THE UNIVERSITY OF CALGARY

A SEQUENTIAL OPTIMIZATION PROCEDURE IN THE INVERSE MODELLING OF THE HYDRAULIC TRANSMISSIVITIES FOR THE FORMER CANADA CREOSOTE SITE IN CALGARY, ALBERTA

BY

NGOK-MING TO

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE STUDIES IN PARTIAL

FULFILMENT OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF CIVIL ENGINEERING

CALGARY, ALBERTA

SEPTEMBER, 1998

ABSTRACT

Six scenarios of flow data have been captured for the former Canada Creosote site over several years. Six corresponding MODFLOW steady state numerical models for the site have been constructed to analyse the data.

A new, automated sequential optimization procedure has been developed to maximize the usage of the information in the scenarios so as to reach a common set of optimized transmissivities for the model zones. The procedure controls the variances, and minimizes the overall residuals' sums of squares of the lack of fit. An application of industrial optimization concepts enables well-conditioned matrices to be formed by the use of variable weights in the optimization. The optimized values produce lower residuals than the existing "expert input" from hydrogeologists.

Different optimized quantities with unlike units and magnitudes are compared by the use of a new measure named weighted optimized non-dimensional norm (WONN).

Acknowledgment

The author would like to thank his advisor, Dr Laurence Bentley, for his guidance, and to express appreciation for his patience towards the author who is new to the field of groundwater modelling. He would like to thank Dr Les Smith of UBC, for the helpful information on and insights into the response surface method as it is applied to groundwater modelling.

The kind assistance of the Calgary office personnel of CH2M Gore & Storrie Limited, in supplying the electronic data for the project and providing the details of the work prior to this study, is very much appreciated.

He would also like to thank his wife Cecilia for her support and patience, and his daughters Beatrice and Emily for proof reading and help in preparing the flow charts.

Financial assistance from the NSERC operating grant and the Engineering for the Environment program at the University of Calgary is gratefully acknowledged.

DEDICATION

"The harvest of a quiet eye"
- William Wordsworth

This thesis is dedicated to

Dr Sheldon Cherry,

and

to the memory of the late

Dr Noel Nathan, of UBC.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Charts

List of Symbols

Y: vector of the dependent variable of the model output of heads, n by 1
n': number of observations and items of prior information
n: number of observations or measurements
p: number of parameters to be optimized (also number of constraints in penalty functions)
X: sensitivity matrix n by p
system response to unit change in parameter
β: vector of true system parameters, p by 1
ε: true error or disturbance vector n by 1
e: estimator of the true error or error vector
b: the estimator of parameter vector for some parameter vector **ß**
σ^2_{true}: true variance
σ^2_{ref}: reference variance
S(b): sum of square e^Te, which in PEST is SS_{1of}
Φ: S(b) in PEST
εs: vector of true error for sample
εp: vector of true error for prior information
sffp: sequential feed forward procedure
Var( ): is the variance operator on a random variable
σ^2_{comu}: common variance of untransformed disturbance, for samples or measurements only in this study
σ^2_{com}: common variance of transformed disturbance, for samples or measurements only in this study
$\ddot{y}_o$: vector of model generated heads
b_o: parameter set used in optimization process
Ω: generalized diagonal weighting matrix, may include V^{-1} or it may reflect importance of data
V: variance-covariance matrix for variables
f(x): continuous functions of x
c: penalty parameter
P(x): penalty term
$½(g(x))^2$: penalty term in quadratic form
Q, F, G: Hessians of q(x), f(x), g(x) respectively
λ: Lagrange multipliers
ss: sample size for Monte Carlo simulation
pf: estimated probability for Shooman's Monte Carlo relationship
SS_T: total sum of square
SS_{mean}: Sum of square due to the mean
SS_{conv}: Sum of square corrected for the mean
SS_{fact}: Sum of square due to factors
SSr: Sum of square of residual
SS_{lof}: Sum of square lack of fit

SS_{pe}:	Sum of square, purely experimental error
$\acute{\mathbf{y}}$:	vector of measurement
cov :	Raw coefficient of variation = (variate sd / variate mean)
sd or σ :	Standard deviation of random variable
μ :	mean value of random variable
Σ_p:	the parameter covariance matrix with prior information on the parameters
$\mathbf{C}_p$:	the joint covariance matrix of the prior parameter estimates
Cov( ):	The covariance of variables in the bracket due to sample variance alone (no prior information on parameters)
wonn:	The weighted optimized non-dimensional norm
asp:	Arbitrary starting point simulations

1.0 Introduction

The former Canada Creosote site (Figure 1.1-site plan) is located adjacent to the south bank of the Bow River on the western edge of downtown Calgary, and covers an area of approximately 18 hectares. Wood preserving operations at the site occurred over a period of approximately 38 years from 1924 to 1962, and involved the use of tars, creosote and petroleum oils. Pentachlorophenol was used during the 1950s.

The site is underlain by sand and gravel fill (1 to 6.7m thick) and alluvial sands and gravels (1 to 5 m thick). Bedrock typically occurs at a depth of 6 to 9m and is comprised of sandstone and shales. The water table occurs at a depth of 4 to 7m below ground surface, and is situated within the alluvial sands and gravels. The groundwater flows towards the river. Investigations at the site initiated by Alberta Environment in 1988 to assess the orphaned site showed that the major form of contamination beneath the site is free creosote that occurs as a pool within the alluvial sands and gravels that infills a depression in the bedrock surface. This pool lies below the water table, encompasses an area of approximately 31,000 m^2, and is estimated to contain 4750 m^3 of creosote. (Golder Associates Ltd, 1990)

To minimize contaminant loading to the Bow River from the site, a groundwater contaminant control system is installed at the site. The components of the system include a partially enclosing bentonite slurry and secant pile wall, a set of clean water and dirty water pumping wells and a dirty water treatment plant. The wall is designed as a physical barrier to the direct flow of creosote related nonaqueous phase liquids. The clean and dirty water pumping wells are to control the groundwater mounding behind the wall which would otherwise drive the dissolved phase of the contaminants around the ends of the wall or beneath it through the bedrock (CH2M G&S 1997). The clean water pumping wells are to intercept the inflow of groundwater into the creosote area, so as to minimize the amount of water flowing through the creosote plume area. The dirty water pumping wells are to stop the dissolved phase of the creosote from leaving the containment area. The dirty water will be treated to extract the dissolved creosote before the cleaned water

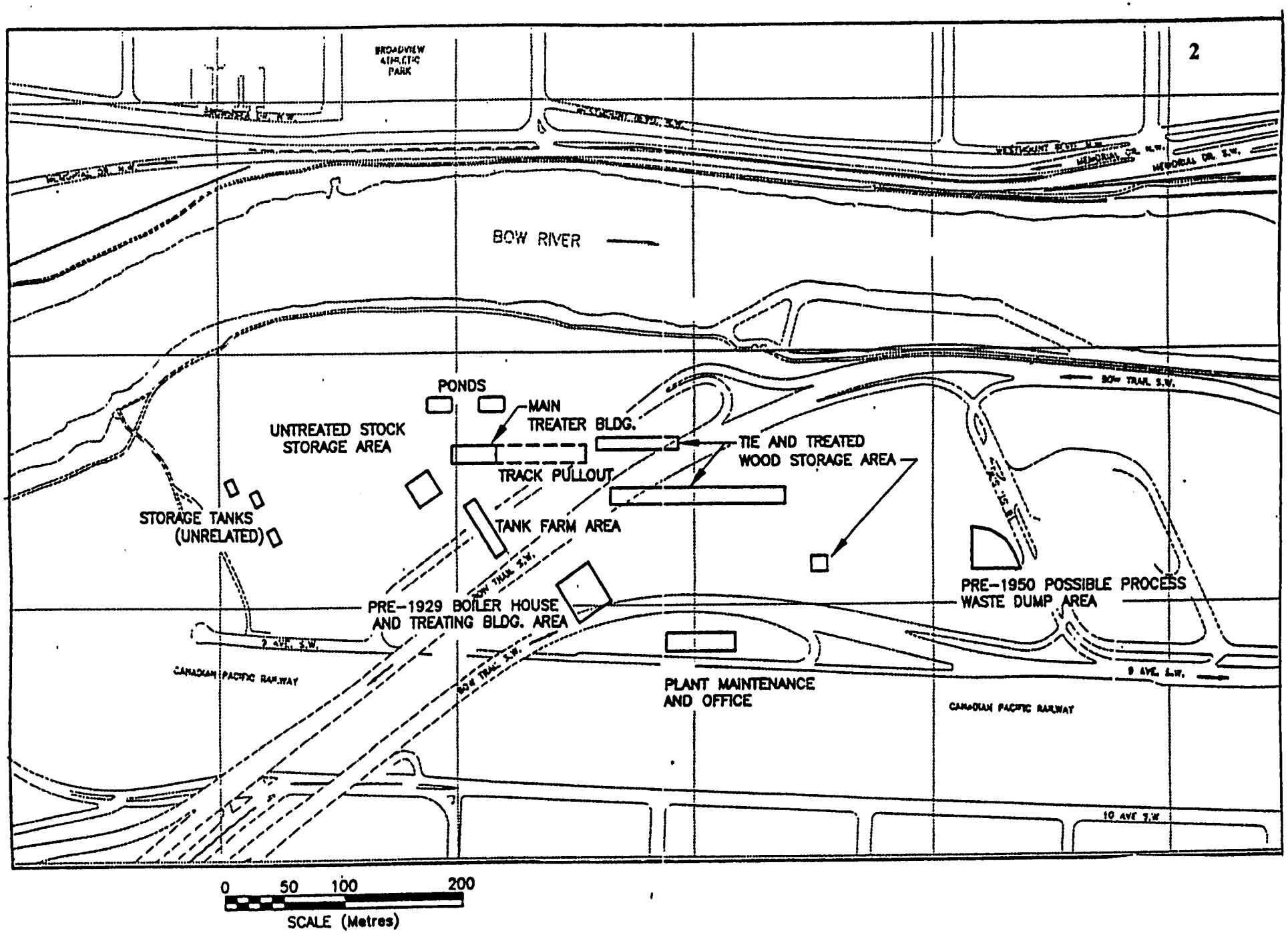
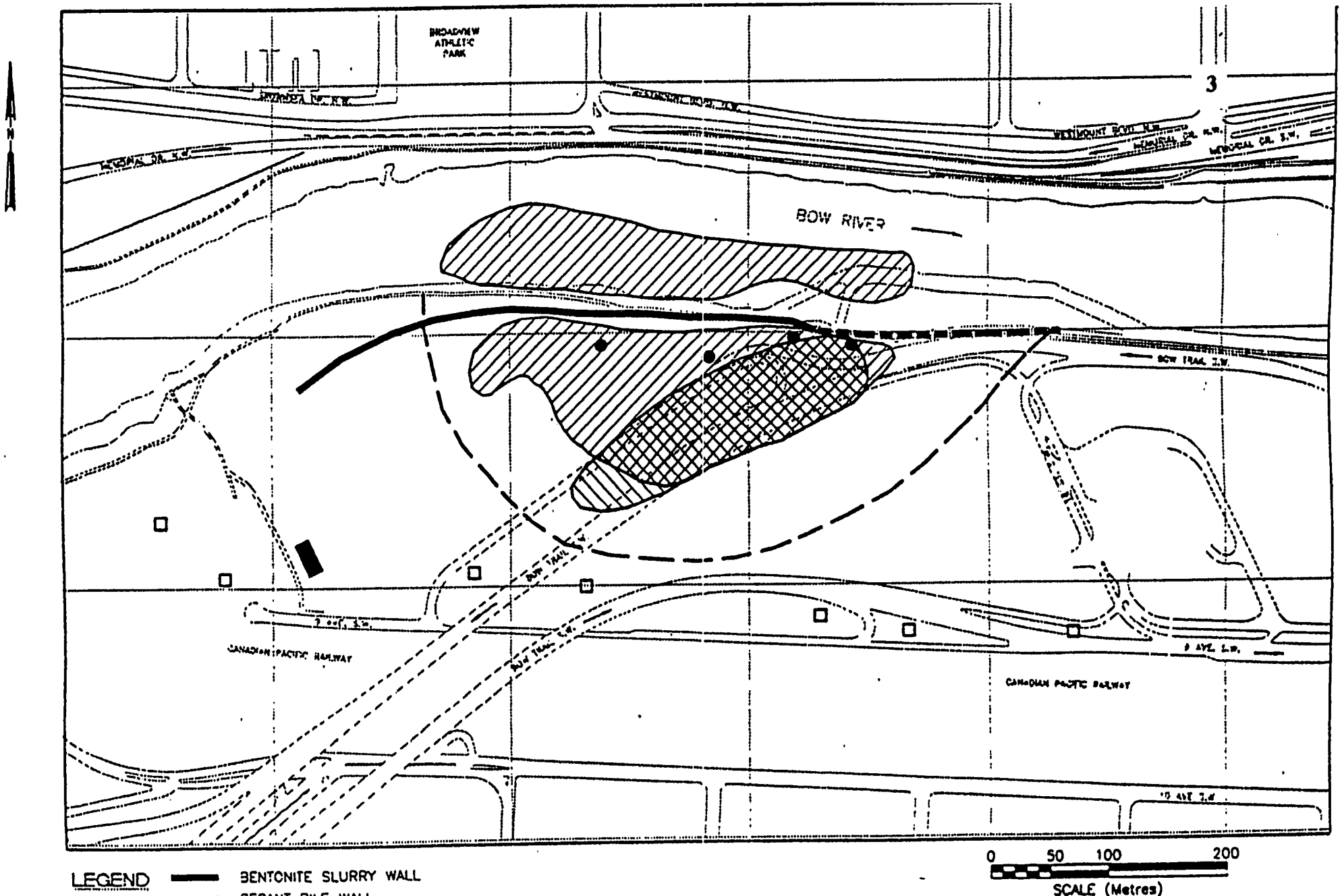


Figure 1.1 Former Canada Creosote site plan, Calgary, Alberta



Main Contaminant Bodies and Management Strategy Components

Figure 1.2 Pollution countermeasures on site

is released back to the environment. The free creosote is slowly being extracted as well, in small quantities, as the economics of creosote disposal become more attractive.

To support the development of an effective groundwater management strategy, CH2M Gore & Storrie Limited developed a groundwater flow model for the site on behalf of Alberta Environmental Protection. The model was developed to provide the site manager with a reliable decision support tool with which to assess current and future pumping scenarios (Figure 1.2). The United States Geological Survey model MODFLOW was selected as the primary groundwater flow model, and model design and post-processing was done with Visual Modflow by Waterloo Hydrogeologic Software Ltd. Model calibration and validation are reported (CH2M G&S 1996). Since the report, more measurements have been made and additional measuring stations installed.

The northern boundary of the model is the Bow River. The boundary is modelled by fixing the river elevation reference to a measured level in a monitoring well directly adjacent to the river. This level is projected up and down stream at a gradient of .002. The southern boundary corresponds to the southern escarpment of the Bow River floodplain. A single constant value of 1050.3m above sea level is chosen as the fixed hydraulic head of this boundary. The bottom boundary is impermeable bedrock, and the east and west boundary of the model domain are modelled by no flow cells.

For studies performed todate, recharge into the region is regarded as a second order process, and its magnitude for the simulation is zero. A value of 0.15 is chosen for the specific yield of the region, and 1e-6 is chosen for specific storage.

The model is divided into regions or zones of constant hydraulic parameters. These zones have been adjusted in calibration procedure and are shown in Figure 1.3. Hydraulic transmissivities are extremely difficult to measure in the field, and the determination of the transmissivities to use in the zones is a time consuming process. Efforts are made to derive the proper values from geological information of the site and pumping tests results. Then these values have to be modified in calibration procedure based on measurements. After the calibration, the model predictions to new events at the site have been reported to be satisfactory (CH2M G&S 1997).

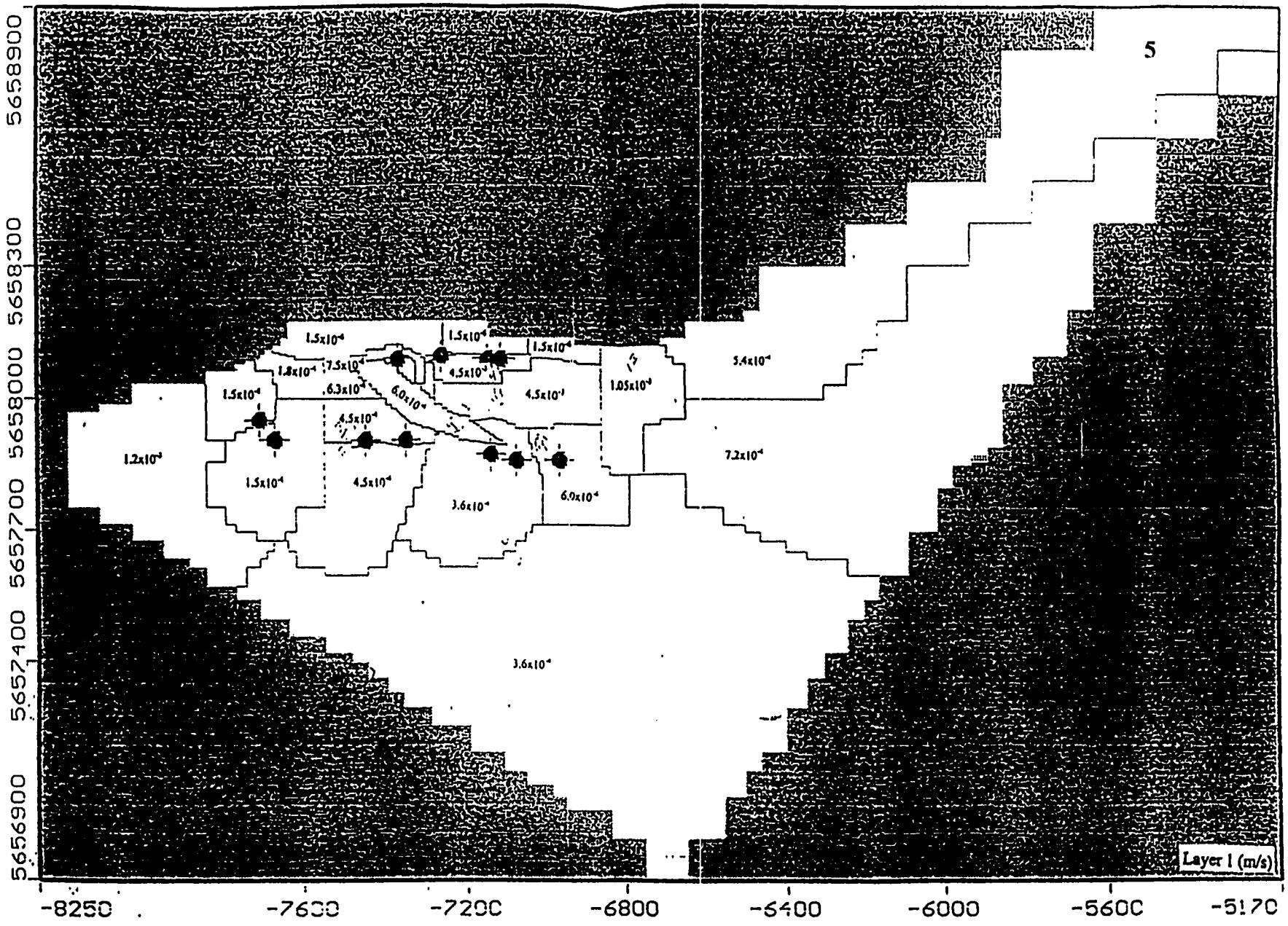


Figure 1.3 Examples of zones of hydraulic transmissivity

1.1 Model calibration

The computer model calibration and validation against monitoring well measurements has been described in the 1996 CH2M G&S report. Some of the assumptions of the model include:

- Recharge is not included in the model
- Storage values are not used in steady state simulations

The state of the groundwater level at the site can be broadly classified into a high level and a low level due to the different river stages in the summer and winter. During these two states it is assumed that the water level does not change significantly. Thus steady state simulations are appropriate. The present study does not deal with the transient states in between the high and the low levels, and events during ice-jam situations. Ice jam may cause the river level to rise during low flow conditions. The rise in river level may occur in a relatively short time (within hours) and may cause water to flow into the contaminated area from the river. Such an occurrence is not a steady state but a transient event.

Six sets of head measurements for the steady state scenarios have been provided for use in this present study. They are:

a. Scenario 1 or Scen1, May 1993: natural site conditions where river stage is high due to spring flood, and site groundwater elevations are high due to spring recharge. This is called the pre-wall, high season scenario with no wells pumping.
b. Scenario 2 or Scen2, September 1992: site conditions where river stage is low and groundwater elevations are at or near seasonal lows. This is the pre-wall, low season scenario, no wells pumping
c. Scenario 3 or Scen3, June 1995: High river stage season after the wall was finished. This is the postwall, high season scenario with no pumping wells
d. Scenario 4 or Scen4, December, 1995: Low river stage season 8 months after the wall was installed. This is the post wall low season scenario no wells pumping

e. Scenario 5 or Scen5, June17 1997: High river stage season after the wall was installed and clean and dirty wells are pumping. This is the post wall, with wells high season scenario

f. Scenario 6 or Scen6, October , 1996: Low river stage season after wall was finished with the wells pumping. This is the post wall with wells, low season scenario.

The hydraulic head measurement well location for each scenario, and the clean water and dirty water pump well locations for the last two scenarios are shown in a series of 6 maps, figures 1.4a to 1.4f , each corresponding to a scenario.

These head measurements, together with model domain grid information, boundary conditions, and transmissivity values obtained from calibration against the head measurements for the zones in the domain, are provided by CH2MG&S for this study.

The models are used with the Visual Modflow pre and post processors. They have been converted for use with a different pre- and post-processing set of programs called Groundwater Vistas for this present study. Groundwater Vistas is a product by James and Douglas Rambaugh of Environmental Simulations Inc., of Herndon, Virginia. Figure 1.5 shows a sample of Groundwater Vistas output of hydraulic head contour map with flow direction arrows for scenario 5.

1.2 Review of Inverse Modelling, Parameter Estimation and Optimization with Prior Information and Weights

To obtain a good fit between measured head data and computed results from flow models such as MODFLOW, the model parameters, such as hydraulic conductivity, are calibrated against the observed head data. The method of calibration used is manual and labour intensive. This thesis presents a methodology to automatically calibrate system parameters of hydraulic conductivity (Carrera and Neuman, 1986a) for the scenarios of the Canada Creosote site.

Parameters such as hydraulic conductivity or vertically averaged transmissivity tensors are constitutive variables, and as such should be calculated, as oppose to field

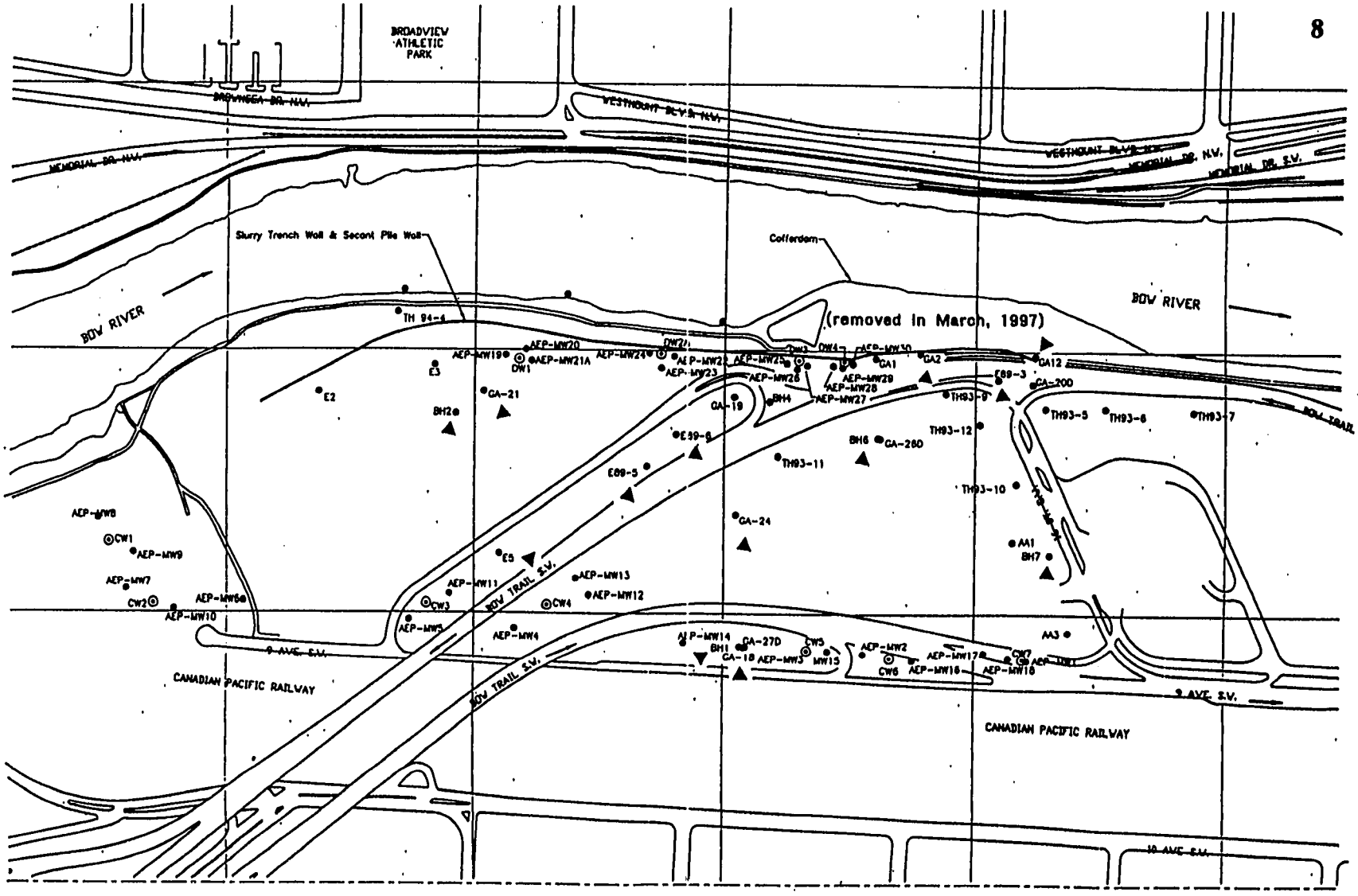


Figure 1.4a Scenario 1

Note: Arrow heads show monitoring well locations for scenario

Wells not shown: GA8, GA11, GA22, GA23, E89-4, GA17, GA28, GA9

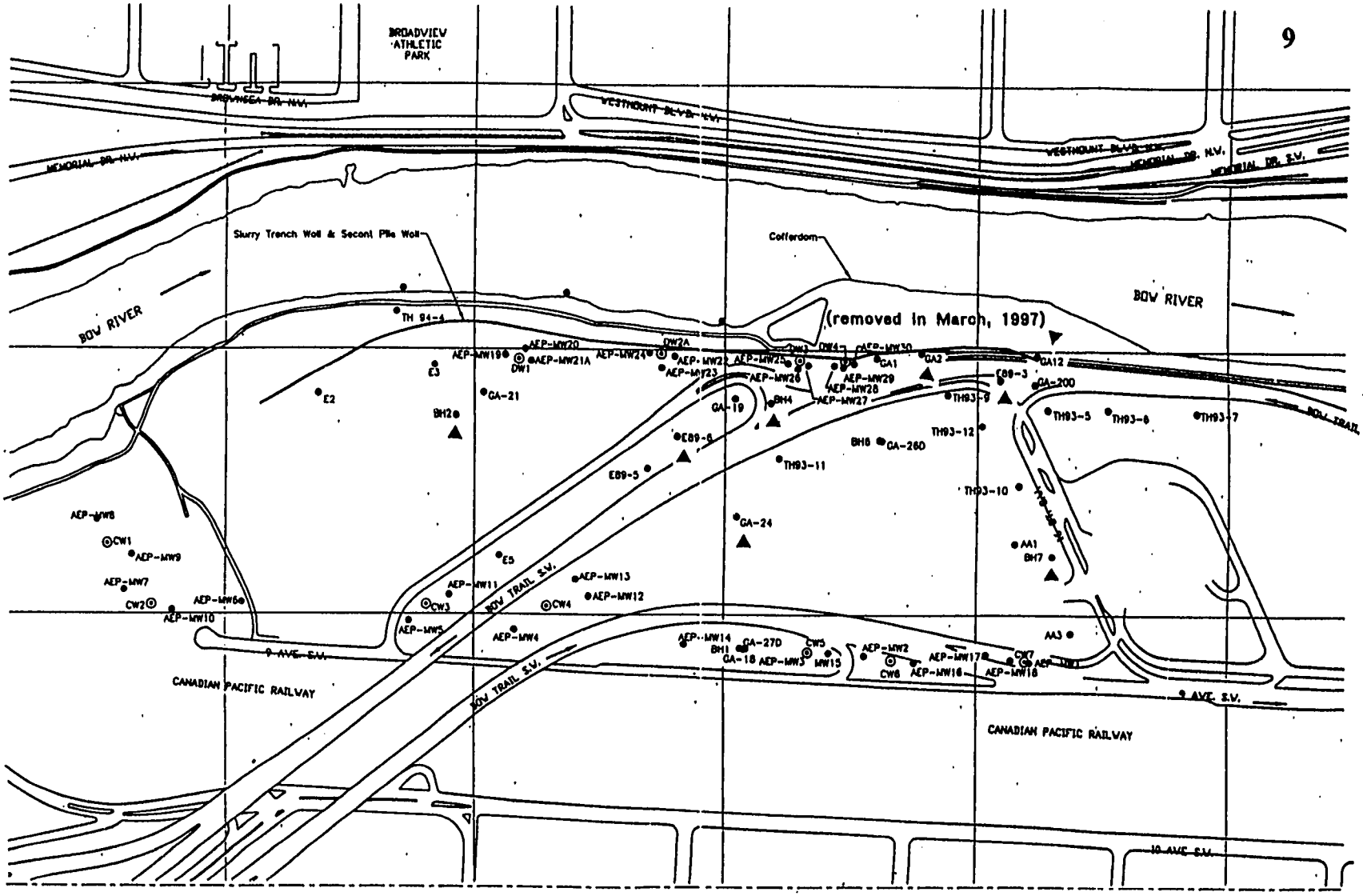


Figure 1.4b Scenario 2

Note: Arrow heads show monitoring well locations for scenario

Wells not shown: GA10, GA22, GA9, GA8, E89-4, GA11, GA21, GA23, E89-5

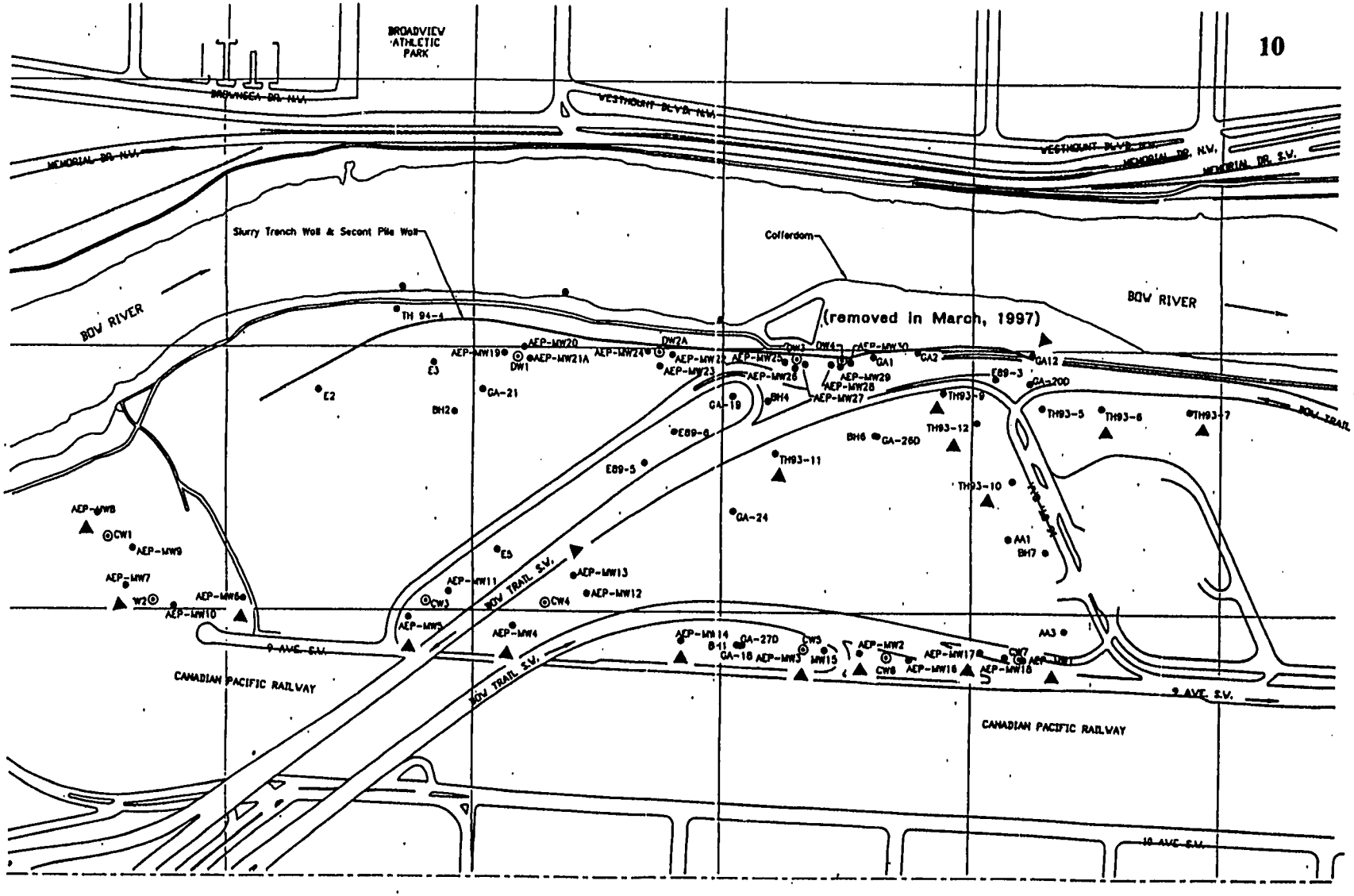


Figure 1.4c Scenario 3

Note: Arrow heads show monitoring well locations for scenario

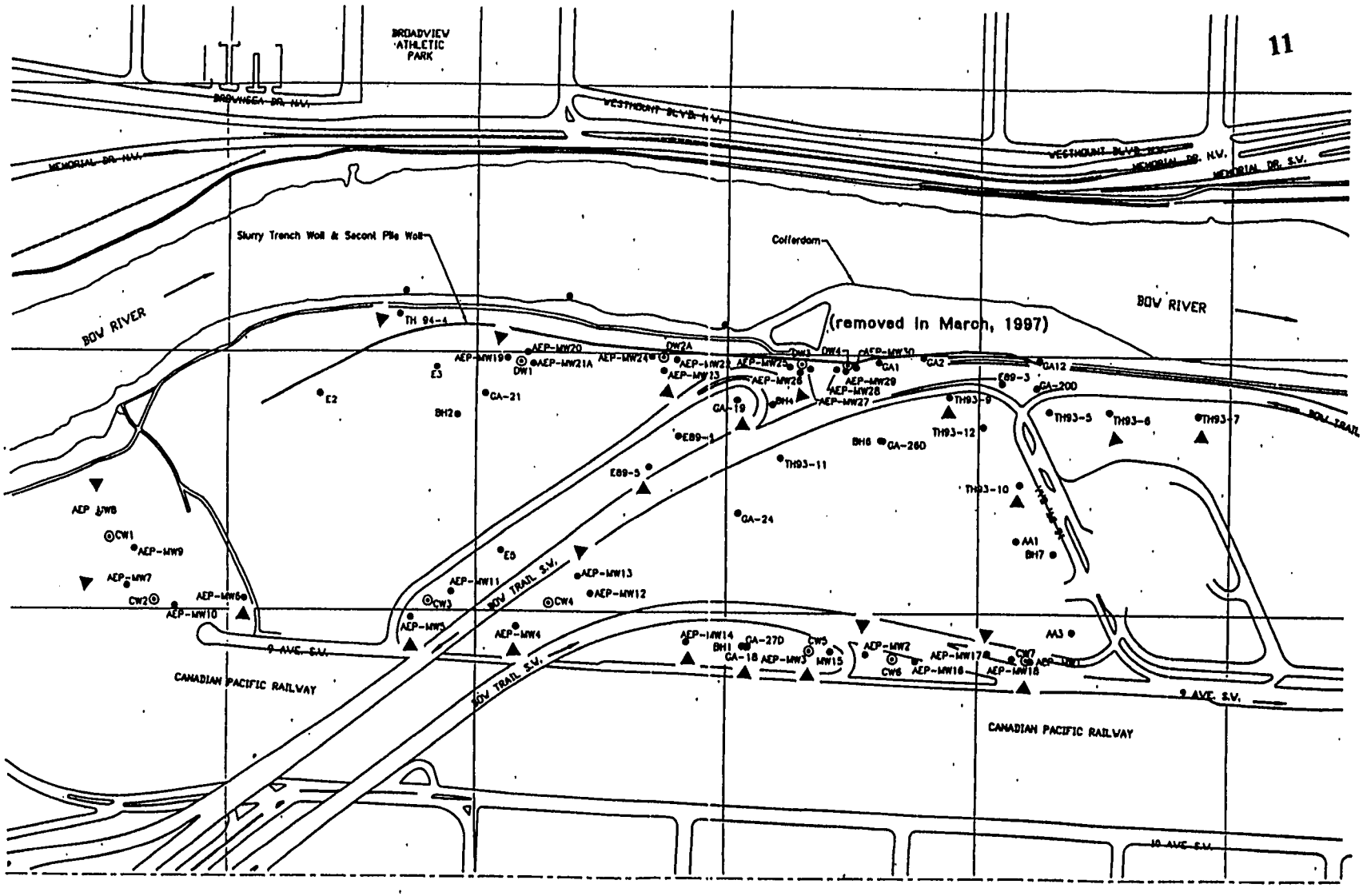


Figure 1.4d Scenario 4

Note: Arrow heads show monitoring well locations for scenario

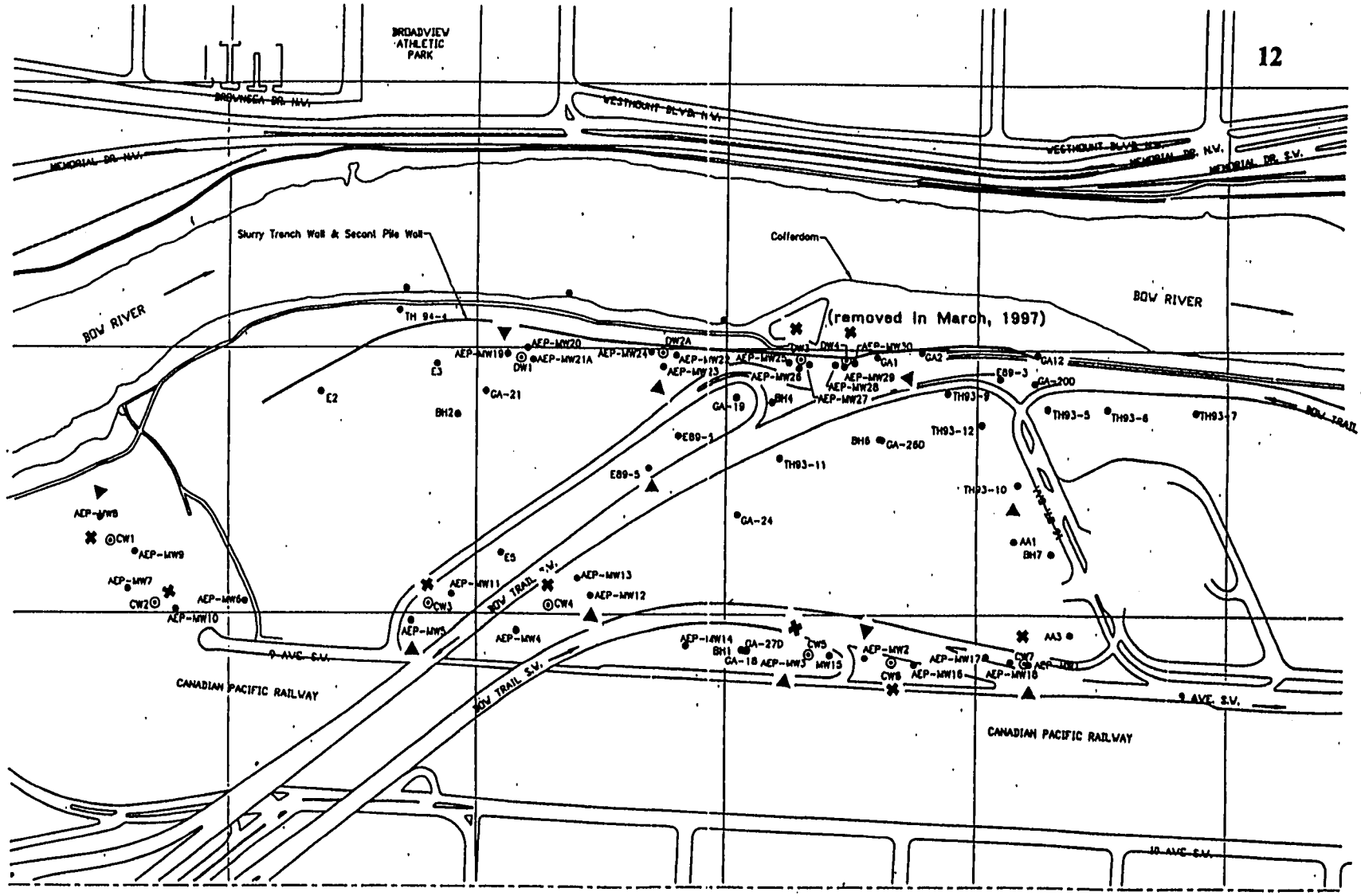


Figure 1.4e Scenario 5

Note: Arrow heads show monitoring well locations for scenario
Crosses show locations of pumping wells

