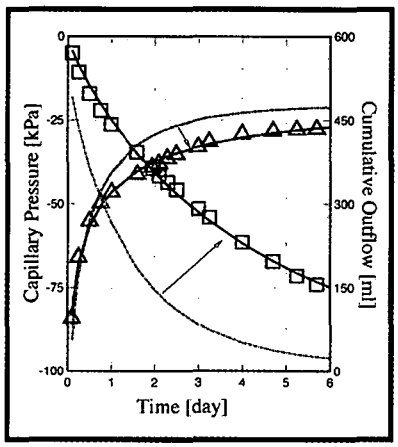
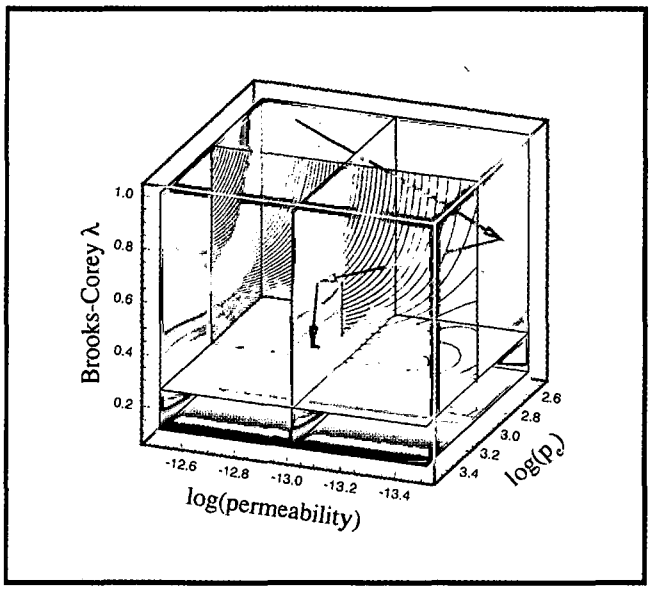
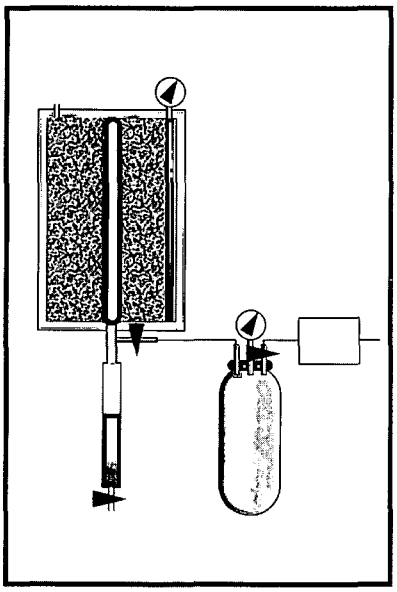
LBNL-40042
UC-800

ITOUGH2
Sample Problems







S. Finsterle

Earth Sciences Division
Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720

November 1997

This work was supported, in part, by the Director, Office of Civilian Radioactive Waste Management, Yucca Mountain Site Characerization Project Office, Regulatory and Site Evaluation Division, by the WIPP Project, Sandia National Laboratories, under Document Number 129847, and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Geothermal Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

REFERENCE COPY
Does Not
Circulate
Bldg. 50 Library - Ref.
Lawrence Berkeley National Laboratory
LBNL-40042
Copy 1

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

LBNL-40042
UC-800

ITOUGH2
Sample Problems

Stefan Finsterle

Earth Sciences Division
Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720

November 1997

This work was supported, in part, by the Director, Office of Civilian Radioactive Waste Management, U.S. Department of Energy, through Memorandum Purchase Order EA9013MC5X between TRW Environmental Safety Systems, Inc. and the Ernest Orlando Lawrence Berkeley National Laboratory, by a grant from the Swiss National Cooperative for the Disposal of Radioactive Waste (Nagra), Wettingen, Switzerland, and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Geothermal Technologies, of the U.S. Department of Energy, under Contract No. DE-AC03-76SF00098.

TABLE OF CONTENTS

1. INTRODUCTION 1

2. PROBLEM 1: SENSITIVITY ANALYSIS, PARAMETER ESTIMATION, AND UNCERTAINTY ANALYSIS 3
   2.1 Introduction 3
   2.2 Solving the forward problem with ITOUGH2 5
   2.3 Sensitivity analysis 8
   2.4 Parameter estimation 11
   2.5 Uncertainty propagation analysis 16

3. PROBLEM 2: ANALYSIS OF A LABORATORY EXPERIMENT WITH KLINKENBERG EFFECT 21
   3.1 Modeling the gas-pressure-pulse-decay experiment 21
   3.2 An ill-posed inverse problem 24
   3.3 A well-posed inverse problem 32
   3.4 Parameterizing systematic errors 35

4. PROBLEM 3: CALIBRATION OF GEOTHERMAL RESERVOIR MODEL 39
   4.1 Problem statement 39
   4.2 Automatic parameter selection 40
   4.3 Calibration results and total sensitivity 43

5. PROBLEM 4: MULTISTEP DESATURATION EXPERIMENT 46
   5.1 Introduction 46
   5.2 Specifying step changes in boundary conditions 48
   5.3 Minimization algorithms 49
   5.4 Model identification and prior information 51

6. PROBLEM 5: COMBINED INVERSION OF STEADY-STATE AND TRANSIENT DATA 54
   6.1 Introduction 54
   6.2 The forward problem 55
   6.3 The inverse problem 57
   6.4 Sensitivity measures and residual statistics 60

7. PROBLEM 6: VENTILATION EXPERIMENT 64
7.1 Introduction 64
7.2 Discussion of selected modeling issues 65
7.3 Minimization, residual and error analysis 66

8. PROBLEM 7: NUMERICAL DISPERSION 76
8.1 Capillary diffusion 64
8.2 Compensation for numerical dispersion 65

ACKNOWLEDGMENT 83
REFERENCES 84
INDEX 86

LIST OF FIGURES

Figure 2.1.1. Schematic of synthetic laboratory experiment. 3
Figure 2.1.2. TOUGH2 input file *sam1*. 4
Figure 2.2.1. ITOUGH2 input file *sam1p1i* used to solve direct problem. 5
Figure 2.2.2. ITOUGH2 input file *sam1p2i* used to solve forward problem and to generate column file *sam1p2i.col* with flow rates and pressures as a function of time. 7
Figure 2.2.3. Excerpt from ITOUGH2 plot file *sam1p2i.col*. Time is in the first column, followed by the simulated state variables for each data set. This file will be used as the data file in Part 4. 7
Figure 2.3.1. ITOUGH2 input file *sam1p3i* used to perform sensitivity analysis of 5 observations with respect to 3 parameters. 9
Figure 2.3.2. Excerpt from ITOUGH2 output file *sam1p3i.out*, showing global sensitivity measures. 10
Figure 2.4.1. ITOUGH2 input file *sam1p4i* used to perform parameter estimation by inverse modeling. 12
Figure 2.4.2. Pressure transient at center of column calculated with initial parameter set and after optimization. 15
Figure 2.4.3. Flow rates at inlet calculated with initial parameter set and after optimization. 15
Figure 2.5.1. ITOUGH2 input file *sam1p5i* used to examine prediction uncertainty using first-order-second-moment (FOSM) error propagation analysis. 17
Figure 2.5.2. Excerpt from ITOUGH2 input file *sam1p6i* used to examine prediction uncertainty using Monte Carlo simulations. 18
Figure 2.5.3. Excerpt from ITOUGH2 output file *sam1p6i.out*, showing the distribution of the uncertain input parameters. 19
Figure 2.5.4. Comparison between FOSM and Monte Carlo error propagation analysis. 20
Figure 3.1.1. Schematic of gas-pressure-pulse-decay apparatus. 21
Figure 3.1.2. TOUGH2 input file *sam2*. 23
Figure 3.2.1. ITOUGH2 input file *sam2p1i*. Includes input Parts 2 and 3. 25
Figure 3.2.2. Comparison between measured and calculated pressures. 29
Figure 3.2.3. Sensitivity of pressure in upstream and downstream reservoir with respect to permeability, Klinkenberg factor, and porosity. 29

Figure 3.3.1. Comparison between measured and calculated pressure transient curves from three simultaneously inverted gas-pressure-pulse-decay experiments. 34
Figure 3.3.2. Residuals as a function of time, showing systematic overprediction of pressures at late times. 34
Figure 3.4.1. Comparison between measured and calculated pressure transient curves from three simultaneously inverted gas-pressure-pulse-decay experiments. 36
Figure 3.4.2. Residuals as a function of time. Residuals exhibit a random structure. 36
Figure 4.1.1. Five-spot well pattern with grid for modeling 1/8 of symmetric domain. 39
Figure 4.2.1. Excerpt from file *sam3p1i* showing block for automatic parameter selection. 42
Figure 4.2.2. Excerpt from file *sam3p1i.out* showing selection criteria 42
Figure 4.3.1. Calibration and prediction of pressures, temperatures, water and vapor flow rates. 44
Figure 5.1.1. Schematic of apparatus for radial flow experiment to determine unsaturated hydraulic properties. 46
Figure 5.2.1. Excerpt from ITOUGH2 input file *sam4i*, showing block `TIMES` with `RESTART` option for changing saturation boundary condition. . 48
Figure 5.3.1. Excerpt from ITOUGH2 input file *sam4i*, showing block `OPTIONS` with different minimization algorithms. 49
Figure 5.3.2. Contours of the objective function in the three parameter planes (a) $\log(k_{abs})$-$\log(p_e)$, (b) $\log(k_{abs})$-λ, and (c) λ-$\log(p_e)$. The solution path taken by the Levenberg-Marquardt minimization algorithm is shown as a bold line. 51
Figure 5.4.1. Comparison between observed (symbols) and simulated (lines) system response for the multi-step radial flow experiment. The prescribed suction pressure at the extraction cylinder is shown as a dashed line. 52
Figure 5.4.2. Relative permeability function derived by inverse modeling for the Brooks-Corey and the van Genuchten model. 52
Figure 6.1.1. Combining a steady-state run with subsequent transient simulation. . . 54
Figure 6.2.1. Excerpt from subroutine USERBC for assigning atmospheric pressure fluctuations as user-specified boundary conditions. 56

Figure 6.3.1. Excerpt from ITOUGH2 input file *sam5i* showing blocks with pneumatic pressure, saturation, and water potential data. 58
Figure 6.3.2. Excerpt from ITOUGH2 input file *sam5i* showing block for the estimation of a constant added to the pneumatic pressure data 59
Figure 6.3.3. Excerpt from ITOUGH2 input file showing block `COMPUTATION`. . 60
Figure 6.4.1. Match of pneumatic pressures at three elevations in one borehole. . . . 61
Figure 6.4.2. Excerpt from ITOUGH2 output file *sam5i.out* with measures of total sensitivity. 61
Figure 6.4.3. Excerpt from ITOUGH2 output file *sam5i.out* with residual statistics and relative contribution to objective function for each observation type. 63
Figure 7.1.1. Schematic of model domain and instrumentation 64
Figure 7.2.1. Excerpt from ITOUGH2 input file *sam6i*, showing block `OBSERVATION` for specifying an average evaporation rate. 66
Figure 7.3.1. Comparison between computed and measured water potentials. 68
Figure 7.3.2. Excerpt from ITOUGH2 output file *sam6i.out.* Residual plot. 68
Figure 7.3.3. Excerpt from ITOUGH2 output file *sam6i.out.* Summary of residual analysis. 69
Figure 7.3.4. Excerpt from ITOUGH2 output file *sam6i.out.* with linear regression analysis of plot of the calculated versus observed system response. . . 70
Figure 7.3.5. Excerpt from ITOUGH2 input file *sam6i*, showing block `COMPUTATION`. 71
Figure 7.3.6. Excerpt from ITOUGH2 output file showing correction procedure along first eigenvector of parameter covariance matrix. 75
Figure 8.1.1. TOUGH2 input file *sam7*. 78
Figure 8.1.2. Verification of TOUGH2 solution. Only minor discretization effects are seen for grid spacing $h = 0.01$ m, maximum time step $\Delta t_{max} = 100$ seconds, and $A = 1000$ Pa. 78
Figure 8.2.1. Subroutine USERDATA. Liquid saturation is calculated analytically as a function of space and time, and provided as the data points to be matched. 80
Figure 8.2.2. ITOUGH2 input file *sam7i*. 81
Figure 8.2.3. Discretization effects are compensated for by reducing parameter A from 100 Pa to 48 Pa. 82

LIST OF TABLES

Table 1.1.	Summary of Sample Problems	1
Table 1.2.	Summary of Issues Addressed by Sample Problems	2
Table 3.1.1.	Overview of Inverse Modeling Runs	22
Table 3.2.1.	Summary of Inverse Modeling Results: Initial Guess, Best Estimate, Standard Deviation, and Ratio of Conditional and Joint Standard Deviation, Part 1	31
Table 3.2.2.	Estimation Covariance Matrix, Part 1	31
Table 3.2.3.	Eigenvectors and Eigenvalues of Estimation Covariance Matrix	31
Table 3.3.1.	Summary of Inverse Modeling Results: Initial Guess, Best Estimate, Standard Deviation, and Ratio of Conditional and Joint Standard Deviation, Part 2	32
Table 3.3.2.	Estimation Covariance Matrix, Part 2	33
Table 3.4.1.	Summary of Inverse Modeling Results: Initial Guess, Best Estimate, Standard Deviation, and Ratio of Conditional and Joint Standard Deviation, Part 3	37
Table 3.4.2.	Matrix of Direct (Lower Triangle) and Overall (Upper Triangle) Parameter Correlations, Part 3	37
Table 3.4.3.	Goodness-of-Fit, Residual Statistics, and Model Identification Criteria	37
Table 4.1.1.	Observations Used for Model Calibration	40
Table 4.1.2.	True, Initial, and Estimated Parameter Set	40
Table 4.2.1.	Estimation Error, Correlation, and Sensitivity	40
Table 4.3.1.	Total Sensitivity of Observations, Standard Deviation of Residuals, and Contribution to Objective Function (COF)	45
Table 5.3.1.	Comparison of Minimization Algorithm Efficiency	50
Table 5.4.1.	Inverse Modeling Results for the Brooks-Corey and van Genuchten Model	51
Table 7.3.1.	Estimates Obtained Starting From Five Different Initial Parameter Sets	67
Table 7.3.2.	Covariance (Diagonal and Lower Triangle) and Correlation (Upper Triangle) Matrices From Linear Error Analysis	72
Table 7.3.3.	Joint and Conditional Standard Deviations	72
Table 7.3.4.	Matrix of Direct (Lower Triangle) and Total (Upper Triangle) Correlation Coefficients	72
Table 7.3.5.	Eigenanalysis of Covariance Matrix	73

1. INTRODUCTION

This report contains a collection of ITOUGH2 sample problems. It complements the "ITOUGH2 User's Guide" [*Finsterle*, 1997a], and the "ITOUGH2 Command Reference" [*Finsterle*, 1997b].

ITOUGH2 is a program for parameter estimation, sensitivity analysis, and uncertainty propagation analysis. It is based on the TOUGH2 simulator for non-isothermal multiphase flow in fractured and porous media [*Pruess*, 1987, 1991a]. The report "ITOUGH2 User's Guide" [*Finsterle*, 1997a] describes the inverse modeling framework and provides the theoretical background. The report "ITOUGH2 Command Reference" [*Finsterle*, 1997b] contains the syntax of all ITOUGH2 commands. This report describes a variety of sample problems solved by ITOUGH2.

Table 1.1 contains a short description of the seven sample problems discussed in this report. The TOUGH2 equation-of-state (EOS) module that needs to be linked to ITOUGH2 is also indicated. Each sample problem focuses on a few selected issues shown in Table 1.2. ITOUGH2 input features and the usage of program options are described. Furthermore, interpretations of selected inverse modeling results are given. Problem 1 is a multipart tutorial, describing basic ITOUGH2 input files for the main ITOUGH2 application modes; no interpretation of results is given. Problem 2 focuses on non-uniqueness, residual analysis, and correlation structure. Problem 3 illustrates a variety of parameter and observation types, and describes parameter selection strategies. Problem 4 compares the performance of minimization algorithms and discusses model identification. Problem 5 explains how to set up a combined inversion of steady-state and transient data. Problem 6 provides a detailed residual and error analysis. Finally, Problem 7 illustrates how the estimation of model-related parameters may help compensate for errors in that model.

Table 1.1. Summary of Sample Problems

#	EOS	Description
1	3	Tutorial; introduction to main ITOUGH2 applications
2	3	Analysis of data from laboratory gas-pressure-pulse-decay experiment; parameter correlation, non-uniqueness, systematic errors
3	1	Calibration of geothermal reservoir model; automatic parameter selection
4	9	Analysis of multi-step radial desaturation experiment; step changes in boundary conditions; minimization algorithm
5	3	Pneumatic pressures; user-specified boundary conditions; matching transient data after steady-state equilibrium calculation
6	4	Ventilation experiment; error analysis
7	9	Examines numerical diffusion by comparison to analytical solution

Table 1.2. Summary of Issues Addressed by Sample Problems

#	Part	CPU time @	Application/ITOUGH2 features
1	1	4 sec	Solves forward problem using ITOUGH2
	2	6 sec	Generates time series plots
	3	15 sec	Sensitivity analysis
	4	1 min	Parameter estimation
	5	20 sec	FOSM uncertainty propagation analysis
	6	6 min	Monte Carlo simulations
2	1	2 min	Non-uniqueness
	2	2 min	Resolve non-uniqueness by adding observations
	3	4 min	Parameterization of systematic errors
3	1	15 min	Automatic parameter selection
	2	5 min	FOSM uncertainty propagation analysis
4	1	40 sec	Gauss-Newton
	2	1 min	Levenberg-Marquardt
	3	5 min	Simplex
	4	1 hour	Simulated Annealing
	5	20 min	Grid Search
5	-	45 min	Matching transient data after steady-state simulation
6	-	20 min	Inversion; check of linearity assumption
7	-	4 min	User-specified data function

@ Approximate CPU time on Sun Ultra 1, Model 170, 167 MHz Ultra SPARC Processor

All input and data files needed to run the sample problems are part of the ITOUGH2 distribution package. The TOUGH2 and ITOUGH2 input file names are *sam#* and *sam#i*, respectively, where # stands for the problem number. Multipart problems have file names of the form *sam#p%i*, where # is the problem number and % is the part number. Data files are identified by a "*.dat*" file extension.

The key to a successful application of ITOUGH2 is (1) a good understanding of multiphase flow processes, (2) the ability to conceptualize the given flow and transport problem and to develop a corresponding TOUGH2 model, (3) detailed knowledge about the data used for calibration, (4) an understanding of parameter estimation theory and the correct interpretation of inverse modeling results, and (5) proficiency in using ITOUGH2 options. This report addresses issues (4) and (5).

2. PROBLEM 1: SENSITIVITY ANALYSIS, PARAMETER ESTIMATION, AND UNCERTAINTY ANALYSIS

2.1 Introduction

The purpose of Problem 1 is to develop a sequence of simulations that demonstrate the three key applications of ITOUGH2, namely (1) sensitivity analysis for experimental design, (2) parameter estimation by data inversion, and (3) uncertainty propagation analysis. We focus here on a step-by-step description of the development of ITOUGH2 input files rather than on the interpretation of the inverse modeling results.

A synthetic laboratory experiment is chosen for simplicity. A schematic of the experimental layout is shown in Figure 2.1.1. Water is injected at constant pressure into a one-dimensional, horizontal column filled with uniform, partially saturated sand. This setup is similar to the one used for a standard steady-state Darcy experiment. However, there is a certain amount of free gas initially present in the column, and information about the transient behavior of pressures and flow rates are used to determine two-phase flow parameters.

The corresponding TOUGH2 input file is shown in Figure 2.1.2.

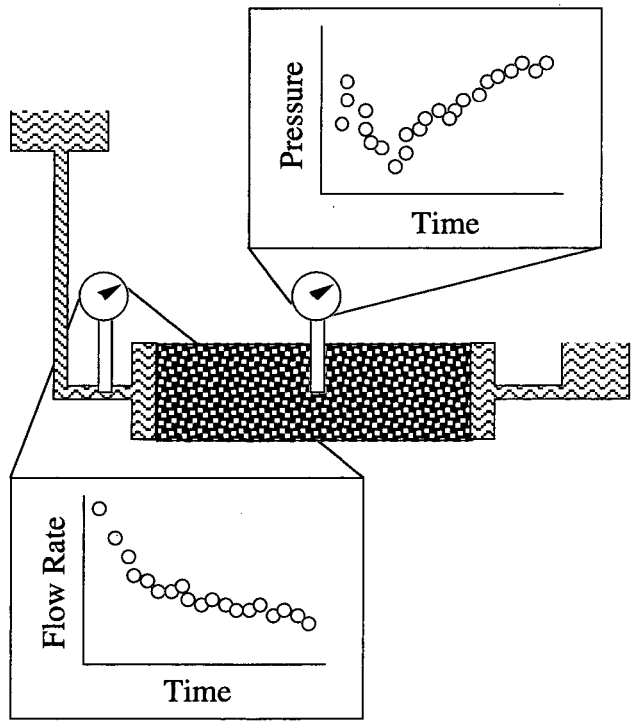


Figure 2.1.1. Schematic of synthetic laboratory experiment.

```
sam1: TOUGH2 input file for simulating two-phase, transient "Darcy" experiment
ROCKS----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
SAND     1     2650.     .3500 2.000E-12                          2.51      920.
 1.000E-08
BOUND    0     2650.     .9900 2.000E-12                          2.51   100000.

RPCAP----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
    3      0.200E+00 0.050E+00
    1      0.000E+00 0.000E+00 1.000E+00
PARAM----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
   2 500    999910000010000000400003000
 0.000E+00 6.000E+02 2.500E+00 1.000E+01

 100000.000000000000 10.30000000000000  20.000000000000000
MULTI----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
    2    2    2    6
ELEME----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
IN             BOUND .1000E+50                     .0000E+00
A11 1   49    1SAND  .1000E-03                     .1000E-01
OUT            BOUND .1000E+50                     .5000E+00

CONNE----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
IN   A11 1                   1 .1000E-10 .5000E-02 .1000E-01
A11 1A11 2   48    1    1    1 .5000E-02 .5000E-02 .1000E-01
A1150OUT                     1 .5000E-02 .1000E-10 .1000E-01

START----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
INCON----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
IN
 110000.000000000000 0.00000000000000000  20.000000000000000
OUT
 100000.000000000000 0.00000000000000000  20.000000000000000

ENDCY----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
```

Figure 2.1.2. TOUGH2 input file *sam1*.

We assume that the objective of the experiment is to estimate the permeability and the porosity of the sand as well as the initial gas saturation. Furthermore, we presume that only one flow meter and one pressure transducer are available for data collection. The measurement uncertainties of the two instruments are 5 ml/min and 200 Pa, respectively.

Step-by-step instructions will be given for this sample problem. In the first two parts (Section 2.2), we use ITOUGH2 to simply solve the forward problem, producing a plot file, and at the same time generating the synthetic database for the subsequent inversions. In Part 3 (Section 2.3), a sensitivity analysis is performed to determine the optimum location of the flow meter and pressure transducer. This demonstrates how ITOUGH2 can be used for experimental design. The estimation of the three parameters of interest based on synthetic flow rate and pressure data is discussed in Part 4 (Section 2.4). Finally in Section 2.5, we compare the uncertainty of the model predictions by using either linear uncertainty propagation analysis (Part 5) and Monte Carlo simulations (Part 6).

2.2 Solving the Forward Problem with ITOUGH2

It is very important for any inverse modeling effort to solve the forward problem in a stable, robust, and efficient manner. A standard TOUGH2 forward run should always be performed to check the appropriateness of the simulation before more time-consuming inversions are initiated. Plotting the simulation results obtained with the initial parameter set against the data is also strongly recommended because frequent errors such as wrong conversion factors, opposite signs of measured and calculated flow rates, time lags and data shifts, etc. can readily be detected. Furthermore, it provides a first assessment of the model, and allows one to estimate the CPU-time required to solve the inverse problem.

A single TOUGH2 simulation can be performed by typing the following command:

```
tough2 sam1 3 &
```

This command makes use of the UNIX script file *tough2*, which automatically takes file *~/itough2/invdir* as a dummy ITOUGH2 input file. If using this approach, potential error messages are printed to file *t2.msg*. An alternative way is to write a short ITOUGH2 input file (*sam1p1i*) as shown in Figure 2.2.1.

```
> COMPUTATION
  >> OPTION
     >>> solve FORWARD problem only
     <<<
  <<
<
```

Figure 2.2.1. ITOUGH2 input file *sam1p1i* used to solve direct problem.

The following command must be used to solve all parts of Problem 1:

```
itough2 sam1p#i sam1 3 &
```

where "#" is the part number.

Since no parameters or observations are provided, the ITOUGH2 output file *sam1p1i.out* contains warning messages which can be ignored in this specific case. In general, however, one should always consult the following files to check for potential errors:

(1) the ITOUGH2 output file (e.g., *sam1p1i.out*)
(2) the TOUGH2 output file (e.g., *sam1.out*)
(3) the ITOUGH2 message file (e.g., *sam1p1i.msg*)

In Part 2, we use ITOUGH2 to generate a time plot of pressure and flow rates at two selected points within the column. The plotfile will later be used as the synthetic database for the inversion.

ITOUGH2 provides convenient options in block `> OBSERVATIONS` to pick points in space and time at which the value of selected TOUGH2 output variables can be examined. Here, we are interested in the pressures [Pa] in the center of the column, and the flow rates [ml/min] at the inlet. Sixty equally spaced points in time between 10 and 600 seconds are selected to generate the time series. Columns of time versus calculated flow rates and pressures are written to file *sam1p2i.col*, which can be processed by most visualization packages for plotting. The corresponding ITOUGH2 input file is shown in Figure 2.2.2.

The pressure is observed at the center of the column, identified by the corresponding gridblock name (note that blanks in a gridblock name must be replaced by underscores). For flow rates, the two gridblock names defining the connection at the inlet are specified. Since no measured data are available, command `>>>> NO DATA` is added, which automatically generates dummy data points of value 10^{-50}. A multiplication factor of -1.6667E-5 is specified, converting positive flow rates in units of [ml/min] to the flow rates calculated by TOUGH2, which are negative and in [kg/sec]. Note that the sign of the calculated flow rates is the result of a convention, i.e., it is arbitrarily defined by the ordering of elements in a connection. This convention has to be accounted for by choosing the correct sign of the conversion factor.

By default, the format of the plotfile concurs with the TECPLOT visualization software. In order to arrange the plotfile in columns, command `>>> FORMAT: COLUMN` is used. The first column in file *sam1p2i.col* (see Figure 2.2.3) holds the selected times in the chosen time units; the second column contains the TOUGH2 output for the first data set (e.g., pressures at element A1125); the third column contains the calculated system response for the second data set (e.g., liquid flow rate across interface IN__0 A11_1), and so on. If measurements are available, additional columns will be generated with the observed data. (File *sam1p2i.col* will be used in Part 4 as the file from which the synthetic data are extracted.)

