

**A Practical Perspective on Ill-Posed Inverse Problems:
A Review with Some New Developments***

by

Finbarr O'Sullivan

**Department of Statistics
University of California
Berkeley, California**

Technical Report No. 52

November 1985

***Research Supported by the National Science Foundation
under Grant No. MCS-8403239.**

**Department of Statistics
University of California
Berkeley, California**

A Practical Perspective on Ill-Posed Inverse Problems: A Review with Some New Developments

Finbarr O'Sullivan

Department of Statistics
University of California
Berkeley, CA 94720.

ABSTRACT

Ill-posed inverse problems arise in many branches of science and engineering. In the typical situation one is interested in recovering a whole function given a finite number of noisy measurements on functionals. Performance characteristics of an inversion algorithm are studied via the mean square error which is decomposed into bias and variability. Variability calculations are often straightforward, but useful bias measures are more difficult to obtain. An appropriate definition of what geophysicists call the Backus-Gilbert averaging kernel leads to a natural way of measuring bias characteristics. Moreover, the ideas give rise to some important experimental design criteria. It can be shown that the optimal inversion algorithms are methods of regularization procedures, but to completely specify these algorithms the signal to noise ratio must be supplied. Statistical approaches to the empirical determination of the signal to noise ratio are discussed; cross-validation and unbiased risk methods are reviewed and some extensions, which seem particularly appropriate in the inverse problem context, are indicated. Linear and non-linear examples from medicine, meteorology and geophysics are used for illustration.

Key-words: ill-posed inverse problems, mean square error, averaging kernel, bias, variability, optimal inversion algorithm, experimental design, B-splines, cross-validation, unbiased risk, predictive mean square error, estimable loss, random spheres, integral equations, radiative transfer, distributed parameter identification.

This work was supported by the National Science Foundation under Grant Number MCS-8403239.

A Practical Perspective on Ill-Posed Inverse Problems:

A Review with Some New Developments

Finbarr O'Sullivan

Department of Statistics

University of California

Berkeley, CA 94720.

1. Introduction

Inverse problems have got to do with situations where one is interested in making inferences about a phenomenon from partial or incomplete information. Accordingly, statistical estimation and model building are both inverse problems. In modern science there is an increasingly important class of inverse problems which are not amenable to classical statistical estimation procedures and such problems are termed ill-posed. The notion of ill-posedness is due to Tikhonov[22] and an extended treatment of the concept appears in Tikhonov and Arsenin[23]. In an ill-posed inverse problem, a classical least squares, minimum distance or maximum likelihood solution may not be uniquely defined, moreover the sensitivity of such solutions to slight perturbations in the data is often unacceptably large.

A typical example of an ill-posed inverse problem is described by Nychka et al.[19]. Here one is interested in the estimation of a 3-dimensional tumor size distribution in liver tissue, from measurements on cross-sectional slices. A schematic for the experiment is given in Figure 1.1.

Figure 1.1 ABOUT HERE

Modeling tumors as spheres randomly distributed in the tissue an approximate integral relationship between the 3-dimensional distribution of tumor radii and the 2-dimensional distribution of radii observed in cross-sectional slices may be derived. Letting z_i be the observed proportion of 2-dimensional slices with radii in the interval $[z_i, z_{i+1}]$ one has that

$$z_i = \int K(z_i, r) f_3(r) dr + \epsilon_i \quad i = 1, 2, \dots, m,$$

where ϵ_i are measurement/modeling errors. f_3 is the density of 3-dimensional tumor radii and the kernels $K(z_i, r)$ are given by

$$K(z_i, r) = \begin{cases} 0 & \epsilon \leq r < z_i \\ \sqrt{r^2 - z_i^2} & z_i \leq r < z_{i+1} \\ \sqrt{r^2 - z_i^2} - \sqrt{r^2 - z_{i+1}^2} & z_{i+1} \leq r < R \end{cases}$$

Physically, ϵ is the smallest detectable tumor radius and R is the largest possible tumor radius in the given section of tissue. Since $\epsilon \leq z_1 < z_2 < \dots < z_{m+1} < R$, the estimation of f_3 must be restricted to the interval $[\epsilon, R]$. It is easy to appreciate the ill-posedness of this inverse problem. The kernels $K(z_i, \cdot)$ are smooth, and as a result relatively large perturbations in f_3 can give rise to very slight perturbations in the data and conversely. It follows from this that least squares, minimum X^2 , or maximum likelihood solutions will be very sensitive to slight changes in the data.

Ill-posed inverse problems have become recurrent theme in modern science, see for example Crystallography[23], Geophysics[2, 8, 25], Medical Electrocardiograms[19], Meteorology[46, 47], Microfluorimetry[35], Radio Astronomy[24], Reservoir Engineering[27, 37], and Tomography[10, 55]. Corresponding to this broad spectrum of fields of application, there is a wide literature on different kinds of inversion algorithms, that is, techniques for solving the inverse problem. The basic principle common to all such methods is as follows; seek a solution that is consistent both with the observed data and *prior* notions about the physical behavior of the phenomenon under study. Different practical problems have led to unique strategies for implementation of this principle, such as the Method of Regularization[53], Maximum Entropy[24] and Quasi-Reversibility[28]. Understanding the performance characteristics of a given inversion method is

an important issue: Firstly, such information has obvious intrinsic value and secondly, it can critically influence the choice of experimental design, see section 2.

The primary goal of this paper is to identify some tools for assessing the finite sample performance characteristics of an inversion algorithm. These tools, most of which can be found scattered throughout the diverse inverse problem literature, are considered in the context of the following generalized non-linear regression model: Measurements, z_i , are of the form

$$z_i = \eta(x_i, \theta) + \epsilon_i, \quad i=1, 2, \dots, m,$$

where θ is in Θ (the nominal parameter space), $\eta(x_i, \cdot)$ are linear or non-linear functionals of θ , and the ϵ_i are measurement errors assumed to have mean zero. In all that we discuss, there will be an underlying assumption that the unknown true function, θ , is well approximated by a *smooth* function. Though this assumption does not allow us talk about highly discontinuous functions, such as those that arise in typical pattern recognition problems, the model is still quite general, and includes for example linear and non-linear integral equations of the first kind. Examples used later on include the temperature retrieval problem in satellite meteorology, and the system identification problem of reservoir engineering.

1.1. Inversion Algorithms and the Method of Regularization

An inversion algorithm, S , is a mapping that takes data into parameter estimates.

$$\hat{\theta} = S z$$

When S is linear then $\hat{\theta}$ can be written as a linear combination of *impulse response functions*, s_i ; i.e.

$$\hat{\theta}(t) = \sum_{i=1}^m s_i(t) z_i$$

where $s_i = S e_i$, and e_i is the i 'th unit vector in R^m . In statistical terms, an inversion algorithm would be simply called an estimator, but here the terms estimator and inversion algorithm are used interchangeably.

One of the most useful techniques for generating inversion algorithms or estimators is the Method of Regularization (MOR), see Titterton[54]. The MOR procedure is due to Tikhonov[52, 53]. There are various possible implementations of the method but they all amount to choosing $\hat{\theta}$ to be the minimizer of a weighted combination of two functionals: The first functional measures lack of fit to the observed data and the second measures physical plausibility of the estimate. For example, we might choose $\hat{\theta}$ to be the minimizer of a criterion of the form

$$\frac{1}{m} \sum_{i=1}^m [z_i - \eta(x_i, \theta)]^2 + \lambda J(\theta) \quad , \quad \lambda > 0. \quad (1.1)$$

The functional J is chosen so that highly *irregular* or *physically implausible* θ 's have large values. Statisticians will recognize this method as a version of Penalized Likelihood Estimation described by Good and Gaskins[21]; the sum of squares is the likelihood part and the functional J the penalty term. The method is equivalent to the Method of Sieves introduced by Grenander[22]. Also, if J is chosen by some information principle such as entropy, $J(\theta) = -\int \theta(t) \log \theta(t) dt$, then the method of regularization yields a procedure equivalent to the Method of Maximum Entropy pioneered by Jaynes, see Jaynes[24] and Chapter 4 of McLaughlin[34].

When the functionals, $\eta(x_i, \cdot)$, are linear in θ , and J is quadratic with $J(\theta) \geq 0$ ($= 0$ for $\theta = 0$) then the solution to (1.1) is linear in the observed data. Moreover, in this case the MOR has an interesting Bayesian interpretation. To see this, first suppose that Θ is finite dimensional, i.e. $\Theta = \text{span}_{1 \leq k \leq K} \{\phi_k\}$ with ϕ_k linearly independent. The elements of Θ can be written as $\sum_k \beta_k \phi_k$ for $\beta = (\beta_1, \beta_2, \dots, \beta_K)'$ in R^K , so that Θ can be identified with R^K . Since J is quadratic and $J(0) = 0$, $J(\theta)$ can be expressed as a quadratic form in β , $J(\theta) = \beta' \Omega \beta$, for some positive semi-definite matrix Ω . It follows that the MOR estimator is $\sum_k \hat{\beta}_k \phi_k$ where $\hat{\beta}$ minimizes

$$\frac{1}{m} \sum_{i=1}^m [z_i - X_i' \beta]^2 + \lambda \beta' \Omega \beta \quad .$$

Here X is a design matrix, $X_{ik} = \eta(x_i, \phi_k)$. Thus $\hat{\beta}$ is given by

$$\hat{\beta} = S z \quad \text{where} \quad S = [X' X + m \lambda \Omega]^{-1} X' \quad .$$

Obviously, the MOR estimator is linear in the observed data. A Bayesian interpretation for $\hat{\beta}$ is obtained by specifying a Gaussian prior with mean zero and covariance matrix proportional to Ω^{-1} . Then, if the ϵ_i 's are independent and identically distributed Gaussian random variables with mean zero, the MOR estimator, $\hat{\beta}$, is the posterior mean of β given the data.

The foregoing statements carry through to more general settings. If Θ is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, the $\eta(x_i, \cdot)$ are bounded linear functionals and J is quadratic, $J(\theta) = \langle \theta, W\theta \rangle$, where W is positive semi-definite, then it can be shown, see Cox[13] for example, that the MOR estimators have the form

$$\hat{\theta}_\lambda = S s \quad \text{where} \quad S = [X' X + m \lambda W]^{-1} X'$$

where the *design matrix*, X , is an operator derived from the functionals $\eta(x_i, \cdot)$. Thus S is a linear operator from the data space into Θ . General Bayesian interpretations for the method are also available, these are discussed by Kimeldorf and Wahba[26]. Further results on the optimality of the MOR are described in section 2.

1.2. Outline

Performance characteristics of an inversion algorithm are studied via the mean square error which can be split into bias and variability components. Bias measures the systematic error while variability measures the random error. In section 2 we consider linear inversion algorithms and describe some ways of measuring bias and variability. Variability is calculated in a direct manner. For bias it is convenient to introduce a generalized version of what geophysicists call the Backus-Gilbert averaging kernel, see Backus and Gilbert[6, 7]. The generalized notion of the averaging kernel allows one compute the maximum or average expected bias and also leads to some natural design criteria. These are described and illustrated in section 2.4. Optimal inversion algorithms can be found and these turn out to be MOR procedures. The use of B-splines for obtaining numerically convenient and accurate approximations to the averaging kernel and bias is described in section 3. Though the theory of linear inverse problems is fairly well established, the field of non-linear inverse problems is in its infancy. There are many exciting and challenging

problems that need to be tackled in this area. The performance of MOR estimators when applied to two interesting non-linear inverse problems is discussed in section 4. One of these problems arises in satellite meteorology and is concerned with the estimation of atmospheric characteristics such as temperature from upwelling radiance measurements. The second problem is of major interest in reservoir engineering. It concerns the estimation of reservoir characteristics, such as the ease of flow of fluid in the reservoir, from pressure-history data measured at distributed well sites. As described in section 2, optimal inversion algorithms though MOR procedures are not fully specified without supplying the signal-to-noise ratio. For a given problem this will not be known and so has to be empirically determined. The final section of the paper deals with this issue; the methods of cross validation and unbiased risk are described and some relevant extensions to ill-posed inverse problems are developed.

2. Finite Sample Performance of an Inversion Algorithm

The quality of an inversion algorithm at some point, t , is measured by comparing the estimate, $\hat{\theta}(t)$, to the true value, $\theta(t)$. This difference can be decomposed into systematic and random components as

$$\theta(t) - \hat{\theta}(t) = \theta(t) - E\hat{\theta}(t) + E\hat{\theta}(t) - \hat{\theta}(t) ,$$

where the expectation is over the distribution of possible errors. The average performance of the inversion at t is measured by the mean square error (MSE).

$$\begin{aligned} \text{MSE}(t) &= E[\theta(t) - \hat{\theta}(t)]^2 = [\theta(t) - E\hat{\theta}(t)]^2 + E[\hat{\theta}(t) - E\hat{\theta}(t)]^2 \\ &= \text{bias}^2(t, \theta) + \text{var}(t, \theta) . \end{aligned}$$

The mean square error depends both on θ and the assumed error distribution. It is sum of the squared bias, $\text{bias}^2(t, \theta)$, and the variance, $\text{var}(t, \theta)$. If the inversion algorithm is designed solely to minimize bias then the variance dominates the mean square error and vice-versa. Thus a good inversion algorithm must balance bias and variability, i.e. unbiasedness may be a desirable property in classical statistical estimation, but in the context of ill-posed inverse problems it is not.

Mean square error performance of an inversion algorithm can, in principle, be found by Monte-Carlo simulation. Modern computing resources are making this a very viable and practical approach. For linear problems one can avoid direct Monte-Carlo simulation and in the process obtain some useful insights which can be applied to more complex situations.

