

PROCESSING AND INTERPRETATION OF PRESSURE TRANSIENT DATA
FROM PERMANENT DOWNHOLE GAUGES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ENERGY RESOURCES
ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Masahiko Nomura
September 2006

© Copyright by Masahiko Nomura 2007
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Dr. Roland N. Horne) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Dr. Anthony R. Kovsky)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

(Dr. Hamdi Tchelepi)

Approved for the University Committee on Graduate Studies.

Abstract

Reservoir pressure has always been the most useful type of data to obtain reservoir parameters, monitor reservoir conditions, develop recovery schemes, and forecast future well and reservoir performances. Since 1990, many wells have been equipped with permanent downhole gauges (PDG) to monitor the well performance in real time. This continuous monitoring enables engineers to observe ongoing changes in the well and to make operating adjustments to optimize recovery. The long-term pressure record is also useful for parameter estimation, since the common pressure transient tests such as drawdown or build-up tests are conducted during relatively short periods.

However, the PDG data have several complexities, since the pressure data are not measured under well-designed conditions as in the conventional well testing schemes. The pressure data may have various types of noise and may exhibit aberrant behavior that is inconsistent with the corresponding flow rate. The flow rate may change irregularly, since the flow rate is not controlled based on a designed scenario.

This study investigated methods to analyze the long term pressure data acquired from permanent downhole gauges. The study addressed both the data processing and the parameter estimation problem with the connection to the physical behavior of the reservoir. The methods enabled us to address specific issues: (1) model identification, (2) flow rate estimation, (3) transient identification, and (4) data smoothing.

Acknowledgements

This dissertation leaves me greatly indebted to Professor Roland N. Horne, my principal advisor, for his advice, guidance, and encouragement during the course of this work.

My sincere thanks are also due to Professor Anthony R. Kovscek and Hamdi Tchelepi, who served on the reading committee and Professor Yinyu Ye and Louis J. Durlofsky, who participated in the examination committee.

I appreciate Professor Trevor Hastie and James Ramsay for sharing their insights and knowledge to help broaden my understanding of data smoothing, nonparametric regression, and statistical inference.

I am also thankful for all my friends in the Department of Energy Resources Engineering and Stanford University. They are more helpful than they realize.

Financial support for this work was provided by Teikoku Oil Company (TOC) and Japan Oil, Gas, and Metals National Corporation (JOGMEC). My special gratitude goes to my colleagues in TOC for helping me to get academic leave.

Contents

Abstract	v
Acknowledgements	vi
1 Introduction	1
1.1 Background	1
1.2 Previous work	3
1.3 Problem statement	7
1.4 Dissertation outline	7
2 Review of data smoothing techniques	9
2.1 Single variable smoothing technique	10
2.1.1 Parametric regression	10
2.1.2 Preliminary definition of smoother and operator	12
2.1.3 Local smoother	13
2.1.4 Regression spline	15
2.1.5 Smoothing spline	16
2.1.6 Nonlinear smoother	19
2.1.7 Wavelet smoothing	19
2.2 Model assessment and selection	21
2.2.1 Degrees of freedom	21
2.2.2 Prediction error estimate	23
2.3 Multivariate data smoothing technique	27
2.3.1 ACE algorithm	27
2.3.2 Some theoretical aspects of ACE	34

2.4	Summary	37
3	Single transient data smoothing	39
3.1	Constrained smoother	39
3.1.1	Functional representation	41
3.1.2	Least square fitting	45
3.1.3	Knot insertion and preliminary check	50
3.1.4	Smoothing effect of derivative constraints	51
3.2	Smoothing control	57
3.3	Multiple smoothing parameter control	65
3.4	Summary	75
4	Multitransient data analysis	80
4.1	Model identification and flow rate recovery	81
4.2	Comparison with the existing method	89
4.3	Transient identification	98
4.3.1	Wavelet processing	98
4.3.2	Detection algorithm	106
4.4	Field application	125
4.5	Summary	129
5	Conclusions and Future work	134
	Nomenclature	142
	Bibliography	143

List of Tables

4.1	MSE of pressure derivative (psi^2) and flow rate estimation error.	96
4.2	The number of break points and GCV score (insertion scheme).	116
4.3	The number of break points and GCV score (after deletion scheme).	121

List of Figures

2.1	Projection onto column space of a design matrix.	11
2.2	Model complexities (degrees of freedom) and the corresponding prediction squared error (PSE) curve.	22
2.3	K -fold cross validation. (a) original data, (b) data extraction and model fitting to the remaining data, and (c) predict the value at the extracted data points.	26
2.4	Scatter plot of response variable Y and predictor variables X_1 , X_2 , and X_3	31
2.5	Transformations of response variable Y and predictor variables X_1 , X_2 , and X_3	32
2.6	Prediction quality of $\theta(Y)$ and Y	33
2.7	Hilbert space setting in ACE algorithm.	35
3.1	Hat function and its first and second integral.	42
3.2	Function representation with hat functions and its integrated function.	43
3.3	Functional representation. Upper: integrated function, Middle: shifted function, and Lower: flipped function.	46
3.4	Schematic of active set method.	48
3.5	Example fitting results with 0.3 log interval (sixth order derivative constraints). Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model. open circle: estimated derivative and solid line: true solution.	52
3.6	An example synthetic data for 2 % error case.	53
3.7	Pressure derivative estimates with higher order derivative constraints. Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model.	54

3.8	MSE, bias, and variance (psi^2) for pressure estimate with higher order derivative constraints. Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model.	55
3.9	MSE, bias, and variance (psi^2) for pressure derivative estimate with higher order derivative constraints. Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model.	56
3.10	Degrees of freedom of the smoother.	57
3.11	MSE of pressure derivative (psi^2) for various noise levels. Open marks: w/o derivative constraints.	58
3.12	Pressure derivative estimates for various noise levels.	59
3.13	Definition of roughness and its control points.	60
3.14	MSE, bias, and variance for various smoothing parameters (model: infinite-acting radial flow for noise level 2 %). Horizontal line shows the MSE value without smoothing control.	64
3.15	GCV score for various smoothing parameters.	65
3.16	The effect of smoothing parameter (infinite-acting radial flow model). Thin line: true solution and bold line: the estimate.	66
3.17	The effect of smoothing parameter (dual porosity model). Thin line: true solution and bold line: the estimate.	67
3.18	The effect of smoothing parameter (closed boundary). Thin line: true solution and bold line: the estimate.	68
3.19	MSE of pressure derivative (psi^2). open marks: w/o smoothing control. . .	69
3.20	Pressure derivative estimates for various noise levels. Solid line: true solution.	70
3.21	Local MSE, bias, and variance (psi^2) for various smoothing control parameter (infinite-acting radial flow model). Thin line: true solution and bold line: estimates.	71
3.22	Local MSE, bias, and variance (psi^2) for various smoothing control parameter (dual porosity model). Thin line: true solution and bold line: the estimate.	72
3.23	Local MSE, bias, and variance (psi^2) for various smoothing control parameter (closed boundary model). Thin line: true solution and bold line: the estimate.	73
3.24	The concept of multiple smoothing parameter control.	74
3.25	The estimated λ profile (knots are placed at every 10 control points).	76
3.26	Local MSE, bias, and variance (psi^2).	77

3.27	The estimated λ profile (knots are placed at every 2 control points).	78
3.28	Local MSE, bias, and variance (psi^2).	79
4.1	An example synthetic data with 1 % noise.	82
4.2	GCV score and degrees of freedom for the various smoothing parameters (pressure error 1%).	85
4.3	MSE of the pressure derivative (psi^2). Open circles: without smoothing control	86
4.4	GCV and degrees of freedom for various control parameters.	86
4.5	MSE of the pressure derivative (psi^2). Open circles: w/o regularization. . .	87
4.6	MSE of the pressure derivative (psi^2).	88
4.7	The error of the estimated flow rate.	89
4.8	The deconvolved response with the exact flow rate profile. Solid line: true solution.	90
4.9	The deconvolved response with 1 % flow rate error. Solid line: true solution.	91
4.10	The deconvolved response with 10 % flow rate error. Solid line: true solution.	92
4.11	The estimated flow rate profile for pressure noise 3 % and rate noise 10 % (infinite-acting radial flow).	93
4.12	Schematic of Schroeter's formulation.	95
4.13	The derivative plot for 1 % pressure error and 5 % flow rate error.	97
4.14	A synthetic pressure data and an expanded view (0.3 % noise).	99
4.15	Wavelet transformation of the original signal (1). Left: the approximated signal. Right: the detailed signal.	101
4.16	Wavelet transformation of the original signal (2). Left: the approximated signal. Right: the detailed signal.	102
4.17	Wavelet processing results (1). Upper: true break point locations. Lower: the detected break point locations with 2 psi threshold.	103
4.18	Wavelet processing results (2). Upper: the detected break point locations with 0.5 psi threshold. Lower: the detected break point locations with 0.01 psi threshold.	104
4.19	The expanded view of wavelet processing results with 0.5 psi threshold. . .	105
4.20	The deconvolved response using the break points with wavelet. Circle: w/ wavelet-detected break points and triangle: w/ adjusted break points. . . .	106

4.21	Pressure fitting result after the first iteration and the selected regions for break point insertion.	110
4.22	The expanded view of the pressure fitting results after first iteration.	111
4.23	Pressure fitting result after the second iteration and the selected regions for break point insertion.	112
4.24	The expanded view of the pressure fitting results after the second iteration.	113
4.25	GCV score during the break point insertion.	115
4.26	Final fitting results.	116
4.27	The estimated break point location. Upper: the estimated location of the break points. Lower: the location of the true break points	117
4.28	The expanded view of the estimated break point location. Open circles: true location.	118
4.29	GCV score during the whole sequence.	119
4.30	The estimated break point location. Upper: the estimated location of the break points. Lower: location of the true break points.	120
4.31	GCV score plotted versus the smoothing parameter in case 1.	122
4.32	The estimated pressure derivative	122
4.33	A procedure for the transient identification.	123
4.34	The procedure for the data analysis.	124
4.35	The original data and the wavelet processing results (field data 1).	125
4.36	The GCV score and the number of break points.	126
4.37	The final fitting results.	127
4.38	The GCV score for the smoothing parameters.	128
4.39	The derivative estimates with the selected smoothing parameter.	128
4.40	The original data and the wavelet processing results (field data 2).	129
4.41	The GCV score and the number of break points.	130
4.42	The final fitting results.	131
4.43	The GCV score for the smoothing parameters.	132
4.44	The derivative estimates with the selected smoothing parameter.	132

Chapter 1

Introduction

1.1 Background

Reservoir pressure data are useful for obtaining reservoir parameters, monitoring reservoir conditions, developing recovery schemes, and forecasting future well and reservoir performance. Reservoir properties can be inferred by matching the pressure response to a reservoir model, since the alteration of production conditions such as well shut-in, production and injection rate increase or decrease are reflecting in the changes of wellbore or reservoir pressure. Then the inferred reservoir parameters and reservoir models can be used for future reservoir management. For this purpose, well-designed pressure transient tests such as drawdown or build-up test are conducted to observe the pressure response by changing well flow rates for a relatively short period.

Since the 1990s, many wells have been equipped with permanent downhole pressure gauges to monitor the well pressure in real time. This continuous monitoring enables engineers to observe ongoing changes in the well and to make operating adjustment to prevent accidents and optimize recovery. Unneland and Haugland [31] reported experience in permanent downhole gauges in the North Sea and demonstrated their utility and cost-effectiveness for reservoir management. Ouyang and Sawiris [24] proposed a novel application of permanent downhole gauges for production and injection profiling instead of production logging which could be cost-expensive for horizontal or multilateral wells offshore.

Several problems related to the data acquisition system itself have been discussed. Veneruso, Economides and Akmansoy [34] discussed some of the potential noise and nonsignal components of the permanent downhole gauge data which are acquired under uncontrolled

in-situ environment or sometimes under hostile conditions. Kikani, Fair, and Hite [19] pointed out overall pressure data resolution for a gauge system could be lower than the gauge specifications and this may lead to misinterpretation of pressure data. Veneruso, Hiron, Bhavsar, and Bernard [35] reported that short-circuit connections for data transmission causes the most observed failures of gauge systems in the field and that temperature is an important factor for their lives.

Despite the wide usage of permanent downhole gauges, the application of permanent downhole gauges requires special processing and interpretation techniques due to the following complexities [23, 1].

1. Extremely large volume of data

In some cases, pressure is measured at 1 second or 10 second intervals for a period of several years. One year of data consists of millions of measurements. Usually it is impossible to include the entire data set in one processing or interpretation due to the limitations of computer resources.

2. Different type of errors

Compared to the data from well-designed pressure transient tests, permanent downhole gauge data are prone to different types of errors. In the case of long-term monitoring, the well and reservoir may undergo dynamic changes throughout their lives. The well may be stimulated or worked over due to failure in the wellbore. Reservoir pressure may fall below the bubble point because of oil and/or gas production, resulting in two-phase or even three-phase flow in the reservoir. Because of these changes, the permanent downhole gauge data may contain erroneous measurements such as noise and outliers. Abrupt changes in temperature can also cause erroneous recordings. Sometimes the data acquisition system simply malfunctions. These factors creates noise or outliers in the pressure signal. These noise and outliers can give uncertainty in the interpretation.

3. Aberrant pressure behavior

Since the continuous long-term pressure monitoring is under an uncontrolled environment, the recorded pressure data may be inconsistent with flow rate changes for the same reasons mentioned previously. Aberrant pressure behavior during a transient may lead to large uncertainties in reservoir parameter estimation or even an

unreasonable interpretation.

4. Incomplete flow rate history

In most cases, a complete record of flow rate history is not available. In general, the flow rate is not measured continuously. The flow rate may sometimes be measured once a week or only once a month, although there are unmeasured rate changes in between. This incomplete information makes the data analysis difficult.

5. Transient identification

In order to analyze the pressure data, one needs to know the break point, which is the starting point of each transient. Due to the incomplete flow rate history, each break point often has to be located only from the pressure signal.

6. Data analysis

In the data analysis, one needs to recognize a reservoir model and estimate the parameters. Long-term pressure data helps the model recognition process significantly, since the conventional well testing lasts during a relatively short period. However the data processing issues mentioned previously makes this process quite difficult. Moreover the changes in reservoir properties associated with long-term production (for example reservoir compaction) may not be negligible. Since the properties change with time, it is not accurate to interpret all the data at once.

Although these complexities make interpretation of permanent downhole gauge data quite difficult, these data have the potential to provide the more reservoir information than the traditional pressure transient test data. Also continuously monitored data can provide information on the temporal change of reservoir properties associated with long-term production.

1.2 Previous work

Permanent downhole pressure gauges have been installed in many wells since 1990s. However, there have been a limited number of studies on permanent downhole gauge data applications. Many of the studies have focused on the hardware involved in permanent downhole gauge installations [35, 33, 19].

Unneland, Manin, and Kuchuk [32] presented a decline curve analysis using the data acquired with permanent downhole pressure gauges. The authors reported that the analysis of long-term data overcame the ambiguity associated with traditional well test analysis.

Athichanagorn, Horne, and Kikani [1, 2] proposed a seven-step procedure for data processing and interpretation of permanent downhole gauges using a wavelet approach. This was the first study to tackle the issues associated with data from permanent downhole gauges in a comprehensive manner. The seven steps consist of (1) outlier removal, (2) denoising, (3) transient identification, (4) data reduction, (5) flow rate history reconstruction, (6) behavioral filtering, and (7) data interpretation. A brief description of the steps is given here to give more insight into the characteristics of permanent downhole gauge data and one example of data analysis.

1. Outlier removal.

Outliers are data points lying away from the general data trend. Each outlier creates two consecutive singularities corresponding to the departure and the arrival in the wavelet detailed signal. The outlier is identified and removed by checking these singularities.

2. Denoising.

Noise may be defined as the scatter around the general trend of data. Denoising is done using wavelet analysis by setting low wavelet detail coefficients to zero.

3. Transient identification.

Pressure data exhibits rapid changes when a new transient begins (changes in flow rate), creating singularities in the pressure signal. Wavelet analysis is utilized to locate these singularities.

4. Data reduction.

The number of data is reduced based on the pressure thresholding method. In this method, the data are sampled where a certain pressure difference is observed between data points within the maximum sampling interval.

5. Flow rate history reconstruction.

Flow rate history is reconstructed by parameterizing the unknown flow rates as regression model parameters and constraining the regression match to certain known well rates or cumulative production.

6. Behavioral filtering.

Nonlinear regression is applied to determine matched and unmatched sections of pressure data based on the magnitudes of the variances and thus to filter erroneous or aberrant data.

7. Data interpretation.

The processed data is analyzed using a moving window based technique. This technique takes into account the possibility of reservoir and fluid properties changes as production proceeds.

There are a couple of issues in these steps.

1. Denoising.

Denoising requires specification of a threshold for the wavelet detailed signal. Khong [18] determined the noise level using local linear fitting for the data interval over which noise level needs to be estimated. However, the validity of this procedure is dependent on the assumption that pressure varies linearly with time over the time interval the noise level is to be estimated. Ouyang and Kikani [23] used the logarithm function to determine the noise level. They concluded their approach gave a drastic improvement of noise level determination.

2. Transient identification.

Transient identification is one of the major issues in the data processing. Although the wavelet approach drastically reduces human intervention for this purpose, a wavelet approach has been found not to detect break points reliably. Depending on the user-specified threshold, false break points are detected and true break points are missed. A difficulty is choosing the threshold value which can select the valid break points while avoiding any false break point. Khong [18] utilized the feature of break points in the Fourier domain for discriminating true and false break points. Although some improvement was attained, this method failed to screen false break points completely. Ouyang and Kikani [23] showed a limitation of the wavelet approach by formulating the minimum flow rate change detectable for a given reservoir properties and threshold values. Since the location and the number of break points affects the estimation results, this processing issue is inseparable from the subsequent data analysis in its

nature. In that sense, all the local filtering approaches such as wavelet transformation have one obvious limitation. It is impossible for one to judge whether the data processing results are sufficient for the subsequent reservoir parameter estimation.

3. Flow rate history reconstruction and aberrant data detection.

The regression analysis in the flow rate history reconstruction is based upon the assumption of a reservoir model. This is a difficulty if the reservoir model is unknown a priori, especially when the reservoir is a new field or history of the reservoir is not available. The removal of aberrant transients from the analysis is based upon the variance between model pressure and measured pressure and is therefore also based upon an assumption of the reservoir model.

To avoid the ambiguity, Thomas [30] employed a parametric regression approach to extract the model response function without assuming the specific reservoir model by matching the pressure data and also utilized a pattern recognition approach using an artificial neural network for detecting aberrant data sections. However, neither approach could overcome the difficulties completely.

A process to identify the model response function from a given pressure signal is called deconvolution. Although several deconvolution techniques have been proposed, they can not be applied directly due to the complexities associated with permanent downhole gauge data [20, 6, 3].

In recent years, for permanent downhole gauge interpretation, Schroeter, Hollaender, and Gringarten [29, 28] have presented a new deconvolution technique to account for uncertainties not only in the pressure, but also in the rate data. Levitan, Crawford, and Hardwick [21] recommended the use of one single buildup section for the application to avoid the difficulties associated with the pressure data and inconsistency such as storage or skin changes during a sequence, which is to be expected in long-term production/injection. Levitan et al. concluded that accurate reconstruction of constant rate drawdown system response is possible with a simplified rate history as long as; (1) the time span of the rate data is preserved, (2) the well rate honors cumulative well production, and (3) the well rate data accurately represent major details of the true rate history immediately before the start of buildup (the length of this time interval is about two times of the duration of buildup). The well rate prior to this detailed rate time interval can be averaged. However, for accurate estimation from the limited data, their proposed method requires longer buildup (not

drawdown) and corresponding detailed flow rate history. Under the common operation in the industry, this gives a large restriction to the selection of available data sections. In the existing literature related to deconvolution, data processing issues have not been addressed.

1.3 Problem statement

Permanent downhole gauge data have a great potential to provide more reservoir information than the traditional pressure transient test data. In particular continuously monitored data can provide information on a temporal change of reservoir properties associated with long-term production. Also these data can be utilized for reservoir and well management on a routine basis.

The objective of this study has been to develop procedures for processing and analyzing the continuously monitored pressure data from permanent downhole gauges. It is important to extract quantitative information as far as possible by utilizing the long-term pressure data fully.

Based on the limitations identified from the literature review, this study aimed at developing methods which meet the following requirements.

1. To identify locations of flow rate changes from pressure response and partition into individual transients.
2. To reconstruct a complete flow rate history from existing rate measurements, production history, and pressure data.
3. To identify a reservoir model.
4. To detect aberrant transient sections.
5. To be able to estimate reservoir parameters that change with time.

The developed methods should be automated to avoid human intervention as far as possible, given the huge number of data points introduced.

1.4 Dissertation outline

The goal of this study has been to develop a method to process and interpret the long term pressure data acquired from permanent downhole pressure gauges. Less subjective and

more reliable methods were investigated to accommodate the difficulties inherent in the data. Toward this end, the approach undertaken was based on nonparametric regression.

The dissertation summarizes the investigation undertaken during this study. The chapter-by-chapter review is as follows.

