Variable Modeling of Spatially Distributed Random Interval Observations

Njeri Wabiri

Thesis Presented for the Degree of

Doctor of Philosophy
In Statistical Sciences
Science Faculty, University of Cape Town

Advisors: A/Prof. Christien Thiart
Prof. Renkuan Guo

2007

Variable Modeling of Spatially Distributed Random Interval Observations

Abstract
by
Njeri Wabiri

Often, data is not observed exactly, but rather an interval range or fuzzy measure of data is observed. In the spatial domain, observed intervals in the same neighborhood influence each other. The assimilation of imprecise data in spatial models is not only appealing but is also becoming an important research topic. However, the development of models, for spatial imprecise (fuzzy) data is still an emerging field in environmental applications. As a part of this emerging field, this thesis addresses the dual of randomness (imprecise observing of data as an interval) and the spatial dependence by (i) extending kriging to handle intervals, (ii) a non-parametric approach using kernels, (iii) Bi-variable copula kriging approach that encompass the modeling of uncertainty (dependence) between the observed lower and upper interval values and (iv) a fuzzy -credibility kriging approach that utilizes a fuzzy variable, the equivalent of observed interval data.

First, we use the support function concept for embedding interval-valued data into Hilbert space. This mirrors the pair wise dissimilarities among the intervals, and generates a univariate configuration of random intervals, interpreted as 2D location in the domain of a suitable Hilbert space. The approach opens an interface between interval-valued data analysis and support function-based data analysis. Each random interval corresponds to a random support function, and the interval space assumes a coordinate system with intervals as bivariable random vectors of lower-upper random variables. A generalized weighted interval metric and corresponding interval kernel is defined.

A kriging approach to model spatial random intervals is proposed. The approach focuses on stationary isotropic random interval models and utilizes the center-radius variable representation of random intervals. Two techniques:- a) composite kriging and b) component-wise kriging are defined based on proposed composite and component-wise (co)variance models.

The proposed non-parametric approach relaxes the stationarity assumptions in kriging, focusing on non-stationary random interval models. The approach allows the observed random spatial-interval data to define the optimal functional form of the spatial random interval process. The problem of providing spatial interval-valued estimates of the spatial interval regression function based on some kernel function defined in a Hilbert space is

addressed via a Kernel interval regression technique in a joint spatial interval domain. Our approach combines a spatial adaptive kernel and a robust interval uncertainty kernel.

The spatial kernel ensures that only samples sharing similar intensity and gradient information are gathered for local approximation. The robust uncertainty kernel adapts to the interval structure and minimizes the influence of outliers. This kernel-based technique is found to be efficient when data is mapped to a high dimensional feature space where algorithms as simple as their linear counterparts in input space are used.

As an alternative to indicator kriging for spatial random interval/vector data, we propose two scalar variable modeling techniques. In additional to spatial interval prediction, a unique contribution to the literature of the spatial data analysis and GIS modeling is made. The two techniques a) bivariable copula grade kriging, and b) credibility distribution grade kriging of spatially distributed random intervals exploit the joint distribution of the spatial random intervals. The copula approach combines analytical distribution (marginal) models and an empirical copula. The credibility approach is based a scalar fuzzy variable, the equivalent of the random interval set, with maximum entropy. This leads to a tractable analytical model for the uncertain phenomena allowing for easy prediction and mapping.

DEDICATION

I dedicate this dissertation to my parents, who taught me that there is no such thing as finished dream. Mom and Dad, I am prepared for new moves, bigger waves and longer rides to achieve new dreams. God bless you.

ACKNOWLEDGMENTS

I am certainly most indebted to my advisors, Associate Professor Christien Thiart, for her continued support and the many helpful discussions throughout my studies; Professor Renkuan Guo for the insight and encouragement to take a road that I dared not to. With your positive attitude and patience, you both created a very pleasant atmosphere and I would like to thank you for the many things I was able to learn from you about mathematical statistics, geostatistics, stochastic processes, fuzzy logic and otherwise.

I am greatly indebted to many sponsors Third World Academy of Women Scientist (TWOWS), International Postgraduate funding office for the financial support without which I would not have managed to support my Self.

Deep thanks goes to Professor Tim Dunne, Head of Department of Statistical Sciences for ensuring I had all the facilities I needed to do this research.

Deep gratitute to the almighty God for the strength and wisdom through this journey.

CONTENTS

FIGURES

TABLES

SYMBOLS

Symbol	Definition
$\mathbb{R}$	set of real numbers
$\mathbb{R}^n$	vector space of n-tuples of real numbers or
$\mathbb{C}^n$	vector space of n-tuples of complex numbers
$\mathbb{R}^+$	set of real positive numbers
$\Re^+$	set of extended positive real space $\mathbb{R}^+ \cup \infty = [0, \infty)$
$\overline{\mathbb{R}}$	extended real line $[-\infty, +\infty]$
$\mathrm{CI}(\mathbb{R})$	Closed interval number $\boldsymbol{a} = [a^l, a^u] = \{\boldsymbol{a} \mid a^l < a < a^u,$ $a^l, a^u \in \mathbb{R}$
$\mathbb{IR}$	Set of proper intervals $\boldsymbol{a} = [a^l, a^u], \quad a^l \leq a^u$
$\overline{\mathbb{IR}}$	Set of improper intervals $\boldsymbol{a} = [a^l, a^u], \quad a^l \geq a^u$
$\mathbb{I}^*\mathbb{R}$	Extended interval set defined as $\mathbb{IR} \cup \overline{\mathbb{IR}}$
$\mathcal{K}(\mathbb{R})$	Set of nonempty closed intervals
$\mathcal{K}_c(\mathbb{R})$	Set of nonempty compact intervals
$\mathcal{K}_{cx}(\mathbb{R})$	Set of nonempty convex compact intervals
$f : X \to Y$	mapping from space X to space Y: f is continuous if and only if (iff), for every open set $V \in Y$, $f^{-1}(V)$ is open
$(\Omega, \mathcal{A}, P)$	probability measure space: Ω-possible outcomes, $\mathcal{A}$- σ-algebra(events/ subsets of Ω) and P- probability measure of events
l^2	space of square summable sequences
l^p	space of sequences summable with p-the power
$L^1(\Omega)$	space of integrable functions on Ω
$L^2(\Omega)$	space of square integrable functions on Ω
$L^p(\Omega)$	space of functions integrable with p-power on Ω
$\mathcal{H}$	Hilbert Space (Appendix: A.1.5)
$a, b, c, \ldots$	points of $\mathbb{R}^n$

Symbol	Definition
$A, B, C, \ldots$	subsets of $\mathbb{R}^n$
$A \cap B$	intersection of A and B subsets of $\mathbb{R}^n$
$A \cup B$	union of A and B subsets of $\mathbb{R}^n$
$A \wedge B$	minimum or intersection of events A and B
$A \vee B$	maximum or union of events A and B
$x \in X$	x is an element of the set X
$x \notin X$	x is not in X
$A \subset X$	A is a subset of X
$A \setminus X$	complement of A in X
$\overline{A}$	closure of a set A i.e. smallest closed set containing A
A^{int}	interior of a set A i.e. largest open set contained in A
$s_A(u)$	support function of a non-empty compact convex set A
$X \times Y$	Cartesian product of X and Y
$\forall x \in X$	for all x in X
$\exists x \in X$	there exists an x in X such that
D	fixed spatial region or a subset of $\mathbb{R}^d$
$Z(\boldsymbol{x})$ or $z(\boldsymbol{x})$	regionalized variable
$A \Rightarrow B$	A implies B
$A \Leftarrow B$	A is implied by B
$A \Longleftrightarrow B$	A is implied by B and B is implied by A
iff	if and only if
a.s. or a.e.	almost surely i.e. an event E occurs almost surely if $Pr\{E\} = 1$ or property holds almost everywhere if the set of points where it fails to hold is of measure zero.
$\cong$	equivalent
$\triangleq$	equal by definition
sup	suprema (supremum)
inf	infima (infinum)
$\|\|x\|\|_2$	norm of x or Euclidean vector
$\mathrm{St}(A)$	Steiner point of a convex body
$b(m, r)$	ball centered at point m with a radius of r
§	denotes a section

ACRONYMS

Acronym	Definition
SRF	Stationary random process (or field, function)
SCRI	Spatial closed random interval
SIRF	Spatial Interval random function
OLS	Ordinary least squares
GLS	Generalized least squares
WLS	Weighted least squares
OK	Ordinary Kriging
OIK	Ordinary interval kriging
KDE	Kernel Density Estimation
CRIG	Classical Random Interval Geostatistics
FCG	Fuzzy Credibility-based Geostatistics
BCG	Bivariable Copula-based Geostatistics

CHAPTER 1

INTRODUCTION

Scope

The chapter introduces the problem addressed in this dissertation. A review of the developments of spatial prediction in particular application of geostatistics is given. Limits of the geostatistical approach to modeling spatially distribute random vectors are highlighted. This leads to the proposed research problem addressing specific objectives.

1.1 Introduction

In the early 1950s, mining engineers were faced with the problem of estimating disseminated ore reserves. Available classical methods were found to be insufficient. Krige (1951), a South African mining engineer and Sichel, a statistician designed alternative optimal interpolation methods for estimating the sparsely distributed ore reserves (Journel and Huijbregts, 1978). A French mathematician, Matheron (1963, 1975, 1989) developed Krige's concepts and formalized them within a single framework. Matheron introduced a tool to analyze the spatial continuity of sample values called *variogram*.

The concept of a variogram was referred to as a structural function in meteorology (Gandin, 1965) and in probability (Yaglom, 1987), a mean-squared difference in time series (Jowett, 1952) or an autocovariance function in the study of Brownian motion on $\mathbb{R}^n$ (Lévy, 1948). For other references see Cressie (1988). Kolmogorov (1941) introduced it to study the local structure of turbulence in fluid. Nevertheless, it has been Matheron's mining terminology that has persisted. Based on the variogram, Matheron introduced an estimation method and called it *kriging*. The phrase kriging was coined in recognition of Krige. Though originally developed for use in the mining industry, the methods eventually lead to the birth of a scientific field known as Geostatistics.

Within five decades, the field of geostatistics has expanded its horizon, both on theoretical aspects and practical applications. Beginning with the concepts of spatial correlation, variograms and the simple kriging formalism, it rapidly extended to take into account, e.g., constant unknown mean (ordinary kriging), non-stationary mean (kriging with a trend and/or a varying local mean (universal kriging and intrinsic random function of order k (IRF-k))), auxiliary variables (cokriging) and categorical variables (indicator kriging). Chiles and Delfiner (1999); Cressie (1993); Goovaerts (1997) provides details of the methods. The ability of kriging to quantify and model the spatial structure means that predictions are tailored to the intrinsic structure of the variable of interest and not only the sample numbers and sampling patterns.

Geostatistics has also been applied in wide range of applications among them: remote sensing (Atkinson and Curran, 1998; Curran et al., 1989; Woodcook et al., 1988a,b); soils science (Ver Hoef and Cressie, 1993; Webster, 1985; Webster and Oliver, 2001); hydro-geology (Kitanidis, 1997), public health (Carrat and Valleron, 1992); agriculture (especially, in precision farming) (Cahn et al., 1994); hydrology (Ver Hoef and Cressie, 1993) and petroleum geology (Hohn, 2000). In spite of its

adaptability to wide range of applications, kriging methodology still faces several shortcomings:

- Kriging is only a best linear unbiased estimator in sense of minimum squared error (Cressie, 1993; Stein, 1999); it is a common least squares solution to the common regression problem.
- Kriging is based on the wide spread assumption of stationarity, often modeled with a global spatial covariance model across the spatial region of interest. Intuitively, kriging is a parametric approach to spatial prediction.
- Kriging produces a map of estimates which is "best" (and hence unique map) in a statistical sense. It also produces an associated map of spatial uncertainty (Goovaerts, 1997). This implies that we cannot generate plausible realizations or maps to aid in better assessment of spatial uncertainty. The latter can be achieved by using simulation (Chiles and Delfiner, 1999; Goovaerts, 1997)
- Kriging, originally designed to handle scalar(real)-valued spatial measures cannot automatically handle fuzzy ("imprecise ") spatial data. For example for spatial random vectors (intervals) incorporating 'dual uncertain' information influenced by randomness and lack of knowledge (fuzziness), the approach breaks down. Often a simple naive approach is to use rough approximations of spatial random vectors/intervals i.e. the mean, median and indicator values. This implicitly leads to loss of information.

Among the widely used alternatives to geostatistical techniques are: regression (generalized linear models, generalized additive models (Hastie and Tibshirani, 1990), radial basis function interpolation (Webster and Oliver, 2001), global polynomial regression and inverse distance weighting methods (Myers, 1990)). Webster and Oliver (2001) notes that the methods do not take into account (local) spatial autocorrelation. In addition, the methods are optimal with even distribution of sample points in the study area.

Proposed solutions relate to combining kriging and other methods (e.g. Fuzzy sets (Bandemer and Gebhardt, 2000; Bárdossy et al., 1988, 1990b; Burrough, 1989; Kacewiez, 1994; Piotrowski et al., 1994); Regression kriging (Fuentes, 2001; Hengl et al., 2003; Kleinschmidt et al., 2000; Stacey et al., 2006; Thomson et al., 1998)). These approaches are often customized for specific applications. Other methods worthy mentioning include autoregressive spatial smoothing (Ying and Dean, 2001); Splines (Hutchinson, 1995; Wahba, 1992); Geographically weighted regression (GWR)

(Brunsdon et al., 1998a,b; Fotheringham et al., 2002) and Geostatistical simulation (Chiles and Delfiner, 1999; Christian, 2002; Goovaerts, 1997).

1.2 Research Problem

Often spatial data is not observed exactly but rather an interval (range) of the data is observed. The observed sample of regionalized interval variables incorporates dual uncertain information influenced by randomness and lack of knowledge (imprecision/fuzziness).

Major breakthroughs in information technology, means that large amounts of varied data are easily collected from almost everywhere on earth and at different scales. This varied nature of the data poses additional challenges to geostatistics. There is need to adapt or design spatial prediction methods that are able to combine varied data rigorously using sound theoretical concepts. Designed approaches should easily adapt to fuzziness/imprecision inherent in spatial data.

Except for Fuzzy set based methods and Geostatistical Simulation, the spatial prediction methods in §1.1 lack the capability to handle the dual uncertainty (randomness and imprecision (fuzziness)) in spatial data. The method treats the fuzziness as a random error.

Geostatistical simulation, alternatively called spatial stochastic simulation (Ripley, 1987b) has been used to deal with sample spatial uncertainty and is beyond the scope of the study. Briefly, geostatistical simulation deals with spatial uncertainty by generating several realizations of an assumed "spatial model" of phenomena under study. Estimate is then extracted from the realization. If realization honors the sample values (conditional simulation) the estimate is more accurate.

Developed Fuzzy-based methods (Bandemer and Gebhardt, 2000; Bárdossy et al., 1988, 1990a; Kacewiez, 1994) utilize Kaufmann (1975) fuzzy variable concept. Fuzzy geostatistical modeling in Kaufmanns sense, is still a set modeling based on fuzzy set variograms and leads to fuzzy sets predictions. New modeling techniques should ultimately adopt a "variable modeling approach" for fuzzy (imprecise) spatial data to ensure real-valued estimates rather than set-valued estimates.

In this study, the research problem is to handle spatial imprecise (interval-valued) data. We aim to address this by developing new innovative techniques for spatial estimation with interval data namely:-

→ Classical Random Interval Geostatistics (CRIG) methodology
- Composite and Component-wise kriging

- Spatial Non-parametric Kernel Interval Regression

→ Bi-variable Copula-based Geostatistics (BCG) methodology

→ Fuzzy Credibility-based Geostatistics (FCG) methodology

The Classical Random Interval Geostatistics exploits probability theory for modeling randomness and random interval theory for modeling fuzziness/imprecision. Kriging and spatial non-parametric kernel interval approaches are proposed. The two approaches, applies linear algorithms to modeling spatial random intervals based on the lower/upper or center/radius coordinate variables. The approaches lead to interval outputs represented using the lower and the upper prediction map plus a prediction error map.

A new approach that introduces 'variable modeling idea' to modeling spatial random intervals fields is proposed. The approach constitutes the Bi-variable Copula Geostatistics (BCG) (Chapter 5) and Fuzzy Credibility-based Geostatistics (FCG) (Chapter 6) methodologies. The two variable-based approaches are based on the joint distribution function of the random interval data. They are information-intensive and benefits considerably from the advantages of probability density (pdf)-based inference. The BCG approach combines analytical (marginal) models and an empirical copula in defining the marginal and copula grade marks associated with random interval/vectors at given spatial locations. The FCG approach is based on a flexible and tractable analytical model, the credibility distribution model of a fuzzy variable induced by a random interval with maximum entropy. Compared to BCG, the FCG approach is a more direct approach that simplifies numerical calculations and can be easily applied to a general case.

The BCG and FCG approaches offer flexible output in form of joint probability prediction, from which further elaborate statistics can be derived. Output is easily integrated in available commercial GIS software (e.g. ArcGIS).

By associating the observed spatial random intervals with unique descriptive measures: the copula, marginal and credibility grade marks, the BCG and FCG methods leads to a random vector marked point process. The "points" are the random vector locations and the "marks" are copula, marginal and credibility grades. Integrating statistical copula based methods with the approach to analysis of a marked point process can provide valuable information on the spatial interaction of spatially distributed random interval observations. The same principle applies if we integrate fuzzy variable modeling approaches to marked point process. In both cases, estimation can be easily done via maximum likelihood estimation

Ultimately, the question addressed in this study is whether a scalar-variable modeling idea is applicable to analysis of spatially distributed random interval or fuzzy observations. The answer is "yes", and it is the objective of this study to demonstrate this using the proposed BCG (Chapter 5) and FCG (Chapter 6) methodologies. The study also aims at demonstrating the robustness and the flexibility of the approaches and presenting them in a way acceptable to most users. Section 1.3 provides a summary of the specific research objectives.

1.3 Research Objectives

(i) To extend real-valued parametric stationary models to modeling spatially distributed interval-valued data;

(ii) To extend nonparametric methods based on kernels to facilitate analysis on non-stationary spatial random interval process;

(iii) To introduce the bi-variable copula approach as a first attempt to scalar variable modeling of spatially distributed random vectors;

(iv) To introduce credibility distribution theory, utilizing a maximum entropy approach, for seeking variable modeling approach to random interval events;

(v) To illustrate and test the flexibility of the credibility distribution and copula based method using real-life application.

1.4 Thesis Outline

This dissertation consists of three theoretical sections and an application part. The theoretical part covers the fundamental concepts underlying the proposed techniques (Figure 1.1).

Part A, Figure 1.1, constitute Chapters 2, 3 and 4 of the thesis and is an extension of classical kriging and the non-parametric kernel methods to handle spatial imprecise (interval) data.

Chapter 2 reviews classical geostatistics and random closed interval theory, covering the basic concepts underlying the two theories and defining the classical random interval geostatistics process.

Chapter 3 extends the classical kriging approach to regionalized random intervals. It address objective (i) (Wabiri et al., 2005, attached as Appendix P.2).

Chapter 4 discusses the flaws with parametric interval kriging and highlights how the non-parametric kernel interval regression may provide a much more flexible way that overcomes the stationarity assumptions in interval kriging. It addresses objectives (ii), (Wabiri et al., 2005 attached as Appendix P.3).

In Chapter 5 (Figure 1.1, Part B) we introduce the Bi-variable copula-based geostatistics. The technique exploits a model of the joint distribution of random intervals through its copula and its marginals. The copula captures the interrelations between the random interval features, and intuitively reveals the spatial distribution of the random interval data. The marginals provide information on spatial distribution of the lower and upper variables.

In Chapter 6 (Figure 1.1, Part C) we introduce the fuzzy credibility distribution-based geostatistics. The approach exploits the maximum entropy formalism in defining a fuzzy variable set induced by random interval set. Similar to exact spatial modeling with a unique (additive) probability measure, the approach provides a unique (additive) credibility measure for modeling uncertain/fuzzy events. This leads to a flexible and tractable analytical model of the joint distribution of the spatial random interval (fuzzy) phenomena.

In Chapter 7 we apply the techniques to real-life data. Conclusions and recommendation for future research are given in Chapter 8.

Appendix A provides a review of the complete metric and linear normed spaces including the Hilbert space. Appendix B is a review of the basics of interval arithmetic as used in this study.

In Appendix C.1 and C.2 we give the algorithms for non-parametric kernel interval regression. Algorithms for generating kernel joint density, kernel copula, kernel marginals and kernel membership functions are given in Appendix C.3. Bi-variable copula-based kriging algorithm is outlined in Appendix C.4 with fuzzy credibility-based kriging algorithm in Appendix C.5.

Three published papers accruing from this study are provided in Appendix P.

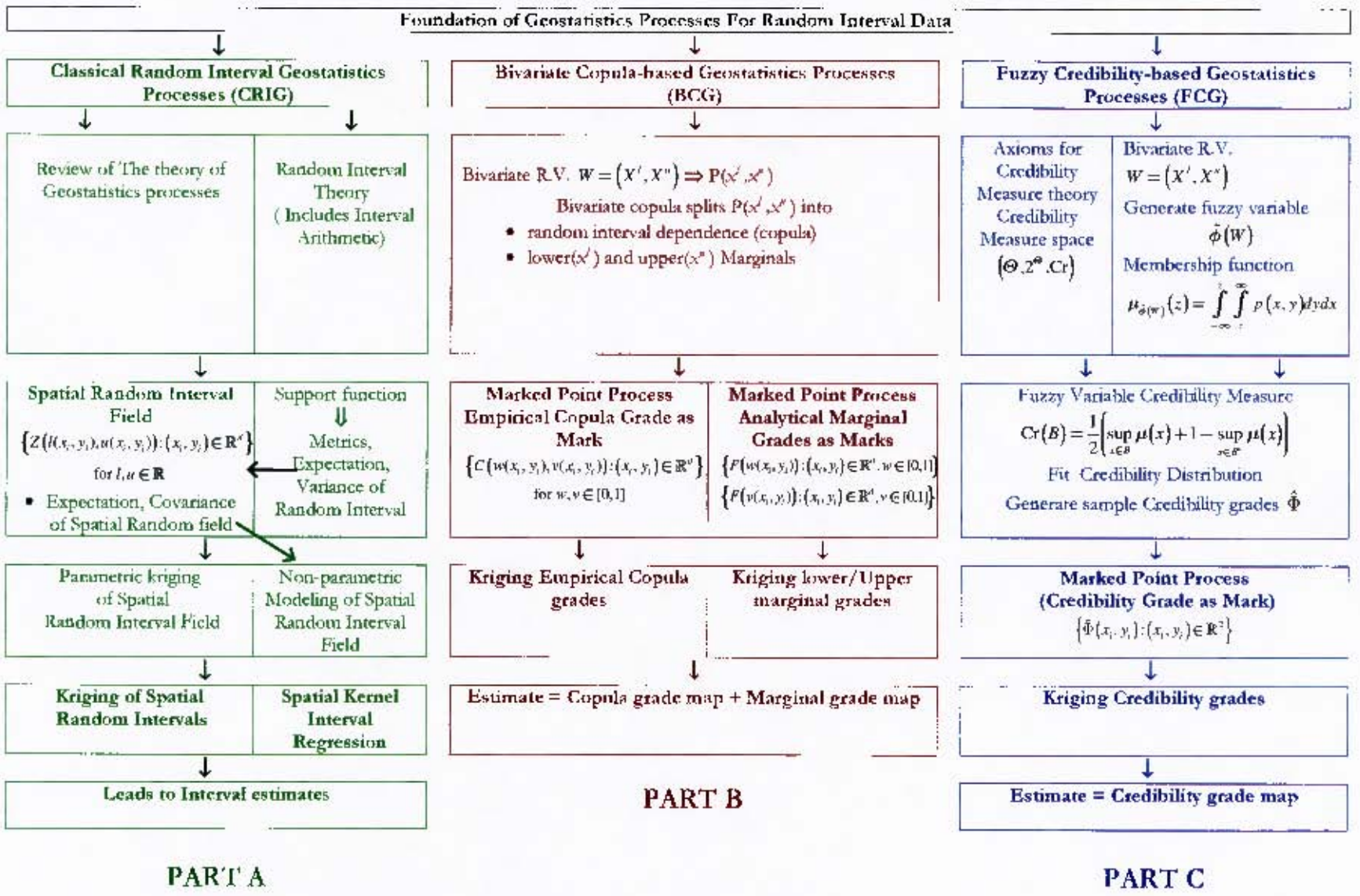


Figure 1.1. Illustrative Structure of the Research Problem

CHAPTER 2

THE SPATIAL CLOSED RANDOM INTERVAL APPROACH

Scope

We devote this chapter to development of the theoretical framework of spatial closed random interval approach as applied to spatial random interval process. A preview and introduction of spatial random process and geostatistical analysis is given. Definitions and detailed concepts of random closed intervals as used in the dissertation are given.

2.1 Spatial Random Process

In geostatistics, observed data represents natural phenomena distributed in space or in time and space e.g., gold deposits or SO_2 pollutants concentrations within a given region in space. Often, the observed phenomena exhibit immense complexity of detail that cannot be fully described using deterministic models (Webster and Oliver, 2001). The deterministic model would involve large number of parameters that consequently require data to be available for most of the locations. This is a major limitation when dealing with scarcely sampled data. An alternative approach is to apply a random process, (called a spatial random process in geostatistics (Chiles and Delfiner, 1999; Goovaerts, 1997)) where series of measurements of an attribute are assumed known.

Briefly, a random process is a sequence of random quantities identified or located (indexed) according to some metric. For example, we can think of the price of a stock denoted as Z_t at time $t = 1, 2, \ldots$, and the sequence of stock prices $\{Z_t : t \geq 0\}$ is a time indexed random process. Similarly, a spatial process is a collection of random variables that are indexed by some set $D \subseteq \mathbb{R}^d$ containing spatial coordinates $\boldsymbol{x} = x_1, x_2, \ldots, x_d$.

Application of random process to geostatistics is possible since the observed random variables (called regionalized random variables by Matheron (1963)) are seen as a plausible realization of an unobserved random process. For example, an observed gold deposit is viewed as only one possible realization of the underlying spatial random process that could have generated a whole set of gold deposits with the same characteristics. Also, a case of disease incidence at a given location is viewed as a realization of a smooth underlying disease risk surface. The underlying smooth risk surface forms a spatial random process.

2.1.1 Basics of Spatial Random Process

Spatial data refers to a data value with the associated location in space (i.e. d-dimensional Euclidean space $\mathbb{R}^d$ with $d \geq 1$). Locations may refer to a point, an area, or volume in space usually called the sample support. For example, we could have the area average disease prevalence in a given region (district or county). The notation $Z(\boldsymbol{x})$ indicates the association of measurement Z with location $\boldsymbol{x}$. In 3D, such that $d = 3$, $\boldsymbol{x}$ is a vector of coordinates (x, y, z) with $x-$latitude, $y-$longitude and $z-$altitude. The random variable Z assumes a basic structure $Z(\boldsymbol{x}) : \boldsymbol{x} \in D \subseteq \mathbb{R}^d$ over a given spatial region D in $\mathbb{R}^d$. In principle, Z could be

determined from any location $\boldsymbol{x}_i$, $i = 1, \ldots, n$ within D, and the collection of realized values; $Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n)$ or $z(\boldsymbol{x}_1), \ldots, z(\boldsymbol{x}_n)$ at locations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ relate to an unobserved spatial random field:

$$\{Z(\boldsymbol{x}) : \boldsymbol{x} \in D \subseteq \mathbb{R}^d\}. \tag{2.1}$$

Important distinctions of spatial data types relate to whether D, is continuous, fixed, or random (Cressie, 1993). For D, a fixed, continuous region, with $\boldsymbol{x}$ varying continuously over D, the spatial random process 2.1 is a geostatistical process. If D, is a fixed, regular lattice, the random process is a lattice process; whereas for D, a random set, the process is a spatial points process. A distinction between the three types of spatial data is not clear-cut. Geostatistics data can be transformed to lattice data by merging data over large finite regions. By conceptually changing the index set D, from discrete to continuous, one can move from lattice data to geostatistical data. Aggregating point pattern over large regions results into lattice data.

In this research, we will deal with the geostatistical process. For spatial point process, interested readers should refer to the following texts: Ripley (1987a); Stoyan et al. (1987); Venables and Ripley (1994); Ver Hoef and Cressie (1993) for theoretical contributions, and Cressie (1993); Diggle et al. (1998); Moller (2003); Moller and Waagepetersen (2003) for introduction to applications of spatial point processes.

2.1.2 Properties of a Spatial Random Process

The spatial random field is characterized by the joint spatial distribution of the vector $Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n)$ of realized values at locations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$. However, due to limited number of realizations, the joint spatial distribution can not be uniquely defined, except on the assumption of stationarity.

2.1.2.1 Second-order Stationarity

A spatial random process is said to be **strictly stationary** if the joint spatial distribution of $(Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n))$ is the same as that of $(Z(\boldsymbol{x}_1 + \boldsymbol{h}), \ldots, Z(\boldsymbol{x}_n + \boldsymbol{h}))$ for any n spatial locations $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in D\}$ and any distance vector $\boldsymbol{h}$, provided all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, $\boldsymbol{x}_1 + \boldsymbol{h}, \ldots, \boldsymbol{x}_n + \boldsymbol{h}$ are within the spatial domain D. Then, the mean

is constant across the region D,

$$\begin{aligned} E(Z(\boldsymbol{x})) &= m(\boldsymbol{x}) = m \\ E(Z(\boldsymbol{x}+\boldsymbol{h})) &= m(\boldsymbol{x}+\boldsymbol{h}) = m \end{aligned} \tag{2.2}$$

and the covariance between spatial random variables $Z(\boldsymbol{x})$ and $Z(\boldsymbol{x}+\boldsymbol{h})$ separated by a distance vector $\boldsymbol{h}$ is

$$\begin{aligned} C(\boldsymbol{x}, \boldsymbol{x}+\boldsymbol{h}) &= E[Z(\boldsymbol{x}) - m(\boldsymbol{x})][Z(\boldsymbol{x}+\boldsymbol{h}) - m(\boldsymbol{x}+\boldsymbol{h})] \\ &= E[Z(\boldsymbol{x})Z(\boldsymbol{x}+\boldsymbol{h})] - m(\boldsymbol{x})m(\boldsymbol{x}+\boldsymbol{h}) \qquad \text{using 2.2} \\ &= E[Z(\boldsymbol{x})Z(\boldsymbol{x}+\boldsymbol{h})] - m^2 \\ &= C(\boldsymbol{h}). \end{aligned} \tag{2.3}$$

The covariance is only a function of the separation vector $\boldsymbol{h}$. Further, if the mean 2.2 and the covariance 2.3 of a spatial process exists, and fully characterizes the spatial process, then the process is said to be **second-order stationary or weakly stationary**. It is therefore not uncommon that the terms second-order stationarity and covariance are used interchangeably in some of the geostatistics literature.

Often, $\boldsymbol{h}$ represents both the distance and direction between locations $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{h}$. Directional aspect means for example that the process may be different in the north-south direction compared to east-west direction. If $C(\boldsymbol{h})$ depends on both the distance and direction, the spatial process is said to be **anisotropic**, otherwise it is **isotropic**. We will use h to denote $\boldsymbol{h}$ except where it is otherwise or explicitly stated.

The covariance function is symmetrical i.e., $C(h) = C(-h)$, and generally decreases as h increases, eventually approaching the limit zero for large h i.e.

$$\lim_{h \to \infty} C(h) = 0. \tag{2.4}$$

The distance h with limit zero covariance is called the "**range**", which we denote by R. Normally, R represents the transition from the state in which a spatial correlation exists ($h \leq R$) to the state in which there is absence of spatial correlation ($h > R$).

Let $Z(\boldsymbol{x}_i)$ represent the regionalized variable at location $\boldsymbol{x}_i$, $i = 1, \ldots, n$, and $\text{Cov}\,(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) = C(\boldsymbol{x}_i, \boldsymbol{x}_j)$ the covariance between the regionalized variables at locations $\boldsymbol{x}_i, \boldsymbol{x}_j, i, j = 1, \ldots, n$. Since variance is always positive, the spatial

covariance satisfies the positive semi-definiteness condition (Cressie, 1993; Stein, 1999) i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0 \tag{2.5}$$

for any finite set of locations $\boldsymbol{x}_i, \ldots, \boldsymbol{x}_n$ and arbitrary real coefficients $a_i, a_j, i, j = 1, \ldots, n$ with at least one being different from zero. This follows noting that the left hand side of Eq. 2.5 is the variance of $\sum_{i=1}^{n} a_i Z(\boldsymbol{x}_i)$ (Chiles and Delfiner, 1999; Stein, 1999).

The classical estimator for covariance is given as

$$\hat{C}(h) = \frac{1}{|n(h)|}\sum_{i,j=1}^{n(h)} (z(\boldsymbol{x}_i) - \hat{m})(z(\boldsymbol{x}_j) - \hat{m}) \tag{2.6}$$

where $\hat{m}$ is the mean estimate, $n(h) = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) : \boldsymbol{x}_i - \boldsymbol{x}_j = h + \Delta h, i, j, \ldots, n\}$, $|n(h)|$ is the cardinality, and Δh is the distance lag tolerance within which to combine the pairs $\boldsymbol{x}_i, \boldsymbol{x}_j$ to create lag classes for unequally spaced data. As a rule of thumb, there should be at least 30 distinct pairs at each lag distance to stabilize $\hat{C}(h)$, and it should not be calculated for distances larger than half the maximal distance between the points (Journel and Huijbregts, 1978).

2.1.2.2 Intrinsic Stationarity

Often, the mean of a spatial random process varies across the region, and the variance tends to increase without bound as the area of interest increases. Consequently, second order stationarity no longer holds, and the covariance is not defined. Therefore, an alternative measure of spatial correlation is necessary for such processes. This also means having a wider class of spatial process. In this regard, Goovaerts (1997) highlights a class of spatial processes called the "incremental stationary process". They consist of processes whose increments $Z(\boldsymbol{x}) - Z(\boldsymbol{x} + h)$ are stationary and the mean is assumed constant for small h. For this class of processes,

$$\gamma(h) = \frac{1}{2}\text{Var}\left(Z(\boldsymbol{x}) - Z(\boldsymbol{x} + h)\right) \tag{2.7}$$

forms a measure of degree of dissimilarity between any two values, and is the alternative to covariance function. The measure $\gamma(h)$ is referred to as the **(semi)-variogram**. Assuming second-order stationarity of increments, the variogram 2.7

reduces to

$$\gamma(h) = \frac{1}{2}E[Z(\boldsymbol{x}) - Z(\boldsymbol{x}+h)]^2 \tag{2.8}$$

The incremental stationary process satisfying Eq. 2.8 are referred to as intrinsically stationary processes, and hence the term **intrinsic stationarity** (Chiles and Delfiner, 1999; Goovaerts, 1997). It is not uncommon that the terms intrinsic stationarity and variogram are used interchangeably.

Parallel to covariance, the variogram is symmetric, $\gamma(h) = \gamma(-h)$, and generally increases with corresponding increase in h. Let C_1 denote the a priori variance of the spatial random process. In practice, the variogram stops increasing beyond the **range** and becomes more or less stable around a value C_1, also called the **sill** value,

$$\lim_{h\to\infty} \gamma(h) = \mathrm{Var}\,(Z(\boldsymbol{x})) = C_1 (= \sigma^2) \tag{2.9}$$

A **nugget effect** is said to occur if

$$\lim_{h\to 0} \gamma(h) = C_0 > 0 \tag{2.10}$$

By definition $\gamma(0) = 0$. However, Eq.2.10 implies that $\gamma(h)$ approaches a constant, C_0, as h tends to zero. The constant C_0 is the **nugget effect** (Cressie, 1993; Stein, 1999). This implies that microscale variation (small nuggets) is causing discontinuity at the origin, hence, the spatial process is fundamentally discontinuous. If continuity is expected at the origin, Cressie (1993) notes that the only possible reason for $C_0 > 0$ is measurement error: that is, if measurements were made several times, the results would fluctuate around the true value with a measurement error variance denoted as C_{ME}. However, for a spatial process with only a single set of values, the only cause for discontinuity at origin is the existence of a microscale process whose variance, say, C_{MS} is the nugget effect. With only one observation at each location effects of measurement error and nugget effect are confounded. Nonetheless measurement error and nugget effect are different concepts. Most often, kriging equations implicitly assume that $C_{\mathrm{ME}} = 0$ and deal only with the microscale variance as the nugget effect. A high nugget effect is an indication of a highly irregular spatial process, whereas small or zero nugget indicates a very regular and smooth spatial process (Stein, 1999). By this, if the origin is efficiently modeled, then we can effectively identify the nature of the spatial random process.

Figure 2.1 (Thiart, 2005, Figure 6.2) shows an example of a variogram.

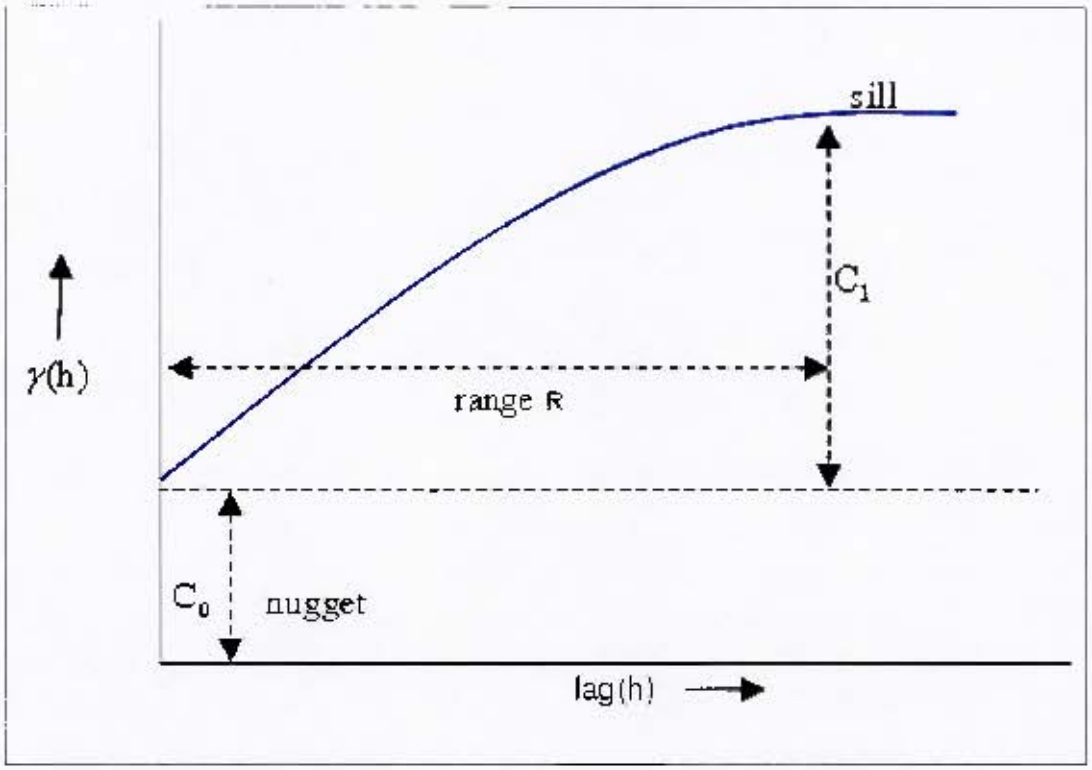


Figure 2.1. An Example of a variogram with range, nugget and sill and range, (Thiart, 2005, Figure 6.2).

The classical variogram estimator, also called the methods of moment estimator (MOM) is defined as

$$\hat{\gamma}(h) = \frac{1}{2|n(h)|} \sum_{i,j=1}^{n(h)} [z(\boldsymbol{x}_i) - z(\boldsymbol{x}_j)]^2 \tag{2.11}$$

where $n(h) = \{(\boldsymbol{x}_i, \boldsymbol{x}_j) : \boldsymbol{x}_i - \boldsymbol{x}_j = h\}$ i.e., pairs at lag h apart and $|n(h)|$ is the cardinality. The estimate does not require knowledge of the mean. If there is an unknown trend, the variogram filters the trend (Schabenberger and Gotway, 2005, pp 137). This is an advantage of use of the variogram over the covariance function.

Theoretically, the class of intrinsic random fields includes the class of second-order stationary random processes. Under second-order stationarity, if the nugget effect is zero: $\gamma(0) = 0$ or $C_0 = 0$ for $h = 0$, then $C(0) = \mathrm{Var}\,(Z(\boldsymbol{x}))$. In addition

using 2.3 and 2.8, we have the relationship

$$\gamma(h) = C(0) - C(h) \tag{2.12}$$

The variogram of second-order stationary process needs to be conditionally negative semi-definite (Chiles and Delfiner, 1999):

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(\boldsymbol{x}_i - \boldsymbol{x}_j) \leq 0 \tag{2.13}$$

for any finite set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and any constants $a_1, \ldots, a_n$ with $\sum_{i=1}^{n} a_i = 0$. This is a necessary condition for $\gamma(h)$ to be a valid semi-variogram in the general intrinsically stationary case (Cressie, 1993).

The variogram, like the covariance is isotropic only if it is a function of distance, otherwise it is anisotropic. Isotropic processes are convenient to deal with since there are a number of widely accepted parametric forms for covariance $C(h)$ and variogram $\gamma(h)$ (Chiles and Delfiner, 1999). Examples of isotropic semi-variogram functions:

Linear:

$$\gamma(\boldsymbol{h}) = \begin{cases} C_0 + bh & \text{if} \quad h > 0 \\ C_0 & \text{otherwise} \end{cases} \tag{2.14}$$

$C_0 \geq 0$ is the nugget effect; $b = C_1/R$, is the slope, C_1 is the sill, and R the range

Spherical:

$$\gamma(h) = \begin{cases} C_0 + C_1 & \text{if} \quad h > R \\ C_0 + C_1\left(\frac{3}{2}\frac{h}{R} - \frac{1}{2}\left(\frac{h}{R}\right)^3\right) & \text{if} \quad 0 < h \leq R \\ 0 & \text{otherwise} \end{cases} \tag{2.15}$$

The model is valid for $d = 1, 2, 3$, but fails the conditionally negative semi-definite condition for higher dimensions. The linear and spherical models reach a finite sill value and are described as transitive.

Exponential:

$$\gamma(h) = \begin{cases} C_0 + C_1\left(1 - e^{-\frac{h}{R}}\right) & \text{if} \quad h > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.16}$$

Though simpler in functional form to spherical models, it does not have a finite

range. It is valid for all h, but for 2-dimensional case, the Matérn function (2.20) is a better choice.

Gaussian:

$$\gamma(h) = \begin{cases} C_0 + C_1\left(1 - e^{-\left(\frac{h}{R}\right)^2}\right) & \text{if } h > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

Exponential power model:

$$\gamma(h) = \begin{cases} C_0 + C_1\left(1 - e^{-\left|\left|\frac{h}{R}\right|\right|^p}\right) & \text{if } h > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.18}$$

For $p = 1$, the model is Exponential, and for $p = 2$ it is Gaussian.

Power:

$$\gamma(h) = \begin{cases} C_0 + C_1 h^\lambda & \text{if } h > 0 \text{ and } 0 \leq \lambda < 2 \\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

It is the only non-monotonic function. The model forms a general case of linear model and does not represent any random process.

Matérn: Defined in terms of correlation

$$\rho(h) = \begin{cases} \frac{1}{2^{v-1}\Gamma(v)}\left(\frac{2h\sqrt{v}}{R}\right)^v K_v\left(\frac{2h\sqrt{v}}{R}\right) & \text{if } h > 0 \\ C_0 + C_1 & \text{if } h = 0 \end{cases} \tag{2.20}$$

where v is the smoothness parameter; $\Gamma(\cdot)$ a gamma function; and $K_v(\cdot)$ a modified Bessel function of order v (Abramowitz and Stegun, 1972). The Matérn class includes the exponential case when $v = 1/2$, and the Gaussian as a limiting case when $v \to \infty$. Being able to strike a balance between these two extremes, the Matérn class is well suited for modeling isotropic random processes. For $v = 3/2$, Handcock and Stein (1993); Jennifer et al. (2004); Stein (1999) gives the convenient closed form for $\gamma(h)$ and $C(h)$ defined by:

$$\gamma(h) = \begin{cases} C_0 + C_1\left[1 - \left(1 + \frac{h}{R}\right)e^{-\left(\frac{h}{R}\right)}\right] & \text{if } \quad h > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.21}$$

The isotropic models form the core of classical geostatistics and are used in translating anisotropic covariance or variogram models to isotropic models (Goovaerts,

1997; Webster and Oliver, 2001). Let $\gamma_0, \ldots, \gamma_p$ denote isotropic variogram functions and $\boldsymbol{A}_0, \ldots, \boldsymbol{A}_p$ denote transformation matrices for transforming the spatial coordinates of the observations to restore isotropy by removing the directional effect. Then, we can transform simple non-homogeneous models e.g., the **geometric anisotropy model,** (same sill and different ranges) to isotropic model given as

$$\gamma(h) = \gamma_0(||\boldsymbol{A}_0 h||$$

and the **zonal anisotropy model,** (different sill and different ranges) to

$$\gamma(h) = \sum_{i=0}^{p} \gamma_i(||\boldsymbol{A}_i h||)$$

2.2 Variogram Estimation

For large sample ($n \geq 30$), the classical variogram estimator 2.11 is approximately normal, and it is not robust in the presence of outliers. Suggested robust variogram estimators include: Genton (1998a,b) highly robust variogram estimator and Cressie and Hawkins (1984) variogram estimator also called **madogram** by Goovaerts (1997).

Both the MOM estimator 2.11, and the robust estimator fail the conditional non-positive-definiteness condition, leading to spatial predictions with negative variances. The most common solution is to replace the empirical variogram $\gamma(h)$ by the isotropic parametric models known to be conditionally positive definite.

Among the proposed methods for fitting the isotropic models, least squares methods is the most commonly applied method but not without limitations. Suppose we wish to fit a parametric model $\gamma(h; \boldsymbol{b})$, parameterized with finite parameter vector $\boldsymbol{b}$, to the estimated variogram $\gamma(h)$. We let γ_{mom} denote the vector of MOM estimates at given set of lag values h, and $\gamma_{\boldsymbol{b}}$ the vector of model predicted values at same lag values of h. Cressie (1993) has outlined three versions of non-linear least-squares estimators:

Ordinary least squares(OLS): Choose $\boldsymbol{b}$ to minimize

$$\{\gamma_{\text{mom}} - \gamma_{\boldsymbol{b}}\}^T \{\gamma_{\text{mom}} - \gamma_{\boldsymbol{b}}\}$$

assuming observations are uncorrelated. With correlated data the method is biased.

Generalized Least Squares (GLS): Choose $\boldsymbol{b}$ to minimize

$$\{\gamma_{\text{mom}} - \gamma_{\boldsymbol{b}}\}^T V(\boldsymbol{b})^{-1}\{\gamma_{\text{mom}} - \gamma_{\boldsymbol{b}}\}$$

where $V(\boldsymbol{b})$ denotes the variance-covariance matrix of $\gamma_{\boldsymbol{b}}$.

Weighted least squares (WLS): Choose $\boldsymbol{b}$ to minimize

$$\{\gamma_{\text{mom}} - \gamma_{\boldsymbol{b}}\}^T W(\boldsymbol{b})^{-1}\{\gamma_{\text{mom}} - \gamma_{\boldsymbol{b}}\}$$

where $W(\boldsymbol{b})$ is a diagonal matrix whose diagonal elements are the variances of the entries of $\gamma_{\boldsymbol{b}}$.

The weighted least square is a compromise between efficiency (generalized least squares) and simplicity (ordinary least squares). The procedure gives more weight to observations that have small variance, and down weights observations that have large variances. Proposed methods for fitting least squares include Genton (1998b) generalized least squared with a known covariance structure and Cressie (1985) weighted least squared criterion. Once variogram is optimally fitted, estimates of the spatial random field are obtained via kriging. Prior to addressing the problem of spatial prediction (or kriging), we revise the link between random functions space and a Hilbert spaces, building on Kolmogorov (1933) theory of probability.

2.3 Kriging Methodology

2.3.1 Spatial Random Fields and Hilbert Spaces

A "Hilbert space" is an infinite dimensional vector spaces, the next best thing to a Euclidean space–a finite dimensional inner product space. The single most useful property of these spaces is that they permit the introduction of complete (and by assumption orthonormal) sets of deterministic functions which constitute a basis. If continuous spatial random fields can be defined as continuous areas in a Hilbert space, then we can easily apply the projection theorem given in Luenberger (1987); Stein (1999), to predict an unknown spatial random field.

Let Ω be an open set in $\mathbb{R}^n$ (Ω can be the whole $\mathbb{R}^n$) and V be a finite-dimensional vector space. For the space $L^2(V, \Omega)$, a set of vector valued functions $f = (f_1, \ldots, f_n)$

on Ω such that $\sum_{i=1}^{n} \int_{\Omega} |f_i(\boldsymbol{x})|^2 < \infty$, $L^2(V, \Omega)$ is a **Hilbert space** with inner product

$$\langle f, g \rangle = \sum_{i=1}^{n} \int_{\Omega} f_i(\boldsymbol{x}) g_i(\boldsymbol{x}) d\boldsymbol{x} \tag{2.22}$$

Alternatively, a space consisting of random variables Z and W with finite quadratic mean, $E[Z^2] < \infty$, is (almost) a Hilbert space with inner product $E[ZW]$ (Stein, 1999). The induced norm, $||Z|| = \sqrt{EZ^2}$, and convergence in norm $||\cdot||$ is equivalent to convergence in quadratic mean of the random variables. Distance between the two random variables Z and W is

$$d^2(Z, W) = E(Z - W)^2, \tag{2.23}$$

and equality of the random variables $Z = W$ implies $Pr[Z = W] = 1$, hence $||Z - W||^2 = E(Z - W)^2 = 0$.

The mathematical properties of the inner product $E[ZW]$ are similar to covariance properties between random variables whose mean and covariance exists and are finite. Stein (1999) gives a special case of random variables with zero mean, finite variance, where the inner product

$$E[ZW] = \begin{cases} \text{Cov}\,(ZW) \\ \text{Var}\,(Z) \qquad \text{iif} \quad Z = W \end{cases} \tag{2.24}$$

Let $D = \Omega \subseteq \mathbb{R}^2$ be a spatial region of interest, and $Z(\boldsymbol{x}) : \boldsymbol{x} \in D$ be a spatial random field such that $E[Z^2(\boldsymbol{x})] < \infty$. Then, for each $\boldsymbol{x}$, $Z(\boldsymbol{x})$ is an element of the Hilbert space just defined. Parametrically, since the spatial random field $Z(\boldsymbol{x})$ is a family of random variables in a two dimensional parameter space $\boldsymbol{x}$, it can viewed as an "area" in the Hilbert space. The area is continuous if

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}'} E\left[Z(\boldsymbol{x}) - Z(\boldsymbol{x}'))^2\right] = 0, \qquad \boldsymbol{x}' \in D$$

Spatial random fields satisfying this condition are said to be **continuous in quadratic mean.** Consequently, continuity in quadratic mean corresponds to continuity in Hilbert space. Then, spatial random fields are elements of Hilbert space and projection in Hilbert space parallels spatial prediction.

Replacing Ω with D, the spatial region of interest, $L^2(V, D)$ parallels the vector space of greatest importance. A collection of vector valued random fields $Z(\boldsymbol{x})$ in

$L^2(V, D)$ for which

$$\int_D E[Z^2(\boldsymbol{x})]d\boldsymbol{x} < \infty \tag{2.25}$$

constitute a linear vector space, with the inner product

$$E\langle Z(\boldsymbol{x}), W(\boldsymbol{x})\rangle = E\left[\int_D Z(\boldsymbol{x})W(\boldsymbol{x})d\boldsymbol{x}\right] \tag{2.26}$$

By letting $\langle Z(\boldsymbol{x}), W(\boldsymbol{x})\rangle$ denote the 2D space-domain inner product; the deterministic portion of the inner product (2.26) and the expected value portion are explicitly indicated. This representation allows certain theoretical manipulations to be performed easily.

One of the most interesting results of the theory of random fields is that the normed vector space for the random fields previously defined is separable. Consequently, there exist a complete (and, by assumption, orthonormal) set $\{h_i(\boldsymbol{x})\}, i = \{1, \ldots\}$ of deterministic functions which constitute a basis. Then, a random field in the space of random fields can be represented as

$$Z(\boldsymbol{x}) = \sum_{i=1}^{\infty} Z_i h_i(\boldsymbol{x}) \forall \boldsymbol{x} \in D \tag{2.27}$$

where $\{Z_i\}$, the representation of $Z(\boldsymbol{x})$, is a sequence of random variables given by

$$\begin{aligned} Z_i &= \langle Z(\boldsymbol{x}), h_i(\boldsymbol{x})\rangle \\ &= \int_D Z(\boldsymbol{x})h_i(\boldsymbol{x})d\boldsymbol{x} \end{aligned}$$

A special class of random fields, the second-order stationary random fields, with finite mean and covariance, form elements of the Hilbert space. This allows us to use projection theorem (Luenberger, 1987; Stein, 1999) to approximate any second-order stationary random field.

2.3.2 Kriging Methodology

Kriging methodology is a spatial prediction approach. Assuming a stationary random field $Z(\cdot)$ with finite first and second order moments on compact set D in $\mathbb{R}^d$, we define the random field model

$$Z(\boldsymbol{x}) = m(\boldsymbol{x}) + e(\boldsymbol{x}) \qquad \boldsymbol{x} \in D \tag{2.28}$$

where $m(\cdot)$ is the mean of $Z(\cdot)$ and $e(\cdot)$ a random field with mean zero. The goal is to predict linear functional of $Z(\cdot)$ on D based on observed $Z(\boldsymbol{x}_i) \in D, i = 1, \ldots, n$. Assuming $EZ(\boldsymbol{x}) < \infty$ and $E[Z(\boldsymbol{x})]^2 < \infty$, the mean and covariance of the random field $Z(\cdot)$, are respectively defined as

$$\begin{aligned} E(Z(\boldsymbol{x})) &= m(\boldsymbol{x}) \\ \text{Cov}\,(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) &= E\left[Z(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)\right]\left[Z(\boldsymbol{x}_j) - m(\boldsymbol{x}_j)\right] \\ &= \begin{cases} EZ(\boldsymbol{x}_i)Z(\boldsymbol{x}_j) & \text{if } E[Z(\boldsymbol{x}_i)] = 0 \\ E[Z(\boldsymbol{x})^2] = \text{Var}\,(Z(\boldsymbol{x})) & \text{if } \boldsymbol{x}_i = \boldsymbol{x}_j = \boldsymbol{x} \end{cases}, \end{aligned}$$

For example, for $d = 2$, the spatial random field $Z(\cdot)$ represents a family of random variables in a two dimensional parameter space $\boldsymbol{x} \in \mathbb{R}^2$, and it corresponds to an area in Hilbert space, and hence the use of projection theorem to approximate the spatial random field.

Statistically, if $Z(\boldsymbol{x}_0)$ (for $\boldsymbol{x}_0 \in \mathbb{R}^2$) is a random variable and $Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n)$ is a finite set of observed random variables, then, for $c_i \in \mathbb{R}, i = 1, \ldots, n$, its projection, $\hat{Z}(\boldsymbol{x}_0) = c_1 Z(\boldsymbol{x}_1) + \cdots + c_n Z(\boldsymbol{x}_n)$, is a unique linear combination closest to $Z(\boldsymbol{x}_0)$ in the $||\cdot||$-norm such that

$$\left|\left|\sum c_j Z(\boldsymbol{x}_j) - Z(\boldsymbol{x}_0)\right|\right|^2 = E\left(\sum c_j Z(\boldsymbol{x}_j) - Z(\boldsymbol{x}_0)\right)^2 \tag{2.29}$$

is minimal. The projection is characterized by the requirement that the residual $\sum c_j Z(\boldsymbol{x}_j) - Z(\boldsymbol{x}_0)$ is uncorrelated with all the $Z(\boldsymbol{x}_j)$ i.e., $\langle Z(\boldsymbol{x}_0) - \sum c_j Z(\boldsymbol{x}_j), Z(\boldsymbol{x}_j)\rangle = 0$. This is the least squares solution to the common known regression problem (or kriging predictor in geostatistics).

Formally, let,

$Z^*(\boldsymbol{x}_0)$	estimator of $Z(\boldsymbol{x}_0)$
$\boldsymbol{m}_i$	Mean of $Z(\boldsymbol{x}_i)$
$\boldsymbol{m}_0$	Mean of $Z(\boldsymbol{x}_0)$
λ_i	Weights associated with $Z(\boldsymbol{x}_i)$
$C_{ij} = \text{Cov}\,(\boldsymbol{x}_i, \boldsymbol{x}_j)$	Covariance between $Z(\boldsymbol{x}_i)$, $Z(\boldsymbol{x}_j)$
$C_{i0} = \text{Cov}\,(\boldsymbol{x}_i, \boldsymbol{x}_0)$	Covariance between $Z(\boldsymbol{x}_i)$, $Z(\boldsymbol{x}_0)$
$C_{00} = \text{Var}\,(Z(\boldsymbol{x}_0))$	(Co)variance of $Z(\boldsymbol{x}_0)$

All kriging estimators are variants of the basic linear regression estimator:

$$Z^*(\boldsymbol{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i) \tag{2.30}$$

subject to condition that $\sum_{i=1}^{n} \lambda_i = 1$, except for simple kriging where weights do not have to sum to one. To estimate (2.30), we seek to minimize the expected mean prediction error (mpe), (2.29), given as

$$E(Z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0))^2 = \text{Var}\,(Z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)) + [E(Z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0))]^2 \tag{2.31}$$

where the first term is the error variance, σ^2, and the last term the squared bias.

Among the class of linear estimators, the kriging estimator is the Best Linear Unbiased Predictor (BLUP) (Cressie, 1993; Stein, 1999). **"Best"** means that the estimator is unbiased and optimal in the sense of minimum mean squared error. The unbiased condition requires that

$$E(Z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)) = 0, \tag{2.32}$$

which eliminates the second term in (2.31). Then optimal condition requires that weights λ_i are chosen to minimize error variance, σ^2 defined as

$$\begin{aligned} \sigma^2 &= \text{Var}\,(Z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)) \\ &= \text{Var}\,(Z^*(\boldsymbol{x}_0)) - 2\text{Cov}\,(Z^*(\boldsymbol{x}_0), Z(\boldsymbol{x}_0)) + \text{Var}\,(Z(\boldsymbol{x}_0)) \end{aligned} \tag{2.33}$$

Substituting (2.30), the error variance (2.33) becomes

$$\begin{aligned} \sigma^2 &= \text{Var}\left(\sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i)\right) - 2\text{Cov}\left(\sum_{i=1}^{n} \lambda_i Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_0)\right) + \text{Var}(Z(\boldsymbol{x}_0)) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C_{ij} - 2\sum_{i=1}^{n} \lambda_i C_{i0} + C_{00} \end{aligned} \tag{2.34}$$

under unbiasedness condition 2.32

Three main forms of kriging are defined based on the form of the mean structure (Goovaerts, 1997):

Simple Kriging(SK) : the mean of the process is known and constant over the whole study area.

Ordinary Kriging(OK) : the mean of the process is unknown but constant within a local neighborhood.

Kriging with a trend (or drift) : the unknown mean is not constant even within a local neighborhood. Examples include Universal kriging(UK) and a more general approach, the Intrinsic Random Function of Order k (IRF-k) (Chiles and Delfiner, 1999). Briefly, under UK, mean is modeled as a linear combination of functions, $f_k(\boldsymbol{x})$ of the spatial coordinates $\boldsymbol{x} = (x_1, x_2)$ in a $2D$: $\mu(\boldsymbol{x}) = \sum_{k=0}^{K} \beta_k f_k(\boldsymbol{x})$, with β_k the unknown coefficients. Usually the first function, $f_0(\boldsymbol{x}) = 1$ to guarantee the inclusion of the constant-mean case in the spatial model. The other functions are monomials of low degree in the coordinates of $\boldsymbol{x}$.

The Intrinsic Random Function of Order k (IRF-k) accommodates varying nature of the trend. The term k represents the order of polynomial trends- k = 0 means constant drift, and the IRF-k is equivalent to ordinary kriging; if k = 1, we have linear drift; k = 2 yields quadratic drift. The IRF-k models are defined by having increments of a sufficiently high order for stationarity to be reached, thus overcomes the problem of non-stationarity affecting the variogram estimation.

Cokriging is a multivariate extension of kriging (Goovaerts, 1997, p.203-248). Different forms of cokriging depend on whether the secondary regionalized variable is known everywhere and if it varies smoothly across the study area. If both cases are satisfied, then there is no loss of information if we retain in the cokriging system only the secondary datum collocated with the location $\boldsymbol{x}$ being estimated (Goovaerts, 1998). In deed, the collocated value tends to screen the influence of further away secondary variables introducing negative weights for secondary variables.

Cokriging and IRF-k approaches are not addressed is this thesis. We extending ordinary kriging to modeling spatial random intervals based on the corresponding random bivariate variable. Extension of kriging with trend is given only for illustration purposes.

Ordinary Kriging (OK)

In ordinary kriging, stationarity is assumed within a local neighborhood $n(h)$. Here $n(h)$ refers to the number of points within a given distance (h) or area of the

location $\boldsymbol{x}_0$ being estimated. For local constant mean $m(\boldsymbol{x}) = m$, and conditional to the unbiasedness condition (2.32), the OK estimator is given as

$$Z^*_{ok}(\boldsymbol{x}_0) = \sum_{i=1}^{n(h)} \lambda_i Z(\boldsymbol{x}_i) \tag{2.35}$$

The unbiasedness condition 2.32 reduces to

$$E(Z^*(\boldsymbol{x}_0) - Z(\boldsymbol{x}_0)) = E\Big[\sum_{i=1}^{n(h)} \lambda_i Z(\boldsymbol{x}_i)\Big] - Z(\boldsymbol{x}_0) = m\left(\sum_{i=1}^{n(h)} \lambda_i - 1\right) = 0, \tag{2.36}$$

Implying that weights sum to one. The optimization problem is then solved using a method of Lagrange multipliers by defining

$$L(\lambda_i,\ i = 1, \ldots, n(h); \psi) = \sigma^2 + 2\psi\left(\sum_{i=1}^{n(h)} \lambda_i - 1\right) \tag{2.37}$$

where ψ is a Lagrange multiplier. A positive Lagrange multiplier means that the samples are not efficiently spread around the estimated point. For example, it might be the case of clustering, then we are using points behind other points which introduce a screening effect. A negative Lagrange multiplier leads to smaller variance and implicitly means that the samples are well spread out relative to how close they are to the estimator. Thus, a large negative Lagrange multiplier implies that the samples are spread out evenly and close to the prediction location.

Substituting 2.34 for σ^2 in 2.37 we get

$$\begin{aligned} L(\lambda_i,\ i = 1, \ldots, n(h); \psi) \ = \ & \sum_{i=1}^{n(h)} \sum_{j=1}^{n(h)} \lambda_i \lambda_j C_{ij} - 2\sum_{i=1}^{n(h)} \lambda_i C_{i0} + C_{00} \\ & + 2\psi\left(\sum_{i=1}^{n(h)} \lambda_i - 1\right) \end{aligned} \tag{2.38}$$

and the error variance is minimized by setting to zero each of the $n + 1$ partial first

derivatives:

$$\frac{\delta L}{\delta \lambda_i} = \sum_{j=1}^{n(h)} \lambda_j C_{ij} - C_{i0} + \psi = 0 \qquad i = 1, \ldots, n(h)$$

$$\frac{\delta L}{\delta \psi} = \sum_{i=1}^{n(h)} \lambda_i - 1 = 0$$

resulting to the ordinary kriging system of linear equations

$$\begin{aligned} \sum_{j=1}^{n(h)} \lambda_j C_{ij} + \psi &= C_{i0} \qquad i = 1, \ldots, n(h) \\ \sum_{j=1}^{n(h)} \lambda_j &= 1 \end{aligned} \tag{2.39}$$

and an ordinary kriging variance

$$\sigma^2_{ok}(Z^*(\boldsymbol{x}_0)) = C_{00} - \sum_{i=1}^{n(h)} \lambda_i C_{i0} - \psi$$

Let

$\gamma_{ij} = C_{00} - C_{ij}$	variogram between Z_i and Z_j
$\gamma_{i0} = C_{00} - C_{i0}$	variogram between Z_i and Z_0
γ_{00}	variogram of Z_0

In terms of variogram the ordinary kriging system (Cressie, 1993), 2.39 reduces to

$$\begin{aligned} \sum_{j=1}^{n(h)} \lambda_j \gamma_{ij} - \psi &= \gamma_{i0} \qquad i = 1, \cdots, n(h) \\ \sum_{j=1}^{n(h)} \lambda_j &= 1, \end{aligned} \tag{2.40}$$

which in matrix notation is written as

$$\boldsymbol{\Gamma \lambda} = \boldsymbol{\gamma}, \tag{2.41}$$

where

$$\Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n(h)} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{n(h)1} & \cdots & \gamma_{n(h)n(h)} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}_{p\times p}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_{n(h)} \\ -\psi \end{bmatrix}_{p\times 1}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{n(h)0} \\ 1 \end{bmatrix}_{p\times 1}$$

where $p = n(h) + 1$. The OK variance is

$$\sigma^2_{ok}(Z(\boldsymbol{x}_0)) = \boldsymbol{\gamma}'\Gamma^{-1}\boldsymbol{\gamma} - \frac{\mathbf{1}'\Gamma^{-1}\boldsymbol{\gamma} - 1}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}. \tag{2.42}$$

Since only $\boldsymbol{\gamma}$ changes in (2.41) for different $\boldsymbol{x}_0$, Γ only has to be calculated once for each observation in the neighborhood window, it is easy to predict the process for several locations once the variogram has been estimated.

Important points :

- the kriging system and kriging variance depend only on the covariance function and on spatial lay-out of the sampled support and not the actual data values.
- the semi-variogram, $\boldsymbol{\gamma}$, performs statistical distance weighting of the data points in the support, while Γ^{-1} rescales the weights in $\boldsymbol{\gamma}$ to add to one and also, the semi-variogram, Γ^{-1} allows for possible redundancies in the support, i.e., it attempts to compensate for any possible clustering of the support points. This combination is the key power of kriging.

More details about kriging techniques are found in a series of geostatistics books: Isaaks and Srivastava (1989) for basic introduction; Chiles and Delfiner (1999), Goovaerts (1997), Stein (1999), and Cressie (1993) for detailed developments and discussions.

We also note that kriging is by construction designed to handle single or real-valued measurements, hence if the data is plagued with imprecision (fuzziness), the method breaks down. To address this problem, we introduce the spatial random closed interval approach, and extend the ordinary kriging to handle spatial random intervals.

2.4 The Spatial Closed Random Interval Approach

2.4.1 Closed Random Intervals

Assume that we are representing uncertainty about a system in the world through reference to a universe of discourse, denoted by Θ, with $2^{\Theta} = \{A : A \subseteq \Theta\}$ the set of all objects on which uncertainty can be valued. Θ may be finite, countable/uncountable, and can take different forms either real values, sets or intervals.

We recall the definition of a random variable.

Definition 2.4.1 Given a probability space $(\Omega, \mathcal{B}_\Omega, Pr)$, a real-valued function $X : \Omega \to \Theta$ is a **random variable** if X is Pr-measurable, i.e., $\forall \theta \in \Theta, X^{-1}(\theta) = \{\omega \in \Omega : X(\omega) \in \theta\} \in \mathcal{B}_\Omega$. In this case, Θ consist of real values. The random variable X induces probabilities, $P(= Pr * X^{-1})(\theta)$ to each $\theta \in \Theta$. Similarly,

Definition 2.4.2 Given the probability space $(\Omega, \mathcal{B}_\Omega, Pr)$, the set-valued function $X : \Omega \to 2^{\Theta} \setminus \emptyset$ is a **random set** if X is Pr-measurable, such that $\forall A \subseteq 2^{\Theta} \setminus \emptyset, X^{-1}(A) \in \mathcal{B}_\Omega$. In this case, Θ consist of set values. The random set X associates a probability
$P(= Pr * X^{-1})(A)$ to each $A \subseteq 2^{\Theta} \setminus \emptyset$.

