DEPARTMENT OF STATISTICS
University of Wisconsin
1210 West Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 902

March 12, 1993

On Combining Data from Multiple Sources with Unknown Relative Weights (Thesis)

by

Feng Gao

ON COMBINING DATA FROM MULTIPLE SOURCES WITH UNKNOWN RELATIVE WEIGHTS

By

Feng Gao

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy
(Statistics)

at the
UNIVERSITY OF WISCONSIN – MADISON
1993

Acknowledgments

I would like to express my sincerest gratitude to my advisor, Professor Grace Wahba, for her invaluable advice and continuing encouragement during my years as a graduate student.

Appreciation is extended to Professors Donald Johnson, Brian Yandell, Douglas Bates and Wei-Yin Loh for their service on my final examination comittee, especially to Professors Donald Johnson and Brian Yandell for their careful reading of this thesis and their valuable comments.

I would like to thank Dr. Fred Reames for his assistance in the Monte Carlo experiments.

Finally, special thanks go to my parents, my sister and my wife Xiuling Guo for their love and support.

This research was partially supported by NSF under Grant DMS9121003 and NASA under Contract NAGW 2961. Some of the simulations for this study were performed at NASA Centers For Computational Sciences and at the San Diego Supercomputer Center.

Abstract

[1] The problem of using the "direct" variational methods in a statistical model that merges data from different sources with unknown relative weights is considered. The model is initially proposed for data assimilation in numerical weather prediction, but could also be useful in other scientific fields where data from different sources need to be combined. To carry out this merging optimally, it is necessary to provide an estimate of the relative weights to be given to data from different sources. In Wahba, Johnson and Reames (1990), two methods are suggested to estimate the weighting parameters and the smoothing parameters simultaneously. These two methods are the generalized maximum likelihood estimate and the risk cross validation estimate. We denote these two estimates by GML-r and RCV, where r represents the weighting parameter. In this dissertation, a new form of generalized cross validation estimate is developed to simultaneously estimate the weighting parameters and the smoothing parameter. We name this new estimate GCV-r. We prove the weak consistency and the asymptotic normality of all these estimators. The convergence rates for these estimators are obtained under some conditions. We also conduct some Monte Carlo experiments that simulate meteorological realities. It is shown, both theoretically and by simulation, that the GCV-r is substantially better than the RCV estimate in estimating the weighting parameter r. Also from the simulation results, the GCV-r appears to be more robust than the GML-r when there is model misspecification, which is often inevitable in practice.

Key words and phrases: smoothing spline, weighting parameter, GCV, GML.

[1]This report incorporates all of the results of Technical Report 894. Chapter 5 is a joint work with Professor Grace Wahba, Dept. of Statistics, and Professor Donald Johnson, Dept. of Atmospheric and Oceanic Sciences.

Contents

Chapter 1

Introduction

1.1 Introduction

The problem of simultaneously combining data from many different sources with unknown relative weights in a variational analysis is considered. This is a very important problem in data assimilation in numerical weather prediction, where information from forecast, direct data, indirect data (such as satellite radiances), and prior meteorological information of various sorts, e.g. the frequency spectrum of a parameter decays at a certain rate, etc. are available and need to be combined to get a good initialization for numerical weather prediction. See Wahba (1982a), Wahba (1985b), Wahba et al. (1990), Lorenc (1986), Parrish and Derber (1992), and others. In this dissertation, we consider a model that combines data from two different sources. The methods and results derived here can be easily extended to models that combines data from more than two sources. This technique can also be very useful in other scientific fields where data from different sources need to be combined.

The model we will be using in this thesis first appears in Wahba et al. (1990) and Bates, Reames and Wahba (1990). It is of the following form:

$$\begin{cases} y_{1i} = L_{1i}f + \varepsilon_{1i} \\ y_{2j} = L_{2j}f + \varepsilon_{2j} \end{cases} \tag{1.1.1}$$

where $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$, f is a smooth function, which for example, could represent the true but unknown state in meteorology, for instance, the 500 millibar heights of the atmosphere. We may assume this f to be a function in a reproducing kernel Hilbert

space $\mathcal{H}$. L_{1i} and L_{2j}, $i = 1, \ldots, n_1$, $j = 1, \ldots, n_2$, are bounded linear functionals on $\mathcal{H}$. $\varepsilon_1 = (\varepsilon_{11}, \ldots, \varepsilon_{1n_1})' \sim \mathcal{N}(0, \sigma_1^2 I)$, $\varepsilon_2 = (\varepsilon_{21}, \ldots, \varepsilon_{2n_2})' \sim \mathcal{N}(0, \sigma_2^2 I)$, where σ_1^2 and σ_2^2 are unknown, ε_1 and ε_2 are independent.

The ultimate goal is to estimate f. If we knew σ_1^2 and σ_2^2, we would do this by solving the following variational problem:

$$\min_{f \in \mathcal{H}} \frac{1}{n_1 + n_2} \left(\frac{1}{\sigma_1^2} \sum_{i=1}^{n_1} (y_{1i} - L_{1i} f)^2 + \frac{1}{\sigma_2^2} \sum_{j=1}^{n_2} (y_{2j} - L_{2j} f)^2 \right) + \lambda J(f) \tag{1.1.2}$$

where $J(f)$ is a roughness penalty and λ is the smoothing parameter that controls the trade-off between the weighted residual sum of squares $(1/(n_1 + n_2))((1/\sigma_1^2) \sum_{i=1}^{n_1} (y_{1i} - L_{1i} f)^2 + (1/\sigma_2^2) \sum_{j=1}^{n_2} (y_{2j} - L_{2j} f)^2)$ and the penalty $J(f)$. The solution $\hat{f}$ is called a spline in a general sense (see Wahba (1990)).

In meteorology, f is often a function on the sphere, and can be approximated by a finite Fourier expansion, for example the spherical harmonic expansion

$$f \approx f_{00} + \sum_{l=1}^{M} \sum_{s=-l}^{l} f_{ls} Y_{ls},$$

which is the analog of a trigonometric expansion of a periodic function, i.e. function on the circle. Here Y_{ls} are the spherical harmonics and f_{ls} are the Fourier-Bessel coefficients. For simplicity, in this thesis in Chapters 3 and 4, we only consider the cases where f is a function on the circle, and we assume n is even, i.e. f has the expansion

$$f \approx a_0 + \sum_{\nu=1}^{n/2-1} a_\nu \sqrt{2} \cos 2\pi\nu t + \sum_{\nu=1}^{n/2-1} b_\nu \sqrt{2} \sin 2\pi\nu t + a_{n/2} \cos \pi n t.$$

The result for function on the circle can be easily extended to function on the sphere.

After approximating f by a finite Fourier expansion, the model (1.1.1) can be written as

$$\begin{cases} y_1 = X_1 f + \varepsilon_1 \\ y_2 = X_2 f + \varepsilon_2 \end{cases} \tag{1.1.3}$$

where y_i is of dimension n_i, $i = 1, 2$, $f = (f_0, f_1, \ldots, f_{n-1})'$ is of dimension n and contains the Fourier-Bessel coefficients of function f, X_i are $n_i \times n$ matrices, $i = 1, 2$ and $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2 I)$, $i = 1, 2$ and they are independent.

If we let $\int_0^1 (f^{(m)}(u))^2 du$ be the penalty, then the variational problem (1.1.2) becomes one of finding the minimizer f of

$$\frac{1}{n_1+n_2}\left(\frac{1}{\sigma_1^2}\|y_1 - X_1 f\|^2 + \frac{1}{\sigma_2^2}\|y_2 - X_2 f\|^2\right) + \lambda f'\Sigma^{-1} f$$

or, equivalently

$$\frac{1}{\theta(n_1+n_2)}\left\{\frac{1}{r}\|y_1 - X_1 f\|^2 + r\|y_2 - X_2 f\|^2 + \alpha f'\Sigma^{-1} f\right\} \tag{1.1.4}$$

where $\theta = \sigma_1\sigma_2$ is a nuisance parameter, $r = \sigma_1/\sigma_2$ is the weighting parameter, α is the smoothing parameter and Σ is such a diagonal matrix that for cases where f is a function on the circle, $\Sigma = \text{diag}[\xi_\nu], \nu = 1, \ldots, n$, $\xi_1 \to \infty$, $\xi_\nu = [\nu/2]^{-2m}, \nu = 2, \ldots, n$, where $[x]$ is the integer part of x; for cases where f is a function on the sphere, $\Sigma = \text{diag}[\xi_{ls}], s = -l, \ldots, l; l = 0, \ldots, M$, $\xi_{00} \to \infty$, $\xi_{ls} = [l(l+1)]^{-m}, s = -l, \ldots, l; l = 1, \ldots, M$.

This is so because the eigenvalues of the reproducing kernel of the Hilbert space of periodic smooth functions are all of multiplicity 2 and are $\lambda_\nu = (2\pi\nu)^{-2m}$. Similar result is also true for functions on the sphere. See page 21-27 of Wahba (1990). $m \geq 1$ is the number of continuous derivatives f is assumed to possess. m is assumed to be known.

If we know r and α, we can get the estimate:

$$\hat{f}_{r,\alpha} = (\frac{1}{r}X_1'X_1 + rX_2'X_2 + \alpha\Sigma^{-1})^{-1}(\frac{1}{r}X_1'y_1 + rX_2'y_2) \tag{1.1.5}$$

So, finding a good estimate of r and α becomes very important.

Finding a good estimate of r alone is also of interest, because it will give us some knowledge on the relative accuracies of different instruments from which we get the data.

In order to derive the generalized maximum likelihood estimate, we need to look at the problem from a Bayesian point of view. The stochastic model behind the maximum likelihood estimate is

$$f \sim \mathcal{N}(0, b\Sigma), \tag{1.1.6}$$

where $b > 0$.

In meteorology, the data from real life usually behave as if they were from such a stochastic model with some $m \geq 1$. See Figure 5.1.3 in Chapter 5. Also see Wahba (1982b) and Stanford (1979) for more examples. So, it is appropriate to assume such a stochastic

model as prior information on f. The log of posterior distribution is proportional to

$$-\left\{\frac{1}{\sigma_1^2}\|y_1 - X_1 f\|^2 + \frac{1}{\sigma_2^2}\|y_2 - X_2 f\|^2 + \frac{1}{b} f'\Sigma^{-1} f\right\},$$

and finding the maximizer of it is equivalent to finding the minimizer of

$$\frac{1}{\theta}\left\{\frac{1}{r}\|y_1 - X_1 f\|^2 + r\|y_2 - X_2 f\|^2 + \alpha f'\Sigma^{-1} f\right\} \tag{1.1.7}$$

where θ and r are as before and the smoothing parameter $\alpha = \sigma_1\sigma_2/b$ is the noise to signal ratio. Since minimizing (1.1.7) is the same as minimizing (1.1.4), we get the same form of the estimate $\hat{f}_{r,\alpha}$ as in (1.1.5). For more discussion on the relationship between spline smoothing and Bayesian estimates, see Wahba (1990), Kimeldorf and Wahba (1970) and Wahba (1978).

Our interest is in estimating r, α and finally f. We will describe several estimation methods and study their properties in the later chapters. In order to study the properties of different estimators, we need to specify certain conditions on the "design" matrices X_1 and X_2 in (1.1.3). We will do this in the next section.

1.2 Several Interesting Cases

In this thesis, we will consider several models that describe some of the most interesting situations arising in meteorology. Our y_1 and y_2 are data from different sources. These sources could include i).direct observations; ii).satellite soundings; and iii).forecast data from a forecast model. Now let us look into these three different sources in a bit more detail:

i). If one data set, say y_1, is from the direct observations of the true state at certain points in the spatial domain, for example, the 500 millibar heights observed from the radiosonde stations all over the world, then L_{1i}'s in (1.1.1) are the evaluation functionals. If we assume the observational points are uniformly distributed on the sphere or on the circle, then after approximating f by a finite Fourier expansion with $n = n_1$, the columns of X_1 would be perpendicular to each other, and we have $X_1'X_1 = nI$. We will prove our main results under this assumption. For more general cases, we believe that similar results are still true under some regularity conditions.

ii). If one data set, say y_2, is from remote soundings at a satellite sensor, then L_{2j}'s

in (1.1.1) can be approximated by an integral equation, which, in an idealized situation, can be expressed in a matrix form as in (1.1.3), where, if we assume $n = n_2$, X_2 can be written as $X_2 = \sqrt{n}\Gamma K$, where Γ is an orthogonal matrix and K is a diagonal matrix of eigenvalues of the kernel function in the integral equation. Since the kernel function is squared integrable, k_ν, the $\nu\nu$th entry of K, should satisfy $\sum_{\nu=1}^{\infty} k_\nu^2 < \infty$. For simplicity, we may assume that for cases where f is a function on the circle, $K = \text{diag}[k_\nu], \nu = 1, \ldots, n$, where $k_\nu = [\nu/2]^{-p}, \nu = 2, \ldots, n$, for some $p > 1/2$; for cases where f is a function on the sphere, $K = \text{diag}[k_{ls}], s = -l, \ldots, l; l = 0, \ldots, M$, where $k_{ls} = [l(l+1)]^{-p}, s = -l, \ldots, l; l = 1, \ldots, M$, for some $p > 1/2$.

iii). If one data set, say y_2, is from a forecast model, then in our Monte Carlo studies, we have the data in the frequency domain, i.e. we have the vector of Fourier coefficients as data. In practice, the forecast errors in the spatial domain are usually correlated. See Hollingsworth and Lonnberg (1986), Lonnberg and Hollingsworth (1986) and Wahba (1989) for examples. In Wahba (1989), a mathematical model is used to describe the forecast error correlation. According to this model, the forecast error can be viewed as a stochastic process. This stochastic process should be "rougher" than the signal if we also view the signal f as a sample function from some other stochastic process. By "rougher" we mean the eigenvalues for the error, which correspond to the energy spectrums in a Karhunen-Loeve expansion, decay at a slower rate than that of signal f. For example, we may assume $n = n_2$ and for cases where f is a function on the circle, we may assume the forecast error has eigenvalues $K^{-2} = \text{diag}[k_\nu], \nu = 1, \ldots, n$, where $k_\nu = [\nu/2]^{-2p}, \nu = 2, \ldots, n$, for some $0 < p < m$; for cases where f is a function on the sphere, we may assume the forecast error has eigenvalues $K = \text{diag}[k_{ls}], s = -l, \ldots, l; l = 0, \ldots, M$, where $k_{ls} = [l(l+1)]^{-p}, s = -l, \ldots, l; l = 1, \ldots, M$, for some $0 < p < m$. Thus, we have $\tilde{y} = f + \tilde{\varepsilon}$, where the vector f contains Fourier coefficients and $\tilde{\varepsilon} \sim \mathcal{N}(0, \sigma^2 K^{-2})$. Multiplying both side by K, we get $y \equiv K\tilde{y} = Kf + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. The model with this kind of correlated error is of practical interest, but has not been studied in this context before.

In all the 3 cases above, we have $y = Xf + \varepsilon$ with X of the form ΓK, where Γ is an orthogonal matrix, K is a diagonal matrix, and ε is *i.i.d.*. Multiplying Γ' on both side, we get the model:

$$\begin{cases} y_1 = K_1 f + \varepsilon_1 \\ y_2 = K_2 f + \varepsilon_2 \end{cases} \tag{1.2.1}$$

where K_1 and K_2 are diagonal matrices.

In Chapters 3 and 4, we will only consider cases where f is a function on the circle. We will assume $n = n_1 = n_2$ and (1.2.1) to be our model. We will consider the following 3 cases of practical interests:

case 1: $K_i = \sqrt{n}\text{diag}[k_{i\nu}], \nu = 1, \ldots, n; k_{i\nu} = [\nu/2]^{-p_i}, \nu = 2, \ldots, n$, where $p_i \geq 0$, $i = 1, 2$. This approximates cases where we have direct observations and/or satellite soundings.

case 2: $K_1 = \sqrt{n}\text{diag}[k_{1\nu}], \nu = 1, \ldots, n; k_{1\nu} = [\nu/2]^{-p_1}, \nu = 2, \ldots, n$, where $p_1 \geq 0$. $K_2 = \text{diag}[k_{2\nu}], \nu = 1, \ldots, n; k_{2\nu} = [\nu/2]^{p_2}, \nu = 2, \ldots, n$, where $0 \leq p_2 < m$. In this case, data set one could be direct observations or satellite soundings, data set two is from a forecast model.

case 3: $K_i = \text{diag}[k_{i\nu}], \nu = 1, \ldots, n; k_{i\nu} = [\nu/2]^{p_i}, \nu = 2, \ldots, n$, where $0 \leq p_i < m$, $i = 1, 2$. In this case, the two data sets may be from two different forecast models.

1.3 Models with More Than Two Sources

We can extend our model to models that combine data from more than two sources. In this section, we will see how to do this.

We can consider the following model that combines data from J sources:

$$y_j = K_j f + \varepsilon_j, \ j = 1, \ldots, J \tag{1.3.1}$$

where $\varepsilon_j \sim \mathcal{N}(0, \sigma_j^2), j = 1, \ldots, J$.

Then we can estimate f by finding the minimizer f of

$$\sum_{j=1}^{J} \frac{1}{\sigma_j^2} \|y_j - K_j f\|^2 + \frac{1}{b} f' \Sigma^{-1} f \tag{1.3.2}$$

or equivalently

$$\sum_{j=1}^{J} \frac{1}{r_j} \|y_j - K_j f\|^2 + \alpha f' \Sigma^{-1} f \tag{1.3.3}$$

where $r_j = \sigma_j^2 / (\prod_{l=1}^{J} \sigma_l^2)^{1/J}$ are the weighting parameters and $\alpha = (\prod_{l=1}^{J} \sigma_l^2)^{1/J} / b$ is the smoothing parameter. Note that $\prod_{j=1}^{J} r_j = 1$. For more on models that combine data from more than two sources, please see Wahba et al. (1990).

1.4 Outline of Thesis

We now give an outline of the remainder of this thesis. In Chapter 2, we first give the formula of the generalized maximum likelihood estimate and the risk cross validation estimate for estimating the weighting and smoothing parameters simultaneously. Both of these two estimates appear in Wahba et al. (1990). We then introduce a new form of generalized cross validation estimate for estimating the weighting and smoothing parameters simultaneously. We also give formula for calculating the confidence intervals in our model. In Chapter 3, we study the properties of all the estimators described in Chapter 2. In Chapter 4, we report some results from simulation studies for functions on the circle. In Chapter 5, we present Monte Carlo experiment results for functions on the sphere. We conclude the thesis in Chapter 6.

Chapter 2

Estimation of Weighting and Smoothing Parameters

As we have seen in Chapter 1 that in order to estimate the underlying unknown function f, it is necessary to provide a good estimate of the weighting parameter r and the smoothing parameter α. In this chapter, we will first introduce two methods for estimating r and α simultaneously. These are the generalized maximum likelihood estimate and the risk cross validation estimate, both of which were proposed in Wahba et al. (1990). Then we will develop a new generalized cross validation method for simultaneously estimating r and α. Later on in Chapter 3, we will show that in theory, this new generalized cross validation is better than the risk cross validation, and in Chapters 4 and 5, we will demonstrate by simulation studies that this new generalized cross validation is significantly better than the risk cross validation in estimating r.

2.1 GML-r Estimate

In this section, we derive the generalized maximum likelihood estimate under the stochastic assumption (1.1.6).

We can write model (1.2.1) as $y = Kf + \varepsilon$, where

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \; K = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}, \; \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

After rescaling it, we can write the model as $y^r = K^r f + \varepsilon^r$, where $y^r = I^{-1}(r)y$, $K^r = I^{-1}(r)K$, $\varepsilon^r = I^{-1}(r)\varepsilon$ and

$$I(r) = \begin{pmatrix} \sqrt{r}I_{n_1} & 0 \\ 0 & (1/\sqrt{r})I_{n_2} \end{pmatrix}.$$

In this rescaled model, $\varepsilon^r \sim \mathcal{N}(0, \theta I)$ and the minimization problem (1.1.7) is equivalent to

$$\min_f \{\|y^r - K^r f\|^2 + \alpha f' \Sigma^{-1} f\} \tag{2.1.1}$$

The solution to (2.1.1) is

$$\hat{f}_{r,\alpha} = (K^{r'} K^r + \alpha\Sigma^{-1})^{-1} K^r y^r$$

or

$$\hat{f}_{r,\alpha} = (\frac{1}{r}K_1'K_1 + rK_2'K_2 + \alpha\Sigma^{-1})^{-1}(\frac{1}{r}K_1'y_1 + rK_2'y_2).$$

Let

$$M = (\frac{1}{r}K_1'K_1 + rK_2'K_2 + \alpha\Sigma^{-1}),$$

and

$$A^r(r,\alpha) = \begin{pmatrix} (1/r)K_1M^{-1}K_1' & K_1M^{-1}K_2' \\ K_2M^{-1}K_1' & rK_2M^{-1}K_2' \end{pmatrix}.$$

This $A^r(r,\alpha)$ is the rescaled influence matrix, i.e. we have $\hat{y}^r = A^r(r,\alpha)y^r$.

Under our model,

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}(0, \theta W(r,\alpha)) \tag{2.1.2}$$

where

$$W(r,\alpha) = \frac{1}{\alpha}\begin{pmatrix} K_1\Sigma K_1' & K_1\Sigma K_2' \\ K_2\Sigma K_1' & K_2\Sigma K_2' \end{pmatrix} + \begin{pmatrix} rI_{n_1} & 0 \\ 0 & (1/r)I_{n_2} \end{pmatrix}.$$

It can be shown that

$$W(r,\alpha) = (I^{-1}(r)[I - A^r(r,\alpha)]I^{-1}(r))^{-1}. \tag{2.1.3}$$

We find the Generalized Maximum Likelihood (GML) estimate of r and α by letting

$\xi_1 \to \infty$ or $\xi_{00} \to \infty$ in Σ depending on whether f is a function on the circle or on the sphere.

Let

$$\begin{pmatrix} z \\ \cdots \\ w \end{pmatrix} = \begin{pmatrix} Q_2' \\ \cdots \\ (1/\sqrt{\xi_1})T' \end{pmatrix} y,$$

where $T = (k_{10}, 0, \ldots, 0, k_{20}, 0, \ldots, 0)'$ is the first column of matrix $(K_1', K_2')'$ and Q_2' is any $(n_1 + n_2 - 1) \times (n_1 + n_2)$ matrix satisfying $Q_2'Q_2 = I_{n_1+n_2-1}$ and $Q_2'T = 0$. For example,

$$Q_2' = \begin{pmatrix} 0 & I_{n_1-1} & 0 & 0 \\ k_{20}/\sqrt{k_{10}^2 + k_{20}^2} & 0 & -k_{10}/\sqrt{k_{10}^2 + k_{20}^2} & 0 \\ 0 & 0 & 0 & I_{n_2-1} \end{pmatrix}.$$

Then z is a $n_1 + n_2 - 1$ dimensional random vector and

$$z \sim \mathcal{N}(0, \theta v(r, \alpha)) \tag{2.1.4}$$

where $v(r, \alpha) = Q_2'W(r, \alpha)Q_2$. It is easy to see that

$$\lim_{\xi_1 \to \infty} Ezw' = 0,$$

$$\lim_{\xi_1 \to \infty} Eww' = b(T'T)(T'T).$$

Following the argument for one data source case as in Wahba (1985a), we claim that since the limit distribution of w does not depend on r and α, it is appropriate to define the GML estimate of r and α as the usual ML estimate based on the distribution of z in (2.1.4). For more details, also see Wahba (1990). A straightforward calculation gives the GML estimate of r and α as the minimizer of

$$M(r, \alpha) = \frac{(1/(n_1 + n_2))z'v^{-1}(r, \alpha)z}{(\det v^{-1}(r, \alpha))^{1/(n_1+n_2-1)}} = \frac{(1/(n_1 + n_2))y'Q_2v^{-1}(r, \alpha)Q_2'y}{(\det v^{-1}(r, \alpha))^{1/(n_1+n_2-1)}}$$

and the GML estimate of θ is $\hat{\theta}_{gml} = (1/(n_1 + n_2 - 1))z'v^{-1}(\hat{r}_{gml}, \hat{\alpha}_{gml})z$.

Since the null space has dimension 1 in this case, it can be shown by straightforward

calculation that

$$Q_2 v^{-1}(r,\alpha) Q_2' = I^{-1}(r)[I - A^r(r,\alpha)]I^{-1}(r)$$

$$\det v^{-1}(r,\alpha) = \det{}^+(I^{-1}(r)[I - A^r(r,\alpha)]I^{-1}(r))$$

where $\det^+$ means the product of the nonzero eigenvalues. Thus the GML estimate of r and α is the minimizer of

$$M(r,\alpha) = \frac{(1/(n_1+n_2))y' I^{-1}(r)[I - A^r(r,\alpha)]I^{-1}(r)y}{(\det{}^+(I^{-1}(r)[I - A^r(r,\alpha)]I^{-1}(r)))^{1/(n_1+n_2-1)}} \tag{2.1.5}$$

We call this estimate GML-r estimate, where r represents the weighting parameter.

Generally, this function has a bowl shape near the true (r,α). Figure 2.1.1 is an example of the contour plot of the generalized likelihood function. The data are from the following model for function on the circle:

$$\begin{cases} y_1 = K_1 f + \varepsilon_1 \\ y_2 = K_2 f + \varepsilon_2 \end{cases}$$

where $f \sim \mathcal{N}(0, b\Sigma)$, $b = 1$, Σ is the same as the one appears in section 1.1 with $m = 2$ for functions on the circle, and $n = n_1 = n_2 = 400$. K_1 and K_2 are of the form described in case 1 of section 1.2, with $p_1 = 0$ and $p_2 = 0.6$. So y_1 could be direct observations and y_2 could be satellite soundings. $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2 I)$, $i = 1, 2$, and $\sigma_1 = 0.5$, $\sigma_2 = 0.4$. Thus the true r is $r_0 = \sigma_1/\sigma_2 = 1.25$ and the true α is $\alpha_0 = \sigma_1\sigma_2/b = 0.2$.

2.2 RCV Estimate

To derive a generalized cross validation (GCV) estimate in our problem with two data sources is not as easy as it looks. A "direct" generalization of the usuall GCV estimate goes as follows:

Using the same notation as in the previous section, we can write our model and the rescaled model as $y = Kf + \varepsilon$ and $y^r = K^r f + \varepsilon^r$.

Now, let $f_{r,\alpha}^{[k]}$ denote the spline estimate of f using all but the kth data point of y. We denote the kth data point by y_k, and the kth row of K and K^r by K_k and K_k^r, respectively. Then a direct generalization of the ordinary cross validation (OCV) from one data source

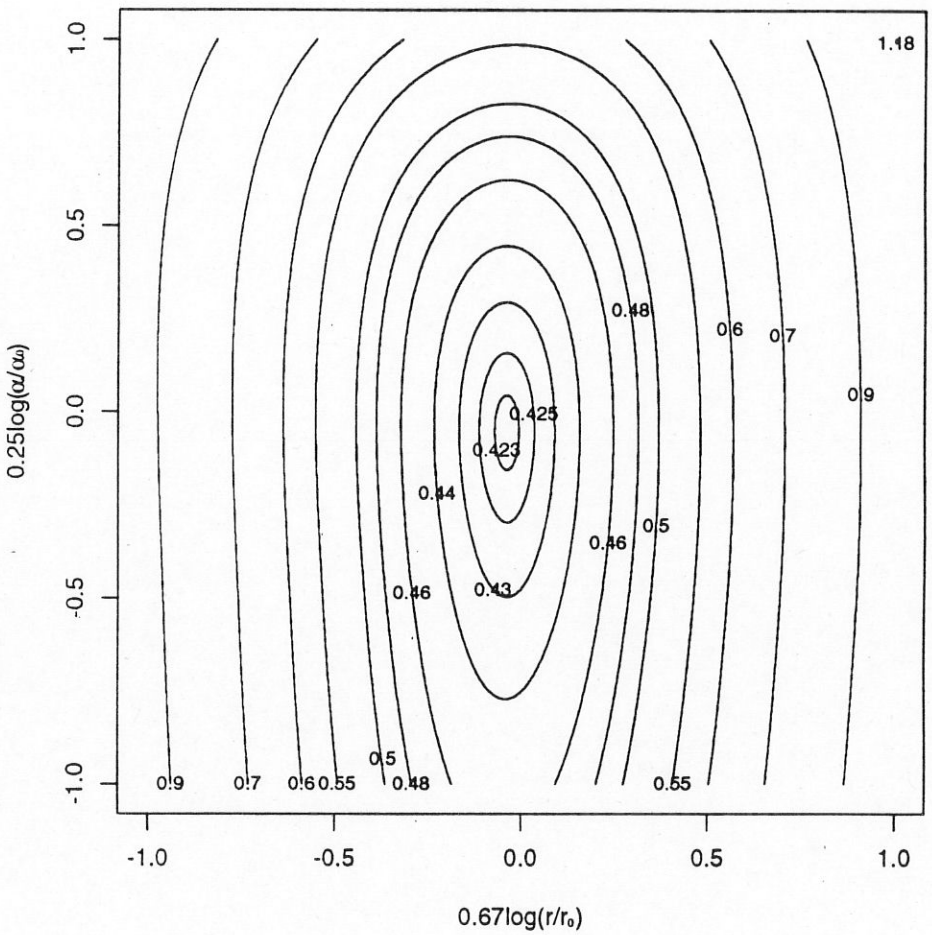


Figure 2.1.1: The GML-r function $M(r, \alpha)$

model would be

$$OCV_1(r,\alpha) = \frac{1}{n_1+n_2}\sum_{k=1}^{n_1+n_2}(y_k^r - K_k^r f_{r,\alpha}^{[k]})^2. \tag{2.2.1}$$

This OCV_1 function measures the overall predictability of the spline estimate $\hat{f}_{r,\alpha}$.

It is then not hard to see that the *leaving-out-one* lemma (Lemma 3.1 of Craven and Wahba (1979), also see Wahba (1990)) is true here, i.e. we have:

Lemma 2.2.1 If we let $h_{r,\alpha}[k,z]$ be the solution to (2.1.1) with the kth data point y_k^r being replaced by z, then $h_{r,\alpha}[k, K_k^r f_{r,\alpha}^{[k]}] = f_{r,\alpha}^{[k]}$. □

Then following the same argument as in Lemma 3.2 of Craven and Wahba (1979), it can be shown that

$$y_k^r - K_k^r f_{r,\alpha}^{[k]} = \frac{y_k^r - K_k^r \hat{f}_{r,\alpha}}{1 - a_{kk}^r(r,\alpha)}$$

where $a_{kk}^r(r,\alpha)$ is the kkth entry of the rescaled influence matrix $A^r(r,\alpha)$. Thus, we have the following OCV *identity*.

Theorem 2.2.1

$$\frac{1}{n_1+n_2}\sum_{k=1}^{n_1+n_2}(y_k^r - K_k^r f_{r,\alpha}^{[k]})^2 \equiv OCV_1(r,\alpha) = \frac{1}{n_1+n_2}\sum_{k=1}^{n_1+n_2}\frac{(y_k^r - K_k^r \hat{f}_{r,\alpha})^2}{(1-a_{kk}^r(r,\alpha))^2}.$$

□

Then we get the generalized cross validation by replacing $a_{kk}^r(r,\alpha)$ by $(1/(n_1+n_2))\sum_{i=1}^{n_1+n_2} a_{ii}^r(r,\alpha) = (1/(n_1+n_2))trA^r(r,\alpha)$ to achieve certain invariance properties that OCV does not have (see Wahba (1990) for more details), and we have

$$GCV(r,\alpha) = \frac{(1/(n_1+n_2))\|(I-A^r(r,\alpha))y^r\|^2}{[(1/(n_1+n_2))tr(I-A^r(r,\alpha))]^2}.$$

Unfortunately, this function is not usable because it will decrease when r tends to zero or infinity. So when searching for minimizer of this function, we often end up at some point far away from the true r.

The above claim can be verified most easily by assuming that $n = n_1 = n_2$, and K_i has the forms in one of the three cases in Section 1.2, so that K_i, M and Σ are all diagonal matrices, therefore they are interchangeable for matrix multiplication. Then it is easy to see that

$$\lim_{r\to 0} A^r(r,\alpha) = \begin{pmatrix} I_n & 0 \\ 0 & 0 \end{pmatrix},$$

$$\lim_{r\to\infty} A^r(r,\alpha) = \begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix}.$$

So

$$\lim_{r\to 0} tr(I_{2n} - A^r(r,\alpha)) = tr\begin{pmatrix} 0 & 0 \\ 0 & I_n \end{pmatrix} = n,$$

$$\lim_{r\to\infty} tr(I_{2n} - A^r(r,\alpha)) = tr\begin{pmatrix} I_n & 0 \\ 0 & 0 \end{pmatrix} = n.$$

On the other hand, it can be shown that

$$\lim_{r\to 0}(I_{2n} - A^r(r,\alpha))y^r = 0, \tag{2.2.2}$$

$$\lim_{r\to\infty}(I_{2n} - A^r(r,\alpha))y^r = 0. \tag{2.2.3}$$

So the above claim holds. To prove (2.2.2), notice

$$\begin{aligned}
&(I_{2n} - A^r(r,\alpha))y^r \\
&= \begin{pmatrix} (I_n - (1/r)K_1^2M^{-1})(1/\sqrt{r})y_1 - K_1K_2M^{-1}\sqrt{r}y_2 \\ -K_1K_2M^{-1}(1/\sqrt{r})y_1 + (I_n - rK_2^2M^{-1})\sqrt{r}y_2 \end{pmatrix} \\
&= \begin{pmatrix} (rK_2^2 + \alpha\Sigma^{-1})M^{-1}(1/\sqrt{r})y_1 - K_1K_2M^{-1}\sqrt{r}y_2 \\ -K_1K_2M^{-1}(1/\sqrt{r})y_1 + ((1/r)K_1^2 + \alpha\Sigma^{-1})M^{-1}\sqrt{r}y_2 \end{pmatrix},
\end{aligned}$$

and

$$\lim_{r\to 0} M^{-1}\frac{1}{\sqrt{r}} = 0,$$

$$\lim_{r\to 0}(\frac{1}{r}K_1^2 + \alpha\Sigma^{-1})M^{-1} = I_n.$$

So (2.2.2) holds. Similarly, we can prove (2.2.3).

To overcome this difficulty, a risk cross validation (RCV) estimate of r and α is proposed in Wahba et al. (1990). It is the minimizer of $RCV(r,\alpha)$ given by

$$RCV(r,\alpha) = \frac{1}{n_1+n_2}\left\{\frac{\|y_1 - \hat{y}_1\|^2}{((1/n_1)tr[I - A_{11}^r(r,\alpha)])^2} + \frac{\|y_2 - \hat{y}_2\|^2}{((1/n_2)tr[I - A_{22}^r(r,\alpha)])^2}\right\}. \tag{2.2.4}$$

where A_{11}^r and A_{22}^r are the diagonal blocks of $A^r(r,\alpha)$.

RCV function has a bowl shape near the optimal point. Figure 2.2.1 is an example of this function using the same data from the example described in the previous section.

2.3 GCV-r Estimate

The RCV function $RCV(r,\alpha)$ is somewhat asymmetric, and is therefore somewhat unnatural. If we let $\hat{r}_{rcv}$ denote the RCV estimate of r and so forth, we will show later that the convergence rate for $\hat{r}_{rcv}$ is very slow compared with that for $\hat{r}_{gml}$. See Chapters 3, 4 and 5 for more details. These drawbacks impel us to seek a better formula for cross validation estimate.

To find a better GCV function, let us take another look at the OCV. OCV is a measurement of the predictability of *data* points by the estimate $\hat{f}_{r,\alpha}$. Since *data* should be known to us completely, there should be no unknown parameters in the *data*. So, the *data* should be y instead of y^r, which contains the unknown parameter r. For OCV_1, which leads to the generalized cross validation that does not work in the previous section, every time r changes, the *leaving-out-one* criterion also changes. That is perhaps one of the reasons why it does not work. Therefore it is only appropriate to define OCV as

$$OCV_2(r,\alpha) = \frac{1}{n_1+n_2}\sum_{k=1}^{n_1+n_2}(y_k - K_k f_{r,\alpha}^{[k]})^2. \tag{2.3.1}$$

Now, consider the following identity:

$$y_k - K_k f_{r,\alpha}^{[k]} = \frac{y_k^r - K_k^r \hat{f}_{r,\alpha}}{I_{kk}^{-1}(r) - a_{kk}^*(r,\alpha)},$$

where $I_{kk}^{-1}(r)$ is the kkth entry of $I^{-1}(r)$ and

$$a_{kk}^*(r,\alpha) = \frac{K_k^r \hat{f}_{r,\alpha} - K_k^r f_{r,\alpha}^{[k]}}{y_k - K_k f_{r,\alpha}^{[k]}}.$$

Since $y^r = I^{-1}(r)y$ and $K^r = I^{-1}(r)K$, if we let $\tilde{y}_k^r = K_k^r f_{r,\alpha}^{[k]}$, then $\tilde{y}_k = K_k f_{r,\alpha}^{[k]}$. By the *leaving-out-one* lemma, we have

$$a_{kk}^*(r,\alpha) = \frac{K_k^r h_{r,\alpha}[k,y_k^r] - K_k^r h_{r,\alpha}[k,\tilde{y}_k^r]}{y_k - \tilde{y}_k}.$$

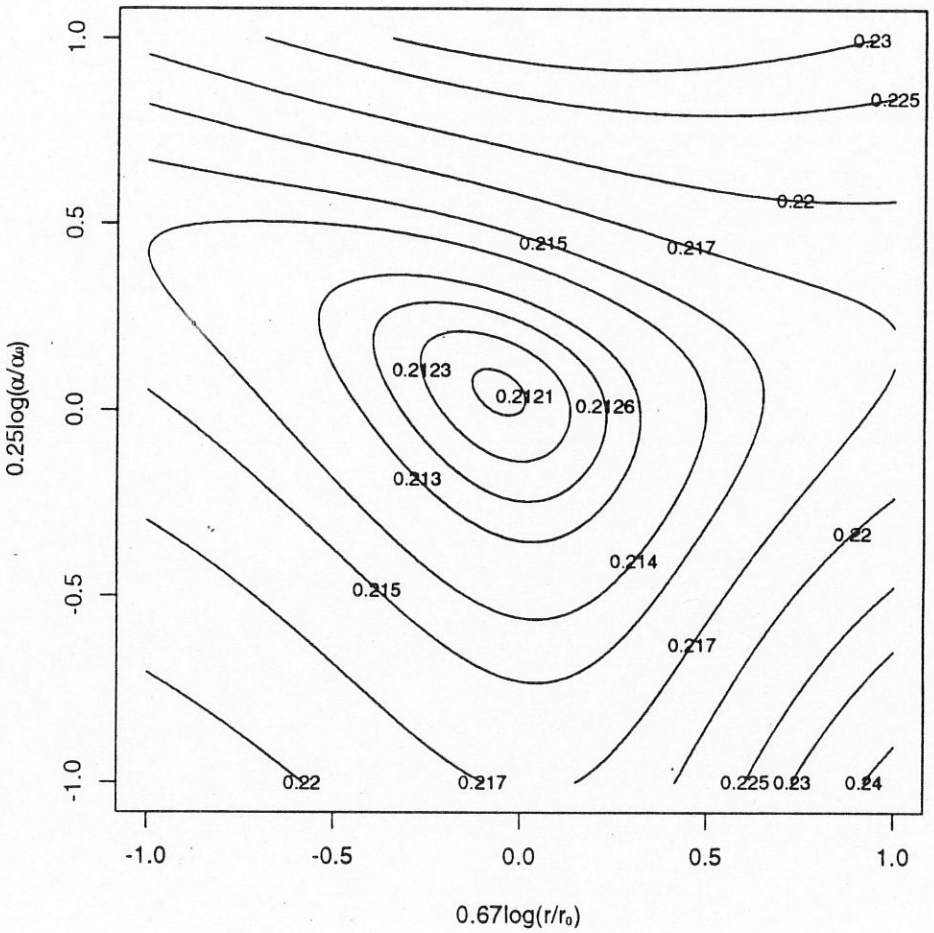


Figure 2.2.1: The RCV function $RCV(r, \alpha)$

Since $K_k^r \hat{f}_{r,\alpha}$ is linear in each data point, we can replace this divided difference by a derivative. So, we have shown that:

$$a_{kk}^*(r,\alpha) = \frac{\partial K_k^r \hat{f}_{r,\alpha}}{\partial y_k} = \frac{\partial K_k^r \hat{f}_{r,\alpha}}{\partial y_k^r}\frac{\partial y_k^r}{\partial y_k} = a_{kk}^r(r,\alpha) I_{kk}^{-1}(r),$$

where $a_{kk}^r(r,\alpha)$ is the kkth entry of $A^r(r,\alpha)$. So we have:

$$y_k - K_k f_{r,\alpha}^{[k]} = \frac{(I_k - A_k^r)I^{-1}(r)y}{(1-a_{kk}^r)I_{kk}^{-1}(r)} = \frac{I_k^{-1}(r)(I-A^r)I^{-1}(r)y}{I_k^{-1}(r)(I-A^r)I_k^{-1}(r)'},$$

where I_k, $I_k^{-1}(r)$ and A_k^r are the kth row of matrices I, $I^{-1}(r)$ and A^r, respectively. The last "=" is obtained by multiplying the numerator and the denominator by $I_{kk}^{-1}(r)$ in an attempt to make the denominator symmetric. Then we have the following OCV *identity*.

Theorem 2.3.1

$$\frac{1}{n_1+n_2}\sum_{k=1}^{n_1+n_2}(y_k - K_k f_{r,\alpha}^{[k]})^2 \equiv OCV_2(r,\alpha) = \frac{1}{n_1+n_2}\sum_{k=1}^{n_1+n_2}\frac{(I_k^{-1}(r)(I-A^r)I^{-1}(r)y)^2}{(I_k^{-1}(r)(I-A^r)I_k^{-1}(r)')^2}$$

□

In order to achieve some invariance properties, we define a new GCV by replacing $I_k^{-1}(r)(I-A^r)I_k^{-1}(r)'$ by $(1/(n_1+n_2))\sum_{i=1}^{n_1+n_2} I_i^{-1}(r)(I-A^r)I_i^{-1}(r)' = (1/(n_1+n_2))tr(I^{-1}(r)(I-A^r)I^{-1}(r)')$, and we have:

$$V(r,\alpha) = \frac{(1/(n_1+n_2))\|I^{-1}(r)[I-A^r(r,\alpha)]I^{-1}(r)y\|^2}{((1/(n_1+n_2))tr(I^{-1}(r)[I-A^r(r,\alpha)]I^{-1}(r)))^2} \tag{2.3.2}$$

This new GCV function is also a natural generalization of the GCV function defined for one source data problem if we look at the problem from a Bayesian point of view.

For a one source data problem, we can write the GCV function as

$$V(\alpha) = \frac{(1/n)\sum_{\nu=1}^{n-M}(z_{\nu n}/(\lambda_{\nu n}+\alpha))^2}{((1/n)\sum_{\nu=1}^{n-M}(1/(\lambda_{\nu n}+\alpha))^2}$$

where, if we assume a stochastic model for the problem, $z_{\nu n} \sim \mathcal{N}(0, b(\lambda_{\nu n}+\alpha))$, for $\nu = 1,\ldots,n-M$. See Chapter 1 and (4.6.2), (6.3.4) and (6.3.6) of Wahba (1990) for more details, and notice that $\alpha = \sigma^2/b = n\lambda$.

For our two source data problem, if we assume that f is from the Bayesian model (1.1.6),

then from (2.1.2) and (2.1.3), we know that

$$y \sim \mathcal{N}(0, \theta(I^{-1}(r)[I - A^r(r,\alpha)]I^{-1}(r))^{-1}).$$

Now, it becomes clear that the new GCV is a generalization of the usual GCV for one source data model from a Bayesian point of view.

We call this estimate GCV-r estimate, where r represents the weighting parameter.

The GCV-r formula looks more symmetric. This function also has a bowl shape near its minimum. Figure 2.3.1 is an example of contour plot of this GCV-r function using the same data from the example described in the end of section 2.1. We can see that this bowl-shaped function is much "sharper" in the r direction. Therefore it is expected to give a "tighter" estimate of r. Later on we will show both theoretically and by simulation that indeed, this GCV-r is substantially better than the RCV in estimating r.

2.4 Bayesian "Confidence Intervals"

It can be shown that the posterior covariance matrix for $\hat{y}^r$ if the stochastic model is correct, is $\sigma_1\sigma_2 A^r(r,\alpha)$, see Wahba (1983). We can estimate σ_k by

$$\hat{\sigma}_k^2 = \frac{\|y_k - \hat{y}_k\|^2}{tr(I_{n_k} - A^r_{kk})},$$

where A^r_{kk} is the kk-th block of A^r, $k = 1, 2$. Once we have $\hat{\sigma}_1$ and $\hat{\sigma}_2$, we can calculate the 95% Bayesian "confidence intervals" by

$$\hat{y}_{1i} \pm 1.96\hat{\sigma}_1\sqrt{A^r_{1i,1i}}$$

and

$$\hat{y}_{2j} \pm 1.96\hat{\sigma}_2\sqrt{A^r_{2j,2j}}$$

We got very good interval estimates in Monte Carlo experiments in Chapter 5.

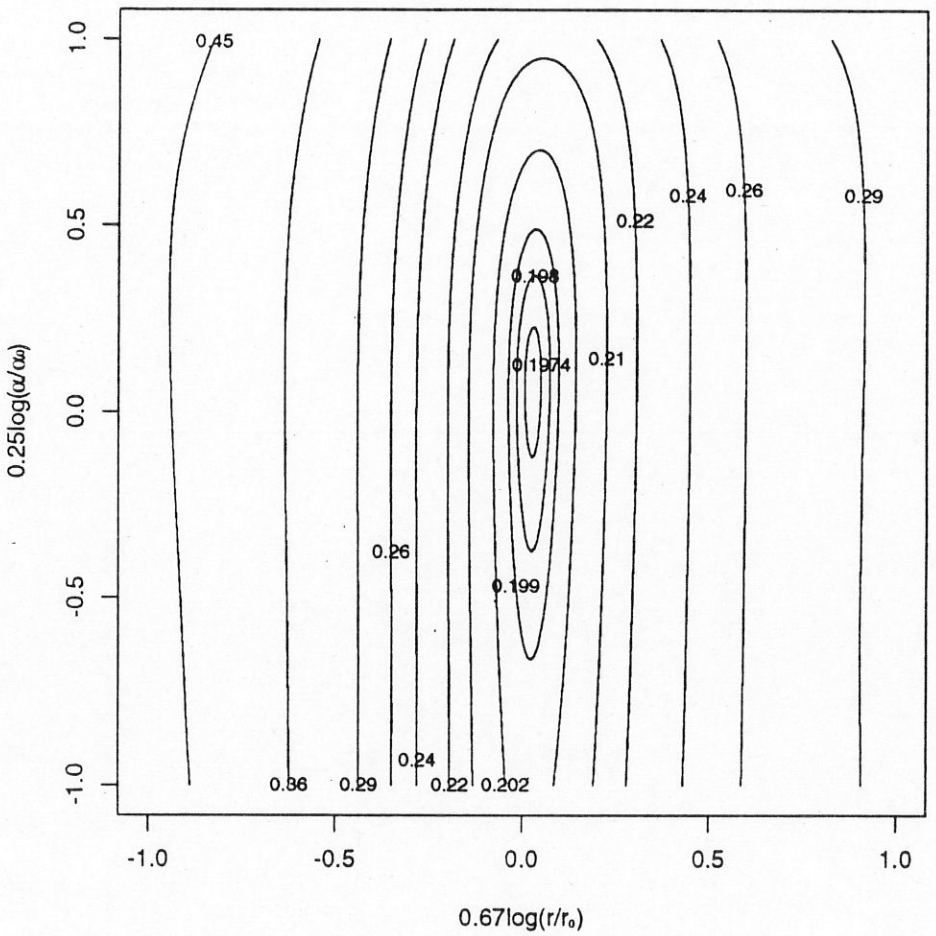


Figure 2.3.1: The GCV-r function $V(r, \alpha)$

Chapter 3

Properties of GML-r, GCV-r and RCV Estimators

In this chapter, we study the properties of GML-r, GCV-r and RCV estimators $\hat{r}_{gml}$, $\hat{\alpha}_{gml}$, etc.. We will obtain the convergence rates for these estimators under certain conditions. We also study the properties of the predictive mean-square error criterion to be defined in Section 3.4. We will give a weak convergence theorem for the GCV-r. And we will briefly discuss the robustness of the estimators.

