USCSIPI #108

Robust Algorithms for Two-Dimensional Spectrum Estimation†

Richard R. Hansen, Jr., Rama Chellappa, and

Govind Sharma

June 29, 1987

Signal and Image Processing Institute

University of Southern California

Los Angeles, California, 90089-0272

Robust Algorithms for Two-Dimensional Spectrum Estimation†

Richard R. Hansen, Jr., Rama Chellappa, and Govind Sharma

June 29, 1987

Signal and Image Processing Institute

University of Southern California

Los Angeles, California, 90089-0272

† Partially supported by the Office of Naval Research under the contract no. N00014-86-K-0302.

Abstract

In this paper we investigate robust estimation of two-dimensional (2-D) power spectra of signals which are adequately represented by Gaussian random field models but for which we have imperfect observations. Two situations of particular interest occur when the contaminating noise is additive and when the contaminating noise appears in the innovations. In these cases the observed data is not Gaussian and conventional procedures are no longer efficient. To estimate the parameters of the signal model from the contaminated data we describe two new procedures which were originally proposed for estimation of scale and location from independent data and adapted to one-dimensional autoregression parameter estimation by previous researchers. The first algorithm is a robustification of least squares and equivalent to an iterated weighted least squares problem where the weights are data dependent. Known as the generalized maximum-likelihood (GM) estimator its analysis is accomplished by the use of a so-called "influence function" or directional derivative of the estimator in the direction of the contamination. We compute expressions for relative efficiency of the estimator using the influence function and specify criteria for selection of the estimator's robustifying functions. The second algorithm is an iterative procedure known as a filter-cleaner. This procedure is shown to be approximately equivalent to an optimal minimization problem.

Experiments using the robust procedures with synthetic data are reported and the results compared with a conventional method of model-based spectrum estimation, i.e., consistent least squares parameter estimation. Finally, we conclude with a summary of the utility and improved performance of the robust procedures over the conventional method and a discussion of the shortcomings of these heuristically derived robust methods.

Contents

List of Figures

List of Tables

1 Introduction

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 and using the parameter estimates in the theoretical spectrum expression derived from the model. The model selection is often one of convenience, yielding optimal results when all assumptions (e.g., normality) are satisfied and having tractable parameter estimation methods. Model based spectrum estimation using a variety of models has been evaluated and good results obtained for data which matches the assumptions [1,2,3].

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.

This paper investigates model based two-dimensional spectrum estimation techniques for imperfectly observed signals. The procedures developed here use the concept of robust parameter estimation which has received extensive consideration in the statistical literature. A major portion of this literature (e.g., see [4] and [5]) treats location and linear regression models with the assumption of independent and identically distributed (IID) data. The literature on robust estimation in the dependent or autoregressive time series case is sparse. Kleiner et al. [6,7,8] have developed several robust estimation techniques for the one-dimensional (1-D) case when the data are representable by an autoregressive model. Very little, if any, work has been published for two-dimensional robust spectrum estimation; and the 1-D techniques are not directly extendable without considering the unique properties of two-dimensional models.

Two general approaches seem to be prevalent in robust estimation. The first, Huber's minimax approach [5], attacks the problem by considering so-called "M-estimators", which are a generalization of maximum likelihood 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 [4] 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 neighbor of a nominal model.

In this paper we employ the second approach and derive a directional derivative which is similar to Hampel's influence function. Asymptotic properties of robust parameter estimators for two-dimensional model-based spectrum estimation are computed using this function, and criteria for designing the form of the estimators are proposed.

2 Contamination in Observed Data

The assumption of normality for the true signal of interest is often proposed from empirical evidence or justified in theory by application of a suitable central limit theorem. But in practical empirical situations measurement errors or isolated phenomena may 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 signal measurement error. Then the observed data is distributed as though it were Gaussian near the mean but having no tails at all.

This paper concentrates on the signal plus noise model, where the observed data $\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$ on the $M \times M$ rectangular lattice $\Omega_M = \{\mathbf{s} = i,j \mid 0 \leq i,j \leq M-1\}$ are the sum of the true signal $y(\mathbf{s})$ and a noise component $n(\mathbf{s})$. Thus, the model is

$$x(\mathbf{s}) = y(\mathbf{s}) + n(\mathbf{s}), \quad \mathbf{s} \in \Omega_M. \tag{1}$$

The primary objective is to estimate the spectrum of $y(\mathbf{s})$ given the observed data $\{x(\mathbf{s}), \mathbf{s} \in \Omega_M\}$ and assuming that the signal $y(\mathbf{s})$ is adequately represented by a spatial interaction model, either a spatial autoregressive or Markov random field, nominally driven by a Gaussian random noise input. We will not elaborate on the methods for specification and conventional estimation of parameters in spatial interaction models since they have been widely studied in the engineering and statistical literature. For example, see [2,3,9].

We are particularly interested in two commonly occuring situations: 1) the innovative outlier (IO) model where $n(\mathbf{s}) \equiv 0$ for all $\mathbf{s}$ but $y(\mathbf{s})$ is a non-Gaussian signal and 2) the additive outlier model where $n(\mathbf{s}) \neq 0$ for at least some $\mathbf{s}$ but $y(\mathbf{s})$ is Gaussian. Other situations exist, for example, $n(\mathbf{s})$ may replace $y(\mathbf{s})$ or $n(\mathbf{s})$ may be correlated with $y(\mathbf{s})$. The latter requires a more general model than (1) and the former, known as substitutive outliers (SO), will only be treated empirically in the experiments reported here.

Martin and Thomson [10] have suggested that isolated measurement errors or outliers be modeled with the mixture distribution

$$\mathcal{F}_e = (1-\gamma)\delta_0 + \gamma\mathcal{N}(0,\sigma_e^2)\ . \tag{2}$$

Here δ_0 is the degenerate distribution having all its mass at zero and $\mathcal{N}(0,\sigma_e^2)$ is the standard normal distribution with mean zero and variance σ_e^2. The IO model will be said to hold whenever, $x(\mathbf{s}) = y(\mathbf{s})$ for all $\mathbf{s}$ 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 (2). The AO model will occur if $n(\mathbf{s})$ in (1) has the distribution given by (2). Then the signal is observed correctly most of the time, i.e., $x(\mathbf{s}) = y(\mathbf{s})$, but $100\,\gamma$ 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 γ, $\mathcal{F}_x$ 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 γ the distribution $\mathcal{F}_x$ is also unknown regardless of whether it is associated with the innovations or an additive effect.

Even when γ is small, say $\gamma \leq .10$, 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 [2] and [3] are vulnerable to even a few outliers, as are their counterparts in 1-D [10]. 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.

Strictly speaking a robust procedure should have the following two properties: i) when the data are "good", i.e., Gaussian, the procedure should be almost as good as the conventional (often optimal) procedure presuming normality, and ii) when outliers are present or the distribution deviates slighty from the assumptions, the procedure should still work well, and in particular work much better than the conventional procedure.

In the next section we describe two procedures which have these properties, and in Section 4 we present the results of limited empirical studies. We conclude in Section 5 with an evaluation of the usefulness of these algorithms.

3 Algorithms for 2-D Robust Spectrum Estimation

3.1 Two-Dimensional Parametric Models

There are two nonequivalent classes of models for two dimensional random fields, the simultaneous models and the conditional Markov models [11]. Here for the spectrum estimation problem we are particularly interested in the simultaneous autoregressive models from the first class and the conditional Gaussian-Markov models from the second class. We briefly describe these models since the details can be found elsewhere [9,12,13,14].

Class 1: Simultaneous autoregressive models are generalizations to multi-dimensions of the familiar one-dimensional time series autoregressions and are characterized by the difference equation

$$y(\mathbf{s}) = \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} y(\mathbf{s}+\mathbf{r}) + \beta w(\mathbf{s}), \quad \mathbf{s} \in \Omega, \tag{3}$$

where $\{w(\mathbf{s})\}$ is an IID Gaussian random noise array with $\mathrm{E}\{w(\mathbf{s})\} = 0$ and $\mathrm{E}\{w^2(\mathbf{s})\} = 1$, and N is the neighbor set which defines the region of support for the model. By restricting the members of the set N to be a subset of the non-symmetric half-plane, unilateral non-symmetric half-plane (NSHP) models are defined, while no restrictions on the members of N result in noncausal autoregressive (NCAR) models. For both models the spectra are given by

$$\mathcal{S}_y(\lambda) = \frac{\beta^2}{|1 - \sum_{\mathbf{r}\in N} \theta_{\mathbf{r}} \exp(-j\,\lambda^t \mathbf{r})|^2}. \tag{4}$$

Class 2: Gaussian-Markov random fields (GMRF) are characterized by the difference equation

$$y(\mathbf{s}) = \sum_{\mathbf{r}\in N} \theta_{\mathbf{r}} y(\mathbf{s}+\mathbf{r}) + e(\mathbf{s}), \quad \mathbf{s} \in \Omega, \tag{5}$$

where the correlated noise array $\{e(\mathbf{s})\}$ has $\mathrm{E}\{e(\mathbf{s})\} = 0$ and correlation structure

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

The Gaussian-Markov random field $\{y(\mathbf{s})\}$ possesses a Markov property with respect to neighbor set N [15,14], namely, the conditional density of $y(\mathbf{s})$ is

$$p(y(\mathbf{s}) \mid \text{all } y(\mathbf{r}),\ \mathbf{r} \neq \mathbf{s}) = p(y(\mathbf{s}) \mid y(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N)\ .$$

In view of the Gaussian nature of $y(\mathbf{s})$, it is also true that

$$\mathrm{E}\{y(\mathbf{s}) \mid \text{all } y(\mathbf{r}),\ \mathbf{r} \neq \mathbf{s}\} = \mathrm{E}\{y(\mathbf{s}) \mid y(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$$

and

$$\mathrm{Var}\{y(\mathbf{s}) \mid \text{all } y(\mathbf{r}),\ \mathbf{r} \neq \mathbf{s}\} = \mathrm{Var}\{y(\mathbf{s}) \mid y(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}\ .$$

The GMRF model is usually defined with a noncausal neighbor set since an equivalent NSHP model can be found when the neighbor set is unilateral. In general, this equivalency is not true for NCAR and noncausal GMRF models.

The GMRF spectrum is computed by

$$\mathcal{S}_y(\lambda) = \frac{\nu}{1 - \sum_{\mathbf{r}\in N} \theta_{\mathbf{r}} \exp(-j\,\lambda^t \mathbf{r})}. \tag{6}$$

3.2 2-D Robust Parameter Estimation

Robust spectrum estimation for time series has been suggested in [6,7,10]. Since least squares estimates are consistent for 2-D non-symmetric half-plane (NSHP) and noncausal Gaussian-Markov random field (GMRF) models, the arguments used in the 1-D case can be followed and a robustified least squares problem can be defined. Suppose the signal is modeled by a NSHP Gaussian random field, then robust parameter estimates $\underline{\hat{\theta}}_{GM}$ of $\underline{\theta} = \text{col}\{\theta_{\mathbf{r}},\ \mathbf{r} \in N\}$ and S_{GM} of β are computed by solving

$$\sum_{\mathbf{s}\in\Omega} \mathbf{x}_{\mathbf{s}} W(\mathbf{x}_{\mathbf{s}}) \psi\left(\frac{x(\mathbf{s}) - \underline{\hat{\theta}}_{GM}^t \mathbf{x}_{\mathbf{s}}}{c_v S_{GM}}\right) = \mathbf{0} \tag{7}$$

and

$$\sum_{\mathbf{s}\in\Omega} W(\mathbf{x}_{\mathbf{s}}) \left[c_v^2 \psi^2\left(\frac{x(\mathbf{s}) - \underline{\hat{\theta}}_{GM}^t \mathbf{x}_{\mathbf{s}}}{c_v S_{GM}}\right) - B\right] = 0\ , \tag{8}$$

where the past history vector is defined by $\mathbf{x}_{\mathbf{s}} = \text{col}\{x(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$. The estimators (7) and (8) are known as the two-dimensional generalized maximum-likelihood (GM) estimators. The equations to be solved are identical for a GMRF model but then S_{GM} is and estimate of $\sqrt{\nu}$.

In (7) and (8) c_v and B are tuning constants selected to adjust robustness and yield consistent estimates when the observed data $\{x(\mathbf{s})\}$ are normally distributed. The function $\psi(\cdot)$ is to limit the influence of those summands of (7) and (8) for which $v(\mathbf{s}) = x(\mathbf{s}) - \underline{\theta}^t \mathbf{x_s}$ 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 $\mathbf{x_s}$. The choice of $\psi(\cdot)$ and $W(\cdot)$ functions with good robustness properties will be discussed in subsequent sections.

Equations (7) and (8) can be solved with an iterated weighted least squares procedure as suggested by Huber [5] and the robust estimates $\underline{\theta} = \hat{\underline{\theta}}_{GM}$ and $\beta = S_{GM}$ then used in (4) to compute the estimated spectrum.

The main tool in the analysis and synthesis of robust estimators for independent data is the influence function [4] 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, [16] and [17]). 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 (7) and (8).

