Figure 6.4c Empirical and fitted experimental covariogram for PAH.
Figure 6.4d Empirical and fitted experimental covariogram for PCB.
6.3 Spatial Prediction

The simple kriging or prediction model is described in Cressie (1993). We assume a zero-mean process of the modeled random variable \( \tilde{u}_i \). We seek to optimize the predictions based upon the squared-error loss function,

\[
L(u(s_0), p(U; s_0)) = (u(s_0) - p(U; s_0))^2
\]

where \( u(s_0) \) is the prediction at location \( s_0 \). \( U = \{\tilde{u}(s_1), ..., \tilde{u}(s_n)\} \) and \( p(U; s_0) \) is the predictor function. The best linear predictor is given by

\[
p(U; s_0) = \sum_{i=1}^{n} l_i \tilde{u}(s_i) + k
\]

such that \( E(u(s_0) - p(U; s_0))^2 \) is minimized. Such an optimal predictor \( p^* \) is given by

\[
p^*(U; s_0) = c^T \Sigma^{-1} U
\]

where \( c = (c(s_0, s_1), ..., c(s_0, s_n))^T \), \( \Sigma \) is an \( n \times n \) whose \( (i, j) \) th element is \( c(s_i, s_j) \), the spatial covariance or covariogram between observations \( \tilde{u}(s_i) \) and \( \tilde{u}(s_j) \). The minimized mean-squared prediction error is given by

\[
\sigma_p^2 = C(s_0, s_0) - c^T \Sigma^{-1} c
\]

For the studied EMAP contaminants, predictions were generated for the entire study area on a 5000m \( \times \) 5000m grid. Prediction neighborhoods were computed by beginning with a 5000m radius and expanding by 1000m increments until each prediction contained at least one observed location. The predictions were then contoured using a smoothing spline with a small penalty parameter of \( \lambda = .0005 \). Prediction errors were also generated for each prediction location.
Performance of the estimated covariogram and of the prediction method in
general is assessed using cross-validation as described in Cressie (1993). The closeness
of the predicted values to the true values can be characterized by two measures, e.g.

\[ CV_A(U) = \frac{1}{n} \sum \frac{\{\hat{u}(s_j) - \hat{u}_{-j}(s_j)\}}{\sigma_{-j}(s_j)} \]  

or alternatively,

\[ CV_B(U) = \left( \frac{1}{n} \sum \left( \frac{\{\hat{u}(s_j) - \hat{u}_{-j}(s_j)\}}{\sigma_{-j}(s_j)} \right)^2 \right)^{1/2} \]

where \( \hat{u}_{-j}(s_j) \) is the predicted contaminant level using the abbreviated observation
vector \( U = \{\hat{u}(s_1), ..., \hat{u}(s_{j-1}), \hat{u}(s_{j+1}), ..., \hat{u}(s_n)\} \) and \( \sigma^2_{-j}(s_j) \) is the prediction error for
\( \hat{u}_{-j}(s_j) \). For assessment of the EMAP contaminant predictions, the covariogram was
estimated using the full ensemble of observations, \( U = \{\hat{u}(s_1), ..., \hat{u}(s_n)\} \). The cross-
validation measure in (3.45) should be approximately 0 and the cross-validation measure
in (3.46) should be approximately 1 for an adequate predictive model. Cross-validation
using both measures for each of the four EMAP contaminants are shown in Table 6.3.

<table>
<thead>
<tr>
<th>Contaminant</th>
<th>( CV_A(U) )</th>
<th>( CV_B(U) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>As</td>
<td>-0.04531</td>
<td>0.99237</td>
</tr>
<tr>
<td>DDT</td>
<td>0.00504</td>
<td>1.01816</td>
</tr>
<tr>
<td>PAH</td>
<td>-0.04192</td>
<td>0.93342</td>
</tr>
<tr>
<td>PCB</td>
<td>0.01183</td>
<td>1.17325</td>
</tr>
</tbody>
</table>
No specific acceptance interval for either $CV_A(U)$ or $CV_B(U)$ is known in the literature. Cressie (1993) suggests "close to zero" for $CV_A(U)$ and "close to one" for $CV_B(U)$.

The crossvalidation results suggest acceptable performance of the experimental covariogram estimator. The worst case is for PCB and crossvalidation measure $CV_B(U)$ in (6.8), but the computed value of 1.17 is still reasonable.

Histograms of the predictions are shown in Figs. 6.5a, 6.6a, 6.7a and 6.8a for each of As, DDT, PAH and PCB respectively. Prediction maps are illustrated in Figs. 6.5b, 6.6b, 6.7b and 6.8b for each of the four modeled contaminants. Only elevated prediction maps representing the upper tails of the histograms are contoured for clarity. Histograms of the prediction errors are shown in Figs. 6.5c, 6.6c, 6.7c and 6.8c. Prediction error maps are shown in Figs. 6.5d, 6.6d, 6.7d and 6.8d. In order to plot closer-to-true predicted values, the expected contaminant level was added to each modeled residual from (6.2). The expected silt-clay composition and total organic carbon for the bay was computed as the means,

\[
\bar{X}_1 = \frac{1}{n_i} \sum_{i=1}^{n_i} x_{i1}
\]

\[
\bar{X}_2 = \frac{1}{n_i} \sum_{i=1}^{n_i} x_{i2}
\]

where $x_{i1}$ is the observed silt-clay content at station $i$ and $x_{i2}$ is the observed total organic carbon at station $i$. Each kriged prediction was then adjusted as follows,

\[
\hat{Z}(s_j) = \hat{u}(s_j) + \mu + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2
\] (6.9)
where $\hat{\mu}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are least squares estimates obtained from fitting the model (6.1). The prediction errors were the approximated by assuming the covariates are measured without error, e.g.

