ABSTRACT

CHAKRABORTY, ARNAB. Contributions to Spatial Statistics and Big Data Analytics. (Under the direction of Soumendra Nath Lahiri).

With the proliferation of sensor-related technologies and advancement of global positioning system in recent years, large-scale weather data can be captured through the microsensors installed in mobile devices. As this information is gathered by various mobile applications, e.g. WeatherSignal, AccuWeather, DarkSky etc., often this type of data is referred as 'crowdsourced' weather data, i.e. information collected through users of various mobile applications. Though these datasets has the potential to be used for weather analysis, especially in 'hyper-local' regions around population centers, due to the amateur quality of the micro-sensors, non-laboratory environement, indoor-outdoor activity of the users and many such reasons, the quality of the observations is compromised. Hence, a robust scalable statistical methodology to analyze these varying-quality spatial data is needed which, along with the variety and velocity of the big-data, incorporates the veracity of the information as well.

In Chapter 2, a method of veracity scoring has been introduced for geostatistical data. In addition, a statistical methodology to analyze noisy spatial data has been proposed which takes the veracity information of the observations into account and thus, provides robust inference and prediction. The advantage of the proposed method has been showcased through simulations and several case-studies involving crowdsourced ambient temperature data over the contiguous USA. Chapter 3 extends the discussion of veracity score based methods in more theoretical details. Large sample properties of the veracity score based parameter estimators has been analyzed and their asymptotic efficiency as compared to the standard method in geostatistics has been established. In Chapter 4, the

attentions has been shifted to anisotropic covariance modeling which is one of the most important research area of interest in spatial statistics. A copula-based method has been proposed that incorporates the directional variogram analysis to construct an admissible joint covariance over the space. Finally, the large sample properties of the covariance parameter estimators have been established and justified through Monte Carlo studies.

Contributions to Spatial Statistics and Big Data Analytics

by
Arnab Chakraborty

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2019

APPROVED BY:

Alyson Wilson	Donald Martin
William Boettcher	Soumendra Nath Lahiri Chair of Advisory Committee

DEDICATION

To Maa, Baba, and Priyanka

for being there for me, always.

BIOGRAPHY

Arnab Chakraborty was born on September 13, 1992 in Chinsurah, West Bengal, India. He completed his secondary education (till 10th standard) in 2008 and his higher secondary education in 2010 from Hooghly Collegiate School, Chinsurah. Having keen interest in mathematical sciences, Arnab joined Indian Statistical Institute (ISI), Kolkata among the top 40 students from all over India for his undergraduate study in Statistics and Mathematics. In 2013, he received B.Stat (Hons.) degree in First Division with *Distinction* from ISI, Kolkata. He continued to pursue his master's degree in Statistics at ISI, Kolkata and graduated with M.Stat with First Division in 2015. Arnab then joined the Department of Statistics at North Carolina State University as a graduate student in 2015. Under the guidance of Dr. Soumendra Nath Lahiri, he successfully earned his Doctor in Philosophy degree in Statistics from NC State University in 2019.

ACKNOWLEDGEMENTS

It has taken me four years to finish this dissertation and it would not have been possible without the support and guidance of a bunch of lovely people, whom I really feel fortunate to have in my life. Though I can never thank them enough for their contributions to my accomplishments, I want to acknowledge their efforts briefly in this section.

First of all, I want to thank my family members for their constant love and support throughout my life. I think I owe all my accomplishments to my parents, Mr. Amit Kumar Chakraborty and Mrs. Lakshmi Chakraborty. Their efforts since my childhood have helped me stay focused towards achieving my goals. Next, I want to mention my lovely girlfriend and my partner in life, Priyanka. Every time I have fallen down, either mentally or physically, even being nearly 9000 miles away from me, she has always been there for me. She has encouraged me through every hurdle of my life in the last three years and has always believed in me, when at times, even I did not believe in myself.

The person I want to thank the most for this dissertation is my advisor Dr. Soumendra Nath Lahiri. I am indebted to him for his care, guidance and encouragement as well as his patience with me. Often I was stuck with doubts about my research and he always clarified those in a way that was best for me to comprehend. His approach of tackling any mathematical or statistical research problem is something that I have always tried to learn and will continue to follow in my future. Apart from his guidance in my research, I am grateful for his teachings in two of the advanced courses that I took in the first two years of my graduate curriculum. It was his way of teaching a very complicated topic fairly simply that made me interested to work in spatial statistics under his supervision. I also want to thank Dr. Alyson Wilson, Dr. Donald Martin and Dr. William Boettcher for agreeing to serve as my advisory committee. Their constructive comments and suggestions have

helped improve this dissertation significantly. Apart from my advisory committee, I want to thank Dr. Eric Laber, Dr. Eric Chi, Dr. Jason Osborne, Dr. Leonard Stefanksi and Dr. Marie Davidian for teaching me important concepts and topics in Statistics during the course of my stay at NC State. I also want to thank our staff at the Department of Statistics, especially Lanakila and Alison, for helping me out with official issues, as well as Terry and Chris for being patient with my computational issues that I have faced during my doctoral degree program.

Before coming to NC State, back in India, I was fortunate to have great teachers and I am indebted to each and every one of them. I want to particularly thank Nimai Kundu, Amitabha Pal and Tapan Kumar, whose teachings in my high school days helped me grow an interest to pursue Mathematics and Statistics in my higher studies. I am and forever will be immensely grateful to Padma Shri Prof. Jayanta K. Ghosh for his care, guidance and encouragement during my summer internship at Department of Statistics in Purdue University. I also want to thank Dr. Sourabh Bhattacharya, Dr. Arijit Chakraborty and Dr. Ayan Basu for their teachings that have helped me get a rigorous training in Statistics during my bachelor's and master's degree program.

In these four years of my life I have come across a bunch of people who has made my stay at Raleigh smooth and enjoyable. First, I want to thank Dr. Indranil Sahoo (Sahoo Da), Dr. Arnab Hazra (Hazra Da), Salil Koner and Dhrubajyoti Ghosh (DjG) – who are probably the best roommates to live with. I think the things that I have enjoyed the most during these four years are things like watching movies or TV series with them, long discussions or debates with them on all types of topics, arranging our regional festivals and departmental events together etc. Their companionship and encouragement have made me get through many tight situations over the years and never made me feel that I am living away from my family. I want to thank Suman (Su-The-Man), Rahul

(Ghoshal), Rahul Chak, Pulama Di, Moumita Di, Debraj Da, Priyam Da (Ala Da), Souvik Da for making my life more enjoyable in Raleigh. Special thanks goes to Jim, Michele, Yun Hee and John for welcoming me to their life, which made me realize I got a home, so far away from home. Not to forget some of my friends from undergraduate and graduate years in ISI, Kolkata: Atanu, Indranil (U), Diganta, Arnab (Chow), Debojyoti Da, Sourav (Chhutku), Sandipan (Mamba), Indrayudh who made the transition of my life from childhood to adulthood smooth and enjoyable.

Finally, I want to thank the Department of Statistics at NC State for giving me the opportunity to pursue a doctoral degree in the topic of my interest and their constant support and guidance throughout my curriculum.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Spatial statistics is a field of science that is concerned with the theory and applications related to spatially indexed physical processes such as ambient temperature, precipitation, humidity, air pollution, soil acidity, soil quality etc. Introduced as a statistical methodology to predict ore reserve in mining engineering in early 1960s, in the last 50 years spatial statistics has emerged as a hybrid discipline of statistics, mathematics, geology and computer science with lots of applications in forestry, agriculture, climatology, meteorology, weather analysis, epidemiology, economics, image processing and many other related fields. In the last two decades, a handful number of works have been added to the literature of spatial statistics, and a significant number of these researches are pertained to meteorology with a major focus on weather analysis. Weather analysis involves structure exploration of the key atmospheric features – for example temperature, wind velocity, air pressure etc. – and prediction of the key weather elements based on the information collected at the meteorological stations or through satellite images. Though prediction of weather conditions in future based on the information in past – which is known as weather forecasting – is more common in literature, spatial prediction of the

weather-related processes has similar importance as temporal prediction because, as mentioned by Cressie (1993), "spatial prediction is just as important as temporal prediction, because people living those cities and rural districts without monitoring stations have the same right to know how little or how much their water or their air is polluted." Most of the spatial prediction of weather-related processes are based on data collected by high-performance sensors at meteorological stations or images captured by high-resolution cameras in satellites. Recently, with the advancement of mobile sensor-related technology, geo-tagged weather information is being collected by micro-sensors installed in mobile devices and gathered by mobile weather applications like AccuWeather, WeatherSignal etc. These datasets are often referred as 'crowdsourced' weather data as the information is coming from the mobile application users. Standard methodologies in geostatistics or spatial statistics is not directly applicable to these mobile sensor-generated data as quality of the observations are often hampered in crowdsourced datatsets due to several factors: the low-quality of the sensors, indoor-outdoor user activity, influence of external and internal processes etc. to name a few. Developing data-driven robust as well as scalable methodologies to analyze noisy spatial data like crowdsourced weather data is broad focus of this dissertation.

In today's big-data scenario, where the data is being collected by automated systems and sensors, in addition to the velocity and variety of the data, veracity of the observations is emerging as one of the most important area of research. Though there are recent works in news, media and communication sciences (for example, see Conroy et al. 2015; Rendon et al. 2018) where veracity of the information is the focus of the research, quality assessment of the observations in spatial statistics is not that common. The standard approach to analyze geostatistical observations involves least squares-based estimation of the regression and covariance parameters which is highly sensitive to the noise in the data.

Cressie (1993) proposed some basic techniques based on explanatory analysis to detect aberrant observations or 'outliers' in spatial data. There are works in spatial statistics literature which incorporate robust estimation of the mean and covariance structure of the process, for example, Cressie and Douglas (1980), Genton (1998), Künsch et al. (2011) etc. But none of these studies discuses any methodology to provide quantitative assessment of the veracity of the spatial observations.

In Chapter 2, we consider a varying-quality crowdsourced weather data, with the aim to improve accuracy of the spatial prediction of the ambient temperature process in 'hyper-local' resolution. In this work, to deal with the noisy nature of the dataset, we introduce a reliability metric, namely Veracity Score (VS), to assess the quality of the crowdsourced observations using a coarser, but high-quality, reference data coming from ground weather stations. In addition, we propose a methodology to analyze noisy spatial data which can incorporate the reliability assessment and thus produce much robust inference and prediction. We refer this approach of analysing geostatistical data as VS-based method. Extensive simulation studies demonstrate the advantage of the VS-based approach over standard practice when some of the observations are associated with high noise. Finally, the merits of the proposed methodology are illustrated through several case studies analyzing crowdsourced daily average ambient temperature readings for one day in the contiguous United States. Additional supplementary materials to this work to support the understanding of the readers have been provided in Chapter 5.

Chapter 3 extends our discussion of veracity scores and VS-based methods to more detailed theoretical analysis. In Chapter 2 – in the analysis of ambient temperature from crowdsourced data – the definition of VS uses a high-quality weather station data as reference which is often not available in practice. Hence, Chapter 3 considers the proposed definition of VS and the VS-based method when there is no reference data available and

provides theoretical justification using spatial asymptotics. In this chapter, we discuss the spatial asymptotic frameworks and assumptions on the underlying spatial process under which the VS-based regression parameter estimators are consistent. Moreover, the mean squared error in the VS-based estimation has been shown to be approximately independent of the high-noise variances associated with the 'bad' observations. But, the accuracy of the VS-based estimation is affected if the proportion of low-quality observations increases. The merits of the VS-based technique have been evaluated through extensive simulation studies and a real-data example involving analysis of coalash (Gomez and Hazen 1970) data.

In Chapter 4, we shift our attention to estimation of the dependence structure of the spatial process. Watson (1972) pointed out that in spatial analysis, it is imperative to consider the small-scale variation of the process which is often included in the model through the second-order structure of the de-trended process. For modeling second-order stationary covariance structure often it is assumed that the underlying process is *isotropic* in nature. This is assumption is restrictive in real-data scenario and misspecification of this assumption may lead to wrong inference and prediction. In practice, presence of anisotropy is ensured through directional variogram analysis: if the marginal variograms along finitely-many different directions behave differently then it is assumed that the underlying covariance structure is anisotropic. But, the directional variograms do not straightforwardly provide an admissible joint covariance over the space. Hence, modeling of anisotropic covariances are mainly restricted to geometric or zonal anisotropy models at the exception of the non-geometric range anisotropy by Ecker and Gelfand (2003), nested modeling by Eriksson and Siska (2000), general range anisotropy model by Allard et al. (2016) etc. But, none of these methods incorporates the directional variogram fits used to confirm the presence of anisotropy prior model fitting. In Chapter 4, we

have introduced a copula-based approach to combine the directional variograms for two specified directions to construct a valid anisotropic covariance over $\mathbb{R}^2$. How marginally fitted variograms for more than two directions can be incorporated to estimate the joint anisotropic covariance is discussed and the consistency and asymptotic normality of the covariance parameter estimators have been established. Finally, the proposed method has been illustrated through several examples and evaluated through simulation studies.

Chapter 2

A Statistical Analysis of Noisy Crowdsourced Weather Data

2.1 Introduction

In recent years there has been a proliferation of weather-related applications for mobile devices such as cellphones, iPods, and laptops. These applications not only provide service to the user but also collect and share spatial data on location, ambient temperature, barometric pressure, humidity, etc., captured by the small-scale sensors installed in the devices. Analyzing and understanding these crowdsourced data sets is becoming an area of increasing interest.

One use of the mobile sensor-generated data is to analyze and understand atmospheric processes at very fine spatial resolution. Most of the methodologies in literature for spatial prediction of weather elements are based on global images coming from satellites or measurements taken at meteorological stations on the ground (for example, see Thornton et al. 1997; Florio et al. 2004 etc.). But none of these sources are dense enough so that

the variability of the process can be analyzed in hyper-local regions. For instance, the ground-stations are generally situated away from localities e.g. at airports or national parks etc. Hence, weather-related analysis solely based on ground-station data does not often provide correct assessment of the variation of the underlying process in the localities. However, in disaster detection, traffic management, and many defense-related activities, prediction of the process in a very localized region (hyper-local) is often more important than the global imputation of the process over a bigger region. Crowdsourced data captured by mobile sensors can serve as a potential source in these scenarios especially in regions where the ground weather stations are sparse but the population density and hence the density of the mobile-devices like cellphones, iPads etc. is relatively high. In a recent article, Sosko and Dalyot (2017) have used a crowdsourced mobile-sensor data in forest fire detection to densify the static geo-sensor network (SGN), which is primarily comprised of meteorological stations with high-performance sensors. In this work, we show that more efficient and reasonable prediction surfaces can be created in hyper-local regions with denser but noisy crowdsourced data as compared to a global prediction surface obtained from high-quality but coarser ground-station data.

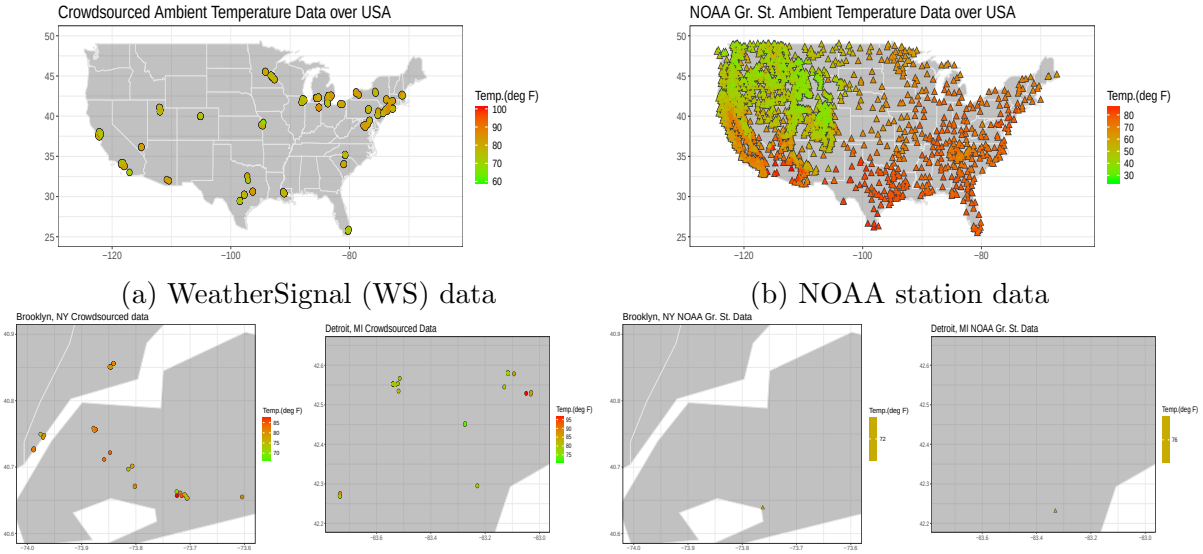
2.1.1 WeatherSignal and NOAA ground-station data

We analyze a static crowdsourced data set consisting of geo-coded daily average ambient temperature readings over the continental United States on April 30, 2013. These data were gathered by a cellphone application named `WeatherSignal`, available both for iOS and Android. In addition to providing information on current weather and forecasts, the app also gathers geographic and weather information using cellphone sensors, leading to a huge amount of crowdsourced spatial weather data from all over the globe. The

`WeatherSignal` application is operated by an organization named OpenSignal. Through the research partnership program of OpenSignal, we were provided real-time (in milliseconds) ambient temperature readings captured by various mobile phones for the above-mentioned day. For each spatial location, we have temporally aggregated the temperature readings to the daily average by taking mean of the regionally estimated hourly temperatures throughout the day. The details of the aggregation are explained elaborately in Section 5.1.1.1 in Chapter 5. After the aggregation, we have the crowdsourced daily average temperature readings at 1879 spatial locations in the United States, as shown in Figure 2.1a. From the figure, it can be seen that the crowdsourced observations are clumped together in high-population density regions like Detroit, Chicago, New York, and Los Angeles etc. In Figure 2.1c and 2.1d we show hyper-local versions of the WeatherSignal data for two nearly square regions at Brooklyn, NY and Detroit, MI.

Along with the crowdsourced data from the WeatherSignal app, we also have ground-station data on the daily average ambient temperature from the National Oceanic and Atmospheric Administration (NOAA). We used the Global Historical Climate Network Daily (GHCND) data access tool to retrieve the daily ambient temperature summaries for April 30, 2013 from 2094 stations in the continental United States. We have plotted the ground-station observations in Figure 2.1b.

Comparing Figure 2.1a and Figure 2.1b, we can see that the NOAA ground-station data provides much more spatial coverage than the crowdsourced data in the entire United States or large parts of United States like east-coast, mid-west etc. are considered and hence for global modeling or building a global prediction surface of the ambient temperature, the ground-station data is clearly a better choice. However, for hyper-local prediction of the spatial process, we believe that crowdsourced data has the potential to capture the local behavior of the spatial process more accurately. For example, in



(a) WeatherSignal (WS) data (b) NOAA station data

(c) WS, Brooklyn (d) WS, Detroit (e) NOAA, Brooklyn (f) NOAA, Detroit

Figure 2.1: Spatial plots of the crowdsourced and NOAA ground-station data. (c) - (f) show zoomed hyper-local versions of the crowdsourced WeatherSignal (c - d) and NOAA station data (e - f).

Figure 2.1e and 2.1f we have plotted the available ground-station observations in the same square neighborhoods as the crowdsourced data in Figure 2.1c and 2.1d. In the area around Brooklyn, NY, there are approximately 90 crowdsourced observations available, where as the number of ground-station observations is only one. Motivated by this observation, in this chapter, we propose a method to improve the accuracy of the hyper-local predictions using the available crowdsourced information in addition to the ground-station data over a bigger surrounding.

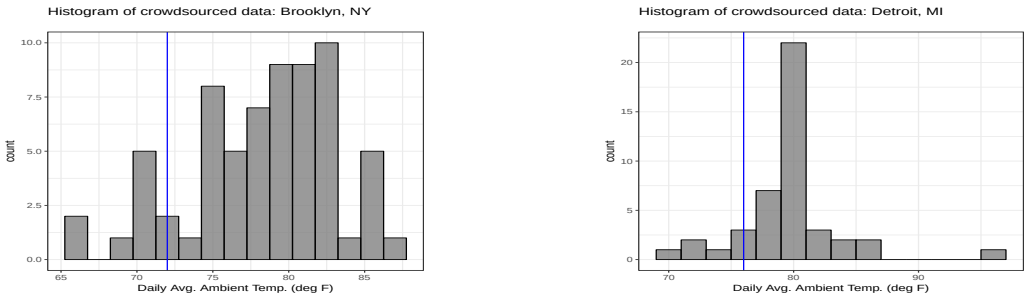


Figure 2.2: Empirical distribution of the crowdsourced average temperatures in the regions from Figure 2.1 for Brooklyn, NY (left) and Detroit, MI (right). Blue vertical lines represent the average ground-station values in the considered regions.

2.1.2 The challenge in analyzing crowdsourced mobile-sensor data

The challenge in analyzing mobile sensor-generated crowdsourced data lies in the low quality and hence poor reliability of an unknown proportion of the data. When data are collected from mobile applications, the readings are prone to contamination for various reasons. The inaccurate observations can occur due external factors, low-resolution sensors, or a combination of these factors. For instance, the temperature readings can be affected by battery temperature, whether the user is inside or outside, the proximity of the device to a hot or cold object, the heterogeneity of the sensors used by different devices, and many other unknown processes.

To illustrate the varying quality of the observations in the WeatherSignal data, Figure 2.2 shows the temperature distribution for the two hyper-local regions shown in Figure 2.1c and 2.1d. The regions are rectangular blocks with the larger side less than or equal to 25 miles. The daily average temperature values in the crowdsourced data set vary from 65°F to 90°F in the Brooklyn region, and from 70°F to 100°F in the region near Detroit, MI for the same day.

These temperature distributions show the nature of the noise involved in the crowdsourced data. Due to the factors associated with the data collection process, a portion of the observations in the crowdsourced data are either contaminated or not representative of the ambient temperature, which is the outdoor air temperature close to the earth's surface. Comparing the histograms with the single ground-station observation in both the regions, we can see that although there are large deviations, a good proportion of the crowdsourced observations are 'close' to the corresponding ground-station observations (72°F in the Brooklyn and 76°F in Detroit), which are collected in laboratory environment with high-quality sensors maintaining World Meteorological Organization (WMO) standards.

Building models based on the noisy crowdsourced data that ignore the reliability of the sensor-generated observations can lead to erroneous prediction. For instance, we used leave-one-out prediction of the observations in the regional block around Brooklyn (Figure 2.1c) using standard techniques of spatial analysis, with a reasonable mean and covariance model (discussed in Section 2.3.1), and the errors in the predictions ranged from -30°F to 40°F. These first-stage analyses motivated us to take the quality of the observations in the WeatherSignal data into consideration. Though "absurd" observations can be identified using existing spatial outlier detection techniques (for example, see Chapter 1 of Cressie 1993; Harris et al. 2014 etc.) and can be omitted from the analysis, it is not straightforward to address observations with small to moderate measurement errors. For instance, using a too strict threshold on the measurement error may lead to deletion of significant number of observations, resulting in a complete loss of information for specific locations.

The new methodology should address the three following challenges. First, in addition to just identifying high-noise observations, a continuous assessment of the veracity of all

the observations in a geostatistical setting is needed. Second, the definition of veracity should take into account the behavior of the process in the study region so that the "misleading" observations can be detected. Third, the veracity assessment of the observations should be incorporated into the subsequent analysis to allow for robust inference and efficient prediction. Though there are studies (for example, Allahbakhsh et al. 2013) in the literature on quality assessment of crowdsourced data coming from volunteers or paid participants, assessment of sensor-generated data quality is not common. Sosko and Dalyot (2017) mention an elementary root mean squared error approach for accuracy measurement using a reference data set from Israeli Meteorological Stations. However, neither of these papers provide full geostatistical inference and prediction using noisy crowdsourced data.

In this chapter, we make several contributions. First, we introduce a Veracity Score (VS) to measure the reliability of crowdsourced observations using a reference data set. Second, we propose a VS-based methodology to incorporate the veracity assessment into standard spatial analysis so that the effect of noisy and misleading observations is reduced, hence making the estimation and prediction more robust and efficient. Third, we show that using the VS-based technique in hyper-local regions with relatively higher number of crowdsourced observations can produce a more accurate and efficient prediction surface as compared to the global prediction surface obtained through the analysis of ground-station data alone. This chapter is organized as follows. In Section 2.2, we introduce the veracity score and describe its elementary properties in a relevant geostatistical setting. Section 2.3 includes a brief description of the standard approach for analyzing geostatistical data, followed by a detailed description of the VS-based methodology for estimation and prediction. In Section 2.4, we describe simulation studies to justify the superiority of VS-based methodology over the standard approach in the analysis of noisy

crowdsourced data. In Section 2.5, we provide details of the analysis, estimation and hyper-local prediction in a case study. Finally, Section 2.6 summarizes our effort and discusses limitations and possible future works.

2.2 Defining and Measuring Veracity

In this section, we provide the intuition and motivation for veracity scoring. We the sample size as n, constants by $C, C_1, C_2, \ldots$, and constants except the parameters in the argument by $C(\cdot)$. We denote the volume of a set $A \subset \mathbb{R}^2$ as $|A|$, i.e., the Lebesgue measure of A if it has nonzero volume and the cardinality of A if A is finite.

2.2.1 Motivation for veracity scoring

To provide motivation for veracity scoring, consider a very simple yet practical example.

Example 2.1. Let $Z_1, \ldots, Z_n$ be independent noisy observations with $E(Z_i) = \mu$ and $\mathrm{Var}(Z_i) = \sigma_i^2$ for $i \in \{1, \ldots, n\}$. The usual sample mean, which is also the o.l.s. estimator for μ, is given by $\hat{\mu}_{\mathrm{ols}} = \bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i$, with $E(\hat{\mu}_{\mathrm{ols}}) = \mu$ and $\mathrm{Var}(\hat{\mu}_{\mathrm{ols}}) = \frac{1}{n} \sum_i^n \sigma_i^2$. If we assume $\sigma_i^2 = C \cdot i^b$ we have

$$\mathrm{Var}(\hat{\mu}_{\mathrm{ols}}) = C(b) \cdot n^b,$$

for some constant $C(b)$. Instead of the generic sample, consider a weighted average of the observations given by $\hat{\mu} = (\sum_{i=1}^n v_i Z_i)/(\sum_{i=1}^n v_i)$, where the weights $v_i = i^{-a}$, i.e. are inversely proportional to the variance of the noisy observations. Then

$$\mathrm{Var}(\hat{\mu}) = C(a, b) \cdot n^{b-1},$$

for some constant $C(a, b)$. A significant gain in efficiency can be achieved by assigning lower weights to high variance observations.

If we can find a formulation of the veracity score that is inversely related to the observation noise variance, we can use it to reduce the effect of the noise in the inference and achieve a more accurate and efficient estimator.

2.2.2 Preliminaries

Let $\{Z(\mathbf{s}_1), \dots Z(\mathbf{s}_n)\}$ be the varying-quality observations – for example, the crowd-sourced data from cellphone sensors – which are observed at irregularly spaced locations $\mathcal{S}_n \coloneqq \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset \mathbb{R}^2$. In addition, at spatial locations $\mathcal{T}_m \coloneqq \{\mathbf{t}_1, \dots, \mathbf{t}_m\} \subset \mathbb{R}^2$, assume that we have $\{Y(\mathbf{t}_1), \dots, Y(\mathbf{t}_m)\}$, which are high-quality, reliable observations of the spatial process – for example, measurements from the ground-stations. It is common to assume (Cressie 1993, Gelfand et al. 2010) that the spatial random field of interest $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ can be represented as

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \tag{2.2.1}$$

where $\mu(\mathbf{s})$ is a deterministic smooth mean function capturing the large scale variation of the process, i.e., $E(Y(\mathbf{s})) = \mu(\mathbf{s})$. Here, $\epsilon(\mathbf{s})$ is a mean zero spatially correlated residual process which addresses the small-scale variations over the space. For the varying-quality Z-process, we write the decomposition in Equation 2.2.1 as

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s}), \tag{2.2.2}$$

where $w(\mathbf{s})$ is the aggregated noise associated with the observation $Z(\mathbf{s})$. For example, if we assume that the varying-quality observations arise from an additive-multiplicative noise model as

$$Z(\mathbf{s}_i) = \epsilon_{M_i} Y(\mathbf{s}_i) + \epsilon_{A_i}(\mathbf{s}_i), \tag{2.2.3}$$

then the associated w-process will have the form $w(\mathbf{s}_i) = \epsilon_{M_i}(\mu(\mathbf{s}_i) - 1) + \epsilon_{M_i}\epsilon(\mathbf{s}_i) + \epsilon_{A_i}$. If there is no multiplicative component ϵ_{M_i} in the contamination, then $w(\mathbf{s}_i) = \epsilon(\mathbf{s}_i) + \epsilon_{A_i}$. In the next subsection, we define a score to assess the quality or reliability of the observation $Z(\mathbf{s}_i)$, namely veracity score.

2.2.3 Veracity score: formulation and properties

A good measure of veracity should not only identify "absurd" observations, but also provide a score for each observation on a continuous scale, so that the effect of the "bad" observations can be reduced automatically, making inference robust against the low-quality observations. Our goal is to formulate a continuous scoring procedure to measure the veracity of the observations in two different scenarios. The first scenario assumes a reference data set containing observations with high-quality but low-density in the concerned regions is available. The second scenario assumes that we do not have any high-quality reference information available.

2.2.3.1 Veracity Score with reference data

Consider a hyper-local regional block like those in Figure 2.1c or 2.1d, and denote it by $\mathcal{R} \subset \mathbb{R}^2$. The observation vector with locations inside $\mathcal{R}$ is given as $\mathbf{Z} \coloneqq (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'$. Consider $\mathcal{R}$ to be the region of interest for analyzing the varying-quality observations $\mathbf{Z}$.

Consider another regional block $\mathcal{D}$ such that $\mathcal{R} \subset \mathcal{D} \subset \mathbb{R}^2$ and $|\mathcal{R}| << |\mathcal{D}|$. Let the reference data vector with locations inside $\mathcal{D}$ be denoted as $\mathbf{Y} \coloneqq (Y(\mathbf{t}_1), \ldots, Y(\mathbf{t}_m))'$. The reference data $\mathbf{Y}$ is high-quality and hence reliable representation of the spatial process of interest, but it has low data-coverage in the hyper-local region of interest $\mathcal{R}$. So, to get a reasonable sample size for the reference data, we need to consider the larger region $\mathcal{D}$. We denote a δ-neighborhood around a spatial point $\mathbf{s} \in \mathbb{R}^2$ as $\mathcal{B}_\delta(\mathbf{s})$, with $\mathcal{B}_\delta(\mathbf{s}) \coloneqq (\mathbf{s}-\delta, \mathbf{s}+\delta]$ for some $\delta \in \mathbb{R}^+$, where the subtraction and addition is component-wise.

Define the VS of the observation $Z(\mathbf{s}_i)$ as

$$V(\mathbf{s}_i) = \phi\left(\frac{|Z(\mathbf{s}_i) - \xi(\mathbf{s}_i)|}{\alpha + D\left(\boldsymbol{\xi}_i\right)}\right), \tag{2.2.4}$$

where $\phi : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+ \cup \{0\}$ is some non-increasing function such that $\sup_x \phi(x) < \infty$. We call $\phi(\cdot)$ the veracity function with $\alpha \in \mathbb{R}^+$ as a regularity parameter. By $\xi(\mathbf{s}_i)$ we denote a reasonable benchmark for the target process at $\mathbf{s}_i$, and $\boldsymbol{\xi}_i \coloneqq \left(\xi(\mathbf{s}_{i_1}), \ldots, \xi(\mathbf{s}_{i_{n(i)}})\right)'$ where $\left\{\mathbf{s}_{i_1}, \ldots, \mathbf{s}_{i_{n(i)}}\right\}$ is the set of observation locations in the small δ-neighborhood $\mathcal{B}_\delta(\mathbf{s}_i)$. Finally, $D(\mathbf{x})$ denotes a robust measure of dispersion of the observations in the vector $\mathbf{x}$.

Now consider the benchmark value, $\xi(\mathbf{s})$, for the target at location $\mathbf{s}$. If we have high-quality observations of the Y-process from the reference data at the varying-quality data sites $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$, then the obvious choice is to take $\xi(\mathbf{s}_i) = Y(\mathbf{s}_i)$. In practice, as we see in Figure 2.1c to 2.1f, the locations of the ground-station measurements (reference data) and the crowdsourced data (varying-quality observations) almost always differ significantly. Hence to define the benchmark at location $\mathbf{s}_i$, we propose to compute a kriging surface, $\left\{\mathbf{s}, \hat{Y}(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\right\}$, of the Y-process using the observation vector $\mathbf{Y}$. Then, we define $\xi(\mathbf{s}_i)$

as

$$\xi(\mathbf{s}_i) = \hat{Y}(\mathbf{s}_i) + (1-\nu)\,\mathcal{C}\left(\mathbf{Z}_i - \hat{\mathbf{Y}}_i\right), \tag{2.2.5}$$

where $\mathbf{Z}_i \coloneqq \left(Z(\mathbf{s}_{i_1}), \ldots, Z(\mathbf{s}_{i_{n(i)}})\right)'$ and $\hat{\mathbf{Y}}_i \coloneqq \left(\hat{Y}(\mathbf{s}_{i_1}), \ldots, \hat{Y}(\mathbf{s}_{i_{n(i)}})\right)'$. Here $\mathcal{C}(\mathbf{x})$ is a robust measure of central tendency of the values in the vector $\mathbf{x}$ and $\nu \in [0,1]$ is a mixing parameter that we discuss in detail later.

If we have a reasonable benchmark, $\xi(\mathbf{s}_i)$, for the spatial process of interest at the location $\mathbf{s}_i$, the definition of the VS in Equation 2.2.4 is a transformed measure of the scaled deviation of the observation $Z(\mathbf{s}_i)$ from the benchmark value. In the definition of VS, the measure of dispersion, $D(\boldsymbol{\xi}_i)$, in the denominator takes the variability in the δ-neighborhood into account. For example, in the analysis of ambient temperature, the variation in a small neighborhood in the mountains is likely to be higher than an area close to the sea-level. Hence, the statistic $\frac{|Z(\mathbf{S}_i)-\xi(\mathbf{S}_i)|}{\alpha+D(\boldsymbol{\xi}_i)}$ measures the deviation of the observation from its benchmark relative to the local variability. In the following sections, we use interquartile range (i.e. $D(\mathbf{x}) = \text{IQR}(\mathbf{x})$) as the robust measure of dispersion in equation 2.2.4 and the sample median (i.e. $\mathcal{C}(\mathbf{x}) = Q_2(\mathbf{x})$, where Q_j is the j-th sample quartile) as the robust measure of central tendency in equation 2.2.5. There are other robust choices as well, but we use the sample quantile based statistic because it is familiar to the practitioners and easy to interpret. Also, these choices are theoretically justified as the sample quantiles are asymptotically consistent under dependence (Ghosh 1971, Sun and Lahiri 2006). The parameter α determines the baseline of the deviation. For lower values of α we penalize more, and for higher values we allow for a larger deviation from the benchmark. We call α the *baseline deviation* of the VS, and its unit is same as the process of interest, which makes the VS unit free.

We require the veracity function ϕ to have the following properties:

1. $\phi(\cdot)$ is a non-increasing function with bounded range, $\phi(x) \leq \phi(0) < \infty$.

2. $\phi(x) \downarrow 0$ as $x \to \infty$.

With this formulation, lower values of the VS correspond to the low-quality or less reliable observations and high values of the VS correspond to the better quality of the observations. We use $\phi(x) = \exp(-x)$ for our analysis in the subsequent sections. The advantage of this function is that the VS lies naturally in $[0, 1]$, and it penalizes exponentially as the scaled deviation from the benchmark value increases. We discuss other possible choices in Section 5.1.2.1 in Chapter 5.

Now we try to interpret the mixing parameter ν in the definition of VS. Under the assumption that the estimated mean process $\hat{\mu}(\mathbf{s})$ is smooth and the kriged-residual process $\hat{\epsilon}(\mathbf{s})$ is a spatially correlated second-order stationary mean-zero process, for a small enough $\delta > 0$, we can write $Q_2\left(\hat{\mathbf{Y}}_i\right) \approx \hat{Y}(\mathbf{s}_i)$, as the variation of the kriged process $\hat{Y}(\mathbf{s})$ inside the δ-neighborhood is negligible. Hence, we can approximately rewrite the benchmark as

$$\xi(\mathbf{s}_i) \approx \nu\, \hat{Y}(\mathbf{s}_i) + (1 - \nu)\, Q_2\left(\mathbf{Z}_i\right).$$

Here, to get a possible approximation the spatial process at location $\mathbf{s}_i$, instead of just using the estimated value $\hat{Y}(\mathbf{s}_i)$ from the high-quality reference data over a bigger surrounding, we want to leverage the available varying-quality observations in the hyper-local region. We propose to use a mixture of an approximation of the spatial process coming from the reference data over a bigger region $\mathcal{D}$, i.e. $\hat{Y}(\mathbf{s}_i)$ and a robust local estimate coming from the varying-quality observations in the small δ-neighborhood $\mathcal{B}_\delta(\mathbf{s}_i)$ around the location of interest $\mathbf{s}_i$, i.e. $Q_2\left(\mathbf{Z}_i\right)$. Due to the smooth mean and spatially correlated

residual process, the spatial observations in a "small" neighborhood are likely to behave "similarly." Therefore, it is sensible to use a robust estimate of the central tendency of the varying-quality observations in that small neighborhood as the locally estimated approximation of the spatial process at $\mathbf{s}_i$. The same approach has been used to detect outliers in literature (e.g., Chapter 1 of Cressie (1993) and Papritz (2018b)). We prefer sample median due to its robustness and asymptotic efficiency (see Sen 1968) as an estimator of location compared to the sample mean when there are outliers present in the data. The mixing parameter ν decides the weight of mixing between the estimated process from the reference data and the local approximation from the varying-quality observations. The optimal ν balances the error in estimation from the reference data and the error in the approximation of the spatial process using the sample median in the δ-neighborhood.

2.2.3.2 Veracity Score without reference data

We propose a similar definition of the VS when we do not have any high-quality reference observations available. In this scenario our definition of VS is

$$V(\mathbf{s}_i) = \phi\left(\frac{|Z(\mathbf{s}_i) - \mathcal{C}(\mathbf{Z}_i)|}{\alpha + D(\mathbf{Z}_i)}\right). \tag{2.2.6}$$

The idea behind the definition given in Equation 2.2.6 is similar to that in Section 2.2.3.1. As we do not have information available from a high-quality reference data set, we use only the locally estimated central tendency as the proxy of the target and the local variation in the denominator to take the regional variability into account. Note that the definition of the VS in Equation 2.2.4 approximately equals the VS as given in Equation 2.2.6 if we take $\nu = 0$.

The formulations of the VS, both with and without reference data, depend on δ, which

is a positive scalar equal to half of the length of the neighborhood $\mathcal{B}_{\boldsymbol{\delta}}(\mathbf{s}_i)$ used to estimate the center and dispersion locally. The choice of δ should be such that the δ-neighborhood $\mathcal{B}_{\boldsymbol{\delta}}(\mathbf{s}_i)$ is small as compared to the region of interest $\mathcal{R}$, but at the same time large enough to have sufficient sample size to provide a good assessment of the quality of the observations. To make the formulation of VS as well-defined, we need the number of points in the δ-neighborhood, $n(i)$, larger than 2 for each $i \in \{1, 2, \ldots, n\}$. If we do not have enough data points to compute the measure of dispersion for an observation, we say that the VS is undefined for those observations.

2.3 Veracity Score Methods

Before going to the VS-based version of the spatial analysis, we briefly describe the standard approach of geostatistical analyses.

2.3.1 Review of standard analysis of spatial data

For this section, we use the model specified in Equations 2.2.1 and 2.2.2 as well as the notations stated in Section 2.2.2. In geostatistics, often the smooth deterministic mean process $\{\mu(\cdot)\}$ is modeled under a spatial regression framework where the mean function is assumed to have a *linear* form, $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$, where $\mathbf{x}(\cdot) = (x_1(\cdot), ..., x_p(\cdot))'$ is a p-dimensional deterministic vector process of known covariates and $\boldsymbol{\beta}$ denotes the unknown regression parameter vector. To make the inference feasible from only one replication of the process over the space, some stationarity assumption on the second-order structure of the residual process $\{\epsilon(\mathbf{s})\}$ is required. One of the most commonly used assumptions is that $\{\epsilon(\mathbf{s})\}$ is an intrinsically stationary process with an admissible parametric variogram function $2\gamma(\mathbf{h}; \boldsymbol{\theta}) = \text{Var}\{\epsilon(\mathbf{s}) - \epsilon(\mathbf{s} + \mathbf{h})\}$, where $\boldsymbol{\theta}$ is the covariance parameter of interest.

For now, the description of the analysis is given without taking the noisy nature of the observations into account, so $\{w(\mathbf{s})\}$ is assumed to be identically equal to $\{\epsilon(\mathbf{s})\}$. Since the covariance parameter is unknown, the standard analysis starts with the estimation of the regression parameters in the linear mean model using ordinary least squares (o.l.s.).

$$\hat{\boldsymbol{\beta}}_{\text{ols}} = \underset{\boldsymbol{\beta}}{\text{argmin}} \sum_{i=1}^{n} \{Z(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}\}^2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z},$$

where $\mathbf{X} \coloneqq (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$. Next, the residual vector is computed using the estimated mean model as $\hat{\boldsymbol{\epsilon}} = \mathbf{Z} - X\hat{\boldsymbol{\beta}}_{\text{ols}}$, and these are used to estimate the covariance structure or variogram of the process. In practice, the variogram is estimated in two steps: first, a nonparametric empirical estimator is computed for discrete lags, e.g. the *classical* or *method-of-moments* semivariogram estimator proposed by Matheron (1962); then the covariance parameter $\boldsymbol{\theta}$ is estimated using a least squares-based variogram model fitting techniques (Cressie 1993). Under the assumption that the semivariogram function $\gamma(\mathbf{h})$ has a parametric form, i.e. $\gamma(\mathbf{h}) \equiv \gamma(\mathbf{h}; \boldsymbol{\theta})$ – for example, the Matérn covariance (Gelfand et al. 2010) – the parameter $\boldsymbol{\theta}$ is estimated using weighted least squares (w.l.s.) as

$$\hat{\boldsymbol{\theta}}_{\text{wls}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \sum_{j=1}^{k} w_j \left\{\hat{\gamma}(\mathbf{h}_j) - \gamma(\mathbf{h}_j; \boldsymbol{\theta})\right\}^2, \tag{2.3.1}$$

where $\hat{\gamma}$ is the empirical semivariogram estimator. For the parametric class of variograms, Matérn is a popular choice as it provides a rich class of variograms to choose from (Haskard 2007). Other valid parametric models for variograms can be found in Cressie (1993) and Gneiting (2013).