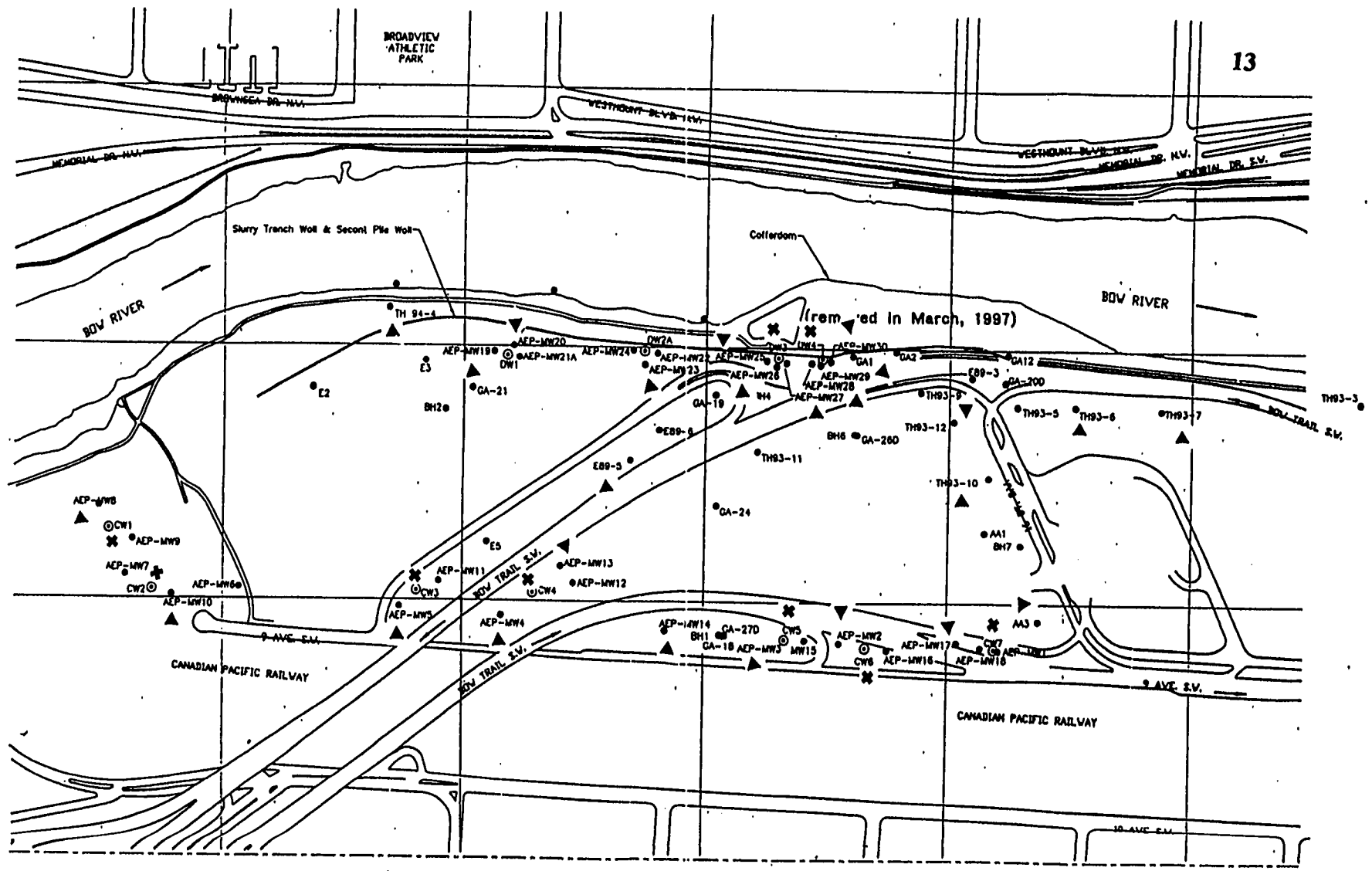
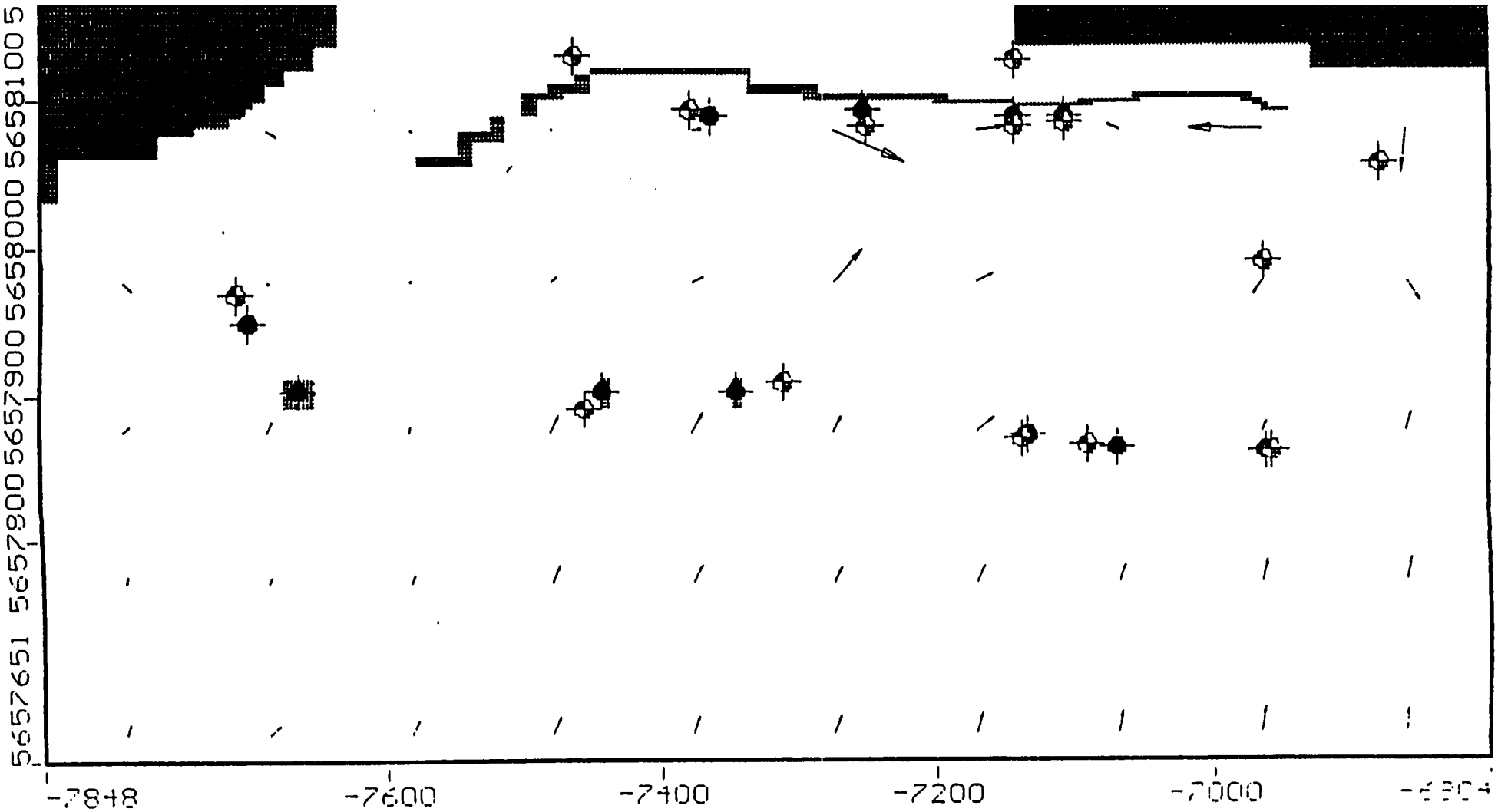


Figure 1.4f Scenario 6

Note: Arrow heads show monitoring well locations for scenario
Crosses show pumping well locations


5657651
5657900
5657900
5658000
5658100
5
-7848
-7600
-7400
-7200
-7000

variables (eg hydraulic heads) which are measured (Cushman, 1986). To calculate the parameters from measured head data, inverse modelling is used. Inverse modelling proceeds in a direction opposite to that of forward modelling.

In forward modelling of subsurface flow, the spatial and temporal distribution of hydraulic heads of a modelled area is defined by function F of parameters $p_i(x)$, that is, $h(x,t) = F(p_i(x))$, where p_i can include parameters such as conductivity or transmissivity (Carrera and Neuman, 1986b). Inverse modelling, on the other hand, calculates the parameters from head distributions, that is, $p_i(x) = R_i(h(x,t))$. The traditional formulation of the steady state Cauchy problem for groundwater flow in mathematical physics gives $R_i = F^{-1}$, and the solution of the inverse modelling for the parameters depends on how well-posed the problem is. In the context of classical mathematical physics, the solution to the inverse modelling problem depends on whether the parameters are identifiable, unique, and stable: For the classical problem the solution is unique if and only if the parameters are identifiable. (Carrera and Neuman, 1986b).

The inverse modelling technique in this thesis takes the indirect approach (Neuman and Yakowitz, 1979) . The indirect approach minimizes head residuals between the measurements and the modelled results. The classical solution technique is no longer used here, and hence the parameter identifiability and uniqueness may or may not be related. The three properties of the parameters can all be associated with the ill-conditioned matrices assembled for the problem, and can be explained thus (Carrera and Neuman, 1986b):

Identifiable parameter: indicated by the rank of the Jacobian matrix being equal to the number of decision variables or parameters to be found;

Unique Parameter: Hessian matrix is positive definite, or the objective function is convex within the domain of definition of parameters. It is difficult to ensure a global minimum for the objective function so that there is global parameter uniqueness. However, when prior information restricts the solution to a minimum in its neighbourhood, the solution is acceptable even without the assurance of a global minimum. Prior information on the

parameters is prior knowledge gained through the modellers' experience, judgment or existing measurements.

Stability: Small changes in head data gives large oscillations in parameter estimates. This is especially true if the data is noisy and the Jacobian matrix is ill-conditioned. An eigenvalue analysis will illustrate this point in subsequent section of this thesis. Instability is sometimes unavoidable in the analysis, but meaningful results can still be obtained by reducing the dimensionality of the parameter space.

These considerations will be relevant to indirect inverse modelling, and will be discussed as they arise. In particular, more recent papers outline assessment methods to quantify parameter identifiability in inverse modelling (Speed and Ahlfeld, 1996) by the use of the raw coefficient of variation. Classical applied analyses enable the assessment of the noise and instability problem by using the eigenvalue methods (Lanczos, 1964).

Other considerations affecting the effectiveness of the inverse modelling of parameter estimation are related to the scale of the head measurements and prior information as compared to that of the model (Ginn and Cushman, 1990). Scale incompatibility is an issue that has to be addressed by modellers.

Also incompatible are the different types of residuals in the objective function in Carrera and Neuman (1986a), consisting of head residual and parameter prior information residuals. These residuals have different units and are kept separate from each other in augmented matrices. The program PEST (see section 1.3) adopts the same separation of the residuals in inverse modelling. In spite of the fact that these residuals are incommensurate, an overall minimum is obtained for the residuals at the optimal point. This study proposes a special non-dimensional measure for different types of optimized quantities with dissimilar units (including head residuals and parameter uncertainties), to facilitate comparison for decision making.

Finally, much research effort in parameter estimation by inverse modelling in the subsurface flow has been spent in identifying the nature and the statistical structure of the errors and numerical algorithms, with little apparent transfer of technology from other

fields of study (such as engineering), which use very similar techniques in optimization. The recent papers on the use of factor space analysis in groundwater applications utilized some of the response surface analysis concepts (Weiss and Smith, 1998a,b). This type of work could prove to be fruitful.

In this thesis, well established techniques in experimental design such as the analysis of variance (ANOVA), and manufacturing optimization strategies used by Genichi Taguchi and other engineers (see for example, Roy, 1990) are applied to the data analyses. Results generated quantitatively demonstrate improvements over the "expert input", in the overall minimization of residuals in the optimization of the scenarios. Scaling of optimized measures could be used to compare different criteria for decision making purposes.

1.3 The purposes of the present study

As has been explained in the last section, hydraulic transmissivities are parameters that are difficult to measure in the field. By the use of computers and automated techniques, this study attempts to obtain a set of groundwater flow parameters such as hydraulic transmissivities, through the use of optimization technique and indirect inverse modelling methodology for the Canada Creosote site in Calgary, Alberta.

The method to optimize transmissivities with the help of numerical models is to form an objective function with the residual between the hydraulic heads measured at the monitoring wells and the computer model output. The value of the objective function is then minimized by the use of some methodology, such as the Gauss-Newton technique. The optimum hydraulic transmissivities are found when a global minimum is reached for the objective function. The theory of this technique is well documented (Cooley and Naff, 1990, Chauvent, 1991). Other relevant statistical techniques can be applied to the analyses to give a measure of uncertainty to the numerical results. Numerous applications of these techniques to groundwater flow applications can be found in the literature (Neuman and Yakowitz, 1979, Yeh, 1986, Cooley, 1977, just to name a few).

The numerical package used in this study for optimization using the result of MODFLOW is the Parameter Estimating software from Watermark Computing (1994), called PEST. This program can be run in conjunction with Groundwater Vistas, and allows the use of weighted prior information for the parameters. Weights for the measurements are also allowed. The use of prior information for the parameters is to reduce the parameter uncertainty, but at the expense of the model fit (Yeh, 1986, page 106). The "proper" use of weights for both the prior information and the measurements will enhance the parameter estimates.

Herein lies a special challenge for this study. The existence of six scenarios with different monitoring well locations and boundary conditions presents a unique opportunity for parameter estimation, which are constitutive variables of the groundwater flow governing equations, and should be the same for all scenarios of the same spatial region.

One of the main objectives of the study is to devise a methodology to optimize the six scenarios in sequence by using different head and prior information weighting methods to achieve minimum overall head residuals. This sequential optimization will also have several benefits over single optimizations:

a) The uncertainty of parameters can be reduced simultaneously as the lack of fit between model output and measurements is minimized.

b) The parameter prior information is generated from one sequential step to another, based on head data from observation wells and scenario numerical models of the same region. The problem due to incompatible scale (Ginn and Cushman, 1990) is minimized.

c) There is room for technical innovation. Concepts from manufacturing optimization can be adopted. Industrial process optimization has two components - one is the search for the optimal point, the other, which is just as important, is to minimize the variance of the process (Taylor, 1991). While the exact mechanics of factory-style minimization of

process variance is not possible in groundwater applications, the concept can certainly be transferred. The coefficient of variation method to determine parameter identifiability (Speed and Ahlfeld, 1996), is refined here by specifying and controlling the parameter standard deviation. A variance control variable weighting method is used in a new sequential optimization process. All are developed and used in this study.

d) Decision support: The different scenarios are modelled with different number of zones of transmissivity. With the use of sequential optimization a model choice can be made with the comparison of "norm" measures of selected optimization criteria that can best represent the modellers' wants and needs.

The results of the sequential optimization will be compared with the results from the manual calibration performed by expert hydrogeologists.

2.0 The Process of Parameter Estimation

This section details some of the background, the mathematics and the optimization techniques used in the program PEST™ or Parameter Estimation, a model independent parameter estimation algorithm produced by Watermark Computing, 1994. PEST is based on the Gauss-Levenberg-Marquardt procedure. The least squares method as applied to the indirect inverse modelling technique in groundwater hydrology is explained in the first subsection 2.1. The use of prior information and weights as they relate to the sequential optimization process, and the concepts of penalty function and the optimization process, Lagrange multipliers, the canonical convergence ratio, eigenvalue, eigenvector and noise analyses, are explained in sections 2.2, 2.3. Section 2.3.1 contains a discussion of the penalty function and the associated eigenvalue analyses. The similarities between techniques used in optimization and experimental design are explained in section 2.4 as well as the use of ANOVA to examine the optimized results.

2.1 The least squares method in inverse modelling

The groundwater hydraulic head distribution at the former Canada Creosote site is monitored by a number of observation wells, at which the water level or hydraulic heads are measured. These measurements are made on a continuous basis throughout the year. The northern boundary of the numerical model is the Bow river, and is modelled as a fixed head boundary (see Figure 2.1). The southern boundary, corresponding to the southern escarpment of the Bow River floodplain, is a fixed head boundary as well, the constant value of which is based on previously calibrated results (CH2MG&S 1997 report). The concept of the numerical model of the groundwater flow at the former Canada Creosote site can be described by a general relationship of indirect inverse modelling (section 1.2):

Measured hydraulic heads = (Transfer or mapping matrix) (site parameters-transmissivities)

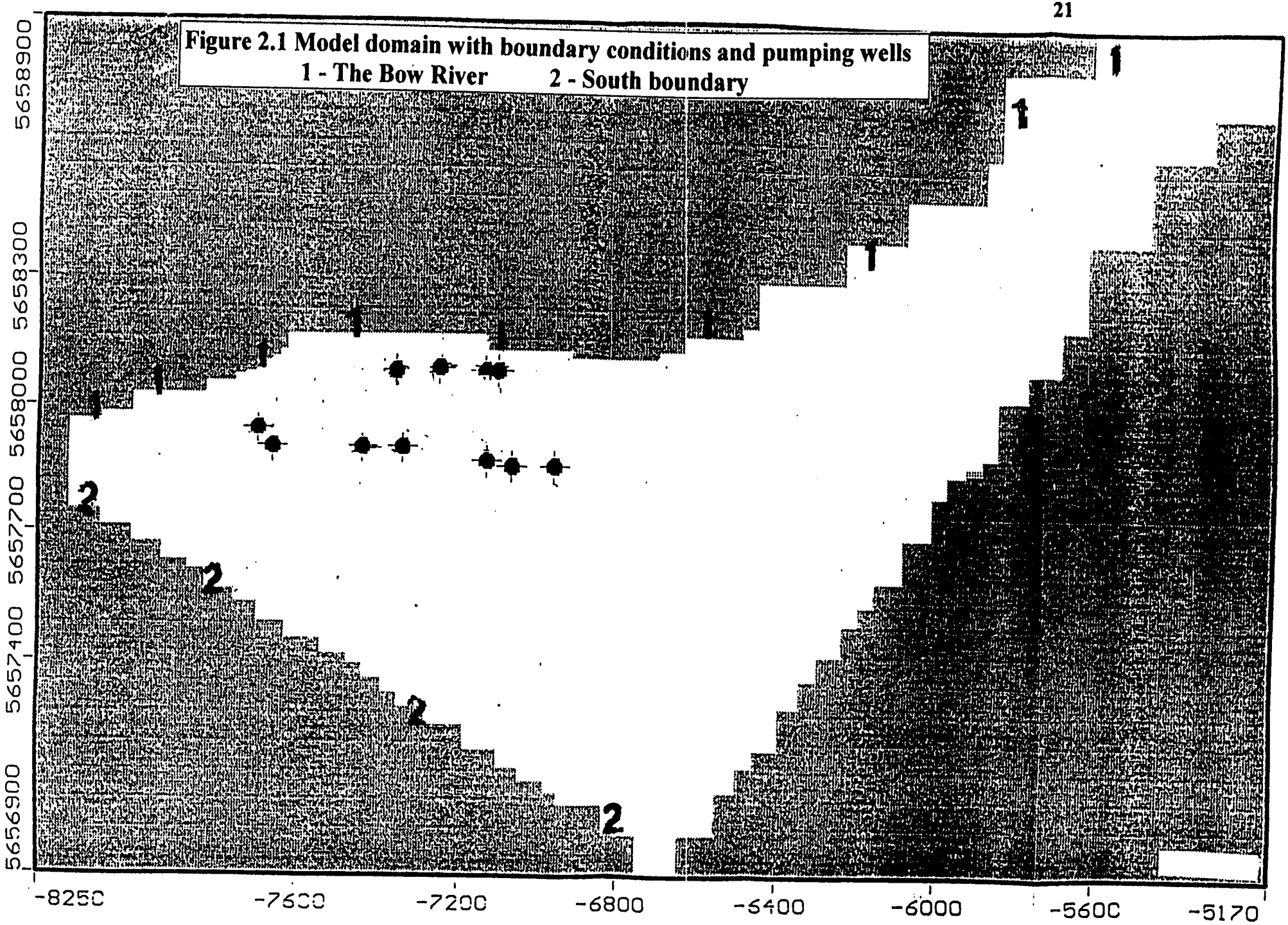


Figure 2.1 Model domain with boundary conditions and pumping wells
1 - The Bow River 2 - South boundary

Which relates "n" measurements of heads to "p" parameters of transmissivities.

Based on a model of the groundwater system with linear parameters, matrix equations between the heads and the parameters are as shown (page 51, Cooley and Naff, 1990):

$$\mathbf{Y} = \mathbf{X}(\beta) + \varepsilon \qquad (2.1)$$

Where $\mathbf{X}$ is a sensitivity matrix of n rows and p columns, each value of $\mathbf{X}$ is the system response to a unit change in the parameter,

β is a vector of true system parameters, dimensioned p by 1,

$\mathbf{Y}$ is a vector of independent variable of hydraulic heads, dimensioned n by 1,

ε is the true error or disturbance vector of the model of dimension n by 1:

the true error is random, and thus it represents the stochastic part of the model.

n is the number of measurements or observations,

p is the number of parameters or factors (transmissivity) to be optimized.

For this study, the system parameters of interest are the hydraulic transmissivities contained in the vector β, and the true system errors in the vector ε, both of which are unknown. However, the mathematical description of the groundwater flow physics through the use of the system of programs MODFLOW provides information to the sensitivity matrix $\mathbf{X}$, and measurements from observation wells at the site constitute the entries into the dependent variable vector $\mathbf{Y}$.

If estimates of β and ε can be found such that the error structure of the true model is duplicated as much as possible, then the resulting model is the best possible approximation. The optimization method used to find the error structure closest to the true model in this study is the least squares method. The least squares method is designed for an overdetermined system of equations such that the number of measurements exceeds the number of parameters to be optimized : $n>p$. Surplus measurements in overdetermined systems of equations can be used to take advantage of all measurement

information. This advantage outweighs the potential disadvantage of having outliers in the measurement data. The least squares method does not emphasize the outliers as does the Chi-squares method.

If estimators of the true parameters in vector β are those in vector **b**, and the true error vector entries in ε are estimated by elements in vector **e**, then the least squares method requires the residual vector **e** (an estimator of the error vector) to be formed (Cooley and Naff, 1990, page 52):

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{b}) + \mathbf{e} \qquad (2.2)$$

Where $\hat{\mathbf{y}}$ is the estimate vector from monitoring locations

e is the estimated residual vector

b is the estimates of the true parameters

The least square method leads to the true variance of the disturbance:

$$\sigma^2_{true} = E(\varepsilon^T \varepsilon)/n \qquad (2.2a)$$

Where E is the expected value operator,

σ^2_{true} is assumed to be unbiassed with zero mean

An estimator for σ^2_{true} is σ^2, which is defined:

$$\sigma^2 = \mathbf{e}^T\mathbf{e}/n \qquad (2.2b)$$

Where **e** is an estimate of ε.

The sum of sqaures term $\mathbf{e}^T\mathbf{e}$ is abbreviated as S(**b**). The least square procedure is one that will find the parameter set **b** that will minimize S(**b**) for all possible **b**. The value of S(**b**) from PEST is equivalent to the sum of sqaures of lack of fit (SS_{lof}) in ANOVA, which will be explained in subsection 2.4.

The S(**b**) (or Φ in PEST) is defined as the objective function and the parameter set that minimizes it will be the optimal parameter set. The indirect inverse modelling equations to find the transmissivities by minimizing head residuals in the objective function are non-linear, and iterations are required to find the parameters (Carrera and Glorioso, 1991). If the non-linear relationship is linearized by means of replacing the original mapping matrix by a Jacobian or sensitivity matrix, obtained by taking the linear term of a Taylor's expansion at some starting values of the parameters, then a system of linear equations (Cooley and Naff, 1990) results.

In the optimization process, the working parameter set has to be systematically altered to dynamically find a global minimum for the objective function. The non-linear mapping in PEST is already mentioned in the subsection, and to distinguish the incremental changes in the model estimates of $\acute{\mathbf{y}}$-the measured heads, and the parameter estimates **b**, the following notations are used:

$\acute{\mathbf{y}}_0$ is the computer model generated vector of heads at the monitoring well locations

$\mathbf{b}_0$ the corresponding parameter set used

If another estimated parameter set **b** is in the neighbourhood of $\mathbf{b}_0$, a first order Taylor's theorem expansion leads to the approximate update in the computed head values,

$$\acute{\mathbf{y}} = \acute{\mathbf{y}}_0 + \mathbf{X}(\mathbf{b} - \mathbf{b}_0) \qquad (2.3)$$

Where X is the Jacobian or sensitivity matrix.

The linearized, unconstrained objective function takes the form

$$S(\mathbf{b}) = (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0 - \mathbf{X}(\mathbf{b} - \mathbf{b}_0))^T (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0 - \mathbf{X}(\mathbf{b} - \mathbf{b}_0)) \qquad (2.4)$$

Substituting equation 2.2 into the equation for S(**b**) (Cooley and Naff, 1990, page 53):

$$S(\mathbf{b}) \approx (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0 - \mathbf{X}(\mathbf{b} - \mathbf{b}_0))^T (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0 - \mathbf{X}(\mathbf{b} - \mathbf{b}_0)) \approx (\mathbf{e}^T\mathbf{e}) \qquad (2.5)$$

$S(\mathbf{b})$ is a quadratic function in parameter (or factor) space spanned by the vector **b**. In parameter space (up to a dimension of p), the square of the length of this vector **b** (inner product) has the geometric interpretation of the square of the radius from the center of the parameter space to the quadratic surface of $S(\mathbf{b})$:

$$\mathbf{b}^T\mathbf{b} = (b_1^2 + b_2^2 + \ldots. b_p^2) \qquad (2.6)$$

Where the b_1^2 , b_2^2 up to b_p^2 are on each of the individual parameter axis. Each axis of the parameter space is equal to the inverse of the corresponding square root of the eigenvalues of the normal matrix (Weiss and Smith, 1998a). The normal matrix, a first order approximation of the Hessian (see section 2.3.1), is obtained by the differentiation of the objective function $S(\mathbf{b})$ with respect to each of the parameters (equation 2.7). There is an inverse relationship between the covariance matrix (equation 2.30a) and the normal matrix, therefore the axes are also proportional to the square root of the eigenvalues of the covariance matrix.

To minimize the objective function in equation 2.4, the quadratic function $S(\mathbf{b})$ in its linearized form above is differentiated with respect to each of the parameters and the resulting equation set to zero:

$$\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}_j} = 0 \qquad j=1.....p \qquad (2.7)$$

The actual mechanics of the differentiation can be found in (Cooley and Naff, 1990), and from the minimization, the result of the linearized upgrade vector **u** for the parameters is shown below:

$$\mathbf{u} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\acute{\mathbf{y}} - \acute{\mathbf{y}}_0) \qquad (2.8)$$

As an example, the jth parameter upgrade vector is used to calculate the next (j+1) value of the parameters:

$$\mathbf{b}_{j+1} = \mathbf{b}_j + \beta \mathbf{u}_j \qquad (2.9)$$

Where β is a scalar calculated by using a ratio containing the Jacobian matrix (PEST manual).

A diagonal weighting matrix can be used for the measurements and the prior information (if available), to reflect the importance of or confidence in the values. Without prior information, the optimization process to find the vector **b** is an unconstrained optimization. When there is prior information available, it can be incorporated into the objective function S(**b**) as linear, equality penalty functions with appropriate weights.

The forms of the weighted objective function and the parameter upgrade vector are shown below, with the generalized weighting matrix Ω :
The unconstrained, weighted objective function takes the form

$$S(\mathbf{b}) = (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0 - \mathbf{X}(\mathbf{b} - \mathbf{b}_0))^T \Omega (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0 - \mathbf{X}(\mathbf{b} - \mathbf{b}_0)) \qquad (2.10)$$

Where Ω is a diagonal weighting matrix
The parameter upgrade vector, **u**, can then be found,

$$\mathbf{u} = (\mathbf{X}^T \Omega \mathbf{X})^{-1} \mathbf{X}^T \Omega (\acute{\mathbf{y}} - \acute{\mathbf{y}}_0) \qquad (2.11)$$

To numerically eliminate the problem due to instability during the optimization which could manifest itself in fluctuations of parameter values, lower and upper bound values for each parameter to be optimized will have to be specified in the input to PEST. The bounds will prevent extremely large values of the parameters that lead to numerical failure of the program, but will not be a solution to instability itself. If there is instability, the parameter will stay at the upper and lower bound, as will be explained below.

If an element of the parameter set in equation 2.11 reaches the maximum or minimum bound values and stays there, then the corresponding component of **u** either

i) becomes large (positively or negatively) such that the parameter is cut off by the bounds,

ii) or it is zero, and the parameter is not upgraded.

The coefficients of the linear combination of $(\acute{\mathbf{y}}-\acute{\mathbf{y}}_0)$ come primarily from the sensitivity matrix **X**, the square of its inverse and their products with the weighting matrices. If the value of the element of the upgrade vector is so large that the parameter increases (or decreases) beyond the bounds, the results of the optimization are considered "insensitive", by groundwater modellers, to that parameter. The corresponding elements of the sensitivity matrix for this parameter will be small.

In the case that an element of the upgrade vector for the parameter becomes zero, the linear combinations of $(\acute{\mathbf{y}}-\acute{\mathbf{y}}_0)$ is very unlikely to be identically zero for different steps of the optimization. Hence the coefficients of $(\acute{\mathbf{y}}-\acute{\mathbf{y}}_0)$ are most likely the quantities to be identically zero for that parameter, in order for that element of the upgrade vector to vanish. In this regard, the parameter which stays at the maximum or minimum bounds is more appropriately termed "over-sensitive" than "insensitive" in the optimization process. The corresponding elements of the sensitivity matrix for this parameter will be large.

Parameter insensitivity (or over-sensitivity) is a symptom, not a cause. The instability of the parameter can be due to identifiability or uniqueness, but it can always be traced back to the system of equations in the normal matrices. One of the most common solutions to instability is to reduce the dimension of the parameter space, that is, to reduce the number of parameters to be optimized and the rank of the normal matrix. The "correct" number of parameters or factors to be included in a model of data has been studied extensively in applied factor analyses. Many rules have been proposed, some of which depend on the judgment of the researcher. For simplicity, when insensitivity of parameters occurs in this study, a reduced number of parameters is usually used for the next optimization until no instability occurs. This is somewhat similar to the "common factor" methodology in applied factor analyses (Rummel, 1970, page 104).

With the inclusion of prior information, the weighting matrix can be further refined by an augmented matrix (Cooley and Naff, 1990), consisting of Ω_s and Ω_p, for the

samples (or measurement) and parameters respectively. Similarly, the residual errors **e** in equation 2.2 will also have components of $\mathbf{e}_s$ and $\mathbf{e}_p$. The form of the objective function S(**b**), with the use of prior information for the parameters, is therefore redefined as follow:

$$S(\mathbf{b}) = \mathbf{e}_s{}^T \Omega_s \mathbf{e}_s + \mathbf{e}_p{}^T \Omega_p \mathbf{e}_p \qquad (2.12)$$

The assumption of zero correlation between the prior information on the parameters and the measurements (or samples) is tacit in the above equation, since the weighting matrices are in block diagonal form.

In the present study, the use of prior information on the hydraulic transmissivity parameters stems from a new screening sequential process. The theoretical part of the use of weights in the optimization process is described in the next subsection. The theory will be applied to scenario data of the former Canadian Creosote site in sections 3, 4 and 5. The weights used are the inverse of the estimated variances (Yeh, 1986) of the parameters obtained from prior optimization step(s). The use of weights is discussed in section 2.2.

2.1.1 Confidence limits and weights for prior information on parameters

This subsection shows the relationship between the PEST output of the confidence limits, and the calculation of the prior information weights for the parameters in the matrix Ω_p .

The variance due to the model lack of fit, which is computed by using the error estimator **e** in the objective function S(**b**) from section 2.1, can be approximated by $\sigma_{ref}{}^2$, which, in its simplest form, is equal to

$$\sigma_{ref}{}^2 = S(\mathbf{b})/(n'-p) = \mathbf{e}^T\mathbf{e}/(n'-p) \qquad (2.13)$$

Where S(**b**) is the objective function at the solution

n' is the number of measurements and prior information

p is the number of parameters to be optimized. For this study, the number

of prior information is equal to the number of parameter to be optimized in a sequential procedure.

σ_{ref}^2 is the "reference variance".

When the measurements are independent and the weighting matrix is an identity matrix, the variance and covariance matrix Cov(**b**) is:

$$\text{Cov}(\mathbf{b}) = \sigma_{ref}^2 (\mathbf{X}^T\mathbf{X})^{-1} \qquad (2.14)$$

The diagonal terms of this matrix Cov (**b**) are the estimates of the variances of each of the p parameters. The diagonal terms are denoted by s_{bi}^2, i from 1 to p. The confidence limits for the parameter b_i, at the 95% confidence level are

$$b_i \pm (F_{95\%}(1,n'-p)\, s_{bi}^2)^{1/2} \qquad (2.15)$$

Where $F_{95\%}(1,n'-p)$ is the value of the F distribution.

The linearized confidence interval, is based on the linearized least square procedure. In spite of being a linear projection of the variance of the parameter, the confidence interval reflects the conditioning of the normal matrix. Large confidence limits imply an ill-conditioned matrix. As a confirmation, the values of the largest and smallest eigenvalues of the normal matrix show whether it is ill-conditioned or not. Ill-conditioned matrices also give rise to severe convergence problems and unstable solutions in the inverse modelling. This will be further explored in section 2.3.

2.2 **Variances and the use of weights**

The special challenge this study faces is the assessment of hydraulic transmissivity for the zones of the region of study based on data collected at different times of the year and at different locations of monitoring wells, in the six scenarios. Since the area of study is the same model region for all six scenarios, the hydraulic transmissivities for the same

zones should be the same for all scenarios (Figures 1.4a to 1.4f). Due to the differences in time and monitoring locations, the optimized hydraulic transmissivity for the same zone is different for individually optimized scenarios.

A new sequential forward feeding procedure (sffp) is used to find a common set of parameters that will minimize the total $S(\mathbf{b})$ or SS_{lof} for all six scenarios. The procedure sequentially optimizes parameters of single scenarios through a random sequence of the six scenarios, beginning with an unconstrained optimization of the first scenario. The unconstrained results of the first scenario are then passed onto a constrained optimization, again, of the first scenario as prior information of the decision variables (parameters). Prior information has been introduced in section 1.2.

There are arguments both for and against generating prior information and inverse modelling of parameters using the same set of hydraulic heads. On the one hand, the unconstrained optimization results are not usually used as prior information again for the same model in a subsequent optimization. This has been explained in (Carrera and Neuman, 1986a, page 203): The prior head errors and prior parameter estimate errors are assumed to lack cross-correlation, and theoretically head values used for inverse modelling should not be used to derive prior estimate for parameters.

However, a procedure similar to the first step of the above sffp is used by Theil (Cooley and Naff, 1990, page 74, Theil, 1963) for a linear model. It was found that the estimates of the parameters obtained by using the unconstrained minimization had a bias in the order of the square root of the reciprocal of the number of measurements (see section 2.2.1). The subsequent constrained optimization using prior information from the first run provided solution to the complete problem. Since the optimization of the sffp is based on a linearized procedure, it should perform well in providing the solution to the constrained problem with the self-generated prior information. So, the same set of hydraulic heads are used for inverse modelling as well as generating prior information of the parameters in the first step of the sffp.

After the first step, the results of the constrained optimization are used as prior information for the optimization of the same set of decision variables of a second sequence scenario, and so on (see schematic chart 2.0 below). The total SS_{lof} for all six

scenarios based on each of the optimized parameter set from the sequential steps are compiled. The set of optimized parameters that gives the lowest overall SS_{lof} will be the chosen set of optimized parameters to be used for all six scenarios.

The true errors ε in equation 2.1 can be separated into measurement or sample error, ε_s and prior information error, or ε_p , in the same manner as the error estimator **e** at the end of section 2.1. The variances for these errors are in augmented matrices shown below:

$$\mathrm{Var}\begin{bmatrix} \varepsilon_s \\ \varepsilon_p \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(\varepsilon_s) & \mathbf{0} \\ \mathbf{0} & \mathrm{Var}(\varepsilon_p) \end{bmatrix} \qquad (2.16)$$

The assumptions for the true errors are that they have a normal distribution with the same variance and zero mean. While some of the errors for the parameters may be correlated, and the actual error statistical distribution may not be truly normal, it is nevertheless a good working assumption and a convenient starting point for analysis. The assumed form of the variances for the two types of true errors are:

$$\mathrm{Var}(\varepsilon_s) = \mathbf{V}_s \sigma^2_{com} ,$$

$$\mathrm{Var}(\varepsilon_p) = \mathbf{V}_p \sigma^2_{com} \qquad (2.17)$$

Where $\mathbf{V}_s$ is the scaled variance-covariance matrix for the samples or measurements, is symmetric and positive definite and of order n,

$\mathbf{V}_p$ is the scaled variance-covariance matrix for the prior information, is symmetric and positive definite and of order p for the sequential forward feeding procedure.

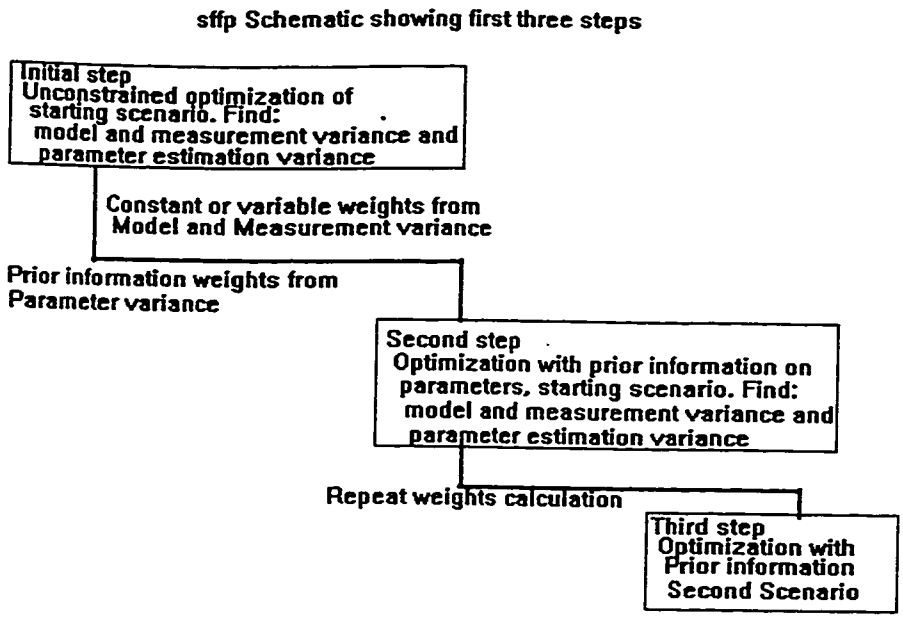


Chart 2.0 Flow chart of the sequential forward feeding procedure (sffp)

Equation 2.16 now becomes

$$\mathrm{Var}\begin{bmatrix}\varepsilon_s\\ \varepsilon_p\end{bmatrix} = \sigma^2_{com}\begin{bmatrix}\mathbf{V_s} & \mathbf{0}\\ \mathbf{0} & \mathbf{V_p}\end{bmatrix} \qquad (2.18)$$

Where σ^2_{com} is defined as the common variance for the transformed disturbance of $\{\mathbf{V}^{-1/2}\varepsilon\}$ (Cooley and Naff, 1990).

In (Cooley and Naff, 1990), the common variance for both the samples or measurements and the items of prior information is the same, and the distinction lies in the variance-covariance matrices which are in blocks of $\mathbf{V_s}$ and $\mathbf{V_p}$ in an augmented matrix $\mathbf{V}$. For this study, the common variances for the transformed or untransformed disturbances refer only to the samples or measurements. The variances for the items of prior information has been introduced in section 2.1.1 as the reciprocals of the elements of Ω_p .

The sample common variance relationship in equation form, is

$$Var(\mathbf{V}^{-1/2}\varepsilon) = \mathbf{I}\,\sigma^2_{com} \qquad (2.19)$$

Where **I** is the identity matrix

V is the variance-covariance matrix (RHS matrix of equation 2.18)

ε is the error vector (LHS vector of equation 2.18)

From equation 2.19, the value for the common variance is

$$\sigma^2_{com} = E((\mathbf{V}^{-1/2}\varepsilon)^T(\mathbf{V}^{-1/2}\varepsilon))/(n'-p)$$

or $$\sigma^2_{com} = E(\varepsilon^T \mathbf{V}^{-1} \varepsilon)/(n'-p) \qquad (2.20)$$

Where E is the expected value operator,

n' is the number of measurements and items of prior information,

p is the number of parameters to be optimized

Equation 2.20 will mean that if an estimator of the error vector is used, then the sum of sqaures term S(**b**) following equation 2.3 and the form of equation 2.20, will become

$$S(\mathbf{b}) = \mathbf{e}^T\ \mathbf{V}^{-1}\ \mathbf{e} \qquad (2.21)$$

The sum of sqaures term can be generalized to

$$S(\mathbf{b}) = \mathbf{e}^T\ \Omega\ \mathbf{e} \qquad (2.22)$$

Where Ω is a general diagonal positive definite weighting matrix

The weighted least square formulation of the objective function is equation 2.22, for optimal weighted least square optimization (Cooley and Naff, 1990),

$$\Omega = \mathbf{V}^{-1} \qquad (2.23)$$

Unfortunately, **V** is never known so we must estimate it to the best of our ability (Carrera and Neuman, 1986a).

Again, the form of the sum of sqaures term showing both sample or measurement and prior information weighting is (also section 2.1) :

$$S(\mathbf{b}) = \mathbf{e}^T_s \, \Omega_s \, \mathbf{e}_s + \mathbf{e}^T_p \, \Omega_p \, \mathbf{e}_p \qquad (2.24)$$

If the measurements are totally uncorrelated, and all the errors have the same normal distribution with the same variance, then equation 2.19 becomes

$$\mathrm{Var}(\varepsilon) = \mathbf{I} \, \sigma^2_{comu} \qquad (2.25)$$

Where the **V** matrix (that is, $\mathbf{V_s} + \mathbf{V_p}$ in equations 2.17) becomes an identity matrix,

σ^2_{comu} is the common variance of the untransformed disturbance, ε.

