

## ABSTRACT

Title of dissertation:      **OUTLIER MODELING FOR  
SPATIAL GAUSSIAN RANDOM FIELDS**

Ekaterina Litvinova-Sotiris  
Doctor of Philosophy, 2013

Dissertation directed by:   **Dr Eric Slud, Department of Mathematics  
Dr William Bell, U.S. Census Bureau**

In this dissertation, we worked on extending time series outlier detection methodology to spatial data. An integral part of the outlier detection algorithm is a hypothesis test for the presence of at least one outlier in the data. The distribution of the corresponding test statistic is not known and as a result the critical value corresponding to a size  $\alpha$  test is estimated by approximating the tail probability for the test statistic. We identified and studied two methods of approximating the tail probability for the test statistic in the case when the parameters of the underlying spatial process are known. These approximations are based on bounds on the tail probability of the maxima of a discretely sampled Gaussian random field. We also study the distribution of the test statistic in the case when the parameters of the underlying spatial process are unknown and are estimated using maximum likelihood.

OUTLIER MODELING FOR SPATIAL GAUSSIAN RANDOM  
FIELDS

by

Ekaterina Litvinova-Sotiris

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2013

Advisory Committee:

Dr. Eric Slud, Chair/Advisor

Dr. William Bell, Co-Chair/Co-Advisor

Dr. Rama Chellappa

Dr. Abram Kagan

Dr. Paul Smith

© Copyright by  
Ekaterina Litvinova-Sotiris  
2013

# Dedication

To Sasha

## Acknowledgments

I am very happy to be writing this part of the thesis, as I feel deep gratitude to many people who have gotten me this far. This is a tribute to all my teachers, my well-wishers, my supporters, my mentors.

First of all, I would like to express my deep love and appreciation to my first teachers: my parents. You gave me life and guided me through the early years. I understand all of the challenges of being a parent so much more now that I am a parent myself! Thank you! I would also like to thank my parents-in-law, for all of their love and support! And, also, I would like to offer my appreciation to my extended family and to my friends.

Secondly, Dr. Bell and Dr. Slud, I feel so lucky for having had you as my mentors. I feel that I have learned so much from both of you. You challenged me, yet showed patience and care. You inspired me to grow and to aim high. You sacrificed your time and energy to see me prosper. I will remain forever grateful!

Thirdly, I want to thank all of the professors at the University who have imparted their wisdom and have been a source of inspiration. To my teachers: Dr. Kagan, Dr. Kedem, Dr. Koralov and Dr. Slud, thank you. And although I have never had you as a teacher, Dr. Smith, you have been a mentor, thank you. Also, I would like to thank Dr. Trivisa for all of her encouragement and support. Lastly, I would like to extend my gratitude to Dr. Balan for his help with the thesis. And, of course, I would like to thank the members of my dissertation committee.

The Mathematics Department staff is truly wonderful. Thanks you so much

for all your help! I would like to extend a special thank you to Celeste Regalado, she always goes out of her way to help the students!

Part of my thesis work was done while I was employed by the U.S. Census Bureau. I am thankful for this opportunity. I would also like to thank Dr. Tucker McElroy of the Census Bureau for his contribution to the thesis.

Last but definitely not least, I want to express my deep love and appreciation to my husband Vasilis. You have been my rock, my confidant, my loving partner. I love you.

## Table of Contents

List of Figures	vii
List of Abbreviations	viii
1 Indicator Variable Method of Outlier Detection	1
1.1 Introduction . . . . .	1
1.2 Indicator Variable Method . . . . .	2
1.2.1 Additive Outliers (AO) . . . . .	3
1.2.1.1 Outlier Detection Algorithm . . . . .	5
1.2.2 Other Types of Outliers . . . . .	7
1.2.3 Contributions of the dissertation . . . . .	9
2 Spatial Setting Considered	13
2.1 Index set $S$ and the sampling scheme . . . . .	14
2.2 Model for $W(\cdot)$ . . . . .	15
2.2.1 Canonical model for kriging . . . . .	16
2.2.1.1 Matérn Model . . . . .	17
2.2.2 ICAR(1) Model . . . . .	19
3 Critical Value	22
3.1 $\mu$ and $\theta$ known, location $s_0$ of a potential outlier known . . . . .	23
3.2 $\sigma^2$ unknown, other parameters known, location $s_0$ of a potential outlier known . . . . .	25
3.3 $\mu$ and $\theta$ known, location $s_0$ of potential outlier unknown . . . . .	26
3.3.1 Approximating $P$ by a function of $P^*$ . . . . .	28
3.3.2 Distribution of $\Lambda^N(\cdot)$ under $H_0$ . . . . .	31
3.3.2.1 Properties of $\Lambda^N(\cdot)$ when $Y(\cdot)$ is from the Matérn model . . . . .	43
3.3.2.2 Properties of $\Lambda^N(\cdot)$ when $Y(\cdot)$ is from the ICAR(1) model . . . . .	48
3.3.3 Bounds on the tail probability of the maximum of a discretely sampled Gaussian random field . . . . .	58

3.3.3.1	Bonferroni bound ( $P_{BON}$ ) . . . . .	59
3.3.3.2	Discrete Local Maxima (DLM) bound ( $P_{DLM}$ ) . . . . .	60
3.3.3.3	Comparison Between $P_{BON}$ and $P_{DLM}$ . . . . .	61
3.3.4	Recommended Critical Values . . . . .	63
3.3.4.1	When $Y(\cdot)$ is from the Matérn model . . . . .	66
3.3.4.2	When $Y(\cdot)$ is from the ICAR(1) model . . . . .	68
3.3.4.3	Bonferroni approximation to $t_\alpha$ , varying $n$ . . . . .	71
3.4	$\mu$ and $\theta$ unknown, location of outlier $s_0$ unknown . . . . .	71
3.4.1	Asymptotic properties of parameter estimates . . . . .	74
3.4.1.1	Increasing-domain vs. Infill asymptotics . . . . .	74
3.4.2	Simulated results . . . . .	79
4	Power . . . . .	95
4.1	How does power vary as we vary the degree of correlation in $Y(\cdot)$ ? . . . . .	97
4.1.1	Single outlier in the middle of the grid . . . . .	99
4.1.2	Single outlier at the edge of the grid . . . . .	100
4.2	Bonferroni and DLM bounds on power . . . . .	102
4.2.1	Bonferroni bound . . . . .	103
4.2.2	DLM bound . . . . .	107
4.2.3	Conclusion . . . . .	109
A	Background Information . . . . .	113
A.1	Measure Theory . . . . .	113
A.2	Properties of Random Variables and Processes . . . . .	113
A.2.1	Martingales . . . . .	114
A.3	Spatial Random Field . . . . .	116
A.3.1	Kriging . . . . .	117
A.3.2	Gaussian Conditional Autoregressions . . . . .	118
A.3.3	Mean Square Continuity and Integrability . . . . .	121
A.3.4	Spectral Domain . . . . .	122
A.3.4.1	Fourier Transform . . . . .	123
A.3.4.2	Spectral Representation of a Continuous Spatial Process . . . . .	124
A.3.4.3	Bochner's Theorem . . . . .	124
A.3.4.4	Spectral Representation of Isotropic Covariance Functions . . . . .	125
A.3.5	Equivalence of probability measures . . . . .	126
A.3.5.1	Equivalent measures and kriging, fixed domain asymptotics . . . . .	127
A.4	Hypothesis Testing . . . . .	129
A.5	Parameter Estimation . . . . .	131
A.5.1	Estimation of $\sigma^2$ . . . . .	133
	Bibliography . . . . .	136



## List of Figures

2.1	Matérn Autocorrelation Function . . . . .	20
3.1	Matérn Autocorrelation Function, varying $\nu$ and $\phi$ . . . . .	46
3.2	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ on $\mathbb{Z}_N^2$ , varying $\nu$ and $\phi$ , $N=3$ . . . . .	47
3.3	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ on $\mathbb{Z}_N^2$ , varying $\nu$ and $\phi$ , $N=11$ . . . . .	48
3.4	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ on $\mathbb{Z}_N^2$ , varying $\nu$ and $\phi$ , $N=25$ . . . . .	49
3.5	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ ; $\nu = 1, \phi = 2$ ; on $\mathbb{Z}_N^2$ (top), $\mathbb{Z}_{N-1}^2$ (bottom), $N = 3$ , $h$ is the Euclidean distance between observations. . . . .	50
3.6	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ ; $\nu = 1, \phi = 2$ ; on $\mathbb{Z}_N^2$ (top), $\mathbb{Z}_{N-1}^2$ (bottom), $N = 4$ , $h$ is the Euclidean distance between observations. . . . .	51
3.7	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ ; $\nu = 1, \phi = 2$ ; on $\mathbb{Z}_N^2$ (top), $\mathbb{Z}_{N-1}^2$ (bottom), $N = 5$ , $h$ is the Euclidean distance between observations. . . . .	52
3.8	Maximum, minimum and median correlations between $\Lambda(s)$ and $\Lambda(s+h)$ ; $\nu = 1, \phi = 2$ ; on $\mathbb{Z}_N^2$ (top), $\mathbb{Z}_{N-1}^2$ (bottom), $N = 11$ , $h$ is the Euclidean distance between observations. . . . .	53
3.9	Maximum absolute differences over $s$ of $\text{cov}(\Lambda(s), \Lambda(s+h))$ , for a fixed $h$ . The max is taken over $\mathbb{Z}_N^2$ , varying $N$ . . . . .	54
3.10	Maximum absolute differences over $s$ of $\text{cov}(\Lambda(s), \Lambda(s+h))$ , for a fixed $h$ . The max is taken over $\mathbb{Z}_{N-1}^2$ , varying $N$ . . . . .	55
3.11	$L(h)$ as a function of lag $h$ . The figures reported on the plots are the values of $L(h)$ corresponding to the highest correlated $Y(\cdot)$ , with $\nu = 1$ and $\phi = 2$ . . . . .	56
3.12	ICAR(1) Autocorrelation Function, $K_\theta(h)$ , as in (2.12), $h = \sqrt{h_x^2 + h_y^2}$ , varying $\rho$ . . . . .	72
3.13	Matern autocorrelation function with $\nu = 0.5$ , corresponds to the exponential autocorrelation function; varying $\phi$ . . . . .	80

4.1	One outlier of magnitude 3 is added to a Matérn process $W(\cdot)$ . Top: $Y(\cdot)$ rough, $\nu = 0.5$ , $\phi = 0.1$ . Bottom: $Y(\cdot)$ smooth, $\nu = 2$ , $\phi = 1$ . $N = 5$ . . . . .	98
4.2	$E(\Lambda^N(s))$ under $H_1$ , single outlier of magnitude $\beta = 3$ at the center of $\mathbb{Z}_2^2$ grid. Top: smooth $Y(\cdot)$ with $\nu = 2$ and $\phi = 1$ , bottom: rough $Y(\cdot)$ with $\nu = 0.5$ and $\phi = 0.1$ . . . . .	106
4.3	$E(\Lambda^N(s))$ and $E(\check{\Lambda}^N(s))$ under $H_1$ , single outlier of magnitude $\beta = 3$ at the center of $\mathbb{Z}_2^2$ grid. Smooth $Y(\cdot)$ , with $\nu = 2$ and $\phi = 1$ . . . . .	112

## Chapter 1: Indicator Variable Method of Outlier Detection

### 1.1 Introduction

Data values which are extreme with respect to the underlying distribution are referred to as *distributional outliers* [1]. The focus of this dissertation is on detection of such outliers in the setting where the underlying random field is a stationary Gaussian random field, defined on  $S \subset \mathbb{R}^d$ .

The purpose of outlier detection is two-fold. First, if we are interested in analyzing the data underlying the assumed, outlier-free distribution, it becomes important to treat observations which are outlying with respect to the assumed distribution. Kriging predictors are affected by the presence of outliers, as can be seen in the inflation of mean squared prediction error. Descriptive statistics, such as mean and standard deviation, as well as parameter estimates are also affected. Second, identification of distributional outliers can lead to the discovery of unusual observations and has a number of practical implications in areas such as credit card fraud, athlete performance, voting irregularity, soil and air pollutant detection, abnormality detection in medical imaging, and severe weather predictions.

In this dissertation, we study the *Indicator Variable* method of outlier detection. This method has been used extensively with time series data [2–10]. This

algorithm is used for outlier detection in the X-12 and TRAMO-SEATS programs, which have been widely used by statistical offices. Our contribution lies in carrying over the methodology from time series to higher dimensional data. Our interest lies in data indexed by two-dimensional spatial location  $s = (x, y) \in \mathbb{R}^2$ . However, where possible, we keep the discussion general on  $S \in \mathbb{R}^d$ .

## 1.2 Indicator Variable Method

Let  $s \in \mathbb{R}^d$  be a data location in a  $d$ -dimensional Euclidean space and suppose  $W(s)$  is a random quantity. Let  $s$  vary over an index set  $S \subset \mathbb{R}^d$ , assumed fixed, so as to generate a random field

$$\{W(s) : s \in S\} \tag{1.1}$$

We restrict our study to the case where  $W(\cdot)$  is stationary Gaussian random field, with

$$E(W(s)) = \mu, \forall s \in S \tag{1.2}$$

and

$$\text{cov}(W(s), W(t)) = C_\theta(s - t), \forall s, t \in S \tag{1.3}$$

where  $C_\theta(\cdot)$  is the *autocovariance function*. See Chapter 2 for our reasoning for choosing a constant mean model. We denote as  $\tau = (\mu, \theta)$  the parameters of  $W(\cdot)$ .

### 1.2.1 Additive Outliers (AO)

Suppose  $W(\cdot)$  is the "true", uncontaminated process and assume for now that  $\tau$  is known. Suppose we observe

$$Y_n = \{Y(s) : s \in S_n \subset S\} \quad (1.4)$$

where  $Y(\cdot)$  is contaminated by an outlier of an unknown magnitude  $\beta$  at a known location  $s_0 \in S_n$ . Thus,

$$Y(s) = \begin{cases} W(s) + \beta & s = s_0 \\ W(s) & s \neq s_0. \end{cases} \quad (1.5)$$

This type of an outlier is known in the time series literature as an *Additive Outlier* (AO). Then we have

$$Y_n \sim N_n(\mu + X_{s_0}\beta_{s_0}, \Sigma_\theta) \quad (1.6)$$

where  $\Sigma_\theta$  is a positive definite matrix defined via  $C_\theta$  and the *Indicator variable*  $X_s$  is defined as

$$X_s(u) = \begin{cases} 1 & u = s \\ 0 & u \neq s. \end{cases} \quad (1.7)$$

Because we estimate the effect of an outlier by regressing on an Indicator variable  $X_s$ , we call this method of outlier detection the *Indicator variable* method. The effect of the outlier can be estimated using one of the standard parameter estimation techniques, such as *Generalized Least Squares* or *Maximum Likelihood* (see Section A.5). The hypotheses for a formal test for the presence of an outlier at

location  $s_0$  are

$$\begin{aligned} H_0 &: \beta(s_0) = 0 \\ H_a &: \beta(s_0) \neq 0. \end{aligned} \tag{1.8}$$

Let  $\lambda^n(s_0)$  be the likelihood ratio test statistic for this hypothesis test. The superscript  $n$  indicates that we are testing for an outlier at location  $s_0$ , with respect to  $Y(\cdot)$  on  $S_n$ . We give a review of hypothesis testing in Section A.4 and give more details about the test statistic  $\lambda^n(s_0)$  in Chapter 3.

Now suppose we speculate that an outlier is present in  $Y_n$ , but we are not sure of its location. The corresponding hypothesis test for testing for an outlier at an *unknown* location is

$$\begin{aligned} H_0 &: \bigcap_{u \in S_n} H_{a,u}^c \\ H_a &: \bigcup_{u \in S_n} H_{a,u} \end{aligned} \tag{1.9}$$

where  $H_{a,u} : \beta(u) \neq 0$  is the alternative hypothesis for testing for the presence of a single outlier at location  $s_0 = u$ , as in (1.8). Because  $H_a$  is the union of  $H_{a,u}$  over  $u$ , we can use the union-intersection method of test construction to construct a test statistic for (1.9) (see Section A.4). The test statistic is

$$\max_{u \in S_n} |\lambda^n(u)| \tag{1.10}$$

where  $\lambda^n(u)$  is the test statistic for testing the presence of an outlier at location  $u$ .

In practice, the number and locations of AOs in the data are usually unknown. In addition, the parameters  $\tau$  of the process  $Y(\cdot)$  are unknown and need to be estimated. Below, we lay out the *Indicator variable* outlier detection algorithm in this general setting.

### 1.2.1.1 Outlier Detection Algorithm

Suppose  $Y(\cdot) = W(\cdot) + O(\cdot)$ , where  $W(\cdot)$  is defined on  $S \in \mathbb{R}^d$  as in (1.1)-(1.3) and  $O(\cdot)$  is the outlier process. We assume that  $O(\cdot)$  is deterministic, in the sense that it does not exhibit variation, however its magnitude is unknown. Let  $S_O \subset S$  be the locations of the outliers, unknown a priori. Then,

$$O(s) = \begin{cases} \beta(s) & s \in S_O \\ 0 & s \notin S_O \end{cases} \quad (1.11)$$

where  $\beta(s)$  is the magnitude of the outlier at  $s \in S_O$ . Modeled this way,  $Y(\cdot)$  is a stationary Gaussian random field, with

$$E(Y(s)) = \mu + \sum_{t \in S_O} \beta(t) X_t(s), \quad \forall s \in S \quad (1.12)$$

and

$$\text{cov}(Y(s), Y(t)) = C_\theta(s - t), \quad \forall s, t \in S \quad (1.13)$$

where  $C_\theta(\cdot)$  is the autocovariance function of  $W(\cdot)$ , as in (1.3). Consider a sample  $Y_n \equiv \{Y(s) : s \in S_n \subset S\}$ . Then,

$$Y_n \sim N_n(X_{S_O} \beta_{S_O}, \Sigma_\theta) \quad (1.14)$$

where  $X_{S_O}$  is the matrix of regressors, whose first column is all 1's, corresponding to the mean  $\mu$  of  $W(\cdot)$  and the other columns are Indicator variables  $X_s$ , as in (1.7), corresponding to each  $s \in S_n \cap S_O$ . In reality, we do not know the locations of the outliers in the sample a priori. Therefore, we do not know a priori the form that the regressor matrix  $X$  takes and we need to infer this from the data. Below, we

detail an algorithm which searches out the locations of the outliers, constructs the regressor matrix  $X$  and estimates the magnitudes  $\beta$  of the outliers.

### Outlier Detection Algorithm

We now assume that  $\tau = (\mu, \theta)$  is unknown. The algorithm is borrowed from time series literature, see [2, 10]. The algorithm starts with  $X = \mathbf{1}$ , the vector corresponding to the constant mean parameter  $\mu$ , which parametrizes the outlier-free process  $W(\cdot)$ . Parameter estimation is done via maximum likelihood, unless otherwise noted.<sup>1</sup>

1. Estimate the current model.
2. Calculate a robust estimate of  $\sigma^2$ , as in (A.42) or (A.45). See Section A.5.1 for more details.
3. Let  $\lambda^n(s)$  be defined as in (3.4) for each  $s \in S_n$  and let  $\hat{\lambda}^n(s)$  be the estimate of  $\lambda^n(s)$  with the MLE  $\hat{\tau}$  in place of  $\tau$ . Compute the estimated test statistics  $\hat{\lambda}^n(s)$  for all  $\{s : X_s \notin \text{col}(X)\}$ , where  $\text{col}(X)$  is the column space of  $X$ .
4. Let

$$\hat{\lambda}_{max}^* = \max_{u: X_u \notin \text{col}(X)} |\hat{\lambda}^n(u)|$$

---

<sup>1</sup>This is in contrast to the "classical geostatistics" method of parameter estimation, which combines method-of-moments and least squares. This method does not require any distributional assumptions, but it is not optimal in any known sense. However, when the Gaussian assumption for  $W(\cdot)$  is justified, it is known that in the increasing-domain asymptotics framework and under certain regularity conditions, ML estimators are consistent and asymptotically normal (see Section 3.4).



Add the Indicator variable corresponding to  $\hat{\lambda}_{max}^*$  into  $X$  if  $\hat{\lambda}_{max}^* > t$ , where  $t$  is the critical value (see Chap. 3 for details).

5. Repeat Steps 1 – 4 until no more outliers are found.
6. Estimate the current model and use the MLE of  $\sigma^2$  in the next step.
7. Delete the Indicator variable corresponding to  $\hat{\lambda}_{min} = \min_{s: X_s \in \text{col}(X)} |\hat{\lambda}^n(s)|$ , if  $\hat{\lambda}_{min} < t$ .
8. Repeat Steps 6 and 7 until no more Indicator variables are deleted.

Notice that the full estimation of the model takes places only at Steps 1 and 6. In the computation of the test statistics at Step 3, the parameters are held constant at their values estimated at Steps 1 and 2, and only the *GLS* regression computations are performed for each outlier detection statistic.

## 1.2.2 Other Types of Outliers

Additive outliers are assumed to occur without a specific pattern. Outliers may also occur in the data in a certain pattern. For example, we may have a *cluster of outliers* centered at some location  $s_0$ , with a radius  $r$ . Then,

$$O(s) = \begin{cases} \beta(s_0) & \|s - s_0\| \leq r \\ 0 & \|s - s_0\| > r \end{cases} \quad (1.15)$$

where  $\beta(s_0)$  is the magnitude of the outlier at all observations which are within  $r$  radius of  $s_0$ . Define

$$X_t^r(s) = \begin{cases} 1 & \|t - s\| \leq r \\ 0 & \|t - s\| > r. \end{cases} \quad (1.16)$$

Let  $S_C$  be the set of cluster centers of the *cluster* outliers. Then we have that  $Y(\cdot)$  is a stationary Gaussian random field, with

$$E(Y(s)) = \mu + \sum_{t \in S_C} \beta(t) X_t^r(s), \quad \forall s \in S \quad (1.17)$$

and with autocovariance function  $C_\theta(\cdot)$  of  $W(\cdot)$ , as in (1.3). Clusters of outliers may be observed in medical images, where a patch of tissue may be considered "abnormal"; in weather modeling, where a natural phenomenon such as an earthquake or a tornado may be observed over a spatial region; or in environmental applications, where the effect of a certain pollutant in the air may be felt over a spatial region.

Outliers may also occur in a cluster in a *decaying* manner. This may happen in all of the examples above, where there is a peak in the cluster and then the effect of an outlier starts diminishing as we move away from the center of the cluster. For example, the epicenter of an earthquake can be the center of the cluster, where the impact of the earthquake diminishes as we move away from the epicenter. Suppose an outlier of magnitude  $\beta(s_0)$  is present at  $s_0$  and as we move away from  $s_0$ , the outliers are still observed, but at a decaying rate  $\delta$ . In other words,

$$O(s) = \begin{cases} \beta(s_0) & s = s_0 \\ \delta^{\|s-s_0\|} \beta(s_0) & s \neq s_0 \end{cases} \quad (1.18)$$

where  $\beta(s_0)$  is the magnitude of the outlier at  $s_0$ . Define

$$X_t^d(s) = \begin{cases} 1 & s = t \\ \delta^{\|s-t\|} & s \neq t \end{cases} \quad (1.19)$$

Let  $S_D$  be the set of cluster centers of the *decaying* outliers. Then  $Y(\cdot)$  is a stationary Gaussian random field, with

$$E(Y(s)) = \mu + \sum_{t \in S_D} \beta(t) X_t^d(s), \quad \forall s \in S \quad (1.20)$$

and having autocovariance function  $C_\theta(\cdot)$  of  $W(\cdot)$ , as in (1.3). The detection algorithm in Section 1.2.1.1 can be used to detect *cluster* and *decaying* outliers. In this dissertation, we focus on establishing solid methodology for the detection of spatial AOs. Other types of outliers will be studied later.

### 1.2.3 Contributions of the dissertation

We commenced this work with the idea of extending the Indicator Variable outlier detection algorithm, as laid out in Section 1.2.1.1, to higher dimensional data. Our interest lies primarily in spatial data, indexed by a two-dimensional geographical location. Further, we assume that  $W(\cdot)$  is a Gaussian process. An integral part of the algorithm is the selection of an appropriate critical value  $t_\alpha$ , corresponding to a size  $\alpha$  test, for the hypothesis test in (1.9). The distribution of the corresponding test statistic  $\max_{s \in S_n} |\lambda^n(s)|$ , as in (1.10), is unknown in this case. For time series data, Chang et al. [4] computed the critical values via Monte Carlo simulation and Ljung [5] studied the use of the asymptotic distribution of  $\max_{s \in S_n} |\lambda^n(s)|$  in critical value selection. The major contribution of this thesis lies in identifying and studying

two bounds on  $\max_{s \in S_n} |\lambda^n(s)|$ , based on an approximation to the tail distribution of  $\max_{s \in S_n} \psi^n(s)$ . Here,  $\psi^n(x, y) = (-1)^{(x+y)} \lambda^n(x, y)$ . The need for this transformation is explained in Section 3.3.1.

We assume increasing-domain asymptotics for  $Y(\cdot)$  sampled on the regular lattice  $\mathbb{Z}_N^2 = \{(x, y) : x, y = 0, \pm 1, \pm 2, \dots, \pm N\}$ . When  $Y(\cdot)$  is sampled on a set  $S_n$ , we label the corresponding test statistic for the presence of an outlier at a location  $s$  as  $\lambda^n(s)$ . When  $S_n$  is  $\mathbb{Z}_N^2$ , with  $n = (2N + 1)^2$ , we label the test statistic for the presence of an outlier at location  $s$  as  $\Lambda^N(s)$ . This avoids the need to specify  $n$  in terms of  $N$ . For asymptotic results, we take  $N \rightarrow \infty$ .

One of the two bounds on the tail probability of  $\max_{s \in \mathbb{Z}_N^2} \psi^N(s)$  that we studied, the DLM bound (see Section 3.3.3.2), requires stationarity of  $\psi^N(\cdot)$ , which follows from stationarity of  $\Lambda^N(\cdot)$ . We show in Theorem 1 that  $\Lambda^N(\cdot)$  is asymptotically stationary. We also provide the form of the corresponding limiting autocovariance function in Corollary 2. Since the result in Theorem 1 is asymptotic, we also study the finite sample distributional properties of  $\Lambda^N(\cdot)$  for the special cases of the underlying process  $Y(\cdot)$  that we were interested in studying. In Section 3.3.2.1, we show that under the popular Matérn model for  $Y(\cdot)$  (see Section 2.2.1.1),  $\Lambda^N(\cdot)$  is approximately stationary inside the lattice that  $Y(\cdot)$  is sampled on. In Section 3.3.2.2, we show that when  $Y(\cdot)$  comes from the ICAR(1) model (see Section 2.2.2),  $\Lambda^N(\cdot)$  is exactly stationary inside the lattice that  $Y(\cdot)$  is sampled on. The attractive feature of this bound is that it accurately approximates the tail probability of  $\max_{s \in \mathbb{Z}_N^2} \psi^N(s)$  regardless of the degree of correlation in  $\psi^N(\cdot)$ .

The Bonferroni bound on the tail probability of  $\max_{s \in \mathbb{Z}_N^2} \psi^N(s)$  (see Sec-

tion 3.3.3.1) accurately approximates the tail probability of  $\max_{s \in \mathbb{Z}_N^2} \psi^N(s)$  when  $\psi^N(\cdot)$  and thus  $\Lambda^N(\cdot)$  is weakly correlated. In Sections 3.3.2.1 and 3.3.2.2, we show that under the Matérn and ICAR(1) models for  $Y(\cdot)$ ,  $\Lambda^N(\cdot)$  exhibits weak correlations, regardless of the correlation structure of  $Y(\cdot)$ . Thus, the use of the Bonferroni bound on the tail probability of the  $\max_{s \in \mathbb{Z}_N^2} \psi^N(s)$  is justified in this case. The attractive feature of this bound is that the corresponding critical values for  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$  are extremely easy and fast to calculate.

In the likely case when the parameters  $\tau$  of  $Y(\cdot)$  are unknown, the test statistic for the presence of at least one outlier in  $Y(\cdot)$  is  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s, \hat{\tau}_N)|$ , where  $\hat{\tau}_N$  is the MLE of  $\tau$ . We show in Theorem 6 that when  $\hat{\tau}_N$  is a consistent estimate of  $\tau$ ,  $\Lambda^N(\hat{\tau}_N, s) \xrightarrow[N \rightarrow \infty]{P} \Lambda(\tau, s)$ , for all  $s \in \mathbb{Z}^2$ , where  $\mathbb{Z}^2 = \{(x, y) : x, y = 0, \pm 1, \pm 2, \dots\}$ . Here  $\Lambda(\tau, s)$  is a stationary Gaussian process, given as in Theorem 1. In Section 3.4.1, we discuss asymptotic properties of parameter estimates of a Gaussian random field  $Y(\cdot)$  defined on  $\mathbb{Z}^d$ . In particular, Theorem 7 and Proposition 6 state conditions under which the MLE  $\hat{\tau}_N$  is consistent for  $\tau$ , assuming increasing-domain asymptotics for  $Y(\cdot)$  sampled on a regular lattice.

Given our results, as  $N$  increases, for a fixed  $t$ ,  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(\hat{\tau}_N, s)| > t)$  should approach  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda(\tau, s)| > t)$ . In turn, in Section 3.3, we demonstrate that when  $Y(\cdot)$  comes from either the Matérn or the ICAR(1) model,  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda(\tau, s)| > t)$  is well approximated by its Bonferroni bound, as in (3.87). Since the Bonferroni bound does not depend on  $\tau$ , we do not need to estimate  $\tau$  to calculate it. Thus, as long as  $N$  is large enough to make  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(\hat{\tau}_N, s)| > t)$  close to  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda(\tau, s)| > t)$ , the critical values for the test statistic in (3.90) should be

well approximated by the Bonferroni critical value approximation, as in (3.88).

We also studied the power of the test in (3.28), where we assume the presence of a single outlier in  $Y(\cdot)$  under the alternative hypothesis  $H_a$ . These results are given in Chapter 4.

## Chapter 2: Spatial Setting Considered

When we can not keep the discussion general, special cases of the Gaussian random field  $W(\cdot)$  need to be considered. Also, we need to establish the index set  $S$  on which the process  $W(\cdot)$  is defined as well as the sampling scheme.

One reason to consider special cases of  $W(\cdot)$  is to be able to derive properties of the outlier detection test statistic  $\max_{s \in S_n} |\lambda^n(s)|$  in (1.10) when model parameters are unknown. When  $\lambda^n(\cdot)$  is a function of estimated parameters, its distribution is typically unknown. However, in some cases, we can appeal to asymptotic properties of parameter estimates under specific models to make statements about the asymptotic distribution of  $\lambda^n(\cdot)$ . Also, under some specifications of  $W(\cdot)$ ,  $\lambda^n(\cdot)$  possesses certain attractive properties, which should be pointed out. For example, the Bonferroni bound on the tail probability of  $\lambda^n(\cdot)$  is accurate when  $\lambda^n(\cdot)$  is weakly correlated (see Section 3.3.3.1). In some cases, such as when  $W(\cdot)$  comes from the ICAR(1) model (see Section 2.2.2),  $\lambda^n(\cdot)$  is weakly correlated, *regardless* of  $\theta$ , the ICAR(1) autocorrelation parameters. Thus, in this case, the Bonferroni bound, being very simple to calculate, can be advocated for critical value estimation.

The asymptotic properties of parameter estimates also depend on the index set  $S$  on which the process is defined, as well as the sampling methodology. Section 2.1

describes the index set  $S$  and the sampling methodology considered in this thesis.

## 2.1 Index set $S$ and the sampling scheme

Generally, spatial data comes in one of the three forms give below:

1. *Geostatistical data.*  $S$  is a fixed subset of  $\mathbb{R}^d$ ;  $W(s)$  is a random vector at location  $s \in S$ .
2. *Lattice data.*  $S$  is a fixed (regular or irregular) collection of countably many points of  $\mathbb{R}^d$ ;  $W(s)$  is a random vector at location  $s \in S$ .
3. *Point Patterns.*  $S$  is a point process in  $\mathbb{R}^d$  or a subset of  $\mathbb{R}^d$ ;  $W(s)$  is a random vector at location  $s \in S$ .

In this thesis, we consider lattice data, with  $S = \mathbb{Z}^2$  where  $\mathbb{Z}^2$  is defined as  $\{(x, y) : x, y = \{0, \pm 1, \pm 2, \dots\}\}$ . One reason to consider lattice data is because there are more results on asymptotic properties of parameter estimates for processes defined on a regular grid as opposed to irregularly sited observations [11, p.85]. For more details on asymptotic properties of parameter estimates see Section 3.4.1. Also, the DLM bound on the tail probability of  $\max_{s \in S_n} \lambda^n(s)$  in (1.10), described in Section 3.3.3.2, is most naturally defined for regular lattice data (although extension to other types of data structures seem possible). Since we were interested in examining the usefulness of this bound, we were further prompted to assume the regular lattice structure.

Next we discuss the sampling scheme. Let  $\mathbb{Z}_N^2 \equiv \{(x, y) : x, y = \{0, \pm 1, \pm 2, \dots, \pm N\}\}$ . Then, the corresponding sample  $W_n \equiv \{W(s) : s \in \mathbb{Z}_N^2\}$  is ordered lexico-



graphically as follows. The  $k$ th element of  $W_n$  corresponds to  $s = (x, y)$ , such that  $k = (2N + 1)(y + N) + (x + N + 1)$ . Notice that  $W_n$  has  $n = (2N + 1)^2$  elements. We increase the sample size  $n$  by increasing  $N$ . Notice that we do not increase the sample size  $n$  by 1, rather we increase  $N$  by 1, which results in a  $8(N + 1)$  increment in  $n$ . However, as  $N \rightarrow \infty$ , so does  $n \rightarrow \infty$ . As  $N \rightarrow \infty$ , the index set  $Z_N^2$  converges to  $Z^2$ . This type of spatial asymptotics is referred to as *increasing domain asymptotics*. We give more details on spatial asymptotics in Section 3.4.1.

## 2.2 Model for $W(\cdot)$

In the most general case, we consider the following model for  $W(\cdot)$ . Let  $s \in \mathbb{R}^d$  be a point in a  $d$ -dimensional Euclidean space and suppose  $W(s)$  is a random quantity. Let  $s$  vary over an index set  $S \subset \mathbb{R}^d$ , assumed fixed, so as to generate a random field

$$\{W(s) : s \in S\} \tag{2.1}$$

We restrict our study to the case where  $W(\cdot)$  is a stationary Gaussian random field, with

$$E(W(s)) = \mu, \forall s \in S \tag{2.2}$$

and

$$\text{cov}(W(s), W(t)) = C_\theta(s - t), \forall s, t \in S \tag{2.3}$$

where  $C_\theta(\cdot)$  is the *autocovariance function*. The Gaussianity assumption is chosen because it allows a large number of results and methods to be derived analytically. The Gaussianity assumption turns out to be reasonable in many practical problems,

especially after a suitable point-wise transformation of the data is taken. The assumption of stationarity is made in order to be able to make inference from a *single* realization of the random field  $W(\cdot)$ . The constant mean model is chosen partly for ease of exposition. Section 2.2.1 provides justification for using a constant mean stationary Gaussian model when the goal is spatial interpolation.

### 2.2.1 Canonical model for kriging

For statistical methodologies to be broadly and effectively employed, it is important to have canonical models which work reasonably well for a wide range of problems. For the purposes of interpolating spatial data in  $d$  dimensions (see Section A.3.1), Stein [12, Sec. 1.6] proposes the following model:  $W(s) = \mu + \epsilon(s)$ , where  $\mu$  is an unknown constant and  $\epsilon(s)$  is a mean 0 stationary isotropic Gaussian random field with autocovariance function from the Matérn class, detailed below. He does not imply that all, or even most, spatial data can be reasonably modeled in this fashion. However, by making prudent extensions to this model where appropriate, by including, for example, geometric anisotropies, measurement errors or by taking a point-wise transformation of the observations, Stein argues the present practice of spatial modeling can be distinctly improved.

The most important reason for adopting the Matérn class of models is the inclusion of the parameter  $\nu$  in the model (see Section 2.2.1.1), which controls the smoothness of the random field. The larger  $\nu$  is, the smoother the random field. Smoothness in this case refers to the degree of differentiability of the autocovariance

function  $C_\theta(\cdot)$  at the origin (see Section A.3.3 for details). In particular,  $W(\cdot)$  will be  $m$  times mean square differentiable if  $\nu > m$ . Stein [12, Chapter 3] argues that when the goal is interpolation at locations surrounded by neighboring observations, the smoothness of the underlying random field plays a crucial role. Many autocovariance functions, such as exponential and Gaussian, for example, *assume* the smoothness as known and fixed and do not allow it to be estimated from the data.

Stein argues against inclusion of mean functions of the form  $X'\beta$ . He claims that commonly, when this is done in practice, the mean functions turn out to be highly smooth and to have little impact on the local behavior of the random field and, in turn, on interpolation. However, in arguing for less emphasis on modeling the mean function when the goal is spatial interpolation, Stein suggests it is nevertheless important to include mean functions which do have a strong effect on the local behavior of the random field. For example, when interpolating monthly average surface temperatures in a region based on scattered observations, one might use altitude as a component of  $X$ . In a mountainous region, variations in altitude may largely explain local variations in average temperature and hence including altitude as a component of  $X$  may have a profound effect on the spatial interpolation of average temperatures.