Once the covariance structure is estimated, one can try to improve the mean parameter estimates using *estimated generalized least squares* (e.g.l.s.) estimator, given by $\hat{\boldsymbol{\beta}}_{\text{egls}} =$

$\left(X'\hat{\Sigma}^{-1}X\right)^{-1}X'\hat{\Sigma}^{-1}\mathbf{Z}$, where $\hat{\Sigma}$ is the estimated variance of $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_n))'$. However, this introduces additional variability due to using the estimated covariance parameters in the mean estimator and is not necessarily more efficient than the o.l.s. estimator.

The most commonly used method to predict the process at new locations is to predict the ϵ-process at the given locations by the *best linear unbiased predictor* (BLUP) given the observed residual vector $\hat{\boldsymbol{\epsilon}}$, also known as *ordinary kriging* estimator (Cressie 1993, p. 122). The standard predictor of $Y(\mathbf{s}_0)$ is

$$\hat{Y}_{\text{std}}(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\hat{\boldsymbol{\beta}}_{\text{ols}} + \hat{\epsilon}_{\text{ok}}(\mathbf{s}_0), \tag{2.3.2}$$

where $\hat{\epsilon}_{\text{ok}}(\mathbf{s}_0)$ is the ordinary kriging predictor for $\epsilon(\mathbf{s}_0)$.

The stationarity assumption on the random error process can be relaxed using regional stationary models (for example Fuentes 2001, Paciorek and Schervish 2006). Another way to fit the spatial regression model to the geostatistical data is using likelihood-based model estimation (Cressie 1993), but that requires additional distributional assumption on the ϵ-process.

The standard approach for estimation and prediction explained is not reliable for analyzing noisy spatial observations, as both the least squares-based mean parameter estimators (Huber and Ronchetti 2009) and the method-of-moments empirical semivariogram estimator are highly sensitive to the noise (Cressie and Douglas 1980) in the data.

2.3.2 Veracity score-based estimation of the mean function

In the standard approach, as described in Section 2.3.1, the regression parameter vector $\boldsymbol{\beta}$ is estimated using the o.l.s. method. For our approach, instead of simple squared error loss, motivated by Ex. 2.1, we propose to minimize a weighted version of the loss function with the veracity scores as the corresponding weights. The VS-based estimator of the mean parameter $\boldsymbol{\beta}$ is given as

$$\hat{\boldsymbol{\beta}}_{\text{vs}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{n} V(\mathbf{s}_i)\mathcal{L}\left(Z(\mathbf{s}_i), \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}\right). \tag{2.3.3}$$

For least squares-based estimators, we have $\mathcal{L}(y, u) = (y - u)^2$, the squared-error loss function. The locally estimated veracity scores lessen the effects of "absurd" observations in the objective function and thus make the estimation of the mean function less sensitive to the noise. The VS-based approach is adaptive to the quality of the observations and thus lessens the impact of outliers in the data. To make the estimation more robust to contamination, one can use any robust loss function instead of squared-error loss in Equation 2.3.3. As an example, consider Huber's loss function (Huber and Ronchetti 2009), $\mathcal{L}(\cdot, \cdot; k)$, given by

$$\mathcal{L}(y, u; k) = \begin{cases} \frac{1}{2}(y-u)^2 & \text{for } |y-u| \leq k, \\ k(|y-u| - \frac{k}{2}), & \text{otherwise.} \end{cases} \tag{2.3.4}$$

We have used an MM-type estimator with a *linear quadratic quadratic* ψ-function for the robust regression as discussed in Koller and Stahel (2011). The advantage of using this estimator is that in addition to penalizing less for high residuals, the parameters associated with the ψ-function can be tuned to improve the asymptotic efficiency for

the estimators. The corresponding optimization to solve Eequation 2.3.3 can be executed using Iterative Re-weighted Least Squares (IRLS) as discussed in Todorov and Filzmoser (2009).

The assessment of goodness of fit for the estimated linear model is essential. The usual Multiple R^2 is not reasonable to use, as the loss function is different from ordinary least squares. Inspired by the pseudo-R^2_{WLS} coined by Willet and Singer (1988), we propose another variant of the coefficient of determination for VS-based regression as

$$R^2_{\text{vs}} = 1 - \frac{\sum_{i=1}^{n} V(\mathbf{s}_i)\mathcal{L}\left(Z(\mathbf{s}_i), \mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}_{\text{vs}}\right)}{\sum_{i=1}^{n} V(\mathbf{s}_i)\mathcal{L}\left(Z(\mathbf{s}_i), \bar{Z}\right)},$$

where $\bar{Z} = n^{-1}\sum_i Z(\mathbf{s}_i)$. The idea behind this measure is that instead of using the squared error loss to compute the total sum of squares and the residual sum of squares, the proposed R^2_{vs} uses the robust loss function to measure the total variability in the data (i.e. $\sum_{i=1}^{n} V(\mathbf{s}_i)\mathcal{L}\left(Z(\mathbf{s}_i), \bar{Z}\right)$) and the variability that is not explained by the model (i.e., $\sum_{i=1}^{n} V(\mathbf{s}_i)\mathcal{L}\left(Z(\mathbf{s}_i), \mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}_{\text{vs}}\right)$). Although we do not provide any theoretical justification, it appears from explanatory analysis with synthetic data and simulations that R^2_{vs} may provide an overly optimistic assessment of the goodness of the fit for the model when the Huber's loss function or MM-type estimation is used.

2.3.3 Veracity score-based estimation of the covariance structure

To explore the second-order structure of the spatial process, we analyze the residuals obtained by de-trending the observations, $\hat{\epsilon}_{\text{vs}}(\mathbf{s}_i) = Z(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}_{\text{vs}}$ for $i \in \{1, 2, ..., n\}$. When conducting analysis with varying-quality geostatistical data, after the robust esti-

mation of the regression parameters, a portion of the residuals are affected by the presence of measurement error in the data, and direct analysis of these residuals can result in misleading and inefficient estimation of the covariance structure. To reduce the noise of the observed residuals, we propose a VS-based modification of residuals using a local smoothing prior estimation of the covariance parameters. When we have a high-quality reference data, we define the VS-based smoothed version of the residuals as

$$\tilde{\epsilon}(\mathbf{s}_i) = V(\mathbf{s}_i)^q \hat{\epsilon}_{\text{vs}}(\mathbf{s}_i) + (1 - V(\mathbf{s}_i)^q) Q_2(\boldsymbol{\xi}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{\text{vs}}), \tag{2.3.5}$$

where $\mathbf{X}_i \coloneqq \left(\mathbf{x}(\mathbf{s}_{i_1}), ..., \mathbf{x}(\mathbf{s}_{i_{n(i)}})\right)'$ is the $n(i) \times p$ matrix of the covariates corresponding to the observations in $\mathcal{B}_{\boldsymbol{\delta}}(\mathbf{s}_i)$.

If we do not have reference data available, then the analogous smoothed version of the residuals is given by

$$\tilde{\epsilon}(\mathbf{s}_i) = V(\mathbf{s}_i)^q \hat{\epsilon}_{\text{vs}}(\mathbf{s}_i) + (1 - V(\mathbf{s}_i)^q) Q_2(\hat{\boldsymbol{\epsilon}}_i), \tag{2.3.6}$$

where $\hat{\boldsymbol{\epsilon}}_i = \left(\hat{\epsilon}_{\text{vs}}(\mathbf{s}_{i_1}), \ldots, \hat{\epsilon}_{\text{vs}}(\mathbf{s}_{i_{n(i)}})\right)'$. Again note that the definition in Equation 2.3.5 approximately simplifies to the one in Equation 2.3.6 if $\nu = 0$.

For poor quality observations, when $V(\mathbf{s}_i)$ is small, the effect of the observed value of the residual $\hat{\boldsymbol{\epsilon}}_{\text{vs}}(\mathbf{s}_i)$ is scaled down by $V(\mathbf{s}_i)^q$ (as $V(\mathbf{s}_i) \in (0, 1]$), and the locally estimated 'benchmark' value of the residual process in the small neighborhood is enforced by $(1 - V(\mathbf{s}_i)^q)$ in Equations 2.3.5 and 2.3.6. Here q is the parameter regulating the degree of the smoothing needed. For instance, $q = 0$ implies no smoothing, and $q = 1$ implies the convex combination of the locally-corrected residual and the observed residual. The effect of VS-based smoothing is illustrated on a synthetic data set in Chapter 5.

We propose to use variogram model fitting with the VS-based smoothed version of the residuals, $\{\tilde{\epsilon}(\mathbf{s}_i)\}_{i=1}^n$, to estimate the covariance parameter $\boldsymbol{\theta}$ robustly. First a generic nonparametric semivariogram is evaluated at discrete lags using the *robust* semivariogram estimator proposed by Cressie and Douglas (1980):

$$\hat{\gamma}_{\text{vs}}(\mathbf{h}_u) = \frac{\left\{\frac{1}{2|N(H_u)|}\sum_{(\mathbf{s}_i,\mathbf{s}_j)\in N(H_u)}|\tilde{\epsilon}(\mathbf{s}_i)-\tilde{\epsilon}(\mathbf{s}_j)|^{\frac{1}{2}}\right\}^4}{0.457+\frac{0.494}{|N(H_u)|}}\,, \qquad \text{for } u\in\{1,\ldots,K\}\,, \tag{2.3.7}$$

where $N(H_u)=\{\mathbf{h}\in\mathcal{H}:\mathbf{h}\in H_u\}$. H_u are small lag classes or *bins* (see p. 34, Gelfand et al. 2010), which are often called *tolerance regions* (see p. 70, Cressie 1993), and these construct a partition of size K of the lag-space $\mathcal{H}=\{\mathbf{s}-\mathbf{s}':\mathbf{s},\mathbf{s}'\in\mathcal{R}\}$. The candidate lag for the tolerance region H_u is denoted by $\mathbf{h}_u$, which is often taken to be the mean of the observed lags in the bin or the centroid of the the class H_u.

The parameters are estimated using method of weighted least squares as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{\text{vs}} &= \underset{\boldsymbol{\theta}}{\text{argmin}}\;\; Q_{\text{wls}}(\boldsymbol{\theta}) \\ &= \underset{\boldsymbol{\theta}}{\text{argmin}}\;\; \sum_{u=1}^{K}\frac{|N(\mathbf{h}_u)|}{\{\gamma(\mathbf{h}_u;\boldsymbol{\theta})\}^2}\{\hat{\gamma}_{\text{vs}}(\mathbf{h}_u)-\gamma(\mathbf{h}_u;\boldsymbol{\theta})\}^2\,,\end{aligned} \tag{2.3.8}$$

where $\gamma(\cdot;\boldsymbol{\theta})$ is some pre-specified parametric admissible semivariogram model, as discussed in Section 2.3.1. Other robust empirical variogram estimators (for example Genton 1998, Lark 2000) can also be used instead of the one proposed by Cressie and Douglas (1980), as given in Equation 2.3.7. Genton (1998) showed that the robustness properties of the empirical semivariogram proposed by Cressie and Douglas (1980) are not enough in the presence of "absurd" outliers in the data. But, due to the VS-based smoothing in the first stage of the covariance estimation, the very large measurement errors have already been addressed and hence, using Cressie and Douglas (1980)'s version of robust

variogram estimator is reasonable here.

2.3.4 Veracity score-based spatial prediction

Often the aim for spatial analysis of geostatistical data is to predict the process at locations of interest or to create a prediction surface over a region of interest. To predict the ϵ-process at a new location $\mathbf{s}_0$, we can use ordinary kriging with the VS-based smoothed residuals $\tilde{\boldsymbol{\epsilon}} = (\tilde{\epsilon}(\mathbf{s}_1), \ldots, \tilde{\epsilon}(\mathbf{s}_n))'$ as

$$\tilde{\epsilon}(\mathbf{s}_0) = \left\{ \boldsymbol{\gamma} + \mathbf{1}\frac{(1 - \mathbf{1}'\Gamma^{-1}\gamma)}{\mathbf{1}'\Gamma^{-1}\mathbf{1}} \right\}' \Gamma^{-1}\tilde{\boldsymbol{\epsilon}}, \tag{2.3.9}$$

where $\boldsymbol{\gamma} = \left(\gamma(\mathbf{s}_0 - \mathbf{s}_1; \hat{\boldsymbol{\theta}}_{\text{vs}}), ..., \gamma(\mathbf{s}_0 - \mathbf{s}_n; \hat{\boldsymbol{\theta}}_{\text{vs}})\right)'$ and $(\Gamma)_{ij} = \gamma(\mathbf{s}_i - \mathbf{s}_j; \hat{\boldsymbol{\theta}}_{\text{vs}})$ (see chapter 3, Cressie 1993). The residual kriging variance, which quantifies the prediction uncertainty, can be estimated as

$$\hat{\text{Var}}\left(\tilde{\epsilon}(\mathbf{s}_0)\right) = \hat{\sigma}^2_{\text{ok}}(\mathbf{s}_0) = \boldsymbol{\gamma}'\Gamma^{-1}\boldsymbol{\gamma} - \frac{\left(\mathbf{1}'\Gamma^{-1}\gamma\right)^2}{\mathbf{1}'\Gamma^{-1}\mathbf{1}}.$$

Finally, we predict the process at $\mathbf{s}_0$ using the modified version of Equation 2.3.2 as,

$$\hat{Y}_{\text{vs}}(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)'\hat{\boldsymbol{\beta}}_{\text{vs}} + \tilde{\epsilon}(\mathbf{s}_0). \tag{2.3.10}$$

In Equation 2.3.10, both the mean and covariance parameters have been robustly estimated using the VS-based procedures. The smoothing parameter q for the VS-based smoothing of the residuals can be chosen from a discrete pre-specified set to maximize the cross-validated prediction accuracy.

There are other robust kriging approaches available in literature, for example, Künsch

et al. (2011) and Papritz (2018a). Both of these techniques require distributional assumption on the ϵ-process. It is not straightforward to determine how to reduce the effects of observations that are not noisy but represent some other spatial process. For example, if in a local region most of the crowdsourced ambient temperatures are captured in indoor settings, applying the robust procedures directly may lead to misleading estimation of the model parameters and hence bad prediction of the outdoor ambient temperature. On the other hand, the VS-based technique can use a benchmark value, possibly obtained from a high-quality but low-density reference data, to reduce the effects of the 'misleading' observations and thus estimate and predict the process of interest robustly and efficiently. We have compared the VS-based technique with the robust-REML (Künsch et al., 2011) through empirical studies in Chapter 5. More extensive theoretical comparison of other robust kriging methodologies with the VS-based technique in case of no available reference data is beyond the scope of this dissertation.

2.4 Simulation Study

Our simulation study aims to justify the superiority of the VS-based estimation and prediction methods as compared to the standard approach for analyzing noisy geostatistical data. We have considered two scenarios here: the first one is when no reference data is available and the second one is when a coarser but better quality reference data is present.

2.4.1 Without reference data

We take the sampling region for the varying-quality observations to be $\mathcal{R} \equiv \mathcal{R}_n := [0, \lambda_n]^2$, where $\{\lambda_n\}_n$ is a sequence of positive real numbers determining the size of the

sampling region. We have assumed that the varying-quality observations, i.e. $\{Z(\mathbf{s}_1), \ldots$ $\ldots, Z(\mathbf{s}_n)\}$ are coming from an additive-multiplicative noise model as given in Equation 2.2.3. To generate the "true" process for simulation purposes, we use the following spatial linear model:

$$Y(\mathbf{s}_i) = \beta_0 + (\beta_x, \beta_y)' \, \mathbf{s}_i + \beta_h \; h(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{2.4.1}$$

where $\boldsymbol{\beta} \coloneqq (\beta_0, \beta_x, \beta_y, \beta_h)'$ is the vector of regression parameters; $h(\mathbf{s})$ is the altitude of the location $\mathbf{s}$; and $\{\epsilon(\mathbf{s})\}$ is a second-order stationary spatially correlated process.

To define the altitude function over the sampling region, we use the deterministic function $h(\mathbf{s}) = H_1 \cdot \sum_{j=1}^{H_2} w_h(j) f(\mathbf{s};\ \boldsymbol{\mu}_j, \Sigma_j) + H_3$, where $f(\cdot; \boldsymbol{\mu}, \Sigma)$ denotes the bivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix Σ and $\{(\boldsymbol{\mu}_j, \Sigma_j) :\ j \in \{1, \ldots, H_2\}\}$ are fixed set of vectors and matrices. The residual vector $(\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_n))'$ are sampled from a second-order stationary mean-zero Gaussian process with isotropic Matérn covariance given by

$$C(d; \boldsymbol{\theta}) = \sigma_\epsilon^2 \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{d}{\rho}\right)^{\kappa} K_\kappa \left(\frac{d}{\rho}\right) + \tau^2 \mathbb{1}(d = 0), \tag{2.4.2}$$

where Γ is the gamma function, K_κ is the modified Bessel function of the third kind with order κ (Abramowitz and Stegun 1972). The covariance parameter vector of interest is $\boldsymbol{\theta} = (\tau^2, \sigma_\epsilon^2, \rho, \kappa)'$, where τ^2 is the nugget effect, $\sigma_\epsilon^2, \rho, \kappa$ are the partial sill, range and smoothness parameters respectively (Haskard 2007, Gelfand et al. 2010).

To generate noise for the varying-quality observations, we use the following model for the additive and multiplicative components, denoted by $\boldsymbol{\epsilon}_A \coloneqq (\epsilon_{A_1}, \ldots, \epsilon_{A_n})'$ and

$\boldsymbol{\epsilon}_M \coloneqq (\epsilon_{M_1}, \ldots, \epsilon_{M_n})'$ respectively:

$$\epsilon_{M_i} \underset{\text{indep.}}{\sim} 2 \times \text{Beta}(\alpha_{M_i},\alpha_{M_i});\ \epsilon_{A_i} \underset{\text{indep.}}{\sim} N(0, \sigma^2_{A_i});\ \{\epsilon_{M_i}\}_{i=1}^{\infty} \text{ and } \{\epsilon_{A_i}\}_{i=1}^{\infty} \text{ are indep., and;}$$

$$\sigma^2_{M_i} = \begin{cases} 0 & \text{if } i \in G_n \\ \sigma^2_M & \text{o.w.} \end{cases} ;\ \sigma^2_{A_i} = \begin{cases} 0 & \text{if } i \in G_n \\ \sigma^2_A & \text{o.w. ,} \end{cases} \tag{2.4.3}$$

where $\sigma^2_{M_i} = \frac{1}{2\alpha_{M_i}+1}$; $G_n \subset \{1, \ldots, n\}$ is a subset of indices and σ_M, σ_A are positive constants. With this model, if $i \in G_n$, we have no noise associated with the observation, i.e., $Z(\mathbf{s}_i) = Y(\mathbf{s}_i)$. If $i \notin G_n$, then $Z(\mathbf{s}_i) = \epsilon_{M_i} Y(\mathbf{s}_i) + \epsilon_{A_i}$, where ϵ_{M_i} and ϵ_{A_i} have positive variance. We further assume that the proportion of "good" observations is a constant q_e, i.e., $|G_n|/n \approx q_e$, and $1 - q_e$ is the proportion of noisy observations in the data. This model is inspired by the crowdsourced data analysis scenario where only a proportion of observations are "bad". The choice of multiplicative error distribution in Equation 2.4.3 restricts its realizations to be in $[0, 2]$ and also ensures that the multiplicative errors are symmetric around 1.

We set $\boldsymbol{\beta} = (55, 1.5, -1, -0.08)'$, $\boldsymbol{\theta} = (0, 6, 0.5, 3)'$. To investigate the robustness of the VS with increasing noise in the data, we consider three contamination models specified by the following parameters: (a) $\sigma_A = 5, \alpha_M = 2, q_e = 0.95$, (b) $\sigma_A = 50, \alpha_M = .5, q_e = 0.9$ and (c) $\sigma_A = 100, \alpha_M = 0.05, q_e = 0.8$. As we go from model (a) to (c), the noise in the data increases both in extent and magnitude. For example, with model (a), the variance of a noisy observation at location $\mathbf{s}$ is $0.2\,(\mathbf{x}(\mathbf{s})'\boldsymbol{\beta})^2 + 28.6$, and the proportion of such observations is 5%; with model (c), the same variance will be $0.91\,(\mathbf{x}(\mathbf{s})'\boldsymbol{\beta})^2 + 10005.73$, and the proportion of noisy observations rises to 20%.

Next we analyze the simulation results to compare the performances of VS-based and

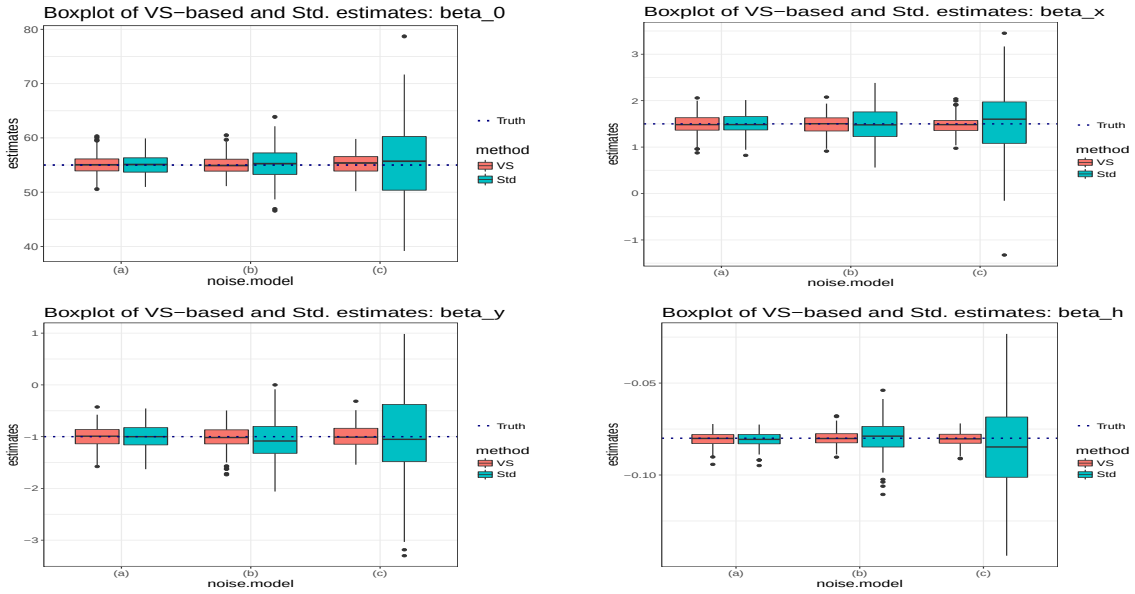


Figure 2.3: Performance of the VS-based and standard regression parameter estimators for analyzing varying-quality observations (sample size $n = 500$) without reference data.

standard approach. The choices of the regularity parameters in the VS-based estimation like the baseline deviation α and the smoothing parameter q are discussed in Chapter 5.

In Figure 2.3 we show boxplots of the VS-based estimator $\hat{\beta}_{\text{vs}}$ and the standard estimator $\hat{\beta}_{\text{ols}}$ for the four regression parameters based on $B = 200$ simulations with $n = 500$ samples. The VS-based technique shows more robustness towards the added noise in the observations. As we move from noise model (a) to (c), the efficiency of the o.l.s. estimator is heavily compromised, where as the spread of the VS-based estimates is hardly increased. Chapter 5 contains additional simulation results for regression parameter estimation including tables with bias and standard errors (s.e.) of the estimates as well as boxplots for $n = 100, 3000$. All of these simulations show similar results to justify the superiority of VS-based mean parameter estimation in the analysis of noisy spatial data as compared to the standard o.l.s method.

We also evaluate the VS-based and standard covariance parameter estimation and

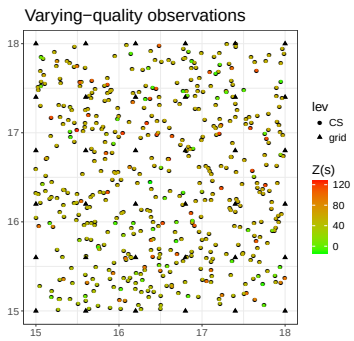
show the results in Table 2.1. In each of the cases, the estimates of the sill parameter

Table 2.1: Performance of the VS-based methodology and standard approach in estimating covariance parameters on varying-quality observations.

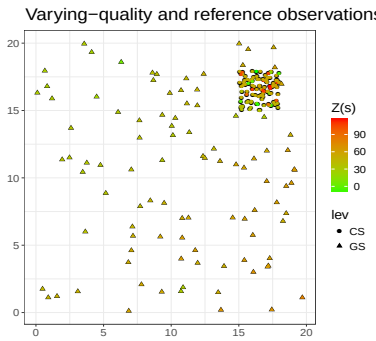
Noise Model	n	bias.sill.VS	bias.sill.Std	bias.range.VS	bias.range.Std
(a)	100	-0.313 (3.31)	3837.513 (9867.28)	-0.296 (0.13)	6.671 (16.1)
	500	0.23 (1.16)	623.629 (1644.56)	-0.114 (0.06)	3.778 (9.91)
	3000	0.344 (0.62)	36.098 (82.01)	-0.026 (0.05)	0.307 (3.2)
(b)	100	7.657 (8.13)	17545.465 (58680)	-0.357 (0.08)	69.945 (454.78)
	500	1.747 (1.52)	5135.181 (14207.51)	-0.158 (0.06)	3.711 (11.05)
	3000	0.48 (0.96)	1108.544 (3515.95)	-0.06 (0.05)	8.377 (52.63)
(c)	100	32.774 (9.51)	6606.713 (27599.4)	-0.39 (0.03)	130.833 (463.05)
	500	15.352 (6.23)	21915.533 (63507.44)	-0.241 (0.05)	6.222 (47.09)
	3000	2.933 (1.14)	5289.832 (12192.3)	-0.111 (0.04)	4.624 (19.92)

($\sigma_\epsilon^2 + \tau^2$, the total variance the residual process) obtained by the VS-based methodology is more accurate by large margins as compared to standard variogram estimation. As the sample size increases both the bias and standard deviations of the VS-based estimators are closing towards 0 under all the considered noise models. Table 2.1 clearly establishes the efficiency of VS-based covariance estimation as compared to the standard approach when some of the observations are corrupted. In Chapter 5, we have provided tables to compare the performance of the VS-based methodology and the robust Gaussian REML method proposed by Künsch et al. (2011) and implemented by Papritz (2018a) in the estimation of the covariance parameters. The results are largely in favor of VS-based methodology. For a fixed n, if we move from noise model (a) to noise model (c) the increase in bias and standard errors of the VS-based sill parameter estimator is prominent, though the magnitude of increment is negligible as compared to the standard method of estimation.

Next we evaluate the VS-based spatial prediction using a $4\lceil\lambda_n\rceil \times 4\lceil\lambda_n\rceil$ grid over the



(a) Varying-quality observations and the grid to validate prediction.



(b) Varying-quality hyper-local observations with reference data.

Figure 2.4: Example sampling points for the simulations.

sampling region $\mathcal{R}$ as shown in Figure 2.4a. We make predictions at these grid points using both the VS-based and standard approach and evaluate the predictions and kriging by the following metrics:

$$\text{RMSPE} = \sqrt{\frac{1}{n}\sum_{\mathbf{S}^*}\left(\hat{Y}_{\text{vs}}(\mathbf{s}^*) - Y(\mathbf{s}^*)\right)^2}\;;\quad \text{ResRMSPE} \;= \sqrt{\frac{1}{n}\sum_{\mathbf{S}^*}\left(\tilde{\epsilon}(\mathbf{s}^*) - \epsilon(\mathbf{s}^*)\right)^2}\,,$$

where the sum $\sum_{\mathbf{S}^*}$ is over the grid points. We define the performance metrics for the standard methods analogously. The Root-Mean-Squared-Prediction-Error (RMSPE) measures the average prediction error over the selected grid; and the Residual-Root-Mean-Squared-Prediction-Error (ResRMSPE) evaluates the accuracy and efficiency of the kriging on the selected grid for the spatially correlated residual process: $\{\epsilon(\mathbf{s})\}$. By Av.RMSPE we denote $\frac{1}{B}\sum_b \text{RMSPE}(b)$ where RMSPE(b) is the prediction error in the b-th simulation iteration. We define Av.ResRMSPE similarly.

Table 2.2 summarizes the results which show that the VS-based predictions are much better than the standard analysis in almost all the cases. As we go from model (a) to model (c) the prediction accuracy has compromised for both the VS-based as well as

Table 2.2: Prediction performance of the VS-based methodology and standard approach on varying-quality observations without any reference data.

		VS		Std. App.	
Noise Model	n	Av.RMSPE	Av.ResRMSPE	Av.RMSPE	Av.ResRMSPE
(a)	100	5.29 (4.04)	0.703 (0.22)	8.61 (15.37)	3.637 (1.82)
	500	4.046 (1.03)	0.281 (0.03)	4.826 (8.89)	4.416 (1.33)
	3000	3.927 (1.07)	0.141 (0.02)	3.228 (1.37)	5.306 (0.48)
(b)	100	9.67 (6.11)	1.796 (0.77)	37.38 (92.72)	14.717 (6.99)
	500	8.478 (5.04)	0.358 (0.07)	28.911 (75.28)	14.267 (7.56)
	3000	5.196 (3)	0.15 (0.02)	20.546 (33.01)	14.902 (8.07)
(c)	100	21.071 (11.29)	5.833 (1.39)	98.585 (206.89)	38.74 (19.03)
	500	26.325 (14.35)	1.376 (1.44)	66.6 (152.04)	36.354 (20.06)
	3000	13.722 (6.55)	0.23 (0.04)	94.429 (193.5)	31.606 (23.25)

the standard approach with much higher impact for the later one. However, in terms of residual kriging efficiency the VS-based methodology is highly robust as compared to the ordinary kriging using the residuals obtained from o.l.s.

2.4.2 With reference data

In this subsection, we consider a situation that is more similar to our case study. In addition to the n varying-quality observations in the hyper-local region $\mathcal{R} = [0, \lambda_n]^2$, we have m-many high-quality observations available over a larger region $\mathcal{D} = [0, \Lambda_m]^2$. One example of the sampling points is shown in Figure 2.4b. Our goal is to predict the process within the hyper-local region $\mathcal{R}$ using the varying-quality observations. We again use a $4\lceil\lambda_n\rceil \times 4\lceil\lambda_n\rceil$ grid over the hyper-local region of interest $\mathcal{R}$ to evaluate the predictions. In addition to the predictions obtained by the VS-based and standard methodology on the varying-quality observations, we also consider the global predictions obtained by using only the reference data on the larger region as shown in Figure 2.4b. For this simulations we have considered the sample sizes for varying-quality observations to be equal to $50, 100$

and 500 because the hyper-local regions in our case studies do not contain very 'large' (not more than 300) number of crowdsourced observations. For the reference data the sample sizes we have considered are: $m = 100, 500$.

We only show the results for evaluating the predictions and kriging under the above-said scenario as that include estimation of both the mean and the covariance parameters. We call the predictions obtained by using the standard methodology on the reference data

Table 2.3: Performance of hyper-local predictions using the VS-based methodology, the standard approach and global predictions using reference data only. For these simulations we used reference data with sample size $m = 100$.

		VS		Std. App.		Ref. Only	
Noise Model	n	Av.RMSPE	Av.ResRMSPE	Av.RMSPE	Av.ResRMSPE	Av.RMSPE	Av.ResRMSPE
(a)	50	12.26 (12.71)	7.084 (5.08)	1740.696 (9518.03)	1745.244 (9604.81)		
	100	10.877 (11.61)	6.104 (5.24)	230.117 (918.91)	224.56 (934.89)		
	500	8.787 (8.05)	6.287 (6.47)	358.694 (1976.29)	352.86 (1975.49)		
(b)	50	12.933 (13.1)	8.206 (6.76)	52829.372 (662485.91)	52946.917 (664727.7)		
	100	9.907 (10.81)	6.439 (5)	115.222 (923.6)	387.071 (915.98)	9.711 (8.54)	9.017 (7.46)
	500	9.005 (8.66)	6.72 (5.06)	26.31 (19.18)	217.784 (15.88)		
(c)	50	12.33 (18.51)	8.7 (16.61)	10198.908 (85831.41)	9740.72 (85082.65)		
	100	10.131 (10.93)	7.093 (5.04)	155.796 (126.08)	412.788 (31.68)		
	500	9.786 (8.45)	6.402 (5.24)	239.728 (29.49)	27.335 (8.35)		

only as 'Ref. Only'. From Table 2.3, we see that for all noise models, the performance of the VS-based predictor in terms of Av.RMSPE using varying-quality observations is similar to the predictor using standard methodology on the reference data only, when we have enough number of varying-quality observations. However, if the standard methodology is implemented on the varying-quality observations, the predictions and kriging at the grid points are absurd in many occurrences resulting in high RMSPE and ResRMSPE. If we consider the kriging performance i.e. the ResRMSPE, the VS-based technique has outperformed even the kriging using high-quality reference data. Clearly, using the VS-based

methodology on the varying-quality observations in a hyper-local region one can improve the kriging efficiency as compared to the global kriging using reference data only. The model we used to generate the true process over the study region is fixed over the space and hence the RMSPE of the hyper-local VS-based analysis is similar or higher by a slight extent than the global analysis using the good-quality reference observations in a bigger region. But in practice, where model misspecification error is involved, VS-based analysis using crowdsourced observations can produce better and efficient predictions at hyper-local regions as compared to the global analysis using high-quality observations over a much bigger region. Additional details regarding the simulation results are reported in Chapter 5.

2.5 Case Study: Spatial Analysis of WeatherSignal Data

In this section, we analyze the WeatherSignal data described in Section 2.1.1 using the VS-based methodology (Section 2.3). Our goal for this noisy crowdsourced data set is to perform structure exploration and then prediction of the daily average ambient temperature process in hyper-local regions of interest.

2.5.1 Building hyper-local prediction surfaces

Here we describe the VS-based analysis of the crowdsourced WeatherSignal data using the NOAA ground-station data as reference. We first select a hyper-local region, as denoted by $\mathcal{R}$ in Section 2.2.3.1, around Los Angeles, CA, as shown in Figure 2.5d. The analysis starts by defining a region large enough to have sufficient NOAA ground-station

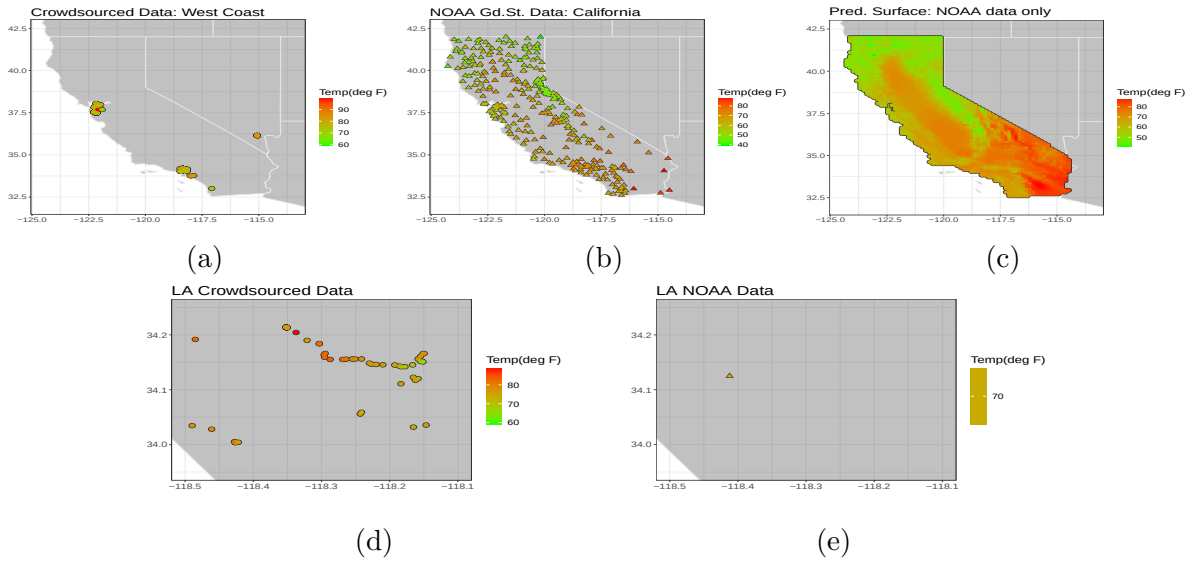


Figure 2.5: (a) Crowdsourced observations in CA; (b) Available ground-station observations; (c) Prediction surface using the standard approach on the ground-station data; (d) Crowdsourced observations in a hyper-local region around Los Angeles; (e) ground-station observations in a hyper-local region around Los Angeles.

observations to build a reasonable global prediction surface around the region of interest. In Figure 2.5b, we plot the $m = 310$ ground-station observations in California. Using the standard approach on the NOAA ground-station data, as described in Section 2.3.1, we build a prediction surface for California and plot it in Figure 2.5c. The model we use to estimate the mean is given by

$$\mu(\mathbf{s}) = \beta_0 + \beta_x \, s_x + \beta_y \, s_y + \beta_{xy} \, s_x s_y + \beta_h \, h(\mathbf{s}), \tag{2.5.1}$$

where $\mathbf{s} := (s_x, s_y)'$ and $h(\mathbf{s})$ denotes the elevation of the point $\mathbf{s}$. The mean model explains 79% (adjusted R^2) of the variability in the ground-station ambient temperatures in California.

We then fit a Matérn covariance to the observed residuals from the mean model

estimation. Details of the variogram estimation are given in Table 2.4 and Figure 2.6. We then use standard kriging methodology with the estimated mean and covariance model to create the prediction surface $\left\{(\mathbf{s}, \hat{Y}(\mathbf{s})) : \mathbf{s} \in \mathcal{D}\right\}$, as shown in Figure 2.5c.

Table 2.4: Estimated Matérn parameters.

Parameters	**Estiamtes**
partial sill (σ^2)	13.78
range (ρ)	0.36
nugget (τ^2)	7.95
smoothness (κ)	2.45

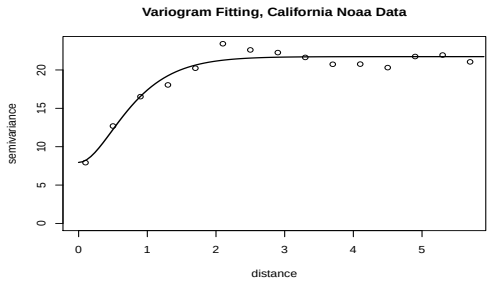
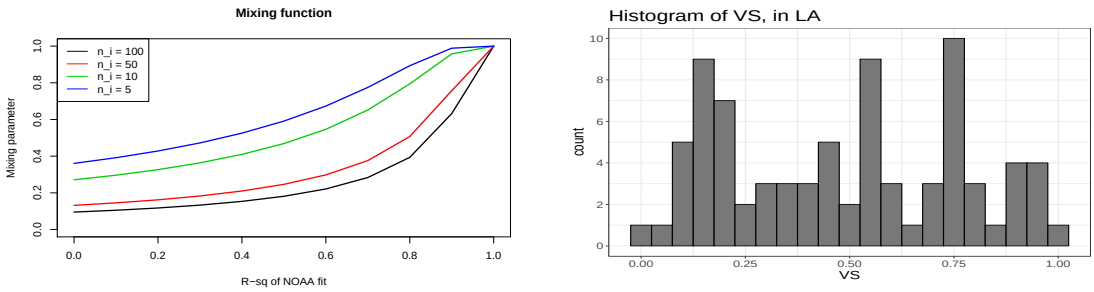


Figure 2.6: Variogram estimation

As we can see in Figure 2.5a, the spatial coverage of the crowdsourced data does not support a global prediction surface over California or even the coast of California. However, if we consider the 25×25 mile region ($\mathcal{R}$) in LA, as shown in Figure 2.5d, the density of crowdsourced data is much higher as compared to only one ground-station observation (Figure 2.5e). While there is only one ground-station available at Los Angeles International Airport, the number of crowdsourced observations, $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$, in $\mathcal{R}$ is $n = 80$.

The next part of the analysis examines whether we can leverage the additional crowdsourced information through the VS-based methodology. We want to explore whether we can create a more reasonable and efficient prediction surface $\left\{(\mathbf{s}, \hat{Y}_{\text{vs}}(\mathbf{s})) : \mathbf{s} \in \mathcal{R}\right\}$ over the region $\mathcal{R}$ in Los Angeles as compared to the surface obtained from the analysis of the ground-station data only, $\left\{(\mathbf{s}, \hat{Y}(\mathbf{s})) : \mathbf{s} \in \mathcal{R}\right\}$.

The VS-based analysis starts by computing the veracity score of the crowdsourced observations using the definition in Equation 2.2.4. We set the baseline deviation $\alpha = 3$. In an ideal scenario, when the corresponding δ-neighborhood has very little variation and $\text{IQR}(\boldsymbol{\xi}_i) \approx 0$, an observation with 3°F deviation from the corresponding benchmark



(a) (b)

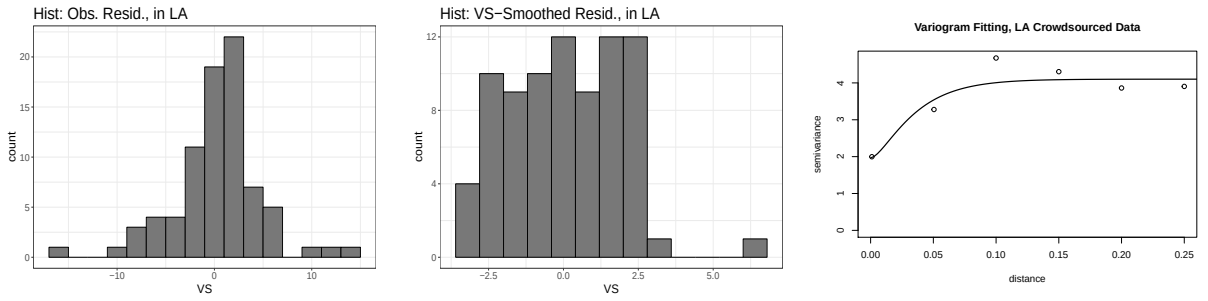
Figure 2.7: Mixing function (a) and the histogram of the veracity scores (b) for the crowdsourced observations in Los Angeles.

value has a VS approximately equal to $\exp(-1) \approx 0.368$, while an observation with a 1°F deviation has a VS ≈ 0.716. To define the neighborhood for computation of the VS, we take $\delta = 0.08$ in the units of latitude and longitude. To choose a suitable mixing parameter ν, we use the function

$$\nu(\mathbf{s}_i) = 1 - \exp\left(\frac{-1}{(1 - \mathrm{R}^2)\sqrt{n(i)}}\right),$$

where R^2 is the adjusted R-squared for the estimation of the mean surface using NOAA ground-station data only and $n(i)$ is the number of crowdsourced data in the δ-neighborhood. As Figure 2.7a shows, this function is increasing in R^2 and decreasing in $n(i)$. $\nu(\mathbf{s}_i) = 1$ if $\mathrm{R}^2 = 1$ and $\nu(\mathbf{s}_i) = 0$ if $n(i) = \infty$. With this formulation, the mixing parameter takes both the goodness of fit for the ground-station data and the number of crowdsourced observations used for local approximation of the target value into account. Using the specified parameters, we compute the VS for the crowdsourced observations in $\mathcal{R}$ and plot their empirical distribution in Figure 2.7b.

