ESTIMATION OF PROPERTIES

IN

PETROLEUM RESERVOIRS

Thesis by

Piyush Chimanlal Shah

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1977

(Submitted October 1, 1976)

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Professors G. R. Gavalas and J. H. Seinfeld for their constant guidance during the course of this research.

Financial support from the California Institute of Technology in the form of Fellowships, Teaching and Research Assistantships is gratefully acknowledged.

I wish to thank the faculty and the staff of the Department of Aeronautics, CIT for their kind and understanding support during my stay here as a graduate student. The aid of Virginia Conner, Elizabeth Fox, Susan Pinter and others, who painstakingly typed this thesis, is gratefully acknowledged.

Finally, I wish to thank my wife, Varsha, for her understanding patience and constant encouragement during my work on this research.

ABSTRACT

The determination of parameters in a dynamical system, on the basis of noisy observations of its state is variously known as parameter estimation, identification or the inverse problem. In this work, the determination of porous rock property distribution in a petroleum reservoir using the production rate records and observed pressures (the history matching problem) is considered.

The history matching problem is inherently underdetermined because of the large number of unknown parameters relative to the available data. The number of unknowns can be reduced by representing the distributions by a small number of parameters (parameterization). The commonly used zonation approach involves a parameterization, but introduces a considerable modeling error. In chapter 1, Bayesian estimation theory is extended to history matching as an alternative to zonation; it is sought to alleviate the underdeterminacy through specification of a priori statistical information about the unknown parameters. Application of Bayesian estimation and zonation to the problem of porosity and permeability estimation in a one-dimensional single-phase reservoir indicates that the former yields superior estimates; this holds true even when the prior statistics involve large errors. The application of the conjugate gradient and the Gauss-Newton (or Marquardt's) algorithms for history matching is investigated, and the numerical effort for zonation and Bayesian estimation in one- and two-dimensional reservoirs is estimated in detail.

In chapter 2, analytic expressions are derived for the sensitivities of an observed oil pressure to small, arbitrary changes in the porosity and permeability distributions in a one-dimensional reservoir. The results indicate that highly oscillatory components of either have very small influence on the pressure and thus cannot be determined by history matching. Further, the dependence of all the observed pressures on the unknown parameters is linearized, for small deviation, about two reference property distributions. The linear relation is analyzed to yield quantitative information concerning the statistical properties of the problem. Iterative corrections in the history matching algorithms are identified with various pseudo-inverses of the linear relation, thus explaining the properties of the resulting estimates. The nature of the linear relation is found to be not strongly dependent on the reference property distributions used for linearization; thus such analysis can be performed prior to estimation. It is discussed how the linearized analysis can be used to determine the determinacy of any given parameterization.

The information derived from the linearized analysis and that in the a priori statistics is synthesized in chapter 3 to predict covariances for the zonation and Bayesian estimates. Since the results of the linearized analysis depend only weakly on the reference distribution, the predicted covariances are valid for a class of reservoirs having "true" property distributions with identical prior statistics. A good agreement is found when the predicted variances are compared with actual mean square estimate errors in simulations

with four distributions with given prior statistics. The sensitivity of the estimates and their covariance to changes and errors in the specification of the prior statistics are investigated in considerable detail. The determination of zonation with smallest trace of estimate covariance for a given problem is considered. The design of Marquardt's algorithm to yield the smallest expected total estimate error for a given zonation is discussed.

TABLE OF CONTENTS

Table of Contents (Cont'd)

Table of Contents (Cont'd)

Table of Contents (Cont'd)

Table of Contents (Cont'd)

List of Figures

List of Figures (cont.)

List of Figures (cont.)

List of Figures (cont.)

LIST OF TABLES

LIST OF TABLES (Cont.)

INTRODUCTION

Estimation of porous rock properties in a petroleum reservoir is important for the purpose of accurate modeling and performance optimization. The estimation of properties based on well pressure and production rate data belongs to the class of parameter estimation problems in distributed systems. Such problems are also known as identification or inverse problems.

In this dissertation, we consider estimation of the spatial distributions of transmissibility (assumed isotropic) and storage in a two-dimensional petroleum reservoir. For simplicity, we consider a single-phase oil reservoir with impermeable boundary. It is emphasized that the methodology developed here is easily extended to multi-phase reservoirs with arbitrary boundary conditions and, indeed, to the problem of identification of any distributed parameter dynamic system. For the single phase reservoir, the oil pressure as a function of space and time is governed by a linear, parabolic partial differential equation. From the geological and drilling data, the boundary location, the flow conditions at the boundary, and the initial oil pressure are usually known. The available data consist of the production rate history records and observations of the well pressures at a finite number of instants. It is assumed that the production rate history is perfectly known whereas the pressure observations are contaminated by

random observation errors with known statistics.

The problem of property estimation is usually posed as minimization of the total deviation of the calculated model pressures from the observations with respect to the unknown rock properties. This approach (Jacquard and Jain 1965, Jahns 1966, Carter et al. 1974) is commonly known as pressure history matching. Although in principle the property distributions are specified by an infinite number of parameters, a computational model can only contain a finite number. For computational purposes, the partial differential equation is approximated by finite differencing with a uniform spatial grid covering the reservoir. The most detailed description of the unknown distributions is obtained by allowing the properties to vary independently at each block of the grid. While minimizing the modeling error, this approach entails a great deal of uncertainty in the estimates due to a large number of unknowns compared to the limited data available.

The uncertainty in the estimates can be reduced either by reducing the number of unknowns through constraints on the unknowns (parameterization), or by incorporating in estimation auxiliary information, or both. The commonly used zonation approach (Jahns 1966, Thomas et al. 1972) involves a parameterization; it reduces the number of parameters by restricting the unknown property distributions to be uniform within each of several contiguous regions of the reservoir called zones. The larger the number of zones, the more detailed is the description of the property distributions; but, at the same time, the statistical uncertainty in the

parameter estimates is higher. On the other hand, very few and large zones yield accurate estimates but entail large modeling errors. Thus an intermediate number of zones yields a minimal overall estimation error.

As an alternative to zonation, we may attempt to reduce the uncertainty in the estimates by taking into account additional information about the unknowns, which is based on certain geological considerations. For example, the property variations, deriving from random sedimentation processes, should have a randomness encompassing a certain degree of smoothness. In Chapter 1, we seek to incorporate such auxiliary statistical information in the estimation process through extension of Bayesian estimation theory. Due to nonlinearity of the property estimation problem, direct application of classical Bayesian estimation is not feasible; instead, its penalty function formulation is utilized to incorporate prior statistical geological information.

The statistical information is derived from geological observations concerning typical property variations that occur in the sedimentary rock constituting the reservoir. Such observations need not be on the particular reservoir being treated, but may encompass a whole class of reservoirs formed under similar sedimentation conditions. However, laboratory measurements of local property values on reservoir samples (for example, obtained by coring) may also be taken into account while specifying the prior statistics.

We have suggested a statistical model for the property variations, and a method of specifying the statistics based on the radius of auto-correlation reflecting the degree of smoothness of distributions, to serve in absence of detailed geological information. However, scope for considerable improvement exists as more geological observations become available.

For illustration, Bayesian estimation and zonation are applied to the problem of estimating porosity and permeability distributions in a hypothetical one-dimensional reservoir, and their performance is compared. The numerical simulations are used to investigate such questions as the optimum number of parameters in zonation, the effect of erroneous prior statistics in Bayesian estimation. Considerable attention is given to computational aspects such as convergence rate and computer time required by two of the most commonly used minimization algorithms, Marquardt's and conjugate gradient.

The nonlinear parameter estimation problem is solved iteratively, the corrections in the current estimates being obtained through the solution of an associated linear subproblem. An analysis of this linear subproblem is apt to yield a considerable amount of information about the nature of the nonlinear problem. In Chapter 2, we carry out such an analysis of the linearized relation between small deviations in the unknown parameters and the observed pressures. It yields quantitative information about the ill-conditioned and underdetermined nature of the problem. It is shown that

iterative corrections in the different minimization algorithms can be identified with different pseudoinverses of the linear relation; thus the nature of the resulting estimates of the property distributions can be explained. Furthermore, such analysis can be used to determine the conditioning and determinacy of any given parameterization.

It is observed that the characteristics of the linear relation do not depend strongly on the reference distribution used for linearization. This is the key feature that makes it possible to perform the analysis before performing extensive estimation. It is also of cardinal importance in the determination of a posteriori covariance associated with the property estimates. The information about the unknown rock properties derives from two sources: the observations and the prior statistics. The observations yield information mainly about the components of the unknowns which are significantly sensitive to them. The information about the less sensitive components is derived from the prior statistics. The linearized analysis aids the separation between the two types of components, and thus the determination of the covariance. The weak dependence of the linear relation on the reference distribution allows prediction of the covariance before performing estimation, using only the prior estimates in analysis. Furthermore, the predictions are valid for an ensemble of distributions which do not differ significantly from those used in the analysis.

The total modeling error in zonation can be estimated using the results of the above analysis. When compared with the statistical uncertainty in the estimates, it provides a theoretical basis to determine a priori the optimum number of zones commensurate with the available data. As the prior statistical information is taken into account while predicting the estimate covariance, a change or error in the statistics can be expected to influence the predictions. Furthermore, since the Bayesian approach utilizes the prior statistics in the estimation process, the resulting estimates are also affected by such changes or errors. These aspects are investigated in considerable detail in Chapter 3.

CHAPTER 1

ESTIMATION OF ROCK PROPERTIES IN A SINGLE PHASE PETROLEUM RESERVOIR

In this chapter, we consider the estimation of porous rock property distributions (e.g. those of permeability and porosity) in a petroleum reservoir from well production and pressure data, which is commonly referred to as history matching. The reservoir geometry, boundary and initial conditions, and fluid properties are assumed to be known.

The estimation of reservoir properties is inherently an underdetermined problem because of the large number of unknown parameters relative to the available data; hence, its solution is non-unique. A commonly used approach to reducing the number of parameters is zonation, where the reservoir is divided into several zones and the properties within each zone are assumed uniform. The zonation entails considerable modeling errors due to arbitrary assignment of zone boundaries and the insistence that the properties are uniform within each zone. In this chapter Bayesian estimation theory is extended to history matching as an alternative to zonation. The underdetermined nature of the general history matching problem can be alleviated through specification of a priori statistical information on the unknown rock properties. The Bayesian estimation and zonation are compared through their application to the problem of porosity and permeability estimation in a one-dimensional, single-phase reservoir.

The mathematical model for a single-phase oil reservoir is formulated in section 1.1. In section 1.2, the history matching problem is formulated in continuous and discrete terms, and relative merits of the two formulations are considered. Further, non-uniqueness of the solution and consequent need for additional information or parameterization are discussed, citing zonation as an example of the latter. In section 1.3, we introduce Bayesian estimation theory and extend it to the history matching problem. The nature of the prior statistical information is discussed from a geological viewpoint, and a simple homogeneous random process model for the property distributions is developed to serve in the absence of any definite information. The Bayesian estimation is shown to result in a significant reduction in the number of parameters. The homogeneous random process model is used in section 1.4 to produce realizations, which serve as "true" property distributions in simulated test problems. The details of simulation are also included in this section. In section 1.5, we consider, including their derivation, two minimization algorithms for history matching: conjugate gradient and Gauss-Newton (or Marquardt) algorithms. Their convergence characteristics are discussed and detailed estimates are derived for computation effort per iteration of each algorithm, for zonation and Bayesian estimation in one-dimensional and two-dimensional reservoirs. Finally, in section 1.6, the numerical simulations with one-dimensional reservoir are used to investigate questions such as the optimum number of parameters in zonation, the effect of erroneous prior statistics in Bayesian esimation and to compare the performance of the two methods.

1.1 Mathematical Model of a Single Phase Oil Reservoir

In a single phase reservoir, the oil flows through the porous rock under the influence of pressure gradients towards opened wells, where it is removed. The energy required for driving the oil flow against the frictional resistance to the motion, and against the gravitational forces during the primary recovery, is provided by expansion of the oil and the porous rock as the ambient pressure declines. Whenever the reservoir has a contact at the boundary with an aquifer or a gas reservoir, the expansion of the fluid in the latter accompanied by motion of the front separating the two fluids supplies part of the energy necessary for driving the motion.

In order to describe the flow of oil in a single phase oil reservoir, one has to take into account the properties of the medium, the thermodynamic and transport properties of the oil and the mechanics of the fluid motion. However, to reduce the complexity of the problem, we shall make several simplifying assumptions. Firstly, we assume that the flow velocities are small enough everywhere that the flow description with the low Reynolds number approximation is accurate. In such a situation, for a flow in a horizontal plane, the flow velocity at any point in the reservoir is proportional to the gradient of the pressure at that point. This law of fluid motion is commonly known as Darcy's law (Collins, 1961). The constant of proportionality depends on local fluid and rock properties. We shall assume that the rock properties are isotropic. Then Darcy's law yields the flow velocity,

$$\underset{\sim}{u} = -\frac{k}{\mu}\underset{\sim}{\nabla} p \qquad (1.1.1)$$

where k and μ are the local rock permeability and oil viscosity respectively.

We shall assume that the vertical thickness h of the reservoir is small compared with other dimensions, and varies slowly with position and that the reservoir lies in a horizontal plane. Then we may describe the oil flow by a two-dimensional model and disregard the effect of gravity. The law of conservation of mass applied to the oil in a small prismatic volume element with height h of the reservoir leads to the continuity equation,

$$\underset{\sim}{\nabla}\cdot(h\rho\underset{\sim}{u}) = -\frac{\partial}{\partial t}(h\phi\rho) \qquad (1.1.2)$$

where ϕ is the local rock porosity and ρ is the local oil density. It will be assumed that the rock compressibility is sufficiently small, so that the porosity and permeability are independent of pressure, and consequently time invariant.

Introducing (1.1.1) into (1.1.2) we obtain

$$\underset{\sim}{\nabla}\cdot\left(\frac{h\rho k}{\mu}\underset{\sim}{\nabla}p\right) = h\phi\frac{\partial\rho}{\partial t}\,. \qquad (1.1.3)$$

The thermodynamic properties of the oil are assumed to be of a very simple nature: the oil density ρ is only a function of pressure p, and the compressibility of oil c is constant. Then we have,

$$c = \frac{1}{\rho}\frac{d\rho}{dp} = \text{const.} \qquad (1.1.4)$$

$$\underset{\sim}{\nabla}\rho = \underset{\sim}{\nabla}p\frac{d\rho}{dp} = c\rho\underset{\sim}{\nabla}p \qquad (1.1.5)$$

$$\frac{\partial\rho}{\partial t} = \frac{\partial p}{\partial t}\cdot\frac{d\rho}{dp} = c\rho\frac{\partial p}{\partial t} \qquad (1.1.6)$$

Introducing the last two relations into (1.1.3), we get

$$\rho \underset{\sim}{\nabla}\cdot \left(\frac{kh}{\mu} \underset{\sim}{\nabla} p\right) + c\ \rho \frac{kh}{\mu} \underset{\sim}{\nabla} p \cdot \underset{\sim}{\nabla} p = h\phi c \rho \frac{\partial p}{\partial t} \qquad (1.1.7)$$

If the fluid motion is assumed to be so slow as to yield very small pressure gradients, then the second term on the left of (1.1.7) may be neglected in comparison with the first. Consequently, we obtain a linear parabolic partial differential equation for pressure $p(x, y;t)$,

$$\phi c h \frac{\partial p}{\partial t} = \underset{\sim}{\nabla}\cdot \left(\frac{kh}{\mu} \underset{\sim}{\nabla} p\right) \qquad (1.1.8)$$

The parameter combinations ϕch and $\frac{kh}{\mu}$ are called respectively transmissibility v and storage w at a given point in the plane of the reservoir.

In order to complete the mathematical model of the physical reservoir, we need to prescribe boundary and initial conditions. The condition at the boundary ∂D of the domain D of the reservoir depends on what adjoins the reservoir. If the reservoir under consideration is surrounded by impermeable rock, then the appropriate condition is zero flux normal to the boundary; it leads to, through the use of Darcy's law,

$$\underset{\sim}{\nabla} p(x, y;t) \cdot \underset{\sim}{n} = 0 \qquad (x, y) \in \partial D \qquad (1.1.9)$$

where $\underset{\sim}{n}$ is the unit vector at (x, y) normal to the boundary. Alternatively, if the reservoir is surrounded by a large aquifer, pressure of which remains constant, then the boundary conditions for the initial stage of oil production is,

$$p(x, y;t) = \text{constant} = p_o \qquad (x, y) \in \partial D \qquad (1.1.10)$$

More complex flow conditions at the boundary due to a motion of the water front are possible. For the sake of simplicity, in most of this work we shall only consider the case of a reservoir with impermeable boundary described by (1.1.9).

The initial condition for (1.1.8), describing the state of the reservoir at the beginning of the oil flow, is clearly that of the fluid at rest with a uniform pressure throughout the horizontal domain D of the reservoir. The value of this initial pressure p_o will be assumed known, giving

$$p(x, y; 0) = p_o \qquad (x, y) \in D \qquad (1.1.11).$$

Figure (1.1.1) shows a schematic drawing of the reservoir.

The description of the producing reservoir is completed by including the producing wells. The wells opened in the reservoir will be assumed to traverse the whole of the thickness h. Then the flow in their vicinity may be approximated by a two-dimensional flow in the horizontal plane. The boundary condition at a well is the mathematical statement of the fact that the total flow of oil towards the well at any time must account for its current production rate. Thus, we have the additional boundary conditions,

$$\frac{kh}{\mu} \oint_{i^{th}\,\text{well}} \nabla p \cdot \underset{\sim}{n}\, dp = q_i(t) \qquad i = 1, 2, \ldots, I_p \qquad (1.1.12)$$

where the integral is over the periphery of the i^{th} well and I_p is the total number of producing wells. In (1.1.12), $q_i(t)$ is the volumetric flow rate of i^{th} well; it is negative for production.

An alternative description of the wells contained within the reservoir boundary is obtained by approximating the wells by point

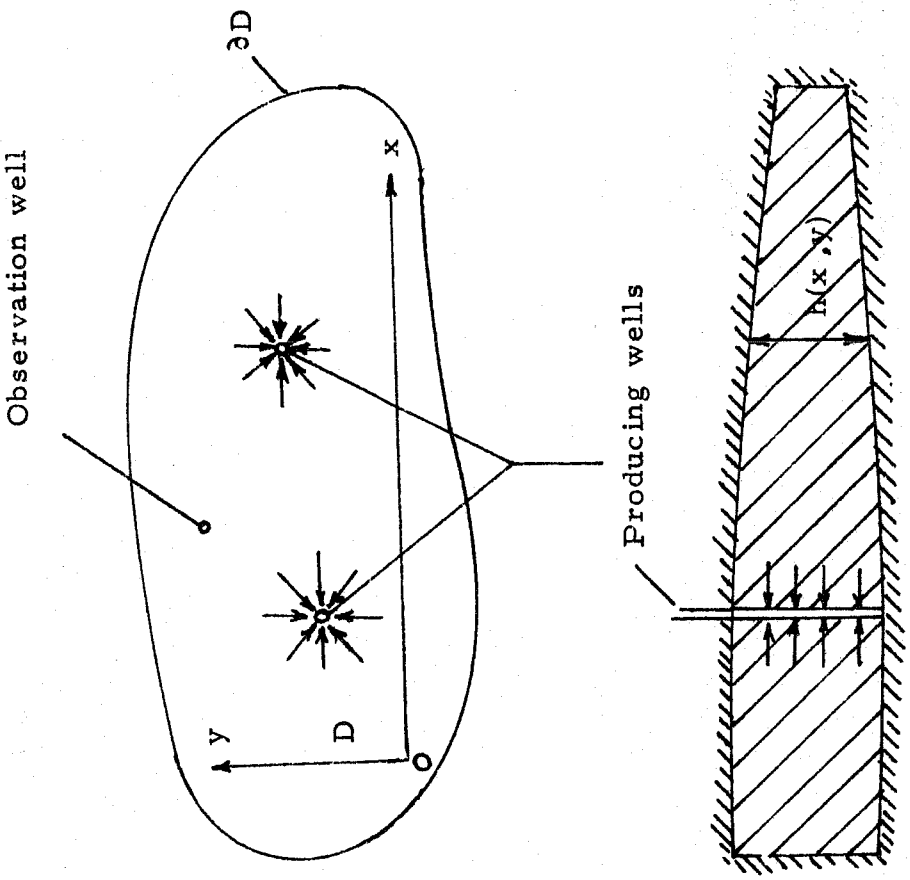


Figure 1.1.1

Schematic Drawing of a Two-Dimensional Reservoir

sources with strength $q_i(t)$ situated at the center of the respective wells. This representation makes the mathematical treatment more convenient than that given above and leads to additional terms containing delta functions on the right hand side of (1.1.8),

$$\phi ch \frac{\partial p}{\partial t} = \underset{\sim}{\nabla}\cdot\left(\frac{kh}{\mu} \underset{\sim}{\nabla} p\right) + \sum_{i=1}^{I_p} q_i(t)\delta(x-x_i)\delta(y-y_i) \qquad (1.1.13)$$

Equations (1.1.8, 9, 11, 12) or (1.1.13, 9, 11) constitute the mathematical model describing the oil flow in the reservoir.

1.2 Rock Property Estimation

1.2.1 The Property Estimation Problem

It is necessary to know the distributions of the transmissibility $v(x, y) = \frac{kh}{\mu}$ and the storage $w(x, y) = \phi ch$, to model accurately a given two-dimensional single phase oil reservoir. These functions are never known completely a priori, and must therefore be estimated from the available information about the reservoir. Usually, this information is in the form of records of production rate history for the different wells and the pressure measurements at different time instants at some or all of the wells. In addition, the analysis of the core samples provides information about v and w at the well locations. Geological information about the type of the reservoir, and information about the extent of the reservoir including the location of the boundary are also usually available.

A generally used method of utilizing the pressure and production history records involves seeking functions $v(x, y)$ and $w(x, y)$ which yield model pressures that match the measurements. This procedure of estimation is commonly referred to as history matching.

If the measurements of the production rates and the pressures are corrupted by unknown observation errors, as is invariably the case, insistence on an exact match between the model pressures and the measurements is not justifiable. In this situation, we may attempt a match in the least square sense, using weighting that depends on statistics of the different observation errors. Such a formulation of the history matching problem is,

$$\underset{v,w}{\text{Min}} \; J = J_p \equiv \sum_i \frac{1}{2\sigma_i^2}\left[p^{cal}(x_i, y_i; t_i) - p^{obs}(x_i, y_i; t_i)\right]^2 \qquad (1.2.1)$$

where the summation extends over all the observations. The quantities $p^{cal}(x_i, y_i; t_i)$ are the values of the solution of the model pressure equation (1.1.8) using the estimated functions $v(x, y)$ and $w(x, y)$ for the respective property distributions, and the recorded production history.

1.2.2 General Approaches to Estimation of Distributed Parameters

The task of determining the two functions $v(x, y)$ and $w(x, y)$ from a finite number of noisy observations is clearly impossible; there would be infinitely many sets of these functions which satisfy the data to the same degree of accuracy, making it impossible to choose any one of them as the correct solution. Furthermore, a computational model, of necessity, has to be finite dimensional.

However, the problem of distributed parameter estimation can be approached conceptually in two different ways, depending on at which stage a finite dimensional approximation is introduced. In figure (1.2.1) we show the two alternative approaches. In the left branch of the loop, the problem of minimization of J is posed in

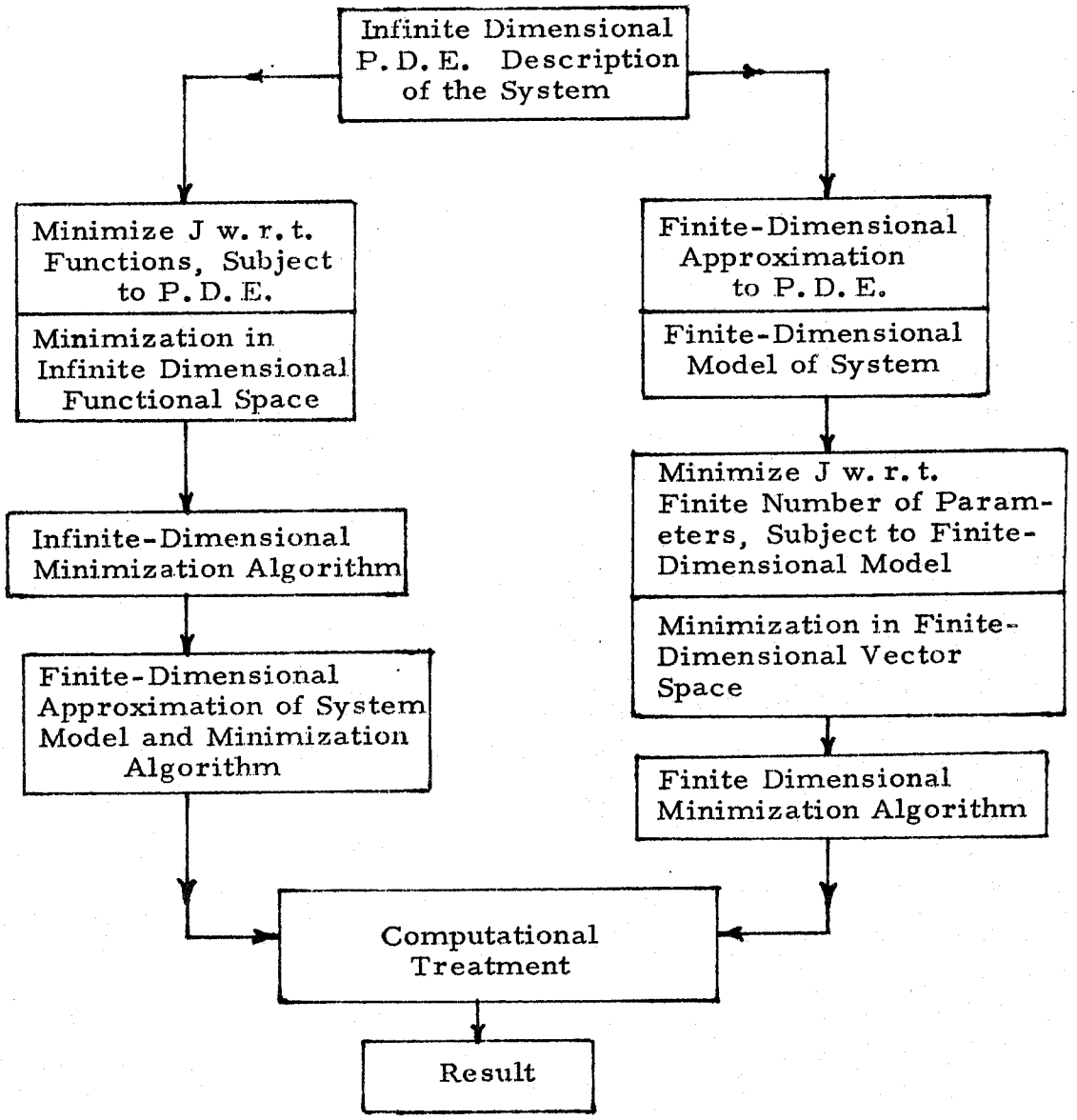


Figure 1.2.1

Alternative Approaches to Parameter Estimation in Distributed Systems

infinite dimensions, an algorithm is derived, and at the end it is approximated by finite differences or any other suitable device for computational purposes. In the right branch of the loop, the original infinite dimensional system model (partial differential equation) is approximated at the beginning by a finite dimensional model (discrete algebraic equations, for example) through finite differencing or other suitable methods; subsequently, an appropriate algorithm is derived for the finite dimensional minimization problem, which can be exactly implemented numerically. Thus, in this approach, no implementation errors are introduced while executing the minimization procedures for history-matching. Due to the exact numerical implementation of the minimization algorithm in the latter, this approach performs better computationally, yielding a better history match with less effort. Hence, we shall adopt this approach in the sequel.

We shall use a finite difference scheme employing a uniform spatial grid covering the domain of the reservoir for the purpose of the finite dimensional approximation. This would reduce the original partial differential equation to a set of coupled ordinary differential equations for the oil pressure at the grid points. These will be further approximated by differencing in time, using constant time steps to yield a set of discrete algebraic equations, which will be taken as the model for the system. The minimization of J will be carried out subject to these model equations, with respect to the grid point values of the rock properties v and w.

1.2.3 Non-uniqueness of Solution and Parameterization

The most detailed description of the unknown rock properties is obtained by allowing v and w to vary independently at each block of the spatial grid used in the finite differencing. For accurate modeling, a fine spatial grid is essential; this would lead to a large number of unknowns compared to the limited data available. Thus, while minimizing the modeling error, this approach entails a great deal of uncertainty, or even leads to a situation where the number of unknowns exceeds the number of data points, resulting in non-uniqueness of the solution to the property estimation problem.

The problem of the non-uniqueness can be dealt with by reduction of the number of unknowns or by injection of additional information, or by a combination of the two. The reduction of the unknowns can be affected by parameterization, i.e. introduce a set of fewer parameters along with a rule (mapping) which uniquely determines the values of v and w at all the grid points corresponding to a given value for this parameter set. The property estimation problem then reduces to that of minimizing J with respect to this parameter set. Zonation, described in the following subsection, is an example of parameterization.

We note that any parameterization that reduces the number of unknowns, necessarily introduces some constraints on the unknown rock properties. If the mapping between the rock properties and the parameters is linear, then the parameterization can be viewed as a projection of the unknowns onto a linear-subspace of their Euclidean space. For example, let $\underset{\sim}{\pi}$ be an M-vector of parameters and let the

linear mapping be,

$$\begin{bmatrix} \underset{\sim}{v} \\ --- \\ \underset{\sim}{w} \end{bmatrix} = \underset{\sim}{G}\,\underset{\sim}{\pi} \qquad (1.2.2)$$

where $\underset{\sim}{v}$ and $\underset{\sim}{w}$ are N-vectors respectively of the values of v and w at the N-grid points, and $\underset{\sim}{G}$ is a (2NxM) matrix of rank M. Then the above parameterization is equivalent to a mapping of the 2N-space of rock properties onto an M-space ($M \leq 2N$) of the parameters π_i. Alternatively, this parameterization is equivalent to the restriction that the component of the 2N-vector of rock properties be zero in a (2N-M) dimensional subspace of E^{2N}, which is the orthogonal complement of the subspace spanned by the columns of $\underset{\sim}{G}$. This is tantamount to (2N-M) linear equality constraints that the 2N property values must satisfy. These constraints are inflexible, and hence are often called "hard" constraints. Furthermore, each parameterization may be viewed as an introduction of some additional information into the property estimation problem; for example, (1.2.2) injects information that the components of the 2N-property vector $[\underset{\sim}{v}^T | \underset{\sim}{w}^T]^T$ along a (2N-M) dimensional subspace of E^{2N} is zero. We shall consider two different parameterizations involving linear mapping in the sequel.

Additional information may be injected into the property estimation problem in several ways. The estimated properties may be required to satisfy some "hard" inequality constraints on their values; for example, $w_{\ell_i} \leq w_i \leq w_{u_i}$, where w_{ℓ_i} and w_{u_i} are respectively the lower and the upper limits on the allowable values of w_i (Carter et al., 1974). Alternatively, the minimization index J may be modified by

the addition of a positive definite term involving the unknown property distributions. For example, the possibility of highly oscillatory estimates of the distributions of v(x, y) and w(x, y) can be restricted by the introduction of a penalty term in the index,

$$J = J_p + \int\int \left[a_1 \|\underset{\sim}{\nabla} v\|^2 + a_2 \|\underset{\sim}{\nabla} w\|^2 \right] dxdy \qquad (1.2.3)$$

where a_1 and a_2 are positive weighting factors. Similarly, penalty may be placed on higher derivatives of v and w. This approach to the injection of additional information was attempted but led to no significant improvement in the accuracy of the estimates. Contrary to the inequality constraints, this approach can be thought of as involving "soft" (i.e. flexible) constraints on the unknown distributions.

The former method of hard inequality constraints without any additional parameterization, does not yield satisfactory results; the property estimates often tend to lie along the boundary of the feasible region in E^{2N} (Carter et al, 1974). In addition, the minimization of J, subject to these hard constraints, is more difficult. The latter method of derivative penalty constraints has the shortcoming that the information introduced by it is rather arbitrary in nature; and it may lead to severe errors in the estimates. In section 1.3, we present a new approach, which we call Bayesian approach, that introduces additional geological information through a (soft) penalty constraint. It may be thought of as an extension of the derivative penalty method.

1.2.4 Zonation

In zonation approach to parameterization, the unknown property distributions are described by piece-wise constant approximations.

The values of a given rock property are assumed constant over contiguous regions of the reservoir, called "zones." Clearly, for uniform zone size, the larger the number of zones, the better is the approximation to the original distributions. Conversely, large zones yield less accurate approximation, but involve fewer parameters and consequently are conducive to a unique and more precise determination of the parameters. Thus the selection of the number of zones employed in the zonation approach, is a matter of compromise between two errors: the parameterization error and the error due to statistical uncertainty of the estimates. The question of optimum zonation commensurate with the available data will be discussed in chapter 3.

A major disadvantage of the zonation approach is that the zone boundaries are arbitrarily chosen, usually on the basis of well locations and have no bearing on the actual property variations in the reservoir. In addition, the inflexible constraint that the rock properties be uniform within zones and suffer sudden jumps across the zone boundaries is clearly artificial. The zonation approximation is not usually in harmony with the geologist's experience with property variations; deriving from random sedimentation processes, they should have a randomness rather distinct from piecewise constancy. The Bayesian approach, introduced in section 1.3, restricts the statistical uncertainty by incorporating probabilistic information concerning this randomness.

1.3 Bayesian Approach to Property Estimation

1.3.1 Bayesian Estimation

The Bayesian approach to estimation, involves a probabilistic

description of the parameters to be estimated. The estimation process consists of updating the prior probability density for the parameters using the observations, yielding a posteriori probability density. The estimates and their statistical properties (for example, their covariance) are determined from the a posteriori probability density (Bryson and Ho 1969, Sage and Melsa 1971).

1.3.1.1 Linear Bayesian Estimation

For clarity we present in this subsection Bayesian estimation applied to a linear estimation problem; the reader is cautioned that the notation in this exposition is different from the main text of this chapter. (See the box.)

The typical Bayesian estimation problem is concerned with the linear model (Bryson and Ho 1969, Sage and Melsa 1971)

$$\underset{\sim}{z} = \underset{\sim}{H}\,\underset{\sim}{x} + \underset{\sim}{\epsilon} \tag{1.3.1}$$

where $\underset{\sim}{x} = (x_1, \ldots, x_M)$ is the random vector to be estimated, $\underset{\sim}{z} = (z_1, \ldots, z_N)$ the measurement vector, $\underset{\sim}{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)$ is the random error vector and $\underset{\sim}{H}$ a known N x M matrix. Let the $\underset{\sim}{x}$ and $\underset{\sim}{\epsilon}$ be normally distributed vectors with means $\underset{\sim}{\mu}$, $\underset{\sim}{0}$ and covariances $\underset{\sim}{V}_x$, $\underset{\sim}{V}_\epsilon$ respectively. The parameters $\underset{\sim}{\mu}$ and $\underset{\sim}{V}_x$ constitute the prior knowledge about the vector $\underset{\sim}{x}$.

In Bayesian estimation the estimates of $\underset{\sim}{x}$ are defined by the maximization of the conditional probability density $p(\underset{\sim}{x}|\underset{\sim}{z})$. This density $p(\underset{\sim}{x}|\underset{\sim}{z})$ is given by Bayes rule:

$$p(\underset{\sim}{x}|\underset{\sim}{z}) = \frac{p(\underset{\sim}{z}|\underset{\sim}{x})\,p(\underset{\sim}{x})}{p(\underset{\sim}{z})} \tag{1.3.2}$$

Using (1.3.1) we find that $p(\underset{\sim}{z}|\underset{\sim}{x})$ is normal with mean $\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{x}$ and variance $\underset{\sim}{V}_\epsilon$; and $p(z)$ is normal with mean $\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{\mu}$ and variance $\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{\underset{\sim}{V}}_x\underset{\sim}{H}^T + \underset{\sim}{V}_\epsilon$ so that

$$p(\underset{\sim}{x}|\underset{\sim}{z}) = \frac{1}{(2\pi)^{M/2}} \frac{\det(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{\underset{\sim}{V}}_x\underset{\sim}{H}^T + \underset{\sim}{V}_\epsilon)^{\frac{1}{2}}}{\det\underset{\sim}{\underset{\sim}{V}}_x \det\underset{\sim}{V}_\epsilon}$$

$$\cdot \exp\{-\tfrac{1}{2}(\underset{\sim}{z}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{x})^T\underset{\sim}{V}_\epsilon^{-1}(\underset{\sim}{z}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{x})-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T\underset{\sim}{\underset{\sim}{V}}_x^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$$

$$+\tfrac{1}{2}(\underset{\sim}{z}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{\mu})^T(\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{\underset{\sim}{V}}_x\underset{\sim}{H}^T+\underset{\sim}{V}_\epsilon)^{-1}(\underset{\sim}{z}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{\mu})\} \qquad (1.3.3)$$

The maximization of $p(\underset{\sim}{x}|\underset{\sim}{z})$ with respect to $\underset{\sim}{x}$ is equivalent to the minimization of the quadratic function

$$J = (\underset{\sim}{z}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{x})^T\underset{\sim}{V}_\epsilon^{-1}(\underset{\sim}{z}-\underset{\sim}{\underset{\sim}{H}}\underset{\sim}{x}) + (\underset{\sim}{x}-\underset{\sim}{\mu})^T\underset{\sim}{\underset{\sim}{V}}_x^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) \qquad (1.3.4)$$

The first term in J represents the weighted mismatch between the model and observations while the second represents the weighted deviation of $\underset{\sim}{x}$ from its a priori mean $\underset{\sim}{\mu}$.

1.3.1.2 Extension to Reservoir Estimation

As shown in subsection (1.3.1.1), the Bayesian estimation reduces to a quadratic minimization problem provided the parameters and the measurement errors are normally distributed and the model is linear in the parameters. When these conditions are not satisfied, the rigorous application of Bayesian estimation is extremely tedious and is used very rarely. Since the reservoir estimation problem is nonlinear in the parameters, a rigorous application of Bayesian estimation is impractical. However, the following practical approach,

akin to least squares estimation, is feasible. By analogy to (1.3.4) we define an objective function

$$J = J_p + J_2 \tag{1.3.5}$$

where J_p is the customary objective function based on the difference between the observed and calculated pressures defined in (1.2.1), and

$$J_2 = \underset{\sim}{u}^T \underset{\sim}{P}_o^{-1} \underset{\sim}{u} \tag{1.3.6}$$

where,

$$\underset{\sim}{u}^T = [\underset{\sim}{v}-\overline{\underset{\sim}{v}})^T | (\underset{\sim}{w}-\overline{\underset{\sim}{w}})^T] \tag{1.3.7}$$

and

$$\underset{\sim}{P}_o = E\{\underset{\sim}{u}\,\underset{\sim}{u}^T\} \tag{1.3.8}$$

with

$$\overline{\underset{\sim}{v}} = E\{\underset{\sim}{v}\}, \quad \overline{\underset{\sim}{w}} = E\{\underset{\sim}{w}\} \quad . \tag{1.3.9}$$

Implicit in the above formulation is the assumption that the unknown property vectors can be described by an a priori probability distribution. In the following subsection we describe a method for arriving at such a prior probability distribution, through the use of geological information. The expectations in (1.3.8-9) are taken with respect to this probability distribution. The quantities $\overline{\underset{\sim}{v}}$ and $\overline{\underset{\sim}{w}}$ are the prior mean values of the property vectors. The matrix $\underset{\sim}{P}_o$ is the prior covariance associated with the 2N-vector $[\underset{\sim}{v}^T | \underset{\sim}{w}^T]^T$.

The estimation is then defined by the minimization of the composite index $J = J_p + J_2$. The term J_p represents the measurement error, σ_i^2 being the variance of the i^{th} measurement, J_2 is a Bayesian

term penalizing the weighted deviations of the parameters from their prior mean values. Minimization of J_p results in utilization of the information in the observations; J_2 introduces the prior information about the parameters into the estimation procedure. By adding J_2 we require that the parameters follow some preconceived pattern. This requirement reduces the statistical uncertainty at the cost of increasing the residual observation error. The relative weight in the two terms J_p, J_2 is determined by the variance of the observation error, σ_i^2, and by $\underset{\sim}{P}_o^{-1}$. When little is known a priori about the rock properties, $\underset{\sim}{P}_o$ will have large diagonal elements, and the elements of $\underset{\sim}{P}_o^{-1}$ will be small so that the term J_2 will be given a small weight. The weighting of the prior information can be reduced, if desired, by modifying the term J_2 of (1.3.6) by

$$J_2 = \beta \, \underset{\sim}{u}^T \underset{\sim}{P}_o^{-1} \underset{\sim}{u} \qquad (1.3.6a)$$

where $0 < \beta \leq 1$.

The above formulation assumes prior knowledge about the means of transmissibility and storage. If it is desired to use prior information about the correlation of these properties but not about their means, $\underset{\sim}{\bar{v}}$, $\underset{\sim}{\bar{w}}$, the latter quantities can also be considered as parameters to be estimated. As discussed later in the subsection (1.3.2) these vectors may be parameterized by a few scalars; thus, such an alteration will not increase the total number of unknowns significantly.

Lastly, we make an important point about the nomenclature. Although the approach to property estimation suggested here is called

Bayesian approach, it should be more properly called pseudo-Bayesian approach, since it is not rigorous. However, for the sake of brevity, we will continue to call it Bayesian approach; it is hoped that no misunderstanding will result for this somewhat inexact nomenclature.

1.3.2 Nature of Prior Geological Information

The application of probabilistic models in geology is the subject of a recent review (Marriam, 1976). In applications discussed, a given class of geological structures is modeled by a random process with a certain probability distribution. A specific structure is regarded as a realization of the random process. A realization is then the outcome of a random event which in our case is the sedimentation process responsible for the formation of the porous medium. The articles of Preston and Davis in the above reference (Marriam, 1976) treat sedimentary formations with regard to microstructure. The distribution of rock properties such as porosity and permeability pertaining to the macrostructure, i.e., the reservoir as a whole, can also be considered as random processes resulting from the joint action of several random conditions during sedimentation. Pryor (1972) has measured porosities and permeabilities as functions of position in recent sand bodies. Similar measurement in actual petroleum reservoirs require extensive drilling, and have not been carried out.

To simplify notation we consider a linear or radially symmetric reservoir so that the rock properties v and w vary in one direction only, although the following discussion can easily be generalized to multidimensional problems. In the discrete description, these properties

are characterized by random N-vectors $\underset{\sim}{v} = (v_1, \ldots, v_N)$, $\underset{\sim}{w} = (w_1, \ldots, w_N)$.

The geological information required for Bayesian estimation includes the means and covariance matrices of $\underset{\sim}{v}$ and $\underset{\sim}{w}$. Let these be given by,

$$E\{v_i\} = \bar{v} \tag{1.3.10}$$

$$E\{w_i\} = \bar{w} \tag{1.3.11}$$

$$E\{(v_i-\bar{v})(v_j-\bar{v})\} = P_{ij} = f(|i-j|) \tag{1.3.12}$$

$$E\{(w_i-\bar{w})(w_j-\bar{w})\} = Q_{ij} = g(|i-j|) \tag{1.3.13}$$

$$E\{(v_i-\bar{v})(w_j-\bar{w})\} = R_{ij} = h(|i-j|) \tag{1.3.14}$$

The reader is cautioned against confusing the function $h(|i-j|)$ with the depth h(x) (or h) of the reservoir. The function $f(|i-j|)$ and $g(|i-j|)$ are autocorrelations of transmissibility and storage respectively; $h(|i-j|)$ is their cross-correlation function.

The definitions (1.3.10-14) imply that the processes v and w are homogeneous. This simple description is appropriate in the absence of information suggesting systematic spatial variations. If such variations are known to exist, they can be treated by parameterizing the mean of v_i and w_i by smooth functions, for example:

$$E\{v_i\} = \bar{v}_i = ax_i + b \tag{1.3.15}$$

$$E\{w_i\} = \bar{w}_i = a'x_i + b' \tag{1.3.16}$$

The parameters P_{ij}, Q_{ij}, and R_{ij} are then defined by equations (1.3.12-14) with $\bar{v}$, $\bar{w}$ replaced by $\bar{v}_i$ and $\bar{w}_i$.

In a purely Bayesian approach, the parameters $\bar{v}$, $\bar{w}$ (or a, b, a', b') and $\underset{\sim}{P}$, $\underset{\sim}{Q}$, $\underset{\sim}{R}$, are considered as known, representing the prior information. It is possible, however, to regard $\bar{v}$ and $\bar{w}$ (or a, b, a', b') as parameters to be estimated. In all cases $\underset{\sim}{P}$, $\underset{\sim}{Q}$, and $\underset{\sim}{R}$ must be treated as known.

The determination of $\underset{\sim}{P}$, $\underset{\sim}{Q}$, and $\underset{\sim}{R}$ can in principle be carried out using past measurements on geologically similar reservoirs that have been subjected to extensive core sampling. When such data are not available, we must proceed in a more qualitative fashion. Consider, for example $P_{ij} = f(|i-j|)$. This function, being an autocorrelation, attains its maximum at i = j and the quantity f(0), which is the variance of v_i, may be specified using previous information concerning the size of fluctuations in the transmissibility. Experience usually suggests upper and lower limits for reservoirs of a given type, e.g., $v_\ell < v_i < v_u$. Then we may set

$$f(0) = \left(\frac{v_u - v_\ell}{2}\right)^2 \qquad (1.3.17a)$$

$$g(0) = \left(\frac{w_u - w_\ell}{2}\right)^2 \qquad (1.3.17b)$$

While f(0) and g(0) depend on the magnitude of the fluctuations at a single location, $f(|i-j|)$ and $g(|i-j|)$, for $i \neq j$, depend on the magnitude and correlation of fluctuations at two locations i and j. In the case of rapid lateral variability of properties, $f(|i-j|)$ and $g(|i-j|)$ are rapidly decreasing (in absolute value) functions of $|i-j|$, while in the case of smooth property variations, $f(|i-j|)$ and $g(|i-j|)$ decline slowly with increasing $|i-j|$.

Very little geological information is at present available regarding the lateral variability of reservoir properties. We are thus led to adopt empirical expressions for the spatial correlations, for example

$$f(|i-j|) = f(0)\, e^{-|i-j|^2/s^2} \tag{1.3.18}$$

$$g(|i-j|) = g(0)\, e^{-|i-j|^2/s^2} \tag{1.3.19}$$

making the reasonable assumption that the spatial correlations of storage and transmissibility have the same functional form. Furthermore, transmissibility and storage are correlated with each other as shown by laboratory measurements on core samples (Pryor, 1972). One possible representation of this correlation is

$$h(|i-j|) = \rho f^{\frac{1}{2}}(|i-j|) g^{\frac{1}{2}}(|i-j|) \tag{1.3.20}$$

where ρ is independent of i and j and satisfies $0 < \rho < 1$.

The parameter s introduced in equations (1.3.18-19) is a measure of the length scale of property variations. Thus, if Δx is the grid size in the finite difference model of the reservoir, $s\Delta x$ is the correlation length, i.e., the distance beyond which the property values at some location are essentially uncorrelated with those at another. Figure (1.3.1) illustrates the function $f(|i-j|)$ for various values of s while figure (1.3.2) gives property distributions obtained as realizations from spatial correlations with different values of s. The details of the computational procedure for obtaining the realizations are discussed later in section 1.4. A comparison of figures (1.3.1) and (1.3.2) illustrates how the nature of the correlation function $f(|i-j|)$

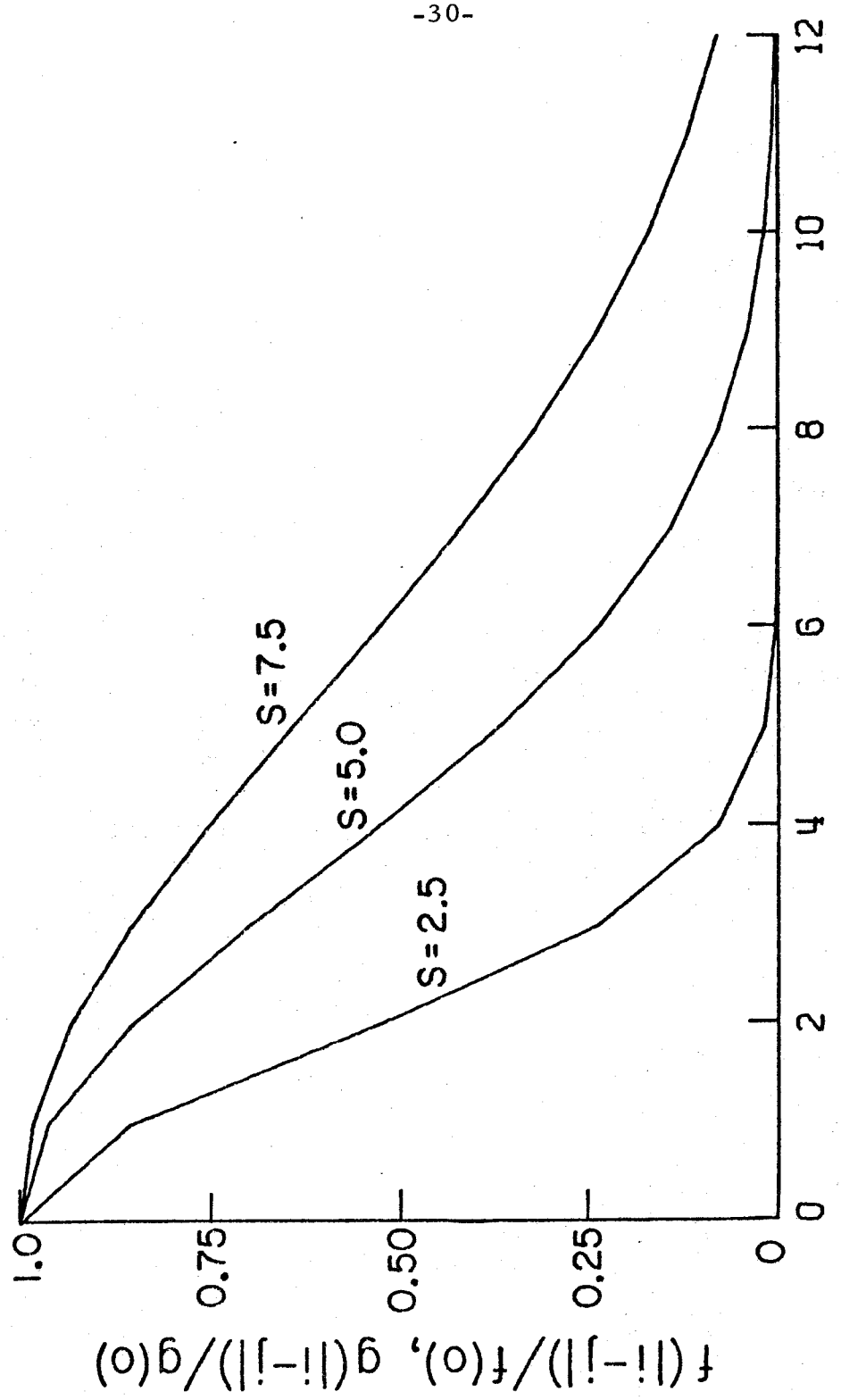
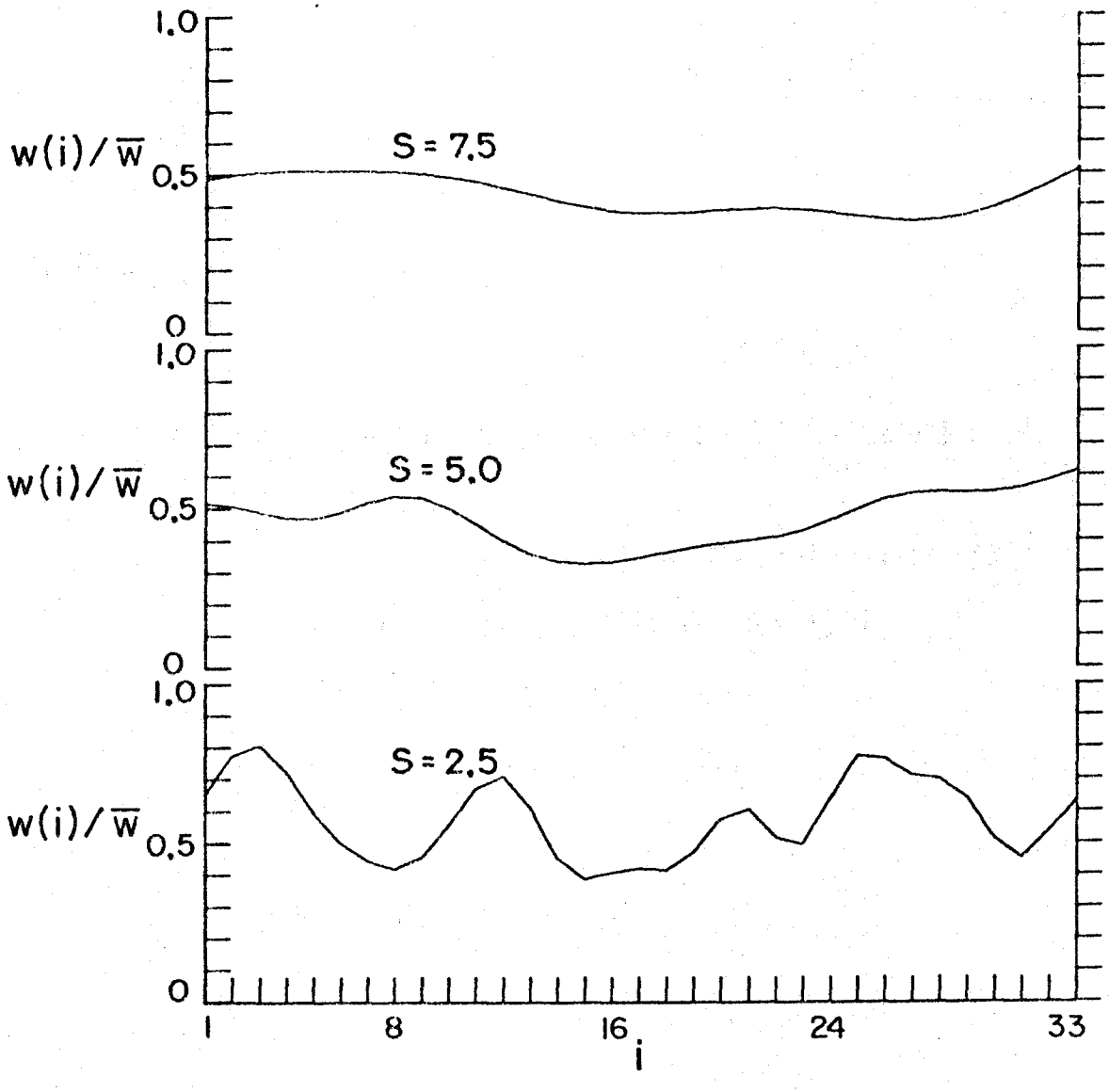


Figure 1.3.1 The Autocorrelation Function for Transmissibility and Storage



REALIZATIONS FOR DIFFERENT AUTOCORRELATIONS

Figure 1.3.2

relates to the oscillatory content of the property distributions. The choice of s hopefully can be made with some guidance from the geologist. As will be shown later, Bayesian estimation provides reasonable results even when the values of f(0), g(0) and s used contain considerable errors.

With these definitions of $\underset{\sim}{P}$, $\underset{\sim}{Q}$ and $\underset{\sim}{R}$, the (2Nx2N) matrix of prior covariance is given by,

$$\underset{\sim}{P}_o = \begin{bmatrix} \underset{\sim}{P} & \underset{\sim}{R} \\ \underset{\sim}{R}^T & \underset{\sim}{Q} \end{bmatrix} \quad . \tag{1.3.21}$$

We note that in a two or three-dimensional reservoir, the assumption of homogeneity of the random process would lead to, instead of the second relations in (1.3.12-14),

$$P_{ij} = f(d_{ij}) \tag{1.3.22}$$

$$Q_{ij} = g(d_{ij}) \tag{1.3.23}$$

$$R_{ij} = h(d_{ij}) \tag{1.3.24}$$

where d_{ij} is the absolute distance (positive) between the grid points i and j. Furthermore, (1.3.18-19) may be replaced by

$$f(d) = f(0)\, e^{-d^2/(s\Delta x)^2} \tag{1.3.25}$$

$$g(d) = g(0)\, e^{-d^2/(s\Delta x)^2} \quad . \tag{1.3.26}$$

When more detailed geological information is available, the above random process model for v and w can be refined. For example, when geological information suggests that the reservoir consists of two or more sections with distinct properties, the mean values $\bar{v}$, $\bar{w}$ can be chosen independently in each of these sections.

1.3.3 Reduction of the Number of Parameters

The Bayesian penalty term defined by equation (1.3.6) can be rewritten as a sum of the squares as follows. Let $\lambda_1 > \lambda_2 > \ldots > \lambda_{2N} > 0$ be the characteristic values of $\underset{\sim}{P}_o$, $\underset{\sim}{z}(1), \ldots, \underset{\sim}{z}(2N)$ be the corresponding normalized characteristic vectors and $\underset{\sim}{Z}$ be defined by:

$$\underset{\sim}{Z} = (\underset{\sim}{z}(1) \ \ldots \ \underset{\sim}{z}(2N)) \tag{1.3.27}$$

Since $\underset{\sim}{P}_o$ is symmetric, $\underset{\sim}{Z}$ is orthogonal and $\underset{\sim}{Z}^T \underset{\sim}{P}_o^{-1} \underset{\sim}{Z}$ is diagonal with entries $1/\lambda_i$. Thus, from (1.3.6)

$$J_2 = \sum_{i=1}^{2N} \frac{\xi_i^2}{\lambda_i} \tag{1.3.28}$$

where $\underset{\sim}{\xi} = (\xi_1, \ldots, \xi_{2N})$ is the vector

$$\underset{\sim}{\xi} = \underset{\sim}{Z} \underset{\sim}{u} \tag{1.3.29}$$

Then $\underset{\sim}{u}$ is given in terms of $\underset{\sim}{\xi}$ by

$$\underset{\sim}{u} = \underset{\sim}{Z}^T \underset{\sim}{\xi} = \sum_{i=1}^{2N} \xi_i \underset{\sim}{z}(i) \tag{1.3.30}$$

In most cases of practical interest the characteristic values λ_i decline very rapidly with increasing i as a result of the spatial correlation of the reservoir properties. From equation (1.3.28) it follows that those ξ_i corresponding to very small λ_i will be effectively suppressed in the minimization. It is thus computationally convenient to retain a limited number M of terms in equation (1.3.30), setting

$$\underset{\sim}{u} \cong \sum_{i=1}^{M} \xi_i \underset{\sim}{z}(i) \tag{1.3.31}$$

and then minimize the modified index

$$J = J_p + \sum_{i=1}^{M} \frac{\xi_i^2}{\lambda_i} \qquad (1.3.32)$$

with respect to $\xi_1, \ldots, \xi_M$. The number of unknown parameters has now been reduced from 2N to M. An additional advantage accruing from this is alleviation of ill-conditioning due to the inverse of the nearly singular matrix $\underset{\sim}{P}_o$ in (1.3.6). Both of these are very important from a numerical standpoint. As before, the weight placed on prior information may be reduced by setting

$$J = J_p + \beta \sum_{i=1}^{M} \frac{\xi_i^2}{\lambda_i} \qquad (1.3.32a)$$

where $0 < \beta < 1$.

Expression (1.3.29) indicates that, in effect, the Bayesian approach to estimation results in a parameterization, the property vector being a linearly related to the parameter vector.

1.3.4 Non-Gaussian Probability Density for Parameters

Clearly, actual reservoir properties are not normally distributed. For example, based on the Gaussian distribution, there is a finite probability that any grid point value of a rock property is negative. Such a situation is not physically meaningful. In practice, however, this may not cause any difficulty because the probability density is not used directly in the estimation procedure, and the penalty term (1.3.6) usually leads to property estimates which are close to their prior mean values. On the other hand, some special statistical relation between the properties may make a non-Gaussian representation more attractive.

In general, maximization of non-Gaussian probabilities do not

lead to simple quadratic minimization criteria. Hence, it is preferable to define a suitable nonlinear parameterization, such that the parameters themselves have a Gaussian distribution. For example, let

$$\begin{bmatrix} \underset{\sim}{v} \\ \hline \underset{\sim}{w} \end{bmatrix} = \underset{\sim}{f}(\underset{\sim}{\zeta}) \qquad (1.3.33)$$

where $f_i(\underset{\sim}{\zeta}) > 0 \qquad \forall\, \underset{\sim}{\zeta}, \quad i = 1, 2, \ldots, 2N.$

and let $\underset{\sim}{\zeta}$ be Gaussian with mean $\overline{\underset{\sim}{\zeta}}$ and prior covariance $\underset{\sim}{P}_{\zeta}$. Then the arguments similar to those employed in arriving at (1.3.5) may be used to obtain the minimization criterion,

$$J = J_p + \beta(\underset{\sim}{\zeta} - \overline{\underset{\sim}{\zeta}})^T \underset{\sim}{P}_{\zeta}^{-1}(\underset{\sim}{\zeta} - \overline{\underset{\sim}{\zeta}}) \qquad (1.3.34)$$

where the minimization is to be done with respect to the parameter vector $\underset{\sim}{\zeta}$.

1.4 Conditions of Simulation

1.4.1 Reservoir Descriptions

The procedure for parameter estimation described in the previous section was tested and compared to zonation by means of a hypothetical one-dimensional reservoir ($0 \leq x \leq 3,200$ ft.) with impermeable boundaries and a known uniform thickness. The compressibility and viscosity of the oil were taken to have constant values, $c = 5 \times 10^{-6}$ $(\text{psi})^{-1}$ and $\mu = 1$ cp; the reservoir volume factor was taken to be unity.

Since the thickness is uniform and known, the rock properties to be estimated may be taken as the permeability and porosity. For

the purpose of numerical calculations the reservoir was divided by a uniform grid into 32 intervals; the 33 grid points were numbered 1 through 33 (Figure 1.4.1). The property values were taken as constant over a region extending half a mesh size on either side of each grid point; at either end point, this region is of only half a mesh size, extending into the reservoir. All the wells were located at grid points. This discretization results in 33 vectors, $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ of the unknown properties (N = 33). For simplicity the composite 66-vector $[\underset{\sim}{k}^T \mid \underset{\sim}{\phi}^T]^T$ will be represented by the symbol $\underset{\sim}{\pi}$.

1.4.2 Simulated Property Distributions

The rock property vectors $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ were simulated by random vectors with uniform mean values $\overline{k} = 5.0$ md, $\overline{\phi} = 0.2$ and covariances given by

$$E\left\{\frac{(k_i-\overline{k})(k_j-\overline{k})}{\overline{k}^2}\right\} = P_{ij} = f(0)e^{-|i-j|^2/s^2} \qquad (1.4.1)$$

$$E\left\{\frac{(\phi_i-\overline{\phi})(\phi_j-\overline{\phi})}{\overline{\phi}^2}\right\} = Q_{ij} = g(0)e^{-|i-j|^2/s^2} \qquad (1.4.2)$$

$$E\left\{\frac{(k_i-\overline{k})(\phi_j-\overline{\phi})}{\overline{k}\,\overline{\phi}}\right\} = R_{ij} = \rho f^{\frac{1}{2}}(0)g^{\frac{1}{2}}(0)e^{-|i-j|^2/s^2} \qquad (1.4.3)$$

where $f(0) = 0.0625$, $g(0) = 0.0625$, and $\rho = 0.5$. The covariance matrices in Equations (1.4.1 - 3) are similar to those of Equations (1.3.12 - 14), since for uniform thickness h, ϕ, and k are proportional to v and w. The normalization in Equations (1.4.1 - 3), while not essential, has been found computationally convenient and efficient. (See appendix 1.1.)

Using the above means and covariance matrices, several reali-

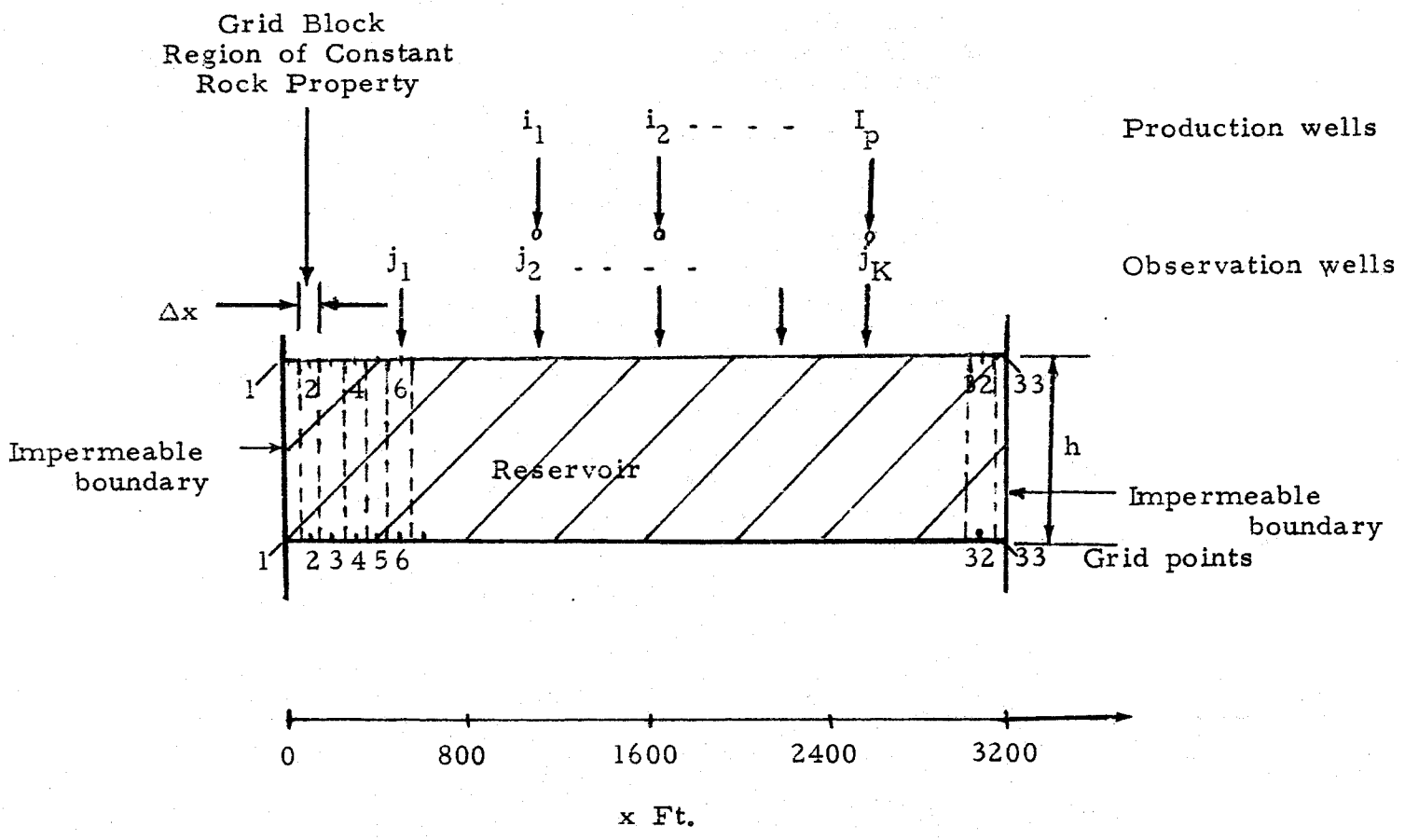


Figure 1.4.1 Discretization of One-dimensional Reservoir

zations of the random vectors $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ were generated numerically by means of a procedure described in appendix 1.1. Eight such realizations R1 - R8 are shown in Figure 1.4.2 . Some of these were used as the "true" property distributions in simulations, as described later.

1.4.3 Discretization of Pressure Equation

The partial differential equation (1.1.13) reduces to, for a one-dimensional reservoir with constant c, μ and uniform h ,

$$\phi(x) \frac{\partial p}{\partial t} = \frac{1}{\mu c} \frac{\partial}{\partial x} (k(x) \frac{\partial p}{\partial x}) + \frac{1}{ch} \sum_{i=1}^{I_p} q_i(t) \delta(x-x_i) \quad . \tag{1.4.4}$$

The boundary conditions are

$$\frac{\partial p}{\partial x} = 0 \qquad x = 0, L \qquad \text{for all } t \quad . \tag{1.4.5}$$

The initial condition was taken to be

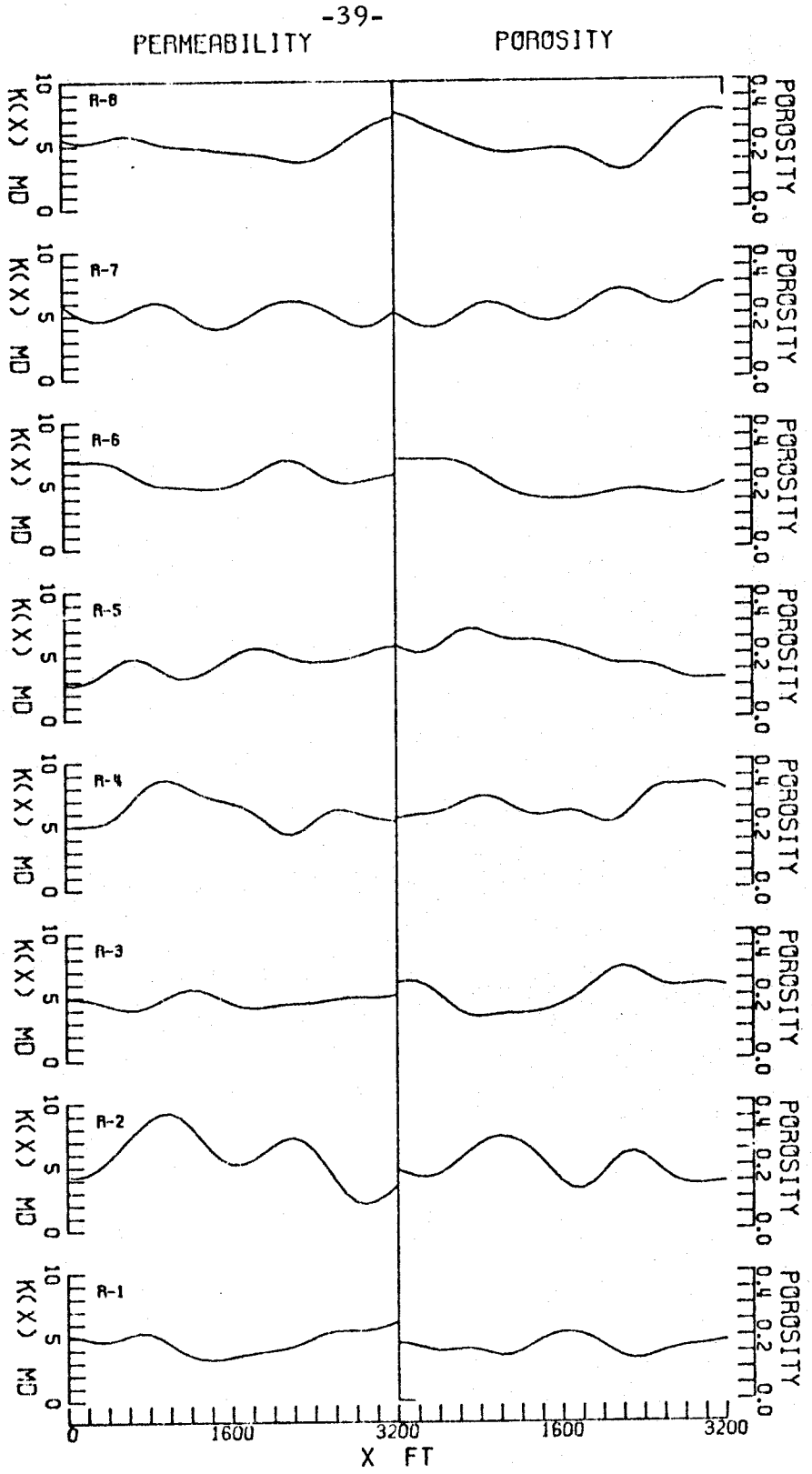
$$p(x, 0) = 0 \quad . \tag{1.4.6}$$

These equations were spatially discretized using the 33-point uniform grid and second order finite differences, resulting in a set of 33 coupled ordinary differential equations. These were further discretized with respect to time using the implicit Crank-Nicholson scheme (the trapezoidal rule) with uniform step size. The derivation and the resulting vector ordinary differential equations and the discrete algebraic equations are presented below.

Let the pressure distribution at t have the value $p_i(t)$ at the grid point i. Let us define the N-vector of these grid point values,

$$\underset{\sim}{p}(t) = [p_1(t), p_2(t), \ldots, p_N(t)]^T \quad . \tag{1.4.7}$$

The differential operator $\frac{\partial}{\partial x} (k(x) \frac{\partial p}{\partial x})$ at t and $x = x_i$, the location of the i^{th} grid point, can be approximated by the central dif-



8 REALIZATIONS OF HOMOGEN. RANDOM PROCESS
S=5.RHO=.5 MEANS=5MD..2 STD.DEVS=1.25MD..05

Figure 1.4.2

ference scheme,

$$\frac{\partial}{\partial x}\left(k(x)\frac{\partial p}{\partial x}\right) = \frac{\partial k}{\partial x}\frac{\partial p}{\partial x} + k(x)\frac{\partial^2 p}{\partial x^2}$$

$$= \frac{k_{i+1}-k_{i-1}}{2\Delta x} \cdot \frac{p_{i+1}(t)-p_{i-1}(t)}{2\Delta x}$$

$$+ k_i \frac{p_{i+1}(t)-2p_i(t)+p_{i-1}(t)}{(\Delta x)^2} \quad . \qquad (1.4.8)$$

It is easy to verify that provided $\frac{\partial^3 k}{\partial x^3}$ and $\frac{\partial^4 k}{\partial x^4}$ exist at $x = x_i$, the above approximation involves an error $O(\Delta x)^2$. The boundary condition (1.4.5) can be approximated to the same degree of accuracy as follows. Let $x_1 = 0$ and $x_N = L$. Then, (1.4.5) implies at $x = x_1$,

$$\frac{\partial}{\partial x}\left(k\frac{\partial p}{\partial x}\right) = k(1)\frac{\partial^2 p}{\partial x^2}\Big|_{x=x_1} \qquad (1.4.9)$$

Let us extend the domain of the reservoir beyond the boundary point x_1 by including a hypothetical point x_o at $x = -\Delta x$. Then $\frac{\partial p}{\partial x} = 0$ at $x = x_1$ yields the pressure at the hypothetical point,

$$p_o(t) = p_2(t) \quad . \qquad (1.4.10)$$

Then the finite difference approximation of the right hand side of (1.4.9) yields,

$$\frac{\partial}{\partial x}\left(k\frac{\partial p}{\partial x}\right)\Big|_{x=x_1} = k_1\frac{p_2-2p_1+p_o}{(\Delta x)^2} = 2k_1\frac{p_2-p_1}{(\Delta x)^2} \quad . \qquad (1.4.11)$$

It is easy to verify that if $\frac{\partial^4 p}{\partial x^4}$ exists at $x = x_1$, this approximation involves an error $O(\Delta x)^2$. A similar treatment can be applied to the boundary condition at $x = x_N$. Then, substituting these approximations in (1.4.4) and arranging the equation in a vector form we obtain,

$$\underset{\sim}{\Phi}\frac{d\underset{\sim}{p}}{dt} = \underset{\sim}{F}(k)\underset{\sim}{p}(t) + \underset{\sim}{\bar{q}}(t) \qquad (1.4.12)$$

where

$$\underset{\sim}{\Phi} = \text{diag}(\phi_1, \phi_2, \ldots, \phi_N) \tag{1.4.13}$$

$$\underset{\sim}{F}(\underset{\sim}{k}) = \frac{1}{(\Delta x)^2 \mu c} \begin{bmatrix} -2k_1 & 2k_1 & & \underset{\sim}{0} \\ & \ddots & \ddots & \\ & \frac{1}{4}(k_{i-1}+4k_i-k_{i+1}) & -2k_i & \frac{1}{4}(-k_{i-1}+4k_i+k_{i+1}) \\ & & \ddots & \ddots \\ & & 2k_N & -2k_N \end{bmatrix} \tag{1.4.14}$$

The elements of the vector $\underset{\sim}{\bar{q}}(t)$ are zero except those corresponding to the I_p grid points where the producing wells are located, and these are respectively given by $\frac{1}{ch\Delta x} q_i(t)$, $i = 1, 2, \ldots, I_p$. The initial condition is

$$\underset{\sim}{p}(0) = \underset{\sim}{0} \quad . \tag{1.4.15}$$

The original parabolic system has infinite eigenvalues varying from 0 to ∞. The vector ordinary differential equation system (1.4.12) has N eigenvalues, but the spread between the smallest and the largest (in absolute value) eigenvalue is usually very large. Such a system is called a stiff system; it is well known that explicit methods of time differencing require very small time steps for stability (Gear, 1971). On the other hand, the implicit scheme employing the trapezoidal rule is unconditionally stable and is of second order accuracy ($O(\Delta t)^2$ error), (Isaacson and Keller, 1966) and thus can be expected to yield results close to the true solution even for fairly large time steps. Hence, we select it for time discretization in the following.