### 2.2.1.1 Matérn Model

Let  $h_s = (h_x, h_y)$  denote an increment in  $s$  along each axis and let  $h = \|h_s\| = \sqrt{h_x^2 + h_y^2}$  denote the Euclidean distance between the corresponding points. The

isotropic Matérn autocovariance function takes the following form:

$$C_\theta(h) = \sigma^2 K_\theta(h) \quad (2.4)$$

where  $K_\theta(h)$  is the autocorrelation function, defined as

$$K_\theta(h) = \frac{2}{\Gamma(\nu)} \left( \frac{h}{2\phi} \right)^\nu J_\nu \left( \frac{h}{\phi} \right) \quad (2.5)$$

where  $(\phi, \nu) \in \theta$  are the *range* and *smoothness* parameters, respectively.  $J_\nu(z)$  is the modified Bessel function of the third kind of order  $x$  and  $\Gamma(\cdot)$  is the gamma function. The Matérn class is flexible and includes:

- Exponential family (with  $\nu = 1/2$ ), where

$$K_\theta(h) = e^{-h/\phi}. \quad (2.6)$$

- In general, when  $\nu$  is of the form  $m + \frac{1}{2}$ , where  $m$  is a non-negative integer, the autocovariance function is of the form  $e^{-h/\phi}$  times a polynomial in  $h$  of degree  $m$ . For example, when  $\nu = \frac{3}{2}$ ,

$$K_\theta(h) = e^{-h/\phi} \left( \frac{h}{\phi} + 1 \right); \quad (2.7)$$

when  $\nu = \frac{5}{2}$ ,

$$K_\theta(h) = e^{-h/\phi} \left\{ \frac{1}{3} \left( \frac{h}{\phi} \right)^2 + \frac{h}{\phi} + 1 \right\}. \quad (2.8)$$

- Gaussian family (as  $\nu \rightarrow \infty$ ), where

$$K_\theta(h) = e^{-(h/\phi)^2}. \quad (2.9)$$

The parameter  $\nu$  controls the smoothness of the random field. The range parameter  $\phi$  rescales the distance axis, but it does not control the local behavior (smoothness) of

the random field. In Figure 2.1 (top), the Matérn autocorrelation function is plotted with  $\phi$  fixed at 1 and varying  $\nu$ . The smoothness parameter  $\nu$  affects the slope of the autocorrelation function at 0, moreover, its effect is seen on correlations at multiple lags. In Figure 2.1 (bottom), the Matérn autocorrelation function is plotted with  $\nu$  fixed at 0.5, thus resulting in the exponential covariance function, and varying  $\phi$ . We see that the range parameter controls how quickly the autocorrelation decays with distance.

### 2.2.2 ICAR(1) Model

In addition to the canonical model, proposed by Stein [12] and described in Section 2.2.1, we consider a special case of Gaussian conditional autoregressions, which are discussed in Section A.3.2.

**Definition 1.** Consider  $s = (x, y) \in \mathbb{Z}^2$ . Define  $\mathcal{N}(s)$ , the set of nearest neighbors of  $s$ , as

$$\mathcal{N}(s) = \{(x - 1, y), (x + 1, y), (x, y - 1), (x, y + 1)\}. \quad (2.10)$$

**Definition 2 (ICAR(1)).** Let  $\{W(s) : s \in \mathbb{Z}^2\}$  be a stationary Gaussian random field with conditional moments specified as:

$$E(W(s)|W_{(s)}) = \mu + \rho \sum_{t \in \mathcal{N}(s)} (W(t) - \mu), \quad \text{var}(W(s)|W_{(s)}) = \sigma^2, \quad (2.11)$$

with  $|\rho| < 1/4$ . Here  $W_{(s)} = \{W(t) : t \in \mathbb{Z}^2 \setminus s\}$ . We refer to such random fields as *Isotropic Conditionally Autoregressive Model of Order 1 (ICAR(1))*.

The existence of such processes was demonstrated by Rozanov [13]. The re-

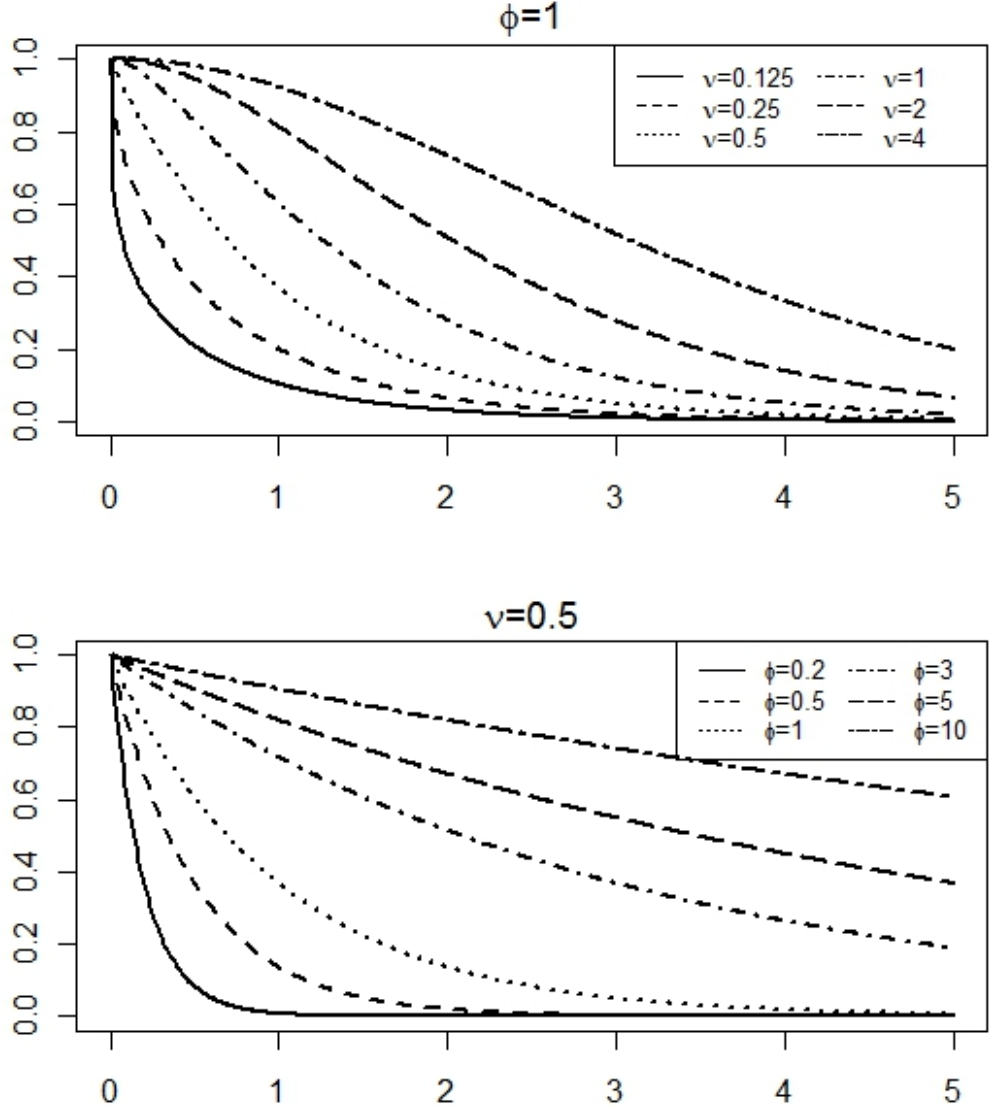


Figure 2.1: Matérn Autocorrelation Function

sulting autocovariance function  $C_\theta(h_x, h_y)$  is given in [14], as

$$C_\theta(h_x, h_y) = \frac{\sigma^2}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\cos(\omega_1 h_x + \omega_2 h_y) d\omega_1 d\omega_2}{1 - 2\rho(\cos(\omega_1) + \cos(\omega_2))} \quad (2.12)$$

with  $\theta = (\sigma^2, \rho)$ . The joint distribution of a sample  $W_n = \{W(s) : s \in \mathbb{Z}_N^2\}$  is then

given as

$$W_n \sim N_n(\mu \mathbf{1}, \Sigma_\theta) \quad (2.13)$$

where  $\Sigma_\theta$  is generated from (2.12). Then [15, Sec. 6.4] shows that  $\Sigma_{\theta,n}^{-1}$  takes the form:

$$\Sigma_{\theta,n}^{-1} = \frac{1}{\sigma^2} (I - \rho M_n) \quad (2.14)$$

where  $M_n \equiv (m_{st}^n)_{s,t \in \mathbb{Z}_N^2}$  is a matrix of weights, which does not depend on the parameter  $\theta$ . Notice the dependence on  $n$  of  $\Sigma_{n,\theta}^{-1}$ . Then for all  $s \in \mathbb{Z}_{N-1}^2$  and all  $t \in \mathbb{Z}_N^2$

$$m_{st}^n = \begin{cases} 1 & t \in \mathcal{N}(s) \\ 0 & \text{otherwise.} \end{cases} \quad (2.15)$$

Let  $\mathcal{B}_N \equiv \mathbb{Z}_N^2 \setminus \mathbb{Z}_{N-1}^2$  be the set of edge points of  $\mathbb{Z}_N^2$ . The weights  $m_{st}^n$  for  $s \in \mathcal{B}_N$  can be found by inverting  $\Sigma_{\theta,n}$ . We initially considered this class of models with the hope of simplifying calculations, as the precision matrix in this case is mostly sparse. The time series equivalent of the ICAR(1) process is the Autoregressive Order 1 (AR(1)) process, which does result in simplified calculations due to the sparsity in the precision matrix. However, in the two-dimensional case, complications arise due to the non-negligible effect of the boundary  $\mathcal{B}$ , where the weights  $m_{st}^n$  are not specified as in (2.15) and can only be calculated by inverting the covariance matrix  $\Sigma_\theta$ . The computation of  $\Sigma_\theta$  requires computation of  $C_\theta$  in (2.12) at relevant lags. However, the computation of  $C_\theta(h_s)$ , is very time consuming at large lags  $h_s$ . In Section 3.3.2.2, we propose a way to approximate  $\Sigma_{\theta,n}^{-1}$  with a circulant matrix, which is sparse and fast to compute.

## Chapter 3: Critical Value

For a review of hypothesis testing, see Section A.4. When constructing a hypothesis test, the goal is to have the largest possible power when the Type I error probability is uniformly constrained by a specified significance level. Let  $T$  be the test statistic. Let  $t_\alpha$  be the critical value corresponding to a *size*  $\alpha$  test, that is,

$$P_{H_0}(T > t_\alpha) = \alpha \tag{3.1}$$

In this chapter, we discuss how the critical value for a *size*  $\alpha$  test is derived or estimated under different scenarios. We start with the most trivial case, when all of the model parameters are known and also the location of the outlier is known. The scenario that is the most realistic is that neither the parameter values nor the location of the outlier is known a priori. However, in this case, the distribution of the corresponding test statistic is unknown and we must rely on approximating distributions and asymptotic results on parameter estimates to approximate  $t_\alpha$ . Under the null hypothesis of "no outliers", we assume that the sampled data  $Y_n \equiv \{Y(s) : s \in S_n \subset \mathbb{R}^d\}$  is

$$Y_n \sim N_n(\mu, \Sigma_\theta) \tag{3.2}$$

where the entries in  $\Sigma_\theta$  are defined by  $C_\theta(s - t) = \text{cov}(Y(s), Y(t))$ . We also assume that  $\Sigma_\theta = \sigma^2 D_\gamma$ , where  $\theta = (\sigma^2, \gamma)$ .



### 3.1 $\mu$ and $\theta$ known, location $s_0$ of a potential outlier known

Suppose that the parameters  $\tau = (\mu, \theta)$  are known and the location of the outlier  $s_0 \in S_n$  is also known. Then to test whether the outlier effect is statistically significant, we form the following hypothesis test:

$$H_0 : \beta(s_0) = 0 \tag{3.3}$$

$$H_a : \beta(s_0) \neq 0$$

and the corresponding test statistic:

$$\lambda^n(s_0) = \frac{\hat{\beta}(s_0)}{\sqrt{\text{var}(\hat{\beta}(s_0))}} \tag{3.4}$$

where  $\beta(s_0)$  is the regression coefficient corresponding to  $X_{s_0}$ , defined as in (1.7). The superscript  $n$  indicates that we are testing for an outlier at location  $s_0$ , relative to  $Y(\cdot)$  on  $S_n$ . The statistic  $\lambda^n(s)$  is the Wald statistic and we show below that a hypothesis test based on  $\lambda^n(s)$  is equivalent to the likelihood ratio test. Let  $Z_n = Y_n - \mu \mathbf{1}$  and let  $L(\beta|z_n)$  be the likelihood function of  $\beta$  as a function of observed data  $z_n$ . Consider the likelihood ratio statistic for this hypothesis test:

$$LR = \frac{L(0|z_n)}{\sup_{\beta} L(\beta|z_n)} \tag{3.5}$$

The numerator

$$L(0|z_n) \propto |\Sigma|^{-1/2} \exp\left(-\frac{z'\Sigma^{-1}z}{2}\right) \tag{3.6}$$

The denominator

$$\sup_{\beta} L(\beta|z_n) \propto |\Sigma|^{-1/2} \exp\left(-\frac{(z - X\hat{\beta})'\Sigma^{-1}(z - X\hat{\beta})}{2}\right) \tag{3.7}$$

where  $\hat{\beta}$ , the MLE of  $\beta$  under  $H_a$ , is given by:

$$\hat{\beta}(s_0) = (X'_{s_0}\Sigma^{-1}X_{s_0})^{-1}X'_{s_0}\Sigma^{-1}z_n \quad (3.8)$$

Now consider

$$\begin{aligned} -2\log(\text{LR}) &= z'\Sigma^{-1}z - (z - X\hat{\beta})'\Sigma^{-1}(z - X\hat{\beta}) \\ &= 2z'\Sigma^{-1}X\hat{\beta} - (X\hat{\beta})'\Sigma^{-1}X\hat{\beta} = \frac{\hat{\beta}^2}{\text{var}(\hat{\beta})} \end{aligned} \quad (3.9)$$

where  $\text{var}(\hat{\beta}) = (X'\Sigma^{-1}X)^{-1}$ . This is because

$$z'\Sigma^{-1}X\hat{\beta} = \frac{z'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}\hat{\beta}}{(X'\Sigma^{-1}X)^{-1}} = \frac{\hat{\beta}^2}{\text{var}(\hat{\beta})} \quad (3.10)$$

and

$$(X\hat{\beta})'\Sigma^{-1}X\hat{\beta} = \hat{\beta}^2 X'\Sigma^{-1}X = \frac{\hat{\beta}^2}{\text{var}(\hat{\beta})}. \quad (3.11)$$

Therefore,  $-2\log(\text{LR}) = \lambda^n(s_0)^2$ . Thus the likelihood ratio test, which rejects the null hypothesis for small values of the likelihood ratio test statistic LR, has the same rejection region as a test of the same size which is based on  $\lambda^n(s_0)^2$  (or alternatively  $|\lambda^n(s_0)|$ ), and rejects for large values of either of these statistics. The distribution of the test statistic  $\lambda^n(s_0)$  under  $H_0$  is standard Gaussian:

$$\lambda^n(s_0) = \frac{\hat{\beta}(s_0)}{\sqrt{\text{var}(\hat{\beta}(s_0))}} \sim N(0, 1) \quad (3.12)$$

The critical value  $t_\alpha = z_{\alpha/2}$ , where  $P(Z > z_\alpha) = \alpha$  with  $Z \sim N(0, 1)$ .

### 3.2 $\sigma^2$ unknown, other parameters known, location $s_0$ of a potential outlier known

This scenario differs from Section 3.1 only in that  $\sigma^2$  is unknown. So

$$Y_n \sim N_n(\mu, \sigma^2 D_\gamma) \quad (3.13)$$

and  $\sigma^2$  is unknown. The location of the outlier is still assumed known to be  $s_0$ . Thus, the hypothesis test is as in (3.3). Again, let  $Z_n = Y_n - \mu \mathbf{1}$ . The test statistic is

$$\tilde{\lambda}^n(s_0) = \frac{\hat{\beta}(s_0)}{\sqrt{\widehat{\text{var}}(\hat{\beta}(s_0))}} \quad (3.14)$$

where

$$\widehat{\text{var}}(\hat{\beta}(s_0)) = s_*^2 (X' \Sigma^{-1} X)^{-1} \quad (3.15)$$

and  $s_*^2$  is an unbiased estimator of  $\sigma^2$  under  $H_a$  :

$$s_*^2 = \frac{(z - X\hat{\beta})' D^{-1} (z - X\hat{\beta})}{n - 1}. \quad (3.16)$$

The likelihood ratio test statistic is

$$LR = \frac{\sup_{\sigma^2} L(0, \sigma^2 | z_n)}{\sup_{\beta, \sigma^2} L(\beta, \sigma^2 | z_n)} = \frac{L(0, \hat{\sigma}_0^2)}{L(\hat{\beta}, \hat{\sigma}_1^2)} \quad (3.17)$$

where  $\hat{\sigma}_0^2$  is the MLE of  $\sigma^2$  under  $H_0$ ;  $\hat{\beta}$  and  $\hat{\sigma}_1^2$  are MLEs of  $\beta$  and  $\sigma^2$  respectively under  $H_1$ . These are given below:

$$\hat{\sigma}_0^2 = \frac{z' D^{-1} z}{n} \quad (3.18)$$

$$\hat{\beta} = (X' D^{-1} X)^{-1} X' D^{-1} z \quad (3.19)$$

$$\hat{\sigma}_1^2 = \frac{(z - X\hat{\beta})' D^{-1} (z - X\hat{\beta})}{n}. \quad (3.20)$$

We then have

$$L(0, \hat{\sigma}_0^2) \propto |\hat{\sigma}_0^2 D|^{-1/2} \exp\left(-\frac{z' D^{-1} z}{2\hat{\sigma}_0^2}\right) \quad (3.21)$$

$$L(\hat{\beta}, \hat{\sigma}_1^2) \propto |\hat{\sigma}_1^2 D|^{-1/2} \exp\left(-\frac{(z - X\hat{\beta})' D^{-1} (z - X\hat{\beta})}{2\hat{\sigma}_1^2}\right). \quad (3.22)$$

In both Equations (3.21) and (3.22), the exponentials simplify to  $-n/2$ . Therefore

$$LR = \frac{|\hat{\sigma}_0^2 D|^{-1/2}}{|\hat{\sigma}_1^2 D|^{-1/2}} = \left[\frac{(\hat{\sigma}_0^2)}{(\hat{\sigma}_1^2)}\right]^{-n/2}. \quad (3.23)$$

The algebraic reductions in (3.9) imply that

$$\hat{\sigma}_0^2 - \hat{\sigma}_1^2 = \frac{n\hat{\beta}^2}{(X'D^{-1}X)^{-1}}. \quad (3.24)$$

Thus,

$$(\tilde{\lambda}^n)^2 = \frac{\hat{\beta}^2}{\widehat{\text{var}}\hat{\beta}} = \frac{\hat{\beta}^2}{(X'D^{-1}X)^{-1}s_*^2} = \frac{\hat{\sigma}_0^2 - \hat{\sigma}_1^2}{ns_*^2} = \frac{(\hat{\sigma}_0^2 - \hat{\sigma}_1^2)(n-1)}{n\hat{\sigma}_1^2} = \quad (3.25)$$

$$= \frac{n-1}{n} [(LR)^{-2/n} - 1] \quad (3.26)$$

Since LR is a monotone function of  $(\tilde{\lambda}^n)^2$ , the likelihood ratio test is equivalent to

a test based on  $|\tilde{\lambda}^n|$ . Under  $H_0$ ,

$$\tilde{\lambda}^n(s) \sim t_{n-1}. \quad (3.27)$$

The critical value is  $t_\alpha = t_{n-1, \alpha/2}$ , where  $P(Q > t_{n-1, \alpha}) = \alpha$  with  $Q \sim t_{n-1}$ .

### 3.3 $\mu$ and $\theta$ known, location $s_0$ of potential outlier unknown

Suppose now that we do not know the location of the outlier. Then the hypothesis test is

$$H_0 : \beta(s) = 0 \quad \forall s \in S_n \quad (3.28)$$

$$H_a : \exists t \in S_n : \beta(t) \neq 0$$

Another way of stating this is

$$H_0 : \cap_{t \in S_n} H_{a,t}^c \tag{3.29}$$

$$H_a : \cup_{t \in S_n} H_{a,t}$$

where

$$H_{0,t} : \beta(t) = 0 \tag{3.30}$$

$$H_{a,t} : \beta(t) \neq 0$$

is the hypothesis test for testing for the presence of a single outlier at location  $t$ , as in (3.3). The test statistic corresponding to the hypothesis test in (3.29) is

$$\max_{s \in S_n} (|\lambda^n(s)|) \tag{3.31}$$

where  $\lambda^n(s)$  is defined as in (3.4). In this case, the distribution of  $\max_{s \in S_n} (|\lambda^n(s)|)$  is unknown.

In the rest of this section, we assume that the process  $W(\cdot)$  is defined on the lattice  $\mathbb{Z}^2$ , and that the sample  $W_n$  is taken on  $\mathbb{Z}_N^2 \equiv \{(x, y) : x, y = 0, \pm 1, \pm 2, \dots, \pm N\}$ , with  $n = (2N + 1)^2$ . The sample  $W_n$  is indexed in lexicographical order. See Section 2.1 for more details. When we consider  $S_n$  to be the lattice  $\mathbb{Z}_N^2$ , with  $n = (2N + 1)^2$ , we denote  $\lambda^n(\cdot)$  as  $\Lambda^N(\cdot)$ . This avoids the need to define  $n$  in terms of  $N$ . Let

$$P = P_{H_0} \left[ \max_{s \in \mathbb{Z}_N^2} (|\Lambda^N(s)|) > t \right]. \tag{3.32}$$

Let  $\{\psi^N(x, y) : (x, y) \in \mathbb{Z}_N^2\}$  be defined as follows:

$$\psi^N(x, y) = (-1)^{x+y} \Lambda^N(x, y) \tag{3.33}$$

and let

$$P^* = P_{H_0} \left[ \max_{s \in \mathbb{Z}_N^2} (\psi^N(s)) > t \right]. \tag{3.34}$$

In Section 3.3.1, we discuss our reasoning for introducing  $\psi^N(\cdot)$  and discuss how we approximate  $P$  by  $P^*$ . In turn,  $P^*$  is approximated by two upper bounds, which are discussed in Sections 3.3.3.1 and 3.3.3.2. The performance of the two upper bounds in approximating  $P^*$  depends on the second moment properties of the vector  $\psi^N(\cdot)$ . In Section 3.3.2 we discuss the second moment properties of  $\psi^N(\cdot)$ , specifically when  $W_n$  comes from the two models we consider: Matérn and ICAR(1).

### 3.3.1 Approximating $P$ by a function of $P^*$

In this section, we discuss our approach to approximating  $P$  in (3.32) as a function of  $P^*$ , as given in (3.34). First, to take advantage of available theory on the tail probabilities of the maximum of a Gaussian random field, we rewrite  $P$  in terms of the Gaussian random field  $\Lambda^N$ , instead of the non-Gaussian  $|\Lambda^N|$ .

$$\begin{aligned}
P &= P \left[ \max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t \right] \\
&= P \left[ \max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t \right] + P \left[ \min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t \right] \\
&= -P \left[ \max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t, \min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t \right] \\
&= 2P \left[ \max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t \right] - P \left[ \max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t, \min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t \right].
\end{aligned} \tag{3.35}$$

The last equality is due to the fact that the distribution of  $\Lambda^N(\cdot)$  is symmetric about 0 under  $H_0$  and thus  $P(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t)$  and  $P(\min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t)$  are the same.

Now, to facilitate the approximation of  $P(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t)$  and  $P(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t, \min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t)$  in (3.35), we consider a random vector  $\psi^N(\cdot)$  defined as in (3.33). Since  $|\psi^N(\cdot)| = |\Lambda^N(\cdot)|$ ,  $P$  in (3.32) can be written as  $P(\max_{s \in \mathbb{Z}_N^2}$

$|\psi^N(s)| > t$ ) and each term of (3.35) can be written in terms of  $\psi^N(\cdot)$  instead of  $\Lambda^N(\cdot)$ .

The reason we express  $P$  in (3.35) as a function of  $\psi^N(\cdot)$ , instead of  $\Lambda^N(\cdot)$ , is the following. Let  $h_s = (h_x, h_y) \in \mathbb{Z}^2$  and let  $\tilde{h} = |h_x| + |h_y|$ . As will be shown in Section 3.3.2, for the processes  $W(\cdot)$  we consider,  $\text{cov}(\Lambda^N(s), \Lambda^N(s + h_s))$  is negative when  $\tilde{h} = 1$  and small for  $\tilde{h} > 1$ . However, one of the bounds we present for the tail probability of the maximum of a Gaussian random field (see Section 3.3.3) depends on the underlying field being positively correlated. The random field  $\psi^N(\cdot)$  is thus better suited for our analysis, since, as Proposition 1 implies,  $\text{cov}(\psi^N(s), \psi^N(s + h_s))$  is positive and small for  $\tilde{h} > 1$ .

**Proposition 1** (Distribution of  $\psi^N(\cdot)$ ). *For a fixed  $N$ , let  $\psi^N(x, y) = (-1)^{x+y}\Lambda^N(x, y)$  for all  $(x, y) \in \mathbb{Z}_N^2$ . Then  $\psi^N(\cdot)$  is a Gaussian mean-zero, unit-variance random field, with*

$$\text{cov}(\psi^N(s), \psi^N(s + h_s)) = (-1)^h \text{cov}(\Lambda^N(s), \Lambda^N(s + h_s)) \quad (3.36)$$

where  $h_s = (h_x, h_y) \in \mathbb{Z}^2$  and  $\tilde{h} = |h_x| + |h_y|$ .

Replacing  $\Lambda^N(\cdot)$  with  $\psi^N(\cdot)$  in (3.35), we have

$$P = P \left[ \max_{s \in \mathbb{Z}_N^2} |\psi^N(s)| > t \right] \quad (3.37)$$

$$= 2P \left[ \max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t \right] - P \left[ \max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t, \min_{s \in \mathbb{Z}_N^2} \psi^N(s) < -t \right]. \quad (3.38)$$

To approximate the first term in (3.37),  $P(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t)$ , we consider two bounds on the tail probability of the maximum of a Gaussian random field, as

discussed in Section 3.3.3. Now, consider the second term in (3.37):

$$P \left[ \max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t, \min_{s \in \mathbb{Z}_N^2} \psi^N(s) < -t \right]. \quad (3.39)$$

Let

$$B = \left\{ \max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t \right\}$$

$$C = \left\{ \min_{s \in \mathbb{Z}_N^2} \psi^N(s) < -t \right\}.$$

Then (3.39) can be written as  $P(B \cap C)$ . Next we explain why we think in this case

$P(B \cap C)$  is bounded above by  $P(B)P(C)$ . We can write

$$B \cap C = \{ \psi^N(s_i) > t \text{ and } \psi^N(s_j) < -t \text{ for some } s_i, s_j \in \mathbb{Z}_N^2 \}.$$

Now imagine a realization of  $\psi^N(\cdot)$  where  $\psi^N(s_i)$  exceeds  $t$  for some  $i$ . Then, if  $\psi^N(\cdot)$  is highly correlated, the random field will tend to linger around the threshold  $t$  and not venture down to  $-t$ . However, if  $\psi^N(\cdot)$  is weakly correlated, there is a higher chance that the process will venture down to  $-t$ . This heuristic argument supports the hypothesis that  $P(B \cap C)$  is maximized when  $\psi^N(\cdot)$  is uncorrelated. In turn, when  $\psi^N(\cdot)$  is uncorrelated,  $B$  and  $C$  are independent and thus  $P(B \cap C) = P(B)P(C)$ . This gives justification for using  $P(B)P(C)$  as an upper bound on  $P(B \cap C)$ . Also, since in this case  $P(B) = P(C)$  due to symmetry of the distribution of  $\psi^N(\cdot)$ , we have  $P(B \cap C) = P(B)^2$  and thus we can write (3.37) as

$$P = P(\max_{s \in \mathbb{Z}_N^2} |\psi^N(s)| > t) \approx 2P(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t) - P(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t)^2 \quad (3.40)$$

Next we justify omitting the second term in the approximation of  $P$  in (3.40). We are interested in using (3.40) to approximate the critical value  $t_\alpha$  for a size  $P = \alpha$  test,



by using the right hand side of (3.40) as an approximation for  $P$ . Typical values of  $\alpha$  are in the range of  $[0.01, 0.10]$ . Take  $\alpha = 0.05$  and let  $x = P_{H_0}(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t_\alpha)$ . Then, we have  $2x - x^2 \approx 0.05$ , which results in  $x = 0.0253$  and  $x^2 = 0.00064$ . Since  $x^2 = 0.00064$  is of smaller order of magnitude than  $2x = 0.0506$ , we drop the second term in (3.40) and approximate  $P$  as in

$$P = P(\max_{s \in \mathbb{Z}_N^2} |\psi^N(s)| > t) \approx 2P(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t) \quad (3.41)$$

### 3.3.2 Distribution of $\Lambda^N(\cdot)$ under $H_0$

Let  $W(\cdot)$  be defined as in (2.2) and (2.3) and consider a sample  $W_n \equiv \{W(s) : s \in \mathbb{Z}_N^2\}$ , with  $n = (2N + 1)^2$ . Let  $Z(\cdot) = Y(\cdot) - \mu$ . Then, under  $H_0$  in (3.28),

$$E(Z(s)) = 0, \quad \text{cov}(Z(s), Z(t)) = C_\theta(s - t), \quad \text{for all } s, t \in \mathbb{Z}^2 \quad (3.42)$$

and

$$Z_n \sim N_n(0, \Sigma_\theta) \quad (3.43)$$

with  $\Sigma_\theta$  defined lexicographically from  $C_\theta$ , as in Section 2.1. Let  $\Lambda^N(s)$  be the outlier detection statistic for location  $s$ . Proposition 2 gives the distribution of  $\Lambda^N(\cdot)$  under  $H_0$  in (3.28).

**Proposition 2 (Distribution of  $\Lambda^N(\cdot)$  under  $H_0$  in (3.28)).** *Consider the hypothesis test in (3.28) where, under  $H_0$ ,  $E(Y(\cdot)) = \mu$ . Let*

$$\Lambda^N(s) = \frac{\hat{\beta}(s)}{\sqrt{\text{var}(\hat{\beta}(s))}} \quad (3.44)$$

where  $\hat{\beta}$  is the generalized least squares estimator of  $\beta$ . Let  $\Sigma_{\theta,N}^{-1} \equiv (\sigma_N^{st})_{s,t \in \mathbb{Z}_N^2}$  be the inverse covariance matrix of  $\{Y(s) : s \in \mathbb{Z}_N^2\}$ . Then  $\Lambda^N(\cdot)$  is Gaussian with

$$\Lambda^N(s) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} (Y(t) - \mu) \quad (3.45)$$

where

$$E(\Lambda^N(s)) = 0, \text{ for all } s \quad (3.46)$$

and

$$\text{cov}(\Lambda^N(s), \Lambda^N(t)) = (\sigma_N^{ss})^{-1/2} (\sigma_N^{tt})^{-1/2} \sigma_N^{ts}. \quad (3.47)$$

*Proof.* Let  $Y_n \equiv \{Y(s) : s \in \mathbb{Z}_N^2\}$ , with  $n = (2N + 1)^2$ . For each  $s \in \mathbb{Z}_N^2$ ,  $\Lambda^N(s) = \frac{\hat{\beta}(s)}{\sqrt{\text{var}(\hat{\beta}(s))}}$ , where  $\hat{\beta}$  is the generalized least squares estimate of  $\beta$ , i.e.  $\hat{\beta}(s) = (X'_s \Sigma_N^{-1} X_s)^{-1} X'_s \Sigma_N^{-1} (Y_n - \mu \mathbf{1}_n)$  and having variance  $\text{var}(\hat{\beta}(s)) = (X'_s \Sigma_N^{-1} X_s)^{-1}$ , with  $X_s$  defined as in (1.7). Therefore,

$$\Lambda^N(s) = \frac{(X'_s \Sigma_N^{-1} X_s)^{-1} X'_s \Sigma_N^{-1} (Y_n - \mu \mathbf{1}_n)}{\sqrt{(X'_s \Sigma_N^{-1} X_s)^{-1}}} = (X'_s \Sigma_N^{-1} X_s)^{-1/2} X'_s \Sigma_N^{-1} (Y_n - \mu \mathbf{1}_n). \quad (3.48)$$

Now,

$$X'_s \Sigma_N^{-1} = (\sigma_N^{s1}, \dots, \sigma_N^{sn}), \quad X'_s \Sigma_N^{-1} X_s = \sigma_N^{ss}, \quad X'_s \Sigma_N^{-1} X_t = \sigma_N^{st}. \quad (3.49)$$

Then,

$$\Lambda^N(s) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} (Y(t) - \mu).$$

Since each  $\Lambda^N(s)$  is a linear combination of elements of the Gaussian random vector

$Z_n \equiv \{Z(s) : s \in \mathbb{Z}_N^2\}$ ,  $\Lambda^N(\cdot)$  is Gaussian. Since  $E(Z_n) = 0$ ,  $E(\Lambda^N(\cdot)) = 0$ . Now,

$$\begin{aligned} \text{cov}(\Lambda^N(s), \Lambda^N(t)) &= (X'_s \Sigma_N^{-1} X_s)^{-1/2} X'_s \Sigma_N^{-1} \Sigma_N ((X'_t \Sigma_N^{-1} X_t)^{-1/2} X'_t \Sigma_N^{-1})' \\ &= (X'_s \Sigma_N^{-1} X_s)^{-1/2} X'_s \Sigma_N^{-1} X_t (X'_t \Sigma_N^{-1} X_t)^{-1/2} \\ &= (\sigma_N^{ss})^{-1/2} \sigma_N^{st} (\sigma_N^{tt})^{-1/2}. \end{aligned} \tag{3.50}$$

□

It is important to note that the distribution of  $\Lambda^N(\cdot)$  does not depend on  $\mu$ . Thus, we do not need to vary  $\mu$  when simulating  $\Lambda^N(\cdot)$ . Also, in both the Matérn and ICAR(1) models that we consider,  $\Sigma_\theta$  can be written as  $\sigma^2 D_\gamma$ , for some matrix  $D_\gamma$  with  $\theta = (\sigma^2, \gamma)$ . Proposition 3 states that in this case, the distribution of  $\Lambda^N(\cdot)$  is invariant to the choice of  $\sigma^2$ . This is important to point out, because it implies that we do not need to vary  $\sigma^2$  when simulating  $\Lambda^N(\cdot)$ .

**Proposition 3** (Invariance of  $\Lambda^N(\cdot)$  to  $\sigma^2$ ). *Let  $\Sigma_\theta$  in Proposition 2 be expressed as  $\sigma^2 D_\gamma$ , for some matrix  $D_\gamma$ , with  $\theta = (\sigma^2, \gamma)$ . Then the distribution of  $\Lambda^N(\cdot)$  is invariant to the choice of  $\sigma^2$ .*

*Proof.* Let  $D_\gamma^{-1} \equiv (d^{st})_{s,t \in \mathbb{Z}_N^2}$ . Then

$$\xi_{st} = (\sigma^{ss})^{-1/2} (\sigma^{tt})^{-1/2} \sigma^{ts} = (\sigma^2 d^{ss})^{-1/2} (\sigma^2 d^{tt})^{-1/2} \sigma^2 d^{ts} = (d^{ss})^{-1/2} (d^{tt})^{-1/2} d^{ts}$$

.

□

In Section 3.3.3.2 we present the Discrete Local Maxima (DLM) bound on  $P^*$  [16]. This bound assumes stationarity of  $\psi^N(\cdot)$ , which would follow from stationarity of  $\Lambda^N(\cdot)$ , via Proposition 1. Although exact stationarity of  $\Lambda^N(\cdot)$  does

not generally hold, Theorem 1 states that  $\Lambda^N$  is asymptotically stationary. Recall that  $\Lambda^N(s)$  denotes the outlier detection statistic at location  $s$ , having observed  $Y(\cdot)$ , which is defined on  $\mathbb{Z}_N^2$ . Before proving Theorem 1, we present two supporting lemmas. Lemma 1 is due to [5].

**Lemma 1.** *Consider  $W_n \equiv \{W(u) : u \in S_n \subset \mathbb{R}^2\}$ . Let  $W_n \sim N_n(0, \Sigma_\theta)$ , with  $\theta$  assumed known. Let  $Y_n = W_n + X_s \beta(s)$ , where  $X_s$  is defined as in (1.7) and  $\beta(s)$  is the regression coefficient corresponding to  $X_s$ , with  $s$  known. Denote the maximum likelihood estimator of  $\beta(s)$  as  $\hat{\beta}(s) = (X_s' \Sigma_\theta^{-1} X_s)^{-1} X_s' \Sigma_\theta^{-1} Y_n$ . Then,  $\hat{\beta}(s)$  can be written as*

$$\hat{\beta}(s) = Y(s) - E(W(s)|W_{(s)}) \quad (3.51)$$

where  $W_{(s)} = \{W(t) : t \in S_n \setminus s\}$ .