Let the universe of discourse Θ be the real space, i.e. $\Theta = \mathbb{R}$. Instead of working with arbitrary sets of $\mathbb{R}$, we limit ourselves to set of closed intervals $\boldsymbol{a} = [a^l, a^u] \in \mathbb{IR}$, $a^l, a^u \in \mathbb{R}$, and define

$$\mathcal{B}_{\mathbf{R}} := [a^l, a^u] \in \mathbb{IR} : a^l, a^u \in \mathbb{R}, a^l \leq a^u$$

as the Borel field of closed intervals defined on $\mathbb{R}$. Then,

Definition 2.4.3 Given the probability space $(\Omega, \mathcal{B}_\Omega, Pr)$, a **closed random interval**, denoted as ζ, is an interval-valued function $\zeta : \Omega \to \mathcal{B}_{\mathbf{R}}$, with the property that if $\boldsymbol{a} \in \mathcal{B}_{\mathbf{R}}$, then, $\zeta^{-1}(\boldsymbol{a}) = \{\omega \in \Omega : \zeta(\omega) \in \mathcal{B}_{\mathbf{R}}\} \in \mathcal{B}_\Omega$. It follows that, $P(= Pr * \zeta^{-1})(\boldsymbol{a})$ is the probability associated with each closed random interval.

In general, given random variables $X, Y : \Omega \to \mathbb{R}$ such that $X \leq Y$, let $\zeta = [X(\omega), Y(\omega)]$ and $\zeta_1 = (X(\omega), Y(\omega))$ respectively defines closed and open random intervals, the following statements are equivalent:

(a) ζ is strongly measurable
(b) X, Y are measurable
(c) ζ_1 is strongly measurable

and (c) $\Rightarrow$ (a) (Miranda et al., 2005).

Respectively, derived random variables $Z = (X + Y)/2$ and $W = (Y - X)/2$ represents the interval center (denoted as "mid") and interval radius (denoted as "rad" or "spr"). The "mid" is a real-valued random variable, while the "rad" is always a positive random variable (Körner, 1995). Since any number $a \in \mathbb{R}$ can be represented as the degenerate interval $[a, a]$, closed random intervals form extension of real numbers.

Under closed random interval theory, available information of the outcome is represented by the probability measure space $(\Omega, \mathcal{B}_\Omega, Pr)$ and a random interval, ζ, such that for all $B \in \mathcal{B}_\Omega$, what is observed is random interval $\zeta(B)$ within which the true value lies (Stainslaw, 1990). The observed random interval is a general form of knowledge that describes two essential features of the knowledge about the value of a parameter. First, an interval expresses the imprecision of the knowledge about parameter value, and second, the randomness captures the uncertainty about this value.

From definition 2.4.3, it is clear that each closed random interval induces a probability measure P on the space $(\mathbb{R}, \mathcal{B}_\mathbb{R}, P)$, which leads to a class $\mathcal{P}$ of all the possible probability measures for each random interval. Congar (2000) indicates that this class of probability measures completely summarizes the situation of uncertainty, and contains all the possible probability measure governing the situation of uncertainty represented by the closed random interval. Dempster (1968) represent this class of probability measures by a system of **upper** and **lower** probability measures or respectively the **belief** (Bel) and **plausibility** (Pl) functions (Stainslaw, 1990).

2.4.2 Closed Intervals as Elements of Two Dimensional Space

Applying the x-y coordinates system concept, a closed random interval can be viewed as a two-dimensional random vector with lower-upper(l-u) or center-radius (c-r) coordinates random variables defined on a common probability space. Hence, for closed intervals assume $l - u$ or $c - r$ coordinates systems.

Let $\Re^+ = \mathbb{R}^+ \cup \infty = [0, \infty]$, denote the set of extended real positive numbers with $\mathbb{R}^+$ the set of real positive numbers. The set of closed random intervals defined on $\Re^+$ given as $\mathcal{I} = \{[a, b] : a, b \in \mathbb{R}^+, a \geq b\}$ can be represented as points

$$\mathcal{T} = \{(a, b) : a, b \in \Re^+, a \geq b\}, \quad \text{with } [a, b] = \emptyset \quad \text{for } a \geq b. \tag{2.43}$$

in an extended two dimensional space,$\mathcal{T}$, (Figure 2.2).

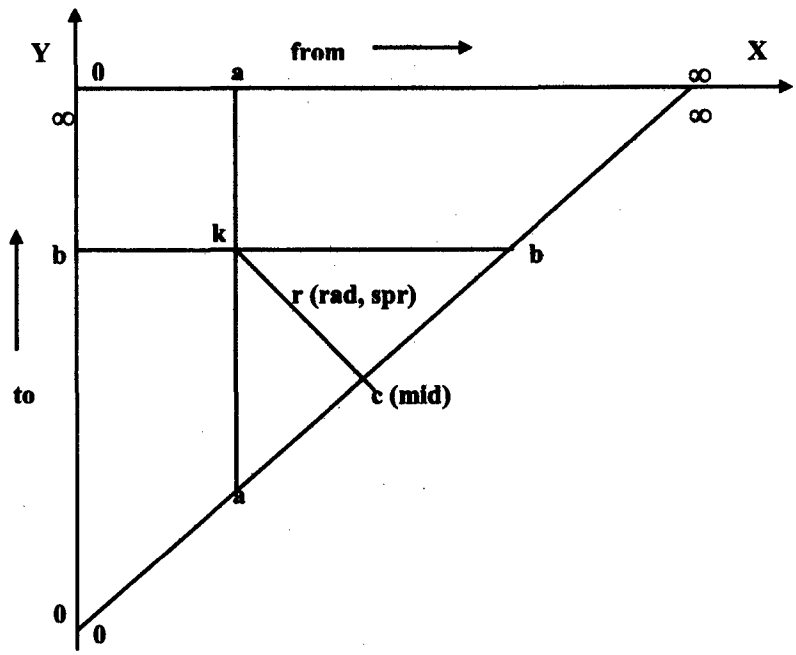


Figure 2.2. Unique display of a closed interval $[a, b] \subseteq \Re^+$ as a point $k = (a, b)$ also given in terms of center(c)-radius(r) (c, r)) (Smets, 2005).

The radius $r = (b-a)/2$ (Figure 2.2) is the distance from the point $k = (a, b)$ to its perpendicular projection, the center $c = (b+a)/2$, on the diagonal domain $\Re^+$. Thus, the center forms the best approximation of the closed interval $\boldsymbol{k} = [a, b]$, with the radius as a measure of imprecision/uncertainty. The closed interval $\boldsymbol{k} = [a, b]$ is then represented as a pair (c, r) in the set

$$\mathcal{V} = \{(c, r) : c, r \in \Re^+, r \geq 0\} \tag{2.44}$$

The relationship between the pair (l, u) in the set $\mathcal{T}$ and the pair c, r in the set $\mathcal{V}$ is given as

$$\begin{pmatrix} r \\ r \end{pmatrix} = \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} b \\ a \end{pmatrix} \tag{2.45}$$

Non-empty elements of $\mathcal{I}$ are in one-to-one correspondence with those of $\mathcal{T}$ and $\mathcal{V}$. In particular, using $\mathcal{V}$ representation, it is possible to explore the distribution of uncertainty around the interval central value (center) and weight the influence of extreme interval factors according to their relative importance to center.

Characterization of the interval using center-radius (c-r) coordinate variables is

valid. It has been underlined in the literature (for instance, Kulpa (2001)) that the center-radius representation has advantages in various applications and theoretical considerations.

- first, the (c, r) representation allows developed interval-based models to take into account a central aspect of the random intervals (the center), as well as their imprecision (radius). In principle, the true value is close to the center and the radius is the imprecision of the true value.

- second, applying the concept of distribution of order statistics on random intervals, the center has the closest distribution to the interval data distribution (Nader et al., 2004). Therefore, the center contains more information relative to other interval features. Besides, the degree of dependence between consecutive interval feature values is maximum and symmetrical about the median. Therefore large interval radius (measure of imprecision) means decreased dependence among interval feature values.

- third, correlation conclusions for interval-based models, using (c, r) interval representation, leads to same correlational conclusions to those based on separate "r"−regression and "c"−regression, (Gil et al., 2001, 2002). On the contrary, conclusions of interval-based model applying (l, u) representation are not coherent with those based on separate "l"−regression and "u"−regression.

2.4.3 The Spatial Closed Random Interval Approach

Naturally, spatial data can constitute interval-valued observations, where observed regionalized variable consist of sample of values indicating a given level of uncertainty. Moore (1979) emphasizes that "uncertainty" in initial data would be a more accurate description for many applications than "error" in initial data. The uncertainty is also a indication of lack of full knowledge of phenomena under study. Uncertainty in spatial data can be coded in form of inequalities constraints, with actual data $Z(\boldsymbol{x})$ at location $\boldsymbol{x}$ lying between lower $(Z(\boldsymbol{x})^l)$ and upper $(Z(\boldsymbol{x})^u)$ limit values such that $Z(\boldsymbol{x})^l \leq Z(\boldsymbol{x}) \leq Z(\boldsymbol{x})^u$. Journel (1986) refers to such data as "soft data" with given prior probabilities and implements a soft kriging approach based on indicator simulation. In contrast single valued data is referred to as "hard data". Chiles and Delfiner (1999) notes that, the information carried by inequality depends on the tightness of the bounds, the wider the bounds the higher the uncertainty.

The data plus the uncertainty can then be associated with the concept of **closed random intervals**, (Congar, 2000; Dempster, 1968). The spatial random interval, $Z(\boldsymbol{x})$, is treated as spatial random vector: $\left(Z(\boldsymbol{x})^l, Z(\boldsymbol{x})^u\right)$ or $\left(Z(\boldsymbol{x})^c, Z(\boldsymbol{x})^r\right)$. This implies that $Z(\boldsymbol{x})$ is a spatial bivariate random variable taking values in the two dimensional plane (Figure 2.2). For the purpose of this study, spatial data is classified as:

***Precise data*:** if the spatial measurements are real-valued (called *hard data* in geostatistical literature). This implies that $Z(\boldsymbol{x})^l = Z(\boldsymbol{x})^u = Z(\boldsymbol{x})$. Then, available information can be modeled by a unique (additive) probability measure as defined in probability and statistics theory. It is also important to note that any uncertainty in this case usually refers to future data; those that have been observed have a exact or precise value.

***Imprecise or uncertain* :** if the spatial measurement are not precise, meaning they cannot be represented by a single value, but by a set or an interval of values. Besides random error, data is plagued with additional uncertainty due to vague information or lack of knowledge, introducing a qualitative aspect to spatial data. We will refer to the such data as *imprecise (or fuzzy)* spatial data or imprecise(fuzzy) regionalized variables characterized by uncertainty which comprises of both randomness and lack of knowledge (vagueness).

An example of interval data is the concentration of SO_2 pollutant reported as possible range of values, mineral reserves given as ranging within a given set of values. Such data can be presented by spatial random intervals/vectors, defined as $\left(Z(\boldsymbol{x})^l, Z(\boldsymbol{x})^u\right)$ or $\left(Z(\boldsymbol{x})^c, Z(\boldsymbol{x})^r\right)$. The width of the intervals of values defines the degree of uncertainty (accuracy) and may change from one sample to another. Xavier (2003), Dubrule and Kostov (1986), Kostov and Dubrule (1986) and Journel (1986) refers to interval data as inequality constraints (soft data).

Sometimes, the target variable may not be observable, but some connected variables are directly measurable. Using physical laws, observed measurements could be converted into a set of possible values, coded as intervals, for the variable of interest.

For the sake of clarity, for this study, the terms "fuzziness", "uncertainty" and "imprecision" will be used interchangeably and the meaning should be derived from the context.

Similar to precise data modeling with a unique (additive) probability measure, the challenge is to define a unique additive measure, within distributional framework, to model the imprecise (or fuzzy) spatial data represented as spatial random intervals/vectors.

We also emphasize that the study considers error fuzziness in measurements. Errors in observed locations (coordinates) is beyond the scope of this study

2.5 Discussion

In this chapter we have reviewed the basics of spatial random process. Further, utilizing the sound theoretical framework of closed random intervals, we represent random intervals as bivariate random vectors characterized by either lower(l)-upper(u) or center(c)-radius(r) coordinate variables. Specifically, based on c-r coordinates, a closed interval is transformed to its optimal approximation (center(c)) plus a variable measure of interval uncertainty the radius(r). Treating intervals as bivariate vectors we are able to apply linear algorithms in modeling random interval related data. In particular, we propose parametric (kriging) and non-parametric (kernel interval regression) approaches to model spatial random intervals.

In addition, the closed random interval theory is linked to the theory of interval analysis (see Moore (1979) for details). This provides us with the necessary operations and properties with respect to intervals, called interval arithmetic (Appendix B).

CHAPTER 3

PARAMETRIC MODELING OF SPATIAL RANDOM INTERVALS

Scope

We devote this chapter to theoretical aspects of parametric modeling of spatial random interval process of second order. Beginning with concepts of a complete interval space, characteristics of random intervals and hence the properties of spatial random interval process together with the corresponding kriging methodology are outlined. Based on a well outlined constraints, we arrive at two approaches of random interval kriging: Composite kriging and component-wise kriging. The chapters also gives credit to work done by Diamond (1988); Dubrule and Kostov (1986); Kostov and Dubrule (1986). Application and comparative results are given in chapter 7.

3.1 Characteristics of Random Intervals

Let $\mathcal{K}(\mathbb{R})$ denoted the collection of all non-empty closed intervals on $\mathbb{R}$, $\mathcal{K}_c(\mathbb{R})$ denoted the collection of all non-empty compact intervals on $\mathbb{R}$, and $\mathcal{K}_{cx}(\mathbb{R})$ the collection of all non-empty convex compact intervals on $\mathbb{R}$. Associated with the space $\mathcal{K}_c(\mathbb{R})$ is a natural non-negative symmetrical function, $d : \mathcal{K}_c(\mathbb{R}) \times \mathcal{K}_c(\mathbb{R}) \rightarrow \mathbb{R}([0, \infty))$. The function d represents a distance metric and is complete if and only if (iff), given any non-empty compact intervals $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathcal{K}_c(\mathbb{R})$, the following conditions

$$\begin{aligned} &i) \quad d(\boldsymbol{a}, \boldsymbol{b}) \leq d(\boldsymbol{a}, \boldsymbol{c}) + d(\boldsymbol{c}, \boldsymbol{b}) \quad \text{(the triangle inequality)} \\ &ii) \quad d(\boldsymbol{a}, \boldsymbol{a}) = 0 \\ &iii) \quad d(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{b}, \boldsymbol{a}) = 0 \quad \text{iff } \boldsymbol{a} = \boldsymbol{b} \end{aligned} \tag{3.1}$$

hold (Ash, 1972; Robert, 2000). The most natural distance d refers to the Hausdorff metric defined in Eq. 3.12. The space $\mathcal{K}_{cx}(\mathbb{R})$ of non-empty compact convex intervals on $\mathbb{R}$ has a semi-linear structure induced by vector (or Minikowiski) addition and scalar multiplication

$$\boldsymbol{a} + \boldsymbol{b} = \{a + b : a \in \boldsymbol{a}, b \in \boldsymbol{b}\}, \quad \lambda \boldsymbol{a} = \{\lambda a : a \in \boldsymbol{a}, \ \lambda \in \mathbb{R}\} \tag{3.2}$$

However, $\mathcal{K}_{cx}(\mathbb{R})$ is not a vector space. To define the properties of its elements we need to translate it into a vector structure with well defined algebraic and metric structure. This is possible using support function embedding discussed in §3.1.1

3.1.1 Support Function and Steiner Point of Convex Intervals

Let d denote the dimension of a Euclidean space, $\mathbb{R}^d$. The set $\mathcal{K}_{cx}(\mathbb{R}^d)$ denotes the collection of all non-empty compact convex sets on $\mathbb{R}^d$. Any closed convex random set $A \in \mathcal{K}_{cx}(\mathbb{R}^d)$ can be uniquely defined by a support function mapping (Körner, 1995; Vitale, 1988),

$$S_A(v) = \sup_{a \in A} \langle a, v \rangle; \qquad a \in S^{d-1} \tag{3.3}$$

where S^{d-1} is the $(d-1)$ dimensional unit sphere, v is a unit vector in $\mathbb{R}^d$, and $\langle a, v \rangle$ is the scalar inner product in $\mathbb{R}^d$. The support function of convex set measures the extent of the convex set in the direction of the unit vector $\boldsymbol{v}$ (Figure 3.1).

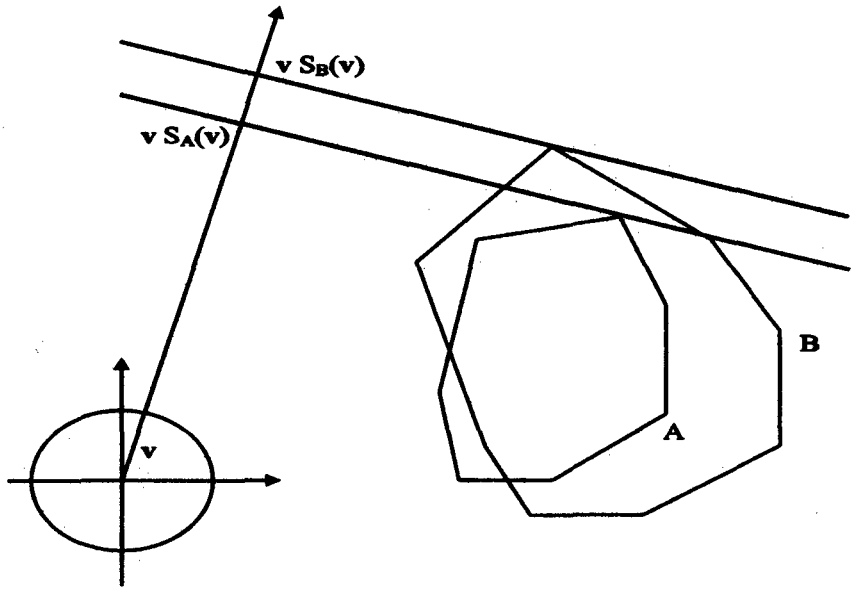


Figure 3.1. The support functions $S_A(\mathbf{v})$, $S_B(\mathbf{v})$ of two convex bodies $A; B$

The support function preserves the semi-linear structure in Eq. 3.2

$$S_{\alpha A+\beta B} = \alpha S_A + \beta S_B \quad \text{for } \alpha, \beta \geq 0. \tag{3.4}$$

and metric structure: $d(A, B) = d(S_A, S_B)$. Thus, linear operations with convex random sets corresponds to point-wise operations on the support functions which are easier to compute. Further, a random convex set (or a random interval) by a single element (value), usually the Steiner point (Körner, 1995; Meyer, 1970). Justified as follows:

Let r denote the radius of a ball, $\mathbb{B}^d(o, r)$, centered at the origin, in a $d-$dimensional space $\mathbb{R}^d$. For a constant

$$c_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)} \tag{3.5}$$

we denote by

$$\text{Vol}\left(\mathbb{B}^d(o,r)\right) = c_d r^d, \text{ and Area } \left(\mathbb{B}^d(o,r)\right) = dc_d r^{d-1},$$

the volume and the surface area of the $d-$dimensional ball, $\mathbb{B}^d(o, r)$, respectively. For unit ball, i.e. $r = 1$, Vol $\left(\mathbb{B}^d(o,1)\right) = c_d$ and Area $\left(\mathbb{B}^d(o,r)\right) = dc_d$.

The constant c_d increases up to $d = 5$ and then decreases toward zero (Table 3.1). Hence, no matter how large the radius, r, increasing the dimension, d, will ultimately produce a sphere of arbitrarily small measure. Therefore, representation

Table 3.1. Constant c_d representing the unit volume of a $d-$dimensional sphere.

Dimension (d)	Constant c_d	Value
0	1	
1	2	$= 2.00000$
2	π	$= 3.14159$
3	$\frac{4}{3}\pi$	$= 4.11879.....$
4	$\frac{\pi^2}{2}$	$= 4.93480.....$
5	$\frac{8\pi^2}{5}$	$= 5.26379....$
6	$\frac{\pi^3}{6}$	$= 5.16771....$
7	$\frac{16\pi^3}{105}$	$= 4.72477....$
8	$\frac{\pi^4}{24}$	$= 4.05871....$
9	$\frac{32\pi^4}{945}$	$= 3.29850....$
$\vdots$	$\vdots$	$\vdots$
2k	$\frac{\pi^k}{k!}$	$\rightarrow 0$ with $k = d/2$

of any convex body will depend more on the surface (boundary values, vertices) of the convex body rather than the interior values. Hence a function of the boundary values gives the single optimal value representation of the convex body. An example of such a value is the Steiner point:

Definition 3.1.1 *The "Steiner point",* $\mathrm{St}(A)$, *of a convex set* A *is*

$$\mathrm{St}(A) = \frac{1}{\mathrm{Vol}(\mathbb{B}^d)} \int_{S^{d-1}} v S_A(v) w(dv) \tag{3.6}$$

with w *the normalized measure on elements of* S^{d-1}. *The Steiner point corresponds to a Lipschitz-continuous mapping,* $\mathrm{St} : (\mathcal{K}(\mathbb{R}^d), d_H) \rightarrow (\mathbb{R}^d, ||\cdot||)$ *such that*

$$||\mathrm{St}(A) - \mathrm{St}(B)|| \leq c\, d_H(A, B), \tag{3.7}$$

where d_H *is the Hausdorff metric, (Eq. 3.12), and* c *is the Lipschitz constant, (Helmut et al., 1994, 1997). The constant* c *is a modification of 3.5 given as*

$$c = \frac{2\Gamma(d/2+1)}{\pi^{1/2}\Gamma(d/2+1/2)} \tag{3.8}$$

The metric d_H can be replaced by normed metrics such as δ_2, (Eq. 3.16) or the weighted metric d_G, (Eq. 3.17). Since $d_G(\boldsymbol{a}, \boldsymbol{b}) \leq \delta_2(\boldsymbol{a}, \boldsymbol{b}) \leq \delta_H(\boldsymbol{a}, \boldsymbol{b})$, the normed metrics give better representation of similarity between intervals, by transferring the properties of an L_2 space to the interval space.

Steiner point preserves the convex arithmetic (Butnariu et al., 2004). Darinka (1998) and Maxwell (2003) confirms the Steiner point as the special generalized selector (or reference point) as given by Congar (2000) and Vitale (1988) for a convex body. Vetterlein and Navara (2006) refers to the Steiner point as a defuzzification method.

Remark 3.1.1 In general, let r be a positive real valued random variable with $E[r^2] < \infty$ and $\boldsymbol{m}$ be a random vector in $\mathbb{R}^d$ (not necessarily independent from r) with $E||\boldsymbol{m}||_2^2 < \infty$. Let $b(\boldsymbol{m}, r)$ denote a ball centered at point $\boldsymbol{m}$, with radius r. Then A with

$$A = b(\boldsymbol{m}, r) = \boldsymbol{m} + (r \times b(\mathrm{o}, 1)) \tag{3.9}$$

is a random convex set with $\boldsymbol{m}$ the Steiner point.

Example 3.1.1 *Let $d = 1$, then $\mathcal{K}_{cx}(\mathbb{R})$, is the set of convex random intervals. Let $\boldsymbol{b} = [b^l, b^u]$, be an element of $\mathcal{K}_{cx}(\mathbb{R})$ with support function*

$$S_{\boldsymbol{b}}(v) = \begin{cases} b^l & \text{if } v = -1 \\ b^u & \text{if } v = 1 \end{cases} \tag{3.10}$$

Let w be a normalized uniform measures on S^{d-1}, $\left(w\left(S^{d-1}\right) = 1/2\right)$. Using 3.6 and 3.10, we have $\mathrm{St}(\boldsymbol{b}) = \left(b^l + b^u\right)/2$, such that $\mathrm{St}(\boldsymbol{b}) \in \boldsymbol{b}$ (Helmut et al., 1994).

Respectively, let $b^c = \mathrm{St}(\boldsymbol{b})$ and $b^r = (b^u - b^l)/2$ denote the center and radius. Applying 3.9,

$$\boldsymbol{b} = b^c + (b^r \times b(\mathrm{o}, 1)) = b^c + [-b^r, b^r] = [b^c - b^r, b^c + b^r] \tag{3.11}$$

This partitions the interval into a part that represents the single value and a part that represents the ambiguity around the single value. A parallel representation of convex intervals is given by Smets (2005) and Kulpa (2001).

It follows that, using the support function concept, we can embed the interval space $\mathcal{K}_{cx}(\mathbb{R})$ into the function space $L_2(S^{d-1})$, for $d = 1$, and results in the function (Hilbert) space $L_2(S^{d-1})$ are transferred into the space of non-empty compact convex intervals, $\mathcal{K}_{cx}(\mathbb{R})$. Then, each convex compact interval is identified with a special Hilbert space-valued random variable (the random support function). Since the independence of random support functions, (Näther, 2000), is well defined, it is not necessary to develop a special concept of independence for non-empty compact convex intervals.

Concepts for expectation and variance of compact convex intervals are then derived from the corresponding well defined notions for Hilbert space-valued random variables. The completeness and separability properties of the Hilbert space ensures that it is either isomorphic to $d-$dimensional Euclidean space, $\mathbb{R}^d$, for some finite d or to the L_2 space.

In §3.1.2 we provide a review interval metrics and propose a weighted interval metric that will be used in this dissertation.

3.1.2 Metrics Between Closed Random Intervals

The Hausdorff metric

Let $\boldsymbol{a}, \boldsymbol{b}$ be non-empty compact intervals in $\mathcal{K}_c(\mathbb{R})$ for $d = 1$

$$d_H(\boldsymbol{a}, \boldsymbol{b}) = \max \left\{ \sup_{a \in \boldsymbol{a}} \inf_{b \in \boldsymbol{b}} |a - b|, \sup_{b \in \boldsymbol{b}} \inf_{a \in \boldsymbol{a}} |a - b| \right\} \tag{3.12}$$

The closed and bounded property in $\mathcal{K}_c(\mathbb{R})$, insures that $d_H(\boldsymbol{a}, \boldsymbol{b}) = 0$ implies $\boldsymbol{a} = \boldsymbol{b}$. The space $(\mathcal{K}_c(\mathbb{R}), d_H)$ is a complete separable metric space. Separable means that $\mathcal{K}_c(\mathbb{R})$ contains a countable dense interval $\mathcal{V}$, such that, the closure of $\mathcal{V}$ is $\mathcal{K}_c(\mathbb{R})$, while complete means that $\mathcal{K}_c(\mathbb{R})$ contains d_H-convergent intervals: there exist a real interval $\boldsymbol{a}_0 \in \mathcal{K}_c(\mathbb{R})$ such that every sequence of intervals $\{\boldsymbol{a}_n\}$ converges to $\boldsymbol{a}_0$ i.e.,

$$\lim_{n(h)} d_H(\boldsymbol{a}_n, \boldsymbol{a}_0) = 0 \quad \text{for sequences of intervals} \quad (\boldsymbol{a}_n)_{n=1}^{\infty} \in \mathcal{K}_c(\mathbb{R}).$$

For point intervals, $d_H(\boldsymbol{a}, \boldsymbol{b})$ reduces to usual distance $d([a, a], [b, b]) = \sqrt{|a - b|^2}$ for real numbers.

The δ_2-metric

Let $\boldsymbol{a}, \boldsymbol{b}$ be non-empty compact convex intervals in $\mathcal{K}_{cx}(\mathbb{R})$ for $d = 1$. Using support function, the non-empty compact intervals are mapped into the Hilbert space with corresponding norm given as

$$||\boldsymbol{a}||_2 = \sqrt{\int_{S^{d-1}} |S_{\boldsymbol{a}}(v)|^2 w(dv)} \tag{3.13}$$

where w is the normalized $(d-1)-$dimensional Lebesgue measure on elements of unit sphere S^{d-1}, $(w(S^{d-1}) = 1)$.

The interval metric in $\mathcal{K}_{cx}(\mathbb{R})$ is equivalent to the Hilbert distance between support functions, and is given as

$$\delta_2(\boldsymbol{a},\boldsymbol{b}) = \delta_2(S_{\boldsymbol{a}}, S_{\boldsymbol{b}}) = ||S_{\boldsymbol{a}} - S_{\boldsymbol{b}}||_2 = \sqrt{\int_{S^{d-1}} |S_{\boldsymbol{a}}(v) - S_{\boldsymbol{b}}(v)|^2 w(dv)} \tag{3.14}$$

Since $d = 1$, $S^{d-1} = \{-1, 1\}$, and the integral becomes summation over S^{d-1}, and the distance 3.14 becomes

$$\begin{aligned} \delta_2(\boldsymbol{a},\boldsymbol{b}) &= \sqrt{\sum_{v\in\{-1,1\}} |S_{\boldsymbol{a}}(v) - S_{\boldsymbol{b}}(v)|^2 w(v)} \\ &= \sqrt{\big(S_{\boldsymbol{a}}(-v) - S_{\boldsymbol{b}}(-v)\big)^2 w(-v) + \big(S_{\boldsymbol{a}}(v) - S_{\boldsymbol{b}}(v)\big)^2 w(v)} \end{aligned}$$

With a uniform weight measure, $w(v) = w(-v) = 0.5$, and applying 3.10,

$$\delta_2(\boldsymbol{a},\boldsymbol{b}) = \sqrt{.5\big[(a^l - b^l)^2 + (a^u - b^u)^2\big]} \tag{3.15}$$

Let $a^c = (a^u + a^l)/2$ and $a^r = (a^u - a^l)/2$ denote the center and radius coordinates, accordingly, $a^u = a^c + a^r$ and $a^l = a^c - a^r$ and 3.15 reduces to

$$\delta_2(\boldsymbol{a},\boldsymbol{b}) = \sqrt{(a^c - b^c)^2 + (a^r - b^r)^2} \tag{3.16}$$

which is similar to Körner (1995) metric.

The δ_2 metric is less sensitive to any transformations of the non-empty compact intervals compared to the Hausdorff metric, and the space $(\mathcal{K}_{cx}(\mathbb{R}), \delta_2)$ is complete and separable (Diamond and Kloeden, 1990, 1994).

The d_W-metric

This is a generalization of the Hausdorff metric. Let $\boldsymbol{a}, \boldsymbol{b}$ be non-empty compact convex intervals in $\mathcal{K}_{cx}(\mathbb{R})$,the d_W-metric defined as

$$d_W(\boldsymbol{a},\boldsymbol{b}) = \sqrt{\int_{[0,1]} [\lambda(a^u - b^u) + 1 - \lambda(a^l - b^l)]^2 dW(\lambda)}$$

where $W(\lambda), \forall\lambda \in [0, 1]$, is a normalized symmetrical weight measure or a probability measure on the measurable space, $([0, 1], \mathcal{B}_{[0,1]})$, with $\mathcal{B}_{[0,1]}$ the Borel σ-field on

$[0,1]$. The weights $W(\lambda)$ indicate the contribution of interval features to the interval distance.

By fixing $W(0) = W(1) = 0.5$ for $\lambda = \{0,1\}$, and $W(\lambda) = 0$ for all $\lambda \in (0,1)$, the d_W-metric reduces to δ_2 metric (Diamond, 1988; D'Urso and Gastaldi, 2000). Lubiano (1999) refers to the $W-$metric as the $S-$mean squared dispersion. Bertoluzza and Cariolaro (1997) gave a special case of d_W-metric for $\lambda = W(0) = W(0.5) = W(1)$.

Proposed Metric- The General Weighted d_G- interval metric

Reviewed metrics, are similar and assigns equal weight to squared Euclidean distances between the interval features. In reality, we would expect the Steiner center to contain more information while the spread is an indication of interval imprecision or ambiguity. There is a need to define some measure of relative importance to each value within an interval. This allows us to weight each interval feature by its mutual information within the interval, which will have an influence on the defined distance measure. This new metric reflects the near and overlap degree between two intervals; it is a simplified inclusion of all the (co)variances.

Theorem 3.1.1 The general interval metric is defined as

$$d_G^2(\boldsymbol{a},\boldsymbol{b}) = (a^c - b^c)^2 + \frac{\alpha}{\alpha+\beta}(a^r - b^r)^2 \tag{3.17}$$

where $\boldsymbol{a},\boldsymbol{b}$ are non-empty compact convex intervals in $\mathcal{K}_{cx}(\mathbb{R})$ with $\boldsymbol{a} = (a^c, a^r)$ and $\boldsymbol{b} = (b^c, b^r)$ using 2.44; α and β are the parameters of a Beta probability density,

$$g(\lambda) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1-\lambda)^{\beta-1} \tag{3.18}$$

Proof. Let $\delta_2^2(\boldsymbol{a},\boldsymbol{b})$ denote the metric (3.16), and $\lambda \in (0,1)$. Further, let $([0,1], \mathcal{B}([0,1]), G)$ denote a probability space, with G a normalized weight on the measurable space, $([0,1], \mathcal{B}_{[0,1]})$, where $\mathcal{B}_{[0,1]}$ is the Borel σ-field on $[0,1]$. We define a weighted interval metric as

$$d_G(\boldsymbol{a},\boldsymbol{b}) = \sqrt{\int_0^1 [\lambda\delta_2^2(\boldsymbol{a},\boldsymbol{b}) + (1-\lambda)(a^c - b^c)^2] dG(\lambda)}.$$

Thus $d_G(\boldsymbol{a},\boldsymbol{b})$ is a weighted distance metric. The weights $\lambda \in (0,1)$, reflects the contribution of the δ_2-metric and of the squared Euclidean difference between in-

terval midpoint to the distance. Instead of fixing the weights, we let λ be distributed continuously within $(0,1)$ based on a continuous probability density function. Because of its flexibility, the Beta probability density function; a continuous and bounded probability density function within $(0,1)$ and parameterized by parameters $(\alpha, \beta > 1)$ is ideally suited for the choice of λ.

Applying 3.18, $dG(\lambda) = g(\lambda)d\lambda$, and substituting 3.16 for δ^2,

$$\begin{aligned} d_G^2(\boldsymbol{a},\boldsymbol{b}) &= \int_0^1 \Big[\lambda\Big((a^c-b^c)^2+(a^r-b^r)^2\Big) \\ &\quad +(1-\lambda)(a^c-b^c)^2\Big]\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1-\lambda)^{\beta-1}d\lambda \\ &= \int_0^1 \Big[(a^c-b^c)^2+\lambda(a^r-b^r)^2\Big]\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1-\lambda)^{\beta-1}d\lambda \\ &= (a^c-b^c)^2+(a^r-b^r)^2\int_0^1 \lambda\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1-\lambda)^{\beta-1}d\lambda \\ &= (a^c-b^c)^2+(a^r-b^r)^2\frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)} \\ &= (a^c-b^c)^2+\frac{\alpha}{\alpha+\beta}(a^r-b^r)^2 \end{aligned}$$

■

Theorem 3.1.2 The generalized metric d_G^2 between any intervals is a complete metric in the interval space $\mathcal{K}_{cx}(\mathbb{R})$.

Proof. For metric $d_G^2(\boldsymbol{a},\boldsymbol{b})$ to be a complete metric, it has to satisfy the metric properties defined in 3.1. It is easy to verify property (ii) and (iii). For the triangle inequality property (i) let $\boldsymbol{a},\boldsymbol{b},\boldsymbol{e} \in \mathcal{K}_{cx}(\mathbb{R})$,

$$d_G^2(\boldsymbol{a},\boldsymbol{b}) \leq d_G^2(\boldsymbol{a},\boldsymbol{e}) + d_G^2(\boldsymbol{e},\boldsymbol{b}) \tag{3.19}$$

For the triangular property of Euclidean distance for interval center and for interval radius, we can assert that:

$$(a^c-b^c)^2 \leq (a^c-e^c)^2+(e^c-b^c)^2 \quad \text{and} \quad (a^r-b^r)^2 \leq (a^r-e^r)^2+(e^r-b^r)^2$$

Then, it follows that:

$$\begin{aligned} (a^c-b^c)^2+\frac{\alpha}{\alpha+\beta}(a^r-b^r)^2 \;\leq\; &\Big((a^c-e^c)^2+\frac{\alpha}{\alpha+\beta}(a^r-e^r)^2\Big) \\ &+\Big((e^c-b^c)^2+\frac{\alpha}{\alpha+\beta}(e^r-b^r)^2\Big), \end{aligned}$$

implying that

$$d_G^2(\boldsymbol{a},\boldsymbol{b}) \leq d_G^2(\boldsymbol{a},\boldsymbol{e}) + d_G^2(\boldsymbol{e},\boldsymbol{b}). \tag{3.20}$$

When applied to point intervals, $d_G^2(\boldsymbol{a},\boldsymbol{b})$ reduces to usual distance $d([a,a],[b,b]) = |a-b|$ for real numbers.

Since, the space $(\mathcal{K}_{cx}(\mathbb{R}), d_G(\cdot))$ is a complete and separable metric space, it makes sense to talk of independently distributed random intervals/vectors, (Klement et al., 1986). ■

Comparing the metrics

Let

$$\begin{array}{llll} \boldsymbol{a}_1 = [a_1^l, a_1^u] = [1,2], & a_1^c = \dfrac{3}{2}, & a_1^r = \dfrac{1}{2} & \boldsymbol{b}_1 = [b_1^l, b_1^u] = [0,3], \quad b_1^c = \dfrac{3}{2}, \quad b_1^r = \dfrac{3}{2} \\ \boldsymbol{a}_2 = [a_2^l, a_2^u] = [0,2], & a_2^c = 1, & a_2^r = 1 & \boldsymbol{b}_2 = [b_2^l, b_2^u] = [1,3], \quad b_2^c = 2, \quad b_2^r = 1 \end{array}$$

then

$$\begin{aligned} d_G^2(\boldsymbol{a}_1,\boldsymbol{b}_1) &= (a_1^c - b_1^c)^2 + \frac{\alpha}{\alpha+\beta}(a_1^r - b_1^r)^2 \\ &= 0 + \frac{\alpha}{\alpha+\beta} \times 1 \\ &= \frac{\alpha}{\alpha+\beta} \end{aligned}$$

while

$$\begin{aligned} \delta_2^2(\boldsymbol{a}_1,\boldsymbol{b}_1) &= (a_1^c - b_1^c)^2 + (a_1^r - b_1^r)^2 \\ &= 1 \end{aligned}$$

and

$$\begin{aligned} d_H(\boldsymbol{a}_1,\boldsymbol{b}_1) &= \max\left\{ \sup_{a\in\boldsymbol{a}_1} \inf_{b\in\boldsymbol{b}_1} |a-b|, \inf_{a\in\boldsymbol{a}_1} \sup_{b\in\boldsymbol{b}_1} |a-b| \right\} \\ &= \max\left\{(a_1^u - b_1^l), (a_1^l - b_1^u)\right\} \\ &= \max\left\{(2,-2)\right\} = 2 \end{aligned}$$

Clearly, for any α and β, $\frac{\alpha}{\alpha+\beta}$ approaches one, thus

$$d_G^2(\boldsymbol{a}_1,\boldsymbol{b}_1) \leq \delta_2^2(\boldsymbol{a}_1,\boldsymbol{b}_1) \leq d_H(\boldsymbol{a}_1,\boldsymbol{b}_1) \tag{3.21}$$

This is also true distance between $\boldsymbol{a}_2$ and $\boldsymbol{b}_2$. The weighted metric the d_G metric is a better option as it allows us to penalize extreme interval features, giving more weight to interval features close to the center. Using the weighted metric defined in Eq. 3.17, the properties (expectation and (co)variance) of closed convex interval are explored next.

3.1.3 Expectation and Variance of Random Intervals

Let d denote a metric in a any given metric space denoted as (M, d). Under Frechets principle, expectation EX of a random variable X taking values in (M, d) is the solution to the following problem

$$Ed^2(X, EX) = \inf_{p \in M} Ed^2(X, p) \tag{3.22}$$

where $Ed^2(X, p)$ is the usual expectation of the real-valued variable $d^2(X, p)$. The Frechets variance of X defined as

$$\mathbf{Var}\, X = Ed^2(X, EX) \tag{3.23}$$

is a generalization of the known fact that for a real-valued variable X the expectation EX minimizes $E|X - p|^2$, and $\mathrm{Var}\, X = E|X - EX|^2$. Frechets variance, parallels Stein (1999) variance of random elements in a Hilbert space. A parallel emphasis can be found in Lyashenko (1983).

Since the Frechet expectation is a nonlinear operator, it makes it difficult to calculate and an obstacle in obtaining the variances. An expectation with better linear properties is the Aumann (1965) expectation (3.24). Based on a general L_2 metric, the Aumann-expectation parallels Frechets-expectation (Körner, 1995; Körner and Näther, 2001; Näther, 2000), which leads to:

Definition 3.1.2 Let (Ω, Σ, Pr) be a probability space, an interval-valued function $\boldsymbol{a} : \Omega \to \mathcal{K}_{cx}(\mathbb{R})$ is a interval-valued random variable. The Aumann (1965) expectation of $\boldsymbol{a}$ is given as

$$E[\boldsymbol{a}] \;=\; \{E\boldsymbol{a}' : \boldsymbol{a}'(\omega) \in \boldsymbol{a}(\omega)\ a.s.(Pr),\ \text{and } \boldsymbol{a}' \text{ is integrable}\} \tag{3.24}$$

where $E\boldsymbol{a}'$ is the usual expectation of random vectors $\boldsymbol{a}' \in \mathbb{R}$ such that $\boldsymbol{a}' \subseteq \boldsymbol{a}$ almost surely (a.s) in probability. $\boldsymbol{a}'$ is called a selection of $\boldsymbol{a}$. For example, Congar (2000) defines $\boldsymbol{a}' = (a^l, a^u)$ which is the largest selection of $\boldsymbol{a}$. If $\boldsymbol{a}$ is $\mathcal{K}_{cx}(\mathbb{R})$−valued

and if $E[\boldsymbol{a}] < \infty$, then $E[\boldsymbol{a}] \in \mathcal{K}_{cx}(\mathbb{R})$.

Alternatively, taking expectation of Eq. 3.11 for interval variable $\boldsymbol{a}$,

$$E[\boldsymbol{a}] = \left[E\left(a^c - a^r\right), E\left(a^c + a^r\right)\right] = \left[Ea^l, Ea^u\right]$$

In vector notation, either $\boldsymbol{a} = \left(a^l, a^u\right)$ or $\boldsymbol{a} = \left(a^c, a^r\right)$. Then, the Aumann expectation can be written as

$$E[\boldsymbol{a}] = \left(Ea^l, Ea^u\right) \quad \text{or} \quad \left(Ea^c, Ea^r\right) \tag{3.25}$$

Further, Vitale (1988) has given Aumann' expectation as Boucher expectation of the corresponding support function i.e. $S_{E\boldsymbol{a}}(v) = E_{S_{\boldsymbol{a}}}(v)$.

Contrary to Frechet's expectation, Aumman's expectation is linear with respect to minikowiski addition $\oplus$ i.e., for $\boldsymbol{b}$ an interval-valued random variable

$$E(\lambda_1 \boldsymbol{a} \oplus \lambda_2 \boldsymbol{b}) = \lambda_1 E\boldsymbol{a} \oplus \lambda_2 E\boldsymbol{b}$$

Remark 3.1.2 *The Aumman expectation and associated variance can both be estimated in a similar way as in classical sense. Körner (1995); Körner and Näther (2001); Näther (2000) justified Aumann-expectation as a Frechets-expectation with respect to the general L_2-distance metric, δ_2. Besides, for quantifying error and variance associated with random interval $\boldsymbol{a}$, the expected value of $\boldsymbol{a}$ is the Frechets expectation of $\boldsymbol{a}$ with respect to δ_2, but not δ_H (Lubiano, 1999). Therefore, (co)-variance of random intervals can be defined using the δ_2-metric or a generalized δ_2 metric, for example the d_G metric.*

Let E denote Aumman's expectation. Assume a general L_2-metric d, and an interval-valued random variable $\boldsymbol{a}$, if $E[\boldsymbol{a}] < \infty$, and $E||\boldsymbol{a}||_d^2 < \infty$, then the Frechet's variance of $\boldsymbol{a}$, is

$$\mathbf{Var}\, \boldsymbol{a} = Ed^2(\boldsymbol{a}, E\boldsymbol{a}) \tag{3.26}$$

In principle, the variance of random intervals can be deduced from a more general point of view using the notion of variance for random support functions in a Hilbert space. Let, $p = 1$ denote the dimension, the variance of random interval $\boldsymbol{a}$ if it exists is given as

$$\begin{aligned} \mathrm{Var}\,(\boldsymbol{a}) &= Ed^2\left(\boldsymbol{a}, E\boldsymbol{a}\right) = Ed^2\left(S_{\boldsymbol{a}}(v), S_{E(\boldsymbol{a})}(v)\right) \\ &= \int_{S^{p-1}} E\left(S_{\boldsymbol{a}}(v) - S_{E(\boldsymbol{a})}(v)\right)^2 w dv \end{aligned} \tag{3.27}$$

where v is the unit vector taking values in unit sphere $S^{p-1} = \{-1, 1\}$, and w is Lebesgue measure on unit sphere space $(S^{p-1}, \mathcal{B}_{S^{p-1}}, w)$. Since $S^{p-1} = \{-1, 1\}$ the integration reduces to summation and variance 3.27 becomes

$$\text{Var}\,(\boldsymbol{a}) = \sum_{v\in\{-1,1\}} E\left(S_{\boldsymbol{a}}(v) - S_{E(\boldsymbol{a})}(v)\right)^2 wdv \tag{3.28}$$

Using standard arguments from random sets theory by Stoyan and Molchanov (1997), $S_{E\boldsymbol{a}}(v) = ES_{\boldsymbol{a}}(v)$, and 3.28 reduces to

$$\begin{aligned}\text{Var}\,(\boldsymbol{a}) &= \sum_{v\in\{-1,1\}} E\left(S_{\boldsymbol{a}}(v) - ES_{\boldsymbol{a}}(v)\right)^2 wdv &(3.29)\\ &= \sum_{v\in\{-1,1\}} \text{Var}\, S_{\boldsymbol{a}}(v)wdv &(3.30)\end{aligned}$$

which presents the connection between the Var $(\boldsymbol{a})$ and the classical defined Var $S_{\boldsymbol{a}}(v)$. We use for both the same notation since a random number is a special random interval. Given two random intervals $\boldsymbol{a}, \boldsymbol{b}$, the corresponding scalar product in the Hilbert $(L_2(S^{p-1}))$ space is given as

$$\langle \boldsymbol{a}, \boldsymbol{b}\rangle = \sum_{v\in\{-1,1\}} S_{\boldsymbol{a}}(v)S_{\boldsymbol{b}}(v)wdv \tag{3.31}$$

Expanding 3.29, and using 3.31, we can rewrite the variance as

$$\begin{aligned}\text{Var}\,(\boldsymbol{a}) &= \sum_{v\in\{-1,1\}} E\left[S_{\boldsymbol{a}}(v)S_{\boldsymbol{a}}(v) - 2S_{\boldsymbol{a}}(v)ES_{\boldsymbol{a}}(v) + ES_{\boldsymbol{a}}(v)ES_{\boldsymbol{a}}(v)\right]wdv\\ &= \sum_{v\in\{-1,1\}} ES_{\boldsymbol{a}}(v)S_{\boldsymbol{a}}(v)wdv - \sum_{v\in\{-1,1\}} ES_{\boldsymbol{a}}(v)ES_{\boldsymbol{a}}(v)wdv\\ &= E\sum_{v\in\{-1,1\}} S_{\boldsymbol{a}}(v)S_{\boldsymbol{a}}(v)wdv - \sum_{v\in\{-1,1\}} ES_{\boldsymbol{a}}(v)ES_{\boldsymbol{a}}(v)wdv\\ &= E\langle \boldsymbol{a}, \boldsymbol{a}\rangle - \langle E\boldsymbol{a}, E\boldsymbol{a}\rangle &(3.32)\end{aligned}$$

Note that

$$\text{Var}\,(\boldsymbol{a}) = \text{Cov}\,(\boldsymbol{a}, \boldsymbol{a}) = E\left(\boldsymbol{a} - E\boldsymbol{a}\right)(\boldsymbol{a} - E\boldsymbol{a})) = E(\boldsymbol{a} - E\boldsymbol{a})^2 = Ed^2(\boldsymbol{a}, E\boldsymbol{a}) \tag{3.33}$$

is basically the expectation of the squared distance between $\boldsymbol{a}$ and its mean $E\boldsymbol{a}$.

Extending 3.29 to the case of two random intervals $\boldsymbol{a}$ and $\boldsymbol{b}$,the covariance becomes

$$\begin{aligned}
\text{Cov}(\boldsymbol{a},\boldsymbol{b}) &= \sum_{v\in\{-1,1\}} E\left[\left(S_{\boldsymbol{a}}(v)-ES_{\boldsymbol{a}}(v)\right)\left(S_{\boldsymbol{b}}(v)-ES_{\boldsymbol{b}}(v)\right)\right] wdv \qquad (3.34)\\
&= \sum_{v\in\{-1,1\}} \left(ES_{\boldsymbol{a}}(v)S_{\boldsymbol{b}}(v)-ES_{\boldsymbol{a}}(v)ES_{\boldsymbol{b}}(v)\right) wdv\\
&= E\left(\sum_{v\in\{-1,1\}} S_{\boldsymbol{a}}(v)S_{\boldsymbol{b}}(v)wdv\right) - \sum_{v\in\{-1,1\}} ES_{\boldsymbol{a}}(v)ES_{\boldsymbol{b}}(v)wdv\\
&= E\langle \boldsymbol{a},\boldsymbol{b}\rangle - \langle E\boldsymbol{a}, E\boldsymbol{b}\rangle \qquad \text{using 3.32} \qquad (3.35)
\end{aligned}$$

The correlation coefficient denoted by $\rho(\boldsymbol{a},\boldsymbol{b})$, is defined as

$$\rho(\boldsymbol{a},\boldsymbol{b}) = \frac{\text{Cov}(\boldsymbol{a},\boldsymbol{b})}{\sqrt{\text{Var}(\boldsymbol{a})\text{Var}(\boldsymbol{b})}} \qquad (3.36)$$

Example 3.1.2 Explicitly, given random intervals $\boldsymbol{a}_i(\omega) = [a_i^l(\omega), a_i^u(\omega)]$ such that a_i^l, a_i^u are integrable random variables with $a_i^l(\omega) \le a_i^u(\omega)$ for all $\omega \in \Omega$, $i = 1,2$. Aumann expectation, $E\boldsymbol{a}_i(\omega) = [Ea_i^l, Ea_i^u]$. Assuming the δ_2 interval metric, and given that $S^{p-1} = \{-1,1\}$ where -1 and 1 respectively denote the interval lower and upper values,

$$\begin{aligned}
\text{Var}(\boldsymbol{a}_i) &= E\left[\left(a_i^l(\omega)-E(a_i^l)\right)^2 + \left(a_i^u(\omega)-E(a_i^u)\right)^2\right] = \text{Var}(a_i^l) + \text{Var}(a_i^u)\\
&\qquad\qquad \text{using 3.29}\\
\text{Cov}(\boldsymbol{a}_1,\boldsymbol{a}_2) &= E\left[\left(a_1^l(\omega)-E(a_1^l)\right)\left(a_2^l(\omega)-E(a_2^l)\right)\right.\\
&\quad \left. + \left(a_1^u(\omega)-E(a_1^u)\right)\left(a_2^u(\omega)-E(a_2^u)\right)\right] \qquad \text{using 3.34}\\
&= \text{Cov}(a_1^l, a_2^l) + \text{Cov}(a_1^u, a_2^u)
\end{aligned}$$

$$\rho(\boldsymbol{a}_1,\boldsymbol{a}_2) = \frac{\text{Cov}(a_1^l, a_2^l) + \text{Cov}(a_1^u, a_2^u)}{\left(\sqrt{\text{Var}(a_1^l)+\text{Var}(a_1^u)}\cdot\sqrt{\text{Var}(a_2^l)+\text{Var}(a_2^u)}\right)}$$

In general, for any two random intervals $\boldsymbol{a},\boldsymbol{b}$, and using Eq. 3.25, $E\boldsymbol{a} = (Ea^c, Ea^r)$ or $E\boldsymbol{b} = (Eb^c, Eb^r)$. Using the general weighted interval metric d_G,

if $Ed_G^2(\boldsymbol{a}, \{0\}) = E||\boldsymbol{a}||_{d_G}^2 < \infty$, then, the variance of $\boldsymbol{a}$

$$\begin{aligned}
\text{Var}(\boldsymbol{a}) &= Ed_G^2(\boldsymbol{a}, E\boldsymbol{a}) \\
&= E\left[[a^c - Ea^c]^2 + \frac{\alpha}{\alpha+\beta}[a^r - Ea^r]^2\right] \qquad \text{(using theorem 3.1.1)} \\
&= E[a^c - Ea^c]^2 + \frac{\alpha}{\alpha+\beta}E[a^r - Ea^r]^2 \\
&= \text{Var}(a^c) + \frac{\alpha}{\alpha+\beta}\text{Var}(a^r) \qquad (3.37)
\end{aligned}$$

This is an extension of the Körner (1995) definition. Similarly, using the d_G metric and Eq. 3.37, the covariance between $\boldsymbol{a} = (a^c, a^r)$ and $\boldsymbol{b} = (b^c, b^r)$ is

$$\begin{aligned}
\text{Cov}(\boldsymbol{a}, \boldsymbol{b}) &= E\left[(a^c - E(a^c))(b^c - E(b^c)) + \frac{\alpha}{\alpha+\beta}(a^r - E(a^r))(b^r - E(b^r))\right] \\
&= E[(a^c - E(a^c))(b^c - E(b^c))] + \frac{\alpha}{\alpha+\beta}[(a^r - E(a^r))(b^r - E(b^r))] \\
&= \text{Cov}(a^c, b^c) + \frac{\alpha}{\alpha+\beta}\text{Cov}(a^r, b^r) \qquad (3.38)
\end{aligned}$$

As a generalized L_2-metric, the d_G metric induces additive variance, $\text{Var}^{d_G}(\boldsymbol{a}_1 \oplus \boldsymbol{a}_2) = \text{Var}^{d_G}(\boldsymbol{a}_1) \oplus \text{Var}^{d_G}(\boldsymbol{a}_2)$, for independent random intervals, contrary to the non-normed generalized Hausdorff metric d_W. Therefore, variance of sum of independent random intervals is similar to variance of real-valued random variable. We note that, while the variance is real-valued the expectation is interval-valued.

Remark 3.1.3 Having defined random interval variables of second order, i.e. random intervals with existing expectation and variance, we proceed to define the spatial interval random function characterized by regionalized random intervals of second order, taking values in $(\mathcal{K}_{cx}(\mathbb{R}), d_G)$. The space $\mathcal{K}_{cx}(\mathbb{R})$ is isomorphic to two-dimensional half plane $(\Re^+)$, with either the interval endpoint coordinate system or the midpoint-radius coordinate system. Though, the most straightforward coordinate system is the interval endpoints, it has certain drawbacks (Gil et al., 2001, 2002; Kulpa, 2001) as discussed in §2.4.2. This makes it awkward for practical uses. For our case we will use intervals represented via center-radius coordinate system.

Next applying the above concepts to spatial imprecise data we define a spatial random interval process and proceed to define corresponding kriging approaches.

3.2 Spatial Interval Random Function

3.2.1 Properties

The spatial random interval function is characterized by second order regionalized random intervals $\boldsymbol{Z}(\boldsymbol{x}_i) = (Z(\boldsymbol{x}_i)^l, Z(\boldsymbol{x}_i)^u)$, $(i = 1, \ldots, n)$ observed at locations $\boldsymbol{x}_i$ over the spatial region $D \subseteq \mathbb{R}^d$. Based on the center-radius coordinates $\boldsymbol{Z}(\boldsymbol{x}_i) = \Big(Z^c(\boldsymbol{x}_i), Z^r(\boldsymbol{x}_i)\Big)$, and using 2.45, we can revert to $\boldsymbol{Z}(\boldsymbol{x}_i) = \Big(Z^l(\boldsymbol{x}_i), Z^u(\boldsymbol{x}_i)\Big)$.

The second order spatial random interval function is given as

$$\boldsymbol{Z}(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x}) + \boldsymbol{e}(\boldsymbol{x}) \tag{3.39}$$

with $\boldsymbol{Z}(\boldsymbol{x}) = \Big(Z^c(\boldsymbol{x}), Z^r(\boldsymbol{x})\Big)$ the regionalized random interval, $\boldsymbol{m}(\boldsymbol{x}) = \Big(m_c(\boldsymbol{x}),\ m_r(\boldsymbol{x})\Big)$ the mean, and $\boldsymbol{e}(\boldsymbol{x}) = \Big(e^c(\boldsymbol{x}), e^r(\boldsymbol{x})\Big)$ the zero mean random function that captures the erratic fluctuations of the process. A particular concern is the dependence or independence between Z^c and Z^r, or Z^l and Z^u. Note, due to the concept of support function embedding of intervals to a positive cone and hence the independence of support function Näther (2000), we proceed by not making assumption about the pairwise dependence or independence between Z^c and Z^r, or Z^l and Z^u.

Let $\boldsymbol{h}$ or h denote the spatial distance between two locations. Both $\boldsymbol{h}$ or h will be used interchangeably. Parallel to real-valued spatial process, the isotropic interval covariance is a function of spatial distance h and is given as

$$\begin{aligned} \boldsymbol{C}(h) &= E[\boldsymbol{Z}(\boldsymbol{x})\boldsymbol{Z}(\boldsymbol{x}+h)] - E[\boldsymbol{Z}(\boldsymbol{x})]E[\boldsymbol{Z}(\boldsymbol{x}+h)] \\ &= E[\boldsymbol{Z}(\boldsymbol{x})\boldsymbol{Z}(\boldsymbol{x}+h)] - \boldsymbol{m}(\boldsymbol{x})\boldsymbol{m}(\boldsymbol{x}+h) \end{aligned} \tag{3.40}$$

If the mean function is a constant, i.e. $\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x}+h) = \boldsymbol{m}$, as is the case with simple and ordinary kriging (with local neighborhood), interval covariance reduces to

$$\boldsymbol{C}(h) = E[\boldsymbol{Z}(\boldsymbol{x})\boldsymbol{Z}(\boldsymbol{x}+h)] - \boldsymbol{m}^2 \text{ with } \boldsymbol{m}^2 = (m_c^2, m_r^2) \tag{3.41}$$

Separately, the center and radius covariances are

$$\begin{aligned} C^c(h) &= E[Z^c(\boldsymbol{x}+h), Z^c(\boldsymbol{x})] - m_c^2 \\ C^r(h) &= E[Z^r(\boldsymbol{x}+h), Z^r(\boldsymbol{x})] - m_r^2, \end{aligned} \tag{3.42}$$

respectively. For $h = 0$, interval covariance 3.41 becomes

$$\boldsymbol{C}(0) = E[\boldsymbol{Z}(\boldsymbol{x})\boldsymbol{Z}(\boldsymbol{x})] - \boldsymbol{m}^2 \tag{3.43}$$

and

$$\begin{aligned} C^c(0) &= E[Z^c(\boldsymbol{x}), Z^c(\boldsymbol{x})] - m_c^2 \\ C^r(0) &= E[Z^r(\boldsymbol{x}), Z^r(\boldsymbol{x})] - m_r^2 \end{aligned} \tag{3.44}$$

Application of d_G norm, helps in obtaining a tractable form of interval covariance. Let $k_{\alpha,\beta} = \alpha/(\alpha + \beta)$; using the general d_G metric, the variance of regionalized random interval $\boldsymbol{Z}(\boldsymbol{x})$ with expectation $E[\boldsymbol{Z}(\boldsymbol{x})] = \boldsymbol{m}$ is defined as

$$\begin{aligned} \mathrm{Var}(\boldsymbol{Z}(\boldsymbol{x}))(=\boldsymbol{C}(0)) &= E d_G^2[\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{m}] \\ &= E\left\{[Z^c(\boldsymbol{x}) - m_c]^2 + k_{\alpha,\beta}[Z^r(\boldsymbol{x}) - m_r]^2\right\} \\ & \qquad\qquad \text{applying 3.17} \\ &= E[Z^c(\boldsymbol{x}) - m_c]^2 + k_{\alpha,\beta} E[Z^r(\boldsymbol{x}) - m_r]^2 \\ &= E\left[Z^c(\boldsymbol{x})^2\right] - m_c^2 + k_{\alpha,\beta}\left(E\left[Z^r(\boldsymbol{x})^2\right] - m_r^2\right) \\ &= C^c(0) + k_{\alpha,\beta} C^r(0) \qquad \text{using 3.44} \end{aligned} \tag{3.45}$$

The variance is real-valued, the sum of the center and radius covariances. It is positive definite since it is a sum of covariances of center and radius. Similarly, extending to the case where $h \neq 0$,

$$\boldsymbol{C}(h) = C^c(h) + k_{\alpha,\beta} C^r(h) \tag{3.46}$$

Hence, the interval covariance 3.41 reduces to the sum of center and radius covariances at distance lag h.

Assuming stationarity, the expected value of the d_G weighted distance between

regionalized random intervals $\boldsymbol{Z}(\boldsymbol{x}+h)$ and $\boldsymbol{Z}(\boldsymbol{x})$ is

$$\begin{aligned} Ed_G^2[\boldsymbol{Z}(\boldsymbol{x}+h),\boldsymbol{Z}(\boldsymbol{x})] &= E\Big[\big(Z^c(\boldsymbol{x}+h)-Z^c(\boldsymbol{x})\big)^2\Big] \\ &\quad +k_{\alpha,\beta}E\Big[\big(Z^r(\boldsymbol{x}+h)-Z^r(\boldsymbol{x})\big)^2\Big] \\ &= EZ^c(\boldsymbol{x}+h)^2-2EZ^c(\boldsymbol{x}+h)Z^c(\boldsymbol{x})+EZ^c(\boldsymbol{x})^2 \\ &\quad +k_{\alpha,\beta}\left(EZ^r(\boldsymbol{x}+h)^2-2EZ^r(\boldsymbol{x}+h)Z^r(\boldsymbol{x})+EZ^r(\boldsymbol{x})^2\right) \\ &= \left[C^c(0)+m_c^2\right]-2\left[C^c(h)+m_c^2\right]+\left[C^c(0)+m_c^2\right] \\ &\quad +k_{\alpha,\beta}\Big(\left[C^r(0)+m_r^2\right]-2\left[C^r(h)+m_r^2\right]+\left[C^r(0)+m_r^2\right]\Big) \\ &\qquad\qquad \text{applying 3.42 and 3.44} \\ &= 2\Big[C^c(0)-C^c(h)+k_{\alpha,\beta}\left(C^r(0)-C^r(h)\right)\Big] \qquad (3.47) \end{aligned}$$

Then,

$$\begin{aligned} \frac{1}{2}Ed_G^2[\boldsymbol{Z}(\boldsymbol{x}+h),\boldsymbol{Z}(\boldsymbol{x})] &= \left[C^c(0)+k_{\alpha,\beta}C^r(0)\right]-\left[C^c(h)+k_{\alpha,\beta}C^r(h)\right]\Big] \\ &= \boldsymbol{C}(0)-\boldsymbol{C}(h) \qquad \text{using 3.45 and 3.46} \qquad (3.48) \end{aligned}$$

The original definition of classical variogram in chapter 2 Eq. 2.8 can be easily extended to random interval process. Thus, the interval variogram can be written as

$$\boldsymbol{\Gamma}(h) = \frac{1}{2}E(\boldsymbol{Z}(\boldsymbol{x}+h)-\boldsymbol{Z}(\boldsymbol{x}))^2, \qquad (3.49)$$

and equals half the expectation of squared distance between regionalized random intervals $\boldsymbol{Z}(\boldsymbol{x})$ and $\boldsymbol{Z}(\boldsymbol{x}+h)$. Assuming the given distance to be d_G, interval variogram 3.49 reduces to

$$\boldsymbol{\Gamma}(h) = \frac{1}{2}Ed_G^2(\boldsymbol{Z}(\boldsymbol{x}),\boldsymbol{Z}(\boldsymbol{x}+h)=\boldsymbol{C}(0)-\boldsymbol{C}(h) \qquad (3.50)$$

which parallels the well defined relation 2.12 (Goovaerts, 1997).

3.2.2 Variance of Linear Combination of Spatial Random Intervals

Let λ_i denote weights associated with each spatial random interval $\boldsymbol{Z}(\boldsymbol{x}_i)$. For a given set of neighborhood points $n(h)$,

$$\boldsymbol{Z} = \sum_{i=1}^{n(h)} \lambda_i \boldsymbol{Z}(\boldsymbol{x}_i) \tag{3.51}$$

denotes the linear combination of observed spatial random intervals in the neighborhood $n(h)$. For simplicity of notation let, $\boldsymbol{Z}_i$ denote the spatial random interval at location $\boldsymbol{x}_i$ i.e. $\boldsymbol{Z}_i = \boldsymbol{Z}(\boldsymbol{x}_i)$.

Prior to defining the variance of $\boldsymbol{Z}$, we make the following assumptions about the (co)variance of the radius data; the measure of random interval imprecision ("ambiguity" or "uncertainty"):

$$(i) \quad \text{The radius data has zero correlation structure.} \tag{3.52}$$

$$(ii) \quad \text{The radius data has full correlation structure.} \tag{3.53}$$

Based on the two assumptions, we arrive at two different variances of random interval.

3.2.2.1 The Composite Variance

Let $\boldsymbol{Z}_i = (Z_i^c, Z_i^r)$. Since $Z_i^r > 0$, the radius data weights must be positive. Cressie (1993) notes that although non-negative weights imply that $Z_i^r > 0$, it is a constraint that is too heavy-handed unless there are good reasons to use it. Note that $\boldsymbol{Z} = \left(\sum_{i=1}^{n(h)} \lambda_i Z_i^c, \sum_{i=1}^{n(h)} \lambda_i Z_i^r \right)$ can still be valid, $\left(\text{i.e. } \sum_{i=1}^{n(h)} \lambda_i Z_i^c \geq 0 \text{ and } \sum_{i=1}^{n(h)} \lambda_i Z_i^r \geq 0 \right)$ but still have negative weights. Szidarovszky et al. (1987) suggested a version of kriging with only positive weights. While kriging with positive weights ensures that predicted values lie within the minimum and maximum values of observed values, the associated cost is highly inflated kriging standard errors (Schabenberger and Gotway, 2005, pp 232). According to Diamond (1988), if the process exhibits a high nugget effect, negative weights are automatically eliminated. If of utmost importance to eliminate weights, we propose the use of Deutsch (1996) approach.

The expectation,

$$E\boldsymbol{Z} = \sum_{i=1}^{n(h)} \lambda_i \left(EZ_i^c, EZ_i^r\right) = \sum_{i=1}^{n(h)} \lambda_i \left(m_c, m_r\right) = \sum_{i=1}^{n(h)} \lambda_i \boldsymbol{m}.$$

Then,

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{Z}) &= E\left[d_G^2[\boldsymbol{Z}, E\boldsymbol{Z}]\right] \\
&= E\Big[d_G^2\Big(\sum_{i=1}^{n(h)} \lambda_i \boldsymbol{Z}_i, \sum_{i=1}^{n(h)} \lambda_i \boldsymbol{m}\Big)\Big] = E\Big[d_G^2\Big(\sum_{i=1}^{n(h)} \lambda_i \left(Z_i^c, Z_i^r\right), \sum_{i=1}^{n(h)} \lambda_i \left(m_c, m_r\right)\Big)\Big] \\
&= E\Bigg\{\Big(\sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j (Z_i^c - m_c)(Z_j^c - m_c)\Big) \\
&\qquad + k_{\alpha,\beta}\Big(\sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j (Z_i^r - m_r)(Z_j^r - m_r)\Big)\Bigg\} \qquad \text{(using 3.17)} \\
&= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j \big(E[Z_i^c Z_j^c] - m_c^2\big) + k_{\alpha,\beta}\Big\{\sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j \big(E[Z_i^r Z_j^r] - m_r^2\big)\Big\} \\
&\qquad \text{(after expanding and taking expectation)} \\
&= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j C_{ij}^c + k_{\alpha,\beta} \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j C_{ij}^r \qquad (3.54)
\end{aligned}$$

Let

$$\boldsymbol{C}_{ij} = C_{ij}^c + k_{\alpha,\beta} C_{ij}^r. \qquad (3.55)$$

Based on the first assumption 3.52, C_{ij}^r reduces to a diagonal covariance matrix, and the variance, 3.54 becomes

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{Z}) &= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j \left[C_{ij}^c + k_{\alpha,\beta} C_{ij}^r\right] \\
&= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i\lambda_j \boldsymbol{C}_{ij} \qquad \text{using 3.55.} \qquad (3.56)
\end{aligned}$$

with weights associated to center variable only. This assumption leads to "Composite kriging" approach discussed in §3.3.1. It is the classical BLUP- problem for linear regression with the random interval observations related by $\boldsymbol{C}_{ij}$ conditional on weights summing to one i.e., $\sum_{i=1}^{n(h)} \lambda_i = 1$ based on the center values only. An

additional constraint $\lambda_i \geq 1$ may be introduced when dealing with a special case that requires weights to be non-negative, but not without cost of high prediction errors (Schabenberger and Gotway, 2005).

