USC SIPI Report #134

Parametric Spectrum Estimation for Contaminated Random Fields

Richard Russell Hansen, Jr.

December 7, 1988

Dedication

To my wife, Carolyn.

She encouraged me to begin this work and then gave of herself so that

I might complete it.

To my son, Rusty.

I hope this work is an inspiration for him to accomplish good things.

To the loving memory of my mother,

Mary Elizabeth Hansen.

She instilled in me the courage to pursue a seemingly impossible task.

Acknowledgements

I can not convey enough appreciation to Professor Rama Chellappa, my advisor and dissertation committee chairman, for giving me the opportunity to contribute this work. His support and friendship were without qualification. His enthusiasm for and understanding of image and signal processing, statistics, mathematics, and engineering is unsurpassed and he is a mentor for me to emulate.

I would like to thank the members of my dissertation committee, Professor Alexander Sawchuk and Professor Louis Gordon for committing their time to guide and evaluate my work.

A special thanks goes to Dr. Allan Weber. His remarkable knowledge of the computer and software systems at the Signal and Image Processing Institute was the key to my successful battle with those beasts of technology. He unselfishly gave his time and effort to fix many apparent system problems, often a result of my own ignorance, and create new software to accomplish minor cosmetic changes in my published work for the sake of what he probably thought was a whim.

I must extend my appreciation to the staff of the Signal and Image Processing Institute, especially to Erlinda Varilla who kept the money coming, to Gloria Bullock and Jerene Carey who went out of their way to provide typing and administrative assistance, and to Toy Mayeda who kept me awake on many a sleepy morning with his "coffee club".

I must thank my fellow graduate students, both past and present, at the Signal and Image Processing Institute. They are a great group, a literal melting pot of backgrounds, ideas, and interests. In particular, I want to thank my friend and past roommate, Professor Shankar Chatterjee, who helped me with understanding many key concepts in image processing. Dr. Jinchi Hao, Professor Georgios Gianakis, Yitong Zhao, Quinfen Zheng, Anand Rangarajan, Dr. Chellappa's students, the guys on second floor and the other SIPI students deserve mention for being friends and establishing a quality environment in which to learn and discover.

I would like to acknowledge the support of the Office of Naval Research through Contract N00014-86-K-0302.

Contents

List of Figures

List of Tables

Abstract

In this dissertation we propose and investigate new methods of parametric spectrum estimation for two-dimensional random fields which are adequately represented with spatial interaction models. The maximum likelihood (ML) estimator is considered optimal for strictly Gaussian fields and, because of its invariance property, yields the ML estimate of the spectrum. However, many observed fields do not conform to strict distributional assumptions because of inherent contamination (noise) or other isolated imperfections (outliers) resulting from the data measurement and recording processes.

When the observed signal is Gaussian we show that a noncausal autoregressive signal plus noise model that accounts for possible contamination yields better spectrum estimates than conventional signal only models. The theoretical properties of the ML parameter estimates for noncausal autoregressive plus noise model are stated and proved, and the numerical properties are experimentally evaluated and compared with the conventional signal only model for direction of arrival estimation in planar array signal processing.

When the Gaussian assumption is suspect we propose robust techniques for estimating the parameters of spatial interaction model spectra. First, we extend the time series generalized M-estimator to two-dimensional nonsymmetric half-plane and Gaussian-Markov random field models, and then we analyze this

robust estimator's performance in a series of parameter and spectrum estimation experiments.

The generalized M-estimator is essentially a heuristic weighted LS procedure with no optimality property. Therefore, we formulate an optimal robust problem and solve it for the nonsymmetric half-plane model. The result is an optimal robust estimator which performs significantly better than conventional LS and ML when confronted with data having various contaminations.

The generalized M-estimator will not perform well for noncausal autoregressive models. Additionally, the covariance and conditional probability structure of noncausal models preclude a solution to the optimal robust problem. Thus, we work directly from the observed data and propose empirical estimators for the noncausal autoregressive and Gaussian-Markov random field models. The robust empirical estimators performance in the contaminated situation is shown through a series of experiments to be better than the performance of optimal methods based on a strict Gaussian assumption.

Chapter 1

Introduction

This dissertation presents research on model based spectrum estimation techniques for imperfectly observed two-dimensional signals. The observations or, equivalently, data are assumed to be a two-dimensional (2-D) random field [1, 2, 3, 4]. Therefore, the general problem addressed is that of determining the spectral density function of a 2-D random field given a finite set of observations from that field. If we denote the autocovariance function of a real stationary (homogeneous) random field $\mathcal{X} = \{x(k,l) \mid \text{integers } k\&l \in (-\infty, \infty)\}$ by $\mathrm{E}\{x(k,l)x(i,j)\} = q_x(k-i, l-j)$ for all integers k, l, i, and j and where $\mathrm{E}\{\cdot\}$ denotes expectation, then the Fourier transform pair

$$q_x(k,l) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathcal{S}_x(\omega_1, \omega_2) \exp[\jmath(\omega_1 k + \omega_2 l)] d\omega_1 d\omega_2 \tag{1.1}$$

and

$$\mathcal{S}_x(\omega_1, \omega_2) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} q_x(k,l) \exp[-\jmath(\omega_1 k + \omega_2 l)] \tag{1.2}$$

describes the relationship between the covariance function $q_x(k,l)$ and the spectral density function or spectrum $\mathcal{S}_x(\omega_1, \omega_2)$, where ω_1 and ω_2 are spatial frequency variables.

Given exact knowledge of $q_x(k,l)$ for all k and l the problem would be to simply compute the spectrum from (1.2). However, exact knowledge of $q_x(k,l)$ implies one of two things, we either have $x(k,l)$ for all k and l, from which we can compute $q_x(k,l)$, or we have an exact 2nd-order description of the probability distribution $\mathcal{F}_x$ of $x(k,l)$. In all practical cases we have neither. Usually, we have only a finite set of data $\mathcal{X}_{MN} = \{x(k,l) \mid 0 \leq k \leq M-1 \text{ and } 0 \leq l \leq N-1\}$ and at best a functional description of an approximate $\mathcal{F}_x$. Consequently, the spectrum must be "estimated" from the finite observation set $\mathcal{X}_{MN}$.

This problem has additional complications if we truly want the spectrum of a process or signal which cannot be observed directly. That is, if $\mathcal{Y}$ denotes the signal of interest and we only have a set of observations $\mathcal{X}_{MN} = \{x(k,l) = g(y(k,l),\eta(k,l)) \mid 0 \leq k \leq M-1 \text{ and } 0 \leq l \leq N-1\}$, where $g(y,\eta)$ denotes a scalar function of y and η, then good estimates of $\mathcal{S}_x(\omega_1,\omega_2)$ may not be sufficient since $\mathcal{S}_\eta(\omega_1,\omega_2)$ can taint $\mathcal{S}_x(\omega_1,\omega_2)$ and completely obscure $\mathcal{S}_y(\omega_1,\omega_2)$. The estimation methods investigated in this dissertation concentrate on this aspect of the 2-D spectrum estimation problem.

1.1 Notation and Other Conventions

In this dissertation boldface lowercase letters, $\mathbf{r}$, $\mathbf{s}$, $\mathbf{t}$ etc., and subscripted boldface lowercase letters, e.g., $\mathbf{s}_i$, will denote elements of the 2-D $M \times M$ square lattice $\Omega_M \subset \mathcal{Z}^2$, $\mathcal{Z}^2$ is the set of all integer pairs k,l such that k and l are elements of the closed set $[-\infty,\infty]$. Hence, $\Omega_M = \{\mathbf{s} = (i,j) \mid 0 \leq i,j \leq M-1\}$.

Let $\mathbf{s} = (i,j)$ and $\mathbf{t} = (k,l)$, then $\mathbf{s} < \mathbf{t}$ if and only if $i < k$ or $i = k$ and $j < l$, and $\mathbf{s} = \mathbf{t}$ if and only if $i = k$ and $j = l$. Moreover, $i < j$ if and only if $\mathbf{s}_i < \mathbf{s}_j$,

and $i = j$ if and only if $\mathbf{s}_i = \mathbf{s}_j$. If $\mathbf{r} = (i,j)$ and $\mathbf{s} = (k,l)$, then by $\mathbf{t} = \mathbf{s} + \mathbf{r}$ we mean $\mathbf{t} = (m,n)$ where $m = i + k$ and $n = j + l$.

We write the summation of a function $g : G \times \Re^m \rightarrow \Re^n$, $G \subset \Re^2$, over all lattice points in Ω_M by

$$\sum_{\mathbf{s}\in\Omega_M} g(\mathbf{s}) \text{ or } \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} g(i,j) \text{ or } \sum_{k=0}^{M^2-1} g(\mathbf{s}_k),$$

which are equivalent.

Column vectors will be be denoted by boldface roman and Greek letters, which may include subscripts and superscripts, such as $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{x}_i$, $\boldsymbol{\theta}$ etc. For example, let N be a finite set whose elements are $\mathbf{r}_i$, $i = 1,\ldots,p$, where $\mathbf{r} \neq (0,0)$ but otherwise arbitrary. We will often make use of the p-dimensional column vector $\mathbf{x}_\mathbf{s} = \text{col } \{x(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$ and the $(p+1)$-dimensional column vector $\mathbf{x}_\mathbf{s}^0 = \text{col } \{x(\mathbf{s}), \mathbf{x}_\mathbf{s}\}$ with $N_0 = N \cup (0,0)\}$. The transpose of a column vector $\mathbf{x}$ is $\mathbf{x}^t$, which is a row vector. An asterisk $(*)$ will denote the complex conjugation.

Matrices will usually be denoted by capital letters such as A, B, Q etc. with possible subscripts, Q_x, for example. Superscripts on matrices will usually denote a matrix operation such as inversion, A^{-1}. When the individual elements of a $n \times n$ matrix A are required to be explicitly displayed we will use the notation $A = \text{mat } \{a(i,j),\ i,j = 1,\ldots,n\}$.

Calligraphic letters will be used for several purposes: spectral density functions as in $\mathcal{S}_x(\boldsymbol{\omega})$ where $\boldsymbol{\omega} = (\omega_1,\omega_2)$ and $-\pi \leq \omega_1,\omega_2 < \pi$, probability distribution functions such as in $\mathcal{F}(x)$, and sets of observations or data such as in $\mathcal{X}_M = \{x(\mathbf{s}), \mathbf{s} \in \Omega_M\}$.

Additonal notation and deviations from these conventions will be described where necessary.

1.2 Motivation and Research Objectives

1.2.1 Practical Applications

Spectrum estimation has much application in the sciences and engineering. Here we highlight three practical problems which require good spectral estimates.

Direction Finding with Planar Arrays.

An important use of spectral density functions is that of determining the wave-number or direction of arrival of plane waves propagating in space and impinging on an array of receiving sensors. This important estimation problem is common to radar, sonar, seismology, radio astronomy, and other similar array processing applications. Consequently, an extensive part of signal processing research conducted today is directed towards finding improved spectrum estimation methods or optimizing existing techniques. We describe this problem here for sources of narrow-band signals and show how the spectrum of the array output signal can be used to determine the directions of arrival.

Consider Figure 1.1 which shows a $M \times M$ array of receiving antennae equally distributed in a square lattice pattern over a plane region in space. Let $\mathbf{s} = (i, j)$ denote the i, jth sensor of the array and d be the distance between any two sensors in the same row or column. Impinging on the array are L plane waves generated by narrow band sources with the same carrier frequency f_c which are all located on one side of the array and at distinct angles ϕ_k, θ_k, and γ_k relative to the array plane as shown in the figure. Each source is assumed to be at a sufficient distance from the array that only far field effects are important. Thus, the wave from each

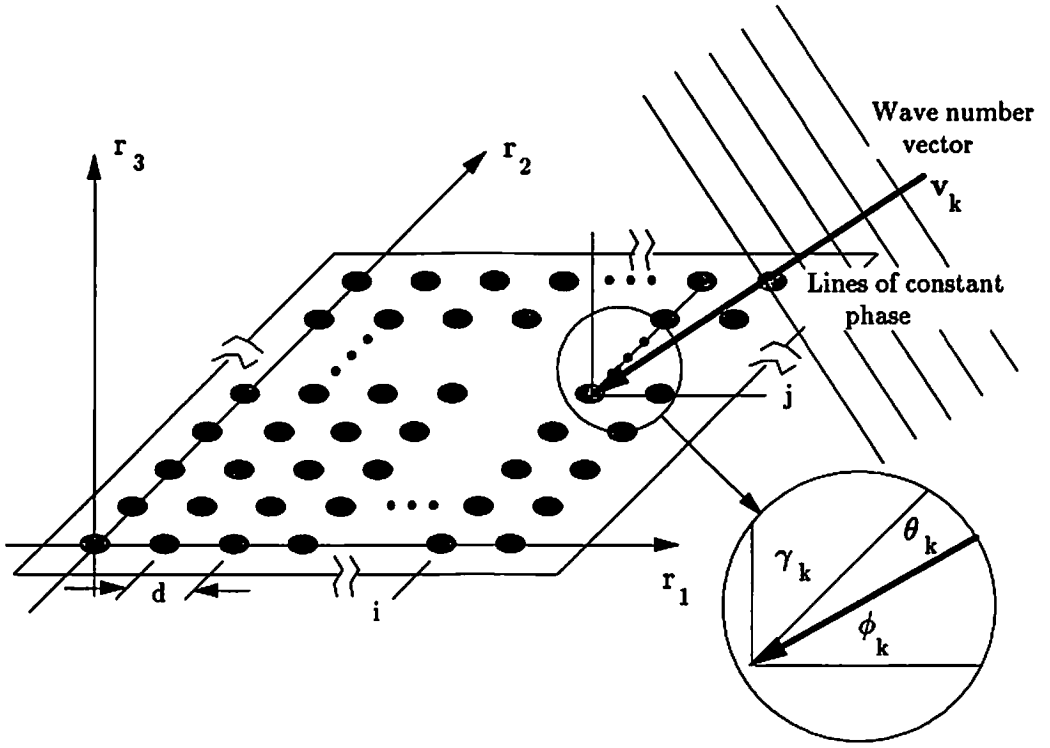


Figure 1.1: Geometry of planar array with a ray of an incident plane wave.

source is essentially planar as it impinges on the array. Figure 1.1 shows a ray of the kth incident plane wave.

Assuming that the array lies in the r_1, r_2-plane ($r_3 = 0$) of a Cartesian coordinate system with the origin of the array at the coordinate system center, the kth traveling plane wave [5] can be represented by

$$g(t, \mathbf{r}) = A_k \cos[2\pi(f_c t + \mathbf{v}_k^t \mathbf{r})],$$

where

$$\mathbf{v}_k = \frac{1}{\lambda_c} \begin{bmatrix} \cos(\phi_k) \\ \cos(\theta_k) \\ \cos(\gamma_k) \end{bmatrix}$$

is the wave number vector, λ_c the wave length, and A_k the amplitude of the kth incident plane wave.

$$\mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}$$

is a vector from the origin of the coordinate system to a point in space.

Using this notation the vector from the array origin to the i,jth sensor is $\mathbf{r_s} = d \cdot [i\ j\ 0]^t$. The vector product of $\mathbf{v}_k$ and $\mathbf{r_s}$ may be expressed as

$$\mathbf{v}_k^t \mathbf{r_s} = \frac{d}{\lambda_c}[i \cdot \cos(\phi_k) + j \cdot \cos(\theta_k)],$$

and $s_k(\mathbf{r_s}, t)$, the noiseless signal produced at the input of the i,jth sensor due to the kth plane wave, is

$$s_k(\mathbf{r_s}, t) = A_k \cos\left\{2\pi f_c t + \frac{2\pi d}{\lambda_c}[i \cdot \cos(\phi_k) + j \cdot \cos(\theta_k)] + \alpha_k\right\},$$

where α_k is the phase of the kth plane wave measured at the origin. The signal at the input of the i,jth sensor due to all plane waves is

$$S(\mathbf{s}, t) = \sum_{l=1}^{L} s_l(\mathbf{r_s}, t).$$

However, the signal at the output of each sensor differs from $S(\mathbf{s}, t)$ because of noise from the sensor as well as noise from undesirable signals impinging on the array. This noise, in general, can be represented by a narrow-band signal [5] given by

$$\eta(\mathbf{s}, t) = B_s \cos(2\pi f_c t + \beta_s),$$

where B_s is the amplitude and β_s is the phase of the noise at the i,jth sensor. B_s is Rayleigh distributed and β_s is distributed uniformly over the interval $(0, 2\pi)$.

Hence, the true i,jth sensor output, denoted by $x(\mathbf{s},t)$, is

$$x(\mathbf{s},t) = S(\mathbf{s},t) + \eta(\mathbf{s},t).$$

We note that $x(\mathbf{s},t)$ is a function of both the sensor location and the time variable. At any instant in time, say $t = t_0$, we have the output of M^2 sensors, called a *snapshot*, which we denote as $\mathcal{X}_M(t_0) = \{x(\mathbf{s},t_0), \mathbf{s} \in \Omega_M\}$. If the array and the signal sources are relatively stationary over a period of time, then several snapshots can be taken and used for computational purposes. On the other hand, one would also like to be able to determine the direction cosines for each signal source from only one snapshot. That is the problem we concentrate on here. Thus, we take $t_0 = 0$ without loss of generality, and the observed signal at the output of the i,jth sensor is written simply as

$$x(\mathbf{s}) = S(\mathbf{s}) + \eta(\mathbf{s}), \tag{1.3}$$

where the signal $S(\mathbf{s})$ is a sum of L cosinusoidal signals given by

$$S(\mathbf{s}) = \sum_{l=1}^{L} A_l \cos(\boldsymbol{\varphi}_l^t \mathbf{s} + \alpha_l). \tag{1.4}$$

$\boldsymbol{\varphi}_l = \text{col}\,\{\varphi_{1l}, \varphi_{2l}\}$ is the direction cosine vector with

$$\begin{aligned} \varphi_{1l} &= \frac{2\pi d}{\lambda_c}\cos(\phi_l) \\ \varphi_{2l} &= \frac{2\pi d}{\lambda_c}\cos(\theta_l), \end{aligned}$$

and $\{\eta(\mathbf{s})\}$ is a random Gaussian noise field.

The direction finding problem is to estimate the parameters $\boldsymbol{\varphi}_l$, $l = 1,\ldots,L$, in equation (1.4) from a finite set of observations $\mathcal{X}_M = \{x(\mathbf{s}), \mathbf{s} \in \Omega_M\}$, which is one snapshot of planar array data. The power spectrum of $\{S(\mathbf{s})\}$, the signal of

interest, consists of L impulses located at spatial frequencies equal to φ_l. Thus, the 2-D power spectrum can be used to estimate the direction of arrival of the L plane waves. The noise spectrum, however, is also included in the spectrum of the observations. When $\eta(\mathbf{s})$ is independent of $S(\mathbf{s})$ the spectrum of the observations is

$$\mathcal{S}_x(\omega) = \mathcal{S}_S(\omega) + \mathcal{S}_\eta(\omega) \tag{1.5}$$

for $\omega = (\omega_1, \omega_2)$, $-\pi \leq \omega_1, \omega_2 < \pi$. Consequently, if the signal-to-noise ratio is low, the interesting details of $\mathcal{S}_S(\omega)$ may be obscured by $\mathcal{S}_\eta(\omega)$.

The direction finding problem is solved by detecting and resolving peaks in the spectrum which correspond to the direction of arrival of the plane waves impinging on the array. The problem is complicated by the fact that in practical applications the received signals are corrupted by noise from several sources. This noise degrades the estimated spectrum, obscuring the fine structure and making it difficult to determine and localize the spectral peaks corresponding to the signals. Generally speaking, researchers have concentrated on the one-dimensional (1-D) problem, i.e., estimation of signal spectra from linear array data [5]. The 2-D or planar array problem has usually been approached by adapting 1-D techniques to the 2-D situation [6, 7]. This is especially true for the model based approach where 2-D parametric modeling and spectrum estimation reported in the literature have assumed known covariances [8] or used strictly causal (quarter-plane or nonsymmetric half-plane) 2-D models [9, 10]. Since causality in two dimensions has little meaning, causal models usually require large model orders for good performance and result in estimated spectra with directional biases and extraneous spectral structure. To eliminate this phenomena the extended techniques need to employ some method of combining multiple estimates to eliminate the biases

and smooth the spectrum [11]. One would suspect that the elimination of the causality constraint might permit procedures with improved performance. For example, Sharma and Chellappa showed in [12] that the noncausal autoregressive (NCAR) model having far fewer parameters provided results comparable to those obtained in [13] where a causal model was used. It is generally agreed that applicable models with fewer parameters (the concept of parsimony) are preferable in any model based procedure.

Texture Analysis.

Texture analysis is an important area of study in image processing where the spectrum is a useful tool [14]. The spectrum of a texture is computed and various characteristics (or features) are obtained from the spectrum. These features are used to describe or characterize the texture for the purposes of pattern classification and analysis. For a reliable classifier an accurate spectrum is required. Known classification procedures assume that the texture is observed cleanly or preprocessing of the texture is performed, i.e., the data is filtered before computing the spectrum followed by feature extraction. Most commonly occurring textures are conveniently described by Gaussian random field models [15, 16]. However, the observed process, in this case the recorded texture image, may be contaminated with imperfections during the recording process. These imperfections can be regarded as additive noise. As one example, consider a texture image recorded on film: the grain of the film emulsion is a noise added to the real texture signal. In this case the observed data may not be Gaussian even though the texture is accurately described by the Gaussian model. Consequently, an accurate estimate of the signal spectra is required using contaminated texture data.

Image Restoration.

Consider the $M \times M$-pixel image model

$$x(\mathbf{s}) = h(\mathbf{s}) \star y(\mathbf{s}) + n(\mathbf{s}), \quad \mathbf{s} \in \Omega_M$$

where $\mathcal{Y}_M$ is the real image, $\mathcal{N}_M$ is an additive noise independent of $y(\mathbf{s})$, $h(\mathbf{s})$ is an impulse response function associated with a spatial degradation such as blurring or motion, and "$\star$" denotes convolution. Given the observation $\mathcal{X}_M$ of the image the restoration problem is to form an estimate of $\mathcal{Y}_M$. A common performance criteria for the restoration is to minimize the mean square error [17], $e_{MSE} = \mathrm{E}\{[y(\mathbf{s}) - \hat{y}(\mathbf{s})]^2\}$, where $\hat{y}(\mathbf{s})$ is the real image pixel estimate. It is straightforward to derive the minimizing solution to this problem and show that the minimum variance filter (Wiener filter) resulting from the solution is

$$W(\omega) = \frac{H^*(\omega)}{|H(\omega)|^2 + \frac{\mathcal{S}_n(\omega)}{\mathcal{S}_y(\omega)}}$$

where $H(\omega)$ is the Fourier transform of $h(\mathbf{s})$. Exact knowledge of $\mathcal{S}_n(\omega)$ and $\mathcal{S}_y(\omega)$ is required for optimal results.

In practice one only has an estimate of $\mathcal{S}_n(\omega)$ and a nominal model for $\mathcal{S}_y(\omega)$. Thus, the best restoration may not be achieved. For example, Kassam and Poor [18] have shown that a "trivial filter", which corresponds to all-pass filtering if the SNR is positive and no-pass filtering if the SNR is negative, will often do better than the optimal $W(\omega)$ filter when the true $\mathcal{S}_n(\omega)$ and $\mathcal{S}_y(\omega)$ deviate from their estimates.

1.2.2 Contamination in Observed Data

In the foregoing examples, as in much of signal processing, a Gaussian assumption is often made regarding the observed data since this assumption usually reduces the complexity of the problem from both theoretical and empirical standpoints. The assumption of normality is often based on empirical evidence or justified in theory by application of a suitable central limit theorem. But in practical empirical situations, the observed signals contain undesirable imperfections or noise which is inherent to the system under study or which occur because of measurement errors or isolated phenomena. In many situations the corrupting noise itself can be considered Gaussian with the result that the observations remain Gaussian but with a more complicated structure. However, measurement errors and isolated errors can cause observed data sets to contain small fractions of unusual data points, or "outliers", which are not consistent with a strictly Gaussian assumption. Such data in principle can be modeled as having a distribution which is nearly Gaussian in the central region but with heavier tails. In other situations the rounding or grouping caused by finite bit quantization and computation of signals can also be viewed as a signal measurement error. Then the observed data is distributed as though it were Gaussian near the mean but having no tails at all.

A simple model for a contaminated random field is $x(\mathbf{s}) = y(\mathbf{s}) + n(\mathbf{s})$, $\mathbf{s} \in \Omega_M$. Two commonly occurring situations are 1) the innovative outlier (IO) model where $n(\mathbf{s}) \equiv 0$ for all Ω_M but $y(\mathbf{s})$ is a non-Gaussian signal and 2) the additive outlier model where $n(\mathbf{s}) \neq 0$ for a small fraction of Ω_M but $y(\mathbf{s})$ is Gaussian. Other situations exist, for example, $n(\mathbf{s})$ may replace $y(\mathbf{s})$, substitutive outliers (SO), or $n(\mathbf{s})$ may be correlated with $y(\mathbf{s})$.

Martin and Thomson [19] have suggested that isolated measurement errors or outliers in time series be modeled with the mixture distribution

$$\mathcal{F} = (1-h)\Delta(0) + h\mathcal{N}(0,\sigma^2)\,. \tag{1.6}$$

Here $\Delta(0)$ is the degenerate distribution having all its mass at zero and $\mathcal{N}(0,\sigma^2)$ is the standard normal distribution with mean zero and variance σ^2. The 2-D IO model will be said to hold whenever, $x(\mathbf{s}) = y(\mathbf{s})$ for all Ω_M and the innovations process deviates from a nominal Gaussian distribution. For example, the innovations may have a heavy-tailed non-Gaussian common distribution which results from the sum of a normal random variable and a random variable distributed according to (1.6). The AO model will occur if $n(\mathbf{s})$ has the distribution given by (1.6). Then the signal is observed correctly most of the time, i.e., $x(\mathbf{s}) = y(\mathbf{s})$, but $100\,h$ percent of the time $y(\mathbf{s})$ is observed with error. We note that this contamination leads to a non-Gaussian heavy-tailed distribution for the $x(\mathbf{s})$, although for small h, $\mathcal{F}(x)$ will be nearly $\mathcal{F}(y) = \mathcal{N}(0,\sigma_y^2)$. In general, these are only two of several possible contaminations, and for unknown h the distribution $\mathcal{F}(x)$ is also unknown regardless of whether it is associated with the innovations or an additive effect.

1.2.3 Spectrum Estimation–Past and Present

Spectrum estimation most likely had its origins in the 17th century work of Sir Isaac Newton [20, 21]. Since that time the subject has been undertaken by many renowned scientists and engineers. Almost as many approaches to a solution have been investigated as there have been investigators. The 1982 *IEEE Proceedings* paper by Robinson [20] provides an excellent "historical perspective of spectrum

estimation" and the 1981 *IEEE Proceedings* paper by Kay and Marple [22] summarizes many of the new techniques developed in the last two decades. Two collections of papers, [23] and [24] contain much of the current thought relating to spectrum estimation and its application in practical situations. Two recent books [21] and [25] are good exposes on the fundamentals of spectrum estimation and analysis, especially their application to time series.

Work on the 2-D problem, i.e., spectrum estimation for random fields, is comparatively sparse. It turns out that large parts of the 2-D spectrum estimation and analysis work are attempts at extending 1-D techniques to the 2-D problem [7]. For instance, The Fourier based methods, in which the fast Fourier transform (FFT) and periodogram play such a vital part, extend directly with extra care taken to choose smoothing windows that give an everywhere positive spectrum [26]. Techniques have also been developed that employ 1-D estimators directly by processing the 2-D data along each dimension. These estimators are usually called separable estimators [6]. The Capon maximum likelihood method (MLM) [10] and Pisarenko's method [7], so often associated with spectral analysis of periodic time series, have also been extended to the 2-D situation. The autoregressive (AR) time series model has also found application to spectrum estimation of random fields in the form of the quarter-plane (QP) and nonsymmetric half-plane (NSHP) random field models [11, 8, 9]. By lexicographically ordering the lattice data the 2-D results are very similar to the 1-D AR spectrum estimator.

In one dimension, the AR spectrum estimator is equivalent to the maximum entropy (MEM) estimator [27], but the correspondence does not carry over to two dimensions, a consequence of the fact that 2-D polynomials need not factorize. Moreover, the notion of causal and anti-causal ordering is not meaningful in two

dimensions. However, a 2-D maximum entropy problem can be formulated and a solution determined. The 2-D MEM estimate is the spectrum maximizing the entropy of a Gaussian random field subject to a correlation matching constraint [7]. Several methods have been suggested for the nonlinear MEM optimization problem [28].

In [29] and [30] the MEM spectrum estimate has been shown to be equivalent to the spectrum of a stationary Gaussian-Markov random field (GMRF). Thus, the MEM spectrum estimate can be obtained by computing the GMRF spectrum [31].

Only recently have 2-D NCAR random field models received attention as 2-D spectrum estimators [12].

1.2.4 Model Based Spectrum Estimation

In model based spectrum estimation, a parametric model, typically a constant coefficient difference equation driven by a random input, is proposed to represent the observed data. An estimate of the spectrum is obtained by fitting the model to the observed data (estimating parameters) and using the parameter estimates in the theoretical spectrum expression derived from the model.

The derivation of the model theoretical spectrum is an exercise in system theory and is relatively straight forward. Consequently, there are two paramount issues in model based spectrum estimation:

1. Selection of an appropriate model and

2. Selection of a parameter estimation method.

Both items are critical for good spectrum estimates as will be borne out by the experimental results in this dissertation. The model is often suggested by the mechanism which generates the observed process. For example, if the mechanism is cyclic in structure then the model should be able to accurately represent a cyclic process. However, the model also dictates, to a large degree, the estimation method required for accurate parameter estimates. In general, parameter estimators should yield consistent and asymptotically efficient estimates. But often estimators with these characteristics require significant computational and memory storage capacity. Hence, the model selection is often one of convenience, yielding optimal results when all assumptions (e.g., normality) are satisfied and possessing an estimator which can be implemented with minimum use of computing resources.

In engineering, especially in the field of image processing, random field models have found wide application. In particular, the spatial interaction models [32, 2, 4, 33] are popular. This class of models includes the Markov random field models [34], of which the GMRF model [35, 36] is of interest, and also the class of simultaneous models [1, 32]. The simultaneous model class includes the QP, NSHP, and NCAR models [4, 37, 16], which are the generalizations to two dimensions of the familiar time series autoregressive models. Model based spectrum estimation using a variety of these models has been evaluated and good results obtained for data which matches the distributional assumptions about the observed data [30, 31, 12].

But what happens when the assumptions are not satisfied, especially when the fine structure of the spectrum is of primary interest? Large deviations from the assumptions are obviously going to cause problems. But in this case one did a

predictably poor job of model selection and could do better by choosing a second model. It is not clear, though, what happens when large deviations occur in a tiny fraction of the data because of errors in observation or when small deviations occur in all the data because of strict distributional assumptions. Note that all measured data are of limited accuracy and are basically discrete; and, therefore, can only be described approximately by a continuous distribution. In these last two situations the true signal can be thought of as having been generated by an assumed nominal model but observed with error.

The concept of robust parameter estimation has received extensive consideration in the statistical literature. A major portion of this literature (e.g., see [38] and [39]) treats location and linear regression models for independent and identically distributed (IID) data. The literature on robust spectrum estimation in the dependent or autoregressive time series case is sparse. Kleiner et al. [40, 41, 42] have developed several robust estimation techniques for the 1-D case when the data are represented by an autoregressive model. Very little, if any, work has been published for 2-D robust model based spectrum estimation; and the 1-D techniques are not directly extendable without considering the unique properties of 2-D models.

Two general approaches seem to be prevalent in the statistical literature. The first, Huber's minimax approach [39], attacks the problem by considering so-called "M-estimators", which are a generalization of maximum likelihood (ML) estimators for location and scale of independent data. Huber's method is to optimize the worst that can happen in a neighborhood of a nominal model, as measured by the asymptotic variance of the M-estimator. The second, proposed by Hampel [38] and often leading to estimators similar to Huber's, is known as

the "infinitesimal approach" whose main tool is the so-called "influence function". The influence function is a directional derivative which describes the effect of an additional observation on a statistic (or parameter estimate). The influence function can be used to linearize an estimator and predict its performance in a neighborhood of a nominal model.

1.3 Thesis

When the primary objective is to estimate the spectrum of $\mathcal{Y}$ given the observed data $\mathcal{X}_M$ by the model based methods, care must be taken to choose a model that accurately represents the observed data. In most previous works [6, 11, 8, 13, 9, 31, 12], models have been proposed for $\mathcal{Y}$ and estimates made from $\mathcal{X}_M$ with an assumption that the signal $y(\mathbf{s})$ is adequately represented by a spatial interaction model, either a spatial autoregressive or Markov random field, driven by a Gaussian random noise. Clearly, a better job of estimation can be done if one accounts for the differences between $\mathcal{X}_M$ and $\mathcal{Y}$. Thus, instead of assuming that $x(\mathbf{s}) = y(\mathbf{s})$ for all $\mathbf{s} \in \Omega_M$, consider a model for the observations given by

$$x(\mathbf{s}) = g(y(\mathbf{s}), n(\mathbf{s}))$$

for some function $g(\cdot, \cdot)$.

If one knows *a priori* the distributions of both $y(\mathbf{s})$ and $n(\mathbf{s})$ then this information can be used, for example, to write the likelihood function for the observations $\mathcal{X}_M$, thus leading to an optimal ML estimator. Consider the case when $g(y(\mathbf{s}), n(\mathbf{s})) = y(\mathbf{s}) + n(\mathbf{s})$, where the distributions for both $y(\mathbf{s})$ and $n(\mathbf{s})$ are known to be Gaussian and $\mathcal{Y}$ is adequately modeled by a spatial interaction model. If $\mathcal{X}_M$ contains contamination then this model should perform better for

spectrum estimation than the conventional $g(y(\mathbf{s}), n(\mathbf{s})) = y(\mathbf{s})$. If $\mathcal{X}_M$ is error free then little is lost since the parameter estimates should suggest that the mean and variance of $n(\mathbf{s})$ are zero. For the 2-D direction finding problem we claim that a NCAR plus noise model will perform better than a NCAR only model [12, 13] because:

1. The error, $e(\mathbf{s}) = S(\mathbf{s}) - y(\mathbf{s})$, in fitting the model can be included in $n(\mathbf{s})$, thereby reducing the complexity of the NCAR model required to duplicate the nonrandom $S(\mathbf{s})$, and

2. The spectrum $\mathcal{S}_y(\omega)$ approximating $\mathcal{S}_S(\omega)$ can be computed explicitly without the obscuring $\mathcal{S}_\eta(\omega)$ or $\mathcal{S}_n(\omega)$.

If, on the other hand, one can only make an assumption concerning the distribution and structure of the random field $\mathcal{Y}$ and knows neither the distribution of $n(\mathbf{s})$ nor the functional relationship $g(\cdot,\cdot)$ relating $x(\mathbf{s})$ to $y(\mathbf{s})$ and $n(\mathbf{s})$, then strict assumptions on the Gaussianity of $x(\mathbf{s})$ can lead to predictably poor spectrum estimation results. Robust alternatives which consider a nominal model for $\mathcal{X}$ should perform better than optimal procedures requiring a strict Gaussian assumption.

1.4 Contributions

The main contribution resulting from the research reported in this dissertation is the synthesis and analysis of new model based spectrum estimation methods for 2-D random fields. The methods reported here perform significantly better under a wide variety of practical empirical situations than the model based methods

previously reported in the technical literature. These contributions include the following:

- A new NCAR plus noise model is defined and applied for 2-D random field spectrum estimation. Specifically, we make the following contributions:
  - The NCAR plus noise model is shown to differ from the linear NCAR model because of its affine structure.
  - The theoretical properties of the NCAR plus noise model are stated and proved.
  - The NCAR plus noise is shown to outperform the NCAR only model in a variety of practical situations, including the important problem of direction finding associated with planar array signal processing.
- The generalized maximum likelihood (GM) estimator, which has been studied for 1-D time series, is extended to the 2-D problem. We make the following contributions in this regard:
  - The 2-D GM-estimator is defined for both NSHP and GMRF models.
  - We define an influence function for the 2-D GM-estimator and show how to use it as a tool for selecting the critical parameters of the estimator that control its performance.
  - We suggest an algorithm for solving the 2-D GM-estimator equations.
  - We apply the 2-D GM-estimator to the 2-D spectrum estimation problem and compare its performance with the conventional methods.
- We define and evaluate an optimal robust estimator for NSHP random field models. Specific contributions include:

- We generalize the familiar 2-D likelihood function for spatial interaction models and use this generalization to define a general class of estimators for 2-D random field models.
- An influence function applicable to a wide variety of contaminated data problems is defined and used to define Hampel's optimality problem for the NSHP model.
- Hampel's optimality problem is solved for the NSHP model. The result is an optimal robust estimator.
- The optimal robust estimator is evaluated through a series of experiments and shown to be superior to the conventional estimator when observation data contain contamination.

- Empirical approaches are suggested for noncausal model estimation. These models are shown through a series of experiments to yield results superior to conventional methods of parameter estimation from contaminated data.

1.5 Organization of Dissertation

In Chapter 2 we present a short review of the theory of robust statistics. In Chapter 3 we describe the two general classes of spatial interaction models, the GMRF and NCAR models, which are useful for 2-D spectrum estimation. Here we give the details of the conventional means of parameter estimation and derive special model characteristics required in later chapters. Chapter 4 presents the signal plus noise model for spectrum estimation. We discuss the attributes of this model and investigate the theoretical bases of the ML parameter estimates for the NCAR

plus noise model. The utility of the model is demonstrated through a series of experiments in which we simulate planar array data and compute spectrum estimates for comparing the NCAR plus noise model with a conventional NCAR only model. Chapter 5 is the first of three chapters in which we investigate 2-D robust parameter estimation. In this chapter we extend the generalized M-estimator, employed for robustly estimating parameters in time series, to the 2-D situation. We develop the necessary tools to evaluate this heuristic estimator and we compare its performance through a series of experiments with that of conventional parameter estimators. A general notation is derived for the likelihood function of the NCAR and GMRF models and an influence function, useful for designing robust estimators, is derived in Chapter 6. By specializing to the NSHP model, we specify and solve and optimality problem which leads to an optimal robust estimator. Through experiments we compare this estimator with the conventional NSHP estimator and with the 2-D GM-estimator derived in Chapter 5. In the next chapter, Chapter 7, the ideas of Chapters 5 and 6 are used to suggest robust estimators for the noncausal models. These "empirical" estimators are shown to perform much better than the conventional means of parameter estimation when confronted with contaminated signal data. Chapter 8 summarizes the research and suggests topics requiring additional investigation.

Chapter 2

Robust Statistics

In this chapter we define robustness and give an overview of the theory of robust statistics and its application to parameter estimation in time series models.

Robustness signifies insensitivity to small deviations from assumptions. Here, we will primarily be concerned with distributional robustness where the shape of the true underlying distribution which generated observed data deviates slightly from an assumed model. The first use of this notion of robustness may properly be attributed to G.E.P. Box although eminent statisticians, such as Newcomb, K. Pearson, Gosset, Jeffreys, and E.S. Pearson, were clearly aware of it [38].

There are three notions of robustness which have been introduced in the literature: 1) *efficiency robustness* which requires that an estimator have a high efficiency, say ≥ 90 percent, at a nominal distribution, and high efficiency at a variety of strategically chosen distributions [43], 2) *minimax robustness* introduced by Huber [39] which requires that a robust estimator's asymptotic variance be minimized at the worst model in a neighborhood of models, and 3) *qualitative robustness* introduced by Hampel [44, 38] which requires that large changes in a

small amount of the data and small changes in a large amount of the data both produce only small changes in an estimate. Qualitative robustness is a continuity criteria where infinitesimal changes are measured through the use of an appropriate metric on the space of distributions for the data and the estimate, respectively. It will be this last notion of Hampel's that will be of particular concern to us. This infinitesimal theory, which has the influence function as its central tool for analysis and synthesis of algorithms, yields useful and computationally attractive procedures for spectrum estimation. In general, however, a good robust estimator should clearly possess at least qualitative and quantitative (efficiency) robustness, with minimax robustness highly desirable.

Strictly speaking a robust procedure should have the following properties: 1) when the data are "good", e.g., Gaussian, the procedure should be almost as good as the conventional (often "optimal") procedure presuming normality, and 2) when outliers are present, the procedure should still work well and, in particular, work much better than the conventional procedure.

2.1 The Influence Function

Suppose we have a set of observations $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ which are independent and identically distributed. The empirical distribution of the sample will be denoted by $\mathcal{F}_n(x)$ and is given by

$$\mathcal{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \Delta(x_i) \tag{2.1}$$

where $\Delta(x)$ is the distribution which places point mass 1 at x. We will assume that the true underlying distribution which generated the data is $\mathcal{F}(x)$ and that

$$\lim_{n\to\infty} \mathcal{F}_n(x) = \mathcal{F}(x)\,. \tag{2.2}$$

An explanation of the notation is appropriate here. $\Delta(x)$ denotes the probability distribution for the random variate X. This function is mathematically equivalent to the generalized unit-step function,

$$\Delta(x) = \begin{cases} 0, & X < x \\ 1, & X \geq x \end{cases}$$

We will denote the generalized unit-impulse function as $\delta(x)$ given by

$$\delta(x) = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Consider some *parametric model* $\{\mathcal{F}_\phi;\ \phi \in \Phi\}$ characterized by a family of distributions $\mathcal{F}_\phi$ parameterized by ϕ which belongs to some p-dimensional parameter space Φ. In classical estimation we assume that the observations $\mathcal{X}_n$ come from one of the $\mathcal{F}_\phi$ (e.g., $\mathcal{F}(x) = \mathcal{F}_{\phi_0}$), then estimate ϕ_0 based upon a finite set of observed data. In the robust approach we assume that the parametric model $\{\mathcal{F}_\phi;\ \phi \in \Phi\}$ is only an approximation to reality and that $\mathcal{F}(x)$ may not equal $\mathcal{F}_\phi$ for any ϕ. The task, however, is still to determine a value of ϕ which in some robust sense is "optimal". For example, $\mathcal{F}_\phi$ may be the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ where $\phi = \text{col}\ \{\mu, \sigma\}$, but the data x_i are generated by a non-normal distribution.

As an estimator for ϕ consider a sequence of real-valued statistics $T_n = T_n(x_1, \ldots, x_n) = T_n(\mathcal{F}_n)$, where the form of the estimator is data independent

and can be replaced by functionals, i.e., $T_n(\mathcal{F}_n(x)) = T(\mathcal{F}_n(x))$. Furthermore, assume that the functional is "Fisher consistent", i.e.,

$$T(\mathcal{F}_\phi) = \phi\,.$$

In otherwords, at the model when $\mathcal{F}(x) = \mathcal{F}_\phi$ the estimator asymptotically measures the right quantity.

In Hampel's qualitative approach to robust statistics he begins by defining the influence function.

Definition 2.1 *The* (p-dimensional) influence function (IF) *of* T *at a distribution* $\mathcal{F}$ *is given by*

$$IF(x;T,\mathcal{F}) = \lim_{t\to 0} \frac{T\left((1-t)\mathcal{F} + t\Delta(x)\right) - T(\mathcal{F})}{t}$$

for all x *where the limit exists.*

The heuristic interpretation of this function is that it describes the standardized effect of an infinitesimal contamination at the point x on the estimate T. The IF is usually evaluated at the model distribution $\mathcal{F}_\phi$.

Von Mises has shown [45] that T may be expanded at a distribution $\mathcal{F}$ in terms of a Taylor series. If distribution $\mathcal{G}$ is "near" $\mathcal{F}$, then

$$T(\mathcal{G}) = T(\mathcal{F}) + \int IF(x;T,\mathcal{F})d(\mathcal{G}-\mathcal{F})(x) + \text{remainder.} \qquad (2.3)$$

If we replace $\mathcal{G}$ by $\mathcal{F}_n(x)$ in (2.3) and note that

$$\int IF(x;T,\mathcal{F})d\mathcal{F}(x) = 0\,,$$

then

$$T(\mathcal{F}_n) \approx T(\mathcal{F}) + \int IF(x;T,\mathcal{F})d\mathcal{F}_n(x) + \text{remainder.}$$

or using (2.1) to evaluate the integral, then rearranging,

$$\sqrt{n}(T_n - T(\mathcal{F})) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(x_i;\ T, \mathcal{F}) + \text{remainder.}$$

Since the x_i are IID, so also are the $IF(x_i;\ T, \mathcal{F})$; and hence, this term is asymptotically normal by the central limit theorem. In most cases of interest the remainder becomes negligible as $n \to \infty$ so that $\sqrt{n}(T_n - T(\mathcal{F}))$ is asymptotically normal with an asymptotic covariance matrix given by

$$V(T, \mathcal{F}) = \int IF(x; T, \mathcal{F})\, IF^t(x;\ T, \mathcal{F}) d\mathcal{F}(x)\,. \tag{2.4}$$

Rigorous mathematical conditions for this development are available (see [38]); but as Hampel points out the influence function is mainly a heuristic tool, and it is often easier to verify the asymptotic properties by methods other than the necessary regularity conditions required in the above development.

The influence function can also be used to compute the asymptotic Cramer-Rao inequality for a sequence of estimators. That is, if

$$I(\phi_0) = \int \left[\frac{\partial}{\partial \phi} \log f_\phi(x)\right] \left[\frac{\partial}{\partial \phi} \log f_\phi(x)\right]^t_{\phi=\phi_0} d\mathcal{F}_0(x) \tag{2.5}$$

is the Fisher information matrix at a distribution $\mathcal{F}_0(x) = \mathcal{F}_{\phi_0}$, where $\phi_0 \in \Phi$ is fixed, then

$$\mathbf{d}^t V(T, \mathcal{F}_0)\, \mathbf{d} \ \geq\ \mathbf{d}^t I^{-1}(\phi_0)\, \mathbf{d} \ \text{ for all } \mathbf{d} \in R^p \tag{2.6}$$

where equality holds if and only if

$$IF(x; T, \mathcal{F}_0) \text{ is proportional to } \left.\frac{\partial}{\partial \phi} \log f_\phi(x)\right|_{\phi=\phi_0}.$$

Here, f_ϕ is the density function corresponding to $\mathcal{F}_\phi$. It then follows that the estimator is asymptotically efficient if

$$IF(x; T, \mathcal{F}_0) = I^{-1}(\phi_0) \left.\frac{\partial}{\partial \phi} \log f_\phi(x)\right|_{\phi=\phi_0}.$$

2.2 Robustness Measures

Although the influence function is useful for evaluating the asymptotic properties of a sequence of estimators, its real utility comes from the several robustness measures which can be derived from it. These properties were introduced by Hampel et al. [38].

Gross-error sensitivity.

Definition 2.2 *The* (unstandardized) gross-error sensitivity *of T at a distribution $\mathcal{F}$ is given by*

$$\gamma^* = \sup_x \|IF(x; T, \mathcal{F})\| .$$

where $\|\cdot\|$ is the Euclidean norm.

The gross-error sensitivity measures the worst influence that a small amount of contamination can have on the value of an estimator. Consequently, we can regard it as an upper-bound on the asymptotic bias. It seems intuitively clear that it should be finite for all good robust procedures.

Local-shift sensitivity.

Definition 2.3 *The* local-shift sensitivity *of T at a distribution $\mathcal{F}$ is given by*

$$\lambda^* = \sup_{x \neq y} \frac{\|IF(y;\, T, \mathcal{F}) - IF(x; T, \mathcal{F})\|}{|y - x|} .$$

Not only should robust procedures be insensitive to gross errors, but small fluctuations in the data, such as rounding or grouping, should also have minimal effect. The local-shift sensitivity describes the worst effect of a small deviation about a

point x, or equivalently, the addition of an observation at y and removal of one at x.

Rejection point.

Definition 2.4 *The* rejection point *of T at a distribution $\mathcal{F}$ is given by*

$$\rho^* = \inf_r \{r > 0 \mid IF(x; T, \mathcal{F}) = 0 \text{ when } |x| > r\}$$

if r exists, otherwise $\rho^* = \infty$.

A classical robustness concept is to reject extreme points entirely. Often statisticians manually prune a data set, throwing way points which obviously do not belong due to their relative magnitude, prior to performing a statistical procedure. This pruning concept is embodied in the rejection point, since if it exists for an estimator T, i.e., $\rho^* < \infty$, then all observations farther away than ρ^* are rejected completely by the estimator and have no "influence" on the estimate.

Breakdown point. By definition the influence function is a local concept usually evaluated at the model distribution $\mathcal{F}_\phi$. Consequently, the preceding robustness measures can only be considered good within a small distance from $\mathcal{F}_\phi$. Thus, to complement these local properties Hampel defines a global property, the breakdown point. Several definitions are possible, although the following one seems easiest to apply and often leads to the same value as the more rigorously defined.

Definition 2.5 *The* asymptotic breakdown point *[39] (or* gross-error breakdown point *[38]) of T at a distribution $\mathcal{F}$ is*

$$\varepsilon^* = \sup_\varepsilon \{\varepsilon \leq 1 \mid b(\varepsilon) < b(1)\}$$

where

$$b(\varepsilon) = \sup_{\mathcal{G} \in P_\varepsilon(\mathcal{F})} |T(\mathcal{G}) - T(\mathcal{F})|$$

and the contamination "neighborhood" $P_\varepsilon(\mathcal{F})$ is given by

$$P_\varepsilon(\mathcal{F}) = \{\mathcal{G} \mid \mathcal{G} = (1-\varepsilon)\mathcal{F} + \varepsilon\mathcal{H}\} .$$

Loosely speaking, ε^* is the smallest fraction of grossly contaminated data points which makes the estimate completely unreliable, and it gives an idea of the distance from the assumed model that the local linearization provided by the influence function can be used. Note that ε^* is not necessarily dependent on the distribution $\mathcal{F}$, and that $b(1)$, the maximum bias of the estimator T in a contamination neighborhood, is the worst possible value of $b(\varepsilon)$ (usually infinite).

2.3 Robust Parameter Estimation

2.3.1 M-estimators

Huber introduced the class of estimators called M-estimators [39, 38], a generalization of classical maximum likelihood.

Definition 2.6 *An* M-estimator *is the value $T_n = T_n(x_i, \ldots, x_n)$ such that*

$$\sum_{i=1}^{n} \rho(x_i, T_n) = \min_{T_n}! \tag{2.7}$$

where ρ is some function on $X \times \Phi$. If ρ has a derivative

$$\psi(x, \phi) = \frac{\partial}{\partial \phi} \rho(x, \phi) ,$$

then T_n satifies the implicit equation

$$\sum_{i=1}^{n} \psi(x_i, T_n) = 0\,. \tag{2.8}$$

Let $\mathcal{G}_n$ be the empirical distribution function generated by the sample $\mathcal{X}_n = \{x_i, \ldots, x_n\}$ and assume that the real-valued statistic T_n can be replaced by the functional $T(\mathcal{G}_n)$, where T is given by

$$\int \psi(x, T(\mathcal{G}))\, d\mathcal{G}(x) = 0\,. \tag{2.9}$$

Now, if $\mathcal{G}$ is a contaminated distribution, that is, $\mathcal{G} = (1-t)\mathcal{F} + t\Delta(x)$, where t is small but fixed, it can be shown by differentiating (2.9) at $t = 0$ that

$$IF(x; \psi, \mathcal{F}) = M^{-1}(\psi, \mathcal{F})\, \psi(x, T(\mathcal{F})) \tag{2.10}$$

where the $p \times p$ matrix M is given by

$$M(\psi, \mathcal{F}) = -\int \frac{\partial}{\partial \phi} \psi(y, \phi)\Big|_{\phi = T(\mathcal{F})} d\mathcal{F}(y)$$

and we have replaced T by the defining ψ-function. Note that M is independent of x and $IF(x; \psi, \mathcal{F})$ is proportional to $\psi(x, T(\mathcal{F}))$. Therefore, for M-estimators the influence function can be used for defining the ψ-function. The asymptotic covariance matrix for M-estimators is

$$V(T, \mathcal{F}) = M^{-1}(\psi, \mathcal{F})\, Q(\psi, \mathcal{F})\, M^{-t}(\psi, \mathcal{F})$$

with

$$Q(\psi, \mathcal{F}) = \int \psi(x, T(\mathcal{F}))\, \psi^t(x, T(\mathcal{F}))\, d\mathcal{F}(x)\,.$$

M-estimators can easily be considered in the context of regression models. Suppose that the data are described by

$$x_i = \mathbf{z}_i^t \theta + r_i, \quad i = 1, \ldots, n \tag{2.11}$$

where $\mathbf{z}_i = \text{col}\,\{z_{i1}, \ldots, z_{ip}\}$, $\boldsymbol{\theta} = \text{col}\,\{\theta_1, \ldots, \theta_p\}$, and the r_i, $i = 1, \ldots, n$, are IID with $\mathrm{E}\{r_i^2\} = \sigma_r^2$.

M-estimates of the parameters $\boldsymbol{\theta}$ and scale σ_r [39] [38] are obtained for fixed B as a solution of

$$\min_{\boldsymbol{\theta}, \sigma_r} \left\{ \sum_{i=1}^{n} \left[\rho\left(\frac{x_i - \boldsymbol{\theta}^t \mathbf{z}_i}{\sigma_r} \right) + B \right] \sigma_r \right\} . \tag{2.12}$$

B is chosen to make the estimates consistent at the nominal distribution [41]. By differentiating (2.12) with respect to $\boldsymbol{\theta}$ and σ_r we obtain for the estimators $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and S of σ_r the implicit equations

$$\sum_{i=1}^{n} \mathbf{z}_i \psi\left(\frac{x_i - \hat{\boldsymbol{\theta}}^t \mathbf{z}_i}{S} \right) = \mathbf{0} \tag{2.13}$$

and

$$\sum_{i=1}^{n} \chi\left(\frac{x_i - \hat{\boldsymbol{\theta}}^t \mathbf{z}_i}{S} \right) = 0, \tag{2.14}$$

where

$$\psi(x) = \frac{\partial}{\partial x} \rho(x)$$

and

$$\chi(x) = x\,\psi(x) - \rho(x) - B\,.$$

Selection of the ψ-function to have good robust characteristics, such as being bounded ($\gamma^* < \infty$), yields robust estimates of the parameters, so long as the $\mathbf{z}$ portion of the model is correctly specified [41].

2.3.2 GM-estimators

M-estimators are good when the data $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ are independent. In the dependent case, for example when the data, $x(1), \ldots, x(n)$, are an autoregressive

time series

$$x(k) = \sum_{l=1}^{p} \theta_l x(k-l) + r(l), \quad k = 1 \text{ to } n.$$

where $\theta = \text{col } \{\theta_1, \ldots, \theta_p\}$ and $\mathbf{x}_k = \text{col } \{x(k-1), \ldots, x(k-p)\}$, then the estimation procedure must be modified. Martin et al. [41, 42] have considered M-estimators for 1-D autoregressive time series models. Direct application of (2.13) and (2.14) with $\mathbf{z}_i = \mathbf{x}_i = \text{col } \{x(i-1), \ldots, x(i-p)\}$, may give poor results since $\mathbf{x}_i \psi((x(i) - \theta^t \mathbf{x}_i)/\sigma_r)$ is not necessarily bounded due to contamination in the $\mathbf{x}_i$. It can be shown that M-estimates lack qualitative robustness and have zero breakdown points for autoregressions.

This leads to the following modifications to (2.13) and (2.14).

Definition 2.7 Generalized M-estimates (GM-estimators) $\hat{\theta}$ *of* θ *and* S *of* σ_r *are solutions to the following implicit equations.*

$$\sum_{i=p+1}^{n} \mathbf{x}_i W(\mathbf{x}_i) \psi \left(\frac{x(i) - \hat{\theta}^t \mathbf{x}_i}{S} \right) = 0 \tag{2.15}$$

and

$$\frac{1}{n-2p-1} \sum_{i=p+1}^{n} W(\mathbf{x}_i) \psi^2 \left(\frac{x(i) - \hat{\theta}^t \mathbf{x}_i}{S} \right) = A \tag{2.16}$$

where $\mathbf{x}W(\mathbf{x})$ *is a bounded and continuous function of* $\mathbf{x}$.

For the GM-estimate, if $\psi(\cdot)$ is bounded and continuous, then the same is true of the summands in (2.15). In this case it can be shown that the GM-estimates are qualitatively robust and have positive breakdown points. However, the breakdown point is bounded above by $\frac{1}{p+1}$ [42], and therefore, for large order autoregressions the breakdown point will be significantly smaller than $\frac{1}{2}$, the theoretical maximum.

A good choice for $W(\cdot)$ is

$$W(\mathbf{x}_i) = w\left(\frac{1}{p}\mathbf{x}_i^t \hat{C}_{\mathbf{x}}^{-1} \mathbf{x}_i\right)$$

where $w(\cdot)$ is a non-negative continuous function and $\hat{C}_{\mathbf{x}}$ is an estimate of the $p \times p$ covariance matrix for the past history vector $\mathbf{x}_i$. If $w(\cdot)$ is a redescending function such that $w(x) = 0$ whenever $|x| > b$, then $\mathbf{x}W(\mathbf{x})$ will be bounded.

Chapter 3

Spatial Models for Spectrum Estimation

The model based approach assumes that a data value at a given point on a 2-D lattice is statistically dependent upon data at neighboring locations of the lattice. One way of specifying this dependency is through the use of a 2-D random field model. There are two nonequivalent classes of models for 2-D random fields yielding rational spectra which have received much attention in the statistical and engineering literature [4, 33, 16, 36]: the simultaneous models and the conditional Markov models. Here, for the spectrum estimation problem we are particularly interested in the autoregressive models from the first class and the conditional Gaussian-Markov models from the second class. We describe these models in some detail since they will be used extensively throughout this dissertation.

3.1 Class 1: Autoregressive Random Field Models

Autoregressive models are generalizations to two dimensions of the familiar time series autoregressions. Suppose we want to model a set of data $\mathcal{Y}_M = \{y(\mathbf{s}), \mathbf{s} \in \Omega_M\}$ on the 2-D square lattice $\Omega_M = \{\mathbf{s} = s_1, s_2 | 0 \leq s_1, s_2 \leq M-1\}$. The

autoregressive model of this data has the difference equation formulation

$$y(\mathbf{s}) = \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} y(\mathbf{s}+\mathbf{r}) + w(\mathbf{s}), \quad \mathbf{s} \in \Omega_M, \tag{3.1}$$

The input array $\{w(\mathbf{s})\}$ is an independent and identically distributed (IID) random noise field with $\mathrm{E}\{w(\mathbf{s})\} = 0$ and $\mathrm{E}\{w^2(\mathbf{s})\} = \beta^2$. N, the neighbor set, has elements $\mathbf{r} = r_1, r_2$ where r_1 and r_2 are integers. Noncausal autoregressive (NCAR) models have no restrictions on the elements of N, except that $(0,0) \notin N$. If N is a subset of the nonsymmetric half-plane Ω^+ nonuniquely defined by

1. $\mathbf{s}_1 \in \Omega^+, \mathbf{s}_2 \in \Omega^+ \Rightarrow \mathbf{s}_1 + \mathbf{s}_2 \in \Omega^+$,

2. $\mathbf{s} \in \Omega^+ \Rightarrow -\mathbf{s}$ not $\in \Omega^+$, and

3. $(0,0)$ not $\in \Omega^+$,

then the model is known as a nonsymmetric half-plane (NSHP), semicausal, or unilateral autoregressive model. Figure 3.1 shows this region on the two-dimensional lattice Ω_M. Quarter-plane (QP) models are obtained with the further restriction that $N = \{\mathbf{r} = r_1, r_2 \mid r_1 \leq 0,\ r_2 \leq 0, \text{ and } (r_1, r_2) \neq (0,0)\}$. Obviously, NCAR models can be considered to include the causal models as a subclass.

In NCAR models when N includes symmetric neighbors, i.e., $\mathbf{r} \in N$ implies $-\mathbf{r} \in N$, then $\theta_{\mathbf{r}}$ must equal $\theta_{-\mathbf{r}}$ to ensure that the parameters are identifiable [16]. Moreover, a necessary and sufficient condition on $\theta = \mathrm{col}\ \{\theta_{\mathbf{r}},\ \mathbf{r} \in N\}$ to ensure stationarity of $\mathcal{Y}_\infty$ is that

$$1 - \sum_{(r_1, r_2) \in N} \theta_{r_1, r_2} z_1^{r_1} z_2^{r_2} \neq 0$$

for all z_1 and z_2 such that $|z_1| = |z_2| = 1$.

In view of the symmetry requirement for the elements of the neighbor set for a NCAR model with noncausal support we define a *model order* according to Figure

3.2. For example, a 2nd-order symmetric NCAR model has a neighbor set which contains not only the points labeled with "2" but also those labeled with "1", i.e., $N^2 = \{(1,0),(-1,0),(0,1),(0,-1),(1,1),(-1,-1),(-1,1),(1,-1)\}$. For convenience we may also specify the 2nd-order symmetric NCAR model through the use of the asymmetric neighbor set N_s such that for all $\mathbf{r} \in N$, $\mathbf{r} \in N_s$ implies $-\mathbf{r} \notin N_s$. However, this representation precludes the sublcass of NSHP models.

The spectrum for $\mathcal{Y}_\infty$ obeying (3.1) is given by

$$\mathcal{S}_y(\omega) = \frac{\beta^2}{|1 - \sum_{\mathbf{r}\in N} \theta_{\mathbf{r}} \exp(-\jmath\omega^t\mathbf{r})|^2} \tag{3.2}$$

for $\omega = (\omega_1, \omega_2), -\pi \leq \omega_1, \omega_2 < \pi$.

An alternate representation for the noncausal autoregression on a finite lattice is obtained by using the $M^2 \times 1$ vectors $\mathbf{y} = \text{col}\ \{y(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$ and $\mathbf{w} = \text{col}\ \{w(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$ of the lexicographically ordered arrays $\{y(\mathbf{s}), \mathbf{s} \in \Omega_M\}$ and $\{w(\mathbf{s}), \mathbf{s} \in \Omega_M\}$, respectively. In this representation (3.1) is written as

$$B(\boldsymbol{\theta})\mathbf{y} = \mathbf{w} \tag{3.3}$$

where $B(\boldsymbol{\theta})$ is a $M^2 \times M^2$ block-Toeplitz matrix which depends on $\boldsymbol{\theta}$ and the specific boundary conditions. Using this notation, $Q_y = \beta^2[B^t(\boldsymbol{\theta})B(\boldsymbol{\theta})]^{-1}$.