*Proof.* Denote by  $W_0$  the vector  $W_n$ , with 0 substituted for  $W(s)$ . Then we can write  $W_n = W_0 + X_s W(s)$ , where  $X_s$  is defined as in (1.7). Consider the kriging predictor of  $W(s)$  given  $W_{(s)}$ :  $\hat{W}(s) = E(W(s)|W_{(s)})$ . We show next that:

$$\hat{W}(s) = -(X_s' \Sigma^{-1} X_s)^{-1} X_s' \Sigma^{-1} W_0 \quad (3.52)$$

Let  $Q_{ij}$  be an  $n \times n$  matrix which switches rows  $i$  and  $j$  of matrix  $M$ , when multiplied on the left of  $M$ . Then, when  $Q_{ij}$  is multiplied on the right of  $M$ , it switches columns  $i$  and  $j$  of  $M$ . We also have  $Q_{ij} = Q'_{ij}$  and  $Q_{ij}^{-1} = Q_{ij}$ . Let  $m$  be the index of  $Y(s)$  in  $Y_n$ . Let  $\tilde{\Sigma}_s = Q_{mn} \Sigma Q_{mn}$ ,  $\tilde{X} = Q_{mn} X_s$ ,  $\tilde{W}_0 = Q_{mn} W_0$ . It then follows that

$$\hat{W}(s) = -(X_s' \Sigma^{-1} X_s)^{-1} X_s' \Sigma^{-1} W_0 = -(\tilde{X}' \tilde{\Sigma}_s^{-1} \tilde{X})^{-1} \tilde{X}' \tilde{\Sigma}_s^{-1} \tilde{W}_0$$

Now,

$$\tilde{\Sigma}_s = \begin{pmatrix} \Sigma_{(s)} & \sigma_{(s),s} \\ \sigma_{(s),s} & \sigma_s \end{pmatrix}$$

with

$$\tilde{\Sigma}_s^{-1} = \begin{pmatrix} (\Sigma_{(s)} - \frac{1}{\sigma_s} \sigma_{(s),s} \sigma'_{(s),s})^{-1} & -\frac{1}{k} \Sigma_{(s)}^{-1} \sigma_{(s),s} \\ -\frac{1}{k} \sigma'_{(s),s} \Sigma_{(s)}^{-1} & \frac{1}{k} \end{pmatrix}$$

where  $k = \sigma_s - \sigma'_{(s),s} \Sigma_{(s)}^{-1} \sigma_{(s),s}$ . Now,

$$\tilde{X}' \tilde{\Sigma}_s^{-1} \tilde{X} = \frac{1}{k}, \quad \tilde{X}' \tilde{\Sigma}_s^{-1} \tilde{W}_0 = -\frac{1}{k} \Sigma_{(s)}^{-1} \sigma_{(s),s} W_{(s)}$$

Therefore,

$$\begin{aligned} \hat{W}(s) &= -(\tilde{X}' \tilde{\Sigma}_s^{-1} \tilde{X})^{-1} \tilde{X}' \tilde{\Sigma}_s^{-1} \tilde{W}_0 \\ &= -k \left( -\frac{1}{k} \Sigma_{(s)}^{-1} \sigma_{(s),s} W_{(s)} \right) \\ &= \Sigma_{(s)}^{-1} \sigma_{(s),s} W_{(s)} \\ &= E(W(s) | W_{(s)}) \end{aligned} \tag{3.53}$$

Now consider the GLS estimate of  $\beta(s)$ :

$$\hat{\beta}(s) = (X'_s \Sigma^{-1} X_s)^{-1} X'_s \Sigma^{-1} Y_n \tag{3.54}$$

Also, observe that  $Y_n = W_0 + X_s Y(s)$ . Then, replacing  $Y_n$  in (3.52) with  $W_0 + X_s Y(s)$ , we get:

$$\begin{aligned} \hat{\beta}(s) &= (X'_s \Sigma^{-1} X_s)^{-1} X'_s \Sigma^{-1} Y_n \\ &= (X'_s \Sigma^{-1} X_s)^{-1} X'_s \Sigma^{-1} (W_0 + X_s Y(s)) \\ &= (X'_s \Sigma^{-1} X_s)^{-1} X'_s \Sigma^{-1} W_0 + (X'_s \Sigma^{-1} X_s)^{-1} X'_s \Sigma^{-1} X_s Y(s) \\ &= Y(s) - \hat{W}(s) \end{aligned} \tag{3.55}$$

□

This Lemma carries significance on its own. It implies that the effect of an outlier at location  $s$ ,  $\beta(s)$ , can be estimated equivalently in two different ways. We can either obtain the GLS estimate, as in (3.54), or we can obtain the estimate of  $\beta(s)$  as the difference between the observation  $Y(s)$  and its interpolated value  $\hat{W}(s)$ .

**Lemma 2.** Let  $\Lambda^N(s) = \hat{\beta}(s)/\sqrt{\text{var}(\hat{\beta}(s))}$ . Then, under  $H_0$ ,  $\Lambda^N(s)$  can be written as

$$\Lambda^N(s) = \frac{Y(s) - E(Y(s)|Y_{(s)}^N)}{\sqrt{\text{var}(Y(s)|Y_{(s)}^N)}} \quad (3.56)$$

where  $Y_{(s)}^N = \{Y(t) : t \in \mathbb{Z}_N^2 \setminus s\}$ .

*Proof.* Under  $H_0$  in the hypothesis test (3.3),  $W(\cdot)$  and  $Y(\cdot)$  are identical. Then, Lemma 1 implies that, under  $H_0$ ,

$$\hat{\beta}(s) = Y(s) - E(Y(s)|Y_{(s)}^N). \quad (3.57)$$

The denominator of  $\Lambda^N(s)$  is

$$\sqrt{\text{var}(\hat{\beta}(s))} = \sqrt{\text{var}(Y(s) - E(Y(s)|Y_{(s)}^N))} = \sqrt{\text{var}(Y(s)|Y_{(s)}^N)}. \quad (3.58)$$

The last equality is due to the following statement regarding any two jointly Gaussian random variables  $X$  and  $Y$ :

$$\begin{aligned} \text{var}(Y - E(Y|X)) &= E[\text{var}(Y - E(Y|X)|X)] + \text{var}[E(Y - E(Y|X)|X)] \\ &= E[\text{var}(Y|X)] \\ &= \text{var}(Y|X) \end{aligned} \quad (3.59)$$

The last equality is due to the fact that for Gaussian random variables  $X$  and  $Y$ ,  $\text{var}(Y|X)$  does not involve  $X$ . The term  $\text{var}[E(Y - E(Y|X)|X)]$  is 0, because  $E(Y - E(Y|X)|X) = 0$ .  $\square$

**Theorem 1** (Asymptotic stationarity of  $\Lambda^N(\cdot)$ ). *Suppose  $Y = \{Y(s) : s \in \mathbb{Z}^2\}$  is a strictly stationary Gaussian process and let  $Y_{(s)}^N \equiv \{Y(t) : t \in \mathbb{Z}_N^2 \setminus \{s\}\}$ . For all  $s \in \mathbb{Z}^2$ , define*

$$\Lambda^N(s) = \frac{Y(s) - E(Y(s)|Y_{(s)}^N)}{\sqrt{\text{var}(Y(s)|Y_{(s)}^N)}} \mathbf{1}_{\mathbb{Z}_N^2}(s). \quad (3.60)$$

*Also assume that  $\text{var}(Y(0)|Y_{(0)}) = \gamma^2 > 0$ , where  $Y_{(s)} = \{Y(t) : t \in \mathbb{Z}^2 \setminus \{s\}\}$ . Then  $\Lambda^N \xrightarrow[N \rightarrow \infty]{a.s.} \Lambda$ , where  $\Lambda$  is a strictly stationary process.*

*Proof.* For each  $s \in \mathbb{Z}^2$ ,  $\{E(Y(s)|Y_{(s)}^N)\}_{N \in \mathbb{Z}}$  is a martingale in  $N$ , since

1.  $E(E(Y(s)|Y_{(s)}^{N+1})|Y_{(s)}^N) = E(Y(s)|Y_{(s)}^N)$ ,
2.  $E([E(Y(s)|Y_{(s)}^N)]^2) \leq E(E([Y(s)]^2|Y_{(s)}^N)) = E([Y(s)]^2) < \infty$ .

The first inequality in 2. follows from Jensen's Inequality for Conditional Expectations and the last from the fact that  $Y(\cdot)$  is a strictly stationary Gaussian process. Since 2. does not depend on  $N$ ,  $\{E(Y(s)|Y_{(s)}^N)\}_{N \in \mathbb{Z}}$  is a  $L^1$  bounded martingale and Doob's martingale convergence result holds (see Theorem 13), which implies that, for all  $s$ :

$$E(Y(s)|Y_{(s)}^N) \xrightarrow[N \rightarrow \infty]{a.s.} E(Y(s)|Y_{(s)}) \in L^1 \quad (3.61)$$

which, in turn, implies

$$Y(s) - E(Y(s)|Y_{(s)}^N) \xrightarrow[N \rightarrow \infty]{a.s.} Y(s) - E(Y(s)|Y_{(s)}) \in L^1 \quad (3.62)$$

Now, consider the denominator of (3.60). Since  $Y$  is Gaussian,  $\text{var}(Y(s)|Y_{(s)}^N)$  is a constant for all  $N$  and  $s$ . Moreover, consider a transformation  $T_s : \mathbb{Z}^2 \rightarrow \mathbb{Z}^2$ , such that for all  $t \in \mathbb{Z}^2$ ,  $T_s Y(t) = Y(t + s)$ . Then for all  $s$ ,  $T_s Y(0) = Y(s)$  and  $T_s Y_{(0)} = Y_{(s)}$ . Since  $Y$  is strictly stationary,  $T_s$  is measure-preserving (see Section A.1). Therefore, for all  $s$ ,  $\text{var}[Y(s)|Y_{(s)}^N] = \text{var}[T_s Y(0)|T_s Y_{(0)}^N] = T_s \text{var}[Y(0)|Y_{(0)}^N] = \text{var}[Y(0)|Y_{(0)}^N]$ . The last equality follows if we note that  $\text{var}[Y(0)|Y_{(0)}^N]$  is a constant in  $Y_{(0)}^N$  and thus  $T_s$  can be dropped. Since  $E(Y_0^2)$  is bounded, uniformly in  $N$ ,  $\{E(Y(0)|Y_{(0)}^N)\}_{N \geq 0}$  is a square integrable martingale, uniformly in  $N$  (see Section A.2.1). Thus,  $\lim_{N \rightarrow \infty} \text{var}(Y(0)|Y_{(0)}^N) = \text{var}(Y(0)|Y_{(0)})$  and thus we have, for all  $s$ ,

$$\lim_{N \rightarrow \infty} \text{var}(Y(s)|Y_{(s)}^N) = \text{var}(Y(0)|Y_{(0)}) = \gamma^2 > 0. \quad (3.63)$$

We then have, for each  $s$ ,

$$\Lambda^N(s) \xrightarrow[N \rightarrow \infty]{a.s.} \Lambda(s) \quad (3.64)$$

where

$$\Lambda(s) = \frac{Y(s) - E(Y(s)|Y_{(s)})}{\gamma} \quad (3.65)$$

Strict stationarity of  $\{\Lambda(s) : s \in \mathbb{Z}^2\}$  follows if we can show that for any  $s_1, \dots, s_k \in \mathbb{Z}^2$  and any  $A_1, \dots, A_k$ ,

$$P[E(Y(s_1 + h)|Y_{(s_1+h)}) \in A_1, \dots, E(Y(s_k + h)|Y_{(s_k+h)}) \in A_k]$$

does not depend on  $h$ . Consider a transformation  $T_s : \mathbb{Z}^2 \rightarrow \mathbb{Z}^2$ , such that for all  $t \in \mathbb{Z}^2$ ,  $T_s Y(t) = Y(t + s)$ . Since  $Y$  is strictly stationary,  $T_s$  is measure-preserving



(see Section A.1). Then we have,

$$\begin{aligned}
& P(E(Y(s_1 + h)|Y_{(s_1+h)}) \in A_1, \dots, E(Y(s_k + h)|Y_{(s_k+h)}) \in A_k) \\
&= P(E(T_h Y(s_1)|T_h Y_{(s_1)}) \in A_1, \dots, E(T_h Y(s_k)|T_h Y_{(s_k)}) \in A_k) \\
&= P(T_h E(Y(s_1)|Y_{(s_1)}) \in A_1, \dots, T_h E(Y(s_k)|Y_{(s_k)}) \in A_k) \\
&= P(E(Y(s_1)|Y_{(s_1)}) \in A_1, \dots, E(Y(s_k)|Y_{(s_k)}) \in A_k).
\end{aligned}$$

□

This theorem gives the asymptotic distribution of  $\{\Lambda^N(s) : s \in \mathbb{Z}^2\}$  as  $N \rightarrow \infty$  and its stationarity. Then, for any  $K$ , asymptotic stationarity of  $\{\Lambda^N(s) : s \in \mathbb{Z}_K^2\}$  follows.

**Corollary 1 (Asymptotic stationarity of  $\psi^N$ ).** *Let  $\psi^N(x, y) = (-1)^{x+y}\Lambda^N(x, y)$  be defined on  $\{s \in \mathbb{Z}^2\}$  with  $\Lambda^N(\cdot)$  defined as in Theorem 1. Then  $\psi^N \xrightarrow[N \rightarrow \infty]{a.s.} \psi \in L^1$ , where  $\psi$  is a strictly stationary process.*

This corollary gives the asymptotic distribution of  $\{\psi^N(s) : s \in \mathbb{Z}^2\}$  as  $N \rightarrow \infty$  and its stationarity. Then, for any  $K$ , asymptotic stationarity of  $\{\psi^N(s) : s \in \mathbb{Z}_K^2\}$  follows. In Section 3.3.2.1, we demonstrate that when  $Y(\cdot)$  comes from the Matérn model,  $\psi^N(\cdot)$  is approximately stationary on the interior points of the lattice. Moreover, when  $Y(\cdot)$  comes from the ICAR(1) model, we demonstrate in Section 3.3.2.2 exact stationarity of the subset of  $\psi^N(\cdot)$  that is defined on the interior points of the lattice.

We have shown that  $\Lambda^N(s) \xrightarrow[N \rightarrow \infty]{a.s.} \Lambda(s)$ , for all  $s \in \mathbb{Z}^2$ , where  $\Lambda(\cdot)$  is a stationary Gaussian process. We now state what form the stationary autocovariance

function  $\xi(\cdot)$ , corresponding to  $\Lambda(\cdot)$ , takes. In fact, it turns out that  $\xi(\cdot)$  can be expressed in terms of  $C'$ , the autocovariance function corresponding to  $f^{-1}$ , where  $f$  is the spectral density of  $Y(\cdot)$ . Note that when  $Y(\cdot)$  comes from either the Matérn or the ICAR(1) model, the spectral density  $f$  exists.

Let  $f$  be the spectral density corresponding to  $Y(\cdot)$  and let  $\Sigma_N(f)$  be the stationary covariance matrix corresponding to  $f$ , defined on  $\mathbb{Z}_N^2$ . Let  $\Sigma_N^{-1}(f) \equiv (\sigma_N^{st})_{s,t \in \mathbb{Z}_N^2}$  be the inverse of  $\Sigma_N(f)$ . Generally,  $\Sigma_N^{-1}(f)$  is non-stationary. However, Theorem 2 states that for all  $s, t \in \mathbb{Z}^2$ ,  $|\sigma_N^{st} - C'(s - t)| \xrightarrow{N \rightarrow \infty} 0$ , where  $C'$  is the autocovariance function corresponding to  $f^{-1}$ .

**Theorem 2.** (from [17]) *Let  $\mathcal{F}$  denote the set of admissible spectra for two dimensional random fields, defined as follows. For any spatial autocovariance function  $C(h_1, h_2)$ , consider the sums  $S_{h_1, \cdot} = \sum_{h_2} |h_2| |C(h_1, h_2)|$ ,  $S_{\cdot, h_2} = \sum_{h_1} |h_1| |C(h_1, h_2)|$  and  $S_{\cdot, \cdot} = \sum_{h_1, h_2} |h_1| |h_2| |C(h_1, h_2)|$ . Define the set*

$$\mathcal{F} = \{f : [-\pi, \pi]^2 \rightarrow \mathbb{R}^+, S_{h_1, \cdot} < \infty \forall h_1, S_{\cdot, h_2} < \infty \forall h_2, S_{\cdot, \cdot} < \infty\}.$$

*Let  $\Sigma_N(f)$  be a  $(2N+1)^2 \times (2N+1)^2$  block-Toeplitz covariance matrix associated with spectral density  $f$  and defined on  $\mathbb{Z}_N^2$ . Assume  $f, f^{-1} \in \mathcal{F}$ . Let  $\Sigma_N^{-1} f = (\sigma_N^{st})_{s,t \in \mathbb{Z}_N^2}$  and let  $C'$  be the autocovariance function corresponding to  $f^{-1}$ . Then for all  $s, t \in \mathbb{Z}^2$ , we have*

$$|\sigma_N^{st} - C'(s - t)| \xrightarrow{N \rightarrow \infty} 0.$$

The result of Theorem 2 was confirmed through personal correspondence with Dr. Tucker McElroy. Dr. McElroy confirmed that the result follows as an extension of the proof of Lemma 4.1 of [17]. Moreover, the conditions on spectral densities

required in Theorem 2 hold for Matérn and ICAR(1) models due to the smoothness of  $f$  and  $f^{-1}$  at the origin.

Now, consider the processes  $\Lambda^N(\cdot)$  and  $\Lambda(\cdot)$  defined in Theorem 1. Corollary 2 states that for all  $s, t \in \mathbb{Z}^2$ ,  $\text{cov}(\Lambda^N(s), \Lambda^N(t)) \xrightarrow{N \rightarrow \infty} \text{cov}(\Lambda(s), \Lambda(t))$ . Moreover, Corollary 2 gives the form of the autocovariance function of  $\Lambda(\cdot)$ , in terms of  $C'$ , the autocovariance function corresponding to the inverse spectral density  $f^{-1}$  of the underlying process  $Y(\cdot)$ .

**Corollary 2.** *Let  $\xi(\cdot)$  be the autocovariance function corresponding to  $\Lambda(\cdot)$  in Theorem 1. Then, for all  $h_s \in \mathbb{Z}^2$ ,  $\xi(h_s) = C'(0)^{-1}C'(h_s)$ , where  $C'(\cdot)$  is defined in Theorem 2.*

*Proof.* Let  $\Lambda^N(\cdot)$  be defined as in (3.60). Then, for all  $s, h_s \in \mathbb{Z}^2$ ,

$$\text{cov}(\Lambda^N(s), \Lambda^N(s + h_s)) = (\sigma_N^{ss})^{-1/2} (\sigma_N^{s+h_s, s+h_s})^{-1/2} \sigma_N^{s, s+h_s} \mathbf{1}_{Z_N^2}(s) \mathbf{1}_{Z_N^2}(s + h_s)$$

as in (3.47). Then, Theorem 2 implies that

$$\text{cov}(\Lambda^N(s), \Lambda^N(s + h_s)) \xrightarrow{N \rightarrow \infty} (C'(0))^{-1/2} (C'(0))^{-1/2} C'(h_s) = C'(0)^{-1} C'(h_s).$$

In turn, since  $\Lambda^N(\cdot)$  is Gaussian, Corollary 4, together with the Cauchy-Schwartz inequality imply that  $\text{cov}(\Lambda^N(s), \Lambda^N(s + h_s)) \xrightarrow{N \rightarrow \infty} \xi(h_s) = C'(0)^{-1} C'(h_s)$  for all  $h_s \in \mathbb{Z}^2$ . □

Now, assuming that  $f(\omega, \theta)$  is the spectral density corresponding to  $Y(\cdot)$ ,  $C'(h_s)$  takes the following form:

$$C'(h_s) = \int_{(-\pi, \pi]^2} e^{ih_s^t \omega} f(\omega, \theta)^{-1} d\omega, \text{ for all } h_s \in \mathbb{Z}^2. \quad (3.66)$$

Therefore,  $\xi(h_s)$  in Corollary 2 takes the form:

$$\xi(h_s) = \frac{\int_{(-\pi, \pi]^2} e^{ih_s^t \omega} f(\omega, \theta)^{-1} d\omega}{\int_{(-\pi, \pi]^2} f(\omega, \theta)^{-1} d\omega}, \text{ for all } h_s \in \mathbb{Z}^2. \quad (3.67)$$

When  $Y(\cdot)$  comes from the Matérn model, the spectral density takes the following form, as given in [12, p.71]:

$$f(\omega) = \sigma^2(\phi^{-2} + \omega^2)^{-\nu}, \text{ with } \nu > 0, \phi > 0 \quad (3.68)$$

When  $Y(\cdot)$  comes from the ICAR(1) model, Cressie states in [15, p.449] that the spectral density is proportional to:

$$f(\omega) \propto \frac{1}{1 - 2\rho(\cos(\omega_1) + \cos(\omega_2))} \quad (3.69)$$

The result in Corollary 2 is significant, because it provides a comparison point for the results on magnitudes of correlations in  $\Lambda^N(\cdot)$  given in Section 3.3.2.1. In this section, we make claims that  $\Lambda^N(\cdot)$  is stationary inside the lattice  $\mathbb{Z}_N^2$  and that the magnitudes of correlations of  $\Lambda^N(\cdot)$  inside the lattice are low. We can compute the autocovariance function  $\xi(\cdot)$  of the limiting process  $\Lambda(\cdot)$ , using the formula given in Corollary 2. We can then compare the autocovariances of  $\Lambda(\cdot)$  to the autocovariances of  $\Lambda^N(\cdot)$  inside the lattice  $\mathbb{Z}_N^2$ , where we claim  $\Lambda^N(\cdot)$  is approximately stationary.

The behavior of the Bonferroni bound (see Section 3.3.3.1) depends on the degree of correlation in the process  $\psi^N(\cdot)$  and therefore in the process  $\Lambda^N(\cdot)$ . In fact, the Bonferroni bound accurately approximates the tail probability of  $\max_{s \in \mathbb{Z}_N^2} \psi^N(s)$  when  $\psi^N(\cdot)$  is weakly correlated. In Sections 3.3.2.1 and 3.3.2.2, we consider the correlation structure of  $\Lambda^N(\cdot)$ , when the underlying process  $W(\cdot)$  comes from either

the Matérn or the ICAR(1) model. It is worthwhile to note that when  $W_n$  is independent, then so is  $\Lambda^N(\cdot)$  (and therefore,  $\psi^N(\cdot)$ ). For if  $W_n$  is independent, then  $\Sigma = \sigma^2 I$  and  $\Sigma^{-1} = \sigma^{-2} I$  and  $\xi_{st} = (\sigma^{ss})^{-1/2} (\sigma^{tt})^{-1/2} \sigma^{ts} = \mathbf{1}(s = t)$ .

### 3.3.2.1 Properties of $\Lambda^N(\cdot)$ when $Y(\cdot)$ is from the Matérn model

In this section we point out some relevant features of the correlation structure of  $\Lambda^N(\cdot)$  when the underlying process  $Y(\cdot)$  comes from the Matérn model. The conclusions in this section are based on high accuracy numerical computations.

**Approximate stationarity of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ .** Theorem 1 states that for a fixed  $K$ ,  $\{\Lambda^N(s) : s \in \mathbb{Z}_K^2\}$  becomes stationary as  $N \rightarrow \infty$ . In this section we show that when  $Y(\cdot)$  comes from the Matérn model,  $\Lambda^N(\cdot)$  is approximately stationary on  $\mathbb{Z}_{N-1}^2$ , regardless of the value of  $N$ . To demonstrate this, we compute the maximum, minimum and median correlations in  $\Lambda^N(\cdot)$  at different lags. We do this first over the entire sample space  $\mathbb{Z}_N^2$  and over  $\mathbb{Z}_{N-1}^2$ . More formally, let  $A_K(h) = \{(s, t) \in \mathbb{Z}_K^2 : |s - t| = h\}$  and define

$$M_K^N(h) = \max_{A_K(h)} (\text{cov}(\Lambda^N(s), \Lambda^N(t)))$$

$$m_K^N(h) = \min_{A_K(h)} (\text{cov}(\Lambda^N(s), \Lambda^N(t)))$$

$$\text{med}_K^N(h) = \text{median}_{A_K(h)} (\text{cov}(\Lambda^N(s), \Lambda^N(t)))$$

Generally, we observe that  $M_N^N$ ,  $m_N^N$  and  $\text{med}_N^N$  differ, especially at smaller lags, indicating non-stationarity of  $\Lambda^N(\cdot)$  over  $\mathbb{Z}_N^2$ . However,  $M_{N-1}^N$ ,  $m_{N-1}^N$  and  $\text{med}_{N-1}^N$  are very close to each other, which indicates that  $\Lambda^N(\cdot)$  is approximately stationary on  $\mathbb{Z}_{N-1}^2$ . We observe this trend regardless of the value of  $N$ . Figures 3.5-3.8

illustrate this finding. Notice that  $med_{N-1}^N$  serves as a proxy for the stationary autocorrelation function of  $\Lambda^N(\cdot)$  inside the lattice, whereas  $M_N^N$  and  $m_N^N$  reflect correlations on the boundary  $\mathcal{B}_N$ .

In the paragraphs to follow, we illustrate the above described finding in more detail. To simplify presentation, we draw attention to the following. The degree of correlation in  $Y(\cdot)$  is positively related to the magnitudes of correlation in  $\Lambda^N(\cdot)$ . In turn, higher magnitudes of correlation in  $\Lambda^N(\cdot)$  result in greater differences between  $M_K^N$ ,  $m_K^N$  and  $med_K^N$ , for a fixed  $N$  and  $K = N$  or  $K = N - 1$ . Therefore, it suffices to convince the reader that  $M_{N-1}^N$ ,  $m_{N-1}^N$  and  $med_{N-1}^N$  are approximately the same for those  $\Lambda^N(\cdot)$  which correspond to highly correlated  $Y(\cdot)$ . This is because weakly correlated  $\Lambda^N(\cdot)$  will naturally exhibit less variation in correlations and should converge to stationarity quicker than higher correlated  $\Lambda^N(\cdot)$ . To illustrate this point, we consider Matérn autocorrelation functions of various smoothness, as plotted in Figure 3.1. Figures 3.2, 3.3 and 3.4 present the correlations in the corresponding  $\Lambda^N(\cdot)$ , for varying sample sizes ( $N = 3, 11, 25$ ). The observation to make is that a higher correlated  $Y(\cdot)$  corresponds to higher magnitudes in correlation of  $\Lambda^N(\cdot)$ , regardless of  $N$ . In fact, Figures 3.2, 3.3 and 3.4 are nearly identical, indicating that correlations in  $\Lambda^N(\cdot)$  do not vary much with  $N$ . Secondly, the higher the magnitudes of correlation in  $\Lambda^N(\cdot)$ , the more difference there is between correlations on the boundary (reflected in the  $m_N^N$  and  $M_N^N$  numbers) and inside the boundary (reflected in the  $med_N^N$  numbers). Thus if we show that  $\Lambda^N(\cdot)$  corresponding to a highly correlated  $Y(\cdot)$  is approximately stationary on  $\mathbb{Z}_{N-1}^2$ , we can conclude the same about  $\Lambda^N(\cdot)$  corresponding to a weakly correlated  $Y(\cdot)$ .

Now, we are ready to illustrate our main finding, which is that  $\Lambda^N(\cdot)$  is approximately stationary on  $\mathbb{Z}_{N-1}^2$ . Consider Figures 3.5-3.8. These are plots of  $M_N^N$ ,  $m_N^N$  and  $med_N^N$  (top) and  $M_{N-1}^N$ ,  $m_{N-1}^N$  and  $med_{N-1}^N$  (bottom). In these plots,  $\Lambda^N(\cdot)$  corresponds to a highly correlated process  $Y(\cdot)$ , modeled by  $\nu = 1$  and  $\phi = 2$ . The six figures correspond to three different sample sizes, with  $N = (3, 4, 5, 11)$ . In red, are the median correlations. Observe that as we compare Figures 3.5-3.8 the only difference between the median correlations we observe is when passing from  $N = 3$  to  $N = 4$ . Even then, the difference is very marginal. For  $N \geq 4$ , the median correlations do not change with  $N$  and are the same on  $\mathbb{Z}_N^2$  and  $\mathbb{Z}_{N-1}^2$ .

We observe that on  $\mathbb{Z}_N^2$ , there is variation in correlations of  $\Lambda^N(\cdot)$  at fixed lags, thus indicating non-stationarity of  $\Lambda^N(\cdot)$ . However, on  $\mathbb{Z}_{N-1}^2$ , the variation in correlations of  $\Lambda^N(\cdot)$  is very small, thus indicating approximate stationarity of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ .

Figures 3.9 and 3.10 serve to quantify the variation in  $\text{cov}(\Lambda(s), \Lambda(s+h))$  over  $s \in \mathbb{Z}_N^2$  and  $s \in \mathbb{Z}_{N-1}^2$  for various  $h$ . From Figure 3.9, we see that the variation in  $\text{cov}(\Lambda(s), \Lambda(s+h))$  over  $s \in \mathbb{Z}_N^2$  is considerable for  $h = 1$  and  $h = 1.41$  and that it does not change for  $N \geq 4$ . The variation in  $\text{cov}(\Lambda(s), \Lambda(s+h))$  over  $s \in \mathbb{Z}_N^2$  is small for  $h \geq 2$ . From Figure 3.10, we see that the variation in  $\text{cov}(\Lambda(s), \Lambda(s+h))$  over  $s \in \mathbb{Z}_{N-1}^2$  is small for all  $h$  and does not vary for  $N \geq 3$ .

**Magnitude of correlations in  $\Lambda^N(\cdot)$ .** In this section, we would like to summarize our findings regarding the magnitudes of correlations in  $\Lambda^N(\cdot)$ . In the previous section, we showed that  $\Lambda^N(\cdot) = \{\Lambda^N(s) : s \in \mathbb{Z}_N^2\}$  is approximately stationary on  $\mathbb{Z}_{N-1}^2$ , regardless of the value of  $N$ . We also showed that the median

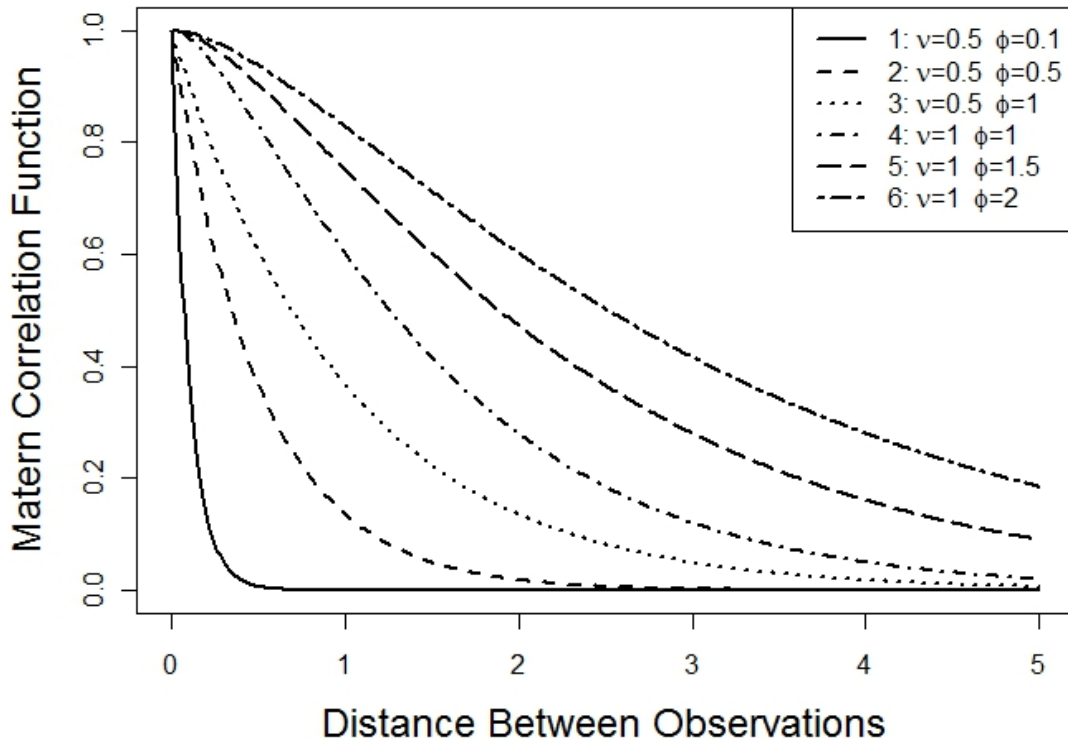


Figure 3.1: Matérn Autocorrelation Function, varying  $\nu$  and  $\phi$

correlations in  $\Lambda^N(\cdot)$  do not change for  $N \geq 4$ . These median correlations are equivalent to the stationary autocorrelation function of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ . Let  $h_s = (h_x, h_y)$  be an increment in  $s$  along each axis and let  $h = \sqrt{h_x^2 + h_y^2}$ . Let  $L(h)$  be the autocovariance function of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ , where  $\Lambda^N(\cdot)$  has been shown to be approximately stationary. Figure 3.11 plots  $L(h)$  for the six models for  $Y(\cdot)$  in Figure 3.1 and for  $N = (3, 11, 25)$ . As we discussed in the previous section,  $L(h)$  stabilizes at  $N \geq 4$ . For  $N \geq 4$ , for the six models we consider for  $Y(\cdot)$ ,  $-0.35 \leq L(1) \leq 0$ ,  $-0.06 \leq L(\sqrt{2}) \leq 0.01$ ,  $0 \leq L(2) \leq 0.08$ , with  $|L(h)| \leq 0.02$  for



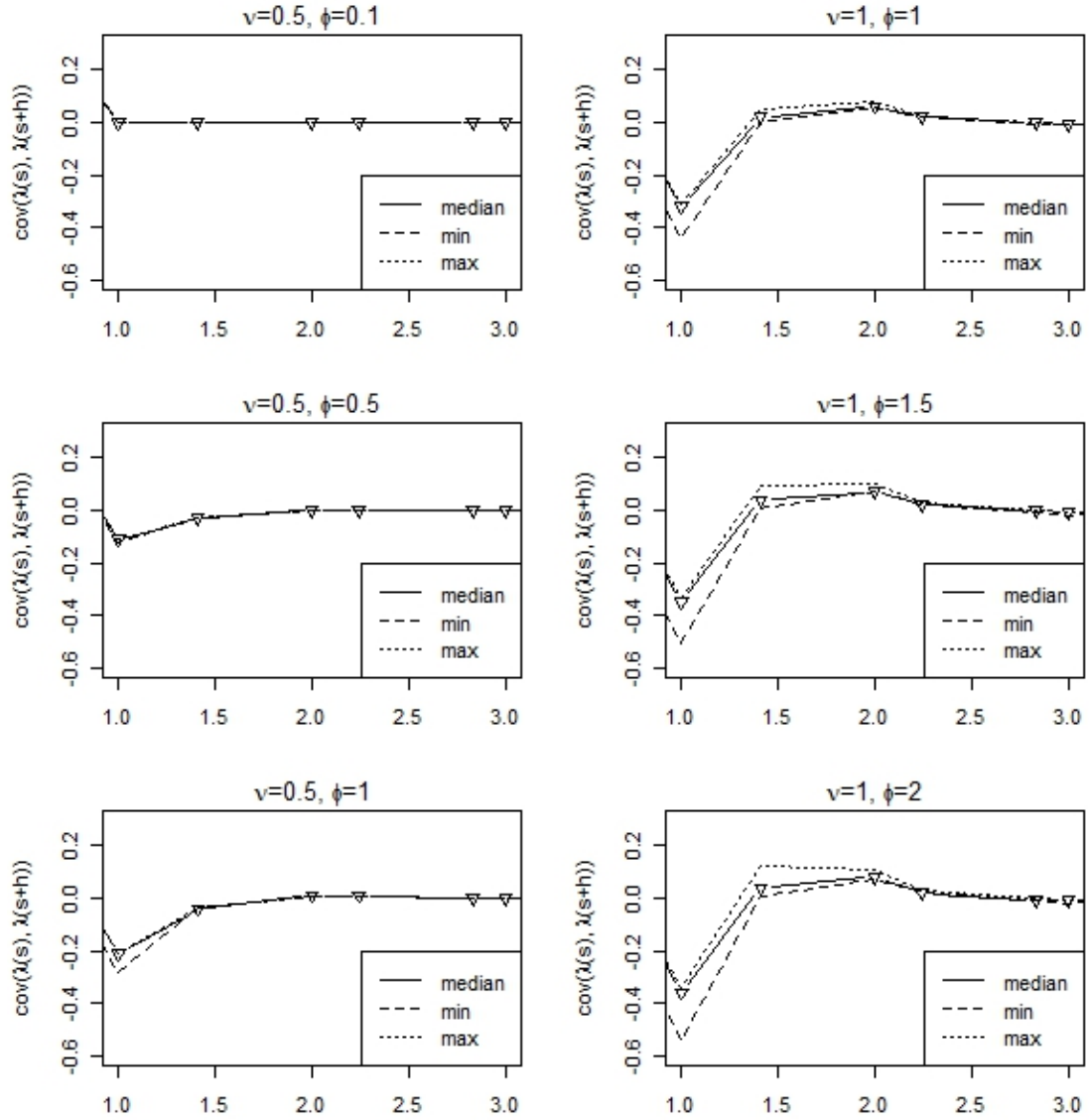


Figure 3.2: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$  on  $\mathbb{Z}_N^2$ , varying  $\nu$  and  $\phi$ ,  $N=3$ .