3.2.2.2 The Component-wise Variance

Respectively, we let λ_i^c and λ_i^r denotes the center and radius weights. For $\boldsymbol{Z}_i = (Z_i^c, Z_i^r)$, Eq. 3.51 reduces to;

$$\boldsymbol{Z} = \sum_{i=1}^{n(h)} (\lambda_i^c Z_i^c, \lambda_i^r Z_i^r) \text{ with a mean } E\boldsymbol{Z} = \sum_{i=1}^{n(h)} (\lambda_i^c m_c, \lambda_i^r m_r), \ \lambda_i^r \geq 0. \tag{3.57}$$

Then, applying the second assumption, 3.53,

$$\begin{aligned}
\text{Var}(\boldsymbol{Z}) = E\big[d_G^2[\boldsymbol{Z}, E\boldsymbol{Z}]\big] &= E\Big[d_G^2\big[\sum_{i=1}^{n(h)} (\lambda_i^c Z_i^c, \lambda_i^r Z_i^r), \sum_{i=1}^{n(h)} (\lambda_i^c m_c, \lambda_i^r m_r)\big]\Big] \\
&= E\Big[\Big(\sum_{i=1}^{n(h)} \lambda_i^c (Z_i^c - m_c)\Big)^2 + k_{\alpha,\beta}\Big(\sum_{i=1}^{n(h)} \lambda_i^r (Z_i^r - m_r)\Big)^2\Big] \\
&\qquad\qquad \text{(using 3.17)} \\
&= \sum_{i,j=1}^{n(h)} \lambda_i^c \lambda_j^c \Big(E\big[Z_i^c Z_j^c\big] - m_c^2\Big) \\
&\quad + k_{\alpha,\beta} \sum_{i,j=1}^{n(h)} \lambda_i^r \lambda_j^r \Big(E\big[Z_i^r Z_j^r\big] - m_r^2\Big) \\
&= \sum_{i,j=1}^{n(h)} \lambda_i^c \lambda_j^c C_{ij}^c + k_{\alpha,\beta} \sum_{i,j=1}^{n(h)} \lambda_i^r \lambda_j^r C_{ij}^r
\end{aligned} \tag{3.58}$$

The approach leads to "Component-wise kriging" in §3.3.2. The center weights are estimated using the center data conditional on $\sum_{i=1}^{n(h)} \lambda_i^c = 1$ while the radius weights are estimates using the radius data only conditional on $\sum_{i=1}^{n(h)} \lambda_i^r = 1$ and $\lambda_i^r \geq 0$.

3.3 Interval Ordinary Kriging (OIK)

Let $\boldsymbol{x} = (x_1, x_2)^T$ denote the 2-dimensional spatial location and $\boldsymbol{Z}_i$ the observed regionalized random intervals at locations $\boldsymbol{x}_i$, $(i = 1, \ldots, n)$ over a spatial region

$D \subseteq \mathbb{R}^2$. Extending 3.39, the spatial random interval process is

$$\boldsymbol{Z}(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x}) + \boldsymbol{e}(\boldsymbol{x})$$

where $\boldsymbol{Z}(\boldsymbol{x}) = \left(Z^c(\boldsymbol{x}), Z^r(\boldsymbol{x})\right)$ is the regionalized random field, $\boldsymbol{m}(\boldsymbol{x}) = \left(m_c(\boldsymbol{x}),\ m_r(\boldsymbol{x})\right)$ is the mean, and $\boldsymbol{e}(\boldsymbol{x}) = \left(e^c(\boldsymbol{x}), e^r(\boldsymbol{x})\right)$ is spatial error such that $E\boldsymbol{e}(\boldsymbol{x}) = 0$ and $\mathrm{Cov}(\boldsymbol{e}(\boldsymbol{x}_i), \boldsymbol{e}(\boldsymbol{x}_j)) = \sigma^2\phi(||\boldsymbol{x}_i - \boldsymbol{x}_j||; \boldsymbol{\theta})$. The parameter $\boldsymbol{\theta}$ controls the range of spatial association.

Ordinary kriging assumes a non-stationary random process model where stationarity is limited within each search neighborhood. Prediction involves minimizing the error variance of the estimator of spatial random interval process with respect to the d_G-metric.

Normally, spatial measurements are made on finite volumes, rather than points (x, y) in space. The finite volumes, called "sample support" can be a line in 1D, an area in 2D, or volume in 3D. Let $\boldsymbol{z}(\boldsymbol{x})$ be the regionalized random interval, a realization of spatial random interval field $\boldsymbol{Z}$. Consider

$$\boldsymbol{s}_0 = \frac{1}{|D|}\int_d \boldsymbol{z}(\boldsymbol{x}')d\boldsymbol{x}', \quad \boldsymbol{x}' \in D$$

as the average of $\boldsymbol{z}(\boldsymbol{x})$, over a given finite support D. Then, $\boldsymbol{s}_0$ is a realization of interval random field $\boldsymbol{S}_0$ approximated by

$$\boldsymbol{S}_0^* = \sum_{i=1}^{n(h)} \boldsymbol{\lambda}_i \boldsymbol{Z}_i$$

using the interval random field $\{\boldsymbol{Z}_i : i = 1, \ldots, n\}$. Then, based on minimum variance prediction, $\boldsymbol{S}_0 \approx \boldsymbol{S}_0^*$. If Q denotes a set of functions or estimators, then $\boldsymbol{S}_0^* \in Q$ such that

$$d_G^2(\boldsymbol{S}_0^*, \boldsymbol{S}_0) = \inf_{\boldsymbol{S}' \in Q} \{d_G^2(\boldsymbol{S}', \boldsymbol{S}_0)\}$$

is the regression function of the data with respect to the interval norm d_G. The estimator $\boldsymbol{S}_0^*$ is the best approximating function (in terms of minimum variance prediction) in Q of the data set $\boldsymbol{Z}_i$. For example, in general least square regression with d_G norm,

$$d_G^2(\boldsymbol{S}_0^*, \boldsymbol{S}_0) = (S_0^{c*} - S_0^c)^2 + k_{\alpha,\beta}\,(S_0^{r*} - S_0^r)^2 \text{ using 3.17}$$

The same principle applies to ordinary kriging predictor, and it is in-fact a best linear unbiased predictor, BLUP (Cressie, 1993). BLUP means that the estimator is a linear, unbiased and optimal in the sense of minimum squared error (mse). The goal is to obtain weights λ_i such that

$$(i) \quad E(\boldsymbol{S}_0^*) = E\left(\sum_{i=1}^{n(h)} \boldsymbol{\lambda}_i \boldsymbol{Z}_i\right) = \sum_{i=1}^{n(h)} \boldsymbol{\lambda}_i E\boldsymbol{Z}_i = \boldsymbol{m}$$

unbiasedness

$$(ii) \quad \text{Var}\, \boldsymbol{S}_0^* = E d_G^2(\boldsymbol{S}_0^*, \boldsymbol{S}_0) \quad \text{is minimized under a set of constraints}$$

In ordinary interval kriging, condition (i), equivalent to the constraint $\sum_{i=1}^{n(h)} \lambda_i = 1$, secures the unbiased estimator and we need to minimize (ii), the variance of the estimator. Hence, the goal is to minimize

$$\begin{aligned} E d_G^2(\boldsymbol{S}_0^*, \boldsymbol{S}_0) &= E\Big[(S_0^{c*} - S_0^c)^2 + k_{\alpha,\beta}(S_0^{r*} - S_0^r)^2\Big] \\ &= ES_0^{c*2} - 2ES_0^{c*}S_0^c + ES_0^{c2} \\ &\quad + k_{\alpha,\beta}\big[ES_0^{r*2} - 2ES_0^{r*}S_0^r + ES_0^{r2}\big] \end{aligned} \tag{3.59}$$

3.3.1 Composite OIK System

Let $\boldsymbol{x}, \boldsymbol{x}' \in D$, we define

$$\begin{aligned} C_{DD}^c &= \frac{1}{|D|^2} \int_d \int_d C^c(\boldsymbol{x} - \boldsymbol{x}')\, d\boldsymbol{x}\, d\boldsymbol{x}' \quad \text{and} \\ C_{i0}^c &= \frac{1}{|D|^2} \int_d \int_d C^c(\boldsymbol{x}_i - \boldsymbol{x}') d\boldsymbol{x}_i d\boldsymbol{x} \end{aligned} \tag{3.60}$$

and similarly for C_{DD}^r and C_{i0}^r.

Then, using 3.42, terms in 3.59 can be expressed as follows,

$$ES_0^{c2} = C_{DD}^c + m_c^2 \quad \text{and} \quad ES_0^{r2} = C_{DD}^r + m_r^2 \tag{3.61}$$

$$E(S_0^{c*}, S_0^c) = \sum_{i=1}^{n(h)} \lambda_i C_{iD}^c + m_c^2 \quad \text{and} \quad E(S_0^{r*}, S_0^r) = \sum_{i=1}^{n(h)} \lambda_i C_{iD}^r + m_r^2 \tag{3.62}$$

Also,

$$ES_0^{c*2} = \sum_{i=1}^{n(h)} \sum_{j=1}^{n(h)} \lambda_i \lambda_j C_{ij}^c + m_c^2 \qquad ES_0^{r*2} = \sum_{i=1}^{n(h)} \sum_{j=1}^{n(h)} \lambda_i \lambda_j C_{ij}^r + m_r^2 \tag{3.63}$$

For expressions 3.61 and 3.63 to be valid, $\lambda_i \in \mathbb{R}$ for ES_0^{c*2}, and $\lambda_i \geq 0$ for ES_0^{r*2}, i.e. weights can take on values on the extended real space: $[-\infty \cup \mathbb{R}]) \cup (\mathbb{R} \cup +\infty]$. The positive definiteness condition of covariances implies that the weights can be uniquely obtained.
Applying 3.55

$$\begin{aligned} \boldsymbol{C}_{ij} = &\quad C^c_{ij} + k_{\alpha,\beta} C^r_{ij} \quad \text{covariance of } \boldsymbol{Z}_i, \boldsymbol{Z}_j \\ \boldsymbol{C}_{iD} = &\quad C^c_{iD} + k_{\alpha,\beta} C^r_{iD} \quad \text{covariance of } \boldsymbol{Z}_i, \boldsymbol{Z}_D \\ \boldsymbol{C}_{DD} = &\quad C^c_{DD} + k_{\alpha,\beta} C^r_{DD} \quad \text{covariance of } \boldsymbol{Z}_D, \boldsymbol{Z}_D \end{aligned} \tag{3.64}$$

Substituting 3.61, 3.62 and 3.63 in 3.59, and using 3.64,the means cancel out and

$$\begin{aligned} Ed^2_G(\boldsymbol{S}^*_0, \boldsymbol{S}_0) &= \sum_{i,j=1}^{n(h)} \lambda_i \lambda_j C^c_{ij} - 2 \sum_{i=1}^{n(h)} \lambda_i C^c_{iD} + C^c_{DD} \\ &\quad + k_{\alpha,\beta} \Big[\sum_{ij=1}^{n(h)} \lambda_i \lambda_j C^r_{ij} - 2 \sum_{i=1}^{n(h)} \lambda_i C^r_{iD} + C^r_{DD} \Big] \\ &= \sum_{i,j=1}^{n(h)} \lambda_i \lambda_j \boldsymbol{C}_{ij} - 2 \sum_{i=1}^{n(h)} \lambda_i \boldsymbol{C}_{iD} + \boldsymbol{C}_{DD} \end{aligned} \tag{3.65}$$

Note that, the first term of 3.65 equals composite variance given in 3.56. Similarly, the second can be defined as the composite covariance between $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_D$ for locations $\boldsymbol{x}' \in D$, while the third term is the composite variance of $\boldsymbol{Z}_D$ for all locations $\boldsymbol{x}' \in D$. Then, the variance of the interval random field $\boldsymbol{S}^*_0$ (3.65), is minimized conditional on (i) i.e. weights sum to one.

Introducing the Lagrange multiplier ψ for condition (i), such that $\sum_i^{n(h)} \lambda_i = 1$, we have

$$\begin{aligned} L(\lambda_i, i = 1, \ldots, n; \psi) &= \sum_{i,j=1}^{n(h)} \lambda_i \lambda_j \boldsymbol{C}_{ij} - 2 \sum_{i=1}^{n(h)} \lambda_i \boldsymbol{C}_{iD} + \boldsymbol{C}_{DD} \\ &\quad + 2 \Big(\psi \sum_{i=1}^{n(h)} \lambda_i - 1 \Big) \end{aligned} \tag{3.66}$$

Minimizing 3.66 with respect to weights and ψ results to **Composite OIK system**:

$$\begin{aligned}\frac{dL}{d\lambda_i} &= \sum_{j=1}^{n(h)} \lambda_j C_{ij} - C_{iD} + \psi = 0 \qquad i = 1, \ldots, n(h) \\ \frac{dL}{d\psi} &= \sum_{i=1}^{n(h)} \lambda_i - 1 = 0 \end{aligned} \tag{3.67}$$

rewritten as

$$\begin{aligned}\sum_{j=1}^{n(h)} \lambda_j C_{ij} + \psi &= C_{iD} \qquad i = 1, \ldots, n(h) \\ \sum_{i=1}^{n(h)} \lambda_i &= 1 \end{aligned} \tag{3.68}$$

The system 3.68 has $(n(h)+1)$ linear equations with $(n(h)+1)$ unknowns (λ's, and ψ)

Let $\boldsymbol{C_{kk}}$ denote an $n(h) \times n(h)$ matrix consisting of sum of center and scaled radius covariances between sample data locations; and $\boldsymbol{C_{kD}}$ denote an $n(h) \times 1$ vector of sum of midpoint and scaled radius covariances between sample locations and prediction data location. Let $\boldsymbol{\lambda}$ denote vectors of weights; $\mathbf{1}' = (1, \ldots, 1)_{1 \times n(h)}$ vector and $p = n(h) + 1$. In matrix notation, 3.68 becomes

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{C_{kk}} & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}_{p \times p} ; \quad \boldsymbol{L} = \begin{pmatrix} \boldsymbol{\lambda} \\ \psi \end{pmatrix}_{p \times 1} ; \quad \boldsymbol{k} = \begin{pmatrix} \boldsymbol{C_{kD}} \\ 1 \end{pmatrix}_{p \times 1} \tag{3.69}$$

We note that the covariance matrices $\boldsymbol{K}$ and $\boldsymbol{k}$ are positive definite since they are sum of positive definite covariances, therefore the optimization problem can be solved to obtain the weights. As is the common practice in geostatistics, the interval variogram is modeled first. Then, using the relation $\boldsymbol{C(h)} = \boldsymbol{C(0)} - \boldsymbol{\Gamma(h)}$ (Isaaks and Srivastava, 1989, p.289), the covariances are generated and substituted back in 3.69.

3.3.2 Component-wise OIK System

The component-wise kriging is based on separate kriging with center data and separate kriging with radius data. Thus, the weights in equation 3.62 change to

$$\begin{aligned} E(S_0^{c*}, S_0^c) &= E\Big(\sum_{i=1}^{n(h)} \lambda_i^c Z_i^c, D)\Big) = \sum_{i=1}^{n(h)} \lambda_i^c E\big(Z_i^c, D\big) \\ &= \sum_{i=1}^{n(h)} \lambda_i^c C_{iD}^c + m_c^2 \\ E(S_0^{r*} S_0^r) &= \sum_{i=1}^{n(h)} \lambda_i^r C_{iD}^r + m_r^2. \end{aligned} \tag{3.70}$$

Similarly, the terms in 3.63 change to

$$\begin{aligned} ES_0^{c*2} &= E\Big(\sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i^c \lambda_j^c Z_i^c Z_j^c\Big) = \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i^c \lambda_j^c E\big(Z_i^c, Z_j^c\big) \\ &= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i^c \lambda_j^c C_{ij}^c + m_c^2 \\ ES_0^{r*2} &= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i^r \lambda_j^r C_{ij}^r + m_r^2 \end{aligned} \tag{3.71}$$

with the constraint $\lambda_i^r, \lambda_j^r \geq 0$.

Substituting 3.61, 3.70 and 3.71 in RHS of 3.59, the means cancel out and we have

$$\begin{aligned} Ed_G(\boldsymbol{S}_0^*, \boldsymbol{S}_0)^2 &= \sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i^c \lambda_j^c C_{ij}^c - 2\sum_{i=1}^{n(h)} \lambda_i^c C_{iD}^c + C_{DD}^c \\ &\quad + k_{\alpha,\beta}\Big[\sum_{i=1}^{n(h)}\sum_{j=1}^{n(h)} \lambda_i^r \lambda_j^r C_{ij}^r - 2\sum_{i=1}^{n(h)} \lambda_i^r C_{iD}^r + C_{DD}^r\Big] \end{aligned} \tag{3.72}$$

Consider the first part of Eq. 3.72: the first term equals the first term in the component-wise variance, 3.58, but for the center of mass Z_i^c values. Similarly the second term is the covariance of Z_i^c and Z_D^c for all locations $\boldsymbol{x}' \in D$, while the third term is the (co)variance of $Z^c(\boldsymbol{x}')$ for all locations $\boldsymbol{x}' \in D$. The same applies to the second part of 3.72 but using the radius data.

Introducing the Lagrange multiplier $\psi = (\psi^c, \psi^r)$,

$$\begin{aligned} L(\lambda_i^c, \lambda_i^r, i = 1, \ldots, n; \psi^c, \psi^r) &= \sum_{i=1}^{n(h)} \sum_{j=1}^{n(h)} \lambda_i^c \lambda_j^c C_{ij}^c - 2 \sum_{i=1}^{n(h)} \lambda_i^c C_{iD}^c + C_{DD}^c \\ &+ k_{\alpha,\beta} \Big[\sum_{i=1}^{n(h)} \sum_{j=1}^{n(h)} \lambda_i^r \lambda_j^r C_{ij}^r - 2 \sum_{i=1}^{n(h)} \lambda_i^r C_{iD}^r + C_{DD}^r \Big] \\ &+ 2\psi^c \Big(\sum_{i=1}^{n(h)} \lambda_i^c - 1 \Big) + 2\psi^r \Big(\sum_{i=1}^{n(h)} \lambda_i^r - 1 \Big) \end{aligned} \tag{3.73}$$

Minimizing 3.73, with respect to weights, ψ^c and $\psi^r \geq 0$ results to

$$\begin{aligned} \frac{dL}{d\lambda_i^c} &= \sum_{j=1}^{n(h)} \lambda_j^c C_{ij}^c - C_{iD}^c + \psi^c = 0 \qquad i = 1, \ldots, n(h) \\ \frac{dL}{d\psi^c} &= \sum_{i=1}^{n(h)} \lambda_i^c - 1 = 0 \\ \frac{dL}{d\lambda_i^r} &= k_{\alpha,\beta} \sum_{j=1}^{n(h)} \lambda_j^r C_{ij}^r - k_{\alpha,\beta} C_{iD}^r + \psi^r = 0 \qquad i = 1, \ldots, n(h) \\ \frac{dL}{d\psi^r} &= \sum_{i=1}^{n(h)} \lambda_i^r - 1 = 0 \end{aligned} \tag{3.74}$$

rewritten as

$$\begin{aligned} \sum_{j=1}^{n(h)} \lambda_j^c C_{ij}^c + \psi^c &= C_{iD}^c \qquad i = 1, \ldots, n(h) \\ \sum_{i=1}^{n(h)} \lambda_i^c &= 1 \end{aligned} \tag{3.75}$$

$$\begin{aligned} k_{\alpha,\beta} \sum_{j=1}^{n(h)} \lambda_j^r C_{ij}^r + \psi^r &= k_{\alpha,\beta} C_{iD}^r \qquad i = 1, \ldots, n(h) \\ \sum_{i=1}^{n(h)} \lambda_i^r &= 1 \end{aligned} \tag{3.76}$$

Now, 3.75 has $(n(h)+1)$ linear equations with $(n(h)+1)$ unknowns (λ^c's, and ψ^c) estimated using the center of mass data only, and 3.76 has $(n(h)+1)$ linear equations with $(n(h)+1)$ unknowns (λ^r's, and ψ^r, with $\lambda^r \geq 0$) using the radius data only.

Respectively, let C_{kk}^c and C_{kk}^r denote the $n(h) \times n(h)$ matrices of center of mass

and radius covariances between sample data locations; and $\boldsymbol{C}^c_{kD}$ and $\boldsymbol{C}^r_{kD}$ denote the $n(h) \times 1$ vectors of center of mass and radius covariances between sample locations and prediction data location. Then, with $p = n(h) + 1$, components of Eq. 3.75 becomes

$$\boldsymbol{V}^c = \begin{pmatrix} \boldsymbol{C}^c_{kk} & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}_{p\times p}; \quad \boldsymbol{L}^c = \begin{pmatrix} \boldsymbol{\lambda}^c \\ \psi^c \end{pmatrix}_{p\times 1}; \quad \boldsymbol{v}^c = \begin{pmatrix} \boldsymbol{C}^c_{kD} \\ 1 \end{pmatrix}_{p\times 1} \tag{3.77}$$

while 3.76 is

$$\boldsymbol{V}^r = \begin{pmatrix} \boldsymbol{C}^r_{kk} & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}_{p\times p}; \quad \boldsymbol{L}^r = \begin{pmatrix} \boldsymbol{\lambda}^r \\ \psi^r \end{pmatrix}_{p\times 1}; \quad \boldsymbol{v}^r = \begin{pmatrix} \boldsymbol{C}^r_{kD} \\ 1 \end{pmatrix}_{p\times 1} \tag{3.78}$$

giving two kriging systems

$$\boldsymbol{V}^c\boldsymbol{L}^c = \boldsymbol{v}^c, \qquad \boldsymbol{V}^r\boldsymbol{L}^r = \boldsymbol{v}^r \tag{3.79}$$

Alternatively, combining 3.77 and 3.78 results to an ordinary kriging system denoted as

$$\boldsymbol{V}\boldsymbol{L} = \boldsymbol{v} \tag{3.80}$$

$$\boldsymbol{V} = \left(\begin{array}{c|c} \boldsymbol{V}^c_{p\times p} & \mathbf{0}_{p\times p} \\ \hline \mathbf{0}_{p\times p} & \boldsymbol{V}^r_{p\times p} \end{array}\right); \quad \boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}^c_{p\times 1} \\ \hline \boldsymbol{L}^r_{p\times 1} \end{pmatrix}; \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}^c_{p\times 1} \\ \hline \boldsymbol{v}^r_{p\times 1} \end{pmatrix} \tag{3.81}$$

Explicitly, we have

$$\boldsymbol{V} = \left(\begin{array}{cc|cc} \boldsymbol{C}^c_{kk} & \mathbf{1} & & \\ \mathbf{1}' & 0 & & \\ \hline & & \boldsymbol{C}^r_{kk} & \mathbf{1} \\ & & \mathbf{1}' & 0 \end{array}\right)_{2p\times 2p}; \quad \boldsymbol{L} = \begin{pmatrix} \boldsymbol{\lambda}^c \\ \psi^c \\ \hline \boldsymbol{\lambda}^r \\ \psi^r \end{pmatrix}_{2p\times 1}; \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{C}^c_{kD} \\ 1 \\ \hline \boldsymbol{C}^r_{kD} \\ 1 \end{pmatrix}_{2p\times 1} \tag{3.82}$$

The system of equations, 3.80 is derived by combining two systems of equations each with $((n(h)+1))$ sets of equations. Thus we can denote Eq. 3.80 as systems of $(2((n(h)+1)))$ equations. Shary (1997) refers to such, as a system of equations in the Euclidean space of double dimension $\mathbb{R}^{2(n)}$. This double dimension OIK system 3.80 consists of $(2n(h)+2)$ linear equations with $(2n(h)+2)$ unknowns (λ^c's, λ^r's ψ^r and ψ^c).

Solving separate systems in 3.79 leads to similar results as those based on the double dimension OIK system 3.80.

3.3.3 Interval Arithmetic OIK System

To define an interval kriging system with interval-valued equations, we set C_{kk} and C_{kD} into degenerate interval matrices, Weights are then obtained by solving the interval system of equations. The derived interval kriging system parallel the kriging system defined by Diamond (1988). While Diamond solved his system using constrained programming, we apply Kaucher extended interval arithmetic described in §3.3.4. In principle, Kaucher arithmetic, translates linear interval systems into non-interval or point equations in Euclidean space of double dimension $\mathbb{R}^{2n}$ (Shary, 1997), and interval solutions are obtained using normal numerical algorithms (Shary, 1997).

3.3.4 Linear Algebraic Equations in Kaucher Arithmetic

Kaucher arithmetic expands the classical interval arithmetic, (Appendix B), to a wider algebraic system with better algebraic properties, providing richer manipulation techniques, making it easier to seek solution in the new wider algebraic system (Shary, 1997, 2002).

Briefly, the elements of Kaucher arithmetic are pairs of real numbers, not necessarily related by condition $x^l \leq x^u$, (the proper intervals). It includes the set $\overline{\mathbb{IR}}$, of improper intervals given as, $[x^u, x^l]$, as well as real numbers identified as degenerate intervals. The proper, $\mathbb{IR}$, and improper $\overline{\mathbb{IR}}$, intervals; the two halves of Kaucher interval space change places as a result of dualization mapping: dual : $\mathbb{IR} \to \overline{\mathbb{IR}}$ such that dual $(\boldsymbol{x}) = [x^u, x^l]$. Kaucher interval space is isomorphic to $\mathbb{R}^2$ and the properties of $\mathbb{R}^2$.

Using Kaucher arithmetic, the solution to a linear interval system,

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{3.83}$$

where

$$\boldsymbol{A} = (\boldsymbol{a}_{ij})_{n\times n}\,,\ \boldsymbol{a}_{ij} = \left[a_{ij}^l, a_{ij}^u\right]\,,\ i,j = 1,2,\cdots,n$$
$$\boldsymbol{b} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n),\ \boldsymbol{b}_k = \left[b_k^l, b_k^u\right],\ k = 1,2,\cdots,n$$
and
$$\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n),\ \boldsymbol{x}_k = \left[x_k^l, x_k^u\right],\ k = 1,2,\cdots,n$$

is reduced to a problem of solving one non-interval (point) system of equations in the Euclidean space of double dimension $\mathbb{R}^{2n}$, (Shary, 1997). Respectively, let $A^u = (a_{ij}^u)_{n\times n}$ and $A^l = (a_{ij}^l)_{n\times n}$ denote the upper matrix and the lower matrix of the interval matrix $\boldsymbol{A}$. To solve the system 3.83, let $\boldsymbol{A}^u$ and $\boldsymbol{A}^l$ respectively denote

the upper and lower $n \times n$ matrices. Then, we define $2n \times 2n$ real non-singular matrices

$$\widetilde{C} \triangleq \begin{bmatrix} \frac{1}{2}\left(A^u + A^l\right) & 0 \\ 0 & \frac{1}{2}\left(A^u + A^l\right) \end{bmatrix}_{2n \times 2n},$$

$$\widetilde{D} \triangleq \begin{bmatrix} \frac{1}{2}\left(A^u - A^l\right) & 0 \\ 0 & \frac{1}{2}\left(A^u - A^l\right) \end{bmatrix}_{2n \times 2n}$$

and $2n \times 1$ real vectors

$$\widetilde{x} \triangleq \begin{bmatrix} x^u \\ -x^l \end{bmatrix}_{2n \times 1} \quad \widetilde{b} \triangleq \begin{bmatrix} b^u \\ -b^l \end{bmatrix}_{2n \times 1}$$

The algebraic solution to the system 3.83 is obtained by applying classical rules of interval arithmetic to dual system (Lakeyev, 1995; Shary, 1997)

$$\widetilde{C}\widetilde{x} - \widetilde{D}|\widetilde{x}| = \widetilde{b} \tag{3.84}$$

Extending the approach to interval kriging system, we transform the composite OIK system 3.69 into an interval system by rewriting $\boldsymbol{K}$ as an interval matrix, and similarly the vector $\boldsymbol{k}$ as an interval vector. We then define an interval kriging system

$$\boldsymbol{K}\boldsymbol{g} = \boldsymbol{k} \tag{3.85}$$

where the interval vector $\boldsymbol{g}$ is the algebraic solution to the system 3.85, if substituting this vector into the system and executing all interval operations according to rules of interval arithmetic results in the system 3.85, (Shary, 2002). Respectively, let $\boldsymbol{K}_{ij}^u$ and $\boldsymbol{K}_{ij}^l$ denote upper and lower $(n+1) \times (n+1)$ kriging matrices, by defining

$$\boldsymbol{C} = \tfrac{1}{2}\left(\boldsymbol{K}_{ij}^u + \boldsymbol{K}_{ij}^l\right), \quad \boldsymbol{D} = \tfrac{1}{2}\left(\boldsymbol{K}_{ij}^u - \boldsymbol{K}_{ij}^l\right),$$

the optimal solution to interval kriging system 3.85 reduces to problem of obtaining the solution to the dual system

$$\underset{(2n+2)\times(2n+2)}{\begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix}} \times \underset{(2n+2)\times 1}{\begin{bmatrix} g^u \\ -g^l \end{bmatrix}} - \underset{(2n+2)\times(2n+2)}{\begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} \end{bmatrix}} \times \underset{(2n+2)\times 1}{\left|\begin{bmatrix} g^u \\ -g^l \end{bmatrix}\right|} = \underset{(2n+2)\times 1}{\begin{bmatrix} k^u \\ -k^l \end{bmatrix}}$$

For details of Kaucher extended interval arithmetic, see (Lakeyev, 1995; Shary, 1997, 2002) and Appendix B.

3.3.5 Spatial Interval Kriging Without Trend

3.3.5.1 Composite Kriging

The spatial random interval estimate $\hat{\boldsymbol{Z}}_0$ at unknown location $\boldsymbol{x}_0$ is obtained first by solving, the composite kriging system, 3.69, to generate weights λ. Once the weights are obtained the estimate is given as

$$\begin{aligned} \hat{\boldsymbol{Z}}_0 &= \left(\sum_{i=1}^{n(h)} \hat{\lambda}_i Z_i^c, \sum_{i=1}^{n(h)} \hat{\lambda}_i Z_i^r \right) \qquad \lambda_i \geq 0 \\ &= \left(\hat{Z}_0^c, \hat{Z}_0^r \right) \end{aligned} \tag{3.86}$$

The estimate, 3.86, can be transformed back to $\hat{\boldsymbol{Z}}_0 = \left(\hat{Z}_0^l, \hat{Z}_0^u \right)$ using 2.45. Let C_0^c and C_0^r respectively denote the center and radius nugget values, while C_1^c and C_1^r respectively denote center and radius sill values. Also let $\boldsymbol{C}_0 = C_0^c + kC_0^r$, $\boldsymbol{C}_1 = C_1^c + kC_1^r$ and $\boldsymbol{C}_{iD} = C_{iD}^c + kC_{iD}^r$. The OIK variance

$$\sigma_{oik}^2 = \boldsymbol{C}_0 + \boldsymbol{C}_1 - \sum_{i=1}^{n(h)} \hat{\lambda}_i \boldsymbol{C}_{iD} + \psi \tag{3.87}$$

3.3.5.2 Interval Arithmetic Kriging

Solving the interval kriging system 3.85, generates a vector of interval weights. The approach gives similar results to composite system. This is explained by the fact that the interval kriging system is obtained by transforming the composite kriging system to an interval system, which is implementing using Kaucher extended interval arithmetic.

3.3.5.3 Component-wise Kriging

We solve the component-wise kriging systems 3.77 and 3.78 to obtain the kriging weights. Estimate is similarly given by 3.86, but with a slight variation. Note that in this case we solve two separate kriging systems, the center and the radius kriging

systems. The estimate is

$$\begin{aligned}\hat{\boldsymbol{Z}}_0 &= \left(\sum_{i=1}^{n(h)} \hat{\lambda}_i^c Z_i^c, \sum_{i=1}^{n(h)} \hat{\lambda}_i^r Z_i^r\right) \qquad \text{such that } \lambda_i^r \geq 0 \\ &= \left(\hat{Z}_0^c, \hat{Z}_0^r\right) \end{aligned} \tag{3.88}$$

Let $\boldsymbol{C}^c = C_0^c + C_1^c$, $\boldsymbol{C}^c = C_0^r + C_1^r$

$$\sigma_{oik}^2 = \left(\boldsymbol{C}^c - \sum_{i=1}^{n(h)} \hat{\lambda}_i^c C_{iD}^c + \mu^c\right) + k\left(\boldsymbol{C}^r - \sum_{i=1}^{n(h)} \hat{\lambda}_i^r C_{iD}^r\right) + \mu^r \tag{3.89}$$

If there is strong spatial correlation structure on the radius variable, we would expect the component-wise kriging system to optimally model the spatial structure in the interval radius and therefore improving the overall kriging, leading to better estimates. However, the radius model is almost a nugget effects model, implying minimal correlation structure of radius measures. Thus, both the component-wise and the composite give parallel results with slight variations.

3.3.6 Spatial Interval Kriging with Trend

In this study, we demonstrate spatial interval kriging with trend where the trend $t(\boldsymbol{x})$, is a function coordinates defined as

$$\hat{t}(\boldsymbol{x}) = \sum_{k=0}^{2} \hat{\beta}_k h_k(\boldsymbol{x}); \qquad h_0(\boldsymbol{x}) = 1 \tag{3.90}$$

with $h_k(\boldsymbol{x})$ the basis functions of the coordinates, and $\hat{\beta}_k$ the corresponding coefficients. We also obtain the trend prediction error, denoted by $\sigma_{\hat{t}(\boldsymbol{x})}^2$.

We, then fit a variogram to the residuals

$$e_i^{\text{resd}} = Z_i^r - \hat{t}(\boldsymbol{x}) \tag{3.91}$$

Since residuals can take on positive and negative values, we relax the positivity constraint for the weights. For each of the defined kriging systems, estimate at unknown location $\boldsymbol{x}_0$ is obtained additively as follows:

Composite Kriging with Trend

$$\hat{\boldsymbol{Z}}_0 = \left(\sum_{i=1}^{n(h)} \hat{\lambda}_i Z_i^c, \sum_{i=1}^{n(h)} \hat{\lambda}_i e_i^{\text{resd}} + \hat{t}(\boldsymbol{x}) \right) \tag{3.92}$$

and can be translated back to $\hat{\boldsymbol{Z}}_0 = \left(\hat{Z}_0^l, \hat{Z}_0^u\right)$ using 2.45. Let $\boldsymbol{C}_0 = C_0^c + kC_0^{\text{resd}}$, $\boldsymbol{C}_1 = C_1^c + kC_1^{\text{resd}}$ and $\boldsymbol{C}_{iD} = C_{iD}^c + kC_{iD}^{\text{resd}}$. Similarly, by additivity property, the kriging variance is

$$\sigma_{oik}^2 = \boldsymbol{C}_0 + \boldsymbol{C}_1 - \sum_{i=1}^{n(h)} \hat{\lambda}_i \boldsymbol{C}_{i0} + \mu + \sigma_{\hat{t}(\boldsymbol{x})}^2 \tag{3.93}$$

3.3.6.1 Interval Arithmetic Kriging with Trend

Similarly, we transform the composite system with trend into an interval system and implement it using Kaucher extended interval arithmetic.

Component-wise Kriging with Trend

By the additivity property,

$$\hat{\boldsymbol{Z}}_0 = \left(\sum_{i=1}^{n(h)} \hat{\lambda}_i^c Z_i^c, \sum_{i=1}^{n(h)} \hat{\lambda}_i^{\text{resd}} e_i^{\text{resd}} + \hat{t}(\boldsymbol{x}) \right) \tag{3.94}$$

Similarly, the kriging variance

$$\begin{aligned} \sigma_{oik}^2 &= C_0^c + C_1^c - \sum_{i=1}^{n(h)} \hat{\lambda}_i^c C_{i0}^c + \mu^c \\ &\quad + k\left(C_0^{\text{resd}} + C_1^{\text{resd}} - \sum_{i=1}^{n(h)} \hat{\lambda}_i^{\text{resd}} C_{i0}^{\text{resd}} + \mu^{\text{resd}} \right) + \sigma_{\hat{t}(\boldsymbol{x})}^2 \end{aligned} \tag{3.95}$$

3.4 Discussion

In this chapter we developed the framework for parametric modeling of spatial random intervals. Two new approaches: composite kriging and component-wise kriging are defined according to the spatial correlation structure of the data vagueness or imprecision measured by the radius variable.

First, we represented the spatial imprecise(fuzzy) data as spatial random intervals and hence defined a spatial random interval process. Applying concepts of closed random interval theory, the defined spatial random interval were represented as spatial bivariate variables consisting of either lower-upper or center-radius regionalized variables.

Utilizing the center-radius regionalized variables we defined the two ordinary kriging approaches:

Composite kriging, based on assumption of zero spatial correlation structure of the radius regionalized variables. Hence, the main contribution of the correlation structure for modeling the spatial random interval process is the center variables. The approach coincides with the classical best linear unbiased estimate (BLUE) in linear regression, with observations correlated by the composite (co)variance (= Center (co)variance plus radius variance).

Based on the composite covariance, kriging is performed using center data conditional on sum of center weights equals 1. The fact that no condition is imposed on the radius weights is a disadvantage of the composite kriging; the uncontrolled radius weight values leads to large radius values for the estimated values, hence wider intervals. Because of this weakness, we refer to the estimate as a WEAK COMPOSITE "BLUE".

Component-wise kriging, based on assumption of a full spatial correlation structure of the radius regionalized variables. The approach involve splitting up the problem into separate kriging for the center and the radius data. Where, necessary, additional constraint of non-negative kriging weights for radius variable is imposed.

Kriging on radius data leads to a linear unbiased estimates for radius spatial field. Similarly Kriging on center data leads to a linear unbiased estimates for "center spatial field". Combination of these two linear unbiased estimates implies a linear unbiased combined estimate of the spatial random interval process. We refer to this estimate as the COMPONENT-WISE BLUE. The estimate is optimal and reduces to composite BLUE when radius data have no significant correlation structure. The "composite BLUE" parallels Diamond (1988) approach to kriging of intervals.

In ordinary kriging, a common solution to deal with non-stationarity is to apply regression technique's so as to obtain stationary random field of residuals before prediction (kriging with trend). For illustration purposes, the approach was extended

to interval-valued data, leading to composite interval regression and component-wise interval regression. Extension to the general approach, intrinsic random functions of order k (IRF-k), was not addressed, rather a kernel based non-parametric spatial interval regression approach (Chapter4), that does not impose the stationarity assumption.

CHAPTER 4

NON-PARAMETRIC MODELING OF SPATIAL RANDOM INTERVALS APPLICATION OF A SPATIALLY-WEIGHTED INTERVAL KERNEL

Scope

In this chapter we propose a non-parametric kernel based regression method that relaxes the first and second order stationarity assumptions. The approach utilizes and allows the spatial random interval data to define the optimal functional form of the estimate of the random field.

A detailed review of non-parametric kernel-based algorithms is given with particular emphasis on concepts that apply to modeling of spatial random interval data. Taking advantage of support function embedding of intervals into Hilbert's space, we propose a "spatial-interval" working domain space. A spatial adaptive kernel and a robust interval uncertainty kernel are defined for the new domain. The spatial kernel creates the local spatial structure that ensures that only samples sharing similar intensity and gradient information are gathered for local approximation. The robust uncertainty kernel minimizes the influence of outliers caused by occasional miss-registration. On a scale of zero to one, the robust kernel measures the reliability of measurements at neighborhood locations. A zero measure indicates completely untrustworthy data and a one measure represent trustworthy data. Local estimates at selected locations are obtained as a data convolution of the two kernels.

4.1 Nonparametric Estimation: Applications to Interval-valued Spatial Process

Normally, problems in pattern recognition involve obtaining the probability density function describing an observed random quantity. In general, the forms of the underlying density functions are not known. While classical parametric densities are mostly unimodal, practical pattern recognition problems involve multi-modal densities. Further, high-dimensional densities cannot often be simply represented as the product of one-dimensional density functions. While mixture methods (McLachlan and Peel, 2000), partially alleviate this problem, they are still restricted to parametric modeling. For example, mixture methods, require that the designer have extensive knowledge of the problem. Implicitly, one needs to know the form of the models and the model parameters.

An alternative method to parametric approach is the nonparametric approach to density estimation. An attractive feature of nonparametric procedures is that they can be used with arbitrary distributions, without the assumption that the forms of the underlying densities are known. Nonparametric density estimation has experienced a wide explosion of interest over the last 20 years. Texts include Bowman and Azzalini (1997); Loader (1999); Schimek (2000); Wand and Jones (1995).

Among the nonparametric methods is the histogram (Bowman and Azzalini, 1997). Data is grouped into intervals (bins), and each interval is represented with its midpoint. Density function is defined by counting the frequency of data that falls into each bin. The major drawback with histograms is the lack of convergence to the right density function if the data set is small (Comaniciu et al., 2000; Elgammal et al., 2000). Another drawback with histograms, in general, is that they are not suitable for higher dimensional features.

A particular nonparametric technique that estimates the underlying density and is quite general is the kernel density estimation (KDE) technique (Scott, 1992; Silverman, 1986). Briefly, consider an independent identically distributed random sample $x_i, i = 1, \ldots, n$, from one-dimensional space, i.e. $(x_i \in \mathbb{R})$, with an unknown continuous probability density function f, the kernel density estimator,

$$\hat{f}(x) = \frac{1}{n} \sum_{i}^{n} K_b(x - x_i) \tag{4.1}$$

where b is the scale or smoothing parameter, $K_b(\cdot)$ is a symmetric kernel function,

defined in §4.2.2. As a scaled kernel i.e. $K_b(x - x_i) = (1/b)K((x - x_i)/b)$ and the density reduces to

$$\hat{f}(x) = \frac{1}{nb}\sum_{i}^{n} K\left(\frac{x - x_i}{b}\right) \tag{4.2}$$

A good discussion of kernel estimation techniques can be found in Scott (1992) and Wand and Jones (1995).

Unlike histograms, even with a small number of samples, KDE leads to a smooth, continuous and differentiable density estimate. Kernel density estimators asymptotically converge to any density function with enough samples (Bowman and Azzalini, 1997; Scott, 1992). This property makes the technique quite general for estimating the density of any distribution. In reality, other nonparametric density estimation methods, e.g., histograms and mixture methods (McLachlan and Peel, 2000), can be shown to be asymptotically kernel methods (Duda et al., 2001). For higher dimensions, $(\mathbb{R}^d, d > 1)$, products of one-dimensional kernel functions (see §4.4) are used. All these benefits make KDE very appealing to modeling interval-valued distributions.

Though, KDE has been applied to the field of pattern recognition, little has been done in kernel density estimation and regression for set-valued data and even much less for set-valued spatial data. To our knowledge, a short talk by Friel (1999), on set-valued regression based on set metrics is the only reference. We extend Friel's approach, applying KDE techniques to the problem of modeling the set-valued spatial distribution and utilizes a weighted metric between spatial random intervals. A special case of an interval-valued spatial process is given as an example.

4.2 Definition of Kernel and its Profile Function

Modifying Mark and Tomasi (2005) definition:-

Definition 4.2.1 *A* **profile** k*, is a piecewise continuous, monotonically non-increasing function from a non-negative real to a non-negative real;* $k : [0, \infty) \rightarrow [0, \infty)$*, such that the definite integral* $\int_0^\infty k(q)dq < \infty$.

Definition 4.2.2 *Let* X *denote a* d*-dimensional real Euclidean space; a subset of* $\mathbb{R}^d$*. A* **kernel** K *is a function from a vector* $x \in X$ *to a non-negative real, i.e.*

$K : X \rightarrow \mathbb{R}$ *with the following properties:-*

$$K(\boldsymbol{x}) = K(-\boldsymbol{x}), \text{ such that } K(-\boldsymbol{x}), K(\boldsymbol{x}) \geq 0 \tag{4.3}$$

$$\int_X K(\boldsymbol{x})d\boldsymbol{x} = 1 \tag{4.4}$$

$$\int_X \boldsymbol{x}K(\boldsymbol{x})d\boldsymbol{x} = 0 \tag{4.5}$$

$$\int_X \boldsymbol{x}^2K(\boldsymbol{x})d\boldsymbol{x} = \sigma_k^2 > 0 \tag{4.6}$$

The even symmetry (4.3)*, allows us to define the kernel profile* $k(\boldsymbol{x})$*, from*

$$K(\boldsymbol{x}) = c_{k,d}k(||\boldsymbol{x}||^2), x \geq 0 \tag{4.7}$$

such that $c_{k,d}$ *is the normalization constant determined from 4.4, and* $||\cdot||^2$ *is the* L_2 *norm of a point* $\boldsymbol{x}$ *from the origin.*

The importance of the profile is revealed in the case of multivariate kernel density estimation (§4.4), spatial interval density estimation (§4.6) and spatial interval regression in §4.8.

Also we note that, kernels cannot be differentiated, rather profiles are differentiated leading to new kernels referred to as shadow kernels (Cheng, 1995). For example if the derivative of $k(\cdot)$ exists, then, $g(\cdot) = -k'(\cdot)$ can be used as a profile to define a new kernel $G(\boldsymbol{x})$ such that $G(\boldsymbol{x}) = c'_{k,d}g(||\boldsymbol{x}||^2)$ with normalization constant $c'_{k,d}$. $G(\cdot)$ is the shadow kernel of kernel $K(\cdot)$. This property becomes very useful when deriving the kernel interval estimates using gradient decent minimization in §4.8.

4.3 Kernel Density Estimation

Kernel density estimation is a smoothing method that generalizes individual point locations or events, $\boldsymbol{x}_i, i = 1, \ldots, n$, to an entire area and provides density estimates (probability density), $f(\boldsymbol{x})$, at any location within the study region R. From a visual point of view (Figure 4.1) it can be thought of a three-dimensional sliding kernel function k_i that 'visits' every location $\boldsymbol{x}$

Distances to each observed events $\boldsymbol{x}_i$ that lies within a specified distance b_i, referred to as the bandwidth or smoothing parameter, are measured and contribute to the density estimate at $\boldsymbol{x}$ according to how close they are to $\boldsymbol{x}$, (Gatrell et al.,

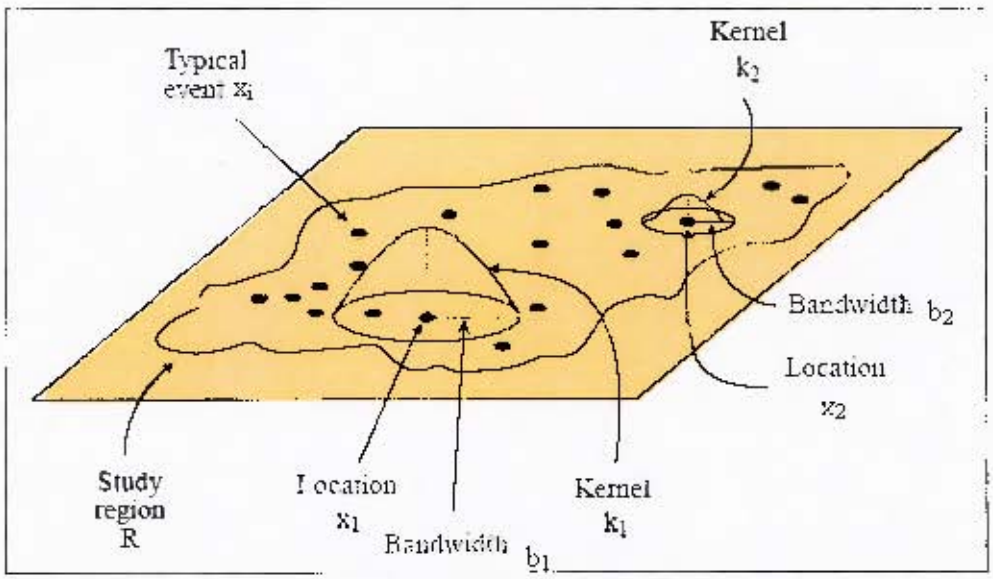


Figure 4.1. Kernel density estimation of a spatial point process, Adapted from Bailey and Gatrell (1995) .

1996). The kernel density estimate produces a more smooth estimate of $f(\boldsymbol{x})$, the sum of kernels centered at each observation.

Since, the kernel contributes to the density estimation of points within that same local neighborhood, we are able to take into account the local correlations, where points close in a kernel are given smaller bandwidths, hence high correlation. The problem of data sparsity can be effectively handled by adapting the kernel bandwidth (b_i) according to the extent of local neighborhood as seen in Figure 4.1.

For simplicity of notation, let b also denote the adaptive bandwidth b_i. The commonly used kernel functions, (Benedetti, 1977; Silverman, 1986) are listed in Table 4.1 and displayed in Figure 4.2.

Table 4.1. Examples of commonly used univariate kernel functions.

Gaussian	$K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$	$-\infty < t < \infty$
Epanechnikov	$K(t) = \frac{3}{4}(1-t^2)$	$\lvert t\rvert \leq 1$
Bisquare (Quartic)	$K(t) = \frac{15}{16}(1-t^2)^2$	$\lvert t\rvert \leq 1$
Tricube	$K(t) = \frac{35}{32}(1-t^2)^3$	$\lvert t\rvert \leq 1$

$t = (x - x_i)/b$

The Gaussian kernel extends to infinity in all directions, and therefore is evaluated for each point in the region (Leitner and Staufer-Steinnocher, 2001). It is simple and a popular choice for kernels with infinite support, (Scott, 1992). On the

contrary, the symmetric Beta kernels (Epanechnikov, Bisquare, Tricube), defined by a circumscribed radius b, which is also the bandwidth, applies to a limited or finite area around each event.

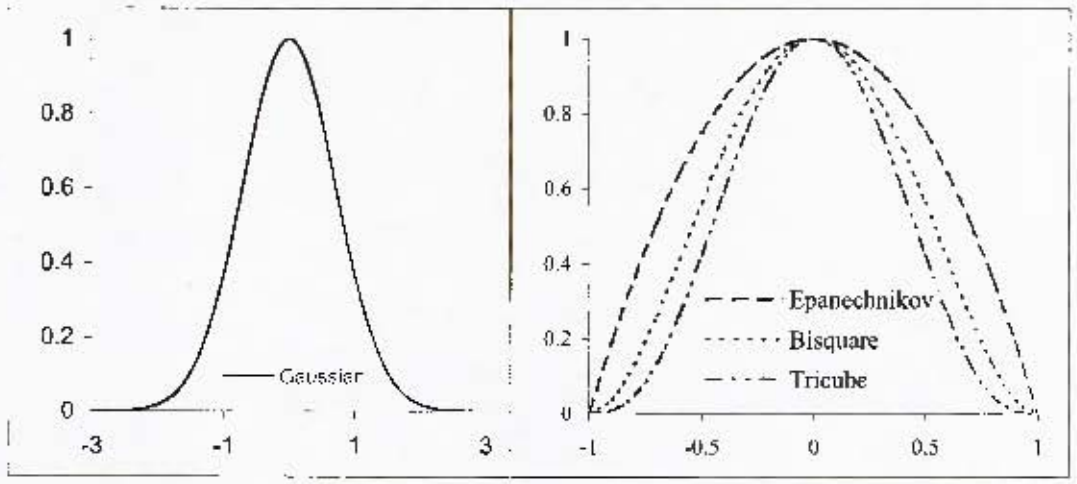


Figure 4.2. A graph of commonly used univariate kernel functions. Left graph is the Gaussian kernel and right graph is the symmetric Beta family kernels on finite support $[-1, 1]$.

This ensures that only data, local to the point at which $\hat{f}(\boldsymbol{x})$ is estimated, is used in the fit. The Epanechnikov kernel is optimal as it minimizes the mean squared error (mse) (Scott, 1992). A truncated Gaussian kernel could also be used. Often, the Gaussian kernel is used for its continuity, differentiability, and locality properties. In addition, choosing the Gaussian as a kernel function is different from fitting the distribution to a Gaussian model (normal distribution) or to a mixture of Gaussian models (McLachlan and Peel, 2000). Here, the Gaussian is only used as a function to weight the data points. Unlike parametric fitting of a mixture of Gaussians, kernel density estimation is a more general approach that does not assume any specific shape for the density function.

4.4 Multivariate Kernel Density Estimation

Let $\boldsymbol{x}$ denote a vector in a $d-$dimensional region $\mathbb{R}^d$. Multivariate extensions of univariate kernel $K_1(x)$ to $\mathbb{R}^d$ is obvious. A multivariate kernel is defined in two ways:

$$\begin{aligned} &\text{Product kernel:} && K^{\mathrm{P}}(\boldsymbol{x}) = K_1(x_1) \cdots K_1(x_d) \\ &\text{Radially symmetric kernel:} && K^{\mathrm{R}}(\boldsymbol{x}) = a_{k,d} K_1(||\boldsymbol{x}||), \ \boldsymbol{x} \geq 0, \end{aligned} \tag{4.8}$$

where $||\cdot||$ is an appropriate norm in $\mathbb{R}^d$ and $a_{k,d} = 1/\int_{\mathbb{R}^d} K_1(||\boldsymbol{x}||)d\boldsymbol{x}$ is a normalization constant, ensuring that $K^{\mathrm{R}}(\boldsymbol{x})$ integrates to one (Comaniciu and Meer, 2002). Zheng et al. (2004) refers to $a_{k,d}$ as the density edge correction term where $\int_{\mathbb{R}^d} K_1(||\boldsymbol{x}||)d\boldsymbol{x}$ is the unit volume of a $d-$dimensional sphere. Hence, using Zheng et al. (2004) definition and Table 3.1 in chapter 3, $a_{k,1} = c_1^{-1} = 1/2$ for 1D space, $a_{k,2} = c_2^{-1} = 1/\pi$ for 2D space and $a_{k,3} = c_3^{-1} = 3/4\pi$ for 3D space.

We will deal with a special class of multivariate radially kernels in Table 4.2, based on the L_2 norm $||\cdot||^2$, and previously defined in Eq. 4.7 as

$$K(\boldsymbol{x}) = c_{k,d}k(||\boldsymbol{x}||^2), x \geq 0$$

The strictly positive normalization constant $c_{k,d}$ ensures that $K(\boldsymbol{x})$ integrates to one. We let K denote both the product and radial kernels, and without confusion, respective meaning should arise from the context.

Table 4.2. Examples of multivariate radial symmetrical kernel functions (derived using Bowman and Azzalini, 1997).

Gaussian	$K(t) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{\|t\|^2}{2}\right)$
Epanechnikov	$K(t) = \frac{1}{2c_d}(d+2)\left(1-\|t\|^2\right)$
Bisquare (Quartic)	$K(t) = \frac{1}{8c_d}(d+2)(d+4)\left(1-\|t\|^2\right)^2$
Tricube	$K(t) = \frac{1}{48c_d}(d+2)(d+4)(d+6)\left(1-\|t\|^2\right)^3$

$t = (x - x_i)/b$

c_d is unit volume of a $d-$ dimensional sphere

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ denote a vector of data points where $\boldsymbol{x}_i$ lie in a $d-$dimensional study region: $R \subset \mathbb{R}^d$. The multivariate density $f(\boldsymbol{x})$, at a point $\boldsymbol{x}$ is estimated as follows (Burt and Barber, 1996):

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{b^d} K\left(\frac{\boldsymbol{x}_i - \boldsymbol{x}}{b}\right) \tag{4.9}$$

where $K(\cdot)$ is a multivariate kernel function (see examples in Table 4.2) with b the common bandwidth in all dimensions d. Rather than have a constant bandwidth b for all dimensions, we could have different bandwidth $b_1, \ldots, b_d$ for each d dimensions. For simplicity, we will assume a constant bandwidth in all dimensions i.e. $b = b_1 = \cdots = b_d$.

Using the multivariate kernel in Eq. 4.7, the corresponding multivariate density

estimator (4.9) can be rewritten as

$$\hat{f}(\boldsymbol{x}) = \frac{1}{nb^d}\sum_{i=1}^{n} c_{k,d}k\left(\left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}}{b}\right\|^2\right) \tag{4.10}$$

normalized by the constant $c_{k,d}$. For the Gaussian multivariate kernel, $c_{k,d} = 1/(2\pi)^{d/2}$, (see Table 4.2), and the density estimate 4.10 reduces to

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(2\pi b^2)^{d/2}} \exp\left(-\frac{1}{2}\left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}}{b}\right\|^2\right) \tag{4.11}$$

Example 4.4.1 *Consider a 2-dimensional spatial random process characterized by the pairs* $\{\boldsymbol{x}_i, Z_i\}_{i=1}^n$, *with* Z_i *a real-valued measurement at each* $\boldsymbol{x}_i \in R \subset \mathbb{R}^2$. *The Gaussian bivariate spatial density estimate,* $\hat{f}(\boldsymbol{x})$, *at a location* $\boldsymbol{x}$, *is given as*

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i}^{n} \frac{1}{2\pi b^2} \exp\left(-\frac{1}{2}\left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}}{b}\right\|^2\right) \tag{4.12}$$

and with Epanechnikov bivariate radial kernel,

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i}^{n} \frac{2}{\pi b^2}\left(1 - \left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}}{b}\right\|^2\right) \tag{4.13}$$

while for Bisquare bivariate radial kernel,

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n}\sum_{i}^{n} \frac{3}{\pi b^2}\left(1 - \left\|\frac{\boldsymbol{x}_i - \boldsymbol{x}}{b}\right\|^2\right)^2 \tag{4.14}$$

4.4.1 Selection of Kernel Bandwidth b

The choice of kernel function is of secondary importance to selection of bandwidth parameter (Silverman, 1986). Zheng et al. (2004) and Scott (1992) indicates that the shape of the kernel does not significantly affect the estimate mean square error (mse). By varying the bandwidth, different kernels can be made equivalent in terms of mean square error.

The effect of increasing the bandwidth b is to stretch the region around $\boldsymbol{x}$ within which observed events influence the density estimate at $\boldsymbol{x}$. For large b, $f(\boldsymbol{x})$ will appear flat and local features will be obscured. If b is small, then $f(\boldsymbol{x})$ tends to a collection of spikes centered on $\boldsymbol{x}$ (Bowman and Azzalini, 1997).

Much has been written on choice of the bandwidth parameters. For example, in $d-$dimension,

$$b_j = (4/(d+2))^{1/(d+4)} \times n^{-1/(d+4)}\sigma_j \text{ or } b_j = n^{-1/(d+4)}\hat{\sigma}_j \tag{4.15}$$

respectively denotes Silverman (1986) and Scott (1992) bandwidths with σ_j the variance of the j^{th} dimension. In ArcGIS software approach to kernel estimation, the bandwidth is given by $\min(x, y)/30$- the minimum dimension of the extent of spatial point pattern divided by 30.

Nonetheless, these methods do not take into account the spatial distribution of the points. In addition the arbitrary nature of coefficients (e.g. the value 30 in ArcGIS) has no statistical interpretation.

Alternative automatic bandwidth selection methods include cross-validation, generalized cross-validation, minimum AIC, and direct plug-in methods. Excellent overviews of these various methods are given by Bowman and Azzalini (1997); Gatrell et al. (1996); Loader (1999); Wand and Jones (1995). The cross-validation method, exploits the adequacy of fit of the model under consideration. An example is the Least Square Cross Validation (CV) (Loader, 1999). At each observation point i, we estimate the phenomena, Z_i , using all the available observations except the observation at point i. A cross validated sum of squares of errors is then formed between the estimate $\hat{Z}_i$ and the true value Z_i as shown in Eq. 4.16.

$$CV = \sum_{i=1}^{n} \left[\frac{\sum_{i \neq j=1}^{n} K_b(\boldsymbol{x} - \boldsymbol{x}_j) \times Z_j}{\sum_{i \neq j=1}^{n} K_b(\boldsymbol{x} - \boldsymbol{x}_j)} - Z_i \right]^2 \tag{4.16}$$

where $K_b(\boldsymbol{x} - \boldsymbol{x}_j)$ is the kernel function, either the Bisquare or the Epanechnikov kernels.

A global optimal bandwidth is the one that corresponds to the lowest CV score. A global bandwidth will cover a variable number of observations depending on the location of the estimate. A variable (local) bandwidth may be chosen to cover a certain number of observations. Given k nearby observations, the local bandwidth at point i extend to the kth Nearest Neighbor (KNN) of point i, (Silverman, 1986). The optimal KNN is then the one that minimizes the CV score in Eq. 4.16.

The resulting estimator is called the adaptive kernel estimate where the bandwidth is dependent on: a) the process spatial variation; and b) the observations density around the point of estimate. The method also adds flexibility to ker-

nel estimation by allowing the user to vary k depending on the desired degree of smoothing. And unlike Kriging, the underlying stochastic process is not assumed to be stationary. The approach has been used by among others Zhuang et al. (2002) and Choi and Hall (1998).

Recently developed method utilizing the concept of KNN approach and not covered in this work is the work on Geographically Weighted Regression (Fotheringham et al., 2002).

4.5 Limits of Kernel Based Methods

Kernel methods exploits the linear structure of vector spaces, however, the interval space, $\mathcal{K}_{cx}(\mathbb{R})$ is not a vector space and is also bounded. This limits the application of KDE. A solution is to transform or isometrically embed the interval space to an unbounded vector space, the Hilbert space. Embedding $\mathcal{K}_{cx}(\mathbb{R})$ into a Hilbert space implies that results and properties in the L_2 space can be transferred to $\mathcal{K}_{cx}(\mathbb{R})$ space. Indirectly, using the kernel trick (see definition 4.5.1 in §4.5.1) we are able to define an appropriate kernel for interval space, making it possible to perform kernel density estimation and regression for interval-valued data. Since interval space is low in dimensionality, efficient computation of kernel density estimation for interval-valued spatial probability density functions can be achieved.

4.5.1 Kernel Methods for Nonlinear Input Spaces

Most algorithms for data analysis (e.g. least squares regression, multivariate regression) are based on the assumption that the data can be represented as vectors in a finite dimensional vector space. These methods make extensive use of the linear structure.

However, in real life, data is often collected from varied input spaces e.g. interval space for interval-valued data, and color space or image space for image data. These input spaces are not vector spaces and could also be bounded, which limits the application of nonparametric KDE method. We are faced with the following question "How do we exploit the linear algorithms for analysis in nonlinear spaces?" Using kernels, we can derive nonlinear versions of the linear algorithms.

The approach can be described as follows: Let $\mathcal{F}$ denote a nonlinear input space, and $\mathcal{H}$ a vector space (the Hilbert space). Given a linear algorithm (an algorithm that works in $\mathcal{H}$), one first maps the data in nonlinear input space to a Hilbert space using a nonlinear mapping $\phi(\cdot) : \mathcal{F} \rightarrow \mathcal{H}$, and then run the algorithm on the vector

representation $\phi(\boldsymbol{x})$ of the data. In other words, we perform nonlinear analysis of the data using a linear method.

The purpose of the map $\phi(\cdot)$ is to translate nonlinear structures of the data into linear ones in $\mathcal{H}$, preserving the natural basic structure in input space. But then, "Which is the right map $\phi(\cdot)$ to achieve this?" and "Can we use it to define a kernel in the input space?" We will look at maps or embeddings that correspond to kernels in input spaces.

Note that, if while executing algorithms in Hilbert space $\mathcal{H}$, only inner products between data vectors are considered, i.e. data appears only in expressions like $\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle = \phi(\boldsymbol{x})^T \phi(\boldsymbol{x}')$, we can make use of the fact that for certain specific maps, $\phi(\cdot)$, this inner product can be computed directly from $\boldsymbol{x}$ and $\boldsymbol{x}'$ without explicitly computing $\phi(\boldsymbol{x})$ and $\phi(\boldsymbol{x}')$. This is referred to as the kernel trick (see definition 4.5.1).

Definition 4.5.1 ***Kernel "trick"*** *(Shawe-Taylor and Cristianini, 2004, p. 60) Let k_1 denote any symmetrical kernel function on input space $\mathcal{F}$, such that the kernel matrix with elements $k_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is positive definite for all $\boldsymbol{x}_i \in \mathcal{F}, i = 1, \ldots, n$; then, there exists a Hilbert space $\mathcal{H}$ and a nonlinear mapping $\phi : \mathcal{F} \rightarrow \mathcal{H}$ such that for $\phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \in \mathcal{H}$,*

$$k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$$

Since kernels correspond to inner products in some spaces, they can be considered as measures of similarities between data points for any input space. We demonstrate this for a general case in §4.5.2 with an example of an interval input space.

4.5.2 General Input Space Kernels

Let $\mathcal{F}$ denote an input space, and k_1 denote a kernel in $\mathcal{F}$ corresponding to map $\phi(\cdot)$ into a Hilbert space, $\mathcal{H}$. Then, for $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{F}$, $\phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \in \mathcal{H}$, such that $k_1(\boldsymbol{x}, \boldsymbol{z}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \rangle$. The Hilbert norm defined as

$$||\phi(\boldsymbol{x})||^2 = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}) \rangle = k_1(\boldsymbol{x}, \boldsymbol{x}), \tag{4.17}$$

coincides with Scholkopf et al. (2004) definition. The associated Hilbert distance is

$$\begin{aligned} ||\phi(\boldsymbol{x}) - \phi(\boldsymbol{z})||^2 &= \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}) \rangle - 2\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \rangle + \langle \phi(\boldsymbol{z}), \phi(\boldsymbol{z}) \rangle \\ &= k_1(\boldsymbol{x}, \boldsymbol{x}) - 2k_1(\boldsymbol{x}, \boldsymbol{z}) + k_1(\boldsymbol{z}, \boldsymbol{z}) \text{ using 4.17} \end{aligned} \tag{4.18}$$

Both Hilbert norm and distance are functions of kernels and hence are also kernels. By making $k_1(\boldsymbol{x}, \boldsymbol{z})$, the subject, we can rewrite 4.18 as

$$\begin{aligned} k_1(\boldsymbol{x}, \boldsymbol{z}) &= \left(k_1(\boldsymbol{x}, \boldsymbol{x}) + k_1(\boldsymbol{z}, \boldsymbol{z}) - ||\phi(\boldsymbol{x}) - \phi(\boldsymbol{z})||^2\right)/2 \\ &= \left(||\phi(\boldsymbol{x})||^2 + ||\phi(\boldsymbol{z})||^2 - ||\phi(\boldsymbol{x}) - \phi(\boldsymbol{z})||^2\right)/2 \end{aligned} \tag{4.19}$$

From 4.19, the kernel as a measure of similarity between objects is given by the opposite of the squared distance $||\phi(\boldsymbol{x}) - \phi(\boldsymbol{z})||^2$ between their images in the Hilbert space, plus the norms. Often, norms in Hilbert space equals one (Shawe-Taylor and Cristianini, 2004), which reduces 4.19 to

$$k_1(\boldsymbol{x}, \boldsymbol{z}) = 1 - ||\phi(\boldsymbol{x}) - \phi(\boldsymbol{z})||^2/2 \tag{4.20}$$

Example 4.5.1 *Let $\mathcal{K}_{cx}(\mathbb{R})$ denote an interval input space, and k_1 a kernel in the $\mathcal{K}_{cx}(\mathbb{R})$ corresponding to support function mapping $S(\cdot)$ (see chapter 3, Eq. 3.10) into the Hilbert space $\mathcal{H}$. Respectively, for intervals $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathcal{K}_{cx}(\mathbb{R})$, we have support functions $S_{\boldsymbol{a}}$ and $S_{\boldsymbol{b}}$ in $\mathcal{H}$. Extending the kernel trick to interval space, the interval kernel becomes*

$$k_1(\boldsymbol{a}, \boldsymbol{b}) = \langle S_{\boldsymbol{a}}, S_{\boldsymbol{b}} \rangle, \tag{4.21}$$

with interval norm

$$||S_{\boldsymbol{a}}||^2 = \langle S_{\boldsymbol{a}}, S_{\boldsymbol{a}} \rangle = k_1(\boldsymbol{a}, \boldsymbol{a}). \tag{4.22}$$

Applying 4.18, the interval distance is

$$||S_{\boldsymbol{a}} - S_{\boldsymbol{b}}||^2 = k_1(\boldsymbol{a}, \boldsymbol{a}) - 2k_1(\boldsymbol{a}, \boldsymbol{b}) + k_1(\boldsymbol{b}, \boldsymbol{b}) \tag{4.23}$$

and thus

$$k_1(\boldsymbol{a}, \boldsymbol{b}) = \left(||S_{\boldsymbol{a}}||^2 + ||S_{\boldsymbol{b}}||^2 - ||S_{\boldsymbol{a}} - S_{\boldsymbol{b}}||^2\right)/2 \tag{4.24}$$

Kernels can also be derived from existing kernels. The easiest way to derive a kernel is to normalize an existing kernel (Shawe-Taylor and Cristianini, 2004, p.60). Let $\boldsymbol{x}, \boldsymbol{z}$ be elements of space X, and $k_1(\boldsymbol{x}, \boldsymbol{z})$ a kernel in X corresponding to a map $\phi(\boldsymbol{x})$ into $\mathcal{H}$, such that $k_1(\boldsymbol{x}, \boldsymbol{z}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \rangle$. Similarly, the normalized kernel $k(\boldsymbol{x}, \boldsymbol{z})$, corresponds to the map

$$\frac{\phi(\boldsymbol{x})}{||\phi(\boldsymbol{x})||}$$

such that

$$k(\boldsymbol{x},\boldsymbol{z}) = \left\langle \frac{\phi(\boldsymbol{x})}{||\phi(\boldsymbol{x})||}, \frac{\phi(\boldsymbol{z})}{||\phi(\boldsymbol{z})||} \right\rangle = \frac{k_1(\boldsymbol{x},\boldsymbol{z})}{\sqrt{k_1(\boldsymbol{x},\boldsymbol{x})k_1(\boldsymbol{z},\boldsymbol{z})}} \tag{4.25}$$

Let c denote a constant in $\mathbb{R}^+$, other valid kernels include:

$$k(\boldsymbol{x},\boldsymbol{z}) = ck_1(\boldsymbol{x},\boldsymbol{z}) \quad \text{and} \quad k(\boldsymbol{x},\boldsymbol{z}) = \exp(k_1(\boldsymbol{x},\boldsymbol{z})) \tag{4.26}$$

The simplest kernel is a linear kernel which we denote as $k_2(\boldsymbol{x},\boldsymbol{z}) = \langle \boldsymbol{x},\boldsymbol{z} \rangle$, implying that $\boldsymbol{x} = \phi(\boldsymbol{x})$. Given a constant $c = 1/\sigma^2$, we define an exponential kernel

$$k(\boldsymbol{x},\boldsymbol{z}) = \exp(ck_2(\boldsymbol{x},\boldsymbol{z})) = \exp(\langle \boldsymbol{x},\boldsymbol{z} \rangle/\sigma^2).$$

We also know that,

$$\sqrt{\exp(x)} = \exp\left(\frac{x}{2}\right) \Longleftrightarrow \frac{1}{\sqrt{\exp(x)}} = \exp\left(-\frac{x}{2}\right)$$

Thus, normalizing the exponential kernel, we have

$$\begin{aligned} \frac{\exp(\langle \boldsymbol{x},\boldsymbol{z} \rangle/\sigma^2)}{\sqrt{\exp(||\boldsymbol{x}||^2/\sigma^2)\exp(||\boldsymbol{z}||^2/\sigma^2)}} &= \exp\left(\frac{\langle \boldsymbol{x},\boldsymbol{z} \rangle}{\sigma^2} - \frac{\langle \boldsymbol{x},\boldsymbol{x} \rangle}{2\sigma^2} - \frac{\langle \boldsymbol{z},\boldsymbol{z} \rangle}{2\sigma^2}\right) \\ &= \exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{z}||^2}{2\sigma^2}\right) \end{aligned} \tag{4.27}$$

which is the Gaussian kernel given by Shawe-Taylor and Cristianini (2004, p.77) and Vapnik (2000). The Gaussian bandwidth parameter (σ^2) specifies how quickly the kernel vanishes as points move further away from each other. For simplicity, we use $k_G(\boldsymbol{x},\boldsymbol{z})$ to denote the Gaussian kernel, i.e.

$$k_G(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{z}||^2}{2\sigma^2}\right) \tag{4.28}$$

with σ^2 the Gaussian bandwidth.