2.1. Linear Problems

By a linear problem we shall mean that both the functionals, $\eta(z, \cdot)$, are linear in θ , and the inversion algorithm, S , is linear in the data.

2.1.1. Variability

Variability computations for a linear inversion algorithms are very straightforward. Here, variability does not depend on θ and

$$\text{var}(t, \theta) \equiv \text{var}(t) = \text{var} \left[\sum_{i=1}^m s_i(t) \epsilon_i \right]$$

where s_i is the impulse response function. Thus if the errors, ϵ_i , have covariance Σ_{ϵ} , then

$$\text{var}(t) = \mathbf{s}(t)' \Sigma_{\epsilon} \mathbf{s}(t)$$

where $\mathbf{s}(t) = (s_1(t), s_2(t), \dots, s_m(t))'$. In particular, if the errors are independent with constant variance σ^2 , then the variability is simply

$$\text{var}(t) = \sigma^2 \mathbf{s}(t)' \mathbf{s}(t) .$$

2.1.2. Bias

Bias properties are best understood by introducing the notion of the Averaging Kernel. In certain circumstances the averaging kernel yields a representation of the form

$$E\hat{\theta}(t) = \int A(t, s) \theta(s) ds , \quad (2.1)$$

for the systematic part of the estimate. The function $A(t, \cdot)$ is known as the averaging kernel at t and $A(t, \cdot)$ determines the nature of the bias incurred at t . The averaging kernel is related to what engineers and astronomers call the **point-spread** function. The point-spread function at t is defined to be the solution obtained by the inversion algorithm when the true function is a Dirac δ -function at t and there is no measurement error. Thus from (2.1) if the averaging kernel at t is $A(t, \cdot)$ then the point-spread at t is the function $A(\cdot, t)$.

The representation in (2.1) is an L_2 representation for the averaging kernel. In the Geophysics literature this representation is known as the Backus-Gilbert averaging kernel, after two geophysicists Backus and Gilbert[6, 7]. Alternative representations for the averaging kernel are also possible and these alternative representations are more useful when it comes to computing bias. The averaging kernel and its generalizations are described next.

2.2. Averaging Kernel

2.2.1. Backus-Gilbert Formulation

Backus and Gilbert worked in an integral equation context.

$$\eta(x, \theta) = \int K_i(s) \theta(s) ds \quad i=1, 2, \dots, m$$

and the kernels $K_i(\cdot)$ are known smooth functions. For a linear inversion algorithm the $E\hat{\theta}(t)$ can be written as

$$E\hat{\theta}(t) = \sum_{i=1}^m s_i(t) \int K_i(s) \theta(s) ds .$$

So, taking the summation inside the integral sign, we have

$$E\hat{\theta}(t) = \int A(t, s) \theta(s) ds$$

where

$$A(t, s) = \sum_{i=1}^m s_i(t) K_i(s) .$$

The function $A(t, \cdot)$ is the Backus-Gilbert averaging kernel for the inversion algorithm S at t . For illustration, an averaging kernel corresponding to a method of regularization procedure applied to the tumor size distribution problem, described in section 1, is given in Figure 2.1.

Figure 2.1 ABOUT HERE

One can see that the kernel is well centered about the point of interest, $r = .4$. Moreover the kernel seems fairly symmetric so that if the 3-dimensional tumor radius density were locally linear in the neighborhood of this point then the inversion algorithm would be locally unbiased. Properties of the averaging kernel can be varied by changing the regularization parameter, λ - large values of λ cause the averaging kernel to be more spread out. Techniques for empirically selecting this parameter are discussed in section 5.

The center, spread and skewness of the averaging kernel give a rough appreciation for its behavior. Assuming they exist, these are defined to be the first, second and third moment of the absolute value of the averaging kernel when suitably normalized. Letting

$$\bar{A}_\lambda(t, \cdot) = \frac{A_\lambda(t, \cdot)}{\int |A_\lambda(t, s)| ds}, \text{ then}$$

Characteristics of the Averaging Kernel

center:
$$c(t) = \int |\bar{A}_\lambda(t, s)| s ds$$

spread:
$$sp(t) = \sqrt{\int |\bar{A}_\lambda(t, s)| (s - c(t))^2 ds}$$

skewness:
$$sk(t) = \int |\bar{A}_\lambda(t, s)| \left| \frac{s - c(t)}{sp(t)} \right|^3 ds$$

The skewness is dimensionless while the center and spread are in t -units. Skewness is important since a symmetric averaging kernel ($sk(t)=0$.) will exactly pass a linear trend.

Intuitively, the bias at a point, t , is determined by the how close the averaging kernel is to a Dirac δ -function at t . Backus and Gilbert tried to develop some direct measures of the nearness of the averaging kernel to a Dirac δ -function - " δ -ness of the averaging kernel". By choosing the inversion algorithm, so that the averaging kernel is as δ -like as possible, subject to some upper bound on the size of the standard error, one obtains so called Backus-Gilbert inversion algorithms. The idea seems perfectly reasonable, however, there is some degree of arbitrariness in the δ -ness criteria defined by Backus and Gilbert, moreover in general it is not true that the function maximizing a δ -ness criterion will necessarily be the Dirac δ -function. Problems with the δ -ness criteria really arise because the Backus-Gilbert calculus takes place in an L_2 setting where evaluation is not a continuous linear functional. If we work in a space where evaluation is continuous, we can derive a more refined formulation of the averaging kernel and use a straightforward calculus to assess δ -ness. The refined definition of the averaging kernel also allows one deal with

more general linear functionals $\eta(x, \theta)$.

3.3.3. Refined Formulation of the Averaging Kernel

Preliminaries: Linear Functionals and Representer

We shall need the notion of a representer of a continuous linear functional. To motivate this concept, consider first the case where $\Theta = \text{span}_{1 \leq k \leq K} \{\phi_k\}$ and the ϕ_k are linearly independent. Here, elements of Θ are identified by a K -vector of coefficients, β in R^K . Moreover the usual inner product on R^K determines an inner product on Θ by

$$\langle \theta_1, \theta_2 \rangle = \beta_1' \beta_2$$

where $\theta_1 = \sum_{k=1}^K \beta_{1k} \phi_k$, and $\theta_2 = \sum_{k=1}^K \beta_{2k} \phi_k$. If $\eta(x, \cdot)$ is linear then for any θ in Θ

$$\eta(x, \theta) = X_i' \beta \quad (2.2)$$

where $\theta = \sum_{k=1}^K \beta_k \phi_k$ and $X_i = (\eta(x, \phi_1), \eta(x, \phi_2), \dots, \eta(x, \phi_K))'$. X_i determines an element,

$\xi_x = \sum_{k=1}^K X_{ik} \phi_k$, in Θ with the property that for all θ in Θ

$$\eta(x, \theta) = \langle \xi_x, \theta \rangle$$

In functional analysis terms, ξ_x is the *representer* of the linear functional $\eta(x, \cdot)$. An important linear functional is evaluation at point. The representer in Θ of evaluation at t is given by

$$e_t = \sum_{k=1}^K \phi_k(t) \phi_k \quad .$$

One can easily verified that

$$\theta(t) = \langle e_t, \theta \rangle \quad , \quad \text{for all } \theta \text{ in } \Theta .$$

The notion of a representer can be extended to more elaborate function spaces. The level of functional analysis needed to understand this is very elementary and the interested reader might consult Rudin[45]. It is important to realize that, in general, the form of the representer depends

on which inner product is used. Let Θ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. A linear functional l is continuous if there is a constant M such that

$$|l(\theta)| \leq M\|\theta\| \quad \text{for all } \theta \text{ in } \Theta .$$

Corresponding to any continuous linear functional, l , there exists a representer ξ in Θ such that

$$l(\theta) = \langle \xi, \theta \rangle \quad \text{for all } \theta \text{ in } \Theta .$$

This is known as the Riesz representation theorem, Rudin[45]. A Hilbert space of real valued functions in which evaluation is continuous is known as Reproducing Kernel Hilbert Space (RKHS). Reproducing kernel Hilbert spaces play an important role in applied mathematics. The importance of RKHS in the study of ill-posed inverse problems has long been emphasised by Professor Wahba, see Wahba[59] and the references cited therein. Evaluation is not continuous in L_2 , but is continuous in the space of functions whose first derivatives are square integrable.

Representer of the Averaging Kernel

Given the notion of a representer, the generalization of the averaging kernel is very simple. Let Θ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and suppose that the functionals $\eta(x, \cdot)$ are continuous with representers ξ_x . Since

$$E\hat{\theta}(t) = \sum_{i=1}^m s_i(t) \eta(x_i, \theta) ,$$

the representer of the *averaging kernel*, $A(t)$, is given by

$$A(t) = \sum_{i=1}^m s_i(t) \xi_{x_i} .$$

Thus the averaging kernel is a linear combination of the representers of the functional $\eta(x, \cdot)$. The linear combination is determined by the impulse response functions of the inversion algorithm.

2.3. Bias Measures and Some Design Criteria

The more general formulation of the averaging kernel leads to natural ways of measuring bias and this in turn motivates some useful design criteria. Let Θ be such that evaluation at t is continuous and let e_t denote the corresponding representer. From the averaging kernel, the bias at t may be written as

$$\theta(t) - \hat{\theta}(t) = \langle e_t - A(t), \theta \rangle .$$

Average Bias

The representation, $\langle e_t - A(t), \theta \rangle$, can be used to compute the expected squared bias with respect to a prior distribution on possible θ values. Thus, if $\Theta = \text{span}_{1 \leq k \leq K} \{\phi_k\}$ and a prior mean and covariance for θ is specified by means of the coefficients of the ϕ_k 's, β ,

$$E_\beta[\beta] = \beta_0 \quad , \quad \text{Var}_\beta[\beta] = \Sigma_\beta \quad ,$$

then the average expected bias is

$$b^2(t) = E_\beta[\theta(t) - E\hat{\theta}(t)]^2 = E_\beta[\langle e_t - A(t), \theta \rangle^2] = \mathbf{c}_t' \Sigma_\beta \mathbf{c}_t + \langle e_t - A(t), \theta_0 \rangle^2$$

where $\mathbf{c}_t = (\langle e_t - A(t), \phi_1 \rangle, \langle e_t - A(t), \phi_2 \rangle, \dots, \langle e_t - A(t), \phi_K \rangle)'$ and $\theta_0 = \sum_{k=1}^K \beta_{0k} \phi_k$.

Maximum Bias

A less sophisticated measure of bias is the maximum bias over all functions in Θ whose norm is less than some specified value, μ . From the averaging kernel representation for bias and the Cauchy-Schwartz inequality, this is given by

$$\sup_{\|\theta\|^2 \leq \mu^2} [\theta(t) - E\hat{\theta}(t)]^2 = \sup_{\|\theta\|^2 \leq \mu^2} \langle e_t - A(t), \theta \rangle^2 = \|e_t - A(t)\|^2 \mu^2 .$$

Thus letting $b_M(t) = \|e_t - A(t)\|$, $b_M^2(t)$ is proportional to the maximum bias over any ball in Θ .

Figure 2.2 plots the maximum bias, $b_M(t)$, and standard error, square root of variability, for a MOR inversion applied to the tumor size distribution problem. Both the bias and standard error have poor behavior for small radii, suggesting that there is difficulty in getting reliable

estimates of the size distribution for small radii. This is largely a consequence of the fact that small tumors are hard to detect, see Nychka et al.[38] for more discussion. The ripples in the plot are due to the finite sample size, $m=50$, in this illustration. The vertical scales on the bias and variability plots are left unspecified. If particular values for the μ and σ are assigned then an appropriate scaling can be setup.

Figure 2.2 ABOUT HERE

Design Criteria

Combining bias with variability gives an overall assessment of the performance of the inversion algorithm which can be used for design purposes. The average bias measure gives rise to an Average Mean Square Error design criterion (AMSE):

$$AMSE(t) = \overline{b^2}(t) + var(t) \quad .$$

Also, from the maximum bias measure, we obtain the Maximum Mean Square Error design criterion (MMSE):

$$MMSE(t) = \mu^2 b_M^2(t) + var(t) \quad .$$

2.4. Optimal Inversion Algorithms and Experimental Design

Since both the average and maximum mean square errors depend on the inversion algorithm, we can select the algorithm which performs best with respect to either of these criteria. Interestingly, the solution one obtains in each case is a MOR procedure. This analysis can be carried out in a very general setting, however the structure of these results becomes most clear when Θ is finite dimensional and the measurement errors are mean zero uncorrelated with constant variance σ^2 .

Minimizing the Average Mean Square Error

Let $\mathbf{a}(t) = (s_1(t), s_2(t), \dots, s_K(t))$. For simplicity suppose that the prior for β has zero mean and covariance $\tau^2 \Sigma_\beta$. The average mean square error can be written as

$$AMSE(t) = \tau^2 (X_i - X \mathbf{a}(t))' \Sigma_\beta (X_i - X \mathbf{a}(t)) + \sigma^2 \mathbf{a}(t)' \mathbf{a}(t) ,$$

where $X_i = (\phi_1(t), \phi_2(t), \dots, \phi_K(t))'$ and the i 'th row of X is given in equation (2.2). Minimizing with respect to $\mathbf{a}(t)$ we have

$$[X' \Sigma_\beta X + \frac{\sigma^2}{\tau^2} I] \hat{\mathbf{a}}(t) = X' \Sigma_\beta X_i .$$

This holds for any t . If Σ_β is invertible, then, by straightforward algebra, the optimal inversion algorithm can be expressed as

$$\hat{S} = [X' X + \frac{\sigma^2}{\tau^2} \Sigma_\beta^{-1}]^{-1} X'$$

Comparing this with the forms given in section 1.1, it can be seen that the optimal algorithm corresponds to a MOR estimator with $\lambda = \frac{\sigma^2}{m \tau^2}$ and $J(\theta) = \beta \Sigma_\beta^{-1} \beta$. λ^{-1} is interpreted as the signal-to-noise ratio.