$h \geq 2$ . Another way of summarizing this is to state the following.  $L(h)$  is highest in magnitude and is negative at  $h = 1$ . At lags greater than 1, the magnitude of  $L(h)$  is small.

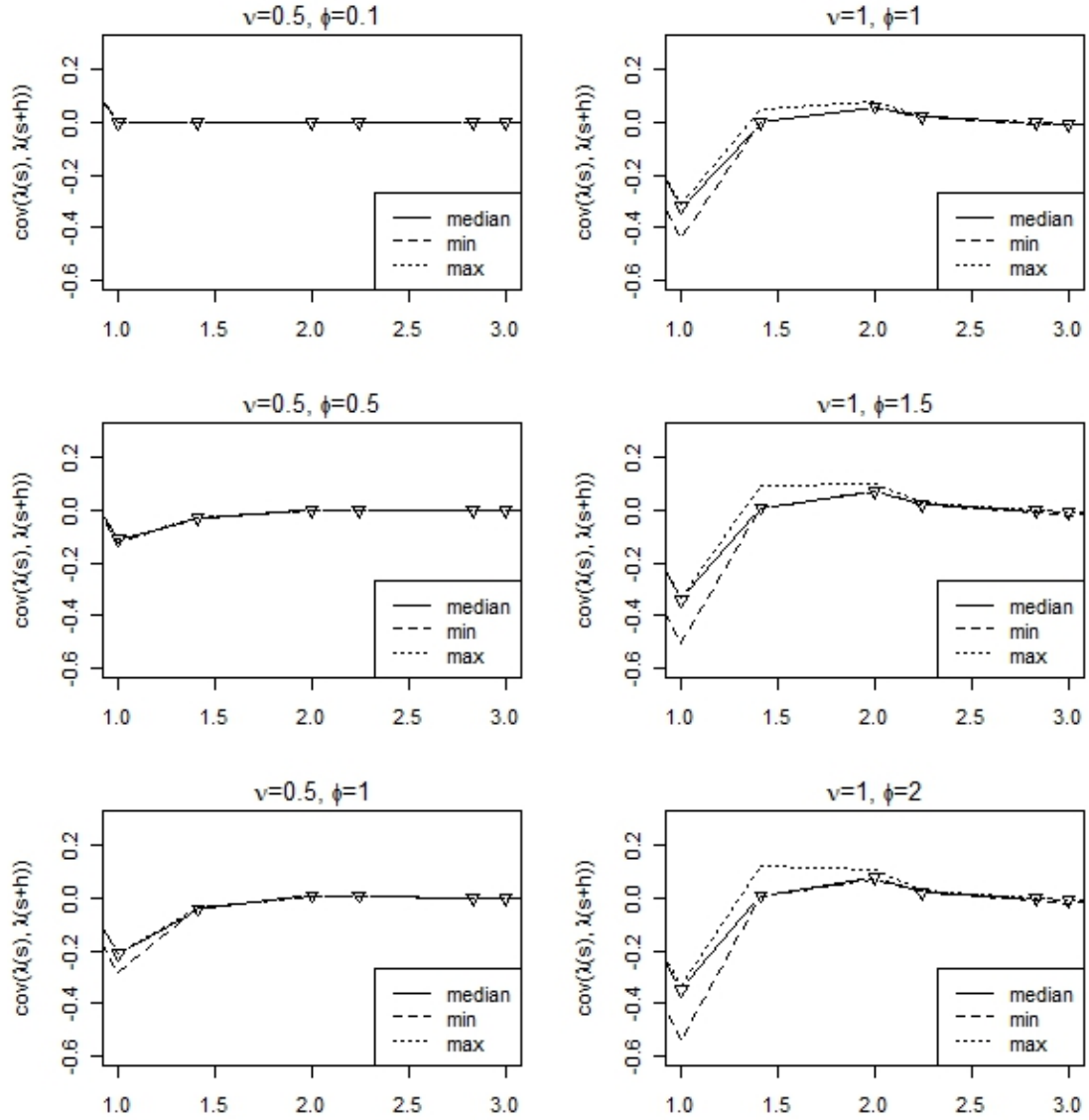


Figure 3.3: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$  on  $\mathbb{Z}_N^2$ , varying  $\nu$  and  $\phi$ ,  $N=11$ .

### 3.3.2.2 Properties of $\Lambda^N(\cdot)$ when $Y(\cdot)$ is from the ICAR(1) model

#### 1. Stationarity and correlation structure of $\Lambda^N(\cdot)$ on $\mathbb{Z}_{N-1}^2$

In this section we demonstrate stationarity of the outlier detection vector  $\Lambda^N(\cdot)$

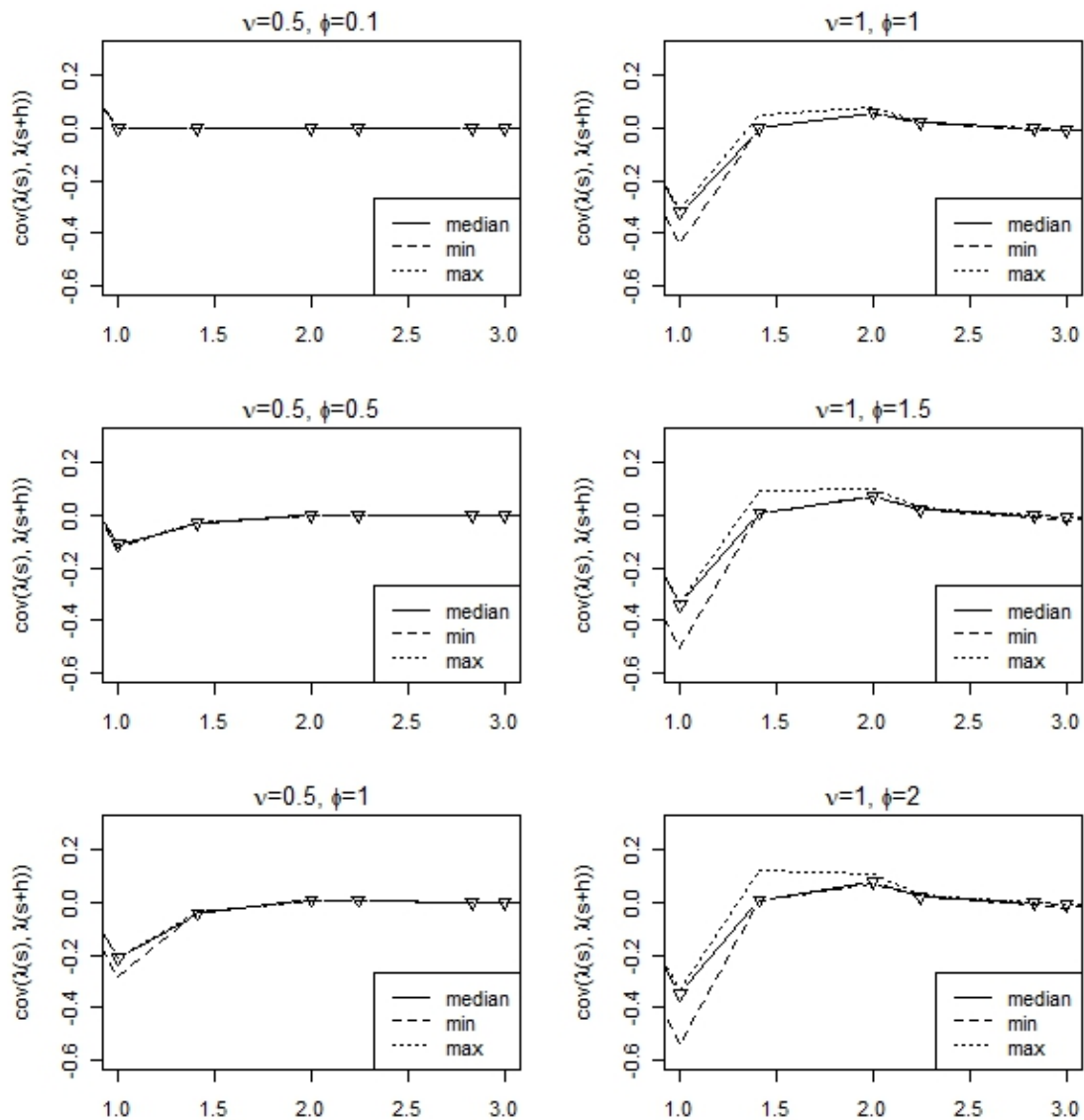


Figure 3.4: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$  on  $\mathbb{Z}_N^2$ , varying  $\nu$  and  $\phi$ ,  $N=25$ .

on the subset  $\mathbb{Z}_{N-1}^2$  of the sample space  $\mathbb{Z}_N^2$ , when the underlying process  $Y(\cdot)$  comes from the ICAR(1) model. We also give a simplified form for the autocovariance function  $L_\theta(\cdot)$  of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ .

**Proposition 4** (Distribution of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ ). *Let  $\{W(s) : s \in \mathbb{Z}^2\}$  be*

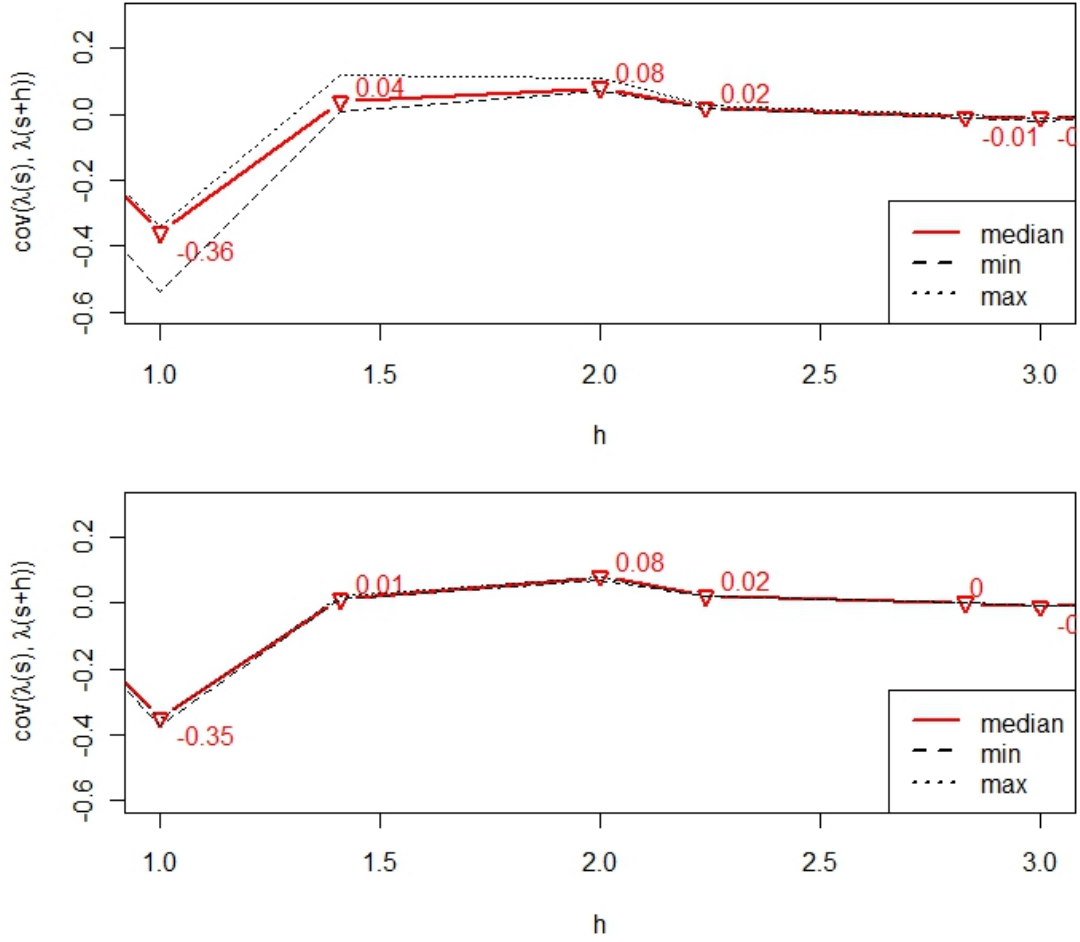


Figure 3.5: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$ ;  $\nu = 1, \phi = 2$ ; on  $\mathbb{Z}_N^2$  (top),  $\mathbb{Z}_{N-1}^2$  (bottom),  $N = 3$ ,  $h$  is the Euclidean distance between observations.

*distributed as in Definition 2 and consider a sample  $W_n \equiv \{W(s) : s \in \mathbb{Z}_N^2\}$ . Consider the outlier detection vector  $\Lambda^N(s)$ , defined as in (3.44). Then  $\Lambda^N(\cdot)$  is stationary on  $\mathbb{Z}_{N-1}^2$ . Furthermore, let  $L_\theta(\cdot)$  be the autocovariance function of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ . Then,*

$$L_\theta(h) = \mathbf{1}(h = 0) - \rho \mathbf{1}(h = 1) \quad (3.70)$$

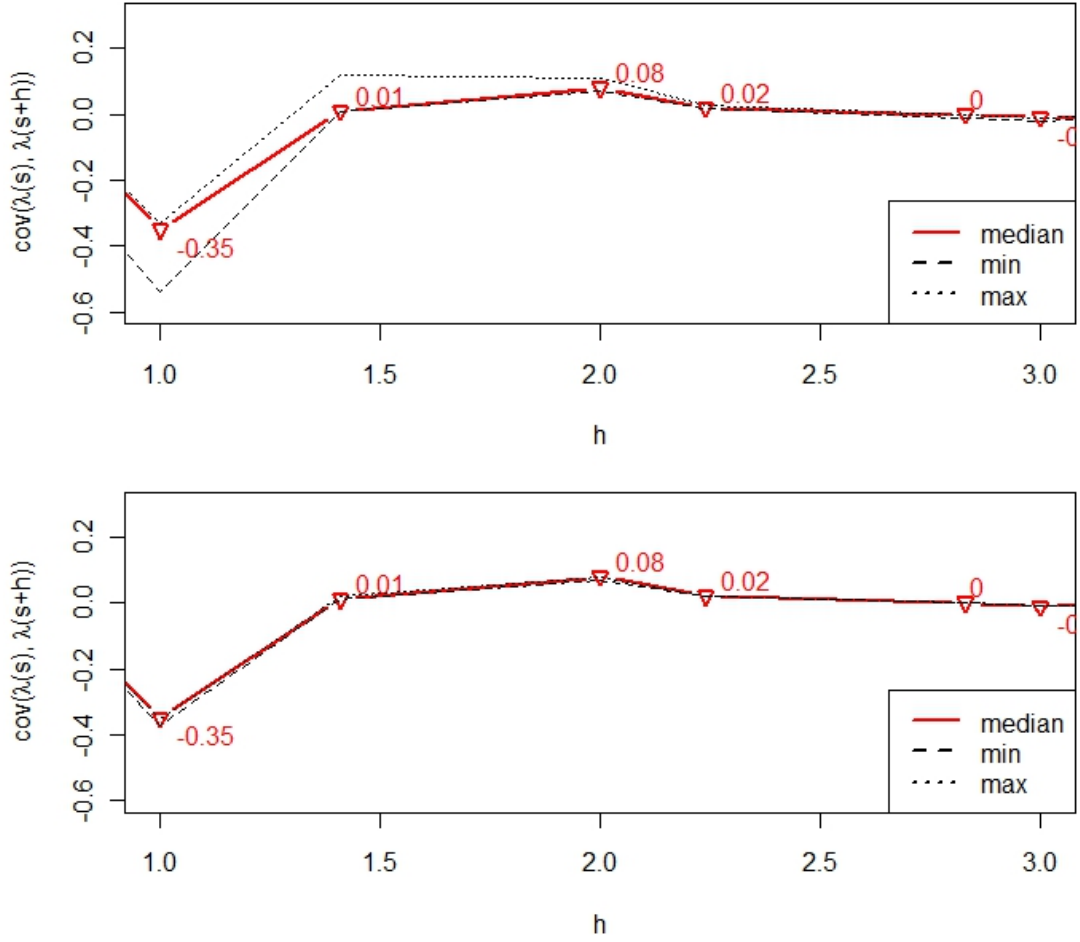


Figure 3.6: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$ ;  $\nu = 1, \phi = 2$ ; on  $\mathbb{Z}_N^2$  (top),  $\mathbb{Z}_{N-1}^2$  (bottom),  $N = 4$ ,  $h$  is the Euclidean distance between observations.

*Proof.* According to Proposition 2,  $\text{cov}(\Lambda^N(s), \Lambda^N(t)) = (\sigma_N^{ss})^{-1/2}(\sigma_N^{tt})^{-1/2}\sigma_N^{st}$ , where  $\sigma_N^{st}$  are elements of  $\Sigma_N^{-1}$ , the precision matrix of  $W_n$ . Now,  $\Sigma_N^{-1} = I - \rho M_N$ , as in (2.14). Let  $m_{st}^N$  be elements of  $M_N$ . Then, according to (2.15), for all  $s, t \in \mathbb{Z}_{N-1}^2$ ,

$$m_{st}^N = \mathbf{1}(|s - t| = 1) \quad (3.71)$$

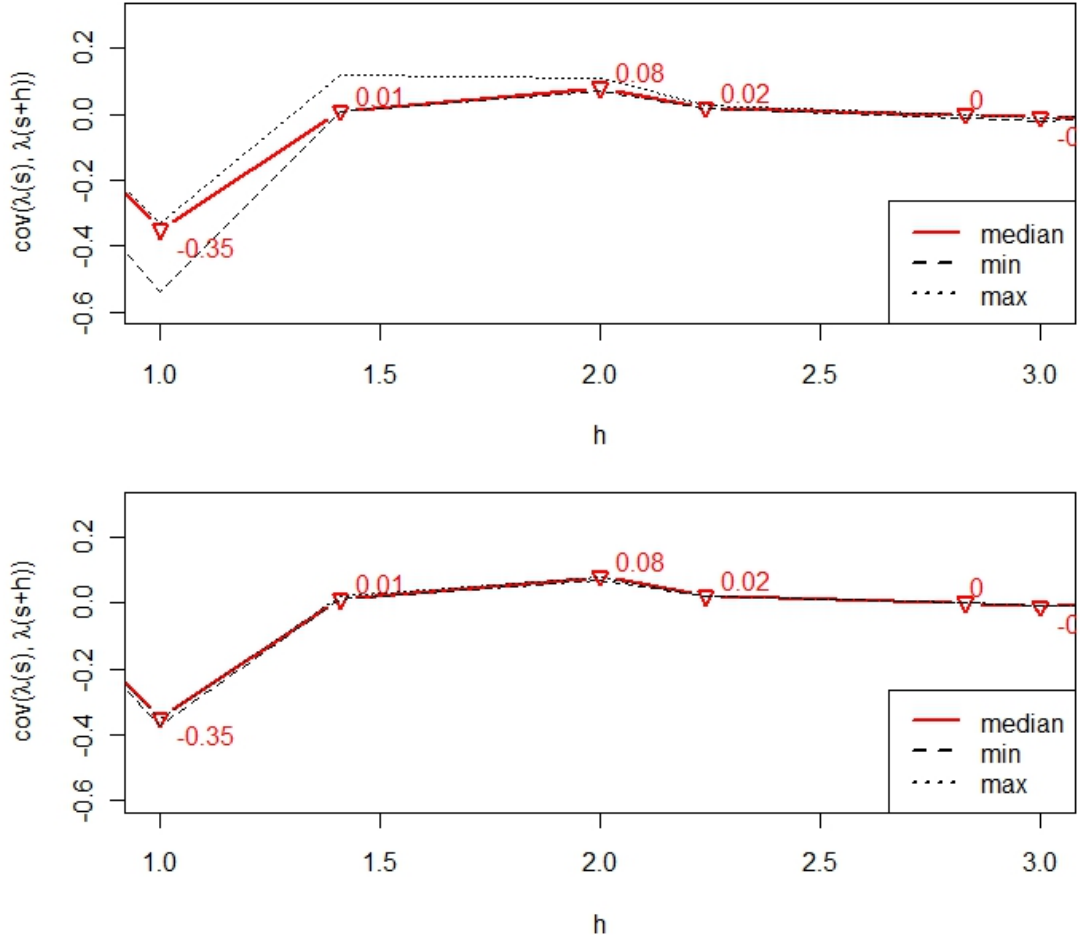


Figure 3.7: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$ ;  $\nu = 1, \phi = 2$ ; on  $\mathbb{Z}_N^2$  (top),  $\mathbb{Z}_{N-1}^2$  (bottom),  $N = 5$ ,  $h$  is the Euclidean distance between observations.

Therefore, for all  $s, t \in \mathbb{Z}_{N-1}^2$ ,

$$\sigma_N^{st} = \mathbf{1}(|s - t| = 0) - \rho \mathbf{1}(|s - t| = 1) \quad (3.72)$$

Since for all  $s \in \mathbb{Z}_{N-1}^2$ ,  $\sigma_N^{ss} = 1$ , we have for all  $s, t \in \mathbb{Z}_{N-1}^2$ ,  $\text{cov}(\Lambda^N(s), \Lambda^N(t)) = \sigma_N^{st}$ . This, in turn, implies that for all  $s, t \in \mathbb{Z}_{N-1}^2$ ,

$$\text{cov}(\Lambda^N(s), \Lambda^N(t)) = \mathbf{1}(|s - t| = 0) - \rho \mathbf{1}(|s - t| = 1) \quad (3.73)$$

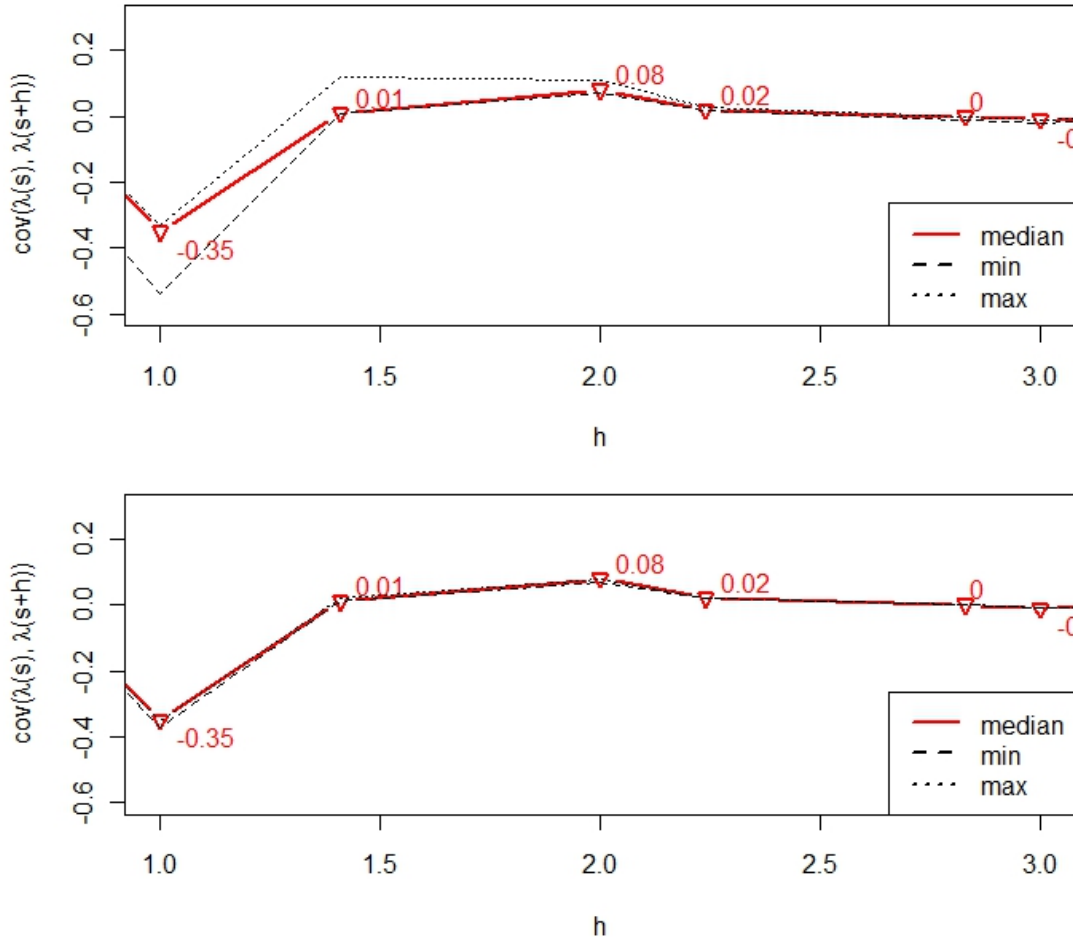


Figure 3.8: Maximum, minimum and median correlations between  $\Lambda(s)$  and  $\Lambda(s+h)$ ;  $\nu = 1, \phi = 2$ ; on  $\mathbb{Z}_N^2$  (top),  $\mathbb{Z}_{N-1}^2$  (bottom),  $N = 11$ ,  $h$  is the Euclidean distance between observations.

This proves stationarity of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ . Let  $L_\theta(\cdot)$  be the autocovariance function of  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$ . Then (3.73) implies that

$$L_\theta(h) = \mathbf{1}(h = 0) - \rho \mathbf{1}(h = 1) \quad \square$$

We have shown that  $\Lambda^N(\cdot)$  is stationary on  $\mathbb{Z}_N^2 \setminus \mathcal{B} \equiv \mathbb{Z}_{N-1}^2$  when  $Y(\cdot)$  is modeled as ICAR(1). This is important, since the *DLM* bound (see Sec-

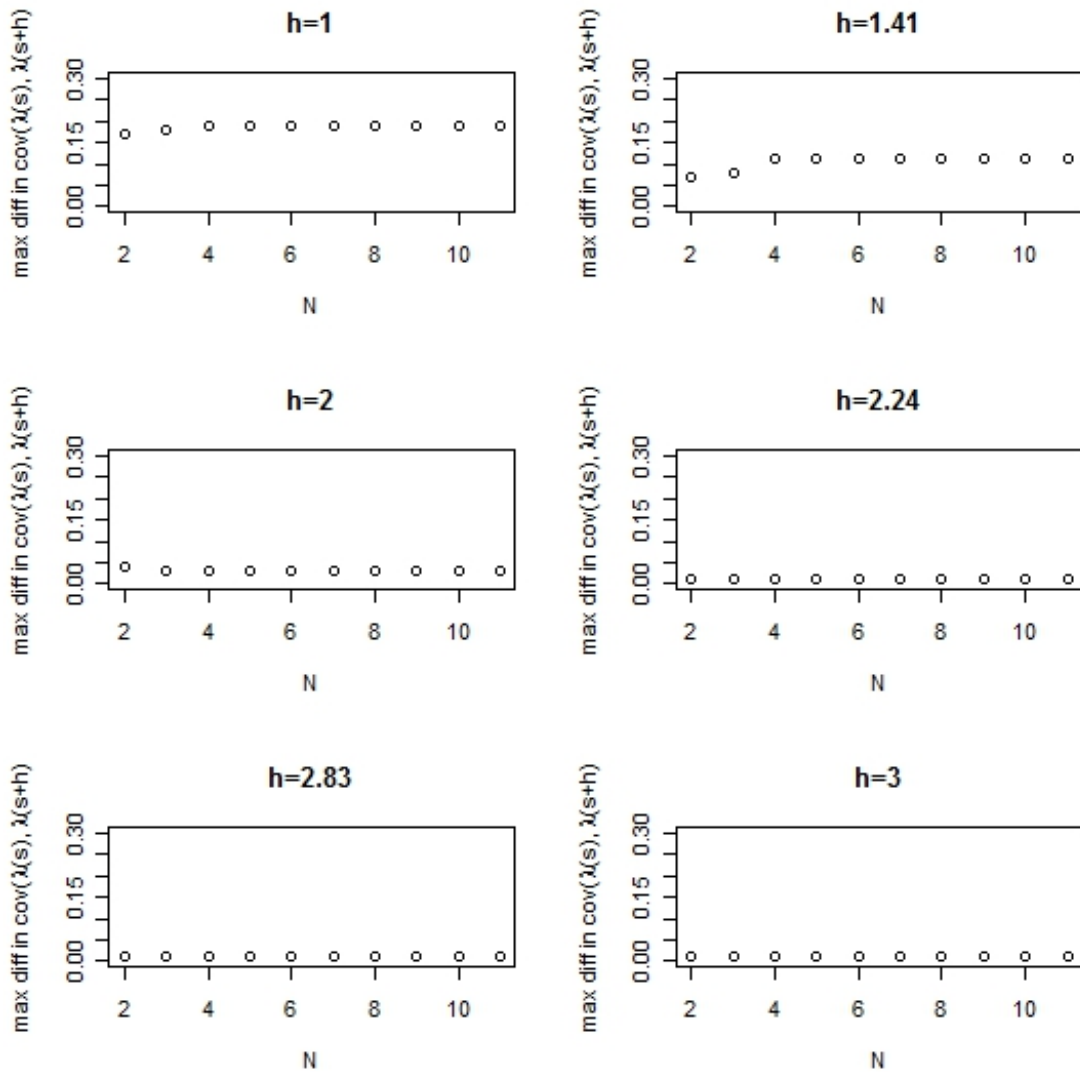


Figure 3.9: Maximum absolute differences over  $s$  of  $\text{cov}(\Lambda(s), \Lambda(s+h))$ , for a fixed  $h$ . The max is taken over  $\mathbb{Z}_N^2$ , varying  $N$ .

tion 3.3.3.2) requires stationarity of  $\Lambda^N(\cdot)$ . Furthermore, since  $|\rho| < 1/4$ , pairwise correlations in  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$  are at most  $1/4$  in magnitude. Therefore,  $\Lambda^N(\cdot)$  and  $\psi^N(\cdot)$  are weakly correlated on  $\mathbb{Z}_{N-1}^2$ . This is important, because for weakly correlated processes, the Bonferroni bound is accurate (see



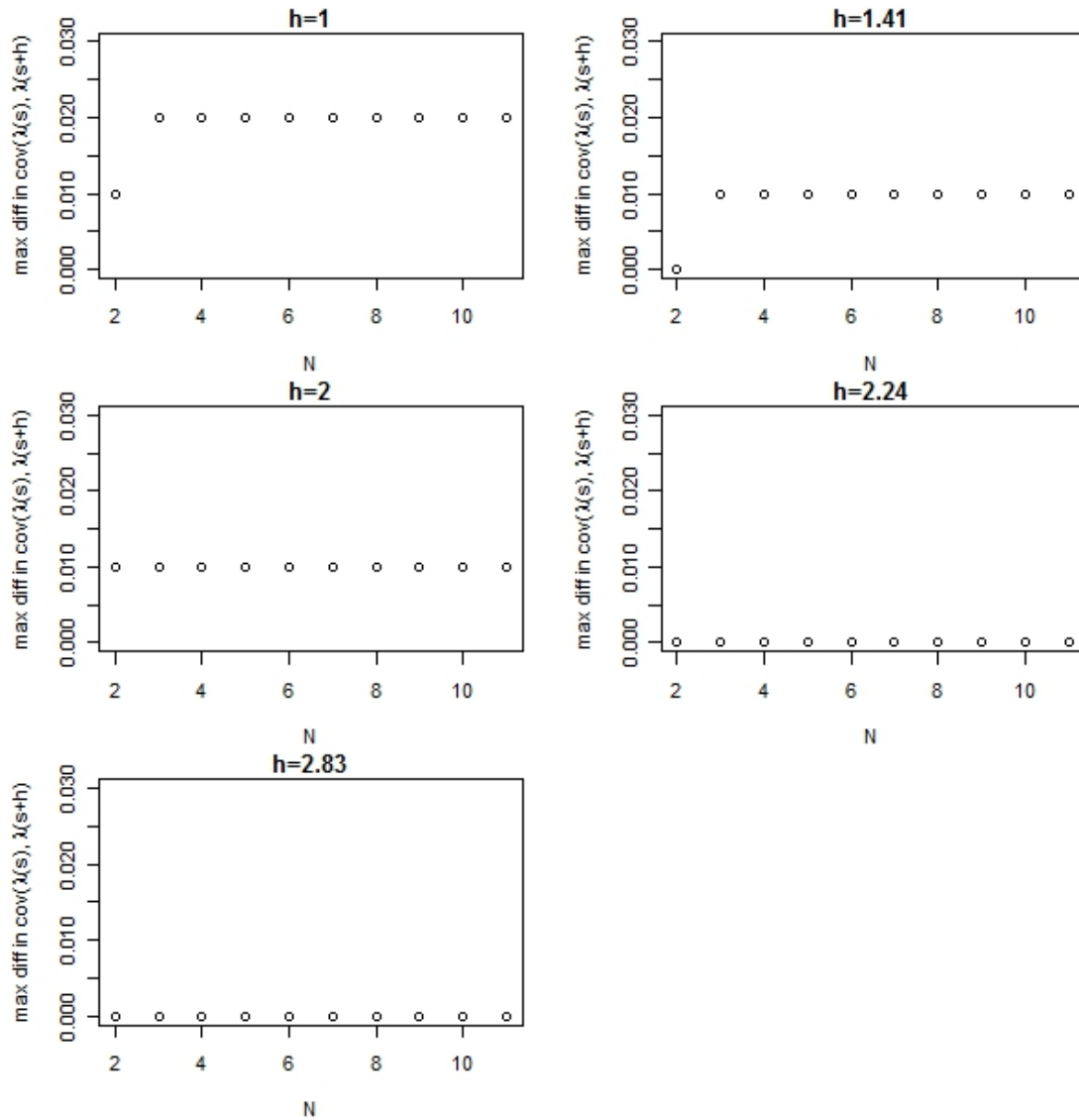


Figure 3.10: Maximum absolute differences over  $s$  of  $\text{cov}(\Lambda(s), \Lambda(s+h))$ , for a fixed  $h$ . The max is taken over  $\mathbb{Z}_{N-1}^2$ , varying  $N$ .

Section 3.3.3.1). Since the Bonferroni bound does not require the knowledge of the correlation structure of  $Y(\cdot)$  and is extremely easy to calculate, cases where it performs well should be pointed out.

*Circulant approximation to  $\Sigma_N^{-1}$*

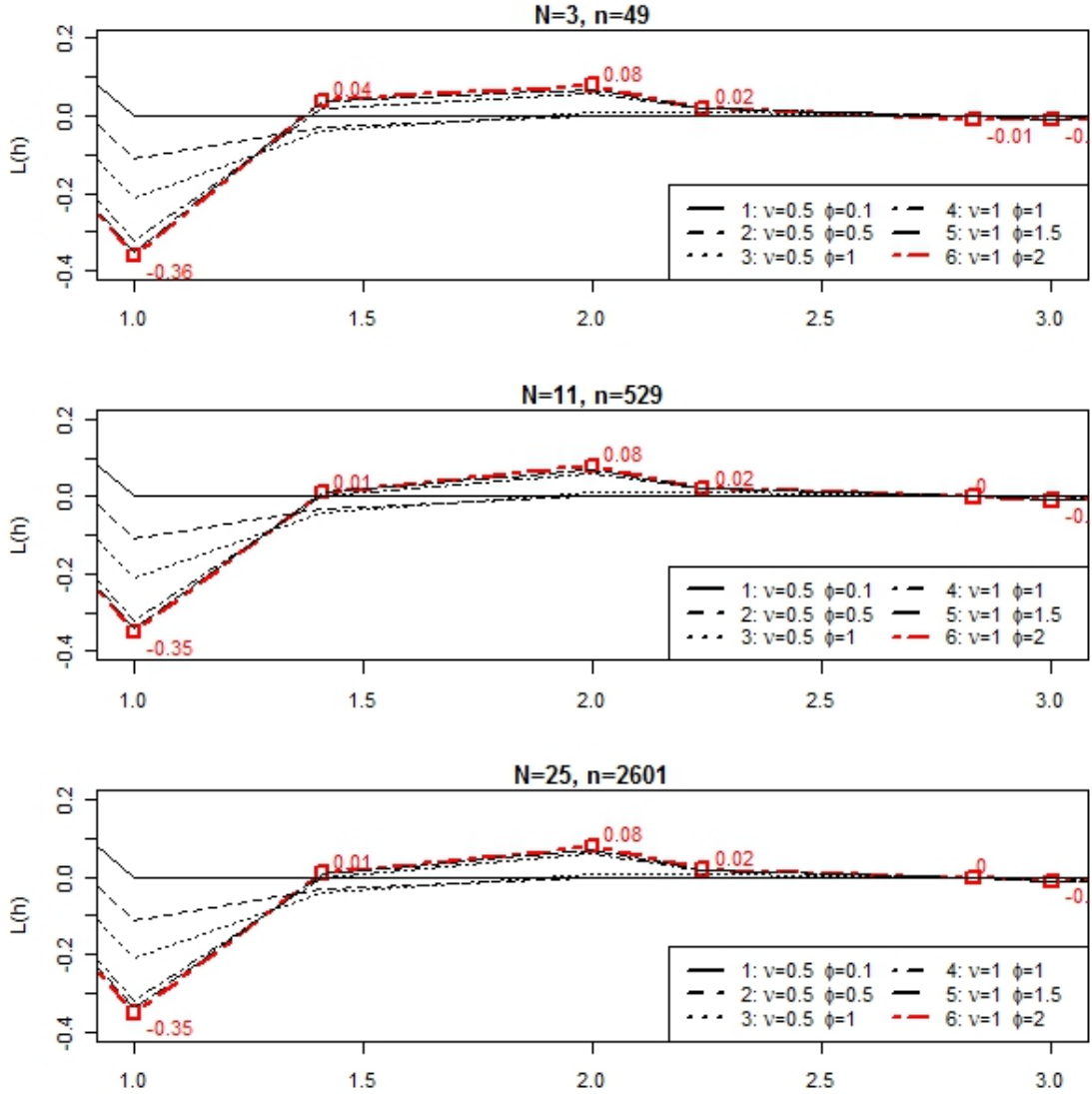


Figure 3.11:  $L(h)$  as a function of lag  $h$ . The figures reported on the plots are the values of  $L(h)$  corresponding to the highest correlated  $Y(\cdot)$ , with  $\nu = 1$  and  $\phi = 2$ .