The transformed disturbance is shown in equation 2.19.

The σ^2_{comu} is equal to σ^2_{com} if the variance-covariance matrix is equal to the identity matrix (assumption of independence).

2.2.1 Weighting Matrices for the Sequential Feed Forward Procedure (sffp)

With the foregoing discussion, the weighting strategy for the sequential feed forward procedure (sffp) can now be addressed. The weighting for the prior information and measurements are discussed separately.

I) Weights for prior information on the decision variables of logarithmic hydraulic transmissivities, log Kx

The logarithmic values of the hydraulic transmissivities are used in the optimization to increase the speed and stability of the process (PEST manual). The logarithmic hydraulic transmissivities are assumed to have a normal distribution (a

property of the lognormal distribution, Ang and Tang, 1990). The error for the log Kx is assumed to be normally distributed with zero mean and equal variances. At the start of any random sequence in the sffp, the magnitude of the variances of the log Kx is unknown. The first step in the sffp is to perform an unconstrained (with no prior information) optimization to estimate both σ^2_{com} for the samples and the variance for each parameter. These estimates are outputs available from the program PEST.

A similar procedure was performed by (Theil 1963) to estimate the common variance and to solve the normal equation from least square. Theil found that the unconstrained optimization using least square (ie the first step in sffp) may create a bias on the estimates of the order of $(1/n)^{1/2}$. With the measurement sample size range for the six scenarios in this study, the possible bias range could be 0.2 to 0.3 m2/day in the parameters values.

After the initial step, the weighting matrix for the prior information on the decision variables of subsequent sffp steps is set to the reciprocal of the parameter variance from the previous step.

II) Weights for the measurements of hydraulic head at monitoring wells

The magnitudes of $\mathbf{V_s}$ and therefore Ω_s must be approximated. Two approaches are explained for approximating Ω_s.

The first approach- constant Ω_s:

The variance $\mathbf{V_s}$ of all the measurements is the same and related to the accuracy of the physical measuring devices. The measurement error is assumed to be N(0, 0.01) and uncorrelated between measurement points. The chosen variance reflects the physical limit of the accuracy of the measuring devices, and the use of normal distribution with zero mean conforms to previous assumptions . The weighting matrix is :

Ω_s for all steps in first approach = $(1/0.1)^2$ **(I)**

$$\Omega_s = 100(\mathbf{I}) \qquad (2.26)$$

The weighting matrix will therefore have diagonal terms of 100, for all measurements and all steps in the sffp.

The assumption of uncorrelated measurements is a good one, since spatial correlation of measurements is weak at worst (Carrera and Neuman, 1986a).

However, applying the reciprocal of the assumed measurement error variance to the measurement term in S(**b**) in equation 2.24 as weight is not completely accurate. The so called measurement error is actually the sum of model and measurement error (see equation 2.3). As well, the mathematical process in equation 2.3 filters out the measurement noise, which is assumed to be normal with zero mean (Carrera, 1988, page 567, Carrera and Neuman, 1986a), and thus de-emphasizes the measurement error variance.

The second approach- Variable weighting matrix Ω_s:

Step 1) For the initial step in the sffp, the measurement or sample common variance is unknown, but with the assumption of measurement independence, equation 2.25 is used. This also means that the variance matrix $\mathbf{V}_s$ for this approach is equal to the identity matrix **I**. A unit weight is used for the measurements in this step.

For ease of reference, a counting index k is used to indicate the steps of the second approach. For the unconstrained optimization step, k = 0.

Step2 to 6) From the results of step 1, the variance of the measurements is estimated by :

$$\mathrm{Var}(\mathbf{e}_s) = \mathbf{V}_s \sigma^2_{comu} = \mathbf{I}\, s^2 \qquad (2.27)$$

Where $\mathbf{e}_s$ is the estimator of the true error vector,

σ^2_{comu} is the common variance of the untransformed disturbance ε,

$\mathbf{V_s = I}$, covariance matrix of measurements is equal to the identity matrix for independent measurements

$$s^2 = (S(\mathbf{b})/n'-p) \approx \sigma^2_{comu} \qquad (2.28)$$

$S(\mathbf{b})$ is the value of the objective function at the optimum point,

n' and p are defined as in equation 2.20

The inverse of the estimated sample error variance $s^2\,\mathbf{I}$, from the first step, is utilized as a diagonal weighting matrix in the second step. The weighting matrix is then updated for subsequent steps:

$$\Omega_{s\,k+1} = (\mathbf{I}\, s^2{}_k)^{-1} \qquad (2.29)$$

For $k = 0$ to 5

Prior to discussing the variable weights used for the samples, the covariance matrix of the model parameter estimates from a linearized error analysis, with prior estimate for the parameters, is presented (Carrera and Neuman, 1986a, page 205):

$$\Sigma_p = [\,(1/s^2)\,(\mathbf{X}^T\,\mathbf{V_s}^{-1}\,\mathbf{X}) + \mathbf{C_p}^{-1}]^{-1} \qquad (2.30)$$

Where Σ_p is the parameter covariance matrix with prior information on the parameters

$\mathbf{V_s}$ is defined in equation 2.27 for the measurements or samples

$\mathbf{X}$ is the Jacobian matrix

$\mathbf{C_p}$ is the joint covariance matrix of the prior parameter estimates

The parameter covariance matrix Σ_p is made up of two components: the sample or measurement variance and the parameter variance respectively . For the sequential procedure sffp, an estimate of the second component of the parameter prior information covariance matrix is known at the beginning of the optimization. It is the covariance matrix of the estimated parameters of the previous step. The diagonal terms of the parameter covariance matrix of the previous step are used to form weights

for the prior information input to the present step - a procedure to be described in section 4.

To examine the first component due to the sample variance, the first term in the bracket in equation 2.30 is singled out. This term is denoted by Cov(**b**). In looking at this term alone, the overall parameter estimate covariance is equivalent to having no prior parameter information.

The importance of the weighting in the sequential procedure for the samples is shown by the formula for Cov(**b**). It is a measure of the quality of parameter estimation. Its value should be small when the estimates have low uncertainties. Hence, with reference to the sequential process sffp,

$$\text{Cov}(\mathbf{b}) = s^2_{k+1} (\mathbf{X}^T \mathbf{V_s}^{-1}_{k+1} \mathbf{X})^{-1} \qquad (2.30a)$$

Where Cov(**b**) is the estimate of the covariance matrix of the parameters based on sample variance alone

$\mathbf{V_s}$ is defined above for equation 2.30

k = 0 to 5

However, with the measurements assumed independent, $\mathbf{V}_{s\,k+1}$, for k=0 to 5, is equal to the identity matrix (see equation 2.27).

When the weight of equation 2.29 is applied to 2.30a, the covariance matrix of estimated parameters is:

$$\begin{aligned} \text{Cov}(\mathbf{b}) &= s^2_{k+1} (\mathbf{X}^T \Omega_{s\,k+1} \mathbf{X})^{-1} \\ &= s^2_{k+1} (\mathbf{X}^T (\mathbf{I} s^2_k)^{-1} \mathbf{X})^{-1} \\ &= (s^2_{k+1} / s^2_k) (\mathbf{X}^T \mathbf{X})^{-1} \end{aligned} \qquad (2.31)$$

This procedure has a normalizing or scaling effect on Cov(**b**). If the estimated sample common variance for the untransformed disturbance σ^2_{comu} (or its estimate, s^2)of the present step (or k+1) is the same as that for the prior step, k, (which is unlikely as the number and locations of the monitoring wells are different from scenario to scenario), then the error variance estimate for the present step will be normalized to unity:

When this happens, Cov(**b**) of the estimated parameters depends only on the inverse of the product of the Jacobian matrix and its transpose in equation 2.31. If the elements of the Jacobian matrix are large, then the uncertainties of the estimated parameters are small. The effects of the common variance of the samples are controlled.

The overall parameter covariance, Σ_p, based on both the sample and the prior information covariance, will be reduced when both RHS components of equation 2.30 are controlled. In addition to the above variable weights for the samples, separate weights equal to the inverse of the parameter prior information covariance matrix diagonal terms of the previous step are used for the parameter estimation, in section 4. The overall values of Σ_p, the parameter covariance, are small, and an overall improvement of the parameter estimate is achieved.

This is an innovation in applying a concept from manufacturing optimization (Taylor, 1991) to groundwater modelling. The variance of a manufacturing process is reduced to improve on the process optimum (see also section 1.2). Even though groundwater flow is not a factory process that can be controlled to a large extent, the sequential optimization in this study provides an ideal situation to control the variance through the steps. Even when the s^2 for the two steps are different, the method is still thought to be somewhat effective because the sample common variance for the untransformed disturbance from scenario to scenario is not expected to be drastically different.

The normalizing procedure will achieve a scaling effect in bringing the magnitude of the weighted error variance estimate closer to unity. This may enhance the parameter estimation accuracy of the calculations better than the constant value of 0.01 for $\mathbf{V_s}$ (see 2.4.1-I) from the first approach: The magnitude of the weight here changes with the steps in sffp, according to the sum of squares of the particular scenario run. More importantly, the uncertainty of the estimated parameters is controlled while the optimization algorithm searches for the smallest SS_{lof}, thus both objectives of decreasing the lack of fit error and the parameter uncertainty can be achieved. This is an improvement over individual optimization, in which the use of prior information only improve the parameter accuracy,

but at the expense of the model fit to the data. In fact, the use of prior information can "only make the model fit worse" (Yeh, 1986).

The differences between the two approaches are obvious from the above discussion. The use of a constant $\mathbf{V_s}$ for the first approach is to use the best understanding of the measurement accuracy to maintain an unbiased estimate of the variance of the transformed disturbance. The second approach attempts to control the weighted estimate of the variance of the untransformed disturbance from step to step.

The effectiveness of the approaches will be tested empirically in sections 3 and 4.

A scaling procedure using s^2 in equation 2.28 above has been applied to a first order approximation of the Hessian matrix in an independent study by (Weiss and Smith 1998b). It should be pointed out here that the Weiss and Smith method is used in the scaling and the construction of an approximation of the Hessian and its inverse, which follows classical theories in the improvement of convergence and eigenvalue structure of the approximation (see Luenberger, 1984, page 275). The application is fundamentally different from the intention of the method applied here in the sequential process. It will be useful to compare the eigenvalue ratios for approaches I (constant weights) and II (variable weights) in sequence 1 to see if there is an improvement in the eigenvalue structure. This will be carried out in section 6.

2.3 **The eigenvalue and eigenvector analysis and the noise problem**

Recall in section 2.1 the equations used in PEST are outlined, and both the parameter upgrade vector and the covariance matrix depend on the inverse of the normal matrix. The key equations are repeated here for completeness:

The parameter upgrade vector is:

$$\mathbf{u} = (\mathbf{X}^T\Omega\,\mathbf{X})^{-1}\,\mathbf{X}^T\Omega(\acute{\mathbf{y}} - \acute{\mathbf{y}}_0)$$

The parameter covariance matrix is

$$C(\mathbf{b}) = \sigma^2_{com} \ (\mathbf{X}^T \Omega \ \mathbf{X})^{-1}$$

Where $(\mathbf{X}^T \Omega \ \mathbf{X})$ is the normal matrix

The parameter upgrade vector is derived directly from the least square formulation. While the **X** matrix (Jacobian) is not symmetric, the normal matrix is symmetric. Its eignevalues will be real and positive (except when the normal matrix is ill-conditioned to the extent that it is almost singular, then no solution can be found).

Solution of matrix equations that requires inversion of a large matrix is a cumbersome procedure, especially when the procedure involves physical measurements such as $\acute{\mathbf{y}}$. These physical measurements which are often limited to an accuracy of not more than 0.1m. Small fluctuations of the measurements (noise) could be exaggerated to an unacceptable degree when the upgrade vector is obtained by multiplying the inverse of the normal matrix.

Lanczos (1964) gives an eigenvalue checking procedure to see whether there is a potential noise problem. If the weighting matrix Ω in the normal matrix is an identity matrix, the upgrade vector becomes

$$\mathbf{u} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T(\acute{\mathbf{y}} - \acute{\mathbf{y}}_0) \qquad (2.32)$$

Suppose **X** and **u** are both transformed into an orthogonal reference system by means of the matrix U which contains the eigenvectors of $\mathbf{X}^T\mathbf{X}$ as columns (the eigenvectors are also output of PEST):

$$\underline{\mathbf{u}} = \mathbf{U}^{-1} \mathbf{u} \qquad (2.33)$$

$$\underline{\mathbf{X}} = \mathbf{U}^{-1}\mathbf{X} \mathbf{U} \qquad (2.34)$$

Where $\underline{u}$ and $\underline{X}$ are now both in the orthogonal reference system. Equation 2.32 becomes

$$\underline{\mathbf{u}} = (\underline{\mathbf{X}^T\mathbf{X}})^{-1} \underline{\mathbf{X}^T}(\acute{\mathbf{y}} - \acute{\mathbf{y}}_0) \qquad (2.35)$$

Where $(\underline{\mathbf{X}^T\mathbf{X}})$ is now a diagonal matrix with the eigenvalues $e_i^2 = e_1^2 \ldots\ldots e_p^2$,

p being the number of parameters to be optimized.

If $\underline{X}_i$ is a column of $\underline{X}$, then the i^{th} component in the product of $\underline{X}^T(\acute{y} - \acute{y}_0)$ becomes,

$$\underline{X}_i(\acute{y} - \acute{y}_0) = e_i (\acute{y} - \acute{y}_0) \cos \theta_i \qquad i = 1 \ldots p \qquad (2.36)$$

Where θi is the angle between $\underline{X}_i$ and $(\acute{y} - \acute{y}_0)$. Equation 2.35 becomes

$$\underline{u} = (\acute{y} - \acute{y}_0) \cos \theta_i / e_i \qquad (2.37)$$

Consequently, the value of the upgrade vector is dependent on the magnitude of the eigenvalue e_i. If the small eigenvalue is extremely small compared to the largest eigenvalue, then a small fluctuation in $(\acute{y} - \acute{y}_0)$ will cause a large change in the upgrade vector. The criterion for determining whether there is a potential noise problem is the square root of the ratio between the largest and smallest eigenvalues. The magnitude of the square root term, when compared to the accuracy of the measurements, can indicate whether noise will be amplified. For instance, in our present problem, assuming the accuracy of the measurement is in the order of 0.1m, and the magnitude of the heads at the monitoring wells are at approximately the 1000 m mark, there could be a noise problem if the square root of the eigenvalue ratio is 10^4 or higher. A slight error could be blown up to the same order of magnitude as the head measurements.

An example of the eigenvalue ratio is shown in section 6.

2.3.1 Penalty functions and the associated eigenvalue analyses

An understanding of the background for the use of prior information and weights for the sequential procedure can be obtained by examining the classical theories (from Luenberger,1984, Lanczos 1964, Gill et al. 1981) of penalty functions, eigenvalue and eigenvector analyses of the resulting matrices.

Penalty functions can be used to form an unconstrained optimization which approximates the effects of a constrained optimization. The approximation is accomplished by adding to the objective function a term that prescribes a high cost of for

the violation of the constraint. The general form of a penalty function optimization process is:

$$\text{minimize } q(c,\mathbf{x}) = f(\mathbf{x}) + cP(\mathbf{x}) \qquad (2.38)$$
$$\text{wrt } \mathbf{x}$$

Where $f(\mathbf{x})$ is a continuous function

c is a penalty parameter that determines the severity of the penalty, and the degree of the approximation to the constrained problem. As the magnitude of c approaches large values, the approximation to the original constrained problem becomes more accurate.

$P(\mathbf{x})$ is the penalty term that is a positive, continuous function, and it becomes zero inside of the feasible region where the condition is not violated.

The above equations can be used as a theoretical basis for a discussion on the sequential process to be detailed in chapters 3 and 4. The prior information that was used from file to file for the hydraulic transmissivities, carries with it penalty parameters. When the penalty parameter 'c' increases in value in the sequential process (see illustration below, excerpted from section 6), the unconstrained optimization approximates the constrained optimization closely:

The penalty parameter c (Wp^2) input to steps, seq 1, 14 zones model

log Kx no.	step 1	step 2	step 3	step 4	step 5	step 6
6	33.99	55.25	398.76	739.86	333.10	127.47

The sums of squares residual for the model fit can be easily obtained by applying the analysis of variance (section 2.4) to the resulting model outputs from the later sequential steps.

2.4 Assessing the Results of the Optimization

In the design of experiments it is common to assess the results of the experiments and the model performance by comparing them to the actual measurements. The analyses of the variance (ANOVA) between the experimental results and observations will then show whether the model assumed by the experimenter is adequate to describe the physics.

In the present application of inverse modelling techniques, different sets of optimized parameters produce different model outputs at the monitoring locations, which can then be compared to the observations. It is logical, then, to examine the model output compared to the observations in a similar manner to the design of experiments, so that information regarding how well the model results fit the measurements, the precision of the measurements and the adequacy of the parameters used in the model can be analysed. By the analyses of these variances, the best parameter set can be found, by inverse modelling, on the basis that it can reduce the overall sum of squares of the lack of fit between model results and measurements, in well-defined ways.

This approach can be visualized by assuming the design of experiments proceeds in one direction, ending in the ANOVA and the verification of the model. Proceeding in a different direction is the inverse modelling technique, optimizing the parameter sets to produce a minimum residual, then ending at the ANOVA task of examining the performance of the model and the parameters:

This section summarizes the techniques used in the ANOVA, and is excerpted from an experimental design text (Deming and Morgan, 1987).

2.4.1 Analysis of Variance (ANOVA) for linear models - Sums of squares

The different sums of squares of the data and model output can be classified into the following categories, each category represents a stage in response space (Deming and Morgan, 1987). The relationship of these stages will be summarized at the end. A schematic relationship of ANOVA to optimization and experimental designs is shown in chart 2.1.

I) The total sum of squares, SS_T

SS_T is defined as the sum of squares of the measurements or observations. If $\acute{y}$ is the vector of measurements, then

$$SS_T = \acute{\mathbf{y}}^T\acute{\mathbf{y}} \qquad (2.39)$$

The degree of freedom associated with SS_T is equal to the number of observations.

II) The sum of squares due to the mean, SS_{mean}

This is defined as the sum of squares of the mean measurements. If $\underline{\mathbf{Y}}$ is the vector of mean measurements, obtained when the sum of the measurement values is divided by the number of measurements, then,

$$SS_{mean} = \underline{\mathbf{Y}}^T\underline{\mathbf{Y}} \qquad (2.40)$$

This sum of squares has one degree of freedom.

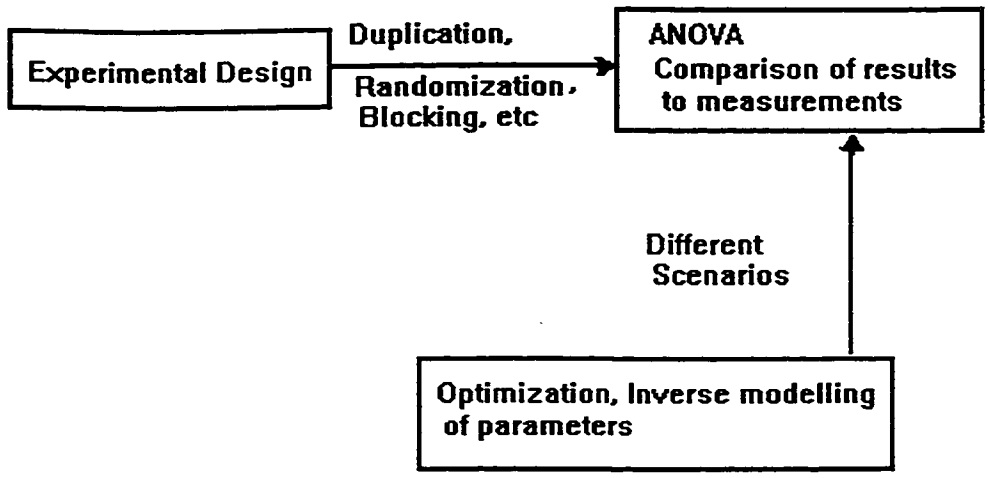


Chart 2.1 The use of ANOVA

III) The sum of squares corrected for the mean, SS_{corr}

The mean of the measurements can be subtracted from each of the individual measurements to produce the measurements corrected for the mean:

$$SS_{corr} = (\acute{\mathbf{y}}-\underline{\mathbf{Y}})^T(\acute{\mathbf{y}}-\underline{\mathbf{Y}}) \qquad (2.41)$$

The degree of freedom for this sum of squares is the number of observations minus one.

The sum of squares and their degrees of freedom are additive:

$$SS_T = SS_{mean} + SS_{corr} \qquad (2.42)$$

IV) The sum of squares due to factors or parameters, SS_{fact}

The vector of model outputs of hydraulic heads is $\ddot{\mathbf{y}}$. The sum of squares of factors or parameters is:

$$SS_{fact} = (\ddot{\mathbf{y}}-\underline{\mathbf{Y}})^T(\ddot{\mathbf{y}}-\underline{\mathbf{Y}}) \qquad (2.43)$$

For linear models the number of degrees of freedom could be the number of parameters minus one, p-1, or p, depending on whether the model contains a constant term or not.

V) The sum of squares of residuals, SS_r

The sum of squares of the residuals is made up of two parts: the sum of squares of lack of fit, SS_{lof} ,and the sum of squares of purely experimental uncertainty, SS_{pe}. SS_{pe} requires results from duplicated experiments to be properly assessed. Based on stable, identical simulations conducted in this study, results from identical simulations are the same, so SS_{pe} is zero. Assuming SS_{pe} to be negligible, the sum of squares of residuals is therefore identical to that for the lack of fit:

$$SS_r = SS_{lof} = (\acute{\mathbf{y}} - \ddot{\mathbf{y}})^T(\acute{\mathbf{y}} - \ddot{\mathbf{y}}) \qquad (2.44)$$

The degree of freedom associated with this term is the difference between the number of observations and the number of parameters.

The sums of squares are again additive, as are their degrees of freedom (Chart 2.2):

$$SS_{corr} = SS_{fact} + SS_r \qquad (2.45)$$

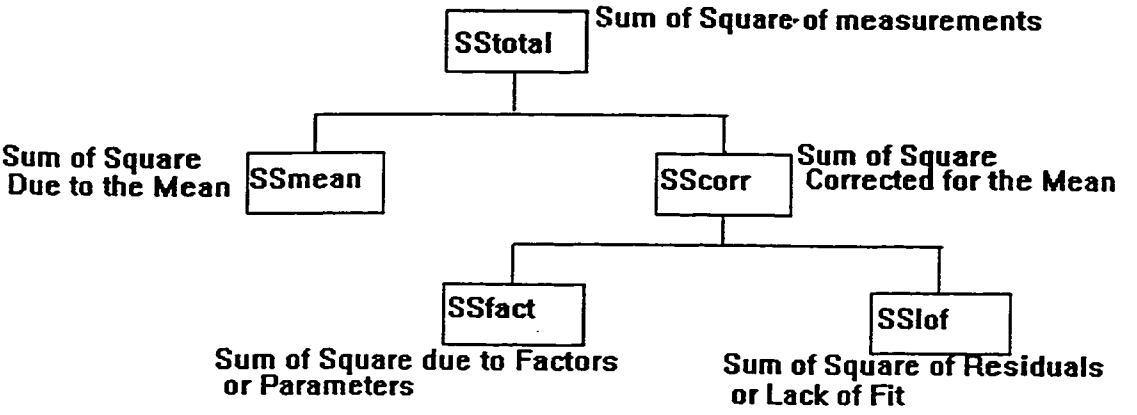


Chart 2.2 illustrates the relationship of the sums of squares:

3.0 The optimization of the hydraulic transmissivity parameters of the former Canada Creosote site

The former Canada Creosote site is underlain by sand and gravel fill (1 to 6.7 m thick), and alluvial sands and gravel up to 5 m thick. Bedrock is at a typical depth of 6 to 9 m and is comprised of sandstone and shale. The water table is situated within the alluvial sands and gravel, at a depth of 4 to 7m. The Bow River borders the northern edge of the site and the direction of groundwater flow is towards the Bow, except during ice jamming events on the river. The creosote (Non Aqueous-Phase Liquids), occurs in both heavier and lighter than water phases, and lies within the alluvial sands and gravel. The total volume of the NAPL is approximately 4750 cu m (Golder and Associates, 1990).

The containment system for the pollutants consists of a cofferdam, which was removed in March, 1997, a partially enclosing bentonite-slurry and secant-pile wall and a pumping system which controls the groundwater flow into the area. The pumping system is comprised of clean water (CW) pumping wells, which minimize the inflow of clean groundwater into the area in which the plume of creosote is located; and dirty water (DW) wells which extract the groundwater from the contaminated area for a dirty water treatment plant (DWTP) (CH2M Gore & Storrie Ltd, 1997). The treatment plant removes the creosote from the water and discharges the cleaned water. The efficient operation of the system depends on the groundwater flow which in turn depends on the hydraulic transmissivity of the natural and fill material at the site. Despite efforts to directly measure the hydraulic transmissivities, much interpretation is required for the assignment of hydraulic transmissivities to the zones across the site. An alternative is to calculate these constitutive variables from inverse modelling as outlined in section 1.2.

In this inverse modelling study, numerical models are used to estimate the hydraulic transmissivities of the aquifer. As mentioned in section 1.2, parameter identification through inverse modelling is the process of selecting an optimal set of parameters by using observations of the dependent variable (hydraulic heads at monitoring wells). The number of observations is limited and finite, whereas the spatial domain is continuous and the dimension of its hydraulic transmissivity parameter is

theoretically infinite. The reduction of the number of parameters from an infinite dimension to a finite number of zones is the process of parameterization (Yeh, 1986).

The optimization program PEST, when faced with a large number of parameters to be optimized against a limited number of measurements or observations, will conclude the search of the optimized parameter set with a number of the parameters fixed at the upper or lower bounds. The parameters are unidentifiable with the available data. This is an indication that the parameters are insensitive to the available data. The background of this occurrence is explained in section 2. There is a need to screen and to limit the number of hydraulic transmissivities used as decision variables for the optimization process.

There are six scenarios of steady state events at different times of the year, which were collected over several years:

The analyses of the optimized results begin with screening test runs for the different scenarios. These cases can be regarded as the models of the same spatial domain, each having different boundary conditions and stresses for the particular time during which the head measurements were made. The number and spatial locations of the measurements are different for each scenario due to changes in operations and installation and destruction of wells over the years. A meticulous manual calibration was performed on the well data of six separate sampling events (table 3.1) by the staff and consultants of CH2M G&S to correspond to high and low river stages (CH2M G&S reports, 1996,1997).

The six scenarios are: before the installation of the containment wall, after the containment wall and after the wall with clean and dirty water pumping wells. Each corresponds to distinct sampling event. All scenarios to be discussed are modelled with steady state simulations.

Table 3.1 Models of the same spatial domain at the former Canada Creosote Site

Scen1	high river stage before the wall was installed. Also known as the hi-pr-wa file
Scen2	low river stage before the wall was installed. Also known as the

	lo-pr-wa file
Scen3	high river stage after the wall was installed. Also known as the hi-wal file
Scen4	low river stage after the wall was installed . Also known as the lo-wal file
Scen5	high river stage after the wall was installed with CWs and DWs pumping*. Also known as the june17-h file
Scen6	low river stage after the wall was installed with CWs and DWs pumping*. Also known as the oct-p-hk file

*CW denotes clean water wells; DW denotes dirty water wells

Scenarios Scen1 and 2 contain data collected before there was any remedial measure to contain the creosote, and at a time when the data collection system was not as complete as the other scenarios. Therefore the quality and the quantity of data were somewhat inferior to those of Scen3 to 6. In the simulation and analyses that follow, all six files are used, with an emphasis on the last four.

3.1 The expert input, log transformation, identifiability of transmissivities and the analyses of variance (ANOVA)

All the simulation scenarios have transmissivities produced by hydrogeologists after examining the site information. These values are the expert input, to distinguish them from the values generated by optimization. The screening and systematic reduction of the number of parameters to be optimized takes several steps.

The first step is to reduce the number of transmissivity zones to be optimized in order to eliminate the unidentifiable and insensitive parameters (section 1, 2). Parameter or factor selection and factor dimension reduction is not a simple task, and often it is achieved by subjective judgement of the modeller (Rummel, 1970, "common factor analyses"). There is a quantitative screening process for the identifiable parameters proposed by Speed (1996) based on the coefficient of variation (cov), which can be

applied to the transmissivity values from unconstrained optimization. The cov is a ratio of the parameter standard deviation to its mean (σ/μ, see Ang and Tang, 1975). The ratio is often used as a measure of the dispersion or variability of the parameter about its mean.

The transmissivities that have high cov values will not be identifiable by the optimization process, and the modeller's judgment will be required to either take these transmissivities out of the decision variable set or to incorporate them in other meaningful ways to be discussed later. Transmissivities with low cov values are identifiable by the indirect inverse modelling.

From the manual calibration process and through expert analyses, 24 zones of transmissivity have been identified for the Canada Creosote site 2-layer model (Figures 3.0a,b). Based on the expert calibration, some of the 24 zones have identical transmissivities (transmissivities), as shown below:

Table 3.2 Zones of Transmissivity with "expert" estimates- 24 zones

Index	Transmissivity zone numbers	Kx in m2 /day
1	1,2,3	0.1296
2	4	0.648
3	5	10.368
4	6,7	12.96
5	8,9	15.552
6	10,11	31.04
7	12,13	38.88
8	14,15	46.656
9	16,17	51.84
10	18	54.432
11	19	62.208
12	20,21	90.72

13	22	103.68
14	23,24	388.8

Both the program Groundwater Vistas and the template file of PEST treat input transmissivities with the same values as one zone, so whenever a 24 zone model is used, the identical transmissivities above are made slightly different to separate them into different zones.

An important aspect of the optimization process is the log transformation (base 10) of the transmissivities. The transmissivities are assumed to have a lognormal distribution. The program PEST manipulates the log values to make use of the normal or Gaussian distribution properties, one of which is that the mean and the standard deviation as parameters of the distribution are the same as those of the transmissivity - the log transmissivities. When the cov is calculated by the ratio of the standard deviation to the mean, the calculation is performed with transmissivities and not log-transmissivities. The conversion from lognormal distribution parameters to the cov of the variate (transmissivities) can be found in (Ang and Tang, 1975), and is summarized by the following formula:

$$\text{cov of transmissivity} = [\exp(\text{Var}(\ln T)) - 1]^{0.5} \qquad (3.1)$$

Where cov is the coefficient of variation of transmissivity, denoted by T,

Var (lnT) is the variance of the natural log transformed transmissivity. The variance is a parameter of the lognormal distribution, and is obtained from the diagonal terms of the PEST output variance and covariance matrix,

Log T from PEST is converted to ln T by means of a change of base.

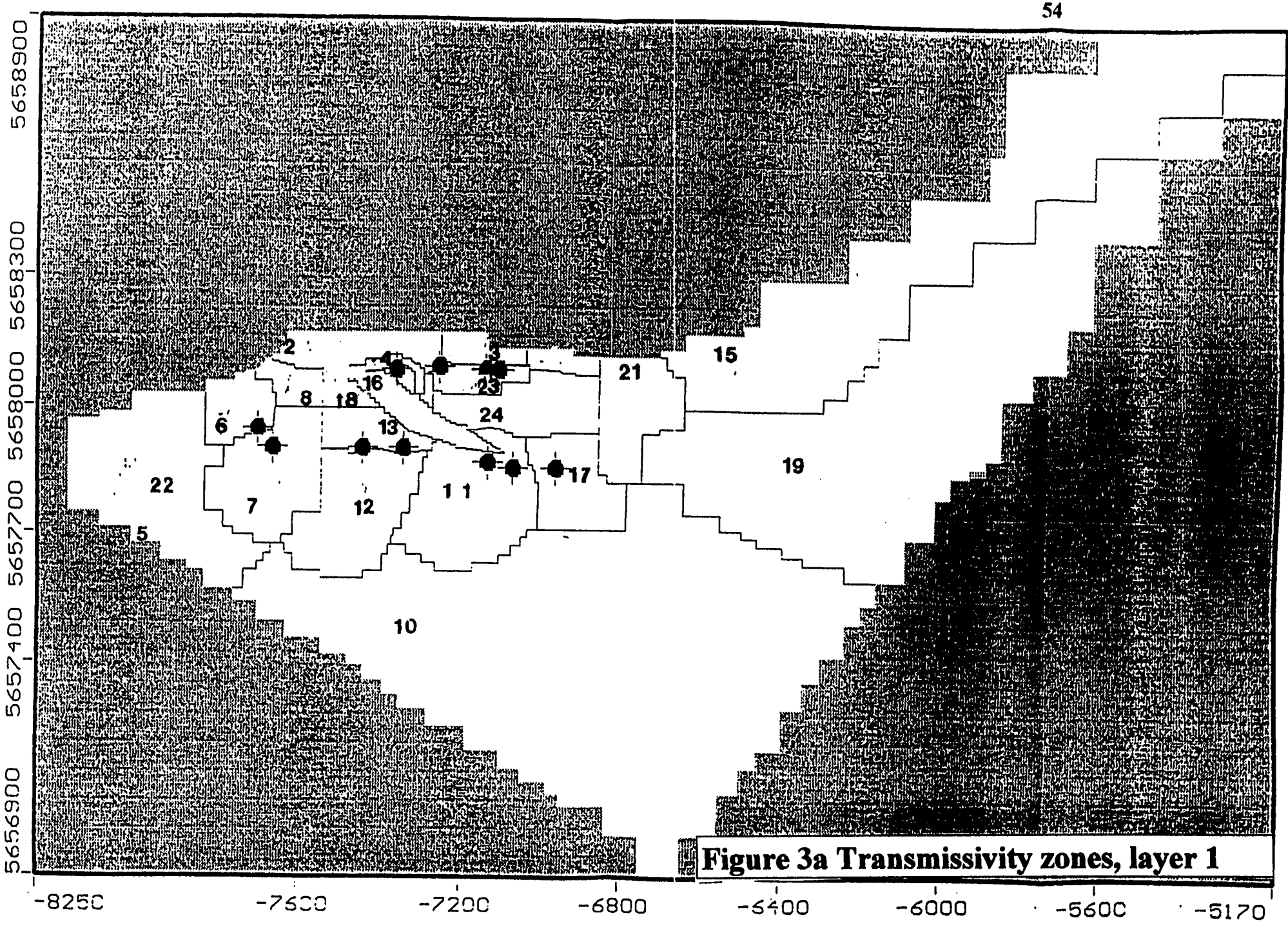


Figure 3a Transmissivity zones, layer 1

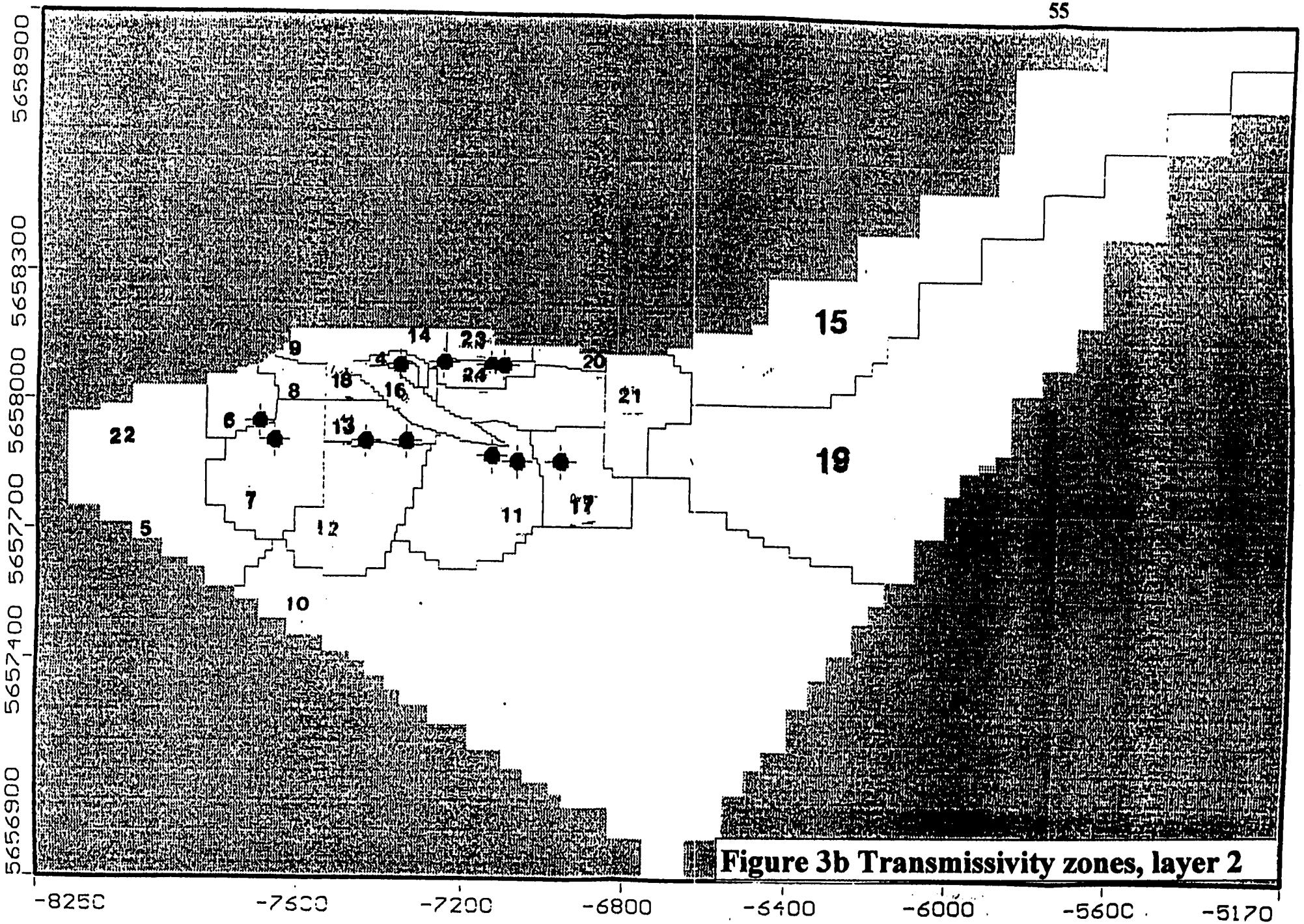


Figure 3b Transmissivity zones, layer 2

It is often assumed that the distribution of errors of the log transmissivities estimates are Gaussian with zero mean (Carrera and Neuman, 1986a). In real life there are exceptions to the assumption of Gaussian, zero mean of the log transmissivity errors (Ginn and Cushman, 1990). However, for this study normality is assumed. This is highly convenient for the calculations. The normal distribution of log-transformed transmissivity errors is also equivalent to the untransformed errors having a lognormal distribution.

To assess the identifiability of the zones, a table for "raw coefficient of variation" (Speed et al, 1996) is constructed below for the 24 transmissivities in table 3.3. Parameters with low cov values are identifiable and therefore used as decision variables. The standard deviation and mean estimates for the parameters are obtained by unconstrained optimization of the 24 zone models of the scenarios. The optimizations are performed for groups of 5 zones or fewer each time to reduce instability by reducing the dimensionality of the parameters. The resulting condition numbers (Speed et al, 1996, Weiss and Smith, 1998b), which are the ratios of the largest eigenvalue to the least eigenvalue of the normal matrix, are below 10^5, indicating that matrices are not ill-conditioned and instability due to noise is not significant (see section 2.3). Upper and lower bounds are required input to parameters on the decision variable list. The bounds are to prevent fluctuations of the parameter values in an unstable situation (see section 2:3). The upper and lower bounds used were specified as input to PEST in terms of hydraulic transmissivity. The upper bounds were 1000 m2 /day and later changed to and maintained at 4000 m2 /day, and the lower bound is 10^{-4} m2 /day for all parameters to be optimized.

The diagonal terms of the covariance matrix of the parameters from the optimization are used as the log variance for the parameter (denoted by variance in table 3.3). The cov for each of the parameters is calculated by means of equation 3.1. The identifiability of the parameter is based on an assessment of the cov values.
To use the table 3.3 for identifiable parameter selection, there are several relevant considerations:
Firstly, instability of the parameter has to be determined. When the mean values of decision variables stay at the upper (4000.m2 /day) or lower (1e-4 m2 /day) bounds, then

the decision variables are insensitive to the optimization (see detail discussion in section 2.3). Transmissivity zone 18 and 22 are both at the upper bound of 4000 m2 /day, and are thus excluded from the decision variables.

Tables 3.3 Parameter estimation and coefficients of variation for the 24 zones

zone	1	2	3	4	5	6
variance	2.94e2	402.4	178.7	16.41	327.6	3.137
cov	9.99e50	9.99e50	9.99e50	7.81e18	9.99e50	4088.88

zone	7	8	9	10	11	12
variance	3.311	4.712	22.78	7.41e-3	.2078	2.78e-2
cov	6485.29	2.66e5	1.68e26	.2002	1.4175	.3981

zone	13	14	15	16	17	18
variance	.1109	68.16	9.474	2.72e-2	.2088	.878
cov	.8946	9.99e50	8.08e10	.3935	1.4232	-

zone	19	20	21	22	23	24
variance	.3236	4.566	.7763	14.63	.2168	4.71e-2
cov	2.1355	1.81e5	7.7656	-	1.4685	.5324

Note: '-' denotes mean value at upper or lower bounds, no cov calculated

9.99e50 denotes a cap on cov value if log variance exceeds 50.

Secondly, transmissivities with the lowest cov are chosen for further optimization. But the choice of the magnitude of cut-off cov is arbitrary; other than the obviously large cov values, there is not a "standard" magnitude of the cov that can be used as a "cut-off".

There is also the balance between a large cov magnitude for cut-off and the reasonably small number of decision variables required for a stable optimization.