Minimizing the Maximum Mean Square Error

The maximum mean square error in a ball of radius μ is given by

$$MMSE(t) = \mu^2 [X_i - \sum_{i=1}^m s_i(t) X_i]' [X_i - \sum_{i=1}^m s_i(t) X_i] + \sigma^2 \mathbf{a}(t)' \mathbf{a}(t)$$

Minimizing with respect to $\mathbf{a}(t)$, we get that the optimal vector satisfies

$$[X' X + \frac{\sigma^2}{\mu^2} I] \hat{\mathbf{a}}(t) = X' X_i$$

Thus the optimal inversion algorithm is given by

$$\hat{S} = [X' X + \frac{\sigma^2}{\mu^2} I]^{-1} X'$$

Again, this is a MOR inversion algorithm with $\lambda = \frac{\sigma^2}{m \mu^2}$ and $J(\theta) = \|\theta\|^2$. λ^{-1} is again

interpreted as a signal-to-noise ratio.

Versions of the above results have appeared at several times in the literature. Weinreb and Crosby[60] credit Foster[18], see also Strand and Westwater[51]. The optimal inversion algorithm from the point of view of average mean square error is called the minimum-rms solution while the optimal solution for maximum mean square error is the minimum information solution. Generalizations of these results have appeared in the statistics literature. Kimeldorf and Wahba[26] show that the minimum-rms solution is sometimes interpretable as an optimal Bayesian procedure. Minimum information solutions are also been termed minimax. Results on the minimaxity of the MOR are given by Li[29], and also Speckman[48]. In practice the signal-to-noise ratio is not known so that the parameter λ needs to be set empirically. Some statistical methods available for doing this are given in section 5.

Experimental Design

The MMSE and the AMSE are also functions of the the design points, $x, i=1,2,\dots,m$, and an optimal experimental design can be defined as the design making the MMSE or the AMSE minimum. Since one may not be interested in performance at just a single point, t , an integrated AMSE or MMSE over some region of interest is often more appropriate. Crosby and Weinreb[60] have carried out a program of this kind in connection with the selection of spectral wavelengths for satellite radiometers. Their criterion is reduced to a simple trace criterion, see eqn. (10) of their paper. More recently, Wahba[58] has also discussed the design issue. She presents two design criterion: one of which is akin to Weinreb and Crosby's criterion, see eqn. (13), and a second which is termed the "degrees of freedom for signal" criterion. The use of integrated mean square error as a design criterion is not new to statisticians. Box and Draper[9] proposed this as a design criterion for *model-robust* response surface designs. Further discussion of this literature can be found in section 5 of the recent paper by Steinberg and Hunter[50].

3. Numerical Approximation of the Averaging Kernel with B-splines

Exact computation of averaging kernels requires the manipulation of the representers of the functionals $\eta(x_i, \cdot)$ for $i=1, 2, \dots, m$. In reproducing kernel Hilbert spaces there are theoretical formulas available for the evaluation of representers, see Nychka et al.[38] for example. However, direct evaluation of averaging kernels by means of such formulas is extremely inefficient, computationally, and other approaches are needed. Approximating the elements of Θ by simpler forms, it is often possible to obtain highly efficient methods for evaluating the averaging kernels at a negligible loss in accuracy. To illustrate this we consider one very popular MOR procedure for estimating a one-dimensional real valued function, θ defined on an interval $[a, b]$: $\hat{\theta}$ is the minimizer over the Sobolev space $W_2^2[a, b] = \Theta$ of

$$\frac{1}{m} \sum_{i=1}^m [z_i - \eta(x_i, \theta)]^2 + \lambda \int_a^b [\ddot{\theta}(t)]^2 dt \quad \lambda > 0.$$

$W_2^2[a, b]$ is an infinite dimensional Hilbert space with inner product

$$\langle \theta, \phi \rangle = \int \theta(t) \phi(t) dt + \int \ddot{\theta}(t) \ddot{\phi}(t) dt \quad \theta, \phi \text{ in } W_2^2[a, b].$$

Sobolev spaces are discussed at length in the book by Adams[1]. It is well known that the elements of Θ can be approximated to arbitrarily high degree of accuracy by cubic B-splines and we can make good advantage of this in the computation of the averaging kernels. Before describing this we pause briefly to describe cubic B-splines. The standard reference on B-splines is the book by de Boor[16]. Throughout this section we continue to assume that the functionals $\eta(x_i, \cdot)$ are linear.

3.1. B-splines

Consider a set of distinct knot points $[t_1 < t_2 < \dots < t_{K+1}]$ with $t_1 \leq a$ and $b \leq t_{K+1}$. (Multiple knots are also allowed, see de Boor[16] for more details). A cubic spline on $[a, b]$ is a cubic polynomial between successive knot points joined up at the knots so as to have continuous second derivative over the interval $[a, b]$. By increasing the knot density in $[a, b]$, functions in $W_2^2[a, b]$ can be very closely approximated by cubic splines. Cubic splines can be expressed as a

linear combination of basis elements $\{B_k\}_{k=1}^K$. The elements, B_k , of the basis are called B-splines (B stands for basis) and the entire basis is called the B-spline basis. Each element of the basis is non-negative and has very local support: in fact B_k is zero outside of the interval $[t_k, t_{k+4}]$. Over $[t_k, t_{k+4}]$, B_k is proportional to the probability density for the sum of four independent uniform random variables, U_i , $i=1,2,3,4$, where U_i is defined on $[t_{k+i-1}, t_{k+i}]$. The local support property of B-splines can be used to great advantage in computations. de Boor[16] has developed a set of Fortran programs for manipulating B-splines and these are now available in most modern mathematical software libraries.

3.2. Computation of the Averaging Kernel in a B-spline basis

Let $\Theta_K = \{B_k\}_{k=1}^K$. The MOR estimate $\hat{\theta}_\lambda$ is approximated as $\hat{\theta}_\lambda \approx \sum_{k=1}^K \hat{\beta}_k B_k$ where

$$\hat{\beta} = [X' X + m \lambda \Omega_2]^{-1} X' s \quad (3.1)$$

and $X_{i,k} = \eta(x_i, B_k)$ and $\Omega_2(j, k) = \int_a^b \ddot{B}_j(t) \ddot{B}_k(t) dt$.

Three inner products on Θ_K will be defined. For $\theta = \sum_{k=1}^K \theta_k B_k$, and $\phi = \sum_{k=1}^K \phi_k B_k$ in Θ ,

we have

Euclidean Inner Product:

$$\langle \theta, \phi \rangle_E = \sum_{k=1}^K \theta_k \phi_k = \theta' \phi \quad \text{where} \quad \theta = (\theta_1, \theta_2, \dots, \theta_K)' \quad \text{and} \quad \phi = (\phi_1, \phi_2, \dots, \phi_K)' ,$$

L_2 Inner Product:

$$\langle \theta, \phi \rangle_2 = \int \theta(t) \phi(t) dt = \theta' \Omega_0 \phi \quad \text{where} \quad \Omega_0(j, k) = \int B_j(t) B_k(t) dt ,$$

Sobolev Inner Product:

$$\langle \theta, \phi \rangle_S = \int \theta(t) \phi(t) dt + \int \ddot{\theta}(t) \ddot{\phi}(t) dt = \theta' [\Omega_0 + \Omega_2] \phi .$$

Corresponding to each inner product there is an approximate representation for the averaging kernel at t . Let these be denoted $A_E(t)$, $A_2(t)$, and $A_S(t)$,

$$E\hat{\theta}_\lambda(t) \approx \langle A_E(t), \theta \rangle_E = \langle A_2(t), \theta \rangle_2 = \langle A_S(t), \theta \rangle_S \quad .$$

$A_E(t)$ is most easily computed; from (3.1), its B-spline coefficients are given by

$$\mathbf{a}_e(t) = X' X [X' X + m \lambda \Omega_2]^{-1} \mathbf{a}(t)$$

where $\mathbf{a}(t) = (B_1(t), B_2(t), \dots, B_K(t))'$ are the B-spline coefficients of the representer of evaluation at t (with respect to the Euclidean inner product). The B-spline coefficients, $\mathbf{a}_2(t)$ and $\mathbf{a}_s(t)$, of $A_2(t)$ and $A_S(t)$ are directly obtained from the $\mathbf{a}_e(t)$.

$$\mathbf{a}_2(t) = \Omega_0^{-1} \mathbf{a}_e(t) \quad ; \quad \mathbf{a}_s(t) = [\Omega_0 + \Omega_2]^{-1} \mathbf{a}_e(t) \quad .$$

A word of caution: The matrices Ω_0 and $[\Omega_0 + \Omega_2]$ are poorly conditioned for K large. As a result a very stable method such as a singular value decomposition, see Dongarra et al.[17], should be used to compute the inverses. From the L_2 representation, we have that the Backus-Gilbert averaging kernel at t is approximately given by

$$A_2(t, s) \approx \sum_{k=1}^K a_{2k}(t) B_k(s) \quad .$$

The Sobolev representation for the averaging kernel is

$$A_s(t, s) \approx \sum_{k=1}^K a_{sk}(t) B_k(s) \quad .$$

Bias Computations

The average and maximum bias can be approximated directly in terms of the Euclidean representation for the averaging kernel in $\{B_k\}_{k=1}^K$. For example, the **maximum bias** in a ball of radius μ in $\Theta = W_2^2[a, b]$ is $\mu^2 b_M^2(t)$ where

$$b_M^2(t) \approx [\mathbf{a}(t) - \mathbf{a}_e(t)]' [\Omega_0 + \Omega_2]^{-1} [\mathbf{a}(t) - \mathbf{a}_e(t)] \quad .$$

4. Application to Non-Linear Inverse Problems

We shall now apply the techniques described in section 2 and 3 to the study of two non-linear ill-posed inverse problems taken from satellite meteorology and reservoir engineering. In both cases the function, θ , of interest is restricted to be one-dimensional and we consider the following MOR procedure: θ is the minimizer in $W_2^2[a, b]$ of

$$\frac{1}{m} \sum_{i=1}^m |z_i - \eta(z_i, \theta)|^2 + \lambda \int_a^b |\ddot{\theta}(t)|^2 dt \quad \lambda > 0 .$$

In order to study performance characteristics, we linearize the functionals $\eta(z_i, \cdot)$ about some value θ_0

$$\eta(z_i, \theta) \approx \eta(z_i, \theta_0) + \nabla_i \eta(z_i, \theta_0)(\theta - \theta_0) ,$$

and consider the properties of a modified MOR procedure: θ is the minimizer of

$$\frac{1}{m} \sum_{i=1}^m |z_i' - \nabla_i \eta(z_i, \theta_0) \theta|^2 + \lambda \int_a^b |\ddot{\theta}(t)|^2 dt \quad \lambda > 0 .$$

where $z_i' = z_i - \eta(z_i, \theta_0) + \nabla_i \eta(z_i, \theta_0) \theta_0$. The averaging kernel calculus can be applied to the modified MOR procedure and the results of this *linearized* analysis are presented below. A rigorous justification for the linearization is not attempted. (This is a very challenging problem and even an asymptotic analysis seems to be quite difficult. The theory presented in Cox and O'Sullivan[14] may provide a starting point for further investigation of this topic). It is assumed that the analysis will give reliable results whenever the degree of non-linearity in the functionals η is low.

To compute averaging kernels, bias and variability for the linearized problem, we use a B-spline basis and follow the development in section 3. The number of basis elements is chosen so that any plots of averaging kernels, and bias and variability characteristics, are visually unchanged by the addition of extra basis elements. The design matrix, X , has the form

$$X_{ik} = \nabla_i \eta(z_i, \theta_0) B_k \quad i=1, 2, \dots, m ,$$

where B_k is the k 'th element of the B-spline basis. Two sets of system libraries were employed:

computations for the satellite meteorology example were carried out on a DEC VAX 11/750 machine using the Bspline and Linpack routines which are part of the publicly available CMLIB; computations for the reservoir engineering example, required repeated numerical solutions of a diffusion equation, and these were carried out on a Boeing Computer Services Cray 1 machine using routines available in the BCSLIB.

4.1. Temperature Retrieval from Satellite Radiance Data

Intensities of radiation measured by modern meteorological satellites provide information about atmospheric characteristics such as temperature and moisture. This new database is becoming an increasingly important tool in the process of *nowcasting*, i.e. specifying the current state of the atmosphere, and *forecasting*, i.e. describing the future states of the atmosphere. Smith *et al* [47] and Smith[46] describe the basic features of these measurement systems. A typical climatological temperature profile is given in Figure 4.1. The vertical axis is decreasing in pressure while the horizontal axis gives temperature. It is standard practice for meteorologists to plot things in this way because moving up the vertical axis then corresponds to going higher up in the atmosphere. The plot shows a temperature inversion near 20 mb, the temperature is generally increasing as one moves away from this point. Inversions are a characteristic feature of atmospheric temperature profiles. The location of this upper atmosphere temperature inversion is known as the tropopause height.