4.5.3 Interval-valued Kernel

For interval input space, applying the concept of support function mapping, the distance between intervals is equivalent to the Hilbert (L_2) distance. This distance has already been calculated in chapter 3, Eq. 3.16, and we recall it here. Given

intervals $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{K}_{cx}(\mathbb{R})$, the interval distance is

$$\delta_2^2(\boldsymbol{a}, \boldsymbol{b}) \;=\; (a^c - b^c)^2 + (a^r - b^r)^2 \tag{4.29}$$

In chapter 3, Eq. 3.17, we defined a general weighted interval metric, given as

$$\boldsymbol{d}_G^2(\boldsymbol{a}, \boldsymbol{b}) \;=\; (a^c - b^c)^2 + \frac{\alpha}{\alpha + \beta}(a^r - b^r)^2 \tag{4.30}$$

Extending 4.28 to interval space, the Gaussian kernel for interval-valued data is

$$\boldsymbol{k}_G(\boldsymbol{a}, \boldsymbol{b}) = \exp\Big(-\frac{||\boldsymbol{a} - \boldsymbol{b}||^2}{2\sigma^2}\Big) \tag{4.31}$$

where $||\boldsymbol{a} - \boldsymbol{b}||^2$ is either 4.30 or 4.29; and σ^2 is the Gaussian kernel bandwidth determining the extent of interval dependence or correlation. Parameter σ^2 controls the flexibility of the kernel. While small σ^2 risks over fitting or existence of bounds / irregular estimate, large values gradually reduces the kernel to a constant or flat function (oversmoothing). This is because large values of the bandwidth reduce the locality of the kernel and consequently overfitting.

Other non-Gaussian profiles for interval-valued data include

$$\text{Bisquare: } \left(1 - \left(\frac{d_G(\boldsymbol{a}, \boldsymbol{b})}{b}\right)^2\right)^2, \qquad \text{Epanechnikov: } 1 - \left(\frac{d_G(\boldsymbol{a}, \boldsymbol{b})}{b}\right)^2 \tag{4.32}$$

where $d_G(\boldsymbol{a}, \boldsymbol{b}) \leq b$ is the interval distance between intervals and b is the bandwidth determined using KNN approach.

4.6 Spatial Interval Kernel Density Estimation

The spatial-interval domain space is generated by concatenating two independent domains: the 'spatial domain" due to the spatial coordinates, and the "interval domain" due to interval coordinates. Hence, the spatial-interval domain space assumes a coordinate system of dimension $d = s + q$, with q representing the interval dimension and s representing the spatial dimension. For the interval domain, $q = 2$ representing the lower-upper or the center-radius coordinate variables. For 1D spatial domain $s = 1$, for 2D $s = 2$ and for $3D$ $s = 3$. While we work in a $2D$ space extensions to $3D$ is straight forward. Given a 2D spatial domain $\boldsymbol{x} = (x, y)$ and an interval domain $\boldsymbol{Z}(\boldsymbol{x}) = \big(Z^l(\boldsymbol{x}), Z^u(\boldsymbol{x})\big)$, we define the joint space denoted as $\boldsymbol{p} = \{\boldsymbol{x}, \boldsymbol{Z}(\boldsymbol{x})\}$ with dimension $d = 2 + 2 = 4$.

Due to the different nature of the two domains, proper normalization is needed prior to analysis or estimation (Comaniciu and Meer, 2002). We will achieve this through the use of normalized kernels in the respective domains.

Let b^s and b^q respectively denote the spatial and interval domain bandwidth parameters. For kernel density estimation in the joint spatial-interval space, we use product kernel that consists of a spatial radially bivariate kernel $k^s; b^s$ which corresponds to a regular 2-dimensional non-negative kernel function bounded by bandwidth b^s, and an interval radially kernel $k^q; b^q$ which corresponds non-negative kernel function bounded by bandwidth b^q. The multivariate kernel for the joint spatial-interval space is a product of the two given as

$$K_{b^s,b^q}(\boldsymbol{p} = \{\boldsymbol{x}, \boldsymbol{Z}(\boldsymbol{x})\}) = \frac{C}{(b^s)^2(b^q)^2} k^s \left(\left\| \frac{\boldsymbol{x}}{b^s} \right\|^2 \right) k^q \left(\left\| \frac{\boldsymbol{Z}(\boldsymbol{x})}{b^q} \right\|^2 \right) \tag{4.33}$$

where $\boldsymbol{x}$ and $\boldsymbol{Z}(\boldsymbol{x})$ are respectively the spatial and interval parts; and C is the normalization constant. Then, the interval kernel density estimate is defined as

$$\hat{f}(\boldsymbol{p} = \{\boldsymbol{x}, \boldsymbol{Z}(\boldsymbol{x})\}) = \frac{C}{n(b^s)^2(b^q)^2} \sum_{i=1}^{n} k^s \left(\left\| \frac{\boldsymbol{x}_i - \boldsymbol{x}}{b^s} \right\|^2 \right) k^q \left(\left\| \frac{\boldsymbol{Z}(\boldsymbol{x}_i) - \boldsymbol{Z}(\boldsymbol{x})}{b^q} \right\|^2 \right) \tag{4.34}$$

If using, adaptive bandwidth selection, the density estimator reduces to

$$\hat{f}(\boldsymbol{p} = \{\boldsymbol{x}, \boldsymbol{Z}(\boldsymbol{x})\}) = \frac{1}{n} \sum_{i=1}^{n} \frac{C}{(b_i^s)^2(b_i^q)^2} k^s \left(\left\| \frac{\boldsymbol{x}_i - \boldsymbol{x}}{b_i^s} \right\|^2 \right) k^q \left(\left\| \frac{\boldsymbol{Z}(\boldsymbol{x}_i) - \boldsymbol{Z}(\boldsymbol{x})}{b_i^q} \right\|^2 \right) \tag{4.35}$$

where b_i^s and b_i^q are the spatial and interval adaptive bandwidths respectively. For kernel regression, the normalization constant introduced to both numerator and denominator will cancel out.

Before defining the spatial interval kernel regressions, we review the basics of spatial kernel regression for real-valued data.

4.7 Spatial Real-valued Kernel Regression

Spatial real-valued kernel regression assumes existence of a smooth bivariate mean function $m(\boldsymbol{x}_i)$ relating response vector Z_i to location predictor $\boldsymbol{x}_i \in D, i = 1, \ldots, n$, such that

$$Z(\boldsymbol{x}_i) = m(\boldsymbol{x}_i) + \boldsymbol{e}(\boldsymbol{x}_i) \tag{4.36}$$

with $E[\boldsymbol{e}(\boldsymbol{x}_i)] = 0$ and $\text{Cov}(\boldsymbol{e}(\boldsymbol{x}_i), \boldsymbol{e}(\boldsymbol{x}_j)) = \sigma^2\boldsymbol{\phi}(||\boldsymbol{x}_i - \boldsymbol{x}_j||; \delta)$ a non-negative covariance function. The parameter δ controls the range of spatial association. Kern II (2001) has detailed the relationship between kernel functions and the covariance (or the correlogram) functions, identifying a one-to-one relationship between kernel and covariance functions for isotropic spatial processes.

Therefore, for stationary isotropic spatial processes, application of kernel approach is analogous to modeling $\boldsymbol{\phi}(\cdot)$ with a global isotropic covariance model parameterized with a global parameter $\boldsymbol{\theta}$, while the smooth mean function $m(\boldsymbol{x}_i)$ is modeled with an unknown global constant (i.e., $m(\boldsymbol{x}_i) = m$) in case of simple kriging, or unknown local constant in case of ordinary kriging.

For non-stationary process, we apply the kernel approach. Under the kernel approach, the local mean structure models both the trend and spatial structure; while any micro-scale covariance (called nugget effect in kriging) is part of the independent distributed errors. The mean function $m(\boldsymbol{x}_i)$ varies with location, with $E[\boldsymbol{e}(\boldsymbol{x}_i)] = 0$ and

$$\text{Cov}(\boldsymbol{e}(\boldsymbol{x}_i), \boldsymbol{e}(\boldsymbol{x}_j)) = \begin{cases} \sigma^2\boldsymbol{I} & i = j \\ 0 & i \neq j \end{cases} \tag{4.37}$$

In principle, we can say that, local kernel smoothing is equivalent to non-stationary covariance function modeling.

To obtain the underlying function $m(\boldsymbol{x})$, we proceed as follows: let $D \subseteq \mathbb{R}^2$ denote a two dimensional spatial region of interest; and $\boldsymbol{H}_{2\times 2}$ the two dimensional diagonal bandwidth matrix defining the extent of spatial dependence in the x and y direction. Assume that $\hat{m}(\boldsymbol{x}; \boldsymbol{H}, p)$ is the spatial regression estimate at location $\boldsymbol{x}$, with p the degree of the local polynomial. For $p \geq 1$, Rupper and Wand (1994) gives a good example.

Let $p = 0$, and for simplicity of notation we let $\hat{m}(\boldsymbol{x}; \boldsymbol{H}, p = 0) = \hat{m}(\boldsymbol{x})$. The estimate

$$\hat{m}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K(\boldsymbol{x}_i - \boldsymbol{x}; \boldsymbol{H}) Z(\boldsymbol{x}_i) \tag{4.38}$$

is the weighted average of $Z(\boldsymbol{x}_i)$ such that $\boldsymbol{x}_i$ lies in a region around location $\boldsymbol{x}$ defined by the kernel function K. With unevenly spaced data the estimate leads to poor results. We solve the problem using an alternative estimate, $\hat{m}(\boldsymbol{x})$, a solution to weighted least squares problem

$$\underset{\theta_x}{\text{argmin}} \sum_{i=1}^{n} K(\boldsymbol{x}_i - \boldsymbol{x}; \boldsymbol{H}) \left(Z(\boldsymbol{x}_i) - \theta_x\right)^2 \tag{4.39}$$

where θ_x is an assumed local regression fit of a constant of the unknown function, $m(\boldsymbol{x})$, and K is a smoother kernel function, e.g. the Gaussian kernel. The optimal local regression fit $\hat{m}(\boldsymbol{x})$ that minimizes 4.39, is given as

$$\hat{m}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} K(\boldsymbol{x}_i - \boldsymbol{x}; \boldsymbol{H}) Z(\boldsymbol{x}_i)}{\sum_{i=1}^{n} K(\boldsymbol{x}_i - \boldsymbol{x}; \boldsymbol{H})}, \tag{4.40}$$

which is the famous Nadaraya-Watson (NW) kernel estimator (Silverman, 1986). The estimate modifies 4.38 so that it is a true weighted average where the weights for each $Z(\boldsymbol{x}_i)$ will sum to one.

The bounded symmetric Beta family kernels (Epanechnikov, Bisquare, Tricube) are preferred since they ensure that only data, local to the point at which $m(\boldsymbol{x})$ is estimated, is used in the fit. This means that Gaussian kernel is less desirable, because although it is light in the tails, it is not zero, meaning in principle that the contribution of every point to the fit must be computed.

The approach used in 4.39 suggest similar estimators for multi-valued data, with expression $(Z(\boldsymbol{x}_i) - \theta_x)^2$ replaced with an appropriate metric in the input data space. Example is the kernel regression estimator for interval-valued data in (§4.8).

4.8 Spatial Interval Kernel Regression

Let the pairs $\{\boldsymbol{x}_i, \boldsymbol{Z}_i\}_{i=1}^n$, denote the interval-valued measurement $\boldsymbol{Z}_i$ at each location $\boldsymbol{x}_i \in D \subset \mathbb{R}^2$. The spatial kernel regression function 4.36 reduces to a spatial interval kernel regression function:

$$\boldsymbol{Z}(\boldsymbol{x}_i) = \boldsymbol{m}(\boldsymbol{x}_i) + \boldsymbol{e}_i(\boldsymbol{x}_i), \tag{4.41}$$

with $\boldsymbol{m}(\boldsymbol{x}_i)$ the nonparametric interval regression function at location $\boldsymbol{x}_i$, and $\boldsymbol{e}_i(\boldsymbol{x}_i)$ the independent distributed (iid) random interval errors. The nonparametric interval regression function $\boldsymbol{m}(\boldsymbol{x}_i)$ models both the trend and the spatial correlation; while any micro-scale variation is part of the iid errors such that $E\boldsymbol{e}_i(\boldsymbol{x}_i) = [0, 0]$ with

$$\text{Cov}(\boldsymbol{e}(\boldsymbol{x}_i), \boldsymbol{e}(\boldsymbol{x}_j)) = \begin{cases} [\sigma^2 \boldsymbol{I}, \sigma^2 \boldsymbol{I}] & i = j \\ [0, 0] & i \neq j \end{cases} \tag{4.42}$$

The goal is to estimate the nonparametric interval regression function $\boldsymbol{m}(\boldsymbol{x}_i)$ conditional on the available interval data $\boldsymbol{Z}(\boldsymbol{x}_i)$. The basic approach is to assume that

the variance and mean of the interval random field $\boldsymbol{Z}$ exists and are constant. Then, a good estimate of the process at any given location $\boldsymbol{x}_i$ is the average of all samples given by $\hat{\boldsymbol{m}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i$. A common characteristics of the estimator, $\hat{\boldsymbol{m}}$, is that it minimizes the mean prediction error (mpe)

$$E(\hat{\boldsymbol{m}}) = \sum_{i=1}^{n}(\hat{\boldsymbol{m}} - \boldsymbol{Z}_i)^2$$

The mpe, however, is not localized for similarity in the interval structure, all data values have equal importance relative to estimate. It leads to very smooth estimates which are not robust to outliers. We need to define a more robust error norm to take into account outliers, errors and any form of discontinues in the process. Let, $\hat{\boldsymbol{m}}_i = \hat{\boldsymbol{Z}}_i$ denote the estimate at location $\boldsymbol{x}_i$. Rather than minimize the mpe, we minimize a robust uncertainty function, $\boldsymbol{\phi}(\cdot)$, given as

$$E(\hat{\boldsymbol{m}}_i) = \sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.43}$$

Alternatively, we can minimize

$$E(\hat{\boldsymbol{m}}_i) = \sum_{i=1}^{n}\sum_{j=1}^{n}||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2\boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.44}$$

with residual error $||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2$ weighted by the robust uncertainty function, $\boldsymbol{\phi}(\cdot)$. Each $\boldsymbol{Z}_j$ is assigned a certainty measure $\boldsymbol{\phi}(\cdot)$ for its reliability in the prediction.

Estimate, $\hat{\boldsymbol{m}}_i$ at a given location $\boldsymbol{x}_i$ is obtained by solving the objective functions

$$\underset{\hat{\boldsymbol{m}}_i}{\text{argmin}}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.45}$$

or

$$\underset{\hat{\boldsymbol{m}}_i}{\text{argmin}}\sum_{i=1}^{n}\sum_{j=1}^{n}||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2\boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.46}$$

using gradient descent method. The method requires initial estimate of $\hat{\boldsymbol{m}}_i$ and it can be solved by an iterative weighted least squares minimization. Choice of initial estimate is critical; although the weighted median is generally a robust choice as an initial estimate, the mean is also often used.

Let $\hat{\boldsymbol{m}}_i^k$ denote the initial estimate at location $\boldsymbol{x}_i$; the gradient decent minimization of Eq. 4.45 is

$$\begin{aligned}
\hat{\boldsymbol{m}}_i^{k+1} &= \hat{\boldsymbol{m}}_i^k - \nu \frac{\partial E}{\partial \hat{\boldsymbol{m}}_i^k} \\
&= \hat{\boldsymbol{m}}_i^k - \nu \sum_{j=1}^{n} \phi'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) 2(\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j) \qquad (4.47) \\
&= \Big(1 - 2\nu \sum_{j=1}^{n} \phi'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2)\Big) \hat{\boldsymbol{m}}_i^k + 2\nu \sum_{j=1}^{n} \phi'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) \boldsymbol{Z}_j
\end{aligned}$$

with ν the step size chosen so that we do not take too big or too small of a step. Too big of a step will overshoot the function minimum, and too small of a step will result in a long convergence time. ν defines the convergence criteria for solving (4.47). Based on, the first term of 4.47, an optimal value of ν is set to,

$$\nu = \frac{1}{2 \sum_{j=1}^{n} \phi'(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2)} \qquad (4.48)$$

If the derivative of $\phi(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2)$ exists for $\hat{\boldsymbol{m}}_i \in [0, \infty)$, then, $\boldsymbol{g}(\cdot) = -\phi'(\cdot)$ is a symmetrical positive definite function belonging to a class of radial symmetrical kernels (Cheng, 1995). $\boldsymbol{g}(\cdot)$ forms the profile of the shadow kernel to the kernel with the profile $\phi(\cdot)$. Applying (4.48), (4.47) becomes

$$\hat{\boldsymbol{m}}_i^{k+1} = \frac{\sum_{j=1}^{n} \boldsymbol{g}(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) \boldsymbol{Z}_j}{\sum_{j=1}^{n} \boldsymbol{g}(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2)} \qquad (4.49)$$

The estimate $\hat{\boldsymbol{m}}_i^{k+1}$ is the interval valued Nadaraya-Watson estimator at location $\boldsymbol{x}_i$, with $\boldsymbol{g}$ an interval kernel function. Possible choices of interval kernel functions are defined in §4.9, Table 4.3.

The weight function $\boldsymbol{g}(\cdot)$ minimizes the influence of outliers caused by occasional miss-registration. The kernel function adapts to local interval structures. This leads to more samples of the same modality being gathered for the analysis. Essentially, the function assigns low weights to potential outliers effectively excluding them from the analysis. For example, for a Gaussian kernel, the bandwidth parameter σ^2 defines acceptable range of the residual error $||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||$. Samples with residual error less that σ^2 get a certainty close to one, whereas those with residual error

larger that $2\sigma^2$ gets extremely low certainty.

Similarly, solution to 4.46, is modified Nadaraya-Watson estimate given as

$$\hat{\boldsymbol{m}}^{k+1} = \frac{\sum_{j=1}^{n} g(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) \left[1 + ||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2\right] \boldsymbol{Z}_j}{\sum_{j=1}^{n} g(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) \left[1 + ||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2\right]} \tag{4.50}$$

Estimates 4.49 and 4.50 are relatively the same, but the last one gives slightly narrow bounds. The estimate, $\hat{\boldsymbol{m}}_i^{k+1}$, can be regarded as the unconditional estimate of $\boldsymbol{Z}_i$, and easily obtained using the idea of mean shift procedure, which has been studied for the unconditional mode estimation (Comaniciu and Meer, 2002; Comaniciu et al., 2000, 2001). To solve 4.50 we use the algorithm in Appendix C.1 implemented in C++.

However, the method ignores the fact that natural spatial data is often comprised of directional structures, and that the derived random fields can be integrated along these structures to improve their estimation. It is also a fact that observations within a local spatial neighborhood are approximately similar, (Tobler's First Law of geography (Tobler (1970)). Hence, rather than having both i and its neighborhood points j running from 1 to n, we define the local neighborhood j of i to run from 1 to $n(i)$. This will allows us to exploit the local spatial structure leading to an optimal conditional estimate.

The conditional estimate is obtained by localizing the robust uncertainty function to the local spatial structure using an additional weight function $\boldsymbol{w}(\cdot)$ such that

$$E(\hat{\boldsymbol{m}}) = \sum_{i=1}^{n} \sum_{j}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.51}$$

with $n(i)$ is the local spatial neighborhood of $\boldsymbol{x}_i$, and $\boldsymbol{w}(\cdot)$ is some weight function, usually a kernel function, not depending on $\boldsymbol{Z}_j$. The weight function $\boldsymbol{w}(\cdot)$ forms the spatial kernel that models the local spatial structure between sample observations (Anselin et al., 2004; Brunsdon et al., 1998b; Cressie, 1993). If $\boldsymbol{w}(\cdot) = 1$ we obtain the unconditional estimate in Eq. 4.49.

An alternative spatial localized robust uncertainty function is defined as

$$E(\hat{\boldsymbol{m}}) = \sum_{i=1}^{n} \sum_{j}^{n(i)} ||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2 \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.52}$$

The residual error $||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2$ is weighted by a product of two kernels, the uncertainty kernel, $\boldsymbol{\phi}(\cdot)$, and spatial kernel, $w(\cdot)$. Both systems 4.51 and 4.52 defines a local spatial weighted robust loss function. The first kernel $\boldsymbol{w}$ is the traditional spatial weighted regression kernel with KNN ($n(i)$) bandwidth, defining the local spatial structure. Within the defined spatial window, a second kernel, $\boldsymbol{\phi}$, the robust uncertainty kernel based on the interval distance between the neighbors and prediction point values is defined. Neighbors with greater geographical distance from prediction point are assigned smaller weight; and at a given geographical distance neighbors with similar sizes to the initial estimate at prediction point are assigned large weights.

The estimate is obtained by solving the objective functions

$$\underset{\hat{\boldsymbol{m}}_i}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.53}$$

or

$$\underset{\hat{\boldsymbol{m}}_i}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j}^{n(i)} ||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2 \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}(||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2) \tag{4.54}$$

using gradient descent method. An initial estimate of $\hat{\boldsymbol{m}}_i$ is required. Besides using the weighted median and the mean, close sample (in geographical distance) are sometimes used instead. The latter is applicable when minute details are of interest. Let $\hat{\boldsymbol{m}}_i^k$, denote the initial estimate, then

$$\begin{aligned}
\hat{\boldsymbol{m}}_i^{k+1} &= \hat{\boldsymbol{m}}_i^k - \nu \frac{\partial E}{\partial \hat{\boldsymbol{m}}_i^k} \\
&= \hat{\boldsymbol{m}}_i^k - \nu \sum_{i=1}^{n} \sum_{j=1}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) \times 2(\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j) \\
&= \left(1 - 2\nu \sum_{j=1}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2)\right) \hat{\boldsymbol{m}}_i^k \\
&\quad + 2\nu \sum_{j=1}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2) \boldsymbol{\phi}'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2) \boldsymbol{Z}_j
\end{aligned} \tag{4.55}$$

with ν, the step size chosen optimally to avoid too big or too small of a step to

convergence. The optimal ν based on data is given as

$$\nu = \frac{1}{2\sum_{j}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)\boldsymbol{\phi}'(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2)} \tag{4.56}$$

Let $\boldsymbol{g}(\cdot) = -\boldsymbol{\phi}'(\cdot)$ is a kernel belonging to a class of radial symmetric kernels (Cheng, 1995). Using the $\boldsymbol{g}$ and Eq. 4.56, reduces 4.55 to a spatially weighted interval estimator, parallel to the Nadaraya-Watson kernel estimator given as

$$\hat{\boldsymbol{m}}_i^{k+1} = \frac{\sum_{j}^{n(i)} \boldsymbol{w}\left(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\right) \boldsymbol{g}\left(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2\right) \boldsymbol{Z}_j}{\sum_{j}^{n(i)} \boldsymbol{w}\left(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\right) \boldsymbol{g}\left(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2\right)} \tag{4.57}$$

with $\boldsymbol{w}$ a radial symmetrical spatial kernel, and $\boldsymbol{g}$ a radial symmetrical interval kernel.

Similarly, applying gradient descent minimization to 4.54, the corresponding estimate is given as

$$\hat{\boldsymbol{m}}^{k+1} = \frac{\sum_{j}^{n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)\boldsymbol{g}(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2)\left[1 + ||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2\right] \boldsymbol{Z}_j}{\sum_{j \in n(i)} \boldsymbol{w}(||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)\boldsymbol{g}(||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2)\left[1 + ||\hat{\boldsymbol{m}}_i^k - \boldsymbol{Z}_j||^2\right]} \tag{4.58}$$

The estimates 4.57 and 4.58 are relatively the same, but the second gives slightly narrow bounds. Solution to 4.57 is obtained iteratively using the algorithm in Appendix C.2, and implemented in C++.

The estimate 4.57 parallels Mrazek et al. (2004) "sigma-filter" in digital image smoothing, and is one step approximation to conditional mode estimator given by Einbeck and Tutz (2004). Atkinson (2004) has also explored similar approach for image classification in remote sensing. Application of 4.57 for spatial interval modeling allows us to identify regions of high intensity or hot spots while effectively smoothing out the process.

4.9 Model Selection: Kernels

Let $d_{ij}^S = \sqrt{||\boldsymbol{x}_i - \boldsymbol{x}_j||^2}$ denote the geographical distance between prediction point (i) and neighborhood locations (j), and h_i the spatial bandwidth correspond-

ing to the j^{th} nearest neighbor in the spatial domain. Similarly, let $d_{ij}^G = \sqrt{||\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j||^2}$ denote the distance between intervals at prediction point (i) and intervals neighborhood points (j), and b_i the interval bandwidth corresponding to the j^{th} nearest neighbor in the interval domain. For the Gaussian kernel, let σ_S and σ_G denotes the spatial domain and interval domain bandwidth parameters. Then, applicable radial symmetric kernels for the spatial and interval weights include Gaussian, Epanechnikov and Bisquare radial kernels (Table 4.3).

Table 4.3. Spatial and interval kernels profiles applicable for non-parametric spatial-interval kernel regression.

Kernel	Spatial kernel, $\boldsymbol{w}(\cdot)$	Interval kernel, $\boldsymbol{g}(\cdot)$
Gaussian	$\exp\left(-\frac{(d_{ij}^S)^2}{2\sigma_S^2}\right)$	$\exp\left(-\frac{(d_{ij}^G)^2}{2\sigma_G^2}\right)$
Bisquare	$\left(1-\left(\frac{d_{ij}^S}{h_i}\right)^2\right)^2$	$\left(1-\left(\frac{d_{ij}^G}{b_i}\right)^2\right)^2$ for $d_{ij}^S \leq h_i, d_{ij}^G \leq b_i$
Epanechnikov	$1-\left(\frac{d_{ij}^S}{h_i}\right)^2$	$1-\left(\frac{d_{ij}^G}{b_i}\right)^2$ for $d_{ij}^S \leq h_i, d_{ij}^G \leq b_i$
$d_{ij}^S = \sqrt{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}$	$d_{ij}^G = \sqrt{\|\hat{\boldsymbol{m}}_i - \boldsymbol{Z}_j\|^2}$	
σ_S^2–spatial bandwidth	σ_G^2–the interval bandwidth	

Except for the Gaussian kernel that extends to infinity in all directions, the rest have a circumscribed radius, which is also the bandwidth, and is determined using k nearest neighbors. Intuitively, this leads to adaptive kernels, with large bandwidths in sub regions with scarce data and small bandwidth in regions of high intensity.

The Gaussian spatial smoothing scale, σ_S, is the geographical distance beyond which there is no spatial association relative to the prediction point. The scale σ_S plays a decisive role in the quality of smoothing. It is directly related to the local variability of the data: the more variable the data is the smaller it should be. Therefore, the scale must be large enough to cover sufficient samples for a stable local analysis. Unless the sample density is high everywhere in the region, a normal choice of spatial weight function is a Gaussian kernel function with a scale equivalent of variance. The Gaussian function introduces minimal smoothing while its support is still large to cover enough samples.

The Gaussian interval smoothing scale, σ_G is the interval bandwidth defined within the spatial neighborhood. The scale σ_G determines the degree of certainty of values within the spatial neighborhood. Values separated by σ_G are given zero certainty and are regarded as untrustworthy data. Both σ_S and σ_G may be defined using the variance of the samples in the respective domains.

4.10 Discussion

In this chapter we have developed a framework for non-parametric modeling of spatial random intervals. Unlike random interval kriging (Chapter 3), the underlying stochastic process is not assumed to be stationary. The approach utilizes a spatial-interval kernel, a product of spatial adaptive radial kernel and a robust interval radial kernel. The spatial kernel ensures that only samples sharing similar intensity and gradient information are gathered for local approximation. The robust interval kernel minimizes the influence of outliers or extreme values that significantly differ from the local neighborhood values. The choice of initial estimate at the prediction location adds flexibility to the approach. Use of close sample (in geographical distance) applies when minute details are of interest. Otherwise, the average or weighted median of neighborhood values is used. This approach adapts well depending on the nature of the application .

With obtained interval predicted values at unknown location, the approach offers a decision maker a possible bounded range of values from which to base their decision.

CHAPTER 5

BI-VARIABLE COPULA-BASED GEOSTATISTICS APPLICATION TO SPATIAL RANDOM INTERVAL DATA

Scope

In this chapter, we propose an approach that utilizes joint distribution function to model interval uncertainty (imprecision) in spatial random interval data. Specifically, we extend the bivariate copula approach to model the joint distribution using its copula and its marginals. The copula models the dependence between random variables without influencing information on the marginals, and hence is invariant under any monotonic transformation of the associated marginals. Therefore, frequently used data transformations (e.g. natural logarithms) do not influence the copulas. This makes copula(s) attractive models for the dependence structure in random functions characterized by multi-point properties. We investigate the theoretical and practical application of bivariate copula and extend it to the modeling of spatial random intervals. Integrating copula logic, geostatistics and GIS, we derive corresponding marginal and copula prediction maps which are estimates of the equivalent spatial distribution functions. In this new direction, the copula methodology forms a significant new technique to handle the co-movement of the random interval features in a given spatial regions. With copula map we can identify the frequency and persistence of high or lows values of the observed fuzzy phenomena across space. Risk areas are easily identified.

5.1 Introduction

Random intervals form the most basic representation of fuzzy (imprecise) data. In chapter 2, we showed random intervals to be random bivariate variables. The fact that vector presentation of intervals is in nature a bivariate random variable analysis, suggests for application of bivariate joint probability distribution of a pair of random variables as a model of the random interval data. Within spatial context, random interval data is also dependent across space. Therefore, the challenge is how to effectively handle the joint distribution plus the spatial arrangements.

The bivariate probability distribution of a pair of random variables contains the full information on the structure of bivariate random variable. For example, the full information on the random interval $Z = (L, U)$ can be modeled by the joint distribution $F_{LU}(l, u)$ of the lower (L) and upper (U) variables. Observe that each of the lower and upper variables is fully described by its cumulative distribution function $F_L(l) = P(L \leq l)$ and $F_U(u) = P(U \leq u)$ (the so-called marginals). However, and this is important to note, the marginals give us no information about the joint behavior. If the variables L and U are independent, the joint distribution function is simply the product of the marginals, $P(L \leq l; U \leq u) = F_L(l) \times F_U(u)$. Hence, to obtain a full description of random variables L and U together, we use the two ingredients: the marginals and the type of interrelation, in this case independence.

The question is: Can this kind of separation between marginals and dependence also be realized in a more general framework? The right concept for this is the bivariate copulas, which are joint distribution functions with uniform marginals of the bivariate random variable. Copulas describe the dependence structure between random variables without the information on the marginal distributions. Hence, they can be seen as the essential representation of dependence between random variables over the range of quartiles. Intuitively then, copulas allows to express whether the corresponding dependence is different for different quartiles of the variable. For example, high values may exhibit strong dependence and low values weak dependence. Unlike correlation-based inference, the copula extracts the way in which variables co-move, regardless of the scale with which the variables are measured.

Copulas have been widely applied in the financial sector (Embrechts et al., 2002; Hu, 2004; Mikosch, 2003). Schmitz (2003a,b) have explored the relationship between copulas and stochastic processes. A relationship between copulas and markov process is given by Darsow et al. (1992). Application of copulas in the spatial context is a recent concept (Bárdossy, 2006; De Michele and Salvadori, 2003), and is only

applied in the case of scalar-valued regionalized variables where correlation structure is estimated using a copula function rather than the variogram. This emphasizes need for new models of spatial dependence structure.

In this chapter, we integrate copula logic, geostatistics and GIS for spatial prediction on regionalized random intervals. The copula logic splits the joint distribution into its copula and its marginal distribution functions. Based on thresholds that correspond to values observed in the sample, each spatial random interval is assigned a copula value (or grade) based on derived empirical copula and marginal values (grades), derived from corresponding marginal values. Generated marginal and copula vectors are then treated as random functions. Thus variographic analysis and ordinary kriging is applicable resulting in prediction maps of the equivalent spatial distribution functions.

The copula map indicates the frequency and persistence of large or small values of phenomena across the spatial domain. The marginal maps independently show the spatial pattern of the lower and upper vector values. Unique estimates at unknown location are obtained by combining the copula and the marginal maps.

This chapter is structured as follows:- The bivariate copula is introduced in §5.2. Section 5.3 address the methods of fitting copulas to observed data. Section 5.4 details the integration of copula logic and geostatistics for spatial prediction on regionalized random intervals. Spatially distributed copula and marginal measures (referred to as sample spatial copula and marginal grades) are defined in §5.4.1. In §5.4.2 we define the copula and marginal variogram models and the copula and marginal kriging.

5.2 Bivariate Copula Basic Concepts

Let (X, Y) denote a bivariate random variable with random variables X and Y defined on a common probability space $(\Omega, \mathcal{A}, P)$ where Ω is a non-empty set, $\mathcal{A}$ a σ-field consisting of some subsets of Ω, and P a probability measure on $\mathcal{A}$. The cumulative distribution functions (the marginals) of X and Y are defined as $F_X(x) = \Pr(X \leq x)$, $x \in \mathbb{R}$, and $F_Y(y) = \Pr(Y \leq y)$, $y \in \mathbb{R}$ respectively. Their joint distribution function is defined as $F_{XY}(x, y) = \Pr[X \leq x, Y \leq y]$. Hence, to each possible pair of real numbers (x, y) we can associate three distributional measures: - $F_X(x)$, $F_Y(y)$ and $F_{XY}(x, y)$, which all lie in the unit interval $\mathbb{I} = [0, 1]$. Therefore we can observe that each pair of real numbers (x, y) will lead to a point $(F_X(x), F_Y(y))$ in the unit plane $\mathbb{I}^2$.

Now, a function $C : \mathbb{I}^2 \rightarrow \mathbb{I}$ given as

$$C(F_X(x), F_Y(y)) = F_{XY}(x, y) \tag{5.1}$$

is called a two-dimensional copula. In other words, a two-dimensional copula implies a couple of two univariate distributions. General concept of copulas and developments can be found in Cherubini et al. (2005); Joe (1997) and Nelsen (1999). For the properties of the two-dimensional copula, we first review some definitions given by Nelsen (1999).

Definition 5.2.1 [H-VOLUME] Let S_1 and S_2 be nonempty subsets of $\overline{\mathbb{R}}$, and let H be a function such that the domain of $H = S_1 \times S_2$. We denote by $Dom(H)$ the domain of H, while $Ran(H)$ denotes its range. Thus $Dom(H) = S_1 \times S_2$. Let $B = [x_1, x_2] \times [y_1, y_2]$ be a rectangle all of whose vertices are in $Dom(H)$. The H-Volume of B is given by:

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

Definition 5.2.2 [TWO-INCREASING] A two-variate real-valued function H is two-increasing if $V_H(B) \geq 0$ for all rectangles B whose vertices lie in $Dom(H)$.

Definition 5.2.3 (GROUNDED) Suppose $a_1 = \min\{S_1\}$ and $a_2 = \min\{S_2\}$. We say that a function H:$S_1 \times S_2 \rightarrow \mathbb{R}$ is grounded if $H(x, a_2) = H(a_1, y) = 0$ for all (x, y) in $S_1 \times S_2$.

Properties of a two-dimensional copula

Summary properties of the two-dimensional copula defined in Eq. 5.1 are:

1. $Dom(C) = \mathbb{I}^2$;
2. C is grounded and two-increasing;
3. For every w in S_1 and every v in S_2; $C(w, 1) = w$ and $C(1, v) = v$ The values w and v form the marginal values for specifying the cumulative distribution functions respectively.

Remark 5.2.1 Note that $0 \leq C(w, v) \leq 1$ for every (w, v) in $Dom(C)$. Therefore, $Ran(C) \subset \mathbb{I}$.

Theorem 5.2.1 *(FRECHET HOEFFDING BOUNDS) Let C be a bivariate copula. Then for $\forall w, v \in \mathbb{I}$,*

$$Q(w,v) = max(w+v-1,0) \leq C(w,v) \leq min(w,z) = M(w,v)$$

The bivariate functions Q and M are called lower and upper Fréchet Hoeffding bounds.

Proof. Let (w,v) be an arbitrary point in $Dom(C)$. Recall that

$$C(w,v) \leq C(w,1) = w$$

and

$$C(w,v) \leq C(1,v) = v.$$

Hence

$$C(w,v) \leq \min(w,v).$$

Furthermore, $V_c([w,1] \times [v,1]) \geq 0$ implies $C(w,v) \geq w+v-1$ since

$$\begin{aligned} V_c([w,1] \times [v,1]) &= C(1,1) - C(1,v) - C(w,1) + C(w,v) \\ &= 1 - v - w + C(w,v) \Rightarrow C(w,v) \geq v + w - 1 \geq 0 \\ &\qquad \text{by definition of copula} \end{aligned}$$

Therefore $C(w,v) \geq \max(w+v-1,0)$. Hence we observe that copula functions have both upper and lower bounds. ■

The importance of copulas to mathematical statistics is described in

Sklar's Theorem (Sklar, 1959): *Given random variables X and Y with joint distribution function F_{XY} and marginals F_X and F_Y, respectively, there exists a copula C such that, for all $x, y \in \bar{\mathbb{R}}$,*

$$F_{XY}(x,y) = C(F_X(x), F_Y(y))$$

If the marginal distributions are continuous, then the copula C is unique; otherwise, C is uniquely determined on $Ran(F_X) \times Ran(F_Y)$. Conversely if C is a copula and F_X and F_Y are marginal distribution functions, then the function F_{XY} is a joint distribution function with marginals F_X and F_Y.

A detailed proof of Sklar' Theorem is found in Nelsen (1999). While the copula can be regarded as the pure expression of the dependence without the influence of

the marginal distributions, it is itself a bivariate distribution with uniform marginals on $\mathbb{I}$. The bivariate distribution of random variables is completely determined by its copula and its marginal distributions. For independent random variables, we have the product copula given as $C(w,v) = wv$, while for dependent random variables the copula becomes $C(w,v) = \min(w,v)$.

Remark 5.2.2 Several applications involve the shape of the graph of the copula i.e., of the surface $z = C(w,v)$. The surface can be viewed as the joint distribution function i.e. $z = F_{XY}(x,y)$ in which the x and y axes have been relabeled in units of $w = F_X(x)$ and $v = F_Y(y)$

Copulas, similarly to distributions functions, admit the notion of density (Bárdossy, 2006; Cherubini et al., 2005; Käärik, 2006). By derivation, the copula density $c(w,v)$ for the copula $C(w,v)$ is given as

$$c(w,v) = \frac{\partial^2 C(w,v)}{\partial w \partial v} \tag{5.2}$$

In terms of probability density functions (Cherubini et al., 2005, p. 66),

$$\begin{aligned} f_{XY}(x,y) &= \frac{\partial^2 C(F_X(x),F_Y(y))}{\partial x \partial y} = \frac{\partial^2 C(w,v)}{\partial w \partial v}\frac{\partial F_X(x)}{\partial x}\frac{\partial F_Y(y)}{\partial y} \\ &= c(w,v) f_X(x) f_Y(y) \end{aligned} \tag{5.3}$$

with c the copula density and f the univariate probability density functions. This is an important result, because it states that under appropriate conditions, the joint density can be written as a product of the marginal densities and the copula density. If for example the random variables X and Y are independent, then $c(w,v) = 1$ and

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

which is a familiar formula for independent variables.

The generalized inverse concept given by Nelsen, 1999, p.19 provides a method of reconstructing a copula from the margin's and the joint distribution. Let F_X^{-1} and F_Y^{-1} be quasi-inverses of F_X and F_Y respectively. Then for any (w,v) in $Dom(C)$,

$$C(w,v) = F_{XY}(F_X^{-1}(w), F_Y^{-1}(v)). \tag{5.4}$$

Example 5.2.1 *Assume a Gumbel's logistic bivariate distribution*

$$F_{XY}(X,Y) = \frac{1}{1+\exp(-x)+\exp(-y)}$$

with marginals

$$F_X(x) = \frac{1}{1+\exp(-x)} \text{ and } F_Y(y) = \frac{1}{1+\exp(-y)} \tag{5.5}$$

Applying 5.4, the bivariate copula is

$$C(w,v) = F_{XY}(F_X^{-1}(w), F_Y^{-1}(v)$$

where

$$F_X^{-1}(w) = \ln\frac{w}{1-w} \text{ and } F_Y^{-1}(v) = \ln\frac{v}{1-v}$$

Hence

$$\begin{aligned} C(w,v) &= F(F_X^{-1}(w), F_Y^{-1}(v)) \\ &= \frac{1}{1+\exp\left(-\ln\frac{w}{1-w}\right)+\exp\left(-\ln\frac{v}{1-v}\right)} \\ &= \frac{1}{1+\frac{1-w}{w}+\frac{1-v}{v}} \\ &= \frac{wv}{w+v-wv} \end{aligned}$$

Example 5.2.2 *Gaussian copula (Nelsen, 1999): Let N_ρ denote a standard bivariate normal distribution with Pearson's correlation coefficient ρ. Also let Φ denote the standard (univariate) normal distribution function with mean zero and unit variance. Then the Gaussian copula is given by*

$$C(w,v;\rho) = N_\rho\left(\Phi^{-1}(w), \Phi^{-1}(v)\right)$$

where $w, v \in [0,1]$.

In terms of Sklar's theorem, for any two marginal distribution functions F_X and F_Y, the distribution defined as

$$F_{XY}(x,y) = C(F_X(x), F_X(y);\rho) = N_\rho\left(\Phi^{-1}(w), \Phi^{-1}(v)\right)$$

is a bivariate distribution function whose marginals are F_X and F_Y respectively, and the copula that connects F_{XY} to F_X and F_Y is the Gaussian copula. Hence, Sklar's theorem allows one to construct bivariate distributions with non-normal marginal distributions and the Gaussian copula.

5.3 Fitting bivariate copulas to data

In this section we illustrate some of the methods in which parameters of copulas can be estimated. We are focusing on single-parameterized copulas (Joe, 1997, p.60). Firstly we illustrate the method of Maximum Likelihood estimation using Canonical Maximum Likelihood. Secondly a semi-parametric approach in which the marginal distributions functions are determined empirically and a parametric family of a copula is chosen.

5.3.1 Maximum Likelihood Estimation

Application of Canonical Maximum Likelihood is favorable since we do not need to assume any parametric families of marginal distributions and hence are free from any misspecification error.

Let X be a random variable and $\{x_1, x_2, \cdots, x_n\}$ be a random sample of size n drawn from X. The estimate to the true marginal distribution of the random variable X is

$$\hat{F}_X(x) = \frac{\#\{x_t \leq x\}}{n},$$

where $\#\{x_t \leq x\}$ refers to the number of observations satisfying the condition $x_t \leq x$ for $t = 1, 2, \cdots, n$. Similarly, let Y be a random variable and $\{y_1, y_2, \cdots, y_n\}$ be random sample of size n drawn from Y. The estimate of the true marginal distribution of the random variable Y is

$$\hat{F}_Y(y) = \frac{\#\{y_t \leq y\}}{n}$$

The likelihood function is defined as

$$L(\theta) = \prod_{t=1}^{n} \hat{c}(w_t, v_t; \theta),$$

where $w_t = \hat{F}_X(x)$, $v_t = \hat{F}_Y(y)$, $\hat{c}(w_t, v_t; \theta)$ is the density of the copula model, with θ the copula parameter (Joe, 1997).

Note here that:

$$c(w_t, v_t; \theta) = \left[\frac{\partial^2 C(w, v; \theta)}{\partial w \partial v}\right]_{(w_t, v_t)}$$

The alternative to maximum likelihood estimation, is to employ the Kernel Density Estimation method (Silverman, 1986), reviewed in chapter 4. In §5.3.2 we provide the kernel distribution estimator and an estimator of the kernel-based copula as applied in this work.

5.3.2 Kernel Distribution Estimator

Let X be a random variable in one-dimensional space and $\{x_1, x_2, \cdots, x_n\}$ be a random sample of size n drawn from X. The kernel density estimate takes the form

$$\hat{f}(x; b) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - x_i}{b}\right)$$

where b represents the bandwidth parameter, and K represent a univariate kernel function (see examples in table 4.1). For bandwidth selection methods, see §4.4.1. The empirical kernel distribution estimator of the true marginal distribution of the random variable X is

$$\hat{F}_X(x; b) = \int_{-\infty}^{x} \hat{f}(x; b)\, dx$$

Similarly, we can define the empirical distribution estimator of the true marginal distribution of the random variable Y.

Let $Z = (X, Y)$ be a bivariate random variable and $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ be a random sample of size n drawn from Z. The kernel density estimate of Z takes the form

$$\hat{f}(x, y; b) = \frac{1}{nb^2} \sum_{i=1}^{n} K(x - x_i, y - y_i; b) \tag{5.6}$$

where b is the bandwidth parameter, and K is a multivariate kernel function (see examples in table 4.2). We define the empirical bivariate kernel joint distribution estimator as

$$\hat{F}_{XY}(x, y; b) = \int_{-\infty}^{x} \int_{-\infty}^{y} \hat{f}(x, y; b)\, dy dx \tag{5.7}$$

Empirical kernel marginal distribution estimators can also be obtained from the joint density. For example, for random variable X, the kernel marginal distribution

estimator is defined as

$$\hat{F}_X(x;b) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} \hat{f}(x,y;b)\,dydx \tag{5.8}$$

and similarly for Y.

Example 5.3.1 *Let $Z = (Z^l, Z^u)$ denote the random interval (or the bivariate random variable) with $\{(z_1^l, z_1^u), (z_2^l, z_2^u), \cdots, (z_n^l, z_n^u)\}$ the random sample of size n drawn from $Z = (Z^l, Z^u)$. The kernel joint density estimate is*

$$\hat{f}\left(z^l, z^u; b\right) = \frac{1}{nb^2}\sum_{i=1}^{n} K\left(z^l - z_i^l, z^u - z_i^u; b\right) \tag{5.9}$$

The kernel joint distribution estimator becomes

$$\hat{F}_{Z^lZ^u}\left(z^l, z^u; b\right) = \int_{-\infty}^{z^l}\int_{-\infty}^{z^u} \hat{f}\left(z^l, z^u; b\right) dz^u dz^l \tag{5.10}$$

and the kernel marginal distribution estimator for Z^l is

$$\hat{F}_{Z^l}\left(z^l; b\right) = \int_{-\infty}^{z^l}\int_{-\infty}^{\infty} \hat{f}\left(z^l, z^u; b\right) dz^u dz^l \tag{5.11}$$

while for Z^u is given as

$$\hat{F}_{Z^u}\left(z^u; b\right) = \int_{-\infty}^{z^u}\int_{-\infty}^{\infty} \hat{f}\left(z^l, z^u; b\right) dz^l dz^u \tag{5.12}$$

The empirical kernel copula is obtained by estimating values based on the cdf at n distinct points in $\mathbb{R}^2$, and is given by

$$\hat{C}(w,v) = \hat{F}_{Z^lZ^u}\left(F_{Z^l}^{-1}, F_{Z^u}^{-1}\right) \tag{5.13}$$

where $F_{Z^l}^{-1}$ and $F_{Z^u}^{-1}$ are quasi-inverses of F_{Z^l} and F_{Z^u}.

The advantage of kernel estimation of the copula implies that we do not have to choose a copula from a large family of copulas.

The semi-parametric method utilizes a parametric family of a copula. Although, we must be careful in choosing a copula by considering the characteristics of the data such as tail dependence, asymmetry of the data, and non-normality, we can find the best choice of the copula by using well-known Akaike's Information Criterion (AIC)

or Bayesian information criteria (BIC) (Akaike, 1974; Schwarz, 1978). Since we are dealing with single-parameter copulas only, we have that:

$$AIC = -2l(\theta) + 2$$

and

$$BIC = -2l(\theta) + \ln(n)$$

where $l(\theta)$ is the log of the likelihood function. The best choice of the copula is one with lowest AIC or BIC value. Cherubini et al. (2005) and Joe (1997) have provided different range of copulas that can be applied.

5.4 Copula modeling of Spatial Random Intervals

For any given spatial locations $\boldsymbol{x}_i = (x_i, y_i), i = 1, \ldots, n$, in $D \subseteq \mathbb{R}^2$, $\boldsymbol{Z} = (Z^l, Z^u) = \left(\left(Z_1^l(\boldsymbol{x}_1), Z_1^u(\boldsymbol{x}_1)\right), \left(Z_2^l(\boldsymbol{x}_2), Z_2^u(\boldsymbol{x}_2)\right), \cdots, \left(Z_n^l(\boldsymbol{x}_n), Z_n^u(\boldsymbol{x}_n)\right) \right)$ is a random vector of regionalized bivariate random variables. Separately, $Z^l = \{Z^l(\boldsymbol{x}_i)\}_i^n$ represents the random vector of lower regionalized random variables while $Z^u = \{Z^u(\boldsymbol{x}_i)\}_i^n$ is the random vector of upper regionalized random variables. The bivariate random variable (Z^l, Z^u) assumes vector values $\{(z_i^l, z_i^u)\}_i^n = \left(z^l(\boldsymbol{x}_i), z^u(\boldsymbol{x}_i)\right)_i^n$, i.e.

$$(Z^l, Z^u) = (z_i^l, z_i^u) \tag{5.14}$$

We denote by $F_{Z^l}(z^l) = P(Z^l \leq z^l)$ and $F_{Z^u}(z^u) = P(Z^u \leq z^u)$ the theoretical univariate cumulative distribution functions of the lower and upper values respectively. $F_{Z^l Z^u}(z^l, z^l) = P(Z^l \leq z^l, Z^u \leq z^u)$ denotes the joint cumulative distribution function.

Copula approach splits the joint distribution, $F_{Z^l Z^u}(z^l, z^l)$ into its copula, $C\left(F_{Z^l}(z^l), F_{Z^u}(z^u)\right)$ and its marginals $F_{Z^l}(z^l)$ and $F_{Z^u}(z^u)$. The copula and the marginals are smooth differentiable distribution functions monotonically increasing from 0 to 1, and can be estimated from the data.

5.4.1 Spatially Distributed Copula and Marginal Grades

In classical geostatistics, indicator kriging is based on a preliminary coding of each observation $z(\boldsymbol{x}_i)$ into a vector of indicators (Eq. 5.15), defined for a set of K

thresholds z_k discretisizing the range of variation of the attribute:

$$\mathbf{I}(\boldsymbol{x}_i; z_c) = \begin{cases} 1 & z(\boldsymbol{x}_i) \leq z_k, k = 1, \ldots, K \\ 0 & \text{otherwise} \end{cases} \tag{5.15}$$

As a generalization to indicator function, the copula approach utilizes the joint distribution(copula) and the marginals for coding each observation $z(\boldsymbol{x}_i) = (z_i^l, z_i^u)$ into three measures (5.16),

$$\begin{aligned} F_{Z^l}(z_i^l) &\subset (0,1) \\ F_{Z^u}(z_i^u) &\subset (0,1) \\ C(w_i, v_i) &\subset (0,1), \text{ with } w_i = F_{Z^l}(z_i^l) \text{ and } v_i = F_{Z^u}(z_i^u) \end{aligned} \tag{5.16}$$

for all thresholds corresponding to values observed in the sample. Hence, the copula kriging approach can be regarded as indicator kriging of the marginal and joint (copula) probabilities when all data are retained in the kriging system.

Definition 5.4.1 *For a given bivariate random variable (Z^l, Z^u) with copula function $C(w, v)$ and the marginal distributions F_{Z^l} and F_{Z^u} for Z^l and Z^u respectively, if $(Z^l, Z^u) = (z_i^l, z_i^u)$ at location $\boldsymbol{x}_i = (x_i, y_i)$, then $C(w_i, v_i)$ is called the copula grade for bivariate random variable (Z^l, Z^u) at location $\boldsymbol{x}_i$, and similarly $F_{Z^l}(z_i^l)$ and $F_{Z^u}(z_i^u)$ the marginal distribution grades at location $\boldsymbol{x}_i$ respectively.*

The collection of spatially distributed copula grades, denoted as

$$\left\{C(w_i, v_i),\ \boldsymbol{x}_i \in D \subset \mathbb{R}^2,\ i = 1, 2, \cdots, n\right\} \tag{5.17}$$

is called sampled copula grades over region D and corresponds to a copula random field. Accordingly, the collections

$$\left\{F_{Z^l}(z_i^l),\ \boldsymbol{x}_i \in D \subset \mathbb{R}^2,\ i = 1, 2, \cdots, n\right\} \tag{5.18}$$

and

$$\left\{F_{Z^u}(z_i^u),\ \boldsymbol{x}_i \in D \subset \mathbb{R}^2,\ i = 1, 2, \cdots, n\right\} \tag{5.19}$$

are called the sampled lower and upper marginal distribution grades over region D. Respectively, they correspond to the lower and upper marginal random fields.

The spatial copula and marginal random fields may also be respectively referred to as COPULA GRADE MARKED POINT PROCESS *and* MARGINAL GRADE MARKED

POINT PROCESS. *The "points" are the random vector spatial locations and the "marks" are copula and the marginal grades.*

Since the copula distribution of a bivariate random variable contains the full information on the dependence structure of the bivariate random variable, it is reasonable to say that the copula grades reveal the information of bivariate random variable's spatial distribution, i.e., $\left\{\left(Z^l, Z^u\right), \boldsymbol{x}_i \in D\right\}$.

Just similar to the general treatment in spatial data analysis, the available information on $\left(Z^l, Z^u\right)$ is the sampled copula grades over region D, so that we can perform the kriging on the sampled copula grades over region D and therefore obtain the copula grade predictor 5.20 at any unsampled location, $\boldsymbol{x}_0 \in D$.

$$\hat{C}(w_0, v_0) = \sum_{i=1}^{n(h)} \lambda_i C\left(w_i, v_i\right), \quad \sum_{i=1}^{n(h)} \lambda_i = 1 \tag{5.20}$$

$n(h)$ is the neighborhood sample and λ_i are the weights accounting for data configuration, the proximity of data to unsampled location $\boldsymbol{x}_0$, as well as the spatial pattern modeled by a semivariogram model.

For the unique determination of the estimate for $\left(z_0^l, z_0^u\right)$ at location $\boldsymbol{x}_0 \in D$, the copula grade map alone is not adequate although it is supposed to reveal the full distributional information on bivariate random variate $\left(Z^l, Z^u\right)$. The reason underlying is that, the copula kriging map ignores the marginal information (w_i, v_i) but keeps the coordinate information (x_0, y_0). Therefore, we need the additional information modeled by the lower and upper marginal grade predictors respectively defined as

$$\hat{F}_{Z^l}(z_0^l) = \sum_{i=1}^{n(h)} \lambda_i F_{Z^l}(z_i^l), \quad \sum_{i=1}^{n(h)} \lambda_i = 1 \tag{5.21}$$

$$\hat{F}_{Z^u}(z_0^u) = \sum_{i=1}^{n(h)} \lambda_i F_{Z^u}(z_i^u), \quad \sum_{i=1}^{n(h)} \lambda_i = 1 \tag{5.22}$$

Then estimate at location $\boldsymbol{x}_0$ is interpreted using the triplet $\left(w_0, v_0, , \hat{C}(w_0, v_0)\right)$ where $w_0 = \hat{F}_{Z^u}(z_0^u)$ and $v_0 = \hat{F}_{Z^u}(z_0^u)$. Despite, the copula kriging grade map still reveals the spatial distribution of the bivariate distribution of bivariate random variable $\left(Z^l, Z^u\right)$.

5.4.2 Copula and Marginal Grade Kriging

The basic stationarity assumptions for the copula and marginal grade kriging are similar to those of ordinary kriging (Chiles and Delfiner, 1999; Goovaerts, 1997). Parallel to the experimental semivariogram for ordinary and/or indicator kriging, we define the experimental copula grade semivariogram, $\hat{\gamma}_{C(w,v)}(h)$ as follows:

$$\hat{\gamma}_{C(w,v)}(h) = \frac{1}{2|n(h)|}\sum_{i=1}^{n(h)}\left[\left(C(w(\boldsymbol{x}_i+h), v(\boldsymbol{x}_i+h)) - C(w(\boldsymbol{x}_i), v(\boldsymbol{x}_i))\right)^2\right] \quad (5.23)$$

where h is the distance separation vector between spatial locations $\boldsymbol{x}_i$ and $\boldsymbol{x}_i + h$; $C(w(\boldsymbol{x}_i), v(\boldsymbol{x}_i))$ and $C(w(\boldsymbol{x}_i+h), v(\boldsymbol{x}_i+h))$ the sample copula grades at locations $\boldsymbol{x}_i$ and $\boldsymbol{x}_i+h$ derived using the empirical copula; and $n(h) = \{(\boldsymbol{x}_i, \boldsymbol{x}_i+h), i = 1, \ldots, n\}$, the set of data pairs separated by distance vector h, with $|n(h)|$ the cardinality.

Similarly, based on marginal grade values associated with the lower and upper marginal random functions, Eq.s 5.24 and 5.25 respectively denotes the lower and upper experimental marginal grade semivariograms.

$$\hat{\gamma}_{F_{Z^l}(z_i^l)}(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)}\left[\left(F_{Z^l}(\boldsymbol{x}_i+h) - F_{Z^l}(\boldsymbol{x}_i)\right)^2\right] \quad (5.24)$$

$$\hat{\gamma}_{F_{Z^u}(z_i^u)}(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)}\left[\left(F_{Z^u}(\boldsymbol{x}_i+h) - F_{Z^u}(\boldsymbol{x}_i)\right)^2\right] \quad (5.25)$$

As an ordinary kriging system on copula grades, the copula grade kriging system is defined as

$$\begin{aligned}\sum_{j=1}^{n(h)}\lambda_j\gamma_{C(w,v)}(\boldsymbol{x}_i-\boldsymbol{x}_j) - \psi &= \gamma_{C(w,v)}(\boldsymbol{x}_0-\boldsymbol{x}_i) \;\; i = 1,\ldots,n(h)\\ \sum_{j}^{n(h)}\lambda_j &= 1\end{aligned} \quad (5.26)$$

where ψ is the Lagrange multiplier. Solution to the system 5.26 leads to the weights for the copula predictor in 5.20.

Similarly, solution to lower marginal kriging system (5.27)

$$\begin{aligned}\sum_{j=1}^{n(h)}\lambda_j\gamma_{F_{Z^l}}(\boldsymbol{x}_i-\boldsymbol{x}_j) - \psi &= \gamma_{F_{Z^l}}(\boldsymbol{x}_0-\boldsymbol{x}_i) \;\; i = 1,\ldots,n(h)\\ \sum_{j}^{n(h)}\lambda_j &= 1\end{aligned} \quad (5.27)$$

results to weights for the lower marginal grade predictor (5.21), while solution to upper marginal kriging system (5.28)

$$\begin{aligned} \sum_{j=1}^{n(h)} \lambda_j \gamma_{F_{Z^u}}(\boldsymbol{x}_i - \boldsymbol{x}_j) - \psi &= \gamma_{F_{Z^u}}(\boldsymbol{x}_0 - \boldsymbol{x}_i) \;\; i = 1, \ldots, n(h) \\ \sum_{j}^{n(h)} \lambda_j &= 1 \end{aligned} \tag{5.28}$$

provides weights for upper marginal grade predictors in Eq. 5.22.

5.5 Discussion

In this chapter we have combined bivariate kernel copula and geostatistics for variable modeling of spatially distributed random intervals. Copulas possess desirable features which make them very useful for dependence analysis. We show how to use copulas to model random vector dependence structure prior to performing spatial prediction.

In the copula approach, the joint distribution of random intervals is replaced by its copula and its marginal distributions. Each spatial random interval is characterized using three distributional measures: the copula, lower and upper distributional measures (referred to as grades). Copula grades model the interval dependence structure without any disturbing effect coming form the marginal distributions. This provides better insight into the laws governing the random interval and consequently the spatial behavior of the spatial random interval process.

With defined copula and marginal grades at each spatial location, it is easier to perform kriging and generate the appropriate maps, which are estimates of the equivalent spatial distribution functions. The copula map indicates the frequency and persistence of the minimum and maximum values across space. The marginal maps independently show the predicted marginal grades of minimum and maximum.

As a summary, the bi-variable copula-based geostatistics is seen as an alternative of indicator kriging (IK). It is equivalent to indicator kriging of the marginal and joint probability when all data are retained in the kriging system. In addition to the advantage of jointly handling of data, the copula approach requires less parameterization as compared to IK. IK requires as many variograms as there are chosen threshold levels and often leads to order relation problems. The copula approach requires only the copula and marginal variograms, on average well behaved, brings no order relation problems, and transforms prediction with spatial random interval

into an easy task.

Because it is based on the distribution of the random interval data, the copula approach has additional advantages over alternative approaches i.e. interval kriging and interval kernel regression based on observed random interval data. The derived copula and analytical marginal models are forward-looking. They are capable of incorporating a wide range of future eventualities that simply are not captured using observed data. They do not require a large amount of data in order to be estimated accurately, and furthermore are instantly capable of reflecting a change in the observed phenomena. A sudden shift in observed phenomena and/or other related factors could be immediately captured in observed data and the implied bivariate distribution (copula + marginal distributions).

While there are dangers to over-inference from derived distribution, they can be applied to virtually any spatial phenomena and still provide several advantages over the methods based solely on spatial data e.g. interval kriging.

Defined copula-based kriging takes care of the interval uncertainty through the copula and facilitates a scalar variable kriging spatially distributed random interval data. The predicted copula grade map alone is not able to provide the realized value at locations because the copula is only a model of the interval dependence structure. We need the additional information modeled by the marginal distributions. This means that additional kriging on marginal distributional grades is necessary. In the next chapter, we develop a framework that utilizes a scalar fuzzy variable with a distribution function that models both randomness and imprecision in random interval data via conversion modeling.

CHAPTER 6

FUZZY CREDIBILITY-BASED GEOSTATISTICS: APPLICATION TO SPATIAL RANDOM INTERVAL DATA

Scope

This chapter defines an alternative fuzzy modeling approach for spatial random interval that is variable-based rather than set-based. Beginning with a highlight of Kaufmann 1975 approach, we generate a scalar fuzzy variable equivalent of the random interval data, with a maximum entropy data-assimilated membership function. Lius Credibility distribution theory (Liu, 2004) for modeling random fuzzy phenomena is then applied to membership function to define the distribution of the random fuzzy variable with an associated credibility measure. Ordinary kriging is then performed on fuzzy variable credibility grades.

6.1 Introduction to Fuzzy Variable Modeling of Random Intervals

Random fuzzy uncertainty is a common problem in spatial models where spatial data fuzzy (vague or imprecise) and often exist in form of intervals of values (random intervals). Accordingly, the optimal approach to model spatial random interval processes should be developed in terms of the basic concept of fuzzy mathematics (Kaufmann, 1975; Zadeh, 1965, 1978). However, mathematical treatments for uncertain phenomena at subset or event level, except for that in probability theory introduces complexity of having to deal with set-based operations and outputs,

Naturally, dealing with real-valued numbers is much easier than dealing with subsets or events. In standard probability theory, random variable and the associated distribution function play important roles of converting set-based arguments into variable-based arguments. Converting set-based arguments into variable-based arguments results in great conveniences in applications. As a mathematical operation, we would prefer to deal with real-valued numbers. We refer to such a modeling idea as "variable modeling". Thus, if we can have a variable oriented approach for every uncertainty case, then the modeling efforts will be greatly simplified.

Unfortunately, except for random uncertainty theory or probability theory, the mathematical treatments on other type of uncertain phenomena, say, fuzzy events, random interval-events, or random fuzzy events, lacked such consciousness of uncertainty variable modeling until the general uncertainty theoretical framework proposed by Liu (2004).

For example, the fuzzy mathematics initiated by Zadeh (1965, 1978) facilitated a foundation for dealing with vague phenomena, which was based on a membership function and possibility measure of fuzzy events. The possibility measure was designed to play the role of probability measure in probability theory. However, it failed to because, possibility measure does not possess self-duality property as that in probability theory. Kaufmann (1975) also proposed the concept of fuzzy variable with the intention to create a counterpart in probability theory. However, Kaufmanns fuzzy variable is just another name for a fuzzy subset and the mathematical operations are set based, difficult to handle. Existing fuzzy geostatistical approaches (Bandemer and Gebhardt, 2000; Bárdossy et al., 1988, 1990b; Burrough and McDonnell, 1998; Kacewiez, 1994), are based on fuzzy set operations and utilize fuzzy variograms generating fuzzy kriged values and fuzzy kriging variances, rather than real-valued predictions and error.

The axiomatic foundation proposed by Liu (2004), named as (standard) cred-

ibility theory (i.e.$(\vee, \wedge)-$credibility measure theory), provides a sound framework for scalar fuzzy variable modeling of fuzzy set, random fuzzy set, or fuzzy random set phenomena. The credibility measure possesses self-duality property and is able to play a similar role to that of a probability measure in probability theory. Thus, parallel to probability theory, where a random variable is modeled with a probability distribution function plus a probability measure, Lius Credibility theory models a random fuzzy set, or fuzzy random phenomena with a credibility distribution function plus a credibility measure. A review of credibility measure theory is given in sections §6.3. For a detailed review of classical and non-classical credibility theories readers should consult Guo and Li (2006).

We build on Liu (2004) credibility measure theory and adapt it to kriging with spatial random interval data. First, we note that random interval data constitute both random and fuzzy uncertainty. Applying an empirical kernel-based approach, the random interval is first converted to a fuzzy set with maximum entropy. Then, derived fuzzy set is represented as a scalar fuzzy variable under credibility measure theory. Associated with the scalar fuzzy variable is a maximum entropy data-assimilated membership function and its counterpart the data-assimilated credibility function, non-decreasing function from 0 to 1. This conversion simultaneously models both random and fuzzy uncertainty inherent in random intervals. The credibility distribution generalizes the principle of membership function, widely used in describing fuzzy phenomena.