$$
\hat{\sigma}_p^2 = \hat{\sigma}_{\mu}^2 + \hat{\sigma}_{\beta}^2 + \left(1/n_x^2\right) \sum_{i=1}^{n} \left(\hat{\sigma}_{\mu_i}^2 x_{i1}^2 + \hat{\sigma}_{\beta_i}^2 x_{i2}^2\right)
$$

(6.10)

where $\hat{\sigma}_p^2$ is the prediction error estimate of (6.6), $\hat{\sigma}_{\mu}^2$, $\hat{\sigma}_{\beta_1}^2$ and $\hat{\sigma}_{\beta_2}^2$ are the prediction error estimates of are the estimation error estimates for $\mu$, $\beta_1$ and $\beta_2$ respectively.

### 6.4 Interpretation

Arsenic concentrations are highest in the center of Trinity Bay followed by elevated patches of high concentration in East bay and western Galveston Bay. Slightly elevated concentrations are also seen around the city of Galveston on Galveston Island. Slightly elevated concentrations are also observed around the mouth of Clear Lake. The source of Arsenic contamination is unknown. Arsenic may be more evenly distributed in the Bay region, but dredging within the high concentration zone maintains elevated levels in the surface sediment.

DDT concentrations are observed to be elevated at the mouth of the Trinity River, and in Trinity Bay, the source of which is likely agricultural runoff. The high concentration observed in the center of Galveston Bay might be a result of current or tidal patterns combined with the relative geographical bottleneck in the area. The source is still assumed to be agricultural contributed by both the Trinity River and Buffalo Bayou drainage system. Also evident is an elevated concentration around the city of Galveston, again agriculturally related.
PCB contamination is elevated throughout the northern end of the Bay. A likely source is the Buffalo Bayou drainage system from Houston. Elevated levels are also observed around the Clear Lake area and again, the city of Galveston. Sources are probably industrial effluent, possibly refineries.

A notable feature of the kriging map for PAH is the location of elevated contours surrounding the Trinity Bay Oil Field. Drilling production activities in the region are the most likely cause of high PAH levels in Trinity Bay. The specific source may be transfer spillage, but is most likely related to effluent produced by past drilling activities as it is assumed the oil is transported primarily by pipeline as opposed to tanker.
Figure 6.5a Prediction histogram for As.
Figure 6.5b Prediction map for As.
Figure 6.5c Prediction error histogram for As.
Figure 6.5d Prediction error map for As.
Figure 6.6a Prediction histogram for DDT.
Figure 6.6b Prediction map for DDT.
Figure 6.6c Prediction error histogram for DDT.
Figure 6.6d  Prediction error map for DDT.
Figure 6.7a Prediction histogram for PAH.
**Figure 6.7b** Prediction map for PAH.
Figure 6.7c Prediction error histogram for PAH.
Figure 6.7d Prediction error map for PAH.
**Figure 6.8a** Prediction histogram for PCB.
Figure 6.8b Prediction map for PCB.
Figure 6.8c  Prediction error histogram for PCB.
Figure 6.8d Prediction error map for PCB.
7. APPLICATION: SPATIOTEMPORAL MODELING OF SEA SURFACE TEMPERATURES OFF SOUTHERN CALIFORNIA (CalCOFI)

The California Cooperative Fisheries Investigations Study (CalCOFI) was begun in 1949 in the Pacific Ocean off southern California. The existence of such a long-term spatial study is rare as fieldwork of the nature of the CalCOFI study is expensive. However, the ability to monitor important fisheries and climatic events associated with the California current have proven productive.

The region contains an upwelling zone of nutrient rich water that supports a large phytoplankton biomass and consequently important economic fisheries controlled by the fluctuation of environmental parameters. The predominant circulation feature of the area is the California Current that is greatly influenced by variation in atmospheric pressure and the prevailing wind field. The physical properties of the region are greatly influenced by climatic events that include the El Nino – Southern Oscillation or ENSO. This externally induced physical variability greatly influences the biology of the region. Present macrozooplankton concentration is trending downward and is lower than has been observed for the 45 years of the study (Roemmich and McGowan 1995). This contingent includes a number of economically important larval species. This is attributed to an increase in observed water temperature patterns across the region and is a matter of great concern since the macrozooplankton component of the region ultimately supports the economically important fish species at higher trophic levels. In addition, abundance of oceanic birds in the CalCOFI study region has declined steadily since 1988 (Veit et al. 1996).
The El Niño Southern Oscillation (ENSO) is a pronounced disturbance of the atmospheric circulation over lower latitudes of the Pacific sector associated with a redistribution of mass during the course of the El Niño warming and La Niña cooling of the equatorial Pacific. ENSO events are particularly characteristic of the study region and greatly influence cyclic temperature fluctuations. These occur on roughly 7 and 10 year cycles and vary in intensity. The primary purpose of generating contoured SST anomaly maps is to identify sources of hypothesized long-term warming in the study region. Therefore, the observed SST’s must be adjusted for ENSO events. In order to accomplish this, we utilize the Southern Oscillation Index (SOI) as a covariate in the trend model described below. The SOI is defined as the normalized pressure difference between Tahiti and Darwin. There are several slight variations in the SOI values calculated at various centers. Here we calculate the SOI based on the method given by Ropelewski and Jones (1987). The reader is also referred to Allan et al. (1991) and Können et al. (1998) for details of the early pressure sources and methods used to compile the series from 1866 onwards.