With the number of observation wells ranging from 12 to 26 for the six scenarios, five decision variables were observed to work well, in terms of stability. Using the above tables as illustration, the cov less than 8.0 is used as a cut-off, ten zones (number 10, 11, 12, 13, 16, 17, 19, 21, 23 and 24) are identifiable. The number of zones is more than 5.

A compromise is made to determine which zones to select for screening tests. Nine zones out of the ten above will be used : 10, 11,12, 13, 16, 17, 20, 23, 24. Many of these zones have the same expert calibrated transmissivity (see table 3.2). To reduce the number of decision variables to 5, the zones with identical transmissivities are grouped together. They will be optimized as groups. With this arrangement, a 14 transmissivity zone model is created, in which 5 zones with index number, in table 3.2, of 6, 7, 9, 12 and 14 will be the decision variables. These decision variables will contain the transmissivity zones of 10,11,12,13,16,17, 20, 21, 23 and 24, of the original 24 zone model, in groups. After using the 14 zone model for the screening optimization, the 24 zone model will be used again to seek the optimal point of minimum SS_{lof} for all six scenarios combined.

The set of 14 transmissivity zones used in all six scenarios are listed in table 3.4 below. With the measurements from each scenario and the model outputs, a series of calculations can be carried out to find the sums of square and the corresponding variance by means of ANOVA.

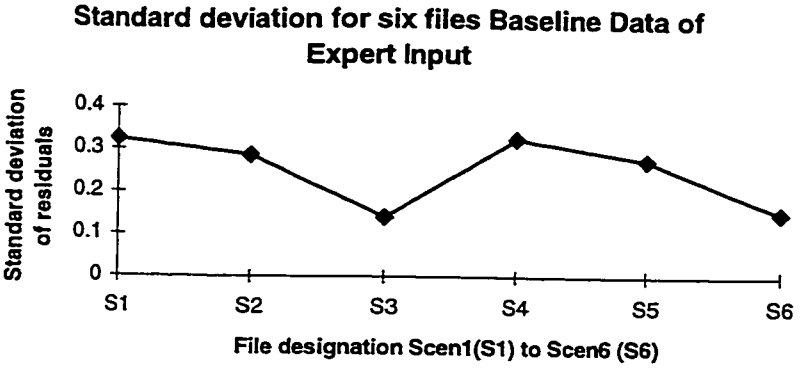
A general classification of the types of variances, already discussed in section 2.4, is:

SS_T (total sum of squares)
SS_{mean} (sum of squares of the mean)
SS_{corr} (sum of squares corrected for the mean)
SS_{FACT} (sum of squares of the factors or parameters)
SS_R (sum of squares of the residuals or the lack of fit) = SS_{lof}
Assuming that the pure experimental error is negligible for computer simulations.

Each type of variance is calculated by dividing the appropriate sum of squares by its degree of freedom. The procedure for the calculation of the variances were discussed in

section 2.4. The square root of the variances for the residual (or lack of fit) for the six scenarios using the expert calibrated parameters are shown in figure 3.1.

Figure 3.1 Standard deviation of lack of fit with expert calibrated parameters for six scenarios



The expert calibration produced reasonably uniform performance between the six scenarios with somewhat lower values of the standard deviation of the lack of fit in scenarios S3 and S6.

3.2 Screening tests for transmissivities, model with reduced parameters

The transmissivities for simulation in the last section (expert calibration) are interpreted from the site hydrostratigraphic units and hydrogeologic information and then calibrated to a best expert fit to the observations of the six scenarios. To better understand their effects in the simulation and the optimization process to follow, more screening tests are performed on them.

Scen6 was used to delineate which transmissivity (Kx) zones are identifiable using the given head measurements, have stable estimates of the parameter values, and do not have large uncertainties (Weiss and Smith 1998a,b). The large uncertainties are an

indication of ill-conditioned matrices. The ill-conditioning will worsen the noise problems in the data and will cause unstable estimates, as explained in section 2.3. The screening tests attempted to optimize all 14 zones. The 14 transmissivity zones with the corresponding expert calibration transmissivity values for the two layer model are shown in table 3.4:

Table 3.4 Zones of Transmissivity Chart with "expert" estimates

Transmissivity zone numbers	Kx m2 /day	Kx m2 /s
1	.1296	.0000015
2	.6480	.0000075
3	10.368	.00012
4	12.96	.00015
5	15.552	.00018
6	31.04	.00036
7	38.88	.00045
8	46.656	.00054
9	51.84	.0006
10	54.432	.00063
11	62.208	.00072
12	90.72	.00105
13	103.68	.0012
14	388.8	.0045

The spatial distribution of these zones is illustrated in Figure 3.1. In preliminary analyses, parameter zones Kx1 and Kx3 were determined to be insensitive to the head data in scen6, and delayed convergence in the optimization algorithm. Consequently, the

expert parameter estimates for Kx1 and Kx3 are used as prior information with a weight of one.

Parameter Kx4 has an effect on scenario Scen6. The presence of pumping wells (Table 3.1 CW and DW) causes draw down of the water level close to the wells. If Kx4 is adjusted in the optimization to unrealistic levels, then the observation well would be dry in the MODFLOW simulation. The "expert" input for this Kx prevents this occurrence, thus Kx4 is kept at the "expert" input value and out of the optimization decision variables set.

For this optimization, the upper and lower bounds for all parameters are set to 1000.0 and 1e-4 m2 /day respectively. Results of the optimization of 12 unconstrained parameters and Kx1 and Kx3 with prior information are shown in table 3.5.

3.2.1 The Decision Variables for the 14 Zone, 2 Layer Model - cov with magnitude of parameter standard deviation

The discussion in section 3.1 dealt with the selection of the decision variables from the 24 zone, two layers model to the formation of the 14 zone two layer model. This section examines the decision variables to be used in the 14 zone, two layers model to

Table 3.5 Optimized results of Transmissivities (Scen6) with 95% Confidence Limits

Parameters Kx	Estimated value m2 /day	95 % Lower Limit m2 /day	95 % Upper Limit m2 /day
Kx1	.129520	7.852656e-2	.213626
Kx2	17.2213	1.599289e-9	1.854398e11
Kx3	10.3728	6.2884	17.1102
Kx4	13.0985	12.699	13.5105
Kx5	18.2793	8.08488e-5	4.132813e6
Kx6	21.0574	9.13883	48.5199

Kx7	22.6541	2.48818	206.258
Kx8	.107649	4.348048e-6	2665.16
Kx9	13.1952	1.2451	139.838
Kx10	621.045	80.5147	4790.39
Kx11	563.744	8.319523e-4	3.820622e8
Kx12	59.2753	.786072	4469.77
Kx13	109.637	1.074439e-5	1.118756e9
Kx14	112.585	9.20596	1376.87

Table 3.6 cov for 14 zone model from scen6

Kx No	1	2	3	4	5	6	7
ln Variance	.0544	115.95	.0544	.0002	33.036	.1514	1.0603
ln mean	-2.0439	2.8461	2.3392	2.5725	2.9058	3.0473	3.1203
cov	-	1.51e25	-	-	1.49e7	.4044	1.3739

Kx No	8	9	10	11	12	13	14
ln Variance	22.2467	1.211	.9072	39.1810	4.0618	56.6243	1.3626
ln mean	-2.2289	2.5799	6.4314	6.3346	4.0822	4.6972	4.7237
cov	6.77e4	1.5351	1.2154	3.22e8	7.5550	1.98e12	1.7048

Note: "-" denotes transmissivities excluded from this cov calculation, see explanation in section 3.2

verify that the same variables are indeed identifiable. With reference to table 3.6, transmissivities with cov values higher than 8.0 will be excluded from being the decision variables. Six transmissivities are below the cut off: Kx 6, 7, 9, 10, 12, 14.

The cov is not an absolute measure of the dispersion of transmissivities from the mean. For the same magnitude of cov, transmissivity with a higher mean can tolerate a

higher standard deviation than a transmissivity with a lower mean. This can be demonstrated by examining the estimated transmissivity mean and standard deviation from the lognormal distribution parameters of ln variance and ln mean in table 3.6 above.

The cov, which is the ratio of estimated transmissivity standard deviation to the estimated transmissivity mean, is given by equation 3.1 and tabulated in table 3.6. The conversion of the distribution ln mean to the transmissivity mean is given by (Ang and Tang, 1975):

$$\text{transmissivity mean} = \exp(\ \ln \text{mean} + 0.5(\text{var}(\ln T))) \qquad (3.2)$$

Where ln mean is the natural log of the mean as in table 3.6, it is a distribution parameter,

Var(ln T) is defined in equation 3.1.

The transmissivity mean and standard deviation for the six zones are calculated from equation 3.1 and 3.2 and tabulated in table 3.7:

Table 3.7 Potential decision variables - transmissivity mean and standard deviation

Decision variable Kx No.	6	7	9	10	12	14
Transmissivity mean estimate m2 /day	21.360	45.027	14.791	676.51	86.942	128.02
Transmissivity sd estimate m2 /day	8.6371	61.860	22.707	822.25	66.606	218.25
Decision	accept	accept	accept	reject	accept	accept

The decision to accept five of the six decision variables is based on the magnitude of the parameter standard deviations.

After this process of elimination, the number of parameters to be optimized is reduced to 5, namely, Kx 6,7, 9,12,14. Other scenarios were also tested. Eight runs in total are made: Seven with Scen4, one with Scen3, resulting in the same five parameters

being identified again as the ones to be included in the decision variable set. These results agree with the 24 zone results in section 3.1, in which the compromise between the number of decision variables and the stability of the parameter estimates is attained.

3.2.2 Individual optimization for the six scenarios

The objective of the individual optimization is to optimize the 5 selected transmissivity zones values (parameters) for all six scenarios, independently. No prior information is used in the individual optimization runs, and all the weights for the measurements are set to unity. After the optimization runs, the sets of optimized parameters for Kx 6,7,9,12 and 14 and their 95% confidence limits, are obtained and listed below in Table 3.8. The section on the lower right hand corner is a listing of the values of σ_{ref}^2 (sigma sq in the table) of equation 2.13 for each of the individual optimization.

The individual optimization are the best fit we can expect from any given scenario, using the current set of decision variables. The variances of lack of fit will be the lowest that could be expected from the process, for each individual scenario. It is also evident that the Kx values changed from file to file, and their mean values showed significant deviation even on a log scale plot. Log scale plots for Kx6 to Kx14 based on the above table are shown in Figures 3.2a to 3.2e. Figure 3.3 compares the individual sum of squares to those determined from the expert values.

3.2.4 The search for an "average" or "central" set of parameter values: the Monte Carlo simulation

This subsection details efforts to find an "average" set of parameter values which are suitable for and common to all six scenarios (Scen1 to Scen6). From the different values of the Kx in table 3.4, an attempt is made to obtain one set of common values that can be used throughout the simulations to produce overall low lack of fit. Two common

means of averaging Kx are to take the mean or the median value of the six optimization runs.

Table 3.8 Individually optimized results of 5 Kx in m2 /day for each of the six scenarios:

Scenario	Kx6	95%lo	95%hi	Kx7	95%lo	95%hi
S1	118.677	1.00804	13971.8	210.714	8.71574	5094.29
S2	368.775	9.05E-02	1.60E+06	34.5877	7.66E-02	1.56E+04
S3	0.402728	5.94E-02	2.73097	7.08819	6.69E-01	75.0498
S4	1.20327	7.83E-03	184.872	8.33855	1.74E-01	398.686
S5	32.8448	11.9935	89.9477	87.6232	4.57063	1679.82
S6	35.5833	14.5318	87.1313	36.0534	3.25243	399.656
	Kx 9	95%lo	95%hi	Kx 12	95%lo	95%hi
S1	100.733	1.10539	9179.73	200.613	0.197815	203451
S2	412.46	1.95E-01	8.73E+05	299.391	7.77E-02	1.15E+06
S3	6.45113	8.76E-01	47.5342	4.18459	3.10E-01	56.4825
S4	1.14496	4.79E-03	273.801	2.25338	6.75E-03	751.835
S5	40.8816	0.831702	2009.5	1000	2.60844	383370
S6	56.3888	6.76935	469.72	86.573	14.5844	513.899
	Kx 14	95%lo	95%hi	Expert	Scenario	sigma sq
S1	1000	13.256	75437.4	Kx6=31.04	S1	0.033
S2	981.739	1.09E+01	8.84E+04	Kx7=38.88	S2	0.0308
S3	77.1418	8.48E+00	701.905	Kx9=51.84	S3	0.0062
S4	16.0114	1.44E-01	1779.05	Kx12=90.72	S4	0.045
S5	932.035	13.5095	64302.2	Kx14=388.8	S5	0.019
S6	383.37	66.6656	2204.62		S6	0.0201

Note: Sigma sq is σ_{ref}^2, see above discussion

Expert denotes expert calibration results

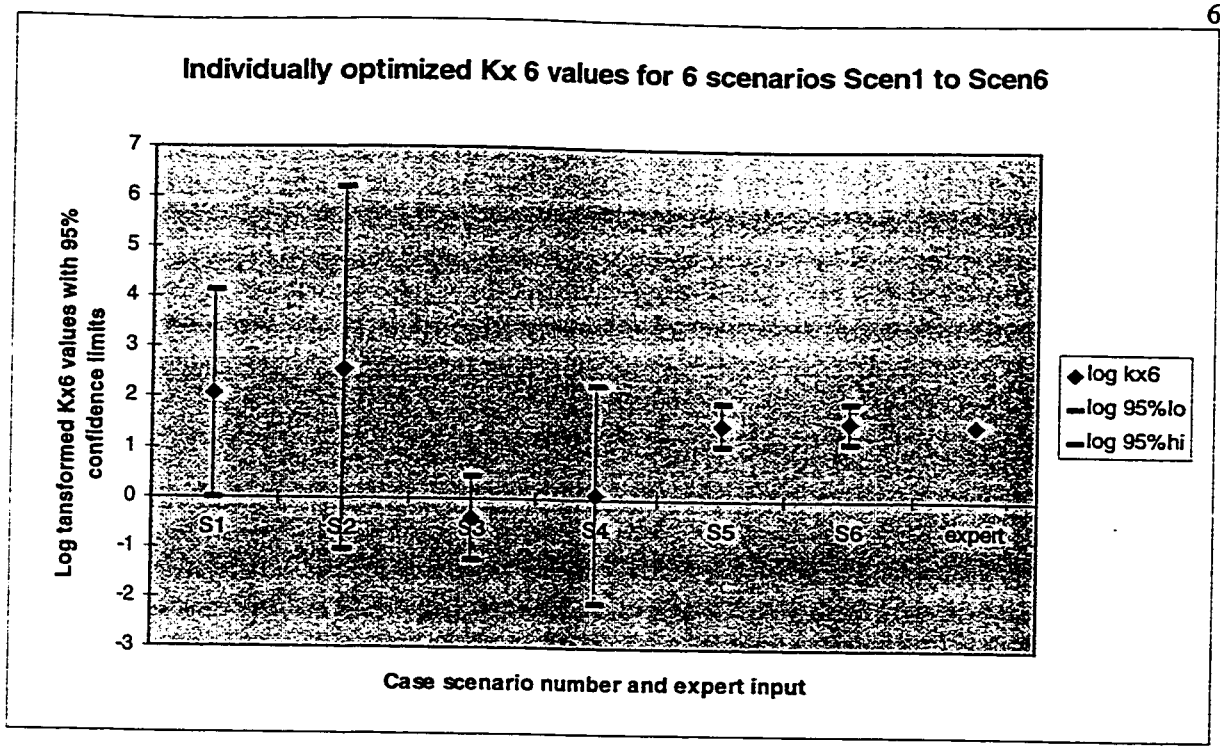


Figure 3.2a Log transformed plot of individually optimized Kx6 values for each scenario

Figure 3.2b Log transformed individually optimized Kx7 values for each scenario

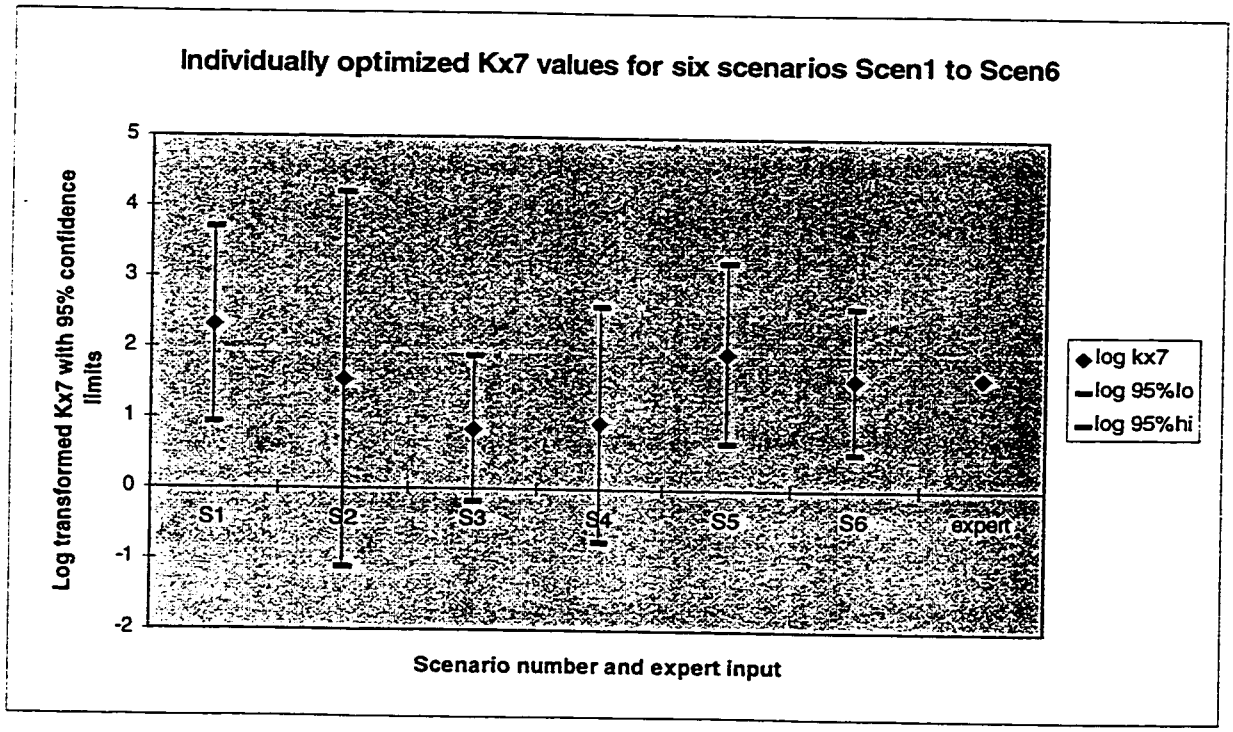


Figure 3.2c Log transformed individually optimized Kx9 values for each scenario

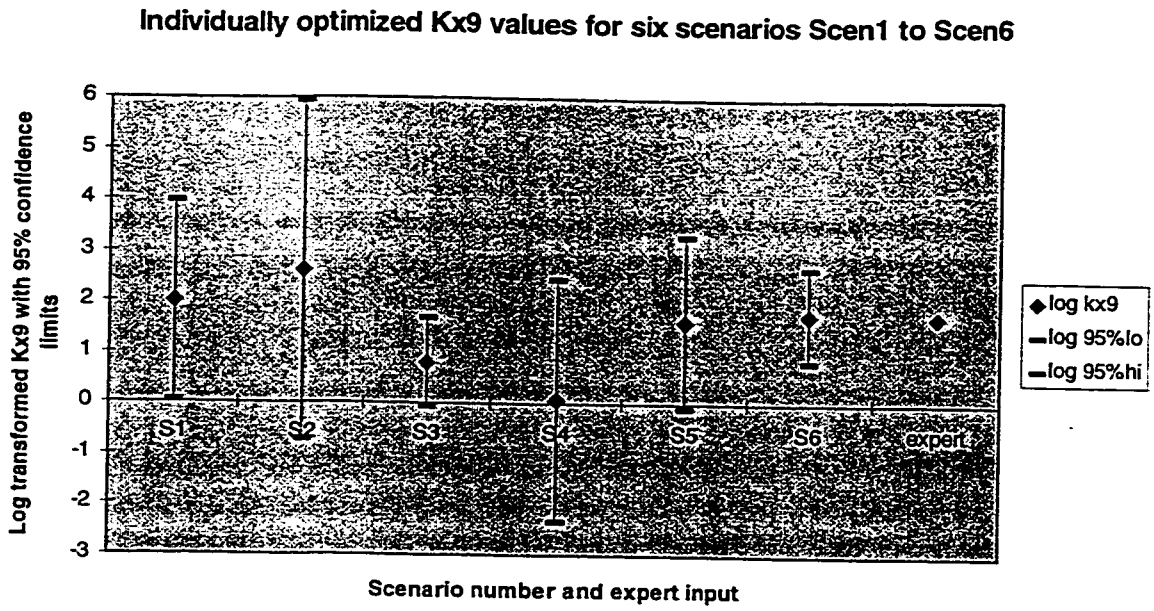


Figure 3.2d Log transformed individually optimized Kx12 values for six scenarios

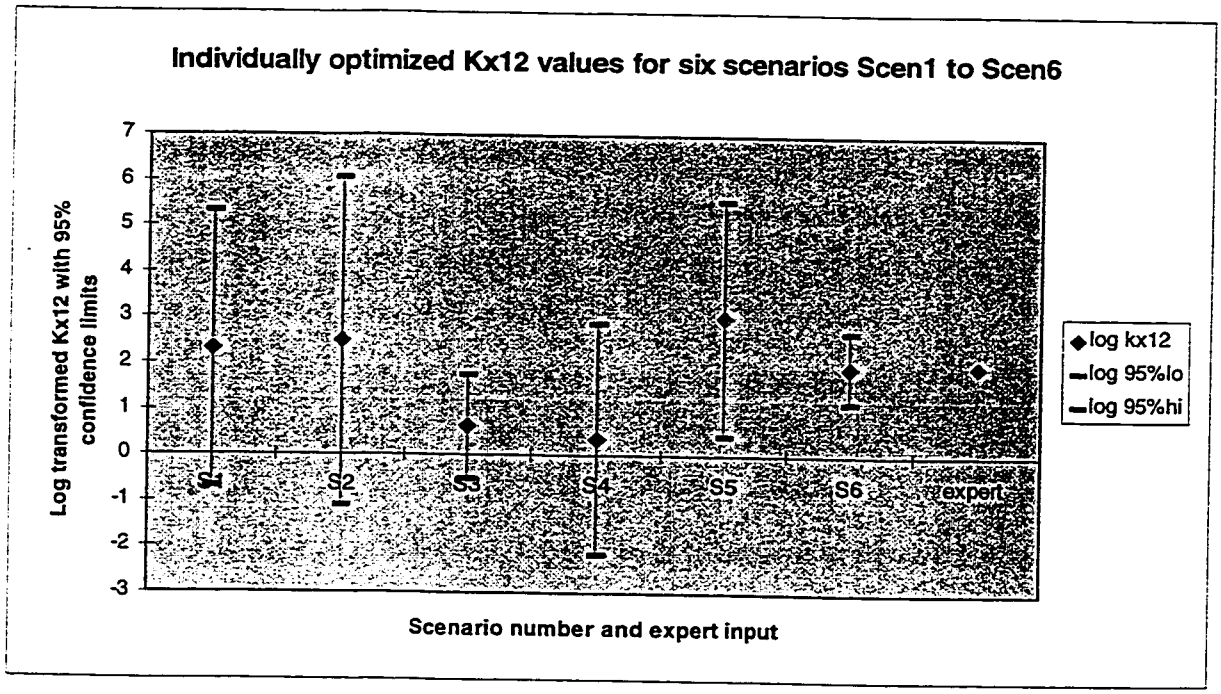


Figure 3.2e Log transformed individually optimized Kx14 values for each scenario

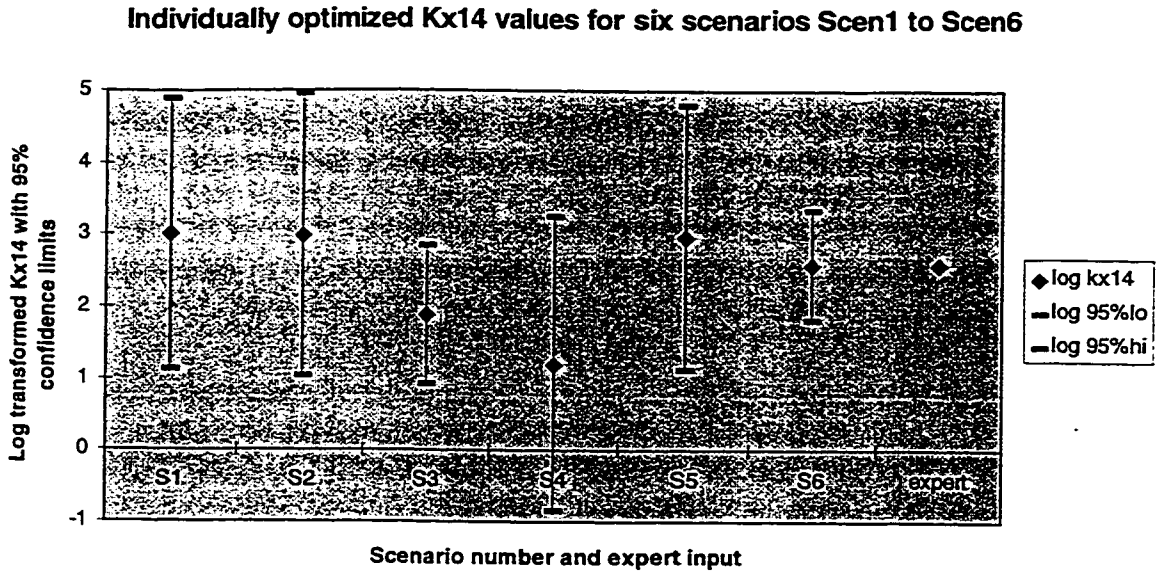
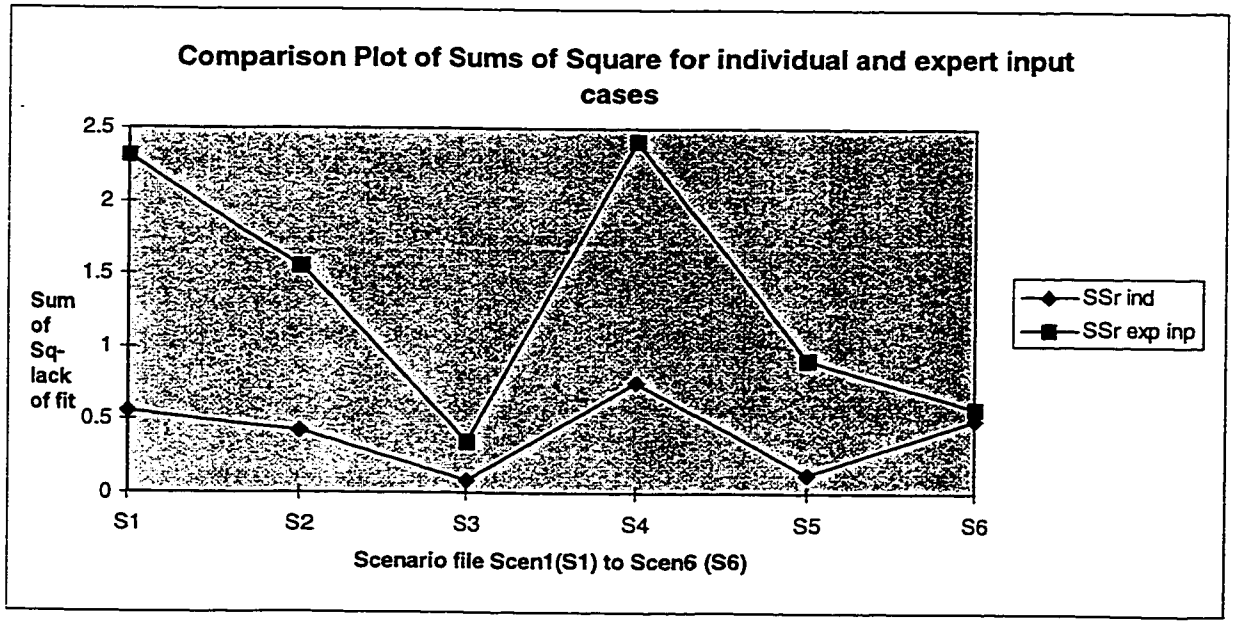


Figure 3.3 Comparison plot: SS_r ind is based on individually optimized values

SS_r exp inp is based on expert input



An arithmatic mean of the Kx can be obtained as an output of Monte Carlo simulation by means of the program Crystal Ball. This measure is affected by the magnitudes of the extreme values within the set of Kx, and the calculation does not take into account the uncertainty in the value of Kx indicated by s_b^2. A measure less affected by the large range of values of Kx is the geometric mean, which can be obtained by dividing the sum of the log transformed values of Kx by the number of Kx (that is, 6). But this measure still does not account for the uncertainty. A more appropriate average measure in this case is the median.

Medians are not affected by extreme values in the set, and the means and standard deviations of the Kx can be used in a Monte Carlo simulation of the particular Kx. The position of the median will be affected by the spread of the resulting Kx realizations in the Monte Carlo simulation, thus accounting for the uncertainties indicated by the PEST output.

The assumption used in the Monte Carlo simulation is that all of the log transformed Kx belong to normal distributions with means and standard deviations taken from the individual optimization runs for the six scenarios. The program Crystal Ball is used to generate the output statistics for the combined input of the six lognormal inputs from the individually optimized runs for each Kx.

The method of Shooman (1968) is used to determine how many simulations are required to obtain a 1% error in the mean value of the parameter at the 95% confidence level (Ang and Tang, 1990):

$$\text{Percentage error of simulation} = 200((1-pf)/(ss(pf)))^{0.5} \qquad (3.3)$$

Where the percentage error is 1,

pf has a value of 0.5 for the estimated probability,

ss is the number of simulations

In 19 out of 20 simulation trials (a confidence of 95%), the means have an accuracy of $\pm$ 1% of their estimated values, if the number of trial samples ss is equal to 40,000.

Each of the Kx is input into the program with six lognormal probability distribution curves, one from each of the six scenarios. The means and the standard deviations of the curves are furnished by the individual optimization of the last section. 40,000 values from each of the input lognormal distributions curves are produced, using the random number sampling within the computer program. This makes a total of 240,000 trials for each combined Kx estimate. The combined estimate for each Kx is compiled one at a time, on an output distribution chart (Figure 3.4). This chart shows the frequency of the realizations. From this output, the mean and the median are found. The final Kx output are tabulated in table 3.9.

The remaining results for the other Kx are presented in APPENDIX A. The sums of square due to lack of fit for the median and the mean(or average) are shown in table 3.9.

Table 3.9 values from the Monte Carlo simulations and other "average" measures:

Type / Kx	Kx6	Kx7	Kx9	Kx12	Kx14
Median	34.22	35.15	51.45	138.46	911.74
Mean	92.85	64.12	103.11	265.82	565.09
Geo mean	17.08	33.30	29.85	60.50	275.1
Expert	31.04	38.88	51.84	90.72	388.8

Figure 3.4 Combined value for Kx6 from Crystal Ball simulations

Crystal Ball Report
Simulation started on 12/6/97 at 19:13:24
Simulation stopped on 12/6/97 at 19:35:53

Forecast: Combined value for Kx6 **Cell: C7**

Summary:
Certainty Level is 96.84%
Certainty Range is from -Infinity to +Infinity m/day
Display Range is from 0.00 to 450.00 m/day
Entire Range is from 0.00 to 1,204.53 m/day
After 240,000 Trials, the Std. Error of the Mean is 0.27

Statistics:

	Value
Trials	240000
Mean	92.85
Median	34.22
Mode	---
Standard Deviation	131.08
Variance	17,180.97
Skewness	1.56
Kurtosis	4.03
Coeff. of Variability	1.41
Range Minimum	0.00
Range Maximum	1,204.53
Range Width	1,204.53
Mean Std. Error	0.27

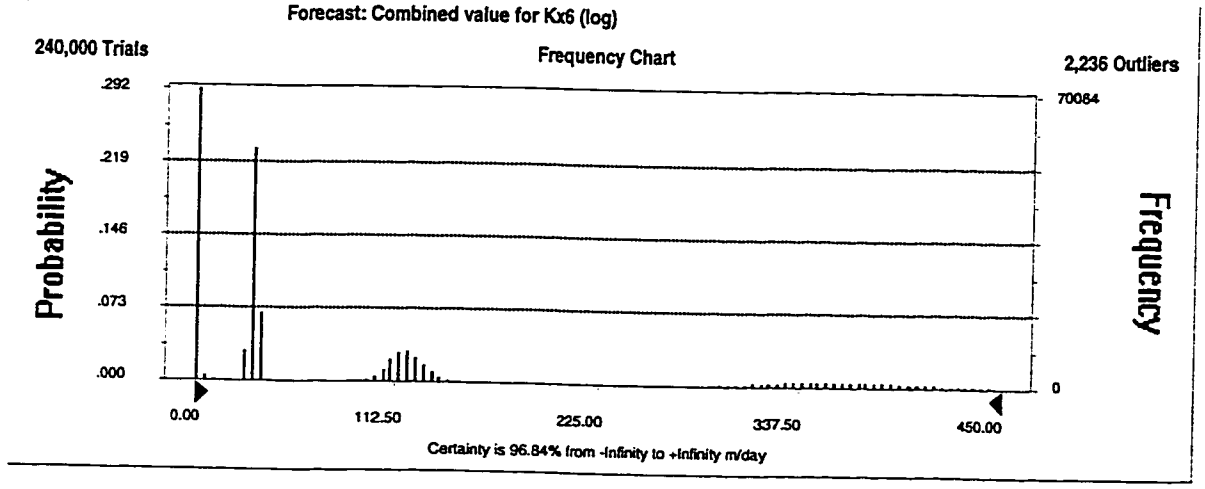
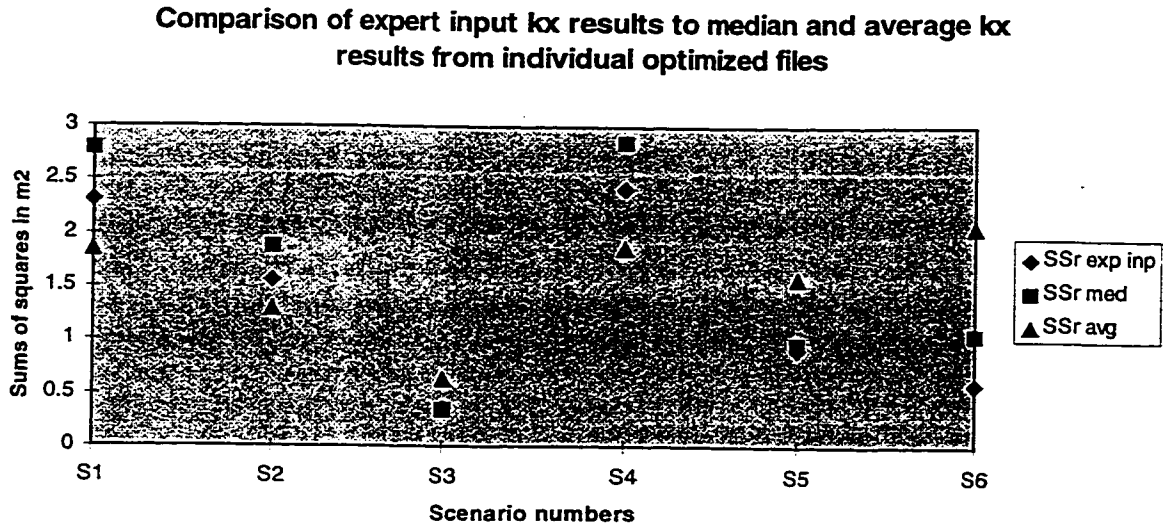


Figure 3.5 Comparison plot for median and mean against expert input



As can be seen, the median values produced sum of squares of lack of fit similar to the trend of those from the expert input, however, the sums of square for both the median and the average inputs are higher than those from the expert input (eg the total SS_R for the six scenarios with expert input is 8.133, as compared to 9.243 from averaged input and 9.85 from median input). Figure 3.5, shows that the averaged input seems to produce the most even sum of squares of lack of fit for all six scenarios.

The "average measures" approach failed to improve on the overall variances of the expert input, therefore some other method has to be devised to seek an improvement. It can be argued, though, that results from the expert input values are similar in trend to those from "average measures" input, and that the experts chose transmissivities that are similar to an averaged measure.

4.0 The optimization of model parameters by sequential feed forward procedure

The last section describes the use of an "average" measure designed to generate results that will be acceptable over the suite of six scenarios. In this section, with the use of prior information and weights, a new sequential feed forward procedure (sffp) is introduced to reduce the overall sums of squares of the lack of fit.

4.1 The use of prior information and weights

The use of prior information with optimization in inverse groundwater modelling is to alleviate the instability problem and to improve the conditioning of the normal matrix (Neuman and Yakowitz, 1979, see also section 2.1). Prior information and its weights are introduced into the system of matrix equations as penalty terms, and their anticipated effects on the system of equations have been explained in section 2.1.

Recalling the discussion in section 2.2, weighting is separated into sample (or measurement) and prior information weights. The sequential procedure, which will be discussed in the next section, begins with an unconstrained optimization without prior information. So the first step in the sequence has sample weights only.

There are two approaches to obtaining the sample weights. This has already been explained in section 2.2, but is reproduced here for completeness: The first approach is when the sample weights are selected as the inverse of the estimated variance of the head measurements. The head measurement errors are assumed to be independent, normally distributed with a standard deviation of 0.1m. Consequently, the weighting matrix is constant for all the optimizations and recalling equation 2.26 :

$\Omega_s = 100(\mathbf{I})$, a diagonal matrix

The second approach for weighting is based on a concept of the control of process variance (see section 2.2.1). The assumption is that the measurements errors are independent and of equal magnitude, thereby making the variance-covariance matrix Vs

an identity matrix (equation 2.27). The variable weights, calculated by the inverse of an estimate of the common variance for the untransformed disturbance (equations 2.27, 2.29):

$$s^2 = (S(\mathbf{b})/n\text{-}p) \approx \sigma^2_{comu}$$

$$\mathbf{I}\, s^2_k = \Omega_s^{-1}{}_{k+1} \qquad k = 0 \text{ to } 5$$

These weights are used purely to scale the common variance of the present step, to near unity (section 2.2.1, equation 2.31).

The anticipated result is that the variance of the sffp calculation is controlled, and the propagation of the variance to the parameter estimate will also be controlled. If the estimation of Cov(**b**) of equation 2.30a has less uncertainty, then the estimation of the parameters will be closer to their true values, and the overall SS_{lof} will be minimized. The first step of the unconstrained optimization uses 1.0 as weights in accordance with the identity matrix.

As a prelude to the empirical tests with the two approaches, unconstrained optimization with two randomly selected scenarios (S1 and S3) are performed with PEST sample weights of 1.0 and 10.0 to examine the differences in results. All other variables are the same for the scenarios selected. The weights of 10 and 1 have no appreciable effect on the unconstrained optimization.

While the unconstrained optimization parameter estimates are not affected by the sample weights, the estimates of the error in the parameter estimates will be affected and Cov(**b**) is used to calculate the prior information weights of the next step.

As an illustration of the use of prior information and weights, the following exercise with the case file Scen4 is conducted to examine the variances of the optimized values. The starting values for Scen4 for the five parameters are the expert input in table 3.2. Weights for the PEST input for both the Kx and the measurements are set to unity. The optimized results for Scen4, and the 95% confidence levels are in table 4.1:

Table 4.1 Optimized values for Scen4, with the 95% confidence limits

Kx	mean	95%lo	95%hi
6	31.7183	13.8764	72.5009
7	38.7328	6.34517	236.436
9	31.5772	7.69387	129.599
12	72.1418	8.17619	636.536
14	270.822	41.7305	1757.58

4.2 The sequential forward feeding procedure (sffp) and ANOVA

The objective of this optimization is to show the effects of the optimization of the six steady state files in sequence, in which the 5 optimized parameters are updated for each optimization and passed onto the next file for use as prior estimates, with the appropriate weights. Again, the discussion in section 2.2 applies.

Some of the sequential optimization effects to be studied include:

1. Importance of the order of the sequential optimization,
2. The importance of sample and prior information weights,
3. The importance of the starting scenario,
4. The retention of the optimized information from scenario to scenario,
5. The value of the optimized parameters during and at the end of the sequential procedure.

Results generated during the feed forward procedure are compared to the best possible results computed from individual scenarios (section 3.2.2).

4.2.1 The reference sequence in chronological order

The starting sequence follows the chronological order of the six scenarios to examine the overall sum of squares of the lack of fit for the sequence. The chronological order is logically appealing because the site measurement and control systems evolve with time and more and better data should be collected towards the last steps of the sequence. Whether this is the case remains to be determined. If the overall sums of

squares do decrease with the reference sequence order, then the claim on improvement over time can be substantiated.

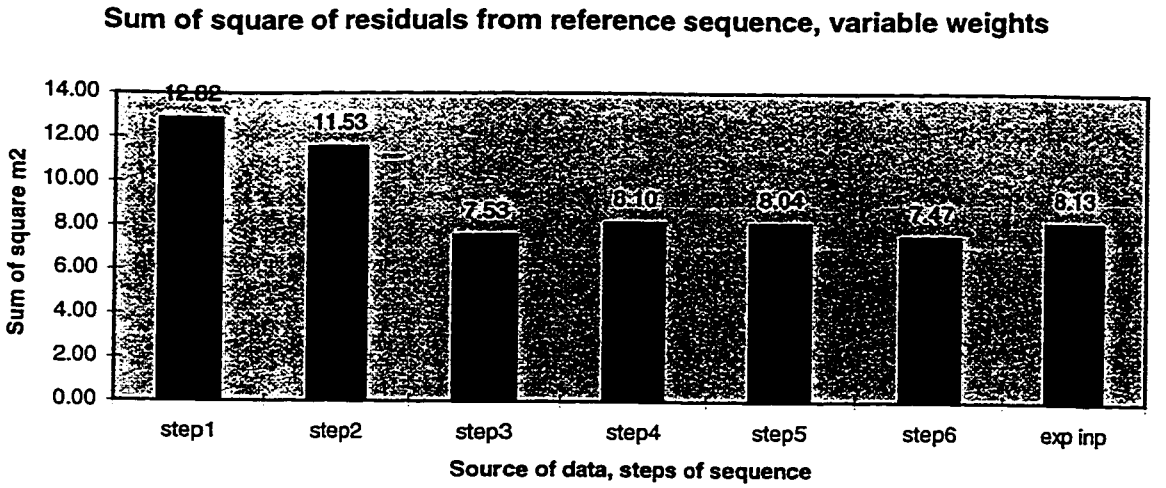
The reference sequence begins with Scen1 and ends with Scen6, and an unconstrained optimization of Scen1 provides the input to the first sequence step of constrained optimization of the same scenario.