The representation in (3.1) is not valid for the data lying on the boundary if the array dimension M is finite; as for some $\mathbf{s}$, the value $y(\mathbf{s}+\mathbf{r})$ may not be defined. The alternative is to assume the specific finite lattice model

$$y(\mathbf{s}) = \sum_{\mathbf{r}\in N} \theta_{\mathbf{r}} y(\mathbf{s} \oplus \mathbf{r}) + \beta w(\mathbf{s}),\ \ \mathbf{s} \in \Omega_M, \tag{3.4}$$

where $\oplus$ indicates sum modulo M along both the coordinate axes. Equation (3.4) means that the data are represented on a toroidal lattice, and as a result of this assumption the autocorrelation is periodic; hence, the power spectral density

is the discrete Fourier transform of the periodic correlation sequence. Then for $\omega_{\mathbf{s}} = \frac{2\pi}{M}\mathbf{s}$ the discrete spectral density for $\mathcal{Y}_M$ described by the toroidal NCAR model is

$$S_y^o(\omega_{\mathbf{s}}) = \frac{\beta^2}{|\mu_{\mathbf{s}}|^2}, \ \mathbf{s} \in \Omega_M, \tag{3.5}$$

The $\mu_{\mathbf{s}} = 1 - \theta^t \psi_{\mathbf{s}}$, where $\psi_{\mathbf{s}} = \text{col}\ \{\exp(-\jmath\omega_{\mathbf{s}}^t \mathbf{r}),\ \mathbf{r} \in N\}$, for $\mathbf{s} \in \Omega_M$, are the eigenvalues of $B^t(\theta)B(\theta)$. $B(\theta)$ for the toroidal model is a $M^2 \times M^2$ block-circulant matrix with each $M \times M$ block also a circulant. It follows that the coefficients θ must obey $|\mu_{\mathbf{s}}|^2 \neq 0$ for all $\mathbf{s} \in \Omega_M$ for $\mathcal{Y}_M$ to be stationary.

3.2 Class 2: Gaussian-Markov Random Field Models

Assume that the random field $\mathcal{Y}_M = \{y(\mathbf{s}), \mathbf{s} \in \Omega\}$ is to have zero mean, Gaussian distribution, and obey the Gaussian-Markov random field (GMRF) model difference equation

$$y(\mathbf{s}) = \sum_{\mathbf{r} \in N_o} \theta_{\mathbf{r}}[y(\mathbf{s}+\mathbf{r}) + y(\mathbf{s}-\mathbf{r})] + e(\mathbf{s}), \ \ \mathbf{s} \in \Omega_M, \tag{3.6}$$

where the stationary Gaussian noise field $\{e(\mathbf{s}), \mathbf{s} \in \Omega_M\}$ has the property that

$$\mathrm{E}\{e(\mathbf{s})e(\mathbf{t})\} = \begin{cases} -\theta_{\mathbf{s}-\mathbf{t}}\nu, & (\mathbf{s}-\mathbf{t}) \in N \\ \nu, & \mathbf{s} = \mathbf{t} \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

We assume that (3.6) has input-output stability. Using (3.6) and (3.7) one can prove [34] that the conditional density of $y(\mathbf{s})$ given all other values is

$$f(y(\mathbf{s})|\text{all } y(\mathbf{t}), \mathbf{t} \neq \mathbf{s}) = f(y(\mathbf{s})|\text{all } y(\mathbf{s}+\mathbf{r}), \mathbf{r} \in N). \tag{3.8}$$

Alternatively, one can show [32] that the conditional density of (3.8) leads to the model (3.6).

The spectral density function for the GMRF model (3.6) is

$$S_y(\omega) = \frac{\nu}{1 - 2\sum_{\mathbf{r}\in N_s} \theta_{\mathbf{r}} \cos(\omega^t \mathbf{r})}. \tag{3.9}$$

Note that $y(\cdot)$ obeying the NCAR model (3.1) is not Markov with respect to N if N is not unilateral, i.e.,

$$f(y(\mathbf{s})|\text{all } y(\mathbf{s}+\mathbf{r}), \mathbf{r} \neq \mathbf{s}) \neq f(y(\mathbf{s})|\text{all } y(\mathbf{s}+\mathbf{r}), \mathbf{r} \in N); \tag{3.10}$$

hence, the NCAR models are different than the 2-D Gaussian-Markov models. However, for all NCAR models

$$f(y(\mathbf{s})|\text{all } y(\mathbf{s}+\mathbf{r}), \mathbf{r} \neq \mathbf{s}) = f(y(\mathbf{s})|\text{all } y(\mathbf{s}+\mathbf{r}), \mathbf{r} \in N') \tag{3.11}$$

for some N' where $N \subset N'$. Thus, some GMRF models have spectra which factorize; and hence, for every NCAR model there exists a GMRF model having the same spectral density function.

Gaussian-Markov random field models can also be represented on a toroidal lattice. In vector-matrix notation the toroidal GMRF model is

$$A(\theta)\mathbf{y} = \mathbf{e}, \tag{3.12}$$

where $\mathbf{y}$ and $\mathbf{e}$ are column vectors of the respective lexicographically ordered arrays, and $A(\theta)$ is the block/block-circulant $M^2 \times M^2$ matrix of coefficients $\theta_{\mathbf{r}}, \mathbf{r} \in N_s$. The spectral density function for the toroidal GMRF model is

$$S_y^o(\omega_{\mathbf{s}}) = \frac{\nu}{\mu_{\mathbf{s}}}, \ \mathbf{s} \in \Omega_M, \tag{3.13}$$

where $\mu_{\mathbf{s}} = 1 - 2\theta^t \mathbf{c}_{\mathbf{s}}$ and $\mathbf{c}_{\mathbf{s}} = \text{col}\ \{\cos(\omega_{\mathbf{s}}^t \mathbf{r}), \mathbf{r} \in N_s\}$.

3.3 Parameter Estimation for Spatial Models

Suppose that the random field $\mathcal{Y}_\infty = \{y(\mathbf{s}), \mathbf{s} \in \Omega_\infty\}$ is Gaussian, mean zero, and has a spectral density function given by either (3.2) or (3.9) and that we have available a set of observations from the $M \times M$ lattice Ω_M. Furthermore, assume that the appropriate choice regarding the model structure has been made, i.e., the neighbor set N is known. Now, the spectrum estimation problem is reduced to a two step parameter estimation problem: (1) the parameters $\boldsymbol{\theta}$ and β^2 or ν are estimated and (2) these estimates are then used in the theoretical expression for the spectrum, (3.2) or (3.9) as the case maybe.

A popular estimator in 1-D and 2-D is the least squares (LS) estimator. This estimator is consistent for 1-D and 2-D causal AR models and for the noncausal GMRF model; but, $\hat{\boldsymbol{\theta}}_{LS}$ is not consistent for NCAR models with bilateral neighbor sets [33]. The LS estimator for the toroidal AR model is

$$\hat{\boldsymbol{\theta}}_{LS} = \left[\sum_{\mathbf{s}\in\Omega_M} \mathbf{y}_\mathbf{s}\mathbf{y}_\mathbf{s}^t\right]^{-1} \left[\sum_{\mathbf{s}\in\Omega_M} \mathbf{y}_\mathbf{s} y(\mathbf{s})\right] \tag{3.14}$$

and

$$\hat{\beta}^2_{LS} = \frac{1}{M^2} \sum_{\mathbf{s}\in\Omega_M} [y(\mathbf{s}) - \hat{\boldsymbol{\theta}}^t_{LS}\mathbf{y}_\mathbf{s}]^2 \tag{3.15}$$

where $\mathbf{y}_\mathbf{s} = \text{col}\,\{y(\mathbf{s}\oplus\mathbf{r}), \mathbf{r} \in N\}$. A similar expression is obtained for the GMRF model. A theorem regarding the consistency of the estimate $\hat{\boldsymbol{\theta}}_{LS}$ and $\hat{\nu}_{LS}$ and an expression for the asymptotic variance of the estimate $\hat{\boldsymbol{\theta}}_{LS}$ can be found in [33]. These estimates are not fully efficient.

An asymptotically consistent and efficient estimate can be obtained for all NCAR and GMRF models by the maximum likelihood (ML) method. Since the Jacobian of the transformation matrix from the noise variable $w(\mathbf{s})$ (or $e(\mathbf{s})$) to

$y(\mathbf{s})$ is not unity in NCAR (or GMRF) models with symmetric neighbor sets, the log-likelihood function is a complicated function of the parameters even for the Gaussian case. Assuming the availability of boundary values, the negative log-likelihood function is

$$l_M(\phi) = \frac{M^2}{2}\log(2\pi) + \frac{1}{2}\log[\det\{Q_{\mathbf{y}}\}] + \frac{1}{2}\mathbf{y}^t Q_{\mathbf{y}}^{-1}\mathbf{y} \tag{3.16}$$

where $Q_{\mathbf{y}} = \mathrm{E}\{\mathbf{y}\mathbf{y}^t\}$ and $\phi = \mathrm{col}\ \{\theta, \beta\}$ (or $\phi = \mathrm{col}\ \{\theta, \nu\}$) for the NCAR (or GMRF model). The ML estimates are obtained by minimizing (3.16) with respect to the parameters ϕ.

However, the solution of this problem requires extensive computation due to the terms $\det\{Q_{\mathbf{y}}\}$ and $Q_{\mathbf{y}}^{-1}$; and an explicit expression for $\det\{Q_{\mathbf{y}}\}$ is difficult to derive. Using the approximation [1, 35, 37, 46]

$$\lim_{M\to\infty}\left\{\frac{1}{M^2}\log[\det\{Q_y\}]\right\} = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log[S_y(\omega,\phi)]d\omega, \tag{3.17}$$

one can show that an asymptotically equivalent expression for the negative log-likelihood function is

$$\tilde{l}_M(\phi) = \frac{M^2}{2}\log(2\pi) + \frac{M^2}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log[S_y(\omega,\phi)]d\omega + \frac{1}{2}\mathbf{y}^t Q_y^{-1}\mathbf{y}, \tag{3.18}$$

with $S_y(\omega,\phi)$ given by (3.2) or (3.9) for the NCAR or GMRF model, respectively.

For the NCAR model the quadratic form in (3.18) can easily be shown to be

$$\mathbf{y}^t Q_y^{-1}\mathbf{y} = \frac{1}{\beta^2}\sum_{\mathbf{s}\in\Omega_M}[y(\mathbf{s}) - \theta^t\mathbf{y}_{\mathbf{s}}]^2 \tag{3.19}$$

and for the GMRF model

$$\mathbf{y}^t Q_y^{-1}\mathbf{y} = \frac{1}{\nu}\sum_{\mathbf{s}\in\Omega_M} y(\mathbf{s})[y(\mathbf{s}) - \theta^t\mathbf{y}_{\mathbf{s}}]. \tag{3.20}$$

A convenient form of the GMRF quadratic term, which is equivalent to (3.20) [31], is

$$\mathbf{y}^t Q_y^{-1} \mathbf{y} = \frac{M^2}{\nu} \left[c_y(0) - 2 \sum_{\mathbf{r} \in N_s} \theta_{\mathbf{r}} c_x(\mathbf{r}) \right], \tag{3.21}$$

where the sample covariances are

$$c_y(\mathbf{r}) = \frac{1}{M^2} \sum_{\mathbf{s} \in \Omega_M} y(\mathbf{s}) y(\mathbf{s}+\mathbf{r}), \quad \mathbf{r} \in N_s$$

and

$$c_y(0) = \frac{1}{M^2} \sum_{\mathbf{s} \in \Omega_M} x^2(\mathbf{s}).$$

By introducing the toroidal lattice, an approximation for $l_M(\phi)$, requiring many fewer computations to determine a global minimum, is

$$l_M^o(\phi) = \frac{1}{2M^2} \sum_{\mathbf{s} \in \Omega_M} \left\{ \log[\mathcal{S}_y^o(\omega_{\mathbf{s}}, \phi)] + \frac{\mathcal{P}_{y,M}(\omega_{\mathbf{s}})}{\mathcal{S}_y^o(\omega_{\mathbf{s}}, \phi)} \right\}, \tag{3.22}$$

where $\mathcal{P}_{y,M}(\omega_{\mathbf{s}})$ is the periodogram of the observations, i.e.,

$$\mathcal{P}_{y,M}(\omega_{\mathbf{s}}) = |Y(\omega_{\mathbf{s}})|^2 \tag{3.23}$$

and

$$Y(\omega_{\mathbf{s}}) = \frac{1}{M} \sum_{\mathbf{t} \in \Omega_M} y(\mathbf{t}) \exp(-j \frac{2\pi}{M} \mathbf{s}^t \mathbf{t}). \tag{3.24}$$

It turns out that as $M \to \infty$ there exists a unique $\hat{\phi}_{ML}$ that yields the global minimum of either (3.16), (3.18), and (3.22). Moreover, $M[\hat{\phi}_{ML} - \phi_0]$ is distributed as $\mathcal{N}(0, \Gamma^{-1}(\phi_0))$ where

$$\Gamma(\phi) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left[\frac{\partial}{\partial \phi} \log \mathcal{S}_y(\omega, \phi) \right] \left[\frac{\partial}{\partial \phi} \log \mathcal{S}_y(\omega, \phi) \right]^t d\omega. \tag{3.25}$$

The matrix $\Gamma(\phi)$ is simply related to the Fisher information matrix $I_M(\phi)$, defined as

$$I_M(\phi) = \frac{1}{M^2} \mathrm{E} \left\{ \left[\frac{\partial}{\partial \phi} l_M(\phi) \right] \left[\frac{\partial}{\partial \phi} l_M(\phi) \right]^t \right\}, \tag{3.26}$$

by the asymptotic relationship

$$\lim_{M \to \infty} I_M(\phi) = \Gamma(\phi). \tag{3.27}$$

Thus, we conclude that the ML estimator for the NCAR or GMRF model is consistent and asymptotically efficient whenever the data are distributed normally.

The NSHP model has characteristics which makes it the spatial analog of the 1-D time-series model. Specifically, $B(\theta)$ in (3.3) is a lower triangular matrix whose main diagonal elements are unity. Therefore, $\det\{B(\theta)\} = 1$ and

$$\begin{aligned} \det\{Q_y\} &= \det\{\beta^2[B^t(\theta)B(\theta)]^{-1}\} \\ &= \frac{\beta^{2M^2}}{\det^2\{B(\theta)\}} \\ &= 1. \end{aligned}$$

This simplifies the likelihood function for the NSHP model which is

$$f(\mathbf{y}|\phi) = \prod_{\mathbf{s}\in\Omega_M} (2\pi\beta^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\beta^2}[y(\mathbf{s}) - \theta^t \mathbf{y}_\mathbf{s}]^2\right\}. \tag{3.28}$$

Note that this likelihood function is the product over all lattice points of the conditional probability density functions. The random field possessing this property is said to have a one-sided Markov property. In particular,

$$\mathrm{E}\{y(\mathbf{s})|y(\mathbf{t}), \mathbf{t} < \mathbf{s}\} = \mathrm{E}\{y(\mathbf{s})|\mathbf{y}_\mathbf{s}\} = \sum_{\mathbf{r}\in N} \theta_\mathbf{r} y(\mathbf{s}+\mathbf{r}), \tag{3.29}$$

because $\mathrm{E}\{w(\mathbf{s})|y(\mathbf{t}), \mathbf{t} < \mathbf{s}\} = 0$. This property of NSHP models is identical to that of the causal time-series models. The conditional variance of $y(\mathbf{s})$ is

$$\begin{aligned} \mathrm{E}\left\{[y(\mathbf{s}) - \mathrm{E}\{y(\mathbf{s})|\mathbf{y}_\mathbf{s}\}]^2|\mathbf{y}_\mathbf{s}\right\} &= \mathrm{E}\left\{[y(\mathbf{s}) - \theta^t\mathbf{y}_\mathbf{s}]^2|\mathbf{y}_\mathbf{s}\right\} \\ &= \mathrm{E}\left\{w^2(\mathbf{s})|\mathbf{y}_\mathbf{s}\right\} \\ &= \beta^2. \end{aligned} \tag{3.30}$$

3.4 Rules for Model Selection

In the previous section we assumed that the appropriate model from the class of NCAR and GMRF models was known. From 1-D time series analysis, it is known [47] that the use of an appropriate model leads to good results in spectrum estimation. The model selection problem comes under the category of a multiple decision problem. The procedure is as follows. A priori, we choose a number of different models of the same kind (e.g., NCAR models with differing neighbor sets N), compute a test statistic for each model and choose the neighbor set or model having the minimum value of test statistic. Two different ways have been employed for computing the test statistic, namely, the AIC criterion [48, 49] and the Bayes' method [50]. In the Bayes method the test statistic is chosen so that the decision rule minimizes the probability of error, i.e., choosing the wrong model order. It is also a weakly consistent method. Alternatively, the AIC rule has no optimality property and is not consistent, i.e., even when the number of observations tends to infinity the probability of error does not go to zero [51].

Suppose we have three sets N_1, N_2, and N_k of neighbors containing p_1, p_2, and p_k members, respectively. Corresponding to each N_k, we write the toroidal NCAR models as

$$y(\mathrm{s}) = \sum_{\mathrm{r} \in N_k} \theta_{\mathrm{r}} y(\mathrm{s} \oplus \mathrm{r}) + w(\mathrm{s}).$$

Then the decision rule for the choice of appropriate neighbors [33] is: choose the neighbor set N_k^* where

1. $k^* = \arg\{\min_k B_k\}$.

2. $B_k = \sum_{\mathrm{s} \in \Omega_M} \log[S_y^{\circ}(\omega_{\mathrm{s}}, \phi_k^*)] + p_k \log(M^2)$.

3. θ_k^* and β_k^{*2} are the maximum-likelihood estimates.

4. $\psi_{sk} = \text{col}\ \{\exp(-\jmath\omega_s^t \mathrm{r}), \mathrm{r} \in N_k\}$ and p_k is the number of elements in N_k.

This decision rule can be derived by using asymptotic Bayes decision theoretic methods as in [50]. A similar decision rule has been developed in [33] for GMRF models. Alternatively, a strongly consistent decision rule is given in [52]. The test statistic for this rule is given by

$$H_k = \sum_{s \in \Omega_M} \log[\mathcal{S}_y^o(\omega_s, \phi_k^*)] + 2p_k \log[\log(M^2)],$$

and one chooses the neighbor set N_k^* where $k^* = \arg\{\min_k H_k\}$.

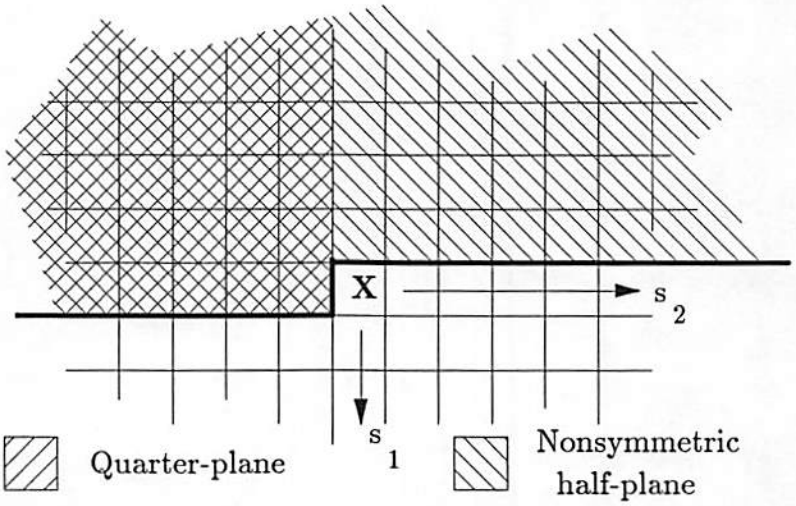


Figure 3.1: Nonsymmetric half-plane and quarter-plane on the two-dimensional lattice.

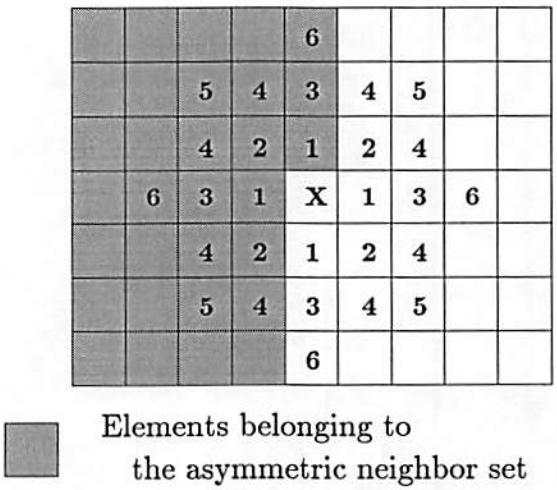


Figure 3.2: Model orders for the symmetric NCAR and GMRF models.

Chapter 4

2-D Noncausal Signal Plus Noise Estimators

In this chapter we introduce the NCAR plus noise random field model and the ML method of estimating its parameters from a finite set of data. We then demonstrate the NCAR plus noise model's superior performance, compared with the NCAR only model, in the practical application of detecting and resolving sinusoidal signals in noise. These signals are typical of those encountered in radar, sonar, seismology, radio astronomy, and other similar planar array signal processing applications.

4.1 Noncausal Autoregressive Plus Noise Model

Let $\mathcal{X}_M = \{x(\mathbf{s}), \mathbf{s} \in \Omega_M\}$ be a set of observations of a 2-D random field. The NCAR signal plus noise model representation for $\mathcal{X}$ is

$$x(\mathbf{s}) = y(\mathbf{s}) + n(\mathbf{s}), \tag{4.1}$$

where $y(\mathbf{s})$ obeys a NCAR model, given by (3.1) of Chapter 3, and $n(\mathbf{s})$ is a random noise field. The spectrum for $\mathcal{Y}_\infty$ obeying (3.1) with parameters $\boldsymbol{\phi} = \text{col}\,\{\boldsymbol{\theta}, \beta\}$

and $\boldsymbol{\theta} = \text{col}\,\{\theta_{\mathbf{r}}, \mathbf{r} \in N\}$ is given by

$$S_y(\omega, \phi) = \frac{\beta^2}{|1 - \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} \exp(-j\omega^t \mathbf{r})|^2}. \tag{4.2}$$

If $n(\mathbf{s})$ in (4.1) is Gaussian white noise with $\mathrm{E}\{n(\mathbf{s})\}$ and $\mathrm{E}\{n^2(\mathbf{s})\} = \gamma^2$, then $S_n(\omega) = \gamma^2$, a constant for all ω. If $n(\mathbf{s})$ is independent of the process $y(\mathbf{s})$, then the spectrum for the observations $\mathcal{X}_\infty$ is given by

$$S_x(\omega, \phi) = \frac{\beta^2}{|1 - \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} \exp(-j\omega^t \mathbf{r})|^2} + \gamma^2. \tag{4.3}$$

An alternate representation of the noncausal autoregression on a finite lattice is obtained by using the $M^2 \times 1$ vectors $\mathbf{y} = \text{col}\,\{y(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$, $\mathbf{w} = \text{col}\,\{w(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$, $\mathbf{x} = \text{col}\,\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$, and $\mathbf{n} = \text{col}\,\{n(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$ of the lexicographically ordered arrays $\{y(\mathbf{s}), \mathbf{s} \in \Omega_M\}$, $\{w(\mathbf{s}), \mathbf{s} \in \Omega_M\}$, $\{x(\mathbf{s}), \mathbf{s} \in \Omega_M\}$, and $\{n(\mathbf{s}), \mathbf{s} \in \Omega_M\}$, respectively. In this representation (3.1) is written as

$$B(\boldsymbol{\theta})\mathbf{y} = \mathbf{w}$$

and the signal plus noise model (4.1) is

$$\mathbf{x} = \mathbf{y} + \mathbf{n}, \tag{4.4}$$

where $B(\boldsymbol{\theta})$ is a $M^2 \times M^2$ block Toeplitz matrix depending on $\boldsymbol{\theta}$ and the specific boundary conditions (see, for example, [33]). If $w(\mathbf{s})$ is Gaussian with mean zero, then $y(\mathbf{s})$ is Gaussian with $\mathrm{E}\{\mathbf{y}\} = \mathbf{0}$. Moreover, if $\{n(\mathbf{s})\}$ is an IID Gaussian random noise field with common distribution $\mathcal{N}(0, \gamma^2)$ and uncorrelated with $\{y(\mathbf{s})\}$, then it follows that

$$\begin{aligned} Q_w &= \mathrm{E}\left\{\mathbf{w}\mathbf{w}^t\right\} = \beta^2 I_{M^2} \\ Q_y &= \mathrm{E}\left\{\mathbf{y}\mathbf{y}^t\right\} = \beta^2 [B^t(\boldsymbol{\theta}) B(\boldsymbol{\theta})]^{-1} \\ Q_n &= \mathrm{E}\left\{\mathbf{n}\mathbf{n}^t\right\} = \gamma^2 I_{M^2} \end{aligned}$$

where I_{M^2} is the $M^2 \times M^2$ identitiy matrix. Also, it follows that $\mathrm{E}\{\mathbf{x}\} = \mathbf{0}$ and that

$$Q_x = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^t\right\} = \beta^2[B^t(\boldsymbol{\theta})B(\boldsymbol{\theta})]^{-1} + \gamma^2 I_{M^2}.$$

Since it is known that LS estimates are not even consistent for noncausal autoregressions [33, 1], we should expect no better results for LS estimates of the parameters of the NCAR plus noise model. Hence, we concentrate on the ML method for estimating the spectrum parameters. We begin with the following definition.

Definition 4.1 *The true likelihood function of a finite set of observations, denoted by the $M^2 \times 1$ column vector $\mathbf{x} = \mathrm{col}\,\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$, of a stationary Gaussian random field with spectral density $S_x(\omega, \phi)$, $\omega = (\omega_1, \omega_2)$, $-\pi \leq \omega_1, \omega_2 \leq \pi$, is given by*

$$\begin{aligned} L_M(\phi) &= p(\mathbf{x} \mid \phi) \\ &= (2\pi)^{-\frac{M^2}{2}}[\det\{Q_x\}]^{-\frac{1}{2}} \exp(-\frac{1}{2}\mathbf{x}^t Q_x^{-1}\mathbf{x}), \end{aligned} \tag{4.5}$$

where Q_x is the $M^2 \times M^2$ matrix whose i,j-th element is $q_x(\mathbf{t}_i - \mathbf{t}_j)$.

Maximizing (4.5) is equivalent to minimizing the negative-log likelihood function

$$l_M(\phi) = \frac{1}{2}\left\{\log[\det\{Q_x\}] + \mathbf{x}^t Q_x^{-1}\mathbf{x}\right\}. \tag{4.6}$$

The *maximum likelihood estimate* of the parameters is the value of ϕ which maximizes (4.5), or equivalently, minimizes (4.6). This estimate will be denoted by $\bar{\phi}_M$ and is formally

$$\bar{\phi}_M = \arg\left\{\min_{\phi} l_M(\phi)\right\}. \tag{4.7}$$

When $\bar{\phi}_M$ belongs to an open set a necessary condition is that $\bar{\phi}_M$ solves the p-dimensional set of equations

$$l'_M(\phi) = \frac{\partial}{\partial \phi} l_M(\phi) = 0. \tag{4.8}$$

For the NCAR plus noise model (4.8) is not a simple linear equation in the parameters ϕ; and it is not readily apparent that a solution of (4.8), even if it exists, is unique. So, the solution must be determined by finding the global minimum of (4.6) using a nonlinear optimization algorithm. However, these procedures require extensive computation because of the terms $\det\{Q_x\}$ and Q_x^{-1}. Thus one is motivated to consider an approximate version of (4.5) requiring fewer computations, but one that maintains desirable asymptotic properties. The existence and uniqueness questions will be addressed again in Section 4.2.

The approximation that we consider here is known as Whittle's approximation (see [1]) and is given in the following definition.

Definition 4.2 *The approximate (negative-log) likelihood function of a finite set of observations,* $\mathbf{x} = \text{col}\,\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$, *of a stationary Gaussian random field with spectral density* $S_x(\omega, \phi)$ *and periodogram*

$$\mathcal{P}_{x,M}(\omega) = \frac{1}{M^2}|X(\omega)|^2 \tag{4.9}$$

with

$$X(\omega) = \sum_{\mathbf{s}\in\Omega_M} x(\mathbf{s}) \exp(-j\omega^t \mathbf{s}), \tag{4.10}$$

is given by

$$\tilde{l}_M(\phi) = \frac{M^2}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left\{ \log[S_x(\omega, \phi)] + \frac{\mathcal{P}_{x,M}(\omega)}{S_x(\omega, \phi)} \right\} d\omega. \tag{4.11}$$

The *approximate maximum likelihood estimate* of the parameters is the value of ϕ which minimizes (4.11) and is denoted by $\tilde{\phi}_M$, i.e.,

$$\tilde{\phi}_M = \arg\left\{\min_{\phi} \tilde{l}_M(\phi)\right\}. \tag{4.12}$$

A practical assumption that reduces the required computations further is that the lattice Ω is toroidal. With this assumption one may consider the model strictly on the finite lattice Ω_M with doubly periodic boundary conditions instead of the zero-valued boundary conditions assumed for (4.11). In this situation the Fourier transform (4.10) can be replaced with the discrete Fourier transform $X(\mathbf{s})$ given by

$$X(\mathbf{s}) = \frac{1}{M} \sum_{\mathbf{t}\in\Omega_M} x(\mathbf{t}) \exp(-j\frac{2\pi}{M}\mathbf{s}^t\mathbf{t}) \,. \tag{4.13}$$

Now the approximate ML function for the toroidal model reduces to

$$l_M^o(\theta,\beta,\gamma) = \frac{1}{2M^2} \sum_{\mathbf{s}\in\Omega_M} \left[\log\left(\frac{\beta^2 + |\mu_\mathbf{s}|^2\gamma^2}{|\mu_\mathbf{s}|^2}\right) + \frac{|X(\mathbf{s})|^2|\mu_\mathbf{s}|^2}{\beta^2 + |\mu_\mathbf{s}|^2\gamma^2}\right] \tag{4.14}$$

where the spectrum for the toroidal lattice NCAR only model is

$$\mathcal{S}_y^o(\omega_\mathbf{s}) = \frac{\beta^2}{|\mu_\mathbf{s}|^2}, \ \mathbf{s} \in \Omega_M, \tag{4.15}$$

and for the NCAR plus noise model

$$\mathcal{S}_x^o(\omega_\mathbf{s}) = \frac{\beta^2}{|\mu_\mathbf{s}|^2} + \gamma^2, \ \mathbf{s} \in \Omega_M. \tag{4.16}$$

The $\mu_\mathbf{s} = 1 - \theta^t\psi_\mathbf{s}$, $\psi_\mathbf{s} = \text{col}\ \{\exp(-j\frac{2\pi}{M}\mathbf{s}^t\mathbf{r}),\ \mathbf{r} \in N\}$, $\mathbf{s} \in \Omega_M$, are the eigenvalues of $B^t(\theta)B(\theta)$, which for the toroidal model is a $M^2 \times M^2$ block-circulant matrix. It follows that the coefficients θ must obey $|\mu_\mathbf{s}|^2 \neq 0$ for all $\mathbf{s} \in \Omega_M$ for $\mathcal{Y}_M$ to be stationary.

4.2 Approximate ML Asymptotic Properties

We know from earlier work [12, 1, 35, 37] that the approximate ML procedure provides parameter estimates for the NCAR only model with good asymptotic properties (e.g., consistency, efficiency, normality etc.). We suspect that these properties also hold for NCAR plus noise model approximate ML estimates. In this section we provide a theoretical foundation for the approximate ML estimator of NCAR plus noise model parameters by stating and proving theorems which delineate these asymptotic properties.

Although, consistency, asymptotic efficiency, and asymptotic normality of ML estimates of parameters for independent observations are well known [53, 54, 55], these statistical facts for dependent situations are not well known because of additional technical details required for their proof. Furthermore, to the best of our knowledge rigorous proofs of these results for higher dimensional signal plus noise processes are not available.

The explicit evaluation of the true ML estimator for the NCAR model is a difficult task even for Gaussian data and is even more difficult for the NCAR plus noise model. Therefore, one is interested in obtaining a more tractable problem by simplifying the likelihood equation. But under what circumstances can one simplify the likelihood function and still maintain an estimator with good properties? The answer is provided by the notion of a "principal part". If $l_n(\phi)$ is the true likelihood function given a value of the parameter ϕ, see Definition 4.1, then the *principal part* is any function $\tilde{l}_n(\phi)$ such that $n^{-1}[l_n(\phi) - \tilde{l}_n(\phi)]$ converges in weak probability to zero as the number of observations n increases without bound. Dzhaparidze (see [56]) indicates that Whittle [1] was the first to use the principal

part of the likelihood function for parameter estimation purposes in a time series case.

The approximate ML function (4.11) for the NCAR plus noise model, as we have shown, is simpler than the exact likelihood function; and at least empirically, it provides estimates with desirable error statistics. It turns out, as shown below, that the approximate ML function fits the definition of the principal part of the ML and that the properties of the estimates derived from it include consistency, asymptotic efficiency and normality.

4.2.1 Consistency and Asymptotic Efficiency

Previously presented asymptotic properties of approximate ML type estimators for d-dimensional random fields (for example, see [35, 37]) have depended upon a linearity property of the process which permits the factorization of the spectral density function. That is, if $S_x(\omega)$ is the spectral density function of the process $\mathcal{X}_\infty$, then

$$\mathcal{G}(\omega) = \frac{S_x(\omega)}{\sigma^2} \tag{4.17}$$

does not depend on σ^2 where $\sigma^2 = \mathrm{V}\{\varepsilon(\mathbf{s})\}$ is the variance of the independent and identically distributed random field $\{\varepsilon(\mathbf{s}), \mathbf{s} \in \Omega_\infty\}$ in the Wold decompostion [57] of the process:

$$x(\mathbf{s}) = \sum_{\mathbf{r} \geq \mathbf{0}} g_{\mathbf{r}} \varepsilon(\mathbf{s} - \mathbf{r}) \tag{4.18}$$

and

$$\mathcal{G}(\omega) = |\sum_{\mathbf{r} \geq \mathbf{0}} g_{\mathbf{r}} \exp(-\jmath \omega^t \mathbf{r})|^2. \tag{4.19}$$

For example, (4.2) permits such a factorization, but on the other hand, (4.3) does not.

In this section we relax the linearity assumption and show that parameter estimates for the spectral density function (4.3), when computed from a suitable principal part of the likelihood function, are consistent, asymptotically efficient, and asymptotically normal. Specifically, we are concerned with the situation where the 2-D random field is a noncausal autoregression with independent additive noise whose spectral density function $S_x(\omega, \phi)$ is known except for the finite set of parameters $\phi = \text{col}\,\{\phi_i,\ 1 \leq i \leq p\}$. The proofs for the asymptotic properties of the NCAR plus noise model approximate ML estimates we give here follow the line of reasoning for times series models given in [56].

To prove these properties for the NCAR plus noise model we need the following assumptions regarding the process $\mathcal{X}_\infty$, its spectral density function $S_x(\omega) = S_x(\omega, \phi)$, and covariance function $q_x(\mathbf{t})$, which are related by (1.1) and (1.2) in Chapter 1.

Assumption 4.1 *ϕ_0, the true value of the parameters, belongs to the closed set Φ^c contained in an open set $\Phi \subseteq \Re^p$, the p-dimensional Euclidean space.*

Assumption 4.2 *If $\phi_1, \phi_2 \in \Phi^c$ and $\phi_1 \neq \phi_2$, then*

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |S_x(\omega, \phi_1) - S_x(\omega, \phi_2)| d\omega \neq 0.$$

Assumption 4.3 *The first and second derivatives of $S_x(\omega, \phi)$ and $S_x^{-1}(\omega, \phi)$ exist and are continuous in ω for all $\phi \in \Phi$.*

Assumption 4.4 *For finite k and K, $0 < k \leq S_x(\omega, \phi) \leq K$ for all ω and $\phi \in \Phi$.*

Assumption 4.5 $\sum_{t\in\Omega_\infty} |q_x(t)|^2|t|^2 < \infty$, *where*

$$q_x(t) = \mathrm{E}\{x(s)x(s+t)\}.$$

The first property we consider is that of the accuracy of the approximate ML function (4.11). A similar result for the NCAR only model is given by [37, Proposition 1], which was not proven due to the difficulty of using classical Hilbert space arguments. The proof presented here is different because it uses a more stringent condition (Assumption 4.5) on the covariances of the random process, yet does not rely on (4.17) and is applicable to the NCAR plus noise model.

Theorem 4.1 *Under Assumptions 4.4 and 4.5 the approximate ML function* $\tilde{l}_M(\phi)$ *given by (4.11) satisfies*

$$\operatorname*{plim}_{M\to\infty} \frac{1}{M}[l_M(\phi) - \tilde{l}_M(\phi)] = 0, \tag{4.20}$$

where $l_M(\phi)$*, the true likelihood is given by (4.5).*

Proof: We write

$$l_M(\phi) - \tilde{l}_M(\phi) = \frac{1}{2}(l_1 + l_2) \tag{4.21}$$

where

$$l_1 = \log[\det\{Q_x\}] - \frac{M^2}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \log \mathcal{S}_x(\omega,\phi)d\omega \tag{4.22}$$

$$l_2 = \mathbf{x}^t Q_x^{-1}\mathbf{x} - \frac{M^2}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi)}d\omega. \tag{4.23}$$

First, we have that

$$\begin{aligned}\mathbf{x}^t A_x \mathbf{x} &= \sum_{t\in\Omega_M}\sum_{s\in\Omega_M} a_x(t-s)x(t)x(s) \\ &= \frac{M^2}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi)}d\omega \end{aligned} \tag{4.24}$$

where A_x is the $M^2 \times M^2$ matrix whose i,j-th entry is $a_x(\mathrm{t}_i - \mathrm{t}_j)$, $\mathrm{t}_i, \mathrm{t}_j \in \Omega_M$, and $a_x(\mathrm{t})$ is given by

$$a_x(\mathrm{t}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} S_x^{-1}(\omega, \phi) \exp(\jmath \omega^t \mathrm{t}) d\omega. \tag{4.25}$$

Substituting (4.24) in (4.23), l_2 becomes

$$l_2 = \mathbf{x}^t [Q_x^{-1} - A_x] \mathbf{x}. \tag{4.26}$$

It is a simple matter to show that the expected value and variance of the quadratic form (4.26) are

$$\mathrm{E}\{l_2\} = \mathrm{tr}\{I_{M^2} - Q_x A_x\} \tag{4.27}$$

and

$$\mathrm{V}\{l_2\} = 2\mathrm{tr}\left\{[I_{M^2} - Q_x A_x]^2\right\}, \tag{4.28}$$

respectively. Consequently, it follows from Lemma 4.3, Section 4.5, that for the random part of (4.21)

$$\lim_{M \to \infty} \frac{1}{M} \mathrm{E}\{l_2\} = 0 \tag{4.29}$$

and

$$\lim_{M \to \infty} \frac{1}{M} \mathrm{V}\{l_2\} = 0. \tag{4.30}$$

Next we use a result in Künsch [35, Theorem 2.5] which (shown under conditions less strict than Assumptions 4.4 and 4.5) proves that

$$\lim_{M \to \infty} \frac{1}{M^2} \log[\det\{Q_x\}] = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log S_x(\omega, \phi) d\omega. \tag{4.31}$$

It follows from (4.31) that for the deterministic part of (4.21)

$$\lim_{M \to \infty} l_1 = 0. \tag{4.32}$$

Combining the results (4.29), (4.30), and (4.32) and using a result in [54, Theorem 7-2, pg 60], it follows that $\frac{1}{M}[l_M(\phi_0) - \tilde{l}_M(\phi_0)]$ converges in the mean square to zero as $M \to \infty$. Consequently, since convergence in mean square implies convergence in probability, we have

$$\operatorname*{plim}_{M \to \infty} \left\{ \frac{1}{M}[l_M(\phi_0) - \tilde{l}_M(\phi_0)] \right\} = 0, \tag{4.33}$$

and the theorem is proved. □

Theorem 4.1 suggests that the approximate ML estimator $\tilde{\phi}_M$ of ϕ_0 possesses the same asymptotic properties as the estimator $\bar{\phi}_M$ derived from (4.5). In fact, we have proved here a stronger result than [35] or [37], i.e., convergence in the mean-square sense. The next theorem addresses the convergence of $\tilde{\phi}_M$.

Theorem 4.2 *Under Assumptions 4.1 through 4.5 the approximate ML estimator* $\tilde{\phi}_M$ *converges in weak probability to the true value* ϕ_0, *i.e.,* $\operatorname*{plim}_{M \to \infty} \tilde{\phi}_M = \phi_0$.

Proof: Let $\phi \in \Phi^c$ and $\phi \neq \phi_0$ and consider the function

$$\begin{aligned} \frac{1}{M^2}[\tilde{l}_M(\phi_0) - \tilde{l}_M(\phi)] &= \frac{1}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left\{ \log \left[\frac{S_x(\omega, \phi_0)}{S_x(\omega, \phi)} \right] \right. \\ &\qquad \left. + \mathcal{P}_{x,M}(\omega) \left[1 - \frac{S_x(\omega, \phi_0)}{S_x(\omega, \phi)} \right] \right\} d\omega. \end{aligned} \tag{4.34}$$

It follows from (4.24) in the proof of Theorem 4.1 that

$$\mathbf{x}^t[A_x^0 - A_x]\mathbf{x} = \frac{M^2}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathcal{P}_{x,M}(\omega)[S_x^{-1}(\omega, \phi_0) - S_x^{-1}(\omega, \phi)] d\omega, \tag{4.35}$$

where A_x^0 and A_x are $M^2 \times M^2$ matrices whose elements are $a_x^0(\mathbf{t}_i - \mathbf{t}_j)$ and $a_x(\mathbf{t}_i - \mathbf{t}_j)$, respectively, and

$$\begin{aligned} a_x^0(\mathbf{t}) &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} S_x^{-1}(\omega, \phi_0) \exp(\jmath \omega^t \mathbf{t}) d\omega \\ a_x(\mathbf{t}) &= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} S_x^{-1}(\omega, \phi) \exp(\jmath \omega^t \mathbf{t}) d\omega. \end{aligned}$$

The quadratic function $\mathbf{x}^t[A_x^0 - A_x]\mathbf{x}$ is the random part of (4.34) and has mean and variance

$$\mathrm{E}\left\{\mathbf{x}^t[A_x^0 - A_x]\mathbf{x}\right\} = \mathrm{tr}\left\{Q_x(A_x^0 - A_x)\right\} \tag{4.36}$$

and

$$\mathrm{V}\left\{\mathbf{x}^t[A_x^0 - A_x]\mathbf{x}\right\} = 2\mathrm{tr}\left\{[Q_x(A_x^0 - A_x)]^2\right\}, \tag{4.37}$$

respectively. The trace in (4.36) is computed by using Lemma 4.1, Section 4.5. This computation yields

$$\begin{aligned}\lim_{M\to\infty} \frac{1}{M^2}\mathrm{tr}\left\{Q_x(A_x^0 - A_x)\right\} &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi_0)[\mathcal{S}_x^{-1}(\omega,\phi_0) - \mathcal{S}_x^{-1}(\omega,\phi)]d\omega \\ &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[1 - \frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\right]d\omega. \end{aligned} \tag{4.38}$$

Applying Lemma 4.1 to (4.37) yields

$$\lim_{M\to\infty} \frac{2}{M^2}\mathrm{tr}\left\{[Q_x(A_x^0 - A_x)]^2\right\} = \frac{2}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[1 - \frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\right]^2 d\omega. \tag{4.39}$$

Combining the result (4.38) with the deterministic part of (4.34)

$$\begin{aligned}\lim_{M\to\infty} \mathrm{E}\left\{\frac{1}{M^2}[\tilde{l}_M(\phi_0) - \tilde{l}_M(\phi)]\right\} &= \frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\log\left[\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\right] + 1 - \frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\right\}d\omega. \end{aligned} \tag{4.40}$$

However, for any two positive functions f_1 and f_2 it is true that

$$\log\frac{f_1(x)}{f_2(x)} \leq \frac{f_1(x)}{f_2(x)} - 1,$$

and the condition is strict for all x such that $f_1(x) \neq f_2(x)$. Since $\phi \neq \phi_0$ implies $\mathcal{S}_x(\omega,\phi) \neq \mathcal{S}_x(\omega,\phi_0)$ almost everywhere (Assumption 4.3), it follows that

$$\lim_{M\to\infty} \mathrm{E}\left\{\frac{1}{M^2}[\tilde{l}_M(\phi_0) - \tilde{l}_M(\phi)]\right\} < 0. \tag{4.41}$$

On the other hand, the variance does not depend on the deterministic part, and therefore

$$\begin{aligned}\lim_{M\to\infty} \mathrm{V}\left\{\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]\right\} &= \lim_{M\to\infty}\frac{1}{2M^2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[1-\frac{S_x(\omega,\phi_0)}{S_x(\omega,\phi)}\right]^2 d\omega \\ &= 0. \end{aligned} \tag{4.42}$$

Since ϕ_0 and ϕ belong to the closed set Φ^c, there exists positive μ depending on ϕ_0 and ϕ such that

$$\lim_{M\to\infty} \mathrm{E}\left\{\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]\right\} = -\mu(\phi_0,\phi). \tag{4.43}$$

For $\delta > 0$ an application of Tchebycheff's inequality results in

$$\begin{aligned}\lim_{M\to\infty} \mathrm{P}\left\{\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]+\mu(\phi_0,\phi)<\delta\right\} &\geq \lim_{M\to\infty} \mathrm{P}\left\{\left|\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]+\mu(\phi_0,\phi)\right|<\delta\right\} \\ &\geq \lim_{M\to\infty}\left[1-\frac{1}{\delta^2}\mathrm{E}\left\{\left|\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]+\mu(\phi_0,\phi)\right|^2\right\}\right]. \end{aligned} \tag{4.44}$$

Letting $-k(\phi_0,\phi) = \delta - \mu(\phi_0,\phi)$, (4.44) is equivalent to

$$\begin{aligned}\lim_{M\to\infty} \mathrm{P}\left\{\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]<-k(\phi_0,\phi)\right\} &\geq 1-\frac{1}{\delta^2}\lim_{M\to\infty}\mathrm{V}\left\{\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]\right\}. \end{aligned} \tag{4.45}$$

Using the result of (4.42) we have that

$$\lim_{M\to\infty} \mathrm{P}\left\{\frac{1}{M^2}[\tilde{I}_M(\phi_0)-\tilde{I}_M(\phi)]<-k(\phi_0,\phi)\right\} = 1 \tag{4.46}$$

for $\phi \neq \phi_0$.

For each $\phi_1 \in \Phi^c$ choose $\delta(\phi_1)$, a positive constant possibly depending on ϕ_1, such that $|\phi_1 - \phi_2| < \delta(\phi_1)$ for all $\phi_2 \in \Phi$. Then compute

$$\begin{aligned}
\left|\frac{1}{M^2}[\tilde{I}_M(\phi_1) - \tilde{I}_M(\phi_2)]\right| &= \frac{1}{2(2\pi)^2}\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\log\left[\frac{\mathcal{S}_x(\omega,\phi_1)}{\mathcal{S}_x(\omega,\phi_2)}\right]\right.\right. \\
&\qquad \left.\left. +\mathcal{P}_{x,M}(\omega)[\mathcal{S}_x^{-1}(\omega,\phi_2) - \mathcal{S}_x^{-1}(\omega,\phi_1)]\right\} d\omega\right| \\
&\leq \frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\left|\log\left[\frac{\mathcal{S}_x(\omega,\phi_1)}{\mathcal{S}_x(\omega,\phi_2)}\right]\right|\right. \\
&\qquad \left. +\mathcal{P}_{x,M}(\omega)\left|\mathcal{S}_x^{-1}(\omega,\phi_2) - \mathcal{S}_x^{-1}(\omega,\phi_1)\right|\right\} d\omega \\
&\leq \frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{\left|\frac{\mathcal{S}_x(\omega,\phi_1)}{\mathcal{S}_x(\omega,\phi_2)} - 1\right|\right. \\
&\qquad \left. +\mathcal{P}_{x,M}(\omega)\left|\mathcal{S}_x^{-1}(\omega,\phi_2) - \mathcal{S}_x^{-1}(\omega,\phi_1)\right|\right\} d\omega \\
&= \frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{S}_x(\omega,\phi_1) + \mathcal{P}_{x,M}(\omega)] \\
&\qquad \cdot\left|\mathcal{S}_x^{-1}(\omega,\phi_2) - \mathcal{S}_x^{-1}(\omega,\phi_1)\right| d\omega. \qquad (4.47)
\end{aligned}$$

Now, define the set $\Delta(\phi_1) = \{\phi : |\phi_1 - \phi| < \delta(\phi_1)\}$ and let

$$H(\phi_1, \delta(\phi_1)) = \sup_{\phi\in\Delta(\phi_1)}\left\{\left|\frac{\partial}{\partial\phi}\mathcal{S}_x^{-1}(\omega,\phi)\right|\right\}. \qquad (4.48)$$

Combining (4.47) and (4.48) it follows from the mean value theorem that

$$\left|\frac{1}{M^2}[\tilde{I}_M(\phi_1) - \tilde{I}_M(\phi_2)]\right| \leq H_{\delta,M}(\phi_1) \qquad (4.49)$$

where

$$H_{\delta,M}(\phi_1) = \frac{\delta H(\phi_1, \delta(\phi_1))}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{S}_x(\omega,\phi_1) + \mathcal{P}_{x,M}(\omega)]d\omega. \qquad (4.50)$$

Furthermore, we have that

$$\begin{aligned}
&\mathrm{E}\left\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{S}_x(\omega,\phi_1) + \mathcal{P}_{x,M}(\omega)]d\omega\right\} \\
&\quad = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathcal{S}_x(\omega,\phi_1)d\omega + \frac{1}{M^2}\mathrm{E}\left\{\mathbf{x}^t I_{M^2}\mathbf{x}\right\} \\
&\quad = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathcal{S}_x(\omega,\phi_1)d\omega + \frac{1}{M^2}\mathrm{tr}\left\{Q_x\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi_1)d\omega + q_x(0) \\
&= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi_1)d\omega + \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi_0)d\omega \\
&< \infty
\end{aligned} \tag{4.51}
$$

and

$$
\begin{aligned}
&\lim_{M\to\infty} \mathrm{V}\left\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{S}_x(\omega,\phi_1)+\mathcal{P}_{x,M}(\omega)]d\omega\right\} \\
&\quad = \lim_{M\to\infty} \mathrm{V}\left\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathcal{P}_{x,M}(\omega)d\omega\right\} \\
&\quad = \lim_{M\to\infty} \mathrm{V}\left\{\frac{1}{M^2}\mathbf{x}^t I_{M^2}\mathbf{x}\right\} \\
&\quad = \lim_{M\to\infty} \frac{2}{M^4}\mathrm{tr}\left\{Q_x^2\right\} \\
&\quad = \lim_{M\to\infty} \frac{2}{M^2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathcal{S}_x^2(\omega,\phi_0)d\omega \\
&\quad = 0.
\end{aligned} \tag{4.52}
$$

From which it follows that

$$
\begin{aligned}
\lim_{\delta\to 0} \mathrm{E}\{H_{\delta,M}(\phi_1)\} &= \lim_{\delta\to 0}\frac{\delta H(\phi_1,\delta(\phi_1))}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{S}_x(\omega,\phi_1)+\mathcal{S}_x(\omega,\phi_0)]d\omega \\
&= 0
\end{aligned} \tag{4.53}
$$

and

$$
\begin{aligned}
\lim_{M\to\infty} \mathrm{V}\{H_{\delta,M}(\phi_1)\} &= \lim_{M\to\infty} \delta^2 H^2(\phi_1,\delta(\phi_1) \\
&\qquad \cdot \mathrm{V}\left\{\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{S}_x(\omega,\phi_1)+\mathcal{S}_x(\omega,\phi_0)]d\omega\right\} \\
&= 0.
\end{aligned} \tag{4.54}
$$

The theorem follows from the result contained in [58, Lemma 3, pg 106] which is applicable under the conditions (4.46), (4.53), and (4.54). The conclusion is that $\tilde{\phi}_M \to \phi_0$ in probability as $M \to \infty$. □

Next, we establish the asymptotic distribution properties of the approximate ML estimator $\tilde{\phi}_M$. The Fisher information matrix is defined to be

$$I_M(\phi) = \frac{1}{M^2}\mathrm{E}\left\{\left[\frac{\partial}{\partial\phi}l_M(\phi)\right]\left[\frac{\partial}{\partial\phi}l_M(\phi)\right]^t\right\}, \tag{4.55}$$

or, equivalently [54],

$$I_M(\phi) = -\frac{1}{M^2}\mathrm{E}\left\{\frac{\partial^2}{\partial\phi^2}l_M(\phi)\right\}. \tag{4.56}$$

Theorem 4.1 suggests that

$$\lim_{M\to\infty} I_M(\phi) = \lim_{M\to\infty}\left[-\frac{1}{M^2}\mathrm{E}\left\{\frac{\partial^2}{\partial\phi^2}l_M(\phi)\right\}\right].$$

Let us also define the matrix

$$\Gamma(\phi) = \frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]^t d\omega. \tag{4.57}$$

Taking the derivative of (4.11) with respect to the parameters ϕ we get

$$\frac{\partial}{\partial\phi}\tilde{l}_M(\phi) = \frac{M^2}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]\left[1-\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi)}\right]d\omega. \tag{4.58}$$

Differentiating (4.58) we obtain the matrix of second derivatives

$$\begin{aligned}
&\frac{\partial^2}{\partial\phi^2}\tilde{l}_M(\phi)\\
&= \frac{M^2}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi)}\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]^t d\omega\\
&\quad+\frac{M^2}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[1-\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi)}\right]\frac{\partial^2}{\partial\phi^2}\log\mathcal{S}_x(\omega,\phi)d\omega.
\end{aligned} \tag{4.59}$$

Under the assumed regularity conditions it follows that

$$\begin{aligned}
&\lim_{M\to\infty}\frac{1}{M^2}\mathrm{E}\left\{\frac{\partial}{\partial\phi}\tilde{l}_M(\phi)\right\}\\
&\quad= \frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\right]d\omega
\end{aligned} \tag{4.60}$$

and

$$\lim_{M\to\infty}\frac{1}{M^2}\mathrm{E}\left\{\frac{\partial^2}{\partial\phi^2}\tilde{l}_M(\phi)\right\}$$
$$= -\frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\phi)\right]^t d\omega$$
$$+\frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi)}\right]\frac{\partial^2}{\partial\phi^2}\log\mathcal{S}_x(\omega,\phi)d\omega. \qquad (4.61)$$

Here, we used

$$\lim_{M\to\infty}\mathrm{E}\{\mathcal{P}_{x,M}(\omega)\} = \lim_{M\to\infty}\sum_{\mathbf{t}\in\Omega_M}\mathrm{E}\{C_x(\mathbf{t})\}\exp(-\jmath\omega^t\mathbf{t})$$
$$= \mathcal{S}_x(\omega,\phi_0),$$

where $\mathcal{P}_{x,M}(\omega)$ is defined in (4.9) and the biased sample correlations $C_x(\mathbf{t})$ are given by

$$C_x(\mathbf{t}) = \frac{1}{M^2}\sum_{\mathbf{s}+\mathbf{t}\in\Omega_M}x(\mathbf{s})x(\mathbf{s}+\mathbf{t}).$$

Letting $\phi=\phi_0$ in (4.60) and (4.61), it follows that

$$\lim_{M\to\infty}\frac{1}{M^2}\mathrm{E}\left\{\frac{\partial}{\partial\phi}\tilde{l}_M(\phi_0)\right\} = \mathbf{0} \qquad (4.62)$$

and

$$\lim_{M\to\infty}\frac{1}{M^2}\mathrm{E}\left\{\frac{\partial^2}{\partial\phi^2}\tilde{l}_M(\phi_0)\right\} = -\Gamma(\phi_0). \qquad (4.63)$$

Consequently,

$$\lim_{M\to\infty}I_M(\phi_0) = \Gamma(\phi_0) \qquad (4.64)$$

in the sense of convergence in weak probability. We now give the following theorem.

Theorem 4.3 *Let Assumptions 4.1 through 4.5 apply, and furthermore, assume that* $\frac{\partial^2}{\partial\phi^2}\mathcal{S}_x(\omega,\phi)$ *exists for all* $\phi\in\Phi$ *and is continuous in* ω. *Then as* $M\to\infty$ *the vector* $M[\tilde{\phi}_M-\phi_0]$ *converges in distribution to* $\mathcal{N}(\mathbf{0},\Gamma^{-1}(\phi_0))$ *where* $\Gamma(\phi)$ *is given by (4.57).*

Proof: Let

$$a(\omega,\phi) = \mathbf{h}^t\left[\frac{\partial}{\partial\phi}\log \mathcal{S}_x(\omega,\phi)\right] \tag{4.65}$$

and

$$\Delta_M(\phi) = -\frac{1}{M}\left[\frac{\partial}{\partial\phi}\tilde{l}_M(\phi)\right], \tag{4.66}$$

where $\mathbf{h} = \text{col}\,\{h_1,\ldots,h_p\}$ is an arbitrary p-dimensional vector such that $|\mathbf{h}| \neq 0$. Combining (4.58) and (4.66) we compute

$$\begin{aligned}
&\mathbf{h}^t\Delta_M(\tilde{\phi}_M) - \mathbf{h}^t\Delta_M(\phi_0)\\
&\quad= \frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{h}^t\left[\frac{\partial}{\partial\phi}\log \mathcal{S}_x(\omega,\phi_0)\right]\left[1-\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi_0)}\right]d\omega\\
&\qquad-\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathbf{h}^t\left[\frac{\partial}{\partial\phi}\log \mathcal{S}_x(\omega,\tilde{\phi}_M)\right]\left[1-\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]d\omega\\
&\quad= \frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left\{a(\omega,\phi_0)\left[1-\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\phi_0)}\right]\right.\\
&\qquad\left.-a(\omega,\tilde{\phi}_M)\left[1-\frac{\mathcal{P}_{x,M}(\omega)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]\right\}d\omega,
\end{aligned} \tag{4.67}$$

and after some algebraic manipulation this becomes

$$\begin{aligned}
&\mathbf{h}^t\Delta_M(\tilde{\phi}_M) - \mathbf{h}^t\Delta_M(\phi_0)\\
&\quad= \frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}[\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]\left[\frac{a(\omega,\tilde{\phi}_M)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}-\frac{a(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi_0)}\right]d\omega\\
&\qquad-\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}a(\omega,\tilde{\phi}_M)\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]d\omega.
\end{aligned}$$

But $\tilde{\phi}_M$ is a solution of $\frac{\partial}{\partial\phi}\tilde{l}_M(\phi) = \mathbf{0}$, hence $\Delta_M(\tilde{\phi}_M) = \mathbf{0}$ and for all $\mathbf{h}$

$$\begin{aligned}
\mathbf{h}^t\Delta_M(\phi_0) \;=\; &\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}a(\omega,\tilde{\phi}_M)\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]d\omega\\
&-\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{a(\omega,\tilde{\phi}_M)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}-\frac{a(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi_0)}\right]\\
&\cdot[\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]d\omega.
\end{aligned} \tag{4.68}$$

Consider only the first term in (4.68). The mean value theorem implies that there exists $\phi^* = (1-\delta)\phi_0 + \delta\tilde{\phi}_M$, for some $0 \leq \delta \leq 1$, such that

$$\begin{aligned}
&\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} a(\omega,\tilde{\phi}_M)\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]d\omega \\
&\quad = \frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{a(\omega,\tilde{\phi}_M)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}[\mathcal{S}_x(\omega,\tilde{\phi}_M)-\mathcal{S}_x(\omega,\phi_0)]d\omega \\
&\quad = \frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{a(\omega,\tilde{\phi}_M)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\left[\frac{\partial}{\partial\phi}\mathcal{S}_x(\omega,\phi^*)\right]^t[\tilde{\phi}_M-\phi_0]d\omega,
\end{aligned}$$

which upon again using the definition (4.65) yields

$$\begin{aligned}
&\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} a(\omega,\tilde{\phi}_M)\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]d\omega \\
&\quad = M[\tilde{\phi}_M-\phi_0]^t\left\{\frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{1}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\frac{\partial}{\partial\phi}\mathcal{S}_x(\omega,\phi^*)\right]\right. \\
&\qquad \left.\cdot\left[\frac{\partial}{\partial\phi}\log\mathcal{S}_x(\omega,\tilde{\phi}_M)\right]^t d\omega\right\}\mathbf{h}. \qquad (4.69)
\end{aligned}$$

Theorem 4.2 showed that $\operatorname{plim}_{M\to\infty}\tilde{\phi}_M = \phi_0$ and this implies $\operatorname{plim}_{M\to\infty}\phi^* = \phi_0$. Clearly,

$$\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} a(\omega,\tilde{\phi}_M)\left[1-\frac{\mathcal{S}_x(\omega,\phi_0)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}\right]d\omega \to M[\tilde{\phi}_M-\phi_0]^t\Gamma(\phi_0)\mathbf{h} \qquad (4.70)$$

in probability as $M\to\infty$.

Now consider the second term in (4.68). Again, the mean value theorem implies that for some ϕ^*, $0 \leq \delta \leq 1$,

$$\begin{aligned}
&\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{a(\omega,\tilde{\phi}_M)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}-\frac{a(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi_0)}\right][\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]d\omega \\
&\quad = M[\tilde{\phi}_M-\phi_0]^t\left\{\frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\phi}\left[\frac{a(\omega,\phi^*)}{\mathcal{S}_x(\omega,\phi^*)}\right]\right. \\
&\qquad \left.\cdot[\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]d\omega\right\}. \qquad (4.71)
\end{aligned}$$

It is a simple matter using (4.65) to show that

$$
\begin{aligned}
-\frac{\partial}{\partial\phi}\left[\frac{a(\omega,\phi^*)}{\mathcal{S}_x(\omega,\phi^*)}\right] &= \left\{\frac{1}{\mathcal{S}_x^2(\omega,\phi^*)}\frac{\partial^2}{\partial\phi^2}\mathcal{S}_x(\omega,\phi^*)\right. \\
&\qquad \left. -\frac{2}{\mathcal{S}_x^3(\omega,\phi^*)}\left[\frac{\partial}{\partial\phi}\mathcal{S}_x(\omega,\phi^*)\right]\left[\frac{\partial}{\partial\phi}\mathcal{S}_x(\omega,\phi^*)\right]^t\right\}\mathbf{h} \\
&= \frac{\partial^2}{\partial\phi^2}\mathcal{S}_x^{-1}(\omega,\phi^*)\mathbf{h},
\end{aligned}
$$

and if we use this relationship in (4.71), we get that

$$
\begin{aligned}
&\frac{M}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{a(\omega,\tilde{\phi}_M)}{\mathcal{S}_x(\omega,\tilde{\phi}_M)}-\frac{a(\omega,\phi_0)}{\mathcal{S}_x(\omega,\phi_0)}\right][\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]d\omega \\
&\qquad = M[\tilde{\phi}_M-\phi_0]^t\left\{\frac{1}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\partial^2}{\partial\phi^2}\mathcal{S}_x^{-1}(\omega,\phi^*)\right. \\
&\qquad\qquad \left.\cdot[\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]d\omega\right\}. \qquad (4.72)
\end{aligned}
$$

Now let

$$
b_x(\mathrm{t}) = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\partial^2}{\partial\phi^2}\mathcal{S}_x^{-1}(\omega,\phi^*)\exp(j\omega^t\mathrm{t})d\omega \qquad (4.73)
$$

and let B_x^* be the matrix whose i,jth element is $b_x(\mathrm{t}_i-\mathrm{t}_j)$, $\mathrm{t}_i,\mathrm{t}_j\in\Omega_M$. Then, using the relationship (4.24) and Lemma 4.1, Section 4.5, it follows that

$$
\begin{aligned}
&\lim_{M\to\infty}\frac{1}{M^2}[\mathbf{x}^tB_x^*\mathbf{x}-\mathrm{tr}\,\{Q_xB_x^*\}] \\
&\qquad = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{\partial^2}{\partial\phi^2}\mathcal{S}_x^{-1}(\omega,\phi^*) \\
&\qquad\qquad \cdot[\mathcal{P}_{x,M}(\omega)-\mathcal{S}_x(\omega,\phi_0)]d\omega. \qquad (4.74)
\end{aligned}
$$

Moreover, we have that

$$
\begin{aligned}
\lim_{M\to\infty}\frac{1}{M^2}\mathrm{E}\left\{\mathbf{x}^tB_x^*\mathbf{x}-\mathrm{tr}\,\{Q_xB_x^*\}\right\} &= \lim_{M\to\infty}\frac{1}{M^2}\mathrm{tr}\,\{Q_xB_x^*-Q_xB_x^*\} \\
&= 0 \qquad (4.75)
\end{aligned}
$$

and

$$\begin{aligned}\lim_{M\to\infty}\frac{1}{M^4}\mathrm{V}\left\{\mathrm{x}^t B_x^*\mathrm{x}-\mathrm{tr}\left\{Q_x B_x^*\right\}\right\} &= \lim_{M\to\infty}\frac{1}{M^4}\mathrm{E}\left\{[\mathrm{x}^t B_x^*\mathrm{x}]^2\right\}\\ &= \frac{2}{M^4}\mathrm{tr}\left\{[Q_x B_x^*]^2\right\}\\ &= 0 \qquad (4.76)\end{aligned}$$

Consequently, since $\tilde{\phi}_M$ is a consistent estimator of ϕ_0, the left side of (4.71) and the second term of (4.68) converges in probability to zero as $M \to \infty$. Hence, $M\Gamma(\phi_0)[\tilde{\phi}_M - \phi_0] - \Delta_M(\phi_0)$ converges to zero in probability.