Figure 4.1 ABOUT HERE

The processing of radiance data to get temperature estimates involves the solution of an interesting inverse problem. The radiative transfer equations, Liou[31], are used to model how the intensity of radiation, z , at frequency ν , depends on the temperature profile T , temperature as a function of pressure, in the column beneath the satellite;

$$z_i = R_{\nu_i}(T) + \epsilon_i, \quad i=1, \dots, m$$

where

$$R_{\nu}(T) = B_{\nu}(T(x_0))\tau_{\nu}(x_0) - \int_{x_0}^{x_s} B_{\nu}(T(x))\dot{\tau}_{\nu}(x)dx$$

and x is some monotone transformation of pressure p ; x_s corresponds to the surface and x_0 corresponds to the top of the atmosphere. Meteorologists usually work in *kappa* -units, i.e. $x(p) = p^{5/8}$, because atmospheric variations are believed to be slowly varying in this scale. $\tau_{\nu}(x)$ is the transmittance of the atmosphere above x at wavenumber ν , and B_{ν} is Planck's function given by:

$$B_{\nu}(T) = c_1 \nu^3 / [\exp(c_2 \nu / T) - 1]$$

where

$$c_1 = 1.1906 \times 10^{-5} \text{ erg-cm}^2\text{-sec}^{-1}$$

and

$$c_2 = 1.43868 \text{ cm-deg}(K).$$

The measurement errors, ϵ_i , are roughly mean zero and uncorrelated, however, different channels have different noise levels so by dividing through by relative noise levels in channels we obtain a generalized non-linear regression model as in section 1. The TIROS-N system, see Smith *et al* [47], has fifteen channels ($m=15$). Linearizing the $R_{\nu_i}(T)$ about the climatological profile, T_0 , given in Figure 4.1, we get

$$\nabla_T R_{\nu_i}(T_0)T = k_s(\nu_i)T(x_s) + \int_{x_0}^{x_s} k(\nu_i, x)T(x)dx,$$

where $k_s(\nu) = \dot{B}_{\nu}(T_0(x_s))\tau_{\nu}(x_s)$ and $k(\nu, x) = -\dot{B}_{\nu}(T_0(x))\dot{\tau}_{\nu}(x)$. Since the linearized functionals have an explicit form, the linearized design matrix is very easy to compute by numerical integration.

$$X_{ik} = \nabla_T R_{\nu_i}(T_0)B_k = k_s(\nu_i)B_k(x_s) + \int_{x_0}^{x_s} k(\nu_i, x)B_k(x)dx,$$

where B_k is the k 'th element of the B-spline basis.

An averaging kernel at 700 millibars for the MOR inversion is given in Figure 4.2. The corresponding bias and variability characteristics are given in Figure 4.3. Again these plots correspond to a particular value for the regularization parameter λ . Larger values for the smoothing result in broader averaging kernels, more bias and less variance.

Figure 4.2 ABOUT HERE

Figure 4.3 ABOUT HERE

Notice that the averaging kernel has sharp behavior near the surface. This is attributable to the microwave channels. The data obtained from these are nearly direct measurements of the surface *skin* temperature, $T(z_s)$. As a result the L_2 representers of the functionals corresponding to these channels are very spiked at the surface and since the averaging kernel is a linear combination of these representers, the behavior at the surface is to be expected. The effect of the microwave channels on the bias is also quite dramatic. The variability profile indicates regions near 600mb and 200 mb where sampling density might be improved. However, there are physical constraints on selection of spectral wavelengths which make it difficult to get good coverage throughout the atmosphere, see pp 250 ff of Liou[31]. The operating characteristics given in Figure 4.3 relate to maximum bias. In a meteorological setting, where there is a huge database of prior information on atmospheric variation, an average bias measure with respect to a climatological prior would be more appropriate. In spite of the fact that there is a dependence on the initial climatology, T_s , the retrieval characteristics predicted by the averaging kernel calculus is largely in agreement with those found by Monte-Carlo simulation in O'Sullivan and Wahba[39].

4.2. The History Matching Problem of Reservoir Engineering

The dynamic flow of fluid through a porous medium is usually modeled by a diffusion equation

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} - \frac{\partial}{\partial \mathbf{x}} \left\{ a(\mathbf{x}) u(\mathbf{x}, t) \right\} = q(\mathbf{x}, t) \quad \mathbf{x} \text{ in } \Omega, t \text{ in } [0, T]$$

subject to prescribed initial and boundary conditions. Here u is pressure, q accounts for the withdrawal or injection of fluid into the region Ω , and a is the transmissivity or conductance which determines the ease with which fluid flows through the medium. The initial condition is $u(\mathbf{x}, 0) = u_0(\mathbf{x})$ and a typical boundary condition is no fluid flow across the boundary of the region, i.e. $\frac{\partial u}{\partial \omega} = 0$, where ω represents the direction normal to the boundary. The history matching problem arises as one tries to use scattered well data on $u(\mathbf{x}_i, t_j)$ and $q(\mathbf{x}_j, t_j)$, $i=1, 2, \dots, m$, $j=1, 2, \dots, l$, to infer the diffusion parameter, a , see Cooley[11, 12], Kravaris and Seinfeld[27] and Neuman and Yakowitz[37]. This problem is an example of a broad class of inverse problems which arise in connection with partial differential equations. These problems have attracted an amount of pure mathematical interest. See Anger[4], Lions[30], McLaughlin[34], and especially Payne[41].

For a simplified version of the history matching problem let $\Omega = [0, 1]$ and $T=1$. Assume there is no injection or withdrawal of fluid and that there is no flow across the boundary. Suppose that we have 10 measurement sites at $x_i = (5i-3)/49$, $i=1, 2, \dots, 10$, and readings on u are made for $t_j = 0.0(0.007)0.5$. These data are modeled by a non-linear regression

$$z_{i,j} = u(x_i, t_j; a) + \epsilon_i$$

where the errors are mean zero with constant variance.

The dependence of $u(x_i, t_j; a)$ on a is again non-linear. By linearizing $u(x_i, t_j; a)$ about a plausible transmissivity profile, a_0 , such as the one given in Figure 4.4, we can compute averaging kernels, etc. The true pressure history plays a significant role in determining the information recovered about transmissivity. Roughly speaking, gradients in the pressure history generate

information about the transmissivity. A pressure history corresponding to the transmissivity profile in Figure 4.4 is given in Figure 4.5. This pressure history is driven by an initial pressure distribution which ranges from 10 at $x=0$. to 100 at $x=1$.

Figure 4.4 ABOUT HERE

Figure 4.5 ABOUT HERE

Computation of the Linearized Design Matrix

Unlike the temperature retrieval problem, we do not have an explicit analytical representation for the observed functionals, and this makes the computation of the design matrix a bit more complicated.

$$X_{i,k} = \nabla_{a_k} u(x_i, t_i, a_0) B_k = \frac{\partial \hat{u}}{\partial a_k}(x_i, t_i; a_0)$$

where the gradient is taken in the direction of functions of the form $a(x) = \sum_{k=1}^K a_k B_k(x)$ and \hat{u}

is the solution to

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x} \left\{ a(x) \frac{\partial u}{\partial x} \right\} = q(x, t) \quad \text{with} \quad (x, t) \text{ in } [0, 1] \times [0, 5] ,$$

$$\text{subject to} \quad \begin{cases} u(x, 0) = u_0(x) \\ \frac{\partial u(x, t)}{\partial x} = 0 \text{ for } x = 0, 1 \end{cases}$$

Let $D(a) : U \rightarrow Q \times U_0 \times B_0 \times B_1$ denote the mapping

$$D(a)u = \left[\frac{\partial u}{\partial t} - \frac{\partial}{\partial x} \left\{ a(x) \frac{\partial u}{\partial x} \right\}, u(\cdot, 0), \frac{\partial u(0, \cdot)}{\partial x}, \frac{\partial u(1, \cdot)}{\partial x} \right] .$$

Under regularity, by applying the implicit function theorem, the inverse of $D(a)$, denoted $G(a)$, exists and is differentiable in a neighborhood of a_0 .

$$G(a) [q, u, 0, 0] = \hat{u}(\cdot; a) .$$

The relation $G(a)D(a) = I$ may be used to obtain $\frac{\partial \hat{u}}{\partial a_k}(x, t; a_0)$. Differentiating, we have

$$\frac{\partial}{\partial a_k} \{ G(a)D(a) \} = \frac{\partial G(a)}{\partial a_k} D(a) + G(a) \frac{\partial D(a)}{\partial a_k} = 0 .$$

Thus

$$\frac{\partial G(a)}{\partial a_k} = - G(a) \frac{\partial D(a)}{\partial a_k} G(a) ,$$

which implies

$$\frac{\partial \hat{u}}{\partial a_k} = - G(a) \left\{ \frac{\partial D(a)}{\partial a_k} \hat{u} \right\} .$$

But $\frac{\partial D(a)}{\partial a_k} u = \left[-\frac{\partial}{\partial x} \{ B_k(x) \frac{\partial u}{\partial x} \}, 0, 0, 0 \right]$, and the expression simplifies to

$$\frac{\partial \hat{u}(x, t; a_0)}{\partial a_k} = G(a_0) \left[\frac{\partial}{\partial x} \{ B_k(x) \frac{\partial u}{\partial x} \}, 0, 0, 0 \right] .$$

Hence $X_{i,k}$ can be found by solving the original diffusion equation with the forcing term, q , replaced by $\frac{\partial}{\partial x} \{ B_k(x) \frac{\partial u}{\partial x} \}$ and the initial pressure distribution, u_0 , replaced by the constant 0.

By this method, the computation of the entire linearized design matrix requires K separate numerical solutions of the diffusion equation (K being the number of basis elements). Although this is a generally applicable technique, it is rather inelegant. Since the solution corresponding to B_k is likely to be "near" the solution corresponding to B_{k+1} , it may be possible to improve computational efficiency by using some form of relaxation. This is currently being investigated. For the time invariant problem, Neuman and Yakowitz[37] use properties of a particular finite difference scheme to develop a fast method for computing the analogue of the design matrix. A further approach to this problem, relying on an optimal control formulation, is employed by Kravaris and Seinfeld[27].

Linearized Averaging Kernels and Retrieval Characteristics

Figures 4.6 and 4.7 give linearized averaging kernels and retrieval characteristics.

Figure 4.6 ABOUT HERE

Figure 4.7 ABOUT HERE

One would suspect that information about α should depend critically on the true pressure history. If the initial pressure were constant, then, since there would be no pressure gradients, there would be no lateral flow. However, the functional parameter is a transmissivity and information about transmissivity can only be generated by lateral flow. The averaging kernel and the bias and variability characteristics show that for the given pressure history, the greatest detail on transmissivity is recovered near the middle of the x -range. Changing the pressure history results in different retrieval properties - the bias and variability calculus seems to make very good physical sense.

5. Empirical Selection of Smoothing Parameters

Nearly all inversion algorithms, and in particular the MOR, have explicit or implicit smoothing/tuning parameters, corresponding to the signal-to-noise ratio which, in a practical setting, have to be selected by the user. Two of the more popular techniques to emerge from the statistical literature on this problem are the methods of cross-validation and unbiased risk estimation. Typically, these techniques try to find that value of the smoothing parameter which minimizes the predictive mean square error (PMSE).

$$PMSE(\lambda) = \frac{1}{m} \sum_{i=1}^m [\eta(z_i, \theta) - \eta(z_i, \hat{\theta}_\lambda)]^2$$

where θ and $\hat{\theta}_\lambda$ are the true and estimated functions. (Throughout this section we will use λ to denote the smoothing/tuning parameters of the inversion algorithm). The predictive mean square (PMSE) is a convenient criterion but not necessarily a criterion of real intrinsic interest. Inverse problems are focused on a particular function, so it seems more appropriate to phrase a loss directly in terms of that function. Thus it might be of interest to consider a loss of the form

$$L(\lambda) = \frac{1}{T} \sum_{i=1}^T [\theta(t_i) - \hat{\theta}_\lambda(t_i)]^2$$

where the t_i 's are some points of direct interest to the investigator. For certain loss functions, which are *estimable* in the sense described below, it is possible to develop refined versions of the cross-validation and unbiased risk assessments. It should be pointed out that the PMSE has a certain robustness property; it is often the case that the best choice of the smoothing parameter, from the point of view of PMSE loss, is very nearly optimal from the point of view of other loss functions also, see Cox [13], Lukas[32], Ragozin[42] and Wahba[57]. Thus the practical need for refined procedures will only arise in situations where the PMSE and the loss function of interest have very different minimizers.

We assume throughout that we are dealing with linear inversion algorithms and that the functionals $\eta(z, \theta)$ are themselves linear in θ . For some extensions to non-linear inversion algorithms, see O'Sullivan and Wahba[39], O'Sullivan et al.[40], Villalobos and Wahba[56] and

Wahba[59]. A brief review of cross-validation and unbiased risk in the standard predictive mean square error context is given next. For a more detailed discussion of these methods see the recent review of Titterton[54].

5.1. Empirical Assessment of PMSE

The *Hat-matrix*, $H(\lambda)$, is defined to be the matrix that maps data \mathbf{s} into predictions $\hat{\mathbf{s}}$,

$$\hat{\mathbf{s}} = H(\lambda)\mathbf{s}$$

For a linear inversion algorithm the Hat-matrix is obtained from the impulse response functions, $s_i^{(\lambda)}$, via

$$H_{ij}(\lambda) = \eta(x_i, s_j^{(\lambda)})$$

Here the dependence of the inversion on the smoothing/tuning parameters λ is highlighted- the impulse response functions are functions of λ . The trace of the Hat-matrix occurs in both the cross-validation and unbiased risk assessments of the PMSE. We begin with the unbiased risk method, which is easier to describe.