In summary, the chapter is structured as follows: Section 6.2 established the relationship between the theory of random interval sets and fuzzy set variables. Section 6.3 looks at the credibility measure theory for its role in defining scalar fuzzy variable for modeling on fuzzy set systems. In §6.4 we integrate concepts developed in §6.2 and §6.3 and define a maximum entropy data-assimilated credibility distribution model of the derived scalar fuzzy variable, based on its counterpart the maximum entropy data-assimilated membership function. Based on the credibility distribution, we generate credibility grades for the fuzzy variable. Hence, we reduce the two dimensional random interval data process to a one-dimensional fuzzy data process. The derived vector of credibility grades defines spatial credibility random function on which ordinary kriging is performed. In §6.5 ordinary kriging on spatial credibility grades is proposed. An algorithm outlining the steps in credibility ordinary kriging is given in Appendix C.5.

6.2 Theory of Random Interval set and Fuzzy Variable set

In this section, we seek to give a clear relationship between the random interval set and fuzzy variable set. First, with notational changes to suite our application, we review a random interval set theory as proposed by Heilpern (1990).

Let $\mathbb{R}$ be the set of all real numbers. A random interval of the universal space of closed intervals, $\mathbb{IR}$ is a measurable mapping from the probability space $(\Omega, 2^{\Omega}, P)$ to the power set $2^{\mathbb{IR}}$ of $\mathbb{IR}$. Let Z be a random interval taking value in $\mathbb{IR}$, i.e.,

$$Z(\omega) = \left[Z^l(\omega), Z^u(\omega)\right], \omega \in \Omega, Z^l(\omega) \leq Z^u(\omega) \tag{6.1}$$

A typical treatment is to convert the closed interval Z into a point in the two-dimensional space (Chapter 2, Equation 2.43) recalled here as

$$\mathcal{T} = \left\{\left(Z^l, Z^u\right),\ Z^l, Z^u \in \Re^+, Z^l \leq Z^u\right\} \subset \mathbb{R}^2.$$

Each of the random intervals is treated as a (bivariate) random variable taking values on $\mathcal{T}$.

Let X and Y denote continuous random variables such that $X \leq Y$. Then, $f(x, y)$ denotes the bivariate density function such that:

$$\int_{-\infty}^{\infty}\int_{x}^{\infty} f(x, y)dydx = 1 \tag{6.2}$$

For a random interval $Z = \left(Z^l, Z^u\right)$, where Z^l and Z^u are continuous random variables such that $Z^l \leq Z^u$, the bivariate density $f(z^l, z^u)$ satisfies

$$\int_{-\infty}^{\infty}\int_{z^l}^{\infty} f(z^l, z^u)dz^u dz^l = 1 \tag{6.3}$$

Second, let us define a fuzzy variable set.

Definition 6.2.1 *Let W denote set of all possible elements (or the universe of discourse) and w be a generic element of W. A fuzzy variable set $\tilde{A}$ in W is defined as a set of ordered pairs*

$$\{w; \mu_{\tilde{A}}(w) | w \in W\},$$

where $\mu_{\tilde{A}}(w) : W \rightarrow [0, 1]$ is called the membership function for fuzzy variable set. The membership function maps each element of W to a membership grade between 0 and 1. The membership grades indicates the extent to which each w belongs to A.

Definition 6.2.2 *A fuzzy number is a kind of fuzzy variable set of the real number i.e.* $A \subseteq \mathbb{R}$. *In general a fuzzy number* $\tilde{A}$ *with membership function given as*

$$\mu_{\tilde{A}}(w) = \begin{cases} l(w) & \text{if } a \leq w < b \\ 1 & \text{if } b \leq w \leq c \\ r(w) & \text{if } c < w \leq d \\ 0 & \text{otherwise} \end{cases} \tag{6.4}$$

where $l(\cdot)$ *is monotone-increasing from zero to one and* $r(\cdot)$ *is monotone-decreasing function is called a "normal fuzzy number", and has a maximum value 1 on interval* $[b, c]$.

Now, let $\tilde{\varphi}(Z)$ denote the fuzzy variable generated by the random interval Z, then its membership function is expressed by:

$$\mu_{\tilde{\varphi}(Z)}(z) = \int_{-\infty}^{z} \int_{z}^{\infty} f(z^l, z^u) dz^u dz^l \ , \forall z \in \tilde{\varphi}(Z) \tag{6.5}$$

The fuzzy variable can assume different fuzzy numbers, and the degree of difficulty of predicting the specified value that the fuzzy variable will take is referred to as the fuzzy entropy (a measure of uncertainty) (Liu, 2004).

A normal fuzzy variable can be associated with a random interval based on the condition defined in Eq.6.2; the density function, $f(x, y)$ of the random interval Z associated with a normal fuzzy variable will takes positive values for points in the rectangle $D = [a, b] \times [c, d]$ only. The associated joint distribution function with the largest entropy is given by Heilpern (1990, pg. 216):

$$F_{X,Y}(x, y) = \begin{cases} l(x) & \text{for } y = b \\ 1 - r(y) & \text{for } x = c \\ l(x)(1 - r(y)) & \text{for } (x, y) \in \text{Int}D \\ 0 & \text{otherwise} \end{cases} \tag{6.6}$$

where Int$D$ denotes the interior of rectangular region D.

Therefore, the link between random interval set and fuzzy variable set is mathematically unavoidable and intrinsic. To utilize fuzzy variable modeling for random interval field, for its role in variable modeling on fuzzy set systems, we review Liu (2006) credibility measure theory in §6.3.

6.3 A Review of $(\vee, \cdot)$-Credibility Measure Theory

The $(\vee, \cdot)$-Credibility measure theory is built on the axiomatic credibility measure theory as well as probability theory (Guo and Li, 2006; Liu, 2006). We briefly review the Credibility measure theory here because of its role in variable modeling on fuzzy systems.

Definition 6.3.1 Let Θ denote a nonempty set, with corresponding power set $\mathcal{B}(\Theta) = 2^{\Theta}$. We refer to the elements $B \subset \Theta$, such that $B \in \mathcal{B}(\Theta)$ as events. In addition, let $\mathrm{Cr}(B)$ denote a number assigned to event B such that $0 \leq \mathrm{Cr}(B) \leq 1$. The number $\mathrm{Cr}(B)$ indicates the credibility that the event B occurs. $\mathrm{Cr}(B)$ satisfies the following axioms Liu (2006):

Axiom 1. $\mathrm{Cr}(\Theta) = 1$

Axiom 2. $\mathrm{Cr}(\cdot)$ is non-decreasing, i.e. $\mathrm{Cr}(B) \leq \mathrm{Cr}(C)$ for $B \subseteq C$.

Axiom 3. $\mathrm{Cr}(\cdot)$ is self-dual, i.e. $\mathrm{Cr}(B) + \mathrm{Cr}(B^c) = 1$ for $B \in \mathcal{B}(\Theta)$.

Axiom 4. $\mathrm{Cr}\{\cup_i B_i\} \wedge 0.5 = \sup[\mathrm{Cr}\{B_i\}]$ for any $\{B_i\}$ with $\mathrm{Cr}(B_i) \leq 0.5$

Axiom 5. Assume that a given set of functions $\mathrm{Cr}_k(\cdot) : 2^{\Theta_k} \to [0,1]$ satisfy Axioms 1-4, and $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_q$, then, for each $(\theta_1, \theta_2, \ldots, \theta_q) \in \Theta$,

$$\mathrm{Cr}(\theta_1, \theta_2, \ldots, \theta_q) = \begin{cases} \frac{1}{2}\Pi_{k=1}^{q}(\mathrm{Cr}(\theta_k) \wedge 1) & \text{if } \min_{1\leq k\leq q}\{\mathrm{Cr}(\theta_k)\} < 0.5 \\ \min_{1\leq k\leq q}\{\mathrm{Cr}(\theta_k)\} & \text{if } \min_{1\leq k\leq q}\{\mathrm{Cr}(\theta_k)\} \geq 0.5 \end{cases} \tag{6.7}$$

Any set function $\mathrm{Cr} : \mathcal{B}(\Theta) \to [0,1]$ satisfying Axioms 1-5 is called a $(\vee, \cdot)$-Credibility measure. Parallel to probability space in probability theory, the triplet, $(\Theta, \mathcal{B}(\Theta), \mathrm{Cr})$ is called the $(\vee, \cdot)$-credibility measure space.

Definition 6.3.2 A fuzzy variable, ξ is a mapping from the credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$ to a set of real numbers. The membership function associated with fuzzy variable ξ in the credibility space is given as

$$\mu(z) = (2\mathrm{Cr}\{\xi = z\}) \wedge 1, \quad \text{for } z \in \mathbb{R}. \tag{6.8}$$

Therefore, given a membership function we can derive the corresponding credibility measure and vise-versa. Similar to membership grades, credibility grades indicates the likelihood or favorability of a value belonging to a fuzzy variable set or a fuzzy

variable assuming a given value. Thus, members of a fuzzy set can be assigned either the membership grades or the credibility grades.

Definition 6.3.3 Let ξ be a fuzzy variable with membership function μ. Then for any set B of real numbers (Liu, 2004) :-

$$\mathrm{Cr}\{\xi \in B\} = \frac{1}{2}\left(\sup_{z\in B}\mu(z) + 1 - \sup_{z\in B^c}\mu(z)\right). \tag{6.9}$$

It follows that

$$\mathrm{Cr}\{\xi = z\} = \frac{1}{2}\left(\mu(z) + 1 - \sup_{y\neq z}\mu(y)\right), \quad \forall z, y \in \mathbb{R} \tag{6.10}$$

$$\mathrm{Cr}\{\xi \leq z\} = \frac{1}{2}\left(\sup_{y\leq z}\mu(y) + 1 - \sup_{y>z}\mu(y)\right), \ \forall z, y \in \mathbb{R} \tag{6.11}$$

$$\mathrm{Cr}\{\xi \geq z\} = \frac{1}{2}\left(\sup_{y\geq z}\mu(y) + 1 - \sup_{y<z}\mu(y)\right), \ \forall z, y \in \mathbb{R} \tag{6.12}$$

If μ is a continuous function,

$$\mathrm{Cr}\{\xi = z\} = \mu(z)/2 \tag{6.13}$$

Formally, credibility distribution is defined as :

Definition 6.3.4 The credibility distribution $\Phi : \mathbb{R} \to [0, 1]$ that a fuzzy variable ξ takes value less than or equal to a real-number $z \in \mathbb{R}$ is given as

$$\begin{aligned}\Phi_\xi(z) &= \mathrm{Cr}\{\theta : \xi(\theta) \leq z\} \\ &= \frac{1}{2}\left(\sup_{y\leq z}\mu(y) + 1 - \sup_{y>z}\mu(y)\right) \ \forall z, y \in \mathbb{R}\end{aligned} \tag{6.14}$$

$\Phi_\xi(z)$ is the credibility that the fuzzy variable ξ takes a value less than or equal to z.

Example 6.3.1 Let ξ follow normal distribution of mean ϖ and variance σ^2 with ϖ being a fuzzy variable with a membership function:

$$\mu_\varpi(z) = \begin{cases} 1 - e^{-\beta z} & z \geq 0, \beta \in \mathbb{R} \\ 0 & \text{otherwise} \end{cases} \tag{6.15}$$

parameterized by β. Then ξ is a fuzzy variable. This example hints an important fact about the relation between random fuzzy variable (event) and fuzzy variable,

stated as a theorem 6.3.1, (Liu, 2004):

Theorem 6.3.1 Let ξ be a random fuzzy variable. If the expectation $E_P[\xi(\theta_0)]$ exists for any given $\theta_0 \in \Theta$, then $E_P[\xi(\cdot)]$ is a fuzzy variable. It is obvious that in Example 6.3.1, $E_P[\xi] = \varpi$ is the fuzzy variable with membership function defined by Equation 6.15. This theorem pave a way toward a data-assimilated fuzzy membership obtained using a set of observations from a random fuzzy variable, denoted as $\{x_1, x_2, \ldots, x_n\}$.

Definition 6.3.5 The expected value of a fuzzy random variable is (Liu and Liu, 2003):

$$E[\xi] = \int_0^{+\infty} \mathrm{Cr}\{\theta \in \Theta | E[\xi(\theta)] \geq r\} dr - \int_{-\infty}^{0} \mathrm{Cr}\{\theta \in \Theta | E[\xi(\theta)] \leq r\} dr \quad (6.16)$$

(6.17)

provided at least one of the two integrals are finite. The variance of a fuzzy variable ξ is defined as $\mathrm{Var}[\xi] = E[\xi - E[\xi]]^2$. We should emphasize again that the expectation and variance of a fuzzy variable are real numbers. On the contrary, the expectation of random interval variables is also a random interval. Therefore, application of Credibility distribution theory reduces random interval sets to scalar fuzzy variable, and hence the scalar variable modeling of random interval events.

Finally, since the probability of a random interval is a set mapping from a power set to unit interval: $P_W : 2^U \rightarrow [0,1]$, defined on space $(U, 2^U, P_W)$, while the credibility measure for fuzzy subsets is set mapping from a power set to unit interval: $\mathrm{Cr} : 2^U \rightarrow [0,1]$, defined on credibility space $(U, 2^U, \mathrm{Cr})$, the link between probability measure and credibility measure exists and a measure called probability-credibility consistence measure is defined for the degree of link.

In the next section, §6.4, we theoretically derive the membership function and the credibility distribution for the induced fuzzy variable.

6.4 Credibility Distribution for Induced Fuzzy Variable via Kernel Estimation

Based on the arguments in §6.2 and §6.3, as long as a set of spatial inequality constraints or spatial random interval observations of spatial process, $\left\{\left(z_i^l, z_i^u\right), i = 1, 2, \ldots, n\right\}$, can be collected, we can estimate the joint density of spatial interval variable $\left(Z^l, Z^u\right)$. This joint density can be denoted by $p_k(z^l, z^u)$. A typical nonparametric

approach is **kernel density estimation under maximum entropy principle**. The concept of fuzzy entropy is well-defined in the fuzzy mathematical literature (**De Luca and Termini, 1972; Rudas and Kaynak, 1998**).

Definition 6.4.1 Let ξ be a continuous fuzzy variable defined on credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$ then the fuzzy entropy, $H[\xi]$ is defined as:

$$H[\xi] = \int_{-\infty}^{\infty} S\left(\mathrm{Cr}\left(\{\theta : \xi(\theta) = u\}\right)\right) du \tag{6.18}$$

where: $S(t) = -tlnt - (1-t)ln(1-t)$, (De Luca and Termini, 1972).

Let $x_1, \ldots, x_n$ denote a sample of possible values assumed by a fuzzy variable. The sample fuzzy entropy, $J[\xi]$ is given by:

$$J[\xi] = \frac{1}{n} \sum_{i=1}^{n} S\left(\mathrm{Cr}\left(\{\theta : \xi(\theta) = x_i\}\right)\right) \tag{6.19}$$

The kernel related parameters are thus chosen by maximizing the sample entropy $J[\xi]$, and then the data-assimilated membership function of the generated fuzzy set $\tilde{\varphi}\left(Z^l, Z^u\right)$ is established:

$$\mu_{\tilde{\varphi}\left([Z^l, Z^u]\right)}(z) = \int_{-\infty}^{z} \int_{z}^{\infty} p_k(z^l, z^u) dz^u dz^l \tag{6.20}$$

The counterpart membership function, the data-assimilated credibility distribution for the induced fuzzy variable is given as

$$\Phi_{\xi}(z) = \frac{1}{2}\left(\mu_{\tilde{\varphi}\left([Z^l, Z^u]\right)}(z) + 1 - \sup_{y \neq z}\left[\mu_{\tilde{\varphi}\left([Z^l, Z^u]\right)}(y)\right]\right) \tag{6.21}$$

Remark 6.4.1 *The data-assimilated credibility distribution may be viewed as the marginal increase of a person's strength of belief that 'z is $\tilde{\varphi}\left([Z^l, Z^u]\right)$' and is assumed to be proportional to the strength of his/her belief that 'z is $\tilde{\varphi}\left([Z^l, Z^u]\right)$' and to the strength of his/her belief that 'z is not in $\tilde{\varphi}\left([Z^l, Z^u]\right)$'. This relationship between the membership function and the credibility distribution justifies the use of either.*

As to the probability measure of event $P\left[\xi(\theta) \in B\right]$, we have:

$$P\left[\xi(\theta) \in B\right] = \int_B \int p_k(z^l, z^u) dz^u dz^l \tag{6.22}$$

6.5 Ordinary Credibility Grade kriging

6.5.1 Credibility Distributional Grades

Note that for any given location $\boldsymbol{x}_i = (x_i, y_i) \in D$, the fuzzy variable ξ will take a value $z_i = z(\boldsymbol{x}_i)$ at this location, and thus a value of credibility distribution (Eq. 6.21), $\Phi_\xi(z_i)$ is assigned to location $\boldsymbol{x}_i$.

Definition 6.5.1 *For a given fuzzy variable ξ with credibility distribution Φ_ξ, if $\xi = z_i$ at location $\boldsymbol{x}_i = (x_i, y_i)$, then $\Phi_\xi(z_i)$ is called the credibility grade for fuzzy variable ξ at location (x_i, y_i). The collection of spatially distributed credibility grades, denoted as*
$\{\Phi_\xi(z_i),\ \boldsymbol{x}_i \in D \subset \mathbb{R}^2,\ i = 1, 2, \cdots, n\}$, *is called sampled credibility grades over region D. The credibility grades range from 0 to 1 and forms an alternative generalization to 0/1 indicator codes as used in indicator kriging.*

Since the credibility distribution of a fuzzy variable ξ contains the full information on the fuzzy variable, it is logical to say that the credibility grades reveal the information of fuzzy variable ξ spatial distribution, i.e., $\{\xi(\boldsymbol{x}), \boldsymbol{x}_i \in D\}$. Similar to the general treatment in spatial data analysis, the available information on $\xi(\boldsymbol{x})$ is the sampled credibility grades over region D, so that we can perform the kriging on the sampled credibility grades over region D and therefore obtain the credibility grade at any unsampled location, say, $(\boldsymbol{x}_0) \in D$, denoted as $\Phi_\xi(z_0)$.

6.5.2 Credibility Grade Kriging

The basic stationarity assumptions for the credibility grade kriging are similar to those of ordinary kriging cases discussed in chapter 2. Therefore, we extend the definition of ordinary semivariogram to credibility grades and define the credibility grade semivariogram $\gamma_\Phi(h)$ as

$$\gamma_\Phi(h) = \frac{1}{2}\mathbb{E}\Big[\left(\Phi(z(\boldsymbol{x}+h)) - \Phi(z(\boldsymbol{x}))\right)^2\Big] \tag{6.23}$$

where h is the distance separation vector between spatial locations $\boldsymbol{x}$ and $\boldsymbol{x}+h$ with $\Phi(z(x))$ and $\Phi(z(x+h))$ the corresponding credibility grades.

The sample credibility grade semivariogram (6.24) used for credibility grade kriging is given as

$$\hat{\gamma}_\Phi(h) = \frac{1}{2n(h)}\sum_{i=1}^{n(h)}\left[\left(\Phi(z(\boldsymbol{x}_i+h)) - \Phi(z(\boldsymbol{x}_i))\right)^2\right] \tag{6.24}$$

where $n(h)$ is the total samples within a distance h.

Applying the experimental semivariogram (6.24), we derive the credibility grade kriging system 6.25 as an ordinary kriging on credibility grades defined as

$$\begin{aligned} \sum_{j=1}^{n(h)} \lambda_j \gamma_\Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) + \psi &= \gamma_\Phi(\boldsymbol{x}_0 - \boldsymbol{x}_i) \;\; i = 1, \ldots, n(h) \\ \sum_{j}^{n(h)} \lambda_j &= 1 \end{aligned} \tag{6.25}$$

where ψ is the Lagrange multiplier. Solution to the system 6.25 results to weights λ_i for the credibility grade predictor defined as

$$\hat{\Phi}(z(\boldsymbol{x}_0)) = \sum_{i=1}^{n(h)} \lambda_i \Phi(z(\boldsymbol{x}_i)), \quad \sum_{i}^{n(h)} \lambda_i = 1 \tag{6.26}$$

with $n(h)$ the $\boldsymbol{x}_0$ neighborhood sample.

6.6 Discussion

This chapter has theoretically looked at scalar fuzzy variable approach of modeling spatial random intervals. A scalar fuzzy variable equivalent of random interval data is defined with an associated distribution function, the credibility distribution function and a self-duality measure, the credibility measure. Contrary to fuzzy sets variable methods with fuzzy set estimates/ predictions the proposed approach, leads to scalar valued estimates.

Applying a bivariate kernel-based empirical approach, the random interval is first converted to a fuzzy set with maximum entropy. Then the fuzzy set is represented as a scalar fuzzy variable under credibility measure theory. This conversion simultaneously models both random and fuzzy uncertainty inherent is random interval data. Associated with the scalar fuzzy variable is a maximum entropy data-assimilated membership function and its counterpart the data-assimilated credibility function, non-decreasing function from 0 to 1. Based on the credibility distribution we generate fuzzy variable credibility grades. The derived vector of credibility grades defines credibility random function on which ordinary kriging is performed.

The methodology offers several advantages. First, by defining non-parametric kernel based membership function (Eq. 6.20) based on the data, we avoid the problem of having to assume or define a prior functional form of the membership function. This is a move from the often subjective expert opinion approach of defining

membership functions.

Second the approach is robust in the presence of small and ill-conditioned samples; and does not require high dimensional integration of marginal probabilities. Therefore, it can be easily implemented in both small as well as very large samples.

Third, for modeling spatial random fields with imprecise data, we are able to move from the basic random interval analysis approach to variable-based modeling approach. This is of great convenience and easily integrates with most GIS mapping software.

The approach also offers flexible output in form of real-valued estimates based on joint distributions. This makes it a potential tool for spatial analysis of non-scalar spatial data given as intervals.

Further, we note that credibility grades are unique descriptive marks of the spatial random interval/vector objects in the study region. Such model can be refereed to as a CREDIBILITY GRADE MARKED POINT PROCESS and can be studied within the framework of marked point process. Here, the "points" are the random vector spatial locations and the "marks" are credibility grade.

CHAPTER 7

APPLICATION OF PROPOSED METHODS TO POLLUTION DATA

Scope

We devote this chapter to the application of proposed methodologies as outlined in this thesis (Chapter 3 - 6) to the Spatial interpolation competition SIC2004 dataset (Dubois, 2005) on Gamma dose pollution across Germany.

7.1 Introduction

In this chapter, we apply all the proposed methods developed in chapters 2 -6 to real-life data, the Spatial interpolation competition SIC2004 dataset (Dubois, 2005) on Ambient gamma dose pollution across Germany. The goal of the project is to assess the magnitude and geographical extent of a radioactive contamination in the air and on the ground. This implies converting reported local observations of a gamma dose rates into information with spatial continuity, i.e. prediction maps which are crucial for decision-making and/or further modeling. The prediction maps provide early warnings about any increase of gamma dose levels above certain thresholds, to necessitate any national emergency plan where necessary.

Generating optimal prediction maps involves application of well trained algorithms. The task of training the algorithms is not trivia (Dubois and Shibli, 2003). First of all, no uniformly best prediction mapping technique exists because each associated model has its advantages and downsides. Second, nearly all prediction techniques require number of parameters to be set or estimated, as shown in the implementation of techniques below. In summary, the chapter is structured as follows:

Section	Related Chapter	Objective addressed	Results
7.2 Data description			
7.2.1 Training set			
7.3 Criteria for comparing methods			
7.4 Classical Random Interval Geostatistics	2		
7.4.1 Kriging of Spatial Random Intervals	3	(i)	Fig. 7.9
7.4.2 Spatial Kernel Interval Regression	4	(ii)	Fig. 7.13
7.5 Bi-variable Copula-based Geostatistics	5	(iii, iv)	Fig. 7.17
7.5.1 Preliminary Analysis			
7.5.2 Structural Analysis			
7.5.3 Prediction			
7.6 Fuzzy Credibility-based Geostatistics	6	(iii, v)	Fig. 7.21
7.6.1 Preliminary Analysis			
7.6.2 Structural Analysis			
7.6.3 Prediction			
7.7 Comparison of Bi-variable Copula-based and Fuzzy Credibility-based Geostatistics	5, 6	(iv, v)	Fig. 7.23
7.8 Overall assessments of all techniques			

7.2 Data description

The SIC2004 dataset measurements of ambient gamma dose rates were extracted from the European Radiological Data Exchange Platform database (EURDEP, see http://eurdeppub.jrc.it/, De Cort and de Vries (1997) developed by the Radioactivity Environmental Monitoring (REM) group to make European radiological monitoring data available to decision-makers. From this database, 10 sets of mean daily values were selected from year 2003, by drawing roughly one day at random from each month for 10 months. Then, using the 10 sets at each of location a random interval is derived as described in §7.2.1.1. A further filtering of these data was applied to select only measurements reported by the German national automatic monitoring network (IMIS) of the Federal Office for Radiation Protection (BfS, http://www.bfs.de/). This selection ensured that the data were homogeneous in terms of measurement technique and included the densest monitoring network in Europe, i.e. the German one. More details regarding the datasets can be found in Dubois and Galmarini (2005). From around 2000 monitoring stations in Germany, 1008 stations (relative locations Figure 7.1, right) were selected by drawing a rectangular window. These stations were common to each of the 10 sets and all reported values for each day selected.

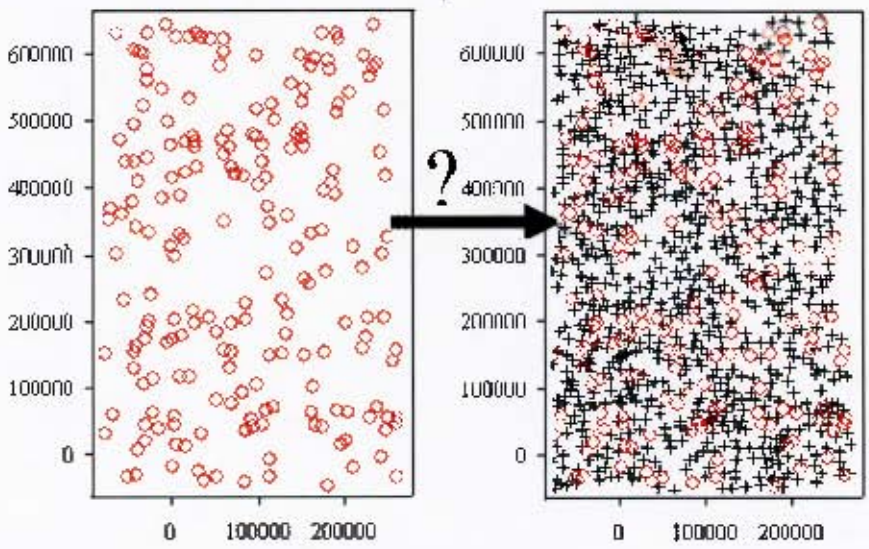


Figure 7.1. A map of relative locations for the 200 sampling locations for the SIC2004 training dataset (left). Right panel is relative locations of the 808 prediction locations (crosses) and the 200 datasets locations (circles). Horizontal axis is x(the easting in meters). Vertical axis is y (the northing in meters)

7.2.1 Training sets

From the selected 1008 monitoring stations, a single sampling scheme of 200 monitoring locations (Figure 7.1, left) was selected randomly and extracted. Each of the 10 sets at the $n = 200$ sampling locations consists of:

ID -an integer value that uniquely identifies each monitoring station
x -the easting of the monitoring station in meters
y -the northing of the monitoring station in meters
Z -the ambient gamma dose rate measured in nanoSievert per hour (nSv/h).

Each of the $N - n = 808$ prediction locations consist of:

ID -an integer value that uniquely identifies each monitoring station
x -the easting of the monitoring station in meters
y -the northing of the monitoring station in meters

At the end of the Spatial interpolation competition SIC2004 competition, the observed values at the 808 locations were given. These observed values will be used to calculate the prediction error for each of the proposed techniques.

Table 7.1. Summary statistics for the 10 training datasets for the 200 sampling locations in Germany.

Dataset No.	Min	Mean	Median	Max	Std	Skewness	kurtosis
1	55.0	97.6	98.0	150.0	19.1	0.03	-0.52
2	55.9	97.4	97.9	155.0	19.3	0.08	-0.45
3	59.9	98.8	100	157.0	18.5	0.12	-0.27
4	56.1	93.8	94.8	152.0	16.8	0.20	0.00
5	56.4	92.4	92.0	143.0	16.6	0.21	-0.16
6	54.4	89.9	90.4	133.0	15.9	0.11	-0.47
7	56.1	91.7	91.7	140.0	16.2	0.12	-0.37
8	54.9	92.4	92.5	148.0	16.6	0.15	-0.05
9	56.5	96.6	97.0	149.0	18.2	0.05	-0.38
10	54.9	95.4	95.7	152.0	17.2	0.12	-0.18

Comparing the 10 datasets (Table 7.1, Figure 7.2) at the sample 200 stations, the variable (gamma dose rates) does not vary much in time. Based on routine monitoring, the variable does not show strong spatio-temporal fluctuations for time intervals measured, as is usually the case, for example, with ozone levels, natural background radioactivity and rainfall fields. In addition, normal gamma dose rates vary with local geology and typically range from 50 to 120 nSv/h (CTIF et al., 2006;

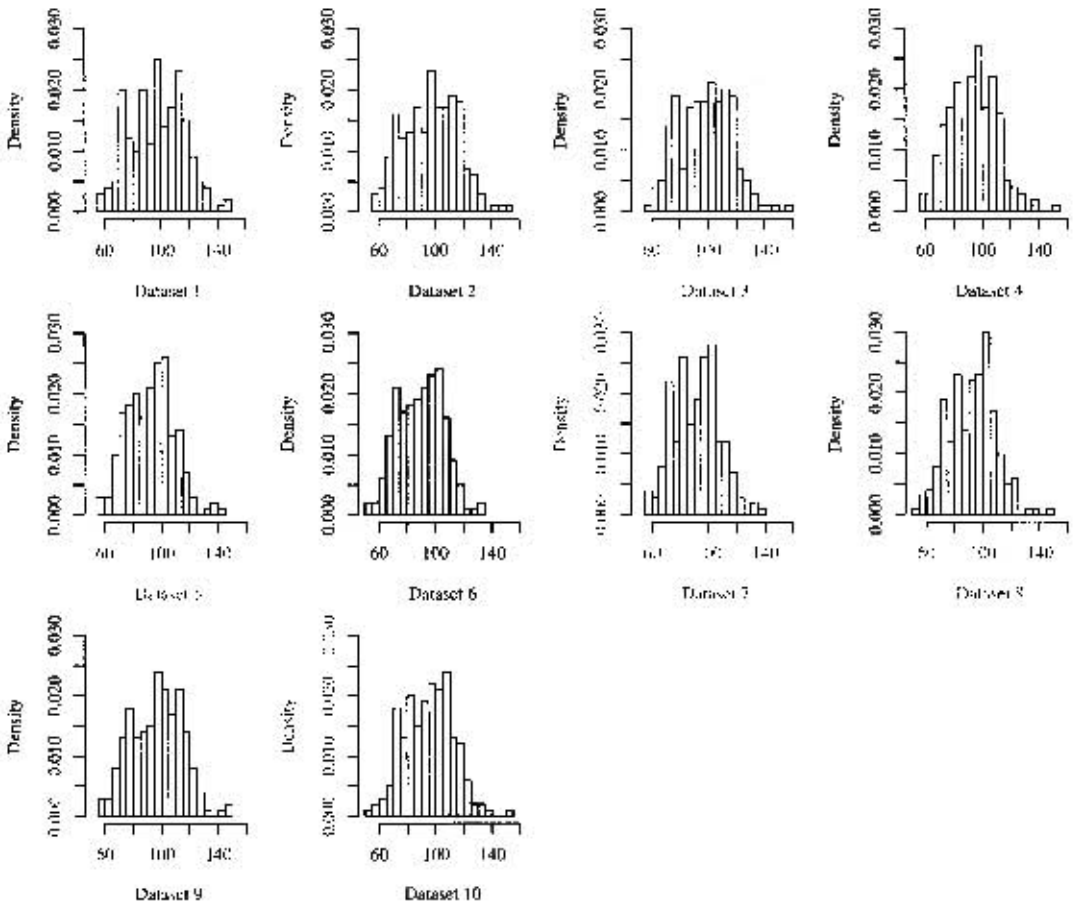


Figure 7.2. Histograms for the 10 datasets from the 200 sample locations. All the datasets are mostly symmetric. No obvious outliers are present.

Somerville and McMahon, 2006). Increases in the dose rate occasionally occur due to washout of radon decay products from the atmosphere or due to a dangerous source that is leaking or potentially involved in a terrorist act or explosion. More details regarding the datasets can be found in Dubois and Galmarini (2005).

7.2.1.1 Defining regionalized random intervals

Since the data do not vary significantly, one may choose to use only one of the 10 datasets to train the algorithms. However, to ensure that all imprecision/fuzziness in the data is incorporated in the analysis, it would be better to train and test the algorithms using the whole interval data range (10 dataset values) at each location. This way, one incorporates all the possible values at a given location including any sudden shift in the observed phenomena. Generated random intervals at the 200 sample spatial locations are called regionalized random intervals (Figure 7.3b).

In Figure 7.3, we summarizes the analysis steps beginning with the definition of the regionalized random intervals, training and testing the algorithms followed by prediction and finally the calculation of prediction error using observed and predicted values at the 808 locations

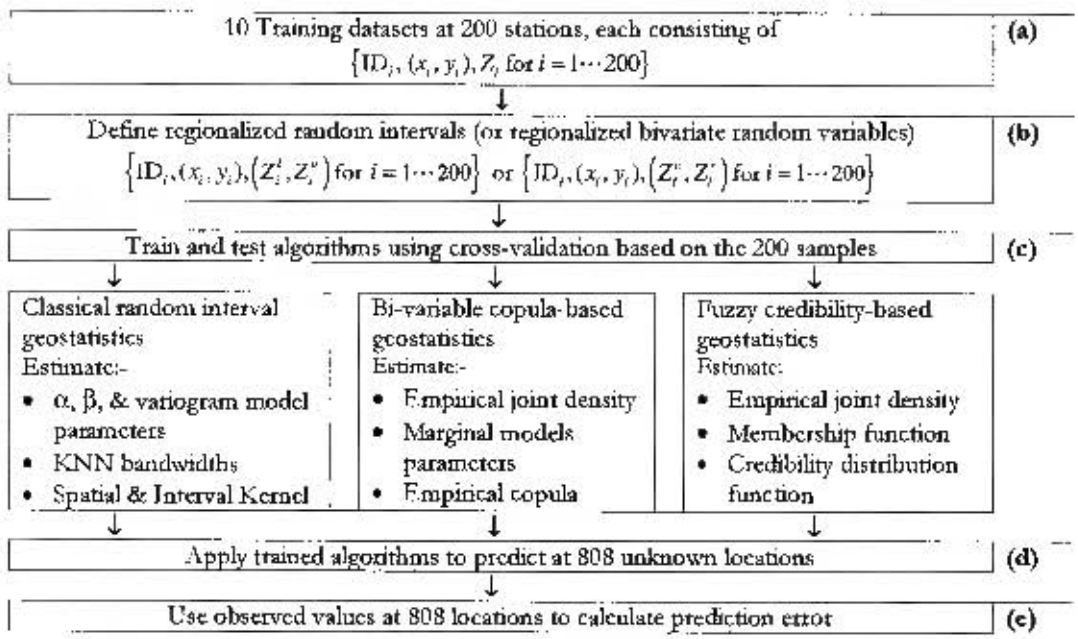


Figure 7.3. A chart showing the general steps from the (a) 10 original training datasets to (b) defining the regionalized random intervals, (c) training and testing algorithms using cross-validation based on the 200 samples (d) applying the trained algorithms to predict at 808 locations, and (e) calculating prediction error using the observed values at 808 prediction locations.

To define the regionalized random intervals (Figure 7.3b) we proceed as follows. For each of the 200 sample locations, we define a random interval, $\boldsymbol{Z}_i = [Z_i^l, Z_i^u], i = 1, \ldots, 200$ such that

$$Z_i^l = \min\left([\text{datavalue1 - datavalue 10}] \text{ at spatial location}(i)\right)$$
$$Z_i^u = \max\left([\text{datavalue1 - datavalue 10}] \text{ at spatial location}(i)\right)$$

The lower bound variable, Z_i^l is the lowest recorded value at location (i) across the selected 10 data values while the upper bound variable, Z_i^u is the largest recorded value at location (i) across the selected 10 data values. Since we are dealing with spatial data, we then have spatial random intervals or regionalized random intervals (Figure 7.3b). The regionalized random intervals represent the fuzzy measures of the observed phenomena. This scenario mimics naturally occurring real world data

which is often not exact, but exists in form of interval values bounded by minimum and maximum values.

Respectively, $Z^c = (Z^u + Z^l)/2$ and $Z^r = (Z^u - Z^l)/2$ denotes the interval center of mass and interval uncertainty (radius). As shown in chapter 2, the center is the optimal single-value approximation of the random interval. In addition, applying the theory of closed random interval (§2.4.2), the random interval $\boldsymbol{Z} = [Z^l, Z^u]$ is viewed as a bivariate random vector, (Smets, 2005), of two coordinates random variables i.e.,

$$\boldsymbol{Z} = \begin{cases} (Z^l, Z^u) \text{ with } Z^l \leq Z^u & \text{lower(l) and upper (u) variables} \\ (Z^c, Z^r) \ni Z^c \in \boldsymbol{Z} \text{ and } Z^r \geq 0 & \text{center (c) and radius (r) variables} \end{cases}$$

in l-u and c-r coordinate systems. Thus spatial random intervals are basically spatial bivariate random vectors or regionalized bivariate random vectors. Applying equation Eq. 2.45, we can move from l-u to c-r coordinates system and vice-versa.

In Table 7.2 we provide the summary statistics for the spatial bivariate random vector variables:- the lower, upper, center and radius.

Table 7.2. Summary statistics for the spatial bivariate random vector variables: lower and upper values as well as the center(midpoint) and radius values. The spatial random intervals will be used to train and test the algorithms.

Variable	Min	Mean	Median	Max	Std	Skewness	kurtosis
lower(l)	54.4	90.2	90.4	143.0	15.9	0.11	-0.49
upper(u)	59.9	99.9	101.0	157.0	18.8	0.08	-0.40
center(c)	57.2	95.1	95.2	150.0	17.3	0.09	-0.45
radius(r)	1.3	4.9	4.7	18.3	2.2	2.55	9.48

As expected the upper values consist of larger dose values relative to lower sample values. Analysis based on the two datasets (lower and Upper) may be slightly different in terms of magnitude, but overall, we expect insignificant difference in the predicted spatial patterns.

Figure 7.4, show the size distribution of lower (minimum) and upper (maximum) values of gamma dose rates at the 200 sample monitoring stations.

The center regionalized random variables are the optimal single-value approximations of the regionalized random intervals. Figure 7.5, left, shows the size distribution of center values of the data set at 200 sample locations, and the right shows the observed values at 808 prediction locations.

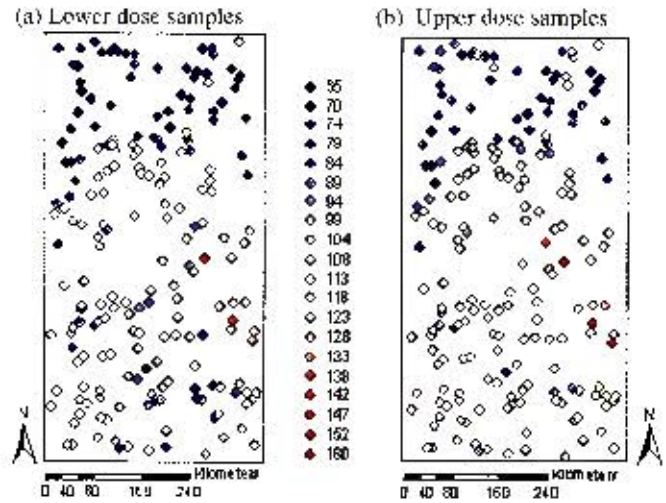


Figure 7.4. A map of the size distribution of gamma dose rate samples at the 200 monitoring stations in Germany; a) is the lower dose sample values and b) the upper dose sample values. In both, small values dominate in the north compared to a mixture of medium to large values in the south east. Note the unsampled area around the middle and slightly left of the center of the map.

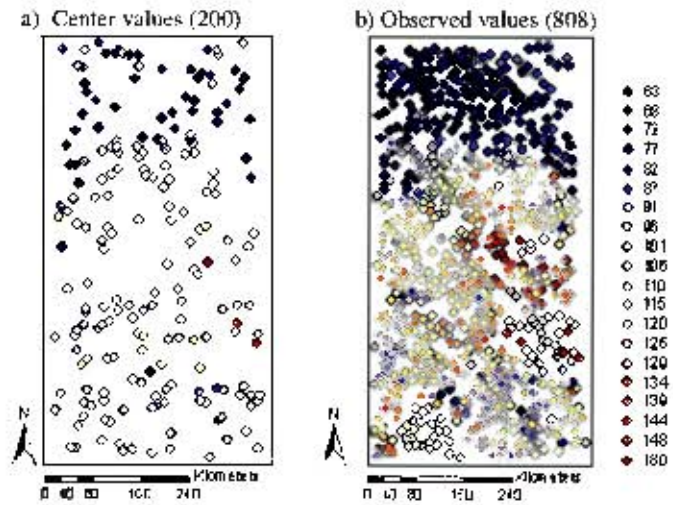


Figure 7.5. A map of the size distribution of gamma dose rates for a) center values at the 200 monitoring stations and b) observed values at the 808 prediction locations in Germany. In both cases, small values occur in the north. Similar to Figure 7.4, panel (a) shows a large central region plus pockets of regions in north and south with no data. Mixture of large to medium values occur in the south-east direction. Panel b is the expected patterns from predictions based on the 200 samples.

7.3 Criteria to compare the approaches

To compare the different approaches we use sets of criteria grouped according to scope:

Global Criteria

Global criteria are used in order to evaluate the methods ability for yielding predictions as close as possible to the observations. The criteria are based on the error distribution, which allows discriminating the methods on the basis of their capacity to predict Gamma dose rates. Let $\boldsymbol{Z}_i$ and $\hat{\boldsymbol{Z}}_i$ respectively denote the observed and the predicted interval at location (i), we use the mean error (ME); the mean absolute error (MAE); the root mean square error (RMSE) and the Pearsons correlation coefficient (r) summary statistic respectively computed as

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{Z}_i - \hat{\boldsymbol{Z}}_i\right) \tag{7.1}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\boldsymbol{Z}_i - \hat{\boldsymbol{Z}}_i\right| \tag{7.2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{Z}_i - \hat{\boldsymbol{Z}}_i\right)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_G^2\left(\boldsymbol{Z}_i, \hat{\boldsymbol{Z}}_i\right)} \tag{7.3}$$

$$\text{r} = \frac{\text{Cov}\left(\boldsymbol{Z}, \hat{\boldsymbol{Z}}\right)}{\sqrt{\text{Var}\left(\boldsymbol{Z}\right)\text{Var}\left(\hat{\boldsymbol{Z}}\right)}} \tag{7.4}$$

The mean absolute error (MAE) is computed as the mean of the absolute magnitude difference between individual interval coordinates, while the RMSE is the average squared distance (e.g. weighted interval metric, d_G) between the observed and predicted intervals.

The ME is an indicator of bias, and should be close to zero. The RMSE and MAE are measures of accuracy of the prediction, should be approximately the same and are expected to approach zero. The closest they are to zero, the closer the predicted values are to the actual values. If errors are normally distributed, the RMSE is thought to be the best measure. MAE, similar to RMSE, incorporates the bias as well as the variance of the error distribution, and is more robust with respect to extreme values than RMSE (Journel, 1994).

The correlation coefficient measures the relationship between the observed and the predicted values. High correlation indicates better prediction. However, care should be taken when interpreting the correlation coefficient since it often produces an inflated index. To complement for this problem, RMSE or the MAE is used.

Local Criteria

Local Criteria is based on the patterns shown on the prediction maps, together with associated prediction error map. It helps to visualize how the methods perform at a local scale. Predicted maps give an indication of the methods ability to reproduce the expected pattern including the local features. The prediction error maps indicate the associated prediction errors. Usually, high prediction errors dominate in regions with no sample values.

7.4 Classical Random Interval Geostatistics (CRIG) methodologies

7.4.1 Kriging of Spatial Random Intervals

7.4.1.1 Explanatory Spatial Interval Data Analysis

For explanatory analysis of spatial interval data we investigate for directional trends (Figure 7.6) for the lower, upper, center and radius regionalized variables.

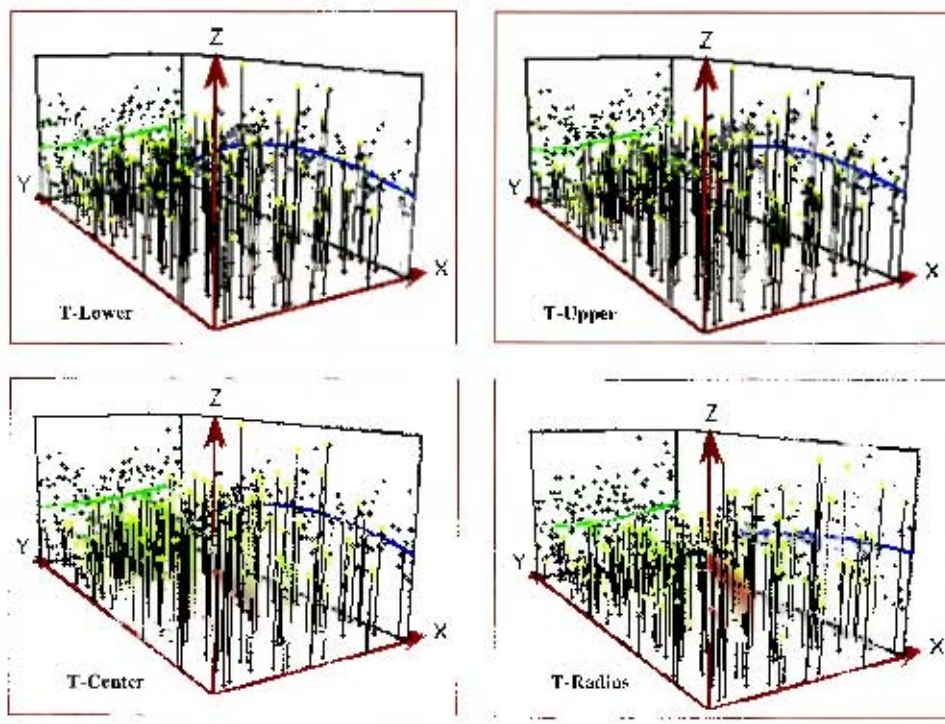


Figure 7.6. Spatial trend analysis for random vector variables: T-Lower, T-Upper, T-Center and T-Radius, representing the minimum, maximum, center and radius directional trends. There is a no significant general directional trend.

There is no significant trend in the data. The insignificant north-south global trend in the radius merely indicates the increase in uncertainty associated with

large values relative to the rest of the data. Although not always the case, there is likelihood that large observed values have also high level of uncertainty/imprecision relative to the rest of the data.

7.4.1.2 Structural Analysis (Variography)

Structural analysis is often an iterative process aimed at identifying the optimal model of spatial structure. For structural analysis we will use the center-radius representation of the regionalized random vectors. This way we are able to 'partially' reduce interval modeling to variable modeling using the center data and a measure of uncertainty given by the radius data. The center and radius data directional variograms indicated little evidence of anisotropy, so analysis was based on omni-directional variograms. In Figure 7.7 we shows the center and radius isotropic spherical models based on lag size of 20,000 meters. Although an arbitrary number, 20,000m produced variogram that clearly shows increasing semi-variance with increasing separation distance.

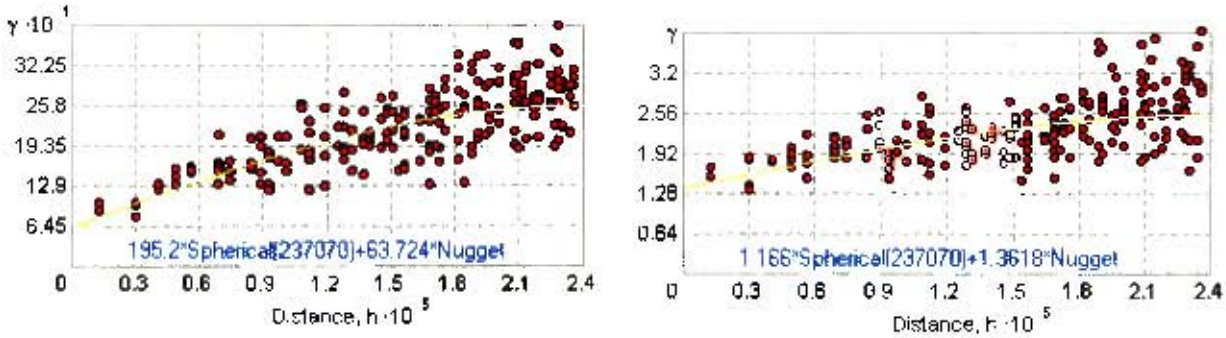


Figure 7.7. Fitted isotropic spherical model for Center values (left panel) with nugget=63.724; sill=195.2; range=237070 meters. Right panel shows the radius spherical model with nugget=1.3618; sill=1.166; range=237070 meters. The models slowly increases from the origin, an indicate of a smooth process.

Radius data has a relatively higher nugget effect of 1.3618, compared to its sill value (1.166). This is an indication of quite low correlation structure of the radius data, the measure of interval uncertainty. So much of variance contributed by radius data is either due to measurement error variance or a micro-scale structure variance. Still, the radius does contribute to spatial structure as indicated by the small sill value. Simply, the radius model can be regarded as nugget effect model.

Kriging on intervals utilizes the interval variogram (Eq. 3.50) which equals half the expectation of squared distance between regionalized random intervals. The

applied distance corresponds to the weighted interval metric (Eq. 3.17):

$$d_G^2(\boldsymbol{a},\boldsymbol{b}) = (a^c - b^c)^2 + \frac{\alpha}{\alpha+\beta}(a^r - b^r)^2$$

However, to apply the interval metric, we need to estimate the α and β parameters which are associated with a beta probability density function as given in Eq. 3.18. In §7.4.1.3 we show how the parameters were estimated.

7.4.1.3 Estimation of α and β based on Training Random Interval Data

Since spatial random intervals can be approximated with a single value (the center), the distance between random intervals is proportional to the distance between their center values (see Eq. 3.7). Thus utilizing the absolute distance between the center values, we can optimize on the α and β parameters. We use the absolute distance measure to avoid the influence of possible extreme values. The optimal α and β parameters, were given by the average of 10 sample parameters estimates of beta and alpha in Table 7.3, Figure 7.8.

Table 7.3. The estimated beta and alpha distribution parameters together with the distribution goodness of fit statistics (Kolmogorov-smirnov statistic) at critical values 0.0272 (5%) and 0.0362(1%) significant levels, for each of the 10 samples of distances between center values. Note (*) and (**) denote 5% and 1% significance levels. Min and max are the minimum and maximum values fitted distributions for each sample.

Sample	Alpha	Beta	Min	Max	Kolmogorov-smirnov statistic	
1	1.1908	4.4646	-0.0136	93.534	0.0300	*
2	1.1909	4.5152	-0.0009	93.524	0.0289	*
3	1.1560	4.0304	-0.0200	88.635	0.0171	**
4	1.1677	4.1939	-0.0020	90.575	0.0142	**
5	1.2156	4.3599	-0.0169	90.956	0.0208	**
6	1.1871	4.9207	-0.0068	100.200	0.0285	*
7	1.1874	4.0606	-0.0007	90.309	0.0181	**
8	1.1194	4.3823	-0.0017	96.118	0.0291	*
9	1.1488	3.7258	-0.0019	83.373	0.0328	
10	1.2083	3.9652	-0.0156	85.574	0.0232	*

The 10 sample beta and alpha parameters estimates (Table 7.3) estimated as follows:-

i) calculate the absolute distance for all the pairs of regionalized center variables.

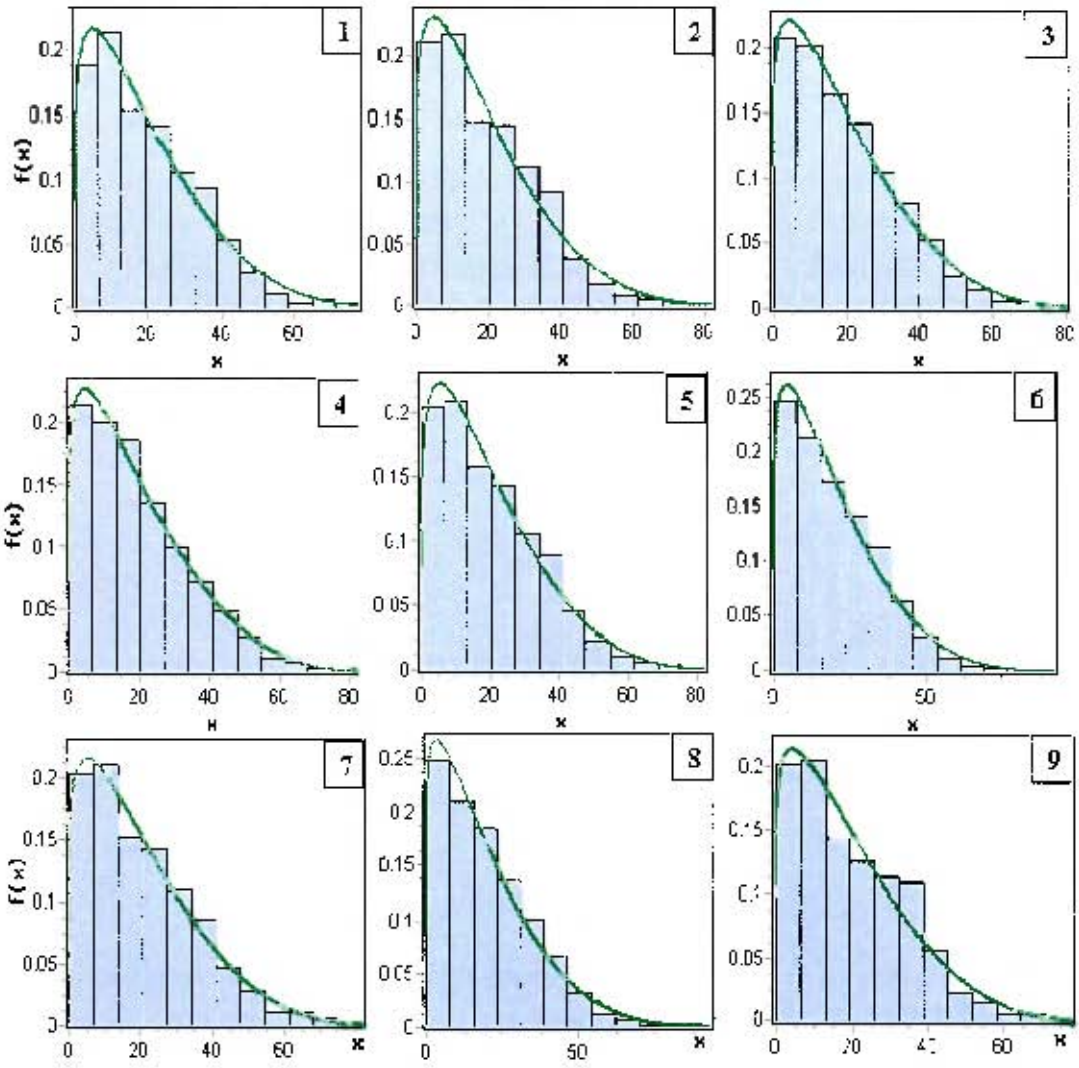


Figure 7.8. The fitted beta probability density estimates based on the samples of absolute distances between the interval center values. The distributions are not significantly different. This is also indicated by the parameters (Table 7.3), which are very similar.

ii) from the collection of absolute distances, we select a random sample, with replacement, of size $n = 2500$. In total we obtain $k = 10$ random samples

iii) for each of the $k = 10$ samples, we fit the Beta probability density

$$\hat{f}(\lambda) = \frac{1}{\mathrm{B}(\alpha, \beta)} \frac{(\lambda - a)^{\alpha-1}(b - \lambda)^{\beta-1}}{(b - a)^{\alpha+\beta-1}}$$

with a(min) and b(max) respectively the minimum and the maximum distance values in each of the k samples

iv) α and β estimates parameters are given in Table 7.3 together with summary

statistics for each of the 10 samples. The min and max (Table 7.3) values represent the minimum and maximum values of fitted distributions for each of the 10 samples. Figure 7.8 shows the fitted Beta probability densities.

v) to obtain optimal beta and alpha parameter estimates, we average the 10 sample parameters estimates of beta and alpha which leads to $\alpha = 1.1772$ and $\beta = 4.2619$, and the interval distance becomes

$$d_G^2(\boldsymbol{a}, \boldsymbol{b}) = (a^c - b^c)^2 + 0.2164(a^r - b^r)^2 \tag{7.5}$$

The distance (7.5) is used in deriving the composite variance (Eq. 3.56) and the component-wise variance (Eq. 3.58) which are used in defining the corresponding kriging systems referenced in §7.4.1.4 below.

7.4.1.4 Kriging: Composite and Component-wise

Kriging weights are obtained by solving:-

i) composite kriging system, (Eq. 3.69)

ii) component-wise kriging systems (Eq's 3.77 and 3.78)

Estimates at unknown locations are obtained using,

i) composite kriging by applying Eq. 3.86, and for the variance applying Eq. 3.87

ii) component-wise kriging by applying Eq. 3.88, and for the variance applying Eq. 3.89

The results for the global criteria (cross-validation) are given in Table 7.4. We also compare our results with the SIC2004 most accurate results 'best SIC2004 results' provided in Dubois, 2005, p.13. Note that the algorithms listed in SIC2004 competition were based on scalar-valued measures (Dubois, 2005).

It is clear that between the two new interval kriging methods, we do not see much difference in the bias (ME) or the correlation coefficient, r. For the MAE and RMSE however, composite method outperforms the component-wise method. This can be explained by the high nugget effect (greater than the sill value) for the radius data. High radius nugget effect optimally captures the fuzzy uncertainty and implies that the two approaches shares approximately the same covariance structure, which is better modeled with the composite variance used in composite kriging. This makes the composite kriging a better model (for this particular dataset).

Table 7.4. Cross-validation prediction errors for Composite kriging and Component-wise Kriging compared to Best SIC2004 results (Dubois, 2005). Bold values indicate the optimal results.

Statistics	Composite $\boldsymbol{Z} = (Z_0^c, Z_0^r)$	Component-wise $\boldsymbol{Z} = (Z_0^c, Z_0^r)$	Best SIC2004 results (Dubois, 2005, p.13)
Mean Observed	(95.00, 4.83)	(95.00, 4.83)	
Mean Predicted	(94.87, 4.83)	(94.87, 4.85)	
ME	-0.13	-0.11	**-0.04**
MAE	9.92	9.99	**9.05**
RMSE	**11.97**	11.98	12.43
r	0.73	0.73	**0.79**

If we compare our results with that of "best" results obtained in SIC2004 competition, our new technique in general does not outperform "conventional" prediction methods. In 3 out of the 4 global criteria, it is only composite kriging that have the lowest RMSE. In our view, the RMSE simultaneously captured the bias plus the variance making both the new kriging systems better for this particular dataset.

To assess the local uncertainty, Figure 7.9, provides predicted lower and upper maps for both composite and component-wise kriging, compared with a map of the observed values.

Overall, both methods lead to similar prediction patterns (Figure 7.9a, b for Composite kriging and Figure 7.9e, f for Component-wise kriging). Composite kriging tend to have notably positive errors (red areas) while component-wise method tend to have errors spread over a wide spectrum (ranging from large negative(blue) to large positive(red). This could be due to the use of uncontrolled radius weight values in composite kriging §3.4, which leads to large weights being applied to radius values and hence higher variance. Comparing with Figure 7.4, the prediction errors are high (red) in the middle and slightly left of the center map (agreeing with the area of no sampled values). The interpolated maps show that there are other areas of high uncertainty and variability that may be in need of additional sampling as shown by the error maps. These regions are characterized by high predicted values, situation that could be explained by the sparse large to medium sampled values(Figure 7.4). Therefore, closest neighbor may be dissimilar although with a large weight. There is also the screening effect combined with effect of neighborhood size. A small neighborhood leads to an unbiased estimate which suffers from large variability. Increasing the neighborhood, the estimate becomes smoother, less variable and more biased.

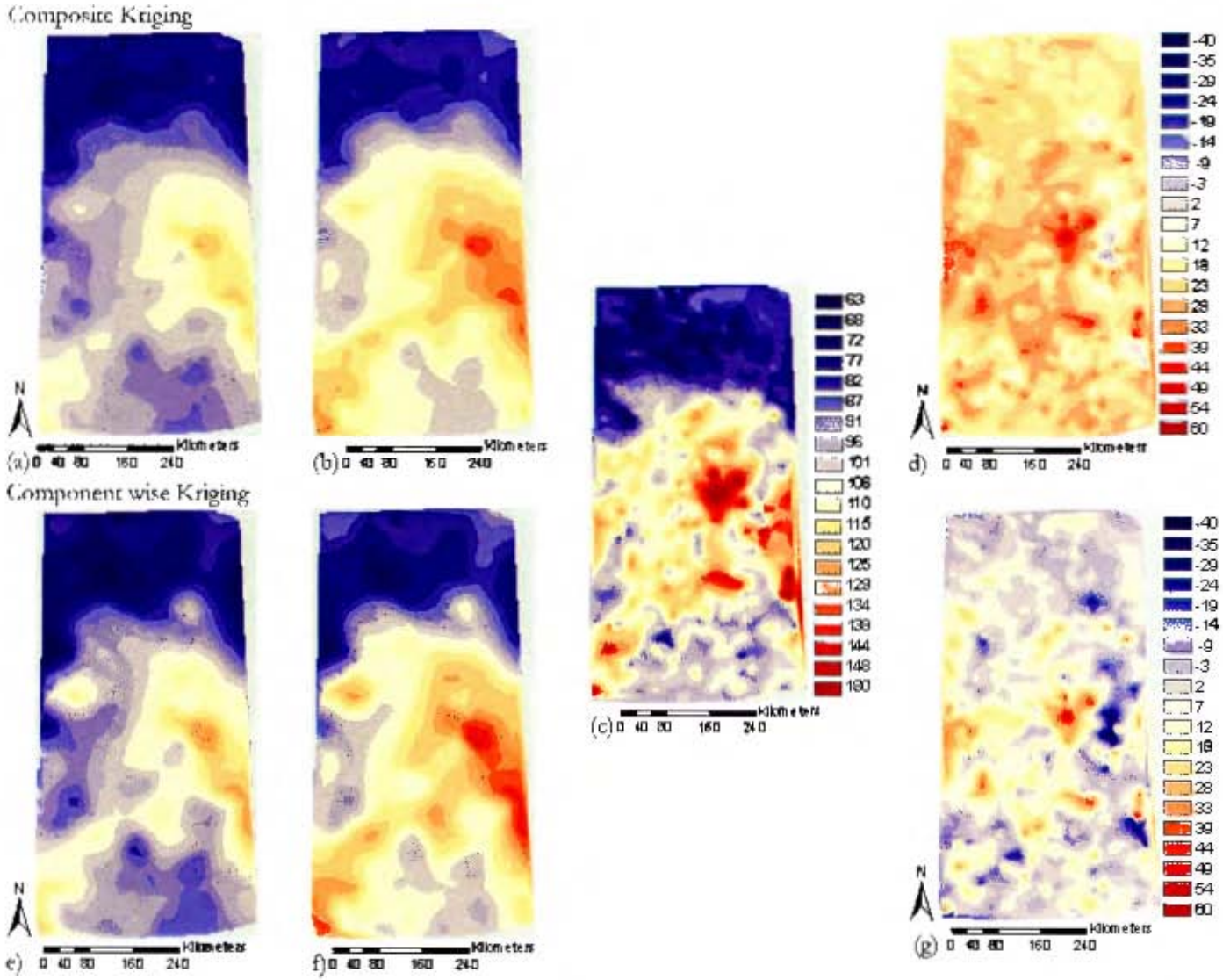


Figure 7.9. Maps of predicted gamma dose rate based on composite kriging:- a) is the lower predicted, b) the upper predicted, c) the observed values and d) the prediction error = observed- predicted; component-wise:- e) lower predicted, f) upper predicted and g) the prediction error. Despite the overall similar prediction patterns, f has a high intensity compared to b. Also the uncertainty is quite different in d (mostly red: a case of underprediction (observed value - predicted value is positive) while g has a greater range of values ranging from negative(blue) to positive (red)).

On a global criteria (RMSE) the two techniques are slightly different with composite method having a RMSE of 11.97 and component-wise a RMSE of 11.98. Composite kriging assume zero correlation structure of the radius data, so globally the error associated with the approach is smaller. In summary, while one methods performs better on global criteria the other performs better in local criteria (§7.3 page 7-10). For this particular data set the methods should be used to complement each other.

Prediction errors, based on center values ($\hat{Y}$) of the predicted intervals and the observed scalar values (Y) at 808 estimation locations are given in Table 7.5 and Figure 7.9d and g. Again the composite kriging outperforms the component-wise method on all 4 criteria.

Table 7.5. Prediction errors for Composite kriging and Component-wise kriging based on 808 observed and predicted (center) values. The composite approach performs better relative to component-wise approach.

Statistics	Composite $\boldsymbol{Z} = (Z_0^c, Z_0^r)$	Component-wise $\boldsymbol{Z} = (Z_0^c, Z_0^r)$
Mean Observed	(98.01)	(98.01)
Mean Predicted	(95.37, 4.87)	(95.45, 4.87)
ME	**-2.641**	-2.55
MAE	**9.44**	9.59
RMSE	**12.87**	13.09
r	**0.78**	0.76

7.4.2 Spatial Kernel Interval Regression

7.4.2.1 Preliminary Analysis

Prior to performing non-parametric spatial interval prediction, we investigate the probability distribution (pdf) of the random interval/vector data. The pdf reveals important features characterizing the distribution of gamma dose levels e.g. multiple modes. The pdf is estimated using bivariate kernel density estimation, with bandwidth parameter, h obtained by the cross-validation (CV) method discussed in §4.4.1. Except for the radius data, the CV method results to two bandwidths (h1, h2) for lower, upper and center data (Figure 7.10).

Bandwidth parameters (Table 7.6) were also estimated by using the mean integrated squared error approach (Schimek, 2000, MISE bandwidths) and the Silverman (1986), Scott (1992) approach (normal bandwidths) based on the assumption

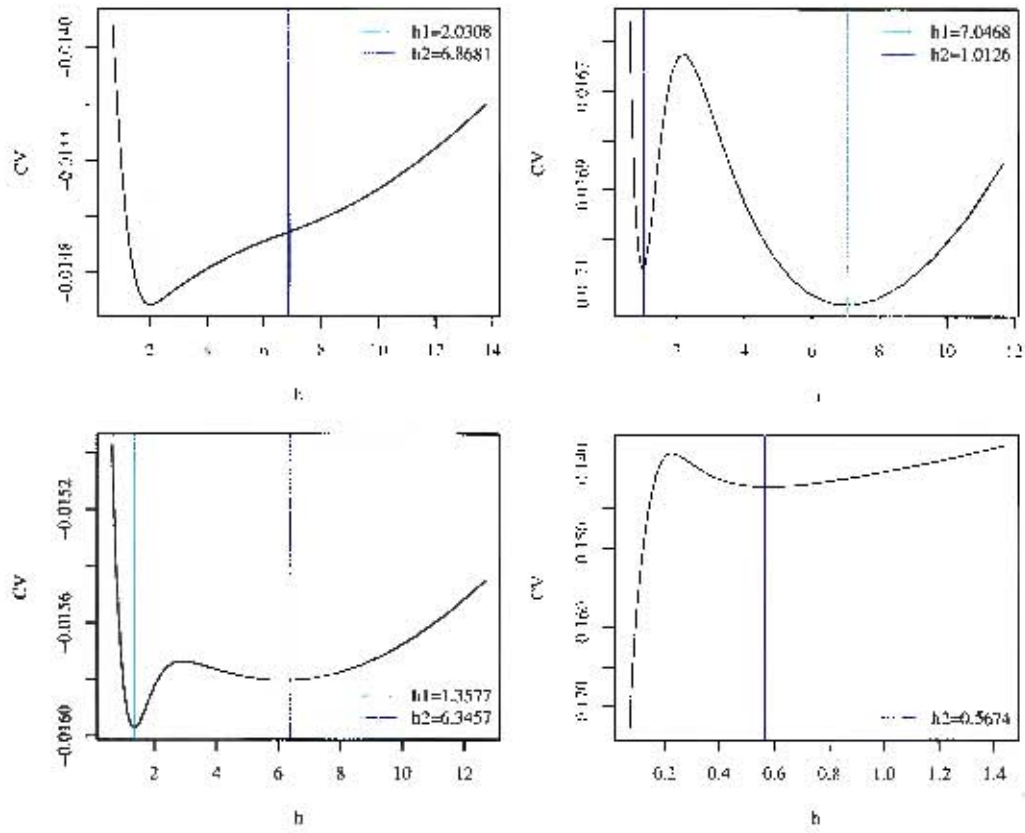


Figure 7.10. A graph of cross-validation (CV) scores vs. the global bandwidth parameter (h) for upper variable (top left) with bandwidths: h1= 2.0308; h2= 6.8681; lower variable (top right) with bandwidths: h1= 1.0126; h2= 7.0468; center variable (bottom left) with bandwidths: h1= 1.3577; h2= 6.3457 and radius variable (bottom right) with bandwidth: h2= 0.5674.

that data is normal. For all cases, the larger of the CV bandwidths (h1 or h2) are approximately equal to normal bandwidths which are approximately equal to the MISE bandwidths (Table 7.6). However, in practice the kth nearest neighbor (KNN) bandwidths (§7.4.2.2) are used. This leads to an adaptive bandwidth and hence adaptive kernel density estimation and regression. The effects of using different bandwidths in estimating a density surface are shown and discussed in Figure 7.11.