Now, to complete the proof we need only show that $\Delta_M(\phi_0)$ is an asymptotically normal p-dimensional random vector distributed as $\mathcal{N}(0, \Gamma(\phi_0))$. We digress for the moment to discuss the log likelihood ratio, a device first suggested by LeCam [59] to replace the artificial requirements of 2nd and 3rd-order differentiability of the likelihood function used to prove asymptotic normality of ML estimates. For example, see [60, 61, 62]. The likelihood ratio has been thoroughly studied by Ibragimov and Has'minskii [63] for ML and Bayesian parameter estimation from independent data. They also suggest that much of their analysis of the likelihood ratio is applicable to dependent situations. Dzhaparidze [56] applies this device to prove asymptotic normality of ML parameter estimates in the more general 1-D time series setting. The results we present below for our 2-D problem follow Dzhaparidze's arguments, and so we omit the details. Instead, for readers interested in the complete proof, which is extensive and encompasses many preliminary results, we suggest [56] for review.

To conclude the proof of the asymptotic normality of $M(\tilde{\phi}_M - \phi_0)$ we begin by defining the log likelihood ratio. Denote by $\mathcal{F}_M(\mathcal{S}_x(\omega, \phi))$ or equivalently by $\mathcal{F}_{M,\phi}$ the Gaussian probability distribution which corresponds to the set of observations

$\mathcal{X}_M$ whose spectral density function is $\mathcal{S}_x(\omega, \phi)$. The log of the likelihood ratio is defined to be the log of the Radon-Nikodym derivative given by

$$\Lambda(\phi_1, \phi_2) = \log \left[\frac{d\mathcal{F}_{M,\phi_2}}{d\mathcal{F}_{M,\phi_1}} \right], \tag{4.77}$$

which in the case of the absolutely continuous Gaussian distributions is equivalent to

$$\Lambda(\phi_1, \phi_2) = \frac{1}{2} \left\{ \log[\det\{Q_{S_1}\}] - \log[\det\{Q_{S_2}\}] + \mathbf{x}^t [Q_{S_1}^{-1} - Q_{S_2}^{-1}]\mathbf{x} \right\} \tag{4.78}$$

where Q_{S_1} and Q_{S_2} are the $M^2 \times M^2$ covariance matrices for $\mathcal{S}_x(\omega, \phi_1)$ and $\mathcal{S}_x(\omega, \phi_2)$, respectively.

One then proves that if $\mathcal{S}$ and $\mathcal{G}$ are spectral density functions and $\mathcal{G}_M \to \mathcal{G}$ as $M \to \infty$, then

$$\Lambda(\mathcal{S}, \mathcal{G}) - \Lambda(\mathcal{S}, \mathcal{G}_M) \to 0 \tag{4.79}$$

in $\mathcal{F}_{M,\phi}$ probability as $M \to \infty$. In particular, if $\mathcal{S} = \mathcal{S}_x(\omega, \phi_0)$ and $\mathcal{G}_M = \mathcal{S}_x(\omega, \tilde{\phi}_M)$, then

$$\begin{aligned} \Lambda(\phi_0, \tilde{\phi}_M) - \frac{M}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\mathcal{P}_{x,M}(\omega) - \mathcal{S}_x(\omega, \phi_0)}{\mathcal{S}_x(\omega, \phi_0)} \mathbf{h}^t \left[\frac{\partial}{\partial \phi} \log \mathcal{S}_x(\omega, \phi_0) \right] d\omega \\ + \frac{M^2}{4(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \left[\frac{\partial}{\partial \phi} \log \mathcal{S}_x(\omega, \phi_0) \right]^2 d\omega \\ = \Lambda(\phi_0, \tilde{\phi}_M) - \mathbf{h}^t \boldsymbol{\Delta}_M(\phi_0) + \frac{1}{2} \mathbf{h}^t \Gamma(\phi_0) \mathbf{h} \to 0 \end{aligned} \tag{4.80}$$

in $\mathcal{F}_{M,\phi_0}$ probability as $M \to \infty$.

This last result requires for matrices of Fourier coefficients Q_S, Q_a, and $Q_{a/S}$ corresponding to the functions $\mathcal{S}_x(\omega, \phi_0)$, $a(\omega, \phi_0)$, and $\frac{a(\omega, \phi_0)}{\mathcal{S}_x(\omega, \phi_0)}$, respectively, that the mean and variance of the quadratic form

$$\mathbf{x}^t [Q_S^{-1} Q_a - Q_{a/S}] \mathbf{x}$$

are bounded. This fact is guaranteed if $q_x(\mathbf{t})$ and $p_x(\mathbf{t})$, given by

$$q_x(\mathbf{t}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \mathcal{S}_x(\omega, \phi_0) \exp(\jmath \omega^t \mathbf{t}) d\omega \tag{4.81}$$

and

$$p_x(\mathbf{t}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{a(\omega, \phi_0)}{\mathcal{S}_x(\omega, \phi_0)} \exp(\jmath \omega^t \mathbf{t}) d\omega, \tag{4.82}$$

satisfy the conditions

$$\sum_{\mathbf{t} \in \Omega_\infty} |q_x(\mathbf{t})|^2 |\mathbf{t}|^2 < \infty \tag{4.83}$$

and

$$\sum_{\mathbf{t} \in \Omega_\infty} |p_x(\mathbf{t})|^2 |\mathbf{t}|^2 < \infty. \tag{4.84}$$

That (4.83) implies (4.84) is an open question, but this assumption seems to be no more a severe restriction than those already imposed by Assumptions 4.1 through 4.5 since the NCAR signal plus noise model satisfies these conditions.

A parametric family of distributions $\{\mathcal{F}_{M,\phi},\ \phi \in \Phi\}$ and $\Phi \subset \Re^p$ is called locally asymptotically normal (at a fixed point ϕ_0) if there exists a sequence of p-dimensional random vectors $\{\Delta_M(\phi_0),\ M = 1, 2, \ldots\}$ and a positive definite $p \times p$ matrix $\Gamma(\phi_0)$ such that

1. $\Lambda(\phi_0, \phi + \frac{1}{M}\mathbf{h}) - \mathbf{h}^t \Delta_M(\phi) + \frac{1}{2}\mathbf{h}^t \Gamma(\phi)\mathbf{h} \rightarrow 0$ in $\mathcal{F}_{M,\phi}$ probability as $M \rightarrow \infty$ for any p-dimensional vector $\mathbf{h}$, and

2. $\Delta_M(\phi_0) \rightarrow \mathcal{N}(\mathbf{0}, \Gamma(\phi_0))$ in distribution as $M \rightarrow \infty$.

That the family of Gaussian distributions $\{\mathcal{F}_{M,\phi},\ \phi \in \Phi\}$ resulting from the Gaussian random field $\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$ with spectral density function $\mathcal{S}_x(\omega, \phi)$ is locally asymptotically normal in the sense that the vector $\Delta_M(\phi)$ in (4.80) is distributed as $\mathcal{N}(\mathbf{0}, \Gamma(\phi))$ as $M \rightarrow \infty$ is a consequence of [56, Theorem A1.2, Appendix 1 to Chapter 1].

The theorem follows. □

Theorem 4.4 *Under the conditions of Theorem 4.3 and if $x(\mathbf{s})$ is Gaussian, then $\tilde{\phi}_M$ is an asymptotically efficient estimator of ϕ_0.*

Proof: The theorem follows from (4.64) and the fact that the true ML estimator is an efficient estimator when $\{x(\mathbf{s})\}$ is a Gaussian process. □

4.2.2 Existence and Uniqueness

Existence and uniqueness questions naturally arise in view of the possibilities that the approximate ML equation

$$\lambda_M(\phi) = \frac{\partial}{\partial \phi} \tilde{l}_M(\phi) = 0. \tag{4.85}$$

may have no solution or more than one. The existence question is easily answered by the following argument. The approximate ML estimate $\tilde{\phi}_M$, being consistent (Theorem 4.2), converges in weak probability to ϕ_0. Since ϕ_0 lies within an open set (Assumption 4.1), $\tilde{l}_M(\phi)$ takes on its minimum value within an open set. The existence in probability of a root of (4.85) is guaranteed by the fact that when a differentiable function reaches a minimum in an open set, the derivatives must vanish at that point [64]. Thus, at the least, $\tilde{\phi}_M$ solves (4.85) asymptotically.

On the other hand, uniqueness is not so easily established. For linear 2-D models uniqueness has been established by showing that the approximate ML function is convex [12, 35]. However, Sharma and Chellappa's convexity proof [12, Theorem 1] has errors even though their conclusion appears correct in view of the correspondence between Guassian-Markov random field (GMRF) models [35] and NCAR models. Here, for the NCAR plus noise model convexity of $\tilde{l}_M(\phi)$

is not readily apparent. But even if the solution is unique in the Gaussian case, it may not remain that way when working with real data, which is rarely Gaussian and is certainly not for sinusoidal signals with additive Gaussian noise. Hence, we only suggest here that the theoretical asymptotic results of consistency, efficiency, and normality apply to $\tilde{\phi}_M$ and not to just any solution of (4.85). Consequently, for the empirical solution one must guarantee that the true optimum point of the approximate ML function is computed and not rely on any stationary point when computing the spectrum estimate.

4.2.3 The Asymptotic Spectrum Estimate

In this section we have proven the theoretically important asymptotic properties of consistency, normality, and efficiency of the approximate ML parameter estimates for a Gaussian random field having the regularity properties described in Assumptions 4.1 through 4.5. Since the NCAR plus noise model (4.1) and (3.1) has these properties, the approximate ML yields a consistent estimate of the spectral density function $\mathcal{S}_x(\omega, \phi)$ of $\mathcal{X}_\infty$. This follows from the property of consistent estimators that any continuous function of a consistent estimate is also consistent [54]. That these properties also apply to the toroidal lattice model is clear from the fact that the discreet Fourier transfrom (4.13) converges to the Fourier transform (4.10) as the number of observations tends to infinity.

4.3 Solving for the Approximate ML Estimates

Equation (4.14) is highly nonlinear in the parameters, and a good nonlinear minimization algorithm is required to solve for the spectrum parameters. There

are many gradient search algorithms available for this purpose such as Newton-Raphson, conjugate gradient, or modified Newton method. Regardless of the algorithm the optimization procedure to minimize (4.14) requires initial starting values for the parameters. These initial values must be chosen judiciously in view of the fact that (4.14) may have local minimum or other stationary points where the algorithm may prematurely terminate and result in parameter estimates yielding inaccurate spectra. The initial values should be as near the global minimum as possible to avoid local stationarities.

There are several possibilities. For example, least squares (LS) are often used for initial conditions, but these are not consistent for NCAR models and for low SNRs the least squares estimate of β^2 will be unrealistically large. In [65] bias-compensated least squares (BCLS) estimates for the signal plus noise model were proposed. Let $\mathbf{x_s} = \text{col}\ \{x(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$ be the p-dimensional neighborhood vector, then these estimates for the NCAR plus noise model are

$$\hat{\theta}_{BCLS} = \left[\sum_{\mathbf{s}\in\Omega_M} \mathbf{x_s}\mathbf{x_s}^t - \hat{\gamma}^2 M^2 I_p\right]^{-1} \left[\sum_{\mathbf{s}\in\Omega_M} \mathbf{x_s} x(\mathbf{s})\right] \tag{4.86}$$

and

$$\hat{\beta}^2_{BCLS} = \frac{1}{M^2}\sum_{\mathbf{s}\in\Omega_M} [x(\mathbf{s}) - \hat{\theta}^t_{BCLS}\mathbf{x_s}]^2 - \hat{\gamma}^2(1 + \hat{\theta}^t_{BCLS}\hat{\theta}_{BCLS}). \tag{4.87}$$

To compute the bias-compensated least squares estimate an initial guess for γ^2 is needed. In most array processing applications the signal noise can be measured during periods when the signal is absent, thus yielding an estimate $\hat{\gamma}^2_0$ to be used in (4.86) and (4.87). For low pass signals a value of $|X(\mathbf{s})|^2$ at large spatial frequencies may be used since the noise power is distributed evenly over the whole spectrum and the signal power is concentrated near the origin [66]. Alternatively, the following procedure will yield a good starting value for γ^2 since $\{|X(\mathbf{s})|^2,\ \mathbf{s} \in$

$\Omega_M\}$ for sinusoids in noise contains regions between peaks with contributions from only noise.

1. Compute the periodogram $|X(\mathbf{s})|^2$,

2. Smooth the periodogram with a $n \times n$ averaging window filter to obtain $\overline{|X(\mathbf{s})|}^2$, and

3. Let $\hat{\gamma}_0^2 = \min \overline{|X(\mathbf{s})|}^2$.

Another alternative for initial conditions is to compute estimates of $\boldsymbol{\theta}$ based on estimated correlations similar to the procedure in [65]. Define $\mathbf{y_s} = \text{col}\,\{y(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$, $\mathbf{x_s} = \text{col}\,\{x(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$, and $\mathbf{n_s} = \text{col}\,\{n(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$. Then

$$x(\mathbf{s}) = \boldsymbol{\theta}^t(\mathbf{x_s} - \mathbf{n_s}) + w(\mathbf{s}) + n(\mathbf{s}) \tag{4.88}$$

and

$$\begin{aligned} \mathrm{E}\{x(\mathbf{s})x(\mathbf{s}+\mathbf{t})\} &= \boldsymbol{\theta}^t\mathrm{E}\{\mathbf{x_s}x(\mathbf{s}+\mathbf{t})\} - \boldsymbol{\theta}^t\mathrm{E}\{\mathbf{n_s}x(\mathbf{s}+\mathbf{t})\} & (4.89)\\ &\quad + \mathrm{E}\{w(\mathbf{s})x(\mathbf{s}+\mathbf{t})\} + \mathrm{E}\{n(\mathbf{s})x(\mathbf{s}+\mathbf{t})\}. & (4.90) \end{aligned}$$

Now let $p(\mathbf{t}) = \sum_{\mathbf{s}\in\Omega_M} x(\mathbf{s})x(\mathbf{s}+\mathbf{t})$ and $\mathbf{q_t} = \sum_{\mathbf{s}\in\Omega_M} \mathbf{x_s}x(\mathbf{s}+\mathbf{t})$. Equation (4.90) implies that

$$p(\mathbf{t}) \approx \mathbf{q_t^t}\boldsymbol{\theta}. \tag{4.91}$$

for values of $\mathbf{t} = t_1, t_2$ outside the local region where the data is correlated. Generally, for the NCAR model this region is defined by the set $\{\mathbf{t}|\mathbf{t} \notin N'\}$ where $N' = \{\mathbf{r}, 2\mathbf{r}|\mathbf{r} \in N \cup (0,0)\}$. Now define

$$\mathbf{p} = \text{col}\,\{p(1), \ldots, p(T)\}$$

and

$$Q = \begin{bmatrix} \mathbf{q}_1^t \\ \vdots \\ \mathbf{q}_T^t \end{bmatrix}$$

where T is the number of elements t. Then

$$\mathbf{p} \approx Q\boldsymbol{\theta}.$$

If T is equal to the number of parameters, then take as an estimate of $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = Q^{-1}\mathbf{p} \tag{4.92}$$

assuming Q^{-1} exists. Better results will be obtained if T is greater than the number of parameters. In this case let $\mathbf{e} = \mathbf{p} - Q\boldsymbol{\theta}$ and $l^2 = \frac{1}{2}\mathbf{e}^t\mathbf{e}$. To estimate $\boldsymbol{\theta}$, take the value minimizing l^2 which is given by

$$\hat{\boldsymbol{\theta}}_{LSCor} = [Q^tQ]^{-1}Q^t\mathbf{p} \tag{4.93}$$

or

$$\hat{\boldsymbol{\theta}}_{LSCor} = \left[\sum_{t \notin N'} \mathbf{q}_t\mathbf{q}_t^t\right]^{-1}\left[\sum_{t \notin N'} \mathbf{q}_t p(\mathrm{t})\right]. \tag{4.94}$$

β^2 can then be computed by (4.87) given a value for γ^2. These estimates will be designated as the least squares correlation (LSCor) estimates.

Finally, one can compute the ordinary LS estimate for $\boldsymbol{\theta}$ and compensate the LS estimate of β^2 by using (4.87). These estimates will be designated as the compensated least squares (CompLS) estimates.

The model order must also be determined. Generally, for Gaussian data the Akaike criteria [48] can be used to find the optimum order. However, in theory it is possible to use second order autoregressive models for modeling the spectra of sinusoids [25]. In the presence of added noise an NCAR only model will require

larger model orders; but, this should not be the case with the NCAR plus noise model. The added noise should have no impact on the order required of the autoregressive part of the model. Thus, the model order will be a direct function of the number of sinusoidal signals present.

4.4 Experimental Results

To compare NCAR plus noise modeling with NCAR only modeling for spectrum estimation we conducted several experiments using both Gaussian data and simulated planar array data.

Experiment 1: In this experiment we evaluated and compared the performance of NCAR plus noise spectrum modeling with that of NCAR only spectrum modeling when confronted with Gaussian data. We began by generating a 64×64 array of equally spaced Gaussian data obeying a 2nd-order NCAR random field having the parameter values shown in Table 4.1. The procedure in [16] was used to generate the data on a toroidal lattice.

$\theta_{1,0} = \theta_{-1,0}$	$\theta_{0,1} = \theta_{0,-1}$	$\theta_{1,1} = \theta_{-1,-1}$	$\theta_{-1,1} = \theta_{1,-1}$	β^2
.19450	.05710	−.13600	.23470	1.0

Table 4.1: Model parameters used to generate NCAR signal data used in the experiments.

In part 1 of this experiment we computed spectrum estimates by fitting (estimating parameters) a 2nd-order NCAR model and a 2nd-order NCAR plus noise model to the toroidal lattice noise free NCAR data, i.e., $\gamma^2 = \sigma_n^2 = 0.0$. Parameter

estimates were computed using a gradient descent algorithm to minimize the approximate likelihood given in (4.14). Least squares estimates were used as initial conditions to start the NCAR only minimization while bias-compensated LS estimates were used for the NCAR plus noise minimization. The initial additive noise variance estimate was arbitrarily selected as $\hat{\sigma}_n^2 = 0.11064$. Estimated spectra for both models were then computed by substituting the estimated values of β and θ, shown in Table 4.2 into (4.15) and computing the spectrum $\mathcal{S}_x(\omega_s) = \mathcal{S}_y(\omega_s)$ at the lattice points in the quarter spatial frequency plane. A plot of the unsmoothed periodogram $\{10 \log |X(s)|^2\}$ of the noise free NCAR data is shown in Figure 4.1, and the theoretical spectrum $\{10 \log \mathcal{S}_y(\omega_s)\}$ using the true values of the parameters from Table 4.1 is shown in Figure 4.2. Figure 4.3 shows a plot of the estimated signal spectrum $\{10 \log \mathcal{S}_y(\omega_s)\}$ using NCAR only model parameter estimates, and Figure 4.4 shows a plot of the estimated signal spectrum $\{10 \log \mathcal{S}_y(\omega_s)\}$ using the NCAR plus noise model parameter estimates. For more exact comparison Figure 4.5 is a plot comparing a slice of each spectrum along the column $\left\{\left(\omega_1 = \frac{k\pi}{32}, \omega_2 = \frac{12\pi}{32}\right) | k = 0, \ldots, 31\right\}$ and Figure 4.6 is a plot comparing a slice of each spectrum along the row $\left\{\left(\omega_1 = \frac{10\pi}{32}, \omega_2 = \frac{k\pi}{32}\right) | k = 0, \ldots, 31\right\}$.

For this noise free data both procedures, using identical model orders, computed essentially the same values of the signal model coefficients and yielded similar estimated signal spectra. The estimated noise variance computed with the NCAR plus noise model was $\hat{\sigma}_n^2 = .00409$, essentially zero, as expected.

In the second part of this experiment we added IID Gaussian noise with mean zero and variance $\sigma_n^2 = 3.87363$ to the NCAR signal data, yielding data with an

approximate signal-to-noise ratio (SNR) of 0db. Here, SNR is defined by

$$\text{SNR} = 10 \log \left(\frac{\sigma_y^2}{\sigma_n^2} \right), \tag{4.95}$$

where $\sigma_y^2 = \text{V}\{y(\text{s})\}$ and $\sigma_n^2 = \text{V}\{n(\text{s})\}$.

We computed the toroidal ML estimates of β and $\boldsymbol{\theta}$ for the NCAR only model and β, $\boldsymbol{\theta}$, and σ_n for the NCAR plus noise model. These estimates are shown in Table 4.2. The unsmoothed periodogram $\{10 \log |X(\text{s})|^2\}$ of the signal plus noise is shown in Figure 4.7 and the estimated signal plus noise spectrum $\{10 \log \mathcal{S}_x(\omega_\text{s})\}$ and estimated signal only spectrum $\{10 \log \mathcal{S}_y(\omega_\text{s})\}$ using 2nd-order models are shown in Figures 4.8 and 4.9, respectively. Here, $\mathcal{S}_x(\omega_\text{s}) \neq \mathcal{S}_y(\omega_\text{s})$. These results clearly show that a 2nd-order NCAR only model is not sufficient to estimate the spectrum well, so the spectrum was estimated using increasingly higher order NCAR only models. The result of a 5th-order NCAR only model is shown in comparison to the theoretical signal plus noise spectrum and the 2nd-order NCAR only model spectrum in the spectra-slice plots of Figures 4.10 and 4.11.

Model	SNR	$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2	σ_n^2
True Values	N/A	.19450	.05710	−.13600	.23470	1.0	0.0/3.87363
NCAR only	no noise	..19020	.06168	−.13321	.23432	1.01166	N/A
NCAR+N	no noise	.19052	.06162	−.13335	.23454	1.00560	0.00409
NCAR only	0db	.09247	.04638	−.07035	.13178	5.98819	N/A
NCAR+N	0db	.19040	.04472	−.13033	.23165	1.02169	3.86090

Table 4.2: Toroidal ML estimates of 2nd-order NCAR only and NCAR plus noise model parameters.

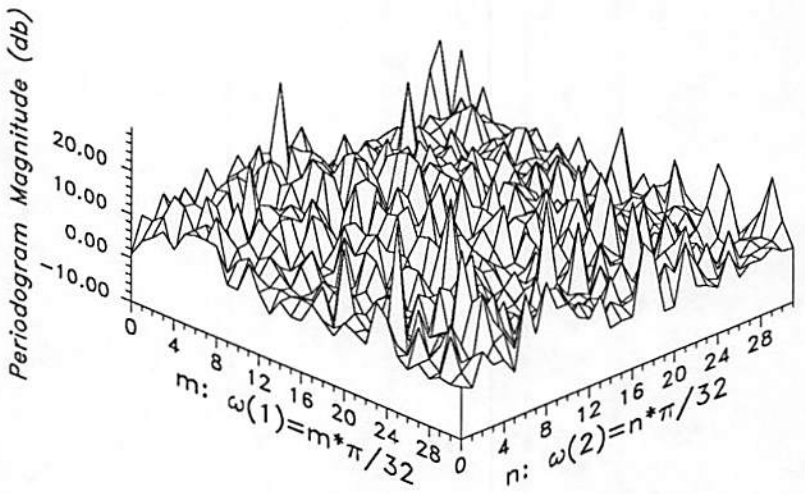


Figure 4.1: Unsmoothed periodogram of 2nd-order NCAR signal with no noise.

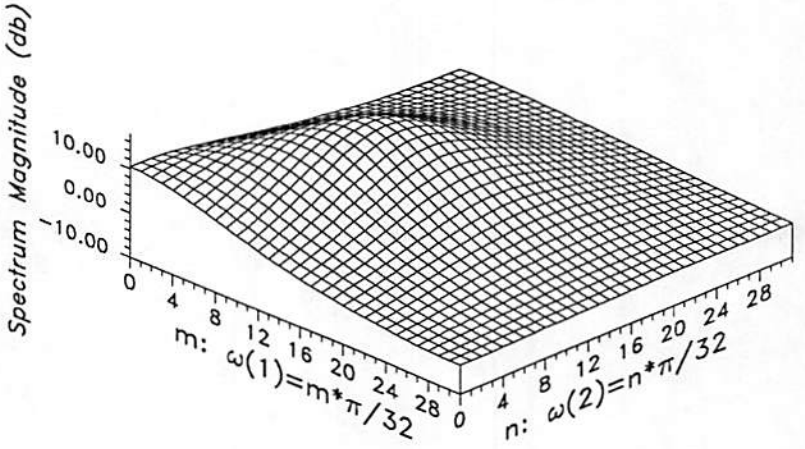


Figure 4.2: Theoretical spectrum of 2nd-order NCAR signal with no noise.

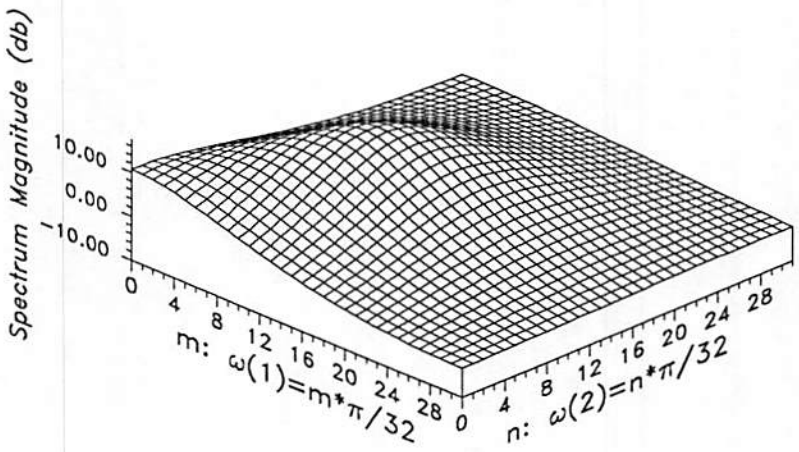


Figure 4.3: Estimated spectrum of NCAR signal with no noise using 2nd-order NCAR only model.

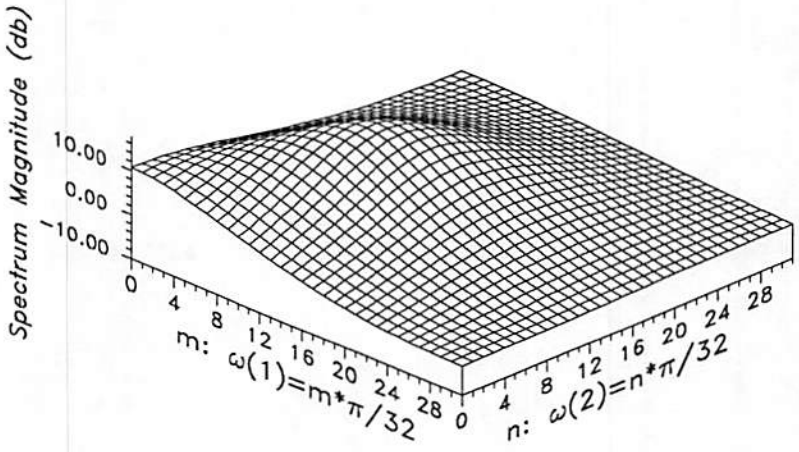


Figure 4.4: Estimated spectrum of NCAR signal with no noise using 2nd-order NCAR plus noise model.

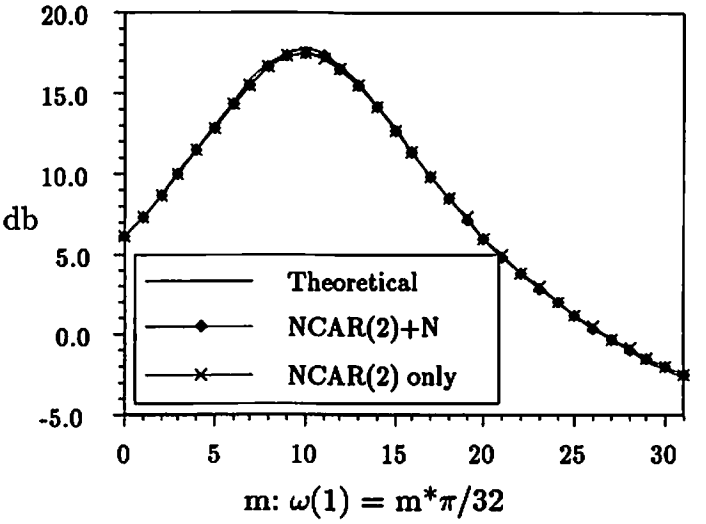


Figure 4.5: Column slices of 2-D spectra of 2nd-order NCAR signal with no noise.

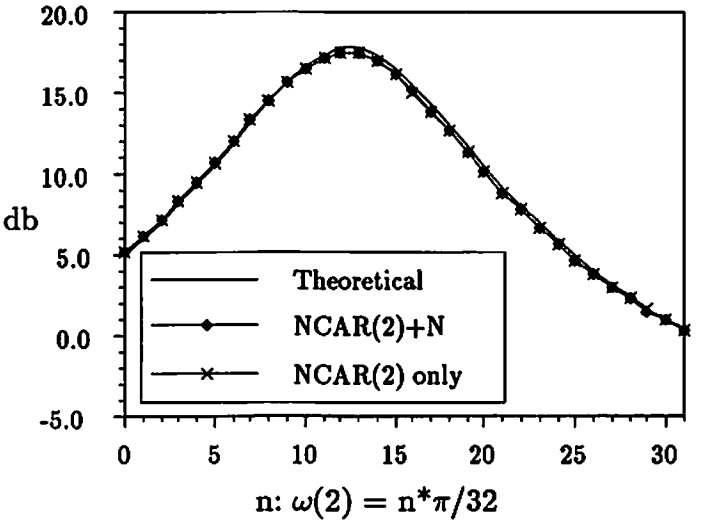


Figure 4.6: Row slices of 2-D spectra of 2nd-order NCAR signal with no noise.

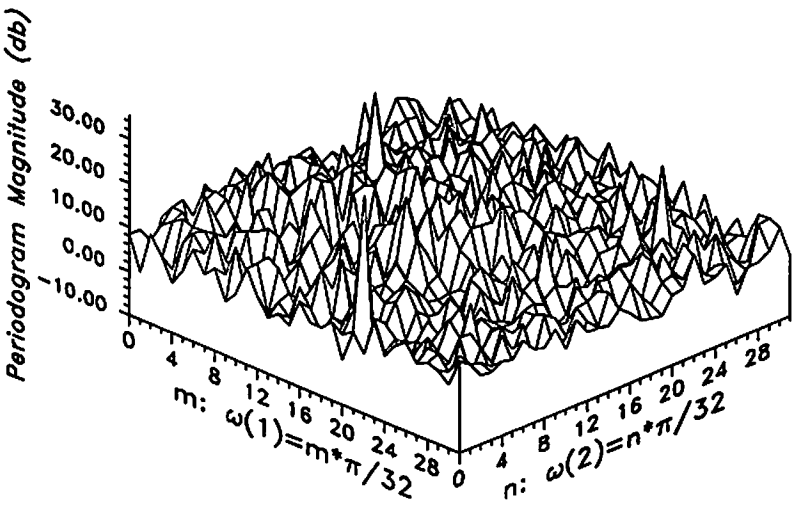


Figure 4.7: Unsmoothed periodogram of 2nd-order NCAR signal in Gaussian noise with SNR = 0db.

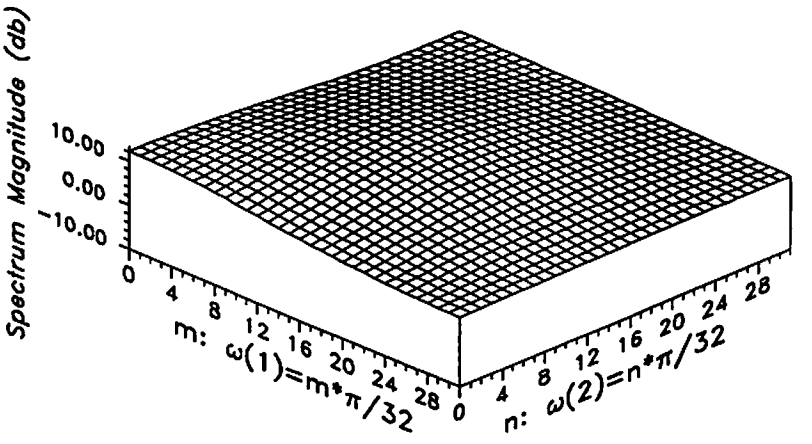


Figure 4.8: Estimated spectrum of NCAR signal in Gaussian noise with SNR = 0db using 2nd-order NCAR only model.

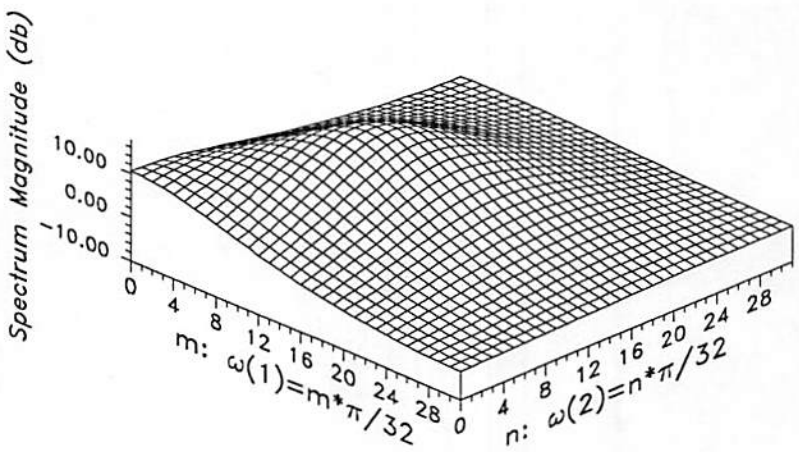


Figure 4.9: Estimated spectrum of NCAR signal in Gaussian noise with SNR = 0db using 2nd-order NCAR plus noise model.

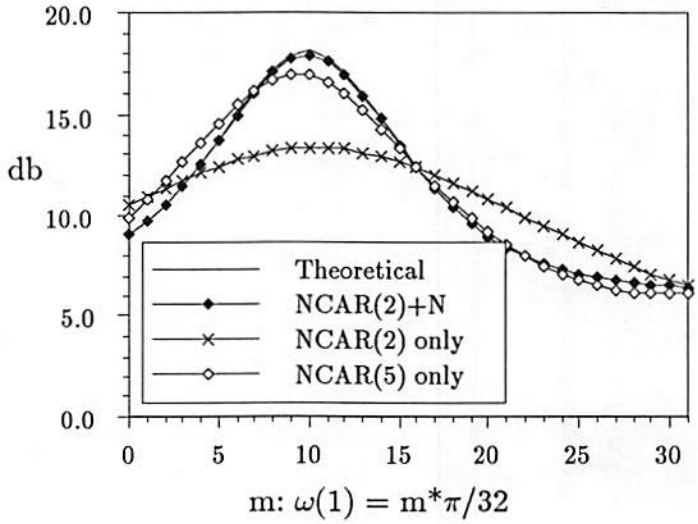


Figure 4.10: Column slices of 2-D spectra of 2nd-order NCAR signal in Gaussian noise with SNR = 0db.

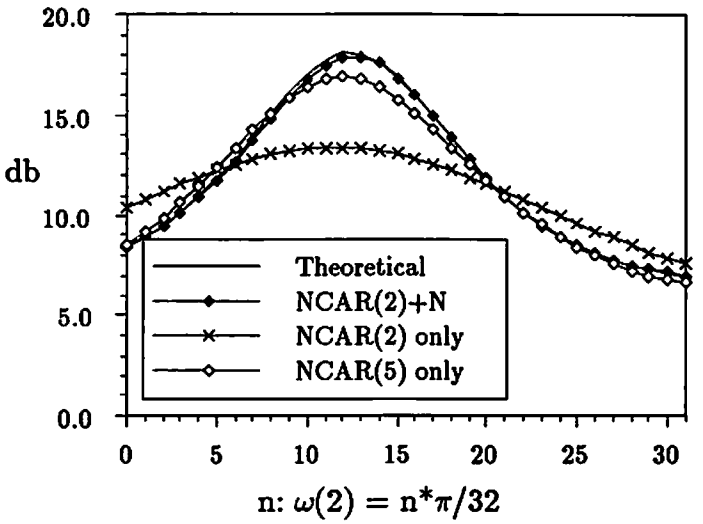


Figure 4.11: Row slices of 2-D spectra of 2nd-order NCAR signal in Gaussian noise with SNR = 0db.

Experiment 2: To obtain the performance statistics of signal plus noise spectrum estimation we generated ensembles of 32×32 arrays of Gaussian data. The first ensemble consisted of 64 arrays of toroidal noise free data $E_y = \{\mathcal{Y}^k = \{y_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64\}$, each array $\mathcal{Y}^k$ generated in the manner of [16] and according to a 2nd-order NCAR model having the parameters displayed in Table 4.1. To form two additional ensembles of signal plus noise data, E_{x1} and E_{x2}, we added two IID random noise fields having normal distributions with means of zero and variances $\sigma_n^2 = .16$ and $\sigma_n^2 = 1.0$, thus yielding signal plus noise data with SNRs, computed by (4.95), of approximately 20db and 5db, respectively. Using the toroidal ML procedure we computed parameter estimates from all sets of data in each ensemble for both 2nd-order NCAR only and 2nd-order NCAR plus noise models. The results of Experiment 1 indicated that 2nd-order was sufficient for estimating the spectra well with a NCAR plus noise model. The estimate average, average error (ε), and average squared-error (ς^2) statistics of the toroidal ML parameter estimates are shown in Tables 4.3, 4.4, and 4.5, respectively. The average error was computed by

$$\varepsilon_{\mathbf{r}} = \frac{1}{64} \sum_{k=1}^{64} (\hat{\theta}_{\mathbf{r}}^k - \theta_{\mathbf{r}}^0), \quad \mathbf{r} \in N_s \tag{4.96}$$

and the average squared-error was computed by

$$\varsigma_{\mathbf{r}}^2 = \frac{1}{64} \sum_{k=1}^{64} (\hat{\theta}_{\mathbf{r}}^k - \theta_{\mathbf{r}}^0)^2, \quad \mathbf{r} \in N_s, \tag{4.97}$$

where $\theta_{\mathbf{r}}^0, \mathbf{r} \in N_s$, are the true values of the coefficients from Table 4.1. Similar equations were used to compute the statistics for β.

Model	SNR	$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2	σ_n^2
true values	N/A	.19450	.05710	−.13600	.23470	1.0	0.0/0.16/1.0
NCAR only	no noise	.19466	.05615	−.13710	.23315	0.99233	N/A
NCAR+N	no noise	.19658	.05586	−.13801	.23448	0.95634	0.02409
NCAR+N	20db	.19296	.05420	−.13528	.23484	1.00667	0.41096
NCAR+N	5db	.19162	.04648	−.12978	.23607	1.07997	0.91227

Table 4.3: Average toroidal ML estimates of 2nd-order NCAR only and NCAR plus noise model parameters for data with various SNRs

Model	SNR	$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2	σ_n^2
NCAR only	no noise	.00016	−.00095	−.00110	−.00155	−.00767	N/A
NCAR+N	no noise	.00208	−.00124	−.00201	−.00022	−.04366	.02409
NCAR+N	20db	−.00154	−.00290	.00072	.00014	.00667	.25096
NCAR+N	5db	−.00288	−.01062	.00622	.00137	.07997	−.08773

Table 4.4: Average toroidal ML estimate errors of 2nd-order NCAR only and NCAR plus noise model parameters for data with various SNRs

Model	SNR	$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2	σ_n^2
NCAR only	no noise	.00013	.00019	.00012	.00014	.00282	N/A
NCAR+N	no noise	.00015	.00020	.00012	.00015	.00771	.00208
NCAR+N	20db	.00019	.00025	.00014	.00019	.01178	.10922
NCAR+N	5db	.00030	.00059	.00028	.00035	.03575	.02371

Table 4.5: Average toroidal ML estimate squared-errors of 2nd-order NCAR only and NCAR plus noise model parameters for data with various SNRs

Experiment 3: Using the data ensembles from Experiment 2, we evaluated the impact of the various types of initial conditions on determination of the toroidal ML function global minimum. Table 4.6 shows the average errors of the BCLS, LSCor, CompLS, and LS parameter estimates obtained from E_{x2}, the ensemble of 5db data. In all cases $\sigma_n^2 = 1.0$ was used as the initial estimate of the additive noise variance. Figure 4.12 compares the average errors of these initial parameter estimates. Note that the LS estimates are the same as the CompLS estimates except for β_{LS} which is significantly larger than the compensated value β_{CompLS}.

Clearly, the CompLS and the ordinary LS estimates have smaller errors than the BCLS and LS correlation estimates suggested by Woods, et. al., in [65]. One would suspect that the CompLS estimate should provide the best set of initial conditions. To test this notion we used the initial conditions from Table 4.3 as the starting values for the toroidal ML procedure. The conjugate gradient minimization routine available in the IMSL library [67] was used to minimize (4.14). Figures 4.13, 4.14, 4.15, and 4.16 plot the values of the toroidal ML function at the convergence of the gradient descent algorithm and the global minimum for each of the 64 arrays of 5db NCAR signal data. From these figures we note that LS initial conditions lead to the global minimum[1] in 64 of the 64 minimizations and CompLS led to the minimum in 63 of the 64 minimizations while BCLS and LSCor estimates did poorly, with the apparent global minimum being reached in only 24 of 64 and 40 of 64 cases, respectively.

[1]Theoretically, the global minimum cannot be determined with certainty. Thus, the apparent global minima were established by selecting the minimum value of the likelihood function from all the optimization runs for each set of data and then comparing the values of the parameters at these optimum points to the true known parameter values. If the estimates were very close to the true values, the minimum computed likelihood function value was taken to be the global minimum.

To further analyze this situation we repeated this last experiment with an initial value of $\sigma_n^2 = 1.2$. Initial conditions for the 5db data were then computed using this value of σ_n^2. Table 4.7 shows the initial condition values and Figure 4.17 plots the initial condition average errors. Figures 4.18 through 4.21 plot the values of the toroidal likelihood function at convergence and at the global minimum for each of the 64 arrays of 5db NCAR signal data. Here the global minimum was found in 63 of the 64 cases starting from the LS estimates and 61 of the 64 cases starting from CompLS estimates. On the other hand, the apparent global minimum was reached only 10 of 64 cases from BCLS estimates and 37 of 64 cases from LSCor estimates. It appears that even with a poor estimate of the additive noise variance LS and CompLS provide better initial conditions than bias-compensated or LSCor estimates. Moreover, the results suggest accurate coefficient estimates are more important than accurate β estimates for putting the toroidal ML function in the vicinity of the global minimum.

	$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2	σ_n^2
True Values	.19450	.05710	−.13600	.23470	1.0	1.0
Bias-Comp LS	.25406	.03320	−.16334	.31601	0.80196	1.0
LS Cor	.25103	.03199	−.13823	.30709	1.00920	1.0
Comp LS	.20930	.06068	−.14666	.27807	0.84650	1.0
LS	.20930	.06068	−.14666	.27807	2.14269	1.0

Table 4.6: Average initial conditions for SNR = 5db data computed using various estimators with high precision noise variance estimate.

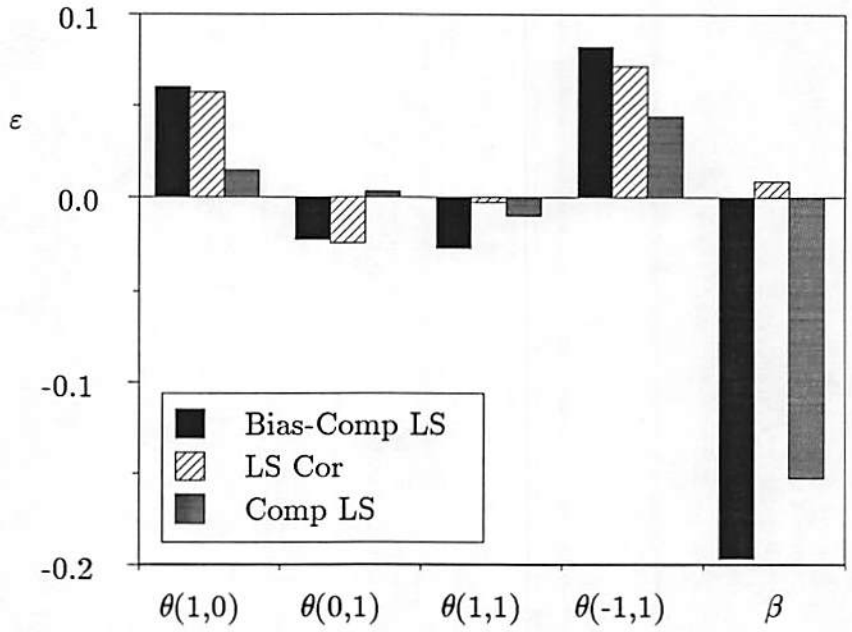


Figure 4.12: Comparison of average initial condition errors for SNR = 5db data computed using various estimators with high precision noise variance estimate.

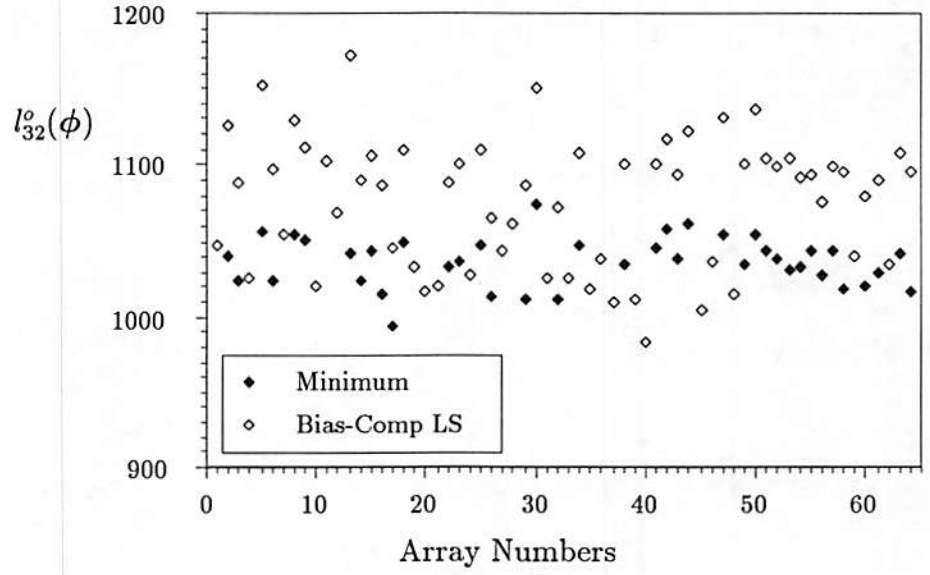


Figure 4.13: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from BCLS initial conditions with high precision noise variance estimate.

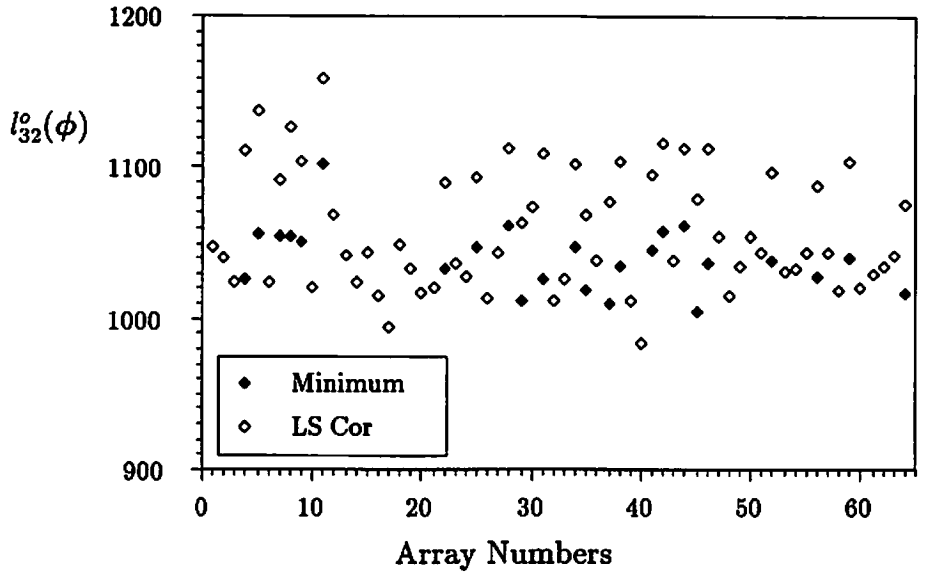


Figure 4.14: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from LSCor initial conditions with high precision noise variance estimate.

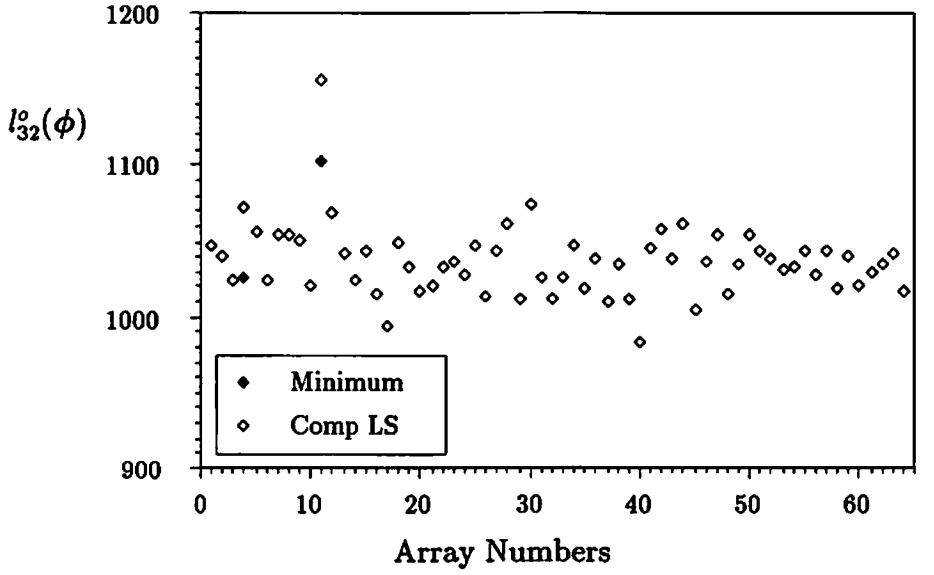


Figure 4.15: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from CompLS initial conditions with high precision noise variance estimate.

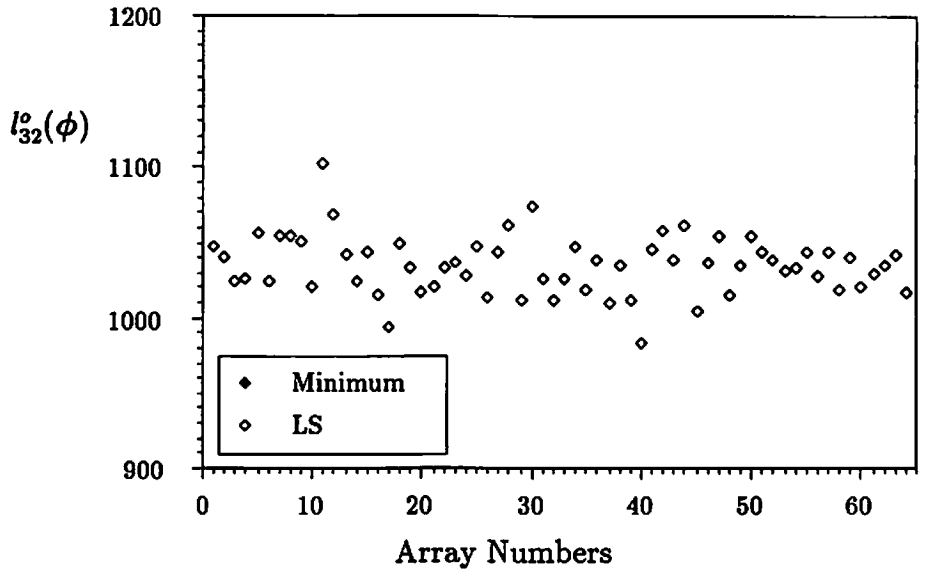


Figure 4.16: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from LS initial conditions with high precision noise variance estimate.

	$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2	σ_n^2
True Values	.19450	.05710	−.13600	.23470	1.0	1.0
Bias-Comp LS	.26851	.01715	−.16331	.32809	0.51878	1.2
LS Cor	.25103	.03199	−.13823	.30709	0.77216	1.2
Comp LS	.20930	.06068	−.14666	.27807	0.58726	1.2
LS	.20930	.06068	−.14666	.27807	2.14269	1.2

Table 4.7: Average initial conditions for SNR = 5db data computed using various estimators with poorly estimated noise variance.

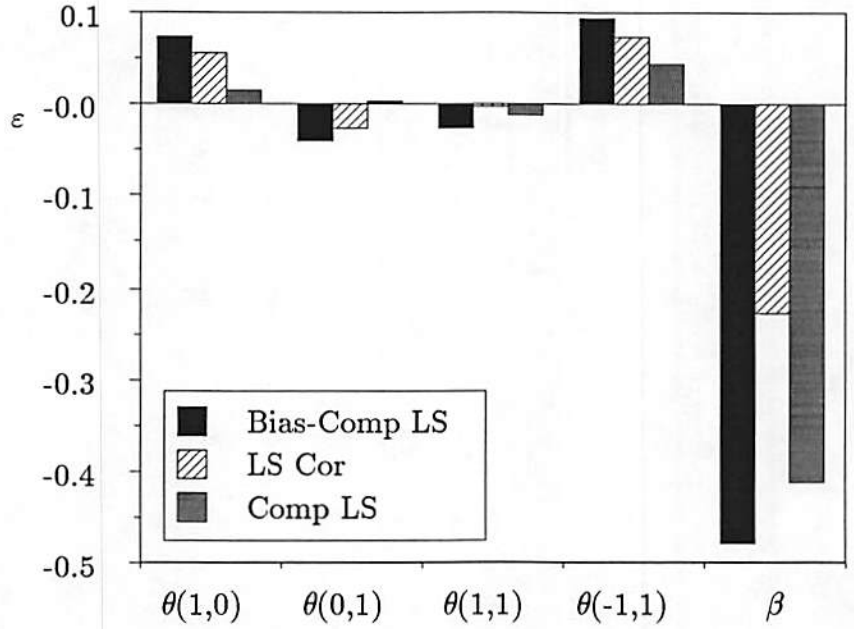


Figure 4.17: Comparison of average initial condition errors for SNR = 5db data computed using various estimators with poorly estimated noise variance.

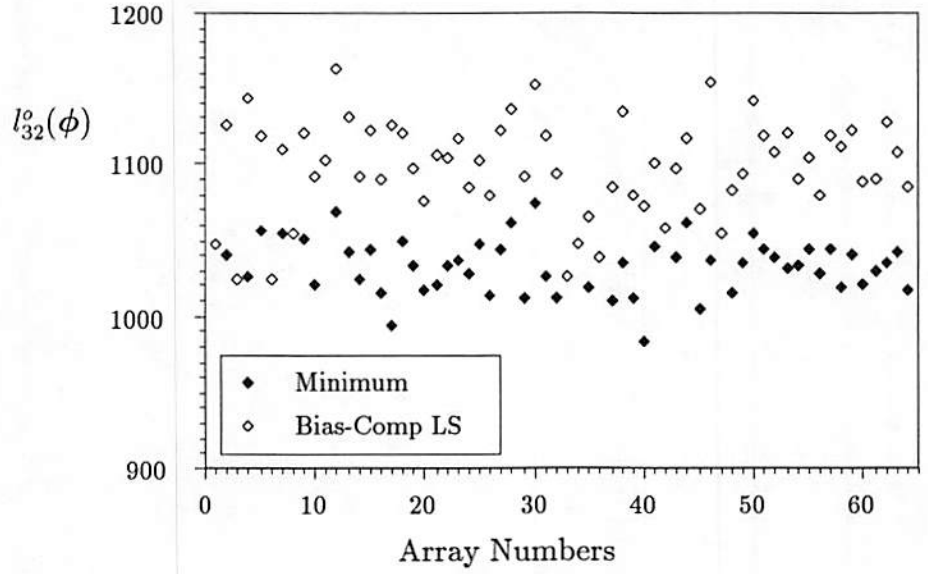


Figure 4.18: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from BCLS initial conditions with poorly estimated noise variance.

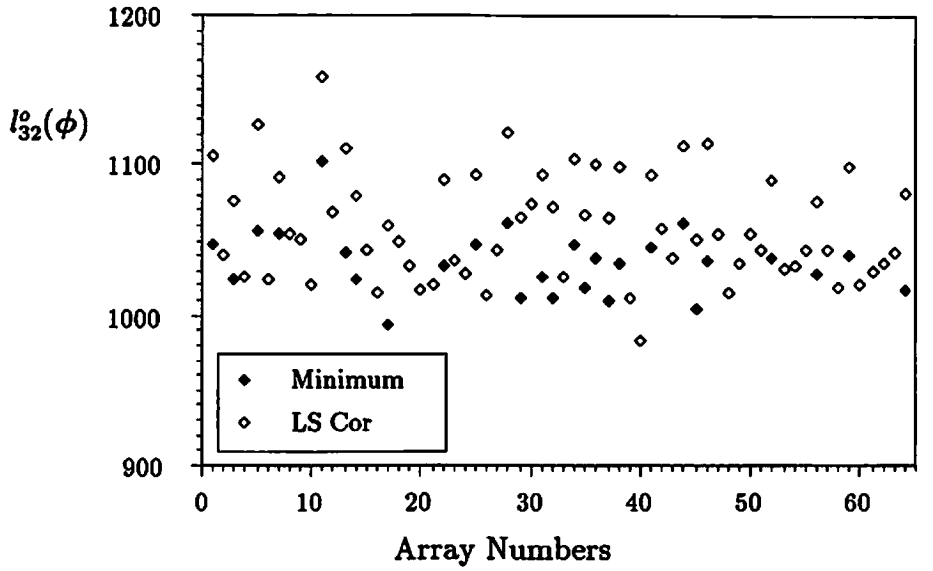


Figure 4.19: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from LSCor initial conditions and poorly estimated noise variance.

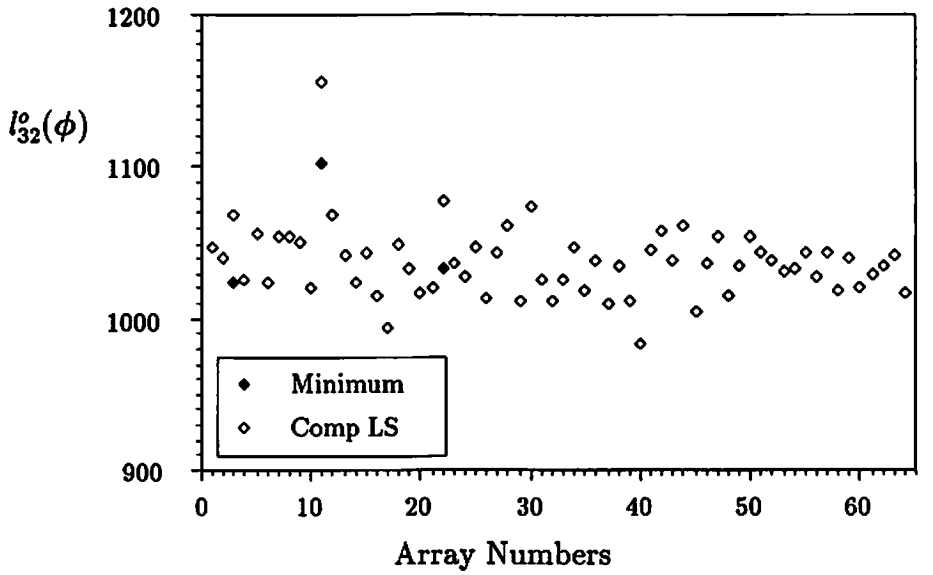


Figure 4.20: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from CompLS initial conditions and poorly estimated noise variance.

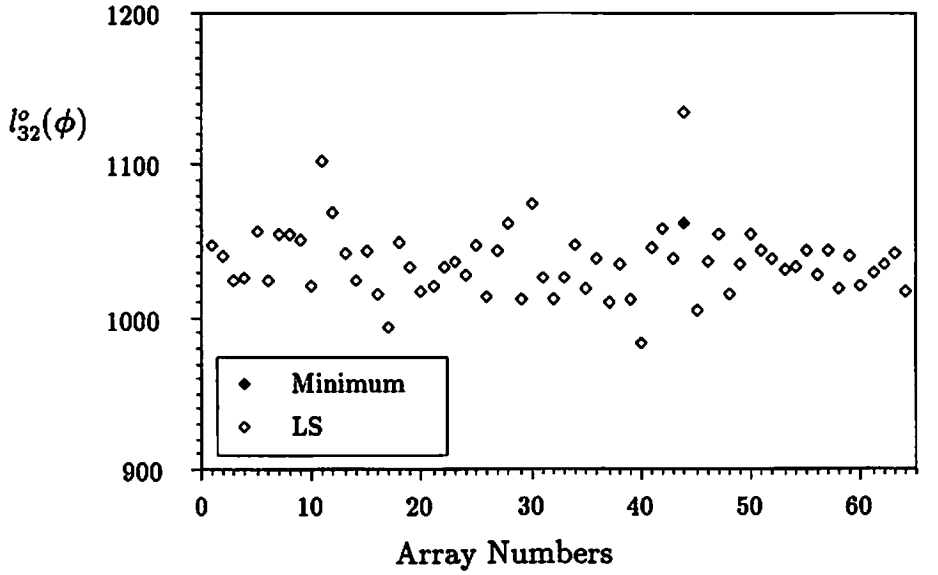


Figure 4.21: 2nd-Order NCAR signal plus noise model toroidal ML global minima and values computed from LS initial conditions and poorly estimated noise variance.

Experiment 4: This experiment was conducted to evaluate and compare the performance of the NCAR only and the NCAR plus noise models to detect and resolve 2-D sinusoids in additive Gaussian noise.