7.1 Data Description

The study area is defined as an 800km square region off Point Conception and the Channel Islands. The region currently includes 65 stations located along evenly spaced transects perpendicular to the southern California shoreline. The sampling for each station is currently 4 times per year. This basic sample schedule and design has been maintained since the onset of the study, however station locations and sampling intervals have varied throughout the study. There are no single stations with long-term regular
sampling intervals. In addition, stations were moved periodically and stations were irregularly added or removed from the area. Only one year, 1979 is absent from the study. Sampling locations for each year are shown in Figure 1.
CalCOFI station locations

Figure 7.1 CalCOFI station locations by year.
CalCOFI station locations

Figure 7.1 CalCOFI station locations by year (cont’d).
CalCOFI station locations

Figure 7.1 CalCOFI station locations by year (cont'd).
Figure 7.1 CalCOFI station locations by year (cont’d).
Figure 7.1 CalCOFI station locations by year (cont’d).
CalCOFI station locations

Figure 7.1 CalCOFI station locations by year (cont'd).
CalCOFI station locations

Figure 7.1 CalCOFI station locations by year (cont’d).
CalCOFI station locations

Figure 7.1 CalCOFI station locations by year (cont’d).
The data were obtained from the CalCOFI data archives located at ftp://www-
mrlg.ucsd.edu/pub/IEH/. The coordinates are originally formatted as latitude and longitude. These were transformed to Universal Transverse Mercator (UTM) in order for the area to be defined as a region in two-dimensional cartesian coordinate space with units in kilometers. Sea surface temperatures are reported in degrees celcius.

In this analysis, anomalous or unexpected variability is sought to be identified. With this in mind, the known or expected trend is removed. More importantly, mesoscale trend removal is done in order to satisfy the mean part of the intrinsic stationarity assumptions. When plotted, the sea-surface temperatures show a strong seasonal trend. Figure 7.2 shows gridded SST means by month for the study region. The map levels are obtained by aggregating all SST's over the study period by spatial bin location and then smoothing the binned means with a spline. In addition to seasonal variation Figure 7.2 indicates a strong trend in northing or north-south direction, and lesser trend in easting or the east-west direction. Seasonal variation is also characteristic of the area. Figure 7.3 shows gridded SST variation by month for the study region. In order to accomplish trend removal, we estimated the following model

\[
Z(s_i, t_j) = \mu + \beta_1 x_i + \beta_2 x_i^2 + \\
\delta_1 \cos(d_j) + \delta_2 x_i \cos(d_j) + \delta_3 y_i \cos(d_j) + \delta_4 x_i y_i \cos(d_j) + \\
\gamma_1 \sin(d_j) + \gamma_2 x_i \sin(d_j) + \gamma_3 y_i \sin(d_j) + \gamma_4 x_i y_i \sin(d_j) + \\
\zeta_i s_{oi} + \eta(s_i, t_j).
\]  

(7.1)
where $s_i = (x_i, y_i)$, $d_j = \pi \left( \frac{2 \text{day}(j)}{365} - 1 \right)$, $j \in \{1, \ldots, 365\}$ and $soi_j$ = southern oscillation index for $j$th day. Such a class of spatiotemporal trend models is discussed in Dimitrakopoulos and Luo (1997). Spatiotemporal trend models of the form in (7.1) must have component functions that are linearly independent over any set of data in order for the corresponding spatiotemporal kriging system to have a unique solution. Furthermore, in order to assure coordinate invariance, the trend form must demonstrate tensorial invariance as defined in Dimitrakopoulos and Luo (1997). In this case, the mixed polynomial-Fourier trend in (7.1) is tensorial invariant since by definition all lower order terms less than the maximum spatial and temporal orders are present.

Analyses are based on the model residuals $\tilde{\eta}(s_i, t_j)$ estimated from (7.1) as a zero mean process defined herein as the observed SST-anomaly. A frequency histogram of $\tilde{\eta}(s_i, t_j)$ is shown in Figure 7.4. The observed empirical form of the distribution does not suggest deviations from normality. The effectiveness of gross deseasonalization and spatial detrending of the data is shown in Figs. 7.5 and 7.6 for easting and northing respectively. The deseasonalized grid plots indicate the model in (7.1) to be effective in removing seasonal effects and directional trend. The observed SST-anomaly for the aggregated region by year is shown in Figure 7.7. Evident is the gradual warming trend that began during the early 1970’s. Also apparent is cyclic temperature trend with an observed oscillation period of about 25 years. Two high-temperature peaks are seen at 1958-1959 and 1983. There appears to have been a gradual escalation and decline surrounding each of these two observed warming events.
local smoothed SST mean by month

Figure 7.2 Gridded SST mean by month.
local smoothed SST mean by month

Figure 7.2 Gridded SST mean by month (cont’d).
local smoothed SST variance by month

Figure 7.3 Gridded SST variance by month.
Figure 7.3 Gridded SST variance by month (cont'd).
Figure 7.4 Frequency histogram of observed SST anomaly for entire study.
**Figure 7.5** Grid plot by easting of trended and detrended SST anomaly.
seasonal SST trend by northing

![Graph showing seasonal SST trend by northing.]

deseasonalized SST trend by northing

![Graph showing deseasonalized SST trend by northing.]