We next estimate the mean and covariance of the process. For robust estimation of the mean function, we use the weighted MM-type estimator, as discussed in Section 2.3.2 with



(a) (b) (c)

Figure 2.8: Histograms of the observed residuals (a) and VS-based smoothed residuals (b) and the VS-based variogram fitting (c) for optimal $q = 0.8$.

the VS of the observations as the corresponding weights. Once the regression parameters are estimated, for a given smoothing parameter q in Equation 2.3.5, we use the VS-based smoothing technique to reduce the effects of noise in the residual process as discussed in Section 2.3.3. Using the smoothed residuals, we estimate the covariance parameters and use the estimates to create a prediction surface using VS-based kriging as discussed in Section 2.3.4.

To make an optimal choice for q, we use the reference data. For a pre-specified set of values of $q \in [0.05, 3]$ the covariance estimation and kriging are executed at the ground-station locations that are inside the hyper-local region $\mathcal{R}$, and the q that minimizes the mean squared error of prediction at the stations is chosen to be optimal. In the analysis for the hyper-local region around Los Angeles, there is only one station available, so we use the set of points with VS greater than or equal to 0.8 as test data and minimize leave-one-out cross-validated mean squared prediction error, i.e. $n_*^{-1} \sum_j \left(Z(\mathbf{s}_j) - \hat{Y}_{\text{vs}}^{(-j)}(\mathbf{s}_j) \right)^2$ where $\hat{Y}_{\text{vs}}^{(-j)}(\mathbf{s}_j)$ is the predicted value at $\mathbf{s}_j$ obtained using $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_{j-1}), Z(\mathbf{s}_{j+1}), \ldots Z(\mathbf{s}_n)\}$ as the training data and the sum is over the test data set whose cardinality is denoted by n_*. In Figures 2.8a and 2.8b, we plot the histograms of the observed residuals from the VS-based robust regression and the residuals

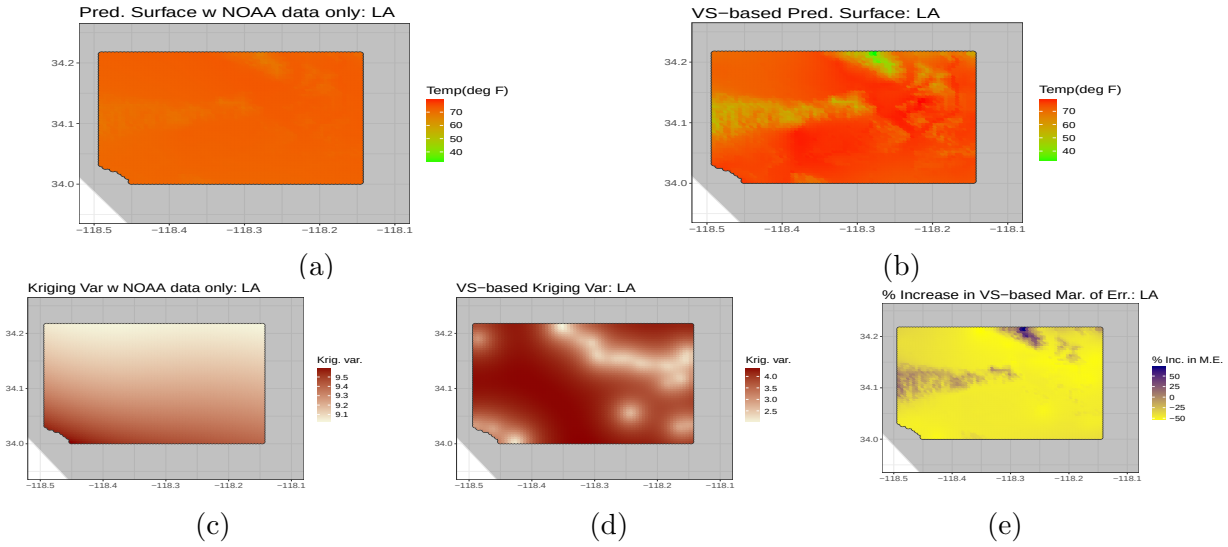


Figure 2.9: (a) Hyper-local version of the same surface as in Figure 2.5c; (b) Prediction surface obtained by the VS-based technique on the crowdsourced data in Los Angeles; (c) Residual kriging variance for the predictions using NOAA data only (d) Residual kriging variance for the predictions using the VS-based predictions with crowdsourced data; (e) the % increase in the margin of error for the VS-based predictions as compared to the predictions with NOAA data

after the VS-based smoothing. The VS-based smoothing clearly reduces the spread of the residual values by smoothing out the large errors. In Figure 2.8c, we show the robust variogram fitting of the VS-based smoothed residuals for the optimal choice of the smoothing parameter $q = 0.8$.

Given these analyses, we construct a prediction surface over the region $\mathcal{R}$ using Equation 2.3.2. In Figure 2.9, we plot the hyper-local prediction surfaces obtained by the standard analysis with the NOAA ground-station data only as well as the one obtained by implementing the VS-based technique on the crowdsourced observations with the ground-station data as the reference. Clearly the prediction surface obtained from standard analysis of the ground-station data (Figure 2.9a) is too smooth to capture the local variability accurately. The prediction surface obtained by the VS-based analysis on

crowdsourced data shows more variation across the space. To highlight the advantage of having crowdsourced observations, we compare the residual kriging variance surfaces in Figures 2.9c and 2.9d. Around the crowdsourced observations, the kriging variance is much smaller as compared to the global kriging using only the ground-station data over a larger region.

In addition, we illustrate the gain in efficiency by plotting the percentage increase in margin of error (at 95% confidence) for the VS-based predictions from the hyper-local crowdsourced information as compared to the global prediction using ground-station data only, i.e., $100 \times \Big(\text{M.E.}(\hat{Y}_{\text{vs}}(\mathbf{s})) - \text{M.E.}(\hat{Y}(\mathbf{s}))\Big) \,/\, \Big(\text{M.E.}(\hat{Y}(\mathbf{s}))\Big)$, where M.E. denotes the 'margin of error' (half of the length of the prediction interval) to predict the target response $Y(\mathbf{s})$. To compute the margin of error, we use ad hoc confidence intervals for the residual kriging predictor with ± 1.96 as the corresponding quantiles and then add the margin of error of the mean $(1.96 \times \text{s.e.}(\mathbf{x}(\mathbf{s})'\hat{\boldsymbol{\beta}}_{\text{vs}}))$ and the margin of error of the residual kriging predictor $(1.96 \times \sqrt{\text{Krig.Var.}(\tilde{\epsilon}(\mathbf{s}))})$. The margin of error for the standard predictor is computed similarly. A more theoretically justifiable interval can be obtained through spatial re-sampling technique as discussed in Lahiri (2003b), but that requires further research and is beyond the scope of this study. In Figure 2.9e, for most of the locations where the predictions have been carried out, there are decrease in the margin of errors for the VS-based predictions as compared to the global predictions using ground-station data only. At the locations that are close the crowdsourced observations, the VS-based prediction technique has achieved up to a 50% gain in efficiency.

The disadvantage of VS-based hyper-local analysis is that the model is estimated very regionally and hence extrapolation of the estimated mean model outside the sample space is likely to give misleading and inefficient predictions. For example, in Figure 2.9b there are locations with elevations of more than 500 meters, while the maximum elevation

in the crowdsourced sample is 350 meters. This leads to poor predictions (e.g., ambient temperature less than 50°F) at some locations as can be seen in Figure 2.9e. Note that, though in those regions the efficiency of VS-based predictions fall short, the residual kriging variance (Figure 2.9c and 2.9d) for the VS-based kriging predictor is still less than the global kriging with NOAA data only. So, the loss in efficiency in VS-based predictions is solely due to the the extrapolation of the hyper-locally estimated mean function at points outside the covariate sample space.

We conduct a similar analysis for another hyper-local region close to Brooklyn, NY and plot the results in Figure 2.10. The prediction surface in Figure 2.10c is obtained by using standard methodology on 120 ground-station observations over the east-coast; and the suface in Figure 2.10d is generated through VS-based hyper-local analysis of the crowdsourced observations in Figure 2.10b. Comparing these two prediction surfaces, we again see that the regional variation is prominent for the prediction surface obtained from VS-vased hyper-local analysis where as, the global analysis generates a surface that is too smooth to accurately capture local variations. In Figure 2.10f, the advantage of having crowdsourced data for hyper-local prediction of the process is visible, as the kriging variance of the VS-based methodology is much smaller compared to Figure 2.10e, especially in locations close to the crowdsourced observations. In Figure 2.10g, we see up to 33% gain in margin of error by implementing the VS-based methodology on the crowdsourced data in locations close to the crowdsourced observations. Similar to the previous analysis of the Los Angeles data, the advantage of the VS-based hyper-local predictions is lost if the predictions are attempted at locations too far from the crowdsourced observations or at locations with elevations outside the range of crowdsourced sample.

In addition to the VS-based hyper-local analysis, we have also conducted the analysis for both of the hyper-local regions in Los Angeles and Brooklyn with the standard

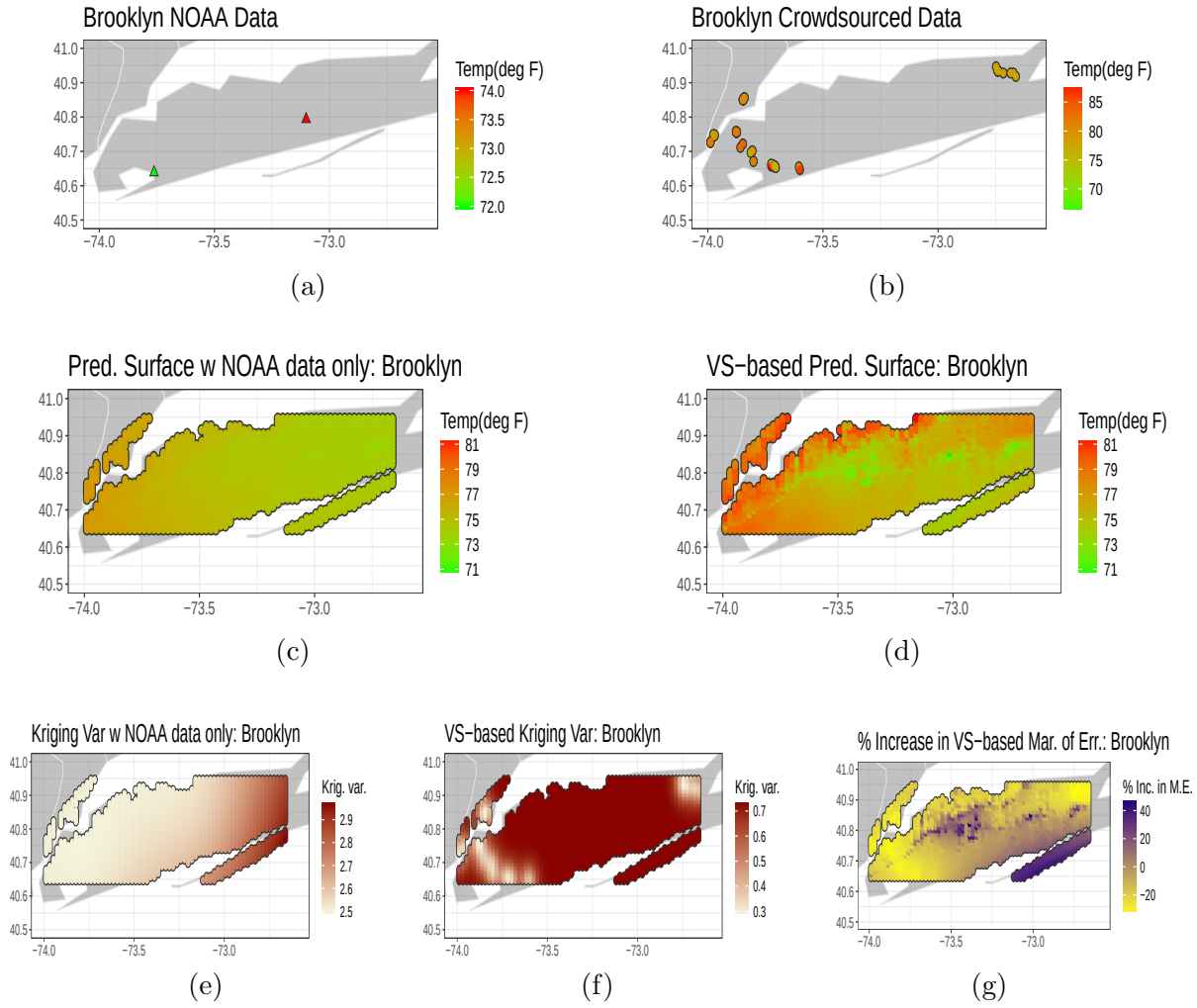


Figure 2.10: (a) Ground-station observations in the selected hyper-local region; (b) Crowdsourced observations in the same region (c) Prediction surface obtained by standard analysis of NOAA ground-station data; (d) Prediction surface obtained by the VS-based technique on the crowdsourced data; (e) Residual kriging variance for predictions using NOAA data only; (f) Residual kriging variances for the predictions using the crowdsourced data; (g) Percent increase in the margin of error for the VS-based predictions compared to the predictions with NOAA data

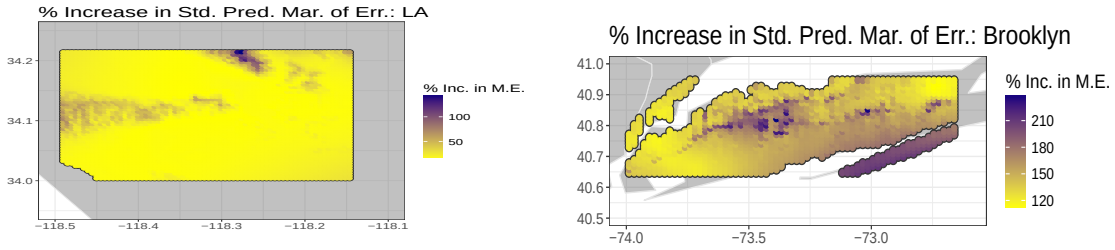


Figure 2.11: The increase in margin of error for the standard approach in hyper-local regions in Los Angeles (left) and Brooklyn (right).

approach without considering the veracity of the crowdsourced observations and then compared the predictions with the global prediction surface obtained using reference data only. As we can see in Figure 2.11, in both Los Angeles and Brooklyn, the margins of error for the predictions using the standard approach are larger in all the locations as compared to the global predictions using ground-station data. In Brooklyn, even at the locations around the crowdsourced observations, with reference to the global prediction using ground-station data, the margin of error of standard predictions using the crowdsourced observations have increased by at the least 120%, where as, as we have mentioned already, the VS-based methodology has achieved a decrease in the margin of error up to 33% (Figure 2.10g).

2.5.2 Validation at the ground-stations

The goal of the analysis in this section is to validate the predictions obtained by hyper-local analysis of crowdsourced data using VS-based methodology. To do so, we have selected a set of 14 ground-stations that satisfy the following criteria: (1) there are at least 30 crowdsourced data points available nearby with at least 20 observations with a VS greater or equal to 0.4; (2) the elevation of those stations is not too far from the range

of the local crowdsourced samples. We have conducted 14 hyper-local analyses, as described in Section 2.5.1, for hyper-local structure exploration of the ambient temperature and then predicted at those selected ground-station locations to validate the VS-based predictions. We have omitted these 14 stations before-hand so that these are not used in defining the 'benchmark' value at the crowdsourced data locations to compute VS; this way the validation data has no effect on the training phase of the predictions. We have also conducted the same hyper-local analyses using the standard technique without taking the quality of the observations into account. The results are compiled in Table 2.5. The

Table 2.5: Predictions using both the VS-based and standard approach at the ground-stations with crowdsourced observations in proximity.

STATION_NAME	Target Temp.	PredTemp.VS	VS.ME	PredTemp.Std	Std.ME
CHICAGO OHARE INTERNATIONAL AIRPORT IL US	76	76.01	6.22	76.75	8.74
WASHINGTON DULLES INTERNATIONAL AIRPORT VA US	79	82.80	5.13	75.33	18.35
WASHINGTON REAGAN NATIONAL AIRPORT VA US	80	81.95	7.77	75.52	31.64
MIAMI INTERNATIONAL AIRPORT FL US	79	77.81	0.50	78.57	1.23
LITTLE TUJUNGA CALIFORNIA CA US	68	64.78	6.01	63.74	7.41
LOS ANGELES INTERNATIONAL AIRPORT CA US	68	68.91	3.31	67.87	4.26
BEVERLY HILLS CALIFORNIA CA US	70	67.94	6.27	68.12	7.54
TOLEDO EXPRESS AIRPORT OH US	75	79.45	5.72	79.39	8.18
DETROIT METROPOLITAN AIRPORT MI US	76	78.79	6.66	80.74	9.73
MINNEAPOLIS ST PAUL INTERNATIONAL AIRPORT MN US	70	77.03	6.67	76.97	10.96
CARLOS AVERY MINNESOTA MN US	69	73.33	11.46	74.65	19.05
JFK INTERNATIONAL AIRPORT NY US	72	78.24	3.06	80.82	3.82
ISLIP LI MACARTHUR AIRPORT NY US	74	75.10	6.29	75.61	8.79
AUSTIN BERGSTROM INTERNATIONAL AIRPORT TX US	81	78.87	2.24	86.50	10.19

advantage of using the VS-based techniques as compared to the standard methodology is clear from the results. The RMSPE of the VS-based predictor for these 14 ground-stations is 3.71, while for the standard approach it is 4.54. More importantly, the average margin of error (at 95% confidence) for standard predictor is 13.61 and for the VS-based methodology it is 6.28. Relative to the standard methodology, on average, the VS-based technique has achieved approximately 54% gain in efficiency of the predictions.

2.6 Summary and Conclusions

In this chapter, we have introduced the veracity score to assess the quality of observations in geostatistical setting. The VS is defined using a benchmark to compare the varying quality observations with. We used the ground-station data as our reference to define the benchmark values in the case studies. The similar scoring approach to assess the veracity of the observations can be used in other contexts as well. We have also discussed the case when no other reference information is available and propose a version of VS using locally and robustly estimated measure of center as the benchmark. A robust approach for modeling varying-quality spatial data using the VS has been proposed and evaluated. We have illustrated the VS-based methodology on a crowdsourced data set coming from the mobile app WeatherSignal using NOAA ground-station data as the reference. Both the simulation studies in section 2.4 and the case studies in section 2.5.1 show the advantage of the VS-based methodology over the standard geostatistical approach when dealing with noisy spatial data. In addition, by implementing the VS-based methodology on the varying-quality local crowdsourced data, we can achieve a more accurate and efficient hyper-local predictions as compared to the global prediction obtained from the analysis of ground-station data only.

In the analysis of crowdsourced data using the VS-based methodology, the model is estimated using observations in a local region. Predicting at more distant locations or with covariates outside the range of the sample may provide misleading and inefficient predictions. Though global prediction surfaces using the ground-station data or satellite images have advantages in this regard, many organizations are interested in providing cost-effective hyper-local predictions of weather using sensor-generated geographical information through weather-related mobile apps. For example, the global leader in weather

information, AccuWeather, launched AccUcast in 2015 (AccuWeather 2015), a feature that allows each user to share their local weather information as captured by the built-in mobile sensors. Other applications include Sunshine (Moynihan 2015) and Dark Sky (Dalton 2016), which turn each app-user into a "meteorological station" for gathering and sharing hyper-local weather information. Spatial prediction of daily weather is generally based on satellite imagery or data from weather stations (Thornton et al. (1997), Vancutsem et al. (2010)). However, the availability of crowdsourced data at more granular spatial resolution, especially in high-population density regions, may provide a way to answer with more reliability whether you need to wear a jacket or bring an umbrella for a 10 block walk inside the city.

There are several interesting future directions for this work. In this chapter, we have not provided theoretical justification for the superiority of the VS-based methodology as compared to the standard approach in the analysis of noisy spatial data. Inspired by the simulations executed in this work, we believe that under a suitable asymptotic framework and a very general noise model, we can theoretically justify the robustness and efficiency of the VS-based methodology. The mean and covariance models used to explore the structures of the average temperature process are quite simple, yet reasonable and effective. More complex models like nonlinear regression models and anisotropic covariance can be incorporated in the VS-based technique to increase flexibility of the analysis. The margins of error reported in this study are ad hoc. To get theoretically justifiable prediction intervals, distributional assumption is needed or re-sampling technique for dependent data (Lahiri 2003b) needs to be incorporated.

Acknowledgements

This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). This research is also partially funded by National Science Foundation (NSF) grant DMS-1613192. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government. Special thanks are extended to the OpenSignal team that was in-charge of the academic partnership program in 2015 for making the data available to the authors.

Chapter 3

Large Sample Properties of VS-based Estimation

3.1 Introduction

As discussed in the previous chapter, with the advancement of technologies, in the field of data science and big-data analytics, 'veracity' is becoming one of the most important V's, along with velocity, variety and volume of the data. In a recent article, Lukoianovaand and Rubin (2014) have mentioned how the data source, data collection technologies and data processing methodologies can induce bias, ambiguity and noise in real-world big-data scenario. Practitioners often not only want to detect the noisy observations but also want to assess a reliability score to each of the observations which efficiently indicates the 'relative' magnitude of the associated noise with it. For example, studies conducted by Evans (1997) show that dynamic display of veracity of the contents can help the map-users (e.g. google map) to take advantage of the reliability information in making decisions. Though there are some works on veracity analysis in media and communications (for

example, see Conroy et al. 2015; Rendon et al. 2018 etc.) reliability assessment in the analysis of spatial data is not that common in literature. Standard geostatistical analysis of the corrupted observations without taking the noisy nature of the data into account can result in erroneous inference and prediction. Moreover, in standard geostatistical framework, often we do not have replications of the observed process and hence, the standard outlier detection techniques like extreme value analysis or principle component analysis can not be used directly. As mentioned by Saha and Srivastava (2014), with the advancement of global positioning system and micro-sensors used to collect spatial data in large volume, a scalable and timely methodology in geostatistics is needed to assess the reliability of the observations as well as incorporate the veracity assessment in the subsequent analysis to produce robust inferences.

Following the notation introduced in Chapter 2, let $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{D} \subseteq \mathbb{R}^2\}$ be the random field of our interest which is assumed to have a decomposition of the form,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}), \quad \text{for } \mathbf{s} \in \mathcal{D}, \tag{3.1.1}$$

where $\mu(\mathbf{s})$ is a deterministic smooth mean function capturing the large scale variation of the process and $\epsilon(\mathbf{s})$ is a mean-zero spatially correlated 'residual' process for addressing the small scale variations over the space (see chapter 2 Cressie 1993, chapter 3 Gelfand et al. 2010). Often in practice, for example the crowdsourced weather data considered in Chapter 2, instead of observing realizations from the true process $\{Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)\}$, we observe corrupted observations $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$. For this chapter also, we assume that the corrupted observations can be written as,

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + w(\mathbf{s}_i), \quad \text{for } \mathbf{s} \in \mathcal{D}, \tag{3.1.2}$$

where $\{w(\mathbf{s})\}$ is a mean-zero spatially correlated process, possibly nonstationary, which contains noise in addition to the small-scale local variations induced by $\epsilon(\mathbf{s})$. In Chapter 2, we have introduced a reliability metric, namely Veracity Score (VS) to assess the quality of the observations $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$. In addition, we have proposed and demonstrated a VS-based method to analyze the varying-quality observations coming from mobile sensors through crowdsourcing. Though the merits of the VS-based method have been evaluated through simulations, we have not analyzed asymptotic properties of the VS or the VS-based estimators theoretically. Moreover, in the real-data example, we have used a NOAA ground-station data as reference to define the VS for the analysis of crwodsourced temperature data. Any example of the VS-based methodology when no reference data is available has not been demonstrated in Chapter 2. In this chapter, we have established asymptotic consistency of the VS-based regression parameter estimator under the assumption that a 'small' proportion of the observations are corrupted through the additive-multiplicative noise model as given in Equation 3.1.2. In addition, efficiency of the VS-based method has been established through the comparison of the asymptotic mean squared errors (MSE) of the VS-based and the OLS estimators. Finally, the VS-based estimation and prediction has been evaluated and compared with other robust methods in geostatistics through simulation and a real-data example.

The organizations of this chapter is as follows. In Section 3.2, we briefly review the VS-based methodology as given in Chapter 2 for better understanding of the results of this chapter. Section 3.3 defines the spatial framework, states the assumptions and the results regarding asymptotic properties of the VS-based estimators. In Section 3.4, we evaluate the VS-based methodology through extensive simulations. Section 3.5 provides a real-data example that has been used previously by Cressie (1993) and Papritz (2018a) to showcase the robust properties of their methods. Finally, in Section 3.6, we summarize

our findings and future directions to this work.

3.2 Review of Veracity Score (VS) Methods

In this section, we briefly review the VS-based method in geostatistics and also set the notations that will be used for stating the results later.

3.2.1 Formulation of VS

The aim of the measure of veracity is not only to identify the 'outliers' but also quantify the amount of variation associated with each of the observations due to noise, possibly on a continuous scale, so that the effect of the noise can be reduced in the concurrent analysis by down-weighting the effect of the low-quality observations.

Before going to the definition of VS, here we introduce the required notations and assumptions. Suppose $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$ are the varying-quality observations at irregularly spaced locations $\mathcal{S}_n \coloneqq \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. As discussed above, $Z(\mathbf{s}_i)$ is corrupted version of the 'true' realization $Y(\mathbf{s}_i)$. The true process $\{Y(\mathbf{s})\}$ is assumed to have a decomposition as given in Equation 3.1.1. Also, a δ-neighborhood around a spatial point $\mathbf{s} \in \mathbb{R}^2$ is denoted as $\mathcal{B}_\delta(\mathbf{s})$, i.e. $\mathcal{B}_\delta(\mathbf{s}) \coloneqq (\mathbf{s} - \delta, \mathbf{s} + \delta]$ for some $\delta \in \mathbb{R}^+$, where the subtraction and addition is component-wise.

Under the notations described above, the VS of the observation $Z(\mathbf{s}_i)$ is given by,

$$V(\mathbf{s}_i) = \phi\left(\frac{|Z(\mathbf{s}_i) - \mathcal{C}(\mathbf{Z}_i)|}{\alpha + D(\mathbf{Z}_i)}\right), \tag{3.2.1}$$

where $\phi : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+ \cup \{0\}$ is some non-increasing bounded above function, referred as veracity function; $\alpha \geq 0$ is a regularity parameter, called the *baseline deviation*; $\mathbf{Z}_i \coloneqq$

$\Big(Z(\mathbf{s}_{i_1}), \ldots, Z(\mathbf{s}_{i_{n(i)}})\Big)'$ where $\{\mathbf{s}_{i_1}, \ldots, \mathbf{s}_{i_{n(i)}}\}$ is the set of observation locations in the small δ-neighborhood $\mathcal{B}_\delta(\mathbf{s}_i)$; and finally, $\mathcal{C}(\mathbf{x})$ is a measure of central tendency and $D(\mathbf{x})$ is a measure of dispersion of the values in the vector $\mathbf{x}$. Clearly, the definition of VS given in Equation 3.2.1 measures the amount of deviation of the observation $Z(\mathbf{s}_i)$ from the locally estimated measure of center $\mathcal{C}(\mathbf{Z}_i)$ relative to the variation in that neighborhood $D(\mathbf{Z}_i)$. The veracity function $\phi(\cdot)$ being a non-increasing function, larger the value of the scaled deviation $\frac{|Z(\mathbf{S}_i)-\mathcal{C}(\mathbf{Z}_i)|}{\alpha+D(\mathbf{Z}_i)}$ lower is the VS and for lower values of the scaled deviation, VS will be higher.

For the measure of center and dispersion we have considered the following two choices that are widely used by practitioners:

1. $\mathcal{C}(\mathbf{Z}_i) = Q_2(\mathbf{Z}_i)$ and $D(\mathbf{Z}_i) = \text{IQR}(\mathbf{Z}_i) = Q_3(\mathbf{Z}_i) - Q_1(\mathbf{Z}_i)$, where $Q_j(\mathbf{x})$ denotes the j-th quartile of the observations in $\mathbf{x}$;

2. $\mathcal{C}(\mathbf{Z}_i) = \bar{Z}_{i.} = \frac{1}{n(i)} \sum_{j=1}^{n(i)} Z(\mathbf{s}_{i_j})$, and $D(\mathbf{Z}_i) = \text{sd}(\mathbf{Z}_i) = \sqrt{\frac{1}{n(i)-1} \sum_j \left(Z(\mathbf{s}_{i_j}) - \bar{Z}_{i.}\right)^2}$.

Clearly, the first set of choices are based on sample quantiles and are expected to be more robust and efficient as compared to the sample mean and standard deviation in the analysis of noisy data, even under dependence (for details, see Sen 1968). In this chapter, we refer the VS with choice (a) as the measure of location and scale as 'Medain-VS' and denote it by $V^{(m)}(\mathbf{s}_i)$. Similarly, for choice (b) we denote the VS as $V^{(a)}(\mathbf{s}_i)$ and refer it as 'Mean-VS'. For veracity function we have again used $\phi(x) = \exp(-x)$ as in Chapter 2. The results in this chapter can be generalized to the other choices provided in Section 5.1.2.1 in Chapter 5.

3.2.2 VS-based estimation in spatial regression

Before going to the estimation of the model parameters, we need to discuss the spatial linear regression setup in brief. In spatial regression, the mean structure $\mu(\cdot)$ in Equation 3.1.1 is assumed to have a *linear* form in terms of a set of known covariates, i.e. $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$, where $\mathbf{x}(\cdot) = (x_1(\cdot), ..., x_p(\cdot))'$ is the p-dimensional deterministic vector process of covariates and $\boldsymbol{\beta}$ denotes the unknown regression parameter vector. Hence, the spatial regression model can be written as,

$$Y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + \epsilon(\mathbf{s}). \tag{3.2.2}$$

To make the inference feasible form only one replication of the process over the space, some assumption on the second-order structure of the ϵ-process is needed. One of the most commonly used assumptions is that $\{\epsilon(\cdot)\}$ is an intrinsically stationary process with an admissible parametric variogram function $2\gamma(\mathbf{h};\boldsymbol{\theta}) = \text{Var}\{\epsilon(\mathbf{s}) - \epsilon(\mathbf{s}+\mathbf{h})\}$ and $\boldsymbol{\theta}$ is the covariance parameter of interest. Under second-order stationarity, it can be shown that $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$ where $C(\mathbf{h}) = \text{Cov}(\epsilon(\mathbf{s}), \epsilon(\mathbf{s}+\mathbf{h}))$ is the covariogram or covariance function of the ϵ-process. For more details about variograms and covariograms see Chapter 2, Cressie (1993). Though often in literature distributional assumptions like Gaussian Process (GP) are made, in this study, we do not impose strict distributional restriction on the random field $\{\epsilon(\cdot)\}$.

Under the setting discussed above, as the covariance parameter is unknown, the standard practice is to estimate the regression parameter using *ordinary least squares* (OLS): $\hat{\boldsymbol{\beta}}_{\text{ols}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, where $\mathbf{X} \coloneqq (\mathbf{x}(\mathbf{s}_1), \ldots, \mathbf{x}(\mathbf{s}_n))'$. Next, the residuals from the OLS fit are used in *weighted least squares*-based variogram estimation (for details,

see page 90, Chapter 2 in Cressie 1993) to obtain the covariance parameter estimator $\hat{\boldsymbol{\theta}}_{\text{wls}}$. Once the covariance parameter is estimated, one can try to improve the mean parameter estimates by using *estimated generalized least squares* (EGLS) estimator, given by $\hat{\boldsymbol{\beta}}_{\text{egls}} = \left(X'\Sigma^{-1}(\hat{\boldsymbol{\theta}}_{\text{wls}})X\right)^{-1} X'\Sigma^{-1}(\hat{\boldsymbol{\theta}}_{\text{wls}})\mathbf{Z}$, where $\Sigma(\hat{\boldsymbol{\theta}}_{\text{wls}})$ is the estimated covariance matrix. But, this last step introduces additional variability due to the estimation of the covariance parameters and hence not necessarily preferable than the OLS estimator. Finally, the estimated mean and covariance structure is used to predict the process at a new location $\mathbf{s}_0$ using *best linear unbiased predictor* or kriging predictor (for details, see Chapter 3 in Cressie 1993).

The standard approach of estimation and kriging is not robust in nature and hence, may produce erroneous inference and prediction in case there are outliers in the data. To estimate the mean parameter $\boldsymbol{\beta}$ robustly, instead of using ordinary squared error loss, in VS-based estimation, we minimize weighted squared error loss with VS as the corresponding weights as shown below:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{vs}} &= \underset{\boldsymbol{\beta}}{\text{argmin}} \sum_{i=1}^{n} V(\mathbf{s}_i)\left(Z(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}\right)^2 \\ &= \left(\mathbf{X}'\mathbf{D}_v\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}_v\mathbf{Z},\end{aligned} \tag{3.2.3}$$

where $\mathbf{D}_v = \text{diag}(V(\mathbf{s}_1), \ldots, V(\mathbf{s}_n))$. Clearly, as the VS is expected to be smaller for bad-quality observations, the VS-based estimator given in Equation 3.2.3 is expected to be affected less by the corrupted observations in the data. If the VS can capture the noisy observations perfectly, i.e. $V(\mathbf{s}_i) \approx 0$ if $Z(\mathbf{s}_i)$ is contaminated and $V(\mathbf{s}_i) = 1$ otherwise, then $\hat{\boldsymbol{\beta}}_{\text{vs}}$ is approximately equal to the OLS estimator computed only from the 'noiseless' observations in the data.

Depending on the versions of VS used to compute the VS-based estimator, we have

considered two versions of the VS-based estimator in this chapter: 'Mean-VS', denoted by $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(a)}$, where the sample mean and standard deviations are used to assess veracity and 'Median-VS', denoted by $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(m)}$, with sample median and IQR in the definition of VS. Though the focus of this work is mainly on the properties of the 'Median-VS' regression estimator, in the following sections we have remarked about the properties and asymptotic behaviors of the 'Mean-VS' version as well.

Once the regression parameter is estimated robustly, the observed residuals from the VS-based WLS, i.e. $\hat{\epsilon}_{\text{vs}}(\mathbf{s}_i) = Z(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_i)'\hat{\boldsymbol{\beta}}_{\text{vs}}$ for $i \in \{1, 2, ..., n\}$, can not be used directly in the covariance parameter estimation due to the presence of noise in a part of these residuals. In Chapter 2, we have proposed a VS-based smoothing of the residuals:

$$\tilde{\epsilon}(\mathbf{s}_i) = V(\mathbf{s}_i)^q \hat{\epsilon}_{\text{vs}}(\mathbf{s}_i) + (1 - V(\mathbf{s}_i)^q) Q_2(\mathbf{Z}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}_{\text{vs}}), \tag{3.2.4}$$

where $\mathbf{X}_i \coloneqq \left(\mathbf{x}(\mathbf{s}_{i_1}), ..., \mathbf{x}(\mathbf{s}_{i_{n(i)}})\right)'$ is the $n(i) \times p$ matrix of the covariates corresponding to the observations in $\mathcal{B}_{\boldsymbol{\delta}}(\mathbf{s}_i)$; and $q \geq 0$ is another regularity parameter which determines the smoothness of the residuals defined in Equation 3.2.4. Next, these VS-based smoothed residual vector $\tilde{\boldsymbol{\epsilon}} \coloneqq (\tilde{\epsilon}(\mathbf{s}_1), \ldots, \tilde{\epsilon}(\mathbf{s}_n))'$ is used to get the VS-based covariance parameter estimator $\hat{\boldsymbol{\theta}}_{\text{vs}}$ using the robust variogram model fitting proposed by Cressie and Douglas (1980). Finally, the prediction of the ϵ-process at a new location $\mathbf{s}_0$, denoted by $\tilde{\epsilon}(\mathbf{s}_0)$, is obtained using ordinary kriging on the VS-based smoothed residuals as discussed in Chapter 2.

In the next section, we have focused on the asymptotic behavior of VS-based regression parameter estimator, i.e. $\hat{\boldsymbol{\beta}}_{\text{vs}}$. Though investigation of asymptotic properties VS-based covariance estimation is beyond the scope of this chapter, we have evaluated the same using simulations under suitable spatial frameworks in Section 3.4.

3.3 Asymptotic Properties of the VS-based Regression Estimator

In this section, we first establish asymptotic consistency of the VS-based estimator: $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(m)}$. Next, we show that for analyzing noisy spatial data it is asymptotically more efficient than the standard OLS estimator. But, before stating the main results of this section we need to introduce the spatial framework, the regularity conditions on the spatial process and a noise model under which the results of this chapter have been proved. From this section onward, by $C(\cdot), C, C_1, C_2, \ldots$ we denote constants with respect to the sample size n. For $d \geq 2$, we denote the volume of a set $A \subset \mathbb{R}^d$ as $|A|$, i.e., the Lebesgue measure of A if it has nonzero volume and the cardinality of A if A is finite.

3.3.1 Model specification

As discussed before, the process of our interest is denoted as $\{Y(\mathbf{s}) : \mathbf{s} \in \mathcal{D} \subset \mathbb{R}^2\}$ and instead of observing realizations from the 'true' process, we observe corrupted observations $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$ where, we assume that the noise is inserted in the observations through the additive-multiplicative error structure defined in Equation 2.2.3. For this section as well, we consider the spatial regression framework as defined in Equation 3.2.2. For the mean-zero second-order stationary residual process $\{\epsilon(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, let us denote the marginal distribution function of $\epsilon(\mathbf{s})$ as $F_\epsilon(\frac{x}{\sigma_\epsilon})$ and the dependence structure is given as,

$$\text{cov}\left(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)\right) = \begin{cases} \sigma_\epsilon^2 \, \rho_\epsilon\left(\mathbf{s}_i - \mathbf{s}_j\right) & \text{if } \mathbf{s}_i \neq \mathbf{s}_j \\ \sigma_\epsilon^2 & \text{otherwise.} \end{cases}$$

Here, $F_\epsilon(\cdot)$ is a distribution function of a mean-zero unit-variance random variable and $\rho_\epsilon(\cdot)$ is a non-negative definite function on $\mathbb{R}^2$ with $\rho_\epsilon(\mathbf{0}) = 1$. To preserve the identifiability, we have assumed that the supports of $\{\epsilon_{M_i} : i \in \{1, \ldots, n\}\}$ are contained in $[0, \infty)$.

We assume that among $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$ only a 'small' portion is 'bad-quality data' and the corrupted observations are coming through the additive-multiplicative transformation of the data, as we have already defined in Equation 2.2.3. Mathematically, we formulate this assumption as follows:

$$\epsilon_{M_i} \underset{\text{indep.}}{\sim} F_M\left(\frac{x-1}{\sigma_{M_i}}\right);\ \epsilon_{A_i} \underset{\text{indep.}}{\sim} F_A\left(\frac{x}{\sigma_{A_i}}\right);\ \{\epsilon_{M_i}\}_{i=i}^{\infty} \text{ and } \{\epsilon_{A_i}\}_{i=i}^{\infty} \text{ are independent; and}$$

$$\sigma^2_{M_i} = \begin{cases} 0 & \text{if } i \in G_n \\ \sigma^2_M & \text{o.w.} \end{cases} ;\ \sigma^2_{A_i} = \begin{cases} 0 & \text{if } i \in G_n \\ \sigma^2_A & \text{o.w.}\ , \end{cases} \tag{3.3.1}$$

where $G_n \subset \{1, \ldots, n\}$ is a subset of indices; $\sigma_M, \sigma_A > 0$; F_M, F_A are distribution functions of centered and scaled random variables. So, under this framework if $i \in G_n$ we have no noise associated with the observation, i.e. $Z(\mathbf{s}_i) = Y(\mathbf{s}_i)$. We assume that the proportion of good observations is approximately a constant and denoted by q_e, i.e. $|G_n|/n \approx q_e$ and $1 - q_e$ is the proportion of noisy observations in the data, which is not 'too large' (e.g., $1 - q_e \not> 0.25$). Under the specified model we can rewrite our observations as,

$$Z(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + w(\mathbf{s}_i), \tag{3.3.2}$$

where $E(w(\mathbf{s}_i)) = 0$ and $\text{Var}(w(\mathbf{s}_i)) = \sigma_i^2$ and

$$\sigma_i^2 = \begin{cases} \sigma_\epsilon^2 & \text{if } i \in G_n, \\ \sigma_\epsilon^2 + \tau_i^2 & \text{o.w.;} \end{cases}$$

where, $\tau_i^2 \equiv \tau(\mathbf{s}_i)^2 = (\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta})^2\sigma_M^2 + \sigma_\epsilon^2\sigma_M^2 + \sigma_A^2$ is the additional noise variance associated with the corrupted observations. We denote the marginal distribution function of $w(\mathbf{s}_i)$ as $F_i(x)$ for $i \in \{1, \ldots, n\}$, for any $n \geq 2$. Clearly, if $i \in G_n$, $w(\mathbf{s}_i) = \epsilon(\mathbf{s}_i)$ and $F_i(x) = F_\epsilon(x/\sigma_\epsilon)$ and if $i \in G_n^c$, we denote the marginal distribution function of $w(\mathbf{s}_i)$ as $F_2(\frac{x}{\sigma_i})$, where $F_2(\cdot)$ is again a distribution function of a centered and scaled random variable. Finally, without loss of any generality we have taken the baseline deviation parameter α in the definition of VS to be equal to 0 for the results of this Chapter. This can be straightforwardly extended for any constant $\alpha > 0$.