This simulation produces a relatively large TOUGH2 output file *sam1.out* even for this small problem, because full results are printed at 60 printout times. Specifying a negative number for parameter *KDATA* in TOUGH2 block PARAM.1 suppresses the printout of state variables in the TOUGH2 output file, saving disk space and making the run slightly faster.

```
> OBSERVATION

  >> select : 60 points in TIME, EQUALLY spaced between
     10.0 600.0 seconds

  >> PRESSURE
     >>> ELEMENT          : A1125
         >>>> NO DATA needed
         <<<<
     <<<

  >> LIQUID FLOW RATE
     >>> CONNECTION       : IN__0  A11_1
         >>>> FACTOR      :-1.666667E-05 (ml/min - kg/sec)
         >>>> NO DATA needed
         <<<<
      <<<
  <<

> COMPUTATION

  >> OPTION
     >>> solve FORWARD problem only
     <<<

  >> OUTPUT
     >>> LIST all available plot file FORMATS and select: COLUMNS
     >>> output in MINUTES
     <<<
  <<
<
```

Figure 2.2.2. ITOUGH2 input file *sam1p2i*. Solves forward problem and generates column file *sam1p2i.col* with flow rates and pressures as a function of time.

```
      TIME            SIM0            SIM0
      [min]        P(GAS) A1125  F-L IN  0 A11 1
  .16666667E+00    .10302017E+06    .85780174E+02
  .33333333E+00    .10351308E+06    .65323231E+02
  .50000000E+00    .10353113E+06    .57637722E+02
  .66666667E+00    .10340909E+06    .53653780E+02
  .83333333E+00    .10325474E+06    .50982470E+02
  .10000000E+01    .10310265E+06    .48876966E+02
  .11666667E+01    .10296287E+06    .47075537E+02
  .............    .............    .............
  .10000000E+02    .10491346E+06    .21109160E+02
```

Figure 2.2.3. Excerpt from ITOUGH2 plot file *sam1p2i.col*. Time is in the first column, followed by the simulated state variables for each data set. This file will be used as the data file in Part 4.

2.3 Sensitivity Analysis

Part 3 demonstrates how ITOUGH2 is used to perform a sensitivity analysis. By calculating global sensitivity measures for each parameter and each observation type, alternative experimental configurations can be tested against each other, and the design most suitable for estimating parameters in a subsequent inversion can be identified. As in any sensitivity analyses involving non-linear processes, prior knowledge about the parameters must be available, or the analysis has to be repeated for different parameter combinations. Furthermore, the parameters of interest and their potential variations must be specified, as well as the type, location, and expected uncertainty of the measurements. Since the experiment has to be optimized with respect to the estimation of permeability, porosity, and initial gas saturation, these three parameters and their respective variations are defined in the `> PARAMETER` block of the ITOUGH2 input file (Figure 2.3.1).

Since the permeability of the column and the adjacent boundary gridblocks are considered a single parameter to be varied, both material names, i.e., SAND and BOUND, are specified on the same line. Note that the initial gas saturation is referred to as primary variable no. 2 of the TOUGH2 block PARAM.4, which holds the default initial conditions. Potential parameter variations are specified, scaling the sensitivity coefficients of the Jacobian matrix. For permeability, the variation of the logarithm is given rather than permeability itself. No parameter annotation is provided for porosity, prompting ITOUGH2 to automatically generate an identifier.

In the `> OBSERVATION` block, pressures in elements A1112, A1125, and A1138 are examined. The element names correspond to three potential measuring locations within the column, at a distance of 1/4, 1/2, and 3/4 of the total column length from the inlet, respectively. Furthermore, the connections defining the inlet and the outlet are given as the two alternative points for flow rate measurements.

The program option `>>> SENSITIVITY ANALYSIS` is invoked, which makes ITOUGH2 run $n+1=4$ TOUGH2 simulations to calculate the Jacobian matrix, where n is the number of parameters. Note that option `>>> FORWARD` used in file *sam1p2i* (Figure 2.2.2) must be disabled. This can be done by (1) deleting the corresponding line, (2) replacing the command level indicator "`>>>`" by blanks (or any other character), (3) surrounding the line with "`/*`" and "`*/`" comment characters, or (4) adding a comment character ("#") in the first column.

The Jacobian matrix and the covariance matrix of the parameters provide the basis for a detailed sensitivity analysis. We notice from the global sensitivity measures shown in Figure 2.3.2 that the highest sensitivity is realized for the pressure measurements in the center of the column, and that the flow rate data at the inlet contain significantly more information than the flow rate data at the outlet, suggesting implementation of the experimental configuration shown in Figure 2.1.1.

```
> PARAMETER

  >> ABSOLUTE permeability
     >>> MATERIAL         : SAND_ BOUND
         >>>> ANNOTATION : log(abs. perm.)
         >>>> LOGARITHM
         >>>> VARIATION  : 0.3
         <<<<
      <<<

  >> POROSITY
     >>> MATERIAL         : SAND_
         >>>> VALUE
         >>>> VARIATION  : 0.10
         <<<<
     <<<

  >> INITIAL condition for primary variable No.: 2
     >>> DEFAULT
         >>>> ANNOTATION : Gas entrapped
         >>>> VALUE
         >>>> VARIATION  : 0.10
         <<<<
     <<<
  <<

> OBSERVATION

  >> select : 60 points in TIME, EQUALLY spaced between
     10.0 600.0 seconds

  >> PRESSURE

     >>> ELEMENT          : A1112
         >>>> ANNOTATION : Pressure 1/4
         >>>> NO DATA available
         >>>> DEVIATION  : 200.0 Pa  (expected measurement error)
         <<<<

     >>> ELEMENT          : A1125
         >>>> ANNOTATION : Pressure 1/2
         >>>> NO DATA available
         >>>> DEVIATION  : 200.0 Pa
         <<<<

     >>> ELEMENT          : A1138
         >>>> ANNOTATION : Pressure 3/4
         >>>> NO DATA available
         >>>> DEVIATION  : 200.0 Pa
         <<<<
     <<<
```

Figure 2.3.1. ITOUGH2 input file *sam1p3i* used to perform sensitivity analysis of 5 observations with respect to 3 parameters.

```
  >> LIQUID FLOW RATE
     >>> CONNECTION       : IN__0  A11_1
         >>>> ANNOTATION : Flow inlet
         >>>> FACTOR      :-1.666667E-05 (ml/min - kg/sec)
         >>>> NO DATA available
         >>>> DEVIATION  : 5.0 ml/min  (expected measurement error)
         <<<<
     >>> CONNECTION       : A1150  OUT_0
         >>>> ANNOTATION : Flow outlet
         >>>> FACTOR      :-1.666667E-05 (ml/min - kg/sec)
         >>>> NO DATA available
         >>>> DEVIATION  : 5.0 ml/min
         <<<<
     <<<
  <<


> COMPUTATION
  >> OPTION
     ... solve FORWARD problem only
     >>> perform SENSITIVITY analysis
     <<<

  >> OUTPUT
     >>> PLOTFILE contains : COLUMNS
     <<<
  <<
<
```

Figure 2.3.1 (cont.). ITOUGH2 input file *sam1p3i* used to perform sensitivity analysis of 5 observations with respect to 3 parameters.

		log(abs. perm.)	POROSITY SAND	Gas entrapped	Total	
Total from data	Pressure 1/4	274.9	113.3	262.4	650.5	
Total from data	Pressure 1/2	440.1	180.9	370.1	991.0	←
Total from data	Pressure 3/4	289.3	116.9	255.8	662.1	
Total from data	Flow inlet	168.2	41.5	119.6	329.4	←
Total from data	Flow outlet	9.2	2.4	30.5	42.2	
Total parameter sensitivity		1181.7	455.0	1038.4		

Figure 2.3.2. Excerpt from ITOUGH2 output file *sam1p3i.out*, showing global sensitivity measures.

2.4 Parameter Estimation

Part 4 demonstrates the main application of ITOUGH2, i.e., the estimation of TOUGH2 input parameters by automatically calibrating the model against discrete observations in space and time. We concluded from the sensitivity analysis discussed in Section 2.3 that it is most advantageous to use pressure data measured at the center of the column and flow rate data at the inlet to estimate the three parameters of interest. In this synthetic experiment, the data were not actually measured, but rather obtained by running a forward simulation (see Part 2). The second and third columns in file *sam1p2i.col* contain the pressure and flow rate data, respectively, as a function of time (which is stored in Column 1).

While the true parameters $\mathbf{p}_{true}{}^T = (-11.7, 0.35, 10.30)$ are given by the TOUGH2 input file *sam1* used to generate the data, we pretend not to know their values. An initial guess $\mathbf{p}_o{}^T = (-12.0, 0.25, 10.25)$ is provided for each parameter through the ITOUGH2 input file *sam1p4i* (see Figure 2.4.1). Starting from this initial parameter set, correct identification of the true parameter set may serve as a verification of the ITOUGH2 minimization algorithm.

An admissible range is specified for the third parameter in order to prevent ITOUGH2 from suggesting values that lead to an error in the TOUGH2 simulation. The third parameter represents initial gas saturation with physical values in the interval $0 \le S_{gi} \le 1$; however, the admissible range for the corresponding TOUGH2 input parameter is $10.0 < \mathrm{DEP}(2) < 11.0$.

The simulation results are compared at 60 points in time to the data provided on file *sam1p2i.col.* Data could also be directly supplied through the ITOUGH2 input file *sam1p4i.* If reading from an external file, keyword `FILE` must be present, and the file name and time units have to be given on the `>>>> DATA` command line. The data file *sam1p2i.col* (see Figure 2.2.3) contains two header lines that have to be skipped before actual data can be read (command `>>>> HEADER`). Furthermore, the columns holding the times (default: column 1) and observed values (default: Column 2) are specified using command `>>>> COLUMN`.

In the `> COMPUTATION` block, the number of ITOUGH2 iterations is being limited to 5. The input file generates 3 warning messages, which can be ignored.

```
> PARAMETER
  >> ABSOLUTE permeability
     >>> MATERIAL                : SAND_ BOUND
         >>>> ANNOTATION         : log(abs. perm.)
         >>>> LOGARITHM
         >>>> initial GUESS      : -12.0
         >>>> VARIATION          : 0.3
         <<<<
      <<<
  >> POROSITY
     >>> MATERIAL                : SAND_
         >>>> initial GUESS      : 0.25
         >>>> VARIATION          : 0.10
         <<<<
     <<<
  >> INITIAL condition for primary variable No.:2
     >>> DEFAULT
         >>>> ANNOTATION         : Gas entrapped
         >>>> VALUE
         >>>> initial GUESS      : 10.25
         >>>> admissible RANGE : 10.01 10.99
         >>>> VARIATION          : 0.10
         <<<<
     <<<
  <<

> OBSERVATION

  >> select : 60 points in TIME, EQUALLY spaced between
     10.0 600.0 seconds

  >> PRESSURE
     >>> ELEMENT                    : A1125
         >>>> ANNOTATION            : Pressure 1/2
         >>>> HEADER contains       : 2 lines
         >>>> COLUMNS               : 1 2          (time vs. pressure)
         >>>> Read DATA from FILE : sam1p2i.col  (time is in MINUTES)
         >>>> standard DEVIATION   : 200.0 Pa     (measurement error)
         <<<<
     <<<

  >> LIQUID FLOW RATE
     >>> CONNECTION defining inlet: IN__0  A11_1
         >>>> ANNOTATION            : Flow inlet
         >>>> FACTOR                :-1.666667E-05 (ml/min - kg/sec)
         >>>> HEADER contains       : 2 lines
         >>>> COLUMNS               : 1 3          (time vs. flow rate)
         >>>> Read DATA from FILE : sam1p2i.col  (time is in MINUTES)
         >>>> standard DEVIATION   : 5.0 ml/min   (measurement error)
         <<<<
     <<<
  <<
```

Figure 2.4.1. ITOUGH2 input file *sam1p4i*. Performs parameter estimation by inverse modeling.

```
> COMPUTATION

  >> STOP
     >>> after : 5  ITERATIONS
     >>> ignore WARNINGS
     <<<

  >> ERROR
     >>> use A PRIORI error variance for error analysis
     <<<

/*
  >> OPTION
     ... solve FORWARD problem only
     >>> perform SENSITIVITY analysis
     <<<

  >> OUTPUT
     >>> PLOTFILE contains : COLUMNS
     <<<
*/
  <<
<
```

Figure 2.4.1 (cont.). ITOUGH2 input file *sam1p4i*. Performs parameter estimation by inverse modeling.

By default, the error analysis is based on the *a posteriori* or estimated error variance which is calculated from the final residuals. In our case, however, the estimated error variance would be very close to zero because no random noise representing measurement errors has been added to the synthetic data. The command `>>> A PRIORI` makes ITOUGH2 use the *a priori* defined error variance for the error analysis, i.e., it is assumed that the final residuals exhibit a standard deviation of 200 Pa and 5 ml/min, respectively. Blocks `>> OPTION` and `>> OUTPUT` are deactivated by surrounding them with "`/*`" and "`*/`" comment characters, making ITOUGH2 perform the default application, i.e., parameter estimation by means of the Levenberg-Marquardt minimization algorithm. The plot file will be generated using the default time units (seconds) and the default format instead of columns. The default plot file format, which is TECPLOT (plot file extension *.tec*), can be changed by redefining variable `IPLOTFMT` in `BLOCK DATA IT`, file *it2main.f*.

The inversion is started by typing

```
itough2 sam1p4i sam1 3 &
```

The optimization process can be followed during execution by typing the UNIX command `prista` which displays the current status of the inversion, i.e., the number of TOUGH2 runs and ITOUGH2 iterations completed, parameter updates and current parameter values,

reduction and current value of objective function, etc. Repeated use of `prista` may suggest termination of the inversion before the specified maximum number of ITOUGH2 iterations has been reached because no significant reduction of the objective function can be achieved. Termination is supported by the `kit` command, which ensures that complete output is generated before execution of ITOUGH2 is stopped. The installation and usage of UNIX commands `prista` and `kit` are described in *Finsterle* [1997b].

After completion of the inversion, results are written to various output files. The main ITOUGH2 output file is named *sam1p4i.out* and contains optimization statistics, error and residual analyses, and the best estimate parameter set. File *sam1.out* contains the TOUGH2 output from the last simulation, which is in most cases the run with the best estimate parameter set. Additional messages can be found in file *sam1p4i.msg*. The values of the best estimate parameter set are also written to file *sam1p4i.par* for convenient restarting of an ITOUGH2 run (see Part 5). The plot file *sam1p4i.tec* contains the interpolated data at the calibration points, the simulated system response using the initial parameter set, and the simulated system response using the best estimate parameter set.

The symbols in Figures 2.4.2 and 2.4.3 represent the synthetic pressure and flow rate data, respectively, as read from file *sam1p2i.col*. The simulation results obtained with the initial and final parameter sets are shown as dashed and solid lines, respectively. The perfect match demonstrates that the minimum of the objective function is accurately identified within 5 iterations. The estimated and true parameter sets are identical, verifying parameter estimation by ITOUGH2 for this well-posed inverse problem.

Part 4 of Problem 1 is convenient to explore many ITOUGH2 features. It is suggested to perform a variety of additional inversions to test the capability of ITOUGH2. For example, minimization could be started from different initial parameter guesses, noisy data could be generated and used for inversion, and systematic errors can be introduced to study their impact on the estimates. Additional or different parameters can be determined, such as the pore space compressibility (instead of or in addition to the initial gas saturation), boundary pressure at the inlet, or parameters of the relative permeability and capillary pressure functions can be subjected to the estimation process. Furthermore, the user should experiment with different options for defining parameters, observations, and data. The use of `prista` and `kit` can also be practiced, i.e., runs can be prematurely terminated and restarted, etc.

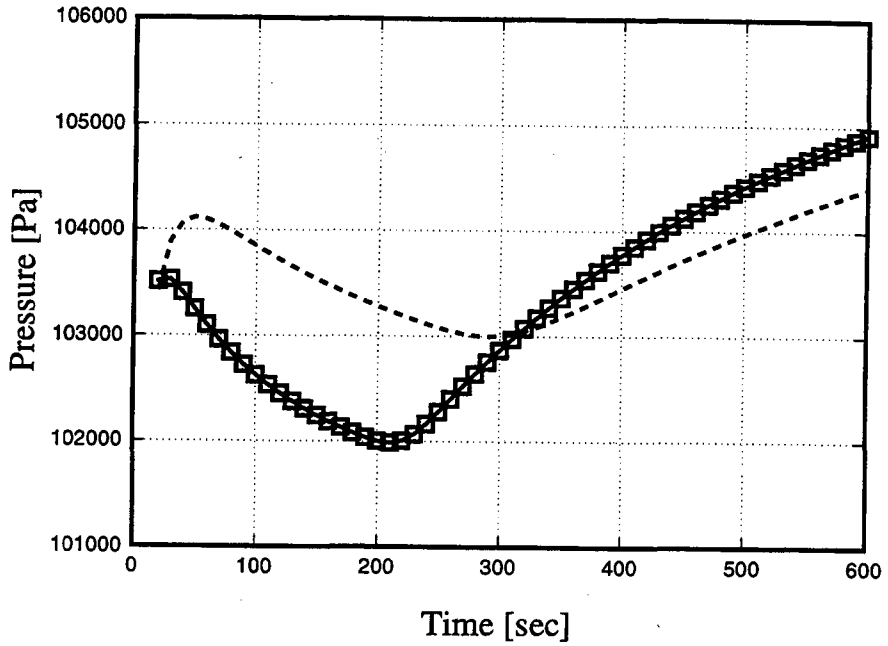


Figure 2.4.2. Pressure transient at center of column calculated with initial parameter set (dashed line) and after optimization (solid line). Synthetic data are shown as squares.

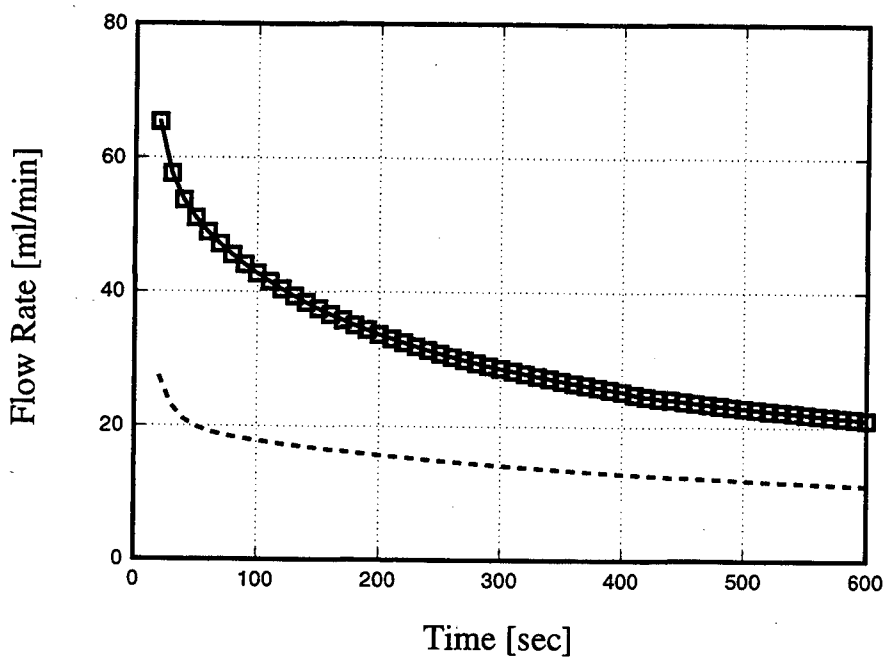


Figure 2.4.3. Flow rates at inlet calculated with initial parameter set (dashed line) and after optimization (solid line). Synthetic data are shown as squares.

2.5 Uncertainty Propagation Analysis

ITOUGH2 offers two methods for studying the effects of parameter uncertainty on model predictions: (1) first-order-second-moment (FOSM) error propagation analysis, and (2) Monte Carlo simulations. For small standard deviations of the input parameters, if the model output can be approximated by a linear function of the parameters within the range of the error band, FOSM is a fast method to calculate a measure of prediction uncertainty that is easy to report. If the model is highly non-linear, and the uncertainties of the input parameters are large, Monte Carlo simulations have to be performed to examine many parameter combinations. For outputs that do not show a normal or log-normal distribution, Monte Carlo simulations provide the full distribution of the system output at the selected points in space and time. The Monte Carlo method is very flexible in handling non-Gaussian distributions of both input parameters and output variables, but they are computationally expensive, and results are difficult to report. In this sample problem we compare both approaches, and at the same time introduce a few additional ITOUGH2 options that are useful for many other ITOUGH2 applications.

The standard deviations of three uncorrelated TOUGH2 input parameters, $\log(k)$, ϕ, and S_{gi}, are assumed to be 0.1, 0.05, and 0.05, respectively. The best estimates of the three parameters are taken from the previous inversion, and are directly read from file *sam1p4i.par*. This ITOUGH2 option (see Figure 2.5.1) allows the execution of a sequence of problems in series as a batch job.

Performing a simulation of a synthetic laboratory experiment, we are interested in the reliability of the model predictions, e.g., the uncertainty of the pressure in the center of the column. The laboratory experiment consists of three parts: (1) injection of water into a partially saturated sand column for 5 minutes under constant pressure, (2) injection of gas for 2.5 minutes, followed by (3) a 2.5 minute shut-in recovery period. In standard TOUGH2, the three test events would have to be run separately in sequence, where the simulation is stopped after 5 and 7.5 minutes, and restarted after adjustment of the boundary condition at the inlet. In ITOUGH2, however, it is necessary to handle all three test events in a single TOUGH2 simulation. This requires automatic adjustment of boundary conditions at $t = 5.0$ and $t = 7.5$ minutes. While general, time-dependent boundary conditions can be supplied through subroutine USERBC (see Problem 5), simple changes of primary variables and element volumes can be conveniently specified directly in the ITOUGH2 input file using the `>>>> RESTART` option (see Figure 2.5.1).

```
> PARAMETER

  >> Take first GUESS (= mean) from FILE: sam1p4i.par

  >> ABSOLUTE permeability
     >>> MATERIAL                  : SAND_ BOUND
         >>>> ANNOTATION           : log(abs. perm.)
         >>>> LOGARITHM
         >>>> standard DEVIATION : 0.10
         <<<<
      <<<

  >> POROSITY
     >>> MATERIAL                  : SAND_
         >>>> VALUE
         >>>> admissible RANGE    : 0.01 0.99
         >>>> standard DEVIATION : 0.05
         <<<<
     <<<

  >> INITIAL condition for primary variable No.:2
     >>> DEFAULT
         >>>> ANNOTATION           : Gas entrapped
         >>>> VALUE
         >>>> admissible RANGE    : 10.01 10.99
         >>>> standard DEVIATION : 0.05
         <<<<
     <<<
  <<

> OBSERVATION

  >> select : 60 points in TIME, EQUALLY spaced between
     10.0 600.0 seconds

  >> RESTART TIME: 1  in [MINUTE]
     5.0
     IN__0  2  10.99   (replace water by air in injection grid block)

  >> RESTART TIME: 1  in [MINUTE]
     7.5
     IN__0  0  1.0E-06 (reduce volume for shut-in recovery)

  >> PRESSURE
     >>> ELEMENT                   : A1125
         >>>> ANNOTATION           : Pressure 1/2
         >>>> NO DATA (this is a prediction)
         <<<<
     <<<
  <<
```

Figure 2.5.1. ITOUGH2 input file *sam1p5i* used to examine prediction uncertainty using first-order-second-moment (FOSM) error propagation analysis.

```
> COMPUTATION

  >> JACOBIAN
     >>> use CENTERED finite difference quotient and
     >>> PERTURB by as much as: 5 %
     <<<

  >> ERROR
     >>> draw error band on (1-ALPHA)=: 95 % confidence level
     >>> First-Order-Second-Moment (FOSM) error propagation analysis
     <<<
  <<
<
```

Figure 2.5.1 (cont.). ITOUGH2 input file *sam1p5i* used to examine prediction uncertainty using first-order-second-moment (FOSM) error propagation analysis.

When performing a FOSM analysis, it is suggested to use a relatively large perturbation factor of 5% in combination with a centered finite difference quotient. The plotfile *sam1p5i.tec* contains the predicted pressure for the mean parameter values as well as the upper and lower bounds of the error band on the confidence level specified by the `>>> ALPHA` command.

In order to invoke Monte Carlo simulations, the `> COMPUTATION` block has to be adjusted as shown in Figure 2.5.2. Various seed numbers should be tried in combination with keyword `GENERATE only`, until a satisfactory distribution of the input parameters is achieved as shown in Figure 2.5.3. Then, keyword `GENERATE` can be deleted to invoke the actual Monte Carlo simulations. Make sure that the parameter range is specified in the `> PARAMETER` block.

```
> COMPUTATION

  >> STOP
     >>> after :100 Monte Carlo SIMULATIONS
     <<<

  >> ERROR
     >>> MONTE CARLO (SEED: 777, g-ENERATE only)
     <<<
  <<
<
```

Figure 2.5.2. Excerpt from ITOUGH2 input file *sam1p6i*. Examines prediction uncertainty using Monte Carlo simulations.

```
==================================================================================
 Parameter No.  1 : log(abs. perm.)   Distribution   :          Normal
 Mean             :    -.117208E+02   Std. Deviation :     .100016E+00
 Lower bound      :    -.120347E+02   Upper bound    :    -.114310E+02

  -.120045E+02     1   *
  -.119441E+02     1   *
  -.118837E+02     3   ***
  -.118234E+02    19   *******************
  -.117630E+02    25   *************************
  -.117026E+02    21   *********************
  -.116423E+02    18   ******************
  -.115819E+02     8   ********
  -.115215E+02     2   **
  -.114612E+02     2   **
==================================================================================
 Parameter No.  2 :   POROSITY SAND   Distribution   :          Normal
 Mean             :     .345833E+00   Std. Deviation :     .496882E-01
 Lower bound      :     .226377E+00   Upper bound    :     .458681E+00

   .237992E+00     3   ***
   .261222E+00     5   *****
   .284453E+00     9   *********
   .307683E+00    15   ***************
   .330914E+00    13   *************
   .354144E+00    19   *******************
   .377374E+00    16   ****************
   .400605E+00     9   *********
   .423835E+00     8   ********
   .447065E+00     3   ***
==================================================================================
 Parameter No.  3 :   Gas entrapped   Distribution   :          Normal
 Mean             :     .102915E+02   Std. Deviation :     .498949E-01
 Lower bound      :     .101773E+02   Upper bound    :     .104157E+02

   .101892E+02     2   **
   .102130E+02     5   *****
   .102369E+02    14   **************
   .102607E+02    20   ********************
   .102845E+02    16   ****************
   .103084E+02    11   ***********
   .103322E+02    14   **************
   .103561E+02     9   *********
   .103799E+02     7   *******
   .104037E+02     2   **
==================================================================================
```

Figure 2.5.3. Excerpt from ITOUGH2 output file *sam1p6i.out*, showing the distribution of the uncertain input parameters.

The results from both the FOSM and Monte Carlo uncertainty analyses are visualized in Figure 2.5.4. While the linear FOSM analysis gives a reasonable estimate of prediction uncertainty for most parts of the experiment, the Monte Carlo simulations reveal an asymmetry of the output distribution in the period where non-linearities prevail. Note that FOSM analysis assigns a certain probability to pressure responses that are below 1 bar, which is physically not possible. The Monte Carlo simulations stay away from that lower

bound. A parameter combination of low permeability, high porosity, and low initial gas saturation yielded the highest pressures.