**Figure 7.6** Grid plot by northing of trended and detrended SST anomaly.
Figure 7.7 Regional SST anomaly by year.
7.2 Covariogram Estimation

From Chapter 5, the $d$-indexed Fourier-Bessel series covariogram estimator for $Z(s,t)$ where $s \in \mathbb{R}^2$ and $t \in T$ is given by

$$c_{2,M,N}(r,\tau; \hat{\theta}) = \frac{4\pi}{r \tau_c} \sum_{m=1}^{M} \sum_{n=1}^{N} \left( \frac{J_0(i_m.h/r_c)}{J_1(i_m)} \right) \left( \frac{J_{-1/2}(j_n.\tau/r_c)}{J_{1/2}(j_n)} \right) C(i_m, j_n; \hat{\theta}). \quad (7.2)$$

Where $C(i_m, j_n; \hat{\theta})$ is the estimated bivariate Johnson $S_U$ distribution described in Chapter 5. Use of the experimental Fourier-Bessel series covariogram estimator specifically written as (7.2) requires the assumption of spatial isotropy, an attribute of the true unknown random field. Should sufficient evidence of anisotropy be present, the estimator is easily rewritten to accommodate directional invariance in the same manner as other common parametric forms described in Cressie (1993). Although tests for isotropy are unknown in the literature, the most common method of verification is graphically with directional covariograms. In the case of space and time indexed observations, binning or regular sampling through time allows the practitioner to explore spatial isotropy through time. Figure 7.7 illustrates north-south and east-west binned empirical covariograms by year. The plots offer no reason to use an anisotropic form of (7.2). The isotropic covariograms indexed by year are shown in Figure 7.8.

The experimental covariogram model is fit to an empirical binned covariogram estimator. Spatial bins are defined every 10 km from 0 to 400 km. The mean distances for each bin are used as the computed spatial ordinate values. The temporal ordinate values are mid-year or 0.5 for example for the first year of the study, 1.5 for the second
year etc. The space-isotropic time-stationary spatiotemporal empirical covariogram estimator has form given by,

$$
\hat{c}_z(r, \tau) = \frac{1}{N(r, \tau)} \sum_{N(h, r)} \tilde{\eta}(s_i, t_i) \tilde{\eta}(s_j, t_j),
$$

where

$$
N(r, \tau) = \{(s_i, s_j) : \|s_i - s_j\| = r, t_i - t_j = \tau, t_i - t_j \geq 0; i, j = 1, ..., n\},
$$

and \( \tilde{\eta}(s_i, t_i) \) is the residual obtained by estimation of the model in (7.1). \( r \) and \( \tau \) above are defined as space and time increments. Our \( r \) and \( \tau \) are binned such that \( r \in B(r(l)) \) where \( l = 1, ..., H \) and \( \tau \in B(\tau(k)) \) where \( k = 1, ..., T \) where \( H = 80 \) and \( T = 25 \) as discussed above. The bin counts for the spatiotemporal empirical covariogram are all greater than 30,000. The empirical spatiotemporal covariogram computed from (7.3) is shown in Figure 7.8. By assuming the true process \( \eta(s_i, t_i) \) to be zero mean, the empirical covariogram in (7.3) is unbiased.

For the spatiotemporal Fourier-Bessel series covariogram estimator, the computational time for estimation of the parameters is influenced considerably by the starting values. In order to derive reasonable starting values, the estimates were first obtained using a smaller truncated series of \( M = N = 20 \). These estimates were obtained with minimal computational effort and then used as starting values for the series truncated at \( M = N = 100 \). The Nelder-Mead simplex algorithm was again used to estimate the parameters based upon an unweighted mean squared error loss function. Weighted loss functions generally resulted in poorer estimates. Parameters estimated by the Nelder-Mead simplex algorithm are given in Table 7.1.
Table 7.1  Final parameter estimates for experimental spatiotemporal covariogram (CalCOFI study).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>143.25</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.90</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>32.65</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>58.52</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>134.08</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>229.83</td>
</tr>
<tr>
<td>$r_c$</td>
<td>131.11</td>
</tr>
<tr>
<td>$t_c$</td>
<td>242.70</td>
</tr>
<tr>
<td>MSE</td>
<td>0.01687</td>
</tr>
</tbody>
</table>
Figure 7.7  Directional empirical spatial covariograms indexed by year.
Figure 7.7 Directional empirical spatial covariograms indexed by year (cont’d).
Figure 7.7  Directional empirical spatial covariograms indexed by year (cont’d).
Figure 7.8 Empirical spatiotemporal covariogram.
Figure 7.9 Fitted experimental spatiotemporal covariogram.
7.3 Spatiotemporal Prediction

Construction of countoured maps of SST anomaly are generated by spatiotemporal kriging as an extension of the spatial case. This involves estimation of the underlying spatiotemporal process at unmeasured locations and times. Of equal importance to process estimation is the mapping of estimated prediction errors such that the accuracy of the model may be locally assessed. Various prediction methods are described in Cressie (1993). As in the spatial case, we assume a zero mean process and formally utilize ordinary kriging in an extended spatiotemporal context.

The spatiotemporal kriging system is similar to that of the purely spatial case. We again assume a zero-mean process of the modeled random variable $\tilde{\eta}(s_i, t_j)$. We seek to optimize the predictions based upon the squared-error loss function,

$$L(\eta(s_0, t_0), p(U; s_0, t_0)) = (\eta(s_0, t_0) - p(U; s_0, t_0))^2$$

(7.4)

where $\eta(s_0, t_0)$ is the prediction at location $s_0$ and time $t_0$, $U = \{\tilde{\eta}(s_1, t_1), \ldots, \tilde{\eta}(s_n, t_n)\}$ and $p(U; s_0, t_0)$ is the predictor function. The best linear predictor is given by

$$p(U; s_0, t_0) = \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{l}(s_i, t_j) + k$$