Table 4.2 Reference sequence has the following six steps (Steps 1 to 6):

Step in sequence	Scenario optimized	Comment
1	S1	Input to S1 from unconstrained optimization of S1
2	S2	Input to S2 from step 1
3	S3	Input to S3 from step 2
4	S4	Input to S4 from step 3
5	S5	Input to S5 from step 4
6	S6	Input to S6 from step 5
		Output from step 6 (S6)

The sole purpose of this sequence is to examine if there is a decrease of overall sums of squares for the chronological ordered reference sequence steps. The plot of the sums of squares of the reference sequence and the expert input is shown in figure 4.1. Variable weighting for the samples is used.

Figure 4.1



The reference sequence results show that the sffp does improve on the expert input overall sum of squares, and there is a general trend of a decrease in the sums from the first scenario to the last. However, the decrease is not monotonic. It is evident that in steps 4 and 5 the sums increase over that of step 3. This fluctuation of the overall sums of squares supports the idea that by mixing up the order of other sequences to non-chronological steps may achieve lower sums of squares than what the reference sequence can produce.

The sequences that follow (sequence 1 to 4) use randomly mixed sequences of the six scenarios. The results will show that the order of the sequence is important to reduce the overall sums of squares, and depending on the order, better or worse results than the expert input can be obtained.

4.2.2 The first sequence

The first sequence starts with the individually optimized parameter values and variances of file S4 and ended with the file S2. The sequence begins with a scenario data file from the period after the wall is installed. The measurement data is considered to be of better quality than the two scenarios before the wall (S1 and S2). The sequence ends with one of the before the wall scenarios (table 4.3):

Table 4.3 Sequence 1 has the following six steps (Steps 1 to 6):

Step in sequence	Scenario optimized	Comment
1	S4	Input to S4 from unconstrained optimization of S4
2	S5	Input to S5 from step 1
3	S6	Input to S6 from step 2
4	S1	Input to S1 from step 3
5	S3	Input to S3 from step 4
6	S2	Input to S2 from step 5
		Output from step 6 (S2)

Table 4 shows the source of the input to each step of the optimization. From the results of each prior step, the sample and the prior information weights are calculated and put forward to the next scenario optimization. Since there are two approaches in finding the sample weights (section 2.2), there are two sets of results for sequence 1. The following is a method to assess the sequence results for the two approaches:

For each of the sffp steps above, including the output from step 6, the optimized parameter sets are found by PEST. The parameter set from each step is then fed back into the scenario files to find the total SSr -sum of squares of the lack of fit, for the six scenarios. The total SSr for sequence steps are found in Figures 4.2 and 4.3. The expert input (exp inp) is also plotted for reference. As can be seen, the variable weights approach outperforms the constant ωs approach for each step of the optimization. The total SSr for the variable weights approach exhibits less fluctuations than the constant weight approach.

The variable sample weights performs better than the constant weights, and details of it are examined further. For the variable weights, the evolution of the decision variables as the sffp proceeds are plotted in figures 4.4 to 4.8. The discussion on the series of weights used for the steps in the sffp are shown in section 6.

The ANOVA analysis follows the formulae in section 2.4.1. The ratio of SS_{fact} to SS_{corr} is termed the coefficient of multiple determination (R^2), and is a measure of how much of the variance in the data has been accounted for by the factors or parameters in the model. The higher the value of this ratio, the better the parameter performance. If the value of the SS_{fact} was small with a high R^2, then a large proportion of the overall variance was captured by the parameters or factors, but the actual magnitude of the sum of squares of the factors is small. This will indicate a set of parameters that work really well.

Figure 4.2 The total SSr for approach I - Constant Ωs at 10- Sequence 1

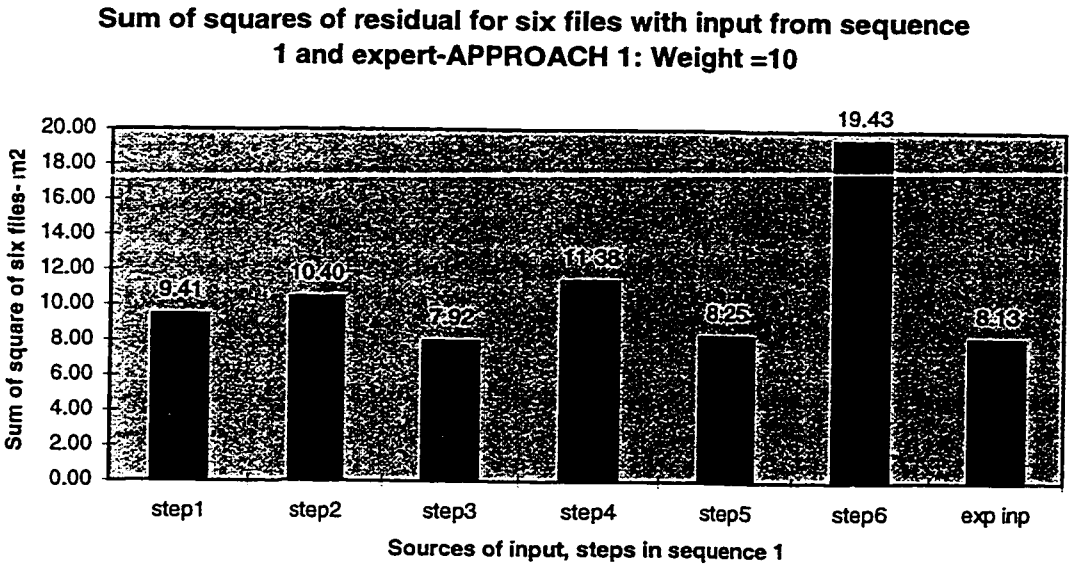
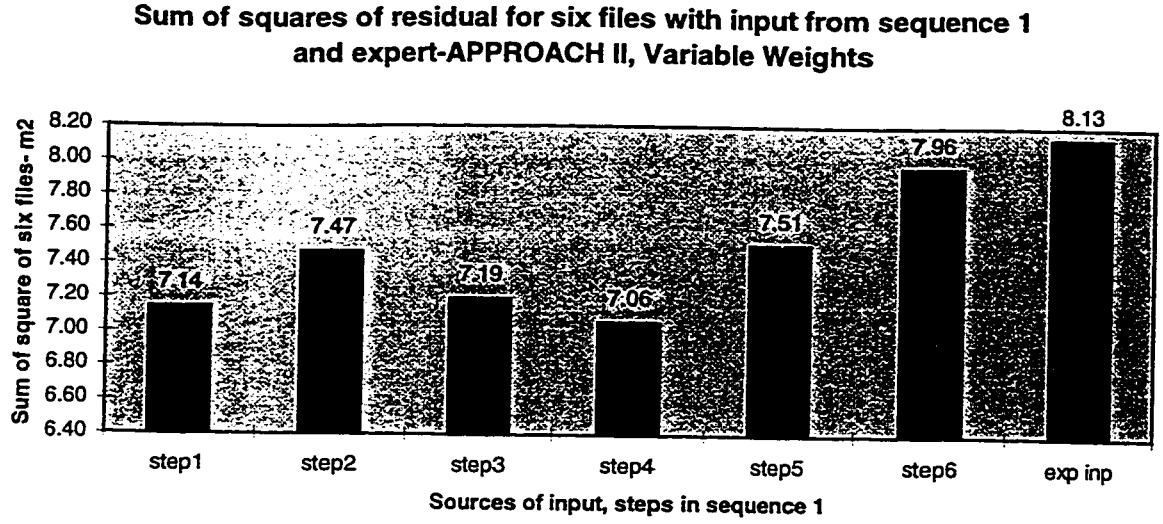
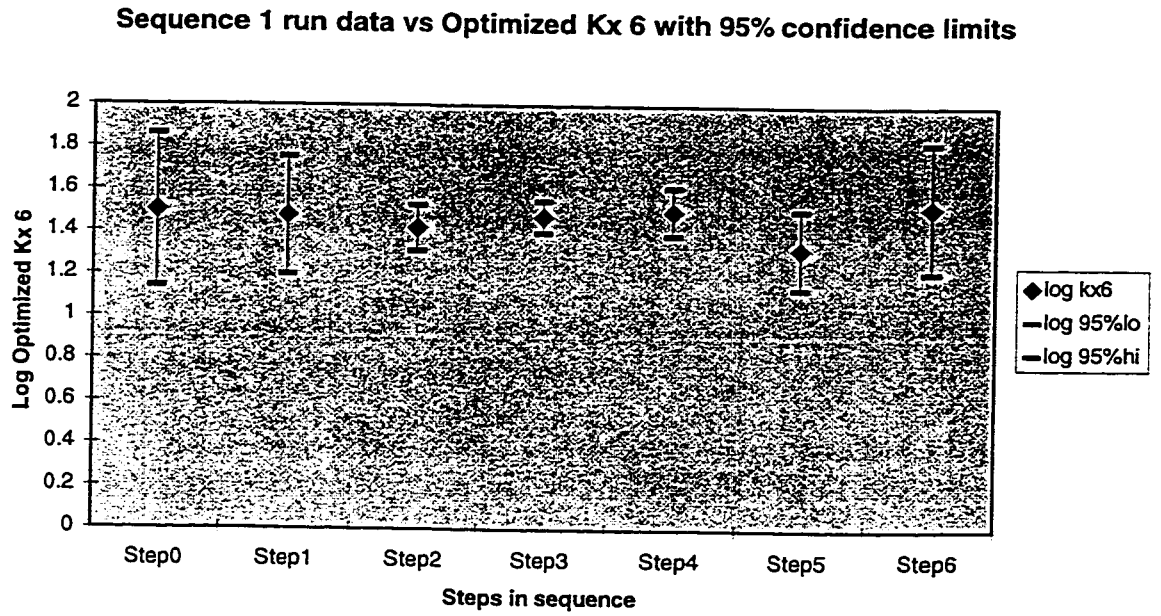


Figure 4.3 Total SSr for approach II - variable Ωs - Sequence 1



The following plots are from the results of sequence 1, using the second (variable weight) approach:

Figure 4.4 Plot of Kx6



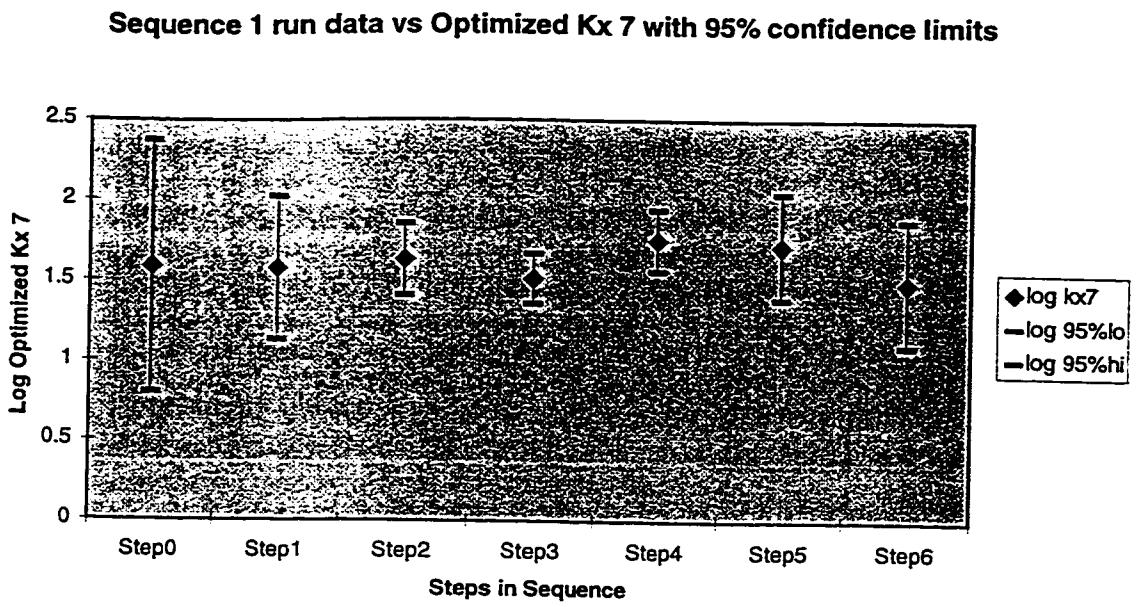


Figure 4.5 Plot of Kx7

Figure 4.6 Plot of Kx9

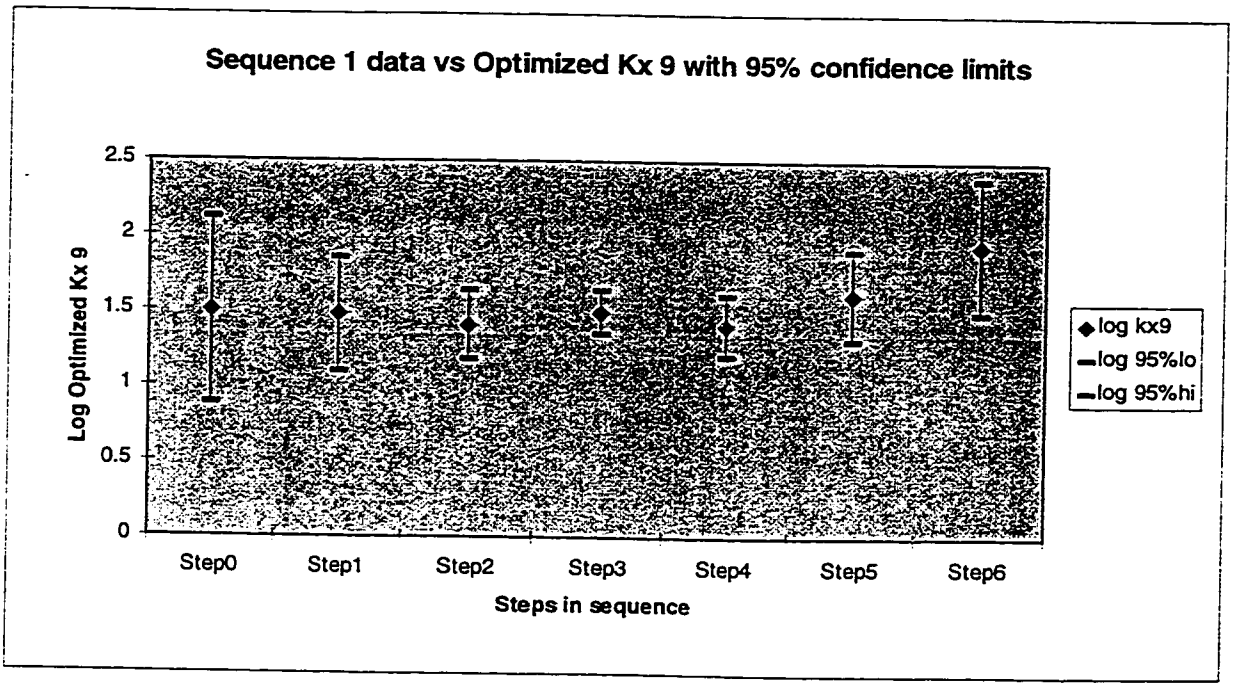


Figure 4.7 Plot of Kx12

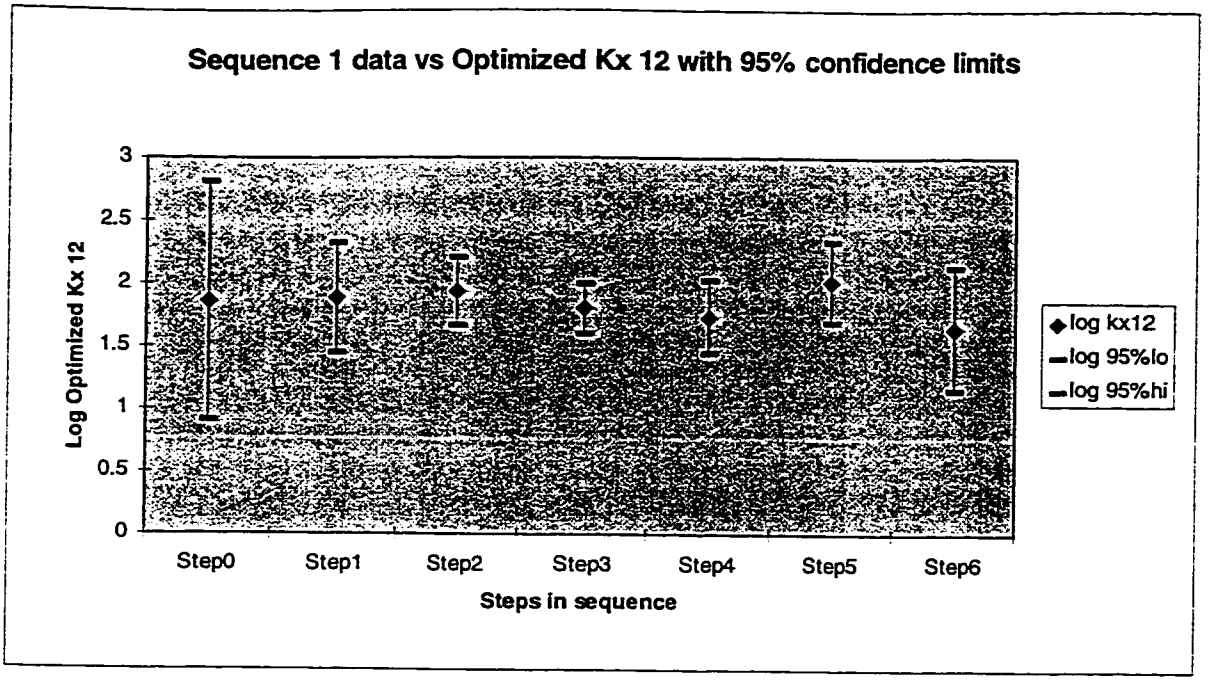
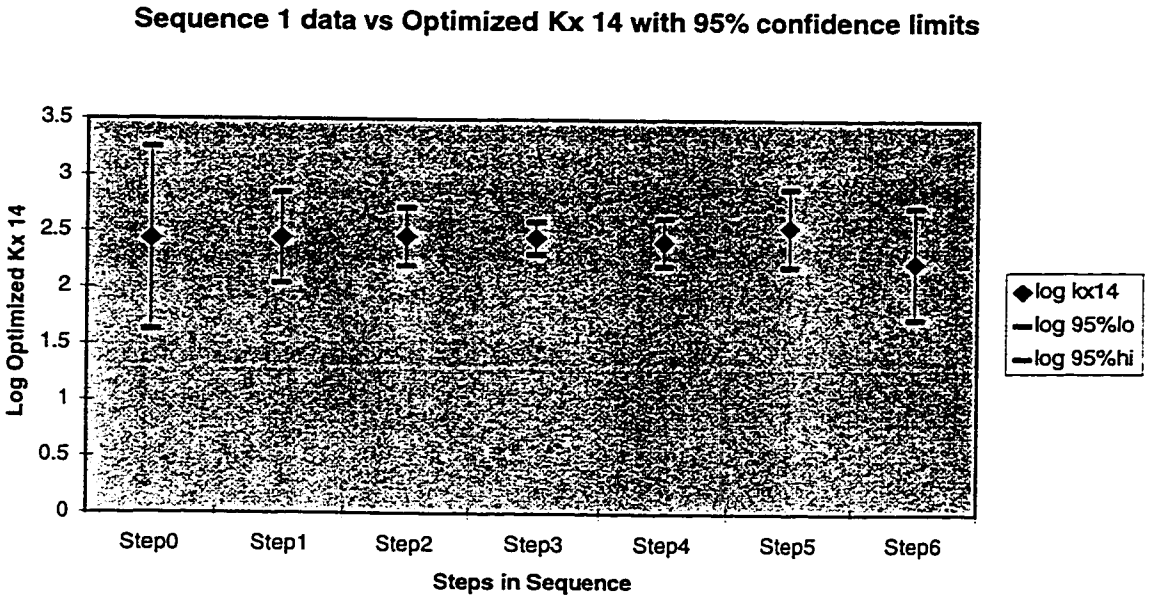


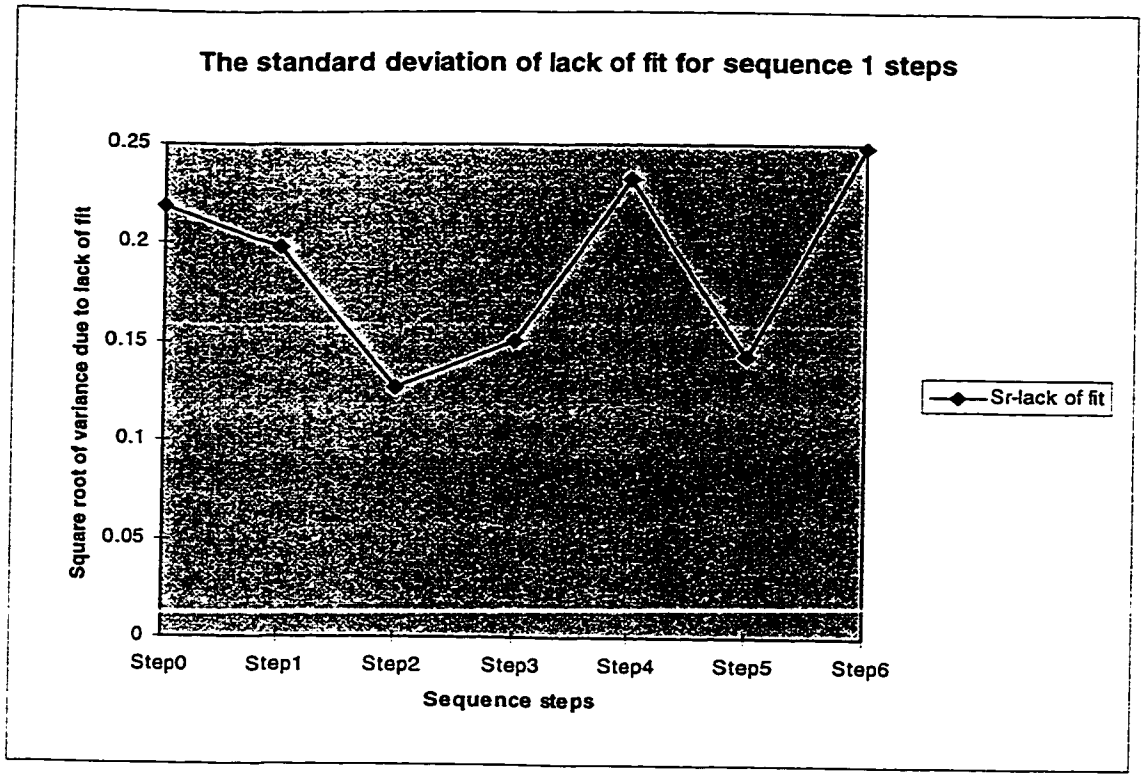
Figure 4.8 Plot of Kx14



Figures 4.9 to 4.11 for sequence 1 show the magnitudes of the squares root of the variances during each step of the procedure. The plots show the relative increase or decrease in standard deviation for the lack of fit, standard deviation for the parameters

and the coefficients of multiple determination for each of the steps in the sequence respectively:

Figure 4.9 The plot of standard deviation for the lack of fit, sequence 1, for each step



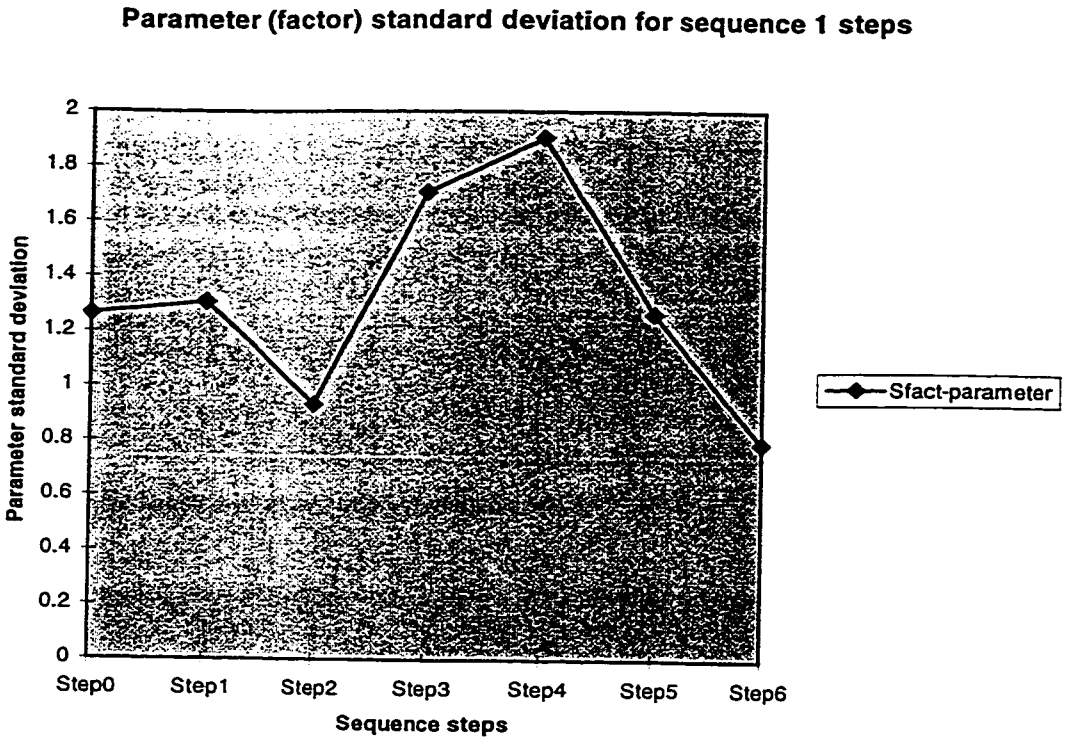
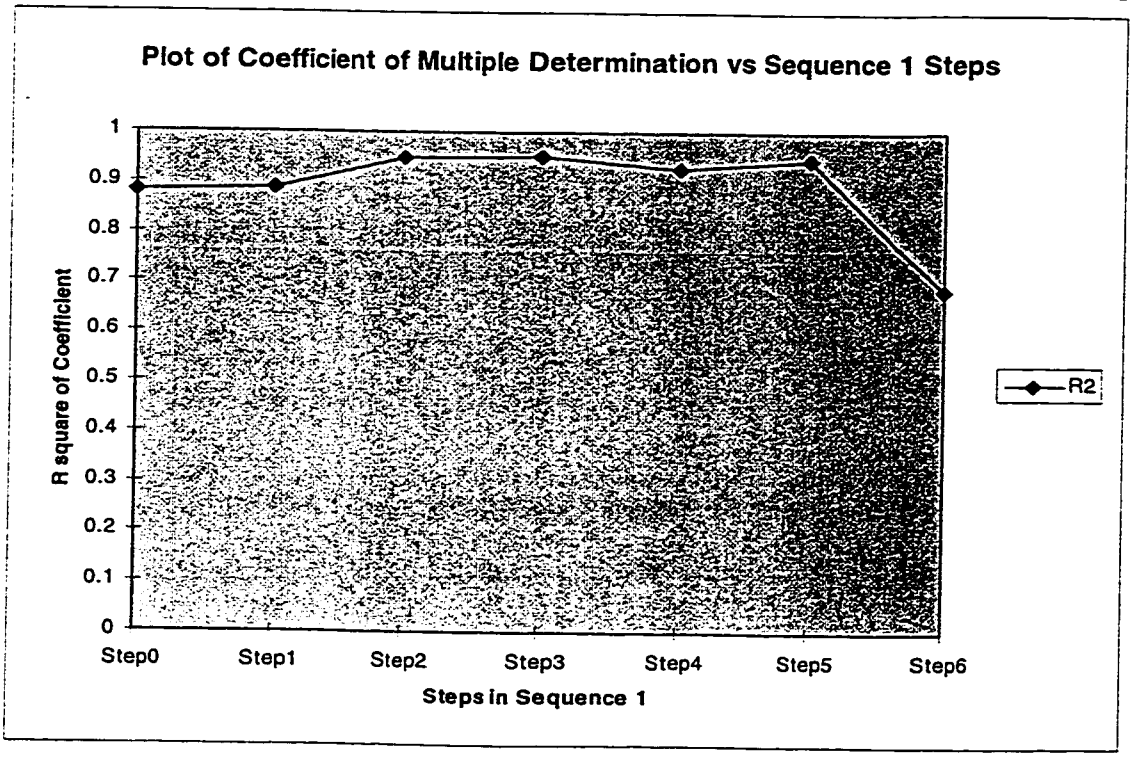


Figure 4.10 The plot of standard deviation for the factors, sequence 1, for each step

Figure 4.11 Plot of coefficient of multiple determination, sequence 1, for each step



It can be seen from figures 4.9 to 4.11 that the lack of fit standard deviation increased where the scenarios S1 and S2 were involved. Both the factor (parameter) standard deviation and the coefficient R^2 decreases towards the end of the sequence for S2.

For sequence 1, the optimized parameters from the output of step six are compared to results from the individual optimization (section 3.2.2) and the expert input cases are plotted for comparison in figure 4.12. The total sum of squares of lack of fit for sequence 1 step 6 is lower than that for the expert input. The errors are more evenly distributed over the six scenarios in the sequential optimization results as compared to those from the expert input. The last step of sequence 1 is the optimization of file S2, which causes a marked decrease in the lack of fit sum of squares. This brings the interesting question to mind: does this sequential procedure retain some "memory" of the optimization steps it went through?

To investigate this question, the results from step 5 are used to calculate the sum of squares for the lack of fit, and is compared to those from step 6 in figure 4.13. It is found that indeed the results from step 5 of sequence 1 does have a higher sum of squares for file S2, prior to the optimization of S2 in step 6, indicating that the sequential procedure retains the some effects of the last step in the process of optimization. Just how much is retained from the previous steps will depend on many factors. The area of search for a minimum in parameter space will ultimately determine the overall sums of squares magnitude. The area of search in turn depends on the starting values (or default values of transmissivities, which in this case are the expert input), the parameters being optimized and the weights used in the prior information. It is also noted that the sum of squares for scenario S1 increases significantly with the optimization of S2 in step 6, while the remaining sums of squares stay more or less the same.

This indicates that for any particular optimization with one of the six scenarios, because of the "unique" locations of the monitoring wells of each scenario and consequently a different objective function, the prior information and weights from a previous file may not ensure the best results in generating the lowest sum of squares for the next optimization. This is acceptable as the objective of the sequential optimization is to obtain a global best fit which is preferred to any individual fit.

Figure 4.12 Comparison plot-sums of squares of lack of fit from end results of sequence1,step 6

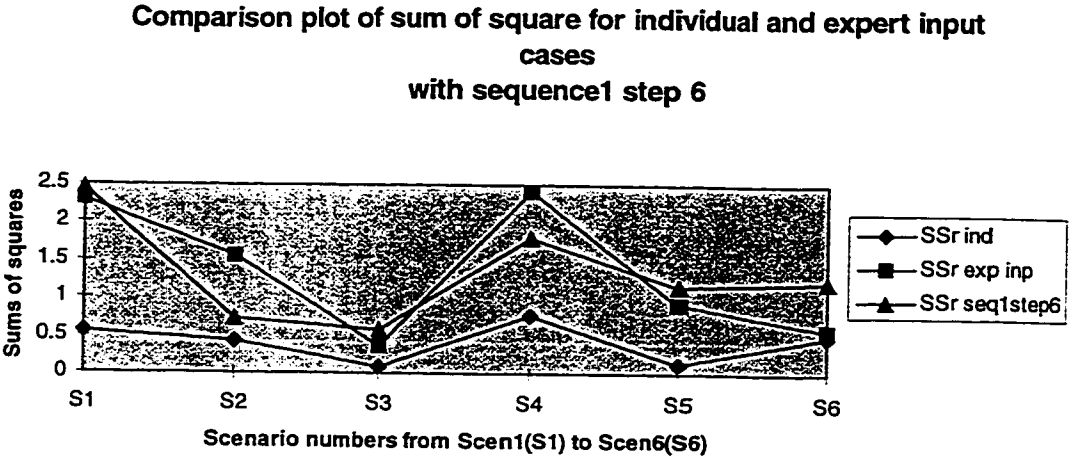
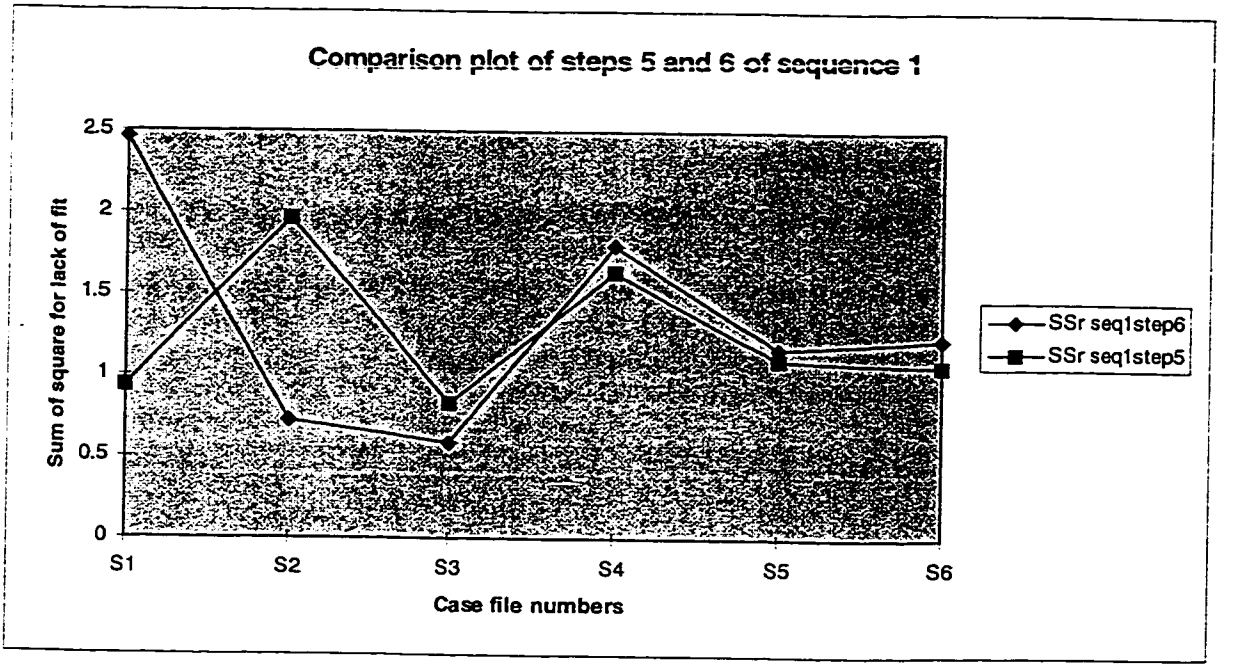


Figure 4.13 Comparison plot of step 5 sequence 1 results to step 6 sequence 1



Thus how well the final optimized results will apply to all six scenarios depends on

1. The order of the sequence:

This is important because when sequence 1 results are compared to those from the reference sequence, the differences are entirely due to the different order of the sequences.

2. The variable weighting of the samples and prior information,
3. The files themselves (how dissimilar they are)
4. The starting file(s) before the penalty parameter becomes large. This factor is especially important for the application of the sffp in section 5. The changes in the penalty parameter (section 2.3.1) for sequence 1 is shown in table 6.2 in section 6, and an illustration is reproduced below:

The penalty parameter c (Wp^2) input to steps, seq 1, 14 zones model

Kx	step 1	step 2	step 3	step 4	step 5	step 6
6	33.99	55.25	398.76	739.86	333.10	127.47

A comparison of the sample and parameter weights is included in section 6.

Both of the total sums of squares for the lack of fit in figure 4.13 are lower than those for the expert input. If SS_{lof} is the only criterion for optimization, the parameters associated with the lowest sum of squares will be taken as the "best" set of optimized parameters.

Other measures of a "best" solution exist. There are the considerations pertaining to the parameter uncertainties, which can become a major objective for optimization in itself (Weiss and Smith, 1998b). A new method is devised and explained in the latter part of this section to help with the decision-making and selection criteria.

With reference to Figure 4.3, the improvements of the sffp sequence 1 on the sums of squares of lack of fit over the expert input are obvious in all step results. Step 4 of sequence 1 reduce the overall sum of squares of the expert input by 13.2%.

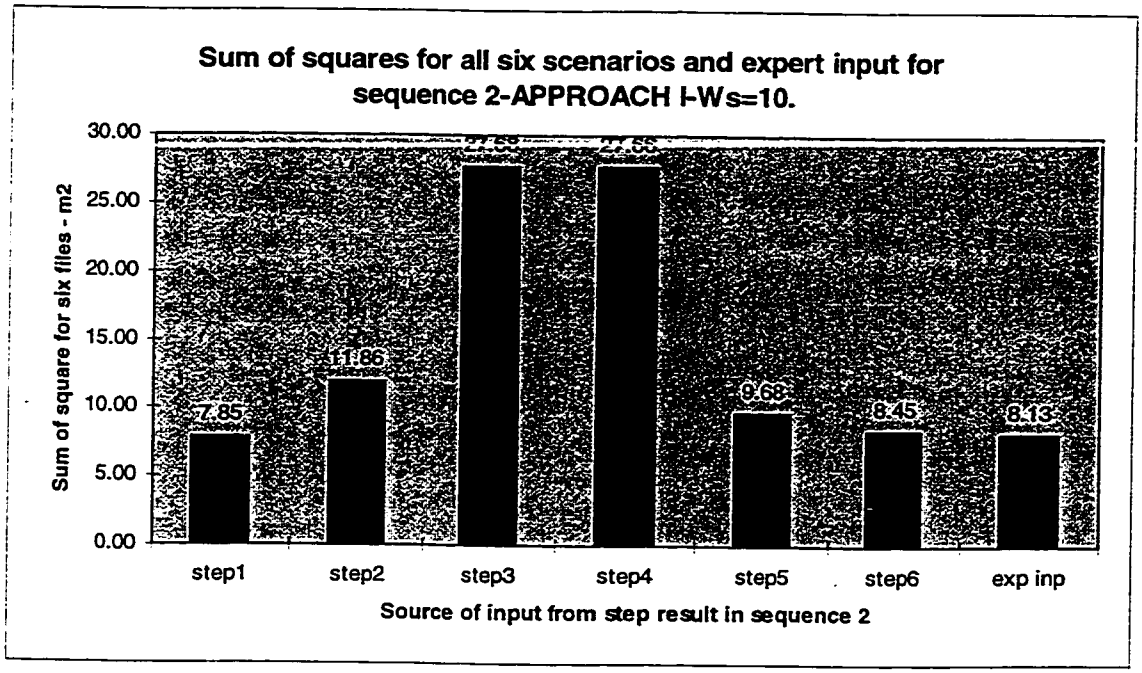
4.2.3 The second sequence

The second sequence is found in table 4.4, and the results in Figures 4.14 and 4.15:

Table 4.4 The steps in sequence 2

Step in sequence	Scenario optimized	Comment
1	S6	Input to S6 from unconstrained optimization of S6
2	S1	Input to S1 from step 1
3	S2	Input to S2 from step 2
4	S4	Input to S4 from step 3
5	S5	Input to S5 from step 4
6	S3	Input to S3 from step 5
		Output from step 6 (S3)

Figure 4.14 Approach I - constant Vs, sequence 2



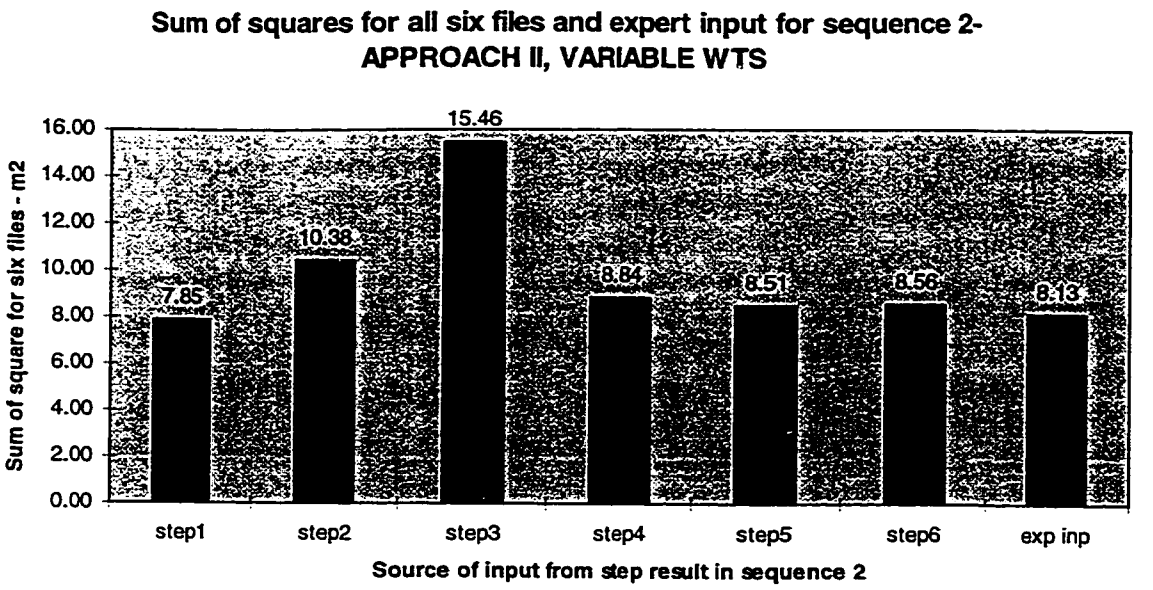


Figure 4.15 Approach II - variable weights, sequence 2

Again, the fluctuations in the SSr from approach I exceed those from approach II. Both sequences reach the lowest sum of squares of the lack of fit at the initial step. The variable weights achieve a much more controlled overall sums of squares than the constant sample weights.