Let pressure vector $\underset{\sim}{p}(t)$ at time $t = i\Delta t$ be denoted by $\underset{\sim}{p}_i$. Then the trapezoidal rule gives

$$\underset{\sim}{\Phi}\frac{\underset{\sim}{p}_{i+1}-\underset{\sim}{p}_i}{\Delta t} = \frac{\underset{\sim}{F}}{2}(\underset{\sim}{p}_{i+1}+\underset{\sim}{p}_i) + \tfrac{1}{2}(\bar{\underset{\sim}{q}}_{i+1}+\bar{\underset{\sim}{q}}_i) \quad . \tag{1.4.16}$$

After rearrangement we obtain

$$\underset{\sim}{G}\,\underset{\sim}{p}_{i+1} = \underset{\sim}{H}\,\underset{\sim}{p}_i + \underset{\sim}{q}_i \qquad i = 0, 1, 2, \ldots, T \tag{1.4.17}$$

where for convenience we have defined

$$\underset{\sim}{G} = \underset{\sim}{\Phi} - \frac{\Delta t}{2}\underset{\sim}{F} \tag{1.4.18}$$

$$\underset{\sim}{H} = \underset{\sim}{\Phi} + \frac{\Delta t}{2}\underset{\sim}{F} \tag{1.4.19}$$

$$\underset{\sim}{q}_i = \tfrac{1}{2}\Delta t(\bar{\underset{\sim}{q}}_{i+1} + \bar{\underset{\sim}{q}}_i) \tag{1.4.20}$$

In (1.4.17), we have assumed that the total period of interest is T time steps long. For the purpose of history matching, $t = T\Delta t$ may be the time of the last pressure observation. The initial condition is

$$\underset{\sim}{p}_o = \underset{\sim}{0} \quad . \tag{1.4.21}$$

We used $\Delta t = 1.75$ days in our numerical computation.

As discussed in section 1.2, we shall treat the set of discrete algebraic equations (1.4.17, 21) as the description for our system and derive finite-dimensional history matching algorithms for it.

1.4.4 Conditions of the Estimation Problem

Two sets of conditions for the property estimation problem were investigated using simulations. We shall refer to these as S_1 and S_2. The conditions of these sets are described in table (1.4.1) and in figures (1.4.3-4). As detailed therein, there are K observation sites where the pressures are observed; there are R time instants when the observations are taken simultaneously at all the sites.

Table 1.4.1: Conditions of Simulation

Set: S_1

K = Number of observation sites = 10

Observation sites situated at grid points 3, 6, 8, 11, 14, 17, 20, 23, 26, 29.

R = Number of observations at each site = 8

Observations at all sites are taken simultaneously, at intervals of 42 days, starting at t = 38.5 days.

I_p = Number of production wells = 1

Production well at grid point 17

$q(t)$ = Production rate = 0.16×10^{-2} STB/day/ft^2 constant over the observed history. (See figure 1.4.3)

Set: S_2

K = 3

Observation sites situated at grid points 8, 17, 25.

R = 25

Observations at all sites are taken simultaneously, at intervals of 14 days, starting at t = 10.5 days.

I_p = 2

Production wells at grid points 8, 17.

$q_i \cdot(t)$: Production rate history--piecewise--constant as shown in figure (1.4.4).

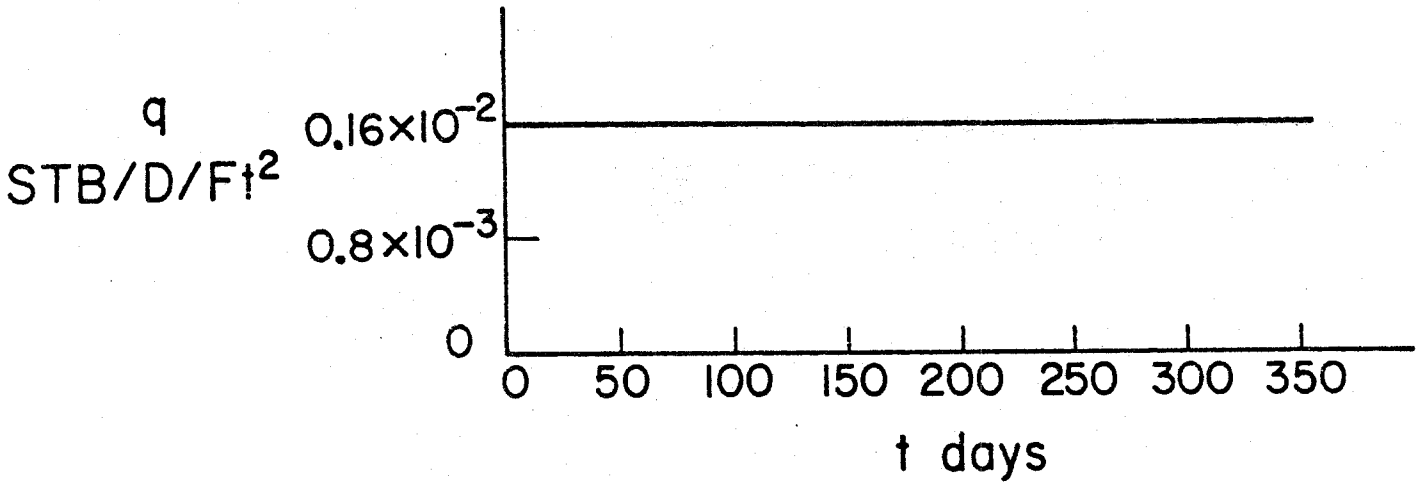
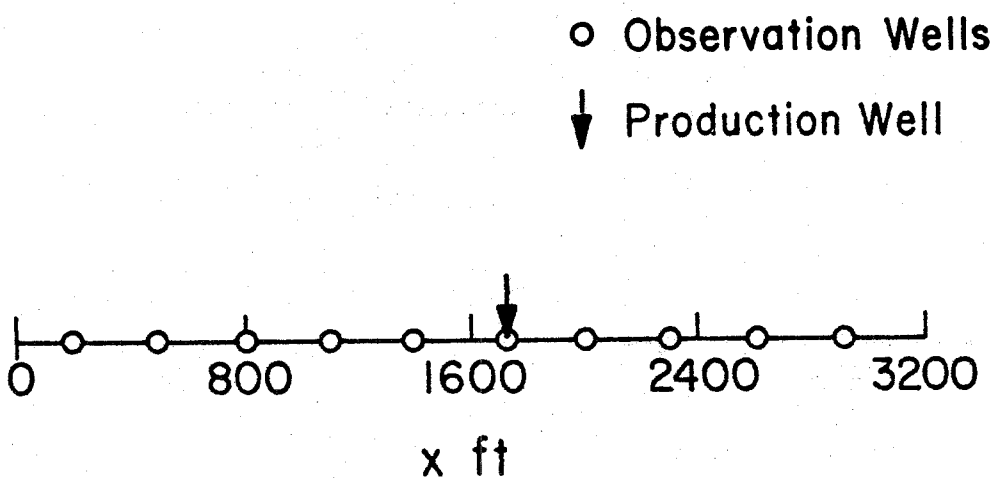
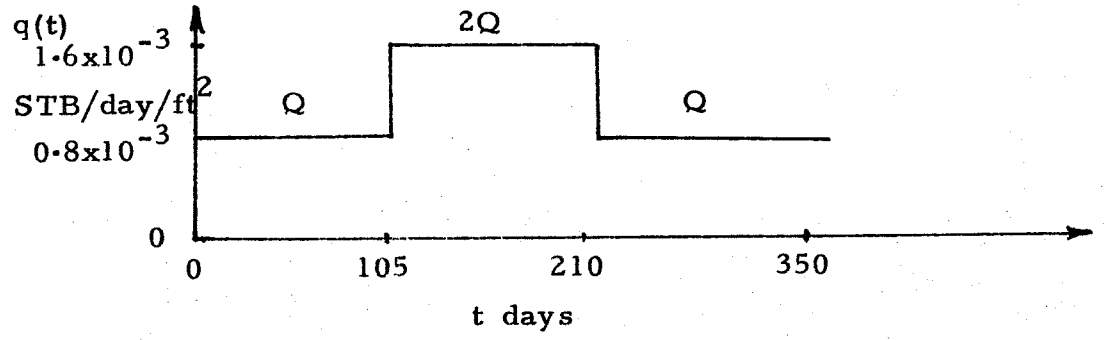
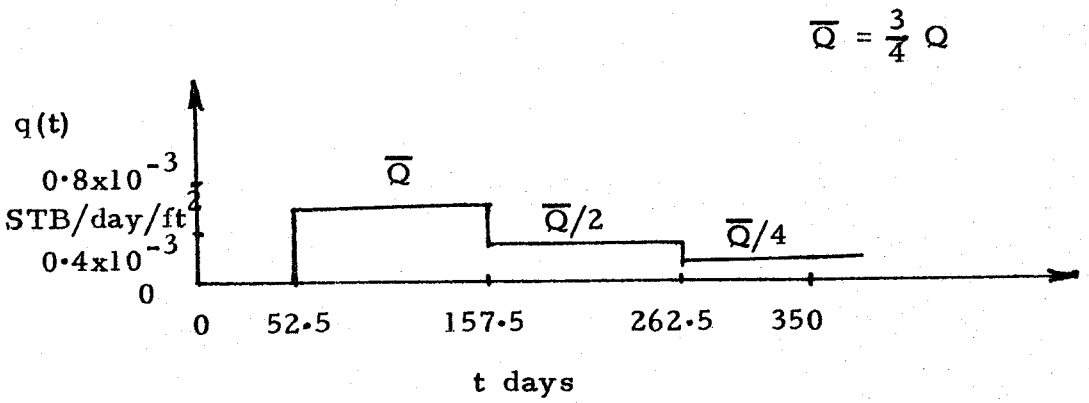


Figure 1.4.3
Conditions of Simulation
Set S_1.





Production from well at grid point 17.



Production from well at grid point 8.

Figure 1.4.4

Condition of Simulation for Set S_2.

Thus, there are a total of $L = RK$ observed pressures available for history matching. Under these circumstances we can write the summation in the expression (1.2.1) for J_p more explicitly as

$$J_p = \sum_{m=1}^{R} \sum_{n=1}^{K} \frac{1}{\sigma^2_{mn}} (p_{i_m, j_n} - p^o_{mn})^2 \qquad (1.4.22)$$

where $p_{i,j}$ is the calculated (model) pressure at time step i and grid point j ($= (p_i)_j$); p^o_{mn} is the observed pressure at the site located at j_n and at time step i_m.

For simulations with conditions S_1, four realizations R1 - R4 of the random process with the statistics given in subsection (1.4.2), and an additional non-random property distribution, R-NR, were used as the "true" property distributions. These five distributions are shown in figure (1.4.5). With set S_2, four realizations R1, R2, R3, R5 (see figure 1.4.2) were used for this purpose.

For each property distribution, equations (1.4.17, 21) were solved using the values of the "true" distribution for k and ϕ. The pressure histories obtained from the difference equations were corrupted by normally distributed measurement error with zero mean and 1 psi^2 variance to produce the simulated observations. Since the difference equations were used both for the generation of the simulated observations and for the subsequent estimation, the truncation errors due to discretization were excluded, and the results of estimation depended only on observation errors and on the inherent statistical structure of the problem. In real applications, several other types of errors would contribute to the final estimate errors. These include

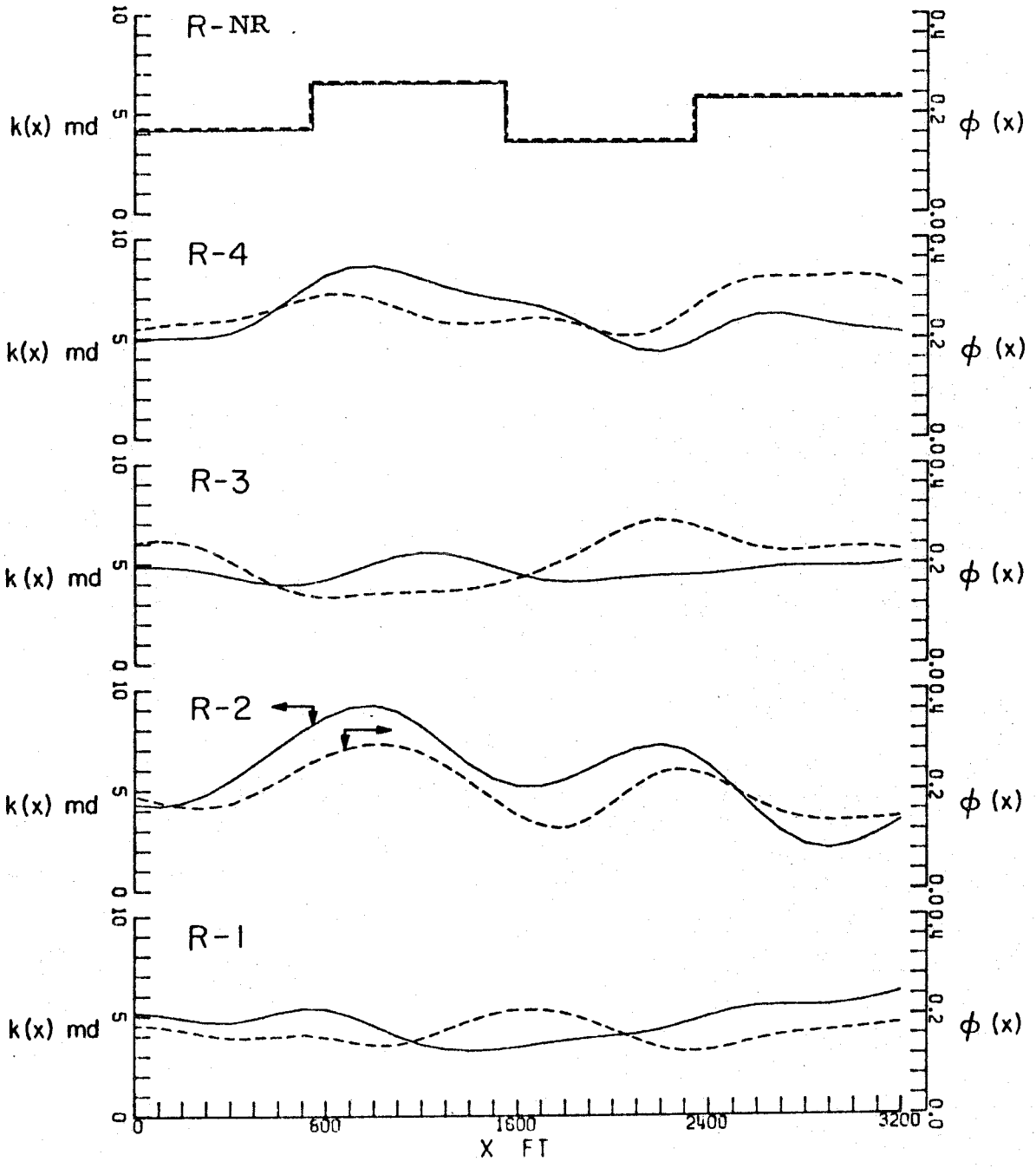


Figure 1.4.5

Property Distributions Used in Simulations With Conditions S_1.

errors in the production rates, discretization errors due to the finite differencing, and modeling errors of various types, for example, regarding the location of the reservoir boundaries. Since the object of our study is to compare the zonation and Bayesian approaches using various minimization algorithms, it is not necessary to include errors other than observation errors and errors arising because of the random nature of the unknown parameters.

1.4.5 Parameterizations Used in Simulation

The simulated property estimation problem for each property distribution and each set of conditions (S_1 or S_2) was numerically worked out using several different parameterizations, indicated below.

Firstly, the problem was treated with the zonation approach. Three different zonations involving different number of zones, NZ, with almost uniform zone size were attempted; in particular, the values NZ = 4, 8, 16 were used. Table (1.4.2) details the configuration of each of these zonations. Since we have two rock properties to be estimated, the number of parameters in zonation is $M \equiv 2(NZ)$.

The problem was also treated without any parameterization, leaving the values of all the elements of the 33-vectors $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ to be independently determined. This can be thought of as a special case of zonation with NZ = N(=33).

Lastly, the problem was worked out via the Bayesian approach. For this purpose, the normalized rock properties, $\underset{\sim}{k}/\overline{k}$ and $\underset{\sim}{\phi}/\overline{\phi}$, along with their covariances as defined in (1.4.1-3) were used. Thus, the Bayesian parameters are the components of $[(\underset{\sim}{k}-\overline{\underset{\sim}{k}})^T/\overline{k},\ (\underset{\sim}{\phi}-\overline{\underset{\sim}{\phi}})^T/\overline{\phi}]^T$ along M eigenvectors of their composite covariance matrix $\underset{\sim}{P}_o$ with

Table 1.4.2: Zonations Used in Simulations

No. of Zones (NZ)	Grid points included in the different zones
4	(1-8), (9-16), (17-24), (25-33)
8	(1-4), (5-8), (9-12), (13-16), (17-20), (21-24), (25-28), (29-33)
16	(1-2), (3-4), , (29-30), (31-33)
33	(1), (2), , (32), (33)

largest eigenvalues. Several values of the parameters s, β and M were used for a comparative study. The results of the simulations are described in section 1.6.

1.5 Minimization Algorithms

For generality of discussion, we shall denote the M-vector of parameters in all of the different parameterizations by $\underset{\sim}{\pi}$. From the context and the dimension of $\underset{\sim}{\pi}$, it will be clear which parameterization is under consideration.

Due to the non-quadratic and implicit dependence of J on $\underset{\sim}{\pi}$, it is necessary to carry out the minimization in an iterative manner. For this purpose, we start with a prior estimate or "initial guess" of the parameters and start the iterations. An iteration consists of determining corrections in the current estimate, which leads to a reduction in J. We employed two alternative algorithms for determination of the iterative corrections: the conjugate gradient algorithm and the Gauss-Newton algorithm or its variant called Marquardt's algorithm. In the present section we describe details of derivation of these algorithms and present some results regarding convergence characteristics in the framework of the minimization problems defined in sections 1.2 and 1.3. As discussed in section 1.2, we shall consider only finite-dimensional algorithms.

1.5.1 Conjugate Gradient Algorithm

This algorithm requires the gradient of J with respect to the parameter vector $\underset{\sim}{\pi}$ for determination of the iterative correction. We have in general,

$$\frac{\partial J}{\partial \underset{\sim}{\pi}} = \frac{\partial J_p}{\partial \underset{\sim}{\pi}} + \frac{\partial J_2}{\partial \underset{\sim}{\pi}}$$

$$= \sum_{m=1}^{K} \sum_{n=1}^{R} (p_{i_m, j_n} - p^0_{mn}) \frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} + \frac{\partial J_2}{\partial \underset{\sim}{\pi}} \qquad (1.5.1)$$

where we have used definition (1.4.22) of J_p. The quantities $\partial p_{i,j}/\partial \pi_\ell$ are called the sensitivity coefficients. They can be evaluated by solution of the sensitivity equations, which are derived from the system equations by differentiating them w.r.t. π_ℓ. Another method for the calculation of the sensitivity coefficients will be discussed in the next subsection. Both of these methods are detailed in appendix 1.4. The evaluation of the sensitivity coefficients requires very extensive computational effort. However, as shown in appendix 1.2, it is possible to evaluate $\partial J_p/\partial \underset{\sim}{\pi}$ more efficiently without recourse to the sensitivity coefficients by the formula,

$$\frac{\partial J_p}{\partial \pi_\ell} = \sum_{i=0}^{T-1} \underset{\sim}{\psi}_i^T \left[\frac{\partial \underset{\sim}{G}}{\partial \pi_\ell} \underset{\sim}{p}_{i+1} - \frac{\partial \underset{\sim}{H}}{\partial \pi_\ell} \underset{\sim}{p}_i \right] \qquad \ell = 1, 2, \ldots, M \qquad (1.5.2)$$

The sequence $\{\underset{\sim}{\psi}_i\}$ of the N-vectors in (1.5.2) is the solution of the forced initial value problem,

$$\underset{\sim}{G}^T \underset{\sim}{\psi}_{i-1} = \underset{\sim}{H}^T \underset{\sim}{\psi}_i + 2 \sum_{n=1}^{K} (p_{i_m, j_n} - p^0_{mn}) \delta_{i, i_m} \underset{\sim}{e}_{j_n} \qquad (1.5.3)$$

$$\underset{\sim}{\psi}_T = \underset{\sim}{0} \qquad (1.5.4)$$

where $\underset{\sim}{e}_{j_n}$ is the j_nth column of the (NxN) identity matrix.

The equations (1.5.3-4) are called adjoint system equations and $\{\underset{\sim}{\psi}_i\}$ are called adjoint variables. The derivation of the above formulation, based on a variational approach, is detailed in appendix

1.2. This method of calculation $\partial J/\partial \underset{\sim}{\pi}$ for the discrete system is a finite-dimensional analogue of a method for an infinite dimensional system (p.d.e model) devised by Chen et al. (1974) and Chevant et al. (1975).

Although it is possible to use expression (1.5.2) directly to compute the gradient for any parameterization, usually it is simpler, and more efficient from the computational standpoint, to first evaluate $\partial J/\partial \underset{\sim}{k}$ and $\partial J/\partial \underset{\sim}{\phi}$, and subsequently transform these into $\partial J/\partial \underset{\sim}{\pi}$. In appendix 1.2 we present the transformations appropriate for zonation and Bayesian parameterizations. Note that the adjoint equation can be separately integrated backward in time <u>after</u> the pressure equation is integrated forward.

The minimization scheme using the conjugate gradient algorithm is detailed in the flow chart in figure (1.5.1). In our computation, the conjugate search directions $\underset{\sim}{h}_i$ were determined using the Polak-Ribiere-Polyak-Sorensen formulation (Polak, 1973), as it is known to be generally superior to several other more commonly used formulations for the conjugate directions. The superiority of its performance in application to the reservoir parameter estimation problem was confirmed through numerical experimentation. The unidirectional search for the local minimum of J at each iteration is the most time consuming computational step, as it involves repeated solution of the system equation. Consequently, a careful comparison between several procedures was carried out. The procedure ultimately used is detailed in appendix 1.3.

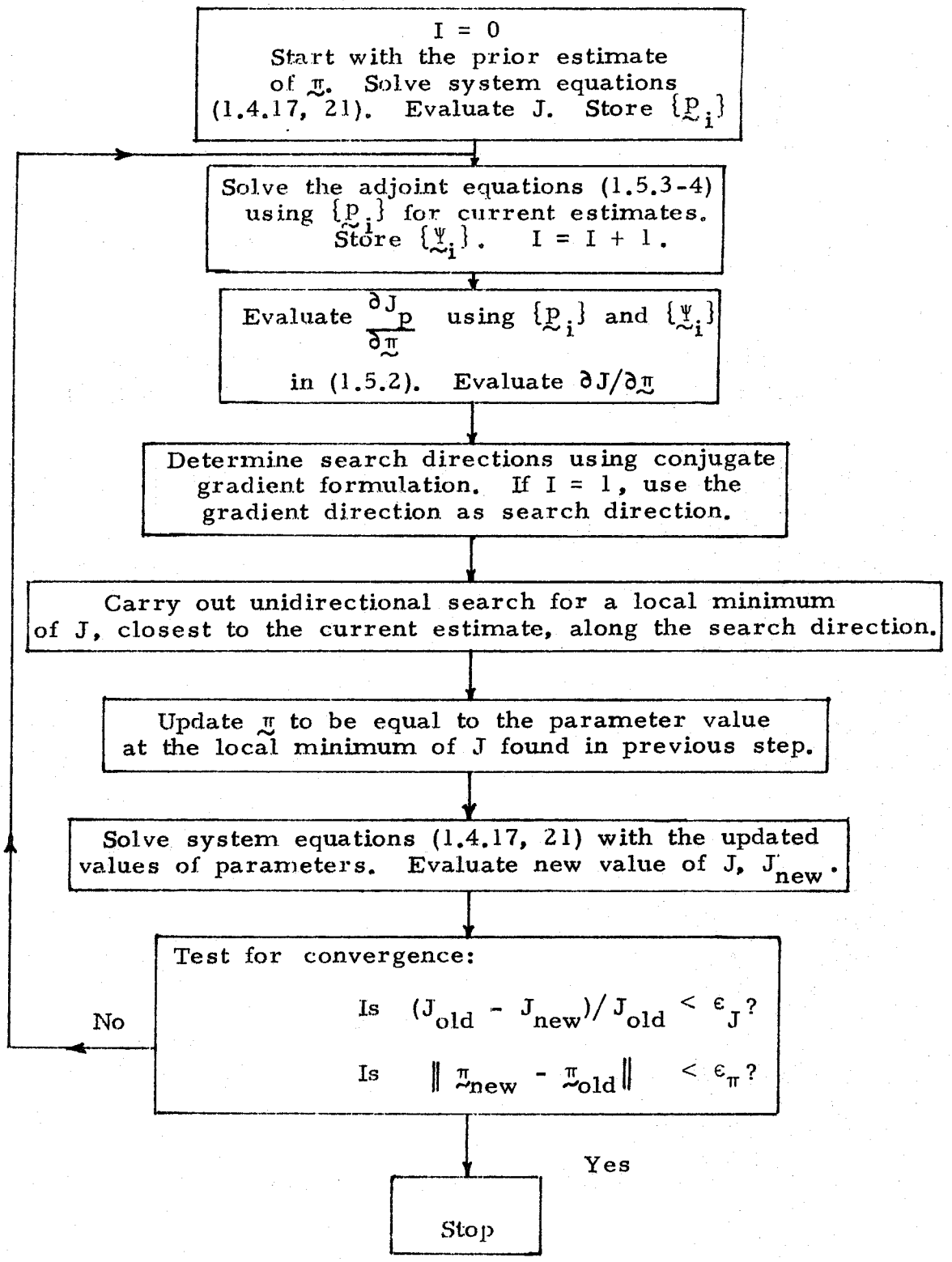


Figure 1.5.1 Conjugate Gradient Algorithm

1.5.1.1 Convergence and Estimates Resulting from Conjugate Gradient Algorithm

At first, the conjugate gradient algorithm was implemented using single-precision computation. The resulting convergence was very slow and the final values of J reached were usually much larger than the expected residual based on the observation errors. (Appendix 1.6 contains an analysis that leads to a probabilistic description of the residual value of J.) The numerical difficulty arises from large errors in the computed gradient $\partial J/\partial \underset{\sim}{\pi}$. The formula (1.5.2) for the gradient involves a repeated sum of the differences of almost equal numbers. This step leads to a loss of three to four significant digits and the single precision results are sometimes meaningless. This numerical difficulty prompted the use of double precision in all subsequent computation. Then the convergence characteristics improved substantially, and it was possible to reduce J to the expected residual level.

The reduction in J is very rapid during the first few iterations. However, as the minimum is approached, the convergence becomes slower. Figures (1.5.2a-d) illustrate typical convergence characteristics of the conjugate gradient algorithm. (Due to the very slow convergence near the minimum, it was found necessary to terminate the minimization according to some criterion. The computation was stopped when the fractional decrease in J in one iteration became smaller than a predetermined level, usually 5×10^{-4}.)

When sensitivities of the minimization index J with respect to some of the parameters are much smaller as compared to those of others, the numerical minimization is difficult. Determination of

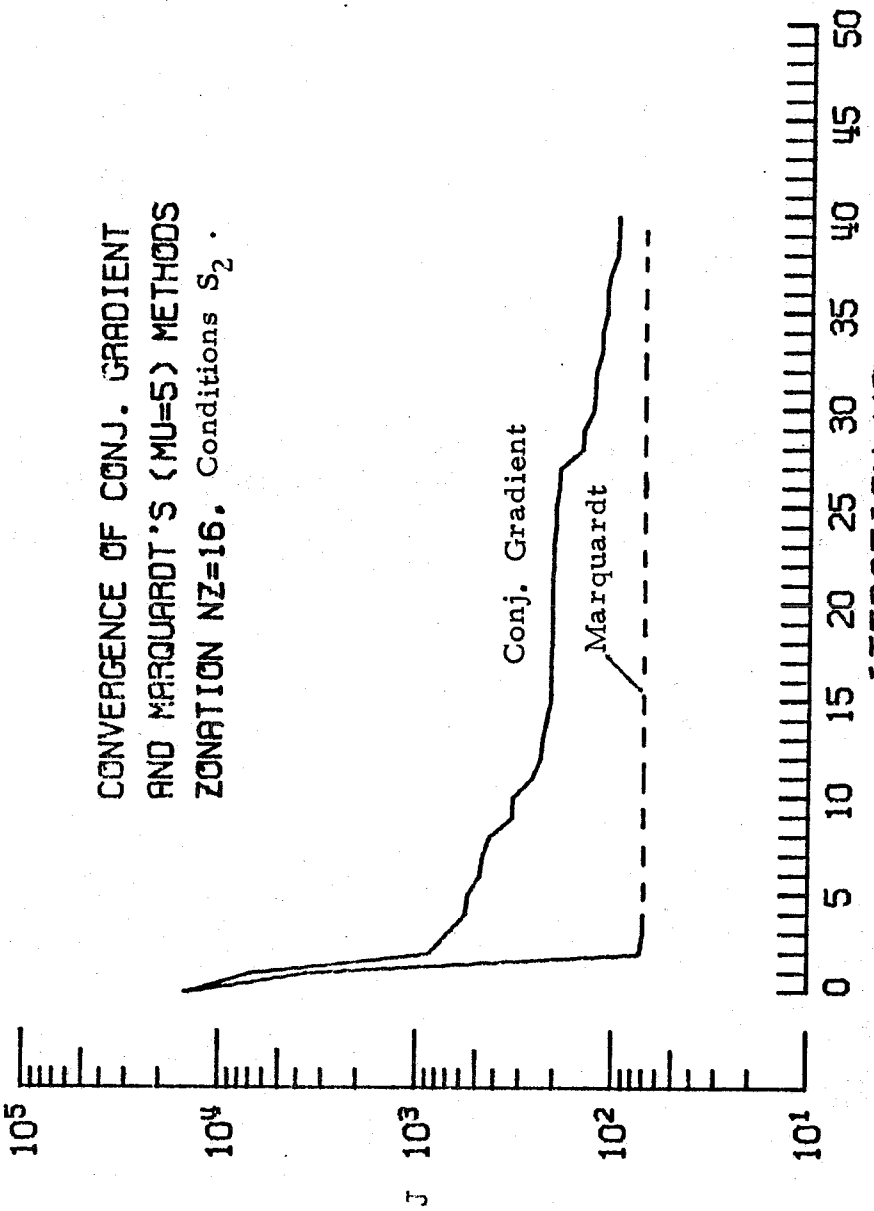


Figure 1.5.2a

Convergence of Minimization Algorithms

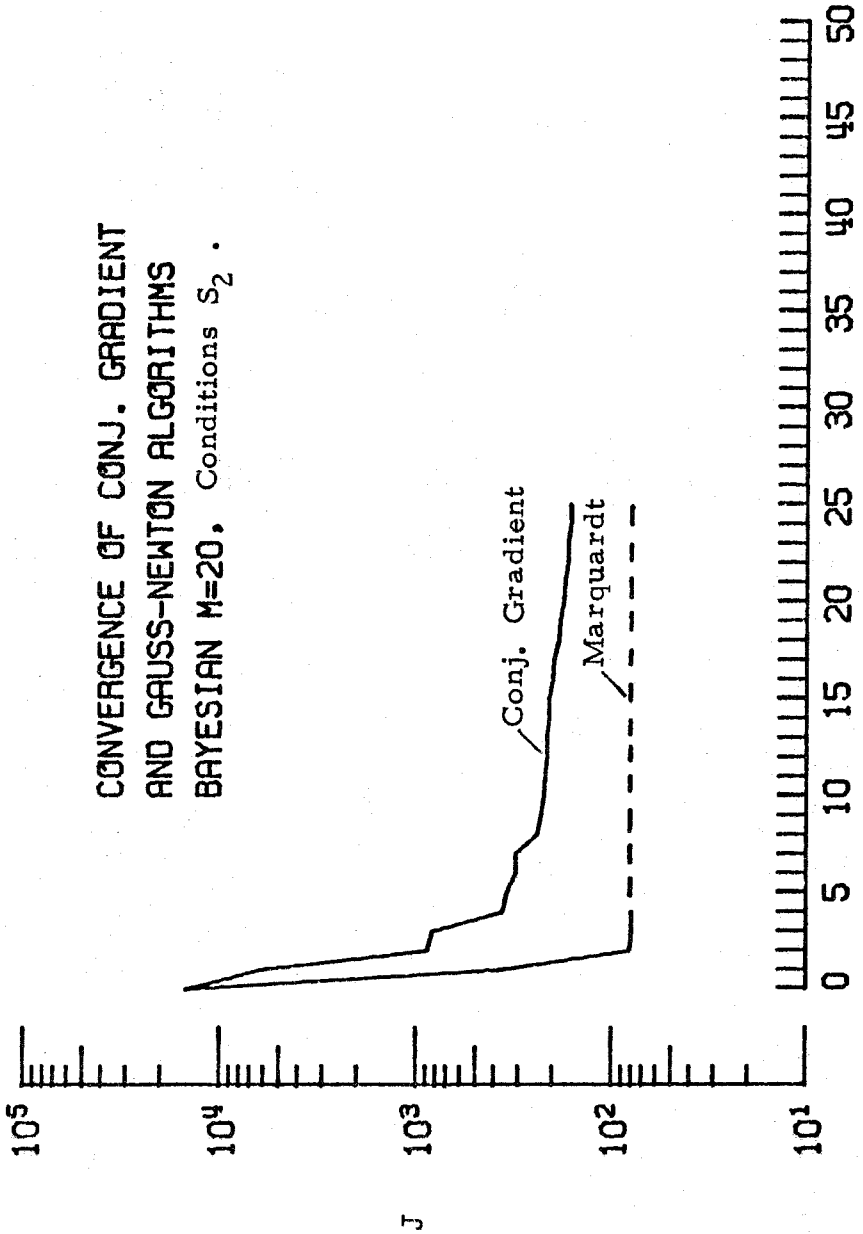


Figure 1.5.2b

Convergence of Minimization Algorithms

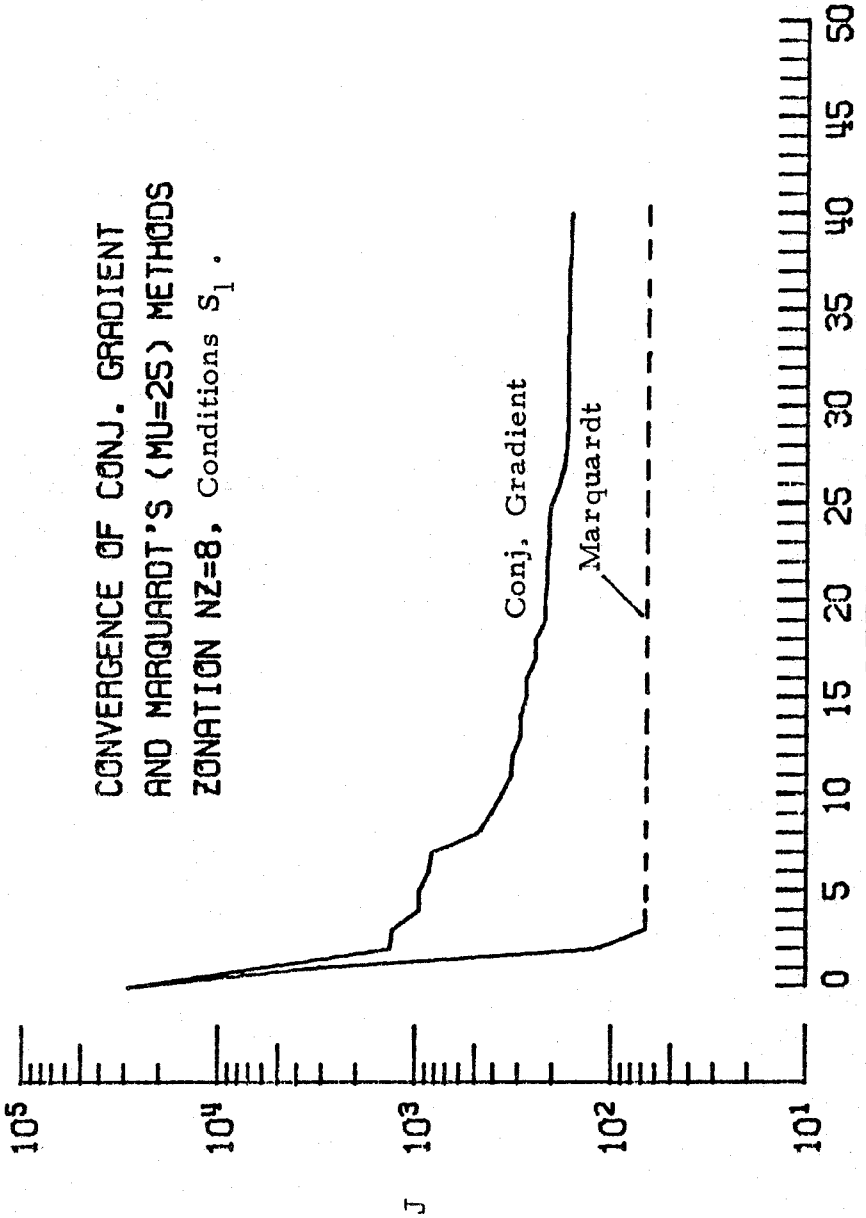


Figure 1.5.2c

Convergence of Minimization Algorithms

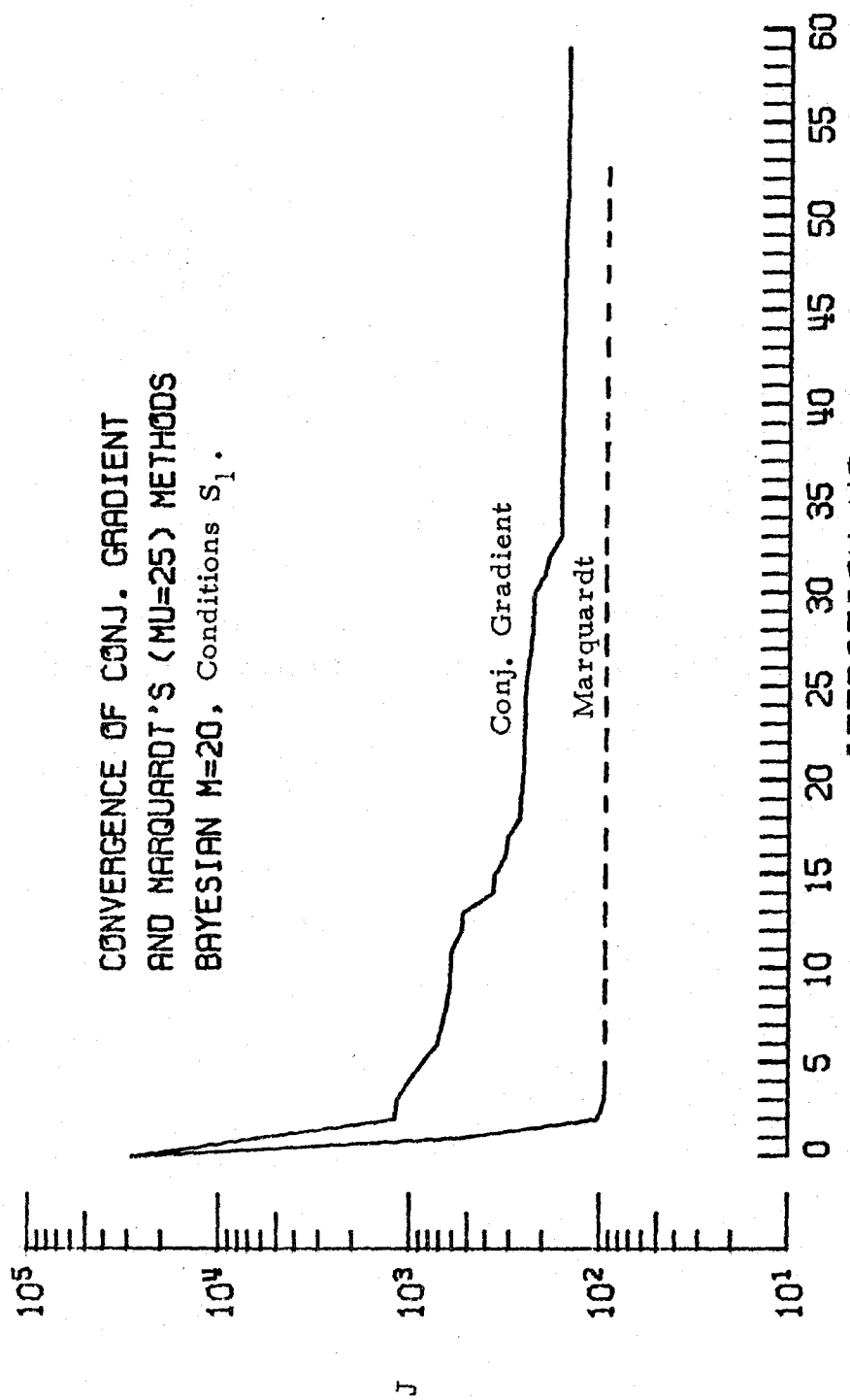


Figure 1.5.2d

Convergence of Minimization Algorithms

the minimum in the subspace of the less sensitive parameters is the difficult step. The value of J may be close to the minimum value and yet the estimates of these parameters may be far from the minimum. In these circumstances, the estimates numerically obtained depend on the algorithm used for minimization.

The first order gradient algorithms, a class to which the conjugate gradient algorithm belongs, make iterative corrections in the parameter estimates in the direction of the gradient $\partial J/\partial \underset{\sim}{\pi}$. The component of the gradient in a given parameter direction is proportional to the sensitivity of J w.r.t. that parameter. Thus, generally at each iteration, the components of the correction along the more sensitive parameter directions are dominant. Consequently, the estimates of the less sensitive parameters change little from their starting values and only the more sensitive parameters are estimated accurately.

The situation just described is invariably encountered in the reservoir parameter estimation problem. The permeability in regions away from the production and observation sites and especially in the regions close to impermeable boundaries where pressure gradients are small has little effect on the observations (figure (1.53)). Consequently, the final estimates of the permeability in such regions differ little from their initial estimates (see figure (1.5.4)).

In the case of porosity estimation, its mean value over the reservoir domain has a very large influence on the pressures. The mean ϕ determines the total volume and consequently the average density of oil (for any given cumulative production) in the reservoir. For small compressibility of the oil, the pressure is a very strong

function of the density. As a result, the pressure level at all the stages of production are strongly determined by the mean value of ϕ. The actual distribution of ϕ only weakly influences J. As a result, the estimate of the mean, $\overline{\phi}$, is very accurate but that of the actual distribution is poor, the estimated profile being quite flat around the mean. This feature is clearly illustrated by figure (1.5.4). (We note that these remarks apply only to the case of reservoir with impermeable boundaries. When the constant pressure boundary conditions exist, there will be intrusion of surrounding fluid in the reservoir and the mean, $\overline{\phi}$, does not influence the pressure levels to as great an extent. We present a more detailed discussion of sensitivities in chapter 2.)

The convergence characteristics of the conjugate gradient algorithm are also influenced by the large range of parameter sensitivities. This range is indeed very large in zonation approach as the zonation parameters represent the values of k and ϕ in local regions of the reservoir. Consequently, estimates only of the zonation parameters representing high sensitivity regions are corrected. This limits the effective dimensionality of search space, the subspace of the parameter space where unidirectional search is carried out. For a given problem, parameterizations using different numbers of zones result in approximately equal effective dimensionality of the search space. The conditions that lead to this situation are somewhat complex, involving two opposing influences; we discuss the details in subsection 1.5.1.2. Furthermore, for different zonations, even among the more sensitive parameters, the sensitivities are widely

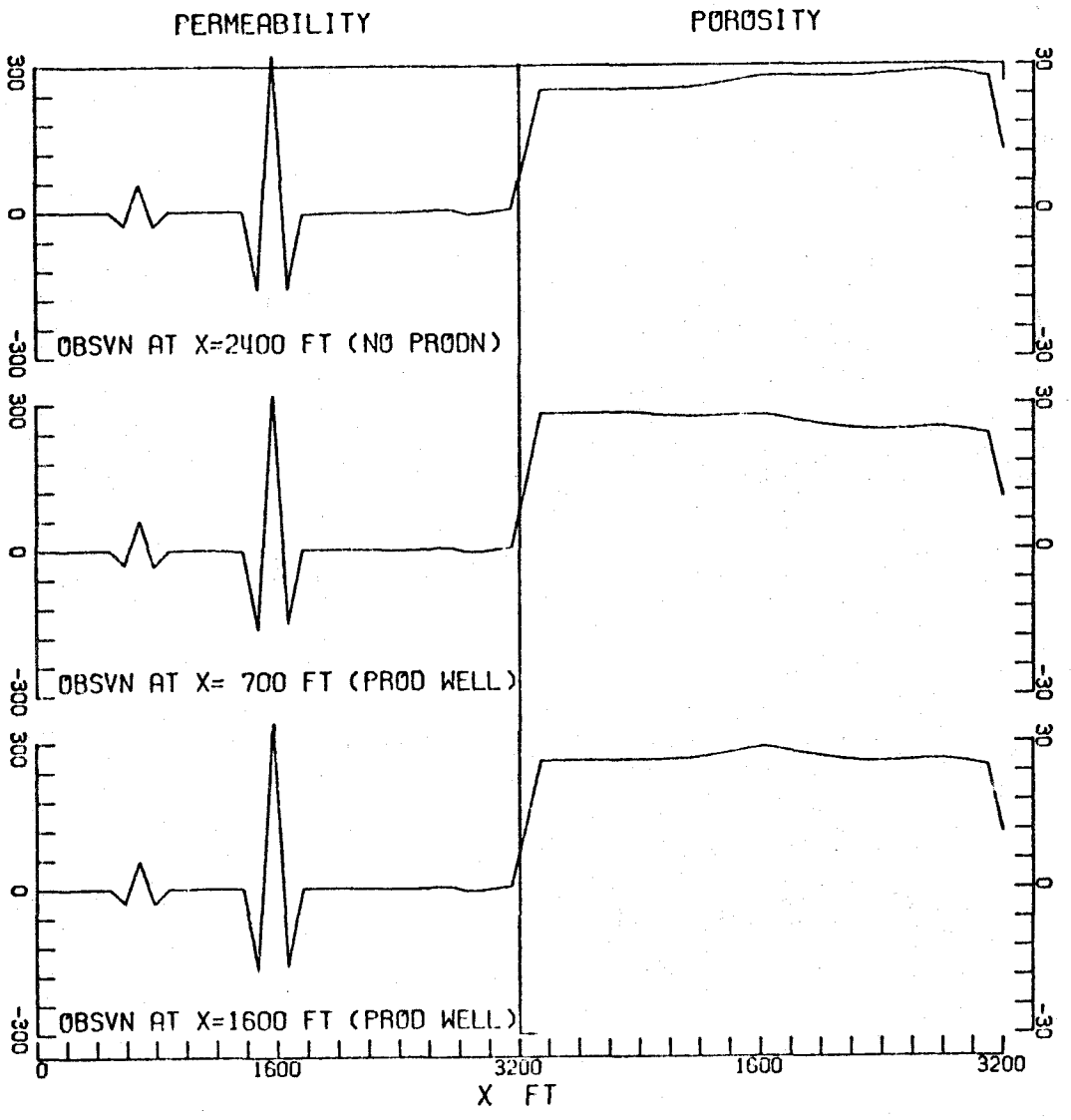
spread. As a result, the hypersurface of J, when viewed in the subspace of these parameters, has very long narrow valleys. Under these conditions the first order gradient algorithms are known to perform poorly even when the number of parameters is relatively small, and a large number of iterations are required for minimization (Bryson and Ho, 1969). Consequently, the number of iterations required for achieving a given reduction in J did not change appreciably as the number of zones ranged over the set $\{4, 8, 16, 33\}$.

The convergence properties are more favorable in the case of Bayesian estimation. The Bayesian parameters represent the values of k and ϕ over the whole reservoir rather than in any local region; this causes them to have more uniform sensitivities. Furthermore, the Bayesian penalty, J_2, limits the size of the corrections along the less important directions (more spatially oscillatory components of the distributions) to a minimum. Consequently the conjugate gradient algorithm in the Bayesian approach requires fewer iterations for a given reduction in J as compared to zonation. When additional parameters are introduced in the Bayesian estimation, the situation is not changed significantly. The new parameters are the components of $(\underset{\sim}{k}-\underset{\sim}{\bar{k}})^T \mid (\underset{\sim}{\phi}-\underset{\sim}{\bar{\phi}})^T)$ along the eigenvectors of $\underset{\sim}{P}_o$ with smaller eigenvalues; as evidenced by expression (1.3.32), nonzero values of these parameters are most heavily penalized, resulting in very little participation by the new parameters. Thus, the effective dimensionality of the search space does not increase with the number of parameters M beyond a certain limit, and the number of iterations required for the Bayesian estimation does not change significantly with M.

1.5.1.2 Zonation Parameter Sensitivities and Convergence

In a given property estimation problem, zonation has a strong influence on the sensitivities of the resulting parameters. There are two possibilities having opposite influence, and depending on zonation and region of the reservoir, either of them may occur. For example, if two or more grid points with property sensitivities, which are large and of the same sign, are lumped into a single zone, the resulting zonation parameter has a very large sensitivity. On the other hand, if the grid points have very large sensitivities with opposite signs, these may cancel each other, resulting in a small sensitivity of the zonation parameter.

For illustration, we plot in figure (1.5.3) $(\partial p_{i_m, j_n} / \partial k_\ell)$ and $(\partial p_{i_m, j_n} / \phi_\ell)$ for the same observation time i_m, and three different observation locations j_n, for conditions of the set S_2. The association of an element of $\underset{\sim}{k}$ or $\underset{\sim}{\phi}$ with a region in the one-dimensional reservoir is utilized for the identification of the abscissae. The sensitivity with respect to k is large near a production well (for example, at x = 1600 ft), but its sign changes from positive to negative as x changes from the well location to an adjoining grid point on either side. For distances from the producing well larger than one grid spacing, the k-sensitivities are very small in absolute value. A zonation permeability parameter that lumps the well location with one grid point on either side will have a net sensitivity much smaller than the individual grid point sensitivities due to cancellation. The opposite situation occurs when many contiguous grid points away from the wells are lumped into a single zone; the sensitivities have the same sign at



SENSITIVITY OF OBSERVED PRESSURES

Figure 1.5.3

all the grid points and thus the resulting zonation parameter has a higher sensitivity.

In the former case, zonation entails a severe loss of parameter sensitivities, resulting in a poor convergence. On the other hand, in the latter case the zonation improves sensitivities and aids convergence. Thus there are opposing influences of zonation on convergence of the conjugate gradient algorithm.

1.5.2 Gauss-Newton and Marquardt Algorithm

The Gauss-Newton minimization algorithm has been applied by several authors to history matching (Jahns (1966), Thomas et al. (1972)). In this algorithm, Newton's method (Issacson and Keller (1966)) is used to find the "zero" of $\partial J/\partial \underset{\sim}{\pi}$, but the second derivative $\partial^2 J/\partial \underset{\sim}{\pi}\, \partial \underset{\sim}{\pi}$ required for this purpose is evaluated approximately. The approximation to the second derivative is obtained in terms of the first derivatives $\partial p_{i_m, j_n}/\partial \underset{\sim}{\pi}$; details of the approximation are contained in appendix 1.5. The resulting iterative correction for the general problem including J_2 is,

$$\underset{\sim}{\pi}^{\ell+1} = \underset{\sim}{\pi}^{\ell} - \left[\sum_{m=1}^{R} \sum_{n=1}^{K} \frac{2}{\sigma^2_{mn}} \left(\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} \right)\left(\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} \right)^T + \frac{\partial^2 J_2}{\partial \underset{\sim}{\pi}\, \partial \underset{\sim}{\pi}} \right]^{-1}_{\underset{\sim}{\pi}^{\ell}} \left(\frac{\partial J}{\partial \underset{\sim}{\pi}} \right)_{\underset{\sim}{\pi}^{\ell}} \qquad (1.5.5)$$

where $\underset{\sim}{\pi}^{\ell}$ is the estimate after the ℓ^{th} iteration. The chief advantage of this method is that it requires a much smaller amount of computation as compared to the exact second-order Newton-Raphson method, while maintaining the convergence characteristics of the latter in the

vicinity of the minimum. It has the disadvantage that it may diverge even when the exact second derivative of J is positive-definite. The divergence more commonly occurs when the approximation to the second derivative is poor. Such circumstances arise in history matching when the prior estimates are much different from the "true" values and the associated history match errors are large.

When the term J_2 is absent (e.g. in zonation) or small, the reservoir estimation problem is generally ill-conditioned; this is reflected in this algorithm in the form of a nearly singular Hessian matrix in expression (1.5.5). In these circumstances, the behavior of the correction in (1.5.5) becomes unstable and the iterative process may diverge. Marquardt's modification remedies this problem by adding a positive definite matrix $\mu \underset{\sim}{I}$ to the Hessian before inverting it. This result can be arrived at rigorously by constraining the size of iterative correction $\|\underset{\sim}{\pi}^{\ell+1} - \underset{\sim}{\pi}^{\ell}\|$ to be equal to a given constant. Appendix 1.5 contains the details of this derivation. Systematic methods for determination of μ are available (appendix 1.5, Eykhoff (1974)), but are computationally expensive; hence a value of μ that led to satisfactory convergence for a given simulation was determined by trial and subsequently was held fixed. In chapters 2 and 3 we shall approach the question of determination of μ from an altogether different standpoint.

The value of μ used determines the convergence characteristics of Marquardt's algorithm. In a given problem convergence is guaranteed for sufficiently large μ. Large values of μ yield slow convergence because, as discussed in appendix 1.5, Marquardt's algorithm

approaches the gradient algorithm. On the other hand, as $\mu \rightarrow 0$ the Gauss-Newton algorithm is approached. Thus for intermediate value of μ Marquardt's algorithm offers a compromise between convergence rate and insurance against divergence. Typical convergence characteristics of Marquardt's method are depicted in figures (1.5.2a-d). As the minimum is approached, the Gauss-Newton approximation becomes more accurate and the convergence rate improves.

In the application of Marquardt's algorithm to zonation (J_2=0), large values of μ yield results akin to those of the gradient algorithm, i.e. the resulting property distributions are less oscillatory and do not differ significantly from the prior estimates in the regions of low sensitivity. On the other hand, small values of μ produce large corrections along the insensitive parameter directions(corresponding to smaller eigenvalues of the Hessian). Therefore, the final estimates are spatially oscillatory as shown in figure (1.5.4), and involve large estimate errors as indicated in table (1.5.1). Further analysis of this effect will be carried out in chapter 2.

In the Bayesian estimation, the term $\partial^2 J_2 / \partial \underset{\sim}{\pi} \partial \underset{\sim}{\pi}$ in the Hessian is positive definite, and exerts a significant stabilizing influence on the Gauss-Newton iterations. Thus, addition of the term $\mu \underset{\sim}{I}$ was found to be not always necessary for convergence. Furthermore, when Marquardt's modification is employed, the value of μ does not have significant influence on the nature of the corrections, as the more oscillatory components are considerably suppressed by the penalty term J_2.

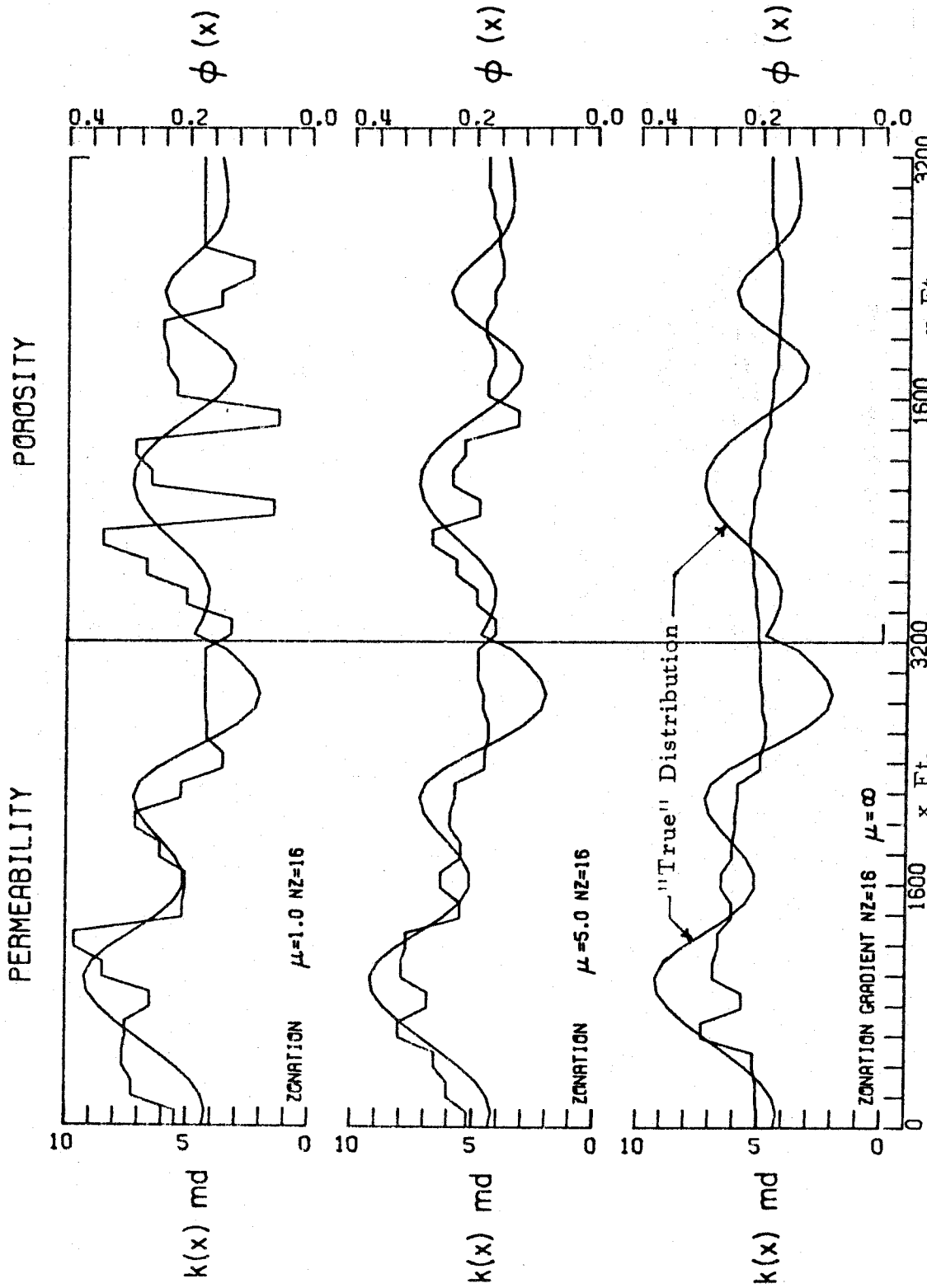


Figure 1.5.4 Final Property Estimates for Different Values of Marquardt Parameter μ.

Table 1.5.1

Effect of Minimization Algorithm

Property Distribution: R-2. Conditions of Estimation Problem: S_2

Initial Values: J_p = 14,900; J_K^* = 11.73; J_ϕ^* = 7.22

Zonation: No. of zones = 16

Algorithm	Final Values		
	J_p	J_K	J_ϕ
Gradient ($\mu = \infty$)	98.9	8.39	6.82
Marquardt (μ = 5.0)	70.5	7.52	7.04
Marquardt (μ = 1.0)	69.8	8.65	12.49

---

* J_K and J_ϕ are defined by,

$$J_K = \frac{1}{\bar{k}} \sum_{i=1}^{N} (k_i^{est} - k_i^{true})$$

$$J_\phi = \frac{1}{\bar{\phi}} \sum_{i=1}^{N} (\phi_i^{est} - \phi_i^{true})$$

Setting aside the question of computational time, a comparison between the conjugate gradient and Marquardt's algorithms, in terms of the errors J_p, J_k and J_ϕ, favors the latter algorithm as shown in table (1.5.2). The computational time requirements are discussed in detail in subsection 1.5.3.

1.5.2.1 Calculation of Sensitivity Coefficients

There are two options available for computing the derivatives, $\partial p_{i_m, j_n} / \partial \underset{\sim}{\pi}$. First, one may simply solve the sensitivity equations obtained by differentiating the original pressure equations with respect to each of the M unknown parameters. This method is detailed in appendix 1.4. Second, the calculation of $\partial p_{i_m, j_n} / \partial \underset{\sim}{\pi}$ can be carried out by evaluating the formula,

$$\frac{\partial p_{i_m, j_n}}{\partial \pi_\ell} = - \sum_{i=0}^{i_m - 1} \underset{\sim}{\psi}_i^T \frac{\partial \underset{\sim}{G}}{\partial \pi_\ell} \underset{\sim}{p}_{i+1} + \sum_{i=1}^{i_m - 1} \underset{\sim}{\psi}_i^T \frac{\partial \underset{\sim}{H}}{\partial \pi_\ell} \underset{\sim}{p}_i \qquad (1.5.6)$$

where the N-vector sequence $\{\underset{\sim}{\psi}_i\}$ is the solution of the initial value problem,

$$\underset{\sim}{G}^T \underset{\sim}{\psi}_{i-1} = \underset{\sim}{H}^T \underset{\sim}{\psi}_i \quad i = 0, 1, 2, \ldots, i_m - 1 \qquad (1.5.7)$$

$$\underset{\sim}{G}^T \underset{\sim}{\psi}_{i_m - 1} = \underset{\sim}{e}_{j_n} \qquad (1.5.8)$$

where $\underset{\sim}{e}_{j_n}$ is the j_nth column of the (NxN) identity matrix. Again, (1.5.7-8) are called adjoint equations and $\{\underset{\sim}{\psi}_i\}$ are called adjoint variables. This is an exact formulation for the sensitivity coefficients of the grid point pressures determined by the discrete equations (1.4.17, 21). We present the derivation of this formulation, based on a simple variational approach, in appendix 1.4. This formulation is a discrete analogue of those obtained by Jacquard and Jain (1965)

Table 1.5.2. Comparison of Conjugate Gradient and Marquardt's Algorithms.

Property Distribution: R-2

Minimization algorithm	Initial J_p	Final J_p	Final J_k	Final J_ϕ	No. of Iterations	Computer time, sec. *
Zonation (No. of zones = 16) 3 obs. wells; 2 prod. wells, no. of obs. = 75. (S_2)						
Conjugate gradient	0.149×10^5	98.9	8.39	6.82	40	136
Marquardt's (μ=5)	0.149×10^5	70.5	7.24	6.85	4	124
Bayesian M = 20, β = 1, s = 5, 3 obs. wells, 2 prod. wells, no. of obs. = 75. (S_2)						
Conjugate gradient	0.149×10^5	163	7.73	6.11	25	106
Gauss-Newton (μ=0)	0.149×10^5	79.8	7.51	5.73	4	121
Zonation (No. of zones = 8) 10 obs. wells, 1 prod. well, no. of obs. = 80. (S_1)						
Conjugate gradient	0.285×10^5	161	5.86	6.74	40	151
Marquardt's (μ=25)	0.285×10^5	67.0	6.37	9.18	4	129
Bayesian M = 20, β = 1, s = 5, 10 obs. wells, 1 prod. well, no. of obs. = 80. (S_1)						
Conjugate gradient	0.285×10^5	125	5.84	6.26	59	223
Marquardt's (μ=25)	0.285×10^5	69.7	3.07	4.60	5	177

* Computer time reported is for IBM 370-158.

for linear ordinary differential equation model, and by Carter et al. (1974) for linear partial differential model of the reservoir. Although our results are analogous to those of others, the method of deviation in appendix 1.4 is more general and applicable to a wide variety of problems—those described by linear or nonlinear discrete equations, ordinary differential equations, and partial differential equations (hyperbolic or parabolic)—whereas the derivations by Jacquard and Jain or Carter et al. are not easily extended to other problems. (See section 2.4 for more discussion.) Thus, in this method we have a powerful tool. In chapter 2 we shall present its application to a partial differential equation description of the reservoir.

We note that in the case of time invariant systems ($\underset{\sim}{G}$, $\underset{\sim}{H}$ independent of t), the adjoint equation needs to be integrated only <u>once</u> for each observation site with $i_m = T$. This solution can be used, after appropriate changes in the subscripts, to evaluate the sensitivity coefficients of pressures at <u>all</u> the observation times at that location. Thus, when the observations are available from K different sites, K integrations of the adjoint equations are required, one for each site with a different initial condition. This results in a very important saving of the computational effort.

As mentioned earlier in connection with the calculation of the gradient $\partial J/\partial \underset{\sim}{\pi}$, it is simpler and usually more efficient to first calculate the derivatives $\partial p_{i_m, j_n}/\partial \underset{\sim}{k}$ and $\partial p_{i_m, j_n}/\partial \underset{\sim}{\phi}$, and from these evaluate $\partial p_{i_m, j_n}/\partial \underset{\sim}{\pi}$ subsequently for Bayesian or zonation parameterization. Details of these transformations are contained in appendix 1.4.

Furthermore, use of double-precision arithmetic is recommended for the same reasons as before.

We compare the computational effort required by the two methods of sensitivity coefficient calculation in subsection 1.5.3.

1.5.3 Computational Time Requirements

We have just discussed two numerical minimization algorithms commonly used in parameter estimation problems. From the point of view of implementation the principal question is: which algorithm should be employed in a specific problem? The total computational effort required to obtain convergence in a parameter estimation problem depends on the number of iterations and the time required per iteration. The number of iterations is difficult to predict for any given method and problem, whereas the time required per iteration can be exactly determined. In general, the first-order methods, such as the conjugate gradient method, converge slowly but require little computation per iteration. The quasi-second-order methods, the Gauss-Newton and Marquardt, converge more rapidly but require more computation per iteration than the first-order methods.

It is useful to attempt to estimate the computational effort per iteration required in each of the three minimization methods as a function of the number of unknown parameters, the number of data points, and the number of mesh points employed in the numerical solution of the pressure equation. With such an estimate it is then possible to select the most efficient method to use in a particular problem. As noted above, the number of iterations required for convergence cannot be predicted, although from computational experience

one can estimate the relative number of iterations required by different methods for the same problem. For example, our experience has indicated that the conjugate gradient method requires about ten times as many iterations for convergence as the Gauss-Newton or Marquardt methods. On this basis the choice of the conjugate gradient method over the Gauss-Newton or Marquardt methods would require it to be at least ten times faster per iteration than the other two methods.

We will take the number of multiplications required for each method as a measure of the computational effort per iteration. (The number of additions is approximately the same as the number of multiplications for each of the methods and therefore will not be considered.)

We shall analyze the discrete versions of the two minimization algorithms in conjunction with two types of parameterization, zonation and Bayesian. In each of the four cases we shall report, for both one-dimensional and two-dimensional reservoirs, estimates of the number of multiplicative operations per iteration required for the discrete formulation of the history matching problem. We shall consider both of the two alternative procedures available for calculation of the sensitivity coefficients required in the Gauss-Newton algorithm.

The discrete equations(1.4.17,21) are assumed to model one-dimensional reservoirs. We assume that the implicit equations at each time step are solved via L-U decomposition of G obtained by direct factorization of the tridiagonal matrix (Isaacson and Keller, 1966). Discrete pressure equations obtained by application of the alternating direction implicit (ADI) method (Peaceman and Rachford,

1955) are assumed to model the two-dimensional reservoir (See appendix 1.7 for details of discretization). Again, at each time step for each row of grid points, the implicit equations will be assumed to be solved by L-U decomposition.

The considerations involved in arriving at the estimates of computational effort for all the cases are detailed in appendix 1.7. Since the ratio of effort involved in two alternative algorithms is important while making a choice between them, we present the results in the form of such ratios.

Let us define τ_1 as the ratio of the number of multiplications per iteration in the conjugate gradient algorithm to that in the Gauss-Newton (or Marquardt) algorithm with the adjoint equation solution for the sensitivities. Then we also define τ_2 as the same ratio, but with direct integration of the sensitivity equations in the Gauss-Newton (or Marquardt) method. Table (1.5.3) shows the expressions for τ_1 and τ_2 for a one-dimensional reservoir for the zonation and Bayesian approaches; and table (1.5.4) shows ratios for a two-dimensional reservoir.

The parameter S represents the number of times it is necessary to solve the pressure equation while carrying out the unidirectional search, for the minimum of J along the search direction, at each iteration of the conjugate gradient algorithm. Its value depends on the search algorithm used and the degree of accuracy of the search. The accuracy of the search also influences the number of iterations required for convergence; a greater accuracy will reduce the number of iterations. However, quantitative prediction of this is not possible

Table 1.5.3: Ratios τ_1 and τ_2 for One-Dimensional Reservoir

	Zonation Approach	Bayesian Approach
τ_1 sensitivities computed via adjoint variables	$\frac{(6S + 16)NT}{\{2(3K + L + 3)NT + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML\}}$	$\frac{(6S + 16)NT + 2MN}{\{2(3K + L+3)NT + \frac{1}{2}M^2L + 2NLM + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML\}}$
τ_2 sensitivity equations used directly	$\frac{(6S + 16)NT}{\{(6M + 14)NT + 4MT + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML\}}$	$\frac{(6S + 16)NT + 2MN}{\{6(2M +1)NT + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML\}}$

N = number of mesh points in spatial grid

T = number of time steps in pressure equation integration over observed history

K = number of observation locations

M = number of unknown parameters

S = number of pressure equation solutions in one-dimensional search for local minimum of J

L = total number of data points

Table 1.5.4: Ratios τ_1 and τ_2 for Two-dimensional Reservoir

	Zonation Approach	Bayesian Approach
τ_1	$\dfrac{6(S+2)NT + 6NT}{3(2K+L+2)NT + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML}$	$\dfrac{6(S+2)NT + 6NT + 2MN}{3(2K+L+2)NT + \frac{1}{2}M^2L + 2NLM + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML}$
τ_1	$\dfrac{6(S+2)NT + 6NT}{6(M+1)NT + 2NT\sqrt{2M} + 2MT\sqrt{N} + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML}$	$\dfrac{6(S+2)NT + 6NT + 2MN}{6(2M+1)NT + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML}$

and a search routine may be selected by experimentation. For the search routine we used (appendix 1.3), a typical value of S is 4.

We further define θ as the ratio of the number of multiplications per iteration in the Gauss-Newton (or Marquardt) method when the adjoint equation formulation is used to that when the sensitivity equations are solved. Clearly $\theta = \tau_2/\tau_1$. When $\theta > 1$, direct integration of the sensitivity equation is preferable in the Gauss-Newton algorithm.

Figure (1.5.5) shows τ_1 and τ_2 as functions of the number of parameters M for fixed values of N, T, L, and K (N = 30, T = 200, K = 3, L = 60) and for S = 3, 4, and 5 for zonation in one-dimensional reservoirs. The behavior of τ_1 and τ_2 in the Bayesian case is similar to that shown and is presented in figure (1.5.6). Although one may consider the variation of τ_1 and τ_2 with each of the parameters, the most informative, and most relevant, case is that in which N, T, L, and K are fixed, and the number of unknown parameters M and the number of simulations in the unidirectional search, S are varied.

If, as noted, roughly ten times as many iterations are required in the conjugate gradient method as in the Gauss-Newton method, the total effort required by the two algorithms will be equal when τ_1 or τ_2 is 0.1. Figure (1.5.5) indicates that τ_1 is quite insensitive to M, decreasing only very slowly as M increases. Up to M = 60, $0.1 < \tau_1 < 1.0$, indicating that the Gauss-Newton method is to be favored. When the sensitivity equations are used (i.e., τ_2), for S = 3 the Gauss-Newton method is preferable to the conjugate gradient method when $M < 52$; for S = 4 this conclusion holds for $M < 60$. Thus, under the conditions of figure (1.5.5), the Gauss-Newton method appears to be

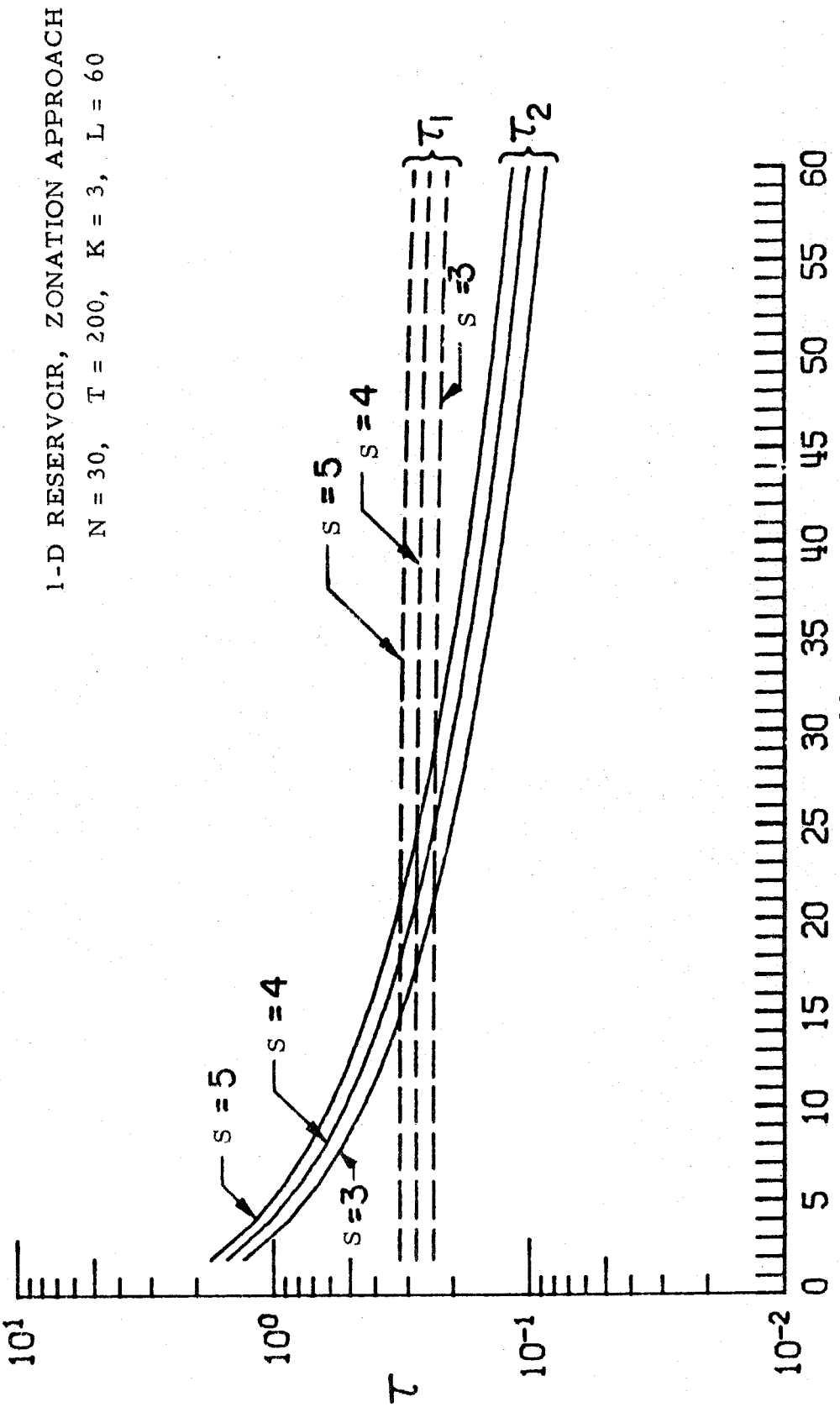
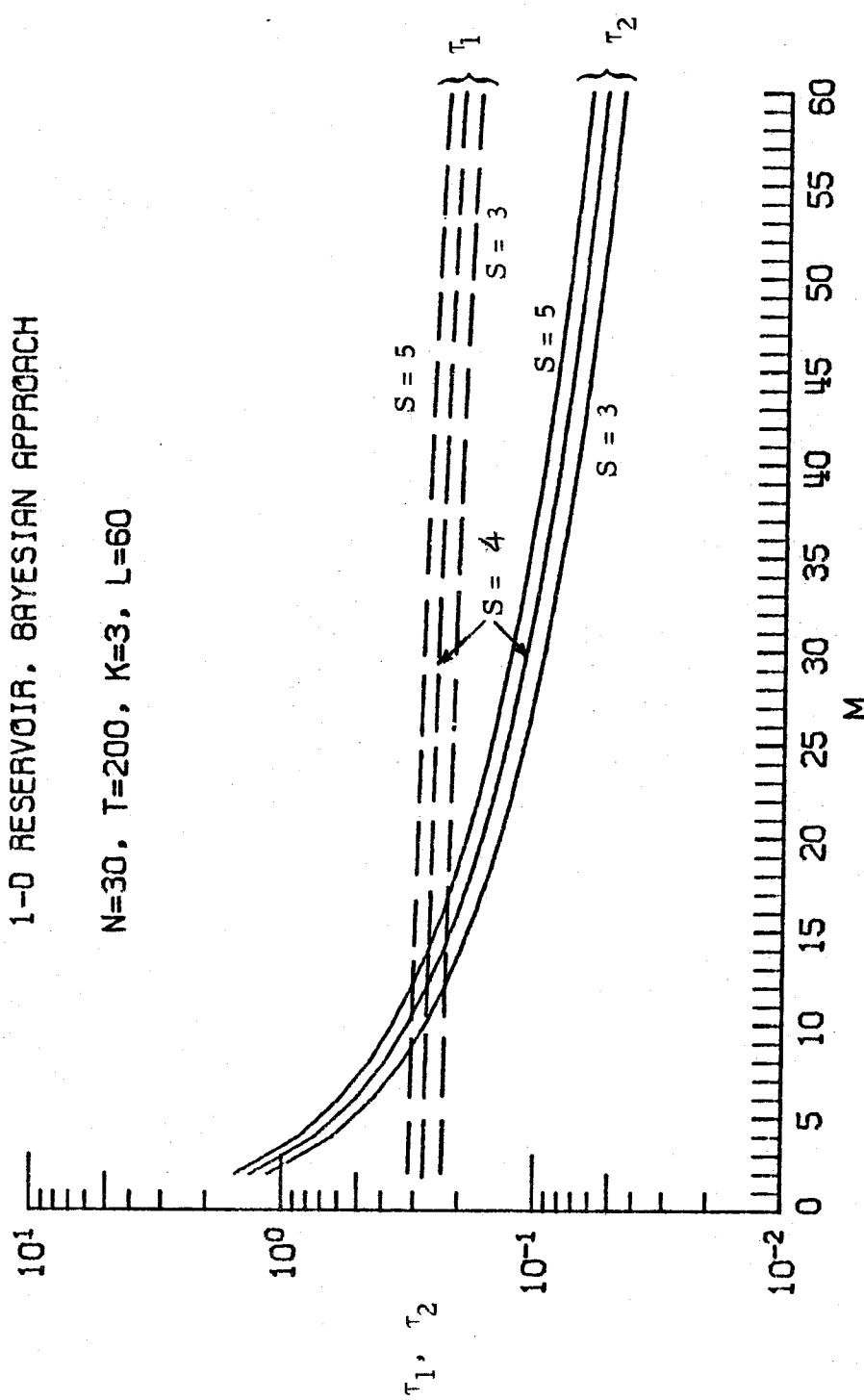


Figure 1.5.5 Ratios τ_1 and τ_2 for Zonation in One-Dimensional Reservoir

preferable to the conjugate gradient method, for zonation. Similarly, figure (1.5.6) indicates that when adjoint variable method is used in the Gauss-Newton algorithm, it is preferable to the conjugate gradient algorithm for all values of M and S in the Bayesian approach. On the other hand, when the sensitivity equations are integrated, the Gauss-Newton algorithm is more efficient than the conjugate gradient algorithm in Bayesian estimation when M is less than 26, 30 and 36 respectively for S = 3, 4 and 5. Thus from the figures we can decide the most efficient algorithm for any parameterization, for a given ratio of the number of iterations.