In this chapter, we always assume (1.2.1) to be our model with $n = n_1 = n_2$. When $n_1 \neq n_2$, it is still possible to obtain similar results, but the derivations and the formulas would become much more complicated. We only study cases where f is a function on the circle. The results can be easily extended to cases where f is a function on the sphere. For example, in Section 3.1 we will get the following result: for cases where f is a function on the circle, and under the assumptions of case 1 described in Section 1.2 and model (1.1.6), we have that as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ n^{1/(4m+4p)}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} \theta_0^2 & 0 & 0 \\ 0 & r_0^2 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}),$$

where $p = \min\{p_1, p_2\}$, p_1 and p_2 are the decay rates of eigenvalues of matrices K_1 and K_2. Now when f is a function on the sphere, $\Sigma = \mathrm{diag}[\xi_{ls}]$, where $\xi_{ls} = [l(l+1)]^{-m}$, for $s = -l, \ldots, l; l = 1, \ldots, M$, and $n = \sum_{l=0}^{M} \sum_{s=-l}^{l} 1$, see Chapter 2 of Wahba (1990). Roughly

speaking, in this case the νth entry of Σ would behave like ν^{-m}, $\nu = 1, \ldots, n$. So, once we replace m with $m/2$, and similarly, replace p_i with $p_i/2$, we will have the following result for f on the sphere:

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ n^{1/(2m+2p)}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} \theta_0^2 & 0 & 0 \\ 0 & r_0^2 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}),$$

Generally speaking, it is not very easy to get estimators of the smoothing parameter with fast convergence rate. There have been some results on the convergence rates of estimators of the smoothing parameter in the literature, all of them are very slow. See, for example, Hardle, Hall and Marron (1988), Stein (1989), Stein (1990), and Hall and Johnstone (1992).

In Wahba (1989), a different kind of method to estimate r is given. Roughly speaking, that method requires that ε_1 and ε_2 have sufficiently different covariance structures and assumes that there exist two matrices B_1 and B_2 such that $B_1 K_1 = B_2 K_2$. Then if we let

$$w = B_2 y_2 - B_1 y_1 = B_2 \varepsilon_2 - B_1 \varepsilon_1 \tag{3.0.1}$$

r can be treated as "smoothing parameter" in the above model, where $B_2\varepsilon_2$ is the "signal" part and $B_1\varepsilon_1$ is the "noise" part. Since we know, generally, it is not easy to estimate the smoothing parameter, and if we assume $B_2\varepsilon_2$ has mth continuous derivatives, it is expected that the estimator of r from this model would have convergence rate $n^{1/(4m)}$.

. From what we shall see about the GML-r and the GCV-r estimators of r, we know that actually we can do better in estimating r without requiring that ε_1 and ε_2 have different covariance structures. Certain amount of information is lost by subtracting data set 1 from data set 2 in the above model (3.0.1).

We will present properties of GML-r, GCV-r and RCV estimators under the three cases described in section 1.2. Since the methods to prove these results are similar, we will only give proofs under case 1.

3.1 Properties of GML-r Estimators

Case 1:

The main result for this section is:

Theorem 3.1.1 For case 1, suppose f comes from the stochastic model (1.1.6), and let $p = \min\{p_1, p_2\}$. Then the GML-r estimator $\hat{\phi}_{gml} \equiv (\hat{\theta}_{gml}, \hat{r}_{gml}, \hat{\alpha}_{gml})'$ is weakly consistent and asymptotically normally distributed; that is as $n \to \infty$, $\hat{\phi}_{gml} \sim \mathcal{N}(\phi, \mathcal{I}_n^{-1})$. The convergence rates for $\hat{r}_{gml}$ and $\hat{\alpha}_{gml}$ are $1/\sqrt{n}$ and $1/n^{1/(4m+4p)}$, respectively. In particular, as $n \to \infty$, we have:

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ n^{1/(4m+4p)}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} \theta_0^2 & 0 & 0 \\ 0 & r_0^2 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}), \tag{3.1.1}$$

where $\theta_0 = \sigma_1\sigma_2$ is the true θ, $r_0 = \sigma_1/\sigma_2$ is the true weighting parameter, $\alpha_0 = \sigma_1\sigma_2/b$ is the true smoothing parameter and c_{33} is a constant which depends on $\theta_0, r_0, \alpha_0, p_1, p_2$ and m.

The GML-r estimate of r and α is based on the distribution of z of (2.1.4), the components of which are not independent. So the standard results on maximum likelihood estimators cannot be applied directly. However, we can use a similar method used by Stein in Stein (1989) and Stein (1990). In Stein (1989), a result from Mardia and Marshall (1984) is used to investigate the properties of maximum likelihood estimators in a linear model when residuals are correlated and when the covariance among the residuals is determined by a parametric model like our case here.

In order to use the result in Mardia and Marshall (1984), we first calculate the Fisher information of model (2.1.4). We have the following:

Lemma 3.1.1 Let $\mathcal{I}$ be the expected Fisher information matrix for (θ, r, α) in the distribution of z in (2.1.4), and let

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_{\theta\theta} & \mathcal{I}_{\theta r} & \mathcal{I}_{\theta\alpha} \\ \mathcal{I}_{\theta r} & \mathcal{I}_{rr} & \mathcal{I}_{r\alpha} \\ \mathcal{I}_{\theta\alpha} & \mathcal{I}_{r\alpha} & \mathcal{I}_{\alpha\alpha} \end{pmatrix}.$$

Then we have:

$$\mathcal{I}_{\theta\theta} = \frac{1}{2\theta^2}(2n-1),$$

$$\mathcal{I}_{\theta r} = \frac{1}{2\theta r} tr[-\frac{1}{r}K_1' K_1 M^{-1} + r K_2' K_2 M^{-1}],$$

$$\mathcal{I}_{\theta\alpha} = \frac{1}{2\theta\alpha}(1 - tr[\frac{1}{r}K_1'K_1M^{-1} + rK_2'K_2M^{-1}]),$$

$$\mathcal{I}_{rr} = \frac{1}{2r^2}(2n + tr[(\frac{1}{r}K_1'K_1M^{-1} - rK_2'K_2M^{-1})^2 - 2(\frac{1}{r}K_1'K_1M^{-1} + rK_2'K_2M^{-1})]),$$

$$\mathcal{I}_{r\alpha} = \frac{1}{2r\alpha}tr[-\frac{1}{r}K_1'K_1M^{-1}\alpha\Sigma^{-1}M^{-1} + rK_2'K_2M^{-1}\alpha\Sigma^{-1}M^{-1}],$$

$$\mathcal{I}_{\alpha\alpha} = \frac{1}{2\alpha^2}(-1 + tr[(\frac{1}{r}K_1'K_1M^{-1} + rK_2'K_2M^{-1})^2]),$$

where $M = (\frac{1}{r}K_1'K_1 + rK_2'K_2 + \alpha\Sigma^{-1})$.

Proof: The derivations of the above results are straightforward, but very lengthy. We discuss the details in Appendix A. □

Note that the above results are true for all the three cases.

For the purpose of evaluating the Fisher information in all the 3 cases described in section 1.2, and also for later use, we need the following lemma:

Lemma 3.1.2 Suppose $0 \leq p < q$, $q > 1$, l_1 and l_2 are integers satisfying $l_1 \geq 0$ and $l_2 \geq 0$. Then as $n \to \infty$, we have:

i). $$\sum_{i=1}^{n} \frac{n^{l_1} i^{l_2 q}}{(n + i^q)^{l_1 + l_2}} = \begin{cases} O(n^{1/q}) & \text{if } l_1 > 0 \\ O(n) & \text{if } l_1 = 0, \end{cases}$$

ii). $$\sum_{i=1}^{n} \left(\frac{i^p}{n + i^q}\right)^{l_1} = \begin{cases} O(n^{1 - l_1(q-p)}) & \text{if } l_1(q-p) < 1 \\ O(\log n) & \text{if } l_1(q-p) = 1 \\ O(1) & \text{if } l_1(q-p) > 1, \end{cases}$$

iii). $$\sum_{i=1}^{n} \frac{1}{i^p} = \begin{cases} O(n^{1-p}) & \text{if } 0 \leq p < 1 \\ O(\log n) & \text{if } p = 1 \\ O(1) & \text{if } p > 1. \end{cases}$$

Proof: Using some quadrature formulas to approximate the sums, we will get the results. A similar kind of calculation appeared in Cox (1988), where a result similar to part of our results in the above lemma was given. Details of the proof will be shown in Appendix A. □

By applying Lemma 3.1.1 and Lemma 3.1.2 to case 1, and letting

$$\mathcal{I}^{-1} = \begin{pmatrix} \mathcal{I}^{\theta\theta} & \mathcal{I}^{\theta r} & \mathcal{I}^{\theta\alpha} \\ \mathcal{I}^{\theta r} & \mathcal{I}^{rr} & \mathcal{I}^{r\alpha} \\ \mathcal{I}^{\theta\alpha} & \mathcal{I}^{r\alpha} & \mathcal{I}^{\alpha\alpha} \end{pmatrix},$$

we have:

Lemma 3.1.3 In case 1, letting $p = \min\{p_1, p_2\}$, we have:

$$\mathcal{I}_{\theta\theta} = (2n-1)/(2\theta^2), \quad \mathcal{I}_{\theta r} = O(n^{1/(2m+2p)}), \quad \mathcal{I}_{\theta\alpha} = O(n^{1/(2m+2p)}),$$
$$\mathcal{I}_{rr} = (2n + o(n))/(2r^2), \quad \mathcal{I}_{r\alpha} = O(n^{1/(2m+2p)}),$$
$$\mathcal{I}_{\alpha\alpha} = O(n^{1/(2m+2p)}).$$

And therefore

$$\mathcal{I}^{\theta\theta} = \theta^2/n + o(n^{-1}), \quad \mathcal{I}^{\theta r} = O(n^{-2+1/(2m+2p)}), \quad \mathcal{I}^{\theta\alpha} = O(n^{-1}),$$
$$\mathcal{I}^{rr} = r^2/n + o(n^{-1}), \quad \mathcal{I}^{r\alpha} = O(n^{-1}),$$
$$\mathcal{I}^{\alpha\alpha} = O(n^{-1/(2m+2p)}).$$

Proof: According to Lemma 3.1.1, all the entries in $\mathcal{I}$ are traces of certain matrices, and these traces can be seen to be equivalent to the sums in Lemma 3.1.2. Also notice that under our model (1.2.1), the matrices K_1, K_2 and Σ are diagonal. This makes the multiplication of the matrices interchangeable. For more details, please see Appendix A. □

Now we are ready to prove our main result, Theorem 3.1.1:

Proof: We can apply Theorem 1 of Mardia and Marshall (1984) directly. Using our Lemma 3.1.1 and Lemma 3.1.3, we can verify conditions (3.1) and (3.3) of Theorem 1 of Mardia and Marshall (1984). We do not need to check condition (3.4), because the mean function is identically 0 here. These conditions are sufficient for the asymptotic normality and weak consistency of $\hat{\phi}_{gml}$. These conditions actually guarantee the growth and convergence of the Fisher information. For more details on the verification of condition (3.1) and (3.3) of Theorem 1 of Mardia and Marshall (1984), please see Appendix A. □

If we had only one data source with direct observations, then there would be no weighting parameter, and $p = 0$, the convergence rate for the smoothing parameter becomes $n^{1/(4m)}$,

which is exactly the same rate Stein (1989) and Stein (1990) get.

In our result here, we see that the smaller the p, the faster the convergence rate is for $\hat{\alpha}$. This is intuitively correct, because if we define the energy of signal in a set of data by $(b/n)tr(K\Sigma K')$, then the smaller the p is, the slower the decay rate of the energy spectrum in the signal would be, and therefore the more information about the signal we will have from that data set, and the better we can estimate α.

In the above result, the asymptotic variance of $\hat{\theta}$ is θ^2. This is similar to the simple case when $y \sim \mathcal{N}(0, \sigma^2)$, the ML estimator of σ^2 satisfies $\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, 2\sigma^4)$. Here we have 2 data sets of n observations, so n becomes $2n$, and hence we have the above result.

We have not been able to get a closed form for the constant c_{33} in the asymptotic covariance matrix. However, we can get approximate values of it numerically by using Lemma 3.1.1 once we know $\theta, r, \alpha, p_1, p_2$ and m.

Case 2:

Applying Lemma 3.1.1 and Lemma 3.1.2 to case 2, and using a similar method, we can get the weak consistency and asymptotic normality results for case 2. We have:

Theorem 3.1.2 For case 2, under the stochastic assumption, we have:

i). If $1 - 4(m - p_2) < 1/(2m + 2p_1)$, then as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ n^{1/(4m+4p_1)}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} \theta_0^2 & 0 & 0 \\ 0 & r_0^2 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}). \tag{3.1.2}$$

This is similar to case 1.

ii). If $1/(2m + 2p_1) \leq 1 - 4(m - p_2)$, then as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ n^{1/2-2(m-p_2)}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{pmatrix}), \tag{3.1.3}$$

where c_{ij} are constants depending on $\theta_0, r_0, \alpha_0, p_1, p_2$ and m. □

We have a different kind of result here. Asymptotically, $\hat{\theta}, \hat{r}$, and $\hat{\alpha}$ are not independent any more, and the convergence rate of $\hat{\alpha}$ depends on p_2 in a different manner. Now recall that in case 2, the "design" matrix for data set two has a different form (see section 1.2), so the larger p_2 is, the slower the decay rate of the energy spectrum in the signal, and therefore

the more information we get from data set two, and the better we can estimate α. That is why the larger is p_2, the faster is the convergence rate of $\hat{\alpha}$. However, on the other hand, intuitively we know that if p_2 is too close to m or $p_2 = m$, then in the second data set, the signal and the noise will have covariance structures that are too close to each other or even exactly the same. In this case we can not tell the signal from the noise in a sensible way. Therefore it should be impossible to estimate α. In fact, it can be shown that when $p_2 \nearrow m$, those constants c_{ij} will increase and become very large. So, in order to have a good estimate of α, we should not only look at the convergence rate, but also should require that the constants c_{ij} be not too large, that is, p_2 should not be too close to m.

We have not been able to get closed forms for those constants c_{ij} in the asymptotic covariance matrices. However, we can get approximate values of them numerically by using Lemma 3.1.1 once we know $\theta, r, \alpha, p_1, p_2$ and m. Below, after the theorem for case 3, is a simplified example in which the closed forms for these c_{ij}'s are obtained. It can give us some idea on how these c_{ij}'s would change along with p_1, p_2 and m.

Case 3:

Similarly, we have:

Theorem 3.1.3 For case 3, let $p = \max\{p_1, p_2\}$. Then under the stochastic assumption, we have:

i). If $m - 1/4 = p$, then as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ \sqrt{\log n}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} \theta_0^2 & 0 & 0 \\ 0 & r_0^2 & 0 \\ 0 & 0 & c_{33} \end{pmatrix}). \tag{3.1.4}$$

ii). If $m - 1/4 < p < m$, then as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta_0) \\ \sqrt{n}(\hat{r}_{gml} - r_0) \\ n^{1/2-2(m-p)}(\hat{\alpha}_{gml} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{pmatrix}). \tag{3.1.5}$$

This is similar to the second result in case 2. □

We do not consider the case when $p < m - 1/4$, because if $p < m - 1/4$, then $\mathcal{I}^{\alpha\alpha}$ does not go to zero as $n \to \infty$. Therefore $\hat{\alpha}_{gml}$ does not have the weak consistency. This is because when p is too small, the energy spectra in the signal decay too fast, so we do not

get enough information about the signal from the data, and therefore we cannot get a good estimate of α.

So, in order to have a good estimate of α, p has to be big enough (at least $p = m - 1/4$) to assure enough information, and hence a good convergence rate of $\hat{\alpha}$. On the other hand, as $p \nearrow m$, the constants c_{ij} will increase.

We are unable to get closed forms for those constants c_{ij} in the asymptotic covariance matrices. However, the following simplified example illustrates how those c_{ij} in the simplified situation change along with p and m. This example can give us some idea about how the c_{ij} in our Theorem 3.1.3 might change when p changes from $m - 1/4 = p$ to $m - 1/4 < p < m$.

Suppose

$$y_\nu = f_\nu + \varepsilon_\nu,$$

where f_ν's are independent, ε_ν's are *i.i.d.*, f_ν's and ε_ν's are independent, and $f_\nu \sim \mathcal{N}(0, b/\nu^q)$, $\varepsilon_\nu \sim \mathcal{N}(0, \theta)$. Then the q in this example corresponds to $2m - 2p$ in our Theorem 3.1.3. We will only consider cases where $0 < q \leq 1/2$. It is easy to see

$$y_\nu \sim \mathcal{N}(0, \theta\left(\frac{1 + \alpha\nu^q}{\alpha\nu^q}\right)),$$

where $\alpha = \theta/b$.

The log likelihood of $y = (y_1, \ldots, y_n)'$ is

$$L = C - \frac{n}{2}\log\theta - \frac{1}{2}\sum_{\nu=1}^{n}\log\frac{1 + \alpha\nu^q}{\alpha\nu^q} - \frac{1}{2\theta}\sum_{\nu=1}^{n}\frac{\alpha\nu^q}{1 + \alpha\nu^q}y_\nu^2,$$

where C is a constant. By definition, the Fisher information is

$$\mathcal{I} = \begin{pmatrix} -E(\partial^2 L/\partial\theta^2) & -E(\partial^2 L/\partial\theta\partial\alpha) \\ -E(\partial^2 L/\partial\theta\alpha) & -E(\partial^2 L/\partial\alpha^2) \end{pmatrix}.$$

It is not hard to show that

$$-E(\frac{\partial^2 L}{\partial\theta^2}) = \frac{n}{2\theta^2},$$

$$-E(\frac{\partial^2 L}{\partial\theta\partial\alpha}) = -\frac{1}{2\theta\alpha}\sum_{\nu=1}^{n}\frac{1}{1 + \alpha\nu^q},$$

$$-E(\frac{\partial^2 L}{\partial \alpha^2}) = \frac{1}{2\alpha^2}\sum_{\nu=1}^{n}\frac{1}{(1+\alpha\nu^q)^2}.$$

For simplicity and without loss of generality, we assume $\alpha = 1$. Then

$$\mathcal{I} = \begin{pmatrix} n/(2\theta^2) & -(1/(2\theta))\sum_{\nu=1}^{n} 1/(1+\nu^q) \\ -(1/(2\theta))\sum_{\nu=1}^{n} 1/(1+\nu^q) & (1/2)\sum_{\nu=1}^{n} 1/(1+\nu^q)^2 \end{pmatrix}.$$

Now let

$$a_{1n}(q) = \frac{\sum_{\nu=1}^{n} 1/(1+\nu^q)}{\sum_{\nu=1}^{n} 1/\nu^q},$$

$$a_{2n}(q) = \frac{\sum_{\nu=1}^{n} 1/(1+\nu^q)^2}{\sum_{\nu=1}^{n} 1/\nu^{2q}}.$$

Then

$$\sum_{\nu=1}^{n}\frac{1}{1+\nu^q} = a_{1n}(q)\sum_{\nu=1}^{n}\frac{1}{\nu^q},$$

$$\sum_{\nu=1}^{n}\frac{1}{(1+\nu^q)^2} = a_{2n}(q)\sum_{\nu=1}^{n}\frac{1}{\nu^{2q}}.$$

It can be shown that for any $0 < q \leq 1/2$, we have

$$\lim_{n\to\infty} a_{1n}(q) = 1,$$

$$\lim_{n\to\infty} a_{2n}(q) = 1.$$

We also have

$$\sum_{\nu=1}^{n}\frac{1}{\nu^q} = n^{1-q}\sum_{\nu=1}^{n}\left(\frac{\nu}{n}\right)^{-q}\frac{1}{n} = n^{1-q}b_{1n}(q), \qquad 0 < q \leq \frac{1}{2}$$

$$\sum_{\nu=1}^{n}\frac{1}{\nu^{2q}} = \begin{cases} n^{1-2q}\sum_{\nu=1}^{n}(\nu/n)^{-2q}(1/n) \\ \log n + C_n \end{cases} = \begin{cases} n^{1-2q}b_{2n}(q), & 0 < q < 1/2 \\ (\log n)b_{3n}, & q = 1/2 \end{cases}$$

where

$$\lim_{n\to\infty} b_{1n}(q) = \int_0^1 x^{-q}dx = \frac{1}{1-q}, \text{ for } 0 < q \leq \frac{1}{2}$$

$$\lim_{n\to\infty} b_{2n}(q) = \int_0^1 x^{-2q}dx = \frac{1}{1-2q}, \text{ for } 0 < q < \frac{1}{2}$$

and

$$\lim_{n\to\infty} b_{3n} = 1, \text{ for } q = \frac{1}{2}.$$

So, we have

$$\mathcal{I} \approx \begin{pmatrix} n/(2\theta^2) & -(1/(2\theta(1-q)))n^{1-q} \\ -(1/(2\theta(1-q)))n^{1-q} & (1/2)\log n \end{pmatrix}, \text{ for } q = \frac{1}{2}$$

$$\mathcal{I} \approx \begin{pmatrix} n/(2\theta^2) & -(1/(2\theta(1-q)))n^{1-q} \\ -(1/(2\theta(1-q)))n^{1-q} & (1/(2(1-2q)))n^{1-2q} \end{pmatrix}, \text{ for } 0 < q < \frac{1}{2}$$

and as $n \to \infty$, "$\approx$" will become "=".

Therefore, we have

$$\mathcal{I}^{-1} \approx \begin{pmatrix} 2\theta^2/n & 2\theta/((1-q)\sqrt{n}\log n) \\ 2\theta/((1-q)\sqrt{n}\log n) & 2/\log n \end{pmatrix}, \text{ for } q = \frac{1}{2}$$

$$\mathcal{I}^{-1} \approx \begin{pmatrix} 2\theta^2(1-q)^2/(q^2n) & 2\theta(1-2q)(1-q)/(q^2n^{1-q}) \\ 2\theta(1-2q)(1-q)/(q^2n^{1-q}) & 2(1-2q)(1-q)^2/(q^2n^{1-2q}) \end{pmatrix},$$

$$\text{for } 0 < q < \frac{1}{2}$$

and as $n \to \infty$, "$\approx$" will become "=".

So, as $n \to \infty$, we have:

i). For $q = 1/2$, which corresponds to $m - 1/4 = p$ in Theorem 3.1.3,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta) \\ \sqrt{\log n}(\hat{\alpha}_{gml} - \alpha) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} 2\theta^2 & 0 \\ 0 & 2 \end{pmatrix}),$$

ii). For $0 < q < 1/2$, which corresponds to $m - 1/4 < p < m$ in Theorem 3.1.3,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_{gml} - \theta) \\ n^{1/2-q}(\hat{\alpha}_{gml} - \alpha) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} 2\theta^2(1-q)^2/q^2 & 2\theta(1-2q)(1-q)/q^2 \\ 2\theta(1-2q)(1-q)/q^2 & 2(1-2q)(1-q)^2/q^2 \end{pmatrix}).$$

We can see that when $q = 1/2$, $2\theta^2(1-q)^2/q^2 = 2\theta^2$, $2\theta(1-2q)(1-q)/q^2 = 0$. However, when $q = 1/2$, $2(1-2q)(1-q)^2/q^2 = 0 \neq 2$. Also notice that when $q \to 0$, which corresponds to $p \nearrow m$ in our Theorem 3.1.3, c_{ij} will increase.

Finally, according to the above theorems and by Slutsky's Theorem, we have:

Theorem 3.1.4 For all the cases considered above, as $n \to \infty$, we have

$$\hat{r}_{gml} \xrightarrow{P} r_0, \qquad \hat{\alpha}_{gml} \xrightarrow{P} \alpha_0. \tag{3.1.6}$$

□

3.2 Properties of GCV-r Estimators

In this section we study the properties of the GCV-r estimators. The method of proving these properties is similar to the method used by Stein in Stein (1989) and Stein (1990). As before, we will only prove the results for case 1. A similar method can be used to get the results for case 2 and case 3.

Case 1:

The main result for this section is:

Theorem 3.2.1 For case 1, suppose f comes from the stochastic model (1.1.6), and let $p = \min\{p_1, p_2\}$. Then the GCV-r estimators $\hat{r}_{gcv}$ and $\hat{\alpha}_{gcv}$ are weakly consistent and asymptotically normally distributed. The convergence rates for $\hat{r}_{gcv}$ and $\hat{\alpha}_{gcv}$ are $1/\sqrt{n}$ and $1/n^{1/(4m+4p)}$, respectively. In particular, as $n \to \infty$ we have:

$$\begin{cases} \sqrt{n}(\hat{r}_{gcv} - r_0) \xrightarrow{d} \mathcal{N}(0, c_1) \\ n^{1/(4m+4p)}(\hat{\alpha}_{gcv} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_2) \end{cases} \tag{3.2.1}$$

where c_1 and c_2 are constants depending on $\theta_0, r_0, \alpha_0, p_1, p_2$ and m.

We are unable to get closed forms for the constants c_1 and c_2. It is expected that when the stochastic model is true, these two constants would be larger than the constants in the covariance matrix for GML-r estimators (see Stein (1989) and Stein (1990)). However, in Chapters 4 and 5, we will show some simulation results, which indicate that in some cases, c_1 and c_2 are not much bigger.

Sketch of proof of Theorem 3.2.1: Here and afterward, we will use $\hat{r}$ for $\hat{r}_{gcv}$ and $\hat{\alpha}$ for $\hat{\alpha}_{gcv}$ whenever it will not cause ambiguity. Also notice that under our assumptions, the matrices K_1, K_2 and Σ are all diagonal matrices, and therefore the multiplication of these matrices are interchangeable.

Our GCV-r function can be written as

$$V(r,\alpha) = \frac{(1/(2n))\|W^{-1}(r,\alpha)y\|^2}{((1/(2n))tr(W^{-1}(r,\alpha)))^2}$$

where $W^{-1}(r,\alpha) = I^{-1}(r)[I - A^r]I^{-1}(r)$. The GCV-$r$ estimators $(\hat{r},\hat{\alpha})$ must satisfy

$$\begin{cases} \frac{\partial}{\partial r}V(\hat{r},\hat{\alpha}) = 0 \\ \frac{\partial}{\partial \alpha}V(\hat{r},\hat{\alpha}) = 0 \end{cases}$$

which, by straightforward calculation, can be shown to be equivalent to

$$\begin{cases} w_r(\hat{r},\hat{\alpha}) = 0 \\ w_\alpha(\hat{r},\hat{\alpha}) = 0 \end{cases}$$

where

$$\begin{cases} w_r(r,\alpha) = y'\{W^{-1}\frac{\partial}{\partial r}W^{-1}tr(W^{-1}) - W^{-2}tr(\frac{\partial}{\partial r}W^{-1})\}y \\ w_\alpha(r,\alpha) = y'\{W^{-1}\frac{\partial}{\partial \alpha}W^{-1}tr(W^{-1}) - W^{-2}tr(\frac{\partial}{\partial \alpha}W^{-1})\}y \end{cases} \tag{3.2.2}$$

By mean-value theorem, we have:

$$\begin{cases} 0 = w_r(\hat{r},\hat{\alpha}) = w_r(r_0,\alpha_0) + \frac{\partial}{\partial r}w_r(r_*,\hat{\alpha})(\hat{r} - r_0) + \frac{\partial}{\partial \alpha}w_r(r_0,\alpha_*)(\hat{\alpha} - \alpha_0) \\ 0 = w_\alpha(\hat{r},\hat{\alpha}) = w_\alpha(r_0,\alpha_0) + \frac{\partial}{\partial r}w_\alpha(r_{**},\hat{\alpha})(\hat{r} - r_0) + \frac{\partial}{\partial \alpha}w_\alpha(r_0,\alpha_{**})(\hat{\alpha} - \alpha_0) \end{cases}$$

where r_* and r_{**} are points between r_0 and $\hat{r}$, and α_* and α_{**} are points between α_0 and $\hat{\alpha}$. This gives us (provided that the matrix is non-singular):

$$\begin{pmatrix} \hat{r} - r_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix} = -\begin{pmatrix} \frac{\partial}{\partial r}w_r(r_*,\hat{\alpha}) & \frac{\partial}{\partial \alpha}w_r(r_0,\alpha_*) \\ \frac{\partial}{\partial r}w_\alpha(r_{**},\hat{\alpha}) & \frac{\partial}{\partial \alpha}w_\alpha(r_0,\alpha_{**}) \end{pmatrix}^{-1}\begin{pmatrix} w_r(r_0,\alpha_0) \\ w_\alpha(r_0,\alpha_0) \end{pmatrix}.$$

We then need the following two lemmas:

Lemma 3.2.1 In case 1, as $n \to \infty$, we have:

$$\frac{\partial}{\partial r} w_r(r_*, \hat{\alpha}) = E(\frac{\partial}{\partial r} w_r)(1 + o_p(n^{-\epsilon})), \quad \frac{\partial}{\partial \alpha} w_r(r_0, \alpha_*) = E(\frac{\partial}{\partial \alpha} w_r)(1 + o_p(n^{-\epsilon})),$$

$$\frac{\partial}{\partial r} w_\alpha(r_{**}, \hat{\alpha}) = E(\frac{\partial}{\partial r} w_\alpha)(1 + o_p(n^{-\epsilon})), \quad \frac{\partial}{\partial \alpha} w_\alpha(r_0, \alpha_{**}) = E(\frac{\partial}{\partial \alpha} w_\alpha)(1 + o_p(n^{-\epsilon})).$$

where ϵ is some positive number, and $o_p(n^{-\epsilon})$ is some random variable, say x, such that $n^{\epsilon} x \xrightarrow{P} 0$, as $n \to \infty$.

Proof: See Appendix A. □

Lemma 3.2.2 In case 1, as $n \to \infty$, we have:

$$E\frac{\partial}{\partial r} w_r(r_*, \hat{\alpha}) = O(n^2), \qquad E\frac{\partial}{\partial \alpha} w_r(r_0, \alpha_*) = O(n^{1+1/(2m+2p)}),$$

$$E\frac{\partial}{\partial r} w_\alpha(r_{**}, \hat{\alpha}) = O(n^{1+1/(2m+2p)}), \quad E\frac{\partial}{\partial \alpha} w_\alpha(r_0, \alpha_{**}) = O(n^{1+1/(2m+2p)}).$$

Proof: See Appendix A. □

From those two lemmas we can see that when n is sufficiently large, the above matrix is (with probability 1) non-singular. And as $n \to \infty$ we have:

$$\begin{pmatrix} \hat{r} - r_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix} = \begin{pmatrix} O(1/n^2) & O(1/n^2) \\ O(1/n^2) & O(1/n^{1+1/(2m+2p)}) \end{pmatrix} \begin{pmatrix} w_r(r_0, \alpha_0) \\ w_\alpha(r_0, \alpha_0) \end{pmatrix}.$$

By definition, $w_r(r_0, \alpha_0)$ and $w_\alpha(r_0, \alpha_0)$ are quadratic forms, which can be written as

$$w_r(r_0, \alpha_0) = y' V_r y$$

and

$$w_\alpha(r_0, \alpha_0) = y' V_\alpha y$$

where $y = (y_{11}, \ldots, y_{1n}, y_{21}, \ldots, y_{2n})'$, and V_r and V_α are some $2n$ by $2n$ matrices (see (3.2.2) for closed forms of V_r and V_α). Notice that since we assume K_1 and K_2 to be diagonal matrices, matrix A^r is of the form:

$$A^r = \begin{pmatrix} A^r_{11} & A^r_{12} \\ A^r_{21} & A^r_{22} \end{pmatrix}$$

with each block being a diagonal matrix, and so are the matrices V_r and V_α, i.e. V_r and V_α can be written as

$$V_r = \begin{pmatrix} V_r^{11} & V_r^{12} \\ V_r^{21} & V_r^{22} \end{pmatrix}$$

and

$$V_\alpha = \begin{pmatrix} V_\alpha^{11} & V_\alpha^{12} \\ V_\alpha^{21} & V_\alpha^{22} \end{pmatrix}$$

where each block is a diagonal matrix. Also notice that

$$y \sim \mathcal{N}(0, \theta(I^{-1}(r)[I - A^r]I^{-1}(r))^{-1})$$

where $(I^{-1}(r)[I-A^r]I^{-1}(r))^{-1}$ is also a 2×2 block matrix with each block a diagonal matrix. This means $(y_{11}, y_{21})', (y_{12}, y_{22})', \ldots, (y_{1n}, y_{2n})'$ are independent of each other. Therefore $w_r(r_0, \alpha_0)$ and $w_\alpha(r_0, \alpha_0)$ can be written as

$$w_r(r_0, \alpha_0) = \sum_{i=1}^{n} (y_{1i}, y_{2i}) \begin{pmatrix} V_r^{11}(i) & V_r^{12}(i) \\ V_r^{21}(i) & V_r^{22}(i) \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \equiv \sum_{i=1}^{n} x_{ri}$$

and

$$w_\alpha(r_0, \alpha_0) = \sum_{i=1}^{n} (y_{1i}, y_{2i}) \begin{pmatrix} V_\alpha^{11}(i) & V_\alpha^{12}(i) \\ V_\alpha^{21}(i) & V_\alpha^{22}(i) \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \equiv \sum_{i=1}^{n} x_{\alpha i}$$

where $V_r^{11}(i)$ is the ith diagonal element of the diagonal matrix V_r^{11} and so forth, and x_{ri}, $i = 1, \ldots, n$ are independent and $x_{\alpha i}$, $i = 1, \ldots, n$ are also independent. So, we see that $w_r(r_0, \alpha_0)$ and $w_\alpha(r_0, \alpha_0)$ are sums of independent random variables. We can apply Liapounov's theorem (see Chung (1974)) to $w_r(r_0, \alpha_0)$ and $w_\alpha(r_0, \alpha_0)$ to get the following lemma:

Lemma 3.2.3 In case 1, as $n \to \infty$, we have:

$$n^{-3/2} w_r(r_0, \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_1),$$

$$n^{-(1+1/(4m+4p))} w_\alpha(r_0, \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_2).$$

Proof: See Appendix A. □

Using this lemma and assuming that $p \geq 0$ and $m > 1/2$, which is usually true, the

result of Theorem 3.2.1 will follow. □

Case 2:

By the same methods, we can get the following:

Theorem 3.2.2 For case 2, under the stochastic assumption, we have:

i). If $1 - 4(m - p_2) < 1/(2m + 2p_1)$, then as $n \to \infty$,

$$\begin{cases} \sqrt{n}(\hat{r}_{gcv} - r_0) \xrightarrow{d} \mathcal{N}(0, c_1) \\ n^{1/(4m+4p_1)}(\hat{\alpha}_{gcv} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_2) \end{cases} \tag{3.2.3}$$

ii). If $1/(2m + 2p_1) \leq 1 - 4(m - p_2)$, then as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{r}_{gcv} - r_0) \\ n^{1/2-2(m-p_2)}(\hat{\alpha}_{gcv} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.2.4}$$

□

Case 3:

Similarly, we have:

Theorem 3.2.3 For case 3, let $p = \max\{p_1, p_2\}$. Then under the stochastic assumption, we have:

i). If $m - 1/4 = p$, then as $n \to \infty$,

$$\begin{cases} \sqrt{n}(\hat{r}_{gcv} - r_0) \xrightarrow{d} \mathcal{N}(0, c_1) \\ \sqrt{\log n}(\hat{\alpha}_{gcv} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_2) \end{cases} \tag{3.2.5}$$

ii). If $m - 1/4 < p < m$, then as $n \to \infty$,

$$\begin{pmatrix} \sqrt{n}(\hat{r}_{gcv} - r_0) \\ n^{1/2-2(m-p)}(\hat{\alpha}_{gcv} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.2.6}$$

□

Finally, according to the above theorems and by Slutsky's Theorem, we have:

Theorem 3.2.4 For all the cases considered above, as $n \to \infty$, we have

$$\hat{r}_{gcv} \xrightarrow{P} r_0, \qquad \hat{\alpha}_{gcv} \xrightarrow{P} \alpha_0. \tag{3.2.7}$$

□

The above results show that the GCV-r estimators have the same convergence rates as those for GML-r estimators under the stochastic model.

3.3 Properties of RCV Estimators

In this section, we study the properties of RCV estimators. Since the method of proving the properties is similar to the method used in the previous section, we will only briefly outline the basic steps for proving the following Theorem 3.3.1.

Case 1:

The main result for this section is:

Theorem 3.3.1 For case 1, suppose f comes from the stochastic model (1.1.6), and let $p = \min\{p_1, p_2\}$. Then as $n \to \infty$ we have:

$$\begin{pmatrix} n^{1/(4m+4p)}(\hat{r}_{rcv} - r_0) \\ n^{1/(4m+4p)}(\hat{\alpha}_{rcv} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.3.1}$$

where c_{ij} are constants depending on $\theta_0, r_0, \alpha_0, p_1, p_2$ and m.

Proof: We will only loosely go through the basic steps in the proof, from which one can see why the result in the theorem is true. A rigorous proof is possible by applying the same method in the proof of Theorem 3.2.1. As before, we will use $\hat{r}$ for $\hat{r}_{rcv}$ and $\hat{\alpha}$ for $\hat{\alpha}_{rcv}$.

Recall that the RCV function is:

$$RCV(r, \alpha) = \frac{1}{2n}\left\{\frac{\|y_1 - \hat{y}_1\|^2}{((1/n)tr[I - A_{11}^r])^2} + \frac{\|y_2 - \hat{y}_2\|^2}{((1/n)tr[I - A_{22}^r])^2}\right\}.$$

and

$$\begin{cases} \hat{y}_1 = A_{11}^r y_1 + r A_{12}^r y_2 \\ \hat{y}_2 = (1/r) A_{21}^r y_1 + A_{22}^r y_2. \end{cases}$$

Also notice that (remembering that the multiplications of K_1, K_2 and M^{-1} are interchangeable because they are diagonal matrices):

$$\begin{pmatrix} A_{11}^r & A_{12}^r \\ A_{21}^r & A_{22}^r \end{pmatrix} = \begin{pmatrix} (1/r)M^{-1}K_1^2 & M^{-1}K_1K_2 \\ M^{-1}K_1K_2 & rM^{-1}K_2^2 \end{pmatrix},$$

where $M = ((1/r)K_1^2 + rK_2^2 + \alpha\Sigma^{-1})$.

So we have:

$$y_1 - \hat{y}_1 = [(I_n, 0) - (A_{11}^r,\ rA_{12}^r)]y$$

$$= [rM^{-1}K_2^2 + \alpha\Sigma^{-1}M^{-1},\ -rM^{-1}K_1K_2]\left(\begin{bmatrix} K_1 f \\ K_2 f \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}\right)$$

$$= \alpha\Sigma^{-1}M^{-1}K_1 f + [rM^{-1}K_2^2 + \alpha\Sigma^{-1}M^{-1},\ -rM^{-1}K_1K_2]\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix},$$

and similarly,

$$y_2 - \hat{y}_2 = \alpha\Sigma^{-1}M^{-1}K_2 f + [-\frac{1}{r}M^{-1}K_1K_2,\ \frac{1}{r}M^{-1}K_1^2 + \alpha M^{-1}\Sigma^{-1}]\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

To get rid of the cross terms (of f and ε) in the RCV function, and therefore to make the argument simple, let us take the mathematical expectation of RCV with respect to f, and we have:

$$E_f RCV(r,\alpha) = \frac{1}{2n}\{tr(b\alpha^2 M^{-2}\Sigma^{-1}K_1^2)/(\frac{1}{n}tr(I_n - A_{11}^r))^2$$

$$+tr(b\alpha^2 M^{-2}\Sigma^{-1}K_2^2)/(\frac{1}{n}tr(I_n - A_{22}^r))^2$$

$$+\varepsilon'[\frac{1}{((1/n)tr(I_n - A_{11}^r))^2}\begin{pmatrix} I_n - 2A_{11}^r + A_{11}^{r2} & -rA_{12}^r + rA_{11}^r A_{12}^r \\ -rA_{12}^r + rA_{11}^r A_{12}^r & r^2 A_{12}^{r2} \end{pmatrix}$$

$$+\frac{1}{((1/n)tr(I_n - A_{22}^r))^2}\begin{pmatrix} (1/r^2)A_{21}^{r2} & -\frac{1}{r}A_{21}^r + \frac{1}{r}A_{21}^r A_{22}^r \\ -\frac{1}{r}A_{21}^r + \frac{1}{r}A_{21}^r A_{22}^r & I_n - 2A_{22}^r + A_{22}^{r2} \end{pmatrix}]\varepsilon\}$$

$$\equiv \frac{1}{2n}\{B^*(r,\alpha) + \varepsilon' V^*(r,\alpha)\varepsilon\}$$

where $B^*(r,\alpha)$ is the bias term and $\varepsilon' V^*(r,\alpha)\varepsilon$ is the variance term.

Now, the randomness only comes from the noise ε. Roughly speaking, $(\hat{r}, \hat{\alpha})$ should satisfy

$$\begin{cases} \frac{\partial}{\partial r}E_f RCV(\hat{r},\hat{\alpha}) = 0 \\ \frac{\partial}{\partial \alpha}E_f RCV(\hat{r},\hat{\alpha}) = 0 \end{cases}$$

which, is equivalent to

$$\begin{cases} R_r(\hat{r},\hat{\alpha}) = 0 \\ R_\alpha(\hat{r},\hat{\alpha}) = 0 \end{cases}$$

where

$$\begin{cases} R_r(r,\alpha) = \{\frac{\partial}{\partial r}B^* + \varepsilon'\frac{\partial}{\partial r}V^*\varepsilon\} \\ R_\alpha(r,\alpha) = \{\frac{\partial}{\partial \alpha}B^* + \varepsilon'\frac{\partial}{\partial \alpha}V^*\varepsilon\}. \end{cases}$$

By first order Taylor expansion, we have:

$$\begin{cases} 0 = R_r(\hat{r},\hat{\alpha}) \approx R_r(r_0,\alpha_0) + \frac{\partial}{\partial r}R_r(r_0,\alpha_0)(\hat{r}-r_0) + \frac{\partial}{\partial \alpha}R_r(r_0,\alpha_0)(\hat{\alpha}-\alpha_0) \\ 0 = R_\alpha(\hat{r},\hat{\alpha}) \approx R_\alpha(r_0,\alpha_0) + \frac{\partial}{\partial r}R_\alpha(r_0,\alpha_0)(\hat{r}-r_0) + \frac{\partial}{\partial \alpha}R_\alpha(r_0,\alpha_0)(\hat{\alpha}-\alpha_0). \end{cases}$$

This gives us

$$\begin{aligned} \begin{pmatrix} \hat{r}-r_0 \\ \hat{\alpha}-\alpha_0 \end{pmatrix} &\approx -\begin{pmatrix} \frac{\partial}{\partial r}R_r(r_0,\alpha_0) & \frac{\partial}{\partial \alpha}R_r(r_0,\alpha_0) \\ \frac{\partial}{\partial r}R_\alpha(r_0,\alpha_0) & \frac{\partial}{\partial \alpha}R_\alpha(r_0,\alpha_0) \end{pmatrix}^{-1} \begin{pmatrix} R_r(r_0,\alpha_0) \\ R_\alpha(r_0,\alpha_0) \end{pmatrix} \\ &\approx -\begin{pmatrix} E_\varepsilon\frac{\partial}{\partial r}R_r(r_0,\alpha_0) & E_\varepsilon\frac{\partial}{\partial \alpha}R_r(r_0,\alpha_0) \\ E_\varepsilon\frac{\partial}{\partial r}R_\alpha(r_0,\alpha_0) & E_\varepsilon\frac{\partial}{\partial \alpha}R_\alpha(r_0,\alpha_0) \end{pmatrix}^{-1} \begin{pmatrix} R_r(r_0,\alpha_0) \\ R_\alpha(r_0,\alpha_0) \end{pmatrix}. \end{aligned}$$

Now recall that in case 1, $K_i = \sqrt{n}\mathrm{diag}[k_{i\nu}], \nu = 1,\ldots,n; k_{i\nu} = [\nu/2]^{-p_i}, \nu = 2,\ldots,n$, where $p_i \geq 0$, $i = 1,2$. Suppose $p_1 \leq p_2$, i.e. $p = p_1$, and let "$\sim$" mean that two terms have the same order as $n \to \infty$. It can be seen that

$$\begin{aligned} tr(A_{11}^r) &= tr(\frac{1}{r_0}M^{-1}K_1^2) \\ &\sim \sum_{i=1}^{n} \frac{ni^{-2p_1}}{(1/r_0)ni^{-2p_1} + r_0 ni^{-2p_2} + \alpha_0 i^{2m}} \\ &\sim \sum_{i=1}^{n} \frac{n}{n + i^{2m+2p_1}} \\ &= O(n^{1/(2m+2p)}). \end{aligned}$$

The last "=" above is from i) of lemma 3.1.2. So it is easy to see that $(1/n)tr(I_n - A_{11}^r) =$

$O(1)$, and similarly $(1/n)tr(I_n - A_{22}^r) = O(1)$.

It is also not hard to see that $E_\epsilon \frac{\partial}{\partial r} R_r(r_0, \alpha_0)$ is a *trace* of some matrix that can be written as $\sum_j tr((M^{-1}K^2)^{m_j})$, where m_j are positive integers and K means either K_1 or K_2. And by i) of lemma 3.1.2, all these *traces* in the sum have order $O(n^{1/(2m+2p)})$. So $E_\epsilon \frac{\partial}{\partial r} R_r(r_0, \alpha_0)$ has order $O(n^{1/(2m+2p)})$. Similarly $E_\epsilon \frac{\partial}{\partial \alpha} R_r(r_0, \alpha_0)$, $E_\epsilon \frac{\partial}{\partial r} R_\alpha(r_0, \alpha_0)$, and $E_\epsilon \frac{\partial}{\partial \alpha} R_\alpha(r_0, \alpha_0)$ all have order $O(n^{1/(2m+2p)})$. So we have

$$\begin{pmatrix} \hat{r} - r_0 \\ \hat{\alpha} - \alpha_0 \end{pmatrix} \approx \begin{pmatrix} O(n^{-1/(2m+2p)}) & O(n^{-1/(2m+2p)}) \\ O(n^{-1/(2m+2p)}) & O(n^{-1/(2m+2p)}) \end{pmatrix} \begin{pmatrix} R_r(r_0, \alpha_0) \\ R_\alpha(r_0, \alpha_0) \end{pmatrix}.$$

It is easy to see that the matrix $V^*(r, \alpha)$ in the variance term above is a 2×2 block matrix with each block being a diagonal matrix. The matrices $\frac{\partial}{\partial r} V^*(r, \alpha)$ and $\frac{\partial}{\partial \alpha} V^*(r, \alpha)$ also have the same structure. So

$$\varepsilon' \frac{\partial}{\partial r} V^*(r_0, \alpha_0)\varepsilon = \sum_{\nu=1}^{n} x_{r\nu}$$

$$\varepsilon' \frac{\partial}{\partial \alpha} V^*(r_0, \alpha_0)\varepsilon = \sum_{\nu=1}^{n} x_{\alpha\nu}$$

where $x_{r\nu}$ and $x_{\alpha\nu}$ are quadratic forms of $(\varepsilon_{1\nu}, \varepsilon_{2\nu})'$, $\nu = 1, \ldots, n$, and therefore $x_{r\nu}$, $\nu = 1, \ldots, n$ are independent and $x_{\alpha\nu}$, $\nu = 1, \ldots, n$ are independent.