We computed spectrum estimates from simulated planar array data. To simulate this data we generated 64×64 arrays of sinusoidal data in additive IID Gaussian noise in accordance with (1.3) and (1.4) having parameter values $\varphi_1 = \frac{\pi}{32}[15, 17]^t$ and $\varphi_2 = \frac{\pi}{32}[19, 21]^t$ and various values of σ_n^2. Here, $\gamma^2 \neq \sigma_n^2$ and SNR is defined by

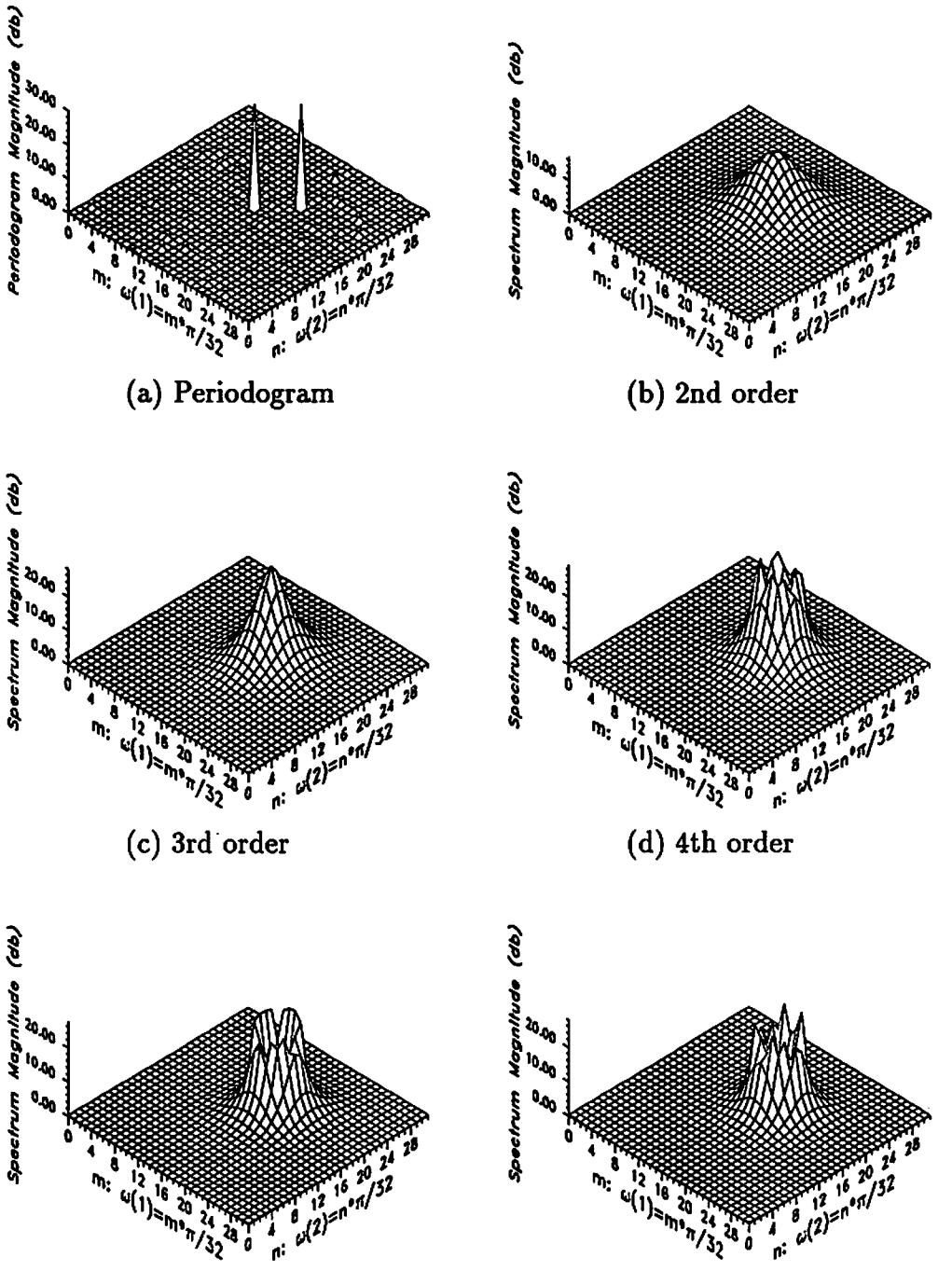
$$\text{SNR}_j = 10 \log \left(\frac{A_j^2}{2\sigma_n^2} \right), \quad j = 1, \ldots, L. \tag{4.98}$$

Spectra were estimated by using the toroidal ML procedure for both NCAR only and NCAR plus noise models of various orders. Least squares and BCLS estimates, with the initial estimate of γ^2 computed by selecting the minimum value of the periodogram averaged over a 7×7 window, were used as initial conditions for the minimization procedure. The results are shown in Figures 4.22 through 4.30. For example, Figure 4.22 shows the unsmoothed periodogram and the estimated spectra resulting from application of the toroidal ML procedure with 2nd through 6th-order NCAR only models when the SNR = 10db. Figure 4.23 is similar to Figure 4.22, but here NCAR plus noise models were used. Figures 4.24 through 4.30 are similar plots for SNRs ranging from 5db to -20db.

These estimates were computed by a conjugate gradient minimization routine, and in all cases stationary points were found, i.e., the likelihood function converged and at convergence the gradient of the likelihood function was approximately zero. Many of the NCAR only results were local minima; but local minima were easily avoided by the conjugate gradient algorithm when using the NCAR plus

noise model. In several cases, specifically, Figures 4.28(e), 4.29(c), and 4.30(d), the estimates of θ yielded exact poles of the spectrum, and we were unable to compute the spectrum. However, the poles were located correctly.

Additional experiments were conducted to evaluate performance with sinusoidal signals having closer spatial frequencies and differing amplitudes. Figure 4.31 is the estimated spectrum using a 2nd order NCAR plus noise model on data with $\varphi_1 = \frac{\pi}{32}[8,9]^t$ and $\varphi_2 = \frac{\pi}{32}[10,11]^t$ and SNR = 0db. Figure 4.32 shows the resulting estimated spectrum of data consisting of two unequal amplitude sinusoids, $\varphi_1 = \frac{\pi}{32}[15,17]^t$ and $\varphi_2 = \frac{\pi}{32}[19,21]^t$ with $\text{SNR}_1 = 10\text{db}$ and $\text{SNR}_2 = 0\text{db}$.

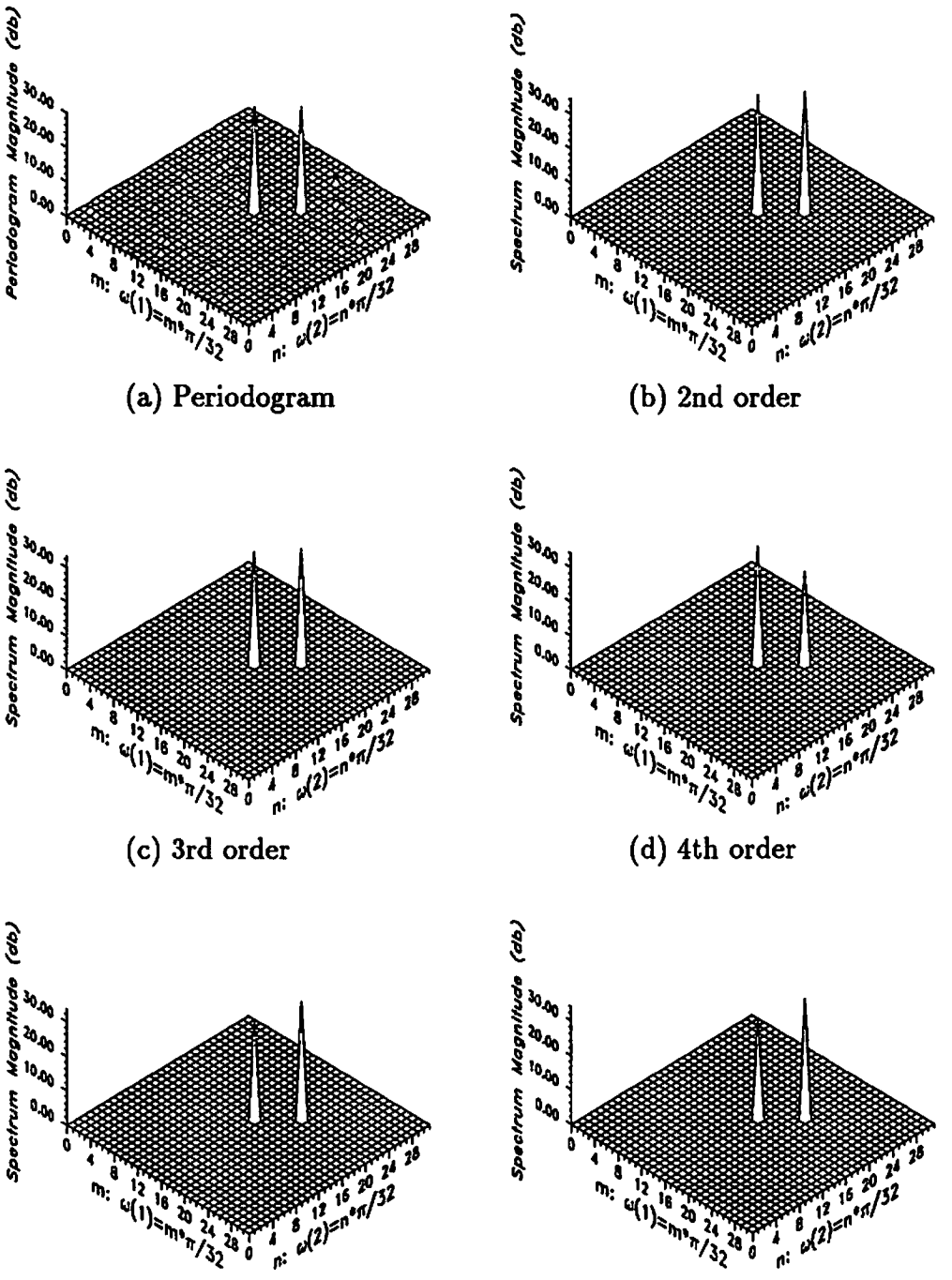


(a) Periodogram (b) 2nd order

(c) 3rd order (d) 4th order

(e) 5th order (f) 6th order

Figure 4.22: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = 10db and using NCAR only model.

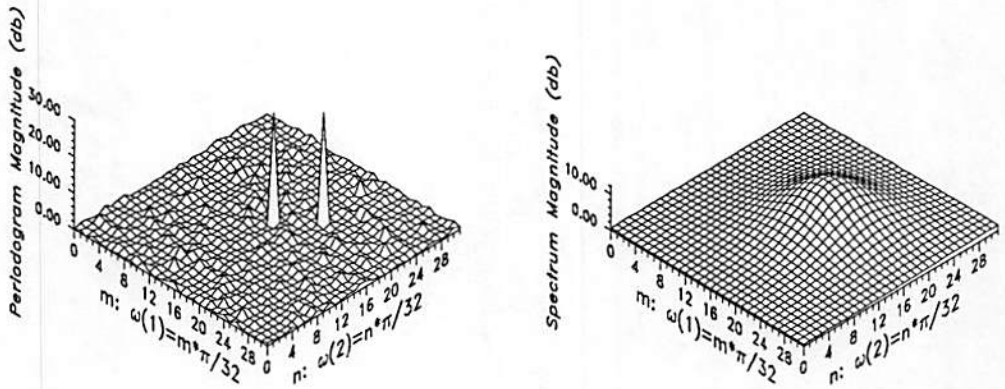


(a) Periodogram (b) 2nd order

(c) 3rd order (d) 4th order

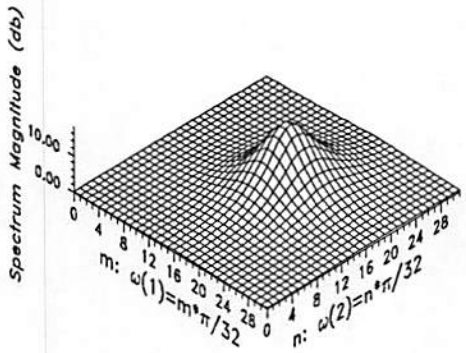
(e) 5th order (f) 6th order

Figure 4.23: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = 10db and using NCAR plus noise model.

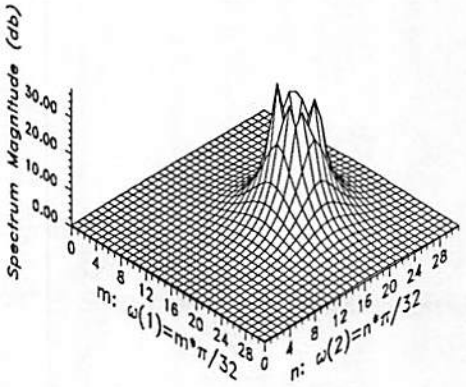


(a) Periodogram

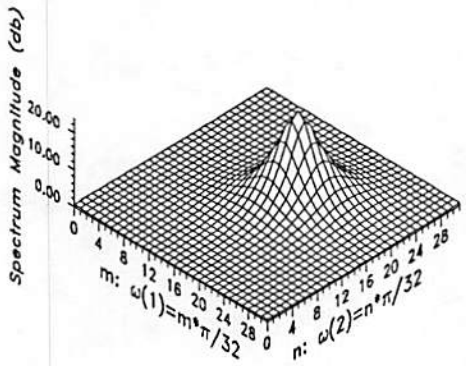
(b) 2nd order



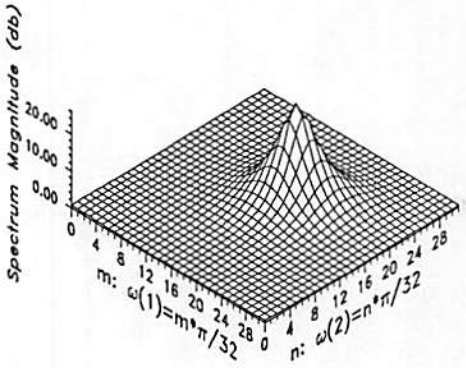
(c) 3rd order



(d) 4th order

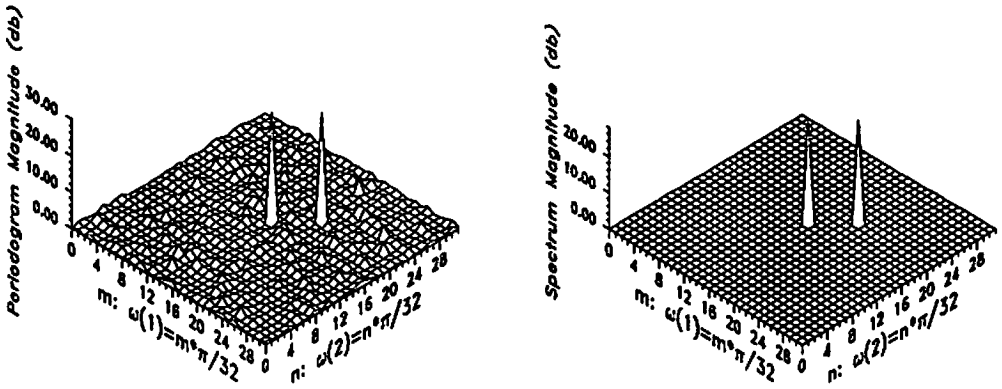


(e) 5th order



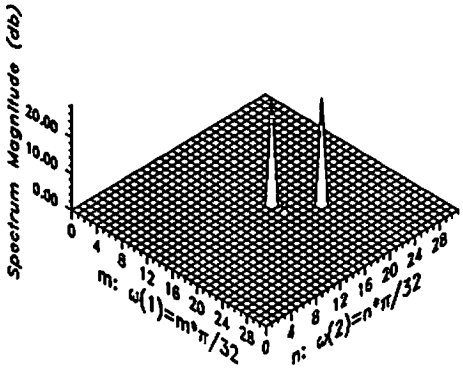
(f) 6th order

Figure 4.24: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = 5db and using NCAR only model.

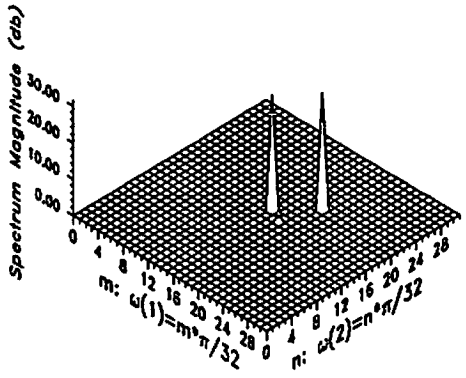


(a) Periodogram

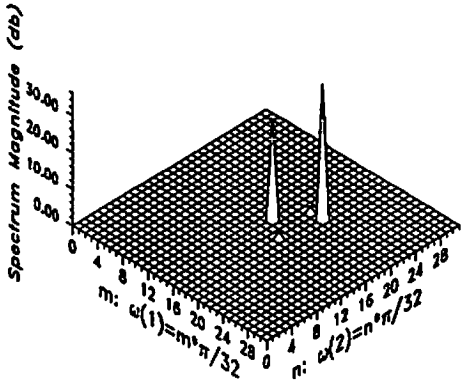
(b) 2nd order



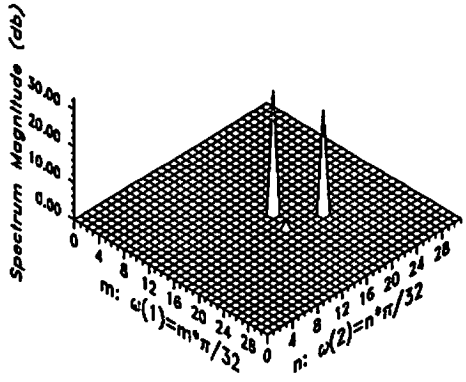
(c) 3rd order



(d) 4th order

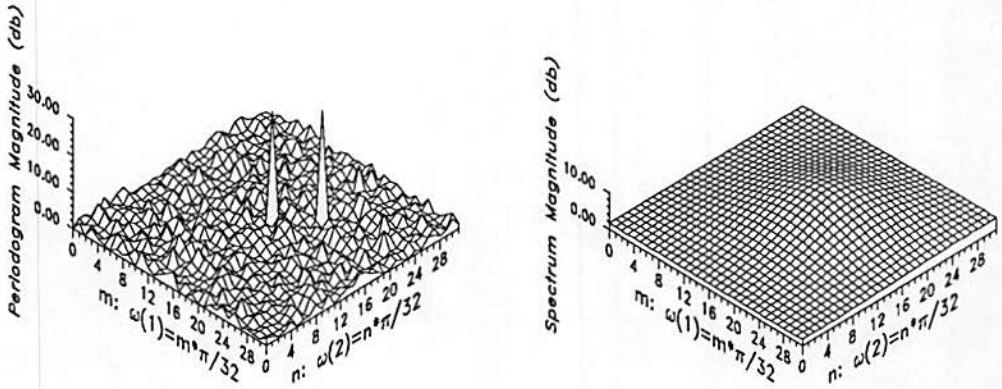


(e) 5th order



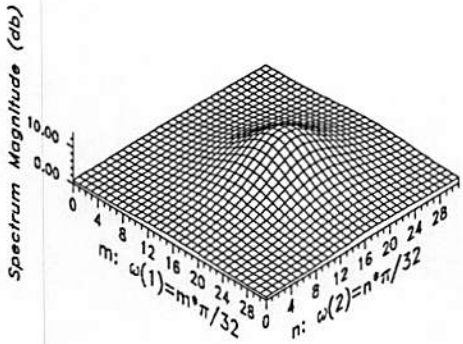
(f) 6th order

Figure 4.25: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = 5db and using NCAR plus noise model.

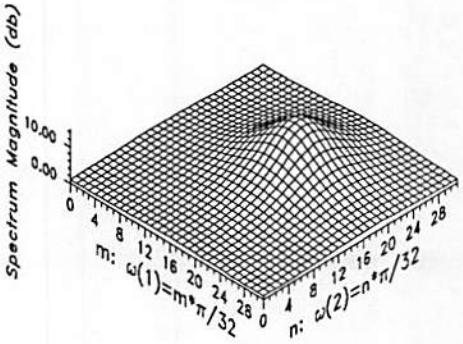


(a) Periodogram

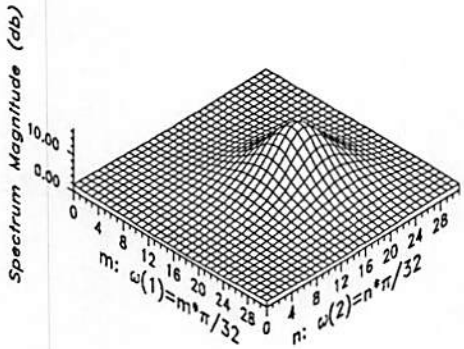
(b) 2nd order



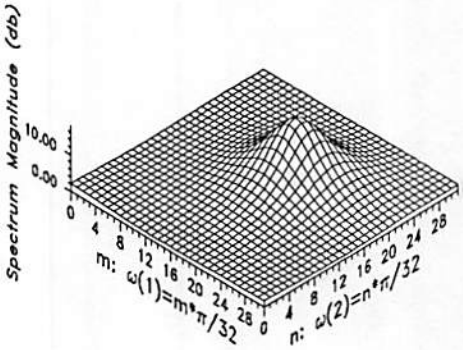
(c) 3rd order



(d) 4th order

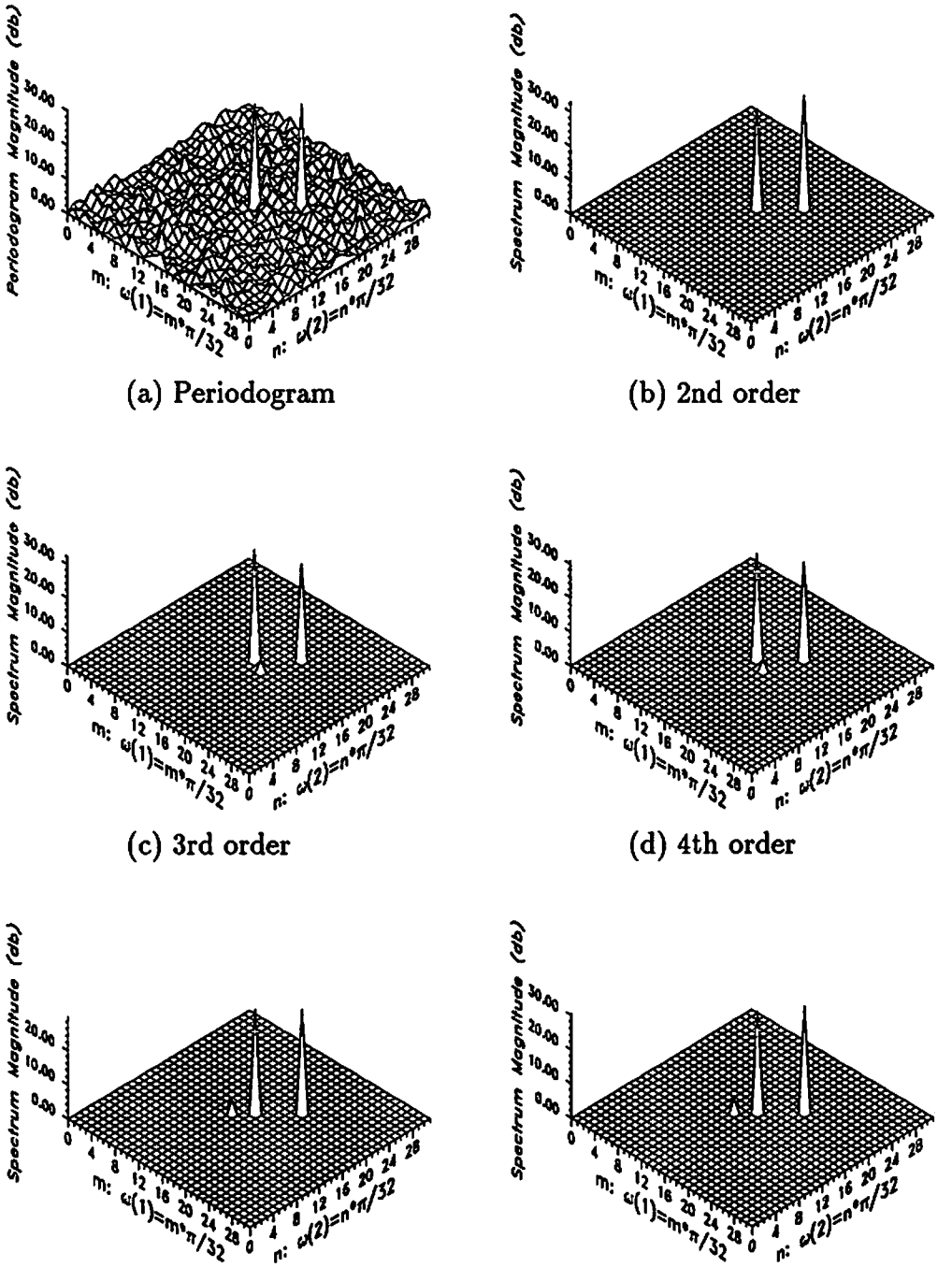


(e) 5th order



(f) 6th order

Figure 4.26: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = 0db and using NCAR only model.



(a) Periodogram

(b) 2nd order

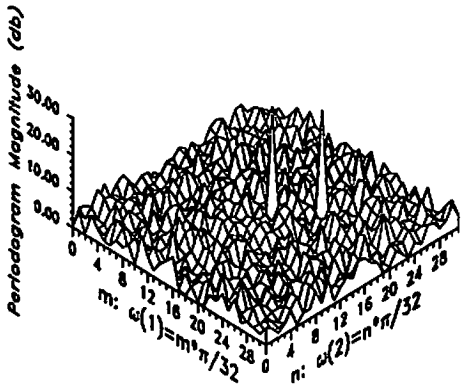
(c) 3rd order

(d) 4th order

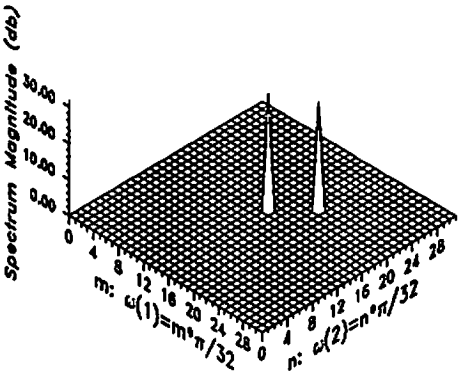
(e) 5th order

(f) 6th order

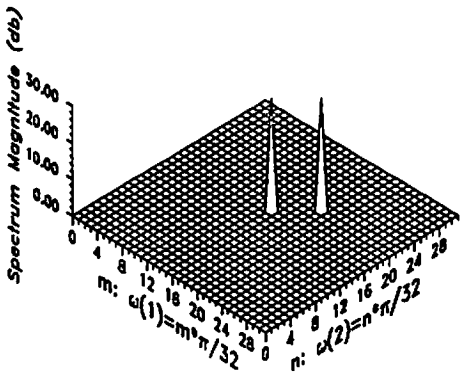
Figure 4.27: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = 0db and using NCAR plus noise model.



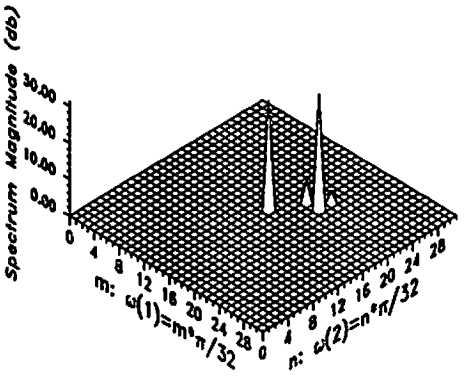
(a) Periodogram



(b) 2nd order

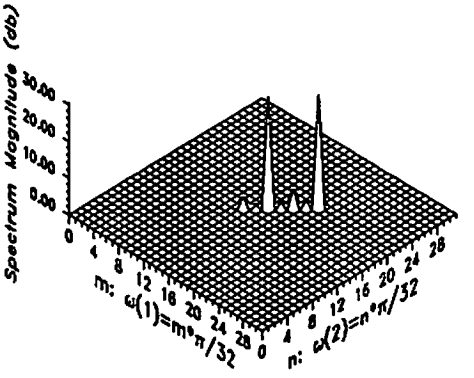


(c) 3rd order



(d) 4th order

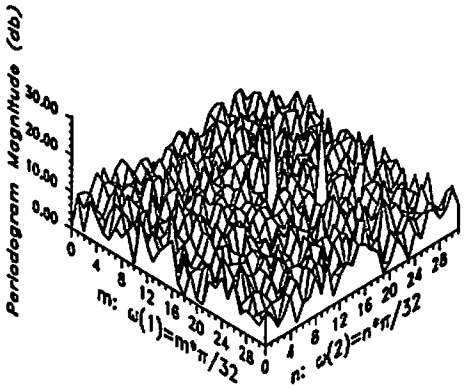
No plot available



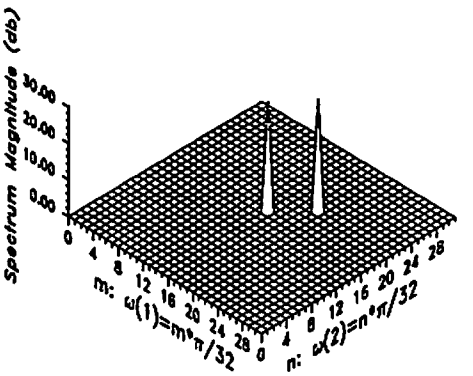
(e) 5th order

(f) 6th order

Figure 4.28: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = −5db and using NCAR plus noise model.



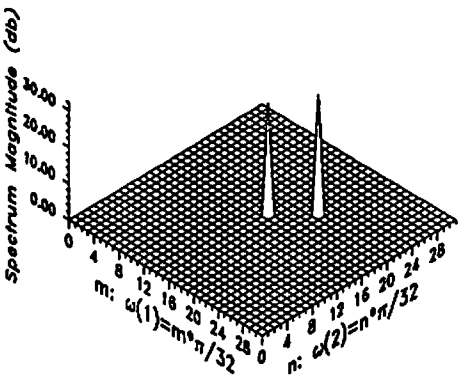
(a) Periodogram



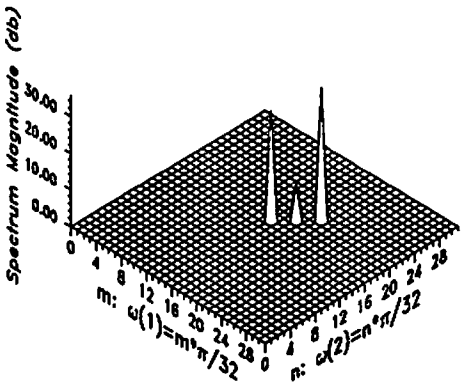
(b) 2nd order

No plot available

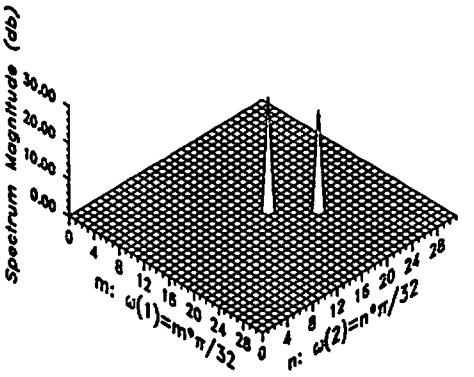
(c) 3rd order



(d) 4th order

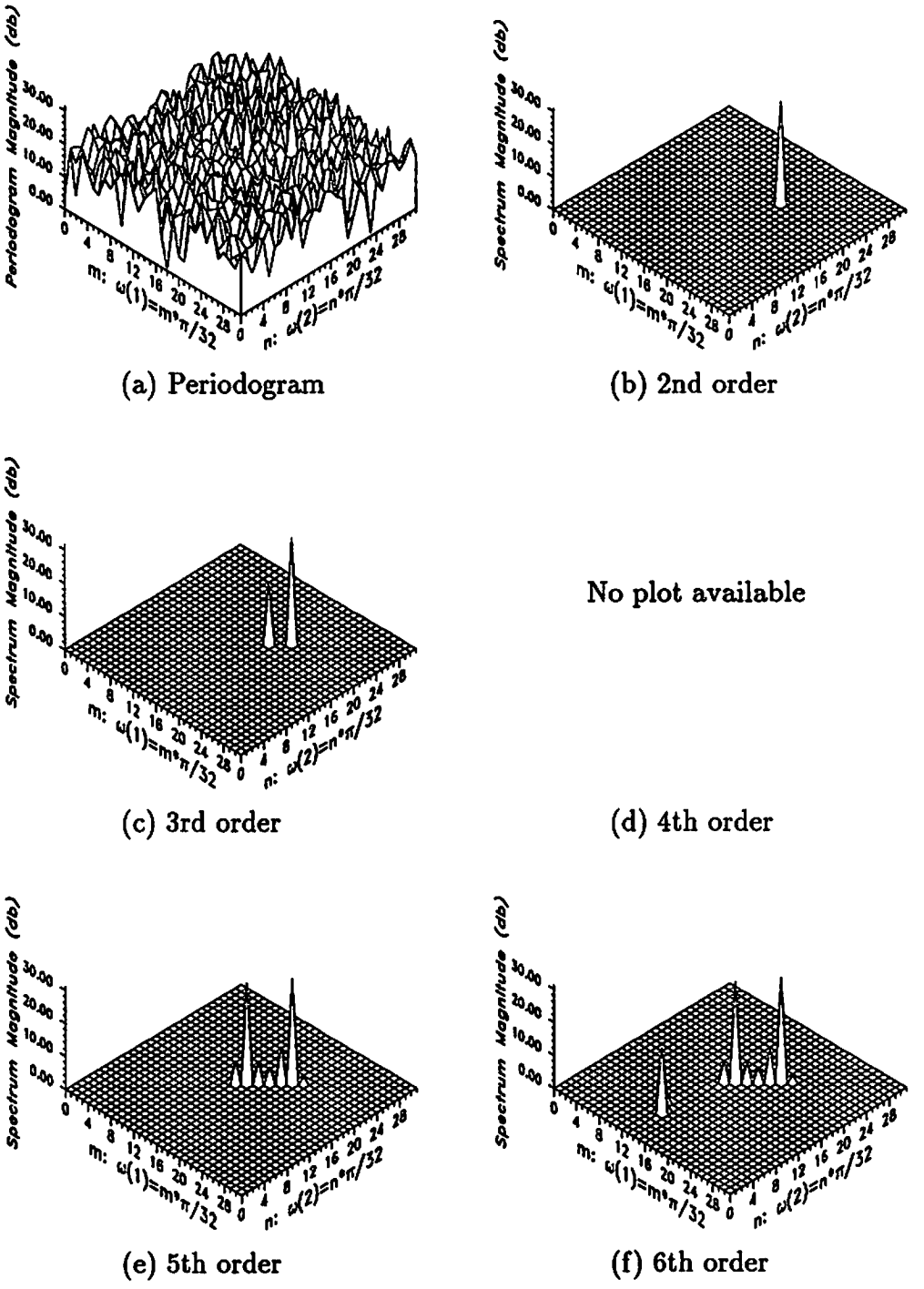


(e) 5th order



(f) 6th order

Figure 4.29: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = -10db and using NCAR plus noise model.



(a) Periodogram

(b) 2nd order

(c) 3rd order

(d) 4th order

(e) 5th order

(f) 6th order

Figure 4.30: Periodogram and estimated spectra for two sinusoids in Gaussian noise with SNR = −20db and using NCAR plus noise model.

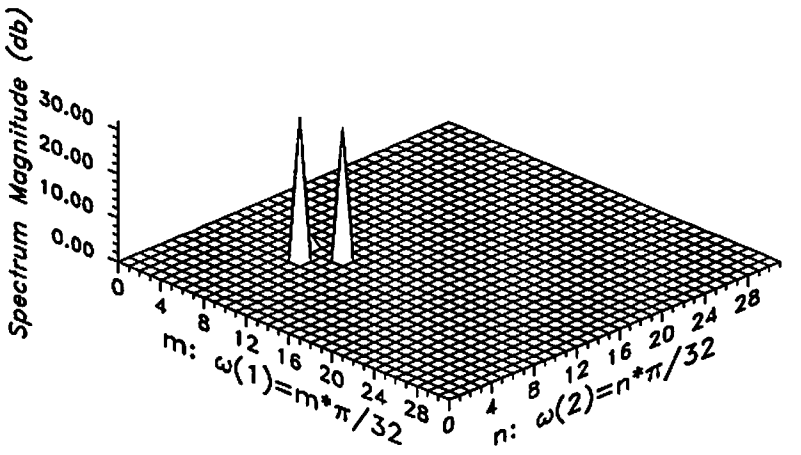


Figure 4.31: Estimated spectrum for two closely spaced sinusoids in Gaussian noise with SNR = 0db and using NCAR plus noise model.

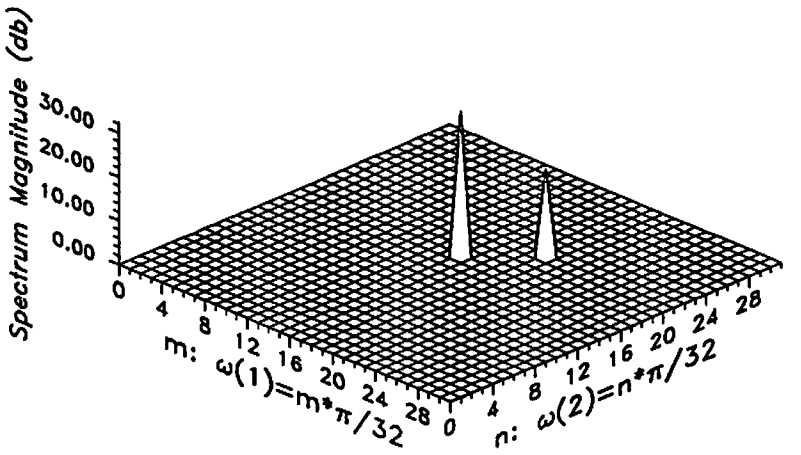


Figure 4.32: Estimated spectrum for two unequal amplitude sinusoids in Gaussian noise with SNR = 10db and SNR = 0db and using NCAR plus noise model.

4.5 Useful Lemmas

We prove the three lemmas used in the proofs of Theorems 4.1 through 4.4.

Lemma 4.1 *Let $A_x(\omega)$ and $B_x(\omega)$ be two strictly positive and continuous functions of ω. Futhermore, let A and B be the $M^2 \times M^2$ matrices whose i,j-th elements are $a_x(\mathrm{t}_i - \mathrm{t}_j)$ and $b_x(\mathrm{t}_i - \mathrm{t}_j)$, respectively, and where*

$$\begin{aligned} a_x(\mathrm{t}) &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)\exp(\jmath\omega^t\mathrm{t})d\omega \\ b_x(\mathrm{t}) &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} B_x(\omega)\exp(\jmath\omega^t\mathrm{t})d\omega. \end{aligned}$$

Then,

$$\lim_{M\to\infty}\frac{1}{M^2}\mathrm{tr}\{AB\} = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)B_x(\omega)d\omega.$$

Proof: The i,j-th element of the product of the matrices A and B is given by

$$[AB]_{ij} = \sum_{\mathrm{t}\in\Omega_M} a_x(\mathrm{t}_i - \mathrm{t})b_x(\mathrm{t} - \mathrm{t}_j)$$

so the trace is

$$\mathrm{tr}\{AB\} = \sum_{\mathrm{s}\in\Omega_M}\sum_{\mathrm{t}\in\Omega_M} a_x(\mathrm{s} - \mathrm{t})b_x(\mathrm{t} - \mathrm{s}).$$

Consequently, we have

$$\begin{aligned} &\lim_{M\to\infty}\frac{1}{M^2}\mathrm{tr}\{AB\} \\ &\quad= \lim_{M\to\infty}\frac{1}{M^2}\sum_{\mathrm{s}\in\Omega_M}\sum_{\mathrm{t}\in\Omega_M} a_x(\mathrm{s} - \mathrm{t})b_x(\mathrm{t} - \mathrm{s}) \\ &\quad= \lim_{M\to\infty}\frac{1}{M^2}\sum_{\mathrm{s}\in\Omega_M}\sum_{\mathrm{t}\in\Omega_M}\frac{1}{(2\pi)^4}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)B_x(\tau) \\ &\qquad\cdot\exp[\jmath\omega^t(\mathrm{s} - \mathrm{t})]\exp[\jmath\tau^t(\mathrm{t} - \mathrm{s})]d\omega d\tau \\ &\quad= \lim_{M\to\infty}\frac{1}{M^2}\sum_{\mathrm{t}\in\Omega_M}\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)B_x(\tau) \end{aligned}$$

$$\cdot \exp[\jmath(\omega-\tau)^t \mathrm{t}] \frac{1}{(2\pi)^2} \sum_{\mathrm{s}\in\Omega_M} \exp[-\jmath(\omega-\tau)^t \mathrm{t}] d\tau d\omega$$

$$= \lim_{M\to\infty} \frac{1}{M^2} \sum_{\mathrm{t}\in\Omega_M} \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)B_x(\tau) \cdot \exp[\jmath(\omega-\tau)^t \mathrm{t}]\delta(\omega-\tau) d\tau d\omega$$

$$= \lim_{M\to\infty} \frac{1}{M^2} \sum_{\mathrm{t}\in\Omega_M} \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)B_x(\omega) d\omega$$

$$= \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A_x(\omega)B_x(\omega) d\omega.$$

□

Lemma 4.2 *Let* $\mathcal{S}_x(\omega) \geq k > 0$ *and define*

$$q_x(\mathrm{t}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega) \exp(\jmath\omega^t \mathrm{t}) d\omega \tag{4.99}$$

and

$$a_x(\mathrm{t}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x^{-1}(\omega) \exp(\jmath\omega^t \mathrm{t}) d\omega, \tag{4.100}$$

then $\sum_{\mathrm{t}\in\Omega_\infty} |q_x(\mathrm{t})|^2|\mathrm{t}|^2 < \infty$ *implies* $\sum_{\mathrm{t}\in\Omega_\infty} |a_x(\mathrm{t})|^2|\mathrm{t}|^2 < \infty$.

Proof: The inverse Fourier transform gives

$$\mathcal{S}_x(\omega) = \sum_{\mathrm{t}\in\Omega_\infty} q_x(\mathrm{t}) \exp(-\jmath\omega^t \mathrm{t}). \tag{4.101}$$

Differentiating (4.101) with respect to ω,

$$\frac{d}{d\omega}\mathcal{S}_x(\omega) = \sum_{\mathrm{t}\in\Omega_\infty} -\jmath \mathrm{t} q_x(\mathrm{t}) \exp(-\jmath\omega^t \mathrm{t}) \tag{4.102}$$

Hence, $\jmath\frac{d}{d\omega}\mathcal{S}_x(\omega)$ and $\mathrm{t}q_x(\mathrm{t})$ are a Fourier transform pair, and by Parseval's theorem

$$\frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \left|\frac{d}{d\omega}\mathcal{S}_x(\omega)\right|^2 d\omega = \sum_{\mathrm{t}\in\Omega_\infty} |q_x(\mathrm{t})|^2|\mathrm{t}|^2. \tag{4.103}$$

Similarly, since

$$\mathcal{S}_x^{-1}(\omega) = \sum_{\mathrm{t}\in\Omega_\infty} a_x(\mathrm{t}) \exp(-\jmath\omega^{\mathrm{t}}\mathrm{t}), \tag{4.104}$$

we have that

$$\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\frac{d}{d\omega}\mathcal{S}_x^{-1}(\omega)\right|^2 d\omega = \sum_{\mathrm{t}\in\Omega_\infty}|a_x(\mathrm{t})|^2|\mathrm{t}|^2. \tag{4.105}$$

But

$$\left|\frac{d}{d\omega}\mathcal{S}_x^{-1}(\omega)\right|^2 = \frac{1}{\mathcal{S}_x^4(\omega)}\left|\frac{d}{d\omega}\mathcal{S}_x(\omega)\right|^2, \tag{4.106}$$

and consequently,

$$\begin{aligned}
\sum_{\mathrm{t}\in\Omega_\infty}|a_x(\mathrm{t})|^2|\mathrm{t}|^2 &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\frac{d}{d\omega}\mathcal{S}_x^{-1}(\omega)\right|^2 d\omega \\
&= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\frac{1}{\mathcal{S}_x^4(\omega)}\left|\frac{d}{d\omega}\mathcal{S}_x(\omega)\right|^2 d\omega \\
&\leq \frac{1}{k^4}\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left|\frac{d}{d\omega}\mathcal{S}_x(\omega)\right|^2 d\omega \\
&= \frac{1}{k^4}\sum_{\mathrm{t}\in\Omega_\infty}|q_x(\mathrm{t})|^2|\mathrm{t}|^2 < \infty.
\end{aligned} \tag{4.107}$$

□

Lemma 4.3 *Let* $\sum_{\mathrm{t}\in\Omega_\infty}|q_x(\mathrm{t})|^2|\mathrm{t}|^2 < \infty$ *and* $\sum_{\mathrm{t}\in\Omega_\infty}|a_x(\mathrm{t})|^2|\mathrm{t}|^2 < \infty$, *where* $q_x(\mathrm{t})$ *and* $a_x(\mathrm{t})$ *are defined as in Lemma 4.2. Then*

$$|\mathrm{tr}\left\{I_{M^2} - Q_x A_x\right\}| < \infty \tag{4.108}$$

and

$$\mathrm{tr}\left\{\left[I_{M^2} - Q_x A_x\right]^2\right\} < \infty. \tag{4.109}$$

Proof: On the first hand, the (i,j)-th entry of I_{M^2} is $\delta(\mathrm{t}_i - \mathrm{t}_j) = \sum_{\mathrm{t}\in\Omega_\infty} q_x(\mathrm{t} - \mathrm{t}_i)a_x(\mathrm{t} - \mathrm{t}_j)$. This is shown as follows:

$$\begin{aligned}
&\sum_{\mathrm{t}\in\Omega_\infty} q_x(\mathrm{t}-\mathrm{t}_i)a_x(\mathrm{t}-\mathrm{t}_j) \\
&\quad= \sum_{\mathrm{t}\in\Omega_\infty} \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega)\exp[\jmath\omega^t(\mathrm{t}-\mathrm{t}_i)]d\omega \\
&\qquad\cdot\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x^{-1}(\tau)\exp[\jmath\tau^t(\mathrm{t}-\mathrm{t}_j)]d\tau \\
&\quad= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega)\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x^{-1}(\tau)\exp[-\jmath(\omega^t\mathrm{t}_i+\tau^t\mathrm{t}_j)] \\
&\qquad\cdot\left\{\frac{1}{(2\pi)^2}\sum_{\mathrm{t}\in\Omega_\infty}\exp[\jmath(\omega+\tau)^t\mathrm{t}]\right\}d\tau d\omega \\
&\quad= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega)\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x^{-1}(\tau)\exp[-\jmath(\omega^t\mathrm{t}_i+\tau^t\mathrm{t}_j)] \\
&\qquad\cdot\delta(\omega+\tau)d\tau d\omega \\
&\quad= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\exp[-\jmath\omega^t(\mathrm{t}_i+\mathrm{t}_j)]d\omega \\
&\quad= \delta(\mathrm{t}_i-\mathrm{t}_j). \qquad (4.110)
\end{aligned}$$

On the second hand, the i,j-th element of Q_xA_x is

$$[Q_xA_x]_{ij} = \sum_{\mathrm{t}\in\Omega_M} q_x(\mathrm{t}_i-\mathrm{t})a_x(\mathrm{t}-\mathrm{t}_j) \qquad (4.111)$$

Combining (4.110) and (4.111) we have the i,j-th element of $I_{M^2} - Q_xA_x$ as

$$[I_{M^2} - Q_xA_x]_{ij} = \sum_{\mathrm{t}\in\Omega_\infty} q_x(\mathrm{t}_i-\mathrm{t})a_x(\mathrm{t}-\mathrm{t}_j) - \sum_{\mathrm{t}\in\Omega_M} q_x(\mathrm{t}_i-\mathrm{t})a_x(\mathrm{t}-\mathrm{t}_j), \quad (4.112)$$

and since $q_x(\mathrm{t}) = q_x(-\mathrm{t})$ and $a_x(\mathrm{t}) = a_x(-\mathrm{t})$, (4.112) can be written as

$$\begin{aligned}
[I_{M^2} - Q_xA_x]_{ij} &= \sum_{\mathrm{t}\in\Omega_\infty} q_x(\mathrm{t}-\mathrm{t}_i)a_x(\mathrm{t}-\mathrm{t}_j) - \sum_{\mathrm{t}\in\Omega_M} q_x(\mathrm{t}_i-\mathrm{t})a_x(\mathrm{t}-\mathrm{t}_j) \\
&= \sum_{\mathrm{t}\notin\Omega_M} q_x(\mathrm{t}-\mathrm{t}_i)a_x(\mathrm{t}-\mathrm{t}_j). \qquad (4.113)
\end{aligned}$$

The trace can now be written as

$$\mathrm{tr}\,\{I_{M^2} - Q_xA_x\}$$

$$
\begin{aligned}
&= \sum_{s\in\Omega_M}\sum_{t\notin\Omega_M} q_x(t-s)a_x(t-s) \\
&= \sum_{s\in\Omega_M}\sum_{t\in\Omega_\infty} q_x(t-s)a_x(t-s) \\
&\qquad - \sum_{s\in\Omega_M}\sum_{t\in\Omega_M} q_x(t-s)a_x(t-s) \\
&= \sum_{s\in\Omega_M}\sum_{t\in\Omega_\infty} q_x(t)a_x(t) - \sum_{s\in\Omega_M}\sum_{t\in\Omega_M} q_x(t-s)a_x(t-s) \\
&= M^2 \sum_{t\in\Omega_\infty} q_x(t)a_x(t) \\
&\qquad - \sum_{k=0}^{M-1}\sum_{l=0}^{M-1}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1} q_x(i-k,j-l)a_x(i-k,j-l) \\
&= M^2 \sum_{t\in\Omega_\infty} q_x(t)a_x(t) \\
&\qquad - \sum_{t=(k,l)\in\Omega_M} (M-|k|)(M-|l|)q_x(t)a_x(t) \\
&= \sum_{t=k,l\in\Omega_\infty} \min\{M^2,(|k|+|l|)M-|kl|\}q_x(t)a_x(t) \qquad (4.114)
\end{aligned}
$$

A simple application of Schwarz's inequality yields

$$
\begin{aligned}
|\mathrm{tr}\{I_{M^2} - Q_xA_x\}|^2 &= \left|\sum_{t=k,l\in\Omega_\infty} \min\{M^2,(|k|+|l|)M-|kl|\}q_x(t)a_x(t)\right|^2 \\
&\le \sum_{t=k,l\in\Omega_\infty} \min\{M^2,(|k|+|l|)M-|kl|\}|q_x(t)|^2 \\
&\qquad \cdot \sum_{t=k,l\in\Omega_\infty} \min\{M^2,(|k|+|l|)M-|kl|\}|a_x(t)|^2 \\
&\le \sum_{t\in\Omega_\infty} |q_x(t)|^2|t|^2 \cdot \sum_{t\in\Omega_\infty} |a_x(t)|^2|t|^2 \\
&< \infty. \qquad (4.115)
\end{aligned}
$$

The second part of the lemma follows from the fact

$$
\mathrm{tr}\left\{A^2\right\} = \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2
$$

for any symmetric matrix A. Therefore, from (4.113), we have

$$
\mathrm{tr}\left\{[I_{M^2} - Q_xA_x]^2\right\} = \sum_{s\in\Omega_M}\sum_{r\in\Omega_M}\left|\sum_{t\notin\Omega_M} q_x(t-s)a_x(t-r)\right|^2 \qquad (4.116)
$$

and, applying the Schwarz inequality and (4.114),

$$\begin{aligned}
\mathrm{tr}\left\{[I_{M^2} - Q_x A_x]^2\right\} &\leq \sum_{s\in\Omega_M}\sum_{r\in\Omega_M}\left[\sum_{t\notin\Omega_M}|q_x(t-s)|^2 \cdot \sum_{t\notin\Omega_M}|a_x(t-r)|^2\right] \\
&\leq \sum_{s\in\Omega_M}\sum_{t\notin\Omega_M}|q_x(t-s)|^2 \cdot \sum_{r\in\Omega_M}\sum_{t\notin\Omega_M}|a_x(t-r)|^2 \\
&\leq \sum_{s\in\Omega_M}\sum_{t\notin\Omega_M}|q_x(t)|^2|t|^2 \cdot \sum_{s\in\Omega_M}\sum_{t\notin\Omega_M}|a_x(t)|^2|t|^2 \\
&< \infty. \qquad (4.117)
\end{aligned}$$

□

Chapter 5

2-D Generalized M-Estimators

In many situations contamination in observed data is not Gaussian or its distribution is unknown. This situation can occur when measurement errors or isolated errors cause observed data sets to contain small fractions of unusual data points, or "outliers", which are not consistent with a strictly Gaussian assumption. When the "outliers" have a large scale relative to the aggregate the observed data can be modeled in principle with a distribution that is nearly Gaussian in the central region but having heavier tails. In other situations the rounding or grouping caused by finite bit quantization and computation of signals can also be viewed as signal measurement error. Then the observed data is distributed as though it were Gaussian near the mean but having no tails at all. The signal plus noise model considered in Chapter 4, where strict distributions were assumed, may not perform well in this instance. In this chapter we propose a heuristic parameter estimator, the 2-D generalized maximum likelihood (GM) estimator, for NSHP and GMRF models. No strict assumption on the distribution of the observed data is made.

Martin and Thomson [19] have suggested that isolated measurement errors or outliers in time series be modeled with the mixture distribution

$$\mathcal{F} = (1-h)\Delta(0) + h\mathcal{N}(0,\sigma^2), \tag{5.1}$$

where $0 \leq h \leq 1$. Here $\Delta(0)$ is the degenerate distribution having all its mass at zero and $\mathcal{N}(0,\sigma^2)$ is the standard normal distribution with mean zero and variance σ^2. Let $y(\mathbf{s})$ represent the signal of interest. The 2-D innovative outlier (IO) model will be said to hold whenever, $x(\mathbf{s}) = y(\mathbf{s})$ for all Ω_M and the innovations process deviates from a nominal Gaussian distribution. For example, the innovations may have a heavy-tailed non-Gaussian common distribution which results from the sum of a normal random variable and a random variable distributed according to (5.1). The additive outlier (AO) model will occur if $x(\mathbf{s}) = y(\mathbf{s}) + n(\mathbf{s})$ and $n(\mathbf{s})$ has the distribution given by (5.1). Then the signal is observed correctly most of the time, i.e., $x(\mathbf{s}) = y(\mathbf{s})$, but $100\,h$ percent of the time $y(\mathbf{s})$ is observed with error. We note that this contamination leads to a non-Gaussian heavy-tailed probability distribution $\mathcal{F}(x)$, although for small h, $\mathcal{F}(x)$ will be nearly $\mathcal{F}(y) = \mathcal{N}(0,\sigma_y^2)$.

Even when h is small, say $h \leq .10$, the outliers may have a detrimental effect on parameter estimates; and consequently, the model based spectrum estimate. In this situation the optimally designed estimation procedure based on the assumption of strict normality is not fully efficient. For example, the parametric methods of 2-D spectrum estimation described in [31] and [12] are vulnerable to even a few outliers, as are their counterparts in 1-D [19]. Therefore, a procedure is required whose performance remains quite good for a broad class of underlying distributions (in the neighborhood of the Gaussian distribution) but which may not necessarily be best for any of them. Such procedures are called robust.

5.1 2-D Robust Parameter Estimation

Robust spectrum estimation for time series has been suggested in [40, 41, 19]. Since LS estimates are consistent for NSHP and GMRF models, the arguments used in the 1-D case can be followed and a robustified least squares problem defined. Suppose the signal is a NSHP Gaussian random field and modeled by (3.1) of Chapter 3 with a neighbor set $N \subset \Omega^+$. Then robust parameter estimates $\hat{\theta}_{GM}$ of $\theta = \text{col}\{\theta_r,\ r \in N\}$ and S_{GM} of β are computed by solving

$$\sum_{s \in \Omega_M} x_s W(x_s) \psi \left(\frac{x(s) - \hat{\theta}_{GM}^t x_s}{c_v S_{GM}} \right) = 0 \tag{5.2}$$

and

$$\sum_{s \in \Omega_M} W(x_s) \left[c_v^2 \psi^2 \left(\frac{x(s) - \hat{\theta}_{GM}^t x_s}{c_v S_{GM}} \right) - B \right] = 0\,, \tag{5.3}$$

where the past history vector is defined by $x_s = \text{col}\{x(s+r),\ r \in N\}$. The estimators (5.2) and (5.3) are known as the 2-D generalized maximum-likelihood (GM) estimators. The equations to be solved are identical for a GMRF model, (3.6) of Chapter 3 with a possibly noncausal neighbor set N, but then S_{GM} is an estimate of $\sqrt{\nu}$. Since both models will be treated identically in the following development we shall denote the scale for either the NSHP or GMRF innovations process by S, i.e., $S = \beta$ for the NSHP model or $S = \sqrt{\nu}$ for the GMRF model.

In (5.2) and (5.3) c_v and B are tuning constants selected to adjust robustness and yield consistent estimates when the observed data $\mathcal{X}_M = \{x(s), s \in \Omega_M\}$ are normally distributed. The function $\psi(\cdot)$ is to limit the influence of those summands of (5.2) and (5.3) for which $v(s) = x(s) - \theta^t x_s$ is a poor estimate of the residual and $W(\cdot)$ is a weight function to down-weight those summands with outliers in the components of x_s. The choice of $\psi(\cdot)$ and $W(\cdot)$ functions with good robustness properties will be discussed in subsequent sections.

Equations (5.2) and (5.3) can be solved with an iterated weighted least squares procedure as suggested by Huber [39] and the robust estimates $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{GM}$ and $\beta = S_{GM}$ then used to compute the estimated spectrum from the model's theoretical spectral density function.

The main tool in the analysis and synthesis of robust estimators for independent data is the influence function [38] which has been proposed on heuristic grounds but, nevertheless, contains information on the asymptotic bias and variance of robust maximum-likelihood (M) estimators of location and scale. For dependent data the situation is complicated by several technical arguments which have not yet been clearly resolved (see, for example, [68] and [69]). In the next section we generalize the 2-D GM-estimator as an asymptotic statistical functional for which we define a directional derivative (viz., an influence function) that can be used to guide the selection of the functions $\psi(\cdot)$ and $W(\cdot)$ in (5.2) and (5.3).

5.1.1 An Asymptotic GM-Estimator

In the following assume $\mathrm{E}\{x(\mathbf{s})\} = 0$, otherwise, robustly center the observations by replacing them with $x(\mathbf{s}) - \hat{x}_M$ where $\hat{x}_M$ is an ordinary M-estimate (see [39]) of the mean of $x(\mathbf{s})$. Let $\mathbf{x}$ and $\mathbf{x}^0$ represent dummy variables for the observed data neighbor set vectors $\mathbf{x}_\mathbf{s}$ and $\mathbf{x}^0_\mathbf{s}$, respectively, as defined in Section 1.1 of Chapter 1. The joint marginal probability distribution function of $\mathbf{x}^0_\mathbf{s}$ will be denoted by $\mathcal{F}^{N_0}(x)$. Then the asymptotic GM-estimates $\hat{\boldsymbol{\theta}}_{AGM}$ of $\boldsymbol{\theta}$ and S_{AGM} of β are defined by the functional $T(\mathcal{F}^{N_0}) = \mathrm{col}\{\boldsymbol{\theta}(\mathcal{F}^{N_0}), S(\mathcal{F}^{N_0})\}$ and computed as a root of the $p+1$ equations

$$\mathrm{E}_{\mathcal{F}^{N_0}}\{\mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0; \boldsymbol{\theta}(\mathcal{F}^{N_0}), S(\mathcal{F}^{N_0}))\} = \mathbf{0} \tag{5.4}$$

and

$$\mathrm{E}_{\mathcal{F}^{N_0}}\{W(\mathbf{x})[\psi^2(\mathbf{x}^0;\boldsymbol{\theta}(\mathcal{F}^{N_0}),S(\mathcal{F}^{N_0}))-B]\}=0. \tag{5.5}$$

Here, the constant c_v has been included in the definition of $\psi(\cdot)$ for simplicity.

Note that if the empirical distribution function of the observed data on an $M \times M$ lattice is given by

$$\mathcal{F}_M^{N_0}(x)=\frac{1}{M_2}\sum_{\mathbf{s}\in\Omega_M}\Delta(\mathbf{x}_\mathbf{s}^0)\,, \tag{5.6}$$

then the solution of (5.4) and (5.5) for $\mathcal{F}^{N_0}=\mathcal{F}_M^{N_0}(x)$ yields the estimates $\hat{\boldsymbol{\theta}}_{GM}$ and S_{GM} defined by (5.2) and (5.3).

By a suitable definition of the boundary conditions, $\mathcal{F}_M^{N_0}(x)$ is a reasonable estimate of $\mathcal{F}^{N_0}(x)$ in a variety of ways [70]. Furthermore, we expect that $T(\mathcal{F}_M^{N_0})$ relates to $T(\mathcal{F}^{N_0})$ in a similar fashion if $T(\cdot)$ is sufficiently well behaved. Hence, analysis of (5.4) and (5.5) should lead to reasonable prediction of the behavior of (5.2) and (5.3) for sufficiently large M.

Consider the following minimization problem:

$$\min_{\boldsymbol{\theta}}\left\{\int W(\mathbf{x})\rho(\mathbf{x}^0;\boldsymbol{\theta},S)d\mathcal{F}^{N_0}\right\} \tag{5.7}$$

for fixed S where $\rho(\mathbf{x}^0;\boldsymbol{\theta},S)$ is related to $\psi(\mathbf{x}^0;\boldsymbol{\theta},S)$ by

$$\frac{\partial}{\partial\boldsymbol{\theta}}\rho(\mathbf{x}^0;\boldsymbol{\theta},S)=\mathbf{x}\psi(\mathbf{x}^0;\boldsymbol{\theta},S). \tag{5.8}$$

The solution of (5.7) is also a root of (5.4). Moreover, if $\rho(v)$ is a convex function in v, implying that $\psi(v)$ must be strictly monotone, then the solution to (5.4) is unique. This follows from the fact that if $\rho(v)$ is continuously differentiable and convex, then (5.4) is both necessary and sufficient for the solution to be a global minimizing point [71]. Existence of the solution is guaranteed if $\rho(v)$ is symmetric.

5.1.2 An Influence Function for the GM-Estimator

Von Mises [45] has shown that a statistical functional $T(\cdot)$ at a distribution $\mathcal{G}$, which is "near" a distribution $\mathcal{F}$, can be written as a Taylor series expansion at $\mathcal{F}$ as in

$$T(\mathcal{G}) = T(\mathcal{F}) + \int t(x)d(\mathcal{G} - \mathcal{F}) + \text{remainder}, \tag{5.9}$$

if there exists a real function $t(\cdot)$ such that for all $\mathcal{G}$ in domain$\{T\}$ it holds that

$$T'(\mathcal{F}; \mathcal{G} - \mathcal{F}) \doteq \lim_{h \to 0} \frac{T(\mathcal{F} + t(\mathcal{G} - \mathcal{F})) - T(\mathcal{F})}{h} = \int t(x)d\mathcal{G}. \tag{5.10}$$

$T'(\mathcal{F}; \mathcal{G} - \mathcal{F})$ is known as the 1st-order Gâteaux derivative of $T(\mathcal{F})$. Note that $T'(\mathcal{F}; \mathcal{G} - \mathcal{F})$ is simply the ordinary right-hand derivative in the direction of $\mathcal{G}$, at $t = 0$, of the functional $T((1 - h)\mathcal{F} + h\mathcal{G})$. The derivative $T'(\mathcal{F}; \mathcal{G} - \mathcal{F})$ depends, in general, not only on $\mathcal{F}$ but also on the measure $\mathcal{G} - \mathcal{F}$. If $T'(\mathcal{F}; \mathcal{G} - \mathcal{F})$ is evaluated at $\mathcal{G} = \mathcal{F}$, then $T'(\mathcal{F}) = 0$, and consequently, $\int t(x)d\mathcal{F} = 0$.