To simulate  $\Lambda^N(\cdot)$ , we first need to compute the corresponding covariance matrix  $\Xi_N$ . This, in turn, requires knowledge of  $\Sigma_N^{-1}$ . To compute  $\Sigma_N^{-1}$ , we first compute the entries in  $\Sigma_N$  via (2.12) and then invert  $\Sigma_N$ . When  $N$  is large, the computation of  $\Sigma_N$  is very time consuming. Instead, we propose

replacing  $\Sigma_N^{-1}$  with its circulant approximation  $C_N$ , obtained by treating the  $\mathbb{Z}_N^2$  lattice as a torus with opposite sides adjacent. In this case, all sites will have four neighbors and the specification of  $C_N$  is trivial and results in a sparse matrix. Below, we give a formal definition of  $C_N$ . First, we need to define  $\tilde{\mathcal{N}}(\cdot)$ , the set of nearest neighbors on the torus of  $\mathbb{Z}_N^2$ .

**Definition 3.** Consider the lattice  $\mathbb{Z}_N^2$  wrapped on the torus, with opposite sides of the lattice adjacent. Let  $s = (x, y) \in \mathbb{Z}_N^2$  and let  $\tilde{\mathcal{N}}(s)$  be the set of nearest neighbors of  $s$  on the  $\mathbb{Z}_N^2$  torus. Let  $\mathcal{B} \equiv \mathbb{Z}_N^2 \setminus \mathbb{Z}_{N-1}^2$  be the set of boundary points of  $\mathbb{Z}_N^2$ . Then,

$$\tilde{\mathcal{N}}(s) = \begin{cases} \mathcal{N}(s) & s \notin \mathcal{B} \\ \{\mathcal{N}(s) \cap \mathbb{Z}_N^2\} \cup (-x, y) & s \in \mathcal{B}, x = \pm N \\ \{\mathcal{N}(s) \cap \mathbb{Z}_N^2\} \cup (x, -y) & s \in \mathcal{B}, y = \pm N \end{cases} \quad (3.75)$$

where  $\mathcal{N}(\cdot)$  is defined as in Definition 1.

Now, we define the circulant approximation  $C_N$  to  $\Sigma_N^{-1}$ , formed by treating the opposite sides of the lattice  $\mathbb{Z}_N^2$  as adjacent.

**Definition 4.** Let  $W(\cdot)$  be defined as in Definition 2, which implies that a sample  $W_n \equiv \{W(t) : t \in \mathbb{Z}_N^2\}$  is distributed as in (2.13) with the precision matrix  $\Sigma_N^{-1} \equiv (\sigma_N^{st})_{s,t \in \mathbb{Z}_N^2}$  taking the form (2.14). Let  $C_N \equiv (c_N^{st})_{s,t \in \mathbb{Z}_N^2}$  be the circulant approximation of  $\Sigma_n^{-1}$ , obtained by treating the opposite sides of the lattice  $\mathbb{Z}_N^2$  as adjacent. Then  $C_N = I - \rho \tilde{M}_N$  with  $\tilde{M}_N \equiv (\tilde{m}_N^{st})_{s,t \in \mathbb{Z}_N^2}$  where

for all  $s, t \in \mathbb{Z}_N^2$ ,

$$\tilde{m}_N^{st} = \mathbf{1}(t \in \tilde{\mathcal{N}}(s)) \quad (3.76)$$

Notice that  $\Sigma_N^{-1}$  and  $C_N$  are equivalent on  $\mathbb{Z}_{N-1}^2$ , as follows from Definition 4. Let  $\tilde{\Lambda}^N(s)$  be the approximation to the test statistic  $\Lambda^N(s)$ , obtained by replacing  $\Sigma_N^{-1}$  with  $C_N$ . Then it follows easily that  $\Lambda^N(s)$  and  $\tilde{\Lambda}^N(s)$  are equivalent on  $\mathbb{Z}_{N-1}^2$ .  $\Lambda^N(s)$  and  $\tilde{\Lambda}^N(s)$  differ on  $s \in \mathcal{B}_N$ . As  $N$  increases, the effect of the boundary in many problems diminishes. For example, for us it is of interest to study the tail probabilities of  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$ . As  $N \rightarrow \infty$ , the proportion of observations on the boundary  $\mathcal{B}_N$  shrinks and so does the impact of  $\{\Lambda^N(s) : s \in \mathcal{B}_N\}$  on  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$ . Thus, if we want to study the behavior of  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$  for large  $N$ , the use of the approximation  $\tilde{\Lambda}^N(s)$  in place of  $\Lambda^N(s)$  is justified. However, when  $N$  is small, the percentage of the observations on the boundary  $\mathcal{B}_N$  is large and thus replacing  $\Lambda^N(s)$  with  $\tilde{\Lambda}^N(s)$  is questionable.

### 3.3.3 Bounds on the tail probability of the maximum of a discretely sampled Gaussian random field

Consider  $Z_n \equiv \{Z(s) : s \in \mathbb{Z}_N^2\}$ , a stationary, zero-mean, unit-variance Gaussian random field, with autocovariance function  $C_\theta(\cdot)$ . Let

$$P_{max} = P(\max_{s \in \mathbb{Z}_N^2} Z(s) > t) \quad (3.77)$$

be the tail probability of the maximum of  $Z_n$ . In this section, we present two bounds on  $P_{max}$ , the Bonferroni bound and the Discrete Local Maxima (DLM) bound. The Bonferroni bound is attractive because it is very easy to calculate and it does not depend on  $\theta$ . However, its drawback is that it performs well only when the process  $Z_n$  is weakly correlated. The DLM bound, on the other hand, performs well regardless of the degree of correlation in  $Z_n$ . However, it does depend on  $\theta$  and is more difficult to calculate than the Bonferroni bound.

### 3.3.3.1 Bonferroni bound ( $P_{BON}$ )

Boole's Inequality, also known as the *union bound*, says that for any finite or countable set of events, the probability that at least one of the events happens is no greater than the sum of the probabilities of individual events.

**Proposition 5 (Boole's Inequality).** *Let  $\{A_i\}$  be a countable set of events.*

*Then*

$$P(\cup_i A_i) \leq P(\sum_i A_i) \tag{3.78}$$

Boole's Inequality is a special case of the Bonferroni Inequalities. We refer to  $P(\sum_i A_i)$  as  $P_{BON}$ , the Bonferroni upper bound on  $P(\cup_i A_i)$ .

**Corollary 3.** *Let  $Z_n \equiv \{Z(s) : s \in \mathbb{Z}_N^2\}$  be a Gaussian zero-mean, unit-variance random field. Let,*

$$P_{BON} = \sum_{s \in \mathbb{Z}_N^2} P(Z(s) > t) \tag{3.79}$$

*Let  $P_{max}$  be defined as in (3.77). Then  $P_{max} \leq P_{BON}$ .*

*Proof.* The result follows if we note that

$$P\{\max_{s \in \mathbb{Z}_N^2} Z(s) > t\} = P\{\cup_{s \in \mathbb{Z}_N^2} (Z(s) > t)\}.$$

□

Corollary 3 shows that  $P_{BON}$  is an upper bound on  $P_{max}$ . Theorem 3 implies that  $P_{BON} - P_{max} \rightarrow 0$  as  $t \rightarrow \infty$ .

**Theorem 3** (from [16]). *Let  $\{Z(s)\}_{s \in S}$  be a mean-zero stationary Gaussian process defined on a finite set  $S \subset \mathbb{Z}^2$ . Let  $P_{BON}$  be defined as in (3.79). Let  $K(h)$  be the autocorrelation function of  $Z(s)$ , as a function of the lag  $h$ . Then,*

$$\liminf_{t \rightarrow \infty} \left\{ -\frac{2}{t^2} \log(P_{BON} - P_{max}) \right\} = 1 + \frac{1}{\sigma_{c,BON}^2} \quad (3.80)$$

where the critical variance  $\sigma_{c,BON}^2$  is defined as

$$\sigma_{c,BON}^2 = \frac{1 + K(1)}{1 - K(1)} \quad (3.81)$$

### 3.3.3.2 Discrete Local Maxima (DLM) bound ( $P_{DLM}$ )

Taylor et al. [16] present an upper bound on  $P_{max}$ , which they refer to as an improved Bonferroni-type bound based on discrete local maxima. The attractive feature of this bound is that, unlike the Bonferroni bound, it works well regardless of the degree of correlation in the process  $Z_n$ .

**Definition 5 (Discrete Local Maxima (DLM) bound, from [16]).** *Let  $P_{DLM}$  be the expected number of discrete local maxima of  $Z_n$  on  $\mathbb{Z}_N^2$ , defined*

as

$$P_{DLM} = \sum_{u \in \mathbb{Z}_N^2} P(Z(u) > t, Z(v) < Z(u), v \in \mathcal{N}(u) \cap \mathbb{Z}_N^2) \quad (3.82)$$

where  $\mathcal{N}(u)$  is defined as in Definition 1.

Theorems 4 and 5, from [16], state that  $P_{DLM}$  is an upper bound on  $P_{max}$  and that  $P_{DLM} - P_{max} \rightarrow 0$  as  $t \rightarrow \infty$ .

**Theorem 4.** *Let  $P_{max}$  be defined as in (3.77). Then  $P_{max} \leq P_{DLM}$ .*

**Theorem 5.** *Let  $Z_n \equiv \{Z(s) : s \in \mathbb{Z}_N^2\}$  be a stationary, zero-mean, unit-variance Gaussian random field. Let  $P_{DLM}$  be defined as in Definition 5. Let  $K_\theta(\cdot)$  be the autocorrelation function of  $Z_n$ . Then,*

$$\liminf_{t \rightarrow \infty} \left\{ -\frac{2}{t^2} \log(P_{DLM} - P_{max}) \right\} = 1 + \frac{1}{\sigma_{c,DLM}^2} \quad (3.83)$$

where the critical variance  $\sigma_{c,DLM}^2$  is defined as

$$\sigma_{c,DLM}^2 = \frac{1 + K(\sqrt{2})}{1 - K(\sqrt{2})} \quad (3.84)$$

Notice that in order for  $P_{DLM}$  to outperform  $P_{BON}$ , we must have  $\sigma_{c,DLM}^2 < \sigma_{c,BON}^2$ . This happens when  $K(\sqrt{2}) < K(1)$ .

### 3.3.3.3 Comparison Between $P_{BON}$ and $P_{DLM}$

In this section, we compare the performance of  $P_{DLM}$  and  $P_{BON}$  in approximating  $P_{max}$  in (3.77). We have made claims that  $P_{BON}$  works well when the underlying process  $Z(\cdot)$  is weakly correlated. We claim that  $P_{DLM}$  agrees with  $P_{max}$  better than  $P_{BON}$ , as the correlation structure of  $Z(\cdot)$  changes. We

simulate processes  $Z_n = \{Z(s) : s \in \mathbb{Z}_N^2\}$  of varying degree of correlation and compute  $P_{DLM}$  and  $P_{BON}$  in each case. For illustration, we consider processes  $Z(\cdot)$  with Matérn autocorrelation functions and vary parameters  $\nu$  and  $\phi$  to generate processes of varying degree of correlation. Figure 3.1 presents the six types of Matérn autocorrelation function we consider, varying from very weakly correlated (1) to highly correlated (6). We also vary the sample size, considering the index set  $\mathbb{Z}_N^2$  with  $N = (3, 11, 25)$ , which corresponds to  $n = (49, 529, 2601)$ .

Table 3.1 reports the Bonferroni bound ( $P_B$ ) and the DLM bound ( $P_D$ ) on  $P_{max} * 100$ , as well as the Monte Carlo approximation of  $P_{max} * 100$  based on simulation ( $P_S$ ), for  $N = (3, 11, 25)$ . The simulated value serves as a proxy for the true value, as we do not know the true value of  $P_{max}$ . Table 3.2 reports the Monte Carlo sampling error for values of  $P_{SIM}$  in the neighborhood of  $(0.01, 0.05, 0.10)$ .

In Table 3.1, we choose critical values so that the corresponding  $P_{max}$  is in the neighborhood of 0.05. This is because we propose the bounds  $P_{BON}$  and  $P_{DLM}$  with the goal of approximating critical values corresponding to a size  $\alpha$  test and  $\alpha = 0.05$  is a typical value.  $P_{SIM}$  and  $P_{DLM}$  are reported for each of the six models in Figure 3.1. However, the Bonferroni bound ( $P_{BON}$ ) does not change with model parameters and is thus reported only once.

In Tables 3.3-3.5, we report the differences  $P_S - P_B$  and  $P_S - P_D$ , corresponding to values in Table 3.1. Notice that in Tables 3.3-3.5,  $P_{SIM} - P_{BON}$  is small



for models (1) – (3) of Figure 3.1, regardless of the value of  $t$ . These models correspond to weakly correlated  $Z(\cdot)$ , with  $K_\theta(h) \leq 0.4$  for all  $h$ . However, we see that as the correlation in  $Z(\cdot)$  increases,  $P_S - P_B$  starts deviating from 0, as evident in the results for models (4) – (6) of Tables 3.3-3.5. Notice that as the correlation in  $Z(\cdot)$  increases,  $P_S - P_D$  is small for all values of  $t$  and for all models (1) – (6). Thus, these results support the hypothesis that  $P_{DLM}$  approximates  $P_{max}$  closer than does  $P_{BON}$ , as the correlation in  $Z(\cdot)$  changes. This happens regardless of the value of  $N$ .

The conclusion that we draw from these results is the following. When  $Z(\cdot)$  is weakly correlated,  $P_{max}$  (approximated by  $P_{SIM}$ ) does not change much as the correlation in  $Z(\cdot)$  changes, and either  $P_{BON}$  or  $P_{DLM}$  can be used to approximate  $P_{max}$ . The advantage to  $P_{BON}$  is that it is extremely simple and fast to calculate. We feel confident that as long as  $K_\theta(h) \leq 0.4$  for all  $h$ , we can define  $Z(\cdot)$  as weakly correlated for our purposes. Note that this is a soft bound and is more of a heuristic figure. When  $Z(\cdot)$  is moderately to highly correlated,  $P_{DLM}$  traces  $P_{max}$  better than  $P_{BON}$  and we recommend using  $P_{DLM}$  to approximate  $P_{max}$ .

### 3.3.4 Recommended Critical Values

Based on the material in Sections 3.3.1, 3.3.2 and 3.3.3 we are now in position to make recommendations on critical value selection for the test statistic  $\Lambda^N$  (3.31). We make the following broad recommendations:

- When  $\Lambda^N(\cdot)$  (and therefore  $\psi^N(\cdot)$ ) is weakly correlated, the use of the Bonferroni bound to approximate  $P^*$  in (3.34) is justified. Care should be taken as to how *weak* correlation is defined. Our experience suggests that the use of the Bonferroni bound is justified when the magnitude of correlations in  $\Lambda^N(\cdot)$  is at most 0.4. However, it is important to keep in mind that this recommendation is based on observation only and is not supported by rigorous theory. Thus, whenever possible, the DLM bound should be computed as well and the critical values corresponding to the two bounds should be compared.
- The DLM bound approximates  $P^*$  consistently well, regardless of the correlation structure of  $\Lambda^N(\cdot)$ . Thus, when the user does not know the correlation structure of  $\Lambda^N(\cdot)$ , or if the  $\Lambda^N(\cdot)$  exhibits high (in magnitude) correlations, the DLM bound is advocated for critical value estimation.

In Sections 3.3.4.1 and 3.3.4.2 we give recommendations on critical value selection for  $\lambda$  when the underlying processes are Matérn and ICAR(1) (respectively). To achieve a size  $\alpha$  test, we need to find  $t_\alpha$ , such that

$$P_\alpha = P_{H_0}(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t_\alpha) = \alpha. \quad (3.85)$$

In Section 3.3.1, we presented an argument for approximating  $P_\alpha$  as

$$P_\alpha \approx 2P(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t_\alpha). \quad (3.86)$$

Next, we detail the computations involved in obtaining  $t_\alpha$  based on the Bonferroni and DLM approximations to  $P(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t_\alpha)$ , as well as the

simulated value of  $t_\alpha$ .

- **Bonferroni** Since the Bonferroni bound on  $P_{H_0}(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t_\alpha)$  is

$$P_{\alpha, BON}^N = nP(\psi^N(s) > t_\alpha) = n(1 - \Phi(t_\alpha)), \quad (3.87)$$

the Bonferroni critical value approximation is

$$t_{\alpha, BON}^n \approx \Phi^{-1}\left(1 - \frac{\alpha}{2n}\right) \quad (3.88)$$

where  $n = (2N + 1)^2$ .

- **DLM** The Discrete Local Maxima (DLM) bound on  $P_{H_0}(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t_\alpha)$  is

$$P_{\alpha, DLM}^N = \sum_{s \in \mathbb{Z}_N^2} P(\psi^N(s) > t_\alpha, \psi^N(t) < \psi^N(s), t \in \mathcal{N}(s) \cap \mathbb{Z}_N^2) \quad (3.89)$$

To obtain the *DLM* approximation to  $t_\alpha$ ,  $t_{\alpha, DLM}$ , we need to set  $P_{\alpha, DLM}$  to  $\frac{\alpha}{2}$  and solve for  $t_\alpha$ . Since it is not possible to obtain a closed-form analytical expression for  $t_\alpha$  in this case, we do the following. We evaluate  $2P_{\alpha, DLM}$  for a range of  $t_\alpha$ 's and fit a smoothing spline to the pair  $(2P_{\alpha, DLM}, t_\alpha)$ . Then, using the fitted equation, we predict  $t_\alpha$  for  $\alpha = \alpha_0$ , where  $\alpha_0$  is the hypothesis test level we have in mind.

- **Simulation** To obtain a simulated approximation to  $t_\alpha$ ,  $t_{\alpha, SIM}$ , we do the following. Let  $M$  be the simulation size and fix  $N$ .
  - (a) For  $m = 1, \dots, M$ , simulate  $\Lambda_m^N(\cdot)$  and obtain  $\zeta_m = \max_{s \in \mathbb{Z}_N^2} |\Lambda_m^N(s)|$ .
  - (b) Let  $t_\alpha = \zeta_{(k)}$ , the  $k$ -th order statistic of  $\zeta$ , where  $k = (1 - \alpha)M$ .

### 3.3.4.1 When $Y(\cdot)$ is from the Matérn model

In this section, we discuss the behavior of  $t_\alpha$  when  $Y(\cdot)$  comes from the Matérn model. We consider the six models for  $Y(\cdot)$  presented in Figure 3.1, to cover a range of correlation scenarios. We also vary the index space  $\mathbb{Z}_N^2$  on which observations are sampled and consider  $N = (2, 3, 4, 5, 7, 11)$ , corresponding to sample sizes of  $n = (25, 49, 81, 121, 225, 529)$ .

First, we explain why we consider small to moderate sample sizes. The goal of this exposition is to convince the reader of two things. First, that  $t_{\alpha,DLM}$ , the DLM critical value approximation to  $t_\alpha$ , is a consistently good approximation, regardless of the degree of correlation in  $\Lambda^N(\cdot)$ . Second, that  $t_{\alpha,BON}$ , the Bonferroni critical value approximation to  $t_\alpha$  is a good approximation when  $\Lambda^N(\cdot)$  is weakly correlated.

As we saw in Section 3.3.2.1, when  $Y(\cdot)$  comes from the Matérn model, the magnitude of correlations in  $\Lambda^N(\cdot)$  is positively related to the degree of correlation in  $Y(\cdot)$ . We also saw that even for highly correlated  $Y(\cdot)$ ,  $\Lambda^N(\cdot)$  is weakly correlated, *except* on and around the boundary  $\mathcal{B}$ , where the magnitude of correlations is higher than inside the boundary (see Figures 3.5-3.8). Only when  $N$  is small do we have a considerable percentage of the observations on the boundary  $\mathcal{B}$ , thus resulting in a more strongly correlated  $\Lambda^N(\cdot)$ . Therefore, only when  $N$  is small should we expect some variation in the  $t_\alpha$ 's as we vary the correlation in  $Y(\cdot)$ . When this is the case, we should expect  $t_{\alpha,DLM}$  to

match  $t_\alpha$  better than  $t_{\alpha,BON}$ . As we increase  $N$ , the percentage of observations on  $\mathcal{B}_N$  decreases and thus the variation in the correlations of  $\Lambda^N(\cdot)$  decreases. In turn, the critical values  $t_\alpha$  do not vary much with the degree of correlation in  $Y(\cdot)$ , since  $\Lambda^N(\cdot)$  remains consistently weakly correlated. In this case,  $t_\alpha$  is well approximated by  $t_{\alpha,BON}$ , regardless of the degree of correlation in  $Y(\cdot)$ .

The percentage of observations on the boundary decreases rapidly with  $N$ . When  $N = (2, 3, 4, 5, 7, 11)$ , the percentage of observations on  $\mathcal{B}$  is (64%, 49%, 40%, 33%, 25%, 17%), respectively. Thus, as we will see in the discussion to follow, we only see variability in  $t_\alpha$ s for very small  $N$ .

Table 3.6 lists critical values corresponding to size  $\alpha = (0.01, 0.05, 0.10)$  test and to the six models for  $Y(\cdot)$  labeled as in Figure 3.1, when  $N = 2$  ( $n = 25$ ). We see that even when the sample size is this small, there is not much variation in the simulated critical values ( $t_{\alpha,SIM}$ ) between the six models for  $Y(\cdot)$  that we consider. To gauge whether this magnitude of variation in the critical values has an effect on the corresponding size of the test, we calculated an approximation to  $P_{H_0}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t_\alpha)$  for the range of critical values reported in Table 3.6. These values are reported in Table 3.7. As we see, there is little variability in these numbers, with the greatest being for  $t_\alpha \in [2.83, 2.88]$ . These t-values were calculated for a size  $\alpha = 0.1$  test. We see that the approximated  $P_{H_0}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t)$  varies between 0.09 and 0.11, thus exhibiting some variation around 0.10. If the researcher wants to be very precise about the size of the test corresponding to the estimated critical

values, we recommend calculating  $t_{\alpha,DLM}$ , since it traces  $t_{\alpha,SIM}$  more closely than  $t_{\alpha,BON}$ .

As can be seen from Tables 3.8-3.11, for  $N \geq 3$  there is even less variation in  $t_{\alpha}$ s than for  $N = 2$ . We see practically no variation in the  $t_{\alpha,DLM}$  values and we suspect that the variation in the  $t_{\alpha,SIM}$  values is a simulation error. Therefore, it makes no difference for  $N \geq 3$  whether we use  $t_{\alpha,BON}$  or  $t_{\alpha,DLM}$  to approximate  $t_{\alpha}$ . Since  $t_{\alpha,BON}$  is much easier and faster to calculate, we advocate calculating this bound only.

The conclusion of this discussion is the following. Due to the fact that  $\Lambda^N(\cdot)$  corresponding to  $Y(\cdot)$  from the Matérn model is consistently weakly correlated, the Bonferroni approximation  $t_{\alpha,BON}$  is nearly as good as the DLM approximation  $t_{\alpha,DLM}$ . Since  $t_{\alpha,BON}$  is easier and faster to calculate, we advocate calculating this bound only. Only for very small sample sizes, with  $N \leq 3$ , do we justify the use of  $t_{\alpha,DLM}$ , since there does appear to be some variation in  $t_{\alpha}$ s in this case, due to effect of the boundary of  $\mathbb{Z}_N^2$ . With such a small sample size,  $t_{\alpha,DLM}$  does not take much time to calculate, thus justifying its use further.

### 3.3.4.2 When $Y(\cdot)$ is from the ICAR(1) model

Let  $C_{\theta}(h_x, h_y)$  be the ICAR(1) autocovariance function, specified as in (2.12).

Let  $h = \sqrt{h_x^2 + h_y^2}$ . Figure 3.12 presents  $C_{\theta}(h_x, h_y)$  as a function of  $h$ . Notice that since  $C_{\theta}(h_x, h_y)$  is isotropic, it is a function only of  $h$ . Figure 3.12 presents

autocorrelation functions of varying degree of correlation, corresponding to various values of  $\rho$ , with  $\sigma^2 = 1$ . Besag and Kooperberg [14] point out that in practice moderate to substantial autocorrelations are typical. As can be seen from Figure 3.12, these autocorrelations correspond to  $\rho$  very close to the boundary value of 0.25. Now suppose we observe  $Y_n \equiv \{Y(s) : s \in \mathbb{Z}_N^2\}$  and consider the outlier detection vector  $\Lambda^N(\cdot)$  corresponding to  $Y_n$ . As discussed in Section 3.3.2.2, the autocorrelations in  $\Lambda^N(\cdot)$  on  $\mathbb{Z}_{N-1}^2$  are at most  $\rho$  in magnitude. Since  $\rho$  is less than 0.25, we conclude that  $\Lambda^N(\cdot)$  is weakly correlated on the interior point of the lattice (i.e.  $\mathbb{Z}_{N-1}^2$ ), regardless of the value of  $\rho$ . On the boundary, the correlation structure changes somewhat. However, the effect of the boundary diminishes as  $N$  grows. Since  $\Lambda^N(\cdot)$  is weakly correlated inside the boundary of the lattice, we show critical values  $t_\alpha$  are well approximated by  $t_{\alpha, BON}$ , regardless of the correlation structure of the underlying ICAR(1) process  $Y(\cdot)$ .

As discussed in more detail in Section 3.3.2.2, to compute the estimated critical values  $t_\alpha$  for a given value of  $\rho$  and sample size  $n$ , we need to calculate the covariance matrix of the corresponding  $\Lambda^N(\cdot)$ . This, in turn, requires the computation and inversion of  $\Sigma_N$ , the covariance matrix of the corresponding ICAR(1) process  $Y(\cdot)$ , whose entries are calculated using (2.12). It turns out that the calculations involved are very time consuming. Even for the small samples sizes we present here,  $N = (2, 3, 4)$ , the time required was in magnitude of days. This motivated us to investigate the circulant approximation to

$\Lambda^N(\cdot)$ ,  $\tilde{\Lambda}^N(\cdot)$ , as discussed in Section 3.3.2.2. We expect the critical values  $\tilde{t}_\alpha$ , based on the circulant approximation  $\tilde{\Lambda}^N(\cdot)$ , to be close to  $t_\alpha$  when  $N$  is large. This is because  $\Lambda^N(\cdot)$  and  $\tilde{\Lambda}^N(\cdot)$  are equivalent on  $\mathbb{Z}_{N-1}^2$  and the percentage of observations on  $\mathbb{Z}_{N-1}^2$  relative to  $\mathbb{Z}_N^2$  goes to 1 as  $N \rightarrow \infty$ . We did not test this empirically, however, since for large  $N$ , the computation time involved in simulating  $\Lambda^N(\cdot)$  is prohibitively long.

We did compare  $\tilde{t}_\alpha$  and  $t_\alpha$  for small values of  $N$ . The results are presented in Tables 3.12-3.14. In these tables,  $t_\alpha$  are reported at the top and  $\tilde{t}_\alpha$ , at the bottom. We were encouraged to see that even for such small sample sizes, the values of  $t_\alpha$  and  $\tilde{t}_\alpha$  are very close, regardless of  $\alpha$ ,  $\rho$  or  $N$ . Since the circulant approximation only gets more accurate as  $N$  grows, we see no reason why  $t_\alpha$  and  $\tilde{t}_\alpha$  should not be close for  $N$  larger than the values we tested. Therefore, to save computing time, we advocate the use of the circulant approximation to  $\Sigma_N^{-1}$  for critical value estimation.

Tables 3.12-3.14 reveal another important trend. Even for small  $N$ , the simulated values  $t_{\alpha,SIM}$  do not vary much as we vary the correlations in  $Y(\cdot)$ . Thus the Bonferroni bound  $t_{\alpha,BON}$  is a good approximation to  $t_{\alpha,SIM}$  even for small  $N$ , regardless of the correlation structure of  $Y(\cdot)$ . Remember, it is only for small  $N$  that we expect a higher degree of correlation in  $\Lambda^N(\cdot)$ , due to the boundary effect, and thus some variation in the critical values, as we vary the correlation structure of  $Y(\cdot)$ . However, as our results indicate, even for small  $N$ , the critical values do not vary much, indicating that  $\Lambda^N(\cdot)$  remains weakly



correlated, irrespective of the degree of correlation in  $Y(\cdot)$ .

We also calculated  $\tilde{t}_\alpha$  for  $N = (7, 11)$ , the results are in Tables 3.15 and 3.16.

These results are consistent with the finding that  $\tilde{t}_\alpha$  do not change with the correlations in  $Y(\cdot)$ . Therefore, for ICAR models, we advocate the use of the Bonferroni approximation to  $t_\alpha$ , since it is easier to calculate than the DLM approximation and results in comparable approximations to  $t_\alpha$ .

### 3.3.4.3 Bonferroni approximation to $t_\alpha$ , varying $n$ .

In Sections 3.3.4.1 and 3.3.4.2 we have shown that the critical values  $t_\alpha$  do not change significantly as we vary the correlations in  $Y(\cdot)$ , when  $Y(\cdot)$  comes from the Matérn or ICAR(1) model. In turn, this justifies the use of the Bonferroni approximation to  $t_\alpha$ . The Bonferroni approximation,  $t_{\alpha, BON}$  does vary with  $N$ , and thus with  $n$ . In Table 3.17, we report  $t_{\alpha, BON}^n \approx \Phi^{-1}(1 - \frac{\alpha}{2n})$ , for various values of  $\alpha$  and  $n$ . The remarkable fact is  $t_{\alpha, BON}$  can be calculated quickly for any  $n$ . For example, Table 3.17 took seconds to generate.

## 3.4 $\mu$ and $\theta$ unknown, location of outlier $s_0$ unknown

In Section 3.3, we considered the scenario when the location of the outlier,  $s_0$ , is unknown, but the parameters  $\tau = (\mu, \theta)$  of the process  $W(\cdot)$  are known. In this section, we consider the more realistic scenario when neither the parameters  $\tau$  nor the location of the outlier  $s_0$  are known. We assume the parameters are estimated via maximum likelihood and consider the asymptotic properties

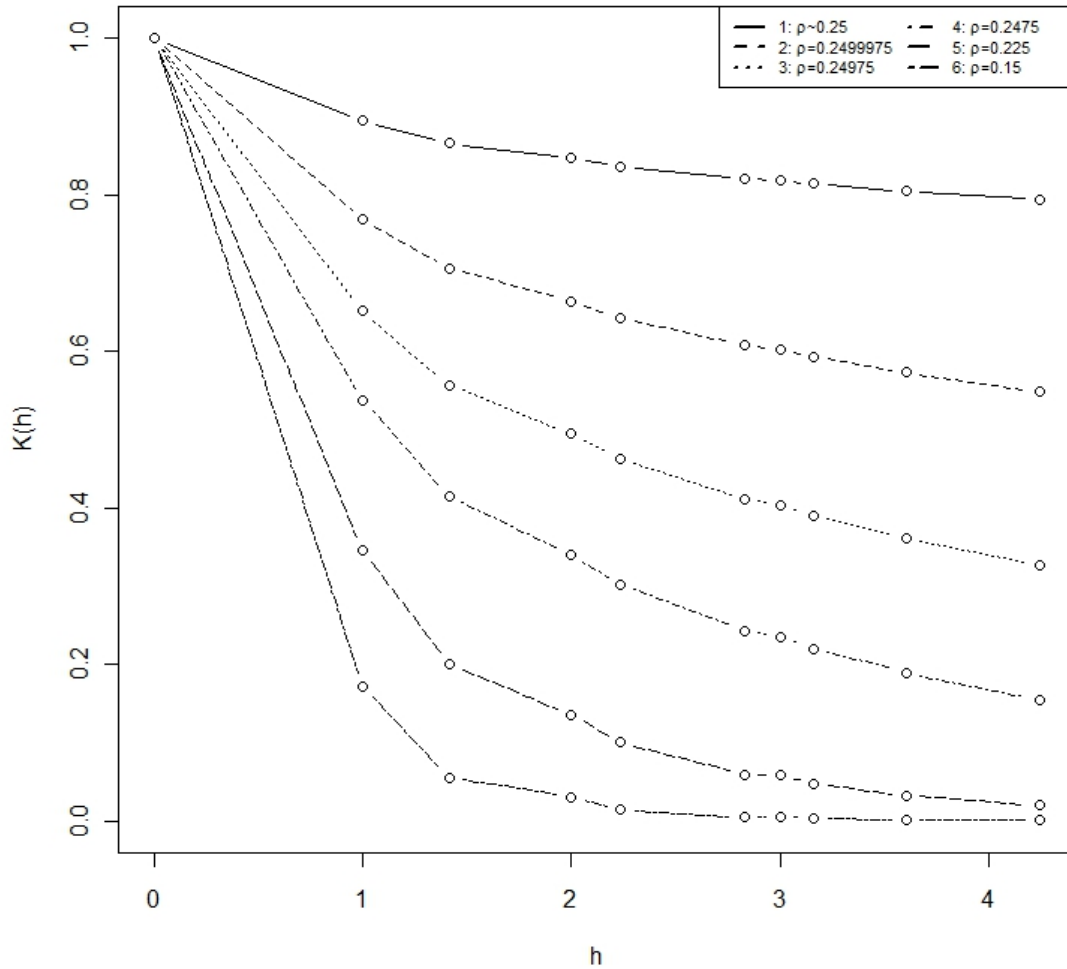


Figure 3.12: ICAR(1) Autocorrelation Function,  $K_\theta(h)$ , as in (2.12),  $h = \sqrt{h_x^2 + h_y^2}$ , varying  $\rho$ .

of the parameter estimates, as well as the asymptotic behavior of the estimated outlier detection statistic  $\Lambda^N(\hat{\tau}_N)$  as  $N \rightarrow \infty$ .

Let  $\Lambda^N(s, \tau)$  be defined as in (3.45). Let  $\Lambda^N(\hat{\tau}_N, \tau, s)$  be obtained by replacing  $\tau$  in  $\Lambda^N(s, \tau)$  with  $\hat{\tau}_N$ , the MLE of  $\tau$ . Notice that  $\Lambda^N(\hat{\tau}_N, \tau, s)$  still depends on  $\tau$ , since  $Y(\cdot)$  is parametrized by  $\tau$ . When  $\hat{\tau}_N$ , the MLE of  $\tau$ , is consistent, The-

orem 6 states that for all  $s \in \mathbb{Z}^2$ ,  $\Lambda^N(\hat{\tau}_N, \tau, s) \xrightarrow[N \rightarrow \infty]{P_\tau} \Lambda(\tau, s)$ . In Section 3.4.1, we discuss conditions under which  $\hat{\tau}_N$  is consistent for  $\tau$ .

Theorem 6 has a practical implication when estimating critical values. When  $\tau$  is unknown, the test statistic for the hypothesis in (3.28) becomes

$$\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(\hat{\tau}_N, s)| \quad (3.90)$$

As  $N$  increases, for a fixed  $t$ ,  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(\hat{\tau}_N, s)| > t)$  should approach  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda(\tau, s)| > t)$ . In turn, in Section 3.3, we demonstrated that when  $Y(\cdot)$  comes from either the Matérn or the ICAR(1) model,  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda(\tau, s)| > t)$  is well approximated by its Bonferroni bound, as in (3.87). Since the Bonferroni bound does not depend on  $\tau$ , we do not need to estimate  $\tau$  to calculate it. Thus, as long as  $N$  is large enough to make  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(\hat{\tau}_N, s)| > t)$  close to  $P(\max_{s \in \mathbb{Z}_N^2} |\Lambda(\tau, s)| > t)$ , the critical values for the test statistic in (3.90) should be well approximated by the Bonferroni critical value approximation, as in (3.88).