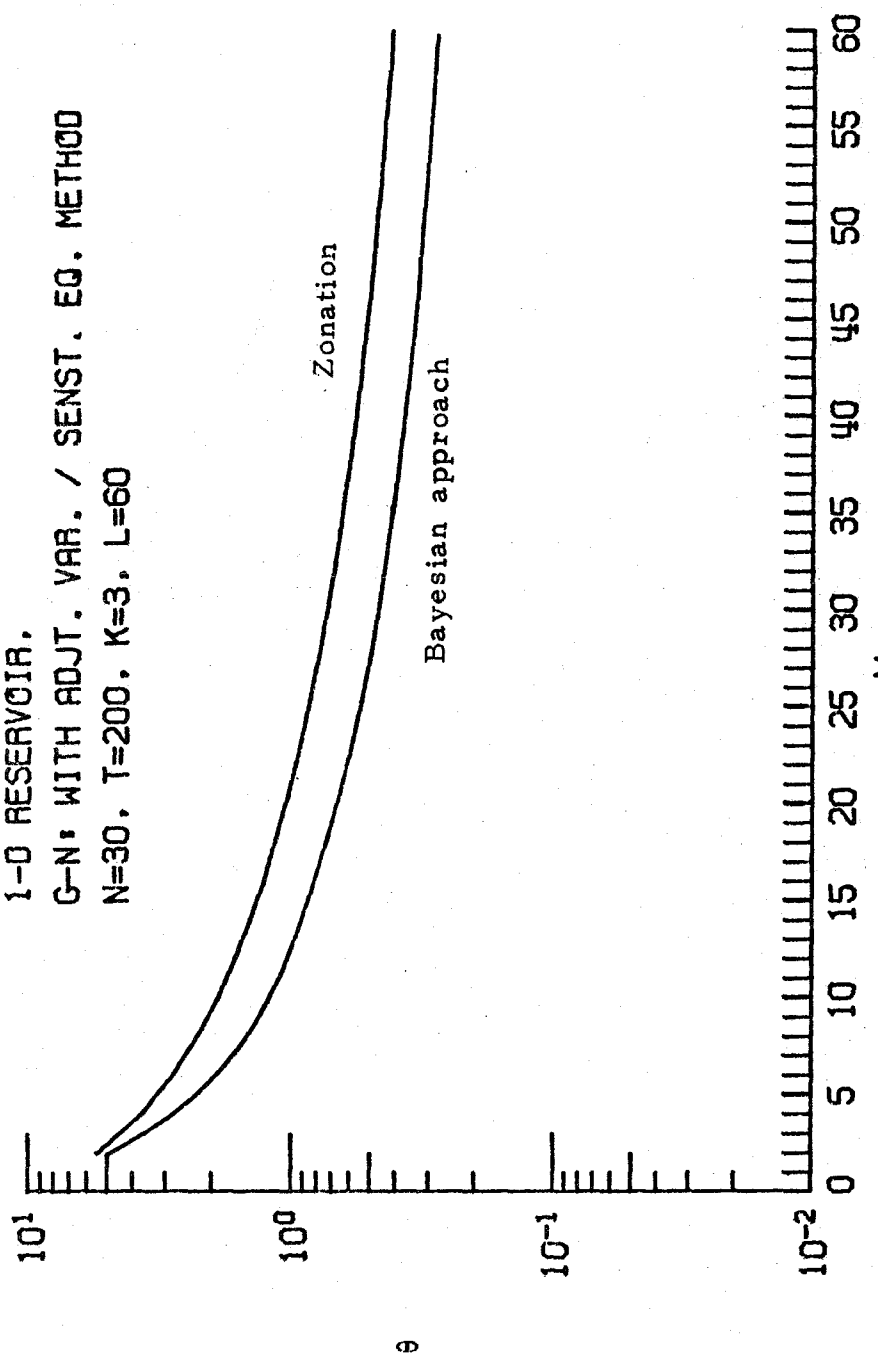
Figure (1.5.7) shows the variation of θ with M for zonation and Bayesian parameterizations under the same conditions as figure (1.5.5). When $\theta < 1$, the adjoint variable formulation is preferable to direct the solution of the sensitivity equations. For $M > 22$, the adjoint formulation is preferable for zonation; whereas in the Bayesian approach, $M \geq 12$ for this conclusion to hold. Obviously, the value of M at which it becomes more efficient to employ the adjoint variable approach as opposed to direct solution of the sensitivity equations will depend on the particular conditions (N, K, L, T) of the problem.

Figures (1.5.8-9) contain plots of τ_1 and τ_2 against M for zonation and Bayesian estimation respectively in two-dimensional reservoir, for N = 900, T = 200, K = 9, L = 180. Since N and L are large for this problem, the adjoint variable approach for sensitivity calculation requires a very large computational effort. (See the term 3LNT in table (1.5.4).) The conjugate gradient is preferable to the Gauss-Newton for $M \leq 180$ for both parameterizations. However, when



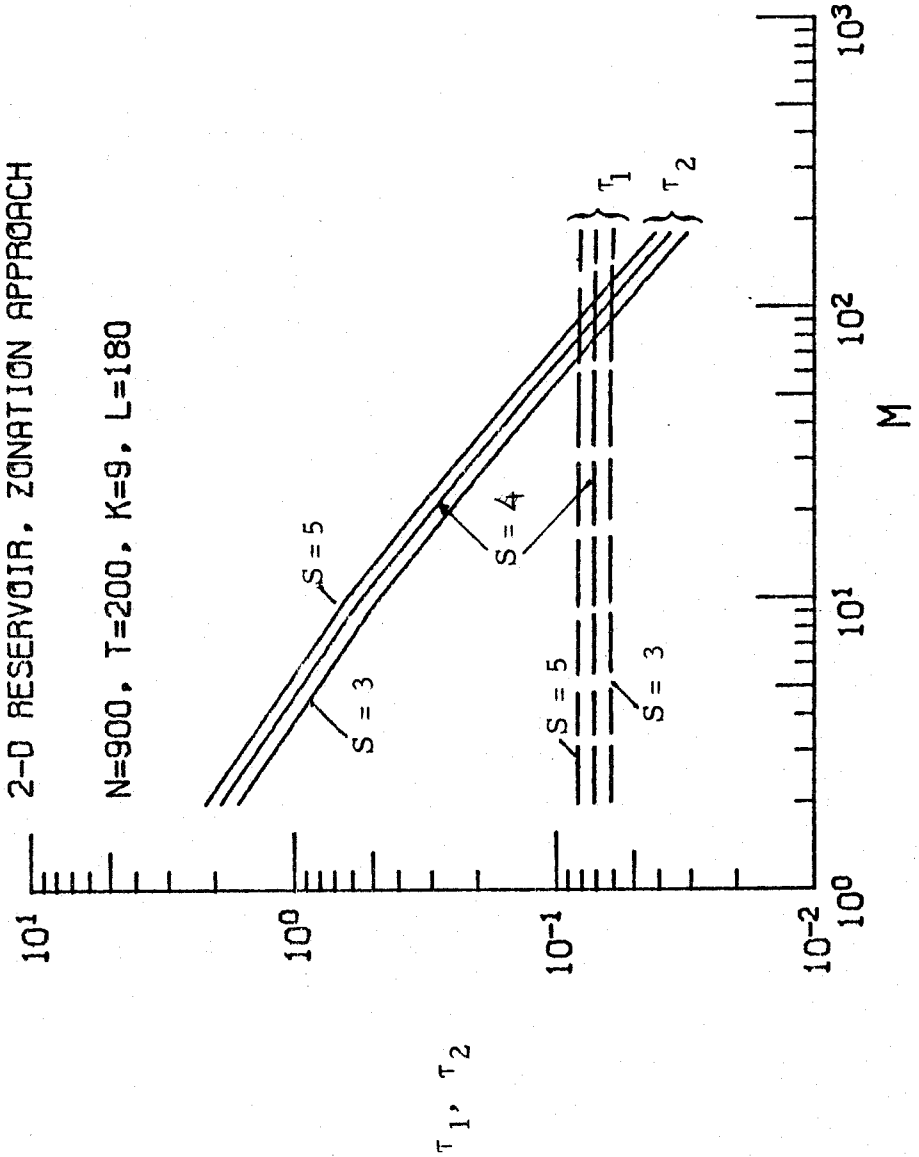
Ratios τ_1 and τ_2 for Bayesian Estimation in One-Dimensional Reservoir

Figure 1.5.6



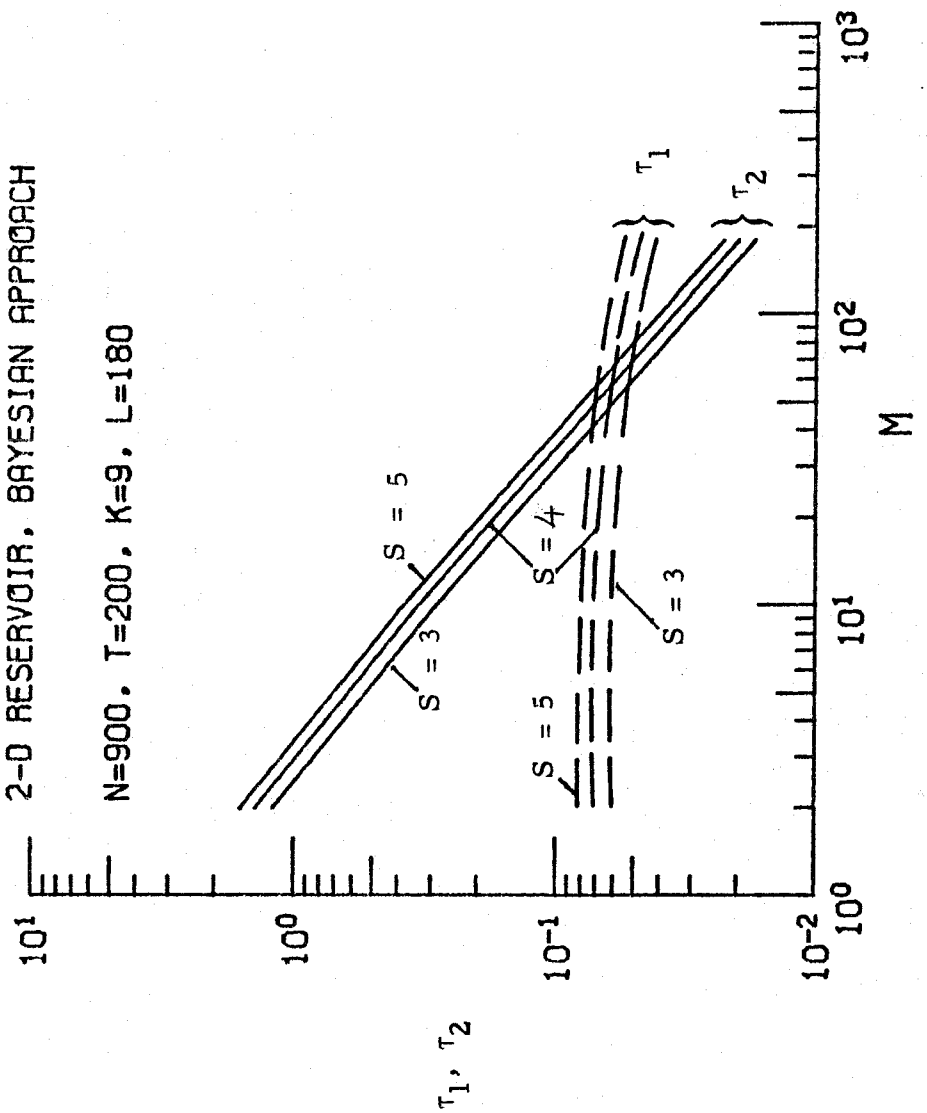
Ratios θ for Zonation and Bayesian Estimation in One-Dimensional Reservoir

Figure 1.5.7



Ratios τ_1 and τ_2 for Zonation in Two-Dimensional Reservoir

Figure 1.5.8



Ratios τ_1 and τ_2 for Bayesian Estimation in Two-Dimensional Reservoir

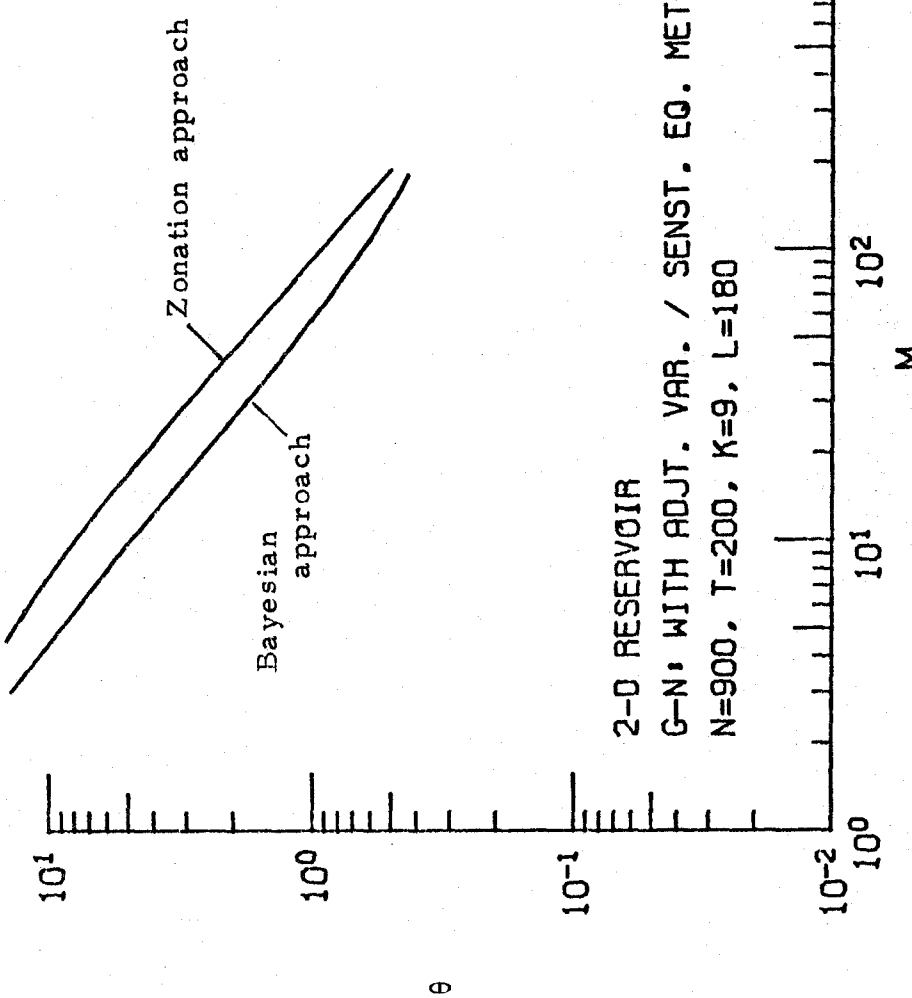
Figure 1.5.9

sensitivity equations are directly integrated, the Gauss-Newton algorithm is more efficient than the conjugate gradient for $M < 70$ in zonation and for $M < 35$ for Bayesian estimation, when $S = 3$.

Figure (1.5.10) shows the variation of θ with M for two-dimensional reservoir, for the same conditions as in figures (1.5.8-9). It is evident from these plots that, in the Gauss-Newton algorithm, direct integration of sensitivity equations is more efficient than adjoint variable method when M is smaller than 100 for zonation and 60 for Bayesian approach. Again, these conclusions are true for only the mentioned values of N, T, K and L; and for a given set of conditions plots similar to these may be drawn for a similar analysis.

1.5.4 Combination of Two Algorithms

Under certain circumstances, a combination of conjugate gradient and Marquardt algorithms may be more efficient than either of them alone. This is especially true when the initial estimates of the parameters involve large errors. In this situation, the initial history match error is very large and the Gauss-Newton approximation is poor. Then a large value of μ is required to ensure convergence of the Marquardt algorithm; as noted earlier, this results in a slow convergence and inefficiency due to the large computational effort per iteration. On the other hand, the conjugate gradient algorithm reduces J significantly in the first few iterations, but the convergence slows down considerably as J approaches its minimum value (see figures 1.5.2a-d). Thus, reduction of J to a satisfactory value may require a very large number of iterations if the initial guess is far from the "true" values of the parameters.



Ratios θ for Zonation and Bayesian Estimation in Two-Dimensional Reservoir

Figure 1.5.10

In this situation it is beneficial to utilize the initial rapid convergence of the conjugate gradient algorithm, and subsequently when its convergence slows down switch to the Marquardt algorithm which has a rapid convergence in the vicinity of the minimum. In addition to being efficient from a computational standpoint, this scheme has another advantage over the conjugate gradient algorithm when the zonation approach is used. As we shall discuss in chapters 2 and 3, for a given zonation, an "optimal" value of μ can be determined which yields the smallest total expected error (trace of the covariance) in the resulting estimates. When the combination of the two algorithms is used, this "optimal" value of μ can be used in the Marquardt algorithm, thus improving the quality of the estimates.

1.6 Results and Discussion

This section is concerned with the errors in the estimates obtained by zonation and Bayesian estimation. The criteria used to evaluate the estimate errors are :

$$J_k = \frac{1}{\overline{k}} \sum_{j=1}^{N} |\hat{k}_j - k_j^*| \qquad (1.6.1)$$

$$J_\phi = \frac{1}{\overline{\phi}} \sum_{j=1}^{N} |\hat{\phi} - \phi_j^*| \qquad (1.6.2)$$

where $\hat{k}_j$, $\hat{\phi}_j$ are the estimates and k_j^*, ϕ_j^* are the simulated or "true" values. In an actual field study these two errors cannot be determined since k_j^*, ϕ_j^* are not known. The success of estimation can only be evaluated by making predictions and comparing them with new measurements as they become available. A simulation study, on the other hand, is useful in providing comparisons based on errors J_k, J_ϕ. The

final data fit error, or history match error,

$$J_p = \sum_{m=1}^{R} \sum_{n=1}^{K} \left[p_{i_m, j_n}^{cal} - p_{mn}^{0} \right]^2 \qquad (1.6.3)$$

evaluated at $\underset{\sim}{\hat{k}}$ and $\underset{\sim}{\hat{\phi}}$ will also be followed. As will be seen below, however, this error alone is not an adequate indication of the quality of estimation.

The property estimates $\underset{\sim}{\hat{k}}$, $\underset{\sim}{\hat{\phi}}$ have a random component, for they depend on the random observation error and on the "true" property distributions, which are taken as a realization of a random process. A study of the effect of various conditions of estimation and a comparison of Bayesian estimation with zonation would thus require a sufficiently large sample of simulations. Each member of this sample would employ different observation errors and "true" property distributions. Since such a rigorous comparison would be quite expensive in computer time, we have limited the simulations to four property realizations generated using the statistics defined by the expressions (1.4.1-3). Two sets, S_1 and S_2, of conditions of estimation problem (see table (1.4.1)) were used in the simulations. The realizations (see figure (1.4.2)) R1-R4 were used in simulation with conditions S_1, whereas the realizations R1, R2, R3, R5 were used with S_2. In the numerical results reported here, all the simulations for a given set of conditions (S_1 or S_2) employed the same set of observation errors. Additional simulations, utilizing conditions S_1, with different sets of observation errors were also performed; these indicated that the randomness introduced by the observation error is substantially lower than the one resulting from a change in the "true" profile.

Hence, most of the comparisons were made on the basis of the same set of observation errors.

1.6.1 Ill-conditioning or "Non-uniqueness"

Reservoir estimation problems are known to be ill-conditioned in the sense that widely different parameter estimates produce essentially the same fit J_p (Jahns (1966), Chevant et al. (1975)). This behavior is often referred to as "non-uniqueness," although the term should not be understood in its strict mathematical sense. The ill-conditioned nature of estimation will be evaluated from the matrix of sensitivity coefficients in chapter 2. Some results of the present simulation study exhibiting ill-conditioning are presented in tables (1.6.1-2). Although the data fit error J_p has essentially attained its irreducible level, considerable estimate errors J_k and J_ϕ persist, and a lower value of J_p does not always imply lower estimate errors. The differences between the values of J_k and J_ϕ from case to case reflect considerable differences in the estimated property distributions.

1.6.2 Effect of the Number of Unknown Parameters

1.6.2.1 Zonation

The effect of the number of zones on the errors J_p, J_k, and J_ϕ for estimation problem conditions S_1 is illustrated by the results of tables (1-6.3-6). In all cases the data-fit error J_p decreases when the number of zones is increased from four to sixteen, with most of the change occurring from four to eight zones. Further increase in the number of zones does not seem to have any effect on J_p other than a small random variation. The estimate errors J_k and J_ϕ, on the

Table 1.6.1 Comparison of Data-fit and Estimate Errors.

Property Distribution: R-2 Estimation Problem Conditions: S_1

Initial values: $J_p = 0.285 \times 10^5$; $J_k = 11.73$; $J_\phi = 7.22$

Conditions of Estimation	Final Values		
	J_p	J_k	J_ϕ
Zonation: NZ = 8 Marquardt (μ=25)	67.0	6.37	9.18
Zonation: NZ = 16 Marquardt (μ=100)	65.0	6.55	6.40
Zonation NZ = 33 Marquardt (μ=100)	64.7	7.09	7.22
Bayesian: M = 20; β = 0.1; s = 5.0 Marquardt (μ=25)	65.8	4.92	6.27
Bayesian: M = 20, β = 1.0, s = 2.5 Marquardt (μ=25)	65.8	5.01	5.01

Table 1.6.2: Comparison of Data-Fit and Estimate Errors

Property Distribution: R-2; Estimation Problem Conditions: S_2

Initial Values: J_p = 14,900; J_k = 11.73; J_ϕ = 7.22

Zonation, Marquardt (μ = 5)

No. of Zones	Final Values		
	J_p	J_k	J_ϕ
4	77.0	11.07	6.43
8	71.4	7.22	6.29
16	70.5	7.52	7.04
33	70.7	7.89	6.82

Bayesian (M = 20, β = 0.1), Gauss-Newton (μ = 0)

Length of Autocorr.	Final Values		
s	J_p	J_k	J_ϕ
5.0	71.9	7.42	5.04
2.5	70.8	6.26	5.03
7.5	70.9	8.67	5.80

Table 1.6.3 Results of Zonation and Bayesian Estimation

Property Distribution: R-1 Estimation Problem Conditions: S_1

Initial values: $J_p = 0.105 \times 10^7$; $J_k = 4.70$; $J_\phi = 5.78$

Algorithm: Marquardt

Zonation: ($\mu = 500$ for NZ = 4; $\mu = 100$ otherwise)

No. of zones	Final values		
	J_p	J_k	J_ϕ
4	148	5.36	6.89
8	66.0	2.44	2.71
16	64.1	3.24	4.43
33	64.9	3.74	5.18

Bayesian Estimation: ($\mu = 25$)

M	β	s	Final values			
			J_p	J_2	J_k	J_ϕ
20	0.1	5.0	64.2	0.837	2.02	3.26
20	1.0	5.0	64.9	6.6	2.28	3.35
20	10.0	5.0	70.2	50.9	2.58	3.46
10	1.0	5.0	77.5	7.49	2.62	4.03
20	1.0	5.0	64.9	6.6	2.28	3.35
25	1.0	5.0	64.7	6.53	2.32	3.35
30	1.0	5.0	64.7	6.53	2.31	3.35
40	1.0	5.0	64.7	6.53	2.31	3.35
20	1.0	2.5	64.0	7.75	2.58	3.62
20	1.0	5.0	64.9	6.6	2.28	3.35
20	1.0	7.5	72.8	9.51	2.30	3.19

Table 1.6.4 Results of Zonation and Bayesian Estimation.

Property Distribution: R-2 Estimation Problem Conditions: S_1

Initial values: $J_p = 0.285 \times 10^5$; $J_k = 11.73$; $J_\phi = 7.22$
Algorithm: Marquardt

Zonation: ($\mu = 500$ for NZ = 4; $\mu = 100$ otherwise)

No. of zones	Final values J_p	J_k	J_ϕ
4	223	8.79	6.53
8	68.2	5.42	8.98
16	65.0	6.55	6.40
33	64.7	7.09	7.22

Bayesian Estimation: ($\mu = 25$)

M	β	s	Final values J_p	J_2	J_k	J_ϕ
20	0.1	5.0	65.8	4.32	4.92	6.27
20	1.0	5.0	69.7	93.4	3.07	4.60
20	10.0	5.0	103	153	3.51	4.19
10	1.0	5.0	118	36.4	5.14	4.51
20	1.0	5.0	69.7	93.4	3.07	4.60
30	1.0	5.0	69.4	24.4	3.64	4.89
40	1.0	5.0	69.4	24.4	3.64	4.89
20	1.0	2.5	65.8	25.0	5.01	5.01
20	1.0	5.0	69.7	93.4	3.07	4.60
20	1.0	7.5	74.9	42.5	3.27	4.91

Table 1.6.5 : Results of Zonation and Bayesian Estimation.

Property Distribution: R-3 Estimation Problem Conditions: S_1

Initial values: $J_p = 0.306 \times 10^5$; $J_k = 2.99$; $J_\phi = 7.33$
Algorithm: Marquardt

Zonation: ($\mu = 500$ for NZ = 4; $\mu = 100$ for NZ = 8, 16, 33)

	Final values		
No. of zones	J_p	J_k	J_ϕ
4	75.2	2.05	5.15
8	65.9	1.80	3.01
16	65.4	2.52	2.95
33	67.3	2.29	3.71

Bayesian Estimation: ($\mu = 25$)

			Final values			
M	β	s	J_p	J_2	J_k	J_ϕ
20	0.1	5.0	65.4	1.82	2.08	2.42
20	1.0	5.0	68.0	9.55	1.05	1.96
20	10.0	5.0	79.2	63.3	1.79	3.56
10	1.0	5.0	76.2	10.8	2.30	3.76
20	1.0	5.0	68.0	9.55	1.05	1.96
30	1.0	5.0	69.7	9.64	1.11	2.08
40	1.0	5.0	69.7	9.64	1.11	2.08
20	1.0	2.5	67.5	9.95	2.07	3.27
20	1.0	5.0	68.0	9.55	1.05	1.96
20	1.0	7.5	71.8	15.2	1.23	2.03

Table 1.6.6 Results of Zonation and Bayesian Estimation.

Property Distribution: R-4 Estimation Problem Conditions: S_1

Initial values: $J_p = 0.161 \times 10^7$; $J_k = 8.28$; $J_\phi = 10.41$
Algorithm: Marquardt

Zonation: ($\mu = 500$ for NZ = 4; $\mu = 100$ for NZ = 8, 16, 33)

No. of Zones	Final values J_p	J_k	J_ϕ
4	173	4.12	4.46
8	66.4	2.44	3.14
16	65.4	3.96	3.98
33	66.1	7.10	8.56

Bayesian Estimation: ($\mu = 25$)

M	β	s	Final values J_p	J_2	J_k	J_ϕ
20	0.1	5.0	66.3	3.35	1.47	2.64
20	1.0	5.0	69.6	20.4	1.12	2.04
20	10.0	5.0	88.8	148	2.43	2.00
10	1.0	5.0	80.9	19.0	3.92	2.96
20	1.0	5.0	69.6	20.4	1.12	2.04
30	1.0	5.0	69.2	20.3	1.03	2.05
40	1.0	5.0	69.2	20.3	1.03	2.05
20	1.0	2.5	66.0	25.7	2.01	2.39
20	1.0	5.0	69.6	20.4	1.12	2.04
20	1.0	7.5	78.3	21.0	3.06	2.92

other hand, pass through a minimum at eight zones. The minimum is rather flat since there are only small differences between the results obtained with eight and sixteen zones. To explain this minimum we observe that the estimate error has two components. One is due to the error of parameterization and decreases with an increasing number of zones. The other is due to statistical scatter and increases as the number of zones increases. Evidently, the sum of the two errors attains its minimum at an intermediate number of zones. These results emphasize that decreasing J_p does not necessarily imply improved estimates. We shall provide a theoretical analysis of estimate errors in zonation including a procedure for predicting the optimum number of zones in chapter 3.

The effect of the number of zones on J_p, J_k and J_ϕ is further illustrated by the simulations with conditions S_2, the results of which are presented in tables (1.6.7-10). In this set, the data appear to contain less information about the parameters than in set S_1 (we present a quantitative analysis concerning this aspect in chapters 2 and 3); consequently the error J_p approaches closely the irreducible level even for NZ = 4. Also, J_p increases in two cases as NZ is changed from 8 to 16; this implies that for this set, zonation with 16 zones results in too high a dimensionality of the parameter space to carry out an extensive minimization efficiently. Again, the total error $J_k + J_\phi$ passes through a flat minimum at NZ = 16; however, the statistical scatter is too large to allow conclusions about the optimal number of zones from this limited sample.

Table 1.6.7: Results of Zonation and Bayesian Estimation

Property Distribution: R-1 Estimation Problem Conditions: S_2

Initial Values: $J_p = 0.800 \times 10^6$; $J_k = 4.70$; $J_\phi = 5.78$

Algorithm: Marquardt

Zonation: ($\mu = 100$)

No. of zones	Final Values J_p	J_k	J_ϕ
4	78.7	1.02	3.19
8	70.0	3.47	3.05
16	70.4	4.17	3.30
33	70.9	4.14	5.07
Bayesian Estimation ($\mu = 50$)	70.7	3.39	2.49

$\beta = 1$, $s = 5$, $M = 20$

Table 1.6.8: Results of Zonation and Bayesian Estimation

Property Distribution: R-2; Estimation Problem Conditions: S_2

Initial values: $J_p = 0.149 \times 10^5$, $J_k = 11.73$, $J_\phi = 7.22$

Algorithm: Marquardt

Zonation: ($\mu = 100$)

No. of zones	Final Values		
	J_p	J_k	J_ϕ
4	78.3	9.05	6.29
8	71.5	8.53	6.27
16	71.8	7.88	6.32
33	72.0	8.26	6.35

Bayesian Estimation: ($\mu = 25$)

M	β	s	Fianl Values		
			J_p	J_k	J_ϕ
20	0.1	5.0	71.9	7.42	5.04
20	0.25	5.0	73.4	7.57	5.44
20	1.0	5.0	79.8	7.51	5.73
16	0.1	5.0	71.9	7.67	5.08
20	0.1	5.0	71.9	7.42	5.04
20	0.1	2.5	70.8	6.26	5.03
20	0.1	5.0	71.9	7.42	5.04
20	0.1	7.5	70.9	8.67	5.80

Table 1.6.9: Results of Zonation and Bayesian Estimation

Proberty Distribution: R-3; Estimation Problem Conditions: S_2

Initial values: $J_p = 0.256 \times 10^5$; $J_k = 2.99$; $J_\phi = 7.33$

Algorithm: Marquardt

Zonation: ($\mu = 100$)

No. of zones	Final Values		
	J_p	J_k	J_ϕ
4	70.5	1.89	4.49
8	70.5	1.80	2.88
16	71.3	1.48	2.17
33	73.3	1.89	2.56
Bayesian Estimation ($\mu = 50$)	73.1	1.51	1.84

$\beta = 1$, $s = 5$, $M = 20$

Table 1.6.10: Results of Zonation and Bayesian Estimation

Property Distribution: R-5; Estimation Problem Conditons: S_2

Initial values: $J_p = 0.437 \times 10^5$, $J_k = 5.49$, $J_\phi = 7.12$

Algorithm: Marquardt

Zonation: ($\mu = 100$)

No. of zones	Final Values		
	J_p	J_k	J_ϕ
4	70.9	5.34	3.18
8	69.5	3.98	3.19
16	70.1	3.74	2.39
33	72.8	4.20	2.71
Bayesian Estimation ($\mu = 50$)	70.7	3.86	2.63

$\beta = 1$, $s = 5$, $M = 20$

1.6.2.2 Bayesian Estimation

Tables (1.6.3-6) indicate that an increase in the number of the basis vectors used in Bayesian estimation results in a decrease in all three errors J_p, J_k, and J_ϕ. The decrease becomes insignificant beyond 20 vectors; hence for the problem with conditions S_1, the optimum number of unknowns is about 20. This is close to the sixteen unknowns (eight zones) which were found optimal in zonation for S_1. However, unlike zonation, the estimate errors in the Bayesian estimation do not increase when the number of parameters is increased beyond the optimum. This is an attractive feature of Bayesian estimation since in a real application the optimum number of unknowns would be quite uncertain a priori.

1.6.3 The Effect of Incorrect Prior Statistics in Bayesian Estimation

The weighting factor β, $0 < \beta < 1$ was introduced in equations (1.3.6a) and (1.3.22a) to allow a reduced weighting of prior informatior However, this factor can also be varied to study the effect of error in prior statistics. In the latter case β should be allowed to assume values larger, as well as smaller, than unity. The use of $\beta > 1$ is tantamount to placing a larger confidence in the prior estimates of all the parameters. The other parameter that can be varied to study the effect of errors in prior statistics is the dimensionless correlation length s.

Tables (1.6.3-6,8) present results obtained by using erroneous values of β and s. As discussed in section 5, the "true" property

distributions R1 - R5 were generated using the "correct" statistics ($\beta = 1$, $s = 5$). The use of erroneous β, e.g., $\beta = 0.1$ or $\beta = 10$ instead of $\beta = 1$ produces significant variation in the estimate errors J_k and J_ϕ. However, the differences are partly due to statistical scatter since the correct value of β does not always result in the smallest estimate errors. Similar behavior is observed when s is varied in the range $2.5 < s < 7.5$ (see also figure (1.3.1)). In view of the rather large errors in β and s explored, the resulting estimate variations are considered to be quite acceptable and, as will be shown below, they compare favorably with the results of zonation.

1.6.4 Comparison between Zonation and Bayesian Estimation

1.6.4.1 Simulations with Random Property Distributions

Tables (1.6.3-6) present the estimate errors obtained using zonation and Bayesian estimation for the "true" property distributions R1 - R4 and a single set of estimations S_1. In a real application the Bayesian parameters β, s will not be accurately known. Similarly, the optimum number of zones will also be uncertain. Hence, a comparison of the two approaches must encompass results obtained with various values of β, s and with different numbers of zones. If we limit the comparison to cases with $\beta = 0.1, 1, 10$; $s = 2.5, 5, 7.5$ and with 8 and 16 zones, we find that the Bayesian estimation yields results considerably better than zonation. For example, the J_k, J_ϕ errors averaged over all the cases specified (with conditions S_1) are 2.74, 3.36 in Bayesian estimation and 3.55, 4.45 in zonation. Incidentally, the Bayesian results are superior for each of the four

property distributions R1 - R4.

Using the results in tables (1.6.7-10), a similar but less extensive comparison can be carried out for simulations using four realizations R1, R2, R3, R5 under conditions S_2. Due to limited data, we shall restrict the comparison to cases with $\beta = 1$, $s = 5$ and with 8 and 16 zones for this set of conditions. The values of J_k and J_ϕ averaged over the four realizations are, respectively, 4.07 and 3.17 in Bayesian estimation, 4.45 and 3.35 for 8 zones, and 4.32 and 3.55 for 16 zones. Thus, for these conditions too, the Bayesian estimation appears to be superior to zonation.

1.6.4.2 Simulations with Non-random Property Distributions

As already discussed, the "true" distributions R1 - R5 used in tables (1.6.3-10) were obtained as realizations of a random process with the covariance matrices given by equations (1.4.1-3) with $\beta = 1$, $s = 5$. The Bayesian penalty term used in the estimation was formulated using the same functional form for the covariances but different values of β and s. The mere fact, however, that the same functional form for the covariance matrices is utilized both for the generation of "true" property distributions and for the estimation may introduce a bias favoring the Bayesian approach in comparison to zonation. To investigate the importance of this bias, we carried out further simulations with the property distribution R-NR which, as seen in figure (1.4.5), is piecewise constant. Estimation computations were then carried out using two zonation procedures. In the first, the true zone

boundaries were assumed known, i.e., three of them were taken to coincide with the three points of discontinuity in R-NR. In the second, the zone boundaries were selected arbitrarily. The results of these simulations with conditions S_1 are presented in table (1.6.11). Table (1.6.12) presents similar results for fewer simulations with estimation conditions S_2. For the case of known zone boundaries zonation, as expected, this provides considerably better results than the Bayesian approach. However, for the case of unknown zone boundaries, the Bayesian approach provides much better estimates than zonation. Since the latter situation would be more in accord with real field studies, it is concluded that the Bayesian estimation provides generally better estimates than zonation.

1.6.4.3 Convergence of Minimization Procedure

In addition to yielding superior estimates, the Bayesian parameterization results in superior convergence characteristics of minimization algorithms. This aspect is especially important when the "true" property distributions are not smooth, as in R-NR. As evidenced by the multiple values of μ reported in tables (1.6.11-12), in zonation the Marquardt algorithm required many iterations beginning with large values of μ. On the other hand, in all the reported cases of Bayesian estimation, except one, a single value of μ sufficed to yield a satisfactory value of J_p in 5 to 7 iterations. Thus, for these distributions generally two to three times as much computational effort was required for estimation using zonation as that for the Bayesian estimation. For smooth "true" property distributions (R1-R5), the difference was not very large, but generally Bayesian parameterization led to

Table 1.6.11 Results of Zonation and Bayesian Estimation.

Property Distribution: R-NR Estimation Problem Conditions: S_1

Initial values: $J_p = 0.408 \times 10^5$; $J_k = 7.51$; $J_\phi = 7.51$

Algorithm: Marquardt

Zonation: zone boundaries coincide with discontinuities in "true" distributions

No. of zones	μ^*	Final values J_p	J_k	J_ϕ
4	10^5, 10^4, 2×10^3	75.0	2.25	3.41
8	10^5, 10^3	73.0	2.92	4.59
16	10^2	68.0	3.22	4.14
33	10^2	69.8	3.47	4.54

Zonation: zone boundaries do not coincide with discontinuities in "true" distributions

No. of zones	μ^*	Final values J_p	J_k	J_ϕ
4	10^5, 10^4, 5×10^2	413	6.63	10.39
8	10^5, 10^4, 5×10^2	106	4.43	6.66
16	10^5, 10^4, 10^3	76.5	4.05	5.86

Bayesian Estimation: ($\mu = 25$)

M	β	s	Final values J_p	J_2	J_k	J_ϕ
20	0.1	5.0	66.5	3.11	4.36	3.19
20	1.0	5.0	69.8	15.3	4.27	3.03
20	10.0	5.0	97.7	74.4	3.54	3.90
20	1.0	2.5	67.6	9.44	3.74	3.32
20	1.0	5.0	69.8	15.3	4.27	3.03
20	1.0	7.5	80.0	23.0	4.83	4.14

* When multiple values of μ are reported, they were used successively as minimization of J_p proceeded.

Table 1.6.12: Results of Zonation and Bayesian Estimation

Property Distribution: R-NR; Estimation Problem Conditions: S_2

Initial values: $J_p = 0.106 \times 10^7$; $J_k = 7.51$; $J_\phi = 7.51$

Algorithm: Marquardt

Zonation: zone boundaries coincide with discontinuities in "true" distribution.

No. of zones	μ^*	Final Values J_p	J_k	J_ϕ
4	10^3	77.8	3.65	3.94
8	10^5, 10^2	71.0	3.46	4.03
16	10^5, 10^2	71.7	3.72	4.94

Zonation: zone boudnaries do not coincide with discontinuities in "true" distribution.

No. of zones	μ^*	Final Values J_p	J_k	J_ϕ
8	10^4, 2×10^3	82.4	7.45	6.50

Bayesian Estimation: $(\mu = 10^2, 25)^*$

M	β	s	Final Values J_p	J_k	J_ϕ
20	1.0	5.0	74.1	5.29	6.38

* When multiple values of μ are reported, they were used successively as minimization of J_p proceeded.

quicker convergence. Thus, also from the standpoint of computational effort, the Bayesian parameterization is preferable, especially for ill-behaved problems.

While closing this section we note that the general problem of estimate errors for the Bayesian estimation as well as zonation is treated by means of a linearized analysis in the following chapters. The theoretical results confirm most of the numerical results of this section relative to the significance of the number of unknowns, the effect of errors in prior statistics and the comparison between Bayesian estimation and zonation.

1.7 Conclusions

(1) The use of well pressure data to estimate the porosity and permeability distributions in a petroleum reservoir is an underdetermined statistical problem. Its solution necessitates the reduction of the number of unknowns by some form of parameterization.

(2) When zonation is used to reduce the number of unknowns, the best estimates are obtained for an intermediate number of zones.

(3) Another way of reducing the number of unknowns and decreasing the error due to statistical scatter is by Bayesian estimation. This involves the addition to the objective function of a penalty term incorporating prior geological information. This approach results, in effect, in a parameterization with fewer unknowns and better conditioned minimization problems. The results of the Bayesian estimation were found to be more accurate than those of zonation in a simulated case of a one-dimensional reservoir. The Bayesian formulation also resulted in improved convergence of the minimization algorithms.

(4) The accuracy of the Bayesian estimates depends on the accuracy of the prior statistics used; therefore, the Bayesian estimation can be greatly improved as additional geological information or spatial autocorrelations of rock properties become available.

(5) Both the Bayesian and the zonation approaches to estimation were implemented by the conjugate gradient and the Gauss-Newton (or Marquardt's) minimization algorithms. Detailed comparisons of the computational effort per iteration for the conjugate gradient and Gauss-Newton (or Marquardt) methods, for one-dimensional and two-dimensional reservoirs, were carried out. Formulas have been developed to enable one to determine which method is expected to be more efficient for a particular problem. In addition, comparisons of the computational effort in the Gauss-Newton (or Marquardt) method when the adjoint and sensitivity equation formulations are used were carried out.

CHAPTER 2

LINEAR ANALYSIS OF THE RESERVOIR PARAMETER ESTIMATION PROBLEM

The nonlinear parameter estimation problem is usually solved iteratively. At each iteration, the corrections in the current parameter estimates are obtained through the solution of an associated linear subproblem. An analysis of the linear subproblem is apt to yield a considerable amount of information about the nature of the nonlinear problem. Such information can be used to guide the iterative estimation process in a more efficient and fruitful manner. In this chapter we present an analysis of the linear subproblem and illustrate some ways in which the resulting information can be utilized. We shall postpone until chapter 3 the exploitation of the information about the nature of the problem to obtain approximate but realistic covariances for the estimates.

Having introduced the linear subproblem in section 2.1, we proceed to describe in section 2.2 its decomposition and the conclusions about the existence and uniqueness of its solution under various conditions. In section 2.3, the inverses of the linear subproblem corresponding to the corrections in the commonly used first and quasi-second order iterative minimization algorithms are discussed; they are analyzed in the light of the decomposition and several conclusions are made about the nature of the resulting corrections. In section 2.4, we present analytic results about the sensitivity of an observed pressure with respect to the porosity and permeability in a one-dimensional

single phase reservoir. The results provide a basis for making an order-of-magnitude comparison of the influence on the observation of the various Fourier components of perturbations in the porosity and permeability, leading to some conclusions about the identifiability of these components of the unknown rock properties. This is followed up in section 2.5 by a numerical analysis of the linear subproblem for two different sets of k and ϕ distributions in a one-dimensional reservoir, illucidating the behavior of the simulations described in chapter 1. In section 2.6, an approach to simplification of the parameter estimation problem, based on the results of the linear analysis, is described; its merits and demerits relative to the Bayesian approach are discussed. Another application of the linearized analysis is described in section 2.7, which may be used for a priori evaluation of the identifiability of the rock properties in the different regions of a reservoir, and may aid a rational choice of parameterization. We close the chapter by a discussion about the uses of the linearized analysis in section 2.8, addressed in particular to the practicing reservoir engineer.

2.1 The Linear Subproblem in Reservoir Property Estimation

The observed pressures are functions of the unknown rock properties in the reservoir for a given set of initial and boundary conditions and a given production history. Similarly, the calculated model pressures depend only on the estimates of these rock properties when the rest of the conditions remain unchanged. Let the pressures $\{p(\mathbf{x}_i, t_j); i = 1, 2, \ldots, R,\ j = 1, 2, \ldots, K\}$ be the observed quantities. For the discretization described in chapter 1, these are represented by the grid point values $\{p_{i_m, j_n}; m=1, 2\ldots, R, n=1, 2, \ldots K\}$. Let us define

an L-vector (L = RK) of the calculated pressures, $\underset{\sim}{y}^T = (p_{i_1,j_1}, p_{i_2,j_1}, \ldots, p_{i_R,j_1}, p_{i_1,j_2}, \ldots, p_{i_R,j_K})$. Let us denote the L-vector of the observations $\{p^o_{mn}\}$ as $\underset{\sim}{y}^o$, and the M-vector of unknown parameters as $\underset{\sim}{\pi}$. The vector $\underset{\sim}{\pi}$ is linearly related to the composite vector of the grid point values of the unknown rock properties. When these are the spatial distributions of the permeability $k(\underset{\sim}{r})$, and the porosity $\phi(\underset{\sim}{r})$,

$$\underset{\sim}{\pi} = \underset{\sim}{G} \begin{bmatrix} \underset{\sim}{k} \\ --- \\ \underset{\sim}{\phi} \end{bmatrix} \tag{2.1.1}$$

where $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ are the N-vectors of the grid point values and $\underset{\sim}{G}$ is an (Mx2N) matrix depending on the parameterization used, N being the number of grid points. Then we can express the functional relation as,

$$\underset{\sim}{y} = \underset{\sim}{y}(\underset{\sim}{\pi}) \tag{2.1.2}$$

This is a nonlinear implicit relation. The associated linear subproblem can be derived as follows. Consider a variation $\delta\underset{\sim}{\pi}$ in $\underset{\sim}{\pi}$; let $\delta\underset{\sim}{y}$ be the resulting variation in $\underset{\sim}{y}$. Then we have the Taylor series expansion of (2.1.2) as

$$\delta\underset{\sim}{y} = \left(\frac{\partial \underset{\sim}{y}}{\partial \underset{\sim}{\pi}}\right)\delta\underset{\sim}{\pi} + \frac{1}{2}\sum_{ij} \frac{\partial^2 \underset{\sim}{y}}{\partial\pi_i\,\partial\pi_j}\,\delta\pi_i\,\delta\pi_j + \ldots \tag{2.1.3}$$

For small variations we may neglect the terms quadratic and higher order in $\{\delta\pi_i\}$ to obtain,

$$\delta\underset{\sim}{y} = \left(\frac{\partial \underset{\sim}{y}}{\partial \underset{\sim}{\pi}}\right)\delta\underset{\sim}{\pi} \equiv \underset{\sim}{A}\,\delta\underset{\sim}{\pi} \tag{2.1.4}$$

where we have denoted the LxM matrix of the derivatives by $\underset{\sim}{A}$. The

element a_{ij} of $\underset{\sim}{A}$ is the sensitivity of the i^{th} observation with respect to the j^{th} parameter; hence $\underset{\sim}{A}$ is commonly called the sensitivity matrix.

2.2 Analysis of the Linear Problem

Lanczos (1961) introduced a diagonalization of a rectangular matrix called singular value decomposition. For an (LxM) real matrix $\underset{\sim}{A}$, two sets of orthonormal eigenvectors $\{\underset{\sim}{u}_i\}$ and $\{\underset{\sim}{v}_i\}$ may be found such that

$$\underset{\sim}{A}\,\underset{\sim}{v}_j = \lambda_j \underset{\sim}{u}_j \qquad (2.2.1a)$$

$$\underset{\sim}{A}^T \underset{\sim}{u}_j = \lambda_j \underset{\sim}{v}_j \qquad (2.2.1b)$$

Equivalently,

$$\underset{\sim}{A}^T \underset{\sim}{A}\,\underset{\sim}{v}_j = \lambda_j^2 \underset{\sim}{v}_j \qquad j = 1, 2, \ldots, M \qquad (2.2.2a)$$

$$\underset{\sim}{A}\,\underset{\sim}{A}^T \underset{\sim}{u}_j = \lambda_j^2 \underset{\sim}{u}_j \qquad j = 1, 2, \ldots, L \qquad (2.2.2b)$$

It can be assumed without loss of generality that all eigenvalues $\{\lambda_i\}$ are positive and are arranged in a non-increasing order of magnitude. It can be shown (Noble, 1969) that for some integer $r \leq \min(M, L)$,

$$\lambda_i = 0 \qquad i > r \qquad (2.2.3)$$

The integer r is the rank of $\underset{\sim}{A}$ and may be interpreted as the number of independent relations existing between the elements of the vectors $\delta\underset{\sim}{y}$ and $\delta\underset{\sim}{\pi}$ in (2.1.4). In terms of the estimation problems, it is called the number of degrees of freedom in the data.

Defining the (Lxr) matrix $\underset{\sim}{U} = (\underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots, \underset{\sim}{u}_r)$, (Mxr) matrix $\underset{\sim}{V} = (\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots \underset{\sim}{v}_r)$ and (rxr) diagonal matrix $\underset{\sim}{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, we can write (2.2.1) as,

$$\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}^T \qquad (2.2.4)$$

where we have used the orthogonality within the sets of $\{\underset{\sim}{u}_i\}$ and $\{\underset{\sim}{v}_i\}$. We can find (L-r) orthogonal L-vectors, $\underset{\sim}{u}_{r+1}, \ldots, \underset{\sim}{u}_L$, each of which is orthogonal to all the columns of $\underset{\sim}{U}$. Similarly, a set of (M-r) orthonormal M-vectors $\{\underset{\sim}{v}_{r+1}, \ldots, \underset{\sim}{v}_K\}$, each orthogonal to the columns of $\underset{\sim}{V}$ can be found. Using these vectors as columns, the Lx(L-r) matrix $\underset{\sim}{U}_o$ and Mx(M-r) matrix $\underset{\sim}{V}_o$ can be formed. Then the linear vector spaces spanned by the columns of $\underset{\sim}{U}$ and $\underset{\sim}{U}_o$ are orthogonal complements of each other; together, they span the L-dimensional Euclidean space E^L. Similarly, the columns of $\underset{\sim}{V}$ and $\underset{\sim}{V}_o$ together span E^M. The foregoing discussion can be summarized in the following mathematical statements:

$$\underset{\sim}{U}^T\underset{\sim}{U} = \underset{\sim}{I}_r \qquad (2.2.5a)$$

$$\underset{\sim}{U}_o^T\underset{\sim}{U}_o = \underset{\sim}{I}_{(L-r)} \qquad (2.2.5b)$$

$$\underset{\sim}{U}_o^T\underset{\sim}{U} = \underset{\sim}{0} \quad (L-r)xr \qquad (2.2.5c)$$

$$\underset{\sim}{V}^T\underset{\sim}{V} = \underset{\sim}{I}_r \qquad (2.2.6a)$$

$$\underset{\sim}{V}_o^T\underset{\sim}{V}_o = \underset{\sim}{I}_{(M-r)} \qquad (2.2.6b)$$

$$\underset{\sim}{V}^T\underset{\sim}{V}_o = \underset{\sim}{0} \quad rx(M-r) \qquad (2.2.6c)$$

If $r = L \leq M$, then $\underset{\sim}{U}$ is an orthogonal square matrix with full rank, and we have

$$\underset{\sim}{U}^T \underset{\sim}{U} = \underset{\sim}{U}\,\underset{\sim}{U}^T = \underset{\sim}{I}_L \qquad (2.2.7)$$

Similarly if $r = M \leq L$, the matrix $\underset{\sim}{V}$ is orthogonal,

$$\underset{\sim}{V}^T \underset{\sim}{V} = \underset{\sim}{V}\,\underset{\sim}{V}^T = \underset{\sim}{I}_M \qquad (2.2.8)$$

Lanczos (1961) demonstrated that the condition $r = L$ guarantees the existence of a solution to linear equation (2.1.4), and the condition $r = M$ guarantees uniqueness if a solution exists.

However, we shall follow Jackson (1972) for the analysis presented below. Since L-vectors $\{\underset{\sim}{u}_i : i=1, \ldots, L\}$ form a complete orthonormal set, we may express $\delta\underset{\sim}{y}$ as their linear combination,

$$\delta\underset{\sim}{y} = \sum_{i=1}^{L} \beta_i \underset{\sim}{u}_i = \sum_{i=1}^{r} \beta_i \underset{\sim}{u}_i + \sum_{r+1}^{L} \beta_i \underset{\sim}{u}_i = \underset{\sim}{U}\underset{\sim}{\beta} + \underset{\sim}{U}_o\underset{\sim}{\beta}_o \qquad (2.2.9)$$

where $\underset{\sim}{\beta}$ is an r-vector and $\underset{\sim}{\beta}_o$ is an (L-r) vector. Similarly, we can express $\delta\underset{\sim}{\pi}$ as,

$$\delta\underset{\sim}{\pi} = \underset{\sim}{V}\,\underset{\sim}{a} + \underset{\sim}{V}_o\,\underset{\sim}{a}_o \qquad (2.2.10)$$

Substituting these into (2.1.4),

$$\underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}^T(\underset{\sim}{V}\underset{\sim}{a} + \underset{\sim}{V}_o\,\underset{\sim}{a}_o) = \underset{\sim}{U}\,\underset{\sim}{\beta} + \underset{\sim}{U}_o\,\underset{\sim}{\beta}_o \qquad (2.2.11)$$

An exact solution of (2.2.11) exists if the residual vector $\underset{\sim}{\epsilon} = \underset{\sim}{A}\,\delta\underset{\sim}{\pi} - \delta\underset{\sim}{y}$ vanishes. We have,

$$\|\underset{\sim}{\epsilon}\|^2 = \underset{\sim}{\epsilon}^T\underset{\sim}{\epsilon} = \|\underset{\sim}{\Lambda}\underset{\sim}{a} - \underset{\sim}{\beta}\|^2 + \|\underset{\sim}{\beta}_o\|^2 \qquad (2.2.12)$$

The least square solution, which minimizes $\|\underset{\sim}{\epsilon}\|^2$, is given by

$$\underset{\sim}{a} = \underset{\sim}{\Lambda}^{-1}\underset{\sim}{\beta} \qquad (2.2.13)$$

and it has the least square error $\|\underset{\sim}{\beta}_o\|^2 = \|\underset{\sim}{U}_o^T \delta\underset{\sim}{y}\|^2$. This least square error is zero if $\delta\underset{\sim}{y}$ is orthogonal to all the columns of $\underset{\sim}{U}_o$. When $r = L$, this condition holds trivially, since $\underset{\sim}{U}_o$ is nonexistent. When $r < L$, this condition for existence of an exact solution leads to (L-r) linear constraints on $\delta\underset{\sim}{y}$,

$$\underset{\sim}{u}_i^T \delta\underset{\sim}{y} = 0 \qquad i = r+1, \ldots, L \qquad (2.2.14).$$

In such circumstances the system is called overconstrained.

The vector $\underset{\sim}{a}_o$ is not determined from the data by equation (2.2.11), since $\underset{\sim}{U}\ \underset{\sim}{\Lambda}\ \underset{\sim}{V}^T \underset{\sim}{V}_o \equiv \underset{\sim}{0}$, and can be arbitrarily selected without affecting the data fit. Thus, unless $r = M$, there is an infinite number of solutions, each corresponding to a different choice of $\underset{\sim}{a}_o$. Uniqueness of the solution is guaranteed when $r = M$, since then $\underset{\sim}{a}_o$ is nonexistent. Hence when $r < M$, the linear system is said to be underdetermined.

In the case when $r < L$ and $r < M$, the linear system is both overconstrained and underdetermined. Then no exact solutions exist, and an infinite number of solutions exist which yield equal residuals satisfying the least square criterion.

We shall illustrate the foregoing ideas by a few simple examples. As an example of an underdetermined system, let us consider the (2x3) system $\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$ with,

$$\underset{\sim}{A} = \begin{bmatrix} (2+\sqrt{3})/3 & (2-\sqrt{3})/3 & 2/3 \\ \sqrt{2}(4-\sqrt{3})/6 & \sqrt{2}(4+\sqrt{3})/6 & 2\sqrt{2}/3 \end{bmatrix}$$

Then the singular value decomposition of $\underset{\sim}{A}$ is,

$$\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}^T$$

with,

$$\underset{\sim}{U} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \sqrt{2/3} \\ \sqrt{2/3} & -1/\sqrt{3} \end{bmatrix}$$

$$\underset{\sim}{\Lambda} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

and,

$$\underset{\sim}{V} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \end{bmatrix}$$

Then, $\underset{\sim}{U}_o$ is nonexistent, and $\underset{\sim}{V}_o = [1/\sqrt{6}, 1/\sqrt{6}, -\sqrt{2/3}]^T$, a 3-vector. From the given data $\underset{\sim}{b}$, the component of $\underset{\sim}{x}$ along $\underset{\sim}{V}_o$ cannot be determined, and any value of this component will not alter the data fit, since $\underset{\sim}{A}\,\underset{\sim}{V}_o = \underset{\sim}{0}$.

Now consider the (3x2) overconstrained system $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$, with,

$$\underset{\sim}{A} = \begin{bmatrix} (2+\sqrt{3})/3 & \sqrt{2}\,(4-\sqrt{3})/6 \\ (2-\sqrt{3})/3 & \sqrt{2}\,(4+\sqrt{3})/6 \\ 2/3 & 2\sqrt{2}/3 \end{bmatrix}$$

The singular value decomposition is,

$$\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}^T = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & \sqrt{2/3} \\ \sqrt{2/3} & -1/\sqrt{3} \end{bmatrix}$$

For this system we obtain $\underset{\sim}{U}_o = [1/\sqrt{6}, 1/\sqrt{6}, -\sqrt{2/3}]^T$, a 3-vector, and

$\underset{\sim}{V}_o$ is nonexistent. Since we have L = 3, M = 2, and r = 2, it is overconstrained but not underdetermined. For existence of an exact solution, the data $\underset{\sim}{b}$ must be orthogonal to the vector $\underset{\sim}{U}_o$. If $\underset{\sim}{U}_o^T \underset{\sim}{b} = 0$, a unique exact solution exists, otherwise we obtain a unique least square solution,

$$\underset{\sim}{x} = (\underset{\sim}{A}^T\underset{\sim}{A})^{-1} \underset{\sim}{A}^T\underset{\sim}{b}$$

Lastly, as an example of the underdetermined and overconstrained system, consider the (3x2) system $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ with

$$\underset{\sim}{A} = \begin{bmatrix} 1/\sqrt{3} & \sqrt{2/3} \\ 2/\sqrt{3} & \sqrt{8/3} \\ \sqrt{2/3} & 2/\sqrt{3} \end{bmatrix}$$

yielding

$$\underset{\sim}{A} = \begin{bmatrix} 1/\sqrt{7} \\ 2/\sqrt{7} \\ \sqrt{2/7} \end{bmatrix} [\sqrt{7}\,] \; [1/\sqrt{3} \;\; \sqrt{2/3}\,]$$

In this case we have,

$$\underset{\sim}{U}_o = \begin{bmatrix} 2/\sqrt{5} & \sqrt{2/35} \\ -1/\sqrt{3} & \sqrt{8/35} \\ 0 & -\sqrt{5/7} \end{bmatrix} \quad \text{and} \quad \underset{\sim}{V}_o = \begin{bmatrix} -\sqrt{2/3} \\ 1/\sqrt{3} \end{bmatrix}$$

No exact solutions of this system exist. An infinite number of least square solutions with equal residuals exist, all differing from each other by multiples of $\underset{\sim}{V}_o$. The least square residual will be zero if $\underset{\sim}{b}$ satisfies the two linear constraints, $\underset{\sim}{U}_o^T \underset{\sim}{b} = \underset{\sim}{0}$.

After multiplying by $\underset{\sim}{U}^T$, the equation (2.2.11) reduces to

$$\underset{\sim}{\Lambda}\,\underset{\sim}{\alpha} = \underset{\sim}{\beta} \qquad (2.2.15a)$$

Noting that $\alpha_i = \underset{\sim}{v}_i^T\,\delta\underset{\sim}{\pi}$ and $\beta_i = \underset{\sim}{u}_i^T\,\delta\underset{\sim}{y}$, we obtain,

$$\lambda_i(\underset{\sim}{v}_i^T\,\delta\underset{\sim}{\pi}) = \underset{\sim}{u}_i^T\,\delta\underset{\sim}{y} \qquad i = 1, 2, \ldots, r \qquad (2.2.15b)$$

Thus, for small variations, a component of the variation $\delta\underset{\sim}{y}$ in the simulated observations along $\underset{\sim}{u}_i$ is linearly related to the component of the variation $\delta\underset{\sim}{\pi}$ in the estimate along $\underset{\sim}{v}_i$, the constant of proportionality being λ_i. For a given α_i, the larger λ_i is, the larger is β_i. Equivalently, a larger value of λ_i yields a smaller value of α_i for a given β_i. Thus, the component of the parameter vector $\underset{\sim}{\pi}$ along a vector $\underset{\sim}{v}_i$ corresponding to a small eigenvalue λ_i is poorly determined by the observed data; a small error in the observations along $\underset{\sim}{u}_i$ will cause a large error in the component of the estimate along $\underset{\sim}{v}_i$.

The perturbation equation, $\delta\underset{\sim}{y} = \underset{\sim}{A}\,\delta\underset{\sim}{\pi}$ can be interpreted as a step in the iterative process aimed at achieving a match between the observations $\underset{\sim}{y}^o$ and the computed pressures $\underset{\sim}{y}$. In this situation, we wish to make small correction $\delta\underset{\sim}{\pi}$ in $\underset{\sim}{\pi}$, leading to a variation $\delta\underset{\sim}{y}$ in $\underset{\sim}{y}$ which reduces the differences $\Delta\underset{\sim}{y} \equiv (\underset{\sim}{y} - \underset{\sim}{y}^o)$, at the current iteration. Consequently, $\delta\underset{\sim}{y} = -c\Delta\underset{\sim}{y}$ where $c \leq 1$ is a positive constant. Once such a $\delta\underset{\sim}{y}$ is selected, the simultaneous equations (2.1.3) can be solved to yield the correction $\delta\underset{\sim}{\pi}$ in the estimates at the current iteration. The foregoing analysis indicates that the components of $\Delta\underset{\sim}{y}$ along the columns of $\underset{\sim}{U}_o$ cannot be reduced by making

a small correction in the existing parameter estimate. In addition, an attempt for reduction in the component of $\Delta\underset{\sim}{y}$ along $\underset{\sim}{u}_i$ using (2.2.15) will result in a correction in $\underset{\sim}{\pi}$ along $\underset{\sim}{v}_i$ of magnitude proportional to i/λ_i . Thus the correction $\delta\underset{\sim}{\pi}$ will have dominant components along the vectors $\underset{\sim}{v}_i$ with small λ_i . If the components along $\{\underset{\sim}{v}_i\}$ with very small eigenvalues are included in the solution, it may not be acceptable due to large uncertainty. In addition, the iterative procedure for history matching would yield large alterations in the estimate at each iteration, and thus may fail to converge.

Thus, clearly, the components of $\underset{\sim}{\pi}$ along the less sensitive directions $\{\underset{\sim}{v}_i : \lambda_i < d_o,\ 0 < d_o \ll 1\}$ cannot be determined from the pressure observations alone; some additional information is necessary for their satisfactory determination. In absence of any such information, it may be expedient to refrain from making any corrections along such directions in the parameter space.

2.3 Different Inverses of the Linear Relation

As described in the foregoing sections, nonlinear parameter estimation problems are solved iteratively, the corrections at each iteration being determined through the solution of a linear subproblem. As discussed in section 2.2, the solution of the linear algebraic problem depends on the characteristics of the linear matrix operator $\underset{\sim}{A}$. In the case of the reservoir parameter estimation problem, our computational experience suggests that the linear subproblem is usually underdetermined and overconstrained, in the sense defined in section 2.2, for most ordinary parameterizations. This point is quantitatively substantiated in section 2.5. This is due to the fact

that the rank of the sensitivity matrix $\underset{\sim}{A}$ is usually much smaller than the number of parameters or the number of observations for most parameterizations. As a result, the inverse of the linear algebraic problem is not unique, and the final estimates depend on the choice of the inverse used. In view of this, we shall discuss several alternative approaches to the solution of the linear subproblem arising at each iteration and their effects on the final parameter estimates.

2.3.1 Inverse Used in the Gradient Algorithm

This algorithm attempts to minimize a scalar positive criterion, J, representing the history match error by making the estimate corrections in the direction that would yield locally the most rapid decrease in J. This is the direction of the negative gradient of the function, J, in the parameter space. Assuming the observation noise to be zero-mean with independent elements and with a uniform variance, the criterion J may be written as,

$$J = \frac{1}{2} \sum_{m=1}^{R} \sum_{n=1}^{K} \left(p_{i_m, j_n} - p^{o}_{mn} \right)^2 = \frac{1}{2} [\underset{\sim}{y}(\underset{\sim}{\pi}) - \underset{\sim}{y}^{o}]^T [\underset{\sim}{y}(\underset{\sim}{\pi}) - \underset{\sim}{y}^{o}] \qquad (2.3.1)$$

Then the gradient of J with respect to $\underset{\sim}{\pi}$ is,

$$\frac{\partial J}{\partial \underset{\sim}{\pi}} = \left[\frac{\partial \underset{\sim}{y}}{\partial \underset{\sim}{\pi}} \right]^T [\underset{\sim}{y}(\underset{\sim}{\pi}) - \underset{\sim}{y}^{o}] = \underset{\sim}{A}^T [\underset{\sim}{y}(\underset{\sim}{\pi}) - \underset{\sim}{y}^{o}] \qquad (2.3.2)$$

and the correction for the gradient algorithm is,

$$\delta \underset{\sim}{\pi} = -d \frac{\partial J}{\partial \underset{\sim}{\pi}} = -d \, \underset{\sim}{A}^T \underset{\sim}{\Delta y} \quad d > 0 \qquad (2.3.3)$$

where $\Delta \underset{\sim}{y} = [\underset{\sim}{y}(\underset{\sim}{\pi}) - \underset{\sim}{y}^{o}]$. Again, writing $\delta \underset{\sim}{y} = -c \, \Delta \underset{\sim}{y}$, and defining a new constant $d_1 = d/c$, (2.3.3) can be written as,

$$\delta \underset{\sim}{\pi} = d/c \, \underset{\sim}{A}^T \delta \underset{\sim}{y} = d_1 \, \underset{\sim}{A}^T \delta \underset{\sim}{y} \qquad (2.3.4)$$

The last expression may be taken as the inverse of the linear sub-problem (2.1.4) employed by the gradient algorithm. The linear theory suggests that a small value of the constant of proportionality d_1 be used; otherwise, a decrease in J cannot be guaranteed. In practical applications, d_1 is usually determined by a unidirectional search for the first local minimum of J in the direction of the negative gradient. The search procedure incorporates more information about the function J than is included in the gradient and its results cannot be interpreted in the framework of linearized analysis.

Using the decomposition of the sensitivity matrix $\underset{\sim}{A}$ discussed in section 2.2, we obtain

$$\delta\underset{\sim}{\pi} = d_1 \underset{\sim}{V} \underset{\sim}{\Lambda} \underset{\sim}{U}^T \delta\underset{\sim}{y} \qquad (2.3.5)$$

This expression indicates that the correction in $\underset{\sim}{\pi}$ given by the gradient algorithm contains dominant contributions along a few vectors $\underset{\sim}{v}_i$ in the $\underset{\sim}{\pi}$-space with the largest singular values λ_i. The correction made along the direction $\underset{\sim}{v}_i$ is proportional to λ_i. This scheme results in very small corrections along the vectors corresponding to very small singular values and none along the columns of $\underset{\sim}{V}_o$ It is observed, and quantitatively demonstrated later in this chapter, that the linear sub-space spanned by the less sensitive vectors of $\{\underset{\sim}{v}_i\}$ does not change significantly as the estimate, about which the linearization is done, is changed. As a result, the final estimates obtained through the gradient algorithm differ from the initial estimates only along those few directions with high sensitivity.

2.3.2 Lanczos Inverse and the Gauss-Newton Algorithm

Lanczos (1961) introduced an inverse of the linear problem (2.1.4), which is equivalent to the generalized inverse of Penrose (1955),

$$\delta\underset{\sim}{\pi} = \underset{\sim}{V}\,\underset{\sim}{\Lambda}^{-1}\,\underset{\sim}{U}^{T}\,\delta\underset{\sim}{y} \qquad (2.3.6)$$

Jackson (1972) further discussed the application of this solution to the ill-conditioned problems. Their solution is,

$$\delta\underset{\sim}{\pi} = \overline{\underset{\sim}{V}}\,\overline{\underset{\sim}{\Lambda}}^{-1}\overline{\underset{\sim}{U}}^{T}\,\delta\underset{\sim}{y} \qquad (2.3.7)$$

where $\overline{\underset{\sim}{V}}$ and $\overline{\underset{\sim}{U}}$ are matrices formed by the first ℓ colums of $\underset{\sim}{V}$ and $\underset{\sim}{U}$ respectively, and $\overline{\underset{\sim}{\Lambda}}$ is the diagonal matrix of the largest ℓ eigenvalues in $\underset{\sim}{\Lambda}$. This adaptation leads to a better conditioned solution; however, the solution lacks entirely in the components along the vectors $\{\underset{\sim}{v}_{\ell+1}, \ldots, \underset{\sim}{v}_{r}\}$ and the columns of $\underset{\sim}{V}_{o}$. The solution is dominated by the components along the vectors $\underset{\sim}{v}_{i}$ associated with the smaller among the ℓ singular values in $\overline{\underset{\sim}{\Lambda}}$. Furthermore, (2.3.7) is the solution with the smallest norm which takes into account the inverse of the largest ℓ sensitivities. The integer ℓ is selected from the consideration of the resolution and the uncertainty of the correction $\delta\underset{\sim}{\pi}$ (Jackson, 1972). The larger ℓ is, the better is the resolution as more orthogonal directions in the parameter space are included in the solution. At the same time, a larger value of ℓ implies inclusion of some additional vectors $\underset{\sim}{v}_{i}$ with even smaller eigenvalues; this results in a greater uncertainty in $\delta\underset{\sim}{\pi}$ for a given uncertainty in $\delta\underset{\sim}{y}$, and the covariance of the solution increases. (See chapter 3 for details.)

Lanczos (1961) demonstrated the equivalence of his inverse discussed above and the ordinary least square solution when the problem is over-constrained but not under-determined ($\ell = M \leq L$). Jackson (1972) demonstrated that the Lanczos inverse is equivalent to the inverse of Smith and Franklin (1969) when the problem is under-determined but not over-constrained ($\ell = L \leq M$). These statements can be readily verified by substituting the decomposition $\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Lambda}\,\underset{\sim}{V}^T$ in the respective solutions,

$$\delta\underset{\sim}{\pi} = (\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\,\underset{\sim}{A}^T\,\delta\underset{\sim}{y} \tag{2.3.8}$$

$$\delta\underset{\sim}{\pi} = \underset{\sim}{A}^T(\underset{\sim}{A}\underset{\sim}{A}^T)^{-1}\,\delta\underset{\sim}{y} \tag{2.3.9}$$

In the following, we shall demonstrate that, when $\ell = M \leq L$, the Lanczos inverse of the linearized subproblem is equivalent to the iterative correction of the Gauss-Newton algorithm for the history matching problem. In appendix 1.5 we have described the second order minimization algorithm and its Gauss-Newton approximation. It leads to the corrections determined by the linear system of algebraic equations,

$$\underset{\sim}{A}^T\underset{\sim}{A}\,\delta\underset{\sim}{\pi} = -\,\underset{\sim}{A}^T\,\Delta\underset{\sim}{y} \tag{2.3.10}$$

If we identify $\delta\underset{\sim}{y} = -\Delta\underset{\sim}{y}$, and use the singular value decomposition of $\underset{\sim}{A}$ and the orthogonality of the columns of $\underset{\sim}{U}$ and $\underset{\sim}{V}$, we obtain the solution of (2.3.10)

$$\delta\underset{\sim}{\pi} = \underset{\sim}{V}\,\underset{\sim}{\Lambda}^{-1}\,\underset{\sim}{U}^T\,\delta\underset{\sim}{y} \tag{2.3.11}$$

This is identical with the Lanczos inverse.

Now we shall briefly explore Marquardt's modification of the Gauss-Newton algorithm. This modification, also discussed at length in appendix 1.5, is designed to aid the convergence of the algorithm. It consists of adding a fixed positive constant μ to each of the diagonal elements of $\underset{\sim}{A}^T\underset{\sim}{A}$ in (2.3.10) before its solution. The result is,

$$\delta\underset{\sim}{\pi} = (\underset{\sim}{A}^T\underset{\sim}{A} + \mu\, \underset{\sim}{I})^{-1}\, \underset{\sim}{A}^T\, \delta\underset{\sim}{y} \qquad (2.3.12)$$

Using the singular value decomposition of $\underset{\sim}{A}$ and the orthogonality of $\underset{\sim}{V}$ when $\ell = M$, we obtain

$$\underset{\sim}{V}^T\, \delta\underset{\sim}{\pi} = (\underset{\sim}{\Lambda}^2+\mu\underset{\sim}{I})^{-1}\, \underset{\sim}{\Lambda}\, \underset{\sim}{U}^T\, \delta\underset{\sim}{y} \qquad (2.3.13a)$$

or component-wise,

$$\underset{\sim}{v}_i^T\, \delta\underset{\sim}{\pi} = \frac{\lambda_i}{\lambda_i^2+\mu}\, \underset{\sim}{u}_i^T\, \delta\underset{\sim}{y} \qquad (2.3.13b)$$

From the last relation it is clear that the component of the correction $\delta\underset{\sim}{\pi}$ along the vector $\underset{\sim}{v}_i$ in the parameter space will be small if either λ_i is very small or very large. For a given $\delta\underset{\sim}{y}$, the component of $\delta\pi$ along the vector $\underset{\sim}{v}_i$ with eigenvalue closest to $\sqrt{\mu}$ will be dominant. Thus the results of Marquardt's method are intermediate to those obtained from Jackson's modification (2.3.7) of the Lanczos inverse by rounding off the smaller eigenvalues to zero, and the Gauss-Newton method or equivalently the unmodified inverse. The components of Marquardt's correction along $\{\underset{\sim}{v}_i\}$ corresponding to very small $\{\lambda_i\}$ will be close to zero, whereas both the Gauss-Newton and the Lanczos solutions have dominant components along the $\underset{\sim}{v}_i$

with the smallest of the non-zero eigenvalues. In Marquardt's method, we can "tune" the corrections at each iteration by choosing the parameter μ, and thus have some control over the final estimates of the history matching process. The parameter μ, however, is not totally free because it has to be such that the iterative process converges. We shall have more to say about the choice of the parameter μ in chapter 3.

2.4 Some Analytical Results Concerning Sensitivity

In this section we shall attempt to gain some qualitative understanding of the sensitivity of the observed pressure with respect to the rock properties at different locations in a given reservoir. For this purpose, we shall derive analytical expressions for the sensitivity of the pressure at a given location and time, in a single phase reservoir with uniform thickness h, with respect to the permeability k and porosity ϕ. For simplicity, we shall assume the reservoir to be one-dimensional, with spatial domain $\{x \in (0, L)\}$. A production well with constant production rate is situated at $x = a < L$, and the pressure observation site is at $x = b < L$. In order to gain insight into the differences arising out of different boundary conditions, we shall treat reservoirs with two kinds of boundary conditions: impermeable boundaries and constant pressure boundaries.

As mentioned above, we wish to determine the sensitivity coefficients $\partial p(b, t)/\partial k(x_1)$ and $\partial p(b, t)/\partial \phi(x_1)$ for the values of $x_1 \in (0, L)$. If we let x_1 be a variable, and denote it by $x \in (0, L)$, then these quantities are functions of x, defined over (0, L). These functions are

called the functional derivatives of p(b, t) w. r. t. k(x) and $\phi(x)$ respectively, and are denoted by the shorthand notations

$$\frac{\delta p(b,t)}{\delta k(x)} \text{ and } \frac{\delta p(b,t)}{\delta \phi(x)} .$$

The functional derivatives have a property that, the change $\delta p(b,t)$ in p(b,t) due to a change $\delta k(x)$ is given by:

$$\delta p(b,t)\Big|_{\delta k(x)} = \int_o^L \frac{\delta p(b,t)}{\delta k(x)} \delta k(x) dx \qquad (2.4.1)$$

A similar expression involving $\frac{\delta p(b,t)}{\delta \phi(x)}$ yields $\delta p(b, t)$ due to a change $\delta\phi(x)$ in $\phi(x)$.

The expressions for the sensitivity coefficients for a spatially discrete ordinary differential equation model of a single-phase reservoir were first developed by Jacquard and Jain (1965) by analogy with the electric circuit theory. Carter et al. (1974) treated the continuous model described by a partial differential equation and derived the expressions for the functional derivatives, using a reciprocity property of certain convolution integrals of the solution of a forced linear parabolic differential equation. We shall rederive their results using a simpler variational approach, which is applicable to all linear and nonlinear dynamic systems—parabolic, hyperbolic, as well as the lumped parameter systems described by the ordinary differential equations. An application of this approach to a discrete system has already appeared in appendix 1.4. In the following we present the derivation for the one-dimensional single-phase reservoir.

2.4.1 Reservoir with Impermeable Boundaries

At first, we shall derive expressions for the sensitivity coefficients.