(7.5)

such that $E(\eta(s_0) - p(U; s_0, t_0))^2$ is minimized. Such an optimal predictor $p^*$ is given by

$$p^*(U; s_0, t_0) = c^T \Sigma^{-1} U$$

(7.6)
where $\mathbf{c} = \left( c(s_0, t_0; s_1, t_1), \ldots, c(s_0, t_0; s_n, t_n) \right)^T$, $\Sigma$ is an $n \times n$ matrix whose $(i, j)$th element is $c(s_i, t_i; s_j, t_j)$ or the spatial covariance or covariogram between observations $\tilde{\eta}(s_i, t_i)$ and $\tilde{\eta}(s_j, t_j)$, and $n$ is the total number of observations. The minimized mean-squared prediction error is given by

$$\sigma_p^2 = C(s_0, t_0; s_0, t_0) - \mathbf{c}^T \Sigma^{-1} \mathbf{c}.$$  

(7.7)

In practice, we do not use the entire complement of $n$ observations to construct the predictor but instead utilize a spatial and temporal local neighborhood of $n' < n$ observations. The kriging neighborhood is defined in order to uniformly utilize contributions by spatial and temporal observations nearby the prediction locations. Each neighborhood is defined to have the same number of members as follows: for each of the first three time lags with reference to the prediction time, e.g. lag 0 through lag 2, the $k$th nearest neighbors are chosen where $k = 5$ for a total of $3 \times 5 = 15$ neighbors. Observation suggested that increasing the neighborhood size beyond that defined did not significantly improve the estimates but did increase computational time considerably. Prediction maps are shown in Fig. 7.10 and prediction error maps are shown in Fig. 7.11.

Prediction locations are generated on a $50km \times 50km$ regular grid system for the $800km \times 800km$ square study region and each year beginning in 1953. The boundary or edge points are excluded. For each year, there are a total of $13 \times 13 = 169$ gridded prediction points. For the entire 45 year time period, there are a total of 7,605 spatiotemporal predictions.

Examination of the spatiotemporal prediction errors suggests they are underestimated in comparison to the purely spatial prediction errors generated for the
Galveston Bay contaminants. This is likely attributable to the form of the estimated spatiotemporal covariogram characterized by a relatively small local dependence or slope in both space and time. The issue of underestimation of prediction errors in the classical kriging sense is discussed in Cressie (1993), where the problem lies within implementation of the estimated covariance as a parameter in computation of the estimated prediction errors. Therefore, for crossvalidation of the predicted SST's, we propose the following modification to the crossvalidation measures given by (6.7) and (6.8);

\[ CV_A(U) = (1/n) \sum_{i=1}^{n} \sum_{j=1}^{n_r} \left\{ \left( \hat{\eta}_{-(i,j)}(s_i, t_j) - \bar{\eta}(s_i, t_j) \right) / \hat{\sigma}_{-(i,j)}(s_i, t_j) \right\} \]  \hspace{1cm} (7.8)

and

\[ CV_B(U) = \left[ (1/n) \sum_{i=1}^{n} \sum_{j=1}^{n_r} \left\{ \left( \hat{\eta}_{-(i,j)}(s_i, t_j) - \bar{\eta}(s_i, t_j) \right) / \hat{\sigma}_{-(i,j)}(s_i, t_j) \right\}^2 \right]^{1/2} \] \hspace{1cm} (7.9)

where \( \hat{\eta}(s_i, t_j) \) is the predicted contaminant level using the abbreviated observation vector \( U = \{ \hat{\eta}(s_i, t_1), ..., \hat{\eta}(s_{i-1}, t_{j-1}), \hat{\eta}(s_{i+1}, t_{j+1}), ..., \hat{\eta}(s_{n_r}, t_{n_r}) \} \) and \( \hat{\sigma}_{-(i,j)}^2(s_i) \) is the pooled maximum likelihood estimator of the variance of \( \hat{\eta}_{-(i,j)}(s_i, t_j) \) and \( n \) is the total number of observations. The cross-validation measure in (7.8) should be approximately 0 and the cross-validation measure in (7.9) should be approximately 1 for an adequate predictive model. Cross-validation scores for the experimental spatiotemporal covariogram are \(-0.015386\) for \( CV_A(U) \) and \(0.97016\) for \( CV_B(U) \). The crossvalidation results suggest acceptable performance of the experimental covariogram.
estimator. However, they should be interpreted with prudence as suggested by Goovaerts (1997).

7.4 Interpretation

The study region is characterized by a complex interaction of events that occur on various time scales. Year to year changes in mesoscale climate variability serve to influence the local physical oceanography of the region. Although the study region is small compared to host domain of the North Pacific Ocean, a study of the comparative microscale changes will help to understand long-term and Pacific-wide changes in hydrographic and atmospheric responses.

The region is characterized by two current systems, the California Current and the North Equatorial Countercurrent. The California current is the easterly component of the clockwise or anticyclonic North Pacific Gyre system. At its northernmost range, the North Pacific Gyre splits to form the Alaskan Gyre to the north and southerly flowing California Current. The North Equatorial Countercurrent flows north of the equator to Panama and the northward along the North American continent. A major influence of temperature regimes in the study region is directly related to the relative dominance of either of these two current systems. During some years, the warm North Equatorial Countercurrent forces the cooler California Current atypically westward with a net effect of warmer SST’s in the region. When the California Current predominates the region, cooler net SST’s are observed. The interaction between the two current systems also typically forms a counterclockwise eddy field south of the Channel Islands.
The region is complicated by almost regular southerly atmospheric forcing making conditions favorable for upwelling. When sustained southerly winds force surface water south, the effect is a subsurface Ekman displacement of water mass to the west that results in upwelling of cooler deep water. Depending upon the predominant current system, variable surface mixing between cool and warm water may be observed.