3.3.2 Spatial framework and notations

We have taken the sampling region to be $\mathcal{D} \equiv \mathcal{D}_n = \lambda_n[0,1]^2$, i.e. a 2-dimensional square region with area λ_n^2 where $\{\lambda_n\}$ is a sequence of positive reals such that $\lambda_n \to \infty$ as $n \to \infty$ and

$$\lambda_n^{-1} + n^{-1}\lambda_n^2 \to 0. \tag{3.3.3}$$

The condition in 3.3.3 is similar to the *mixed increasing domain* asymptotic framework, as discussed in Hall and Patil (1994) and Lahiri et al. (1999) where, the first component of the condition states that the domain has to increase with the sample size and the second component in 3.3.3 allows the possibility of infilling sampling points as well.

The results stated here are not particular to the shape and position of the rectangular region $\mathcal{D}_n$. Our results can be extended to regions of any shape that can be obtained by inflating a prototype region by a scaling constant and the prototype region is contained in $(-1/2, 1/2]^2$ with center at the origin. For more details of this spatial asymptotic framework see Lahiri et al. (2002).

Now, we will introduce a set of notations required to state the regularity conditions, the results as well as the proofs. For any $A \subset \mathbb{R}^2$ and a random field $\{T(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ let us denote $\mathcal{F}_T(A) = \sigma\langle T(\mathbf{s}) : \mathbf{s} \in A\rangle$, the σ-field generated by the random variables $\{T(\mathbf{s}) : \mathbf{s} \in A\}$. For any $d \geq 2$ and for any $\mathbf{x} \in \mathbb{R}^d$ the L_1 and L_2 norms are denoted by $|\mathbf{x}|$ and $\|\mathbf{x}\|$ respectively. For any two sets A and B in $\mathbb{R}^2$ we denote the the distance between them as $d(A, B) = \inf\{|\mathbf{s} - \mathbf{s}'| : \mathbf{s} \in A, \mathbf{s}' \in B\}$. For the random field $\{T(\mathbf{s})\}$ we define the strong-mixing coefficient as

$$\alpha_T(u, v) = \sup\{\tilde{\alpha}_T(A, B) : \ d(A, B) \geq u; \ \ |A| \leq v, \ \ |B| \leq v\}, \tag{3.3.4}$$

where, the supremum is taken over the set of all 2-dimensional rectangles A, $B \subset \mathbb{R}^2$, $u, v > 0$; and

$$\tilde{\alpha}_T(A, B) = \sup\{|P(V_1 \cap V_2) - P(V_1)P(V_2)| : V_1 \in \mathcal{F}_T(A), \ V_2 \in \mathcal{F}_T(B)\}.$$

For a distribution function $F(\cdot)$, we define the inverse of it as $F^{-1}(p) = \inf\{x : F(x) \geq p\}$, for $p \in [0, 1]$.

To define the VS in Section 3.2 we used a square neighborhood $\mathcal{B}_\delta(\mathbf{s}_i)$ around the spatial point $\mathbf{s}_i$ with side 2δ units. From this section we change the notation as: $\delta \equiv \delta_n$ and $\mathcal{B}_\delta(\cdot) \equiv \mathcal{B}_{\delta_n}(\cdot)$ to ensure that the size of the 'local' neighborhood varies with the

sample size n. Let $M^p_{n(i)}$ be the smallest p-th sample quantile of $\left\{w(\mathbf{s}_{i_1}), \ldots, w(\mathbf{s}_{i_{n(i)}})\right\}$, the residuals associated with the observations in the square-neighborhood $\mathcal{B}_{\delta_n}(\mathbf{s}_i)$. We denote $\hat{F}_{n(i)}(x) = \left\{\text{No. of } w(s_{i_j}) \leq x\right\}/n(i)$, the empirical distribution function of the realizations from the w-process in the neighborhood around $\mathbf{s}_i$. Clearly, $M^p_{n(i)} = \hat{F}^{-1}_{n(i)}(p)$. By $\bar{F}_{n(i)}(x)$ we denote the distribution function defined as $\bar{F}_{n(i)}(x) = \left(\sum_{j=1}^{n(i)} F_{i_j}(x)\right)/n(i)$, where, F_{i_j} denotes the distribution function of $w(\mathbf{s}_{i_j})$. The smallest p-th quantile of the distribution function $\bar{F}_{n(i)}$ is denoted by $\xi^p_{n(i)}$, i.e. $\xi^p_{n(i)} = \bar{F}^{-1}_{n(i)}(p)$.

3.3.3 Consistency of the VS-based regression parameter estimator

Here we focus on proving the consistency of $\hat{\boldsymbol{\beta}}^{(m)}_{\text{vs}}$, the Median-VS version of VS-based estimator. The conditions we need for that are following.

(C.1) Number of observations in any unit square in $\mathcal{D}_n$ is in between $[C_1\frac{n}{\lambda_n^2}, C_2\frac{n}{\lambda_n^2}]$ for some positive constants C_1 and C_2 such that $C_1 < C_2$.

(C.2) Number of non-noisy observations in any unit square of the sampling region is bounded below by $C_1 q_e n \lambda_n^{-2}$.

(C.3) The covariate process $\mathbf{x}(\mathbf{s})$ is such that, $\mathbf{x}(\mathbf{s}) = \mathbf{x}_0(\lambda_n^{-1}\mathbf{s})$ where $\mathbf{x}_0 : [0,1]^2 \to \mathbb{R}^p$ is a differentiable function with bounded partial derivatives in $(0,1)^2$ and $\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}(\mathbf{s}_i)\|^2 = O(1)$ for any set of $\{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathcal{D}_n$ and for any $n \geq 2$.

(C.4) $\{\epsilon(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^2\}$ is a second-order stationary random field such that, $\int_{\mathbb{R}^2} |\rho_\epsilon(\mathbf{h})| d\mathbf{h} < \infty$; and for some $\kappa > 0$, $E|\epsilon(\mathbf{0})|^{2+\kappa} < \infty$ and $\alpha_\epsilon(u,v) \leq Cu^{-\nu_1}v^{\nu_2}$ with $\nu_1 > 4/\kappa$ and $\nu_2 > 0$.

(C.5) $\delta_n^{-1} + \lambda_n^{-1}\delta_n + \sqrt{n}\lambda_n^{-1}\delta_n^{-1} \to 0$.

(C.6) $\sup_n \sup_{i\in\{1,\ldots,n\}} |\xi^p_{n(i)}| < \infty$ for all $p \in \{0.25, 0.75\}$.

(C.7) For all $i \in \{1,\ldots,n\}$, $j \in \{1,\ldots,n(i)\}$ and $p \in \{0.25, 0.5, 0.75\}$, $F_{i_j}(\cdot)$ is absolutely continuous in an neighborhood of $\xi^p_{n(i)}$, i.e. $f_{i_j}(x) = (d/dx)\, F_{i_j}(x)$ exists in that neighborhood.

(C.8) For all $i \in \{1,\ldots,n\}$, $\quad 0 < \inf_{j\in\{1,\ldots,n(i)\}} f_{i_j}(\xi^p_{n(i)}) \leq \sup_{j\in\{1,\ldots,n(i)\}} f_{i_j}(\xi^p_{n(i)}) < \infty$.

(C.9) $F_i^{-1}(0.5) = 0$ for all $i \in \{1,\ldots,n\}$.

In a later remark we have discussed how the conditions would be modified in case of consistency of the Mean-VS estimator, i.e. $\hat{\boldsymbol{\beta}}^{(a)}_{\text{vs}}$. We now comment briefly on the conditions. Assumption (C.1) states that the sampling design has to be such that the number of observations in any unit square varies as the same rate as of n/λ_n^2, the 'average' number of observations per unit square. In condition (C.2), we restrict the number of noisy observations in any unit square by $C_1(1-q_e)n\lambda_n^{-2}$ which combining with (C.1) tells that at most $1-q_e$ proportion of the observations in an unit block can be noisy, i.e. corresponding noise variances (denoted by τ_i^2) are non-zero. Assumption (C.3) puts regularity conditions on the covariates; it states that the covariate process is an 'inflated' version of a 'smooth' process $\mathbf{x}_0(\mathbf{s})$ in the prototype sampling region $[0,1]^2$, making $\mathbf{x}(\mathbf{s})$ a slowly varying process over the space. The other condition in (C.3) is standard in regression theory which puts a bound on the magnitude of the covariates as compared to the sample size. (C.4) states the required moment and mixing conditions on the residual process $\{\epsilon(\mathbf{s})\}$. (C.5) makes sure that the size of the blocks used to compute the VS increases resulting in increasing number of observations for VS computation as n increases but at the same time the second term of the condition states that the block has to be small

enough as compared to the whole sampling region retaining the 'local' feature of VS. The third term in (C.5) states that the size of the neighborhood for computation of VS has to grow at faster rate than the rate in which the samples are infilling: this condition is required to prove the consistency of the sample quantiles computed in the neighborhoods. Conditions (C.6) - (C.8) are standard assumptions needed for quantile consistency of non-i.i.d. random variables (for details, see Sen 1968 and Ghosh 1971). (C.9) is needed to make the $w(\mathbf{s}_i)$'s marginal population median is equal to 0.

We now state the consistency results. In the following proposition we provide the asymptotic approximation for the Median-VS.

Proposition 3.1. *Under conditions (C.1) to (C.9), for all* $i \in \{1, 2, \ldots, n\}$,

$$V^{(m)}(\mathbf{s}_i) = \exp\left(-\frac{|w(\mathbf{s}_i)|}{\mathcal{I}_n(\mathbf{s}_i)}\right) + O_p(a_n), \tag{3.3.5}$$

where, $\mathcal{I}_n(\mathbf{s}_i) = \xi^{0.75}_{n(i)} - \xi^{0.25}_{n(i)}$, *i.e. the IQR of the distribution function* $\bar{F}_{n(i)}(x)$ *and* a_n *is a sequence of positive reals* $\downarrow 0$ *given by,* $a_n = n^{-1/2}\lambda_n\delta_n^{-1} + \delta_n\lambda_n^{-1}$.

Before we comment on the importance of Proposition 3.1 let us state the following lemma.

Lemma 3.1. *Under conditions (C.1) to (C.9), for all* $i \in \{1, \ldots, n\}$

$$C^{(l)}_\epsilon(q_e) \leq \mathcal{I}_n(\mathbf{s}_i) = \xi^{0.75}_{n(i)} - \xi^{0.25}_{n(i)} \leq C^{(u)}_\epsilon(q_e),$$

where, $C^{(l)}_\epsilon(q_e) = \sigma_\epsilon\left(F^{-1}_\epsilon(max\{1-(0.25/q_e), 0\}) - F^{-1}_\epsilon(min\{0.25/q_e, 1\})\right)$ *and* $C^{(u)}_\epsilon(q_e) = \sigma_\epsilon\left(F^{-1}_\epsilon(min\{0.75/q_e, 1\}) - F^{-1}_\epsilon(max\{1-(0.75/q_e), 0\})\right)$.

From Lemma 3.1 we conclude that the scaling term, $\mathcal{I}_n(\mathbf{s}_i)$, in the Median-VS approximation in Equation 3.3.5 is bounded above by $C^{(u)}_\epsilon(q_e)$, which is a finite constant

if $q_e > 0.75$, i.e. the proportion of noisy observations is less than 25%. Hence, under the assumption that $q_e > 0.75$, from Proposition 3.1 we can see that if the noise variance associated with $w(\mathbf{s}_i)$ – denoted by τ_i^2 – is large, with higher probability $|w(\mathbf{s}_i)|$ will take larger values making $\exp\left(-|w(\mathbf{s}_i)|/\mathcal{I}_n(\mathbf{s}_i)\right)$ (from here denoted by $\tilde{V}^{(m)}(\mathbf{s}_i)$) – which is bounded above by $\exp\left(-|w(\mathbf{s}_i)|/C_\epsilon^{(u)}(q_e)\right)$ – closer to 0; whereas if $w(\mathbf{s}_i)$ has less variance then with higher probability $|w(\mathbf{s}_i)|$ will take values in neighborhood of 0 making $\tilde{V}^{(m)}(\mathbf{s}_i)$ closer to 1. Also, note that the lower bound $C_\epsilon^{(l)}(q_e) > 0$ if we have $q_e > 0.5$. So, under a set of standard regularity conditions, Proposition 3.1 asymptotically justifies our aim for defining VS: observations with large noise-variance will have VS 'close' to 0 and for observations with little or no noise variance will have VS close to 1.

Next, we use the approximation in Proposition 3.1 to come up with a representation of the VS-based mean parameter estimator $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(m)}$ in the following theorem.

Theorem 3.1. *Under conditions (C.1) to (C.9),*

$$\hat{\boldsymbol{\beta}}_{vs}^{(m)} = \boldsymbol{\beta} + \left(\frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w}\right) + O_p\left(a_n\right), \tag{3.3.6}$$

where $\tilde{\mathbf{D}}_v = diag\left(\tilde{V}^{(m)}(\mathbf{s}_1), \ldots, \tilde{V}^{(m)}(\mathbf{s}_n)\right)$ *and* $\mathbf{w} \coloneqq (w(\mathbf{s}_1), \ldots, w(\mathbf{s}_n))'$.

The representation in Equation 3.3.6 approximates the distribution of the deviation of the Median-VS-based mean parameter estimator $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(m)}$ from it's target value $\boldsymbol{\beta}$ by the random variable given by $\left(\frac{1}{n}\mathbf{X}'E\left(\tilde{D}_v\right)\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\tilde{D}_v\mathbf{w}\right)$. This approximation is used to prove the consistency of $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(m)}$ in the following corollary. It is also used for the efficiency comparison of the VS-based estimation with that of the standard OLS method. Finally, for the consistency of the VS-based estimator we need the following additional conditions.

(C.10) $\frac{1}{n}\mathbf{X}'\mathbf{X} \to \mathbf{C}_X \succ 0$.

(C.11) $E\left(\tilde{D}_v \mathbf{w}\right) = \mathbf{0}$.

(C.12) The proportion of non-noisy observations, i.e. q_e is strictly greater than 0.75.

(C.13) For any real number $a > 0$, $\psi_\epsilon(a) \coloneqq \int_0^\infty e^{-x} P\left(|\epsilon(\mathbf{s})| < ax\right) dx > 0$.

Condition (C.10) is standard assumption needed for the consistency of the least squares based mean parameter estimators. (C.11) is trivially valid if the $w(\mathbf{s}_i)$'s are marginally symmetric around 0. (C.12) is needed because in Lemma 3.1, we have seen that to get a finite upper-bound on the numerator of $\tilde{V}^{(m)}(\mathbf{s}_i)$ we need the proportion of noisy observations to be less than 0.25. (C.13) is required to ensure the invertibility of the matrix $\left(n^{-1}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)$. This condition is trivially true when the marginal distribution of $\epsilon(\mathbf{s})$ has positive mass on a neighborhood around its center 0.

Corollary 3.1. *Under conditions (C.1) to (C.13),* $\hat{\boldsymbol{\beta}}_{vs}^{(m)} = \boldsymbol{\beta} + O_p\left(a_n\right)$.

Corollary 3.1 not only proves the consistency of the Median-VS-based estimator but also provides an upper bound on its rate of convergence in probability. The proofs of all the results stated in this section are given in Section 3.7.

Remark 3.1. Similar consistency results of Mean-VS-based estimator $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(a)}$ can be proved but that requires a set of different conditions than that are stated in (C.1) - (C.9). As the sample standard deviation is used in the definition of Mean-VS we need $\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}(\mathbf{s}_i)\|^4 = O(1)$ as well as all the error components including $\epsilon(\mathbf{s})$ must have finite fourth order moments. In addition, the mixing condition in (C.4) changes as we would require ϵ-process to be fourth-order stationary and all the correlations of the form $\text{cor}\left(\epsilon^a(\cdot), \epsilon^b(\cdot + \mathbf{h})\right)$, for all $a, b \in \{1, \ldots, 4\}$, have to be integrable over the space. Conditions (C.5) - (C.9) are not needed as those are specifically for the existence of the quantiles used in defining the Median-VS-based estimator.

Though the consistency of the VS-based estimators has been proved in this section, that does not necessarily establish the advantage of VS-based estimation over the standard OLS method when it comes to analyzing noisy geostatistical data. In the next subsection, we have investigated the asymptotic efficiency of the VS-based estimators with respect to that of the OLS estimator by comparing their approximated mean squared errors.

3.3.4 Asymptotic efficiency of VS-based estimators

The results of this section is only for the Median-VS-based regression estimator and hence, for simplicity of notations we denote it as $\hat{\boldsymbol{\beta}}_{\text{vs}}$ instead of $\hat{\boldsymbol{\beta}}_{\text{vs}}^{(m)}$. From the representation in Theorem 3.1, we can write $\hat{\boldsymbol{\beta}}_{\text{vs}} - \boldsymbol{\beta} = \mathbf{l}_n^{\text{vs}} + O_p(a_n)$ where,

$$\mathbf{l}_n^{\text{vs}} = \left(n^{-1}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)^{-1}\left(n^{-1}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w}\right)$$

is the leading term of the error associated with VS-based estimator. On the other hand, due to linearity of OLS estimator, $\hat{\boldsymbol{\beta}}_{\text{ols}} - \boldsymbol{\beta} = \mathbf{l}_n^{\text{ols}}$, where it can be easily shown that $\mathbf{l}_n^{\text{ols}} = (n^{-1}\mathbf{X}'\mathbf{X})^{-1}(n^{-1}\mathbf{X}'\mathbf{w})$. The results in this section show that, $E\left(\|\mathbf{l}_n^{\text{vs}}\|^2\right)$ can be bounded above by a quantity that is asymptotically independent of the noise variances (denoted by τ_i^2) associated with the corrupted observations, whereas $E\left(\|\mathbf{l}_n^{\text{ols}}\|^2\right)$ is bounded below by a quantity that is a linear function of the noise variances.

Before stating the results of this section, we need to introduce a few more notations. It can be shown easily that $E\left(\|\mathbf{l}_n^{\text{ols}}\|^2\right) = n^{-1}\text{tr}\left(\mathbf{H}_n\boldsymbol{\Sigma}_w\right)$ (details are given in Section 3.7), where $\mathbf{H}_n := n^{-1}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X})^{-2}\mathbf{X}'$, $\boldsymbol{\Sigma}_w := \text{Var}\left(\mathbf{w}\right)$. For any symmetric matrix $\mathbf{A}$, by $\lambda_{\min}\left(\mathbf{A}\right)$, we denote the minimum eigenvalue of $\mathbf{A}$. For simplicity, let us introduce the following notations: $M_X := \sup_{\mathbf{s}\in\mathcal{D}_n}\max_{1\le j\le p} x_j^2(\mathbf{s})$; $C^{(0)} := 2^{-1}\text{tr}\left(\mathbf{C}_X^{-1}\right)$;

$C^{(1)} \coloneqq 2p\left(\lambda_{\min}\left(\mathbf{C}_X\right)\right)^{-2} M_X$; and finally $\mathcal{M}_n(\varepsilon) \coloneqq \{1 \leq i \leq n : (\mathbf{H}_n)_{ii} > n^{-1}\varepsilon\}$, where by $(\mathbf{A})_{ij}$ we denote the ij-th element of the matrix $\mathbf{A}$. Now we state the main results of this subsection.

Theorem 3.2. *Let assumptions (C.1) to (C.13) hold true. Then, for positive constants $C_1(q_e)$ and C_2,*

$$E\left(\|\mathbf{l}_n^{vs}\|^2\right) = O\left(n^{-1}\left(C_1(q_e)\right)^2 + C_2\lambda_n^{-4}\right), \tag{3.3.7}$$

where $C_1(q_e) = \left(C_\epsilon^{(u)}(q_e)/\psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)\right)$ and C_2 is independent of the noise variance parameters (i.e. τ_i^2 for $i \in G_n^c$ and q_e).

Theorem 3.3. *Let assumptions (C.1) to (C.4) and (C.10) hold true. Then, for any $0 < \varepsilon < min\left(C^{(0)}, C^{(1)}\right)$ and large enough n,*

$$E\left(\|\mathbf{l}_n^{ols}\|^2\right) > n^{-1}\varepsilon\left(\sigma_\epsilon^2\left(C^{(0)} - \varepsilon\right)\left(C^{(1)} - \varepsilon\right)^{-1} + n^{-1}\sum_{i \in \mathcal{M}_n(\varepsilon) \cap G_n^c} \tau_i^2\right) - C_1\lambda_n^{-4}, \tag{3.3.8}$$

with $|\mathcal{M}_n(\varepsilon)| \geq \left(\frac{C^{(0)}-\varepsilon}{C^{(1)}-\varepsilon}\right) n$. Here, C_1 is a positive constant not dependent on the noise variance parameters.

Corollary 3.2. *Let the assumptions (C.1) - (C.4) and (C.10) hold true. In addition, assume $inf_n \min\limits_{1 \leq i \leq n} \|\mathbf{X}[i,]\|^2 > 0$. Then, there exists an $\varepsilon > 0$ such that for large enough n,*

$$E\left(\|\mathbf{l}_n^{ols}\|^2\right) > n^{-1}\left(\varepsilon\sigma_\epsilon^2 + n^{-1}\sum_{i \in G_n^c} \tau_i^2\right) - C_1\lambda_n^{-4}, \tag{3.3.9}$$

for some positive constant C_1 not dependent on the noise model parameters.

From Theorem 3.2 we see that the MSE of the leading error term in the Median-VS-based regression estimation can be bounded above by $n^{-1}\left(C_1(q_e)\right)^2 + C_2\lambda_n^{-4}$ which goes to 0 as $n \to \infty$ and is dependent on the noise model parameters only through the proportion of the good observations q_e. Note that, $\psi_\epsilon(\cdot)$ is an increasing function. We have already mentioned that as q_e increase, $C_\epsilon^{(u)}(q_e)$ decreases and $C_\epsilon^{(l)}(q_e)$ increases, that is $C_1(q_e)$ decreases restricting the range of $E\left(\|\mathbf{l}_n^{\text{vs}}\|^2\right)$ from above. The other term, $C_2\lambda_n^{-4}$ has hardly any effect on $E\left(\|\mathbf{l}_n^{\text{vs}}\|^2\right)$ as λ_n^{-4} is of much smaller order as compared to n^{-1}. Clearly, increasing the noise variances, i.e. τ_i^2 does not have high impact on the MSE of the Median-VS-based regression estimator. On the contrary, the lower-bound for the MSEs of the OLS estimator, as we can see in Theorem 3.3 and Corollary 3.2, is dependent on the additional noise variances of the corrupted observations through the summation $n^{-1}\sum_{i\in G_n^c\cap\mathcal{M}_n(\varepsilon)}\tau_i^2$ and $n^{-1}\sum_{i\in G_n^c}\tau_i^2$ respectively. For example, if we assume that for all $i \in \{1,\ldots,n\}$, $\tau_i^2 \approx \tau^2$, then $n^{-1}\sum_{i\in G_n^c}\tau_i^2 \approx (1-q_e)\tau^2$. Clearly, MSE of the OLS estimator can be hampered heavily if the noise variances associated with some of the observations is arbitrarily 'large', whereas the MSE of the VS-based estimator is bounded above by a quantity which is independent of the noise variance and thus VS-based estimation is more efficient to estimate the regression parameters from noisy spatial data. The proofs of Theorem 3.2, Theorem 3.3 and Corollary 3.2 are given in Section 3.7.

3.4 Simulation Study

In this section, we study the accuracy of the VS-based mean parameter estimators under the multiplicative-additive noise model as described in Section 3.3.1 and compare its performance with two of the most commonly used least squares-based estimators: ordinary

least squares (OLS) and the generalized least squares (GLS).

3.4.1 Simulation setup

We use the following spatial regression model to simulate the realizations from the the 'true' random field $\{Y(\mathbf{s})\}$:

$$Y(\mathbf{s}_i) = \beta_0 + (\beta_x, \beta_y)' \mathbf{s}_i + \beta_h \, h(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{3.4.1}$$

where $\boldsymbol{\beta} \coloneqq (\beta_0, \beta_x, \beta_y, \beta_h)'$ is the vector of regression parameters; $h(\mathbf{s})$ is a deterministic function of location $\mathbf{s}$; and $\{\epsilon(\mathbf{s})\}$ is a second-order stationary spatially correlated process. To define the function $h(\mathbf{s})$ over the sampling region, we use the deterministic function $h(\mathbf{s}) = H_1 \cdot \sum_{j=1}^{H_2} w_h(j) f(\mathbf{s};\ \boldsymbol{\mu}_j, \Sigma_j) + H_3$, where $f(\cdot; \boldsymbol{\mu}, \Sigma)$ denotes the bivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix Σ and $\{(\boldsymbol{\mu}_j, \Sigma_j) :\ j \in \{1, \ldots, H_2\}\}$ are fixed set of vectors and matrices. The choice of this function is motivated from the elevation function used in the simulations of Chapter 2. The residual vector $(\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_n))'$ are sampled from a second-order stationary mean-zero Gaussian process with the following covariance function:

$$C^{\exp}(\mathbf{h}; \boldsymbol{\theta}) = \sigma_\epsilon^2 \exp\left(\frac{\|\mathbf{h}\|}{\rho}\right) + \tau^2 \mathbb{I}(\mathbf{h} = \mathbf{0}), \tag{3.4.2}$$

where, $\mathbb{I}(\cdot)$ denotes the indicator function. The covariance parameter vector of interest is $\boldsymbol{\theta} = (\tau^2, \sigma_\epsilon^2, \rho)'$, where τ^2 is the nugget effect, σ_ϵ^2, ρ are the partial sill and range parameters respectively (for details see Haskard 2007; page 37 Gelfand et al. 2010).

To include noise for the varying-quality observations $(Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))'$ as in Equation 2.2.3, we use the following distributions for the additive and multiplicative com-

ponents, denoted by $\boldsymbol{\epsilon}_A \coloneqq (\epsilon_{A_1}, \ldots, \epsilon_{A_n})'$ and $\boldsymbol{\epsilon}_M \coloneqq (\epsilon_{M_1}, \ldots, \epsilon_{M_n})'$ respectively. For $i \in G_n^c$,

$$\epsilon_{M_i} \underset{\text{indep.}}{\sim} 2 \times \text{Beta}(\alpha_M,\alpha_M);\ \epsilon_{A_i} \underset{\text{indep.}}{\sim} N(0, \sigma_A^2);\ \{\epsilon_{M_i}\}_{i=i}^{\infty} \text{and} \{\epsilon_{A_i}\}_{i=i}^{\infty} \text{ are independent,} \tag{3.4.3}$$

where $\sigma_M^2 = \frac{1}{2\alpha_M+1}$. To select the set of 'good' observations we have randomly selected $\lfloor q_e n \rfloor$ many indices among $\{1, \ldots, n\}$ and defined that set of indices to be G_n. The choice of multiplicative error distribution in (3.4.3) restricts its realizations to be in $[0, 2]$ and also ensures that the multiplicative errors are symmetric around 1.

Estimation of the regression parameters through VS-based or the OLS methods does not take the covariance between the observations into account whereas the GLS estimator uses the covariance information to make the mean parameter estimation more efficient. In practice, the covariance parameters are unknown and hence needs to be estimated from the spatial observations and then the estimated covariance parameters are used to compute GLS estimator. Then the updated mean estimates are used to compute the covariance parameters and we continue until the parameter estimates have converged. This estimator is often called the estimated GLS (EGLS, for details see page 23 in Cressie 1993). To estimate the covariance parameters in computation of the EGLS estimator we use the weighted least squares-based variogram estimation (for details, see Chapter 2 in Cressie 1993).

In Section 3.5, we demonstrate the VS-based methodology – mean estimation, covariance estimation and prediction – by implementing on the coalash data. To estimate the covariance parameters robustly in the analysis of coalash data – where some of the observations are outliers – we use the VS-based covariance estimation as discussed in

Chapter 2 and mentioned again in Section 3.2. To justify the use of VS-based variogram estimation methodology, in addition to the simulation results given in Chapter 2, we conduct an extensive simulation study in this chapter as well. Theoretical justification of the VS-based variogram estimation needs further research and is beyond the scope of this study.

For all the simulation results in this section, we take $\boldsymbol{\beta} = (70, 5, -2, -0.05)'$ and $\boldsymbol{\theta} = (0, 6, 1)'$. To investigate the performances of the different types of estimators with varying noise level, we consider the following set of values of the noise parameters: (a) $\sigma_A \in \{5, 50, 100\}$, (b) $\alpha_M \in \{2, 0.5, 0.05\}$, and (c) $q_e \in \{0.95, 0.85, 0.75\}$. The set of values for each of the above parameters are written in the order of increasing noise variance and proportion. For example, if the first values of all the three sets are considered then the additional noise variance of a corrupted observation at location $\mathbf{s}$ is $0.2\left(\mathbf{x}(\mathbf{s})'\boldsymbol{\beta}\right)^2 + 26.2$, and the proportion of such observations is 5%; whereas the same variance will be $0.91\left(\mathbf{x}(\mathbf{s})'\boldsymbol{\beta}\right)^2 + 10005.46$ if the last values of the sets are considered and the proportion of such noisy observations will be 25%.

3.4.2 Results

Table 3.1 shows the empirical mean square errors (MSE $= B^{-1}\sum_{b=1}^{B}\|\hat{\boldsymbol{\beta}}_b - \boldsymbol{\beta}\|^2$) of the two VS-based estimators – Median VS and Mean VS – and two least squares-based estimators – ordinary least squares (OLS) and estimated generalized least squares (EGLS). In the final column of Table 3.1, the relative efficiency (RE) of the Median-VS-based regression estimator with respect to the OLS one is reported to emphasize the superiority of VS-based estimation in the analysis of noisy spatial data.

It is evident from Table 3.1 that VS-based estimation has performed better than the

Table 3.1: Empirical mean squared errors (MSE) of the VS-based and least squares (LS) based regression parameter estimator; relative efficiency (R.E.) of the Median-VS estimator with respect to that of OLS estimator – for varying noise model parameters $(\sigma_A, \sigma_M, q_e)$ and varying sample sizes (n). For each sub-table, the other noise parameters are fixed at their first value, e.g. for the sub-table (top) with varying σ_A, the other parameters are fixed at $\sigma_m = 0.447$ and $q_e = 0.95$.

		VS-based MSE		**LS-based MSE**		**RE(Med-VS, OLS)**
σ_A	**n**	**Med-VS**	**Mean-VS**	**OLS**	**EGLS**	**MSE(OLS)/MSE(Med-VS)**
	500	1.807	1.959	3.237	5.620	1.791
5	1000	1.166	1.184	1.662	3.416	1.425
	5000	0.567	0.574	0.746	1.436	1.317
	500	1.806	1.906	4.783	14.862	2.648
50	1000	1.134	1.165	2.534	6.112	2.234
	5000	0.701	0.693	1.055	2.114	1.505
	500	2.350	2.905	12.938	36.208	5.505
100	1000	0.969	1.102	7.817	13.895	8.067
	5000	0.618	0.595	1.618	6.064	2.619
		VS-based MSE		**LS-based MSE**		**R.E.(Med-VS, OLS)**
σ_M	**n**	**Med-VS**	**Mean-VS**	**OLS**	**EGLS**	**MSE(OLS)/MSE(Med-VS)**
	500	1.889	1.945	2.771	4.481	1.467
0.447	1000	1.005	1.033	1.559	2.942	1.552
	5000	0.572	0.578	0.699	1.352	1.220
	500	2.219	2.253	4.098	13.248	1.847
0.707	1000	1.063	1.117	2.685	6.031	2.525
	5000	0.647	0.641	0.919	2.527	1.420
	500	2.221	2.422	6.182	20.996	2.784
0.953	1000	1.201	1.206	3.578	9.048	2.978
	5000	0.669	0.671	1.476	4.746	2.206
		VS-based MSE		**LS-based MSE**		**R.E.(Med-VS, OLS)**
q_e	**n**	**Med-VS**	**Mean-VS**	**OLS**	**EGLS**	**MSE(OLS)/MSE(Med-VS)**
	500	2.121	2.160	3.080	5.697	1.452
0.95	1000	1.121	1.110	1.447	2.527	1.291
	5000	0.742	0.741	0.836	1.281	1.126
	500	1.783	2.026	4.274	8.825	2.397
0.9	1000	1.129	1.177	2.362	5.475	2.092
	5000	0.634	0.635	0.833	2.288	1.314
	500	2.017	2.351	6.074	15.545	3.011
0.8	1000	1.087	1.267	3.567	9.290	3.283
	5000	0.550	0.565	1.105	4.618	2.010

least squares-based methods across all scenarios. Moreover, the increasing noise variance has very little effect on the performance of the VS-based estimation in terms of MSE. For instance, consider the case when $n = 500$. If the standard deviation of the associated additive noise is increase from 5 to 100, the MSE of Median-VS-based estimator has only increased by 30%, whereas the MSE of the EGLS and OLS estimators have increased by approximately 546% and 300% in the respective order. The consistency of the VS-based estimator is also prominent from Table 3.1 – in each of the scenarios, as sample size

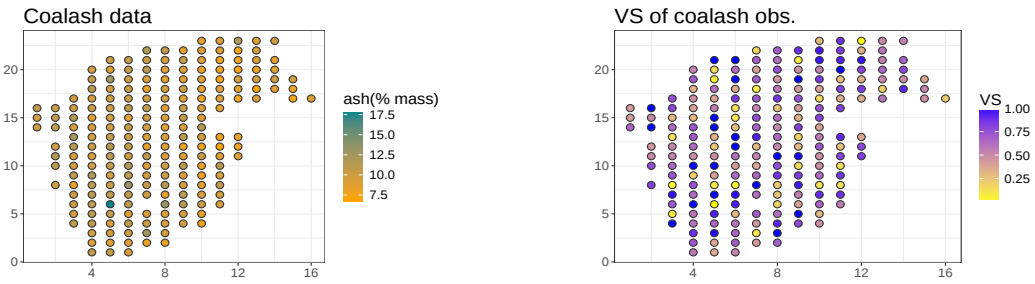
increases, the MSE decreases to 0. The relative efficiency column justifies the advantage of the VS-based estimator to the least squares-based estimators. For example, when $\sigma_A = 100$, i.e. very high noise variance is associated with 5% of the observations – to achieve same accuracy as of the Median-VS-based estimators the OLS and the EGLS methods need approximately 8 and 14.5 times more observations. Also, given a sample size, the relative efficiency increases if the magnitude of the noise variances associated with the 'bad' observations increase (controlled by σ_M and σ_A) or the proportion of the 'bad' observations increases.

As we can see in Table 3.1, the accuracy is hampered severely if instead of the OLS estimator EGLS is used to estimate the regression parameter. One of the possible reasons behind this is inefficient estimation of the covariance parameters using weighted least squares-based methods. In Table 3.2, we have investigated the performance of the VS-based covariance estimation and compared it with that of the standard approach in practice: weighted least squares (WLS) based variogram estimator (for details, see Cressie 1993). Clearly, the WLS approach has failed completely to estimate the covariance parameters – partial sill (or psill, denoted by σ_ϵ^2) and range (denoted by ρ). Due to the added noise variance in the data the optimization in the variogram fitting in WLS method is not stable. On the other hand, VS-based variogram estimation has reasonably good accuracy, given that the number of non-noisy observations is not too 'large', e.g. consider the case when the sample size is 5000 with 5% of the observations are corrupted. Also, in most of the cases, if the noise model parameters are held fixed, the MSE of the VS-based estimators decreases to 0, as sample size $n \to \infty$, which justifies the consistency of the VS-based variogram estimation. But, in the case where the proportion of noisy observations is high the efficiency of the VS-based estimation has hampered severely – for instance, when the proportion of non-noisy observations has decreased from 90% to 80%,

Table 3.2: Empirical mean squared errors (MSE) of the VS-based and weighted least squares (WLS) based covariance parameter (psill : σ^2_ϵ, range : ρ) estimator for varying noise model parameters (σ_A, σ_M, q_e) and varying sample sizes (n). For each sub-table, the other noise parameters are fixed at their first value, e.g. for the sub-table (top) with varying σ_A, the other parameters are fixed at $\sigma_m = 0.447$ and $q_e = 0.95$.

		psill MSE		**range MSE**	
σ_A	**n**	**VS**	**WLS**	**VS**	**WLS**
5	500	19.707	2058.760	2.471	3.099
	1000	1.255	1992.875	0.123	2.697
	5000	0.532	609.690	0.032	17.537
50	500	36.275	1334.724	4.549	0.580
	1000	1.800	1820.486	0.146	0.639
	5000	0.757	1163.299	0.046	11.401
100	500	45.653	682.913	9.477	0.502
	1000	1.675	913.053	0.186	0.506
	5000	0.933	1802.586	0.060	10.526
		psill MSE		**range MSE**	
σ_M	**n**	**VS**	**WLS**	**VS**	**WLS**
0.447	500	7.712	1996.900	0.928	0.757
	1000	1.401	1863.389	0.164	5.779
	5000	0.580	577.577	0.047	6.312
0.707	500	14.324	2045.811	2.097	13.983
	1000	1.663	1566.315	0.170	4.272
	5000	0.820	792.043	0.038	8.928
0.953	500	28.337	1578.990	3.621	54.341
	1000	1.507	1537.173	0.201	16.551
	5000	1.134	1185.805	0.091	1.645
		psill MSE		**range MSE**	
q_e	**n**	**VS**	**WLS**	**VS**	**WLS**
0.95	500	11.786	1803.304	1.302	2.062
	1000	1.230	2385.300	0.146	2.884
	5000	0.665	531.284	0.047	12.393
0.9	500	5.216	2318.057	0.906	5.762
	1000	1.292	1535.030	0.098	5.707
	5000	1.063	817.137	0.026	2.803
0.8	500	36.674	1745.520	1.865	0.658
	1000	6.300	1308.637	0.116	13.993
	5000	11.844	1428.831	0.049	0.790

the MSE of the VS-based partial sill estimator has increased by approximately 350% for both $n = 1000$ and 5000. Clearly, the accuracy of the VS-based covariance estimation is highly sensitive to the proportion of noisy observations in the data – the same fact that we have established theoretically for the VS-based mean estimation in Section 3.3.4. As mentioned previously in Chapter 2, we have reported the results of our comparative analysis of VS-based technique and one of the other existing robust covariance estimation methodology in geostatistics – robust REML proposed by Künsch et al. (2011) in Section 5.1.3.2. The results are surprisingly in favor of the VS-based methodology – both in



(a) Ash % in coal seams (b) VS of the observations

Figure 3.1: Spatial plots of the coalash data (a) and VS of the observations (b).

terms of computational complexity and the accuracy of estimation. In the next section, we are going to apply VS-based estimation technique on the coalash data (Gomez and Hazen 1970) and compare our results with that of Papritz (2018b), where the author has implemented robust REML estimation on the same dataset.

3.5 Example: coalash data

This section provides a demonstration of the VS-based estimation and prediction by implementing it on a coalash data (Gomez and Hazen 1970) which reports measurement (% mass) of ash present in coal seam in Pennsylvania. The dataset is available in R-package gstat (Pebesma 2004) and has been used to demostrate robust geostatistical techniques by Cressie (1993) and Papritz (2018b) previously. Let us denote the target process as $\{Y(\mathbf{s})\}$ where $Y(\mathbf{s})$ is the % mass of ash in the coal seam at location $\mathbf{s}$. The observations plotted in Figure 3.1a are denoted by $\{Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)\}$ where $n = 208$ – some these observations are corrupted due to measurement errors or other types of unknown noises.

The data has been plotted spatially in Figure 3.1a. Next, the VS of the observations is computed and plotted in Figure 3.1b. Cressie (1993) identified a set of observations

as outliers by investigating the deviation of the observations from their overall sample median. VS uses the similar idea – instead of using the overall median, it assigns a score to each of the observations according to its deviation from the 'local' median relative to the 'local' variation. The observations that were identified as outliers by Cressie (1993) (mention in Papritz 2018b as well) received VS less than 0.18 (in a scale of $(0, 1]$). As both in theoretical analysis and simulation studies we have established that VS-based estimation is hampered if proportion of noisy observations is significantly large, we check the proportion of observations which have VS less than 0.22 – which implies the deviation from local median is at least 1.51×local-IQR and we find that approximately 12% of the observations have 'high' noise associated with it.

Once the VS of the observations has been computed we incorporate those score to estimate the mean and covariance structure of the process robustly. Through some robust explanatory analysis as proposed by Cressie (1993), we have considered the following spatial regression model for the target process $\{Y(\mathbf{s})\}$,

$$Y(\mathbf{s}) = \beta_0 + \beta_x s_x + \epsilon(\mathbf{s}), \tag{3.5.1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_x)'$ is the regression parameter, s_x is longitude (i.e. $\mathbf{s} = (s_x, s_y)'$) and $\{\epsilon(\mathbf{s})\}$ is a zero mean second order stationary process with Matérn covariance and the covariance parameter vector is given by $\boldsymbol{\theta} = (\sigma^2_\epsilon, \rho, \eta^2, \nu)'$ where σ^2_ϵ is the partiall sill, ρ is the range, η^2 is the nugget and ν is the smoothness parameter (for details, see Haskard 2007; Gelfand et al. 2010).

VS-based regression yields an estimate of $\hat{\boldsymbol{\beta}}_{\text{vs}} = (11.071, -0.188)'$ for the regression parameter – which suggests that the ash proportion decreases as we go from west to east. Next, the VS-based smoothing is implemented on the observed residuals from

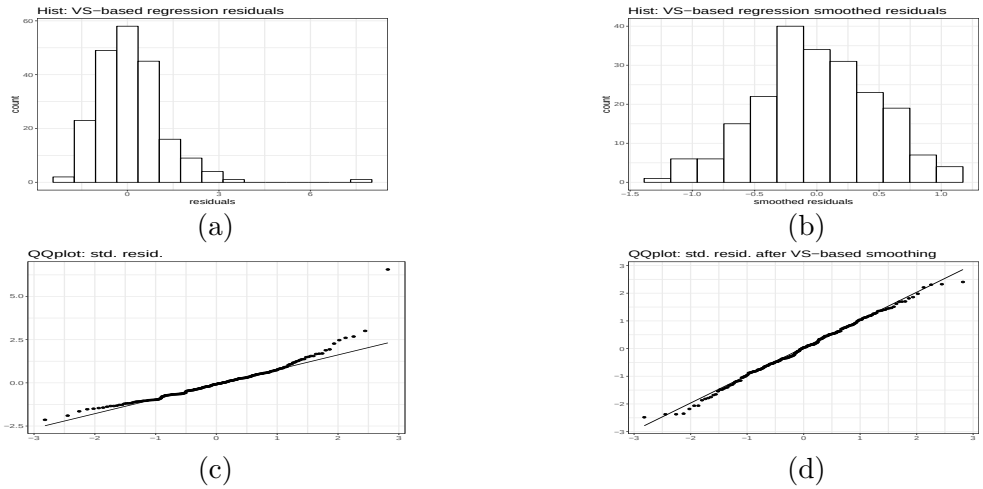


Figure 3.2: VS-based smoothing of residuals: histogram of observed residuals from VS-based regression (a), histogram of smoothed residuals (b), qqplot of observed residuals (c), qqplot of smoothed residuals (d).

the VS-based regression and the effect of smoothing has been displayed through frequency plots and QQplots in Figure 3.2. Finally, the VS-based smooth residuals are used to estimate the covariance parameters through VS-based variogram model fitting. We summarize the details of VS-based covariance estimation in Figure 3.3 and Table 3.3.