3.2.1 An Asymptotic GM-Estimator

First, we introduce some notation. Define the m-dimensional vector $\mathbf{x_s} = \text{col}\{x(\mathbf{s}+\mathbf{r}),\ \mathbf{r} \in N\}$. Let $\mathbf{x}_\mathbf{s}^0 = \text{col}\{x(\mathbf{s}), \mathbf{x_s}\}$, $N_0 = N \cup (0,0)$, and denote the $(m+1)$-dimensional distribution function of $\mathbf{x}_\mathbf{s}^0$ by $\mathcal{F}_x^{N_0}$. We assume that the data are stationary so that $\mathcal{F}_x^{N_0}$ is the same for all $\mathbf{s}$. Now, let x_0, $\mathbf{x}$, and $\mathbf{x}^0$ be dummy variables for $x(\mathbf{s})$, $\mathbf{x_s}$, and $\mathbf{x}_\mathbf{s}^0$, respectively. In the following assume $\text{E}\{x(\mathbf{s})\} = 0$, 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 [5]) of the mean of $x(\mathbf{s})$. Then the asymptotic GM-estimates $\underline{\theta}_{AGM}$ of $\underline{\theta}$ and S_{AGM} of β are defined by the functional $T(\mathcal{F}^{N_0}) = \text{col}\{\underline{\theta}(\mathcal{F}^{N_0}), S(\mathcal{F}^{N_0})\}$ whose value is a root of the $m+1$ equations

$$\text{E}_{\mathcal{F}^{N_0}}\{\mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0; \underline{\theta}(\mathcal{F}^{N_0}), S(\mathcal{F}^{N_0}))\} = \mathbf{0} \tag{9}$$

and

$$\text{E}_{\mathcal{F}^{N_0}}\{W(\mathbf{x})[\psi^2(\mathbf{x}^0; \underline{\theta}(\mathcal{F}^{N_0}), S(\mathcal{F}^{N_0})) - B]\} = 0 \tag{10}$$

evaluated when $\mathcal{F}^{N_0} = \mathcal{F}_x^{N_0}$. 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

$$\mathcal{F}_x^{N_0,M} = \frac{1}{M_2} \sum_{\mathbf{s} \in \Omega_M} \delta_{\mathbf{x}_\mathbf{s}^0} \,, \tag{11}$$

then the solution of (9) and (10) at $\mathcal{F}^{N_0} = \mathcal{F}_x^{N_0,M}$ yields the estimates $\hat{\underline{\theta}}_{GM}$ and S_{GM} defined by (7) and (8).

By a suitable definition of the boundary conditions, $\mathcal{F}_x^{N_0,M}$ is a reasonable estimate of $\mathcal{F}_x^{N_0}$ in a variety of ways [18]. Furthermore, we expect that $T(\mathcal{F}_x^{N_0,M})$ relates to $T(\mathcal{F}_x^{N_0})$ in a similar fashion if $T(\cdot)$ is sufficiently well behaved. Hence, analysis of (9) and (10) should lead to reasonable prediction of the behavior of (7) and (8) for sufficiently large M.

Consider the following minimization problem:

$$\int W(\mathbf{x})\rho(\mathbf{x}^0; \underline{\theta}, S) d\mathcal{F}^{N_0} = \min_{\underline{\theta}}! \tag{12}$$

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

$$\frac{\partial}{\partial \underline{\theta}} \rho(\mathbf{x}^0; \underline{\theta}, S) = \mathbf{x}\psi(\mathbf{x}^0; \underline{\theta}, S). \tag{13}$$

The solution of (12) is also a root of (9). Moreover, if $\rho(v)$ is a convex function in v, implying that $\psi(v)$ must be strictly monotone, then the solution to (9) is unique. This follows from the fact that if $\rho(v)$ is continuously differentiable and convex, then (9) is both necessary and sufficient for the solution to be a global minimizing point [19]. Existence of the solution is guaranteed if $\rho(v)$ is symmetric.

3.2.2 An Influence Function for the GM-Estimator

Von Mises [20] 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

$$T(\mathcal{G}) = T(\mathcal{F}) + \int \phi(x) d(\mathcal{G} - \mathcal{F}) + \text{remainder}, \tag{14}$$

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

$$T'(\mathcal{F}; \mathcal{G} - \mathcal{F}) \stackrel{\text{def}}{=} \lim_{t \to 0} \frac{T(\mathcal{F} + t(\mathcal{G} - \mathcal{F})) - T(\mathcal{F})}{t} = \int \phi(x) d\mathcal{G}. \tag{15}$$

$T'(\mathcal{F}; \mathcal{G} - \mathcal{F})$ 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 $T((1-t)\mathcal{F} + t\mathcal{G})$. 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, $\int \phi(x) d\mathcal{F} = 0$.

Letting $\mathcal{G} = \delta_x$ in (15) we see that $T'(\mathcal{F}; \delta_x - \mathcal{F}) = \phi(x)$, which Hampel [4] calls the influence function (IF) of T at $\mathcal{F}$, usually written as $IF(x; T(\mathcal{F}))$. Thus it is true that

$$T(\mathcal{G}) - T(\mathcal{F}) = \int IF(x; T(\mathcal{F})) d\mathcal{G} + \text{remainder}. \tag{16}$$

When (16) is evaluated with $\mathcal{G}$ equal to the observed sample distribution, $\mathcal{F}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$, in most cases [4] the remainder becomes negligible for $n \to \infty$

so that

$$T(\mathcal{F}_n) - T(\mathcal{F}) \approx \frac{1}{n}\sum_{i=1}^{n} IF(x_i; T(\mathcal{F})), \tag{17}$$

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 (17) 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

$$V(T, \mathcal{F}) = \int IF^2(x; T(\mathcal{F}))d\mathcal{F} \ . \tag{18}$$

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 (17) and (18).

The approach that we take in this paper 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 $(m+1)$-dimensional distribution $\mathcal{F}_t^{N_0} = (1-t)\mathcal{F}^{N_0} + t\delta_{\mathbf{x}^0}$ for $0 \leq t \leq 1$ does not correspond directly with any naturally occuring contamination in the random field [16]. This problem occurs because $\delta_{\mathbf{x}^0}$ is not a stationary distribution in the set of $(m+1)$-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 (17) and that an influence function defined with $\mathcal{F}_t^{N_0}$ is proportional to the kernels of (9) and (10) 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 $\underline{\theta}_t = \underline{\theta}(\mathcal{F}_t^{N_0})$ and $S_t = S(\mathcal{F}_t^{N_0})$ and note that $\underline{\theta}_0$ and S_0 are the asymptotic GM-estimates when the data are distributed with marginal $\mathcal{F}^{N_0}$, i.e., when $t=0$. Substitite $\mathcal{F}_t^{N_0}$ in (9) and (10) and differentiate with respect to t at $t=0$, obtaining

$$\frac{\partial}{\partial t}\int \mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0;\underline{\theta}_t, S_t)d\mathcal{F}_t^{N_0}\bigg|_{t=0} = \mathbf{0} \tag{19}$$

and

$$\frac{\partial}{\partial t}\int W(\mathbf{x})[\psi^2(\mathbf{x}^0;\underline{\theta}_t, S_t) - B]d\mathcal{F}_t^{N_0}\bigg|_{t=0} = 0 \ . \tag{20}$$

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.

By defining an $(m+1)$-dimensional GM-estimator influence function for $\underline{\theta}$

and S under $\mathcal{F}^{N_0}$ as

$$IF_{GM}(\mathbf{x}^0; \underline{\theta}_0, S_0) = \begin{bmatrix} \frac{\partial \underline{\theta}_t}{\partial t} \\ \\ \frac{\partial S_t}{\partial t} \end{bmatrix}_{t=0},$$

and carrying out the indicated differentiation in (19) and (20), the result in matrix-vector form is

$$IF_{GM}(\mathbf{x}^0; \underline{\theta}_0, S_0) = \begin{bmatrix} M_{\underline{\theta}} & \mathbf{m}_{\underline{\theta},S} \\ \\ \mathbf{m}^t_{S,\underline{\theta}} & m_S \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0; \underline{\theta}_0, S_0) \\ \\ W(\mathbf{x})[\psi^2(\mathbf{x}^0; \underline{\theta}_0, S_0) - B] \end{bmatrix}, \qquad (21)$$

assuming the inverse exists and where

$$\begin{aligned}
M_{\underline{\theta}} &= \frac{1}{S_0}\int W(\mathbf{x})\psi'(\mathbf{x}^0; \underline{\theta}_0, S_0)\mathbf{x}\mathbf{x}^t d\mathcal{F}^{N_0} \\
\mathbf{m}_{\underline{\theta},S} &= \frac{1}{S_0}\int \mathbf{x}W(\mathbf{x})\psi'(\mathbf{x}^0; \underline{\theta}_0, S_0)\left(\frac{x_0 - \underline{\theta}_0^t\mathbf{x}}{S_0}\right) d\mathcal{F}^{N_0} \\
\mathbf{m}_{S,\underline{\theta}} &= \frac{2}{S_0}\int \mathbf{x}W(\mathbf{x})\psi'(\mathbf{x}^0; \underline{\theta}_0, S_0)\psi(\mathbf{x}^0; \underline{\theta}_0, S_0) d\mathcal{F}^{N_0} \\
m_S &= \frac{2}{S_0}\int W(\mathbf{x})\psi'(\mathbf{x}^0; \underline{\theta}_0, S_0)\psi(\mathbf{x}^0; \underline{\theta}_0, S_0)\left(\frac{x_0 - \underline{\theta}_0^t\mathbf{x}}{S_0}\right) d\mathcal{F}^{N_0}
\end{aligned}$$

and $\psi'(v) = \frac{d}{dv}\psi(v)$.

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 $\mathbf{m}_{\underline{\theta},S}$ and $\mathbf{m}_{S,\underline{\theta}}$ are both zero. In this case the influence function separates, i.e.,

$$\begin{bmatrix} IF_{GM}(\mathbf{x}^0; \underline{\theta}_0) \\ \\ IF_{GM}(\mathbf{x}^0; S_0) \end{bmatrix} = \begin{bmatrix} M_{\underline{\theta}}^{-1}\mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0; \underline{\theta}_0, S_0) \\ \\ m_S^{-1}W(\mathbf{x})[\psi^2(\mathbf{x}^0; \underline{\theta}_0, S_0) - B] \end{bmatrix} \qquad (22)$$

where $IF_{GM}(\mathbf{x}^0; \underline{\theta}_0)$ and $IF_{GM}(\mathbf{x}^0; S_0)$ are the influence functions for the separate estimates $\underline{\theta}_{AGM}$ with S known and S_{AGM} with $\underline{\theta}$ known, respectively.

When the observed data obey the AO model and $n(\mathbf{s})$ in (1) is not identically zero, then $\mathcal{F}^{N_0} = \mathcal{F}_x^{N_0}$ and $\underline{\theta}_0 \neq \underline{\theta}$ due to bias in the estimate; although, if the amount of contamination is small we should expect that $|\underline{\theta}_0 - \underline{\theta}|$ and $|S_0 - \beta|$ will be small for good choices of $\psi(\cdot)$ and $W(\cdot)$. Then (22) should hold approximately for the AO model as well.

The importance of this separation is that the robustness properties of the estimate $\hat{\underline{\theta}}_{GM}$ do not depend on S_{GM}, and vice versa. This suggests an alternative to simultaneous solution of (7) and (8): compute any robust estimate of scale and then use this in solving (7). Additionally, we can evaluate the asymptotic properties of the parameter estimates $\hat{\underline{\theta}}_{GM}$ by considering a fixed scale S.

By an expansion similar to (14) we obtain

$$T(\mathcal{F}^{N_0,M}) - T(\mathcal{F}^{N_0}) = \frac{1}{M^2} \sum_{\mathbf{s} \in \Omega_M} IF_{GM}(\mathbf{x}_{\mathbf{s}}^0; T(\mathcal{F}^{N_0})) + \text{remainder}. \tag{23}$$

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

$$M[T(\mathcal{F}^{N_0,M}) - T(\mathcal{F}^{N_0})] \longrightarrow \mathcal{N}(0, V(T, \mathcal{F}^{N_0})), \tag{24}$$

where

$$V(T, \mathcal{F}^{N_0}) = \int IF_{GM}(\mathbf{x}^0; T(\mathcal{F}^{N_0})) IF_{GM}^t(\mathbf{x}^0; T(\mathcal{F}^{N_0})) d\mathcal{F}^{N_0}. \tag{25}$$

Proof: Clearly, $\int IF_{GM}(\mathbf{x}^0; \underline{\theta}_0, S_0) d\mathcal{F}^{N_0} = \mathbf{0}$. By a suitable selection of $\psi(\cdot)$ and $W(\cdot)$, $IF_{GM}(\mathbf{x}_\mathbf{s}^0; \underline{\theta}_0, S_0)$ for all $\mathbf{s}$ will be uniformly bounded and a stationary process. When $\mathbf{s} - \mathbf{t} \notin N$, then $IF_{GM}(\mathbf{x}_\mathbf{s}^0; \underline{\theta}_0, S_0)$ and $IF_{GM}(\mathbf{x}_\mathbf{t}^0; \underline{\theta}_0, S_0)$ are independent and $IF_{GM}(\mathbf{x}_\mathbf{s}^0; \underline{\theta}_0, S_0)$ is α-mixing. Under these conditions a central limit theorem for dependent random vectors [21] applies and (24) follows. Since NSHP and GMRF models have a Markov property it is easy to show that

$$\int IF_{GM}(\mathbf{x}^0; \underline{\theta}_0, S_0) d\mathcal{F}(x_0|\mathbf{x}) = \mathbf{0}.$$

Then $IF_{GM}(\mathbf{x}_\mathbf{s}^0; \underline{\theta}_0, S_0)$ is a martingale difference so that $IF_{GM}(\mathbf{x}_\mathbf{s}^0; \underline{\theta}_0, S_0)$ and $IF_{GM}(\mathbf{x}_\mathbf{t}^0; \underline{\theta}_0, S_0)$ are uncorrelated for $\mathbf{s} \neq \mathbf{t}$ and (25) follows.