Letting $\mathcal{G} = \Delta(x)$ in (5.10) we see that $T'(\mathcal{F}; \Delta(x) - \mathcal{F}) = t(x)$, which Hampel [38] calls the influence function (IF) of T at $\mathcal{F}$, which we write as $IF(x; T(\mathcal{F}))$. Thus, it is true that

$$T(\mathcal{G}) - T(\mathcal{F}) = \int IF(x; T(\mathcal{F}))d\mathcal{G} + \text{remainder}. \tag{5.11}$$

When (5.11) is evaluated with $\mathcal{G}$ equal to the observed sample distribution, $\mathcal{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \Delta(x_i)$, in most cases [38] the remainder becomes negligible for $n \to \infty$ so that

$$T(\mathcal{F}_n) - T(\mathcal{F}) \approx \frac{1}{n}\sum_{i=1}^{n} IF(x_i; T(\mathcal{F})), \tag{5.12}$$

which is the estimation error in the estimate $T(\mathcal{F}_n)$ of $T(\mathcal{F})$. It is this last expression that gives $IF(x; T(\mathcal{F}))$ its name, for $IF(x_i; T(\mathcal{F}))$ represents the approximate contribution, or influence, of the observation x_i toward the error. Moreover, if the

x_i are independent, then the terms on the right side of (5.12) are independent, and by the central limit theorem $\sqrt{n}[T(\mathcal{F}_n) - T(\mathcal{F})]$ is asymptotically normal, and it is a simple matter to show that asymptotic variance equals

$$V(T, \mathcal{F}) = \int IF^2(x; T(\mathcal{F}))d\mathcal{F} \,. \tag{5.13}$$

For the IID case we see that the bias and asymptotic efficiency of the estimator $T(\mathcal{F}_n)$ depend explicitly on $IF(x; T(\mathcal{F}))$. It turns out that M-estimators for the IID or regression model, defined by $\int \psi(x; T, \mathcal{F}))d\mathcal{F}$, have an $IF(\cdot)$ which is proportional to $\psi(\cdot)$. Thus, the desired robustness properties for the estimator may be achieved by simply selecting a ψ-function dictated by analysis of the results (5.12) and (5.13).

The approach that we take in this chapter is to define a similar device for $T(\mathcal{F}^{N_0})$, then use it to guide the selection of $\psi(\cdot)$ and $W(\cdot)$ used in the GM-estimator. Unlike the IID case, the $(p+1)$-dimensional distribution $\mathcal{F}^{N_0,h} = (1-h)\mathcal{F}^{N_0} + h\Delta(\mathbf{x}^0)$ for $0 \leq h \leq 1$ does not correspond directly with any naturally occurring contamination in the random field [68]. This problem occurs because $\Delta(\mathbf{x}^0)$ is not a stationary distribution in the set of $(p+1)$-dimensional marginal distributions, and it matters not only the magnitude of $\mathbf{x}^0$ but also the location of the contamination. Patchy outliers will have quite a different effect than will isolated outliers. Nevertheless, we show that we obtain an equation similar to (5.12) and that an influence function defined with $\mathcal{F}^{N_0,h}$ is proportional to the kernels of (5.4) and (5.5) except for a constant multiplying factor. Thus, similar reasoning to the IID case is used for selecting the GM-estimator ψ-function which is composed of $\psi(\cdot)$ and $W(\cdot)$.

Let $\boldsymbol{\theta}_h = \boldsymbol{\theta}(\mathcal{F}^{N_0,h})$ and $S_h = S(\mathcal{F}^{N_0,h})$ and note that $\boldsymbol{\theta}_0$ and S_0 are the asymptotic GM-estimates when the data are distributed with marginal $\mathcal{F}^{N_0}$, i.e., when $h = 0$. Substitute $\mathcal{F}^{N_0,h}$ in (5.4) and (5.5) and differentiate with respect to h at $h = 0$, obtaining

$$\frac{\partial}{\partial h} \int \mathbf{x} W(\mathbf{x}) \psi(\mathbf{x}^0; \boldsymbol{\theta}_h, S_h) d\mathcal{F}^{N_0,h} \Big|_{h=0} = \mathbf{0} \tag{5.14}$$

and

$$\frac{\partial}{\partial h} \int W(\mathbf{x}) [\psi^2(\mathbf{x}^0; \boldsymbol{\theta}_h, S_h) - B] d\mathcal{F}^{N_0,h} \Big|_{h=0} = 0 \,. \tag{5.15}$$

Here we make the assumption that $\psi(\cdot)$ and $W(\cdot)$ are sufficiently well behaved so that the processes of integration and differentiation are interchangeable.

If we define a $(p+1)$-dimensional GM-estimator influence function for $\boldsymbol{\theta}$ and S under $\mathcal{F}^{N_0}$ as

$$IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta}, S) = \left[\begin{array}{c} \frac{\partial \boldsymbol{\theta}_h}{\partial h} \\ \frac{\partial S_h}{\partial h} \end{array} \right]_{h=0}$$

and carry out the indicated differentiation in (5.14) and (5.15), the result in matrix-vector form is

$$IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta}, S) = \left[\begin{array}{cc} M_\theta & \mathbf{m}_{\theta,S} \\ \mathbf{m}_{S,\theta}^t & m_S \end{array} \right]^{-1} \left[\begin{array}{c} \mathbf{x} W(\mathbf{x}) \psi(\mathbf{x}^0; \boldsymbol{\theta}, S) \\ W(\mathbf{x}) [\psi^2(\mathbf{x}^0; \boldsymbol{\theta}, S) - B] \end{array} \right], \tag{5.16}$$

assuming the inverse exists and where

$$\begin{aligned} M_\theta &= \frac{1}{S} \int W(\mathbf{x}) \psi'(\mathbf{x}^0; \boldsymbol{\theta}, S) \mathbf{x}\mathbf{x}^t d\mathcal{F}^{N_0} \\ \mathbf{m}_{\theta,S} &= \frac{1}{S} \int \mathbf{x} W(\mathbf{x}) \psi'(\mathbf{x}^0; \boldsymbol{\theta}, S) \left(\frac{x_0 - \boldsymbol{\theta}^t \mathbf{x}}{S} \right) d\mathcal{F}^{N_0} \\ \mathbf{m}_{S,\theta} &= \frac{2}{S} \int \mathbf{x} W(\mathbf{x}) \psi'(\mathbf{x}^0; \boldsymbol{\theta}, S) \psi(\mathbf{x}^0; \boldsymbol{\theta}, S) d\mathcal{F}^{N_0} \\ m_S &= \frac{2}{S} \int W(\mathbf{x}) \psi'(\mathbf{x}^0; \boldsymbol{\theta}, S) \psi(\mathbf{x}^0; \boldsymbol{\theta}, S) \left(\frac{x_0 - \boldsymbol{\theta}^t \mathbf{x}}{S} \right) d\mathcal{F}^{N_0} \end{aligned}$$

and $\psi'(v) = \frac{d}{dv}\psi(v)$. We have dropped the subscript on $\boldsymbol{\theta}_0$ and S_0, but it is kept in mind that $IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta}, S)$ is usually evaluated at the values of $\boldsymbol{\theta}$ and S corresponding with the model distribution $\mathcal{F}^{N_0}$.

We note here that for symmetry reasons (which will become evident in subsequent sections) if outliers occur only in the innovations and the innovations distribution is symmetric, whether or not Gaussian, then $\mathbf{m}_{\theta,S}$ and $\mathbf{m}_{S,\theta}$ are both zero vectors. In this case the influence function separates, i.e.,

$$\begin{bmatrix} IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta}) \\ IF_{GM}(\mathbf{x}^0; S) \end{bmatrix} = \begin{bmatrix} M_\theta^{-1} \mathbf{x} W(\mathbf{x}) \psi(\mathbf{x}^0; \boldsymbol{\theta}, S) \\ m_S^{-1} W(\mathbf{x})[\psi^2(\mathbf{x}^0; \boldsymbol{\theta}, S) - B] \end{bmatrix} \tag{5.17}$$

where $IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta})$ and $IF_{GM}(\mathbf{x}^0; S)$ are the influence functions for the separate estimates $\hat{\boldsymbol{\theta}}_{AGM}$ with S known and S_{AGM} with $\boldsymbol{\theta}$ known, respectively. When the observed data obey the AO model and $n(\mathbf{s})$ is not identically zero, then (5.17) should hold approximately for the AO model as well.

The importance of this separation is that the asymptotic robustness properties of the estimate $\hat{\boldsymbol{\theta}}_{GM}$ do not depend on S_{GM}, and vice versa. This suggests an alternative to simultaneous solution of (5.2) and (5.3): compute any robust estimate of scale and then use this in solving (5.2). Additionally, we can evaluate the asymptotic properties of the parameter estimates $\hat{\boldsymbol{\theta}}_{GM}$ by considering a fixed scale S.

By an expansion similar to (5.9) we obtain

$$T(\mathcal{F}_M^{N_0}) - T(\mathcal{F}^{N_0}) = \frac{1}{M^2} \sum_{\mathbf{s} \in \Omega_M} IF_{GM}(\mathbf{x}_\mathbf{s}^0; T(\mathcal{F}^{N_0})) + \text{remainder}. \tag{5.18}$$

where the remainder, under suitable regularity conditions [70], becomes negligible when $M \to \infty$. Also, we find that

$$M[T(\mathcal{F}_M^{N_0}) - T(\mathcal{F}^{N_0})] \longrightarrow \mathcal{N}(\mathbf{0}, V(T, \mathcal{F}^{N_0})), \tag{5.19}$$

where

$$V(T, \mathcal{F}^{N_0}) = \int IF_{GM}(\mathbf{x}^0; T(\mathcal{F}^{N_0}))IF'_{GM}(\mathbf{x}^0; T(\mathcal{F}^{N_0}))d\mathcal{F}^{N_0}. \tag{5.20}$$

Proof: Clearly, $\int IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta}, S)d\mathcal{F}^{N_0} = \mathbf{0}$. By a suitable selection of $\psi(\cdot)$ and $W(\cdot)$, $IF_{GM}(\mathbf{x}^0_{\mathbf{s}}; \boldsymbol{\theta}, S)$ for all $\mathbf{s}$ will be uniformly bounded and a stationary process. When $\mathbf{s}-\mathbf{t} \notin N \cup (0,0)$, then $IF_{GM}(\mathbf{x}^0_{\mathbf{s}}; \boldsymbol{\theta}, S)$ and $IF_{GM}(\mathbf{x}^0_{\mathbf{t}}; \boldsymbol{\theta}, S)$ are independent and $IF_{GM}(\mathbf{x}^0_{\mathbf{s}}; \boldsymbol{\theta}, S)$ is α-mixing. Under these conditions a central limit theorem for dependent random vectors [72] applies and (5.19) follows. Since NSHP and GMRF models have a Markov property it is easy to show that

$$\int IF_{GM}(\mathbf{x}^0; \boldsymbol{\theta}, S)d\mathcal{F}(x_0|\mathbf{x}) = \mathbf{0}.$$

Then $IF_{GM}(\mathbf{x}^0_{\mathbf{s}}; \boldsymbol{\theta}, S)$ is a martingale difference so that $IF_{GM}(\mathbf{x}^0_{\mathbf{s}}; \boldsymbol{\theta}, S)$ and $IF_{GM}(\mathbf{x}^0_{\mathbf{t}}; \boldsymbol{\theta}, S)$ are uncorrelated for $\mathbf{s} \neq \mathbf{t}$ and (5.20) follows [38].

5.1.3 Selection of the Robustifying Functions

Note from (5.17) that the 2-D asymptotic GM-estimator influence function for $\boldsymbol{\theta}$ is proportional to $\psi^*(\mathbf{x}^0; \boldsymbol{\theta}, S) = \mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0; \boldsymbol{\theta}, S)$ and for S is proportional to $\chi^*(\mathbf{x}^0; \boldsymbol{\theta}, S) = W(\mathbf{x})[\psi^2(\mathbf{x}^0; \boldsymbol{\theta}, S) - B]$. Hence, robustness criteria for the influence function translate directly into similar requirements for the kernel functions $\psi^*(\mathbf{x}^0; \boldsymbol{\theta}, S)$ and $\chi^*(\mathbf{x}^0; \boldsymbol{\theta}, S)$. Therefore, selection of $\psi(\cdot)$ and $W(\cdot)$ determine the performance of the estimator.

Hampel [38] has suggested that the influence function for M-estimators meet the following robustness criteria:

1. The influence function should be bounded. This guarantees that no single observation can have an unlimited influence on the value of the estimate.

2. The influence function should be continuous so that small perturbations of the data will result in small changes to the estimation error.

3. The influence function should return to zero. Ridiculously large outliers in the data should have no influence at all on the estimate.

These criteria correspond directly with the more technically defined criteria of *gross-error sensitivity*, *local-shift sensitivity*, and *rejection point*, which may be found in [38]. Two functions proposed for M-estimators in the IID case are Huber's ψ_H-function and Tukey's bisquare ψ_B-function. The former, defined by

$$\psi_H(z) = \begin{cases} z, & |z| < 1 \\ 1, & z \geq 1 \\ -1, & z \leq -1 \end{cases} \tag{5.21}$$

is shown graphically in Figure 5.1 and meets the first two basic criteria. Tukey's ψ_B-function, defined by

$$\psi_B(z) = \begin{cases} z(1-z^2)^2, & |z| < 1 \\ 0, & |z| \geq 1 \end{cases} \tag{5.22}$$

and shown graphically in Figure 5.2, is a redescending function which meets all three criteria. Because ψ_B returns to zero, it provides an extra measure of robustness against extremely large outliers while sacrificing efficiency at the nominal model.

In view of the similarity between the expansions (5.9) and (5.18) we take these criteria to also apply to the 2-D GM-estimator influence function. Specifically, the influence function, and consequently the kernel functions $\psi^*(\mathbf{x}^0; \boldsymbol{\theta}, S)$ and $\chi^*(\mathbf{x}^0; \boldsymbol{\theta}, S)$, should also have the same kinds of properties as ψ_H and ψ_B. For example, if $\psi(\cdot) = \psi_H(\cdot)$, then $W(\cdot)$ should be chosen so that $\mathbf{x}W(\mathbf{x})$ is bounded

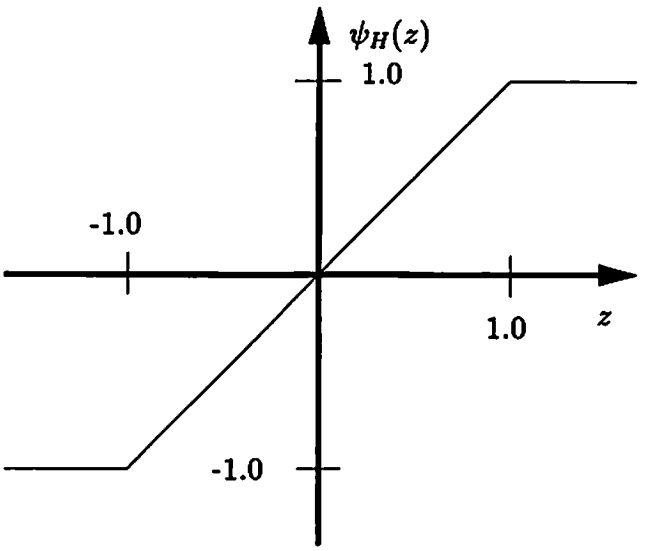


Figure 5.1: Huber's ψ_H-function.

for each element of $\mathbf{x}$, i.e, $W(\cdot)$ should down-weight elements of $\mathbf{x}$ which contain outliers. A natural way of accomplishing this is to let $W(x) = \frac{c}{d}g(\frac{d}{c})$, where d is a measure of the largeness in $\mathbf{x}$ obtained from $d^2 = \frac{1}{p}\mathbf{x}^t C_{\mathbf{x}}^{-1}\mathbf{x}$. Here c is a constant and $C_{\mathbf{x}} = \mathrm{E}\{\mathbf{x}\mathbf{x}^t\}$ is the $p \times p$ covariance matrix for the past history vector $\mathbf{x}_s$ of the clean process. In practice, a function, such as Tukey's bisquare $\psi_H(\cdot)$, which redescends to zero is used for $g(\cdot)$ to insure that $\mathbf{x}W(\mathbf{x})$ remains bounded for arbitrarily large elements in $\mathbf{x}$. Several procedures are available for determining an estimate of $C_{\mathbf{x}}$; for example, see [42]. In this paper we will not discuss this topic but instead assume that $C_{\mathbf{x}}$ is known or estimate it from clean data.

5.1.4 Tuning Constants

The tuning constant B is chosen to make the estimate S_{AGM} consistent when the signal $y(\mathbf{s})$ is observed without error and the innovations process is distributed as

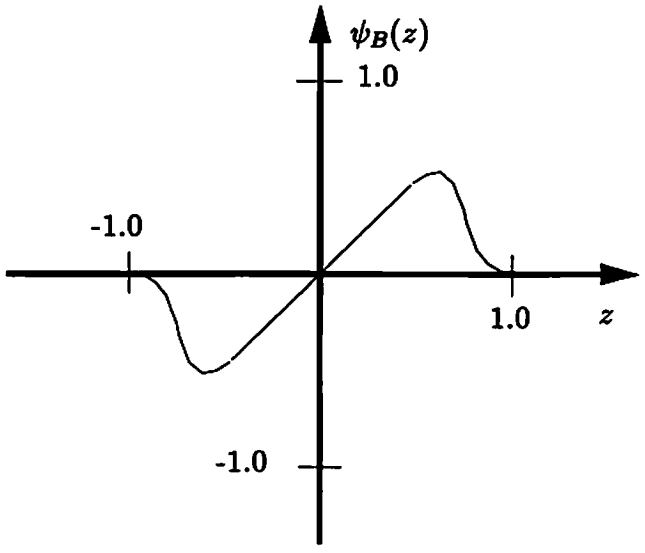


Figure 5.2: Tukey's redescending bisquare ψ_B-function.

$\mathcal{N}(0,1)$. Under these circumstances, (5.5) becomes

$$\int W(\mathbf{x}) \left[c_v^2 \psi^2 \left(\frac{x_0 - \theta^t \mathbf{x}}{c_v S} \right) - B \right] d\mathcal{F}_y^{N_0}(\mathbf{x}^0) = 0. \tag{5.23}$$

The tuning constant is then computed by

$$B = \int c_v^2 \psi^2 \left(\frac{v}{c_v} \right) f_v(v) dv \tag{5.24}$$

where $f_v(v) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{v^2}{2})$.

Proof: Clearly, the data is normal with $\mathcal{F}(x) = \mathcal{F}(y) = \mathcal{N}(0, \sigma_y^2)$. Let $f_y(\mathbf{y}^0)$ be the density function for $\mathcal{F}^{N_0}(y)$, i.e., $\mathcal{F}^{N_0}(y) = \int_{-\infty}^{y} f_y(\mathbf{x}^0) d\mathbf{x}^0$. This joint density can be written in terms of a conditional density and a p-dimensional joint density as $f_y(\mathbf{y}) = f_y(y_0|\mathbf{y}) f_y(\mathbf{y})$. Using these relationships (5.5) becomes

$$\int W(\mathbf{x}) \left\{ \int \left[c_v^2 \psi^2 \left(\frac{x_0 - \theta^t \mathbf{x}}{c_v S} \right) - B \right] f_y(x_0|\mathbf{x}) dx_0 \right\} f_y(\mathbf{x}) d\mathbf{x} = 0. \tag{5.25}$$

Since the NSHP and GMRF models possess the Markov property the conditional density $f_y(y_0|\mathbf{y})$ can be shown to be

$$f_y(y_0|\mathbf{y}) = (2\pi S^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\frac{y_0 - \theta^t \mathbf{y}}{S}\right)^2\right], \tag{5.26}$$

where $S = \beta$ for the NSHP model and $S = \sqrt{\nu}$ for the GMRF model. Now, let $v = \dfrac{x_0 - \theta^t \mathbf{x}}{S}$ and make this change of variable in (5.25), yielding

$$\int W(\mathbf{x}) \left\{ \int \left[c_v^2 \psi^2 \left(\frac{v}{c_v}\right) - B \right] (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{v^2}{2}\right) dv \right\} f_y(\mathbf{x}) d\mathbf{x} = 0. \tag{5.27}$$

The integral in braces is independent of $\mathbf{x}$, therefore, if $W(\cdot)$ is a non-negative and symmetric function, then

$$\int \left[c_v^2 \psi^2 \left(\frac{v}{c_v}\right) - B \right] (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{v^2}{2}\right) dv = 0,$$

and the result (5.24) follows. The constant B, evaluated for selected values of c_v using both $\psi_H(\cdot)$ and $\psi_B(\cdot)$, is shown in Table 5.1.

The constant c_v adjusts the robustness properties of the GM-estimator. Generally, for smaller values of c_v the estimates are more robust with respect to additive effects in the observed data. The compromise, however, is a reduction in the efficiency of the estimator at the nominal, or Gaussian, model. To show this and to obtain insight into the selection of values for c_v we compute the asymptotic variances for the GM-estimator under a Gaussian IO model. Let

$$M(\psi, \mathcal{F}^{N_0}) = \begin{bmatrix} M_\theta & \mathbf{m}_{\theta,S} \\ \mathbf{m}_{S,\theta}^t & m_S \end{bmatrix} \tag{5.28}$$

and rewrite the influence function as

$$IF_{GM}(\mathbf{x}^0; T(\mathcal{F}^{N_0})) = M^{-1}(\psi, \mathcal{F}^{N_0}) \begin{bmatrix} \psi^*(\mathbf{x}^0; \theta, S) \\ \chi^*(\mathbf{x}^0; \theta, S) \end{bmatrix}. \tag{5.29}$$

The asymptotic variance in (5.20) becomes

$$V(T,\mathcal{F}^{N_0}) = M^{-1}(\psi,\mathcal{F}^{N_0}) \begin{bmatrix} \int \psi^*\psi^{*t} d\mathcal{F}^{N_0} & \int \psi^*\chi^* d\mathcal{F}^{N_0} \\ \int \chi^*\psi^{*t} d\mathcal{F}^{N_0} & \int \chi^{*2} d\mathcal{F}^{N_0} \end{bmatrix} M^{-t}(\psi,\mathcal{F}^{N_0}). \tag{5.30}$$

For the IO model with Gaussian distribution the asymptotic estimates $\hat{\theta}_{AGM}$ and S_{AGM} equal the true values, and $v(\mathbf{s}) = \dfrac{x(\mathbf{s}) - \theta^t \mathbf{x_s}}{S}$ is distributed $f_v(v) = \mathcal{N}(0,1)$. It is now a simple matter to show that

$$V(T,\mathcal{F}^{N_0}) = \begin{bmatrix} M_\theta^{-1} Q(\psi^*,\mathcal{F}^{N_0}) M_\theta^{-t} & \mathbf{0} \\ \mathbf{0}^t & m_S^{-2} P(\chi^*,\mathcal{F}^{N_0}) \end{bmatrix} \tag{5.31}$$

where

$$\begin{aligned} Q(\psi^*,\mathcal{F}^{N_0}) &= C_{W_{\mathbf{x}}^2} \int \psi^2(v) f_v(v) dv \\ P(\chi^*,\mathcal{F}^{N_0}) &= \overline{W}_2 \int [\psi^2(v) - B]^2 f_v(v) dv, \end{aligned}$$

and

$$\begin{aligned} C_{W_{\mathbf{x}}^2} &= \mathrm{E}\{W^2(\mathbf{x})\mathbf{x}\mathbf{x}^t\} \\ \overline{W}_2 &= \mathrm{E}\{W^2(\mathbf{x})\}. \end{aligned}$$

The expressions for M_θ and m_S can also be simplified to

$$\begin{aligned} M_\theta &= C_{W\mathbf{x}} \int \psi'(v) f_v(v) dv \\ m_S &= 2\overline{W}_1 \int \psi'(v)\psi(v) v f_v(v) dv, \end{aligned}$$

where

$$\begin{aligned} C_{W\mathbf{x}} &= \mathrm{E}\{W(\mathbf{x})\mathbf{x}\mathbf{x}^t\} \\ \overline{W}_1 &= \mathrm{E}\{W(\mathbf{x})\}. \end{aligned}$$

It is clearly the case that under the same conditions and the fact that if $\psi(\cdot)$ is an odd function, then $\mathbf{m}_{\theta,S} = \mathbf{0}$ and $\mathbf{m}_{S,\theta} = \mathbf{0}$.

Using these results in (5.31) the asymptotic covariance matrix for the estimate $\hat{\theta}_{AGM}$ is then

$$V(\theta, \mathcal{F}^{N_0}) = S^2 \frac{\int \psi^2(v) f_v(v) dv}{\left[\int \psi'(v) f_v(v) dv\right]^2} C_{W_{\mathbf{x}}}^{-1} C_{W_{\mathbf{x}}^2} C_{W_{\mathbf{x}}}^{-1}, \tag{5.32}$$

and similarly the asymptotic variance of the estimate S is

$$V(S, \mathcal{F}^{N_0}) = \frac{S^2}{4} \frac{\int [\psi^2(v) - B]^2 f_v(v) dv}{\left[\int \psi'(v) \psi(v) v f_v(v) dv\right]^2} \frac{\overline{W}_2}{\overline{W}_1^2}. \tag{5.33}$$

Ordinary M-estimators are defined when $W(\mathbf{x}) \equiv 1$ for all $\mathbf{x}$. In this case $C_{W_{\mathbf{x}}} = C_{W_{\mathbf{x}}^2} = C_{\mathbf{x}}$, the covariance matrix of the p-dimensional data vector $\mathbf{x}_s$, $\overline{W}_1 = \overline{W}_2 = 1$, and (5.32) and (5.33) are the asymptotic variances of the M-estimators for θ and S.

If in addition $\psi(v) = v$ for all v, then the GM-estimator reduces to the LS estimator of θ and S. Carrying out the computations in (5.32) and (5.33) for the Gaussian situation yields

$$V_{LS}(\theta, \mathcal{N}(0, \sigma_y^2)) = S^2 C_{\mathbf{x}}^{-1} \tag{5.34}$$

and

$$V_{LS}(S, \mathcal{N}(0, \sigma_y^2)) = \frac{S^2}{2}. \tag{5.35}$$

When the data comes from the Gaussian IO model by definition there are no outliers. The optimal function $W(\cdot)$ should be $W(\mathbf{x}) \equiv 1$ for all $\mathbf{x}$. In this case the asymptotic GM-estimator covariance matrix for θ reduces to

$$V(\theta, \mathcal{N}(0, \sigma_y^2)) = S^2 \frac{\int \psi^2(v) f_v(v) dv}{\left[\int \psi'(v) f_v(v) dv\right]^2} C_{\mathbf{x}}^{-1}. \tag{5.36}$$

In most situations the true values of the parameters are not known. Consequently, the evaluation of a Cramer-Rao bound is not possible, although an analytical expression can easily be computed using the joint distribution $\mathcal{F}^{N_0}$ or the influence function in the Gaussian case. The limiting Fisher information matrix $\Gamma(\theta)$ turns out to be given by

$$\Gamma(\theta) = \frac{1}{S^2} C_{\mathbf{x}} + \frac{f'(\mathbf{x})}{f(\mathbf{x})} \tag{5.37}$$

where $f'(\mathbf{x}) = \frac{\partial}{\partial \theta} f(\mathbf{x})$ which is a complicated function of the parameters θ. Our stated purpose though is to determine guides for selection of the tuning constant c_v. Actual efficiencies are best obtained from Monte Carlo results. In fact, analytical results for the AO model are extremely difficult since the distribution function is not even available.

We are motivated then to define an asymptotic efficiency relative to the LS estimates, which are consistent for the NSHP and GMRF models,. Thus,

$$\text{Eff}_{GM} \doteq \frac{\text{tr}\left\{V_{LS}(\theta, \mathcal{F}^{N_0})\right\}}{\text{tr}\left\{V(\theta, \mathcal{F}^{N_0})\right\}}. \tag{5.38}$$

Substituting the expressions for $V_{LS}(\theta, \mathcal{N}(0, \sigma_y^2))$ and $V(\theta, \mathcal{N}(0, \sigma_y^2))$ in (5.38) yields the relative GM-estimator efficiency

$$\text{REff}_{GM} = \frac{[\int \psi'(v) f_v(v) dv]^2}{\int \psi^2(v) f_v(v) dv}. \tag{5.39}$$

Values of REff_{GM}, computed for various values of the tuning constant c_v are shown in Table 5.1.

	ψ_H		ψ_B	
c_v	B	REff_{GM}	B	REff_{GM}
1.0	0.5161	0.9031	0.0266	0.1250
1.1	0.5777	0.9191	0.0343	0.1368
1.2	0.6352	0.9330	0.0433	0.1632
1.3	0.6880	0.9451	0.0537	0.1954
1.4	0.7358	0.9555	0.0652	0.2309
1.5	0.7785	0.9642	0.0778	0.2685
2.0	0.9205	0.9897	0.1552	0.4666
3.0	0.9950	0.9996	0.3434	0.7727
4.0	0.9999	1.0000	0.5134	0.9100
5.0	1.0000	1.0000	0.6395	0.9611
6.0	1.0000	1.0000	0.7277	0.9810
10.0	1.0000	1.0000	0.8886	0.9976
20.0	1.0000	1.0000	0.9706	0.9998
∞	1.0000	1.0000	1.0000	1.0000

Table 5.1: Tuning constant B and relative efficiency for selected values of c_v using ψ_H and ψ_B.

5.2 Experimental Results

Experiments with synthetic data were conducted to evaluate the performance of the 2-D GM-estimator. The first experiment compares robust spectrum estimation results for two sinusoids in noise with the conventional approach of LS estimation. It is well known that NSHP modeling of complex spectra requires large model order to resolve details [12]; and since the robust technique requires extensive computational capacity due to its iterative nature, the GM-estimator was evaluated on smaller data sets and low order models in Experiments 2, 3, and 4. In these experiments many runs were made against different data sets to obtain a feel for the statistical performance of the GM-estimator. The raw data from which the graphs in this paper were made can be found in [73]. No spectra were computed in Experiments 2, 3, and 4 since the spectrum for low order models has no interesting detail.

Experiment 1: An important application of spectrum estimation is the detection and resolution of two closely spaced sinusoids in noise. In one and two dimensional studies this problem is usually formulated as sinusoids in Gaussian white noise. Here, for evaluating the performance of the GM-estimator we compare the Gaussian white case to the case when the contaminating noise has a heavy tailed non-Gaussian distribution. The LS and GM procedures were used to compute estimates from both Gaussian and non-Gaussian data, thus yielding four spectra for comparison.

We generated a 64 × 64 set of lattice data according to

$$y(\mathbf{s}) = A_1 \cos(\boldsymbol{\varphi}_1^t \mathbf{s} + \alpha_1) + A_2 \cos(\boldsymbol{\varphi}_2^t \mathbf{s} + \alpha_2) + \eta(\mathbf{s})$$

with $\boldsymbol{\varphi}_1 = \frac{\pi}{64}[16.0, 16.0]^t$, $\boldsymbol{\varphi}_2 = \frac{\pi}{64}[16.0, 20.0]^t$, $A_1 = A_2 = 1.0$, $\alpha_1 = .2$, $\alpha_2 = .3$ and $\eta(\mathbf{s})$ IID with common distribution $\mathcal{N}(0, .05)$, equivalent to a SNR = 10db. In this experiment a NSHP model with a 40 element neighbor set, shown in Figure 5.3, was used.

The estimated spectrum of the signal plus Gaussian noise using conventional LS estimates is shown in Figure 5.4, and the spectrum using the GM-estimator on the same data is shown in Figure 5.5. Huber's ψ_H-function was chosen for $\psi(\mathbf{x}^0; \boldsymbol{\theta}, S)$ and $W(x) = \frac{c}{d} g(\frac{d}{c})$ with Tukey's bisquare function used for $g(\cdot)$. The tuning constant $B = 0.7785$ was chosen from Table 5.1 for $c_v = 1.5$. The constant $c = 6.0$ was used for $W(\cdot)$. The LS and the GM-estimates result in similar spectra, both resolving two peaks at $\omega_1 = \left[\frac{8\pi}{32}, \frac{8\pi}{32}\right]^t$ and $\omega_2 = \left[\frac{8\pi}{32}, \frac{10\pi}{32}\right]^t$, which are correct. Thus, the GM-estimator does almost equally well as the conventional procedure in the Gaussian situation.

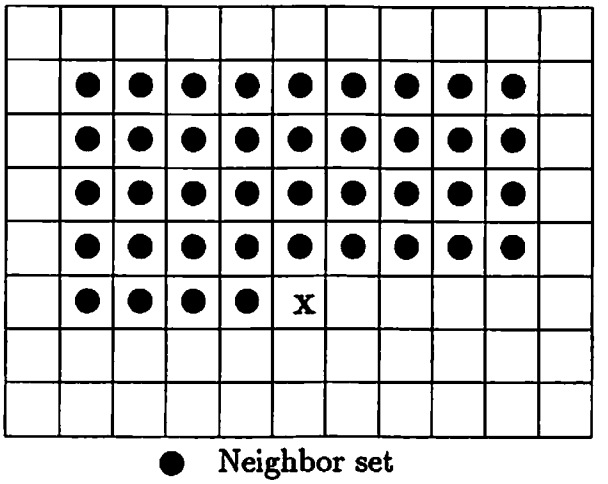


Figure 5.3: Neighbor set.

Next, data contaminated with a heavy-tailed distribution were formed by adding outliers from a distribution $\mathcal{N}(0, 1.0)$ to 10 percent of the signal plus Gaussian white noise at uniformly distributed lattice sites. The estimated spectrum using LS estimates from the contaminated data is shown in Figure 5.6, and the spectrum computed using the GM-estimate is shown in Figure 5.7. Here, the spectrum computed using LS-estimates from the heavy-tailed data is relatively poor; but the GM-estimator does almost as well as when the data are Gaussian.

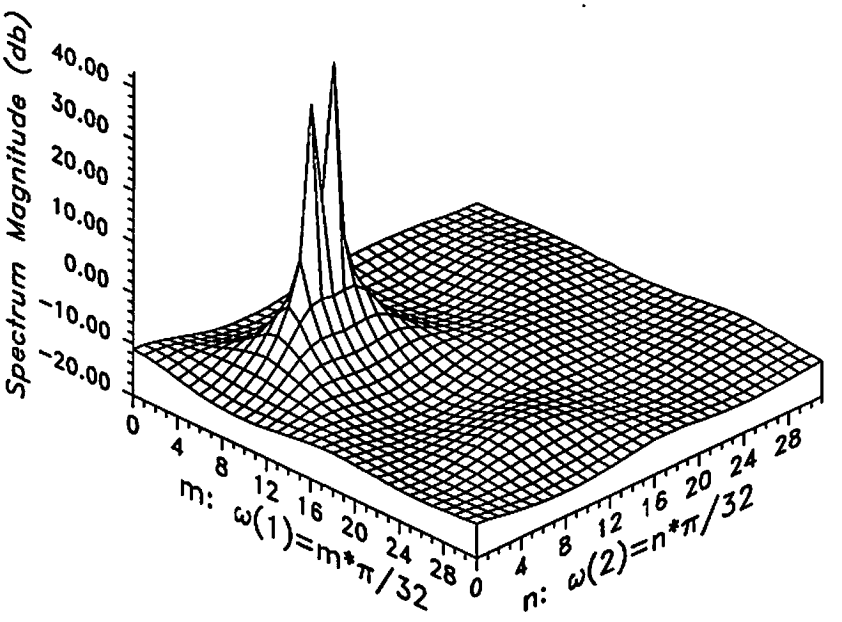


Figure 5.4: LS estimate of the spectrum of two sinusoids in Gaussian noise with SNR = 10db.

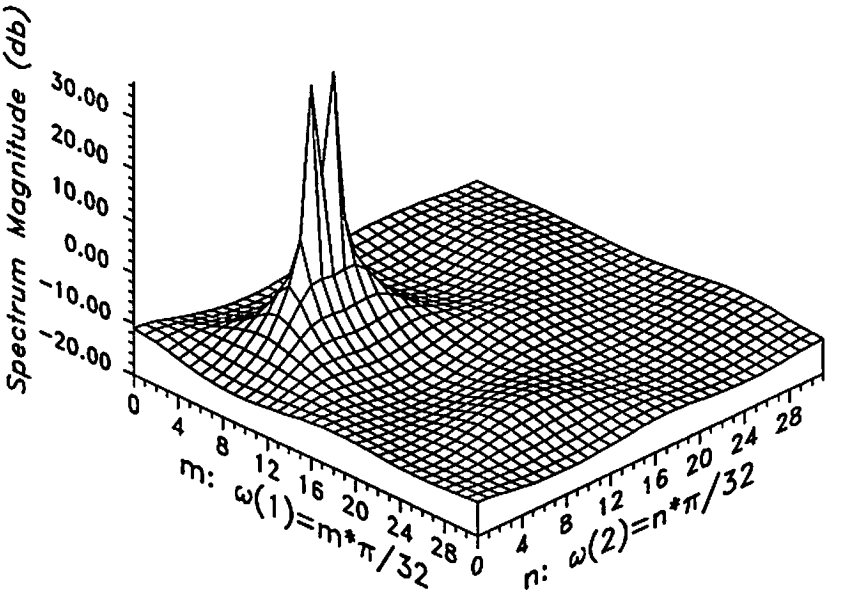


Figure 5.5: GM-estimate of the spectrum of two sinusoids in Gaussian noise with SNR = 10db.

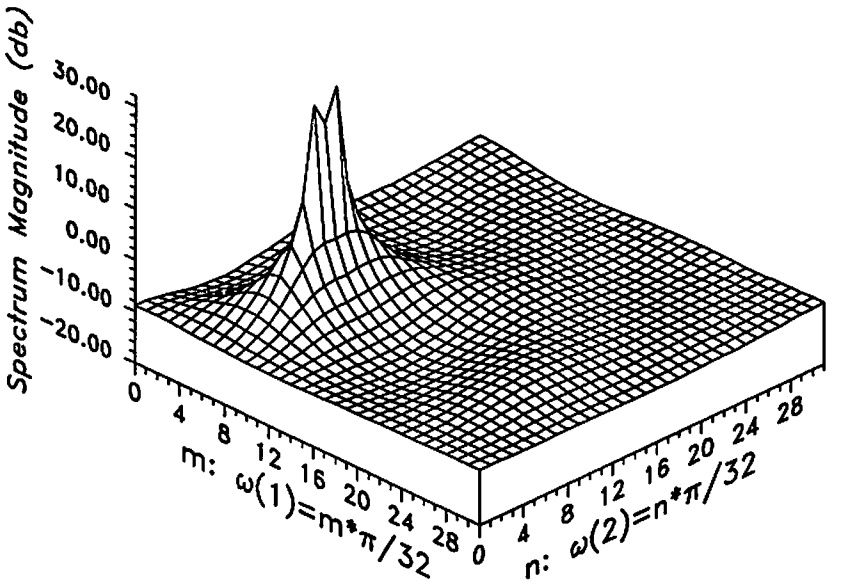


Figure 5.6: LS estimate of the spectrum of two sinusoids in non-Gaussian noise with SNR = 10db and 10 percent additive outliers.

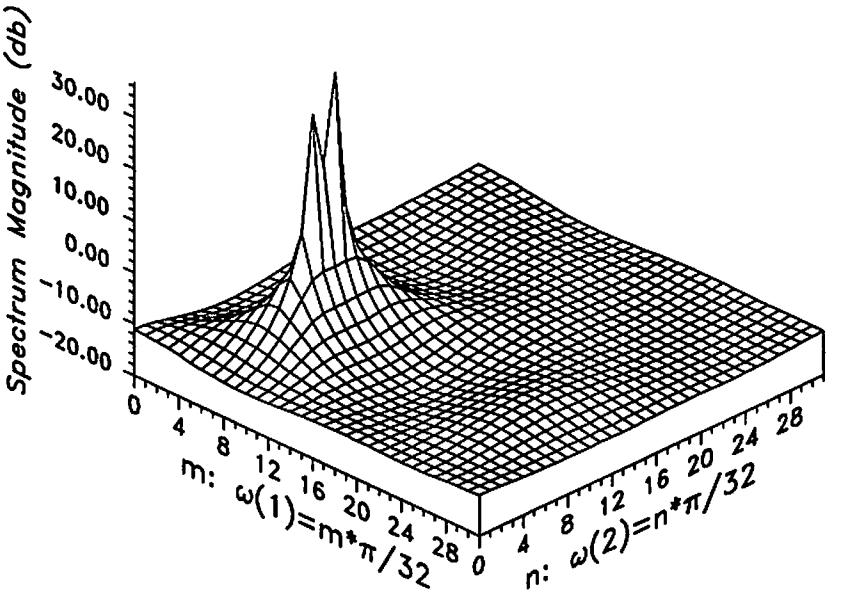


Figure 5.7: GM-estimate of the spectrum of two sinusoids in non-Gaussian noise with SNR = 10db and 10 percent additive outliers.

Experiment 2: In the first part of this experiment we generated data using a 1st-order GMRF model on a 32 × 32 toroidal lattice with parameter values shown in Table 5.2. Contaminated data were formed by adding outliers from a distribution $\mathcal{N}(0, 20.25)$ to 5 percent of the data at uniformly distributed lattice sites. Both LS estimates and robust GM-estimates from the contaminated data sets were computed and compared to the theoretical parameter values and the LS estimates from the clean data. Twelve cases were run by using different sets of outliers. The ψ-functions and tuning constants were chosen to be same as in Experiment 1. Results are shown in Table 5.3. Figure 5.8 compares graphically the squared errors, given by

$$\sigma_r^2 = \frac{1}{p} \sum_{r \in N} (\hat{\theta}_r - \theta_r^0)^2,$$

where θ_r^0, $r \in N$, are the true values of the parameters from Table 5.2.

$\theta_{1,0} = \theta_{-1,0}$	$\theta_{0,1} = \theta_{0,-1}$	ν
.23400	.10110	1.00000

Table 5.2: Neighbor set and coefficients for the GMRF model.

One easily sees that the GM-estimates for both θ and ν in the non-Gaussian AO model situation are better than LS-estimates for the same data in every simulation run. On the other hand the LS and GM-estimators yield similar results for Gaussian data.

In the second part of this experiment we generated 12 sets of contaminated data by using 12 different GMRF data sets and adding 5 percent outliers to each set of data. The LS and GM-estimation errors for each run are shown in Figure 5.9. The conclusions are identical to the first part of this experiment.

Run No.	Est Method	$\theta_{(1,0)} = \theta_{(-1,0)}$	$\theta_{(0,1)} = \theta_{(0,-1)}$	ν	Sq Err
	Theoretical Value	.234000	.101100	1.00000	
	LS Clean Data	.236163	.106680	1.00565	
	GM Clean Data	.229264	.103671	1.03459	.000028
1	LS Contaminated Data	.133592	.045695	2.48139	.007120
	GM Contaminated Data	.159626	.076479	1.27725	.003385
2	LS Contaminated Data	.134956	.062564	1.70262	.006095
	GM Contaminated Data	.167096	.077300	1.17206	.002817
3	LS Contaminated Data	.179539	.076275	1.72544	.002065
	GM Contaminated Data	.182444	.085116	1.16808	.001675
4	LS Contaminated Data	.137559	.093601	2.05112	.004947
	GM Contaminated Data	.157125	.090598	1.17712	.003253
5	LS Contaminated Data	.205881	.067976	1.66953	.001208
	GM Contaminated Data	.197621	.079790	1.18175	.001104
6	LS Contaminated Data	.162913	.065956	1.84768	.003512
	GM Contaminated Data	.175196	.077595	1.18580	.002281
7	LS Contaminated Data	.121111	.102253	2.08793	.006628
	GM Contaminated Data	.160067	.086082	1.19599	.003107
8	LS Contaminated Data	.115475	.090131	2.21251	.007420
	GM Contaminated Data	.160979	.103696	1.18925	.002831
9	LS Contaminated Data	.135594	.106215	2.02817	.005057
	GM Contaminated Data	.177840	.077823	1.22309	.002117
10	LS Contaminated Data	.175925	.046927	2.07145	.003600
	GM Contaminated Data	.180178	.085476	1.21113	.001792
11	LS Contaminated Data	.169912	.062514	1.88382	.003170
	GM Contaminated Data	.186321	.088134	1.16140	.001414
12	LS Contaminated Data	.188894	.080131	1.70692	.001470
	GM Contaminated Data	.186448	.084844	1.20755	.001474

Table 5.3: GM-estimator results for additive outliers in a GMRF with changing contamination.

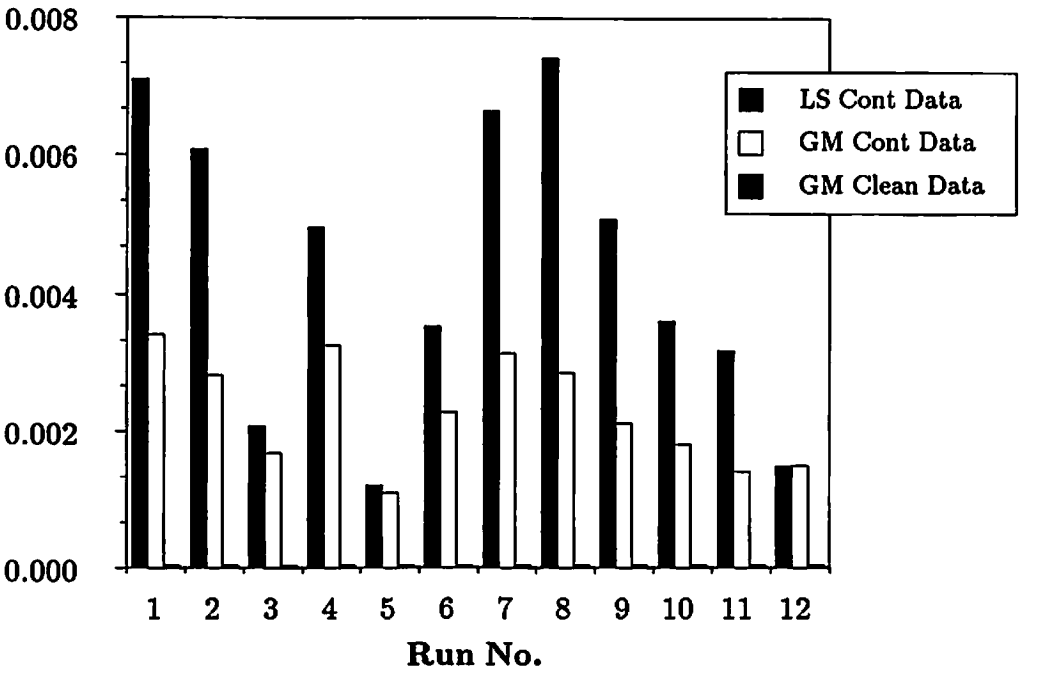


Figure 5.8: Errors for the results shown in Table 5.3.

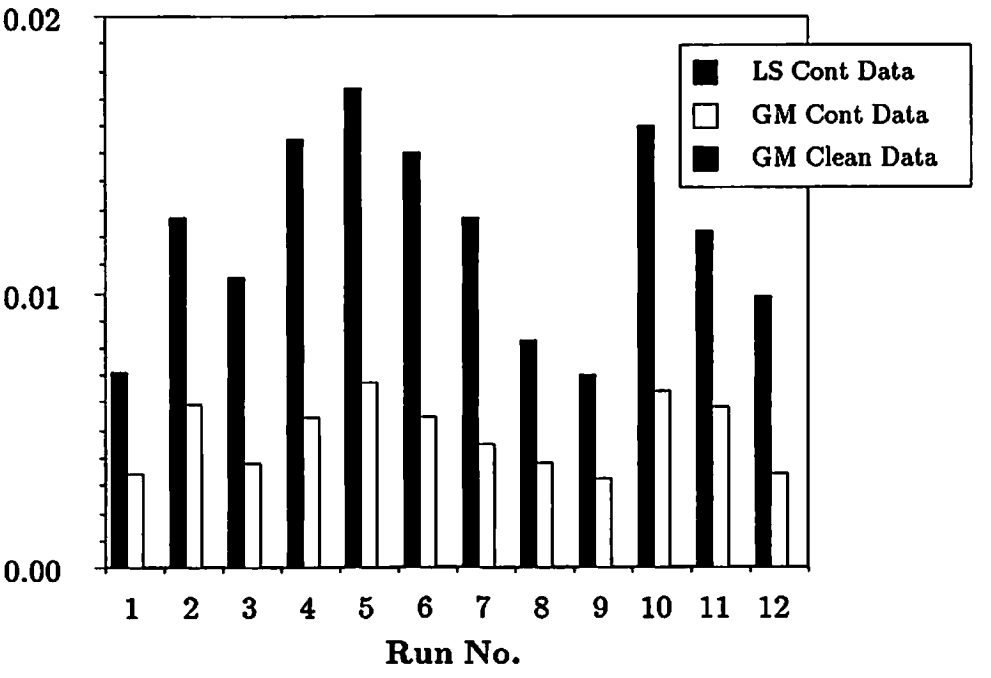


Figure 5.9: Estimation errors for additive outliers in a changing GMRF.

Experiment 3: We evaluated the GM-estimator's performance under different types of contamination, i.e., innovative and substitutive outliers. First we generated 12 sets of 2-D NSHP data using a causal neighbor set and 12 sets of IID Gaussian random noise. The model parameters were are shown in Table 5.4. Next, we simulated innovative outliers by taking the same IID Gaussian random noise fields and adding outliers from a distribution $\mathcal{N}(0, 20.25)$ to 5 percent of the noise data at uniformly spaced lattice sites. The autoregressive data was then regenerated using this driving noise with a heavy-tailed distribution. The ψ-functions and tuning constants were chosen to be same as in Experiment 1. Least squares and GM-estimates of the clean data and contaminated data are listed in Table 5.5 and the squared errors graphed in Figure 5.10.

$\theta_{-1,0}$	$\theta_{0,-1}$	$\theta_{-1,-1}$	ν
.97040	.97350	−.96860	1.00000

Table 5.4: Neighbor set and coefficients for the NSHP model.

Note that for the IO model LS and GM-estimators do equally well in estimating θ in both Gaussian and non-Gaussian situations. However, the GM-estimator outperforms the LS-estimator for estimating the scale β. This is as expected since symmetrically distributed innovative outliers effect only the scale and not the structure of the spectrum, and both the LS and GM-estimator are consistent estimators of θ for the IO model.

In the second part of this experiment we used the 12 sets of 2-D GMRF data used in the second part of Experiment 2. Data contaminated by substitutive outliers were formed by substituting outliers from a distribution $\mathcal{N}(0.20.25)$ for

5 percent of the clean data at uniformly distributed lattice sites. Errors for the LS and GM-estimates were computed and compared with the LS estimates from the clean data. The squared error results are shown in Figure 5.11. Again the GM-estimator outperformed the LS-estimator in the non-Gaussian situation.

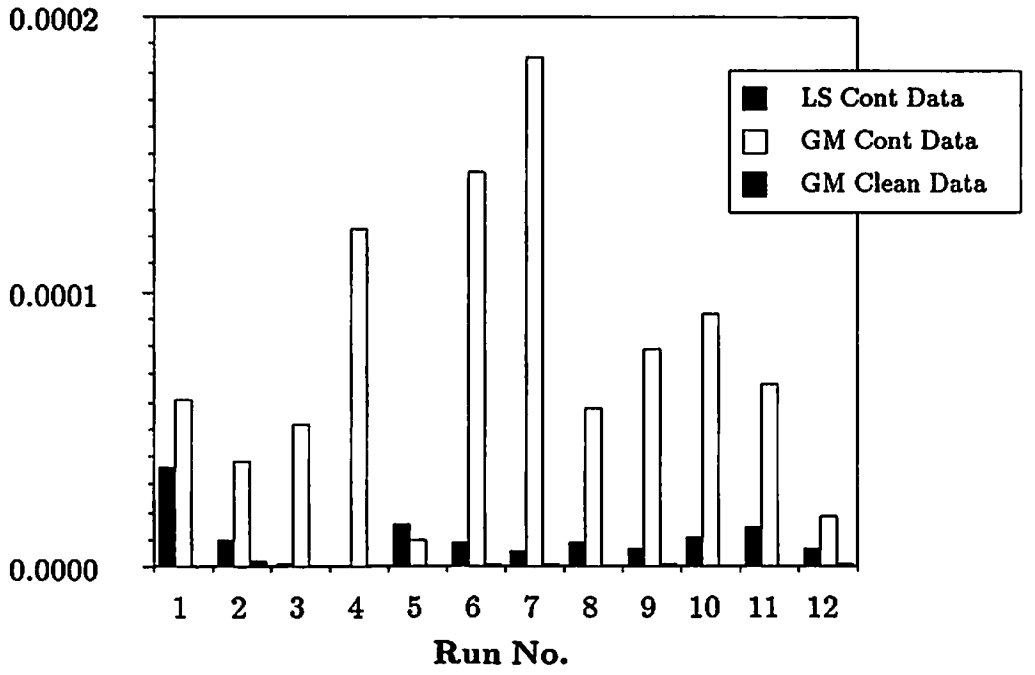


Figure 5.10: Errors for the results shown in Table 5.5.

Run No.	Est Method	$\theta_{(-1,0)}$	$\theta_{(0,-1)}$	$\theta_{(-1,-1)}$	β^2	Sq Err
	Theoretical Value	.970400	.973500	−0.968600	1.00000	
	LS Clean Data	.987543	.990750	−0.989295	1.00068	
1	LS Contaminated Data	.979687	.984730	−0.986289	2.49886	.000036
	GM Contaminated Data	.989580	.989391	−1.002531	1.22117	.000060
	GM Clean Data	.986765	.990741	−0.989043	1.04411	.000000
	LS Clean Data	.978444	.983040	−0.980719	0.98394	
2	LS Contaminated Data	.979591	.988116	−0.978982	1.54501	.000010
	GM Contaminated Data	.980168	.990954	−0.987684	1.13906	.000038
	GM Clean Data	.976672	.983041	−0.978850	1.01879	.000002
	LS Clean Data	.985621	.987905	−0.985635	0.99660	
3	LS Contaminated Data	.984006	.987575	−0.985954	1.62352	.000001
	GM Contaminated Data	.989386	.991061	−0.997082	1.11063	.000052
	GM Clean Data	.985942	.987097	−0.985748	1.00305	.000000
	LS Clean Data	.983659	.987190	−0.982041	1.04070	
4	LS Contaminated Data	.982791	.987659	−0.982092	1.92547	.000000
	GM Contaminated Data	.991139	.994523	−0.998166	1.15460	.000123
	GM Clean Data	.984271	.987338	−0.982968	1.00193	.000000
	LS Clean Data	.985335	.991773	−0.987939	1.03691	
5	LS Contaminated Data	.980742	.988022	−0.984732	1.64153	.000015
	GM Contaminated Data	.984896	.990145	−0.993127	1.25468	.000010
	GM Clean Data	.984787	.992368	−0.988579	1.09527	.000000
	LS Clean Data	.985653	.985634	−0.983150	0.95784	
6	LS Contaminated Data	.987582	.989621	−0.985784	1.77971	.000009
	GM Contaminated Data	.992080	.995766	−1.000052	1.12274	.000143
	GM Clean Data	.985511	.987048	−0.984666	0.97672	.000001
	LS Clean Data	.983893	.986260	−0.984334	0.89570	
7	LS Contaminated Data	.986830	.986725	−0.987343	1.96666	.000006
	GM Contaminated Data	.994321	.992332	−1.004607	1.07080	.000186
	GM Clean Data	.982811	.985403	−0.982872	0.92354	.000001
	LS Clean Data	.986759	.987601	−0.990171	1.00272	
8	LS Contaminated Data	.982307	.986697	−0.987574	2.05374	.000009
	GM Contaminated Data	.989451	.993037	−1.001807	1.21356	.000057
	GM Clean Data	.987400	.987676	−0.990468	1.02488	.000000
	LS Clean Data	.988944	.987261	−0.987261	1.00454	
9	LS Contaminated Data	.991466	.991042	−0.988385	1.95325	.000007
	GM Contaminated Data	.993732	.994287	−1.000144	1.15843	.000075
	GM Clean Data	.989334	.988319	−0.988145	0.97535	.000000
	LS Clean Data	.985354	.986986	−0.985492	0.97958	
10	LS Contaminated Data	.979887	.985451	−0.984810	2.06648	.000011
	GM Contaminated Data	.988242	.991869	−1.001101	1.15539	.000092
	GM Clean Data	.984484	.986933	−0.985738	0.99613	.000000
	LS Clean Data	.985522	.986145	−0.989582	0.97128	
11	LS Contaminated Data	.988188	.991835	−0.987793	1.69048	.000014
	GM Contaminated Data	.990950	.995895	−0.998284	1.10968	.000067
	GM Clean Data	.985292	.986050	−0.989459	0.99013	.000000
	LS Clean Data	.986957	.989961	−0.990111	1.02466	
12	LS Contaminated Data	.984518	.986245	−0.990286	1.55146	.000007
	GM Contaminated Data	.988375	.988803	−0.997312	1.16567	.000018
	GM Clean Data	.985825	.991247	−0.990510	0.99350	.000001

Table 5.5: GM-estimator results for innovative outliers in a changing NSHP field.

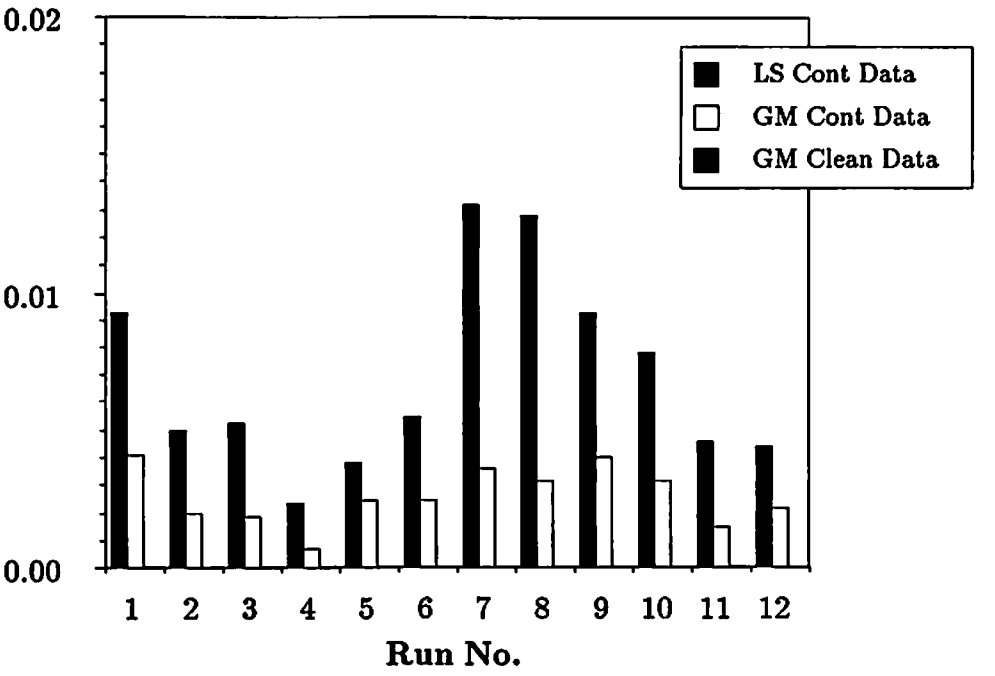


Figure 5.11: Estimation errors for substitutive outliers in a changing GMRF.

Experiment 4: Next we evaluated the GM-estimator when both the $\psi(\cdot)$ and $W(\cdot)$ functions redescend. Here we repeated both parts of Experiment 2 using Tukey's bisquare function (5.22) for both $\psi(\cdot)$ and $g(\cdot)$. The tuning constants were $B = 0.7277$ and $c_v = c = 6.0$ in accordance with Table 5.1. The estimate squared errors are shown in Figure 5.12 and Figure 5.13 for parts one and two of the experiment, respectively. No conclusions can be drawn regarding improvement in estimates of $\boldsymbol{\theta}$ using a redescending ψ-function.

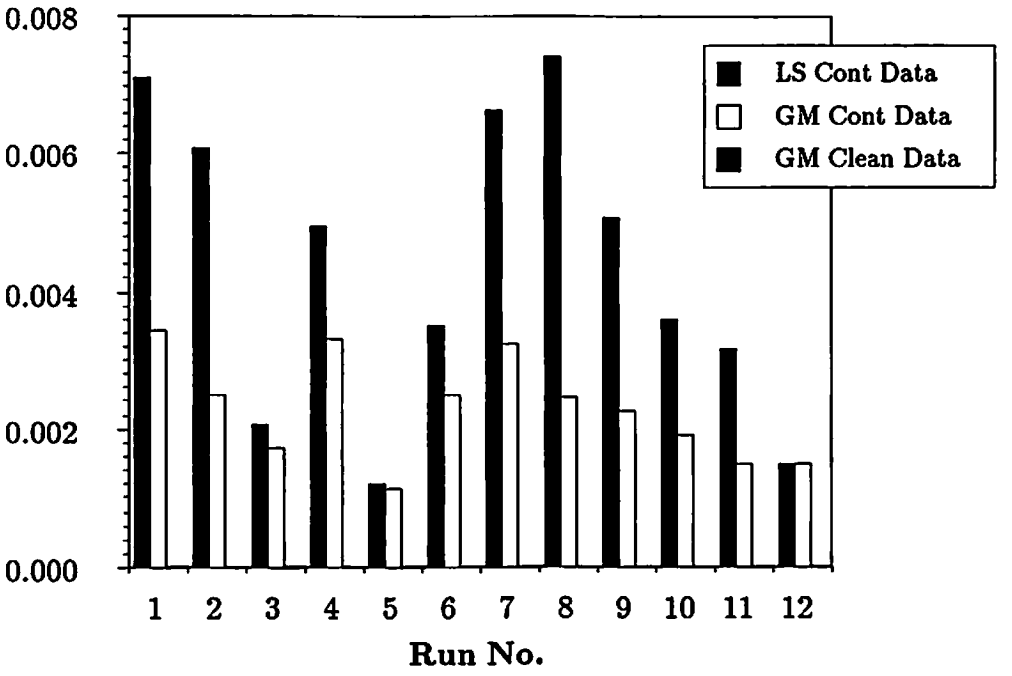


Figure 5.12: Estimation errors for additive outliers in a GMRF with changing contamination and redescending ψ_B-function.