Unbiased Risk

This procedure assumes that one has a reliable estimate of the noise level σ . Given σ , the procedure uses the residual sum of squares (RSS) to construct an unbiased estimate of the PMSE risk. The basic steps are as follows:

$$\text{PMSE}(\lambda) = \frac{1}{m} \sum_{i=1}^m [\eta(x_i, \theta) - \hat{z}_i]^2 \equiv \frac{1}{m} \|\eta(\theta) - \eta(\hat{\theta}_\lambda)\|_m^2$$

i.e. $\eta(\cdot) = (\eta(x_1, \cdot), \eta(x_2, \cdot), \dots, \eta(x_m, \cdot))'$. By linearity, the expected value of the predictive mean square error is

$$\begin{aligned} m E\{\text{PMSE}(\lambda)\} &= \|[I - H(\lambda)]\eta(\theta)\|_m^2 + \sigma^2 \text{tr}[H(\lambda)' H(\lambda)] \\ &= \text{BIAS}^2(\lambda) + \sigma^2 \text{tr}[H(\lambda)' H(\lambda)] \end{aligned}$$

Meanwhile,

$$RSS(\lambda) = \sum_{i=1}^n |z_i - \hat{z}_i|^2 = \| [I - H(\lambda)] \mathbf{z} \|^2$$

So the expected value is

$$\begin{aligned} E\{RSS(\lambda)\} &= \| [I - H(\lambda)] \eta(\theta) \|^2 + \sigma^2 \text{tr} \{ [I - H(\lambda)]' [I - H(\lambda)] \} \\ &= BIAS^2(\lambda) + \sigma^2 \text{tr} \{ [I - H(\lambda)]' [I - H(\lambda)] \} \end{aligned}$$

Combining these formulae we have that

$$PMSE(\lambda) = \frac{1}{n} RSS(\lambda) - \sigma^2 + 2\sigma^2 \frac{\text{tr} [H(\lambda)]}{n}$$

is an unbiased estimate of the predictive mean square error. In the standard regression context $H = X(X'X)^{-1}X'$, and $PMSE$ reduces to Mallows' C_p statistic, see [33].

Cross-Validation

In cross-validation one considers, \hat{z}_{-i} , which is defined to be the prediction of $\eta(z_i, \theta)$ from an estimator constructed from data with the i 'th data point, z_i , omitted. The idea being that if the prediction rule is really good, that is, λ well chosen, then \hat{z}_{-i} should be reasonably close to z_i on average. Ordinary cross-validation or Allen's predictive sum of squares (PRESS), see Allen [3], is defined to be

$$V_o(\lambda) = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_{-i}|^2$$

For the MOR estimators in section 1.1, where $H(\lambda) = X[X'X + m\lambda W]^{-1}X'$, a rank one update formula, can be used to show that

$$\hat{z}_{-i} = z_i - \frac{(z_i - \hat{z}_i)}{(1 - h_{ii}(\lambda))}$$

where $h_{ii}(\lambda)$ is the i 'th diagonal element of the Hat-matrix, also known as the i 'th leverage value. It follows that the ordinary cross-validation assessment is a weighted residual sum of squares

$$V_o(\lambda) = \frac{1}{n} \sum_{i=1}^n \left[\frac{z_i - \hat{z}_i}{1 - h_{ii}(\lambda)} \right]^2$$

If instead of dividing residuals by $1-h_{ii}(\lambda)$ we divide by the mean value, $1-\bar{h}_{ii}(\lambda)$, then we obtain the Generalized Cross Validation (GCV) of Craven and Wahba[15]. The GCV score is usually written as

$$V(\lambda) = \frac{\frac{1}{m} \sum_{i=1}^m |z_i - \hat{z}_i|^2}{\left[1 - \frac{1}{m} \text{tr} H(\lambda)\right]^2} .$$

We point out that, since

$$\frac{V(\lambda)}{m} = \frac{RSS(\lambda)}{\left[1 - \frac{1}{m} \text{tr} H(\lambda)\right]^2} ,$$

in a regression context, where $H = X(X'X)^{-1}X'$, the GCV score is proportional to the residual mean square divided by the degrees of freedom for error. So in this context, the GCV score reduces to a model selection statistic proposed by Anscombe[5] and simplified by Tukey, see pp. 386 ff Mosteller and Tukey[36].

A great deal is known about the asymptotic behavior of the above empirical assessments methods. The typical result says that the minimizer of the empirical assessment tends to minimize the PMSE, in large samples; Monte carlo simulation results show that a similar property tends to hold in finite samples, see for example Craven and Wahba[15], Golub Heath and Wahba[20], Nychka et al[38], Rice[43], and Speckman[49].

5.2. Empirical Assessment of Estimable Losses

Borrowing from the terminology of the standard linear model we can characterize an estimable functional as follows

Definition: A continuous linear functional $\langle \xi, \cdot \rangle$ is estimable if there exists c in R^m such that $E c' z = \langle \xi, \theta \rangle$ for all θ in Θ .

Clearly, since $\eta(z, \cdot)$ are continuous linear functionals, $\langle \xi, \cdot \rangle$ is estimable iff $\exists c$ in R^m such that

$$\xi = \sum_{i=1}^m c_i \xi_{z_i}$$

where ξ_{z_i} are representers of $\eta(z_i, \cdot)$ in Θ . With this we say that a loss function is *estimable* if it defined in terms of *estimable* functionals.

To illustrate how to empirically assess estimable losses we consider a particular estimable loss of the form

$$L(\lambda) = \sum_{j=1}^T \langle \xi_j, \theta - \hat{\theta}_\lambda \rangle^2$$

where ξ_j are all estimable. Let \mathbf{e}_j be such that

$$\xi_j = \sum_{i=1}^T c_{ji} \xi_{z_i} \quad j=1, \dots, T.$$

A simple modification of the cross-validation and unbiased risks techniques can be formulated to directly assess the loss $L(\lambda)$.

Unbiased Risk Assessment of $L(\lambda)$

By a development similar to that used in Rice[44], the expected value of $L(\lambda)$ as

$$\begin{aligned} E[L(\lambda)] &= \sum_{j=1}^T \langle \xi_j, \theta - E\hat{\theta}_\lambda \rangle^2 + \sigma^2 \sum_{j=1}^T \mathbf{e}_j' [I - H(\lambda)]' [I - H(\lambda)] \mathbf{e}_j \\ &= BIAS^2(\lambda) + \sigma^2 \sum_{j=1}^T \mathbf{e}_j' [I - H(\lambda)]' [I - H(\lambda)] \mathbf{e}_j. \end{aligned}$$

The sum of squares

$$SS = \sum_{j=1}^T [\mathbf{e}_j' \mathbf{z} - \mathbf{e}_j' \hat{\mathbf{z}}]^2$$

has expected value

$$E[SS] = BIAS^2(\lambda) + \sigma^2 \sum_{j=1}^T \mathbf{e}_j' H(\lambda)' H(\lambda) \mathbf{e}_j.$$

Thus

$$\hat{L}(\lambda) = SS - \sigma^2 \sum_{j=1}^T \mathbf{e}_j' \mathbf{e}_j + 2\sigma^2 \sum_{j=1}^T \mathbf{e}_j' H(\lambda)' \mathbf{e}_j,$$

is an unbiased estimate of $L(\lambda)$. Again note that a reliable estimate of σ is necessary in order to be able to use this assessment.

Cross-Validation Assessment of $L(\lambda)$

The cross-validation procedure is slightly more complicated to derive. Instead of leaving out one data point, we now omit the e_j 'th component of the data and use the remaining data to develop a prediction, \hat{c}_{s_j} , of $\langle \xi_j, \theta \rangle$. The cross-validation assessment then compares e_j 's to \hat{c}_{s_j} , i.e.

$$V_s(\lambda) = \sum_{j=1}^T |e_j' s - \hat{c}_{s_j}|^2$$

In the context of MOR estimators the situation becomes clearer. To make the notation less complicated, suppose the e_j are normalized so that $e_j' e_j = 1$ for $j=1, 2, \dots, T$. Let

$$P_j = I - e_j e_j'$$

P_j is a projection onto the space orthogonal to e_j . The estimator obtained by removing the e_j 'th component minimizes

$$\frac{1}{m} [P_j (s - \eta(\theta))]' P_j^- [P_j (s - \eta(\theta))] + \lambda \langle \theta, W \theta \rangle$$

where P_j^- is the generalized inverse of P_j . Since $P_j P_j^- P_j = P_j$, it follows that

$$\hat{c}_{s_j} = e_j' [X' P_j X + m \lambda W]^{-1} X' P_j s$$

Again, using a rank one update formula and some algebra we have that

$$\hat{c}_{s_j} = e_j' s - \frac{e_j' s - e_j' \hat{s}}{[1 - e_j' H(\lambda) e_j]}$$

where $H(\lambda) = X [X' X + m \lambda W]^{-1} X'$. Thus the ordinary cross-validation score is

$$V_s(\lambda) = \sum_{j=1}^T \left[\frac{e_j' s - e_j' \hat{s}}{1 - e_j' H(\lambda) e_j} \right]^2$$

and the GCV extension is

$$V(\lambda) = \frac{\frac{1}{T} \sum_{j=1}^T |\mathbf{e}_j' \mathbf{s} - \mathbf{e}_j' \hat{\mathbf{z}}|^2}{|\mathbf{1} - \overline{\mathbf{e}_j' H(\lambda) \mathbf{e}_j}|^2} .$$

where

$$\overline{\mathbf{e}_j' H(\lambda) \mathbf{e}_j} = \frac{1}{T} \sum_{j=1}^T \mathbf{e}_j' H(\lambda) \mathbf{e}_j$$

The advantage of the cross-validation assessment over the unbiased risk assessment is that the cross-validation method does not require knowledge of σ .

Acknowledgements

The motivation for this article arose out of a series of presentations I gave in a seminar on "Inverse Problems and Statistical Signal Models" jointly organized with David Donoho at Berkeley in the Spring of 1985. I would like to thank the participants in the seminar and also D. L. Banks, B. A. Bolt, D. R. Brillinger, M. H. DeGroot, J. A. Rice, G. Wahba, and two referees for many useful comments which have lead to substantial improvements in the paper.

- [1] Adams, R. (1975). *Sobolev Spaces*, Academic Press. New York.
- [2] Aki, K. and Richards, G. (1980). *Quantitative Seismology: Theory and Methods*, W. H. Freeman. San Francisco.
- [3] Allen, D. M. (1974). "The relationship between variable selection and data augmentation and a method of prediction," *Technometrics* 16, pp. 125-127.
- [4] Anger, G. (1979). *Inverse and Improperly Posed Problems in Differential Equations: Proceedings of a Conference on Mathematical and Numerical Methods*, Akademie-Verlag. Berlin.
- [5] Anscombe, F. J. (1967). "Topics in the investigation of linear relations fitted by least squares," *J. Roy. Statist. Soc. B* 29, pp. 1-52.
- [6] Backus, G. and Gilbert, F. (1968). "The resolving power of gross earth data," *Geophysics Journal of the Royal Astronomical Society* 266, pp. 169-205.
- [7] Backus, G. and Gilbert, F. (1970). "Uniqueness in the inversion of inaccurate gross earth data," *Philos. Trans. Royal Soc. Ser. A*. 266, pp. 123-192.
- [8] Bolt, B. A. (1980). "What can inverse problems do for applied mathematics and the sciences," *Search* 11, 6,
- [9] Box, G. E. P. and Draper, N. R. (1959). "A basis for the selection of a response surface design," *Amer. Statist. Assoc.* 54, pp. 622-653.
- [10] Budinger, T. F. (1980). "Physical attributes of single-photon tomography," *J. Nucl. Med.* 21, 6,
- [11] Cooley, R. L. (1982). "Incorporation of prior information on parameters into non-linear regression groundwater flow models, 1 Theory," *Water Resour. Res.* 18, pp.

965-976.

- [12] Cooley, R. L. (1983). "Incorporation of prior information on parameters into non-linear regression groundwater flow models, 2 Applications," *Water Resour. Res.* **19**, pp. 662-676.
- [13] Cox, D. D. (1983). "Approximation of the method of regularization estimators." Tech. Rep. No. 723, Statistics Dept., University of Wisconsin-Madison .
- [14] Cox, D. D. and O'Sullivan, F., (1985). Analysis of penalized likelihood type estimators with application to generalized smoothing in Sobolev Spaces, Manuscript.
- [15] Craven, P. and Wahba, G. (1979). "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation," *Numer. Math.* **31**, pp. 377-403.
- [16] DeBoor, C. (1978). *A Practical Guide to B-Splines*, Springer-Verlag, New York.
- [17] Dongarra, J. J., Bunch, J. R., Moler, C. B., and Stewart, G. W. (1979). *Linspack User's Guide*, SIAM, Philadelphia.
- [18] Foster, M. R. (1961). "An application of the Wiener-Kolmogorov smoothing theory to matrix inversion," *J. SIAM* **9**, pp. 387-392.
- [19] Franzone, P. Colli, Taccardi, B., and Viganotti, C. (1977). "An approach to inverse calculation of epi-cardial potentials from body surface maps," *Adv. Cardiol.* **21**, pp. 167-170.
- [20] Golub, G., Heath, M., and Wahba, G. (1979). "Generalized cross validation as a method for choosing a good ridge parameter," *Technometrics* **31**, pp. 315-224.