3.2.3 Selection of the Robustifying Functions

Note from (22) that the 2-D asymptotic GM-estimator influence function for $\underline{\theta}$ is proportional to $\underline{\psi}^*(\mathbf{x}^0; \underline{\theta}_0, S_0) = \mathbf{x}W(\mathbf{x})\psi(\mathbf{x}^0; \underline{\theta}_0, S_0)$ and for S is proportional to $\chi^*(\mathbf{x}^0; \underline{\theta}_0, S_0) = W(\mathbf{x})[\psi^2(\mathbf{x}^0; \underline{\theta}_0, S_0) - B]$. Hence, robustness criteria for the influence function translate directly into similar requirements for the kernel functions $\underline{\psi}^*(\mathbf{x}^0; \underline{\theta}_0, S_0)$ and $\chi^*(\mathbf{x}^0; \underline{\theta}_0, S_0)$. Therefore, selection of $\psi(\cdot)$ and $W(\cdot)$ determine the performance of the estimator.

Hampel [4] 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. Thus, 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*, respectively, which may be found in [4]. 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

$$\psi_H(t) = \begin{cases} t, & |t| < 1 \\ 1, & t \geq 1 \\ -1, & t \leq -1 \end{cases} \tag{26}$$

is shown graphically in Figure 1 and meets the first two basic criteria. Tukey's ψ_B-function, defined by

$$\psi_B(t) = \begin{cases} t(1-t^2)^2, & |t| < 1 \\ 0, & |t| \geq 1 \end{cases} \tag{27}$$

and shown graphically in Figure 2, is a redescending function which meets all three criteria. Because ψ_B returns to zero, it provides an extra measure of

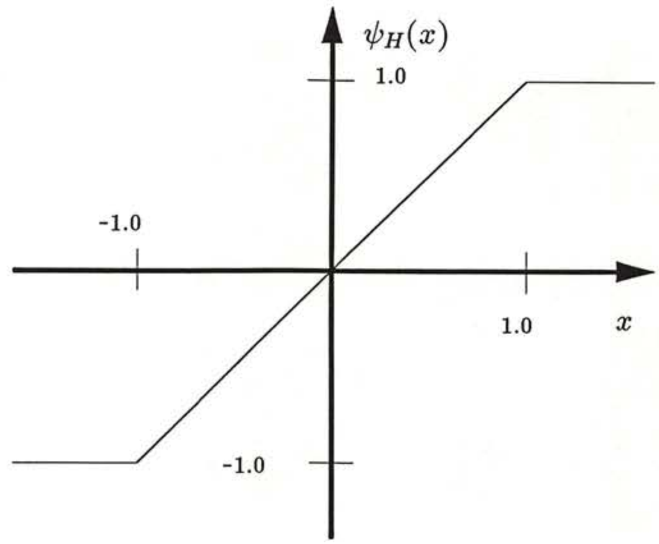

Figure 1: Huber's ψ_H-function.

robustness against extremely large outliers while sacrificing efficiency at the nominal model.

In view of the similarity between the expansions (14) and (23) we take these criteria to also apply to the two-dimensional GM-estimator influence function. Specifically, the influence function, and consequently the kernel functions $\underline{\psi}^*(\mathbf{x}^0;\underline{\theta}_0, S_0)$ and $\chi^*(\mathbf{x}^0;\underline{\theta}_0, S_0)$, 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 for each element of $\mathbf{x}$, i.e, $W(\cdot)$ should downweight 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

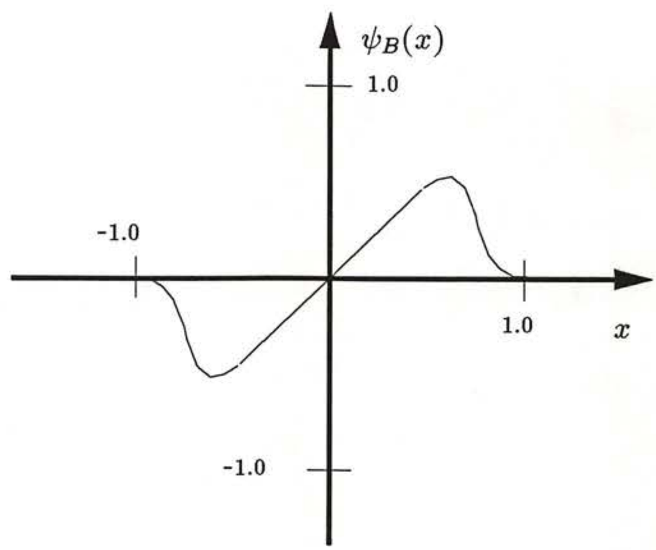

Figure 2: Tukey's redescending bisquare ψ_B-function.

$d^2 = \frac{1}{m}\mathbf{x}^t C_{\mathbf{x}}^{-1}\mathbf{x}$. Here c is a constant and $C_{\mathbf{x}} = \mathrm{E}\{\mathbf{x}\mathbf{x}^t\}$ is the $m \times m$ 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 [8]. In this paper we will not discuss this topic but instead assume that $C_{\mathbf{x}}$ is known or estimate it from clean data.

3.2.4 Tuning Constants

The tuning constant B is chosen to make the estimate $S_{AGM} = \beta$ when the signal $y(\mathbf{s})$ is observed without error and $w(\mathbf{s})$ is distributed as $\mathcal{N}(0,1)$ in (3). Under these circumstances (10) becomes

$$\int W(\mathbf{x}) \left[c_v^2 \psi^2 \left(\frac{x_0 - \underline{\theta}^t \mathbf{x}}{c_v \beta} \right) - B \right] d\mathcal{F}_y^{N_0}(\mathbf{x}^0) = 0. \tag{28}$$

since for the IO model $\underline{\theta}_{AGM} = \underline{\theta}$ so long as $\mathcal{F}_w$ is symmetric. The tuning constant is then computed by

$$B = \int c_v^2 \psi^2 \left(\frac{v}{c_v} \right) f_v(v) dv \tag{29}$$

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

Proof: Clearly, the data is also 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 $\mathcal{F}_y^{N_0}$, i.e., $\mathcal{F}_y^{N_0} = \int_{-\infty}^{\underline{\mathbf{X}}^0} f_y(\mathbf{x}^0) d\mathbf{x}^0$. Additionally, this joint density can be written in terms of a conditional density and a m-dimensional joint density as $f_y(\mathbf{y}) = f_y(y_0|\mathbf{y}) f_y(\mathbf{y})$. Using these relationships (29) becomes

$$\int W(\mathbf{x}) \left\{ \int \left[c_v^2 \psi^2 \left(\frac{x_0 - \underline{\theta}^t \mathbf{x}}{c_v \beta} \right) - B \right] f_y(x_0|\mathbf{x}) dx_0 \right\} f_y(\mathbf{x}) d\mathbf{x} = 0. \tag{30}$$

The NSHP model driven by Gaussian white noise is a unilateral Gaussian-Markov model and the conditional density $f_y(y_0|\mathbf{y})$ can be shown to be

$$f_y(y_0|\mathbf{y}) = (2\pi\beta^2)^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left(\frac{y_0 - \underline{\theta}^t \mathbf{y}}{\beta} \right)^2 \right]. \tag{31}$$

Now, let $v = \frac{x_0 - \underline{\theta}^t \mathbf{x}}{\beta}$ and make this change of variable in (30), yielding

$$\int W(\mathbf{x}) \left\{ \int \left[c_v^2 \psi^2 \left(\frac{v}{c_v} \right) - B \right] (2\pi)^{-\frac{1}{2}} \exp \left(-\frac{v^2}{2} \right) dv \right\} f_y(\mathbf{x}) d\mathbf{x} = 0. \tag{32}$$

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 (29) follows. The constant B, evaluated for selected values of c_v using both $\psi_H(\cdot)$ and $\psi_B(\cdot)$, are shown in Table 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

$$M(\psi, \mathcal{F}^{N_0}) = \begin{bmatrix} M_{\underline{\theta}} & \underline{m}_{\underline{\theta},S} \\ \underline{m}_{S,\underline{\theta}}^t & m_S \end{bmatrix} \tag{33}$$

and rewrite the influence function as

$$IF_{GM}(\mathbf{x}^0; T(\mathcal{F}^{N_0})) = M^{-1}(\psi, \mathcal{F}^{N_0}) \begin{bmatrix} \underline{\psi}^*(\mathbf{x}^0; \underline{\theta}_0, S_0) \\ \chi^*(\mathbf{x}^0; \underline{\theta}_0, S_0) \end{bmatrix}. \tag{34}$$

The asymptotic variance in (25) becomes

$$V(T,\mathcal{F}^{N_0}) = M^{-1}(\psi,\mathcal{F}^{N_0}) \begin{bmatrix} \int \underline{\psi}^* \underline{\psi}^{*t} d\mathcal{F}^{N_0} & \int \underline{\psi}^* \chi^* d\mathcal{F}^{N_0} \\ \int \chi^* \underline{\psi}^{*t} d\mathcal{F}^{N_0} & \int \chi^{*2} d\mathcal{F}^{N_0} \end{bmatrix} M^{-t}(\psi,\mathcal{F}^{N_0}). \quad (35)$$

For the IO model with Gaussian distribution, i.e., $w(\mathbf{s})$ in (3), IID with common distribution $\mathcal{N}(0,1)$, the asymptotic estimates $\underline{\theta}_0$ and S_0 equal the true values $\underline{\theta}$ and β, respectively. Moreover, if $v(\mathbf{s}) = \frac{x(\mathbf{s}) - \underline{\theta}_0^t \mathbf{x_s}}{S_0}$, then $v(\mathbf{s})$ is IID with common distribution $f_v(v) = \mathcal{N}(0,1)$. It is now a simple matter to show that

$$V(T,\mathcal{F}^{N_0}) = \begin{bmatrix} M_{\underline{\theta}}^{-1} Q(\underline{\psi}^*,\mathcal{F}^{N_0}) M_{\underline{\theta}}^{-t} & \mathbf{0} \\ \mathbf{0}^t & m_S^{-2} P(\chi^*,\mathcal{F}^{N_0}) \end{bmatrix} \quad (36)$$

where

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

and

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

The expressions for $M_{\underline{\theta}}$ and m_S can also be simplified to

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

where

$$C_{W\mathbf{x}} = \mathrm{E}\{W(\mathbf{x})\mathbf{x}\mathbf{x}^t\}$$

$$\overline{W}_1 = \mathrm{E}\{W(\mathbf{x})\}.$$

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

Using these results in (36) the asymptotic covariance matrix for the estimate $\underline{\theta}_0$ is then

$$V(\underline{\theta}, \mathcal{F}^{N_0}) = \beta^2 \frac{\int \psi^2(v) f_v(v) dv}{[\int \psi'(v) f_v(v) dv]^2} C_{W\mathbf{x}}^{-1} C_{W^2\mathbf{x}} C_{W\mathbf{x}}^{-1}, \tag{37}$$

and similarly the asymptotic variance of the estimate S_0 is

$$V(S, \mathcal{F}^{N_0}) = \frac{\beta^2}{4} \frac{\int [\psi^2(v) - B]^2 f_v(v) dv}{[\int \psi'(v)\psi(v) v f_v(v) dv]^2} \frac{\overline{W}_2}{\overline{W}_1^2}. \tag{38}$$

Ordinary M-estimators are defined when $W(\mathbf{x}) \equiv 1$ for all $\mathbf{x}$. In this case $C_{W\mathbf{x}} = C_{W^2\mathbf{x}} = C_{\mathbf{x}}$, the covariance matrix of the m-dimensional data vector $\mathbf{x_s}$, $\overline{W}_1 = \overline{W}_2 = 1$, and (37) and (38) are the asymptotic variances of the M-estimators for $\underline{\theta}$ and β.

If in addition $\psi(v) = v$ for all v, then the GM-estimator reduces to the least squares estimator of $\underline{\theta}$ and β. Carrying out the computations in (37) and (38) for the Gaussian situation yields

$$V_{LS}(\underline{\theta}, \mathcal{N}(0, \sigma_y^2)) = \beta^2 C_{\mathbf{x}}^{-1} \tag{39}$$

and

$$V_{LS}(S, \mathcal{N}(0, \sigma_y^2)) = \frac{\beta^2}{2}. \tag{40}$$

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 $\underline{\theta}$ reduces to

$$V(\underline{\theta}, \mathcal{N}(0, \sigma_y^2)) = \beta^2 \frac{\int \psi^2(v) f_v(v) dv}{\left[\int \psi'(v) f_v(v) dv\right]^2} C_{\mathbf{x}}^{-1}. \tag{41}$$

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 Fisher information matrix $I(\underline{\theta})$ turns out to be given by

$$I(\underline{\theta}) = \frac{1}{\beta^2} C_{\mathbf{x}} + \frac{f'(\mathbf{x})}{f(\mathbf{x})} \tag{42}$$

where $f'(\mathbf{x}) = \frac{\partial}{\partial \underline{\theta}} f(\mathbf{x})$ which is a complicated function of the parameters $\underline{\theta}$. 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 then the distribution function is not even available.