7.4.2.2 Spatial weighted-interval Prediction

Since the random intervals are distributed in a spatial region, estimation should be defined within the spatial context. Spatial joint interval kernel prediction com-

Table 7.6. MISE-based and normal-based bandwidths for lower, upper, center and radius variables.

Bandwidth	Lower	Upper	Center	Radius
Normal	5.8393	6.8717	6.3359	0.7177
MISE	5.8884	7.2403	6.7896	0.9312

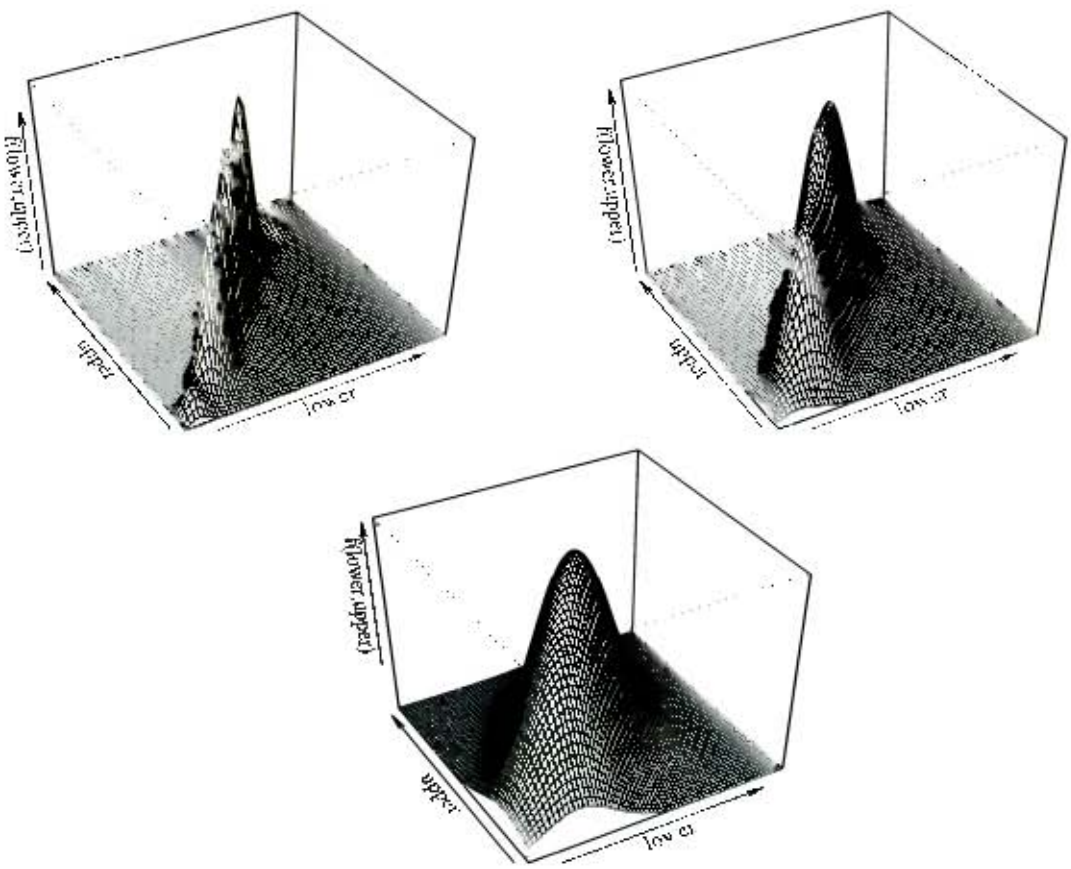


Figure 7.11. The joint probability density surface $f(\text{lower}, \text{upper})$ of random interval data. Irregular surface (top left) with half MISE bandwidths: lower =2.9441, upper = 3.6202; Optimal surface (top right) with MISE bandwidths: lower =5.8883, upper = 7.2403, and Over smoothed surface (bottom) with double MISE bandwidths: lower =11.7766, upper = 14.4806. The top right map identifies two density modes for the distribution of gamma dose rates.

bines spatial distance weighting metric (spatial kernel) with feature-space distance weighting metric (interval kernel) to estimate values at unknown location. As is the common practice, we use the kth nearest neighbor (KNN) bandwidths instead of

the global bandwidths.

Choice of kth nearest neighbor (KNN) Bandwidth Parameters

For the weight functions, Bisquare and Epanechnikov kernels were used. The two are defined on compact support, thus we only work with data in the local neighborhood, relative to Gaussian kernel where all data values are involved in the local estimation. The corresponding spatial bandwidths were estimated by applying K-th nearest neighbor (KNN) cross-validation (CV) approach (§4.4.1, Eq. 4.16). A graph of CV scores vs. K-th nearest neighbors (KNN) for the spatial kernel is given in Figure 7.12. Within the defined K-th nearest neighbors (KNN) spatial neighborhood, the interval bandwidths is given as the largest distance between the neighborhood interval values relative to the estimation point.

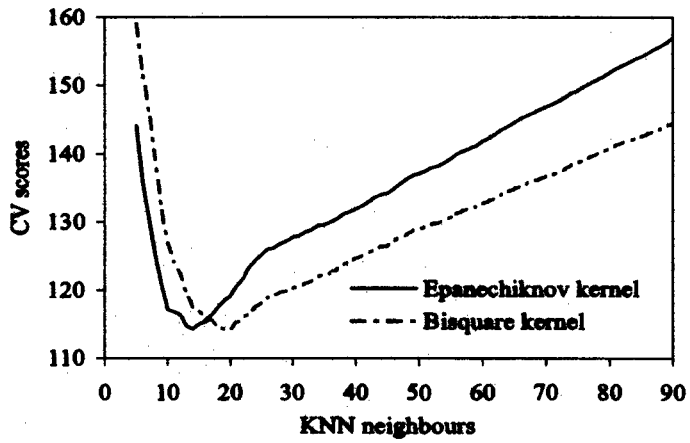


Figure 7.12. A graph of spatial KNN bandwidths for Epanechnikov kernel (14 neighbors with spatial KNN bandwidth = 53775.3 meters) and Bisquare kernel (20 neighbors with spatial KNN bandwidth = 64738.1 meters).

Spatial random interval estimates at unknown spatial locations were obtained using the mean update algorithm defined in Appendix C.1. Table 7.7 provides the cross-validation prediction errors associated with Bisquare and Epanechnikov kernel-based predictions.

The non-parametric methods gives lower RMSE and MAE relative to parametric kriging methods (Table 7.4), particularly for the Epanechnikov kernel based on 14 neighbors. The Epanechnikov kernel approach with the lowest RMSE provides better predictions. The bias of ME= 0.06 based on the Epanechnikov kernel approximately equal to the lowest ME for techniques in Dubois (2005).

Table 7.7. Cross-validation prediction errors for the spatial interval kernel predictions using Bisquare and Epanechnikov kernels.

Statistics	Nonparametric interval regression	
	Bisquare Kernel	Epanechnikov Kernel
ME	**0.05**	**0.06**
MAE	**9.12**	9.20
RMSE	11.02	**10.95**
r	0.76	**0.77**

The difference in Bisquare and Epanechnikov kernel approaches may be explained by the difference in the numbers of neighbors used. The accuracy of the non-parametric methods is further demonstrated by the slightly higher correlation coefficient (0.76 for Bisquare and 0.77 for Epanechnikov kernels) between the observed intervals and the predicted intervals as compared to correlation coefficient of 0.73 for the parametric kriging methods. But they are low compared to Dubois (2005) correlation of 0.79.

The choice of the initial estimate at the prediction location also affects the output. We demonstrate this using two initial estimates: 1) the close sample (in geographical distance) and 2) the average of the selected neighborhood values. The latter leads to slightly higher MAE = 9.19 for Bisquare kernel. Both gives a RMSE =11.10 which is smaller than the best SIC2004 RMSE of 12.43.

The slightly better performance for the non-parametric approach can be attributed to the fact that, the method optimally exploits the local spatial structure using a spatial kernel, rather than an assumed stationary covariance model (as is done in kriging). In addition the use of uncertainty feature-based kernel helps to penalize values that are very dissimilar to the local neighborhood values. This reduces the effect of any abrupt changes in the data values. Thus obtained estimates are akin to the neighborhood values. Also, there is no screening effect as encountered in kriging.

Figure 7.13 shows the predicted maps plus the prediction error maps. Both methods display prediction errors spreading over a blue to red spectrum. Bisquare has large negative errors. In both cases, high errors dominate in central region with no observed data (Figure 7.4). The smoothing nature of kernel based methods, leads to very smoothed maps that misses out the central region with large values.

Table 7.8 summarizes the prediction errors for the non-parametric methods. Calculations are based on the predicted center values at the 808 estimation locations

and the observed values at 808 estimation locations.

Table 7.8. Prediction errors for Bisquare Kernel and Epanechnikov Kernels based on 808 observed and 808 (center) predicted values using the close sample (in geographical distance (Sample)) or the average of neighborhood (Neigbr) as the initial estimate. Overall, the methods have near similar prediction errors. However, the Bisquare kernel performs slightly better than the Epanechnikov kernel for all the global criteria except the RMSE

	Bisquare Kernel		Epanechnikov Kernel	
Statistics	Sample	Neigbr	Sample	Neighbr
ME	**-2.56**	-2.59	-2.67	-2.65
MAE	**9.81**	9.84	9.83	9.87
RMSE	13.39	13.39	**13.38**	13.41
r	0.75	**0.8**	0.76	0.76

Comparing Kriging and Spatial Interval Regresssion

Based on results in Table 7.9, the method displays different levels of accuracy. Some have low bias (low ME) but less accurate as revealed by the high RMSE.

Table 7.9. Prediction errors for Kriging (Composite & Component-wise); Bisquare & Epanechnikov based on observed and predicted center values at the 808 prediction locations. The non-parametric method uses the either the close sample (in geographical distance (Sample)) or the average of neighborhood (Neigbr) as the initial estimate. Overall the kriging methods seem to perform better in predicting values (RMSE of 12.87 and MAE of 9.44), then the Bisquare approach. The insignificant difference may be due to oversmoothing nature of kernel based methods.

	Kriging		Bisquare		Epanechnikov	
Statistics	Composite	Component-wise	Sample	Neigbr	Sample	Neigbr
Mean Observed	98.01	98.01	98.01	98.01	98.01	98.01
Mean Predicted	95.37	95.45	95.45	95.42	95.34	95.36
ME	-2.64	**-2.55**	-2.56	-2.59	-2.67	-2.65
MAE	**9.44**	9.60	9.81	9.84	9.83	9.87
RMSE	**12.87**	13.09	13.39	13.39	13.38	13.41
r	0.78	0.77	0.75	**0.8**	0.76	0.76

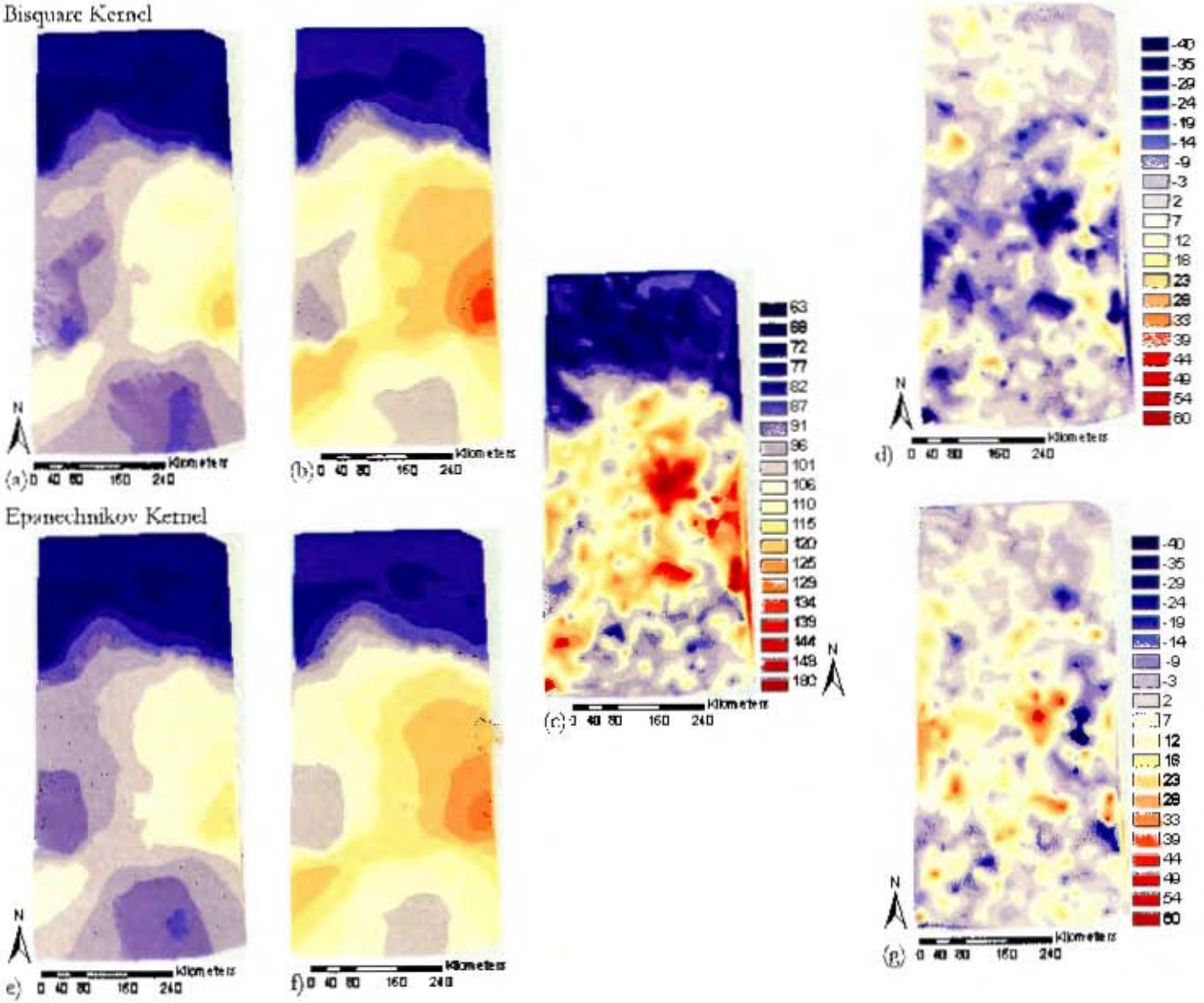


Figure 7.13. Maps of predicted gamma dose rate using Bisquare kernel:- a) the lower predicted, b) the upper predicted c) the observed values and d) prediction error = observed-predicted. Using Epanechnikov kernel:- e) the lower predicted, f) the upper predicted and g) the prediction error. The kernel based methods produces very similar smoothed patterns, typical of their nature as smoothing methods.

7.5 Bi-variable Copula-based Geostatistics

7.5.1 Preliminary Analysis: Bi-variable Copula-based Geostatistics

Prior to spatial prediction we need to model the dependence structure between the lower and upper interval values using a copula. First, we estimate the kernel joint density $\hat{f}_{Z^l Z^u}$ of the random intervals by applying Eq. 5.9. Numerical integration of $\hat{f}_{Z^l Z^u}$ with respect to Z^l and Z^u lead to the estimates of joint distribution $\hat{F}_{Z^l Z^u}$ (Eq. 5.10) and the marginals $\hat{F}_{Z^l}$ (Eq. 5.11), $\hat{F}_{Z^u}$ (Eq. 5.12). The numerical integration was done using computer code implemented in C++.

From remark 5.2.2, pg. 5-6, the graph of copula, $z = C(w, v)$ can be viewed as the graph of the joint distribution function i.e. $z = F_{Z^l Z^u}(z^l, z^u)$ in which the z^l and z^u axes have been relabeled in units of $w = F_{Z^l}(z^l)$ and $v = F_{Z^u}(z^u)$. Figure 7.14 shows the empirical kernel copula based on observed random intervals.

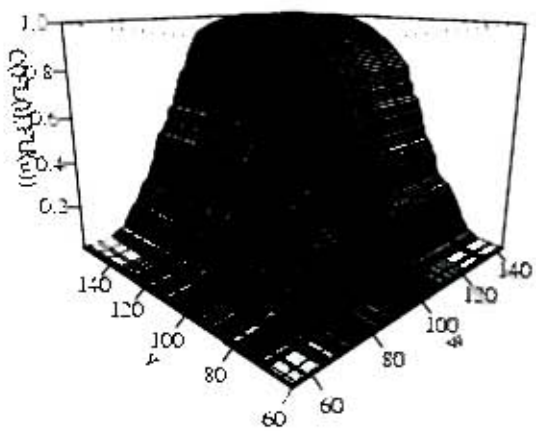


Figure 7.14. A graph of empirical copula, $C(F_{Z^l}(z^l), F_{Z^u}(z^u))$ for regionalized random intervals. The copula surface $C(F_{Z^l}(z^l), F_{Z^u}(z^u))$ is the joint distribution function surface with x and y axes as the l and u values. Copula measures the dependence between the lower and the upper values.

Applying nonlinear regression analysis, we fitted analytical model in Eq. 7.6 to the lower and upper empirical distributions.

$$F_X(x) = \frac{a}{(1 + \exp(b - cx))^{1/d}} \quad \text{for } x = \{z^l, z^u\},\ X = \{Z^l, Z^u\} \tag{7.6}$$

The model parameters a,b,c,d plus their standard errors are given in Table 7.10. Also given is the model standard error. The model standard error indicates the goodness of fit of the fitted regression line. The further away the empirical estimates are from the fitted regression line the greater the value of the model standard error. The

low model standard errors (Table 7.10) indicate good fit. Figure 7.15 represents the fitted marginal models.

Table 7.10. The lower and upper marginal distribution model parameter estimates with standard errors in brackets, for the training random intervals.

Parameters		Lower		Upper
a	1.0218	(0.0059)	1.0018	(0.0057)
b	11.0019	(0.5579)	10.3271	(0.6125)
c	0.1335	(0.0050)	0.1161	(0.0050)
d	0.7899	(0.0850)	0.7801	(0.1304)
Model stderr	0.0189	0.0191		

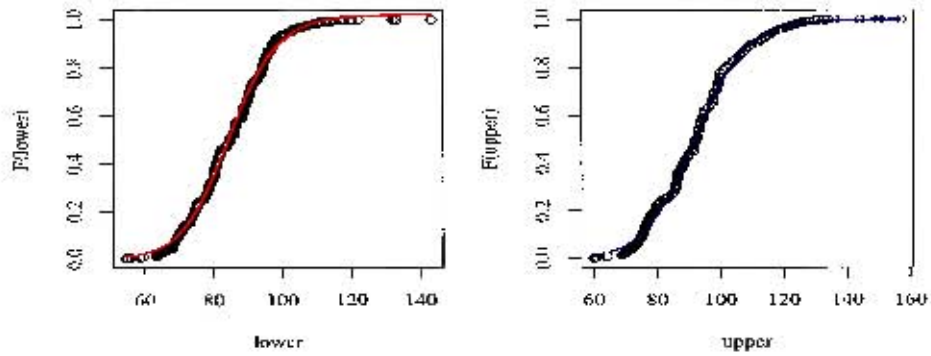


Figure 7.15. Empirical and fitted marginal functions. Left graph is the lower estimated and fitted marginals and right graph is the upper estimated and fitted marginals. The fitted marginal models accurately explain the lower and upper empirical distributions, also confirmed by the low model standard error in Table 7.10.

Using the fitted marginal models and observed random intervals at each location, we generate corresponding sample spatial marginal grades. Sample spatial copula grades are obtained from the empirical copula.

At a given spatial location, the lower marginal grade defines the marginal probability of the observed phenomena being less or equal the observed lower value and similarly for the upper marginal grades. Cumulatively, they provide the estimate of the lower and upper spatial distribution grades at given locations. The sample copula grades define the joint probability of the lower and upper values. Copula + marginal grades provide the full information associated with each spatial bivariate random vector. The collection of sample spatial copula grades results to a copula random function which is then subjected to spatial prediction. The same applies to the spatial samples of marginal grades.

7.5.2 Structural Analysis: Bi-variable Copula-based Geostatistics

Structural analysis on sample spatial copula grades and sample spatial marginal grades, based on lag size of 20,000 meters lead to variogram models (Figure 7.16).

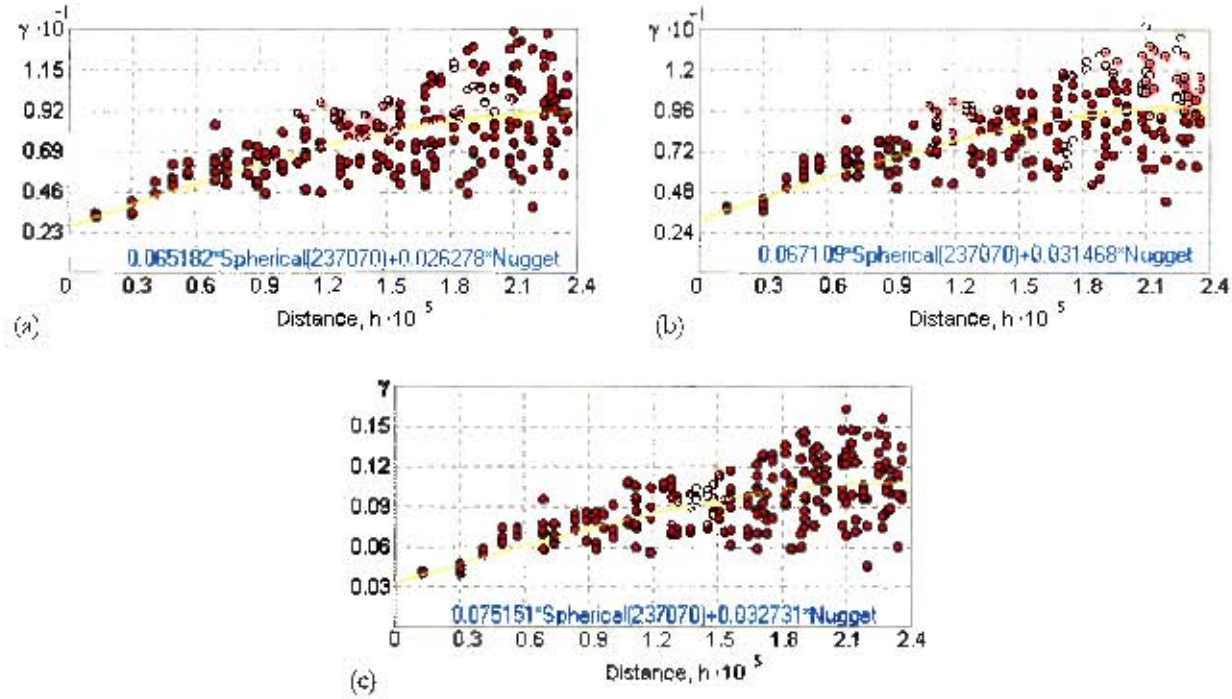


Figure 7.16. A graph of fitted spherical variogram models with model parameters inset for :-a) upper marginal grades, b) lower marginal grades and c) copula grades. The models are used for kriging of associated marginal and copula grade random fields. Though the variogram models display very similar basic structures, the copula-variogram indicates a slightly higher correlation structure, which might lead to better prediction. The upper model has lower nugget effect compared to lower and copula models. This has the effect of reducing uncertainty associated with upper grade predictions. The models also increases slowly from the origin indicating a smooth process.

7.5.3 Prediction: Bi-variable Copula-based Geostatistics

Kriging on copula and marginal grades, based on a neighborhood of 20 observations leads to predicted maps (Figure 7.17).

The probability prediction of gamma dose rate (Figure 7.17) is displayed on a scale from low probability (blue) to high probability (Red). Three prediction maps are generated, one showing the prediction probabilities for the lower values, one

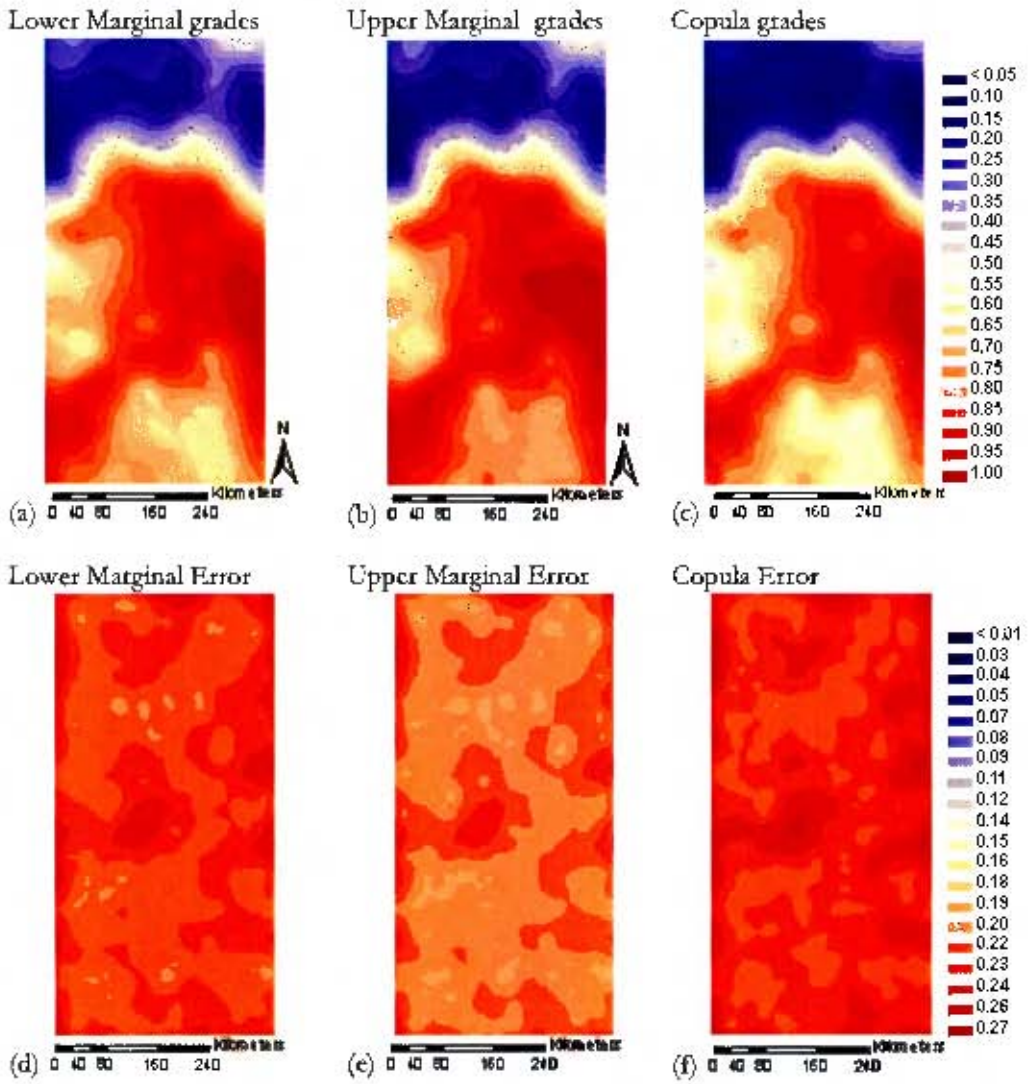


Figure 7.17. Maps of predicted probabilities of gamma dose rates at the 200 monitoring stations in Germany based on a) Lower marginal grades (b) Upper marginal grades (c) Copula grades with corresponding prediction errors d)- lower, e)- upper and f) copula. Lower map (a) is the spatial distribution of the lower interval values; upper map (b), the spatial distribution of the upper values; and copula map (c) the spatial joint predicted probabilities of the lower/ upper values. The maps indicate a high likelihood of lower gamma dose values in the north. Pockets of large probabilities in the south-east region indicate regions likely to be exposed to large gamma dose values. Copula has the higher uncertainty (f) and hence is more unstable compared to the lower and upper map. Large prediction errors occur in regions with no observed data, (see Figure 7.4). Similar to above methods, pockets of high values occurring together with high errors (see central region).

showing the predicted probabilities of the upper values and one map showing the predicted joint probabilities (copula). The interpretation of the three maps should

be done in triplet (lower, upper and copula) plus the associated uncertainty maps.

For lower variable values, low predicated probabilities in the north (Figure 7.17a) indicate low gamma dose values. High gamma concentration occurs towards the south as revealed by corresponding high predicted probabilities in red. This is in line with the observed lower variable values in Figure 7.4, left. A similar pattern is shown for the upper marginal grade map (Figure 7.17b). In addition, the upper marginal map identifies more regions of high gamma concentrations compared to the lower marginal grade map (Figure 7.17a). Though the lower and upper maps were modeled independently, they are complementary maps.

The copula map (Figure 7.17c) indicates the predicted joint probabilities of lower and upper variable values. It reveals the spatial joint distribution of the lower and the upper variable values and show better prediction compared to marginal grade maps. High predicted joint probabilities in south-east direction, reveals strong persistence in occurrence of high values. Low predicted joint probabilities in the north shows strong persistence of low gamma dose values in the north. Therefore the probability of observing a large value towards the south is high. As a joint distribution, the copula map clearly identifies the spatial behavior of the random interval, but with high uncertainty. This confirms the separate patterns by lower and upper grade maps. Relatively, the copula map provides more distinct patterns of the spatial distribution of the gamma dose.

Based on prediction error maps, Figure 7.17d,e,f for lower, upper and copula predicted probabilities, High error (red) indicates a high prediction errors, and occur in the central region with no sampled data (see Figure 7.4). However, overall upper marginal predictions has lower prediction errors (Figure 7.17e). Copula (Figure 7.17f) has the highest prediction error. We also note pockets of high errors associated with high predicted probabilities. This is mostly in the with mixed data and/or no data.

We also generated the associated cross-validation prediction errors. Cross-validation was performed by removing one observation from the set of sample locations, performing the kriging, comparing the kriging to the left out observation, and then cycling through all the observations in all events. Summary statistics of the cross-validation prediction errors for lower, copula and upper predictions are given in Table 7.11. The copula has large RMSE compared to the rest. The upper grade provides slightly accurate predictions, as indicated by the more narrow error distribution and a lower RMSE (this is confirming our observations and conclusions from Figure 7.17).

Table 7.11. Cross-validation prediction errors for lower, copula and upper marginal grade predictions based on observed random intervals.

Statistics	Lower	Copula	Upper
ME	-0.0006	**-0.0004**	-0.0005
RMSE	0.2180	0.2081	**0.1925**

7.6 Fuzzy Credibility-based Geostatistics

7.6.1 Preliminary Analysis: Fuzzy Credibility-based Geostatistics

The fuzzy credibility approach defines a single flexible joint distribution model of distribution of the random interval. The approach is based on fuzzy variable set ξ corresponding to the random intervals. The fuzzy variable set consist of all values of observed gamma dose rates with membership grades assigned using an empirically defined data-assimilated membership function.

Applying Eq. 6.20, the membership function, $\mu_\xi(\cdot)$ for the induced fuzzy variable set ξ is first estimated through numerical integration (using a C++ computer code) of the joint density function under the constraint: $\sup\{\mu_\xi(\cdot)\} = 1$. Obtained membership function is then modeled with a lower function, $l(z)$ increasing from zero to one, the core with $\mu_\xi(\cdot) = 1$ and the upper function, $r(z)$ decreasing from one to zero. In Table 7.12 we give the model parameters with standard errors in brackets of the fitted lower $l(z)$ and upper $l(u)$ membership functions. The functions were estimated using linear and non-linear regression methods (using r-code)

Table 7.12. Empirical lower $l(z)$ and upper $l(u)$ data-assimilated membership function parameter estimates with standard errors in bracket.

	Lower	Upper
Parameters	$l(z) = a + bz + bz^2 + bz^3$	$u(z) = a\exp(-(b-z)^2/(2c^2))$
a	1.9517 (0.3535)	1.0003(0.0034)
b	-0.1097(0.0139)	103.7329 (0.1483)
c	0.0018 (0.0002)	14.2983 (0.0927)
d	-0.000008 (0.0000007)	
Model Stderr	(0.0124)	(0.0068)

Applying Eq. 6.4, together with $l(z)$ and $u(z)$ membership function (Table 7.12),

the data-assimilated membership model is given as

$$\mu_{\xi}(z) = \begin{cases} 1.9517 - 0.1097z + 0.0018z^2 - 0.000008z^3 & \text{if } 52.795 \leq z < 102.186 \\ 1 & \text{if } 102.186 \leq z \leq 103.780 \\ 1.0003 \exp\left(-\frac{(z-103.7329)^2}{408.89}\right) & \text{if } 103.780 < z \leq 157 \\ 0 & \text{otherwise} \end{cases} \tag{7.7}$$

Figure 7.18 shows the modeled membership function.

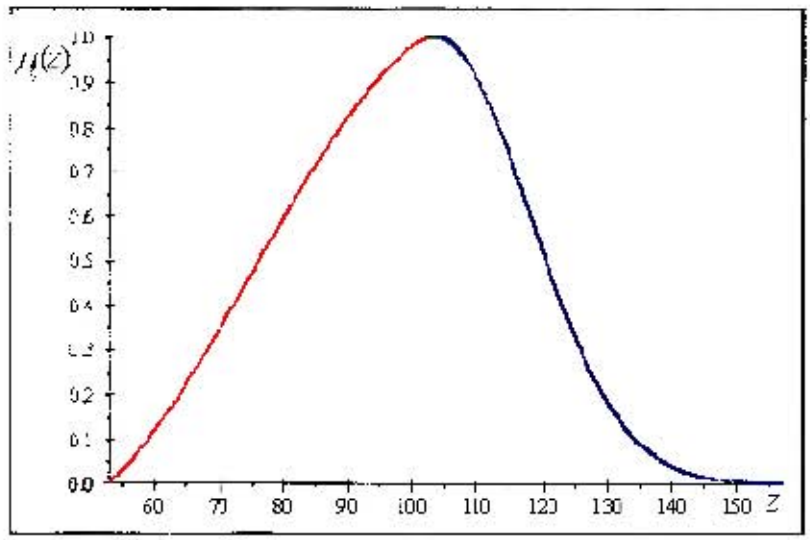


Figure 7.18. Membership function for the induced fuzzy variable ξ. Red curve is monotone increasing to 1, green is the core and blue curve the monotone decreasing from 1 to zero. Low and high values are assigned low membership grades, with the most frequent and probably acceptable values having high membership grades.

The membership function assigns membership grades to observed fuzzy variable events $z \in \xi$. Extremely large (unlikely) and/or no events are assigned low and/or zero membership grades. The most likely events assume large membership grades. Since unlikely and/or no events are assigned almost similar membership grades, we have a problem in distinguishing between low and high gamma dose values.

Therefore, we need to define another function which is non-decreasing from zero to one and is based on the derived membership function. The non-decreasing function will enable us to assign zero grade to small and/or no fuzzy events and maximum grade of 1 to extreme large values. This will allow us to separate the low from the high gamma dose levels.

Eq. 6.14 establishes the general relationship between the membership function

and an increasing function 'the credibility distribution' of the fuzzy variable set. Eq. 6.21 provides the derived data-assimilated credibility distribution derived from data-assimilated membership function of the induced fuzzy variable set. We can use the credibility function just as we can use the membership function (Remark 6.4.1, p.6-9)

Applying Eq. 6.14, together with membership function, Eq. 7.7, the data-assimilated credibility distribution becomes:

a) For $z \in [52.795, 102.186)$, $\mu(z) = 1.9517 - 0.1097z + 0.0018z^2 - 0.000008z^3$ (Figure 7.18, red curve) is monotone-increasing from 0 to 1. Therefore,

$$\begin{aligned} & \sup_{y \le z} \left(1.9517 - 0.1097y + 0.0018y^2 - 0.000008y^3\right) \\ = \quad & 1.9517 - 0.1097z + 0.0018z^2 - 0.000008z^3, \ \forall z \in [52.795, 102.186) \end{aligned} \tag{7.8}$$

and

$$\sup_{y > z} \left(1.9517 - 0.1097y + 0.0018y^2 - 0.000008y^3\right) = 1, \quad \forall z \in [52,795, 102.186) \tag{7.9}$$

Applying 7.8 and 7.9,

$$\begin{aligned} \Phi_\xi(z) &= \frac{1}{2}\left(\sup_{y \le z} \mu(y) + 1 - \sup_{y > z} \mu(y)\right) \\ &= \frac{1}{2}\left(1.9517 - 0.1097z + 0.0018z^2 - 0.000008z^3\right) \end{aligned} \tag{7.10}$$

b) For $z \in (102.186, 103.780]$, $\mu(z) = 1$. Applying 6.10

$$\Phi_\xi(z) = \mathrm{Cr}\left(\{\theta : \xi(\theta) = z\}\right) = \frac{1}{2} \tag{7.11}$$

c) For $\forall z \in [103.780, 157)$,

$$\mu(z) = 1.0003 \times \exp\left(-\frac{(z - 103.7329)^2}{408.89}\right)$$

is monotone-decreasing, then

$$\sup_{y \le z} \left(1.0003 \times \exp\left(-\frac{(z - 103.7329)^2}{408.89} \right) \right) = 1 \tag{7.12}$$

and

$$\begin{aligned} & \sup_{y > z} \left(1.0003126 \times \exp\left(-\frac{(z - 103.73288)^2}{408.89} \right) \right) \\ = \quad & 1.0003126 \times \exp\left(-\frac{(z - 103.73288)^2}{408.89} \right) \end{aligned}$$

thus, applying 7.12 and 7.13

$$\begin{aligned} \Phi_\xi(z) &= \mathbf{Cr}\left(\{\theta : \xi(\theta) \le z\}\right) \\ &= 1 - 0.500\,16 \exp\left(-\frac{(z - 103.73288)^2}{408.89} \right) \end{aligned} \tag{7.13}$$

To summarize, combining 7.10, 7.11 and 7.13, the data-assimilated credibility distribution for the induced fuzzy variable is

$$\Phi_\xi(z) = \begin{cases} 0 & \text{if } z < 52.795 \\ \frac{1}{2}\left(1.9517 - 0.1097z + 0.0018z^2 - 0.000008z^3\right) & \text{if } z \in [52.795, 102.186) \\ \frac{1}{2} & \text{if } z \in [102.186, 103.780] \\ 1 - 0.500\,16 \exp\left(-\frac{(z-103.73288)^2}{408.89} \right) & \text{if } z \in (103.780, 157.000] \\ 1 & \text{if } z > 157.000 \end{cases}$$

Figure 7.19 is a graphed representation of the data-assimilated credibility distribution for the fuzzy variable set ξ induced by the random interval $\left(Z^l, Z^u\right)$. High dose values are assigned high credibility grades. The credibility grade of 0.5 corresponds to a membership grade of 1 and denotes level beyond which gamma dose rates starts posing a risk (according to the normal acceptable threshold levels of 50 to 120 nSv/h CTIF et al., 2006; Somerville and McMahon, 2006).

The data-assimilated credibility distribution indicates the marginal increase of ones strength of belief that observed gamma dose rates belongs the defined fuzzy variable set i.e. it is acceptable and to what degree. Simply put, it indicates the likelihood or favorability of a value belonging to a fuzzy variable set or a fuzzy variable assuming a given value.

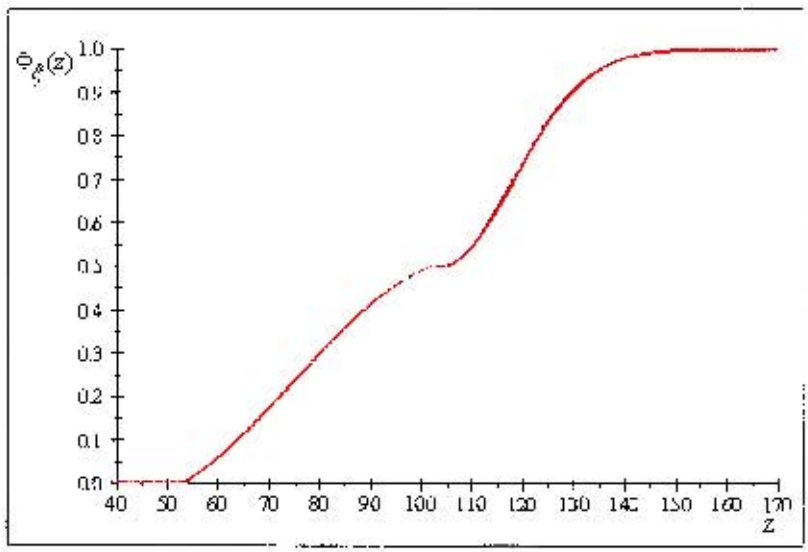


Figure 7.19. The data-assimilated credibility distribution $\Phi_\xi(z)$ for the induced fuzzy variable set ξ.

High credibility grades indicate the extent of the fuzzy phenomena being less or equal to the extreme large fuzzy measures of gamma dose rates. Thus, high credibility grades indicate hazardous conditions while low credibility grades indicate safe conditions.

7.6.2 Structural Analysis: Fuzzy Credibility-based Geostatistics

The center values optimally approximate the random intervals, and therefore provide scalar-valued fuzzy measures of the random intervals. Using ArcGIS, structural analysis on sample spatial credibility grades (Eq. 6.24) associated with center values based on a lag size of 20,000 meters lead to variogram models (Figure 7.20).

7.6.3 Prediction: Fuzzy Credibility-based Geostatistics

Kriging on sample spatial credibility grades based on a neighborhood of 20 observations leads to predicted maps (Figure 7.21). The credibility of gamma dose rate in Figure 7.21 is displayed on a scale from low credibility (blue) to high credibility(Red). High predicted credibility indicate regions with high marginal increase in concentrations of gamma dose while low predicted credibility indicate safe regions with low gamma concentrations. Pockets of large credibility grades (red) in the

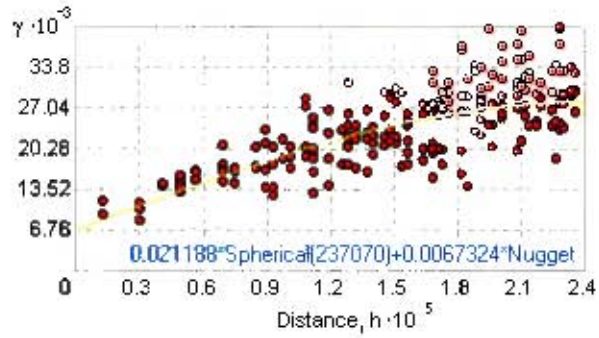


Figure 7.20. A graph of credibility grade spherical variogram models defined on sample spatial credibility grades generated using data-assimilated credibility function on the center values. The model parameters nugget = 0.006734, sill = 0.02118 and range=237070. The models increase slowly from the origin, an indication of the smoothness of the random process.

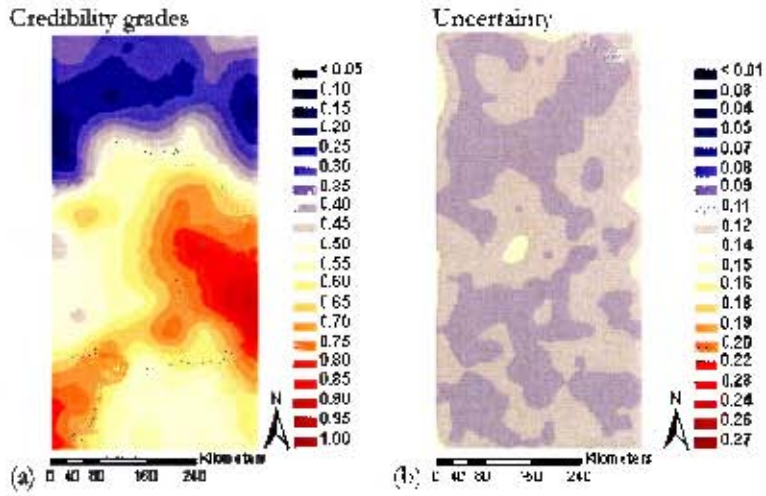


Figure 7.21. A map of predicted credibility grades of gamma dose rates based on credibility grades defined on fuzzy measures using the data-assimilated credibility distribution model. a) is the credibility grade map with b) the uncertainty map. The credibility grades indicate the marginal increase in the extent of pollution. High grades indicate regions of prone to high pollution. Pockets of large credibility grades (red) in the south east region indicate large concentration of dose levels. Low credibility grades in the north are attributed to low gamma dose levels. Predicted map is more variable (clear pattern) in the more densely sampled areas than the sparsely sampled areas. Prediction errors are quite low with highest prediction error (as expected) in the no sample area (yellow pocket).

southern region indicate large concentration of dose levels. Low credibility grades in the north are attributed to low gamma dose levels.

Credibility prediction has much lower uncertainty (Figure 7.21b) compared to Bi-variable copula-errors (Figure 7.17 d, e,f). Hence credibility-based method pro-

vides more accurate predictions. The improvement over the copula approach is further demonstrated by the low RMSE using the cross-validation prediction errors (Table 7.13).

Table 7.13. Cross-validation prediction errors for credibility prediction on center fuzzy measures. The method predicted well with low bias (low mean error) and high accuracy as revealed by low RMSE

Statistics	Center
ME	-0.0003
RMSE	0.1145

Generality of credibility grade kriging

The generality of credibility approach means that, if we have another set of fuzzy measures of gamma dose at the 200 sampling locations, we can use the defined credibility distribution model for the process, generate credibility grades and perform prediction. To demonstrate this, we apply the credibility approach on a provided extra routine dataset for the 200 sampling locations, in addition to the 10 training datasets. Predicted credibility grade map based on credibility grades derived from the routine dataset is given in Figure 7.22.

7.7 Bi-variable Copula-based vs. Fuzzy Credibility-based Geostatistics

To compare the bi-variable copula-based and fuzzy credibility-based predictions Figure 7.23 looks at gamma dose prediction in Germany applying bi-variable copula kriging and credibility grade kriging standardized to a common classification.

Figure 7.23d based on credibility approach give more separation of the gamma dose polluted regions compared with the copula and marginal maps (Figure 7.23a,b and c). The copula and marginal maps seem to be overestimating in some regions (dark red areas). This may be due to over-inference (a characteristic of probability distribution-based inference).

Predictions based on the routine dataset (Figure 7.23e) provide more distinct demarcation of polluted regions. This demonstrates the flexibility of use of the credibility based approach. Using inverse credibility function predicted credibility grades can be converted back to original scale and results interpreted using original values.

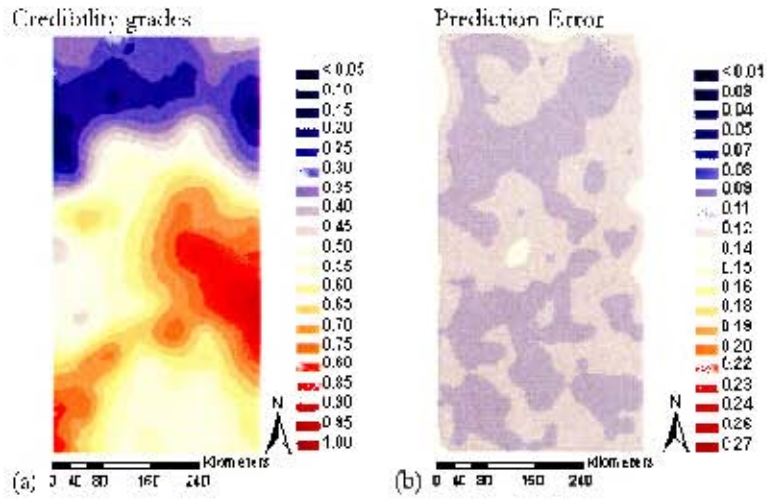


Figure 7.22. A map of predicted credibility grades of gamma dose rates based on credibility grades defined on routine dataset by applying the data-assimilated credibility distribution model. a) is the credibility grade with b) uncertainty map. This map provide clear pattern of the distribution of gamma dose rates, confirming the generality of fuzzy credibility approach. Predictions are accurate with low uncertainty (prediction error). High predicted grades (red) indicate regions of prone to high pollution. Pockets of large credibility grades (red) in the south-east region indicate large concentration of dose levels, while low credibility grades dominate in the north indicate low gamma dose levels. Overall, prediction errors are quite low, except for the regions with no sampled data.

The credibility approach simplifies the calculations with better predictions compared to the copula based approach and applies well for this particular dataset.

Cross-validation prediction errors for the bi-variable copula-based and fuzzy credibility-based approaches are summarized in Table 7.14.

Table 7.14. Cross-validation prediction errors for copula, lower, upper and credibility grade kriging techniques. The methods performed well as revealed by the low error distributions. The copula has the largest RMSE. The credibility based method provides the most accurate predictions, with the lowest RMSE

Statistics	copula	lower	upper	credibility
ME	-0.0006	-0.0004	-0.0005	**-0.0003**
RMSE	0.2180	0.2081	0.1925	**0.1145**

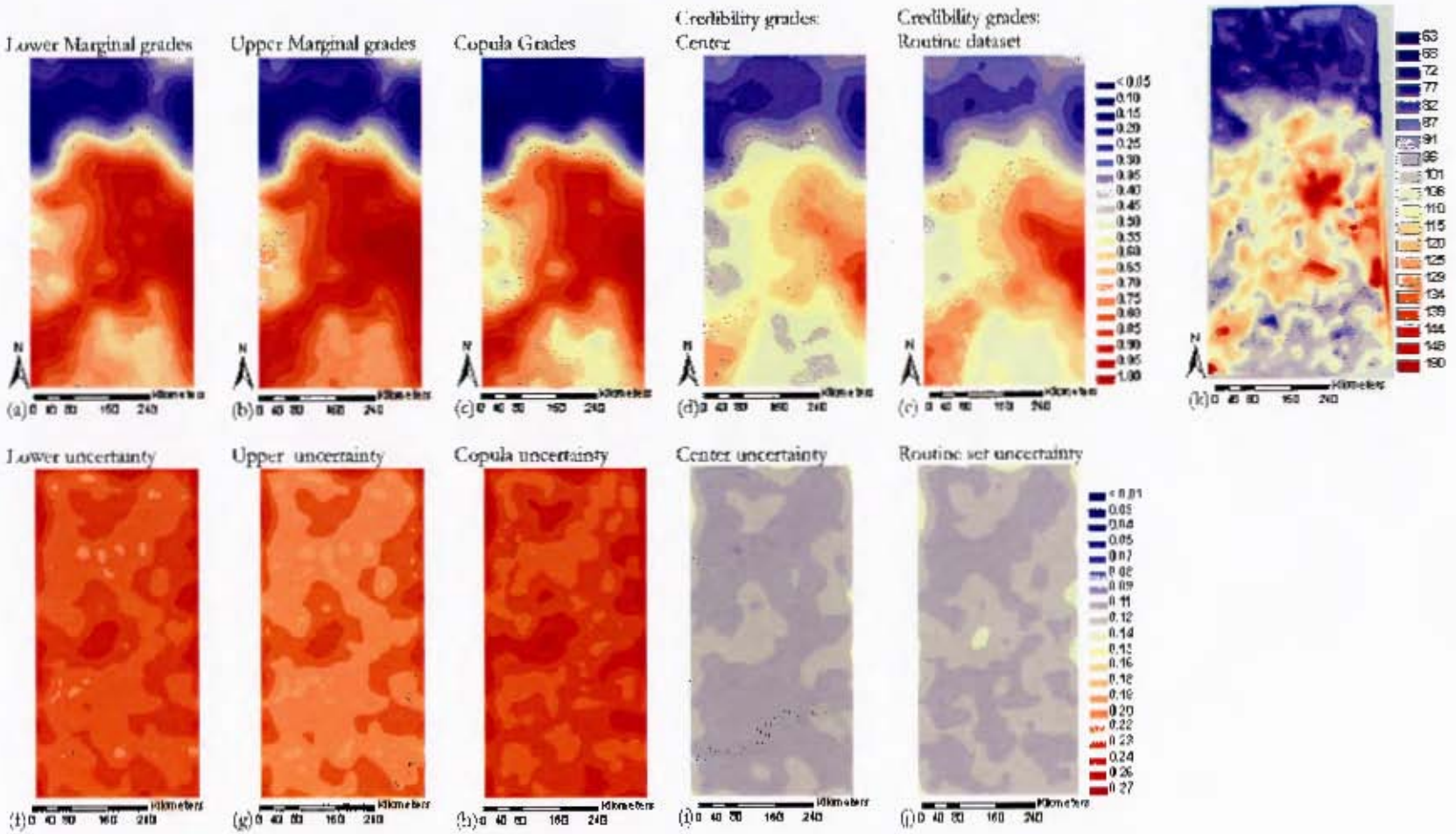


Figure 7.23. Comparison of the gamma dose predicted probabilities:- a) Lower, (b) Upper, (c) Copula (joint) predicted probabilities with predicted credibility grades based on:- d) center sample measures and (e) routine set. Respectively the prediction errors are f) Lower, g) upper, h) copula, i) center credibility and j) routine set credibility prediction errors. Top right panel (k) is a kriged map of the observed values at 808 estimation locations. Relative to panel (k), the credibility maps (d,e) are superior to the rest in alienating regions prone to high gamma dose rates and have the lowest uncertainty (i,j). In all cases pockets of high error occur in areas with no sample values (see Figure 7.4).

7.8 Overall assessment of methods

These comments are purely based on the application of methods on one single dataset.

- Proposed composite and component-wise kriging methods intuitively reduced uncertain interval data to scalar data values, the center and radius the interval uncertainty measure. This allows exploring the distribution of uncertainty around the center values, weighting the influence of extreme features according to their relative importance to center.
- Intuitively, the composite and component-wise methods reduces interval-based methods to combined and separate variable based methods using the center and radius data. The approaches are easier and leads to same correlation conclusions to interval-based methods. This was demonstrated noting that composite kriging is the same as interval-based kriging (parallel to Diamond (1988) approach), while component-wise is a generalization of the two. With little or no vague information (low uncertainty) hence narrow interval bounds, the methods lead to very similar results.
- The spatially-weighted interval kernel approach adapts well to the nature of the application with no stationarity assumptions.
- Empirical copula map (Figures 7.17c) indicates asymmetrical random interval dependence structure across space, with strong high dependence of large gamma dose rates in varying regions and strong low dependence structure for small gamma dose rates in the north. Hence, large values often persists together and similarly for low gamma dose rates. Regions with a mixture of large and small dose rates have weak dependence structure. These are the regions with sudden shift in gamma dose levels.
- The marginal grade maps independently show the predicted cumulative probabilities of lower and upper gamma dose rates. A decision maker can then decide on the acceptable cumulative probabilities.
- In terms of prediction errors, the copula uncertainty (Figures 7.23h) is higher compared to that of lower and upper marginal maps (Figures 7.23f,g) which are higher than for the credibility approach with the lowest uncertainty (Figures 7.23i). The credibility approach is thus more stable.
- The difficult associated with copula approach is the interpretation of 3 complementary predicted maps (lower, upper and copula).

- The fuzzy credibility-based approach utilizes a single full and flexible joint distribution model of distribution of the random fuzzy variable, the equivalent of random interval. It is therefore simple and flexible to use as opposed to bi-variable copula approach that separate the joint distribution function into copula and marginal components.
- Since the copula, marginal and credibility maps are base on joint distribution of the random interval/vector data, they have much to recommend over the alternatives e.g. interval kriging and interval kernel regression which are limited to the observed data. The derived distribution-based models
  - → are information-intensive and forward-looking. Hence, they are capable of incorporating a wide range of future eventualities that simply are not captured using observed data.
  - → do not require a large amount of data in order to be estimated accurately
  - → are capable of reflecting a change in the observed phenomena. A sudden shift in observed phenomena and/or other related factors could be immediately captured in observed data and the implied distributions.
- The copula, marginal and credibility methods do compensate each other and easily integrates with most spatial and GIS mapping software which is often based on scalar variables. This is of great convenience and provides effective communication of spatial patterns for fuzzy data given as intervals of values. The approaches also generalize the indicator kriging approach that transforms data into indicator codes.
- It is obvious there are some differences between the maps according to the technique applied. This difference may be attributed to the way the marginal, copula and membership functions are estimated from the respective data values. For the purpose of this work we have shown that each can be effectively applied to effective spatial mapping of the fuzzy observations given as intervals. In future, the choice, effectiveness and appropriateness of each approach will depend on the
  - → application (type of data)
  - → complexity of the implementation
  - → problem that the decision maker wants to address.

CHAPTER 8

CONCLUSIONS AND FUTURE DIRECTIONS

Scope

This chapter summarizes our achievements in view of the stated objectives. It also highlights the limitations and provides perspectives for future research.

8.1 Introduction

The main motivation for this study was to find innovative approaches to jointly model data and uncertainty (in the form of randomness and fuzziness). The model aids decision making in the presence of uncertainty characterized by random interval and fuzziness (or imprecision). Specifically, we addressed the following objectives:

(i) extended kriging to modeling spatial random intervals/ vectors;

(ii) introduced a non-parametric kernel method for analysis of non-stationary spatial random interval process;

(iii) introduced bi-variable copula-based geostatistics for variable modeling of spatial random intervals;

(iv) introduced fuzzy credibility-based geostatistics for variable modeling of spatial random interval events;

(v) illustrated and tested the flexibility of the proposed techniques to real-life applications.

We summarize the major contributions toward achieving the stated objectives under the headings:- Classical random interval geostatistics; Bi-variable copula-based geostatistics and Fuzzy credibility-based geostatistics.

Our main contribution was to apply the proposed techniques to modeling ambient gamma dose pollution based on uncertain/fuzzy spatial data. The goal is to

→ to estimate values at unknown locations and assess the uncertainty attached to the prediction

→ compute indicators e.g. probability of exceeding any critical threshold levels that help in accurate delineation of polluted regions for risk assessment and remediation.

Often the probability of pollutants concentrations lower than a threshold level is in some cases more important that the best estimate of the pollutant concentrations for the unsampled locations in delineating polluted areas.

8.2 Summary and Conclusions

8.2.1 Classical Random Interval Geostatistics

We examined (i) kriging of spatial random intervals and (ii) non-parametric spatial-weighted interval kernel regression methods, for spatial prediction of gamma dose pollution, with a goal of assessing the magnitude and geographical extent of a radioactive contamination in the air and on the ground.

Kriging of Spatial Random Intervals

We proposed and developed composite and component-wise kriging. Composite kriging models spatial random intervals when the measure of uncertainty has zero spatial correlation structure while Component-wise kriging applies when data uncertainty displays well defined spatial correlations structure. Therefore, composite approach, which parralels Diamond (1988) interval kriging method is special case of component-wise kriging. For this study, the data uncertainty was almost a nugget effects model, favoring composite kriging. Results based on the composite and component-wise were accurate as revealed by the low mean error close to zero, low mean absolute error and low root mean square error.

For decision making, predicted maps identified regions prone to high gamma dose concentrations. Regions with no sample data were also clearly highlighted by high prediction error.

If we compare our results with that of "best" results (Dubois, 2005), the new technique in general does not outperform "conventional" prediction methods, for this particular dataset. However, we note that, the separation of imprecise fuzzy data into center data and radius (uncertainty) measure, forms a first attempt in search of variable-based modeling spatial random intervals. The techniques reduced interval-based methods to combined and/or separate center and radius-based methods, easier to handle and leads to similar correlation conclusions to those based on interval-based methods.

Spatial-weighted Interval Kernel Regression

Developed spatial-weighted interval kernel regression focused on non-stationary random interval fields, addressing the stationarity problems associated with kriging. The approach gave better predictions as revealed by slightly lower root mean square error (increased accuracy) and low mean error (low bias). Further, the method

reported a higher correlation coefficient between the observed and the predicted spatial random intervals. Typical of smoothing methods, the methods produced smoothed prediction maps.

A key advantage associated with the spatial-weighted interval kernel approach is the combined spatially-weighted interval kernel, a product of a spatial kernel and an interval feature kernel. The spatial kernel defines the local spatial structure. This ensured that only samples sharing similar intensity and gradient information are gathered for local approximation. The interval kernel is a robust uncertainty kernel that penalizes outliers, minimizing their influence to local prediction. The result is optimized estimates akin to neighborhood values.

Choice of the initial estimate at the prediction location provided an added flexibility to the approach. Using nearest neighbor sample value, only applicable when minute details are of interest, as initial estimate lead to bias (low ME) compared to use of average of the selected neighborhood values. This identifies regions of high dose concentrations while effectively smoothing out the process.

In summary, these two approaches applies a very subtle form of scalar-variable modeling of spatial random intervals via the lower and upper or the center and radius coordinate variables. However, both lead to interval outputs, given in form of lower and upper prediction maps plus an prediction error map.

8.2.2 Variable Modeling of Spatial Random

The variable-based approaches, constituting the bi-variable copula-based and fuzzy credibility-based geostatistics utilized the joint distribution function of the random interval data for optimal prediction.

Bi-variable Copula-based Geostatistics

The copula logic exploited the joint distribution of random interval data through its copula and its marginal distribution functions. For threshold values corresponding to sample values, each random interval had an associated copula and marginal (lower and upper) value (grade) in the range 0 to 1. Generated marginal (lower and upper) and copula grade vectors were treated as random functions, thus variographic analysis and ordinary kriging was applicable. This yielded copula and marginal prediction maps of the equivalent spatial distribution functions.

Predicted maps identified regions of high dose concentrations. The methods were also unstable as shown by the high uncertainty maps. One may claim that estimated

spatial marginal distributions could have sufficiently described the observed spatial patterns. However, the high correlation coefficient between the observed lower and upper gamma dose, modeled by the copula, revealed a significant relationship between them. High dose values persist together, and similar to low values. Intuitively then, the copula map reveals the spatial distribution of the joint distribution of random interval data.

Bi-variable copula approach can be regarded as alternative to indicator kriging (IK) for random intervals; It is equivalent to indicator kriging of the marginal and joint probability when all data are retained in the kriging system, and offers several benefits:-

- it requires less parameterization. The copula approach requires only the copula and marginal variograms, on average well behaved, brings no order relation problems, and transforms prediction with spatial random interval into an easy task.
- the direct coding of the conditioning data using analytical marginal models allows for dynamic relocation of predicted grades to the conditioning data values using inverse functions.
- by splitting conditions up into a marginal and dependence part, the approach gives better insight into the laws governing the spatial behavior of the random interval/vector process.

Fuzzy Credibility-based Geostatistics

We demonstrated the ease of application of fuzzy credibility kriging method for modeling samples of imprecise measures (intervals) of gamma dose rates. The conversion of a random interval data to a scale random fuzzy variable with maximum entropy and fully characterized by a credibility distribution function, at same footage of probability space, is a key achievement.

The approach lead to real (scalar)-value spatial predictions. Existing fuzzy geostatistical methods are based on fuzzy set and lead to fuzzy set(subset) estimates/predictions.

Results were accurate as revealed by the low standard prediction error map and low RMSE. The approach lead to predictions (based on a single map) that are more realistic and simplified the numerical calculations. The predicted credibility grades indicate the relative importance of each value on the prediction map. Zero credibility grades indicated regions of low gamma dose pollution in the north, while pockets

of large credibility grade in the southern region identify high gamma dose concentrations. For a decision maker, generated single map is easier to interpret, compared to multiple maps generated via the alternative fuzzy methods. For example once can decide on an acceptable credibility threshold level to delineate gamma dose polluted region.

The fuzzy credibility approach defined offers additional advantages:

- it generalize indicator kriging; indicator kriging results (probability map) represent a membership function or credibility distribution of the probability of being less than a specific credibility threshold level.
- it generalizes the membership function, which is a merely an induced function and a conventional mathematical language for describing fuzzy phenomena, by its counterpart the credibility distribution function for continuous random fuzzy data.
- it includes the conventional kriging approach with crisp values supplemented by imprecise degenerate intervals.
- it reflects the real-world conditions more closely than traditional geostatistics, since it dose not impose exactness artificially on phenomena which are diffuse by their nature. It can be extended to mapping other pollutants.

8.2.3 Power of Variable Modeling of Spatial Random Intervals

We summarize the potential of the variable-based approaches:- bi-variable copula-based geostatistics and fuzzy credibility-based geostatistics in modeling of regionalized random intervals.

- Both approaches allows for two step prediction process: a modeling step and a prediction step. For credibility approach, we first estimated the data-assimilated credibility function leading to fuzzy credibility distribution prediction map. For copula method, marginal and copula are first estimated followed with prediction maps. This allows for goal-based predictions and thus better fit the requirements of the stakeholders.
- Since they are based on the joint distribution of random interval data, they have much to recommend over the alternatives like interval kriging and interval kernel regression which are limited to observed data. First, they are forward-looking, thus are capable of incorporating a wide range of future eventualities that simply are not captured using observed data.

Second, they do not require a large amount of data in order to be estimated accurately, and furthermore they are capable of reflecting a change in the observed phenomena. Sudden shift in observed phenomena and/or other related factors could be immediately captured in observed data and the implied distribution models. Despite the danger of over-inference associated with joint distributions, they can be applied to general case and still provide these advantages.

- The non-parametric kernel estimation of joint distribution is parameter-free which offers great generality allowing the methods to capture various expectations like multi-modality and asymmetries existing in the imprecise data.
- By splitting conditions up into a marginal part and a dependence part, the copula-based approach leads to better insight into the laws governing the spatial behavior of the random vector process. The approach also opens the possibility to build a variety of dependence structures based on existing parametric or non-parametric models of the marginal distributions and the copula.
- Contrary to conventional indicator geostatistics, the methods are not the result of discretization as is the case of indicator kriging. They can be described as probability or indicator kriging on regionalized interval data.
- Credibility fuzzy geostatistics provides a sound framework for modeling a spatial fuzzy variable, the equivalent of the spatial random interval sets. The approach avoid the general fuzzy set approach proposed by Zadeh, which leads to fuzzy sets or interval outputs based on alpha cut level.
- Credibility distributional can be used to model various properties of the random interval process such as the smoothness and impulsiveness which can be described using fuzzy sets. Because of its attractive properties, scalar-valued outputs based on fuzzy inputs, the fuzzy logic implemented via credibility measure theory has a tremendous potential in the geostatistics field.

8.3 Recommendations for Future Work

This work brings forth new perspective for modeling spatial random uncertain data. There are aspects of this work that could be further developed. Some relate to limitations of the methods themselves while others relate to how the methods were tested and evaluated. We highlight the limitations and some recommendations for future research.

- The techniques discussed in this thesis were applied to one dataset. Before we can generalize these results it should be applied to a wider scope of data (e.g. geology and social data).
- Composite and component-wise methods reduced interval-based methods to separate center-based and radius-based methods, easier to handle, and with same correlation conclusions to those based on interval-based methods. Further approaches to visualization of vector output should be explored.
- Application/ Concepts of IRF-k should be exploited as a future contribution to modeling of spatial random intervals.
- The spatially-weighted interval kernel could be extended to multivalued data with a general weighted feature kernel. Care should be taken in the choice of bandwidth for the different dimensions.
- The application of copula faces two limitations a) the existing analytical forms of copula are not appropriate for estimation and b) alternative approximation procedures are often not easy to implement. However, the empirical copula does capture key aspects of random vector dependence structure and hence the random interval/vector spatial structure.
- Still, it would be easier to deal with an analytical model of the copula. Application of two separate maximization of the univariate likelihood, followed by maximization of the bivariate likelihood is a possible approach. Several families of copula could also be used. This would increase the generality of copula toward modeling dependence structure of the bivariate correlated random variables that are often met in environmental research.
- We also note that interpreting information given by three maps (copula and marginals) might pose some difficulty.
- We have used copula as a model of random interval dependence structure. Future exploits should focus on theoretical developments of a combined spatial-feature copula that simultaneously model both the joint distribution and the spatial dependence structure.
- The copula approach constitute very promising avenues for future research in developing variable-based methods for modeling random fields with distinct non-Gaussian characteristics (asymmetry). Defined copula method in this study should be extended to other application for proof of consistency.