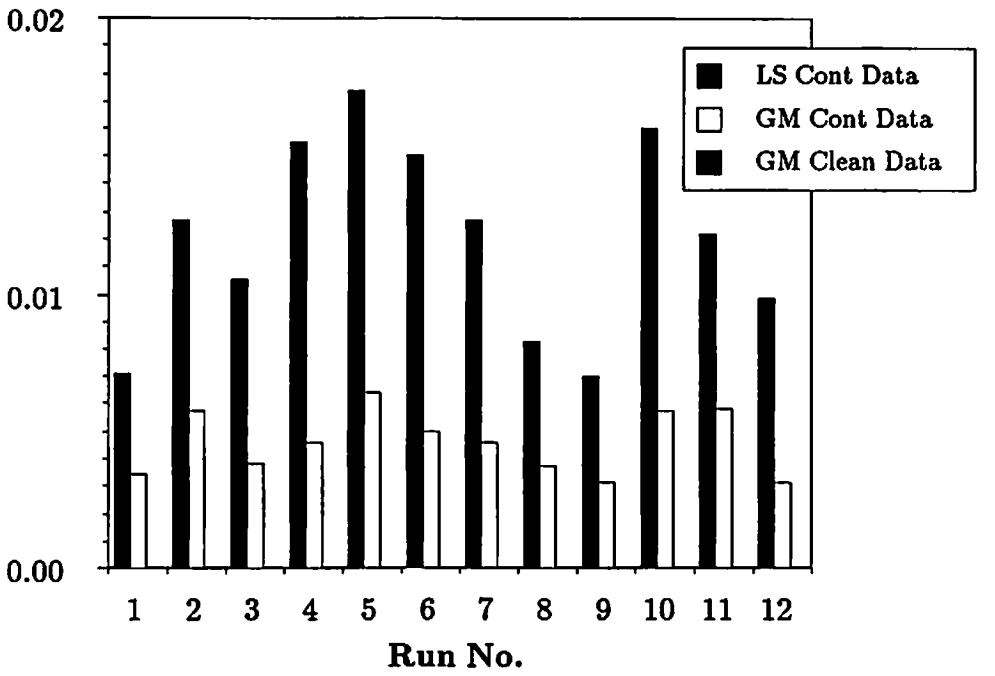


Figure 5.13: Estimation errors for additive outliers in a changing GMRF with redescending ψ_B-function.

Summary. Tables 5.6 and 5.7 summarize the results of Experiments 2 , 3 and 4. Table 5.6 shows the average of the squared errors for the coefficients $\{\theta_{\mathbf{r}},\ \mathbf{r} \in N\}$ estimates, and Table 5.7 shows the average of the absolute values of the errors for the residual's variance β^2 estimates. All errors are computed relative to the LS estimates of the clean data.

The GM-estimator yields estimates from the contaminated data which are closer to the true (LS Clean Data) values than the non-robust LS estimator in both the additive and substitutive outlier cases. This is true for both the coefficients and scale estimates. For the innovative outlier case the LS and generalized M-estimators do equally well for the coefficients, but the robust procedure does much better for the variance estimates. This is as expected since innovative outliers (from symmetric distributions) have little effect on the shape of the spectrum (see Section 1.2.2) but will effect scale estimates. Too few experiments were run to draw any significant conclusions regarding the use of redescending ψ-functions.

Estimation Method	GMRF/AO			NSHP/IO	GMRF/SO
	Exp. 2.1 ψ_H	Exp. 4.1 ψ_B	Exp. 2.2 ψ_H	Exp. 3.1 ψ_H	Exp. 3.2 ψ_H
LS Contaminated Data	.004358	.004358	.012020	.000010	.006959
GM Contaminated Data	.002271	.002282	.004832	.000077	.002581
GM Clean Data	.000028	.000018	.000017	.000000	.000017

Table 5.6: Estimated coefficient error summary for the GM-estimator. Entries are the averages of the squared errors from each run.

Estimation Method	GMRF/AO			NSHP/IO	GMRF/SO
	Exp. 2.1 ψ_H	Exp. 4.1 ψ_B	Exp. 2.2 ψ_H	Exp. 3.1 ψ_H	Exp. 3.2 ψ_H
LS Contaminated Data	0.95007	.95007	1.50748	0.86675	0.90353
GM Contaminated Data	0.19022	.11854	0.26068	0.16510	0.18419
GM Clean Data	0.02894	.01949	0.02519	0.02917	0.02519

Table 5.7: Estimated scale error summary for the GM-estimator. Entries are the averages of the absolute value of the errors from each run.

Chapter 6

2-D Optimal Robust Estimators

The influence function has been proposed on strictly heuristic grounds as a tool for analysis and synthesis of robust location and scale estimators of independent observations [38]. The importance of the influence function for estimators in the IID case lies in its heuristic interpretation: it describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of the contamination [38]. For independent data the location of an additional observation is unimportant since data are processed without consideration for temporal or spatial ordering. Hence, the distribution $\mathcal{F}^h = (1-h)\mathcal{F} + h\Delta(x)$ used in defining the influence function, where $\Delta(x)$ is the degenerate distribution with all of its mass at x, is satisfactory for describing the contamination.

For random fields the contamination model $\mathcal{F}^{N_0,h} = (1-h)\mathcal{F}^{N_0} + h\Delta(\mathbf{x}^0)$, where $\mathcal{F}^{N_0}$ is a $(p+1)$-dimensional distribution and h is the fraction of the data which is contaminated, used in Chapter 5 provides a tool for gauging the performance of the heuristic 2-D GM-estimator. In general, this distribution is not

satisfactory for data with contamination more complex than additive outliers occurring at widely separated and uncorrelated lattice points. For example, the same value of h is obtained for two data sets with 5 percent contamination, one set having additive outliers at widely separated lattice points and other with additive outliers occuring in strings or small patches. In other words, the value of $\mathbf{x}$ matters and also where $\mathbf{x}$ is placed.

For 1-D autoregressions having general contamination, Künsch [69] has proposed an influence function for a class of estimators depending only on a finite p-dimensional marginal distribution denoted by $\mathcal{F}^p$. This influence function is unique up to an equivalence relation, but there is a unique version which has the interpretation as the influence of an additional observation on the estimate. The unique version depends upon a conditioning on past observations–namely, the influence of an additional observation should have conditional expectation zero given the past, which is faithful to the heuristic interpretation attached by Hampel [38] to the influence function for independent data. For 1-D dependent data sets the additional observation can be placed at the end of the observations, motivating Künsch's use of the conditional expectation given past values to isolate a unique influence function.

In this chapter we take a similar approach to the specification, design, and analysis of a 2-D random field model robust parameter estimator. First, we generalize the formulation of the ML equation for spatial interaction models to create a general class of ML estimators. Then, using the new formulation we specify an optimality problem which we solve to yield a new robust estimator for NSHP models.

6.1 The ML Estimator for Spatial Interaction Models

Let $\mathcal{X} = \{x(\mathbf{s}), \mathbf{s} \in \Omega\}$ be a 2-D field of random variables obeying the spatial interaction model

$$x(\mathbf{s}) = \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} x(\mathbf{s}+\mathbf{r}) + \zeta(\mathbf{s}) \tag{6.1}$$

where $\zeta(\mathbf{s})$ is an identically distributed, stationary, possibly correlated, Gaussian random noise array with variance $V\{\zeta\}$. Thus, both the simultaneous autoregressive and Gaussian-Markov models are included in this description of $\mathcal{X}$.

When N in (6.1) is symmetric, i.e., $\mathbf{r} \in N$ implies $-\mathbf{r} \in N$, (6.1) can be written in terms of the asymmetric set N_s where $\mathbf{r} \in N_s$ implies $-\mathbf{r} \notin N_s$. In this representation (6.1) is

$$x(\mathbf{s}) = \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} [x(\mathbf{s}+\mathbf{r}) + x(\mathbf{s}-\mathbf{r})] + \zeta(\mathbf{s}). \tag{6.2}$$

We shall use this representation for the GMRF model in the development below, but shall retain the representation (6.1) for the NCAR model since this allows us to include causal NSHP models as a subset of the NCAR models. Thus, it should be clear that $\boldsymbol{\theta} = \text{col}\ \{\theta_{\mathbf{r}}, \mathbf{r} \in N_s\}$ and $\mathbf{x}_{\mathbf{s}} = \text{col}\ \{x(\mathbf{s}+\mathbf{r}) + x(\mathbf{s}-\mathbf{r}), \mathbf{r} \in N_s\}$ are p-dimensional vectors for the GMRF model and $\boldsymbol{\theta} = \text{col}\ \{\theta_{\mathbf{r}}, \mathbf{r} \in N\}$ and $\mathbf{x}_{\mathbf{s}} = \text{col}\ \{x(\mathbf{s}+\mathbf{r}), \mathbf{r} \in N\}$ are p-dimensional vectors for the NCAR model where p is the number of elemets in N_s or N.

The negative of the log of the likelihood function for the random field $\mathcal{X}_M$ with respect to the $(p+1)$-dimensional parameter vector $\boldsymbol{\phi} = \text{col}\ \{\boldsymbol{\theta}, \text{V}\ \{\zeta\}\}$ given observations from the finite lattice Ω_M is asymptotically equivalent to

$$l_M(\boldsymbol{\phi}) = \frac{M^2}{2}\log(2\pi) + \frac{M^2}{2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \log[S_x(\omega, \boldsymbol{\phi})]d\omega + \frac{1}{2}\mathbf{x}^t Q_{\mathbf{x}}^{-1}\mathbf{x}.$$

Here, for the GMRF model

$$S_x(\omega,\phi) = \frac{\nu}{1 - 2\theta^t \mathbf{c}_\omega},$$

where $\nu = V(\zeta)$ and $\mathbf{c}_\omega = \text{col}\ \{\cos(\omega^t \mathbf{r}), \mathbf{r} \in N_s\}$. For the NCAR model

$$S_x(\omega,\phi) = \frac{\beta^2}{|1 - \theta^t \psi_\omega)|^2},$$

where $\beta^2 = V(\zeta)$ and $\psi_\omega = \text{col}\ \{\exp(-\jmath\omega^t \mathbf{r}), \mathbf{r} \in N\}$.

Denoting the derivative of $l_M(\phi)$ with respect to the parameters ϕ by the $(p+1)$-dimensional vector $\lambda_M(\phi)$ and using the equivalences (3.19) and (3.20) given in Chapter 3 for the quadratic term $\mathbf{x}^t Q_{\mathbf{x}}^{-1} \mathbf{x}$, the derivative is written as

$$\lambda_M(\phi) = \sum_{\mathbf{s} \in \Omega_M} [\kappa(\mathbf{x}_\mathbf{s}^0; \phi) - \mathrm{h}(\phi)],$$

where for the GMRF model

$$\kappa(\mathbf{x}_\mathbf{s}^0; \phi) = \frac{\partial}{\partial \phi} \left\{ -\frac{1}{2\nu} x(\mathbf{s}) [x(\mathbf{s}) - \theta^t \mathbf{x}_\mathbf{s}] \right\} \tag{6.3}$$

$$\mathrm{h}(\phi) = \frac{1}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \phi} \log[S_x(\omega,\phi)] d\omega, \tag{6.4}$$

and for the NCAR model

$$\kappa(\mathbf{x}_\mathbf{s}^0; \phi) = \frac{\partial}{\partial \phi} \left\{ -\frac{1}{2\beta^2} [x(\mathbf{s}) - \theta^t \mathbf{x}_\mathbf{s}]^2 \right\} \tag{6.5}$$

$$\mathrm{h}(\phi) = \frac{1}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \phi} \log[S_x(\omega,\phi)] d\omega. \tag{6.6}$$

Letting subscripts denote the derivatives with respect to the specific parameters, (6.3) and (6.4) become

$$\kappa_\theta(\mathbf{x}_\mathbf{s}^0; \phi) = \frac{1}{2\nu} \mathbf{x}_\mathbf{s} x(\mathbf{s}) \tag{6.7}$$

$$\kappa_\nu(\mathbf{x}_\mathbf{s}^0; \phi) = \frac{1}{2\nu^2} x(\mathbf{s}) [x(\mathbf{s}) - \theta^t \mathbf{x}_\mathbf{s}] \tag{6.8}$$

$$\mathrm{h}_\theta(\phi) = \frac{1}{\nu(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} S_x(\omega,\phi) \mathbf{c}_\omega d\omega \tag{6.9}$$

$$h_\nu(\phi) = \frac{1}{2\nu} \tag{6.10}$$

for the GMRF model, where $\mathbf{c}_\omega = \text{col}\ \{\cos(\omega^t \mathbf{r}), \mathbf{r} \in N_s\}$; and (6.5) and (6.6) become

$$\kappa_\theta(\mathbf{x}_s^0; \phi) = \frac{1}{\beta^2}\mathbf{x}_s[x(\mathbf{s}) - \theta^t \mathbf{x}_s] \tag{6.11}$$

$$\kappa_\beta(\mathbf{x}_s^0; \phi) = \frac{1}{\beta^3}[x(\mathbf{s}) - \theta^t \mathbf{x}_s]^2 \tag{6.12}$$

$$\mathbf{h}_\theta(\phi) = \frac{1}{\beta^2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega, \phi)[\mathbf{c}_\omega - C_\omega \theta]d\omega \tag{6.13}$$

$$h_\beta(\phi) = \frac{1}{\beta} \tag{6.14}$$

for the NCAR model, where $\mathbf{c}_\omega = \text{col}\ \{\cos(\omega^t \mathbf{r}), \mathbf{r} \in N\}$ and $C_\omega = \text{mat}\ \{\cos[\omega^t(\mathbf{r}-\mathbf{t})], \mathbf{r}\&\mathbf{t} \in N\}$.

We conclude from the asymptotic properties of the ML estimator (see Chapter 3) that ϕ_{ML} is a solution of the equation

$$\lambda_M(\phi) = \mathbf{0}. \tag{6.15}$$

Moreover, as $M \to \infty$ this estimate for both the NCAR and GMRF model is unique and distributed such that $M[\phi_{ML} - \phi_0]$ is $\mathcal{N}(\mathbf{0}, \Gamma^{-1}(\phi_0))$ where $\Gamma(\phi_0)$ is the asymptotic value of the Fisher information matrix $I_M(\phi)$ defined by

$$I_M(\phi) = \frac{1}{M^2}\mathrm{E}\left\{\left[\frac{\partial}{\partial \phi}\log f(\mathbf{x}|\phi)\right]\left[\frac{\partial}{\partial \phi}\log f(\mathbf{x}|\phi)\right]^t\right\} \tag{6.16}$$

and evaluated at the true value of the parameters ϕ_0. The asymptotic value is easily shown to be

$$\Gamma(\phi) = \frac{1}{2(\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\frac{\partial}{\partial \phi}\log \mathcal{S}_x(\omega, \phi)\right]\left[\frac{\partial}{\partial \phi}\log \mathcal{S}_x(\omega, \phi)\right]^t d\omega. \tag{6.17}$$

Evaluation of $\lim_{M\to\infty} I_M(\phi)$ or $\Gamma(\phi)$ for the GMRF model gives

$$\Gamma(\theta) = \frac{2}{\nu^2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathcal{S}_x^2(\omega, \phi)\mathbf{c}_\omega \mathbf{c}_\omega^t d\omega$$

$$\gamma(\nu) = \frac{1}{2\nu^2}$$

$$\gamma(\theta, \nu) = \frac{1}{\nu^2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\mathcal{S}_x(\omega, \phi)\mathbf{c}_\omega d\omega$$

where

$$\Gamma(\phi) = \begin{bmatrix} \Gamma(\theta) & \gamma(\theta,\nu) \\ \gamma^t(\theta,\nu) & \gamma(\nu) \end{bmatrix}.$$

A similar evaluation for the NCAR model gives

$$\begin{aligned} \Gamma(\theta) &= \frac{2}{\beta^4(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x^2(\omega,\phi)[\mathbf{c}_\omega - C_\omega\theta][\mathbf{c}_\omega - C_\omega\theta]^t d\omega \\ \gamma(\beta) &= \frac{2}{\beta^2} \\ \gamma(\theta,\beta) &= \frac{2}{\beta^3(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi)[\mathbf{c}_\omega - C_\omega\theta]d\omega. \end{aligned}$$

When $\mathcal{X}$ is modeled by a NSHP model, the limiting Fisher information matrix has the special form

$$\begin{aligned} \Gamma(\theta) &= \frac{1}{\beta^2}Q_{\mathbf{x}} \\ \gamma(\beta) &= \frac{2}{\beta^2} \\ \gamma(\theta,\beta) &= \mathbf{0}, \end{aligned}$$

where $Q_{\mathbf{x}} = \text{mat}\,\{\mathrm{E}\{x(\mathbf{s}+\mathbf{r})x(\mathbf{s}+\mathbf{t})\}, \mathbf{r}\&\mathbf{t} \in N\}$. This results from the fact that $\mathbf{q}_{\mathbf{x}} - Q_{\mathbf{x}}\theta = \mathbf{0}$ if $\mathbf{q}_{\mathbf{x}} = \text{col}\,\{\mathrm{E}\{x(\mathbf{s})x(\mathbf{s}+\mathbf{r})\}, \mathbf{r} \in N\}$; and hence, $\mathbf{h}_\theta(\phi) = \mathbf{0}$. Therefore, with $h(\phi) \doteq 0$ the derivative of the log likelihood is

$$\boldsymbol{\lambda}_M(\phi) = \sum_{\mathbf{s}\in\Omega_M} \boldsymbol{\kappa}(\mathbf{x}_{\mathbf{s}}^0;\phi)$$

and where now

$$\kappa_\theta(\mathbf{x}_{\mathbf{s}};\phi) = \frac{1}{\beta^2}\mathbf{x}_{\mathbf{s}}[x(\mathbf{s}) - \theta^t\mathbf{x}_{\mathbf{s}}] \tag{6.18}$$

$$\kappa_\beta(\mathbf{x}_{\mathbf{s}};\phi) = \frac{1}{\beta}\left\{\frac{1}{\beta^2}[x(\mathbf{s}) - \theta^t\mathbf{x}_{\mathbf{s}}]^2 - 1\right\}. \tag{6.19}$$

Consequently, a solution to (6.15) is easily computed by

$$\hat{\theta}_{LS} = \left[\sum_{\mathbf{s}\in\Omega_M}\mathbf{x}_{\mathbf{s}}\mathbf{x}_{\mathbf{s}}^t\right]^{-1}\left[\sum_{\mathbf{s}\in\Omega_M}\mathbf{x}_{\mathbf{s}}x(\mathbf{s})\right] \tag{6.20}$$

and

$$\hat{\beta}_{LS}^2 = \frac{1}{M^2} \sum_{\mathbf{s} \in \Omega_M} [x(\mathbf{s}) - \boldsymbol{\theta}_{LS}^t \mathbf{x}_\mathbf{s}]^2, \tag{6.21}$$

which are the classic LS estimators.

We offer the following proposition which will be useful in latter discussions.

Proposition 6.1 *For the NSHP model the limiting Fisher information matrix,* $\lim_{M \to \infty} I_M(\boldsymbol{\phi})$, *is*

$$\Gamma(\boldsymbol{\phi}) = \mathrm{E}\left\{\boldsymbol{\kappa}(\mathbf{x}^0; \boldsymbol{\phi}) \boldsymbol{\kappa}^t(\mathbf{x}^0; \boldsymbol{\phi})\right\}. \tag{6.22}$$

The proof of this proposition is as follows. First, in view of the results (3.29) and (3.30), the conditional mean of $\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi})$ given all past observations $\{x(\mathbf{u}); \mathbf{u} < \mathbf{s}\}$ is $\mathrm{E}\{\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) | x(\mathbf{u}), \mathbf{u} < \mathbf{s}\} = \mathrm{E}\{\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) | \mathbf{x}_\mathbf{s}\} = \mathbf{0}$. Next, suppose $\mathbf{t} < \mathbf{s}$. Then

$$\begin{aligned} &\mathrm{E}\left\{\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) \boldsymbol{\kappa}^t(\mathbf{x}_\mathbf{t}^0; \boldsymbol{\phi})\right\} \\ &\quad = \int\int \boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) \boldsymbol{\kappa}^t(\mathbf{x}_\mathbf{t}^0; \boldsymbol{\phi}) f(x(\mathbf{s}) | x(\mathbf{u}), \mathbf{u} < \mathbf{s}) f(x(\mathbf{u}), \mathbf{u} < \mathbf{s}) dx(\mathbf{s}) d\mathbf{x}_\mathbf{u}, \end{aligned}$$

where $\mathbf{x}_\mathbf{u} = \mathrm{col}\ \{x(\mathbf{u}), \mathbf{u} < \mathbf{s}\}$; and since $\mathbf{x}_\mathbf{t}^0$ does not include $x(\mathbf{s})$ we conclude that

$$\begin{aligned} \mathrm{E}\left\{\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) \boldsymbol{\kappa}^t(\mathbf{x}_\mathbf{t}^0; \boldsymbol{\phi})\right\} &= \int \mathrm{E}\left\{\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) | \mathbf{x}_\mathbf{s}\right\} \boldsymbol{\kappa}^t(\mathbf{x}_\mathbf{t}^0; \boldsymbol{\phi}) f(x(\mathbf{u}), \mathbf{u} < \mathbf{s}) d\mathbf{x}_\mathbf{u} \\ &= 0. \end{aligned}$$

If $\mathbf{t} < \mathbf{s}$, we draw the same conclusion. Finally, the NSHP Fisher information matrix is

$$\begin{aligned} I_M(\boldsymbol{\phi}) &= \frac{1}{M^2} \mathrm{E}\left\{\boldsymbol{\lambda}_M(\boldsymbol{\phi}) \boldsymbol{\lambda}_M^t(\boldsymbol{\phi})\right\} \\ &= \frac{1}{M^2} \sum_{\mathbf{s} \in \Omega_M} \sum_{\mathbf{t} \in \Omega_M} \mathrm{E}\left\{\boldsymbol{\kappa}(\mathbf{x}_\mathbf{s}^0; \boldsymbol{\phi}) \boldsymbol{\kappa}^t(\mathbf{x}_\mathbf{t}^0; \boldsymbol{\phi})\right\} \end{aligned}$$

$$= \frac{1}{M^2} \sum_{s \in \Omega_M} \mathrm{E}\left\{\kappa(\mathbf{x}_s^0; \phi)\kappa^t(\mathbf{x}_s^0; \phi)\right\}$$
$$= \mathrm{E}\left\{\kappa(\mathbf{x}^0; \phi)\kappa^t(\mathbf{x}^0; \phi)\right\}.$$

Thus, the proposition is proven. Note, that for the bilateral NCAR and GMRF models, which contain symmetric neighbors, this proposition is generally not true.

6.2 A General Estimator for Spatial Interaction Models

We can define a more general class of estimators which includes the estimators defined by (6.15). Consider an observation set $\mathcal{X}_M$ and let D be an arbitrary neighbor set such that $N_0 \subseteq D$ and D may include $(0,0)$. Let d be the size of D. Thus, $d \geq p+1$. The empirical d-dimensional marginal distribution

$$\mathcal{F}_M^D(x) = \frac{1}{M^2} \sum_{s \in \Omega_M} \Delta(\mathbf{x}_s^d), \tag{6.23}$$

where $\Delta(\mathbf{x}_s^d)$ is the distribution having all of its mass at $\mathbf{x}_s^d = \mathrm{col}\,\{x(\mathbf{s}+\mathbf{r}), \mathbf{r} \in D\}$. Then a general estimator $\hat{\phi}_M = T(\mathcal{F}_M^D)$ is defined as any functional $T(\cdot)$ of the empirical marginal distribution where

$$T(\mathcal{F}^D) = \phi \text{ if and only if } \int \psi(\mathbf{x}^d; \phi) d\mathcal{F}^D = \mathbf{0}, \tag{6.24}$$

and ψ is chosen to provide "good" estimates, consistent at the true underlying d-dimensional marginal distribution $\mathcal{F}^D(x)$. For example, if $D = N_0$ then substituting the empirical distribution (6.23) into (6.24) and choosing $\psi(\mathbf{x}^d; \phi) = \kappa(\mathbf{x}^0; \phi) - \mathbf{h}(\phi)$ yields (6.15).

6.3 A General Contamination for Lattice Data

In Chapter 5, Section 5.1.2 we derived an influence function for the GM-estimator using the $(p+1)$-dimensional mixture distribution $\mathcal{F}^{N_0,h}$, which may not have any direct relationship with the contamination of the observed data [38]. Consequently, the GM-estimator influence function may not generate a good approximation for the estimator at the distribution of generally contaminated data, for example, data that includes outliers more complex than occurring only at widely separated points on the lattice. However, the $(p+1)$-dimensional marginal distribution $\mathcal{F}^{N_0}(x)$ of the observed data in the more complex case is not a simple mixture of the $(p+1)$-dimensional marginal distribution of the clean data and some contamination; because, unlike in the independent data case, the location of the contamination is also important. A more representative model of the observed data which includes the additive and innovative outlier models as subsets. We then give an approximation for the d-dimensional marginal distribution $\mathcal{F}^D(x)$ of the observed data.

Let the random field data $\mathcal{X} = \{x(\mathbf{s}), \mathbf{s} \in \Omega$ be modeled as

$$x(\mathbf{s}) = [1 - z(\mathbf{s})]y(\mathbf{s}) + z(\mathbf{s})n(\mathbf{s}) \tag{6.25}$$

where $\mathcal{Y} = \{y(\mathbf{s}), \mathbf{s} \in \Omega\}$ is a stationary random field (NCAR or GMRF) with common distribution $\mathcal{F}(y)$, $\mathcal{N} = \{n(\mathbf{s}), \mathbf{s} \in \Omega\}$ is a random noise field, and $z(\mathbf{s})$ is a random variable whose range is on $[0,1]$. If $z(\mathbf{s}) = z$ is a constant for all $\mathbf{s}$ and $n(\mathbf{s}) = y(\mathbf{s}) + v(\mathbf{s})$, then (6.25) becomes the signal plus noise model with additive noise $n(\mathbf{s}) \cdot v(\mathbf{s})$. If $z > 0$, then the AO model holds with a suitable distribution $\mathcal{F}(v)$. In the latter case patchy outliers occur when $v(\mathbf{s})$ is dependent on $y(\mathbf{s})$, otherwise isolated outliers occur. On the otherhand, if $z(\mathbf{s})$ is a 0-1 process with

$\Pr[z(\mathbf{s}) = 1] = h$, then we have substitutive outliers, i.e., we observe $y(\mathbf{s})$ perfectly $100(1-h)$ percent of the time while the other $100h$ percent of the time we observe only noise. Clearly, this model includes many types of contamination.

The determination of a common distribution $\mathcal{F}(x)$ for the $x(\mathbf{s})$ in (6.25) is generally a straight forward procedure when the distributions $\mathcal{F}(z)$ and $\mathcal{F}(n)$ are known. With more effort one could also derive the d-dimensional marginal distribution. But in the spirit of robustness we do not have exact knowledge of the corrupting processes, nor are we assured that the data follows (6.25). This leads us to consider an approximation to the marginal distribution of the observed data, that is, we will approximate the true marginal distribution $\mathcal{F}^{D,h}(x)$ of the observed data $x(\mathbf{s})$ by

$$\tilde{\mathcal{F}}^{D,h}(x) = (1-h)\mathcal{F}^{D}(x) + h\mathcal{G}^{D}(x) \tag{6.26}$$

where $\mathcal{G}^{D}(x)$ is a d-dimensional distribution depending on the $y(\mathbf{s})$, $z(\mathbf{s})$, and $n(\mathbf{s})$. (See [38], [69], and [68]).

6.4 Asymptotic Properties of the General Estimator

The typical approach to designing robust estimators by the infinitesimal method [38] is to compute the asymptotic bias and error variance via an influence function which can be interpreted as the influence of an additional observation on the value of the estimate. For estimators given by (6.24) the design procedure then involves selecting a ψ-function that yields the smallest error variance subject to a bound on the bias. In this section we proceed in this manner to design a robust estimator of the parameters ϕ for the models (6.1) and (6.2). However, as we shall see, a

simple expression, tractable to a solution, requires a simplifying assumption not applicable in general to 2-D models having bilateral neighbor sets. This is because these models have correlations which extend in all directions. The unilateral 2-D model, on the other hand, has statistical properties similar to 1-D time series, i.e., a one-sided Markov property. This property can be exploited to identify a unique influence function leading to a simple expression for the asymptotic error variance. In this case the analysis is similar to [69] where robust estimators for causal autoregressive time series are proposed. We initially carry out the analysis without restricting the neighbor set; and proofs of results which are analogous to the 1-D situation are not provided since they are simple extensions to those contained in [69].

6.4.1 Influence Function

To simplify notation in the following analysis we drop the superscript "d" on the data vector $\mathbf{x}$. The dimension of $\mathbf{x}$ is easily determined from the context in which it appears, and unless otherwise noted $\mathbf{x_s} = \text{col}\,\{x(\mathbf{s}+\mathbf{r}), \mathbf{r} \in D\}$. To derive an influence function for the estimator (6.24) consider the functional $T(\cdot)$ evaluated at the distribution $\tilde{\mathcal{F}}^{D,h}$ given by (6.26), i.e.,

$$\int \psi(\mathbf{x}; T(\tilde{\mathcal{F}}^{D,h}))d\tilde{\mathcal{F}}^{D,h} = \mathbf{0}. \tag{6.27}$$

Taking the ordinary derivative of (6.27) with respect to h and evaluating at $h = 0$, we obtain

$$\begin{aligned} \mathbf{0} &= \frac{\partial}{\partial h}\int \psi(\mathbf{x}; T(\tilde{\mathcal{F}}^{D,h}))d\tilde{\mathcal{F}}^{D,h}\bigg|_{h=0} \\ &= \int \frac{\partial}{\partial h}\psi(\mathbf{x}; T(\tilde{\mathcal{F}}^{D,h}))\bigg|_{h=0} d\mathcal{F}^{D} + \int \psi(\mathbf{x};\phi)d\left[\frac{\partial}{\partial h}\tilde{\mathcal{F}}^{D,h}\right]_{h=0} \end{aligned}$$

$$= \int \frac{\partial}{\partial \phi}\psi(\mathbf{x};\phi)\left[\frac{\partial}{\partial h}T(\tilde{\mathcal{F}}^{D,h})\right]_{h=0} d\mathcal{F}^D + \int \psi(\mathbf{x};\phi)d\left[\frac{\partial}{\partial h}\tilde{\mathcal{F}}^{D,h}\right]_{h=0}$$

or

$$\left[\frac{\partial}{\partial h}T(\tilde{\mathcal{F}}^{D,h})\right]_{h=0}\int\left[\frac{\partial}{\partial \phi}\psi(\mathbf{x};\phi)\right]d\mathcal{F}^D = -\int \psi(\mathbf{x};\phi)d(\mathcal{G}^D - \mathcal{F}^D)$$

Let $\mathcal{G}^D$ be a d-dimensional distribution which is "near" $\mathcal{F}^D$. The *von Mises derivative* of $T(\cdot)$ at the distribution $\mathcal{F}^D$ [74] is defined to be

$$\begin{aligned} T'(\phi;\mathcal{G}^D - \mathcal{F}^D) &= \frac{\partial}{\partial h}T(\tilde{\mathcal{F}}^{D,h})\Big|_{h=0} \\ &= \lim_{h\to 0}\frac{1}{h}[T((1-h)\mathcal{F}^D + h\mathcal{G}^D) - T(\mathcal{F}^D)], \end{aligned}$$

if there exists a real valued function $t(\mathbf{x},\mathcal{F}^D)$ independent of $\mathcal{G}^D$ such that

$$T'(\phi;\mathcal{G}^D - \mathcal{F}^D) = \int t(\mathbf{x},\mathcal{F}^D)d(\mathcal{G}^D - \mathcal{F}^D). \tag{6.28}$$

Clearly, the von Mises derivative of $T(\cdot)$ is

$$T'(\phi;\mathcal{G}^D - \mathcal{F}^D) = M^{-1}(\psi,\mathcal{F}^D)\int \psi(\mathbf{x};\phi)d\mathcal{G}^D$$

where

$$M(\psi,\mathcal{F}^D) = -\int\left[\frac{\partial}{\partial \phi}\psi(\mathbf{x};\phi)\right]d\mathcal{F}^D. \tag{6.29}$$

Another expression for the matrix $M(\psi,\mathcal{F}^D)$ is obtained by differentiating the defining equation (6.24) with respect to ϕ as follows:

$$\frac{\partial}{\partial \phi}\int \psi(\mathbf{x};\phi)d\mathcal{F}^D = 0_d,$$

$$\int\left[\frac{\partial}{\partial \phi}\psi(\mathbf{x};\phi)\right]d\mathcal{F}^D + \int \psi(\mathbf{x};\phi)d\left[\frac{\partial}{\partial \phi}\mathcal{F}^D\right]^t = 0_d,$$

or

$$\begin{aligned} \int\left[\frac{\partial}{\partial \phi}\psi(\mathbf{x};\phi)\right]d\mathcal{F}^D &= -\int \psi(\mathbf{x};\phi)\left[\frac{\partial}{\partial \phi}f^D(\mathbf{x})\right]^t d\mathbf{x} \\ &= -\int \psi(\mathbf{x};\phi)\left[\frac{\partial}{\partial \phi}\log(f^D(\mathbf{x}))\right]^t f^D(\mathbf{x})d\mathbf{x} \\ &= -\int \psi(\mathbf{x};\phi)\left[\boldsymbol{\lambda}^d(\mathbf{x};\phi)\right]^t d\mathcal{F}^D, \end{aligned}$$

where $\boldsymbol{\lambda}^d(\mathbf{x};\boldsymbol{\phi}) = \dfrac{\partial}{\partial\boldsymbol{\phi}}\log(f^D(\mathbf{x}))$, $f^D(\mathbf{x})$ is the joint probability density function associated with $\mathcal{F}^D$, and 0_d is the $d \times d$ matrix of zeroes. Comparing this last equation with (6.29) we see that

$$M(\boldsymbol{\psi},\mathcal{F}^D) = \int \boldsymbol{\psi}(\mathbf{x};\boldsymbol{\phi})\left[\boldsymbol{\lambda}^d(\mathbf{x};\boldsymbol{\phi})\right]^t d\mathcal{F}^D. \tag{6.30}$$

Any function $IF(\mathbf{x};\boldsymbol{\phi})$ such that

$$T'(\boldsymbol{\phi},\mathcal{G}^D - \mathcal{F}^D) = \int IF(\mathbf{x};\boldsymbol{\phi})d\mathcal{G}^D$$

will be called an *influence function* for the functional $T(\cdot)$. Thus, an influence function for the general estimator defined by (6.24) is

$$IF(\mathbf{x};\boldsymbol{\phi}) = M^{-1}(\boldsymbol{\psi},\mathcal{F}^D)\boldsymbol{\psi}(\mathbf{x};\boldsymbol{\phi}). \tag{6.31}$$

Note that

$$\int IF(\mathbf{x};\boldsymbol{\phi})d\mathcal{F}^D = \mathbf{0} \tag{6.32}$$

and that the influence function is determined by the $\boldsymbol{\psi}$-function alone.

In particular, an influence function for the NCAR and GMRF models ML estimator, where $\boldsymbol{\psi}(\mathbf{x}^d;\boldsymbol{\phi}) = \boldsymbol{\kappa}(\mathbf{x}^0;\boldsymbol{\phi}) - \mathbf{h}(\boldsymbol{\phi})$, is easily shown to be $IF(\mathbf{x}^0;\boldsymbol{\phi}) = \Gamma^{-1}(\boldsymbol{\phi})[\boldsymbol{\kappa}(\mathbf{x}^0;\boldsymbol{\phi}) - \mathbf{h}(\boldsymbol{\phi})]$.

6.4.2 Bias and Variance

Let us compute the bias and asymptotic covariance matrix of the estimator defined by $T(\cdot)$. Clearly, if $\hat{\boldsymbol{\phi}}_M = T(\mathcal{F}_M^D)$, then for sufficiently large M a von Mises expansion of the functional $T(\cdot)$ about $\mathcal{F}^D$ yields the bias in the estimate, i.e.,

$$T(\mathcal{F}_M^D) - T(\mathcal{F}^D) = \int IF(\mathbf{x};\boldsymbol{\phi})d(\mathcal{F}_M^D - \mathcal{F}^D) + \text{ remainder},$$

where the remainder becomes negligible as $M \to \infty$ [70, for example]. Evaluating the integral over $\mathcal{F}_M^D$ and making use of (6.32) yields the bias in the estimate from an observation set $\mathcal{X}_M$:

$$M[\hat{\phi}_M - \phi_0] \approx \frac{1}{M} \sum_{s \in \Omega_M} IF(\mathbf{x}_s; \phi). \tag{6.33}$$

The asymptotic bias is then computed as

$$\begin{aligned} \mathbf{b}(T, \mathcal{F}^D) &= \lim_{M \to \infty} \mathrm{E}\left\{\hat{\phi}_M - \phi_0\right\} \\ &= \lim_{M \to \infty} \int IF(\mathbf{x}; \phi) d\mathcal{F}_M^D \\ &= \int IF(\mathbf{x}; \phi) d\mathcal{F}^D \\ &= \mathbf{0}. \end{aligned}$$

The asymptotic covariance can now be computed as

$$\begin{aligned} V(T, \mathcal{F}^D) &= \lim_{M \to \infty} \mathrm{E}\left\{M^2[\hat{\phi}_M - \phi_0][\hat{\phi}_M - \phi_0]^t\right\} \\ &= \lim_{M \to \infty} \int \left[\frac{1}{M} \sum_{s \in \Omega_M} IF(\mathbf{x}_s; \phi)\right] \left[\frac{1}{M} \sum_{t \in \Omega_M} IF(\mathbf{x}_t; \phi)\right]^t d\mathcal{F}^D \\ &= \lim_{M \to \infty} \frac{1}{M^2} \sum_{s \in \Omega_M} \sum_{t \in \Omega_M} \int IF(\mathbf{x}_s; \phi) IF^t(\mathbf{x}_t; \phi) d\mathcal{F}^D. \end{aligned} \tag{6.34}$$

At this juncture in order to simplify (6.34) we specialize to a particular form of the influence function. For example, if

$$\int IF(\mathbf{x}_s; \phi) d\mathcal{F}(x(\mathbf{s}) | x(\mathbf{u}), \mathbf{u} < \mathbf{s}) = \mathbf{0}, \tag{6.35}$$

then by using the arguments supporting the proof of Proposition 6.1 we have that

$$\begin{aligned} V(T, \mathcal{F}^D) &= \lim_{M \to \infty} \frac{1}{M^2} \sum_{s \in \Omega_M} \int IF(\mathbf{x}_s; \phi) IF^t(\mathbf{x}_s; \phi) d\mathcal{F}^D \\ &= \mathrm{E}\left\{IF(\mathbf{x}; \phi) IF^t(\mathbf{x}; \phi)\right\}. \end{aligned}$$

However, we note that of the models considered in Section 6.1, only the NSHP model has this property, in which case

$$\int IF(\mathbf{x}^0;\phi)d\mathcal{F}(x_0|\mathbf{x}) = 0 \tag{6.36}$$

and

$$\begin{aligned} V(T,\mathcal{F}^{N_0}) &= \mathrm{E}\Big\{IF(\mathbf{x}^0;\phi)IF^t(\mathbf{x}^0;\phi)\Big\} \\ &= M^{-1}(\psi,\mathcal{F}^{N_0})Q(\psi,\phi)M^{-t}(\psi,\mathcal{F}^{N_0}), \end{aligned} \tag{6.37}$$

where

$$Q(\psi,\phi) = \int \psi(\mathbf{x}^0;\phi)\psi^t(\mathbf{x}^0;\phi)d\mathcal{F}^{N_0} \tag{6.38}$$

and

$$M(\psi,\mathcal{F}^{N_0}) = \int \psi(\mathbf{x}^0;\phi)\kappa^t(\mathbf{x}^0;\phi)d\mathcal{F}^{N_0}. \tag{6.39}$$

At this time we know of no property of either the bilateral GMRF or NCAR model which simplifies the covariance matrix expression (6.34). Consequently, in the remainder of this chapter we specialize to estimators whose influence function satisfies (6.35) which, in particular, includes the NSHP model. Optimal estimators for the bilateral GMRF and NCAR models will be addressed further in the next chapter.

6.4.3 Asymptotic Cramer-Rao Inequality

Let ϕ be fixed and $\tilde{\phi} = T(\tilde{\mathcal{F}}^D)$, then

$$\tilde{\phi} - \phi = \int IF(\mathbf{x};\phi)d\tilde{\mathcal{F}}^D + \text{ remainder.}$$

Differentiating with respect to $\tilde{\phi}$ at $\tilde{\phi} = \phi$ yields

$$I_d = \int IF(\mathbf{x};\phi)\left[\frac{\partial}{\partial\tilde{\phi}}\tilde{f}^D(\mathbf{x})\right]^t_{\tilde{\phi}=\phi} d\mathbf{x}$$

$$= \int IF(\mathbf{x};\boldsymbol{\phi}) \left[\frac{\partial}{\partial \boldsymbol{\phi}} \log(f^D(\mathbf{x}))\right]^t d\mathcal{F}^D$$
$$= \int IF(\mathbf{x};\boldsymbol{\phi}) \left[\boldsymbol{\lambda}^d(\mathbf{x};\boldsymbol{\phi})\right]^t d\mathcal{F}^D \tag{6.40}$$

where I_d is the $d \times d$-dimensional identity matrix.

Now, let $\mathbf{a}(\mathbf{x};\boldsymbol{\phi}) = \text{col}\,\{IF(\mathbf{x};\boldsymbol{\phi}), \boldsymbol{\lambda}^d(\mathbf{x};\boldsymbol{\phi})\}$. The covariance matrix of $\mathbf{a}(\mathbf{x};\boldsymbol{\phi})$ is

$$\mathrm{E}\left\{\mathbf{a}(\mathbf{x};\boldsymbol{\phi})\mathbf{a}^t(\mathbf{x};\boldsymbol{\phi})\right\} = \begin{bmatrix} V(T,\mathcal{F}^D) & I_d \\ I_d & \Gamma(\boldsymbol{\phi}) \end{bmatrix}, \tag{6.41}$$

which is a nonnegative-definite matrix. Therefore, for any $\mathbf{b} \in \Re^d$ and letting $\bar{\mathbf{b}} = \text{col}\,\{\mathbf{b}, -\Gamma^{-1}(\boldsymbol{\phi})\mathbf{b}\}$, it follows that

$$0 \leq \bar{\mathbf{b}}^t \begin{bmatrix} V(T,\mathcal{F}^D) & I_d \\ I_d & \Gamma(\boldsymbol{\phi}) \end{bmatrix} \bar{\mathbf{b}}$$
$$= \mathbf{b}^t V(T,\mathcal{F}^D)\mathbf{b} - \mathbf{b}^t\Gamma^{-1}(\boldsymbol{\phi})\mathbf{b}.$$

Consequently, $\mathbf{b}^t V(T,\mathcal{F}^D)\mathbf{b} \geq \mathbf{b}^t\Gamma^{-1}(\boldsymbol{\phi})\mathbf{b}$ for all $\mathbf{b} \in \Re^d$.

6.5 An Optimal Robust Estimator

6.5.1 Optimality Problem

For multi-dimensional estimators Hampel et al [38] have suggested the following extremal problem: minimize the trace of the asymptotic covariance matrix among all estimators of the type (6.24) which have an influence function and for which $\gamma_u^* = \sup_{\mathbf{x}} \{|\mathrm{IF}(\mathbf{x},\boldsymbol{\phi})|\} \leq c(\boldsymbol{\phi})$. γ_u^* is called the *gross-error sensitivity*. In the present context (6.35) and (6.40) are additional constraints. Thus, an *optimal robust estimator* is a solution to the following constrained optimization problem:

$$\min\left\{\mathrm{tr}\left\{\int IF(\mathbf{x};\boldsymbol{\phi})IF^t(\mathbf{x};\boldsymbol{\phi})d\mathcal{F}^D)\right\}\right\}$$

subject to

$$\begin{aligned}
&\text{Condition 1:} && \int IF(\mathbf{x}_s;\phi)d\mathcal{F}(x(\mathbf{s})|x(\mathbf{t}),\mathbf{t}<\mathbf{s}) = \mathbf{0} \\
&\text{Condition 2:} && \int IF(\mathbf{x};\phi)\left[\boldsymbol{\lambda}^d(\mathbf{x};\phi)\right]^t d\mathcal{F}^D = I_d \\
&\text{Condition 3:} && \sup_{\mathbf{x}}\{|IF(\mathbf{x};\phi)|\} \leq c(\phi).
\end{aligned}$$

In a practical sense, though, one would want the third constraint to be independent of ϕ. This can be accomplished by considering the *information-standardized sensitivity*

$$\begin{aligned}
\gamma_i^* &= \sup_{\mathbf{x}}\left\{IF^t(\mathbf{x};\phi)\Gamma(\phi)IF(\mathbf{x};\phi)\right\}^{\frac{1}{2}} \\
&= \sup_{\mathbf{x}}\left\{|J^t(\phi)IF(\mathbf{x};\phi)|\right\} \leq c, \qquad (6.42)
\end{aligned}$$

where $J(\phi)J^t(\phi) = \Gamma(\phi)$, or by considering the *self-standardized sensitivity*

$$\gamma_s^* = \sup_{\mathbf{x}}\left\{IF^t(\mathbf{x};\phi)V^{-1}(T,\mathcal{F}^D)IF^t(\mathbf{x};\phi)\right\}^{\frac{1}{2}}. \qquad (6.43)$$

Proof of the invariance of γ_i^* and γ_s^* to parameter transformations can be found in [38, Chapter 4].

However, these standardizations of the sensitivity are not consistent with use of $\mathrm{tr}\left\{V(T,\mathcal{F}^D)\right\}$ as the criterion to optimize [38]. To see this, note that the criterion can be interpreted as the asymptotic mean-squared error, viz.,

$$\begin{aligned}
\mathrm{tr}\left\{V(T,\mathcal{F}^D)\right\} &= \mathrm{tr}\left\{\int IF(\mathbf{x};\phi)IF^t(\mathbf{x};\phi)d\mathcal{F}^D\right\} \\
&= \int |IF(\mathbf{x};\phi)|^2 d\mathcal{F}^D \\
&= \mathrm{E}\left\{|T(\mathcal{F}_M^D) - \phi_0|^2\right\}.
\end{aligned}$$

Consequently, the same metric should be used to measure the asymptotic mean-squared error. For example, if γ_i^* is used, then the criterion to optimize should

be

$$\int |J^t(\phi)IF(\mathbf{x};\phi)|^2 d\mathcal{F}^D = \mathrm{tr}\left\{J^t(\phi)\left[\int IF(\mathbf{x};\phi)IF^t(\mathbf{x};\phi)d\mathcal{F}^D\right]J(\phi)\right\}$$
$$= \mathrm{tr}\left\{\Gamma(\phi)\left[\int IF(\mathbf{x};\phi)IF^t(\mathbf{x};\phi)d\mathcal{F}^D\right]\right\}. \quad (6.44)$$

The optimality criterion using the self-standardized senstitivity is

$$\min_{\psi} \mathrm{tr}\left\{\Gamma(\phi)\left[\int IF(\mathbf{x};\phi)IF^t(\mathbf{x};\phi)d\mathcal{F}^{N_0}\right]\right\}.$$

Since the influence function is completely determined by the ψ-function, we replace $IF(\mathbf{x};\phi)$ with $\psi(\mathbf{x};\phi)$ where the matrix $M(\psi,\mathcal{F}^{N_0})$ has been absorbed in $\psi(\mathbf{x};\phi)$. For the NSHP model D can be taken as N_0, in which case $d = p+1$ and $\boldsymbol{\lambda}^{p+1}(\mathbf{x};\phi) = \kappa(\mathbf{x}^0;\phi)$. The asymptotic variance expression has the simple representation (6.37). With these changes the optimality problem becomes

$$\min_{\psi} \mathrm{tr}\left\{\Gamma(\phi)\left[\int \psi(\mathbf{x}^0;\phi)\psi^t(\mathbf{x}^0;\phi)d\mathcal{F}^{N_0}\right]\right\} \quad (6.45)$$

subject to

$$\text{Condition 1:} \quad \int \psi(\mathbf{x}^0;\phi)d\mathcal{F}(x_0|\mathbf{x}) = \mathbf{0} \quad (6.46)$$

$$\text{Condition 2:} \quad \int \psi(\mathbf{x}^0;\phi)\kappa^t(\mathbf{x}^0;\phi)d\mathcal{F}^{N_0} = I_{p+1} \quad (6.47)$$

$$\text{Condition 3:} \quad \sup_{\mathbf{x}}\{|J^t(\phi)\psi(\mathbf{x}^0;\phi)|\} \leq c. \quad (6.48)$$

The solution to this problem is an optimal robust estimator for the NSHP autoregressive model.

6.5.2 Optimal Solution

Consider (6.45). Let $A(\phi)$ be an arbitrary matrix depending only on the parameters ϕ. In view of (6.44) solving (6.45) is the same as minimizing

$$\int |J^t(\phi)\psi(\mathbf{x}^0;\phi)|^2 d\mathcal{F}^{N_0}$$

$$
\begin{aligned}
&= \int |J^t(\phi)\psi(\mathbf{x}^0;\phi) - A(\phi)\kappa(\mathbf{x}^0;\phi)|^2 d\mathcal{F}^{N_0} \\
&\qquad +\mathrm{tr}\left\{\int J^t(\phi)\psi(\mathbf{x}^0;\phi)\kappa^t(\mathbf{x}^0;\phi)A^t(\phi)d\mathcal{F}^{N_0}\right\} \\
&\qquad +\mathrm{tr}\left\{\int A(\phi)\kappa(\mathbf{x}^0;\phi)\psi^t(\mathbf{x}^0;\phi)J^t(\phi)d\mathcal{F}^{N_0}\right\} \\
&\qquad -\mathrm{tr}\left\{\int A(\phi)\kappa(\mathbf{x}^0;\phi)\kappa^t(\mathbf{x}^0;\phi)A^t(\phi)d\mathcal{F}^{N_0}\right\} \\
&= \int |J^t(\phi)\psi(\mathbf{x}^0;\phi) - A(\phi)\kappa(\mathbf{x}^0;\phi)|^2 d\mathcal{F}^{N_0} \\
&\qquad +2\mathrm{tr}\left\{A(\phi)J(\phi)\right\} - \mathrm{tr}\left\{A(\phi)\Gamma(\phi)A^t(\phi)\right\}
\end{aligned}
$$

This last expression is minimized when $J^t(\phi)\psi(\mathbf{x}^0;\phi) = A(\phi)\kappa(\mathbf{x}^0;\phi)$. This yields, in the ML case, that $A(\phi) = J^{-1}(\phi)$, as expected. In the robust situation, though, $\kappa(\mathbf{x}^0;\phi)$ may be unbounded due to contamination in the data, leading to an unbounded influence function and, possibly, poor estimates. The robust $\psi(\mathbf{x}^0;\phi)$ must be chosen such that $J^t(\phi)\psi(\mathbf{x}^0;\phi)$ is as close as possible to $A(\phi)\kappa(\mathbf{x}^0;\phi)$ but yet have the properties that yield good estimates when the data is contaminated. This suggests that the optimal ψ be taken so that

$$
J^t(\phi)\psi(\mathbf{x}^0;\phi) = \mathbf{h}_c(A(\phi)\kappa(\mathbf{x}^0;\phi)) \tag{6.49}
$$

where $\mathbf{h}_c$ is the $(p+1)$-dimensional Huber function given by

$$
\mathbf{h}_c(\mathbf{z}) = \mathbf{z} \cdot \min\left\{1, \frac{c}{|\mathbf{z}|}\right\} \tag{6.50}
$$

and $|\mathbf{z}|^2 = \mathbf{z}^t\mathbf{z}$ for $\mathbf{z} \in \Re^{p+1}$.

By design this ψ-function satisfies the sensitivity constraint in (6.48). Equation (6.46) is also satisfied for any symmetric distribution $\mathcal{F}(x_0|\mathbf{x})$ and, in particular, the Gaussian NSHP case where $f(x_0|\mathbf{x}) = (2\pi\beta^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\beta^2}(x_0 - \theta^t\mathbf{x})^2\right]$.

The matrix $A(\phi)$ can be determined from (6.47) which is

$$
\int \mathbf{h}_c(A(\phi)\kappa(\mathbf{x}^0;\phi))[J^{-1}(\phi)\kappa(\mathbf{x}^0;\phi)]^t d\mathcal{F}^{N_0} = I_{p+1}.
$$

We make the initial guess that $A(\phi)$ is a constant times $J^{-1}(\phi)$, which is reasonable in view of the fact that for the unstandardized case

$$\Gamma^{-1}(\phi)\int \kappa(\mathbf{x}^0;\phi)\kappa^t(\mathbf{x}^0;\phi)d\mathcal{F}^{N_0} = I_{p+1}.$$

To determine the constant we adopt the method of Künsch [69, Lemma 2]. Consequently, we assume β is known and define the p-dimensional vector

$$\mathbf{y}_s = \frac{1}{\beta}J^{-1}(\theta)\mathbf{x}_s, \tag{6.51}$$

where $J(\theta)$ is any $p \times p$ matrix such that $J(\theta)J^t(\theta) = \Gamma(\theta) = \frac{1}{\beta^2}Q_{\mathbf{x}}$. The vector $\mathbf{y}_s$ is Gaussian with zero mean and covariance matrix

$$\begin{aligned}
\mathrm{E}\left\{\mathbf{y}_s\mathbf{y}_s^t\right\} &= \frac{1}{\beta^2}J^{-1}(\theta)\mathrm{E}\left\{\mathbf{x}_s\mathbf{x}_s^t\right\}J^{-t}(\theta) \\
&= \frac{1}{\beta^2}J^{-1}(\theta)\Gamma(\theta)J^{-t}(\theta) \\
&= I_p.
\end{aligned}$$

Thus, the joint density function of the elements of $\mathbf{y}_s$ is $\eta(\mathbf{y}_s) = \prod_{\mathbf{r}\in N}\eta(y_s(\mathbf{r}))$, where $\eta(\cdot)$ is the standard normal density function.

Let $c(b)$ be a positive scalar function of a positive scalar b, and consider (6.47) with $A(\theta) = [c(b)J(\theta)]^{-1}$, Thus,

$$\int \mathrm{h}_c(c^{-1}(b)J^{-1}(\theta)\kappa_\theta(\mathbf{x}^0;\phi))[J^{-1}(\theta)\kappa_\theta(\mathbf{x}^0;\phi)]^t d\mathcal{F}^{N_0} = I_p. \tag{6.52}$$

Using (6.51) we note that

$$J^{-1}(\theta)\kappa_\theta(\mathbf{x}^0;\phi) = \mathbf{y}w,$$

where $w = \frac{1}{\beta}(x_0 - \theta^t\mathbf{x})$, which has the standard normal distribution $\eta(w)$. Clearly, $d\mathcal{F}^{N_0} = f(x_0|\mathbf{x})f(\mathbf{x})dx_0d\mathbf{x}$; and with the change of variables from x_0 and $\mathbf{x}$ to w

and $\mathbf{y}$, respectively, $d\mathcal{F}^{N_0} = \eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y}$. So, consequently

$$c(b)I_p = \int \mathbf{h}_b(\mathbf{y}w) \cdot w\mathbf{y}^t \eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y},$$

since $\mathbf{h}_c(a\mathbf{z}) = a\mathbf{h}_{c/a}(\mathbf{z})$ for any positive scalar a and where $b = c \cdot c(b)$. The Huber function in the above integral is explicitly given by

$$\mathbf{h}_b(\mathbf{y}w) = \begin{cases} \mathbf{y}w, & |w| \leq \frac{b}{|\mathbf{y}|} \\ b \cdot \text{sign}\{w\} \cdot \frac{\mathbf{y}}{|\mathbf{y}|}, & |w| \geq \frac{b}{|\mathbf{y}|}. \end{cases} \tag{6.53}$$

Then, using this explicit expression the integral can be split so that

$$\begin{aligned} c(b)I_p &= \int \int_{-\frac{b}{|\mathbf{y}|}}^{\frac{b}{|\mathbf{y}|}} w^2 \mathbf{y}\mathbf{y}^t \eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y} \\ &\quad + \frac{b}{|\mathbf{y}|} \int \int_{\frac{b}{|\mathbf{y}|}}^{\infty} w\mathbf{y}\mathbf{y}^t \eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y} \\ &\quad - \frac{b}{|\mathbf{y}|} \int \int_{-\infty}^{-\frac{b}{|\mathbf{y}|}} w\mathbf{y}\mathbf{y}^t \eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y} \\ &= 2 \int \mathbf{y}\mathbf{y}^t \left[\int_0^{\frac{b}{|\mathbf{y}|}} w^2 \eta(w)dw + \frac{b}{|\mathbf{y}|} \int_{\frac{b}{|\mathbf{y}|}}^{\infty} w\eta(w)dw \right] \eta(\mathbf{y})d\mathbf{y}. \end{aligned}$$

By taking the trace of both sides of this last equation and replacing $\text{tr}\{\mathbf{y}\mathbf{y}^t\}$ with $|\mathbf{y}|^2$ we have

$$c(b) = \frac{2}{p} \int |\mathbf{y}|^2 \left[\int_0^{\frac{b}{|\mathbf{y}|}} w^2 \eta(w)dw + \frac{b}{|\mathbf{y}|} \int_{\frac{b}{|\mathbf{y}|}}^{\infty} w\eta(w)dw \right] \eta(\mathbf{y})d\mathbf{y}.$$

This expression can be further simplified by noting that $|\mathbf{y}|^2 = \sum_{\mathbf{r} \in N} y^2(\mathbf{r})$ is a chi-squared distributed random variable with p-degrees of freedom. Therefore, let $z = |\mathbf{y}|^2$ so that

$$c(b) = \frac{2}{p} \int_0^{\infty} z \left[\int_0^{bz^{-\frac{1}{2}}} w^2 \eta(w)dw + bz^{-\frac{1}{2}} \int_{bz^{-\frac{1}{2}}}^{\infty} w\eta(w)dw \right] \chi_p^2(z)dz. \tag{6.54}$$

The constant $c(b)$ for various values of b has been computed for a NSHP model with neighbor set consisting of the three nearest neighbors ($p = 3$). The results are shown in Table 6.1.

b	$c(b)$	$c^*(b)$	Efficiency	γ_i^*
1.0	0.39525	0.20945	0.74584	2.53008
1.2	0.46161	0.27389	0.77798	2.59960
1.5	0.55121	0.37046	0.82015	2.72130
2.0	0.67444	0.51943	0.87570	2.96544
2.5	0.76812	0.64430	0.91574	3.25468
2.8	0.81224	0.70650	0.93380	3.44727
3.0	0.83732	0.74285	0.94381	3.58286
3.2	0.85934	0.77537	0.95240	3.72381
4.0	0.92274	0.87242	0.97596	4.33493
5.0	0.96462	0.93979	0.99011	5.18337
6.0	0.98424	0.97258	0.99604	6.09606
8.0	0.99706	0.99472	0.99940	8.02363
9.5	0.99919	0.99853	0.99986	9.50766

Table 6.1: Efficiency and gross-error constants for NSHP optimal robust estimator, $p = 3$.

Let the optimal ψ-function be defined by (6.49). Then

$$\begin{aligned}\psi(\mathbf{x}^0;\phi) &= c^{-1}(b)J^{-t}(\phi)\mathbf{h}_b(J^{-1}(\phi)\kappa(\mathbf{x}^0;\phi)) \\ &= c^{-1}(b)\Gamma^{-1}(\phi)\kappa(\mathbf{x}^0;\phi)\cdot\min\left\{1,\frac{b}{|J^{-1}(\phi)\kappa(\mathbf{x}^0;\phi)|}\right\},\end{aligned}$$

where

$$\kappa(\mathbf{x}^0;\phi) = \begin{bmatrix} \frac{1}{\beta}\mathbf{x}w \\ \frac{1}{\beta}(w^2-\alpha) \end{bmatrix}$$

and α is selected to make the estimate of β consistent when the data are Gaussian.

Clearly, any matrix multiple of $\psi(\mathbf{x}^0;\phi)$ yields the same estimator, so take

$$\psi(\mathbf{x}^0;\phi) = \begin{bmatrix} \mathbf{x}w \\ (w^2-\alpha) \end{bmatrix}\cdot\min\left\{1,\frac{b}{d_{\mathbf{x}^0}}\right\},$$

where

$$d^2_{\mathbf{x}^0} = \left|J^{-1}(\phi)\begin{bmatrix} \frac{1}{\beta}\mathbf{x}w \\ \frac{1}{\beta}(w^2-\alpha) \end{bmatrix}\right|^2.$$

Substitute this ψ in the defining equation (6.24) and evaluate at the empirical distribution to obtain the optimal robust estimator for the NSHP model:

$$\sum_{s\in\Omega_M} J^{-t}(\phi)\mathrm{h}_c(c^{-1}(b)J^{-1}(\phi)\kappa(\mathrm{x}_s^0;\phi)) = 0.$$

Using (6.18), (6.19), and (6.50) and upon simplification, the NSHP optimal robust estimator becomes

$$\sum_{s\in\Omega_M} \mathrm{x}_s w(s)\cdot W_b(d_{\mathrm{x}_s^0}) = \mathbf{0} \tag{6.55}$$

$$\sum_{s\in\Omega_M} [w^2(s)-\alpha]\cdot W_b(d_{\mathrm{x}_s^0}) = 0 \tag{6.56}$$

where

$$w(s) = \frac{1}{\beta}[x(s)-\theta^t\mathrm{x}_s] \tag{6.57}$$

$$W_b(d_{\mathrm{x}_s^0}) = \min\left\{1, \frac{b}{|d_{\mathrm{x}_s^0}|}\right\} \tag{6.58}$$

$$d^2_{\mathrm{x}_s^0} = w^2(s)\cdot \mathrm{x}_s^t Q_{\mathrm{x}}^{-1}\mathrm{x}_s + \frac{1}{2}[w^2(s)-\alpha]^2 \tag{6.59}$$

$$Q_{\mathrm{x}} = \mathrm{E}\left\{\mathrm{x}_s\mathrm{x}_s^t\right\}. \tag{6.60}$$

and the consistency constant α is determined from

$$\alpha = \frac{a_1(\alpha,b)}{a_2(\alpha,b)} \tag{6.61}$$

where

$$a_1(\alpha,b) = \int_0^\infty v\left[\int_0^{\frac{1}{v}\left[b^2-\frac{1}{2}(v-\alpha)^2\right]} \chi_p^2(z)dz + b\int_{\frac{1}{v}\left[b^2-\frac{1}{2}(v-\alpha)^2\right]}^\infty \frac{1}{|d_{v,z}|}\chi_p^2(z)dz\right]\chi_1^2(v)dv$$

$$a_2(\alpha,b) = \int_0^\infty \left[\int_0^{\frac{1}{v}\left[b^2-\frac{1}{2}(v-\alpha)^2\right]} \chi_p^2(z)dz + b\int_{\frac{1}{v}\left[b^2-\frac{1}{2}(v-\alpha)^2\right]}^\infty \frac{1}{|d_{v,z}|}\chi_p^2(z)dz\right]\chi_1^2(v)dv$$

$$d_{v,z} = d_{\mathrm{x}^0}\big|_{|\mathrm{y}|^2=z,\, w^2=v}.$$

The consistency constant α for several values of b has been computed for a NSHP model ($p = 3$) by an iterative technique and the results are shown in Table 6.2. Figure 6.2 plots the consistency constant versus cutoff for several NSHP models.

b	α
1.0	0.58435
1.2	0.63285
1.5	0.69540
2.0	0.77747
2.5	0.83797
2.8	0.86627
3.0	0.88237
4.0	0.93852
5.0	0.96809
6.0	0.98352
7.0	0.99153
8.0	0.99564
9.5	0.99835

Table 6.2: Consistency constants for the optimal robust NSHP estimator.