1. Chapter 2: Review of data smoothing techniques.

The existing data smoothing techniques are reviewed. The concepts described here are the basis of the methodology development in this study.

2. Chapter 3: Single transient data smoothing.

A data smoother is often said to be a tool for nonparametric regression. In this chapter, the single transient smoothing problem is investigated. Through the numerical experiment and investigation, a more effective smoother algorithm was developed.

3. Chapter 4: Multitransient data analysis.

The multitransient data smoothing problem is described. The algorithm developed enables us to address three technical issues at the same time: (1) flow rate recovery, (2) transient identification, and (3) model identification. The developed procedure was tested for synthetic data and the actual field data sets acquired from conventional well testing and permanent downhole gauges.

4. Chapter 5: Conclusions and future work.

The final chapter summarizes the results and suggestions of future work to be addressed based on the results of this study.

Chapter 2

Review of data smoothing techniques

One of the most popular and useful tools in data analysis is the linear regression model. In the simplest case, we have n measurements of a response variable Y and a single predictor variable X . In some situations we assume that the mean of Y is a linear function of X ,

$$E(Y|X) = \alpha + X\beta \tag{2.1}$$

The parameters α and β are usually estimated by least squares, namely by finding the values α and β that minimize the residual sum of squares. However, if the dependence of Y on X is far from linear, we would not want to summarize it with a straight line.

The idea of single variable (scatter plot) smoothing is describing the dependence of the mean of predictor as a function of X .

A smoother is a tool for summarizing the trend of a response measurement Y as a function of one or more predictor measurements $X_1, \dots, X_p$. The smoother produces an estimate of the trend that is less variable than Y itself, hence the name smoother.

An important property of a smoother is its nonparametric nature: it does not assume a rigid form for the dependence of Y on $X_1, \dots, X_p$. For this reason, a smoother is often referred to as a tool for nonparametric regression.

We can estimate the functional dependence without imposing a rigid parametric assumption using a smoothing technique.

$$E(Y|X) = f(X) \quad (2.2)$$

The running mean (moving average) is a simple example of a smoother, while a regression line (Equation 2.1) is not strictly thought of as a smoother because of its rigid parametric form over the entire domain.

In this chapter, the existing single- and multivariate smoothing techniques and the related concepts are described.

2.1 Single variable smoothing technique

In this section, the various single variable smoothing techniques and their mathematical properties are described. These are the basis for multivariate nonparametric regression techniques discussed in a later section.

2.1.1 Parametric regression

First a parametric regression technique is outlined by limiting ourselves to general linear parametric regression to introduce several concepts. The general form of this kind of model is:

$$y(x) = \sum_{k=1}^m a_k X_k(x) \quad (2.3)$$

where $X_1(x), X_2(x), \dots, X_m(x)$ are arbitrary fixed functions of x , called basis functions. The functions $X_k(x)$ can be nonlinear functions of x . Here linear refers only to the model dependence on its parameters a_k .

Parameters a_k are determined through minimization of the residual sum of squares:

$$RSS(\vec{a}) = \sum_{i=1}^n \left(\sum_{k=1}^m a_k X_k(x_i) - y(x_i) \right)^2 \quad (2.4)$$

In a matrix form,

$$RSS(\vec{a}) = \|\underline{A}\vec{a} - \vec{y}\|^2 \quad (2.5)$$

The matrix $\underline{A}$ is a $(n \times m)$ matrix called the design matrix ($\text{rank}(\underline{A})=m$). $\underline{A}\vec{a}$ represents a linear subspace consisting of column vectors of $\underline{A}$, then the minimum value in Equation 2.4 is given by projecting vector $\vec{y}$ onto this subspace determined a priori. The orthogonality of the residual vector to this subspace formulates the following equation called a normal equation (Figure 2.1).

$$\underline{A}^T(\underline{A}\vec{a} - \vec{y}) = 0 \quad (2.6)$$

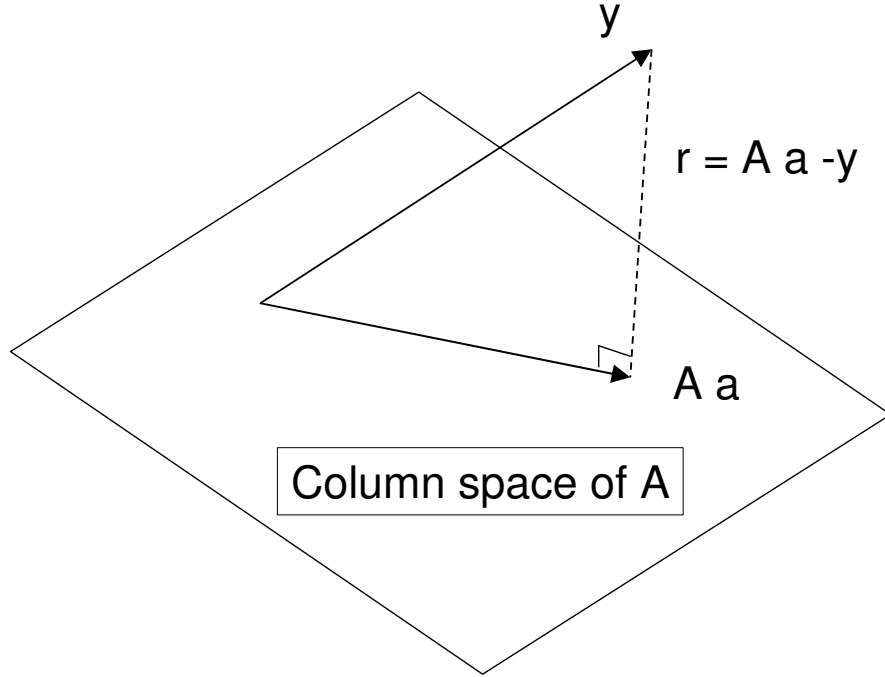


Figure 2.1: Projection onto column space of a design matrix.

Then, the parameters and function value are given by:

$$\vec{a} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \vec{y} \quad (2.7)$$

$$\vec{f} = \underline{A}\vec{a} = \underline{A}(\underline{A}^T \underline{A})^{-1} \underline{A}^T \vec{y} = \underline{H}\vec{y} \quad (2.8)$$

The $(n \times n)$ matrix $\underline{H}(= \underline{A}(\underline{A}^T \underline{A})^{-1} \underline{A}^T)$ independent of $\vec{y}$ is called a hat matrix or projection matrix in statistics. It has several mathematical properties worth noting.

1. $\underline{H}$ is a symmetric, nonnegative definite, and idempotent matrix.

$$\underline{H}\underline{H} = \underline{H} \quad (\text{idempotent}) \quad (2.9)$$

2. Eigenvalues of $\underline{H}$ are 0 and 1.
3. $\underline{I} - \underline{H}$ is also a projection matrix and $(\underline{I} - \underline{H})\vec{y}$ becomes the residual vector $\vec{r}$.
4. Rank of $\underline{H}$ is m (the number of parameters to fit).

Due to its idempotent property, a projection matrix (hat matrix) does not change the smooth $\vec{f}$ through iterative application. Polynomial fitting falls into this category, whose basis functions are defined over entire domain X . Polynomial fitting has a global nature which means tweaking the coefficients to achieve a functional form in one region can cause the function to deviate in a remote region.

2.1.2 Preliminary definition of smoother and operator

Here several definitions for smoothers are described for preparation.

1. Linear smoother

If a smoother can be written down as $\vec{f} = \underline{S}\vec{y}$ for any $\underline{S}$ independent of $\vec{y}$, it is called linear smoother. Then the smoother satisfies

$$\underline{S}(\vec{x} + \vec{y}) = \underline{S}(\vec{x}) + \underline{S}(\vec{y}) \quad (2.10)$$

If $\underline{S}$ is dependent on $\vec{y}$, it is called nonlinear smoother.

2. Constant preserving smoother

If a smoother preserves a constant vector, then it is called constant preserving smoother. Its mathematical notation is:

$$\underline{S}\vec{1} = \vec{1} \quad (2.11)$$

Here $\vec{1}$ is a vector whose components are all 1. Therefore, the constant preserving smoother has at least eigenvalue of 1 and corresponding eigenvector $\vec{1}$.

3. Centered smoother

If a smoother produces zero mean vectors, it is called centered smoother. It can be expressed as a multiplication of two linear operators; centering operator $\underline{S}_c$ and smoothing operator $\underline{S}$.

$$\underline{S}^* = \underline{S}_c \underline{S} = (\underline{I} - \frac{1}{n} \vec{1} \vec{1}^T) \underline{S} \quad (2.12)$$

4. Permutation operator

A permutation matrix $\underline{P}$ is a matrix that has exactly one in each row and each column, all other entries being zero. It exchanges the components of the vector for data sorting etc. It is an orthogonal matrix, and does not change l_2 norm of a vector.

$$\underline{P}^T \underline{P} = \underline{I} \quad (2.13)$$

5. Normalization operator

The normalization operator makes a vector norm one.

2.1.3 Local smoother

In this subsection, three local smoothers are described; (1) running mean, (2) running line, and (3) kernel smoother. All are linear smoothers. Their general definitions are as follows.

1. Running mean

A running mean smoother produces a fit at points x_i by averaging the data in a neighborhood around x_i .

$$S(x_i) = \text{ave}_{j \in N(x_i)} y_j \quad (2.14)$$

Here $N(x_i)$ is called nearest neighborhood, where x_i itself as well as the k points to the left and the k points to the right are included. k is called the span. If it is not possible to take k points to the left or right of x_i , we take as many as we can.

A formal definition of a symmetric nearest neighborhood is:

$$N(x_i) = \{max(i - k, 1), ..., i - 1, i, i + 1, ..., min(i + k, n)\} \quad (2.15)$$

A running mean is the simplest smoother, but the result tends to be wiggly, because each data point has equal and discontinuous weight (zero weight outside of the neighborhood), thus is more affected by the data values entering the neighborhood and exiting out of it. This tends to flatten out the result near the data boundaries. The smoother matrix becomes close to a band matrix and all entries in the same row have the same positive value.

2. Running line

A running line smoother fits a line by least squares to the data in a neighborhood around x_i and alleviates the end effect in the running mean but still tends to give wiggly results due to its discontinuous weighting. The nonzero elements in the i th row of the smoother matrix are given by:

$$s_{i,j} = \frac{1}{n_i} + \frac{(x_i - \bar{x}_i)(x_j - \bar{x}_i)}{\sum_{k \in N_i} (x_k - \bar{x}_i)^2} \quad (2.16)$$

where n_i denotes the number of observations in the neighborhood of the i th point, j subscripts the points in this neighborhood and $\bar{x}_i$ denotes their mean. The smoother matrix becomes close to a band matrix and each component can be positive and negative.

3. Kernel smoother

The locally weighted running line smoother is a linear smoother, whose weights are determined only by x values. Kernel smoothers were developed to improve the jagged appearance in simple running mean or line smoother results by adjusting the weights such that they smoothly disappear. The locally weighted running line smoother is one representative smoother [7]. As a weighting function (kernel), the tricube weighting function is employed.

$$W(u) = (1 - u^3)^3 \quad (2.17)$$

$$u = \frac{|x_i - x_j|}{\Delta(x_i)} \quad (2.18)$$

where x_i is a target point and $\Delta(x_i)$ is the maximum distance between the target point and the point in the neighborhood with span k .

Then the target point is heavily weighted and right and left ends are less weighted. The smoother matrix is close to a band matrix. Although not described here, there are other popular smoothers with different kernels such as Epanechnikov kernel [9].

2.1.4 Regression spline

The polynomial regression has a limited appeal due to the global nature of its fit, while in contrast the smoothers described so far have an explicit local nature. The regression spline offers a compromise by representing the fit as a piecewise polynomials.

Definition

A function $f(x)$, defined on a finite interval $[a, b]$, is called a spline function of degree M (order $M + 1$), having knots as the strictly increasing sequence $k_p, \{p = 0, 1, \dots, g + 1 \}$ ($k_0 = a, k_{g+1} = b$), if the following two conditions are satisfied.

1. On each knot interval $[k_j, k_{j+1}]$, $f(x)$ is given by a polynomial of degree M at most.
2. The function $f(x)$ and its derivative up to order $M - 1$ are all continuous on $[a, b]$.

Here knots are defined as fixed points partitioning the entire domain. This spline function can be expressed using piecewise polynomials. In general, a popular choice is the use of piecewise cubic polynomial basis ($M = 2$) having continuous first and second derivatives at the knots. The reason is simply because if there is discontinuity of the first and second derivative the appearance of the function is jagged even for human eyes. If we are interested in the derivative estimation the higher order spline function is preferable.

A simple choice of basis function is truncated power series.

$$f(x) = \sum_{j=1}^K \beta_j (x - k_j)_+^3 + \beta_{K+1} + \beta_{K+2}x + \beta_{K+3}x^2 + \beta_{K+4}x^3 \quad (2.19)$$

where the notation a_+ denotes the positive part of a . It is clear this basis function representation follows the definition of a spline. We can represent or approximate any function using the predefined basis $(X_j(x), j = 1, 2, \dots, m)$ if the knot position is chosen appropriately. A resulting function can be described using the summation of each basis.

$$f(x) = \sum_{j=1}^m \beta_j X_j(x) \quad (2.20)$$

This is similar to Equation 2.3. Therefore, by fitting this function to data, the resulting function vector $\vec{f}$ can be written down using a hat matrix as:

$$\vec{f} = \underline{A}\vec{\beta} = \underline{A}(\underline{A}^T \underline{A})^{-1} \underline{A}^T \vec{y} = \underline{H}\vec{y} \quad (2.21)$$

In the context of nonparametric regression, one general issue regarding regression splines is determining the number and position of the knots. A smaller number of knots disables the ability of a smoother to detect the local behavior. On the other hand, too many knots tend to give the interpolating function, masking the global nature due to its jagged appearance. We often do not need too many degrees of freedom in the function for the estimation and computation purposes.

2.1.5 Smoothing spline

Here, the smoothing spline and its mathematical properties are described. The difference between the regression spline and the smoothing spline is that the smoothing spline avoids the knot selection problem by using a maximal set of knots (= the number of data points). This yields an interpolating function that is unfavorable in many situations. Then, the resulting complexity of the fit is controlled by regularization. Consider the following minimization problem: among all the functions $f(x)$ with continuous first and second derivatives, find the one that minimizes the penalized residual sum of squares

$$RSS(f, \lambda) = \sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b \{f''(x)\}^2 dx \quad (2.22)$$

where λ is a fixed parameter called the smoothing parameter. The first term measures the closeness to the data, while the second term penalizes the curvature in the function, and λ establishes a tradeoff between the two. The two special cases are:

1. $\lambda = 0$

This yields an interpolating function.

2. $\lambda = \infty$

This gives the simple least squares line fit, since no second derivative values can be tolerated.

The problem is to find a function $f(x)$ which minimizes $RSS(f, \lambda)$.

If we represent $f(x)$ using a set of cubic basis functions, Equation 2.22 can be written down:

$$RSS(\vec{\beta}, \lambda) = (\vec{y} - \underline{A}\vec{\beta})^T(\vec{y} - \underline{A}\vec{\beta}) + \lambda\vec{\beta}^T \underline{\Omega} \vec{\beta} \quad (2.23)$$

where $\underline{A}$ is a design matrix ($A_{ij} = X_j(x_i)$), and Ω is a penalty matrix and its component is $\Omega_{jk} = \int_a^b X_j''(x)X_k''(x)dx$.

The solution is easily seen to be:

$$\vec{\beta} = (\underline{A}^T \underline{A} + \lambda \underline{\Omega})^{-1} \underline{A}^T \vec{y} \quad (2.24)$$

Then the fitted smoothing spline is given as:

$$\vec{f} = \underline{A}(\underline{A}^T \underline{A} + \lambda \underline{\Omega})^{-1} \underline{A}^T \vec{y} \quad (2.25)$$

These equations can be converted into the simple form by replacing $\underline{A}\vec{\beta}$ with $\vec{f}$.

$$RSS(\vec{f}, \lambda) = (\vec{y} - \vec{f})^T(\vec{y} - \vec{f}) + \lambda \vec{f}^T \underline{K} \vec{f} \quad (2.26)$$

$$\vec{f} = \{\underline{I} + \lambda \underline{K}\}^{-1} \vec{y} = \underline{S}_\lambda \vec{y} \quad (2.27)$$

where $\underline{K} = \underline{N}^T \underline{\Omega} \underline{N}^{-1}$ ($\underline{N} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T$).

The following are the interesting properties of the smoother matrix $S_\lambda (= \{\underline{I} + \lambda \underline{K}\}^{-1})$.

1. $\underline{S}_\lambda$ is not a projection matrix.

$$\underline{S}_\lambda \underline{S}_\lambda \neq \underline{S}_\lambda \quad (2.28)$$

2. $\underline{S}_\lambda$ is a symmetric, nonnegative definite matrix. Therefore it has real eigenvalue decomposition.

$$\underline{S}_\lambda = \sum_{k=1}^n \rho_k(\lambda) u_k u_k^T \quad (2.29)$$

where $\rho_k(\lambda)$ and u_k are the k th eigenvalue and the corresponding eigenvector.

3. Rank of $\underline{S}_\lambda$ is n .
4. Eigenvalue $\rho_k(\lambda)$ can be expressed as

$$\rho_k(\lambda) = \frac{1}{1 + \lambda d_k} \quad (2.30)$$

where d_k is an eigenvalue of $\underline{K}$.

5. $\underline{K}$ is a nonnegative definite matrix. Then $d_k \geq 0$ and $\rho_k(\lambda) \in [0, 1]$. Therefore

$$\|\underline{S}_\lambda \underline{S}_\lambda\| \leq \|\underline{S}_\lambda\| \quad (2.31)$$

Here $\|\cdot\|$ is a l_2 matrix norm.

6. Eigenvectors of $\underline{S}_\lambda$ are not affected by the λ value.
7. The smoothing spline preserves any constant and linear function. Therefore the first two eigenvalues of the smoother matrix are 1. The corresponding eigenvectors are constant and linear vectors over x .
8. The sequence of eigenvectors, ordered by decreasing $\rho_k(\lambda)$, exhibits increasing the polynomial behavior in the sense that the number of zero crossing is increased [14].

Based on these properties, when λ increases, polynomial behavior of the fit is drastically reduced, by preferably downweighting the higher order polynomial components of eigenvectors (Equation 2.29). For that reason, the spline smoother is sometimes referred to as a shrinking smoother.

2.1.6 Nonlinear smoother

If a smooth $\vec{f}$ can not be written down as $\vec{f} = \underline{S}\vec{y}$ for any $\underline{S}$ independent of $\vec{y}$, the smoother is called a nonlinear smoother.

An example of a nonlinear smoother is the variable span smoother called super smoother [10]. This smoother is an enhancement of the running line smoother, the difference being that it chooses a different span at each observation. It does so in order to adapt to the changes in the curvature of the underlying function and the variance of Y . In regions where the curvature to variance ratio is higher a smaller span is selected.

In this technique, the span is determined at every data location through cross validation as described later. Define a cross validation score (CV score) as:

$$I^2(x_i|k) = \frac{1}{L} \sum_{j=i-L/2}^{j=i+L/2} \{y_j - s_j^{-1}(x_j|k)\}^2 \quad (2.32)$$

In Equation 2.32, span L is different from span k . $s_j^{-1}(x_j|k)$ is the linear fitting value at x_j , which is calculated from $2k$ observations excluding x_i in the neighborhood.

The optimal span $k(i)$ at each data point x_i is determined by minimizing the equation 2.32,

$$k(i) = \min^{-1} I^2(x_i|k) \quad (2.33)$$

Although span L is selected in a similar way, the L value from $0.2n$ to $0.3n$ is usually reasonable [10]. Here n is the number of data points. If we take L equal to n , the same span k is applied to the entire domain. To account for the local features as much as possible, this algorithm uses the definition of the local CV score instead of global CV score.

All the smoothers we have been looking at become nonlinear if smoothing parameter λ or span k are determined based on the response value y through cross validation etc.

2.1.7 Wavelet smoothing

Wavelet smoothing is another category of data smoothing. With regression splines, we select a set of bases, using either subject-matter knowledge or automatically. With smoothing splines, we use a complete basis, but then adjust the coefficients toward the smoothness of the fit. Wavelets typically use a complete orthonormal basis to represent functions, but then shrink and select the coefficients toward a sparse representation. Just as a smooth function

can be represented by a few spline basis functions, a mostly flat function with a few isolated bumps can be represented with a few basis functions. Wavelet bases are very popular in signal processing and compression, since they are able to represent both smooth and/or locally bumpy functions in an efficient way. Atichanagorn [1, 2] utilized this property of wavelet for denoising and edge detection purposes.

Wavelet smoothing fits the coefficients for the basis by least squares and then thresholds the smaller coefficients.

In a mathematical notation, the smooth result can be written down as:

$$\vec{f} = \underline{W}^T \underline{T} \underline{W} \vec{y} \quad (2.34)$$

Here $\underline{W}$ is a wavelet matrix and $\underline{T}$ is a diagonal thresholding matrix, whose elements are 0 or 1.