By lengthy, but straightforward calculation, it can be shown that

$$E_\epsilon R_r(r_0, \alpha_0) = 0,$$

$$E_\epsilon R_\alpha(r_0, \alpha_0) = 0.$$

This means

$$R_r(r_0, \alpha_0) = \sum_{\nu=1}^{n} (x_{r\nu} - E_\epsilon x_{r\nu}),$$

and

$$R_\alpha(r_0, \alpha_0) = \sum_{\nu=1}^{n} (x_{\alpha\nu} - E_\epsilon x_{\alpha\nu}).$$

It is then easy to see that $\sum_{\nu=1}^{n} E_\epsilon(x_{r\nu} - E_\epsilon x_{r\nu})^2$ can be written as sum of finite number of sums of the forms as in i) of lemma 3.1.2 with $l_1 > 0$. So $\sum_{\nu=1}^{n} E_\epsilon(x_{r\nu} - E_\epsilon x_{r\nu})^2 =$

$O(n^{1/(2m+2p)})$. Similarly, $\sum_{\nu=1}^{n} E_\epsilon|x_{r\nu} - E_\epsilon x_{r\nu}|^3 = O(n^{1/(2m+2p)})$. This means that the sum

$$\sum_{\nu=1}^{n} n^{-1/(4m+4p)}(x_{r\nu} - E_\epsilon x_{r\nu})$$

satisfies the conditions of Liapounov's theorem (Chung (1974)), i.e.

$$\sum_{\nu=1}^{n} E_\epsilon[n^{-1/(4m+4p)}(x_{r\nu} - E_\epsilon x_{r\nu})]^2 = O(1),$$

and

$$\sum_{\nu=1}^{n} E_\epsilon|n^{-1/(4m+4p)}(x_{r\nu} - E_\epsilon x_{r\nu})|^3 = o(1).$$

So by Liapounov's theorem, we have that as $n \to \infty$

$$n^{-1/(4m+4p)} R_r(r_0, \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_1),$$

and similarly

$$n^{-1/(4m+4p)} R_\alpha(r_0, \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_2),$$

and

$$n^{-1/(4m+4p)}[R_r(r_0, \alpha_0) + R_\alpha(r_0, \alpha_0)] \xrightarrow{d} \mathcal{N}(0, c_3).$$

The result of the theorem follows. □

From this result, we see that the convergence rate of $\hat{r}_{rcv}$ is much slower than that of $\hat{r}_{gcv}$.

Case 2:

Similarly, we can get the convergence rates for RCV under case 2. Recall that in case 2, we require that $p_2 < m$ and p_2 be not too close to m, because in practice, noise should be rougher than the signal. So we only consider the following two practically interesting cases:

Theorem 3.3.2 For case 2, under the stochastic assumption, we have:

i). If $1 - 2(m - p_2) < 1/(2m + 2p_1)$, then as $n \to \infty$,

$$\begin{pmatrix} n^{1/(4m+4p_1)}(\hat{r}_{rcv} - r_0) \\ n^{1/(4m+4p_1)}(\hat{\alpha}_{rcv} - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.3.2}$$

ii). If $1-4(m-p_2) < 1/(2m+2p_1) \leq 1-2(m-p_2)$, then as $n \to \infty$,

$$\begin{pmatrix} n^{1-(1/(4m+4p_1)+2(m-p_2))}(\hat{r}_{rcv}-r_0) \\ n^{1-(1/(4m+4p_1)+2(m-p_2))}(\hat{\alpha}_{rcv}-\alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.3.3}$$

□

Note here that when $1-4(m-p_2) < 1/(2m+2p_1)$, we have $1/2 < 1/(4m+4p_1)+2(m-p_2)$, so $1-(1/(4m+4p_1)+2(m-p_2)) < 1/2$. We see that in case 2, the convergence rate of $\hat{r}_{rcv}$ is also slower than that of $\hat{r}_{gcv}$. Also notice if $1/(2m+2p_1) < 1-2(m-p_2)$, we have $1/(4m+4p_1) < 1-(1/(4m+4p_1)+2(m-p_2))$, that means when $1-4(m-p_2) < 1/(2m+2p_1) < 1-2(m-p_2)$ is true, $\hat{\alpha}_{rcv}$ has a faster convergence rate than $\hat{\alpha}_{gcv}$.

Case 3:

For case 3, we also require that both p_1 and p_2 be not too close to m. So we only consider the following case:

Theorem 3.3.3 For case 3, let $p = \max\{p_1, p_2\}$ and assume $m-p = 1/4$. Then under the stochastic assumption and as $n \to \infty$ we have:

$$\begin{pmatrix} (\sqrt{n}/\sqrt{\log n})(\hat{r}_{rcv}-r_0) \\ (\sqrt{n}/\sqrt{\log n})(\hat{\alpha}_{rcv}-\alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.3.4}$$

□

The convergence rate of r is slower than $\sqrt{n}$. Notice $\sqrt{n}/\sqrt{\log n} > \sqrt{\log n}$, so in this case, $\hat{\alpha}_{rcv}$ has a faster convergence rate than $\hat{\alpha}_{gcv}$.

From the above results, and by Slutsky's Theorem, we have:

Theorem 3.3.4 For all the cases considered above, as $n \to \infty$, we have

$$\hat{r}_{rcv} \xrightarrow{P} r_0, \qquad \hat{\alpha}_{rcv} \xrightarrow{P} \alpha_0. \tag{3.3.5}$$

□

We see that, the GCV-r estimator $\hat{r}_{gcv}$ has a faster convergence rate than RCV estimator $\hat{r}_{rcv}$. So in this sense, the GCV-r is better than the RCV in estimating r. Later on in Chapters 4 and 5, our simulation results confirm this.

Remarks:

So far, we have proved all our results about the GML-r, the GCV-r, and the RCV estimators under the assumption that the stochastic model (1.1.6) is true. A natural question is how robust they are? Suppose f is a smooth function from a reproducing kernel Hilbert space, how do these estimators perform under this situation? First of all, under this situation, there is no such a thing as the "true" α_0. So, what should be the "optimal" smoothing parameter α in this case? Our ultimate goal is to estimate f. Perhaps in this case, a more interesting question should be: how do these estimators do in estimating f? To answer these questions, we will give a weak convergence theorem for the GCV-r in Section 3.5. In Section 3.6, we will briefly discuss the robustness of the estimators. In Chapters 4 and 5, we will also try to answer some of the above questions by Monte Carlo studies. But first, let us look at some properties of the predictive mean-square error criterion.

3.4 Properties of the Predictive Mean-Square Error Criterion

Our ultimate goal is to estimate f. To measure how good our estimation of f is, we use the predictive mean-square error defined by

$$T_p(r,\alpha) = \frac{1}{n_1+n_2}\left\{\frac{\sigma_2}{\sigma_1}\|\hat{y}_1 - K_1 f\|^2 + \frac{\sigma_1}{\sigma_2}\|\hat{y}_2 - K_2 f\|^2\right\} \tag{3.4.1}$$

Let $(\hat{r}_t, \hat{\alpha}_t)$ denote the minimizer of $T_p(r,\alpha)$, and assume that the stochastic model (1.1.6) is true. Upon using the same method as in the previous section, we have:

Case 1:

Theorem 3.4.1 For case 1, letting $p = \min\{p_1, p_2\}$, as $n \to \infty$ we have:

$$\begin{pmatrix} n^{1/(4m+4p)}(\hat{r}_t - r_0) \\ n^{1/(4m+4p)}(\hat{\alpha}_t - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.4.2}$$

□

We see that in this case, $\hat{r}_t$ has the same convergence rate as that of $\hat{r}_{rcv}$. Both are slower than $\sqrt{n}$.

Case 2:

Theorem 3.4.2 For case 2, as $n \to \infty$ we have:

i). If $1-2(m-p_2) < 1/(2m+2p_1)$, then

$$\begin{pmatrix} n^{1/(4m+4p_1)}(\hat{r}_t - r_0) \\ n^{1/(4m+4p_1)}(\hat{\alpha}_t - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.4.3}$$

ii). If $1-4(m-p_2) < 1/(2m+2p_1) \leq 1-2(m-p_2)$, then

$$\begin{pmatrix} n^{1/(4m+4p_1)}(\hat{r}_t - r_0) \\ n^{1-(1/(4m+4p_1)+2(m-p_2))}(\hat{\alpha}_t - \alpha_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}). \tag{3.4.4}$$

□

We see that in this case, the convergence rate of $\hat{r}_t$ is not exactly the same as that of $\hat{r}_{rcv}$, it is also slower than $\sqrt{n}$.

Case 3:

For case 3, for the sake of comparison, we only consider one case:

Theorem 3.4.3 For case 3, letting $p = \max\{p_1, p_2\}$, and assume $m - p = 1/4$, then as $n \to \infty$ we have:

$$\begin{cases} \hat{r}_t - r_0 = O_p(1/\log n) \\ \hat{\alpha}_t - \alpha_0 = O_p(1/\sqrt{n}) \end{cases} \tag{3.4.5}$$

□

We see that in this case, the convergence rate of $\hat{r}_t$ is slower than $\sqrt{n}$.

From the above results, and by Slutsky's Theorem, we have:

Theorem 3.4.4 For all the cases considered above, as $n \to \infty$,

$$\hat{r}_t \xrightarrow{P} r_0, \qquad \hat{\alpha}_t \xrightarrow{P} \alpha_0. \tag{3.4.6}$$

□

The above results indicate that the predictive mean-square error is not a good criterion for estimating the weighting parameter r. Simulation results from Chapters 4 and 5 also show this.

3.5 A Weak Convergence Theorem for GCV-r

In this section, we will study properties of the GCV-r estimators when f is a smooth

function from a reproducing kernel Hilbert space. We want to see as $n \to \infty$, whether the GCV-r estimate is efficient in estimating f, i.e. we want to know whether the relative inefficiency with respect to the predictive mean-square error T_p defined by

$$I_p.\text{GCV-}r = \frac{T_p(\hat{r}_{gcv}, \hat{\alpha}_{gcv})}{T_p(\hat{r}_t, \hat{\alpha}_t)}$$

goes down to 1.

We will use a method similar to the method used in Craven and Wahba (1979), and show a weaker result under some assumptions. First, let $\lambda = \alpha/(2n)$, we shall use λ instead of α as the smoothing parameter in this section. We will show that as $n \to \infty$,

$$I_p^* = \frac{E_\epsilon T_p(\tilde{r}, \tilde{\lambda})}{E_\epsilon T_p(r^*, \lambda^*)} \downarrow 1 \tag{3.5.1}$$

where E_ϵ means taking the mathematical expectation with respect to ϵ, and $(\tilde{r}, \tilde{\lambda})$ is the minimizer of $E_\epsilon V(r, \lambda)$ and (r^*, λ^*) is the "optimal" (r, λ), i.e. the minimizer of $E_\epsilon T_p(r, \lambda)$.

As before, we will only show it for case 1. From the previous section, we see that under the stochastic model, $\hat{r}_t \xrightarrow{P} r_0$ and $\hat{r}_{gcv} \xrightarrow{P} r_0$, as $n \to \infty$ (see (3.4.6) and (3.2.7)). We will assume these still to be true for $\tilde{r}$ and r^* here. That is, we assume when n is sufficiently large, $\tilde{r} \approx r_0$ and $r^* \approx r_0$. Since $E_\epsilon T_p(r, \lambda)$ is a continuous function of r and λ, and also since $T_p(r, \lambda)$ is very insensitive to r, we have, as $n \to \infty$,

$$I_p^* \approx \frac{E_\epsilon T_p(r_0, \tilde{\lambda})}{E_\epsilon T_p(r_0, \lambda^*)}$$

Below, we will fix $r = r_0$, and use r for r_0.

Now, let

$$A = \begin{pmatrix} (1/r)M^{-1}K_1^2 & rM^{-1}K_1K_2 \\ (1/r)M^{-1}K_1K_2 & rM^{-1}K_2^2 \end{pmatrix}$$

where as before, $M = (1/r)K_1^2 + rK_2^2 + 2n\lambda\Sigma^{-1}$. Then $\hat{y} = Ay = K\hat{f}$, where $K = (K_1', K_2')'$. Also recall that

$$I^{-1}(r) = \begin{pmatrix} (1/\sqrt{r})I_n & 0 \\ 0 & \sqrt{r}I_n \end{pmatrix},$$

and $r = \sigma_1/\sigma_2$, $\theta = \sigma_1\sigma_2$. So we can write $T_p(r,\lambda)$ as

$$T_p(r,\lambda) = \frac{1}{2n}\left\{\frac{\sigma_2}{\sigma_1}\|\hat{y}_1 - K_1 f\|^2 + \frac{\sigma_1}{\sigma_2}\|\hat{y}_2 - K_2 f\|^2\right\}$$

$$= \frac{1}{2n}\|I^{-1}(r)(Kf - K\hat{f})\|^2$$

and

$$\begin{aligned} E_\varepsilon T_p(r,\lambda) &= \frac{1}{2n}E_\varepsilon\|I^{-1}(r)(Kf - A(Kf+\varepsilon))\|^2 \\ &= \frac{1}{2n}\{\|I^{-1}(r)(I-A)Kf\|^2 + E_\varepsilon\|I^{-1}(r)A\varepsilon\|^2\} \\ &= \frac{1}{2n}\|I^{-1}(r)(I-A)Kf\|^2 + \frac{\theta}{2n}tr(I^{-1}(r)AI^2(r)A'I^{-1}(r)) \\ &= b_t^2(r,\lambda) + \theta\mu_{t2}(r,\lambda) \end{aligned}$$

where

$$b_t^2(r,\lambda) = \frac{1}{2n}\|I^{-1}(r)(I-A)Kf\|^2$$

and (since $I^{-1}(r)AI(r) = A^r$)

$$\mu_{t2}(r,\lambda) = \frac{1}{2n}tr(A^{r2})$$

are the bias and variance terms for $E_\varepsilon T_p(r,\lambda)$, respectively.

Since $I_n = M^{-1}M = M^{-1}((1/r)K_1^2 + rK_2^2 + 2n\lambda\Sigma^{-1})$, we have

$$I_{2n} - A = \begin{pmatrix} rM^{-1}K_2^2 + 2n\lambda M^{-1}\Sigma^{-1} & -rM^{-1}K_1K_2 \\ -(1/r)M^{-1}K_1K_2 & (1/r)M^{-1}K_1^2 + 2n\lambda M^{-1}\Sigma^{-1} \end{pmatrix}.$$

So

$$I^{-1}(r)(I-A)Kf = \begin{pmatrix} (1/\sqrt{r})2n\lambda M^{-1}\Sigma^{-1}K_1 f \\ \sqrt{r}2n\lambda M^{-1}\Sigma^{-1}K_2 f \end{pmatrix}$$

Now it is easy to see that for fixed r, $b_t^2(r,\lambda)$ is a monotone increasing function of λ, with $\frac{\partial}{\partial\lambda}b_t^2|_{\lambda=0} = 0$, and $\mu_{t2}(r,\lambda)$ is a monotone decreasing function of λ, with $\frac{\partial}{\partial\lambda}\mu_{t2}|_{\lambda=0} < 0$.

If $f = (f_0, 0, \ldots, 0)'$, then because the first diagonal entry of Σ^{-1} is zero (see Chapter 1), $b_t^2(r,\lambda) = 0$, so $E_\varepsilon T_p(r,\lambda)$ has a minimizer at $\lambda^* = \infty$.

If $f \neq (f_0, 0, \ldots, 0)'$, then $E_\varepsilon T_p(r,\lambda)$ has a minimizer at $\lambda^* > 0$.

Following the same argument as in Lemma 4.1 of Craven and Wahba (1979), it can be shown that $b_t^2(r,\lambda) \leq O(\lambda)$. It is also easy to see that if $f \neq (f_0, 0, \ldots, 0)'$, and λ is bounded away from zero, then $b_t^2(r,\lambda)$ is also bounded away from zero (as $n \to \infty$). So, $b_t^2(r,\lambda) \to 0$ $(n \to \infty)$ if and only if $\lambda \to 0$ $(n \to \infty)$. In order to let $E_\epsilon T_p(r,\lambda) \to 0$ $(n \to \infty)$, we should let $b_t^2(r,\lambda) \to 0$ and $\mu_{t2}(r,\lambda) \to 0$ $(n \to \infty)$. That is, when $f \neq (f_0, 0, \ldots, 0)'$, we should have $\lambda^* \to 0$.

Now,

$$E_\epsilon V(r,\lambda) = [b^2(r,\lambda) + \theta(\frac{1}{2}(\frac{1}{r}+r) - 2\mu_1(r,\lambda) + \mu_2(r,\lambda))]$$

$$/(\frac{1}{2}(\frac{1}{r}+r) - \mu_1(r,\lambda))^2$$

$$= \frac{1}{((1/2)(1/r+r))^2}\{(\frac{1}{2}(\frac{1}{r}+r))^2 b^2/(\frac{1}{2}(\frac{1}{r}+r) - \mu_1)^2 + \frac{1}{2}(\frac{1}{r}+r)\theta$$

$$+[-\frac{1}{2}(\frac{1}{r}+r)\theta\mu_1^2 + (\frac{1}{2}(\frac{1}{r}+r))^2\theta\mu_2]/(\frac{1}{2}(\frac{1}{r}+r) - \mu_1)^2\}$$

where

$$b^2(r,\lambda) = \frac{1}{2n}\|I^{-1}(r)(I - A^r)I^{-1}(r)Kf\|^2,$$

$$\mu_1(r,\lambda) = \frac{1}{2n}tr(I^{-1}(r)A^r I^{-1}(r)),$$

and

$$\mu_2(r,\lambda) = \frac{1}{2n}tr(I^{-1}(r)A^{r2}I^{-1}(r)).$$

Similarly we have

$$I^{-1}(r)(I - A^r)I^{-1}(r)Kf = \begin{pmatrix} (1/r)2n\lambda M^{-1}\Sigma^{-1}K_1 f \\ r2n\lambda M^{-1}\Sigma^{-1}K_2 f \end{pmatrix}$$

and it can be shown that μ_1 and μ_2 are monotone decreasing functions of λ.

By Cauchy-Schwarz inequality, we have $\mu_1^2 \leq (1/2)(1/r+r)\mu_2$, where "=" is possible only when $K_1 = 0$ or $K_2 = 0$. Also notice that $0 < \mu_1(r,\lambda) < (1/2)(1/r+r)$. So for our problem where $K_1 \neq 0$ and $K_2 \neq 0$, we have,

$$[-\frac{1}{2}(\frac{1}{r}+r)\theta\mu_1^2 + (\frac{1}{2}(\frac{1}{r}+r))^2\theta\mu_2]/(\frac{1}{2}(\frac{1}{r}+r) - \mu_1)^2 > 0.$$

And for any fixed n, it reaches its minimum at $\lambda = \infty$.

If $f = (f_0, 0, \ldots, 0)'$, as before, it is easy to see $b^2(r, \lambda) = 0$. So in this case, $\tilde{\lambda} = \infty$. Therefore $I_p^* = 1$.

Now for the general case when $f \neq (f_0, 0, \ldots, 0)'$, first notice that for b^2 we also have $b^2 \leq O(\lambda)$ and $b^2 \to 0$ if and only if $\lambda \to 0$ $(n \to \infty)$. So we have $\tilde{\lambda} \to 0$.

Also we have

$$\{E_\varepsilon T_p - C[(\frac{1}{2}(\frac{1}{r}+r))^2 E_\varepsilon V - \frac{1}{2}(\frac{1}{r}+r)\theta]\}/E_\varepsilon T_p$$

$$= 1 - [C(\frac{1}{2}(\frac{1}{r}+r))^2(b^2+\theta\mu_2) - C\frac{1}{2}(\frac{1}{r}+r)\theta\mu_1^2]$$

$$/[(b_t^2+\theta\mu_{t2})(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2]$$

$$= 1 - [C(b^2+\theta\mu_2)/(b_t^2+\theta\mu_{t2})][(\frac{1}{2}(\frac{1}{r}+r))^2/(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2]$$

$$+C\frac{1}{2}(\frac{1}{r}+r)\theta\mu_1^2/[(b_t^2+\theta\mu_{t2})(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2]$$

$$= [-(\frac{1}{r}+r)\mu_1 + \mu_1^2 + (1 - C(b^2+\theta\mu_2)/(b_t^2+\theta\mu_{t2}))(\frac{1}{2}(\frac{1}{r}+r))^2]$$

$$/(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2$$

$$+[C\frac{1}{2}(\frac{1}{r}+r)\theta/(b_t^2+\theta\mu_{t2})][\mu_1^2/(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2]$$

where C is a constant to be specified later.

Now, using a similar method that has been used in Craven and Wahba (1979), and letting $p = \min\{p_1, p_2\}$ in case 1, we have as $n \to \infty$:

$$\mu_1 = O(1/(n\lambda^{1/(2m+2p)})),$$

$$\mu_2 = O(1/(n\lambda^{1/(2m+2p)})),$$

and

$$\mu_{t2} = O(1/(n\lambda^{1/(2m+2p)})).$$

So, we must have that as $n \to \infty$, $\lambda^* \to 0$, $\tilde{\lambda} \to 0$, $n\lambda^{*1/(2m+2p)} \to \infty$ and $n\tilde{\lambda}^{1/(2m+2p)} \to \infty$.

By the definitions of b_t^2, b^2, μ_{t2} and μ_2, it can be shown that

$$\lim_{n\to\infty} \frac{b^2(r,\lambda^*)+\theta\mu_2(r,\lambda^*)}{b_t^2(r,\lambda^*)+\theta\mu_{t2}(r,\lambda^*)} = \lim_{n\to\infty} \frac{b^2(r,\tilde{\lambda})+\theta\mu_2(r,\tilde{\lambda})}{b_t^2(r,\tilde{\lambda})+\theta\mu_{t2}(r,\tilde{\lambda})}.$$

Letting

$$C^{-1} = \lim_{n\to\infty} \frac{b^2(r,\lambda^*)+\theta\mu_2(r,\lambda^*)}{b_t^2(r,\lambda^*)+\theta\mu_{t2}(r,\lambda^*)},$$

then $0 < C < \infty$, and we have that as $n \to \infty$

$$(1 - \frac{C(b^2+\theta\mu_2)}{b_t^2+\theta\mu_{t2}})(\frac{1}{2}(\frac{1}{r}+r))^2/(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2 = o(1)$$

for $\lambda = \lambda^*$, or $\lambda = \tilde{\lambda}$.

Now we have

$$\begin{aligned}
&|E_\epsilon T_p - C[(\frac{1}{2}(\frac{1}{r}+r))^2 E_\epsilon V - \frac{1}{2}(\frac{1}{r}+r)\theta]|/E_\epsilon T_p \\
&\leq o(1) + [(\frac{1}{r}+r)\mu_1 + C\frac{1}{2}(\frac{1}{r}+r)\mu_1^2/\mu_{t2}]/(\frac{1}{2}(\frac{1}{r}+r)-\mu_1)^2 \\
&\equiv h(r,\lambda)
\end{aligned}$$

and it is clear that

$$\lim_{n\to\infty} h(r,\lambda^*) = 0,$$

$$\lim_{n\to\infty} h(r,\tilde{\lambda}) = 0.$$

And we have

$$-h(r,\lambda) \leq 1 - C[(\frac{1}{2}(\frac{1}{r}+r))^2 E_\epsilon V - \frac{1}{2}(\frac{1}{r}+r)\theta]/E_\epsilon T_p \leq h(r,\lambda).$$

So

$$E_\epsilon T_p(1-h) \leq C[(\frac{1}{2}(\frac{1}{r}+r))^2 E_\epsilon V - \frac{1}{2}(\frac{1}{r}+r)\theta] \leq E_\epsilon T_p(1+h)$$

and then

$$\begin{aligned}
&E_\epsilon T_p(r,\tilde{\lambda})(1-h(r,\tilde{\lambda})) \\
&\leq C((\frac{1}{2}(\frac{1}{r}+r))^2 E_\epsilon V(r,\tilde{\lambda}) - \frac{1}{2}(\frac{1}{r}+r)\theta)
\end{aligned}$$

$$\leq C((\frac{1}{2}(\frac{1}{r}+r))^2 E_\epsilon V(r,\lambda^*) - \frac{1}{2}(\frac{1}{r}+r)\theta)$$
$$\leq E_\epsilon T_p(r,\lambda^*)(1+h(r,\lambda_*)).$$

So,

$$I_p^* \approx \frac{E_\epsilon T_p(r,\tilde{\lambda})}{E_\epsilon T_p(r,\lambda^*)} \leq \frac{1+h(r,\lambda^*)}{1-h(r,\tilde{\lambda})}.$$

and therefore

$$I_p^* \downarrow 1.$$

□

Similarly, it can be shown that for case 2 and case 3, the weak convergence theorem is true.

From the above proof, we see that for case 1,

$$E_\epsilon T_p(r,\lambda) \leq O(\lambda) + O(1/(n\lambda^{1/(2m+2p)})).$$

If we let

$$\lambda = O(1/n^{(2m+2p)/(2m+2p+1)}),$$

then

$$E_\epsilon T_p(r,\lambda^*) \leq O(1/n^{(2m+2p)/(2m+2p+1)}).$$

If f is very smooth and we have

$$b_t^2(r,\lambda) \leq O(\lambda^q)$$

for some $q \in (1,2]$, then by taking

$$\lambda = O(1/n^{(2m+2p)/((2m+2p)q+1)})$$

we get

$$E_\epsilon T_p(r,\lambda^*) \leq O(1/n^{(2m+2p)q/((2m+2p)q+1)}).$$

Similarly, we can get the convergence rates for case 2 and case 3.

3.6 Comparison of GML-r and GCV-r for Estimating f

From the results in the previous sections, we see that under some conditions, as n increases, $\hat{r}_{gml}$ and $\hat{r}_{gcv}$ will be very close to the true r_0. From the Monte Carlo results in Chapters 4 and 5, we see that this is also more or less true for various misspecification cases. Now, suppose n is large, then both $\hat{r}_{gml}$ and $\hat{r}_{gcv}$ will be very close to the true r_0. The problem of finding the minimizers of the GML-r function and the GCV-r function will then roughly be equivalent to the problem of finding the minimizers of the GML-r function and the GCV-r function as functions of α alone with r fixed at the true r_0. This problem is similar to the problem of estimating the smoothing parameter by the GML and the GCV for one data source problem, which has been studied in Wahba (1985a). It is shown in Wahba (1985a) that if f is from a reproducing kernel Hilbert space, and satisfies some extra smooth condition, then the GML estimator undersmooths relative to the GCV estimator and the predictive mean-square error using the GML estimate goes to zero at a slower rate than that using the GCV estimate. If f is "rough", then the GML and the GCV estimators have asymptotically similar behavior. Results from Stein (1989) and Stein (1990) indicate that when the stochastic model is true, GML estimator is better than GCV estimator, however, when the stochastic model is not true, GML estimator is not robust. Similar results should be true for our problem with two data sources. In our Monte Carlo studies in Chapters 4 and 5, the GCV-r estimator appears to be more robust than the GML-r estimator.

Chapter 4

Simulation Studies for Functions on the Circle

In this chapter, we study the performance of GML-r, GCV-r and RCV estimates under a simplified setting, where the underlying unknown function f is a function on the circle. We assume $n = n_1 = n_2$.

We should keep in mind that there are two goals in our study: we are interested in estimating *both* the weighting parameter r and the unknown function f. So, in section 4.2, we will see how GML-r, GCV-r and RCV perform in estimating f. In section 4.3, we will see how they do in estimating r, and finally in section 4.4, we will see how they do in estimating α. We study all these under the stochastic model as well as under various misspecification cases.

4.1 Experiment Set-Up

We choose the following model

$$\begin{cases} y_1 = K_1 f + \varepsilon_1 \\ y_2 = K_2 f + \varepsilon_2 \end{cases}$$

where K_1 and K_2 are of the form of case 1 (see section 1.2), with $p_1 = 0$ and $p_1 = 0.6$, so $p = \min\{p_1, p_2\} = 0$. Here y_1 could be direct observations and y_2 could be satellite data. We take $\sigma_1 = 0.5$, $\sigma_2 = 0.4$, so the true $r = 1.25$. If there is a stochastic model, we take

$b = 1$, so the true $\alpha = 0.2$.

We first use the stochastic model (1.1.6) with $m = 2$ to generate f, and see how these estimates do when the stochastic model is correctly specified. We choose $n = 400$ and $n = 800$, and do 100 replications. For each replicate, f is a different realization from (1.1.6), and so are the noises.

We then look at the following five different misspecification cases that are likely to occur in real life, especially in meteorology. These cases are:

Misspecification 1: The true f is from stochastic model (1.1.6) with $m = 1$, but f is misspecified as from (1.1.6) with $m = 2$.

Misspecification 2: The true f is from stochastic model (1.1.6) with $m = 2$, but f is misspecified as from (1.1.6) with $m = 1$.

Misspecification 3: The true f is a deterministic function with the Fourier-Bessel coefficients $f = (1, 1, -1, 1, 0, \ldots, 0)'$, but f is misspecified as from (1.1.6) with $m = 2$. In this case, for each replicate, f is the same.

Misspecification 4: In meteorological data, the energy spectra often have such a pattern that the first few eigenvalues (of Σ) will increase, and reach a peak, and then decrease at a certain rate, see, for example, Wahba (1982b), Stanford (1979), and Figure 5.1.3. So, we consider the following misspecification case: the true peak is at the 2nd pair of the eigenvalues (remember for functions on the circle, the eigenvalues are of multiplicity 2), but we misspecify the peak to be at the 10th pair.

Misspecification 5: The true peak is at the 10th pair, but we misspecify it to be at the 2nd pair.

For each of the above five misspecification cases, we choose $n = 400$ and $n = 800$, and do 100 replications.

Now, let us see how these estimates perform.

4.2 Estimation of f

For each case, we run 100 replicates. In each replicate, we use a grid search to obtain $\hat{r}$'s and $\hat{\alpha}$'s from GML-r, GCV-r, RCV and the predictive mean-square error T_p. Then we get the $\hat{f}$. We then calculate I_p.GML-r $= T_p(\hat{r}_{gml}, \hat{\alpha}_{gml})/T_p(\hat{r}_t, \hat{\alpha}_t)$, and similarly, I_p.GCV-r, and I_p.RCV. First, let us compare the GML-r and the GCV-r. We summarize the results in Figure 4.2.1.

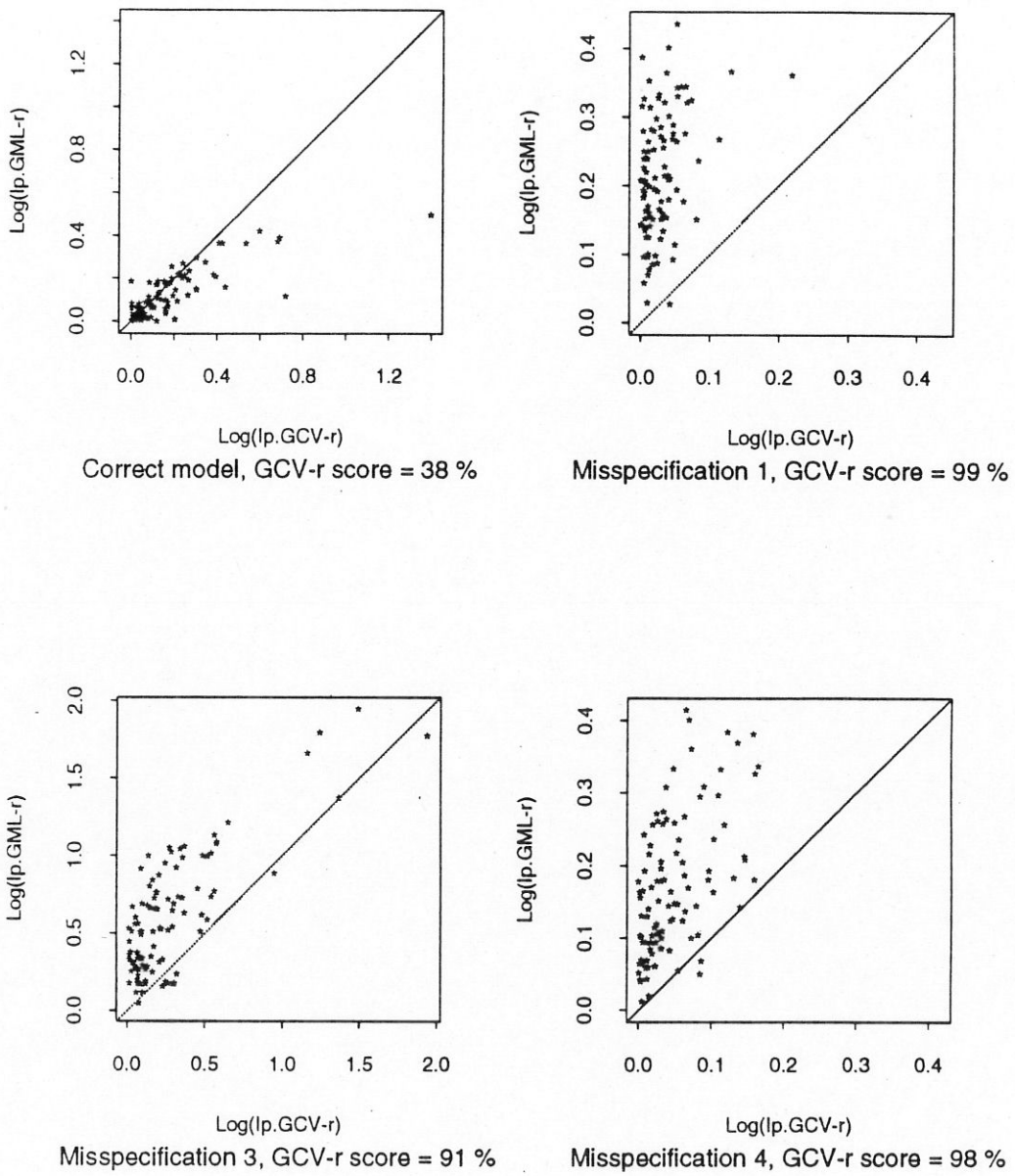


Figure 4.2.1: Log(I_p.GML-r) vs. Log(I_p.GCV-r). n=800 with 100 replicates.

In Figure 4.2.1, we only include plots for the correct model and for misspecifications 1, 3 and 4. For misspecifications 2 and 5, we have very similar plots as misspecification 1. We also only include results for $n = 800$, because for $n = 400$, the results are similar. We see from Figure 4.2.1 that when the stochastic model is true, GML-r does better than GCV-r in estimating the function f. However, when the model is misspecified, GCV-r is superior to GML-r. When reading the plots, please notice that they are on Log scale.

For Log(I_p.GML-r) vs. Log(I_p.RCV), the results are very similar to those for Log(I_p.GML-r) vs. Log(I_p.GCV-r). So when the stochastic model is not true, RCV is also better than GML-r in estimating the unknown function f.

It is of interest to compare GCV-r and RCV, and we summarize the results in Figure 4.2.2.

In Figure 4.2.2, we only include plots for the correct model and for misspecifications 1, 3 and 4. For misspecifications 2 and 5, we have similar plots as that for misspecification 1. We also only include results for $n = 800$, because for $n = 400$, the results are similar. From Figure 4.2.2, we see that GCV-r is slightly better than RCV in estimating f when the model is correct and for misspecifications 1 and 3. For misspecification 4, RCV is slightly better. In misspecification 4, the true f is smoother than specified. However, GCV-r undersmooths a little bit for this particular kind of misspecification, while RCV gives an estimate of the smoothing parameter that is closer to the "optimal" smoothing parameter (see Figure 4.4.4 in Section 4.4). Although GCV-r does much better in estimating the weighting parameter r for misspecification 4 (see results in Section 4.3), RCV is slightly better in estimating the function f for this misspecification. Also notice that Figure 4.2.2 is in Log scale.

In Chapter 3, we show that asymptotically, GCV-r does a better job than RCV in estimating the weighting parameter r, and in the next section, we will present some simulation results that confirm this theoretical result. However, since the predictive mean-square error function is fairly insensitive to r, and also since in some cases, RCV does a slightly better job estimating the smoothing parameter α (see Theorem 3.3.2 and Theorem 3.3.3 compared with Theorem 3.2.2 and Theorem 3.2.3, also see results in Section 4.4), so, RCV sometimes may do slightly better in estimating f than GCV-r in terms of the predictive mean-square error criterion.

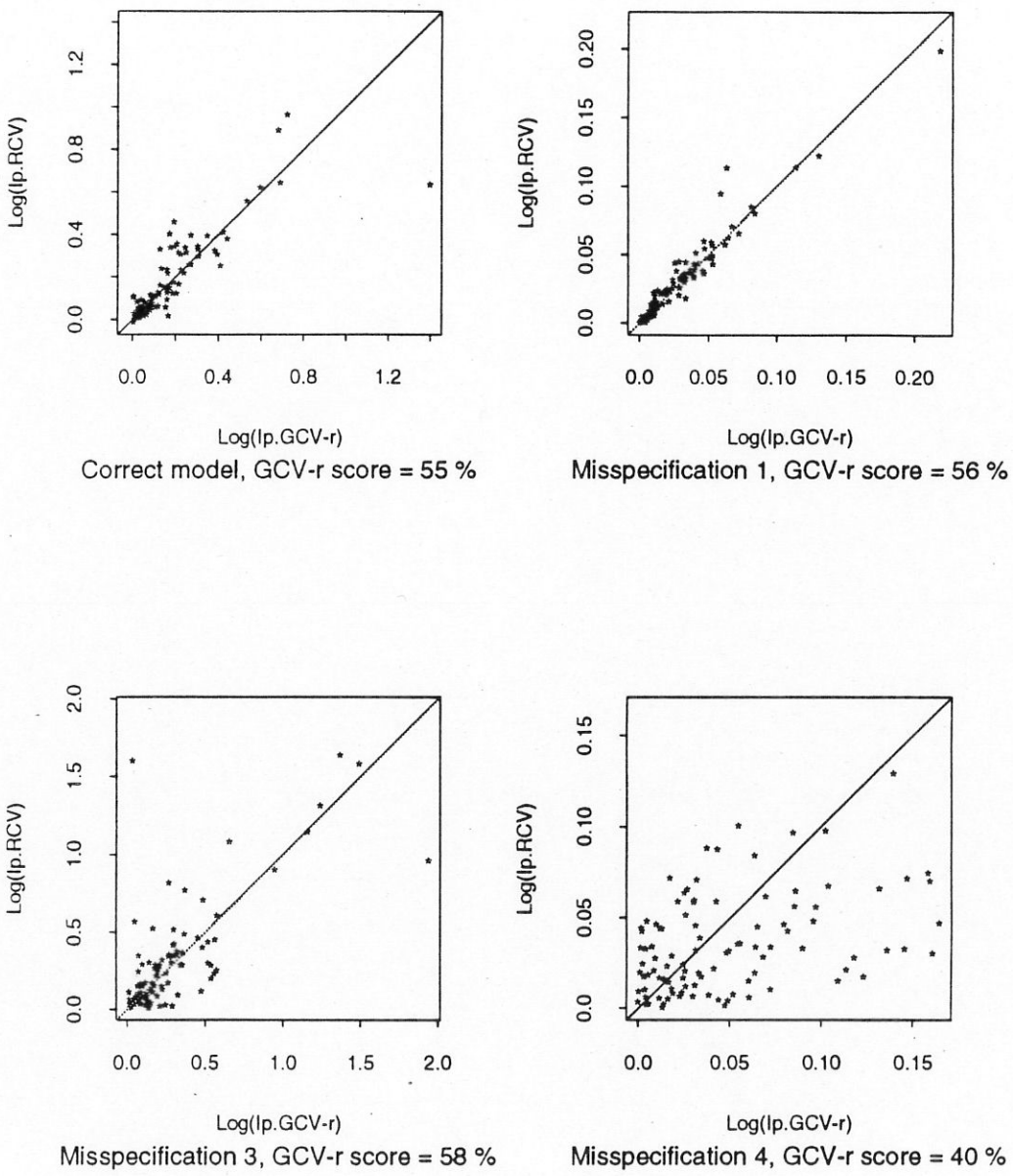


Figure 4.2.2: Log(I_p.RCV) vs. Log(I_p.GCV-r). n=800 with 100 replicates.

4.3 Estimation of r

In this section we will compare the three methods in estimating the weighting parameter r under different cases.

In the simulation study, we used a grid search to find $(\hat{r}, \hat{\alpha})$. The range for search on r was set to be very large to make sure we find the right $\hat{r}$. We do the plots in Log scale. The true $r_0 = 1.25$, so $\text{Log}(r_0)$=0.2231.

First, let us compare GML-r and GCV-r. We summarize the results in Figure 4.3.1.

In Figure 4.3.1, we only include the correct model and misspecifications 1, 3 and 4. For misspecifications 2 and 5, the plots are actually quite similar to that of misspecification 3. From Figure 4.3.1, we see that when the stochastic model is right, GCV-r and GML-r do equally well in estimating r. For misspecification 1, where the true f is rougher than specified, both GCV-r and GML-r estimators are biased, however GCV-r is better than GML-r. For misspecifications 3 and 4, where the true f is smoother than specified, GCV-r and GML-r do equally well.

Now, let us compare GCV-r and RCV. We summarize the results in Figure 4.3.2.

For misspecifications 2 and 5, the plots have the same pattern as those appear in Figure 4.3.2. From Figure 4.3.2, we see that GCV-r is always better than RCV in estimating r. This is in agreement with our theoretical results in Chapter 3 that $\hat{r}_{gcv}$ has a faster convergence rate than $\hat{r}_{rcv}$.

We know that $\hat{r}_t$ has very slow convergence rates (see section 3.4), that means T_p is very insensitive to r. To see how $\hat{r}_t$ performs, let us compare $\hat{r}_{rcv}$ and $\hat{r}_t$. We do this in Figure 4.3.3.

For misspecifications 2 and 5, the plots look similar to that of misspecification 1. From Figure 4.3.3, we see that $\hat{r}_t$ behaves poorly, and in some cases worse than $\hat{r}_{rcv}$. So the predictive mean-square error T_p is not a good criterion for estimating the weighting parameter r.

Looking at histograms of these estimates side by side can also give us some idea on the performance of the estimators. We put these histograms in Figures 4.3.4, 4.3.5, 4.3.6 and 4.3.7. From these histograms, we see clearly that GCV-r and GML-r are better than RCV in estimating the weighting parameter r. We also see that the predictive mean-square error T_p behaves poorly for estimating the weighting parameter r. All these findings are in good agreement with our theoretical results in Chapter 3.

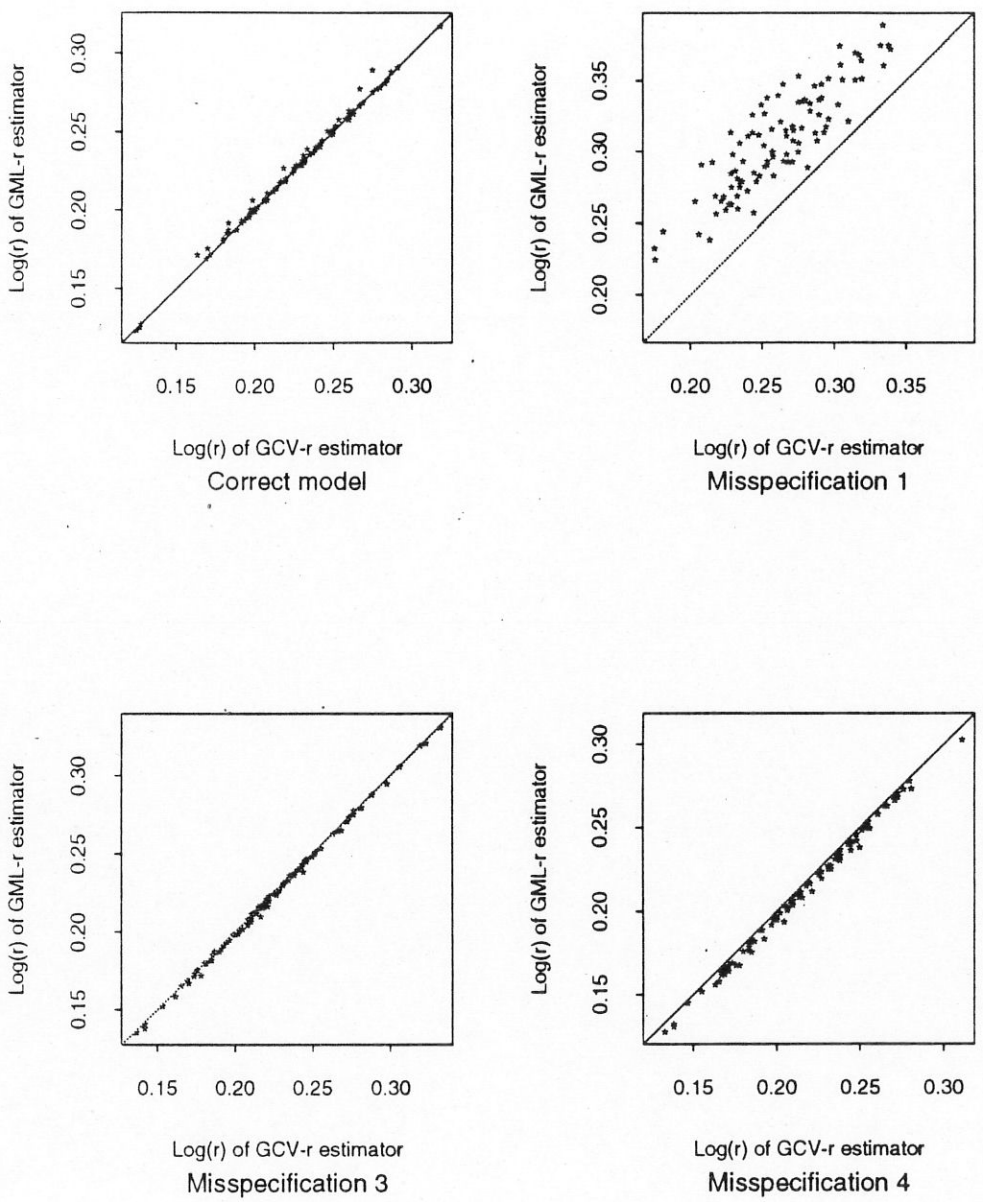


Figure 4.3.1: Log($\hat{r}_{gml}$) vs.Log($\hat{r}_{gcv}$).True Log(r)=0.2231. n=800. 100 replicates.

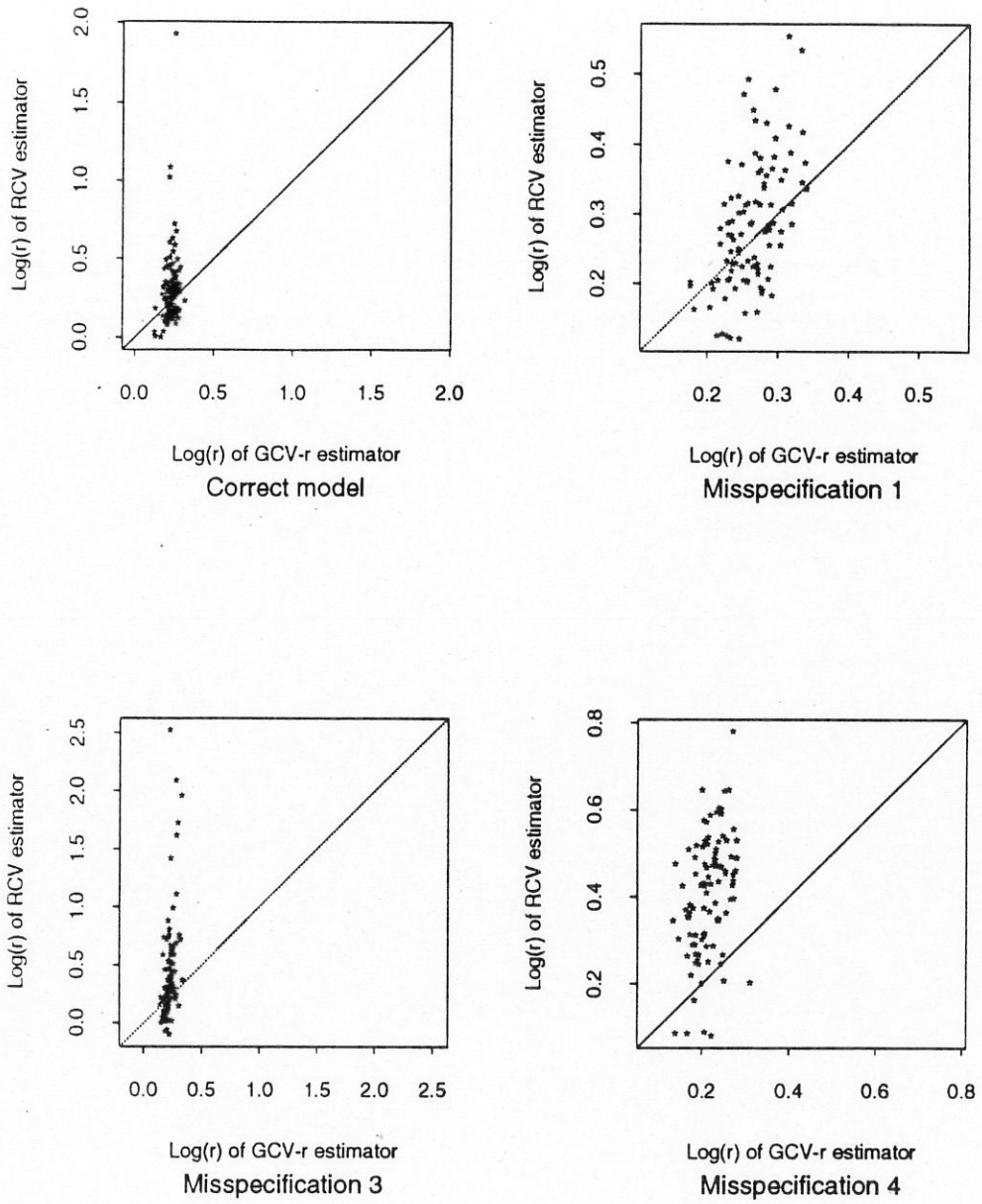


Figure 4.3.2: Log($\hat{r}_{rcv}$) vs. Log($\hat{r}_{gcv}$). True Log(r)=0.2231. n=800. 100 replicates.