We are motivated then to define an aysmptotic efficiency relative to the least squares estimates, although consistent for the NSHP and GMRF models, are

	ψ_H		ψ_B	
c_v	B	REff_{GM}	B	REff_{GM}
1.0	0.5161	0.9031	0.0266	
1.1	0.5777	0.9191	0.0343	
1.2	0.6352	0.9330	0.0433	
1.3	0.6880	0.9451	0.0537	
1.4	0.7358	0.9555	0.0652	
1.5	0.7785	0.9642	0.0778	
2.0	0.9205	0.9897	0.1552	0.5017
3.0	0.9950	0.9996	0.3434	0.7748
4.0	0.9999	1.0000	0.5134	0.9101
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 1: Tuning constant B and relative efficiency for selected values of c_v using ψ_H and ψ_B. (REff_{GM} for ψ_B and values of $c_v < 2.0$ are not shown because of computational inaccuracies.)

not efficient. Thus,

$$\text{Eff}_{GM} \stackrel{\text{def}}{=} \frac{\text{tr}\{V_{LS}(\underline{\theta}, \mathcal{F}^{N_0})\}}{\text{tr}\{V(\underline{\theta}, \mathcal{F}^{N_0})\}}. \tag{43}$$

Substituting the expressions for $V_{LS}(\underline{\theta}, \mathcal{N}(0, \sigma_y^2))$ and $V(\underline{\theta}, \mathcal{N}(0, \sigma_y^2))$ in (43) yields the relative GM-estimator efficiency

$$\text{REff}_{GM} = \frac{\int \psi^2(v) f_v(v) dv}{[\int \psi'(v) f_v(v) dv]^2}. \tag{44}$$

Values of REff_{GM}, computed for various values of the tuning constant c_v are shown in Table 1.

3.3 2-D Filter-Cleaner

An alternative to robust parameter estimation and an intuitively appealing idea is to "clean" the possibly contaminated data $\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$ before computing the spectrum. If $x(\mathbf{s})$ differs too much from a robust prediction $\hat{x}(\mathbf{s})$ based on the other values, then $x(\mathbf{s})$ is replaced by a value closer to $\hat{x}(\mathbf{s})$. This procedure is applicable to NCAR as well as NSHP and GMRF models, since consistent maximum likelihood estimation can be incorporated in the algorithm. We describe this procedure assuming $y(\mathbf{s})$ in (1) to be modeled by a NCAR model.

The algorithm is begun by first fitting a NCAR model, (45), to the observed data $\{x(\mathbf{s}),\ \mathbf{s} \in \Omega_M\}$.

$$x(\mathbf{s}) = \sum_{\mathbf{r} \in N} \theta_{\mathbf{r}} x(\mathbf{s}+\mathbf{r}) + w(\mathbf{s}) \tag{45}$$

To estimate $\underline{\theta} = \mathrm{col}\{\theta_{\mathbf{r}}, \mathbf{r} \in N\}$ a consistent estimator, such as maximum-likelihood (ML) for the NCAR model, is used.

The variance of the residuals $v(\mathbf{s})$ resulting from the fitting of (45) to the data is computed using a robust estimator for scale. For example, the median absolute deviation (MAD) works well (see [4] or [5]).

$$S_{MAD} = \frac{\mathrm{median}\{|v(\mathbf{s}) - \mathrm{median}\{v(\mathbf{s})\}|\}}{0.6745}. \tag{46}$$

The factor of 0.6745 in (46) is to make the estimate of scale consistent when the data is truly Gaussian [5].

The residuals are then passed through a nonlinear filter for cleaning:

$$\hat{v}(\mathbf{s}) = c_v S_{MAD} \psi \left[\frac{v(\mathbf{s})}{c_v S_{MAD}} . \right] \tag{47}$$

The function $\psi(\cdot)$ is one of the ψ-functions described in Section 3.2.1 and c_v is a tuning constant to adjust robustness. From these cleaned residuals we generate an estimate of $y(\mathbf{s})$ as

$$\hat{y}(\mathbf{s}) = \sum_{\mathbf{r} \in N} \hat{\theta}_{\mathbf{r}} \hat{y}(\mathbf{s} + \mathbf{r}) + \hat{v}(\mathbf{s}). \tag{48}$$

To solve this equation we need to make assumptions regarding the boundary conditions. If a 2-D toroidal lattice is assumed, then

$$\hat{\mathbf{y}} = B^{-1}(\hat{\underline{\theta}}) \hat{\mathbf{v}} \tag{49}$$

where $\hat{\mathbf{y}}$ and $\hat{\mathbf{v}}$ are vectors whose components are the lexicographically ordered $\hat{y}(\mathbf{s})$ and $\hat{v}(\mathbf{s})$, respectively, and $B(\hat{\underline{\theta}})$ is a block-circulant transformation matrix. Solution of this equation is efficiently done using the 2-D discrete Fourier transform. See [13] for details. This completes the first iteration.

The estimated values of $\hat{y}(\mathbf{s})$ are used in (45) instead of $x(\mathbf{s})$ to start a new iteration, and this process is continued until small changes in the values of $\hat{\underline{\theta}}$ occur.

Finally, the NCAR spectrum is computed from (4) using $\beta = S_{MAD}$ and $\underline{\theta} = \hat{\underline{\theta}}$. Alternatively, one may replace β^2 in (4) with $\mathcal{S}_v(\lambda)$, the periodogram of the estimated residuals, as suggested in [10].

The filter-cleaner algorithm is schematically shown in Figure 3. In the next section we give an interpretation of this algorithm.

3.4 An Interpretation of the Iterative Algorithm

Kleiner et al. [6] showed that the 1-D filter-cleaner iterative algorithm can be interpreted as an approximate solution to a minimization problem. This interpretation is easily extended to two dimensions by considering (50).

$$\min_{\underline{\theta}} \left\{ \sum_{\mathbf{s}\in\Omega} c_v^2 \rho \left[\frac{x(\mathbf{s}) - \underline{\theta}^t \hat{\mathbf{y}}_{\mathbf{s}}(\underline{\theta})}{c_v S} \right] - \log[\det B(\underline{\theta})] \right\} \tag{50}$$

where

$$\hat{\mathbf{y}}_{\mathbf{s}}(\underline{\theta}) = \operatorname{col}\left[\hat{y}(\mathbf{s}+\mathbf{r}),\ \ \mathbf{r} \in N\right].$$

$\hat{y}(\cdot)$ is an estimate of the underlying NCAR process. Differentiating (50) with respect to $\underline{\theta}$ at $\underline{\theta} = \hat{\underline{\theta}}$ we have

$$\sum_{\mathbf{s}\in\Omega} \frac{c_v}{S} \psi \left[\frac{x(\mathbf{s}) - \hat{\underline{\theta}}^t \hat{\mathbf{y}}_{\mathbf{s}}(\hat{\underline{\theta}})}{c_v S} \right] \left[\hat{\mathbf{y}}_{\mathbf{s}}(\hat{\underline{\theta}}) + G(\mathbf{s}, \hat{\underline{\theta}}) \hat{\underline{\theta}} \right] + \frac{\partial}{\partial \underline{\theta}} \log[\det B(\hat{\underline{\theta}})] = \mathbf{0} \tag{51}$$

where

$$G(\mathbf{s}, \underline{\theta})_{k,l} = \frac{\partial \hat{\mathbf{y}}(\mathbf{s} + \mathbf{r}_k, \underline{\theta})}{\partial \theta_{\mathbf{r}_l}},\ \ \mathbf{r}_k, \mathbf{r}_l \in N$$

and

$$\psi(x) = \rho'(x).$$

If $G(\mathbf{s}, \hat{\underline{\theta}})\hat{\underline{\theta}}$ is small compared to $\hat{\mathbf{y}}_{\mathbf{s}}(\hat{\underline{\theta}})$, then (51) reduces to

$$\sum_{\mathbf{s}\in\Omega} \frac{c_v}{S} \psi \left[\frac{\hat{x}(\mathbf{s}) - \hat{\underline{\theta}}^t \hat{\mathbf{y}}_{\mathbf{s}}(\hat{\theta})}{c_v S} \right] \hat{\mathbf{y}}_{\mathbf{s}}(\hat{\underline{\theta}}) + \frac{\partial}{\partial \underline{\theta}} \log[\det B(\hat{\underline{\theta}})] = \mathbf{0}. \tag{52}$$

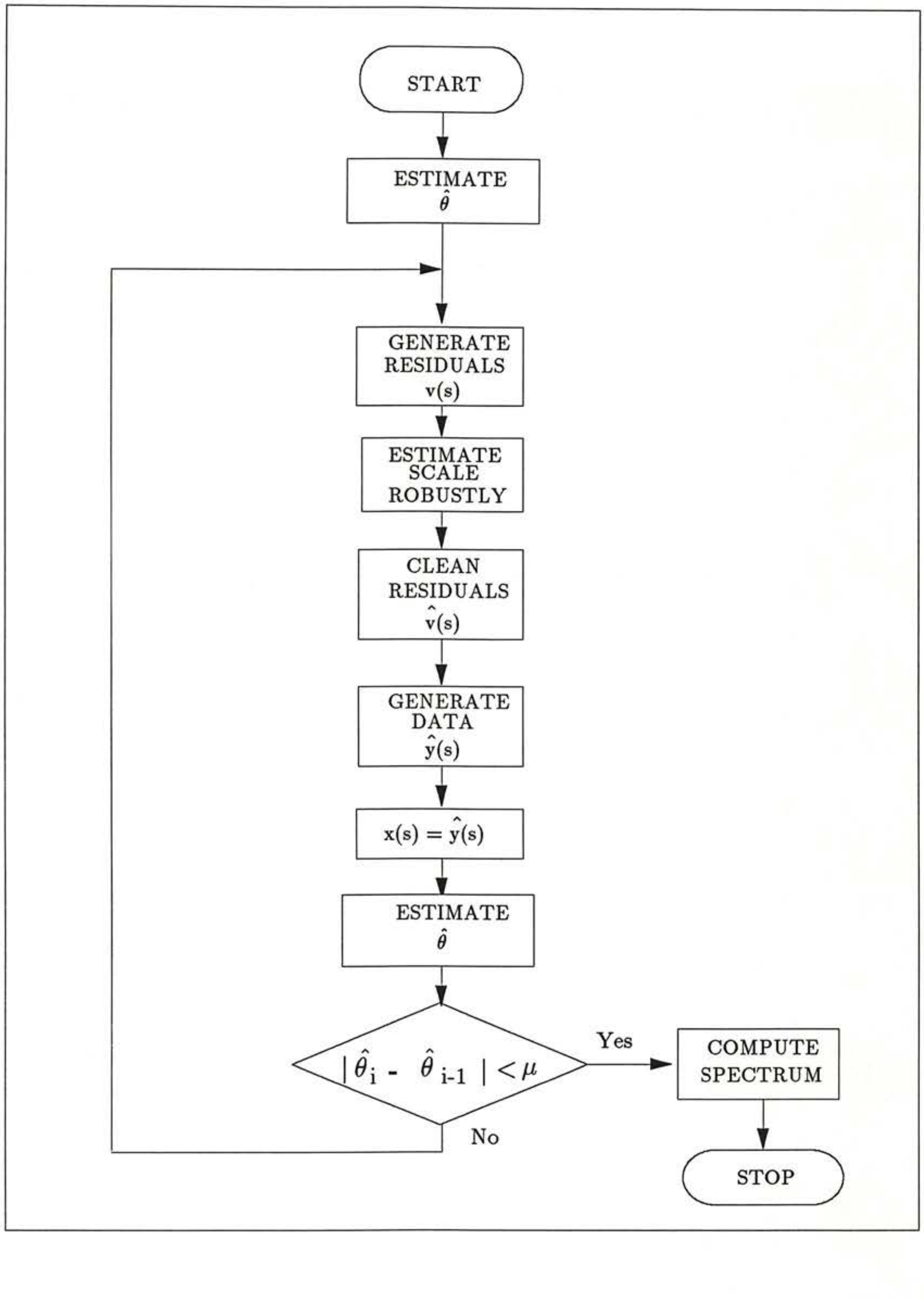

Figure 3: Filter-cleaner algorithm.

But for a fixed scale S, the ML estimate [9] is a solution of the following minimization problem

$$\min_{\underline{\theta}} \left\{ \sum_{\mathbf{s}\in\Omega} \frac{1}{2} \left[\frac{\hat{y}(\mathbf{s}) - \underline{\theta}^t \hat{\mathbf{y}}_{\mathbf{s}}}{S} \right]^2 - \log[\det B(\underline{\theta})] \right\}. \tag{53}$$

Differentiating (53) with respect to $\underline{\theta}$ we get

$$\sum_{\mathbf{s}\in\Omega} \left[\frac{\hat{y}(\mathbf{s}) - \underline{\theta}^t \hat{\mathbf{y}}_{\mathbf{s}}}{S} \right] \frac{\hat{\mathbf{y}}_{\mathbf{s}}}{S} + \frac{\partial}{\partial \underline{\theta}} \log[\det B(\underline{\theta})] = \mathbf{0}. \tag{54}$$

Using (47), (54) can be written as

$$\sum_{\mathbf{s}\in\Omega} \frac{c_v}{S} \psi \left[\frac{x(\mathbf{s}) - \underline{\theta}^t \mathbf{x}_{\mathbf{s}}}{c_v S} \right] \hat{\mathbf{y}}_{\mathbf{s}}(\underline{\theta}) + \frac{\partial}{\partial \underline{\theta}} \log[\det B(\underline{\theta})] = \mathbf{0} \tag{55}$$

which is an approximation to (52). Thus, the robust filtering algorithm is approximately equivalent to computing the estimates (52) until convergence is observed. Equation (52) is itself an approximation to the minimization problem in (50).

4 Simulation Examples

Experiments with synthetic data were conducted to evaluate the performance of the two-dimensional GM-estimator. The first experiment compares robust spectrum estimation results for two sinusoids in noise with the conventional approach of consistent least squares estimation. It is well known that NSHP modeling of complex spectra requires large model order to resolve details [3];

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. No spectra were computed in Experiments 2, 3, and 4 since the spectrum for low order models has no interesting detail. Experiment 5 provides similar results for evaluating the two-dimensional filter-cleaner algorithm.