The results for sequence 1 and 2 show that in general, the best and the most stable results in the sffp are obtained from a variable sample weighting approach, because the common variance for the untransformed disturbance of the previous step can match and control that of the present step.

The remainder of this section will dedicated to the results from the second approach only. The sum of squares for the end results (step 6) of this sequence is shown in figure 4.16. The sum of squares of lack of fit for sequence 2 step 6 is slightly higher than that of the expert input.

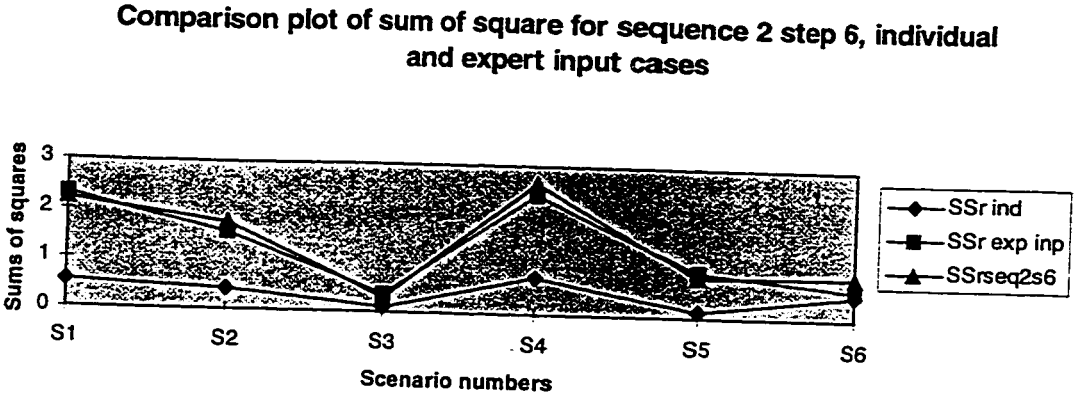


Figure 4.16 Sum of squares for lack of fit for sequence 2 step 6

4.2.4 The third and fourth sequences

The third sequence has scenario S2 as a starting point, as shown in table 4.5, the sum of squares results are shown in figure 4.18:

Table 4.5 structure of the third sequence

Step in sequence	Scenario optimized	Comment
1	S2	Input to S2 from unconstrained optimization of S2
2	S1	Input to S1 from step 1
3	S3	Input to S3 from step 2
4	S4	Input to S4 from step 3
5	S5	Input to S5 from step 4
6	S6	Input to S6 from step 5
		Output from step 6 (S6)

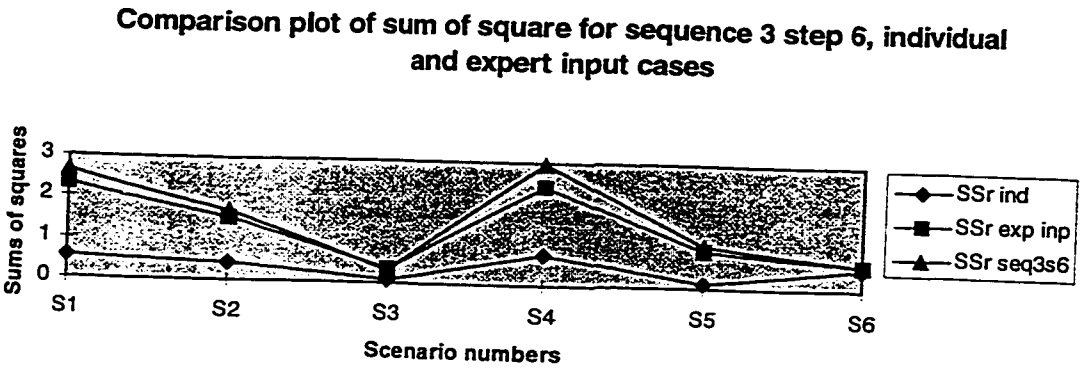
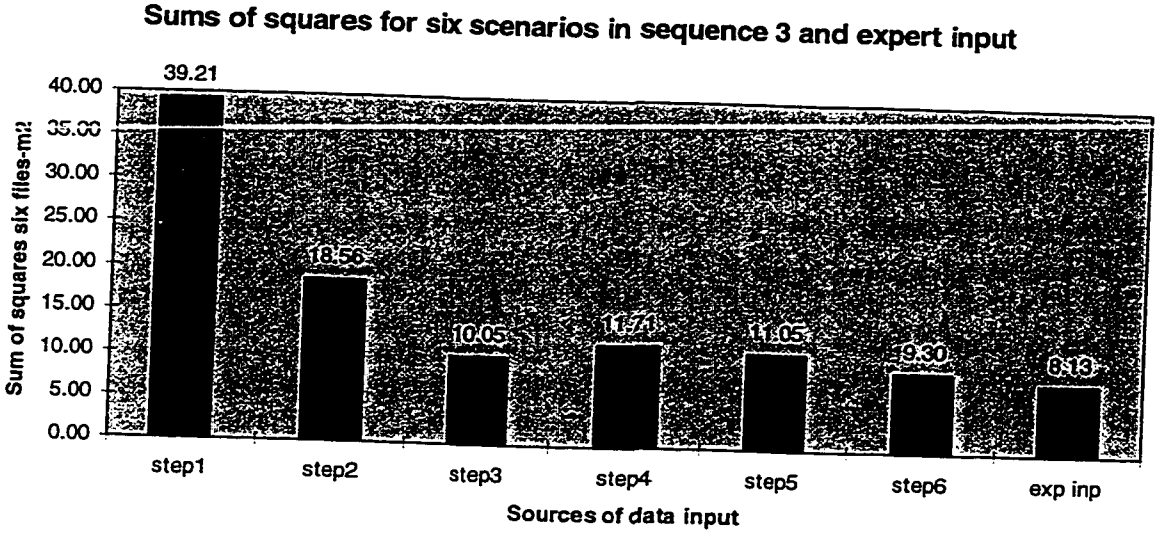


Figure 4.17

Figure 4.18 Overall sums of squares of lack of fit with sequence 3 steps as input



The starting values of some of the parameters in sequence 3 (Kx14,12) remain relatively unchanged during the optimization process. As shown in figures 4.19 and 4.20, the final sum of squares for the lack of fit of sequences 3 is higher than that in the expert input case.

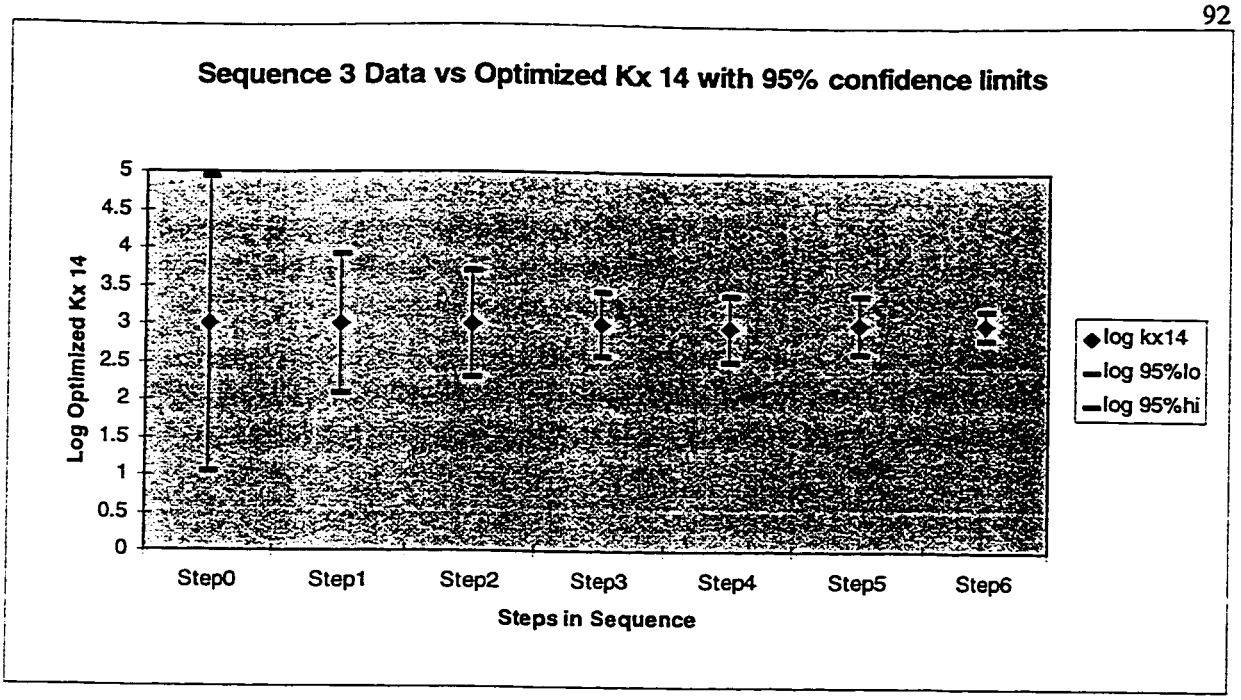
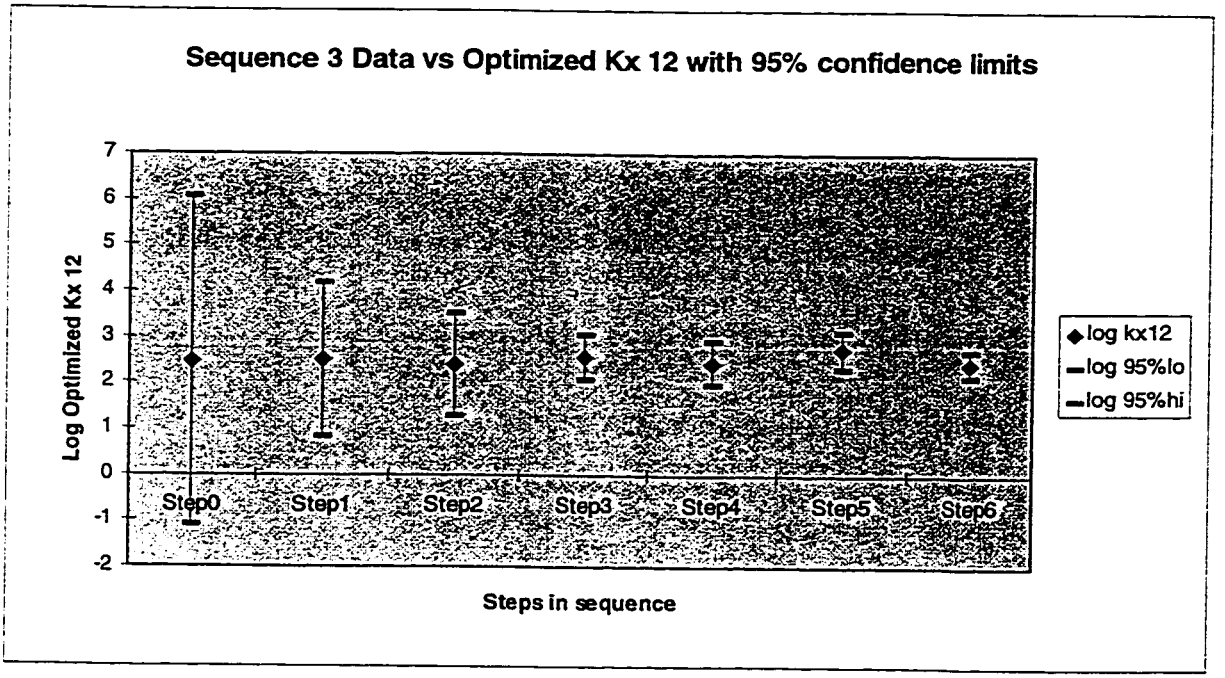


Figure 4.19 Kx14 in sequence 3 optimization

Figure 4.20 Kx12 in sequence 3 optimization



For the other sequences these parameters do exhibit changes in value during the sffp steps by as much as an order of magnitude.

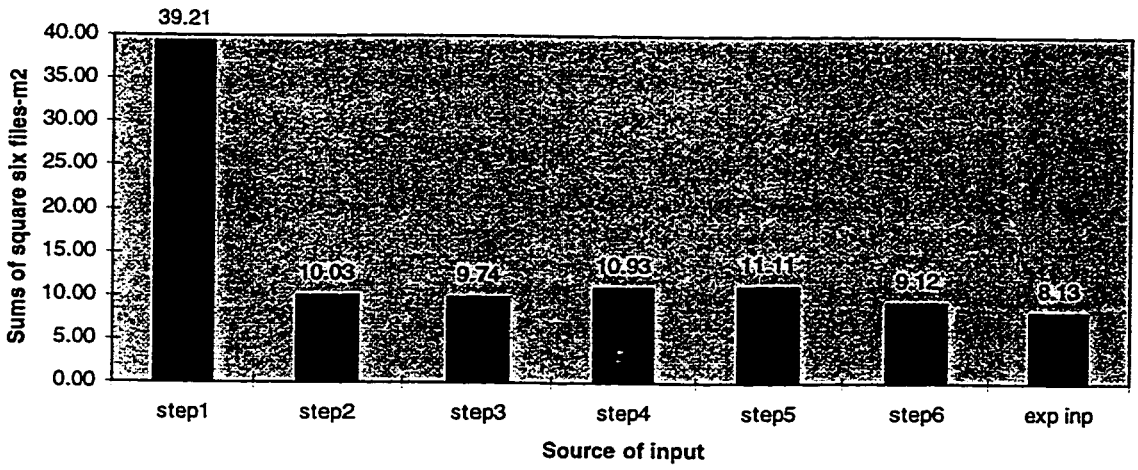
The steps of sequence 4 are listed in table 4.6, and the overall sum of squares in figure 4.21:

Table 4.6 Steps in sequence 4

Step in sequence	Scenario optimized	Comment
1	S2	Input to S2 from unconstrained optimization of S2
2	S3	Input to S3 from step 1
3	S4	Input to S4 from step 2
4	S6	Input to S6 from step 3
5	S5	Input to S5 from step 4
6	S1	Input to S1 from step 5
		Output from step 6 (S1)

The summaries of total sums of squares for sequences 3 and 4 are shown below:

Figure 4.21: Sum of squares for all six files and expert input for sequence 4



The overall sums of squares in sequence 4 are all more than the expert input. The above are the results of the five sequences of optimization for the six scenarios. References to the results will be made in sections that follow for comparison and analyses purposes.

4.5 The use 24 Kx zones for sffp

The discussion in section 1.2 shows the importance of the reduction of the number of parameters in indirect inverse modelling. The use of the coefficient of variation (cov) as a guideline to reduce the dimensionality of decision variables from 24 to 14 and a subsequent verification of the "reduced dimension" model.

The techniques used in the optimization of the 24 zones are based on the findings of the sffp for the 14 zones. It is shown in the 14 zones' sequence 3 optimization that it is possible for the values of the decision variables to remain essentially unchanged during the sequential steps. It is also shown that the sequential starting scenario values of the parameters are important in the subsequent optimization.

With the similarities between the results of section 3.1 and 3.2 on the identifiable parameters, the optimization of the 24 zone model is performed using the 14 zone model results as a guide: The linked zones in the 14 zone model are now de-coupled, and will be optimized as separate zones in the 24 zone model.

4.5.1 The 24 zone optimization strategy

With the results of the previous sections, the optimization process has been streamlined in the following way:

The cov guideline is again used to find the parameters which will be used as decision variables. With the cut-off cov at 8.0, nine zones selected in section 3.1 as the compromise will be used. In addition, zone 19 will be added to the nine, making a total of 10 decision variables - number 10, 11, 12, 13,16,17, 19, 20, 23 and 24. With the exception of zone 19, all were present in the previous 14 zone models as decision variables, although some were linked as one zone (see table 3.2). These will be optimized in a sequential procedure in the same order of sequence 1.

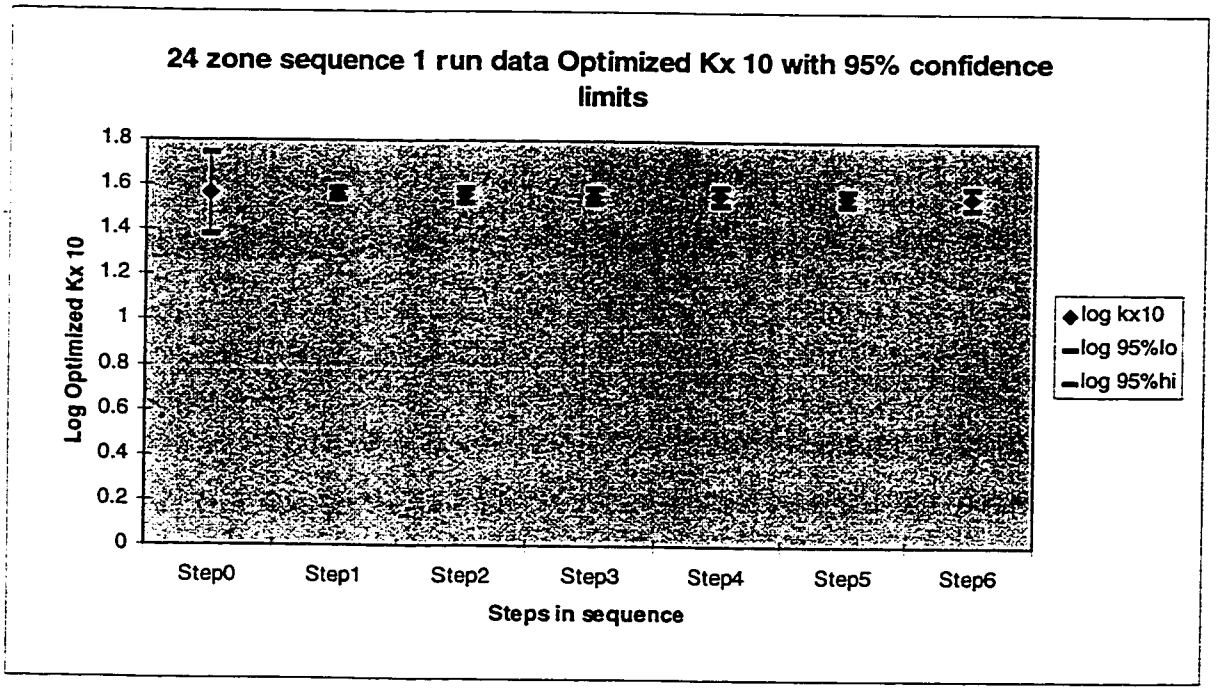
Table 4.7 The 24 transmissivity zones

Hydraulic Transmissivity Zone Number	Kx in m2/day Expert Input
1	.1296
2	.14688
3	.16416
4	.648
5	10.368
6	12.96
7	12.969
8	15.552
9	15.561
10	31.104
11	31.113
12	38.88
13	38.889
14	46.656
15	46.665
16	51.84
17	51.849
18	54.432
19	62.208
20	90.72
21	90.729
22	103.68
23	388.8
24	388.81

The values of the Kx which occupies the same zone in table 3.2 have been modified slightly so that the MODFLOW preprocessor Groundwater Vistas and PEST template file can recognize them as separate zones. The sequence steps and order are the same as those in section 4.2.2. As in the 14 zone model, the optimization begins with an unconstrained optimization, then step by step using variable weights (approach II) and results of the previous step as prior information for the decision variables of the next step. In the unconstrained optimization, the 24 zones are optimized in groups of 5 or fewer zones, to insure stability. Beginning in step 1, all ten zones are used as decision variables in the optimization, with prior information.

The results for the ten decision variables are shown in figures 4.22a to 4.22j, and the sum of square plot is figure 4.23. with step zero denoting the unconstrained optimization step. The actual magnitudes of the variables do not change significantly from step to step.

Figure 4.22a 24 zone model optimization, sequence 1, Kx10



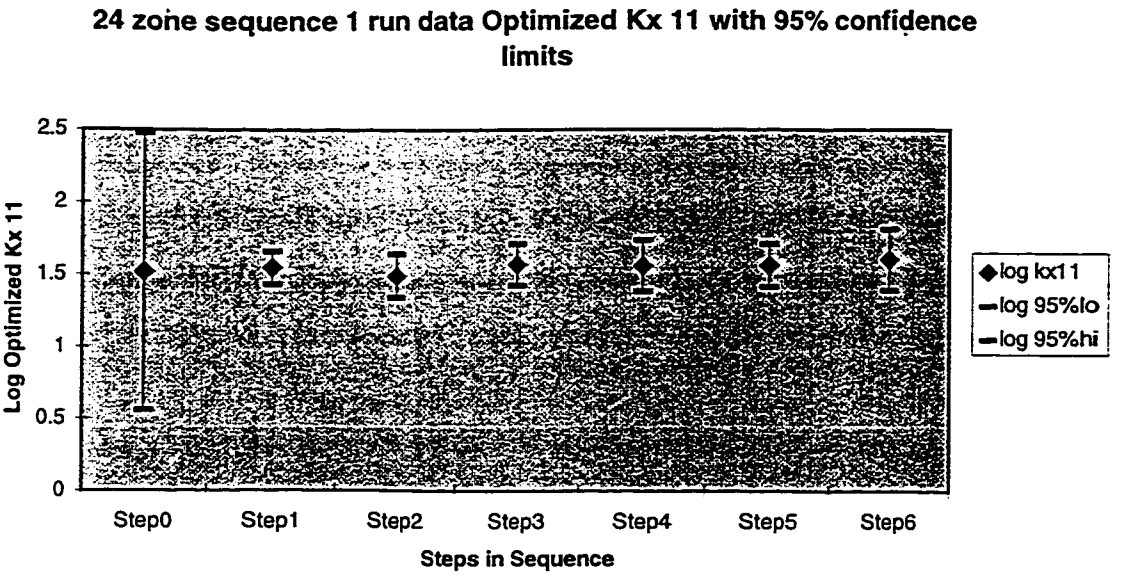
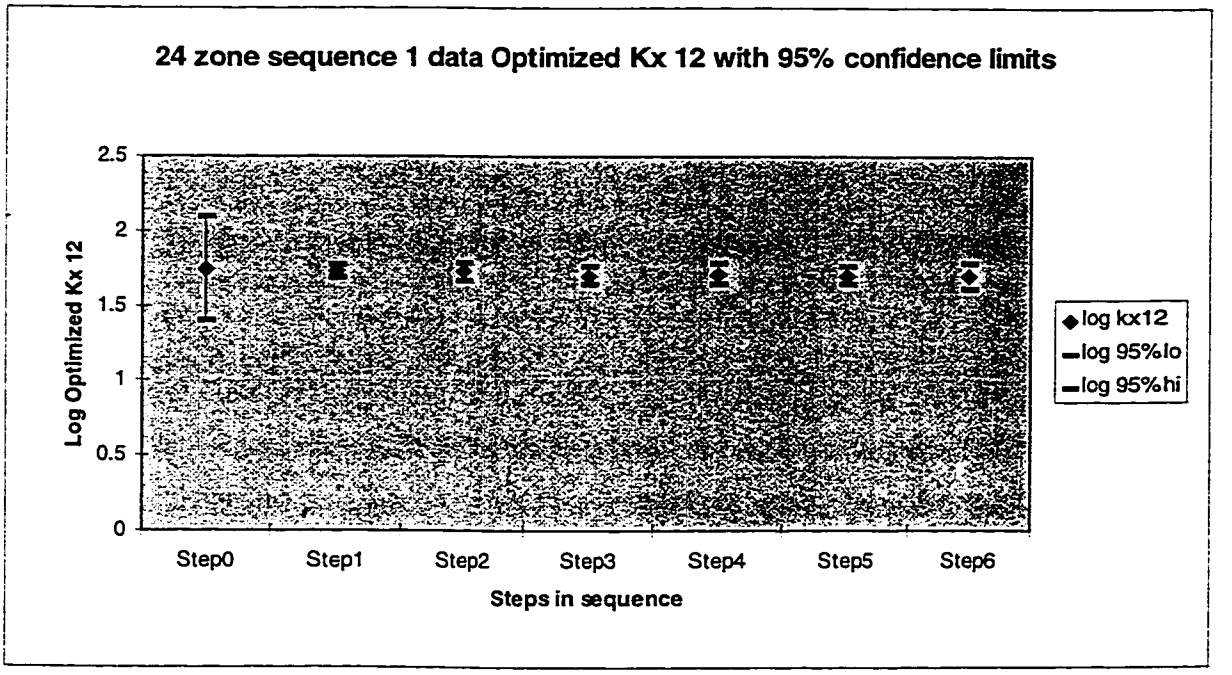


Figure 4.22b 24 zone model optimization, sequence 1, Kx11

Figure 4.22c 24 zone model optimization, sequence 1, Kx12



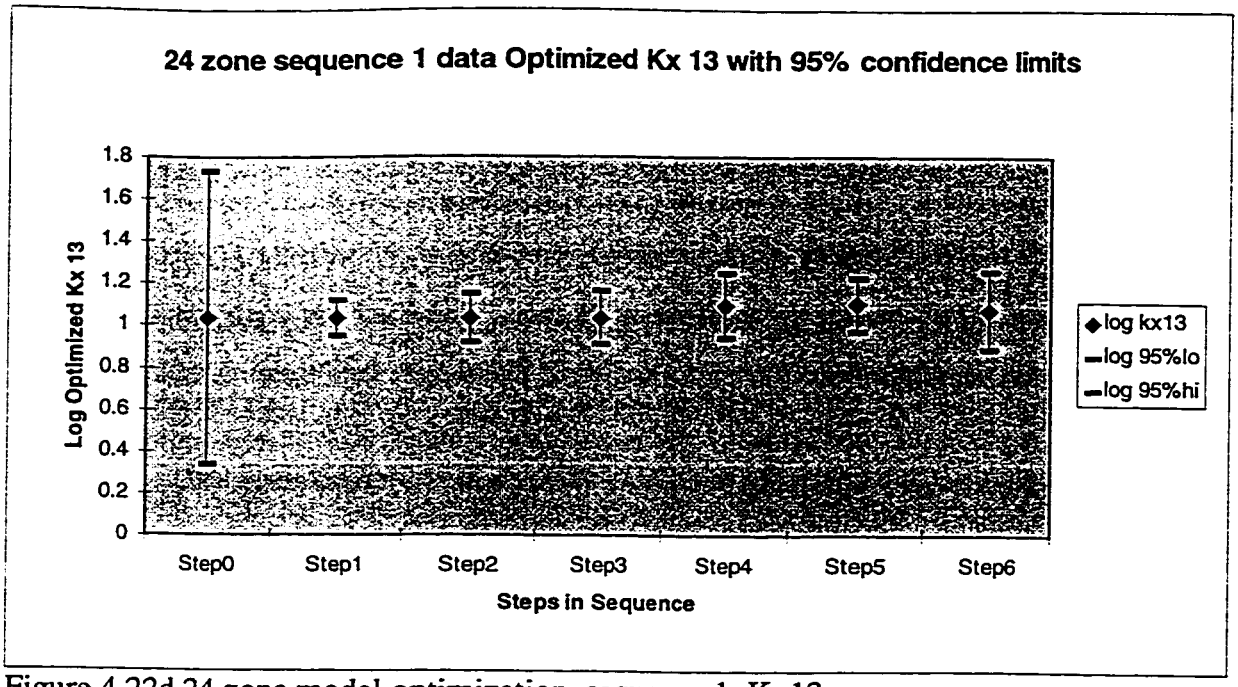
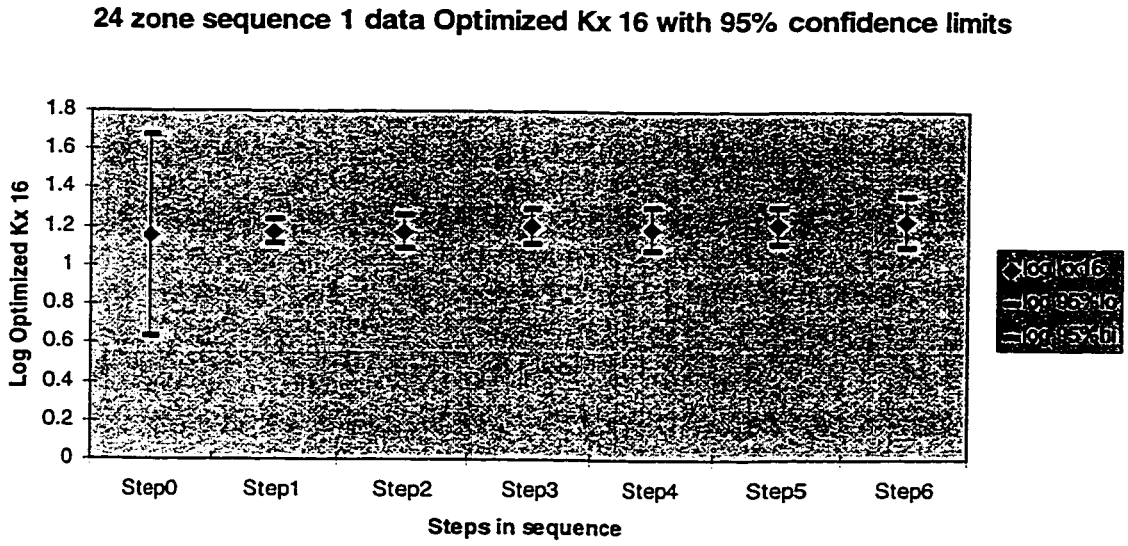


Figure 4.22d 24 zone model optimization, sequence 1, Kx13

Figure 4.22e 24 zone model optimization, sequence 1, Kx16



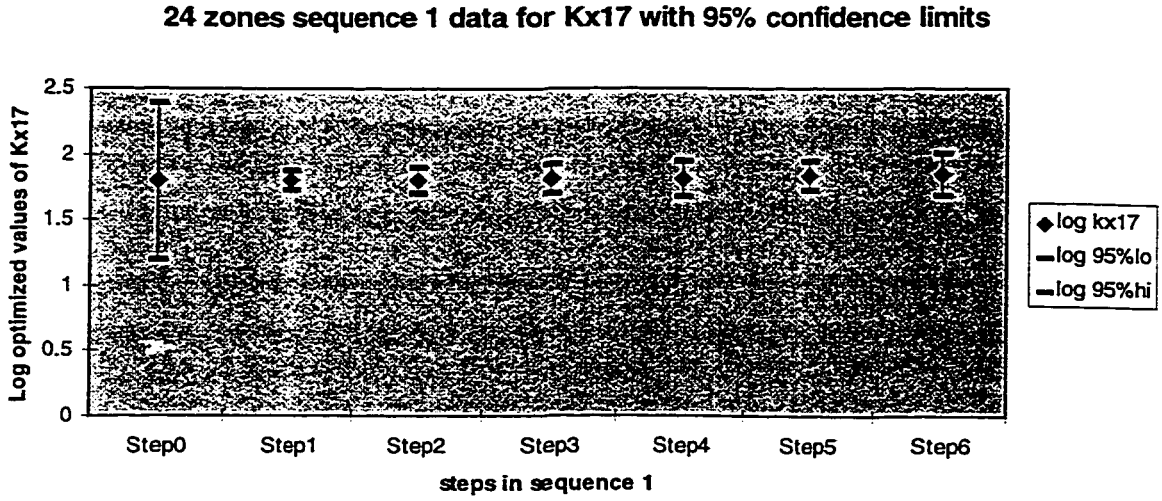
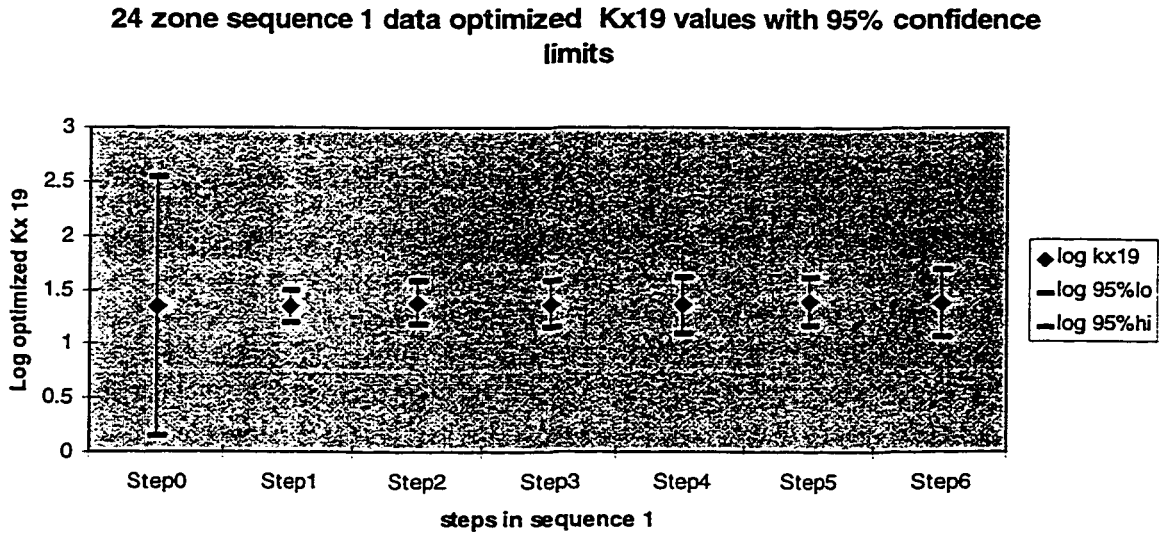


Figure 4.22f 24 zone model optimization, sequence 1, Kx17

Figure 4.22g 24 zone model optimization, sequence 1, Kx19



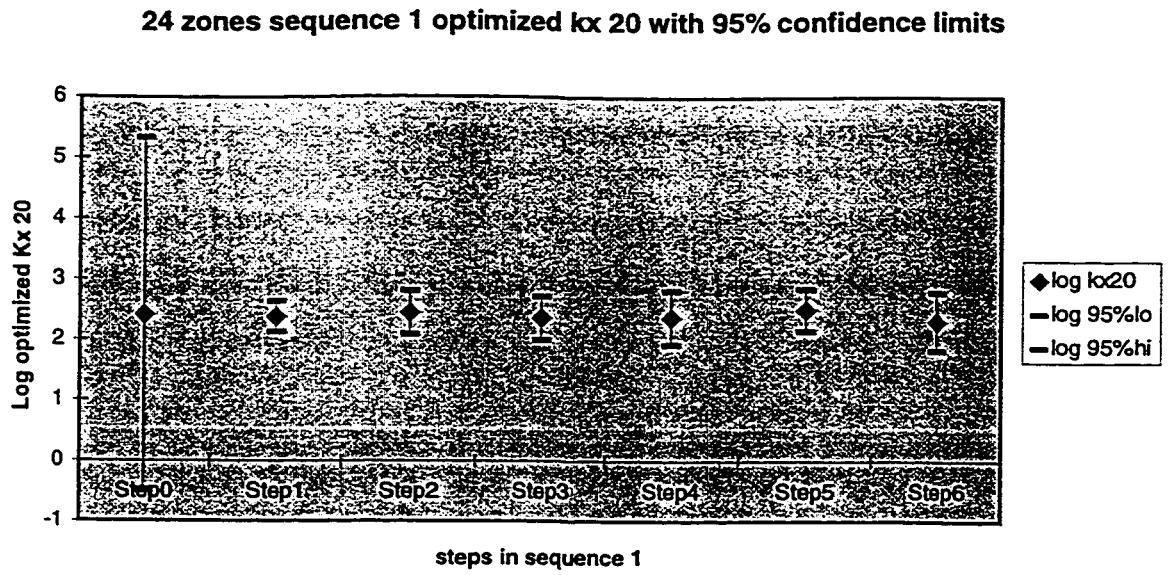


Figure 4.22h 24 zone model optimization, sequence 1, Kx20

Figure 4.22i 24 zone model optimization, sequence 1, Kx23

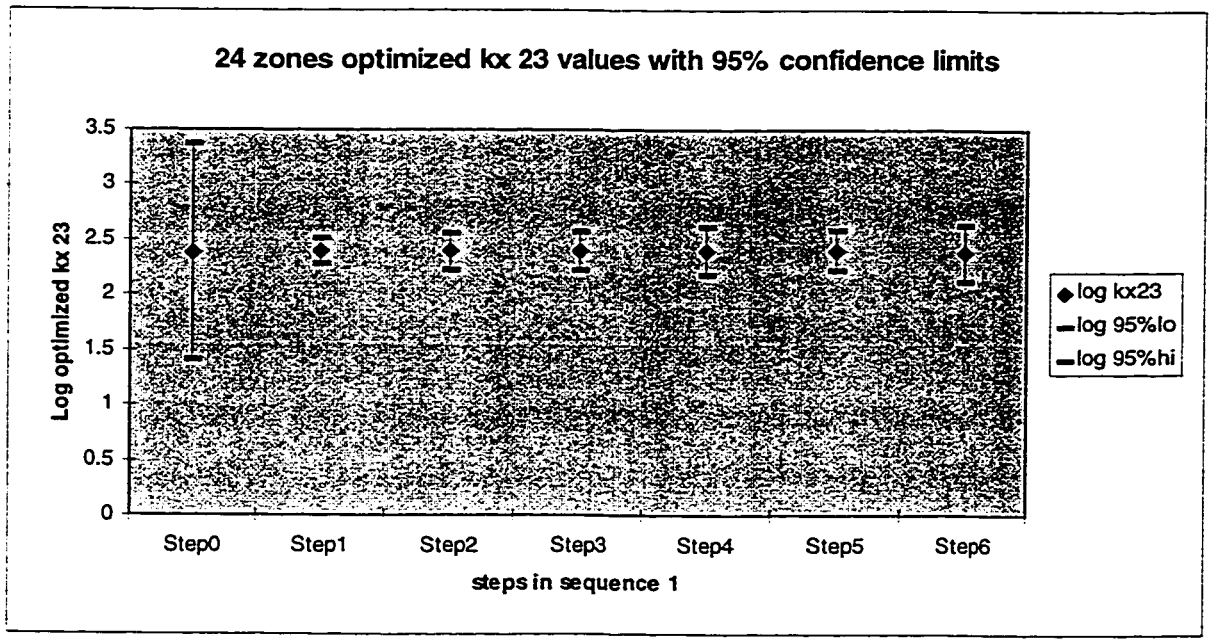


Figure 4.22j 24 zone model optimization, sequence 1, Kx24

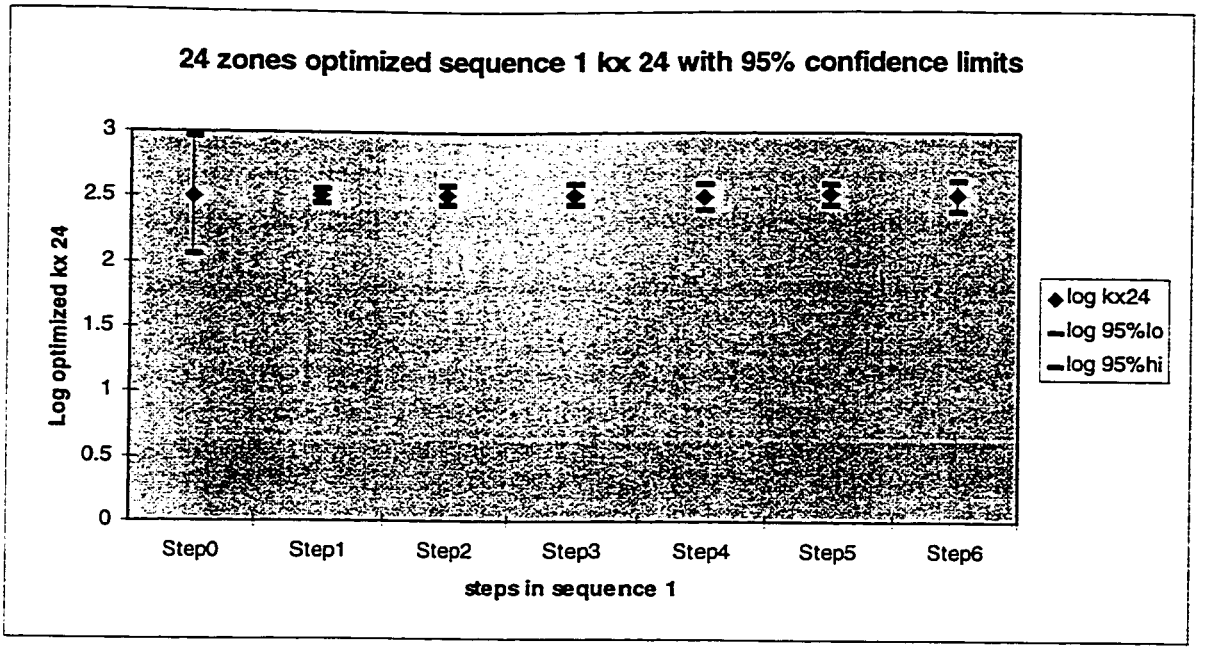
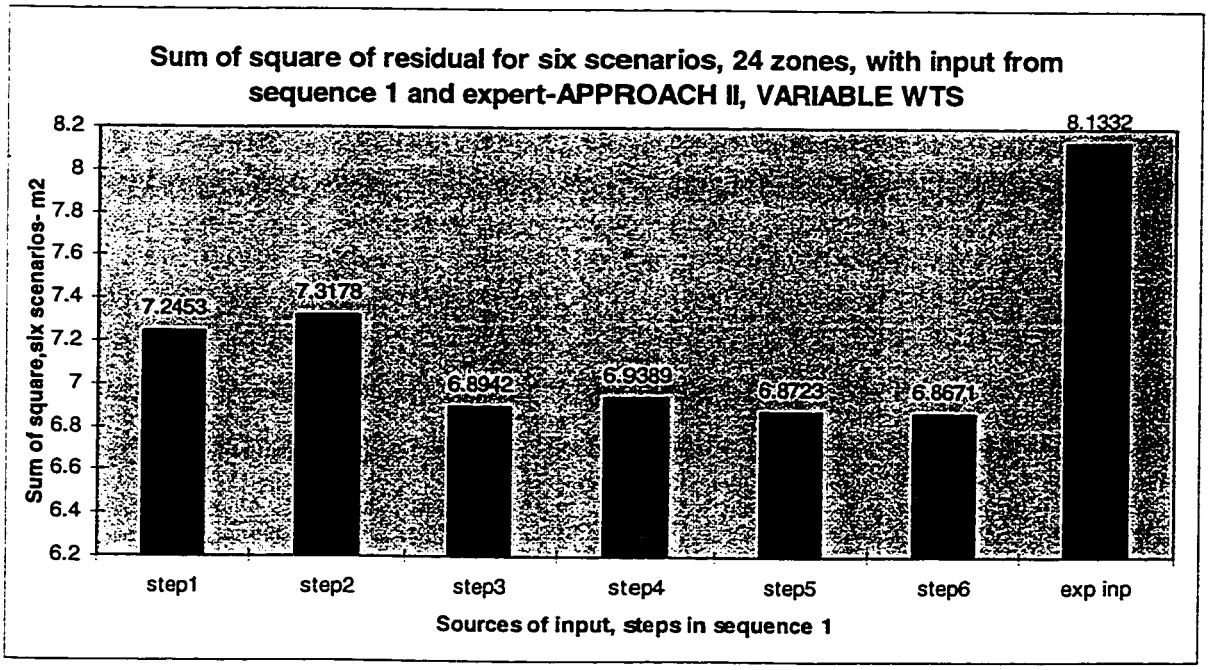


Figure 4.23 Overall sums of squares of residuals with sequence 1 steps as input - 24 zone model



The sums of square of six scenarios of the 24 zone model can be compared to those in figure 4.3 of the 14 zone model (same sequence). The total sums of square for the 24

zone model are very close for all six steps, and the lowest sum of square of 6.8671 in step 6 is an improvement over that of the figure 4.3, which is 7.06 for step 4 of the 14 zone model in the same sequence.