**Theorem 6.** *Let  $\Lambda^N(s, \tau)$  be defined as in (3.45). Let  $\Lambda^N(\hat{\tau}_N, \tau, s)$  be obtained by replacing  $\tau$  in  $\Lambda^N(s, \tau)$  with  $\hat{\tau}_N$ , the MLE of  $\tau$ . Let  $\hat{\tau}_N$  be a function of  $\{Y(s) : s \in \mathbb{Z}_N^2\}$ , with  $\hat{\tau}_N \xrightarrow[N \rightarrow \infty]{P} \tau$ . Then for all  $s \in \mathbb{Z}^2$ ,*

$$\Lambda^N(\hat{\tau}_N, \tau, s) \xrightarrow[N \rightarrow \infty]{P_\tau} \Lambda(\tau, s). \quad (3.91)$$

### 3.4.1 Asymptotic properties of parameter estimates

In this section, we discuss asymptotic properties of parameter estimates of a Gaussian random field  $W(\cdot)$  defined  $\mathbb{Z}^d$ . Assume  $W(s) = X(s)' \beta + \epsilon(s)$ , where  $X(\cdot)$  is a known vector valued function,  $\beta \in B$  is a  $q \times 1$  vector of unknown coefficients,  $B$  is an open subset of  $\mathbb{R}^q$ ,  $\epsilon(\cdot)$  has mean 0 and covariance function  $\text{cov}(\epsilon(s), \epsilon(t)) = C_\theta(s, t)$ ,  $\theta \in \Theta$  is a  $p \times 1$  parameter vector, and  $\Theta$  is an open subset of  $\mathbb{R}^p$ . We assume that  $C_\theta(s, t)$  is twice differentiable with respect to  $\theta$  at all points  $\mathbb{Z}^d \times \Theta$  and that it is positive definite in the sense that for every finite subset  $S_n \in \mathbb{Z}^d$ , the covariance matrix  $\Sigma_n = (\sigma_{st})_{s, t \in S_n}$  is positive-definite.

#### 3.4.1.1 Increasing-domain vs. Infill asymptotics

This section is partially taken from [11].

In this thesis we assume that the sample set,  $S_n$ , is predetermined and nonrandom with the restriction that  $\|s_i - s_j\| \geq \delta > 0$ , for all pairs  $i, j = 1, \dots, n$ , to ensure that the sampling domain increases as  $n$  increases. This type of spatial asymptotics is referred to as *increasing domain asymptotics* [15]. This is in contrast to *infill* [15], also referred to as *fixed-domain* [12] asymptotics, where more and more observations are sampled in a fixed domain  $D$ . When lattice data have a spacing between observations that is fixed, increasing-domain asymptotics is more appropriate. We chose the increasing domain setting, be-

cause there are many more asymptotic results on parameter estimates in this setting than under fixed-domain asymptotics.

One of the difficulties in obtaining fixed-domain asymptotic results even for Gaussian processes observed on a grid is that, under any model including something like a range parameter, there will generally be at least one function of the parameters that cannot be consistently estimated as the number of observations increases. For example, if  $D$  is a bounded subset of  $\mathbb{R}^d$  for  $d \leq 3$  and  $K^0$  and  $K^1$  are two Matérn covariance functions (see Section 2.2.1.1), then the ratio of corresponding spectral densities tending to 1 as the frequency tends to  $\infty$  is necessary and sufficient for the corresponding measures to be equivalent [18, Theorem 2]. As a special case of this result, consider  $K_{\theta,\phi}(h) = \theta e^{-\phi h}$  with  $\theta, \phi$  unknown positive parameters. Then the corresponding Gaussian measures are equivalent on any bounded infinite set  $D$  in 3 or fewer dimensions if and only if  $\theta_0\phi_0 = \theta_1\phi_1$ . It immediately follows that it is not possible to estimate either  $\theta$  or  $\phi$  consistently, based on observations in  $D$ , but it may be possible to estimate their product consistently [18].

It is important to note the following feature of fixed domain asymptotics, discussed in more detail in Section A.3.5.1. If the goal is spatial prediction (kriging), then as long as the corresponding measures are equivalent, the kriging errors are asymptotically the same. This is not the same as saying that if one plugs in the estimated covariance function into the standard kriging formulas, then the resulting predictors are asymptotically optimal or the presumed

mean squared prediction errors are asymptotically correct. Stein [12] claims that such a result is true under fixed-domain asymptotics in broad generality for Gaussian processes with homogeneous covariance functions. However, no substantial progress has been made to formally assess this problem.

Under increasing-domain asymptotics, one might generally expect that under parametric models for the covariance structure, all parameters can be consistently estimated and that the ML estimators should obey the usual asymptotics. Specifically, we might expect that these estimators are asymptotically normal with asymptotic mean given by the true value of the parameters and asymptotic covariance matrix given by the inverse of the Fisher information matrix. Mardia and Marshall [19] give a general result to this effect about ML estimators, but despite a strong effort by the authors to give results that can be verified in practice, it is not easy to verify the conditions for observations that are not on a regular lattice.

Theorem 7, from [19], gives conditions under which parameters  $\beta$  and  $\theta$  are consistent and asymptotically normal, provided we are in the increasing-domain setting.

**Theorem 7** (Mardia and Marshall, 1984). *Consider a weakly stationary Gaussian process  $Y_n \sim N_n(X\beta, \Sigma_\theta)$ , sampled on a regular lattice  $\mathbb{Z}^d$ , with  $\theta = (\theta_1, \dots, \theta_k)$ . Let  $L_n^{(2)}$  be the second-derivative matrix of the negative log-likelihood*

of  $Y_n$  and  $J_n = E(L_n^{(2)})$ . Then,

$$J_n = \begin{pmatrix} J_\beta & 0 \\ 0 & J_\theta \end{pmatrix} \quad (3.92)$$

where  $J_\beta = X'\Sigma^{-1}X$ , and the  $(i, j)$ th element of  $J_\theta$  is

$$(1/2)t_{ij} = (1/2)\text{tr}(\Sigma^{-1}\Sigma_i\Sigma^{-1}\Sigma_j) = (1/2)\text{tr}(\Sigma\Sigma^i\Sigma\Sigma^j) \quad (3.93)$$

where  $\Sigma_i = \frac{\partial\Sigma}{\partial\theta_i}$  and  $\Sigma^i = \frac{\partial\Sigma^{-1}}{\partial\theta_i}$ . Suppose, as  $n \rightarrow \infty$ , we have

(a)

$$t_{ij}/(t_{ii}t_{jj})^{1/2} \rightarrow a_{ij} \quad (3.94)$$

for all  $i, j = 1, \dots, k$  where  $A = (a_{ij})$  is a nonsingular matrix.

(b)  $(X'X)^{-1} \rightarrow 0$ , a  $q \times q$  matrix whose elements are 0.

Further assume that the autocovariance function  $C_\theta(h) = \sigma^2 K_\theta(h)$ . Let  $K_i(h) = \frac{\partial K_\theta(h)}{\partial\theta_i}$  and  $K_{ij}(h) = \frac{\partial K_\theta(h)}{\partial\theta_i\partial\theta_j}$ . Suppose  $K$ ,  $K_i$  and  $K_{ij}$  are absolutely summable over the set  $\{h : h \in \mathbb{Z}^d\} \forall i, j = 1, \dots, k$ . Then the m.l. estimator  $\hat{\tau}_n$  of  $\tau = (\beta, \theta)$  satisfies

$$\hat{\tau}_n \xrightarrow{P} \tau \quad (3.95)$$

and

$$J_n^{1/2}(\hat{\tau}_n - \tau) \xrightarrow{D} N(0, I). \quad (3.96)$$

According to Cressie [15], these results are general enough to have as a special case the ICAR(1) model, as in (2), defined on a regular lattice of the torus.

When an ICAR(1) process on a regular lattice is approximated by wrapping

the lattice points on a torus, the corresponding circulant precision matrix  $\Sigma^{-1}$  has a simple form and is sparse. This has a tremendous effect in reducing the computational burden involved in parameter estimation. Since the parameter estimates obtained using the circulant approximation are consistent, we recommend using the circulant approximation to the precision matrix when the sample size is large.

We have not been able to verify the conditions of Theorem 7 for the Matérn model. This prompted us to look for other sources for consistency of the autocovariance function parameters under the Matérn model. Guyon [20] states conditions on the spectral density of a mean-zero, second-order stationary process which guarantee consistency of the autocovariance function parameters. The author assumes that sampling is done on an increasing sequence of sets  $P_N$ , with  $c_N$  points, which tends to infinity with the same speed in all directions. This assumption is satisfied under our sampling scheme, where  $P_N = \mathbb{Z}_N^2$ .

Let  $f = f(\theta, \omega)$  be the spectral density corresponding to a zero-mean, second-order stationary process  $Z$  with  $\omega$  in a  $d$ -dimensional torus  $T^d$  and  $\theta$  belonging to a bounded open set  $\Theta \subset \mathbb{R}^d$ . Proposition 6 gives conditions on  $f$  under which the MLE  $\hat{\theta}_N$  of  $\theta$  is consistent.

**Proposition 6.** *Suppose the spectral density  $f(\theta, \omega)$  satisfies the following three conditions:*

- (a) *if  $\theta \neq \theta'$ , the first derivative  $f_\theta$  of  $f$  is such that  $f_\theta \neq f_{\theta'}$  in  $L^1(T^d)$ .*



(b)  $f(\omega, \theta)$  is continuous and nonzero on  $T^d \times \Theta$ ; also  $Z$  is a linear process and we suppose that  $\epsilon$  has fourth cumulant  $\kappa_4(\theta)$ .

(c)  $f$ , its first and second derivatives with respect to  $\theta$ , have derivatives with respect to  $\omega_i$  of order  $d$ , in  $L_1(T^d)$  ( $i = 1, \dots, d$ ).

Then the MLE of  $\theta$  is consistent.

Guyon [20, p.100] states that all classical models comply with the conditions above. Theorem 1.1 of [25] supplies the proof of Guyon's assertions in Proposition 6.

### 3.4.2 Simulated results

In this section, we compare critical values for  $\max_{s \in \mathbb{Z}_N^2} |\hat{\Lambda}^N(s)|$  to critical values for  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$ , when  $Y(\cdot)$  comes from the Matérn model with various degrees of correlation. Proposition 7 implies that we do not need to vary  $\mu$  and  $\sigma^2$  when estimating critical values for the test statistic  $\max_{s \in \mathbb{Z}_N^2} |\hat{\Lambda}^N(s)|$ . Also, estimation of Matérn covariance parameters is prone to computational issues. For this reason, in this presentation, we consider the exponential model, which is a special case of the Matérn model, with the smoothness parameter  $\nu$  fixed at 0.5. We vary the range parameter  $\phi$ , as in Figure 3.13, to produce a variety of correlation scenarios for  $Y(\cdot)$ .

Table 3.18 reports the critical values for  $\max_{s \in \mathbb{Z}_N^2} |\hat{\Lambda}^N(s)|$  and for  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$ , for  $N = (2, 3, 4, 7)$  and size  $\alpha = (0.01, 0.05, 0.1)$ . We consider small to moder-

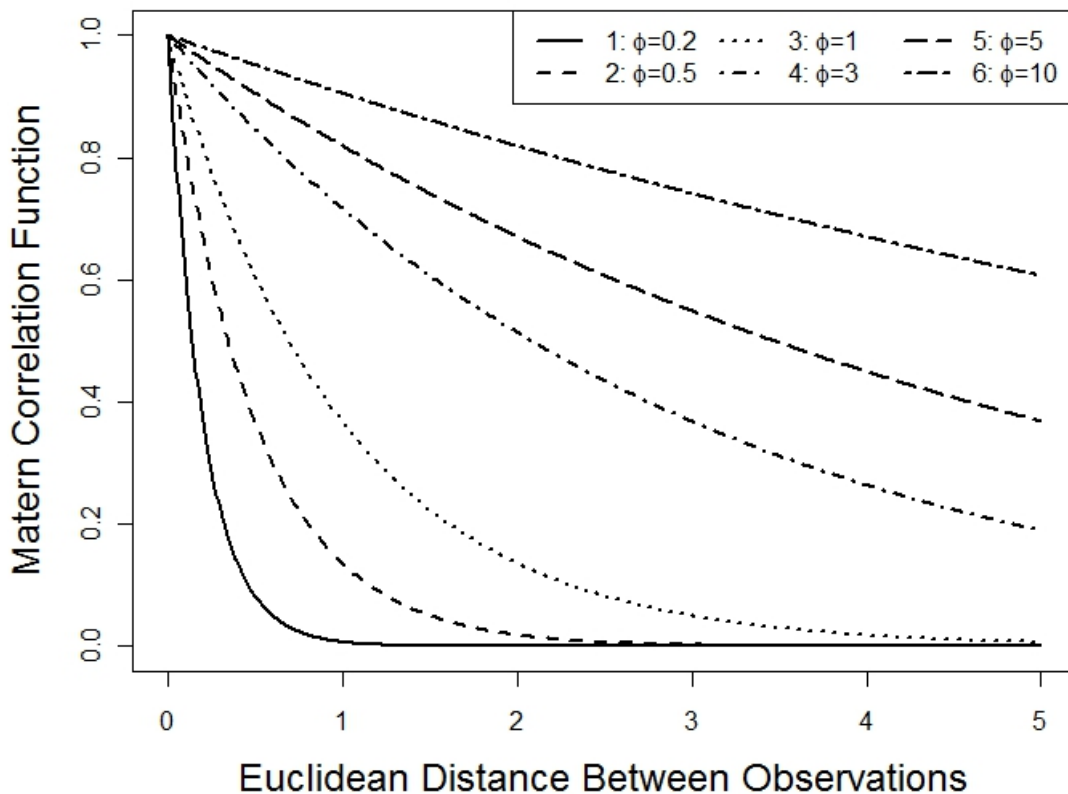


Figure 3.13: Matern autocorrelation function with  $\nu = 0.5$ , corresponds to the exponential autocorrelation function; varying  $\phi$ .

ate sample sizes mainly due to time constraints. However, we do expect the critical values for  $\max_{s \in \mathbb{Z}_N^2} |\hat{\Lambda}^N(s)|$  and for  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$  to converge as  $N$  grows, due to Theorem 6. We observe some difference between critical values for  $\max_{s \in \mathbb{Z}_N^2} |\hat{\Lambda}^N(s)|$  and for  $\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)|$  for all sample sizes studied. However, the differences diminish as  $N$  increases.

**Proposition 7.** *Let  $Y \sim N(\mu, \Sigma)$  and let  $\Lambda = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(Y - \mu\mathbf{1})$ , where  $X$  is a full-rank matrix. Let  $\hat{\Lambda} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}(Y - \hat{\mu}\mathbf{1})$ , where  $\hat{\mu}$*

and  $\hat{\Sigma}$  are the MLEs of  $\mu$  and  $\Sigma$ . Further, let  $\Sigma = \sigma^2 D$ . Then,  $\hat{\Lambda}$  is invariant to the values of  $\mu$  and  $\sigma^2$ .

*Proof.* The MLEs of  $\mu$  and  $\sigma^2$  can be expressed as

$$\hat{\mu} = (\mathbf{1}'\hat{D}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{D}^{-1}Y \quad (3.97)$$

and

$$\hat{\sigma}^2 = \frac{(Y - \hat{\mu}\mathbf{1})'\hat{D}^{-1}(Y - \hat{\mu}\mathbf{1})}{n} \quad (3.98)$$

where  $\hat{D}$  is the MLE of  $D$ . The likelihood  $L(\mu, \sigma^2, D|Y)$  is

$$L(\mu, \sigma^2, D|Y) \propto |\sigma^2 D|^{-1/2} \exp\left(-\frac{(Y - \mu)'\hat{D}^{-1}(Y - \mu)}{2\sigma^2}\right) \quad (3.99)$$

Now plug in (3.97) and (3.98) into (3.99), to obtain the concentrated likelihood for  $\gamma$ , parameters of  $D$ . Then, the term in the exponent simplifies to  $-n/2$ .

And we have

$$L(\hat{\mu}, \hat{\sigma}^2, D|Y) \propto |\hat{\sigma}^2 D|^{-1/2} \propto |(Y - \hat{\mu})'\hat{D}^{-1}(Y - \hat{\mu})D| \quad (3.100)$$

We show next that  $Y - \hat{\mu}$  is invariant to  $\mu$ . Consider  $Y$  with  $\mu = 0$  and let

$Y' = Y + k\mathbf{1}$ . We show that  $Y - \hat{\mu} = Y' - \hat{\mu}'$ .

$$\begin{aligned} Y' - \hat{\mu}' &= Y' - (\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\Sigma}^{-1}Y' \\ &= (Y - k\mathbf{1}) - (\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\Sigma}^{-1}(Y - k\mathbf{1}) \\ &= Y - \hat{\mu} - k\mathbf{1} + k(\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1} \\ &= Y - \hat{\mu} \end{aligned}$$

Therefore, neither  $\hat{\sigma}^2$ , nor  $L(\hat{\mu}, \hat{\sigma}^2, D|Y)$  (and thus  $\hat{D}$ ), being functions of  $Y - \hat{\mu}$ , vary with  $k$ . Therefore,  $\hat{\Lambda} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}(Y - \hat{\mu}1)$  is invariant to the choice of  $k$  (and thus  $\mu$ ).

Now to demonstrate invariance of  $\hat{\Lambda}$  to  $\sigma^2$ , consider  $Y$  with variance 1 and  $Y' = kY$ . Then, it is trivial to see that  $Y' - \hat{\mu}' = k(Y - \hat{\mu})$ . Now

$$L(\hat{\mu}', \hat{\sigma}'^2, D|Y) \propto |(Y' - \hat{\mu}')'D^{-1}(Y' - \hat{\mu}')D| = k^2|(Y - \hat{\mu})'D^{-1}(Y - \hat{\mu})D| \quad (3.101)$$

Thus  $\hat{D}$  is invariant to  $k$ . It is easy to see that  $\hat{\sigma}'^2 = k^2\hat{\sigma}^2$ . But the  $k^2$  multiple of  $\sigma^2$  cancels out in  $\hat{\Lambda}$ . This demonstrates that  $\hat{\Lambda}$  is invariant to the choice of  $\sigma^2$ .

□

$N = 3$		1		2		3		4		5		6	
$t$	$P_B$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$
2.90	9.14	8.09	9.11	8.64	9.04	8.41	8.65	6.68	7.51	5.13	6.10	4.36	5.05
2.95	7.79	7.37	7.76	7.12	7.71	7.42	7.40	5.53	6.45	5.07	5.26	3.52	4.36
3.00	6.61	6.34	6.60	5.81	6.56	6.05	6.31	5.08	5.53	3.77	4.52	3.08	3.75
3.05	5.61	4.92	5.60	5.47	5.56	5.02	5.37	4.74	4.73	3.31	3.88	2.52	3.22
3.10	4.74	4.58	4.73	4.54	4.71	4.31	4.56	3.63	4.03	3.05	3.32	2.23	2.76

$N = 11$		1		2		3		4		5		6	
$t$	$P_B$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$
3.60	8.42	8.66	8.41	7.92	8.40	8.01	8.26	6.51	7.56	6.00	6.33	4.54	5.25
3.65	6.94	6.97	6.93	6.74	6.93	6.63	6.82	5.85	6.26	4.72	5.27	3.53	4.38
3.70	5.70	5.76	5.70	5.79	5.69	5.15	5.62	4.79	5.18	3.83	4.37	2.88	3.64
3.75	4.68	4.68	4.68	4.56	4.67	4.32	4.61	4.23	4.27	3.03	3.62	2.73	3.02
3.80	3.83	3.82	3.83	3.62	3.82	3.86	3.78	3.44	3.51	2.54	2.99	2.31	2.50

$N = 25$		1		2		3		4		5		6	
$t$	$P_B$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$	$P_S$	$P_D$
4.00	8.24	7.64	8.24	7.53	8.23	7.96	8.17	7.25	7.68	5.77	6.61	4.91	5.56
4.05	6.66	6.86	6.66	6.65	6.66	5.97	6.61	5.95	6.23	5.43	5.39	3.97	4.54
4.10	5.37	5.68	5.37	5.27	5.37	5.29	5.34	4.49	5.05	4.26	4.38	3.13	3.70
4.15	4.32	4.22	4.32	4.08	4.32	4.05	4.30	3.99	4.08	3.62	3.55	2.72	3.01
4.20	3.47	3.33	3.47	3.57	3.47	3.58	3.45	3.12	3.28	2.68	2.87	2.29	2.44

Table 3.1: Monte Carlo Estimates of  $P_{max} * 100$  ( $P_S$ ) and approximate values of  $P_{max} * 100$  as a function of  $t$ :  $P_{BON}$  ( $P_B$ ) and  $P_{DLM}$  ( $P_D$ ). Simulation size  $M=10,000$ ; sampled on  $\mathbb{Z}_N^2$ ,  $N$  as labeled. Models 1 – 6 are as labeled in Figure 3.1.

$P_{SIM}$	0.0050	0.0100	0.0150	0.0450	0.0500	0.0550	0.0950	0.1000	0.1050
$SE(P_{SIM})$	0.0014	0.0020	0.0024	0.0041	0.0043	0.0045	0.0057	0.0059	0.0060

Table 3.2: Monte Carlo sampling errors for typical values of  $P_{SIM}$ , with simulation size  $M = 10,000$ .

$N = 3$	$P_S - P_B$						$P_S - P_D$					
$t$	1	2	3	4	5	6	1	2	3	4	5	6
2.90	-1.05	-0.50	-0.73	-2.46	-4.01	-4.78	-1.02	-0.40	-0.24	-0.83	-0.97	-0.69
2.95	-0.42	-0.67	-0.37	-2.26	-2.72	-4.27	-0.39	-0.59	0.02	-0.92	-0.19	-0.84
3.00	-0.27	-0.80	-0.56	-1.53	-2.84	-3.53	-0.26	-0.75	-0.26	-0.45	-0.75	-0.67
3.05	-0.69	-0.14	-0.59	-0.87	-2.30	-3.09	-0.68	-0.09	-0.35	0.01	-0.57	-0.70
3.10	-0.16	-0.20	-0.43	-1.11	-1.69	-2.51	-0.15	-0.17	-0.25	-0.40	-0.27	-0.53

Table 3.3:  $(P_S - P_B)$  and  $(P_S - P_D)$  for  $N = 3$ , in Table 3.1

$N = 11$	$P_S - P_B$						$P_S - P_D$					
$t$	1	2	3	4	5	6	1	2	3	4	5	6
3.60	0.24	-0.50	-0.41	-1.91	-2.42	-3.88	0.25	-0.48	-0.25	-1.05	-0.33	-0.71
3.65	0.03	-0.20	-0.31	-1.09	-2.22	-3.41	0.04	-0.19	-0.19	-0.41	-0.55	-0.85
3.70	0.06	0.09	-0.55	-0.91	-1.87	-2.82	0.06	0.10	-0.47	-0.39	-0.54	-0.76
3.75	0.00	-0.12	-0.36	-0.45	-1.65	-1.95	0.00	-0.11	-0.29	-0.04	-0.59	-0.29
3.80	-0.01	-0.21	0.03	-0.39	-1.29	-1.52	-0.01	-0.20	0.08	-0.07	-0.45	-0.19

Table 3.4:  $P_S - P_B$  and  $P_S - P_D$  for  $N = 11$ , in Table 3.1

$N = 25$	$P_S - P_B$						$P_S - P_D$					
$t$	1	2	3	4	5	6	1	2	3	4	5	6
4.00	-0.60	-0.71	-0.28	-0.99	-2.47	-3.33	-0.60	-0.70	-0.21	-0.43	-0.84	-0.65
4.05	0.20	-0.01	-0.69	-0.71	-1.23	-2.69	0.20	-0.01	-0.64	-0.28	0.04	-0.57
4.10	0.31	-0.10	-0.08	-0.88	-1.11	-2.24	0.31	-0.10	-0.05	-0.56	-0.12	-0.57
4.15	-0.10	-0.24	-0.27	-0.33	-0.70	-1.60	-0.10	-0.24	-0.25	-0.09	0.07	-0.29
4.20	-0.14	0.10	0.11	-0.35	-0.79	-1.18	-0.14	0.10	0.13	-0.16	-0.19	-0.15

Table 3.5:  $P_S - P_B$  and  $P_S - P_D$  for  $N = 25$ , in Table 3.1

		$t_{\alpha,DLM}$						$t_{\alpha,SIM}$					
$\alpha$	$t_{\alpha,BON}$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	3.54	3.54	3.54	3.54	3.54	3.53	3.53	3.56	3.53	3.54	3.53	3.53	3.52
0.05	3.09	3.09	3.09	3.09	3.08	3.07	3.07	3.08	3.08	3.08	3.07	3.07	3.06
0.1	2.88	2.88	2.88	2.87	2.86	2.85	2.85	2.86	2.86	2.85	2.84	2.84	2.83

Table 3.6: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 2$ ,  $n = 25$ . Models 1 – 6 are as labeled in Figure 3.1.



$t_\alpha$	1	2	3	4	5	6
2.83	0.10	0.11	0.11	0.11	0.10	0.10
2.84	0.11	0.10	0.11	0.10	0.10	0.10
2.85	0.10	0.10	0.10	0.09	0.09	0.10
2.86	0.10	0.10	0.10	0.10	0.10	0.09
2.87	0.10	0.10	0.10	0.10	0.09	0.09
2.88	0.10	0.10	0.09	0.09	0.09	0.09
3.06	0.05	0.05	0.05	0.05	0.05	0.05
3.07	0.05	0.05	0.05	0.06	0.05	0.05
3.08	0.05	0.05	0.05	0.05	0.05	0.05
3.09	0.05	0.05	0.05	0.05	0.05	0.05
3.52	0.01	0.01	0.01	0.01	0.01	0.01
3.53	0.01	0.01	0.01	0.01	0.01	0.01
3.54	0.01	0.01	0.01	0.01	0.01	0.01
3.55	0.01	0.01	0.01	0.01	0.01	0.01
3.56	0.01	0.01	0.01	0.01	0.01	0.01

Table 3.7: Approximated values of  $P_{H_0}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t)$ ;  $N = 2$ ,  $n = 25$ .

Simulation size  $M = 10,000$ . Models 1 – 6 are as labeled in Figure 3.1.

		$t_{\alpha,DLM}$						$t_{\alpha,SIM}$					
$\alpha$	$t_{\alpha,BON}$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	3.71	3.71	3.71	3.71	3.71	3.71	3.71	3.71	3.72	3.71	3.71	3.70	3.71
0.05	3.28	3.28	3.28	3.28	3.28	3.27	3.27	3.28	3.28	3.28	3.27	3.26	3.26
0.1	3.08	3.08	3.08	3.08	3.07	3.07	3.07	3.07	3.07	3.06	3.06	3.05	3.05

Table 3.8: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 3$ ,  $n = 49$ . Models 1 – 6 are as labeled in Figure 3.1.

		$t_{\alpha,DLM}$						$t_{\alpha,SIM}$					
$\alpha$	$t_{\alpha,BON}$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	3.94	3.94	3.94	3.94	3.93	3.93	3.93	3.95	3.94	3.94	3.93	3.93	3.93
0.05	3.53	3.53	3.53	3.53	3.53	3.53	3.52	3.52	3.53	3.52	3.52	3.52	3.51
0.1	3.34	3.34	3.34	3.34	3.34	3.33	3.33	3.33	3.32	3.33	3.32	3.32	3.32

Table 3.9: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 5$ ,  $n = 121$ . Models 1 – 6 are as labeled in Figure 3.1.

		$t_{\alpha,DLM}$						$t_{\alpha,SIM}$					
$\alpha$	$t_{\alpha,BON}$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.09	4.10	4.09	4.07	4.08	4.07
0.05	3.69	3.69	3.69	3.69	3.69	3.69	3.69	3.69	3.69	3.68	3.68	3.68	3.68
0.1	3.51	3.51	3.51	3.51	3.51	3.51	3.51	3.50	3.50	3.50	3.49	3.49	3.49

Table 3.10: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 7$ ,  $n = 121$ . Models 1 – 6 are as labeled in Figure 3.1.

		$t_{\alpha,DLM}$						$t_{\alpha,SIM}$					
$\alpha$	$t_{\alpha,BON}$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.27	4.26
0.05	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.90	3.89	3.89
0.1	3.73	3.73	3.73	3.73	3.73	3.73	3.73	3.72	3.72	3.72	3.72	3.72	3.71

Table 3.11: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 11$ ,  $n = 225$ . Models 1 – 6 are as labeled in Figure 3.1.

		$t_{\alpha,DLM}$					$t_{\alpha,SIM}$				
$\alpha$	$t_{\alpha,BON}$	2	3	4	5	6	2	3	4	5	6
0.01	3.54	3.54	3.54	3.54	3.54	3.54	3.54	3.55	3.54	3.54	3.53
0.05	3.09	3.08	3.08	3.08	3.09	3.09	3.07	3.08	3.08	3.09	3.08
0.1	2.88	2.87	2.87	2.87	2.87	2.87	2.85	2.86	2.85	2.86	2.86
0.01	3.54	3.54	3.54	3.54	3.54	3.54	3.52	3.53	3.55	3.53	3.54
0.05	3.09	3.09	3.09	3.09	3.09	3.09	3.07	3.07	3.09	3.07	3.08
0.1	2.88	2.87	2.87	2.87	2.87	2.87	2.85	2.86	2.86	2.85	2.86

Table 3.12: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 2$ ,  $n = 25$ . Models 2 – 6 are as labeled in Figure 3.12. Exact (top), circulant approximation (bottom)

		$t_{\alpha,DLM}$					$t_{\alpha,SIM}$				
$\alpha$	$t_{\alpha,BON}$	2	3	4	5	6	2	3	4	5	6
0.01	3.71	3.71	3.71	3.71	3.71	3.71	3.71	3.71	3.71	3.71	3.71
0.05	3.28	3.28	3.28	3.28	3.28	3.28	3.27	3.27	3.28	3.28	3.28
0.1	3.08	3.08	3.08	3.08	3.08	3.08	3.07	3.06	3.06	3.07	3.07
0.01	3.71	3.71	3.71	3.71	3.71	3.71	3.72	3.71	3.70	3.71	3.71
0.05	3.28	3.28	3.28	3.28	3.28	3.28	3.28	3.27	3.27	3.27	3.28
0.1	3.08	3.08	3.08	3.08	3.08	3.08	3.06	3.06	3.06	3.07	3.07

Table 3.13: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000$ ;  $N = 3$ ,  $n = 49$ . Models 2 – 6 are as labeled in Figure 3.12. Exact (top), circulant approximation (bottom)

		$t_{\alpha,DLM}$					$t_{\alpha,SIM}$				
$\alpha$	$t_{\alpha,BON}$	2	3	4	5	6	2	3	4	5	6
0.01	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.83	3.84	3.84	3.84
0.05	3.42	3.42	3.42	3.42	3.42	3.42	3.41	3.41	3.41	3.42	3.41
0.1	3.23	3.23	3.23	3.23	3.23	3.23	3.21	3.21	3.21	3.21	3.21
0.01	3.84	3.84	3.84	3.84	3.84	3.84	3.82	3.84	3.85	3.84	3.84
0.05	3.42	3.42	3.42	3.42	3.42	3.42	3.41	3.41	3.42	3.41	3.42
0.1	3.23	3.23	3.23	3.23	3.23	3.23	3.21	3.21	3.21	3.21	3.22

Table 3.14: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM). Simulation size  $M = 100,000; N = 4, n = 81$ . Models 2 – 6 are as labeled in Figure 3.12. Exact (top), circulant approximation (bottom)

		$t_{\alpha,DLM}$					$t_{\alpha,SIM}$				
$\alpha$	$t_{\alpha,BON}$	2	3	4	5	6	2	3	4	5	6
0.01	4.08	4.08	4.08	4.08	4.08	4.08	4.09	4.09	4.09	4.08	4.08
0.05	3.69	3.69	3.69	3.69	3.69	3.69	3.69	3.68	3.69	3.69	3.68
0.1	3.51	3.51	3.51	3.51	3.51	3.51	3.50	3.50	3.50	3.49	3.50

Table 3.15: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM), based on the circulant approximation to  $\Sigma_N^{-1}$ . Simulation size  $M = 100,000; N = 7, n = 225$ . Models 2 – 6 are as labeled in Figure 3.12.

		$t_{\alpha,DLM}$					$t_{\alpha,SIM}$				
$\alpha$	$t_{\alpha,BON}$	2	3	4	5	6	2	3	4	5	6
0.01	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.28	4.27	4.27	4.29
0.05	3.90	3.90	3.90	3.90	3.90	3.90	3.89	3.90	3.90	3.90	3.90
0.1	3.73	3.73	3.73	3.73	3.73	3.73	3.72	3.72	3.72	3.72	3.72

Table 3.16: Approximated values of  $t_\alpha$ :  $t_{\alpha,BON}$  (Bonferroni),  $t_{\alpha,SIM}$  (simulated),  $t_{\alpha,DLM}$  (DLM), based on the circulant approximation to  $\Sigma_N^{-1}$ . Simulation size  $M = 100,000$ ;  $N = 11, n = 529$ . Models 2 – 6 are as labeled in Figure 3.12.

$\alpha \backslash n$	49	225	961	2601	8281	22801	40401	361201	1002001
0.01	3.71	4.08	4.41	4.62	4.85	5.05	5.16	5.56	5.73
0.05	3.28	3.69	4.05	4.27	4.53	4.73	4.85	5.27	5.45
0.1	3.08	3.51	3.88	4.12	4.38	4.59	4.71	5.14	5.33

Table 3.17:  $t_{\alpha,BON}$ , varying  $n$

$\alpha$	Estimated Parameters						True Parameters					
$N = 2$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	3.47	3.49	3.52	3.51	3.53	3.55	3.51	3.55	3.59	3.60	3.58	3.57
0.05	3.03	3.03	3.07	3.07	3.07	3.07	3.06	3.08	3.10	3.08	3.09	3.08
0.1	2.81	2.82	2.84	2.84	2.85	2.85	2.87	2.85	2.86	2.85	2.85	2.85
$N = 3$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	3.65	3.72	3.77	3.71	3.69	3.72	3.74	3.68	3.70	3.67	3.74	3.70
0.05	3.23	3.27	3.28	3.29	3.26	3.27	3.28	3.27	3.27	3.27	3.28	3.26
0.1	3.03	3.05	3.06	3.06	3.05	3.06	3.07	3.06	3.07	3.07	3.06	3.06
$N = 4$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	3.80	3.82	3.82	3.84	3.87	3.84	3.82	3.79	3.82	3.80	3.8	3.83
0.05	3.39	3.41	3.40	3.43	3.42	3.42	3.40	3.39	3.40	3.43	3.4	3.40
0.1	3.19	3.21	3.20	3.21	3.22	3.21	3.20	3.21	3.21	3.23	3.2	3.20
$N=7$	1	2	3	4	5	6	1	2	3	4	5	6
0.01	4.04	4.07	4.08	4.12	4.06	4.08	4.10	4.09	4.09	4.08	4.06	4.12
0.05	3.68	3.68	3.70	3.70	3.67	3.69	3.69	3.70	3.68	3.68	3.68	3.70
0.1	3.50	3.50	3.50	3.50	3.49	3.48	3.50	3.50	3.49	3.50	3.50	3.50

Table 3.18: Critical values for a size  $\alpha$  test obtained via simulation, with  $M = 10,000$ . The six models correspond to models in Figure 3.13. Parameters  $\mu$ ,  $\sigma^2$  and  $\phi$  were either estimated or fixed at true values.



## Chapter 4: Power

Consider the hypothesis test in (3.28), which tests for the presence of an outlier at an unknown location. Assume that  $\tau$  is known. In this chapter, we present a discussion on the power of this test, that is,

$$P_{H_1}(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t_\alpha) \quad (4.1)$$

where  $t_\alpha$  is an estimated critical value for a size  $\alpha$  test, as discussed in Section 3.3. We assume that  $Y(\cdot)$  comes from the Matérn model. First, we state the distribution of  $\Lambda^N(\cdot)$  under the alternative hypothesis in (3.28).