Table 3.3: Estimated Matérn parameters.

Parameters	**Estiamtes**
partial sill (σ^2_ϵ)	0.219
range (ρ)	0.486
nugget (η^2)	0.021
smoothness (κ)	5.00

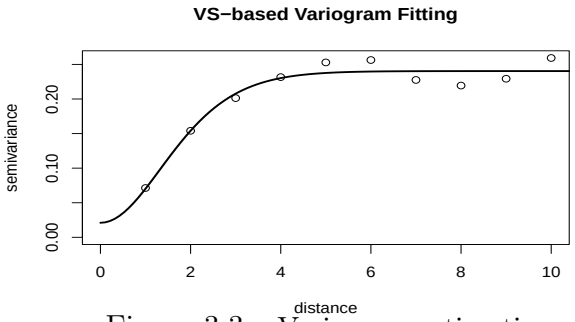


Figure 3.3: Variogram estimation

Next, we compare the VS-based analysis with the robust-REML approach proposed by Künsch et al. (2011). To implement the robust-REML methodology on the coalash data R-package 'georob' (Papritz 2018a) has been used. We used leave-one-out cross-validation technique to compare the above-mentioned two approaches. In a previous analysis of this

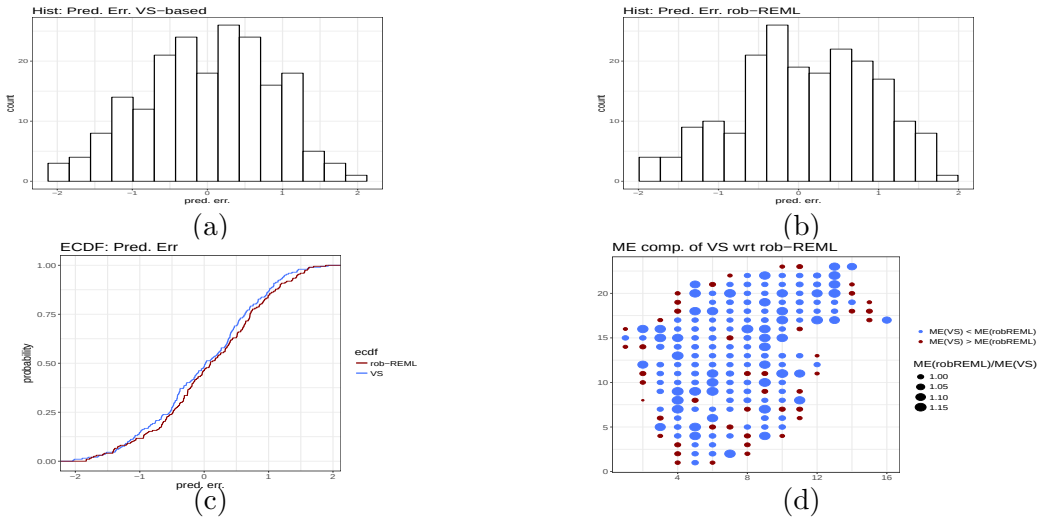


Figure 3.4: Prediction comparison between VS and robust-REML: histogram of prediction errors of VS-based (a) and robust-REML (b) approach; empirical c.d.f. of prediction errors (c) and relative efficiency in terms of margin of prediction errors of VS w.r.t. robust-REML (d).

dataset, Papritz (2018b) has identified some observations – in addition to the list given by Cressie (1993) – as outliers. Hence, we consider each observation of the coalash data excluding the ones marked as outliers by Papritz (2018b) (11 such observations) as test data. For each of these test data point, we consider all other observations (including those marked as outliers by Papritz (2018b)) as our training set for estimation of the mean and covariance parameters and then we predict (kriging) the value of the process at the test data location – for both the VS-based and the robust REML approach. Instead of using VS, we use the analysis done by Papritz (2018b) to identify the outliers for considering the test dataset so that, the choice of test data in the cross-validated analysis is not biased towards the VS-based approach.

In Figure 3.4, we have summarized the results of comparative analysis between VS-based and robust-REML approach. From the histograms in Figure 3.4a-3.4b and the

empirical distribution function in Figure 3.4c it is evident that the prediction errors corresponding to the robust-REML predictor has higher mass in the tails than the VS-based predictions. Moreover, as we can see in FIgure 3.4d, the margin of prediction error (ME), that is the half of the length of the prediction interval, for the VS-based approach is smaller as compared to the robust-REML for most of the test data points. We find that 98.5% of the test data points were inside the 95% VS-based prediction intervals. In terms of cross-validated mean squared prediction error (MSPE), the VS-based MSPE is 0.716 whereas the robust-REML prediction has an MSPE of 0.743, which is 1.038 times the MSPE of VS-based approach. In addition to the efficient estimation and prediction, in terms of the computational time VS-based approach has a huge advantage as compared to the robust-REML, especially in the analysis of large spatial datasets, as demonstrated in Table 3.4.

Table 3.4: Time comparison between VS-based and robust-REML. Machine configuration: DELL R7425 Dual Processor AMD Epyc 32 core 2.2 GHz machines with 512GB RAM each running 64Bit Ubuntu Linux Version 18.04.

Computation Time	VS-based	robust-REML
Min.	0.45	1.47
1st Qu.	0.54	1.99
Median	0.60	2.18
Mean	0.61	8.24
3rd Qu.	0.67	14.22
Max.	0.93	29.75

3.6 Conclusion

In this chapter, we have investigated the large sample behavior of the proposed veracity scoring technique in geostatistics. The asymptotic approximation of the VS suggests that

VS of an observation is expected to take 'small' value (close to 0) if the noise variance associated with the observation is high and it is going to take larger values if the associated noise variance is less. Moreover, the mixed increasing domain asymptotic analysis investigates the circumstances under which the weighted least squares regression estimator with veracity scores as the corresponding weights – which are data-driven and hence, spatially correlated – is asymptotically consistent and also, provides a probabilistic order on the error of estimation. The consistency result is not restricted to the additive-multiplicative noise model – the proofs can be generalized to other nonstationary noise models as well. Theoretical investigation of the mean squared errors shows that the VS-based estimation is highly robust to the magnitudes of the noise variances associated with the corrupted observations, but the accuracy is affected if the proportion of noisy observations increases. On the other hand, ordinary least squares estimator is highly sensitive to both the proportion of noisy observations as well as the magnitudes of the noise variances. We have provided empirical justification of the consistency of the both VS-based mean and covariance parameter estimators through extensive simulations. Finally, the VS-based approach has been applied on a real dataset and the results of the comparative analysis with robust-REML – another robust approach to analyze geostatistical data – shows that the VS-based approach has advantage both in terms of prediction accuracy and computational time.

With the growing engineering development of sensor-based technologies and consequent growing need for big data analytics, veracity scoring of the observations is becoming an important research area of interest. This chapter provides mathematical foundation of the veracity scoring technique introduced in Chapter 2 for the analysis of noisy geostatistical data. Though the consistency of the VS-based regression estimator has been thoroughly analyzed in this chapter, theoretical justification of the consistency and inves-

tigation of other relevant properties of the VS-based covariance estimation needs further research which will be the focus of our future studies. We will also try to extend the VS for spatio-temporal data by considering a neighborhood in both spatial and temporal domain and implement the methodology on a real-time large sensor-generated data, e.g. GPS sensor data.

3.7 Proofs

Let us define $\hat{G}_{n(i)}(x) = 1 - \hat{F}_{n(i)}(x)$ and $\bar{G}_{n(i)}(x) = 1 - \bar{F}_{n(i)}(x)$. To prove Proposition 3.1 we need the following lemmas.

Lemma 3.2. *The strong-mixing coefficient* $\alpha_w(\cdot,\cdot)$ *of* $\{w(\mathbf{s})\}$ *is bounded above by the strong-mixing coefficient* $\alpha_\epsilon(\cdot,\cdot)$ *of* $\{\epsilon(\mathbf{s})\}$*, i.e.*

$$\alpha_w(u,v) \leq \alpha_\epsilon(u,v),$$

for any $u > 0$ *,* $v > 0$*.*

Proof of Lemma 3.2. Let us denote the strong-mixing coefficient between two σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$ as,

$$\alpha^*(\mathcal{A}_1, \mathcal{A}_2) = \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : \ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}.$$

Let us denote the σ-field generated by two σ-fields $\mathcal{A}_1$ and $\mathcal{A}_2$ by $\mathcal{A}_1 \vee \mathcal{A}_2$. Now consider $\alpha_w(u,v) = \sup\{\alpha^*(\mathcal{F}_w(A), \mathcal{F}_w(B)) : \ d(A,B) \geq u; |A|, |B| \leq v\}$. For simplification of notation, we denote the multiplicative error random variables ϵ_M as $\epsilon_M(\mathbf{s})$ to emphasize that this is the multiplicative error associated with the observed process at location

$\mathbf{s}$. Similarly we denote ϵ_A as $\epsilon_A(\mathbf{s})$ and under this notation we have

$$w(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}\left(\epsilon_M(\mathbf{s}) - 1\right) + \epsilon_m\epsilon(\mathbf{s}) + \epsilon_A(\mathbf{s}).$$

Under the specified error model considered here, both $\{\epsilon_M(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ and $\{\epsilon_A(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$ are independent set of random variables.

For a given $u,\ v > 0$ let fix any two rectangles in $\mathbb{R}^2$ such that $d(A,B) \geq u$ and $|A|\ ,\ |B| \leq v$. Now from the definition of $\{w(\mathbf{s})\}$ it is obvious that $\mathcal{F}_w(A) \subseteq \mathcal{F}_{\epsilon_M}(A) \vee \mathcal{F}_{\epsilon_A}(A) \vee \mathcal{F}_{\epsilon}(A)$ where $\mathcal{F}_{\epsilon_M}(A)$ and $\mathcal{F}_{\epsilon_A}(A)$ are the σ-fields generated by the set of random variables $\{\epsilon_M(\mathbf{s}) : \mathbf{s} \in A\}$ and $\{\epsilon_A(\mathbf{s}) : \mathbf{s} \in A\}$ respectively. So,

$$\begin{aligned}\alpha^*\left(\mathcal{F}_w(A), \mathcal{F}_w(B)\right) &\leq \alpha^*\left(\mathcal{F}_{\epsilon_M}(A) \vee \mathcal{F}_{\epsilon_A}(A) \vee \mathcal{F}_{\epsilon}(A)\,,\ \mathcal{F}_{\epsilon_M}(B) \vee \mathcal{F}_{\epsilon_A}(B) \vee \mathcal{F}_{\epsilon}(B)\right) \\ &\leq \alpha^*\left(\mathcal{F}_{\epsilon_M}(A), \mathcal{F}_{\epsilon_M}(B)\right) + \alpha^*\left(\mathcal{F}_{\epsilon_A}(A), \mathcal{F}_{\epsilon_A}(B)\right) + \alpha^*\left(\mathcal{F}_{\epsilon}(A), \mathcal{F}_{\epsilon}(B)\right)\end{aligned}$$

as $\{\mathcal{F}_{\epsilon_M}(A), \mathcal{F}_{\epsilon_A}(A), \mathcal{F}_{\epsilon}(A)\}$ and $\{\mathcal{F}_{\epsilon_M}(B), \mathcal{F}_{\epsilon_A}(B), \mathcal{F}_{\epsilon}(B)\}$ are sets of σ-algebras such that $\mathcal{F}_{\epsilon_M}(A) \vee \mathcal{F}_{\epsilon_M}(B)$, $\mathcal{F}_{\epsilon_A}(A) \vee \mathcal{F}_{\epsilon_A}(B)$ and $\mathcal{F}_{\epsilon}(A) \vee \mathcal{F}_{\epsilon}(B)$ are independent of each other (see page 5, Doukhan 1994). Now as $d(A,B) \geq u > 0$ we have $\mathcal{F}_{\epsilon_M}(A)$ and $\mathcal{F}_{\epsilon_M}(B)$ are independent and hence $\alpha^*\left(\mathcal{F}_{\epsilon_M}(A), \mathcal{F}_{\epsilon_M}(B)\right) = 0$ and similarly $\alpha^*\left(\mathcal{F}_{\epsilon_A}(A), \mathcal{F}_{\epsilon_A}(B)\right) = 0$. So we have $\alpha^*\left(\mathcal{F}_w(A), \mathcal{F}_w(B)\right) \leq \alpha^*\left(\mathcal{F}_{\epsilon}(A), \mathcal{F}_{\epsilon}(B)\right)$ and taking supremum over A and B on both sides we have $\alpha_w(u,v) \leq \alpha_{\epsilon}(u,v)$ and hence proved. ■

Lemma 3.3. *For any* $0 < p < 1$,

$$M^p_{n(i)} = \xi^p_{n(i)} + \frac{p - \hat{F}_{n(i)}(\xi^p_{n(i)})}{\bar{f}_{n(i)}(\xi^p_{n(i)})} + R_{n(i)}, \tag{3.7.1}$$

where $\bar{f}_{n(i)}(x) = \left(\sum_{j=1}^{n(i)} f_{i_j}(x)\right)/n(i)$ *and* $\sqrt{n(i)}R_{n(i)} \to 0$ *in probability.*

Proof of Lemma 3.3. Clearly, the representation in equation 3.7.1 is the Ghosh-Bahadur representation in our setting. We replicate the proof given by Ghosh (1971) for non-i.i.d data.

Let $(\Omega, \mathcal{A}, P)$ be the probability space on which all the random variables considered here are defined. Define $V_n^i = \sqrt{n(i)}\left(M_{n(i)}^p - \xi_{n(i)}^p\right)$. Note that, for any $t \in \mathbb{R}$,

$$\left\{\omega : V_n^i \leq t\right\} = \left\{\omega : W_n^i(t) \leq t_n^i\right\},$$

where,

$$W_n^i(t) = \frac{\sqrt{n(i)}\left(\hat{G}_{n(i)}(\xi_{n(i)}^p + t/\sqrt{n(i)}) - \bar{G}_{n(i)}(\xi_{n(i)}^p + t/\sqrt{n(i)})\right)}{\bar{f}_{n(i)}(\xi_{n(i)}^p)},$$

$$t_n^i = \frac{\sqrt{n(i)}\left(\bar{F}_{n(i)}(\xi_{n(i)}^p + t/\sqrt{n(i)}) - p\right)}{\bar{f}_{n(i)}(\xi_{n(i)}^p)}.$$

By definition of $n(i)$ and from condition (C.1), $C_1\delta_n^2 n\lambda_n^{-2} < n(i) < C_2\delta_n^2 n\lambda_n^{-2}$ and hence, as $n \to \infty$, $n(i) \to \infty$ as well from condition (C.5). Next, define

$$U_n^i = \sqrt{n(i)}\left(p - \hat{F}_{n(i)}(\xi_{n(i)}^p)\right)\left(\bar{f}_{n(i)}(\xi_{n(i)}^p)\right)^{-1}.$$

We can write U_n^i as,

$$U_n^i = \sqrt{n(i)}\left(\bar{f}_{n(i)}(\xi_{n(i)}^p)\right)^{-1}\left(\frac{1}{n(i)}\sum_j\left(\underbrace{F_{i_j}(\xi_{n(i)}^p) - \mathbb{I}\left\{w(\mathbf{s}_{i_j}) \leq \xi_{n(i)}^p\right\}}_{=\ Y_n(\mathbf{s}_{i_j})\ \text{(say)}}\right)\right),$$

where $Y_n(\mathbf{s}_{i_j})$ are mean-zero random variables with variance $\gamma_{j,n} = p_{j,n}(1 - p_{j,n})$ where

$p_{j,n} = F_{i_j}(\xi^p_{n(i)})$; note that, $p = \frac{1}{n(i)} \sum_j p_{j,n}$. Moreover, simple calculation shows

$$\mathrm{Cov}(w(\mathbf{s}_i), w(\mathbf{s}_j)) = \mathrm{Cov}(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)),$$

if $\mathbf{s}_i \neq \mathbf{s}_j$. Define, $Y^*_n(\mathbf{s}_{i_j}) = \gamma^{-1/2}_{j,n} Y_n(\mathbf{s}_{i_j})$; then $\{Y^*_n(\mathbf{s})\}$ is a zero mean stationary random field and,

$$U^i_n = (n(i))^{-1/2} \left(\bar{f}_{n(i)}(\xi^p_{n(i)}) \right)^{-1} \left(\sum_j \gamma^{1/2}_{j,n} Y^*_n(\mathbf{s}_{i_j}) \right)$$

has an asymptotic normal distribution from Theorem 4.2 of Lahiri (2003a).

Next we will show that, for any $t \in \mathbb{R}$, $E\left(U^i_n - W^i_n(t)\right)^2 \to 0$ as $n \to \infty$. It is easy to show that, $U^i_n - W^i_n(t) = \sqrt{n(i)} \left(\bar{f}_{n(i)}(\xi^p_{n(i)}) \right)^{-1} \left(\frac{1}{n(i)} \sum_j \left(T_n(\mathbf{s}_{i_j}) \right) \right)$, where

$$T_n(\mathbf{s}_{i_j}) = \mathbb{I}\left\{ w(\mathbf{s}_{i_j}) \leq \xi^p_{n(i)} \right\} - \mathbb{I}\left\{ w(\mathbf{s}_{i_j}) \leq \xi^p_{t,n(i)} \right\} - \left(F_{i_j}(\xi^p_{n(i)}) - F_{i_j}(\xi^p_{t,n(i)}) \right),$$

where $\xi^p_{t,n(i)} = \xi^p_{n(i)} + t/\sqrt{n(i)}$. Let us denote $\gamma^j_{t,n} = |F_{i_j}(\xi^p_{n(i)}) - F_{i_j}(\xi^p_{t,n(i)})|$. Clearly, $T_n(\mathbf{s}i_j)$ is a sequence of zero mean random variables with variance $\gamma^j_{t,n}\left(1 - \gamma^j_{t,n}\right)$. Also note that, as F_{i_j} is absolutely continuous around $\xi^p_{n(i)}$, $\gamma^j_{t,n} = O(t/\sqrt{n(i)})$.

So, $E\left(U^i_n - W^i_n(t)\right)^2 = n(i) \left(\bar{f}_{n(i)}(\xi^p_{n(i)}) \right)^{-2} E\left(\frac{1}{n(i)} \sum_j T_n(\mathbf{s}_{i_j}) \right)^2$. Now to find an upper bound on $E\left(\frac{1}{n(i)} \sum_j T_n(\mathbf{s}_{i_j}) \right)^2$ let us divide the square neighborhood $\mathcal{B}_{\delta_n}(\mathbf{s}_i)$ into disjoint unit squares $\{B_1, \ldots, B_{b_n}\}$ and b_n is the minimal natural number such that $\mathcal{B}_{\delta_n}(\mathbf{s}_i) \subset \cup_{m=1}^{b_n} B_m$. Clearly $b_n = \lceil 4\delta_n \rceil^2$ and we can rewrite the set of unit squares as $\{B_m : m \in \{1, \ldots, b_n\}\} = \left\{B_{m_1,m_2} : m_1, m_2 \in \left\{1, \ldots, \sqrt{b_n} = \lceil \lambda_n \rceil\right\}\right\}$, i.e. by m_1 and m_2

we are indexing the squares row and column wise respectively. Then,

$$
\begin{aligned}
& E\left(\frac{1}{n(i)}\sum_{j=1}^{n(i)} T_n(\mathbf{s}_{i_j})\right)^2 \\
&= \frac{1}{n(i)^2} E\left(\sum_{m_1}\sum_{m_2} T_n(B_{m_1,m_2})\right)^2 \quad \text{where} \quad T_n(B_{m_1,m_2}) = \sum_{\mathbf{s}_{i_j}\in B_{m_1,m_2}} T_n(\mathbf{s}_{i_j}) \\
&= \frac{1}{n(i)^2} E\left\{\left(\sum_{m_1 \text{ even}}\sum_{m_2 \text{ even}} T_n(B_{m_1,m_2})\right) + \cdots + \left(\sum_{m_1 \text{ odd}}\sum_{m_2 \text{ odd}} T_n(B_{m_1,m_2})\right)\right\}^2 \\
&= \frac{1}{n(i)^2} E\left\{\sum_1 + \sum_2 + \sum_3 + \sum_4\right\}^2,
\end{aligned}
$$

where $\sum_1 = \sum_{m_1 \text{ even}}\sum_{m_2 \text{ even}} T_n(B_{m_1,m_2})$, $\sum_2 = \sum_{m_1 \text{ even}}\sum_{m_2 \text{ odd}} T_n(B_{m_1,m_2})$ and so on. By Jensen's inequality,

$$
E\left\{\sum_1 + \sum_2 + \sum_3 + \sum_4\right\}^2 \le 4\left\{E\left(\sum_1\right)^2 + E\left(\sum_2\right)^2 + E\left(\sum_3\right)^2 + E\left(\sum_4\right)^2\right\}.
$$

Now we will find bounds on each of the expected values on the right hand side.

$$
\begin{aligned}
E\left(\sum_1\right)^2 &= E\left(\sum_{m_1 \text{ even}}\sum_{m_2 \text{ even}} T_n(B_{m_1,m_2})\right)^2 \\
&= \sum_{m_1 \text{ even}}\sum_{m_2 \text{ even}} E\left(T_n(B_{m_1,m_2})\right)^2 + \\
&\quad \sum_{\substack{m_{11},m_{12} \\ \text{even} \\ m_{11}\neq m_{12} \text{ or } m_{21}\neq m_{22}}}\sum_{\substack{m_{21},m_{22} \\ \text{even}}} \operatorname{Cov}\left\{T_n(B_{m_{11},m_{21}}), T_{kl}(B_{m_{12},m_{22}})\right\}
\end{aligned}
$$

Now, the first term

$$E\left(T_n(B_{m_1,m_2})\right)^2 \leq C_2 \frac{n}{\lambda_n^2} \sum_{\mathbf{s}_{i_j} \in B_{m_1,m_2}} E\left(T_n(\mathbf{s}_{i_j})\right)^2 \quad \text{[from condition (C.1)]}$$

$$= O(n^2 \lambda_n^{-4} \gamma_n^*(t)),$$

where $\gamma_n^*(t) = \sup_j \gamma_{t,n}^j$. So, $\sum_{m_1 \text{ even}} \sum_{m_2 \text{ even}} E\left(T_n(B_{m_1,m_2})\right)^2 = O(b_n n^2 \lambda_n^{-4} \gamma_n^*(t))$. For the covariance term note that,

$$\begin{aligned} \left|\text{Cov}\left\{T_n(B_{m_{11},m_{21}}), T_n(B_{m_{12},m_{22}})\right\}\right| &\leq 4 \frac{C_2^2 n^2}{\lambda_n^4} \left\{\sup_{\mathbf{s}} |T_n(\mathbf{s})|\right\} \alpha_w\left(d(B_{m_{11},m_{21}}, B_{m_{21},m_{22}}), 1\right) \\ &\leq C \frac{n^2}{\lambda_n^4} \alpha_\epsilon\left(\sqrt{(m_{12}-m_{11})^2 + (m_{22}-m_{21})^2}, 1\right). \end{aligned}$$

So,

$$\begin{aligned} &\Big| \sum_{\substack{m_{11},m_{12} \\ \text{even} \\ m_{11} \neq m_{12} \text{ or } m_{21} \neq m_{22}}} \sum_{\substack{m_{21},m_{22} \\ \text{even}}} \text{Cov}\left\{T_n(B_{m_{11},m_{21}}), T_{kl}(B_{m_{12},m_{22}})\right\}\Big| \\ &\leq C \frac{n^2}{\lambda_n^4} \sum_{\substack{m_{11},m_{12} \\ \text{even} \\ m_{11} \neq m_{12} \text{ or } m_{21} \neq m_{22}}} \sum_{\substack{m_{21},m_{22} \\ \text{even}}} \alpha_\epsilon\left(\sqrt{(m_{12}-m_{11})^2 + (m_{22}-m_{21})^2}, 1\right) \\ &\leq C \frac{n^2}{\lambda_n^4} \sum_{\substack{p_1=-b_n \\ p_1 \neq 0 \text{ or } p_2 \neq 0}}^{b_n} \sum_{p_2=-b_n}^{b_n} \alpha_\epsilon\left(\sqrt{p_1^2 + p_2^2}, 1\right) \\ &= O\left(n^2 \lambda_n^{-4}\right) \quad \text{[by condition (C.4)]}. \end{aligned}$$

Putting together, we have $E\left(\sum_1\right)^2 = O\left(n^2 \lambda_n^{-4}\left(1 + b_n \gamma_n^*(t)\right)\right) = O\left(n^2 \lambda_n^{-4} \delta_n^2 \gamma_n^*(t)\right)$. Similarly, we can show for the other three sums the bound holds as well. So, for any $t \in \mathbb{R}$, some constants C and C', $E\left(U_n^i - W_n^i(t)\right)^2 = O\left(\frac{n(i) n^2 \delta_n^2 \gamma_n^*(t)}{\left(\bar{f}_{n(i)}(\xi_{n(i)}^p)\right)^2 n(i)^2 \lambda_n^4}\right) = O(\sqrt{n} \lambda_n^{-1} \delta_n^{-1}) \to 0$ by condition (C.5). So, by Lemma 1 of Ghosh (1971), we have the representation in equation 3.7.1. ■

Next, we state the proof of proposition 3.1.

Proof of Proposition 3.1. Note that,

$$\begin{aligned} Q_2\left(\mathbf{Z}_i\right) &= Q_2\left(\left(\mathbf{x}(\mathbf{s}_{i_1})'\boldsymbol{\beta} + w(\mathbf{s}_{i_1}), \ldots, \mathbf{x}(\mathbf{s}_{i_{n(i)}})'\boldsymbol{\beta} + w(\mathbf{s}_{i_{n(i)}})\right)'\right) \\ &= \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + Q_2\left(\underbrace{\left(w(\mathbf{s}_{i_1}), \ldots, w(\mathbf{s}_{i_{n(i)}})\right)'}_{=\mathbf{w}_i}\right) + O(\delta_n\lambda_n^{-1}), \end{aligned} \tag{3.7.2}$$

as, from condition (C.3),

$$\begin{aligned} \left(\mathbf{x}(\mathbf{s}_i) - \mathbf{x}(\mathbf{s}_{i_j})\right)'\boldsymbol{\beta} &= \sum_{k=1}^{p} \beta_k \left(x_k(\mathbf{s}_i) - x_k(\mathbf{s}_{i_j})\right) \\ &= \sum_{k=1}^{p} \beta_k \left(\nabla x_k(\mathbf{s}_{i_j}^*)\right)' \left(\mathbf{s}_i - \mathbf{s}_{i_j}\right) = O\left(\delta_n\lambda_n^{-1}\right), \end{aligned} \tag{3.7.3}$$

for some $\mathbf{s}_{i_j}^* \in \mathbb{R}^2$ in between $\mathbf{s}_i$ and $\mathbf{s}_{i_j}$ component wise. Now, from Lemma 3.3 and condition (C.9), we have $Q_2(\mathbf{w}_i) = O_p\left((n(i))^{-1/2}\right) = O_p(n^{-1/2}\lambda_n\delta_n^{-1})$. So, from equation 3.7.2, we have $Z(\mathbf{s}_i) - Q_2(\mathbf{Z}_i) = w(\mathbf{s}_i) + O_p(n^{-1/2}\lambda_n\delta_n^{-1} + \delta_n\lambda_n^{-1})$. Similar argument using the first and third sample quartile we can show that $\mathrm{IQR}(\mathbf{Z}_i) = \mathcal{I}_n(\mathbf{s}_i) + O_p(n^{-1/2}\lambda_n\delta_n^{-1} + \delta_n\lambda_n^{-1})$. Now, through simple probability calculations we have

$$\begin{aligned} V^{(m)}(\mathbf{s}_i) &= \exp\left(-\frac{|Z(\mathbf{s}_i) - Q_2(\mathbf{Z}_i)|}{\mathrm{IQR}(\mathbf{Z}_i)}\right) \\ &= \exp\left(-\frac{|w(\mathbf{s}_i) + O_p(a_n)|}{\mathcal{I}_n(\mathbf{s}_i) + O_p(a_n)}\right) \\ &= \exp\left(-\frac{|w(\mathbf{s}_i)|}{\mathcal{I}_n(\mathbf{s}_i)}\right) + O_p(a_n), \end{aligned}$$

and hence proved. ■

Next, we prove Lemma 3.1 as stated below which provides an uniform (over the space)

finite upper bound on the quantity $\mathcal{I}_n(\mathbf{s}_i)$.

Proof of Lemma 3.1. By definition of $\xi^p_{n(i)}$, $\bar{F}_{n(i)}(\xi^p_{n(i)}) = p$. Now, observe that, $\frac{1}{n(i)}\sum_{j=1}^{n(i)} F_{i_j}(\xi^p_{n(i)}) = \frac{1}{n(i)}\left\{\sum_{j\in G_{i,n}} F_{i_j}(\xi^p_{n(i)}) + \sum_{j\in G^c_{i,n}} F_{i_j}(\xi^p_{n(i)})\right\}$, where $G_{i,n} \subset G_n$ is the set of indices of the observations in the square neighborhood around $\mathbf{s}_i$ which are not affected by the additive-multiplicative noise. From condition (C.1) and (C.2), $|G_{i,n}|/n(i) \geq q_e$. Hence,

$$p \geq \frac{1}{n(i)} \sum_{j\in G_{i,n}} F_{i_j}(\xi^p_{n(i)}) \geq q_e F_\epsilon(\xi^p_{n(i)}/\sigma_\epsilon),$$

which proves that $\xi^p_{n(i)} \leq \sigma_\epsilon \left(F_\epsilon^{-1}(\min\{p/q_e, 1\})\right)$. In place of p, using the similar argument for the $(1-p)$-th quantile we can show that $\xi^p_{n(i)} \geq \sigma_\epsilon \left(F_\epsilon^{-1}(\max\left\{1 - \frac{1-p}{q_e}, 0\right\})\right)$. Taking $p = 0.75$ and 0.25 we have the proof. ■

Now we will prove the representation theorem for the Median-VS-based regression parameter estimator as given below.

Proof of Theorem 3.1. From definition of Median-VS-based regression estimator,

$$\hat{\boldsymbol{\beta}}^{(m)}_{\text{vs}} - \boldsymbol{\beta} \;=\; \left(\frac{1}{n}\mathbf{X}'\mathbf{D}^{(m)}_v\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\mathbf{D}^{(m)}_v\mathbf{w}\right),$$

where $\mathbf{D}^{(m)}_v = \text{diag}\left(V^{(m)}(\mathbf{s}_1), \ldots, V^{(m)}(\mathbf{s}_n)\right)$. For simplicity of notation, as we are only considering Median-VS here, let us denote $\mathbf{D}^{(m)}_v$ as $\mathbf{D}_v$. Now, from Proposition 3.1, $\frac{1}{n}\mathbf{X}'\mathbf{D}_v\mathbf{w} = \frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w} + O_p(a_n)$ and $\frac{1}{n}\mathbf{X}'\mathbf{D}_v\mathbf{X} = \frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X} + \frac{1}{n}\mathbf{X}'\left(\tilde{\mathbf{D}}_v - E\left(\tilde{\mathbf{D}}_v\right)\right)\mathbf{X} + O_p(a_n)$. Next we will show that $\frac{1}{n}\mathbf{X}'\left(\tilde{\mathbf{D}}_v - E\left(\tilde{\mathbf{D}}_v\right)\right)\mathbf{X} = O_p(\lambda_n^{-1})$.

Notice that $\frac{1}{n}\mathbf{X}'\left(\tilde{\mathbf{D}}_v - E\left(\tilde{\mathbf{D}}_v\right)\right)\mathbf{X}$ is a $p \times p$ matrix and hence enough to show that a general element of of the matrix say (k,l)-th element is $O_p(\lambda_n^{-1})$. Now, from simple

calculations,

$$\left(\frac{1}{n}\mathbf{X}'\left(\tilde{\mathbf{D}}_v - E\left(\tilde{\mathbf{D}}_v\right)\right)\mathbf{X}\right)_{kl} = \frac{1}{n}\sum_{i=1}^{n} x_k(\mathbf{s}_i)\, x_l(\mathbf{s}_i)\left\{\tilde{V}^{(m)}(\mathbf{s}_i) - E\left(\tilde{V}^{(m)}(\mathbf{s}_i)\right)\right\}$$
$$= \frac{1}{n}\sum_{i} T_{kl}(\mathbf{s}_i),$$

where, $T_{kl}(\mathbf{s}_i) = x_k(\mathbf{s}_i)\, x_l(\mathbf{s}_i)\left\{\tilde{V}^{(m)}(\mathbf{s}_i) - E\left(\tilde{V}^{(m)}(\mathbf{s}_i)\right)\right\}$. As, $E\left(T_{kl}(\mathbf{s}_i)\right) = 0$ enough to show that $\mathrm{Var}\left(\frac{1}{n}\sum_i T_{kl}(\mathbf{s}_i)\right) = O(\lambda_n^{-2})$. To do so we will adopt the similar strategy used in the proof of Lemma 3.3. We will divide the whole sampling region $\mathcal{D}_n = \lambda_n[0,1]^2$ into disjoint unit squares $\{B_1, \ldots, B_{b_n}\}$ and b_n is the minimal natural number such that $\mathcal{D}_n \subset \cup_{m=1}^{b_n} B_m$. Clearly $b_n = \lceil \lambda_n \rceil^2$ and we can rewrite the set of unit squares as $\{B_m : m \in \{1, \ldots, b_n\}\} = \left\{B_{m_1,m_2} : m_1, m_2 \in \left\{1, \ldots, \sqrt{b_n} = \lceil \lambda_n \rceil\right\}\right\}$, i.e. by m_1 and m_2 we are indexing the squares row and column wise respectively. Then following similar arguments we can show that $E\left(\frac{1}{n}\sum_i T_{kl}(\mathbf{s}_i)\right)^2 = O(\lambda_n^{-2})$. Hence, $\frac{1}{n}\mathbf{X}'\mathbf{D}_v\mathbf{X} = \frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X} + O_p(\lambda_n^{-1} + a_n) = \frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X} + O_p(a_n)$. So,

$$\hat{\boldsymbol{\beta}}_{\mathrm{vs}}^{(m)} - \boldsymbol{\beta} = \left(\frac{1}{n}\mathbf{X}'\mathbf{D}_v\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\mathbf{D}_v\mathbf{w}\right)$$
$$= \left(\frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X} + O_p(a_n)\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w} + O_p(a_n)\right)$$
$$= \left(\frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)^{-1}\left(\frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w}\right) + O_p(a_n),$$

and hence proved. ■

Next we provide the proof of the Corollary 3.1 as stated below.

Proof of Corollary 3.1. From condition (C.11), $\frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w} = O_p\left(\sqrt{\mathrm{Var}\left(\frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w}\right)}\right)$. Using similar argument as in the proof of Theorem 3.1, we can show that $\mathrm{Var}\left(\frac{1}{n}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w}\right) = O(\lambda_n^{-2})$. Hence, $\hat{\boldsymbol{\beta}}_{\mathrm{VS}}^{(a)} = \boldsymbol{\beta} + \left(\frac{1}{n}\mathbf{X}'E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)^{-1} O_p\left(\lambda_n^{-1}\right) + O_p(a_n)$. Now, for all $i \in$

$\{1, \ldots, n\}$,

$$\begin{aligned}
\left(E\left(\tilde{\mathbf{D}}_v\right)\right)_{ii} &= E\left(\exp\left(-\frac{|w(\mathbf{s}_i)|}{\mathcal{I}_n(\mathbf{s}_i)}\right)\right) \\
&= \int_0^\infty P\left(\exp\left(-\frac{|w(\mathbf{s}_i)|}{\mathcal{I}_n(\mathbf{s}_i)}\right) > x\right) dx \\
&= \int_0^\infty e^{-y} P\left(|w(\mathbf{s}_i)| < \mathcal{I}_n(\mathbf{s}_i) y\right) dy \quad [\text{taking } x = e^{-y}] \\
&> \int_0^\infty e^{-y} P(|\epsilon(\mathbf{s}_i)| < y C_\epsilon^{(l)}(q_e)) \; = \psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right).
\end{aligned}$$

Clearly, as from condition (C.12) $q_e > 0.75$, $C_\epsilon^{(l)}(q_e) > 0$ and under assumption (C.13) $\psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)$ is strictly positive. Also, note that $\frac{1}{n}\mathbf{X}'\mathbf{X} \succeq \frac{1}{n}\mathbf{X}' E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X} \succ \frac{\psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)}{n}\mathbf{X}'\mathbf{X}$ where from condition (C.10), $\frac{1}{n}\mathbf{X}'\mathbf{X} \to \mathbf{C}_X$ for some positive definite matrix $\mathbf{C}_X$. Clearly, for large enough n, $\left(\frac{1}{n}\mathbf{X}' E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)^{-1} \preceq \psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)\mathbf{C}_X$ and hence we have the proof. ■

Our next focus is the efficiency of the Median-VS-based regression estimator as compared to the o.l.s estimator.

Proof of Theorem 3.2. Observe that,

$$\begin{aligned}
E\left(\|\mathbf{l}_n^{\text{vs}}\|^2\right) &= E\left(\mathbf{w}'\tilde{\mathbf{D}}_v\mathbf{X}\left(\mathbf{X}' E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X}\right)^{-2}\mathbf{X}'\tilde{\mathbf{D}}_v\mathbf{w}\right) \\
&= \frac{1}{n}E\left(\mathbf{w}'\tilde{\mathbf{D}}_v\mathbf{A}_n\tilde{\mathbf{D}}_v\mathbf{w}\right) \quad \text{where, } \frac{1}{n}\mathbf{X}' E\left(\tilde{\mathbf{D}}_v\right)\mathbf{X} = \mathbf{B}_n, \frac{1}{n}\mathbf{X}\mathbf{B}_n^{-2}\mathbf{X}' = \mathbf{A}_n \\
&= \frac{1}{n}\text{tr}\left(\mathbf{A}_n E\left(\tilde{\mathbf{D}}_v\mathbf{w}\mathbf{w}'\tilde{\mathbf{D}}_v\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^n (\mathbf{A}_n)_{ii} E\left(\tilde{V}(\mathbf{s}_i)^2 w(\mathbf{s}_i)^2\right) + \frac{1}{n}\sum_{i\neq j}(\mathbf{A}_n)_{ij} E\left(\tilde{V}(\mathbf{s}_i)w(\mathbf{s}_i)\tilde{V}(\mathbf{s}_j)w(\mathbf{s}_j)\right)
\end{aligned}$$

We can show that $\frac{1}{n}\mathbf{X}'\mathbf{X} \succeq \mathbf{B}_n \succ \psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)\frac{1}{n}\mathbf{X}'\mathbf{X}$ where $\psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right) > 0$ and from

condition (C.10) $\frac{1}{n}\mathbf{X}'\mathbf{X} \to \mathbf{C}_X$ as $n \to \infty$. Hence, for large enough n,

$$\lambda_{\min}\left(\mathbf{B}_n\right) \geq \psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)\lambda_{\min}\left(\mathbf{C}_X\right)/2,$$

and so,

$$(\mathbf{A}_n)_{ii} \leq \frac{4\,\sup_i\|\mathbf{X}[i,]\|^2}{n\left(\psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)\lambda_{\min}\left(\mathbf{C}_X\right)\right)^2};$$

and

$$\begin{aligned} E\left(\tilde{V}(\mathbf{s}_i)^2 w(\mathbf{s}_i)^2\right) &= E\left(\exp\left(\frac{-2|w(\mathbf{s}_i)|}{\mathcal{I}_n(\mathbf{s}_i)}\right) w(\mathbf{s}_i)^2\right) \\ &= \sigma_i^2 E\left(\exp\left(-t_{i,n}|u_i|\right) u_i^2\right) \\ &\leq \sigma_i^2 \exp(-2) 4/t_{i,n}^2 \;\text{ where } t_{i,n} = 2\sigma_i/\mathcal{I}_n(\mathbf{s}_i) \\ &\leq \exp(-2)(C_\epsilon^{(u)}(q_e))^2. \end{aligned}$$

So, we have shown that,

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{A}_n)_{ii} E\left(\tilde{V}(\mathbf{s}_i)^2 w(\mathbf{s}_i)^2\right) = O\left(n^{-1}\left(C_\epsilon^{(u)}(q_e)\right)^2\left(\psi_\epsilon\left(C_\epsilon^{(l)}(q_e)\right)\right)^{-2}\right).$$

Using the similar trick as used in the proof of Lemma 3.3, i.e. dividing the region into unit blocks and also using the fact that $|(\mathbf{A}_n)_{ij}| = O(n^{-1})$, we can show that

$$|\frac{1}{n}\sum_{i\neq j}(\mathbf{A}_n)_{ij} E\left(\tilde{V}(\mathbf{s}_i) w(\mathbf{s}_i) \tilde{V}(\mathbf{s}_j) w(\mathbf{s}_j)\right)| = O(\lambda_n^{-4}),$$

which concludes the proof. ■

Our next objective is to prove the lower-bound result for the o.l.s estimator for the considered additive-multiplicative noise model.