4.6 The ANOVA for the lowest sum of square of lack of fit, 14 and 24 zone models

The focus is now on the optimal parameter sets from the sffp modelling. Two models have been used, a 14 zone model which contains linked zones with identical transmissivities from the expert input, and a 24 zone model which contains all zones created by experts, zones with identical transmissivities have been de-coupled. The models have been explained in detail in section 3.

From the results of the sffp, the lowest sum of square of lack of fit (SS_{lof}) is from sequence 1, step four of the 14 zone model. The lowest SS_{lof} for the 24 zone model is from step six. These results are summarized in table 4.8:

Table 4.8 Lowest SS_{lof} results from optimized 14 and 24 zone models:

Comparison between 14- and 24 zone models
sequence 1 step 4 for 14 zone, step 6 for 24 zone

	14 zone	24 zone
SSfact	3.8438	7.6846
R^2	0.8527	0.8849
Overall SSlof six scenarios	7.06	6.8671
Standard dev lof six scenarios*	1.4758	1.4475
Max parameter est variance	1.88E-02	5.30E-02
Max parameter est stand dev	1.37E-01	2.30E-01
eigenvalue ratio largest/least	6.57	106.48

* Note: The standard deviation of the lack of fit for all six scenarios is calculated by dividing the corresponding SSlof by the degree of freedom and taking the square root. The parameters for the above optimized 14-and 24 zone models are shown below.

The table shows that the 24 zone model has the higher value of R^2, and SSfact. This will mean that the 24 zone transmissivities account for more of the variance in the

data than the 14 zone model, but the small difference in the R^2 value shows that the added parameters in the 24 zone model are better, though not by much. Overall, the 24 zone model parameters are better in reducing the SSlof for all six scenarios than the 14 zone model, but again, not by much (less than 3 %).

The maximum parameter variance for the decision variables is higher in the 24 zone model than the 14 zone model, indicating a higher uncertainty for the 24 zone model parameters. The eigenvalue ratio for the normal matrix is also higher for the 24 zone model, the matrix in the 24 zone model is therefore not as well conditioned. There are both favourable and unfavourable aspects of using the models. The parameters sets are:

Table 4.9 24 zone optimized parameters

Hydraulic Conductivity Zone Number	Kx in m2/day
1	.130*
2	.147*
3	.164*
4	.648
5	10.368
6	12.960
7	12.969
8	15.552
9	15.561
10	35.436
11	40.194
12	49.931
13	11.843
14	46.656
15	46.665
16	16.972
17	70.321

18	54.432
19	23.636
20	206.131
21	90.729
22	103.68
23	236.917
24	327.028

Note:* denotes transmissivities that have been modified slightly so as to separate them. Slight changes were made in the original data in the least significant digit in the m2/sec values, and have been transferred into PEST and shown now in m2/day values.

Table 4.10 Optimized parameter set for lowest SSlof - 14 zones model

Transmissivity zone numbers	Kx m2/day
1	.1296
2	.6480
3	10.368
4	12.96
5	15.552
6	31.4881
7	56.5948
8	46.656
9	25.1224
10	54.432
11	62.208
12	54.4468
13	103.68
14	255.32

4.7 Decision support tool: The Weighted Optimized Non-dimensional Norm (WONN)

This section describes a flexible new tool to select the "best" model and parameter values, not on a single measure such as the lowest sums of squares of the lack of fit, but on a combination of a number of measures that may be relevant to the modeller. A typical collection of these measures is shown in table 4.8 above. Since these measures have unlike units, combining them will require some scaling.

The simplest way to describe this straight forward method is to use the example of the 24 zone versus the 14 zone optimized results:

The 24 zone result (table 4.8), has a total SSlof of 6.8671 m^2 or a standard deviation of lack of fit of 1.4475 m, and the highest parameter standard deviation of 2.3e-1 m2/day.

The 14 zone result has a total SSlof of 7.06 m^2 or a standard deviation of lack of fit of 1.4758 m, and the highest parameter estimate standard deviation of 1.37e-1 m2/day.

Both the total SSlof (or standard deviation, lof) and the variance of parameter estimate are measures to be minimized, and the lower their values, the better. The non-dimensional quantity that can be used to combine these two measures is calculated using the following methodology.

The WONN procedure:

Step 1- choose the measures of "goodness' of solution. In this example, the standard deviation of the lack of fit and the highest standard deviation are selected.

Step 2- The chosen measures of estimated parameters should be better when smaller in order to use the scaling procedure in the next step. If the measures are better when maximized, then the suggested scaling step next should be modified.

Step 3- Form a vector of the same type of measure. In this example, there are two vectors, one for the standard deviation of the lack of fit for the 14 zone model, the other for the highest parameter estimate standard deviation. Each vector will have two elements: one each for the 14 and 24 zone models. Normalize both vectors by using the largest value in that vector. Entries in both vectors will consist of the highest number of 1 and decimals. The elements in the vector are now dimensionless.

Step 4- Extract the dimensionless numbers for each model and form two new vectors, one for the 14 zone, the other for the 24 zone model. Find the length of the vector by taking the root sum of square for the elements of the new vectors. Weights can be added to emphasize or de-emphasize any particular element in the new vectors. If weights are used, the same weights have to be applied to the same elements in both vectors.

Step 5- The weighted or un-weighted shortest length vector will be chosen.

An example using the above numbers are used to perform the above procedure:

1. Scale the standard deviation of the lack of fit and the highest parameter variance by means of the largest value in the set. For the lack of fit it is (1.4475, 1.4758) or (0.9808, 1) for (24 zone model, 14 zone model). For the parameter standard deviation it is (2.3e-1, 1.37e-1) or (1, 0.5957) for (24 zone model, 14 zone model).
2. The non-dimensional norm is the inner product of the non-dimensional vector containing the lack of fit standard deviation and the parameter standard deviation elements for the particular model:

   For the 24 zone model:

$$(0.9808,\ 1)\ .\ \begin{pmatrix} 0.9808 \\ 1 \end{pmatrix} = (0.962 + 1) = 1.962 \qquad (4.1)$$

   For the 14 zone model:

$$(1, 0.5957) \cdot \begin{pmatrix} 1 \\ 0.5957 \end{pmatrix} = (1 + 0.3549) = 1.3549 \qquad (4.2)$$

The weighted optimized non-dimensional norm for the 24 zone model is the square root of equation (4.1), or 1.401, and that for the 14 zone model from equation (4.2) is 1.164. The better choice, taking the lack of fit and parameter standard deviations into account, is the 14 zone model which has a lower norm.

The choice for the 14 zone model is obvious, because the model with the lowest norm will have the best combination of sum of square of lack of fit and parameter uncertainty.

The above example treats both elements in the vector with equal emphasis, and the norm is calculated as if the vector elements are orthogonal to each other. While the elements of the vector are different aspects of the same optimization, and thus may not be truly orthogonal in a strict mathematical sense, they are, however, important indicators of the quality of the optimization, and as such can be treated equally by the modeller.

In the case when the modeller assigns more importance to one element of the vector than the others, it can be achieved by giving weights to the elements. Again, using the above example, if the standard deviation of the lack of fit is deemed 5 times as important as the parameter standard deviation, a factor of 5 is used:

For the 24 zone model:

$$(0.9808*5, 1) \cdot \begin{pmatrix} 0.9808*5 \\ 1 \end{pmatrix} = (0.962*25 + 1) = 25.05$$

For the 14 zone model:

$$(1*5, 0.5957) \cdot \begin{pmatrix} 1*5 \\ 0.5957 \end{pmatrix} = (1*25 + 0.3549) = 25.3549$$

The 24 zone model weighted optimized non-dimensional norm (WONN) is now 5.005, and that for the 14 zone model is 5.035, thus making the 24 zone the better choice. The use of weights can place emphasis on any element at all in the vector.

The vector may contain more than two elements. In addition to the above, the ratio of eigenvalues can be used as an indication of the conditioning of the matrices, and can be included in the vector (ratios from table 4.8). Again, the lower the eigenvalue ratio gives the better conditioned matrices:

For the 24 zone model:

$$(0.9808*5,\ 1,\ 1)\ .\ \begin{pmatrix} 0.9808*5 \\ 1 \\ 1 \end{pmatrix} = (0.962*25 + 1 + 1) = 26.05$$

For the 14 zone model:

$$(1*5,\ 0.5957,\ 0.0617)\ .\ \begin{pmatrix} 1*5 \\ 0.5957 \\ 0.0617 \end{pmatrix} = (1*25 + 0.3549 + 0.0038) = 25.3587$$

Calculating the wonn's will make the 14 zone model a better choice again. Other measures such as computer computation time required can be added to this calculation of norm to find the best model. More than two model choices can be compared at once. Any norm with 3 components or fewer can be presented graphically.

5.0 Application of the sequential feed forward procedure (sffp) to an arbitrary starting point simulation (asp) of the six scenarios

The optimization results of the last sections are derived, using the expert input as starting points for the transmissivities of the numerical model. The purpose of this section is to start the optimization of all the transmissivities at some arbitrary point, with minor differences for each of the 14 zones so that the PEST template program can distinguish each zone, to examine the effects of the starting point on the sffp.

The starting points for the transmissivities are arbitrary, yet reasonable. They represent some average physical transmissivity values of the soil medium in the area. In this case, the geometric mean of the expert input transmissivities is adopted, then it is varied by the least significant digit for each of the 14 zones to form the starting points. A table of the starting values is shown in table 5.1.

This simulation is important for the modelling of groundwater flow because, if successful, the sffp will enable a modeller to obtain a set of transmissivities for the zones of the model from the measurements, through an automated procedure and with significantly reduced input from the expert hydrogeologists.

5.1 The sffp algorithm for the arbitrary starting point (asp) scenarios

The following algorithm is developed for the asp simulations:

1) The 14 zone model is used in the asp simulations. 14 transmissivities at or close to the geometric mean of the original expert input values (of 22.0304 m2/day) are tabulated in table 5.1:

Table 5.1 14 zones of transmissivity for unconstrained optimization, asp simulations

Zone number	Transmissivity, m2/day	Log transmissivity
1	22.0304	1.343022

2	22.0305	1.343024
3	22.0306	1.343026
4	22.0307	1.343028
5	22.0308	1.343030
6	22.0309	1.343032
7	22.0310	1.343034
8	22.0311	1.343036
9	22.0312	1.343038
10	22.0313	1.343040
11	22.0314	1.343042
12	22.0315	1.343044
13	22.0316	1.343046
14	22.0317	1.343048

These will be used as starting values throughout the unconstrained optimization, and will only be replaced by selected results from the optimized transmissivities from the sffp.

2) A detailed screening process for all parameters in the 14 zone model is performed using unconstrained optimization. Sets of decision variable with 5 parameters or fewer for each of the scenarios 3 to 6 will be optimized each time, to cover all 14 zones. This is necessary in order to have a small dimensioned decision variable space to avoid optimization-related dry cell problems described in section 3, and to make use of all of the information in the data sets.

Recall the screening procedure in section 3, in which the screening is performed with only scenario 4. Scenario 4 was selected previously because the unconstrained optimization was on groups of up to 10 of decision variables. Due to this higher dimensionality of the variables, the optimization encountered difficulties with both dry cell occurrences which terminated unconstrained optimization attempts, and many insensitive parameter values which stayed at the upper or lower bounds of the optimized parameter range. The dimension of decision variables was not reduced in that

optimization, therefore only scenario 4, which performed the best under those conditions, was used as the screening scenario in section 3.

In the present case, when dry cells occur in the unconstrained optimization of 5 decision variables, the dimensionality of the optimization is again reduced: two groups containing 3 and 2 decision variables are used in two separate unconstrained optimizations. By successive reduction of the dimensionality of decision variables, dry cell occurrence is avoided.

3) The unconstrained optimization results are tabulated and screened by using the log-transformed values of the means and variances of the parameters. The log-transformed values of the transmissivities are assumed to obey the normal distribution, and hence are used as numbers in subsequent calculations.

The criterion for selecting "suitable" optimized transmissivities to replace those in table 5.1, and (or) to be used as prior information in the sffp steps to follow, is based on the raw coefficient of variation and the variance of the parameter. The raw coefficient of variation (cov) is calculated by using equation (3.1).

The cov is not an absolute measure of the variation of the parameter. For the same cov, a parameter with a higher mean can tolerate a higher standard deviation. Therefore, an additional requirement is imposed on the magnitude of the log standard deviation estimate of the parameter, to ensure the proper selection of parameter values. To require the parameter log standard deviation to be zero is not realistic, therefore an arbitrary reference is set. It is the square root of the reference variance of the measurements for the starting scenario of sequence 1, which is S4. The numerical value of this reference is 0.39. The advantages of using a reference like this are that the level of the reference is low, and that the parameter log standard deviation estimates close to this value, both higher and lower than 0.39, can be of equal importance.

The difference between 0.39 and the log parameter standard deviation, together with the values of cov, will form the axes in the weighted optimized non-dimensional norm (WONN) for the choice of the parameter as a decision variable.

The application of the selection criterion will be illustrated below with reference to the optimized values from the unconstrained optimizations.

4) After applying the WONN selection process in step 3, selective starting values for transmissivities are identified from the unconstrained optimization results. These starting values take the places of the default values in table 5.1. Also, prior information for the first 5 decision variables are obtained from the same process.

The order of the first sequence (see section 4) is used for the cycle 1 optimization. After the sequence of optimization is completed, the optimized set of parameters that gives the lowest SSlof for all six scenarios will replace the starting values of the parameters in the table of transmissivities as the new default values. This is the end of cycle 1 of the optimization.

5) Cycle 2 of the process begins the same as cycle 1, with a higher cov cutoff and the log standard deviation requirement. 5 different parameters are selected and optimized again with the order of sequence 1. At the end of cycle 2, five more optimized parameters of this cycle which produce the lowest sum of squares of the lack of fit will replace the existing default values of transmissivities.

6) Cycle 3 or any subsequent cycles that are necessary, are performed in the same manner as for cycle 2, on the remaining decision variables. If, at the end of cycle 3, the 14 zones are updated completely, then the new transmissivities can replace the arbitrary starting point database of transmissivities in table 5.1.

7) If desired, knowledge from the first pass that optimize all 14 zones can be used as prior information on a second pass of optimization, in which cycles are repeated as in the above steps, to obtain a lower overall sum of squares of lack of fit. This will improve on the performance of the asp simulations.

The new set of 14 transmissivities will be compared to the expert input.

5.2Asp simulation, first pass, cycle 1

Following steps 2 3 and 4 of the above algorithm, the unconstrained optimization results for the parameters and their log transformed values, the cov calculation with equation 3.1, results are in tables 5.2, 5.3.

The selection of the parameters using the cov in table 5.2 and the log standard deviation is described below:

The cut off cov is set at 1.2. This value is chosen to include approximately five of the parameters for decision variables. The additional requirement is that the parameters that

Table 5.2 Log mean and Var(log T) of parameters from unconstrained optimization

Source scenario for log mean						table of var (log T)			
Kx	S3	S4	S5	S6	Kx	S3	S4	S5	S6
1	3.6021	4.5899	3.6021	3.1461	1	6.54E+00	4.31E+00	1.70E+02	3.42E+00
2	-2.0516	3.6021	-1.1107	-0.1568	2	451.8	177.9	10.77	3.398
3	2.7340	1.1741	3.6021	1.4821	3	96.88	25.97	198.7	4.527
4	3.2923	-1.9547	1.2695	2.5682	4	49.67	17.29	0.4979	0.4409
5	3.6021	3.6021	0.0826	3.5884	5	6.26	6.6	3.404	0.9475
6	0.2147	0.8320	1.3070	1.2850	6	3.04E-02	0.327	5.95E-03	1.04E-03
7	0.8086	2.1847	1.3825	1.0686	7	8.79E-02	0.5773	4.66E-02	7.17E-03
8	-1.3936	0.9522	1.9058	2.3284	8	4.134	0.3914	2.883	0.7749
9	1.1015	-0.4947	1.3045	0.9673	9	0.1723	44.12	6.80E-02	3.42E-02
10	-3.2840	1.2994	-1.4672	0.1072	10	1.59E+01	6.40E+01	1.51E+00	3.49E+00
11	3.1180	1.1338	-0.4336	1.0660	11	1.485	0.1849	7.052	0.5209
12	1.5136	3.3127	2.3338	1.4231	12	2.37E-01	2.13E+00	1.07E+01	1.50E-01
13	1.8899	1.4041	2.6410	1.0669	13	1.016	6.867	1.219	0.1857
14	2.5865	2.6923	1.7445	2.0204	14	5.43E-02	3.80E-02	2.08E-02	1.04E-02

Table 5.3 Log standard deviation of parameters, cov using transmissivity standard deviation and mean (equation 3.1)

Source scenario for log standard deviation-sd					Cov using ln values and equation 3.1				
Kx	S3	S4	S5	S6	Kx	S3	S4	S5	S6
1	2.5573	2.0755	13.0212	1.8494	1	[3.38e7]	[9.16e4]	[9.99e52]	8.6810E+03
2	21.2556	13.3385	3.2810	1.8432	2	[9.99e52]	[9.99e52]	2.5085E+12	8.1676E+03
3	9.8428	5.0969	14.0969	2.1271	3	9.9900E+52	7.9262E+29	[9.99e52]	1.6289E+05
4	7.0477	5.1584	0.7056	0.6640	4	1.5300E+57	8.0510E+19	3.6070E+00	3.0588E+00
5	2.5020	2.5694	1.8451	0.9734	5	[1.61e7]	[3.97e7]	8.2985E+03	[1.23e1]
6	0.1745	0.5718	0.0771	0.0322	6	4.1847E-01	2.1591E+00	1.7895E-01	7.4179E-02
7	0.2965	0.7598	0.2160	0.0847	7	7.7050E-01	4.5105E+00	5.2972E-01	1.9690E-01
8	2.0332	0.6256	1.6981	0.8803	8	5.7470E+04	2.6393E+00	2.0853E+03	7.7364E+00
9	0.4151	6.6425	0.2608	0.1848	9	1.2219E+00	6.2377E+50	6.5896E-01	4.4555E-01
10	3.9862	7.9983	1.2279	1.8675	10	[1.97e18]	9.9900E+52	5.4460E+01	1.0341E+04
11	1.2186	0.4300	2.6551	0.7217	11	5.1237E+01	1.2905E+00	1.3150E+08	3.8507E+00
12	0.4871	1.4594	3.2672	0.3877	12	1.5871E+00	2.8405E+02	1.9244E+12	1.1039E+00
13	1.0080	2.6201	1.1038	0.4309	13	1.4747E+01	8.0524E+07	2.5298E+01	1.2949E+00
14	0.2330	0.1959	0.1430	0.1038	14	5.7753E-01	4.7253E-01	3.4124E-01	2.3853E-01

Note: [ ] denotes parameters at maximum or minimum bounds of optimization, and are thus eliminated from selection process.

9.99 e52 is the maximum value for cov with log variance greater than 50.

are below the cov cut off should have a log standard deviation closest to 0.39 (see step 3 above). The parameter below the cov cutoff and with log sd value closest to 0.39 will be chosen. A parameter by parameter discussion follows (default values are those in table 5.1), with reference to tables 5.2, 5.3.

Kx1 : Use default value to start.

Kx2 : Use default value to start.

Kx3 : Use default value to start.

Kx4: Use default value to start.

Kx5: Use default value to start.

Kx6 : 3 with cov below cutoff - S3, S5, S6. Use WONN to choose decision variable and start value:

The WONN matrix for Kx 6

Source of cov	S3	S5	S6
cov vector	.41847	.17895	.074179
normalized cov	1	.42763	.17726
(0.39-log sd) vector	.2155	.3129	.3578
normalized sd diff	.60229	.87451	1.
WONN	1.16737	.97023	1.01559
Comment	reject	accept	reject

The S5 unconstrained optimization result is chosen to be starting value and decision variable. Value of parameter is 20.2791 m2/day, log value is 1.307049, in table 5.2.

Kx7 : 3 cov below cutoff: S3, S5, S6. Use WONN to choose decision variable and start value:

The WONN Matrix for Kx7

Source of cov	S3	S5	S6
cov vector	.77050	.52972	.1969
normalized cov	1	.6875	.2555
(0.39-log sd) vector	.0935	.174	.3578
normalized sd diff	.2613	.4863	1
WONN	1.0284	.8337	1.0321
Comment	reject	accept	reject

The S5 unconstrained optimization result is chosen to be the starting value and decision variable. Value of parameter is 24.1245 m2/day, log value is 1.3825 in table 5.2.

Kx8 : Use default starting value.

Kx9 : Two cov values below cutoff, but their magnitudes are close together (.659, .446). Place more emphasis on the log sd's closeness to 0.39 by either using just the

(.39-log sd) axis for wonn or place a weight of 2 on the log sd difference to use both axes. The latter is adopted:

The WONN Matrix, Kx 9

Source of cov	S5	S6
cov vector	.659	.4456
normalized cov	1	.6762
(0.39-log sd) vector	.1292	.2052
normalized sd diff	.6296	1
WONN- wt of 2 on sd difference	1.6080	2.1112
Comment	accept	reject

Decision variable and default value chosen to be 20.1583, log value is 1.304454 in table 5.2.

Kx10 : Use default value to start.

Kx11 : Use default value to start.

Kx12 : 1 with acceptable cov and log sd. Use as starting value and decision variable. Value is 26.4914, log mean=1.423105 in table 5.2.

Kx13 : Use default to start.

Kx14 : All cov and log sd acceptable and cov values are close in value. Place more emphasis on the log sd difference by applying a weight of 2, same as for Kx 9.

The WONN Matrix, Kx 14

Source of cov	S3	S4	S5	S6
cov vector	.5775	.4725	.3412	.2385
normalized cov	1	.8181	.5908	.4130
(0.39-log sd) vector	.157	.1941	.247	.2862

normalized sd diff	.5486	.6782	.8630	1
WONN- wt of 2 on sd difference	1.4845	1.5841	1.8244	2.042
Comment	accept	reject	reject	reject

Choose S3 unconstrained value as starting value and decision variable: value is 385.914, log mean = 2.586491 in table 5.2.

To summarize, the starting values of transmissivities for sequence 1, cycle 1 are in table 5.4. The decision variables have been identified above, with their weights for the prior information of the parameters. The decision variables are, Kx6, 7, 9, 12 and 14, same as in section 3 and 4.

Sequence 1 is executed with the starting values in table 5.4, and the 5 decision variables. The results of the sequential optimization from each step are then used to replace the starting values in table 5.4 for each of the six scenarios to calculate the SSlof. The total SSlof for each set of input is plotted against the expert input total SSlof for comparison, and is shown in figure 5.1.

The parameter set that produced the lowest SSlof (step 4) is used to replace the values of Kx 6, 7, 9, 12, 14 for cycle 2. The starting values of cycle 2 will utilize the optimization results of cycle 1.

5.3 Asp simulation, first pass, cycle 2

The starting values of cycle 2 is in table 5.5 below. The selection of the decision variables for this cycle is based on a higher cov cutoff of 10. The value of the 0.39 for log sd is not enforced in this cycle, except in the case where more than one choice of parameter value is available from table 5.3, the parameter with log sd value closest to 0.39 will be chosen.

After a selection process similar to cycle 1, the decision variables with the prior information weights for the parameters are described below.

First pass, cycle 1 starting values		
Kx	Values	log values
1	22.0304	1.3430
2	22.0305	1.3430
3	22.0306	1.3430
4	22.0307	1.3430
5	22.0308	1.3430
6	20.2791	1.3070
7	24.1245	1.3825
8	22.0311	1.3430
9	20.1583	1.3045
10	22.0313	1.3430
11	22.0314	1.3430
12	26.4914	1.4231
13	22.0316	1.3430
14	385.914	2.5865

Table 5.4

Figure 5.1

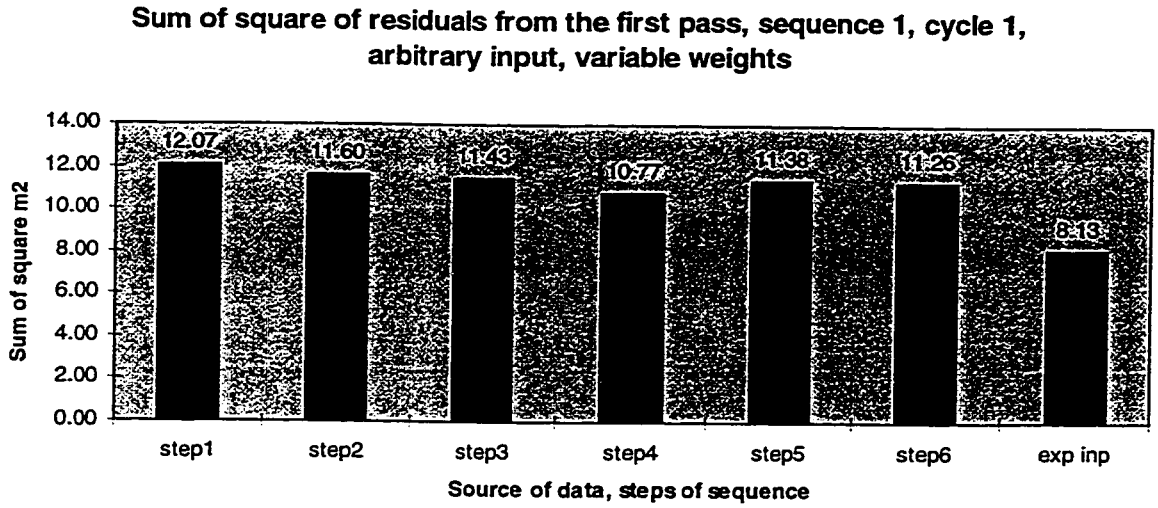


Table 5.5

First pass,cycle 2 starting values		
Kx	Values	Log values
1	22.0304	1.3430
2	22.0305	1.3430
3	22.0306	1.3430
4	22.0307	1.3430
5	22.0308	1.3430
6	21.7833	1.3381
7	36.2496	1.5593
8	22.0311	1.3430
9	21.9249	1.3409
10	22.0313	1.3430
11	22.0314	1.3430
12	16.5103	1.2178
13	22.0316	1.3430
14	360.696	2.5571

4 decision variables for cycle 2:

Kx 4: Log mean is 2.568202.

Kx 8 : Log mean is 0.9522.

Kx 11 : There are two values that have cov below cutoff. The WONN matrix for Kx11 is

Source of cov	S4	S6
cov vector	1.2905	3.8507
normalized cov	.3351	1.
(0.39-log sd) vector	-.04	-.3317
normalized sd diff	.1206	1
WONN- wt of 2 on sd difference	.4129	2.2361
Comment	accept	reject

Log mean of decision variable for Kx 11 is 1.1338.

Kx 13 : log mean is 1.0669.

Zone 10 could have been included in this cycle, but it caused scenario 5 to terminate in the sffp by the occurrence of dry cells. Thus zone 10 is left to the end to be optimized by itself.

Sequence 1 order is again used in the sffp, and the total SSlof is plotted in figure 5.2. Parameters Kx 4, 8, 11 and 13 that produced the lowest SSlof in step 4 of sequence 1 replace the existing starting values of those parameters in table 5.5, and form the starting values for cycle 3.

5.4 Asp simulation, first pass, cycle 3

The starting values of cycle 3 is shown in table 5.6 below. The 4 parameters that become the decision variables are those remaining ones with the most reasonable cov and log sd in tables 5.2 and 5.3. The exception is Kx 10, which creates a problem of dry cell for scenario 5, will be optimized by itself.

Kx 1 : log mean =3.1461.

kx 2 : log mean = -.1568.

Kx 3 : log mean= 1.4821.

Kx 5 : log mean=0.0826.

Figure 5.2

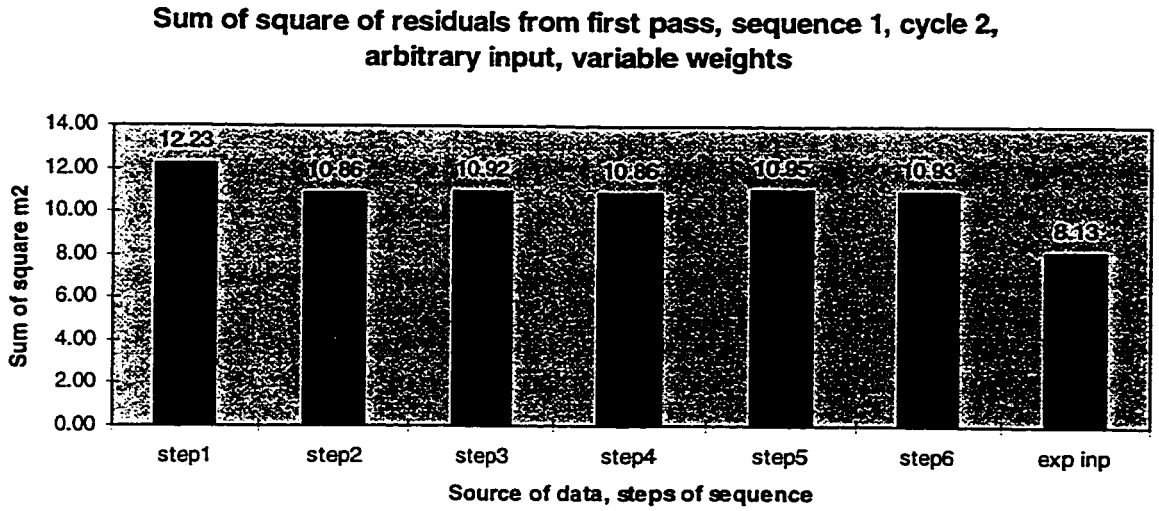
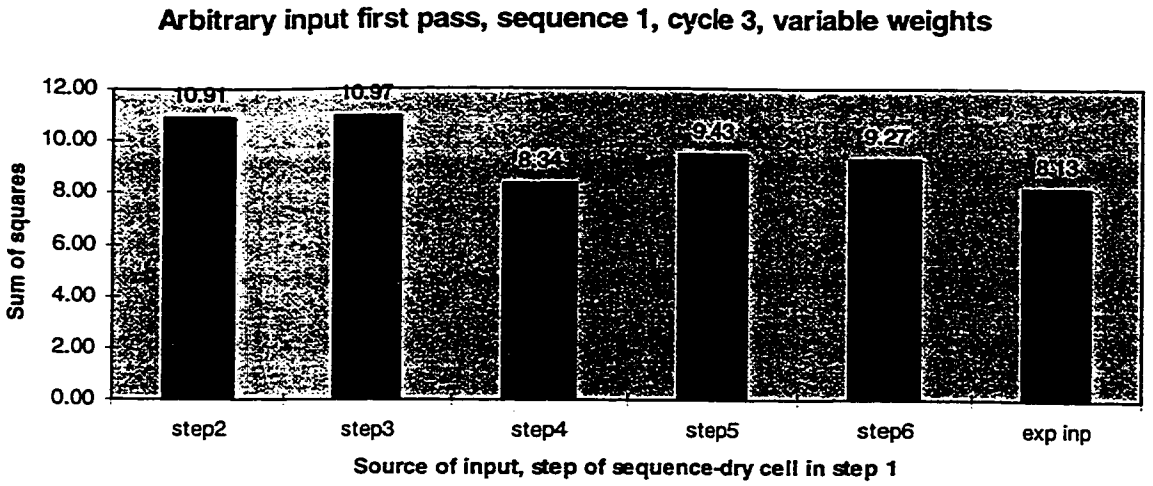


Table 5.6

First pass, cycle 3 starting values

Kx	Values	Log values
1	22.0304	1.3430
2	22.0305	1.3430
3	22.0306	1.3430
4	23.2605	1.3666
5	22.0308	1.3430
6	21.7833	1.3381
7	36.2496	1.5593
8	18.9886	1.2785
9	21.9249	1.3409
10	22.0313	1.3430
11	17.811	1.2507
12	16.5103	1.2178
13	27.0172	1.4316
14	360.696	2.5571

Figure 5.3



The sffp sequence 1 order is used to find the lowest SSlof, and the results are plotted in figure 5.3. The parameter values of step 4 of cycle 3 for kx 1, 2, 3, 5 replace the existing default values in table 5.6.

5.5 Asp simulation, first pass, cycle 4

The starting values for cycle 4 is in table 5.7.

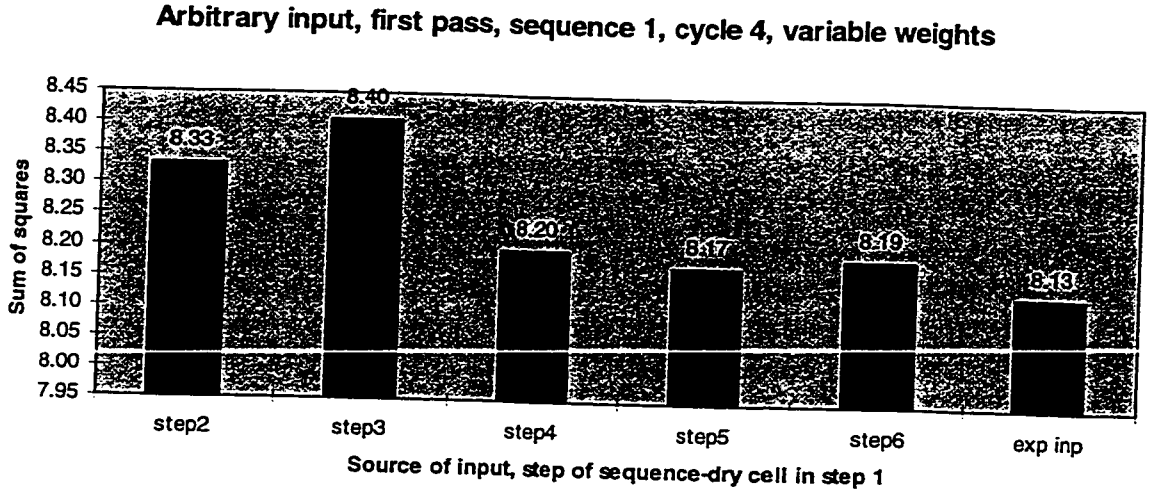
Table 5.7 starting values for cycle 4

First pass, cycle 4 starting values

Kx	Values	Log values
1	15.0233	1.1768
2	1.93603	0.2869
3	22.0822	1.3440
4	23.2605	1.3666
5	14.1031	1.1493
6	21.7833	1.3381
7	36.2496	1.5593
8	18.9886	1.2785
9	21.9249	1.3409
10	22.0313	1.3430
11	17.811	1.2507
12	16.5103	1.2178
13	27.0172	1.4316
14	360.696	2.5571

The decision is Kx 10 by itself, and its log mean is 1.4672. The sffp sequence 1 order is again used, but dry cell occurred in step 2 (scenario 5). The sequence is re-started again using an unconstrained optimization for scenario 5 at that point, and the sequence is resumed after that point. The overall sums of square of lack of fit for the steps are in figure 5.4.

Figure 5.4 Asp simulation, first pass, cycle 4



The value of Kx 10 from step 5 is selected to complete the set of 14 zones transmissivity values. As can be seen, the asp simulation results in an overall sum of square of lack of fit approaching that of the expert input, at 8.17 m^2. The 14 zones of optimized values that produce this result are in table 5.8. The comparison of these results to the expert inputs are shown in table 5.12 in the next subsection.

5.6 Asp simulation, second pass, cycle 1

In this second pass of asp optimization, information from the first asp pass is used to maximize the effect of the choice of decision variables for each cycle, and to attempt to lower the sum of square of lack of fit further. The prior information is not used as weights, but as guidelines to select decision variables and starting values for

First pass, cycle 4 end values

Kx	Values	Log values
1	15.023	1.1768
2	1.936	0.2869
3	22.082	1.3440
4	23.261	1.3666
5	14.103	1.1493
6	21.783	1.3381
7	36.25	1.5593
8	18.989	1.2785
9	21.925	1.3409
10	13.744	1.1381
11	17.811	1.2507
12	16.51	1.2177
13	27.017	1.4316
14	360.696	2.5571

Table 5.8 End values of first pass, cycle 4 of asp simulation

optimizations in the cycles of the second pass. A distinctive feature of the previous asp optimization is that cycles 1 and 2 of the first pass used the most identifiable parameters of the 14 zones based on the criterion of modified cov with variance control. For these two cycles, the overall sums of squares are higher than those for the expert input. The third cycle, with the relatively not so identifiable parameters, brought the overall sum of squares much closer to the expert input standard.

Based on this observation, a parameter that affected the first pass, cycle 3 optimization (Kx 2) will be utilized in the following way: The unconstrained optimization value for Kx2 will be used as starting value for the new cycle 1 optimization - a value of log mean of Kx2 is -0.1568. This starting value of Kx2 will be used until it is optimized as a decision variable. Then its optimized value will replace the starting value. The 5 decision variables will remain the same as those in the first pass. Similarly, the starting values for these five parameters (Kx 6,7,9,12,14) will be the same as those in cycle 1 of the first pass.

A parameter by parameter discussion can be found in section 5.2 for zones other than Kx2. The starting values and decision variables are summarized:

Second pass, cycle 1 starting values

Kx	Values	Log values
1	22.0304	1.3430
2	0.697	-0.1568
3	22.0306	1.3430
4	22.0307	1.3430
5	22.0308	1.3430
6	20.2791	1.3070
7	24.1245	1.3825
8	22.0311	1.3430
9	20.1583	1.3045
10	22.0313	1.3430
11	22.0314	1.3430
12	26.4914	1.4231
13	22.0316	1.3430
14	385.914	2.5865

Table 5.9 Starting values for second pass, cycle 1

the decision variables are the same as those in section 5.2, with Kx 6, 7, 9, 12 and 14.

The sequence 1 order is used with the starting values in table 5.9, and the 5 decision variables. The total SSlof for each set of input is plotted against the expert input total SSlof for comparison, and is shown in figure 5.5.

The parameter set that produced the lowest SSlof (step 4) is used to replace the values of Kx 6, 7, 9, 12, 14 for cycle 2. The starting values of cycle 2 will utilize the optimization results of second pass, cycle 1.

5.7 Asp simulation, second pass, cycle 2

The starting values of cycle 2 is in table 5.10 below. The selection of the decision variables are based on observation of results from the first pass, cycle 3, together with the values of the cov. Kx1, 2, 4, 8, 13 are selected as decision variables.