As is common in scientific endeavors, the prediction plots raise new questions regarding the dynamics of the region. The prediction plots do not suggest a clear warming pattern indicated by the mean temperature field in Figure 7.7. However, the years 1953 through 1969 show more regional variation in SST than the years 1970 through 1998 or the warming period indicated in Figure 7.7 and confirmed by CalCOFI investigators. For the cooler years, patches of warm and cool water seen in the prediction plots, Figure 7.10. The source of local SST variability during these years is uncertain. Strengthened upwelling may have been present resulting in an increase in surface mixing. Sustained wind data may help support this conjecture. Another possible cause of patchy SST contours during the earlier years may be attributed to the level of mixing in the immediate region between the two current systems. The boundary between the two currents is not sharply defined and is characterized by small scale eddies that form and maintain themselves within the foreign water mass for a short time period until mixed or dissipated. These eddies are common on all scales and may be warm or cold. A combination of upwelling and current boundary location may well have been observed during the earlier cool years as well.

The years from 1970 through 1998 are characterized by a generally homogenous temperature field. Although seasonal variations in the influence of either current system
can be seen in the seasonal data, the overall year to year effect of mixing by either current boundary, upwelling or both is suggested to be negligible. The study region in more recent years has in fact been dominated by the Equatorial Countercurrent forcing the California Current offshore (Hayward et al., 1996). However, we suggest that although displacement of the California Current certainly contributes to warming in the region, perhaps general mixing due to the added effect of upwelling or boundary eddies also plays a significant role in determining the SST character of the area.

Further analyses should include covariates, in particular atmospheric forcing or sustained wind stress is significant. An equally important covariate is the upwelling index, itself computed from sustained wind stress. In addition, the differentiation between the two current systems becomes more evident with salinity measures, e.g. the source of a water parcel is identifiable by temperature and salinity. This information may help in determining the influence of current boundary location on the temperature regime of the region.
Figure 7.10 Predicted SST anomaly by year.
Figure 7.10 Predicted SST anomaly by year (cont'd).
Figure 7.10 Predicted SST anomaly by year (cont'd).
Figure 7.10 Predicted SST anomaly by year (cont’d).
Figure 7.10 Predicted SST anomaly by year (cont’d).
Figure 7.10 Predicted SST anomaly by year (cont’d).
Figure 7.10 Predicted SST anomaly by year (cont’d).
Figure 7.10 Predicted SST anomaly by year (cont’d).
Figure 7.11  SST anomaly prediction error by year.
Figure 7.11 SST anomaly prediction error by year (cont'd).
Figure 7.11 SST anomaly prediction error by year (cont’d).
Figure 7.11 SST anomaly prediction error by year (cont'd).
Figure 7.11 SST anomaly prediction error by year (cont'd).
Figure 7.11 SST anomaly prediction error by year (cont’d).
Figure 7.11 SST anomaly prediction error by year (cont’d).
Figure 7.11 SST anomaly prediction error by year (cont’d).
8. CONCLUSIONS

We have proposed and evaluated an experimental covariance series estimator that does not require separability. The ideas presented here are based primarily on the premise that estimation of the spectrum is more general than estimation of the covariance itself. We suggest that the modality of the spectrum characterizes the covariance structure. For example, a single mode spectrum, as is most common, can characterize a number of covariance shapes. For this reason, we begin by conjecturing a general spectral shape and optimize its fit to the empirical covariance.

The proposed method assumes that a parametric function constrained to positivity is easier to estimate than one constrained to be positive definite. Given the heuristic definition of positive definiteness, this is inarguably true. Bochner’s theorem significantly eases validation of covariogram models. For the experimental Fourier-Bessel series covariogram estimator, Bochner’s theorem was actually applied in the development. Although the experimental estimator is not positive definite with probability one, we found no instance of negative prediction variances throughout the simulations and analyses. Other models in the literature, particularly spatiotemporal dependence models, are known to be inadmissible as valid covariances whereas our experimental model has the advantage of having been estimated from the proper parameter space in the limit.

In a timely paper, Kyriakidis and Journel (1999) describe shortcomings of separable models. Given the separable decompositions

\[ c(r, \tau) = c_1(r) + c_2(\tau) \]  

(8.1)

and
\[ c(r, \tau) = c_1(r)c_2(\tau). \] (8.2)

although mathematically agreeable, weaknesses to representations (8.1) and/or (8.2) are readily concluded. The zonal anisotropy implied by (8.1) indicates the random field \( Z(s, t) \) is the same at all time instants. Also, changes in the temporal pattern of \( Z(s, t) \) from location to location cannot be accommodated with (8.1). Models (8.1) nor (8.2) account for space and time interaction. Such models arise from processes acting independently in space and time (Jones and Zhang, 1997). Furthermore, no guideline exists for inferring (separating) the two component structures \( c_1(r) \) and \( c_2(\tau) \) from the estimable covariances \( c(r,0) \) and \( c(0,\tau) \) as discussed in Kyriakidis and Journel (1999).

As noted by Myers and Journel (1990) and discussed above, zonal anisotropies can lead to invalid kriging systems for certain data configurations. Specifically, covariance models built from a sum of one-dimensional structures may only be conditionally positive definite in higher dimensional spaces. Rouhani and Myers (1990) and Myers (personal communication) confirmed a number of problems associated with zonal anisotropy when performing kriging, one example being that of singularity of the kriging system. Similar problems arose when samples are taken at the same or approximately same locations and close time intervals. Indeed, Dimitrakopoulos and Luo (1994) showed that the problem is eliminated if the positive definite covariance function is replaced by a strictly positive covariance function as we have done here with the positive-scaled Johnson \( S_v \) representation of the spectrum.

An important limitation of the estimator may lie in the form of the spectrum.