**Proposition 8 (Distribution of  $\Lambda^N(\cdot)$  under  $H_a$  in (3.28)).** *Consider the hypothesis test in (3.28) where, under  $H_a$ ,  $E(Y(t)) = \mu + \beta(t)$ , where  $\beta(t)$  is the magnitude of the outlier at site  $t$ . Let  $\Lambda^N(s)$  be defined as in Proposition 2. Let  $\Sigma_{\theta,N}^{-1} \equiv (\sigma_N^{st})_{s,t \in \mathbb{Z}_N^2}$  be the inverse covariance matrix of  $\{Y(s) : s \in \mathbb{Z}_N^2\}$ . Then  $\Lambda^N(\cdot)$  is Gaussian with*

$$\Lambda^N(s) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} (Y(t) - \mu) \quad (4.2)$$

where

$$E(\Lambda^N(s)) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} \beta(t) \quad (4.3)$$

and

$$\text{cov}(\Lambda^N(s), \Lambda^N(t)) = (\sigma_N^{ss})^{-1/2} (\sigma_N^{tt})^{-1/2} \sigma_N^{ts} \quad (4.4)$$

*Proof.* The proof follows from the proof of Proposition 2, if we note that

$$E(\Lambda^N(s)) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} E(Y(t) - \mu) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} \beta(t). \quad (4.5)$$

□

Notice that the distribution of  $\Lambda^N(\cdot)$  is invariant to the choice of  $\mu$ . Therefore, we fix  $\mu$  at 0 in our analysis. Remember, under  $H_0$ , the distribution of the test statistic  $\Lambda^N(\cdot)$  was also invariant to  $\sigma^2$  (see Proposition 3). However, this is not the case under  $H_a$ . Let  $\Sigma_\theta = \sigma^2 D_\gamma$ , for some matrix  $D_\gamma$  with  $\theta = (\sigma^2, \gamma)$ .

Then,

$$E(\Lambda^N(s)) = (\sigma_N^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} \sigma_N^{st} \beta(t) = \sigma^{-1} (d^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} d^{st} \beta(t),$$

where  $D_\gamma^{-1} \equiv (d^{st})$ . Thus,  $E(\Lambda^N(\cdot))$  is inversely proportional to  $\sigma^2$ . When performing power calculations, we will be interested in answering the following question: How does the power of the test vary with outlier magnitude? However, in answering this question, we need to consider the variability in the data, controlled by the parameter  $\sigma^2$ , since the effect of outlier(s) diminishes as the variability in the data grows. Since the variance parameter  $\sigma^2$  rescales the mean of  $\Lambda^N(\cdot)$ , it also rescales outlier magnitude. Consider  $\tilde{\beta}(\cdot) = \beta(\cdot)/\sigma$ , the rescaled outlier magnitude. We then have

$$E(\Lambda^N(s)) = (d^{ss})^{-1/2} \sum_{t \in \mathbb{Z}_N^2} d^{st} \tilde{\beta}(t),$$

which no longer depends on  $\sigma^2$ . It makes sense in a simulation study to answer the following question: How does the power of the test vary with the rescaled outlier magnitude  $\tilde{\beta}(\cdot)$ ? We can do this by fixing  $\sigma^2 = 1$ , which results in  $\tilde{\beta}(\cdot) = \beta(\cdot)$ . Thus, in our presentation, we fix  $\mu = 0$  and  $\sigma^2 = 1$  and vary the Matérn covariance parameters of  $Y(\cdot)$ ,  $\nu$  and  $\phi$ .

#### 4.1 How does power vary as we vary the degree of correlation in $Y(\cdot)$ ?

When  $Y(\cdot)$  comes from the Matérn model, the parameters  $\nu$  and  $\phi$  control the degree of correlation in  $Y(\cdot)$ . Naturally, if the underlying process  $Y(\cdot)$  is weakly correlated, the realizations of  $Y(\cdot)$  are rough and the effect of outlier(s) is more likely to be masked. This, in turn, results in diminished power of a hypothesis test which tests for the presence of outliers, as in (3.28). On the other hand, when  $Y(\cdot)$  is highly correlated, its realizations tend to be smooth and, as a result, outliers tend to stand out more from the rest of the data. This is clearly seen in Figure 4.1. In this figure, we plot two realizations of  $Y(\cdot)$  on  $\mathbb{Z}_5^2$ : one, rough, at the top, with  $\nu = 0.5$  and  $\phi = 0.1$  and one, smooth, at the bottom, with  $\nu = 2$  and  $\phi = 1$ . We also add a single outlier of magnitude 3 in the middle of the grid. Clearly, the outlier is more distinguishable in the smooth process than in the rough process.

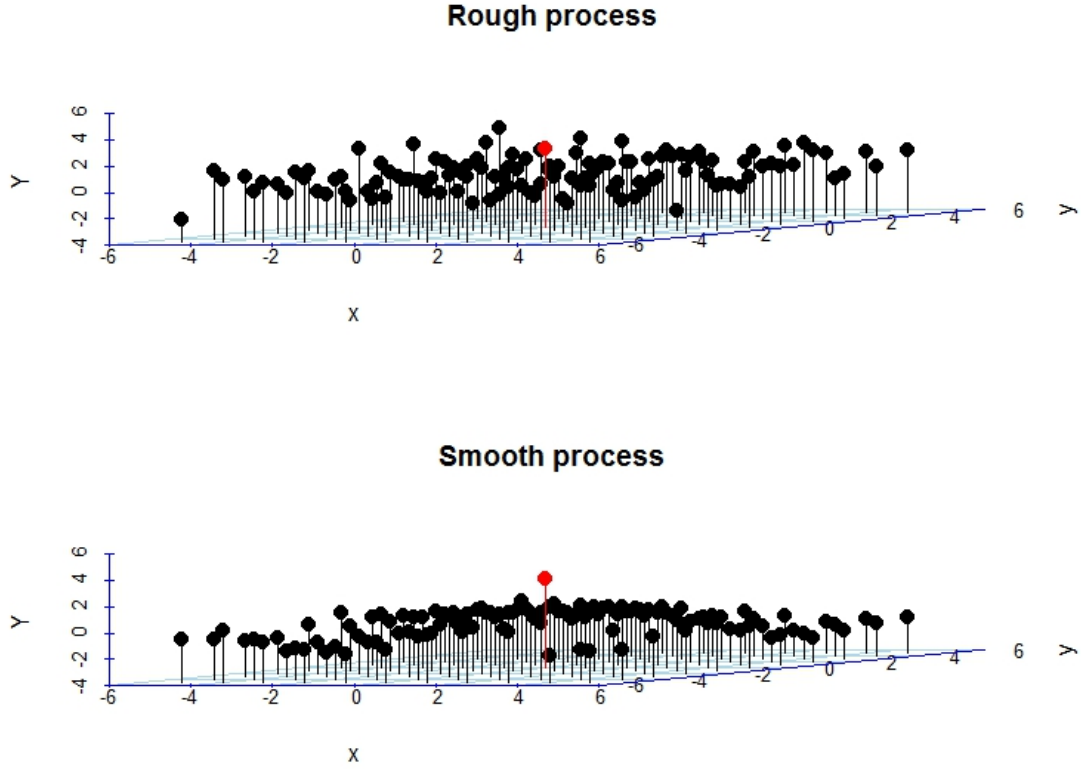


Figure 4.1: One outlier of magnitude 3 is added to a Matérn process  $W(\cdot)$ . Top:  $Y(\cdot)$  rough,  $\nu = 0.5$ ,  $\phi = 0.1$ . Bottom:  $Y(\cdot)$  smooth,  $\nu = 2$ ,  $\phi = 1$ .  $N = 5$

Now, recall that the power of the hypothesis test in (3.28) is

$$P_1 = P_{H_1}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t) = P_{H_1} \left[ (\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t) \cup (\min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t) \right]. \quad (4.6)$$

Clearly  $P_1$  increases as the magnitude of the mean of  $\Lambda^N(\cdot)$  increases. For, if for some  $s$ ,  $E(\Lambda^N(s))$  is large and positive,  $P_{H_1}(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t)$  will be high. On the other hand, if for some  $s$ ,  $E(\Lambda^N(s))$  is large and positive,  $P_{H_1}(\min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t)$  will be high.

### 4.1.1 Single outlier in the middle of the grid

In this section, we assume that a single outlier of magnitude  $\beta$  is present in  $Y(\cdot)$  at the center of the grid  $\mathbb{Z}_N^2$  and provide a discussion on the power of the corresponding test. As we mentioned above,  $P_1$  in (4.6) increases as the magnitude of  $E(\Lambda^N(s))_{s \in \mathbb{Z}_N^2}$  increases. When a single outlier of magnitude  $\beta(t)$  is present at location  $t$ , the mean of  $\Lambda^N(s)$  at  $s \in \mathbb{Z}_N^2$  is

$$E(\Lambda^N(s)) = \frac{\sigma_N^{st}}{\sqrt{\sigma_N^{ss}}} \beta(t). \quad (4.7)$$

Now, if  $Y(\cdot)$  is uncorrelated,  $\Sigma_N = \Sigma_N^{-1} = I$  and thus  $E(\Lambda^N(s)) = 1(s = t)\beta(t)$ . This is also approximately true for weakly correlated processes  $Y(\cdot)$ . Consider Figure 4.2. This figure plots  $E(\Lambda^N(s))_{s \in \mathbb{Z}_2^2}$  corresponding to a process  $Y(\cdot)$  from the Matérn model, with a single outlier of magnitude  $\beta = 3$  added at the center of the grid. The top figure corresponds to a highly correlated  $Y(\cdot)$ , with  $\nu = 2$  and  $\phi = 1$  and the bottom figure corresponds to a weakly correlated  $Y(\cdot)$ , with  $\nu = 0.5$  and  $\phi = 0.1$ . Notice that the mean vector of  $\Lambda^N(\cdot)$  corresponding to a highly correlated  $Y(\cdot)$  exhibits much higher magnitudes than the one corresponding to a weakly correlated  $Y(\cdot)$ . In fact, on the bottom figure, the mean of  $\Lambda^N(\cdot)$  is approximately 3 at the location of the outlier and approximately 0, everywhere else. This is to be expected, since in this case  $Y(\cdot)$  is nearly uncorrelated (see Figure 3.1).

Now we are ready to present some results. In Tables 4.5-4.8, we present simulated power estimates ( $P_S$ ), varying the correlations in the underlying

Matérn process  $W(\cdot)$  and the index set  $\mathbb{Z}_N^2$ . These tables correspond to  $N = (3, 5, 9, 15)$ . For a given size  $\alpha$  test, critical values were estimated via simulation with  $M = 10,000$ . Then, in turn, the power of the test with the estimated critical values was approximated via simulation, with simulation size of  $M = 10,000$ .

As expected, the power increases as the underlying process  $Y(\cdot)$  becomes more correlated. Also, the power decreases with sample size. The reason for this is that searching over more test statistics for locations that are not outliers reduces the rejection probability because the critical values increase and are applied across observations not known to be outliers.

We also wanted to know how the power changes with the magnitude of the outlier,  $\beta$ . The magnitude of the outlier is varied in Tables 4.9-4.11. These numbers were generated for  $N = 3$ , just to get an idea of the growth in power with  $\beta$ .

#### 4.1.2 Single outlier at the edge of the grid

Another question of interest was whether the location of the outlier on the lattice affects the power of the test in (4.1). Instead of placing the outlier at the center of the grid, as we did in Section 4.1.1, we now place a single outlier of magnitude 3 at the bottom left corner of the lattice and compute the corresponding power. The results are reported in Tables 4.1-4.4, where we also list the corresponding numbers for outliers placed in the middle of the

grid, for ease of comparison.

We do not observe much difference in power under the two scenarios when  $Y(\cdot)$  is weakly correlated. This is evident by comparing the numbers in Tables 4.1-4.4 for models 1 and 2. We also do not see much difference in power under the two scenarios for highly correlated  $Y(\cdot)$ , which corresponds to models 5 and 6. Interestingly enough, the biggest difference in power under the two scenarios is seen when  $Y(\cdot)$  is moderately correlated, as in models 3 and 4.

One reason why power under the two scenarios may differ is the following. There are fewer observations surrounding an observation at the corner of the lattice than there are surrounding an observation at the center of the lattice. Therefore, it is more difficult to identify an outlier at the corner of the lattice, where there are fewer neighboring observations to compare to than in the middle of the lattice. This intuitive argument can be made rigorous by noting the following. Consider the form of  $\Lambda^N(\cdot)$  given in (3.56). Since the conditional variance in the denominator of this expression will be larger for observations in the corner of the grid than in the middle, the corresponding statistic  $\Lambda^N(s)$  will be smaller, if all other factors are kept equal. Thus, the power of the corresponding test will be lower.

When the underlying process  $Y(\cdot)$  is weakly correlated, we do not gain much information about an observation being an outlier based on the neighboring observations. Therefore, it makes sense that the power under the two scenarios is similar when  $Y(\cdot)$  is weakly correlated. On the other hand, when  $Y(\cdot)$  is

highly correlated, an unusual observation is likely to stand out regardless of its placement (corner or middle), since the rest of the observations will tend to look similar. When  $Y(\cdot)$  is moderately correlated, however, the additional observations surrounding the outlying observation at the middle of the grid help in determining its “outlyingness”, thus resulting in an increase in the power of the corresponding test over the power of the test for outlier at the corner of the lattice, where fewer neighboring observations are available to help in identifying the corner observation as an outlier.

$\alpha$	Edge of the lattice						Middle of the lattice					
	1	2	3	4	5	6	1	2	3	4	5	6
0.01	0.25	0.27	0.39	0.76	0.98	1.00	0.25	0.28	0.49	0.95	1.00	1.00
0.05	0.42	0.45	0.58	0.87	1.00	1.00	0.41	0.46	0.68	0.98	1.00	1.00
0.1	0.54	0.55	0.66	0.91	1.00	1.00	0.54	0.58	0.77	0.99	1.00	1.00

Table 4.1: Outlier of magnitude  $\beta = 3$  at the edge of the lattice and the middle of the lattice.  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The simulated power numbers correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 3$ .

## 4.2 Bonferroni and DLM bounds on power

In Section 3.3, we studied two bounds on  $P_{H_0}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t)$ , the Bonferroni bound and the DLM bound. Similarly, it is of interest to establish good



$\alpha$	Edge of the lattice						Middle of the lattice					
	1	2	3	4	5	6	1	2	3	4	5	6
0.01	0.19	0.20	0.30	0.64	0.97	1.00	0.18	0.22	0.41	0.92	1.00	1.00
0.05	0.33	0.35	0.47	0.78	0.99	1.00	0.33	0.37	0.66	0.96	1.00	1.00
0.1	0.43	0.45	0.58	0.85	1.00	1.00	0.43	0.47	0.69	0.98	1.00	1.00

Table 4.2: Outlier of magnitude  $\beta = 3$  at the edge of the lattice and the middle of the lattice.  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The simulated power numbers correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 5$ .

bounds on the power,  $P_{H_1}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t)$ . We now present a discussion on both of these bounds.

#### 4.2.1 Bonferroni bound

Recall from Section 3.3 that we approximated

$$P_{H_0}(\max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t) = P_{H_0}(\max_{s \in \mathbb{Z}_N^2} |\psi^N(s)| > t) \approx 2P_{H_0}(\max_{s \in \mathbb{Z}_N^2} \psi^N(s) > t) \quad (4.8)$$

as given in (3.41). In fact, the transition from  $\Lambda^N(\cdot)$  to  $\psi^N(\cdot)$  is only relevant when using the DLM bound. This is because the advantage of  $\psi^N(\cdot)$  is its correlation structure, which is irrelevant when using the Bonferroni approximation. Under  $H_a$ , the mean of  $\Lambda^N(\cdot)$ , as given in (4.7), is no longer 0. Thus  $\Lambda^N(\cdot)$  is no longer symmetric around 0 and thus  $P_{H_a}(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t) \neq$

$\alpha$	Edge of the lattice						Middle of the lattice					
	1	2	3	4	5	6	1	2	3	4	5	6
0.01	0.13	0.14	0.22	0.57	0.95	1.00	0.14	0.15	0.32	0.87	1.00	1.00
0.05	0.26	0.26	0.38	0.73	0.98	1.00	0.25	0.28	0.50	0.94	1.00	1.00
0.1	0.33	0.36	0.46	0.80	0.99	1.00	0.35	0.36	0.58	0.96	1.00	1.00

Table 4.3: Outlier of magnitude  $\beta = 3$  at the edge of the lattice and the middle of the lattice.  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The simulated power numbers,  $P_S$ , correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 9$ .

$P_{H_a}(\min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t)$ . Thus, to approximate power, we use

$$\begin{aligned}
P_1 &= P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} |\Lambda^N(s)| > t \right] \approx P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t \right] + P_{H_1} \left[ \min_{s \in \mathbb{Z}_N^2} \Lambda^N(s) < -t \right] \\
&= P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t \right] + P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} -\Lambda^N(s) > t \right].
\end{aligned} \tag{4.9}$$

The transition from the third to the fourth expression in (4.9) is necessary to take advantage of the theory on tail probabilities of the maxima of Gaussian random fields. Then we have the following approximation to  $P_1$ , using the Bonferroni bound:

$$\begin{aligned}
P_1 &\approx P_{H_1}(\max_{s \in \mathbb{Z}_N^2} \Lambda^N(s) > t) + P_{H_1}(\max_{s \in \mathbb{Z}_N^2} -\Lambda^N(s) > t) \\
&\approx \sum_{s \in \mathbb{Z}_N^2} P(\Lambda^N(s) > t) + \sum_{s \in \mathbb{Z}_N^2} P(-\Lambda^N(s) > t)
\end{aligned} \tag{4.10}$$

$\alpha$	Edge of the lattice						Middle of the lattice					
	1	2	3	4	5	6	1	2	3	4	5	6
0.01	0.09	0.09	0.17	0.47	0.93	1.00	0.09	0.10	0.26	0.82	1.00	1.00
0.05	0.19	0.21	0.30	0.63	0.97	1.00	0.20	0.21	0.39	0.91	1.00	1.00
0.1	0.28	0.29	0.39	0.71	0.98	1.00	0.27	0.30	0.48	0.94	1.00	1.00

Table 4.4: Outlier of magnitude  $\beta = 3$  at the edge of the lattice and the middle of the lattice.  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The simulated power numbers,  $P_S$ , correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 15$ .

However, when we used this approximation, we found that  $P_1$  was overestimated and the severity of the bias increased with the correlations in the underlying process  $Y(\cdot)$ . We attribute this to the mean structure of  $\Lambda^N(\cdot)$ . Recall Figure 4.2, where  $E(\Lambda^N(s))_{s \in \mathbb{Z}_N^2}$  was plotted for two types of  $Y(\cdot)$  processes, a weakly correlated one and a highly correlated one. When  $Y(\cdot)$  is highly correlated,  $E(\Lambda^N(s))_{s \in \mathbb{Z}_N^2}$  exhibits much higher magnitudes in the mean, resulting in inflated  $\sum_{s \in \mathbb{Z}_N^2} P(\Lambda^N(s) > t)$  and  $\sum_{s \in \mathbb{Z}_N^2} P(-\Lambda^N(s) > t)$ . To rectify this problem, based purely on speculation, we constructed a modified process  $\check{\Lambda}^N(\cdot)$  in the following manner. Let  $q$  be the location of the outlier (assumed here to be in the middle of the grid  $\mathbb{Z}_N^2$ ) and define  $\check{\Lambda}^N(s)$  for all  $s \in \mathbb{Z}_N^2$  as

$$\begin{aligned}
E(\check{\Lambda}^N(s)) &= \mathbf{1}(s = q)E(\Lambda^N(q)), \text{ for all } s \in \mathbb{Z}_N^2, \\
\text{cov}(\check{\Lambda}^N(s), \check{\Lambda}^N(t)) &= \text{cov}(\Lambda^N(s), \Lambda^N(t)), \text{ for all } s, t \in \mathbb{Z}_N^2.
\end{aligned} \tag{4.11}$$

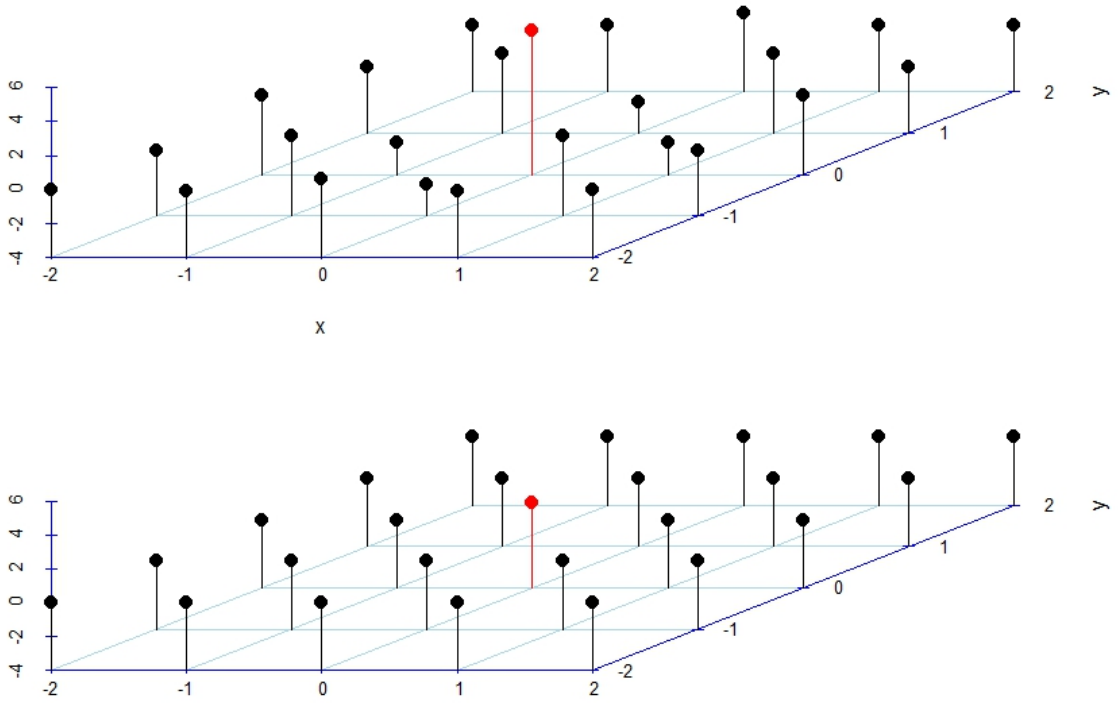


Figure 4.2:  $E(\Lambda^N(s))$  under  $H_1$ , single outlier of magnitude  $\beta = 3$  at the center of  $\mathbb{Z}_2^2$  grid. Top: smooth  $Y(\cdot)$  with  $\nu = 2$  and  $\phi = 1$ , bottom: rough  $Y(\cdot)$  with  $\nu = 0.5$  and  $\phi = 0.1$ .

Figure 4.3 plots the mean of  $\Lambda^N(\cdot)$  and  $\check{\Lambda}^N(\cdot)$ , corresponding to a highly correlated  $Y(\cdot)$  on  $\mathbb{Z}_2^2$ . The mean of  $\check{\Lambda}^N(\cdot)$  is non-zero only at the location of the outlier. Now, consider the approximation to  $P_1$  based on  $\check{\Lambda}^N(\cdot)$ :

$$P_1 \approx \sum_{s \in \mathbb{Z}_N^2} P(\check{\Lambda}^N(s) > t) + \sum_{s \in \mathbb{Z}_N^2} P(-\check{\Lambda}^N(s) > t) \quad (4.12)$$

We found that this approximation works much better. Consider again Tables 4.5-4.8, where the Bonferroni approximation is denoted as  $P_B$ . We see that the difference between  $P_S$  and  $P_B$  is consistently small. Where  $P_B$  exceeds

	1			2			3			4			5			6		
$\alpha$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$
0.01	0.25	0.25	0.25	0.28	0.28	0.27	0.50	0.50	0.49	0.96	0.96	0.95	1.01	1.01	1	1.01	1.01	1
0.05	0.44	0.44	0.41	0.49	0.49	0.46	0.72	0.72	0.68	1.03	1.03	0.98	1.05	1.05	1	1.06	1.05	1
0.1	0.58	0.58	0.54	0.63	0.63	0.58	0.84	0.84	0.77	1.10	1.09	0.99	1.12	1.11	1	1.12	1.10	1

Table 4.5:  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The approximated power numbers,  $P_S$  (simulated),  $P_D$  (DLM) and  $P_B$  (Bonferroni) correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 3$ ,  $\beta = 3$ .

1, in Tables 4.5-4.8, we could assume it is 1. The advantage to the Bonferroni approximation, of course, is that is much faster to calculate than the simulated value. However, we repeat, that these are only speculative results, and have only been tested in the special case where, under  $H_a$ , a single outlier is present at the center of the grid.

#### 4.2.2 DLM bound

When using the DLM bound, we do need the transition from  $\Lambda^N(\cdot)$  to  $\psi^N(\cdot)$ .

Define  $\check{\psi}^N(s)$  for all  $s \in \mathbb{Z}_N^2$  as

$$\begin{aligned}
 E(\check{\psi}^N(s)) &= \mathbf{1}(s = q)E(\psi^N(q)), \text{ for all } s \in \mathbb{Z}_N^2, \\
 \text{cov}(\check{\psi}^N(s), \check{\psi}^N(t)) &= \text{cov}(\psi^N(s), \psi^N(t)), \text{ for all } s, t \in \mathbb{Z}_N^2.
 \end{aligned} \tag{4.13}$$

	1			2			3			4			5			6		
$\alpha$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$
0.01	0.18	0.18	0.18	0.21	0.21	0.22	0.41	0.41	0.41	0.93	0.93	0.92	1.01	1.01	1.00	1.01	1.01	1.00
0.05	0.34	0.34	0.33	0.39	0.39	0.37	0.63	0.62	0.60	1.01	1.01	0.96	1.05	1.05	1.00	1.05	1.05	1.00
0.1	0.47	0.47	0.43	0.52	0.51	0.47	0.76	0.76	0.69	1.08	1.08	0.98	1.10	1.10	1.00	1.10	1.10	1.00

Table 4.6:  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The approximated power numbers,  $P_S$  (simulated),  $P_D$  (DLM) and  $P_B$  (Bonferroni) correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 5$ ,  $\beta = 3$ .

Based on the same logic as in Section 4.2.1, we then approximate:

$$P_1 \approx P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} \check{\psi}^N(s) > t \right] + P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} -\check{\psi}^N(s) > t \right]. \quad (4.14)$$

where we in turn use the DLM approximation for the two terms in (4.14), as in:

$$P_{H_1} \left[ \max_{s \in \mathbb{Z}_N^2} \check{\psi}^N(s) > t \right] = \sum_{u \in \mathbb{Z}_N^2} P \left[ \check{\psi}^N(u) > t, \check{\psi}^N(v) < \check{\psi}^N(u), v \in \mathcal{N}(u) \cap \mathbb{Z}_N^2 \right] \quad (4.15)$$

and, analogously, for  $P_{H_1}(\max_{s \in \mathbb{Z}_N^2} -\check{\psi}^N(s) > t)$ . Here  $\mathcal{N}(u)$  is defined as in Definition 1. Consider again Tables 4.5-4.8, where the DLM approximation is labeled  $P_D$ . Again, where power exceeds 1, in Tables 4.5-4.8, we could assume it is 1. The advantage to the DLM approximation over simulation is that it is faster to compute. However, these are only speculative results, and have only been tested in the special case where, under  $H_a$ , a single outlier is present at

	1			2			3			4			5			6		
$\alpha$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$
0.01	0.13	0.13	0.14	0.15	0.15	0.15	0.33	0.33	0.32	0.89	0.89	0.87	1.01	1.01	1.00	1.01	1.01	1.00
0.05	0.26	0.26	0.25	0.29	0.29	0.28	0.51	0.51	0.50	1.00	1.00	0.94	1.05	1.05	1.00	1.05	1.05	1.00
0.1	0.37	0.37	0.35	0.40	0.40	0.36	0.64	0.64	0.58	1.07	1.07	0.96	1.11	1.10	1.00	1.11	1.11	1.00

Table 4.7:  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The approximated power numbers,  $P_S$  (simulated),  $P_D$  (DLM) and  $P_B$  (Bonferroni) correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 9$ ,  $\beta = 3$ .

the center of the grid.

### 4.2.3 Conclusion

We have found that the Bonferroni and DLM approximations in (4.12) and (4.14) work well in the scenario where, under  $H_a$ , a single outlier of magnitude  $\beta = 3$  is present at the center of the grid  $\mathbb{Z}_N^2$ . We only tested this for small to moderate sample sizes, corresponding to  $N = (3, 5, 9, 15)$ . However, the validity of these approximations need further testing. This would be a good topic for further study.

	1			2			3			4			5			6		
$\alpha$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$	$P_B$	$P_D$	$P_S$
0.01	0.09	0.09	0.09	0.10	0.10	0.10	0.25	0.25	0.26	0.82	0.82	0.82	1.01	1.01	1.00	1.01	1.01	1.00
0.05	0.20	0.20	0.20	0.22	0.22	0.21	0.41	0.41	0.39	0.95	0.95	0.91	1.05	1.05	1.00	1.06	1.06	1.00
0.1	0.30	0.30	0.27	0.33	0.33	0.30	0.53	0.53	0.48	1.04	1.04	0.94	1.11	1.11	1.00	1.12	1.12	1.00

Table 4.8:  $\alpha$  is the size of the test, the corresponding critical value was computed via simulation with  $M = 10000$ . The approximated power numbers,  $P_S$  (simulated),  $P_D$  (DLM) and  $P_B$  (Bonferroni) correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 15$ ,  $\beta = 3$ .

		Simulated power					
$\alpha$	$t_{BON}$	1	2	3	4	5	6
0.01	3.71	0.61	0.67	0.89	1.00	1.00	1.00
0.05	3.28	0.78	0.81	0.95	1.00	1.00	1.00
0.1	3.08	0.84	0.87	0.97	1.00	1.00	1.00

Table 4.9:  $\alpha$  is the size of the test,  $t_{BON}$  is the corresponding Bonferroni approximation to the critical value. The simulated power numbers correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 3$ ,  $\beta = 4$ .



		Simulated power					
$\alpha$	$t_{BON}$	1	2	3	4	5	6
0.01	3.71	0.90	0.93	0.99	1.00	1.00	1.00
0.05	3.28	0.96	0.97	1.00	1.00	1.00	1.00
0.1	3.08	0.97	0.99	1.00	1.00	1.00	1.00

Table 4.10:  $\alpha$  is the size of the test,  $t_{BON}$  is the corresponding Bonferroni approximation to the critical value. The simulated power numbers correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 3$ ,  $\beta = 5$ .

		Simulated power					
$\alpha$	$t_{BON}$	1	2	3	4	5	6
0.01	3.71	0.99	0.99	1.00	1.00	1.00	1.00
0.05	3.28	1.00	1.00	1.00	1.00	1.00	1.00
0.1	3.08	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.11:  $\alpha$  is the size of the test,  $t_{BON}$  is the corresponding Bonferroni approximation to the critical value. The simulated power numbers correspond to the six models for  $Y(\cdot)$  in Figure 3.1, simulation size  $M = 10000$ ,  $N = 3$ ,  $\beta = 6$ .

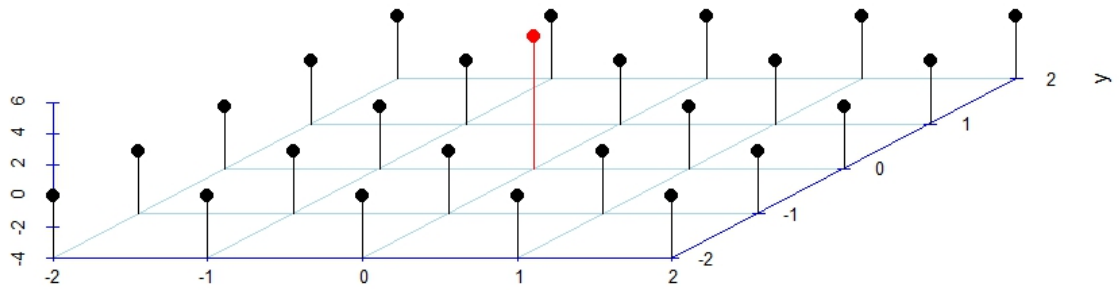
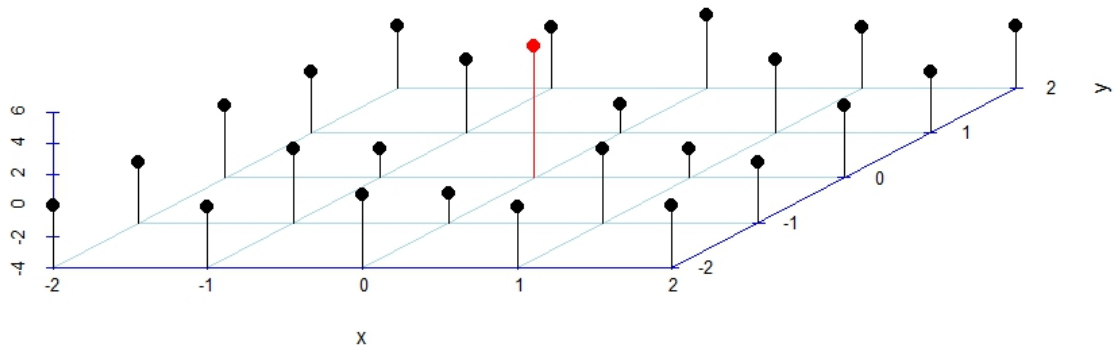


Figure 4.3:  $E(\Lambda^N(s))$  and  $E(\check{\Lambda}^N(s))$  under  $H_1$ , single outlier of magnitude  $\beta = 3$  at the center of  $\mathbb{Z}_2^2$  grid. Smooth  $Y(\cdot)$ , with  $\nu = 2$  and  $\phi = 1$ .

## Chapter A: Background Information

### A.1 Measure Theory

The material in this section is largely taken from [21].

**Definition 6** (measurable transformation). *Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A function  $T : \Omega \rightarrow \Omega$  is called a transformation. Furthermore, if  $T$  is a measurable function, we refer to  $T : \Omega \rightarrow \Omega$  as a measurable transformation.*

**Definition 7** (measure-preserving transformation). *A transformation  $T : \Omega \rightarrow \Omega$  is measure-preserving if for any measurable set  $B \in \mathcal{B}$ ,*

$$P(T^{-1}(B)) = P(B) \tag{A.1}$$

### A.2 Properties of Random Variables and Processes

**Theorem 8.** *Suppose a sequence of Gaussian random variables  $Y_1, Y_2, \dots$ , converges in distribution to a random variable  $Y$ . Then  $Y$  must be Gaussian.*

**Definition 8.** *A set of random variables  $\{Y_s\}_{s \in S}$  is said to be uniformly square integrable if*

$$\lim_{\lambda \rightarrow \infty} \sup_{s \in S} \int_{\{Y_s^2 > \lambda\}} Y_s^2 dP = 0 \tag{A.2}$$

**Theorem 9.** *Suppose  $Y_1, Y_2, \dots$ , is a sequence of uniformly square integrable random variables which converge almost surely to a random variable  $Y$ . Then  $\lim_{n \rightarrow \infty} \text{var}(Y_n) = \text{var}(Y)$ .*

**Theorem 10.** *Suppose  $Y_1, Y_2, \dots$ , is a sequence of Gaussian random variables, whose second moments are uniformly bounded by some constant  $C$ . Then  $Y_1, Y_2, \dots$ , are uniformly square integrable.*

**Corollary 4.** *Suppose  $Y_1, Y_2, \dots$ , is a sequence of Gaussian random variables, whose second moments are uniformly bounded by some constant  $C$ . Further suppose that  $Y_1, Y_2, \dots$ , converge almost surely to a random variable  $Y$ . Then  $\lim_{n \rightarrow \infty} \text{var}(Y_n) = \text{var}(Y)$ .*

**Theorem 11.** *Let  $Y \sim N(\mu, \sigma^2)$ . Then,*

$$E(Y - \mu)^k = \begin{cases} \sigma^k (k+1)!! & k \text{ is even} \\ 0 & k \text{ is odd} \end{cases} \quad (\text{A.3})$$

and

$$E|Y - \mu|^k = \sigma^k 2^{\frac{k}{2}} \frac{\Gamma(\frac{k+1}{2})}{\sqrt{\pi}}. \quad (\text{A.4})$$

### A.2.1 Martingales

This section is largely taken from [21].

Let  $(\Omega, \mathcal{F})$  be a measurable space and  $T$  a subset of  $\mathbb{R}$  or  $\mathbb{Z}$ .

**Definition 9.** *A collection of  $\sigma$ -subalgebras  $\mathcal{F}_t \subset \mathcal{F}$ ,  $t \in T$ , is called a filtration if  $\mathcal{F}_s \subset \mathcal{F}_t$  for all  $s \leq t$ .*

**Definition 10.** A random variable  $X_t$  is adapted to a filtration  $\mathcal{F}_t$  if  $X_t$  is  $\mathcal{F}_t$  measurable for each  $t \in T$ .

**Definition 11.** A family  $(X_t, \mathcal{F}_t)_{t \in T}$  is called a martingale if the process  $X_t$  is adapted to the filtration  $\mathcal{F}_t$ ,  $X_t \in L^1(\Omega, \mathcal{F}, P)$  for all  $t$ , and

$$X_s = E(X_t | \mathcal{F}_s) \text{ for } s \leq t. \quad (\text{A.5})$$

Uniform square integrability of a martingale  $(X_t, \mathcal{F}_t)_{t \in T}$  is equivalent to the uniform square integrability of the sequence of random variables  $(X_t)_{t \in T}$ .