Proof of Theorem 3.3. Recall, $\mathbf{l}_n^{\text{ols}} = (n^{-1}\mathbf{X}'\mathbf{X})^{-1}(n^{-1}\mathbf{X}'\mathbf{w})$. So,

$$\begin{aligned} E\left(\|\mathbf{l}_n^{\text{ols}}\|^2\right) &= E\left(\mathbf{w}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-2}\mathbf{X}'\mathbf{w}\right) \\ &= \frac{1}{n}\text{tr}\left(\mathbf{H}_n\mathbf{\Sigma}_w\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}(\mathbf{H}_n)_{ii}E\left(w(\mathbf{s}_i)^2\right) + \frac{1}{n}\sum_{i\neq j}(\mathbf{H}_n)_{ij}E\left(w(\mathbf{s}_i)w(\mathbf{s}_j)\right), \end{aligned}$$

where $\mathbf{H}_n = n^{-1}\mathbf{X}(n^{-1}\mathbf{X}'\mathbf{X})^{-2}\mathbf{X}'$, $\mathbf{\Sigma}_w = \text{Var}(\mathbf{w})$. Let us also denote $n^{-1}\mathbf{X}'\mathbf{X}$ by $\mathbf{C}_n$ and by condition (C.10), $\mathbf{C}_n \to \mathbf{C}_X$ as $n \to \infty$. Note that $\text{tr}(\mathbf{H}_n) = \text{tr}\left(\mathbf{C}_n^{-1}\right) \to \text{tr}\left(\mathbf{C}_X^{-1}\right) = 2C^{(0)} > 0$. So, for large enough n, $\text{tr}(\mathbf{H}_n) \geq C^{(0)} > 0$. Also, for large enough n,

$$(\mathbf{H}_n)_{ii} \leq \frac{1}{n}\lambda_{\max}\left(\mathbf{C}_n^{-2}\right)\|\mathbf{X}[i,]\|^2 < \frac{C^{(1)}}{n}.$$

For any $\varepsilon > 0$ let us define,

$$\mathcal{M}_n(\varepsilon) := \left\{1 \leq i \leq n : (\mathbf{H}_n)_{ii} > n^{-1}\varepsilon\right\},$$

and let us denote the cardinality of this set by m_n (dependent on ε). Now consider a $\varepsilon > 0$ such that $\varepsilon < \min\left(C^{(0)}, C^{(1)}\right)$. Then,

$$C^{(0)} \leq \text{tr}(\mathbf{H}_n) = \sum_{i=1}^{n}(\mathbf{H}_n)_{ii} \leq m_n\frac{C^{(1)}}{n} + (n - m_n)\frac{\varepsilon}{n},$$

and by simple algebric calculation we have $m_n \geq \left(C^{(0)} - \varepsilon\right)\left(C^{(1)} - \varepsilon\right)^{-1} n$. Recall, $E(w(\mathbf{s}_i)^2)$

is equal to σ_ϵ^2 if $i \in G_n$ and is equal to $\sigma_\epsilon^2 + \tau_i^2$ otherwise. So,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(\mathbf{H}_n)_{ii}E\left(w(\mathbf{s}_i)^2\right) &= \frac{\sigma_\epsilon^2}{n}\sum_{i=1}^{n}(\mathbf{H}_n)_{ii} + \frac{1}{n}\sum_{i\in G_n^c}\tau_i^2\,(\mathbf{H}_n)_{ii} \\
&> \frac{\sigma_\epsilon^2}{n}\sum_{i\in\mathcal{M}_n(\varepsilon)}\frac{\varepsilon}{n} + \frac{1}{n}\sum_{i\in\mathcal{M}_n(\varepsilon)\cap G_n^c}\tau_i^2\frac{\varepsilon}{n} \\
&\geq \frac{\varepsilon}{n}\left(\frac{\sigma_\epsilon^2\left(C^{(0)}-\varepsilon\right)}{\left(C^{(1)}-\varepsilon\right)} + \frac{1}{n}\sum_{i\in\mathcal{M}_n(\varepsilon)\cap G_n^c}\tau_i^2\right).
\end{aligned}$$

Finally, using the similar methodology as in proof of Theorem 3.2 we can show that $|\frac{1}{n}\sum_{i\neq j}(\mathbf{H}_n)_{ij}E\left(w(\mathbf{s}_i)w(\mathbf{s}_j)\right)| = O(\lambda_n^{-4})$. Hence we have shown that, for any $0 < \varepsilon < \min\left(C^0, C^1\right)$ and large enough n,

$$E\left(\|\mathbf{l}_n^{\text{ols}}\|^2\right) > n^{-1}\varepsilon\left(\sigma_\epsilon^2\left(C^{(0)}-\varepsilon\right)\left(C^{(1)}-\varepsilon\right)^{-1} + n^{-1}\sum_{i\in\mathcal{M}_n(\varepsilon)\cap G_n^c}\tau_i^2\right) - C_1\lambda_n^{-4}.$$

■

Proof of Corollary 3.2. Observe that, for large enough n, for all $i \in \{1,\ldots,n\}$,

$$\begin{aligned}
n\,(\mathbf{H}_n)_{ii} &= \left(\mathbf{X}\mathbf{C}_n^{-2}\mathbf{X}'\right)_{ii} \\
&\geq \lambda_{\min}\left(\mathbf{C}_n^{-2}\right)\|\mathbf{X}[i,]\|^2 \\
&> \left(\lambda_{\max}\left(\mathbf{C}_X\right)\right)^{-2}\min_{1\leq i\leq n}\|\mathbf{X}[i,]\|^2/2 \quad (= C_2,\ \text{say}).
\end{aligned}$$

Clearly, from the assumption of the Corollary 3.2, $C_2 > 0$. Take any $\varepsilon > 0$ such that $\varepsilon < C_2$. Then, for all $i \in \{1,\ldots,n\}$, $(\mathbf{H}_n)_{ii} > n^{-1}\varepsilon$ and hence,

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{H}_n)_{ii}E\left(w(\mathbf{s}_i)^2\right) > \frac{1}{n}\left(\varepsilon\sigma_\epsilon^2 + \frac{1}{n}\sum_{i\in G_n^c}\tau_i^2\right).$$

The rest of the proof is similar to that of Theorem 3.3. ■

Acknowledgements

The authors would like to thank Dr. Alyson Wilson for bringing the importance of this work to authors' attention and also for her insightful remarks during the course of this work.

Chapter 4

Anisotropic Covariance Modeling Using Marginal Variograms: A Copula-based Approach

4.1 Introduction

Modeling the dependence structure of the spatial process is one of the major aspects of analyzing geostatistical data. It is often assumed that the underlying spatial process is isotropic in nature, i.e. the spatial covariance between two locations is only dependent on their magnitude of the distance and not on the relative orientation. Though isotropic covariances are vastly common in literature due to its mathematical simplicity and computational ease, the assumption of isotropy is highly restrictive in practice and misspecification of this assumption can result in erroneous inference and prediction. Presence of anisotropy in the process is often verified through directional variogram analysis: if the variograms estimated from the de-trended process behave differently for a finitely-

many distinct directions, then the underlying spatial process is assumed to be anisotropic (e.g. see page 215, Cressie 1993; Ecker and Gelfand 1999). Modeling anisotropic dependence structure are mainly restricted to geometrically anisotropic or zonal anisotropic models; for example, see page 62, Cressie (1993); page 39, Gelfand et al. (2010); Haskard (2007); Ecker and Gelfand (1999) etc. Literature on anisotropic models other than geometric or zonal anisotropy is yet very sparse, at the exception of non-geometric range anisotropy by Ecker and Gelfand (2003); nested variogram models with different range parameters by Zimmerman (1993), Eriksson and Siska (2000) etc. But, as pointed out by Ecker and Gelfand (1999), some of these proposed models do not specify admissible covariance structure on the underlying spatial process. In a recent paper, Allard et al. (2016) has provided a general characterization of admissible anisotropic models where the anisotropic range parameter has been shown to be a mixture of zonal anisotropies – which provides a class of more general form of range anisotropy than the zonal and geometric ones.

The work of this chapter has been motivated by the following questions that are often faced by the practitioners in geospatial sciences. Though the presence of anisotropy is established through estimation of the directional variograms for finitely many specified directions, the resulting covariance model often do not respect the directional or marginal fits. For example, in geometrically anisotropic models, the range parameters are estimated marginally for two perpendicular directions (the ones with maximum and minimum range) and, the estimated parameters along with the orientation of the pairs of directions are included in the final model (for details, see page 215, Cressie 1993; Ecker and Gelfand 1999). But, as mentioned by Cressie (1993), often the directional fits for other directions and the marginal projections of the fitted joint covariance do not agree with each other. Moreover, the directions of maximum and minimum range are

often not perpendicular to each other. On the other hand, consider the situation where in one direction the behavior of the marginal covariance is mostly resembled with an Exponential model where as, in another direction the covariance decays much slowly but it is negligible after a certain distance and hence, probably a Spherical covariance function will be the best fit: how to model such dependence structure where two directional covariances are not only showing different range or irregularity but also they belong to two different parametric classes? Moreover, estimation of the anisotropy parameter for most of the models discussed here so far(e.g. geometric anisotropy, non-geometric range anisotropy by Ecker and Gelfand 1999; nested modeling by Eriksson and Siska 2000) requires least squares based variogram model fitting in $\mathbb{R}^2$ or likelihood-based techniques, which is often computationally restrictive when the sample size is large ($\sim 10^5$ or larger). Hence, a scalable methodology to model anisotropy is needed which can incorporate the directionally fitted variogram models for a finitely many specified directions.

In this chapter, we have introduced a copula-based modeling approach of anisotropic covariances in $\mathbb{R}^2$ that can address the questions mentioned before. The organization of this chapter as follows. In Section 4.2, we have set the notations for this chapter and provided a brief review on the covariance modeling in geostatistics. Section 4.3 introduces a copula-based approach to construct an admissible bivaraite covariance using two marginal covariance models. In particular, it has been shown under some general regularity conditions that any valid bivariate covariance function can be characterized through directional covariances in two directions and a bivariate copula. Moreover, estimation of the parameters of the proposed anisotropy modeling has been discussed and large sample properties of the estimators has been explored and established. In Section 4.4, copula-based anisotropic covariances has been illustrated through several examples. Section 4.5 summarizes the results of the Monte Carlo studies regarding the consistency

of the proposed covariance estimators and flexibility as well as scalability of the copula-based approach as compared to more widely used techniques. In section 4.6, we give some concluding remarks and discuss some potential future directions to this work. Finally, Section 4.7 provides the proofs of all the results stated in this chapter.

4.2 Review of Anisotropic Covariance Models

Let $\{Z(\mathbf{s}) : \mathbf{s} \in \mathcal{D} \subseteq \mathbb{R}^2\}$ be a spatial process with the following decomposition,

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \underbrace{Y(\mathbf{s}) + w(\mathbf{s})}_{\epsilon(\mathbf{S})}, \tag{4.2.1}$$

where $\mu(\mathbf{s})$ $(= E(Z(\mathbf{s})))$ is a smooth function accounting for the large-scale variation of the process, $\{Y(\mathbf{s})\}$ is a spatially correlated random field and $\{w(\mathbf{s})\}$ is a micro-scale random error and $\epsilon(\mathbf{s}) = Y(\mathbf{s}) + w(\mathbf{s})$ is the spatially dependent residual process accounting for the small-scale variation of $Z(\cdot)$. The second-order structure of the random field $Z(\cdot)$ is determined through the covariance structure of the residual process $\epsilon(\cdot)$ as the trend $\mu(\cdot)$ is assumed to be deterministic in nature. The *covariogram* of the random field $\{\epsilon(\cdot)\}$ is defined as a positive-definite function $\varphi(\cdot,\cdot) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ such that $\text{Cov}(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)) = \varphi(\mathbf{s}_i, \mathbf{s}_j)$. Another way to specify the second-order structure of the spatial process is the *variogram* function defined as a function $2\gamma(\cdot,\cdot) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ such that $\gamma(\mathbf{s}_i, \mathbf{s}_j) = \frac{1}{2}(\text{Var}(\epsilon(\mathbf{s}_i) - \epsilon(\mathbf{s}_j)))$. The function γ is often called the *semivariogram* of the process $\{\epsilon(\cdot)\}$ and, it can be shown that to be a valid semivariogram, γ has to be conditionally non-negative definite (for details, see Chapter 2, Cressie 1993). To make the estimation of the model parameters feasible from single replication of the process at finitely many spatial locations further assumptions are needed on the covariogram

and/or variogram functions. In practice, often second-order stationarity of the process is assumed where, the covariance between two points are only dependent on the distance vector between the points, i.e. for any spatial location $\mathbf{s}$ and lag $\mathbf{h}$ in $\mathbb{R}^2$,

$$\text{Cov}\left(\epsilon(\mathbf{s}+\mathbf{h}), \epsilon(\mathbf{s})\right) = \varphi\left(\mathbf{s}+\mathbf{h}, \mathbf{s}\right) = \varphi(\mathbf{h}, \mathbf{0}) \equiv \varphi\left(\mathbf{h}\right), \tag{4.2.2}$$

where $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is also referred as the 'covariogram' or covariance function of the second-order stationary random field $\{\epsilon\left(\cdot\right)\}$. To define the stationarity in terms of the variogram, we need a weaker assumption on the ϵ-process, called *intrinsic stationarity*, where instead of assuming the entire process second-order stationary we only need that the increments, i.e. $\epsilon(\mathbf{s}+\mathbf{h}) - \epsilon(\mathbf{s})$, are weakly stationary and so, under intrinsic stationarity,

$$\text{Var}(\epsilon(\mathbf{s}+\mathbf{h}) - \epsilon(\mathbf{s})) = 2\gamma(\mathbf{s}+\mathbf{h}, \mathbf{s}) = 2\gamma(\mathbf{h}, \mathbf{0}) \equiv 2\gamma(\mathbf{h}), \tag{4.2.3}$$

for all spatial locations $\mathbf{s}$ and lags $\mathbf{h}$ in $\mathbb{R}^2$. Under the stronger condition, i.e. the second-order stationarity, where both φ and γ exists, it can be easily shown that $2\gamma(\mathbf{h}) = 2\left(\varphi(\mathbf{0}) - \varphi(\mathbf{h})\right)$ (Cressie, 1993).

The process $\{\epsilon\left(\cdot\right)\}$ is said to be *isotropic* if the covariance function $\varphi\left(\mathbf{h}\right)$ is only a function of the magnitude of the lag-vector $\mathbf{h}$, and not of the orientation, i.e. $\varphi\left(\mathbf{h}\right) = \varphi^*\left(\|\mathbf{h}\|\right)$ for some valid covariance function $\varphi^*\left(\cdot\right) : \mathbb{R} \to \mathbb{R}$. On the other hand, the ϵ-process is called *anisotropic* if the covariance function is not isotropic, i.e. it is not only a function of the magnitude of the lag-vector, but also the corresponding relative orientation. For details in isotropic covariance modeling see Chapter 2, Cressie (1993). When it comes to modeling anisotropic covariance structure, the most commonly used

approach is the assumption of *geometric anisotropy*. Under geometric anisotropy, it is assumed that the lag-space $\{\mathbf{h} = \mathbf{s} - \mathbf{s}' : \mathbf{s}, \mathbf{s}' \in \mathbb{R}^2\}$ can be rotated by a certain rotation matrix such that the covariance is isotropic in nature in the transformed space and hence, can be modeled using valid univariate covariances. In mathematical formulation, under geometric anisotropy, $\varphi(\mathbf{h}) = \varphi^*(\mathbf{h}'\mathbf{A}\mathbf{h})$ where $\mathbf{A}$ is the matrix deciding the rotation of the space. In this case, the contours of the covariance function, which is circular in case isotropy, are elliptical. Assumption of geometric anisotropy is mathematically and computationally convenient as it only takes an univariate class of parametric covariance and two additional model parameters – the orientation of the ellipse in lag-space and the ratio of the major and minor axes of the ellipse – to completely specify the bivariate covariance.

Geometric anisotropy is a special case of range anisotropy models (Zimmerman, 1993), where the range parameter varies with the direction in lag-space. An example of non-geometric range anisotropy is the zonal anisotropy, where the range parameter is diferent along one single direction in lag-space. It can be shown that the zonal anisotropy is a degenerate case of geometric range anisotropy (Allard et al., 2016). Finite mixtures of zonal anisotropy models has been used in literature to model anisotropic covariance, for details see Chiles and Delfiner (2009). Another example of non-geometric range anisotropy is the separable models, i.e. $\varphi(\mathbf{h}) = \varphi^*(h_1)\,\varphi^*(h_2)$, where $\mathbf{h} = (h_1, h_2)'$. A more flexible non-geometric range anisotropy model has been proposed by Ecker and Gelfand (2003), where the authors have taken product of finitely-many geometric range anisotropy models to construct the anisotropic covariance. In a recent paper, Allard et al. (2016) has provided a general characterization of the range anisotropic models. They have shown that the range parameter can be written as a mixture of zonal anisotropies (where the range parameter is different for a single direction, for details see Allard et al. 2016) and

thus, they have proposed a kernel-based modeling of the anisotropic range, which can capture anisotropic contours that are beyond the scope of geometric anisotropy.

The methods mentioned here do not incorporate the directional variogram analysis into modeling, though often in practice, the presence of anisotropy is confirmed through the directional analysis. Also, anisotropic models that can be generated using isotropic covariances from different parametric classes have not been discussed in literature so far. In the next section, we provide a characterization of such covariance models and discuss anisotropic covariance estimation in details.

4.3 Anisotropy Modeling: A Copula-based Approach

In this section, we discuss how marginally fitted variograms – that are almost always computed in practice to check for anisotropy – can be used to estimate the joint covariance structure of the ϵ-process. Suppose the marginal semivariograms for directions $\Theta = \{\theta_1, \ldots, \theta_L\}$ are given by the class of semivariograms $\{\gamma_1(\cdot), \ldots, \gamma_L(\cdot)\}$ respectively. For identifiability of the directional semivariograms, we assume $0 \leq \theta_1 < \theta_2 < \cdots < \theta_L < \pi$.

4.3.1 Copula-based covariance specification

For the time being, let us consider the case $L = 2$. Then, under the assumption that $\{\epsilon(\mathbf{s})\}$ is a second-order stationary random field, the covariances along the directions θ_1 and θ_2 are given by $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ where,

$$\varphi_i(h) = \begin{cases} \sigma^2 - \gamma_i(h) & \text{if } h \neq 0 \\ \sigma^2 + \tau^2 & \text{if } h = 0. \end{cases}$$

Without loss of any generality, let us assume that there is no nugget component, i.e. as $h \to 0+$, $\sigma^2 - \varphi_i(h) \to 0$. The nugget component can easily be included in the covariograms to construct admissible covariances for second-order stationary processes, for details see page 85, Cressie (1993). We want to construct an admissible bivariate covariance function $\varphi(\mathbf{h}) : \mathbb{R}^2 \to \mathbb{R}$ such that the directional covariances along the orientations θ_1 and θ_2, the marginal covaraiances are given by $\varphi_1(h)$ and $\varphi_2(h)$. For time to time we will use polar coordinate system to represent the covariances in $\mathbb{R}^2$ for easier understanding. Let $(r, \theta)'$ be the polar representation of $\mathbf{h} = (h_x, h_y)$, where $r = \|\mathbf{h}\|_2$ and $\theta = \text{atan2}(h_y, h_x)$ (for details, see, page 42, Organick 1966). Hence, in terms of polar coordinates, we want to construct a covariance function $\varphi(\mathbf{h}) \equiv \varphi(r, \theta)$ on $\mathbb{R}^2$ such that $\varphi(r, \theta_i) = \varphi_i(r)$ for $i = 1, 2$ and all $r \geq 0$.

By Bochner's theorem (Bochner, 2005), for any $d \geq 1$, a continuous function $\varphi(\mathbf{h})$ is an admissible covariance in $\mathbb{R}^d$ (i.e. positive definite) if and only if there exist a positive, symmetric measure G on $\mathbb{R}^d$ such that,

$$\varphi(\mathbf{h}) = \int_{\mathbb{R}^d} \exp(\mathrm{i}\mathbf{h}'\boldsymbol{\omega})\, dG(\boldsymbol{\omega}). \tag{4.3.1}$$

Here, G is called the spectral measure and $F(\boldsymbol{\omega}) = G(\boldsymbol{\omega})/\varphi(0)$ is called the *spectral distribution function* (for details, see page 84, Cressie 1993) and if the density exists, i.e. $dF(\boldsymbol{\omega}) = f(\boldsymbol{\omega})d\boldsymbol{\omega}$ then $f(\cdot)$ is called the *spectral density function*. Clearly, a valid covariance in $\mathbb{R}^2$, $\varphi(\mathbf{h})$, can be thought as the characteristic function of a bivariate random variable $\boldsymbol{\omega} = (\omega_1, \omega_2)'$ with distribution function $F(d\boldsymbol{\omega})$. Let us denote the spectral distributions corresponding to $\varphi_1(h)$ and $\varphi_2(h)$ by $F_1(\cdot)$ and $F_2(\cdot)$ respectively. Also, by $C(\cdot; \boldsymbol{\eta}) : \mathbb{I}^2 \to \mathbb{I}$, we denote a bivariate copula with parameter $\boldsymbol{\eta}$ where $\mathbb{I} = [0, 1]$. Let us denote $F_{12}^*(\mathbf{u}; \boldsymbol{\eta}) = C(\mathbf{F}(\mathbf{u}); \boldsymbol{\eta})$ for $\mathbf{F}(\mathbf{u}) = (F_1(u_1), F_2(u_2))'$ and $\mathbf{u} = (u_1, u_2)' \in \mathbb{R}^2$;

clearly, by Sklar's theorem (Sklar, 1959), F_{12}^* is a bivariate distribution function with F_1 and F_2 as their marginal distribution functions. Also, φ_{12}^* is the bivariate characteristics function corresponding to the distribution function $F_{12}^*(\cdot;\ \boldsymbol{\eta})$, i.e.

$$\varphi_{12}^*(\mathbf{h};\boldsymbol{\eta}) = \int_{\mathbb{R}^2} \exp(\mathrm{i}\mathbf{h}'\mathbf{u})\ dC(\mathbf{F}\left(\mathbf{u}\right);\boldsymbol{\eta}). \tag{4.3.2}$$

Then the following proposition provides an admissible bivariate covariance with marginal covariances equal to φ_i in the direction θ_i for $i = \{1, 2\}$.

Proposition 4.1. *Let φ_1 and φ_2 be two continuous covariance functions on $\mathbb{R}$ along two different directions θ_1 and θ_2 in $[0, \pi)$ with corresponding spectral distributions F_1 and F_2 respectively. Then for any bivariate copula $C(\cdot;\boldsymbol{\eta}) : [0,1]^2 \to [0,1]$ with corresponding density $c(\cdot;\boldsymbol{\eta})$ symmetric around $(0.5, 0.5)'$,*

$$\varphi_{12}(\mathbf{h};\boldsymbol{\eta}) = \varphi_{12}^*\left((\mathbf{A}(\theta_1,\theta_2)')^{-1}\,\mathbf{h};\ \boldsymbol{\eta}\right) \tag{4.3.3}$$

is an admissible covariance on $\mathbb{R}^2$ with φ_i as the directional covariance along θ_i for $i \in \{1, 2\}$. Here, $\varphi_{12}^*(\cdot;\boldsymbol{\eta})$ is defined by Equation 4.3.2 and

$$\mathbf{A}\left(\theta_1,\theta_2\right) = \begin{pmatrix} \cos\theta_1 & \sin\theta_1 \\ \cos\theta_2 & \sin\theta_2 \end{pmatrix}.$$

Moreover, let $\varphi_{12}(\cdot;\boldsymbol{\eta})$ be a bivariate covariance function on $\mathbb{R}^2$ such that the corresponding spectral distribution function F_{12} admits a density denoted by f_{12}. Also, the marginal covariances of φ_{12} along θ_1 and θ_2 ($\theta_1 \neq \theta_2$; $\theta_1, \theta_2 \in [0, \pi)$) are given by φ_1 and φ_2 with corresponding spectral distributions F_1 and F_2 respectively, which are continuous everywhere. Then there exist a unique bivariate copula $C(\cdot;\boldsymbol{\eta})$ such that Equation 4.3.3 holds

true.

The proof of Proposition 4.1 is given in Section 4.7. Clearly, Proposition 4.1 provides us a formulation to construct admissible covariance structures that agree with the directional covariances in two different specified directions $\theta_1, \theta_2 \in [0, \pi)$. As the covariance function is symmetric with respect to origin, it is enough to consider the angles in $[0, \pi)$ only. The advantage of the class of covariances specified by Proposition 4.1 is that, it completely agrees with the directional covariances for the specified two directions; moreover, it is not necessary for the directional covariances to belong to the same class unlike the case of geometric anisotropy or the general range anisotropic models proposed by Allard et al. (2016).

The choice of copula determines the nature of interaction between the spectral random variables ω_1 and ω_2 thus specifies the nature of anisotropy for the underlying spatial process. For example, let $\theta_1 = 0$ and $\theta_2 = \pi/2$ and suppose that the corresponding marginal variograms are Gaussian variograms with range parameters b_1 and b_2. If we use standard bivariate normal copula with correlation parameter ρ, i.e. $C\left((u,v)';\rho\right) = \Phi_2^{(\rho)}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$, where $\Phi_2^{(\rho)}$ is the distribution function of the standard bivariate normal with correlation parameter ρ and Φ is the standard normal distribution function. Then, it can be easily shown that $\varphi_{12}(\mathbf{h};\rho) = \exp\left(-\left(\frac{h_x^2}{b_1^2} + \frac{2\rho h_x h_y}{b_1 b_2} + \frac{h_y^2}{b_2^2}\right)\right)$ which yields a special case of geometrically anisotropic Gaussian covariance structure where the minor axis or the major axis of the ellipse coincide with the Y-axis is the lag-space depending on which one of the b_1 and b_2 is larger. If $\rho = 0$, then $\varphi_{12}(\mathbf{h};\rho)$ becomes the separable covariance model. Clearly, the parameter associated with the copula decides the nature of anisotropy and hence, we call the parameter $\boldsymbol{\eta}$ anisotropy parameter for the covariance model given by Equation 4.3.3.

Now, if number of marginal variograms estimated from the data (i.e. L) is greater than 2, then using a ordered pair of semivariograms $\{\gamma_i, \gamma_j\}$ and a copula $C(\cdot; \boldsymbol{\eta}_{ij})$ we can construct a covariance structure $\varphi_{ij}(\cdot; \boldsymbol{\eta}_{ij})$ on $\mathbb{R}^2$ for $i \neq j \in \{1, \ldots, L\}$. Considering all such possible pairs of marginal semivariograms, the formulation in Proposition 4.1 provides us with a set of candidate anisotropic covariance models denoted by $\Phi = \left\{\varphi_{ij}(\cdot; \boldsymbol{\eta}_{ij}) : 1 \leq i < j \leq L\right\}$. In the next section, we discuss how to estimate the parameters and select the most 'optimal' covariance model from Φ using least squares based marginal variogram fitting.

4.3.2 Spatial covariance modeling

Suppose for direction θ_i, $i \in \{1, \ldots, L\}$, the marginal semivariogram is estimated through least squares based model fitting based on some nonparametric semivariogram estimator (for example, Matheron 1962, Cressie and Douglas 1980, etc.) computed at m_i-many different lags denoted by $\mathcal{H}_i = \left\{\mathbf{h}_{i_1}, \ldots, \mathbf{h}_{i_{m_i}}\right\}$ (for details on directional variogram model fitting, see chapter 3, Gelfand et al. 2010; chapter 2, Cressie 1993). Note that, for each $i \in \{1, \ldots, L\}$, in polar coordinate notation, $\mathbf{h}_{i_k} \equiv (r_{i_k}, \theta_i)'$ for $k \in \{1, \ldots, m_i\}$, where r_{i_k} is the distance of lag-vector $\mathbf{h}_{i_k}$ from origin. For a given class of bivariate copulas $\{C(\cdot; \boldsymbol{\eta}) : \boldsymbol{\eta} \in \Xi\}$, let us denote a (parametric) class of valid semivariograms as,

$$\Gamma_{ij} = \{\gamma_{ij} : \gamma_{ij}(\mathbf{h}; \boldsymbol{\eta}) = \varphi_{ij}(\mathbf{0}; \boldsymbol{\eta}) - \varphi_{ij}(\mathbf{h}; \boldsymbol{\eta}),\ \boldsymbol{\eta} \in \Xi\} \tag{4.3.4}$$

which obey the directional semivariograms in directions θ_i and θ_j. We propose to choose the 'best' variogram model from the class Γ_{ij} using a weighted least squares based

goodness-of-fit criteria as follows:

$$\hat{\boldsymbol{\eta}}_{ij} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}}\, Q_{ij}(\boldsymbol{\eta}) = \underset{\boldsymbol{\eta}}{\operatorname{argmin}} \sum_{l\in\{1,\ldots,L\}} \sum_{k=1}^{m_l} w(\mathbf{h}_{l_k}) \left\{2\gamma^{\#}\left(\mathbf{h}_{l_k}\right) - 2\gamma_{ij}\left(\mathbf{h}_{l_k};\boldsymbol{\eta}\right)\right\}^2, \quad (4.3.5)$$

where $\gamma^{\#}(\cdot)$ is some nonparametric semivariogram estimator, that has been used to fit the directional models as well, e.g. the classical semivariogram estimator by Matheron (1962) or more robust versions of it, e.g. see Cressie and Douglas (1980), Genton (1998) etc. Clearly, the criterion set by Equation 4.3.5 chooses the parametric class of covariance that completely obeys the directional fits along θ_i and θ_j and agrees with the other directional nonparametric semivariogram estimates as 'closely' as possible in terms of the weighted sum of squares distance from its sample analog. It is easy to follow that minimizing $Q_{ij}(\cdot)$ w.r.t. $\boldsymbol{\eta}$ is same as minimizing

$$Q^*_{ij}(\boldsymbol{\eta}) = \sum_{l\in\{1,\ldots,L\}\setminus\{i,j\}} \sum_{k=1}^{m_l} w(\mathbf{h}_{l_k}) \left\{2\gamma^{\#}\left(\mathbf{h}_{l_k}\right) - 2\gamma_{ij}\left(\mathbf{h}_{k_l};\boldsymbol{\eta}\right)\right\}^2,$$

as $Q_{ij}(\boldsymbol{\eta}) = Q^*_{ij}(\boldsymbol{\eta}) + C_{ij}$ and,

$$C_{ij} = \sum_{l\in\{i,j\}} \sum_{k=1}^{m_l} w(\mathbf{h}_{l_k}) \left\{2\gamma^{\#}\left(\mathbf{h}_{l_k}\right) - 2\gamma_{l}\left(\mathbf{h}_{k_l};\boldsymbol{\eta}\right)\right\}^2,$$

is a constant with respect to $\boldsymbol{\eta}$. This shows that in the estimation of the parameter of anisotropy, the directional fits used to construct the class of covariances do not contribute. For different choices of weights in $Q_{ij}(\cdot)$ we get different least squares based estimator of anisotropy parameter, e.g. if all the weights are equal then we get the ordinary least squares (OLS) estimator and for $w(\mathbf{h}_{l_k}) = |N(\mathbf{h}_{l_k})| / \left(\gamma_{ij}\left(\mathbf{h}_{l_k};\boldsymbol{\eta}\right)\right)^2$ yields the widely used weighted least squares (WLS) estimator. More complex variogram model fitting

like generalised least squares (GLS) can be incorporated but, without any distributional assumption on the random field $\{\epsilon(\cdot)\}$ the optimization problem in GLS estimation is not straightforward, for more details in GLS-based variogram model fitting see Chapter 2, Cressie (1993).

Once the anisotropy parameters are estimated through least squares based model fitting for all possible pairs of directional variograms, we have a finite set of candidate covariance structures denoted by $\Phi' = \left\{\varphi_{ij}(\cdot;\hat{\boldsymbol{\eta}}_{ij}) : 1 \leq i < j \leq L\right\}$. To choose from this finite class of covariances we again propose to the goodness-of-fit criteria given in Equation 4.3.5, i.e.,

$$\varphi^*(\cdot) = \underset{\varphi_{ij}\in\Phi'}{\operatorname{argmin}} \left\{Q_{ij}\left(\hat{\boldsymbol{\eta}}_{ij}\right) : 1 \leq i < j \leq L\right\}. \tag{4.3.6}$$

Clearly, the chosen covariance model φ^* is an admissible one, it agress with two of the directional models perfectly, and it is the 'closest' model to the data in the sense of the weighted sum of squares of errors distance. To prevent from overfitting, one can use the various information criterion based model selection techniques as well, e.g. AICc, BIC etc. For this setup, AICc and BIC can be computed as follows:

$$\mathrm{AICc}_{ij} = 2k + m\log\left(\frac{Q_{ij}\left(\hat{\boldsymbol{\eta}}_{ij}\right)}{m}\right) + \frac{2k^2+2k}{m-k-1}; \quad \mathrm{BIC}_{ij} = \log(m)\,k + m\log\left(\frac{Q_{ij}\left(\hat{\boldsymbol{\eta}}_{ij}\right)}{m}\right), \tag{4.3.7}$$

where k is the total number of parameters estimated, i.e. numer parameters estimated in the directional fits along θ_i and θ_j and number of anisotropy parameters in the specified copula $C(\cdot;\boldsymbol{\eta})$. Here, $m = \sum_{l=1}^{L} m_l$ is the total number of discrete lags for which the nonparametric semivariogram $\gamma^{\#}$ has been computed.

Though for given directional variograms for L-many directions the total number of candidate models is $|\Phi'| = L(L-1)/2$, we recommend to combine only the pairs for which the orientations are well separated, i.e. only combine γ_i and γ_j to get the joint covariance φ_{ij} if $\theta_j - \theta_i > \theta_\delta$ for some specified $\theta_\delta \in (0, \pi)$, e.g. say $\theta_\delta = \pi/4$. The reason is that, if two variograms along directions very 'close' to each other is combined then same pairs of observations might be used to compute their corresponding nonparametric semivariograms creating complexities regarding degeneracy, specially in case of irregularly spaced observations where 'tolerance regions' are used to compute the generic variogram estimators (for details, see Chapter 2, Cressie 1993).

In the next subsection, we discuss some large-sample properties of the proposed covariance modeling and model estimators.

4.3.3 Asymptotic properties of the covariance estimator

For the results stated in this subsection, let us introduce some new notations and redefine some of the already used ones. Let the directional univariate covariances (or variograms) along directions θ_1 and θ_2 $(0 \leq \theta_1 < \theta_2 < \pi)$ are given by $\varphi_1(h; \boldsymbol{\alpha}_1)$ (or $2\gamma_1(h; \boldsymbol{\alpha}_1)$) and $\varphi_2(h; \boldsymbol{\alpha}_2)$ (or $2\gamma_2(h; \boldsymbol{\alpha}_2)$), where $\boldsymbol{\alpha}_i$ is the vector of parameters for directional models along θ_i, for $i = 1, 2$. Moreover, suppose the joint covariance function (or variogram) is denoted by $\varphi(\mathbf{h}; \boldsymbol{\nu})$ (or $2\gamma(\mathbf{h}; \boldsymbol{\nu})$), where $\boldsymbol{\nu}$ is the vector of all parameters, i.e. $\boldsymbol{\nu} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', \boldsymbol{\eta}')'$ and by Proposition 4.1, $\varphi(\cdot; \boldsymbol{\nu})$ has the representation in Equation 4.3.3 for some bivariate copula $C(\cdot; \boldsymbol{\eta})$ and univariate spectral distributions $F_i(\cdot; \boldsymbol{\alpha}_i)$ corresponding to covariances $\varphi_i(\cdot; \boldsymbol{\alpha}_i)$, for $i \in \{1, 2\}$. Without loss of any generality, let us assume that $\theta_1 = 0$ and $\theta_2 = \pi/2$, as otherwise we can get $\varphi(\mathbf{h}; \boldsymbol{\nu})$ by multiplying the lag-vector $\mathbf{h}$ by the rotation matrix $(\mathbf{A}'(\theta_1, \theta_2))^{-1}$.

Suppose $2\gamma_n^{\#}(\mathbf{h})$ denotes a generic nonparametric estimator of the variogram at lag $\mathbf{h}$ where n is the number of spatial observations. Then the weighted least squares (WLS) estimator of the directional model parameters are given by,

$$\hat{\boldsymbol{\alpha}}_{i,n} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\, Q_{i,n}^{\#}(\boldsymbol{\alpha}) = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{k=1}^{m_i} w(\mathbf{h}_{i_k};\boldsymbol{\alpha}) \left\{2\gamma_n^{\#}(\mathbf{h}_{i_k}) - 2\gamma_i(\|\mathbf{h}_{i_k}\|;\boldsymbol{\alpha})\right\}^2, \quad (4.3.8)$$

where $\|\mathbf{x}\|$ denotes the norm of vector $\mathbf{x}$. As discussed previously, the weights can involve the covariance parameters as well. From time to time, for better understanding we will refer $\varphi(\cdot;\boldsymbol{\nu})$ as $\varphi(\cdot;\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\eta})$ and similar for variograms. Hence, as mentioned in the last subsection, the anisotropy parameter is estimated as $\hat{\boldsymbol{\eta}}_n = \underset{\boldsymbol{\eta}}{\operatorname{argmin}}\, Q_n^{\#}(\boldsymbol{\eta})$ where,

$$Q_n^{\#}(\boldsymbol{\eta}) = \sum_{l=1}^{L}\sum_{k=1}^{m_l} w(\mathbf{h}_{l_k};\hat{\boldsymbol{\alpha}}_{1,n},\hat{\boldsymbol{\alpha}}_{2,n},\boldsymbol{\eta}) \left\{2\gamma_n^{\#}(\mathbf{h}_{l_k}) - 2\gamma(\mathbf{h}_{l_k};\hat{\boldsymbol{\alpha}}_{1,n},\hat{\boldsymbol{\alpha}}_{2,n},\boldsymbol{\eta})\right\}^2. \quad (4.3.9)$$

Let us denote the set of discrete lags in direction θ_l for which the generic semivariogram estimator is computed as $\mathcal{H}_l = \{\mathbf{h}_{l_k} : 1 \le k \le m_l\}$ for $l \in \{1,\ldots,L\}$; and $\mathcal{H} = \cup_{l=1}^{L}\mathcal{H}_l$ with $|\mathcal{H}| = \sum_{l=1}^{L} m_l = m$. For simplicity of notations we often refer the elements of $\mathcal{H}$ as $\mathbf{h}_i$ for $1 \le i \le m$.

Next, we introduce some more notations that are required to state the conditions under which the consistency and asymptotic normality results have been proved. Observe that, similar to the notations given in Lahiri et al. (2002), $Q_n^{\#}(\boldsymbol{\eta})$ can be written as $Q_n^{\#}(\boldsymbol{\eta}) = \mathbf{g}_n^{\#}(\boldsymbol{\eta})' \mathbf{W}(\hat{\boldsymbol{\alpha}}_{1,n},\hat{\boldsymbol{\alpha}}_{2,n},\boldsymbol{\eta})\, \mathbf{g}_n^{\#}(\boldsymbol{\eta})$ where,

$$\mathbf{g}_n^{\#}(\boldsymbol{\eta}) = \left(2\gamma_n^{\#}(\mathbf{h}_1) - 2\gamma(\mathbf{h}_1;\hat{\boldsymbol{\alpha}}_{1,n},\hat{\boldsymbol{\alpha}}_{2,n},\boldsymbol{\eta}), \ldots, 2\gamma_n^{\#}(\mathbf{h}_m) - 2\gamma(\mathbf{h}_m;\hat{\boldsymbol{\alpha}}_{1,n},\hat{\boldsymbol{\alpha}}_{2,n},\boldsymbol{\eta})\right)',$$

and $\mathbf{W}(\boldsymbol{\nu}) = \operatorname{diag}\left((w(\mathbf{h}_1;\boldsymbol{\nu}),\ldots,w(\mathbf{h}_m;\boldsymbol{\nu}))'\right)$.

Now, we give the conditions under which the WLS estimators of the covariance parameters are consistent and asymptotically normal. As discussed by Lahiri et al. (2002), the asymptotic properties of the least squares based estimators are determined by the asymptotic behaviors of the generic nonparametric semivariogram estimator $\gamma_n^{\#}(\cdot)$. In this work, we provide the asymptotic results under general assumptions of consistency and asymptotic normality of the generic estimator $\gamma_n^{\#}(\cdot)$. For consistency of the WLS covariance estimators that has been discussed in this Chapter so far, we need the following regularity conditions.

(C.1) For any $\varepsilon > 0$, there exist a $\delta > 0$ such that,

i. $\inf\left\{\sum_{\mathbf{h}\in\mathcal{H}_i}\left(2\gamma_i\left(\|\mathbf{h}\|;\boldsymbol{\alpha}_i\right) - 2\gamma_i\left(\|\mathbf{h}\|;\boldsymbol{\alpha}_i^*\right)\right)^2 : \|\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*\| \geq \varepsilon\right\} > \delta$, for $i \in \{1,2\}$,

ii. $\inf\left\{\sum_{\mathbf{h}\in\mathcal{H}}\left(2\gamma\left(\mathbf{h};\boldsymbol{\nu}\right) - 2\gamma\left(\mathbf{h};\boldsymbol{\nu}^*\right)\right)^2 : \|\boldsymbol{\nu} - \boldsymbol{\nu}^*\| \geq \varepsilon\right\} > \delta$.

(C.2) The marginal covariance models $\varphi_i(\cdot;\boldsymbol{\alpha}_i)$ for $i = 1,2$ are such that,

i. densities of the corresponding spectral distributions exists, i.e. $dF_i(h;\boldsymbol{\alpha}_i) = f_i(h;\boldsymbol{\alpha}_i)\,dh$ for some $f_i(h;\boldsymbol{\alpha}_i) \geq 0$.

ii. $\varphi_i(h;\boldsymbol{\alpha}_i)$ is continuous in $\boldsymbol{\alpha}_i \in \mathcal{A}_i$ for all $h \in \mathbb{R}$,

iii. $\sup\left\{\left|\varphi_i(h;\boldsymbol{\alpha}_i)\right| : h \in \mathbb{R}, \boldsymbol{\alpha}_i \in \mathcal{A}_i\right\} < \infty$.

(C.3) $C(\cdot;\boldsymbol{\eta}) : \mathbb{I}^2 \to \mathbb{I}$ is such a copula that,

i. the density of $C(\cdot;\boldsymbol{\eta})$ exists, i.e. $dC(\mathbf{u};\boldsymbol{\eta}) = c(\mathbf{u};\boldsymbol{\eta})\,d\mathbf{u}$,

ii. $\int_{\mathbb{I}^2} \sup_{\boldsymbol{\eta}} c(\mathbf{u};\boldsymbol{\eta})\,d\mathbf{u} < \infty$.

(C.4) The weights used in WLS estimations satisfy the following:

i. $\inf\left\{w\left(\mathbf{h}^*;\boldsymbol{\alpha}_i\right), w\left(\mathbf{h};\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\eta}\right) : \mathbf{h}^* \in \mathcal{H}_i, \boldsymbol{\alpha}_i \in \mathcal{A}_i, i = 1,2; \mathbf{h} \in \mathcal{H}, \boldsymbol{\eta} \in \Xi\right\} > 0$.

ii. $\sup\{w(\mathbf{h}^*;\boldsymbol{\alpha}_i), w(\mathbf{h};\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\eta}) : \mathbf{h}^* \in \mathcal{H}_i, \boldsymbol{\alpha}_i \in \mathcal{A}_i, \mathbf{h} \in \mathcal{H}, \boldsymbol{\eta} \in \Xi\} < \infty$.