This can be viewed as a linear smoother if the threshold is determined a priori. In terms of compression, the smoothing spline achieves compression of the original signal by imposing the smoothness, while wavelets impose the sparsity.

2.2 Model assessment and selection

All the smoothing techniques described in the previous section have their own smoothing parameter (sometimes called complexity parameter) that has to be determined. For example,

1. smoothing parameter λ in the smoothing spline
2. span k in local smoothers
3. the number of basis function in the regression spline
4. the threshold in wavelet smoothing

These parameters should be determined such that the resulting model is appropriate for the target data. In the case of the smoothing spline, the parameter λ indexes models ranging from a straight line to an interpolating function. Similarly a local degree m polynomial ranges from a degree- m global polynomial when the span is infinitely large, to an interpolating fit when the span shrinks.

This indicates that we can not use the residual sum of squares to determine these parameters as well, since we always pick those that give interpolating fits and hence zero residuals. Such a model is unlikely to have high prediction capabilities.

2.2.1 Degrees of freedom

The concept of the degrees of freedom is often utilized to measure the complexities of models in statistics.

Given an estimate $\vec{f}$, it would be useful to know how many degrees of freedom we have fitted to the data. There are several definitions of degrees of freedom in the literature [14]. In any definition, the more degrees of freedom we fit, the rougher will be the function and the higher its variance (Figure 2.2). The simplest definition of the degrees of freedom (DF) is the trace of the smoother matrix $\underline{S}$:

$$DF = \text{trace}(\underline{S}) \tag{2.35}$$

For the linear parametric regression model, the number of parameters to be fitted is the rank of a hat matrix (projection matrix). This definition comes from this analogy. Since $\text{trace}(\underline{S})$ is easiest to compute, it may be the logical choice.

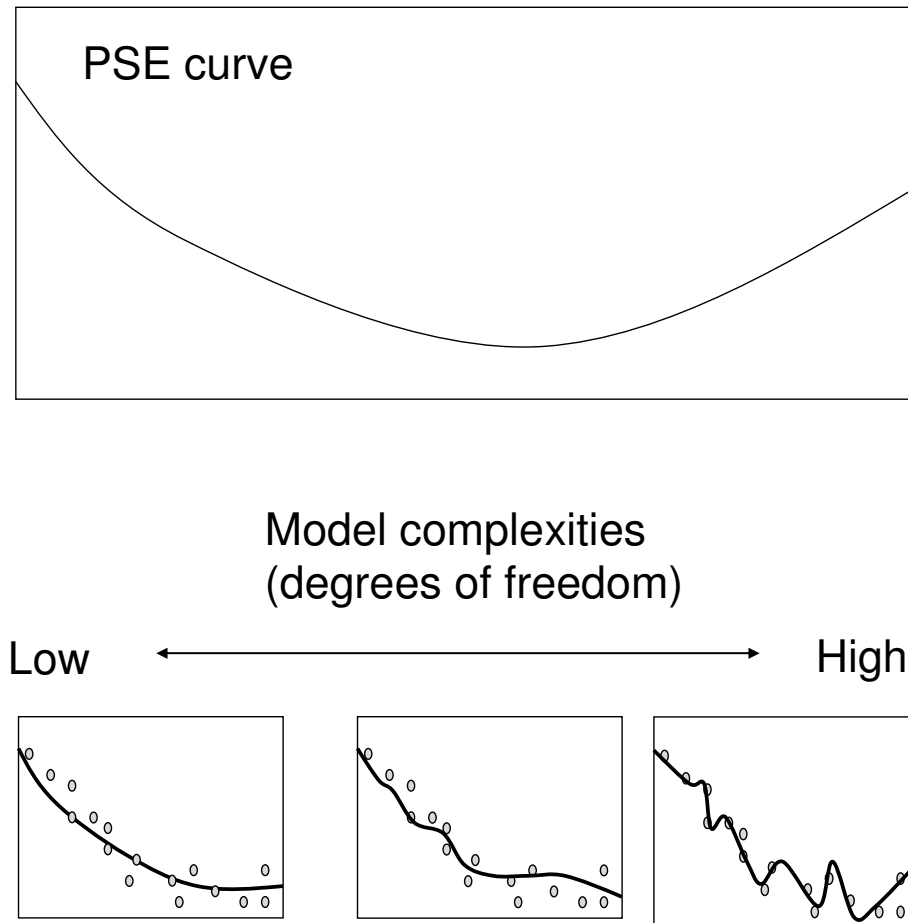


Figure 2.2: Model complexities (degrees of freedom) and the corresponding prediction squared error (PSE) curve.

For the regression spline, the number of knots directly describes the complexities of the model. Larger numbers of knots exhibit more overfitted results and vice versa. For the local moving average, if span k is larger each data point is less weighted, then the diagonal term of the smoother matrix becomes smaller. For the smoothing spline, if we increase the smoothing parameter λ , the eigenvalues of the smoother matrix become small, thus $\text{trace}(\underline{S})$. As in these examples, in many situations this definition can be utilized to explain the model complexities quantitatively.

2.2.2 Prediction error estimate

In this subsection, several concepts related to controlling the model complexities are described.

In order to describe the concepts, we assume:

$$Y = f(X) + \varepsilon \quad (2.36)$$

where $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2$, and the errors ε are independent.

We need to select the degrees of freedom of the model to achieve high predictability rather than accountability for the particular data set. The prediction squared error (PSE) is often utilized as a predictability measure. PSE is decomposed into inherent noise, bias, and variance terms as is well known [14].

$$\begin{aligned} PSE &= \frac{1}{n} \sum_{i=1}^n E(Y_i^* - \hat{f}(x_i))^2 \\ &= \frac{1}{n} \sum_{i=1}^n E(Y_i^* - f(x_i))^2 + \frac{1}{n} \sum_{i=1}^n (E(\hat{f}(x_i)) - f(x_i))^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n E(\hat{f}(x_i) - E(\hat{f}(x_i)))^2 \\ &= \sigma^2 + \text{Bias}^2 + \text{Variance} \end{aligned} \quad (2.37)$$

Here Y_i^* is a realization at x_i , $\hat{f}$ and f are the estimate with the fixed control parameter and the true one respectively. In the final expression, the first term is the error variance of data and cannot be avoided no matter how well we estimate $f(X)$, unless $\sigma^2 = 0$. The second term is the squared bias, the amount by which the average of our estimates differs

from the true mean; the last term is variance, the expected squared deviation of $f(X)$ around its mean. Typically the more complex we make the model $f(X)$, the lower the bias but the higher the variance.

We wish to reduce this PSE while controlling the balance between bias and variance by choosing the control parameter suitably. This aspect is most easily seen in the case of the running mean smoother with span k .

The estimates of running mean with span k is written as:

$$f_k(x_i) = \sum_{j \in N(x_i)} \frac{y_j}{2k+1} \quad (2.38)$$

Its expected value, variance, and bias are:

$$E\{f_k(x_i)\} = \sum_{j \in N(x_i)} \frac{f(x_j)}{2k+1} \quad (2.39)$$

$$\begin{aligned} Var\{f_k(x_i)\} &= E\{f_k(x_i) - E\{f_k(x_i)\}\}^2 \\ &= E\left\{\sum_{j \in N(x_i)} \frac{y_j - f(x_j)}{2k+1}\right\}^2 \\ &= \sum_{j \in N(x_i)} \frac{E\{y_j - f(x_j)\}^2}{2k+1} \\ &= \sum_{j \in N(x_i)} \frac{\sigma^2}{2k+1} \end{aligned} \quad (2.40)$$

$$\begin{aligned} Bias\{f_k(x_i)\} &= E\{f_k(x_i)\} - f(x_i) \\ &= \sum_{j \in N(x_i)} \frac{f(x_j)}{2k+1} - f(x_i) \end{aligned} \quad (2.41)$$

Therefore, increasing span k clearly decreases variance in Equation 2.40 and tends to increase bias since the average of $f_k(x_i)$ is determined from far-away data points. Decreasing the span has opposite effects.

Prediction error is a good measure to select the appropriate model complexity or roughness of the fit. However we do not know the true function, hence the prediction error in

general.

The simplest and most widely used method for estimating prediction error is cross validation. The method tries to estimate the prediction error using the given finite number of data. In that sense, cross validation is sometimes said to be the in-sample prediction error estimate.

K -fold cross validation uses a part of the available data to fit the model, and a different part to test it. Figure 2.3 shows the schematic of the procedure.

The data are split randomly into K roughly equal-sized parts, for the K th part, we fit the model to the other $K - 1$ parts of the data (Figure 2.3 (b)) and calculate the prediction error of the fitted model when predicting the K th part of the data (Figure 2.3 (c)). Then the prediction error called the cross validation score, is estimated by repeating this process K times and averaging the error.

The case of $K = N$ (number of data points) is referred to as *leave-one-out* cross-validation.

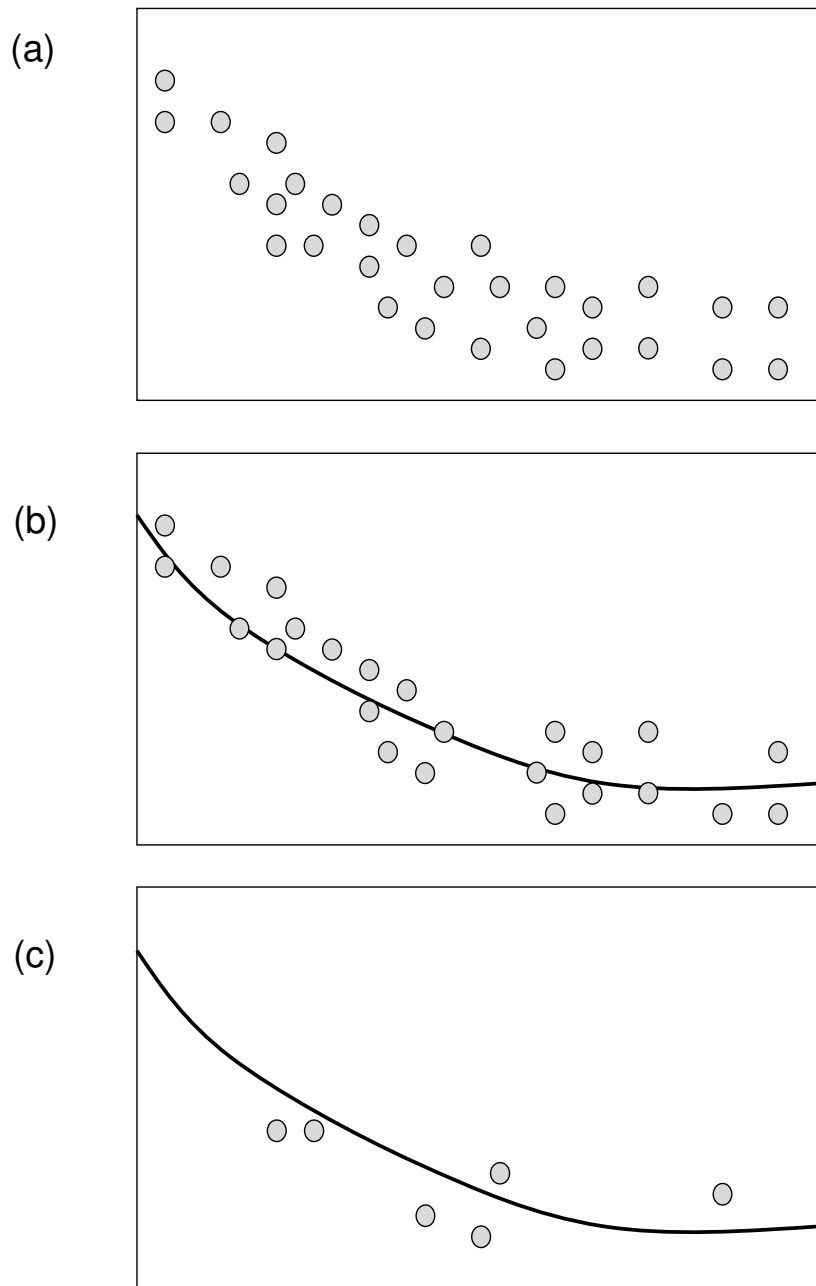


Figure 2.3: K -fold cross validation. (a) original data, (b) data extraction and model fitting to the remaining data, and (c) predict the value at the extracted data points.

2.3 Multivariate data smoothing technique

We have been looking at the single variable data smoothing technique. Here the multivariate data smoothing technique is described. The simplest tool is the multiple linear regression model similarly as in the single variable case.

$$Y = \alpha + X_1\beta_1 + X_2\beta_2 + \dots + X_p\beta_p + \varepsilon \quad (2.42)$$

where $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$.

This model makes a strong assumption about the dependence of $E(Y)$ on $X_1, \dots, X_p$, namely that the dependence is linear in each of the predictors. This assumption makes it difficult for us to achieve the goal of multivariate data analysis:

1. Description

We want a model to describe the dependence of the response on the predictors so that we can learn more about the process that produces Y .

2. Inference

We want to assess the relative contribution of each predictor in explaining Y .

3. Prediction

We wish to predict Y for some sets of values $X_1, \dots, X_p$.

For all these purposes, a restricted class of nonparametric multivariate regression techniques have been developed in the realm of statistics [4, 13]. These techniques have a linear feature in their regression form, but have a striking feature in data mining aspect. In this section, one representative nonparametric regression techniques called the ACE algorithm [4] is described in connection with the data smoothing theory.

2.3.1 ACE algorithm

The conventional multiple regression technique for estimation of response variable (Y) from predictor variables (X_i) requires a functional relationship to be presumed.

However, because of the inexact nature of the relationship between response and predictor variables, it is not always possible to identify the underlying functional form in advance.

ACE (alternating conditional expectations) algorithm, a nonparametric regression technique originally proposed by Breiman and Friedman [4], provides a method for estimating functions (transformations) in multiple regression without prior assumptions of a functional relationship. The method brings objectivity to the choice of transformations in multivariate data analysis. The applicability in the realm of petroleum engineering has been demonstrated by several authors [36, 27].

Generalized additive model [13], and alternating least square method [5] are similar nonparametric regression techniques. They also have the linear feature but ACE enhances the feature by transforming response variables as well.

First, we will outline the algorithm itself and later describe some of theoretical aspects. For further details, see the reference [4].

Let us say that we have response variable Y and predictor variables $X_1, \dots, X_p$. We first define arbitrary zero-mean transformations $\theta(Y), \phi_1(X_1), \dots, \phi_p(X_p)$. A regression of the transformed response variable on the sum of transformed predictor variables results in the following error:

$$e^2(\theta(Y), \phi_1(X_1), \phi_2(X_2), \dots, \phi_p(X_p)) = E\{[\theta(Y) - \sum_i \phi_i(X_i)]^2\} \quad (2.43)$$

Then, ACE finds the optimal transformations $\theta(Y)$ and $\phi_i(X_i)$ that minimize e^2 under $E\{\theta^2(Y)\} = 1$.

For a given set of $\phi_i(X_i)$, the minimization of e^2 with respect to $\theta(Y)$ yields:

$$\theta(Y) = \frac{E\{\sum_i \phi_i(X_i)|Y\}}{\|E\{\sum_i \phi_i(X_i)|Y\}\|} \quad (2.44)$$

Here, $\|\cdot\|$ is a norm (standard deviation). Also, for a given $\theta(Y)$ and a given set of $\phi_j(X_j)$ with $j \neq i$, the minimization of e^2 with respect to $\phi_i(X_i)$ gives:

$$\phi_i(X_i) = E\{[\theta(Y) - \sum_{j \neq i} \phi_j(X_j)]|X_i\} \quad (2.45)$$

Equations 2.44 and 2.45 form the basis of ACE algorithm. These single function minimizations are iterated until one complete pass over the predictor variables fails to reduce e^2 . The error minimization procedure for finding optimal transformations can be summarized as follows.

1. Set starting functions for $\phi_i(X_i)$ and $\theta(Y)$.
(Outer Loop)

2. (Inner loop) Update $\phi_i(X_i)$ for $i = 1, \dots, p$

$$\phi_i(X_i) = E\{[\theta(Y) - \sum_{j \neq i} \phi_j(X_j)]|X_i\} \quad (2.46)$$

(End inner loop)

3. Update $\theta(Y)$

$$\theta(Y) = \frac{E\{\sum_i \phi_i(X_i)|Y\}}{\|E\{\sum_i \phi_i(X_i)|Y\}\|} \quad (2.47)$$

(End outer loop)

This algorithm decreases e^2 at each step by alternately minimizing with respect to one function and holding the others fixed at their previous evaluation. The process begins with an initial guess for the functions and ends when a complete iteration pass fails to decrease e^2 .

In the original algorithm, the starting functions are set as:

$$\begin{aligned} \theta(Y) &= Y/\|Y\| \\ \phi_i(X_i) &= E\{Y|X_i\} \quad (i = 1, \dots, p) \end{aligned} \quad (2.48)$$

By minimizing $E\{[\theta(Y) - \sum f_i(X_i)]^2\}$, ACE provides a regression model,

$$\theta(Y) = \sum \phi_i(X_i) \quad (2.49)$$

$$Y^* = \theta^{-1}(\sum \phi_i(X_i)) \quad (2.50)$$

where Y^* is the prediction of Y . As Equation 2.49 implies, ACE tries to make the relationship of $\theta(Y)$ to $\phi_i(X_i)$ as linear as possible. The resulting transformations are useful for descriptive purposes and for uncovering relationships between Y and X_i . ACE makes it easier to examine how each X_i contributes to Y .

In order to calculate conditional expectations which appear in Equations 2.44 and 2.45, one needs to know the joint probability distribution of Y and X_i . However, such distribution for a finite data set is rarely known. In the ACE algorithm, calculation of conditional expectations are replaced by bivariate scatter plot smoothing.

For smoothing, a local linear fitting technique called the super smoother is employed in the original ACE algorithm [10].

Figure 2.4 shows a scatter plot of response variable Y and predictor variables X_1 , X_2 , and X_3 . This Y is generated using the following equation.

$$Y = X_1 + \log(X_2) + \sin(10X_3) + \varepsilon \quad (2.51)$$

where X_i and ε are sampled from a uniform distribution $U(-1, 1)$. Since Y is a function of X_1, X_2 , and X_3 and also includes noise, it is difficult to observe a clear relationship among response and predictor variables only from each scatter plot. To demonstrate its utility, ACE was applied to this data set. The resultant transformations are shown in Figure 2.5. In the figures, standardized true solutions are also plotted. Approximately, X_1 and Y are transformed to linear functions and X_2 and X_3 are transformed to logarithmic and sinewave functions respectively. As can be seen, ACE captured the relationship among these parameters reasonably well satisfying Equation 2.49 with $\rho = 0.679$ (Figure 2.6). A scatter plot of Y^* versus Y gives $\rho = 0.665$. ACE tries to minimize the error variance in the transformed space. Therefore, lower correlation coefficient is usually obtained in the original space.

The basic limitation of ACE technique is that for prediction purposes transformation of Y is restricted to be monotonic due to invertibility and that it is still linear in the transformed space. However, this algorithm drastically reduces the burden in the multivariate data exploration by generating the first order approximate relationship among parameters in any situation. It is also important to note in the case of well test data applications that the expected pressure transient behavior is in fact a monotonic function.

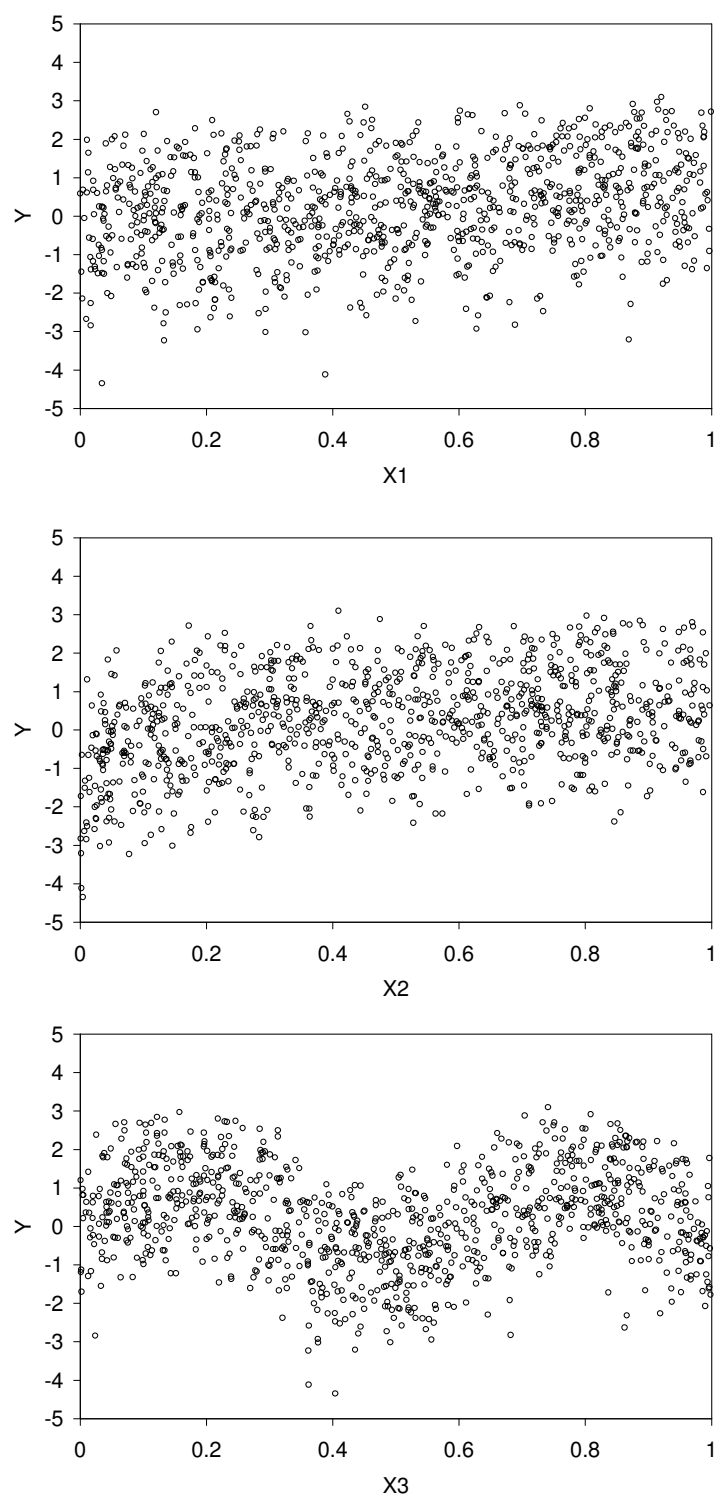


Figure 2.4: Scatter plot of response variable Y and predictor variables X_1 , X_2 , and X_3 .