The equations describing the reservoir are,

$$\phi(x)\frac{\partial p}{\partial t} = \frac{1}{\mu c}\frac{\partial}{\partial x}\left(k(x)\frac{\partial p}{\partial x}\right) + \frac{q(t)}{ch}\delta(x-a) \qquad (2.4.2)$$

$$\frac{\partial p}{\partial x} = 0 \qquad x = 0,\ L,\ \forall t \qquad (2.4.3)$$

$$p(x, 0) = 0 \qquad (2.4.4)$$

Consider the variations $\delta\phi(x)$ and $\delta k(x)$ in the rock properties, and let $\delta p(x, t)$ be the resultant pressure variations. Then for small variations, we have the linear equations,

$$\phi(x)\frac{\partial \delta p}{\partial t} = \frac{1}{\mu c}\frac{\partial}{\partial x}\left(k(x)\frac{\partial \delta p}{\partial x}\right) - \delta\phi(x)\frac{\partial p}{\partial t} + \frac{1}{\mu c}\frac{1}{\partial x}\left(\delta k(x)\frac{\partial p}{\partial x}\right) \qquad (2.4.5a)$$

$$\frac{\partial \delta p}{\partial x} = 0 \qquad x = 0,\ L,\ \forall t \qquad (2.4.5b)$$

$$\delta p(x, 0) = 0 \qquad (2.4.5c)$$

Multiplying (2.4.5a) by an arbitrary function $\psi(x, t)$ and integrating w.r.t. x over (0, L) and w.r.t. t over (0, T), we obtain

$$\int_0^L\int_0^T \phi(x)\frac{\partial \delta p}{\partial t}\psi(x, t)dt\,dx = \frac{1}{\mu c}\int_0^T\int_0^L \frac{\partial}{\partial x}\left(k(x)\frac{\partial \delta p}{\partial x}\right)\psi(x, t)dx\,dt$$
$$-\int_0^T\int_0^L \delta\phi(x)\frac{\partial p}{\partial t}\psi(x,t)dxdt + \frac{1}{\mu c}\int_0^T\int_0^L \frac{\partial}{\partial x}\left(\delta k(x)\frac{\partial p}{\partial x}\right)\psi(x, t)dx\,dt \qquad (2.4.6)$$

Using integration by parts w.r.t. t, we obtain,

$$\int_0^L\int_0^T \phi(x)\frac{\partial \delta p}{\partial t}\psi(x, t)dt\,dx = \int_0^L \phi(x)\,\delta p(x, T)\,\psi(x, T)\,dx$$
$$-\int_0^T\int_0^L \phi(x)\frac{\partial \psi}{\partial t}(x, t)\,\delta p(x, t)dtdx \qquad (2.4.7)$$

where we have used (2.4.5c).

Integrating by parts w.r.t. x twice, we obtain

$$\int_0^T\int_0^L \frac{\partial}{\partial x}\left(k(x)\frac{\partial \delta p}{\partial x}\right)\psi(x,t)\,dx\,dt = \int_0^T \left\{k(x)\frac{\partial \delta p}{\partial x}\psi(x,t)\Big|_{x=L} - k(x)\frac{\partial \delta p}{\partial x}\psi(x,t)\Big|_{x=0}\right\} dt$$

$$- \int_0^T \left\{k(x)\delta p\frac{\partial \psi}{\partial x}\Big|_{x=L} - k(x)\,\delta p\frac{\partial \psi}{\partial x}\Big|_{x=0}\right\} dt$$

$$+ \int_0^T\int_0^L \frac{\partial}{\partial x}\left(k(x)\frac{\partial \psi}{\partial x}\right)\delta p(x,t)\,dt\,dx \qquad (2.4.8)$$

The first two terms on the right of (2.4.8) vanish due to (2.4.5b).

Finally, a single integration by parts w.r.t. x leads to,

$$\int_0^T\int_0^L \frac{\partial}{\partial x}\left(\delta k(x)\frac{\partial p}{\partial x}\right)\psi(x,t)\,dx\,dt = \int_0^T \left\{\delta k(x)\frac{\partial p}{\partial x}\psi\Big|_{x=L} - \delta k(x)\frac{\partial p}{\partial x}\psi\Big|_{x=0}\right\} dt$$

$$- \int_0^T\int_0^L \delta k(x)\frac{\partial p}{\partial x}\frac{\partial \psi}{\partial x}\,dx\,dt \qquad (2.4.9)$$

Again, the first two terms on the right vanish on account of (2.4.3). Note that in problems with more than one spatial dimension, Green's theorem will have to be used instead of the integration by parts in arriving at results analogous to (2.4.8-9). Substituting (2.4.7-9) into (2.4.6) and rearranging, we obtain

$$\int_0^T\int_0^L \left[-\phi(x)\frac{\partial \psi}{\partial t} - \frac{1}{\mu c}\frac{\partial}{\partial x}\left(k(x)\frac{\partial \psi}{\partial x}\right)\right]\delta p(x,t)\,dt\,dx$$

$$+ \int_0^L \phi(x)\delta p(x,T)\psi(x,T)\,dx = -\frac{1}{\mu c}\int_0^T \left\{k(x)\frac{\partial \psi}{\partial x}\,\delta p\Big|_{x=L} - k(x)\frac{\partial \psi}{\partial x}\,\delta p\Big|_{x=0}\right\} dt$$

$$- \int_0^L\int_0^T \delta\phi(x)\frac{\partial p}{\partial t}\psi(x,t)\,dt\,dx - \frac{1}{\mu c}\int_0^T\int_0^L \frac{\partial p}{\partial x}\frac{\partial \psi}{\partial x}\,\delta k(x)\,dx\,dt \qquad (2.4.10)$$

Let us now restrict $\psi(x,t)$ which is entirely arbitrary so far, to satisfy the following system of equations:

$$\phi(x)\frac{\partial\psi}{\partial t} = -\frac{1}{\mu c}\frac{\partial}{\partial x}\left(k(x)\frac{\partial\psi}{\partial x}\right) \qquad (2.4.11)$$

$$\frac{\partial\psi}{\partial x} = 0 \qquad x = 0, L,\ t \in [0, T] \qquad (2.4.12)$$

$$\psi(x, T) = \delta(x-b) \qquad (2.4.13)$$

Note that these define an initial value-boundary problem very similar to the original system problem and may be integrated backwards in time to yield a unique solution $\psi(x, t)$ for $t \in [0, T]$. This problem will be referred to as the adjoint problem; and $\psi(x, t)$ will be called the adjoint variable. With this definition of $\psi(x, t)$, we obtain from (2.4.10),

$$\phi(b)\delta p(b,T) = -\int_0^L \delta\phi(x) \int_0^T \frac{\partial p}{\partial t}\,\psi\, dt\, dx - \frac{1}{\mu c}\int_0^L \delta k(x) \int_0^T \frac{\partial p}{\partial x}\frac{\partial\psi}{\partial x}\, dt\, dx \qquad (2.4.14)$$

From (2.4.14), we can easily obtain the expressions for the functional derivatives for the sensitivity coefficients,

$$\frac{\delta p(b, T)}{\delta\phi(x)} = -\frac{1}{\phi(b)}\int_0^T \frac{\partial p(x, t)}{\partial t}\psi(x, t)\, dt \qquad (2.4.15)$$

and

$$\frac{\delta p(b, T)}{\partial k(x)} = -\frac{1}{\phi(b)\mu c}\int_0^T \frac{\partial p(x, t)}{\partial x}\frac{\partial\psi(x, t)}{\partial x}\, dt \qquad (2.4.16)$$

From these expressions, it is evident that we need the solutions $p(x, t)$ and $\psi(x, t)$ in analytical form, in order to analytically evaluate the sensitivity coefficients. This requirement necessitates that the distributions $k(x)$ and $\phi(x)$ be such that closed form solutions of (2.4.1-3) and (2.4.11-13) can be obtained. Simple distributions such as uniform and linear variation satisfy this requirement. Since the

analytical treatment is the simplest for uniform distributions of k and ϕ, and the results will have the same qualitative behavior as in other cases, these are selected for the following analysis. Let us assume q to be independent of time, and that q and t are scaled such that the system equations reduce to,

$$\frac{\partial p}{\partial t} = \frac{\partial^2 p}{\partial x^2} + q\,\delta(x-a) \qquad (2.4.17)$$

$$\frac{\partial p}{\partial x} = 0 \quad \text{at} \quad x = 0, L \qquad (2.4.18)$$

$$p(x, 0) = 0 \qquad (2.4.19)$$

We wish to obtain the sensitivity coefficients for p(b, T).

The solution of (2.4.17-19) can be easily obtained using the cosine transforms as follows. Defining the transforms,

$$\tilde{p}_i(t) = \int_0^L p(x, t) \cos\left(\frac{i\pi x}{L}\right) dx \qquad i = 0, 1, 2, \ldots. \qquad (2.4.20)$$

the solution p(x, t) can be expressed in a Fourier cosine series,

$$p(x, t) = \frac{2}{L} \sum_{i=1}^{\infty} \tilde{p}_i(t) \cos\frac{i\pi x}{L} + \frac{1}{L}\tilde{p}_o(t) \qquad (2.4.21)$$

Multiplying (2.4.17) by $\cos\frac{i\pi x}{L}$ and integrating over (0, L) and using (2.4.18), we obtain the ordinary differential equation governing $\tilde{p}_i(t)$,

$$\frac{d\tilde{p}_i(t)}{dt} = -\left(\frac{i\pi}{L}\right)^2 \tilde{p}_i + q \cos\frac{i\pi a}{L} \qquad (2.4.22)$$

$$\frac{d\tilde{p}_o}{dt} = q \qquad (2.4.23)$$

The initial conditions are,

$$\tilde{p}_i(0) = 0 \qquad i = 0, 1, 2, \ldots \qquad (2.4.24)$$

The solutions of (2.4.22-24) are,

$$\tilde{p}_i(t) = q\left(\frac{L}{i\pi}\right)^2 \cos\left(\frac{i\pi a}{L}\right)\left[1 - e^{-(i\pi/L)^2 t}\right] \qquad (2.4.25)$$

$$\tilde{p}_o(t) = q\,t \qquad (2.4.26)$$

Then we have

$$p(x,t) = \frac{2qL}{\pi^2}\sum_{i=1}^{\infty}\frac{1}{i^2}\cos\left(\frac{i\pi a}{L}\right)\cos\left(\frac{i\pi x}{L}\right)\left[1-e^{-(i\pi/L)^2 t}\right] + \frac{qt}{L} \qquad (2.4.27)$$

To obtain the solution of the adjoint equations (2.4.11-14) for these ϕ and k, we make a change of variable $\xi = (T-t)$, yielding for $\xi \in (0, T)$,

$$\frac{\partial \psi}{\partial \xi} = \frac{\partial^2 \psi}{\partial x^2}, \qquad \psi(x,0) = \delta(x-b), \frac{\partial \psi}{\partial x} = 0 \quad x = 0, L \qquad (2.4.28)$$

Then the solution of (2.4.28) can be obtained through cosine transforms in a manner very similar to that described for p(x, t). The solution is,

$$\psi(x,\xi) = \frac{2}{L}\sum_{j=1}^{\infty}\cos\frac{j\pi b}{L}\cos\frac{j\pi x}{L}\, e^{-(j\pi/L)^2\xi} + \frac{1}{L} \qquad (2.4.29)$$

or

$$\psi(x,t) = \frac{2}{L}\sum_{j=1}^{\infty}\cos\frac{j\pi b}{L}\cos\frac{j\pi x}{L}\, e^{-(j\pi/L)^2(T-t)} + \frac{1}{L} \qquad (2.4.30)$$

Substituting these expressions in (2.4.15-16), we obtain

$$\begin{aligned}\frac{\delta p(b,T)}{\delta k(x)} = &-\frac{4q}{\pi^2}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{ij}\,\theta(i,j;a,b)\left[1-e^{-(j\pi/L)^2 T}\right] \\ &+\frac{4q}{\pi^2}\sum_{\substack{i=1\\ i\neq j}}^{\infty}\sum_{j=1}^{\infty}\frac{j\,\theta(i,j;a,b)}{i(j^2-i^2)}\left[e^{-(i\pi/L)^2 T} - e^{-(j\pi/L)^2 T}\right] \\ &+\frac{4qT}{L^2}\sum_{i=1}^{\infty}\theta(i,i;a,b)\,e^{-(i\pi/L)^2 T}\end{aligned} \qquad (2.4.31)$$

where,

$$\theta(i,j;a,b) = \sin\frac{inx}{L}\sin\frac{j\pi x}{L}\cos\frac{i\pi a}{L}\cos\frac{j\pi b}{L}$$

and

$$\frac{\delta p(b,T)}{\delta \phi(x)} = \frac{qT}{L^2} + \frac{2q}{\pi^2}\sum_{i=1}^{\infty}\frac{1}{i^2}\left(\cos\frac{i\pi a}{L}+\cos\frac{i\pi b}{L}\right)\cos\frac{i\pi x}{L}\left[1-e^{-(i\pi/L)^2 T}\right]$$

$$+\frac{4q}{\pi^2}\sum_{i=1}^{\infty}\sum_{\substack{j=1\\ l\neq j}}^{\infty}\frac{1}{(j^2-i^2)}\cos\frac{i\pi a}{L}\cos\frac{i\pi x}{L}\cos\frac{j\pi b}{L}\cos\frac{j\pi x}{L}\left[e^{-(i\pi/L)^2 T}-e^{-(j\pi/L)^2 T}\right]$$

$$+\frac{4qT}{L^2}\sum_{i=1}^{\infty}\cos\frac{i\pi a}{L}\cos\frac{i\pi b}{L}\cos^2\left(\frac{i\pi x}{L}\right)e^{-(i\pi/L)^2 T} \qquad (2.4.32)$$

From the property (2.4.1) of the functional derivative, we can obtain the change in the observed pressure due to an arbitrary perturbation $\delta k(x)$ in k. Decomposing $\delta k(x)$ into a Fourier cosine series, we have

$$\delta k(x) = \sum_{\ell=0}^{\infty}\epsilon_\ell \cos\frac{\ell\pi x}{L} \equiv \sum_{\ell=0}^{\infty}\delta k_\ell \qquad (2.4.33)$$

The contribution to $\delta p(b,T)$ due to the ℓth Fourier component δk_ℓ of $\delta k(x)$ in (2.4.33) can be obtained from (2.4.1) and (2.4.31) as,

$$\frac{\pi^2}{q\epsilon_\ell L}\delta p(b,T)\Big|_{\delta k_\ell} = -\sum_{i=1}^{\infty}\frac{1}{i(i+\ell)}\theta_1(i,i+\ell)\left[1-e^{-((i+\ell)\pi/L)^2 T}\right]$$

$$-\sum_{i=1}^{\infty}\frac{1}{i(i+\ell)}\theta_1(i+\ell,i)\left[1-e^{-(i\pi/L)^2 T}\right]$$

$$+\sum_{\substack{i=1\\ i+j=\ell}}^{\ell-1}\sum_{j=1}^{\ell-1}\frac{1}{ij}\theta_1(i,j)\left[1-e^{-(j\pi/L)^2 T}\right]$$

$$+\sum_{i=1}^{\infty}\frac{1}{(\ell^2+2i\ell)}\left[e^{-(i\pi/L)^2 T}-e^{-((i+\ell)\pi/L)^2 T}\right]\left[\frac{i+\ell}{i}\theta_1(i,i+\ell)+\frac{i}{i+\ell}\theta_1(i+\ell,i)\right]$$

$$+\sum_{\substack{i=1\\ i\neq j,\ i+j=\ell}}^{\ell-1}\sum_{j=1}^{\ell-1}\frac{j}{i\ell(j-i)}\theta_1(i,j)\left[e^{-(i\pi/L)^2 T}-e^{-(j\pi/L)^2 T}\right]$$

$$-\frac{\pi^2 T}{L^2}\sum_{i=1}^{\infty}\theta_1(i,i)e^{-(i\pi/L)^2 T}\delta_{2i,\ell} \qquad (\ell\neq 0) \qquad (2.4.34a)$$

$$\frac{\pi^2}{2q\, e_o L}\, \delta p(b, T)\Big|_{\delta k_o} = -\sum_{i=1}^{\infty} \frac{i}{i^2}\, \theta_1(i,i)\,[1-e^{-(i\pi/L)^2 T}]$$

$$-\frac{\pi^2 T}{L^2} \sum_{i=1}^{\infty} \theta_1(i,i) e^{-(i\pi/L)^2 T} \quad (\ell = 0) \qquad (2.4.34b)$$

where $\theta_1(i,j) = \cos\frac{i\pi a}{L}\cos\frac{j\pi b}{L}$.

In the last expression all the terms are smaller than unity in absolute value and have signs alternating in a regular fashion. The series are absolutely convergent for $T > 0$. It follows that for large T,

$$\left.\begin{aligned} \delta p(b, T)\Big|_{\delta k_\ell} &= O(\epsilon_\ell \frac{qL}{\pi^2}\frac{1}{\ell}) \qquad \ell \neq 0 \\ \delta p(b, T)\Big|_{\delta k_o} &= O(\epsilon_o \frac{2qL}{\pi^2}) \qquad \ell = 0 \end{aligned}\right\} \qquad (2.4.35)$$

This indicates that the observed pressures have vanishingly small sensitivity to highly oscillatory components of k(x), and consequently it is not possible to determine accurately such components through history matching.

In order to investigate the component-wise sensitivity of the porosity $\phi(x)$, let us express the variation $\delta\phi(x)$ in a Fourier cosine series,

$$\delta\phi(x) = e_o + \sum_{\ell=1}^{\infty} e_\ell \cos\frac{\ell\pi x}{L} \equiv \sum_{\ell=0}^{\infty} \delta\phi_\ell \qquad (2.4.36)$$

Before we carry out the calculations, from (2.4.32) it is evident that the contribution to $\delta p(b, T)$ due to the component e_o of $\delta\phi(x)$, which represents the mean value of the perturbation in ϕ, will grow linearly with time for large T. Then the sensitivity of the observed pressure to changes in the mean value of ϕ is large and increases with time.

Consequently, the mean value of ϕ can be determined with considerable accuracy through history matching. It is important to note that this conclusion holds only in the case of reservoirs with impermeable boundaries. In such reservoirs the pressure levels are strongly dependent on the cumulative production and the amount of oil originally present, the latter being determined by the mean value of porosity.

A treatment similar to that used for arriving at (2.4.34) leads to the following expression for the change in the observed pressure due to the ℓth Fourier component ($\ell \neq 0$) of $\delta\phi(x)$,

$$\frac{1}{e_\ell}\,\delta p(b,T)\Big|_{\delta\phi_\ell} = \frac{qL}{\pi^2\ell^2}\left(\cos\frac{\ell\pi a}{L}+\cos\frac{\ell\pi b}{L}\right)(1-e^{-(\ell\pi/L)^2T}) + \frac{qT}{L}\sum_{i=1}^{\infty}\cos\frac{i\pi a}{L}\cos\frac{i\pi b}{L}\,e^{-(i\pi/L)^2T}\,\delta_{2i,\ell}$$

$$+\frac{qL}{\pi^2}\sum_{i=1}^{\infty}[e^{-(i\pi/L)^2T}-e^{-((i+\ell)\pi/L)^2T}]\frac{[\cos\frac{i\pi a}{L}\cos\frac{(i+\ell)\pi b}{L}+\cos\frac{(i+\ell)\pi a}{L}\cos\frac{i\pi b}{L}]}{(\ell^2+2i\ell)}$$

$$+\frac{qL}{\pi^2}\sum_{\substack{i=1\\ i\neq j,\ i+j=\ell}}^{\ell-1}\sum_{j=1}^{\ell-1}\frac{\cos\frac{i\pi a}{L}\cos\frac{j\pi b}{L}}{\ell(j-i)}[e^{-(i\pi/L)^2T}-e^{-(j\pi/L)^2T}] \qquad (2.4.37)$$

Then for large T ($\ell \neq 0$),

$$\delta p(b,T)\Big|_{\delta\phi_\ell} = O\left(e_\ell\frac{qL}{\pi^2}\frac{1}{\ell^2}\right) \qquad (2.4.38)$$

Thus, the sensitivity of observed pressure with respect to the highly oscillatory components of $\phi(x)$ is very small. A comparison of (2.4.35) and (2.4.38) reveals that a Fourier component of $\phi(x)$ with non-zero wave number ℓ has much smaller influence than a component of $k(x)$ with the same wave number, the difference increasing with ℓ.

2.4.2 Reservoir with Constant Pressure Boundaries

The pressure in this case is determined by equations (2.4.2) and (2.4.4), with the boundary conditions (2.4.3) replaced by

$$p(x, t) = 0 \qquad x = 0, L, \ \forall\, t \tag{2.4.39}$$

The derivation of the formulas for the functional derivatives $\delta p(b, t)/\delta k(x)$ and $\delta p(b, t)/\delta \phi(x)$ proceeds in a manner very similar to that for reservoir with impermeable boundaries. The adjoint systems equations in this case are (2.4.11), (2.4.13) and with an altered boundary condition,

$$\psi(x, t) = 0 \qquad x = 0, L \qquad t \in [0, T]. \tag{2.4.40}$$

The expressions for the functional derivatives (2.4.15) and (2.4.16) remain unaltered.

Again, for analytic simplicity, we shall treat the case of a one-dimensional reservoir with uniform distributions of k and ϕ. Let the pressure in the reservoir be governed by (2.4.17), (2.4.19) and the boundary condition (2.4.39). The solution can be easily obtained by use of the sine transforms as

$$p(x, t) = \frac{2qL}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{i^2} \left[1 - e^{-(i\pi/L)^2 t}\right] \sin\frac{i\pi a}{L} \sin\frac{i\pi x}{L} \tag{2.4.41}$$

Similarly, the solution of the adjoint system (2.4.11), (2.4.13) and (2.4.40) for this simple case is,

$$\psi(x, t) = \frac{2}{L} \sum_{j=1}^{\infty} \sin\frac{j\pi b}{L} \sin\frac{j\pi x}{L} e^{(j\pi/L)^2 (t-T)} \quad t \in (0, T) \tag{2.4.42}$$

Substitution of these solutions into (2. 4. 15) and (2. 4. 16) and evaluation of the time integrals yields,

$$\frac{\delta p(b,T)}{\delta k(x)} = -\frac{4q}{\pi^2}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{ij}\theta_2(i,j;a,b)\left[1-e^{-(j\pi/L)^2T}\right]$$

$$+\frac{4q}{\pi^2}\sum_{\substack{i=1\\ i\neq j}}^{\infty}\sum_{j=1}^{\infty}\frac{j}{i(j^2-i^2)}\theta_2(i,j;a,b)\left[e^{-(i\pi/L)^2T}-e^{-(j\pi/L)^2T}\right]$$

$$+\frac{4qT}{L^2}\sum_{i=1}^{\infty}\theta_2(i,i;a,b)e^{-(i\pi/L)^2T} \qquad (2.4.43)$$

and

$$\frac{\delta p(b,T)}{\delta \phi(x)} = \frac{4q}{\pi^2}\sum_{\substack{i=1\\ i\neq j}}^{\infty}\sum_{j=1}^{\infty}\frac{1}{(j^2-i^2)}\theta_3(i,j;a,b)\left[e^{-(i\pi/L)^2T}-e^{-(j\pi/L)^2T}\right]$$

$$+\frac{4qT}{L^2}\sum_{i=1}^{\infty}\theta_3(i,i;a,b)e^{-(i\pi/L)^2T} \qquad (2.4.44)$$

where we have defined for brevity

$$\theta_2(i,j;a,b) = \cos\frac{i\pi x}{L}\cos\frac{j\pi x}{L}\sin\frac{i\pi a}{L}\sin\frac{j\pi b}{L}$$

$$\theta_3(i,j;a,b) = \sin\frac{i\pi x}{L}\sin\frac{j\pi x}{L}\sin\frac{i\pi a}{L}\sin\frac{j\pi b}{L}.$$

Then the change $\delta p(b,T)$ due to the ℓth Fourier component in the cosine series (2. 4. 33) for $\delta k(x)$ can be obtained from (2. 4. 1) and (2. 4. 43) as,

$$\frac{1}{\epsilon_\ell}\delta p(b,T)\Big|_{\delta k_\ell} = -\frac{qL}{\pi^2}\sum_{\substack{i=1\\ i+j=\ell}}^{\ell-1}\sum_{j=1}^{\ell-1}\frac{1}{ij}\theta_4(i,j)\left[1-e^{-(j\pi/L)^2T}\right]$$

$$-\frac{qL}{T^2}\sum_{i=1}^{\infty}\frac{1}{(i+\ell)i}\theta_4(i,i+\ell)\left[1-e^{-((i+\ell)\pi/L)^2T}\right.$$

$$-\frac{qL}{\pi^2}\sum_{i=1}^{\infty}\frac{1}{(i+\ell)i}\theta_4(i+\ell,i)\left[1-e^{-(i\pi/L)^2T}\right]$$

$$+\frac{qL}{\pi^2}\sum_{\substack{i=1\\ i\neq j,\ i\neq j=\ell}}^{\ell-1}\sum_{j=1}^{\ell-1}\frac{j}{i\ell(j-i)}\theta_4(i,j)\left[e^{-(i\pi/L)^2T}-e^{-(j\pi/L)^2T}\right]$$

(Cont'd on next page)

$$+\frac{qL}{\pi^2}\sum_{i=1}^{\infty}\left\{\frac{1}{\ell^2+2i\ell}\left[\frac{i+\ell}{i}\,\theta_4(i,i+\ell)+\frac{i}{i+\ell}\,\theta_4(i+\ell,i)\right]\right.$$

$$\left.\left[e^{-(i\pi/L)^2T}-e^{-((i+\ell)\pi/L)^2T}\right]\right\}$$

$$+\frac{qT}{L}\sum_{i=1}^{\infty}\theta_4(i,i)e^{-(i\pi/L)^2T}\,\delta_{2i,\ell}\quad(\ell\neq 0)\qquad(2.4.45a)$$

and

$$\frac{1}{\epsilon_o}\,\delta p(b,T)\Big|_{\delta k_o}=\frac{-2qL}{\pi^2}\sum_{i=1}^{\infty}\frac{1}{i^2}\,\theta_4(i,i)\left[1-e^{-(i\pi/L)^2T}\right]$$

$$+\frac{2qT}{L}\sum_{i=1}^{\infty}\theta_4(i,i)e^{-(i\pi/L)^2T}\qquad(2.4.45b)$$

where

$$\theta_4(i,j)=\sin\frac{i\pi a}{L}\,\sin\frac{j\pi b}{L}\qquad(2.4.46)$$

Similarly from (2.4.44), the pressure change $\delta p(b,T)$ due to the ℓth Fourier component in cosine series (2.4.36) for $\delta\phi$ is,

$$\frac{1}{e_\ell}\,\delta p(b,T)\Big|_{\delta\phi_\ell}=\frac{qL}{\pi^2}\sum_{i=1}^{\infty}\left\{\frac{1}{\ell^2+2i\ell}\left[\theta_4(i,i+\ell)+\theta_4(i+\ell,i)\right]\right.$$

$$\left.\left[e^{-(i\pi/L)^2T}-e^{-((i+\ell)\pi/L)^2T}\right]\right\}$$

$$-\frac{qL}{\pi^2}\sum_{\substack{i=1\\ i\neq j,\ i+j=\ell}}^{\ell-1}\sum_{j=1}^{\ell-1}\frac{1}{\ell(j-i)}\,\theta_4(i,j)\left[e^{-(i\pi/L)^2T}-e^{-(j\pi/L)^2T}\right]$$

$$-\frac{qT}{L}\sum_{i=1}^{\infty}\theta_4(i,i)\delta_{2i,\ell}\,e^{-(i\pi/L)^2T}\quad(\ell\neq 0)\qquad(2.4.47a)$$

and

$$\frac{1}{e_o}\,\delta p(b,T)\Big|_{\delta\phi_o}=\frac{2qT}{L^2}\sum_{i=1}^{\infty}\theta_4(i,j)e^{-(i\pi/L)^2T}\qquad(2.4.47b)$$

where $\theta_4(i,j)$ is the function defined in (2.4.46).

From (2.4.45a, b) we have for $T \to \infty$,

$$\left.\delta p(b, T)\right|_{\delta k_\ell} = O\left(\epsilon_\ell \frac{qL}{\pi^2} \frac{1}{\ell}\right) \ell \neq 0$$
$$= O\left(\epsilon_o \frac{2q\ell}{\pi^2}\right) \quad \ell = 0 \qquad (2.4.48)$$

Similarly from (2.4.47a, b) we obtain for large T,

$$\left.\delta p(b, T)\right|_{\delta \phi_\ell} = O\left(e_\ell \frac{qL}{\pi^2} \frac{1}{\ell^2} e^{-(\pi/L)^2 T}\right) \quad \ell \neq 0$$
$$= O\left(e_o \frac{2qT}{L^2} e^{-(\pi/L)^2 T}\right) \quad \ell = 0 \qquad (2.4.49)$$

A comparison of (2.4.48) and (2.4.35) reveals that the sensitivity of an observed pressure with respect to a Fourier component of k(x) with wavenumber ℓ has identical behavior for the reservoirs with either impermeable boundaries or constant pressure boundaries. The conclusions made earlier about the identifiability of the different component of k(x) from the pressure observations hold for both kinds of reservoirs.

On the other hand, a comparison of (2.4.49) and (2.4.38) reveals that the behaviors of the sensitivity of p(b, T) with respect to different Fourier components of $\phi(x)$ is radically different for the two kinds of reservoirs. In the reservoirs with impermeable boundaries the influence of $\delta\phi_\ell(x)$ on p(b, T) is finite, $O(\ell^{-2})$, for large T; whereas in the constant pressure boundary reservoirs, this influence is vanishing for large T. The difference is even more striking for the component $\ell = 0$, which corresponds to a change in the mean value of $\phi(x)$: in impermeable boundary reservoirs this component has an influence which linearly increases with time and thus dominates the

rest, whereas in the constant pressure boundary reservoirs, this component has vanishingly small influence for large T. The component $\delta\phi_0$ has a larger influence than $\delta\phi_\ell (\ell \neq 0)$ in either case.

From the above analysis we conclude that in the case of the constant pressure boundary reservoir, the problem of estimating k(x) and $\phi(x)$ from observed pressures is more ill-conditioned than the corresponding problem for impermeable boundary reservoirs. The porosity has vanishingly small influence on the observation at large times; and its various Fourier components, including the mean, cannot be determined accurately from the observed pressures alone. The permeability has a finite influence on the observation at all times; however, as the Fourier component with wave number ℓ has $O(\ell^{-1})$ influence at large times, hence the highly oscillatory components of k(x) cannot be estimated accurately from the observations.

Although it is not evident from the foregoing analysis, we note that for the constant pressure boundary reservoirs, the flux and the pressure gradients in regions close to the boundary are of the same order as elsewhere in the reservoir and thus the influence on the observations of the permeability in these regions is not different from elsewhere. On the other hand, as both the flux and the pressure gradients are very small in the vicinity of the impermeable boundaries, k(x) in these regions has small influence on the observed pressures compared to elsewhere in the reservoir. Hence, we expect to be able to determine k(x) more accurately in the region close to constant pressure boundaries compared with the regions near the impermeable boundaries.

In this section, we have attempted to analyze the sensitivity of the observed pressure at a single location and a single time instant. For accurate overall estimation of parameters, it is necessary but not sufficient that the sensitivities of individual pressure observations with respect to the different components of the parameters be large. In such a problem the number of independent degrees of freedom contained in the data is equally important. A sensitivity matrix with large elements can have a smaller rank than another with smaller elements. Hence such analysis of the sensitivity matrix is crucial to the study of the parameter estimation problem. In the next section, we carry out such analysis numerically and present typical results.

2.5 Numerical Results on the Reservoir Parameter Sensitivities

In this section we present numerical results on the linearized problem in order to elucidate more fully the nature of the reservoir parameter estimation problem.

The one-dimensional reservoir covered by a uniform grid with 33 nodes described in chapter 1, with impermeable boundaries is considered. In the case with no constraints, the number of parameters to be estimated is 66, each of the 33 grid point values of the permeability k and the porosity ϕ. To bring out the similarities and the differences arising due to the use of different parameter distributions for linearization, the problems associated with the two different sets of k(x) and ϕ(x) are presented. These distributions are, the realization R-2 of the homogeneous random process described in chapter 1, and the uniform distributions of k(x) = 5.0 md and ϕ(x) = 0.2. The conditions

of the estimation problem are those of the set S_2 defined in chapter 1; they are, two production wells with piece-wise constant production rates, three observation locations, each with 25 observations at uniform time intervals.

There are 75 observations, yielding the 75-vector $\underset{\sim}{y}$, and the sensitivity matrix $\underset{\sim}{A}$ is (75x66). The elements of $\underset{\sim}{A}$, the sensitivity coefficients were computed using the adjoint variable method derived in appendix 1.4. The singular value decomposition of $\underset{\sim}{A}$ was carried out using the double-precision, real version of the program by Bushinger and Golub (1969).

The 66 singular values for each of the two cases, involving linearization about R-2 and the uniform distributions of k and ϕ, are plotted in figure (2.5.1). It is clear from this figure that the variation and the magnitudes of the singular values are characteristic of the general problem and do not change significantly with the actual distribution of k and ϕ as long as their mean values do not change appreciably. It is also evident that the singular values for this problem decline approximately in a geometric progression, each being smaller than the preceding value by a factor of approximately 0.63. Consequently a large number of singular values are much smaller than unity, pointing out the fact that the problem is in effect underdetermined as well as over-constrained. (Also, see section 3.3.)

Figure (2.5.2) contains plots of the vectors $\{\underset{\sim}{v}_i\}$ in the parameter space corresponding to the largest 10 singular values of $\underset{\sim}{A}$ for R-2. Each eigenvector $\underset{\sim}{v}_i$ has 66 elements, one element corresponding to each spatial grid point and each property (k and ϕ). In figure (2.5.2),

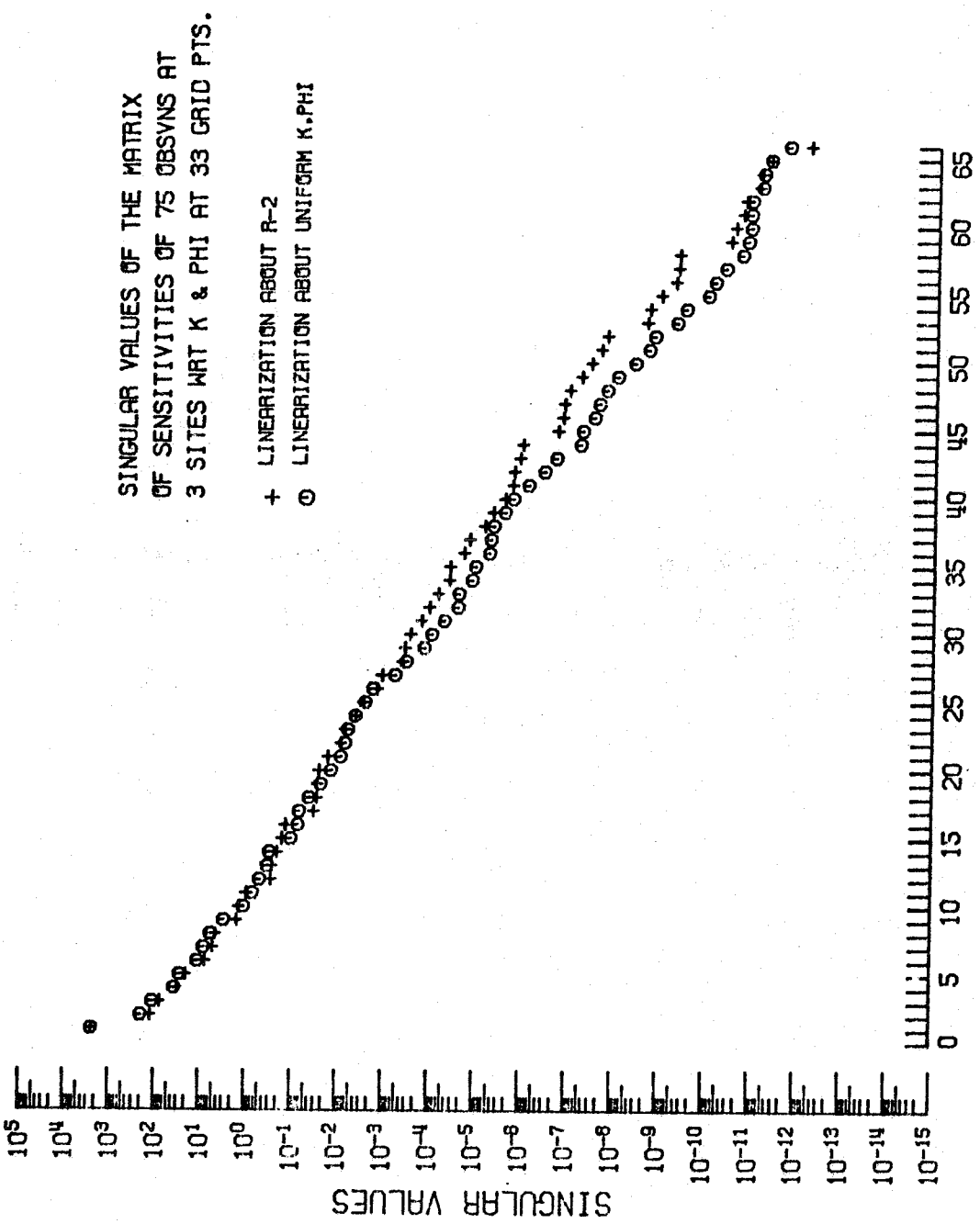


Figure 2.5.1 Singular Values of Sensitivity Matrix.

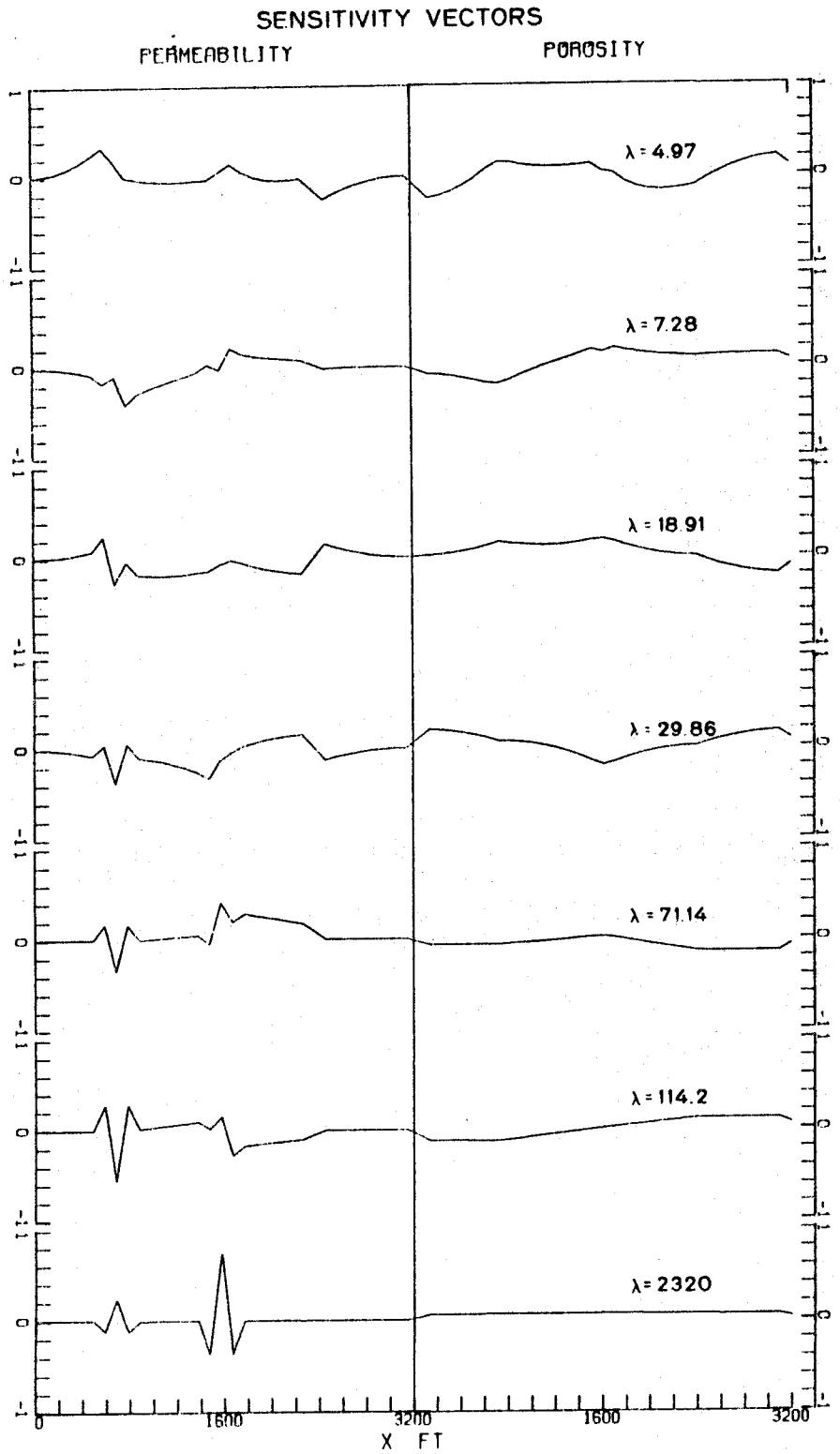


Figure 2.5.2 Eigenvectors of $\underset{\sim}{A}$ in Parameter Space Linearization about R-2.

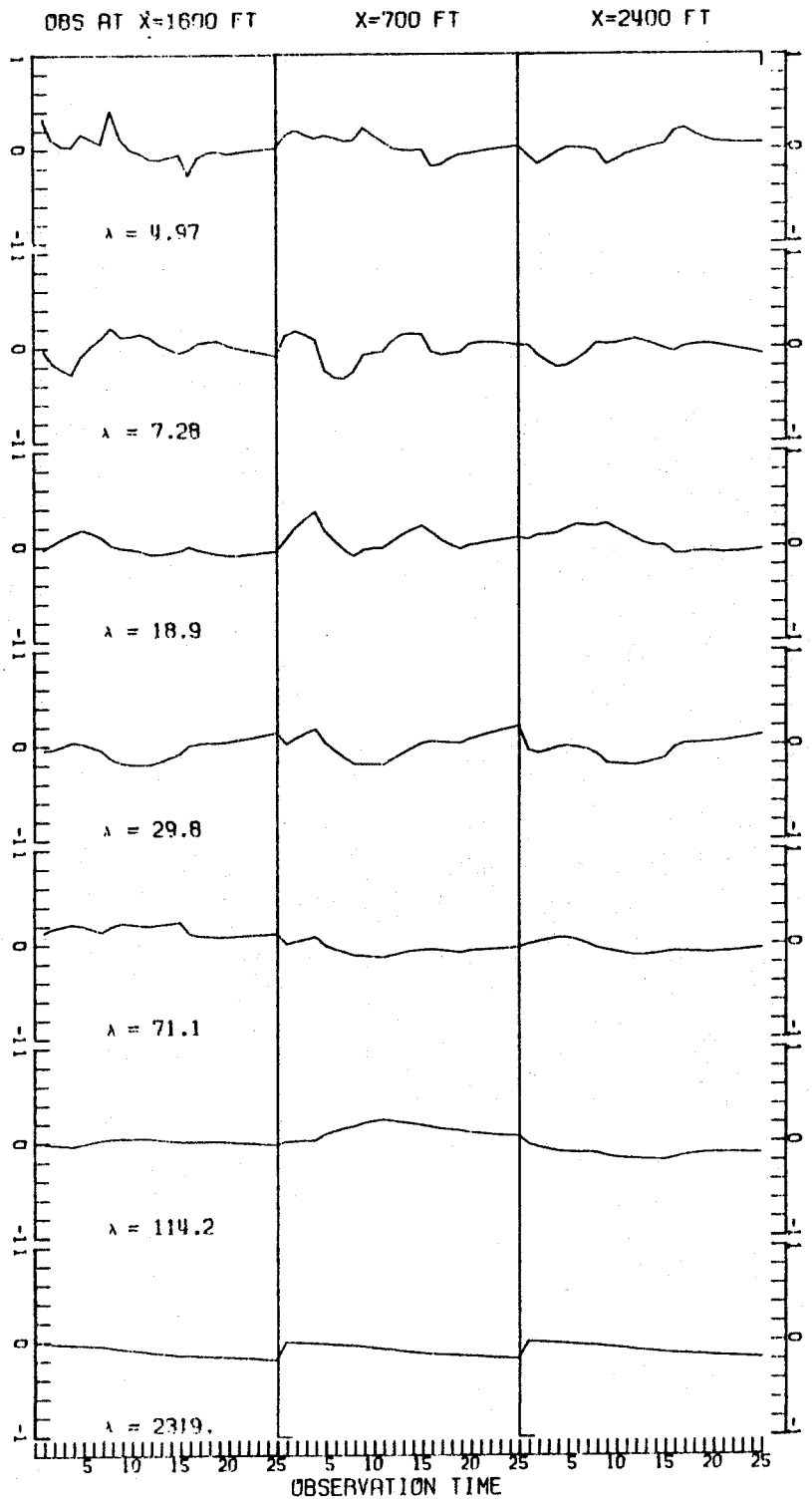
each element of $\underset{\sim}{v}_i$ is shown as a function of location corresponding to the permeability or porosity. It can be easily seen that these vectors become more and more oscillatory as the corresponding singular values decrease. We recall the conclusion of section 2.2 that a singular value λ_i represents the sensitivity of the component of $\underset{\sim}{y}$ along $\underset{\sim}{u}_i$ with respect to the component of $\underset{\sim}{\pi}$ along $\underset{\sim}{v}_i$. Thus this association of decreasing singular values with increasing oscillations in $\{\underset{\sim}{v}_i\}$ is in agreement with the analytical findings of section 2.4. In addition, it can be observed that the first several eigenvectors $\{\underset{\sim}{v}_i\}$ have significant oscillations in the permeability subspace but not in the porosity subspace. (We refer to the two subspaces spanned by the elements of $\underset{\sim}{v}_i$ corresponding to the permeabilities k_i and the porosities ϕ_i as the permeability subspace and the porosity subspace respectively.) Significant oscillations in ϕ-subspace occur in the eigenvectors with singular values much smaller than the largest. This observation is in qualitative agreement with the analytical result in section 2.4 that the components with a wave number ℓ ($\ell \neq 0$) of k(x) and $\varphi(x)$ have respectively $O(\ell^{-1})$ and $O(\ell^{-2})$ influence on the pressure.

Another noteworthy feature is that the vectors with large elements corresponding to the permeability in the vicinity of well locations have large singular values associated with them. This indicates that the permeability in the vicinity of well locations have large influence on the observations. This conclusion is easily verified from the numerical values of the elements of $\underset{\sim}{A}$. The highly oscillatory components of k with almost zero elements in the vicinity of the wells have a weak influence on the observations. Thus, the former

parameters will be determined accurately by history matching, whereas the latter components of $\underset{\sim}{k}$ will be poorly determined.

On the other hand, the eigenvectors $\{\underset{\sim}{v}_i\}$ with large elements corresponding to the permeability in regions away from wells, and especially in the vicinity of impermeable boundaries, have small λ_i, indicating that the permeability in these regions have very small influence on the observations. The porosity sensitivities exhibit a different behavior. The $\underset{\sim}{v}$-vector with the largest eigenvalue has almost equal elements corresponding to the porosities at all the grid points, indicating that the mean value of porosity has a very large influence on the observations. As the eigenvalues λ_i decrease, the associated $\underset{\sim}{v}_i$ show more and more oscillation in the porosity subspace. The effect of the distance of a grid point from any of the wells on the influence of its porosity on the observations is much less pronounced than in the case of the permeability.

Figure (2.5.3) shows the vectors $\{\underset{\sim}{u}_i\}$ in the observation space corresponding to the largest seven singular values for R-2. Since the observation vector has 75 elements, so has each $\underset{\sim}{u}_i$. Each observation corresponds to a specific observation location in the reservoir and a specific time; hence, the corresponding element of $\underset{\sim}{u}_i$ can also be identified by these. Thus, instead of the element number, these more physically meaningful identifications are used for the abscissae on the plots. It is evident that these follow a trend similar to that of $\{\underset{\sim}{v}_i\}$ in the parameter space: for large singular values, the vectors $\{\underset{\sim}{u}_i\}$ have few temporal oscillations over the period of the observed history; as the singular values become smaller, the corresponding



E-VECTRS IN OBSVN SPACE-SNST FOR R-2

Figure 2.5.3

$\underset{\sim}{u}$-vectors have more oscillations in time. This behavior implies that when the observations at a given location have systematic errors, they will have large influences on the parameter estimates. On the other hand, when these observation errors are uncorrelated from one observation instant to the next, they have very little influence on the parameters. Due to this, when the history matching is done, the mismatch of the latter kind will be difficult to eliminate by making corrections in the parameter estimates; consequently, the components with this characteristic will constitute the residual history-match error $\underset{\sim}{y}-\underset{\sim}{y}^{o}$. The rapidly fluctuating history match errors with almost zero mean when used in the forcing term in the adjoint equation, yield solution $\{\underset{\sim}{\psi}_i\}$ that itself fluctuates in space and time with small magnitudes. The use of such a solution for ψ in the expressions (A1.2.10-11) in appendix 1.2 for the gradients $\partial J/\partial \underset{\sim}{k}$ and $\partial J/\partial \underset{\sim}{\phi}$, yields a small norm of the gradient $\partial J/\partial \underset{\sim}{\pi}$ implying that the current estimates are in the vicinity of an extremum of J. In addition, the highly oscillatory behavior of ψ yields large numerical errors in computation with finite precision due to cancellations during the summation for time integration and consequent loss of significant digits. This partly explains why the performance of the gradient algorithm deteriorates as the minimum of J is approached. Similar reasoning can be followed to understand why the residual history match errors with almost zero mean and rapid oscillation fail to produce any further corrections in the estimates when the Gauss-Newton or Marquardt's method is employed.

Figure (2.5.4) shows the first 7 vectors $\{\underset{\sim}{v}_i\}$ obtained by linearization about the uniform distributions of k and ϕ. To facilitate a comparison between the respective $\underset{\sim}{v}$-vectors for linearization about the two different sets of distributions considered, they are plotted superposed on each other in figure (2.5.5). It is evident that the vectors $\{\underset{\sim}{v}_i\}$ with larger sensitivities do not depend very strongly on the actual distributions of k and ϕ used for linearization, as long as their mean values are approximately equal.

A more important question for the purpose of estimation is, how close to each other are the linear subspaces of the parameter space spanned by the highly sensitive vectors $\{\underset{\sim}{v}_i\}$ in the two cases? A quantitative procedure to answer this question is developed in appendix 2.1. The procedure consists of starting with a random vector $\underset{\sim}{\xi}$ in one subspace, with uncorrelated components along a set of orthonormal vectors spanning the subspace. Each of these components is assumed to have a Gaussian distribution with zero mean and a uniform variance σ^2. A projection of $\underset{\sim}{\xi}$ onto the second subspace and the associated residual vector, which are easy to determine, are also random. The ensemble average of the norm of the residual vector divided by the average norm of the random vector $\underset{\sim}{\xi}$ is taken as a scalar measure of the difference in the two subspaces. When one of the subspaces completely contains the other, this ratio will be zero; and for disjoint subspaces, it will be unity.

For the linearization about the two sets of profiles, the values of this measure were found to be 0.35, 0.30 and 0.226, for the dimensionalities of the subspaces of 16, 10 and 6 respectively. These values

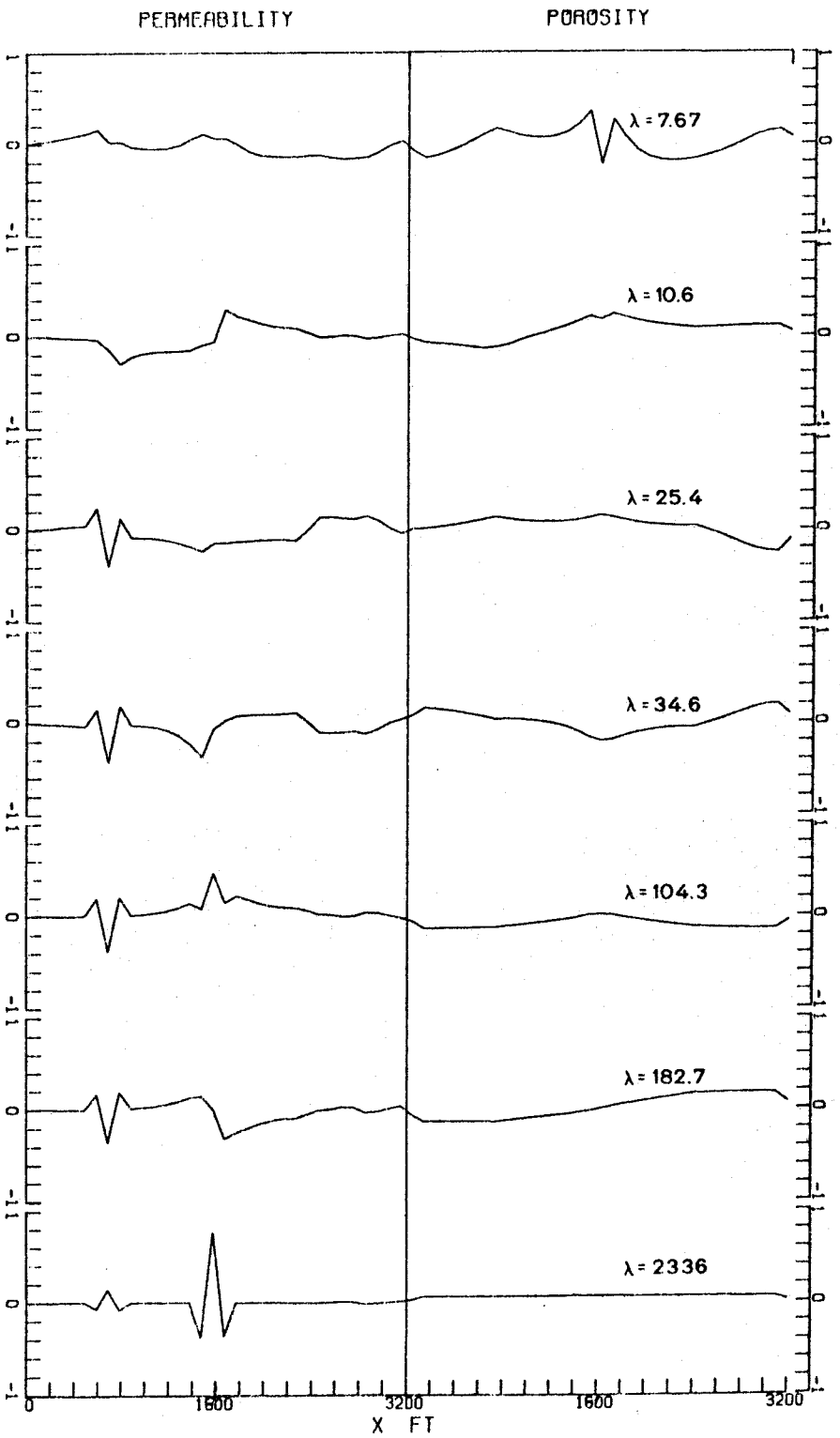


Figure 2.5.4 Eigenvectors of $\underset{\sim}{A}$ in Parameter Space Linearization about Uniform k, φ

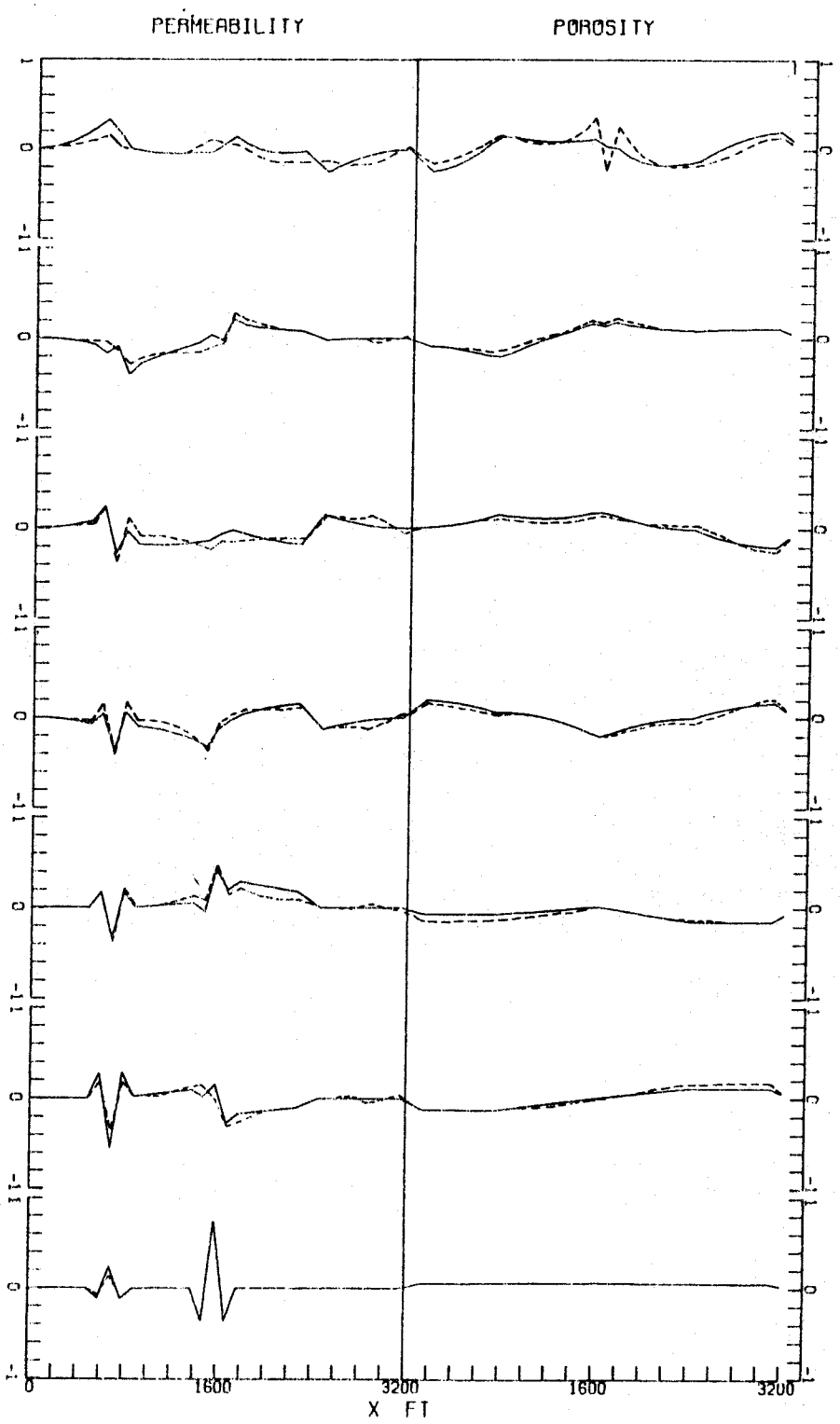


Figure 2.5.5 Comparison of Eigenvectors of $\underset{\sim}{A}$ in Parameter Space for Two Linearizations.

are fairly small compared with unity indicating that the intersection of the two subspaces encompasses a fairly large portion of either of them; a random vector in one of the subspaces has, on the average, a fairly large component in the other. Quantitatively, these values imply that the average length of the component of a completely random unit vector in one of the subspaces, lying outside the other subspace is 0.35, 0.30 and 0.226 for the three cases. We note that these values give only an approximate indication about the equivalence of the two subspaces. In practice, the corrections along the basis vectors in either of the subspaces will not, in general, have equal magnitudes, thus violating the assumption of uniform variance made in the foregoing analysis.

2.6 Estimation Using Only the Parameters with High Sensitivity

As we saw in section 2.5, there are only a few directions locally in the parameter space that significantly influence the observed pressures. We also observed that these directions do not change appreciably as the distributions of k and ϕ used for the linearization are changed, so far as their mean values do not change significantly.

Since only the components of the corrections along these highly sensitive directions are determined accurately by the data, we may estimate them without attempting any corrections in the insensitive components. Later, corrections along the less sensitive components may be made based on some auxilliary information. The Bayesian penalty, introduced in chapter 1, is one example of such additional information. The history match would not be significantly altered by such a modification of the estimates, as the

additional corrections are along the insensitive directions in the parameter space.

The results of such estimation using the 16 most sensitive vectors in the parameter space for the basis of the corrections in the estimates are summarized in table (2.6.1). The estimation is carried out using two different sets of vectors $\{\underset{\sim}{v}_i\}$ obtained from linearizations about R-2 and the uniform distributions. The estimates resulting from the former set are plotted in figure (2.6.1). The details in table (2.6.1) indicate that the results obtained by the two different sets are very close to each other. The history match errors have been reduced to the irreducible level in both cases. The estimate errors J_k and J_ϕ are of the same order as those obtained by Zonation or by the Bayesian approach. However, as can be seen from figure (2.6.1), the estimated distributions have sharp kinks in the vicinity of the production and observation sites. This stems from the fact that the highly sensitive vectors $\{\underset{\sim}{v}_i\}$, utilized as the basis, themselves have such kinks. (See figure 2.5.5.)

We may attempt to smooth out the kinks in the estimates $\hat{\underset{\sim}{\pi}}$ by making appropriate corrections along the insensitive vectors. This smoothing can be done so as to minimize the Bayesian penalty term. Such a smoothing problem is a linear quadratic minimization problem and its solution can be obtained in closed form. The details of the problem and the resulting solution are described in appendix 2.2. However, as is evident from it, the problem and its solution are very ill-conditioned, involving the inverse of a nearly singular matrix $\underset{\sim}{P}_0$. As an alternative, an approximate formulation of the

TABLE 2.6.1

ESTIMATION USING ONLY THE SENSITIVE COMPONENTS

Dimension of the correction subspace = 16.

Conditions of the problem: Set S_2 (See section 1.4)

Initial J_p = 14,900

Initial J_k = 11.73 , Initial J_ϕ = 7.22.

	Final Values		
	J_p	J_k	J_ϕ
(1) Sensitivity vectors obtained by linearization about R-2. Marquardt's Method (μ = 100)	73.2	8.91	6.68
(2) Linearization about uniform distributions of k and ϕ : k(x) = 5.0 md, ϕ(x) = 0.2 Marquardt's Method (μ = 5)	70.6	7.64	6.87

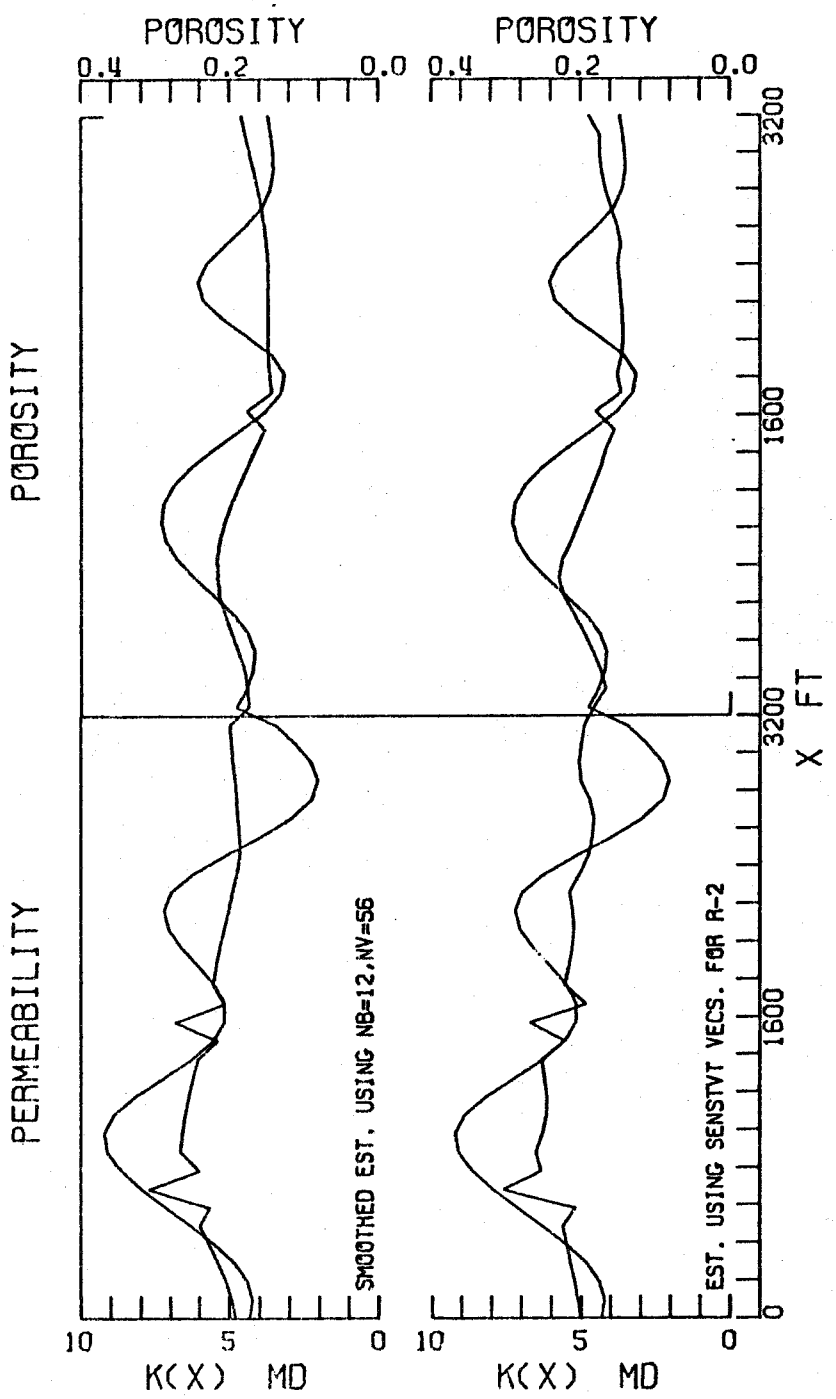


Figure 2.6.1 Estimation Using Only High Sensitivity Components and A Posteriori Smoothing of Estimates.

smoothing problem which is well-conditioned is also described in appendix 2.2. Smoothing was carried out using this approximate scheme, and the results are plotted in figure (2.6.1). The reported results are obtained by removing to the best possible extent, the components of $(\hat{\underset{\sim}{\pi}}-\bar{\underset{\sim}{\pi}})$ along all except the first 12 eigenvectors of the prior covariance matrix $\underset{\sim}{P}_o$, using for correction the 56 least sensitive $\underset{\sim}{y}$-vectors in the parameter space, obtained by linearization about R-2. As can be seen, the kinks in the vicinity of the production well are not significantly smoothed. This is due to the fact that the set of insensitive vectors used for smoothing does not span the parameter space, and therefore cannot annihilate the arbitrary difference between a smooth distribution and the estimate $\hat{\underset{\sim}{\pi}}$.

We may attempt iterative smoothing by successively linearizing about the smoothed estimates. However, due to the approximate invariance of the sensitive subspace described in section 2.5, such iterative attempts may not be successful in completely eliminating the kinks in $\hat{\underset{\sim}{\pi}}$. In addition, if the smoothing is successful after several iterative attempts, the total corrections may be large. Though they are in locally insensitive directions, these large corrections may induce nonlinear effects, not accounted for in the linearized analysis, and significantly alter the history match. If this occurs, further effort would be needed to restore the history match. Thus, this approach appears to require very large computational effort without guaranteeing acceptable results.

As a result, we conclude that the Bayesian approach to estimation is preferable to the one discussed in this section. The

Bayesian approach uses smooth vectors as a basis for the correction subspace, and consequently yields smooth estimates, requiring no a posteriori treatment.

2.7 An Index of Regional Parameter Sensitivity

In this section we develop a scalar measure of the sensitivity of a given rock property in a given region of the reservoir that takes into account all the observations.

The magnitude of the derivative $\partial y_i / \partial \pi_j$ is a measure of the sensitivity of the observation y_i to small variations in the parameter π_j. However, any single derivative does not contain all the information about the sensitivities of all the observations. The matrix $\underset{\sim}{A}$ of all such derivatives contains such information; however, it is in a form that is difficult to delineate. The set of singular values $\{\lambda_i\}$ also contains this information, but it is difficult to assign any regional significance to the associated eigenvectors in parameter space. Hence it is necessary to carry out further analysis to answer the questions about the regional parameter sensitivities.

The singular values $\{\lambda_i\}$ of $\underset{\sim}{A}$ depend on two independent factors: (i) the magnitudes of the element of $\underset{\sim}{A}$, and (ii) the linear independence of the vectors formed by the sensitivities with respect to all the parameters of the different observations (which are the rows of $\underset{\sim}{A}$). The first factor is directly proportional to the observed pressure levels. Since the eigenvectors $\{\underset{\sim}{u}_i\}$ and $\{\underset{\sim}{v}_i\}$ are all normalized to have Euclidean norms of unity, all the information about the magnitude of the elements of $\underset{\sim}{A}$ is contained in the singular values. If all elements of $\underset{\sim}{A}$ were changed by a factor c, the singular

values will also be changed by the same factor, leaving the eigenvectors unaltered. The factor (ii) can be understood from the following. If a row (or a column) of $\underset{\sim}{A}$, that is originally linearly independent of the rest, is replaced by a vector which is linearly dependent on the rest, the rank of $\underset{\sim}{A}$ is reduced by 1. The number of non-zero singular values of $\underset{\sim}{A}$ is equal to its rank. Thus, the singular values are affected by the linear independence of the different rows of $\underset{\sim}{A}$. The second factor depends on the number of independent modes effectively contributing to constitute the observed pressures. The first factor mainly affects the larger of the singular values; whereas the other determines the rank of $\underset{\sim}{A}$ and thus has a large effect on the smaller of the singular values. The second factor is of great importance in the solution of the estimation problem as the smaller singular values have the greater influence on the covariance of the estimate, as is briefly discussed in section 2.2 and developed in detail in chapter 3. The first is a dominant factor in the parabolic problems, the kind to which the reservoir system equation belongs, because the higher modes in the solution decay very quickly with time, thus making it very difficult to alter the second factor significantly. An important exception occurs when an additional observation site is established. The observations at the new site contain fresh information about the eigenfunctions, thus contributing directly to the increment in the effective rank of $\underset{\sim}{A}$. In addition, the rock properties in the vicinity of the new site have a pronounced influence on the observations at that location. Thus a

new observation site contributes the first factor as well. These influences are even more significant when the new site also has a production well.

We shall assume that the eigenvectors of $\underset{\sim}{A}$ in the parameter space can be divided into two groups: $\{\underset{\sim}{v}_1, \underset{\sim}{v}_2, \ldots, \underset{\sim}{v}_\ell\}$ corresponding to relatively large singular values, and the set of remaining $(M-\ell)$ vectors with small singular values. Then the vectors representing the regional rock properties that are effectively contained in the subspace spanned by the first set of vectors will have a high measure of sensitivity. Thus, we may assign the index of sensitivity to the regional rock properties proportional to the degree to which their vector representations are contained in the subspace of the more sensitive $\underset{\sim}{v}$-vectors. This degree may be quantitatively determined by the following procedure.

The i^{th} element of the parameter vector $\underset{\sim}{\pi}^T = [\underset{\sim}{k}^T | \underset{\sim}{\phi}^T]$ represents a rock property in a region of one grid block size in the reservoir. The norm of the component of the M-vector ($M = 2N$) $\underset{\sim}{e}_i$, which is the i^{th} column of the identity matrix, along any unit vector is a measure of the degree to which the vectors coincide. Similarly, the norm of a vector, which is the sum of the vector components of $\underset{\sim}{e}_i$ along a set of unit vectors in E^M is a measure of the degree to which $\underset{\sim}{e}_i$ is contained in the subspace spanned by this set of vectors. Then it follows that the norm of the component of $\underset{\sim}{e}_i$, orthogonal to the above vector-sum is a measure of the degree to which $\underset{\sim}{e}_i$ fails to be contained in the subspace. If we use the set of vectors $\underset{\sim}{v}_i$ associated with large eigenvalues in the above analysis, we have a

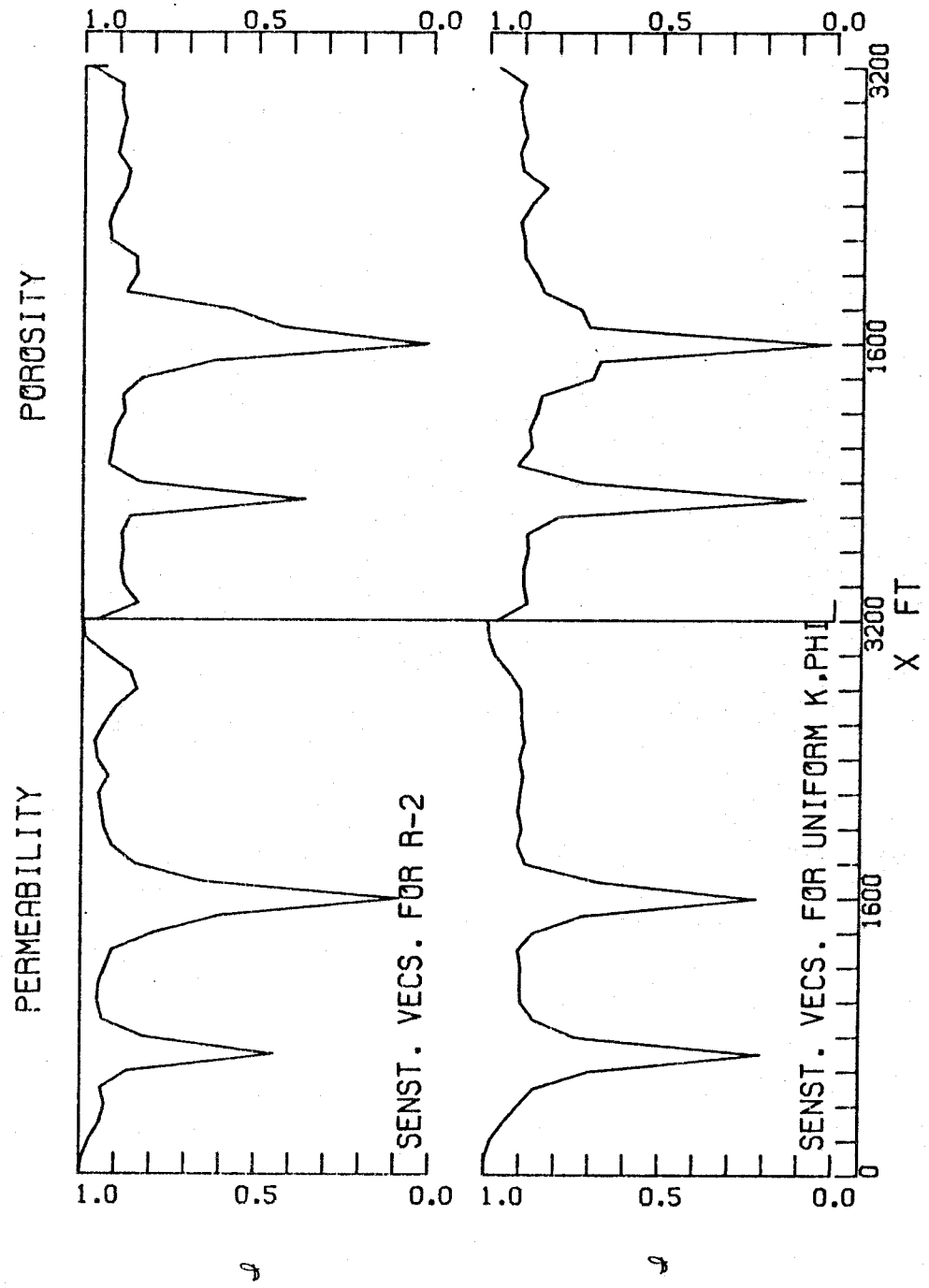
measure of the degree to which the i^{th} component of the parameter vector fails to be contained in the high sensitivity subspace of the parameter space. This component is given by,

$$\underset{\sim}{e}_i - \underset{\sim}{V}_1 \underset{\sim}{V}_1^T \underset{\sim}{e}_i = (\underset{\sim}{I} - \underset{\sim}{V}_1 \underset{\sim}{V}_1^T) \underset{\sim}{e}_i \qquad (2.7.1)$$

where $\underset{\sim}{V}_1$ is the matrix whose columns constitute the set of high sensitivity vectors. Thus we define the index of insensitivity of i^{th} component of the parameter vector as,

$$\mathcal{I}_i = \| (\underset{\sim}{I} - \underset{\sim}{V}_1 \underset{\sim}{V}_1^T) \underset{\sim}{e}_i \| \qquad (2.7.2)$$

where $\| \cdot \|$ is the Euclidean norm. Using this definition of the index of region-wise parameter insensitivity, we can compute this index for all the grid points in the reservoir, considering each rock property separately. Figure (2.7.1) shows this index plotted for all the 66 elements of $\underset{\sim}{\pi}$; two sets of indices are computed, using two different sets of seventeen $\underset{\sim}{v}$-vectors ($\ell = 17$) obtained for the two sets of k and ϕ distributions. Again, the correspondence of a given element of the parameter vector with a region in the reservoir is used to label the abscissa in terms of location within the reservoir, for permeability and porosity. As all the vectors involved in this computation are of unit norm, the residual vector in (2.7.1) will have a maximal norm of unity. A value of the index close to zero corresponds to a high degree of sensitivity and a value close to unity indicates almost no influence on the observations of the appropriate rock property in the corresponding region. The plots in figure (2.7.1) indicate that the permeability and porosity in regions



INDEX OF REGION-WISE INSENSITIVITY

Figure 2.7.1

close to observation wells which also produce (x = 700 ft and x = 1600 ft) have a large influence on the observations. The properties in a region around an observation well with no production (x = 2400 ft) have much lower sensitivity. Properties in regions away from any of the wells have have still smaller sensitivities. The least influential are the sensitivities in regions close to the impermeable boundaries, the regions for the permeability being larger than the porosity. It is noteworthy that the distributions of the insensitivity index for the two sets of k and ϕ distributions do not differ significantly. As a result, we may carry out such analysis, for a given reservoir, in advance of any extensive parameter estimation attempts and consequently be forewarned about the ill-determinancy of the rock properties in specific regions; thus, we may be wary of placing unjustified confidence in the results of the subsequent estimation.

In principle, the distribution of the residuals computed as above, can be used in designing improved parameterization for any specific reservoir parameter estimation problem. For example, we may select the zone boundaries so as to yield zones with approximately equal sensitivities. However, due to the rapid changes in the distribution of the index, indicating high sensitivity only in very small regions and uniformly low sensitivity in the rest of the reservoir, this index is not suitable for a fine grading of the parameters based on the sensitivities. A strict adherence to it will result in a few and grossly unequal zones, which would be unacceptable. A similar sensitivity analysis can also be carried out for these parameters

in Bayesian estimation. Thus, alternative parameterization can be compared. (See section 2.8 for further discussion.)