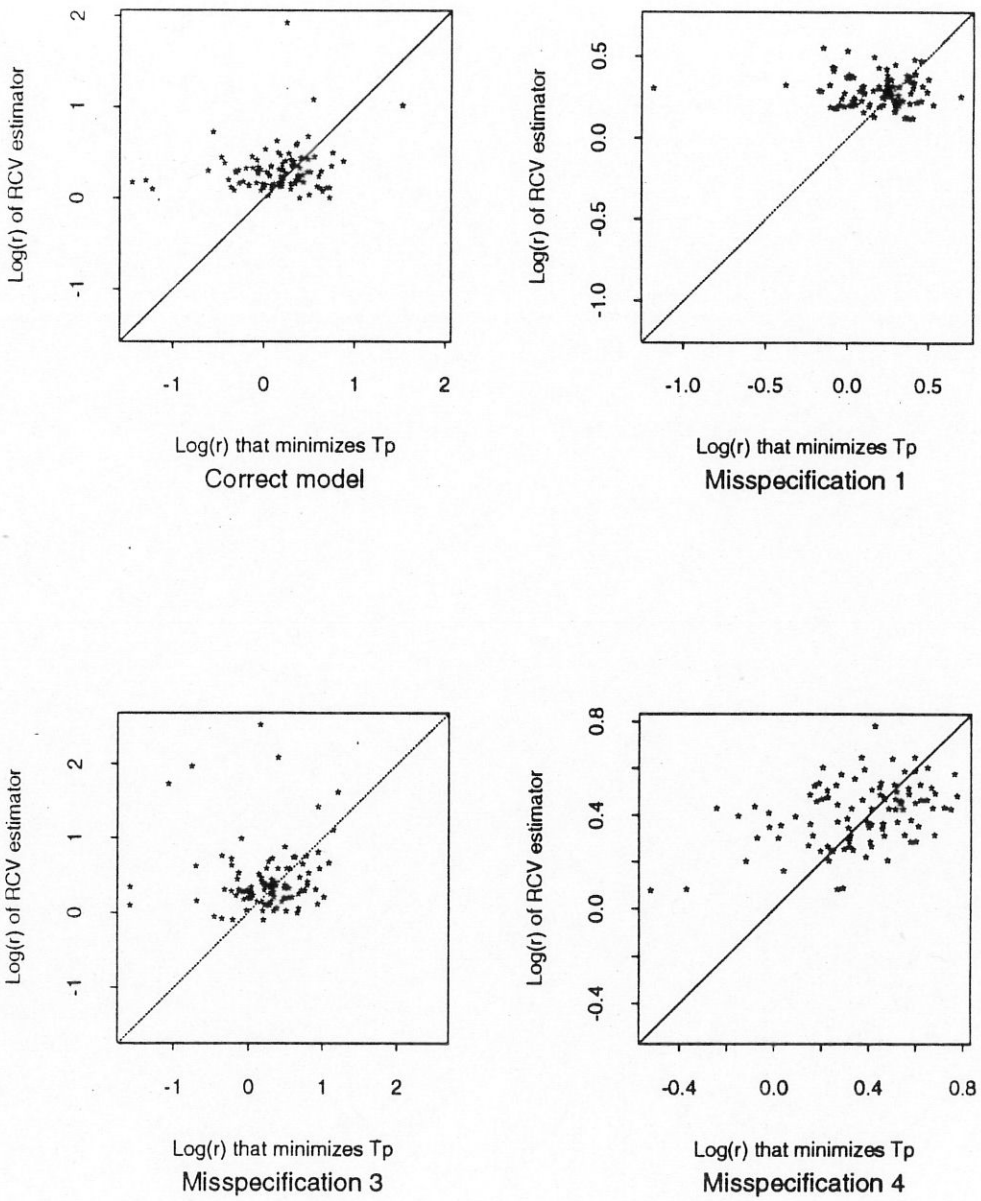


Figure 4.3.3: Log($\hat{r}_{rcv}$) vs. Log($\hat{r}_t$). True Log(r)=0.2231. n=800. 100 replicates.

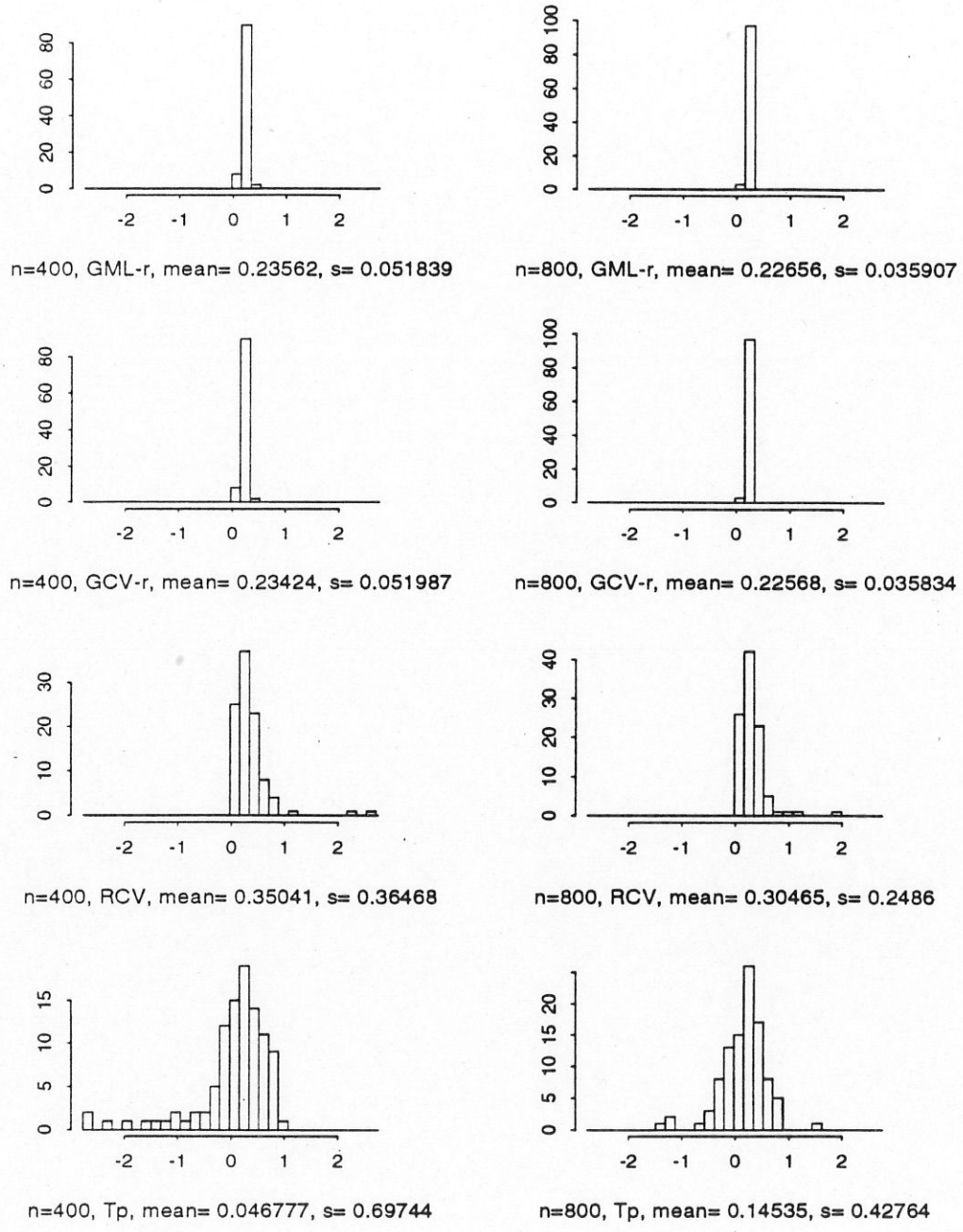


Figure 4.3.4: Distributions of $\mathrm{Log}(\hat{r})$ with 100 replicates when the model is correctly specified. The true $\mathrm{Log}(r)$=0.2231.

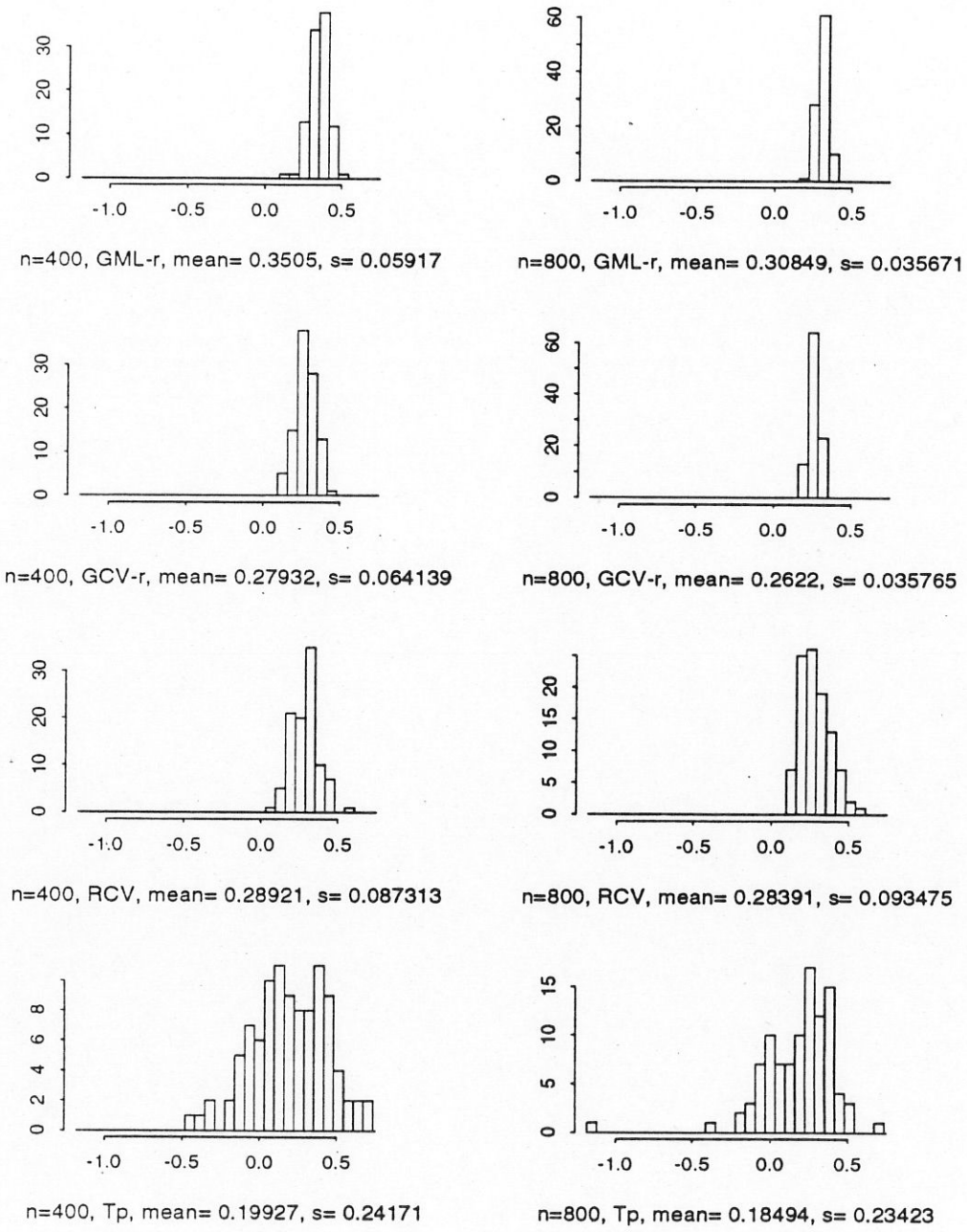


Figure 4.3.5: Distributions of $\mathrm{Log}(\hat{r})$ with 100 replicates for misspecification 1. The true $\mathrm{Log}(r)$=0.2231.

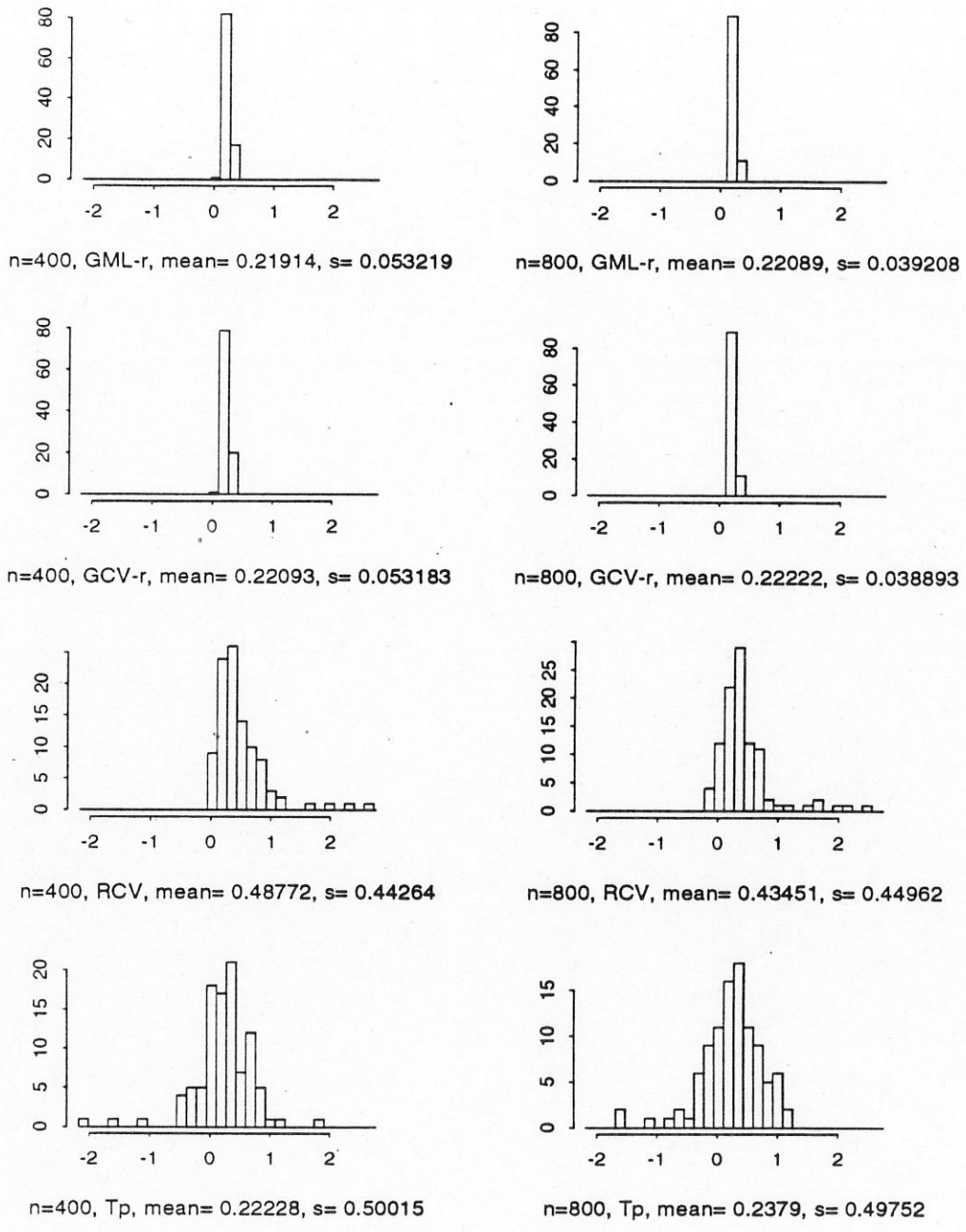


Figure 4.3.6: Distributions of $\mathrm{Log}(\hat{r})$ with 100 replicates for misspecification 3. The true $\mathrm{Log}(r)$=0.2231.

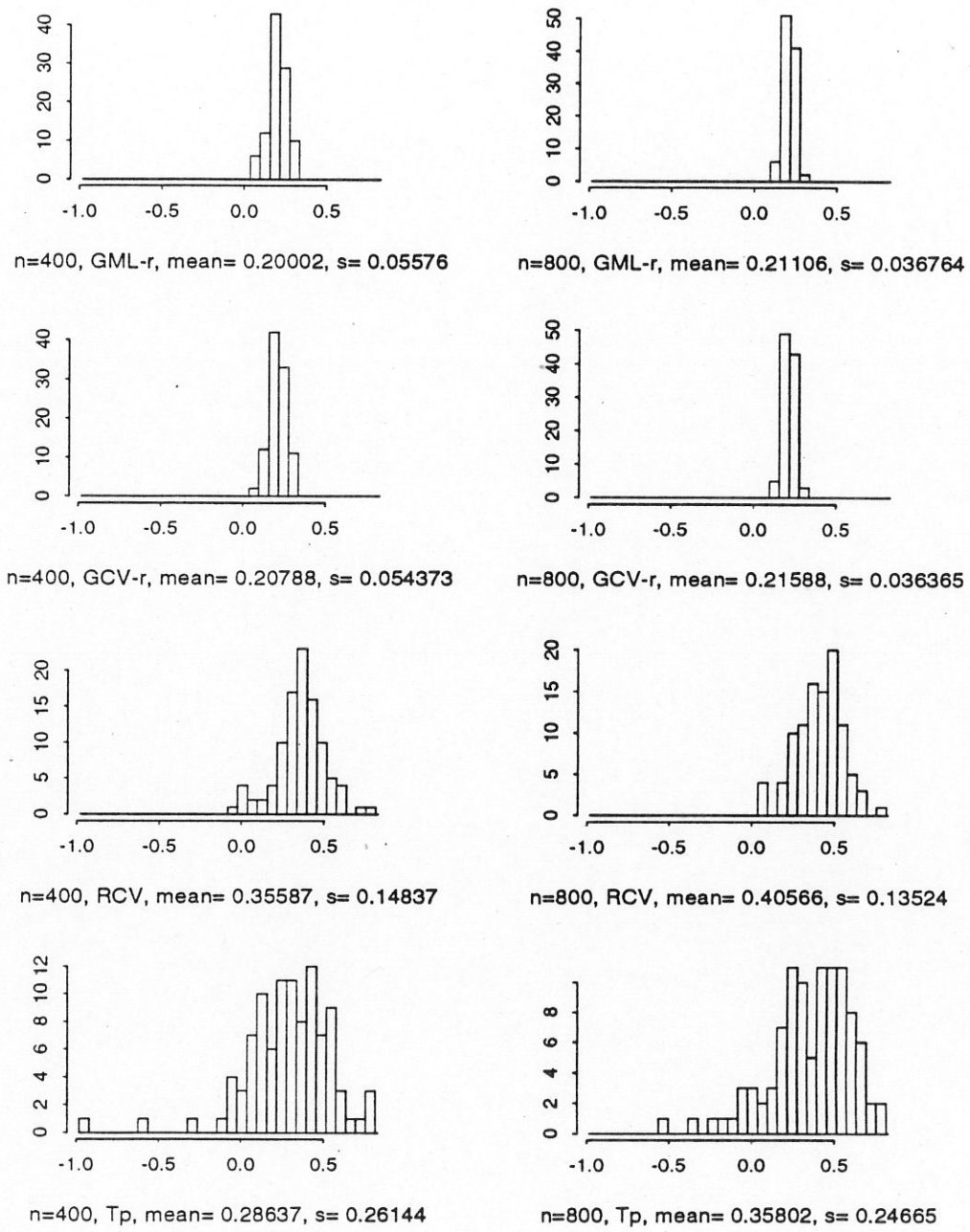


Figure 4.3.7: Distributions of Log($\hat{r}$) with 100 replicates for misspecification 4. The true Log(r)=0.2231.

Finally, let us look at the histograms of $\hat{r}_{gml}$ and $\hat{r}_{gcv}$ in their original scale. We put these histograms in Figure 4.3.8, 4.3.9, 4.3.10 and 4.3.11. From these plots, we see that

i). When the model is correctly specified, the sample means for both estimators are close to the true value of the weighting parameter r, the sample variances for $n = 800$ for both estimators are about half as large as those for $n = 400$. This confirms the theoretical result that the convergence rate for $\hat{r}_{gml}$ and $\hat{r}_{gcv}$ is $1/\sqrt{n}$.

ii). The magnitudes of the sample variance of $\hat{r}_{gcv}$ is almost the same as those of $\hat{r}_{gml}$ when the model is correctly specified as well as when it is misspecified. This shows that even when the stochastic model is right, GCV-r does equally well as GML-r in estimating r. So, estimating r is quite different from estimating α. When estimating α for the correct model, the sample variance of $\hat{\alpha}_{gcv}$ is usually larger than that of $\hat{\alpha}_{gml}$, (see simulation results in the next section, also see Stein (1989) and Stein (1990)).

iii). When the model is misspecified, we get pretty good estimate of r by the GCV-r estimator.

4.4 Estimation of α

In this section, we will show some results on the estimation of α. Remember that when the stochastic model is not right, there is no "true" α. So when we did grid search for $\hat{\alpha}$, the upper limit and lower limit for the search are different under different situations. However, we did make the searching range large enough in each case to make sure that we find the right one. We do the plots in Log scale. We will only present some of the histograms for $\hat{\alpha}$.

From Figure 4.4.1, we see that when the stochastic model is right, the sample variances of $\hat{\alpha}_{gml}$ are smaller than those of $\hat{\alpha}_{gcv}$ or $\hat{\alpha}_{rcv}$. Also remember that when the model is right, the true $\alpha_0 = 0.2$, $\text{Log}(\alpha_0) = -1.609$. So in this case, GML-$r$ does a better job in estimating α. That is why when the model is right, GML-r produces a better estimate of f.

From Figures 4.4.2, 4.4.3 and 4.4.4, we see that when the model is misspecified, $\hat{\alpha}_{gml}$ is very different from the "optimal" smoothing parameter, i.e. $\hat{\alpha}_t$, while $\hat{\alpha}_{gcv}$ and $\hat{\alpha}_{rcv}$ more or less follow the optimal α, $\hat{\alpha}_{rcv}$ follows more closely. That is perhaps why when the model is misspecified, GCV-r produces a better estimate of f than GML-r does, and RCV also produces a better estimate of f although $\hat{r}_{rcv}$ is not as good as $\hat{r}_{gcv}$.

Histograms for misspecification 2 are like those for misspecification 3. Misspecifications 2, 3 and 4 all represent cases where the true f is smoother than specified. We see from the

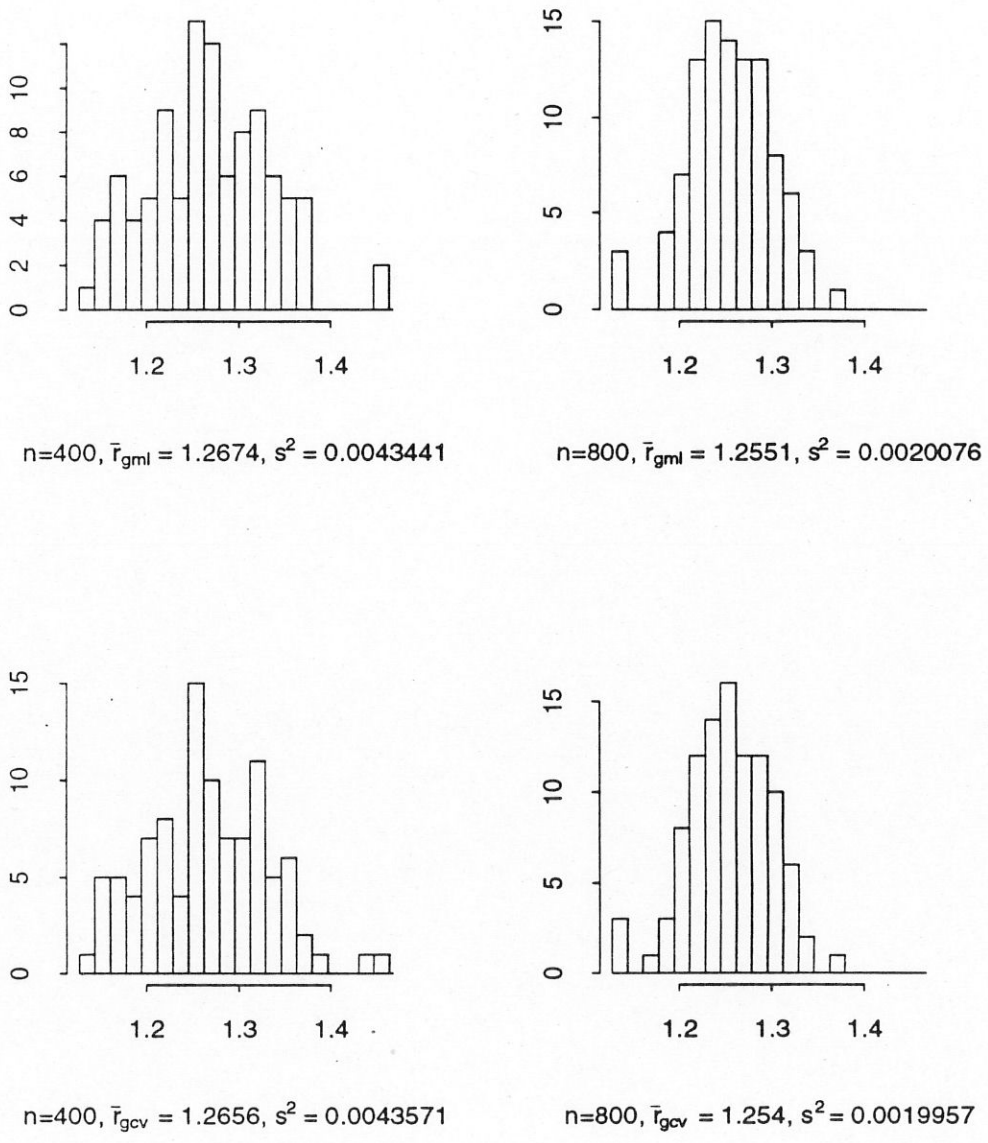


Figure 4.3.8: Distributions of $\hat{r}$ with 100 replicates when the model is correctly specified. The true r=1.25.

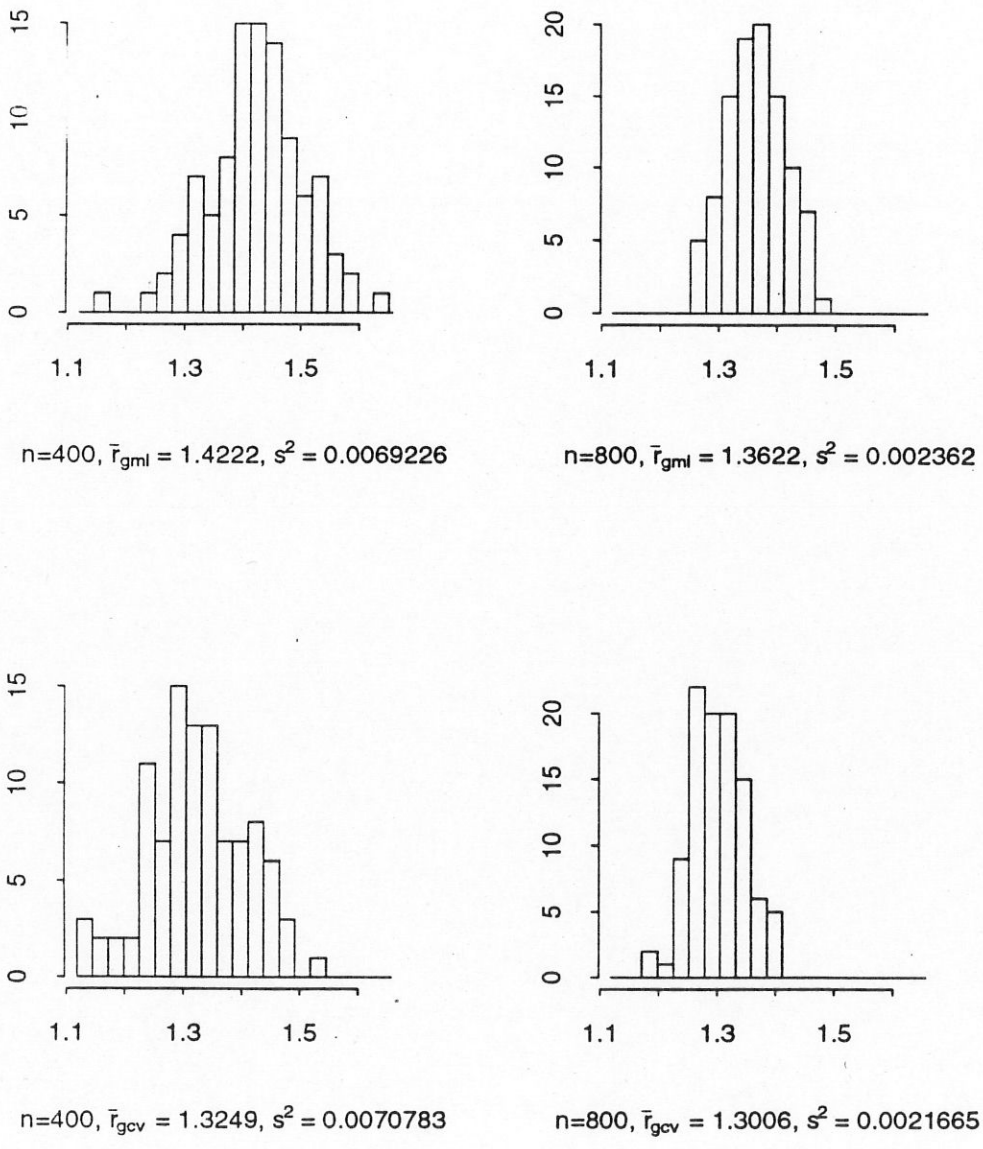


Figure 4.3.9: Distributions of $\hat{r}$ with 100 replicates for misspecification 1. The true r=1.25.

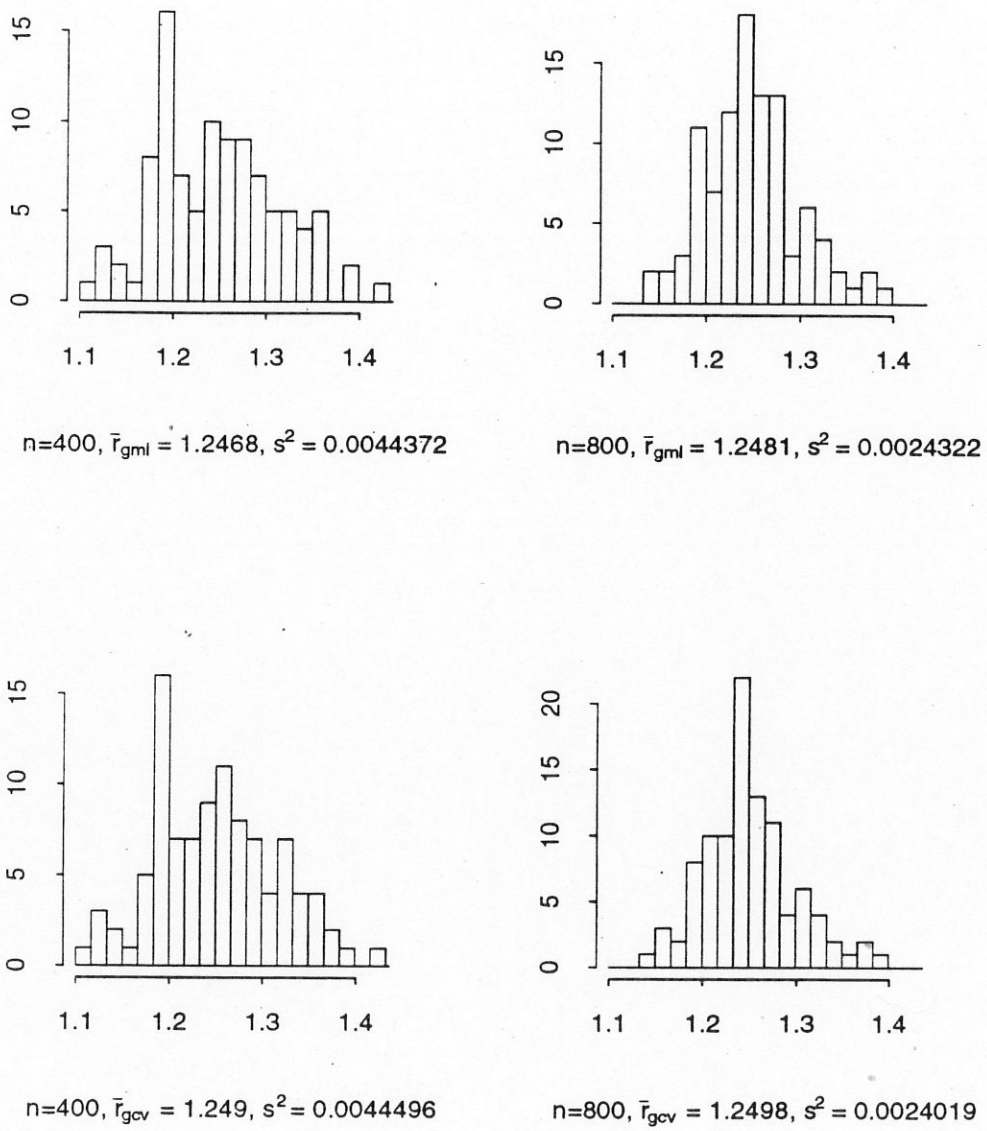


Figure 4.3.10: Distributions of $\hat{r}$ with 100 replicates for misspecification 3. The true r=1.25.

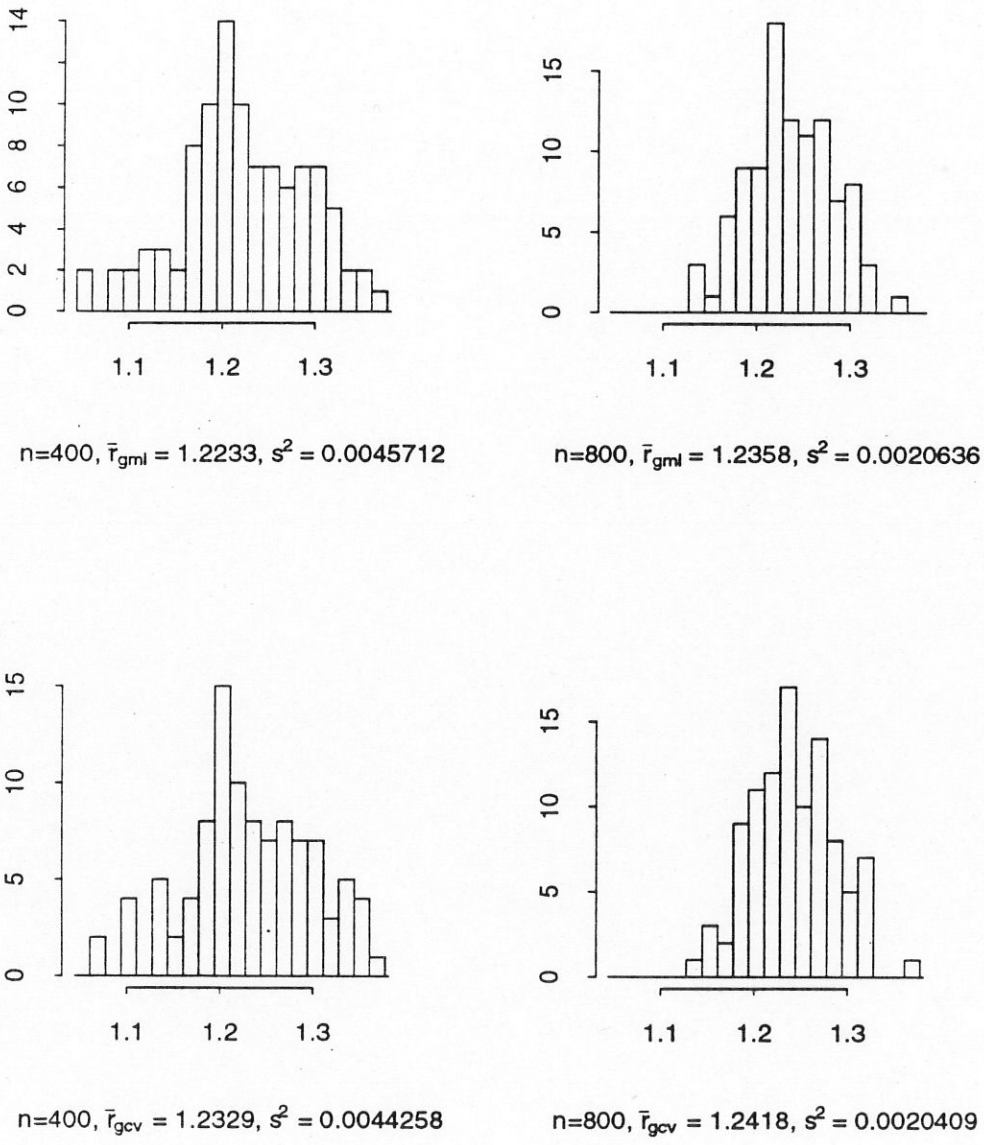


Figure 4.3.11: Distributions of $\hat{r}$ with 100 replicates for misspecification 4. The true r=1.25.

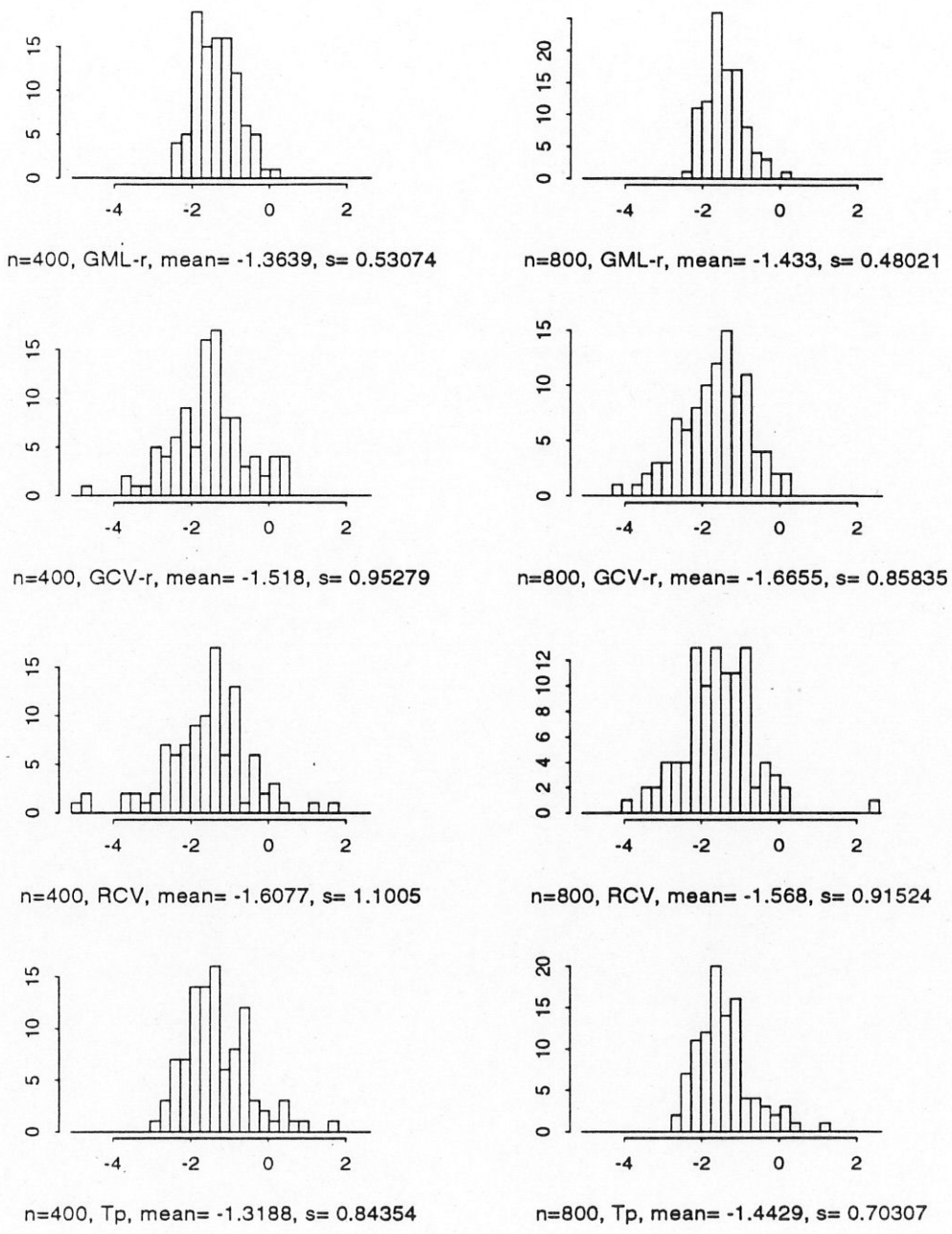


Figure 4.4.1: Distributions of Log($\hat{\alpha}$) with 100 replicates when the model is correctly specified.

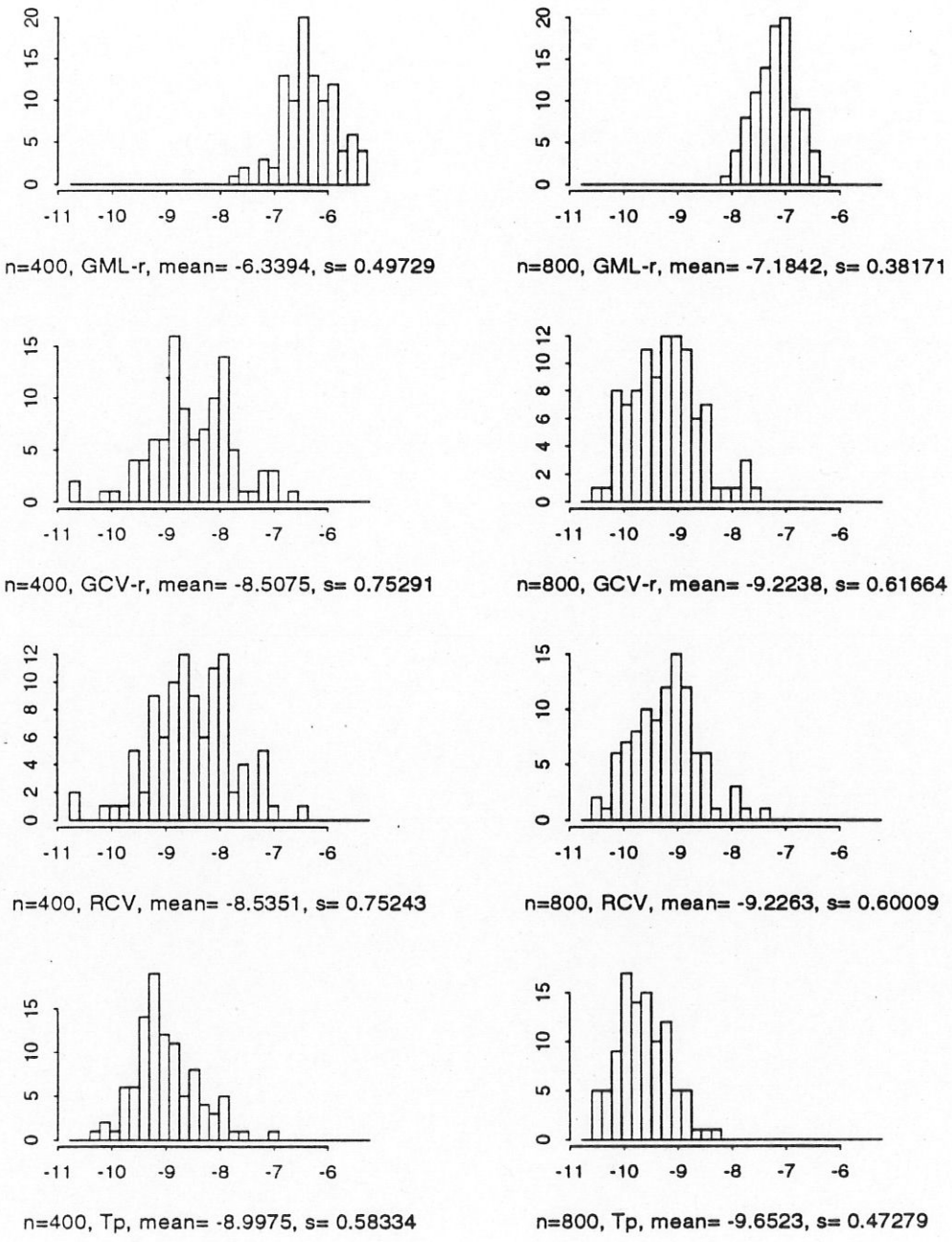


Figure 4.4.2: Distributions of Log($\hat{\alpha}$) with 100 replicates for misspecification 1.

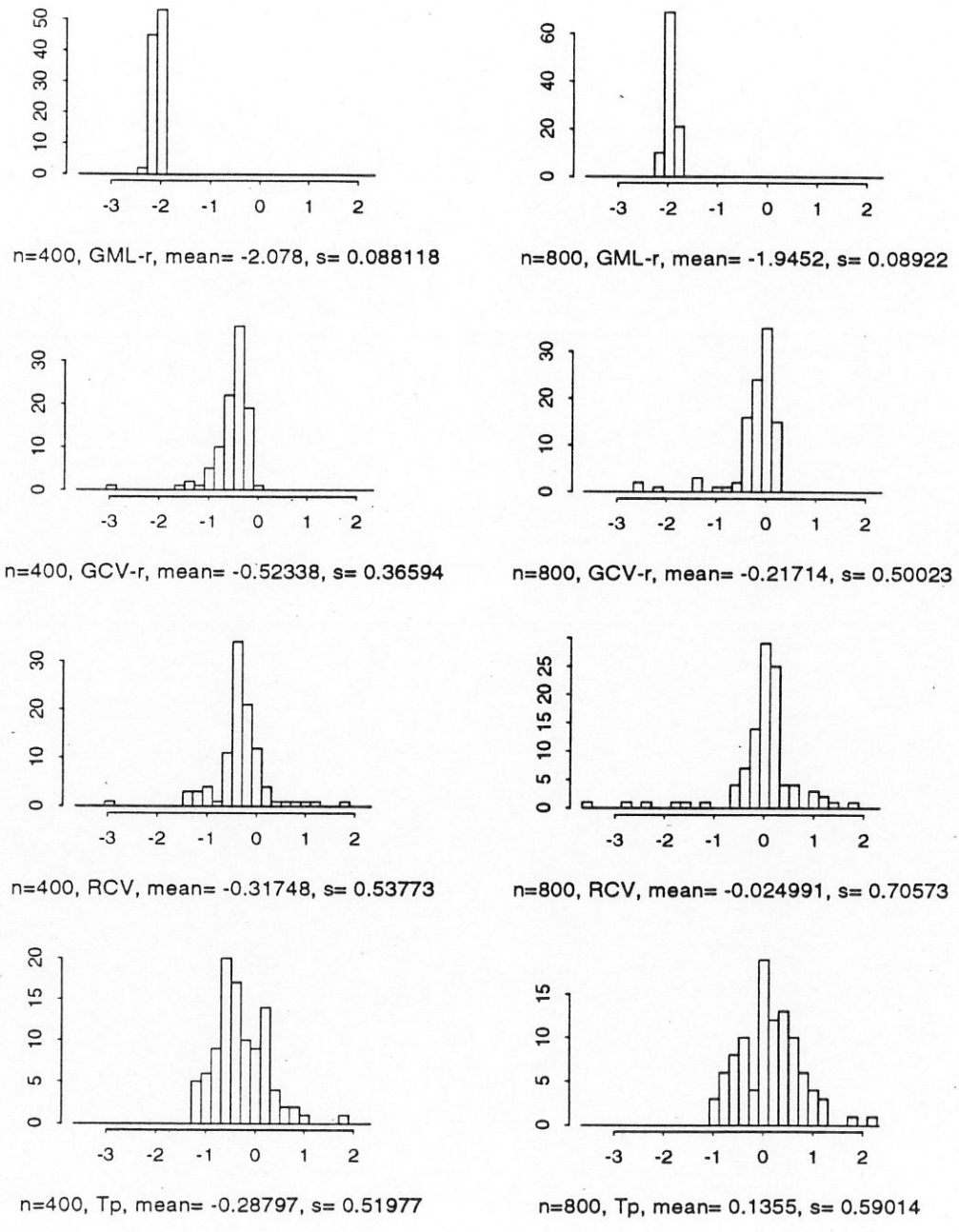


Figure 4.4.3: Distributions of Log($\hat{\alpha}$) with 100 replicates for misspecification 3.