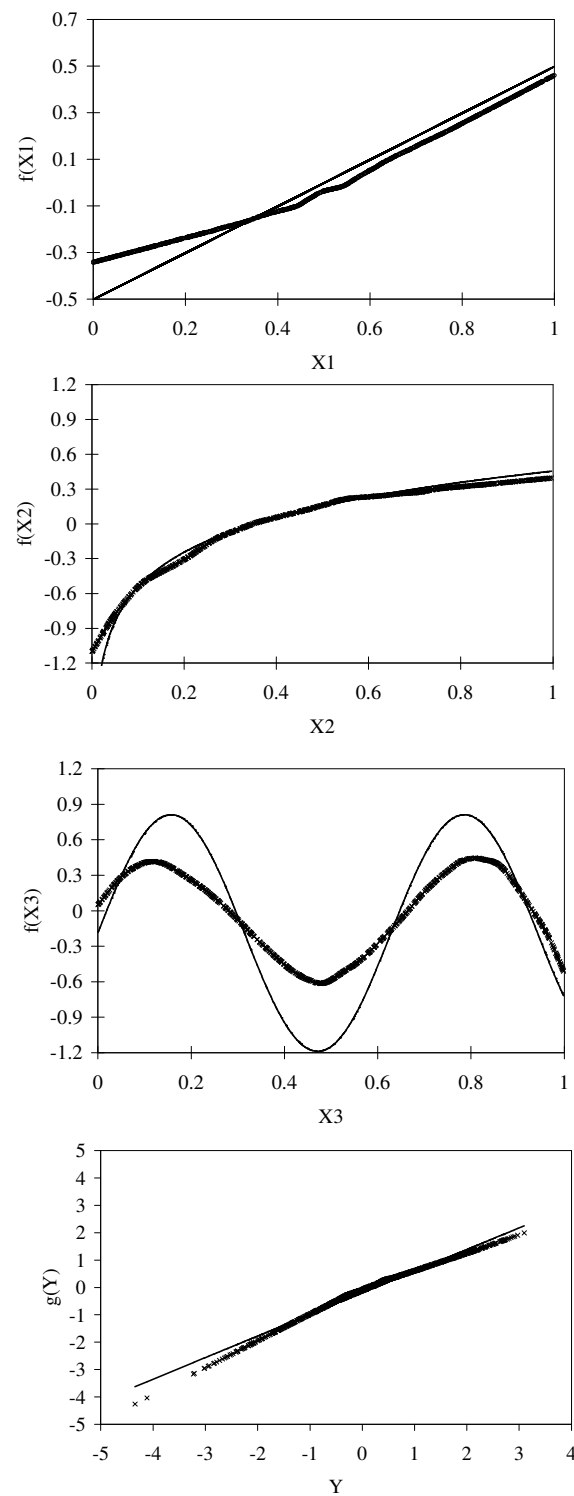


Figure 2.5: Transformations of response variable Y and predictor variables X_1 , X_2 , and X_3 .

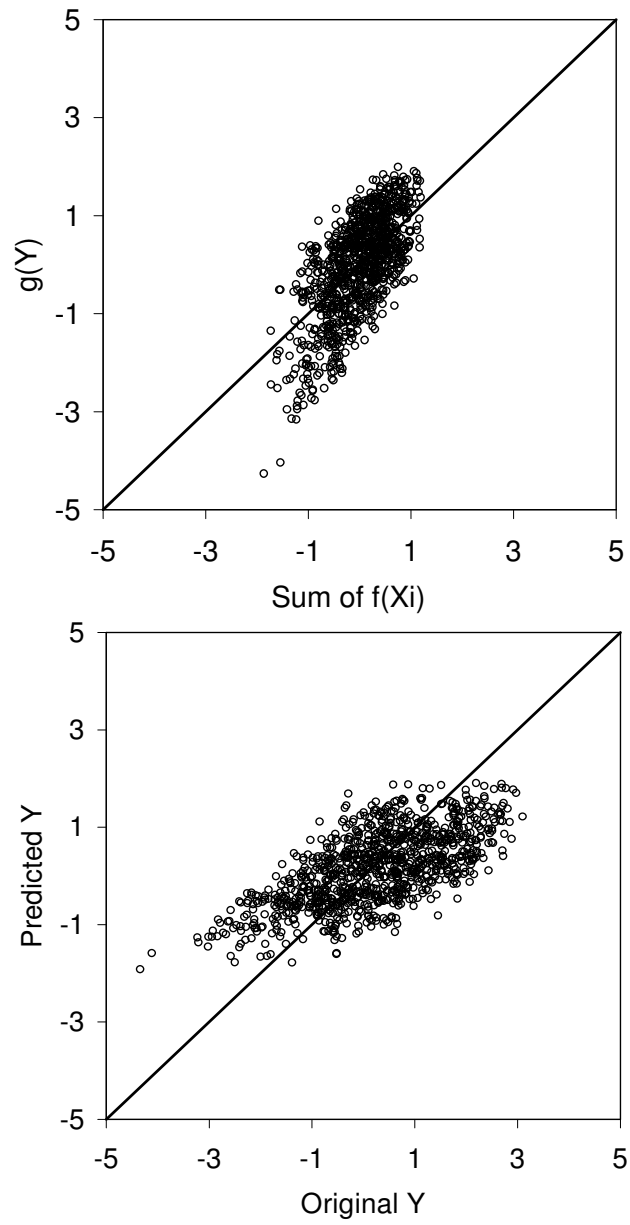


Figure 2.6: Prediction quality of $\theta(Y)$ and Y .

2.3.2 Some theoretical aspects of ACE

In this section, some theoretical aspects of the ACE algorithm are described in population settings and data space.

Let H_j ($j = 1, \dots, p$) denote the Hilbert spaces of measurable functions $\phi_j(X_j)$ with $E\{\phi^2(X_j)\} = 0$, $E\{\phi^2(X_j)\} < \infty$, and inner product $\langle \phi_j(X_j), \phi'_j(X_j) \rangle = E\{\phi_j(X_j)\phi'_j(X_j)\}$. H_Y is the corresponding Hilbert space of functions Y with $E\{\theta(Y)\} = 0$ and $E\{\theta^2(Y)\} = 1$. In addition, denote by H the space of arbitrary centered, square integrable functions of $X_1, \dots, X_p$. Furthermore, denote by $H_{add} \in H$ the linear subspace of additive functions: $H_{add} = H_1 + H_2 + \dots + H_p$. These are all subspaces of H_{YX} , the space of centered square integrable functions of Y and $X_1, \dots, X_p$.

Denote by P_j , P_Y , and P_{add} projection operators onto H_j , H_Y and H_{add} respectively. Then The P_j and P_Y are the conditional expectation operator $E(\cdot|X_j)$ and $E(\cdot|Y)$. Note that P_{add} is not a conditional expectation operator in this setting [4].

The optimization problem in this population setting is to minimize:

$$\begin{aligned} Obj &= E\{\theta(Y) - \bar{\phi}(\underline{X})\}^2 \\ E\{\theta^2(Y)\} &= 1 \\ \bar{\phi}(\underline{X}) &= \sum_{j=1}^p \phi_j(X_j) \in H_{add} \end{aligned} \tag{2.52}$$

Without the additivity restriction, the solution simply becomes:

$$\bar{\phi}(\underline{X}) = E\{\theta(Y) | X_1, X_2, \dots, X_p\} \tag{2.53}$$

We seek the closest additive approximation to this function. The minimizer $\bar{\phi}(\underline{X})$ of Equation 2.52 can be characterized by residuals $\theta(Y) - \bar{\phi}(\underline{X})$, which are orthogonal to the space of fits as in the parametric regression. The main difference from parametric regression is that this algorithm finds the projection space by itself (Figure 2.7).

That is,

$$\theta(Y) - \bar{\phi}(\underline{X}) \perp H_{add} \tag{2.54}$$

Equivalently,

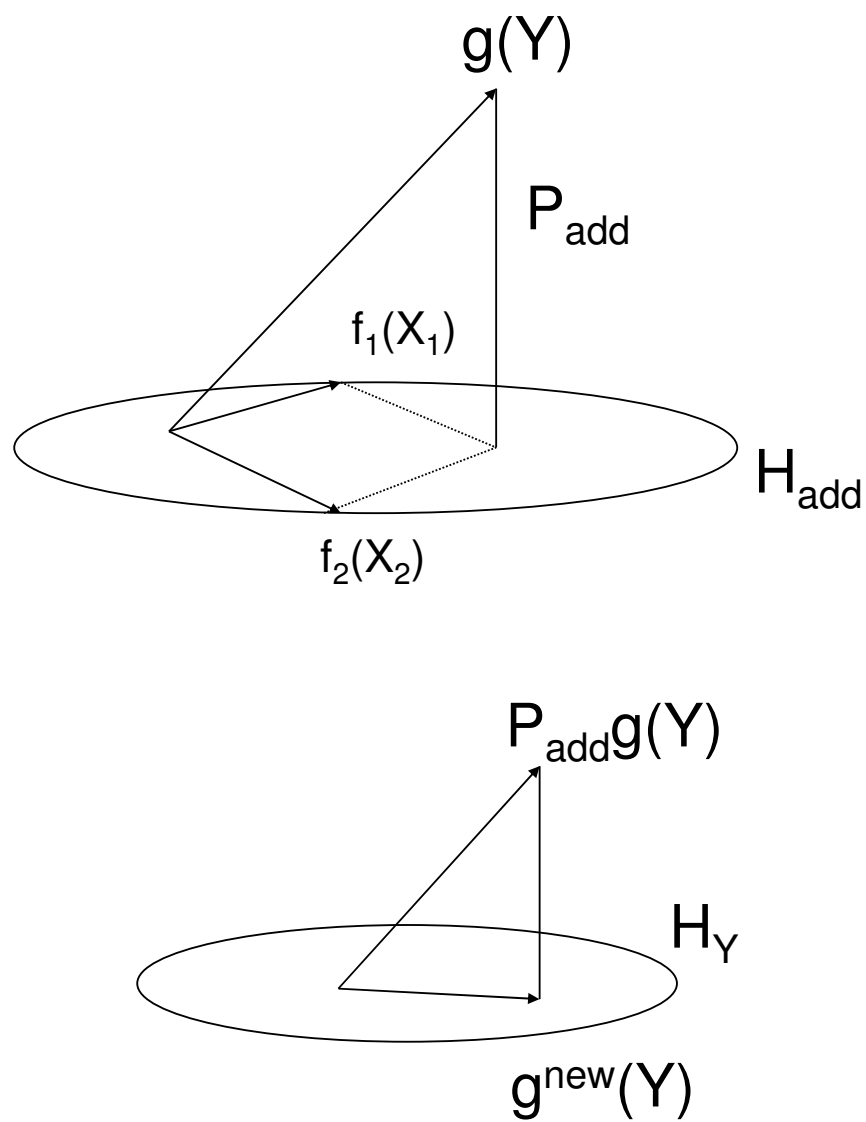


Figure 2.7: Hilbert space setting in ACE algorithm.

$$\theta(Y) - \bar{\phi}(\underline{X}) \perp H_j \quad (j = 1 \dots p) \quad (2.55)$$

By taking a projection of the residual onto the subspaces,

$$P_j(\theta(Y) - \bar{\phi}(\underline{X})) = P_j(\theta(Y) - \sum_{j=1}^p \phi_j(X_j)) = 0 \quad (j = 1 \dots p) \quad (2.56)$$

Since $P_j \phi_j(X_j) = \phi_j(X_j)$, component-wise this can be written as:

$$\phi_j(X_j) = P_j \left\{ \theta(Y) - \sum_{k \neq j} \phi_k(X_k) \right\} \quad (j = 1 \dots p) \quad (2.57)$$

Then the following normal equation is a necessary and sufficient condition for optimality for a fixed $\theta(Y)$.

$$\begin{pmatrix} I & P_1 & P_1 & \dots & P_1 \\ P_2 & I & P_2 & \dots & P_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_p & P_p & P_p & \dots & I \end{pmatrix} \begin{pmatrix} \phi_1(X_1) \\ \phi_2(X_2) \\ \vdots \\ \phi_p(X_p) \end{pmatrix} = \begin{pmatrix} P_1 \theta(Y) \\ P_2 \theta(Y) \\ \vdots \\ P_p \theta(Y) \end{pmatrix} \quad (2.58)$$

Breiman and Frieman [4] proved the row-wise updating of the solution in this normal equation converges to $P_{add} \theta(Y)$.

In practice, the conditional expectation operators P_j ($j = 1 \dots p$) are replaced by smoother matrix $\underline{S}_j$ ($j = 1 \dots p$).

Then, a data version of ACE can be written down as in Equation 2.59.

$$\begin{pmatrix} \underline{I} & \underline{S}_1 & \underline{S}_1 & \dots & \underline{S}_1 \\ \underline{S}_2 & \underline{I} & \underline{S}_2 & \dots & \underline{S}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \underline{S}_p & \underline{S}_p & \underline{S}_p & \dots & \underline{I} \end{pmatrix} \begin{pmatrix} \vec{\phi}_1 \\ \vec{\phi}_2 \\ \vdots \\ \vec{\phi}_p \end{pmatrix} = \begin{pmatrix} \underline{S}_1 \vec{\theta} \\ \underline{S}_2 \vec{\theta} \\ \vdots \\ \underline{S}_p \vec{\theta} \end{pmatrix} \quad (2.59)$$

Once iterative minimization converges to $P_{add} \theta(Y)$ (or the minimum norm), update $\theta(Y)$ by projecting residual onto H_Y .

$$P_Y(\theta(Y) - \phi(\underline{X})) = 0 \quad (2.60)$$

Equivalently,

$$\theta(Y) = E\left\{\sum_{j=1}^p \phi_j(X_j)|Y\right\} / \|E\left\{\sum_{j=1}^p \phi_j(X_j)|Y\right\}\| \quad (2.61)$$

In data space,

$$\theta(Y) = \underline{S}_Y\left\{\sum_{j=1}^p \vec{\phi}_j\right\} / \|\underline{S}_Y\left\{\sum_{j=1}^p \vec{\phi}_j\right\}\| \quad (2.62)$$

These two steps form the basis of the double loop algorithm. Breiman and Friedman proved convergence of this double loop algorithm in function space. In data space, a limited class of smoothers can be applied due to convergence issues. The conditional expectation operator is a projection operator, and thus projection smoother is one applicable smoother, although the problem then becomes a multivariate parametric regression problem. Breiman and Friedman [4] derived necessary conditions on the linear smoother property required for convergence. The condition is that the smoother be required to have the strictly shrinking property ($\|\underline{S}\underline{S}\| < \|\underline{S}\|$). For nonlinear smoothers, it is quite difficult to derive such a condition, since the smoother matrix depends on the data itself. In the original ACE algorithm, the super smoother [10], a nonlinear smoother, was employed based on experimental justification within practical applications.

2.4 Summary

In this chapter, various data smoothing techniques are reviewed including the multivariate version. In this study, these concepts are largely employed for data filtering and mining purposes. Specifically,

1. Filtering (smoothing) pressure data.
2. Finding break points in multitransient pressure data.
3. Identifying a reservoir model.
4. Estimating the flow rate.

These technical issues are mutually related and essentially inseparable. To achieve the goal of this study, we investigated and developed the data analysis method for pressure transient data in a nonparametric manner as described in later chapters.

Chapter 3

Single transient data smoothing

This study sought to develop a method to interpret and analyze long-term pressure data obtained from permanent down-hole gauges. It is important to fully utilize the advantage of such long term pressure data in order to extract quantitative information. A deconvolution approach is a natural candidate for that purpose, since quantitative information tends to be lost when each short-term transient is analyzed separately in a conventional manner.

Time domain deconvolution can be viewed as semiparametric regression in the sense that we describe a unknown response function in a nonparametric manner and enter it into a known convolution equation.

This chapter describes a smoothing algorithm suitable for pressure transient data, that was investigated and developed based on its characteristics.

3.1 Constrained smoother

In many of the applied sciences, it is common that the forms of an empirical relationship are almost completely unknown prior to study. Scatter plot smoothers used in nonparametric regression methods such as ACE algorithm [4] have considerable potential to ease the burden of model specification that a researcher would otherwise face in this situation.

Occasionally the researcher will know the information of the model, then such information should be included in the smoother property to obtain more reliable results with relative ease.

The convolution equation describes the pressure drop at time t as follows.

$$\Delta P(t) = \int_0^t Q'(u)K(t-u)du \quad (3.1)$$

Here $K(t)$ and $Q'(t)$ are the response function and the derivative of flow rate at time t .

In a discrete form, pressure drop at time t_i is given by:

$$\Delta P(t_i) = \sum_{j=1}^n a_j K(t_i - tb_j) \quad (3.2)$$

Here $a_j (= Q_j - Q_{j-1})$ and tb_j are effective flow rate and break point time for the j th transient.

In a deconvolution problem, this discretized convolution equation is fitted to the pressure data, in order to derive a response function. The difficulty is that we need to extract not only pressure but also its derivative with reasonable accuracy, since the pressure derivative is utilized in the reservoir model recognition process. Most existing methods have common oscillation problems in their derivative estimates due to the measurement error in the pressure and flow rate signals.

In order to achieve a reasonable solution, several authors [11, 20] implemented derivative constraints on the solution space. Hutchinson and Sikora [15] argued on physical grounds that the response function (reservoir model) should not only be positive and increasing, but also concave. Katz et al. [16] and Coats et al. [17] made this statement more precisely.

$$K \leq 0, \frac{dK}{dt} \leq 0, \frac{d^2K}{dt^2} \geq 0 \quad (3.3)$$

For single-phase, slightly compressible Darcy flow with initial equilibrium, these constraints were derived rigorously by Coats et al. [17], who showed that in this case there are sign constraints for derivative of any order, namely,

$$K \leq 0, \frac{d^{2n-1}K}{dt^{2n-1}} \leq 0, \frac{d^{2n}K}{dt^{2n}} \geq 0 \quad (3.4)$$

In the existing literature, up to second derivative constraints have been utilized to estimate the response function. Based on the results, their attempts are likely to help remove the unfavorable oscillation from its derivative estimates to some extent [11, 20].

3.1.1 Functional representation

In this section, we derive a functional representation with higher order derivative constraints imposed on a spline basis. For derivation of such a basis function, we start with a hat function defined as:

$$U_p(t) = \begin{cases} \frac{(t-k_{p-1})}{(k_p-k_{p-1})} & : t \in (k_{p-1}, k_p] \\ \frac{(t-k_{p+1})}{(k_p-k_{p+1})} & : t \in (k_p, k_{p+1}] \\ 0 & : \text{otherwise} \end{cases} \quad (3.5)$$

Here k_p is called a knot, a fixed point over the time domain. The function $U_p(t)$ is shaped like a tent, with the peak of the tent at knot k_p and the tent anchored at knots k_{p-1} and k_{p+1} (Figure 3.1). Note that $U_p(t)$ and $U_{p+1}(t)$ overlap on the interval (k_p, k_{p+1}) ; note also that $U_p(t)$ has negative slope and $U_{p+1}(x)$ has positive slope on this interval. Hence, the linear combination is nonnegative, continuous, and piecewise linear on (k_{p-1}, k_{p+1}) . Also the function can have either positive or negative slope on (k_p, k_{p+1}) depending on the relative magnitude of the coefficients. Thus one can approximate any nonnegative function with this linear combination if knots are assigned to the appropriate locations (Figure 3.2). The starting function is described as a linear combination of this basis with nonnegativity constraints.