Condition (C.1) states that both the directional as well as the bivariate veriogram functions have to be identifiable at the considered discrete lags through the set of parameters, i.e. for two different sets of parameters, the variograms can not have same values for all the discrete lag-vectors that has been used in the WLS estimations. Condition (C.2) imposes smoothness and continuity assumptions on the directional variogram models $2\gamma_i(;\boldsymbol{\alpha}_i)$ which are required for consistency of the directional parameters. In assumption (C.3), we state the required regularity conditions on the joint covariance function through the bivariate copula, as the bivariate copula along with the two marginal covariance models completely specifies the joint covariance. The conditions in (C.3) ensure the continuity of the bivariate covariance with respect to the parameter of anisotropy $\boldsymbol{\eta}$. Condition (C.4) states that the weights used in the WLS estimations are positive and finite for all choices of covariance parameters.

Now we will introduce a couple of conditions that are needed in addition to (C.1) - (C.3) to ensure the asymptotic normality of the covariance parameter estimators.

(C.5) $\varphi_i(\cdot;\boldsymbol{\alpha}_i)$ has continuous first order partial derivatives with respect to $\boldsymbol{\alpha}_i$ for all $\boldsymbol{\alpha}_i \in \mathcal{A}_i, i \in \{1,2\}$.

(C.6) i. $c(\cdot;\boldsymbol{\eta})$ has continuous first order partial derivatives with respect to $\boldsymbol{\eta}$ for all $\boldsymbol{\eta} \in \Xi$,

ii. $\int_{\mathbb{I}^2} \sup_{\boldsymbol{\eta}} |\frac{\partial}{\partial \eta^{(a)}} c(\mathbf{u};\boldsymbol{\eta})| d\mathbf{u} < \infty$ for all $1 \le a \le q$,

where $\boldsymbol{\eta} = \left(\eta^{(1)}, \ldots, \eta^{(q)}\right)'$.

(C.5) and (C.6) impose further smoothness conditions on the directional and joint covariance models that is required to establish the asymptotic normality of the covariance

parameter estimators. Now, we will state the main results of this subsection. Let us denote the true values of $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ as $\boldsymbol{\alpha}_{1,0}$, $\boldsymbol{\alpha}_{2,0}$ and $\boldsymbol{\eta}_0$ and $\boldsymbol{\nu}_0$ respectively. Also, we denote the WLS-based covariance parameter estimators as $\hat{\boldsymbol{\nu}}_n = \left(\hat{\boldsymbol{\alpha}}'_{1,n}, \hat{\boldsymbol{\alpha}}'_{2,n}, \hat{\boldsymbol{\eta}}'_n\right)'$. Observe that, under conditions (C.1), (C.2) and (C.4) the consistency, and under (C.1), (C.2), (C.4) and (C.5) the asymptotic normality of $\hat{\boldsymbol{\alpha}}_{i,n}$ directly follows from Theorem 3.1 and Theorem 3.2 of Lahiri et al. (2002). Hence, in this subsection, we only provide asymptotic results for the WLS estimator of anisotropy parameter $\hat{\boldsymbol{\eta}}_n$.

Theorem 4.1. *Under conditions (C.1) - (C.3) and under the assumption*

$$\gamma_n^{\#}(\mathbf{h}) \to \gamma(\mathbf{h}; \boldsymbol{\nu}_0) \quad \text{as } n \to \infty \text{ a.s., for all } \mathbf{h} \in \mathcal{H}, \tag{4.3.10}$$

$\hat{\boldsymbol{\eta}}_n$ is strongly consistent for $\boldsymbol{\eta}$, i.e. $\hat{\boldsymbol{\eta}}_n \to \boldsymbol{\eta}_0$ a.s.. Moreover, in case the consistency result in 4.3.10 holds in probability, $\hat{\boldsymbol{\eta}}_n$ is weakly consistent.

Hence, under conditions (C.1) - (C.3) the WLS estimator $\hat{\boldsymbol{\eta}}_n$ is consistent whenever the corresponding generic semivariogram $\gamma_n^{\#}$ is consistent at the lags $\mathcal{H} = \{\mathbf{h}_1, \ldots, \mathbf{h}_m\}$. Under suitable stationarity and mixing conditions on the random field $\{\epsilon(\mathbf{s})\}$, most of the nonparametric semivariogram estimators – e.g. the method of moments estimator by Matheron (1962), the robust semivariogram estimator by Cressie and Douglas (1980) etc. – are pointwise consistent. Hence, any of these semivariograms can be used to ensure the consistency of the WLS based covariance parameter estimators.

Before stating the asymptotic normality result, let us introduce some more notations. Let us define,

$$\mathbf{g}_n^*(\boldsymbol{\eta}) = \left(2\gamma_n^{\#}(\mathbf{h}_1) - 2\gamma(\mathbf{h}_1; \boldsymbol{\alpha}_{1,0}, \boldsymbol{\alpha}_{2,0}, \boldsymbol{\eta}), \ldots, 2\gamma_n^{\#}(\mathbf{h}_m) - 2\gamma(\mathbf{h}_m; \boldsymbol{\alpha}_{1,0}, \boldsymbol{\alpha}_{2,0}, \boldsymbol{\eta})\right)',$$

and for $1 \leq a \leq q$, $\boldsymbol{\gamma}^{(a)}(\boldsymbol{\eta}_0)$ is the $m \times 1$ vector of partial derivatives of $\{\gamma(\mathbf{h}_1; \boldsymbol{\alpha}_{1,0}, \boldsymbol{\alpha}_{2,0}, \boldsymbol{\eta}), \ldots, \gamma(\mathbf{h}_m; \boldsymbol{\alpha}_{1,0}, \boldsymbol{\alpha}_{2,0}, \boldsymbol{\eta})\}$ with respect to $\eta^{(a)}$ and evaluated at $\boldsymbol{\eta}_0$. Also, let $\Gamma(\boldsymbol{\eta}_0) = \left(-2\boldsymbol{\gamma}^{(1)}(\boldsymbol{\eta}_0), \ldots, -2\boldsymbol{\gamma}^{(q)}(\boldsymbol{\eta}_0)\right)$ is the $m \times q$ matrix containing the first order partial derivatives of $\mathbf{g}_n^*(\boldsymbol{\eta})$ at $\boldsymbol{\eta}_0$.

Theorem 4.2. *Let conditions (C.1)-(C.5) hold true. Moreover, there exists a sequence of positive reals* $\{a_n\}$ *such that* $a_n \to \infty$ *as* $n \to \infty$ *and*

$$a_n \mathbf{g}_n^*(\boldsymbol{\eta}_0) \xrightarrow{d} \boldsymbol{N}(\mathbf{0}, \Sigma(\boldsymbol{\eta}_0)), \tag{4.3.11}$$

for some nonsingular matrix $\Sigma(\boldsymbol{\eta}_0)$*. In addition, let the* $m \times q$ *matrix* $\Gamma(\boldsymbol{\eta}_0)$ *is of full column rank. Then, as* $n \to \infty$*,*

$$a_n(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) \xrightarrow{d} \boldsymbol{N}(\mathbf{0}, \Sigma_W(\boldsymbol{\eta}_0)),$$

where

$$\Sigma_W(\boldsymbol{\eta}_0) = B(\boldsymbol{\eta}_0)\Gamma(\boldsymbol{\eta}_0)' W(\boldsymbol{\nu}_0)\Sigma(\boldsymbol{\eta}_0) W(\boldsymbol{\nu}_0)\Gamma(\boldsymbol{\eta}_0) B(\boldsymbol{\eta}_0), \tag{4.3.12}$$

where $B(\boldsymbol{\eta}_0) = \left(\Gamma(\boldsymbol{\eta}_0)' W(\boldsymbol{\nu}_0)\Gamma(\boldsymbol{\eta}_0)\right)^{-1}$.

Thus, if the nonparametric semivariogram scaled by a scalar a_n follows a normal distribution for large sample size, then the WLS estimator of anisotropic parameter also has an asymptotic normal distribution with the same scaling constant. The asymptotic normality of the nonparametric semivariogram can be established and the scaling constant can be determined by putting sufficient regularity (mixing) conditions on the underlying random field $\{\epsilon(\mathbf{s})\}$ and the spatial sampling design that generates the data locations.

For details of specific cases, e.g. data on integer lattice or irregularly spaced data under a *mixed-increasing-domain* spatial asymptotic framework see Lahiri et al. (2002). The proofs of the results of this subsection are given in Section 4.7.

4.4 Illustration: Bivariate Covariance Modeling Using Directional Variograms

In this section, we will provide some examples of constructing bivariate covariances using given directional valid variograms in two different directions. For the examples discussed below, without any loss of generality we will take $\theta_1 = 0$ and $\theta_2 = \pi/2$. In this case, the rotation matrix $\mathbf{A}(\theta_1, \theta_2) = \mathbf{I}_2$ where $\mathbf{I}_2$ is the 2×2 identity matrix and $\varphi(\cdot) = \varphi^*(\cdot)$ from Equation 4.3.3. For general θ_1 and θ_2 the corresponding bivaraite covariance can be easily computed by rotating the lag-vector $\mathbf{h}$ using the rotation matrix $\mathbf{A}^{-1}$.

4.4.1 Normal copula based covariance models

One of the most commonly used copulas in quantitative sciences is the standard bivariate normal copula given by $C\left((u,v)';\rho\right) = \Phi_2^{(\rho)}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$ where Φ is the distribution function of the univariate standard normal and $\Phi_2^{(\rho)}(\cdot,\cdot)$ is the distribution function of the standard bivariate normal with correlation coefficient $\rho \in (-1,1)$. For $\rho \in \{-1,1\}$ the normal copula can be extended continuously in ρ and calculation involving simple calculus shows that the normal copula attains the lower and upper Fréchet-Hoeffding bounds when $\rho = -1$ and 1 respectively. Moreover, this class of copulas allow '*total concordance ordering*' in terms of the parameter of anisotrpy ρ. For more details in bivariate normal copula see Meyer (2013).

Given two univariate covariances φ_1 and φ_2 with their corresponding spectral distribution functions F_1 and F_2 respectively,

$$\varphi_{12}(\mathbf{h};\rho) = \int_{\mathbb{R}^2} \exp(\mathrm{i}\mathbf{h}'\boldsymbol{\omega})\, d\Phi_2\left(\Phi^{-1}(F_1(\omega_1)), \Phi^{-1}(F_2(\omega_2)); \rho\right) \tag{4.4.1}$$

is the joint covariance with marginal covariograms φ_1 and φ_2 along $\theta_1 = 0$ and $\theta_2 = \pi/2$ respectively. Clearly, for general spectral distributions F_1 and F_2, it is not straightforward to compute the integral in Equation 4.4.1 as the normal distribution functions can only be computed numerically. But using the following lemma, the covariance given in Equation 4.4.1 can be approximated quite accurately using Monte Carlo methods.

Lemma 4.1. *Let ω_i for $i \in \{1,2\}$ be two univariate random variables with cumulative distribution functions F_i, which are continuous everywhere. Then the characteristic function $\varphi_{\boldsymbol{\omega}}(\mathbf{h})$ of the bivariate random variable $\boldsymbol{\omega} = (\omega_1, \omega_2)'$, obtained by combining combining the distributions F_1 and F_2 with a bivariate copula $C(\cdot)$ with corresponding copula density $c(\cdot)$, is same as the characteristic function of the bivariate random variable $\mathbf{F}^{-1}(\mathbf{U}) = \left(F_1^{-1}(U_1), F_2^{-1}(U_2)\right)'$ where $\mathbf{U} = (U_1, U_2)'$ is the bivariate random variable with density $c(\cdot)$, i.e.*

$$\varphi_{\boldsymbol{\omega}}(\mathbf{h}) = \int_{\mathbb{I}^2} \exp\left(\mathrm{i}\mathbf{h}'\mathbf{F}^{-1}(\mathbf{u})\right)\; c(\mathbf{u})\, d\mathbf{u}, \tag{4.4.2}$$

where $\mathbb{I}^2 = [0,1] \times [0,1]$.

Corollary 4.1. *If bivaraite normal copula with correlation parameter ρ is used, i.e. $C\left((u,v)';\rho\right) = \Phi_2^{(\rho)}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$, then*

$$\varphi_{\boldsymbol{\omega}}(\mathbf{h}) = \int_{\mathbb{R}^2} \exp(\mathrm{i}\mathbf{h}'\mathbf{g}(\mathbf{x}))\; \Phi^{(\rho)}(\mathbf{x})\, d\mathbf{x}, \tag{4.4.3}$$

where $\mathbf{g}(\mathbf{x}) = \left(F_1^{-1}(\Phi(x_1)), F_2^{-1}(\Phi(x_2))\right)'$ *and* $\mathbf{x} = (x_1, x_2)'$.

Equation 4.4.3 can be further simplified as $\varphi_{\boldsymbol{\omega}}(\mathbf{h}) = E\left(\cos\left(\mathbf{h}'(\mathbf{g}(\mathbf{X}))'\right)\right)$ if the ω_i's are marginally symmetric around 0 – which is the case here as the corresponding characteristic functions, i.e. the marginal covariance functions are real-valued. Clearly, $\varphi_{12}(\mathbf{h};\rho)$ can be easily approximated through Monte Carlo methods by generating large number (say, B) of random samples $\{\mathbf{X}_1, \ldots, \mathbf{X}_B\}$ from the standard bivariate normal distribution, denoted by $N_2(0,0;1,1;\rho)$ and then $\hat{\varphi}_{12}(\mathbf{h};\rho) = B^{-1}\sum_{b=1}^{B}\cos(\mathbf{h}'\mathbf{g}(\mathbf{X}_b))$ is an $\sqrt{B}$-consistent approximation of $\varphi_{12}(\mathbf{h};\rho)$.

4.4.1.1 Exponential-Gaussian covariance

In this example, we illustrate how a Gaussian and an Exponential marginal covariances can be combined using bivariate normal copula to construct a valid covariance structure over the space. Recall the Exponential and Gaussian covariances,

$$\varphi_{\text{exp}}(h) = \sigma^2 \exp\left(-\frac{|h|}{\alpha_e}\right); \quad \varphi_{\text{gau}}(h) = \sigma^2 \exp\left(-\frac{h^2}{\alpha_g^2}\right),$$

where α_e and α_g are the corresponding range parameters and σ^2 is the variance of the second-order stationary process $\{\epsilon(\mathbf{s})\}$. Without loss of any generality we take $\sigma^2 = 1$ (also, the nugget component $\tau^2 = 0$). Then, it is easy to show that the spectral random variables corresponding to φ_{exp} and φ_{gau} follow Cauchy$(0, \alpha_e^{-1})$ and Normal$(0, 2\alpha_g^{-2})$ respectively. For both of these distributions inverse cumulative distribution functions can be numerically computed with high precision. Using Corollary 4.1 we have illustrated the Exponential-Gaussian covariance for different sets of parameters in Figure 4.1.

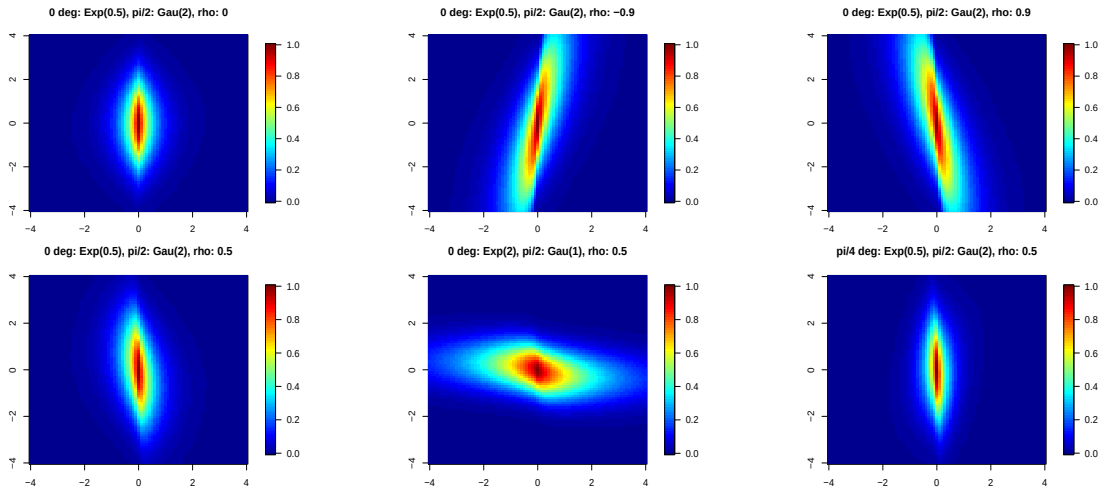


Figure 4.1: Illustration of the covariance functions obtained by combining univariate Exponential and Gaussian covariances.

4.4.1.2 Matérn-Matérn covariance

Matérn is one of the most widely used parametric classes of spatial covariance which we have introduced in Equation 2.4.2 for the simulation study section in Chapter 2. We define it below again to make the notation homogeneous throughout this chapter.

$$\varphi_{\text{mat}}(r;\boldsymbol{\alpha}) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{r}{\alpha_m}\right)^{\kappa} K_{\kappa}\left(\frac{r}{\alpha_m}\right), \tag{4.4.4}$$

where, $\Gamma(\cdot)$ is the Gamma function, K_κ is the Modified Bessel function of third kind with order κ, and $\boldsymbol{\alpha} = (\alpha_m, \kappa)'$ are the range and smoothness parameters respectively. As pointed out by Stein (2012), it is practical to use this class of covariance as it is differentiable everywhere and the degree of smoothness of the process over space can be controlled through the smoothness parameter κ. It can be shown that the Exponential, Gaussian and Whittle covariances are special cases of the Matérn class and thus, Matérn provides a richer parametric class of isotropic covariance models to choose from. For more

details about Matérn covariance see Chapter 4 in Haskard (2007).

When it comes to anisotropy modeling within the Matérn class, the most common practice is to use the geometrically anisotropic Matérn (or often referred as the anisotropic Matérn) where, as discussed in Section 4.2, it is assumed that there exists a rotation matrix such that upon rotation of the lag-space, the covariance function is isotropic in nature. Haskard (2007) have discussed this class of covaraince in details in Chapter 5 and 6 of his dissertation. But as pointed out by Allard et al. (2016), though it is mathematically and computationally convenient to implement, geometrically anisotropic Matérn is only a special case of anisotropy and is restrictive from the assumption that the contours of the covariance function has to be an ellipsoid. Under geometric anisotropy the directions of the maximum range (i.e. the direction along which the spatial correlation is strongest) and the minimum range (i.e. along which the spatial correlation is weakest) has to be seperated by exactly $\frac{\pi}{2}$. A much broader class of anisotropic Matérn can be generated using the kernel-based scaling approach proposed by Allard et al. (2016). But, the covariance proposed by Allard et al. (2016) suffers from the fact that the smoothness parameter can not be different for any two different directions over the space.

With our approach, we can combine two Matérn covariances, say $\varphi_{\text{mat}}(\cdot;\boldsymbol{\alpha}_1)$ and $\varphi_{\text{mat}}(\cdot;\boldsymbol{\alpha}_2)$ with different sets of parameters $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. In Figure 4.2, we have illustrated the resulting covariance functions over the space for different sets of parameters. The marginal covariances along an arbitrary direction $\theta' \in [0, \pi)$, other than the two directions that has been used to construct the joint covariance function, is the characteristic function of a linear combination of two random variables with their corresponding marginal characteristic functions given by univariate Matérns with parameters $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. Clearly, this class of covariances are more general than the geometric anisotropy as the contours can have different shapes and orientations other than ellipsoid and also the

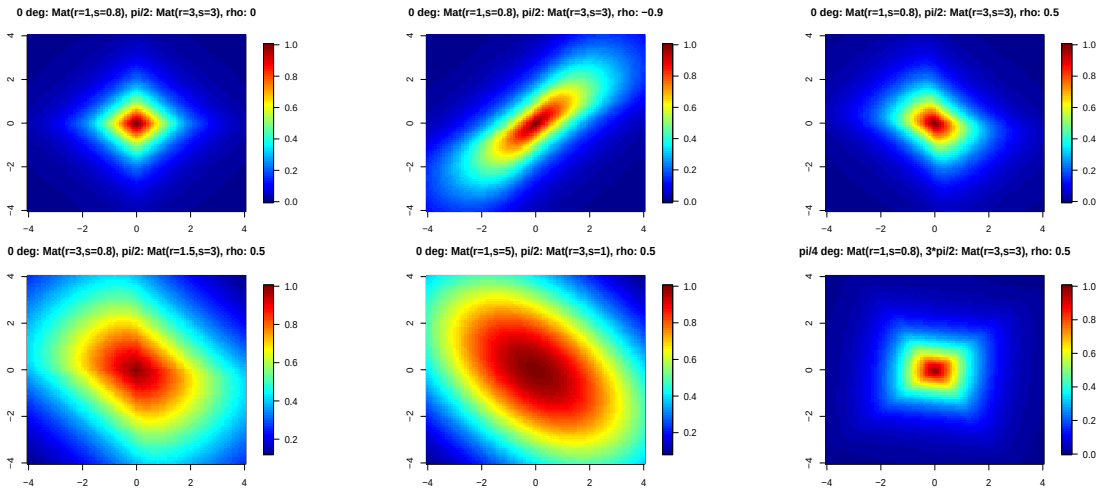


Figure 4.2: Illustration of the covariance functions obtained by combining two univariate Matérn covariances.

smoothness of the process cab be different along different directions over the space.

4.4.1.3 Spherical-Gaussian covariance

Next, we illustrate another class of covariances which has been constructed by combining the univariate Spherical and Gaussian covariances. The univaraite Spherical covariance is given by,

$$\varphi_{\text{sph}}(r;\alpha_s) = \begin{cases} 1 - \frac{3}{2}\left(\frac{r}{\alpha_s}\right) + \frac{1}{2}\left(\frac{r}{\alpha_s}\right)^3 & \text{if } r \leq \alpha_s, \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.5}$$

Spherical model is often used in geostatistics, especially when the covariance between two points decays slowly (at cubic rate as compared to exponential or squared exponential) but it is negligible after a certain lag. To obtain a joint covariance with a Spherical model as the marginal along one direction using our copula-based method, we need to obtain the spectral distribution function of the Spherical covariance given in Equation 4.4.5. With

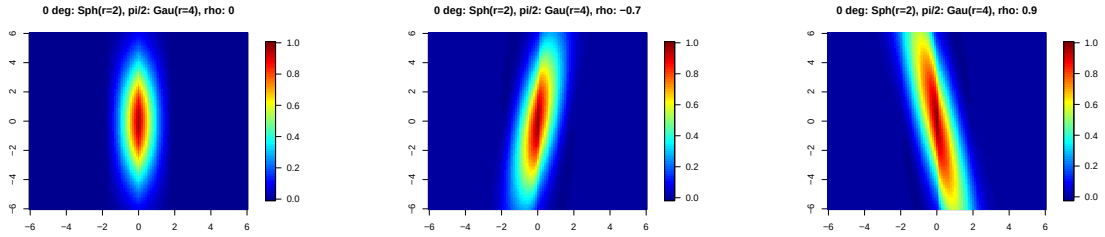


Figure 4.3: Illustration of the covariance functions obtained by combining univariate Spherical and Gaussian covariances.

simple integral calculus it can be shown that the spectral density $f_{\text{sph}}(\cdot;\alpha_s)$ is given by,

$$f_{\text{sph}}(\omega;\alpha_s) = \begin{cases} \frac{\alpha_s}{\pi}\left\{\frac{\sin(\alpha_s\omega)}{\alpha_s\omega} - \frac{3}{2}\left(\frac{\alpha_s\omega\sin(\alpha_s\omega)+\cos(\alpha_s\omega)-1}{(\alpha_s\omega)^2}\right)\right. \\ \left. \quad +\frac{1}{2}\left(\frac{(\alpha_s\omega(\alpha_s^2\omega^2-6)\sin(\alpha_s\omega_s)+3(\alpha_s^2\omega^2-2)\cos(\alpha_s\omega))}{(\alpha_s\omega)^4}\right)\right\} & \text{if } \omega \neq 0, \\ \frac{3\alpha_s}{8\pi} & \text{otherwise.} \end{cases} \tag{4.4.6}$$

Using the formulation given in Equation 4.4.6, we can easily compute the spectral distribution function and the quantile function using numerical methods. In Figure 4.3, we have illustrated the Spherical-Gaussian covariance models for three different parameters of anisotropy.

4.4.2 Other choices of copulas

Apart from the bivariate normal copula, any other family of copulas that satisfies the conditions of Section 4.3.3 can be used to construct bivaraite covariance models. It is easy to follow from Equation 4.4.2 in Lemma 4.1 that, even if the integral does not have an explicit form, the value of the covariance at any lag $\mathbf{h}$ can be approximated with high precision by simulating $\{\mathbf{U}_1,\ldots,\mathbf{U}_N\}$ i.i.d. from Uniform$[0,1]^2$ and then computing

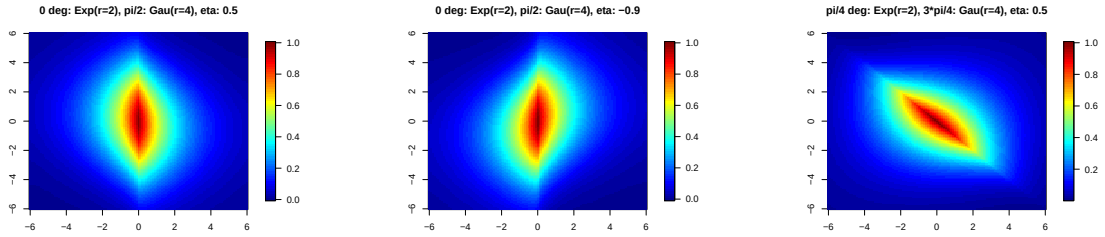


Figure 4.4: Illustration of the covariance functions obtained by combining univariate Exponential and Gaussian covariances using FGM copula.

$N^{-1}\sum_{j=1}^{n}\cos\left(\mathbf{h}'\mathbf{F}^{-1}\left(\mathbf{U}_j\right)\right)c\left(\mathbf{U}_j\right)$. Here, we mention a few families of copulas and discuss their viability to construct bivariate anisotropic covariance functions over the space.

4.4.2.1 Farlie-Gumbel-Morgenstern family

This family of copulas, often referred as FGM family is given by,

$$C\left((u,v)';\eta\right) = uv + \eta u(1-u)v(1-v), \tag{4.4.7}$$

where $\eta \in [-1,1]$ is the parameter of anisotropy in this case. The density of this copula family exists and is given by,

$$c\left((u,v)';\eta\right) = \eta\left(2u-1\right)\left(2v-1\right)+1.$$

In Figure 4.4, we have illustrated the bivariate covariances obtained by combining one Exponential and another Gaussian model using FGM copula. Though FGM family is computationally as well as mathematically convenient to use, as mentioned by Cooray (2018), it is restrictive as the Spearman's Rho (ρ_c) between the two marginals is limited to $[-1/3, 1/3]$.

Many extensions of this family is available in literature, for example, see Huang and

Kotz (1999), Bairamov and Kotz (2002) – where the authors have extended the family of FGM copulas to provide better correlation range between the marginals following the methodology proposed by Sarmanov (1966). Rodrıguez-Lallena and Úbeda-Flores (2004) introduced a new class of copulas motivated from the FGM family. In a recent paper, Cooray (2018) has proposed a computationally convenient and parsimonious extension to the FGM family to model negative dependence between the marginals.

In our context, the choice of copulas depends on the flexibility of the model as well as the convenience in computation of the bivariate covariance from two given directional covariance models. Komelj and Perman (2010) has proposed a construction of the bivariate characteristic function in terms of the univariate marginal characteristic functions using the Rodrıguez-Lallena and Úbeda-Flores (2004) family of copulas. Using the formulation provided by Komelj and Perman (2010), Equation 4.3.2 boils down to,

$$\varphi_{12}^{*}\left(\mathbf{h}\right)=\varphi_{1}\left(h_{1}\right)\varphi_{2}\left(h_{2}\right)+\eta b_{1}b_{2}\left(\varphi_{1}\left(h_{1}\right)-\tilde{\varphi}_{1}\left(h_{1}\right)\right)\left(\varphi_{2}\left(h_{2}\right)-\tilde{\varphi}_{2}\left(h_{2}\right)\right), \tag{4.4.8}$$

where, for $i\in\{1,2\}$, $\tilde{\varphi}_{i}\left(\cdot\right)$ is the 'distortion characteristic function', i.e. the characteristic function corresponding to the cumulative distribution function $\tilde{F}_{i}\left(\omega\right)=h_{i}\circ F_{i}\left(\omega\right)$ where, $h_{i}\left(x\right)=x-b_{i}^{-1}\int_{0}^{x}g_{i}\left(t\right)dt$. Here, $g_{i}\left(\cdot\right):[0,1]\rightarrow\mathbb{R}$ is such a continuous function that,

i. $g_i(t)$ is not identically equal to 0,

ii. $\int_0^1 g_i(t)dt=0$,

iii. $g_i(t)=g_i(1-t)$ for all $0\leq t\leq 1$.

Here, $b_{i}=\max\limits_{0\leq t\leq1}g_{i}\left(t\right)$. Let, $a_{i}=-\min\limits_{0\leq t\leq1}g_{i}\left(t\right)$ for $i\in\{1,2\}$. Define, $M^{a}=\max\left\{a_{1}a_{2},b_{1}b_{2}\right\}$ and $M^{b}=\max\left\{a_{1}b_{2},a_{2}b_{1}\right\}$. Then, it can be shown straightforwardly using Theorem 5

of Komelj and Perman (2010), that for all $\eta \in [-\frac{1}{M^a}, \frac{1}{M^b}]$, Equation 4.4.8 provides an admissible covariance function on $\mathbb{R}^2$.

Though mathematically this formulation proposed by Komelj and Perman (2010) allows us to write the join covariance in terms of the marginal covariances, and not in terms of their corresponding marginal spectral distribution functions, the computation of the distortion characteristic function requires computation of univariate integrals involving their corresponding spectral distributions. Moreover, as shown by Komelj and Perman (2010), like the other copulas from FGM family, this formulation is also restrictive as the resulting joint covariance can only allow the Spearman's Rho between the marginal random variables along the two specified directions to be within $(-3/4, 3/4)$, where as for the bivaraite normal copula, it can be shown that the Spearman's Rho can take any values in $(-1, 1)$ and also as stated in Section 4.4.1, it can be continuously extended continuously as $\rho_c \to -1$ or 1 (for details, see Nelsen 2007).

4.4.2.2 Archimedean family

Every member of Archimedean copula family is generated by a continuous strictly decreasing convex function $\phi(\cdot) : [0, 1] \to [0, \infty]$ such that $\phi(1) = 0$. Given a *phi*-function the corresponding copula is given by,

$$C\left((u, v)'\right) = \phi^{[-1]}\left(\phi(u) + \phi(v)\right), \tag{4.4.9}$$

where $\phi^{[-1]}$ is the pseudo-inverse of ϕ (for more details, see Chapter 4, Nelsen 2007). The function ϕ is referred as the generator of the corresponding copula $C((u, v)')$. There are several parametric family of generators that are available in literature, for example the Clayton family of copulas are generated by $\phi(u) = (u^{-\eta} - 1)/\eta$ for $\eta \in [-1, \infty)$; family of

copulas due to Ali-Mikhail-Haq is given by $\phi(u) = \log\left((1-\eta\,(1-u))\,/u\right)$ for $\eta \in [-1,1]$ etc. For more examples of Archimedean copulas and their properties see Nelsen (2007).

In the next section, we evaluate our proposed covariance model empirically through extensive Monte Carlo studies.

4.5 Simulation Study

In this section, we have used empirical methods to evaluate the WLS-based estimation of the covariance parameters for the copula-based bivariate covariance model. In addition, we also compare the flexibility of the proposed covariance with the widely-used geometrically anisotropic Matérn class of parametric models.

In the first stage of our simulation study, we have generated data ($B = 500$ replications) from a Gaussian Process (referred as GP from here) on a regular grid $\lambda_n[0,1]^2$. As the focus of this chapter is to capture the second-order structure of the process, without loss of any generality, we have taken the mean of the process, $\mu(\mathbf{s})$ to be identically equal to 0. The true covariance structure has been specified by the bivariate normal copula-based covariance model obtained by combining the marginal Exponential correlation with range parameter α_e along direction $\theta_1 = 0$ and the Gaussian correlation (or Squared Exponential) with range parameter α_g along direction $\theta_2 = \pi/2$. The parameter of anisotropy, i.e. the correlation parameter of the bivariate normal copula is denoted by ρ and the overall variance of the process, i.e. the sill parameter is denoted by σ^2. We have not included any micro-scale error or nugget component in these simulations.

In Table 4.1, we have reported the mean squared errors of the WLS-based estimators of the covariance parameters. We have taken λ_n to be equal to $8, 10, 13, 23$ for $n = 484, 961, 4900$ and 10000 respectively. Clearly, the spatial sampling design belongs

Table 4.1: Empirical mean squared errors (MSE) of the WLS-based covariance parameters estimators for the Exponential-Gaussian Bivariate Normal Copula-based model. The true values of the parameters are: $\alpha_e = 2$, $\alpha_g = 4$, $\rho = 0.5$ and $\sigma^2 = 3$.

n	L	**MSE**$(\hat{\alpha}_e)$	**MSE**$(\hat{\alpha}_g)$	**MSE**$(\hat{\rho})$	**MSE**$(\hat{\sigma}^2)$	**Mean Time (sec.)**
484	4	1.508	0.774	0.344	1.482	22.8 (7.4)
961	8	1.350	0.380	0.161	1.067	21.1 (3.1)
4900	12	0.538	0.051	0.057	0.521	29.9 (3.0)
10000	16	0.081	0.004	0.008	0.030	38.3 (5.7)

to the *mixed increasing domain* asymptotic framework – which is required for consistent estimation of the covariance of a second-order stationary mean squared continuous spatial process, as mentioned by Hall and Patil (1994). In Table 4.1, by L we denote the number of directions used in the WLS-based estimation of the parameters. This quantity also increases with n, as with increasing sample size, both the number of distinct orientations observed and its frequency increases. But, as compared to the rate of n, L can be assumed as a fixed value. Clearly, Table 4.1 demonstrates the consistency of the WLS-based estimators, i.e. the MSE of all the estimators tends to 0 as $n \to \infty$. More importantly, the mean time used to obtain the estimators increases at most linearly with n, and the reason is that the optimization in WLS-based estimation only depends on the generic semivariogram values for a fixed set of dicrete lags.

In the second stage of our simulations, we have compared the flexibility of our proposed modeling of the dependence structure with one of the most commonly used parametric class of anisotropic models: the geometrically anisotropic Matérn covariance. We used the same spatial sampling design to simulate observations from a zero mean GP. In the first part of this Monte Carlo study, we have generated the realizations from a random field with covariance structure that is generated by combining a Spherical covarince with range parameter $\alpha_s = 2$ along X-axis and a Gaussian covariance with range param-

eter $\alpha_g = 4$ along Y-axis using a bivaraite normal copula with parameter of anisotropy $\rho = -0.7$ (i.e. the covariance exhibitted in the middle plot of Figure 4.3). The marginal variance of the random field, i.e. the sill parameter σ^2, has been taken to be equal to 3. For each of the $B = 200$ simulations, we have estimated the covariances using two models: the Spherical-Gaussian bivaraite normal copula based model, and the geometrically anisotropic Matérn covariance. The first model is fitted through the WLS-based directional variogram model fitting as proposed in Section 4.3 and the second model is estimated using the maximum likelihood (referred as ML from here) estimation procedure using the `likfit` function in R package *geoR* (Ribeiro Jr and Diggle, 2018). To compare the above-said methodologies, for each of the simulation steps we have computed the average squared difference (referred as ASD from here) between the estimated covariance function and the true covariance over a uniform spatial grid $G \subset [-6, 6] \times [0, 6]$ with minimum distance between two points in the grid equal to 0.4 (i.e. G contains 496 observations), i.e. for some candidate covariance estimator $\hat{\varphi}$,

$$\text{ASD}\left(\hat{\varphi}\right) = \frac{1}{|G|} \sum_{\mathbf{h} \in G} \left(\varphi(\mathbf{h}) - \hat{\varphi}(\mathbf{h})\right)^2 .$$

We have reported the mean and standard deviation of the ASDs for both the above-mentioned models in Table 4.2. Moreover, Table 4.2 summarizes the time taken by the methodologies to fit the corresponding models to the simulated realizations. 0.6 Next,

Table 4.2: Performance under correct model: Spherical-Gaussian-Normal-Copula.

n	Mean.ASD(Sph-Gau-Cop)	Mean.ASD(Geo-Aniso-Mat)	Mean.Time(Sph-Gau-Cop)	Mean.Time(Geo-Aniso-Mat)
484	0.209 (0.335)	12.967 (25.869)	55.53 (20.61)	29.71 (6.85)
961	0.196 (0.280)	12.294 (29.220)	53.57 (18.12)	203.90 (44.82)
4900	0.045 (0.012)	3.197 (2.721)	51.10 (10.39)	22375.81 (5429.40)

we have generated the realizations of the same spatial locations as before from a zero mean GP with covariance specified by a geometrically anisotropic Exponential model with range parameter $\alpha_e = 1.5$, sill $\sigma^2 = 3$, the angle of anisotropy $\psi_A = 0$ (i.e. the angle between the Y-axis and the direction of maximum range), and the ratio of maximum and minimum ranges $\psi_R = 2$ (for more details see documentation of `coords.aniso` function in *geoR*, Ribeiro Jr and Diggle 2018). For this simulation study, we have fitted the two following models to the realizations: Exponential-Exponential bivariate normal copula based model, and the geometrically anisotropic Matérn model. To fit the copula-based model, we have considered L'-many (stated in Table 4.3) possible pairs of angles $\{\theta_1, \theta_2\}$ along which two Exponential covariances with different range parameters are combined to get the joint covariance structure over the space. The fitted model with lowest BIC is selected as the most 'optimal' choice. The results has been reported in the Table 4.3.

Table 4.3: Performance under corresct model: Geometrically Anisotropic Exponential.

n	L'	Mean.ASD(Exp-Exp-Cop)	Mean.ASD(Geo-Aniso-Mat)	Mean.Time(Exp-Exp-Cop)	Mean.Time(Geo-Aniso-Mat)
484	6	0.498 (0.190)	0.182 (0.161)	10.99 (0.33)	39.03 (6.31)
961	21	0.481 (0.178)	0.147 (0.128)	12.52 (0.38)	230.13 (38.01)
4900	36	0.397 (0.048)	0.100 (0.058)	35.91 (2.46)	21362.34 (4963.22)

Table 4.2 clearly demonstrates that the Spherical-Gaussian bivaraite normal copula-based model has better performance to capture the true underlying dependence structure as compared to the anisotropic Matérn, as expected. Though for large n ($n \approx 5000$), the ASD for the geometrically anisotropic Matérn reduces significantly, the average time taken for ML estimation of the geometrically anisotropic Matérn model is nearly 450 times what is needed for the WLS-based model fitting for the copula-based approach. On the

other hand, when the realizations are generated from a random field with true covariance given by a geometrically anisotropic Matérn, the ASD for the Exponential-Exponential bivariate copula-based model is higher than the anisotropic Matérn model but, it is not as 'bad' as the performance of anisotropic Matérn model in Table 4.2. Moreover, in terms of the computational time required for model fitting, the WLS estimation of the copula-based model clearly outperforms the ML estimation for the geometically anisotropic Matérn model. Hence, as suggested by our Monte Carlo study, both in terms of flexibility and scalability, the copula-based modeling of anisotropy is the better choice as compared to the geometrically anisotropic Matérn model, specially when the underlying covariance is misspecified.

4.6 Concluding Remarks

In this chapter, we have introduced a copula-based approach to model anisotropic covariance structure for a mean-squared continuous second-order stationary random field. By combining the spectral distributions of two directional covariances using a bivariate copula we can construct admissible anisotropic covariances on $\mathbb{R}^2$. This copula-based approach allows the practitioners to construct anisotropic covariances that take the directional variogram analysis into account. Next, a least squares based estimation of the model parameters has been proposed. The consistency and asymptotic normality of the estimators have been derived under general conditions on the nonparametric semivariogram estimator used in the least squares optimization. Finally, the proposed approach has been illustrated through several examples and evaluated through Monte Carlo study.

The research presented in this chapter provides a new platform which allows to combine two marginal covariances – even if they are from different class of parametric models

– and thus, goes beyond the classical geometrically anisotropic models or other range anisotropic models as proposed in Zimmerman (1993), Eriksson and Siska (2000), Ecker and Gelfand (1999) etc. Though the copula-based covariance modeling provides a rich class of anisotropic covariances, finding the explicit form of the covariance for any two given directional covariance function is not straightforward. We have proposed a Monte Carlo-based numerical method to approximate the covariance function at required lags. Finding a class of copula, for which the joint covariance can be written explicitly in terms of the marginal ones, requires further research and is going to be the primary focus of our future studies. Also, constructing a copula which corresponds to the general range anisotropic models as proposed in Allard et al. (2016) is another area of our future research interests. Apart from that, this work creates avenue for potential methodological development in the following two directions. Firstly, this copula-based method can be systematically extended for processes on sphere and thus, admissible anisotropic covariances can be constructed to model global processes on planet Earth. Secondly, implementation of this copula-based approach in modeling non-separable space-time covariance, specially in regards to the recent developments in Omidi and Mohammadzadeh (2016), would be of great interest for modeling large scale spatio-temporal processes.

4.7 Proofs

We start this section with the proof of Proposition 4.1.

Proof. First consider the 'if' part, i.e. we have to show that the covariance $\varphi_{12}(\mathbf{h};\boldsymbol{\eta})$ given in Equation 4.3.3 is an admissible covariance on $\mathbb{R}^2$ and the directional covariances along θ_i is given by $\varphi_i(\|\mathbf{h}\|)$ for $i \in \{1,2\}$. Admissibility i.e. positive definiteness of $\varphi_{12}(\cdot)$ follows directly from the fact that $\varphi^*_{12}(\cdot;\boldsymbol{\eta})$ is a valid covariance in $\mathbb{R}^2$ by application of

Bochner (2005)'s theorem. Also, substituting $\mathbf{h} = (r\cos\theta_1, r\sin\theta_1)'$ in Equation 4.3.3 we get $\varphi_{12}(\mathbf{h};\boldsymbol{\eta}) = \varphi_{12}^{*}\left((r,0)';\boldsymbol{\eta}\right) = \varphi_1(r)$ by Sklar (1959)'s theorem. Similarly, putting $\mathbf{h} = (r\cos\theta_2, r\sin\theta_2)'$ we can show that $\varphi_{12}(\mathbf{h};\boldsymbol{\eta}) = \varphi_{12}^{*}\left((0,r)';\boldsymbol{\eta}\right) = \varphi_2(r)$. Hence, the 'if' part is proved.