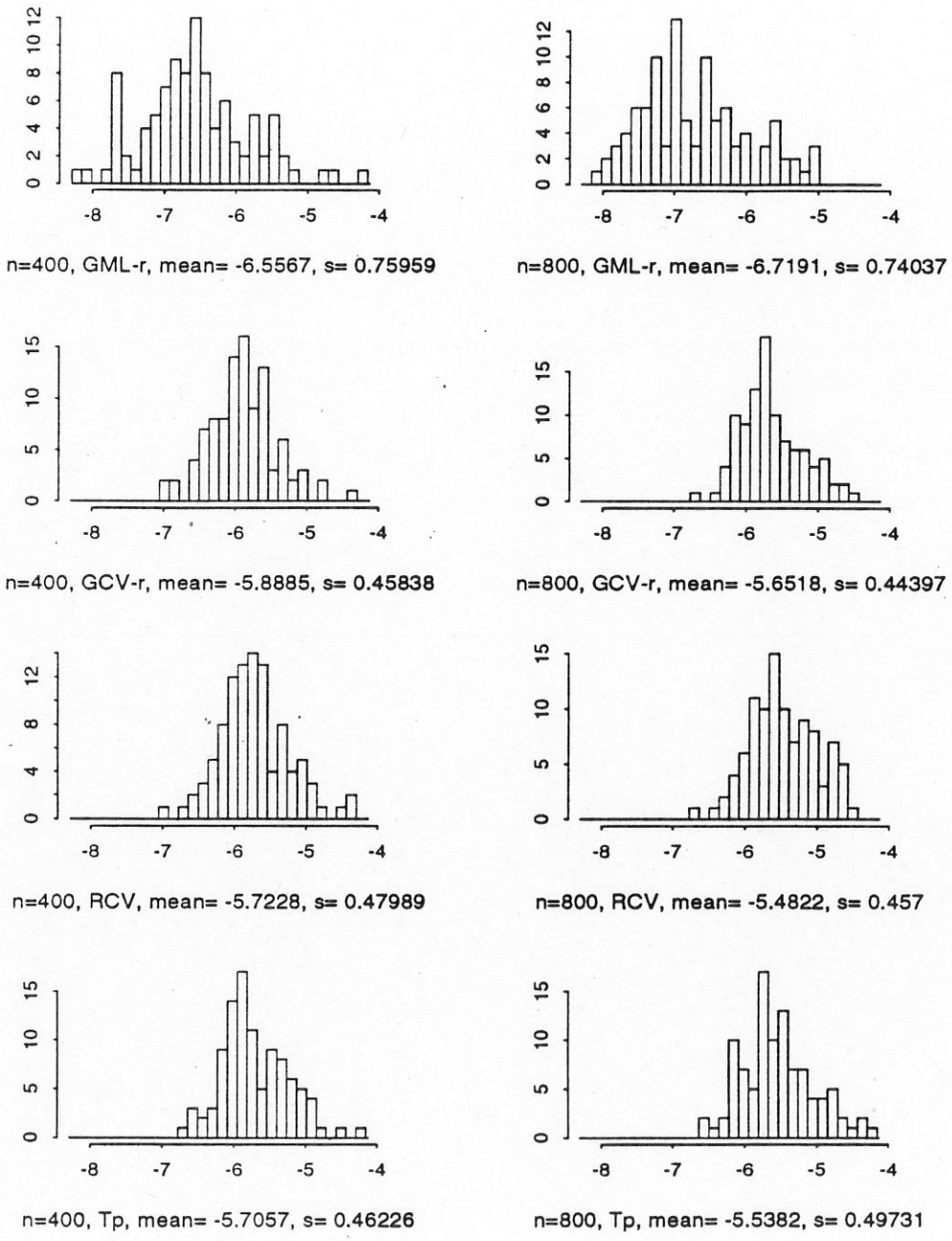


Figure 4.4.4: Distributions of Log($\hat{\alpha}$) with 100 replicates for misspecification 4.

histograms that in these cases, the GML-r estimate undersmooths relative to the GCV-r estimate. This is in agreement with our discussion in section 3.6.

Histograms for misspecification 5 are similar to those for misspecification 4, both cases represent misspecification of the location of peak of the eigenvalues.

4.5 Comparing I_p with $\hat{r}$ and $\hat{\alpha}$

It is also of interest to find out the relationship between getting a good estimate of r and α and getting a good estimate of f. In this section, we will look at some plots of I_p vs. $\hat{r}$ and I_p vs. $\hat{\alpha}$. First, let us look at I_p vs. $\hat{r}$, We put those plots in Figure 4.5.1 and Figure 4.5.2. Again, we do the plots in Log scale and we only do them for n=800.

From the plots, we see that although in some cases, there is a weak trend that better $\hat{r}$ corresponds to better I_p, this trend is not strong and clear. This is perhaps due to the fact that the predictive mean-square error is very insensitive to r, and we are using I_p to measure how good our estimate of f is.

Now let us look at the plots of I_p vs. $\hat{\alpha}$, again in Log scale. Since only when the stochastic model is true, we know what the true α is, for the misspecification cases, we will use the sample mean of the $\hat{\alpha}_t$ as our "optimal" α. We put the plots in Figure 4.5.3 and Figure 4.5.4.

From the plots, we see that in some cases, there is a weak trend that better $\hat{\alpha}$ tends to correspond to better I_p.

So, if we use the predictive mean-square error as a criterion to measure the estimate of f, getting a good estimate of the smoothing parameter α is also important. This is perhaps why for misspecification 4, although RCV is doing badly in estimating the weighting parameter r compared to the GCV-r estimate, it gets slightly better estimate of f in terms of I_p because it gets better estimate of the smoothing parameter α.

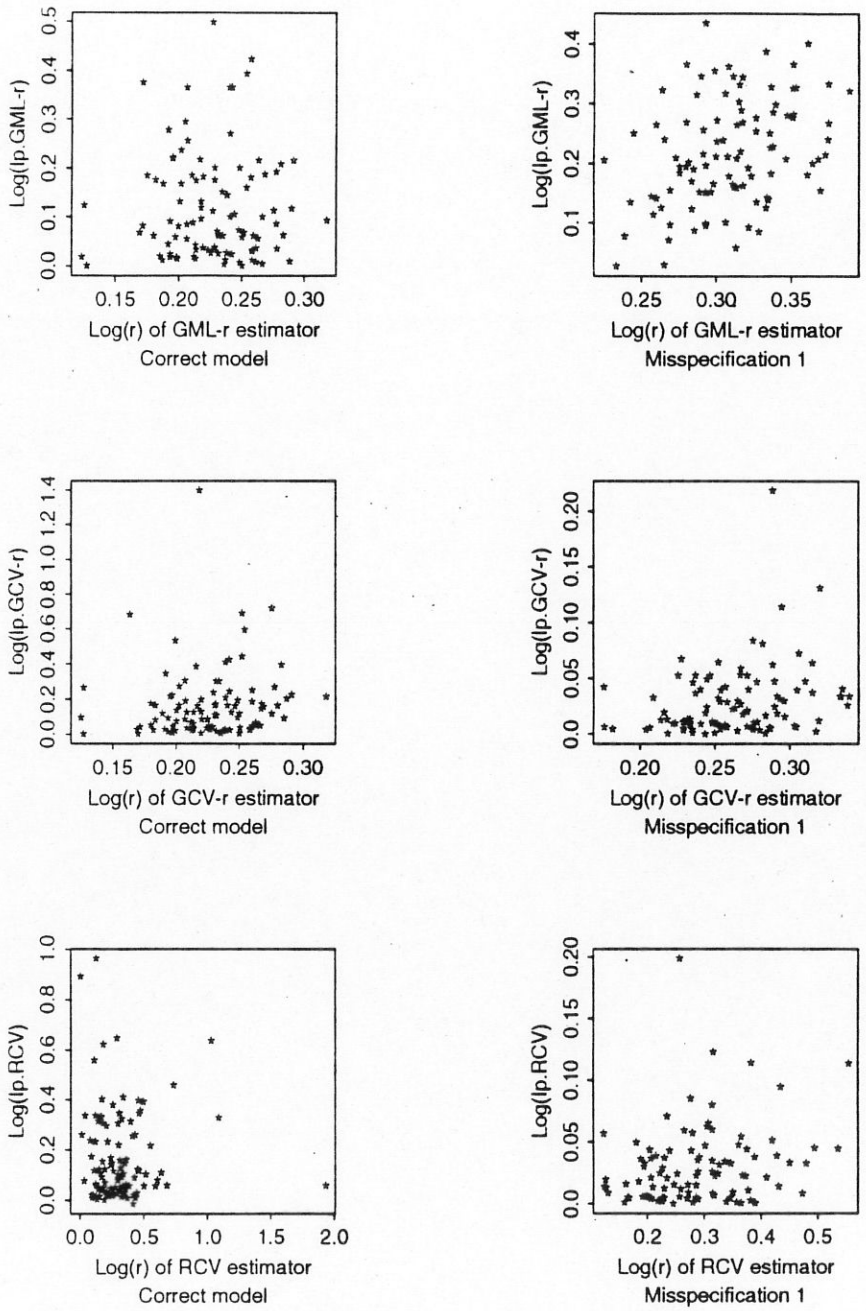


Figure 4.5.1: Log(I_p) vs. Log($\hat{r}$). True Log(r)=0.2231. n=800.

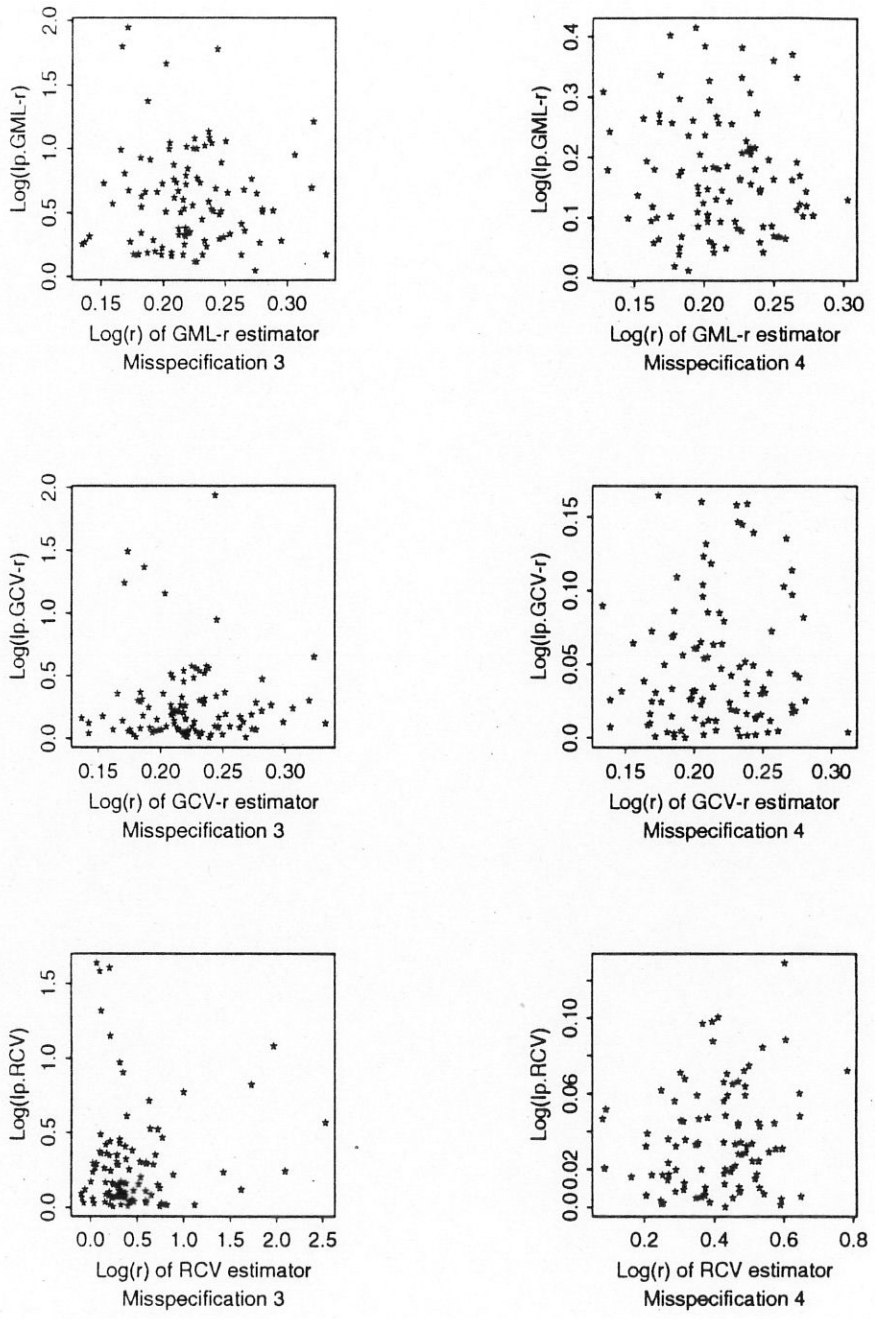


Figure 4.5.2: Log(I_p) vs. Log($\hat{r}$). True Log(r)=0.2231, n=800.

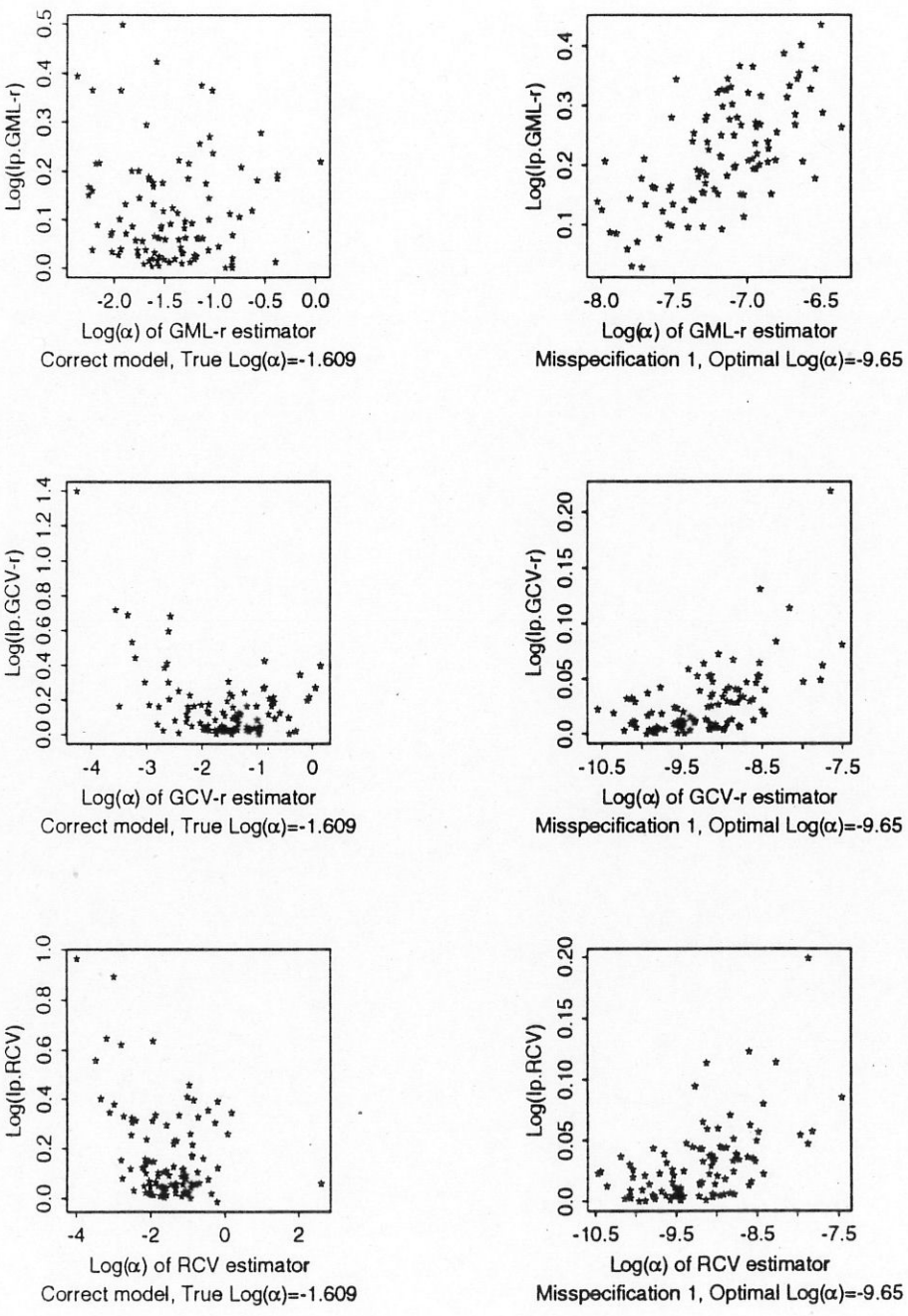


Figure 4.5.3: Log(I_p) vs. Log($\hat{\alpha}$). n=800.

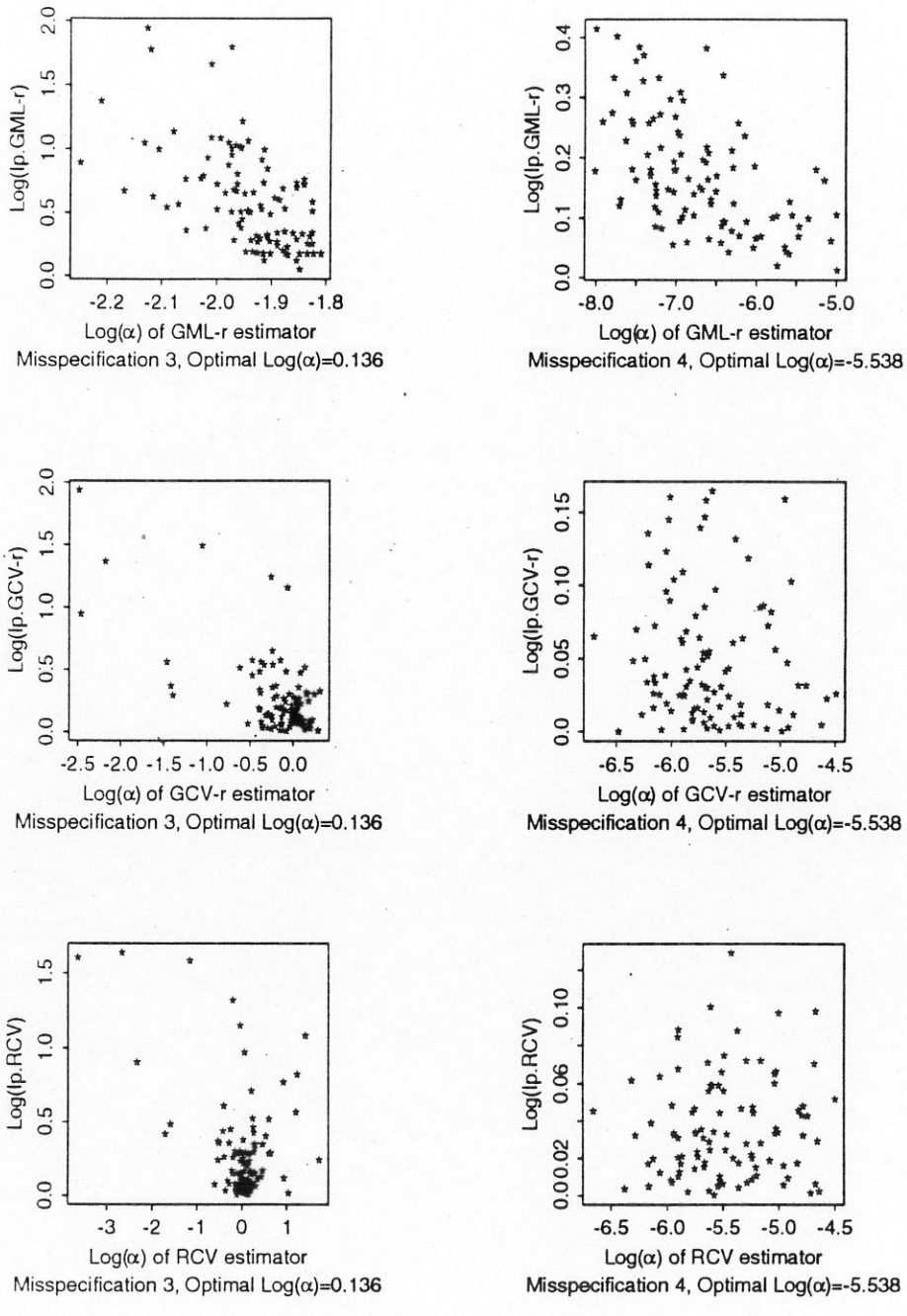


Figure 4.5.4: Log(I_p) vs. Log($\hat{\alpha}$). n=800.

Chapter 5

Simulation Studies for Functions on the Sphere

In this chapter, we present results from some Monte Carlo experiments in a more realistic setting, where f is a function on the sphere. We try to recover the 500 mb height surface from simulated direct observations and forecast of the current 500 mb heights.

5.1 Experiment Set-Up

The 500 mb height surface is a function on the sphere. We can represent this function by a spherical harmonic approximation of the form $f(P) = \sum_{l=0}^{M} \sum_{s=-l}^{l} f_{ls} Y_{ls}(P)$, where Y_{ls} are spherical harmonics, P represents a point on the sphere and f_{ls} are the Fourier-Bessel coefficients. So in order to recover f, we only need to estimate these coefficients f_{ls}. In our experiments, we choose $M = 30$, so there are 961 coefficients.

We have chosen locations of 600 real radiosonde stations over northern hemisphere as our "observation" locations. Figure 5.1.1 shows the locations. The (latitude, longitude)'s in degrees of those 600 radiosonde stations are listed in Appendix C.

We assume the observation error to be white noise. Our direct observations can then be modeled as

$$y_{1i} = \sum_{l=0}^{30} \sum_{s=-l}^{l} f_{ls} Y_{ls}(P_i) + \varepsilon_{1i} \tag{5.1.1}$$

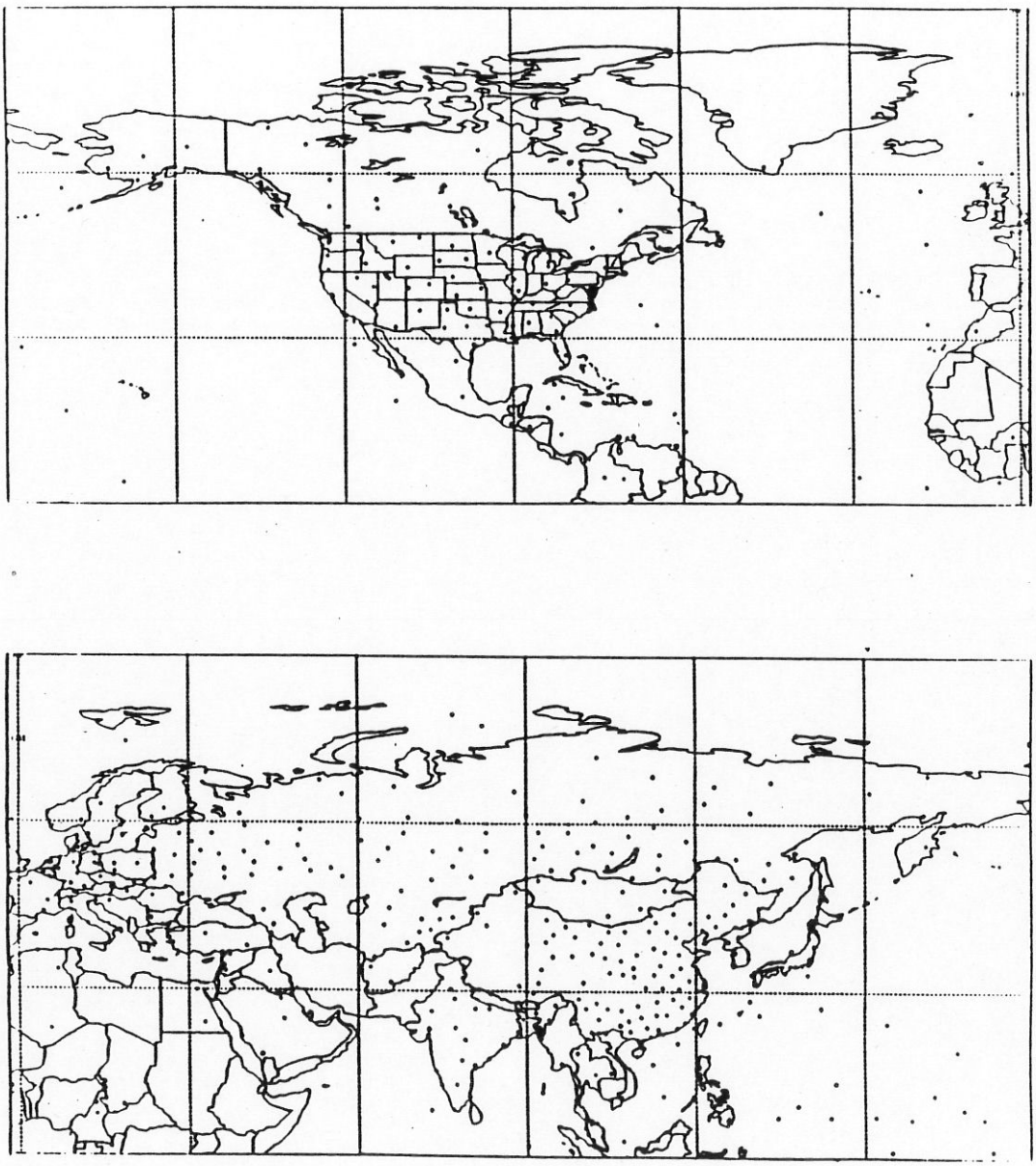


Figure 5.1.1: Locations of the radiosonde stations over northern hemisphere

where $P_i, i = 1, \ldots, 600$ are the observation locations, f_{ls} are the true but unknown Fourier-Bessel coefficients and ε_{1i} are *i.i.d.* random variables, $\varepsilon_{1i} \sim \mathcal{N}(0, \sigma_1^2)$.

We can write (5.1.1) in a matrix form:

$$y_1 = K_1 f + \varepsilon_1 \tag{5.1.2}$$

where y_1 is a 600 dimensional vector of observations, f is a 961 dimensional vector of the Fourier-Bessel coefficients that need to be estimated, ε_1 is a 600 dimensional random vector of noise and K_1 is a 600 by 961 matrix containing the values of spherical harmonics.

The forecast errors in spatial domain are usually correlated and some work has been carried out to study the statistical structure of the forecast error correlation. See Hollingsworth and Lonnberg (1986), Lonnberg and Hollingsworth (1986), Wahba (1989), Mitchell, Charette, Chouinard and Brasnett (1990) and Goerss and Phoebus (1991). We use the model described in Wahba (1989) to approximate the forecast error correlation. According to this model, the forecast error in the frequency domain can be approximated as independent weighted normal random variables. More specifically, the forecast can be modeled as

$$f^F = f + \tilde{\varepsilon}_2 \tag{5.1.3}$$

where f^F is a 961 dimensional vector of forecast, f is a 961 dimensional vector of the true Fourier-Bessel coefficients, $\tilde{\varepsilon}_2$ is the forecast error in the frequency domain, $\tilde{\varepsilon}_2 \sim \mathcal{N}(0, \sigma_2^2 \Lambda)$, where $\Lambda = \mathrm{diag}[\lambda_{ls}^F]$, $s = -l, \ldots, l; l = 0, \ldots, 30$, and λ_{ls}^F only depends on l:

$$\lambda_{00}^F = \frac{1 - (1+\theta)^{-1}}{(1-\theta)^{-1} - (1+\theta)^{-1}}(4\pi) = 2\pi(1-\theta)$$

$$\lambda_{ls}^F = \frac{\theta^l}{(1-\theta)^{-1} - (1+\theta)^{-1}} \frac{4\pi}{2l+1} = \theta^{l-1}(1+\theta)(1-\theta)\frac{2\pi}{2l+1}$$

$$s = -l, \ldots, l; l = 1, \ldots, 30$$

where $\theta = \cos(2\pi L/3R_0) - \sqrt{3}\sin(2\pi L/3R_0)$, R_0 is the circumference of the earth and L represents the half correlation length. Lonnberg and Hollingsworth (1986) found an L of about 600 km. in their work.

This model approximates the forecast error correlation fairly well in the range of our experiments, i.e. in the range of using only the first 961 spherical harmonics.

If we multiply both sides of (5.1.3) by $\Lambda^{-1/2}$, we get

$$y_2 = K_2 f + \varepsilon_2 \tag{5.1.4}$$

where $y_2 = \Lambda^{-1/2} f^F$, $K_2 = \Lambda^{-1/2}$ and $\varepsilon_2 \sim \mathcal{N}(0, \sigma_2^2 I)$.

From (5.1.2) and (5.1.4), we see that our model here is

$$\begin{cases} y_1 = K_1 f + \varepsilon_1 \\ y_2 = K_2 f + \varepsilon_2 \end{cases} \tag{5.1.5}$$

where y_i is of dimension n_i, $i = 1, 2$, f is of dimension n, K_i is an $n_i \times n$ matrix, $i = 1, 2$, $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2 I)$, $i = 1, 2$, they are independent and σ_1 and σ_2 are unknown. $n_1 = 600$, $n_2 = 961$, $n = 961$.

For the true vector f for 500 mb heights we used spherical harmonic coefficients kindly provided by Dr. Fred Reames which approximately described the Northern Hemisphere 500 mb heights at 00GMT for January 2, 1979 and at 00GMT January 14, 1979. Contour plots of the Northern Hemisphere 500 mb heights for both days appear in Figure 5.1.2. Please see Appendix B for a description of how the coefficients were generated.

Stochastic model (1.1.6) with $m = 2$ for functions on the sphere provides a good approximation of meteorological truths. Figure 5.1.3 shows the plots of $\log_{10}(\bar{f}_l^2)$ vs. $\log_{10}(l)$, where $\bar{f}_l^2 = (1/(2l+1)) \sum_{s=-l}^{l} f_{ls}^2, l = 1, \ldots, 30$. The f_{ls}'s for the "prior" come from model (1.1.6) with $m = 2$.

In order to measure the performance of our estimates, we could use the predictive mean-square error $T_p(r, \alpha)$, or the solution mean-square error $T_s(r, \alpha)$ defined by

$$T_s(r, \alpha) = \sum_{l=0}^{30} \sum_{s=-l}^{l} (\hat{f}_{ls} - f_{ls})^2.$$

We can calculate the relative inefficiencies with respect to T_p and T_s defined by

$$I_p = \frac{T_p(\hat{r}, \hat{\alpha})}{\min T_p(r, \alpha)}$$

and

$$I_s = \frac{T_s(\hat{r}, \hat{\alpha})}{\min T_s(r, \alpha)}$$

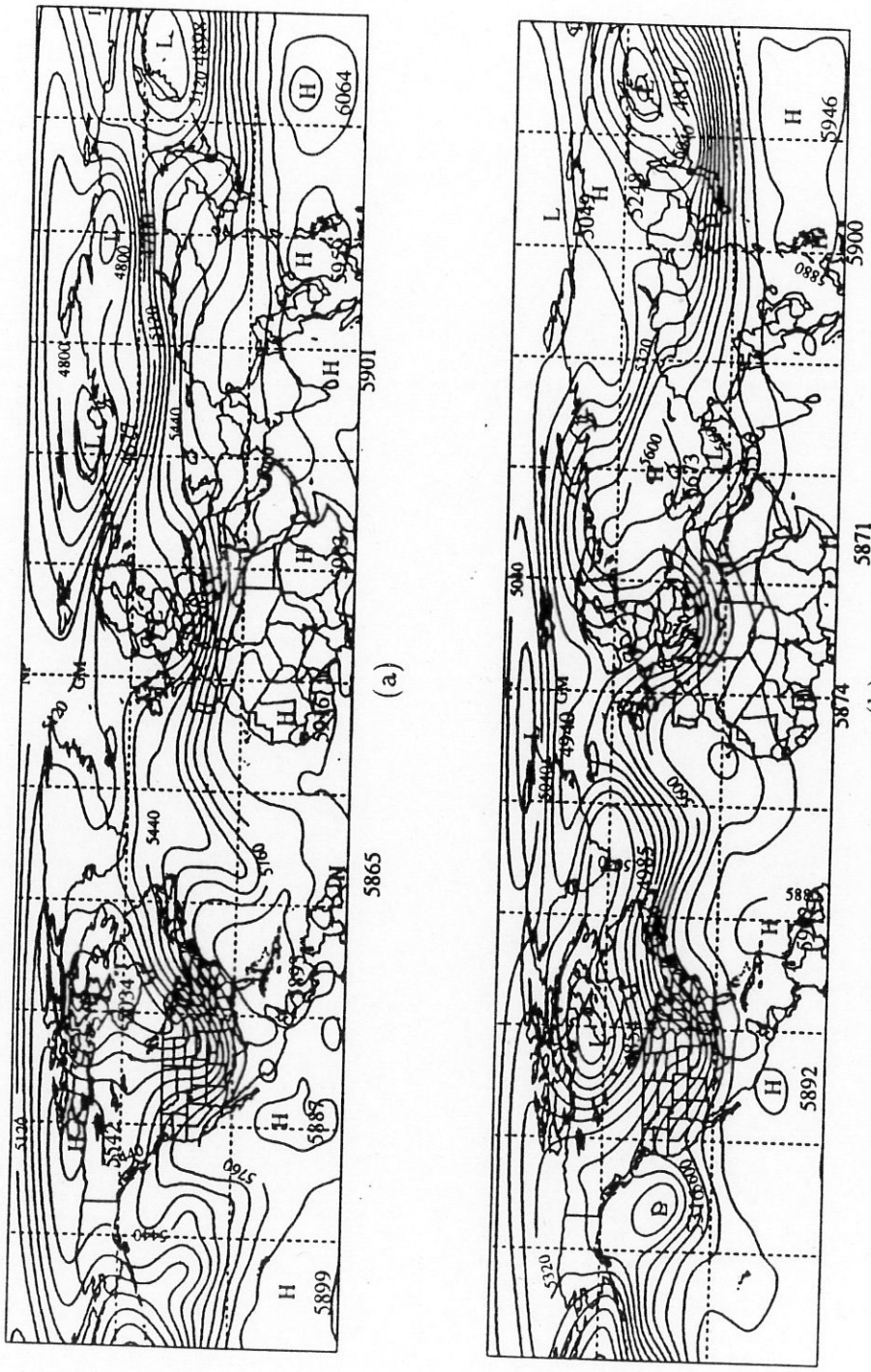


Figure 5.1.2: (a). 500 mb heights of January 2, 1979. (b). 500 mb heights of January 14, 1979.

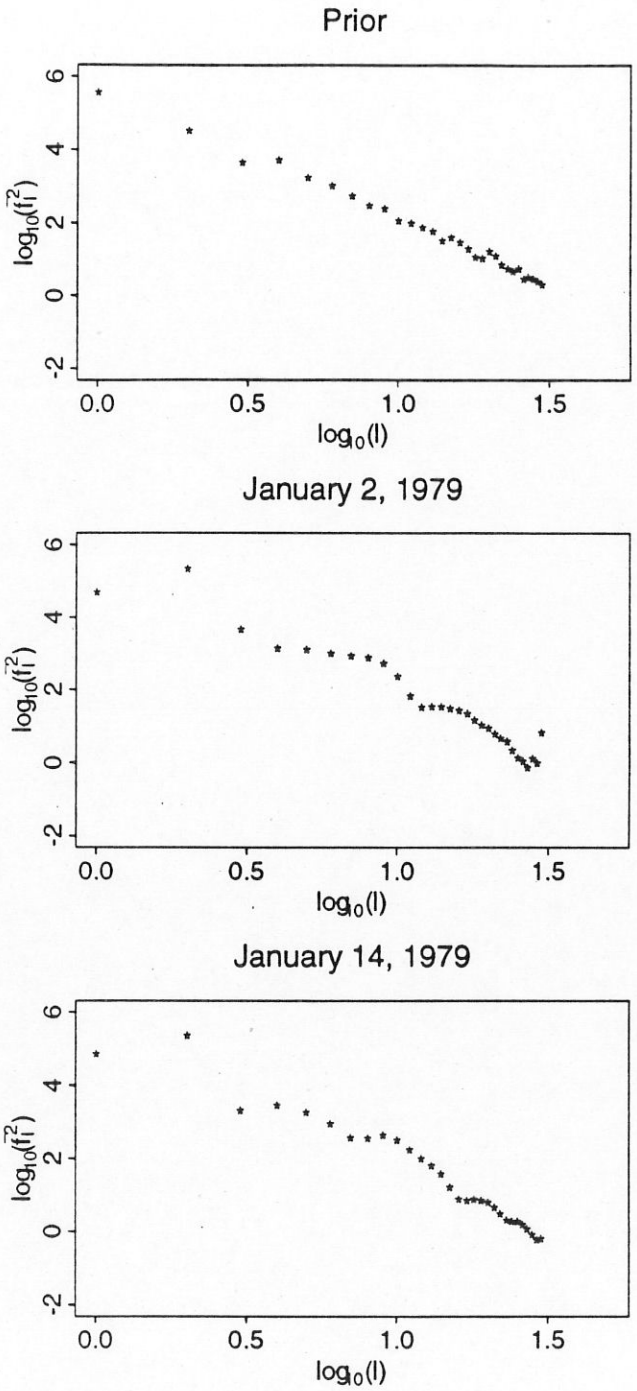


Figure 5.1.3: Comparison of the stochastic model and meteorological truths

where $\hat{r}$ and $\hat{\alpha}$ are estimates of r and α. For more discussions on the relationship between T_p and T_s, please see Wahba and Wang (1990).

5.2 Preliminary Experiment

In our preliminary experiment, we choose $m = 2$, $\sigma_1 = \sigma_2 = 15$m. So $r_0 = 1$. We did 10 replications. We summarize the results in Figure 5.2.1 and Figure 5.2.2.

Figure 5.2.1 shows that GCV-r is much better than RCV for estimating the weighting parameter r in that GCV-r estimate not only has smaller variance, but also is unbiased, while RCV estimates of r for January 2 and January 14 are biased.

Figure 5.2.2 shows that GCV-r is also much better than RCV for estimating the function f in that GCV-r estimate has a much smaller relative inefficiency with respect to both the predictive mean-square error T_p and the solution mean-square error T_s.

5.3 More Results of Monte Carlo Studies

From the results of the preliminary experiment, we see that GCV-r is much better than RCV in estimating r and f. So we should use the GCV-r estimate instead of RCV estimate. In this section, we will concentrate on comparing the GCV-r and the GML-r estimate in some further experiments. We also will study the effects of some possible model misspecifications on the performance of GCV-r and GML-r estimate.

In our experiments, for the true vector f for 500 mb heights we used spherical harmonic coefficients kindly provided by Fred Reames which approximately described the 500 mb heights at 00GMT for January 2, 1979 and January 14, 1979. We also used a random number generator to generate f according to model (1.1.6) as our "prior" case. We used Gaussian random number generator to generate the noises and added them to the signals according to model (5.1.5) to get our "observations" and "forecasts". According to Goerss and Phoebus (1991), σ_1 for 500 mb heights is approximately 9 meters, and σ_2 for 500 mb heights is approximately 11 meters. So in the following experiments, we choose $\sigma_1 = 9$ and $\sigma_2 = 11$, therefore $r_0 = 0.818$. We have conducted the following three experiments:

Experiment 1, $L = 600$ km, $m = 2$. There was no misspecification.

Experiment 2, $L = 600$ km, with a misspecification $m = 3$. In this experiment, for the "prior", f was generated according to (1.1.6) with $m = 2$. We analyzed the data by a

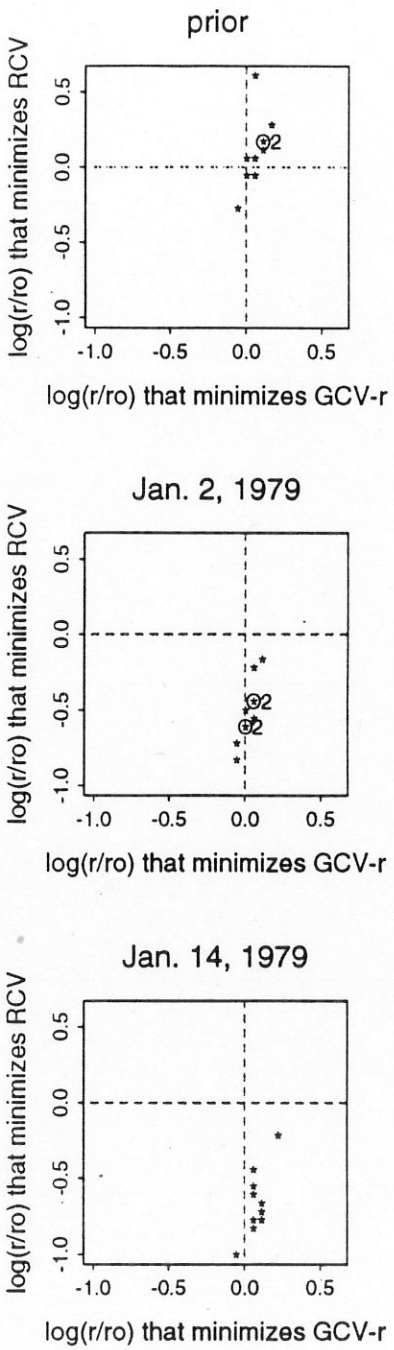


Figure 5.2.1: $\hat{r}_{rcv}$ vs. $\hat{r}_{gcv}$. for Preliminary Experiment.

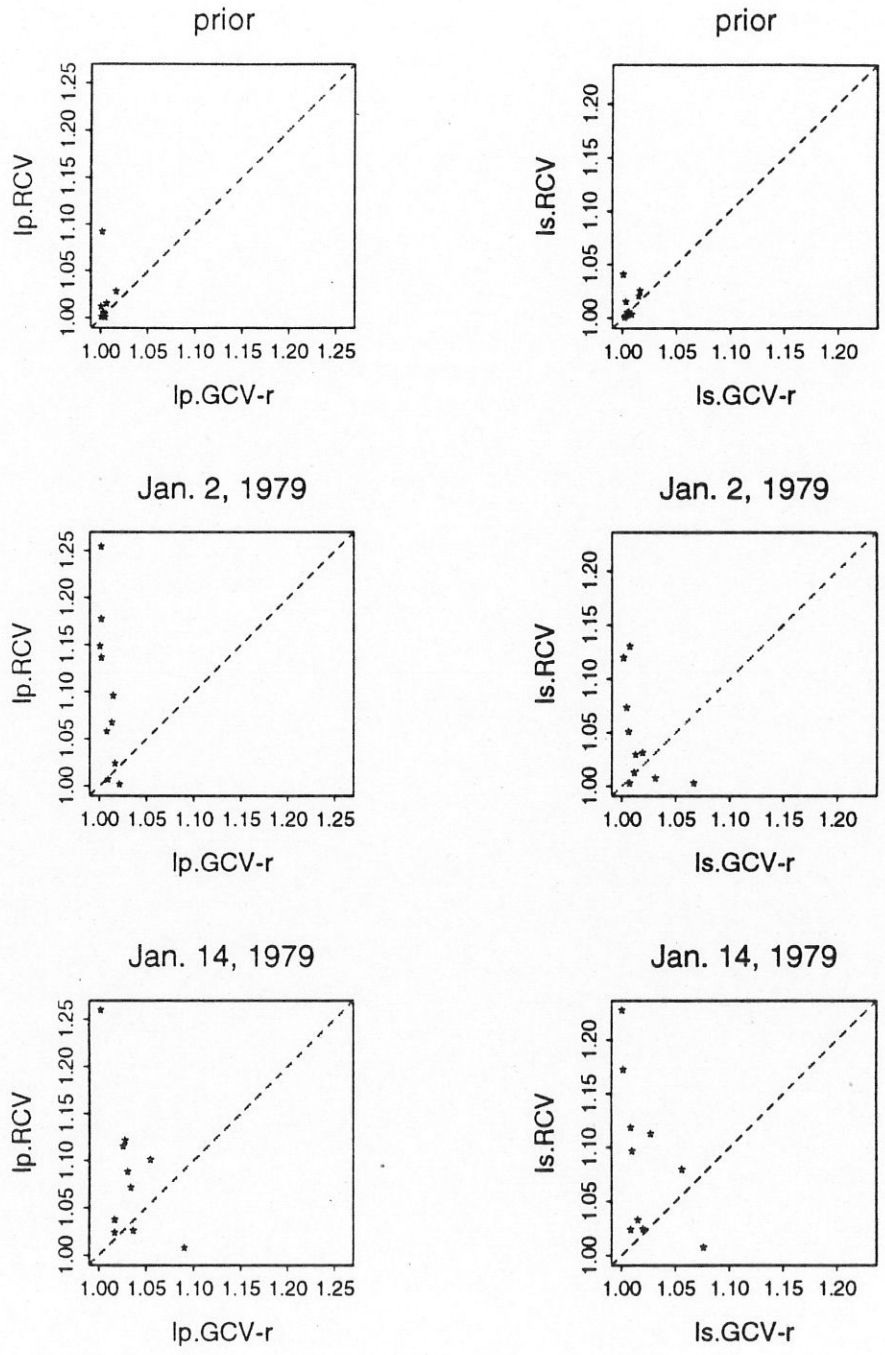


Figure 5.2.2: Relative inefficiencies of RCV vs. GCV-r for Preliminary Experiment.

model with $m = 3$.

Experiment 3, $m = 2$, with a misspecification on the forecast error half correlation length L. We generated the "data" with $L = 500$ km, and we analyzed the data with $L = 600$ km.

For each of the 3 experiments, we estimated r, α and f by GCV-r and GML-r. We compared the GCV-r estimate $\hat{r}_{gcv}$, the GML-r estimate $\hat{r}_{gml}$ with the "optimal" estimate of r that minimizes $T_p(r, \alpha)$ or $T_s(r, \alpha)$. We also compared the performance of GCV-r and GML-r in estimating f in terms of I_p and I_s. We repeated these 10 times with different noises in each experiment.

Next, we present the simulation results for the three experiments separately.