2.8 Practical Uses of the Linear Analysis

The analysis of the linear subproblem presented in this chapter is a very powerful aid in understanding the nature of the parameter estimation problems. By performing this analysis using the prior estimates of the unknown parameters before attempting the estimation, a practicing reservoir engineer can uncover a lot of useful information about a given problem. This information can be subsequently used in carrying out the estimation in an optimal fashion.

The first step is to compute accurately the matrix $\underset{\sim}{A}$ for a suitably fine space grid and the available observations using the prior estimates for linearization. The singular value decomposition of this matrix may be efficiently and accurately accomplished using double precision word-length and the program by Bushinger and Golub (1969). The singular values, which are all positive, when arranged in a decreasing order of magnitude yield a very clear picture about the conditioning of the problem and the number of independent parameters one can expect to estimate in the given problem. An error of one unit in the observations may yield an error of $1/\lambda_i$ units in the component along $\underset{\sim}{v}_i$ of the parameter vector. From this consideration, a positive cutoff in the singular values, such that the singular values smaller than it should not be utilized for estimation, can be selected as follows. Let the standard

deviation of the observation errors be σ_o. Let ϵ be the estimate of the maximum modelling error, based on the grid size and numerical scheme used for simulation. Any uncertainties other than the unknown parameters, about the conditions existing in the reservoir must also be reflected in ϵ. Then the total expected difference between an observed pressure and the corresponding computed pressure after the estimation is completed, is of the order,

$$\sigma = \sigma_o + \epsilon \tag{2.8.1}$$

Let the maximal permissible uncertainty in the final estimate of the parameters be given by the standard deviation $\sigma_{\underset{\sim}{\pi}}$. Then the cutoff c in the singular values may be selected as,

$$c = \sigma/\sigma_{\underset{\sim}{\pi}} \tag{2.8.2}$$

The singular values $\{\lambda_i : \lambda_i < c\}$ should not be used for making corrections in the estimates during the iterative process. This criterion should govern the selection of the history matching scheme. The gradient algorithm, as discussed in section 2.3, will almost always be in compliance of this criterion; but it is usually very restrictive in making corrections along many $\underset{\sim}{v}$-vectors and as a result, the effective cutoff may be higher than in (2.8.2). Marquardt's method on the other hand, offers more control on the corrections through the choice of the parameter μ. The largest correction in Marquardt's method occurs along $\underset{\sim}{v}_i$ closest to $\sqrt{\mu}$. Thus, we may choose $\mu = a c^2$, where $a > 1$ is a positive constant which ensures that only very small corrections are made along the eigenvectors with

$\lambda_i \leq c$. The choice of a depends on the actual parameterization being used; but a = 10 often seems to be satisfactory. The number of singular values greater than the cutoff is approximately the number of parameters that can be estimated with reasonable certainty.

Each of the $\underset{\sim}{v}$-vectors associated with these large singular values should be plotted as a function of the spatial position in the reservoir for each of the unknown rock properties. Then the regions where one or more of these vectors have large elements are where we may expect to determine the appropriate rock property accurately. In the rest of the reservoir domain, the rock property is not well-determined by the data and its final estimate may be expected to be fairly uncertain. If possible, the rest of the $\underset{\sim}{v}$-vectors should also be plotted. The components of the rock properties along these vectors have little influence on the data. Particularly, note that many of the ill-determined components have very rapid spatial oscillations. Similarly, the $\underset{\sim}{u}$-vectors may be plotted against time for each of the observation sites. Then the history match errors following the pattern of any of the $\underset{\sim}{u}$-vectors with large singular values will have a large influence on the parameters and will be more completely eliminated than the others during the history matching attempt. On the other hand, the errors along those with small singular values will be difficult to remove and will constitute the residual history match errors.

The results of this analysis can be further extended to quantify the region-wise parameter sensitivity using the procedure described in section 2.7. One may compute the index of the region-

wise sensitivity for each of the rock properties. The index may be used for a rough indication of the uncertainty to be expected in the corresponding final estimates. It may also be used to determine the optimal zonations as follows. The regions with high sensitivity index may be covered by small zones to obtain a better resolution in the final estimates. The regions with low index of sensitivity may be lumped into a few large zones in order to increase the sensitivity of the zonation parameters to a level such that they can be estimated with acceptable certainty. However, while using this technique it should be borne in mind that this index exhibits a high degree of saturation characteristics. That is, it varies rather abruptly from a high value to a very low value and thus the gradation for the intermediate values of the sensitivity is not very satisfactory. Hence, judgement must be exercised while defining the zones in this manner.

The matrix of the sensitivity of the observations with respect to the zonation parameters can be computed trivially from the matrix $\underset{\sim}{A}$. This is achieved by simply adding the sensitivity values of a rock property at the nodes of the grid contained in the zones. This sensitivity matrix can be similarly analyzed by carrying out its singular value decomposition. Then the singular values will give an indication of how well-conditioned is the zonation parameter estimation problem. If all the singular values are larger than the cutoff, then all the zonation parameters can be determined with the acceptable level of certainty. Thus this analysis can be used to study the effect of the various alternative zonations and consequently to determine the

"optimal" zonation.

The Bayesian approach to estimation, described in chapter 1, has many characteristics that make it preferable to zonation. We may obtain a sensitivity matrix for the Bayesian parameters from $\underset{\sim}{A}$ using the chain rule of differentiation. This merely involves, as in appendix 1.4, the postmultiplication of $\underset{\sim}{A}$ by the transpose of the matrix $\underset{\sim}{Z}$, the columns of which are the eigenvectors, used as the Bayesian parameters, of the prior covariance matrix. Then the singular value decomposition of the resulting matrix may be similarly carried out. The eigenvalues of the prior covariance matrix give the expected degree of participation by the associated eigenvectors in constituting the corrections to the prior estimates. The singular values of the sensitivity matrix, on the other hand, indicate the degree of information available in the observation about the associated $\underset{\sim}{y}$-vectors. Thus, if the number of the basis vectors with high participation factors is less than or equal to the number of singular values larger than the cutoff, the Bayesian estimation is well-conditioned, and the significant components of the corrections are well-determined.

The conditioning of the estimation problem is determined by the spread of the larger singular values. If the singular values larger than the cutoff are spread over many orders of magnitudes, the estimation problem is ill-conditioned. Thus a parameterization which yields the smaller spread of these singular values is the best-conditioned formulation of the problem. The linear analysis can thus be used to compare the conditioning yielded by the alternative

zonation and Bayesian parameterizations; and their relative merits can be judged, leading to an "optimal" parameterization for the given problem.

2.9 Conclusions

(1) The reservoir parameter estimation problem is usually grossly underdetermined, when the values of the rock properties at each grid block is allowed to be independent of others. This leads to a high degree of non-uniqueness of the estimates

(2) Through linearized analysis, it is possible to determine which components of the parameters cannot be estimated from a given set of data. Analytical expressions are derived for the sensitivity coefficients of an observation pressure with respect to small arbitrary perturbations in the porosity and permeability in a single phase one-dimensional reservoir; two different kinds of boundary conditions, for impermeable and constant pressure boundaries, are considered. In reservoirs with either kind of boundary conditions, the Fourier component of the permeability with wave number ℓ ($\ell \neq 0$) has $O(\ell^{-1})$ influence at large times on the pressure; this implies that the highly oscillatory components of the permeability cannot be estimated from the observed pressures. In the reservoir with impermeable boundaries, a similar component ($\ell \neq 0$) of porosity has $O(\ell^{-2})$ influence, whereas for $\ell = 0$ the influence increases linearly with time; thus the mean porosity can be very accurately estimated but the oscillatory components are worse determined by the data than similar components of the permeability. In the constant pressure boundary reservoirs, the influence of the various porosity components vanishes exponentially

with t ; thus accurate determination of porosity in such reservoirs based on observed pressures alone is not possible.

(3) The estimation may be attempted in two steps, estimation of components of parameters in sensitive subspace, followed by the determination of the insensitive components to smooth out the estimated distributions according to some auxilliary a priori information about the parameters. Such an attempt is found to require many iterative steps including linearization and smoothing; it may not be successful even after many iterations due to the approximately invariant nature of the sensitive subspace.

(4) A scalar index of the degree of accuracy to which a given rock property in any region of the reservoir can be estimated from a given pressure data, is developed. This index can be calculated a priori, using nominal estimates of the properties. The index can be used to determine the confidence that can be placed on the results of the subsequent attempts at estimation, and to study the conditioning of the problem under various alternative parameterizations.

CHAPTER 3

COVARIANCE OF THE ESTIMATES

In this chapter we present methods for approximate determination of the covariances for the parameter estimates obtained by the zonation and the Bayesian approaches.

There are two essential ingredients required for the determination of the covariance of an estimate: (a) a measure of information gained from the observations, and (b) a measure of errors in the prior estimates or the "initial guess." The first factor is determined through the linearization of the estimation problem about the available estimates, drawing on the results obtained in Chapter 2. The second factor is used for the estimation of the error in the components of the parameters about which the observations contain inadequate information. It is assumed that the prior information is available in the form of a priori mean and covariance matrix.

In section 3.1, we advance the arguments and derive the procedures for calculating the covariances for three cases: (i) estimation done without constraints, i.e., no attempts are made to group and reduce the number of parameters, (ii) Bayesian estimation, including its parameterization and the "soft" penalty constraint, (iii) estimation with "hard" constraints, an important example being the zonation constraint. In section 3.2, we explore numerically the procedures developed in section 3.1, and illustrate their application for the different approaches to estimation using the one-dimensional single-phase reservoir. A discussion of various aspects of the results and a comparison of the covariances for the different cases are included.

Section 3.3 contains the verification of the predicted covariances for the different parameterizations, using the results of simulated estimation problems with k and ϕ distributions for a sample of four realizations of the random process model. Section 3.4 contains a numerical study of the sensitivity of the computed covariances to changes in the prior covariance matrix. The prior covariance matrix is based on the random process model of the parameter distributions, introduced in the Bayesian estimation. The sensitivity of the covariances in the different cases is investigated in relation to the different parameters determining the random process. Finally, the effect of erroneous prior information on the predicted a posteriori covariance is investigated in section 3.5.

3.1 Derivation of Expressions for the Covariances

In this section we shall discuss the problem of determination of estimate covariance in general terms. Having pointed out the difficulties with the standard approach when applied to the reservoir property estimation, we shall propose a procedure that alleviates these difficulties. Later in this section, we apply the procedure to the zonation and Bayesian approaches to estimation, which are respectively examples of hard and soft constraints on the parameter distributions.

Let $\underset{\sim}{y}$ be the calculated values of the observed pressures corresponding to the estimates $\hat{\underset{\sim}{\pi}}$ at any iteration of the history matching procedure. Let $\underset{\sim}{y}^o$ be the M-vector of observations, and $\underset{\sim}{\pi}^*$ be the M-vector of the unknown "true" values of the parameters. Let $\underset{\sim}{y}^*$ be

the "true" values of the observed pressures, corresponding to the parameter vector $\underset{\sim}{\pi}^*$. Let the observations and $\underset{\sim}{y}^*$ be related by

$$\underset{\sim}{y} = \underset{\sim}{y}^* + \underset{\sim}{\eta} \tag{3.1.1}$$

where $\underset{\sim}{\eta}$ is a Gaussian random vector of observation errors, with zero mean and covariance $\underset{\sim}{\Sigma}$. Let us define the difference of the calculated pressures and the "true" pressures corresponding to the observations as

$$\delta\underset{\sim}{y} = \underset{\sim}{y}^* - \underset{\sim}{y} = \underset{\sim}{y}^o - \underset{\sim}{y} - \underset{\sim}{\eta} \equiv \delta\underset{\sim}{y}_h - \underset{\sim}{\eta} \tag{3.1.2}$$

where $\delta\underset{\sim}{y}_h$ is the history match error. Similarly, we define the difference between the estimate and the unknown "true" value $\underset{\sim}{\pi}^*$ of the parameters as

$$\delta\underset{\sim}{\pi} = \underset{\sim}{\pi}^* - \hat{\underset{\sim}{\pi}} . \tag{3.1.3}$$

We shall assume that the history matching is carried out to such a degree that the residual history match error $\delta\underset{\sim}{y}_{hr}$ is small; let $\hat{\underset{\sim}{\pi}}_h$ be the resulting estimate of $\underset{\sim}{\pi}$. Further, we assume that the observation errors in $\underset{\sim}{\eta}$ are small; then the residual value $\delta\underset{\sim}{y}_r$ of the deviation $\delta\underset{\sim}{y}$ defined in (3.1.2) is also small. We shall assume that the corresponding residual deviation $\delta\underset{\sim}{\pi}_r = \underset{\sim}{\pi}^* - \hat{\underset{\sim}{\pi}}_h$ is also small. Then we have the linearized relation [see (2.1.4)]

$$\delta\underset{\sim}{y}_r = \underset{\sim}{A}\,\delta\underset{\sim}{\pi}_r \tag{3.1.4a}$$

and with (3.1.2) we obtain, for the residual values,

$$\delta\underset{\sim}{y}_{hr} = \underset{\sim}{A}\,\delta\underset{\sim}{\pi}_r + \underset{\sim}{\eta} \quad , \tag{3.1.4b}$$

where $\underset{\sim}{A}$ is the (L × M) sensitivity matrix evaluated at the estimates $\hat{\underset{\sim}{\pi}}_h$,

$$\underset{\sim}{A} = \left(\frac{\partial \underset{\sim}{y}}{\partial \underset{\sim}{\pi}}\right) \hat{\underset{\sim}{\pi}}_h \tag{3.1.5}$$

Then, for any inverse $\underset{\sim}{A}^{\psi}$ of $\underset{\sim}{A}$, the solution of (3.1.4a) is

$$\delta \underset{\sim}{\pi} = \underset{\sim}{A}^{\psi} \delta \underset{\sim}{y}_r \tag{3.1.6}$$

However, the estimate correction $\delta \hat{\underset{\sim}{\pi}}_r$ is obtained by using the residual history match error $\delta \underset{\sim}{y}_{hr}$ in (3.1.6). Then,

$$\delta \hat{\underset{\sim}{\pi}}_r = \underset{\sim}{A}^{\psi} \delta \underset{\sim}{y}_{hr} \tag{3.1.7}$$

We shall define the error in the final estimate as

$$\underset{\sim}{e} = \hat{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^* \tag{3.1.8}$$

where we redefine $\hat{\underset{\sim}{\pi}}$ to be the final estimate of $\underset{\sim}{\pi}$ after the correction $\delta \hat{\underset{\sim}{\pi}}_r$ is made in $\hat{\underset{\sim}{\pi}}_h$:

$$\hat{\underset{\sim}{\pi}} = \hat{\underset{\sim}{\pi}}_h + \delta \hat{\underset{\sim}{\pi}}_r \tag{3.1.9}$$

Then we have

$$\delta \underset{\sim}{\pi}_r - \delta \hat{\underset{\sim}{\pi}}_r = \underset{\sim}{\pi}^* - \hat{\underset{\sim}{\pi}}_h - \hat{\underset{\sim}{\pi}} + \hat{\underset{\sim}{\pi}}_h = \underset{\sim}{\pi}^* - \hat{\underset{\sim}{\pi}} = -\underset{\sim}{e} \tag{3.1.10}$$

Furthermore, we obtain from (3.1.6) and (3.1.7)

$$\delta \underset{\sim}{\pi}_r - \delta \hat{\underset{\sim}{\pi}}_r = \underset{\sim}{A}^{\psi} (\delta \underset{\sim}{y}_r - \delta \underset{\sim}{y}_{hr}) = -\underset{\sim}{A}^{\psi} \underset{\sim}{\eta} \tag{3.1.11}$$

where the relation (3.1.2) is applied to the residuals. Combining (3.1.10) and (3.1.11), and using the statistics of $\underset{\sim}{\eta}$ we obtain,

$$E_\eta \{\underset{\sim}{e}\} = \underset{\sim}{0}$$

$$E_\eta \{\underset{\sim}{e}\underset{\sim}{e}^T\} = \underset{\sim}{A}^\psi E_\eta \{\underset{\sim}{\eta}\underset{\sim}{\eta}^T\} \underset{\sim}{A}^{\psi T} = \underset{\sim}{A}^\psi \underset{\sim}{\Sigma} \underset{\sim}{A}^{\psi T} \tag{3.1.12}$$

where E_η denotes expectation with respect to $\underset{\sim}{\eta}$. Using (3.1.10) we define the covariance of the error in the final estimate as

$$\underset{\sim}{P}_\pi = E_\eta \{\underset{\sim}{e}\underset{\sim}{e}^T\} = E_\eta \{(\hat{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)(\hat{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)^T\} \tag{3.1.13}$$

Then we have

$$\underset{\sim}{P}_\pi = \underset{\sim}{A}^\psi \underset{\sim}{\Sigma} \underset{\sim}{A}^{\psi T} \tag{3.1.14}$$

Thus the covariance of the error in estimate depends on the inverse $\underset{\sim}{A}^\psi$ of $\underset{\sim}{A}$. As discussed in section 2.5, the singular values of $\underset{\sim}{A}$ for an unconstrained reservoir property estimation problem decrease fairly rapidly in a geometric progression, resulting in a situation where only a few of these singular values are significantly larger than unity. This holds for most of the parameterizations normally used, and at least a few of the singular values of $\underset{\sim}{A}$ are usually much smaller than unity. Thus, an inverse of $\underset{\sim}{A}$, which takes into account all of its singular values would usually have very large components yielding a covariance with unacceptably large elements. For example in zonation, when the ordinary least square solution, which is equivalent to the Lanczos inverse (Chapter 2) including all the singular values of $\underset{\sim}{A}$, is used in (3.1.14), the trace of $\underset{\sim}{P}_\pi$ for the simulated problem with conditions of set S_2 (see section 1.4) and $\underset{\sim}{\Sigma} = \underset{\sim}{1}$ is as shown in table (3.1.1). (Parameter means = O(1)).

Table 3.1.1 Effect of Number of Zones on Covariance

No. of Zones	Trace ($\underset{\sim}{P}_{\pi}$)
4	0.843
8	0.427×10^{5}
16	0.504×10^{10}
33	0.406×10^{22}

Thus it is evident that inclusion of all the singular values of $\underset{\sim}{A}$ in $\underset{\sim}{A}^{\psi}$ leads to unacceptably large variances in all of the above zonations except for the four-zone case. However, when the number of zones is small as in the last case the large zones entail a large error due to the parameterization and it must be taken into account also. We shall have more to say about this and other errors later in this section.

On the other hand, ignoring the smaller singular values of $\underset{\sim}{A}$ in constructing $\underset{\sim}{A}^{\psi}$ would imply that $\underset{\sim}{A}$ is in effect rank deficient - a situation in which the problem is underdetermined; then the inverse $\underset{\sim}{A}^{\psi}$ is not unique. Thus, this direct approach to determination of the covariance of the estimates does not yield reliable, acceptable and unique results.

To obtain a realistic and acceptable estimate of $\underset{\sim}{P}_{\pi}$, additional information about $\underset{\sim}{\pi}$ must be incorporated into the covariance determination. Such information may be in the form of a prior probability density of $\underset{\sim}{\pi}$, deriving from core sample data and other field tests, perhaps in conjunction with a random process model for the rock property distribution similar to that used in the Bayesian estimation

discussed in Chapter 1. If more than one kind of prior information is available about $\underset{\sim}{\pi}$, it should be integrated with appropriate weighting during or before its use in the covariance determination procedure. In the following we propose a methodology for accomplishing this. We shall treat individually three types of parameterizations: (i) Estimation with no explicit constraints, (ii) Estimation with soft constraints using Bayesian estimation as an example, (iii) Estimation with hard constraints with zonation as an example.

3.1.1 Covariance for Estimation without Explicit Constraints

In this subsection, we shall deal with estimation, in which the rock properties at each node of the space grid are considered as separate parameters. (This corresponds to the case with NZ = 33 in Chapter 1.) For a given grid, this is the parameterization with the largest number of degrees of freedom in the unknowns; it involves no additional constraints on the property distributions. As pointed out in section 2.5, this problem is generally grossly underdetermined. For example, for the 33-point-grid with 66 unknowns, and for 3 observation locations with a total of 75 observations, the (75 x 66) sensitivity matrix $\underset{\sim}{A}$ has only 9 singular values larger than unity. Let us define the effective rank of $\underset{\sim}{A}$ to be the number of singular values larger than 0.1. Then the effective rank of the above-mentioned sensitivity matrix is 17. Thus, making corrections along more than 17 $\underset{\sim}{v}$-vectors of $\underset{\sim}{A}$ will result in unacceptably large uncertainty in the estimates. All practical algorithms of history matching make iterative corrections in only r out of possible M (= 2N in this case) dimensions; for any algorithm to yield acceptable results r must be close

to the effective rank of $\underset{\sim}{A}$.

The r-dimensional subspace of the parameter space, in which the estimate corrections are made, is different to some extent for different algorithms. Furthermore, the directions change from iteration to iteration due to changes in estimates. However, according to the results of section 2.5, the subspace spanned by a few $\underset{\sim}{v}$-vectors of $\underset{\sim}{A}$ with the largest singular values does not change significantly with the estimates used in linearization. Consequently for a given algorithm, we shall treat the r-dimensional subspace of the parameter space to be invariant. Let the (Mxr) matrix $\underset{\sim}{V}_1$ be formed by using as columns the r $\underset{\sim}{v}$-vectors of $\underset{\sim}{A}$ with the largest singular values. We shall form, using the (M-r) remaining $\underset{\sim}{v}$-vectors of $\underset{\sim}{A}$ as columns, an Mx(M-r) matrix $\underset{\sim}{V}_0$. We shall make the important hypothesis that no corrections in the estimates are made along the columns of $\underset{\sim}{V}_0$. Let $\underset{\sim}{\Lambda}_1$ be an (rxr) diagonal matrix with entries $\lambda_1 \geq \lambda_2 \geq .. \geq \lambda_r$, the largest r singular values of $\underset{\sim}{A}$; and let $\underset{\sim}{\Lambda}_0 = \text{diag}\,(\lambda_{r+1}, \lambda_{r+2}, \ldots, \lambda_M)$. We shall also assume $\lambda_{r+1}, .., \lambda_M << \lambda_1$.

As the columns of $\underset{\sim}{V}_1$ and $\underset{\sim}{V}_0$ together span the parameter space, the residual estimate error $\delta\underset{\sim}{\pi}_r \equiv \underset{\sim}{\pi}^* - \hat{\underset{\sim}{\pi}}_h$ can be expressed as

$$\delta\underset{\sim}{\pi}_r = \underset{\sim}{V}_1\underset{\sim}{a} + \underset{\sim}{V}_0\underset{\sim}{a}_0 \qquad (3.1.15)$$

where $\underset{\sim}{a}$ and $\underset{\sim}{a}_0$ are, respectively, rx1 and (2n-r)x1 vectors defined by $\underset{\sim}{V}_1^T\delta\underset{\sim}{\pi}_r$ and $\underset{\sim}{V}_0^T\delta\underset{\sim}{\pi}_r$. The reader is cautioned that, since the vector $\delta\underset{\sim}{\pi}_r$ is in general different from the deviation $\delta\underset{\sim}{\pi}$ defined in Chapter 2, the vectors $\underset{\sim}{a}$ and $\underset{\sim}{a}_0$ are different from the vectors

bearing the same symbols in Chapter 2. Then the component $\underset{\sim}{V}_1\underset{\sim}{a}$, being along the sensitive subspace, can be attributed to the difference, $\delta\underset{\sim}{y}_r \equiv y^* - y|_{\hat{\pi}_h}$; and the component $\underset{\sim}{V}_0\underset{\sim}{a}_0$ is due to the fact that no corrections are made along columns of $\underset{\sim}{V}_0$. We have from (3.1.4b)

$$\delta\underset{\sim}{y}_{hr} = \underset{\sim}{A}(\underset{\sim}{V}_1\underset{\sim}{a} + \underset{\sim}{V}_0\underset{\sim}{a}_0) + \underset{\sim}{\eta} \tag{3.1.16}$$

$$\delta\underset{\sim}{y}_{hr} = [\underset{\sim}{U}_1\ \underset{\sim}{U}_0]\begin{bmatrix} \underset{\sim}{\Lambda}_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Lambda}_0 \end{bmatrix}\begin{bmatrix} \underset{\sim}{V}_1^T \\ \underset{\sim}{V}_0^T \end{bmatrix} = \underset{\sim}{U}\underset{\sim}{\Lambda}_1\underset{\sim}{a} + \underset{\sim}{U}_0\underset{\sim}{\Lambda}_0\underset{\sim}{a}_0 + \underset{\sim}{\eta} \tag{3.1.17}$$

$$\approx \underset{\sim}{U}\underset{\sim}{\Lambda}_1\underset{\sim}{a} + \underset{\sim}{\eta} \tag{3.1.18}$$

where we have used the assumption about relative magnitudes of $\{\lambda_i\}$. The least-square solution of (3.1.18) is

$$\hat{\underset{\sim}{a}} = \underset{\sim}{\Lambda}_1^{-1}\,\underset{\sim}{U}_1^T\,\delta\underset{\sim}{y}_{hr} \tag{3.1.19}$$

Then the inverse used in the linear problem (3.1.18) is $(\underset{\sim}{U}\underset{\sim}{\Lambda}_1)^{\psi} = \underset{\sim}{\Lambda}^{-1}\underset{\sim}{U}^T$. Then applying the results (3.1.12) we have,

$$E\{\hat{\underset{\sim}{a}} - \underset{\sim}{a}\} = \underset{\sim}{0} \tag{3.1.20}$$

$$E_\eta\{(\hat{\underset{\sim}{a}} - \underset{\sim}{a})(\hat{\underset{\sim}{a}} - \underset{\sim}{a})^T\} = \underset{\sim}{\Lambda}_1^{-1}\,\underset{\sim}{U}_1^T\,\underset{\sim}{\Sigma}\,\underset{\sim}{U}_1\underset{\sim}{\Lambda}_1^{-1} \tag{3.1.21}$$

The error due to the lack of corrections in the insensitive subspace spanned by the columns of $\underset{\sim}{V}_0$ is given by

$$\underset{\sim}{a}_0 = \underset{\sim}{V}_0^T\{\underset{\sim}{\pi}_0 - \underset{\sim}{\pi}^*\} \tag{3.1.22}$$

where $\underset{\sim}{\pi}_0$ is the initial guess with which the iterative estimation

process is started. Since $\underset{\sim}{\pi}^*$ is unknown, only statistical measure of this error can be obtained; for that we need to consider the probability distribution for an appropriate ensemble of the rock properties. Taking the expectation and using the assumed invariance of $\underset{\sim}{V}_0$,

$$E_\pi\{\underset{\sim}{a}_0\} = \underset{\sim}{V}_0^T E_\pi\{\underset{\sim}{\pi}_0 - \underset{\sim}{\pi}^*\} = \underset{\sim}{V}_0^T(\underset{\sim}{\pi}_0 - \overline{\underset{\sim}{\pi}}) \tag{3.1.23}$$

where we define the ensemble mean of $\underset{\sim}{\pi}$,

$$\overline{\underset{\sim}{\pi}} = E_\pi\{\underset{\sim}{\pi}^*\} \tag{3.1.24}$$

If $\underset{\sim}{\pi}_0$ does not coincide with $\overline{\underset{\sim}{\pi}}$, there will be a bias in the estimate; we shall assume that the initial guess is chosen such that no such bias exists. Then

$$E\{\underset{\sim}{a}_0\} = \underset{\sim}{0} \tag{3.1.25}$$

$$E\{\underset{\sim}{a}_0 \underset{\sim}{a}_0^T\} = \underset{\sim}{V}_0^T E_\pi\{(\overline{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)(\overline{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)^T\}\underset{\sim}{V}_0 \tag{3.1.26}$$

We shall assume that the covariance associated with the prior estimate $\overline{\underset{\sim}{\pi}}$ is

$$\underset{\sim}{P}_0 \equiv E\{(\overline{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)(\overline{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)^T\} \tag{3.1.27}$$

Then

$$E_\pi\{\underset{\sim}{a}_0\underset{\sim}{a}_0^T\} = \underset{\sim}{V}_0^T\underset{\sim}{P}_0\underset{\sim}{V}_0 \tag{3.1.28}$$

We shall now assume that $\underset{\sim}{\eta}$ and $\underset{\sim}{\pi}$ are uncorrelated.

In order to determine the total covariance of the error in estimate, we need to add the two components. However, we note

that in (3.1.21) and (3.1.28) we have expectations over two different ensembles of $\underset{\sim}{\eta}$ and $\underset{\sim}{\pi}$. Further, note that the quantity $\underset{\sim}{a}_0$ is independent of $\underset{\sim}{\eta}$ and hence the expectation on the left in (3.1.28) can be interpreted as over both the ensembles. However, the expectation with respect to $\underset{\sim}{\pi}$ of the contribution in (3.1.21) is not easy to determine exactly, since both $\underset{\sim}{\Lambda}_1$ and $\underset{\sim}{U}_1$ depend on $\underset{\sim}{\pi}$ used in linearization. Hence, we shall again make an additional assumption that $\underset{\sim}{\Lambda}_1$ and $\underset{\sim}{U}_1$ are approximately invariant for $\underset{\sim}{\pi}$ belonging to the ensemble. Then within this approximation we can interpret the expectation on the left in (3.1.21) to be over both the ensembles of $\underset{\sim}{\eta}$ and $\underset{\sim}{\pi}$. We then obtain the following expression for the covariance, which is true approximately for all the realizations $\underset{\sim}{\pi}^*$,

$$\underset{\sim}{P}_{\pi} \equiv E_{\eta,\pi}\{(\hat{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)(\hat{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)^T\} = \underset{\sim}{V}_1 E_{\eta}\{(\hat{\underset{\sim}{a}}-\underset{\sim}{a})(\hat{\underset{\sim}{a}}-\underset{\sim}{a})^T\}\underset{\sim}{V}_1^T + \underset{\sim}{V}_0 E_{\pi}\{\underset{\sim}{a}_0\underset{\sim}{a}_0^T\}\underset{\sim}{V}_0^T \tag{3.1.29}$$

$$= \underset{\sim}{V}_1\underset{\sim}{\Lambda}_1^{-1}\underset{\sim}{U}_1^T\underset{\sim}{\Sigma}\underset{\sim}{U}_1\underset{\sim}{\Lambda}_1^{-1}\underset{\sim}{V}_1^T + \underset{\sim}{V}_0\underset{\sim}{V}_0^T\underset{\sim}{P}_0\underset{\sim}{V}_0\underset{\sim}{V}_0^T \tag{3.1.30}$$

$$= \underset{\sim}{V}_1\underset{\sim}{\Lambda}_1^{-1}\underset{\sim}{U}_1^T\underset{\sim}{\Sigma}\underset{\sim}{U}_1\underset{\sim}{\Lambda}_1\underset{\sim}{V}_1^T + (\underset{\sim}{I}-\underset{\sim}{V}_1\underset{\sim}{V}_1^T)\underset{\sim}{P}_0(\underset{\sim}{I}-\underset{\sim}{V}_1\underset{\sim}{V}_1^T) \tag{3.1.31}$$

where the last expression is arrived at by using the orthogonality of $(\underset{\sim}{V}_1 \vdots \underset{\sim}{V}_0)$.

In the absence of detailed statistical information about the observation errors, we may assume that η is zero-mean Gaussian with uniform variance for its elements,

$$\underset{\sim}{\eta} = N(\underset{\sim}{0}, \sigma_0^2\underset{\sim}{I}) \tag{3.1.32}$$

In this case the first term of (3. 1. 31) simplifies,

$$\underset{\sim}{P}_{\pi} = \sigma_0^2 \underset{\sim}{V}_1 \underset{\sim}{\Lambda}_1^{-2} \underset{\sim}{V}_1^T + (\underset{\sim}{I} - \underset{\sim}{V}_1 \underset{\sim}{V}_1^T) \underset{\sim}{P}_0 (\underset{\sim}{I} - \underset{\sim}{V}_1 \underset{\sim}{V}_1^T) \qquad (3.1.33)$$

The considerations involved in arriving at the foregoing results indicate that a statistical knowledge of the true property distribution $\underset{\sim}{\pi}^*$ is essential for the purpose of obtaining a realistic covariance of the estimates, even though such a knowledge is not used in arriving at the estimates themselves.

3. 1. 2 Covariance for Bayesian Estimation

In this approach, we start with an initial guess $\underset{\sim}{\pi}_0 (= \underset{\sim}{\bar{\pi}})$ and minimize

$$J = (\underset{\sim}{y} - \underset{\sim}{y}^0)^T \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{y} - \underset{\sim}{y}^0) + (\underset{\sim}{\pi} - \underset{\sim}{\bar{\pi}})^T \underset{\sim}{P}_0^{-1} (\underset{\sim}{\pi} - \underset{\sim}{\bar{\pi}}) \qquad (3.1.34)$$

where as defined in (3. 1. 24, 27), $\underset{\sim}{\bar{\pi}}$ and $\underset{\sim}{P}_0$ are prior statistics of the "true" property distribution $\underset{\sim}{\pi}^*$ and $\underset{\sim}{\Sigma}$ is a covariance of the observation errors. For small values of variations $\delta\underset{\sim}{y}$ and $\delta\underset{\sim}{\pi}$ in $\underset{\sim}{y}$ and $\underset{\sim}{\pi}$, we have the linear relation (2. 1. 4), $\delta\underset{\sim}{y} = \underset{\sim}{A}\delta\underset{\sim}{\pi}$. At each iteration of the history matching procedure, in the Bayesian approach, the correction $\delta\pi$ is determined by minimization of the quadratic index,

$$J' = (\delta\underset{\sim}{y} - \underset{\sim}{A}\delta\underset{\sim}{\pi})^T \underset{\sim}{\Sigma}^{-1} (\delta\underset{\sim}{y} - \underset{\sim}{A}\delta\underset{\sim}{\pi})$$

$$+ (\underset{\sim}{\pi} + \delta\underset{\sim}{\pi} - \underset{\sim}{\pi})^T \underset{\sim}{P}_0^{-1} (\underset{\sim}{\pi} + \delta\underset{\sim}{\pi} - \underset{\sim}{\pi}) \qquad (3.1.35)$$

The solution of this quadratic minimization problem is

$$\hat{\delta\underset{\sim}{\pi}} = [\underset{\sim}{A}^T \underset{\sim}{\Sigma}^{-1} \underset{\sim}{A} + \underset{\sim}{P}_0^{-1}]^{-1} [\underset{\sim}{A}^T \underset{\sim}{\Sigma}^{-1} \delta\underset{\sim}{y} - \underset{\sim}{P}_0^{-1} (\underset{\sim}{\pi} - \underset{\sim}{\bar{\pi}})] \qquad (3.1.36)$$

Then applying (3.1.36) to the estimate $\hat{\underset{\sim}{\pi}}_h$ and identifying the variation $\delta\underset{\sim}{\pi}$ and $\delta\underset{\sim}{y}$ with their residual values, the estimated correction in $\hat{\underset{\sim}{\pi}}$ is

$$\delta\hat{\underset{\sim}{\pi}}_r = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{P}_0^{-1})^{-1}\,[\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\delta\underset{\sim}{y}_{hr}-\underset{\sim}{P}_0^{-1}(\hat{\underset{\sim}{\pi}}_h-\bar{\underset{\sim}{\pi}})] \qquad (3.1.37)$$

The error in the estimate of $\delta\underset{\sim}{\pi}_r$ is

$$\delta\hat{\underset{\sim}{\pi}}_r-\delta\underset{\sim}{\pi}_r = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{P}_0^{-1})^{-1}\,[\underset{\sim}{A}\underset{\sim}{\Sigma}^{-1}\delta\underset{\sim}{y}_{hr}-\underset{\sim}{P}_0^{-1}(\hat{\underset{\sim}{\pi}}_h-\bar{\underset{\sim}{\pi}})$$

$$-\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}\,\delta\underset{\sim}{\pi}_r-\underset{\sim}{P}_0^{-1}\delta\underset{\sim}{\pi}_r] \qquad (3.1.38)$$

where, as in (3.1.4a)

$$\delta\underset{\sim}{\pi}_r = \underset{\sim}{\pi}^*-\hat{\underset{\sim}{\pi}}_h \qquad (3.1.39)$$

Then using (3.1.4b), (3.1.38) reduces to

$$\delta\hat{\underset{\sim}{\pi}}_r-\delta\underset{\sim}{\pi}_r = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{P}_0^{-1})^{-1}\,[\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\eta}-\underset{\sim}{P}_0^{-1}(\underset{\sim}{\pi}^*-\bar{\underset{\sim}{\pi}})] \qquad (3.1.40)$$

Then

$$E_{\eta,\pi}\{(\delta\hat{\underset{\sim}{\pi}}_r-\delta\underset{\sim}{\pi}_r)(\delta\hat{\underset{\sim}{\pi}}_r-\delta\underset{\sim}{\pi}_r)^T\} = E_\pi\{(\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{P}_0^{-1})^{-1}$$

$$[\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{P}_0^{-1}(\underset{\sim}{\pi}^*-\bar{\underset{\sim}{\pi}})(\underset{\sim}{\pi}^*-\bar{\underset{\sim}{\pi}})^T\underset{\sim}{P}_0^{-1}](\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{P}_0^{-1})^{-1}\} \qquad (3.1.41)$$

where we have used the independence of $\underset{\sim}{\eta}$ and $\underset{\sim}{\pi}$ and the statistics of $\underset{\sim}{\eta}$. Since in the strict sense $\underset{\sim}{A}$ depends on $\underset{\sim}{\pi}^*$, evaluation of the expectation with respect to $\underset{\sim}{\pi}$ is truly a prohibitive task. Hence, as done earlier, it will be determined in an approximate fashion. We recall from section 2.5 that the larger singular values of $\underset{\sim}{A}$ and the associated eigenvectors are almost invariant with the property distribution

$\underset{\sim}{\pi}$ used in linearization, as long as the spatial means of the properties do not change much. In the above expression $\underset{\sim}{A}$ always occurs with additive terms and the smaller singular values do not significantly influence its value. If we restrict our ensemble of $\underset{\sim}{\pi}$ such that its realizations meet this requirement (as they would if they do not have very large prior covariance), then we may treat the matrix $\underset{\sim}{A}$ to be constant, equal to its value at $\hat{\underset{\sim}{\pi}}$ while evaluating the above expectation. Then recalling the definitions of $\bar{\underset{\sim}{\pi}}$ and $\underset{\sim}{P}_0$, we obtain

$$E_{\eta,\pi}\left\{\delta\hat{\underset{\sim}{\pi}}_r - \delta\underset{\sim}{\pi}_r\right\} = \underset{\sim}{0} \tag{3.1.42}$$

$$E_{\eta,\pi}\left\{(\delta\hat{\underset{\sim}{\pi}}_r - \delta\underset{\sim}{\pi}_r)(\delta\hat{\underset{\sim}{\pi}}_r - \delta\underset{\sim}{\pi}_r)^T\right\} = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A} + \underset{\sim}{P}_0^{-1})^{-1} \tag{3.1.43}$$

In light of (3.1.10) the quantity in (3.1.43) is the covariance of the error in estimate, $\underset{\sim}{P}_\pi$:

$$\underset{\sim}{P}_\pi = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A} + \underset{\sim}{P}_0^{-1})^{-1} \tag{3.1.44}$$

We note that in (3.1.44), as in (3.1.29), the definition of $\underset{\sim}{P}_\pi$ includes expectation over both $\underset{\sim}{\eta}$ and $\underset{\sim}{\pi}$. Thus, we expect to find the a posteriori covariances computed for two different property realizations to agree very closely with one another; and we may compute the expected covariance <u>before</u> doing the estimation, using $\underset{\sim}{A}$ evaluated at $\bar{\underset{\sim}{\pi}}$.

The expression for $\underset{\sim}{P}_\pi$ can be written in an alternative form using the matrix inversion lemma (Jazwinski, 1970) as

$$\underset{\sim}{P}_\pi = \underset{\sim}{P}_0 - \underset{\sim}{P}_0\underset{\sim}{A}^T(\underset{\sim}{A}\underset{\sim}{P}_0\underset{\sim}{A}^T + \underset{\sim}{\Sigma})^{-1}\underset{\sim}{A}\underset{\sim}{P}_0 \tag{3.1.45}$$

From a computational standpoint, (3.1.45) is preferable to the

previous expression.

3.1.3 Covariance for Estimation Under Hard Constraints

The zonation approach and the procedure for approximate implementation of the Bayesian estimation discussed in Chapter 1 involve hard constraints on the corrections in the estimates. These constraints confine the corrections in the estimates to an M-dimensional subspace of the parameter space; M depends on the parameterization used; for example, in case of estimation of k and ϕ using NZ zones M = 2NZ, and in the approximate implementation of Bayesian estimation M is simply the number of eigenvectors of $\underset{\sim}{P}_0$ used as a basis for correction in the 2N-dimensional parameter space. Mathematically this can be expressed by (again, assuming $\underset{\sim}{\pi}_0 = \underset{\sim}{\bar{\pi}}$),

$$\underset{\sim}{\pi} - \underset{\sim}{\bar{\pi}} = \underset{\sim}{G}_1 \underset{\sim}{\xi} \qquad (3.1.46)$$

where $\underset{\sim}{G}_1$ is a 2N x M matrix with rank M, its columns being orthonormal vectors which span the admissible parameter space. For example, in the zonation constraint, if the elements $\{\pi_i, \pi_{i+1}, \ldots, \pi_{i+m}\}$ are constrained to be equal, then one column of $\underset{\sim}{G}_1$ will have all elements zero except for those with subscripts $\{i, i+1, .., i+m\}$ which will all be equal to $1/\sqrt{m+1}$. In the numerical implementation of the Bayesian approach, the columns of $\underset{\sim}{G}_1$ will be simply the eigenvectors of $\underset{\sim}{P}_0$ corresponding to the largest M eigenvalues. Due to orthogonality of the columns of $\underset{\sim}{G}_1$, (3.1.46) implies $\underset{\sim}{\xi} = \underset{\sim}{G}_1^T(\underset{\sim}{\pi} - \underset{\sim}{\bar{\pi}})$.

The chain rule of differentiation yields the linear relation between the small variations $\delta\underset{\sim}{y}$ and $\delta\underset{\sim}{\xi}$,

$$\delta\underset{\sim}{y} = \underset{\sim}{A}\,\underset{\sim}{G}_1\,\delta\underset{\sim}{\xi} \qquad (3.1.47)$$

A decomposition of the matrix $(\underset{\sim}{A}\underset{\sim}{G}_1)$ in a manner similar to that used for (2. 2. 4) leads to

$$\underset{\sim}{A}\,\underset{\sim}{G}_1 = \overline{\underset{\sim}{U}}\,\overline{\underset{\sim}{\Lambda}}\,\overline{\underset{\sim}{V}}^T \tag{3. 1. 48}$$

Out of the M singular values in $\overline{\underset{\sim}{\Lambda}}$, let r_1 be large and $(M-r_1)$ be comparatively small; we shall assume as before that no corrections in $\underset{\sim}{\xi}$ are made along the $(M-r_1)$ columns of $\overline{\underset{\sim}{V}}$ associated with the smaller singular values. Let these column vectors in $\overline{\underset{\sim}{V}}$ form an $Mx(M-r_1)$ matrix $\overline{\underset{\sim}{V}}_0$, and let (Mxr_1) matrix $\overline{\underset{\sim}{V}}_1$ be formed from the rest. Let the corresponding columns of $\overline{\underset{\sim}{U}}$ form $\overline{\underset{\sim}{U}}_0$ and $\overline{\underset{\sim}{U}}_1$ respectively. We shall assume (see section 2. 3) that the subspace spanned by the columns of $\overline{\underset{\sim}{V}}_1$ is invariant with $\underset{\sim}{\pi}$ used for linearization. Then we have

$$\delta\underset{\sim}{\xi} = \overline{\underset{\sim}{V}}_1\,\underset{\sim}{\zeta}_1 + \overline{\underset{\sim}{V}}_0\,\underset{\sim}{\zeta}_0 \tag{3. 1. 49}$$

and we may assume that only the component $\overline{\underset{\sim}{V}}_1\underset{\sim}{\zeta}_1$ of $\underset{\sim}{\xi}$ is updated during estimation, whereas no corrections are made in the component $\overline{\underset{\sim}{V}}_0\underset{\sim}{\zeta}_0$. Then using the same reasoning as that for the covariance of the unconstrained case and using analogous notation, we have for the residuals after history matching (i. e. replace $\delta\underset{\sim}{\xi}$ by $\delta\underset{\sim}{\xi}_r$),

$$E_{\eta,\pi}\{(\hat{\underset{\sim}{\zeta}}_1-\underset{\sim}{\zeta}_1)(\hat{\underset{\sim}{\zeta}}_1-\underset{\sim}{\zeta}_1)^T\} = \overline{\underset{\sim}{\Lambda}}_1^{-1}\,\overline{\underset{\sim}{U}}_1^T\,\underset{\sim}{\Sigma}\,\overline{\underset{\sim}{U}}_1\,\overline{\underset{\sim}{\Lambda}}_1^{-1} \tag{3. 1. 50}$$

where $\overline{\underset{\sim}{\Lambda}}_1$ is the r_1xr_1 diagonal matrix of the largest singular values in $\overline{\underset{\sim}{\Lambda}}$. Furthermore, the covariance of the error in the estimate $\hat{\delta\underset{\sim}{\xi}}_r$ is

$$E_{\eta,\pi}\{(\delta\hat{\underset{\sim}{\xi}}_r - \delta\underset{\sim}{\xi}_r)(\delta\hat{\underset{\sim}{\xi}}_r - \delta\underset{\sim}{\xi}_r)\}$$

$$= \overline{\underset{\sim}{V}}_1 \overline{\underset{\sim}{\Lambda}}_1^{-1} \overline{\underset{\sim}{U}}_1^T \underset{\sim}{\Sigma}\, \overline{\underset{\sim}{U}}_1 \overline{\underset{\sim}{\Lambda}}_1^{-1} \overline{\underset{\sim}{V}}_1^T + \overline{\underset{\sim}{V}}_0 E_\pi\{\underset{\sim}{\zeta}_0 \underset{\sim}{\zeta}_0^T\}\, \overline{\underset{\sim}{V}}_0^T \qquad (3.1.51)$$

Using the definitions (3.1.46) and (3.1.49), the orthogonality of the columns of $\underset{\sim}{G}_1$ and $\overline{\underset{\sim}{V}}_0$, and the assumption of noncorrection in the components along the columns of $\overline{\underset{\sim}{V}}_0$, we obtain

$$E_\pi\{\underset{\sim}{\zeta}_0 \underset{\sim}{\zeta}_0^T\} = \overline{\underset{\sim}{V}}_0^T \underset{\sim}{G}_1^T E_\pi\{(\overline{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)(\overline{\underset{\sim}{\pi}} - \underset{\sim}{\pi}^*)^T\}\, \underset{\sim}{G}_1 \overline{\underset{\sim}{V}}_0$$

$$= \overline{\underset{\sim}{V}}_0^T \underset{\sim}{G}_1^T \underset{\sim}{P}_0 \underset{\sim}{G}_1 \overline{\underset{\sim}{V}}_0 \qquad (3.1.52)$$

As in the previous discussion, the second term in (3.1.51) represents the error due to the lack of corrections in $\overline{\underset{\sim}{V}}_0^T(\underset{\sim}{\xi} - \underset{\sim}{\xi}^*)$. In addition to this, another error is introduced due to the hard constraint; it prevents the corrections in $(\hat{\underset{\sim}{\pi}} - \overline{\underset{\sim}{\pi}})$ along the (2N-M) dimensional subspace, which is the orthogonal complement of the subspace spanned by the columns of $\underset{\sim}{G}_1$. The ensemble average of this error is zero; however, it contributes to the covariance. Since this and the error in (3.1.52) derive from the same prior probability distribution for the ensemble of the reservoir properties, the two are correlated. Thus we have the following three additional terms in the expression for the covariance error in the estimate $\hat{\underset{\sim}{\pi}}$,

$$(\underset{\sim}{I} - \underset{\sim}{G}_1 \underset{\sim}{G}_1^T)\, \underset{\sim}{P}_0\, (\underset{\sim}{I} - \underset{\sim}{G}_1 \underset{\sim}{G}_1^T) + (\underset{\sim}{I} - \underset{\sim}{G}_1 \underset{\sim}{G}_1^T) \underset{\sim}{P}_0 \underset{\sim}{G}_1 \overline{\underset{\sim}{V}}_0 \overline{\underset{\sim}{V}}_0^T \underset{\sim}{G}_1^T$$

$$+ \underset{\sim}{G}_1 \overline{\underset{\sim}{V}}_0 \overline{\underset{\sim}{V}}_0^T \underset{\sim}{G}_1^T \underset{\sim}{P}_0\, (\underset{\sim}{I} - \underset{\sim}{G}_1 \underset{\sim}{G}_1^T) \qquad (3.1.53)$$

We note that the first term in the last expression is independent of $\overline{\underset{\sim}{V}}_1$ and $\overline{\underset{\sim}{V}}_0$; it only depends on $\underset{\sim}{P}_0$ and $\underset{\sim}{G}_1$. Hence, once the zonation is decided, for a given $\underset{\sim}{P}_0$, this error is fixed; and it is independent of the property distributions used in linearization for determination of $\underset{\sim}{A}$. We also note that the second and third terms are merely transpositions of each other and represent correlation of the same two errors. Furthermore, as the two errors are orthogonal to each other - one in the subspace spanned by columns $\underset{\sim}{G}_1$ and the other in the orthogonal complement, we may expect them to exhibit some special behaviour. As shown in Appendix 3.1 they have zero trace. Thus they contribute to the cross-correlation elements in the covariance matrix and even to the diagonal elements representing the individual parameter variances, but do not contribute to the trace which is the sum of these variances.

Adding all the contributions up, we obtain

$$\underset{\sim}{P}_\pi = \underset{\sim}{G}_1\overline{\underset{\sim}{V}}_1\overline{\underset{\sim}{\Lambda}}_1^{-1}\overline{\underset{\sim}{U}}_1^T\underset{\sim}{\Sigma}\overline{\underset{\sim}{U}}_1\overline{\underset{\sim}{\Lambda}}_1^{-1}\overline{\underset{\sim}{V}}_1^T\underset{\sim}{G}_1^T+\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\underset{\sim}{P}_0\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T$$

$$+(\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T)\underset{\sim}{P}_0(\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T)+(\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T)\underset{\sim}{P}_0\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T$$

$$+\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\underset{\sim}{P}_0(\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T) \qquad (3.1.54)$$

In the absence of precise information about the observation errors, $\underset{\sim}{\Sigma}$ may be assumed to be $\sigma^2\underset{\sim}{I}$; this yields a simplification of the first term on the right of (3.1.54) to

$$\sigma^2\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_1\overline{\underset{\sim}{\Lambda}}^{-2}\overline{\underset{\sim}{V}}_1^T\underset{\sim}{G}_1^T \qquad (3.1.55)$$

Expression (3.1.54) can be used for the zonation and the Bayesian approach when no penalty is levied along the M admissible vectors. The case of estimation without any constraints, discussed earlier in subsection (3.1.1), is a special case of the approach with hard constraints, the number of constraints being zero and M = 2N. Then the parameters being estimated are simply the N grid point values of each of k and ϕ, implying $\underset{\sim}{G}_1 = \underset{\sim}{I}$; this value of $\underset{\sim}{G}_1$ reduces (3.1.54) to (3.1.31). (Note that for $\underset{\sim}{G}_1 = \underset{\sim}{I}$, $\overline{\underset{\sim}{V}}_1 = \underset{\sim}{V}_1$, $\overline{\underset{\sim}{\Lambda}}_1 = \underset{\sim}{\Lambda}_1$, $\overline{\underset{\sim}{V}}_0 = \underset{\sim}{V}_0$, $\overline{\underset{\sim}{U}}_1 = \underset{\sim}{U}_1$.)

We shall close this section with some general remarks. In the foregoing discussion we assume corrections only in the sensitive subspace and none in its orthogonal complement. This assumption is only approximately satisfied by the different history matching algorithms considered in Chapter 1. As a result, while implementing the above results, a decision has to be made as to where the cutoff lies in the singular values that leads to this division. The covariance will vary with the selected cutoff. In practice, this cutoff can be decided by trial and error as follows. If some additional small singular values are included in the inverse, their contribution will increase the covariance; this is accompanied by some reduction in the error term containing $\underset{\sim}{V}_0$. Thus it seems plausible that an optimum cutoff will exist at the smallest value for a scalar measure of the covariance (the trace, for example). In the following section we investigate this possibility numerically and find that such an optimum value for the cutoff does exist. Then this value of the cutoff may be selected for covariance prediction.

While determining the cutoff, the properties of the algorithm used for history matching may also be taken into account. For the gradient algorithm only a few of the larger singular values must be included in the inverse, yielding a somewhat higher contribution by the error due to noncorrection. Whereas for the Marquardt method, as pointed out in section 2.3, the components of the correction along the $\underset{\sim}{y}$-vector with the singular value closest to $\sqrt{\mu}$ will be dominant; therefore the cutoff in this case should be at singular values substantially smaller than $\sqrt{\mu}$. Looking at it from a different angle, an optimal value for μ may be selected as follows. As noted in the previous paragraph (and verified numerically in section 3.2), an optimum cutoff exists, which leads to the smallest norm or trace of the a posteriori covariance. Then the value of μ may be so selected that the virtual dimensionality of the correction subspace is as close as possible to this optimum cutoff. We present numerical results concerning this in the following section.

As discussed in detail in section 2.5 and briefly in subsection 3.1.2, the subspace spanned by the $\underset{\sim}{y}$-vectors of $\underset{\sim}{A}$ with the larger singular values is approximately independent of the property distributions used for linearization, as long as their spatial means do not change very much; it primarily depends on the conditions of the estimation problem, for example the number and location of the observation sites and production wells, the number and times of the observations at different sites, the production rate histories for different wells. Thus, these vectors can be computed using $\underset{\sim}{A}$ corresponding to the prior estimates and used for the covariance

computation. The computed covariances will be valid approximately for the whole ensemble of the reservoir properties. This makes it possible to compute the covariances for a given estimation problem before doing the estimation.

3.2 Computational Results on Covariances

In this section, we present results concerning numerical computation of the covariances for the different parameterizations considered in the last section.

The conditions of the simulated estimation problem are those of the set S_2 detailed in section 1.4. (They are for a one-dimensional reservoir with three observation sites and two production wells; the production rates are piece-wise constant and there are 25 observations at each site, taken at uniform time intervals from the start of the production.) The observation errors are assumed to be Gaussian random variables independently distributed with zero mean and a variance of 1 psi^2 (implying $\underset{\sim}{\Sigma} = \underset{\sim}{I}$).

The sensitivity matrix $\underset{\sim}{A}$, required in the covariance computation was evaluated for uniform property distributions corresponding to their respective ensemble means of 5.0 md and 0.2. To investigate the effect of the property distribution used in the linearization, some of the computations were repeated using $\underset{\sim}{A}$ for the realization R-2.

The prior covariance $\underset{\sim}{P}_0$, for the computations in this section, is taken to be the "true" matrix which was used in generating the realizations of the property distributions. We shall study the influence of variation in $\underset{\sim}{P}_0$ on the predicted covariances in section 3.4

and the effect of erroneous specification of $\underset{\sim}{P}_0$ in section 3.5.

We shall carry out the analysis in a manner similar to that of section 3.1, taking each parameterization separately. First we shall treat the case of estimation with no explicit constraints; then the Bayesian estimation will be dealt with; finally the zonation will be studied as an example of the hard constraints.

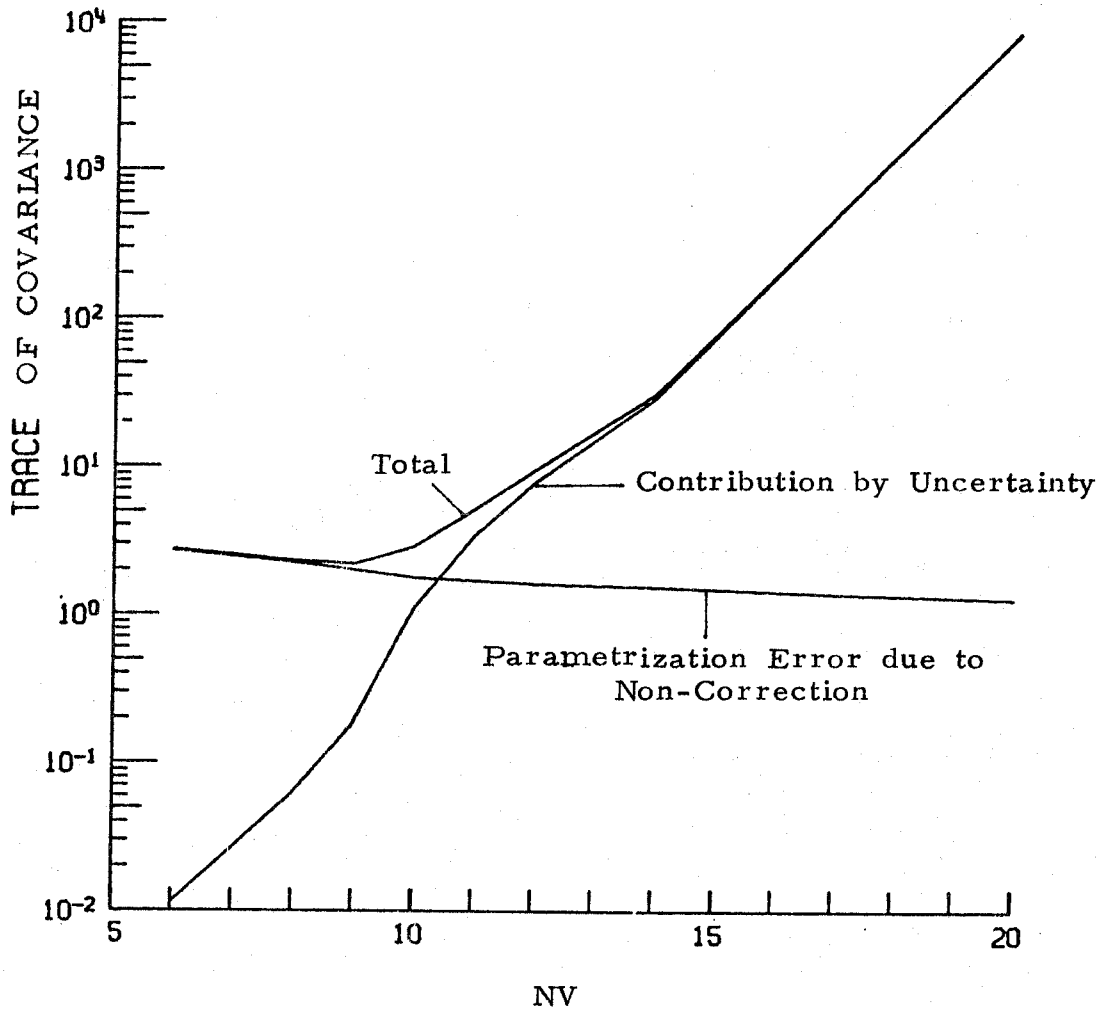
We shall not present the full (66 x 66) covariance matrices, and instead will concentrate on their most important elements, namely those on the principal diagonal. In addition, a comparative study requires a scalar measure for the matrices. We shall use the trace for this purpose; not only is it easy to compute, but also it is a measure of the total uncertainty in the estimates of all the parameters, since it is the sum of their variances. All the variances reported in the remainder of this chapter are for the normalized properties $k_i/\bar{k}$ and $\phi_i/\bar{\phi}$. In this section, we shall attempt to understand and interpret the predicted variances of the 66 parameters, and the contribution to them by the different terms in the expressions of $\underset{\sim}{P}_\pi$ in the respective parameterizations. In the non-Bayesian parameterization, we shall also study the relative importance of the contributions by the uncertainty (term depending on $\underset{\sim}{\Lambda}_1$) and the total error due to parameterization which includes the errors due to noncorrection and the hard constraint and their cross-correlation. This leads to the study of trade-off between the two contributions.

3.2.1 Covariance for Estimation with no Explicit Constraints

Expression (3.1.31) is used in the computation for this parameterization. As it is not evident a priori what value to use for r,

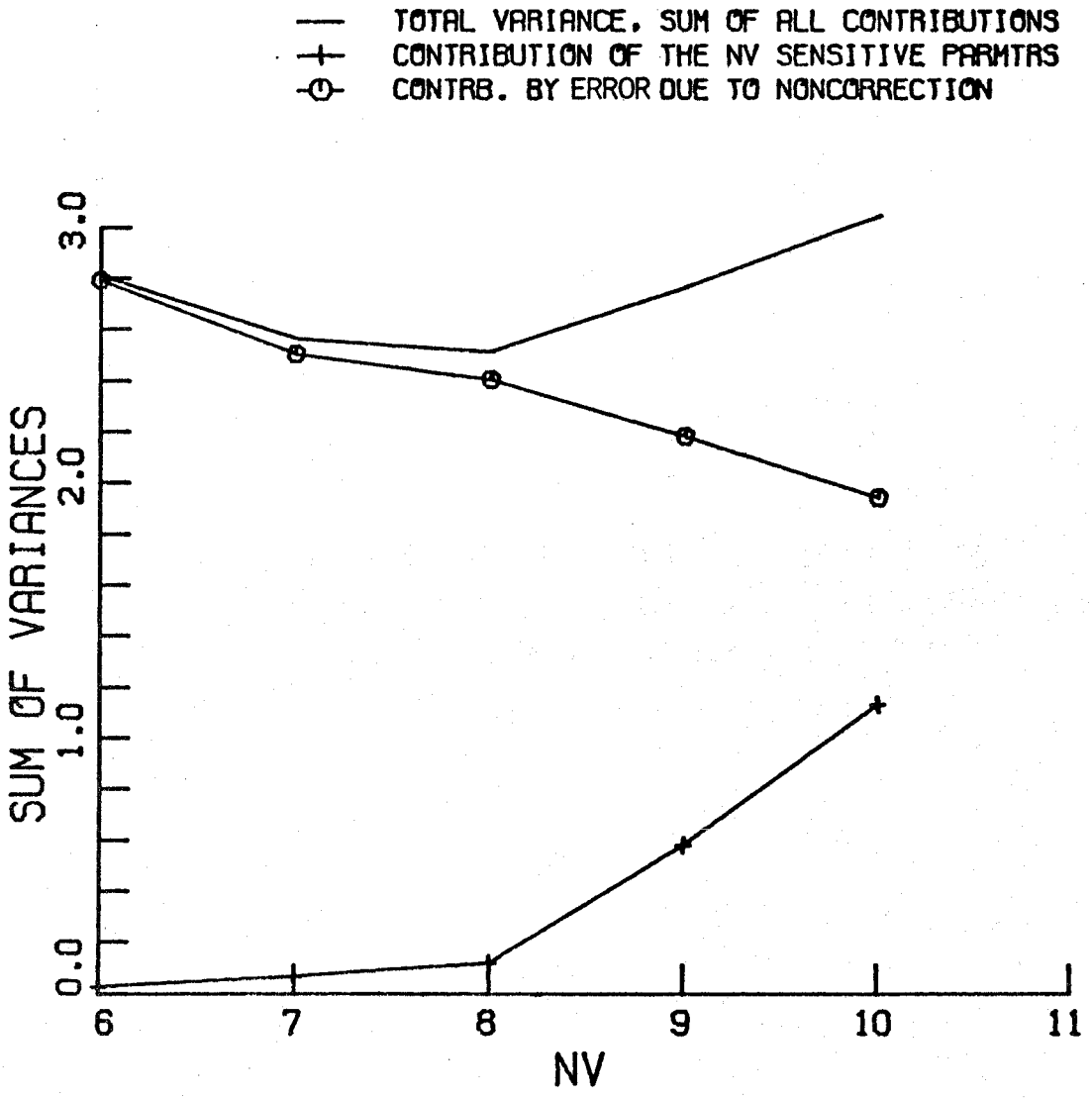
the rank of $\underset{\sim}{V}_1$, we shall let it be a variable, N_v, and study its influence on the different aspects of the result. Figure (3. 2. 1) shows the variation of the trace of $\underset{\sim}{P}_\pi$ corresponding to the normalized rock properties, and the contributions to it by the two terms on the right hand side of (3. 1. 33). The trace of the total covariance is minimum for $N_v = 9$. As N_v increases, the contribution of the first term involving the eigenvalues λ_i (which are the sensitivities of the components of $\underset{\sim}{\pi}$ along $\underset{\sim}{v}_i$) increases rather sharply, whereas the contribution of the term representing the ensemble average of the square of the error due to noncorrection (identified as "error due to noncorrection" in the sequel) diminishes very gradually with increasing N_v. (Note also that in the figures, N_v is written as NV). As a result, the trace of the sum is not symmetric about the minimum. Figure (3.2. 2) shows these results for the case when the sensitivity matrix $\underset{\sim}{A}$ is evaluated for the property distribution R-2. The results in the two cases are very similar. However, the minimum of the trace of the total covariance in the latter case occurs at $N_v = 8$, and the curve is shallower near the minimum. The similarity in the two cases is encouraging, in view of the fact that the R-2 distributions are significantly different from the uniform distributions.

Figure (3. 2. 3) shows the variances of the estimates of the grid point rock properties, which are the diagonal elements of the covariance matrix, for $N_v = 9$ and linearization about the uniform distributions. The association of each element of the parameter vector $\underset{\sim}{\pi} \equiv (\underset{\sim}{k}^T \mid \underset{\sim}{\phi}^T)^T$ with a region in the reservoir is exploited in



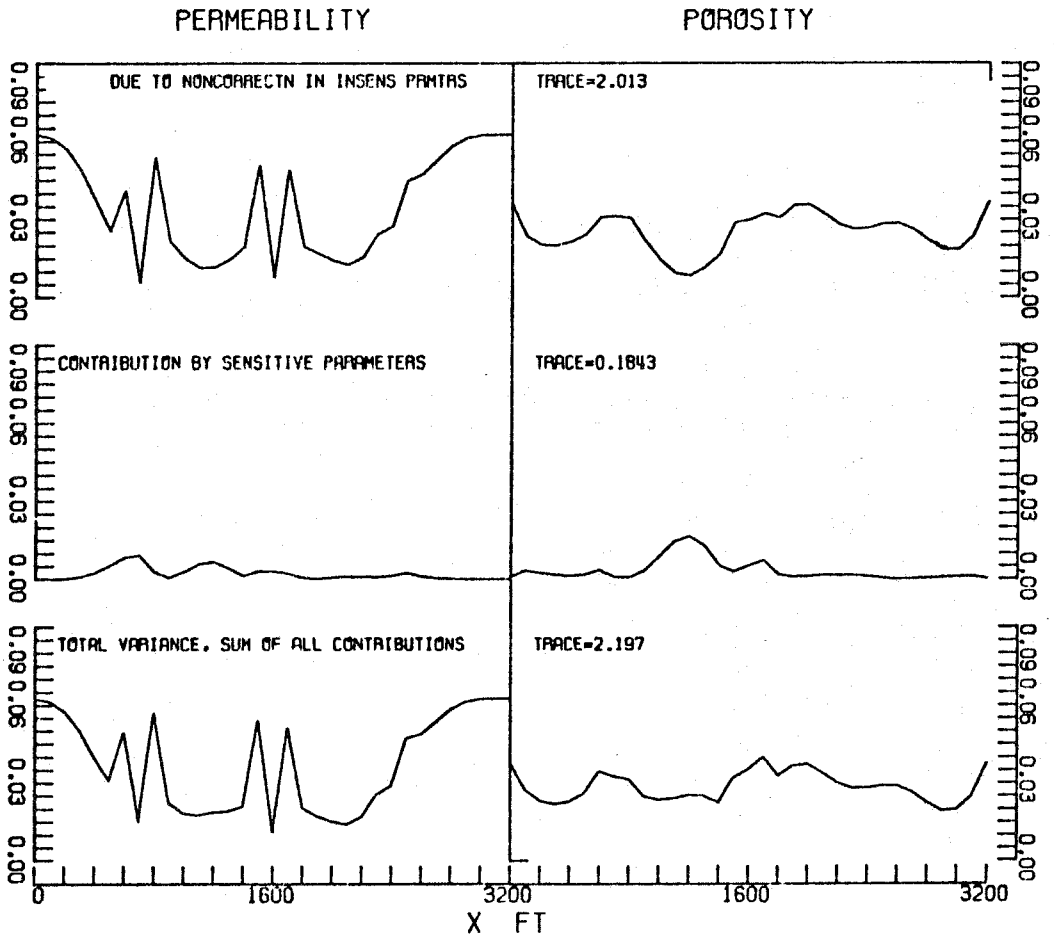
Trade-off between Uncertainty and Parametrization Error due to Non-correction NZ = 33 Linearization about Uniform k, φ.

Figure 3.2.1



TRADEOFF BETWEEN UNCERTAINTY DUE TO
INSENSITIVITY AND ERROR DUE TO NONCORRECTION
NZ=33 LINEARIZATION ABOUT R-2

Figure 3.2.2

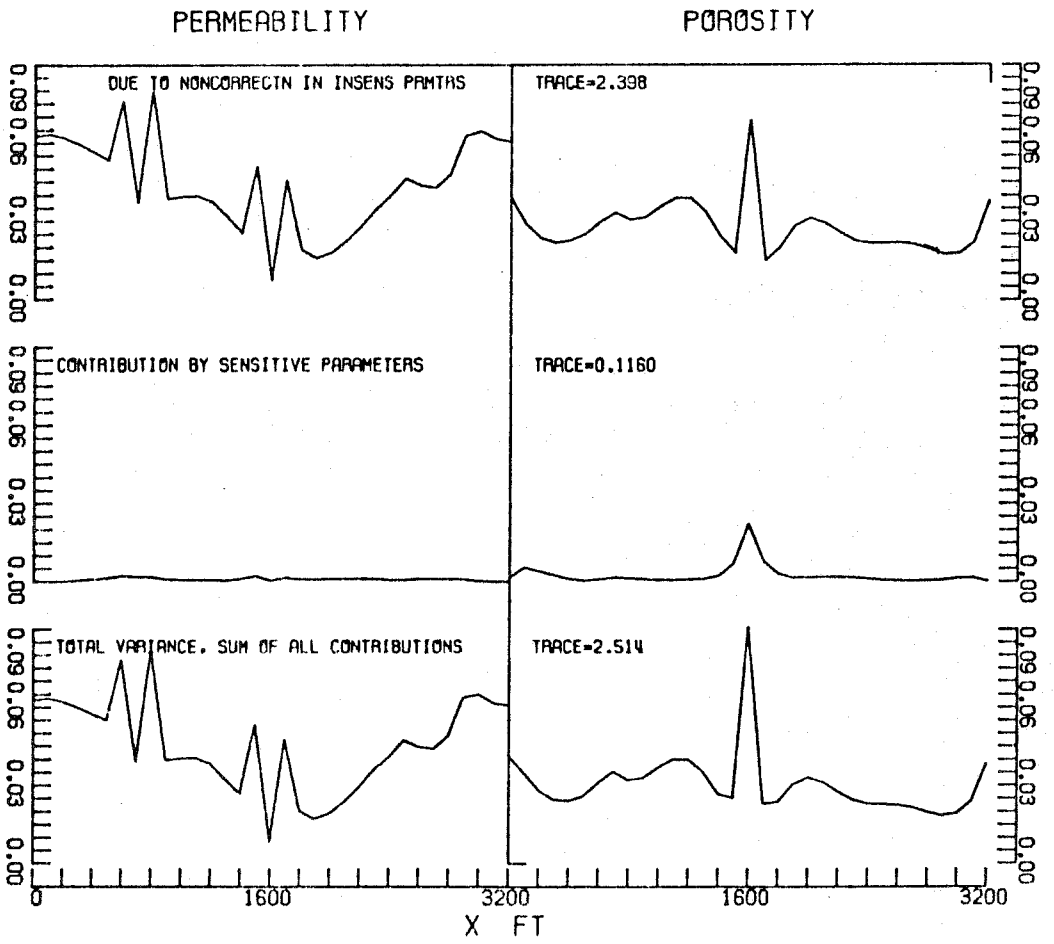


CONTRIBUTION BY VARIOUS TERMS TO VARIANCE
NO CONSTRAINTS, NZ=33, UNIFORM K,PHI

Figure 3.2.3

the identification of the abscissae by the physical location in the reservoir for each property. The figure also shows the contribution to the variance of each π_i by the sensitivity (containing $\{\lambda_j\}$) term and the other term in (3.1.33). The traces of each term and of the total variances are also included. For this problem, the trace of contribution by the error due to noncorrection is much larger than the trace of the other term, for the value of N_v at which the minimum of $tr(\underset{\sim}{P}_\pi)$ occurs. An examination of the distributions in figure (3.2.3) reveals that the contribution of the $\underset{\sim}{V}_0$ term is high in the regions of low sensitivities (for example as defined by the index of insensitivity of section 2.7), whereas the other contribution is very small in these regions, as expected. The contribution by the more sensitive parameter in the uncertainty term is largest in the regions of intermediate sensitivities, as these parameters correspond to the $\underset{\sim}{v}$-vectors with the smaller singular values which are included in this term. The kinks indicating rapid fluctuations in the sensitivity of the permeability in the vicinity of the production wells, fig. (1.5.3), are reflected in the variances through the $\underset{\sim}{V}_0$ term contribution. The variance of the porosity in the region around x = 1200 ft is low because of the high sensitivity due to the existence of the two production wells on its either side. The high sensitivity of the porosity in the regions close to the boundaries, reflected by lower variances, results from the fact that the oil removed from these regions cannot be replenished by any inflow because of the impermeable boundaries.

Figure (3. 2. 4) shows similar results for linearization about R-2 and for $N_v = 8$. The variances of the permeability estimates are very similar to those in the case discussed above. However, the porosity variance has a large "spike" at the central production well when A for R-2 is used, while no such anomaly exists for the other case. This difference is due to two factors. Firstly, the analysis rests on the cutoff N_v that separates rather arbitrarily the eigenvectors into sensitive and insensitive sets; furthermore the factor $\underset{\sim}{\Lambda}^{-2}$ in the first term in the expression for the covariance sharpens the effect of the cutoff, as the λ_i closest to the cutoff yields the largest contribution to this term. The cutoffs are different in the two cases. Furthermore, we note that $\underset{\sim}{v}$-vectors of $\underset{\sim}{A}$ for the two cases do not have an exact one to one relation after the first six singular values. (See figure 2. 5. 5). In particular, the seventh and the eighth vectors for R-2 have pronounced kinks whereas no such kinks exist in the first 9 $\underset{\sim}{v}$-vectors for uniform distributions. These differences become dominant as discussed above. This tendency for abrupt local changes in the computed variances is a shortcoming of the proposed procedure for covariance prediction. However, it must be recalled that the distributions of R-2 are significantly different from the mean values (see figure 1.4.2.) and in this sense this is a rather extreme test. Even in this extreme test, the computed variances for most of the parameters do not differ significantly in the two cases.



CONTRIBUTION BY VARIOUS TERMS TO VARIANCE
NO CONSTRAINTS. NZ=33. LIN. R-2

Figure 3.2.4

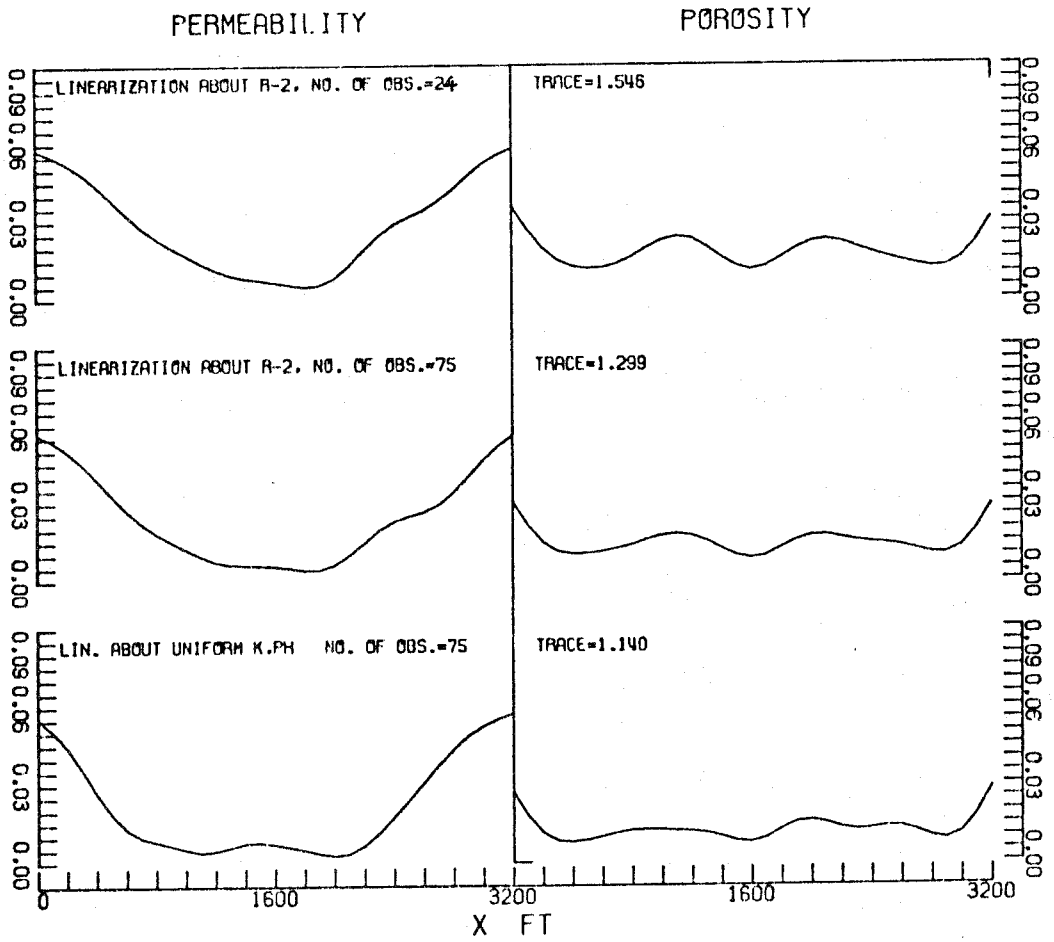
3.2.2 Covariance for Bayesian Estimation

The covariance matrix for the estimate resulting from the Bayesian approach was calculated using (3.1.45).