**Theorem 12** (Doob's Inequality). Let  $(X_t, \mathcal{F}_t)_{t \in T}$  be a submartingale. Let  $S_n = \max_{1 \leq t \leq n} X_t$  be the running maximum of  $X_t$ . Then for any  $l > 0$ ,

$$P(S_n > l) \leq \frac{1}{l} E[X_n^+ 1_{\{S_n \geq l\}}] \leq \frac{1}{l} E[X_n^+], \quad (\text{A.6})$$

where  $X_n^+ = X_n \vee 0$ .

Doob's Inequality implies that for martingales, a uniform bound on second moments with respect to the martingale index implies uniform square integrability.

**Theorem 13** (Doob). Let  $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$  be an  $L^1(\Omega, \mathcal{F}, P)$ -bounded martingale.

Then

$$\lim_{n \rightarrow \infty} X_n = Y \quad (\text{A.7})$$

almost surely, where  $Y$  is some random variable from  $L^1(\Omega, \mathcal{F}, P)$ .

### A.3 Spatial Random Field

Let  $s \in S$  be a point in a  $d$ -dimensional Euclidean space and suppose  $W(s)$  is a random quantity. Let  $s$  vary over the index set  $S \subset \mathbb{R}^d$  as to generate a multivariate random field:

$$\{W(s) : s \in S\}$$

Usually, the index set  $S$  is assumed fixed and this is the scheme that we consider in our work. Further,  $S$  can be assumed a fixed subset of  $\mathbb{R}^d$ , or  $S$  can be a (regular or irregular) collection of countably many points in  $\mathbb{R}^d$ , in which case we refer to the process  $W(\cdot)$  as *lattice data*. In this study, we assume that the process  $W(\cdot)$  is Gaussian, i.e. all of its finite-dimensional distributions are multivariate Gaussian. Let

$$W_n = \{W(s) : s \in S_n \subset S\} \tag{A.8}$$

be a sample from  $W(\cdot)$ . To be able to draw conclusions about  $W(\cdot)$  from  $W_n$ , we need to make simplifying assumptions. A common simplifying assumption is that the probabilistic structure in some sense looks similar in different parts of  $\mathbb{R}^d$ . One way to define this concept is through *weak stationarity* of  $W(\cdot)$ .

**Definition 12** (Weak stationarity). *A random function  $W(s)$  satisfying*

$$E(W(s)) = \mu, \quad \forall s \in S \tag{A.9}$$

and

$$\text{cov}(W(s), W(t)) = C_\theta(s - t), \quad \forall s, t \in S \tag{A.10}$$

is defined to be weakly stationary. Furthermore, if  $C_\theta(s-t)$  is a function only of the Euclidean distance  $\|s-t\|$ , then  $C_\theta(\cdot)$  is called isotropic. We denote by  $\theta$  the parameters of  $C_\theta(\cdot)$ .

**Definition 13** (Autocovariance function). *The quantity  $C_\theta(s-t) = \text{cov}(W(s), W(t))$  is called the autocovariance function.*

**Definition 14** (Strict stationarity). *Consider the finite-dimensional distribution*

$$F_{s_1, \dots, s_m}(w_1, \dots, w_m) = P(W(s_1) \leq w_1, \dots, W(s_m) \leq w_m).$$

*If  $F_{s_1+h, \dots, s_m+h}(w_1, \dots, w_m) \equiv F_{s_1, \dots, s_m}(w_1, \dots, w_m)$  for all  $s_1, \dots, s_m$ , with  $m \geq 1$  and all  $h \in S$ , then  $W(\cdot)$  is called strictly stationary.*

Since a Gaussian distribution is uniquely defined by its first two moments, the concepts of *weak* and *strict stationarity* coincide in this case.

### A.3.1 Kriging

*Spatial prediction*, or *interpolation*, refers to predicting an unobserved  $W(s_0)$  at location  $s_0$  from observed data  $W_n$ . *Kriging* is the minimum-mean-squared-error method of spatial prediction. Label the predictor as  $p(W_n; s_0)$ . Assume that the covariance function of  $W(\cdot)$  is known and fixed.

**Definition 15** (Kriging variance). *The quantity  $\sigma_e^2(s_0) = E(W(s_0) - p(W_n; s_0))^2$  is called the mean-squared prediction error, or the kriging (prediction) variance.*

**Theorem 14.** *Suppose we observe  $W_n$  and would like to predict  $W(s_0)$  at an unobserved location  $s_0$ . The predictor  $p(W_n; s_0)$  which minimizes  $\sigma_e^2(s_0)$  is*

$$p(W_n; s_0) = E(W(s_0)|W_n) \tag{A.11}$$

**Theorem 15.** *Suppose  $W(\cdot)$  is a Gaussian random field. Then  $E(W(s_0)|W_n)$  is linear in  $W_n$ .*

**Definition 16** (Ordinary kriging). *Let  $\epsilon(\cdot)$  be a mean-zero random field, defined on  $S$ . Ordinary kriging refers to kriging under the following two assumptions:*

- $W(s) = \mu + \epsilon(s)$ ,  $s \in S$ ,  $\mu$  unknown.
- $p(W_n; s_0) = \sum_{i=1}^n \lambda_i W(s_i)$      $\sum_{i=1}^n \lambda_i = 1$

The latter condition, that the coefficients of the linear predictor sum up to 1, guarantees uniform unbiasedness, that is  $E(p(W_n; s_0)) = \mu = E(W(s_0))$ . However, when  $W(\cdot)$  is Gaussian, the linear (in  $W_n$ ) predictor and  $E(W(s_0)|W_n)$  coincide and since  $E(W(s_0)|W_n)$  is uniformly unbiased, we do not need the extra condition on the coefficients.

### A.3.2 Gaussian Conditional Autoregressions

This section is taken largely from [14].

Suppose that the random vector  $W_n = \{W(s) : s \in S_n \subset \mathbb{R}^d\}$  has density

$$p_W(w) \propto \exp\left(-\frac{1}{2}w'Qw\right), \quad w \in \mathbb{R}^d \tag{A.12}$$



where  $Q$  is an  $n \times n$  positive definite symmetric matrix. Then,

$$W(s_i)|W_{(i)} \sim N\left(\sum_j \beta_{ij}W(s_j), \kappa_i\right), \quad (\text{A.13})$$

where  $W_{(i)} \equiv \{W(s_j), j \neq i\}$ ,  $\beta_{ii} = 0$ ,  $\beta_{ij} = -Q_{ij}/Q_{ii}$  ( $i \neq j$ ) and  $\kappa_i = 1/Q_{ii} > 0$ . The symmetry of  $Q$  requires that

$$\beta_{ij}\kappa_j = \beta_{ji}\kappa_i. \quad (\text{A.14})$$

When the specification of  $p_W$  is based on (A.13) and thus on the precision matrix  $Q$ , rather than the covariance matrix  $\Sigma$ , it is usually referred to as a *conditional autoregressive* or *auto-Normal* formulation [?]. Gaussian conditional autoregressions have been used in a wide range of applications: human geography, agricultural field experiments, geographical epidemiology, texture analysis and other forms of image processing [14].

Practical applications of Gaussian conditional autoregressions often involve random variables distributed on a regular lattice. Examples include image analysis, where sites represent pixels, and crop experiments, where they equate to plots in the field [14]. Let  $s = (x, y) \in \mathbb{Z}^2$  denote the sites of an infinite regular lattice and suppose that  $\{W(s) : s \in \mathbb{Z}^2\}$  is a stationary Gaussian random field with conditional moments as in (A.13),

$$\begin{aligned} E(W(x, y)|W_{-(x,y)}) &= \sum_{k,l} \eta_{kl}w(x - k, y - l), \\ \text{var}(W(x, y)|W_{-(x,y)}) &= \kappa > 0 \end{aligned} \quad (\text{A.15})$$

where  $W_{-(x,y)} \equiv \{W(t) : t \in \mathbb{Z}^2 \setminus (x, y)\}$  and where

*i.*  $\eta_{00} = 0$ ,

*ii.* the number of nonzero  $\eta_{kl}$ 's is finite,

*iii.*  $\eta_{kl} = \eta_{-k,-l}$ ,

*iv.*  $\sum_{k,l} \eta_{kl} \cos(\omega_1 k + \omega_2 l) < 1$  for all  $\omega_1$  and  $\omega_2$ .

Condition (*iii.*) replaces (A.14) and (*iv.*) that of positive definiteness. Besag and Kooperberg [14] show that the autocovariance function for a stationary random field with conditional moments as in (A.15) is

$$C_\theta(k, l) = \frac{\kappa}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\cos(\omega_1 k + \omega_2 l) d\omega_1 d\omega_2}{1 - \sum_{kl} \eta_{kl} \cos(\omega_1 k + \omega_2 l)}, \quad (\text{A.16})$$

where integrability is ensured by (*iv.*). In order to obtain the restriction of a particular infinite lattice autoregression to a finite array, it remains to identify the conditional means and variances at its boundary  $\mathcal{B}$ ; that is, at sites that have missing neighbors with respect to the infinite system. The problem can be solved in principle by using (A.16) to calculate all  $C_\theta(k, l)$  relevant to the finite array and then invert the corresponding covariance matrix  $\Sigma$  to obtain the precision matrix  $Q$ . However, the numerical integration of (A.16) is usually exceedingly delicate because the moderate to substantial correlations that are typical in practical applications occur when  $\eta_{++}$  is close to 1. Published results are available only for first-order autoregressions, for which  $\eta_{10} = \eta_{-10}$  and  $\eta_{01} = \eta_{0-1}$  are the sole non-zero coefficients.

### A.3.3 Mean Square Continuity and Integrability

Stein [12] argues that when interpolating at locations that are surrounded by neighboring observations, the local behavior of the random field is of primary importance. This local behavior is described in terms of the mean square properties of the random field.

**Definition 17** (Mean-square continuity). *Suppose  $W(\cdot)$  is a random field on  $\mathbb{R}^d$ . Then  $W(\cdot)$  is mean-square continuous at  $s$  if*

$$\lim_{t \rightarrow s} E\{W(t) - W(s)\}^2 = 0. \quad (\text{A.17})$$

For  $W(\cdot)$  weakly stationary with autocovariance function  $C(\cdot)$ ,  $E\{W(t) - W(s)\}^2 = 2\{C(0) - C(t - s)\}$ , so that  $W(\cdot)$  is mean square continuous at  $s$  if and only if  $C(\cdot)$  is continuous at the origin. Since a weakly stationary random field is either mean square continuous everywhere or nowhere, we can say that  $W(\cdot)$  is mean square continuous if and only if  $C(\cdot)$  is continuous at the origin.

**Definition 18** (Mean-square differentiability). *A random process  $W(\cdot)$  is said to have the mean-square derivative  $W'(t)$  at a point  $t \in S$ , provided*

$$\lim_{h \rightarrow 0} E \left\{ \frac{W(t+h) - W(t)}{\Delta t} - W'(t) \right\}^2 = 0. \quad (\text{A.18})$$

**Theorem 16.** *A weakly stationary random field  $W(\cdot)$  is either mean square differentiable everywhere or nowhere. It is mean square differentiable everywhere if and only if  $C''(0)$  exists and is finite, where  $C(\cdot)$  is the autocovariance function of  $W(\cdot)$ .*

By repeated application of Thm 16, it follows that  $W(\cdot)$  is  $m$ -times mean square differentiable if and only if  $C^{(2m)}(0)$  exists and is finite.

### A.3.4 Spectral Domain

The material in this section is taken largely from [11].

Spectral methods are a powerful tool for studying the spatial structure of spatial continuous processes and sometimes offer significant computational benefits. Using the spectral representation of a process, we can easily construct valid (*positive definite*) covariance functions. Likelihood approaches for large spatial datasets are often very difficult to implement, due to computational limitations. Even when we can assume normality, exact calculations of the likelihood for a Gaussian spatial process observed at  $n$  locations requires  $O(n^3)$  operations. The spectral version of the Gaussian log likelihood for gridded data requires  $O(n \log_2 n)$  operations and does not involve calculating determinants.

In this section, we offer a review of the Fourier transform and introduce the spectral representation of a stationary spatial process. We also present Bochner's theorem to obtain the spectral representations of a covariance function, and, in particular, of a  $d$ -dimensional isotropic covariance function. We also describe some commonly used classes of spectral densities.

### A.3.4.1 Fourier Transform

A Fourier analysis of a spatial process, also called a harmonic analysis, is a decomposition of the process into sinusoidal components. The coefficients of these sinusoidal components are the Fourier transform of the process. Suppose that  $g$  is a real or complex-valued function that is integrable over  $\mathbb{R}^d$ . Define:

$$G(\omega) = \int_{\mathbb{R}^d} g(s) \exp(i\omega^t s) ds. \quad (\text{A.19})$$

The function  $G$  is said to be the Fourier transform of  $g$ . Then, if  $G$  is integrable over  $\mathbb{R}^d$ ,  $g$  has the representation

$$g(s) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} G(\omega) \exp(-i\omega^t s) d\omega. \quad (\text{A.20})$$

Euler's identity,  $\exp(-i\phi) = \cos(\phi) + i \sin(\phi)$ , explains why (A.20) is a decomposition of  $g$  into sinusoidal components. When  $d = 2$ , we call  $\omega$  a spatial frequency. The right-hand side of (A.20) is called the Fourier integral representation of  $g$ . The functions  $g$  and  $G$  are said to be a Fourier transform pair. It is often useful to think of functions and their transform as occupying two domains. In our case, the domain of  $g$  is often called the time (space) domain and the domain of  $G$ , the frequency domain. Operations performed in one domain have corresponding operations in the other. For example, the convolution operation in the time (space) domain becomes a multiplication operation in the frequency domain. Such results allow one to move between domains so that operations can be performed where they are easiest or most advantageous.

### A.3.4.2 Spectral Representation of a Continuous Spatial Process

Consider a mean square continuous weakly stationary process  $Z(s)$ , indexed by  $\mathbb{R}$ , with mean 0 and autocovariance function  $C(\cdot)$ . For all such processes, it turns out there can be assigned a complex-valued process  $Y(\omega)$  with orthogonal increments, such that we have for each fixed  $s$ , the following stochastic integral that gives the spectral representation

$$Z(s) = \int_{\mathbb{R}^d} \exp(i\omega^t s) dY(\omega). \quad (\text{A.21})$$

The process  $Y$  is called the spectral process associated with a stationary process  $Z$ . It has the following properties:

$$E(Y(\omega)) = 0 \quad (\text{A.22})$$

$$E[(Y(\omega_3) - Y(\omega_2))(Y(\omega_1) - Y(\omega_0))] = 0, \text{ when } \omega_3 < \omega_2 < \omega_1 < \omega_0$$

An analogous representation to (A.21) exists for a stationary process  $Z(\cdot)$  defined on  $\mathbb{Z}^d$ , in which the domain of integration in (A.21) is  $(-\pi, \pi]^d$ .

### A.3.4.3 Bochner's Theorem

Let  $F$  be a positive finite measure. Consider the function  $C(\cdot)$ , defined as:

$$C(s) = \int_{\mathbb{R}^d} \exp(is^t \omega) F(d\omega). \quad (\text{A.23})$$

Bochner's theorem states that a continuous function  $C(\cdot)$  is nonnegative definite if and only if it can be represented in the form above where  $F$  is a positive definite finite measure. Thus the correlation structure of  $Z$  can be analyzed

with a spectral approach or equivalently by estimating the autocovariance function.

If  $F$  has a density with respect to the Lebesgue measure, this density is the spectral density  $f$ . When the spectral density exists, if the covariance function  $C$  is a continuous function, we have the Fourier inversion formula

$$f(\omega) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp(-i\omega^t x) C(x) dx. \quad (\text{A.24})$$

For the rest of the discussion, we assume that  $f$  exists. Thus  $F(d\omega)$  can be replaced with  $f(\omega)d\omega$ .

#### A.3.4.4 Spectral Representation of Isotropic Covariance Functions

If the  $d$ -dimensional process  $Z$  is isotropic with continuous covariance  $C$  and spectral density  $f$ , then for  $\mathbf{h} = (h_1, \dots, h_d)$ , we have  $C(\mathbf{h}) = C_0(\|\mathbf{h}\|)$ , where  $\|\mathbf{h}\| = [h_1^2 + \dots + h_d^2]^{\frac{1}{2}}$ , for some function  $C_0$  of a univariate argument. We denote  $\|\mathbf{h}\|$  as  $h$ . If the random field  $Z$  has a spectral density  $f(\omega)$ , it can be determined from the known covariance function  $C$  using Bochner's theorem. In particular, when  $d = 2$ , we have

$$f(\omega) = \frac{1}{2\pi} \int_0^\infty J_0(\omega h) h C_0(h) dh \quad (\text{A.25})$$

where  $J_\nu(\cdot)$  denotes the Bessel function of the first kind of order  $\nu$ . A  $d$ -dimensional isotropic covariance function with  $d > 1$  is also a covariance

function of some real stationary random process. Therefore, in looking for examples of isotropic covariance functions, we can examine only the real functions  $C(s)$ . To check whether or not the given function  $C$  is a  $d$ -dimensional isotropic covariance function, one only needs to obtain the corresponding spectral density and examine whether this function is everywhere nonnegative.

### A.3.5 Equivalence of probability measures

This section is largely taken from [11].

Suppose we get to observe a random object  $X$ , generated by a measure  $P_0$ . Roughly speaking,  $P_0$  and  $P_1$  are equivalent if, no matter what value of  $X$  is observed, is it impossible to know for sure which of the two measures is correct. The measures are orthogonal, if no matter what value of  $X$  is observed, it is always possible to determine which measure is correct.

Let us give an example of a class of infinite-dimensional Gaussian measures in which these issues arise. Suppose  $Y(\cdot)$  is a zero mean Gaussian process on the interval  $[0, 1]$  with autocovariance function  $K(h) = \theta e^{-\phi|h|}$ . Let us assume that both  $\theta$  and  $\phi$  are positive, in which case,  $K(\cdot)$  is a positive definite function. For  $j = 0, 1$ , let  $P_j$  be the Gaussian process law with  $(\theta, \phi) = (\theta_j, \phi_j)$ . It is possible to show that if  $\theta_0\phi_0 = \theta_1\phi_1$ , then  $P_0$  and  $P_1$  are equivalent probability measures and otherwise they are orthogonal. Thus, for example, writing  $K_j$  for the autocovariance function under  $P_j$ , if  $K_0(t) = e^{-|t|}$  and  $K_1(t) = 2e^{-|t|/2}$ , then  $P_0$  and  $P_1$  are equivalent. On the other hand, if  $K_0(t) = e^{-|t|}$  and



$K_1(t) = e^{-|t|/2}$ , we can say for sure (with probability 1) which measure is correct.

Suppose  $Y(\cdot)$  is a stationary Gaussian process on  $\mathbb{R}^d$  and the spectral density  $f_0$  satisfies

$$f_0(\omega)|\omega|^\alpha \text{ is bounded away from } 0 \text{ and } \infty \text{ as } |\omega| \rightarrow \infty. \quad (\text{A.26})$$

Define  $G_D(m, K)$  to be the probability measure of a Gaussian process on a domain  $D$  with mean  $m$  and autocovariance function  $K$ . If  $f_0$  satisfies (A.26) and

$$\int_{|\omega|>C} \left\{ \frac{f_1(\omega) - f_0(\omega)}{f_0(\omega)} \right\}^2 d\omega < \infty \quad (\text{A.27})$$

for some  $C < \infty$ , then  $G_D(0, K_0)$  and  $G_D(0, K_1)$  are equivalent on all bounded domains  $D$ . Evidently, these conditions apply only on a bounded domain  $D$  and with mean  $m = 0$ . General conditions for orthogonality of Gaussian measures in more than one dimensions are not so simple. Some are given in Skorohod and Yadrenko [23].

### A.3.5.1 Equivalent measures and kriging, fixed domain asymptotics

The results on equivalent measures are relevant to kriging prediction. Suppose  $P_0$  is the correct model for some Gaussian process on a bounded domain  $D$ , but we instead use an equivalent Gaussian measure  $P_1$  to compute both the kriging predictor and the mean squared errors of these predictors. Stein [24] showed

that as a sequence of observation gets dense in  $D$ , it makes no asymptotic difference whether we use the correct  $P_0$  or incorrect, but equivalent,  $P_1$  to carry out the kriging. This result is summarized in Theorem 17. See Stein [24] for a proof.

**Theorem 17.** *Consider a process  $Y(\cdot)$  defined on a domain  $D$  with finite second moments and  $EY(s) = \beta'm(s)$ . Suppose  $s_0, s_1, \dots$  are in  $D$ ,  $G_D(0, K_0)$  and  $G_D(0, K_1)$  are equivalent probability measures, and  $e_j(s_0, n)$  is the error of the BLUP of  $Y(s_0)$  under  $K_j$  based on  $Y(s_1), \dots, Y(s_n)$ . If  $E_0e_0(s_0, n)^2 \rightarrow 0$  as  $n \rightarrow \infty$ , but  $E_0e_0(s_0, n)^2 > 0$  for all  $n$ , then*

$$\lim_{n \rightarrow \infty} \frac{E_0e_1(s_0, n)^2}{E_0e_0(s_0, n)^2} = 1, \tag{A.28}$$

$$\lim_{n \rightarrow \infty} \frac{E_0e_1(s_0, n)^2}{E_1e_1(s_0, n)^2} = 1, \tag{A.29}$$

The message of this section is that under fixed-domain asymptotics, we get asymptotically equivalent kriging predictors as long as the corresponding measures are equivalent. Thus, under fixed-domain asymptotics, we do not necessarily need consistency of parameter estimates for the corresponding kriging errors to converge to the true errors. If  $s_1, s_2, \dots$  are dense in  $D$ , Stein shows that Theorem 17 holds if the observations include uncorrelated, equal variance measurement errors, as long as the variance of these errors is taken to be the same under  $P_0$  and  $P_1$ .

## A.4 Hypothesis Testing

The material in this section is taken largely from [22].

**Definition 19** (Hypothesis). *A hypothesis is a statement about a population parameter.*

**Definition 20** (Null and Alternative hypotheses). *The two complementary hypotheses in a hypothesis testing problem are called the null hypothesis and the alternative hypothesis. They are denoted as  $H_0$  and  $H_a$ , respectively.*

If  $\theta$  denotes the population parameter, the general format of the null and alternative hypothesis is  $H_0 : \theta \in \Theta_0$  and  $H_a : \theta \in \Theta_0^c$ , where  $\Theta_0$  is a subset of the parameter space and  $\Theta_0^c$  is its complement. We write  $\Theta = \Theta_0 \cup \Theta_0^c$ .

**Definition 21** (Hypothesis test). *A hypothesis test is a procedure which specifies:*

- *For which sample values the decision is made to accept  $H_0$  as true.*
- *For which sample values  $H_0$  is rejected and  $H_a$  is accepted as true.*

**Definition 22** (Rejection region). *The subset  $R$  of the sample space for which  $H_0$  is rejected is called the rejection region. The complement of the rejection region is called the acceptance region.*

Typically, a hypothesis test is specified in terms of a *test statistic*  $\lambda$ , a function of the observed data  $y_n$ . We reject  $H_0$  if  $\lambda \in \mathcal{R}$ , otherwise, we accept  $H_0$ . In the

rest of the discussion, consider the hypothesis test  $H_0 : \theta \in \Theta_0$  vs  $H_a : \theta \in \Theta_0^c$ .

This test might make one of two types of errors:

- If  $\theta \in \Theta_0$ , but the hypothesis test incorrectly decides to reject  $H_0$ , the test has made a *Type I* error.
- If  $\theta \in \Theta_0^c$ , but the hypothesis test incorrectly decides to accept  $H_0$ , the test has made a *Type II* error.

**Definition 23** (Size  $\alpha$  test). *For  $0 \leq \alpha \leq 1$ , a test of size  $\alpha$  is one for which  $P_\theta(Y_n \in R) = \alpha$  for all  $\theta \in \Theta_0$ .*

**Definition 24** (Power). *The power of a hypothesis test is  $P_{\theta \in \Theta_0^c}(Y_n \in R)$ .*

A good test in the class of *size  $\alpha$*  tests would also have a small Type II error probability or large power. A very general method of deriving a test statistic is the *likelihood ratio* method.

**Definition 25** (Likelihood ratio statistic). *The likelihood ratio test statistic for testing  $H_0 : \theta \in \Theta_0$  versus  $H_a : \theta \in \Theta_0^c$  is*

$$LR(y_n) = \frac{\sup_{\Theta_0} L(\theta|y_n)}{\sup_{\Theta} L(\theta|y_n)} \tag{A.30}$$

where  $L(\theta|y_n)$  is the likelihood function of  $\theta$ , given the observed data  $y_n$ .

In some situations, tests for complicated hypotheses can be developed from tests for simpler hypotheses. We discuss one such method here. The *union-intersection method* of test construction might be useful when the null hypoth-

esis is expressed as an intersection, say,

$$H_0 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma. \quad (\text{A.31})$$

Here  $\Gamma$  is an arbitrary index set that may be finite or infinite, depending on the problem. The alternative hypothesis is then written as

$$H_a : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma^c. \quad (\text{A.32})$$

Suppose that tests are available for each of the problems of testing  $H_{0,\gamma} : \theta \in \Theta_\gamma$  vs.  $H_{a,\gamma} : \theta \in \Theta_\gamma^c$ . Say that the rejection region for the rest of  $H_{0,\gamma}$  is  $R_\gamma$ . Then the rejection region for the union-intersection test is  $\bigcup_{\gamma \in \Gamma} R_\gamma$ . The rationale is simple. If any one of the hypotheses  $H_{0,\gamma}$  is rejected, then  $H_0$ , which is true only if  $H_{0,\gamma}$  is true for every  $\gamma$ , is rejected. Only if each of the hypotheses  $H_{0,\gamma}$  is accepted as true, will the intersection  $H_0$  be accepted as true.

Suppose that the test statistic for testing  $H_{0,\gamma}$  is  $g_\gamma(y_n)$  and that the rejection region for  $H_{0,\gamma}$  is  $\{y_n : g_\gamma(y_n) > c\}$ . Then, the rejection region for the union-intersection test can be expressed as

$$\bigcup_{\gamma \in \Gamma} \{y_n : g_\gamma(y_n) > c\} = \left\{ y_n : \sup_{\gamma \in \Gamma} g_\gamma(y_n) > c \right\}. \quad (\text{A.33})$$

Thus, the test statistic for testing  $H_0$  is  $g(y_n) = \sup_{\gamma \in \Gamma} g_\gamma(y_n)$ .

## A.5 Parameter Estimation

We consider a Gaussian random field  $Y(\cdot)$  on  $S \subset \mathbb{R}^d$  with  $Y(s) = X(s)' \beta + \epsilon(s)$ , where  $X(\cdot)$  is a known vector valued function,  $\beta$  is a vector of unknown co-

efficients and  $\epsilon(\cdot)$  is mean 0 with covariance function  $\text{cov}(\epsilon(s), \epsilon(t)) = C_\theta(s, t)$ . Observe  $Y_n \equiv \{Y(s) : s \in S_n \subset \mathbb{R}^d\}$ . The *likelihood function* is just the joint density of the observations viewed as a function of the unknown parameters. A maximum likelihood estimate (MLE) of the unknown parameters is any vector of values for the parameters that maximizes this likelihood function. It is completely equivalent and often easier to maximize the logarithm of the likelihood function, often called log likelihood. Let  $\Sigma_\theta$  be the covariance matrix of  $Y_n$  as a function of  $\theta$  and assume  $\Sigma_\theta$  is non-singular for all  $\theta$ . Define  $X = (X(s_1), \dots, X(s_n))'$  and assume it is of full rank. Then the *log likelihood function* is

$$l(\theta, \beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_\theta| - \frac{1}{2} (Y_n - X\beta)^T \Sigma_\theta^{-1} (Y_n - X\beta). \quad (\text{A.34})$$

Assume further that  $\Sigma_\theta = \sigma^2 D_\gamma$ , where  $\theta = (\sigma^2, \gamma)$ . Let  $\hat{\gamma}$  denote the maximum likelihood estimates of  $\gamma$ , and  $\hat{D}$ , the corresponding maximum likelihood estimate of  $D$ . Then, the maximum likelihood estimates of  $\sigma^2$  and  $\beta$  take the form:

$$\hat{\beta} = (X' \hat{D}^{-1} X)^{-1} X' \hat{D}^{-1} Y_n \quad (\text{A.35})$$

and

$$\hat{\sigma}^2 = \frac{(Y_n - X\hat{\beta})' \hat{D}^{-1} (Y_n - X\hat{\beta})}{n}. \quad (\text{A.36})$$

The estimate in (A.35) is also the *generalized least squares* estimate of  $\beta$ . Ordinary least squares estimation of  $\beta$  assumes that  $\Sigma = I$  and the estimate is gotten by minimizing the sum of squared differences  $(Y_n - X\beta)'(Y_n - X\beta)$ , as a function of  $\beta$ . Generalized least squares estimation assumes correlated data

and the estimate of  $\beta$  is obtained by minimizing the squared estimated Mahalanobis distance between  $Y_n$  and  $X\beta$ :  $(Y_n - X\beta)' \hat{D}^{-1} (Y_n - X\hat{\beta})$ . One way to simplify maximization of (A.34) is to rewrite it as a function of (A.35) and (A.36). The function  $l(\gamma; \hat{\beta}, \hat{\sigma}^2)$  is called the *profile likelihood* for  $\gamma$ . This can be a useful practice when simultaneous maximization of  $l(\theta, \beta)$  proves difficult.

### A.5.1 Estimation of $\sigma^2$

In Section 1.2.1.1, we present the Indicator variable outlier detection algorithm. This algorithm adds Indicator variables to the regressor matrix, one at a time, and re-estimates the current model after each regressor is added. The parameters are estimated via maximum likelihood. Consider the maximum likelihood estimator of  $\sigma^2$  at each step, where  $X$  is the current regressor matrix:

$$\hat{\sigma}_{MLE}^2 = \frac{(Y_n - X\hat{\beta})' \hat{D}^{-1} (Y_n - X\hat{\beta})}{n} \quad (\text{A.37})$$

The problem with the maximum likelihood estimate  $\hat{\sigma}_{MLE}^2$  is that it overestimates  $\sigma^2$ . This is because at each intermediate stage of the outlier detection algorithm,  $X$  is an incomplete matrix, in that it accounts only for some, but not all, of outliers in the data. As a result, the vector  $Y_n - X\hat{\beta}$  has non-zero mean and its expected length is inflated. Below we propose a robust version of  $\hat{\sigma}_{MLE}^2$ , which dampens the effect of  $Y_n - X\hat{\beta}$  on the estimate of  $\sigma^2$ . For ease of exposition, consider the vector  $Z_n$ , with

$$Z_n \sim N(0, \sigma^2 D). \quad (\text{A.38})$$

Assume  $D$  is known. The MLE of  $\sigma^2$  in this case is

$$\hat{\sigma}_{MLE}^2 = \frac{Z'D^{-1}Z}{n}. \quad (\text{A.39})$$

Now consider the Cholesky decomposition of  $D^{-1} = R'R$ , where  $R$  is an upper triangular matrix. We then have  $D = R^{-1}(R^{-1})'$ . The numerator  $Z'D^{-1}Z$  of (A.38) can then be written as

$$Z'D^{-1}Z = Z'R'RZ = (RZ)'RZ = \epsilon'\epsilon. \quad (\text{A.40})$$

where  $\epsilon = RZ$  has  $\text{var}(\epsilon) = R\sigma^2DR' = \sigma^2RR^{-1}(R^{-1})'R' = \sigma^2I$ . So,

$$\hat{\sigma}_{MLE}^2 = \frac{Z'D^{-1}Z}{n} = \frac{1}{n} \sum_{i=1}^n \epsilon_i^2. \quad (\text{A.41})$$

Following the robustifying approach taken with time series [2], we propose the following robust estimator of  $\sigma^2$ :

$$\hat{\sigma}_{MAD1}^2 = 1.5 \text{median}\{|\epsilon_i|\} \quad (\text{A.42})$$

where MAD stands for Median Absolute Deviation. This estimator replaces the average of squared deviations in (A.41) with a scaled median of absolute deviations. As a result, the effect of undetected outliers on  $\hat{\sigma}^2$  is dampened. Alternatively, consider the eigenvalue-eigenvector decomposition of  $D = P\Lambda P'$ , where  $\Lambda$  is the diagonal matrix of eigenvalues and  $P$  is the matrix of eigenvectors. Since  $P'P = I = PP'$ , we have  $D^{-1} = P\Lambda^{-1}P'$ . The numerator  $Z'D^{-1}Z$  of (A.38) can then be written as

$$Z'D^{-1}Z = Z'P\Lambda^{-1}P'Z = (\Lambda^{-\frac{1}{2}}P'Z)'(\Lambda^{-\frac{1}{2}}P'Z) = \xi'\xi \quad (\text{A.43})$$



where  $\xi = \Lambda^{-\frac{1}{2}}P'Z$  has  $\text{var}(\xi) = \Lambda^{-\frac{1}{2}}P'\sigma^2DP\Lambda^{-\frac{1}{2}} = \sigma^2\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} = \sigma^2I$ .

Therefore,

$$\hat{\sigma}_{MLE}^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 \quad (\text{A.44})$$

Thus, another natural robust estimator of  $\sigma^2$  is

$$\hat{\sigma}_{MAD2}^2 = 1.5 \text{median}\{|\xi_i|\} \quad (\text{A.45})$$

## Bibliography

- [1] Haining, R. *Spatial Data Analysis: Theory and Practice*. Cambridge, UK. 2003.
- [2] Bell, W. “A Computer Program for Detecting Outliers in Time Series”. In *American Statistical Association Proceedings of the Business and Economic Statistics Section*, pages 634-639, 1983.
- [3] Fox, A.J. “Outliers in Time Series”. *Journal of the Royal Statistical Society, Series B*, 43:350-363, 1972.
- [4] Chang, I., Tiao, G. and Chen, Chung. “Estimation of Time Series Parameters in the Presence of Outliers”. *Technometrics*, 30:193-204, 1988.
- [5] Ljung, G. “On Outlier Detection in Time Series”. *Journal of the Royal Statistical Society, Series B*, 55:559-567, 1993.

- [6] Ledolter, J. “The Effect of Additive Outliers on the Forecasts from ARIMA Models”. *International Journal of Forecasting*, 5:231-240, 1989.
- [7] Abraham, B. and Box, G.E.P. “Bayesian Analysis of Some Outlier Problems in Time Series”. *Biometrika*, 66:229-236, 1979.
- [8] Gomez, V. and Maravall, A. “Programs TRAMO and SEATS: Instructions for the User”. Working Paper 97001, Ministerio de Economia y Hacienda, Direccion General de Analisis y Programacion Presupuestria. 1997.
- [9] Bell, W., Hillmer, S. and Tiao, G. “Modeling Considerations in the Seasonal Adjustment of Economic Time Series”. *Applied Time Series Analysis of Economic Data, U.S. Department of Commerce, U.S. Census Bureau*, pages 74-100, 1983.
- [10] X-12-ARIMA Reference Manual, Version 0.2.10. 2002. Available at <http://www.census.gov/srd/www/x12a/>.
- [11] “Handbook of Spatial Statistics”. Chapman Hall. New York. 2010.
- [12] Stein, M. “Interpolation of Spatial Data”. Springer. New York. 1999.
- [13] Rozanov, Yu. A. “On the Gaussian Homogeneous Fields with Given Conditional Distribution”, *Theory Probability Appl.*, 12:381-391, 1967.
- [14] Kooperberg, C. and Besag, J. “On Conditional and Intrinsic Autoregression”, *Biometrika*, 85:733-746, 1995.
- [15] Cressie, N. “Statistics for Spatial Data”. John Wiley & Sons. New York. 1991.

- [16] Taylor, J., Worsley, K.J. and Gosselin, F. “Maxima of Discretely Sampled Random Fields, with an Application to ‘bubbles’”. *Biometrika*, pages 1-18, 2007.
- [17] McElroy, T. and Holan, S. “Asymptotic Theory of Cepstral Random Fields”. *Pre-print*. arXiv:112.1977.
- [18] Zhang, H. “Inconsistent Estimation and Asymptotically Equal Interpolation in Model-Based Geostatistics”. *Journal of American Statistical Association*. 2004.
- [19] Mardia, K. and Marshall, R. J. “Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression”. *Biometrika*, 71: 135-146, 1984.
- [20] Guyon, Xavier. “Estimation of a Stationary Process for a  $d$ -dimensional Lattice”. *Biometrika*. 69:95-105, 1982.
- [21] Koralov, L. and Sinai, Ya. “Theory of Probability and Random Processes”. Springer. New York. 2007.
- [22] Casella, G. and Berger, R. “Statistical Inference”. Duxbury Press. Belmont, California, 1990.
- [23] Skorohod, A.V. and Yadrenko, M.I. “On Absolute Continuity of Measure Corresponding to Homogeneous Gaussian Fields”. *Theory of Probability and Its Applications*. 18: 27-40. 1973.
- [24] Stein, M. “Asymptotically Efficient Prediction of a Random Field with a Misspecified Covariance Function”. *Annals of Statistics*. 16: 55-63. 1988.

- [25] Kent, J. and Mardia, K. “Spectral and Circulant Approximations to the Likelihood for Stationary Gaussian Random Fields”. *Journal of Statistical Planning and Inference*. 50: 379-394, 1996.