The solution of the $p + 1$ equations (6.55) and (6.56) is the optimal robust estimate of θ and β for the NSHP model. The constant b determines both the robustness and efficiency of the estimator. In the next section we suggest a procedure for determining b. Note that if $b = \infty$, then (6.55) and (6.56) reduce to the ML or, equivalently, the LS estimator for the NSHP model.

We note that the estimator given by (6.55) through (6.61) is different than Künsch's estimator given in [69] for time series. Künsch's estimator is actually a hybrid estimator, consisting of an estimator in the spirit of (6.45) for θ but utilizing another estimator, Huber's "Proposal 2" [39, pg 137], for the parameter β. This is justified in view of the result [69, Theorem 2.6] that estimates of θ and β are asymptotically independent.

6.5.3 Estimator Efficiency

The critical parameter in (6.55) and (6.56) is b. As noted in the previous section for $b = \infty$ the efficiency of the estimator is optimal since then it corresponds to the ML estimator; but then, robustness suffers. For robust estimators it has been suggested [69] (see also Chapter 5) that, in general, they be designed by choosing an acceptable efficiency.

We define the *efficiency* (Eff) of a $(p+1)$-dimensional estimator $T(\mathcal{F}^{N_0})$ to be

$$\text{Eff}(\hat{\phi}) = \frac{\text{tr}\{\text{Cramer-Rao Bound}\}}{\text{tr}\{\text{Covariance of Estimator}\}} = \frac{\text{tr}\{\Gamma^{-1}(\phi)\}}{\text{tr}\{V(T, \mathcal{F}^{N_0})\}}. \tag{6.62}$$

For the NSHP optimal robust estimator the trace of the asymptotic covariance matrix can easily be computed from the ψ-function via (6.37). This computation shows that

$$\begin{aligned}
\text{tr}\left\{V_\theta(T, \mathcal{F}^{N_0})\right\} &= \text{tr}\left\{\int \psi(\mathbf{x}^0;\theta)\psi^t(\mathbf{x}^0;\theta)\mathcal{F}^{N_0}\right\} \\
&= c^{-2}(b)\text{tr}\left\{J^{-1}(\theta)\left[\int \mathbf{h}_b(\mathbf{y}w)\mathbf{h}_b^t(\mathbf{y}w)\eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y}\right]J^{-t}(\theta)\right\} \\
&= c^{-2}(b)\text{tr}\left\{\Gamma^{-1}(\theta)Q_\mathbf{h}\right\},
\end{aligned}$$

where $Q_\mathbf{h} = \int \mathbf{h}_b(\mathbf{y}w)\mathbf{h}_b^t(\mathbf{y}w)\eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y}$.

Let $[Q_\mathbf{h}]_{\mathbf{r},\mathbf{t}}$ be an element of the $p \times p$ matrix $Q_\mathbf{h}$, and let $y(\mathbf{r}), \mathbf{r} \in N$ be the elements of $\mathbf{y} = \frac{1}{\beta}J^{-1}(\theta)\mathbf{x}$, then

$$\begin{aligned}
[Q_\mathbf{h}]_{\mathbf{r},\mathbf{t}} &= \int [\mathbf{h}_b(\mathbf{y}w)]_\mathbf{r} [\mathbf{h}_b(\mathbf{y}w)]_\mathbf{t} \eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y} \\
&= \int w^2 y(\mathbf{r})y(\mathbf{t})\min^2\left\{1, \frac{b}{|\mathbf{y}w|}\right\}\eta(w)dw \cdot \eta(\mathbf{y})d\mathbf{y} \\
&= 2\int y(\mathbf{r})y(\mathbf{t})\left[\int_0^{\frac{b}{|\mathbf{y}|}} w^2\eta(w)dw + \frac{b^2}{|\mathbf{y}|^2}\int_{\frac{b}{|\mathbf{y}|}}^{\infty}\eta(w)dw\right]\eta(\mathbf{y})d\mathbf{y}
\end{aligned} \tag{6.63}$$

Let $g(b, |\mathbf{y}|)$ denote the integral inside the brackets in (6.63). Then, for $\mathbf{r} \neq \mathbf{t}$, we write

$$[Q_{\mathbf{h}}]_{\mathbf{r},\mathbf{t}} = 2 \int y(\mathbf{t}) \left[\int_{-\infty}^{\infty} y(\mathbf{r}) \cdot g(b, |\mathbf{y}|) \eta(y(\mathbf{r}) dy(\mathbf{r}) \right] \prod_{\mathbf{t} \neq \mathbf{r}} \eta(y(\mathbf{t})) dy(\mathbf{t}).$$

However, the function $y(\mathbf{r}) \cdot g(b, |\mathbf{y}|)$ is an odd function with respect to $y(\mathbf{r})$; and therefore, the integral inside the brackets is zero.

This proves that $Q_{\mathbf{h}}$ is a diagonal matrix whose diagonal elements we denote as $c^*(b)$. Letting $z = |\mathbf{y}|$ as before, we can easily compute the trace of $Q_{\mathbf{h}}$ and determine that

$$c^*(b) = \frac{2}{p} \int_0^{\infty} z \left[\int_0^{bz^{-\frac{1}{2}}} w^2 \eta(w) dw + b^2 z^{-1} \int_{bz^{-\frac{1}{2}}}^{\infty} \eta(w) dw \right] \chi_p^2(z) dz. \tag{6.64}$$

Hence, $Q_{\mathbf{h}} = c^*(b) I_p$ and the trace of the asymptotic covariance matrix is

$$\mathrm{tr}\left\{ V_\theta(T, \mathcal{F}^{N_0}) \right\} = c^{-2}(b) c^*(b) \mathrm{tr}\left\{ \Gamma^{-1}(\theta) \right\}. \tag{6.65}$$

The constant $c^*(b)$ for several values of b has been computed for a NSHP model with neighbor set consisting of the three nearest neighbors ($p = 3$). The results are shown in Table 6.1.

Efficiency for the NSHP optimal robust estimator is determined to be

$$\mathrm{Eff}(\hat{\theta}) = \frac{c^2(b)}{c^*(b)}. \tag{6.66}$$

Values of $\mathrm{Eff}(\hat{\theta})$ for selected values of b and a NSHP model ($p = 3$) are shown in Table 6.1. Figure 6.1 plots the efficiency versus cutoff for several NSHP models. Also, in Table 6.1 are values of $\gamma_i^* = \dfrac{b}{c(b)}$.

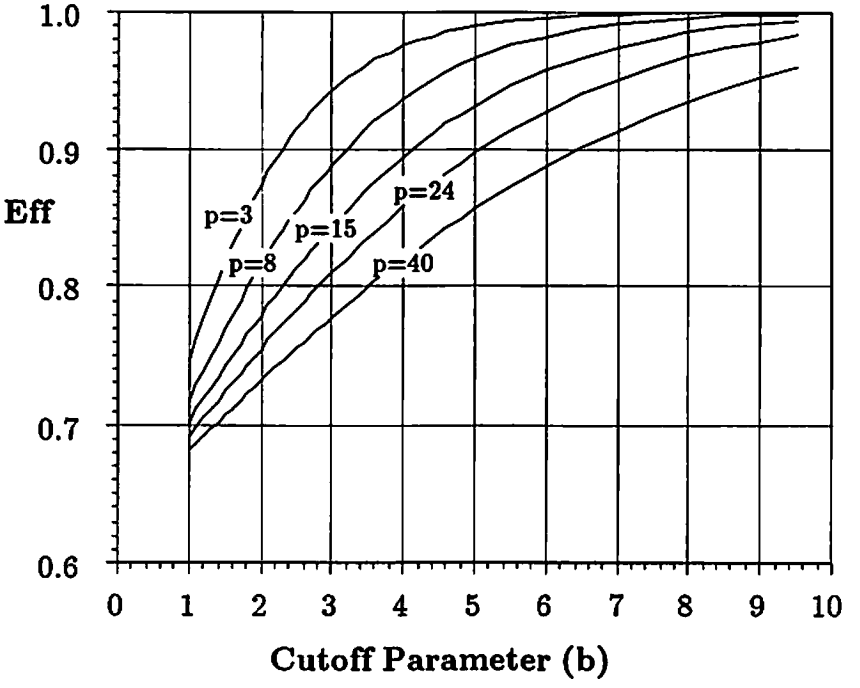


Figure 6.1: NSHP optimal robust estimator efficiencies for selected model orders.

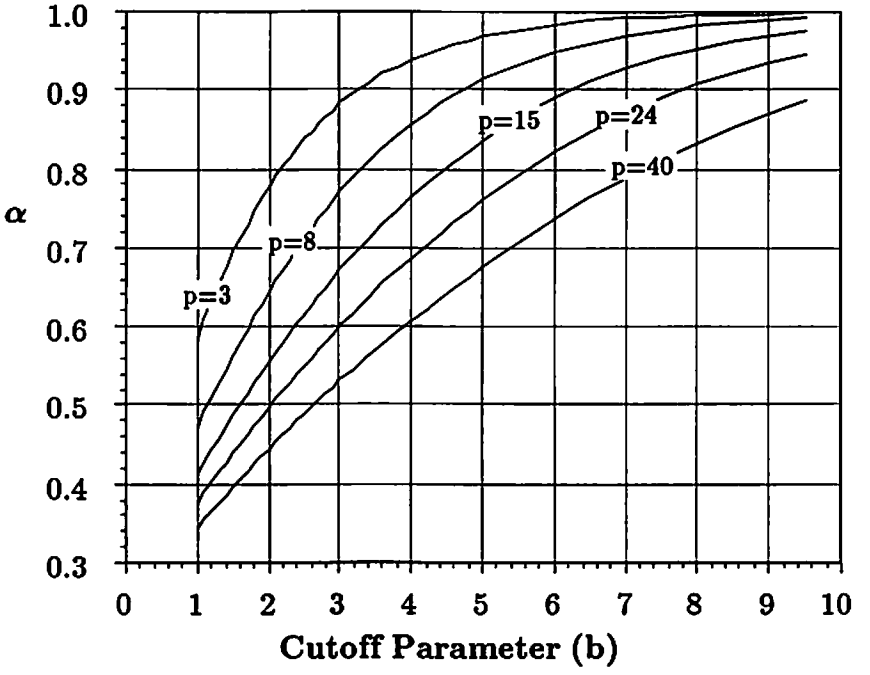


Figure 6.2: NSHP optimal robust estimator consistency constant for selected model orders.

6.6 Experimental Results

Experiments with synthetic data were conducted to evaluate the performance of the 2-D optimal robust estimator for the NSHP model. The optimal robust (OR) estimates were compared with the results obtained from applying the LS estimator, the 2-D GM-estimator of Chapter 5, and the optimal robust estimator of Künsch [69] adapted for the 2-D NSHP model. This estimator, designated "OR(K)", is given by (6.67) through (6.73).

$$\sum_{\mathbf{s}\in\Omega_M} \mathbf{x}_{\mathbf{s}} w(\mathbf{s}) \cdot W_{b_1}(d_{\mathbf{x}_{\mathbf{s}}}) = \mathbf{0} \tag{6.67}$$

$$\sum_{\mathbf{s}\in\Omega_M} [w^2(\mathbf{s}) \cdot W_{b_2}^2(w(\mathbf{s})) - \alpha_H] = 0 \tag{6.68}$$

where

$$w(\mathbf{s}) = \frac{1}{\hat{\beta}_{OR(K)}}[x(\mathbf{s}) - \hat{\boldsymbol{\theta}}_{OR(K)}^t \mathbf{x}_{\mathbf{s}}] \tag{6.69}$$

$$W_b(z) = \min\left\{1, \frac{b}{|z|}\right\} \tag{6.70}$$

$$d_{\mathbf{x}_{\mathbf{s}}}^2 = \mathbf{x}_{\mathbf{s}}^t Q_{\mathbf{x}}^{-1} \mathbf{x}_{\mathbf{s}} \tag{6.71}$$

$$Q_{\mathbf{x}} = \mathrm{E}\left\{\mathbf{x}_{\mathbf{s}} \mathbf{x}_{\mathbf{s}}^t\right\}. \tag{6.72}$$

and the consistency constant α_H is determined from

$$\alpha_H = \int_{-\infty}^{\infty} v^2 W_{b_2}^2(v) \eta(v) dv. \tag{6.73}$$

The consistency constant α_H for several values of b has been computed for a NSHP model ($p = 3$) and the results are shown in Table 6.3.

We used the procedure in [16] to generate an ensemble $E_y = \{\mathcal{Y}^k = \{y_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64\}$ of 64 sets of Gaussian NSHP data on a 32×32 toroidal lattice where each set $\mathcal{Y}^k$ used a unique seed for the Gaussian random number.

b	α_H
1.0	0.51606
1.2	0.63522
1.5	0.77847
2.0	0.92054
2.5	0.97756
2.8	0.99063
3.0	0.99501
3.2	0.99744
4.0	0.99988
5.0	1.00000
6.0	1.00000
8.0	1.00000
9.5	1.00000

Table 6.3: Consistency constants for Huber "Proposal 2" scale estimator.

The parameters of the NSHP model were identical for each set and are shown in Table 6.4.

$\theta_{0,-1}$	$\theta_{-1,0}$	$\theta_{-1,-1}$	β^2
.8	.6	−.5	1.0

Table 6.4: Model parameters used to generate NSHP data used in the experiments.

Least squares parameter estimates, (6.20) and (6.21), were then computed from each set of the ensemble for a NSHP model with neighbor set $N = \{(0,-1),(-1,0),(-1,-1)\}$. These 64 sets of LS estimates of θ and β^2 are denoted as $\phi^k_{LS} = \text{col}\,\{\theta^k_{LS}, \beta^k_{LS}\}, k = 1,\ldots,64$, and used for comparing the performance of the optimal robust and other estimators.

We conducted five experiments using data ensembles composed of E_y and various contaminations. In each experiment the comparisons of the estimators were made by computing the average error, ε, and the average squared errors, ς^2, relative to the LS estimates from E_y for each of the estimators. The coefficient

average errors were computed by

$$\varepsilon_{\mathbf{r}} = \frac{1}{64}\sum_{k=1}^{64}(\hat{\theta}_{\mathbf{r}}^{k} - \theta_{\mathbf{r},LS}^{k}), \quad \mathbf{r} \in N$$

and the coefficient average squared errors by

$$\varsigma_{\mathbf{r}}^{2} = \frac{1}{64}\sum_{k=1}^{64}(\hat{\theta}_{\mathbf{r}}^{k} - \theta_{\mathbf{r},LS}^{k})^{2}, \quad \mathbf{r} \in N.$$

The errors for β^2 were computed in a similar fashion.

The equations (6.55) and (6.56) for the OR estimator and (6.67) and (6.68) for OR(K) were solved by an iterative technique starting from LS estimates of the parameters and continuing until the change in the solution was less than a specified value δ. The iterative technique followed the general outline given below.

Algorithm:

1. Estimate $Q_{\mathbf{x}}$. This was accomplished by computing the sample covariances from the clean data E_y. However, when E_y is not available $Q_{\mathbf{x}}$ can be estimated by any of the robust covariance matrix estimators discussed in [38, Chapter 5] or [39, Chapter 8]. See, also, Chapter 7.

2. Obtain cutoff constants b (b_1, b_2) and compute α (α_h).

3. Compute LS estimates from the given data and set $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_{LS}$ and $\hat{\beta}_0^2 = \beta_{LS}^2$. Set $j = 0$.

4. At Step j, compute for OR

$$\hat{\boldsymbol{\theta}}_{j+1} = \left[\sum_{\mathbf{s}\in\Omega_M} W_b(d_{\mathbf{x}_\mathbf{s}^0})\mathbf{x}_\mathbf{s}\mathbf{x}_\mathbf{s}^t\right]^{-1} \sum_{\mathbf{s}\in\Omega_M} W_b(d_{\mathbf{x}_\mathbf{s}^0})\mathbf{x}_\mathbf{s}x(\mathbf{s})$$

$$\hat{\beta}_{j+1}^2 = \frac{1}{\alpha\sum_{\mathbf{s}\in\Omega_M} W_b(d_{\mathbf{x}_\mathbf{s}^0})} \cdot \sum_{\mathbf{s}\in\Omega_M} W_b(d_{\mathbf{x}_\mathbf{s}^0})[x(\mathbf{s}) - \hat{\boldsymbol{\theta}}_j^t\mathbf{x}_\mathbf{s}]^2.$$

Similar equations are solved for OR(K).

5. If $|\text{col}\,\{\hat{\theta}_{j+1}, \hat{\beta}^2_{j+1}\} - \text{col}\,\{\hat{\theta}_j, \hat{\beta}^2_j\}| > \delta$, $j = j + 1$ and goto Step 4.

6. Otherwise, stop.

In all experimental runs the above algorithm converged in fewer than 10 iterations for both OR and OR(K).

Experiment 1: In this experiment OR and OR(K) were used to estimate the parameters of a NSHP model from the clean Gaussian NSHP data in E_y. The cutoff constant for OR was $b = 2.0$. The cutoff constants for OR(K) were $b_1 = b_2 = 2.0$. The resulting average errors and average squared errors are shown in Figures 6.3 and 6.4.

Experiment 2: An ensemble of data with additive outliers at uncorrelated lattice points was generated by creating 64 sets of Gaussian noise $E_n = \{\mathcal{N}^k = \{n_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, i = 1, \ldots, 64\}$ where $\text{E}\,\{n(\mathbf{s})\} = 0.0$ and $\text{V}\,\{n(\mathbf{s})\} = 5.0$, yielding an approximate signal-to-noise ratio of -5db. Another ensemble of $0-1$ random data $E_u = \{\mathcal{Z}^k = \{z_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64\}$ was created with a binomial distribution having $\text{P}\,\{z(\mathbf{s}) = 1\} = .05$, equivalent to a 5 percent contamination. The data in E_y, E_n, and E_u were then used in (6.25) to compute and ensemble E_x of contaminated NSHP data. Least squares, OR, OR(K), and GM-estimates were computed and errors calculated relative to the LS estimates from E_y. The results are shown in Figures 6.5 and 6.6.

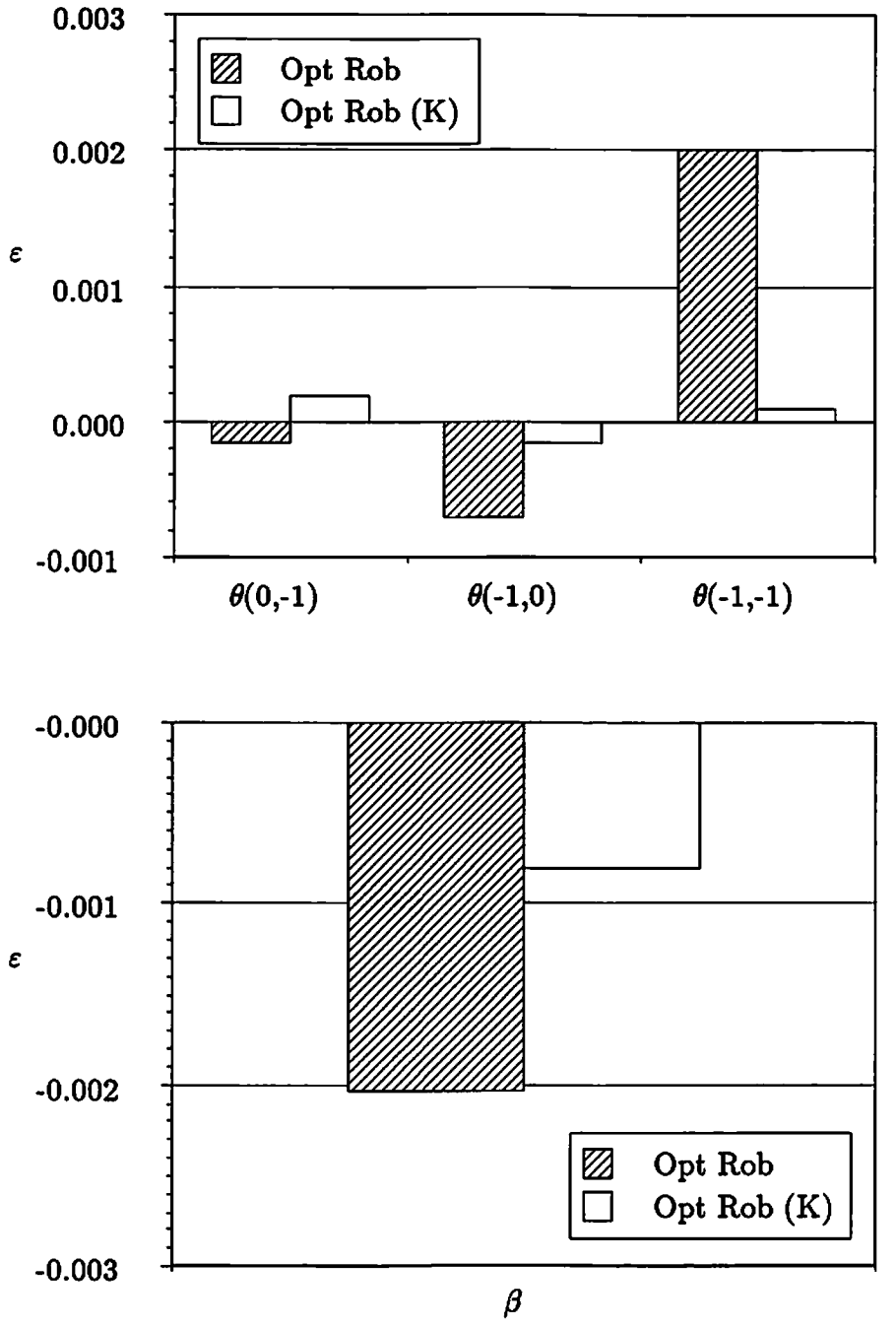


Figure 6.3: Average parameter estimate errors from Gaussian data.

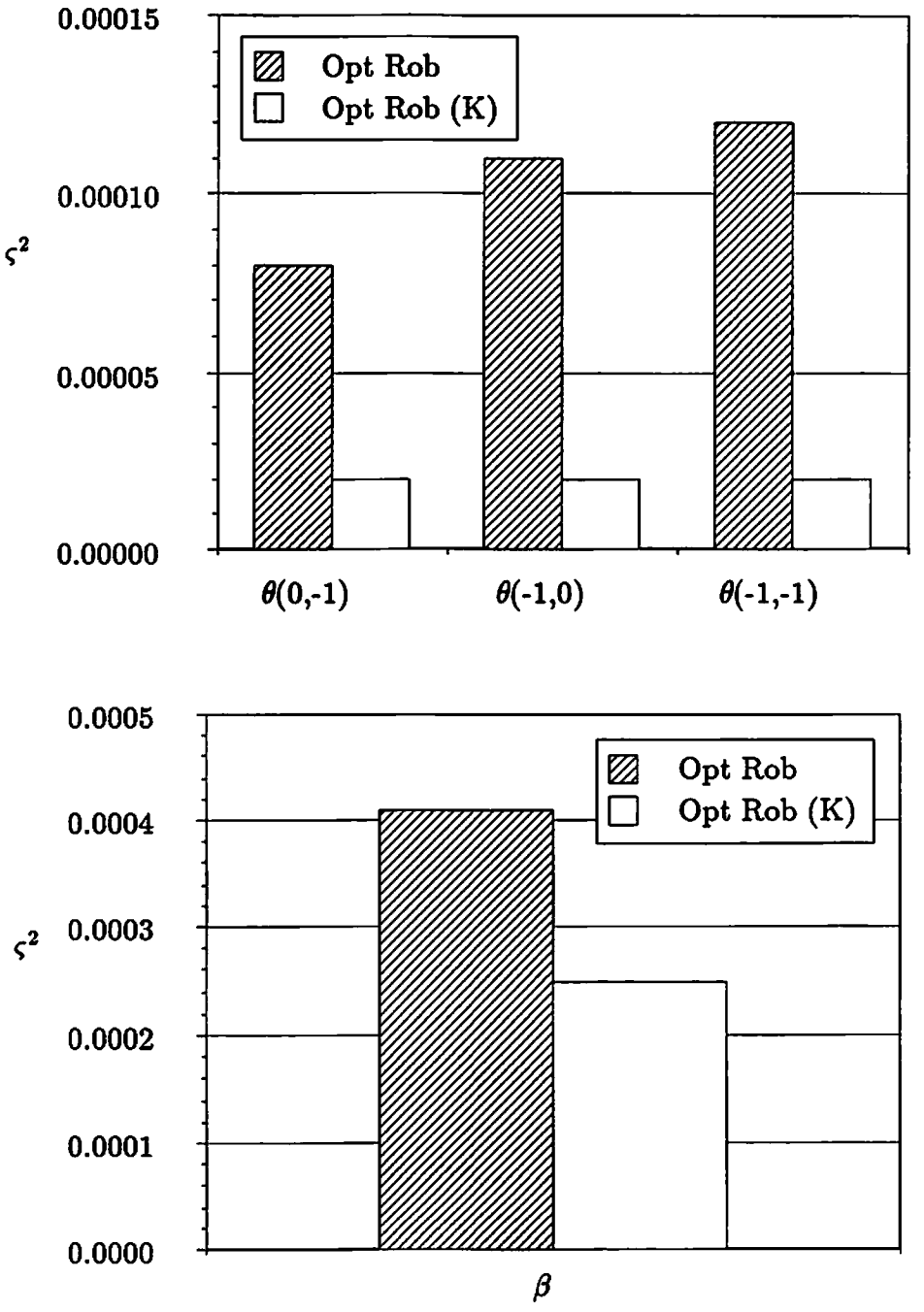


Figure 6.4: Average parameter estimate squared errors from Gaussian data.

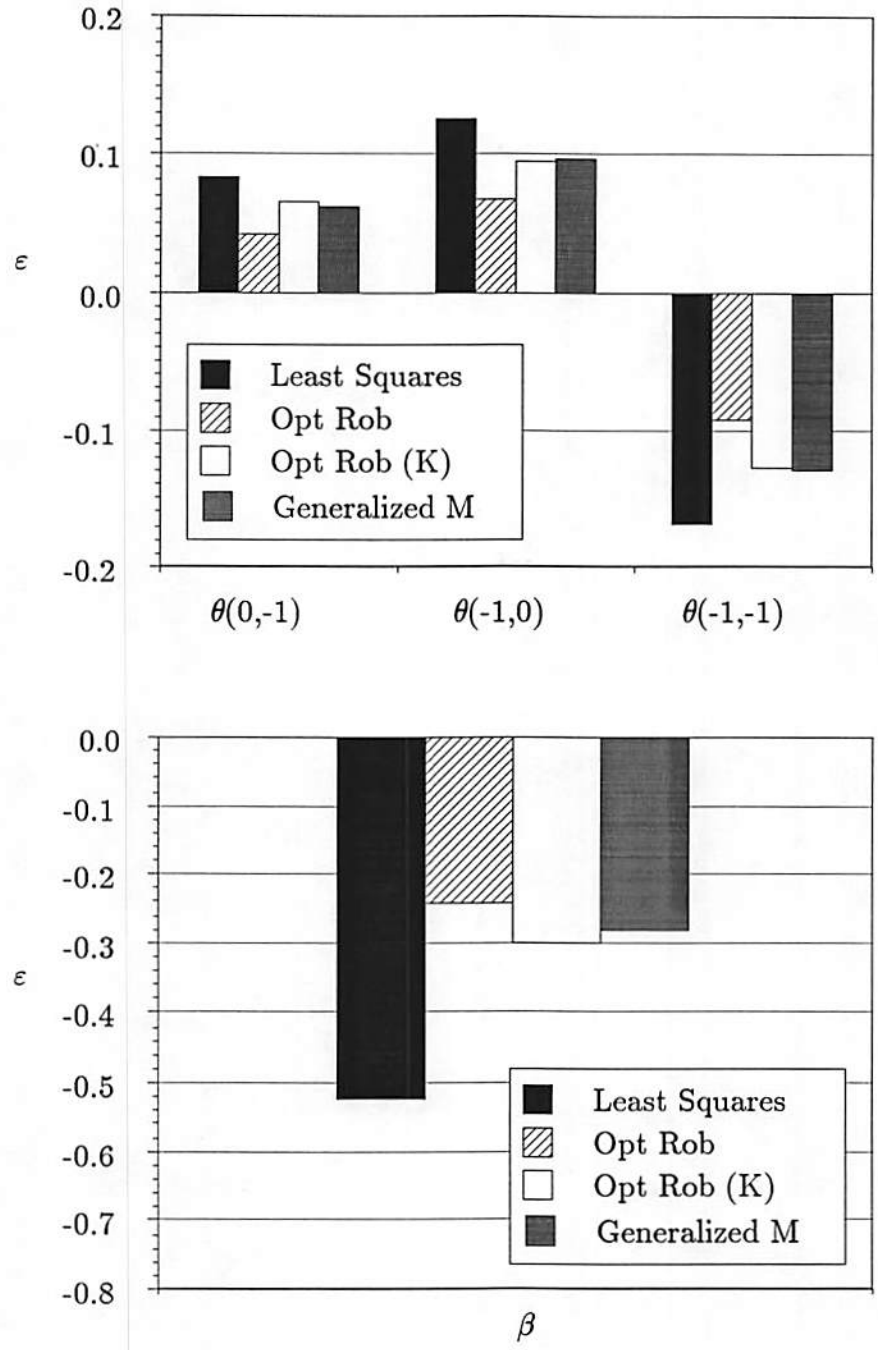
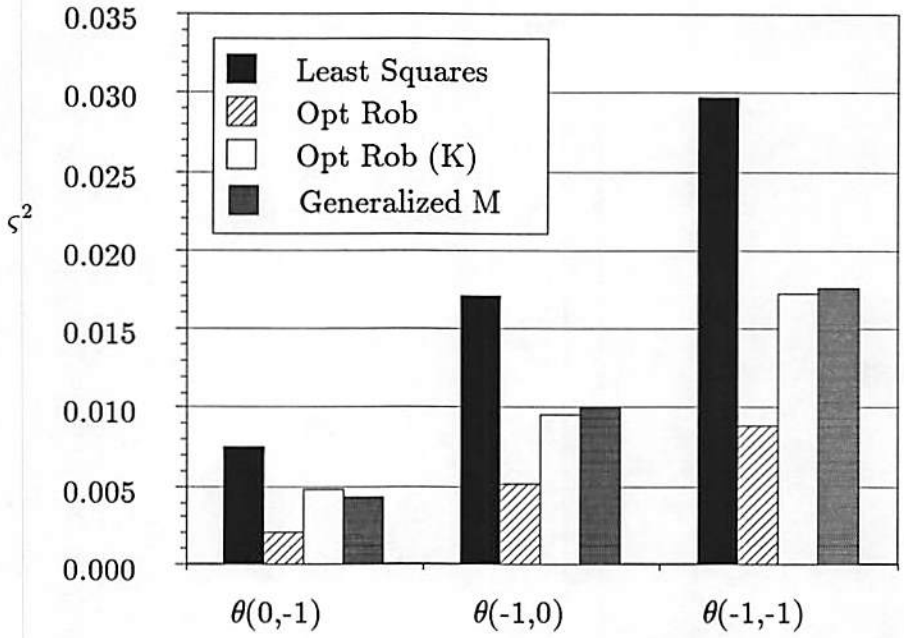


Figure 6.5: Average parameter estimate errors from uncorrelated additive outlier data.



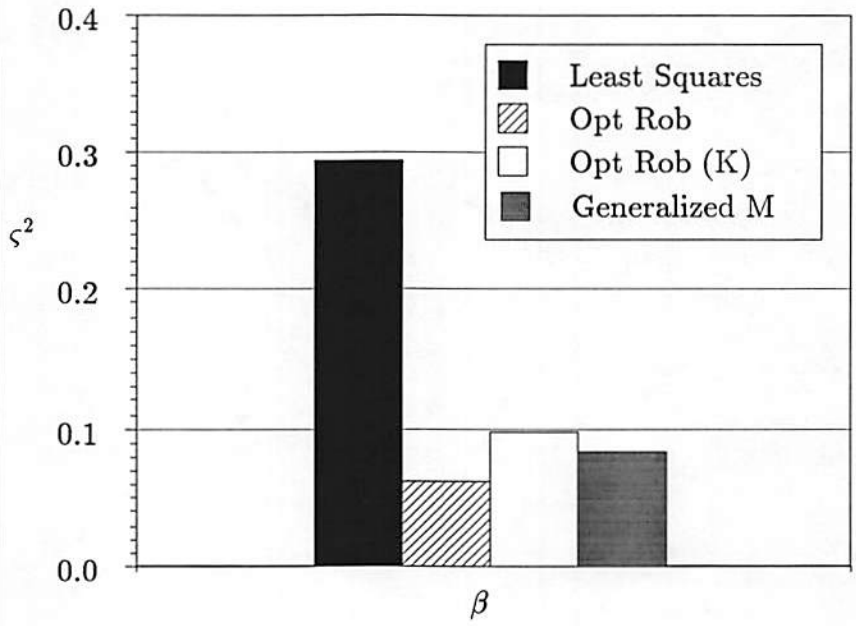


Figure 6.6: Average parameter estimate squared errors from uncorrelated additive outlier data.

Experiment 3: Data with additive outliers at correlated lattice cites were generated by creating an ensemble $E_c = \{\mathcal{Z}^k = \{z_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64\}$ of $0-1$ random data $z(\mathbf{s})$ with a binomial distribution such that

$$\mathrm{P}\{z(\mathbf{s}) = 1\} = .35 \cdot \frac{\sum_{\mathbf{r}\in A} a_{\mathbf{r}} z(\mathbf{s}+\mathbf{r})}{\sum_{\mathbf{r}\in A} |a_{\mathbf{r}}|} + 0.1$$

and the $a_{\mathbf{r}}$ and A are given in Table 6.5. Consequently, lattice sites at which $z(\mathbf{s}) = 1$ occured in clusters. The contaminated ensemble E_x was formed by using E_y, E_n, and E_c in (6.25). Least squares, OR, and OR(K) estimates were computed and errors calculated relative to $\boldsymbol{\theta}_{LS}$ and β^2_{LS}. The results are shown in Figures 6.7 and 6.8.

$a_{0,-1}$	$a_{0,-2}$	$a_{-1,0}$	$a_{-1,-1}$	$a_{-1,-2}$	$a_{-2,0}$	$a_{-2,-1}$	$a_{-2,-2}$
1.0	.6	1.0	.4	.3	.6	.3	.2

Table 6.5: Neighbor set and coefficients for correlated $0-1$ binomial data.

Experiment 4: In this experiment the contaminated data consisted of substitutive outliers at uncorrelated lattice sites. The data was created by using the data E_y, E_n, and E_u in (6.25). The computed errors of LS, OR, and OR(K) estimates relative to the LS estimates from E_y are shown in Figures 6.9 and 6.10.

Experiment 5: Here we formed and ensemble of data E_x with correlated substitutive outliers by using E_y, E_c, and E_n from the preceding experiments and computing (6.25). Least squares, OR, and OR(K) estimate errors are shown in Figures 6.11 and 6.12.

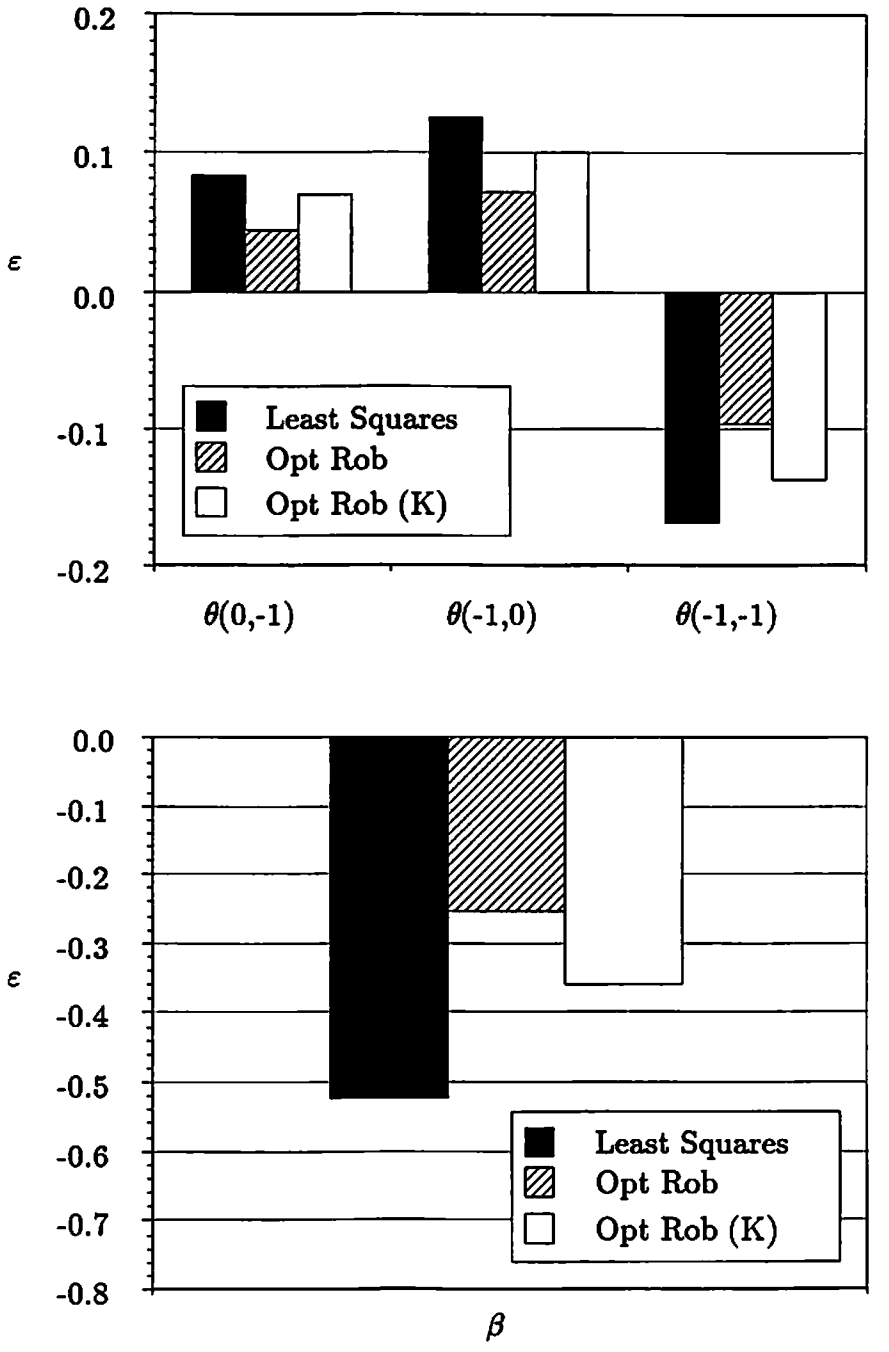


Figure 6.7: Average parameter estimate errors from correlated additive outlier data.

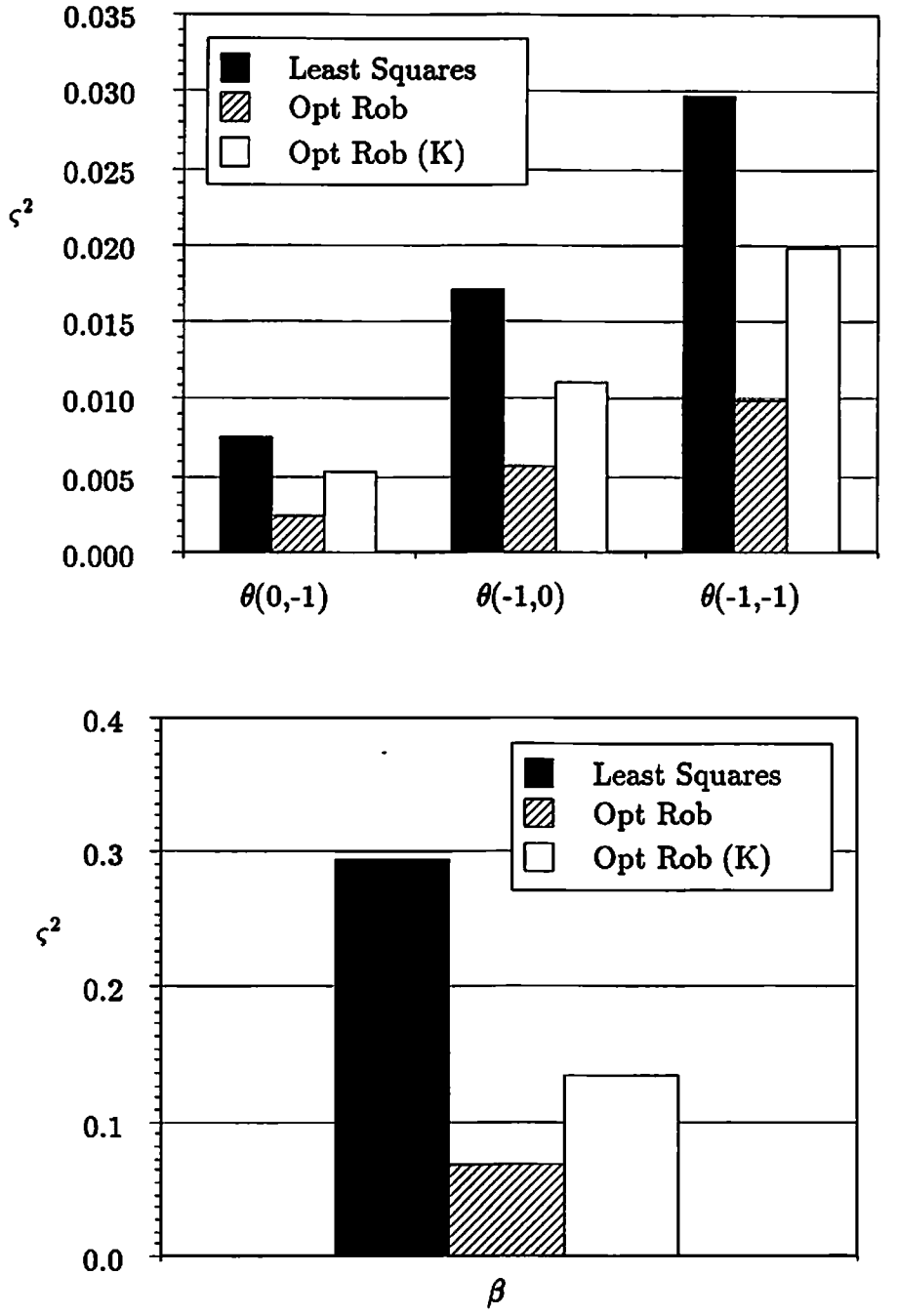


Figure 6.8: Average parameter estimate squared errors from correlated additive outlier data.

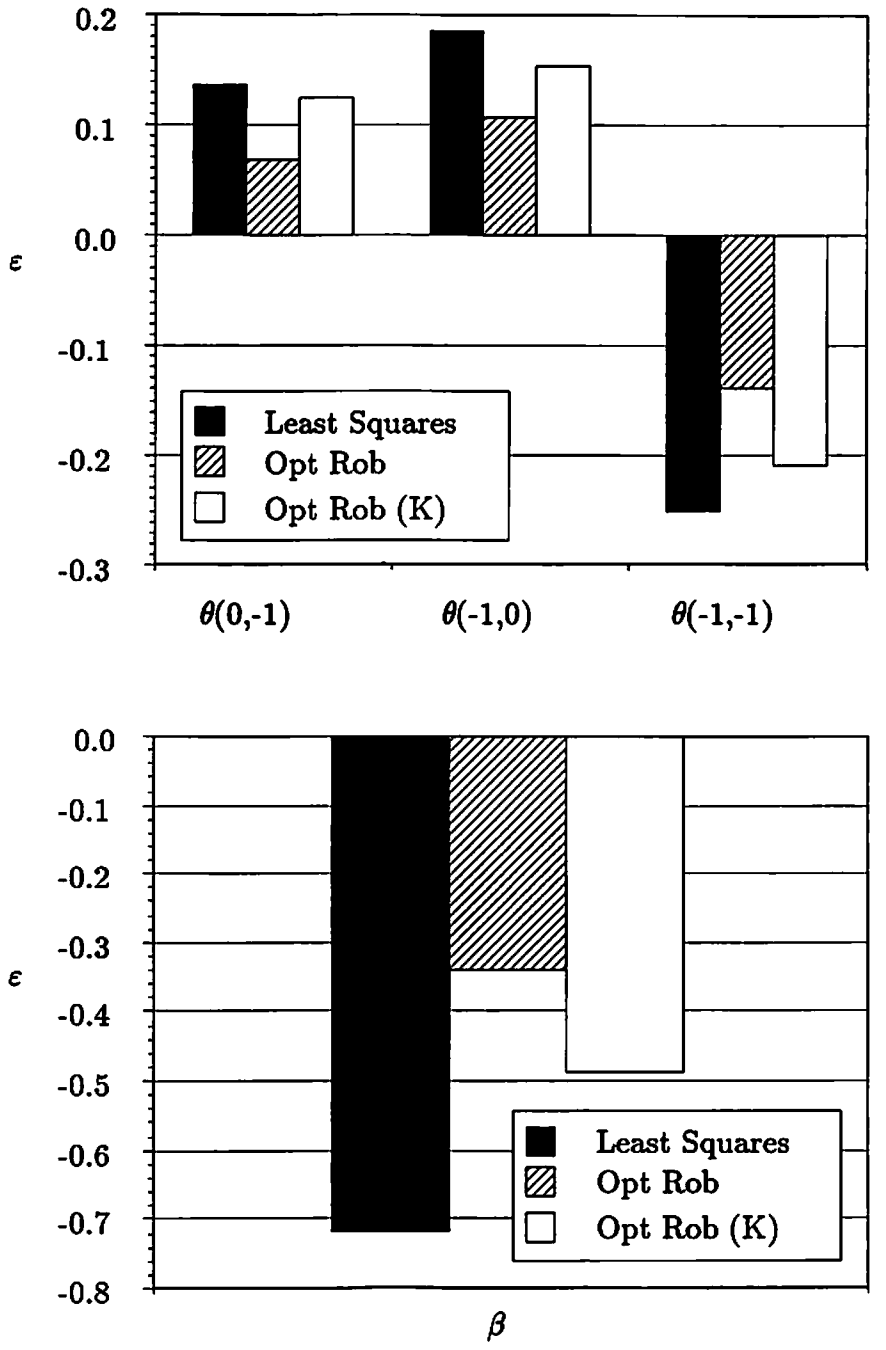


Figure 6.9: Average parameter estimate errors from uncorrelated substitutive outlier data.

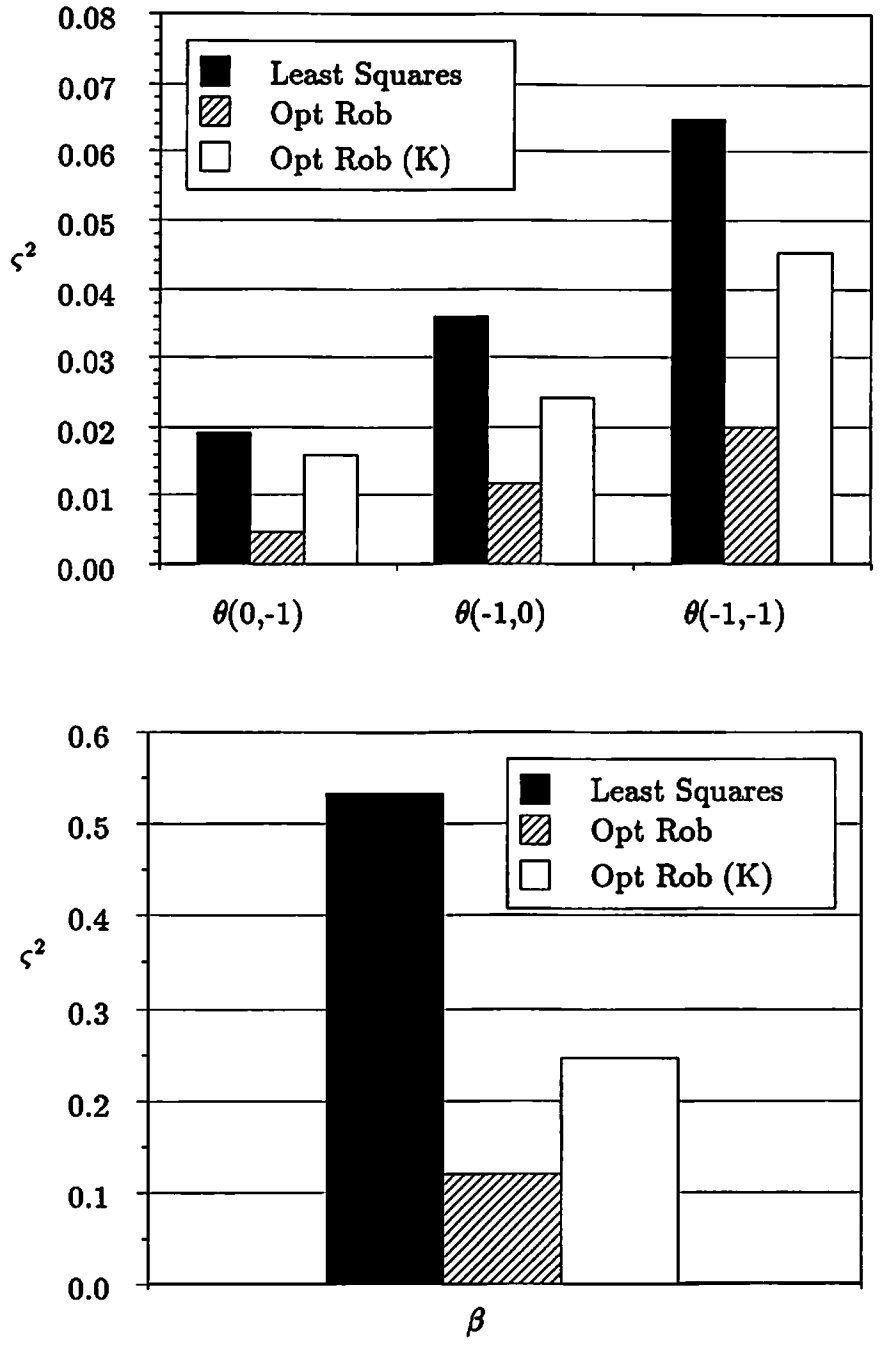


Figure 6.10: Average parameter estimate squared errors from uncorrelated substitutive outlier data.

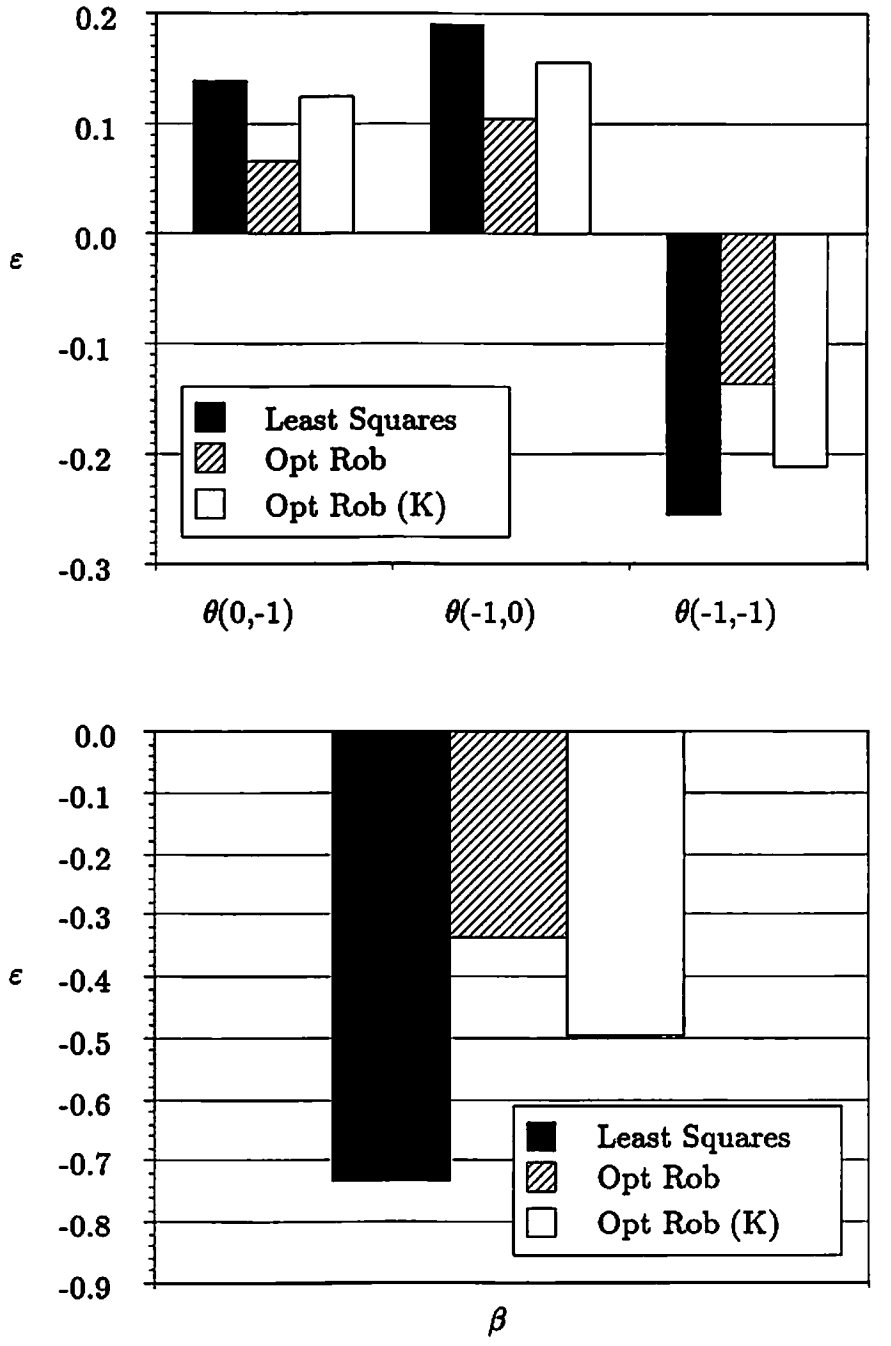


Figure 6.11: Average parameter estimate errors from correlated substitutive outlier data.

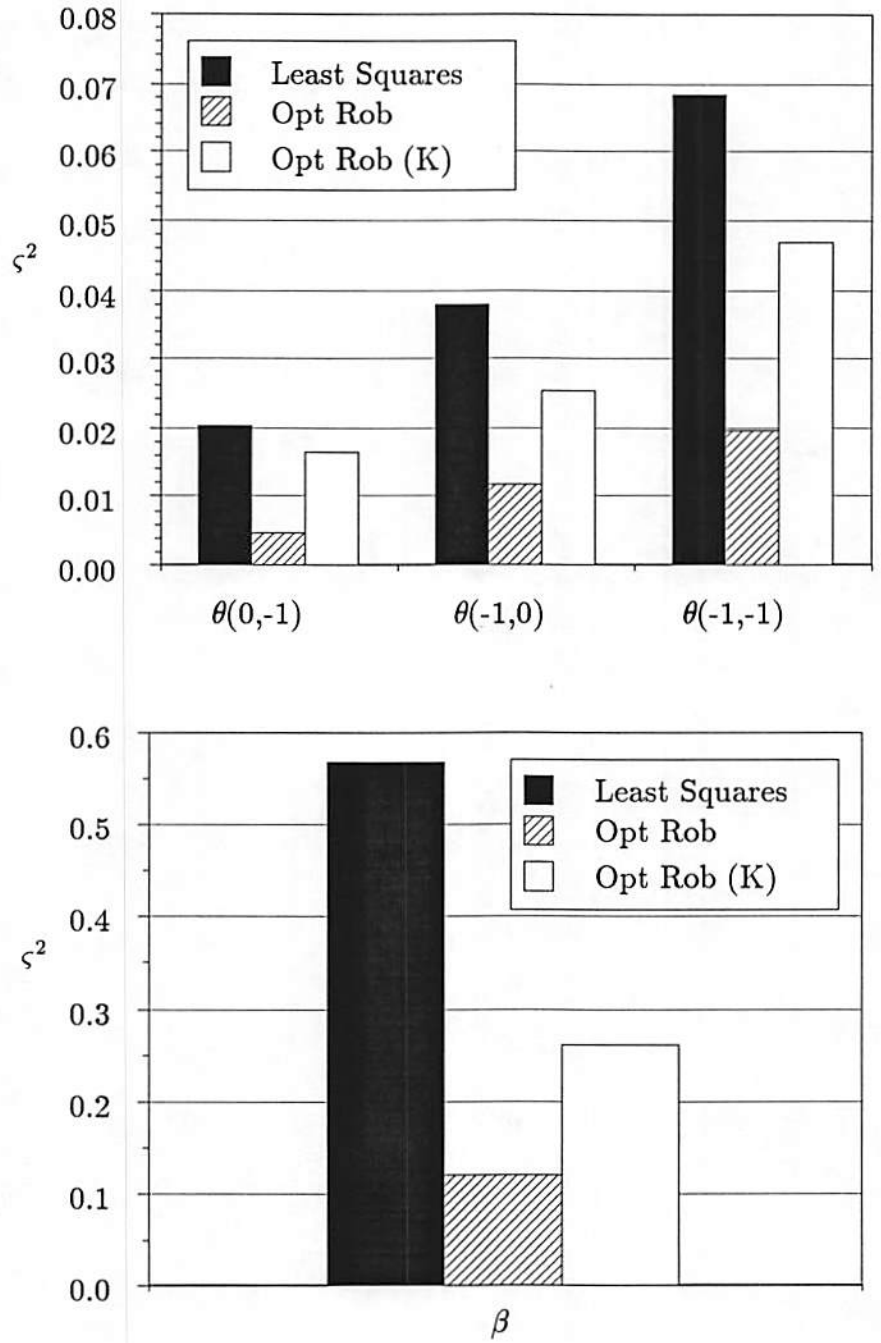


Figure 6.12: Average parameter estimate squared errors from correlated substitutive outlier data.

Chapter 7

Empirical Robust Estimators for Noncausal Models

In Chapter 6 a robust estimator with optimal properties was developed for a specific 2-D random field model, the NSHP autoregressive model. This estimator's structure depends directly on the unique statistical characteristics of the NSHP model and, in particular, on its one sided Markov property. In this chapter we suggest some robust estimators for noncausal models which seem justifiable in view of the developments in Chapters 5 and 6. These estimators do not have any theoretical underpinnings as does the optimal robust estimator for the NSHP model in Chapter 6, but are, more or less, based on the general theme that robust estimators are characterized by a mechanism which down-weights data which seem extreme relative to the aggregate data. Both the 2-D GM-estimator and the NSHP optimal robust (OR) estimator have this mechanism. We note that both the GM-estimator and the optimal robust estimator are essentially weighted LS algorithms whose weights depend directly on the data.

Although one can think of many schemes to accomplish this down-weighting, care must be taken in selecting a method since resulting estimates should be consistent and efficient, or nearly so. Thus, in this chapter we suggest several NCAR

and GMRF robust estimators, each developed from the ML estimator. The ML estimator is known to yield consistent, efficient and asymptotically normal estimates for NCAR and GMRF models when the data are Gaussian. The estimators we suggest can be categorized in two classes, characterized by the technique used to achieve robustness. The first class uses robust covariance estimates in place of sample covariance estimates, while the second class uses the idea that bounding the influence of outlying data produces robustness.

7.1 Robustifying the Likelihood Equation by Robust Covariances

Let β^2 be known and consider the ML equation (6.15) from Chapter 6, rewritten for the NCAR model using (6.11) and (6.13) with $\mathbf{x}_s = \text{col}\,\{x(\mathbf{s}+\mathbf{r})+x(\mathbf{s}-\mathbf{r}), \mathbf{r} \in N_s\}$:

$$\frac{1}{\beta^2} \sum_{\mathbf{s}\in\Omega_M} \left\{\mathbf{x}_s[x(\mathbf{s}) - \boldsymbol{\theta}^t\mathbf{x}_s] - \beta^2\mathbf{h}_\theta(\boldsymbol{\phi})\right\} = \mathbf{0}, \tag{7.1}$$

where

$$\begin{aligned} \mathbf{h}_\theta(\boldsymbol{\phi}) &= \frac{2}{\beta^2(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\boldsymbol{\omega},\boldsymbol{\phi})\mathbf{c}_\omega[1 - 2\boldsymbol{\theta}^t\mathbf{c}_\omega]d\boldsymbol{\omega} \qquad (7.2)\\ \mathcal{S}_x(\boldsymbol{\omega},\boldsymbol{\phi}) &= \frac{\beta^2}{|1 - 2\boldsymbol{\theta}^t\mathbf{c}_\omega|^2} \end{aligned}$$

and $\mathbf{c}_\omega = \text{col}\,\{\cos(\boldsymbol{\omega}^t\mathbf{r}), \mathbf{r} \in N_s\}$. Here, we use the asymmetric neighbor set N_s and enforce the requirement that for every $\mathbf{r} \in N$ implies $-\mathbf{r} \in N$ and that $\theta_\mathbf{r} = \theta_{-\mathbf{r}}$ in order that the parameters be identifiable [16].