In order to reduce the computational requirements, the sensitivity matrix $\underset{\sim}{A}$ was not computed afresh; instead, it was reconstructed from its largest 20 singular values and associated $\underset{\sim}{u}$ and $\underset{\sim}{v}$-vectors. From the plot of the singular values (figure (2.5.1)), it is evident that the ratio λ_{21}/λ_1 is approximately 10^{-6}, indicating a high degree of accuracy in the reconstructed matrix. The reconstructed sensitivities were verified by comparison with their exact values, and except for a few smallest elements four-to-five decimal place agreement was observed. As the term containing $\underset{\sim}{A}$ is added to a positive definite matrix in the second term of (3.1.45), the smaller elements of $\underset{\sim}{A}$ do not contribute significantly to it. The neglect of the smaller singular values of $\underset{\sim}{A}$ while its reconstruction is tantamount to rounding them off to zero, and this will yield a slightly conservative covariance.

The inversion of the (L x L) symmetric matrix in (3.1.45) was carried out accurately via its eigenvalue-eigenvector decomposition using the program of Bushinger and Golub (1969).

The parameter variances for the Bayesian estimation are plotted in figure (3.2.5); it shows two sets of variances resulting from the linearization about R-2 and the uniform property distributions. The variance distributions for the two cases are very similar. The variances of the permeability in the regions around x = 1000 ft and x = 2200 ft are significantly higher for linearization about R-2.



VARIANCE FOR BAYESIAN ESTIMATION

Figure 3.2.5

This is due to substantially higher permeabilities in these regions for R-2 than the mean value, leading to a drop in the pressure gradients in these regions; the sensitivity of observation with respect to the permeability in a region is approximately proportional to the time average of the square of the pressure gradients in that region (see Appendix 1.4). Similarly the porosity estimate variances in the region (800 ft-1400 ft) are larger for linearization about R-2 due to a lower influence on the observation of the porosity in that region. This is due to high porosity values than the mean in these regions for R-2, leading to a smaller average time-rate of pressure decline which is directly proportional to the ϕ-sensitivity.

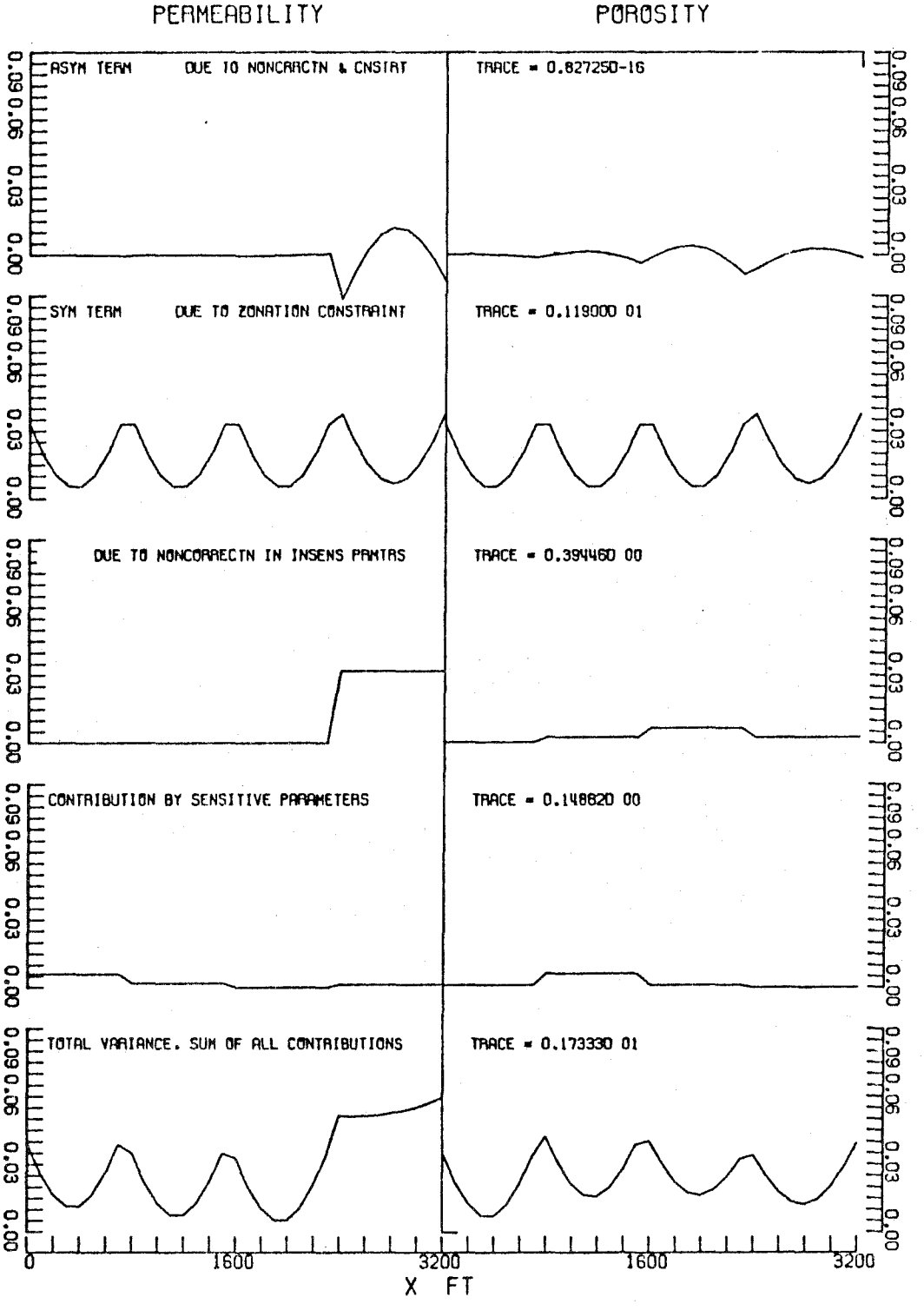
Figure (3.2.5) also shows the variances for the case when the estimation is carried out using a total of 24 observations, 8 at each observation site, taken over the same period at uniform intervals. The linearization in this case is about R-2. The distribution of the variances is similar to that for 75 observations, but they are somewhat higher. However, the small difference in the traces for the two cases indicates that taking three times as many observations at the same location in a given period of time does not increase significantly the information content of the data.

3.2.3 Covariance for Estimation with Hard Constraints

Expressions (3.1.54-55) were used together with $\underset{\sim}{A}$ for linearization about the uniform distributions in this case. The zonations involving 4, 8 and 16 almost uniform zones detailed in Chapter 1 were treated. In each case the calculations were repeated with different

values of N_v, the rank of $\overline{\underset{\sim}{V}}_1$, and the trade-off between the uncertainty and the total parameterization error was examined. As in the case with no explicit constraints (see the foregoing subsection 3.2.1), the trace of the total covariance was minimum for a certain value of N_v; these optimal values were 7, 9 and 9 respectively for NZ = 4, 8 and 16. The details of the distributions of the parameter variances (diagonal elements of $\underset{\sim}{P}_\pi$) including the contributions by the various terms were studied for these values of N_v in each of the zonations. The results and their discussion follow.

The results for the four-zone case are shown in figure (3.2.6). The contribution of the N_v (= 7) sensitive parameters ($\underset{\sim}{v}$-vectors), given by (3.1.55), is small in regions of very high or very low sensitivity. The contribution by the second term which represents the ensemble average of the square of the error due to noncorrection along the remaining ($2NZ-N_v$) $\underset{\sim}{v}$-vectors is largest in the zones of smallest influence. The third term contribution, which is the ensemble average of the square of the error due to the hard constraint of zonation, is the largest near the zone boundaries. This can be explained from the following: the representative value of a rock property in a zone is close to the mean of its "true" distribution in that zone; for a smooth property distribution, implied by the assumed $\underset{\sim}{P}_0$, the mean is close to the value of the distribution at the center of a zone. Thus, on the average, the zonation constraint causes the minimum error at the center and the maximum error at the boundary of a zone. The distribution of this zonation error has a uniform pattern almost all over the reservoir because the pre-



CONTRIBUTION BY VARIOUS TERMS TO VARIANCE
ZONATION APPROACH. NZ=4

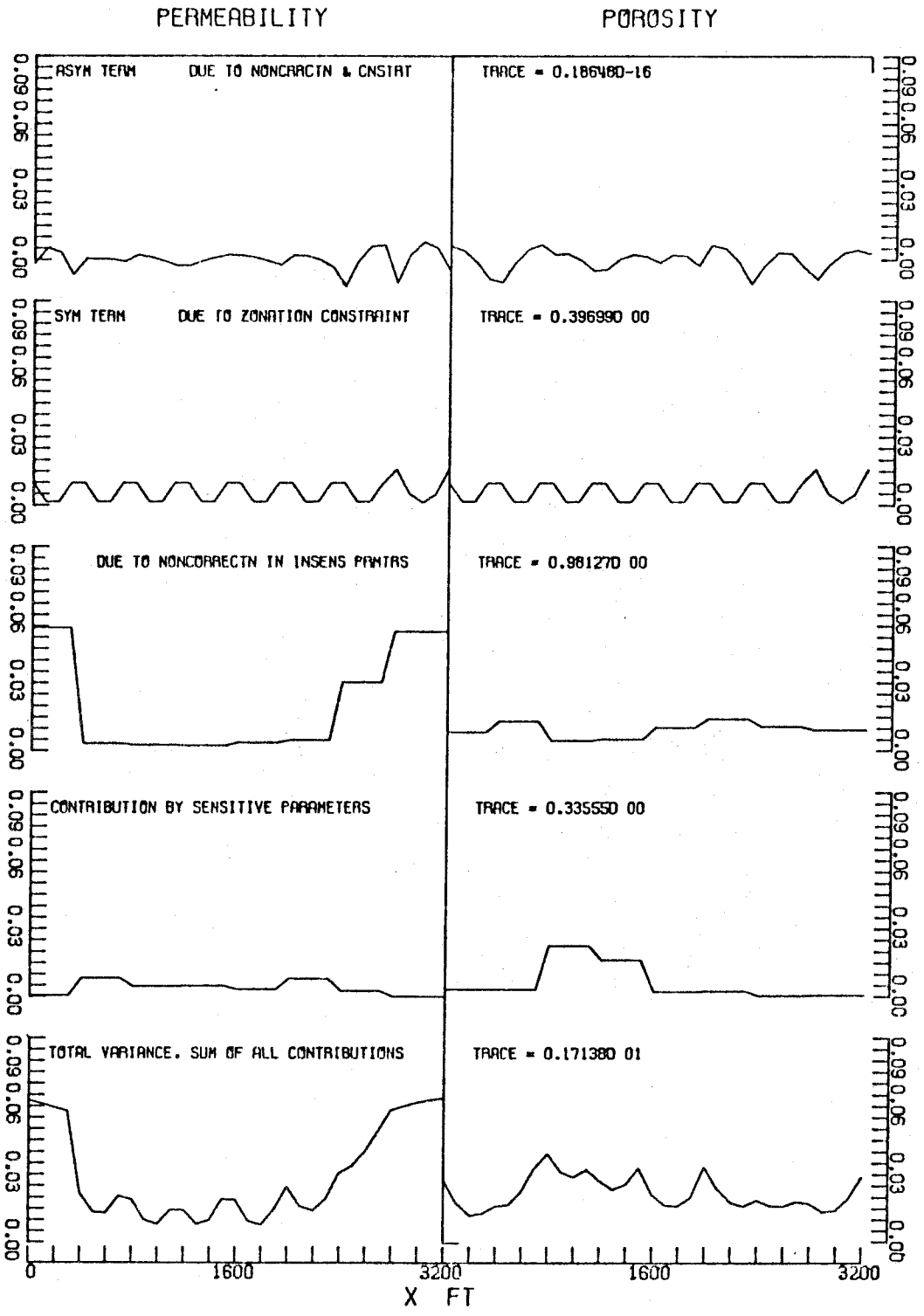
Figure 3.2.6

scribed $\underset{\sim}{P}_0$ represents a homogeneous random process for the property distribution and also all the zones except one are uniform in size. Moreover, since for the normalized property distributions the submatrices in $\underset{\sim}{P}_0$ corresponding to $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ are identical, these error distributions are identical for both the properties.

The last two terms in (3.1.54) are the cross-correlation of the errors due to the hard constraint of zonation and the noncorrection along the columns of $\underset{\sim}{\overline{V}}_0$. They are the transpose of each other, hence only one plot of their diagonal elements is shown in the figure. As discussed in section 3.1, and proved in appendix 3.1, these terms have null traces; this is verified computationally by the reported value. The contribution of the cross-correlation terms is significant only in the regions of low sensitivity, where the noncorrection error is significant. Furthermore this contribution has a negative minimum at the zone boundaries and a positive maximum at the center. It is positive at the center because the error due to noncorrection will be, when the initial guess coincides with the ensemble mean, on the average of the same sign as the error committed at the center of the zone due to the lumping for zonation. The reverse is true close to the zone boundaries.

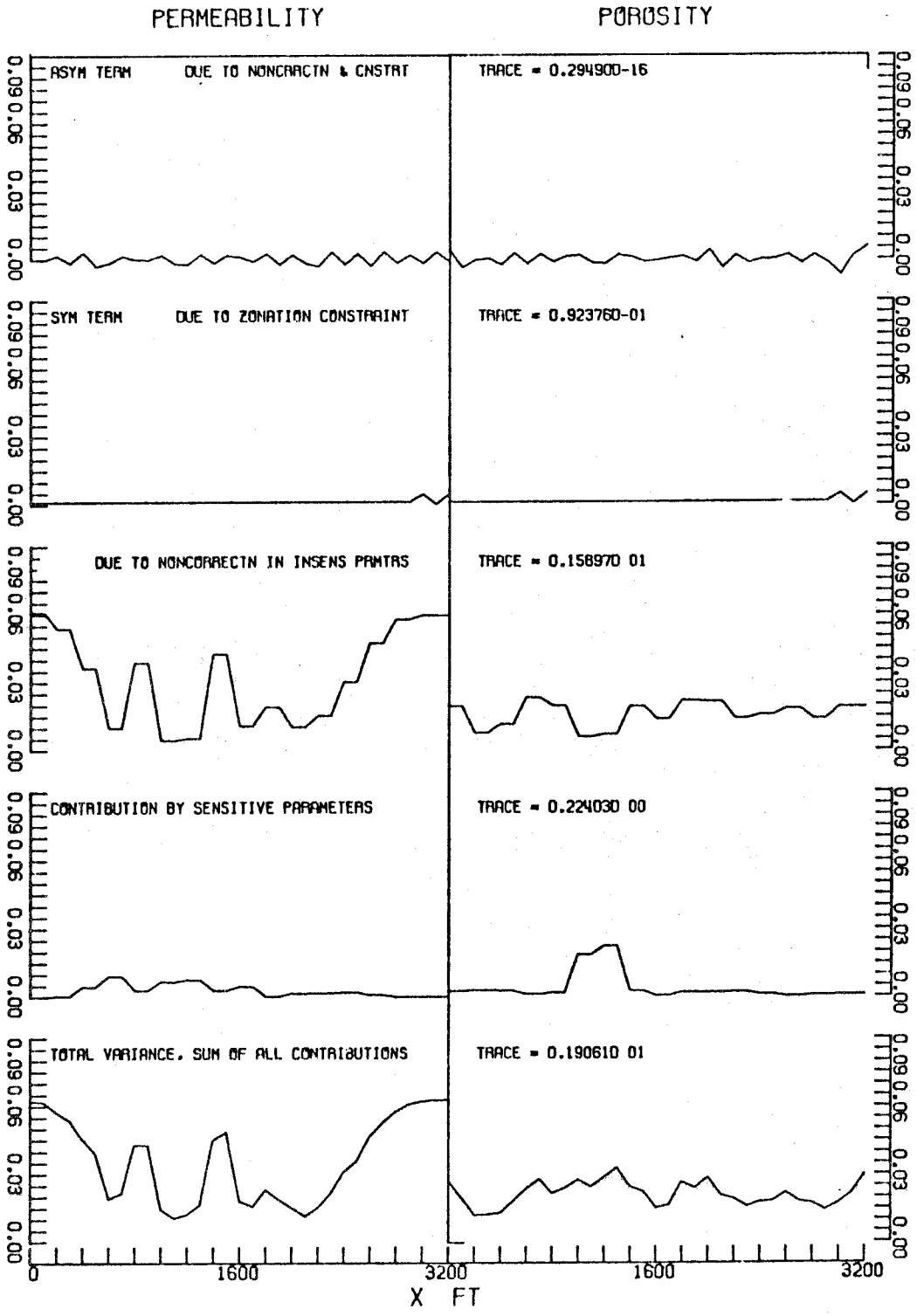
The total variance for k_i is low in the regions close to the production and observation sites because of high sensivity of the observations with respect to those permeabilities. Note that for this case with very few and large zones, the error due to the zonation constraint dominates the rest of the contributions.

Figures (3.2.7) and (3.2.8) contain similar plots for the



CONTRIBUTION BY VARIOUS TERMS TO VARIANCE
ZONATION APPROACH. NZ=8

Figure 3.2.7



CONTRIBUTION BY VARIOUS TERMS TO VARIANCE
ZONATION APPROACH, NZ=16

Figure 3.2.8

cases of 8 and 16 zones respectively. The general features are similar to those of the four-zone case detailed above. The major differences are as follows. The variance distributions for NZ = 4 are dominated by the error due to the hard zonation constraint; as NZ increases this contribution diminishes. At the same time, the error due to the noncorrection in $\overline{\underset{\sim}{V}}_0$ increases, because the dimensionality of the zonation parameter space (= 2NZ) increases while the number of high sensitivity parameters (N_v) remains almost unaltered. For the 16-zone case the error due to the hard constraint is uniform over most of the reservoir, because each zone in this case includes only two grid points which are equidistant from the zone boundaries. The zone closest to the boundary at x = 3200 ft consists of three grid points and hence shows some variability of this error.

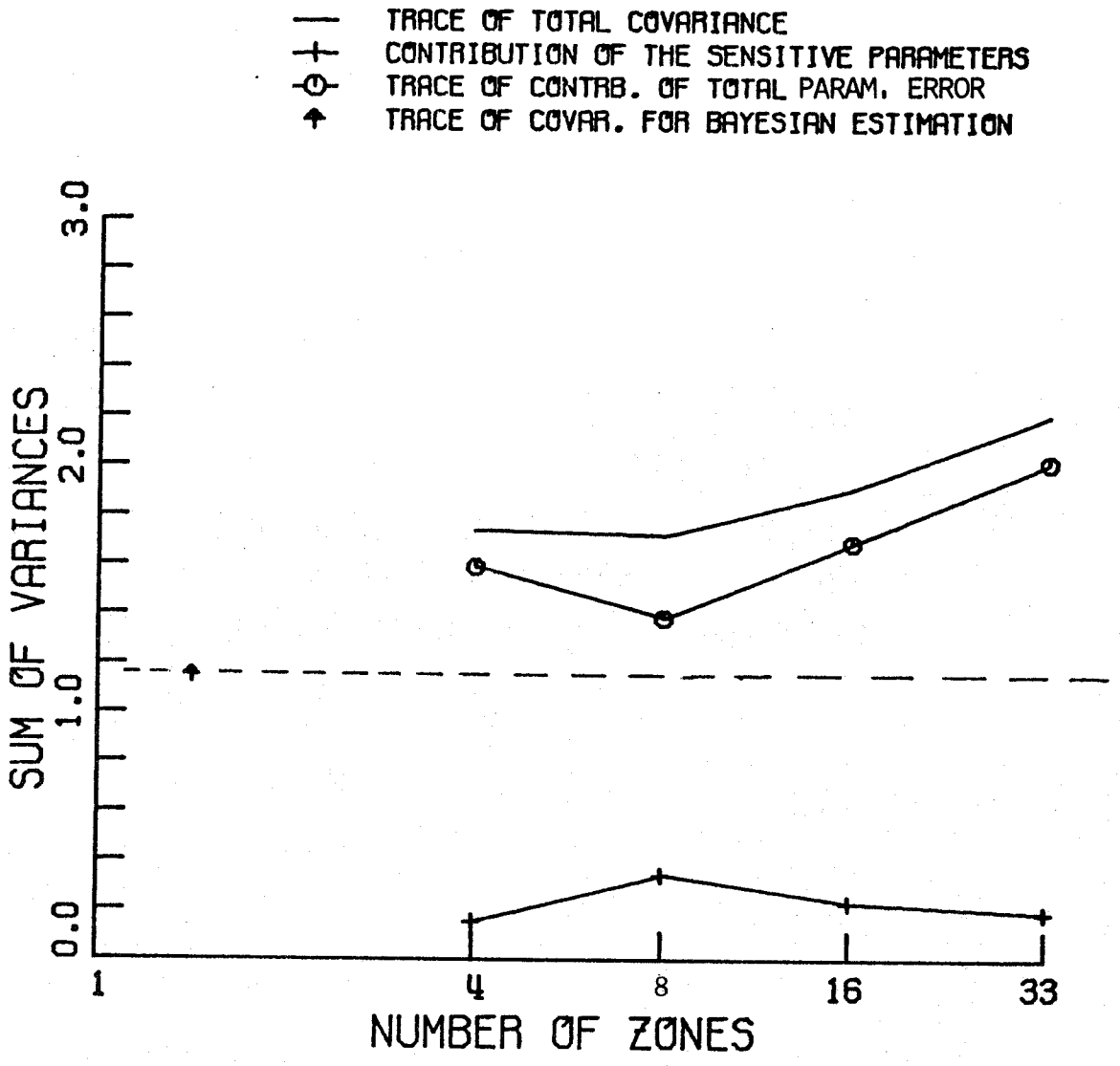
As the number of zones is increased from 8 to 16, the zone size becomes small enough that the kinks in the distribution of the sensitivities of π_i are replicated in that for the zonation parameters. This results in the rapid spatial variations in the error due to noncorrection and consequently in the total variance distribution in figure (3.2.8).

3.2.4 Comparison of Covariances for Different Parameterizations

As discussed at the beginning of this section, we shall compare only the traces of the different matrices. We note that any norm of the matrices would also suffice for this purpose, but the trace is easier to compute. Moreover, the trace of a covariance matrix is an important quantity encompassing the information about

the quality of the estimates of all the parameters; in particular, it is the sum of the variances for the individual parameters (which are all non-negative).

Figure (3.2.9) shows how the trace of $\underset{\sim}{P}_{\pi}$ varies with NZ in the zonation approach. (The case NZ = 33 is included as a special case of zonation.) The values in this plot correspond to the minima in the trade-off curves of $tr(\underset{\sim}{P}_{\pi})$ versus N_v. The figure also shows the contributions by the total parameterization error (which includes all the terms except the first in (3.1.54)) and the uncertainty term involving the sensitivities. The total parameterization error passes through a minimum at NZ = 8; whereas the uncertainty contribution has a maximum there. The parameterization error declines from NZ = 4 to NZ = 8 primarily because of the reduction in the error due to the zonation constraint; the rise from NZ = 8 to NZ = 33 is due to the increase in the error due to noncorrection accompanying the increase in the dimensionality of $\underset{\sim}{\overline{V}}_0$ (= $2NZ-N_v$). The minimum in the total trace is significant. It implies that if the estimation were carried out "optimally" (so as to achieve the corrections commensurate with the appropriate "optimal" value of N_v), the zonation with eight zones would yield the most accurate estimates. Thus, for the given set of zonations, NZ = 8 is the "best" parameterization. We shall compare these predictions with the results detailed in tables (1.6.7-10) of simulations carried out using property distributions for four realizations. The mean values of $J_k + J_\phi$ for this sample of 4 realizations are 8.727, 8.293, 7.863 and 8.795 respectively for NZ = 4, 8, 16 and 33. These



DEPENDENCE OF THE SUM OF ESTIMATE VARIANCES ON PARAMETERIZATION LINEARIZATION ABOUT UNIFORM K,PHI

Figure 3.2.9

results are in a qualitative agreement with our predictions in that a minimum in the total error does occur at an intermediate value of NZ. The minima in two cases do not coincide; it may be due to one or several of the following reasons: (i) The errors J_k and J_ϕ (see table (1.5.1)) are different measures of error than the variances of the estimate errors, (ii) The effective values of N_v in actual simulations may be different than those used for the predictions in this section, (iii) The sample size of four realizations may be too small yielding a large statistical scatter in the mean values of $(J_k + J_\phi)$. (A more extensive simulation using many additional realizations is necessary before reliable estimates of the total errors can be obtained by averaging; however, it would be very expensive.)

Figure (3.2.9) also shows the trace of the covariance for the Bayesian estimation. The trace of $\underset{\sim}{P}_\pi$ is smaller for Bayesian estimation than for any of the zonations; this is expected, because the Bayesian estimation incorporates prior information (in terms of prior mean and $\underset{\sim}{P}_0$) in the estimation process, whereas no such information is exploited by the zonation approach. Thus the Bayesian approach is expected to yield more accurate estimates. This prediction is again qualitatively verified by the results of simulations reported in Chapter 1; the average value of $(J_k + J_\phi)$ for the four realizations is 7.240 for Bayesian estimation, which is lower than that for any of the zonations.

3.3 Verification of the Predicted Variances

The procedure proposed in section 3.1 for determination of the covariances of the property estimates was tested using the results of simulations described in Chapter 1. In this section we present a comparison of the predicted variances (the diagonal elements of the covariance matrix) with the actual estimation errors in the simulations averaged over a sample of four realizations. Admittedly this sample is too small to lead to conclusive results; but it may serve as a qualitative indication of the accuracy of the predicted covariances. The comparison is carried out for all the different parameterizations for which the predicted covariances were presented in the preceding section; the results for the individual parameterization are presented in the same order as before.

We recall that the conditions of the simulations were identical with those used for the predictions. Thus there is no modelling error, or uncertainty about the reservoir geometry, boundary conditions, production rate histories and the observation instants. In the field problems, these factors will be present and our predictive capabilities will, in general, deteriorate. The observation errors are zero-mean, independent Gaussian random variables with a uniform variance of 1 psi^2 ($\underset{\sim}{\Sigma} = \underset{\sim}{I}$).

The minimization was carried out using Marquardt's method, with $\mu = 100$ for all the cases except for the Bayesian estimation in which $\mu = 50$ was used. However, we recall from Chapter 1 that, as long as μ is not changed significantly (say, by one or more orders of magnitude), the results of Bayesian estimation do not suffer any

significant alterations. We recall from the analysis of Chapter 2 that Marquardt's method, in minimization without Bayesian penalty term, yields the largest correction along the $\underset{\sim}{v}$-eigenvector of $\underset{\sim}{A}$ with singular value closest to $\sqrt{\mu}$. Thus the use of $\mu = 100$ in the zonation approaches (NZ = 4, 8, 16, 33) will result in insignificant corrections along the v-vectors with $\lambda_i = O(1)$ or smaller. From figure (2.5.1) it is evident that the number of singular values larger than unity is 10; this implies that for NZ = 33 the dimension of the correction subspace - the effective value of N_v - for the simulations is approximately 10. Furthermore, in the singular value decomposition of $(\underset{\sim}{A}\underset{\sim}{G}_1)$ for NZ = 16, 8 and 4 yielded the result that the number of singular values significantly larger than unity is 9, 9 and 7 respectively. These are precisely the "optimal" values of N_v for these cases observed in the previous section. This indicates that $\mu = 100$ is an "optimal" choice for NZ = 4, 8, 16 and 33; it is expected to yield estimates with the smallest trace of total covariance. Hence we are justified in comparing the results of these simulations with the predictions based on the values of N_v which lead to the smallest trace of the covariance matrix in the respective cases.

We emphasize that the foregoing consideration indicates that an optimal value of μ can indeed be chosen for zonation from an analysis utilizing the prior covariance. We also note that the gradient minimization algorithms do not lend themselves to such analysis of the effective dimensionality of the correction subspace (N_v). Thus, accurate prediction of the covariances is not possible for them; from the considerations of section 2.3, we conclude that the effective

value of N_v for the gradient algorithms are somewhat smaller than those which yield minimal $\mathrm{tr}\{\underset{\sim}{P}_\pi\}$.

The root-mean-square (abbreviated as "rms" in the sequel) estimation error for each property at each grid point was computed for the sample of four realizations, and compared with the square root of the predicted variance (the standard deviations). The two quantities are plotted together for the sake of comparison. Figures (3.3.1-3.3.5) show such plots for all the different parameterizations considered. As done before, the association of each element of $\underset{\sim}{\pi}$ with a grid point in the reservoir is utilized in the identification of the abscissae. We conclude from the comparison that on the whole the predicted standard deviations of the estimate errors and the rms values for the actual simulations are in good agreement in the cases of zonation (NZ = 4, 8, 16) and in estimation without constraints (NZ = 33). The Bayesian estimates have somewhat higher actual estimate errors compared with their predicted values. This may be partly due to the approximations made in arriving at the expression for the predicted covariance and partly due to the limited sample size.

We conclude that when an accurate prior covariance is known, the formulation in section 3.1 yields fairly accurate predictions of the variances for the property estimates obtained via the different approaches to parameterization and estimation considered here.

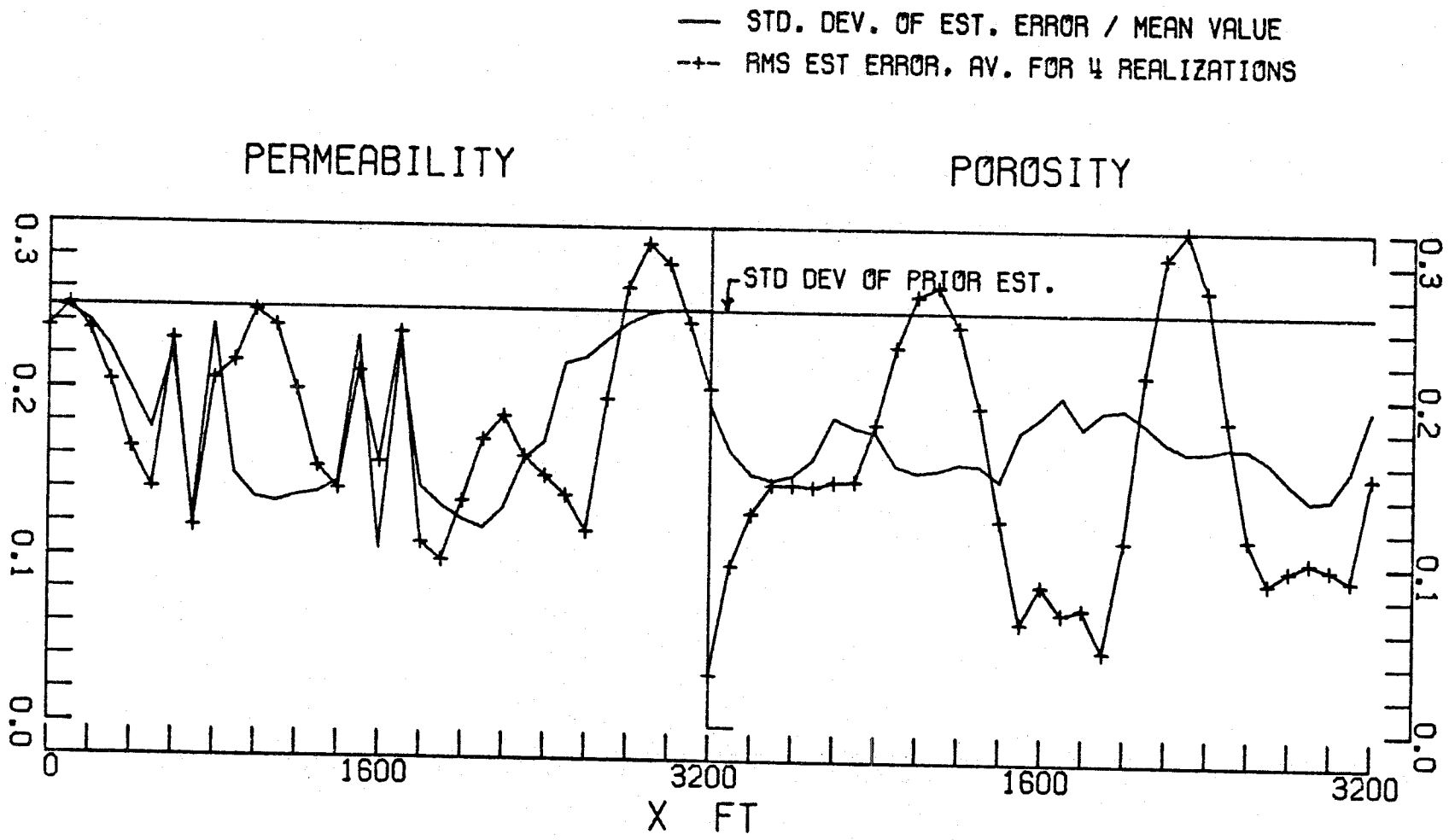


Figure 3.3.1 Verification of Predicted Variances

STD DEV OF EST. ERRORS, NO CNSTRNTS (NZ=33)
LINEARIZATION ABOUT UNIFORM K-PHI

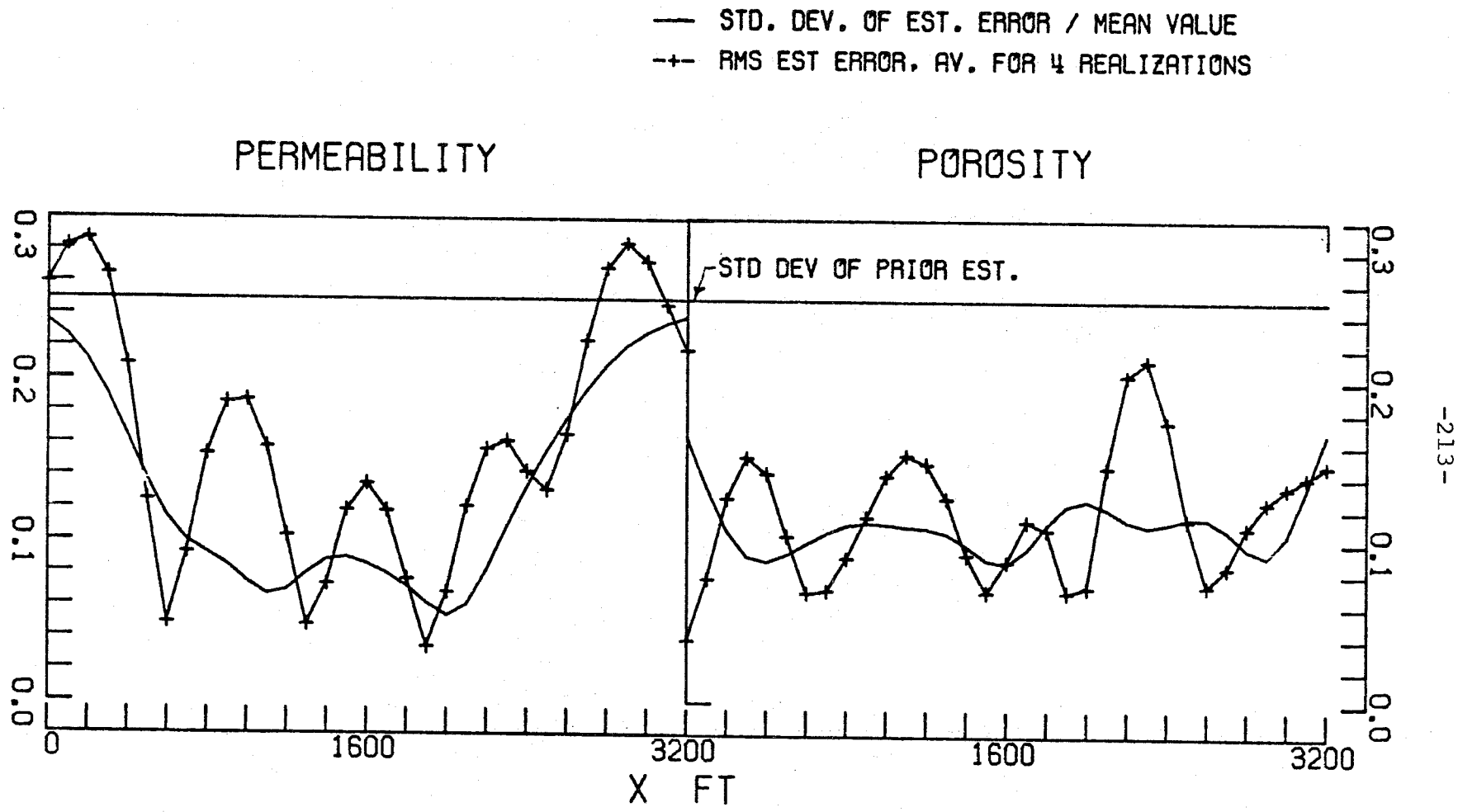


Figure 3.3.2 Verification of Predicted Variances

STD DEV OF EST. ERRORS. BAYESIAN APPROACH
LINEARIZATION ABOUT UNIFORM K-PHI

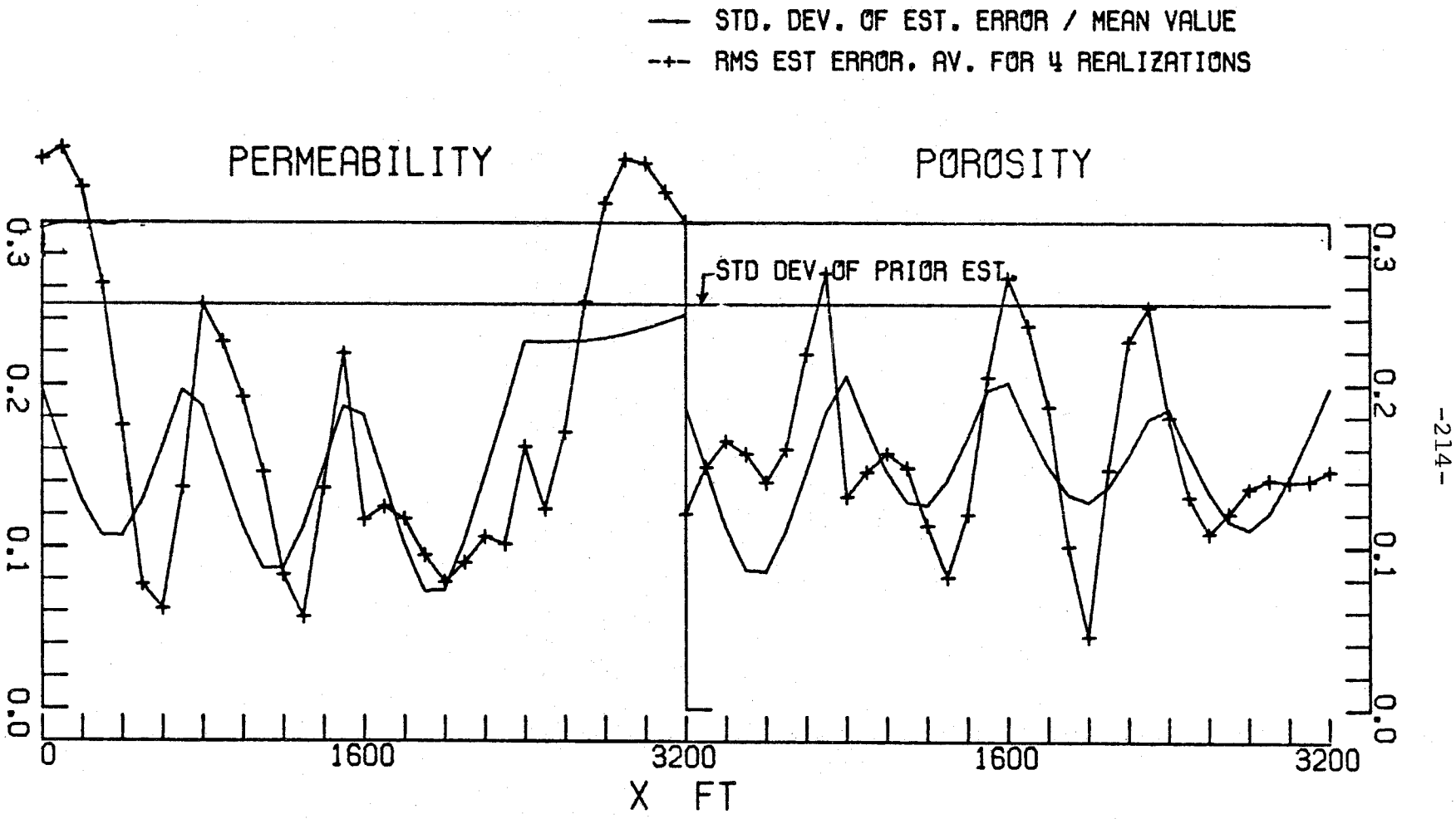


Figure 3.3.3 Verification of Predicted Variances

STD DEV OF EST. ERRORS, ZONATION NZ=4
LINEARIZATION ABOUT UNIFORM K-PHI

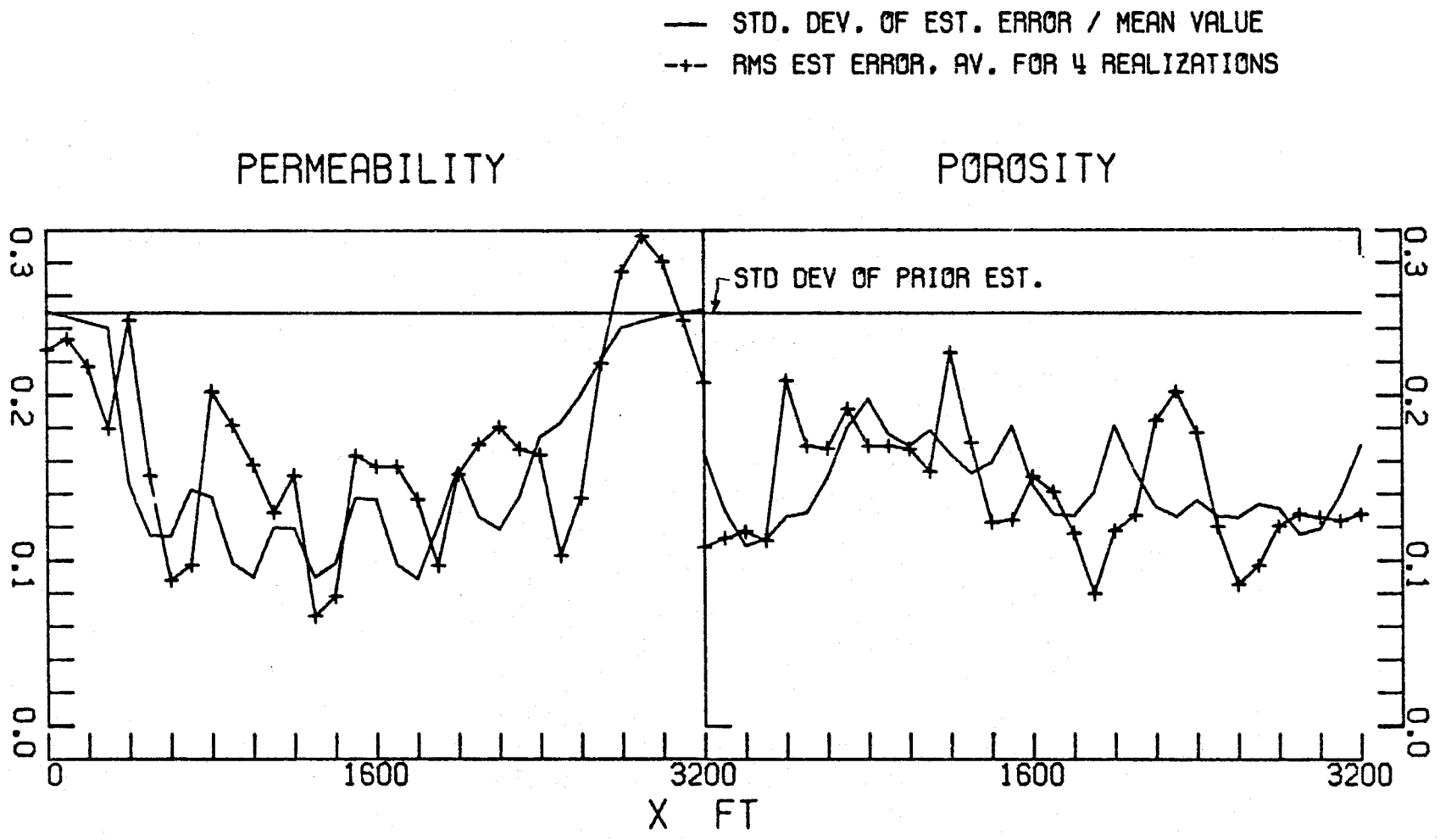


Figure 3.3.4 Verification of Predicted Variances

STD DEV OF EST. ERRORS, ZONATION NZ=8
LINEARIZATION ABOUT UNIFORM K-PHI

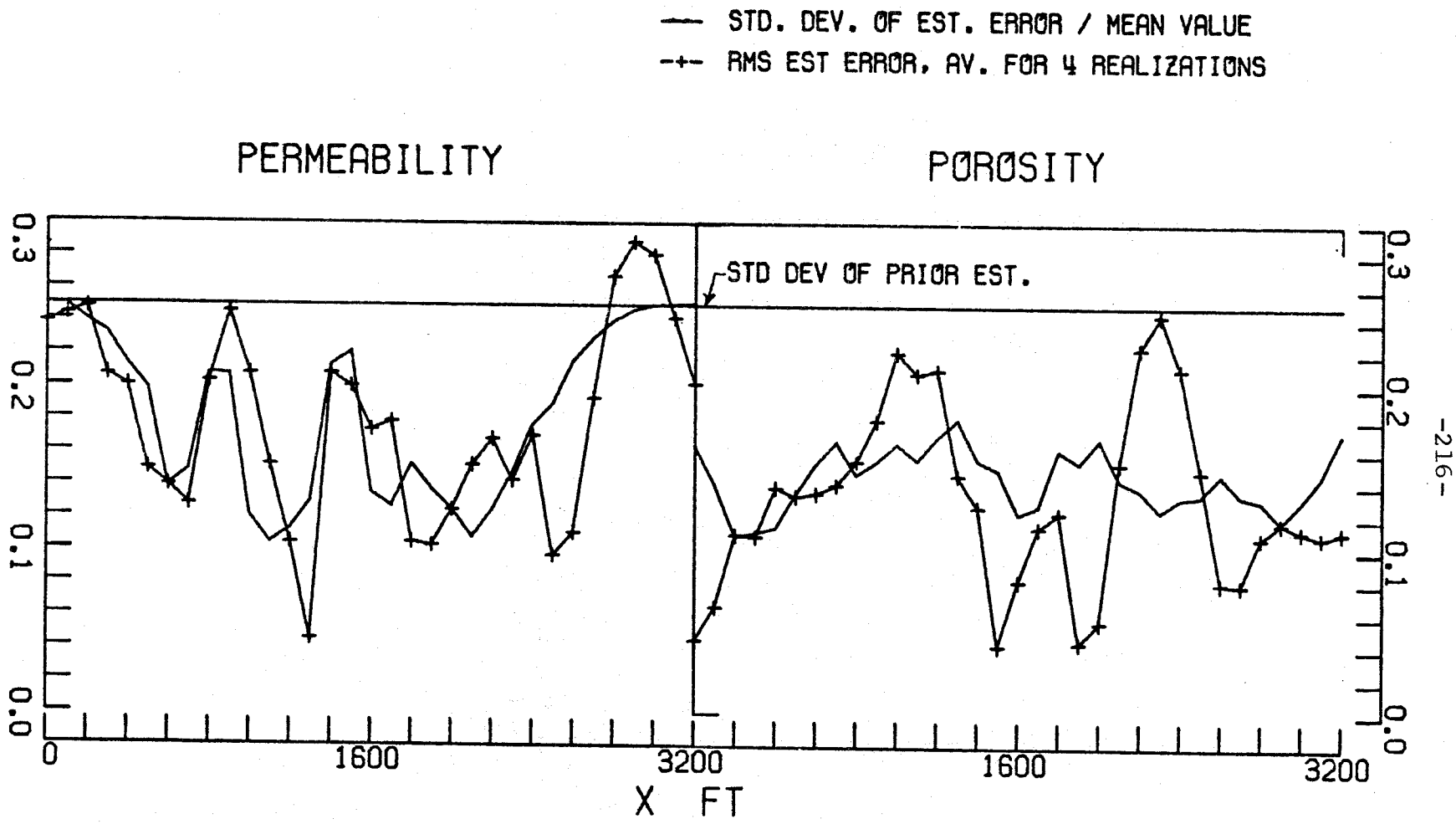


Figure 3.3.5 Verification of Predicted Variances

STD DEV OF EST. ERRORS, ZONATION NZ=16
LINEARIZATION ABOUT UNIFORM K-PHI

3.4 Effect of Prior Covariance on A Posteriori Covariance

In this section we shall study the effect of change in the specified prior covariance on the a posteriori covariance predicted by the results of section 3.1. In addition to indicating the influence of $\underset{\sim}{P}_0$ on $\underset{\sim}{P}_\pi$, such an investigation is important in predicting the effect of an error made while specifying the prior covariance $\underset{\sim}{P}_0$ on the results.

Parameters of Prior Covariance. As detailed in Chapter 1, we can specify the prior covariance matrix for the reservoir parameters based on qualitative or quantitative knowledge about the degree of smoothness of their distributions and the degree of their variability from the specified means. Such a construction hinges on modelling the parameter distributions as a realization of a homogeneous random process. Then the prior covariance matrix for the porosity and permeability distributions can be specified by four scalar parameters, namely s, the parameter associated with the length of the autocorrelation, the uniform variances, σ_k^2 and σ_ϕ^2, characterizing the fluctuations of the respective parameters from their mean values, and ρ, the cross correlation parameter. The parameter s reflects the degree of smoothness of the spatial distributions of k and ϕ. On the other hand, σ_k and σ_ϕ reflect the degree of confidence we have in the specified mean values and also the degree to which the actual local values of the parameters are expected to deviate from the respective means.

We shall investigate the effects of variations in the prior covariance within the framework of this construction. Consequently,

we shall first vary s while keeping the rest of the parameters in $\underset{\sim}{P}_0$ fixed. Then we shall study the effect of the erroneous specification σ_ϕ and σ_k; as the two distributions are correlated, it is reasonable to vary them simultaneously and in the same proportion. Later we shall study the effect of varying ρ.

3.4.1 Change in Specification of s

The covariances for the different parameterizations treated in section 3.2 were recomputed using s = 2.5 and 7.5, while keeping the rest of the parameters in $\underset{\sim}{P}_0$ unchanged as follows: $\sigma_\phi/\overline{\phi} = \sigma_k/\overline{k} = 0.25$, $\overline{\phi} = 0.2$, $\overline{k} = 5.0$ md, $\rho = 0.5$. In the following description of the numerical results, the values of σ_k^2 and σ_ϕ^2 are indicated through a single parameter β, in a manner similar to that of Chapter 1; a variation in these parameters from their nominal values noted above by a factor β will correspond to replacing the prior covariance matrix $\underset{\sim}{P}_0$ by $\beta\underset{\sim}{P}_0$. Denoting the nominal values of the variances quoted above by $\sigma_{\phi 0}^2$ and σ_{k0}^2, the parameter β is given by

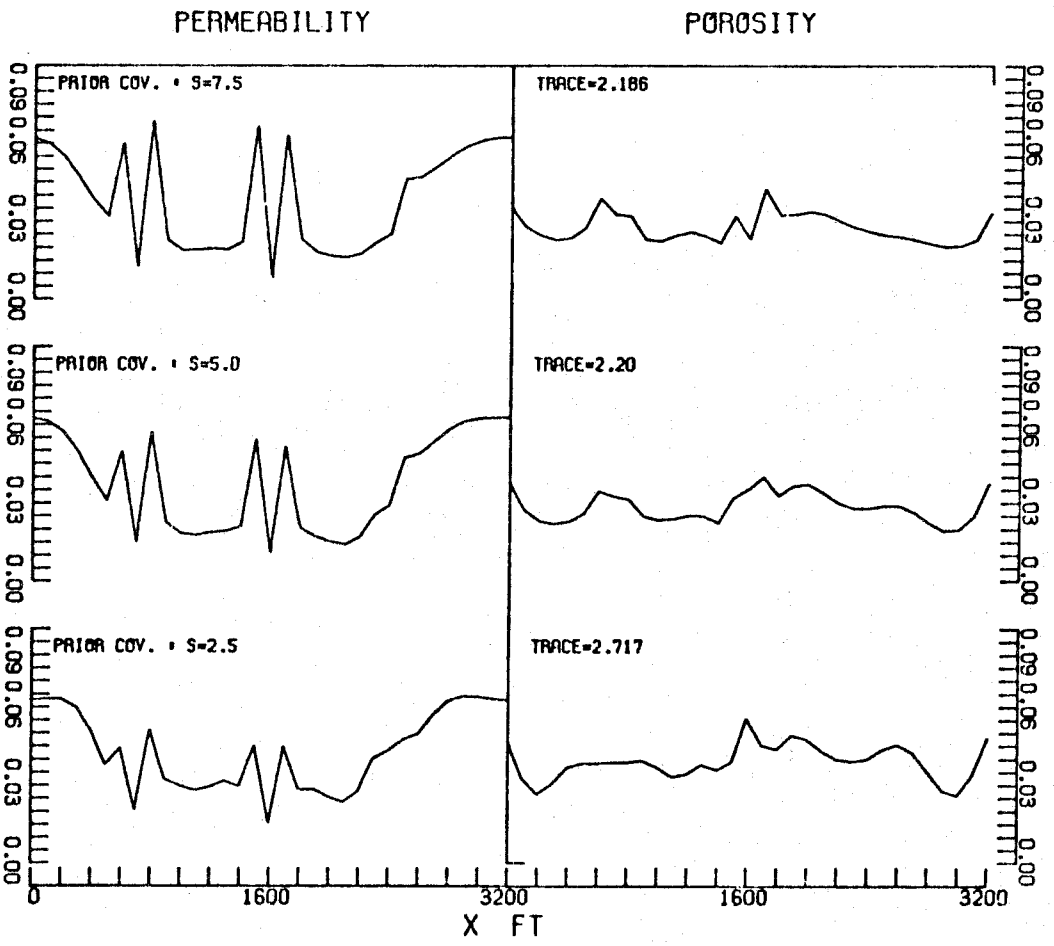
$$\beta = \sigma_k^2/\sigma_{k0}^2 = \sigma_\phi^2/\sigma_0^2 \qquad (3.4.1)$$

The sensitivity matrix $\underset{\sim}{A}$ for the uniform distributions (equal to the respective mean values) was used for all of the following computation. The resulting variances were compared with those obtained for s = 5.0 in section 3.2. The a posteriori variances for these three values of s for each of the five different approaches to parameterization are plotted in figures (3.4.1-3.4.5). The trace of the a posteriori covariance matrix and the relevant parameters of $\underset{\sim}{P}_0$ are also indicated in the figures.

3.4.1.1 Estimation without Constraints

Figure (3.4.1) shows the plots of the variances for estimation done without constraints, i.e. NZ = 33. In this case, trace ($\underset{\sim}{P}_{\pi}$) is minimum for N_v = 9 for all the three values of s. As the value of s is increased, the trace decreases, implying a better overall estimation for a higher value of s. As discussed at the beginning of this chapter, the a posteriori covariance depends on two factors: the prior covariance and the information content in the observations. For a given set of observations, and given $\underset{\sim}{A}$ and N_v, the changes in the posterior covariance are entirely due to those in the prior covariance. Then the above-mentioned trend in trace ($\underset{\sim}{P}_{\pi}$) can be explained from the fact that a larger value of s implies a higher information content in the prior estimates; this assertion follows from the observation made in Chapter 1 that for a higher value of s, the matrix $\underset{\sim}{P}_0$ has a larger number of eigenvalues close to zero. It is well known that a reduction by one in the rank of a covariance matrix of a random vector is tantamount to the existence of an additional deterministic relation between its elements. Thus a larger value of s corresponds to fewer basis vectors (the eigenvectors of $\underset{\sim}{P}_0$) in the parameter space along which the components of the prior estimate vector have significantly large uncertainty.

Now we shall examine the details of the variance distribution for the three values of s. The reduction in the trace with increase in s is mainly brought about by the reduction in the variances of the permeability estimates in the regions between the three observation sites and by an almost uniform reduction in the variances of the



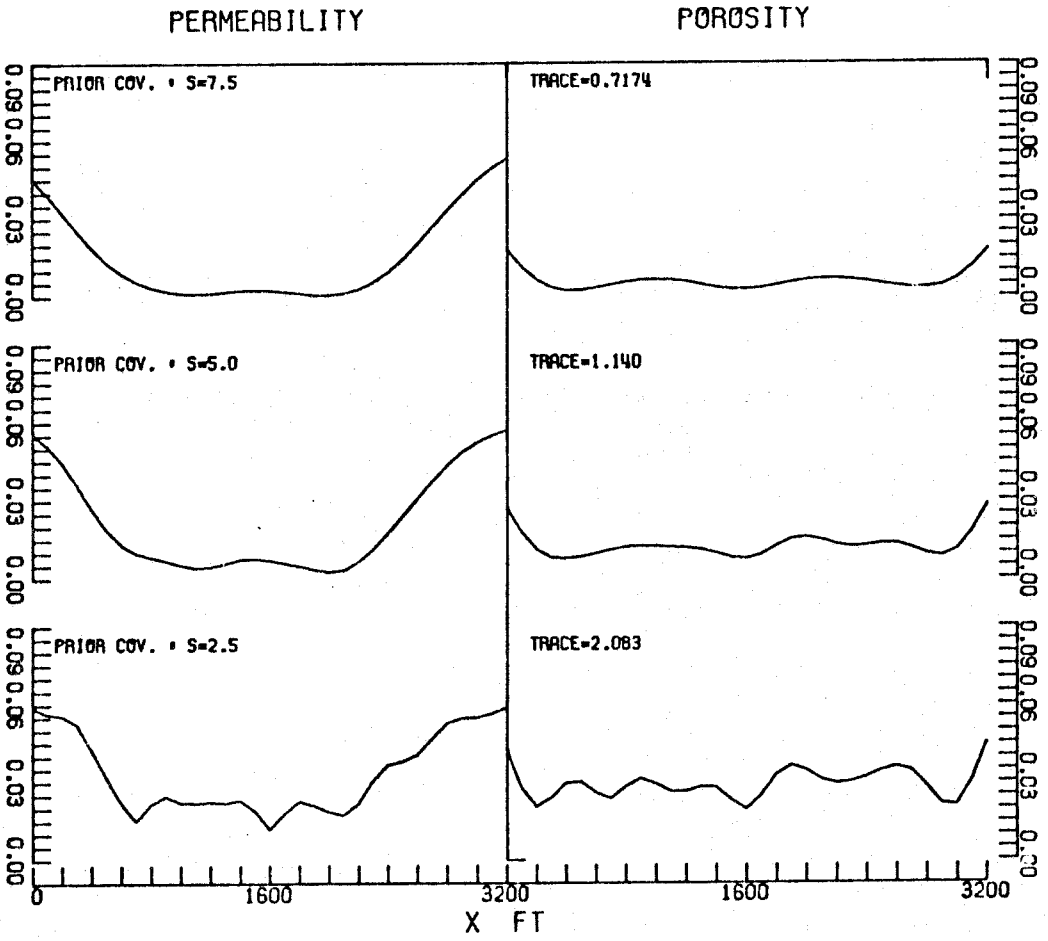
EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ESTMN WITHOUT CONSTRAINTS NZ=33 BETA=1

Figure 3.4.1

porosity estimates all over the reservoir. In addition, as s is increased, the distribution of the variances becomes smoother in general. This is because of the contribution by the error due to noncorrection in the expressions (3.1.30); it is obtained by pre- and post-multiplication of $\underset{\sim}{P}_0$ by the matrix $\underset{\sim}{V}_0$. The larger value of s results in the elements of $\underset{\sim}{P}_0$ dropping off to zero less rapidly as we move away from the diagonal; this has a greater averaging effect in the error term and consequently the results are smoother on the average. An important exception to this trend is, the kinks occurring in the variance distribution of the permeability estimates in the vicinity of the observation sites which contain producing wells become larger as s increases. It appears that the effect of increase in s is to increase the steepness and amplitude of the large and sudden changes in the variance distribution while smoothing out small variations.

3.4.1.2 Bayesian Estimation

Figure (3.4.2) shows the effect of s for the Bayesian estimation. The trace of the a posteriori covariance matrix decreases considerably as s is increased. This is due to the increased information content in $\underset{\sim}{P}_0$ resulting from larger s, as discussed in the foregoing paragraph. An important feature of this case is, the distribution of the a posteriori variances becomes smoother as s is increased, and no sharp changes are present for s = 5 and 7.5. For s = 2.5, the averaging effect of $\underset{\sim}{P}_0$ (discussed above) is sufficiently small to allow the variations in the sensitivity to be reflected in the variance distribution. The rapid decrease in the variances with



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
BAYESIAN ESTIMATION BETA=1

Figure 3.4.2

increase in s indicates that an accurate specification of s is important for a reliable determination of the a posteriori covariance for the Bayesian estimation. In particular, too large a value of s will lead to undue confidence in the estimates.

3.4.1.3 Estimation with Hard Constraints

In the study of the effect of s on the a posteriori covariance for the zonation approach, we shall first investigate how the trade-off between the contributions of the uncertainty and the parameterization error is affected. For each zonation, the trace $\{\underset{\sim}{P}_{\pi}\}$ was found to be minimum at approximately the same value of N_v for the different values of s. For the purpose of illustrating the details, we plot the traces of the various contributions to the total variance against N_v for the three values of s for the eight zone case in figures (3.4.3a, b, c). As explained previously, the contribution by the sensitivity term in the expression (3.1.54) remains unaltered when s is changed. The error due to the hard constraint changes significantly with s but is independent of N_v. The contribution by the error due to noncorrection in the $(2NZ-N_v)$ less sensitive directions in the parameter space changes little with s. On the whole, the minimum in the plot of the trace of total a posteriori covariance does not shift appreciably as s is varied.

Figure (3.4.4) shows the results for the four-zone case. (The values of N_v used in the computations, which yield minimal tr $\{\underset{\sim}{P}_{\pi}\}$, are indicated in this and similar figures throughout this section.) In this case, the dominant contribution to the total variance is that of the error due to the zonation constraint; hence the effect of s on

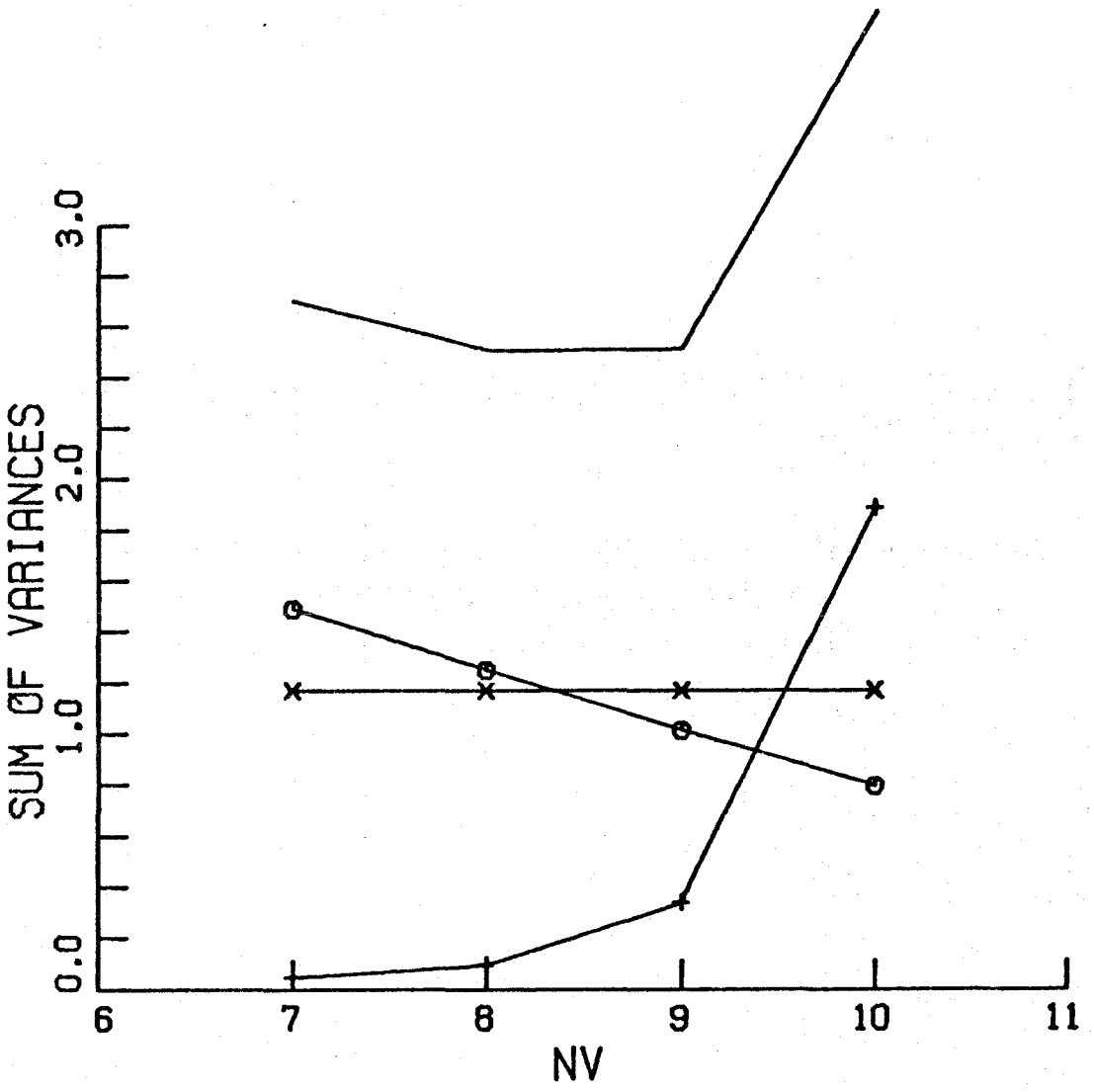


Figure 3.4.3a TRADEOFF BETWEEN UNCERTAINTY AND PARAM. ERROR
EFFECT OF PRIOR COVARIANCE
NZ=8 S=2.5, BETA=1.0

(See Figure 3.4.3b for legend)

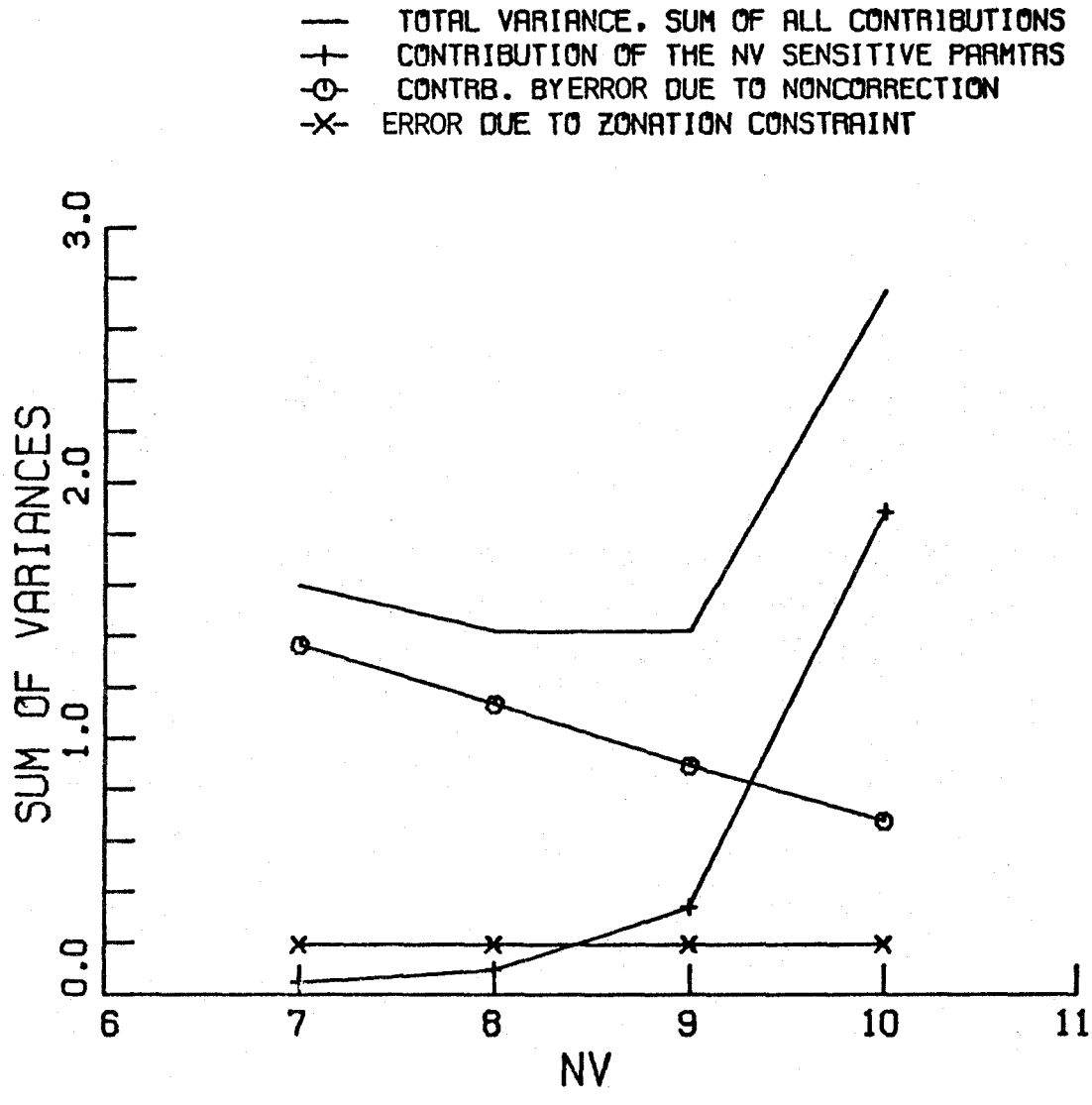
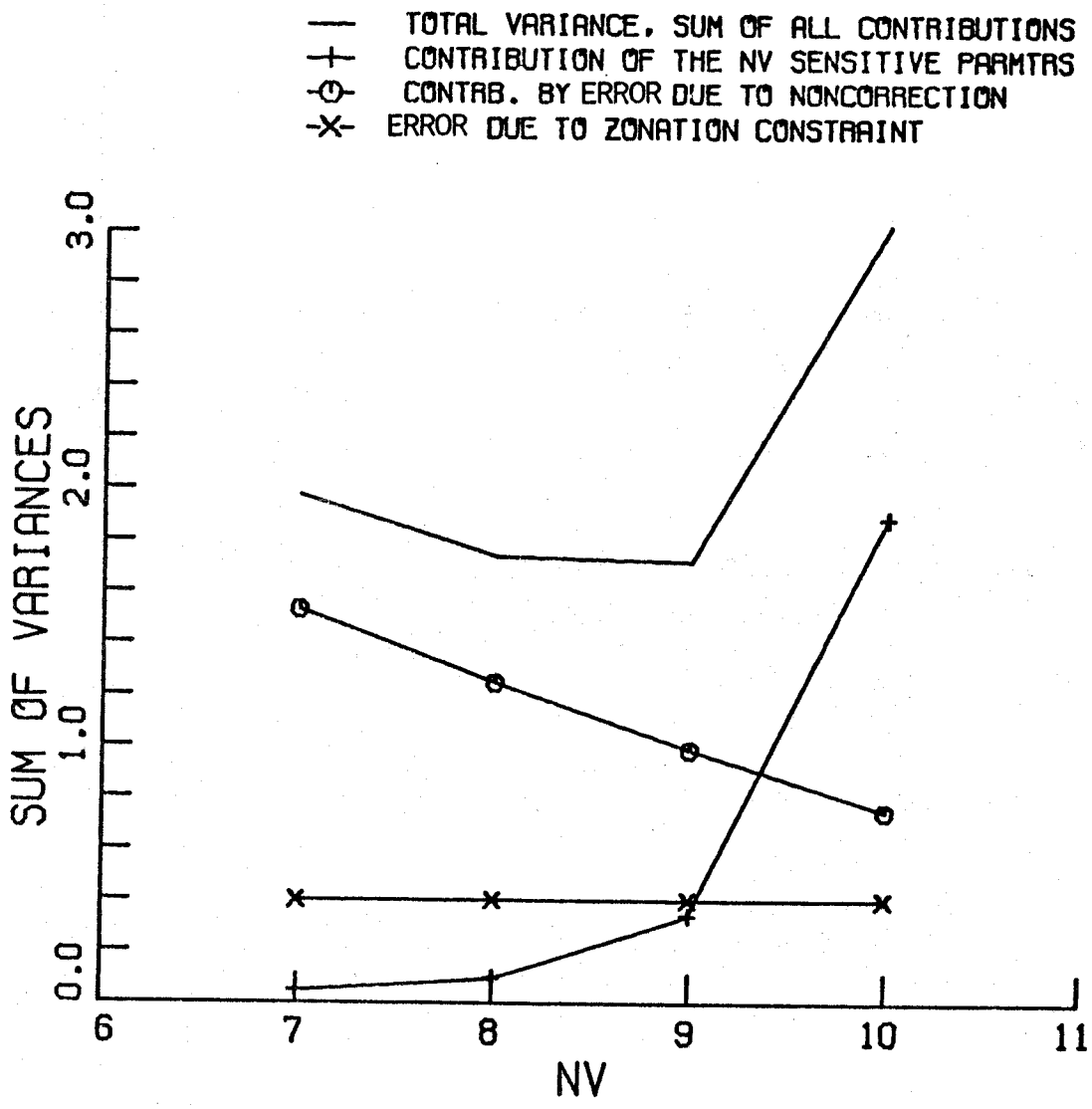
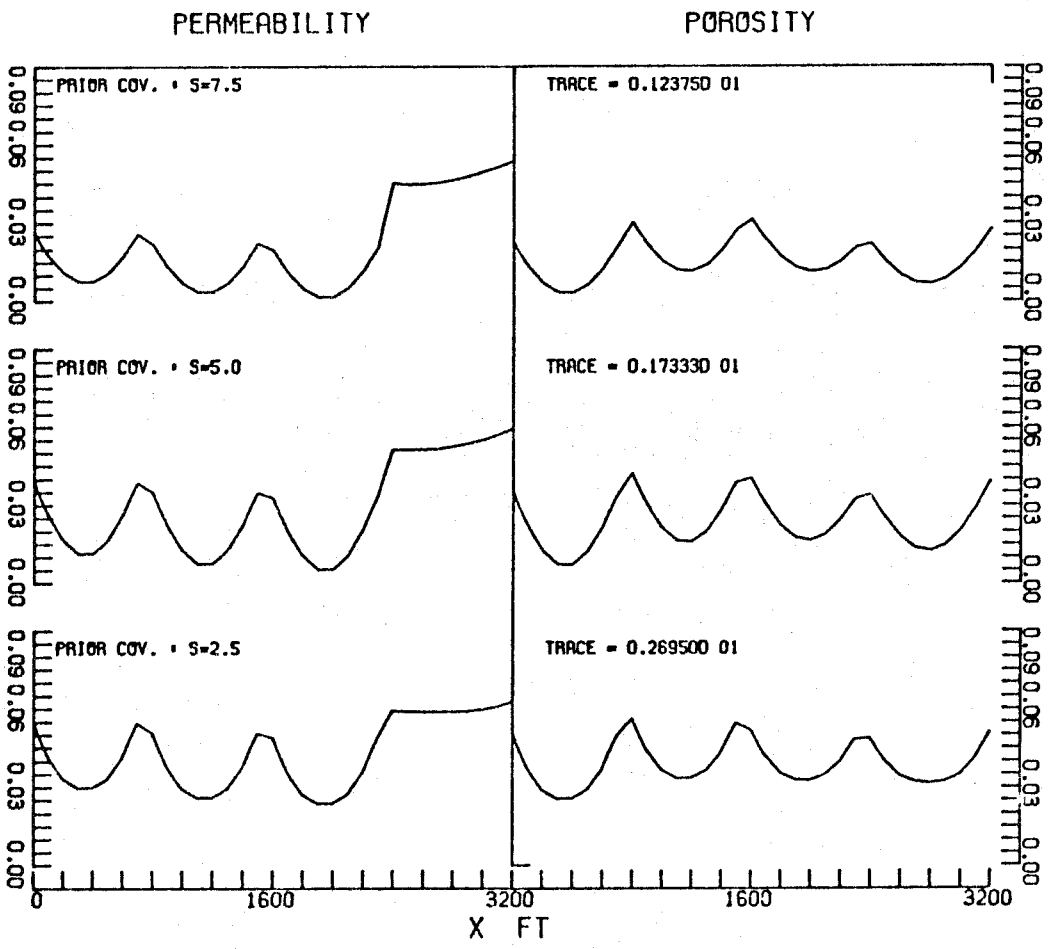


Figure 3.4.3b TRADEOFF BETWEEN UNCERTAINTY AND PARAM. ERROR
EFFECT OF PRIOR COVARIANCE
NZ=8 S=7.5. BETA=1.0



TRADEOFF BETWEEN UNCERTAINTY AND PARAM. ERROR
EFFECT OF PRIOR COVARIANCE
NZ=8 S=5. BETA=1.0

Figure 3.4.3c



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=4 , NV=7 BETA=1

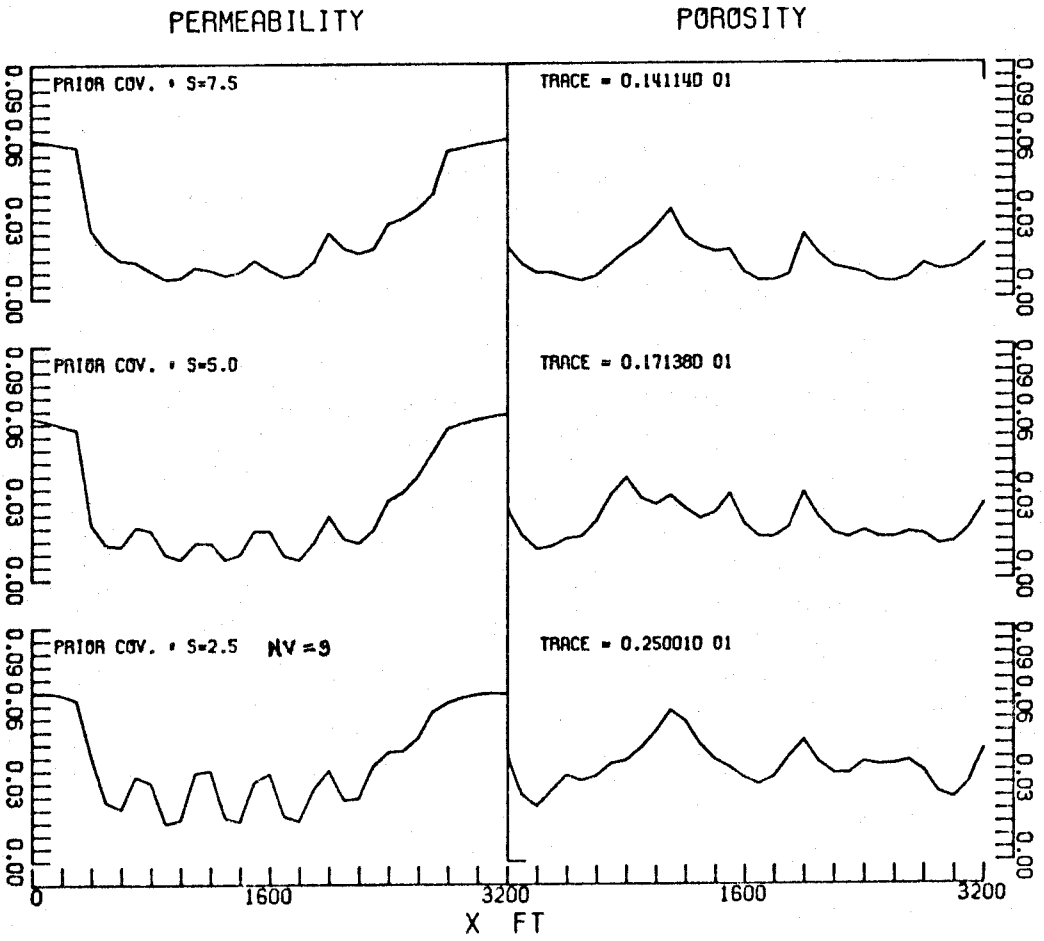
Figure 3.4.4

the total variances can be understood in terms of the variation of this error with s. As s increases, the realizations of the homogeneous random process become less oscillatory, and thus better suited for approximation by zonation; as a result, the error due to constraint decreases considerably as s increases, and with it the trace of the total covariance decreases. The reduction occurs all over the reservoir, but especially near the zone boundaries, because a smoother profile has a smaller average difference between the parameter values at the center and near the boundary of a zone.