4.1 2-D Robust Parameter Estimation

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

$$y(\mathbf{s}) = A_1 \cos(\frac{\pi}{64}\underline{\phi}_1^t \mathbf{s} + \alpha_1) + A_2 \cos(\frac{\pi}{64}\underline{\phi}_2^t \mathbf{s} + \alpha_2) + \zeta(\mathbf{s})$$

with $\underline{\phi}_1 = [16.0, 16.0]^t$, $\underline{\phi}_2 = [16.0, 20.0]^t$, $A_1 = A_2 = 1.0$, $\alpha_1 = .2$, $\alpha_2 = .3$ and $\zeta(\mathbf{s})$ IID with common distribution $\mathcal{N}(0, .05)$, equivalent to a SNR = 10db. In this experiment a NSHP model with the 40 element neighbor set shown in Figure 4 was used.

The estimated spectrum of the signal plus Gaussian noise using conventional consistent least squares estimates is shown in Figure 5, and the spectrum using the GM-estimator on the same data is shown in Figure 6. Huber's ψ_H-function was chosen for $\psi(\mathbf{x}^0; \underline{\theta}_0, S_0)$ 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 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 $\mathbf{s} = [16.0, 16.0]^t$ and $\mathbf{s} = [16.0, 20.0]^t$, which are correct. Thus, the GM-estimator does almost equally well as the conventional procedure in the Gaussian situation.

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 least squares estimates from the contaminated data is shown in Figure 7, and the spectrum computed using the GM-estimate is shown in Figure 8. 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.

Experiment 2. In the first part of this experiment we generated data using a 1st-order Gaussian-Markov random field model on a 32×32 toroidal lattice with parameter values $\theta_{1,0} = \theta_{-1,0} = .2340$, $\theta_{0,1} = \theta_{0,-1} = .1011$, and $\nu = 1.0$. 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 least squares estimates and robust GM-estimates from the contaminated data sets were computed and compared to the theoretical parameter values and the least squares 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 2. Figure 9 shows graphically for comparison the errors tabulated in the last column of Table 2.

One easily sees that the GM-estimates for both $\underline{\theta}$ 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-estimates are shown in Table 3 and errors graphed in Figure 10. The conclusions are identical to the first part of this experiment.

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 two-dimensional non-symmetric half-plane autoregressive data using a causal neighbor set and 12 sets of IID Gaussian random noise fields. The model parameters were $\theta_{-1,0} = .9704$, $\theta_{0,-1} = .9735$, $\theta_{-1,-1} = -.9686$, and $\beta^2 = 1.0$. 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. LS and GM-estimates of the clean data and contaminated data are listed in Table 4 and errors graphed in Figure 11.

Note that for the IO model LS and GM-estimators do equally well estimating $\underline{\theta}$ 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 $\underline{\theta}$ for the IO model.

In the second part of this experiment we used the 12 sets of two-dimensional 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. LS and GM-estimates were computed and the parameter estimates compared with the LS estimates from the clean data. The results are shown in Table 5 and errors graphed in Figure 12.

Again the GM-estimator outperformed the LS-estimator in the non-Gaussian situation.

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 (27) for both $\psi(\cdot)$ and $g(\cdot)$. The tuning constants were $B = 0.7277$ and $c_v = c = 6.0$ in accordance with Table 1. The tabulated estimates are shown in Table 6 with the corresponding errors graphed in Figure 13 and Table 7 with the corresponding errors graphed in Figure 14 for parts one and two of the experiment, respectively. Figures 15 and 16 compare the estimates of $\underline{\theta}$ using a redescending function with the results in Experiment 1. No conclusions can be drawn regarding improvement in estimates of $\underline{\theta}$ using a redescending ψ-function.

Summary. Tables 8 and 9 summarize the results of Experiments 2 , 3 and 4. Table 8 shows the average of the squared errors for the coefficients $\{\theta_{\mathbf{r}},\ \mathbf{r} \in N\}$

estimates, and Table 9 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 least squares 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 least squares 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 least squares 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 2) but will effect scale estimates. Too few experiments were run to draw any significant conclusions regarding the use of redescending ψ-functions. We do note, however, that while the estimates of $\underline{\theta}$ are relatively close for both the redescending and non-redescending situations, the estimates of β^2 seem to be improved with the redescending function.

4.2 2-D Filter-Cleaner

Experiment 5. To evaluate the two-dimensional robust filter-cleaner algorithm simulations for a NCAR model using the filter-cleaner algorithm were carried out using the following first order NCAR model with $\theta_{(1,0)} = \theta_{(-1,0)} = .2340$,

$\theta_{(0,1)} = \theta_{(0,-1)} = .1011$ and $\beta^2 = 1.0$.

Data obeying this model were generated on a toroidal 32×32 lattice. Outliers were generated by first calling a uniform random number generator to give 50 random lattice points. Then 50 independent Gaussian random variates having zero mean and variance of 20.25 were added to the NCAR data at the random lattice points. For comparison purposes ML estimates of parameters were calculated before adding the outliers. Then after adding the outliers we computed the conventional ML estimates again. The iterative procedure using Huber's ψ_H-function for cleaning the residuals was applied with $c_v = 2.0$ as a tuning constant in (47).

This experiment was carried out in two parts. First, for 10 runs we kept the outliers the same but changed the underlying NCAR field. Estimates are shown in Table 10. Second, for another 10 runs we kept the underlying NCAR field constant but changed the outliers. Estimates are show in Table 11. As seen from Table 10 and Table 11, the outliers decreased the numerical value of the ML estimates of parameters and increased the estimated variance in all runs. But the comparisons in Figures 17 and 18 demonstrate that the filter-cleaner improves the estimates. Most runs required only three or four iterations for convergence.

5 Conclusions

These intuitively appealing robust procedures provide simple robust spectrum estimates. The experimental results conclusively show that the GM-estimator and filter-cleaner do better than the conventional least squares estimates for non-Gaussian IO and AO model situations where the non-normality results from outliers distributed uniformly over the observed lattice and with small probability of occurrence. Furthermore, the empirical evidence shows that the GM-estimator also does well when noise has been substituted for the true data.

There remain several problems to be investigated. First, the definition of an influence function for two-dimensional data is not complete since the contaminated data distribution depends not only on the magnitude of the contamination but also its location on the lattice. Second, the robust GM parameter estimator, extended from the 1-D approach, has no theoretical underpinnings, but only a heuristic application of Hampel's qualitative robustness requirements that the influence function of a robust estimator be bounded and continuous. Third, analysis of the GM-estimator is only practical for the IO model since analysis in the presence of additive outliers is extremely difficult as expressions for the ML estimates under a non-Gaussian distribution are not easily computed. Moreover, in the spirit of robustness the true distribution of the observed data is unknown, and without this information we are unable to compute the ML

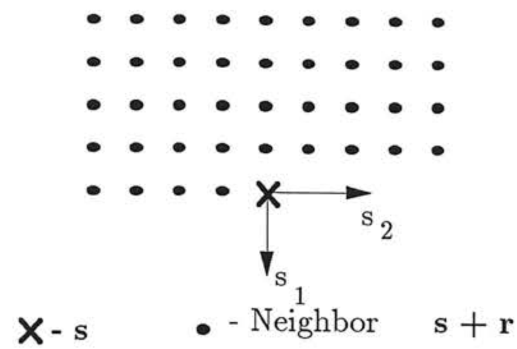

Figure 4: Neighbor set.

estimates. Similar problems occur for the filter-cleaner. For example, to this date it has not even been proved that the cleaned data are stationary [4]. We are currently investigating these problems and hope to present the results soon.

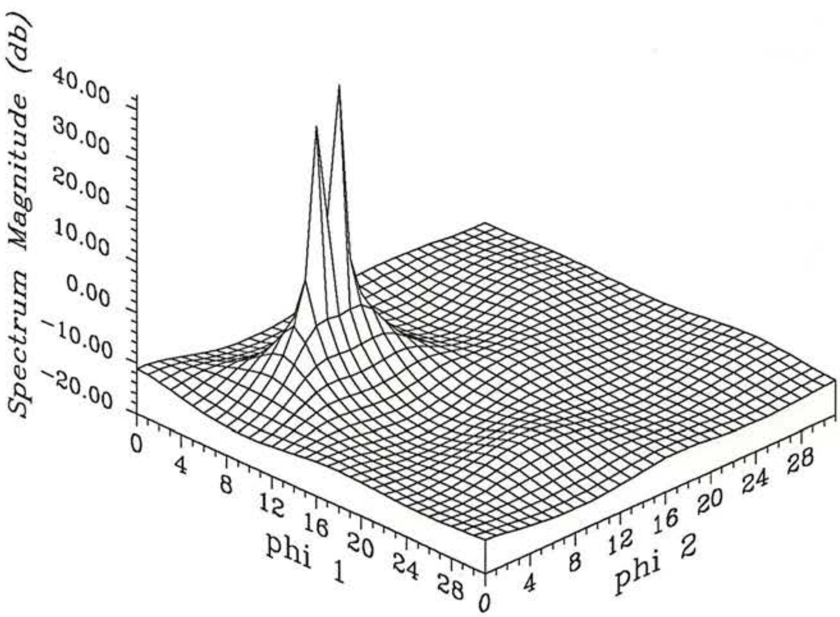

Figure 5: Spectrum of two sinusoids in Gaussian noise. SNR = 10db. Least squares estimate.

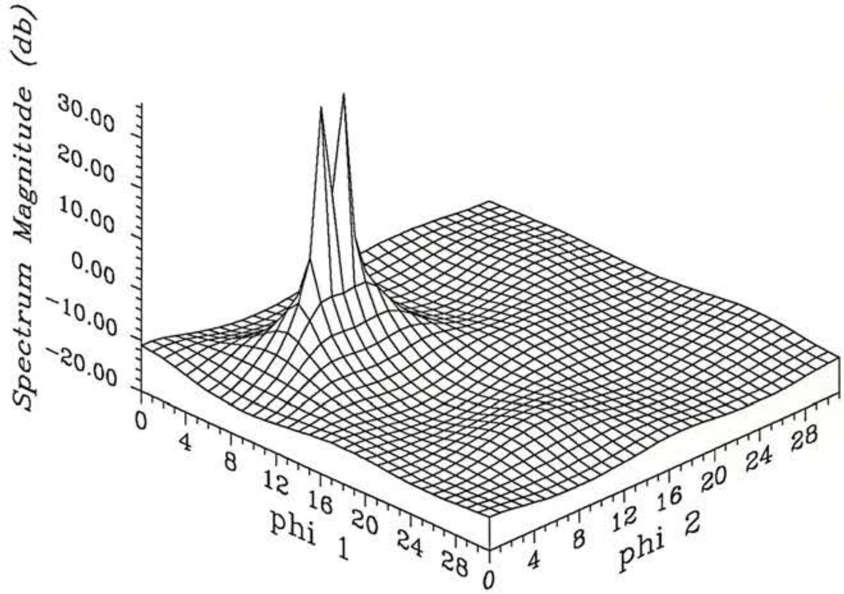

Figure 6: Spectrum of two sinusoids in Gaussian noise. SNR = 10db. Robust estimate.

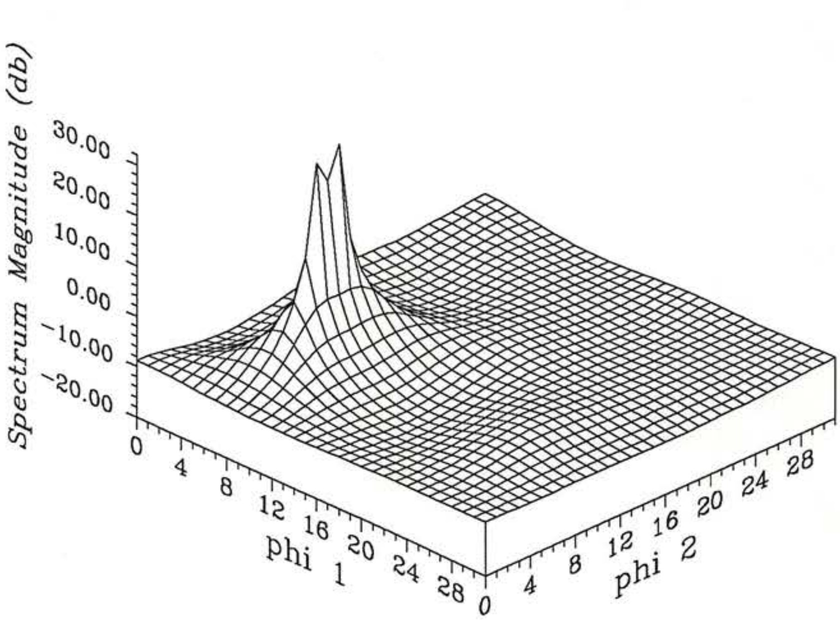

Figure 7: Spectrum of two sinusoids in non-Gaussian noise. SNR = 10db and 10 percent additive outliers. Least squares estimate.

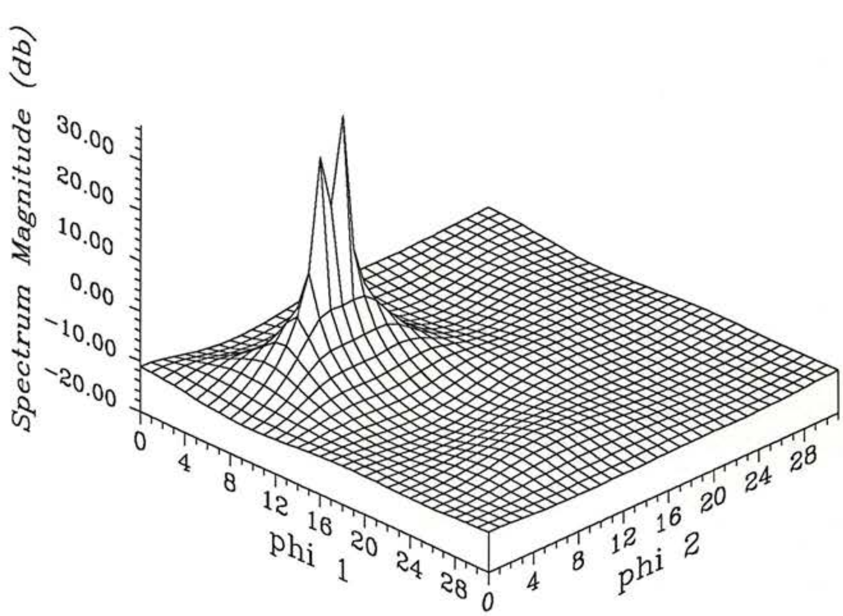

Figure 8: Spectrum of two sinusoids in non-Gaussian noise. SNR = 10db and 10 percent additive outliers. Robust estimate.