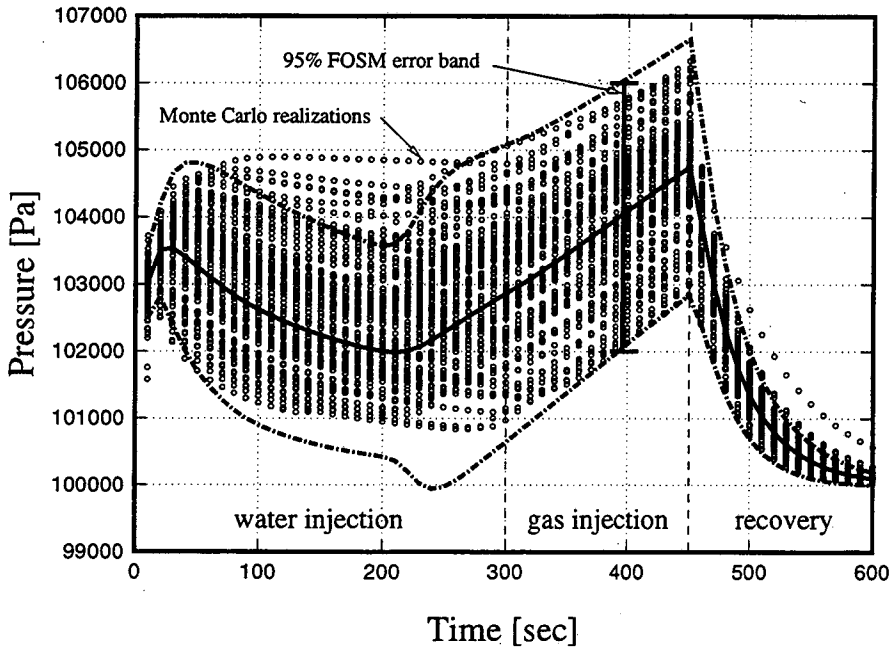


Figure 2.5.4. Comparison between FOSM and Monte Carlo uncertainty propagation analysis.

3. PROBLEM 2: ANALYSIS OF LABORATORY EXPERIMENT WITH KLINKENBERG EFFECT

3.1 Modeling the gas-pressure-pulse-decay experiment

In Problem 2, data from gas-pressure-pulse-decay (GPPD) experiments are analyzed to determine permeability, Klinkenberg slip factor, and porosity of a fine-grained graywacke core plug from the Geysers Coring Project [*Hulen et al.*, 1995]. The experiments were conducted using a specially designed permeameter with small gas reservoirs. A schematic of the experimental apparatus is shown in Figure 3.1.1. To conduct a test, the upstream reservoir is rapidly pressurized to a value about 300 kPa above the initial pressure of the system using nitrogen gas. Gas starts to flow through the sample, and the pressures in both the upstream and downstream reservoirs are monitored as they equilibrate with time.

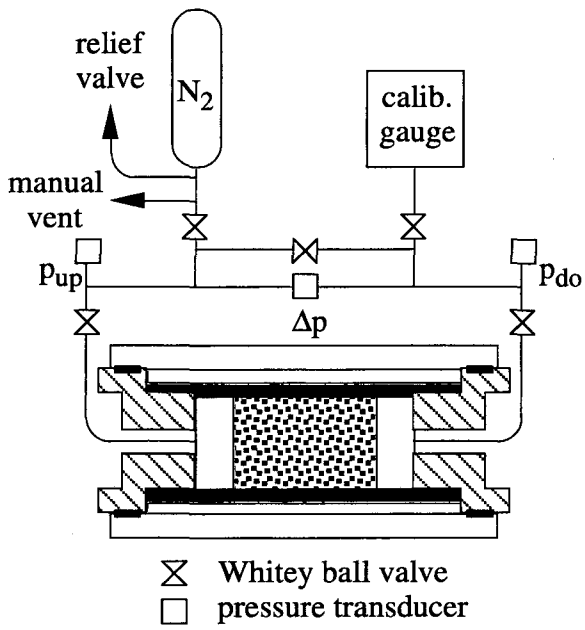


Figure 3.1.1. Schematic of gas-pressure-pulse-decay apparatus.

In porous media with very low permeability and porosity, gas mass flow $\mathbf{F}$ [$kg \cdot s^{-1} \cdot m^{-2}$] may be enhanced as a result of slip flow known as the Klinkenberg effect.

$$\mathbf{F} = -k\left(1+\frac{b}{p}\right)\frac{\rho}{\mu}\nabla p \tag{3.1.1}$$

Here, k is the absolute permeability, ρ is the density, μ is the dynamic viscosity, and p is the gas pressure. The term in parentheses accounts for enhanced gas slip flow, which occurs when the mean free path of the molecules is large relative to the characteristic dimension of

the pores. Slip flow is important at low pressures and in small pores, when a significant fraction of molecular collision is with the pore wall rather than with other gas molecules. In (3.1.1), b is the Klinkenberg slip factor that is a characteristic of both the geometry of the pore space and the thermophysical properties of the gas. It is directly proportional to the mean free path of the molecules [*Klinkenberg*, 1941].

An excerpt from the TOUGH2 input file *sam2* is shown in Figure 3.1.2. A one-dimensional model is connected to two gridblocks representing the gas reservoirs. Core plug, upstream and downstream reservoirs have been assigned to separate model domains (see block ROCKS) for convenient definition of initial conditions (see block INDOM). Furthermore, this arrangement is suitable for subsequent parameter estimation (see Section 3.4). Since data from three experiments will be analyzed simultaneously, there are three such columns merged into the TOUGH2 blocks ELEME and CONNE, where the number of the experiment is indicated by the last digit of the corresponding gridblock names. There is also a GENER block with an extremely small production rate that does not affect the system response. This block will be used in the Part 3 of Problem 2 to simulate and estimate leakage (see Section 3.4).

We perform a stepwise analysis of the GPPD data to illustrate specific aspects of inverse modeling. The consequences of posing a non-unique inverse problem are discussed in Section 3.2. The strong correlation between absolute permeability and Klinkenberg slip factor is resolved by adding data from two additional experiments performed at different pressure levels (Section 3.3). Finally, we discuss the impact of systematic errors, and how they can be parameterized in this specific case. Section 3.4 also addresses the issue of overparameterization and discusses the difference between direct and indirect correlations. Table 3.1.1 summarizes the cases considered. More information can be found in *Finsterle and Persoff* [1997].

Table 3.1.1. Overview of Inverse Modeling Runs

Part	Data	Parameters	Issue
1	pressure level 1	$\log(k)$, $\log(b)$, ϕ	Sensitivity analysis, non-uniqueness, parameter correlation
2	pressure level 1, 2, 3	$\log(k)$, $\log(b)$, ϕ	Joint inversion, biased estimates due to systematic errors
3	pressure level 1, 2, 3	$\log(k)$, $\log(b)$, ϕ, initial pressures p_0, leakage rate q	Joint inversion, parameterization of systematic errors, over-parameterization

```
sam2: Gas-pressure-pulse-decay experiment
ROCKS----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
GEYS1    1     2650.     .0150 2.020E-21 2.020E-21 2.020E-21      2.51      920.
                                         2.106E+07
GEYS2    1     2650.     .0150 2.020E-21 2.020E-21 2.020E-21      2.51      920.
                                         2.106E+07
GEYS3    1     2650.     .0150 2.020E-21 2.020E-21 2.020E-21      2.51      920.
                                         2.106E+07
TOPB1    1     2650.     .9900 2.020E-21 2.020E-21 2.020E-21      2.51   100000.
                                         2.106E+07
BOTB1    1     2650.     .9900 2.020E-21 2.020E-21 2.020E-21      2.51   100000.
                                         2.106E+07
TOPB2    1     2650.     .9900 2.020E-21 2.020E-21 2.020E-21      2.51   100000.
                                         2.106E+07
BOTB2    1     2650.     .9900 2.020E-21 2.020E-21 2.020E-21      2.51   100000.
                                         2.106E+07
TOPB3    1     2650.     .9900 2.020E-21 2.020E-21 2.020E-21      2.51   100000.
                                         2.106E+07
BOTB3    1     2650.     .9900 2.020E-21 2.020E-21 2.020E-21      2.51   100000.
                                         2.106E+07

RPCAP----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
    5
    1      0.000E+00 0.000E+00 0.100E+01
PARAM----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
  -29999    99991000001000000000400003000 0.000E-00 2.334E+00 0.000E+00
          1.00000E09       -1. 0.300E+09           0.0000000
 0.100E+01

 100000.000000000000                 1.0                26.8

ELEME----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
A11 1          TOPB1 .2080E-05 .5091E-03           .2545E-03 .5000E+00-.5000E-03
A21 1          GEYS1 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.1500E-02
A31 1          GEYS1 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.2500E-02
.....          ..... ......... .........           ......... ......... .........
BG1 1          GEYS1 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.5050E-01
BH1 1          GEYS1 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.5150E-01
BI1 1          BOTB1 .2080E-05 .5091E-03           .2545E-03 .5000E+00-.5250E-01
A11 2          TOPB2 .2080E-05 .5091E-03           .2545E-03 .5000E+00-.5000E-03
A21 2          GEYS2 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.1500E-02
A31 2          GEYS2 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.2500E-02
.....          ..... ......... .........           ......... ......... .........
BG1 2          GEYS2 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.5050E-01
BH1 2          GEYS2 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.5150E-01
BI1 2          BOTB2 .2080E-05 .5091E-03           .2545E-03 .5000E+00-.5250E-01
A11 3          TOPB3 .2080E-05 .5091E-03           .2545E-03 .5000E+00-.5000E-03
A21 3          GEYS3 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.1500E-02
A31 3          GEYS3 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.2500E-02
.....          ..... ......... .........           ......... ......... .........
BG1 3          GEYS3 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.5050E-01
BH1 3          GEYS3 .5091E-06 .0000E+00           .2545E-03 .5000E+00-.5150E-01
BI1 3          BOTB3 .2080E-05 .5091E-03           .2545E-03 .5000E+00-.5250E-01

CONNE----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
A11 1A21 1                   3 .5000E-03 .5000E-03 .5091E-03 .1000E+01
A21 1A31 1                   3 .5000E-03 .5000E-03 .5091E-03 .1000E+01
..........                   . ......... ......... ......... .........
```

Figure 3.1.2. TOUGH2 input file *sam2*.

```
INDOM----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
TOPB1
 401000.000000000000 1.0000000000000E+00  .2680000000000E+02
BOTB1
 112500.000000000000 1.0000000000000E+00  .2680000000000E+02
GEYS1
 112500.000000000000 1.0000000000000E+00  .2680000000000E+02
TOPB2
1690000.000000000000 1.0000000000000E+00  .2680000000000E+02
BOTB2
1398900.000000000000 1.0000000000000E+00  .2680000000000E+02
GEYS2
1398900.000000000000 1.0000000000000E+00  .2680000000000E+02
TOPB3
2910000.000000000000 1.0000000000000E+00  .2680000000000E+02
BOTB3
2618300.000000000000 1.0000000000000E+00  .2680000000000E+02
GEYS3
2618300.000000000000 1.0000000000000E+00  .2680000000000E+02

GENER----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
AQ1 2TOP 2                  0     MASS  -1.00E-30
AQ1 3TOP 3                  0     MASS  -1.00E-30

ENDCY----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
```

Figure 3.1.2 (cont.). TOUGH2 input file *sam2*.

3.2 An ill-posed inverse problem

It is obvious from Equation (3.1.1) that the Klinkenberg slip factor b and absolute permeability k are linearly dependent if the average pressure in the core remains constant, i.e., if only late-time data are available from an experiment performed at a single pressure level. If one tries to estimate both parameters based on the pressure data from a single GPPD experiment, the resulting inverse problem is ill-posed because the solution is non-unique. The purpose of this section is to discuss those aspects of the ITOUGH2 error analysis that point towards ill-posedness of the inverse problem.

In Part 1, three parameters are estimated, namely the logarithm of absolute permeability, $\log(k)$, the logarithm of the Klinkenberg slip factor, $\log(b)$, and the porosity ϕ. These three parameters are estimated based on the pressure data from a single GPPD experiment, which was performed at the lowest pressure level, where Klinkenberg effects are expected to be most pronounced. The ITOUGH2 input file is shown in Figure 3.2.1. It also contains input blocks needed for Part 2 and Part 3, surrounded by "/*" and "*/" comment characters.

Note that each parameter refers to multiple materials, i.e., there will be only one estimate of $\log(k)$, $\log(b)$, and ϕ, respectively, which will be assigned to the listed rock types and therefore to all elements with the corresponding material names.

```
> PARAMETER

  >> ABSOLUTE permeability
     >>> MATERIAL: GEYS1 GEYS2 GEYS3 TOPB1 TOPB2 TOPB3 BOTB1 BOTB2 BOTB3
         >>>> estimate LOGARITHM
         >>>> permeability INDEX : 3  i.e. estimate vertical permeability
         >>>> maximum STEP       : 2.0
         >>>> initial GUESS      : -19.0
         >>>> expected VARIATION :   1.0
         >>>> allowable RANGE    : -25.0  -12.0
         <<<<
     <<<

  >> KLINKENBERG parameter
     >>> MATERIAL: GEYS1 GEYS2 GEYS3 TOPB1 TOPB2 TOPB3 BOTB1 BOTB2 BOTB3
         >>>> estimate LOGARITHM
         >>>> initial  GUESS     : 7.0
         >>>> maximum  STEP      : 2.0
         >>>> VARIATION          :   1.0 (scales sensitivity measures)
         <<<<
     <<<

  >> POROSITY
     >>> MATERIAL:  GEYS1 GEYS2 GEYS3
         >>>> estimate VALUE
         >>>> RANGE              : 0.005 0.100
         >>>> maximum STEP       : 0.02        per ITOUGH2 iteration
         >>>> VARIATION          : 0.01
         <<<<
     <<<

/* (delete this line for Part 3, see Section 3.4)

  >> INITIAL PRESSURE

     >>> MATERIAL: TOPB1  = inlet pressure of experiment 1
         >>>> VALUE
         >>>> PRIOR INFORMATION  :  5.0E5
         >>>> VARIATION          :  0.1E5
         <<<<

     >>> MATERIAL: TOPB2 = inlet pressure of experiment 2
         >>>> VALUE
         >>>> PRIOR              : 17.0E5
         >>>> VARIATION          : 0.1E5
         <<<<

     >>> MATERIAL: TOPB3 = inlet pressure of experiment 3
         >>>> VALUE
         >>>> PRIOR              : 30.0E5
         >>>> VARIATION          :  0.1E5
         <<<<
     <<<
```

Figure 3.2.1. ITOUGH2 input file *sam2pli*. Includes input for Part 2 and Part 3.

```
  >> GENERATION RATE

     >>> SOURCE: TOP_2
         >>>> ANNOTATION: Leakage Inlet 2
         >>>> LOGARITHM
         >>>> VARIATION : 1.0
         >>>> GUESS     : -12.0
         >>>> STEP      : 0.25
         <<<<

     >>> SOURCE: TOP_3
         >>>> ANNOTATION: Leakage Inlet 3
         >>>> LOGARITHM
         >>>> VARIATION : 1.0
         >>>> GUESS     : -12.0
         >>>> STEP      : 0.25
         <<<<
     <<<

*/ (delete this line for Part 3, see Section 3.4)

  <<

> OBSERVATION

  >> TIMES: 1
     100.0

  >> TIMES: 30 LOGARITHMICALLY SPACED
     120.0 85200.0

  >> TIME:  2
     67505.0 68605.0

  >> GAS PRESSURE

     >>> ELEMENT: A11_1
         >>>> ANNOTATION  : INLET 1
         >>>> HEADER      : 3
         >>>> SET No.     : 1
         >>>> DATA on File: gppd.dat
         >>>> DEVIATION   : 1000.0  [Pa] (approx. measurement error)
         <<<<

     >>> ELEMENT: BI1_1
         >>>> ANNOTATION  : OUTLET 1
         >>>> HEADER      : 3
         >>>> SET No.     : 1
         >>>> COLUMNS     : 1 3
         >>>> DATA on File: gppd.dat
         >>>> DEVIATION   : 1000.0  [Pa]
         <<<<
```

Figure 3.2.1 (cont.). ITOUGH2 input file *sam2p1i*. Includes input for Parts 2 and 3.

```
/* (delete this line for Part 2, see Section 3.3)

     >>> ELEMENT: A11_2
         >>>> ANNOTATION  : INLET 2
         >>>> HEADER      : 3
         >>>> SET No.     : 2
         >>>> DATA on File: gppd.dat
         >>>> DEVIATION   : 1000.0  [Pa]
         <<<<

     >>> ELEMENT: BI1_2
         >>>> ANNOTATION  : OUTLET 2
         >>>> HEADER      : 3
         >>>> SET No.     : 2
         >>>> COLUMNS     : 1 3
         >>>> DATA on File: gppd.dat
         >>>> DEVIATION   : 1000.0  [Pa]
         <<<<

     >>> ELEMENT: A11_3
         >>>> ANNOTATION  : INLET 3
         >>>> HEADER      : 3
         >>>> SET No.     : 3
         >>>> DATA on File: gppd.dat
         >>>> DEVIATION   : 1000.0  [Pa]
         <<<<

     >>> ELEMENT: BI1_3
         >>>> ANNOTATION  : OUTLET 3
         >>>> HEADER      : 3
         >>>> SET No.     : 3
         >>>> COLUMNS     : 1 3
         >>>> DATA on File: gppd.dat
         >>>> DEVIATION   : 1000.0  [Pa]
         <<<<

*/ (delete this line for Part 2, see Section 3.3)

     <<<
   <<

> COMPUTATION

  >> CONVERGENCE
     >>> number of ITERATIONS:  8
     >>> ignore WARNINGS
     >>> initial value of LEVENBERG parameter: 0.01
     <<<

  >> JACOBIAN
     >>> FORWARD: 5
     <<<
  <<
<
```

Figure 3.2.1 (cont.). ITOUGH2 input file *sam2p1i*. Includes input for Parts 2 and 3.

The three >> TIME statements provide calibration times for Parts 1, 2, and 3. In Part 1, only data from the first GPPD experiment, which lasted for 67,505 seconds, are analyzed. This leads to an automatic adjustment of the time window used for calibration and a corresponding warning message that has to be set aside by command >>> WARNING in block > COMPUTATION.

All the data are read from file *gppd.dat*, which has three header lines. Time is in the first column, and data from the upstream and downstream reservoirs are in the second and third columns, respectively. Data from the three experiments are stored in three sets, separated by a single line of text. The two commands >>>> COLUMN and >>>> SET are used to select the pressure data that correspond to one of the reservoirs and the desired experiment.

The data and the calculated pressures in the upper and lower reservoir are shown in Figure 3.2.2. The dash-dotted lines correspond to the calculated system response with the initial guess for the parameters; the solid lines depict the match after model calibration. The overall system behavior can be described as follows. The gas pressure in the pore space and the two reservoirs is allowed to reach equilibrium prior to testing. After quick injection of a certain amount of gas into the upstream reservoir, gas starts to flow through the sample to the downstream reservoir. Note that the high gas compressibility yields a relatively large storage capacity in the sample itself, leading to a faster pressure decrease in the upstream reservoir and a delayed response in the downstream reservoir. If the experiment were run to steady state, the pressure in the system would be somewhat below the average value of the initial pressure and the applied pressure pulse, the difference being a measure of the amount of gas stored in the pore space of the core.

The elements of the Jacobian matrix **J** provide a means to examine the contribution of each data point to the solution of the inverse problem. In Figure 3.2.3, the scaled sensitivity coefficients are plotted. They are defined as the partial derivatives of the model output with respect to the input parameters, multiplied by the inverse of the respective prior standard deviations:

$$S_{ij} = \frac{\partial y_i}{\partial p_j} \cdot \frac{\sigma_{p_j}}{\sigma_{y_i}} = J_{ij} \cdot \frac{\sigma_{p_j}}{\sigma_{y_i}} \tag{3.2.1}$$

Since only pressure data of equal accuracy are used in this study, an arbitrary value for $\sigma_{y_i} = \sigma_y$ of 1000 Pa can be chosen. The choice of σ_{p_j} can be based on the variance of an independent parameter measurement, for example if porosity was determined by mercury intrusion porosimetry. Prior information in the sense of *Carrera and Neuman* [1986] is introduced by using command >>>> DEVIATION in block > PARAMETER. For this sensitivity analysis, however, σ_{p_j} simply scales the sensitivity coefficients, reflecting the expected variation of a parameter. This interpretation is invoked by command >>>> VARIATION (see Figure 3.2.1).

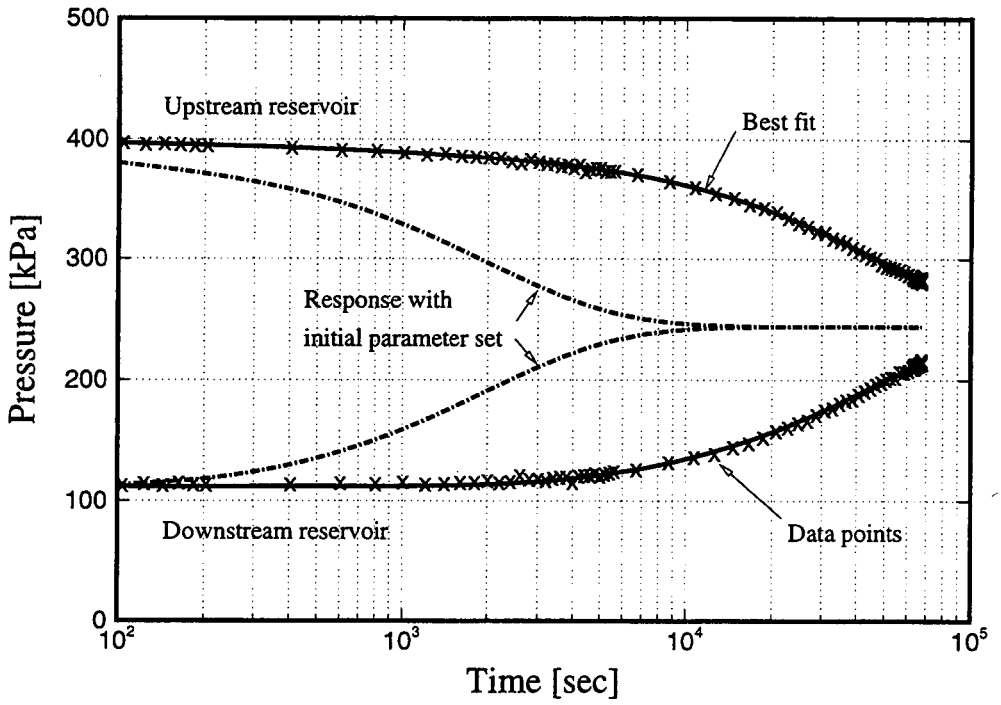


Figure 3.2.2. Comparison between measured and calculated pressures.

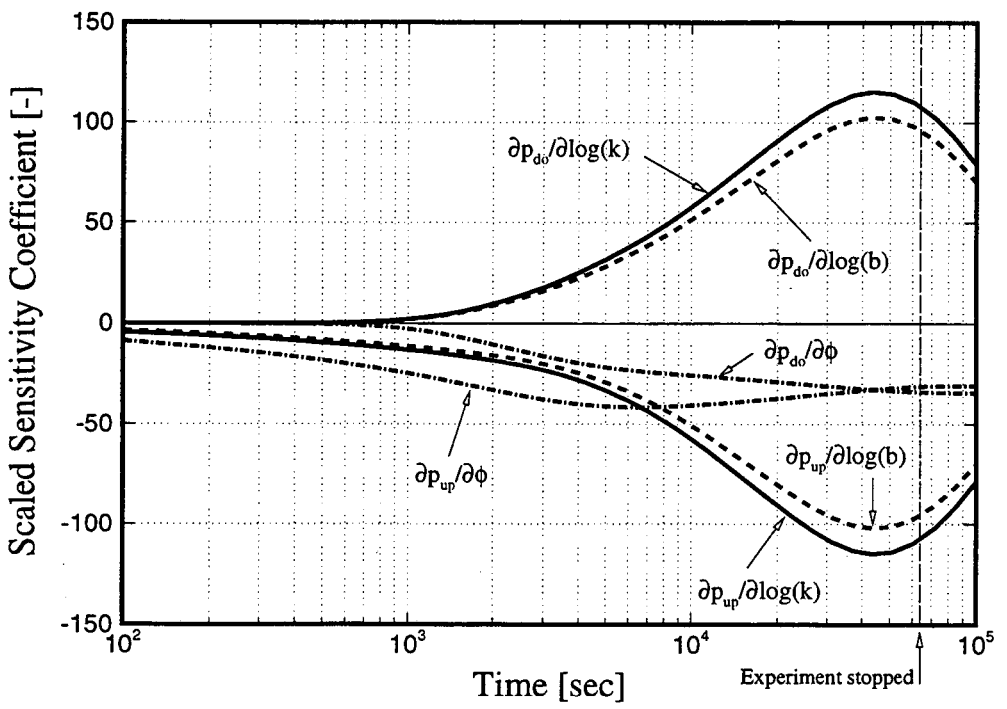


Figure 3.2.3. Sensitivity of pressure in upstream and downstream reservoir with respect to permeability, Klinkenberg factor, and porosity as a function of time.

The curves in Figure 3.2.3 show, for example, that an increase in porosity leads to lower pressures in both reservoirs, whereas an increase in permeability or Klinkenberg factor reduces the upstream pressure, but increases the downstream pressure. This behavior is physically evident. More interesting is the temporal behavior of the sensitivity coefficients. It is obvious that the pressures in the upstream reservoir are immediately affected by changes of the parameters, whereas some time has to pass before the downstream pressure data become sensitive. The absolute sensitivities to $\log(k)$ and $\log(b)$ increase with time and reach a maximum at the inflection point of the pressure transient where gas flow through the sample is sizable and average storativity effects have ceased. The longer the experiment lasts, the less additional information about conductivity can be drawn from the data, since pressure differences and thus flow rates become very small. Eventually the sensitivity coefficients tend to zero. On the other hand, porosity remains sensitive, approaching a constant non-zero value at late time. Porosity could be uniquely identified from the steady-state pressure p_∞ and the initial pressure in the upstream ($p_{0,up}$) and downstream ($p_{0,do}$) reservoirs of volumes V_{up} and V_{do}, respectively, to be:

$$\phi = \frac{(p_{0,up} + p_{0,do})/2 - p_\infty}{p_\infty} \cdot \frac{(V_{up} + V_{do})}{V_{core}} \tag{3.2.2}$$

Analysis of the transient data by inverse modeling techniques allows for a reasonably accurate estimation in a much shorter time, taking advantage of the increased sensitivity of the upstream pressure data at early times.

Note that the information provided by the sensitivity plot can be obtained prior to testing, i.e., in the design stage of an experiment. Test duration and most sensitive periods can be identified, or requirements for sensor accuracy can be derived by comparing the results obtained with different standard deviations σ_{y_i}. Due to the non-linearity of the flow equation, the results of the sensitivity analysis are dependent on the as-yet-unknown parameter values, requiring repetition of the analysis for several potential parameter combinations.

The actual experiment was stopped after about 67,505 seconds. This seems to be a good compromise since the incremental information content of the data with respect to permeability and Klinkenberg factor starts to decrease, and sufficient data have been collected to identify porosity.

The inverse modeling results are summarized in Table 3.2.1. From the perfect match and favorable sensitivities one might expect that an accurate estimation of the three parameters is possible. However, an inspection of the covariance matrix of the estimated parameters reveals a large estimation uncertainty. The standard deviation of both permeability and Klinkenberg factor is about an order of magnitude. This is a result of a high correlation between the two parameters, which yields a non-unique solution.

Table 3.2.1. Summary of Inverse Modeling Results: Initial Guess, Best Estimate, Standard Deviation, and Ratio of Conditional and Joint Standard Deviation, Part 1

Parameter	Initial Guess	Best Estimate	σ_p	$\sigma_p{}^*/\sigma_p$
log(k [m^2])	-19.00	-19.77	1.10	< 0.01
log(b [Pa])	7.00	6.37	1.22	< 0.01
porosity ϕ [%]	1.50	1.02	0.04	0.50

Table 3.2.2. Estimation Covariance Matrix, Part 1

	log(k)	log(b)	porosity
log(k)	1.20	< -0.99	-0.86
log(b)	-1.34	1.50	0.86
porosity	-4.02E-4	4.48E-4	1.79E-7

Diagonal contains variances, lower triangle is covariance matrix, and upper triangle is correlation matrix

Table 3.2.3. Eigenvectors and Eigenvalues of Estimation Covariance Matrix, Part 1

Component	Eigenvectors		
log(k)	0.75	-0.67	1.19E-2
log(b)	0.67	0.75	1.03E-2
porosity	-1.58E-2	2.23E-4	0.9999
Eigenvalues	3.19E-6	2.70	4.46E-8

The covariance matrix with the correlation coefficients in the upper triangle is shown in Table 3.2.2. The correlation coefficient between log(k) and log(b) is very close to -1, i.e., an increase in one parameter can be almost completely compensated by a decrease in the other parameter. The correlation coefficient from the covariance matrix reflects the degree to which the experiment is able to produce independent estimates. Parameter combinations along the eigenvector associated with the largest eigenvalue (see Table 3.2.3) lead to similar pressures and therefore cannot be accurately deduced from the available data. The elements of the corresponding eigenvector indicate the parameters primarily responsible for the ill-conditioning of the inverse problem. In this case, the interdependence between log(k) and log(b) is the dominant source of ill-conditioning as signaled by the large components of the eigenvector. On the other hand, porosity can be estimated relatively independently as indicated by the small eigenvalue for the third eigenvector in Table 3.2.3.