Figure (3.4.5) contains the results for the eight-zone case. As discussed above, the error due to the constraint decreases rapidly with increase in s, yielding a smaller trace. In particular, the high level of error near the zone boundaries decreases rapidly yielding a relatively smooth spatial distribution of the variances for s = 7.5.

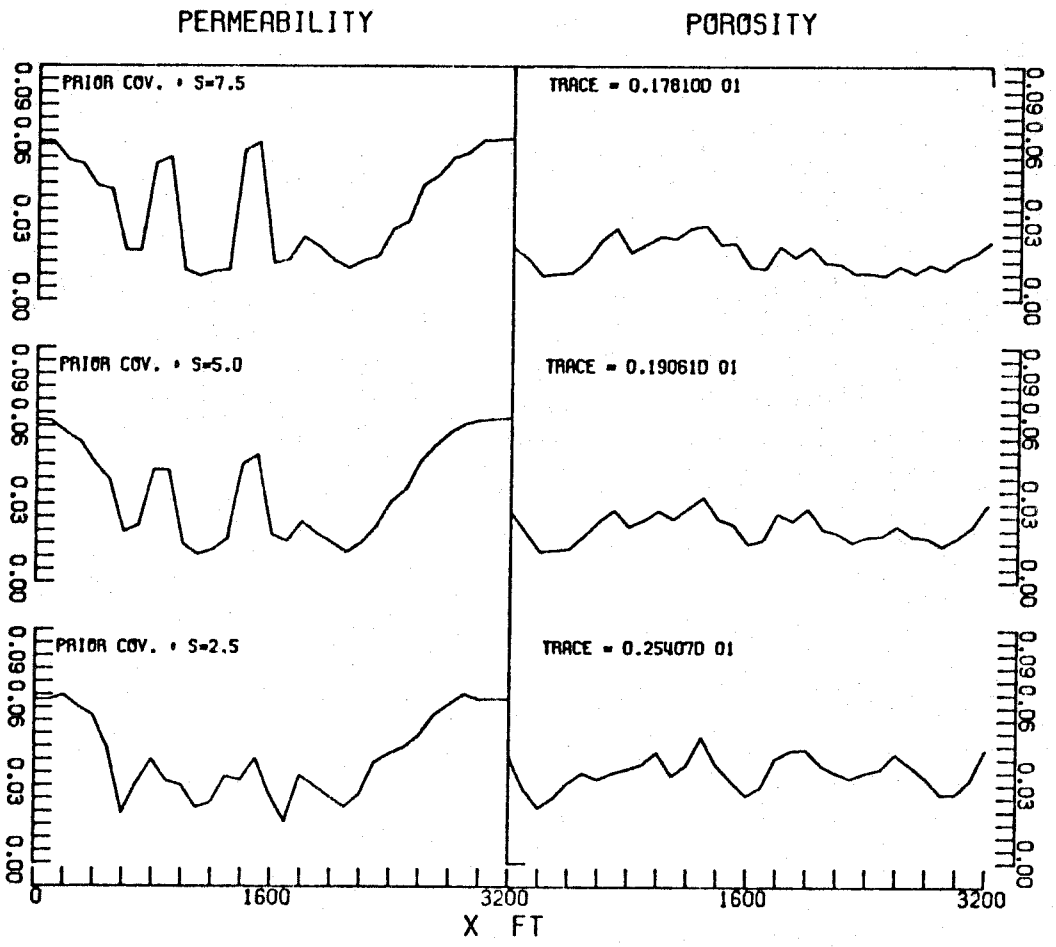
Figure (3.4.6) depicts the results for the 16-zone case. In this case, the trace of the total covariance decreases less dramatically with increasing s; this is because for this case the error due to the zonation constraint is a relatively small contribution to the total covariance. The spatial distributions of the variances become smoother, in general, as s increases but the sudden jumps near the production well sites become sharper. Both these features are very similar to those observed in the case of estimation without constraints (NZ = 33).

In conclusion, we observe that in all the parameterizations studied here, an increase in the parameter s of the prior covariance



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=8 , NV=8 BETA=1

Figure 3.4.5



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=16. NV=9 BETA=1

Figure 3.4.6

leads to smaller variances of the estimates. The largest effects are observed in the cases of Bayesian estimation and zonation with very few zones. The trace of the a posteriori covariance changed at most by a factor of 3 for a 3-fold increase in the value of s. In all the cases studied, a larger value of s leads to a smoother spatial distribution of the variances in most regions of the reservoir. In zonation with many small zones, the variance distribution for the permeability shows sudden changes in the vicinity of the production wells containing observation sites, the kinks becoming larger and sharper with increasing s.

3.4.2 Effect of Change in Prior Variances

An alteration of the specified diagonal elements of the prior covariance matrix has a large influence on the a posteriori covariance. This can be explained from the following. The diagonal elements reflect the degree of confidence we have in the respective prior estimates. Due to the ill-conditioned nature of the problem detailed in Chapter 2, the observations contain almost no information about some of the parameters. Thus the uncertainty in the estimates of these parameters does not decrease due to history matching the observations, and their posterior variances are almost identical with their prior variances. In other words, the a posteriori covariance contains significant error contributions which strongly depend on the prior covariance.

3.4.2.1 Estimation without Constraints

Except in the case of the Bayesian estimation, most of the terms in the expressions for the a posteriori covariance are directly proportional to the prior covariance. This results in large changes in the parameterization error estimates following large variation in the specification of the prior variances. On the other hand, the sensitive term contribution represents the level of information existing in a given set of observations and depends only on the property distributions used for linearization and on the observation policy; in particular, this contribution does not depend on the prior covariance. We recall that the value of the parameter N_v is selected so as to yield a minimum trace of the a posteriori covariance; this involves a trade-off between decreasing parameterization error and increasing uncertainty as N_v is increased. Since the former of the two is significantly altered by changing the prior variances for a given value of s, the optimal value of N_v will also change. This can be simply visualized in the following manner. If the level of prior uncertainty is high, corresponding to a higher value of prior variances, it is advantageous to include the components of the estimate vector about which the observations contain even very little information. On the other hand, if the prior parameter estimates have a relatively small uncertainty, it is disadvantageous to include the less sensitive components of the parameter vector in the correction, as their inclusion will increase the uncertainty of the updated estimates. Thus the optimal value of N_v decreases as the prior variances decrease and vice versa. A quantitative study of this variation

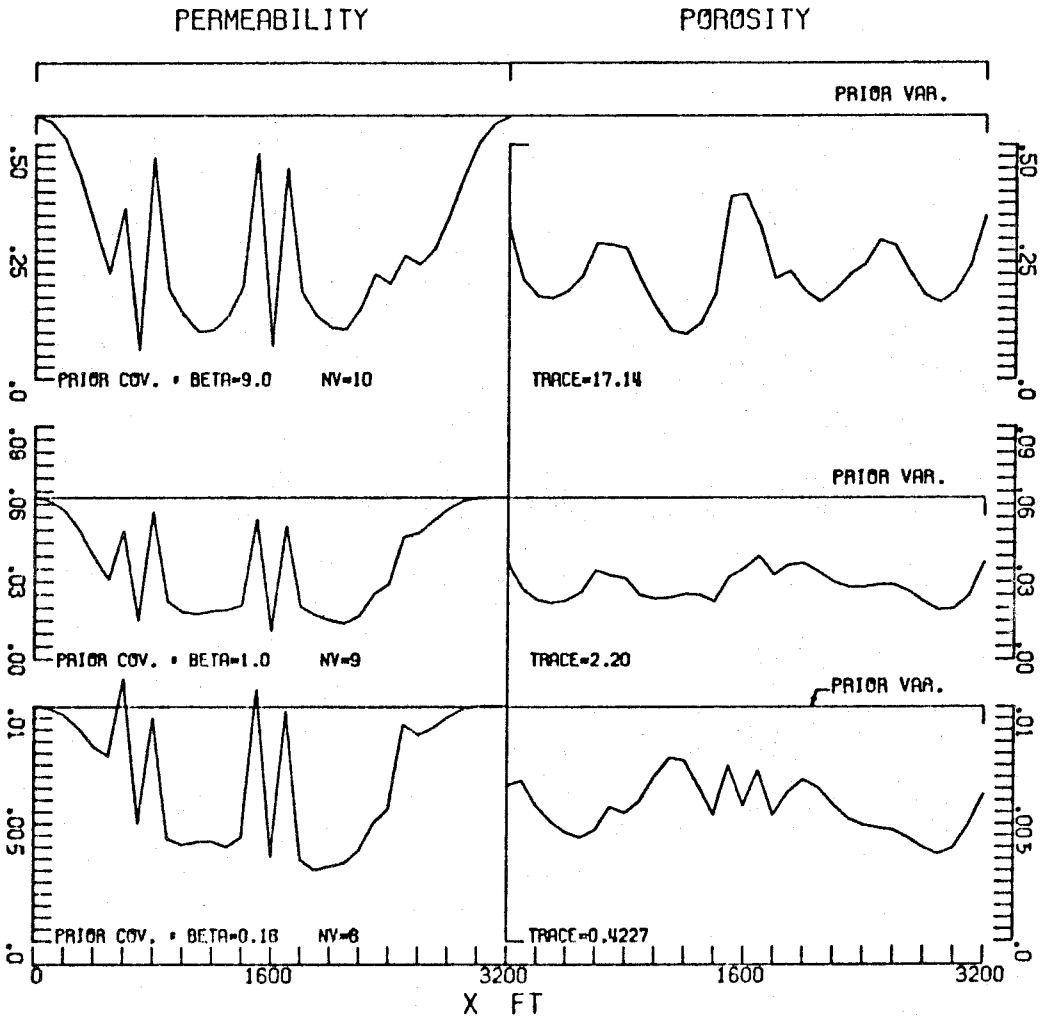
of "optimal" N_v with β is detailed later in relation to the estimation with hard constraints.

The trace of the a posteriori covariance drastically changes as β is increased from 0.16 to 9.0. This is an indication of the degree to which the prior information participates in the determination of the confidence in the a posteriori estimates for this problem. In other words, it is an indication of the ill-determined nature of the parameter estimation problem.

Now we shall examine the details of the spatial distribution of the a posteriori covariances for the different values of β, for the various approaches to the parameterization. Figure (3.4.7) shows this distribution for estimation without constraints, for β = 0.16, 1.0 and 9.0. It also shows the distribution of the prior variances in each case. As noted above, the levels of the variances are altogether different for the different values of β. The details of the distributions are similar for the three cases, because the values of N_v are not significantly different. For a given value of N_v, the various contributions to the parameterization error are simply increased or decreased uniformly by a factor of β.

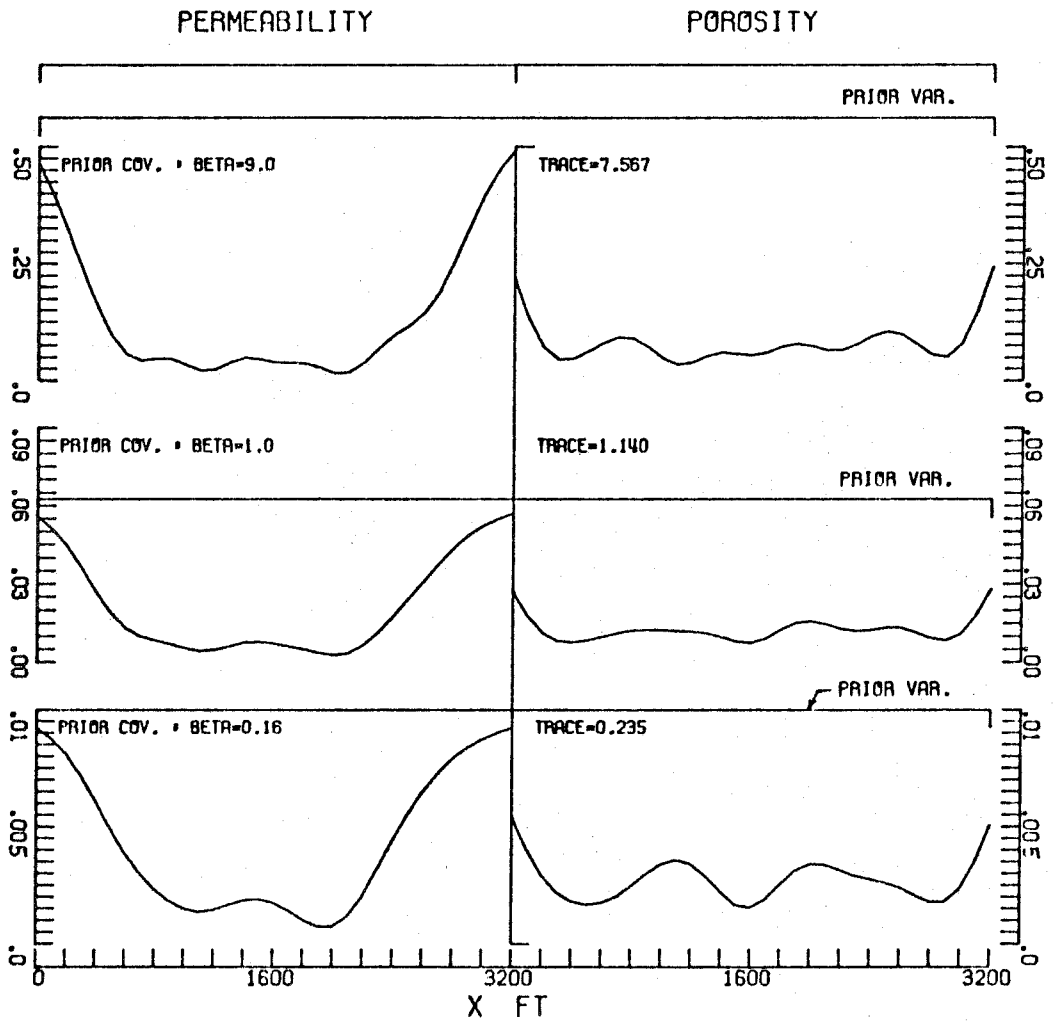
3.4.2.2 Bayesian Estimation

Figure (3.4.8) illustrates the effect of β on the a posteriori variance distribution for the Bayesian estimation. In this case, in addition to altering the levels, the value of β also influences the details of the distribution. This is due to the complicated nonlinear way in which $\underset{\sim}{P}_0$ enters the expression (3.1.45). In this expression,



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ESTMN WITHOUT CONSTRAINTS NZ=33 S=5

Figure 3.4.7



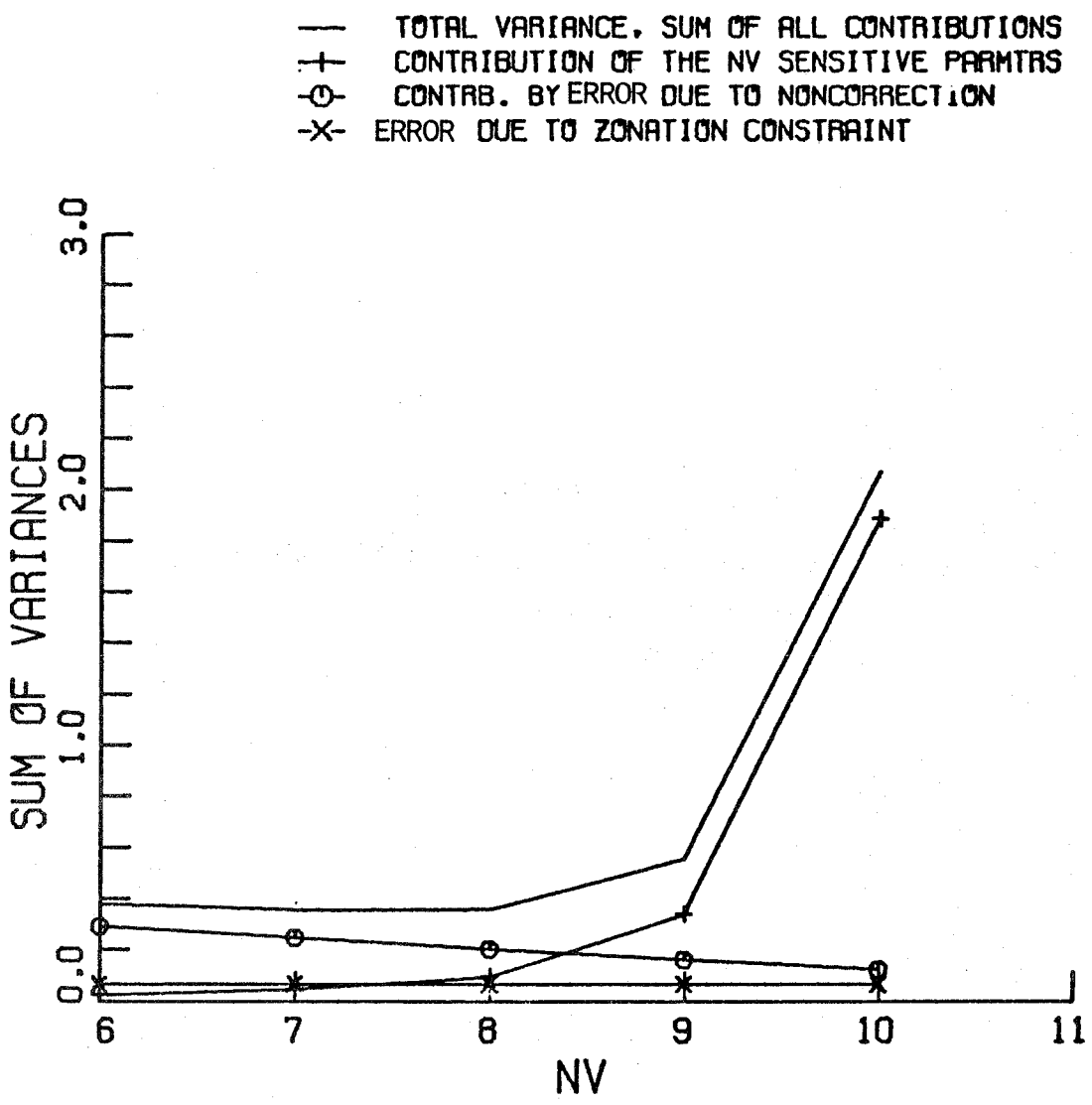
EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
BAYESIAN ESTIMATION S=5

Figure 3.4.8

the a posteriori covariance is obtained by balancing the prior information and the information in the observations. If the prior information is very little (large β), then most of the information is derived from the observations and consequently even small sensitivity of a parameter will have a visible effect. Thus the parameters in a larger portion of the reservoir have substantially smaller variances compared with their prior variances. On the other hand, when the prior estimates are very reliable, corresponding to a small value of β, the information gained from the observations about the low sensitivity parameters has little impact on the a posteriori covariance. As a result, only the regions of reservoir with high parameter sensitivity show appreciable decrease in the variances. As pointed out previously, the variance distributions are relatively smooth for the Bayesian estimation for all values of β.

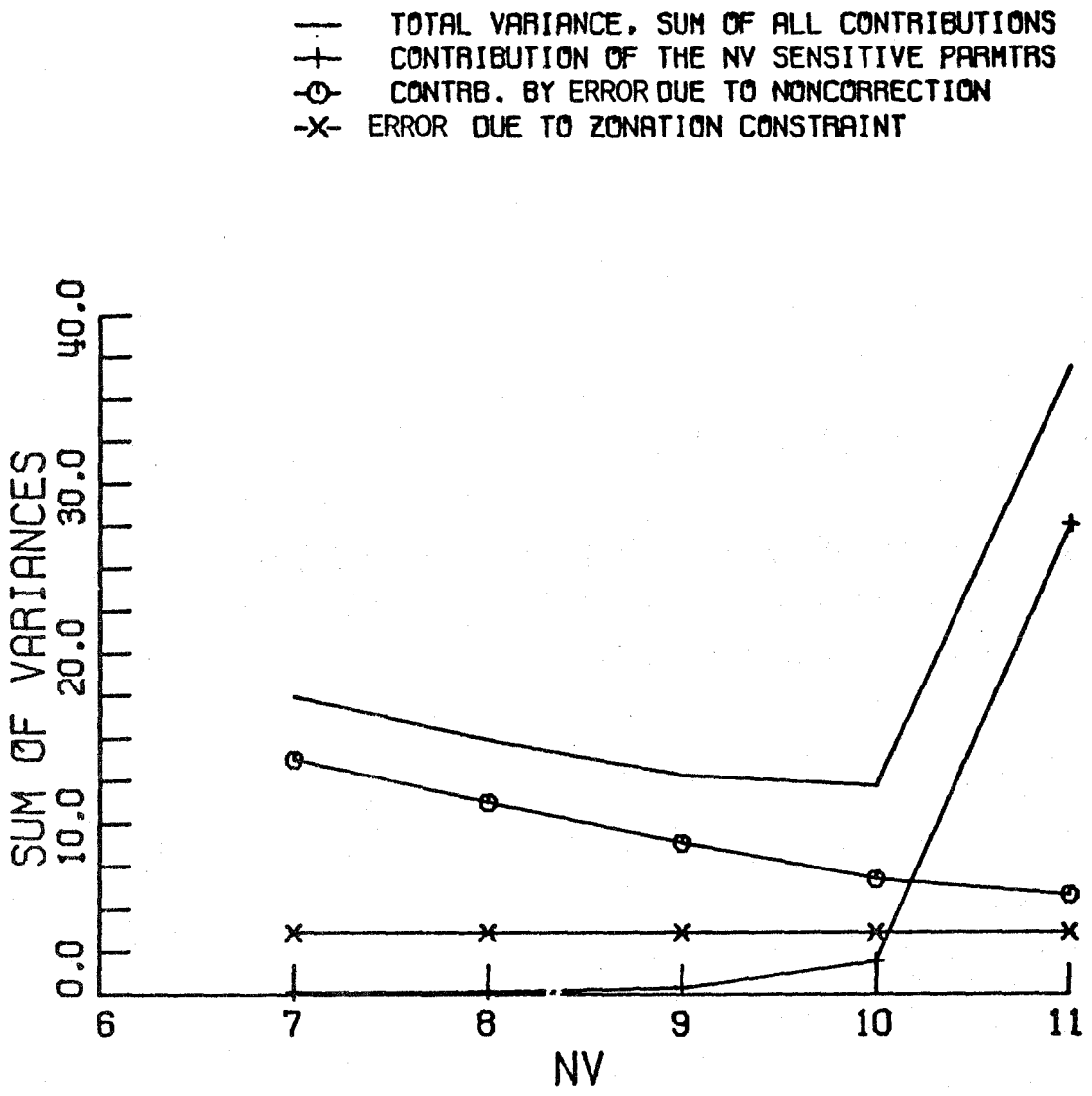
3.4.2.3 Estimation with Hard Constraints

As discussed previously the value of N_v at which the minimum of $tr\{\underset{\sim}{P}_\pi\}$ occurs changes significantly with β. As a quantitative example we present details of the trade-off between the contributions by uncertainty and the parameterization error, for zonation with NZ = 8. While the other parameters are kept fixed at s = 5.0 and ρ = 0.5, the trade-off will be examined for β = 0.16, 1.0 and 0.9. Figures (3.4.9a, b) contain the details of the trade-off determining the optimal N_v for β = 0.16 and β = 9.0 respectively. The trade-off for β = 1.0 is shown in figure (3.4.3b). The optimal values of N_v for β = 0.16, 1.0 and 9.0 occur at 8, 9 and 9 respec-



TRADEOFF BETWEEN UNCERTAINTY AND PARAM. ERROR
EFFECT OF PRIOR COVARIANCE
NZ=8 S=5. BETA=0.16

Figure 3.4.9a



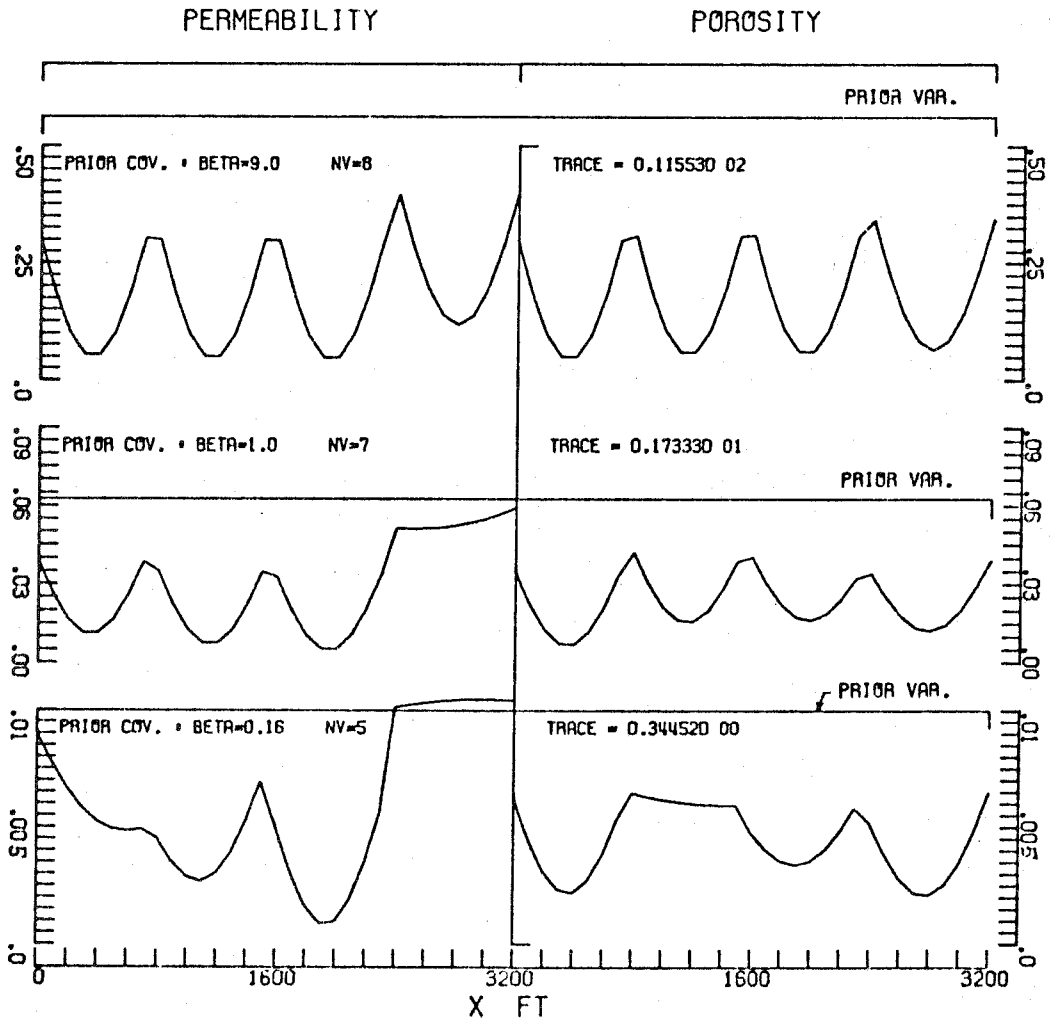
TRADEOFF BETWEEN UNCERTAINTY AND PARAM. ERROR
EFFECT OF PRIOR COVARIANCE
NZ=8 S=5. BETA=9.0

Figure 3.4.9b

tively. The sudden increase with N_v of the contribution by the sensitivity term beyond $N_v = 9$ causes the optimal values for the different cases involving large β to be clustered around $N_v = 9$.

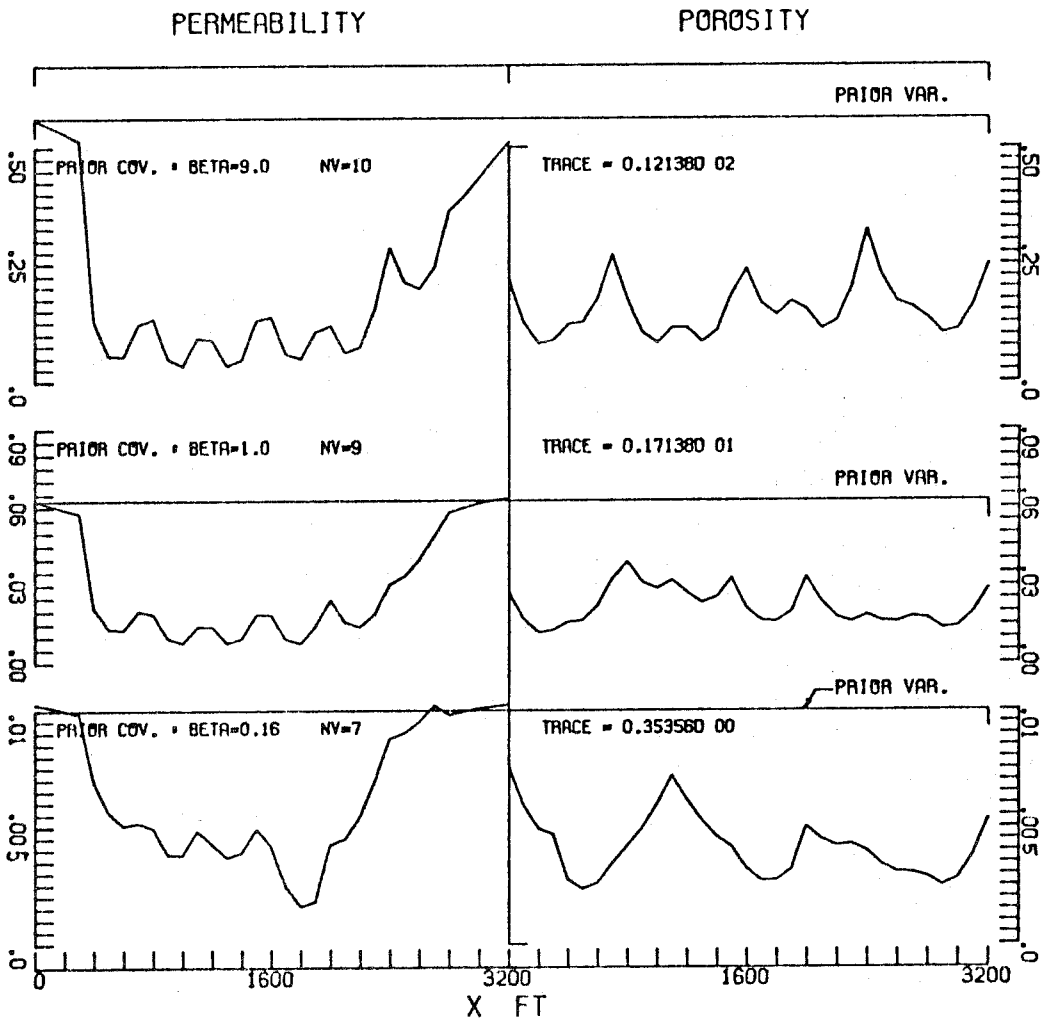
Figure (3.4.10) contains the results for zonation with four zones. In this case, the optimal value of N_v changes considerably with β due to the large contribution of the parameterization errors to the total covariance, and also because the contribution of the sensitivity term does not increase very rapidly with N_v. Consequently, the contributions of the error due to noncorrection and of the cross covariance terms are considerably different for the different values of β, altering significantly the distribution of the a posteriori variances. For small value of β, the uncertainty is kept small by taking a small value of N_v. Hence the error due to noncorrection is large and the variances are relatively large for the parameters with low sensitivity. For $\beta = 9$, the prior information is so little that it is advantageous to include in $\overline{\underset{\sim}{V}}_1$ all eight of the vectors in the parameter space for large β, the error due to the zonation constraint dominates the picture and the variances are large in the regions near the zone boundaries.

Figures (3.4.11) and (3.4.12) show the results for the 8-zone and 16-zone cases. As the number of the zones increases, the contribution of the parameterization errors decreases and the changes in the "optimal" value of N_v become smaller; thus the distributions of the variance are similar for the different values of β for a case with a large number of zones. In the 8-zone case, for $\beta = 0.16$ additional the less sensitive vectors are included in $\overline{\underset{\sim}{V}}_0$ and



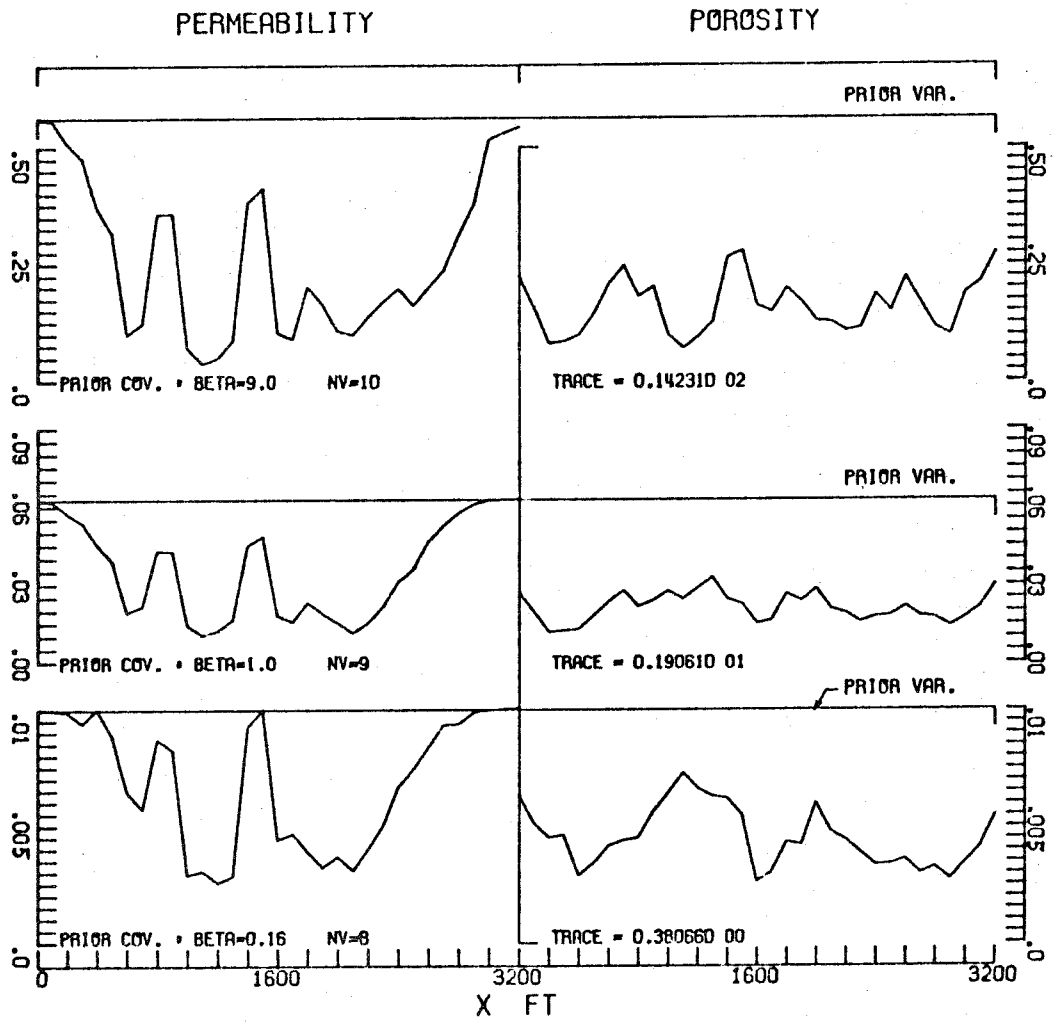
EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=4 S=5

Figure 3.4.10



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=8 S=5

Figure 3.4.11



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=16 S=5

Figure 3.4.12

consequently variances of the permeability in the regions close to the reservoir boundaries are practically unchanged from their prior values. For $\beta = 9$, the error due to the zonation constraint becomes a prominent contribution. In the 16-zone case, these effects are smaller and the distributions of the variance for different values of β are not very dissimilar.

In conclusion, we state that the most important effect of β on the a posteriori variances is in the determination of their magnitude. The variances of the different parameters do not alter significantly relative to each other. Thus an error in specifying the prior variances will cause an approximately uniform misjudgement in the level of confidence in the resulting estimates of the grid point values of the rock properties; the relative degree of confidence in the different parameters will be largely unaffected by such an error. An error in the prior variances by a certain factor will result in a change by approximately the same factor in the a posteriori variances.

3.4.3 Effect of Change in Cross-Correlation in Prior Covariance

The effect of the parameter ρ in the prescribed prior covariance $\underset{\sim}{P}_0$ on the a posteriori covariance for the different cases of parameterization was investigated by repeating the computations with $\rho = 0.25$ and 0.75 while keeping β and s fixed at 1.0 and 5.0 respectively.

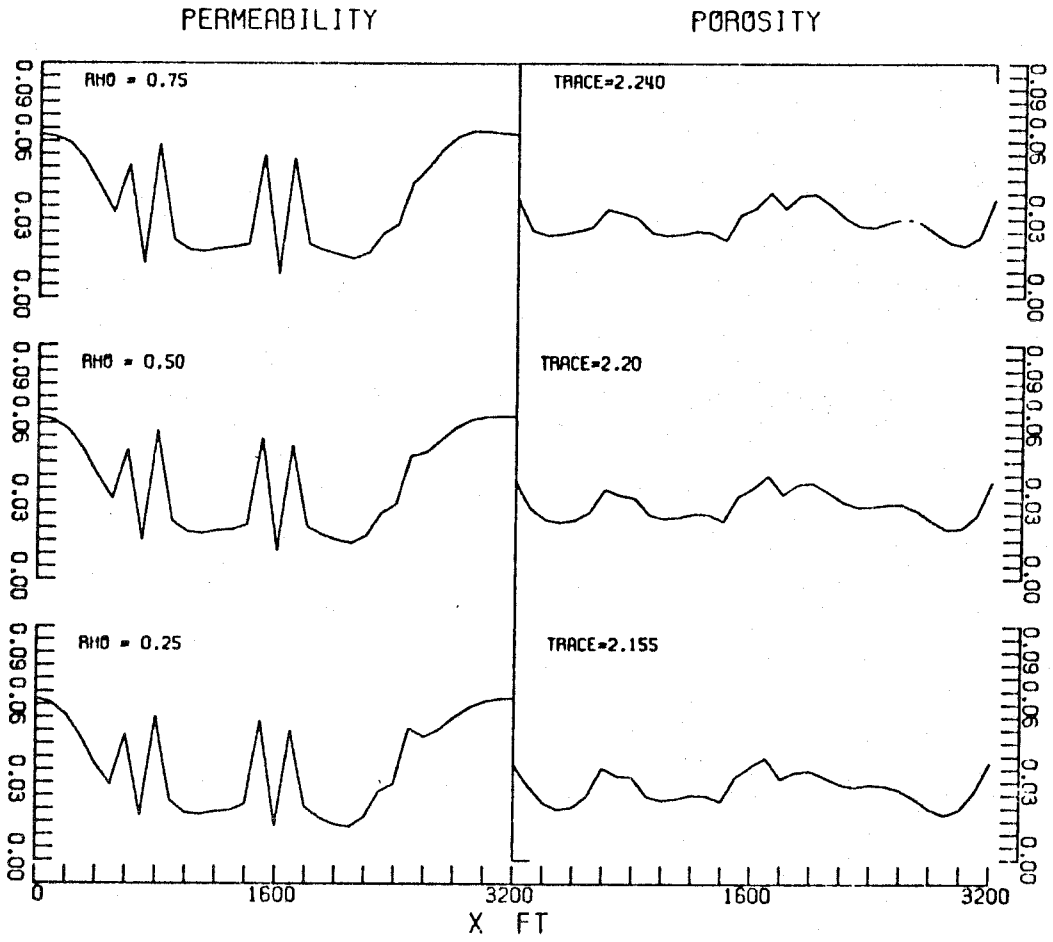
3.4.3.1 Estimation without Constraints

The resulting variances for the case of estimation with-

out constraints are presented in figure (3.4.13). The three-fold change in the value of ρ does not appreciably alter the a posteriori covariance for this case; its trace increases slightly as ρ increases. This is due to the following. A higher value of ρ corresponds to a higher correlation in the distributions of the permeability and porosity. As is evident from figure (2.5.5), the components of the first several v-vectors in the permeability-space and the porosity-space are quite dissimilar; this causes the corrections in their estimates to have much different distributions. This results in a higher error due to noncorrection for a higher value of ρ. The contribution of the sensitivity term does not change as the v-vectors and the value of N_v are the same for all values of ρ. The details of the distributions of the variances do not differ significantly for the three values of ρ.

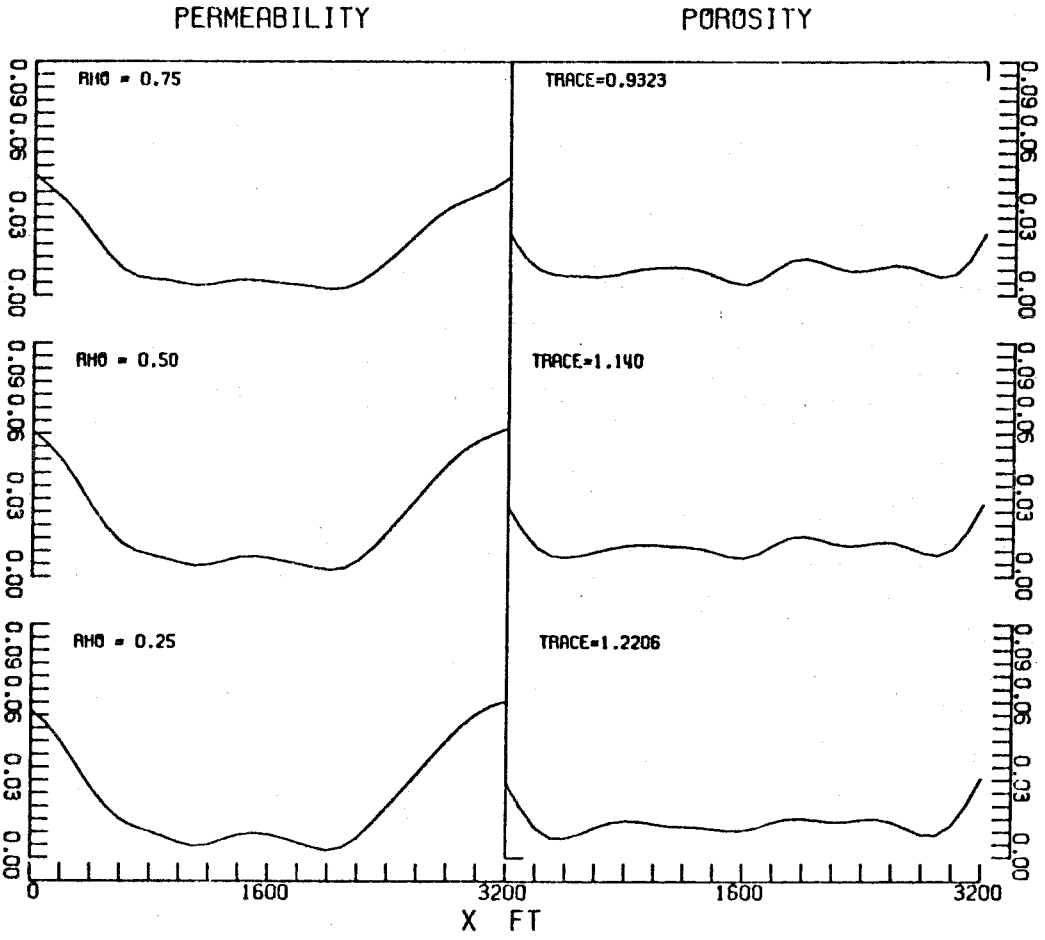
3.4.3.2 Bayesian Estimation

Figure (3.4.14) presents the results for the Bayesian estimation. In this case, the trace of the a posteriori covariance decreases fairly rapidly with increase in ρ. This is due to the fact that this approach utilizes the information in the prior covariance for the purpose of estimation. A larger value of ρ implies greater prior information about the parameter distributions; hence the resulting estimates are less uncertain for a larger value of ρ in $\underset{\sim}{P}_0$. The decrease in the trace is achieved by an approximately uniform decrease in the variances of k and ϕ all over the reservoir. The sensitivity of the observations with respect to the porosity is approximately uniform over the reservoir; for a larger ρ, this results in



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
NO CONSTRAINTS NZ=33, NV=9 BETA=1.S=5

Figure 3.4.13



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
BAYESIAN ESTIMATION BETA=1.S=5

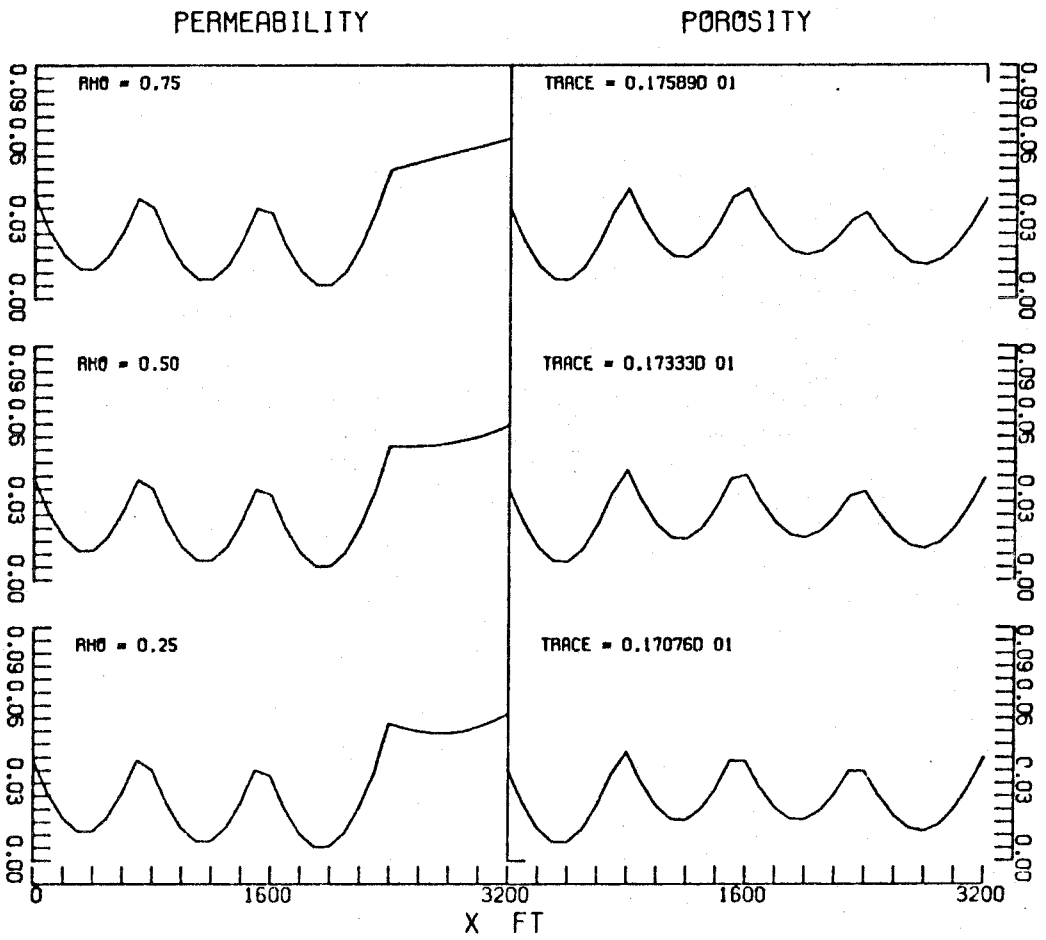
Figure 3.4.14

a greater information in the observations about the permeability in the regions where its sensitivity is low. Hence the variances of the permeability near the impermeable boundaries are significantly smaller and the "trough" in the k-variance distribution becomes wider for a higher value of ρ.

3.4.3.3 Estimation with Hard Constraint

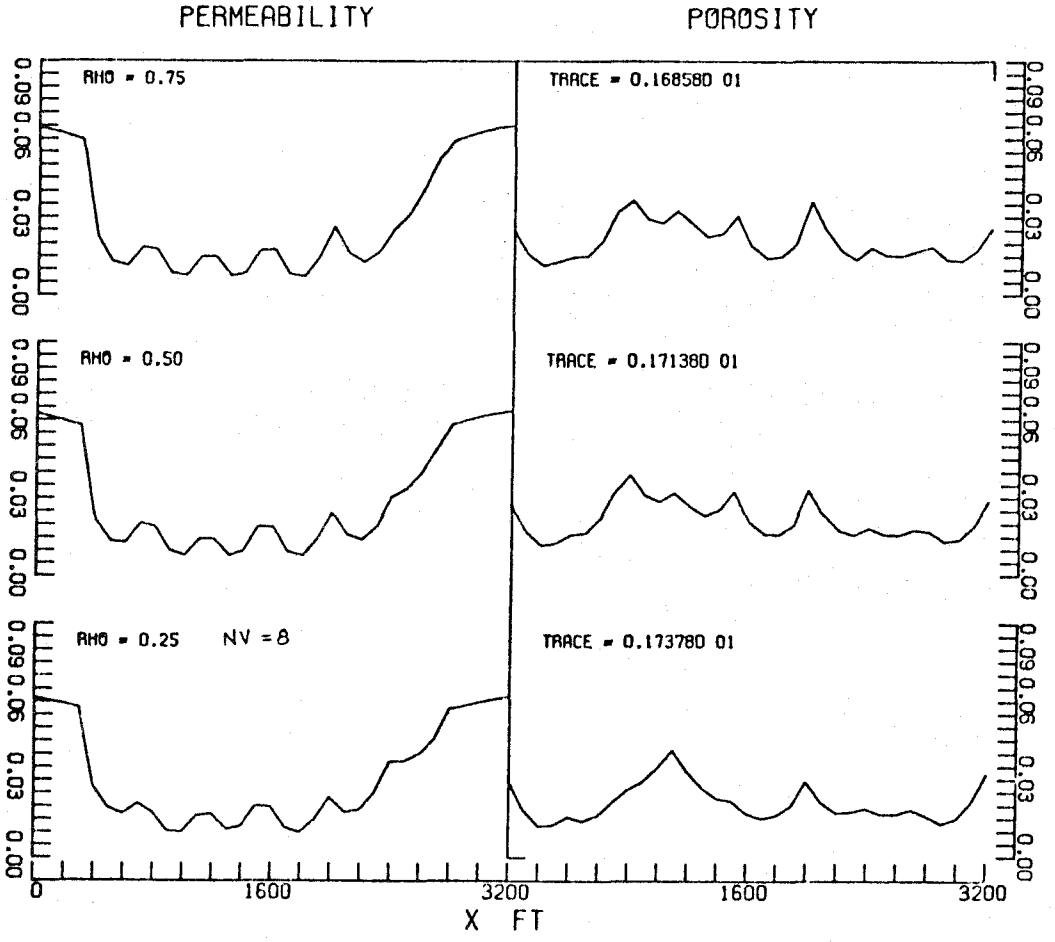
Figure (3.4.15) presents the results for the four-zone case. As in the case of estimation without constraints, the trace increases with increase in ρ because of the increased error due to noncorrection. The variance of the permeability in the region close to the x = 3200 ft boundary is particularly affected. This can be understood from the fact that, for $N_v = 7$, only a single v-vector with the lowest singular value contributes to the noncorrection bias; the largest component of this vector corresponds to the zone covering this region. The error due to the zonation constraints does not change with ρ. As observed earlier, the error due to noncorrection is a very small contribution to the total covariance for the four-zone case. Hence the effect of the three-fold change in ρ is quite small.

Figure (3.4.16) shows the results for the eight-zone case. In this case, too, the trace of the a posteriori covariance does not change significantly with variations in ρ. However, the trend in its variation is contrary to that observed in the rest of the zonation cases. This may be related to the fact that, as presented in figure (3.2.9), the contribution of the parameterization errors to the total covariance is the smallest for NZ = 8. The details of the analysis



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=4 , NV=7 BETA=1.S=5

Figure 3.4.15



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=8 , NV=9 BETA=1,S=5

Figure 3.4.16

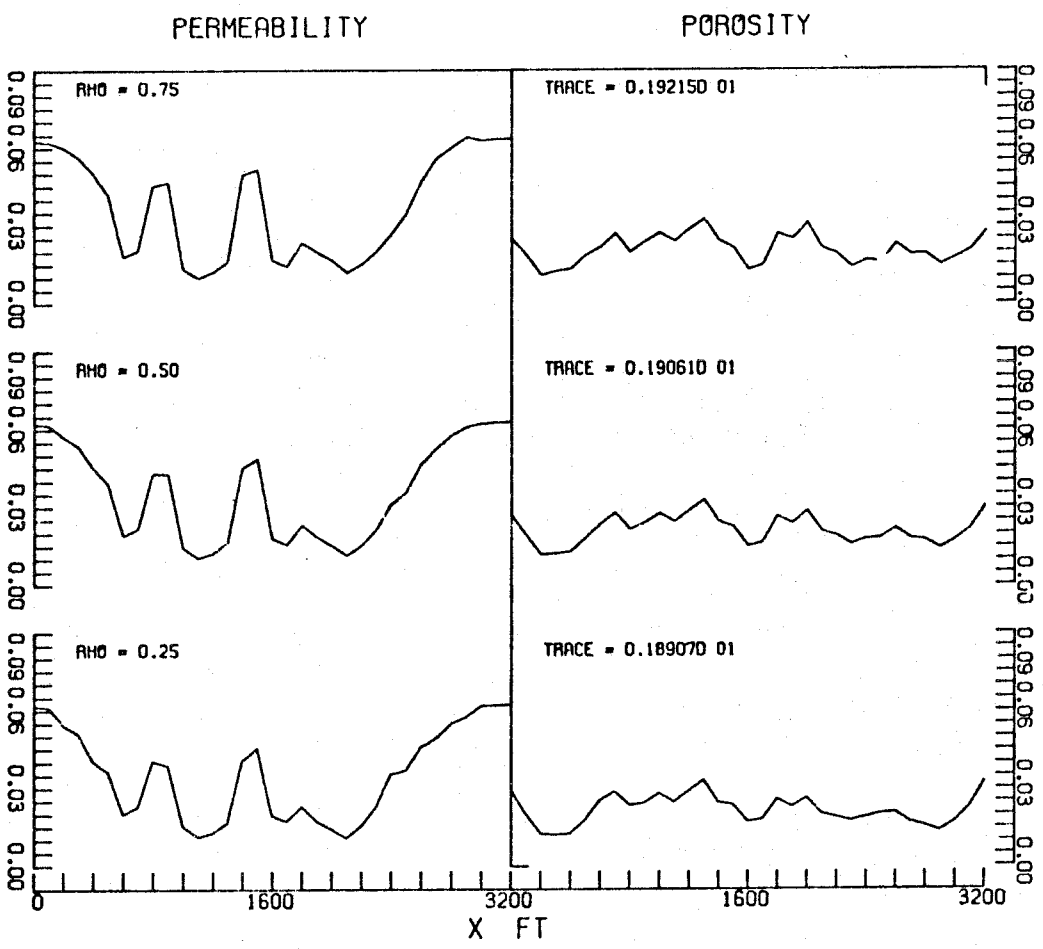
reveal that this error actually decreased with increasing ρ for this case. As indicated in the figure, the "optimal" value of N_v for $\rho = 0.25$ is different from the other two cases, even though the difference in the values of the trace for $N_v = 9$ and $N_v = 8$ is only 0.004. This change of N_v causes the details of the variance distribution for $\rho = 0.25$ to be significantly different from the other two cases.

Figure (3.4.17) contains the results for the 16-zone case. Here again, the a posteriori variances change very little with the three-fold increase in ρ, the trace increasing very slightly with ρ because of the increased error due to noncorrection.

In summary, we conclude that even large changes in the value of the parameter ρ do not alter significantly the a posteriori variances when the process of estimation does not utilize the prior covariance. The effect is significant for the Bayesian estimation, where the trace of the a posteriori covariance decreases by about 30% when ρ is increased from 0.25 to 0.75.

3.5 Effect of Erroneous Prior Covariance

In this section we shall investigate the effect of an error in specifying the prior covariance on the property estimates and the predicted covariances. First we shall treat the non-Bayesian estimation which includes the estimation without constraints and the estimation with hard constraints. (We shall treat the former as a special case of the latter.) Later we shall explore the effect on the Bayesian estimates.



EFFECT OF CHANGES IN PRIOR COV. ON VARIANCES
ZONATION NZ=16, NV=9 BETA=1,S=5

Figure 3.4.17

3.5.1 Non-Bayesian Estimation

As the prior covariance is not utilized in the process of estimation, the estimates are not affected by an error in the prior covariance. However, the predicted covariance of the estimates depends on the prior covariance and hence will be significantly affected by any error made in its specification. The resulting error in predicted covariance is easily calculated when the changes in N_v are overlooked since in that case its dependence on the prior covariance is linear.

Let $\underset{\sim}{P}_0$ be the "true" prior covariance and let $\overline{\underset{\sim}{P}}_0$ be its erroneously specified value. Then the "true" a posteriori covariance of the estimates is given by expression (3.1.54). (Set $\underset{\sim}{G}_1 = \underset{\sim}{I}$ in that and in the following expression for obtaining the results for the special case of estimation without constraints.) The predicted a posteriori covariance based on $\overline{\underset{\sim}{P}}_0$ is obtained by replacing $\underset{\sim}{P}_0$ by $\overline{\underset{\sim}{P}}_0$ in (3.1.54). Subtracting the two, the error in the predicted covariance is given by

$$\begin{aligned}\Delta\underset{\sim}{P}_\pi = (\underset{\sim}{P}_\pi)_{\overline{P}_0} - (\underset{\sim}{P}_\pi)_{P_0} &= \underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\,\Delta\underset{\sim}{P}_0\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T \\ &+ (\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T)\,\Delta\underset{\sim}{P}_0(\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T) + (\underset{\sim}{I}-\underset{\sim}{G}\underset{\sim}{G}_1^T)\Delta\underset{\sim}{P}_0\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T \\ &+ \underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\,\Delta\underset{\sim}{P}_0(\underset{\sim}{I}-\underset{\sim}{G}_1\underset{\sim}{G}_1^T) \qquad (3.5.1)\end{aligned}$$

where $\Delta\underset{\sim}{P}_0 = \overline{\underset{\sim}{P}} - \underset{\sim}{P}_0$.

The effect on the computational results due to the erroneous prior covariance is amply demonstrated by the results in section 2.4. We conclude from these that $\Delta\underset{\sim}{P}_\pi$ is directly proportional to $\Delta\underset{\sim}{P}_0$ and

a large error $\Delta\underset{\sim}{P}_0$ will lead to a comparably large error $\Delta\underset{\sim}{P}_\pi$ in the predicted covariance. We note that this analysis is valid for $\Delta\underset{\sim}{P}_0$ which do not change significantly "optimal" values of N_v used for determining the partition between $\overline{\underset{\sim}{V}}_0$ and $\overline{\underset{\sim}{V}}_1$.

3.5.2 Bayesian Estimation

In the Bayesian approach, on the other hand, the prior covariance $\underset{\sim}{P}_0$ is utilized in obtaining the property estimates as well as their covariance. Thus, both of these are affected by an error in specification of $\underset{\sim}{P}_0$. However, as long as the prior mean $\overline{\underset{\sim}{\pi}}$ is correctly specified, then from (3.1.40, 42) we conclude that the resulting estimate will not be biased, irrespective of an error made in specifying $\underset{\sim}{P}_0$.

In the following we derive the covariance of error in the Bayesian estimate associated with an erroneous prior covariance $\overline{\underset{\sim}{P}}_0$. Expression (3.1.37) yields in this case

$$\hat{\delta\underset{\sim}{\pi}}_r = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\overline{\underset{\sim}{P}}_0^{-1})^{-1}\ [\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{\delta y}_{hr}-\overline{\underset{\sim}{P}}_0^{-1}(\hat{\underset{\sim}{\pi}}_h-\overline{\underset{\sim}{\pi}})] \quad (3.5.2)$$

Then,

$$\hat{\delta\underset{\sim}{\pi}}_r-\delta\underset{\sim}{\pi}_r = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\overline{\underset{\sim}{P}}_0^{-1})^{-1}\ [\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\delta y}_{hr}-\overline{\underset{\sim}{P}}_0^{-1}(\hat{\underset{\sim}{\pi}}_h-\overline{\underset{\sim}{\pi}})$$

$$-\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}\ \underset{\sim}{\delta\pi}_r - \overline{\underset{\sim}{P}}_0^{-1}\underset{\sim}{\delta\pi}_r] \quad (3.5.3)$$

Using (3.1.10), (3.1.39) and (3.1.46) we obtain

$$\hat{\delta\underset{\sim}{\pi}}_r - \delta\underset{\sim}{\pi}_r = \hat{\underset{\sim}{\pi}}_{\overline{B}} - \underset{\sim}{\pi}^* = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\overline{\underset{\sim}{P}}_0^{-1})^{-1}\ [\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\eta} -\overline{\underset{\sim}{P}}_0^{-1}(\underset{\sim}{\pi}^*-\overline{\underset{\sim}{\pi}})]$$

(3.5.4)

where we denote the Bayesian estimate resulting from the use of the erroneous prior covariance $\overline{\underset{\sim}{P}}_0$ by $\hat{\underset{\sim}{\pi}}_{\overline{B}}$. It follows from the definition of the true covariance $\underset{\sim}{P}_0$ that

$$E_{\eta,\pi}\{(\hat{\underset{\sim}{\pi}}_{\overline{B}}-\underset{\sim}{\pi}^*)(\hat{\underset{\sim}{\pi}}_{\overline{B}}-\underset{\sim}{\pi}^*)^T\} = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\overline{\underset{\sim}{P}}_0^{-1})^{-1}[\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A} + \overline{\underset{\sim}{P}}_0^{-1}\underset{\sim}{P}_0\overline{\underset{\sim}{P}}_0^{-1}](\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\overline{\underset{\sim}{P}}_0^{-1})^{-1} \quad (3.5.5)$$

By adding and subtracting $\overline{\underset{\sim}{P}}_0$ from the central factor on the right of (3.5.5), it can be rewritten as

$$\underset{\sim}{P}_{\overline{B}} = (\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A} + \overline{\underset{\sim}{P}}_0^{-1})^{-1} + (\overline{\underset{\sim}{P}}_0\underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}+\underset{\sim}{I})^{-1}(\underset{\sim}{P}_0-\overline{\underset{\sim}{P}}_0)(\underset{\sim}{I} + \underset{\sim}{A}^T\underset{\sim}{\Sigma}^{-1}\underset{\sim}{A}\overline{\underset{\sim}{P}}_0)^{-1} \quad (3.5.6)$$

where we denote the covariance of error in the Bayesian estimate using the erroneous $\overline{\underset{\sim}{P}}_0$ by $\underset{\sim}{P}_{\overline{B}}$. It is evident that for $\overline{\underset{\sim}{P}}_0 = \underset{\sim}{P}_0$, $\underset{\sim}{P}_{\overline{B}} = \underset{\sim}{P}_\pi$. The covariance $\underset{\sim}{P}_{\overline{B}}$ was evaluated for several erroneous values of β and s; the trace of $\underset{\sim}{P}_{\overline{B}}$ for the different cases are shown in table 3.5.1. From the values therein, it follows that even large errors in β and s do not cause very large errors in the Bayesian estimates. This prediction, based on the linearized analysis and some approximations (see Section 3.1), is reasonably well verified by the results of simulations described in Chapter 1. Thus, when accurate $\underset{\sim}{P}_0$ is not available, an approximate value or an estimate for it can be used in the Bayesian estimation without incurring serious errors in the resulting estimates.

Table 3.5.1

Bayesian Estimation: Effect of

Erroneous Prior Covariance

True Parameters: $\beta = 1.0$, $s = 5.0$

β	s	Trace of Error Covariance
1.0	1.0	1.830
1.0	2.5	1.280
1.0	5.0	1.140
1.0	7.5	1.255
1.0	10.0	1.538
1.0	15.0	2.222
0.16	5.0	1.301
1.0	5.0	1.140
9.0	5.0	1.772

3.6 Conclusions

(1) As there is a many-dimensional subspace of the parameter space along which the components of the estimates are ill-determined by the pressure observations and the production rate history data, the variances of the estimates are unacceptably large in absence of any additional information. In this chapter, we derive realistic estimates of the covariances associated with the prior estimates of the parameters. There are two components that make up the a posteriori covariance: (a) The uncertainty in the estimates associated with the errors in observations which depends on the sensitivity of the observations with respect to the parameters. (b) The error associated with the initial guess and the fact that the components of the parameter vector with very little influence on the observations are not updated while history matching. In addition, the covariance determination must take into account the nature of the algorithm and the constraints imposed by the parameterization. In this chapter expressions are derived for the covariance of the error in estimates resulting from the use of the zonation and Bayesian approaches.

(2) The predicted covariances were tested using simulations with a sample of four realizations of a homogeneous Gaussian random process model, with prescribed statistical properties, for porosity and permeability in a one-dimensional reservoir. A good match was observed between the spatial distributions of the predicted variances and the mean square estimate errors

averaged over the four realizations, both for the zonation and the Bayesian approaches to estimation.

(3) The sensitivity of the predicted a posteriori covariance to variation in the specified prior covariance was studied using the homogeneous random process model for the property distributions. It was found that a change in the specification of the length s of the autocorrelation function has little influence on the a posteriori variances, the variances decreasing with increasing s. A change in the prescribed prior variances has a larger effect on the results, the percentage change in the resulting variances being same as that for the prior variances.

(4) It was found that for a given zonation, it is possible to determine an optimal value of μ, the parameter in Marquardt's method, by computing the predicted a posteriori covariance beforehand, by using the prior parameter estimates for linearization. The optimality is in the sense that, the use of the indicated value of μ in Marquardt's method for history matching will yield estimates with the smallest trace of the a posteriori covariance associated with them.

(5) The effect of erroneous prior covariance on the estimates and a posteriori covariance was studied for the different approaches to estimation. In zonation such an error does not influence the estimates, but it significantly affects the predicted covariances; the most important parameter for this is β, which governs the degree of confidence in the prior estimate. On the other hand, in the Bayesian approach, the estimates are affected

by such an error; but for an unbiased prior estimate the resulting estimate is also unbiased. Any error in the prior covariance increases the covariance of error in the resulting estimate; however, even for large errors in $\underset{\sim}{P}_0$, the increase in the latter is relatively small.

CHAPTER 4

CONCLUSIONS

In this thesis, we have dealt with the inverse problem concerning the determination of porous rock property distributions in petroleum reservoirs from the well pressure and production data. The analysis and methodology developed in this work can be easily extended to the general problem of determination of parameter distributions in arbitrary dynamic systems from the input-output data.

The problem of property estimation in a petroleum reservoir is an underdetermined problem. Its solution necessitates the reduction of the number of unknowns by some form of parameterization. When zonation is used for this purpose, the best estimates are obtained for an intermediate number of zones. As an alternative to zonation, we have developed a method of reducing the number of unknowns and decreasing the uncertainty of the estimates by extension of Bayesian estimation theory. It involves the addition to the objective function of a penalty term incorporating statistical, geological information. The results of Bayesian estimation were found to be more accurate than those of zonation in simulation with one-dimensional reservoirs even when the specified prior information contained large errors. Both Bayesian and zonation approaches to property estimation were implemented using the conjugate gradient and the Gauss-

Newton (or Marquardt's) minimization algorithms. Detailed estimates are obtained for the computational effort per iteration of either algorithm for zonation and Bayesian estimation in one-dimensional and two-dimensional reservoirs. Formulas are developed to enable one to determine which method is expected to be more efficient for a particular problem. Two alternative methods for computation of the sensitivity coefficients are considered for Gauss-Newton (or Marquardt's) method and the associated computational effort for one and two-dimensional reservoirs are determined.

The degree of ill-determinacy of the property estimation problem is quantified through its linearized analysis. It is shown that it is possible to determine, through the linearized analysis, which components of the parameters cannot be estimated accurately for a given problem. The highly oscillatory components of the porosity and the permeability are found to have very small influence on the pressure observations, and thus are ill-determined. In reservoirs with impermeable boundaries, the permeability in the regions close to the boundaries is particularly ill-determined. The porosity and permeability in the regions close to the production wells which are also observation sites, are well-determined since they exert a large influence on the observations.

Analytical expressions are derived for the sensitivity coefficients of an observed pressure with respect to small, arbitrary perturbations in the porosity and permeability in

a single phase one-dimensional reservoir, with either impermeable boundary or constant pressure boundary. In reservoirs with either kind of boundary conditions, the Fourier component of the permeability with wave number $\ell(\ell \neq 0)$ has $0(\ell^{-1})$ influence on the pressure at large times from the beginning of production; this implies that highly oscillatory components of the permeability cannot be estimated from the observed pressures. In the reservoir with impermeable boundary conditions, a similar component $(\ell \neq 0)$ of porosity has $0(\ell^{-2})$ influence, whereas for $\ell = 0$ the influence increases linearly with time. Thus the mean porosity can be very accurately determined but its oscillatory components cannot. In the constant pressure boundary reservoir, the influence of the various porosity components on the pressure vanishes exponentially with time (measured from the beginning of the production) and thus accurate determination of the porosity based only on the observed pressures is not possible. These observations are in agreement with the results of numerical analysis discussed in the previous paragraph.

The nature of the linearized relation between the observations and the rock properties is found to be only weakly dependent on the reference distributions used for linearization. Thus the analysis can be performed before extensive estimation attempts. Furthermore, the results of analysis using a set of mean distributions yield information about the property estimation problem concerning an ensemble of reservoirs with properties differing statistically from these means.

A scalar index is developed, from the results of linearized analysis, that indicates the degree of accuracy to which a given rock property at any location in the reservoir can be estimated from a given set of pressure data. This index can be calculated a priori, using nominal mean estimates for linearization. It can be used to study the conditioning of the problem under any parameterization, and thus aids determination of an "optimal" parameterization. These practical uses of the linearized analysis are detailed in section 2.8, which is addressed to a practicing reservoir engineer.

The information about the unknown rock properties comes from two sources: (a) the pressure observations, and (b) the prior statistical information. The observations contain information about only those components of the unknown distributions which significantly influence the observations. The errors in the estimates of these components can be statistically determined from the linearized analysis. The error in estimates of the rest of the components can only be estimated based on the prior information. These two can be synthesized to determine the covariance of errors in estimates. Expressions are developed, based on the properties of the linearized problem, for covariances of estimates from zonation and Bayesian estimation.

The predicted variances for the grid point property values were compared with actual mean square estimate errors in simulations using a limited sample of four property distributions with known statistics. A good agreement was found

for both zonation and Bayesian estimation.

The sensitivity of the predicted estimate covariance to variation in the a priori statistics was studied. Changes in the prescribed length of autocorrelation for spatial property variations and the cross-correlation between the porosity and permeability do not have large influence on the predictions. The degree of confidence in the prior estimates determines to a great extent the covariance of the resulting estimates; large prior variances result in large a posteriori variances. The effect of erroneous specification of the prior statistics on the predicted covariances was also investigated. Again, only the errors in the prior variances significantly affect the a posteriori covariance of the zonation estimates. On the other hand, the predicted covariance for Bayesian estimates is not strongly influenced by errors in the prior statistics.

Using the predicted contributions to the covariance of the parameterization error and the uncertainty in the estimates resulting from the observation errors, it is possible to determine, for a given problem, a zonation that yields the smallest expected total error in the estimates. Thus "optimal" zonation can be determined before detailed estimation. It is also possible, based on the linearized analysis, to determine for a given zonation the parameter μ in Marquardt's algorithm for history matching, such that the expected total error in the resulting estimates is the smallest. Such a design of an "optimal" history matching algorithm for zonation is discussed in detail in chapter 3.

Appendix 1.1

Numerical Procedure for Generating the Samples of a Gaussian Random Vector

Theory:

The procedure followed utilizes the Loeve-Karhunen expansion of a stationary random process into orthogonal functions (Davenport and Root, 1958). We shall describe the discrete version of the procedure.

Let the Gaussian random M-vector $\underset{\sim}{\pi}$ have mean $\overline{\underset{\sim}{\pi}}$ and covariance matrix $\underset{\sim}{P}_o$. Since $\underset{\sim}{P}_o$ is a symmetric and positive definite matrix, all of its eigenvalues are real and positive. Let the eigenvalues be $\lambda_1 > \lambda_2 > \ldots > \lambda_M$ and let the corresponding orthonormal set of eigenvectors of $\underset{\sim}{P}_o$ be $\underset{\sim}{z}(1)$, $\underset{\sim}{z}(2) \ldots . \underset{\sim}{z}(M)$. Defining the matrices

$$\underset{\sim}{Z} = (\underset{\sim}{z}(1), \underset{\sim}{z}(2) \ldots . \underset{\sim}{z}(M)) \qquad \text{(A1.1.1}$$

$$\underset{\sim}{\Lambda} = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_M) \qquad \text{(A1.1.2)}$$

we have,

$$\underset{\sim}{P}_o = \underset{\sim}{Z}\,\underset{\sim}{\Lambda}\,\underset{\sim}{Z}^T. \qquad \text{(A1.1.3)}$$

Consider the decomposition of the random vector $(\underset{\sim}{\pi} - \overline{\underset{\sim}{\pi}})$ along the complete orthonormal set of vectors $\underset{\sim}{z}(1), \ldots . , \underset{\sim}{z}(M)$.

$$\underset{\sim}{\pi} - \overline{\underset{\sim}{\pi}} = \underset{\sim}{Z}\,\underset{\sim}{u} = \sum_{i=1}^{M} u_i\, \underset{\sim}{z}(i) \qquad \text{(A1.1.4)}$$

Then, due to orthogonality of $\underset{\sim}{Z}$,

$$\underset{\sim}{u} = \underset{\sim}{Z}^T(\underset{\sim}{\pi} - \overline{\underset{\sim}{\pi}}) \qquad \text{(A1.1.5)}$$

In the last expression, $\underset{\sim}{u}$ also is a Gaussian random vector with the following mean and covariance:

$$E\{\underset{\sim}{u}\} = \underset{\sim}{0} \tag{A1.1.6}$$

$$E\{\underset{\sim}{u}\underset{\sim}{u}^T\} = \underset{\sim}{Z}^T \underset{\sim}{P}_o \underset{\sim}{Z} = \underset{\sim}{\Lambda} \tag{A1.1.7}$$

Then the elements $u_1, u_2, \ldots, u_M$ of $\underset{\sim}{u}$ are independently distributed Gaussian random variables with zero mean and the variances $\lambda_1, \lambda_2, \ldots, \lambda_M$. Then, the samples of the random variables $u_1, u_2, \ldots, u_M$ can be obtained, one at a time, using a random number subroutine and the corresponding sample of the random parameter vector is given by,

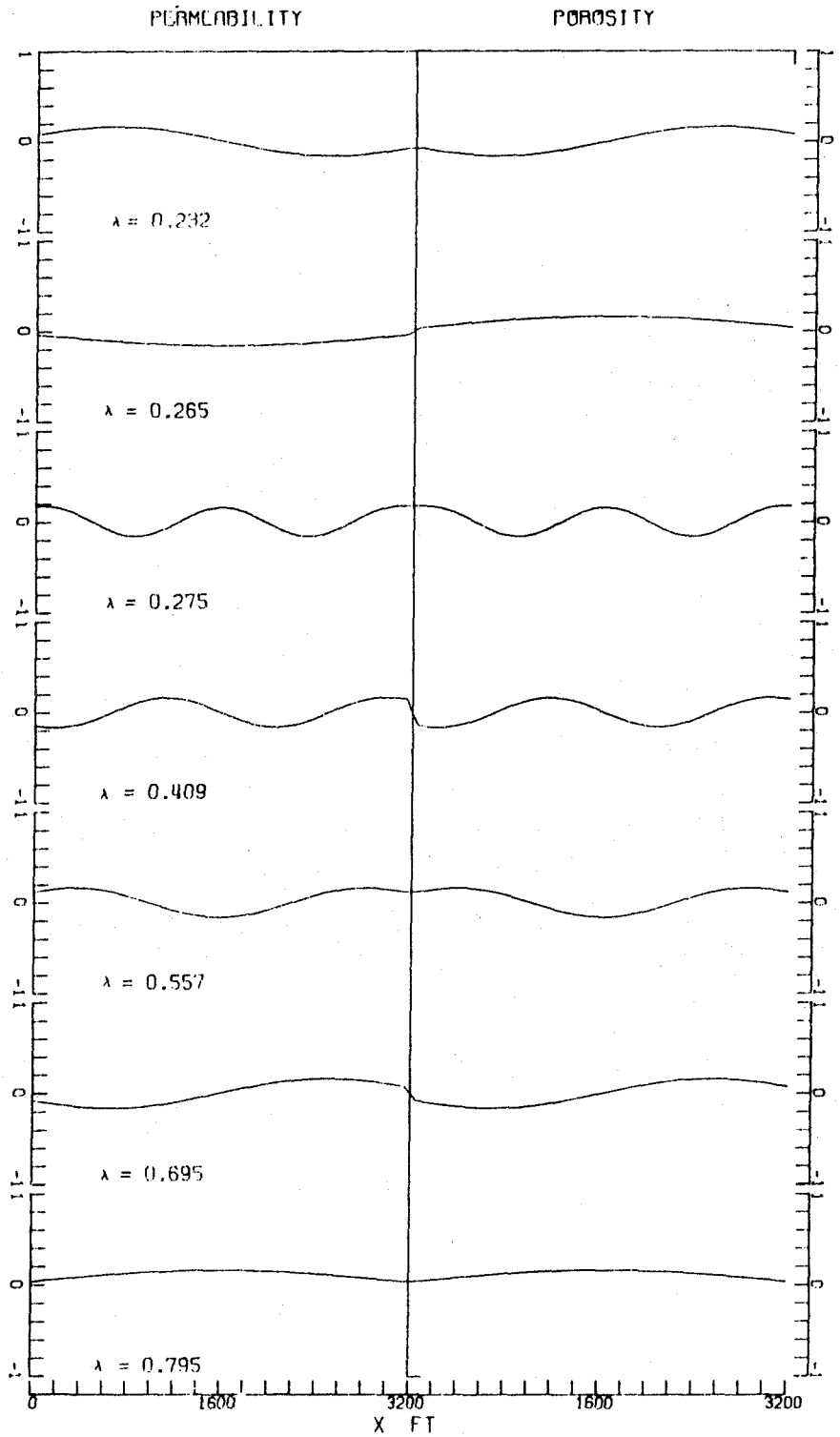
$$\underset{\sim}{\pi} = \underset{\sim}{\pi} + \underset{\sim}{Z}\,\underset{\sim}{u} \tag{A1.1.8}$$

Numerical Details: All computations were performed for constant thickness. The numerical values of k and ϕ in the conventional units are usually one or two orders of magnitude apart. Consequently their variances differ by two to four orders of magnitude. Under these circumstances, when the composite co-variance matrix $\underset{\sim}{P}_o$ is decomposed, the largest few eigenvalues have eigenvectors that lie almost entirely in the k-subspace and the eigenvalues smaller by two to four orders of magnitude have eigenvectors that lie mainly in the ϕ-subspace of the parameter space. This situation implies that we would have to use almost all the eigenvectors of $\underset{\sim}{P}_o$ for a basis while generating the sample functions and in all the subsequent estimation procedures that utilize the prior statistical information. This situation would demand large storage space and computational times. In addition, it would

give rise to numerical difficulties due to widely different orders of magnitude of the errors due to roundoff errors in the k and ϕ subspaces of the parameter space.