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

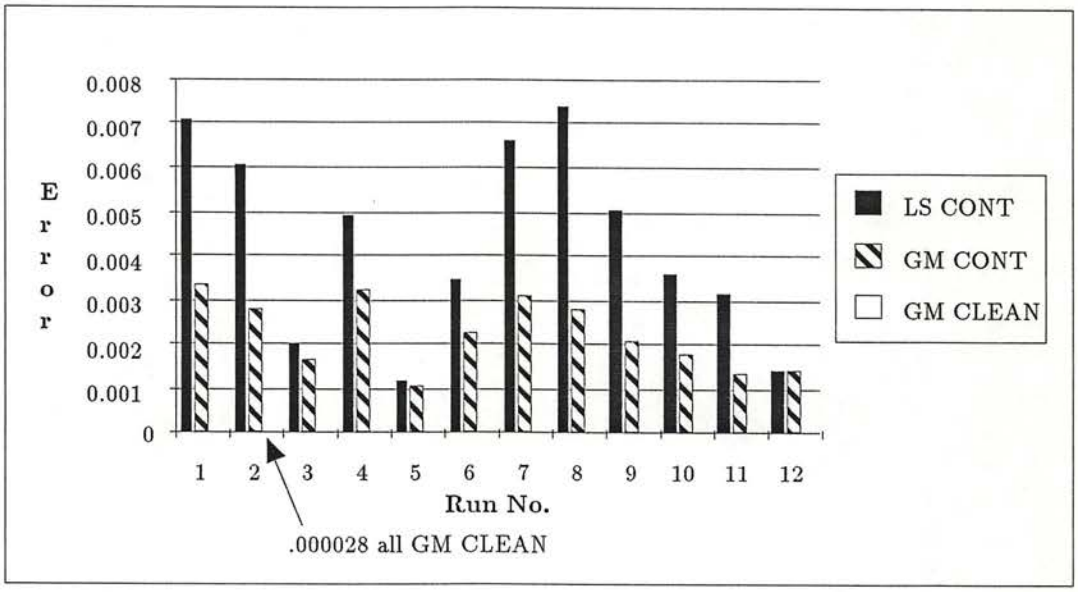

Figure 9: Errors for the results shown in Table 2. CONT signifies contaminated data, i.e., Gaussian data with outliers, and CLEAN signifies Gaussian data without outliers.

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	
1	LS Clean Data	.236163	.106680	1.00565	
	LS Contaminated Data	.133592	.045695	2.48139	.007120
	GM Contaminated Data	.159626	.076479	1.27725	.003385
	GM Clean Data	.229264	.103671	1.03459	.000028
2	LS Clean Data	.240428	.114784	1.03011	
	LS Contaminated Data	.088360	.067536	2.55879	.012679
	GM Contaminated Data	.133752	.092041	1.28817	.005948
	GM Clean Data	.239153	.114014	1.05122	.000001
3	LS Clean Data	.242595	.107649	1.01870	
	LS Contaminated Data	.114572	.039206	2.54770	.010537
	GM Contaminated Data	.162640	.071597	1.26652	.003846
	GM Clean Data	.241548	.115400	1.04944	.000031
4	LS Clean Data	.223647	.073663	0.99701	
	LS Contaminated Data	.056664	.016945	2.51918	.015550
	GM Contaminated Data	.119442	.066373	1.27842	.005456
	GM Clean Data	.226468	.079048	1.03414	.000018
5	LS Clean Data	.282265	.093625	0.93248	
	LS Contaminated Data	.103913	.038853	2.55393	.017405
	GM Contaminated Data	.166528	.085191	1.19616	.006733
	GM Clean Data	.275986	.100556	0.93030	.000044
6	LS Clean Data	.248693	.074178	1.00454	
	LS Contaminated Data	.078250	.044128	2.56529	.014977
	GM Contaminated Data	.145491	.057209	1.28177	.005469
	GM Clean Data	.250922	.075576	1.01768	.000004
7	LS Clean Data	.272333	.063656	0.97107	
	LS Contaminated Data	.121358	.013489	2.50059	.012655
	GM Contaminated Data	.177642	.061307	1.23972	.004486
	GM Clean Data	.271213	.067591	1.02253	.000008
8	LS Clean Data	.214168	.125415	1.08745	
	LS Contaminated Data	.102068	.061869	2.45701	.008302
	GM Contaminated Data	.131741	.096509	1.31228	.003815
	GM Clean Data	.215870	.125856	1.06614	.000002
9	LS Clean Data	.232649	.081906	0.99928	
	LS Contaminated Data	.115411	.064520	2.37098	.007023
	GM Contaminated Data	.152461	.082684	1.24742	.003215
	GM Clean Data	.230616	.081241	1.01009	.000002
10	LS Clean Data	.245452	.108255	0.92095	
	LS Contaminated Data	.080360	.040073	2.47818	.015952
	GM Contaminated Data	.135570	.079535	1.17925	.006449
	GM Clean Data	.241974	.113848	0.95411	.000022
11	LS Clean Data	.269604	.101009	0.97095	
	LS Contaminated Data	.125567	.040974	2.50815	.012176
	GM Contaminated Data	.170548	.058331	1.24076	.005817
	GM Clean Data	.269864	.099759	0.99666	.000001
12	LS Clean Data	.190854	.153230	1.04144	
	LS Contaminated Data	.093560	.051904	2.52522	.009867
	GM Contaminated Data	.121134	.109920	1.30009	.003368
	GM Clean Data	.189772	.162675	1.06802	.000045

Table 3: GM-estimator results for additive outliers with changing GMRF .

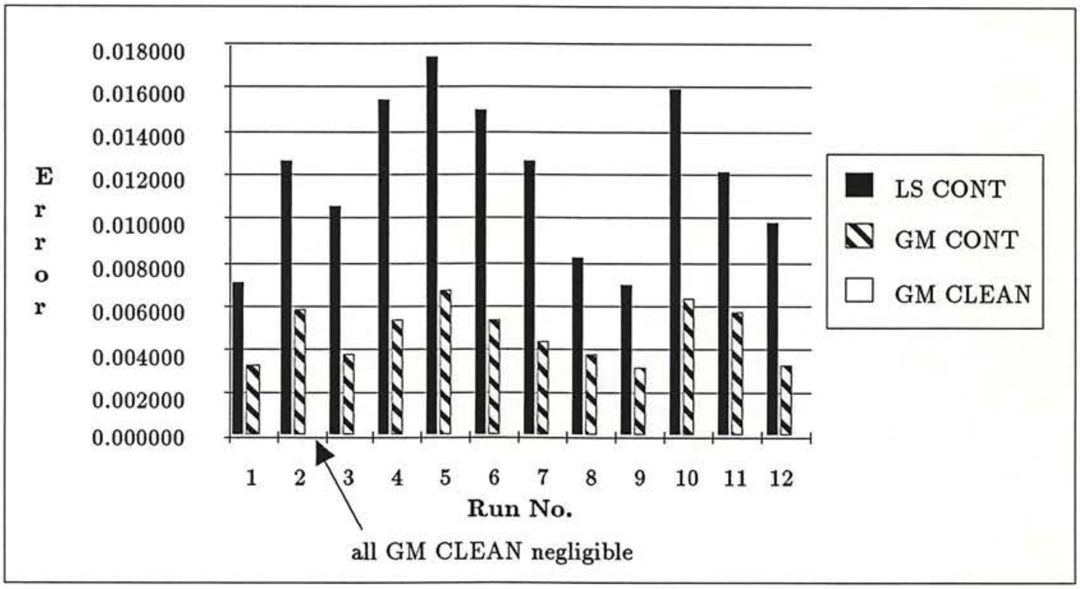

Figure 10: Errors for the results shown in Table 3.

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	
1	LS Clean Data	.987543	.990750	−0.989295	1.00068	
	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
2	LS Clean Data	.978444	.983040	−0.980719	0.98394	
	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
3	LS Clean Data	.985621	.987905	−0.985635	0.99660	
	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
4	LS Clean Data	.983659	.987190	−0.982041	1.04070	
	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
5	LS Clean Data	.985335	.991773	−0.987939	1.03691	
	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
6	LS Clean Data	.985653	.985634	−0.983150	0.95784	
	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
7	LS Clean Data	.983893	.986260	−0.984334	0.89570	
	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
8	LS Clean Data	.986759	.987601	−0.990171	1.00272	
	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
9	LS Clean Data	.988944	.987261	−0.987261	1.00454	
	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
10	LS Clean Data	.985354	.986986	−0.985492	0.97958	
	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
11	LS Clean Data	.985522	.986145	−0.989582	0.97128	
	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
12	LS Clean Data	.986957	.989961	−0.990111	1.02466	
	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 4: GM-estimator results for innovative outliers with changing NSHP field.

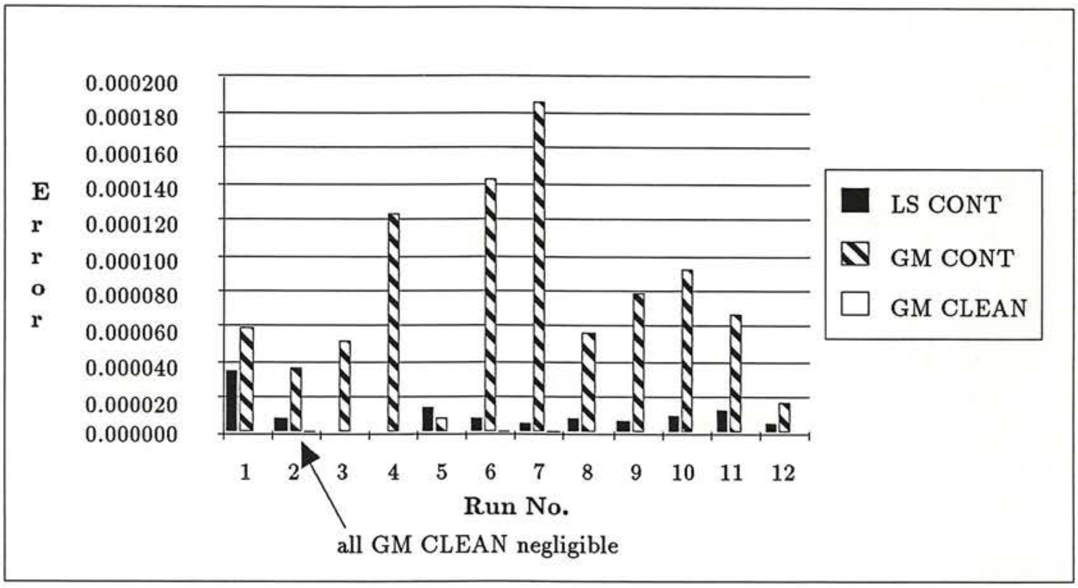

Figure 11: Errors for the results shown in Table 4.

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	
1	LS Clean Data	.236163	.106680	1.00565	
	LS Contaminated Data	.119086	.036672	2.47887	.009304
	GM Contaminated Data	.154036	.069590	1.25918	.004060
	GM Clean Data	.229264	.103671	1.03459	.000028
2	LS Clean Data	.240428	.114784	1.03011	
	LS Contaminated Data	.143098	.095234	1.63618	.004928
	GM Contaminated Data	.178739	.105796	1.18595	.001943
	GM Clean Data	.239153	.114014	1.05122	.000001
3	LS Clean Data	.242595	.107649	1.01870	
	LS Contaminated Data	.155331	.053016	1.70561	.005300
	GM Contaminated Data	.185023	.089586	1.15419	.001820
	GM Clean Data	.241548	.115400	1.04944	.000031
4	LS Clean Data	.223647	.073663	0.99701	
	LS Contaminated Data	.156333	.060597	1.84741	.002351
	GM Contaminated Data	.187209	.075707	1.13710	.000666
	GM Clean Data	.226468	.079048	1.03414	.000018
5	LS Clean Data	.282265	.093625	0.93248	
	LS Contaminated Data	.198769	.067408	1.59788	.003829
	GM Contaminated Data	.214585	.075061	1.11285	.002463
	GM Clean Data	.275986	.100556	0.93030	.000044
6	LS Clean Data	.248693	.074178	1.00454	
	LS Contaminated Data	.163208	.014778	1.83012	.005418
	GM Contaminated Data	.190015	.037523	1.14988	.002393
	GM Clean Data	.250922	.075576	1.01768	.000004
7	LS Clean Data	.272333	.063656	0.97107	
	LS Contaminated Data	.111479	.084193	2.02618	.013148
	GM Contaminated Data	.186868	.067190	1.19518	.003658
	GM Clean Data	.271213	.067591	1.02253	.000008
8	LS Clean Data	.214168	.125415	1.08745	
	LS Contaminated Data	.057887	.091075	2.30465	.012801
	GM Contaminated Data	.137629	.104363	1.25779	.003151
	GM Clean Data	.215870	.125856	1.06614	.000002
9	LS Clean Data	.232649	.081906	0.99928	
	LS Contaminated Data	.103274	.125214	1.99076	.009307
	GM Contaminated Data	.143095	.081648	1.20179	.004010
	GM Clean Data	.230616	.081241	1.01009	.000002
10	LS Clean Data	.245452	.108255	0.92095	
	LS Contaminated Data	.135159	.049256	1.98369	.007823
	GM Contaminated Data	.180199	.062672	1.13009	.003168
	GM Clean Data	.241974	.113848	0.95411	.000022
11	LS Clean Data	.269604	.101009	0.97095	
	LS Contaminated Data	.201832	.034029	1.76589	.004944
	GM Contaminated Data	.224626	.070500	1.13483	.001477
	GM Clean Data	.269864	.099759	0.99666	.000001
12	LS Clean Data	.190854	.153230	1.04144	
	LS Contaminated Data	.125807	.086365	1.65478	.004351
	GM Contaminated Data	.144559	.106571	1.27110	.002160
	GM Clean Data	.189772	.162675	1.06802	.000045

Table 5: GM-estimator results for substitutive outliers with changing GMRF .