The physical explanation for this correlation structure is evident from Equation (3.1.1) where k and b become linearly dependent for a constant average pressure within the sample. In our transient experiment, the average pressure varies slightly with time due to the storage of gas in the pore space, which makes possible the solution of the inverse problem at hand. A more general measure of parameter dependency is the ratio between the joint and the conditional standard deviation, reported in the last column of Table 3.2.1. The standard deviation σ_p is the square root of the diagonal element of matrix $\mathbf{C}_{pp}$ that refers to the joint probability density function, i.e., it takes into account the influence from all correlated parameters. The conditional standard deviation $\sigma_p{}^*$, on the other hand, reflects the uncertainty of an estimate provided that all the other parameters are exactly known. The conditional standard deviation of parameter i is the inverse of the *i-th* diagonal element of the Fisher information matrix $\mathbf{F} = s_0^{-2}(\mathbf{J}^T\mathbf{C}_{yy}^{-1}\mathbf{J})$. We propose to interpret the ratio $\sigma_p{}^*/\sigma_p$ as a comprehensive measure of how independently a parameter can be estimated. A value close to one signifies an independent estimate, whereas values close to zero indicate a loss of parameter identifiability due to its correlation to other uncertain parameters. Part 1 has demonstrated that achieving a good match is a necessary, but not sufficient condition for parameter estimation.

3.3 A well-posed inverse problem

The objective of Part 2 is to reduce the statistical correlation between k and b. The pressure dependence of gas slip flow suggests that the correlation can be reduced by performing experiments at different pressure levels. A simultaneous inversion of all available data should yield a unique solution. We have analyzed data from three GPPD experiments performed using the same core at pressure levels of about 0.3, 1.55, and 2.75 MPa, respectively. The result from the joint inversion is summarized in Table 3.3.1. First, we note the high values for $\sigma_p{}^*/\sigma_p$, which imply that independent estimates have now been achieved. As shown in Table 3.3.2, the correlation between $\log(k)$ and $\log(b)$ is weakened from -0.99 in the previous case to -0.52. As expected, this leads to a significant decrease in the estimation error. The estimated values have changed by an order of magnitude compared to the previous analysis, in accordance with the correlation structure discussed above. Comparisons of the results of Part 1 and Part 2 clearly demonstrate that a good match and high parameter sensitivity are not sufficient to guarantee a meaningful solution of the inverse problem. Omitting a detailed analysis of the estimation uncertainty and correlation structure may lead to erroneous interpretations.

Table 3.3.1. Summary of Inverse Modeling Results: Initial Guess, Best Estimate, Standard Deviation, and Ratio of Conditional and Joint Standard Deviation, Part 2

Parameter	Initial Guess	Best Estimate	σ_p	$\sigma_p{}^*/\sigma_p$
$\log(k\ [\mathrm{m}^2])$	-19.00	-20.68	0.01	0.84
$\log(b\ [\mathrm{Pa}])$	7.00	7.31	0.02	0.85
porosity ϕ [%]	1.50	1.81	0.09	0.99

Table 3.3.2. Estimation Covariance Matrix, Part 2

	log(k)	log(b)	porosity
log(k)	9.64E-5	-0.52	-0.12
log(b)	-9.86E-5	3.68E-4	-0.02
porosity	-1.04E-6	-2.59E-7	8.53E-7

Diagonal contains variances, lower triangle is covariance matrix, and upper triangle is correlation matrix

Figure 3.3.1 shows the agreement between the calculated and observed pressures. While most of the data are reasonably well matched, late-time pressures are systematically underpredicted for the experiment performed at the lowest pressure level, and overpredicted for the two experiments performed at the higher pressure level. This is better illustrated in Figure 3.3.2 where the residuals are plotted as a function of time. Unlike an ideal residual plot that shows random noise around zero with standard deviation σ_y, Figure 3.3.2 reveals a systematic trend in the residuals. The increasing overprediction of pressures with time for the two experiments on the higher pressure level may indicate a gas leak in the apparatus. Since such a leak is not taken into account in the model, a systematic error is introduced leading to an overestimation of porosity that is increased during the optimization process to compensate for the gas volume leaked to the outside environment.

It is very important to acknowledge the difference between systematic and random components of the residuals. Provided that the true system behavior is identified, the residuals become equal to the random measurement errors. While the individual measurement errors are not known *a priori*, they are described in statistical terms, namely through covariance matrix $\mathbf{C}_{zz}$. However, the impact of systematic errors on the parameter estimates is usually much larger than the impact from the random noise in the data, even under well-controlled laboratory conditions. Systematic errors occur in both the data and the numerical simulation. In many cases it is difficult and also irrelevant to distinguish between a systematic modeling error and a systematic error in the data. Systematic errors are simply the result of a conceptual difference between reality and the model. It is more a question of convenience which side of the problem should be addressed to eliminate potential systematic errors.

In some cases, potential systematic errors can be parameterized and subjected to the estimation process. An example of this approach is discussed in Part 3, where uncertainties regarding initial conditions and potential leaking are addressed.

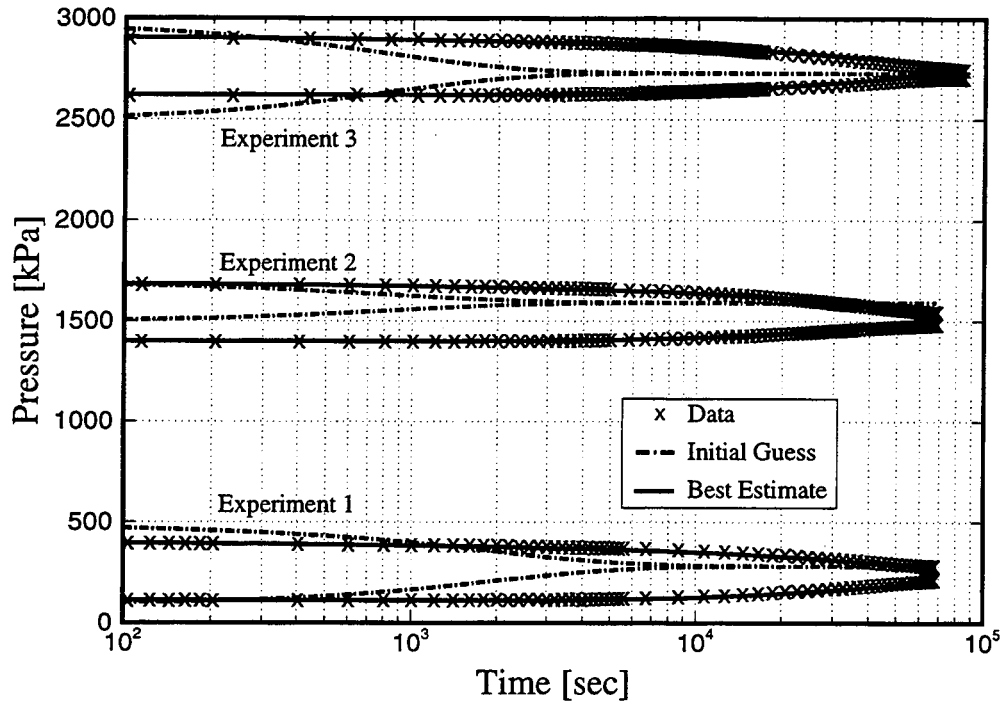


Figure 3.3.1. Comparison between measured and calculated pressure transient curves from three simultaneously inverted gas-pressure-pulse-decay experiments, Part 2.

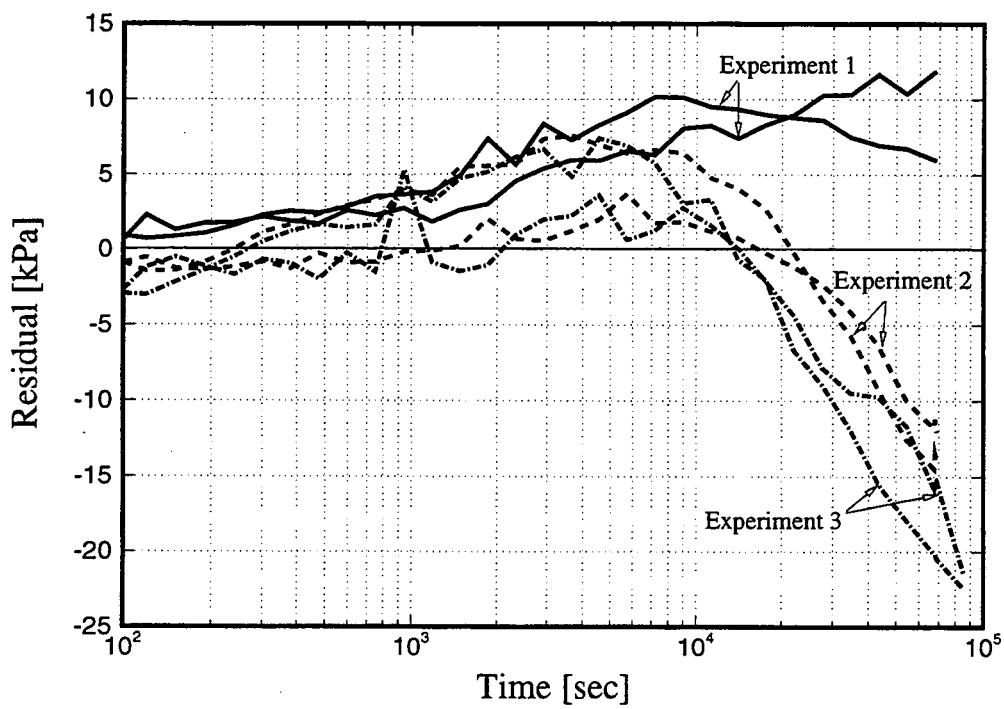


Figure 3.3.2. Residuals as a function of time, showing systematic overprediction of pressures at late times, Part 2.

3.4 Parameterizing systematic errors

In Part 3, we try to reduce the impact of systematic errors and discuss the issue of overparameterization. Recall that the estimated parameters strictly refer to the structure of the model used to invert the data. The fact that a systematic error in the conceptual model leads to biased estimates became apparent in the previous case where the porosity estimate seems to be overpredicted due to leakage. In order to account for potential leakage, we introduce a sink term into the model and estimate its flow rate, which is assumed to be constant. Using a constant mass flux sink term to model the leak seems appropriate since the transient changes in reservoir pressures are relatively small compared to the pressure drop to atmospheric conditions. Furthermore, a test was performed with an impermeable steel plug in the sample holder. Pressure in both the upstream and downstream reservoirs declined exponentially, indicating a constant rate leak to the outside environment.

Besides potential leaks, there is also uncertainty regarding the initial pressure in the upper reservoir. In the previous cases we simply picked the first data point as the initial condition. However, the upstream reservoir undergoes rapid pressurization, causing fluctuations in the data that immediately follow the shut-in of the valves. In order to overcome this problem, we consider the initial pressures in the upstream reservoirs as additional unknown parameters. Accounting for leakage and uncertainty in the initial pressures increases the dimension of parameter vector **p** from 3 to 8. Adding 5 more parameters requires some justification. It is expected that the two new parameter types, initial pressure and leakage rate, can be estimated from independent data, thus mitigating the problem of high parameter correlation due to overparameterization. The initial pressures are most likely determined from early-time data, whereas leakage rate estimates are inferred from late-time data. Moreover, the direct correlations between the three initial pressure estimates are actually zero because they refer to independent experiments. The same argument applies to the two leakage rates. These considerations, which will be assessed further below (see discussion of Tables 3.4.2 and 3.4.3), make it seem unlikely that the inverse problem becomes over-parameterized when increasing the number of parameters from 3 to 8.

With relatively inaccurate initial guesses for all unknown parameters (Table 3.4.1), ITOUGH2 was able to match the data of all three GPPD experiments very accurately within 10 iterations (Figure 3.4.1). The residuals shown in Figure 3.4.2 are much smaller than the ones depicted in Figure 3.3.2. More important, they are devoid of a systematic trend and exhibit a random structure.

The best estimates and their uncertainties are listed in Table 3.4.1. While permeability and Klinkenberg factor are not changed between Parts 2 and 3, a lower porosity value is realized due to the fact that leakage is explicitly modeled. The estimated value is consistent with the result from Part 1 where less leakage is expected due to the low pressure level of that experiment. The total amount of gas leaked out during Experiments 2 and 3 is estimated to be about 0.06 ml and 0.05 ml, respectively.

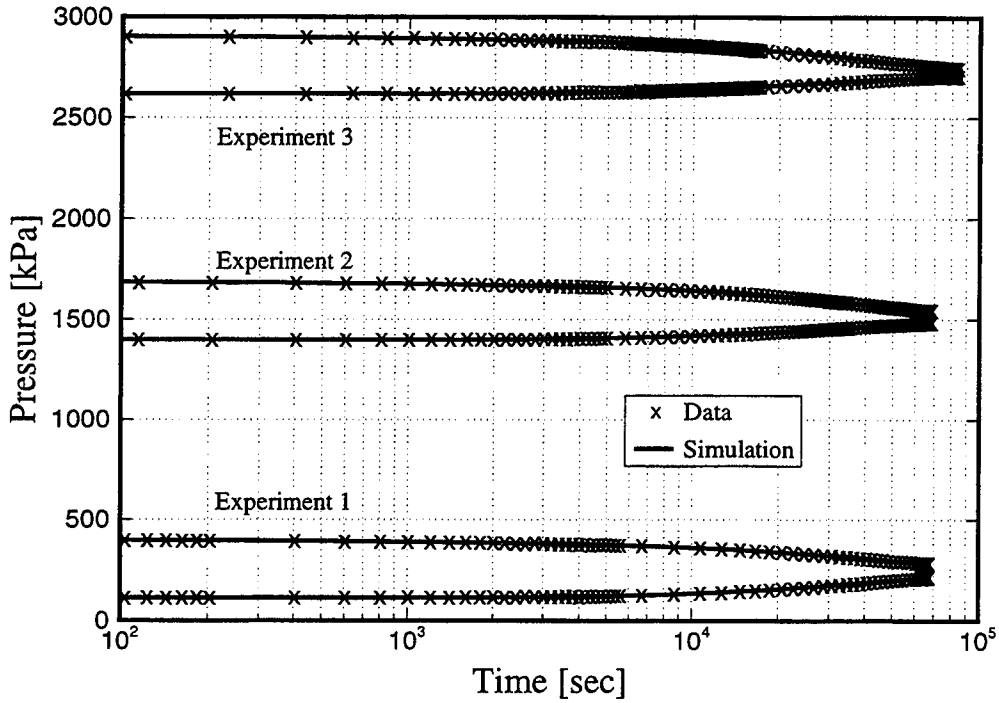


Figure 3.4.1. Comparison between measured and calculated pressure transient curves from three simultaneously inverted gas-pressure-pulse-decay experiments, Part 3.

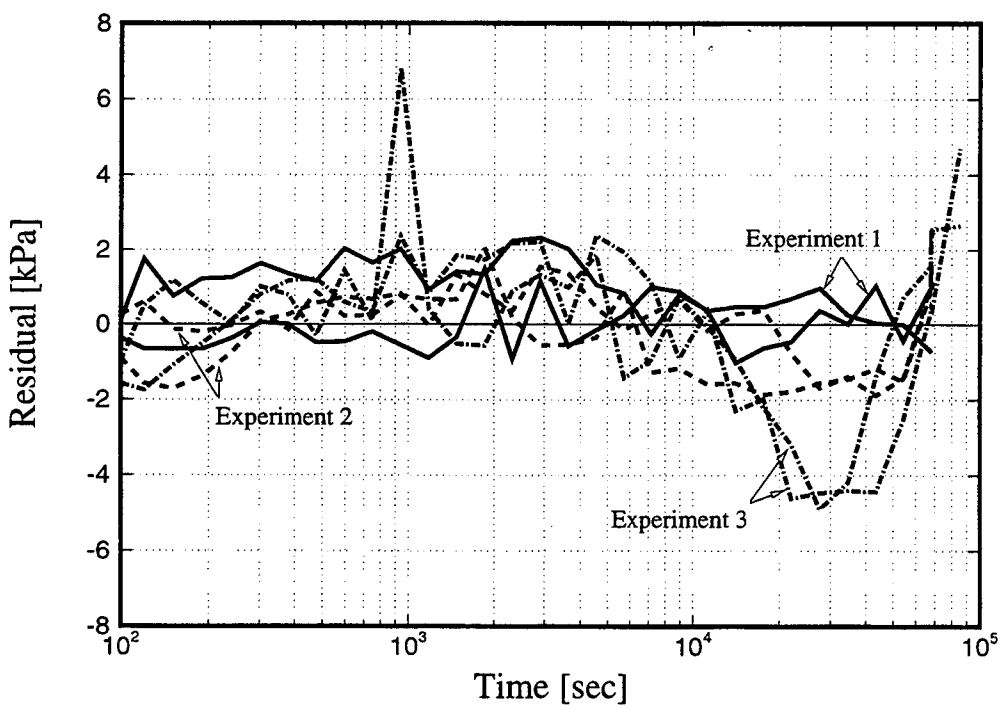


Figure 3.4.2. Residuals as a function of time, Part 3. Residuals exhibit a random structure. Note the difference in scale from Figure 3.3.2.

Table 3.4.1. Summary of Inverse Modeling Results: Initial Guess, Best Estimate, Standard Deviation, and Ratio of Conditional and Joint Standard Deviation, Part 3

Parameter	Initial Guess	Best Estimate	σ_p	$\sigma_p{}^*/\sigma_p$
$\log(k\ [\mathrm{m}^2])$	-19.00	-20.66	< 0.01	0.78
$\log(b\ [\mathrm{Pa}])$	7.00	7.31	< 0.01	0.80
porosity ϕ [%]	1.50	1.05	0.04	0.50
$p_{0,Exp1}$ [bar]	5.00	4.01	0.05	0.69
$p_{0,Exp2}$ [bar]	17.00	16.88	0.04	0.71
$p_{0,Exp3}$ [bar]	30.00	29.07	0.04	0.71
$\log(q_{Exp2}$ [kg/s])	-12.00	-10.79	0.01	0.85
$\log(q_{Exp3}$ [kg/s])	-12.00	-10.71	0.01	0.86

Table 3.4.2. Matrix of Direct (Lower Triangle) and Overall (Upper Triangle) Parameter Correlations, Part 3

	$\log(k)$	$\log(b)$	ϕ	$p_{0,Exp1}$	$p_{0,Exp2}$	$p_{0,Exp3}$	$\log(q_1)$	$\log(q_2)$
$\log(k)$	1.00	-0.51	0.19	0.14	0.32	0.32	0.02	0.02
$\log(b)$	-0.53	1.00	0.05	0.25	-0.08	-0.08	-0.06	-0.06
ϕ	-0.13	-0.08	1.00	0.68	0.58	0.59	-0.34	-0.32
$p_{0,Exp1}$	0.19	0.35	0.47	1.00	0.40	0.41	-0.23	-0.21
$p_{0,Exp2}$	0.20	0.00	0.47	0.00	1.00	0.39	0.12	-0.16
$p_{0,Exp3}$	0.20	0.00	0.47	0.00	0.00	1.00	-0.19	0.14
$\log(q_1)$	-0.02	0.00	-0.37	0.00	0.41	0.00	1.00	0.12
$\log(q_2)$	-0.02	0.00	-0.36	0.00	0.00	0.42	0.00	1.00

Table 3.4.3. Goodness-of-Fit, Residual Statistics, and Model Identification Criteria

Part	Estimated Error Variance s_0^2	Mean of Residuals [Pa]	Std. Dev. of Residuals [Pa]	A-Optimality	Kashyap Criterion
1	1.1	308.5	1007.9	0.041	1031.1
2	43.9	811.6	6591.7	0.003	11235.9
3	2.4	158.4	1525.6	0.001	3465.0

Table 3.4.2 shows the direct correlations in the lower triangular matrix, and the overall correlation coefficients in the upper triangular matrix. The direct correlations are more easily explained from a physical point of view. They indicate to which degree a change in one parameter can be compensated by a change in the other parameter. As mentioned above, the direct correlations between the initial pressure estimates are all zero. Increasing the initial pressure in one experiment cannot be compensated by a change in the initial pressure in any of the other two experiments because the experiments have been performed independently. In the model, there is no connection between the upper reservoir gridblocks of one experiment with the observation points of the other experiment. Despite the apparent independence of the initial pressure estimates as seen in the direct correlation coefficients, there is a substantial overall correlation of about 0.4 between these three parameters. This is a result of indirect correlations. For example, an increase in the initial pressure in the first experiment can be partly compensated by an increase in the absolute permeability, as indicated by a positive direct correlation coefficient. Conversely, this increase in absolute permeability can again be compensated by an increase in the initial pressure of the second experiment. As a result, the two initial pressure estimates are indirectly correlated through permeability (and all the other parameters). Overall correlations are usually difficult to interpret in a inversion involving three or more parameters due to the effects of indirect parameter dependencies.

Parameterization of those aspects of the conceptual model that are most likely to be erroneous is a means to overcome the problem of biased estimation. However, there is a tradeoff between goodness-of-fit and minimum bias on one hand, and estimation uncertainty on the other hand. Increasing the number of parameters always leads to an improvement of the fit, but at the same time increases parameter correlations, resulting in higher estimation uncertainties. This is seen in Part 3 where the standard deviation of porosity is reduced only by a factor of two despite a significant improvement of the fit. All ratios $\sigma_p{}^*/\sigma_p$ indicate higher overall parameter correlations. This is most pronounced for porosity, reflecting its correlation with the initial pressure estimates and leakage parameters. Overparameterization of the inverse problem yields higher parameter uncertainties and thus reduces the capabilities of the predictive model.

Table 3.4.3 summarizes some statistical parameters of the three inversions performed. The estimated error variance s_0^2 is a measure of goodness-of-fit. The first inversion results in the best match. Since the Kashyap model identification criterion has a substantial contribution from the goodness-of-fit measure, it also favors the results from Part 1. However, the A-optimality criterion, which measures the overall parameter uncertainty, clearly indicates that the parameter set from Part 1 is highly ambiguous as compared with the results obtained in Parts 2 and 3. The relatively larger s_0^2-value for Part 3 is a result of actual measurement noise; the standard deviation of the residuals is larger compared with the one of Part 1 despite a smaller bias. According to the A-optimality criterion, Part 2 is an improvement over Part 1. Part 3 performs significantly better than Part 2 with respect to all criteria. This overall performance makes the results from Part 3 the preferred solution. This inversion resulted in a very good match devoid of systematic errors, and yielded accurate estimates. The discussion also shows, however, the an inversion cannot be judged based on a single criteria such as s_0^2 or Kashyap, but that the overall performance should be evaluated.

4. PROBLEM 3: CALIBRATION OF GEOTHERMAL RESERVOIR MODEL

4.1 Problem Statement

In Problem 3, six parameters characterizing the hydraulic, thermophysical, geometric, and natural state conditions of a fractured geothermal reservoir are determined based on synthetic data—pressures, temperatures, vapor and liquid flow rates—obtained from the simulation of steam production and cold water injection.

We consider a two-dimensional five-spot production-injection problem (Figure 4.1.1) previously studied by *Pruess* [1991a] and *Pruess and Wu* [1993]. The problem specifications correspond to conditions typically encountered in deeper zones of two-phase geothermal reservoirs. The medium is assumed to be fractured with embedded impermeable matrix blocks in the shape of cubes with side lengths of 50 m. The permeable volume fraction is 2% with a porosity of 50% for the fracture domain. Reservoir thickness is 305 m. Water with an enthalpy of 500 kJ/kg is injected at a rate of 30 kg/s. Production rate is also 30 kg/s.

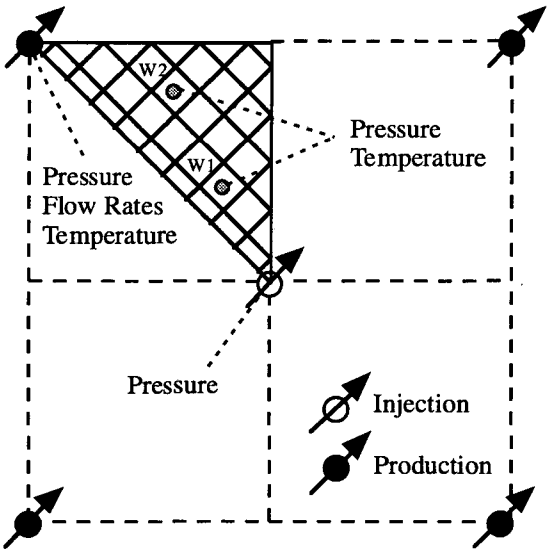


Figure 4.1.1. Five-spot well pattern with grid for modeling 1/8 of symmetric domain. Observation points and type of data measured are also indicated.

We assume that temperature and pressure measurements are taken in the injection (Inj) and production well (Pro), and in two abandoned wells (W1, W2; see Figure 4.1.1). Furthermore, liquid and vapor flow rates are measured in the production well. Note that temperature and pressure measurements are redundant as long as two-phase conditions prevail. TOUGH2 is run in forward mode to generate synthetic data for five years of field performance history, and random noise is added to simulate measurement errors (see Table 4.1.1 for standard deviations).

Table 4.1.1. Observations Used for Model Calibration

Data Type	Location	Std. Dev.
Pressure	Inj/Pro/W1/W2	2.00 bar
Temperature	Pro/W1/W2	5.00 °C
Liquid flow rate	Pro	1.60 kg/s (~5 %)
Vapor flow rate	Pro	0.08 kg/s (~5 %)

Table 4.1.2. True, Initial, and Estimated Parameter Set

Parameter	True Value	Initial Guess	Best Estimate
log (perm. [m^2])	-14.22	-13.00	-14.22
fracture porosity [-]	0.50	0.30	0.60
specific heat [J/kg°C]	1000.00	800.00	993.00
heat cond. [W/m°C]	2.10	2.50	2.31
fracture spacing [m]	50.00	20.00	52.06
temperature [°C]	300.00	250.00	300.21

The model is automatically calibrated against these observations in order to determine the permeability of the fracture system, the porosity of the fracture continuum, the heat conductivity, the specific heat of the rock grains, fracture spacing (which is a parameter of the MINC preprocessor), and the initial reservoir temperature. The true parameter values used to generate the synthetic data are shown in Table 4.1.2.