Experiment 1

Figure 5.3.1 shows the GCV-r and GML-r estimates of r compared with the "optimal estimate" of r with respect to T_p. We see from the plot that for cases where f is a meteorological truth, especially for the case when f is the truth of January 14, 1979, the GML-r estimate is biased. Even the "optimal estimate" is biased. On the other hand, GCV-r gives excellent unbiased estimates.

Figure 5.3.2 shows the GCV-r and GML-r estimates of r compared with the "optimal estimate" of r with respect to T_s. We see the same pattern as in Figure 5.3.1. The GCV-r does a wonderful job estimating the weighting parameter r.

Figure 5.3.3 shows the relative inefficiencies of the estimates with respect to T_p and T_s. We see that for the "prior" cases, the GCV-r and GML-r do equally well, while for the cases of meteorological truths, GCV-r does much better than GML-r.

Figure 5.3.4 shows the box-plot of the 10 replicates of the coverage percentages of the 95% Bayesian "confidence intervals" on data points. The formulas for calculating the Bayesian "confidence intervals" can be found in Section 2.4. The coverages are very close to 95%. These are very good results considering the fact that we know nothing about the noise's variances before we analyze the data.

Figure 5.3.5 shows the plot of $\hat{\sigma}_1/\hat{\sigma}_2$ vs. $\hat{r}$, where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are calculated using the formula in Section 2.4. We see a strong linear correlation, which shows that our estimates of r and σ_1 and σ_2 are consistent.

The improvement over the direct observation or the forecast after combining observational and forecast data according to our methods can be easily seen by looking at the root

mean-square error (RMSE). For example, we can calculate the RMSE for the estimates by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{f}(P_i) - f(P_i))^2}.$$

Similarly, we can calculate the RMSE for the direct observations and the forecasts by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_o(P_i) - f(P_i))^2},$$

and

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_f(P_i) - f(P_i))^2},$$

respectively, where $f(P_i)$ is the true 500 mb height at location P_i since we know the truth in our experiments, $\hat{f}(P_i)$ is the estimated value of 500 mb height at P_i either by GML-r or GCV-r estimate, $f_o(P_i)$ is the observed value, $f_f(P_i)$ is the forecasted value and $P_i, i = 1, \ldots, N$ is a set of points over our experimental region, i.e. the northern hemisphere. For direct observations, we use the 600 radiosonde stations. For forecasts, we use the following 1224 regular latitude/longitude grid of 5×5 degrees: latitude from 0° to 80°, longitude from $-180°$ to 175°. We summarize the results in the following two tables.

Notice that the RMSEs for direct observations are approximately 9 (meter), which is equal to σ_1. This is so because the direct observation errors are $i.i.d.\ \mathcal{N}(0, \sigma_1^2)$. The RMSEs for forecasts are somewhat less than σ_2 which is 11 (meter). This is so because the forecast errors in spatial domain are correlated and $E\sqrt{X} \leq \sqrt{EX}$. Also notice that the RMSEs for analyzed surfaces are smaller over the 600 radiosonde stations than they are over the 1224 grid points. This is not surprising because we have more information over these 600 stations, we should get better estimates over these 600 stations.

From table 5.3.1, it is seen that after combining observational and forecast data according to our methods, the improvement on average over direct observations is about 4 meters. From table 5.3.2, it is seen that on average, the improvement over forecasts is about 3 meters.

Table 5.3.1: RMSE (in meters) over 600 stations for Experiment 1

	January 2, 1979			January 14, 1979		
Runs	Direct Obs.	GCV-anal.	GML-anal.	Direct Obs.	GCV-anal.	GML-anal.
1	8.6981	4.6640	4.6549	8.6801	4.2384	4.2536
2	9.0619	4.5006	4.4927	8.6068	4.4033	4.3320
3	8.8373	4.4443	4.3699	8.7560	4.3871	4.3332
4	8.8280	4.5127	4.4417	8.8672	4.4498	4.4190
5	8.7650	4.2472	4.2295	9.1331	4.3601	4.3373
6	9.2655	4.4876	4.4939	9.0676	4.5677	4.5049
7	8.7035	4.2282	4.1936	9.0496	4.6259	4.5508
8	9.4251	4.5517	4.4993	9.2331	4.4940	4.5438
9	9.1061	4.1863	4.2149	8.8597	4.6344	4.5810
10	8.8997	4.4190	4.4438	8.7835	4.2142	4.0689

Table 5.3.2: RMSE (in meters) over 1224 grid points for Experiment 1

	January 2, 1979			January 14, 1979		
Runs	Forecast	GCV-anal.	GML-anal.	Forecast	GCV-anal.	GML-anal.
1	9.5618	6.6838	6.6499	9.5493	6.6562	6.5646
2	9.2089	6.8074	6.7573	9.8697	7.5520	7.4475
3	9.9623	6.6223	6.4176	9.2657	6.8620	6.5681
4	9.4256	6.3734	6.3306	10.2220	7.4204	7.1957
5	11.0508	7.2269	7.2281	10.3828	7.3223	7.1382
6	9.7560	6.9711	6.9324	10.1805	6.8868	6.5799
7	9.6693	6.5305	6.4713	11.9464	6.8106	6.5055
8	9.9640	6.5120	6.4672	9.7767	7.8665	7.5220
9	9.0028	6.8066	6.6926	10.0572	6.7572	6.5481
10	10.1872	6.7643	6.7587	10.1816	7.2994	6.8685

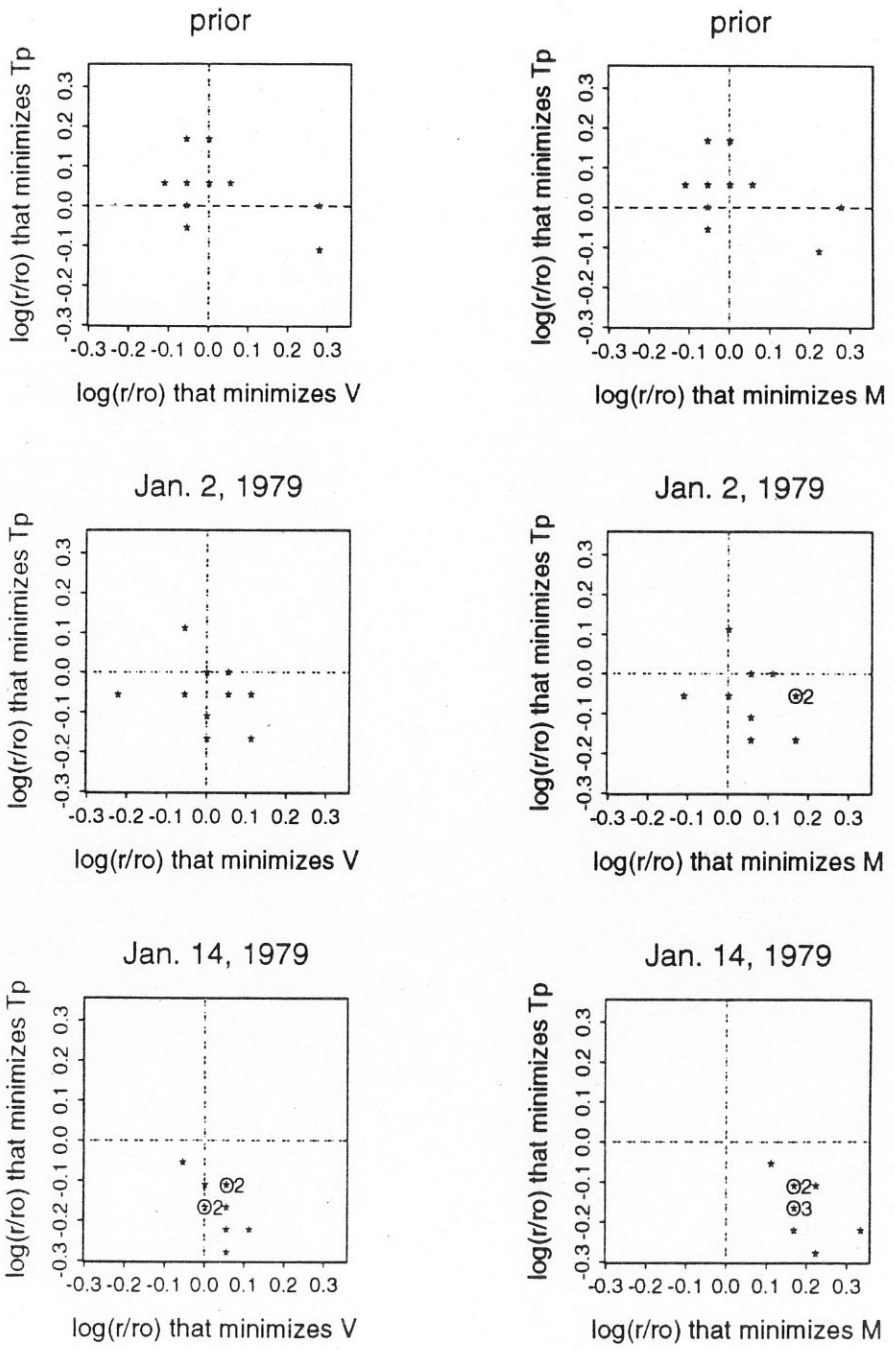


Figure 5.3.1: Estimates of r for Experiment 1.

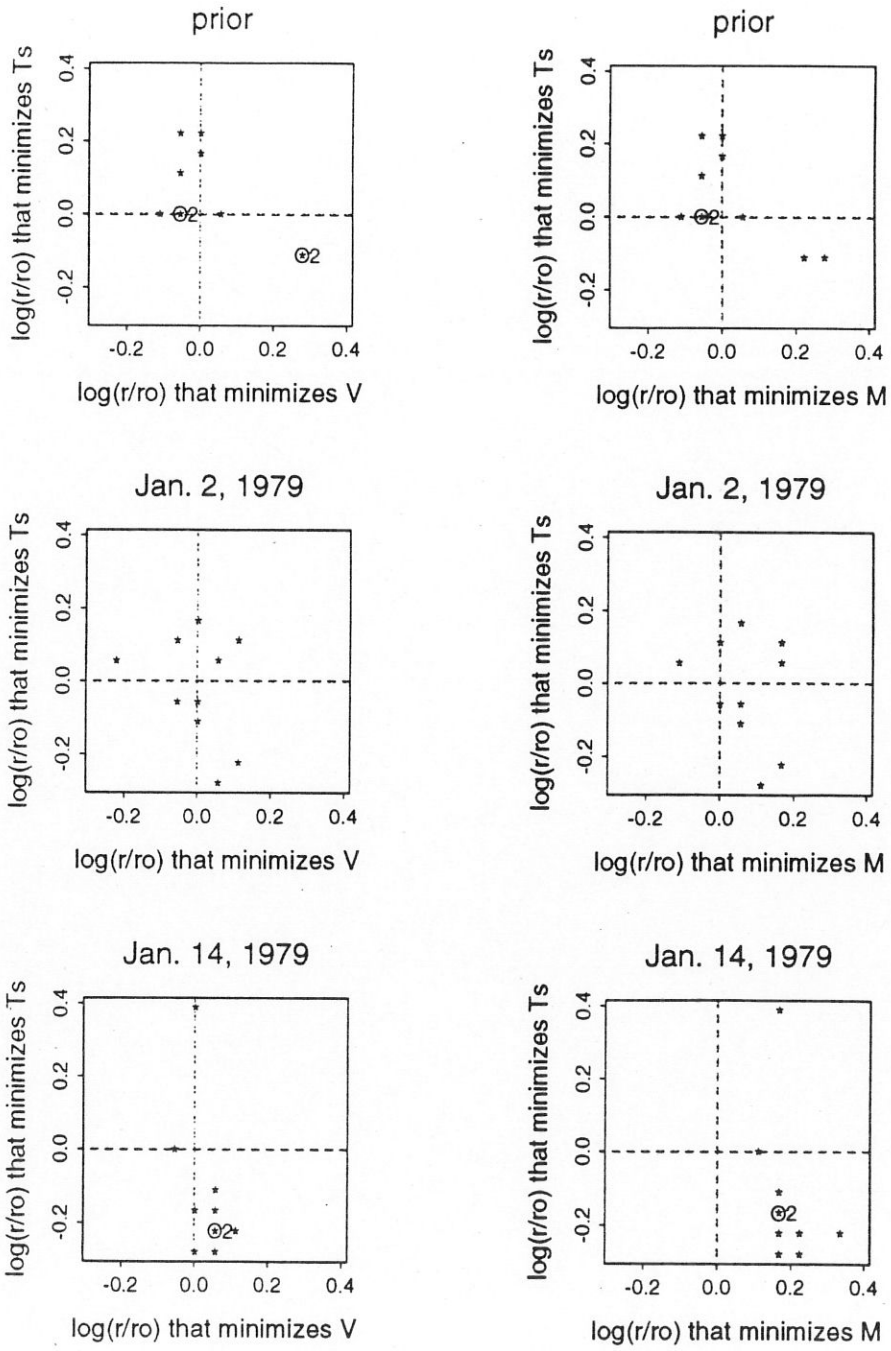


Figure 5.3.2: Estimates of r for Experiment 1.

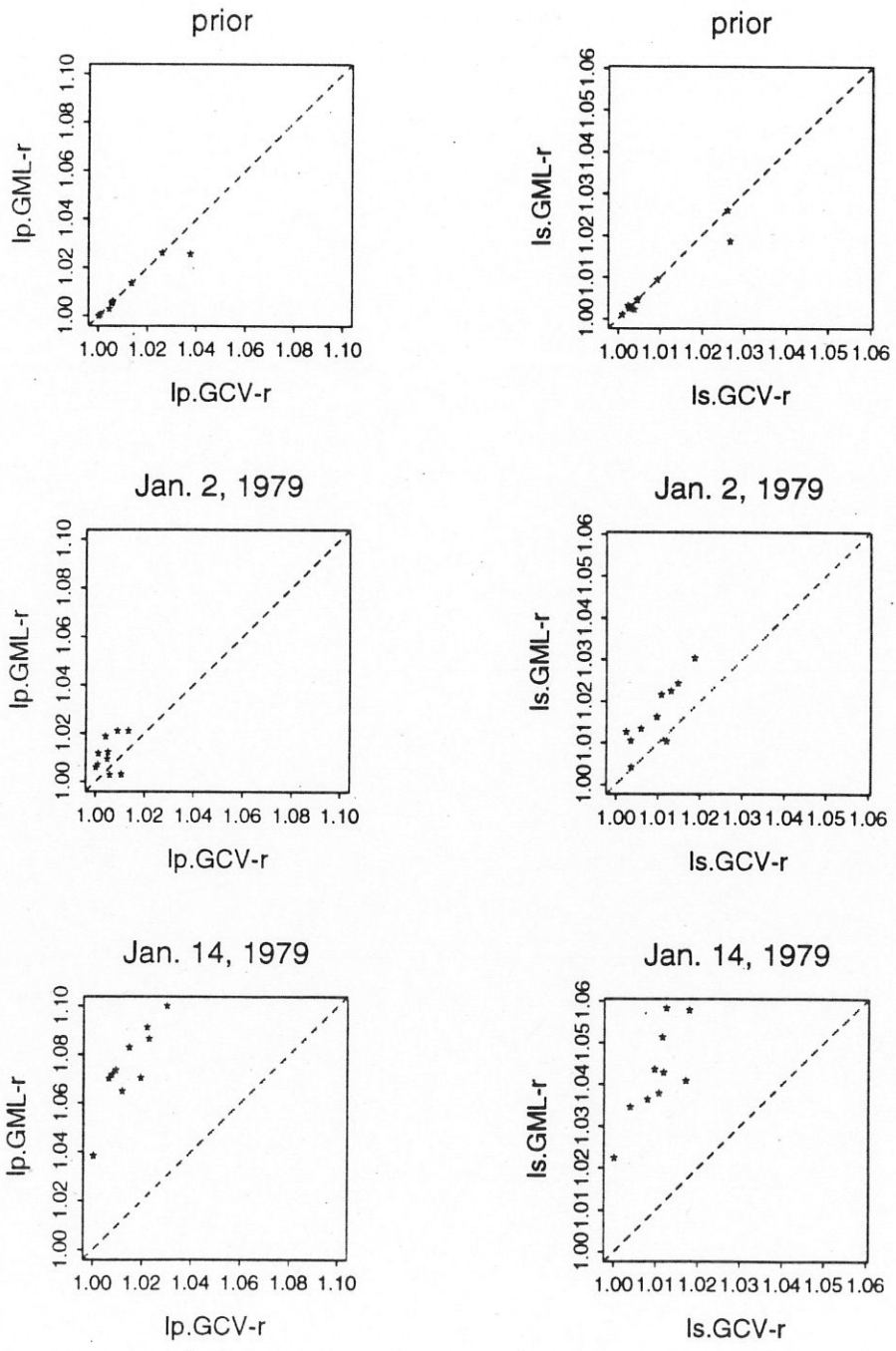
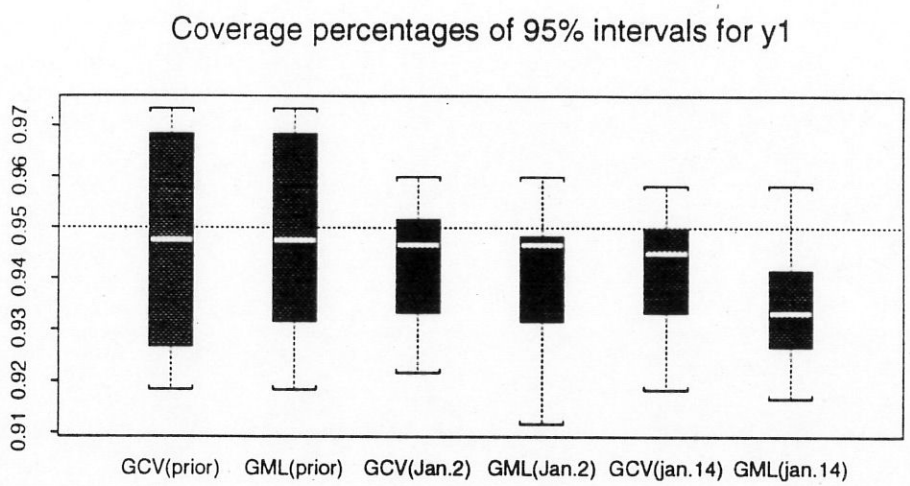


Figure 5.3.3: Relative inefficiencies for Experiment 1.



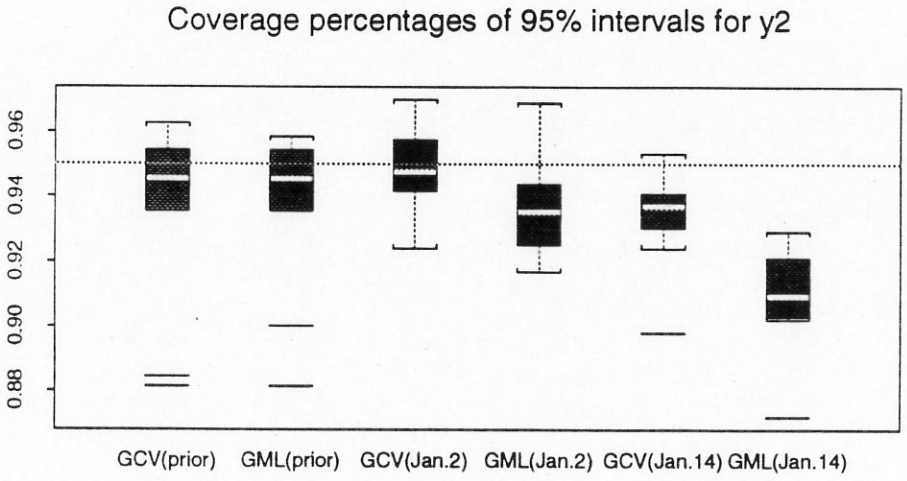


Figure 5.3.4: Bayesian "confidence intervals" for Experiment 1.

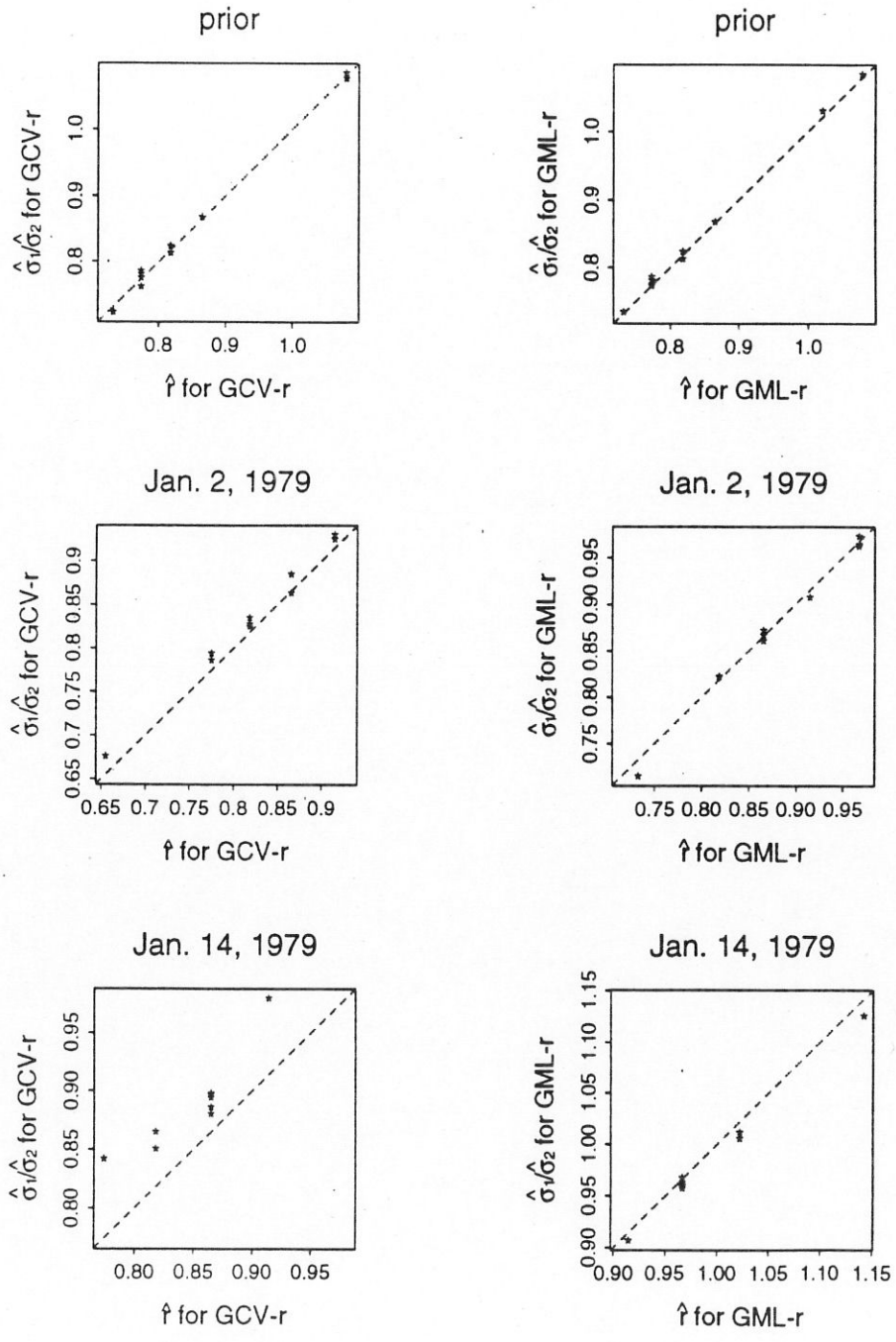


Figure 5.3.5: $\hat{\sigma}_1/\hat{\sigma}_2$ vs. $\hat{r}$ for Experiment 1.

Experiment 2

Figure 5.3.6 shows the GCV-r and GML-r estimates of r compared with the "optimal estimate" of r with respect to T_p. We see from the plot that when there is misspecification of m, GML-r estimate is biased, especially for "data" of prior and January 2, 1979. GCV-r gives very good unbiased estimate of r.

Figure 5.3.7 shows the GCV-r and GML-r estimates of r compared with the "optimal estimate" of r with respect to T_s. We see the same pattern as in Figure 5.3.6. The GCV-r is very robust against this misspecification of m and does a wonderful job estimating the weighting parameter r.

Figure 5.3.8 shows the relative inefficiencies of the GCV-r and GML-r estimates with respect to T_p and T_s. We see that when there is misspecification of m, the GCV-r is superior to the GML-r in estimating the surface f, especially for "data" of prior and January 2, 1979.

From these plots, we see that when we use the model with $m = 3$, GML-r gives better results than it does in experiment 1. This is perhaps due to the fact that the 500 mb height surface for January 14 is smoother than that for January 2. The model with $m = 3$ is more likely to be the correct model for January 14.

Figure 5.3.9 shows the box-plot of the 10 replicates of the coverage percentages of the 95% Bayesian "confidence intervals" on data points. The coverages for the GCV-r estimates are very close to 95% even in the presence of misspecification of m. The coverages for the GML-r estimates for y_2 are somewhat off the 95% line due to the misspecification of m here.

As before, we summarize the improvements after combining observational and forecast data according to our methods in the following two tables. From table 5.3.3, it is seen that on average, the improvement over direct observations is about 4 meters. From table 5.3.4, it is seen that on average, the improvement over forecasts is about 3 meters.

Table 5.3.3: RMSE (in meters) over 600 stations for Experiment 2

	January 2, 1979			January 14, 1979		
Runs	Direct Obs.	GCV-anal.	GML-anal.	Direct Obs.	GCV-anal.	GML-anal.
1	8.8386	4.3906	4.5502	8.7220	4.0981	4.1494
2	8.5642	4.4181	4.8440	9.5355	4.6311	4.6958
3	8.8627	4.3747	4.5852	8.6806	4.5952	4.6105
4	9.2166	4.4955	4.5579	8.7869	4.7389	4.7580
5	8.8304	4.4640	4.6788	8.9777	4.5467	4.5795
6	8.7973	4.5435	4.8005	9.0591	4.2060	4.2913
7	9.0400	4.1128	4.3204	8.7507	4.2530	4.2907
8	9.2087	4.1456	4.3593	8.7572	4.4526	4.4903
9	8.9084	4.3683	4.5945	8.7846	4.6438	4.7048
10	8.9830	4.1022	4.5091	8.7435	4.4902	4.4912

Table 5.3.4: RMSE (in meters) over 1224 grid points for Experiment 2

	January 2, 1979			January 14, 1979		
Runs	Forecast	GCV-anal.	GML-anal.	Forecast	GCV-anal.	GML-anal.
1	10.3307	6.7793	7.2951	9.7993	6.9404	6.9724
2	10.6885	7.0947	7.9308	9.7790	7.2890	7.3936
3	9.4040	6.9808	7.3111	10.7975	7.7539	7.8061
4	11.0332	6.9393	7.1297	9.3461	7.4901	7.5049
5	10.8553	7.0981	7.5949	9.5822	7.1670	7.1915
6	9.9805	7.3584	7.9703	10.0343	6.8072	6.9225
7	9.4557	6.6911	7.1664	9.6213	6.8300	6.8486
8	9.9679	6.9217	7.3261	10.8982	7.2099	7.2851
9	10.5214	7.2824	7.8143	9.6711	7.0495	7.1456
10	9.5974	6.8705	7.7243	9.9840	6.7023	6.7054

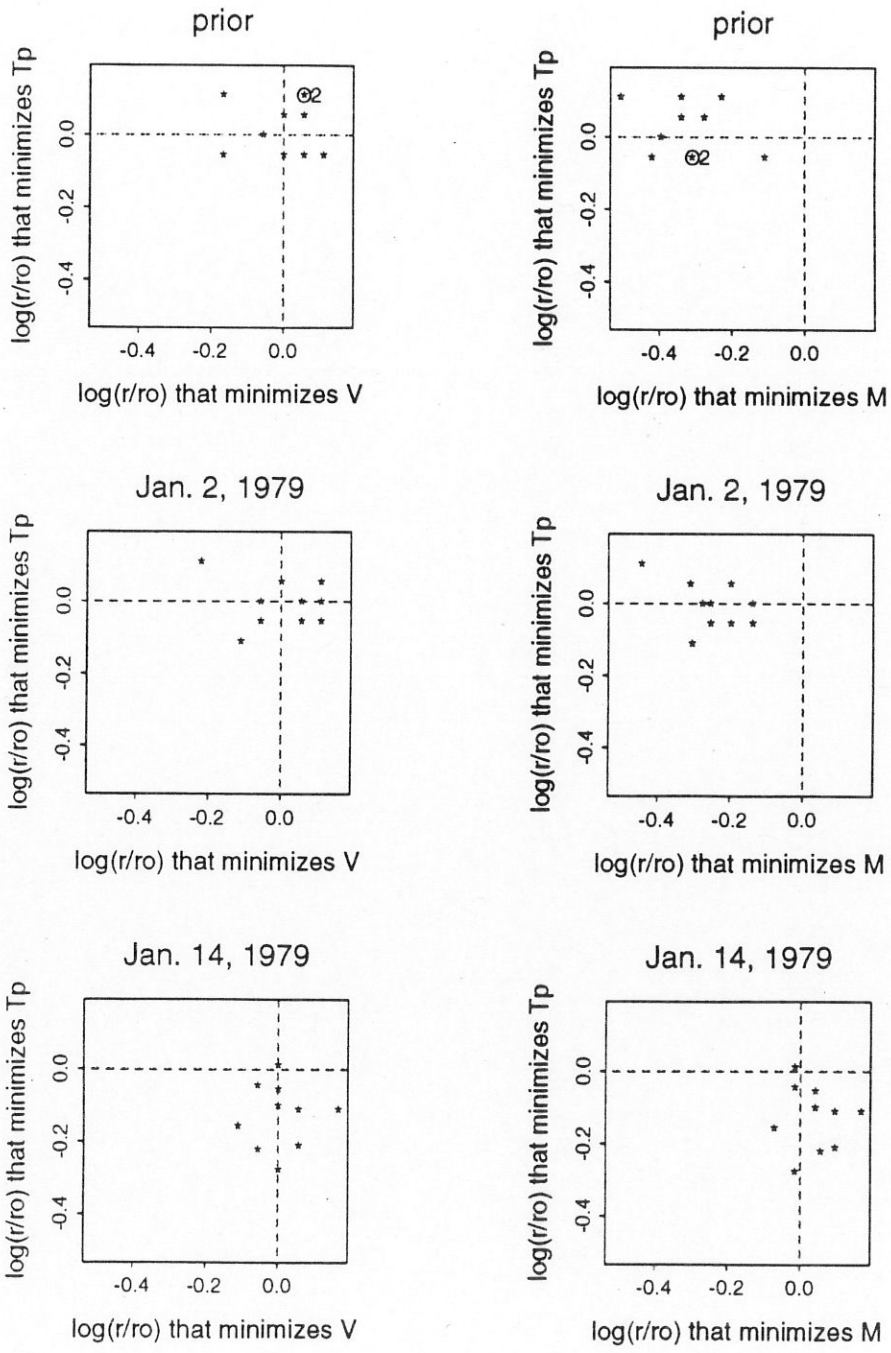


Figure 5.3.6: Estimates of r for Experiment 2.

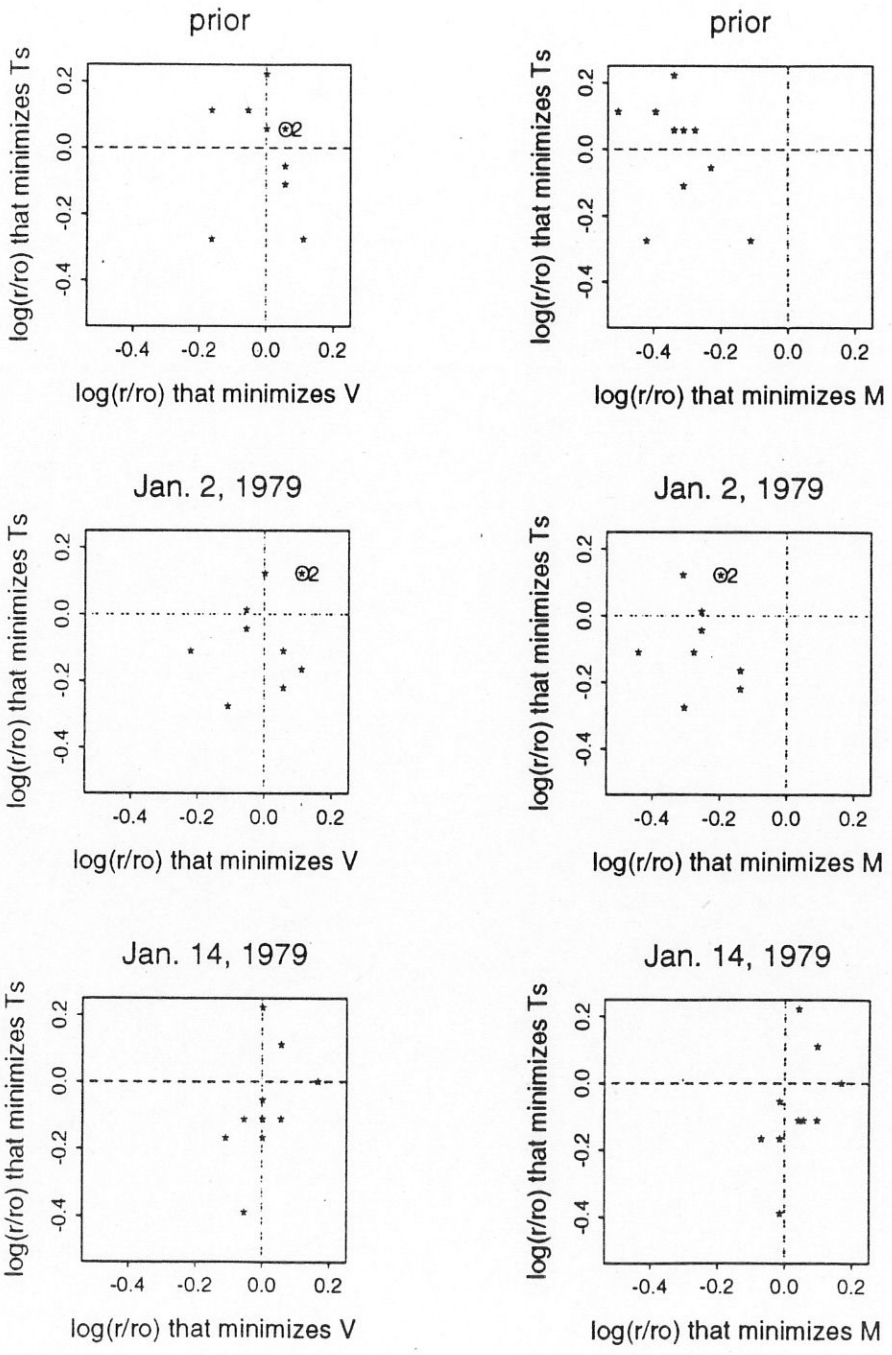


Figure 5.3.7: Estimates of r for Experiment 2.

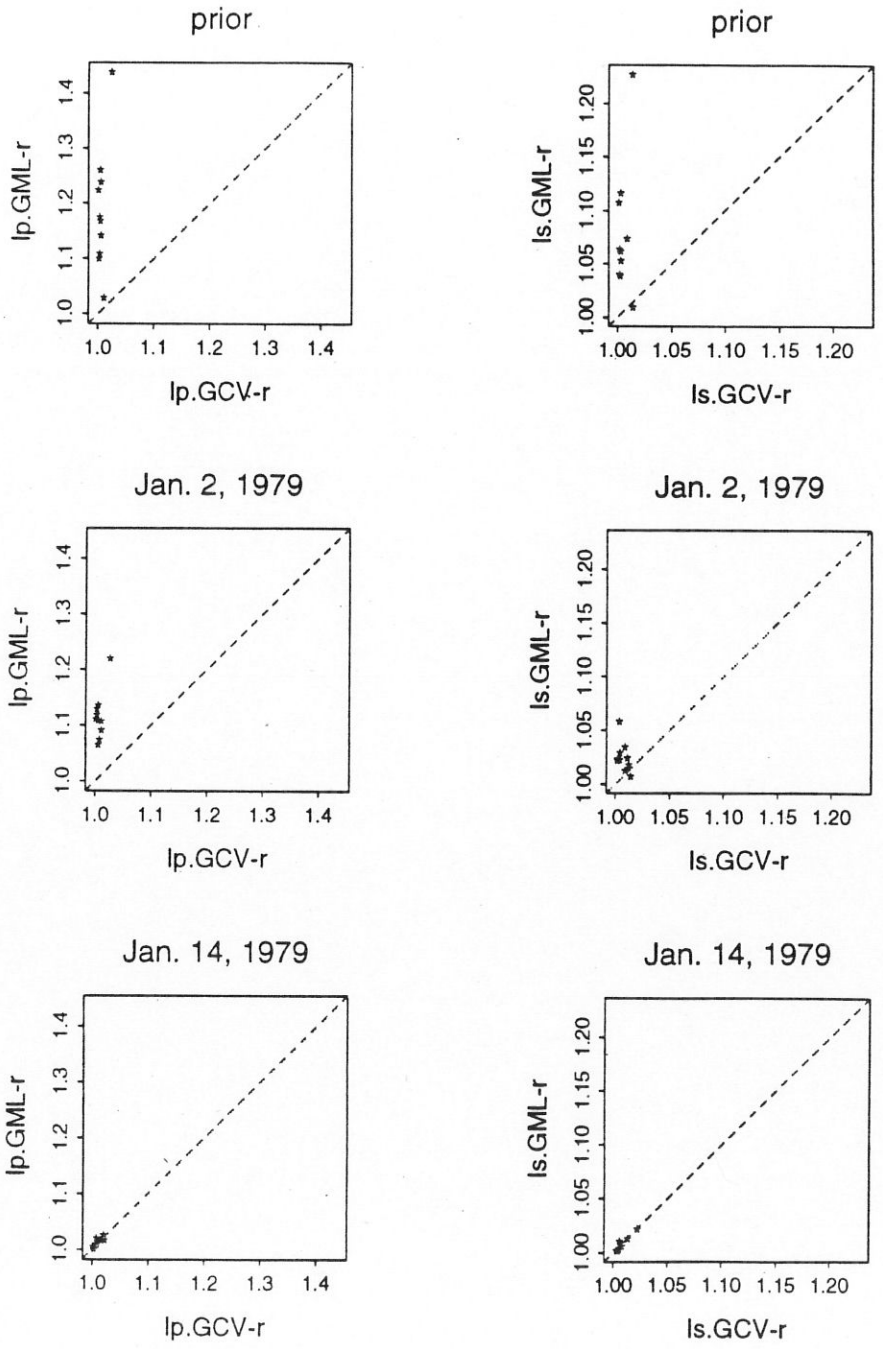
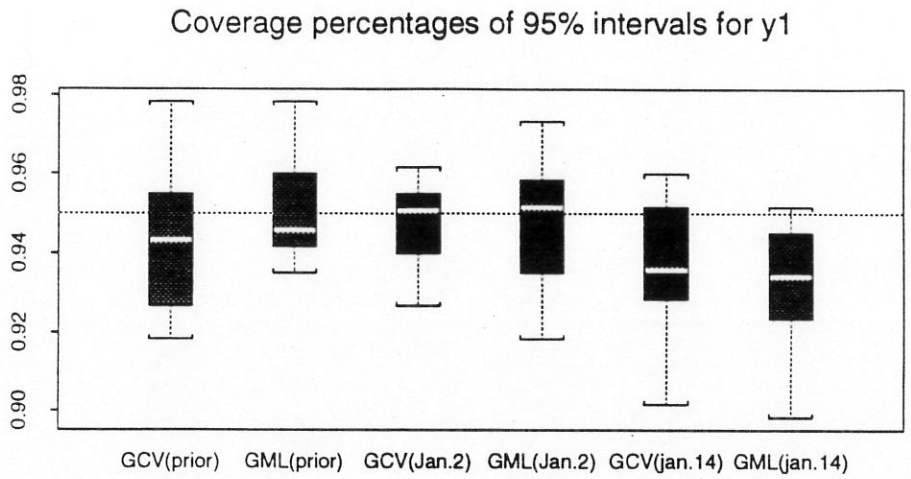


Figure 5.3.8: Relative inefficiencies for Experiment 2.



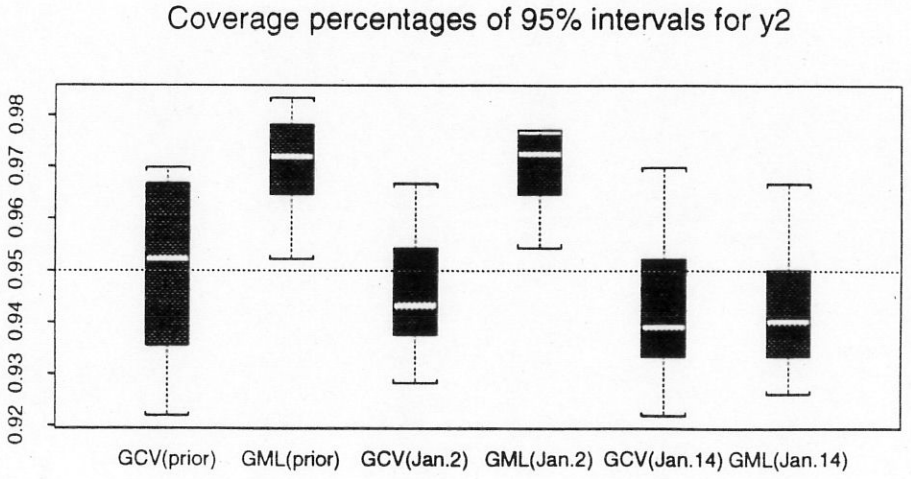


Figure 5.3.9: Bayesian "confidence intervals" for Experiment 2.

Experiment 3

Figure 5.3.10 shows the GCV-r and GML-r estimates of r compared with the "optimal estimate" of r with respect to T_p. We see from the plot that the misspecification of L does not severely affect the performance of GCV-r and GML-r estimate. As in Experiment 1, where there is no misspecification, GML-r is somewhat biased for cases when the "data" are the simulated meteorological truths, while GCV-r gives good unbiased estimates.

Figure 5.3.11 shows the GCV-r and GML-r estimates of r compared with the "optimal estimate" of r with respect to T_s. We see the same pattern as in Figure 5.3.10. The GCV-r does a wonderful job estimating the weighting parameter r.

Figure 5.3.12 shows the relative inefficiencies of the GCV-r and GML-r estimates with respect to T_p and T_s. We see that for "data" which simulate the truths of January 2 and January 14 of 1979, the GCV-r is superior to the GML-r in estimating the surface f.

Figure 5.3.13 shows the box-plot of the 10 replicates of the coverage percentages of the 95% Bayesian "confidence intervals" on data points. The coverages for the GCV-r estimates are close to 95%.

Again, we summarize the improvements after combining observational and forecast data according to our methods in the following two tables. From table 5.3.5, it is seen that on average, the improvement over direct observations is about 4 meters. From table 5.3.6, it is seen that on average, the improvement over forecasts is about 3 meters.

Table 5.3.5: RMSE (in meters) over 600 stations for Experiment 3

	January 2, 1979			January 14, 1979		
Runs	Direct Obs.	GCV-anal.	GML-anal.	Direct Obs.	GCV-anal.	GML-anal.
1	9.3766	4.3756	4.3592	8.8903	4.3981	4.3334
2	8.8351	4.4536	4.4268	8.5986	4.3936	4.3229
3	9.2760	4.2741	4.2890	8.7098	4.2559	4.3656
4	9.0247	4.5392	4.5291	9.4253	4.9217	4.8307
5	8.9753	4.1676	4.1593	9.0724	4.8084	4.9007
6	8.9165	4.2158	4.2225	8.8741	4.0733	4.0717
7	8.9868	4.5911	4.5761	9.2384	4.2192	4.1985
8	8.9059	4.6259	4.6076	8.8816	4.3377	4.2514
9	9.2518	4.5897	4.4935	8.6191	4.5288	4.3735
10	8.8372	4.6560	4.6417	9.0085	4.2838	4.2788

Table 5.3.6: RMSE (in meters) over 1224 grid points for Experiment 3

	January 2, 1979			January 14, 1979		
Runs	Forecast	GCV-anal.	GML-anal.	Forecast	GCV-anal.	GML-anal.
1	10.2090	6.7204	6.7045	9.2822	6.8210	6.5604
2	9.9385	6.8875	6.7907	9.5612	6.5685	6.3466
3	9.3592	7.4215	7.4660	8.8572	6.4069	6.4611
4	9.7465	6.7642	6.6507	10.7349	7.7651	7.6538
5	8.9699	6.6764	6.6129	9.8608	7.3295	7.3687
6	9.6728	6.9955	7.0180	9.0353	7.2531	7.0311
7	9.0908	6.9102	6.8964	8.4513	6.9263	6.3583
8	9.3835	6.6338	6.5825	10.0180	6.8058	6.4347
9	9.8800	6.8831	6.8291	8.9485	6.8253	6.5353
10	10.0739	7.0206	6.9639	9.1386	6.5923	6.4844

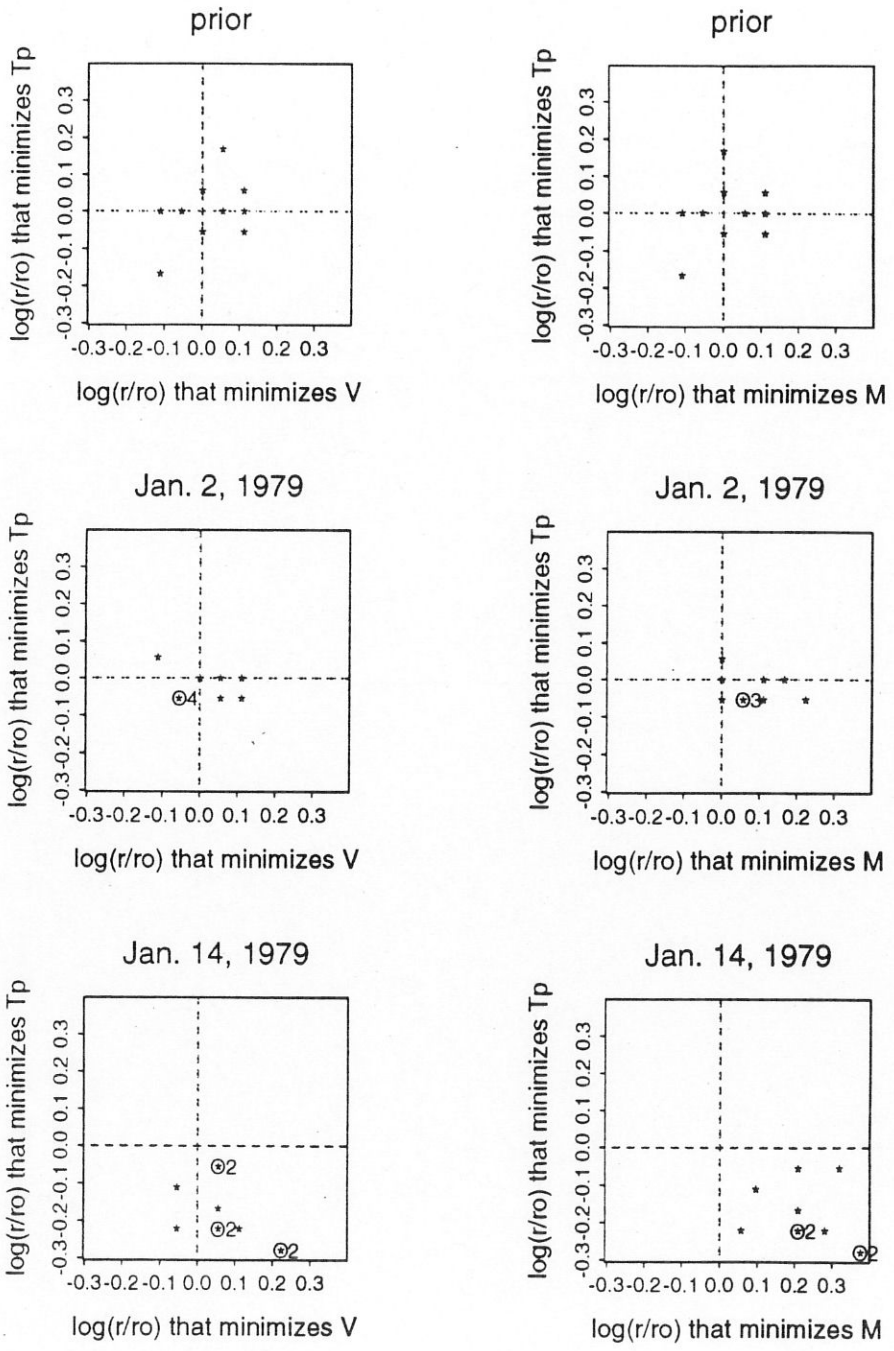


Figure 5.3.10: Estimates of r for Experiment 3.