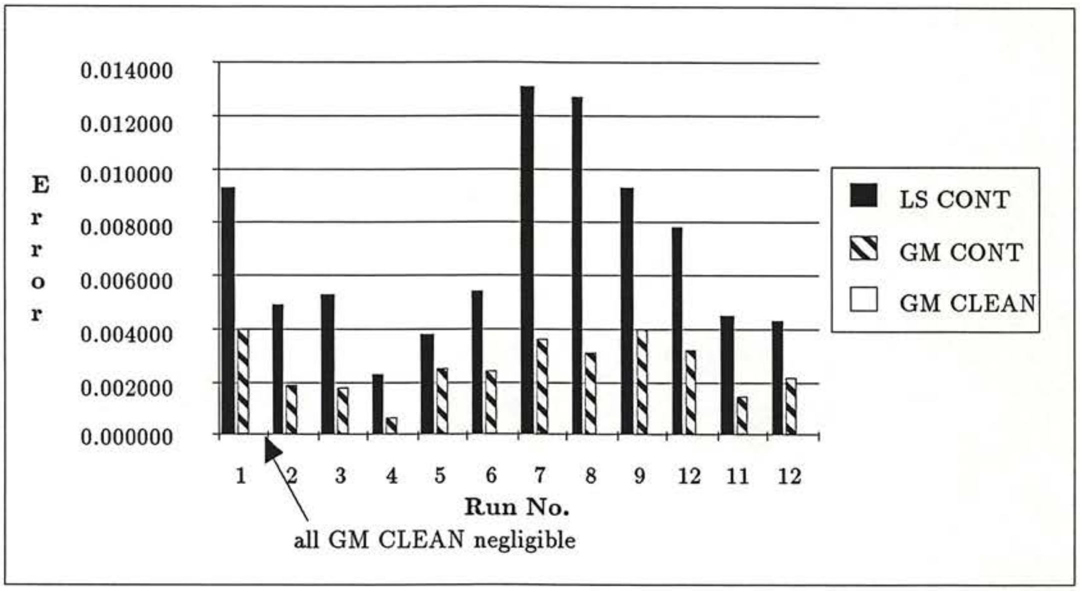

Figure 12: Errors for the results shown in Table 5.

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	.231174	.103447	1.02514	.000018
1	LS Contaminated Data	.133592	.045695	2.48139	.007120
	GM Contaminated Data	.156848	.082112	1.15611	.003447
2	LS Contaminated Data	.134955	.062564	1.70262	.006095
	GM Contaminated Data	.170820	.080460	1.11718	.002479
3	LS Contaminated Data	.179539	.076275	1.72544	.002065
	GM Contaminated Data	.181513	.086274	1.12680	.001702
4	LS Contaminated Data	.137559	.093601	2.05112	.004947
	GM Contaminated Data	.156099	.090561	1.08119	.003335
5	LS Contaminated Data	.205881	.067976	1.66953	.001207
	GM Contaminated Data	.194840	.083588	1.12199	.001120
6	LS Contaminated Data	.162914	.065956	1.84768	.003512
	GM Contaminated Data	.170706	.080176	1.08993	.002494
7	LS Contaminated Data	.121111	.102253	2.08793	.006628
	GM Contaminated Data	.159220	.082725	1.10167	.003247
8	LS Contaminated Data	.115475	.090131	2.21251	.007420
	GM Contaminated Data	.166218	.104251	1.10346	.002449
9	LS Contaminated Data	.135594	.106215	2.02817	.005057
	GM Contaminated Data	.178796	.071914	1.15400	.002250
10	LS Contaminated Data	.175925	.046927	2.07145	.003600
	GM Contaminated Data	.177207	.088759	1.15594	.001898
11	LS Contaminated Data	.169912	.062514	1.88382	.003170
	GM Contaminated Data	.185092	.087290	1.11613	.001492
12	LS Contaminated Data	.188894	.080131	1.70693	.001470
	GM Contaminated Data	.187262	.083213	1.16583	.001471

Table 6: GM-estimator using redescending ψ_B results with changing contamination in a GMRF.

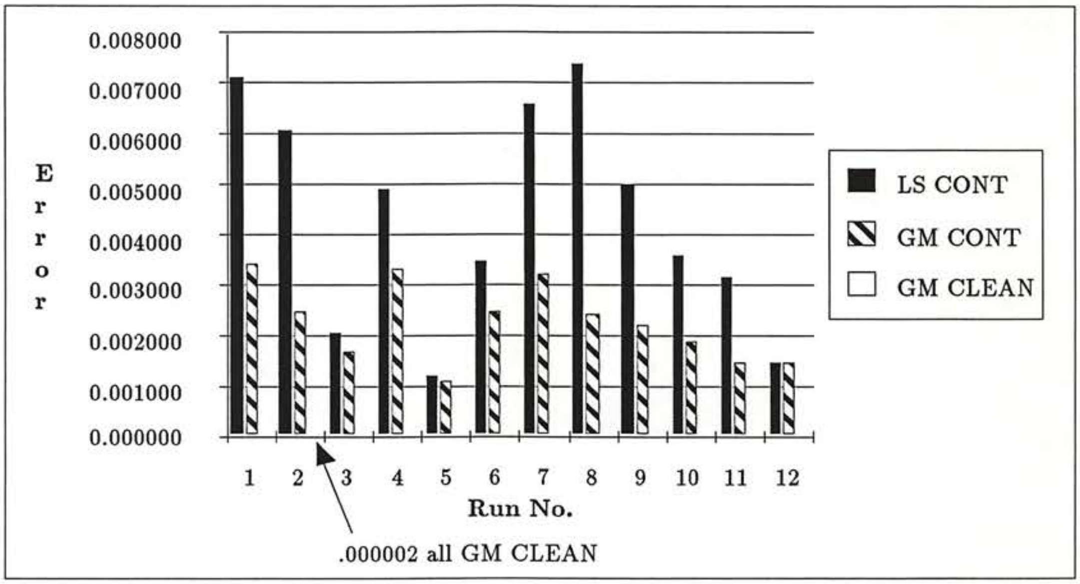

Figure 13: Errors for the results shown in Table 6.

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	
1	LS Clean Data	.236163	.106680	1.00565	
	LS Contaminated Data	.133592	.045695	2.48139	.007120
	GM Contaminated Data	.156848	.082112	1.15611	.003447
	GM Clean Data	.231174	.103447	1.02514	.000018
2	LS Clean Data	.240428	.114784	1.03011	
	LS Contaminated Data	.088360	.067536	2.55879	.012679
	GM Contaminated Data	.135421	.093801	1.15181	.005733
	GM Clean Data	.239569	.114071	1.04459	.000001
3	LS Clean Data	.242595	.107649	1.01869	
	LS Contaminated Data	.114572	.039206	2.54770	.010537
	GM Contaminated Data	.160738	.077252	1.16230	.003812
	GM Clean Data	.242776	.115418	1.04057	.000030
4	LS Clean Data	.223647	.073663	0.99701	
	LS Contaminated Data	.056664	.016945	2.51918	.015550
	GM Contaminated Data	.128238	.072108	1.11820	.004553
	GM Clean Data	.223233	.078112	1.02743	.000010
5	LS Clean Data	.282265	.093625	0.93248	
	LS Contaminated Data	.103913	.038853	2.55393	.017405
	GM Contaminated Data	.169156	.086559	1.08605	.006422
	GM Clean Data	.278590	.098113	0.93577	.000017
6	LS Clean Data	.248693	.074178	1.00454	
	LS Contaminated Data	.078250	.044128	2.56529	.014977
	GM Contaminated Data	.150117	.062442	1.14734	.004927
	GM Clean Data	.249784	.074666	1.01152	.000001
7	LS Clean Data	.272333	.063656	0.97107	
	LS Contaminated Data	.121358	.013489	2.50059	.012655
	GM Contaminated Data	.176811	.071100	1.11739	.004590
	GM Clean Data	.270652	.066584	1.00111	.000006
8	LS Clean Data	.214169	.125415	1.08745	
	LS Contaminated Data	.102068	.061869	2.45701	.008302
	GM Contaminated Data	.131612	.100466	1.20619	.003719
	GM Clean Data	.215578	.126882	1.07509	.000002
9	LS Clean Data	.232649	.081906	0.99928	
	LS Contaminated Data	.115411	.064520	2.37098	.007024
	GM Contaminated Data	.153175	.083887	1.13848	.003160
	GM Clean Data	.231113	.081552	1.01140	.000001
10	LS Clean Data	.245452	.108255	0.92095	
	LS Contaminated Data	.080360	.040073	2.47818	.015952
	GM Contaminated Data	.141314	.083936	1.05435	.005718
	GM Clean Data	.242536	.112349	0.94666	.000013
11	LS Clean Data	.269605	.101009	0.97095	
	LS Contaminated Data	.125567	.040974	2.50815	.012176
	GM Contaminated Data	.171164	.056797	1.09968	.005823
	GM Clean Data	.268690	.099329	0.97564	.000002
12	LS Clean Data	.190854	.153230	1.04144	
	LS Contaminated Data	.093560	.051904	2.52822	.009867
	GM Contaminated Data	.124334	.110919	1.17381	.003108
	GM Clean Data	.191426	.158870	1.06208	.000016

Table 7: GM-estimator using redescending ψ_B results with changing GMRF field.

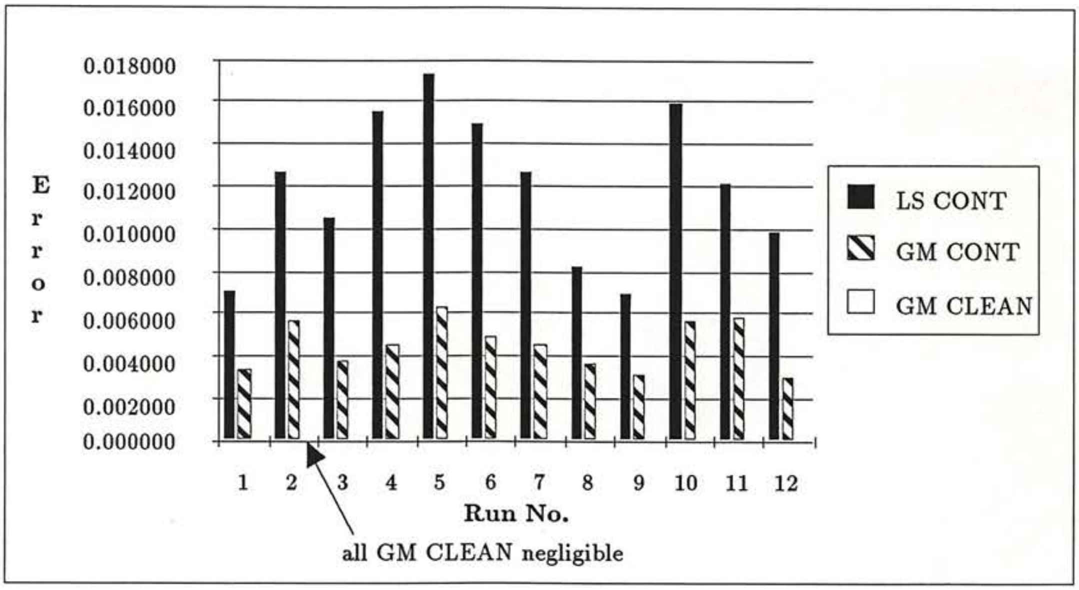

Figure 14: Errors for the results shown in Table 7.

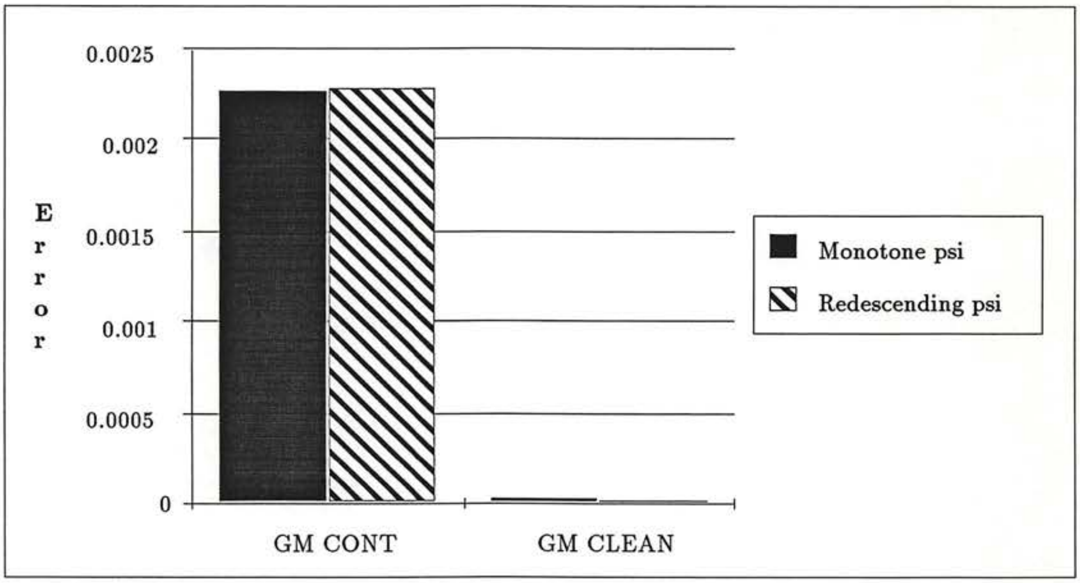

Figure 15: Comparison of the estimation errors for monotone and redescending ψ-functions.