4.2 Automatic Parameter Selection

The parameters to be estimated here are very different in terms of magnitude and physical meaning. Furthermore, some parameters are very sensitive (e.g., initial reservoir temperature and permeability), while others are not sensitive (e.g., fracture porosity and rock specific heat) or highly correlated (e.g., fracture spacing and heat conductivity). Given this disposition, it seems reasonable to first estimate only the most sensitive parameters to improve the match, and continuously add parameters that are less sensitive or more highly correlated. If parameters of low sensitivity are estimated simultaneously with parameters of very high sensitivity, the former tend to be changed drastically according to their correlations with the latter. This may lead to parameter combinations that are physically not reasonable, causing difficulties in the flow simulation. ITOUGH2 offers an automatic parameter selection procedure that makes the inversion both faster and more stable. Figure 4.2.1 shows the corresponding block of the ITOUGH2 input file. The selection criterion used here examines the potential of a parameter to reduce the objective function S:

$$\delta = |\partial S| \tag{4.2.1}$$

Here, ∂S is the change of the objective function if the parameter is perturbed by a small value (typically 1% of the parameter). Normalizing to the maximum value $\delta_{\max}$ yields the selection criterion ω:

$$\omega = \frac{\delta}{\delta_{\max}} \qquad 0 < \omega \leq 1 \tag{4.2.2}$$

Those parameters with an ω-value larger than $|rsens|$=0.05, i.e., the most sensitive parameters, are selected. Parameters unlikely to significantly reduce the objective function are (temporarily) excluded from the optimization process. If a negative value is given for *rsens* as in this example, the selection criterion is relaxed with each iteration k, and reaches zero for the last iteration *miter*, i.e., all parameters are selected for the final step:

$$rsens_k = |rsens| \cdot \left(1 - \frac{k}{miter}\right) \tag{4.2.3}$$

Due to the non-linearity of the inverse problem at hand, sensitivity coefficients change constantly during the optimization. Therefore, the selection criterion has to be reevaluated from time to time, i.e., parameters may be deactivated and reactivated during the course of an inversion. In this example, a full Jacobian matrix is calculated every third iteration, whereas only the derivatives with respect to the selected parameters are evaluated for intermediate iterations.

Figure 4.2.2 shows the set of active (or activated) and inactive (or deactivated) parameters after k=3 iterations. At this point, permeability is the most sensitive parameter. Initially, reservoir temperature was by far the most sensitive parameter. It has been updated during the first two iterations to be close to the true value. From then on, only minor improvements of the fit can be obtained by changing reservoir temperature, and other parameters such as permeability become more important. Specific heat has not been active for the first two iterations, but has been added to the vector of parameters to be updated for the third iteration because its sensitivity criterion ω=0.0431 is larger than the critical value $rsens_3$ which has been relaxed from initially 0.05 to 0.0389. Note that overall parameter correlation (or independence) could be used as an additional criterion.

```
> COMPUTATION
  >> OPTION
     >>> automatic parameter SELECTION
         >>>> revisit all parameters every :  3    ITERATIONS
         >>>> SENSITIVITY criterion  rsens : -0.05
         <<<<
     <<<
  <<
<
```

Figure 4.2.1. Excerpt from file *sam3p1i* for automatic parameter selection.

```
-----------------------------------------------------------------------
                    Automatic Parameter Selection
-----------------------------------------------------------------------
    Parameter        Rel. Sensitivity     Independence          Status
    Critical Value           .0389            .0000
-----------------------------------------------------------------------
  1 PERMEABILITY            1.0000 +          .8973            active
  2 POROSITY FRACT           .0001 -          .9138          inactive
  3 SPECIFIC HEAT            .0431 +          .7521         activated
  4 HEAT COND.               .0008 -          .1477          inactive
  5 FRACT. SPACING           .0023 -          .1412          inactive
  6 RESERVOIR TEMP.          .2230 +          .8998            active
-----------------------------------------------------------------------
```

Figure 4.2.2. Excerpt from file *sam3pli.out* showing selection criteria after third iteration.

Table 4.2.1. Estimation Error, Correlation, and Sensitivity

Parameter	σ	σ^*/σ	$\sum_{i=1}^{m}\left\lvert\frac{\partial z_i}{\partial p}\cdot\frac{\sigma_p}{\sigma_{z_i}}\right\rvert$
log (perm. [m^2])	0.01	0.95	3543
fracture porosity [-]	0.11	0.79	14
specific heat [J/kg°C]	42.98	0.13	55
heat cond. [W/m°C]	0.27	0.09	54
fracture spacing [m]	4.26	0.06	347
temperature [°C]	0.11	0.85	1832

A total of 10 iterations have been performed to arrive at the best-estimate parameter set shown in Table 4.1.2. The estimation uncertainty as well as measures of total parameter correlation, σ^*/σ (discussed in Section 3), and total parameter sensitivity are given in Table 4.2.1. A comparison of the estimated and the true parameter set reveals that permeability and reservoir temperature are accurately identified. They are the most sensitive parameters and can be determined almost independently, as indicated by the ratio of the conditional and the joint standard deviation, which is close to 1. The estimates of fracture spacing, heat conductivity and specific heat exhibit relatively high standard deviations, which are easily explained by the large correlation coefficients among these three parameters. Especially fracture spacing and heat conductivity have a high positive correlation coefficient, i.e., a larger fracture spacing can be almost completely compensated by an increase in heat conductivity. This statement is true for the type and amount of data available, i.e., the correlation between these two parameters may be reduced by taking additional data (e.g., temperature measurements within the matrix at a known distance from the fracture). Finally, the low sensitivity of fracture zone porosity precludes an accurate determination of this parameter, despite its favorable correlation structure.

4.3 Calibration Results and Total Sensitivity

The system response in the injection, production, and observation wells is shown in Figure 4.3.1. The squares are the synthetically generated and perturbed data points used to calibrate the model. The triangles represent the future system response for the true parameter set. The solid lines are the pressures, temperatures, water and vapor flow rates simulated using the estimated parameter set. For the first 5 years, the deviations between the solid lines and the squares minimize the objective function. Beyond 5 years, the solid lines are predictions, i.e., an extrapolation of the system response matched during the calibration period. The model predictions are uncertain due to uncertainties in the estimated parameters. The standard deviation of the calculated system response, i.e., the uncertainty of the predicted temperature in the production well at a certain point in time, is calculated using linear uncertainty propagation analysis (see file *sam3p2i* for details). The resulting 95% error bands on the model predictions are shown as dash-dotted lines in Figure 4.3.1. They have to be considered optimistic because only the uncertainties of the six selected parameters are taken into account. All the other parameters as well as the model structure are assumed to be exactly known. The true system response (triangles) lies within the estimated error band, despite the fact that the parameter set used for the prediction does not exactly correspond to the true one (Table 4.1.2).

The high accuracy of the model prediction can only be achieved by a combined inversion of all available data. It is obvious that the temperature decrease in observation well W1 could not have been predicted by relying only on temperature data during the calibration phase. In our case, the contribution of temperature measurements to the determination of the parameter set is minor. This is mainly due to the fact that a temperature change of 1 °C leads to a vapor pressure change of about 1 bar that can be more easily detected given the assumed accuracy of pressure measurements. Provided that the expected measurement errors (see Table 4.1.1) are reasonable, the bulk of the information about the parameters of interest is contained in the accurate vapor flow rate measurements and the pressure data in the production well. An approximate measure of the contribution of a certain observation (e.g., flow rate data of a given accuracy taken over the entire measurement period) to the solution of the inverse problem can be evaluated by adding all the absolute values of the corresponding sensitivity coefficients, weighted by the expected measurement error and scaled by the inverse of the parameter variation. This qualitative measure is summarized in Table 4.3.1. Comparing total sensitivities of individual observations, one can conclude that accurate measurements of vapor flow rates and pressures and temperatures in the injection and production wells would be sufficient to solve the inverse problem, i.e., data from the observation wells are less sensitive in our example. This kind of an analysis can be performed without actually collecting data, i.e., it can be used to design and optimize monitoring systems. The standard deviations of the final residuals (Table 4.3.1) are on the order of the (assumed) measurement errors (Table 4.1.1), indicating that no significant systematic errors are present. Finally, the contribution of each observation type to the final value of the objective function is evenly distributed among the measurements, again rendering the choice of the prior errors reasonable.

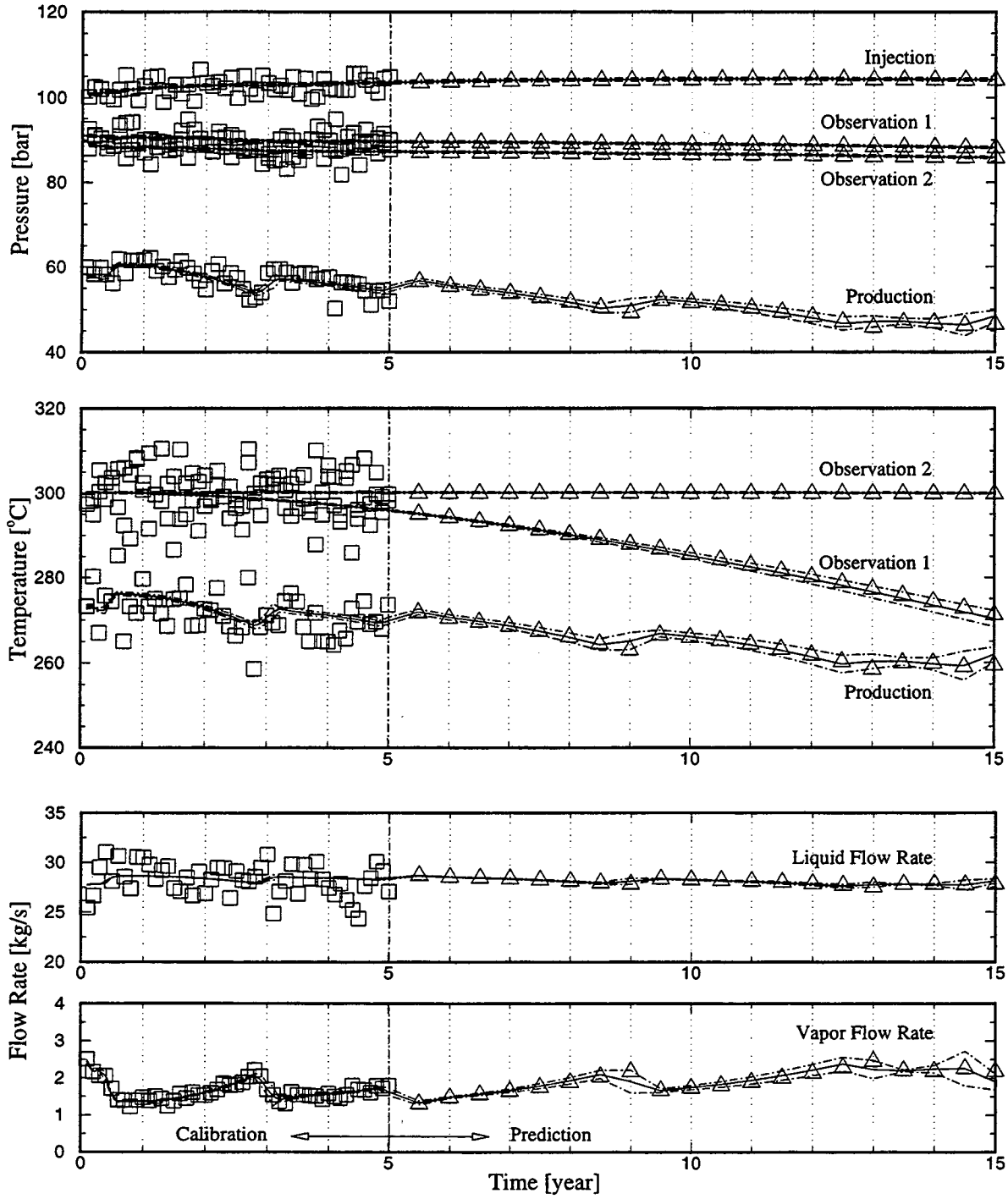


Figure 4.3.1. Calibration and prediction of pressures, temperatures, water and vapor flow rates. Squares are synthetic data points used for calibration. Triangles represent the true system response. Simulation results based on the estimated parameter set are shown as solid lines. Error bands (dash-dotted lines) are calculated using linear uncertainty propagation analysis.

Table 4.3.1. Total Sensitivity of Observations, Standard Deviation of Residuals, and Contribution to Objective Function (COF)

Observation	Total Sensitivity	Std. Dev. of Residuals	COF [%]
Pressure Inj. [bar]	786	1.9	9.8
Pressure Pro. [bar]	1474	1.9	10.1
Pressure W1 [bar]	426	2.2	12.6
Pressure W2 [bar]	358	2.1	11.7
Temp. Pro. [°C]	666	4.6	9.1
Temp. W1 [°C]	107	5.4	12.3
Temp. W2 [°C]	100	5.1	11.3
Water flow rate [kg/s]	87	2.0	10.6
Vapor flow rate [kg/s]	1742	0.1	12.4

5. PROBLEM 4: MULTISTEP DESATURATION EXPERIMENT

5.1 Introduction

Figure 5.1.1 shows a schematic of a flow cell designed for conducting radial desaturation experiments on soil samples. The vacuum applied at the central ceramic extraction cylinder can be adjusted in discrete steps. The apparatus is instrumented with a vial to measure the cumulative water discharge through the central cylinder, and a tensiometer for water potential measurements within the soil sample. The experiment was performed by B. Faybishenko (for more details, see *Finsterle and Faybishenko* [1997]).

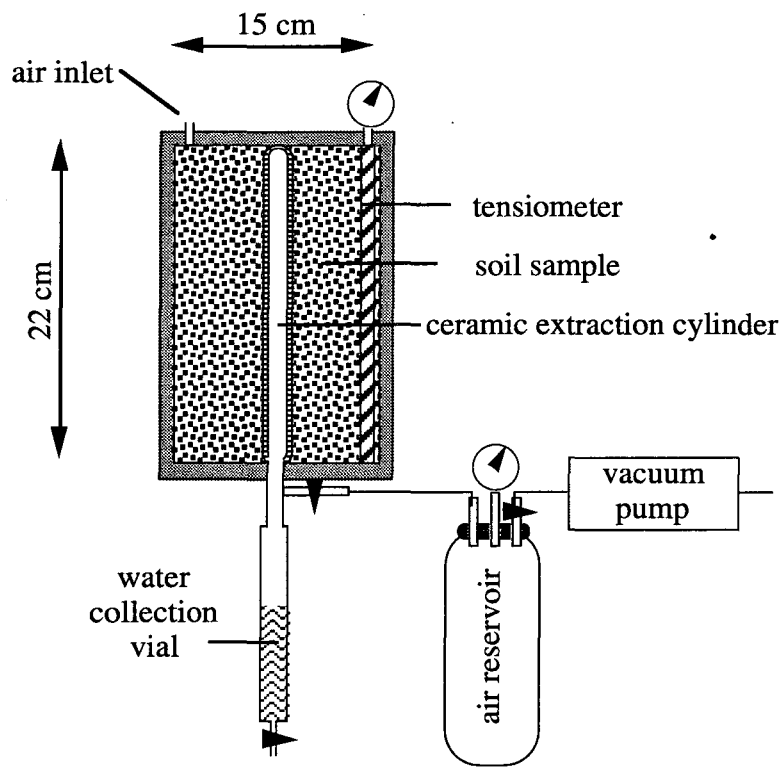


Figure 5.1.1. Schematic of apparatus for radial flow experiment to determine unsaturated hydraulic properties.

Neglecting the minor effects of gravity, a one-dimensional radial model was developed, and EOS9 was used to solve Richards' equation during the stepwise desaturation of the core.

We analyze radial desaturation experiments performed on a soil of clay loam with a dry density of about 1.5 g/cm^3, and a porosity of 0.48. The cumulative water discharge through

the ceramic extraction cylinder and the water potential near the outer wall of the flow cell are used to estimate permeability, pore size distribution index, and air entry pressure.

The estimated parameter values will depend on the functional model used to describe capillary pressure and relative permeability. The estimate of absolute permeability is influenced by the choice of the characteristic curves because permeability is concurrently determined and thus correlated to the parameters of the capillary pressure and relative permeability functions. Therefore, it is crucial to identify the functional model that best represents the true conditions. We consider two models to describe the capillary pressure and relative liquid permeability as a function of liquid saturation S_l. The Brooks-Corey-Burdine (BC) model [*Brooks and Corey*, 1964] is given by:

$$p_c = -p_e \cdot S_e^{-\frac{1}{\lambda}} \tag{5.1.1}$$

$$k_r = S_e^{\frac{2+3\lambda}{\lambda}} \tag{5.1.2}$$

Here, p_e and λ are fitting parameters sometimes referred to as air entry pressure (AEP) and pore size distribution index (PSDI), respectively. The effective liquid saturation S_e is defined as

$$S_e = \frac{S_l - S_{lr}}{1 - S_{lr}} \tag{5.1.3}$$

where S_{lr} is the residual liquid saturation. The Brooks-Corey model is invoked by selecting `IRP=ICP=10`; the TOUGH2 input for this sample problem with the BC model is on file *sam4*.

As an alternative to the BC model, the data were also analyzed using the van Genuchten model (VG) [*van Genuchten*, 1980]:

$$p_c = -\frac{1}{\alpha}\left(S_e^{-1/m} - 1\right)^{1/n} \tag{5.1.4}$$

$$k_r = S_e^{1/2}\left[1 - \left(1 - S_e^{1/m}\right)^m\right]^2 \tag{5.1.5}$$

where

$$m = 1 - \frac{1}{n} \tag{5.1.6}$$

For convenience, and based on a weak analogy to the BC model described by *Morel-Seytoux et al.* [1996], we will refer to parameter n as the pore size distribution index, and $1 / \alpha$ as the air entry pressure. The van Genuchten model is invoked by selecting `IRP=ICP=11`.

While the two models, BC and VG, exhibit only minor differences in the capillary pressure function for intermediate and low liquid saturations, the system behavior differs near full saturation. Because cumulative outflow data contain a significant contribution from the drainage of the larger pores, the two models are expected to perform differently when calibrated against the observed outflow and capillary pressure measurements.

In this sample problem, we demonstrate how step changes in the boundary conditions are specified, and how the objective function can be evaluated on a regular grid. We also discuss the use of prior information and model identification criteria.

5.2 Specifying step changes in boundary condition

In this experiment, the capillary suction in the central extraction cylinder is changed in discrete steps during the experiment. For an inversion of the entire test sequence, changes in the boundary conditions have to be performed automatically. A linear capillary pressure function was chosen that provides the desired capillary pressure between 0 and -100 kPa as a function of liquid saturation, which is the primary variable to be changed using the `RESTART` option in the ITOUGH2 input file as shown in Figure 5.2.1.

```
> OBSERVATIONS

  >> calibration TIMES: 2 [DAYS]
     0.001 0.1

  >> RESTART TIME: 1 [DAYS]
     0.1
     A1__1 1  0.95 (reduces initial liq. sat. in element A1__1 to 0.95
     invoking a capillary force of -5 kPa for second step

  >> calibration TIMES: 4 [DAYS]
     0.25 0.5 0.75 1.1

  >> RESTART TIME: 1 [DAYS]
     1.1
     A1__1 1  0.90 (reduces initial liq. sat. in element A1__1 to 0.90
     invoking a capillary force of -10 kPa for second step
  ...
```

Figure 5.2.1. Excerpt from ITOUGH2 input file *sam4i*, showing block `TIMES` with `RESTART` option for changing saturation boundary condition.

5.3 Minimization Algorithms

As will be verified later, the inverse problem of concurrently estimating permeability, pore size distribution index, and air entry pressure using capillary pressure and cumulative flow rate data from a multistep desaturation experiment has a unique solution, and no local minima exist in the region bounded by the admissible range of parameter values specified in the ITOUGH2 input file. Given these favorable conditions, any of the minimization algorithms implemented in ITOUGH2 should converge to the same solution regardless of the starting point chosen. The different minimization algorithms can be selected simply by pointing the command level marker to the desired option (see Figure 5.3.1). The maximum number of iterations could be set to a very large number to trigger one of the internal stopping criteria. However, a satisfying solution is usually obtained within fewer iterations.

A comparison of the five minimization algorithms currently implemented in ITOUGH2 is given in Table 5.3.1. Note that the performance of each method depends strongly on the problem considered. Furthermore, each of the algorithms could be made more efficient by adjusting some of the method-dependent parameters and convergence criteria. No such fine-tuning has been performed here. Nevertheless, the following observations can be made:

```
> COMPUTATION

  >> STOPPING criterion
     >>> maximum number of ITERATIONS: 1000

  >> OPTIONS

     /* select minimization algorithm here by moving the
     command level marker '>>>' or disabling comment markers */

     >>> GAUSS-NEWTON
         LEVENBERG-MARQUARDT
         downhill SIMPLEX method
         GRID SEARCH - divide parameter space into : 9 9 9 intervals,
                       i.e., evaluate o.f. at 10*10*10=1000 points
 /*  >>> SIMULATED ANNEALING
         >>>> initial TEMPERATURE          : -0.01
         >>>> update after a maximum of    : 50 STEPS
         >>>> annealing SCHEDULE           : 0.95
         >>>> maximum number of ITERATIONS: 1000
         <<<< /*
     <<<
  <<
<
```

Figure 5.3.1. Excerpt from ITOUGH2 input file *sam4i*, showing block `OPTIONS` with different minimization algorithms available.

Table 5.3.1. Comparison of Minimization Algorithm Efficiency

Minimization Algorithm	Number of Iterations	Number of TOUGH2 Runs	Stopping Criterion
Gauss-Newton	7	32	step tolerance
Levenberg-Marquardt	6	41	step tolerance
Downhill Simplex	107	212	fractional range
Simulated Annealing	104	3038	no successful move
Grid Search	-	1000	-

All five minimization algorithms identify the global minimum very accurately. The difference in efficiency is best measured by the total number of TOUGH2 simulations required. For this well-behaved inverse problem, the simple Gauss-Newton algorithm is the most efficient method, followed by the Levenberg-Marquardt algorithm which performs a few unsuccessful uphill steps, increasing the Levenberg parameter, before the scaled step size becomes smaller than the step tolerance. The Levenberg-Marquardt algorithm is expected to be superior to the Gauss-Newton method for highly non-linear problems and if the initial guess is further away from the minimum. The downhill simplex algorithm requires significantly more TOUGH2 simulations than the derivative-based methods. The situation may be improved if a larger initial simplex is specified that encompasses the minimum. Simulated Annealing as a random sampling technique is inefficient in this case, and its advantages over the other algorithms are not required here. Finally, a systematic evaluation of the objective function in the entire three-dimensional parameter space provides a complete description of the topology around the minimum. However, even for a moderate resolution of 10 % of the admissible parameter range, as many as 1000 TOUGH2 solutions have to be calculated.

Figure 5.3.2 shows contour plots of the objective function in three orthogonal parameter planes through the minimum. The plot was created from a subset of the data produced by the grid search method. The shape, size, orientation, and convexity of the minimum provides information about the uniqueness and stability of the inversion, and represents the uncertainty and correlation structure of the estimated parameter set. Furthermore, the presence or absence of local minima can readily be detected. Recall, that the gradient-based minimization algorithms rely on the local examination of the objective function and its derivative, and the linear error analysis is based on the local approximation of the curvature of the objective function at the minimum.

Also depicted in Figure 5.3.2 is the projection of the solution path taken by the Levenberg-Marquardt algorithm that identifies the minimum within a few iterations.

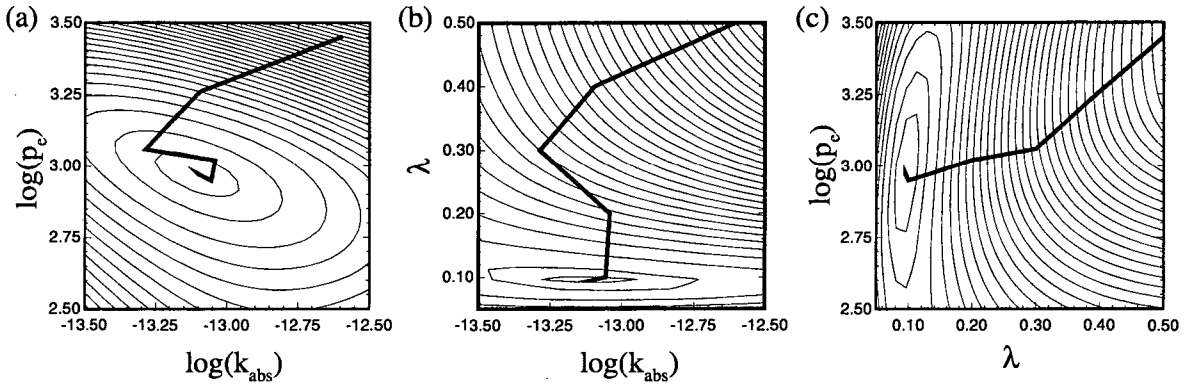


Figure 5.3.2. Contours of the objective function in the three parameter planes (a) $\log(k_{abs})$-$\log(p_e)$, (b) $\log(k_{abs})$-λ, and (c) λ-$\log(p_e)$. The solution path taken by the Levenberg-Marquardt minimization algorithm is shown as a bold line.

5.4 Model Identification and Prior Information

Both the Brooks-Corey (BC) and van Genuchten (VG) model have been calibrated against the capillary pressure and cumulative flow rate data from the multistep desaturation experiment. The results of the inversions performed with the Gauss-Newton algorithm are summarized in Table 5.4.1. The match obtained for the BC model is illustrated in Figure 5.4.1; an almost identical match was obtained with the VG model, indicating that the data does not contain sufficient information for a definite identification of the conceptual model. Note that the estimated permeability value strongly depends on the selected characteristic curves. The VG model requires a significantly larger absolute permeability to match the outflow data. Estimating a high permeability seems necessary to compensate for the sharp decline of the relative permeability curve near full saturation.

Table 5.4.1. Inverse Modeling Results for the Brooks-Corey and van Genuchten Models.

Model	Parameter	Estimate	Std. Dev.
Brooks-Corey	$\log(k\,[\mathrm{m}^2])$	-13.11	0.06
	$\log(p_e\,[\mathrm{Pa}])$	2.97	0.03
	λ	0.10	0.003
van Genuchten	$\log(k\,[\mathrm{m}^2])$	-11.45	0.09
	$\log(1/\alpha\,[\mathrm{Pa}])$	3.14	0.04
	n	1.11	0.004

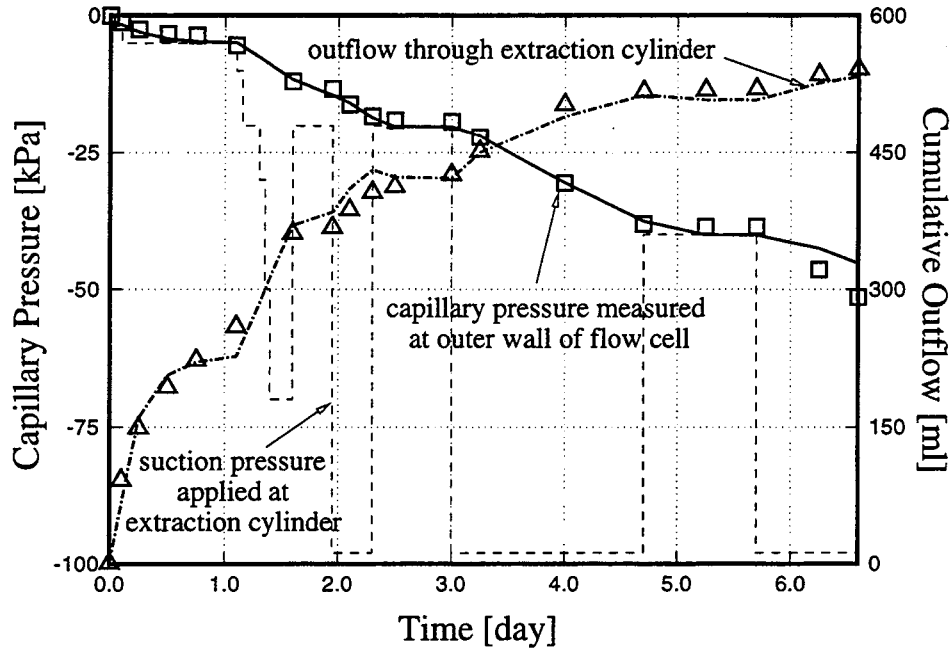


Figure 5.4.1. Comparison between observed (symbols) and simulated (lines) system response for the multi-step radial flow experiment. The prescribed suction pressure at the extraction cylinder is shown as a dashed line.