Although the scaled Johnson \( S_v \) distribution performed well for our simulations and
applications, circumstances may arise such that an alternate spectral form is more appropriate. A number of alternatives may be suggested, including models of the type described by Ratkowsky, e.g.

$$C(\omega; \theta) = \exp\left( f(\omega; \theta) \right)$$

(8.3)

where $f(\omega; \theta)$ is a rational function. Experimentation with these model types showed that models of the type defined in (8.3) are for the most part restrictive in behavior. However, the class of exponentiated rational functions itself represents a variety of spectral shapes including those represented by exponentiated polynomials.

The selection of estimation starting values is particularly important with respect to the experimental covariogram estimator. The estimation stage is computationally intensive compared to standard covariogram models. The reason is attributed to the series form of the estimator combined with the computational overhead of numerical Bessel function evaluation. The estimation method that provided timely estimates, particularly for the spatiotemporal iterated series representation was to begin with low series orders and use estimates obtained from those as starting values for high series orders. Specifically, we first estimated with the spatial order $M = 10$ or spatiotemporal orders $M = N = 10$ and used those computed estimates as starting values for $M = 100$ and $M = N = 100$ for the spatial and spatiotemporal cases respectively.

The precision of our experimental covariogram as a parameter estimate merits some discussion. Theorem 3.3 for the one-variable case and Theorem 5.1 for the two-variable case state convergence properties for the infinite sum. For the finite sum one-variable case, we consider Corollary 3.4.1 as it directly applies to the purely spatial model and derive an expected level of accuracy based upon the order of the summation.
We expect the same to follow for the two-variable case as well although the result is not derived here. One must be reminded that Theorem 3.3 applies to a known function whereas in this case, we are estimating an unknown function, the covariogram. We therefore might expect a stochastic representation of Theorem 3.3 to follow. In practice, numerical accuracy is achieved by incrementally increasing the order of the summation, using the estimates generated from smaller order summations as starting values for higher order summations. Convergence is then attained based upon standard numerical criteria, e.g. relative error. Through this sequential iteration estimation procedure, we experienced progressively better estimated covariograms with increased order of the summation. Based upon observed improvement in fit, we recommend covariogram estimation on the order of $M \geq 100$. Computational requirements escalate for higher order representations due primarily to computation of the Bessel series present in the estimators. This is particularly the case for spatiotemporal covariogram estimation. For this reason, we recommend our sequential method of estimation described above, beginning with perhaps $M = 10$.

The experimental covariogram estimator proposed here is appealing in that a smooth representation of the covariogram is produced as opposed to nonparametric methods of Hall, Fisher and Hoffmann (1994) and Shapiro and Botha (1991) where irregular covariance models are derived. In most cases, there is no physical evidence to support locally "bumpy" covariance representations. With our simulations, the underlying covariance is monotonically non-increasing as a function of distance, but existing nonparametric methods are not guaranteed to provide such representations.
The most common covariogram estimation approach is the OLS loss function, not taking into account dependence in space or space-time of covariogram elements. Weighted Least Squares or WLS as described by Cressie (1985) was tried for the simulations. We found OLS or unweighted least squares to perform better with respect to observed covariogram model residuals.
APPENDIX

A. SAS Code - Spatial Covariogram Estimation

/* ******************************************************* */
/* Fourier-Bessel Series Covariogram Estimation */
/* Spatial Case */
/* Author: L. Scott Baggett */
/* Approach to Spatial and Spatiotemporal */
/* Dependence Modeling", Ph.D. Thesis, */
/* Rice University, Houston, Texas. */
/* */
/* Note: M=100 */
/* */
/* Input File Notes: */
/* */
/* besselj0.dat - text dataset */
/* j0 zeros of Bessel function of first */
/* kind of order zero */
/* */
/* stor.cov - SAS dataset */
/* */
/* h distance */
/* cov empirical covariogram */
/* count bin count */
/* */
/* Output File Notes: */
/* */
/* stor.est - SAS dataset containing parameter */
/* estimates */
/* ******************************************************* */

options linesize=120;

filename file1 'f:\enap\data\besselj0.dat';
libname stor 'f:\enap\data';

title ' ';
data data1;
set stor.cov;
data data2;
infile file1;
input j0;
if _n_ le 100;

proc iml;
start INIT;
use data1;
read all var {h} into R;
read all var {cov} into C;

use data2;
read all var {j0} into j0;
Pi=3.14159265359;

n=nrow(C);
m=nrow(j0);

bessjlsq=j(m,1.0);
do i=1 to m;
bessjlsq[i]=(jbessel(1,j0[i])**2.);
end;
finish INIT;

start COV2D(rx.a) global(j0,bessjlsq.Pi,m);
  u = 0;
  if rx <= a[4] then do;
    t=j(m,1,0);
    g=j(m,1,0);
    do i=1 to m;
      e1 = a[3];
      e2 = (1/Sqrt((j0[i]/a[2])**2+1));
      e3 = Exp(-1/2)*(a[1]**(Log((j0[i]/a[2]) + Sqrt((j0[i]/a[2])**2+1))))**2);
      g[i] = e1*e2*e3;
      t[i]=jbesse1(0,j0[i]*rx/a[4])/bessjlsq[i];
    end;
    k=(4*pi)/a[4];
    u=k*t^'g';
  end;
return(u);
finish COV2D;

start OLS(a) global(R.C.n);
  Chat=j(n,1,0);
  do i=1 to n;
    Chat[i]=COV2D(R[i].a);
  end;
  u=(C-Chat)**(C-Chat);
  u=u/n;
return(u);
finish OLS;

run INIT;

optn = { . 3};
a = { 1 -10 0 40000};
con = { . . 1E-20 1E-20, . . . . };
CALL NLPMNS(rc,psol,"OLS",a,optn,con);