Let us define sample covariances

$$c_x(\mathbf{r}) = \frac{1}{M^2}\sum_{\mathbf{s}\in\Omega_M} 2x(\mathbf{s})[x(\mathbf{s}+\mathbf{r}) + x(\mathbf{s}-\mathbf{r})] \tag{7.3}$$

$$c_x(\mathbf{r},\mathbf{t}) = \frac{1}{M^2}\sum_{\mathbf{s}\in\Omega_M} [x(\mathbf{s}+\mathbf{r}) + x(\mathbf{s}-\mathbf{r})][x(\mathbf{s}+\mathbf{t}) + x(\mathbf{s}-\mathbf{t})], \tag{7.4}$$

the neighbor set sample covariance matrix

$$C_x = \text{mat}\,\{c_x(\mathbf{r},\mathbf{t}), \mathbf{r}\&\mathbf{t} \in N_s\}$$

and the neighbor set sample covariance vector

$$\mathbf{c}_x = \text{col}\,\{c_x(\mathbf{r}), \mathbf{r} \in N_s\}.$$

Then, using these sample covariance expressions in (7.1), the ML equation for the parameter $\boldsymbol{\theta}$ is

$$C_x\boldsymbol{\theta} - \frac{1}{2}\mathbf{c}_x + \beta^2\mathbf{h}_\theta(\boldsymbol{\phi}) = \mathbf{0}. \tag{7.5}$$

Equation (7.5) suggests that C_x and $\mathbf{c}_x$ must be accurate estimates of the true underlying NCAR asymmetric covariances, $Q_x = \text{E}\,\{\mathbf{x}_s\mathbf{x}_s^t\}$ and $\mathbf{q}_x = \text{E}\,\{2\mathbf{x}_s x(\mathbf{s})\}$, for the ML estimate of $\boldsymbol{\theta}$ to be accurate. In the presence of outliers, if (7.3) and (7.4) are used to compute the sample covariances poor results may be obtained. We suggest that robust estimates be used in place of C_x and $\mathbf{c}_x$ in (7.5). Robust estimators of covariance and correlation matrices are adequately discussed in [38, Chapter 5] and [39, Chapter 8]. For our purposes here, we suggest one simple estimator which requires only robust scale estimates. Let x_1 and x_2 be random variables with zero mean and finite variance. Then,

$$\text{cov}\{x_1, x_2\} = \frac{1}{4ab}\left[\text{V}\,\{ax_1 + bx_2\} - \text{V}\,\{ax_1 - bx_2\}\right]. \tag{7.6}$$

If $x_1 = x(\mathbf{s}+\mathbf{r}) + x(\mathbf{s}-\mathbf{r})$ and $x_2 = x(\mathbf{s}+\mathbf{t}) + x(\mathbf{s}-\mathbf{t})$, then $c_x(\mathbf{r},\mathbf{t})$ can be computed from (7.6). By replacing $\text{V}\,\{\cdot\}$ with a robust alternative, $S\{\cdot\}$, (7.6) becomes a robust estimator of the covariance which we denote by $c_x^*(\mathbf{r},\mathbf{t})$. Huber suggests that $a = \dfrac{1}{S\{x_1\}}$ and $b = \dfrac{1}{S\{x_2\}}$; in which case,

$$\frac{1}{4}\left[S^2\left(\frac{x_1}{S\{x_1\}} + \frac{x_2}{S\{x_2\}}\right) - S^2\left(\frac{x_1}{S\{x_1\}} - \frac{x_2}{S\{x_2\}}\right)\right]$$

is a robust correlation estimate which may not be limited to the interval $[-1, +1]$. Therefore, it is further suggested to use

$$r_x^*(\mathbf{r},\mathbf{t}) = \frac{S^2\left(\frac{x_1}{S\{x_1\}} + \frac{x_2}{S\{x_2\}}\right) - S^2\left(\frac{x_1}{S\{x_1\}} - \frac{x_2}{S\{x_2\}}\right)}{S^2\left(\frac{x_1}{S\{x_1\}} + \frac{x_2}{S\{x_2\}}\right) + S^2\left(\frac{x_1}{S\{x_1\}} - \frac{x_2}{S\{x_2\}}\right)},$$

then compute the robust covariances from

$$c_x^*(\mathbf{r},\mathbf{t}) = r_x^*(\mathbf{r},\mathbf{t})S\{x(\mathbf{s}+\mathbf{r}) + x(\mathbf{s}-\mathbf{r})\}S\{x(\mathbf{s}+\mathbf{t}) + x(\mathbf{s}-\mathbf{t})\}. \tag{7.7}$$

Several possible choices exits for the scale estimator. For example, $S\{\cdot\}$ can be computed as the median absolute deviation (MAD) estimate [39]

$$S\{x_i\} = \frac{\text{median}\,\{|x_i - \text{median}\,\{x_i\}\,|\}}{.6745}, \tag{7.8}$$

or $S\{\cdot\}$ can be computed from Huber's "Proposal 2" [39] by solving

$$\sum_{\mathbf{s}\in\Omega_M}\left[\min\left\{b^2, \frac{x_i^2(\mathbf{s})}{S^2}\right\} - \alpha_H\right] = 0 \tag{7.9}$$

for S, where α_H is given in Table 6.3 of Chapter 6. M-estimators and optimal robust estimators for the covariances also exist. See [38] for details.

We note that equation (7.5) is independent of β^2. This suggests that the estimates of $\boldsymbol{\theta}$ and β^2 can be determined independently without affecting the other. With this in mind we have the following algorithm for determining a robust estimate of $\boldsymbol{\phi} = \text{col}\,\{\boldsymbol{\theta}, \beta^2\}$ for the NCAR model.

NCAR Robust Algorithm 1:

1. Compute C_x^* and $\mathbf{c}_x^*$ using (7.7) and a robust estimator of scale.

2. Solve

$$C_x^*\hat{\boldsymbol{\theta}} - \frac{1}{2}\mathbf{c}_x^* + \hat{\beta}^2\mathbf{h}_\theta(\hat{\boldsymbol{\phi}}) = 0$$

and

$$\sum_{\mathbf{s}\in\Omega_M} \left[\min\{b^2, w^2(\mathbf{s})\} - \alpha_H\right] = 0$$

for $\hat{\boldsymbol{\theta}}$ and $\hat{\beta}^2$, where $w(\mathbf{s}) = \frac{1}{\hat{\beta}}[x(\mathbf{s}) - \hat{\boldsymbol{\theta}}^t \mathbf{x}_\mathbf{s}]$.

A similar approach can be taken to develop a robust estimator of GMRF model parameters. The ML estimator equation for the parameter vector $\boldsymbol{\theta}$ of a GMRF model is

$$\frac{1}{2\nu} \sum_{\mathbf{s}\in\Omega_M} [\mathbf{x}_\mathbf{s} x(\mathbf{s}) - 2\nu \mathbf{h}_\theta(\phi)] = \mathbf{0}, \tag{7.10}$$

where

$$\begin{aligned} \mathbf{h}_\theta(\phi) &= \frac{1}{\nu(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi)\mathbf{c}_\omega d\omega \\ \mathcal{S}_x(\omega,\phi) &= \frac{\nu}{1 - 2\boldsymbol{\theta}^t \mathbf{c}_\omega} \end{aligned} \tag{7.11}$$

and $\mathbf{c}_\omega = \text{col}\,\{\cos(\omega^t \mathbf{r}), \mathbf{r} \in N_s\}$. Using the sample covariance definition (7.3) the GMRF ML equation for $\boldsymbol{\theta}$ can be written in the form

$$\frac{1}{2}\mathbf{c}_x - 2\nu \mathbf{h}_\theta(\phi) = \mathbf{0}.$$

This suggests that by replacing $\mathbf{c}_x$ with the robust $\mathbf{c}_x^*$ a robust estimator of $\boldsymbol{\theta}$ can be obtained. However, a simpler, explicit estimator is possible.

Let $\mathbf{q}_x(\boldsymbol{\theta})$ be determined from the inverse Fourier transform of (7.11), that is,

$$\frac{1}{4}\mathbf{q}_x(\boldsymbol{\theta}) = \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathcal{S}_x(\omega,\phi)\mathbf{c}_\omega d\omega.$$

Clearly, $\mathbf{q}_x(\boldsymbol{\theta})$ depends on the parameter vector $\boldsymbol{\theta}$ and is equal to the data covariance vector $\mathbf{q}_x = \mathrm{E}\{2\mathbf{x}_\mathbf{s} x(\mathbf{s})\}$ when $\boldsymbol{\theta}$ is the true value. Consequently, $\mathbf{h}_\theta(\phi) = \frac{1}{4\nu}\mathbf{q}_x(\boldsymbol{\theta})$. But we know for the GMRF model that $\frac{1}{2}\mathbf{q}_x = Q_\mathbf{x}\boldsymbol{\theta}$. Hence,

$\mathbf{h}_\theta(\phi) = \frac{1}{2\nu} Q_{\mathbf{x}}(\theta)\theta$ and the GMRF ML equation can be rewritten and solved for θ to yield

$$\theta = \frac{1}{2} Q_x^{-1}(\theta)\mathbf{c}_x.$$

Whereupon, replacing $Q_x(\theta)$ with C_x yields the familiar LS estimator. Using the same rationale applied for the NCAR model robust estimator, the GMRF LS estimator is robustified by replacing the sample covariances with robust estimates of the covariances and obtaining

$$\theta = \frac{1}{2} C_x^{*-1}\mathbf{c}_x^*.$$

Two approaches can be used for estimating the residual variance ν. First, consider

$$\nu = \frac{1}{M^2} \sum_{\mathbf{s}\in\Omega_M} [x(\mathbf{s}) - \theta^t \mathbf{x}_\mathbf{s}]^2,$$

an estimator proposed by Chellappa [33]. For the robust case this estimator suggests that we again use Huber's "Proposal 2", i.e.

$$\sum_{\mathbf{s}\in\Omega_M} \left[\min\{b^2, w^2(\mathbf{s})\} - \alpha_H\right] = 0 \tag{7.12}$$

where $w(\mathbf{s}) = \frac{1}{\sqrt{\nu}}[x(\mathbf{s}) - \theta^t \mathbf{x}_\mathbf{s}]$.

Alternatively, we can robustify the ML equation for ν,

$$\frac{1}{2\nu^2} \sum_{\mathbf{s}\in\Omega_M} \left\{x(\mathbf{s})[x(\mathbf{s}) - \theta^t \mathbf{x}_\mathbf{s}] - 2\nu^2 h_\nu(\phi)\right\} = 0$$

where $h_\nu(\phi) = \frac{1}{2\nu}$. In terms of the sample covariances,

$$\frac{1}{4} c_x(0) - \frac{1}{2}\theta^t \mathbf{c}_x - \nu = 0.$$

By replacing sample covariances with robust estimates of the covariances we obtain

$$\nu = \frac{1}{2}\left[\frac{1}{2} c_x^*(0) - \theta^t \mathbf{c}_x^*\right].$$

A robust algorithm for computing robust estimates of $\phi = \text{col}\,\{\theta, \nu\}$ for the GMRF model is proposed.

GMRF Robust Algorithm 1:

1. Compute C_x^* and $\mathbf{c}_x^*$ using (7.7) and a robust estimator of scale.

2. Compute

$$\hat{\theta} = \frac{1}{2} C_x^{*-1} \mathbf{c}_x^*$$

and

$$\hat{\nu} = \frac{1}{2}\left[\frac{1}{2} c_x^*(0) - \hat{\theta}^t \mathbf{c}_x^*\right].$$

7.2 Robustification of the Likelihood Function

The influence function for estimators which may be expressed as functionals have especially nice properties for NSHP models. For example, the error covariance matrix for the NSHP model estimator can be expressed as simply the expected value of the outer product of the influence function, viz.,

$$V(T, \mathcal{F}^{N_0}) = \int IF(\mathbf{x}^0, \phi) IF^t(\mathbf{x}^0, \phi) d\mathcal{F}^{N_0}.$$

This is a consequence of the fact that the NSHP random field possesses a one-sided Markov property and the conditional mean (given all past observations) of the influence function is zero. For NCAR and GMRF model estimators, the asymptotic error covariance matrix in terms of the influence function does not reduce to a similar expression and the conditional mean (given all other observations) of the influence function is not necessarily zero. Consequently, the asymptotic error covariance matrix for the NCAR and GMRF estimators is

$$V(T, \mathcal{F}^{N_0}) = \lim_{M \to \infty} \frac{1}{M^2} \sum_{s \in \Omega_M} \sum_{t \in \Omega_M} \int IF(\mathbf{x}_s^0, \phi) IF^t(\mathbf{x}_t^0, \phi) d\mathcal{F}^{N_0}.$$

and a simpler expression does not seem to exist. Solutions to Hampel's optimality problem for the NCAR and GMRF model seem remote. However, in analogy with the NSHP case, it seems reasonable to expect that a bound placed on the bias, or equivalently, an influence function, will result in an estimator yielding parameter estimates closer than ML estimates to the true underlying model values when confronted with data containing outliers. No optimal property regarding the error covariances can be expected.

For instance, let us consider the NCAR model with β^2 known and specializing to the case when $\mathbf{r} \in N$ implies $-\mathbf{r} \in N$. An influence function for the estimator of the NCAR model parameter vector θ of the form (6.24) is

$$IF(\mathbf{x}^0, \theta) = \Gamma^{-1}(\theta)[\kappa_\theta(\mathbf{x}^0, \phi) - \mathbf{h}_\theta(\phi)], \tag{7.13}$$

where $\kappa_\theta(\mathbf{x}^0, \phi)$ and $\mathbf{h}_\theta(\phi)$ are given by (6.11) and (6.13), respectively, and $\mathbf{x}_\mathbf{s} = \text{col}\,\{x(\mathbf{s}+\mathbf{r})+x(\mathbf{s}-\mathbf{r}), \mathbf{r} \in N_s\}$. Obviously, outliers in the data affect only $\kappa_\theta(\mathbf{x}^0, \phi)$ and not $\mathbf{h}_\theta(\phi)$. So, instead of $\kappa_\theta(\mathbf{x}^0, \phi)$ use some robust alternative, $\kappa^*_\theta(\mathbf{x}^0, \phi)$, so that when the data are truly a Gaussian NCAR random field $\kappa^*_\theta(\mathbf{x}^0, \phi)$ looks like $\kappa_\theta(\mathbf{x}^0, \phi)$; but when $\mathbf{x}^0$ contains outliers, $\kappa^*_\theta(\mathbf{x}^0, \phi) \leq \kappa_{MAX} < \infty$.

Therefore, as suggested by the NSHP OR estimator we propose the following robust estimator for the NCAR model parameters.

NCAR Algorithm 2:

1. Solve

$$\sum_{\mathbf{s} \in \Omega_M} \left[\frac{1}{\hat{\beta}} \mathbf{x}_\mathbf{s} w(\mathbf{s}) \cdot W_{b_1}(d_{\mathbf{x}^0_\mathbf{s}}) - \mathbf{h}_\theta(\hat{\phi}) \right] = 0$$

and

$$\sum_{\mathbf{s} \in \Omega_M} \left[\min \left\{ b_2^2, w^2(\mathbf{s}) \right\} - \alpha_H \right] = 0$$

for $\hat{\theta}$ and $\hat{\beta}^2$, where $w(\mathrm{s}) = \frac{1}{\hat{\beta}}[x(\mathrm{s}) - \hat{\theta}^t \mathrm{x_s}]$ and

$$\begin{aligned} W_{b_1}(z) &= \min\left\{1, \frac{b_1}{|z|}\right\} \\ d^2_{\mathrm{x_s^0}} &= \mathrm{x_s^t} Q_x^{-1} \mathrm{x_s} \\ Q_x &= \mathrm{E}\left\{\mathrm{x_s x_s^t}\right\}. \end{aligned}$$

As a second alternative, one can robustify the likelihood function directly. Consider the NCAR negative log-likelihood function rewritten in the form

$$l_M(\phi) = \sum_{\mathrm{s}\in\Omega_M} [H(\phi) + \rho_{ML}(\mathrm{x_s^0}; \phi)], \tag{7.14}$$

where

$$H(\phi) = \frac{1}{2(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log[S_x(\omega, \phi)] d\omega$$

and

$$\rho_{ML}(\mathrm{x_s^0}; \phi) = \frac{1}{2\beta^2}[x(\mathrm{s}) - \theta^t \mathrm{x_s}]^2$$

and where the constant terms not involving the parameters have been neglected. Clearly, only the last term of (7.14), having the functional form $\rho_{ML}(w) = \frac{1}{2}w^2$ for the ML function, involves the possibly contaminated data which can have unlimited influence on $l_M(\phi)$.

As a robustified negative log-likelihood function consider,

$$l^*_M(\phi) = \sum_{\mathrm{s}\in\Omega_M} [H(\phi) + \rho(\mathrm{x_s^0}; \phi)] \tag{7.15}$$

where $\rho(\mathrm{x_s}; \phi)$ is a function which looks like $\rho_{ML}(\mathrm{x_s}; \phi)$ if the data are truly a Gaussian NCAR random field but limits the influence of outliers.

Letting $\psi(\mathrm{x_s^0}; \phi) = \frac{\partial}{\partial \phi}\rho(\mathrm{x_s^0}; \phi)$, the likelihood equation becomes

$$\lambda^*_M(\phi) = \sum_{\mathrm{s}\in\Omega_M} \left[\psi(\mathrm{x_s^0}; \phi) - \mathrm{h}_\theta(\phi)\right] \tag{7.16}$$

with an influence function

$$IF^*(\mathbf{x}^0;\phi) \sim \psi(\mathbf{x}^0;\phi) - \mathbf{h}_\theta(\phi). \tag{7.17}$$

Therefore, choose $\rho(\mathbf{x}^0;\phi)$ such that $\psi(\mathbf{x}^0;\phi)$ yields a bounded $IF(\mathbf{x}^0;\phi)$ as a function of the data.

In keeping with the general rules of robustness, $\psi(w)$ should be a continuous function so that the local-shift sensitivity λ^* (see Definition 2.3, Chapter 2) is small and the gross-error sensitivity γ^* (see Definition 2.2, Chapter 2) is bounded. Thus, if we let $\psi(w) = \dfrac{\partial}{\partial\phi}\psi(w)$ where $\int_{-\pi}^{\pi}\psi(w) = \frac{d}{dw}\rho(w)$, then choose

$$\psi(w) = \begin{cases} c, & w > c \\ w, & |w| \le c \\ -c, & w < -c \end{cases} . \tag{7.18}$$

Integrating this ψ-function yields

$$\rho(w) = \begin{cases} 2cw - \frac{c^2}{2}, & w > c \\ \frac{1}{2}w^2, & |w| \le c \\ -2cw - \frac{c^2}{2}, & w < -c \end{cases} . \tag{7.19}$$

When using this $\rho(w)$ extremely large outliers still affect $l_M^*(\phi)$. For additional robustness a ψ-function which returns to zero could be used, yielding a rejection point ρ^* which is finite (see Definition 2.4, Chapter 2). In this case, possible candidates for $\psi(w)$ and $\rho(w)$ are

$$\psi(w) = \begin{cases} 0, & w > c \\ w(c^2 - w^2)^2, & |w| \le c \\ 0, & w < -c \end{cases} \tag{7.20}$$

and

$$\rho(w) = \begin{cases} \frac{c^6}{6}, & w > c \\ \frac{w^2}{2}\left(c^4 - c^2w^2 + \frac{w^4}{3}\right), & |w| \leq c \\ \frac{c^6}{6}, & w < -c \end{cases} . \tag{7.21}$$

Thus, we have the following algorithm to compute robust estimates of the NCAR model parameters.

NCAR Algorithm 3:

1. Minimize with respect to θ and β

$$l_M^*(\phi) = \sum_{s \in \Omega_M} [H(\phi) + \rho(x_s^0; \phi)],$$

where $\rho(w)$ is given by (7.19) or (7.21).

7.3 Experimental Results

We conducted several experiments to compare the robust estimators with the classical ML and LS estimators for noncausal models.

Experiment 1: We synthesized 64 sets, $\mathcal{Y}^k = \{y_k(s), s \in \Omega_{32}\}, k = 1, \ldots, 64$, of 2nd-order Gaussian NCAR data on a 32×32 square toroidal lattice using the procedure in [16]. The parameter values for the NCAR model are given in Table 7.1. To create contaminated data we generated 64 sets, $\mathcal{N}^k = \{n_k(s), s \in \Omega_{32}\}, k = 1, \ldots, 64$, of Gaussian noise with $\mathrm{E}\{n(s)\} = 0.0$ and $\mathrm{E}\{n^2(s)\} = 25.0$ and 64 sets, $\mathcal{Z}^k = \{z_k(s), s \in \Omega_{32}\}, k = 1, \ldots, 64$, of uncorrelated $0 - 1$ random variates with $\mathrm{P}\{z(s) = 1)\} = 0.02$.

We then computed

$$x_k(s) = x_k(s) + z_k(s) \cdot n_k(s) \tag{7.22}$$

$\theta_{1,0}$	$\theta_{0,1}$	$\theta_{1,1}$	$\theta_{-1,1}$	β^2
.19450	.05710	−.13600	.23470	1.00000

Table 7.1: Neighbor set and coefficients for the NCAR model.

for $\mathbf{s} \in \Omega_{32}$ and $k = 1, \ldots, 64$ to yield 64 sets, $\mathcal{X}^k = \{x_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64$, of 2 percent additive outlier (AO) contaminated data. The ML estimator and NCAR Robust Algorithm 1 with $b = 2.5$ were used to compute parameter estimates from each set $\mathcal{Y}^k$ and $\mathcal{X}^k$ of data. The results are shown in Figure 7.1 and Table 7.2, where the average errors

$$\varepsilon_r = \frac{1}{64} \sum_{k=1}^{64} \left(\hat{\theta}_{\mathbf{r}}^k - \theta_{\mathbf{r},ML}^k \right), \quad \mathbf{r} \in N_s \tag{7.23}$$

and the average squared errors

$$\varsigma_r^2 = \frac{1}{64} \sum_{k=1}^{64} \left(\hat{\theta}_{\mathbf{r}}^k - \theta_{\mathbf{r},ML}^k \right)^2, \quad \mathbf{r} \in N_s, \tag{7.24}$$

relative to the ML estimates, are compared for the ensembles of clean data, $E_y = \{\mathcal{Y}^k, k = 1, \ldots, 64\}$, and contaminated data, $E_x = \{\mathcal{X}^k, k = 1, \ldots, 64\}$.

	ML - AO Data	Robust - Clean Data	Robust - AO Data
ε_β	−0.68920	0.00022	−0.22183
ς_β^2	0.54389	0.00011	0.05526

Table 7.2: NCAR sample average error and squared error of the parameter β^2.

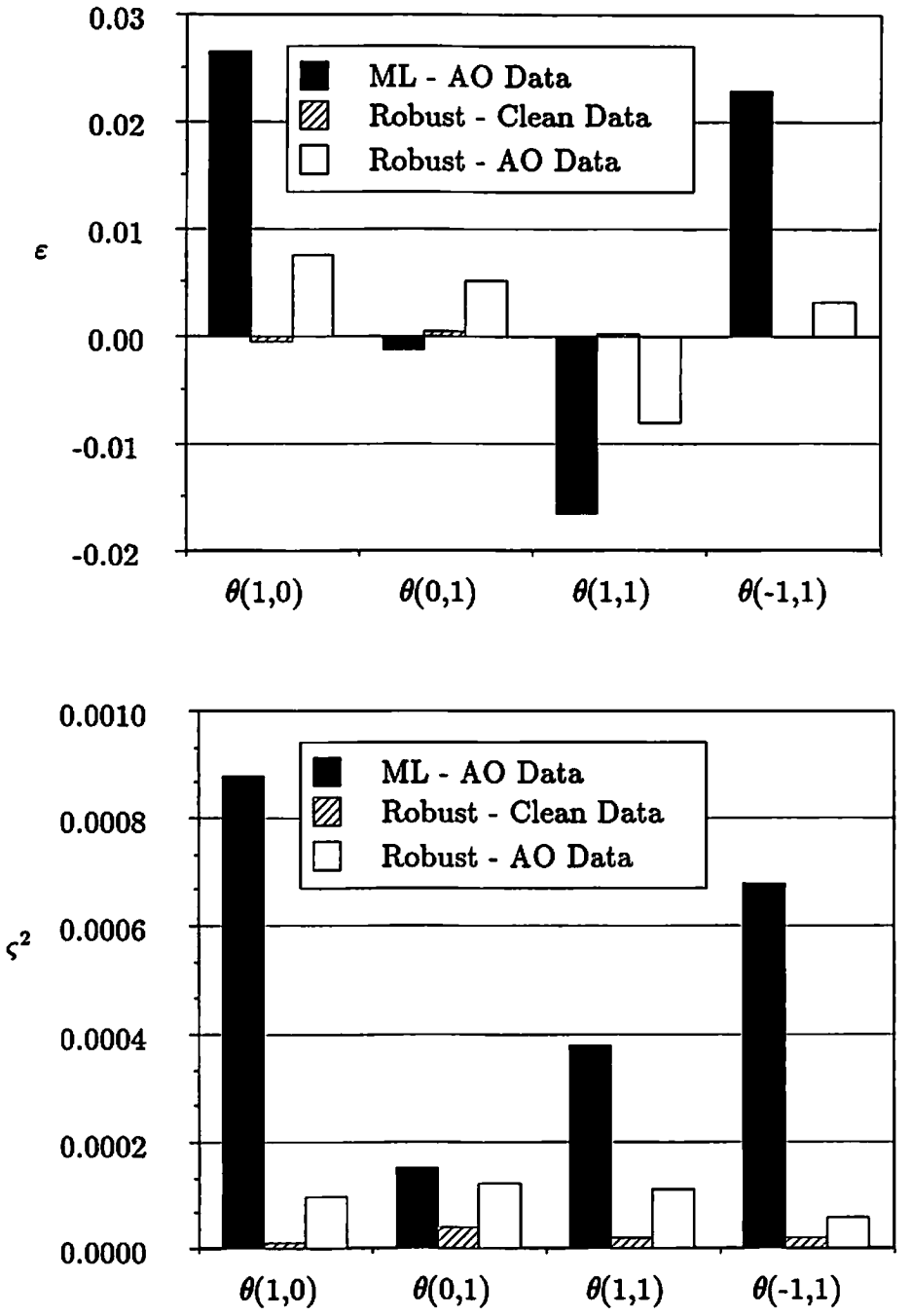


Figure 7.1: NCAR sample average error and squared error of the parameter θ_r.

Experiment 2: As in Experiment 1 we generated an ensemble of clean GMRF data, $E_y = \{\mathcal{Y}^k = \{y_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64\}$, with parameter values given in Table 7.3, and an ensemble of 2 percent additive outlier contaminated data, $E_x = \{\mathcal{X}^k = \{x_k(\mathbf{s}), \mathbf{s} \in \Omega_{32}\}, k = 1, \ldots, 64\}$, using (7.22) with $\mathrm{E}\{n(\mathbf{s})\} = 0.0$ and $\mathrm{E}\{n^2(\mathbf{s})\} = 5.0$. The average errors (7.23) and average squared errors (7.24) of $\boldsymbol{\theta}$ and ν using LS and robustified LS estimates, GMRF Algorithm 1, are shown in Figure 7.2 and Table 7.4. The robust covariance estimates $\mathbf{c}_x^*$ and C_x^* were computed using (7.7) and Huber's "Proposal 2" (7.9) [39] with $b = 2.5$ for the robust scale estimates.

(1,0)	(0,1)	(1,1)	(−1,1)	ν
.19500	.05000	−.13500	.10100	1.00000

Table 7.3: Neighbor set and coefficients for the GMRF model.

	LS - AO Data	Robust - Clean Data	Robust - AO Data
ε_ν	−0.11596	−0.00054	−0.06325
ς_ν^2	0.01641	0.00007	0.00468

Table 7.4: GMRF sample average error and squared error of the parameter ν.

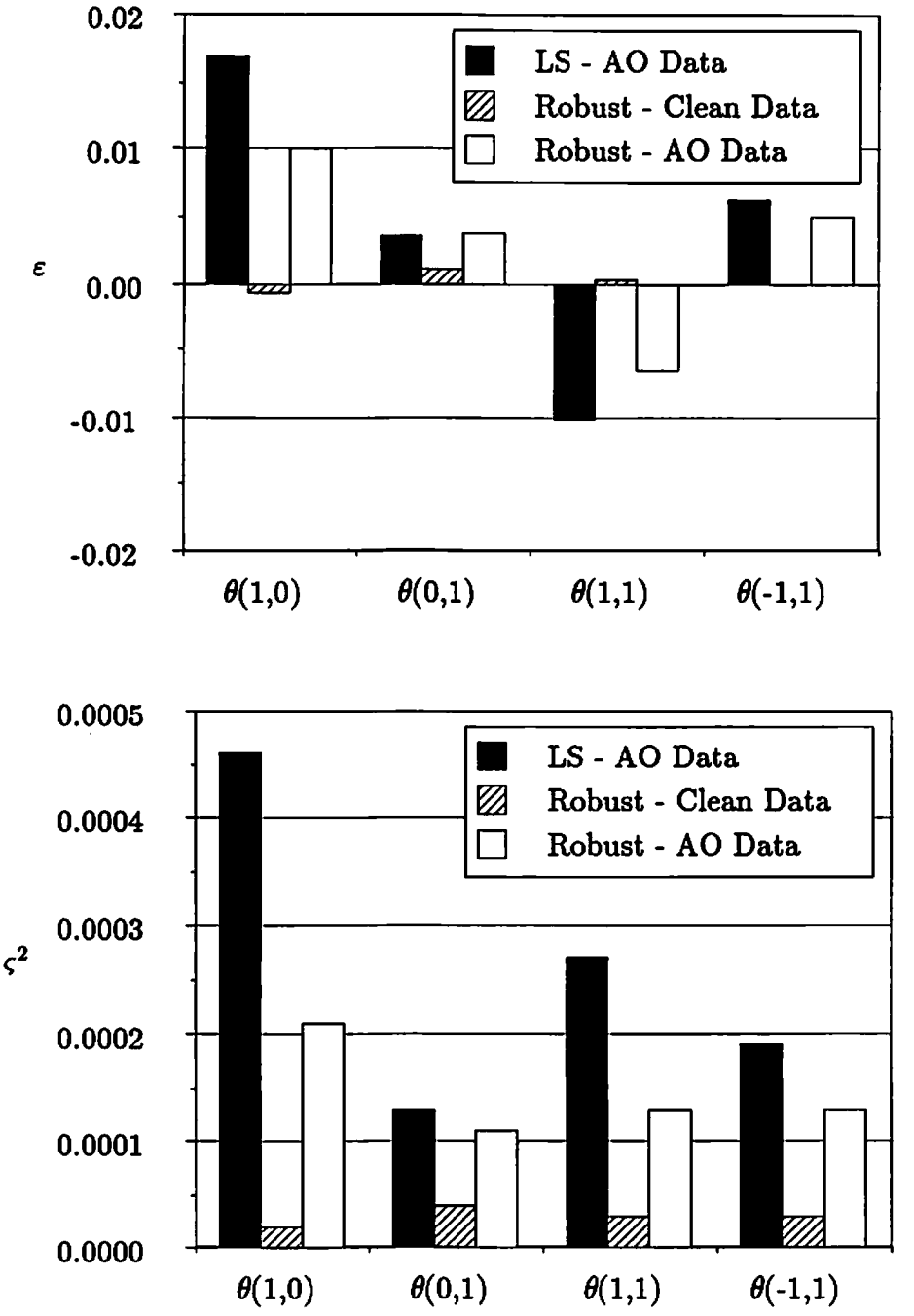


Figure 7.2: GMRF sample average error and squared error of the parameter $\theta_{\mathbf{r}}$.

Experiment 3: To evaluate the statistics of the estimates from NCAR Algorithm 2 we used the clean and contaminated data ensembles from Experiment 1. The average errors (7.23) and average sqaured errors (7.24) of the parameter estimates using this algorithm are compared in Table 7.5 and Figure 7.3. The cutoff constants were $b_1 = 2.5$ and $b_2 = 2.0$.

	ML - AO Data	Robust - Clean Data	Robust - AO Data
ε_β	−0.11596	−0.00054	−0.06325
ς_β^2	0.01641	0.00007	0.00468

Table 7.5: NCAR sample average error and squared error of the parameter β^2.

Experiment 4: NCAR Algorithm 3 was evaluated similarly. Using the clean and contaminated ensembles of data from Experiment 1, estimates of θ and β^2 were made and the error statistics computed and compared with the ML estimates. The results are shown in Figure 7.4 and Table 7.6. Here, we use the ρ-function, given by (7.19), corresponding to a non-redescending ψ-function. The cutoff constant was $c = 1.5$.

	ML - AO Data	Robust - Clean Data	Robust - AO Data
ε_β	−0.11596	−0.00054	−0.06325
ς_β^2	0.01641	0.00007	0.00468

Table 7.6: NCAR sample average error and squared error of the parameter β^2.

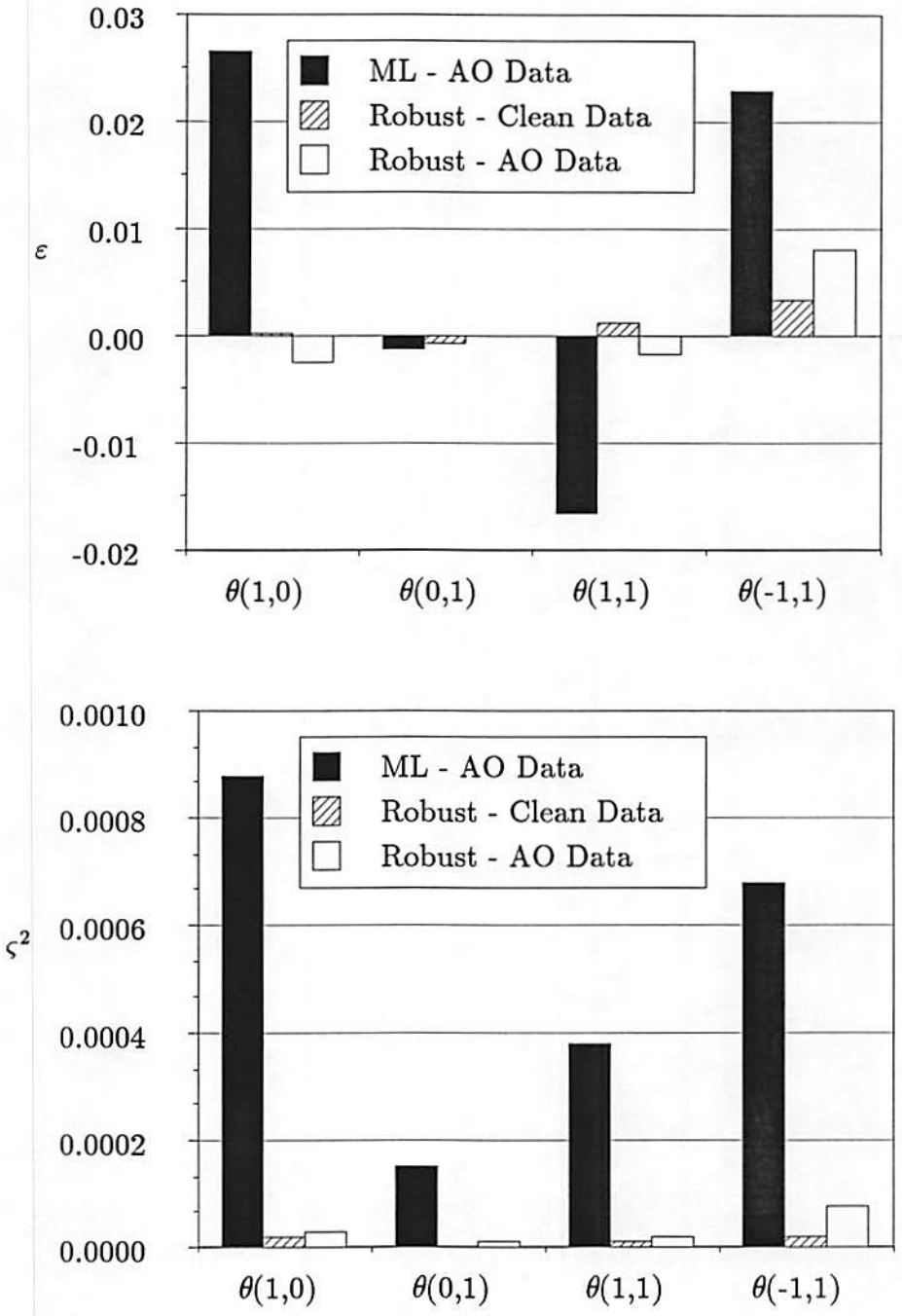


Figure 7.3: NCAR sample average error and squared error of the parameter $\theta_{\mathbf{r}}$.

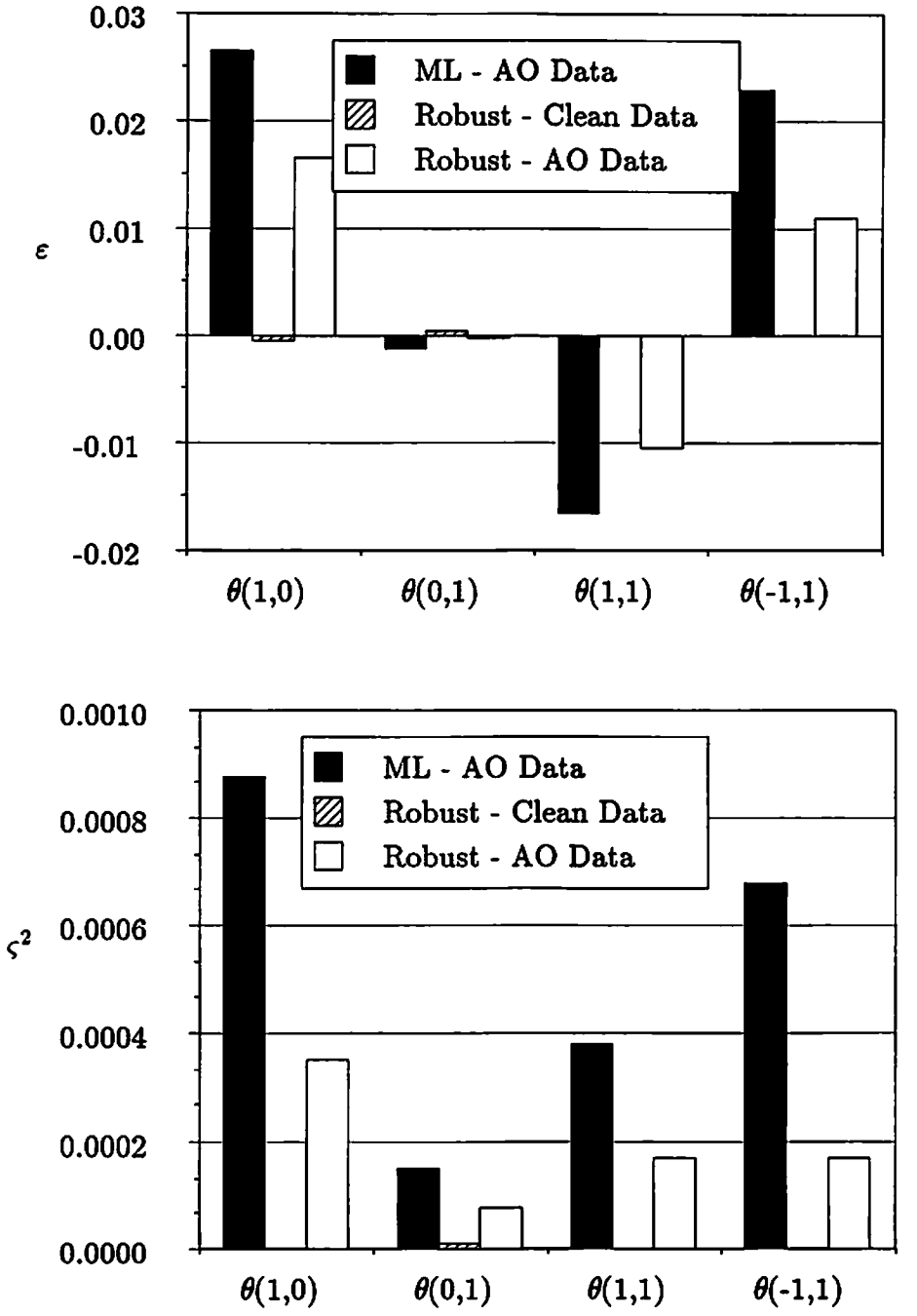


Figure 7.4: NCAR sample average error and squared error of the parameter θ_r.

Summary. The results speak for themselves. In every case the robust estimators perform well on uncontaminated data and do much better than the optimal Gaussian estimator when the data are contaminated. The robust estimates have smaller sample bias and sample variance then the nonrobust estimates in very experiment except for the estimate of $\theta_{0,1}$ from NCAR Algorithm 1, where the robust $\hat{\theta}_{0,1}$ sample error is slightly larger than the non-robust estimate.

Chapter 8

Conclusions and Topics for Future Research

8.1 Conclusions

Two theoretical approaches have been researched to solve the 2-D spectrum estimation from contaminated data problem. These approaches can be classified according to the amount of information available regarding the distribution of the observed data and the structure of the contamination process.

If the distribution functional form is fairly well established as Gaussian and the contaminating noise, if present, is additive to the signal of interest, then the NCAR plus noise model provides superior estimated spectra when compared with the conventional NCAR only model. We have shown that the NCAR plus noise model has practical application to the important planar array processing problem of direction finding. Moreover, the theoretical properties of the NCAR plus noise model ML estimator include optimality in Gaussian situations, and its empirical properties, evaluated by simulation, suggest that numerically the NCAR plus noise model is far superior to the NCAR only model when optimizing the likelihood function.

When the distribution functional form is established to lie only within a neighborhood of a nominal, say Gaussian, distribution, then a robust parametric approach is deemed appropriate. For simple contamination in NSHP and GMRF, such as widely scattered additive and innovative outliers, the weighted LS estimator with data dependent weights, i.e., the 2-D GM-estimator, works well and performs better in the contaminated situation than the conventional LS estimator and performs almost equally as well as the LS when contamination is nonexistent.

However, the 2-D GM-estimator possesses no optimality property and may not perform well when the structure of the contamination is more complicated. Therefore, an optimal robust problem was proposed for a general class of ML type estimators. This optimality problem was solved for the NSHP model and an optimal robust estimator was synthesized. The experiments clearly establish this estimator's superior performance and optimal properties in a variety of contaminated data situations when compared with the performance of the 2-D GM-estimator and the optimal ML estimator based on a strict Gaussian assumption.

Estimators for NCAR and GMRF models were also derived from the basic robust statistical principles. These empirical estimators also perform better than the ML estimator when the observed data has contamination.

8.2 Topics for Future Research

Several topics requiring investigation have become apparent as a consequence of the research conducted in this dissertation. In general, the field of parametric spectrum estimation for contaminated random fields remains open in so far as

the dependent situation is concerned. The optimal robust estimator in Chapter 6 is not the complete answer since it only applies to estimators of the form $\mathrm{E}\{\psi(\mathbf{x}_{\mathbf{s}}^{0};\phi)|x(\mathbf{t}),\mathbf{t}<\mathbf{s}\}=\mathbf{0}$, a class which excludes noncausal random field models. Specifically, we have the following problems which require more analysis and future research.

- For detection and resolution of 2-D spectra with peaks a primary benefit of model based spectrum estimation using noncausal models is a significant reduction in the number of unknown parameters requiring estimation when compared to QP and NSHP models. This advantage is present in both the clean and contaminated data situations. The experimental example in Chapter 5 showed that 40 neighbor NSHP models were necessary to detect and resolve two closely spaced sinusoidal signals in noise. Contrast this with the examples in [12] where only a 5th-order (12 neighbors) NCAR only model and in Chapter 4 where only 2nd-order (4 neighbor) NCAR plus noise model detected and resolved two closely spaced sinusoidal signals in noise.

  All other things being equal one chooses an optimal estimator over others. Although the estimators evaluated in Chapter 7 perform well from an empirical standpoint, nothing much can be said regarding their optimality. Consequently, an optimal robust estimator for the noncausal situation is required.

  Suppose $\mathcal{X}_{\infty}=\{x(\mathbf{s}),\mathbf{s}\in\Omega_{\infty}\}$ is an α-mixing process in the sense given in [72]. Then, we suggest that the asymptotic covariance matrix in terms of an influence function $IF(\mathbf{x}^{0};\phi)\sim\psi(\mathbf{x}^{0};\phi)$ for the ML type NCAR model

estimator is given by

$$V(T, \mathcal{F}^{N_0}) = \mathrm{E}\left\{IF(\mathbf{x}^0;\phi)IF^t(\mathbf{x}^0;\phi)\right\} + \sum_{\mathbf{s}\neq(0,0)} \mathrm{E}\left\{IF(\mathbf{x}^0;\phi)IF^t(\mathbf{x}^0_{\mathbf{s}};\phi)\right\}. \quad (8.1)$$

We suspect that $\mathrm{tr}\left\{V(T, \mathcal{F}^{N_0})\right\}$ is dominated by the trace of the first term in (8.1). Hampel's optimality problem can then be stated as "minimize $\mathrm{tr}\{\mathrm{E}\{IF(\mathbf{x}^0;\phi)IF^t(\mathbf{x}^0;\phi)\}\}$, where $IF(\mathbf{x}^0;\phi) = \psi(\mathbf{x}^0;\phi) - \mathrm{h}(\phi)$, subject to a bound on the bias". The solution to this problem requires further investigation.

- The ML approach to parametric spectrum estimation possesses two major shortcomings:

  1. Computational intensive nonlinear optimization is usually required to solve for the ML parameter estimates, and
  2. Global solutions are usually only possible when the observed data are infinite in extent and the empirical distribution exactly matches the assumed distribution.

  Global optimization algorithms available to date only ameliorate and not eliminate the second shortcoming and usually do so with increased computational requirements. Toroidal lattice assumptions and the fast Fourier transform have reduced computational requirements but not conquered the first shortcoming.

  Attempts have been made by several researchers to adapt the ideas of stochastic relaxation and simulated annealing [75, 76, 77] to the deterministic optimization problem [78]. Interest in this subject is motivated by two

potential benefits: 1) The promise of an algorithm that yields with high probability the global solution and 2) the promise of parallel algorithms solvable on neural network computers. However, the simulated annealing method is generally applicable to optimization of discreet systems with at most a countable number of distinct states. The ML optimization is continuous in nature and application of discreet method will not necessarily guarantee the global solution.

An approach to the continuous problem is by application of stochastic diffusion equation theory [79, 80, 81, 80, 82, 83] which seems on the surface a viable alternative to deterministic nonlinear optimization for determining the global minimum of the negative log-likelihood function.

For example, a possible procedure for determining the global minimum of the negative log-likelihood function, $l_M(\phi) = -\log f(\mathbf{x}|\phi)$, is to compute

$$\phi^{k+1} = \phi^k - a_k \lambda_M(\phi) + a_k \mu^k$$

until convergence, where k is the step count, $\lambda_M(\phi)$ is the gradient of $l_M(\phi)$, $\mu^k = \text{col}\,\{\mu_i^k, i = 1, \ldots, p+1\}$ is a $(p+1)$-dimensional vector of zero mean and identically distributed Gaussian random variables, and the step size a_k is given by

$$a_k = \frac{A_0}{\log(k + A_1)}$$

for $A_0 > 0$ and $A_1 > 0$. The term $a_k \mu^k$ is the "annealing" term to force the sequence $\{\phi^k\}$ to jump around until it eventually settles near a global minimum of $l_M(\phi)$.

Our initial empirical investigation of a diffusion optimization algorithm for minimizing the negative log-likelihood function was less than satisfactory.

The procedure is sensitive to the choice of the step size constants A_0 and A_1, and no theoretical rules are suggested for their selection. This is a topic for future research.

- Having obtained model based spectrum estimates one would also like to know how good they are. This problem remains open, even for the 1-D situation, since statistical analysis of the model based spectrum estimate is difficult and, in general, can only be achieved for asymptotic conditions. Sharma and Chellappa [12] have computed confidence bands for NCAR and GMRF spectra which specify that as $M \rightarrow \infty$ the probability is at least $(1-\alpha)$ that simultaneously for all ω

$$\frac{1}{\mathcal{S}_y^{-1}(\omega,\hat{\phi}) + a(\omega,\hat{\phi})} \leq \mathcal{S}_y(\omega) \leq \frac{1}{\mathcal{S}_y^{-1}(\omega,\hat{\phi}) - a(\omega,\hat{\phi})}, \tag{8.2}$$

where $a^2(\omega,\phi)$ depends on the upper α critical value of a χ^2-distribution having $(p+1)$ degrees of freedom and the asymptotic covariance matrix of the ML parameter estimates. The only requirement is that $\hat{\phi}$ be an asymptotically normal and efficient estimator of ϕ. Such is the case for the NCAR plus noise model estimates of θ, β, and γ. Thus, the confidence bands can be computed similarly for the NCAR plus noise model. For example, use (8.2) with the NCAR plus noise model estimates of θ and β and the associated asymptotic covariance matrix. This gives a confidence band for $\mathcal{S}_y(\omega)$ without the additive $\mathcal{S}_n(\omega) = \gamma^2$.

Another method to compute confidence bands and obtain statistical information regarding $\mathcal{S}(\omega,\hat{\phi})$ is as follows. If $\hat{\mathbf{s}}_y = \text{col}\,\{\mathcal{S}_y(\omega_s,\hat{\phi}), s \in \Omega_M\}$, then as in time series case [25] $\hat{\mathbf{s}}_y$ should be multivariate Gaussian with

$\mathrm{E}\{\hat{\mathbf{s}}_y\} = \mathbf{s}_y$ and covariance matrix

$$\mathrm{E}\left\{[\hat{\mathbf{s}}_y - \mathbf{s}_y][\hat{\mathbf{s}}_y - \mathbf{s}_y]^t\right\} = J\Gamma^{-1}(\phi)J^t,$$

where J is the $M^2 \times (p+1)$ Jacobian matrix of the transformation from θ, β^2 to $\mathbf{s}_y$. Consequently, a confidence band, dependent on ω, can be computed from the covariances. This conjecture requires further analysis and proof. It may also be extendable to $\mathbf{s}_x = \mathbf{s}_y + \gamma^2\mathbf{1}$ with $\mathbf{1} = \mathrm{col}\{1,\ldots,1\}$.

- We suggest that robust techniques can provide improved results in adaptive array signal processing and adaptive array beamforming. Moreover, robust techniques can be applied to Kalman filtering which finds widespread use in radar, sonar, geophysics, and radio astronomy signal processing. All of these practical applications contain their own unique problems which require analysis and solutions, hence, are topics for future research.

- Higher-order statistics have become an important research topic in recent years [84]. For example, cumulant-based parameter estimates of stochastic models [85] have been studied for their promise of good performance in non-Gaussian situations and lower sensitivity to noise. A 3rd-order cumulant,

$$c(\mathbf{s},\mathbf{t}) = \mathrm{E}\{x(\mathbf{s})x(\mathbf{s}+\mathbf{r})x(\mathbf{s}+\mathbf{t})\},$$

can be used to estimate the parameters of a causal model (3.1). However, estimates of $c(\mathbf{s},\mathbf{t})$ directly from the data suffer the same sensitivity to contamination as do the conventional second order sample covariances. We suggest that robust techniques may prove useful in cumulant-based procedures.

References

[1] P. Whittle, "On stationary processes in the plane", *Biometrika*, vol. 41, pp. 434–449, 1954.

[2] M.S. Bartlett, *"The Statistical Analysis of Spatial Pattern"*, Chapman and Hall Ltd., London, 1975.

[3] E. Vanmarcke, *"Random Fields: Analysis and Synthesis"*, The MIT Press, Cambridge, 1983.

[4] R. Chellappa, *"Stochastic Models for Image Analysis and Processing"*, PhD thesis, Purdue University, W. Lafayette, Indiana, 1981.

[5] J.H. Justice, N.L. Owsley, J.L. Yen, and A.C. Kak, *"Array Signal Processing"*, Prentice Hall, Englewood Cliffs, New Jersey, 1985.

[6] L.S. Joyce, "A separable 2-D autoregressive spectral estimation algorithm", *Proceedings ICASSP 79* (Washington DC, April 1979), pp. 677–680, April 1979.

[7] J.H. McClellan, "Multidimensional spectral estimation", *Proc. IEEE*, vol. 70, no. 9, pp. 1029–1039, Sept. 1982.

[8] A.K. Jain and S. Ranganath, "High resolution spectrum estimation in two dimensions", *Proceedings 1st ASSP Workshop on Spectral Estimation* (Hamilton, Ontario, Canada, August 1981), pp. 3.4.1–3.4.5, Aug. 1981.

[9] D. Tjostheim, "Autoregressive modeling and spectral analysis of array data in the plane", *IEEE Trans. on Geosci. Remote Sensing.*, vol. GE-10, no. 1, pp. 15–24, Jan. 1981.

[10] D.E. Dudgeon and R.M. Mersereau, *"Multidimensional Digital Signal Processing"*, Prentice Hall, Englewood Cliffs, New Jersey, 1984.

[11] L.B. Jackson and H.C. Chien, "Frequency and bearing estimation by two-dimensional linear prediction", *Proceedings ICASSP 79* (Washington DC, April 1979), pp. 665–668, April 1979.

[12] G. Sharma and R. Chellappa, "Two-dimensional spectrum estimation using noncausal autoregressive models", *IEEE Trans. on Info. Theory*, vol. IT-32, no. 2, pp. 268–275, Mar. 1986.

[13] J.A. Cadzow and K. Ogino, "Two-dimensional spectral estimation", *IEEE Trans. on Acoust. Speech and Signal Process.*, vol. ASSP-29, no. 3, pp. 396–401, June 1981.

[14] J.S. Weszka, C.R. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification", *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-6, no. 4, pp. 269–285, Apr. 1976.

[15] R. Chellapa and S. Chatterjee, "Classification of textures using Gaussian-Markov random fields", *IEEE Trans. on Acoust. Speech and Signal Process.*, vol. ASSP-33, no. 4, pp. 959–963, Aug. 1985.

[16] R. Chellappa and R.L. Kashyap, "Texture synthesis using 2-D noncausal autoregressive models", *IEEE Trans. on Acoust. Speech and Signal Process.*, vol. ASSP-33, no. 1, pp. 194–203, Feb. 1985.

[17] H.C. Andrews and B.R. Hunt, *"Digital Image Restoration"*, Prentice-Hall, Englewood Cliffs, N.J., 1977.

[18] S.A. Kassam and H.V. Poor, "Robust techniques for signal processing: a survey", *Proc. IEEE*, vol. 73, no. 3, pp. 433–481, Mar. 1985.

[19] R.D. Martin and D.J. Thomson, "Robust-resistant spectrum estimation", *Proc. IEEE*, vol. 70, no. 9, pp. 1097–1114, Sept. 1982.

[20] E.A. Robinson, "A historical perspective of spectrum estimation", *Proc. IEEE*, vol. 70, no. 9, pp. 885–907, Sept. 1982.

[21] Jr. S.L. Marple, *"Modern Spectral Estimation with Applications"*, Prentice Hall, Englewood Cliffs, N.J., 1987.

[22] S.M. Kay and Jr. S.L. Marple, "Spectrum analysis–a modern perspective", *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.

[23] D.G. Childers, ed., *"Modern Spectral Analysis"*, IEEE Press, New York, 1978.

[24] S.B. Kesler, ed., *"Modern Spectral Analysis, II"*, IEEE Press, New York, 1986.

[25] S.M. Kay, *"Modern Spectral Estimation: Theory and Application"*, Prentice Hall, Englewood Cliffs, N.J., 1988.

[26] T.S. Huang, "Two-dimensional windows", *IEEE Trans. on Audio and Electroacoustics*, vol. AU-20, no. 1, pp. 80–90, Mar. 1972.

[27] J.P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra", *Geophysics*, vol. 37, pp. 375–376, April 1972.

[28] J.S. Lim and N.A. Malik, "A new algorithm for two-dimensional maximum entropy power spectrum estimation", *IEEE Trans. on Acoust. Speech and Signal Process.*, vol. ASSP-29, no. 3, pp. 401–413, June 1981.

[29] J.W. Woods, "Two-dimensional Markov spectral estimation", *IEEE Trans. on Info. Theory*, vol. IT-22, no. 5, pp. 552–559, Sept. 1976.

[30] R. Chellappa, Y.H. Hu, and S.Y. Kung, "On two-dimensional Markov spectral estimation", *IEEE Trans. on Acoust. Speech and Signal Process.*, vol. ASSP-31, no. 4, pp. 836–841, Aug. 1983.

[31] G. Sharma and R. Chellappa, "A model-based approach for estimation of two-dimensional maximum entropy power spectra", *IEEE Trans. on Info. Theory*, vol. IT-31, no. 1, pp. 90–99, Jan. 1985.

[32] J.E. Besag, "Spatial interaction and the statistical analysis of lattice systems", *J. Royal Statist. Soc., Series B*, vol. 36, no. 2, pp. 192–236, 1974.

[33] R.L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images", *IEEE Trans. on Info. Theory*, vol. IT-29, no. 1, pp. 60–72, Jan. 1983.

[34] J.W. Woods, "Two-dimensional discrete Markovian random fields", *IEEE Trans. on Info. Theory*, vol. IT-18, no. 2, pp. 232–240, Mar. 1972.

[35] H. Künsch, "Thermodynamics and Statistical Analysis of Gaussian Random Fields", *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 58, pp. 407–421, 1981.

[36] R. Chellappa, "Two-Dimensional Discrete Gaussian Markov Random Field Models for Image Processing," in *"Progress in Pattern Recognition 2"*, L.N. Kanal and A. Rozenfeld, Eds., Elsevier Science Publishers B.V., North Holland, 1985.

[37] X. Guyon, "Parameter estimation for a stationary process on a d-dimensional lattice", *Biometrika*, vol. 69, pp. 95–105, 1982.

[38] F.R. Hampel, E.M. Ronchetti, E.M. Rousseeuw, and W.A. Stahel, *"Robust Statistics: The Approach Based on Influence Functions"*, John Wiley & Sons, New York, 1986.

[39] P.J. Huber, *"Robust Statistics"*, John Wiley & Sons, New York, 1981.

[40] B. Kleiner, R.D. Martin, and D.J. Thomson, "Robust estimation of power spectra", *J. Royal Statist. Soc., Series B*, vol. 41, pp. 313–351, 1979.

[41] R.D. Martin, "Robust Estimation for Time Series Autoregressions," in *"Robustness in Statistics"*, R.L. Launer and G.N. Wilkinson, Eds., Academic Press, New York, 1979.

[42] R.D. Martin and J.E. Zeh, "Robust generalized M-estimates for autoregressive parameters, including small sample behavior", Technical Report 214, Univ. of Washington, Dept. of Elec. Engr., 1978.

[43] J.W. Tukey, "A Survey On Sampling from Contaminated Distributions," in *"Contributions to Probability and Statistics"*, I. Olkin, Eds., Stanford University Press, Stanford, Ca., 1960.

[44] F.R.Hampel, "A general qualitative definition of robustness", *Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, 1971.

[45] R. von Mises, "On the asymptotic distribution of differentiable statistical functions", *Annals of Mathematical Statistics*, vol. 18, pp. 309–348, 1947.

[46] G. Sharma, *"Modern Two-Dimensional Spectrum Estimation Using Noncausal Spatial Models"*, PhD thesis, University of Southern California, Los Angeles, California, 1984.

[47] P.R. Gutowski, E.A Robinson, and S. Treitel, "Spectral estimation: Fact or fiction", *IEEE Trans. on Geoscience Electronics*, vol. GE-16, no. 2, pp. 80–84, Apr. 1978.

[48] H. Akaike, "A new look at the statistical model identification", *IEEE Trans. on Automatic Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[49] W.E. Larimore, "Statistical inference on stationary random fields", *Proc. IEEE*, vol. 65, no. 6, pp. 961–970, June 1977.

[50] R.L. Kashyap, R. Chellappa, and N.Ahuja, "Decision rules for choice of neighbors in random field models of images", *Computer Graphics and Image Processing*, vol. 15, pp. 301–318, Apr. 1981.

[51] R.L. Kashyap, "Inconsistency of the AIC rule for estimating the order of autoregressive models", *IEEE Trans. on Automatic Control*, vol. AC-25, no. 5, pp. 996–998, Oct. 1980.

[52] E.J. Hannan and B.G. Quinn, "The determination of the order of an autoregression", *J. Royal Statist. Soc., Series B*, vol. 41, pp. 190–195, 1979.

[53] H.W. Sorenson, *"Parameter Estimation: Principles and Problems"*, Marcel Dekker, Inc., New York, 1980.

[54] J.M. Mendel, *"Lessons in Digital Estimation Theory"*, Prentice Hall, Englewood Cliffs, N.J., 1987.

[55] E.L. Lehmann, *"Theory of Point Estimation"*, John Wiley & Sons, New York, 1983.

[56] K. Dzhaparidze, *"Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series"*, Springer-Verlag, New York, 1986.

[57] A. Papoulis, *"Probability, Random Variables, and Stochastic Processes"*, McGraw-Hill, Inc., New York, 1984.

[58] M. Rosenblatt, *"Stationary Sequences of Random Fields"*, Birkhäuser Boston, Inc., Boston, 1985.

[59] L. LeCam, "On the assumptions used to prove asymptotic normality of maximum likelihood estimates", *Annals of Mathematical Statistics*, vol. 41, no. 3, pp. 802–828, 1970.

[60] A. Wald, "Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process", *Annals of Mathematical Statistics*, vol. 19, pp. 40–46, 1948.

[61] A.M. Walker, "Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series", *J. Australian Mathematical Soc.*, vol. 4, pp. pp 363–384, 1964.

[62] M.J. Crowder, "Maximum likelihood estimation for dependent observations", *J. Royal Statist. Soc., Series B*, vol. 38, no. 1, pp. pp 45–53, 1976.

[63] I.A. Ibragimov and R.Z. Has'minskii, *"Statistical Estimation–Asymptotic Theory"*, Springer-Verlag, New York, 1981.

[64] C. Kraft and L. LeCam, "A remark on the roots of the maximum likelihood equation", *Annals of Mathematical Statistics*, vol. 27, pp. 1174–1177, 1956.

[65] H. Kaufman, J.W. Woods, S. Dravida, and M. Tekalp, "Estimation and identification of two-dimensional images", *IEEE Trans. on Automatic Control*, vol. AC-28, no. 7, pp. 745–756, July 1983.

[66] H. Jinchi, *"Optimal Image Restoration Algorithms Using Two-Dimensional Gaussian Markov Random Field Models"*, PhD thesis, University of Southern California, Los Angeles, California, 1986.

[67] International Mathematical & Statistical Libraries Inc., *"IMSL Library"*, P.O. Box 4605, Houston, Texas.

[68] R.D. Martin and V.J. Yohai, "Influence functionals for time series", *Annals of Statistics*, vol. 14, pp. 781–818, 1986.

[69] H. Künsch, "Infinitesimal robustness for autoregressive processes", *Annals of Statistics*, vol. 12, pp. 843–863, 1984.

[70] R.J. Serfling, *"Approximation Theorems of Mathematical Statistics"*, John Wiley & Sons, New York, 1980.

[71] D.G. Luenberger, *"Linear and Nonlinear Programming"*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1984.

[72] P. Billingsley, *"Probability and Measure"*, John Wiley & Sons, New York, 1986.

[73] R.R. Hansen and R. Chellappa, "Robust Algorithms for Two-Dimensional Spectrum Estimation", Technical Report USCSIPI #108, Univ. of Southern California, Dept. of Elec. Engr., Systems, 1987.

[74] L.T. Fernholz, *"von Mises Calculus for Statistical Functionals"*, Springer-Verlag, New York, 1983.

[75] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", *IEEE Trans. on Pattern Anal. and Machine Intel.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[76] S. Kirkpatrick, C.D. Gelatt, and M. Vecchi, "Optimization by simulated annealing", *Science*, vol. 220, no. 4598, pp. 671–680, May 1983.

[77] P.J.M. van Laarhoven and E.H.L. Aarts, *"Simulated Annealing: Theory and Applications"*, D. Reidel Publishing Co., Dordrecht, Holland, 1987.

[78] I.O. Bohachevsky, M.E. Johnson, and M.L. Stein, "Generalized simulated annealing for function optimization", *Technometrics*, vol. 28, no. 3, pp. 209–217, Aug. 1986.

[79] F. Aluffi-Pentini, V. Parisi, and F. Zirilli, "Global optimization and stochastic differential equations", *Journal of Optimization Theory and Applications*, vol. 47, no. 1, pp. 1–16, Sept. 1985.

[80] T. Chiang, C. Hwang, and S. Sheu, "Diffusion for global optimization in R^n", *SIAM Journal of Control and Optimization*, vol. 25, no. 3, pp. 737–753, May 1987.

[81] S. Geman and C. Hwang, "Diffusions for global optimization", *SIAM Journal of Control and Optimization*, vol. 24, no. 5, pp. 1031–1043, Sept. 1986.

[82] A.S. Sherman and C.S. Peskin, "A Monte Carlo method for scalar reaction diffusion equations", *SIAM Journal of Sci. Stat. Computations*, vol. 7, no. 4, pp. 1360–1372, Oct. 1986.

[83] H.J. Kushner, "Asymptotic global behavior for stochastic approximation and diffussions with slowly decreasing noise effects: global minimization via Monte Carlo", *SIAM Journal of Applied Mathematics*, vol. 47, no. 1, pp. 169–185, Feb. 1987.

[84] G.B. Gianakis, *"Signal Processing via Higher-Order Statistics"*, PhD thesis, University of Southern California, Los Angeles, California, 1986.

[85] A. Swami and J.M. Mendel, "ARMA parameter estimation using only output cumulants", *Proceedings 4th ASSP Workshop on Spectral Estimation* (Minneapolis, Minnesota, August 1988), pp. 193–198, Aug. 1988.