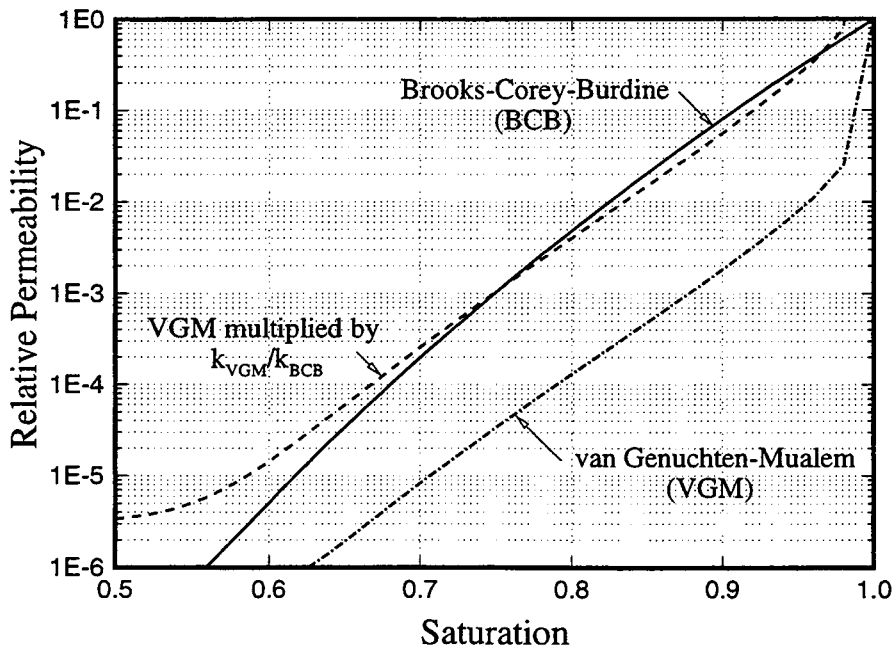


Figure 5.4.2. Relative permeability function derived by inverse modeling for the Brooks-Corey and the van Genuchten model.

Figure 5.4.2 shows the BC and VG relative permeability functions for the respective best estimate parameter sets. Despite the large gap between the two curves, the resulting effective permeabilities, which determine the transient flow behavior, are very similar. To illustrate this, the VG relative permeabilities are multiplied with the ratio of the estimated absolute permeabilities, yielding almost identical effective permeabilities over the range of saturations encountered during the experiment.

The fact that no definitive decision can be made regarding the appropriateness of either the BC or the VG model is an important finding of this analysis. The result becomes more conclusive if prior information about the absolute permeability is introduced. If an independently obtained permeability value is available, this information can be included in the inversion by assigning a penalty to the difference between the prior value and the estimate. In ITOUGH2, this is achieved by replacing the keyword `VARIATION` by `DEVIATION` in the corresponding parameter block. The standard deviation to be specified reflects the relative weight between prior information on one side and the pressure and flow rate data on the other side. If a very accurate permeability measurement is available, this parameter may be excluded from the inversion altogether. If the prior value is uncertain, permeability should be allowed to vary during the inversion according to its relative weight. If the prior permeability value tends to be low, the Brooks-Corey model is likely to perform significantly better than the van Genuchten model, and vice versa.

Finally, if an independent permeability measurement is unavailable, the permeability value concurrently estimated by inverse modeling partly compensates for the error in the model, making the predictions more accurate.

6. PROBLEM 5: COMBINED INVERSION OF STEADY-STATE AND TRANSIENT DATA

6.1 Introduction

Many simulations of transient events assume that the system is initially at equilibrium, from which it evolves after applying a perturbation. Equilibrium conditions are usually obtained by running the model to steady state in a calculation separate from the transient simulation. Steady-state conditions, however, may depend on one or more of the parameters that are to be estimated using the transient data. Furthermore, one might want to concurrently calibrate against steady-state data representing the natural state, and transient data from the test response. Note that the simulation time needed to reach steady state is usually unknown and may change if the parameter set is updated.

ITOUGH2 allows one to perform inversions that require a steady-state run followed by a transient simulation. Once a simulation is terminated by one of the criteria suggesting that steady state has been reached, calibration against the steady-state data points is performed, the time is set to zero, and a second run is invoked for matching the transient test response (Figure 6.1.1). The primary variables at the end of the steady-state period are stored on file *SAVE* and can be used as initial conditions for the following TOUGH2 simulation, which is expected to perform similarly because only one or a few parameters will be updated, i.e., steady state is expected to be reached within a few time steps.

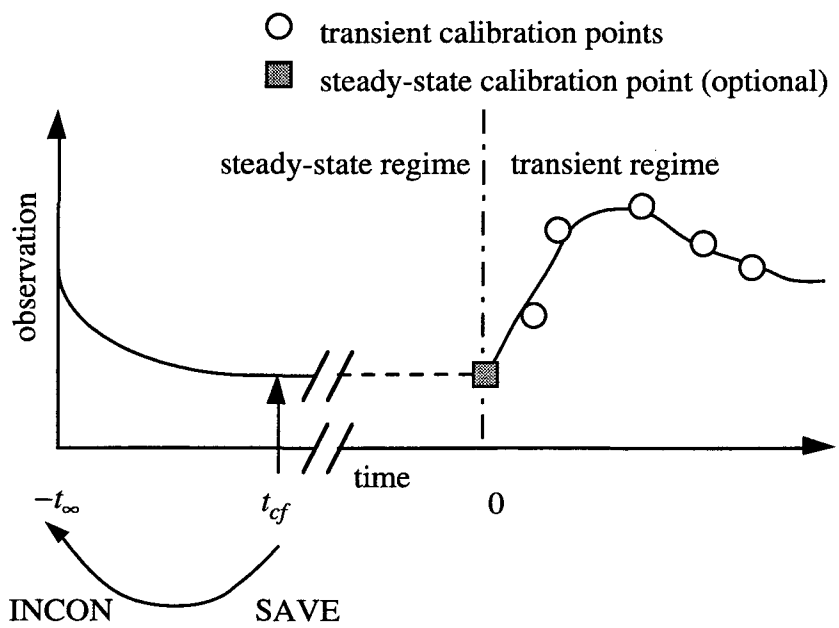


Figure 6.1.1. Combining a steady-state run with a subsequent transient simulation.

Besides demonstrating the combined inversion of steady-state and transient data, this sample problem was also designed to address a number of additional, less frequently used ITOUGH2 features such as:

- Specifying time-dependent Dirichlet boundary conditions
- Performing combined inversions of data from multiple boreholes
- Estimating an unknown constant added to the observed data
- Adjusting units of data to standard TOUGH2 units including shifts
- Specifying a constant perturbation for calculating derivatives
- Using keyword `SAVE` to speed up a series of steady-state simulations
- Using time windows

6.2 The Forward Problem

The forward problem consists of simulating steady-state water percolation in two one-dimensional columns representing boreholes drilled into a thick unsaturated zone. The steady-state run is followed by a transient simulation of atmospheric pressure fluctuations that propagate from the land surface through several hydrogeologic layers to the water table.

Performing a steady-state run prior to a transient simulation is indicated by the ITOUGH2 command `>>> STEADY-STATE` (see Figure 6.3.1), and by setting the starting time of the simulation (TOUGH2 variable *TSTART)* to a large negative number $-t_\infty$. The absolute value $|-t_\infty|$ must be larger than the duration required to reach steady state, e.g., $-t_\infty = -10^{20}$ seconds. The TOUGH2 simulation proceeds until a convergence failure occurs at an unknown point in time t_{cf}. This is considered the steady-state solution to be matched to steady-state data. The primary variables at that point are written to file *SAVE,* which is used as the file with initial conditions for the subsequent TOUGH2 simulation (see Figure 6.1.1). The simulation time is then set to zero, and the transient simulation is started. This requires that the atmospheric pressure fluctuations are initiated as time-varying boundary conditions at time zero. This is achieved by providing code in subroutine USERBC (file *it2user.f*), part of which is shown in Figure 6.2.1. Note that the two boundary elements (`TPA80` and `TPC10`) representing the land surface for the two columns must be active, i.e., they should have a large gridblock volume of 10^{50} m^3 rather than a zero volume. In the first part of the subroutine, the data are read from file *atmos.dat* and stored in the three arrays *DTIME*, *DVALUE1*, and *DVALUE2*. In the second part of the subroutine, the atmospheric pressure at time *TIME* is linearly interpolated between two observed values and assigned to the first primary variable *X(1)*. Note that variable *MOP(22)* is set to 1, and that the data file *atmos.dat* has to be given on the command line as follows:

```
itough2 -fi atmos.dat sam5i sam5 3 &
```

Command option `-fi` copies the specified file to the temporary directory.

We mention here that the user should not specify time-dependent Dirichlet boundary conditions that involve a phase change associated with primary variable switching.

```
*********************************************************************
      SUBROUTINE USERBC(N,CELEM,VOLUME,TIME,X)
*********************************************************************
* User specified boundary condition                                 *
* Set MOP(22).GE.1 to invoke this subroutine                        *
*     MOP(22)=1 don't call EOS                                      *
*     MOP(22)=2 call EOS                                            *
* Return user specified boundary condition (vector X) for element   *
* CELEM and time TIME                                               *
*********************************************************************
C
C$$$$$$$$$ PARAMETERS FOR SPECIFYING THE MAXIMUM PROBLEM SIZE $$$$$$$$$$
      INCLUDE 'maxsize.inc'
C
C$$$$$$$$$ COMMON BLOCK FOR PROBLEM TITLE $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
      INCLUDE 'title.inc'

      DIMENSION X(MAXEQ)
      CHARACTER CELEM*5, FILENAME*40

      PARAMETER (MDATA=2000)

      DIMENSION DTIME(MDATA),DVALUE1(MDATA),DVALUE2(MDATA)

      SAVE DTIME,DVALUE1,DVALUE2
      IREAD/0/,I/0/

      IF (TITLE(:4).EQ.'sam5') THEN
         IF (IREAD.EQ.0) THEN
C --- Read table from a file
            IREAD=IREAD+1
            FILENAME='atmos.dat'
            OPEN(UNIT=39,FILE=FILENAME,STATUS='OLD')
C --- Skip one header line
            READ(39,*)
 1001       I=I+1
            READ(39,*,END=1002) DTIME(I),DVALUE1(I),DVALUE2(I)
            GOTO 1001
 1002       CONTINUE
            NDATA=I-1
            CLOSE(39)
            RETRUN
         ENDIF
         IF (CELEM.EQ.'TPA80') THEN
            CALL INTERP1(TIME,X(1),DTIME,DVALUE1,NDATA)
         ELSE IF (CELEM.EQ.'TPC10') THEN
            CALL INTERP1(TIME,X(1),DTIME,DVALUE2,NDATA)
         ENDIF
      ENDIF
      END
```

Figure 6.2.1. Excerpt from subroutine USERBC for assigning atmospheric pressure fluctuations as user-specified boundary conditions.

6.3 The Inverse Problem

Profiles of saturation and water potentials believed to represent steady-state conditions are available from two boreholes. Furthermore, pneumatic pressures are recorded at five locations within the two boreholes. Calibration occurs at time zero (for steady-state data) and at time intervals of 6 hours (for the pneumatic data).

The profiles of steady-state saturation and water potentials require specifying individual data sets for each measured data point. Each set refers to the element at the appropriate elevation. The steady-state measurement is assigned to time zero, and a small time window is opened, so that calibration occurs at time zero only. This triggers a warning message because the single value is automatically extrapolated over the range of the time window. Water potentials are given as positive numbers in bars, and are converted to negative capillary pressures (the units used by TOUGH2) by specifying a multiplication factor of -10^5. The standard deviation refers to log-cycles because calibration occurs against the logarithm of the observed water potential.

The pneumatic data are provided on external files rather than in the ITOUGH2 input file. The times in the input file are shifted so they match the simulation time, and a time window in the shifted time system is specified, indicating the period for which data are available.

Figure 6.3.1 contains an excerpt from the ITOUGH2 input file showing block `> OBSERVATION` with an example of each observation type.

Block `> PARAMETER` is discussed next. We are primarily interested in estimating absolute permeability for each hydrogeologic layer. It is important to note that the observed saturations and water potentials are strongly affected also by the parameters of the capillary pressure and relative permeability functions, and the pneumatic pressure response is governed by gas diffusivity, which includes porosity. To make this sample problem solvable in a reasonably short time, we fix these parameters and concentrate on absolute permeability.

Information about the absolute permeability is contained in the time lag and attenuation of the pneumatic pressure data, rather than in the absolute value of the observed gas pressure. If the mean pressure at a given elevation is not accurately reproduced by the model, a systematic error is introduced. Since an error in the mean pressure affects all data, the parameters will be adjusted as to minimize the differences in the mean pressure rather than match the time lag and attenuation of the pressure fluctuation. In order to avoid biased estimates, we consider the mean pressure as an additional parameter to be estimated, i.e., we allow the pressure data to be shifted by an unknown constant value. The corresponding ITOUGH2 input block is reproduced in Figure 6.3.2. The initial guess for the data shift is taken from a visual inspection of the match obtained with the initial parameter set. It is not weighted in the objective function. On the other hand, the initial permeability estimates are based on core data and are thus weighted as prior information by specifying the respective standard deviations.

```
> OBSERVATIONS

  >> TIMES: 1
     0.0
     for steady-state data

  >> TIMES (EQUALLY spaced): 481 [DAYS]
     120.0 240.0
     for transient data

  >> Pneumatic PRESSURE data

Data from borehole #1 are shifted in time so that the data from 12/1/95
to 12/31/95 are matched to simulation times 150 days to 180 days.

     >>> ELEMENT: FcC10
         >>>> ANNOTATION: BH#1/N Tpcplnc 1258.2 [masl]
         >>>> TIME SHIFT: -32928.0          [HOURS]
         >>>> FACTOR    : 1000.0            [kPa] - [Pa]
         >>>> DATA FILE : N.dat             [HOURS]
         >>>> DEVIATION :    0.1            [kPa]
         >>>> TIME WINDOW: 36528.0 37248.0  [HOURS]
         <<<<
     ...
     <<<

  >> LIQUID SATURATION
     >>> ELEMENT        :    MaA80
         >>>> ANNOTATION: LIQ. A80 tcwM1
         >>>> DATA [lay midpt = 1340.65,   33 pt(s)]
              0.0        .907000E+00
         >>>> DEVIATION:    .750E-01
         >>>> WINDOW   :    -1.0 1.0
         <<<<
     ...
     <<<

  >> CAPILLARY PRESSURE
     >>> ELEMENT        :    MaA80
         >>>> ANNOTATION: CAP. A80 tcwM1
         >>>> FACTOR    :   -0.100000E+06
         >>>> LOGARITHM
         >>>> DATA [lay midpt = 1340.65,   33 pt(s)]
              0.0       0.310900E+01
         >>>> DEVIATION:   0.607
         >>>> WINDOW   :    -1.0 1.0
         <<<<
     ...
     <<<
  <<
```

Figure 6.3.1. Excerpt from ITOUGH2 input file *sam5i* showing blocks with pneumatic pressure, saturation, and water potential data.

```
> PARAMETER

  >> SHIFT of pneumatic data
     >>> SET No.         : 1
         >>>> ANNOTATION: BH#1/N Mean pneum.
         >>>> GUESS     : 950.0   [Pa]
         >>>> PERTURB   :  -1.0   [Pa]
         >>>> VARIATION : 100.0   [Pa]
         >>>> VALUE
         <<<<
     ...
     <<<

  >> ABSOLUTE permeability
     >>> ROCK type: tcwF1
         >>>> LOGARITHM
         >>>> DEVIATION :   0.61
         >>>> RANGE     : -15.0   -9.0
         <<<<
     ...
     <<<
  <<
```

Figure 6.3.2. Excerpt from ITOUGH2 input file *sam5i* showing block for the estimation of a constant added to the pneumatic pressure data.

The *SHIFT* parameter refers to a data set identified by its ordering number in block OBSERVATION. A negative number is given for variable *PERTURB*, i.e., a constant value of 1 Pa is used to numerically calculate the derivatives. This is preferred over the default relative perturbation because the size of the constant shift is arbitrary. Furthermore, if parameters are automatically selected using the sensitivity criterion (see Eq. (4.1)), the change of the objective function can be appropriately adjusted for this special parameter type.

We finally look at the block with the computational parameters (Figure 6.3.3). To save computer time, this sample problem is a restart of a previous inversion, and only one iteration will be performed. Command >>> STEADY-STATE (SAVE) allows TOUGH2 to run into a convergence failure, at which point conditions are assumed to be at steady state. Furthermore, the *SAVE* file will be transferred to the next TOUGH2 run and used as file *INCON*, holding the initial conditions. The final step size will be transferred as the initial time step size for the next run. This speeds up the simulation if only one parameter is perturbed during the calculation of the Jacobian matrix. However, if all the parameters are updated at the end of an iteration, the initial time step size may be too large. The maximum number of consecutive time step reductions is increased (see command >>> REDUCTION) to avoid a premature termination of the steady-state run. Automatic parameter selection is disabled, but may be useful if performing more iterations or estimating additional parameters. Due to the relatively large number of parameters, overall correlations are expected to be high. To ensure a stable progression during the optimization, the total scaled step size is limited to 0.2. Also, the initial value of the Levenberg parameter is chosen to be relatively high. We

allow the inversion to proceed despite the warning messages discussed earlier, which are understood not to affect the run.

```
 > COMPUTATION

   >> OPTION
      >>> STEADY-STATE (SAVE)
/*
      >>> automatic parameter SELECTION
          >>>> revisit after                      :  3 ITERATIONS
          >>>> SENSITIVITY criterion              : -0.20
*/        <<<<
      <<<

   >> STOPPING CRITERIA
      >>> ignore WARNINGS
      >>> number of ITERATIONS                    :  1
      >>> maximum number of time steb REDUCTIONS : 30
      >>> scaled total STEP size                  :  0.2
      >>> initial value of LEVENBERG parameter    :  0.1
#     >>> solve FORWARD problem
      <<<

   >> OUTPUT
      >>> time units are DAYS
      <<<

  >> JACOBIAN
      >>> FORWARD finite differences for first    : 15 iterations
      >>> parameter PERTURBATION                  :  2.0 %
      <<<
   <<
 <
```

Figure 6.3.3. Excerpt from ITOUGH2 input file showing block COMPUTATION.

6.4 Sensitivity Measures and Residual Statistics

The main purpose of this sample problem is to demonstrate ITOUGH2 input features. Since the formulation of the inverse problem is incomplete, the following short discussion is not meant to be an interpretation of the obtained results; it simply describes some of the measures found in the ITOUGH2 output file.

Figure 6.4.1 shows the match of the pneumatic pressure data at three sensors in one of the boreholes. Note that the estimated parameter set also honors the pneumatics in the second borehole, the saturation and water potential profiles, as well as prior information about the permeabilities, i.e., a better match could be obtained if only pneumatic data in the first borehole were considered. Nevertheless, attenuation and time lag are well reproduced by the model.

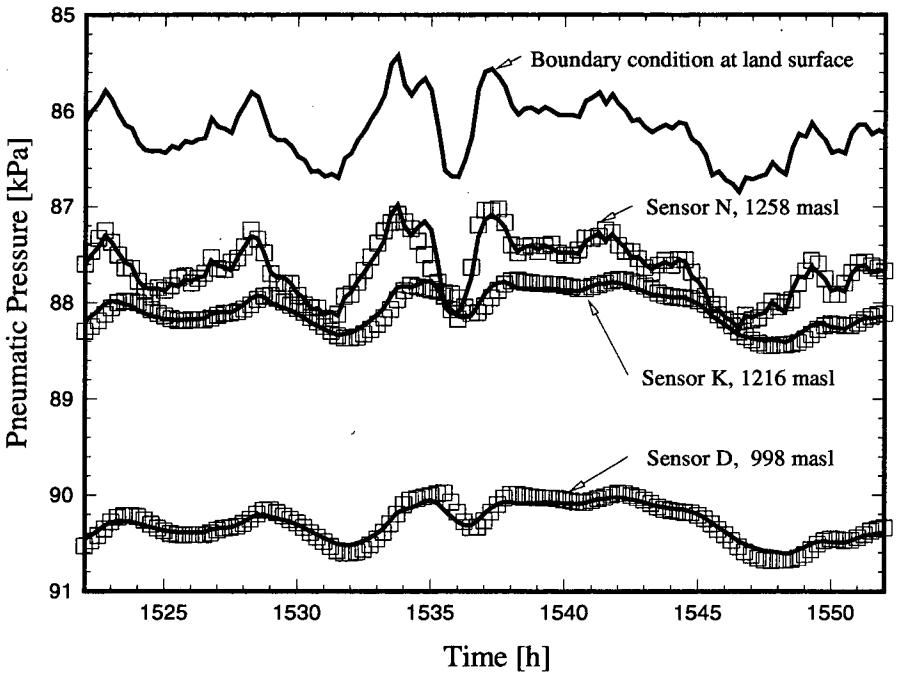


Figure 6.4.1. Match of pneumatic pressures at three elevations in one borehole.

```
Sum of Sensitivity Coefficients
-------------------------------

-------------------------------------------------------------------------------
PARAMETER/OBSERVATION      TOTAL        VARIATION    SENS. OUTPUT   SENS. OBJ. F.
-------------------------------------------------------------------------------
BH#1/N Mean pne:     0.14359E+01     0.10000E+03     0.14359E+03     0.66569E-01
BH#1/K Mean pne:     0.13944E+01     0.10000E+03     0.13944E+03     0.76829E-01
BH#1/D Mean pne:     0.13919E+01     0.10000E+03     0.13919E+03     0.68309E-01
BH#2/400' Mean :     0.13862E+01     0.10000E+03     0.13862E+03     0.67288E-01
BH#2/800' Mean :     0.13763E+01     0.10000E+03     0.13763E+03     0.79238E-01
ABS. K tcwF1   :     0.59516E+02     0.61000E+00     0.36305E+02     0.94505E-01
ABS. K tcwF2   :     0.44942E+02     0.61000E+00     0.27415E+02     0.17524E+00
ABS. K tcwF3   :     0.31218E+03     0.10000E+01     0.31218E+03     0.81973E-01
ABS. K ptnF1   :     0.15197E+03     0.38000E+00     0.57749E+02     0.15692E+00
ABS. K ptnF2   :     0.26584E+02     0.46000E+00     0.12229E+02     0.58988E-01
ABS. K ptnF3   :     0.44825E+02     0.38000E+00     0.17033E+02     0.61901E-01
ABS. K ptnF4   :     0.15002E+03     0.42000E+00     0.63006E+02     0.73299E-01
ABS. K ptnF5   :     0.11681E+03     0.38000E+00     0.44389E+02     0.28752E-01
ABS. K tswF1   :     0.37077E+02     0.10000E+01     0.37077E+02     0.49591E-01
ABS. K tswF2   :     0.65960E+02     0.66000E+00     0.43534E+02     0.18445E+00
ABS. K tswF3   :     0.55330E+02     0.67000E+00     0.37071E+02     0.15757E+00
ABS. K tswF4   :     0.57893E+02     0.56000E+00     0.32420E+02     0.12280E-01
ABS. K tswF5   :     0.11333E+03     0.54000E+00     0.61197E+02     0.23697E+00
ABS. K tswF6   :     0.74579E+02     0.34000E+00     0.25357E+02     0.54305E-01
-------------------------------------------------------------------------------
```

Figure 6.4.2. Excerpt from ITOUGH2 output file *sam5i.out* with sensitivity measures.

Figure 6.4.2 shows an excerpt from the ITOUGH2 output file with the summary of the sensitivity measures for each parameter. The first column is the sum of the absolute values of the weighted sensitivity coefficients for each parameter:

$$\omega_j = \sum_{i=1}^{m} \left| \frac{\partial z_i}{\partial p_j} \cdot \frac{1}{\sigma_{z_i}} \right| \tag{6.4.1}$$

The second column holds the expected parameter variation σ_{p_j}. Multiplying (6.4.1) with σ_{p_j} we obtain a dimensionless, aggregate measure of how sensitive the calculated system response is with respect to parameter j:

$$\Omega_j = \sum_{i=1}^{m} \left| \frac{\partial z_i}{\partial p_j} \cdot \frac{\sigma_{p_j}}{\sigma_{z_i}} \right| \tag{6.4.2}$$

Scaling with σ_{p_j} is necessary to make the sensitivity measures dimensionless and thus comparable with each other. While the magnitude of Ω_j is of little significance, the ratios indicate the relative importance of a given parameter with respect to the other parameters. Here, the permeability of layer *tcwF3* has the largest overall impact on the calculated pneumatic pressures, saturations, and water potentials at the observation points.

The last column in Figure 6.4.2 contains the sensitivity of the objective function with respect to the corresponding parameter:

$$\Xi_j = \left| \frac{\partial S}{\partial p_j} \right| \tag{6.4.3}$$

Note that the observed data do not enter the calculation of Ω, whereas they are part of the objective function used in the calculation of Ξ. The sensitivity ranking of parameters depends on the measure used. A parameter may highly influence the system response (high Ω), but its impact on the objective function may nevertheless be insignificant (low Ξ). This indicates that by changing the corresponding parameter, the match to a subset of the data can be improved, but it gets worse with respect to other data, i.e., the potentially high parameter sensitivity measured by (6.4.2) is not available for reducing the objective function. Note that a measure similar to Ξ is used as the sensitivity criterion for automatic parameter selection (see Section 4.2).

The sensitivity measures discussed here contain a subjective element through the use of the scaling factors σ_{z_i} and σ_{p_j}. Therefore, only large differences in the Ω and Ξ can be considered significant. In this example (Figure 6.4.2), all parameters should be considered of comparable overall sensitivity.

```
MEAN      : Mean of residuals = bias
MEDIAN    : Median of residuals
STD. DEV.: Root mean squared deviation of residuals from bias
AVE. DEV.: Mean absolute deviation of residuals from bias
SKEWNESS : Degree of asymmetry of residuals around bias
KURTOSIS : Relative peakedness of distribution
B/S       : Ratio of bias and standard deviation
C.O.F.    : Relative contribution to final objective function

==================================================================================
TYPE   POINTS    MEAN      MEDIAN  STD. DEV.  AVE. DEV. SKEWNESS KURTOSIS   B/S   C.O.F.
----------------------------------------------------------------------------------
PRIOR I. 19                                                                  5.15 %
PRES.   605 0.740E+00 -0.762E+01  0.866E+02  0.662E+02   0.759    1.231  0.009  83.65 %
CAP. P.  36 0.119E-01  0.422E-02  0.361E+00  0.284E+00   0.379   -0.026  0.033   3.03 %
SAT.     47 0.214E-01  0.423E-01  0.101E+00  0.773E-01   0.066    0.941  0.211   8.17 %
==================================================================================
```

Figure 6.4.3. Excerpt from ITOUGH2 output file *sam5i.out* with residual statistics and relative contribution to objective function for each observation type.

Four different types of data have been used in this inversion: pneumatic pressures, saturations, water potentials, and prior information about the parameters to be estimated. Figure 6.4.3 shows an excerpt from the ITOUGH2 output file that provides statistics of the final residuals and the contribution of each observation type to the objective function. The mean of the residuals is close to zero for all observation types, indicating that the data are on average well matched, and that no significant trade-off between matching data of different types has occurred. The saturation residuals exhibit a slight bias, i.e., the mean of the residuals is relatively large (about 20%, see column B/S) compared with the standard deviation. The final standard deviations are slightly smaller than the assumed measurement errors, leading to an estimated error variance smaller than 1.0.