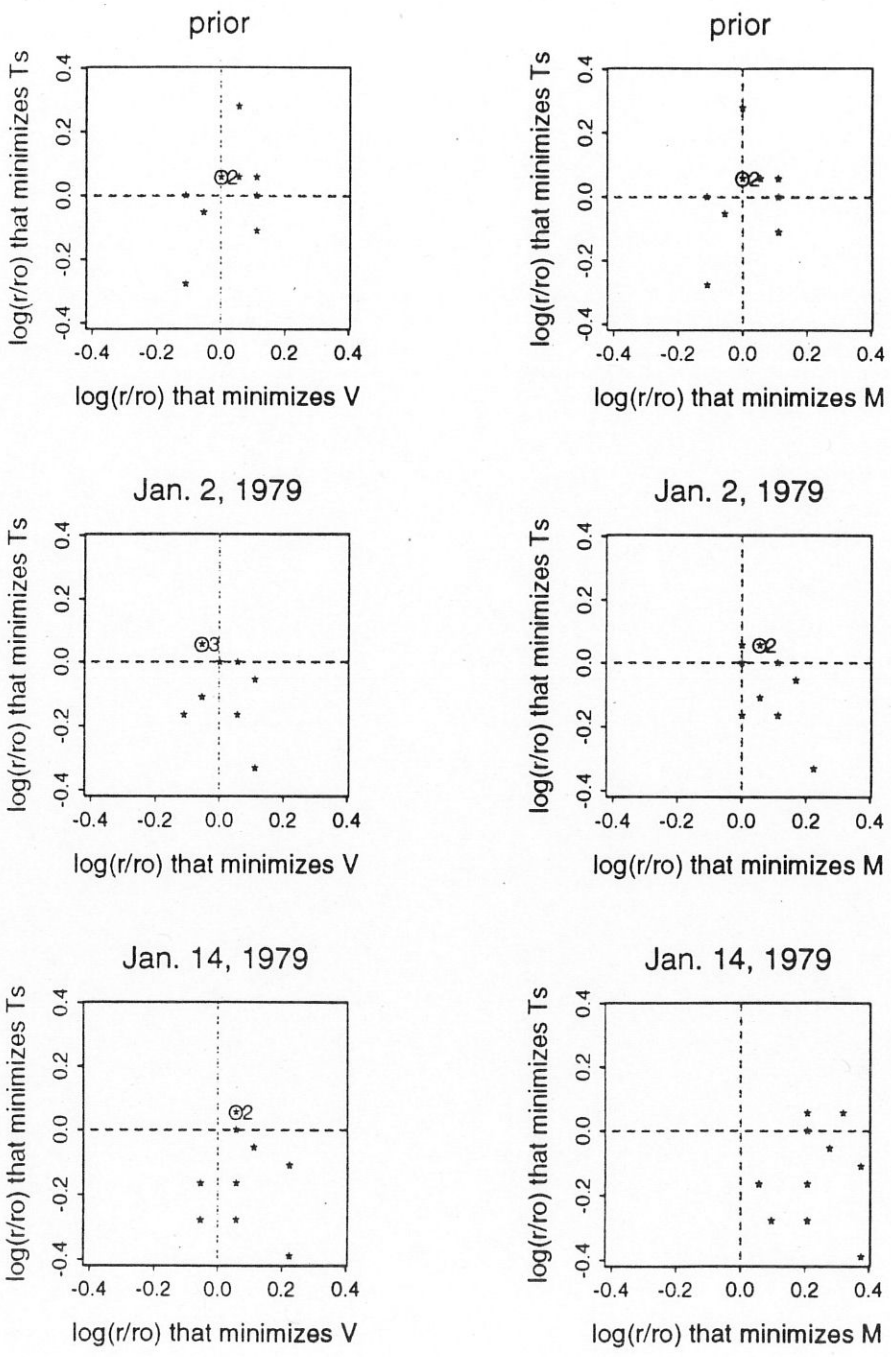


Figure 5.3.11: Estimates of r for Experiment 3.

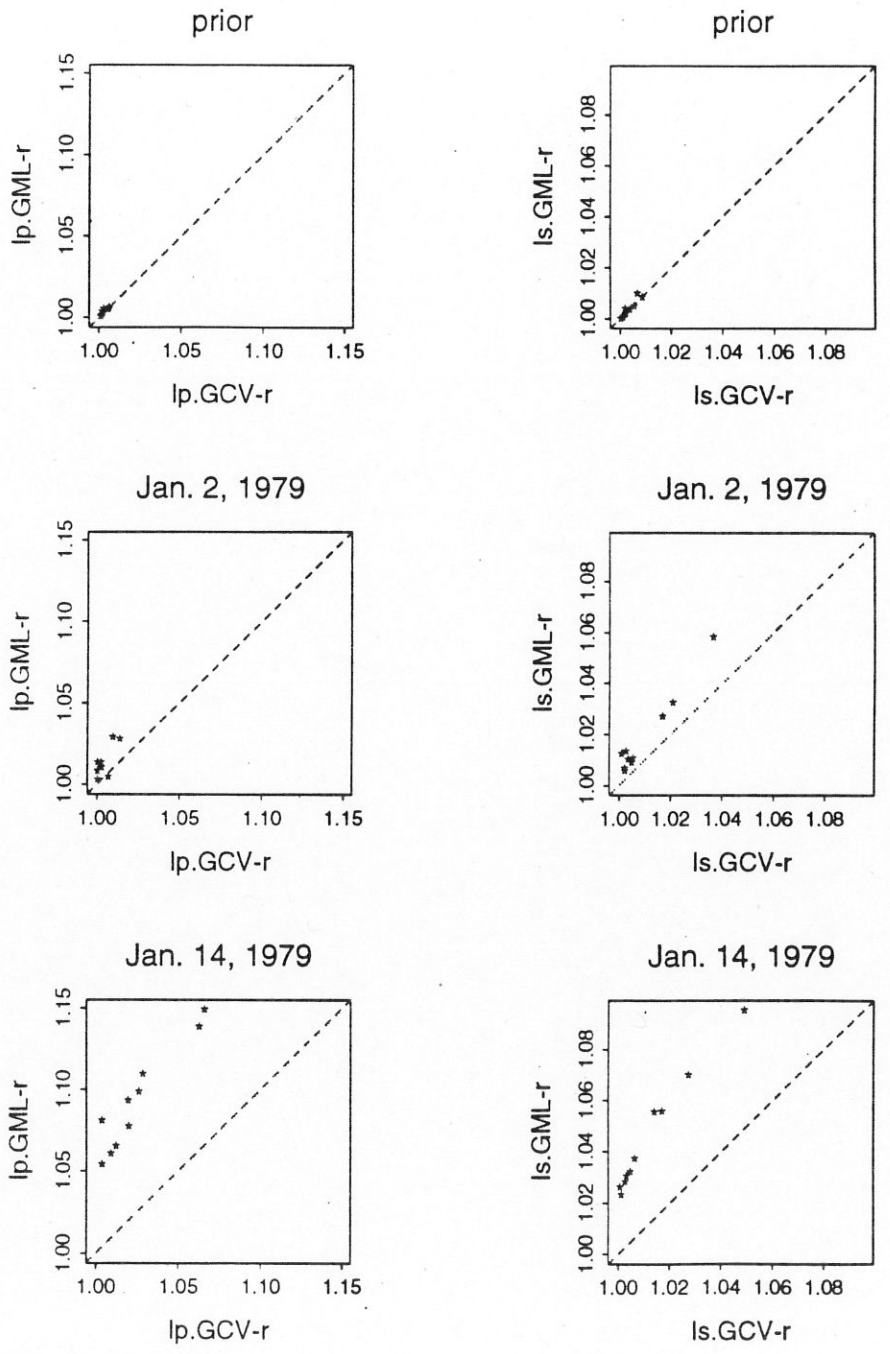
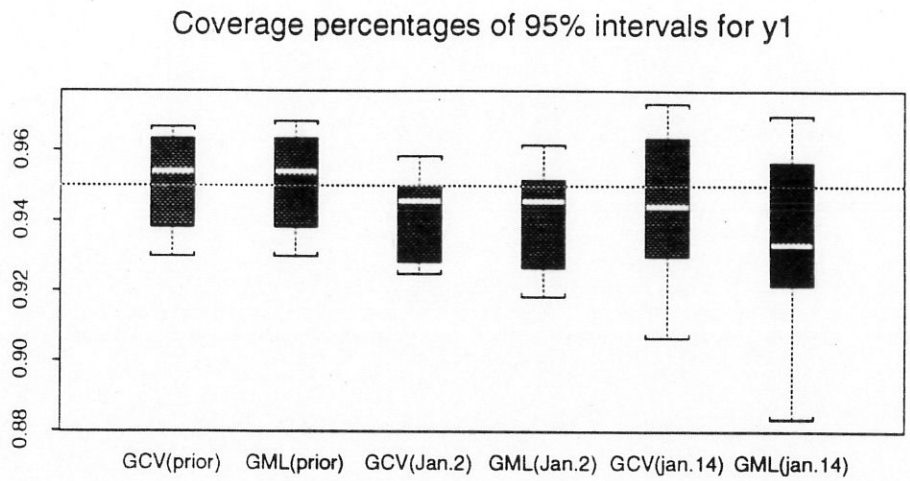


Figure 5.3.12: Relative inefficiencies for Experiment 3.



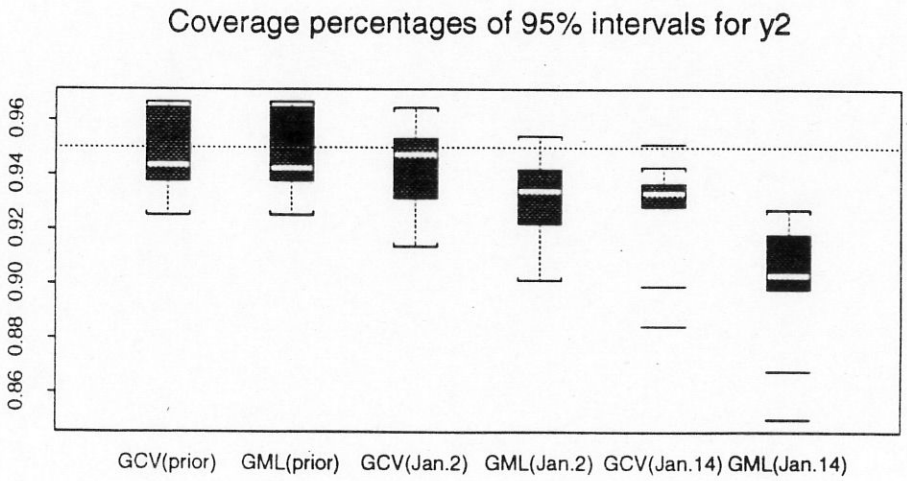


Figure 5.3.13: Bayesian "confidence intervals" for Experiment 3.

Chapter 6

Concluding Remarks

6.1 Summary

In Chapter 2, we introduce a new GCV estimate for simultaneously estimating the weighting parameter and the smoothing parameter in the model described in Chapter 1. We name this estimate GCV-r, where r represents the weighting parameter. We also describe two other methods of estimating the weighting and smoothing parameters simultaneously, namely the GML-r estimate and the RCV estimate.

In Chapter 3, we study the properties of these estimators. We prove the weak consistency and asymptotic normality of these estimators, and obtain the convergence rates for them under some conditions. In particular, we show that the convergence rates for the GML-r estimator $\hat{r}_{gml}$ and the GCV-r estimator $\hat{r}_{gcv}$ are the same, and it is $1/\sqrt{n}$. We also show that the convergence rate for the RCV estimator $\hat{r}_{rcv}$ is much slower. For example, in case 1 if we assume $m = 2$ and $p = 0$, the convergence rate for $\hat{r}_{rcv}$ would be $1/n^{1/8}$. So theoretically, GCV-r is better than RCV for estimating the weighting parameter r. In Chapter 3, we also give a weak convergence theorem for the GCV-r estimate.

In Chapters 4 and 5, we present some simulation results. From these results, we found that the GCV-r estimator $\hat{r}_{gcv}$ is better than the RCV estimator $\hat{r}_{rcv}$ in that $\hat{r}_{gcv}$ has a smaller sample variance than $\hat{r}_{rcv}$ does. From the two Monte Carlo studies that used the "true" 500 mb heights of January 2 and January 14 of 1979 as the "truths" in Chapter 5, we also found that the GCV-r estimator $\hat{r}_{gcv}$ not only has smaller sample variance, but also gets rid of the bias that inflicted the RCV estimator $\hat{r}_{rcv}$.

In the simulation studies in Chapters 4 and 5, we also compared the GCV-r with the GML-r. We found that when the stochastic model is right, the GCV-r appears to do equally well as the GML-r in estimating the weighting parameter r, and the GML-r is better than the GCV-r in estimating the unknown function f. When there is model misspecification, which is often inevitable in practice, the GCV-r gets slightly better results than the GML-r in estimating r, and the GCV-r gets clearly better results than the GML-r in estimating f. So the GCV-r appears to be more robust against departures from the stochastic model than the GML-r.

6.2 Future Research

In order to implement our method of combining data in real life, an efficient algorithm needs to be developed. In our simulation studies in Chapters 4 and 5, since the original goal was to see the performances of GML-r, GCV-r and RCV, and to compare these methods under different situations, not much effort had been devoted to developing an efficient algorithm. Therefore the computations in those simulations were quite costly, especially for the simulations in Chapter 5. To evaluate GML-r, GCV-r and RCV functions in Chapter 5, for every value of r, we had to do an eigenvalue decomposition of a 961×961 symmetric matrix. This was the most expensive part and took about 40 seconds on Cray Y-MP by evoking EISPACK, since at the time when these experiments were carried out, the eigenvalue part of LAPACK had not been installed on the Cray's at San Diego Supercomputer Center and at NASA Supercomputer center. To search for the minimizers of GML-r, GCV-r and RCV functions, we used a grid search. On average, it took about 16 minutes on Cray Y-MP to compute one estimate of (r, α) of GML-r or GCV-r or RCV for the problem in Chapter 5, which was a much reduced version of a real problem. In meteorology, the M in Chapter 5 is usually somewhere around 126, instead of 30. So the number of spherical harmonic coefficients is in the range of 16000, instead of 961. And the number of data in meteorology is usually in the order of 10^6. So to develop an efficient algorithm to compute these estimates is very important.

In real life, due to human error or other reasons, it happens quite often that in a large data set, there will be a few "bad" data. To be able to detect these "bad" data is very important. By checking residuals of the fits from our model, it is possible to detect those "bad" data. It is very interesting to find out how much can this method detect through

some simulation studies. Also, this method might be used to detect system error for a forecast model. We will carry out some Monte Carlo studies to find out these.

Appendix A

More Details for Proofs in Chapter 3

We will give some more details of the proofs in Chapter 3.

Proof of Lemma 3.1.1: Suppose $y \sim \mathcal{N}(0, V(\boldsymbol{\theta}))$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$, $y = (y_1, \ldots, y_n)'$. Let $\lambda(\boldsymbol{\theta}) = \log f_{\boldsymbol{\theta}}(y)$, where $f_{\boldsymbol{\theta}}(y)$ is the density of y. Also let $V_i = \partial V/\partial\theta_i$, $\lambda_i = \partial\lambda/\partial\theta_i$, $i = 1, \ldots, k$, etc. If we ignore a constant, we have

$$\lambda(\boldsymbol{\theta}) = -\frac{1}{2}\log\det V - \frac{1}{2}y'V^{-1}y.$$

By definition, the Fisher information is $\mathcal{I}(\boldsymbol{\theta}) = (\mathcal{I}_{ij})_{k\times k}$, where $\mathcal{I}_{ij} = E\lambda_i\lambda_j$.

It is not hard to see that

$$\lambda_i = \frac{\partial}{\partial\theta_i}\lambda = -\frac{1}{2}tr(V^{-1}V_i) + \frac{1}{2}y'V^{-1}V_iV^{-1}y.$$

Let $Q_i = V^{-1}V_iV^{-1}$, $S_i = tr(V^{-1}V_i)$. Then $Ey'Q_iy = S_i$ and

$$\mathcal{I}_{ij} = \frac{1}{4}E(y'Q_iy - S_i)(y'Q_jy - S_j) = \frac{1}{4}E(y'Q_iyy'Q_jy - S_iS_j).$$

Also let $x = V^{-1/2}y$, $J_i = V^{-1/2}V_iV^{-1/2}$. Then $x \sim \mathcal{N}(0, I)$, $y'Q_iy = x'J_ix$ and $tr(J_i) = S_i$.

For $x \sim \mathcal{N}(0, I)$, and any $n \times n$ matrices A and B, it is easy to see that

$$E(x'Axx'Bx) = tr(A)tr(B) + 2tr(AB).$$

So, we have

$$\mathcal{I}_{ij} = \frac{1}{2} tr(V_i V^{-1} V_j V^{-1}). \tag{A.1}$$

Now for our model, $\theta = (\theta, r, \alpha)'$, $z \sim \mathcal{N}(0, \theta v(r, \alpha))$ as in (2.1.4) and z is of dimension $2n - 1$.

Let $\tilde{\Sigma}$, $\tilde{M}$ and $\tilde{K}_i$, $i = 1, 2$ be defined as

$$\Sigma = \begin{pmatrix} \xi_1 & 0 \\ 0 & \tilde{\Sigma} \end{pmatrix},$$

$$M = \begin{pmatrix} (1/r)k_{10}^2 + rk_{20}^2 & 0 \\ 0 & \tilde{M} \end{pmatrix},$$

and

$$K_i = \begin{pmatrix} k_{i0} & 0 \\ 0 & \tilde{K}_i \end{pmatrix},$$

$i = 1, 2$. Then it is not hard to show that

$$v(r, \alpha) = \begin{pmatrix} \frac{1}{\alpha}\tilde{K}_1\tilde{\Sigma}\tilde{K}_1' + rI_{n-1} & 0 & \frac{1}{\alpha}\tilde{K}_1\tilde{\Sigma}\tilde{K}_2' \\ 0 & (\frac{1}{r}k_{10}^2 + rk_{20}^2)/(k_{10}^2 + k_{20}^2) & 0 \\ \frac{1}{\alpha}\tilde{K}_2\tilde{\Sigma}\tilde{K}_1' & 0 & \frac{1}{\alpha}\tilde{K}_2\tilde{\Sigma}\tilde{K}_2' + \frac{1}{r}I_{n-1} \end{pmatrix},$$

and

$$v^{-1}(r, \alpha) = \begin{pmatrix} \frac{1}{r}I_{n-1} & 0 & 0 \\ 0 & (k_{10}^2 + k_{20}^2)/(\frac{1}{r}k_{10}^2 + rk_{20}^2) & 0 \\ 0 & 0 & rI_{n-1} \end{pmatrix}$$

$$- \begin{pmatrix} (1/r^2)\tilde{K}_1\tilde{M}^{-1}\tilde{K}_1' & 0 & \tilde{K}_1\tilde{M}^{-1}\tilde{K}_2' \\ 0 & 0 & 0 \\ \tilde{K}_2\tilde{M}^{-1}\tilde{K}_1' & 0 & r^2\tilde{K}_2\tilde{M}^{-1}\tilde{K}_2' \end{pmatrix}.$$

Applying (A.1) to our v and v^{-1}, we will get the results in the theorem. We only derive

the formulas for $\mathcal{I}_{\theta\theta}$ and $\mathcal{I}_{rr}$ here as demonstrations, the formulas for other entries can be obtained similarly.

Since $v(r,\alpha)$ does not depend on θ, and V in (A.1) equals $\theta v(r,\alpha)$ in our model, it is easy to see

$$\mathcal{I}_{\theta\theta} = \frac{1}{2}tr(\frac{1}{\theta^2}I_{2n-1}) = \frac{1}{2\theta^2}(2n-1).$$

By straightforward calculation, it is easy to show that

$$v^{-1}\frac{\partial v}{\partial r} =$$

$$\begin{pmatrix} \frac{1}{r}(I_{n-1} - \frac{1}{r}\tilde{K}_1\tilde{M}^{-1}\tilde{K}_1') & 0 & -\tilde{K}_1\tilde{M}^{-1}\tilde{K}_2' \\ 0 & (-\frac{1}{r}k_{10}^2 + rk_{20}^2)/(k_{10}^2 + r^2k_{20}^2) & 0 \\ (1/r^2)\tilde{K}_2\tilde{M}^{-1}\tilde{K}_1' & 0 & -\frac{1}{r}I_{n-1} + \tilde{K}_2\tilde{M}^{-1}\tilde{K}_2' \end{pmatrix}.$$

Then,

$$\mathcal{I}_{rr} = \frac{1}{2}tr(v^{-1}\frac{\partial v}{\partial r}v^{-1}\frac{\partial v}{\partial r})$$

$$= \frac{1}{2r^2}(2n + tr[(\frac{1}{r}K_1'K_1M^{-1} - rK_2'K_2M^{-1})^2 - 2(\frac{1}{r}K_1'K_1M^{-1} + rK_2'K_2M^{-1})]).$$

□

Proof of Lemma 3.1.2: We only prove i). The rest can be proved similarly.

Let

$$S_n = \sum_{i=1}^{n} \frac{n^{l_1}i^{l_2q}}{(n+i^q)^{l_1+l_2}} = \sum_{i=1}^{[n^{1/q}]} \frac{n^{l_1}i^{l_2q}}{(n+i^q)^{l_1+l_2}} + \sum_{i=[n^{1/q}]+1}^{n} \frac{n^{l_1}i^{l_2q}}{(n+i^q)^{l_1+l_2}},$$

where $[x]$ means the integer part of x.

Let $a_n \sim b_n$ mean there exists a constant c such that as $n \to \infty$, $a_n/b_n \to c$. Then we have

$$S_n \sim \sum_{i=1}^{[n^{1/q}]} \frac{n^{l_1}i^{l_2q}}{n^{l_1+l_2}} + \sum_{i=[n^{1/q}]+1}^{n} \frac{n^{l_1}i^{l_2q}}{i^{q(l_1+l_2)}}$$

$$= \sum_{i=1}^{[n^{1/q}]} \left(\frac{i^q}{n}\right)^{l_2} + \sum_{i=[n^{1/q}]+1}^{n} \left(\frac{n}{i^q}\right)^{l_1}.$$

Now

$$\sum_{i=1}^{[n^{1/q}]} \left(\frac{i^q}{n}\right)^{l_2} \leq \sum_{i=1}^{[n^{1/q}]} 1 = O(n^{1/q}),$$

and

$$\begin{aligned}
\sum_{i=[n^{1/q}]+1}^{n} \left(\frac{n}{i^q}\right)^{l_1} &= n^{l_1-ql_1+1} \sum_{i=[n^{1/q}]+1}^{n} \frac{1}{(i/n)^{ql_1}} \frac{1}{n} \\
&\sim n^{l_1-ql_1+1} \int_{\frac{n^{1/q}}{n}}^{1} \frac{1}{x^{ql_1}} dx \\
&\sim \begin{cases} n^{l_1-ql_1+1}(n^{(1-ql_1)/q}/n^{1-ql_1} - 1), & l_1 > 0 \\ O(n), & l_1 = 0 \end{cases} \\
&= \begin{cases} O(n^{1/q}), & l_1 > 0 \\ O(n), & l_1 = 0. \end{cases}
\end{aligned}$$

The last equality is based on the fact that for $q > 1$ and $l_1 \geq 1$ (since l_1 is an integer, $l_1 > 0$ implies $l_1 \geq 1$), we always have $l_1 - ql_1 + 1 < 1/q$. The result of i) in the lemma then follows. □

Proof of Lemma 3.1.3: We will only show $\mathcal{I}_{\theta r} = O(n^{1/2(m+p)})$, the rest can be proved similarly.

Let $a_n \sim b_n$ mean there exists a constant c such that as $n \to \infty$, $a_n/b_n \to c$. By Lemma 3.1.1, and by the assumptions made on K_1 and K_2, we have

$$\begin{aligned}
\mathcal{I}_{\theta r} &= \frac{1}{2\theta r} tr[-\frac{1}{r} K_1' K_1 M^{-1} + r K_2' K_2 M^{-1}] \\
&\sim \sum_{i=1}^{n} \frac{-\frac{1}{r} n i^{-2p_1} + r n i^{-2p_2}}{\frac{1}{r} n i^{-2p_1} + r n i^{-2p_2} + \alpha i^{2m}} \\
&\sim \sum_{i=1}^{n} \frac{n i^{-2p}}{n i^{-2p} + i^{2m}} \\
&= \sum_{i=1}^{n} \frac{n}{n + i^{2m+2p}}.
\end{aligned}$$

Then we can apply i) of Lemma 3.1.2 to the sum by letting $l_1 = 1$, $l_2 = 0$ and $q = 2m + 2p$ to obtain $\mathcal{I}_{\theta r} = O(n^{1/2(m+p)})$. □

Proof of Theorem 3.1.1: We can apply Theorem 1 of Mardia and Marshall (1984)

directly. In Theorem 1 of Mardia and Marshall (1984), sufficient conditions are given for the asymptotic normality and weak consistency of maximum likelihood estimators. Those conditions are (3.1), (3.3) and (3.4) there. We do not need to check (3.4) because the mean function in our model is identically 0. (3.1) is $\lim_{n\to\infty} \mathcal{I}^{-1} = \mathbf{0}$, which is true by noticing Lemma 3.1.3. So we only need to check condition (3.3).

Our model can be written as $z \sim \mathcal{N}(0, V(\theta_1, \theta_2, \theta_3))$, where z is of dimension $2n-1$, $\theta_1 = \theta$, $\theta_2 = r$, $\theta_3 = \alpha$ and $V = \theta v(r, \alpha)$, where $v(r, \alpha)$ appears in (2.1.4) and in the proof of Lemma 3.1.1. Let $\mathcal{I}^{-1} = (\mathcal{I}^{ij})_{3\times 3}$ and $V^{ij} = \partial^2 V^{-1}/\partial\theta_i\partial\theta_j$. Then condition (3.3) is

$$\lim_{n\to\infty} \sum_{i,j,k,l=1}^{3} \mathcal{I}^{ki}\mathcal{I}^{lj} tr(VV^{kj}VV^{li}) = 0.$$

We will only show that

$$\lim_{n\to\infty} \mathcal{I}^{12}\mathcal{I}^{12} tr(VV^{12}VV^{12}) = 0$$

as a demonstration. The other terms in the sum can similarly be shown to go to zero as n goes to infinity.

Since $V = \theta v(r, \alpha)$ and $v(r, \alpha)$ does not depend on θ, it is easy to see

$$VV^{12} = \theta v(r,\alpha)(-\frac{1}{\theta^2}\frac{\partial v^{-1}}{\partial r}) = -\frac{1}{\theta} v \frac{\partial v^{-1}}{\partial r} = \frac{1}{\theta}\frac{\partial v}{\partial r} v^{-1}.$$

By the last equation in the proof of Lemma 3.1.1, we have

$$tr(VV^{12}VV^{12}) = \frac{1}{\theta^2} tr(v^{-1}\frac{\partial v}{\partial r} v^{-1}\frac{\partial v}{\partial r}) = \frac{2}{\theta^2}\mathcal{I}_{rr}.$$

Now, by Lemma 3.1.3, we know $\mathcal{I}_{rr} = O(n)$ and $\mathcal{I}^{12} = \mathcal{I}^{\theta r} = O(n^{-2+1/2(m+p)})$. Since $m \geq 1$, $p \geq 0$, we have

$$\lim_{n\to\infty} \mathcal{I}^{12}\mathcal{I}^{12} tr(VV^{12}VV^{12}) = 0.$$

□

Proof of Lemma 3.2.1 and Lemma 3.2.2: We will need the following two results about multivariate normal distribution:

Suppose

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}(0, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}),$$

and let

$$x = (y_1, y_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Then

$$\begin{cases} Ey_1^4 = 3\sigma_{11}^2 \\ Ey_1^3 y_2 = 3\sigma_{11}\sigma_{12} \\ Ey_1^2 y_2^2 = \sigma_{11}\sigma_{22} + 2\sigma_{12}^2 \end{cases} \tag{A.2}$$

and

$$\begin{cases} E(x) = a_{11}\sigma_{11} + (a_{12} + a_{21})\sigma_{12} + a_{22}\sigma_{22} \\ Var(x) = 2(Ex)^2 + ((a_{12} + a_{21})^2 - 4a_{11}a_{22})(\sigma_{11}\sigma_{22} - \sigma_{12}^2). \end{cases} \tag{A.3}$$

(A.2) can be proved through the use of characteristic function of multivariate normal distribution (see Chapter 2 of Anderson (1984)). (A.3) can be obtained by using (A.2).

We will only prove the result for $\partial w_r/\partial r$ as a demonstration. For other terms, the results can be proved similarly. Note that $w_r(r, \alpha)$ can be written as

$$w_r(r, \alpha) = y' V_r y,$$

here V_r is a 2×2 block matrix with each block being an $n \times n$ diagonal matrix. It is easy to see that

$$w_{rr} = \frac{\partial}{\partial r} w_r = y' V_{rr} y,$$

where

$$V_{rr} = \begin{pmatrix} V_{rr}^{11} & V_{rr}^{12} \\ V_{rr}^{21} & V_{rr}^{22} \end{pmatrix},$$

and each block is an $n \times n$ diagonal matrix. It is also not hard to see that (y_{1i}, y_{2i}) and (y_{1j}, y_{2j}) are independent for $i \neq j$. So

$$w_{rr} = \sum_{i=1}^{n} (y_{1i}, y_{2i}) \begin{pmatrix} V_{rr}^{11}(i) & V_{rr}^{12}(i) \\ V_{rr}^{21}(i) & V_{rr}^{22}(i) \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \sum_{i=1}^{n} x_i,$$

where $V_{rr}^{11}(i)$ is the ith diagonal element of V_{rr}^{11} and so forth, and x_i, $i = 1, \ldots, n$ are

independent.

By Lemma 3.1.2 and Lemma 3.1.3, it is easy to see that

$$tr(I^{-1}(r)A^r(r,\alpha)I^{-1}(r)) = O(n^{1/(2m+2p)}),$$

and therefore

$$tr(W^{-1}(r,\alpha)) = tr(I^{-1}(r)[I - A^r]I^{-1}(r)) = O(n).$$

Similarly,

$$tr(\frac{\partial}{\partial r}W^{-1}) = O(n),$$

$$tr(\frac{\partial^2}{\partial r^2}W^{-1}) = O(n),$$

and so forth.

For any fixed r and α, by taking expectation with respect to y, and using Lemma 3.1.2 and Lemma 3.1.3, it can be shown that

$$E(w_{rr}) = O(n^2).$$

Also, by using (A.3) and Lemma 3.1.2 and Lemma 3.1.3, it can be shown that

$$Var(V_{rr}) = O(n^3).$$

So

$$E\left(\frac{w_{rr} - Ew_{rr}}{Ew_{rr}}\right)^2 = O(n^{-1}).$$

Therefore there is an $\epsilon > 0$, such that

$$E\left(\frac{w_{rr}}{Ew_{rr}} - 1\right)^2 n^{\epsilon} \to 0.$$

So

$$\left(\frac{w_{rr}}{Ew_{rr}} - 1\right) = o_p(n^{-\epsilon}),$$

that is, for any fixed r, and α,

$$w_{rr} = \frac{\partial}{\partial r}w_r = E\left(\frac{\partial}{\partial r}w_r\right)(1 + o_p(n^{-\epsilon})).$$

□

Proof of Lemma 3.2.3: We will only show that as $n \to \infty$,

$$n^{-3/2} w_r(r_0, \alpha_0) \xrightarrow{d} \mathcal{N}(0, c_1).$$

The other result can be obtained similarly.

First, we show that

$$E w_r(r_0, \alpha_0) = 0.$$

From (2.1.2), we know that $y \sim \mathcal{N}(0, \theta W)$. By definition

$$w_r(r, \alpha) = y'\{W^{-1}\frac{\partial W^{-1}}{\partial r} tr(W^{-1}) - W^{-2} tr(\frac{\partial W^{-1}}{\partial r})\}y.$$

So

$$\begin{aligned} E w_r(r, \alpha) &= tr\{(W^{-1}\frac{\partial W^{-1}}{\partial r} tr(W^{-1}) - W^{-2} tr(\frac{\partial W^{-1}}{\partial r}))\theta W\} \\ &= \theta tr(\frac{\partial W^{-1}}{\partial r} tr(W^{-1}) - W^{-1} tr(\frac{\partial W^{-1}}{\partial r})) \\ &= 0. \end{aligned}$$

Thus $w_r(r_0, \alpha_0)$ can be written as

$$w_r(r_0, \alpha_0) = \sum_{i=1}^{n} (x_{ri} - E x_{ri}).$$

Then by a similar method used in the proof of Lemma 3.2.1 and Lemma 3.2.2, it can be shown that

$$\sum_{i=1}^{n} E(x_{ri} - E x_{ri})^2 = O(n^3),$$

$$\sum_{i=1}^{n} E|x_{ri} - E x_{ri}|^3 = O(n^4).$$

So

$$\sum_{i=1}^{n} E(n^{-3/2}(x_{ri} - E x_{ri}))^2 = O(1),$$

$$\sum_{i=1}^{n} E|n^{-3/2}(x_{ri} - E x_{ri})|^3 = o(1).$$

By Liapounov's theorem, the result follows. □

Appendix B

Generation of "True" f's Used in the Experiments

From December 1, 1978 through December 31, 1979, there was an international experiment known as FGGE (First GARP Global Experiment, GARP stands for Global Atmospheric Research Program) or the "Global Weather Experiment", during which nearly every country on earth participated, including the USSR and China, in spite of the Cold War. They agreed to collect and analyze an unusually large quantity of meteorological data. See Johnson (1986) and Proceedings of the First National Workshop on the Global Weather Experiment: Current Achievements and Future Directions, National Academy Press, Washington D.C. 1985 for an overview of this experiment and further references. The European Center for Medium Range Weather Forecasting (ECMWF), in the so-called Level IIIB reanalysis, based on both weather assimilation models as well as all of the data available, provides estimates of the 500 mb heights on a regular latitude/longitude grid of 3.75×3.75 degrees. See Pailleux, Uppala, Illari and Dell'Osso (1986).

Considering only points on this latitude/longitude grid strictly north of the equator results in 2209 ($= 96 \times 23 + 1$) points. To obtain f as described in the text, these 2209 data points were selectively thinned down to 960 points by removing points in the more northerly latitude circles so that the points retained were very roughly relatively equally dense on the hemisphere, some trial and error was used until the resulting contour maps as obtained in the text visually reproduced the ECMWF reanalysis contour maps. The vector f from a particular set of 960 data points $\{y_i, P_i\}$ was obtained by choosing the f_{l_s}

to minimize

$$\sum_{i=1}^{960}(y_i - \sum_{l=0}^{30}\sum_{s=-l}^{l} f_{ls}Y_{ls}(P_i))^2 + \lambda\sum_{l=1}^{30}\sum_{s=-l}^{l}(l(l+1))^2 f_{ls}^2.$$

Since the data had already been smoothed it was not desireable to use GCV to choose λ, this parameter was chosen by inspection to make the resultant contour plots match the ECMWF contour plots visually. We thank Dr. Fred Reames for providing these coefficients, and Todd Schaack and Allen Lenzen for assistance in obtaining the ECMWF FGGE IIIb data.

We put the spherical harmonic coefficients for the 500 mb heights for January 2 and January 14, 1979, which we have used in our experiments, in Table B.1 and Table B.2. Spherical harmonic coefficients f_{ls}'s have double subscripts. In the tables, they appear in the following order: $(l,s) : (0,0),(1,0),(1,-1),(1,1),(2,0),(2,-1),$
$(2,1),(2,-2),(2,2),(3,0),\ldots$. The tables should be read column by column.

Table B.1: Spherical harmonic coefficients for 500 mb heights for January 2, 1979

19504.00544288	37.676192966	5.180566596697	11.25433722849	19.72984107771
377.4718957794	-18.56056671186	-42.2335213329	54.25878624753	-25.38948305174
-20.4264896064	-50.7321378285	3.336583643989	13.95906160787	-21.28513138395
14.51087388655	23.21934145164	17.08505080023	14.61849372596	13.93213650873
-1029.862412103	-11.77600595767	-30.36305395412	7.390236665758	-10.66328314941
-28.69054856137	-2.270496455975	11.04122315444	-47.84886621487	1.231806360952
38.91328167067	9.312406906333	-1.293447355269	4.857037327783	13.63664074829
-8.969050288963	2.078501887511	-4.966086396075	10.40607438949	-1.2669233868
-11.290790114	36.07365290495	-2.812225628695	-22.73243855399	-5.05789623446
-148.5582200685	-52.01782770612	-49.9407916886	27.05850282384	4.672363006433
-85.75554384169	-53.38066651005	-3.16642815616	-28.79902507181	-2.726259837793
42.87706900233	27.84192886653	10.00114787487	-25.78181960454	-10.01408876597
-10.77798111352	15.10883869353	62.51828870943	7.289025592191	6.785598281359
-5.199588391231	-61.87522476614	-12.24983864966	-14.81622438359	8.093146343054
-5.833745839871	24.44990679626	17.22211975406	0.9513268641478	20.19318049315
11.62288649153	-5.367072838153	9.202833597719	14.68540567021	2.97814563999
64.72346812467	-18.06213056125	-56.40294217856	-2.270411129553	0.8119838072883
-77.75630177886	1.084911672273	7.540531018694	-9.016519968654	-1.42593326909
-0.5874859748455	8.991309174045	17.57135241763	-1.39031620533	-8.581303541199
24.28847576245	-18.64136696225	-31.19646447962	10.95843233493	-2.708849562498
-26.37907204453	4.610800231025	22.31649492304	32.13628269023	-0.6063800096975
-19.27119051176	15.28940756307	-20.28175838225	24.73688333292	6.368336100877
15.1061600006	-57.22858308407	-18.25973539617	9.914371504423	-19.74430314754
0.2351670205829	-30.23912862612	-3.257706681815	3.938900884315	-8.505703917908
2.146125201729	14.62849248907	-6.121183018154	-23.64469878311	11.36467654056
60.99075260356	48.52405129018	-8.963416697473	-2.85879846663	-2.172918788574
-55.56669163555	-43.8626391087	-28.58243902134	2.16667360599	2.883745281281
-44.51261088611	20.5863407622	6.034816700272	-11.04985311027	6.979080290267

Table B.1: (Continued)

-1.068864823163	1.933130662398	-6.257358464912	-1.518275583091	-0.2105240282935
-5.554364081259	-3.142052331601	-6.926316746503	-5.241673033729	-2.079594294187
7.2593903715124E-2	-8.850212950402	8.124332875541	-3.469101652444	0.7875687072847
-0.9970644113375	-0.2180584229712	-0.6720305989403	11.07604909707	-0.2960014271954
3.027504605815	-0.4449297687474	-7.041435924067	-1.333639340102	-3.400239212024
-6.042109855792	-11.90191329059	-11.32584705446	-0.5364960677456	-2.318334293948
-8.995851251835	-5.296934297855	-4.378372692342	-10.98587657957	9.503904056977
1.917675382563	-5.68592758674	-3.486325132445	-0.3634162078474	-3.670564261465
-7.923312210332	-3.01722176238	4.589945916063	-3.646518551613	2.078764366669
7.185367672488	7.978495984907	11.06792932271	2.104823288695	-8.390157385769
-3.114144363254	0.2526877804936	-4.272800372747	19.29386476228	4.423177225398
-1.615979702144	-5.255592883136	9.053370648096	-1.332808687514	-3.496335670824
8.354408661197	10.82494025616	-8.71694933342	2.378518962801	0.6261825400556
-7.571772208506	-8.662767775823	-5.345287599218	-6.660987324662	21.18814701537
-4.673324101721	-11.72832999262	12.5473526568	3.4827097763821E-2	2.109866886889
7.072421547777	11.44398316713	-6.102070001575	10.6398000563	-4.679939014319
-4.195787699478	-3.489454738151	-3.767482687291	-4.505889029125	-3.917883363711
-16.06276581622	-2.890911343991	-0.2362376489941	-1.333211990684	1.009443826515
4.311430691314	3.366308288812	-0.4973326672865	-2.55203327162	8.442404127965
3.523681110398	0.2305079704228	-0.1364885568152	-2.084588525975	-1.724058597032
-9.9099278316352E-2	0.3717176943323	2.317904900316	-2.21158651517	1.150832126579
4.500512121879	1.230988137894	0.4254751400587	1.23281099068	-0.7971144643622
1.944932673062	0.8124418663677	0.6915463905965	-0.3115799288562	-4.117246825248
-0.2813533405301	1.707094165989	0.9634071403158	-1.122944331189	-4.595534291874
-2.376375578857	0.6436895765688	3.567521938792	1.911251546144	0.5621090832481
0.5896916179642	3.474200810911	-3.805244311919	2.439986216241	-1.077353458259
0.6696685993441	-2.21165650264	0.1999472046479	-3.960872579688	-2.056862654114
0.3613823620876	1.4028537317	0.7523258882059	-1.205353539537	3.829596692848

Table B.1: (Continued)

1.202363858901	-5.305641880914	-1.619029939409	-2.271901118724	1.704455434064
-2.675170606818	1.811924284744	-6.141402917237	-5.805331341006	0.7560939727414
-1.321707971838	-1.866937621616	4.514272194806	-5.545439313822	-0.6097976393898
-0.6543164194464	-1.988289326045	-1.940863049312	4.166905580101	-0.55977901527
-1.38785661459	5.656328401567	1.016446193125	2.793905977639	-0.5991708595516
0.2443204928352	0.2393285751162	5.30612930496	-1.73035743001	-0.7323376226184
-0.3894115146265	-1.048912427907	-4.056713642007	9.49964281018	-0.6765957618816
-0.4621447751712	-0.1740481429669	-3.882640294367	5.1041858852541E-2	0.1129325328151
-0.1205449747523	-0.3268204490444	3.899690423079	3.108918216302	3.529030324084
-1.913966698329	-0.26725224323	-2.690962196317	-2.614547812348	-1.955934665421
-3.383656798427	-0.2899315963585	-1.670156048826	-8.338690926641	3.977041481644
6.55672719917	-0.4353158199112	5.398648538031	2.155830467262	2.411797755909
-6.060036746544	0.602978904915	-0.9446949376018	-2.917492900239	-5.798633781265
0.1442324752943	-0.1626358297437	-0.2303191215278	-6.1118328102974E-2	-2.887043716156
-6.901556226495	1.393935719191	1.524331269832	3.463720591237	1.460572481566
7.500933733275	-0.9531558225218	-0.518152140205	-1.416388356788	2.742201051423
-2.166447926211	0.3753854982186	0.5869152186847	-1.903158962079	-2.346994216426
-0.4515618315188	-4.237292169733	-0.2860277342135	4.723250639156	7.977584045979
17.59293238746	4.34205414339	-0.8521851438742	-3.121809056966	-1.539297367944
3.664712196395	-5.83129435747	1.52916526863	-1.218851612945	5.364380305379
-6.758337283362	-3.477943791355	0.424439392084	2.359141318601	-3.104623991554
-1.880903761982	-6.562695500207	0.7444450801591	-2.505093375897	-8.260109275965
-2.118291168728	7.143912019452	-1.286233465021	-0.3490659083212	0.3071293230223
6.58720287174	0.4821099866992	0.1383199430904	3.114780492254	-1.919646069322
-0.7860554946679	-1.193278859069	-0.2712893683719	-1.571809195376	-0.9186604413462
1.726445883144	12.63787927555	3.013315053558	1.439037052453	9.3324820624605E-2
3.261746224615	2.490934408266	-3.551582959312	-0.2703967734399	1.081433404116
-5.095955325496	-2.732779119064	3.528706805842	-1.347278323049	-0.9254610190378

Table B.1: (Continued)

3.44140766467	0.2262462189654	-0.6860267451226	-0.9788567783542	3.519989232798
-2.528505025383	-2.960211319357	-0.5224070249487	0.9486762967926	-1.675587912421
-0.2190074812166	6.093973819036	-0.2316931873923	1.642042658392	2.984753896789
-1.571589391863	-1.463291101411	0.7329373650798	-2.620229090265	2.252380562116
-3.682534990045	2.609510303067	-5.2130358611168E-2	-2.337584374106	-3.57370890655
-0.4593390411648	-2.046430749859	0.6817535510178	1.61581063993	-2.974943818664
3.955994685959	-6.915927377585	0.7209235507149	2.120127907321	-0.4550584035376
-2.247624161419	-0.4025284424682	0.2311499804797	-0.2665205029062	-5.4252217523958E-2
2.500828267274	0.3966557692715	-2.393900140732	2.844167067483	-3.815655861989
-0.8233006873249	-0.6894955306435	-7.5687883933117E-2	-2.429353596036	1.330453032563
-0.8853594194481	-1.418479287677	4.791477282184	3.068308446724	-3.52879699562
1.508335828113	1.573473694967	5.049020616353	1.058236434455	-0.5446166422107
6.2423919506843E-2	-0.6881914178092	-2.589703623852	-1.101356936499	2.084449543327
-1.139701847018	1.001741781936	2.881786732342	-4.4572522450261E-2	2.162087965665
1.112259567133	-0.8802405544669	2.061584737626	3.745723749772	0.8346794048235
-1.177003545365	1.083581769727	-2.646633828963	0.6193665846673	-2.02098790386
-0.7621747785326	-3.552981181958	-3.208867904786	-0.6507925695032	1.475914296691
-0.6773048084526	-3.625686260247	2.962530665171	0.3770588594893	-1.422729742202
0.496544215577	0.320966312929	-0.5434866826426	0.7217795579322	1.926250084954
-7.9112038832299E-2	3.552940976384	-1.873141331527	-0.5565745462415	0.8623437551221
0.6491242136087	-1.609121608147	-0.122015883991	0.3095769903288	-0.3366852229267
0.9710391727506	3.156964745282	-5.238071424604	0.9915698309511	-0.6302109373534
-0.620764952534	-1.766343405452	-0.3649084828543	-0.1151506153772	2.045009300203
4.823748250897	0.9453397042868	2.067026439216	-0.2685797635051	0.6519270301399
5.188106809573	1.273249177703	0.851688297205	7.396667785545E-2	0.6326890762686
-4.250329054035	-0.9317521731441	-0.4634265335148	-3.597908145635	2.002408401789
0.5185897126723	-0.6105988579876	-0.1160641513175	-0.2012637676735	-2.350599873832
1.041147907525	2.880872631814	5.422531522026E-2	3.291038770459	3.891376452278

Table B.1: (Continued)