Figure5.5

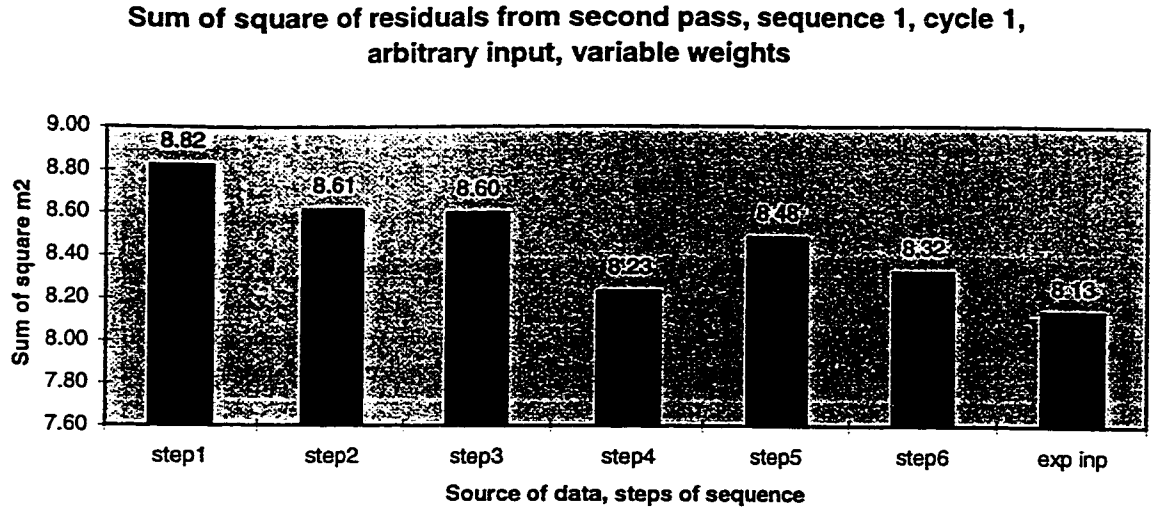


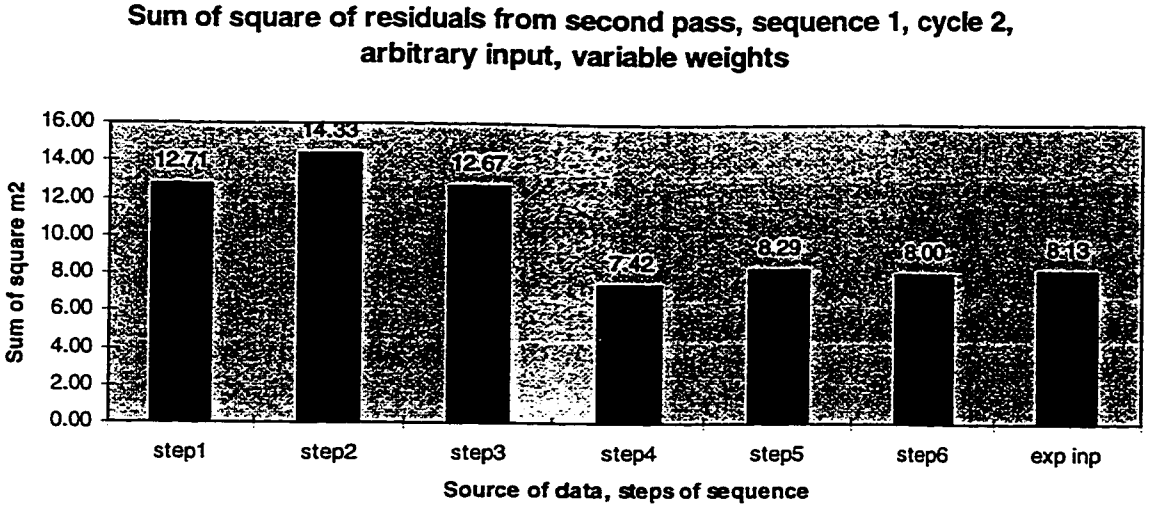
Table 5.10

Second pass, cycle 2 starting values

Kx	Values	Log values
1	22.0304	1.3430
2	0.697	-0.1568
3	22.0306	1.3430
4	22.0307	1.3430
5	22.0308	1.3430
6	23.8273	1.3771
7	39.3631	1.5951
8	22.0311	1.3430
9	17.6245	1.2461
10	22.0313	1.3430
11	22.0314	1.3430
12	22.028	1.3430
13	22.0316	1.3430
14	362.766	2.5596

Sequence 1 order is again used in the sffp, and the total SSlof is plotted in figure 5.6. Parameters Kx 1, 2, 4, 8 and 13 that produced the lowest SSlof in step 4 of sequence 1 replace the existing starting values of those parameters in table 5.10, and form the starting values for cycle 3.

Figure 5.6



5.8 Asp simulation, second pass, cycle 3

The starting values of cycle 3 is shown in table 5.11 below. The remaining 4 parameters become the decision variables, the parameters with the most reasonable cov and log sd are chosen as prior information values.

The sffp sequence 1 order is used to find the lowest SSlof, and the results are plotted in figure 5.7. The parameter values of step 5 of cycle 3 for kx 3, 5, 10, 11 replace those in table 5.11 as the complete set of optimized variables for this second pass of the asp simulation. The complete set of variables has a lower SSlof for all six scenarios than the expert input.

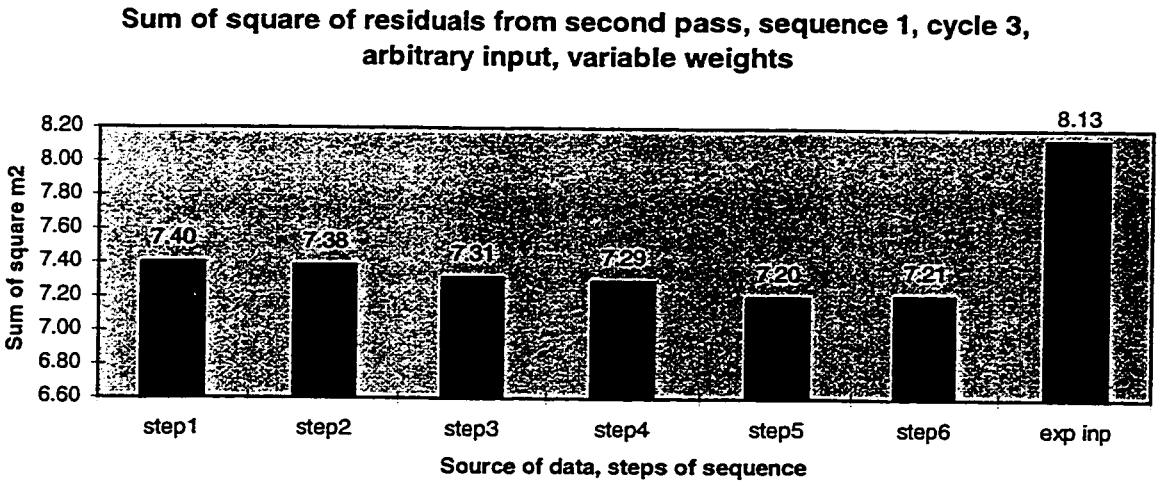
It is also worth noting that the second pass, with its usage of the prior information of the first pass to group the decision variables for the optimization cycles and to choose the starting values, is successful in controlling the dry cell occurrence in the first pass. Only 3 cycles are required in the second pass as compared to 4 in the first.

Second pass, cycle 3 starting values

Kx	values	log values
1	13.796	1.1398
2	0.38472	-0.4149
3	22.0306	1.3430
4	35.2644	1.5473
5	22.0308	1.3430
6	23.8273	1.3771
7	39.3631	1.5951
8	41.1752	1.6146
9	17.6245	1.2461
10	22.0313	1.3430
11	22.0314	1.3430
12	22.028	1.3430
13	15.4587	1.1892
14	362.766	2.5596

Table 5.11

Figure 5.7



Some of the asp results are almost identical to the expert input (eg, Kx2, 7, 8, 14), whereas others are different. Referring to the second pass and expert input results, with the exception of Kx 1 and 10, which have a difference of over two orders of magnitude, the differences among the three sets of transmissivities in general are considered not too

large, considering the natural variability of the transmissivities. For the log transformed values the differences are even smaller.

The arbitrary starting point simulation with the sequential process can enable a set of optimized transmissivities to be found. This set of transmissivities can perform better in reducing the SSlof than the expert input, and has statistical estimates of uncertainties and (linearized) confidence limits associated with each value.

Table 5.12 complete set of second pass asp optimized values for six scenarios, comparison to expert input and first pass results

Second pass, cycle 3 end values, compare to first pass and expert input

Kx	First pass values	Second pass values	exp inp
1	15.0233	13.796	0.1296
2	1.93603	0.385	0.648
3	22.0822	33.265	10.368
4	23.2605	35.264	12.96
5	14.1031	62.572	15.552
6	21.7833	23.827	31.04
7	36.2496	39.363	38.88
8	18.9886	41.175	46.656
9	21.9249	17.625	51.84
10	13.7437	0.251	54.432
11	17.811	19.507	62.208
12	16.5103	22.028	90.72
13	27.0172	15.459	103.68
14	360.696	362.766	388.8

6.0 Analyses, conclusion and future research

The results from the last sections are highlighted here and further analysed. This section begins with a discussion on the noise problem which may hamper efforts in inverse modelling without the benefits of prior information.

6.1 Eigenvalues ratio analysis, noise contamination

From a set of scenario optimization results, which have been obtained under unconstrained and constrained (with prior information) conditions, eigenvalues from the covariance matrices are obtained and listed in table 6.1.

Table 6.1 Eigenvalue analysis

Case run	Largest eigenvalue	Smallest eigenvalue	Square root of ratio
Unconstrained (No Prior Information)	512.8	3.8135e-5	3.667e3
Constrained (With Prior Information)	2.2759e-2	2.2223e-5	3.20e1

The ratios are calculated in accordance with the discussion in section 2.3. It can be seen that the unconstrained case is close to having a potential noise problem due to ill conditioning. In contrast, the problem with prior information, using unit weights, is much better conditioned. An ill-conditioned matrix will take much longer to solve. The time required to find a solution in an unconstrained optimization case is usually three or more times longer than for a constrained case. When the matrices are nearly singular, PEST may fail to find a solution after days of computing.

Equation 2.33 can also be used to examine the weakness of the representation of the parameters in the original reference system (a skewed reference system, which has axes that are not perpendicular to each other). The parameters in the orthogonal reference system (spanned by the eigenvectors) can be expressed as a linear combination of the skewed reference system parameter vector in equation 2.33. If the elements in the column

of the U matrix corresponding to the smallest eigenvalue are examined, the components that made up the orthogonal matrix U in the skewed reference system is weak (particularly in the case where noise is a problem) for that eigenvalue.

The PEST manual recommends the examination of the eigenvector corresponding to the largest eigenvalue. If the vector has more than one dominant element, then these parameters are not identifiable, because a strong correlation exists among them.

The ratios of the largest to the smallest eigenvalues in table 6.1 can also serve to indicate the effectiveness of the weighting scheme. As mentioned in the last paragraph of section 2, the variable weighting (approach II), is numerically similar to classical methods of scaling and preconditioning the construction of an approximation to the Hessian. In other words, the classical scaling is performed to improve the eigenvalue structure of the construction process of the inverse Hessian.

Another way to show that there is numerical similarities between the new sequential process and the classical method is to examine the eigenvalue ratios of the resulting normal matrices for the steps of a sequential process. If the variable weighting is functioning in a similar fashion to the classical methods, then the eigenvalue ratios are expected to be lower than those from the steps of the constant weighting schemes (approach I) of the same sequence.

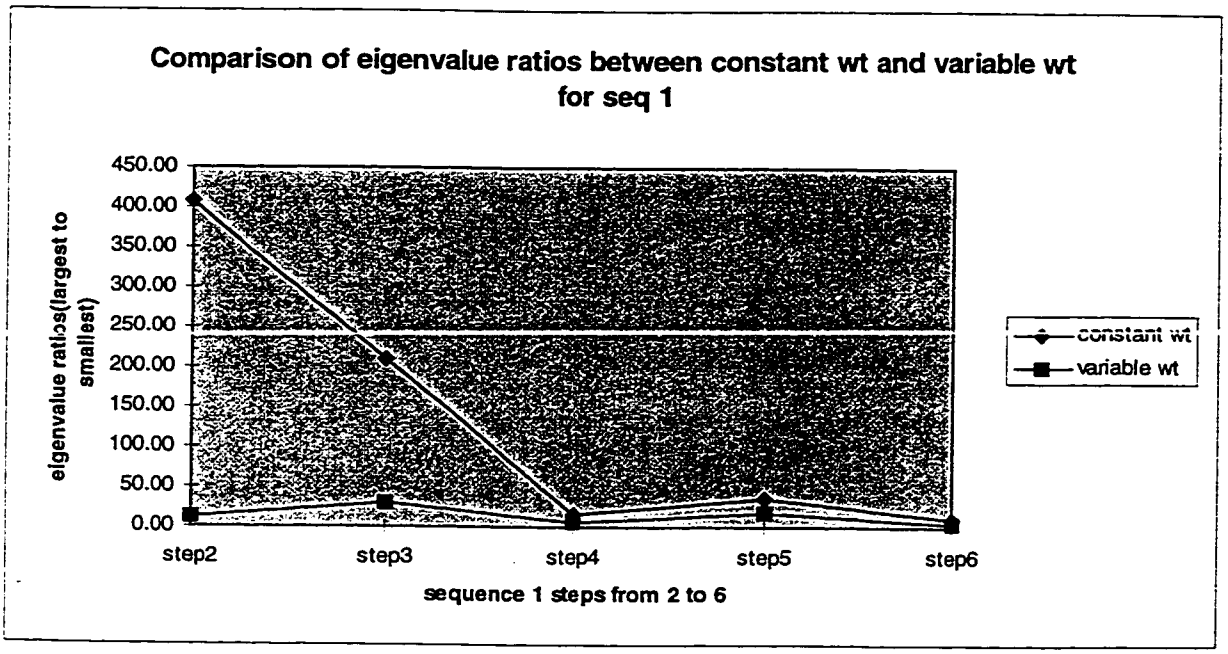
The eigenvalue ratios for the constant and variable weights cases for the steps in sequence 1 are calculated and shown below in table 6.2. As can be seen, the ratios are everywhere lower in the variable weights case, up to over 30 times lower in instances than those obtained by using the constant weights. The ratios are plotted in figure 6.0 for steps 2 to 6.

As was expected, the variable weights do produce better conditioned matrices than constant weights. The proper scaling of the modelling and measurement error variances by the use of variable weights derived from the previous step of the sffp does share numerical similarities to the classical construction methods for the approximation to the inverse of the Hessian.

Table 6.2 Calculation of eigenvalue ratios for constant and variable weights

	constant wt (approach I)			variable wt (approach II)		
steps in seq	largest eigv L	smallest eigv S	ratio L / S	largest eigv L	smallest eigv S	ratio L / S
step1	33.2	2.03E-03	16382.12	34	2.60E-03	13072.40
step2	0.4956	1.21E-03	408.27	1.64E-02	1.29E-03	12.72
step3	9.56E-02	4.57E-04	209.20	1.02E-02	3.52E-04	29.09
step4	9.15E-02	5.78E-03	15.83	1.89E-02	2.87E-03	6.57
step5	6.00E-02	1.62E-03	36.94	3.32E-02	1.88E-03	17.66
step6	0.1615	1.55E-02	10.43	6.83E-02	1.49E-02	4.60

Figure 6.0 eigenvalue ratios for steps 2 to 6 for sequence 1



6.2 The weights as penalty parameters in the prior information

The changes of the value of Ω_p^2 for sequence 1 and 3 in the 14 zone sequential feed forward procedure (section 4) are shown below in table 6.3, 6.4 respectively. Figure 6.1 shows the evolution of the Kx 6 weight ratio for sequence 1 which is typical of all other decision variables. The maximum ratio is at step 4, which means that the parameter has the lowest uncertainty at that step. The lowest SS_{lof} also occurs at that step, therefore the overall minimization of the parameter uncertainty and sum of square of lack

of fit both occur at the same point in the sffp. This is in contrast to Yeh (1986) (see also section 1.3).

Table 6.3 sffp sequence 1, 14 Kx zones - parameter weight Ωp^2

The prior information weight Wp^2 input to steps, seq 1, 14 zones model

Kx	step 1	step 2	step 3	step 4	step 5	step 6
6	33.99	55.25	398.76	739.86	333.10	127.47
7	7.10	21.60	95.10	174.67	117.73	41.00
9	11.65	30.30	92.21	201.62	105.37	50.07
12	4.90	22.95	65.10	110.48	53.26	43.48
14	6.64	26.43	71.35	203.29	94.06	36.69
obs wt- Ws^2	25.00	22.22	36.34	44.15	186.69	49.56
Wp/Ws Kx6	1.17	1.58	3.31	4.09	1.34	1.60
Wp/Ws Kx7	0.53	0.99	1.62	1.99	0.79	0.91
Wp/Ws Kx9	0.68	1.17	1.59	2.14	0.75	1.01
Wp/Ws Kx12	0.44	1.02	1.34	1.58	0.53	0.94
Wp/Ws Kx14	0.52	1.09	1.40	2.15	0.71	0.86

Note: observational weights are the inverse of the reference variance

Figure 6.1 Weight ratio Ωp / Ωs for sequence 1, Kx14

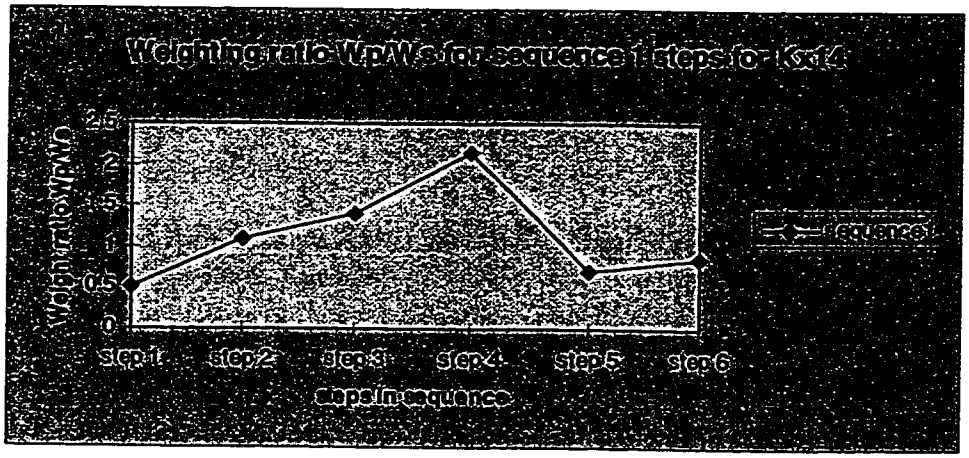
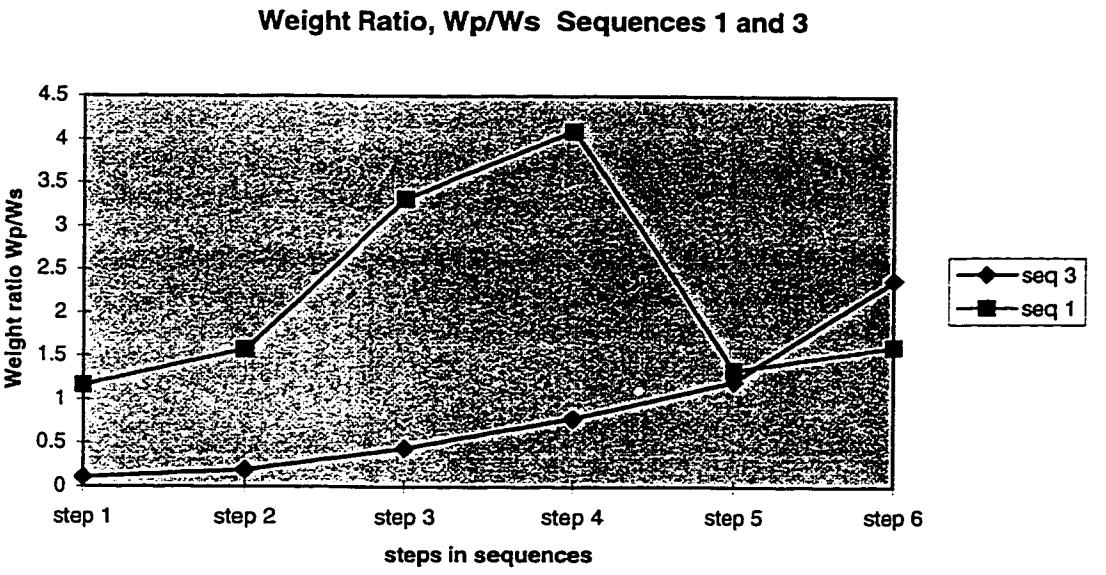


Table 6.4 sffp sequence 3, 14 Kx zones - parameter weight Ωp^2

The penalty parameter c (or Wp^2) input to sequence 3, 14 zones model

Kx	step 1	step 2	step 3	step 4	step 5	step 6
6	0.35	1.58	6.76	37.30	34.87	172.76
7	0.65	2.93	12.54	30.75	24.38	50.19
9	0.42	1.85	6.99	30.12	27.09	43.04
12	0.36	1.56	3.54	18.96	19.13	27.86
14	1.20	5.20	8.54	24.50	22.70	31.69
obs wt Ws^2	32.48	44.09	35.10	60.19	24.01	30.72
kx6Wp/Ws=	0.10	0.19	0.44	0.79	1.21	2.37
Kx7Wp/Ws=	0.14	0.26	0.60	0.71	1.01	1.28
Kx9Wp/Ws=	0.11	0.21	0.45	0.71	1.06	1.18
Kx12Wp/Ws=	0.10	0.19	0.32	0.56	0.89	0.95
Kx14Wp/Ws=	0.19	0.34	0.49	0.64	0.97	1.02

Figure 6.2 Comparison of weight ratios, Ω_p / Ω_s of sequences 1 and 3, Kx 6



It can be seen from the above tables 6.3 and 6.4, and figure 6.2, that the evolution of Ω_p and Ω_s are different for the two sequences. Sequence 1 shows a high Ωp^2 at the beginning, some 100 times the starting Ωp^2 values of sequence 3. Sequence 1 has a narrower confidence band than sequence 3 and it also outperforms sequence 3. This indicates a better conditioned matrix in the starting file of sequence 1. It illustrates the

importance of starting the sequence with a file having the magnitudes of the sample and parameter weights close to each other: They are both present as the sample weight and the inverse of the parameter prior information covariance in the equation for the overall covariance matrix for the parameters (see equation (2.30)).

6.3 Assessing the use of the modified coefficient of variation with variance control as a parameter selection criterion

As indicated in section 5.5, the first pass, cycles 1 and 2 of the asp simulation, based entirely on the modified cov cutoff, did not produce a lower sum of squares of lack of fit than the expert input. This suggests that the most identifiable parameters may not be effective in reducing the lack of fit errors. On closer examination, however, the higher sums of squares errors are due to scenario 1. Figures 6.3, 6.4 for cycles 1 and 2 show the contributions of each scenario to the overall sums of squares.

Figure 6.3

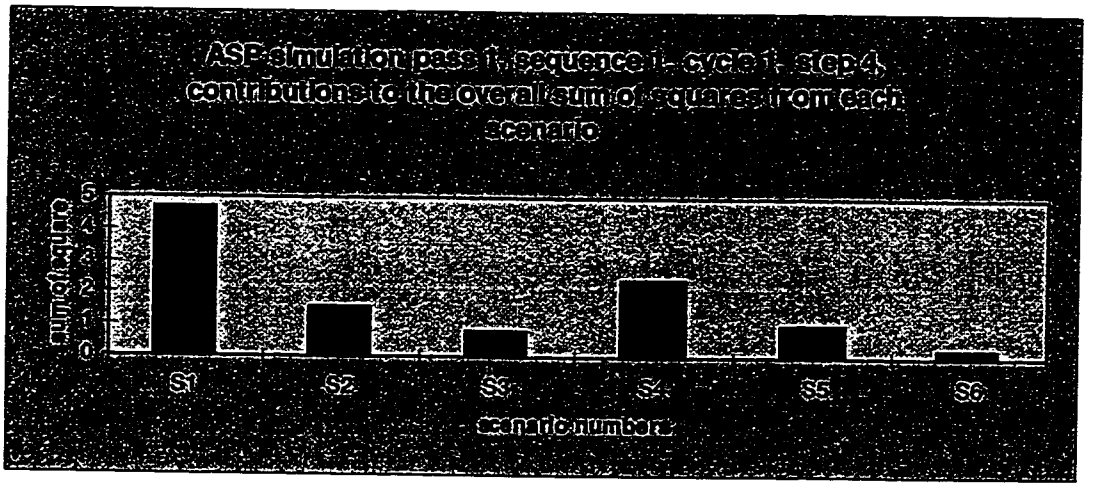
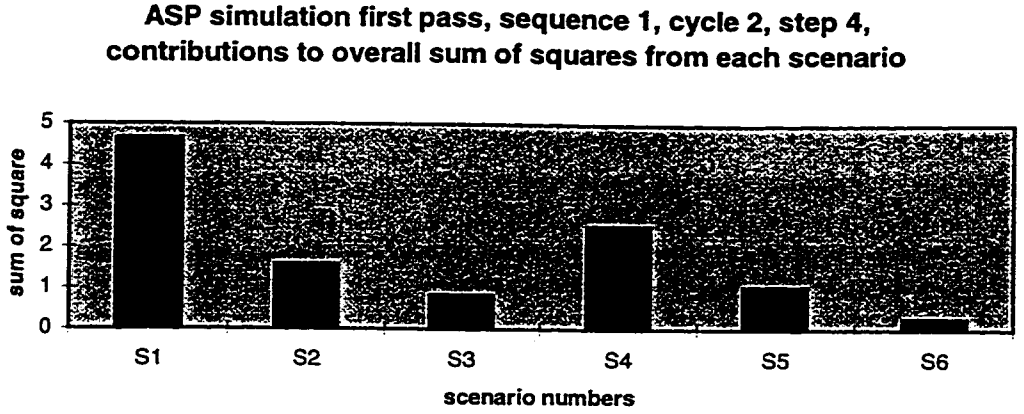


Figure 6.4



It can be seen that scenario 1 has the largest sum of squares contribution for both cycles. By contrast, cycle 3 of the first pass shows a decline of the contribution to the sum of squares from scenario 1 of 2.398 m^2, almost half of the values of cycles 1 and 2.

This is not surprising, since both scenarios 1 and 2 have been excluded from the unconstrained optimization for the calculation of cov and standard deviation at the beginning of the asp process. Data in both scenarios 1 and 2 are deemed to be less accurate than the rest of the suite of six scenarios, therefore they have not been used in the screening process.

However, the same scenarios are part of the overall sum of squares calculations. The net effect is that the most identifiable parameters based on the last four scenarios may differ from the most identifiable of the first two. Had all six scenarios been used in the unconstrained optimization and the calculation of the cov, the need for the second pass of the asp simulations may diminish.

Thus, the modified cov is effective as a means of parameter selection based on parameter identifiability.

6.4 Conclusion and future research areas

The automated search for the optimal set of hydraulic transmissivities for the former Canada Creosote site requires the understanding, analyses and appropriate manipulation of the information in six steady state scenarios of ground water flow. The information contained within the six scenarios of the site has been used to obtain the optimal site transmissivities by means of indirect inverse modelling, through the use of the program systems PEST and MODFLOW, and by a newly implemented method of sequential feed forward procedure (sffp).

An innovative method in variance control transferred from industrial optimization has been adopted for use in the sffp. The control is based on the scaling of the common variance of the measurement and modelling error, by using similar quantities from the previous step of the sffp, as variable weights. The empirical tests conducted with the new variance control show lower overall sum of square of the lack of fit for the six scenarios throughout the steps of the sffp, as compared to the results from the use of a constant common variance method. The variable weighting also produce better conditioned matrices when eigenvalue ratios are examined. This is an expected result, as classical theories in the construction of the approximation of the inverse of the Hessian predict better conditioned matrices due to proper scaling.

The sffp is applied to a model with 24 transmissivity zones, and to the another model of the same spatial region with 14 zones. The latter is obtained from linking zones of the former which had identical transmissivities after expert calibration. Decision variables are selected for optimization by considering the identifiabiliy, uniqueness and stability issues of the parameters through screening processes. The sffp produces optimized sets of parameters from both models which minimizes objective functions of the models. The optimization process produces parameter sets in modelling groundwater flow at the Canada Creosote site which show

i) The overall sum of square of the lack of fit,

ii) The uncertainty of the optimized parameters, and

iii) The conditioning of the matrices produced during the optimization.

The above are all better when their magnitudes are lower.

A flexible, general method of weighted optimized non-dimensional norm (wonn) is introduced to combine measures of the "goodness" of the solution in order to compare the results of the different methods. The 14 zone model with optimized parameters is assessed to be a better model for the site based on the wonn results. The optimized parameter set is thus obtained for overall good results for all six scenarios.

An arbitrary starting point (asp) simulation for the 14 zone model was conducted. The starting values of all the zones are set to the geometric mean of the expert calibration values, which represents a reasonable starting point for the simulations. By unconstrained optimization of scenarios S3 to S6, and by successive application of the sffp in three cycles using a new algorithm developed for the asp, transmissivity values close to the expert input are found. These asp generated transmissivities perform better than the expert input in terms of overall sums of squares. This algorithm will be very useful in automated determination of optimal transmissivities, with a minimum amount of input from experts.

This study presents many innovative methods to assess the optimal parameter set for steady state inverse modelling with more than one data scenario. Future research in this area should concentrate on speeding up the sffp sequence order assessment and streamlining the sequential process. Variance control, the general measure of wonn and the asp algorithm are potentially very powerful tools for the groundwater modelling field. A more rigorous theoretical study should be initiated to explore the exciting new ideas, and to expand them to include automatic zoning for transmissivities.

References:

Ang, A. and W. Tang, 1975, *Probability Concepts in Engineering Planning and Design*, John Wiley and Sons, Vol 1, Basic Principles.

Ang, A. and W. Tang, 1990, *Probability Concepts in Engineering Planning and Design*, Published by Ang and Tang, Vol 2, Decision, Risk and Reliability.

Carrera, J. , 1988, "State of the art of the inverse problem applied to the flow and solute transport equations", Ground Water Flow and Quality Modeling, 549-583.

Carrera, J. and S.P. Neuman, 1986a, "Estimation of aquifer parameters under transient and steady state conditions, 1, Maximum Likelihood Method, Incorporating Prior Information", Water Resources Research, v22, No.2, 199-210.

Carrera, J. and S.P. Neuman, 1986b, "Estimation of aquifer parameters under transient and steady state conditions, 2, Uniqueness, Stability and Solution Algorithms", Water Resources Research, v22, No.2, 211-227.

Carrera, J. and S.P. Neuman, 1986c, "Estimation of aquifer parameters under transient and steady state conditions, 3, Application to Synthetic and Field Data", Water Resources Research, v22, No.2, 228-242.

CH2M Gore & Storrie Limited, 1997, A Groundwater Model to Support Management of the Canada Creosote Site, Calgary, Alberta, Update Report.

CH2M Gore & Storrie Limited, 1996, *Aquifer Analyses Canada Creosote Site, Report to Alberta Environment Protection.*

Chauvent, G.,1991, "On the Theory and Practice of Non-linear Least Squares", Advance Water Resources, v14, No 2, 55-103.

Cooley, R.L., 1977, "A method of estimating parameters and assessing reliability for models of steady state ground water flow", theory and numerical properties, number 1, Water Resources Research, v 13, No2, 318-324.

Cooley, R.l., R.L. Naff, 1990, "Regression Modelling of Ground-water Flow", US Department of the Interior, Techniques of Water-Resources Investigation of the US Geological Survey, book 3.

Cushman, J.H., 1986, " On measurement, scale and scaling ", Water Resources Research, v22, No 2, 129-134.

Deming, S.N. and S.L. Morgan, 1987, *Experimental Design: A Chemometric Approach*, Elsevier.

Gill, P.E., W. Murray and M.H. Wright, 1981, *Practical Optimization*, Academic Press, eighth edition.

Ginn, T.R. and J.H. Cushman, 1990, "Inverse models for subsurface flow: a critical review of stochastic techniques", Stochastic Hydrology and Hydraulics, v4, 1-26.

Golder Associates Ltd., 1990, *Final report to Alberta Environment HELP Project on Remediation Strategies for Canada Creosote Site, Calgary, Alberta*, Report number 892-2803X.

Lanczos, C., 1964, *Applied Analysis*, Prentice Hall, Englewood Cliffs, N.J.,third edition.

Luenberger, D.G., 1984, *Linear and Nonlinear Programming*, Addison-Wesley, second edition

Neuman, S.P. and S. Yakowitz, 1979, "A statistical approach to the inverse problem of aquifer hydrology", Water Resources Research, v15, number 4, 85-860.

Roy, R., 1990, *A Primer on the Taguchi Method*, Van Nostrand Reinhold.

Rummel, R.J., 1970, *Applied Factor Analysis*, Northwestern University Press

Shooman, M.L., 1968, *Probabilistic Reliability : An Engineering Approach,* McGraw-Hill Book Co., New York, NY.

Speed, D.E. and D.P Ahlfeld, 1996, "Diagnosis of structural identifiability in groundwater flow and solute transport equations", Calibration and Reliability in Groundwater Modelling, Proceedings of Modelcare 96 Conference, IAHS, No 237, 287-298.

Taylor, W., 1991, *Optimization and Variation Reduction in Quality*, McGraw Hill.

Theil, H., 1963, " On the use of Incomplete Prior Information in Regression Analysis" American Statistical Association Journal v58, no. 302 401-414.

Weiss, R. and L. Smith, 1998a, "Efficient and Responsible use of prior information in inverse models", Ground Water, 36, 151-163.

Weiss, R. and L. Smith, 1998b, "Parameter space methods in joint parameter estimation for groundwater flow models", Water Resources Research, v34, 4, 647-661.

Yeh, W.W.G., 1986, "Review of parameter Identification Procedures in Ground water Hydrology: The Inverse Problem", Water Resources Research, v22, 95-108.

Appendix A

This appendix contains the results of the Monte Carlo simulations for transmissivity zones 7, 9, 12, 14 of the 14 zone model. The mean and median of the results are used in the calculations

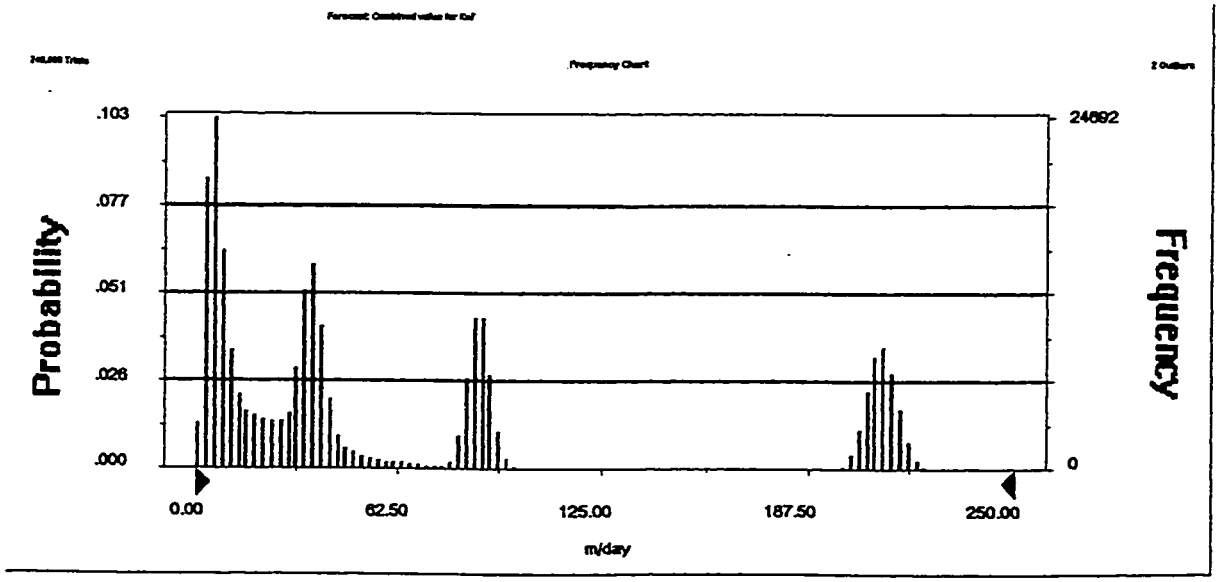
Crystal Ball Report
Simulation started on 12/6/97 at 19:43:18
Simulation stopped on 12/6/97 at 20:05:51

Forecast: Combined value for Kx7

Summary:

Display Range is from 0.00 to 250.00 m/day
Entire Range is from 0.44 to 289.78 m/day
After 240,000 Trials, the Std. Error of the Mean is 0.15

Statistics:	Value
Trials	240000
Mean	64.12
Median	35.15
Mode	---
Standard Deviation	71.30
Variance	5,083.40
Skewness	1.26
Kurtosis	3.16
Coeff. of Variability	1.11
Range Minimum	0.44
Range Maximum	289.78
Range Width	289.34
Mean Std. Error	0.15

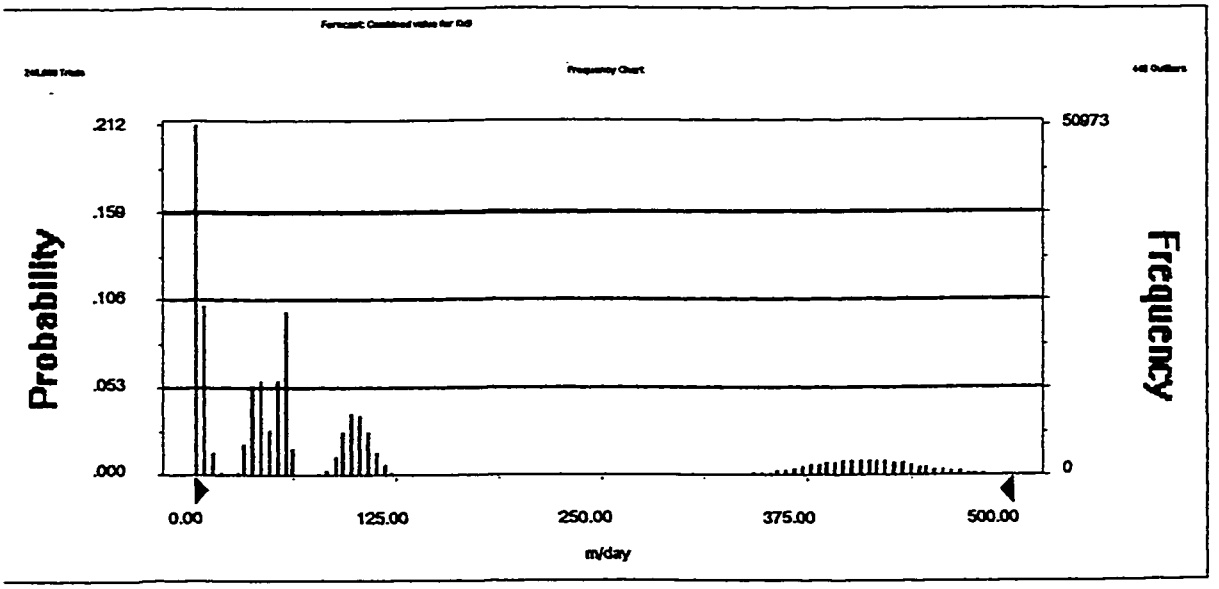


Crystal Ball Report
Simulation started on 12/6/97 at 20:10:03
Simulation stopped on 12/6/97 at 20:32:12

Forecast: Combined value for Kx9

Summary:
Display Range is from 0.00 to 500.00 m/day
Entire Range is from 0.00 to 1,484.23 m/day
After 240,000 Trials, the Std. Error of the Mean is 0.29

Statistics:	Value
Trials	240000
Mean	103.11
Median	51.45
Mode	---
Standard Deviation	143.23
Variance	20,515.69
Skewness	1.62
Kurtosis	4.05
Coeff. of Variability	1.39
Range Minimum	0.00
Range Maximum	1,484.23
Range Width	1,484.23
Mean Std. Error	0.29



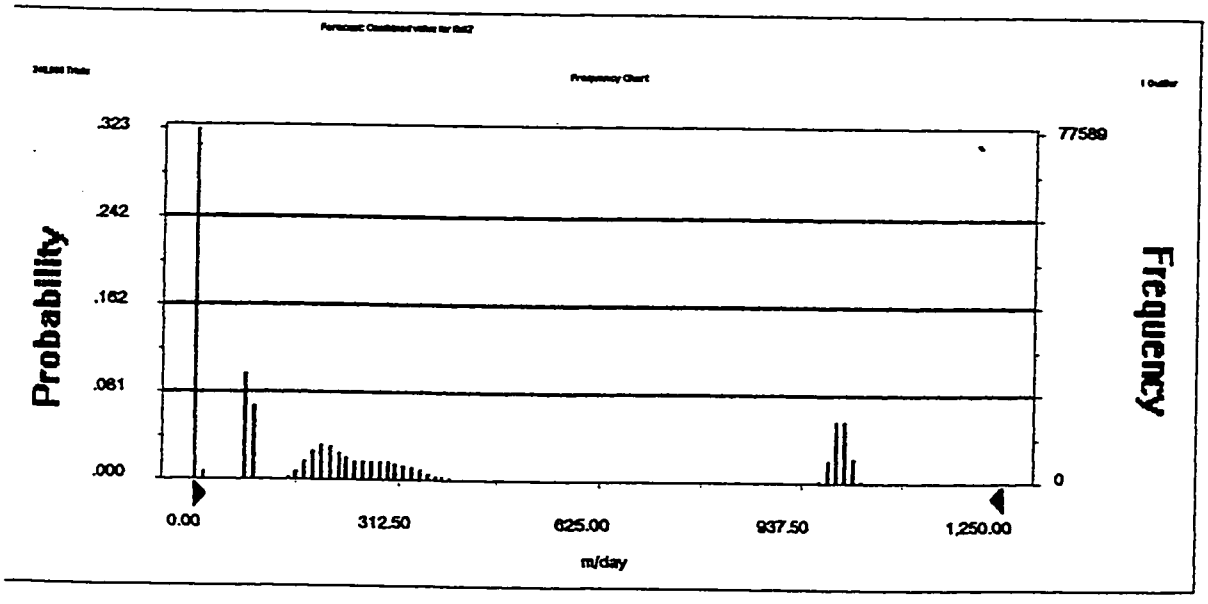
Crystal Ball Report
Simulation started on 12/6/97 at 21:14:43
Simulation stopped on 12/6/97 at 21:36:56

Forecast: Combined value for Kx12

Summary:

Display Range is from 0.00 to 1,250.00 m/day
Entire Range is from 0.00 to 1,695.67 m/day
After 240,000 Trials, the Std. Error of the Mean is 0.71

Statistics:	Value
Trials	240000
Mean	265.82
Median	138.46
Mode	---
Standard Deviation	346.05
Variance	119,749.38
Skewness	1.43
Kurtosis	3.51
Coeff. of Variability	1.30
Range Minimum	0.00
Range Maximum	1,695.67
Range Width	1,695.67
Mean Std. Error	0.71



Crystal Ball Report
Simulation started on 12/6/97 at 20:49:12
Simulation stopped on 12/6/97 at 21:11:33

Forecast: Combined value for Kx14

Summary:
Display Range is from 0.00 to 1,100.00 m/day
Entire Range is from 1.53 to 1,035.81 m/day
After 240,000 Trials, the Std. Error of the Mean is 0.86

Statistics:	Value
Trials	240000
Mean	565.09
Median	911.74
Mode	---
Standard Deviation	422.39
Variance	178,411.45
Skewness	-0.18
Kurtosis	1.22
Coeff. of Variability	0.75
Range Minimum	1.53
Range Maximum	1,035.81
Range Width	1,034.28
Mean Std. Error	0.86