The last column in Figure 6.4.3 shows the relative contribution of each observation type to the objective function. Note that this measure depends on (1) the number of calibration points of the respective observation type, (2) the size of the residuals, and (3) the prior standard deviation assigned to the data. A large contribution may therefore indicate that (1) there are more data available of this type as compared to data of another type, that (2) the match to the respective data is relatively poor, or that (3) the accuracy of the data was overestimated. In this example, the objective function mainly consists of contributions from the pneumatic pressure residuals. We have already noted that the means and standard deviations of the residuals are acceptable and consistent with the expected quality of the data. However, the number of calibration points selected for matching the transient part of this inversion is somewhat arbitrary. The weight given to the pneumatic pressure data may be too large because we have selected 605 transient calibration points as compared to 36 and 47 steady-state calibration points for the water potentials and saturation data, respectively. While this increased weight can be justified by the fact that there actually are much more pressure data than saturation and water potential data, there remains a subjective element through the choice of the number of calibration times selected for the transient simulation. The prior standard deviation may be appropriately adjusted to account for this effect.

7. PROBLEM 6: VENTILATION EXPERIMENT

7.1 Introduction

A series of ventilation tests have been conducted at the Grimsel Rock Laboratory, Switzerland, a research facility operated by the Swiss National Cooperative for the Disposal of Radioactive Waste (Nagra). Ventilation tests were originally conceived to determine the macro-permeability of crystalline rocks by measuring the total inflow into drift sections with controlled ventilation. In these tests, ventilation is simply viewed as a convenient means to convey the incoming moisture to a measuring device. Accordingly, the standard interpretation of these tests is based on assuming that flow toward the drift is single-phase liquid. However, the estimated matrix permeabilities may be affected by partial drying of the drift wall leading to regions that are dominated by two-phase flow effects. In order to quantify the extent of the two-phase region and study its hydraulic properties, a joint project between the Institute of Terrestrial Ecology, ETH Zürich, and Nagra has been initiated. *In-situ* measurements of water potential, water content, temperature, and ambient air humidity were performed during a ventilation test starting November 26, 1991 [*Gimmi et al.*, 1997].

A schematic of the ventilation experiment is shown in Figure 7.1.1. The experimental site is located in mildly deformed granodiorite that is considered homogeneous on the scale of interest. Two boreholes (BOVE 84.011 and BOVE 84.018) were drilled parallel to the drift. They are equipped with conventional pressure transducers to observe the hydraulic head. Thermocouple psychrometers (TP) were installed at six different depths (2, 5, 10, 20, 40, and 80 cm from the drift wall). They measure negative water potentials in the partially saturated region as a function of time. An estimate of the total inflow to large, sealed off sections of the drift is obtained from measurements of the moisture extracted from the circulated air in a cooling trap.

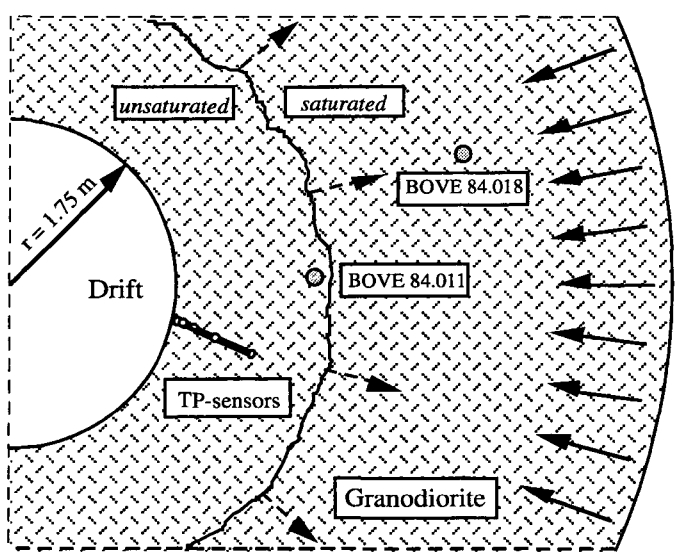


Figure 7.1.1. Schematic of model domain and instrumentation.

A simple radial flow model was developed (file *sam6*). The computational region extends from the drift wall of radius 1.75 m to a presumably unaffected outer boundary at a distance of 6.75 m. A constant pressure of 0.37 MPa is prescribed at the outer boundary, reflecting the undisturbed pressure at drift level. The impact of gravity is neglected. The flow region is partitioned into 200 gridblocks with logarithmically increasing radial distances. Prior to ventilation, the system is run to steady state in order to obtain the initial pressure and saturation distribution. Starting ventilation, formation water evaporates at the surface due to the reduced relative humidity which is the main driving force for the desaturation of the formation. The transfer of moisture at the drift wall is a complicated mechanism that depends on factors such as relative humidity, temperature gradient, wind velocity in the drift, and surface roughness. Rather than explicitly model the moisture transfer across the drift surface, the reduced relative humidity is imposed as a boundary condition at the drift wall, giving rise to an equivalent capillary suction according to Kelvin's equation [*Edlefsen and Anderson*, 1943]:

$$P_{c,equ} = \ln(h)\rho\frac{RT}{M} \tag{7.1.1}$$

where ρ and M are the density and molecular weight of water, and R is the universal gas constant. The relative humidity h in the drift is 68% at a temperature T of 12.5 °C, invoking an equivalent capillary suction $P_{c,equ}$ of -50.0 MPa.

7.2 Discussion of Selected Modeling Issues

In this section we discuss a few aspects of the forward and inverse model to demonstrate how a presumably non-standard observation type (cumulative evaporation rate) can be handled using standard ITOUGH2 features.

First, the drift is modeled as a gridblock with single-phase gas conditions. A linear capillary pressure function is chosen for rock type `DRIFT`, providing the equivalent capillary suction $P_{c,equ}$ of -50.0 MPa at $S_l = 0.0$. The volume of the gridblock representing the drift is set large enough so that the water influx due to ventilation does not change the pressure and saturation conditions in the drift. However, the volume is small enough so that the water mass balance in the gridblock does not suffer from numerical cancellation effects. The rock type `MATRI`, associated with the formation, receives a negative value for the grain specific heat, thus excluding it from the global mass balance calculation. By doing so, the cumulative evaporation flux at the drift surface can be calculated as the change of the total amount of water mass in the gridblock representing the drift. Figure 7.2.1 shows the corresponding block in the ITOUGH2 input file *sam4i*. The observed average evaporation rate of 0.3 microliters per second and square meter of tunnel wall is converted to kilograms of water per meter of tunnel and a ventilation period of 80 days by specifying a multiplication factor of

$$2\pi ra\rho t = 2\pi \cdot 1.75\,[m] \cdot 1.0[m] \cdot 10^{-6}\left[\frac{kg}{\mu l}\right] \cdot 80 \cdot 86400\,[s] = 76\left[\frac{m^2 \cdot s}{\mu l}\right] \tag{7.2.1}$$

```
>> CHANGE of TOTAL MASS of WATER
   >>> MODEL (i.e., DRIFT only because of negative SPHT for MATRI)
       >>>> ANNOTATION: Average evaporation rate
       >>>> FACTOR: 76.0  microliters/s/m^2 -- kg/(m tunnel)/(80 days)
       >>>> DATA                [DAYS]
            80.0 0.3
       >>>> WINDOW: 79.9  80.1 [DAYS]
       >>>> VARIANCE 0.04*2.5E-3 =: 1.0E-4
       <<<<
   <<<
```

Figure 7.2.1. Excerpt from ITOUGH2 input file *sam6i*, showing block OBSERVATION for specifying an average evaporation rate.

The calculated change of water mass in the drift corresponds now to the amount of water evaporated during the 80-day test period. In order to make sure that the calculated and observed water masses are compared only at the end of the simulation, a time window is specified around the last calibration point, which is at $t = 80$ days. Because only one data point is given at $t = 80$ days, a warning message is printed, indicating that the data point is automatically extrapolated to the starting time of the window, and the time of the last calibration point. The warning message could be avoided by providing two data points, one at a time smaller than 79.9 days, and another one with the same observed value at a time larger than 80.1 days. Finally, a variance must be selected to appropriately weigh the evaporation rate data against the matrix potential data. The assumed measurement error variance of $2.5\times10^{-3}\ \mu l^2 \cdot s^{-2} \cdot m^{-4}$ is somewhat arbitrarily multiplied by a factor of 1/25=0.04. This makes the weight of the evaporation rate comparable with one set of water potential measurements, which comprises 25 calibration points in time.

7.3 Minimization, Residual and Error Analysis

In this section we assess the inverse modeling result by performing a detailed error analysis of the estimated parameter set. The three parameters estimated are the absolute permeability k, and the van Genuchten parameters n and $1/\alpha$.

First, we check whether the solution found by the minimization algorithm is likely to be a global minimum. In order to test this, we start minimization from different initial parameter sets. The different starting points can be specified (1) in the TOUGH2 input file, (2) using the fourth-level command >>>> GUESS in each block where the parameters are defined, or (3) using the second-level command >> GUESS (see ITOUGH2 input file *sam4i*). The five initial parameter sets, the best estimates, and the initial and the final values of the objective function are summarized in Table 7.3.1. The five inverse runs result in parameter sets that are almost identical. From this we conclude that the solution is likely to be unique within a parameter space bounded by rather extreme, albeit physically reasonable values, i.e., $-21.0 < \log(k) < -17.0$, $2.0 < n < 5.0$, and $5.7 < \log(1/\alpha) < 6.4$.

Table 7.3.1. Estimates Obtained Starting From Five Different Initial Parameter Sets

Set No.	Parameter	Initial Guess	Initial Objective Function	Final Objective Function	Best Estimate
1	log (k [m^2])	-17.00	67730	123.6	-18.55
	n [-]	3.00			2.46
	log($1/\alpha$ [Pa])	6.00			6.24
2	log (k [m^2])	-18.00	9127	123.5	-18.55
	n [-]	2.00			2.45
	log($1/\alpha$ [Pa])	5.70			6.23
3	log (k [m^2])	-19.00	330	125.3	-18.55
	n [-]	2.50			2.42
	log($1/\alpha$ [Pa])	6.30			6.24
4	log (k [m^2])	-20.00	3297	123.8	-18.54
	n [-]	5.00			2.44
	log($1/\alpha$ [Pa])	6.20			6.22
5	log (k [m^2])	-21.00	4573	124.3	-18.58
	n [-]	4.00			2.47
	log($1/\alpha$ [Pa])	6.40			6.24

In the remainder of this section we discuss some aspects of the residual and error analysis. First, we have to assess whether the match is satisfactory. If the goodness-of-fit criterion suggests that the model is an unlikely match to the data, then both the estimated parameters and the subsequent error analysis are meaningless. A visual inspection of the fit is always recommended. Figure 7.3.1 shows the comparison between the computed and measured water potentials. The actual field data are represented by filled squares. The calibration points that are linearly interpolated between the data are shown as open squares. The calculated solution is depicted as a solid line; the dashed line is the simulation result obtained when only water potentials are matched, i.e., neglecting pressure and evaporation rate measurements. The differences between the two solutions are discussed in *Finsterle and Pruess* [1995]. The match obtained seems reasonable. However, the residual plot printed to the ITOUGH2 output file *sam6i.out* and reproduced in Figure 7.3.2 reveals that certain data sets contain systematic rather than randomly distributed residuals. Data Set No. 4, corresponding to a depth of 20 cm, especially shows a trend in the residuals. Furthermore, water potentials of Data Set No. 1 are systematically overpredicted by the model, as evident by the predominantly negative residuals.

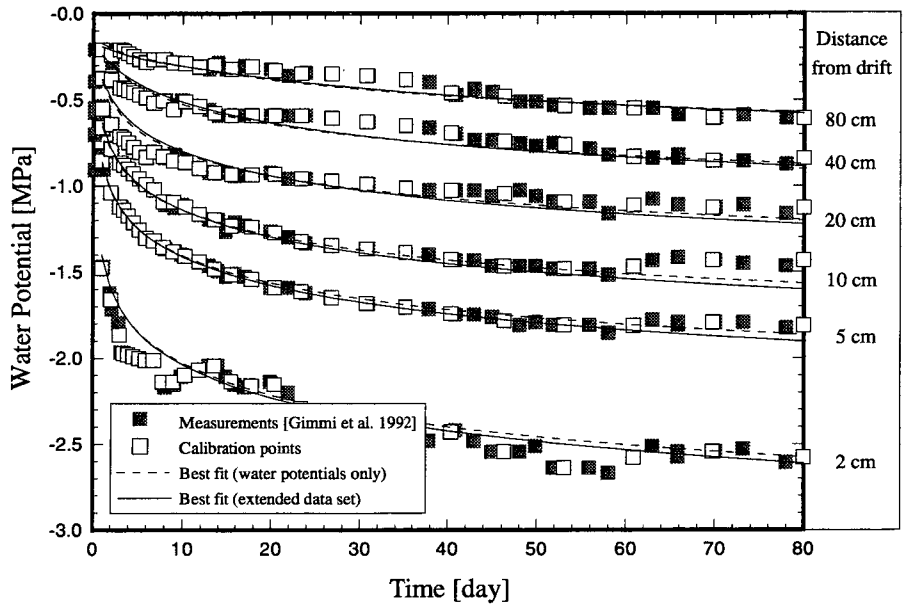


Figure 7.3.1. Comparison between computed and measured water potentials.

```
Residual Plot
-------------

   RESIDUAL     ------------------------------------------------------------
   .155E+06     |                        3                                 |
   .137E+06     |                                                          |
   .119E+06     |                        33                                |
   .102E+06     |                                                          |
   .836E+05     |                                                          |
   .658E+05     |                22                4          555     6    |
   .479E+05     |         1               3        4  4         55    66   |
   .300E+05     |            1     2                444    5  5         6  |
   .121E+05     | 1                        3 33        44    5   5     6 6 |
  -.581E+04     |-1- - - 1 11- - - -2- - - -33333333 -3-334455 - - 5 -6- -6666|
  -.237E+05     | 1                2 22  22 2223   33    4  5      556     6 |
  -.416E+05     |    1 1      1      22 22 22   22 3  23 2     3    56      6|
  -.595E+05     |       1                                  4      3 6 5     |
  -.774E+05     |     1  1                                  4           5    |
  -.953E+05     | 1 1                                        44        5 5   |
  -.113E+06     |     1                   1    1               44        55 |
  -.131E+06     |   1         1  1                                         |
  -.149E+06     |                                                        5 |
  -.167E+06     |  1                                            4 4        |
  -.185E+06     |                  1                               4 4     |
  -.203E+06     |                                                     4    |
  -.221E+06     |               1   1  1                                   |
                ------------------------------------------------------------
          -.259E+07                                                -.188E+06
                                CAPILLARY PRES. [Pa]
```

Figure 7.3.2. Excerpt from ITOUGH2 output file *sam6i.out.* Residual plot.

```
Summary of Residual Analysis
----------------------------
Max weighted residual at observation    :              8
Max weighted residual                   :     -.3597E+01
Max residual                            :     -.1349E+06
Max normalized residual at observation  :              8
Max normalized residual                 :           3.62
Number of large normalized residuals    :              9
Number of poorly controlled observations:              1
```

Figure 7.3.3. Excerpt from ITOUGH2 output file *sam6i.out.* Summary of residual analysis.

The ITOUGH2 output file contains a list of calibration points with the observed value and the corresponding modeling result, the difference between the two (the residual), its weight (the inverse of the prior standard deviation), the squared weighted residual, which is the point's contribution to the objective function, the *a posteriori* standard deviation of the calculated system response (or prediction uncertainty), the local reliability (or measure of influence), and the normalized residual, which can be statistically tested for outliers assuming normality.

A summary regarding large residuals is printed as reproduced in Figure 7.3.3. A total of 9 residuals were identified that are likely to be outliers on a confidence level of 95%. Given the total amount of data available (m=165) and the relatively small value of the largest normalized residual (3.62), it is not expected that these large residuals have a significant influence on the results. This hypothesis could be tested by eliminating the corresponding calibration points or by using one of the robust estimators. ITOUGH2 correctly identified the average evaporation rate as the only poorly controlled observation.

Next, a statistical analysis of the distribution of the residuals is performed, individually for each data set and each observation type. Ideally, the mean of the residuals is zero, and the standard deviation should be consistent with the expected measurement error. In our case, the mean of the residuals for Data Set No. 1 is on the order of the standard deviation, signifying a rather large bias as previously observed on the residual plot. The distribution of all normalized residuals, however, exhibits a bias that is only 10% of the standard deviation.

A linear regression analysis is conducted on a scatter plot in which the calculated pressures are plotted against the observed data (Figure 7.3.4). An intercept of zero and a slope of one are expected. The quantity $(1-R^2)$ expresses the proportion of variance of the prediction that is not attributable to its linear regression on the observation; R is the correlation coefficient. Again, Data Set No. 1, which corresponds to the capillary pressure computed for grid block `RR 19`, exhibits a relatively large intercept, indicating a systematic overprediction of the data. However, only a minor trend seems to be present compared with the one of Data Set No. 4 (grid block `RR 76`) where the slope significantly deviates from one, indicating that the higher, early-time data are underpredicted and the lower, late-time data are overpredicted by the model (see Figure 7.3.1).

```
Linear Regression Analysis Calculated Vs. Observed
--------------------------------------------------

=====================================================================
DATASET                 DATAPOINTS     INTERCEPT       SLOPE          R
---------------------------------------------------------------------
PC(LIQUI) RR 19   [Pa]        27    -0.288E+06   0.897E+00   0.996E+00
PC(LIQUI) RR 39   [Pa]        27    -0.106E+06   0.934E+00   0.999E+00
PC(LIQUI) RR 53   [Pa]        27    -0.153E+06   0.863E+00   0.997E+00
PC(LIQUI) RR 76   [Pa]        27    -0.352E+06   0.641E+00   0.998E+00
PC(LIQUI) RR1 1   [Pa]        27    -0.115E+06   0.810E+00   0.996E+00
PC(LIQUI) RR128   [Pa]        27     0.815E+05   0.117E+01   0.997E+00
P(GAS) RR160      [Pa]         1     0.000E+00   0.000E+00   0.000E+00
P(GAS) RR188      [Pa]         1     0.000E+00   0.000E+00   0.000E+00
Average evapora   [kg]         1     0.000E+00   0.000E+00   0.000E+00
=====================================================================
```

Figure 7.3.4. Excerpt from ITOUGH2 output file *sam6i.out* with linear regression analysis of plot of calculated versus observed system response.

Note that the linear regression analysis does not properly account for differences in measurement quality within a data set. Furthermore, the smaller values have a higher influence on the intercept estimate, and small and large values determine the slope more strongly than intermediate values. Consequently, the results of this analysis should be interpreted with care.

The estimated error variance s_0^2 is an overall goodness-of-fit criterion. If its value is close to one, then the match is—on average—consistent with the expected one, which was previously expressed through the prior covariance matrix $\mathbf{C}_{zz}$. A value slightly less than one was achieved in this sample problem, indicating that the assumption about the measurement errors (i.e., 10% of the water potential measurement) were too pessimistic. This is taken into account in the subsequent error analysis by using s_0^2 as a scaling factor for the covariance matrices.

An uncertainty measure of the estimated parameter values is usually obtained under the assumption of normality and linearity. The normality assumption is based on the fact that the distribution of a sum of random values always tends to normal if the sample size is sufficiently large. The linearity assumption postulates that the model output can be approximated by a linear function of the parameters within the area covered by the confidence region. Both assumptions have to be questioned because the sample size is usually small and the two-phase flow model is highly nonlinear. A linear approximation of estimation uncertainty is given by the covariance matrix $\mathbf{C}_{pp}$:

$$\mathbf{C}_{pp} = s_0^2\left(\mathbf{J}^T\mathbf{C}_{zz}^{-1}\mathbf{J}\right)^{-1} \tag{7.3.1}$$

```
> COMPUTATION

  >> STOPPING criteria
     >>> number of ITERATIONS      : 6
     >>> maximum scaled STEP size : 0.5
     >>> ignore WARNING messages
     <<<

  >> JACOBIAN matrix
     >>> PERTURB parameters by                      : 1 %
     >>> use FORWARD finite differences for first : 5 iterations
         then switch to centered finite difference quotient
     <<<

  >> ERROR analysis
     >>> risk ALPHA is                  :  5 %
     >>> check LINEARITY assumption on : 95 % confidence level
     <<<

  >> OUTPUT
     >>> time units for output is DAYS
     >>> write plot file with CHARACTERISTIC curves
     <<<
  <<
<
```

Figure 7.3.5. Excerpt from ITOUGH2 input file *sam6i*, block `COMPUTATION`.

The Jacobian $\mathbf{J}$ in Eq. (7.3.1) contains the sensitivity coefficients $J_{ij} = \partial z_i / \partial p_j$, evaluated prior to performing the last minimization step. While forward finite differences are used during the first 4 iterations, centered finite differences are calculated for the last two iterations to increase the accuracy of $\mathbf{J}$. The corresponding ITOUGH2 command is shown in block `>> JACOBIAN` of Figure 7.3.5. The parameters are perturbed by 1% for calculating the derivatives numerically.

The covariance and correlation matrices from the linear error analysis are shown in Table 7.3.2. The diagonal elements of matrix $\mathbf{C}_{pp}$ are the variances from the joint probability density function. Since the parameters are correlated, the uncertainty of one parameter affects the uncertainty of the other parameters. The conditional standard deviations $\sigma_p{}^*$, on the other hand, measure the uncertainty of a parameter assuming that all the other parameters are either exactly known or uncorrelated. The conditional standard deviations are shown in Table 7.3.3. They are always smaller than the joint standard deviations σ_p. Therefore, the ratio

$$\chi_k = \frac{\sigma_{p_k}{}^*}{\sigma_{p_k}} \tag{7.3.2}$$

can be regarded as a measure of overall parameter correlation, i.e., of how independently the k-th parameter can be estimated. Small values of χ_k indicate that the uncertainty σ_{p_k} of a parameter is mainly determined by its correlation to other uncertain parameters. A χ_k-value close to one indicates an independent estimate. An experiment should be designed towards small χ_k-values since weak correlations result in lower estimation uncertainties.

Next, we discuss correlation coefficients calculated from the covariance matrix $\mathbf{C}_{pp}$. Correlations among parameters can be viewed as the combined impact of parameter changes on the system behavior. For example, if two parameters are negatively correlated, a similar system response is obtained by concurrently increasing one and decreasing the other parameter. Even though certain pairs of parameters may exhibit preferential correlation structures, correlations are not invariable for a given parameter combination. They depend on the data used for the inversion and also on the number of simultaneously estimated parameters, since indirect correlations may overwhelm the direct correlations.

Table 7.3.2. Covariance (Diagonal and Lower Triangle) and Correlation (Upper Triangle) Matrices From Linear Error Analysis

	log (k [m^2])	n [-]	log($1/\alpha$ [Pa])
log (k [m^2])	6.72×10^{-4}	-0.42	-0.80
n [-]	-3.17×10^{-4}	8.56×10^{-4}	0.02
log($1/\alpha$ [Pa])	-9.70×10^{-5}	2.51×10^{-6}	2.21×10^{-5}

Table 7.3.3. Joint and Conditional Standard Deviations

	Standard Deviation		
	Joint σ_p	Conditional σ_p *	$\chi = \sigma_p{}^*/\sigma_p$
log (k [m^2])	2.59×10^{-2}	1.17×10^{-2}	0.45
n [-]	2.93×10^{-2}	2.18×10^{-2}	0.75
log($1/\alpha$ [Pa])	4.70×10^{-3}	2.33×10^{-3}	0.50

Table 7.3.4. Matrix of Direct (Lower Triangle) and Total (Upper Triangle) Correlation Coefficients

	log (k [m^2])	n [-]	log($1/\alpha$ [Pa])
log (k [m^2])	1.00	-0.42	-0.80
n [-]	-0.67	1.00	0.02
log($1/\alpha$ [Pa])	-0.87	-0.57	1.00

Table 7.3.5. Eigenanalysis of Covariance Matrix

Eigenvalue	Scaled Eigenvalue	Eigenvector Components		
		log (k [m^2])	n [-]	log($1/\alpha$ [Pa])
4.48×10^{-4}	1.55×10^{-4}	0.777	-0.605	0.172
1.10×10^{-3}	1.22×10^{-2}	0.605	0.794	0.061
5.27×10^{-6}	1.46×10^{-5}	-0.174	0.056	0.983

The direct correlation coefficients shown in the lower triangle of Table 7.3.4 can be interpreted more easily based on a physical understanding of the system. For example, since decreasing the value for n reduces the liquid relative permeability, the absolute permeability has to be increased in order to maintain a certain water flow rate. This explains why n and log(k) are negatively correlated. Similarly, the water potentials decrease with higher air entry pressure and higher permeability, leading to a negative correlation between these two parameters. Recall, however, that the correlation coefficients shown in the upper triangle of Tables 7.3.4 and 7.3.2 are the ones describing the statistical correlations among the parameters that are concurrently estimated in an inversion. These correlation coefficients are difficult to assess physically because they include contributions from indirect correlations. Indirect correlations have been previously described (see discussion of Table 3.4.2).

An eigenanalysis of the covariance matrix $\mathbf{C}_{pp}$ is performed (see Table 7.3.5). Parameter combinations along the eigenvector with the largest eigenvalue lead to a match similar to the best fit, whereas a perturbation of similar size from the best-estimate parameter set along the shortest eigenvector yields a significantly poorer agreement with the data. This means that the inverse modeling result is well-constrained in the direction of the shortest eigenvector. Since the parameters considered in the analysis may have different units and magnitudes, the eigenvalues and conditioning number are scaled by the estimated parameter value to allow comparison.

The shape and convexity of the objective function near the minimum determines the accuracy of the estimates. The covariance matrix $\mathbf{C}_{pp}$ approximates the actual surface of the objective function at its minimum by a tangent hyperellipsoid under the assumption of normality and linearity. If the model is nonlinear, the coverage of the confidence region by the linear approximation may be very poor with respect to both its size and its shape. ITOUGH2 offers a correction procedure based on the assumption that the shape of the confidence region is close to ellipsoidal, and that the orientation of the hyperellipsoid in the n-dimensional parameter space is accurately obtained from the linear error analysis. Then, by adjusting only the size of the hyperellipsoid, we can better approximate the confidence region without losing the advantage of producing easily understandable results that are also simple to report. The basic idea of the procedure is to compare the actual likelihood function with the results from the linear approximation at discrete points in the parameter space. These test points are preferably located along the main axis of the hyperellipsoid, that is:

$$\tilde{\mathbf{p}}_{\pm i} = \hat{\mathbf{p}} \pm \sqrt{n \cdot F_{n,m-n,1-\alpha}} \cdot a_i \cdot \mathbf{u}_i \qquad (i = 1 \ldots n) \qquad (7.3.3)$$

Here, $\tilde{\mathbf{p}}_{\pm i}$ are two test parameter sets on the i-th axis, the direction of which is given by the eigenvector $\mathbf{u}_i$ of the covariance matrix $\mathbf{C}_{pp}$. Note that the distance from the optimal parameter set $\hat{\mathbf{p}}$ is selected as a multiple of the corresponding eigenvalue a_i^2 and the quantile of the F-distribution. This means that the correction is tailored to approximate the confidence region on a certain confidence level $1-\alpha$. The eigenvalues a_i^2 representing the length of the semiaxis are now corrected as follows:

$$a'^2_i = a_i^2 \cdot s_0^2 \left(\frac{A_{+i} + A_{-i}}{2} \right) \qquad (7.3.4)$$

with

$$A_{\pm i} = \frac{n \cdot F_{n,m-n,1-\alpha}}{S(\tilde{\mathbf{p}}_{\pm i}) - S(\hat{\mathbf{p}})} \qquad (7.3.5)$$

Here, $S(\mathbf{p})$ is the value of the objective function obtained with parameter vector $\mathbf{p}$. Finally, the new covariance matrix is back-calculated from the eigenvectors $\mathbf{u}_i$ and the updated eigenvalues a'^2_i. The correction procedure requires $2n$ additional solutions of the forward problem and is thus relatively inexpensive. While the resulting confidence region is ellipsoidal by definition, the differences between $S(\tilde{\mathbf{p}}_{+i})$ and $S(\tilde{\mathbf{p}}_{-i})$ provide, as a byproduct of the correction procedure, some insight into the asymmetry of the true confidence region.