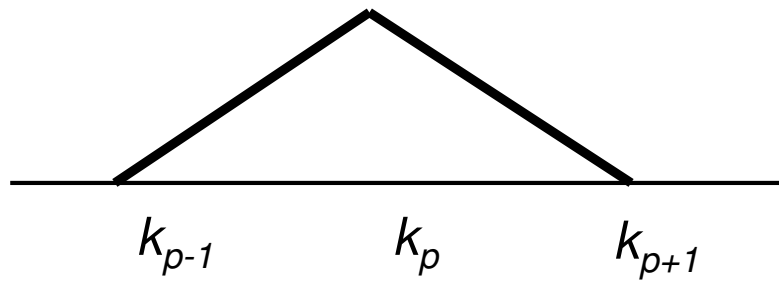
$$f(t) = \sum_{p=1}^m b_p U_p(t) \quad (3.6)$$

Here m is the number of the basis function. The corresponding number of knots is $m+2$. Under nonnegative constraints on the coefficients, the integral of basis function $U_p(t)$ gives a differentiable and nondecreasing function. The resulting function is piecewise quadratic over the interval (k_{p-1}, k_{p+1}) , and constant elsewhere (Figure 3.1). Thus integration of Equation 3.6 becomes:

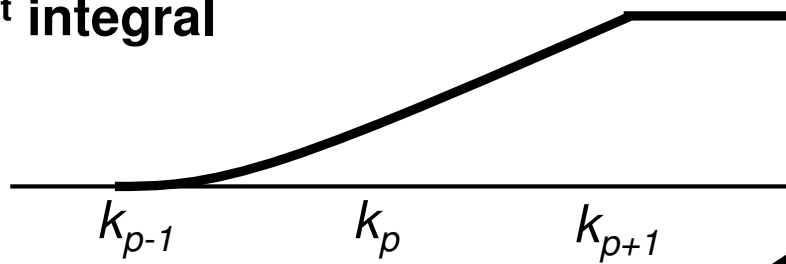
$$D^{-1}f(t) = \sum_{p=1}^m b_p D^{-1}U_p(t) + c_1 \quad (3.7)$$

Here D^{-1} is the integration operator and c_1 is an integral constant. Also further integration gives a twice differentiable and convex, and nondecreasing function, and the resulting function is piecewise cubic over the interval (k_{p-1}, k_{p+1}) and linear elsewhere (Figure 3.1).

original hat function



1st integral



2nd integral

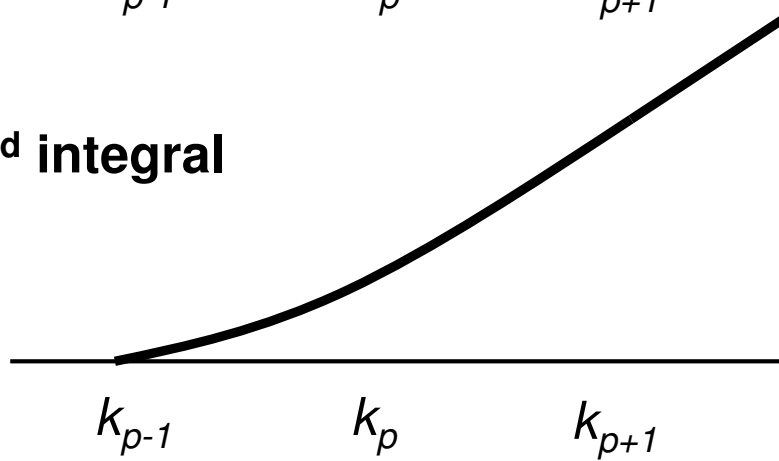


Figure 3.1: Hat function and its first and second integral.

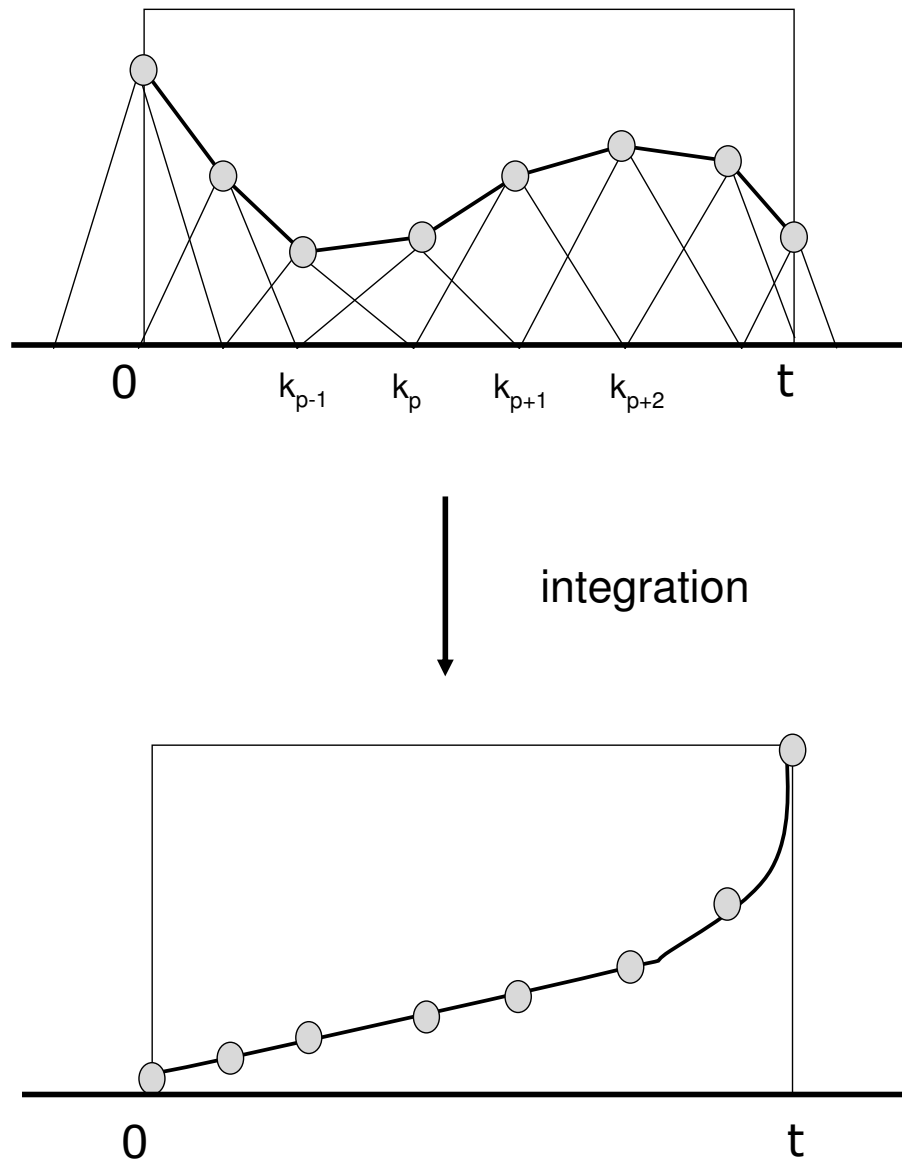


Figure 3.2: Function representation with hat functions and its integrated function.

$$D^{-2}f(t) = \sum_{p=1}^m b_p D^{-2}U_p(t) + c_1 t + c_2 \quad (3.8)$$

We can repeat this integration as many times as we want. n times integration gives the following basis function on the interval (k_{p-1}, k_{p+1}) .

$$D^{-n}U_p(t) = \begin{cases} 0 & : t \in (0, k_{p-1}] \\ \frac{(t-k_{p-1})^{n+1}}{(n+1)!(k_p-k_{p-1})} & : t \in (k_{p-1}, k_p] \\ \frac{(t-k_{p+1})^{n+1}}{(n+1)!(k_p-k_{p+1})} + \frac{(t-k_{p-1})^{n+1} - (t-k_p)^{n+1}}{(n+1)!(k_p-k_{p+1})} & : t \in (k_p, k_{p+1}] \\ \frac{(t-k_{p-1})^{n+1} - (t-k_{p+1})^{n+1}}{(n+1)!(k_p-k_{p-1})} & : t \in (k_{p+1}, \infty] \end{cases} \quad (3.9)$$

Here, the following mathematical formula for iterative integration is employed:

$$\begin{aligned} D^{-n}f(t) &= \int_0^t \int_0^{t_{n-1}} \dots \int_0^{t_2} f(t_1) dt_1 dt_2 \dots dt_n \\ &= \frac{1}{(n-1)!} \int_0^t (t-u)^{n-1} f(u) du \end{aligned} \quad (3.10)$$

The final expression can be written down in the following way:

$$D^{-n}f(t) = \sum_{p=1}^m b_p D^{-n}U_p(t) + \sum_{q=1}^n c_q T_q(t) \quad (3.11)$$

$$T_q(t) = t^{q-1} \quad (3.12)$$

Under the nonnegativity condition, this function satisfies nonnegative derivative of any order over the entire domain.

As shown in Equation 3.4, the exact derivative constraints of the response function is alternating in sign except in the first two constraints. To accommodate this condition and the zero crossing condition at the origin, one device is applied as shown in Figure 3.3. Namely, first shift the entire function such that it passes through 0 value at maximum time (t_{max}), and then flip right and left.

Zero crossing condition;

$$D^{-n}f(t_{max}) = \sum_{p=1}^m b_p D^{-n}U_p(t_{max}) + \sum_{q=1}^n c_q T_q(t_{max}) = 0 \quad (3.13)$$

Flip;

$$K(t) = D^{-n}f(t_{max} - t) = \sum_{p=1}^m b_p D^{-n}U_p(t_{max} - t) + \sum_{q=1}^n c_q T_q(t_{max} - t) \quad (3.14)$$

The resulting function based on Equations 3.13 and 3.14 satisfies the exact properties in Equation 3.4. For simplicity we express the final representation including the second term in the first term,

$$K(t) = \sum_{p=1}^M b_p S_p(t) \quad (3.15)$$

Here the number of parameter is reduced by one due to zero crossing condition ($M = m + n - 1$). The number of constraints and parameters are the same and increase by one if we increase the order of derivative constraints by one. So the original model complexities (degrees of freedom) increase if we set up the higher order derivative constraints.

3.1.2 Least square fitting

The functional representation in Equation 3.15 gives a function with any order of derivative condition under nonnegativity constraints. This fitting problem is written down simply as a constrained least squares problem.

$$\begin{aligned} RSS(\vec{b}) &= \frac{1}{2} \|\underline{H}\vec{b} - \vec{d}\|_2^2 \\ &= \frac{1}{2} \vec{b}^T \underline{H}^T \underline{H} \vec{b} - \vec{d}^T \underline{H} \vec{b} + \frac{1}{2} \vec{d}^T \vec{d} \\ &= \frac{1}{2} \vec{b}^T \underline{G} \vec{b} + \vec{c}^T \vec{b} + \frac{1}{2} \vec{d}^T \vec{d} \\ \underline{A} \vec{b} &\geq 0 \end{aligned} \quad (3.16)$$

where $\underline{H}$ is a design matrix ($N \times M$), $\vec{d}$ is a data vector ($N \times 1$), and $\vec{b}$ ($M \times 1$) is a parameter vector. $\underline{A}$ is a constraint matrix and for this particular problem it is the identity

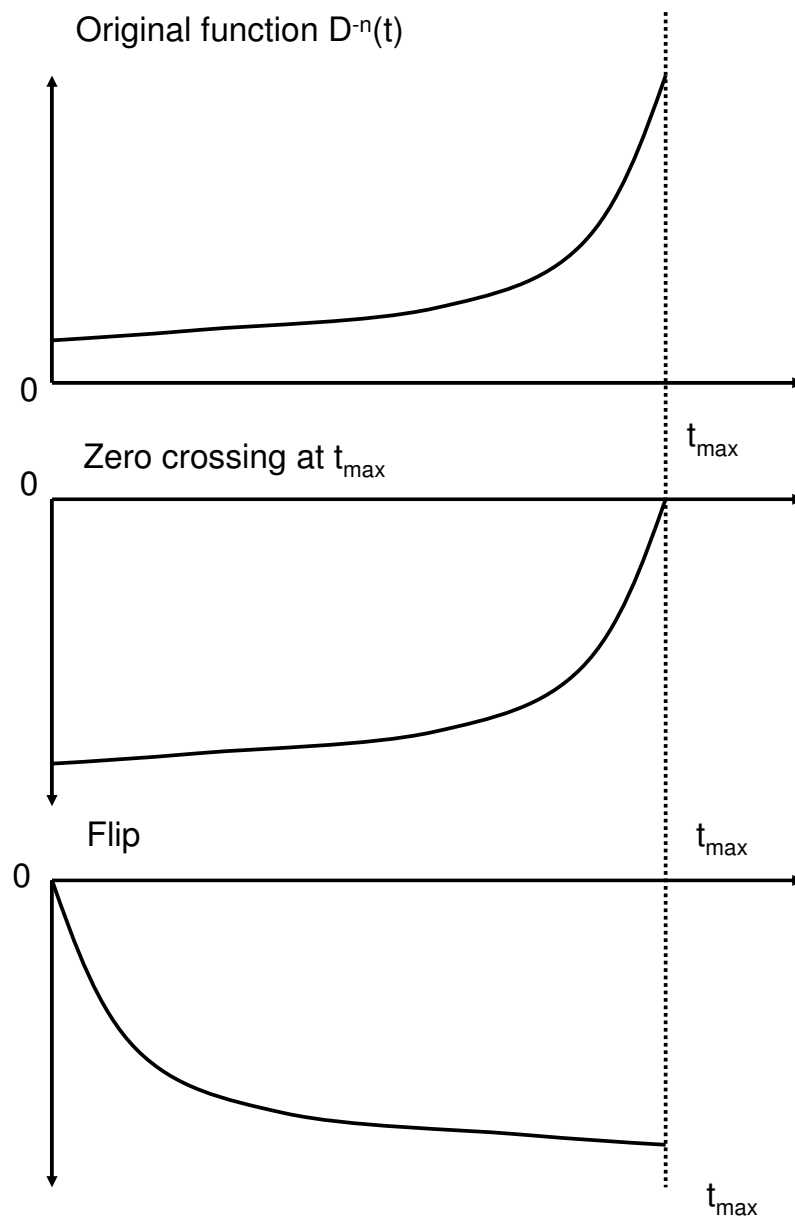


Figure 3.3: Functional representation. Upper: integrated function, Middle: shifted function, and Lower: flipped function.

matrix.

Without constraints, a solution is derived directly by:

$$\vec{b} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \vec{d} \quad (3.17)$$

To achieve nonnegative constraints, one constrained quadratic programming method called the active set method is employed. The active set method seeks the solution within a reduced subspace bounded by constraints starting with any feasible point $\vec{b}_0$. With Hessian $\underline{G}(= \underline{H}^T \underline{H})$ and gradient $\vec{g}(= \underline{G}\vec{b} + \vec{c})$, the first descent direction $\vec{p}$ is obtained by solving:

$$\bar{G}\vec{p} = -\vec{g} \quad (3.18)$$

If a new point $(\vec{b}_0 + \vec{p})$ hits one of the linear constraints, such a boundary is set up as an equality constraint. Then at the next step, the solution is searched in a reduced space. This process is repeated until a minimum solution is obtained (Figure 3.4). Let the active constraint matrix $\underline{A}_k(k \times M)$ include k active constraints. Then any feasible point $\vec{x}$ must satisfy:

$$\underline{A}_k \vec{x} = 0 \quad (3.19)$$

Let the column of the $(M \times M - k)$ matrix $\underline{Z}_k$ form $\underline{A}_k \underline{Z}_k = 0$. Then a feasible point can be described by $\vec{x} = \underline{Z}_k \vec{b}_k$ using any $(M - k \times 1)$ vector $\vec{b}_k$. Under k linear constraints, a problem is reduced to,

$$\begin{aligned} RSS(\vec{b}_k) &= \frac{1}{2} \|\underline{H} \underline{Z}_k \vec{b}_k - \vec{d}\|_2^2 \\ &= \frac{1}{2} \vec{b}_k^T \underline{Z}_k^T \underline{H}^T \underline{H} \underline{Z}_k \vec{b}_k - \vec{d}^T \underline{H} \underline{Z}_k \vec{b}_k + \frac{1}{2} \vec{d}^T \vec{d} \\ &= \frac{1}{2} \vec{b}_k^T \underline{Z}_k^T \underline{G} \underline{Z}_k \vec{b}_k + \vec{c}^T \underline{Z}_k \vec{b}_k + \frac{1}{2} \vec{d}^T \vec{d} \end{aligned} \quad (3.20)$$

A descent direction $\vec{p}_z$ is given by:

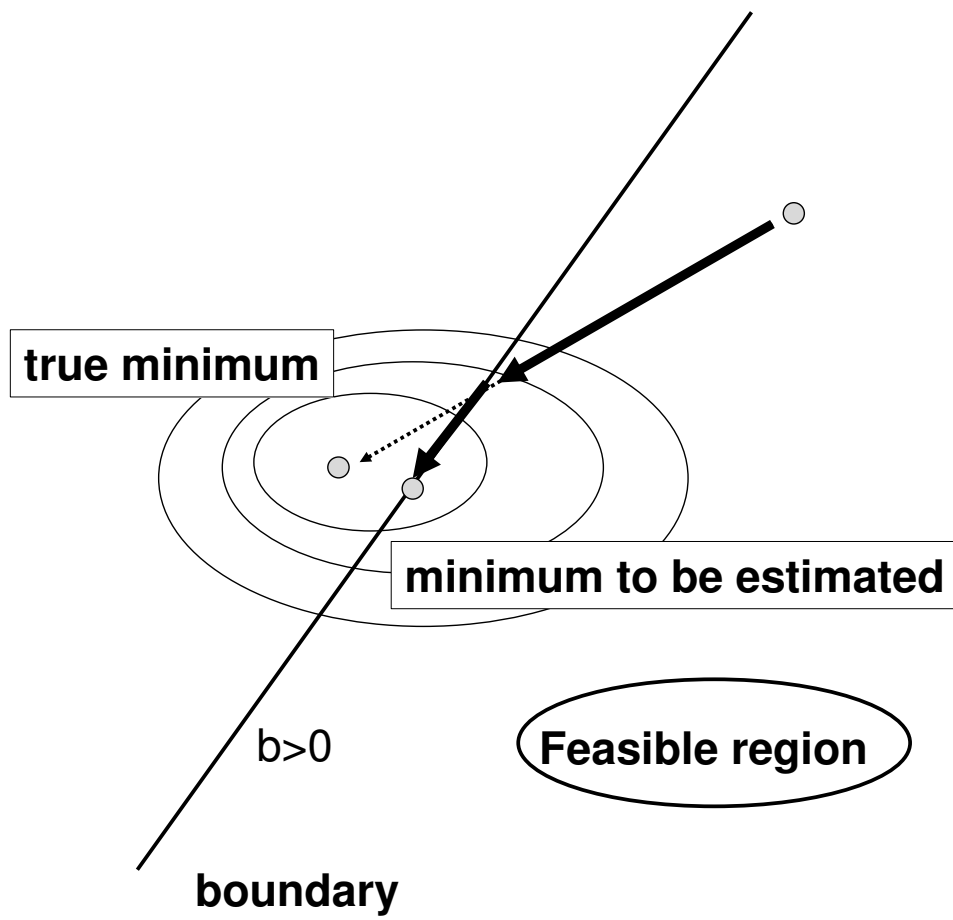


Figure 3.4: Schematic of active set method.

$$\underline{Z}_k^T G \underline{Z}_k \vec{p}_z = -\underline{Z}_k^T \vec{g} \quad (3.21)$$

$\underline{Z}_k^T G \underline{Z}_k \vec{p}_z$ and $\underline{Z}_k^T \vec{g}$ are called a projected Hessian and gradient respectively.

At the minimum point, the projected gradient must vanish.

$$\underline{Z}_k^T \vec{g} = 0 \quad (3.22)$$

This means gradient is a linear combination of row vector of $\underline{A}_k$.

$$\vec{g} = \sum_j^k \lambda_j \vec{a}_j^T = \underline{A}_k^T \vec{\lambda} \quad (3.23)$$

$\vec{\lambda}$ is called the Lagrange parameter (vector). At the minimum these Lagrange parameters must be nonnegative. This is because if one of the Lagrange parameters is negative, the descent direction is such that:

$$\begin{aligned} \vec{a}_j^T \vec{p} &= 1 \\ \vec{a}_i^T \vec{p} &= 0 (i \neq j) \end{aligned} \quad (3.24)$$

For such $\vec{p}$,

$$\vec{g}^T \vec{p} = \lambda_j \vec{a}_j^T \vec{p} = \lambda_j < 0 \quad (3.25)$$

Hence, $\vec{p}$ is a feasible direction. At every iteration, the Lagrange parameter is checked, and if a negative Lagrange parameter is found, the corresponding constraint is removed. This process is repeated until a minimum value is obtained without varying the number of constraints.