outrec=J(1,5,0);
outrec[1,1]=psol[1];
outrec[1,2]=psol[2];
outrec[1,3]=psol[3];
outrec[1,4]=psol[4];
outrec[1,5]=OLS(psol);
create cvec from outrec [colname={a b c rc mse}];
append from outrec;

outobs=J(n,2,0);
do i=1 to n;
  outobs[i,1]=R[i];
  outobs[i,2]=C[i];
end;

create cvec from outobs [colname={r c}];
append from outobs;

outpred=J(40,2,0);
k=0;
do i=1 to 40;
  outpred[i,1]=k;
  outpred[i,2]=COV2D(k,psol);
  k=k+1000;
end;

create bvec from outpred [colname={r c2d}];
append from outpred;
quit;
data stor.est;
set evec;
run;
quit;
B. SAS Code - Spatiotemporal Covariogram Estimation

`/* Fourier-Bessel Series Covariogram Estimation */
/* Spatial Case */
/* Author: L. Scott Baggett */
/* Approach to Spatial and Spatiotemporal */
/* Dependence Modeling”, Ph.D. Thesis. */
/* Rice University, Houston, Texas. */
/* Note: M=100 */
/* */
/* Input File Notes. */
/* */
/* j0d1.dat - text dataset */
/* j1l zeros of Bessel function of first */
/* kind of order zero */
/* */
/* j0d2.dat - text dataset */
/* j2l zeros of Bessel function of first */
/* kind of order one */
/* */
/* dd2.dat - text dataset */
/* */
/* i spatial index */
/* j temporal index */
/* dh spatial bin distance */
/* dt temporal bin distance */
/* n bin count */
/* c empirical covariogram */
/* dbar mean spatial bin distance */
/* */
/* Output File Notes: */
/* */
/* stor.pred100 - SAS dataset containing predictions */
/* */
/* stor.parm100 - SAS dataset containing parameter */
/* estimates */
`
start INIT;

use data1;
read all var {r} into R;
read all var {t} into T;
read all var {c} into C;

use data2;
read all var {jd2} into jd2;

use data3;
read all var {jd1} into jd1;

Pi=3.14159265359;

n=nrow(c);
md1=nrow(jd1);
md2=nrow(jd2);

dmd1=j(md1,1,0);
do i=1 to md1;
   dmd1[i]=(j*bessel(0.5,jd1[i]**2.));
end;

dmd2=j(md2,1,0);
do i=1 to md2;
   dmd2[i]=(j*bessel(1.0,jd2[i]**2.));
end;

finish INIT;

start ARCSINH(s);

u = log(s=sqrt(1-s**2));
return(u);

finish ARCSINH;

start COV2DT(rx,rt,a) global(jd1,jd2,dmd1,dmd2,md1,md2,Pi);

u = 0;
if rx <= a[8] & rt <= a[7] then do:
   td1=j(md1,1,0);
   td2=j(md2,1,0);
   g=j(md1,md2,0);

do i=1 to md1;
   td1[i]=((-sqrt(2/Pi)*cos(jd1[i]*rt/a[7])/sqrt(jd1[i]*rt/a[7]))/dmd1[i]);
end;
do i=1 to md2;
   td2[i]=j*bessel(0,jd2[i]*rx/a[8])/dmd2[i];
end;
do i=1 to md1;
   do j=1 to md2:
      e0 = -1/2*(1-a[2]**2));
      e1 = (a[3]*ARCSINH(jd1[i]/a[5])))**2;
      e2 = -2*a[2]*a[3]*a[4]*ARCSINH(jd1[i]/a[5])*ARCSINH(jd2[j]/a[6]));
      e3 = (a[4]*ARCSINH(jd2[j]/a[6])))**2;
      d1 = sqrt(1-a[2]**2)*sqrt(1-(jd1[i]/a[5])**2)*sqrt(1-(jd2[j]/a[6])**2);
      g[i,j]=j[a[1]/d1]*exp(e0*(e1-e2-e3));
      end;
   end;
k=(4*Pi)/(a[7]*a[8]);
   u=k* td1''*g*td2;
end;
return(u);
finish COV2DT;

start OLS(a) global(R,T,C,n);
    Chat=j(n,1,0);
    do i=1 to n;
        Chat[i]=COV2DT(R[i],T[i],a);
    end;
    u=(C(Chat)''(C-Chat))%^1/2;
    u=u/n;
    return(u);
finish OLS;

run INIT;

optn = ( . 1 );
a = (142.870596  -0.904591  32.588594  58.446311  113.846217  
   229.454810  110.734747  243.572215);
con = ( 1E-20  0.99999999  . . . . 1E-20  1E-20,
   .    0.99999999  . . . . . . . . . .);

CALL NLPSOL(rc,psol,"OLS",a,optn,con);

outpred=J(n,4,0);
    do i=1 to n;
        outpred[i,1]=R[i];
        outpred[i,2]=T[i];
        outpred[i,3]=C[i];
        outpred[i,4]=COV2DT(R[i],T[i],psol);
    end;
create cvec from outpred [colname=(r t c chat)];
append from outpred;

outparm=J(8,1,0);
outparm[1]=psol[1];
outparm[2]=psol[2];
outparm[3]=psol[3];
outparm[4]=psol[4];
outparm[5]=psol[5];
outparm[6]=psol[6];
outparm[7]=psol[7];
outparm[8]=psol[8];
create dvec from outparm [colname=(alpha rho celtal delta2 lambda1 lambda2 rt rc)];
append from outparm;

quit;
data stor.pred100;
    set cvec;

data stor.parm100;
    set dvec;

run;
quit;
REFERENCES