1.027009107243	1.742428067982	-0.3700539418016	1.74397637417	-0.5700037196199
-0.5754290967225	-1.915535721575	2.4536680568752E-2	1.187134851223	-0.4352585735806
-0.2282667744875	-1.297557768446	-0.5736134438035	1.165535015065	-0.1748983416416
3.224913125446	1.041750276939	-0.873876749946	-1.290167205262	-0.1917807899167
1.651955144492	1.388904161454	1.365268115386	1.579117784313	0.6745546529061
-1.248857529952	0.7244367426866	0.9500770673331	0.7172148745825	-0.5934018944558
1.084498186532	-1.986844725987	-0.2602110603121	-0.9628565320749	-0.6989311681219
0.7132031616458	2.35305645191	-0.5817897989517	0.1997422940978	0.6719354040725
-1.46337636962	-0.3324779554295	-0.2610691059998	0.6953006340969	0.2988837257802
-0.1028786675341	1.814746052269	-1.8765714021073E-2	1.091504862763	-1.415678615998
0.6672400239017	-0.5386674789528	-0.8771319601675	-0.7266073188231	-1.659943260126
-0.4400537785766	1.251405102172	-0.4317508154278	1.286292907212	1.668952391893
-0.6565029982388	0.5583132405261	1.956758258378	1.542681114061	-0.3978998651731
0.7263789561384	0.8604201813847	-0.6430281340839	-0.5961892619027	-1.800053031174
0.5146669759641	4.5170537518826E-2	-0.98150012384	-1.149947448171	-0.8321952178798
-0.1347471073098	0.7998548186149	-0.9938159852747	0.4603839369015	2.548171607897
-2.130989447238	1.427152938876	0.380303641798	1.022730238945	0.8057239309174
-0.6203030337718	-1.620423804308	-1.529674805989	-2.040325212895	-0.7032538997723
1.099487545593	2.953690638368	-2.990085203833	2.196172174904	-1.394989920972
2.467937739886	1.287138833597	-0.9854714291	0.854154016451	-1.1384022388988E-2
-1.272416334543	-0.3312412924107	1.281517699831	-1.479502754218	-2.581989172715
1.119189234951	-0.8645876288175	1.415732685141	-0.4888175430258	-0.2996328564812
0.7973537661298	1.725366935328	-0.8110599121826	-0.4103261743547	1.660222865456
-2.063996343379	1.711056588913	-1.899288811519	-0.8615022579647	0.2379312157379
-2.396917022354	-1.821526921248	0.1077324930419	-0.8742560500384	1.554457260184
-2.673127111968	1.838558821959	-1.032621564603	1.32883411162	0.4865021019784
-0.4222800223138	0.8158998927333	-0.2121324579897	1.074833932744	-1.000778294961
-1.959718505004	-1.909653883781	-5.9922721941359E-2	-8.2217691049673E-2	1.305772635307

Table B.1: (Continued)

2.9700514769999E-2	-0.4610698891032	0.4399183Q88369	-0.8209274913139	-0.8656441032049
-1.665285366318	-0.1287010191026	-0.6185332341044	0.9110911579876	5.2597158216759E-2
0.230881464189	0.2083721497557	4.4035965917388E-2	0.4815232161161	0.1828365127185
0.5989738988974	-0.1384617321853	0.1174479120218	-3.604866137459	-3.8045258162404E-2
0.3825992226429	-1.775789801902	0.7864800407272	0.2554687323557	0.4961307759602
-0.1370683572565	0.7580326573392	1.4319997036419E-2	1.814389240903	0.5642041961265
0.2261654891631	0.1322768312304	4.0992028477145E-2	-0.9540858339638	3.803797491697
1.386371405851	-1.308020080009	2.544733118163	-1.079813787407	-0.3174796661599
-0.6140462068965	0.8326520994078	-0.25054850443	1.331839870891	-0.3062724394622
-0.7336145692802	0.7048867450072	-1.389577387979	0.1642093617091	-0.9355756752673
0.7186743373684	-0.3003765328814	0.1938877751677	1.146816404864	0.8449796877271
0.2281290260535	0.1822713957804	0.7689614346479	-0.9961087169943	-0.3702052870634
-1.929070644092	0.9268505478458	-0.4892297446655	0.2098473478572	0.6438769088581
1.601924380338	0.2646066451446	0.1967270335067	-2.470964814463	0.7743502761
0.7679348477818	-1.538872661422	0.1349519525322	1.305196576824	-0.6640669384287
-0.7756756797319	0.1284231410177	-1.508774097677	1.155851646641	-0.2055268165129
-0.3356594472681	0.6078233888125	-0.4313242343041	-0.8831528358591	0.2744472485223
-0.3499317283745	-1.897890149522	0.2287519614363	1.736356579184	-1.044967443223
-1.375147396479	0.839816850278	-0.4296876489671	0.8609315951287	0.949609037027
-0.736938503843	1.662734706996	0.9709976042566	-0.7297620842881	0.9687932506981
0.7144271283716	-0.7827433417464	0.4661914549333	-1.419632523588	0.1150441834891
0.5693083108411	0.1575179816109	-0.277177170175	-0.9566720407359	-1.569858541987
0.4377720370021	-0.2125963641649	-1.572123537728	-1.253403940489	1.311933810097
-0.1603092505525	-0.5783176567984	0.9336488615341	0.1886172311977	-0.2403582858381
-0.4391799570221	0.8909900199661	-0.1473831192223	-0.6011677404322	-0.6328984586581
-0.8635621448796	-0.6807015928275	-0.712643391136	0.6959400115355	0.2988174289884
0.1730975346741	-0.6793849830679	0.1207971459262	-1.100977517894	0.1739886777721
6.7188457908155E-2	-0.3892150573116	0.1968730259099	0.3637702571329	0.4075217704475

Table B.1: (Continued)

1.4808653672204E-2	0.8638752964032	0.7586810023055	0.2052698672482	0.1949574031914
-4.0373339192445E-2	0.3641378222242	-0.5318192160343	-0.9058504680593	-1.31148852813
0.5884358817123	0.1254861065866	0.1806531099726	0.9766707430294	-0.6155163348077
0.5344232690734	-0.4627809435356	6.4410947431694E-2	2.166355509438	0.9416875201986
-2.329525190466	-1.023631820226	1.802738419551	2.317122943345	-1.385528646968
-0.3361007816978	-0.3205438146067	-1.342711883712	0.2678053647559	0.1502144595502
1.094697737138	0.4197143888017	-0.6034780445563	-1.694574606113	0.2025339469461
-0.6563178272999	1.064518891806	2.501547342424	-2.200189772269	-0.3178244450505
-1.59058820729	-0.1137330326198	0.3719430396455	-2.511992947127	0.1488864482239
1.2956803892	0.6402366672796	-2.951792449852	-0.392875615513	
-0.9316493915486	-0.525688303783	0.9749591690956	-7.595661213854	
0.7087028565661	0.2843338702554	1.262401287962	-9.7756509359875E-2	
0.5026637803838	0.3785636795137	0.7498572305273	-3.1146657655801E-2	
0.850944174175	-4.9835151309056E-2	-2.762412653951	1.77280703187	
-2.372197497069	1.139536789881	-1.904600614842	-1.986335464572	
0.7390317939457	0.9451255508426	6.411465546528	1.850535083805	
0.4246774080382	-1.088101070022	1.569695276603	-2.562050132732	
-2.678406046181	0.5916585951132	-0.7319901954993	0.6407022920873	
0.7740262094518	-7.5870938261419E-2	-1.414998243938	2.442876366135	
1.280168001577	-0.2931718751634	-2.372797716605	-4.397018639139	
-0.357880354192	0.6741915725771	-7.593394444038	0.8202158806268	
-0.1523565071601	0.9700871317147	-1.922086964605	3.085611644345	
-1.266787564907	-0.5134812508541	3.248418957635	-3.212381945557	
-3.291175540932	0.4070484159321	-0.9289426237046	-1.100559965742	
-0.736285901784	0.1089042915441	5.630017709137	2.394863266401	
0.1119326478133	-0.3089685242455	-0.4747108980525	1.267415922468	
0.7166535500148	3.4349105707934E-2	-5.590876387905	-2.03092572932	
-0.8054827657959	-0.1282936570832	-1.108048827727	0.8855635554506	

Table B.2: Spherical harmonic coefficients for 500 mb heights for January 14, 1979

19453.364811613	88.593128117599	4.5163466589363	-6.2529845527122	22.814165684445
456.1233042283	9.382572647847	0.68600876146014	25.168255909755	-36.991569857738
5.8335380494982	-60.913838876885	-13.78930063938	21.397995953055	-5.2524621804978
32.193428459104	-19.454616930372	4.8856297910376	28.044124363745	0.12299380835201
-1045.2825646188	4.6889300884921	-10.741077885427	-4.6634948404396	1.8355376793371
-13.729484975423	6.7039899693572	22.728673531067	-28.490702344136	2.2067176305271
33.983826549534	12.389409877995	-9.9585782235467	-33.001580019728	8.6020062230272
9.8992318557482	-4.9257954981502	8.2090152394269	12.274403291161	2.4078811601749
12.103709262447	34.309158883431	-19.210019065235	-11.301426117005	1.7010360763467
-72.381897847357	25.971511359123	-26.220583495558	27.35331565771	-5.893681826731
10.49546742731	-37.33123961172	-25.317046065157	-34.060769075584	-11.274240417237
73.229393556783	62.658134034907	4.1989569183374	-1.7948933071539	0.79476015833138
49.901328360695	4.0721352684791	15.934695996666	-5.0360472872846	-2.1832373387104
-12.228064053275	-63.470017026362	3.2896747380704	2.6672189782719	9.5099061628527
-5.6696053427823	-4.5106033654701	33.295741933444	1.8657363101238	1.2096232464537
-21.767384261804	6.9274601929999	-1.2009145191507	4.7132905562984	-4.8659545242276
106.25561797486	7.5626062556096	-13.09516836865	-20.120310240681	1.4171556685655
39.658721121512	5.4416632825012	-30.000576702798	-22.385816981005	-39.221496022889
53.824193854764	2.2964898088884	7.1229559402012	5.6479908720246	-11.966873316298
85.936495237218	-4.466467811817	-11.797622555569	-6.3839423783718	15.062298786542
4.4990416868452	13.998568658463	27.389690250674	22.586497343422	-13.809402158123
-28.985837916575	8.2958750825662	-24.034560623237	16.299362801802	15.168661097888
-19.122636877018	-4.9754810209991	5.0311778470121	10.998059523587	-32.834644211664
2.8044683304876	-44.230405373442	-10.065411714386	-2.9947657873158	-4.0830214482175
1.4229635243511	28.840394896875	7.7874182598882	-38.557822343795	0.56946712824602
68.79145866108	5.1663045631815	-30.768517362757	-24.317872393491	1.0841677495279
46.475938200905	-33.462230087028	-30.252945827089	16.485333011707	3.4924003794798
2.7765065106622	19.418864949592	-3.1936776407739	-12.521772611926	11.577316781265

Table B.2: (Continued)

0.90238166373612	0.76939497094597	-19.296646914067	0.26206294227349	-0.29937172547391
3.0269344194626	-14.229478637372	3.1691560663499	-14.542354387843	1.196074166546
-3.0171171973887	2.7619941796348	-8.2517666092327	-0.17208538424065	0.34751727710245
1.8928656229559	-11.219177080875	5.6220327957465	-3.0004591456802	0.76867623149538
-5.1606856187993	4.8749935084566	-9.8383249954775	5.1175780437662	-4.6561664883021
-2.2350763389617	-11.112223567779	-5.2814828497628	-5.2542523160226	-4.5106574146863
-7.7570266743444	-9.9468274014781	-11.256254426399	-0.90726476506063	0.17269229513602
2.1856031830938	-13.706539635074	-0.27298111398047	-7.4846607718685	4.3854393466392
-4.4057244371122	3.2191627302546	-9.3621298382593	-3.6208136170636	-2.3045898281223
-9.152949644501	-20.069836520664	0.88235483721487	-2.0191769002363	0.74165286340304
-12.749998556206	1.3539242958491	-4.82011463554604E-2	-1.8307039136803	-3.4193122357503
4.1922777790779	2.7832254259901	-7.3927526225411	-0.52748868155169	-5.0133293991916
-31.549480402153	-9.7183907845807	-1.5708781775644	-6.5011262606687	1.3406223655733
-2.6420797648574	0.93399857830154	-2.5193308330552	-0.95845365356882	-4.1803887874546
8.8709694682856	-11.279825833229	-3.3677531236022	0.31376188649979	-1.0464937472196
-12.660946728919	-2.1958221911106	-3.719628953203	-3.2980642317895	-5.7801082423488
6.9824346485766	-3.6936236989993	-7.5012371116858	-2.2722456997602	1.1005877563686
-23.040494369874	-8.819822583304	1.2188600158005	-2.3632190361156	-1.7541611772126
-2.100968595392	4.063746166953	5.6850397708092	-1.9113909858133	-1.7141317576093
-1.7911911565052	10.444622723868	-4.5908978793341	0.35739035055417	-0.58613308034232
-4.1094962216116	-2.9855410453648	0.99918941110355	-5.3947112310847	0.79616842768125
4.774630732711	2.1957363422193	5.4106894545228	0.40051051012821	-3.5476736349221
12.522027553741	2.7235067778672	-1.9686457276498	5.6871272925407	-2.9884264950149
-1.1854377580267	-0.8703529974469	1.1362230012643	-1.9097662316942	-4.6923456092818
3.1832820191486	0.68878474324867	2.2877792332423	1.9052883810962	0.52150546765712
-1.1405770713512	1.5695483152755	-0.36748128432159	2.7929559617409	4.3650090655593
0.92752077207707	-0.51933096782938	0.14439469983798	-0.20700225657173	-1.0021918076311
0.63390460380231	0.27899730009094	-0.23935790637011	0.23555188854858	2.3226360597394

Table B.2: (Continued)

2.926787312115	-3.0865753833741	3.1029976413525	4.1751490580538	1.2131191341693
-0.37570893200187	1.1364516566496	-7.746479454127	-5.5174247645235	3.64647511536387E-3
0.67719274971926	2.5821803422537	-0.14547428956102	3.01043867613177E-2	-0.6489911402058
0.17579247236871	0.2412328666502	-0.25643242945854	-0.17755686376102	0.57690762196956
2.120168409301	1.9800991043025	-2.057998759614	-1.7773520192131	-0.53324145086302
0.58722231498607	2.5113759323904	-2.362610663349	-2.2145819790592	0.13250662828981
0.58497220008526	-0.48569610825569	-1.5530811435142	1.4292955992816	-0.30056035670648
-1.1611111485856	1.0378061123142	-2.0463055078381	-3.5346726814583	0.27938889235793
0.15664333023886	0.98223761832714	2.0898219125524	0.47309212917639	4.7218123134939
3.8660352674088	2.5365962665015	0.90194500417922	1.9965364371916	4.696580316898
-6.480867508973	0.47765361501952	1.2145877511843	-7.9252556653336	-5.1501614790307
-0.1294554079219	-0.23950244206959	0.92445858941675	-0.91965925851352	3.7188316324161
4.0865171566006	0.57604785959261	1.6094087409815	-1.3011372382793	-3.9856110991211
-2.6736943394928	-0.19049529385691	-3.55226512461149E-2	-3.2643874559288	0.92727255875628
0.45266930904953	0.13111868576511	0.91610636748642	-1.3820017776222	-1.276291353073
-0.35463341473199	-4.0271483403177E-2	1.750359869618	0.40934050175458	-1.3559150676769
-4.3917682778173	8.1137071756391	2.5067411404129	-2.2520818271701	-2.6557163914628
1.3725407761309	-4.4967067956072	-0.29280015292822	3.1937648234252	3.2422939976154
-4.2109245402019	-2.5447494882303	-0.99373989395055	-0.35605753729681	-2.504192074694
-2.3840082432611	4.1764444949623	1.7883571153928	1.2921325868684	1.4297554198747
-4.0374562782316	-4.6672216814527	-0.28726508559532	-0.62382846224614	0.274755801602
2.8046091108899	-0.10916314471325	-0.2203323890447	0.67595468803925	-6.3216595881426
-5.370486702439	0.62473311434111	0.19161625337182	0.77178709940931	-1.6409115919667
-0.27114562771437	-2.9111270543633	-0.26549234534344	0.39370113453033	-1.8551914800587
0.19552473777882	-0.42914868878909	1.44759972160006E-3	2.4759981283217	-3.0413845091515
9.94929256229908E-2	-1.7684352273807	7.9698609657766	2.287276162728	-0.87417357789748
-3.406034905807	-3.5386706140493	0.20297162716766	-1.5226490718041	1.4825765188155
-3.2951957448379	-1.5766833830002	-4.7697002871919	-1.1669340636607	-3.088770530799

Table B.2: (Continued)

4.021415036711	-1.1308119585387	-1.2683043429417	3.529705364558	1.7661067080455
-0.85676381843949	-1.6606661619722	1.17217662455822E-2	5.31084354171769E-2	-0.58851231562743
0.20002998822224	2.7940194868905	-0.50166331004756	-3.2972512828173	2.8080680888218
-2.2806437125438	-1.2428200731558	-0.52931974854614	-3.3723515028519	0.28184387270971
0.27918626479573	1.4285702971221	-9.50947738836447E-2	1.3812767669774	0.80103458921063
1.2094200796229	-1.0828329870091	-0.34097569807277	0.34581332636708	0.82530121416524
-9.76139894587784E-2	-3.9677205364823	6.02975902942393E-2	0.13826928531235	-0.91643094235734
3.2754120516018	-2.0266757280737	7.05065379507355E-2	3.4240236065583	-0.14215617753032
1.9164110498721	-1.1700663059403	-1.1942266834094	0.56278718191354	0.70076594291109
-2.5384855722647	-1.9273601495281	4.7065214852951	-2.4836918821333	-0.98863162814643
-0.88979233522031	-0.69192346436856	-0.59683179573792	-0.54144061172643	-1.0728891298538
0.21849451418788	1.7477148892011	2.1005670186751	0.60260827956834	-2.5445362265651
0.30163063304513	-3.6347591311017	0.2637996255127	-0.15392735554462	2.4796673098976
-0.91971684595291	4.1444184498785	2.8330817860387	-0.71133959974675	-0.22291186168524
1.0751533174899	-0.56920967121791	-0.65376341249948	2.0087325020487	1.74668127015269E-2
-0.81303266292196	-1.6479791211019	-0.43210637567199	-1.8402013212164	1.9324092528412
0.13317145778326	-3.3957555271005	-0.17008169288417	0.10466065071042	-2.8464166889695
-0.56674825573857	0.62124424306861	0.89065223086726	-0.10855833046626	2.4692313618848
-2.83184163337912E-2	0.92377314601743	-0.38557383745024	-0.82581642220589	0.46431127353578
-9.79713264122645E-3	-0.19131500561958	1.0130778901789	-8.02138028541588E-2	-3.8525151407669
-9.21295672559985E-2	3.8175462640456	-1.4874488881786	-0.66299168936668	-2.2020058312713
0.84556821432908	1.3086493139602	-1.9944232067815	0.10030191632429	2.0298564770317
6.4366196815825	-2.8402837494978	-2.3260793043506	-0.11034649815368	0.18423886940771
-3.4138533043683	-0.61040383876681	0.61601657670356	-0.1001530996217	0.58478583710614
2.8854281478773	0.12191400128503	-0.71402523763972	-3.89422274716502E-2	1.8396829610544
-1.2831858421449	0.25818077312471	-0.46571159009931	-0.78233982309635	-1.39718337125033E-2
2.0754108218041	-0.97490639253825	1.8035201720626	0.86063364517324	-1.8046401393528
-1.4439810038086	1.6342270176888	-3.5260607367172	1.6793822672474	-0.55586202845797

Table B.2: (Continued)

0.82212601714894	-0.11561584660267	-0.90996458419943	-0.62169664979925	0.17124347858795
-0.63422318437501	-1.0238271880181	0.31136537172494	-0.46653662590549	0.19309569888426
-5.16395278363433E-2	-2.3053140957308	-0.4480945931113	0.32560398537431	-2.39554432216319E-2
1.8801782908886	3.2413992064548	8.87894547917307E-2	-1.5731521339833	9.99147089789776E-3
-2.0769057684114	-0.58080805653267	-0.27105217672461	0.8215781854676	2.6297663356054
0.59690945372801	0.62765924182136	-1.7813723686126E-2	1.9625611692846	-2.334577646043
0.30652594693007	1.9161659962303	6.41668085156035E-2	1.5158644188243	-0.62328945723787
-0.85007290827111	-1.7741011852155	-0.28661172222981	0.48918681030561	1.4697311094227
5.38560643595609E-2	1.1438341449864	0.31048375647595	-1.8183011706913	-5.7215448132364
-0.89623092080114	0.48901882507595	2.1598792793516	0.14635676032724	-0.13887762906341
0.20898181780485	-3.0839984211785	-3.7382969998447	-0.22218369319868	-0.60725196352655
-0.35490231578435	-0.5365346399872	1.1210689465503	-0.42581158797783	1.7155505618878
-8.1839386554309E-2	2.2347159034105	1.9823061811944	0.60773977692893	1.2602936264173
-0.14276217039258	0.70124002565828	-5.1957816414327	-0.20048968978777	0.16347463721162
3.98149059796848E-2	0.7555559673886	0.91479334043912	1.2061588931708	-1.0098338003249
0.20178599918086	-0.28502309370497	0.20836762394335	0.4308132524336	1.0295703966757
0.8208014993949	-0.15963829167579	2.2972441149397	-0.69775175037151	0.60490477609074
-2.6149390484792	-1.0246289709149	1.0966797707146	1.4906958719355	-0.84144886790584
2.2468636279098	-0.50649779833593	-0.93511077317426	0.51077096958389	-7.53785255971167E-2
1.8857268519633	0.68750847693785	-0.85218823761376	-0.68381945321239	1.2596659630138
-3.0387781352917	-0.71899043209886	1.0119593888557	0.37515729440213	-0.1246804205934
2.0448709896354	0.76201902414243	0.51841250536821	-0.71886167945382	0.94724403229838
0.62319517737808	1.224019281041	-1.1168448174011	0.27441612758312	0.60609677195993
1.9442337982962	-1.6188574908309	-1.3851416537122	-0.33324979003786	0.12989596245324
1.0948447932835	1.2008021851243	2.5743442612093	0.23851392789511	-2.1406724534979
-1.5439018286154	0.50772510315413	-0.8517646842665	-0.40172106913691	0.24062305541045
-0.41028992220627	-0.77326635148932	1.0213699327722	-9.05063042445784E-2	-0.26349863740992
0.74743255539743	0.21712472003145	1.4583897750127	-0.22456829375802	1.4449104488773

Table B.2: (Continued)

1.3232711413087	0.13499299180625	-1.1463601645078	1.1860636474885	1.1344170490523
1.948941117427	2.2069000009021	1.1374145692874	1.4483211587912	1.427961613814
-2.0827962957643E-2	2.86344469654665E-2	1.0914087022727	-1.720246300177	0.15141671891496
-2.0435666194074	-1.7416152202394	-0.21956494329985	-1.013454956976	1.3194537666761
0.69179324975777	0.25462434270279	1.0614287902915	-2.8015084642682	0.63295782310193
0.51781026652311	-4.6121386660392	0.90002932448325	-0.82768100177391	-0.78191221159813
-0.36629955557661	-0.77730545022473	5.48583920885242E-2	-1.3312539698533	0.60973001393008
0.77169833774139	-1.2598802758265	0.10559824638933	-0.17183552114731	3.30092176817955E-2
0.55007751965622	0.67793979627201	0.27870976681209	1.2067784172979	0.63631254922788
0.91828232931273	1.4587422619132	1.0850427798261	0.67326202672605	0.39819239635219
-8.97627507898313E-3	0.84507227215598	0.39892661700416	7.40353480246574E-2	0.60617637688647
6.51736427129666E-2	-0.62681750066501	-6.28160943364l8E-2	-0.76423801065179	0.16737805750516
1.3856887179452	0.40331261578813	0.49811254320969	8.33609184484518E-2	-0.36768503793856
0.44573856274796	0.3078094131314	-0.39190779730481	3.73524254822983E-2	0.32334772595473
-0.48967367060494	-0.31180262957355	9.79006480788365E-2	1.4694021612532	0.47353337415276
0.51245850185979	1.0315752003224	0.16914151732601	2.36194816858012E-2	-0.39746508291125
-0.49315810257747	0.31584369978864	-0.16914034572651	2.0646258064947	-0.1118847846388
0.13341560388062	1.2443019850327	-0.39235676260371	-0.18341746257066	-0.17808261969169
-9.9813293202323E-2	0.44568586420774	-0.16126028903532	-0.84688751198063	-0.53305414590646
0.14480560269477	-0.27485119917418	-0.52628502222265	-6.00036079927218E-2	0.1979893420497
-0.47865207297574	0.21624049125015	0.3263589212594	-2.7751235816209	-0.47128114905813
-0.13542146255267	-3.1528477248095	-0.36817179394075	0.65159671048469	3.42447118027467E-2
-0.44491775372376	0.36183014996947	-7.83063264164632E-2	9.39506239542585E-2	1.66619361685173E-2
0.48152909486738	0.24808645092665	1.05272477807247E-2	0.83553699667606	0.18464644667953
-0.13451918922719	1.3663895464123	1.11975051426901E-2	-0.32676579478199	1.90264660270782E-3
-7.79481062677503E-2	0.47267653376577	0.14019305024376	0.23681084172411	-2.67978858632094E-2
1.21275208174145E-2	1.5096810480932	2.46608071314612E-2	-0.60093563581152	0.11946579243198
-7.37744773741968E-2	-0.43576677685554	9.06227923427734E-2	7.34163875979488E-2	-2.93177623587814E-2

Table B.2: (Continued)

5.28924453677433E-2	0.87361497261454	-8.63489484261213E-2	-0.48015830056261	-0.33277442688854
0.34343081885414	0.48027858614691	5.61478338436594E-2	-2.1085970485815	-8.63989326724623E-2
1.1819068838512	1.4581643993845	-5.82179145252916E-4	-0.71053531490225	8.8567497402114E-2
-0.94685478527714	0.6469151373584	2.85378856387478E-2	1.0106713816142	-0.12964407193844
-1.4252735002798	1.4355136395622	0.55910750432163	-0.7115595422833	3.87058220043042E-2
-1.3271612652449	0.11564740324438	-4.39871533779146E-2	1.1116061178016	-5.92583070315904E-2
-0.30884505936328	-0.94355496594147	-0.14172606290878	0.93920886956555	1.77069978073872E-2
-0.83648231921812	1.172933122381	-0.66785980404188	1.327577719665	-7.45145732054824E-2
-0.51074476796095	-0.37727763266574	-0.73453902122635	-0.18804359713694	4.34938739472632E-2
0.30873905949403	-0.14393082421643	0.5743397033908	-0.17514507826822	
-6.4936328230516E-2	0.43441798475574	-0.23503574587122	1.5005887132284	
0.5306978167741	1.3637459125391	-0.44525455570034	-0.74347665873979	
-1.8432159847161	-0.43602844081111	-0.78928106540557	-1.2644041549334	
0.24965935588597	-0.20587601540387	-0.65088837529821	0.45101056703143	
-3.46665511302882E-2	0.74724684222335	0.32144783861103	1.9500554236162	
1.1202679858767	0.84445026399954	-2.1312731967332	-1.0044657269641	
-1.55567760673865E-2	-0.23157322117088	0.67732260035757	0.19507722966176	
1.7672052815726	0.20445699131489	-0.40978778014085	1.1370543550119	
-0.56092249492724	-0.17080261560311	0.20731719206088	1.2533135419574	
-1.0312855361835	-0.5271707181359	-9.31950219795352E-2	0.33942791060519	
-0.10918250703538	0.48862867716908	0.8621868019789	0.32900635167229	
-1.0841746803079	-0.66779689637881	-0.44311036717536	-0.10019051261185	
0.93367714073994	0.28348822606896	-0.9241319753974	-0.51293598238269	
-0.12340545285329	4.32483957430483E-2	0.29856219032098	0.94327407124405	
9.84041925673287E-2	0.24884318678207	0.90432822337305	-1.0323659708022	
-0.7149832970517	-0.20533297758759	0.96999188230346	0.66262628676507	
-1.2534740194077	-7.31135453258713E-2	5.42334447541701E-2	8.10151166722508E-2	
-0.63382374010774	0.11696544295236	-0.65149427950788	8.98987986783606E-2	

Appendix C

Locations of Radiosonde Stations Used in the Experiments

In our simulation studies, we chose 600 of the radiosonde stations over the northern hemisphere as the locations of our "direct observations". We put pairs of (latitude, longitude) in degrees of those 600 stations in Table C.1.

Table C.1: (latitude, longitude) in degrees for the radiosonde stations used in the experiments

(28.22, -177.35)	(37.75, -122.22)	(53.97, -101.10)	(29.33, -89.400)	(46.38, -75.970)	(61.18, -45.430)
(51.88, -176.65)	(34.75, -120.57)	(29.37, -100.92)	(38.65, -88.970)	(35.27, -75.550)	(52.70, -35.500)
(57.15, -170.22)	(76.23, -119.33)	(46.77, -100.75)	(15.72, -88.600)	(37.85, -75.480)	(4.900, -29.000)
(16.73, -169.52)	(50.23, -119.28)	(41.13, -100.68)	(44.48, -88.130)	(19.90, -75.150)	(38.73, -27.080)
(64.50, -165.43)	(28.88, -118.30)	(25.87, -100.20)	(14.03, -87.250)	(4.700, -74.150)	(70.42, -21.970)
(55.20, -162.72)	(40.90, -117.80)	(37.77, -99.970)	(32.90, -87.250)	(42.75, -73.800)	(56.80, -17.800)
(60.78, -161.80)	(47.63, -117.53)	(19.43, -99.070)	(36.25, -86.570)	(40.65, -73.780)	(3.100, -17.100)
(3.900, -159.38)	(32.73, -117.17)	(44.38, -98.220)	(80.00, -85.930)	(53.20, -70.900)	(47.20, -16.900)
(21.98, -159.35)	(43.57, -116.22)	(32.22, -98.180)	(29.73, -85.030)	(4.470, -70.730)	(32.68, -16.770)
(71.30, -156.78)	(36.62, -116.02)	(35.40, -97.600)	(46.47, -84.370)	(43.65, -70.320)	(6.250, -10.350)
(58.68, -156.65)	(39.28, -114.85)	(25.90, -97.430)	(39.87, -84.120)	(41.67, -69.970)	(51.93, -10.250)
(62.97, -155.62)	(55.30, -114.28)	(28.85, -96.920)	(17.40, -83.930)	(12.18, -68.970)	(38.76, -9.1300)
(19.72, -155.07)	(53.55, -114.10)	(19.17, -96.120)	(42.97, -83.730)	(76.51, -68.830)	(70.93, -8.6600)
(57.75, -152.48)	(60.02, -111.97)	(41.37, -96.020)	(33.95, -83.320)	(63.75, -68.550)	(31.61, -8.0300)
(61.17, -150.02)	(40.77, -111.97)	(64.30, -96.000)	(38.37, -82.550)	(58.10, -68.420)	(33.56, -7.6600)
(64.82, -147.87)	(47.48, -111.35)	(39.07, -95.620)	(27.70, -82.400)	(46.87, -68.020)	(62.02, -6.7700)
(70.13, -143.63)	(18.72, -110.95)	(74.72, -94.980)	(31.25, -82.400)	(36.00, -67.500)	(58.22, -6.3200)
(59.52, -139.67)	(32.12, -110.93)	(32.35, -94.650)	(12.58, -81.700)	(50.22, -66.270)	(54.48, -6.1000)
(60.72, -135.07)	(27.95, -110.80)	(45.55, -94.080)	(24.58, -81.680)	(18.43, -66.000)	(36.15, -5.3300)
(68.30, -133.48)	(42.82, -108.73)	(58.75, -94.070)	(19.28, -81.350)	(43.72, -65.250)	(50.22, -5.3200)
(55.03, -131.57)	(39.12, -108.53)	(36.88, -93.900)	(68.78, -81.250)	(32.37, -64.680)	(48.45, -4.4100)
(50.68, -127.37)	(35.05, -106.62)	(48.57, -93.380)	(40.53, -80.230)	(82.50, -62.330)	(4.100, -4.0000)
(65.28, -126.80)	(48.22, -106.62)	(30.12, -93.220)	(26.68, -80.120)	(16.27, -61.530)	(53.55, -2.9100)
(72.00, -125.27)	(23.18, -106.42)	(34.83, -92.270)	(32.90, -80.030)	(53.32, -60.420)	(56.43, -2.8700)
(47.95, -124.55)	(69.10, -105.12)	(14.53, -90.570)	(36.08, -79.950)	(43.93, -60.020)	(12.35, -1.5200)
(44.92, -123.02)	(39.77, -104.88)	(32.32, -90.080)	(42.93, -78.730)	(13.07, -59.500)	(60.13, -1.1800)
(42.37, -122.87)	(44.05, -103.07)	(53.83, -89.870)	(58.45, -78.120)	(48.53, -58.550)	(44.83, -0.7000)
(53.88, -122.67)	(31.93, -102.20)	(40.67, -89.680)	(38.98, -77.470)	(47.62, -52.750)	(51.08, -0.2200)
(58.83, -122.60)	(35.23, -101.70)	(20.95, -89.650)	(17.93, -76.780)	(68.70, -52.750)	(52.68, 1.6800)

Table C.1: (Continued)

(48.77, 2.0200)	(7.350, 13.570)	(52.12, 23.680)	(32.00, 34.810)	(28.33, 46.120)	(55.47, 65.400)
(39.60, 2.7000)	(52.22, 14.120)	(37.90, 23.730)	(32.00, 34.820)	(24.70, 46.730)	(44.77, 65.530)
(36.71, 3.2500)	(78.07, 14.220)	(54.88, 23.880)	(54.12, 35.330)	(43.02, 47.430)	(31.50, 65.850)
(50.80, 4.3500)	(67.25, 14.400)	(49.82, 23.950)	(51.65, 36.180)	(29.22, 47.980)	(66.53, 66.530)
(43.87, 4.4000)	(50.00, 14.450)	(35.33, 25.180)	(32.36, 36.260)	(46.27, 48.030)	(24.90, 67.130)
(45.73, 5.0800)	(63.18, 14.500)	(41.63, 25.400)	(49.93, 36.280)	(38.73, 48.830)	(47.80, 67.720)
(50.03, 5.4000)	(12.13, 15.030)	(62.40, 25.670)	(41.28, 36.330)	(55.78, 49.180)	(58.15, 68.180)
(22.78, 5.5100)	(45.81, 16.030)	(48.27, 25.970)	(33.41, 36.510)	(58.65, 49.620)	(76.95, 68.580)
(58.86, 5.6600)	(1.610, 16.050)	(50.17, 27.050)	(36.18, 37.220)	(40.65, 49.980)	(38.58, 68.780)
(33.50, 6.7800)	(48.25, 16.370)	(38.43, 27.160)	(61.50, 38.930)	(53.25, 50.450)	(60.97, 69.070)
(46.82, 6.9500)	(52.42, 16.830)	(53.87, 27.530)	(44.10, 39.070)	(51.25, 51.400)	(73.33, 70.030)
(51.40, 6.9700)	(51.13, 16.980)	(57.83, 28.350)	(21.50, 39.200)	(47.02, 51.850)	(40.55, 70.950)
(41.92, 8.8000)	(62.53, 17.450)	(44.21, 28.630)	(54.62, 39.720)	(40.03, 52.980)	(42.85, 71.380)
(48.83, 9.2000)	(40.65, 17.950)	(47.02, 28.870)	(47.25, 39.820)	(67.65, 53.020)	(37.50, 71.500)
(45.43, 9.2800)	(59.35, 17.950)	(59.97, 30.300)	(59.28, 39.870)	(37.47, 53.970)	(23.07, 72.630)
(0.470, 9.4200)	(57.65, 18.350)	(50.40, 30.450)	(37.88, 40.200)	(51.75, 55.100)	(40.92, 72.950)
(54.53, 9.5500)	(9.150, 18.380)	(56.38, 30.600)	(64.58, 40.500)	(65.12, 57.100)	(8.300, 73.000)
(63.70, 9.6000)	(74.52, 19.020)	(61.72, 30.720)	(50.42, 41.050)	(50.28, 57.150)	(26.30, 73.020)
(52.47, 9.7000)	(47.43, 19.180)	(52.45, 31.000)	(42.87, 41.130)	(80.62, 58.050)	(49.80, 73.130)
(36.83, 10.230)	(46.25, 20.100)	(29.86, 31.330)	(52.73, 41.470)	(37.97, 58.330)	(54.93, 73.400)
(50.55, 10.370)	(32.08, 20.260)	(54.75, 32.070)	(41.65, 41.630)	(41.83, 59.980)	(15.48, 73.820)
(60.20, 11.100)	(49.03, 20.320)	(67.13, 32.430)	(18.30, 42.800)	(60.68, 60.430)	(42.83, 74.580)
(48.25, 11.580)	(54.70, 20.620)	(23.96, 32.780)	(44.22, 43.100)	(56.80, 60.630)	(12.95, 74.830)
(41.80, 12.230)	(52.40, 20.970)	(39.95, 32.880)	(56.22, 43.820)	(46.78, 61.670)	(39.47, 75.980)
(57.67, 12.300)	(56.55, 21.020)	(68.97, 33.050)	(67.88, 44.130)	(69.77, 61.680)	(30.33, 76.470)
(55.77, 12.530)	(65.55, 22.130)	(45.02, 33.980)	(33.23, 44.230)	(34.22, 62.220)	(43.23, 76.930)
(9.330, 13.380)	(42.81, 23.380)	(57.90, 34.050)	(40.13, 44.470)	(35.95, 62.900)	(8.480, 76.950)
(54.10, 13.380)	(60.82, 23.500)	(61.82, 34.270)	(41.68, 44.950)	(53.22, 63.620)	(52.28, 76.950)
(52.46, 13.400)	(46.78, 23.560)	(64.98, 34.780)	(51.57, 46.030)	(41.73, 64.620)	(79.50, 76.980)

Table C.1: (Continued)

(28.58, 77.200)	(0.000, 90.000)	(12.43, 98.600)	(22.50, 103.95)	(27.45, 109.68)	(25.83, 114.83)
(23.28, 77.350)	(61.60, 90.000)	(3.570, 98.680)	(26.87, 104.28)	(44.90, 110.12)	(40.78, 114.88)
(12.97, 77.580)	(2.000, 90.100)	(18.78, 98.980)	(52.27, 104.35)	(25.33, 110.30)	(32.93, 115.83)
(60.43, 77.870)	(38.37, 90.150)	(54.88, 99.030)	(43.58, 104.42)	(1.480, 110.33)	(28.67, 115.97)
(55.37, 78.400)	(45.37, 90.530)	(28.65, 99.170)	(40.75, 104.50)	(20.03, 110.35)	(5.950, 116.05)
(17.45, 78.470)	(29.70, 91.130)	(11.80, 99.800)	(33.38, 104.68)	(30.70, 111.08)	(43.95, 116.07)
(21.10, 79.050)	(23.88, 91.250)	(31.63, 99.980)	(15.25, 104.87)	(23.48, 111.30)	(39.80, 116.47)
(10.92, 79.830)	(53.72, 91.400)	(64.17, 100.07)	(54.80, 105.17)	(40.82, 111.68)	(23.40, 116.68)
(44.17, 80.070)	(26.10, 91.580)	(49.63, 100.17)	(38.48, 106.22)	(21.87, 111.97)	(20.67, 116.72)
(13.00, 80.180)	(22.35, 91.820)	(5.300, 100.27)	(23.92, 106.53)	(43.65, 112.00)	(36.68, 116.98)
(73.50, 80.230)	(31.48, 92.050)	(26.87, 100.43)	(35.55, 106.67)	(16.83, 112.33)	(30.52, 117.03)
(50.35, 80.250)	(49.80, 92.080)	(13.73, 100.50)	(10.82, 106.67)	(68.50, 112.43)	(34.28, 117.30)
(16.53, 80.800)	(58.45, 92.150)	(7.200, 100.60)	(26.58, 106.72)	(37.78, 112.55)	(27.33, 117.47)
(26.75, 80.880)	(33.95, 92.620)	(4.220, 100.70)	(47.93, 106.98)	(33.03, 112.58)	(24.45, 118.07)
(43.95, 81.330)	(11.67, 92.720)	(41.98, 101.07)	(2.900, 107.00)	(59.45, 112.58)	(56.92, 118.37)
(77.50, 82.230)	(56.00, 92.880)	(22.67, 101.40)	(33.07, 107.03)	(42.40, 112.90)	(9.750, 118.73)
(58.30, 82.900)	(38.83, 93.380)	(3.120, 101.55)	(40.77, 107.40)	(25.75, 112.98)	(32.00, 118.80)
(55.03, 82.900)	(42.82, 93.520)	(36.75, 101.60)	(0.000, 107.50)	(3.200, 113.03)	(28.97, 118.87)
(41.72, 82.950)	(24.77, 93.900)	(56.07, 101.83)	(61.27, 108.02)	(28.20, 113.07)	(42.27, 118.97)
(46.73, 83.000)	(51.67, 94.380)	(14.97, 102.08)	(24.70, 108.05)	(23.13, 113.32)	(26.08, 119.28)
(17.72, 83.270)	(36.20, 94.630)	(60.33, 102.27)	(57.77, 108.12)	(52.02, 113.33)	(23.52, 119.57)
(26.75, 83.370)	(40.13, 94.780)	(6.170, 102.28)	(16.03, 108.18)	(54.47, 113.58)	(53.73, 119.78)
(20.25, 85.830)	(27.48, 95.020)	(27.88, 102.30)	(22.82, 108.35)	(34.72, 113.65)	(30.23, 120.17)
(65.78, 87.950)	(16.90, 96.180)	(71.98, 102.47)	(50.37, 108.75)	(62.55, 114.00)	(33.77, 120.25)
(47.73, 88.080)	(41.63, 96.880)	(25.02, 102.68)	(34.30, 108.93)	(30.63, 114.07)	(36.07, 120.33)
(26.67, 88.370)	(58.42, 97.400)	(38.72, 103.10)	(53.43, 108.98)	(11.02, 114.17)	(15.17, 120.57)
(22.65, 88.450)	(36.33, 98.030)	(3.780, 103.22)	(32.72, 109.03)	(22.32, 114.17)	(16.42, 120.60)
(30.95, 88.630)	(25.12, 98.480)	(36.05, 103.88)	(36.60, 109.50)	(37.07, 114.50)	(41.13, 121.12)
(24.85, 89.370)	(39.77, 98.520)	(1.370, 103.92)	(18.23, 109.52)	(48.07, 114.50)	(31.17, 121.43)

Table C.1: (Continued)

(63.77, 121.62)	(37.10, 127.03)	(53.07, 132.93)	(49.00, 140.27)	(65.73, 150.90)	(7.080, 171.38)
(38.90, 121.63)	(53.75, 127.23)	(35.43, 133.35)	(53.15, 140.70)	(7.470, 151.85)	(1.350, 172.92)
(6.900, 122.07)	(39.93, 127.55)	(67.55, 133.38)	(38.27, 140.90)	(24.30, 153.97)	(52.72, 174.10)
(43.60, 122.27)	(26.20, 127.66)	(7.330, 134.48)	(43.05, 141.33)	(1.080, 154.77)	
(66.77, 123.40)	(47.72, 128.90)	(48.52, 135.17)	(40.68, 141.38)	(50.00, 155.38)	
(41.82, 123.55)	(71.58, 128.92)	(33.45, 135.77)	(45.42, 141.68)	(6.970, 158.22)	
(10.30, 123.97)	(36.03, 129.38)	(45.03, 136.67)	(50.90, 142.17)	(52.97, 158.75)	
(54.00, 123.97)	(42.88, 129.47)	(37.38, 136.90)	(27.08, 142.18)	(61.85, 160.57)	
(24.33, 124.17)	(28.38, 129.55)	(50.60, 137.08)	(46.92, 142.73)	(56.32, 160.83)	
(43.90, 125.22)	(62.08, 129.75)	(34.73, 137.67)	(7.380, 143.92)	(68.80, 161.28)	
(49.17, 125.23)	(33.58, 130.38)	(76.00, 137.90)	(73.18, 143.93)	(11.35, 162.35)	
(39.03, 125.78)	(31.63, 130.58)	(9.480, 138.08)	(13.55, 144.83)	(55.20, 165.98)	
(45.68, 126.62)	(25.83, 131.23)	(33.12, 139.78)	(43.33, 145.58)	(60.35, 166.00)	
(35.12, 126.82)	(43.12, 131.90)	(39.72, 140.10)	(44.02, 145.82)	(19.28, 166.65)	
(41.72, 126.92)	(50.07, 132.13)	(36.05, 140.13)	(46.20, 150.50)	(8.730, 167.73)	

Bibliography

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd edn, John Wiley & Sons, New York.

Bates, D., Reames, F. and Wahba, G. (1990). Getting better contour plots with s and gcvpack, *Technical Report No. 865.* Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd edn, Academic Press, New York.

Cox, D. (1988). Approximation of method of regularization estimators, *The Annals of Statistics* **16**: 694–712.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions, *Numer. Math.* **31**: 377–403.

Goerss, J. and Phoebus, P. (1991). The multivariate optimum interpolation analysis of meteorological data at FNOC, *Naval Oceanographic and Atmospheric Research Laboratory, Atmospheric Directorate, Monterey, CA 93943-5006.*

Hall, P. and Johnstone, I. (1992). Empirical functionals and efficient smoothing parameter selection, *Journal of the Royal Statistical Society B* **54**: 475–530.

Hardle, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum?, *Journal of the American Statistical Association* **83**: 86–101.

Hollingsworth, A. and Lonnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus* **38**: 111–136.

Johnson, D. (1986). Summary of the Proceedings of the First National Workshop on the Global Weather Experiment, *Bulletin American Meteorological Society* **67**: 1135–1143.

Kimeldorf, G. and Wahba, G. (1970). A correspondence between bayesian estimation on stochastic processes and smoothing by splines, *The Annals of Mathematical Statistics* **41**: 495–502.

Lonnberg, P. and Hollingsworth, A. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data Part II: The covariance of height and wind errors, *Tellus* **38**: 137–161.

Lorenc, A. C. (1986). Analysis methods for numerical weather prediction, *Quarterly Journal of the Royal Meteorological Society* **112**: 1177–1194.

Mardia, K. V. and Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika* **71**: 135–146.

Mitchell, H., Charette, C., Chouinard, C. and Brasnett, B. (1990). Revised Interpolation Statistics for the Canadian Data Assimilation Procedure: Their Derivation and Application, *Monthly Weather Review* **118**: 1591–1614.

Pailleux, J., Uppala, S., Illari, L. and Dell'Osso, L. (1986). Fgge re-analyses and their use at ecmwf, *ECMWF Tech. Memo* **126**: 38.

Parrish, D. F. and Derber, J. C. (1992). The National Meteorological Center's spectral statistical interpolation analysis system, *Monthly Weather Review* **120**: 1747–1763.

Stanford, J. L. (1979). Latitudinal-wavenumber power spectra of stratospheric temperature fluctuations, *Journal of the Atmospheric Sciences* **36**: 921–931.

Stein, M. (1989). Spline smoothing with an estimated order parameter, *Technical Report No. 264*. Department of Statistics, The University of Chicago, Chicago, Ill.

Stein, M. (1990). A comparison of generalized cross validation and modified maximum likelihood for estimating the parameters of a stochastic process, *The Annals of Statistics* **18**: 1139–1157.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression, *Journal of the Royal Statistical Society B* **40**: 364–372.

Wahba, G. (1982a). Variational methods in simultaneous optimum interpolation and initialization, *The Interaction Between Objective Analysis and Initialization* pp. 178–185. Williamson, D. editor, Atmospheric Analysis and Prediction Division, National Center for Atmospheric Research, Boulder, CO.

Wahba, G. (1982b). Vector splines on the sphere, with application on the estimation of vorticity and divergence from discrete, noisy data, *Multivariate Approximation Theory* **2**: 407–429. Schempp, W. and Zeller, K. eds., Birkhauser Verlag, Basel, Boston, Stuttgart.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline, *Journal of the Royal Statistical Society B* **45**: 133–150.

Wahba, G. (1985a). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, *The Annals of Statistics* **13**: 1378–1402.

Wahba, G. (1985b). Variational methods for multidimensional inverse problems, *Remote Sensing Retrieval Methods* pp. 385–408. Deepak, A., Fleming, H., and Chahine, M. eds., Deepak, A. publishing.

Wahba, G. (1989). On the dynamic estimation of relative weights for observations and forecast, *Remote Sensing Retrieval Methods* pp. 347–358. Deepak, A., Fleming, H., and Theon, J. eds., Deepak, A. Publishing, Hampton, VA.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol.59.

Wahba, G. and Wang, Y. (1990). When is the optimal regularization parameter insensitive to the choice of the loss function?, *Communications in Statistics - Theory and Methods* **19**: 1685–1700.

Wahba, G., Johnson, D. R. and Reames, F. (1990). Multiple smoothing and weighting parameters in direct variational methods for objective analysis of meteorological information, *Assimilation of Observations in Meteorology and Oceanography* pp. 448–453. Proceedings of the International Symposium, World Meteorologican Organization, Clermont-Ferrand, France, LeDimet, F.-X., and Talagrand, O. eds.