A solution vector $\vec{b}_k$ and $\vec{f}$ can be expressed as follows:

$$\vec{b}_k = (\underline{Z}_k^T \underline{H}^T \underline{H} \underline{Z}_k)^{-1} \underline{Z}_k^T \underline{H}^T \vec{d} \quad (3.26)$$

$$\vec{f} = \underline{H} \underline{Z}_k (\underline{Z}_k^T \underline{H}^T \underline{H} \underline{Z}_k)^{-1} \underline{Z}_k^T \underline{H}^T \vec{d} = \underline{S} \vec{d} \quad (3.27)$$

The matrix $\underline{S}$ is a reduced projection matrix ($\underline{S}\underline{S} = \underline{S}$) and is categorized as a nonlinear smoother, as matrix $\underline{Z}_k$ depends on data vector $\vec{d}$.

3.1.3 Knot insertion and preliminary check

In the previous sections, we derived a functional representation of the reservoir model. This formulation satisfies the derivative constraints of any order and is expected to approximate any reservoir model behavior if the knot location is chosen appropriately. In general, one can put the knots at every data location, but this is unfavorable due to the increased model complexities and also for the computational burden. But fewer knots tends to give less approximation quality.

One issue in the functional representation is that if we have a wellbore storage in the early time region, the higher derivative constraint (higher than second derivative) does not hold any longer in a strict sense. A storage-dominated region exhibits a unit slope in the log-log derivative plot, where pressure drop is proportional to time.

As a preliminary check, the formulation and knot insertion scheme was tested for three reservoir models; infinite-acting radial flow model, dual porosity model, and closed boundary model. Figure 3.5 shows an example of fitting results. In the figure, the circles show the estimates at the inserted knot locations.

After several experiments, we decided to put the knots at equal log intervals over the time domain. Specifically, starting with the end point (maximum time), we put the knots at equal log interval and stop the knot placement at the time less than 0.001 hrs. The starting knot is placed at time 0 and one additional knot is placed as the ending knot using the same log interval outside of the target time domain.

As shown here, less than 0.3 log interval yielded sufficient approximation quality with any order of derivative constraints. All the models are fitted well to the true solution.

Throughout this study, we employed this functional representation with the knot insertion strategy.

3.1.4 Smoothing effect of derivative constraints

We applied the constrained smoother to noisy data sets in order to investigate its smoothing effect.

We added Gaussian noise of zero mean and various standard deviations to synthetic data sets. In this study, the pressure error E_p was defined as $\frac{\|\vec{\varepsilon}_p\|}{\|\vec{\Delta P}\|}$.

Here, $\vec{\varepsilon}_p$ and $\vec{\Delta P}$ are pressure error and drawdown vectors. Figure 3.6 shows an example synthetic data for 2 % error case. In this example, 2 % error means the standard deviation of 0.78 *psi* for the drawdown 55 *psi*.

Figures 3.7 show an example of fitting results of the constrained smoother and unconstrained cubic regression splines. As can be seen, derivative constraints suppress the noise effect to produce better estimates as the order of derivative constraints increases.

To measure the effectiveness of derivative constraints, MSE (mean squared error) is a natural choice. MSE is defined and decomposed into bias and variance component as is well known [14].

$$\begin{aligned}
 MSE(\vec{f}) &= E\{\vec{f} - \vec{f}\}^2 \\
 &= E\{\vec{f} - E(\vec{f})\}^2 + E\{E(\vec{f}) - \vec{f}\}^2 \\
 &= \{variance\}^2 + \{bias\}^2
 \end{aligned} \tag{3.28}$$

Here $\vec{f}$ and $\vec{f}$ are the estimated signal and true one respectively. Variance is defined as the mean squared difference between each estimate and the mean and bias is defined as the difference between the average estimated signal and true one. We estimated the MSE through numerical simulation with 50 realizations. Figures 3.8 and 3.9 show MSE, bias and variance of pressure and pressure derivative for each derivative constraints in the three reservoir models. Here pressure derivative means pressure derivative multiplied by time ($t \frac{dp}{dt}$).

As the order of derivative constraints increases, MSE and variance decrease while bias is well suppressed, although improvement of prediction quality becomes gradual after the eighth derivative.

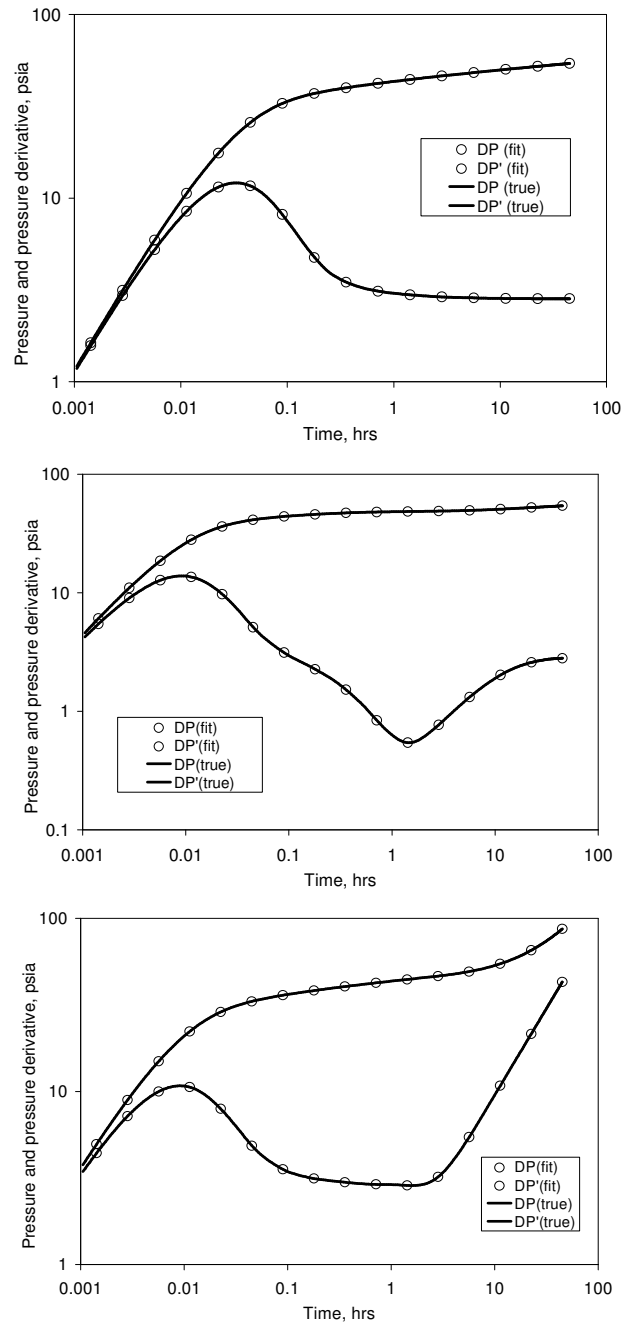


Figure 3.5: Example fitting results with 0.3 log interval (sixth order derivative constraints). Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model. open circle: estimated derivative and solid line: true solution.

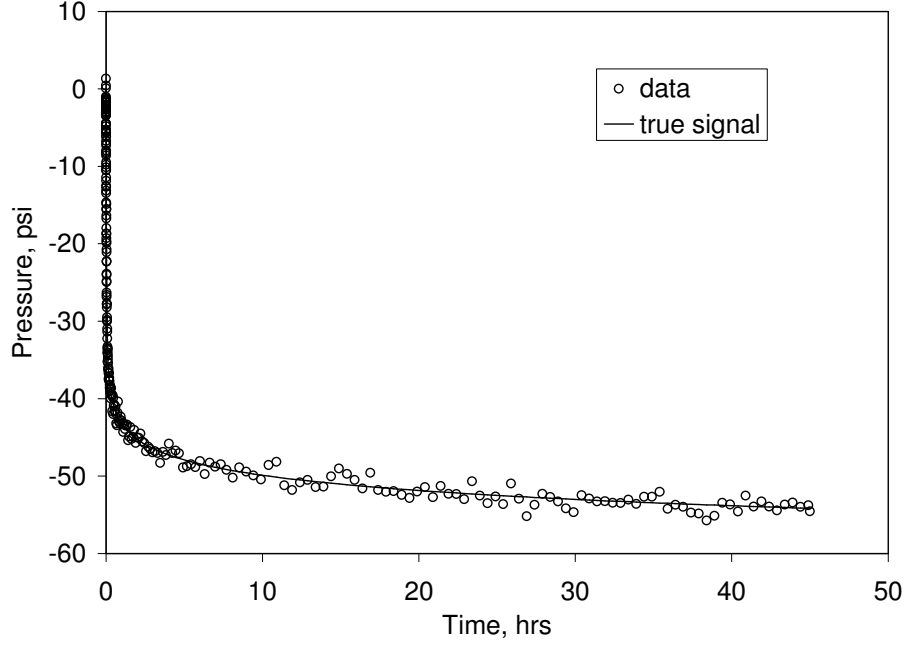


Figure 3.6: An example synthetic data for 2 % error case.

Figure 3.10 shows average degrees of freedom ($trace(S)$) of the smoother. The concept of degrees of freedom is often utilized to measure the model complexity (effective number of parameters) in a nonparametric model [14].

As can be seen in Figure 3.10, the derivative constraints decrease the degrees of freedom effectively and thus the smoothness is achieved.

Figure 3.11 shows the MSE for the various noise level cases with eighth order derivative constraints. Black and white marks represent the cases with and without the derivative constraints respectively. Figure 3.12 shows an example derivative estimate for one realization.

As shown in Figures 3.11 and 3.12, the derivative constraints work to suppress noise effects to some extent, but for the higher noise level large oscillation becomes conspicuous in the estimates.

The derivative constrained formulation considered here achieved partial success to remove unfavorable oscillation in the resulting estimates, but further smoothing control is required for the model recognition.

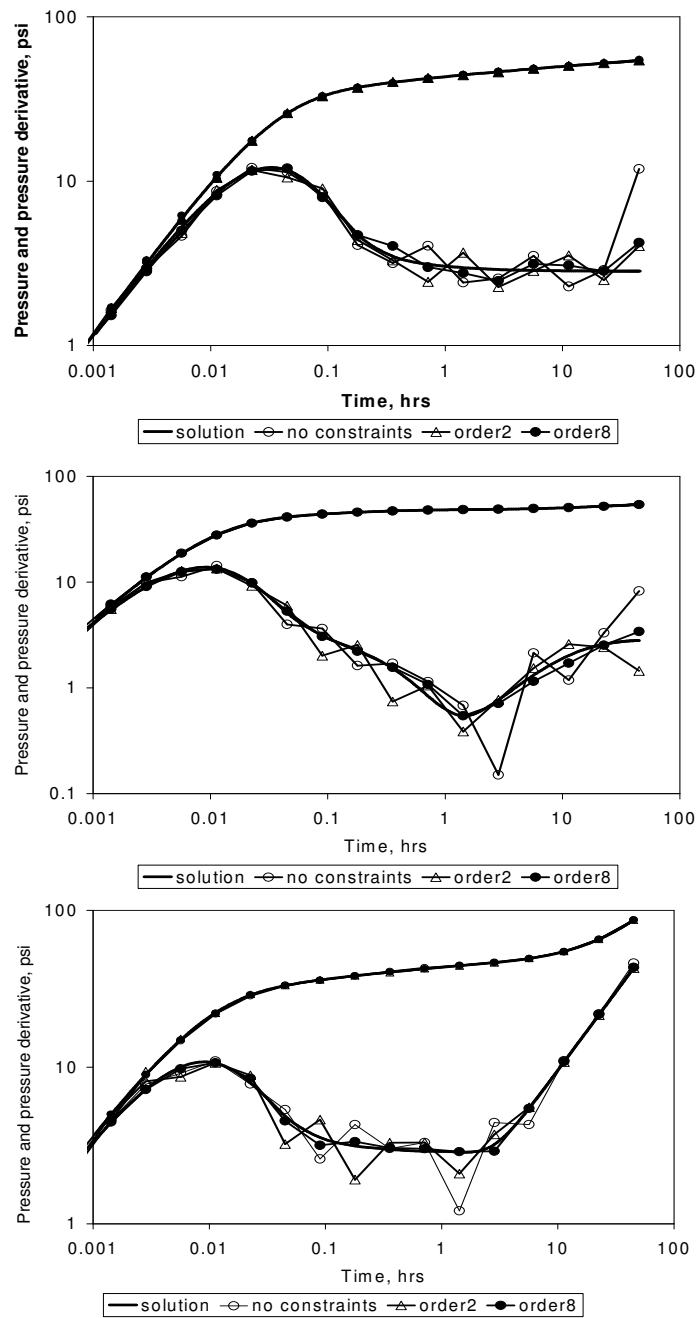


Figure 3.7: Pressure derivative estimates with higher order derivative constraints. Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model.

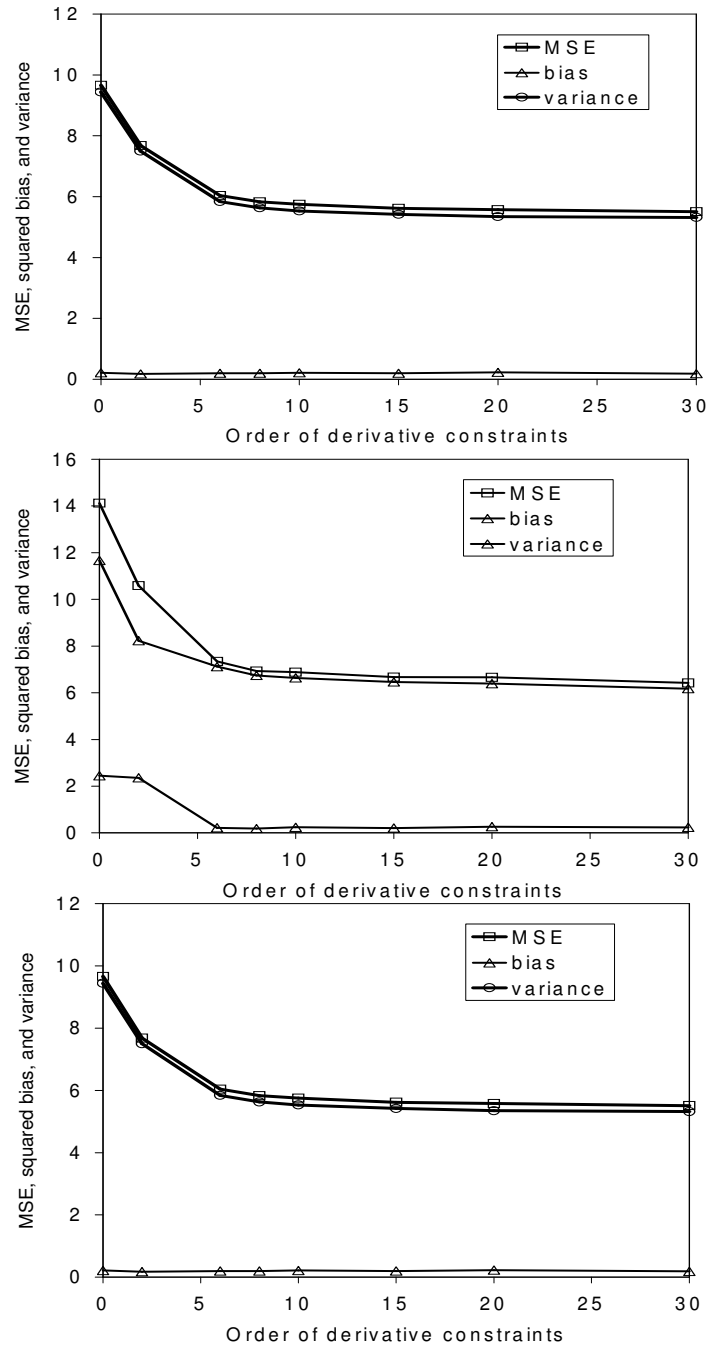


Figure 3.8: MSE, bias, and variance (ψ^2) for pressure estimate with higher order derivative constraints. Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model.

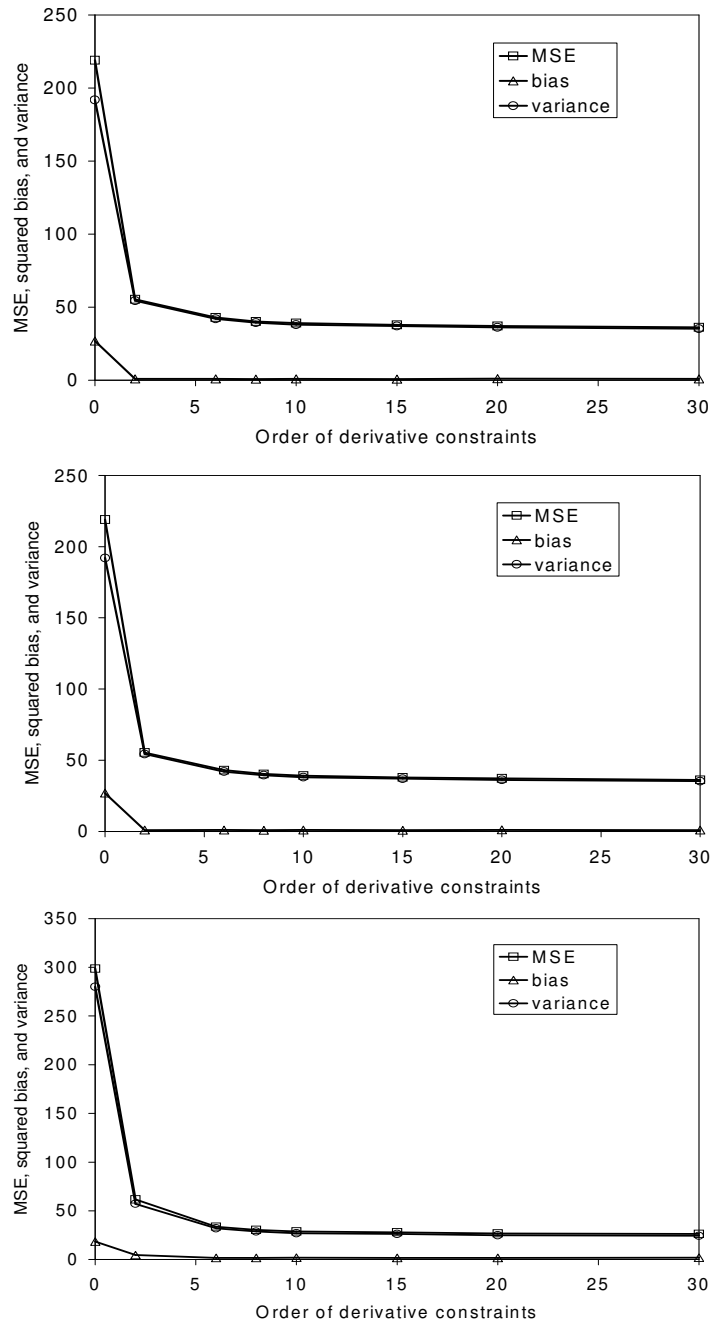


Figure 3.9: MSE, bias, and variance (ψ^2) for pressure derivative estimate with higher order derivative constraints. Upper: infinite-acting radial flow model, Middle: dual porosity model, and Lower: closed boundary model.

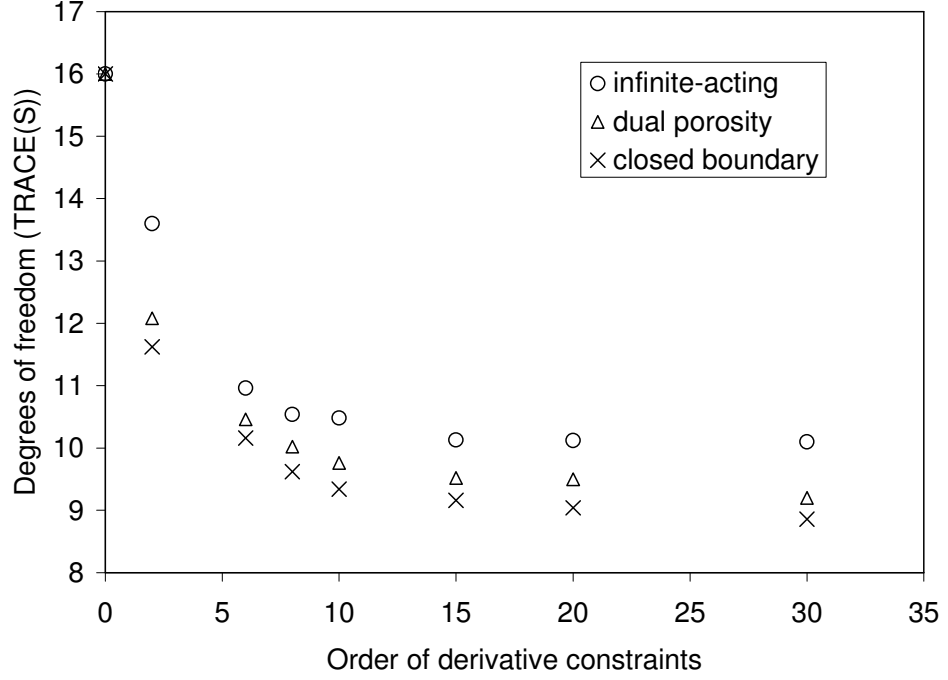


Figure 3.10: Degrees of freedom of the smoother.