The ITOUGH2 command line invoking the correction procedure can be found in Figure 7.3.5. Figure 7.3.6 shows the ITOUGH2 output from the correction along the first eigenvector. For $n = 3$ parameters, $m = 165$ observations, and $\alpha = 0.05$, the value of the F-quantile is about 2.6. Perturbing the best estimate parameter set in both directions along the first eigenvector (see Table 7.3.5) yields the two test points $\tilde{\mathbf{p}}_{+1}$ and $\tilde{\mathbf{p}}_{-1}$. The respective values of the objective function, $S(\tilde{\mathbf{p}}_{+1})$ and $S(\tilde{\mathbf{p}}_{-1})$, are evaluated. They compare well with each other as well as with the value from the linear approximation, which is calculated to be [*Donaldson and Schnabel*, 1987]

$$S(\tilde{\mathbf{p}}_{lin}) = S(\hat{\mathbf{p}}) + s_0^2 \cdot n \cdot F_{n,m-n,1-\alpha} \qquad (7.3.6)$$

From this close agreement it can be concluded that neglecting non-linearity effects in the error analysis is justified in this case. Since the average of $S(\tilde{\mathbf{p}}_{+1})$ and $S(\tilde{\mathbf{p}}_{-1})$ is greater than $S(\tilde{\mathbf{p}}_{lin})$, the eigenvalue is slightly reduced to yield an elliptical region that better represents the true confidence region. The same procedure is applied to all eigenvectors, and the error analysis is repeated for the corrected covariance matrix.

```
Check Linearity Assumption of Error Analysis
--------------------------------------------

Correction along semiaxis No. 1

Eigenvalue corresponding to parameter      : Permeability
Increment of standard deviation k=sqrt(n*F):    0.279149E+01

Step along semiaxis No. 1 leads to new parameter sets:
       best estimate                 step    parameter set #1    parameter set #2
     -0.18552705E+02 +/-  0.45903710E-01     -1.85068015E+01     -1.85986089E+01
      2.45870490E+00 +/-  3.57112260E-02      2.49441613E+00      2.42299367E+00
      6.23540557E+00 +/- -1.02542729E-02      6.22515130E+00      6.24565985E+00

Objective function at minimum, Xmin        :    0.123573E+03
Linear approximation = Xmin + s0^2*k^2     :    0.129517E+03
O.f. at parameter set #1   (numerically)   :    0.130883E+03
O.f. at parameter set #2   (numerically)   :    0.130069E+03
Original eigenvalue                        :    0.447564E-03
Corrected eigenvalue                       :    0.386746E-03
```

Figure 7.3.6. Excerpt from ITOUGH2 output file *sam6i.out* showing correction procedure along first eigenvector of parameter covariance matrix.

8. PROBLEM 7: NUMERICAL DISPERSION

8.1 Capillary Diffusion

In Problem 7, ITOUGH2 is used to examine numerical dispersion effects. This investigation can be considered to be a verification study for both TOUGH2 and ITOUGH2. Furthermore, it demonstrates that numerical dispersion can be partly compensated for by estimating model-related parameters of the capillary pressure function.

Downward flow of liquid in a partially saturated porous medium is calculated numerically and compared with an analytical solution in order to examine possible errors as a result of finite space discretization. The governing equation for unsaturated liquid flow can be written in the form of a convection-diffusion equation for saturation, termed Fokker-Planck equation [*Philip*, 1969], as follows:

$$\frac{\partial S_l}{\partial t} = v\frac{\partial S_l}{\partial z} + div\left(\mathbf{D}_{cap}\nabla S_l\right) \tag{8.1.1}$$

where S_l is liquid saturation, t is time, z is a positive downward space coordinate, and $\mathbf{D}_{cap}$ is a tensor expressing capillary effects. The velocity v for a vertical, gravity-driven propagation of saturation disturbances in the absence of capillary effects can be given by [*Pruess*, 1996]:

$$v = \frac{k \cdot \rho_l \cdot g}{\phi \cdot \mu_l}\frac{\mathrm{d}k_{rl}}{\mathrm{d}S_l} \tag{8.1.2}$$

Here, k is absolute permeability, k_{rl} is liquid relative permeability as a function of liquid saturation S_l, ϕ is porosity, ρ_l and μ_l are density and viscosity of the liquid phase, respectively, and g is gravitational acceleration. A diffusion coefficient expressing capillary effects can be written as follows [*Pruess*, 1996]:

$$D_{cap} = \frac{k \cdot k_{rl}}{\phi \cdot \mu_l}\frac{\mathrm{d}P_{cap}}{\mathrm{d}S_l} \tag{8.1.3}$$

It is seen that in order to obtain a constant velocity and constant diffusivity we must have

$$k_{rl} = c \cdot S_l \tag{8.1.4}$$

$$P_{cap} = A \cdot \ln(k_{rl}) \tag{8.1.5}$$

where A is a coefficient that scales the capillary pressure. Physical constraints require the constant c in Equation (8.1.4) to be one. Equations (8.1.2) and (8.1.4) yield the advective velocity:

$$v = \frac{k \cdot \rho_l \cdot g}{\phi \cdot \mu_l} \tag{8.1.6}$$

and capillary diffusivity is derived from Eqs. (8.1.3), (8.1.4), and (8.1.5) to be:

$$D_{cap} = \frac{A \cdot k}{\phi \cdot \mu_l} \tag{8.1.7}$$

For the characteristic curves given by (8.1.4) and (8.1.5), the coefficients in (8.1.1) are constant which allows applying standard analytical solutions to the convection-diffusion equation for miscible displacement [*Scheidegger*, 1954]:

$$S_l(z,t) = \frac{M(t)}{\sqrt{4\pi D_{cap} t}} \cdot \exp\left[-\frac{(z-vt)^2}{4D_{cap}t}\right] \tag{8.1.8}$$

with

$$M(t) = \int_{-\infty}^{\infty} \left(S_l(z,t) - S_{lr}\right) \mathrm{d}z \tag{8.1.9}$$

where S_{lr} is the residual liquid saturation. Boundary conditions for the solution of (8.1.8) correspond to $S_l \rightarrow S_{lr}$ for $z \rightarrow \infty$, and the initial disturbance is given by a δ-function.

Gravity-driven downflow of liquid under capillary forces is modeled using TOUGH2 in combination with EOS9. A uniform grid spacing of $h = 0.01$ m is chosen. The permeability is $k = 10^{-13}$ m^2, porosity $\phi = 0.1$, residual liquid saturation $S_{lr} = 0$, and the parameter A in (8.1.5) is 1000 Pa. The system is initially dry, except for one gridblock initialized with a liquid saturation of $S_l(t = 0) = 0.9999$. The TOUGH2 input file is shown in Figure 8.1.1. Figure 8.1.2 shows liquid saturation as a function of time at a distance of 0.1, 0.2, 0.3, and 0.5 m from the injection point. The TOUGH2 results (solid lines) are in very good agreement with the analytical solution (symbols). Note that no special effort has been made to accurately model the input δ-function, explaining the somewhat larger spreading of the saturation pulse as it moves through the system. This simulation proves that the TOUGH2 solution is accurate for sufficiently fine space and time discretization.

```
sam7: Examines numerical diffusion
ROCKS----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
MATRI    2                 0.1                    1.000E-13

    1      0.000E+00           1.000E+00
    8      1.000E+02

PARAM----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
 8-11200    12001000001000000000400000000
           1.080E+05 1.000E-01 1.000E+02               9.810
 1.000E-06
            0.000001
MESHM----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
XYZ

NX       1 1.000E+00
NY       1 1.000E+00
NZ     200 0.010E+00


START----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
INCON----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
BF1 1
              0.9999

ENDCY----1----*----2----*----3----*----4----*----5----*----6----*----7----*----8
```

Figure 8.1.1. TOUGH2 input file *sam7*.

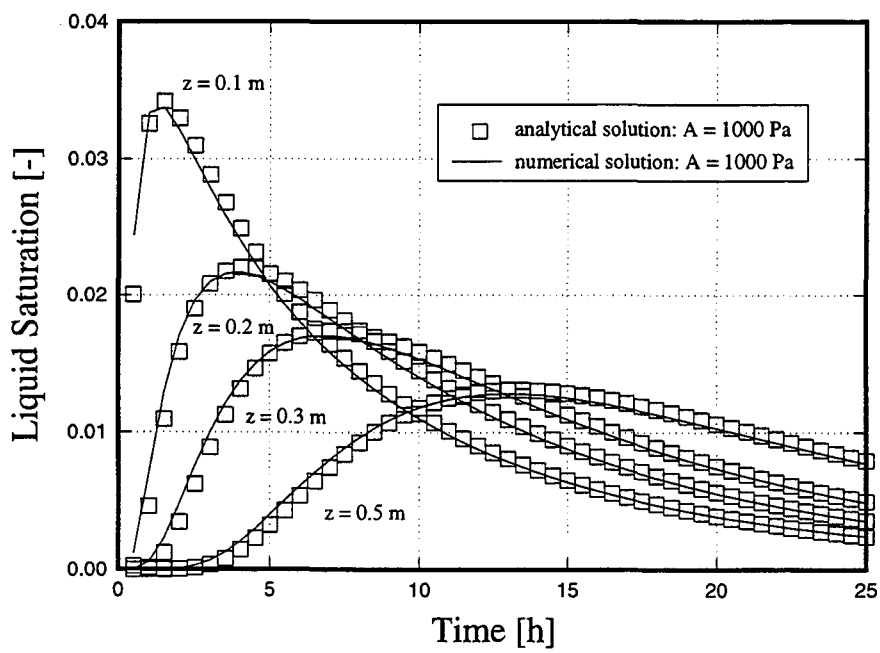


Figure 8.1.2. Verification of TOUGH2 solution. Only minor discretization effects are seen for grid spacing $h = 0.01$ m, maximum time step $\Delta t_{max} = 100$ seconds, and $A = 1000$ Pa.

8.2 Compensating Numerical Dispersion

Space discretization gives rise to artificial numerical dispersion effects. An effective diffusivity due to space discretization is given by [*Pruess*, 1991b]:

$$D_{grid} = \frac{(h/2) \cdot k \cdot \rho_l \cdot g}{\phi \cdot \mu_l} \tag{8.2.1}$$

The total diffusivity in the simulation is therefore a sum of physical (capillary) and numerical (grid) diffusivities, $D_{tot} = D_{cap} + D_{grid}$. From Equations (8.1.7) and (8.2.1) we obtain:

$$A_{tot} = A + A_{grid} = A + (h/2) \cdot \rho_l \cdot g \tag{8.2.2}$$

In our case, the second term on the right hand side, reflecting an effective capillary strength due to space discretization, is 49 Pa. Space discretization effects can be studied either by actually varying the side length of the gridblocks or by changing parameter A in (8.2.2). The smaller the capillary strength A, the more A_{tot} is influenced by space discretization effects.

Let's assume that the true value for A is known to be 100 Pa. If this value were used in a simulation of a liquid pulse propagating through an unsaturated porous medium, the peak liquid saturation would be underpredicted at any point z because numerical diffusion leads to additional spreading of the saturation pulse. If one estimates parameter A based on saturation measurements, an effective value for A will be obtained that partly compensates for discretization errors.

To demonstrate this, we perform an inversion where the data to be matched are obtained from the analytical solution (Equation 8.1.8). While the analytical solution could also be calculated beforehand and prescribed as a list of discrete data points, internal calculation is more convenient and flexible. A user-specified function providing data points is programmed into subroutine USERDATA (Figure 8.2.1). Note that all values needed to evaluate (8.1.8) are directly taken from the input files and the secondary variables calculated by TOUGH2. The ITOUGH2 input file is shown in Figure 8.2.2. Parameter No. 1 of the capillary pressure function is estimated based on analytically calculated saturation data. The annotation "`ANALYT. Z=x.xxx`" is used to identify the data set and to transfer the z-coordinate to subroutine USERDATA.

```
************************************************************************
      SUBROUTINE USERDATA(IDF,TIME,ANNO,F)
************************************************************************
* User specified function to represent observed data                   *
* IDF : data set identifier (input)                                    *
* TIME: time at which data are to be provided (input)                  *
* ANNO: annotation (input)                                             *
* F   : value of observed data at time TIME (output)                   *
************************************************************************
C
C$$$$$$$$$ PARAMETERS FOR SPECIFYING THE MAXIMUM PROBLEM SIZE $$$$$$$$$$
      INCLUDE 'maxsize.inc'
C
C$$$$$$$$$ COMMON BLOCKS FOR SIMULATION PARAMETERS $$$$$$$$$$$$$$$$$$$$$
      INCLUDE 'param.inc'
C
C$$$$$$$$$ COMMON BLOCKS FOR ROCK PROPERTIES $$$$$$$$$$$$$$$$$$$$$$$$$$$
      INCLUDE 'rock.inc'
C
C$$$$$$$$$ COMMON BLOCK FOR SECONDARY VARIABLES $$$$$$$$$$$$$$$$$$$$$$$$
      INCLUDE 'second.inc'
C
C$$$$$$$$$ COMMON BLOCKS FOR TOTAL MASS AND VOLUMES $$$$$$$$$$$$$$$$$$$$
      INCLUDE 'rmasvol.inc'
C
C$$$$$$$$$ COMMON BLOCKS FOR ELEMENTS $$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
      INCLUDE 'elements.inc'

      CHARACTER ANNO*(*)

      SAVE ICALL,A

      DATA ICALL/0/

      ICALL=ICALL+1
      IF (ICALL.EQ.1) WRITE(11,7999)
 7999 FORMAT(6X,'USERDATA 2.5       2 SEPTEMBER 1996',6X,
     &       '#125: USER SPECIFIED DATA DESCRIPTION')

      F=0.0D0
C --- For sample problem 7, analytical solution 1D-pulse
      IF (ICALL.EQ.1) A=CP(1,1)
      IF (ANNO(1:10).EQ.'ANALYT. Z=') THEN
         READ(ANNO(11:15),'(F5.3)',IOSTAT=IOS) Z
         V=PER(3,1)*PAR(4)*GF/POR(1)/PAR(3)
         DIFF=A*PER(3,1)/POR(1)/PAR(3)
         XM=XPVOLU0(2)/POR(1)
         F=XM/(DSQRT(4.0D0*3.1416D0*DIFF*TIME))*
     &     DEXP(-(Z-V*TIME)**2/(4.0D0*DIFF*TIME))
      ENDIF
      END
C --- END of USERDATA
```

Figure 8.2.1. Subroutine USERDATA. Liquid saturation is calculated analytically as a function of space and time, and provided as the data points to be matched.

```
--- Sample Problem 7 --- Sample Problem 7 --- Sample Problem 7

 > PARAMETER

   >> parameter 1 of CAPILLARY pressure function
      >>> ROCK type          : MATRI
          >>>> ANNOTATION    : A in A*ln(krl)
          >>>> PARAMETER CP(: 1 )
          >>>> VALUE
          <<<<
      <<<
   <<

 > OBSERVATIONS

   >> Calibrate at: 50 EQUALLY spaced TIMEs between [HOUR]
      0.5 25.0

   >> LIQUID SATURATION
      >>> GRID BLOCK       : BP1_1
          >>>> ANNOTATION : ANALYT. Z=0.100
          >>>> DEVIATION  : 0.002
          >>>> analytical solution given by USER-specified function
          <<<<

      >>> GRID BLOCK       : BZ1_1
          >>>> ANNOTATION : ANALYT. Z=0.200
          >>>> DEVIATION  : 0.002
          >>>> analytical solution given by USER-specified function
          <<<<

      >>> GRID BLOCK       : CA1_1
          >>>> ANNOTATION : ANALYT. Z=0.300
          >>>> DEVIATION  : 0.002
          >>>> analytical solution given by USER-specified function
          <<<<

      >>> GRID BLOCK       : CU1_1
          >>>> ANNOTATION : ANALYT. Z=0.500
          >>>> DEVIATION  : 0.002
          >>>> analytical solution given by USER-specified function
          <<<<
      <<<
   <<


 > COMPUTATION

   >> STOP after
      >>> : 5 ITERATIONS
      <<<
   <<
 <
```

Figure 8.2.2. ITOUGH2 input file *sam7i*.

The dashed lines in Figure 8.2.3 show the simulated saturations for A = 100 Pa, which gives rise to substantial space discretization effects as high as 50% according to (8.2.2). The solid lines represent the fit to the analytical solution. The optimum, model-related value for A is estimated to be 48 Pa, i.e., discretization effects are compensated for by reducing parameter A from 100 Pa to 48 Pa. The estimate of 48 Pa is in good agreement with the theoretical result from Equation (8.2.2), i.e., $A = A_{tot} - A_{grid} = 100 - 49 = 51$ Pa. The difference of 3 Pa can be attributed to time-discretization effects and inaccuracy in modeling the initial saturation pulse.

This sample problem provides verification for TOUGH2 (see Figure 8.1.1), and ITOUGH2 (the estimated value is consistent with the one calculated by Equation (8.2.2)). In addition, it demonstrates that discretization errors can be partly compensated for by using a model-related parameter—estimated by inverse modeling—rather than the true value. This last remark is only valid, of course, if the same or similar discretization is used for both the inversion and the subsequent prediction run. Moreover, discretization errors are less accurately compensated for if the characteristic curves deviate from the model (see Equations (8.1.4) and (8.1.5)) used in this study.

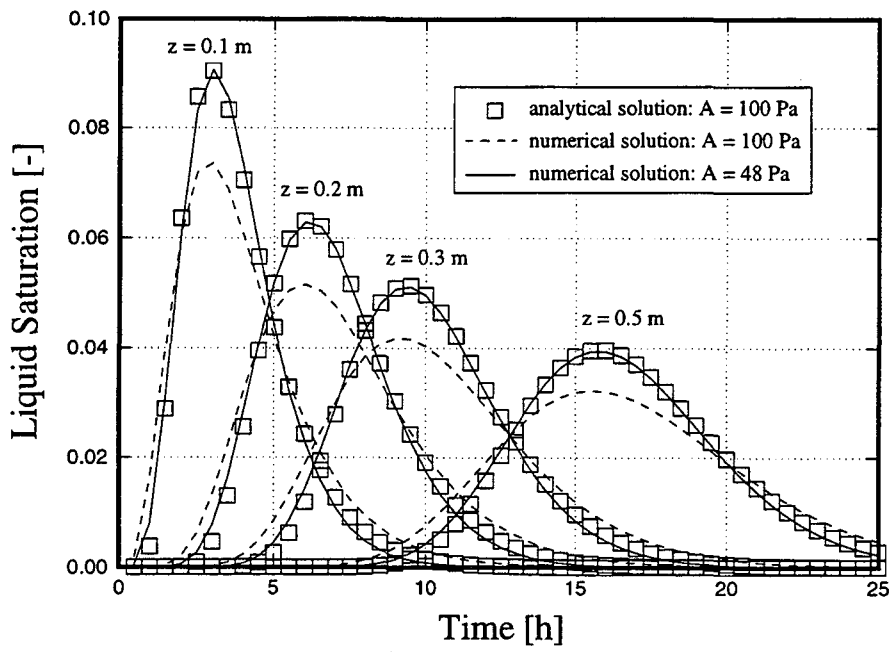


Figure 8.2.3. Discretization effects are compensated for by reducing parameter A from 100 Pa to 48 Pa.

ACKNOWLEDGMENT

ITOUGH2 (up to version 1.1) was developed at the Laboratory of Hydraulics, Hydrology, and Glaciology (VAW) of the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, in collaboration with the Swiss National Cooperative for the Disposal of Radioactive Waste (Nagra), Wettingen, Switzerland. Subsequent versions were supported, in part, by the Director, Office of Civilian Radioactive Waste Management, U.S. Department of Energy, through Memorandum Purchase Order EA9013MC5X between TWR Environmental Safety Systems, Inc. and the Ernest Orlando Lawrence Berkeley National Laboratory, by a grant from the Swiss National Cooperative for the Disposal of Radioactive Waste (Nagra), Wettingen, Switzerland, and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Geothermal Technologies, of the U.S. Department of Energy, under Contract No. DE-AC03-76SF00098.

I would like to acknowledge the people and institutions that provided the laboratory or field data used in the sample problems. P. Persoff (LBNL) performed the gas-pressure-pulse-decay experiments described in Problem 2; the multistep outflow experiment (Problem 4) was conducted by B. Faybishenko (LBNL/UCB); pneumatic pressures, saturations and water potentials used in Problem 5 were collected by various personnel of the Yucca Mountain Project; the ventilation experiment at the Grimsel Rock Laboratory was conducted by T. Gimmi (ETH, Zürich, Switzerland) in collaboration with Nagra (Wettingen, Switzerland).

I would like to thank C. Doughty and A. Mishra for a careful review of the manuscript.

REFERENCES

Brooks, R. H., and A. T. Corey, Hydraulic properties of porous media, Hydrol. Pap., Colorado State Univ., Fort Collins, 3, 1-27, 1964.

Carrera, J. and S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, *22*(2), 199-210, 1986.

Donaldson, J. R., and R. B. Schnabel, Computational experience with confidence regions and confidence intervals for nonlinear least squares, *Technometrics*, *29*(1), 67-82, 1987.

Edlefsen, N. E. and A. B. C. Anderson, Thermodynamics of soil moisture, *Hilgardia*, *15*(2), 31-298, 1943.

Finsterle, S., ITOUGH2 user's guide, Version 3.1, Report LBNL-40400, Lawrence Berkeley National Laboratory, Berkeley, CA , 1997a.

Finsterle, S., ITOUGH2 command reference, Version 3.1, Report LBNL-40401, Lawrence Berkeley National Laboratory, Berkeley, CA , 1997b.

Finsterle, S., and K. Pruess, Solving the estimation-identification problem in two-phase flow modeling, *Water Resour. Res.*, *31* (4), 913-924, 1995.

Finsterle, S., and P. Persoff, Determining permeability of tight rock samples using inverse modeling, *Water Resour. Res.*, *33*(8), 1803-1811, 1997.

Finsterle, S., and B. A. Faybishenko, Design and analysis of an experiment to determine hydraulic parameters of variably saturated porous media, Report LBNL-40245, Lawrence Berkeley National Laboratory, Berkeley, CA , 1997.

Gimmi, T., M. Schneebeli, H. Flühler, H. Wydler, and T. Baer, Field-scale water transport in unsaturated crystalline rock, *Water Resour. Res.*, *33*(4), 589-598, 1997.

Hulen, J. B., B. A. Koenig, and D. L. Nielson, The Geysers Coring Project, Sonoma County, California - Summary and initial results, Proceedings of World Geothermal Congress, Florence, Italy, 1415-1420, 1995.

Klinkenberg, L. J., The permeability of porous media to liquids and gases, *API Drill. and Prod. Prac.*, 200-213, 1941.

Morel-Seytoux, H. J., P. D. Meyer, M. Nachabe, J. Touma, M. Th. van Genuchten, and R. J. Lenhard, Parameter equivalence for the Brooks-Corey and van Genuchten soil characteristics: Preserving the effective capillary drive, *Water Resour. Res.*, *32*(5), 1251-1258, 1996.

Philip, J. R., Theory of infiltration, *Advances in Hydroscience*, 5, 215-296, 1969.

Pruess, K., TOUGH user's guide, Report NUREG/CR-4645, Nuclear Regulatory Commission (also Report LBL-20700, Lawrence Berkeley Laboratory, Berkeley, CA), 1987.

Pruess, K., TOUGH2 - A general-purpose numerical simulator for multiphase fluid and heat flow, Report LBL-29400, Lawrence Berkeley Laboratory, Berkeley, CA, 1991a.

Pruess, K., Grid orientation and capillary pressure effects in the simulation of water injection into depleted vapor zones, *Geothermics*, *20*(5/6), 257-277, 1991b.

Pruess, K., A Fickian diffusion model for the spreading of liquid plumes infiltrating in heterogeneous media, *Transport in Porous Media*, *24*, 1-33, 1996.

Pruess, K. and Y.-S. Wu, A new semi-analytical method for numerical simulation of fluid and heat flow in fractured reservoirs, *SPE Advanced Technology Series*, *1*(2), 63-72, Richardson, Texas, 1993.

Scheidegger, A. E., Statistical hydrodynamics in porous media, *J. Appl. Phys.* *25*(8), 894-1001, 1954.

van Genuchten, M. Th., A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, *44*, 892-898, 1980.

INDEX

A
annotation, 8
A-optimality, 38

B
bias, 35, 38, 57
boundary condition
  estimating, 8, 12,35
  time-dependent, 16, 48, 55

C
COLUMN, 11, 28
compromise, 30
confidence
  level, 18, 69
  region, 74
correlation, 31, 32, 42, 72
  direct/indirect, 38, 73
covariance matrix
  estimated parameters, 31, 70, 73
  measurement errors, 33, 70

D
dispersion, numerical, 79

E
error, systematic, 33, 35, 57, 67
error variance
  a posteriori, 13, 38, 70
  a priori, 13
eigenanalysis, 31, 73
equation-of-state, 1

F
Fisher
  information matrix, 32
forward problem, 5
  FORWARD, 5
FOSM, 16, 42

G
Geysers, 21
Grimsel, 64
goodness-of-fit, 38, 70
GUESS, 66

H
HEADER, 11, 26

I
initial conditions
  estimating, 8, 12, 35
  specifying, 22, 54
inverse problem
  ill-posed, 24
  well-posed, 32
inversion
  simultaneous, 32
invdir, 5
itough2, 5, 13, 55

J
Jacobian, 28, 71

K
Kashyap, 38
KDATA, 6
kit, 14
Klinkenberg effect, 21

L
log-normal distribution,16

M
match
  good, 32
  perfect, 14
  reasonable, 33
MINC, 40
minimization algorithm, 49
minimum, 55, 66
Monte Carlo, 16

N

NO DATA, 6
non-linearity
  error analysis, 73
  minimization algorithm, 50
  sensitivity analysis, 8, 30, 41
  uncertainty propagation, 16, 19

non-unique, 24, 30, 49, 66
normal distribution, 16, 70

O

outlier, 69
output
  file, 5, 14
overparameterization, 35

P

parameter
  estimation, 11
  range, 11, 18
  selection (automatic), 40
  variation, 28
PERTURB, 18, 59
plot
  file, 14
  format, 6, 13
  residual, 67
primary variable, 8, 55
prior information, 28, 53, 57
prista, 13

Q

quantile, 74

R

regression analysis, 69
residual, 33, 63, 69
RESTART, 16, 48

S

seed number, 18
SELECTION, 41
sensitivity
  analysis, 8, 30
  coefficients (scaled), 28
  measures, 10, 42, 43, 62
SET, 28
shut-in, 16
standard deviation
  conditional, 32, 71
  joint, 32, 71
  parameter, 16, 28
  observation, 13, 28, 39, 66
  residual, 33, 43, 63
steady state, 55

T

tough2, 5

U

uncertainty
  estimated parameter, 30, 38
  prediction, 16, 43, 69
unique (see non-unique)
USERBC, 55
USERDATA, 79

V

van Genuchten, 47
VARIATION, 8, 28, 53
ventilation, 64

W

warning messages, 57, 66
  ignore, 11, 13, 28
  message files, 5
WINDOW, 58, 66

X

X(1), 55

Y

Yucca Mountain, 83

Z

Z=x.xxx, 79

Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720