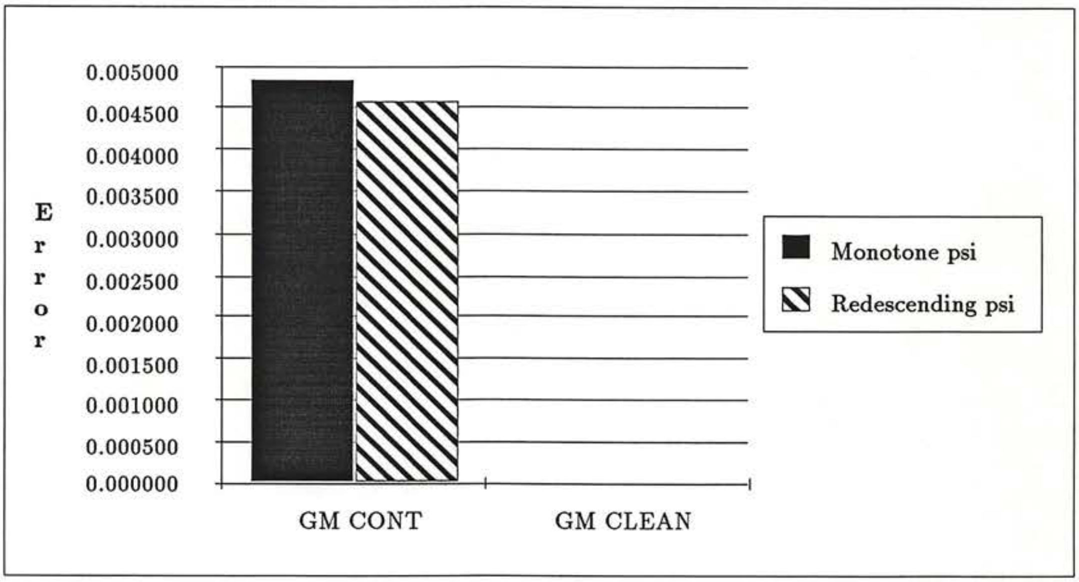

Figure 16: Comparison of the estimation errors for monotone and redescending ψ-functions.

Estimation Method	GMRF/AO			NSHP/IO	GMRF/SO
	Exp. 2.1	Exp. 4.1	Exp. 2.2	Exp. 3.1	Exp. 3.2
	ψ_H	ψ_B	ψ_H	ψ_H	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 8: 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	Exp. 4.1	Exp. 2.2	Exp. 3.1	Exp. 3.2
	ψ_H	ψ_B	ψ_H	ψ_H	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 9: Estimated scale error summary for the GM-estimator. Entries are the averages of the absolute value of the errors from each run.

Run No.	Est Method	$\theta_{(1,0)} = \theta_{(-1,0)}$	$\theta_{(0,1)} = \theta_{(0,-1)}$	β^2	Sq Err
	Theoretical Value	.234000	.101100	1.00000	
1	ML Clean Data	.225644	.121269	1.00170	
	ML Contaminated Data	.154616	.094649	1.91600	.002877
	FC Contaminated Data	.185604	.111985	1.31443	.000845
2	ML Clean Data	.249958	.076669	1.00190	
	ML Contaminated Data	.163430	.058477	2.08880	.003909
	FC Contaminated Data	.209659	.079445	1.25855	.000816
3	ML Clean Data	.218520	.087626	0.98160	
	ML Contaminated Data	.135102	.045927	2.08450	.004349
	FC Contaminated Data	.209613	.071015	1.25949	.000178
4	ML Clean Data	.220391	.110259	1.06010	
	ML Contaminated Data	.139180	.088135	2.08910	.003542
	FC Contaminated Data	.183064	.105823	1.35668	.000706
5	ML Clean Data	.237543	.107537	1.00630	
	ML Contaminated Data	.167581	.086975	2.02010	.002659
	FC Contaminated Data	.201956	.099527	1.31000	.000665
6	ML Clean Data	.239895	.095919	0.94703	
	ML Contaminated Data	.151674	.046862	2.34630	.005095
	FC Contaminated Data	.194207	.082251	1.54170	.001137
7	ML Clean Data	.243742	.103598	0.94607	
	ML Contaminated Data	.150430	.062032	2.17050	.005217
	FC Contaminated Data	.201700	.095267	1.25738	.000918
8	ML Clean Data	.211251	.123445	1.06330	
	ML Contaminated Data	.135504	.085867	2.26500	.003575
	FC Contaminated Data	.176978	.113045	1.53670	.000641
9	ML Clean Data	.222698	.100538	1.01080	
	ML Contaminated Data	.159505	.085977	2.04350	.002103
	FC Contaminated Data	.188891	.096285	1.26087	.000581
10	ML Clean Data	.236326	.112829	0.93604	
	ML Contaminated Data	.153308	.078465	2.04740	.004036
	FC Contaminated Data	.188447	.109484	1.30304	.001152

Table 10: Filter-Cleaner results with changing NCAR field.

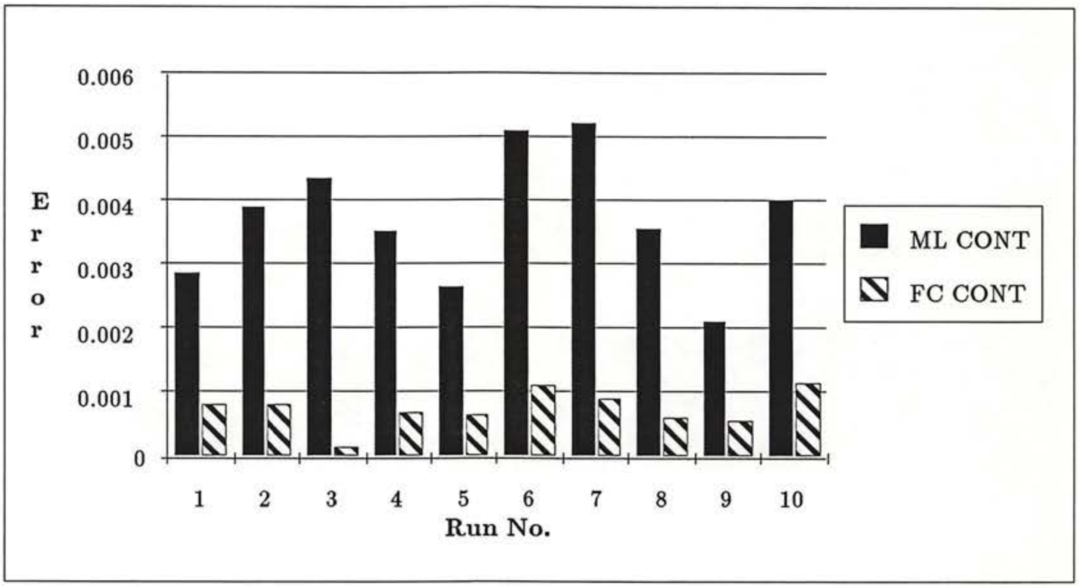

Figure 17: Errors for the results shown in Table 10.

Run No.	Est Method	$\theta_{(1,0)} = \theta_{(-1,0)}$	$\theta_{(0,1)} = \theta_{(0,-1)}$	β^2	Sq Err
	Theoretical Value	.234000	.101100	1.00000	
	ML Clean Data	.236326	.112829	0.93604	
1	ML Contaminated Data	.173117	.062844	1.93850	.003247
	FC Contaminated Data	.211612	.104629	1.23093	.000339
2	ML Contaminated Data	.153344	.069949	2.19160	.004362
	FC Contaminated Data	.202336	.106110	1.33045	.000600
3	ML Contaminated Data	.153827	.067113	2.26490	.0044448
	FC Contaminated Data	.205539	.107336	1.35316	.000489
4	ML Contaminated Data	.166025	.060874	2.17070	.003821
	FC Contaminated Data	.210387	.085819	1.47743	.000701
5	ML Contaminated Data	.153191	.093225	2.01570	.003648
	FC Contaminated Data	.204125	.109471	1.26397	.000524
6	ML Contaminated Data	.183958	.096035	1.73150	.001512
	FC Contaminated Data	.216729	.095992	1.20182	.000334
7	ML Contaminated Data	.133196	.097445	2.09370	.005436
	FC Contaminated Data	.190912	.107058	1.36432	.001048
8	ML Contaminated Data	.122522	.072302	2.34830	.007297
	FC Contaminated Data	.196236	.108129	1.44679	.000815
9	ML Contaminated Data	.135321	.110230	2.09490	.005104
	FC Contaminated Data	.196751	.106330	1.33710	.000804
10	ML Contaminated Data	.178558	.085144	1.77630	.002052
	FC Contaminated Data	.206739	.099702	1.22350	.000524

Table 11: Filter-cleaner results for NCAR field with changing outliers.

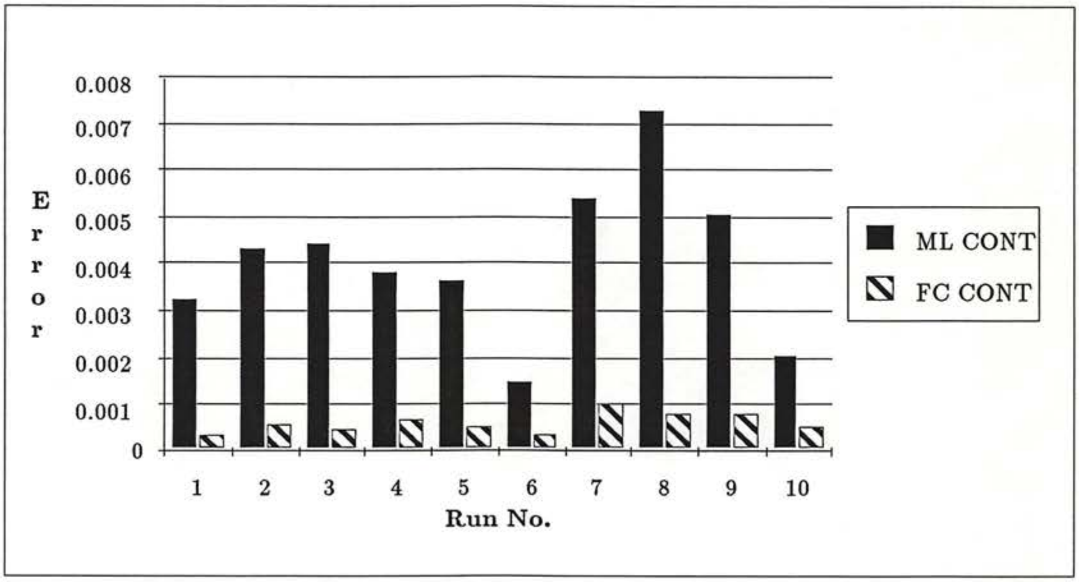

Figure 18: Errors for the results shown in Table 11.

References

[1] R. Chellapa, 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.

[2] 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.

[3] 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.

[4] 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.

[5] P.J. Huber, *"Robust Statistics"*, John Wiley & Sons, New York, 1981.

[6] 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.

[7] 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.

[8] 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. Engg., 1978.

[9] R. Chellappa, *"Stochastic Models for Image Analysis and Processing"*, PhD thesis, Purdue University, W. Lafayette, Indiana, 1981.

[10] R.D. Martin and D.J. Thomson, "Robust-resistant spectrum estimation", *Proc. IEEE*, vol. 70, no. 9, pp. 1097–1114, Sept. 1982.

[11] 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.

[12] 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.

[13] 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.

[14] R. Chellappa, "Two-Dimensional Discrete Gaussian Markov Random Field Models for Image Processing," in *"Progress in Pattern Recognition 2"*, L.N. Kanal and A. Rosenfeld, Eds., Elsevier Science Publishers B.V., North Holland, 1985.

[15] J.W. Woods, "Two-dimensional discrete Markovian random fields", *IEEE Trans. on Info. Theory*, vol. IT-18, no. 2, pp. 232–240, Mar. 1972.

[16] R.D. Martin and V.J. Yohai, "Influence functionals for time series", *Annals of Statistics*, vol. 14, pp. 781–818, 1986.

[17] H. Kunsch, "Infinitesimal robustness for autoregressive processes", *Annals of Statistics*, vol. 12, pp. 843–863, 1984.

[18] R.J. Serfling, *"Approximation Theorems of Mathematical Statistics"*, John Wiley & Sons, New York, 1980.

[19] D.G. Luenberger, *"Linear and Nonlinear Programming"*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1984.

[20] R. von Mises, "On the asymptotic distribution of differentiable statistical functions", *Annals of Mathematical Statistics*, vol. 18, pp. 309–348, 1947.

[21] P. Billingsley, *"Probability and Measure"*, John Wiley & Sons, New York, 1986.