- [21] Good, I. J. and Gaskins, R. A. (1971). "Non-parametric roughness penalties for probability densities," *Biometrika* 58, pp. 255-277.
- [22] Grenander, U. (1981). *Abstract Inference*, J. Wiley.
- [23] Grunbaum, F. A. (1975). "Remark on the phase problem in Crystallography," *Proc. Nat. Acad. Sci. USA* 72, 5, pp. 1699-1701.
- [24] Jaynes, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics*, Synthese Library.
- [25] Jeffreys, H. (1976). *The Earth*, Cambridge University Press.
- [26] Kimeldorf, G. S. and Wahba, G. (1971). "Some results on Tchebycheffian Spline Functions," *Mathematical Analysis and Applications* 33, pp. 82-95.
- [27] Kravaris, C. and Seinfeld, J. H. (1985). "Identification of parameters in distributed parameter systems by regularization," *SIAM J. Control and Optimization* 23, 2, pp. 217-241.
- [28] Lattes, R. and Lions, J. L. (1969). *The Method of Quasi-reversibility, Applications to Partial Differential Equations*, American Elsevier, New York.
- [29] Li, K. C. (1982). "Minimaxity of the method of regularization on stochastic processes," *Ann. Statist.* 10, pp. 937-942.
- [30] Lions, J. L. (1971). *Optimal Control of Systems Governed by Partial Differential Equations*, Springer-Verlag, Berlin.
- [31] Liou, K. (1979). *Introduction to Atmospheric Radiation*, Academic Press, London.

- [32] Lukas, M. (1981). "Regularization of linear operator equations.." Unpublished Ph.D. thesis, Australian National University .
- [33] Mallows, C. L. (1973). "Some comments on C_p ," *Technometrics* 15, pp. 661-675.
- [34] McLaughlin, D. W. (1983). *Inverse Problems: Proceedings of a Symposium in Applied Mathematics*, American Mathematical Society.
- [35] Mendelsohn, J. and Rice, J. (1983). "Deconvolution of Micro-fluometric histograms with B-splines," *J. Amer. Statist. Assoc.* 77, pp. 748-753.
- [36] Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, MA.
- [37] Neuman, S. P. and Yakowitz, S. (1979). "A statistical approach to the inverse problem of aquifer hydrology 1. Theory," *Water Resour. Res.* 15, pp. 845-860.
- [38] Nychka, D., Wahba, G., Goldfarb, S., and Pugh, T. (1984). "Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two dimensional cross sections," *J. Amer. Statist. Assoc.* 79, 388,
- [39] O'Sullivan, F. and Wahba, G. (1985). "A cross validated Bayesian retrieval algorithm for non-linear remote sensing experiments," *Journal of Computat. Physics* 59, 3, pp. 441-455.
- [40] O'Sullivan, F., Yandell, B., and Raynor, W. J. (1986 (in press)). "Automatic smoothing of regression functions in generalized linear models," *Amer. Statist. Assoc.*,
- [41] Payne, L. E. (1975). "Improperly Posed Problems in Partial Differential Equations," *In Regional Conference Series in Applied Mathematics*, SIAM .

- [42] Ragozin, D. (1981). "Error bounds for derivative estimates based on spline smoothing of exact or noisy data." Tech. Rep. No. GN-50, Statistics Dept., University of Washington-Seattle .
- [43] Rice, J. A. (1984). "Bandwidth choice for nonparametric regression," *Ann. Statist.* 12, pp. 1215-1230.
- [44] Rice, J. A. (1984). "Bandwidth choice for differentiation." Manuscript .
- [45] Rudin, W. (1976). *Principles of Mathematical Analysis*, McGraw-Hill, New York.
- [46] Smith, W. (1983). "The retrieval of atmospheric profiles from VAS geostationary radiance observations," *J. Atmospheric Sciences* 40, pp. 2025-2035.
- [47] Smith, W. L., Woolf, H. M., Hayden, C. M., Wark, D. Q., and McMillin, L. M. (October 1979). "The TIROS-N operational vertical sounder," *Bull. Amer. Meteor. Soc.* 50, 10, pp. 1177-1187.
- [48] Speckman, P. (1979). *Minimax estimates of linear functionals in Hilbert space*, Department of Mathematics, University of Oregon.
- [49] Speckman, P. (1982). "Efficient nonparametric regression with cross-validated smoothing splines." Tech. Rep. No. 45, Statistics Dept., University of Missouri-Columbia .
- [50] Steinberg, D. M. and Hunter, W. G. (1984). "Experimental design: review and comment (with discussion)," *Technometrics* 26, 2, pp. 71-130.
- [51] Strand, O. N. and Westwater, E. R. (1968). "Minimum-rms estimation of the numerical solution of a Fredholm integral equation of the first kind," *SIAM J. Numer. Anal.* 5, pp. 287-295.

- [52] Tikhonov, A. (1963). "Solution of incorrectly formulated problems and the regularization method," *Soviet Math Dokl* 5, pp. 1035-1038.
- [53] Tikhonov, A. and Arsenin, V. (1977). *Solutions of Ill-Posed Problems*, Wiley, New York.
- [54] Titterton, D. M. (1985). "Common structure of smoothing techniques in statistics," *International Statistical Review* 53, pp. 141-170.
- [55] Vardi, Y., Shepp, L. A., and Kaufman, L. (1985). "A statistical model for positron emission tomography," *J. Amer Statist. Assoc.* 80, 389,
- [56] Villalobos, M. and Wahba, G. (1983). *Commun. Statist. A* 12, pp. 1449-1480.
- [57] Wahba, G. (1979). "Smoothing and ill-posed problems," pp. 183-194. In *Solution Methods for Integral Equations with Applications*, ed. M. Goldberg, Plenum Press .
- [58] Wahba, G. (1983). "Design criteria and eigensequence plots for satellite computed tomography." Tech. Rep. No. 732, Statistics Dept., University of Wisconsin-Madison .
- [59] Wahba, G. (1984). "Cross-validated spline methods for the estimation of multivariate functions from data on functionals," pp. 205-233. In *Statistics: An Appraisal*, ed. H. A. David and H. T. David, The Iowa State University Press .
- [60] Weinreb, M. P. and Crosby, D. S. (1972). "Optimization of spectral intervals for remote sensing of atmospheric temperature profiles," *Remote Sensing of the Environment* 2, pp. 193-201.

Figure Legends

Figure 1.1 : Reconstruction of the Tumor Size Distribution from data on cross-sectional slices

Figure 2.1 : Sample Averaging Kernel for a MOR procedure applied to the Tumor Problem

Figure 2.2 : Maximum Bias and Standard Error for a MOR procedure applied to the Tumor Problem. The ripples are due to the finite sampling ($m = 50$).

Figure 4.1 : Typical climatological profile, T_p . Vertical axis is pressure in kappa scale. Horizontal axis is in degrees Kelvin. Note the temperature inversion high up in the atmosphere.

Figure 4.2 : Averaging Kernel at 700mb for the Temperature Retrieval Problem. Vertical axis is pressure in kappa scale. Horizontal axis is in degrees Kelvin. The sharp behavior near the surface is attributable to the microwave channels.

Figure 4.3 : Inversion Characteristics of a MOR procedure applied to the Temperature Retrieval Problem

Figure 4.4 : A Transmissivity Distribution, ϵ_0 , for the History Matching Problem

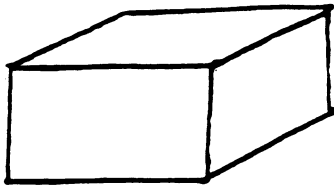
Figure 4.5 : True Pressure History.

Figure 4.6 : Sample Averaging Kernels for a MOR procedure applied to the History Matching Problem.

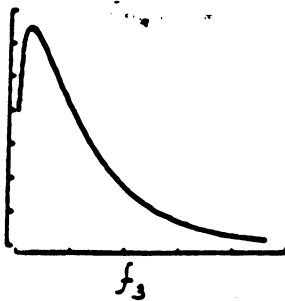
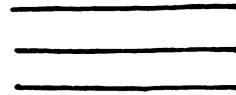
Figure 4.7 : Inversion Characteristics of a MOR procedure applied to the History Matching Problem. Retrieval properties are best near the middle of the x range.

Figure 1.1

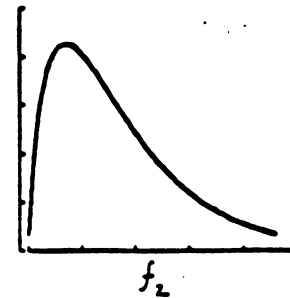
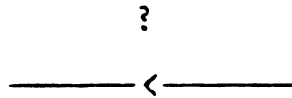
Section of Tissue



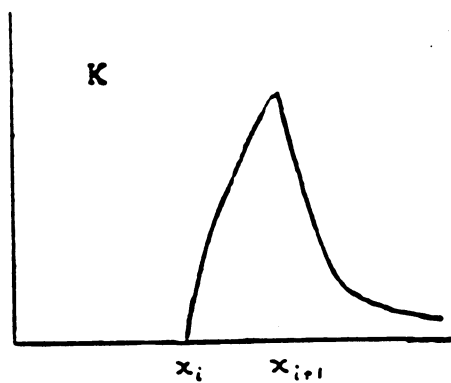
Cross Sectional Slices



distribution of
tumor radii



distribution of
tumor radii
(cdf F_2)



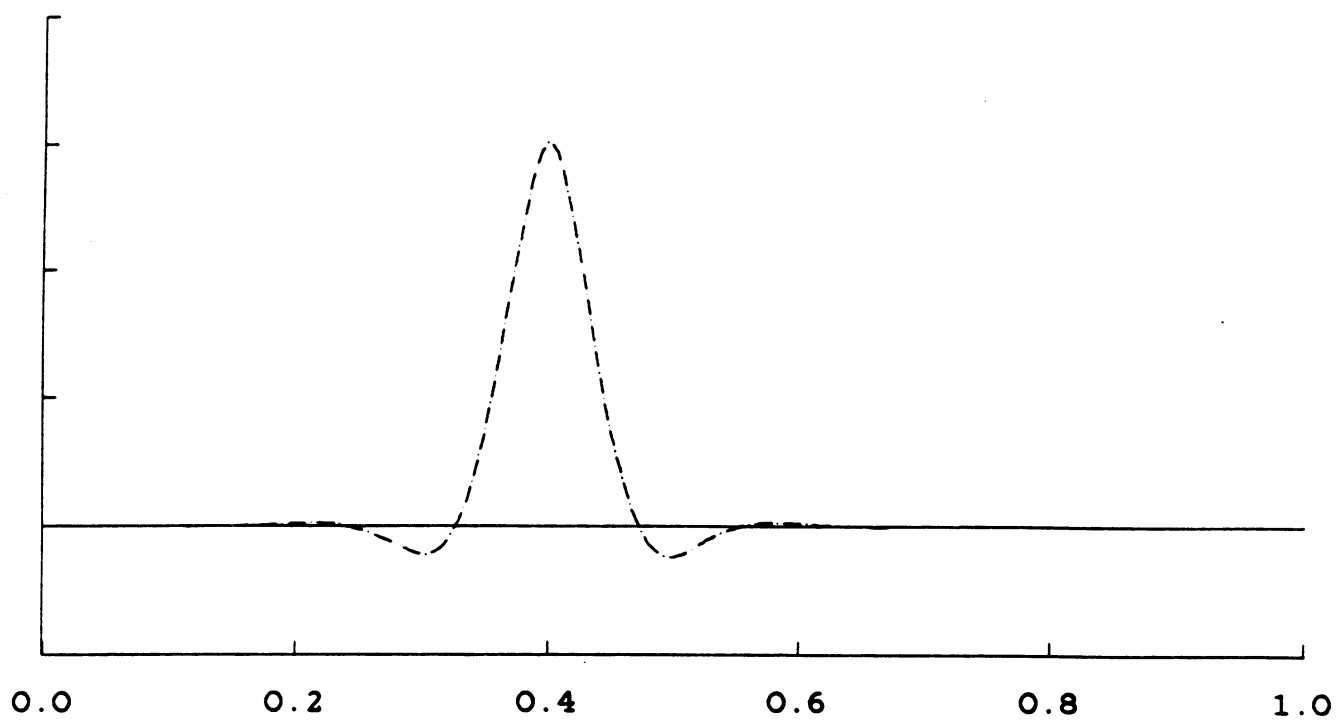
DATA:

$$z_i = \hat{F}_2(x_{i+1}) - \hat{F}_2(x_i) \quad , \quad x_i < x_{i+1}$$

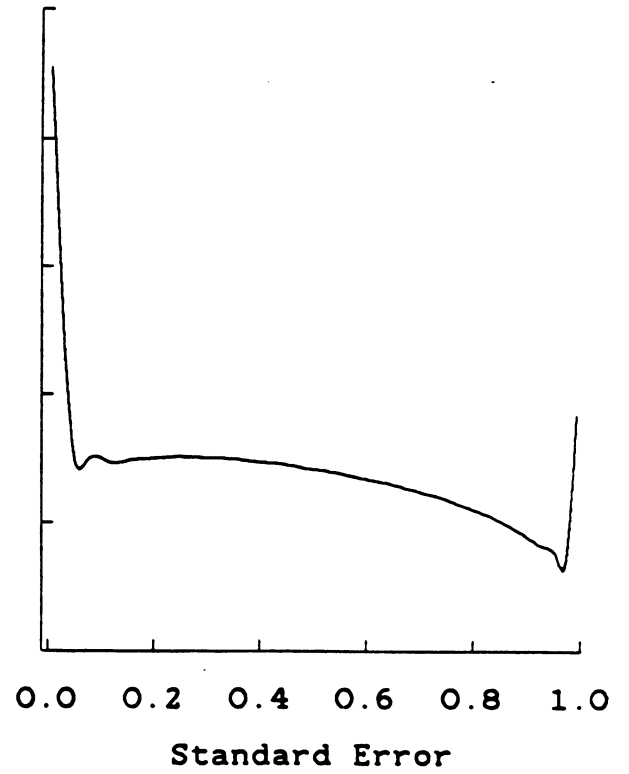
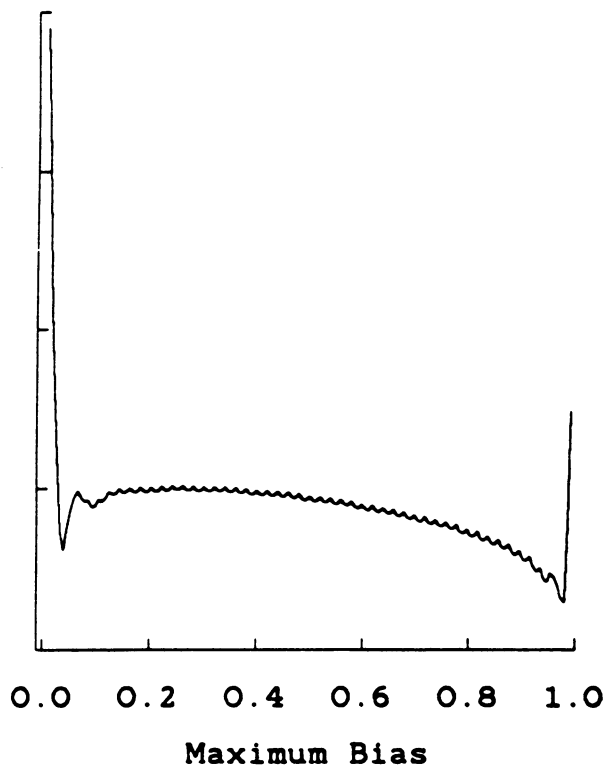
MODEL:

$$z_i \approx \int_{x_i}^{x_{i+1}} K(x_i, r) f_3(r) dr + \epsilon_i$$

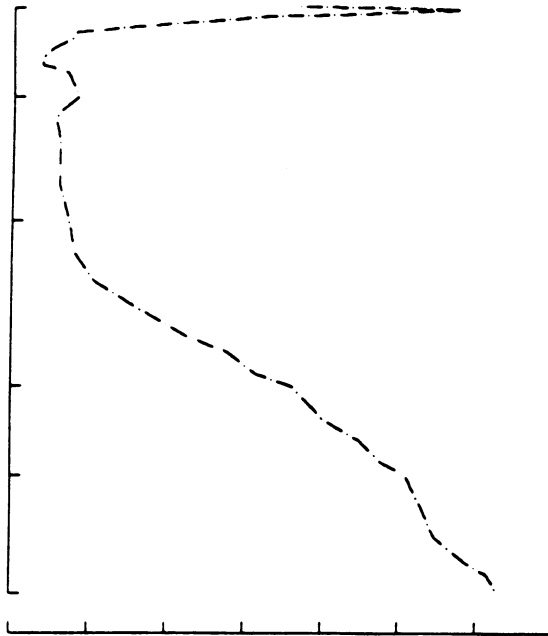
Averaging Kernel at $r = .4$ (Figure 2.1)



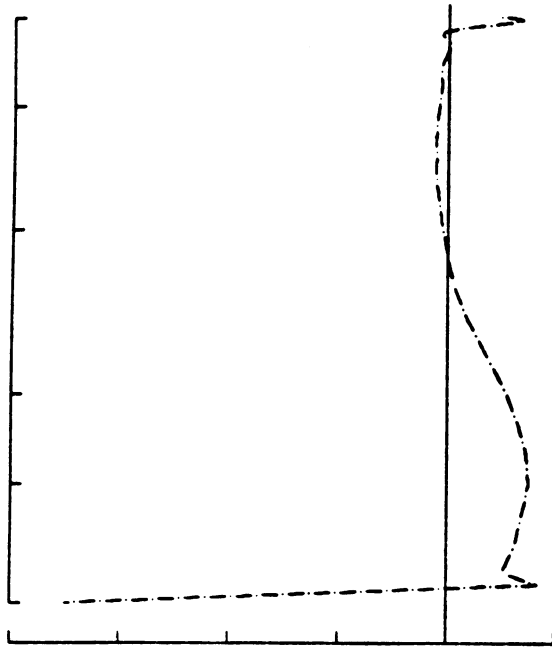
Inversion Characteristics (Figure 2.2)



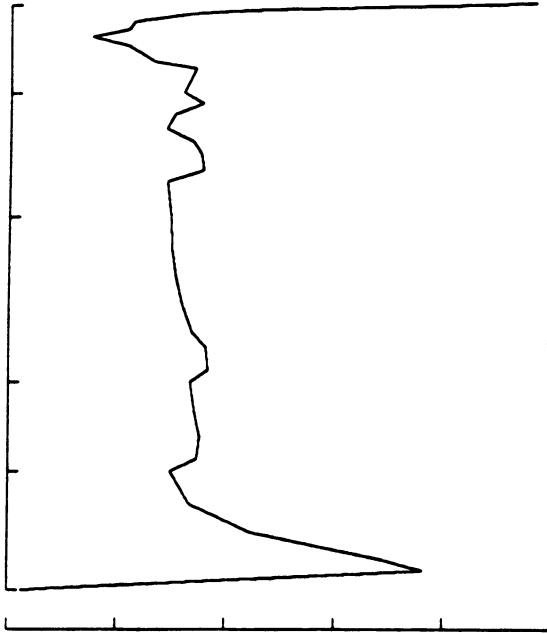
Climatological Temperature Profile (Figure 4.1)



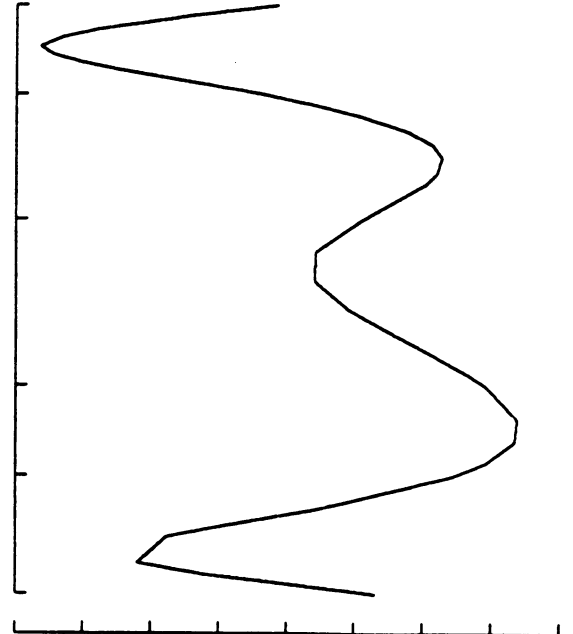
Averaging Kernel at 700 millibars (Figure 4.2)



Inversion Characteristics (Figure 4.3)