3.2 Smoothing control

It is important to remove the oscillation in the estimate in order to achieve interpretable results. Here we try to control the smoothness of the estimates by imposing a roughness penalty on the solution based on the curvature of pressure derivative in the log-log plot.

Specifically, roughness is defined as the slope change (finite second derivative) between consecutive control points. Control points are set at the exact knot locations and their middle points in the logarithm of time (Figure 3.13).

So the objective function becomes:

$$\begin{aligned}
 Obj(\vec{b}) &= \frac{1}{2} \|\underline{H}\vec{b} - \vec{d}\|_2^2 + \lambda \sum_{r=1}^{N_c} R_r^2(\vec{b}) \\
 \underline{A}\vec{b} &\geq 0
 \end{aligned} \tag{3.29}$$

Here $R_r(\vec{b})$ is the roughness at the r th control point defined as:

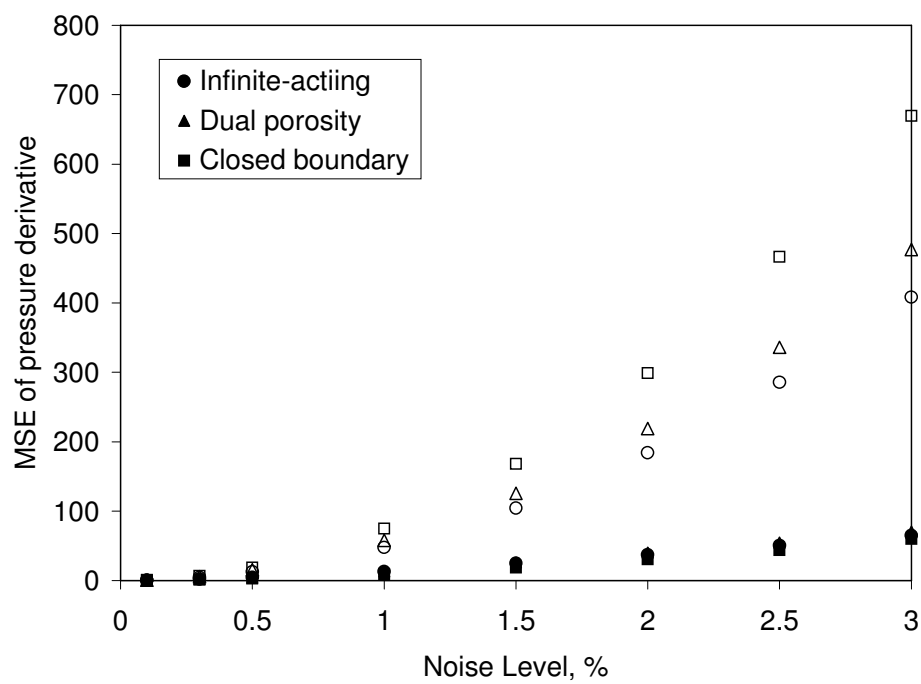


Figure 3.11: MSE of pressure derivative (psi^2) for various noise levels. Open marks: w/o derivative constraints.

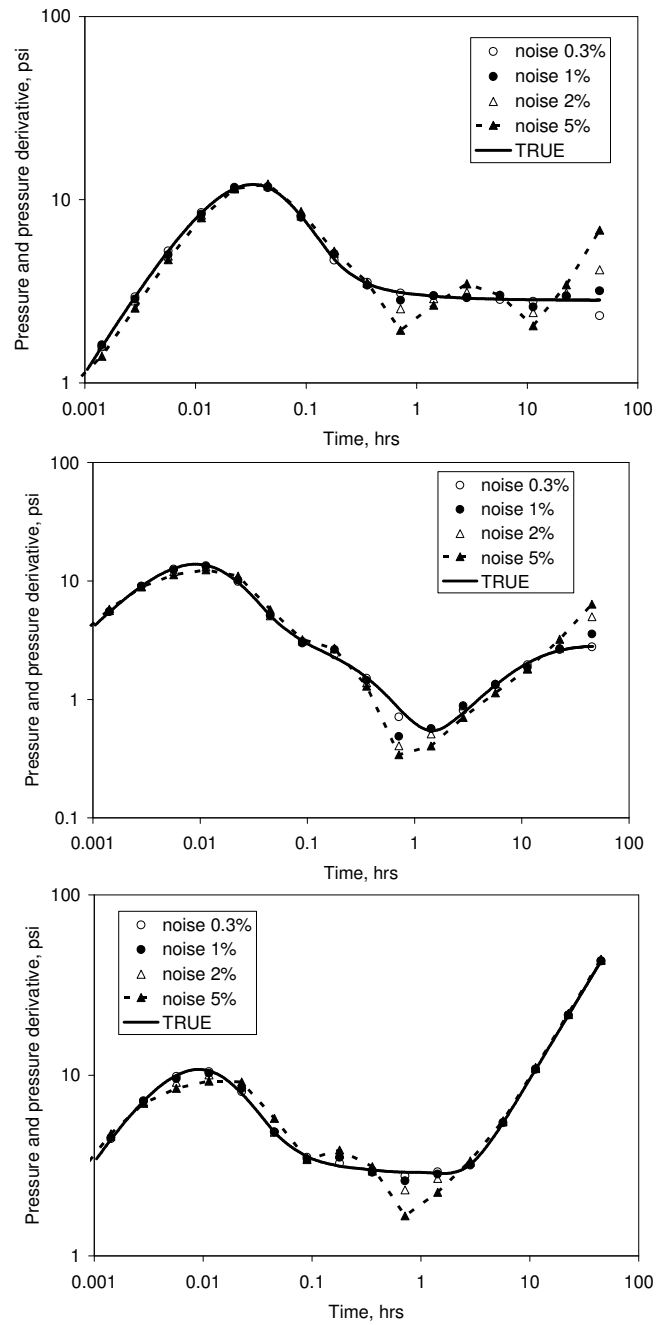


Figure 3.12: Pressure derivative estimates for various noise levels.

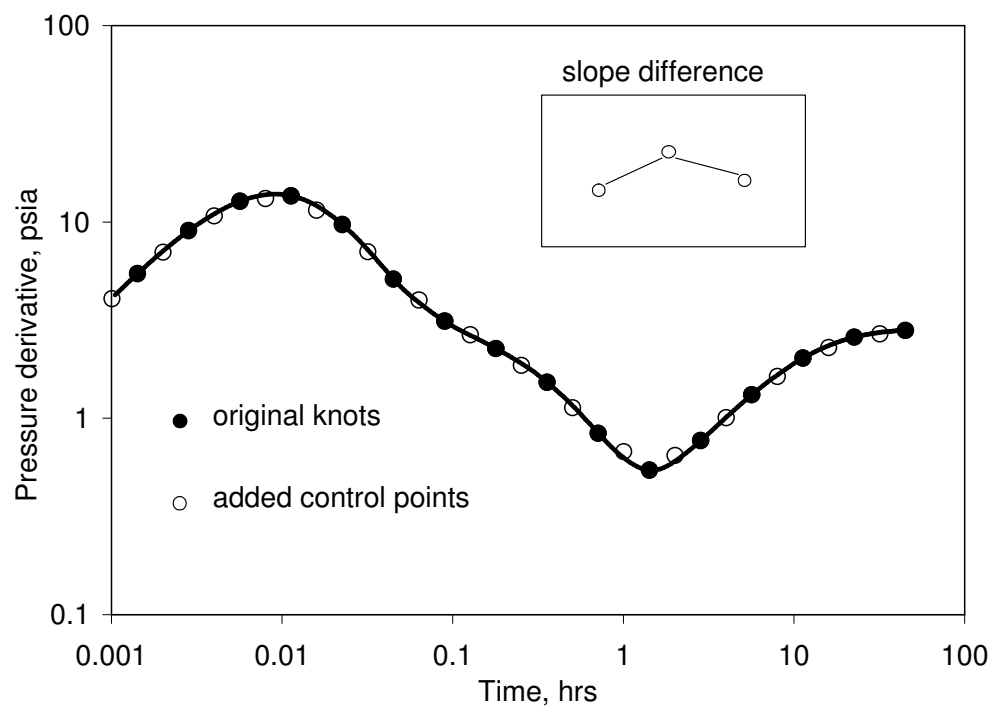


Figure 3.13: Definition of roughness and its control points.

$$\begin{aligned}
R_r(\vec{b}) = & \log(t_{r-1} \sum_{p=1}^M b_p S'_p(t_{r-1})) - 2\log(t_r \sum_{p=1}^M b_p S'_p(t_r)) \\
& + \log(t_{r+1} \sum_{p=1}^M b_p S'_p(t_{r+1}))
\end{aligned} \tag{3.30}$$

This formulation is a nonlinear minimization problem with nonnegativity constraints. The active set method was combined with the Levenberg-Marquardt algorithm to solve this problem. The quadratic approximation to the second nonlinear term yields:

$$\begin{aligned}
Obj(\vec{b}) & \approx \frac{1}{2} \vec{b}^T \underline{H}^T \underline{H} \vec{b} - \vec{d}^T \underline{H} \vec{b} + \frac{1}{2} \vec{d}^T \vec{d} + \frac{\lambda}{2} (\vec{b}^T \underline{X} \vec{b} + \vec{x}^T \vec{b}) \\
& = \frac{1}{2} \vec{b}^T (\underline{H}^T \underline{H} + \lambda \underline{X}) \vec{b} - (\vec{d}^T \underline{H} - \vec{x}^T) \vec{b} + \frac{1}{2} \vec{d}^T \vec{d} \\
& = \frac{1}{2} \vec{b}^T \underline{G} \vec{b} + \vec{c}^T \vec{b} + \frac{1}{2} \vec{d}^T \vec{d}
\end{aligned} \tag{3.31}$$

With the Hessian ($\underline{G}$) and gradient ($\vec{g} = \underline{G}\vec{b} + \vec{c}$), we search the solution with nonnegativity constraints as in the linear least square problem mentioned earlier.

At each iteration, we minimize the quadratic equation using the active set method then check the solution decrease the objective function. If this is not the case, we increase the diagonal term of the Hessian matrix by:

$$\underline{G}' = \underline{G} + w\underline{I} \tag{3.32}$$

Here w is a damping parameter and $\underline{I}$ is an identity matrix.

Suppose we have k active constraints under the damping parameter w at the final stage, the solution and the estimate $\vec{f}$ can be written down as:

$$\vec{b}_k = (\underline{Z}_k^T \underline{G}' \underline{Z}_k)^{-1} \underline{Z}_k^T (\underline{H}^T \vec{d} - \vec{x}^T) \tag{3.33}$$

$$\vec{f} = \underline{H}\underline{Z}_k(\underline{Z}_k^T \underline{G}' \underline{Z}_k)^{-1} \underline{Z}_k^T (\underline{H}^T \vec{d} - \vec{x}^T) \quad (3.34)$$

The main difference between the linear and nonlinear cases is that the additional term appears in the solution ($\vec{x}^T$).

We applied the smoothing control to the same 50 realizations. Figure 3.14 shows MSE, bias, and variance of pressure and pressure derivative for various λ values in the case of the infinite-acting radial flow model. Here the original MSE means the case without smoothing control. As can be seen, if λ increases, the MSE decreases with decreasing variance and increasing bias. Although MSE gradually decreases, it starts to increase again at the minimum point because of the significant bias. The magnitude of λ controls the smoothness of the curve between the two extreme cases. If λ is 0, this minimization gives the original fitting under derivative constraints. If λ is ∞ , this yields a straight line in the log-log plot. As can be seen, this roughness penalty introduces bias in the estimates, and thus its magnitude should be controlled.

To determine the best value of control parameter λ , commonly cross validation (CV) is utilized. Cross validation separates the entire data into training data sets and the corresponding validation data sets. The procedure repeats the model fitting to the training data sets and predicting validation data sets to estimate prediction error for several λ values. However, the direct calculation of prediction error (CV score) requires prohibitive computation. For example, if we employ the direct leave-one-out cross validation, where we leave out one single data point and fit the model to the remaining $n - 1$ data and predict the data left out, it requires n time model fitting to estimate CV score.

To avoid this computational burden, Wahba et al. [12] proposed the generalized cross validation (GCV) to select the smoothing parameter λ . GCV can be expressed using degree of $DF(= \text{trace}(\underline{S}))$ as:

$$\begin{aligned} GCV(\lambda) &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{y_i - \hat{f}_\lambda(x_i)}{1 - \text{trace}(\underline{S}_\lambda)/n} \right\}^2 \\ &= \frac{n}{n - DF} \frac{\sum_{i=1}^n (y_i - \hat{f}_\lambda(x_i))^2}{n - DF} \end{aligned} \quad (3.35)$$

The GCV criterion is a sum of squared error discounted for the degrees of freedom

invested in the model. GCV was originally proposed for linear smoothers, but it has been extended for the nonlinear smoothers as well [22].

From Equation 3.34, we define an approximate smoother matrix $\underline{S}$ as:

$$\underline{S} = \underline{H}\underline{Z}_k(\underline{Z}_k^T \underline{G}' \underline{Z}_k)^{-1} \underline{Z}_k^T \underline{H}^T \quad (3.36)$$

By approximating the degrees of freedom by the trace of this matrix, we tested this GCV criteria for our smoothing parameter selection.

Figure 3.15 shows the GCV profile for one realization. As can be seen, GCV score has a similar shape to the MSE profile and selects a reasonable level of smoothing parameters at its minimum value. The same procedure was repeated for 50 realizations with the same noise level. Although the selected λ values ranged from 0.025 to 2.56 depending on the realization, the median, 25 %, and 75% quartile are 1.28, 0.32, and 1.28. Overall, a reasonable level of the smoothing parameter λ could be selected. From a practical point of view, this GCV is much more efficient compared with other direct cross validation techniques, where we need to fit the model K times to obtain the CV score for one single smoothing parameter. Throughout this study, we used this GCV for any control parameter selection.

Figures 3.16, 3.17, and 3.18 show the pressure derivative estimates for the 2% error case for various λ values including the cross validated value.

For small λ , the oscillation is still observed but for large λ , oversmoothing (bias) becomes dominant in all cases. The cross validated results gave reasonable estimates.

To check the effectiveness of this smoothing control, again MSE, bias, and variance were estimated for the same 50 realizations (Figure 3.19). The black and white marks show the cases with and without smoothing control respectively. Figure 3.20 shows an example derivative estimate for various noise levels.

As can be seen, the smoothing control effectively reduces the MSE and makes it easier to interpret the results. However for the higher noise case, locally poor performance was observed. Figures 3.21 to 3.23 show the local MSE, bias, and variance of pressure derivative estimates.

The local balance between bias and variance is different in all the models. For example, in the case of infinite-acting radial flow, increasing the smoothing parameter has the tendency that the mountain is cut off and the valley is filled around the storage-dominated

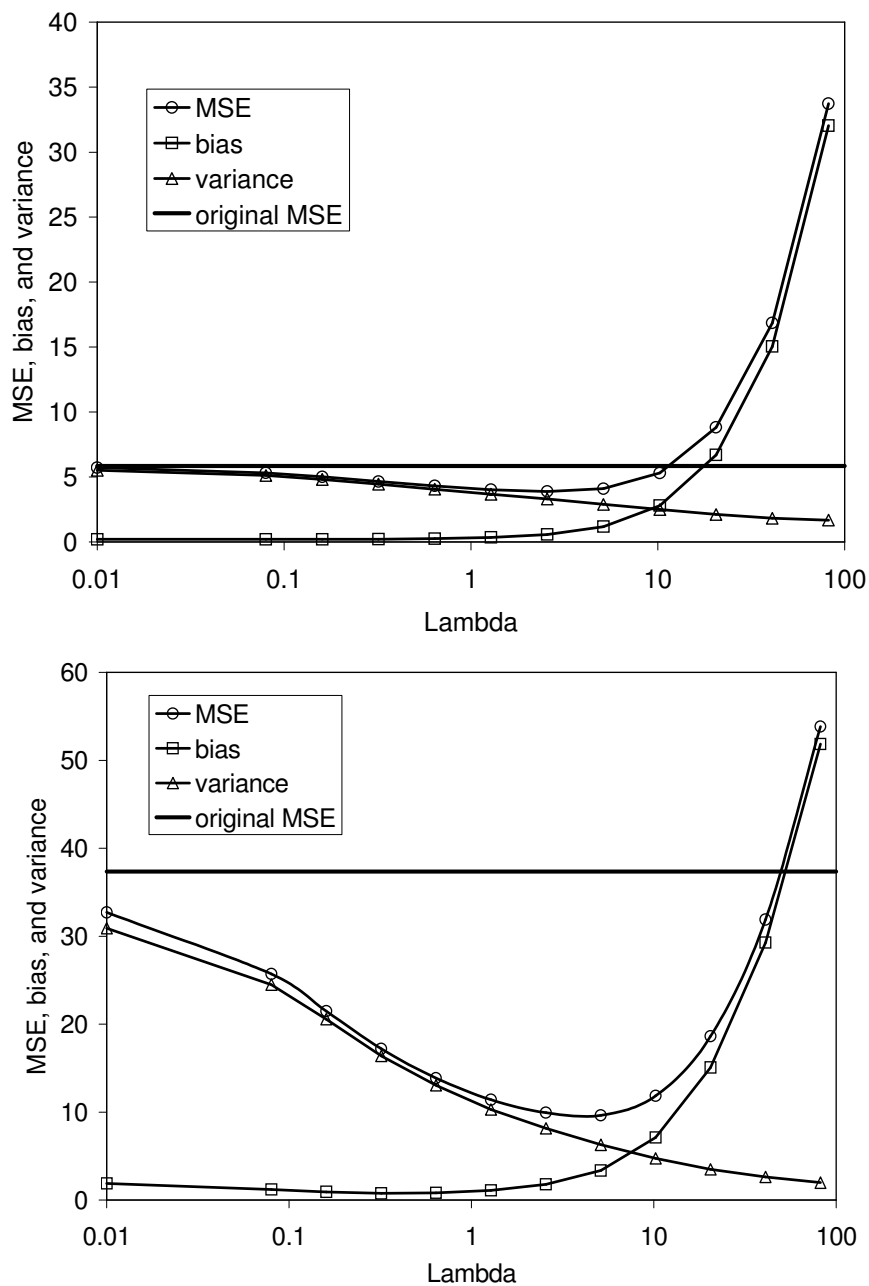


Figure 3.14: MSE, bias, and variance for various smoothing parameters (model: infinite-acting radial flow for noise level 2 %). Horizontal line shows the MSE value without smoothing control.

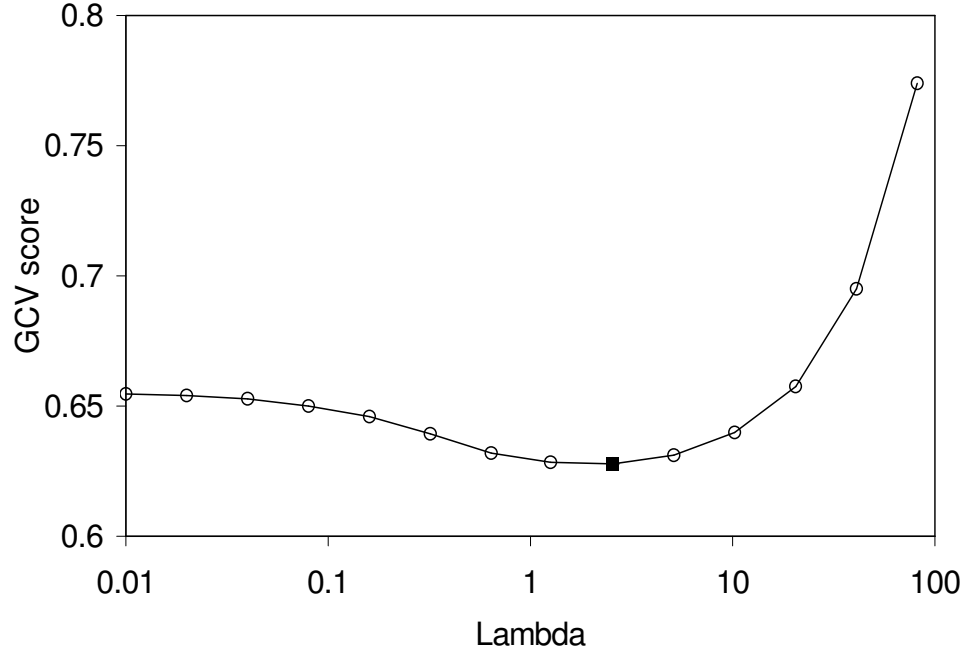


Figure 3.15: GCV score for various smoothing parameters.

region, but it decreases the MSE around the end point. Similar phenomena were observed in the other models.

This indicates that the smoothing control using a single parameter has a limited ability to control the local complexities of the model.

3.3 Multiple smoothing parameter control

The results shown in the last section prompted us to investigate the multiple smoothing parameter control in order to further reduce overall MSE by improving the local balance between bias and variance. To control the local balance we changed the smoothing parameter values at every control point. Figure 3.24 shows the concept.