These difficulties were avoided by scaling the variables k and ϕ. All the computations for generating the sample functions and in the subsequent estimation procedures were carried out in terms of the normalized variables $k/\overline{k}$ and $\phi/\overline{\phi}$. Then, the normalized variables always have root-mean-square values of unity and variances that are numerically quite close to each other.

When the smooth autocorrelation functions given by the expression (1.3.1) are used, the corresponding matrix $\underset{\sim}{P}_o$ for the normalized variables has a few eigenvalues that are O(1) and the rest drop off almost geometrically with uniform rapidity to yield a value of $O(10^{-7})$ for the last eigenvalue. As a very large number of eigenvalues λ_i are several orders of magnitude smaller than the largest eigenvalue λ_1, the amplitudes u_i of the corresponding components will be, on the average, correspondingly smaller than the others. Hence, without loss of much accuracy, we may truncate the summation in (A1.1.4) after a suitable number of terms. In our computations, truncation criterion was $\lambda_i/\lambda_1 < 0.5\times10^{-4}$. For s = 5, this resulted in the use of the first 27 eigenvectors $\underset{\sim}{u}_i$ as a basis of the computed realizations. The figure (A1.1.1) contains the plots of seven eigenvectors of $\underset{\sim}{P}_o$ corresponding to the largest seven eigenvalues, for s = 5, $\rho = 0.5$, $R(0) = 0.25$.



E-VECTORS OF PRIOR COVAR MATRIX S=5

Figure A1.1.1

Appendix 1.2

Algorithm for computing gradient of J w. r. t. parameters

The minimization problem associated with the history matching requirement is,

$$\underset{\underset{\sim}{\pi}}{\text{Minimize}} \quad J = \sum_{m=1}^{R} \sum_{n=1}^{K} (p_{i_m, j_n} - p^{o}_{mn})^2 \tag{A1.2.1}$$

where the observed pressures are at R time instants corresponding to the time steps $\{i_m\}$ and K location corresponding to the grid points $\{j_n\}$. We denote in the above these observed pressures by $\{p^{o}_{mn}\}$ and the corresponding model outputs by $\{p_{i_m, j_n}\}$.

The above minimization is to be performed when $\{p_{i_m, j_n}\}$ are the resulting pressures from the discrete system equations,

$$\underset{\sim}{G}\, \underset{\sim}{p}_{i+1} = \underset{\sim}{H}\, \underset{\sim}{p}_i + \underset{\sim}{q}_i \qquad i = 0, 1, \ldots, T-1 \tag{A1.2.2}$$

$$\underset{\sim}{p}_o = \underset{\sim}{0}. \tag{A1.2.3}$$

In (A1.2.2) the matrices $\underset{\sim}{G}$ and $\underset{\sim}{H}$ are functions of the parameters $\underset{\sim}{\pi}$. Let us define a new performance index by adjoining the constraint equation (A1.2.2) to the original index in (A1.2.1),

$$\bar{J} = J + \sum_{i=0}^{T-1} \underset{\sim}{\psi}_i^T \left[\underset{\sim}{G}\, \underset{\sim}{p}_{i+1} - \underset{\sim}{H}\, \underset{\sim}{p}_i - \underset{\sim}{q}_i\right] \tag{A1.2.4}$$

where $\underset{\sim}{\psi}_i$, $i = 0, 1, 2, \ldots, T-1$ is a sequence of arbitrary N-vectors. We introduce a perturbation $\delta \underset{\sim}{\pi}$ in $\underset{\sim}{\pi}$, which induces the perturbations $\{\delta \underset{\sim}{p}_i\}$ and $\delta \bar{J}$.

$$\begin{aligned} \delta \bar{J} = {} & 2 \sum_{m=1}^{R} \sum_{n=1}^{K} (p_{i_m, j_n} - p^{o}_{mn}) \delta p_{i_m, j_n} \\ & + \sum_{i=0}^{T-1} \underset{\sim}{\psi}_i^T \left[\underset{\sim}{G}\, \delta \underset{\sim}{p}_{i+1} - \underset{\sim}{H}\, \delta \underset{\sim}{p}_i + \delta \underset{\sim}{G}\, \underset{\sim}{p}_{i+1} - \delta \underset{\sim}{H}\, \underset{\sim}{p}_i\right] \end{aligned} \tag{A1.2.5}$$

which can be rewritten as

$$\delta \bar{J} = \sum_{i=1}^{T} \left[\underset{\sim}{\psi}_{i-1}^{T} \underset{\sim}{G} + \underset{\sim}{\psi}_{i}^{T} \underset{\sim}{H} + 2 \sum_{n=1}^{K} (p_{i_m, j_n} - p^{o}_{mn}) \delta_{i, i_m} \underset{\sim}{e}_{j_n}^{T} \right] \delta \underset{\sim}{p}_{i}$$

$$+ \underset{\sim}{\psi}_{T}^{T} \underset{\sim}{H} \delta \underset{\sim}{p}_{T} + \sum_{i=0}^{T-1} \underset{\sim}{\psi}_{i}^{T} \left[\delta \underset{\sim}{G} \underset{\sim}{p}_{i+1} - \delta \underset{\sim}{H} \underset{\sim}{p}_{i} \right] \qquad \text{(A1.2.6)}$$

where $\underset{\sim}{e}_{j_n}$ is j_nth column of the NxN identity matrix, and δ_{i, i_m} is the Kronecker delta.

If we choose $\underset{\sim}{\psi}_{i}$ to satisfy

$$\underset{\sim}{G}^{T} \underset{\sim}{\psi}_{i-1} = \underset{\sim}{H}^{T} \underset{\sim}{\psi}_{i} + 2 \sum_{n=1}^{K} (p_{i_m, j_n} - p^{o}_{mn}) \delta_{i, i_m} \underset{\sim}{e}_{j_n} \qquad \text{(A1.2.7)}$$

$$i = T, T-1, \ldots, 2, 1$$

$$\underset{\sim}{\psi}_{T} = \underset{\sim}{0} \qquad \text{(A1.2.8)}$$

then

$$\delta \bar{J} = \sum_{i=0}^{T-1} \underset{\sim}{\psi}_{i}^{T} \left[\delta \underset{\sim}{G} \underset{\sim}{p}_{i+1} - \delta \underset{\sim}{H} \underset{\sim}{p}_{i} \right]$$

or

$$\frac{\partial \bar{J}}{\partial \pi_{\ell}} = \sum_{i=0}^{T-1} \underset{\sim}{\psi}_{i}^{T} \left[\frac{\partial \underset{\sim}{G}}{\partial \pi_{\ell}} \underset{\sim}{p}_{i+1} - \frac{\partial \underset{\sim}{H}}{\partial \pi_{\ell}} \underset{\sim}{p}_{i} \right] \qquad \text{(A1.2.9)}$$

We note that as long as $\underset{\sim}{p}_{i}$ are calculated to satisfy (A1.2.2-3) at any iteration, $J = \bar{J}$ and (A1.2.9) therefore gives the expression for the components of $\underset{\sim}{J}_{\pi}$.

To compute $\underset{\sim}{J}_{\pi}$ requires one integration of (A1.2.2) in the forward time direction followed by one integration of (A1.2.7) in the backward time direction, followed by evaluation of the summation (A1.2.9) once for each element of $\underset{\sim}{\pi}$.

In particular, when $\underset{\sim}{\pi} = (\underset{\sim}{k}^{T} | \underset{\sim}{\phi}^{T})^{T}$, expression (A1.2.9) gives, for one dimensional reservoir, using the definitions of $\underset{\sim}{G}$ and $\underset{\sim}{H}$ for the implicit scheme,

$$\frac{\partial \bar{J}}{\partial k_{\ell}} = - \frac{\Delta t}{2} \sum_{i=0}^{T-1} \underset{\sim}{\psi}_{i}^{T} \left[\frac{\partial \underset{\sim}{F}}{\partial k_{\ell}} \right] (\underset{\sim}{p}_{i+1} + \underset{\sim}{p}_{i}) \qquad \text{(A1.2.10)}$$

and

$$\frac{\partial J}{\partial \phi_\ell} = \sum_{i=0}^{T-1} \underset{\sim}{\psi}_i^T \left[\frac{\partial \underset{\sim}{G}}{\partial \phi_\ell}\right] (\underset{\sim}{p}_{i+1} - \underset{\sim}{p}_i) = \sum_{i=0}^{T-1} \psi_i^\ell (p_{i+1}^\ell - p_i^\ell)$$

$$\ell = 1.2.... N \qquad \text{(A1.2.11)}$$

where the superscript denotes the element of an N-vector.

Gradient of J w.r.t. the zonation parameters:

In case of zonation with M zones, the 2M-vector $\underset{\sim}{\pi} = [\pi_1^k, \pi_2^k \ldots \pi_M^k, \pi_1^\phi, \ldots, \pi_M^\phi]$ is the parameter vector. Then, for example, a single scalar permeability parameter, π_ℓ^k, determines the values of k at several contiguous grid points which belong to the ℓ^{th} zone. A variation $\delta\pi_\ell^k$ results in the variations of equal magnitudes in the values of k at all these grid points. Then, in (A1.2.9), we have,

$$\frac{\partial \underset{\sim}{G}}{\partial \pi_\ell^k} = \sum_{j \in \ell^{th}\,\text{zone}} \frac{\partial \underset{\sim}{G}}{\partial k_j} \qquad \text{(A1.2.12)}$$

$$\frac{\partial \underset{\sim}{H}}{\partial \pi_\ell^k} = \sum_{j \in \ell^{th}\,\text{zone}} \frac{\partial \underset{\sim}{H}}{\partial k_j} \qquad \text{(A1.2.13)}$$

and consequently,

$$\frac{\partial J}{\partial \pi_\ell^k} = \sum_{j \in \ell^{th}\,\text{zone}} \frac{\partial J}{\partial k_j} \qquad \text{(A.1.2.14)}$$

Similarly, we get the relation,

$$\frac{\partial J}{\partial \pi_\ell^\phi} = \sum_{j \in \ell^{th}\,\text{zone}} \frac{\partial J}{\partial \phi_j} \qquad \text{(A1.2.15)}$$

Gradient of J w.r.t. the Bayesian parameters:

In the Bayesian approach, it is convenient to expand the composite parameter vector $\underset{\sim}{\pi}^T = (\underset{\sim}{k}^T \,\vdots\, \underset{\sim}{\phi}^T)$ in the basis of orthonormal

set $\{\underset{\sim}{z}(i)\}$ of the eigenvectors of the prior covariance matrix $\underset{\sim}{P}_o$. Let,

$$\underset{\sim}{\pi} = \underset{\sim}{Z}\,\underset{\sim}{u} + \overline{\underset{\sim}{\pi}} \qquad \text{(A1.2.16)}$$

Then the vector $\underset{\sim}{u}$ is the parameter vector we wish to estimate, and the gradient $\frac{\partial J}{\partial \underset{\sim}{u}}$ is necessary for this purpose. Let us partition the matrix $\underset{\sim}{Z}$ as,

$$\underset{\sim}{Z}^T = [\underset{\sim}{Z}_k^T \,\vdots\, \underset{\sim}{Z}_\phi^T] \qquad \text{(A1.2.17)}$$

so that we have the N-vectors,

$$\underset{\sim}{k} = \underset{\sim}{Z}_k\,\underset{\sim}{u} + \overline{\underset{\sim}{k}} \qquad \text{(A1.2.18)}$$

$$\underset{\sim}{\phi} = \underset{\sim}{Z}_\phi\,\underset{\sim}{u} + \overline{\underset{\sim}{\phi}} \qquad \text{(A1.2.19)}$$

The chain rule of differentiation yields,

$$\begin{aligned} \frac{\partial J}{\partial \underset{\sim}{u}} &= \left(\frac{\partial \underset{\sim}{k}}{\partial \underset{\sim}{u}}\right)^T \frac{\partial J}{\partial \underset{\sim}{k}} + \left(\frac{\partial \underset{\sim}{\phi}}{\partial \underset{\sim}{u}}\right)^T \frac{\partial J}{\partial \underset{\sim}{\phi}} \\ &= \underset{\sim}{Z}_k^T \frac{\partial J}{\partial \underset{\sim}{k}} + \underset{\sim}{Z}_\phi^T \frac{\partial J}{\partial \underset{\sim}{\phi}} = \underset{\sim}{Z}^T \begin{bmatrix} \partial J/\partial \underset{\sim}{k} \\ \text{- - - -} \\ \partial J/\partial \underset{\sim}{\phi} \end{bmatrix} \end{aligned} \qquad \text{(A1.2.20)}$$

Thus (A1.2.9) can be used to compute the gradient for the Bayesian approach.

Appendix 1.3

Procedure for Unidirectional Search

The unidirectional search is required for the determination of a local minimum of J in the conjugate directions. A basic element in a search procedure is the search step size. It determines the total number of integrations of the system equations required for completing the search in a given iteration, and consequently determines the total computational time required by the estimation procedure employing the conjugate gradient algorithm. We allowed the program to determine automatically an appropriate search step as the minimization proceeded as follows. We started the minimization process with an assigned size for the search step, which was determined by experimentation. After the completion of the first four iterations, we set the search step size to be equal to the average of the lengths of the actual parameter vector corrections that were made during the last four iterations. If this value exceeded a predetermined limit, the search step size was set equal to the limit. These rules seemed to keep the total number of the integration of the system equations to a reasonably low value, usually at 4 per iteration, while given fairly accurate estimates of the minima.

The details of the actual search procedure are summarized in the flow chart in figure A1.3.1.

It may be noted that when it was found necessary to search in the interval enclosed by the starting point and the first step a bisection was used whereas for the search beyond the first step, uniform steps were used. In the former case, the distance between the three

Flow Chart for the Unidirectional Search

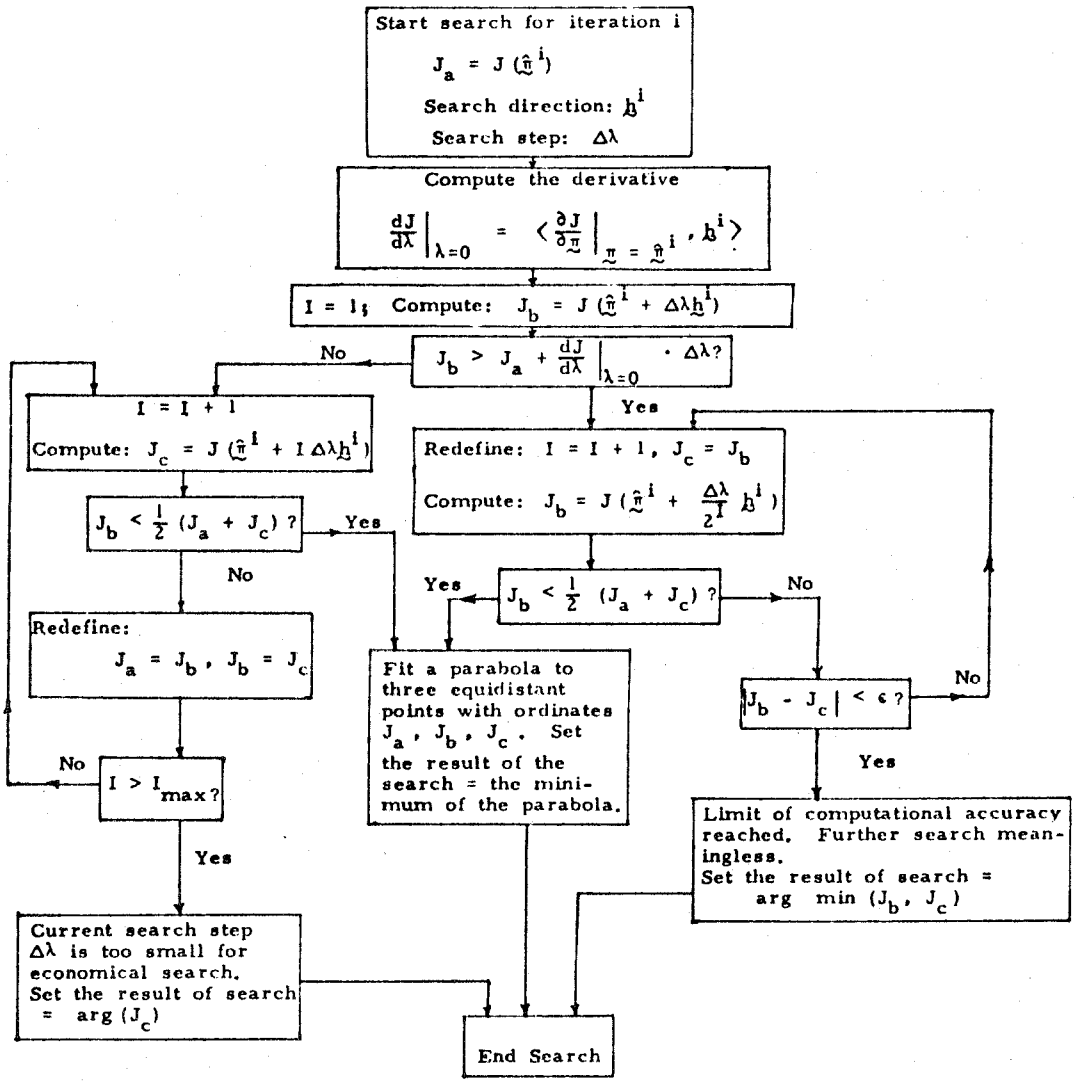


Figure A1.3.1

equidistant points is smaller than the search step size, guaranteeing an accurate determination of the minimum of J in the search direction. On the other hand, in the latter case, if the search step were to be doubled, the three consecutive points would no more be equidistant and at the same time the largest distance between the consecutive points may increase to unreasonably large values, yielding a parabola whose minimum is a poor approximation to the time value. These considerations led us to the strategy in the flow chart.

Appendix 1.4

Sensitivity of an Observation w.r.t. the Parameter Vector $\underset{\sim}{\pi}$

In this appendix, we derive the expressions for the sensitivity of the pressure at a given location and time, computed according to a given discrete scheme.

Method I. This is the discrete analog of the result first obtained by Jacquard and Jain (1965) for the o.d.e. formulation.

The discrete system equations are,

$$\underset{\sim}{G}\ \underset{\sim}{p}_{i+1} = \underset{\sim}{H}\ \underset{\sim}{p}_i + \underset{\sim}{q}_i \qquad i = 0, 1, \ldots, T-1 \tag{A1.4.1}$$

$$\underset{\sim}{p}_o = \underset{\sim}{0} \tag{A1.4.2}$$

Consider a variation $\delta\underset{\sim}{\pi}$ in $\underset{\sim}{\pi}$ with the resulting variations $\delta\underset{\sim}{G}$, $\delta\underset{\sim}{H}$ and $\{\delta\underset{\sim}{p}_i\}$. The first-order variational equations are

$$\underset{\sim}{G}\ \delta\underset{\sim}{p}_{i+1} = \underset{\sim}{H}\ \delta\underset{\sim}{p}_i - \delta\underset{\sim}{G}\ \underset{\sim}{p}_{i+1} + \delta\underset{\sim}{H}\ \underset{\sim}{p}_i \tag{A1.4.3}$$

with

$$\delta\underset{\sim}{p}_o = \underset{\sim}{0} \tag{A1.4.4}$$

Let us multiply (A1.4.3) by a sequence of arbitrary N-dimensional vectors $\{\underset{\sim}{\psi}_i\}$ and sum from i = 0 to $(i_m - 1)$ where i_m is the value of i corresponding to a data point. We obtain

$$\sum_{i=0}^{i_m-1} \underset{\sim}{\psi}_i^T\ \underset{\sim}{G}\ \delta\underset{\sim}{p}_{i+1} = \sum_{i=0}^{i_m-1} \underset{\sim}{\psi}_i^T\ \underset{\sim}{H}\ \delta\underset{\sim}{p}_i - \sum_{i=0}^{i_m-1} \underset{\sim}{\psi}_i^T\ \delta\underset{\sim}{G}\ \underset{\sim}{p}_{i+1} + \sum_{i=0}^{i_m-1} \underset{\sim}{\psi}_i^T\ \delta\underset{\sim}{H}\ \underset{\sim}{p}_i \tag{A1.4.5}$$

Rearranging (A1.4.5) and using (A1.4.4) yields

$$\sum_{i=1}^{i_m-1} \left[\underset{\sim}{\psi}_{i-1}^T\ \underset{\sim}{G} - \underset{\sim}{\psi}_i^T\ \underset{\sim}{H}\right] \delta\underset{\sim}{p}_i = -\sum_{i=0}^{i_m-1} \underset{\sim}{\psi}_i^T\ \delta\underset{\sim}{G}\ \underset{\sim}{p}_{i+1} + \sum_{i=1}^{i_m-1} \underset{\sim}{\psi}_i^T\ \delta\underset{\sim}{H}\ \underset{\sim}{p}_i - \underset{\sim}{\psi}_{i_m-1}^T\ \underset{\sim}{G}\ \delta\underset{\sim}{p}_{i_m} \tag{A1.4.6}$$

Let us choose the sequence $\underset{\sim}{\psi}_i$, $i = 0, 1, 2, \ldots, i_m - 1$ to satisfy

$$\underset{\sim}{G}^T \underset{\sim}{\psi}_{i-1} = \underset{\sim}{H}^T \psi_i \quad i = 0, 1, 2, \ldots, i_m - 1 \tag{A1.4.7}$$

$$\underset{\sim}{G}^T \underset{\sim}{\psi}_{i_m - 1} = \underset{\sim}{e}_{j_n} \tag{A1.4.8}$$

where $\underset{\sim}{e}_{j_n}$ is the j_nth column of I_{NxN}. Using (A1.4.7) and (A1.4.8), (A1.4.6) becomes,

$$\delta p_{i_m, j_n} = - \sum_{i=0}^{i_m - 1} \underset{\sim}{\psi}_i^T \delta \underset{\sim}{G} \underset{\sim}{p}_{i+1} + \sum_{i=1}^{i_m - 1} \underset{\sim}{\psi}_i^T \delta \underset{\sim}{H} \underset{\sim}{p}_i \tag{A1.4.9}$$

Then, we have,

$$\frac{\partial p_{i_m, j_n}}{\partial \pi_\ell} = - \sum_{i=0}^{i_m - 1} \underset{\sim}{\psi}_i^T \frac{\partial \underset{\sim}{G}}{\partial \pi_\ell} \underset{\sim}{p}_{i+1} + \sum_{i=1}^{i_m - 1} \underset{\sim}{\psi}_i^T \frac{\partial \underset{\sim}{H}}{\partial \pi_\ell} \underset{\sim}{p}_i \tag{A1.4.10}$$

Note that it is necessary to solve the adjoint system (A1.4.7) and (A1.4.8) separately for each measurement location, j_n, $n = 1, 2, \ldots, K$. For different observation times, i_m, $m = 1, 2, \ldots, R$, for a given location, the same solution of the adjoint system can be used by merely redefining the subscripts. This requires that the adjoint system be integrated with $i_m = T$ once for each of the K observation locations. The summation in (A1.4.10)has to be evaluated separately for each value of i_m and j_n.

For the one dimensional reservoir, recalling the definitions of $\underset{\sim}{G}$ and $\underset{\sim}{H}$,

$$\frac{\partial p_{i_m, j_n}}{\partial k_\ell} = - \frac{\Delta t}{2} \sum_{i=0}^{i_m - 1} \underset{\sim}{\psi}_i^T \left(\frac{\partial \underset{\sim}{F}}{\partial k_\ell} \right) (\underset{\sim}{p}_{i+1} + \underset{\sim}{p}_i) \tag{A1.4.11}$$

$$\frac{\partial p_{i_m, j_n}}{\partial \phi_\ell} = \sum_{i=0}^{i_m - 1} \psi_i^\ell \, (p_{i+1}^\ell - p_i^\ell) \tag{A1.4.12}$$

where the superscript denotes the element of an N-vector.

The sensitivities with respect to the zonation parameters can be obtained, as described in appendix (1.2), as

$$\frac{\partial p_{i_{m,j_n}}}{\partial \pi_\ell^k} = \sum_{i \in \ell^{th} \text{ zone}} \frac{\partial p_{i_{m,j_n}}}{\partial k_j} \qquad \text{(A1.4.13)}$$

$$\frac{\partial p_{i_{m,j_n}}}{\partial \pi_\ell^\phi} = \sum_{i \in \ell^{th} \text{ zone}} \frac{\partial p_{i_{m,j_n}}}{\partial \phi_j} \qquad \text{(A1.4.14)}$$

The sensitivities with respect to the Bayesian parameters can be similarly obtained as

$$\frac{\partial p_{i_{m,j_n}}}{\partial \underset{\sim}{u}} = Z^T \left[\frac{\partial p_{i_{m,j_n}} / \partial \underset{\sim}{k}}{\partial p_{i_{m,j_n}} / \partial \underset{\sim}{\phi}} \right] \qquad \text{(A1.4.15)}$$

Method II. Sensitivity equations

An alternative method of computing the sensitivities is to solve directly a set of discrete sensitivity equations for each of the parameters.

From the equations (A1.4.3) and (A1.4.4) we obtain the equations for the sensitivities of the pressures w.r.t. the parameter π_ℓ,

$$\underset{\sim}{G}\left[\frac{\partial \underset{\sim}{p}_{i+1}}{\partial \pi_\ell}\right] = \underset{\sim}{H}\left[\frac{\partial \underset{\sim}{p}_i}{\partial \pi_\ell}\right] - \frac{\partial \underset{\sim}{G}}{\partial \pi_\ell} \underset{\sim}{p}_{i+1} + \frac{\partial \underset{\sim}{H}}{\partial \pi_\ell} \underset{\sim}{p}_i \qquad \text{(A1.4.16)}$$

$$\left[\partial \underset{\sim}{p}_o / \partial \pi_\ell\right] = \underset{\sim}{0} \qquad \text{(A1.4.17)}$$

For obtaining the set $\left\{ \frac{\partial p_{i_{m,j_n}}}{\partial \pi_\ell} \right\}$, equations (A1.4.16) and (A1.4.17) need to be solved once for each parameter π_ℓ and then the sensitivities of the desired pressures can be picked out. The sensitivities with respect to the zonation and Bayesian parameters can be obtained according to (A1.4.13, 14, 15).

Appendix 1.5

Gauss-Newton Algorithm and Marquardt's Method

We have the minimization problem,

$$\underset{\underset{\sim}{\pi}}{\text{Min}} \quad J = \sum_{m=1}^{R} \sum_{n=1}^{K} \left(p_{i_m, j_n} - p^{o}_{m,n} \right)^2 \tag{A1.5.1}$$

subject to the system equations.

A necessary condition for a minimum is,

$$\frac{\partial J}{\partial \underset{\sim}{\pi}} = 2 \sum_{m=1}^{R} \sum_{n=1}^{K} \left(p_{i_m, j_n} - p^{o}_{m,n} \right) \frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} = \underset{\sim}{0} \tag{A1.5.2}$$

The second derivative of J w.r.t. $\underset{\sim}{\pi}$ is,

$$\frac{\partial^2 J}{\partial \underset{\sim}{\pi} \, \partial \underset{\sim}{\pi}} = 2 \sum_{m=1}^{R} \sum_{n=1}^{K} \left(p_{i_m, j_n} - p^{o}_{m,n} \right) \frac{\partial^2 p_{i_m, j_n}}{\partial \underset{\sim}{\pi} \, \partial \underset{\sim}{\pi}} + 2 \sum_{m=1}^{R} \sum_{n=1}^{K} \left(\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} \right) \left(\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} \right)^{T} \tag{A1.5.3}$$

The Gauss-Newton approximation involves the assumption that the current estimate of $\underset{\sim}{\pi}^{\ell}$ is close to the true value, yielding a small history matching error $(p_{i_m, j_n} - p^{o}_{m,n})$ for all of the observations. Then we may neglect the first term in (A1.5.3) obtaining the approximation,

$$\frac{\partial^2 J}{\partial \underset{\sim}{\pi} \, \partial \underset{\sim}{\pi}} \approx 2 \sum_{m=1}^{R} \sum_{n=1}^{K} \left(\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} \right) \left(\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}} \right)^{T} \tag{A1.5.4}$$

Then, the updated estimate $\underset{\sim}{\pi}^{\ell+1}$ is obtained from the current estimate $\underset{\sim}{\pi}^{\ell}$ using (A1.5.2), (A1.5.4) and the truncated Taylor series.

$$\left. \frac{\partial J}{\partial \underset{\sim}{\pi}} \right|_{\underset{\sim}{\pi}^{\ell+1}} = \left. \frac{\partial J}{\partial \underset{\sim}{\pi}} \right|_{\underset{\sim}{\pi}^{\ell}} + \left(\frac{\partial^2 J}{\partial \underset{\sim}{\pi} \, \partial \underset{\sim}{\pi}} \right)_{\underset{\sim}{\pi}^{\ell}} \left(\underset{\sim}{\pi}^{\ell+1} - \underset{\sim}{\pi}^{\ell} \right) = \underset{\sim}{0} \tag{A1.5.5}$$

whence,

$$\underset{\sim}{\pi}^{\ell+1} = \underset{\sim}{\pi}^{\ell} - \left(\frac{\partial^2 J}{\partial \underset{\sim}{\pi}\, \partial \underset{\sim}{\pi}}\right)^{-1}_{\pi_\ell} \left.\frac{\partial J}{\partial \underset{\sim}{\pi}}\right|_{\underset{\sim}{\pi}^{\ell}}$$

$$= \underset{\sim}{\pi}^{\ell} - \left[2 \sum_{m,n} \left(\frac{\partial p_{i_{m,j_n}}}{\partial \underset{\sim}{\pi}}\right)\left(\frac{\partial p_{i_{m,j_n}}}{\partial \underset{\sim}{\pi}}\right)^T \Bigg|_{\underset{\sim}{\pi}^{\ell}}\right]^{-1} \left(\frac{\partial J}{\partial \underset{\sim}{\pi}}\right)_{\underset{\sim}{\pi}^{\ell}} \quad \text{(A1.5.6)}$$

When the matrix to be inverted in (A1.5.6) has one or more very small eigenvalues, the corrections in the parameter estimates become large and the iterative process fails to converge. Marquardt's modification alleviates this difficulty by imposing an additional constraint that the correction in parameter estimate have a specified length. Denoting the correction in the parameter estimate by $\Delta\underset{\sim}{\pi}$, the constraint is,

$$\Delta\underset{\sim}{\pi}^T \Delta\underset{\sim}{\pi} = r^2 = \text{constant} \quad \text{(A1.5.7)}$$

Then, our minimization problem can be more expediently posed as the minimization of J in (1.5.1) with respect to $\Delta\underset{\sim}{\pi}$. Adjoining the constraint (A1.5.7) using a scalar Legrange multiplier μ, our problem becomes,

$$\underset{\Delta\underset{\sim}{\pi}}{\text{Min}}\ \bar{J} = J + \mu(\Delta\underset{\sim}{\pi}^T \Delta\underset{\sim}{\pi} - r^2) \quad \text{(A1.5.8)}$$

Then application of the Gauss-Newton procedure to (A1.5.8) yields

$$\Delta\underset{\sim}{\pi} = -\left[2 \sum_{m,n} \left(\frac{\partial p_{i_{m,j_n}}}{\partial \underset{\sim}{\pi}}\right)\left(\frac{\partial p_{i_{m,j_n}}}{\partial \underset{\sim}{\pi}}\right)^T \Bigg|_{\underset{\sim}{\pi}^{\ell}} + \mu \underset{\sim}{I}\right]^{-1} \left(\frac{\partial J}{\partial \underset{\sim}{\pi}}\right)_{\pi^{\ell}} \quad \text{(A1.5.9)}$$

Thus the direct consequence of Marquardt's modification is to strengthen the positive definiteness of the Hessian matrix by addition of the diagonal matrix $\mu \underset{\sim}{I}$. Clearly, $\mu > 0$ for this purpose.

To understand the effect of Marquardt's modification, note that as $\mu \to 0$, (A1.5.9) approaches the result (A1.5.6) of the Gauss-Newton method. On the other hand, as $\mu \to \infty$, the term $\mu \underset{\sim}{I}$ dominates the term approximating the second derivative and the corrections become smaller and smaller in length and parallel to the gradient $(\frac{\partial J}{\partial \underset{\sim}{\pi}})_{\underset{\sim}{\pi}} \ell$. Thus for large values of μ , the Marquardt's method behaves in a manner similar to the gradient algorithm of minimization. For intermediate values of μ , Marquardt's method may be viewed as a method intermediate between the Gauss-Newton method and the gradient method.

The question now remains as to what is an optimum value of μ. In principle this question can be answered by carrying out a unidirectional search. However, such a process is likely to be very expensive as each search step would involve the inversion of a full matrix and an integration of the system equation to compute the resulting value of J. In addition, this search will have to be carried out at each iteration step. In our application, therefore, we used a preselected fixed value of μ for each of the estimation attempts.

The inversion of the matrix in (A1.5.9) needs to be accurately carried out, because the accuracy determines the rate of convergence of the minimization algorithm. In addition, obtaining $\partial p_{i_m, j_n} / \partial \underset{\sim}{\pi}$ for all m, n dominates the total computational effort. So it is worthwhile to expend more effort in the inversion and to minimize the number of iterations. With this in mind, we used the diagonalization of the matrix by its decomposition into eigenvalue and eigenvectors for the purpose of inversion, using a modification of the program by Bushinger and Golub (1969).

Appendix 1.6

Probability density of the irreducible value of J

For the "true" estimates, J will have some irreducible value J^* due to the observation errors. When the observation errors are independent, Gaussian with zero mean and uniform variance, the irreducible J^*, being the sum of their squares is χ^2-distributed

$$J^* = \sum_{m=1}^{R} \sum_{n=1}^{K} \left(p_{i_m, j_n} - p^{o}_{m,n} \right) = \sum_{i=1}^{N_c} \chi_1^2 \qquad \text{(A1.6.1)}$$

where N_o = total number of observation and

$$\chi_i = N(0, \sigma^2) \qquad \text{(A1.6.2)}$$

Then, with $J^* > 0$,

$$p(J^*) = \frac{1}{2^{N_o/2} \sigma^{N_o} \Gamma(\frac{N_o}{2})} J^{*\frac{N_o-2}{2}} e^{-J^*/2\sigma^2} \qquad \text{(A1.6.3)}$$

(Papoulis, 1965)

For $\sigma = 1$, the probability Q, that $J^* > \chi^2$ for $N_o = 70, 80$ are as follows (Abramowitz and Stegun, 1968):

N_o	Q →	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25
70	χ^2 =	43.27	45.44	48.76	51.74	55.33	61.70	69.33	77.58
80	χ^2 =	51.17	53.54	57.15	60.93	64.28	71.14	79.33	88.13

For $N_o = 70$, probability that $61.70 < J^* < 77.58$ is 0.5 and the probabilities are 0.25 for the events, $J^* > 77.58$ and $J^* < 61.70$. Then, for 70 observations with unit variance of observation errors, there

Appendix 1.7

Computational Effort for Gradient and Gauss-Newton Algorithms

In the following, only the number of multiplicative operations (abbreviated as "mults") are reported, the number of additions being approximately equal. Furthermore, we shall estimate the computational effort per iteration in each case, the actual number of iterations required for minimization being kept a variable.

We shall analyze two minimization algorithms, the first order gradient algorithm and the Gauss-Newton algorithm. We shall consider two types of parameterization, the zonation and the Bayesian. In each of the four cases, we shall report for both the one-dimensional and two-dimensional reservoirs the leading-order estimates of the computational effort for the discrete formulation of the problem, the algorithms for which are derived in appendices 1.2 and 1.4.

I. Gradient Algorithm

The numerical effort per iteration for the gradient algorithm with search and the conjugate gradient algorithm are almost identical. For either, the number of simulations (solutions of the pressure equation) per iteration is:

1	pressure	$\{\underset{\sim}{p}_i\}$
1	adjoint	$\{\underset{\sim}{\psi}_i\}$
S	unidirectional search for determining correction	
S+2	total	

The value of s depends on the search algorithm used and the degree of accuracy of the search for the minimum of J along the gradient or conjugate direction. A typical value of s for the procedure in appendix 1.3 is 4.

Determination of Gradient: The effort in determining $\partial J/\partial \underset{\sim}{\pi}$ depends on the number of time steps, T, and the number of grid points, N. In addition, it also depends to some extent on whether the reservoir is one-dimensional (abbreviated as "1-d") or two-dimensional (referred to as "2-d" in the sequel).

(a) 1-d reservoir. The evaluation $\partial J/\partial \underset{\sim}{k}$ can be most economically carried out by computing the necessary elements of the matrix $\sum_{i=1}^{T-1} \underset{\sim}{\psi}_i (\underset{\sim}{p}_{i+1}^T + \underset{\sim}{p}_i^T)$ and subsequently computing their linear combinations occuring in (A1.2.10). For a 1-d reservoir the matrix F is tridiagonal, necessitating the computation of only three principal diagonals of the summation. This requires 3NT mults. The evaluation of $\partial J/\partial \underset{\sim}{\phi}$ according to (A1.2.11) needs additional NT mults. Thus, total effort is 4NT mults.

(b) 2-d reservoir. For a two-dimensional reservoir, the pressure and the adjoint equations of which are solved by the alternating direction implicit (ADI) method, the determination of the number of multiplicative operations is more involved. In order to obtain an accurate estimate of this, we shall work out the structure of the formulation, considering the minimum amount of detail.

For simplicity, let the reservoir be rectangular and be divided into a uniform IxJ (IJ = N) grid. We denote any quantity associated with a grid point by the indices (i, j) and the time value by the superscript ℓ. Then, the application of the ADI method to the pressure equation yields the difference equations

$$\phi_{i,j}\, p_{i,j}^{2\ell+1} - \phi_{i,j}\, p_{i,j}^{2\ell} = a_{i-1,j}\, p_{i-1,j}^{2\ell+1} + a_{i,j}\, p_{i,j}^{2\ell+1} + a_{i+1,j}\, p_{i+1,j}^{2\ell+1} + b_{i,j-1}\, p_{i,j-1}^{2\ell} + c_{i,j}\, p_{i,j}^{2\ell} + d_{i,j+1}\, p_{i,j+1}^{2\ell} \qquad \text{(A1.7.1)}$$

$$\phi_{i,j}\, p_{i,j}^{2\ell+2} - \phi_{i,j}^{2\ell+1}\, p_{i,j}^{2\ell+1} = e_{i,j-1}\, p_{i,j-1}^{2\ell+2} + f_{i,j}\, p_{i,j}^{2\ell+2} + g_{i,j+1}\, p_{i,j+1}^{2\ell+2}$$

$$+ h_{i-1,j} p_{i-1,j}^{2\ell+1} + \mathbf{h}_{i,j} p_{i,j}^{2\ell+1} + h_{i+1,j}\, p_{i+1,j}^{2\ell+1} \qquad \text{(A1.7.2)}$$

where $\{\phi_{i,j}\}$ are the grid porosities and the scalars $\{a_{i,j}, \ldots, h_{i,j}\}$ are linear combinations of the grid permeabilities $\{k_{i,j}\}$. For simplicity, the source terms have been dropped from (A1.7.1-2).

Define the I-vectors

$$\underset{\sim}{p}_j^T = [p_{1,j}, p_{2,j}, \ldots, p_{N,j}] \quad j = 1, 2 \ldots, J \qquad \text{(A1.7.3)}$$

Then (A1.7.1-2) can be written as,

$$\underset{\sim}{\Phi}_j \left(\underset{\sim}{p}_j^{2\ell+1} - \underset{\sim}{p}_j^{2\ell} \right) = \underset{\sim}{A}_i \underset{\sim}{p}_j^{2\ell+1} + \underset{\sim}{B}_{j-1} \underset{\sim}{p}_{j-1}^{2\ell} + \underset{\sim}{C}_j \underset{\sim}{p}_j^{2\ell} + \underset{\sim}{D}_{j+1} \underset{\sim}{p}_{j+1}^{2\ell} \qquad \text{(A1.7.4)}$$

$$\underset{\sim}{\Phi}_j \left(\underset{\sim}{p}_j^{2\ell+1} - \underset{\sim}{p}_j^{2\ell} \right) = \underset{\sim}{E}_{j-1} \underset{\sim}{p}_{j-1}^{2\ell+2} + \underset{\sim}{F}_j \underset{\sim}{p}_j^{2\ell+2} + \underset{\sim}{G}_{j+1} \underset{\sim}{p}_j^{2\ell+2} + \underset{\sim}{H}_j \underset{\sim}{p}_j^{2\ell+1} \qquad \text{(A1.7.5)}$$

The matrices $\{\underset{\sim}{\Phi}_j\}$ $j = 1, 2, \ldots, J$ are diagonal, $\{\underset{\sim}{H}_j\}$ and $\{\underset{\sim}{A}_j\}$ are tridiagonal, and each of the remaining matrices have only a single nonzero element in each row.

If we define the N-vector,

$$\underset{\sim}{p}^T = \left[\underset{\sim}{p}_1^T, \underset{\sim}{p}_2^T, \ldots, \underset{\sim}{p}_J^T \right]$$

then (A1.7.4-5) can be written in the form,

$$\underset{\sim}{\Phi} (\underset{\sim}{p}^{2\ell+2} - \underset{\sim}{p}^{2\ell}) = \underset{\sim}{\bar{A}}\, \underset{\sim}{p}^{2\ell+1} - \underset{\sim}{X}\, \underset{\sim}{p}^{2\ell} \qquad \text{(A1.7.6)}$$

$$\underset{\sim}{\Phi}(\underset{\sim}{p}^{2\ell+2} - \underset{\sim}{p}^{2\ell+1}) = \underset{\sim}{H}\,\underset{\sim}{p}^{2\ell+1} + \underset{\sim}{Y}\,\underset{\sim}{p}^{2\ell+2} \qquad \ell = 0, 1, 2, .., T/2 \tag{A1.7.7}$$

where

$$\underset{\sim}{\Phi} = \begin{bmatrix} \underset{\sim}{\Phi}_1 & & 0 \\ & \underset{\sim}{\Phi}_2 \ddots & \\ 0 & & \underset{\sim}{\Phi}_J \end{bmatrix}, \quad \underset{\sim}{\bar{A}} = \begin{bmatrix} \underset{\sim}{A}_1 & & 0 \\ & \underset{\sim}{A}_2 \ddots & \\ 0 & & \underset{\sim}{A}_J \end{bmatrix}$$

$$\underset{\sim}{H} = \begin{bmatrix} \underset{\sim}{H}_1 & & \\ & \underset{\sim}{H}_2 & 0 \\ 0 & \ddots & \underset{\sim}{H}_J \end{bmatrix}, \quad \underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{C}_1 & \underset{\sim}{D}_2 \ddots & 0 \\ \underset{\sim}{B}_1 & \underset{\sim}{C}_2 \ddots & \underset{\sim}{D}_J \\ 0 & \underset{\sim}{B}_{J-1} & \underset{\sim}{C}_J \end{bmatrix} \tag{A1.7.8}$$

$$\underset{\sim}{Y} = \begin{bmatrix} \underset{\sim}{F}_1 & \underset{\sim}{G}_2 \ddots & 0 \\ \underset{\sim}{E}_1 \ddots & \underset{\sim}{F}_2 \ddots & \underset{\sim}{G}_J \\ 0 & \underset{\sim}{E}_{J-1} & \underset{\sim}{F}_J \end{bmatrix}$$

The matrix $\underset{\sim}{\Phi}$ is diagonal, $\underset{\sim}{A}$ and $\underset{\sim}{H}$ are almost tri-diagonal and each of $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ has 3 nonzero elements in almost each row.

We obtain the modified performance index

$$\begin{aligned} \bar{J} = J &+ \sum_{\ell=0}^{T/2} \underset{\sim}{\psi}^{2\ell+1\,T} \left[\underset{\sim}{\Phi}(\underset{\sim}{p}^{2\ell+1} - \underset{\sim}{p}^{2\ell}) - \underset{\sim}{\bar{A}}\,\underset{\sim}{p}^{2\ell+1} - \underset{\sim}{X}\underset{\sim}{p}^{2\ell} \right] \\ &+ \sum_{\ell=0}^{T/2} \underset{\sim}{\psi}^{2\ell+2} \left[\underset{\sim}{\Phi}(\underset{\sim}{p}^{2\ell+2} - \underset{\sim}{p}^{2\ell+1}) - \underset{\sim}{H}\,\underset{\sim}{p}^{2\ell+1} - \underset{\sim}{Y}\underset{\sim}{p}^{2\ell+2} \right] \end{aligned} \tag{A1.7.9}$$

The usual variational procedure will yield a set of difference equations for the adjoint variable $\{\underset{\sim}{\psi}^{\ell}\}$. When the adjoint and system equations are satisfied, we have the first-order variational relation

$$\delta J = \sum_{\ell=0}^{T/2} \psi^{2\ell+1\,T}\left[\delta\underset{\sim}{\Phi}(\underset{\sim}{p}^{2\ell+1}-\underset{\sim}{p}^{2\ell}) - \delta\underset{\sim}{\bar{A}}\,\underset{\sim}{p}^{2\ell+1} - \delta\underset{\sim}{X}\,\underset{\sim}{p}^{2\ell}\right]$$

$$+ \sum_{\ell=0}^{T/2} \psi^{2\ell+2}\left[\delta\underset{\sim}{\Phi}(\underset{\sim}{p}^{2\ell+2}-\underset{\sim}{p}^{2\ell+1}) - \delta\,\underset{\sim}{H}\,\underset{\sim}{p}^{2\ell+1}\right.$$

$$\left. - \delta\,\underset{\sim}{Y}\,\underset{\sim}{p}^{2\ell+2}\right] \qquad \text{(A1.7.10)}$$

From the last expression, the operation count for computing $\partial J/\partial\underset{\sim}{k}$ and $\partial J/\partial\underset{\sim}{\phi}$ can be easily carried out. These gradients are most economically computed by evaluating the required elements of the summations

$$\sum_{\ell=0}^{T/2} \psi^{2\ell+1}\,\underset{\sim}{p}^{2\ell+1+m\,T} \quad \text{and} \quad \sum_{\ell=0}^{T/2} \underset{\sim}{\psi}^{2\ell}\,\underset{\sim}{p}^{2\ell+m\,T} \qquad \text{for}$$

appropriate values of m, and forming their necessary linear combinations. As each of the matrices $\underset{\sim}{\bar{A}}$, $\underset{\sim}{H}$, $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ has approximately 3N nonzero elements, these sums require

$$4 \times 3NT/2 = 6\,NT \qquad \text{mults.}$$

The terms containing $\delta\underset{\sim}{\Phi}$ do not require any additional computation because the required diagonal terms are already included in the above.

Thus, the evaluation of the expressions for $\partial J/\partial\underset{\sim}{k}$ and $\partial J/\partial\underset{\sim}{\phi}$ requires for a 2-d reservoir,

$$6\,NT + O(N) \qquad \text{mults}$$

when discrete formulation for A. D. I. method is implemented.

For both 1-d and 2-d reservoirs, the determination of $\partial J/\partial\underset{\sim}{\pi}$ from $\partial J/\partial\underset{\sim}{k}$ and $\partial J/\partial\underset{\sim}{\phi}$ requires only O(N) summations in the zonation approach. In the Bayesian approach, this requires the product of (M x 2N) matrix with a (2N x 1) vector, where M is the number of

parameters being estimated. This requires 2MN mults. The rest of the computational steps need O(N, M, T) multiplications.

Each simulation requires approximately 6NT multiplications, when the implicit scheme is used for a 1-d reservoir or the A. D. I. scheme is used for a 2-d reservoir.

Thus, the final estimate for the computational effort per iteration of the conjugate gradient method is,

$$\boxed{6(S+2)NT + \alpha NT + \beta MN + O(M, N, T) \text{ mults}}$$

where $$\alpha = \begin{cases} 4 & \text{1-d reservoir} \\ 6 & \text{2-d reservoir} \end{cases}$$

$$\beta = \begin{cases} 0 & \text{Zonation} \\ 2 & \text{Bayesian approach} \end{cases}$$

II. Gauss-Newton (or Marquardt) Method:

In this method, the quantity $\frac{\partial p_{i_m, j_n}}{\partial \underset{\sim}{\pi}}$ must be computed for $m = 1, 2, \ldots, R$ and $n = 1, 2, \ldots, K$, which determine the time and location of the observations. The total computational per iteration for this method, therefore, depends on the scheme used for computing the sensitivities.

(i) Adjoint Equation method for sensitivity calculation.

For this method, since the adjoint equation has to be integrated separately for each observation location, and since the convolution sum needs to be evaluated separately for each of the observations, the computational effort depends on,

L - the total number of observations

K - the number of observation sites.

In addition, it also depends on the instants when the different observations are obtained. Note that when the pressures at all the K sites are recorded simultaneously for R instants, as assumed in the previous discussions, L = KR. However, we shall assume a more general observation strategy, where there are a total of L observations taken at K different sites. Furthermore, we shall assume for simplicity of analysis that the different observations at any given site are taken at uniform intervals over the observed history period. Then (K+1) simulations are required for the adjoint equations and the original system equation. The sums for the sensitivities of each of the observations w. r. t. $\underset{\sim}{k}$ and $\underset{\sim}{\phi}$ are analogous to those for $\partial J/\partial \underset{\sim}{k}$ and $\partial J/\partial \underset{\sim}{\phi}$, with a difference that now they have to be evaluated for an average time period of T/2. Then the effort for the sensitivity computations can be deduced from that for the gradient computation in the conjugate gradient method. Thus the computation of the (Lx2N) sensitivity matrix requires,

$4 \times NTL/2$ mults for a 1-d reservoir

$6 \times NTL/2$ mults for a 2-d reservoir.

In addition, for the Bayesian approach, the Bayesian parameter sensitivity matrix is obtained from the foregoing (Lx2N) matrix by post-multiplication with (2NxM) matrix Z. This operation needs 2NLM multiplications. The corresponding operation in the zonation approach involves only O(LN) summations.

Thus, total effort for computing the appropriate sensitivity matrix is,

$$6(K+1)NT + \alpha NTL + \beta NLM \text{ multiplications, where}$$

$$\alpha = \begin{cases} 2 & \text{1-d reservoir} \\ 3 & \text{2-d reservoir} \end{cases}$$

$$\beta = \begin{cases} 0 & \text{Zonation} \\ 2 & \text{Bayesian approach.} \end{cases}$$

(ii) Difference Equations Method for Sensitivity Calculations

(a) 1-d reservoir.

Differentiating the system equations (A1.2.2-3) with the parameter π_ℓ, we obtain the sensitivity equations

$$\underset{\sim}{G} \frac{\partial \underset{\sim}{p}_{j+1}}{\partial \pi_\ell} = \underset{\sim}{H} \frac{\partial \underset{\sim}{p}_i}{\partial \pi_\ell} - \left(\frac{\partial \underset{\sim}{G}}{\partial \pi_\ell}\right) \underset{\sim}{p}_{i+1} + \left(\frac{\partial \underset{\sim}{H}}{\partial \pi_\ell}\right) \underset{\sim}{p}_i \qquad i = 1, 2, \ldots \qquad \text{(A1.7.11)}$$

$$\frac{\partial \underset{\sim}{p}_i}{\partial \pi_\ell} = \underset{\sim}{0}. \qquad \text{(A1.7.12)}$$

For each parameter π_ℓ, a separate set of such discrete equations has to be solved. The homogeneous part of each set is identical to the original pressure equations. The forcing terms induce some additional work for the solution. This additional work is different for different parameterizations.

Zonation: Assume M/2 zones, implying M parameters. Each of the M/2 permeability parameters affect on the average (6N/M+4) elements in each of the matrices $\underset{\sim}{G}$ and $\underset{\sim}{H}$. Each of the M/2 porosity parameters affect, on the average, 2N/M diagonal elements in $\underset{\sim}{G}$ and $\underset{\sim}{H}$. Thus, the solution of the M sets of sensitivity equations requires,

including 6NT multiplications for the solution of the pressure equation

$6NT + M(6NT + 8NT/M+4T) = (6M + 14)NT + 4MT$ mults.

Bayesian Approach: Each of the M Bayesian parameters influences the values of k and ϕ at all of the N grid points. This leads to tri-diagonal matrices in both the forcing terms for all the parameters. The evaluation of the forcing terms requires 6NT multiplications for each set of equations. The total effort for M sets of sensitivities is,

$6\,NT + M(6\,NT + 6\,NT) = 6(2M + 1)NT$ mults.

Once the coefficients of the forcing terms are appropriately computed, the resulting sensitivities are w.r.t. the Bayesian parameters, and no more matrix multiplication is necessary, saving 2MNL multiplicative operations.

(b) 2-d reservoir: The discrete sensitivity equations can be derived from the A.D.I. system equations (A1.7.4) and (A1.7.5) by differentiation w.r.t. π_ℓ.

$$\underset{\sim}{\Phi}_j \left(\frac{\partial \underset{\sim}{p}_j^{2\ell+1}}{\partial \pi_\ell} - \frac{\partial \underset{\sim}{p}_j^{2\ell}}{\partial \pi_\ell}\right) = \underset{\sim}{A}_j \frac{\partial \underset{\sim}{p}_j^{2\ell+1}}{\partial \pi_\ell} + \underset{\sim}{B}_{j-1} \frac{\partial \underset{\sim}{p}_j^{2\ell}}{\partial \pi_\ell} + \underset{\sim}{C}_j \frac{\partial \underset{\sim}{p}_j}{\partial \pi_\ell} + \underset{\sim}{D}_{j+1} \frac{\partial \underset{\sim}{p}_{j+1}^{2\ell}}{\partial \pi_\ell}$$

$$- \frac{\partial \underset{\sim}{\Phi}_j}{\partial \pi_\ell}\left(\underset{\sim}{p}_j^{2\ell+1} - \underset{\sim}{p}_j^{2\ell}\right) + \frac{\partial \underset{\sim}{A}_j}{\partial \pi_\ell} \underset{\sim}{p}_j^{2\ell+1} + \frac{\partial \underset{\sim}{B}_{j-1}}{\partial \pi_\ell} \underset{\sim}{p}_{j-1}^{2\ell}$$

$$+ \frac{\partial \underset{\sim}{C}_j}{\partial \pi_\ell} \underset{\sim}{p}_j^{2\ell} + \frac{\partial \underset{\sim}{D}_{j+1}}{\partial \pi_\ell} \underset{\sim}{p}_{j+1}^{2\ell} \qquad \text{(A1.7.13)}$$

and a similar equation for the next time step, implicit in the other direction. These equations differ from the pressure equations only in the forcing terms. Hence, additional work required for their solution is only due to these terms.

(b1) Zonation. Let there be M zonation parameters, implying M/2 zones. For simplicity, we shall assume that the reservoir and each of the zones are rectangular with the aspect ratio (the ratio of length to width) close to unity. If this assumption does not hold, there will be an additional "shape factor" occurring in the expressions for the estimated computational effort. We shall further assume that all the zones are equal in size. Then each zonation parameter determines 2N/M grid block values of either of the rock properties. The average number of brid blocks on the side of a zone in which the properties are assumed constant is approximately $\sqrt{2N/M}$. Then the equation for the sensitivity of the pressure with respect to the porosity in that zone has a forcing term with $\sqrt{2N/M}$ non-zero components. There are $\sqrt{N}$ such equations which determine the sensitivities at all N grid blocks in the reservoir. Thus, the calculation of the porosity sensitivity coefficients requires,

$$\tfrac{1}{2}\, M(6NT + \sqrt{N}\,\sqrt{\frac{2N}{m}}\,T) = 3MNT + NT\sqrt{M/2}$$

Similarly, each permeability zonation parameter determines $\sqrt{2N/M}$ values of k along any line of integration. Then, the matrices $\partial \underset{\sim}{A}_j/\partial \pi_\ell$ have $(3\sqrt{2N/M} + 4)$ non zero elements, and the terms with the derivatives of $\underset{\sim}{B}_{j-1}$, $\underset{\sim}{C}_j$ and $\underset{\sim}{D}_{j+1}$ together have $3\sqrt{2N/M}$ nonzero elements. The same count holds for the equations for the alternative direction.

Thus the total effort for the calculation of the permeability sensitivity coefficients is

$$\tfrac{1}{2}\, M[6NT + T\sqrt{N}\,(6\sqrt{2N/M} + 4)]$$
$$= 3MNT + 3NT\sqrt{M/2} + 2MNT\sqrt{N} \quad \text{mults.}$$

Adding up, total effort for the calculations of the zonation parameter sensitivities is,

$$6(M+1)NT + 2NT\sqrt{2M} + 2MT\sqrt{N} \quad \text{mults.}$$

(b2) Bayesian Approach. Each of the M Bayesian parameters affects the values of k and ϕ at the N grid blocks. Then the forcing terms in the sensitivity equations will have as many nonzero elements as those in the coefficient matrices in the original system equations. This implies $6\sqrt{N}$ nonzero forcing term elements in each of the line equations. Thus, the effort involved in solving the Bayesian parameter sensitivity equations for a 2-d reservoir using the A. D. I. method is,

$$6NT + M(6NT + 6T\sqrt{N}\sqrt{N}) = 6(2M+1)NT \text{ mults.}$$

Computing the Corrections in estimates

After the appropriate sensitivity matrix is computed by either of the two foregoing methods, the corrections $\delta\underset{\sim}{\pi}$ in the estimates are computed by the solution of the linear system,

$$\underset{\sim}{A}^T \underset{\sim}{A}\, \delta\underset{\sim}{\pi} = \underset{\sim}{A}^T \delta\underset{\sim}{y} \qquad \text{(A1.7.14)}$$

where $\delta\underset{\sim}{y}$ is the vector of pressure match errors. The (MxM) symmetric matrix $\underset{\sim}{A}^T\underset{\sim}{A}$ is obtained by $ML/2 + M^2L/2$ multiplications. The computation of the (Mx1) vector $\underset{\sim}{A}^T\delta\underset{\sim}{y}$ needs ML multiplications. The solution of the resulting (MxM) system is most efficiently obtained by Cholesky decomposition, requiring (Isaacson and Keller, 1966)

$$M^3/6 + 3M^2/2 + O(M) \quad \text{mults.}$$

The rest of the computational steps require O(M, N, L, T) multiplicative operations. The computational effort for Marquardt's method is the same as above up to O(N). Adding up, all the steps in a Gauss-Newton or Marquardt iteration require,

$$6(K+1)NT + \alpha NTL + \beta NLM + \frac{1}{2}M^2L + \frac{1}{6}M^3 + \frac{3}{2}M^2 + \frac{3}{2}ML \quad \text{mults}$$

where $\alpha = \begin{cases} 2 & \text{1-d reservoir} \\ 3 & \text{2-d reservoir} \end{cases}$

$\beta = \begin{cases} 0 & \text{Zonation} \\ 2 & \text{Bayesian Approach} \end{cases}$

when the sensitivities are computed through the adjoint variables. When the equations for the sensitivities are solved, the total number of multiplications per iteration is,

$$(6M+14)NT + 4MT + M^2L/2 + M^3/6 + 3M^2/2 + \frac{3}{2}ML \quad \text{(1-d Zonation)}$$

$$6(2M+1)NT + M^2L/2 + M^3/6 + 3M^2/2 + \frac{3}{2}ML \quad \text{(1-d, 2-d Bayesian)}$$

$$6(M+1)NT + 2NT\sqrt{2M} + 2MT\sqrt{N} + M^2L/2 + M^3/6 + 3M^2/2 + \frac{3}{2}ML \quad \text{(2-d, Zonation)}$$

Appendix 2.1

A Measure of Difference in Two Subspaces of a Linear Vector Space.

Let A and B be two linear subspaces of E^n. Let dim(A) = m and dim(B) = ℓ. Let sets of orthonormal vectors $\{\underset{\sim}{v}_i, i=1,2,\ldots,m\}$ and $\{\overline{\underset{\sim}{v}}_i, i=1,2,\ldots,\ell\}$ span the subspaces A and B respectively. Using $\{\underset{\sim}{v}_i\}$ as columns, form an (nxm) matrix $\underset{\sim}{V}$. Similarly form (nxℓ) matrix $\overline{\underset{\sim}{V}}$ from $\{\overline{\underset{\sim}{v}}_i\}$.

Any two vectors $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in A and B can be expressed as

$$\underset{\sim}{x} = \underset{\sim}{V}\,\underset{\sim}{\alpha} \tag{A2.1.1}$$

$$\underset{\sim}{y} = \overline{\underset{\sim}{V}}\,\underset{\sim}{\beta} \tag{A2.1.2}$$

The difference $\underset{\sim}{r}$ between $\underset{\sim}{x}$ and $\underset{\sim}{y}$ is,

$$\underset{\sim}{r} = \underset{\sim}{V}\,\underset{\sim}{\alpha} - \overline{\underset{\sim}{V}}\,\underset{\sim}{\beta} \tag{A2.1.3}$$

For a given $\underset{\sim}{x}$, $\underset{\sim}{y}$ can be determined so as to yield a minimum $\|\underset{\sim}{r}\|$. This can be done by finding the ℓ unknowns $\{\beta_i\}$ that yield the least square solution of the n equations,

$$\underset{\sim}{V}\,\underset{\sim}{\alpha} = \overline{\underset{\sim}{V}}\,\underset{\sim}{\beta} \tag{A2.1.4}$$

The solution is,

$$\underset{\sim}{\beta} = (\overline{\underset{\sim}{V}}^T\overline{\underset{\sim}{V}})^{-1}\,\overline{\underset{\sim}{V}}^T\underset{\sim}{V}\,\underset{\sim}{\alpha} = \overline{\underset{\sim}{V}}^T\underset{\sim}{V}\,\alpha \tag{A2.1.5}$$

This is simply the projection of $\underset{\sim}{V}\,\underset{\sim}{\alpha}$ on to the subspace B. The minimum residual is,

$$\underset{\sim}{r} = (\underset{\sim}{I} - \overline{\underset{\sim}{V}}\,\overline{\underset{\sim}{V}}^T)\underset{\sim}{V}\,\underset{\sim}{\alpha}\ . \tag{A2.1.6}$$

If $\underset{\sim}{\alpha}$ is a random vector, so is $\underset{\sim}{r}$. Then,

$$E\{\underset{\sim}{r}\} = (\underset{\sim}{I} - \bar{\underset{\sim}{V}}\,\bar{\underset{\sim}{V}}^T)\underset{\sim}{V}\, E\{\underset{\sim}{a}\} \qquad \text{(A2.1.7)}$$

and

$$E\{\|\underset{\sim}{r}\|^2\} = E\{\underset{\sim}{r}^T\underset{\sim}{r}\} = E\{\underset{\sim}{a}^T\underset{\sim}{V}^T(\underset{\sim}{I}-\bar{\underset{\sim}{V}}\,\bar{\underset{\sim}{V}}^T)(\underset{\sim}{I}-\bar{\underset{\sim}{V}}\,\bar{\underset{\sim}{V}}^T)\underset{\sim}{V}\,\underset{\sim}{a}\}$$

$$= \mathrm{tr}[E\{\underset{\sim}{a}\,\underset{\sim}{a}^T\}\underset{\sim}{V}^T(\underset{\sim}{I}-\bar{\underset{\sim}{V}}\,\bar{\underset{\sim}{V}}^T)^2\,\underset{\sim}{V}\,] \qquad \text{(A2.1.8)}$$

If the random vector $\underset{\sim}{a}$ is such that,

$$E\{\underset{\sim}{a}\} = \underset{\sim}{0}, \quad E\{\underset{\sim}{a}\,\underset{\sim}{a}^T\} = \sigma^2\underset{\sim}{I} \qquad \text{(A2.1.9)}$$

then

$$E\{\|\underset{\sim}{r}\|^2\} = \sigma^2\,\mathrm{tr}\{\underset{\sim}{V}^T(\underset{\sim}{I} - \bar{\underset{\sim}{V}}\,\bar{\underset{\sim}{V}}^T)^2\,\underset{\sim}{V}\} \qquad \text{(A2.1.10)}$$

and

$$\frac{E\{\|\underset{\sim}{r}\|\}}{E\{\|\underset{\sim}{a}\|\}} = \frac{1}{\sqrt{m}}\,[\mathrm{tr}\{\underset{\sim}{V}^T(\underset{\sim}{I} - \bar{\underset{\sim}{V}}\,\bar{\underset{\sim}{V}}^T)^2\,\underset{\sim}{V}\}]^{\frac{1}{2}} \qquad \text{(A2.1.11)}$$

The values of this measure for different cases are reported in section 2.5 (with $m = \ell$).

Appendix 2.2

A Posteriori Smoothing of the Estimate by Insensitive Corrections

Let the insensitive directions in the parameter space be given by the orthonormal column vectors of the Mx(M-ℓ) matrix $\underset{\sim}{V}_o$.

Let $\underset{\sim}{P}_o$ be the prior covariance matrix of the parameter vector $\underset{\sim}{\pi}$. Then as discussed in chapter 1, $\underset{\sim}{P}_o$ contains information about the auto-correlations and cross-correlations of the rock properties. The correction $\delta\underset{\sim}{\pi}$ in the estimate $\hat{\underset{\sim}{\pi}}$ along the insensitive direction are to be made so as to minimize the Bayesian penalty term,

$$\underset{\delta\underset{\sim}{\pi}}{\text{Min}}\ \bar{J} = (\hat{\underset{\sim}{\pi}} - \bar{\underset{\sim}{\pi}} + \delta\underset{\sim}{\pi})^T \underset{\sim}{P}_o^{-1} (\hat{\underset{\sim}{\pi}} - \bar{\underset{\sim}{\pi}} + \delta\underset{\sim}{\pi}) . \qquad \text{(A2.2.1)}$$

As the corrections lie along the insensitive directions,

$$\delta\underset{\sim}{\pi} = \underset{\sim}{V}_o \underset{\sim}{r} \qquad \text{(A2.2.2)}$$

The solution to this simple quadratic minimization problem is,

$$\underset{\sim}{r} = (\underset{\sim}{V}_o^T \underset{\sim}{P}_o^{-1} \underset{\sim}{V}_o)^{-1} \underset{\sim}{V}_o^T \underset{\sim}{P}_o^{-1} (\hat{\underset{\sim}{\pi}} - \bar{\underset{\sim}{\pi}}) \qquad \text{(A2.2.3)}$$

Since the matrix $\underset{\sim}{P}_o$ has many eigenvalues close to zero, the the evaluation of the corrections from (A2.2.3) entails severe numerical difficulties.

An alternative approximate formulation that avoids these numerical difficulties is as follows.

Let the columns of $\underset{\sim}{B}_o$ be the orthonormal set of eigenvectors of $\underset{\sim}{P}_o$ which correspond to very small eigenvalues. Then the minimization of the Bayesian penalty is approximately equivalent to

finding the corrections such that the components of the updated estimate along the columns of $\underset{\sim}{B}_o$ are as small as possible. Mathematically, we wish to find $\delta\underset{\sim}{\pi} = \underset{\sim}{V}_o \underset{\sim}{r}$ such that it yields the least square solution of the set of linear algebraic equations,

$$\underset{\sim}{B}_o^T (\hat{\underset{\sim}{\pi}} - \bar{\underset{\sim}{\pi}} + \delta\underset{\sim}{\pi}) = \underset{\sim}{0} \qquad \text{(A2.2.4)}$$

The above system of equations is equivalent to,

$$\underset{\sim}{B}_o^T \underset{\sim}{V}_o \underset{\sim}{r} = - \underset{\sim}{B}_o^T (\hat{\underset{\sim}{\pi}} - \bar{\underset{\sim}{\pi}}) \qquad \text{(A2.2.5)}$$

The solution $\underset{\sim}{r}$ of (A2.2.5) is unique if the rank of the matrix $\underset{\sim}{B}_o^T \underset{\sim}{V}_o \geq \dim(\underset{\sim}{r})$. If this condition is not satisfied, a solution $\underset{\sim}{r}$ with a smallest norm can be found through the use of the Lanczos inverse (see section 2.3).

A computationally more convenient formulation of the approximate problem considered above is as follows.

Let the columns of $\underset{\sim}{B}_1$ be the orthonormal eigenvectors of $\underset{\sim}{P}_o$ corresponding to large eigenvalues. The spaces spanned by the columns of $\underset{\sim}{B}_0$ and $\underset{\sim}{B}_1$ are the orthogonal complements of each other. Then the above problem can be posed as that of finding $\underset{\sim}{r}$ such that the component of the updated estimate orthogonal to the subspace spanned by the columns of $\underset{\sim}{B}_1$ is as small as possible. Mathematically, we have the least square problem,

$$(\underset{\sim}{I} - \underset{\sim}{B}_1 \underset{\sim}{B}_1^T)(\hat{\underset{\sim}{\pi}} - \bar{\underset{\sim}{\pi}} + \underset{\sim}{V}_o \underset{\sim}{r}) = \underset{\sim}{0} \qquad \text{(A2.2.6)}$$

The remarks about the solutions of (A2.2.5) also hold for the solution of (A2.2.6).

The formulation (A2.2.6) was used in arriving at the smoothed estimate presented in figure (2.7.1.).

Appendix 3.1

Trace of Cross Covariance Term in Estimation with Hard Constraints

The cross covariance term in expression (3.1.54) is,

$$(\underset{\sim}{I} - \underset{\sim}{G}_1\underset{\sim}{G}_1^T)\ \underset{\sim}{P}_0\ (\underset{\sim}{G}_1\ \overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T) \qquad (A\,3.1.1)$$

Since $\underset{\sim}{P}_0$ is a covariance matrix, it is real and symmetric, and it may be expressed as,

$$\underset{\sim}{P}_0 = \underset{\sim}{T}\underset{\sim}{\Lambda}\underset{\sim}{T}^T \qquad (A\,3.1.2)$$

Where $\underset{\sim}{\Lambda}$ is a diagonal matrix and $\underset{\sim}{T}$ is orthogonal. Let $\{\underset{\sim}{t}_1, \underset{\sim}{t}_2, \ldots, \underset{\sim}{t}_{2\eta}\}$ be the columns of $\underset{\sim}{T}$. Then,

$$\underset{\sim}{P}_0 = \sum_{i=1}^{2N} \lambda_i \underset{\sim}{t}_i\underset{\sim}{t}_i^T \qquad (A\,3.1.3)$$

Using (A 3.1.1) and (A 3.1.3),

$$\begin{aligned} \mathrm{tr}\ \{(\underset{\sim}{I} - \underset{\sim}{G}_1\underset{\sim}{G}_1^T)\ \underset{\sim}{P}_0\underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\} &= \mathrm{tr}\ \{\sum_{i=1}^{2N} \lambda_i\ (\underset{\sim}{I} - \underset{\sim}{G}_1\underset{\sim}{G}_1^T) \\ &\qquad \underset{\sim}{t}_i\underset{\sim}{t}_i^T\ \underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\} \\ &= \sum_{i=1}^{2N} \lambda_i \underset{\sim}{t}_i^T\ (\underset{\sim}{I} - \underset{\sim}{G}_1\underset{\sim}{G}_1^T)\ \underset{\sim}{G}_1\overline{\underset{\sim}{V}}_0\overline{\underset{\sim}{V}}_0^T\underset{\sim}{G}_1^T\underset{\sim}{t}_i \\ &= 0 \qquad (A\,3.1.4) \end{aligned}$$

Where the last result is arrived at by noting that the columns of the matrix $\underset{\sim}{G}_1$ form an orthonormal set of vectors and consequently, the matrices $(\underset{\sim}{I} - \underset{\sim}{G}_1\underset{\sim}{G}_1^T)$ and $\underset{\sim}{G}_1$ are orthogonal to each other.

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A., "Handbook of Mathematical Functions," Dover Publications, Inc., (1968), p. 984.

Bryson, A.E., Jr. and Ho, Y. C., "Applied Optimal Control," Ginn and Company, (1969).

Bushinger, P. A. and Golub, G. H., "Singular Value Decomposition of a Matrix," Communications of the ACM, Vol. 12, No. 10, (October 1969).

Carter, R. D., Kemp, L. F., Jr., Pierce, A. C., and Williams, D. L., "Performance Matching with Constraints," Soc. Pet. Eng. J., (April 1974), pp. 187-196.

Chen, W. H., Gavalas, G. R., Seinfeld, J. H., and Wasserman, M. L., "A New Algorithm for Automatic History Matching," Soc. Pet. Eng. J., (December 1974), pp. 593-608.

Chevant, G., Dupuy, M., Lemmonier, P., "History Matching by Use of Optimal Theory," Soc. Pet. Eng. J., (February 1975), pp. 74-86.

Collins, R. E., "Flow of Fluids Through Porous Materials," Reinhold Publishing Corp., (1961), New York.

Davenport, W. R. and Root, W. L., "An Introduction to the Theory of Random Signals and Noise," McGraw Hill, (1958).

Eykhoff, P., "System Identification: Parameter and State Estimation," Wiley Interscience, (1974), London, p. 162.

Gear, G. W., "Numerical Initial Value Problems in Ordinary Differential Equations," Prentice-Hall, Englewood Cliff, N. J., (1971), p. 210.

Isaacson, E., Keller, H. B., "Analysis of Numerical Methods," John-Wiley & Sons, (1966), New York, p. 54.

Jackson, D. D., "Interpretation of Inaccurate, Insufficient and Inconsistent Data," Geophys. J. R. Astr. Soc., (1972), No. 28, pp. 97-100.

Jacquard, P., and Jain, C., "Permeability Distributions from Field Pressure Data," Soc. Pet. Eng. J., (December 1965), pp. 281-294.

Jahns, H. O., "A Rapid Method for Obtaining a Two-Dimensional Reservoir Description from Well Response Data," Soc. Pet. Eng. J., (December 1966), pp. 315-327.

Jazwinski, A. H., "Stochastic Processes and Filtering Theory," Academic Press, (1970), New York, p. 262.

Lanczos, C., "Linear Differential Operators," D. VanNostrand Co., (1961), London.

Marriam, D. F. (Ed.), "Random Processes in Geology," Springer Verlag, (1976), pp. 63-86.

Noble, B., "Applied Linear Algebra," Prentice-Hall, (1969), Englewood Cliffs, N. J., p. 335.

Papoulis, A., "Probability, Random Variables, and Stochastic Processes," McGraw-Hill, (1965), New York, p. 250.

Peaceman, D. W., and Rachford, H. H., Jr., "The Numerical Solution of Parabolic and Elliptic Differential Equations," J. of SIAM (March 1955), Vol. 3, No. 1, pp. 28-41.

Penrose, R., "A Generalized Inverse for Matrices," Proc. Camb. Phil. Soc., (1955), Vol. 51, pp. 406-413.

Polak, E., "An Historical Survey of Computational Methods in Optimal Control," SIAM Review, (1973), Vol. 15, No. 2, pp. 553-584.

Pryor, W. A., "Reservoir Inhomogencities of Some Recent Sand Bodies," Soc. Pet. Eng. J., (June 1972), pp. 229-245.

Sage, A. P., and Melsa, J. L., "Estimation Theory with Applications to Communications and Control," McGraw-Hill, (1971), New York, p. 176.

Smith, M. L., and Franklin, J. N., "Geophysical Application of Generalized Inverse Theory," J. Geophys. Res., (1969), Vol. 74, pp. 2783-2785.

Thomas, L. K., Hellums, L. J., and Reheis, G. M., "A Nonlinear Automatic History Matching Technique for Reservoir Simulation Models," Soc. Pet. Eng. J., (December 1972), pp. 508-514.