For the 'only if' part, observe that for $i \in \{1,2\}$, $\varphi_{12}\left((r\cos\theta_i, r\sin\theta_i)';\boldsymbol{\eta}\right) = \varphi_i(r)$ for all $r \in \mathbb{R}^+$. We will show that $\varphi_i(\cdot)$ is the characteristic function of the random variable $W_i = \omega_1\cos\theta_i + \omega_2\sin\theta_i$ where $\boldsymbol{\omega} = (\omega_1,\omega_2)'$ has the joint characteristic function $\varphi_{12}(\cdot)$. It is easy to follow that the above statement is trivially true if $\theta_i \in \{0,\pi/2\}$. If $\theta_i \notin \{0,\pi/2\}$,

$$\begin{aligned}
\varphi_{12}\left((r\cos\theta_i, r\sin\theta_i)'\right) &= \int_{\mathbb{R}^2} \exp\left(\mathrm{i}r\left(\omega_1\cos\theta_i + \omega_2\sin\theta_i\right)\right) f_{12}\left((\omega_1,\omega_2)'\right) d\omega_1 d\omega_2 \\
&= \int_{\mathbb{R}^2} \exp\left(\mathrm{i}rw_i\right) |\cos\theta_i| f_{12}\left(\left(\frac{w_i - v\sin\theta_i}{cos\theta_i}, v\right)'\right) dw_i dv \\
&= \int_{\mathbb{R}} \exp\left(\mathrm{i}rw_i\right) \underbrace{\int_{\mathbb{R}} |\cos\theta_i| f_{12}\left(\left(\frac{w_i - v\sin\theta_i}{cos\theta_i}, v\right)'\right) dv}_{f_{W_i}(w_i)}\, dw_i.
\end{aligned}$$

Clearly, φ_i is the characteristic function of W_i for $i \in \{1,2\}$. Again from Bochner (2005)'s theorem, we have the marginal distribution functions of W_i is given by F_i. To get the bivariate covariance function $\varphi_{12}(\mathbf{h})$ we need to find the characteristic function of $\boldsymbol{\omega}$. Now, consider the random variable $\mathbf{W} = (W_1, W_2)'$ and observe that $\boldsymbol{\omega} = \mathbf{A}(\theta_1,\theta_2)^{-1}\mathbf{W}$. Clearly, the joint distribution of $\boldsymbol{\omega}$ is uniquely specified by the joint distribution of $\mathbf{W}$ and by Sklar (1959)'s theorem, there exists a unique bivariate copula $C(\cdot)$ such that the distribution function of $\mathbf{W}$ is given by $C\left((F_1(w_1), F_2(w_2))'\right)$ and the corresponding characteristic function by Equation 4.3.2. The rest of the proof follows immediately from the properties of characteristic function for linear transformation of random variables.

Before stating the proofs of Section 4.3, we first prove Lemma 4.1 as that will be used in the concurrent proofs.

Proof. Recall,

$$\varphi_{\boldsymbol{\omega}}(\mathbf{h}) = \int_{\mathbb{R}^2} \exp(\mathrm{i}\mathbf{h}'\boldsymbol{\omega})\, dC(\mathbf{F}(\boldsymbol{\omega}); \boldsymbol{\eta}).$$

Under the assumption that the density of the copula exists, simple calculations involving elementary differential calculus reveals that,

$$dC(\mathbf{F}(\boldsymbol{\omega})) = f_1(\omega_1) f_2(\omega_2) c(\mathbf{F}(\boldsymbol{\omega}))\, d\boldsymbol{\omega}.$$

Substituting this result in the previous integral and then taking substitution $\mathbf{F}(\boldsymbol{\omega}) = \mathbf{x}$ will yield Equation 4.4.2. ■

The proof of Corollary 4.1 is straightforward from the substitution,

$$\mathbf{x} = \left(\Phi^{-1}(F_1(\omega_1)), \Phi^{-1}(F_2(\omega_2))\right)'.$$

Now, we provide the proofs of the results regarding the large sample properties of the WLS estimator of the parameter of anisotropy. We start with the proof of Theorem 4.1.

Proof. For simplicity of notations we denote

$$\mathbf{W}_n(\boldsymbol{\eta}) = \mathrm{diag}\left((w(\mathbf{h}_1; \hat{\boldsymbol{\alpha}}_{1,n}, \hat{\boldsymbol{\alpha}}_{2,n}, \boldsymbol{\eta}), \ldots, w(\mathbf{h}_m; \hat{\boldsymbol{\alpha}}_{1,n}, \hat{\boldsymbol{\alpha}}_{2,n}, \boldsymbol{\eta}))'\right).$$

Define,

$$Q_n(\boldsymbol{\eta}) = \mathbf{g}(\boldsymbol{\eta})' \mathbf{W}_n(\boldsymbol{\eta}) \mathbf{g}(\boldsymbol{\eta}), \tag{4.7.1}$$

where,

$$\mathbf{g}(\boldsymbol{\eta}) = (2\gamma(\mathbf{h}_1; \boldsymbol{\nu}_0) - 2\gamma(\mathbf{h}_1; \boldsymbol{\alpha}_{10}, \boldsymbol{\alpha}_{20}, \boldsymbol{\eta}), \ldots, 2\gamma(\mathbf{h}_m; \boldsymbol{\nu}_0) - 2\gamma(\mathbf{h}_m; \boldsymbol{\alpha}_{10}, \boldsymbol{\alpha}_{20}, \boldsymbol{\eta}))'.$$

It is easy to follow from conditions (C.1), (C.2) and (C.3) that, $Q_n(\boldsymbol{\eta}_0) = 0$ and if for any $\boldsymbol{\eta} \in \Xi \setminus \{\boldsymbol{\eta}_0\}$ such that $\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \geq \varepsilon$ for some $\varepsilon > 0$, then there exist a $\delta'_\varepsilon > 0$ such that $Q_n(\boldsymbol{\eta}) > \delta'_\varepsilon > 0$. Hence, $Q_n(\cdot)$ has a unique minimum at $\boldsymbol{\eta}_0$.

Now, consider $\Delta_n^{\#} = \sup_{\boldsymbol{\eta}} |Q_n^{\#}(\boldsymbol{\eta}) - Q_n(\boldsymbol{\eta})|$. Note that, under condition (C.2) and (C.3), to show $\Delta_n^{\#} \to 0$ almost surely, it is enough to show that for each $1 \leq i \leq m$, $\Delta_{i,n} = \sup_{\boldsymbol{\eta}} |g_{i,n}^{\#}(\boldsymbol{\eta}) - g_i(\boldsymbol{\eta})| \to 0$ almost surely, where $g_{i,n}^{\#}(\boldsymbol{\eta})$ and $g_i(\boldsymbol{\eta})$ are the i^{th} elements of $\mathbf{g}_n^{\#}(\boldsymbol{\eta})$ and $\mathbf{g}(\boldsymbol{\eta})$ respectively. Observe that,

$$\begin{aligned}
\Delta_{i,n} &= \sup_{\boldsymbol{\eta}} |g_{i,n}^{\#}(\boldsymbol{\eta}) - g_i(\boldsymbol{\eta})| \\
&\leq \underbrace{|2\gamma_n^{\#}(\mathbf{h}_i) - 2\gamma(\mathbf{h}_i; \boldsymbol{\nu}_0)|}_{\Delta_{i,n}^{(1)}} + 2 \underbrace{\sup_{\boldsymbol{\eta}} |\varphi(\mathbf{h}_i; \hat{\boldsymbol{\alpha}}_{1,n}, \hat{\boldsymbol{\alpha}}_{2,n}, \boldsymbol{\eta}) - \varphi(\mathbf{h}_i; \boldsymbol{\alpha}_{10}, \boldsymbol{\alpha}_{20}, \boldsymbol{\eta})|}_{\Delta_{i,n}^{(2)}}.
\end{aligned}$$

Clearly, by assumption 4.3.10, $\Delta_{i,n}^{(1)} \to 0$ almost surely. Now,

$$\begin{aligned}
\Delta_{i,n}^{(2)} &= \sup_{\boldsymbol{\eta}} |\varphi(\mathbf{h}_i; \hat{\boldsymbol{\alpha}}_{1,n}, \hat{\boldsymbol{\alpha}}_{2,n}, \boldsymbol{\eta}) - \varphi(\mathbf{h}_i; \boldsymbol{\alpha}_{10}, \boldsymbol{\alpha}_{20}, \boldsymbol{\eta})| \\
&\leq \sup_{\boldsymbol{\eta}} |\int_{\mathbb{I}^2} \exp\left(\mathrm{i}\mathbf{h}_i' \mathbf{F}^{-1}(\mathbf{u}; \hat{\boldsymbol{\alpha}}_n)\right) - \exp\left(\mathrm{i}\mathbf{h}_i' \mathbf{F}^{-1}(\mathbf{u}; \boldsymbol{\alpha}_0)\right) c(\mathbf{u}; \boldsymbol{\eta})\, d\mathbf{u}| \quad \text{[Lemma 4.1]} \\
&\leq \int_{\mathbb{I}^2} |\exp\left(\mathrm{i}\mathbf{h}_i' \mathbf{F}^{-1}(\mathbf{u}; \hat{\boldsymbol{\alpha}}_n)\right) - \exp\left(\mathrm{i}\mathbf{h}_i' \mathbf{F}^{-1}(\mathbf{u}; \boldsymbol{\alpha}_0)\right)| \sup_{\boldsymbol{\eta}} |c(\mathbf{u}; \boldsymbol{\eta})| d\mathbf{u} \\
&\to 0 \quad \text{a.s.},
\end{aligned}$$

as the integrand is bounded above by the integrable function $\sup_{\boldsymbol{\eta}} |c(\mathbf{u}; \boldsymbol{\eta})|$ and by conditions of Theorem 4.1, $\hat{\boldsymbol{\alpha}}_n = \left(\hat{\boldsymbol{\alpha}}_{1,n}', \hat{\boldsymbol{\alpha}}_{2,n}'\right)' \to \boldsymbol{\alpha}_0 = (\boldsymbol{\alpha}_{10}', \boldsymbol{\alpha}_{20}')'$ almost surely (see Theorem 3.1 in Lahiri et al. 2002) and $\mathbf{F}(\mathbf{u}; \boldsymbol{\alpha}) = (F_1(u_1; \boldsymbol{\alpha}_1), F_2(u_2; \boldsymbol{\alpha}_2))'$ is a continuous function in $\boldsymbol{\alpha}$. Hence, by Dominated Convergence Theorem, we have shown that $\Delta_n^{\#} = \sup_{\boldsymbol{\eta}} |Q_n^{\#}(\boldsymbol{\eta}) - Q_n(\boldsymbol{\eta})| \to 0$ almost surely as $n \to \infty$.

Fix any sample point for which the WLS estimator of the anisotropy parameter is not consistent, i.e. $\hat{\boldsymbol{\eta}}_n \nrightarrow \boldsymbol{\eta}_0$. Then there exists an $\varepsilon > 0$ and a subsequence $\{p_n\}$ such that for all $n \geq 1$, $\|\hat{\boldsymbol{\eta}}_{p_n} - \boldsymbol{\eta}\| \geq \varepsilon$. Then,

$$\begin{aligned} Q^{\#}_{p_n}\left(\hat{\boldsymbol{\eta}}_{p_n}\right) - Q^{\#}_{p_n}\left(\boldsymbol{\eta}_0\right) &\geq Q_{p_n}\left(\hat{\boldsymbol{\eta}}_{p_n}\right) - Q_{p_n}\left(\boldsymbol{\eta}_0\right) - 2\Delta^{\#}_n \\ &\geq \inf\left\{Q_n\left(\boldsymbol{\eta}\right) : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \geq \varepsilon\right\} - 2\Delta^{\#}_n \\ &> 0, \end{aligned}$$

for all $n \geq n_0$ for some natural number n_0 – which implies that $Q^{\#}_{p_n}$ has smaller value at $\boldsymbol{\eta}_0$ than $Q^{\#}_{p_n}\left(\hat{\boldsymbol{\eta}}_{p_n}\right)$ contradicting the fact that $\hat{\boldsymbol{\eta}}_{p_n} = \underset{\boldsymbol{\eta}}{\operatorname{argmin}}\, Q^{\#}_{p_n}\left(\boldsymbol{\eta}\right)$. Hence, by contradiction principle, $\hat{\boldsymbol{\eta}}_n \to \boldsymbol{\eta}_0$ almost surely.

If 4.3.10 holds in probability, then for any given sequence $\{n'\}$, there exists a further subsequence $\{n''\}$ which converges almost surely and for that we can use the similar argument as above to prove the consistency. Hence, it can be shown that in the generic nonparametric semivariogram converges in probability then the corresponding WLS estimator is weakly consistent. ■

Next we shift our attention to the proof of the asymptotic normality result given in Theorem 4.2.

Proof. Let us denote for $1 \leq i \leq m$ and $1 \leq a \leq q$, the partial derivative of $\gamma\left(\mathbf{h}_i; \hat{\boldsymbol{\alpha}}_{1,n}, \hat{\boldsymbol{\alpha}}_{2,n}, \boldsymbol{\eta}\right)$ with respect to $\eta^{(a)}$ evaluated at $\boldsymbol{\eta}^* \in \Xi$ as $\gamma^{(a)}_{i,n}\left(\boldsymbol{\eta}^*\right)$ and the corresponding matrix of partial derivatives of $\mathbf{g}^{\#}_n\left(\boldsymbol{\eta}\right)$ with respect to $\boldsymbol{\eta}$ evaluated at $\boldsymbol{\eta}^*$ as $\boldsymbol{\Gamma}_n\left(\boldsymbol{\eta}^*\right)$, i.e.

$$\boldsymbol{\Gamma}_n\left(\boldsymbol{\eta}^*\right) = \left(-2\boldsymbol{\gamma}^{(1)}_n\left(\boldsymbol{\eta}^*\right), \ldots, -2\boldsymbol{\gamma}^{(q)}_n\left(\boldsymbol{\eta}^*\right)\right),$$

where for $1 \leq a \leq q$, $\boldsymbol{\gamma}^{(a)}_n\left(\boldsymbol{\eta}^*\right) = \left(\gamma^{(a)}_{1,n}\left(\boldsymbol{\eta}^*\right), \ldots, \gamma^{(a)}_{m,n}\left(\boldsymbol{\eta}^*\right)\right)'$. For simplicity of notations, by $\mathbf{W}_n\left(\boldsymbol{\eta}\right)$ we refer $\mathbf{W}\left(\hat{\boldsymbol{\alpha}}_{1,n}, \hat{\boldsymbol{\alpha}}_{2,n}, \boldsymbol{\eta}\right)$ and the partial derivative of $\mathbf{W}_n\left(\boldsymbol{\eta}\right)$ with respect

to $\eta^{(a)}$ evaluated at $\boldsymbol{\eta}^*$ as $\mathbf{W}_n^{(a)}(\boldsymbol{\eta}^*)$.

As $\hat{\boldsymbol{\eta}}_n$ is the minimizer of $Q_n^{\#}(\boldsymbol{\eta})$, we have for $1 \leq a \leq q$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \eta^{(a)}} \left(\mathbf{g}_n^{\#}(\boldsymbol{\eta})' \mathbf{W}_n(\boldsymbol{\eta}) \mathbf{g}_n^{\#}(\boldsymbol{\eta}) \right) \Big|_{\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}_n} \\ &= \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n) + 2 \left(-2\boldsymbol{\gamma}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \right)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n). \end{aligned}$$

Expanding $\mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n)$ in the last term by one step Taylor's series expansion with respect to $\boldsymbol{\eta}_0$ we have,

$$\begin{aligned} 0 = \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n) + 2 \left(-2\boldsymbol{\gamma}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \right)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0) \\ + 2 \left(-2\boldsymbol{\gamma}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \right)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \boldsymbol{\Gamma}_n^*(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0), \end{aligned}$$

where the b^{th} column of $(\boldsymbol{\Gamma}_n^*)_{m\times q}$ is given by $(\boldsymbol{\Gamma}_n^*)_{\cdot,b} = \int_0^1 -2\boldsymbol{\gamma}_n^{(b)}(u\boldsymbol{\eta}_0 + (1-u)\hat{\boldsymbol{\eta}}_n)\, du$, where the integral is componentwise. Clearly, the q equations above can be written in the matrix format as,

$$\begin{aligned} \left(\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \boldsymbol{\Gamma}_n^* \right) (\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = -\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0) \\ - \frac{1}{2} \sum_{a=1}^{q} \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\hat{\boldsymbol{\eta}}_n) \mathbf{e}_a. \end{aligned}$$

Again expanding the last term by one step Taylor's series formula and re-arranging terms we have,

$$\begin{aligned} \left(\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \boldsymbol{\Gamma}_n^* + R_n^* \right) (\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = -\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\eta}}_n)' \mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0) \\ - \frac{1}{2} \sum_{a=1}^{q} \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0)' \mathbf{W}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0) \mathbf{e}_a \\ - \sum_{a=1}^{q} \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0)' \mathbf{W}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \boldsymbol{\Gamma}_n^*(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) \mathbf{e}_a. \end{aligned}$$

where the b^{th} row of $(R_n^*)_{m\times q}$ is given by $(R_n^*)_{b,\cdot} = \frac{1}{2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)' \boldsymbol{\Gamma}_n^{*\prime} \mathbf{W}_n^{(a)}(\hat{\boldsymbol{\eta}}_n) \boldsymbol{\Gamma}_n^*$. Now, by Theorem 4.1, $\hat{\boldsymbol{\eta}}_n \xrightarrow{p} \boldsymbol{\eta}_0$ and by Theorem 3.1 of Lahiri et al. (2002), $\hat{\boldsymbol{\alpha}}_{i,n} \to \boldsymbol{\alpha}_{i,0}$ for $i = 1, 2$. Thus, by continuity of the variogram function with respect to its parameters, from 4.3.11 and Slutsky's theorem we have,

$$a_n \mathbf{g}_n^{\#}(\boldsymbol{\eta}_0) \xrightarrow{\text{d}} \mathbf{N}(\mathbf{0}, \Sigma(\boldsymbol{\eta}_0)).$$

Also, the conditions (C.2)-(C.5) ensures the continuity of the partial derivatives $\boldsymbol{\gamma}^{(a)}(\boldsymbol{\nu})$ and $\mathbf{W}^{(a)}(\boldsymbol{\nu})$ for all $1 \le a \le q$. Hence, as $n \to \infty$, $\mathbf{W}_n(\hat{\boldsymbol{\eta}}_n) \to \mathbf{W}(\boldsymbol{\nu}_0)$, $\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\eta}}_n) \to \boldsymbol{\Gamma}(\boldsymbol{\eta}_0)$ in probability. Moreover, observe that by condition (C.6) the limit and the integration in the definition of $\boldsymbol{\Gamma}_n^*$ can be interchange (by Dominated Convergence Theorem) and hence again by continuity of the partial derivatives, we have $\boldsymbol{\Gamma}_n^* \xrightarrow{p} \boldsymbol{\Gamma}(\boldsymbol{\eta}_0)$. Hence, by repeated use of continuity of the functions with respect to the covariance parameters and Slutsky's theorem we have Theorem 4.2 from 4.3.11.

Chapter 5

Appendix

5.1 Supplementary Material for Chapter 2

5.1.1 Additional data description

In Chapter 2, we have provided necessary data descriptions. Here, we include additional details to support the understanding of the reader.

5.1.1.1 Data Preprocessing: details of aggregation

In the raw data, the time-stamped ambient temperature readings from the sensors in the devices were reported for the particular day as stated in the article. We have decided to aggregate the readings over the day to compute the location-wise average ambient temperature values, as most of the weather information organizations (e.g. NOAA, AccuWeather etc.) use the daily summaries for weather analysis. But as the data collection method depends on the user activity, the observations are irregular in temporal dimension. Hence to convert the real-time readings we took hourly averages of the temperature

readings for each the locations and then fitted a spline regression model to the hourly estimated temperatures and then computed the average daily ambient temperature using the mean of the fitted spline as shown in Figure 5.1 for two locations. Now there were

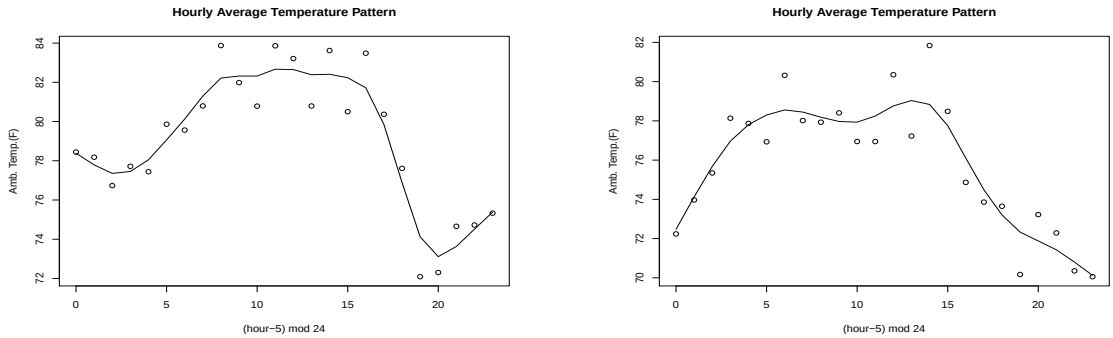


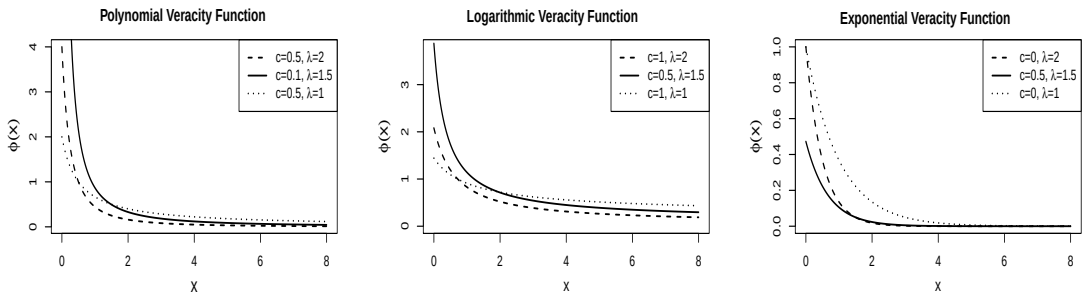
Figure 5.1: In this figure we plotted the hourly averages of the real-time temperature readings for two selected locations and showed how spline regression is fitted to estimate the hourly pattern for that particular day. The x-axis is shifted -5 hours to make the plot look reasonable as the temperature pattern is cyclical over the hours.

a few locations where there were no readings available corresponding to some hours of the day stopping us from computing the average temperature in those hours. For those locations we have estimated the missing hourly average temperatures using information from the 'local' neighborhood.

5.1.2 Additional details of VS-based methodology

5.1.2.1 Other variants of VS

To mention a few other possible variants of veracity functions we have plotted them for three different choices of ϕ in Figure 5.2. For polynomial veracity function $\phi(x) = (x+c)^{-\lambda}$ we need $\lambda > 0$ and $c > 0$ where as for the log-transform veracity function $\phi(x) = (\log(1+x+c))^{-\lambda}$ and exponential veracity function $\phi(x) = \exp(-\lambda(x+c))$,



(a) (b) (c)

Figure 5.2: Possible choices of the ϕ-function in the definition of VS.

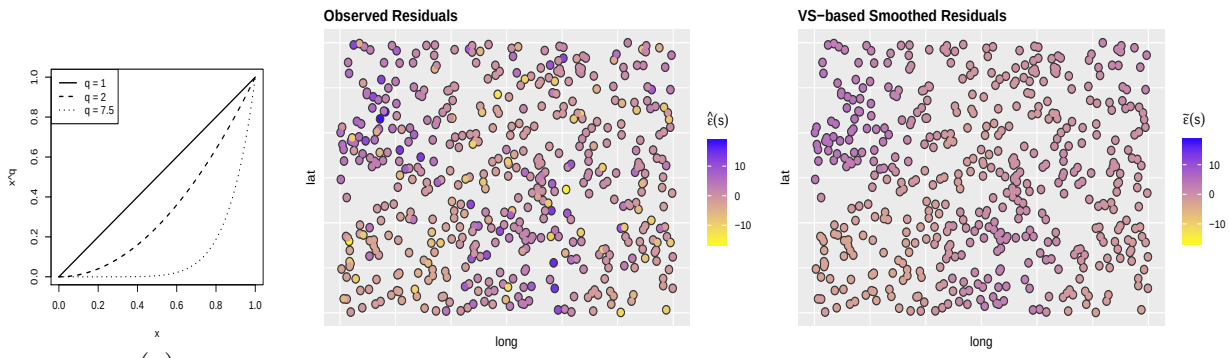
$c \geq 0$ and $\lambda > 0$ make the functions non-increasing and well-defined. Note that all of the veracity functions considered here have bounded range. These functions take maximum value at $x = 0$ and decrease to 0 as $x \to \infty$ and the rate of decay is determined by the parameter λ. The other parameter c determines the shift in the location of the veracity function.

5.1.2.2 VS-based smoothing

Here we have provided an illustration of the VS-based smoothing of the observed residuals as discussed in Section 3 in the article. In Figure 5.3 we used a synthetic data to show how the VS can be incorporated to reduce the effect of the absurd noise in the observed $\epsilon(\mathbf{s}_i)$'s.

5.1.3 Additional details of the simulation study

In the article we have provided the most important results of the simulations we have conducted. Here we provide additional details of the simulation design as well as results strengthening our claims in the article.



(a) (b) (c)

Figure 5.3: VS-based smoothing of the observed residuals. $q = 1.5$ for the smoothed version of the residuals in (c).

5.1.3.1 Parameters for simulation study

In the article we have reported the choices of parameters used for the simulations. Here we have reported the choices of the sampling design parameters as following:

- **Sampling Design Parameters:** these parameters are dependent on number of observations and the set of values we used are following,

  I. Without reference data simulations:

  i. $n = 100, \lambda_n = 3$ i.e. $\mathcal{R} = [0, 3]^2$,

  ii. $n = 500, \lambda_n = 10$

  ii. $n = 3000, \lambda_n = 25$.

  I. With reference data simulations:

  i. $m = 100$, $\Lambda_m = 20$ i.e. $\mathcal{D} = [0, 20]^2$ and $n = 50, 100, 500$ with $\lambda_n = 1.5, 2.5, 3$ respectively.

For all the simulation results reported in the main article we have used our baseline deviation $\alpha = 2$. The mixing parameter is taken to be same as the case-studies given by

$\nu(\mathbf{s}_i) = 1 - \exp\left(\frac{-1}{(1-\mathrm{R}^2)\sqrt{n(i)}}\right)$ where R^2 is the adjusted R-squared for the estimation of the mean surface using reference data only and $n(i)$ is the number of varying-quality observations in the δ-neighborhood. Finally, to choose the smoothing parameter q optimally we have conducted restricted leave-one-out cross-validation where only observations with VS greater or equal to 0.8 is allowed in the test set. So we choose an observation with VS greater or equal to 0.8 and then perform the estimation without that observation and predict at that observation location. Finally we choose our q such that it minimizes the L1O-MSPE which is defined as,

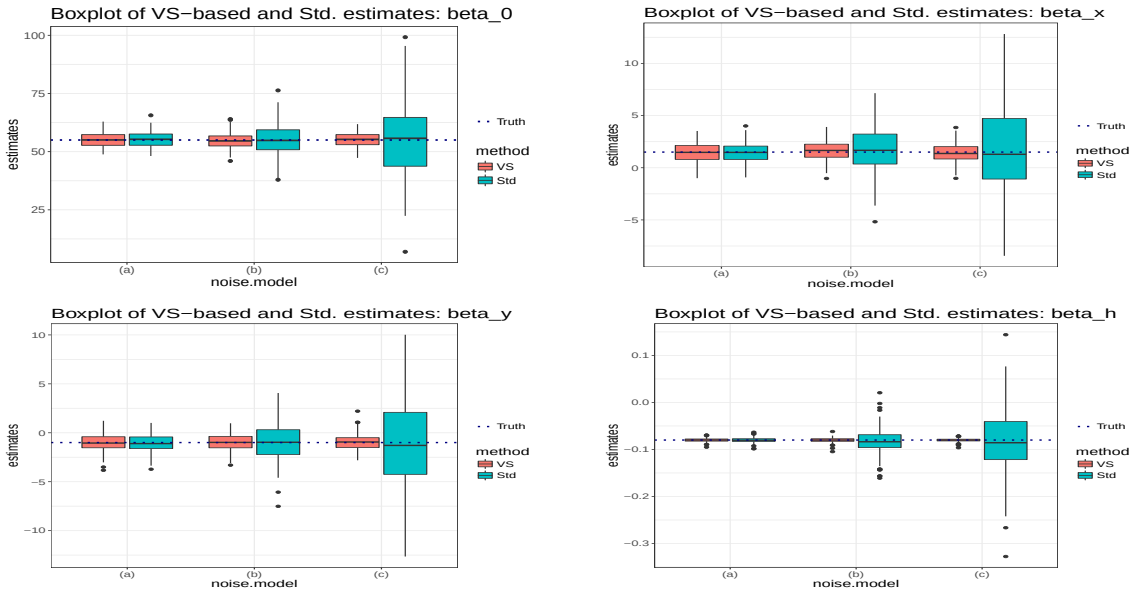
$$\text{L1O-MSPE} = \frac{1}{n^*} \sum_{i \in G_n^*} \left(Y(\mathbf{s}_i) - \hat{Y}_{-i}(\mathbf{s}_i) \right)^2,$$

where G_n^* is the set of indices corresponding to the observations with VS greater or equal to 0.8.

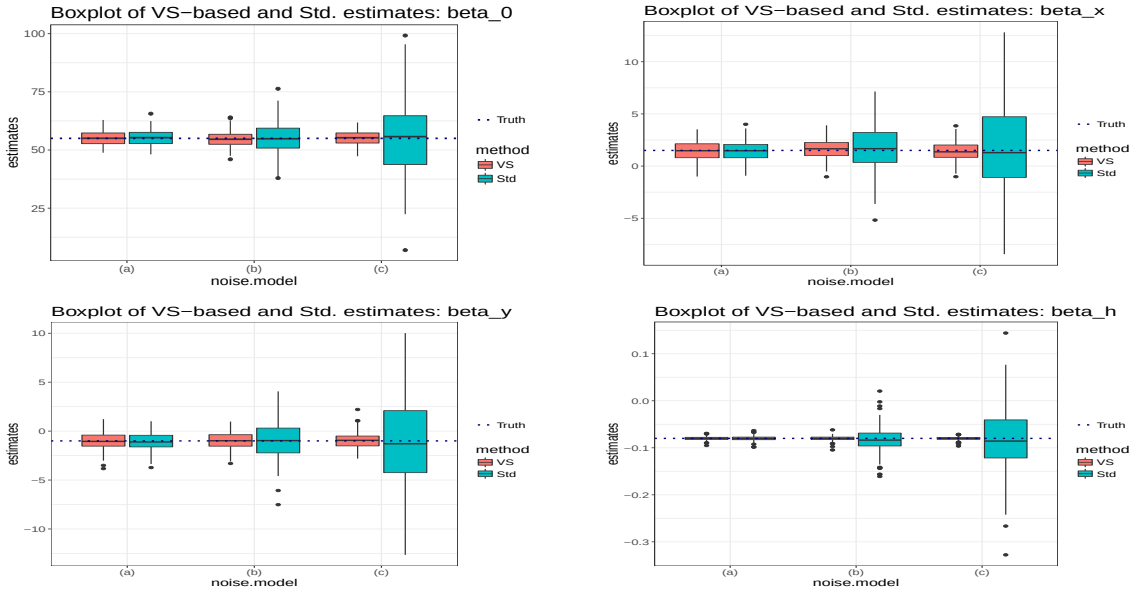
5.1.3.2 Additional results

In Figure 5.4c, we have plotted the boxplots of the VS-based and standard mean parameter estimates for $n = 100$ and 3000. Clearly in all the cases the VS-based estimation is much efficient than the standard o.l.s. estimation.

Next we provide the results for comparative analysis of VS-based and robust REML proposed by Künsch et al. (2011). To implement the later methodology we have used the R-package 'georob' created by Papritz (2018a). In Table 5.1 we have listed the results of a short simulation study to compare the georob (referred as GR henceforth) technique with that of the VS-based approach to estimate the covariance parameters from noisy spatial data. 'Sill.RMSE.VS' denotes the RMSE of the sill parameter computed from $B = 50$ simulated samples and similarly 'Sill.RMSE.GR' refers to the same for the



(a) n = 100



(b) n = 3000

(c) Performance of VS-based and standard approach regression parameter estimators for analyzing varying-quality observations without any reference data.

Table 5.1: Performance of the VS-based methodology and georob proposed by Künsch et al. (2011) and Papritz (2018a) in estimating covariance parameters from noisy spatial observations.

σ_A	n	Sill.RMSE.VS	Sill.RMSE.GR	range.RMSE.VS	range.RMSE.GR
5	500	2.09	3.105e+11	0.27	1.987e+11
	1000	1.00	4.436e+11	0.21	4.213e+11
50	500	2.19	3.507e+11	0.27	3.728e+11
	1000	1.23	2.322e+11	0.26	4.187e+11
100	500	2.59	1.562e+11	0.35	3.033e+11
	1000	1.96	4.032e+11	0.20	4.906e+11

georob technique. We only used 50 simulation iterations as georob uses REML method which requires the covariance matrix inversion in each iteration of the optimization and in addition to that the normal equations of the standard likelihood estimation is modified by a robust influence function and hence this method is highly expensive in terms of computation time. On an average georob requires approximately 145 seconds for one iteration for $n = 1000$ and $\sigma_A = 5$; where as given a q VS-based methodology requires only 2.58 seconds on average. As we can see in the Table 5.1, the robust Gaussian REML, proposed by Künsch et al. (2011) and Papritz (2018a), completely failed to capture the true parameters. One of the probable reason for the absurd results for the georob technique might be the numerical instability in the optimization when dealing with absurd noise in the observations.

REFERENCES

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing*. Dover Publication, New York.

AccuWeather (2015). Accuweather launches accucast, providing exclusive crowdsourced weather feature worldwide. `https://www.accuweather.com/en/press/50601069`. Accessed: 2019-01-30.

Allahbakhsh, M., Benatallah, B., Ignjatovic, A., Motahari-Nezhad, H. R., Bertino, E., and Dustdar, S. (2013). Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, 17(2):76–81.

Allard, D., Senoussi, R., and Porcu, E. (2016). Anisotropy models for spatial data. *Mathematical Geosciences*, 48(3):305–328.

Bairamov, I. and Kotz, S. (2002). Dependence structure and symmetry of huang-kotz fgm distributions and their extensions. *Metrika*, 56(1):55–72.

Bochner, S. (2005). *Harmonic analysis and the theory of probability*. Courier Corporation.

Chiles, J.-P. and Delfiner, P. (2009). *Geostatistics: modeling spatial uncertainty*, volume 497. John Wiley & Sons.

Conroy, N. J., Rubin, V. L., and Chen, Y. (2015). Veracity roadmap: Is big data objective, truthful and credible? *78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, (82).

Cooray, K. (2018). A new extension of the fgm copula for negative association. *Communications in Statistics-Theory and Methods*, pages 1–18.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley series in probability and mathematical statistics,. John Wiley & Sons, Inc.

Cressie, N. and Douglas, H. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, 12(2):115–125.

Dalton, A. (2016). Dark sky's hyperlocal weather app is now available on the web. `https://www.engadget.com/2016/09/20/dark-sky-hyperlocal-weather-app-desktop-web/`. Accessed: 2019-01-30.

Doukhan, P. (1994). *Mixing*. Springer, New York, NY.

Ecker, M. D. and Gelfand, A. E. (1999). Bayesian modeling and inference for geometrically anisotropic spatial data. *Mathematical Geology*, 31(1):67–83.

Ecker, M. D. and Gelfand, A. E. (2003). Spatial modeling and prediction under stationary non-geometric range anisotropy. *Environmental and Ecological Statistics*, 10(2):165–178.

Eriksson, M. and Siska, P. P. (2000). Understanding anisotropy computations. *Mathematical geology*, 32(6):683–700.

Evans, B. J. (1997). Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers & Geoscience*, 23(4):409–422.

Florio, E. N., Lele, S. R., Chang, Y. C., Sterner, R., and Glass, G. E. (2004). Integrating avhrr satellite data and noaa ground observations to predict surface air temperature: a statistical approach. *International Journal of Remote Sensing*, 25(15):2979–2994.

Fuentes, M. (2001). A high frequency kriging approach for nonstationary environmental processes. *Environmetrics*, 12(5):469–483.

Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P. (2010). *Handbook of Spatial Statistics*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press.

Genton, M. G. (1998). Highly robust variogram estimation. *Mathematical Geology*, 30(2):213–221.

Ghosh, J. K. (1971). A new proof of the bahadur representation of quantiles and an application. *Annals of Mathematical Statistics*, 42(6):1957–1961.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19(4):1327–1349.

Gomez, M. and Hazen, K. (1970). Evaluating sulfur and ash distribution in coal seams by statistical response surface regression analysis. Technical report, Bureau of Mines, Denver, Colo.(USA).

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory Related Fields*, 99(3):399–424.

Harris, P., Brunsdon, C., Charlton, M., Juggins, S., and Clarke, A. (2014). Multivariate spatial outlier detection using robust geographically weighted methods. *Mathematical Geosciences*, 46(1):1–31.

Haskard, K. A. (2007). *An anisotropic Matérn spatial covariance model: REML estimation and properties*. PhD thesis, University of Adelaide.

Huang, J. S. and Kotz, S. (1999). Modifications of the farlie-gumbel-morgenstern distributions. a tough hill to climb. *Metrika*, 49(2):135–145.

Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*. Wiley sereis in probability and statistics,. John Wiley & Sons, Inc.

Koller, M. and Stahel, W. A. (2011). Sharpening wald-type inference in robust regression for small samples. *Computational Statistics & Data Analysis*, 55(8):2504–2515.

Komelj, J. and Perman, M. (2010). Joint characteristic functions construction via copulas. *Insurance: Mathematics and Economics*, 47(2):137–143.

Künsch, H. R., Papritz, A., Schwierz, C., and Stahel, A. W. (2011). Robust estimation of the external drift and the variogram of spatial data. *ISI 58th World Statistics Congress of the International Statistical Institute*, (Dublin, Ireland):Aug 21–26.

Lahiri, S. N. (2003a). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, 65(2):356 – 388.

Lahiri, S. N. (2003b). *Resampling methods for dependent data*. Springer, New York.

Lahiri, S. N., Kaiser, M. S., Cressie, N., and Hsu, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association*, 94(445):86–97.

Lahiri, S. N., Lee, Y., and Cressie, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference*, 102(1-2):65–85.

Lark, R. M. (2000). A comparison of some robust estimators of the variogram for use in soil survey. *European journal of soil science*, 51(1):137–157.

Lukoianovaand, T. and Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? *Advances In Classification Research Online*.

Matheron, G. (1962). *Traité de géostatistique appliquée, Tome I*. Number 14. Paris: Technip.

Meyer, C. (2013). The bivariate normal copula. *Communications in Statistics-Theory and Methods*, 42(13):2402–2422.

Moynihan, T. (2015). Clever app turns everyone into a roving weather reporter. https://www.wired.com/2015/10/clever-app-turns-everyone-roving-weather-reporter/. Accessed: 2019-01-30.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Omidi, M. and Mohammadzadeh, M. (2016). A new method to build spatio-temporal covariance functions: analysis of ozone data. *Statistical Papers*, 57(3):689–703.

Organick, E. I. (1966). A fortran iv primer.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5):483–506.

Papritz, A. (2018a). *georob: Robust geostatistical analysis of spatial data,*. R package version 0.3-7.

Papritz, A. (2018b). Tutorial and manual for geostatistical analyses with the r package georob. https://cran.r-project.org/web/packages/georob/vignettes/georob_vignette.pdf. Accessed: 2019-02-12.

Pebesma, E. J. (2004). Multivariable geostatistics in s: the gstat package. *Computers & Geosciences*, 30(7):683–691.

Rendon, H., Wilson, A., and Stegall, J. (2018). Is it fake news? intelligence community expertise and news dissemination as measurements for media reliability. *Intelligence and National Security*, 33(7).

Ribeiro Jr, P. J. and Diggle, P. J. (2018). *geoR: Analysis of Geostatistical Data*. R package version 1.7-5.2.1.

Rodrıguez-Lallena, J. A. and Úbeda-Flores, M. (2004). A new class of bivariate copulas. *Statistics & probability letters*, 66(3):315–325.

Saha, B. and Srivastava, D. (2014). Data quality: The other face of big data. *2014 IEEE 30th International Conference on Data Engineering*.

Sarmanov, O. V. (1966). Generalized normal correlation and two-dimensional frechet classes. In *Doklady Akademii Nauk*, volume 168, pages 32–35. Russian Academy of Sciences.

Sen, P. K. (1968). Asymptotic normality of sample quantiles of m-dependent processes. *Annals of Mathematical Statistics*, 39(5):1724–1730.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231.

Sosko, S. and Dalyot, S. (2017). Crowdsourcing user-generated mobile sensor weather data for densifying static geosensor networks. *ISPRS International Journal of Geo-Information*, 6(3):61.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

Sun, S. and Lahiri, S. N. (2006). Bootstrapping the sample quantile of a weakly dependent sequence. *Sankhya: The Indian Journal of Statistics (2003-2007)*, 68(1):130–166.

Thornton, P. E., Running, S. W., and White, M. A. (1997). Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of hydrology*, 1905(3-4):214–251.

Todorov, V. and Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, 32(3):1–47.

Vancutsem, C., Ceccato, P., Dinku, T., and J.Connor, S. (2010). Evaluation of modis land surface temperature data to estimate air temperature in different ecosystems over africa. *Remote Sensing of Environment*, 114(2):449–465.

Watson, G. S. (1972). Trend surface analysis and spatial correlation. *Geological Society of America, Special Paper*, 146:39–46.

Willet, J. B. and Singer, J. D. (1988). Another cautionary note about r^2: Its use in weighted least-square regression analysis. *The American Statistician*, 42(3):236–238.

Zimmerman, D. L. (1993). Another look at anisotropy in geostatistics. *Mathematical Geology*, 25(4):453–470.