- Since the variable-based approaches (copula and credibility) reduces a random interval process into a marked point process, integrating statistical methods for copula logic, fuzzy credibility logic and the marked point process is an area of further research for modeling spatial interaction of spatially distributed uncertain (fuzzy).
- A future extension worthy of considering is the combination of fuzzy credibility logic, bi-variable copula and geostatistics. Such a combination would lead to CREDIBILITY COPULA-BASED GEOSTATISTICS. The credibility fuzzy logic models random fuzziness while the copula models the spatial dependence structure.
- The membership function is one among the functions associated with random interval set. Other related functions, namely the belief (bel) and the plausibility (pl) functions (Dempster, 1968; Joslyn and Booker, 2003; Smets, 2005) can also be obtained through numerical integration of the joint density of the random interval set. The belief function quantifies the minimal belief that the spatial random interval contains the true value while the plausibility function quantifies the maximal belief that the spatial random interval contains the true value. The two functions define a belief interval, $P_{[bel,pl]}$ that summaries all uncertainty associated with spatial random interval and can be subjected to geostatistical analysis, hence the BELIEF FUNCTIONS-BASED GEOSTATISTICS APPROACH.

In conclusion, a good case has been made that uncertainty (imprecision) in geographical data should be presented in some way. Given that one chooses to do so, modeling and spatially representing the data is an obvious choice. The development of models for spatial imprecise (fuzzy) data is a fast emerging field in environmental applications. Examples include the probabilistic soft (indicator) kriging and Bayesian kriging methods.

As a part of this emerging field, this work has provided unique contribution in the sense of composite kriging; component-wise kriging; spatially-weighted interval kernel regression; and ultimately, the "variable-based" methods: the bi-variable copula-based and the fuzzy credibility-based geostatistics. Ultimately, parallel to exact data modeling with a unique additive probability measure, we have been able to define and apply a unique additive measure, the credibility measure, to model random interval as scalar-valued fuzzy variable. The predictions are real-valued as

opposed to fuzzy sets in existing fuzzy set kriging approaches.

Therefore, variable-based techniques are key to future developments in modeling of environmental applications characterized by fuzzy spatial data, and proposed approaches in this work provide a sound foundation for achieving this. The concept of a FUZZY VARIABLE SIMULATION, may also be exploitated within the context of credibility distribution theory.

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover Publications, New York.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19:255 – 723.

Anselin, L., Florax, R., and Rey, S. J. (2004). *Advances in spatial econometrics : methodology, tools and applications.* Springer-Verlag, Germany.

Ash, R. B. (1972). *Real analysis and probability.* New york Academic press, New York.

Atkinson, P. M. (2004). Spatially weighted supervised classification for remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 5:277–291.

Atkinson, P. M. and Curran, P. (1998). Geostatistics and remote sensing. *Progress in Physical Geography*, 22:61–78.

Aumann, R. J. (1965). Integrals of set valued functions. *Journal of Mathematical Analysis and Applications*, 12:1–12.

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis. Essay.* Longman.

Bandemer, H. and Gebhardt, A. (2000). Bayesian fuzzy kriging. *Fuzzy Sets and Systems*, 112(3):405–418.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42:1– 12.

Bárdossy, A., Bogardi, I., and Kelly, W. E. (1988). Imprecise (fuzzy) information in geostatistics. *Mathematical Geology*, 20:287–311.

Bárdossy, A., Bogardi, I., and Kelly, W. E. (1990a). Kriging with imprecise (fuzzy) variograms. I: Theory. *Mathematical Geology*, 22:63–79.

Bárdossy, A., Bogardi, I., and Kelly, W. E. (1990b). Kriging with imprecise (fuzzy) variograms. II: Application. *Mathematical Geology*, 22:81–94.

Benedetti, J. K. (1977). On the nonparametric estimatioin of regression functions. *Journal of the Royal Statistical Society*, 39:248–253.

Bertoluzza, C. and Cariolaro, D. (1997). On the measure of a fuzzy set based on continuous t-conorms. *Fuzzy Sets and Systems*, 88:355–362.

Bowman, W. A. and Azzalini, A. (1997). *Applied smoothing Techniques for Data Analysis: The Kernel Approach with S-plus Illustrations*. Oxford Press, New York.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. (1998a). Geographically weighted regression- modelling spatial non-stationarity. *Journal of the Royal Statistical Society*, 47:431–443.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. (1998b). Spatial non-stationarity and autoregressive models. *Environment and Planning*, 30:957–973.

Burrough, P. A. (1989). Fuzzy mathematical models for soil survey and evaluation. *Journal of Soil Science*, 40:477–492.

Burrough, P. A. and McDonnell, R. A. (1998). *Principles of Geographical Information Systems*. Oxford, New York.

Burt, J. E. and Barber, G. M. (1996). *Elemantary Statistics for Geographers*. The Guilford Press, New York, 2nd edition.

Butnariu, D., Navara, M., and Vetterlein, T. (2004). Linear space of fuzzy vectors. Technical report, University of Haifa & University of Dortmund & Czech Technical University, Haifa Israel, Dortmund Germany, Czech Republic.

Cahn, M. D., Hummel, J. W., and Brouer, B. H. (1994). Spatial analysis of soil fertility for site-specific crop management. *Soil Science Society of America Journal*, 58:1240–1248.

Carrat, F. and Valleron, A. J. (1992). Epidemiological mapping using the "kriging" method: application to an influenza-like illness epidemic in france. *American Journal of Epidemiology*, 135(11):1293–1300.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799.

Cherubini, U., Luciano, E., and Vechiato, W. (2005). *Copula Methods in Finance*. John Wiley and Sons, New York.

Chiles, J.-P. and Delfiner, P. (1999). *Geostatistics Modelling Spatial Uncertainity*. John Wiley and Sons, New York.

Choi, E. and Hall, P. (1998). Nonparametric approach to analysis of space-time data on earthquake occurrences. *Journal of Computational and Graphical Statistics*, 8:733–748.

Christian, L. (2002). *Geostatistical Simulation.* Springer-Verlag, Berlin.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619.

Comaniciu, D., Ramesh, V., and Meer, P. (2000). Real-time tracking of non-rigid objects using mean shift. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 142–149.

Comaniciu, D., Ramesh, V., and Meer, P. (2001). The variable bandwidth mean shift and data-driven scale selection. In *Proceedings of 8th international conference on computer vision*, pages 438–445, Vancouver BC Canada.

Congar, R. (2000). Models of hard uncertainity for optimal control: The closed random interval approach. Technical report, EUREQUA-ERASME and Université de Rouen, ERASME-Laboratoire d'Economie, Central Paries, France.

Cressie, N. (1988). Variogram. In Kotz, S. and L, J. N., editors, *Entry in Encyclopedia of Statistical Sciences*, pages 489–491. Wiley, New York.

Cressie, N. A. C. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17:563–586.

Cressie, N. A. C. (1993). *Statistics for Spatial data.* John Wiley and Sons, New York, revised edition.

Cressie, N. A. C. and Hawkins, D. M. (1984). Robust kriginga proposal. *Mathematical Geology*, 16:3–18.

CTIF, IAEA, PAHO, and WHO (2006). Upgrading the rpii's national radiation monitoring network. Technical report, Incident and Emergency Centre International Atomic Energy Agency, Vienna, Austria.

Curran, P., Webster, R., and Munden, J. (1989). Spatial correlation in reflected radiation from the ground and its implications for sampling and mapping by ground-based radiometry. *Remote Sensing of Environment*, 29:67–78.

Darinka, D. (1998). Regular castaing representations of multifunctions with applications to stochastic programming. Technical report, Institut für Mathematik, Humboldt Universität Berlin, Berlin, Germany.

Darsow, W. F., Nguyen, B., and Olsen, E. T. (1992). Copulas and markov processes. *Illinois Journal of Mathematics*, 36:600–642.

De Cort, M. and de Vries, G. (1997). The european union radiological data exchange platform (eurdep): Two years of international data exchange experience. *Radiation Protection Dosimetry*, 73:17 – 20.

De Luca, A. and Termini, S. (1972). A definition of non-probabilistic entropy in the setting of fuzzy set theory. *Information and Control*, 20:301–312.

De Michele, C. and Salvadori, G. (2003). A generalized pareto intensity duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysic Research*, 108(D2), 4067, doi:10.1029/2002JD002534:4067.

Dempster, A. P. (1968). Upper and lower probabilities generated by a random closed interval. *The Annals of Mathematical Statistics*, 39(3):957–966.

Deutsch, C. V. (1996). Correcting for negative weights in ordinary kriging. *Computers and Geoscience*, 22:765–773.

Diamond, P. (1988). Interval-valued random functions and the kriging of intervals. *Mathematical Geology*, 20(3):145–165.

Diamond, P. and Kloeden, P. (1990). Metric spaces of fuzzy sets. *Fuzzy Sets and Systems*, 35:241–249.

Diamond, P. and Kloeden, P. (1994). *Metric spaces of fuzzy sets: theory and applications*. World Scientific, Singapore.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Applied Statistics*, 47(3):299–350.

Dubois, G. (2005). Automatic mapping algorithms for routine and emergency monitoring data spatial interpolation comparison 2004. Technical report, Radioactivity Environmental Monitoring, Emissions and Health Unit, Luxembourg.

Dubois, G. and Galmarini, S. (2005). Introduction to the spatial interpolation comparison (sic) 2004 exercise and presentation of the data sets. *Applied GIS*, 2:1 – 11.

Dubois, G. and Shibli, S. A. R. (2003). Monitoring of environmental radioactivity: automatic mapping or expert-dependent systems? In Dubois, G., Malczewski, J., and De Cort, M., editors, *Mapping radioactivity in the environment. Spatial Interpolation Comparison 1997*, pages 253 – 268. Joint Research Centre, European Commission, EUR 20667 EN, EC.

Dubrule, O. and Kostov, C. (1986). An interpolation method taking into account inequality constraints: I-methodology. *Mathematical Geology*, 18:33–51.

Duda, R. O., Stork, D. G., and Hart, P. E. (2001). *Pattern Classification*. John Wiley and Sons, New York.

D'Urso, P. and Gastaldi, T. (2000). A least-squares approach to fuzzy linear regression analysis. *Computational Statistics and Data Analysis*, 34:427–440.

Einbeck, J. and Tutz, G. (2004). Modelling beyond regression functions: an application of multimodal regression to speed-flow data. Technical report, Department of Mathematics, National University of Ireland, Galway, Ireland.

Elgammal, A., Harwood, D., and Davis, L. S. (2000). Non-parametric background model for background subtraction. In *Proceedings of the 6th European Conference of Computer Vision*, pages 751–767.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In M.A.H, editor, *Risk Management: Value at Risk and Beyond*, pages 176–223. University Press, Cambridge.

Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* John Wiley and Sons, USA.

Friel, N. P. (1999). A short talk on set-valued regression. Technical report, Department of Statistics, Uinversity of Glasgow, Scotland. UK.

Fuentes, M. (2001). A high frequency kriging approach for non-stationary environmental processes. *Environmetrics*, 12:469–483.

Gandin, L. S. (1965). *Objective Analysis of Meteorological Fields.* Gidrometeorologicheskoe Izdatel,stvo (GIMIZ) translated by Israel Program for scientific Translations, Jerusalem, Leningrad.

Gatrell, A., Bailey, T., Diggle, P., and Rowlingson, B. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Royal Geographical Society (with the Institute of British Geographers)*, pages 256–274.

Genton, M. G. (1998a). Highly robust variogram estimation. *Mathematical Geology*, 30:213–221.

Genton, M. G. (1998b). Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Mathematical Geology*, 30:323–345.

Gil, M., López-García, M., Lubiano, M., and Montenegro, M. (2001). Regression and correlation analysis of a linear relation between random intervals. *Test*, 10(1):183–201.

Gil, M., Lubiano, M., Montenegro, M., and López, M. (2002). Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika*, 56:97–111.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* Oxford University Press, New York.

Goovaerts, P. (1998). Ordinary cokriging revisited. *Mathematical Geology*, 30:21–42.

Guo, R. and Li, X. (2006). Credibility-hazard function: Theoretical development. In *Proceedings of ICME2006: International Conference on Maintenance Engineering (ICME)*, Cheng Du, China.

Handcock, M. S. and Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, 35:403–410.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Heilpern, S. (1990). Interval random sets and entropy. *Fuzzy Sets and Systems*, 35:213–217.

Helmut, A., Oswin, A., and Günter, R. (1994). Matching shapes with a reference point. In *Proceedings 7th Annual Symposium on Computational Geometry*.

Helmut, A., Oswin, A., and Günter, R. (1997). Matching shapes with a reference point. *International Journal of Computational Geometry and Applications*, 7:349 – 363.

Hengl, T., Geuvelink, G. B. M., and Stein, A. (2003). Comparison of kriging with external drift and regression kriging. Technical report, Department of Earth Systems Analysis, Enschede, Netherlands.

Hohn, M. E. (2000). Geostatistics and petroleum geology. (review). *Technometrics*, 42:444.

Hu, L. (2004). Dependence patterns across financial markets: a mixed copula approach. Technical report, Ohio State University, Ohio State, USA.

Hutchinson, M. F. (1995). Interpolation of mean precipitation using thin plate smoothing splines. *International Journal of Geographic Information Systems*, 9:385–403.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford University Press, New York.

Jennifer, A. H., Davis, R. A., and Merton, A. A. (2004). Model selection for geostatistical models. Technical report, Colorado State University, Pacific Northwest National Lab.

Joe, H. (1997). *Multivariate Dependence and Dependence Models*. Chapman and Hall, New York.

Joslyn, C. and Booker, J. (2003). Generalized information theory for engineering modeling and simulation. Technical report, Knowledge Systems and Computational Biology Team, Computer and Computational Sciences, Los Alamos, NM 87545, USA.

Journel, A. G. (1986). Constrained interpolation and qualitative information– the soft kriging approach. *Mathematical Geology*, 18(3):269–286.

Journel, A. G. (1994). Modeling uncertainty: Some conceptual thoughts. In Dimitrakopoulos, R., editor, *Geostatistics for the Next Century*, pages 30–34. Kluwer, Dordretch.

Journel, A. G. and Huijbregts (1978). *Mining Geostatistics*. Academic Press, New York.

Jowett, G. H. (1952). The accuracy of systematic sampling from conveyor belts. *Applied Statistics*, 1:50–59.

Käärik, E. (2006). Imputation algorithm using copulas. *Metodološki zvezki*, 3(1):109–120.

Kacewiez, M. (1994). Fuzzy geostatistics- an integration of qualitative description into spatial analysis. In Dimitrakopoulos, R., editor, *Geostatistics for the next Century*, pages 448–463. Kluwer, Dortdeecht.

Kaufmann, A. (1975). *Introduction to the Theory of Fuzzy Subsets*. Academic Press, New York.

Kearfott, R. B. (2004). Validated constraint solving: Practicalities, pitfalls, and new developments. Technical report, University of Louisiana at Lafayette, June, 2004 isiCAD.

Kern II, J. (2001). *Bayesian Process Convolution Approaches to Specifying Spatial Dependence Structure*. PhD thesis, Institute of Statistics and Decision Sciences, Duke University.

Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge University Press, New York.

Kleinschmidt, I., Bagayako, M., Clarke, G., Craig, M., and Le Sueur, D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, 29(2):355–361.

Klement, E. P., Puri, M. L., and Ralescu, D. A. (1986). Limit theorems for fuzzy random variables. *Proceedings of Royal Society of London. A*, 407:171–182.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin, English Translation(1950): Foundations of the theory of probability*. Chelsea, New York.

Kolmogorov, A. N. (1941). The local structure of turbulence in an incomprehensible vicious fluid at very large reynolds numbers. In *Doklady Akademii Nauk SSSR, Reprinted 1961, in Turbulence: Classical Papers on statistical theory, 151-155*, volume 30, pages 299– 303. Interscience Publishers, New York.

Körner, R. (1995). A variance of compact convex random sets. Technical report, Freiberg University of Mining and Technology, Freiberg, Germany.

Körner, R. and Näther, W. (2001). On the variance of random fuzzy variables. Technical report, Faculty of Mathematics and Computer Sciences, Freiberg University of Mining, Freiberg, Germany.

Kostov, C. and Dubrule, O. (1986). An interpolation method taking into account inequality constraints:ii- practical approach. *Mathematical Geology*, 18(1):53–73.

Krige, D. (1951). A statistical approach to some mine valuations and allied problems at the witwatersrand. Master's thesis, University of Witwatersrand.

Kulpa, Z. (2001). Diagrammatic representation for interval arithmetic. *Linear Algebra and its Applications*, 324:55–80.

Lakeyev, A. V. (1995). Linear algebraic equations in kaucher arithmetic: Extended abstracts. *A supplement to the international Journal of Reliable computing*, 1:130–133.

Leitner, M. and Staufer-Steinnocher, P. (2001). Kernel density estimation- an innovative exploratory tool for geomarketing. Technical report, Heidelberg, Heidelberg, Herbert Wichmann Verlag.

Lévy, P. (1948). *Processus Stochastiques et Mouvement Brownien. Suivi d'une note de M. Loéve. (French).* Gautier Villars, Paris.

Liu, B. (2004). *Uncertainity Theory: An Intoduction to Its Axiomatic Foundations.* Springer-Verlang, Heidelberg, Berlin.

Liu, B. (2006). A survey of credibility theory. *Fuzzy Optimization and Decision Making*, 5:4.

Liu, Y. K. and Liu, B. (2003). Expected value operator of random fuzzy variable and random fuzzy expected value models. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11:195–215.

Loader, C. (1999). *Local Regression and Likelihood.* Springer-Verlag, New York.

Lubiano, M. A. (1999). Generalised measure of the absolute variation of a random set, definition and statistical applications. In *Eleventh European Young Statisticians Meeting*, Spain.

Luenberger, D. G. (1987). *Optimization by Vector Space Methods.* John Wiley and Sons, New York.

Lyashenko, N. N. (1983). Statistics of random compact in euclidean space. *Journal of Soviet Mathematics*, 21:76–92.

Mark, F. and Tomasi, C. (2005). Meanshift is a bound optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):1–4.

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

Matheron, G. (1975). *Random Sets and Integral Geometry*. John Wiley and Sons, New York.

Matheron, G. (1989). *Estimating and Choosing*. Springer-Verlag, Berlin.

Maxwell, B. (2003). Choice with ambiguity as sets of probabilities. Technical report, Department of Economics, University of Texas at Austin, Austin, TX.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. Wiley, New York.

Meyer, W. (1970). Characterization of steiner point. *Pacific Journal of Mathematics*, 35(3):717–725.

Mikosch, T. (2003). Modeling dependence and tails of financial time series. In Finkenstä dt, B. and Rootzén, H., editors, *Extreme Values in Finance, Telecommunications, and the Environment*, pages 187–286. Chapman and Hall, Boca Raton.

Miranda, E., Couso, I., and Gil, P. (2005). Random intervals as a model for imprecise information. *Fuzzy Sets and Systems*, 154:386–412.

Moller, J. (2003). Aspects of spatial statistics, stochastic geometry and mcmc methods. Technical report, Department Statistics, Aalborg University, Germany.

Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for Spatial point process*. Chapman and Hall/CRC, Boca Raton.

Moore, R. (1979). *Methods and Applications of Interval Analysis*. SIAM Studies in Applied Mathematics, Philadelphia.

Mrazek, P., Weickert, J., and Bruhn, A. (2004). On robust estimation and smoothing with spatial kernels and toner kernels. Technical report, Mathematical Image analysis Group, Saarland University, Saarbrucken, Germany.

Myers, R. H. (1990). *Classical and Modern Regression with Applications*. PWS Kent Publishing, Boston.

Nader, E., Soofi, E. S., and Zahedi, H. (2004). Information properties of order statistics and spacings. *IEE Transactions on Information Theory*, 50(1):177–183.

Näther, W. (2000). On randon fuzzy variables of second order and their application to linear statistical inference with fuzzy data. *Metrika*, 51:201–221.

Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer-Verlag, New York.

Piotrowski, J. A., Bartels, F., Salski, A., and Schmidt, G. (1994). Fuzzy kriging of imprecise hydrogeological data. In *Proc. of IAMG'94 Conf*, Mont Tremblant, Quebec.

Ripley, B. D. (1987a). *Spatial Point Pattern Analysis in Ecology*. Springer-Verlag, Berlin.

Ripley, B. D. (1987b). *Stochastic Simulation*. Wiley, New York.

Robert, B. A. (2000). *Measure and integration theory*. New york Academic press, New York.

Rudas, I. J. and Kaynak, M. O. (1998). Minimum and maximum fuzziness generalized operators. *Fuzzy Sets and Systems*, 98:83 – 94.

Rupper, D. and Wand, M. P. (1994). Multivariate weighted least squares regression. *Annals of Statistics*, 22:1346–1370.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman and Hall, New York.

Schimek, M. G. (2000). *Smoothing and Regression*. John Wiley and Sons, Canada.

Schmitz, V. (2003a). Copulas and stochastic processes. Master's thesis, Institute of Statistics of Aachen University.

Schmitz, V. (2003b). *Copulas and Stochastic Processes*. Shaker-Verlag, Aachen.

Scholkopf, B., Tsuda, K., and Vert, J. (2004). A primer on kernel methods. Technical report, Max Plank Institute fur Biologische Kybernetik, Tubingen, Germany.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461 – 464.

Scott, D. (1992). *Multivariate Density Estimation, Theory, Practice and Visualisation*. John Wiley and Sons, New York.

Shary, S. P. (1997). Algebraic approach in the outer problem for interval linear equations. *Reliable Computing*, 3:103–135.

Shary, S. P. (2002). A new technique in system analysis under interval uncertainty and ambiguity. *Reliable Computing*, 8:321–418.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London.

Sklar, A. (1959). Functions de répartition á n dimensions et leurs marge. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231.

Smets, P. (2005). Belief functions on real numbers. *International Journal of Approximate Reasoning*, 40:181–223.

Somerville, S. and McMahon, C. (2006). Manual for first responders to a radiological emergency. Technical report, Radiological Protection Institute of Ireland, Ireland.

Stacey, K. F., Lark, R. M., Whitmore, A. P., and Milne, A. E. (2006). Using a process model and regression kriging to improve predictions of nitrous oxide emissions from soil. *Geoderma*, 135:107–117.

Stainslaw, H. (1990). Interval random sets and entropy. *Fuzzy Sets and Systems*, 35:213–217.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some theory for Kriging.* Springer-Verlag, New York.

Stoyan, D., Kendall, W. S., and Mecke, J. (1987). *Stochastic Geometry and its applications.* John Wiley and Sons, New York.

Stoyan, D. and Molchanov, I. S. (1997). Set-valued means of random particles. *Journal of Mathematical Imaging and Vision*, 7:111–121.

Szidarovszky, F., Baafi, E. Y., and Kim, Y. C. (1987). Kriging without negative weights. *Mathematical Geology*, 19:549–559.

Thiart, C. (2005). Lecture notes on spatial geostatistical modeling. Lecture series, Department of Statistical Sciences, University of Cape Town, South Africa.

Thomson, M. C., Connor, S. J., Alessandro, U. D., Rowlingson, B., Diggle, P., Cresswell, M., and Greenwood, B. M. (1998). A spatial model of malaria risk in the Gambia for predicting the impact of the insectside treated and untreated bednets on malaria infection. In *Presented at the 10th Colloquium of the spatial Information Research Centre*, University of Otago, New Zealand.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46.

Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory.* Springer-Verlag, New York.

Venables, W. N. and Ripley, B. D. (1994). *Modern applied Statistics with S-plus.* Springer- Verlag, New York.

Ver Hoef, J. M. V. and Cressie, N. (1993). Multivariate spatial prediction. *Mathematical Geology*, 25(2):219–240.

Vetterlein, T. and Navara, M. (2006). Defuzzification using steiner points. *Fuzzy Sets and Systems*, 35:241–249.

Vitale, R. A. (1988). An alternative formulation of mean value for random geometric figures. *Journal of Microscopy*, 151:197–204.

Vladik, K., Nguyen, H. T., Dimuro, G. P., and Bedregal, B. (2003). A new differential formalism for interval-valued functions and its potential use in detecting 1-d landscape features. Technical report, Lagoa Nova, Natal-RN-Brasil.

Wabiri, N., Guo, R., and Thiart, C. (2005). Interval kriging under generalized metric. In *Proceedings of International Workshop on Recent Advances Stochastic Operations Research*, Canmore, Alberta, Canada.

Wahba, G. (1992). *Spline Models for Observational Data*. Capital city Press, Montpeliar, Vermont, 2nd edition.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*. Chapman and Hall, London.

Webster, R. (1985). Quantitative spatial analysis of soil in the field. *Advances in Soil Science*, 3:1–70.

Webster, R. and Oliver, M. (2001). *Geostatistics for Environmental Scientist*. John Wiley and Sons, New York.

Woodcook, C. E., Strahler, A. H., and Jupp, D. L. B. (1988a). The use of variograms in remote sensing: I scene models and simulated images. *Remote Sensing of Environment*, 25:323–348.

Woodcook, C. E., Strahler, A. H., and Jupp, D. L. B. (1988b). The use of variograms in remote sensing: II real digital images. *Remote Sensing of Environment*, 25:349–379.

Xavier, E. (2003). Disjunctive kriging with hard and imprecise data. *Mathematical Geology*, 35(6):699–718.

Yaglom, A. M. (1987). Some classes of random fields in n-dimensional space, related to stationary random process. *Theory of Probability and Its Applications*, 2:273–320.

Ying, C. M. and Dean, C. B. (2001). Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, 57(3):949.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8:338–353.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28.

Zheng, P., Durr, P. A., and Diggle, P. J. (2004). Edge-correction for spatial kernel smoothing when is it necessary? In *Proceedings of the Second International Conference on the Applications of GIS and Spatial Analysis to Veterinary Science, GISVET04*, University of Guelph Ontario, Canada.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, 97:369–380.

APPENDIX A: COMPLETE METRIC AND LINEAR NORMED SPACE

A.1 Metric and Normed Spaces

Definition A.1.1 A **normed linear space** is a vector space X equipped with a mapping $||.|| : X \to \mathbb{R}_+$. $||\cdot||$ defines the norm of elements in the vector space. The minikowiski norms or the l_p is defined as:

$$l_p = ||x||_p = (\sum_i^n |x_i|^p)^{1/p} \qquad \text{for } 1 \leq p \leq \infty, i = 1, \ldots, n$$

Given two points $x, y \in \mathbb{R}^n$ the l_p distance between them is naturally given by $d(x,y) = ||x - y||_p$. If $p = 1$ the l_1 space is defined with "Manhattan metric or city-block metric"

$$d(x,y) \;=\; ||x - y||$$

for $a, b \in \mathbb{R}$

$$d(\boldsymbol{a},\boldsymbol{b}) \;=\; (|a^l - b^l| + |a^u - b^u|)$$

This is the distance viewed as the horizontal and vertical movements between two points.

If $p = 2$, the l_2 is the "Euclidean distance" defined as

$$d(x,y) \;=\; ||x - y||_2$$

In general for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^2$

$$d(\boldsymbol{a},\boldsymbol{b}) \;=\; \left((a^l - b^l)^2 + (a^u - b^u)^2\right)^{1/2}$$

Definition A.1.2 Cauchy Sequence: A sequence of elements (X_n) of a space X is said to be a d-**Cauchy sequence** if $\forall \varepsilon > 0\ \exists m \in n$, such that $d(X_n, X_m) < \varepsilon$, for $m > n$ where d is the distance metric in the space X.

The metric d is said to be **complete** if every d-Cauchy sequence converges i.e. there exists $X_0 \in X$ such that $\lim_n d(X_n, X_0) = 0$.

Definition A.1.3 A **complete metric space** (X, d) is a metric space X equipped with a complete metric d. In most cases the notation X is assumed for (X, d) without specifying the metric d. For a complete metric space X, every Cauchy sequence has a limit in X.

Definition A.1.4 A **Banach space** is a complete normed linear space, i.e. the Cauchy completion of a normed linear space. This forms a space of continuous functions.

Definition A.1.5 A **Hilbert space** $\mathcal{H}$ is a Banach space whose norm is of the form $||X|| = (\langle X, X \rangle)^{1/2}$ where $\langle \cdot, \cdot \rangle$ is the scalar product equivalent in real Euclidean space, i.e. $\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{R}$ such that

$$\begin{aligned} \langle x, y \rangle &= \langle y, x \rangle \text{ for all } x, y \in X \\ \langle x, y \rangle &\quad \text{is a linear functional on } X \text{ for every fixed } y \in X \\ \langle x, x \rangle &\quad \text{is positive for very } 0 \neq x \in X \end{aligned}$$

Remark A.1.1 Every inner product space is a normed linear space with associated norm $||x|| = \langle x, x \rangle^{1/2}$ induced by the inner product. The norm then induces a distance measure function $d(x, y) = ||x - y||$. Hence, the metric topology of complete linear normed space $d(x, y) = ||x - y||$, is similar to metric topology of a complete inner product space $d(x, y) = \langle x, y \rangle = |x - y|$, which draws the analogy between Hilbert spaces and Banach spaces.

The Hilbert space provides a system of coordinates that defines a state of the any system; it is a state space providing a set of all possible values that a system can take on.

Definition A.1.6 A **Topological space** $(X, \mathcal{X})$ is a space X together with its σ-algebra, $\mathcal{X}$, the set of all subsets of X inclusive of the $\emptyset$ and the set X. In most cases, it is understood that X is a topological space. Further, a metric space is automatically a topological space with the topology induced by the associated metric that defines the open and closed sets.

Definition A.1.7 Let $\mathbb{U}$ be a Euclidean space and a set $A \subseteq \mathbb{U}$. A is convex on $\mathbb{U}$ if and only if $\forall u_1, u_2 \in A \; \forall \kappa \in [0, 1]$, $\kappa u_1 + (1 - \kappa) u_2 \in A$. Set A is closed on $\mathbb{U}$ if

$\forall u_n \in A$, $n = 1, 2, \ldots$,

$$\lim_{n \to \infty} u_n = u \in A.$$

If any set A on $\mathbb{U}$ is both closed and convex, A is a closed convex set on $\mathbf{U}$

Theorem A.1.1 A bounded set A is a closed convex set on $\mathbb{R}$ if and only if A is a closed interval number on $CI(\mathbb{R})$.

APPENDIX B: REVIEW OF INTERVAL ARITHMETIC

Interval arithmetic defines algebraic properties for a set of closed real intervals numbers. The basic operations in interval arithmetic as defined by Moore (1979). Let's assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are closed real intervals which include any real numbers x and y in between, and also include the real endpoints a and b for $\boldsymbol{x}$ and c and d for $\boldsymbol{y}$, respectively,

$$\boldsymbol{x} = [a, b] = \{x : a \leq x \leq b\}, \tag{B.1}$$

$$\boldsymbol{y} = [c, d] = \{x : c \leq x \leq d\}. \tag{B.2}$$

Let $\mathcal{K}(\mathbb{R})$ denotes the set of all closed real intervals. The case $a \cong b$ results to $\boldsymbol{x} = [a, a]$, a degenerate or point interval. The operations for addition, subtraction, multiplication and division are

$$\boldsymbol{x} + \boldsymbol{y} = [a + c, b + d], \quad \boldsymbol{x} - \boldsymbol{y} = [a - d, b - c].$$

$$\boldsymbol{x} \times \boldsymbol{y} = \begin{cases} [ac, bd] & \text{if } a \geq 0 \text{ and } c \geq 0, \\ [bc, bd] & \text{if } a \geq 0 \text{ and } c < 0 < d, \\ [bc, ad] & \text{if } a \geq 0 \text{ and } d \leq 0, \\ [ad, bd] & \text{if } a < 0 < b \text{ and } c \geq 0, \\ [bd, ad] & \text{if } a < 0 < b \text{ and } d \leq 0, \\ [ad, bc] & \text{if } b \leq 0 \text{ and } c \geq 0, \\ [ad, ac] & \text{if } b \leq 0 \text{ and } c < 0 < d, \\ [bd, ac] & \text{if } b \leq 0 \text{ and } d \leq 0, \\ [\min\{bc, ad\}, \max\{ac, bd\}] & \text{if } a < 0 < b \text{ and } c < 0 < d \end{cases}$$

$$1/\boldsymbol{y} = [1/d, 1/c], \text{ for } 0 \notin \boldsymbol{y},$$

$$\boldsymbol{x}/\boldsymbol{y} = \boldsymbol{x} \times (1/\boldsymbol{y}), \text{ for } 0 \notin \boldsymbol{y},$$

The n-th power of $\boldsymbol{x}$ is defined as

$$\boldsymbol{x}^n = \begin{cases} [1,1] & \text{if } n = 0, \\ [a^n, b^n] & \text{if } a \geq 0\ , \\ & \text{or if } a < 0 < b \text{ and } n \text{ is odd}, \\ [b^n, a^n] & \text{if } b \geq 0\ , \\ [0, \max\{a^n, b^n\}] & \text{if } a < 0 < b \text{ and } n \text{ is even}, \end{cases} \tag{B.3}$$

For $\boldsymbol{x}/\boldsymbol{y}$ when $0 \in \boldsymbol{y}$, we apply Kaucher extended interval arithmetic which is isomorphic to properties of the $\mathbb{R}^2$ space (§3.3.4).

Commutative and associative algebraic laws, valid for real numbers remain valid for intervals. The distributivity law hold only in a weaker form, and is referred to as sub-distributivity law defined as

$$\boldsymbol{x}(\boldsymbol{y} + \boldsymbol{z}) \subseteq \boldsymbol{x}\boldsymbol{y} + \boldsymbol{x}\boldsymbol{z}. \tag{B.4}$$

For example, for $\boldsymbol{x} = [-2, 2]$; $\boldsymbol{y} = [1, 2]$; and $\boldsymbol{z} = [-2, 1]$,

$$\boldsymbol{x}(\boldsymbol{y} + \boldsymbol{z}) = [-2, 2]([1, 2] + [-2, 1]) = [-2, 2][-1, 3] = [-6, 6]$$

while

$$\boldsymbol{x}\boldsymbol{y} + \boldsymbol{x}\boldsymbol{z} = [-2, 2][1, 2] + [-2, 2][-2, 1]) = [-4, 4] + [-4, 4] = [-8, 8]$$

Clearly the interval number $\boldsymbol{x}(\boldsymbol{y} + \boldsymbol{z}) = [-6, 6]$ is contained in interval number $\boldsymbol{x}\boldsymbol{y} + \boldsymbol{x}\boldsymbol{z} = [-8, 8]$. For any intervals $\boldsymbol{x}$, subtraction contains zero, but is not equal to zero, $0 \in \boldsymbol{x} - \boldsymbol{x}$, while division $\boldsymbol{x}/\boldsymbol{x}$ contains 1, but is not equal to one. If $\boldsymbol{x} = [1, 3]$, then $0 \in \boldsymbol{x} - \boldsymbol{x} = [1, 3] - [1, 3] = [-2, 2]$, and $1 \in \boldsymbol{x}/\boldsymbol{x} = [1, 3]/[1, 3] = [1/3, 3]$.

B.1 Interval functions

An interval function is an interval-valued function, $\boldsymbol{F}$, of one or more interval arguments $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. The interval function $\boldsymbol{F}$ is said to be an *interval extension* of real-valued function f if $\boldsymbol{F}(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$ for all x_i $(i = 1, \ldots, n)$. This implies that if the interval arguments $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are degenerate intervals (i.e. $\boldsymbol{x}_i = x_i$), then $F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is a degenerate interval equal to $f(x_1, \ldots, x_n)$. An interval function is said to be *inclusion monotonic* if $\boldsymbol{x}_i \subset \boldsymbol{y}_i (i = 1, \ldots, n)$, implies $\boldsymbol{F}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \subset \boldsymbol{F}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$.

All basic interval arithmetic operations are inclusion monotonic. Let "$\diamond$" denote the set of operators $\{+, -, \times, /\}$, then $\boldsymbol{x}_i \subset \boldsymbol{y}_i (i = 1, 2)$ implies $(\boldsymbol{x}_1 \diamond \boldsymbol{x}_2) \subset (\boldsymbol{y}_1 \diamond \boldsymbol{y}_2)$. This leads to the fundamental theorem in interval analysis

Theorem B.1.1 *Let $\boldsymbol{F}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be an inclusion monotonic interval extension of real function $f(x_1, \ldots, x_n)$. Then $F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ contains the range of values of $f(x_1, \ldots, x_n)$ for all $x_i \in \boldsymbol{x}_i (i = 1, \ldots, n)$.*

As an illustration of the theorem, consider a real function

$$f(x) = \frac{x}{1-x}, \qquad x \neq 1.$$

For $a = 2, b = 3$, interval $\boldsymbol{x} = [2, 3]$, the range, $R(f; \boldsymbol{x})$, of f over $\boldsymbol{x}$ is

$$R(f; \boldsymbol{x}) = [\frac{2}{1-2}, \frac{3}{1-3}] = [-2, -3/2]$$

while the inclusion monotonic interval extension function

$$F(\boldsymbol{x}) = \frac{[2, 3]}{[1, 1] - [2, 3]} = [-3, -1]$$

Clearly $[-2, -3/2] \subseteq [-3, -1]$ implying that $R(f : \boldsymbol{x}) \subseteq F(\boldsymbol{x})$. The fundamental concept of the inclusion function provides a useful strategy to test intervals for inclusion in interval optimization algorithms.

B.2 Interval Vectors and Matrices

An interval vector is a vector whose components are interval numbers. For example, an n-dimensional interval vector, $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, where $\boldsymbol{x}_j$ are interval values in $\mathbb{I}(\mathbb{R})$. The mid-point, norm, and width values are defined component wise

$$\begin{aligned} \text{mid}\ (\underline{\boldsymbol{x}}) &\triangleq (\text{mid}\ (\boldsymbol{x}_1), \text{mid}\ (\boldsymbol{x}_2), \ldots, \text{mid}\ (\boldsymbol{x}_n)), \\ \text{norm}\ \|\underline{\boldsymbol{x}}\| &\triangleq \max_{k=1,\cdots,n}(|\boldsymbol{x}_k|) \\ \text{wid}\ (\underline{\boldsymbol{x}}) &\triangleq \max_{k=1,\cdots,n}(\text{wid}\ (\boldsymbol{x}_k)) \end{aligned}$$

An interval matrix A is a matrix whose elements are interval numbers i.e.

$\boldsymbol{A} = (\boldsymbol{a}_{ij})_{n\times n}$, with intervals $\boldsymbol{a}_{ij} \in \mathbb{I}(\mathbb{R})$ and

$$\begin{aligned} \text{mid}\ (\boldsymbol{A}) &\triangleq (\text{mid}\ (\boldsymbol{a}_{ij}))_{n\times n}, \\ \text{norm}\ \|\boldsymbol{A}\| &\triangleq \max_{i=1,\cdots,n}\left(\sum_{j=1}^{n}|\boldsymbol{a}_{ij}|\right) \\ \text{wid}\ (\boldsymbol{A}) &\triangleq \max_{i,j=1,\cdots,n}\text{wid}\ (\boldsymbol{a}_{ij}) \end{aligned}$$

For example, 2×2 interval matrix

$$\begin{aligned} A &= \begin{bmatrix} \begin{bmatrix} 2, & 3 \end{bmatrix} & \begin{bmatrix} 1, & 5 \end{bmatrix} \\ \begin{bmatrix} 1, & -1 \end{bmatrix} & \begin{bmatrix} 4, & 6 \end{bmatrix} \end{bmatrix} \\ \text{mid}\ (\boldsymbol{A}) &\triangleq \begin{pmatrix} \text{mid}\begin{bmatrix} 2, & 3 \end{bmatrix} & \text{mid}\begin{bmatrix} 1, & 5 \end{bmatrix} \\ \text{mid}\begin{bmatrix} 1, & -1 \end{bmatrix} & \text{mid}\begin{bmatrix} 4, & 6 \end{bmatrix} \end{pmatrix} \triangleq \begin{bmatrix} 2.5 & 3 \\ 0 & 5 \end{bmatrix} \end{aligned}$$

For detailed algebraic properties of interval matrices see Neumann,(1990).

B.3 Derivatives of interval-valued function

Definition B.3.1 *Assume that* $f^l(t) : \mathbb{T} \to \mathbb{R}$ *and* $f^u(t) : \mathbb{T} \to \mathbb{R}$ *and*

$$f(t) = [f^l(t), f^u(t)]$$

is an interval function on $\mathbb{T} \subseteq \mathbb{R}$ *. If these real-valued functions* $f^l(t)$ *and* $f^u(t)$ *are differentiable at* $t_0 \in T$*, i.e.,* $\frac{d}{dt}f^l(t_0)$ *and* $\frac{d}{dt}f^u(t_0)$ *exist, interval function* $f(t)$ *is differentiable at* $t_0 \in T$*. The derivative of the interval function* $f(t)$ *at* t_0 *is defined as*

$$\frac{d}{dt}f(t_0) = \left[\frac{d}{dt}f^l(t_0) \wedge \frac{d}{dt}f^u(t_0), \frac{d}{dt}f^l(t_0) \vee \frac{d}{dt}f^u(t_0)\right].$$

If $\frac{d}{dt}f^l(t_0) \leq \frac{d}{dt}f^u(t_0)$ *then* $f(t)$ *is said to be differentiable at* t_0 *same-orderly, and*

$$\frac{d}{dt}f(t_0) = \left[\frac{d}{dt}f^l(t_0), \frac{d}{dt}f^u(t_0)\right],\ t_0 \in T;$$

If $\frac{d}{dt}f^l(t_0) \geq \frac{d}{dt}f^u(t_0)$ *then* $f(t)$ *is said to be differentiable at* t_0 *reverse-orderly, and*

$$\frac{d}{dt}f(t_0) = \left[\frac{d}{dt}f^u(t_0), \frac{d}{dt}f^l(t_0)\right], t_0 \in \mathbb{T}$$

Definition B.3.2 *For a general interval function* $f(t) = \left[f^l(t), f^u(t)\right]$ *on* $\mathbb{T}$*, if the derivative of* $f(t)$*,*

$$\frac{d}{dt}f(t) = \left[\frac{d}{dt}f^l(t) \wedge \frac{d}{dt}f^u(t), \frac{d}{dt}f^l(t) \vee \frac{d}{dt}f^u(t)\right]$$

exists for $\forall t \in \mathbb{T}$*, then* $f(t)$ *is differentiable on* $\mathbb{T}$*. Similarly,* $f(t)$ *is differentiable same-orderly on* $\mathbb{T}$ *if*

$$\frac{d}{dt}f(t) = \left[\frac{d}{dt}f^l(t), \frac{d}{dt}f^u(t)\right], \forall t \in \mathbb{T}$$

and $f(t)$ *is differentiable reverse-orderly on* $\mathbb{T}$ *if*

$$\frac{d}{dt}f(t) = \left[\frac{d}{dt}f^u(t), \frac{d}{dt}f^l(t)\right], \forall t \in \mathbb{T}.$$

Further question:- given a class of interval functions $F(\boldsymbol{x})$, where $\boldsymbol{x} \triangleq [x^l, x^u]$, $x^l < x^u$, $x^l, x^u \in \mathbb{R}$ i.e., a real-valued interval, what is the definition of the derivative of $F(\boldsymbol{x})$, $F'(\boldsymbol{x})$? For example, what is the derivative of say $F(\boldsymbol{x}) = [1,2] \cdot \boldsymbol{x}^2$. Kearfott (2004) states that, if the interval $[1,2]$ is "interpreted as a single point value that is known only to lie in the interval $[1,2]$, then the logical derivative would be $2 \cdot [1,2] \cdot \boldsymbol{x}$. This is because $f(x) = ax^2$ for $a \in [1,2]$, so $f'(x) = 2ax$ for the *same* a in $[1,2]$, this implies the definition of an 'interval derivative' or 'interval Jacobin matrix'. Vladik et al. (2003) has given the following new differential formalism for inter-valued functions:

Definition B.3.3 *An interval function* F *is a finite sequence of pairs* $(x_i, \boldsymbol{y}_i)$*,* $i = 1, 2, \cdots, n$*, where for each* i*,* x_i *is a real number, i.e.,* $x_i \in \mathbb{R}$*,* $x_1 < x_2 < \cdots < x_n$*, and* $\boldsymbol{y}_i = [y_i^l, y_i^u]$ *is a non-degenerate interval* ($y_i^l < y_i^u$*,* $y_i^l, y_i^u \in \mathbb{R}$) *enclosing the range of* $\boldsymbol{x}$*.*

Definition B.3.4 *A function* $f : \mathbb{R} \to \mathbb{R}$ *is said to belong to an interval function* $F = \{(x_1, \boldsymbol{y}_1), (x_2, \boldsymbol{y}_2), \cdots, (x_n, \boldsymbol{y}_n)\}$*, i.e.,* $f \in F$*, if* f *is continuously differentiable and for each* i*,* $f(x_i) \in \boldsymbol{y}_i$*.*

Definition B.3.5 *Let* F *be an interval function and* $[a, b]$ *be an interval. A derivative of* F*,*

$$F'([a,b]) \triangleq \bigcap_{f \in F} f'([a,b])$$

where $f'(x)$ is the first order derivative of $f(x)$, and

$$f'([a,b]) \triangleq \{f'(x)|x \in [a,b]\},$$

where $f'([a,b])$ is the range *of the derivative $f'(x)$ over the interval $[a,b]$.*

Remark B.3.1 *The notation $F'([a,b])$ looks like the notation of a range for a real-valued functions and shares some properties of range, say, the range is inclusion-monotonic by which $[a,b] \subseteq [c,d] \Rightarrow f'([a,b]) \subseteq f'([c,d])$. Therefore, derivative $F'([a,b])$ is inclusion-monotonic, i.e., $[a,b] \subseteq [c,d] \Rightarrow F'([a,b]) \subseteq F'([c,d])$. Furthermore, if $[a,b] \cap [c,d] \neq \varnothing$, $[a,b] \cup [c,d] = [a,d]$, an interval, then $F'([a,b] \cup [c,d]) \supseteq F'([a,b]) \cup F'([c,d])$. But it is not a range, if an interval is narrow enough, $F'([a,b]) = \varnothing$.*

Definition B.3.6 *Let $F = \{(x_1, \boldsymbol{y}_1), (x_2, \boldsymbol{y}_2), \cdots, (x_n, \boldsymbol{y}_n)\}$ be an interval function, and let s be a real-valued number. Define an interval function*

$$F - sx \triangleq \{(x_1, \boldsymbol{y}_1 - sx_1), (x_2, \boldsymbol{y}_2 - sx_2), \cdots, (x_n, \boldsymbol{y}_n - sx_n)\}$$

where

$$\boldsymbol{y} - c \triangleq [y^l - c, y^u - c]$$

for the given interval $\boldsymbol{y} = [y^l, y^u]$, $c \in \mathbb{R}$.

Lemma B.3.1 *For $\forall F$ and any interval $[a,b]$, $s \in F'([a,b])$ if and only if (Vladik et al., 2003)*

$$0 \in (F - sx)'([a,b])$$

Theorem B.3.1 *For $\forall F$ and any interval $[a,b]$, let i_l and j_u be the first and last index of the values inside $[a,b]$ (Vladik et al., 2003). Then*

$$F'([a,b]) = [(F')^l, (F')^u]$$

where

$$(F')^l \triangleq \min_{i_l \le i \le j \le j_u} \Delta_{ij}^u$$

$$(F')^u \triangleq \max_{i_l \le i \le j \le j_u} \Delta_{ij}^l$$

$$\Delta_{ij}^l \triangleq \frac{y_i^l - y_j^u}{x_j - x_i}$$

$$\Delta_{ij}^u \triangleq \frac{y_i^u - y_j^l}{x_j - x_i}$$

and

$$[p, q] \triangleq \{x | p \le x \ \mathcal{E} x \le q\}$$

such that if $p > q$, $[p, q] = \varnothing$.

B.4 Interval Matrix operations

By convention an interval matrix $A = (A_{i,j})$ is interpreted as a set of real $m \times n$ matrices (Neumann 1990)

$$A = \{\tilde{A} \in \mathbb{R}^{m \times n} | \tilde{A}_{i,j} \in A_{ij} \quad \text{for } i = 1, \ldots, m; j = 1, \ldots, n\}$$

In short an interval matrix is the one that contains all real matrices, whose elements are obtained from all possible values between the lower and upper bound of its interval elements. A symmetrical interval matrix is one that contains only those real symmetrical matrices whose elements are obtained form all possible values between the lower and upper bound of its interval element i.e.,

$$A_{sym} = \{\tilde{A} \in \mathbb{R}^{n \times n} | \tilde{A}_{i,j} \in A_{ij} \quad \text{with} \tilde{A}^T = \tilde{A}\}$$

Simply, an interval vector is an $n \times 1$ interval matrix. Hansen (1992) refers to interval vector as a box.

Let m and n denote given positive integers and let $\mathbb{I}^*(\mathbb{R}^n)$ and $\mathbb{I}^*(\mathbb{R}^{m \times n})$ denote the sets of all n-dimensional vectors and all $m \times n$ matrices respectively with elements from $\mathbb{I}^*(\mathbb{R})$. Let interval matrix $\boldsymbol{A} \triangleq [\boldsymbol{A}^l, \boldsymbol{A}^u] \in \mathbb{I}^*(\mathbb{R}^{m \times n})$ and interval vector $\boldsymbol{b} \triangleq [\boldsymbol{b}^l, \boldsymbol{b}^u] \in \mathbb{I}^*(\mathbb{R}^n)$.

Definition B.4.1 *For* $\forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{I}^*(\mathbb{R}^{m \times n})$, *then*

(i) $\boldsymbol{A} \triangleq (\boldsymbol{a}_{ij})_{m \times n} = \left([a_{ij}^l, a_{ij}^u]\right)_{m \times n}$, $\boldsymbol{a}_{ij} \in \mathbb{I}^*(\mathbb{R})$;

(ii) $\boldsymbol{A} = \boldsymbol{B}$ *if and only if* $\boldsymbol{a}_{ij} = \boldsymbol{b}_{ij}$, *for* $\forall i, j$;

(iii) $\boldsymbol{A} \subset \boldsymbol{B}$ *if and only if* $\boldsymbol{a}_{ij} \subset \boldsymbol{b}_{ij}$, *for* $\forall i, j$;

(iv) $\boldsymbol{A} \pm \boldsymbol{B} \triangleq (\boldsymbol{a}_{ij} \pm \boldsymbol{b}_{ij})$ *(see theorem B.4.1 (i));*

(v) *for* $\forall \boldsymbol{A} \in \mathbb{I}^*(\mathbb{R}^{m \times v})$ *and* $\forall \boldsymbol{A} \in \mathbb{I}^*(\mathbb{R}^{v \times n})$, $\boldsymbol{A} \cdot \boldsymbol{B} \triangleq \left(\sum_{k=1}^{v} \boldsymbol{a}_{iv} \cdot \boldsymbol{b}_{vj} \right)$;

(vi) *for* $\forall \boldsymbol{A} \in \mathbb{I}^*(\mathbb{R}^{m \times n})$ *and* $\forall \boldsymbol{b} \in \mathbb{I}^*(\mathbb{R}^{n})$, $\boldsymbol{A} \cdot \boldsymbol{b} \triangleq \left(\sum_{k=1}^{v} \boldsymbol{a}_{iv} \cdot \boldsymbol{b}_{v} \right)$;

(vii) *for* $\forall \boldsymbol{A} \in \mathbb{I}^*(\mathbb{R}^{m \times n})$ *and* $\forall \mathbf{c} \in \mathbb{I}^*(\mathbb{R})$, $\boldsymbol{A} \cdot \mathbf{c} = \mathbf{c} \cdot \boldsymbol{A} \triangleq (\mathbf{c} \cdot \boldsymbol{a}_{ij})$;

(viii) $\boldsymbol{I} \triangleq \left(\overline{1}\right)_{m \times n}$, *and* $\mathbf{0} \triangleq \left(\overline{0}\right)_{m \times n}$, *where* $\overline{1} = [1, 1]$ *and* $\overline{0} = [0, 0]$ *respectively;*

(ix) *for* $\forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{I}^*(\mathbb{R}^{m \times n})$, $\boldsymbol{A}^+ \triangleq \left((\overline{a}_{ij})^+\right)$, $\boldsymbol{A}^- \triangleq \left((\overline{a}_{ij})^-\right)$, *and* $\boldsymbol{A} \vee \boldsymbol{B} \triangleq (a_{ij} \vee b_{ij})$

(x) *for* $\forall \mathbf{b} \in \mathbb{I}^*(\mathbb{R}^{n})$, $n \times n$ *matrix* $\operatorname{diag}(\boldsymbol{B})$

$$\operatorname{diag}(\boldsymbol{B}) \triangleq \begin{bmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{bmatrix}.$$

For common real-valued matrices denoted as

$$A = \begin{bmatrix} 4, & -9 \\ 0, & 6 \end{bmatrix} \quad B = \begin{bmatrix} 0, & 2 \\ -1, & -1 \end{bmatrix} \quad C = \begin{bmatrix} 5, & -23 \\ -5, & 1 \end{bmatrix}$$

it is well known that $\{A + B : A \in \boldsymbol{A}, B \in \boldsymbol{B}\} \subseteq \{C \in AB\}$.

Theorem B.4.1 *Let*

$$\boldsymbol{A} = \left(\begin{bmatrix} 4, & 6 \\ 0, & 1 \end{bmatrix} \quad \begin{bmatrix} -9, & 0 \\ 6, & 10 \end{bmatrix} \right)$$

$$\boldsymbol{B} = \left(\begin{bmatrix} 0, & 12 \\ -1, & 1 \end{bmatrix} \quad \begin{bmatrix} 2, & 3 \\ -1, & 3 \end{bmatrix} \right)$$

$$\boldsymbol{C} = \left(\begin{bmatrix} 5, & 9 \\ -5, & 1 \end{bmatrix} \quad \begin{bmatrix} -23, & -9 \\ 1, & 15 \end{bmatrix} \right)$$

be interval matrices. Then

(i) $\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$

Proof.

From the left side we have

$$\begin{aligned}
\boldsymbol{A}+\boldsymbol{B} &= \left(\begin{bmatrix} 4, & 6 \\ 0, & 1 \end{bmatrix} \quad \begin{bmatrix} -9, & 0 \\ 6, & 10 \end{bmatrix}\right)+\left(\begin{bmatrix} 0, & 12 \\ -1, & 1 \end{bmatrix} \quad \begin{bmatrix} 2, & 3 \\ -1, & 3 \end{bmatrix}\right) \\
&= \left(\begin{bmatrix} 4, & 6 \end{bmatrix}+\begin{bmatrix} 0, & 12 \end{bmatrix} \quad \begin{bmatrix} -9, & 0 \end{bmatrix}+\begin{bmatrix} 2, & 3 \end{bmatrix} \right) \\
&= \left(\begin{bmatrix} 4, & 18 \\ -1, & 2 \end{bmatrix} \quad \begin{bmatrix} -7, & 3 \\ 5, & 13 \end{bmatrix}\right) \qquad \text{(B.5)}
\end{aligned}$$

From the right side we have

$$\begin{aligned}
\boldsymbol{B}+\boldsymbol{A} &= \left(\begin{bmatrix} 0, & 12 \\ -1, & 1 \end{bmatrix} \quad \begin{bmatrix} 2, & 3 \\ -1, & 3 \end{bmatrix}\right)+\left(\begin{bmatrix} 4, & 6 \\ 0, & 1 \end{bmatrix} \quad \begin{bmatrix} -9, & 0 \\ 6, & 10 \end{bmatrix}\right) \\
&= \left(\begin{bmatrix} 0, & 12 \end{bmatrix}+\begin{bmatrix} 4, & 6 \end{bmatrix} \quad \begin{bmatrix} 2, & 3 \end{bmatrix}+\begin{bmatrix} -9, & 0 \end{bmatrix} \right) \\
&= \left(\begin{bmatrix} 4, & 18 \\ -1, & 2 \end{bmatrix} \quad \begin{bmatrix} -7, & 3 \\ 5, & 13 \end{bmatrix}\right) \qquad \text{(B.6)}
\end{aligned}$$

From B.6 and B.5 we conclude that $\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{B}+\boldsymbol{A}$. *The rest can be similarly shown.* ■

(ii) $(\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}=\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C})$ *if* $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}\in\mathbb{I}^*(\mathbb{R}^{v\times n})$;

(iii) $\boldsymbol{A}+\boldsymbol{0}=\boldsymbol{0}+\boldsymbol{A}=\boldsymbol{A}$ *if* $\boldsymbol{A},\boldsymbol{0}\in\mathbb{I}^*(\mathbb{R}^{v\times n})$;

(iv) $\boldsymbol{AI}=\boldsymbol{IA}=\boldsymbol{A}$ *if* $\boldsymbol{A},\boldsymbol{I}\in\mathbb{I}^*(\mathbb{R}^{n\times n})$;

(v) *subdistributivity*

$$\begin{aligned}
(\boldsymbol{A}+\boldsymbol{B})\,\boldsymbol{C} &\subseteq \boldsymbol{AC}+\boldsymbol{BC} \\
\boldsymbol{C}\,(\boldsymbol{A}+\boldsymbol{B}) &\subseteq \boldsymbol{CA}+\boldsymbol{CB}
\end{aligned}$$

To illustrate, consider the left hand side

$$
\begin{aligned}
&(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C}\\
&=\left(\begin{bmatrix}4 & 18\\ -1 & 2\end{bmatrix}\ \begin{bmatrix}-7 & 3\\ 5 & 13\end{bmatrix}\right)*\left(\begin{bmatrix}5 & 9\\ -5 & 1\end{bmatrix}\ \begin{bmatrix}-23 & -9\\ 1 & 15\end{bmatrix}\right)\\
&=\left(\begin{matrix}\left(\begin{bmatrix}4 & 18\\ -1 & 2\end{bmatrix}*\begin{bmatrix}5 & 9\\ -5 & 1\end{bmatrix}+\begin{bmatrix}-7 & 3\\ 5 & 13\end{bmatrix}*\begin{bmatrix}-5 & 1\\ -5 & 1\end{bmatrix}\right) & \left(\begin{bmatrix}4 & 18\\ -1 & 2\end{bmatrix}*\begin{bmatrix}-23 & -9\\ -23 & -9\end{bmatrix}+\begin{bmatrix}-7 & 3\\ 5 & 13\end{bmatrix}*\begin{bmatrix}1 & 15\\ 1 & 15\end{bmatrix}\right)\\ \left(\begin{bmatrix}-1 & 2\end{bmatrix}*\begin{bmatrix}5 & 9\end{bmatrix}+\begin{bmatrix}5 & 13\end{bmatrix}*\begin{bmatrix}-5 & 1\end{bmatrix}\right) & \left(\begin{bmatrix}-1 & 2\end{bmatrix}*\begin{bmatrix}-23 & -9\end{bmatrix}+\begin{bmatrix}5 & 13\end{bmatrix}*\begin{bmatrix}1 & 15\end{bmatrix}\right)\end{matrix}\right),\\
&=\left(\begin{matrix}\begin{bmatrix}5 & 197\end{bmatrix} & \begin{bmatrix}-519 & 9\end{bmatrix}\\ \begin{bmatrix}-74 & 31\end{bmatrix} & \begin{bmatrix}-41 & 218\end{bmatrix}\end{matrix}\right)
\end{aligned}
$$

Similarly the right hand side is

$$
\begin{aligned}
&\boldsymbol{AC}+\boldsymbol{BC}\\
&=\left(\begin{bmatrix}4 & 6\\ 0 & 1\end{bmatrix}\ \begin{bmatrix}-9 & 0\\ 6 & 10\end{bmatrix}\right)*\left(\begin{bmatrix}5 & 9\\ -5 & 1\end{bmatrix}\ \begin{bmatrix}-23 & -9\\ 1 & 15\end{bmatrix}\right)\\
&+\left(\begin{bmatrix}0 & 12\\ -1 & 1\end{bmatrix}\ \begin{bmatrix}2 & 3\\ -1 & 3\end{bmatrix}\right)*\left(\begin{bmatrix}5 & 9\\ -5 & 1\end{bmatrix}\ \begin{bmatrix}-23 & -9\\ 1 & 15\end{bmatrix}\right)\\
&=\left(\begin{matrix}\left(\begin{bmatrix}4 & 6\end{bmatrix}*\begin{bmatrix}5 & 9\end{bmatrix}+\begin{bmatrix}-9 & 0\end{bmatrix}*\begin{bmatrix}-5 & 1\end{bmatrix}\right) & \left(\begin{bmatrix}4 & 6\end{bmatrix}*\begin{bmatrix}-23 & -9\end{bmatrix}+\begin{bmatrix}-9 & 0\end{bmatrix}*\begin{bmatrix}1 & 15\end{bmatrix}\right)\\ \left(\begin{bmatrix}0 & 1\end{bmatrix}*\begin{bmatrix}5 & 9\end{bmatrix}+\begin{bmatrix}6 & 10\end{bmatrix}*\begin{bmatrix}-5 & 1\end{bmatrix}\right) & \left(\begin{bmatrix}0 & 1\end{bmatrix}*\begin{bmatrix}-23 & -9\end{bmatrix}+\begin{bmatrix}6 & 10\end{bmatrix}*\begin{bmatrix}1 & 15\end{bmatrix}\right)\end{matrix}\right)\\
&+\left(\begin{matrix}\left(\begin{bmatrix}0 & 12\end{bmatrix}*\begin{bmatrix}5 & 9\end{bmatrix}+\begin{bmatrix}2 & 3\end{bmatrix}*\begin{bmatrix}-5 & 1\end{bmatrix}\right) & \left(\begin{bmatrix}0 & 12\end{bmatrix}*\begin{bmatrix}-23 & -9\end{bmatrix}+\begin{bmatrix}-2 & 3\end{bmatrix}*\begin{bmatrix}1 & 15\end{bmatrix}\right)\\ \left(\begin{bmatrix}-1 & 1\end{bmatrix}*\begin{bmatrix}5 & 9\end{bmatrix}+[1,3]*\begin{bmatrix}-5 & 1\end{bmatrix}\right) & \left(\begin{bmatrix}-1 & 1\end{bmatrix}*\begin{bmatrix}-23 & -9\end{bmatrix}+\begin{bmatrix}-1 & 3\end{bmatrix}*\begin{bmatrix}1 & 15\end{bmatrix}\right)\end{matrix}\right)\\
&=\left(\begin{matrix}\begin{bmatrix}6 & 210\end{bmatrix} & \begin{bmatrix}-573 & 9\end{bmatrix}\\ \begin{bmatrix}-74 & 31\end{bmatrix} & \begin{bmatrix}-55 & 218\end{bmatrix}\end{matrix}\right)
\end{aligned}
$$

Illustrating that $(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} \subseteq \boldsymbol{AC}+\boldsymbol{BC}$. *Similarly* $\boldsymbol{C}(\boldsymbol{A}+\boldsymbol{B}) \subseteq \boldsymbol{CA}+\boldsymbol{CB}$ *can be shown*

(vi) *Equality is satisfied with multiplication by a real-valued matrix* C *i.e.*

$$
\begin{aligned}
(\boldsymbol{A}+\boldsymbol{B})\,C &= \boldsymbol{A}C+\boldsymbol{B}C\\
C\,(\boldsymbol{A}+\boldsymbol{B}) &= C\boldsymbol{A}+C\boldsymbol{B}
\end{aligned}
$$

(vii)

$$
\begin{aligned}
&\boldsymbol{A}(BC) \subseteq (\boldsymbol{AB})\,C\\
&(\boldsymbol{AB})\,\boldsymbol{C} \subseteq A(\boldsymbol{BC}) \quad \textit{if } \boldsymbol{C}=-\boldsymbol{C}\\
&A(\boldsymbol{B}C)=(\boldsymbol{AB})\,C\\
&\boldsymbol{A}(\boldsymbol{BC})=(\boldsymbol{AB})\,\boldsymbol{C} \quad \textit{for } \boldsymbol{A},\boldsymbol{B},\boldsymbol{C} \in \mathbb{I}(\mathbb{R}^{m\times n}),\\
&\qquad\qquad\qquad\qquad \textit{and } \boldsymbol{B}=-\boldsymbol{B},\boldsymbol{C}=-\boldsymbol{C}
\end{aligned}
$$

B.5 Basic operations of Kaucher arithmetic

Assume operator symbol $*$ denotes any operation from the operator set $\{+,-,\cdot\}$. With intervals $\boldsymbol{x}=[x^l,x^u]\in\mathbb{I}^*(\mathbb{R})$, $\boldsymbol{y}=[y^l,y^u]\in\mathbb{I}^*(\mathbb{R})$, $\boldsymbol{t}=[t^l,t^u]\in\mathbb{I}^*(\mathbb{R})$, and a bounded set of extended intervals $\{\boldsymbol{x}_i\}_{i\in J}$, $\boldsymbol{x}_i\in\mathbb{I}^*(\mathbb{R})$ for J an index set, the

$$dual(\boldsymbol{x})\triangleq[x^u,x^l],\ \text{where } x^l>x^u$$

$$pro(\boldsymbol{x})\triangleq\begin{cases}\boldsymbol{x} & \text{if} \quad \boldsymbol{x}\text{ is proper}\\ dual(\boldsymbol{x}) & \text{if} \quad \boldsymbol{x}\text{ is not proper}\end{cases}$$

$$\sup\{\boldsymbol{x}_i|i\in J\}=\left[\inf_{i\in J}\{x_i^l\},\sup_{i\in J}\{x_i^u\}\right],\ \text{largest possible interval}$$

$$\inf\{\boldsymbol{x}_i|i\in J\}=\left[\sup_{i\in J}\{x_i^l\},\inf_{i\in J}\{x_i^u\}\right]\ \text{smallest possible interval}$$

$$\Omega^{\boldsymbol{x}}(\boldsymbol{t})\triangleq\begin{cases}\boldsymbol{t} & \text{if} \quad \boldsymbol{x}\text{ is proper,}\\ dual(\boldsymbol{t}) & \text{if} \quad \boldsymbol{x}\text{ is not proper}\end{cases}$$

$$\boldsymbol{x}*\boldsymbol{y}=\Omega^{\boldsymbol{x}}\left(\Omega^{\boldsymbol{y}}\left(\{\boldsymbol{x}*\boldsymbol{y}|\boldsymbol{x}\in pro(\boldsymbol{x}),\boldsymbol{y}\in pro(\boldsymbol{y})\}\right)\right)$$

Let $x\in\mathbb{R}$, then

$$x^+\triangleq\max\{x,0\}$$
$$x^-\triangleq\max\{-x,0\}$$
$$x\vee y\triangleq\max\{x,y\}$$

and for intervals $\boldsymbol{x},\boldsymbol{y}\in\mathbb{I}^*(\mathbb{R})$,

$$\boldsymbol{x}^+\triangleq\left[\max\{x^l,0\}\wedge\max\{x^u,0\},\max\{x^l,0\}\vee\max\{x^u,0\}\right]$$
$$\boldsymbol{x}^-\triangleq\left[\max\{-x^l,0\}\wedge\max\{-x^u,0\},\max\{-x^l,0\}\vee\max\{-x^u,0\}\right]$$
$$\boldsymbol{x}\vee\boldsymbol{y}\triangleq\left[x^l\vee y^l,x^u\vee y^u\right]$$

Further detailed account of Kaucher interval arithmetic can be found in Lakeyev (1995).