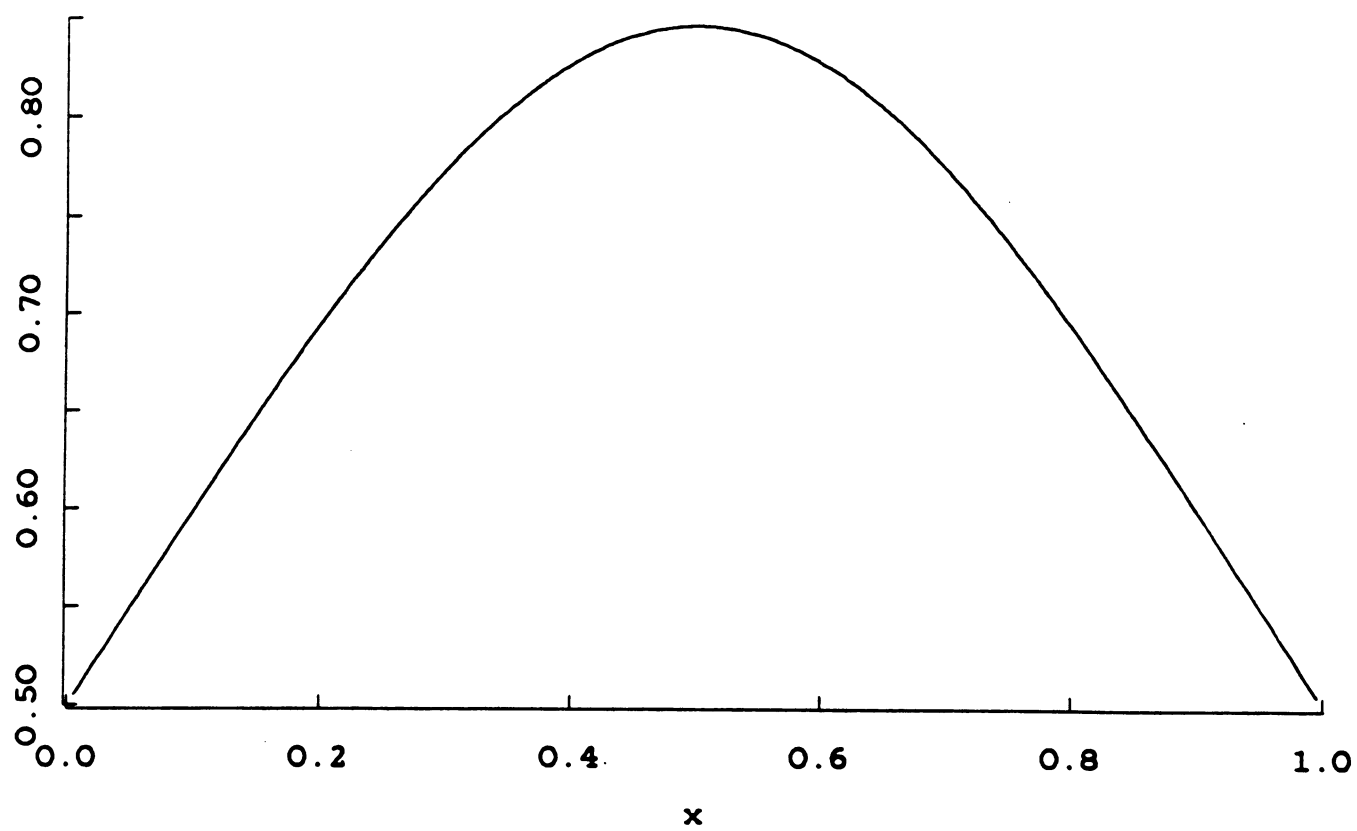


Maximum Bias

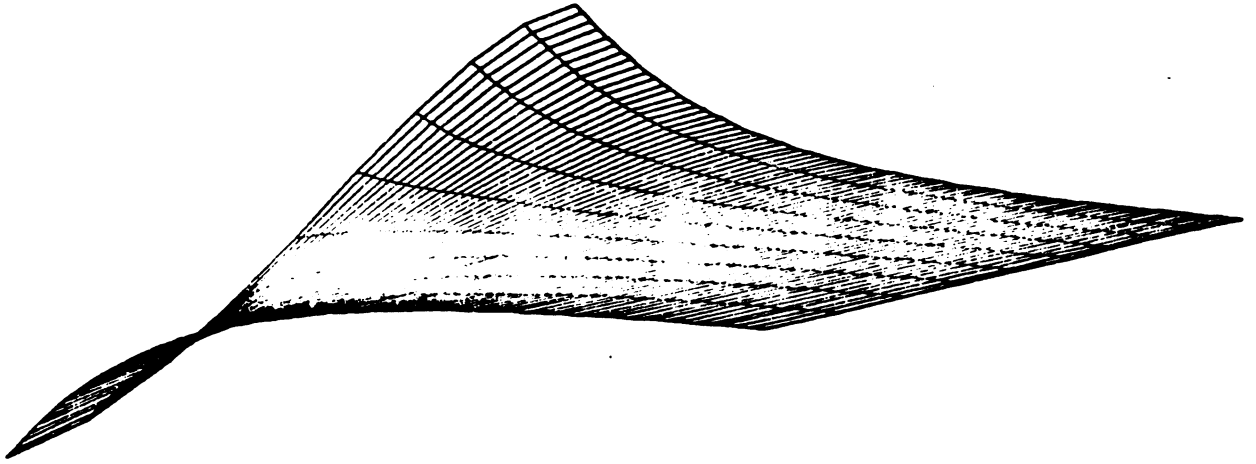


Standard Error

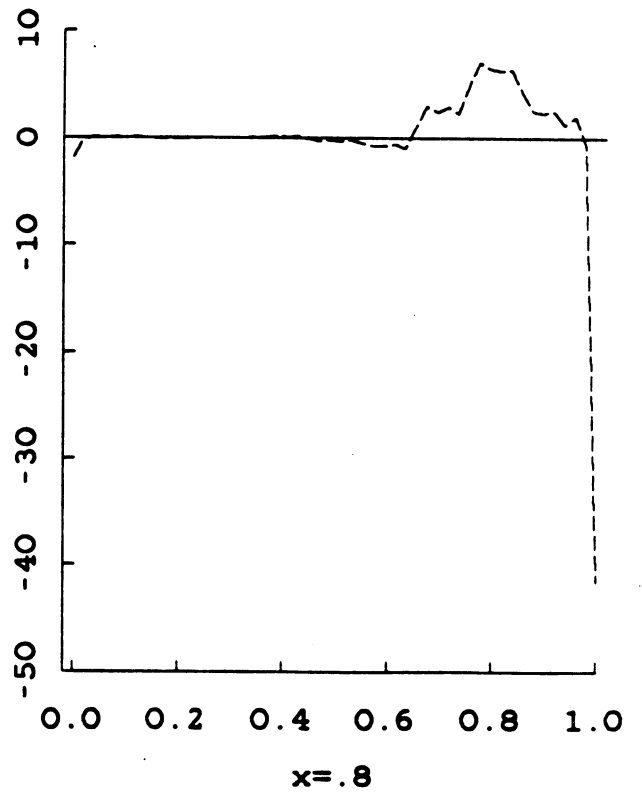
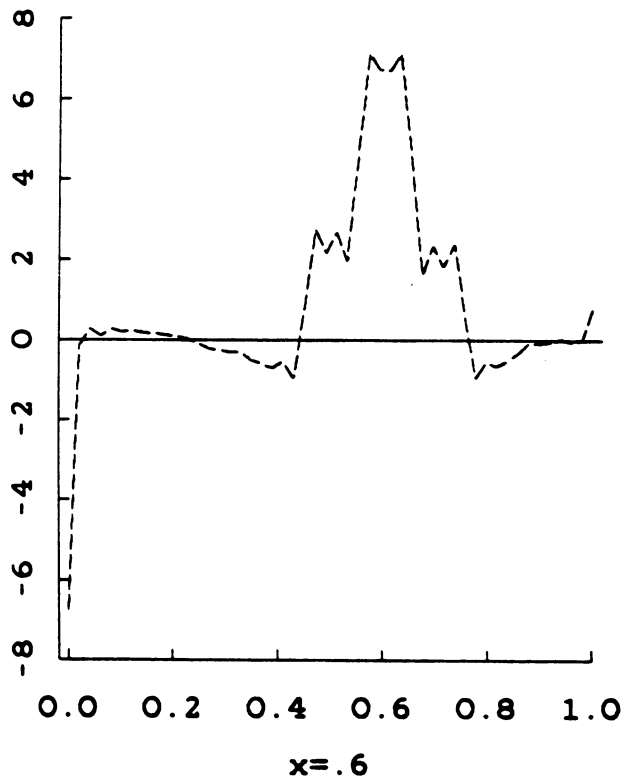
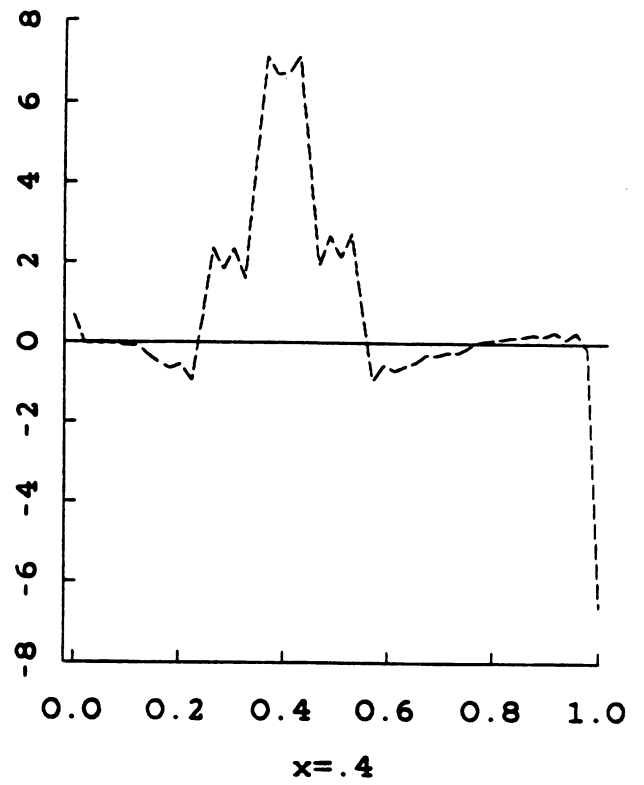
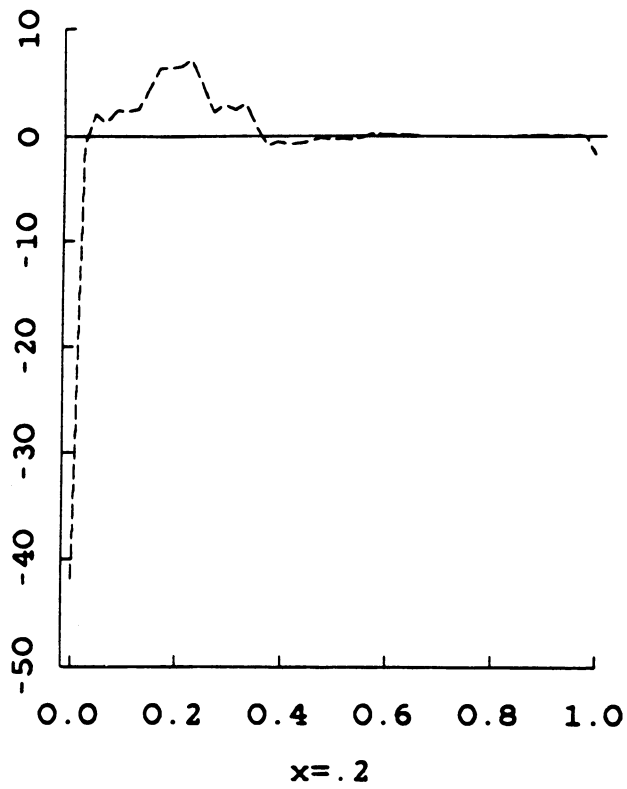
Transmissivity Profile (Figure 4.4)



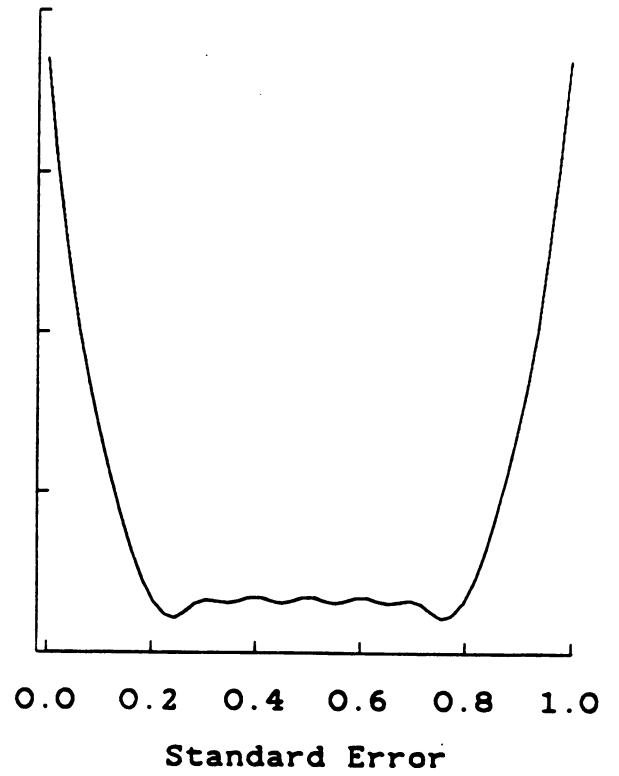
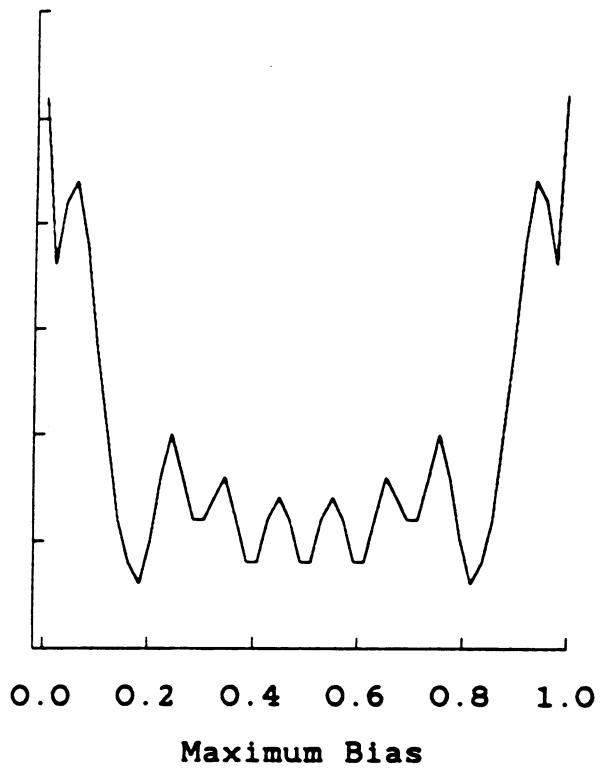
True Pressure History (Figure 4.5)



Sample Averaging Kernels (Figure 4.6)



Inversion Characteristics (Figure 4.7)



TECHNICAL REPORTS
Statistics Department
University of California, Berkeley

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.
2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosociences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.
3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.
4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11, No. 1, 185-214.
5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.
6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.
7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.
8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.
9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.
10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.
11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.
12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.
14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.
15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.
16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.
17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review. Statistical Science, 1988, Vol.3 No. 2 239-271.
18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.
19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).
20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.
21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.
22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.
23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984, 579-611.
25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.
26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.
27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.
28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.
29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.
30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES, 1985 Vol 3 pp. 1-13.
31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.
32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.
33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.
34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.
35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science, Feb 1986, Vol. 1, No. 1, 3-39.
36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.
37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly, Feb 1986, Vol. 93, No. 2, 123-125.
38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.
39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.
40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.
41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.
42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.
43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?
44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.
45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.
46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.
47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data. Scandinavian J. Statist., 1988, 15, 1-23.
48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.
49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.
50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada, January, 1986.
51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.
53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.
54. BLACKWELL, D. (November 1985). Approximate normality of large products.
55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics, 12, 101-128.
56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.
57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.
58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.
59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.
60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.
61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.
62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.
63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.
64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.
65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.
66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.
67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data. Biometrika, 1987, 74, 799-808.
68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).
69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.
71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.
72. LEHMANN, E.L. (July 1986). Statistics - an overview.
73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.
74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.
75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.
76. O'SULLIVAN, F. (September 1986). Relative risk estimation.
77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.
78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.
79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.
80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.
81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.
82. DOKSUM, K.J. and LO, A.Y. (Nov 1986, revised Aug 1988). Consistent and robust Bayes Procedures for Location based on Partial Information.
83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.
85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.
86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.
87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.
88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.
89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model. J. Amer. Statist. Assoc., 1988, 83, 744-749.
90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.
91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.
92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.
- 92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.
93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.
94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.
95. CANCELLED
96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.
97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in Environmental Health Perspectives.
98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.
99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.
100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.
101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.
102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.
103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.
104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.
105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.
106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.
107. CHENG, C-S. (Aug 1987, revised Oct 1988). Some orthogonal main-effect plans for asymmetrical factorials.
108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.
109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \leq k \leq n} S_k^* / ES_n^*$: A prophet inequality for sums of I.I.D. mean zero variates.
110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.
111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.
113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.
114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.
116. RITOV, Y. and BICKEL, P.J. (Sept.1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.
117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.
120. DONOHO, D. and LIU, R. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.
121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. To appear in Statistics a Guide to the Unknown.
122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.
123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov.1987, revised July 1988). Minimax risk for hyperrectangles.
124. ALDOUS, D. (November 1987). Stopping times and tightness II.
125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.
126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.
130. STONE, C.J. (Nov.1987, revised Sept. 1988). Uniform error bounds involving logspline models.
131. Same as No. 140
132. HESSE, C.H. (December 1987). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes
133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.
134. DUBINS, L.E. and SCHWARZ, G. (December 1987). A sharp inequality for martingales and stopping-times.
135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.
136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.
137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.
- 137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.
138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.
139. BERAN, R. (January 1988). Refining simultaneous confidence sets.
140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.
141. BRILLINGER, D.R. (Mar. 1989). a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, *Actas del 2° congreso latinoamericano de probabilidad y estadística matematica*, Caracas, 1985.
142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (Feb. 1988, revised Nov. 1988). Randomization in the two-armed bandit problem.
144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.
145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.
146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.
147. STARK, P.B. (March 1988). Strict bounds and applications.
148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.
149. NOLAN, D. (March 1988). Asymptotics for a multivariate location estimator.
150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.
151. NOLAN, D. (March 1988). Limit theorems for a random convex set.
152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.
153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.
154. DOKSUM, K.A. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.
155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.
156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.
157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.
158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.
159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.
160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.
161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).
162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.
163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform. (Revised by Tech Report No. 180).
164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.
165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.
168. LE CAM, L. (August 1988). Maximum likelihood an introduction.
169. BREIMAN, L. (Aug.1988, revised Feb. 1989). Submodel selection and evaluation in regression I. The X-fixed case and little bootstrap.
170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.
171. STONE, C.J. (September 1988). Large-sample inference for logspline models.
172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.
173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.
174. YOR, M. (October 1988). Interwinings of Bessel processes.
175. ROJO, J. (October 1988). On the concept of tail-heaviness.
176. ABRAHAMAS, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.

177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.
178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subfamilies.
179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signals from data missing low frequencies.
180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).
181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989). A model for step-stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.
182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.
183. BERAN, R. (November 1988). Calibrating prediction regions.
184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.
185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.
186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes
187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.
188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.
189. PITMAN, J.W. and YOR, M. (Mar. 1989). Some extensions of the arcsine law.
190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.
191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.
192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.
193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.
194. PEMANTLE, R. (Feb. 1989). When are touchpoints limits for generalized Pólya urns.
195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.
196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.
197. BREIMAN, L. and SPECTOR, P. (Mar. 1989). Submodel selection and evaluation in regression — the X-random case.
198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.

Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: \$1 per copy.