APPENDIX C: ALGORITHMS

C.1 Algorithm: Interval smoothing

Let

$\boldsymbol{Z}_i = (Z_i^c, Z_i^r), i =, \ldots, n$,

$\boldsymbol{m}_i = \sum_{i=1}^{n} \boldsymbol{Z}_i/n = \left(\sum_{i=1}^{n} Z_i^c/n, \sum_{i=1}^{n} Z_i^r/n\right) = (m_i^c, m_i^r) = \boldsymbol{m}_i$

For each $i = 1, \ldots, n$

1. Initialize $\boldsymbol{m}_i^k = \boldsymbol{m}_i$, and $k = 1$
2. Find $n(i)$ closest interval values to the value at i i.e. according to the distance

$$||\boldsymbol{m}_i^k - \boldsymbol{Z}_j||^2 = (m_i^{k,c} - Z_j^c)^2 + \frac{\alpha}{\alpha+\beta}(m_i^{k,r} - Z_j^r)^2$$

3. Using appropriate kernel function for $\boldsymbol{g}(\cdot)$ (Table 4.3) compute,

$$\begin{aligned} \boldsymbol{m}_i^{k+1} &= \frac{\sum_{j=1}^{n(i)} \boldsymbol{g}\left(||\boldsymbol{m}_i^k - \boldsymbol{Z}_j||^2\right) \times \boldsymbol{Z}_j}{\sum_{j=1}^{n(i)} \boldsymbol{g}\left(||\boldsymbol{m}_i^k - \boldsymbol{Z}_j||^2\right)} \\ e &= \frac{1}{2 * \sum_{j=1}^{n(i)} \boldsymbol{g}\left(||\boldsymbol{m}_i^k - \boldsymbol{Z}_j||^2\right)} \\ k &= k+1 \end{aligned}$$

until $|\boldsymbol{m}_i^{k+1} - \boldsymbol{m}_i^k| < e$

Let $\boldsymbol{m}_i^{\text{conv}}$, denote the convergent point value; the local estimate at given location i

3. Assign $\hat{\boldsymbol{Z}}_i = \boldsymbol{m}_i^{\text{conv}}$

End For

C.2 Algorithm: Spatial Kernel Interval smoothing

Let

$\boldsymbol{Y}_i = \{\boldsymbol{x}_i, \boldsymbol{Z}_i\}$ input data $(x_i, y_i, (Z_i^c, Z_i^r))$,

$\{\boldsymbol{I}_i^k\}_{i=1,2,\ldots,n}, k = 1, \ldots$ evolving sequences in range domain; initial value $\boldsymbol{I}_1^k = \boldsymbol{m}_i$

$\{\boldsymbol{v}_i^k\}_{i=1,2,\ldots,n}, k = 1, \ldots$ evolving sequences in Spatial domain; initial value $\boldsymbol{v}_1^k = \boldsymbol{x}_i$

Define

1. interval distance, $d = ||\boldsymbol{I}_i^k - \boldsymbol{Z}_j||^2 = (I_i^{k,c} - \boldsymbol{Z}_j^c)^2 + \frac{\alpha}{\alpha+\beta}(I_i^{k,r} - \boldsymbol{Z}_j^r)^2$
2. the spatial distance, $s = ||\boldsymbol{v}_i^k - \boldsymbol{x}_j||^2$

   $\boldsymbol{g}(d) = \left(1 - \left(\frac{d}{b_i}\right)^2\right)^2$ the interval kernel function, bandwidth b_i (Table 4.3)

   $\boldsymbol{w}(s) = \left(1 - \left(\frac{s}{h_i}\right)^2\right)^2$ the spatial kernel weight, bandwidth h_i (Table 4.3)

$$\boldsymbol{Y}_i^{k+1} = \begin{pmatrix} \boldsymbol{I}_i^{k+1} \\ \\ \boldsymbol{v}_i^{k+1} \end{pmatrix} = \begin{matrix} \frac{\sum_{j \in n(i)} \boldsymbol{Z}_j \times \boldsymbol{w}(s) \times \boldsymbol{g}(d)}{\sum_{j \in n(i)} \times \boldsymbol{w}(s) \times \boldsymbol{g}(d)} & \text{I} \\ \\ \frac{\sum_{j \in n(i)} \boldsymbol{x}_j \times \boldsymbol{w}(s) \times \boldsymbol{g}(d)}{\sum_{j \in n(i)} \times \boldsymbol{w}(s) \times \boldsymbol{g}(d)} & \text{II} \end{matrix} \qquad \text{(C.1)}$$

For each $i = 1, \ldots, n$

3. Find $n(i)$ nearest neighbors of i according to spatial distance s
4. Compute the neighborhood average $\boldsymbol{m}_i = n^{-1} \sum_{j=1}^{n(i)} \boldsymbol{Z}_j$
5. Initialize $\boldsymbol{I}_i^k = \boldsymbol{m}_i, \boldsymbol{v}_i^k = \boldsymbol{x}_i$, and $k = 1, \ldots$
6. Compute

   i Eq. C.1, I

   ii $e = 1/\left(2 \times \sum_{j \in n(i)} \boldsymbol{w}(s) \times \boldsymbol{g}(d)\right)$

   iii $k = k + 1$

until either the interval shift $|\boldsymbol{I}_i^{k+1} - \boldsymbol{I}_i^k| < e$ or the spatial displacement $|\boldsymbol{v}_i^{k+1} - \boldsymbol{v}_i^k| < e$

Let $\boldsymbol{v}_i^{k+1} = \boldsymbol{v}_i^{\text{conv}}$, be the location convergence and

Let $\boldsymbol{I}_i^{k+1} = \boldsymbol{I}_i^{\text{conv}}$, be the range convergence, then

$\boldsymbol{Y}_i^{\text{conv}} = (\boldsymbol{v}_i^{\text{conv}}, \boldsymbol{I}_i^{\text{conv}})$, forms the convergence of the iterative procedure initialized with $\boldsymbol{Y}_i = (\boldsymbol{v}_i^k, \boldsymbol{I}_i^k)$. Run procedure for all $i = 1, \ldots, n$,

7. Assign , $\hat{\boldsymbol{Y}}_i = (\boldsymbol{x}_i, \boldsymbol{I}_i^{\text{conv}})$. the smoothed value at location $\boldsymbol{x}_i$

End For.

C.3 Algorithms: Kernel Joint distribution (Copula), Kernel Marginal and Kernel Membership functions estimators

Step 1. Applying Eq. 5.6, generate the bivariate kernel density estimate $\hat{f}(l, u)$ with lower(l) and upper(u) the coordinate random variables of the sample observed random intervals/vectors.

Step 2. Find the kernel marginal estimators $\hat{F}_L(l)$ and $\hat{F}_U(u)$ by applying Eq. 5.8 such that $\sup \hat{F}_L(l) = 1$ and $\sup \hat{F}_U(u) = 1$

Step 3. Find the kernel joint distribution estimator $\hat{F}_{LU}(l, u)$ from Eq. 5.7 such that $\sup \hat{F}_{LU}(l, u) = 1$. Applying the copula definition Eq. ??, we have $\hat{F}_{LU}(l, u) = C(F_L(l), F_U(u))$, thus the empirical copula is given by the joint kernel distribution estimator.

Step 4. Find the kernel membership function $\mu_{\tilde{\varphi}(Z)}(z)$ by applying Eq.?? such that $\sup \mu_{\tilde{\varphi}(Z)}(z) = 1$

C.4 Algorithms for Copula and Marginal Grade kriging

Step 1 For the "sampled" bivariate random variate observations, obtain the bivariate kernel density estimate $f_{Z^l Z^u}(z^l, z^u)$

Step 2 Integrate numerically the bivariate density estimate w.r.t. Z^l and Z^u accordingly to obtain the

a) lower marginal, $F_{Z^l}(z^l)$ then fit a lower marginal model

b) upper marginal, $F_{Z^u}(z^u)$ then fit an upper marginal model

c) joint distribution (copula) $F_{Z^l Z^u}(z^l, z^u)$.

Step 3 For each of the "sampled bivariate random variate" observations $\{(z_i^l, z_i^u)\} = \{(z^l(x_i, y_i), z^u(x_i, y_i)), (x_i, y_i) \in D\}$, calculate the sample

a) copula grades $\{C(w_i, v_i)\}$, where

$$C(x_i, y_i) \triangleq C(w_i, v_i) \triangleq F_{Z^l Z^u}(z_i^l, z_i^u) \tag{C.2}$$

b) lower and upper marginal distribution grades

$$\{F_{Z^l}(z_i^l)\} \text{ and } \{F_{Z^u}(z_i^u)\} \tag{C.3}$$

using the data and the fitted margins

Step 4 Use ArcGIS Geostatistical analyst, to generate *appropriate* semi-variogram functions for the sampled copula grades, lower grades and upper grades

Step 5 Use ArcGIS to generate

a) copula grades prediction map as well as prediction error map;

b) lower marginal grade map as well as prediction error map;

c) upper marginal grade map as well as prediction error map;

Step 6 For any unknown location, $(x_0, y_0) \in D$, based on the predicted copula grade map $\widehat{C}(x_i, y_i)$ and one of the predicted marginal distribution grade, say, $\widehat{F}_{Z^l}(z^l(x_i, y_i))$, the predicted value (z_0^l, z_0^u) can be obtained in terms of the inverse transformation

$$\widehat{z}_0^l = F_{Z^l}^{-1}(\widehat{w}_0) \tag{C.4}$$

where

$$\widehat{w}_0 = \widehat{F}_{Z^l}(z^l(x_0, y_0))$$

and

$$\widehat{z}_0^u = F_{Z^u}^{-1}\left(\inf\left\{v : \widehat{C}(\widehat{w}_0, v) \geq \widehat{C}(x_i, y_i)\right\}\right) \tag{C.5}$$

C.5 Algorithms for Credibility grade kriging

Step 1 Use the "sampled bivariate random variate observations", to obtain the joint bivariate kernel density. Integrate numerically the joint density to gen-

erate a kernel data-assimilated membership function for the induced fuzzy variable observations. finally, we define the corresponding data-assimilated credibility distribution function of induced fuzzy variable.

Step 2 For the "sampled fuzzy" observations$\{z_i, (x_i, y_i) \in D\}$, apply credibility distribution to calculate the sampled credibility grades $\{\Phi_\xi(z_i), (x_i, y_i) \in D\}$;

Step 3 Use ArcGIS software to generate an *appropriate* semivariogram model based on the sampled credibility grades$\{\Phi_\xi(z_i), (x_i, y_i) \in D\}$;

Step 4 Use ArcGIS software to generate credibility grade prediction map as well as prediction error map;

Step 5 For any location interested, say, $(x_0, y_0) \in D$, the predicted credibility grade is $\widehat{\Phi}_\xi(z_0)$. Then, the predicted value $\widehat{z}_0$ can be obtained in terms of the inverse transformation

$$\widehat{z}_0 = \Phi_\xi^{-1}\left(\widehat{\Phi}_\xi(z(x_0, y_0))\right)$$

since the credibility distribution Φ_ξ is increasing function on z, the inverse function is

$$\Phi_\xi^{-1}(y) = \inf(z : \Phi_\xi(z) \geq y).$$

APPENDIX P: PAPERS PUBLISHED

P.1 Interval-Spatial methods

→ N. Wabiri, R. Guo, and C. Thiart. **Interval-Spatial methods,** *Proceedings of the GIS Research UK 13th Annual Conference GISRUK2005*, Editors. R. Billen, J. Drummond, D. Forest, and E. Joo. Glasgow, Scotland, April 6-8, 2005 , pp121-129.

Interval–Spatial Methods

N. Wabiri, R. Guo, C. Thiart

Department of Statistical Sciences
University of Cape Town, Private Bag Rhodes Gift
7701, Cape Town, South Africa
nwabiri@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za
Corresponding author: Njeri Wabiri nwabiri@stats.uct.ac.za

KEYWORDS: kriging, algorithm development, statistical analysis

1. Introduction

Kriging is a well known technique of spatial interpolation; known sample values $\{Z(\mathbf{x}_i)$, and $\mathbf{x}_i$, the location of point i , i = I, ..,N) are used to predict unknown sample values ($Z(\mathbf{x}_o)$, and $\mathbf{x}_o$, the location of the unknown point). In ordinary kriging, variables are represented by a single value, for example, at location $\mathbf{x}_i$ we might observe concentration of zinc (measured in ppm) in soil samples. However, in several applications we might not observe a single value at location $\mathbf{x}_i$, but a sample of values. Examples are measurements of pollutant concentrations; SO2, NO, CO_2; daily currency exchange rates in foreign exchange; and in petroleum industries, we may have data in form of inequalities due to conventional drilling practices. In all these cases the exact values is not observable, what is know is that the values lies within a given range of sample values.

Usually the sample values will be summarized by a suitable "single" measure, either the mean or the median. When a "single" measure is used, sample range is ignored and valuable information is thrown away. However, it would be more interesting to take into account the minimum and the maximum values measured at the locations rather than the average or the median because they offer more detailed and complete information about the phenomena under study. Besides, in most cases data may be limited, which makes it even more necessary to utilize all the available information.

In this paper we extended the semivariogram of ordinary kriging to incorporate interval-valued regionalized variables. At location $\mathbf{x}_i$ we are now observing the interval, $Z(\mathbf{x}_i) = [Z^l(\mathbf{x}_i), Z^u(\mathbf{x}_i)]$, where $Z^l(\mathbf{x}_i)$ represent the lower bound and $Z^u(\mathbf{x}_i)$ the upper bound of the interval, with i = 1, ...N.

2. Semivariogram - Ordinary Kriging

Given a regionalized variable $\{Z(x): x \in D \subset \Re^d\}$ with known values at N sample points $\{x_i : i = 1,..., N\}$, it is considered that values at close range take on similar values, (Tobler, 1970), first law of Geography. Therefore to estimate unknown values we can use the strength of the relationship between the known values and the unknown value. This

relationship is quantified by the semivariogram, the average square difference between pairs of sample values separated by distance h:

$$\hat{\gamma}(h) = \frac{1}{2|N_h|}\sum_{i=1}^{N_h}\left[(Z(x_i+h) - Z(x_i))^2\right] \quad (1)$$

with N_h the pair of samples at lag h apart, and $|N_h|$ number of such pairs.The semivariogram is then used in kriging to calculate estimates at unknown locations. Different types of kriging assumes that local means are not necessarily closely related to the population mean, hence uses only the samples in the local neighborhood for the estimate. In particular, in ordinary kriging, the regionalized variable is assumed to be locally stationary thus allowing for an estimate of unknown value, Z at location x_0. $Z(x_o)$ is calculated using the weighted average of the known values in the local neighborhood, i.e. $Z^*(x_0) = \sum \lambda_i \bullet Z(x_i)$. This is possible assuming second-order stationarity with prediction model:

$$Z(x) = \mu + \varepsilon(x), \quad (2)$$

where μ is the local stationary mean and $\varepsilon(x)$ the standard error. Optimal weights are obtained by solving the linear kriging system of equations obtained by minimizing the mean prediction error $E(Z(x_0) - Z^*(x_0))^2$ conditional on $\sum \lambda_i = 1$. (Goovaerts, 1997; Cressie, 1993).

The linear kriging system:

$$\sum \lambda_j \gamma(z(x_i) - z(x_j)) + \psi = \gamma(z(x_i) - z(x_0)), \quad i = 1,...,n$$
$$\sum \lambda_j = 1 \quad (3)$$

With $\gamma(z(x_i) - z(x_j))$, the variogram for known data values, $\gamma(z(x_i) - z(x_0))$ the variogram for know an unknown prediction location and ψ is the Lagrange parameter. In matrix notation Equation 3 reduces to $\Gamma^T \lambda = \gamma$, with $\Gamma_{(n+1)(n+1)}$ real-valued matrix; γ_{n+1} real-valued vector both adjusted for the weights constraints. Weights are obtained by solving the kriging system. This has been implemented using C++ template for geostatistical analysis and R sgeostat.

3. Fitting the semivariogram for Interval –valued data

Kriging of interval is often done naively by analyzing either the median or the average value of the interval range. This reduces to ordinary kriging with substantial loss of information. An alternative method is to estimate variograms for the minimum and the maximum values separately. However, this does not only reduce to ordinary kriging, but it is based on the

assumption that the measurements (i.e. minimum and the maximum values) are independent, which is not the case as they both represent the same phenomena.

A better method would be to looks at the possible ways utilizing all the sample range data (minimum and maximum), with a goal to obtain sharp interval bounds for the estimates. This will result to inner interval estimate which includes all the possible evaluations of the phenomena under study. This paper introduces interval- valued variogram by addressing following issues:

1) estimating the interval-variogram while maintaining the dependence between upper and lower values;
2) Exploring extended interval computation arithmetic methods to extend spatial methods for analysis of interval spatial data.
3) Setting up appropriate interval weight constraints;

To address these issues, we have extended Diamond, 1988 approach, using a more generalized metric on the space of nonempty compact intervals, and finding all the possible optimal solutions for the general case of non-degenerate regionalized interval-values. This has been implemented using Kaucher extended interval arithmetic which allows for the kriging weights to take on any value on the extended real (-∞, +∞) and the problem of obtaining an inner interval estimate of the interval linear kriging system is NP-complete. Much of the work is implemented in C++ .

Interval-valued variogram is a function of the distance between interval-valued sets, the difference between the minimum and maximum values at the corresponding spatial locations. The interval random function model is:

$$\overline{Z}(x) = \overline{m}(x) + \overline{\varepsilon}(x) \tag{4}$$

with $\overline{Z}(x) = [Z^{l}(x), Z^{u}(x)]$, $\overline{m} = [m^{l}(x), m^{u}(x)]$, and and $\overline{\varepsilon} = [\varepsilon^{l}(x), \varepsilon^{u}(x)]$

The interval sample variogram is given as:

$$\hat{\overline{\gamma}}(h) = \frac{1}{2|N_h|}\sum_{i=1}^{N_h}\left[(\overline{\varepsilon}(x_i + h) - \overline{\varepsilon}(x_i))^2\right] \tag{5}$$

Applying interval arithmetic operations to the quadratic part of (5) results to

$$\begin{aligned}\hat{\overline{\gamma}}(h) &= \frac{1}{2|N_h|}\sum_{i=1}^{N_h}[f_{\min}(i,h), f_{Max}(i,h)] \\ &= \left[\frac{1}{2|N_h|}\sum_{i=1}^{N_h} f_{\min}(i,h), \frac{1}{2|N_h|}\sum_{i=1}^{N_h} f_{\max}(i,h)\right] \\ &= \left[\hat{\gamma}^{l}(h), \hat{\gamma}^{u}(h)\right] \text{ , lower and upper variograms}\end{aligned} \tag{6}$$

where $f_{\min}(i,h) = \min(c_{i1}^{\,2}, c_{i1}, c_{i2}, c_{i2}^{\,2})$ and $f_{\max}(i,h) = max(c_{i1}^{\,2}, c_{i1}, c_{i2}, c_{i2}^{\,2})$, with

$$\begin{aligned} c_{i1} &\equiv \varepsilon^{l}(x_i + h) - \varepsilon^{u}(x_i) \\ c_{i2} &\equiv \varepsilon^{u}(x_i + h) - \varepsilon^{l}(x_i) \end{aligned} \tag{7}$$

Assuming second order stationarity, (4) reduces to prediction model:

$$\overline{Z}(x) = \overline{\mu} + \overline{\varepsilon}(x), x \in D \subset \Re^2 \tag{8}$$

With $\overline{\mu}$ the local interval mean. The interval estimate $\overline{Z}^{*}(x_0) = \sum \overline{\lambda}_i \bullet \overline{Z}(x_i)$ is obtained by minimizing mean square prediction error

$$E\left[\overline{Z}(x_0) - \overline{Z}^{*}(x_0)\right]^2 = E\left[\left[Z^{l}(x_0) - Z^{*l}(x_0)\right]^2, \left[Z^{u}(x_0) - Z^{*u}(x_0)\right]^2\right] \tag{9}$$

This leads to linear interval kriging system

$$\begin{aligned} &\sum \overline{\lambda}_j \overline{\gamma}(\overline{z}(x_i) - \overline{z}(x_j)) + \overline{\psi} = \overline{\gamma}(\overline{z}(x_i) - \overline{z}(x_0)), \quad i = 1,\ldots,n \\ &\sum \overline{\lambda}_i = 1 \end{aligned} \tag{10}$$

In matrix notation, we have $\overline{\Gamma}^{T}\,\overline{\lambda} = \overline{\gamma}$, with $\overline{\Gamma}$ *(n +1) x (n+1)* interval matrix; $\overline{\gamma}$ *(n+1)* interval vector both adjusted for the weights constraints.

Using classical interval arithmetic, sample interval-variogram is estimated. The weights will be obtained using Kaucher interval arithmetic, where the interval kriging system Equation 8 reduces to solving:

$$A_{(n+1)x(n+1)}\left[\lambda_i^{\,l}, \lambda_i^{\,u}\right]^{T} - B_{(n+1)x(n+1)}\left|\left[\lambda_i^{\,l}, \lambda_i^{\,u}\right]^{T}\right| = \left[\gamma_{i0}^{\,l}, \gamma_{i0}^{\,u}\right] \tag{11}$$

where A=$\frac{1}{2}\left(\gamma^{u}_{i,j} + \gamma^{l}_{i,j}\right)$, and B= $\frac{1}{2}\left(\gamma^{u}_{i,j} - \gamma^{l}_{i,j}\right)$ are real-valued matrix, respectively, the

midpoints and radius of the elements of the interval kriging matrix. Solution to the kriging system will be done using adjusted verified solution method for interval computation and interval-Cholesky decomposition implemented in c++.

4. Illustrative Example

Data from SIC (2004) is very form a particular suitable case for interval analysis. Data consist of natural ambient radioactivity measured in Germany, given in form of gamma dose rates, reported by means of the national automatic monitoring network (IMIS).
From a rectangular with 1008 monitoring stations, 11 days of measurements have been randomly selected within the a period of 12 months and the average daily dose rates calculated for each day. From the 11 dataset we picked the minimum and the maximum values generating a spatial random interval. The dataset consist of

- *ID*: this integer value is the number (unique value) of the monitoring station chosen by us.
- *X*: X coordinate of the monitoring station indicated in meters
- *Y*: Y coordinate of the monitoring station indicated in meters
- *Z*: mean gamma dose rate measured during 24 hours. Units are nanoSieverts/hour

Using a sample of 200 measurements, the goal is to estimate measurement on the remaining 808 monitoring sites. Figure 1 indicates data locations.

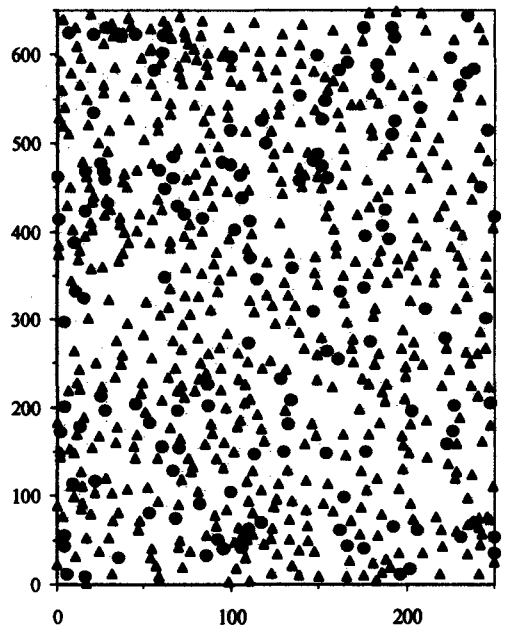


Figure 1. Data locations: ●--sample values
▲-- Estimate locations

Using this data, we applied interval variogram analysis

5. Results and Discussion

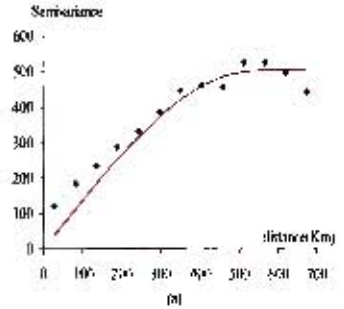
Below are results of variogram analysis, Figure 2 (a), (b), (c) indicate the variograms fitted for median, min, and maximum values.

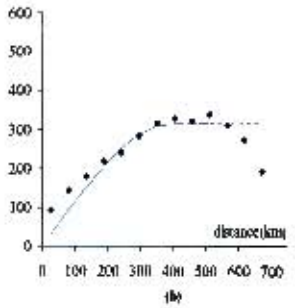
Figure: 2 Sample variogram and fitted models

(a) Maximum values
Nugget = 92.37603,
Sill = 12.71599,
Range = 640615.

(b) Mid values
Nugget = 78.91895
Sill = 312.54907
Range = 476.52823

(c) Minimum values
Nugget = 74.63535
Sill = 241.43437,
Range = 400.0513





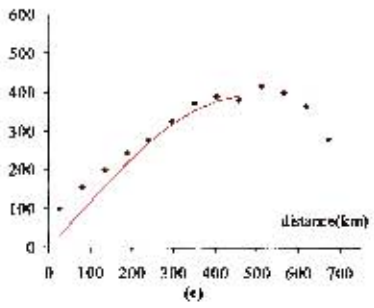


Figure 2. The variograms fitted for median, min, and maximum values.

Using classical interval arithmetic we have also estimated an interval variogram which provides an enclosure for the minimum, mid point and maximum variogram estimates. (Figure 3). The upper and lower values are generated by classical interval arithmetic. Narrowing the interval bands for the interval variogram through verified methods for solving linear system will provide an interval estimate that is guaranteed to include the optimal inner variogram estimate.

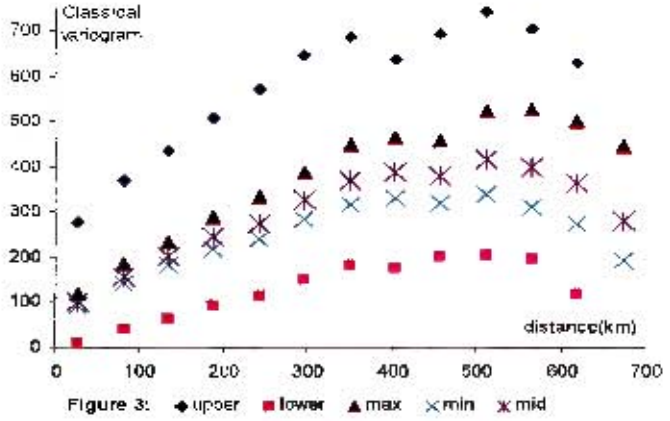


Figure 3. Classical and Interval Variogram.

6. Conclusions

It is clear that classical interval arithmetic will in most cases lead to large interval bounds; hence we have alternatively defined the variogram intermes of well defined metrics is the space of compact closed intervals. This will be an extension of Diamond, 1988 and the well known L2-metric in the Hilbert space of functions.

This represents an ongoing research on extension of interval-uncertainty methods for handling uncertainty in interval spatial data.

References

Cressie, N , 1993. . Statistics for Spatial data. John Wiley and Sons, New York, revised edition.
Diamond, P., 1988 Interval-valued random functions and the kriging of intervals. Mathematical Geology, 20(3):145–165.
Lakeyev, A., 1995. Linear algebraic equations in Kaucher arithmetic. Extended Abstracts, A supplement to the international Journal of Reliable computing.
SIC, 2004, http://www.ai-geostats.org/events/sic2004/sic2004_data.htm

Biography

Njeri Wabiri holds a M.sc. in Mathematical statistics and is currently pursuing PhD studies in GIS with research focus on methods of handling uncertainty inherent in environmental

spatial data by incorporating data from different sources.

P.2 Interval Kriging Under Generalized metric

→ N. Wabiri, R. Guo and C.Thiart. **Interval Kriging Under Generalized metric,** *Proceedings of International Workshop on Recent Advances Stochastic Operations Research*, Edited by Tadashi Dohi, Shunji Osaki and Katsushige Sawaki. Canmore, Alberta, Canada, August 25-26,2005, pp 290-297. (ISBN 4-939146-03-0)

Interval Kriging under Generalized Metric

NJERI WABIRI† RENKUAN GUO † AND CHRISTIEN THIART †

†Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes Gift 7707,
Cape Town, South Africa
E-mail: nwabiri@stats.uct.ac.za, rguo@stats.uct.ac.za, thiart@stats.uct.ac.za

Abstract— Kriging [1] is a well-known and established technique of spatial interpolation and has become a standard procedure or component in the commercial software in spatial statistical analysis or GIS (Geographic Information System), for example ESRI ArcGIS. Naturally, spatial observations are often not represented in the form of a number but an interval-valued number; e.g., pollutant concentrations SO_2, CO_2 cannot be measured in an exact way because of the fluctuation. Therefore, an exact value is not observable, but what is observed is an interval within which the values lies. A simplified approach is to perform kriging on midpoints of interval-valued sample observations. By taking the midpoint, we loss imprecision information contained in each interval-valued observation. In this paper we propose a generalized metric for interval-valued observations and develop a methodology for interval-valued kriging. The method represents a more realistic framework and includes [2] formulation as a special case. An air pollution example, SIC2004 [3] is given as an illustration.

Keywords— Generalised metric, Interval Data kriging, Spatial Process, Geostatistics

1. INTRODUCTION

In the real world, spatial data sometimes cannot be recorded or collected precisely. For instance, pollutant concentrations; SO_2, NO, CO_2 cannot be measured in an exact way because of the fluctuation, and the temperature is also not able to be measured precisely because of similar reason. An appropriate way to describe the SO_2 concentrations in the air is to say that the SO_2 concentration is "below or above certain level " or "bounded" by two values. The phrase "below or above given levels " or "bounded" can be regarded as a interval-valued number. This paper addresses this problem. We explore the random closed interval method as a natural appropriate tool in modeling the statistical models when imprecise data has been observed. We extend metric-based interval methods to analysis of imprecise spatial data.

Among the methods of spatial modelling , kriging [1], is a well-known and established technique for spatial interpolation. However, in its original formulations, kriging was designed to deal with exact measurements, therefore intervals of values or any imprecise data are represented by rough approximations e.g. mean, median and indicator values. It is obvious that this convenience costs the loss of imprecision information contained in each interval-valued observation. [2] investigated interval-valued kriging under Hausdorff metric, and the L_2 metric similar to [4] metric. However, these metrics assign equal weights to squared Euclidean distances between the interval endpoints, midpoints and radii. The methods ignore the variability within the interval bounds. We address this shortcoming by proposing a generalized metric for interval-valued observations.

2. INTERVAL-VALUED NUMBER AND ITS METRIC

Let $\mathbf{R}$ be a real number set, and let $\mathcal{K}_c(\mathbf{R})$ denote a class of nonempty closed compact intervals with semi linear structure

$$a + b = \{a + b : a \in a, b \in b\}, \quad \lambda a = \{\lambda a : a \in a\} \tag{1}$$

for $a, b \in \mathcal{K}_c(\mathbf{R})$, and $\lambda \in \mathbf{R}$. Since support function uniquely defines each compact interval, preserving the semi linear structure, the space $\mathcal{K}_c(\mathbf{R})$ is embedded into a linear space with well

defined linear structure, and well known L_2-metric. Therefore, mean and variance of compact intervals are obtained with respect to the metric.

Explicitly, a nonempty closed compact interval $\boldsymbol{x} = [x^l, x^u]$ for $x^l \leq x^u$, $x^l, x^u \in \mathbf{R}$, where x^l and x^u are the lower and upper interval bounds. For midpoint and range notation: $\boldsymbol{x} = [x^c - x^r, x^c + x^r] = \{x^c, x^r\}$, with

$$\text{midpoint:} \quad x^c = \frac{1}{2}(x^u + x^l) \quad \text{and range:} \quad x^r = \frac{1}{2}(x^u - x^l).$$

Most metrics between intervals, often take account of only some representative points. For example, [2] metric used interval vertices while [4] metric used midpoint and range. The metrics are similar and assigns equal weight to squared Euclidean distances between the corresponding interval points. In reality we would expect the midpoint to contain more information while the range is an indication of interval uncertainty. Hence, a distance measure that optimally weights the influence of the interval range would give a better reflection.

2.1. The Generalised d_G-metric on Interval Space

Theorem 1 *Let a, b be elements of a class $\mathcal{K}_c(\mathbf{R})$ of nonempty closed compact intervals, then*

$$d_G^2(a, b) = |a^c - b^c|^2 + \frac{\alpha}{\alpha + \beta}|a^r - b^r|^2 \tag{2}$$

Proof. Assume that $([0, 1], \mathcal{B}([0, 1]), G)$ is a probability space, the generalized metric

$$d_G(a, b) = \sqrt{\int_0^1 [\lambda \delta_2^2(a, b) + (1 - \lambda)|a^c - b^c|^2] dG(\lambda)}$$

with L_2-metric

$$\begin{aligned} \delta_2^2(a, b) &= \frac{1}{2} \times [|a^u - b^u|^2 + |a^l - b^l|^2] \\ &= |a^c - b^c|^2 + |a^r - b^r|^2 \qquad \text{using midpoint range notation} \end{aligned} \tag{3}$$

Thus, $d_G^2(a, b)$ is a weighted distance metric. The normalised weight measure $G(\lambda)$ for $\lambda \in (0, 1)$, reflects the contribution of the L_2- metric and of the squared Euclidean difference between interval midpoints. Instead of assigning weights to selected interval points, we let λ he distributed continuously within $(0, 1)$ with a continuous symmetric probability density function (pdf). Because of its extreme flexibility, the Beta probability density function (pdf); a continuous and bounded pdf within $(0, 1)$, is ideally suited for the choice of λ. The shape of the Beta pdf depends on the choice of its two parameters $(\alpha, \beta > 1)$.

Therefore, let $dG(\lambda) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1 - \lambda)^{\beta-1}d\lambda$, and using (3)

$$\begin{aligned} d_G^2(a, b) &= \int_0^1 \left[\lambda\left(|a^c - b^c|^2 + |a^r - b^r|^2\right) + (1 - \lambda)|a^c - b^c|^2\right] \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1 - \lambda)^{\beta-1}d\lambda \\ &= |a^c - b^c|^2 + |a^r - b^r|^2 \int_0^1 \lambda \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\lambda^{\alpha-1}(1 - \lambda)^{\beta-1}d\lambda \\ &= |a^c - b^c|^2 + \frac{\alpha}{\alpha + \beta}|a^r - b^r|^2 \end{aligned}$$

When applied to point intervals, $d_G(a, b)$ reduces to usual distance $d([a, a], [b, b]) = |a - b|$ for real numbers. ■

To show that $d_G^2(a, b)$ is a complete metric, we need to show that it satisfies the triangular inequality: $d_G^2(a, b) \leq d_G^2(a, d) + d_G^2(d, b)$, for $a, b, d \in \mathcal{K}(\mathbf{R})$. For the triangular property of Euclidean distance for centers and ranges, we can assert that:

$$|a^c - b^c|^2 \leq |a^c - d^c|^2 + |d^c - b^c|^2 \quad \text{and} \quad |a^r - b^r|^2 \leq |a^r - d^r|^2 + |d^r - b^r|^2$$

Then it follows that:

$$|a^c - b^c|^2 + \frac{\alpha}{\alpha+\beta}|a^r - b^r|^2 \;\leq\; \Big(|a^c - d^c|^2 + \frac{\alpha}{\alpha+\beta}|a^r - d^r|^2\Big) + \Big(|d^c - b^c|^2 + \frac{\alpha}{\alpha+\beta}|d^r - b^r|^2\Big)$$

and the space $(\mathcal{K}_c(\mathbb{R}), d_G(\cdot))$ is complete and separable. Hence, using the $d_G(\cdot)$ metric, we can define the variance and the covariances of interval objects.

3. INTERVAL-VALUED RANDOM FUNCTION

Let $\boldsymbol{Z}(x_i) = [Z(x_i)^l, Z(x_i)^u]$ $(i = 1, \ldots, n)$ represent the interval-valued regionalized variables observed at n spatial locations x_i in spatial region $D \subseteq \mathbb{R}^d$. We modelled the set of interval regionalized variables with a class of nonempty closed compact intervals, $(\mathcal{K}_c(\mathbb{R}), d_G(\cdot))$. For simplicity of notation, let $Z^c(\boldsymbol{x})$, $Z^r(\boldsymbol{x})$ denote the midpoint and range of interval valued regionalized variable, and m_c, m_r the mean of the midpoint and range vectors of interval valued regionalized variables. Under second order spatial stationarity, the interval-valued random function is defined as

$$\boldsymbol{Z}(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x}) + \boldsymbol{e}(\boldsymbol{x}) \tag{4}$$

with $\boldsymbol{e}(\boldsymbol{x}) = \{e^c(\boldsymbol{x}), e^r(\boldsymbol{x})\}$ and

$$\begin{aligned} \text{Mean:} \quad & \boldsymbol{m}(\boldsymbol{x}) = \{\boldsymbol{E}Z^c(\boldsymbol{x}), \boldsymbol{E}Z^r(\boldsymbol{x})\} = \{m_c, m_r\}, \\ \text{Isotropic covariance:} \quad & \boldsymbol{C}(\boldsymbol{h}) = \boldsymbol{E}[\boldsymbol{Z}(\boldsymbol{x})\boldsymbol{Z}(\boldsymbol{x}+\boldsymbol{h})] - \boldsymbol{m}^2 \end{aligned} \tag{5}$$

The isotropic covariance is a function of separation vector $\boldsymbol{h}$, and $\boldsymbol{m}^2 = \{m_c^2, m_r^2\}$. Further,

$$C^c(h) = E[Z^c(x+h), Z^c(x)] - m_c^2 \qquad C^r(h) = E[Z^r(x+h), Z^r(x)] - m_r^2 \tag{6}$$

are the midpoint and spread covariances covariances respectively. For $h = 0$

$$C^c(0) = E[Z^c(x), Z^c(x)] - m_c^2 \qquad C^r(0) = E[Z^r(x), Z^r(x)] - m_r^2$$

The variance of interval-valued regionalized variable $\boldsymbol{Z}(\boldsymbol{x})$ with respect to d_G metric

$$\begin{aligned} \text{Var}\,(\boldsymbol{Z}(\boldsymbol{x})) &= \boldsymbol{E}\big[d_G^2[\boldsymbol{Z}(\boldsymbol{x}), \boldsymbol{m}]\big] = \boldsymbol{E}[Z^c(\boldsymbol{x}) - m_c]^2 + \frac{\alpha}{\alpha+\beta}\boldsymbol{E}[Z^r(\boldsymbol{x}) - m_r]^2 \\ &= C^c(0) + \frac{\alpha}{\alpha+\beta}C^r(0) \end{aligned} \tag{7}$$

is real-valued. Similarly, for a linear combination of observed interval regionalized variables

$$\boldsymbol{Z} = \sum_{i=1}^{n} \lambda_i \boldsymbol{Z}(\boldsymbol{x}_i),$$

the variance with respect to d_G-metric

$$\begin{aligned} \text{Var}\,(\boldsymbol{Z}) &= \boldsymbol{E}[d_G^2[\sum_{i=1}^{n}\lambda_i \boldsymbol{Z}(\boldsymbol{x}_i), \sum_{i=1}^{n}\lambda_i \boldsymbol{m}]] \\ &= \boldsymbol{E}\Big\{\sum_{i=1}^{n}\lambda_i[Z^c(x_i) - m_c]\Big\}^2 + \frac{\alpha}{\alpha+\beta}\boldsymbol{E}\Big\{\sum_{i=1}^{n}\lambda_i[Z^r(x_i) - m_r]\Big\}^2 \\ &= \sum_{i,j=1}^{n}\lambda_i\lambda_j \boldsymbol{E}[Z^c(x_i)Z^c(x_j)] - m_c^2 + \frac{\alpha}{\alpha+\beta}\sum_{i,j=1}^{n}\lambda_i\lambda_j \boldsymbol{E}[Z^r(x_i)Z^r(x_j)] - m_r^2 \\ &= \sum_{i,j=1}^{n}\lambda_i\lambda_j C^c(x_i, x_j) + \frac{\beta}{\alpha+\beta}\sum_{i,j=1}^{n}\lambda_i\lambda_j C^r(x_i, x_j). \end{aligned} \tag{8}$$

ISBN: 4-939146-03-0

Proceedings of International Workshop
on Recent Advances Stochastic
Operations Research,
Canmore, Alberta, Canada, August 25-26,2005

3.1. Ordinary Interval Kriging (OIK)

Ordinary kriging implicitly assumes a non-stationary random function model where stationarity is limited within each search neighbourhood. Prediction involves minimizing the error variance of the estimator of interval-valued random function with respect to the d_G-metric. Normally, spatial measurements are made on finite volumes (i.e., sample support), rather than points $\boldsymbol{x} = (x, y, z)$ in space. Therefore, given $\boldsymbol{z}(\boldsymbol{x})$, a realisation of interval-valued regionalized variable $\boldsymbol{Z}(\boldsymbol{x})$, let

$$s_0 = \frac{1}{|D|}\int_d z(x')dx', \quad \text{for } x' \in D$$

be the average of $\boldsymbol{z}(\boldsymbol{x})$, over the support D. The corresponding random field denoted as $\boldsymbol{S}_0$, is estimated by

$$\boldsymbol{S}_0^* = \sum_{i=1}^{n}\lambda_i \boldsymbol{Z}(\boldsymbol{x}_i)$$

using the dataset $\{\boldsymbol{Z}(\boldsymbol{x}_i) : i = 1, \ldots, n\}$.

Kriging results to the best linear unbiased predictor ,BLUP [1]. "**Best**" possible means that the estimator is unbiased and optimal in the sense of minimum squared error (mse). The goal is to obtain weights λ_i such that

(*i*) $\boldsymbol{E}(\boldsymbol{S}_0^*) = \boldsymbol{E}(\boldsymbol{S}_0) = \boldsymbol{E}(\boldsymbol{Z}) = \boldsymbol{m}$
(*ii*) $\boldsymbol{E}d_G(\boldsymbol{S}_0^*, \boldsymbol{S}_0)^2$ is minimized under a set of constraints

In OIK, condition (*i*) equivalent to the constraint $\sum_{i=1}^{n}\lambda_i = 1$, secures the unbiased estimator and we need to minimise (*ii*) with respect to d_G-metric. Using (2) we minimise

$$\begin{aligned} \boldsymbol{E}d_G(\boldsymbol{S}_0^*, \boldsymbol{S}_0)^2 &= \boldsymbol{E}\Big[(S_0^{c*} - S_0^c)^2 + \frac{\beta}{\alpha+\beta}(S_0^{r*} - S_0^r)^2\Big] \\ &= \boldsymbol{E}S_0^{c*2} - 2\boldsymbol{E}S_0^{c*}S_0^c + \boldsymbol{E}S_0^{c2} + \frac{\beta}{\alpha+\beta}\big[\boldsymbol{E}S_0^{r*2} - 2\boldsymbol{E}S_0^{r*}S_0^r + \boldsymbol{E}S_0^{r2}\big] \end{aligned} \tag{9}$$

Let

$$\begin{aligned} C^c(D, D) &= \frac{1}{|D|^2}\int_d\int_d C^c(x, x')\,dx\,dx' \quad \text{for } x' \in D \text{ and} \\ C^c(x_i, D) &= \frac{1}{|D|^2}\int_d\int_d C^c(x_i, x)dxdx_i \end{aligned}$$

then,

$$\boldsymbol{E}S_0^{c2} = C^c(D, D) + m_c^2 \quad \text{and} \quad \boldsymbol{E}(S_0^{c*}, S_0^c) = \sum_{i=1}^{n}\lambda_i C^c(x_i, D) + m_c^2,$$

and similar for $\boldsymbol{E}(S_0^{r2})$ and $\boldsymbol{E}(S_0^{r*}S_0^r)$. Weights can take any value on the extended real i.e., $[-\infty \cup \mathbb{R}]) \cup (\mathbb{R} \cup +\infty]$, and,

$$\begin{aligned} \boldsymbol{E}S_0^{c*2} &= \sum_{i,j=1}^{n}\lambda_i\lambda_j C^c(x_i, x_j) + m_c^2 \\ \boldsymbol{E}S_0^{r*2} &= \sum_{i,j=1}^{n}\lambda_i\lambda_j C^r(x_i, x_j) + m_r^2 \end{aligned}$$

ISBN: 4-939146-03-0

are defined, ensuring positive definiteness condition. Substituting back to (9)

$$
\begin{aligned}
E d_C(S_0^*, S_0)^2 &= \sum_{i,j=1}^{n} \lambda_i \lambda_j C^*(x_i, x_j) - 2 \sum_{i=1}^{n} \lambda_i C^*(x_i, D) + C^*(D, D) \\
&\quad + k(\alpha, \beta) \Big[\sum_{i,j=1}^{n} \lambda_i \lambda_j C^*(x_i, x_j) - 2 \sum_{i=1}^{n} \lambda_i C^*(x_i, D) - C^*(D, D) \Big] \\
&\quad - \sum_{i,j=1}^{n} \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^{n} \lambda_i C_{i0} + C(D, D) \qquad (10)
\end{aligned}
$$

with

$$
\begin{aligned}
C_{ij} &= C^*(x_i, x_j) + k(\alpha, \beta) C^*(x_i, x_j) \\
C_{i0} &= C^*(x_i, D) - k(\alpha, \beta) C^*(x_i, D) \\
C(D, D) &= C^*(D, D) + k(\alpha, \beta) C^*(D, D), \quad \text{and } k(\alpha, \beta) = \frac{\alpha}{\alpha - \beta}
\end{aligned}
$$

Introducing the Lagrange multiplier for condition (i), and minimising (10), we obtain the OIK system,

$$
\begin{aligned}
\sum_{i=1}^{n} \lambda_i C_{ij} + [\underline{\psi}, \overline{\psi}] &= C_{i0} \qquad i = 1, \ldots, n \\
\sum_{i=1}^{n} [\underline{\lambda}_i, \overline{\lambda}_i] &= [1, 1] \qquad (11)
\end{aligned}
$$

where $[\underline{\psi}, \overline{\psi}]$ is the interval Lagrange parameter. As is the common practice in geostatistics, we modeled the variogram using semivariance function, $\Gamma(x(x_i) - x(x_j)) = \gamma^*(x(x_i) - x(x_j)) - k(\alpha, \beta)\gamma^*(x(x_i) - x(x_j))$, and using the relation $C(h) = C(0) - \Gamma(h)$ [5, p.240], we obtained the covariances which are substituted in (11). The OIK variance is

$$
\sigma_k^2 = C(D, D) - \sum_{i=1}^{n} [\underline{\lambda}_i, \overline{\lambda}_i] C_{i0} + [\underline{\psi}, \overline{\psi}].
$$

4. ILLUSTRATIVE EXAMPLE

To illustrate the methods, we used data from [3]. Data consist of natural ambient radioactivity measured in Germany, given in form of gamma dose rates, reported by means of the national automatic monitoring network (IMIS). The data was collected from rectangular region with 1008 monitoring stations. At each station, 11 days of measurements were randomly selected within a period of 12 months and the average daily dose rates calculated. From the 11 datasets of daily dose rates, we formed a spatial random interval from the minimum and the maximum values. Data consist of a sample of 200 measurements with

ID: this integer value is the number (unique value) for each monitoring station

X: x coordinate of the monitoring station indicated in meters

Y: y coordinate of the monitoring station indicated in meters

Z: z: interval of mean gamma dose rate at each location. Units are nanoSieverts/hour

Figure 1 shows the spatial distribution of sample and prediction locations. We generated midpoint and range values for each location, and figure 2 gives the summary statistics for midpoint(left) and range(right) vectors

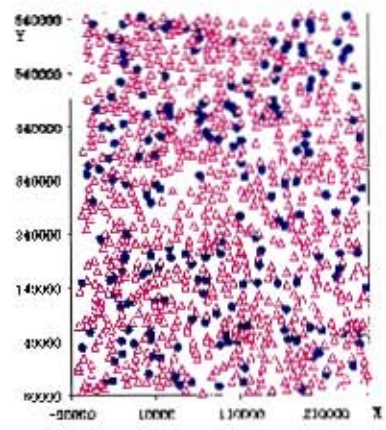


Figure 1: Sample (solid circle) and prediction (empty triangle) locations

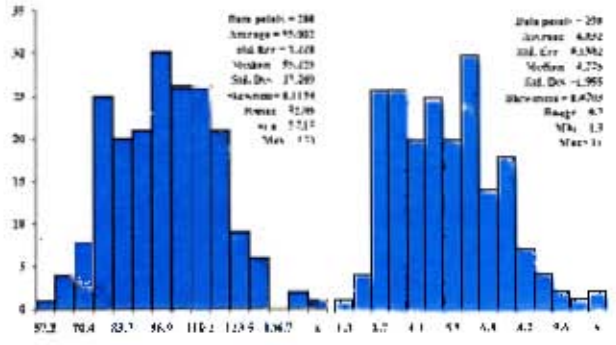


Figure 2: Midpoint and Range Summary Statistics.

4.1. Structural Analysis and Prediction

Using trend analysis, we examined the spatial variability of gamma dose rates of natural ambient radioactivity and its geographical variables. Respectively, figures 3 and 4 indicates semi-significant change in midpoint and range values with east-west and north-south direction.

Spatial structure for the midpoint and range vectors was modelled using spherical models (figures 5 and 6), with C0, C1, and R, the psill, nugget and range parameters respectively. The high nugget effect for the radii data is a natural expectation as we are assuming the interval range to be a measure of uncertainty and randomness. In addition, high points in the variogram models are above the range of spatial uncertainty and are therefore ignored. To use our new kriging system optimal values of α and β are required.

Figure 3: Midpoint Trend Analysis

Figure 4: Range Trend Analysis

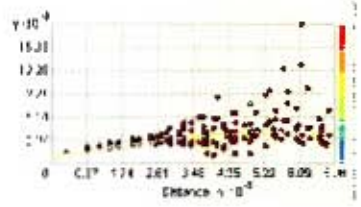


Figure 5: Midpoint spherical model: C0=70.9040, C1= 231.8624, R=332711.4

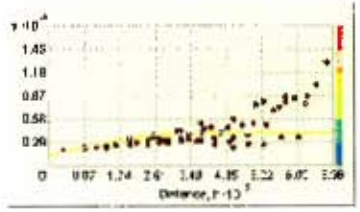


Figure 6: Range spherical model: C0=1.2443, C1= 2.9048, R= 350000

4.1.1. Estimation of α and β based on Data

Parameters for α and β (Table 1) were estimated as follows:

i) *generate* the delta distance and/or squared distance between interval ranges and obtain the mean and standard deviation,

ii) use the mean and standard deviation parameters, to generate a normal random sample, scale it to interval $(0, 1)$, obtain mean, variance and mode which uniquely defines a beta density.

iii) alternatively we used the BestFit 4.5 [7] program to fit a Beta distribution to the scaled data specifying the minimum, and the maximum values.

Table 1: Beta and Alpha parameters.

Distance	Alpha	Beta	d_G-metric $\frac{\alpha}{\alpha+\beta}$
Range Distance (squared)	2.5428	4.2095	0.3766
Delta metric	2.5619	2.5898	0.4073
Range Distance (Absolute)	3.0175	3.2293	0.4834

Then, midpoint and range models were substituted into (11), and the system solved the using a C++ implementation of Kaucher interval arithmetic [8]. We checked the validity of the method using cross-validation for the 200 samples. The cross validation statistics (Table 2) are interval valued with low mean absolute error (MAE) and root mean square error (RMSE), almost equal indicating a better fit. Mean prediction error indicates an equal spread from zero. Prediction at the 808 unknown locations was done and a sample of 100 kriged interval values is shown in figure 7. Note that, the true values (star) are within the bounds of kriged intervals except for some cases.

Table 2: Cross-validation statistics

Statistics	Lower Bound	Upper bound
Mean Observed	90.160	99.834
Mean Predicted	88.437	101.30724
Mean Predicted Error	-11.597	11.137
MAE	2.213	21.089
RMSE	5.666	22.175

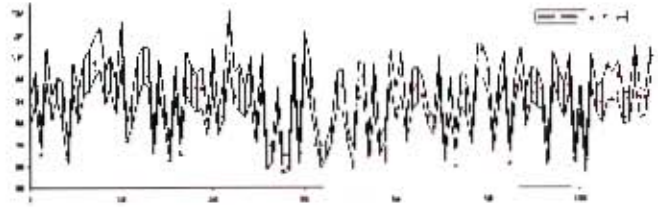


Figure 7: Kriged bounds for a sample of 808 prediction locations. The provided true values are indicated by star and mostly lies within the bounds

5. SUMMARY

With increasingly diverse and abundant data collecting devices. Large amounts of data are collected everywhere on earth and at any scale. In addition, this huge amount of information is of highly variable quality and groups various data types. Hence interval data is a natural occurrence in spatial processes. In this paper, we have explored the methods of analysing interval-valued spatial data by extending the most commonly used L_2- metric into a generalised metric that assigns different weights to values within an interval. This implicitly reflects on the measure of interval uncertainty. Using the distance metric, we have been able to define the corresponding variances and covariances resulting to a kriging system analogous to the traditional kriging system.

Acknowledgements: The research is part of a PhD project on analysis of uncertain spatial data. The author is grateful to the advisers for their valuable suggestions and comments.

REFERENCES

[1] N. A. C. Cressie, *Statistics for Spatial data*, John Wiley and Sons, New York, (1993).

[2] P. Diamond, Interval-Valued Random Functions and the Kriging of Intervals, *Mathematical Geology*, (20), 145–165, (1988).

[3] G. Dubois, Automatic mapping algorithms for routine and emergency monitoring data, *http://www.ai-geostats.org/events/sic2004/sic2004-data.htm*, EUR 21595 EN

[4] D. Pierpaolo and G. Tommaso Gastaldi, A least-squares approach to fuzzy linear regression analysis, *Computational Statistics and Data Analysis*, (34), 427–440, (2000).

[5] E. H. Isaaks and R.M. Srivastava *An Introduction to Applied Geostatistics*, Oxford University Press, New York, (1989).

[6] A. Lakeyev, Linear algebraic equations in Kaucher arithmetic, *Extended Abstracts, A supplement to the international Journal of Reliable computing*, (1995).

[7] BestFit. Version 4.5.1, Palisade Corporation, Newfield, New York, (2004).

P.3 Spatial Interval Kernel Regression

→ N. Wabiri, R. Guo and C. Thiart. **Spatial Interval Kernel Regression,** *Proceedings of 10th International Conference on Industrial Engineering, Theory, Applications and Practice*, Editors: TK Fredericks, SE Butt, A. Pennathur, JE Fernandez, A. Mital. Clearwater, Florida, USA, December 4-6,2005.(ISBN: 0-9654506-1-9).

Proceedings of the 10th Annual
International Conference on Industrial
Engineering-Theory, Applications and Practice
Clearwater, Florida, USA
December 4-7, 2005

SPATIAL INTERVAL KERNEL REGRESSION

N. Wabiri, R. Guo and C. Thiart

Department of Statistical Sciences
University of Cape Town
Private Bag, Rhodes' Gift Rondebosch 7701
Cape Town South Africa
Corresponding author's e-mail: nwabiri@stats.uct.ac.za.

Abstract: In generalized linear regression framework, Kriging is a linear interpolation technique, applied to precise spatial data, with the additional assumptions of stationarity. The assumption of observing a single value at a location and stationarity are not always realistic. We propose a nonparametric kernel regression technique that can handle interval data and relax the assumption of stationarity. The approach proceeds first by mapping spatial random intervals into a spatial-interval feature space. Second, a spatial weighted radial interval kernel is defined, with corresponding spatial and interval bandwidth. Spatial bandwidth extends to the Kth Nearest neighbor (KNN) observation and is determined using cross-validation. Within the given spatial neighbourhood, interval weights based on an interval metric fine-tunes the spatial weights, reducing the effect of dissimilar neighbor values. Estimate at any location is obtained iteratively by gradient descent minimization of a spatially weighted interval kernel. The method is applied on a small dataset.

1. INTRODUCTION

Spatial observations are often not observed as a single value, but rather as an interval observation e.g.: the pollutant concentrations of SO_2 and CO_2. In the traditional methods of interpolation, e.g. Kriging, a further restriction is the assumption of stationarity. Diamond (1988) extended ordinary kriging to interval kriging. Wabiri, et.al.(2005) further improved on Diamond work, using a general weighted interval metric, under the assumption of second order stationarity. The method assumes a known distributional form of spatial random intervals.

We propose a nonparametric kernel based regression method that relaxes the first and second order stationarity assumptions. The approach utilizes the local structure of the data to discover the appropriate (nonparametric) regression functional form of the interval spatial process.

2. SPATIAL INTERVAL KERNEL MODEL

Let the pairs $\{x_i, Z(x_i)\}_{i=1}^n$, with $Z(x_i)=\left[Z^l(x_i), Z^u(x_i)\right]$ the observed spatial random interval at location x_i, $i=1...n$ in a 2D Euclidean space $\Re^2$ represent a spatial interval random process:

$$Z(x_i)=m(x_i)+e(x_i) \tag{1}$$

where $m(x_i)$, is the nonparametric regression function at location x_i and $e(x_i)$ the error term assumed to be independent and homogeneous, (Páez 2004). For ease of notation, let $Z(x_i)=Z_i$ and $m(x_i)=m_i$. The function m_i is localized by observations in the neighborhood of the location x_i defined by a kernel function with an adaptive bandwidth. Possible choices of kernel functions are the positive symmetrical distance decay functions that weights observations, relative to their geographical distance from the regression point. Nearby observations are given more weight than those farther away.

Proceedings of the 10th Annual
International Conference on Industrial
Engineering-Theory, Applications and Practice
Clearwater, Florida, USA
December 4-7, 2005

2.1 Model Formulation

Under global stationarity, a smooth estimate of m_i is the global mean, $\bar{m}=\sum_{i=1}^{n} Z_i \Big/ n$ which minimizes the mean prediction error (MPE), $E(m)=\sum_i (\bar{m}-Z_i)^2$. However, due to spatial heterogeneity; the regression function can not be modeled with a constant value. To allow for high density differences in the process more robustness must be built in the model to take account of outliers, errors and any form of discontinues in the process. Rather than minimize the MPE, we minimize the error loss-function (2) to obtain the so called M-smoothers.

$$E(m)=\sum_{i,j=1}^{n} \Phi\left(\left\|\bar{m}_i - Z_j\right\|^2\right) \tag{2}$$

To localize the global loss function (2), we introduce additional weights to define a local neighbourhood. This yields a local spatial weighted loss function defined over spatial and the interval domain. This joint spatial-interval domain is a product of two kernel functions. The first kernel is the traditional geographical weighted regression kernel, W, with KNN bandwidth and the second is an interval kernel, Φ, defined within the spatial KNN window. The interval kernel is based on interval distance between the neighbors and prediction point values.

With the defined spatial weighted interval kernel function, neighbors with greater geographical distance from prediction point as assigned smaller weights; and at a given geographical distance; neighbors with similar sizes to the initial estimate at prediction point are assigned large weights.

Let $m_i=\bar{m}_i$ denote the initial estimate at location (i) with $n(i)$ the neighborhood set. Instead of minimizing (2), we minimize (3) using gradient descent minimization yielding (4)

$$E(m)=\sum_{i=1}^{n}\sum_{j\in n(i)} W\left(\left\|x_i - x_j\right\|^2\right)\Phi\left(\left\|m_i - Z_j\right\|^2\right) \tag{3}$$

and

$$\begin{aligned} m_i^{t+1} = &\left(1-2\nu\sum_{j\in n(i)} W\left(\left\|x_i - x_j\right\|^2\right)\Phi'\left(\left\|m_i^t - Z_j\right\|^2\right)\right)m_i^t \\ &+2\nu\sum_{j\in n(i)} W\left(\left\|x_i - x_j\right\|^2\right)\Phi'\left(\left\|m_i^t - Z_j\right\|^2\right)Z_j \end{aligned} \tag{4}$$

where ν is the step size chosen so that we do not take too big or too small of a step to convergence. Too big of a step will overshoot the function minimum, and too small step will result in a long convergence. From (4) an optimal value of ν is given as

$$\nu = 1\Big/2\sum_{j\in n(i)} W\left(\left\|x_i - x_j\right\|^2\right)\Phi'\left(\left\|m_i^t - Z_j\right\|^2\right) \tag{5}$$

Let $\Phi(\cdot)$ be a kernel whose derivative exists for $m_i \in [0,\infty)$, then $\eta(t^2)=-\Phi'(t^2)$ is a kernel belonging to a class of radial symmetrical kernels. Kernel η is called the shadow of kernel Φ (Cheng 1995). Using the shadow kernel η and equation (5), (4) yields a spatially weighted interval estimator (6), parallel to the Nadaraya-Watson kernel estimator.

$$m_i^{k+2} = \sum_{j \in n(i)} Z_j * \frac{W\left(\left\|x_i - x_j\right\|^2\right)\eta\left(\left\|m_i^k - Z_j\right\|^2\right)}{\sum_{j \in n(i)} W\left(\left\|x_i - x_j\right\|^2\right)\eta\left(\left\|m_i^k - Z_j\right\|^2\right)} \qquad (6)$$

W is a bivariate radial symmetrical spatial kernel , η a bivariate radial symmetrical interval kernel, the shadow kernel of kernel Φ, and $\|x\|$ the Euclidean norm. Solution to (6) is obtained iteratively.

2.1 Choice of Kernel and the Bandwidth

The choice of kernel function is of secondary importance to selection of bandwidth parameter. Scott (1992) indicates that the shape of the kernel does not significantly affect the estimate mean squared error (MSE). By varying the bandwidth, different kernels can be made equivalent in terms of MSE. For our analysis we chose the bi-square and the Epanechikov radial kernels to model the spatial and interval domains. The kernels are based on local adaptive KNN bandwidth.

To define the interval kernel function, we map the spatial random intervals z_i and z_j to vectors $\phi(z_i)$ and $\phi(z_j)$ in the spatial interval domain (e.g. Hilbert space). Then, the distance between intervals in the input space corresponds to the Hilbert distance $D(z_i, z_j) = \left\|\phi(z_i) - \phi(z_j)\right\|$ between their images in the feature space. Using Wahiri, et.al. (2005[5]) general interval metric, the bi-square interval kernel at location (i) with kernel bandwidth b_i is defined as

$$\eta\left(\left\|m_i^k - Z_j\right\|^2\right) = \frac{15}{16}\left(1 - \left(d_0^{ij}/b_i\right)^2\right)^2 \qquad (7)$$

Where

$$d_0^{ij} = \sqrt{\left(Z_i^c - Z_j^c\right)^2 + \frac{\alpha}{\alpha+\beta}\left(Z_i^r - Z_j^r\right)^2} \qquad (8)$$

is the interval metric between intervals at sample point (j) and prediction point (i) and b_j the maximum interval distance. α , β are predefined **Beta** density parameters. The bi-square spatial kernel is defined as

$$W\left(\left\|x_i - x_j\right\|^2\right) = \frac{15}{16}\left(1 - \left(d_0^s/h_i\right)^2\right)^2 \qquad (9)$$

with d_0^s the geographical distance between prediction point and neighbors, and h_i the bandwidth parameter.

To select the optimal spatial bandwidth we used K^* nearest neighbor (KNN) cross-validation approach with bi-square spatial kernel. The KNN method is based on the inter point distances of the spatial process, and therefore will reflect the spacing of the points rather than the size of study area or the number of points. Besides, the approach adds flexibility to kernel estimation by allowing the user to vary k depending on the desired smoothing. The KNN cross-validation score is

$$CV = 1/n\sum_{i=1}^{n}\left[Z_i - \hat{Z}_{\neq i}(h_i)\right]^2 \qquad (11)$$

with $\hat{Z}_{\neq i}(h_i)$ the fitted value with observation i omitted from the calibration process.

Proceedings of the 10th Annual
International Conference on Industrial
Engineering-Theory, Applications and Practice
Clearwater, Florida, USA
December 4-7, 2005

3. ILLUSTRATIVE EXAMPLE

We illustrate the methods using data from SIC2004 (Dubois, 2004), (Figure 1). Data consist of natural ambient radioactivity measured in Germany, given in form of gamma dose rates, reported by means of the national automatic monitoring network (IMIS). The data was collected from rectangular region with 1008 monitoring stations with samples at 200 stations. We need to estimate values at the remaining 808 stations. For each of 200 stations, 11 days of measurements were randomly selected within a period of 12 months and the average daily dose rates calculated. From the 11 datasets of daily dose rates, we formed a spatial random interval from the minimum and the maximum values. The data variables are, ID, a unique station identity; x, the easting (Meters); y the northing (Meters); and Z the observed interval-valued mean gamma dose rate in nanoSieverts/hour units

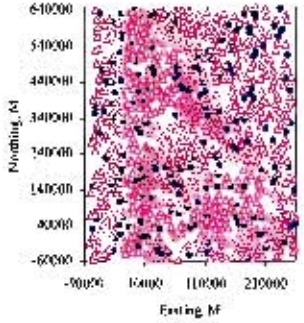


Figure 1. Layout of the dataset: blue indicate the sample, empty triangle the regression points

For each of the spatial random interval we generated the center of mass (center) and radius values. The center of mass vector contains most of the information and was used to perform KNN cross-validation using spatial **Bisquare** and Epanechikov kernels. The optimal KNN (Figure 2) using **Bisquare** kernel is 20 neighbors (bandwidth= 64738.06 M). For Epanechikov kernel optimal KNN is 14 neighbors (bandwidth=53775. 3 M).

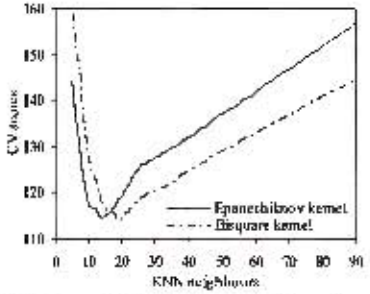


Figure 2. KNN Cross validation for **Bisquare** and Epanechikov kernels

To avoid the boundary problem common in kernel estimation we limit the spatial domain within the interior of observed data. Prior to smoothing out the interval spatial process smoothed out the midpoint to explore possible directional trends. Figure 3.. There is a defined pollution concentrations cluster north-south direction with concentrations more regularly spread is east west direction

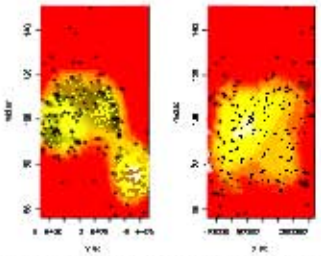


Figure 3. Kernel smoothing of center of mass (mid.sc) in the north–south (y.sc) and east-west (x.sc) direction

To smooth out the interval spatial data, a three step iterative procedure is implemented in C++. The procedure iteratively shifts each location value to the nearby region of local concentration based on 20 neighbors. For locations without initial value a local average is substituted, prior to the iterations.

For each i=1 n

1 Initialize $m_i^k = Z_i$, with initial noisy data or a local mean if there is no initial value ,$x_i^k = x_i, k = 1$

2 Using Bi-square spatial and interval kernels we compute (6), the step size ν (5) and $k = k + 1$
Until either

The interval shift $\left|m_i^{k+1} - m_i^k\right| < \nu$ or

The spatial displacement vector $\left|x_i^{k+1} - x_i^k\right| < \nu$

3 Assign $\hat{Z}_i = m_i^{conv}$ the final estimate where superscript conv denotes the convergence value.

End

Running the procedure for all sample points, each data point is associated with its nearest local minima in the spatial-interval domain. For most points convergence is reached after the first iteration. A plot of a sample of smoothed random intervals is given in Figure 4

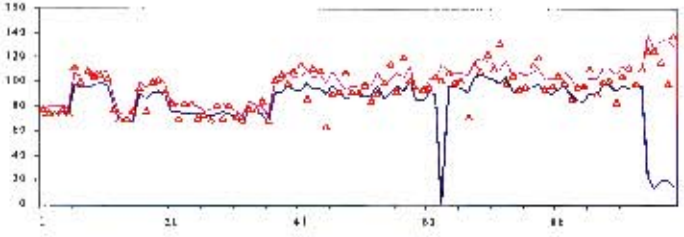


Figure 3. Smoothed random intervals (upper and lower bounds). Triangle are true exact values.

A summary comparison of the prediction errors is compared for general interval kriging and spatial kernel interval regression (Table 1) The nonparametric interval regression produces narrow interval estimate bounds compared to interval kriging

Table 1. Cross-validation statistics for interval kriging and nonparametric interval regression

Interval Kriging			Spatial kernel interval regression (Bisquare kernel)[1] (Epanechnikov)[2]	
Mean Observed	Mean predicted	Mean error	Mean predicted	Mean error
90.169 – 99.834	88.437- 102.307	-11.397- 11.137	90.320–99.969[1] 90.273–99.908[2]	-9.555-9.298[1] -9.562-9.738[2]

4 CONCLUDING REMARKS

With large amounts of varied spatial data obtained at different scales, the challenge to spatial prediction is how to develop and adapt to distribution-free methods that utilizes the local structure of the data to model the true underlying functional form. In this paper, we have m addressed partially the problem by proposing a spatial interval kernel regression method for spatial distributed random intervals. The method based on a spatial–interval domain space fine-tunes the influence of geographical weights by additionally penalizing outliers using a radial interval kernel. With the additional weighting the method performs better relative to kriging.

5. REFERENCES

1. Cressie, N A (1993) Statistics for Spatial data John Wiley and Sons, New York

2. Diamond, P (1988). Interval-Valued Random Functions and the Kriging of Intervals. Mathematical Geology, (20), 145 165

3. Dubois, G. Automatic mapping algorithms for routine and emergency monitoring data http://www.ai-geostats.org/events/sic2004/sic2004-data.htm , EUR 21595 EN

4. Cheng F. (1995). Mean Shift, Mode Seeking, and Clustering. IEEEE TRANSACTION ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. (17), 8, 790-799

5. Páez, A. (2004). Anisotropic Variance Functions in Geographically Weighted Regression Models. Geographical analysis. (36), 4. 145 165,

6. Scott, D. W. (1992) Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons, New York, Chichester

7. Wabiri, N., Gun, R and Thiart, C (2005a) Interval-Spatial Methods, Proceedings of the GIS Research UK 13th Annual Conference. pp121-129.

8. Wabiri, N., Guo, R and Thiart, C (2005b) Interval-Kriging under a Generalized metric, Proceedings of International Workshop on Recent Advances Stochastic Operations Research, pp 290-297. (ISBN 4-939146-03-0)