Then the formulation becomes:

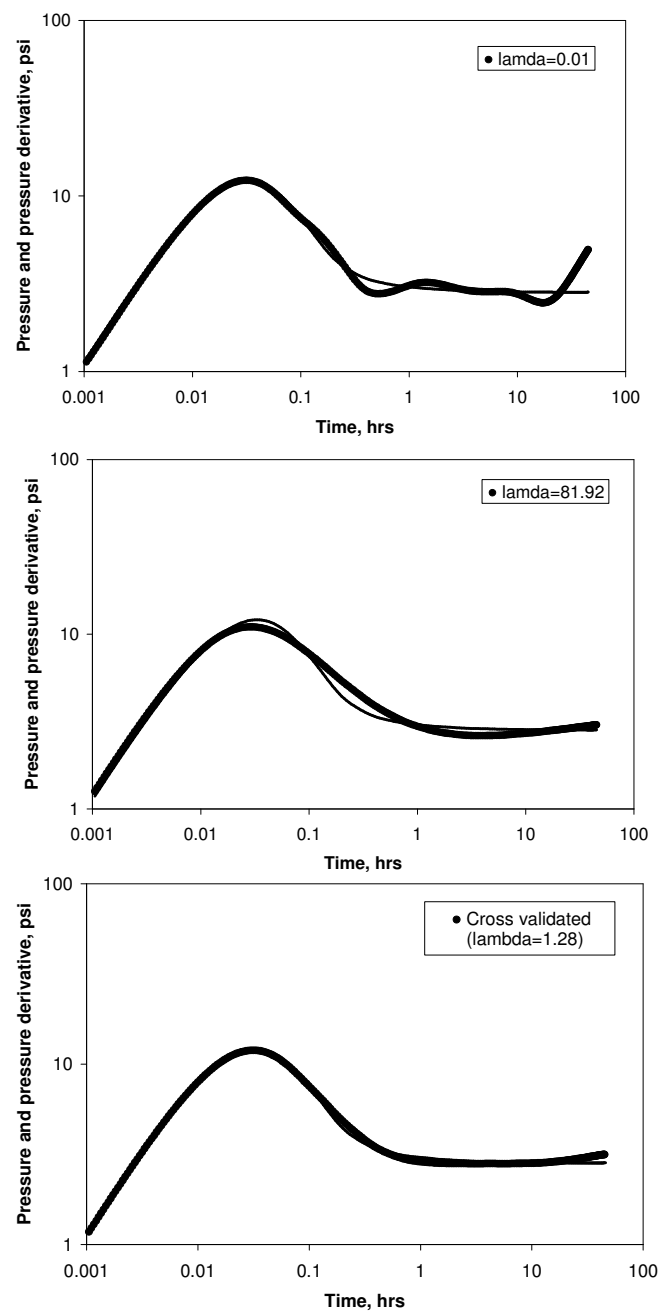


Figure 3.16: The effect of smoothing parameter (infinite-acting radial flow model). Thin line: true solution and bold line: the estimate.

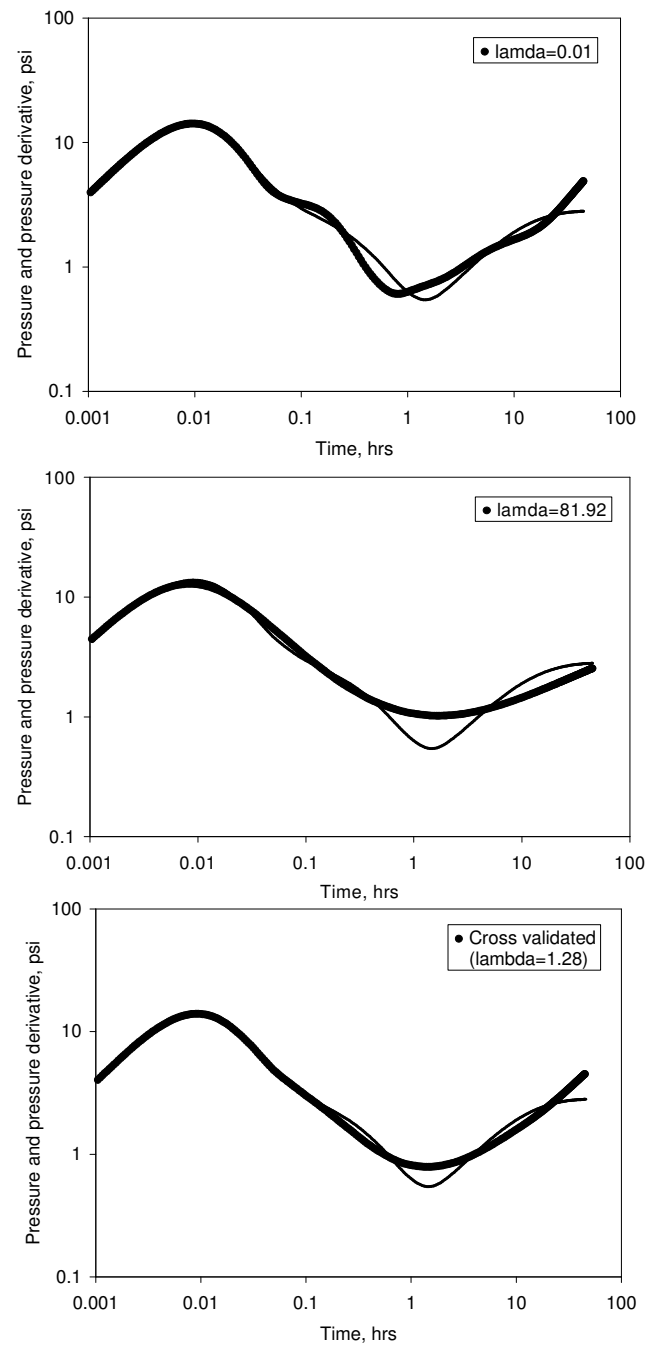


Figure 3.17: The effect of smoothing parameter (dual porosity model). Thin line: true solution and bold line: the estimate.

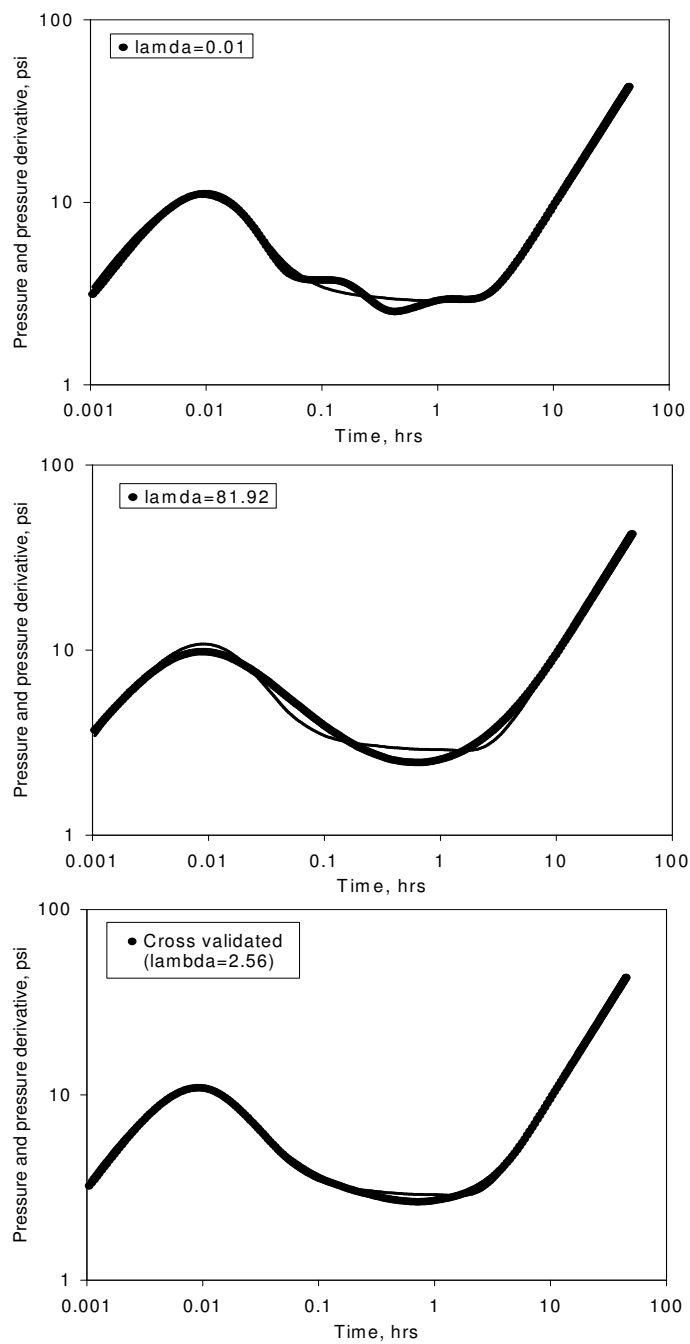


Figure 3.18: The effect of smoothing parameter (closed boundary). Thin line: true solution and bold line: the estimate.

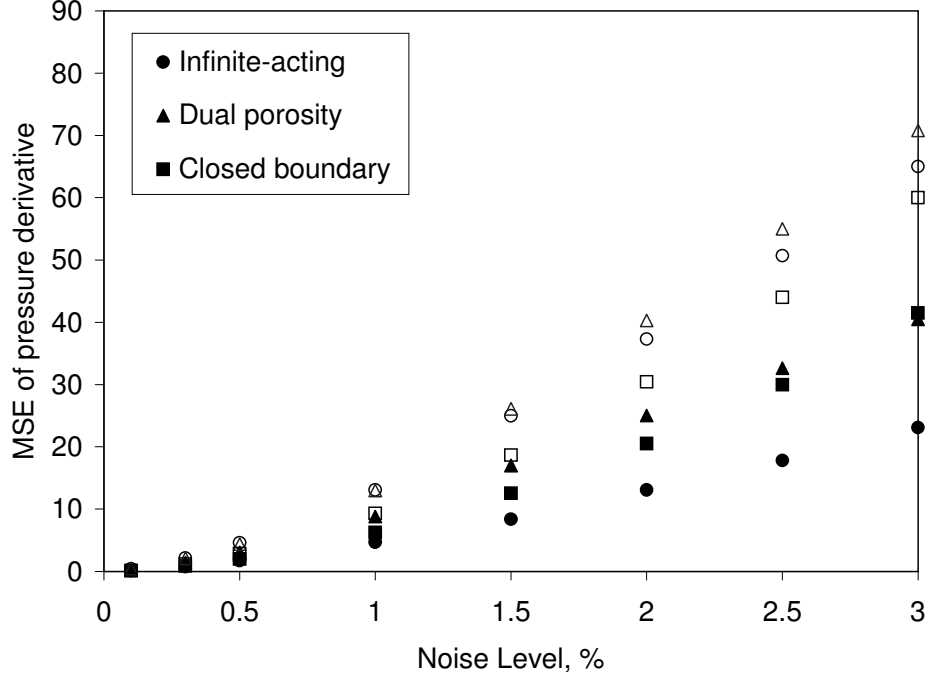


Figure 3.19: MSE of pressure derivative (psi^2). open marks: w/o smoothing control.

$$\begin{aligned}
 Obj(\vec{b}) &= \frac{1}{2} \|\underline{H}\vec{b} - \vec{d}\|_2^2 + \sum_{r=1}^{N_c} \lambda_r R_r^2(\vec{b}) \\
 \underline{A}\vec{b} &\geq 0
 \end{aligned} \tag{3.37}$$

In this formulation, we expressed the λ profile along the time domain with cubic splines by assuming the smoothing control parameters vary smoothly. Namely,

$$\lambda_r = \sum_{p=1}^{N_\lambda} C_p S_p(r) \tag{3.38}$$

$S_p(r)$ and C_p are the cubic spline basis and the coefficient. Here we used the same cubic spline expression developed for the smoother without the nonnegativity constraints. N_λ is the number of knots to represent λ profile and determines the amount of variation of λ values. In Figure 3.24, the black circles show the knot locations to represent the λ profile.

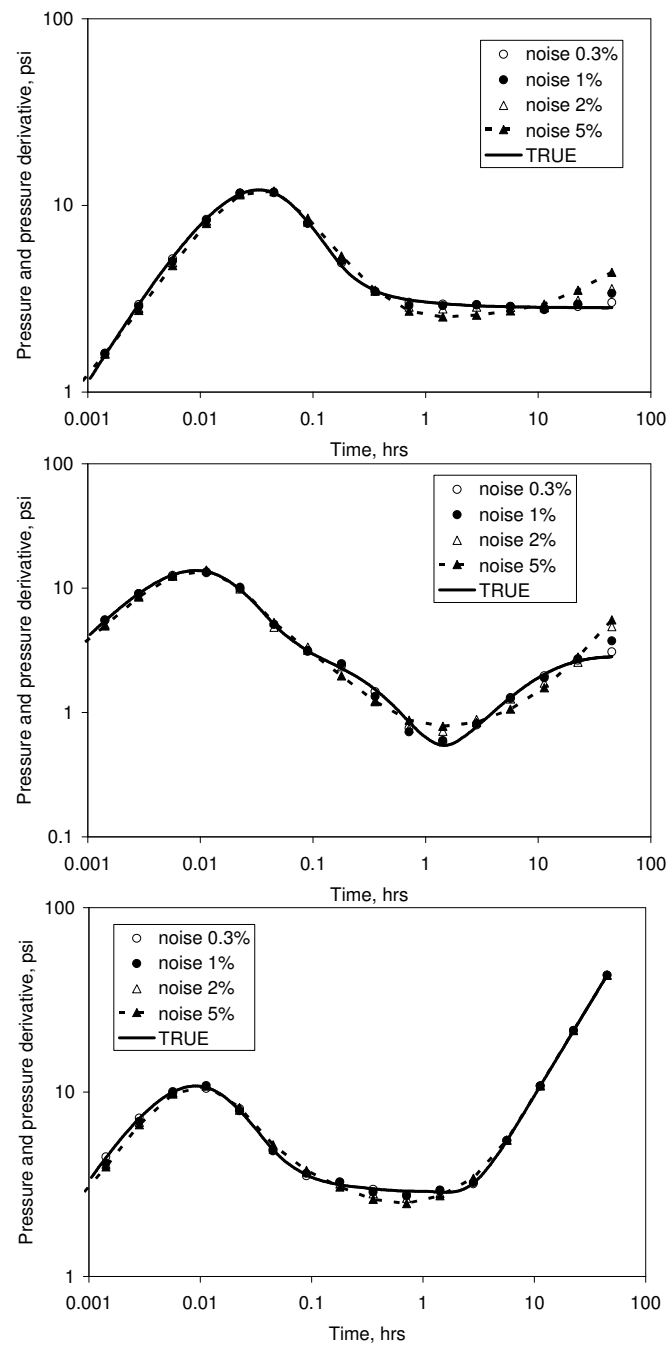


Figure 3.20: Pressure derivative estimates for various noise levels. Solid line: true solution.

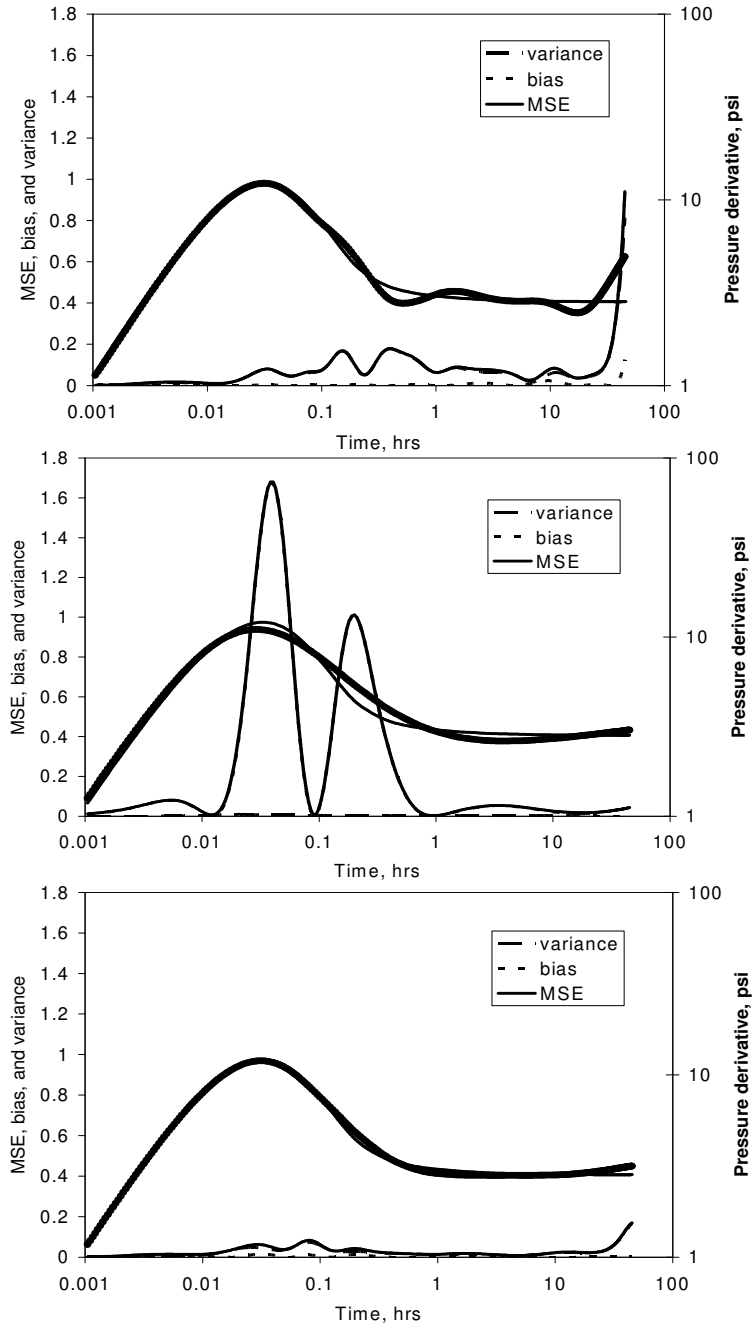


Figure 3.21: Local MSE, bias, and variance (psi^2) for various smoothing control parameter (infinite-acting radial flow model). Thin line: true solution and bold line: estimates.

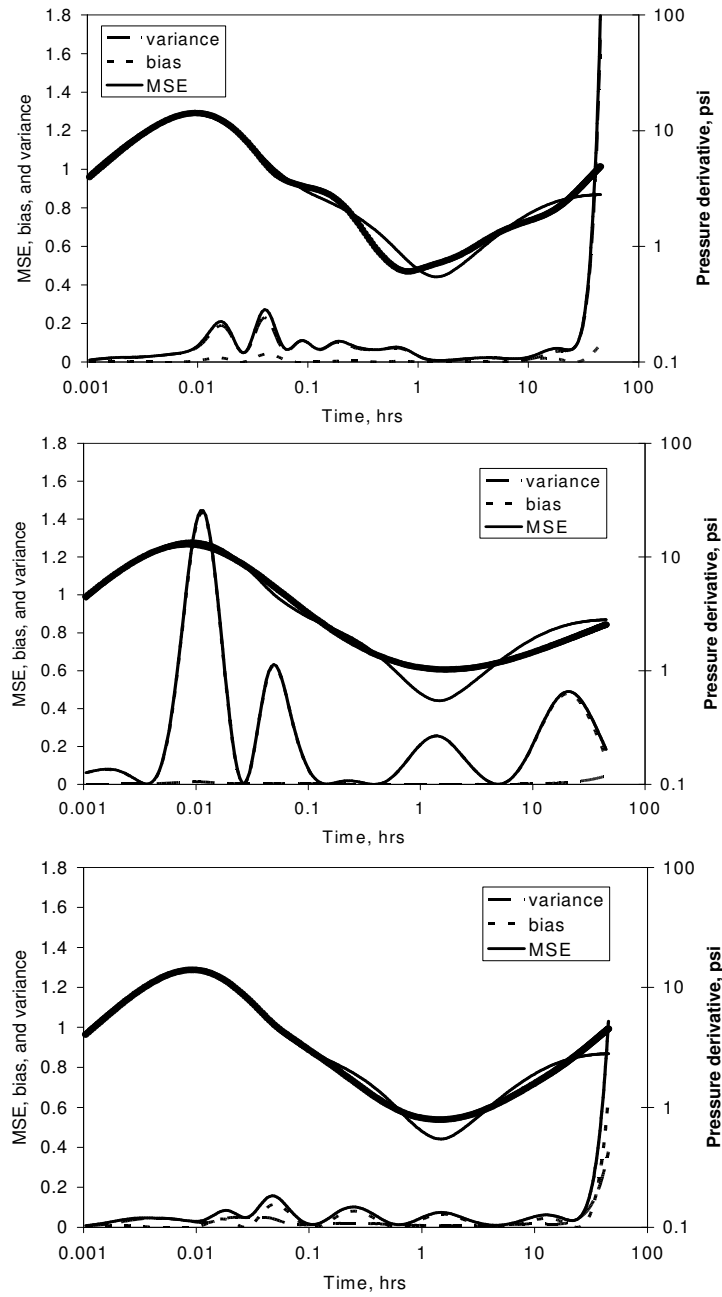


Figure 3.22: Local MSE, bias, and variance (ψ^2) for various smoothing control parameter (dual porosity model). Thin line: true solution and bold line: the estimate.

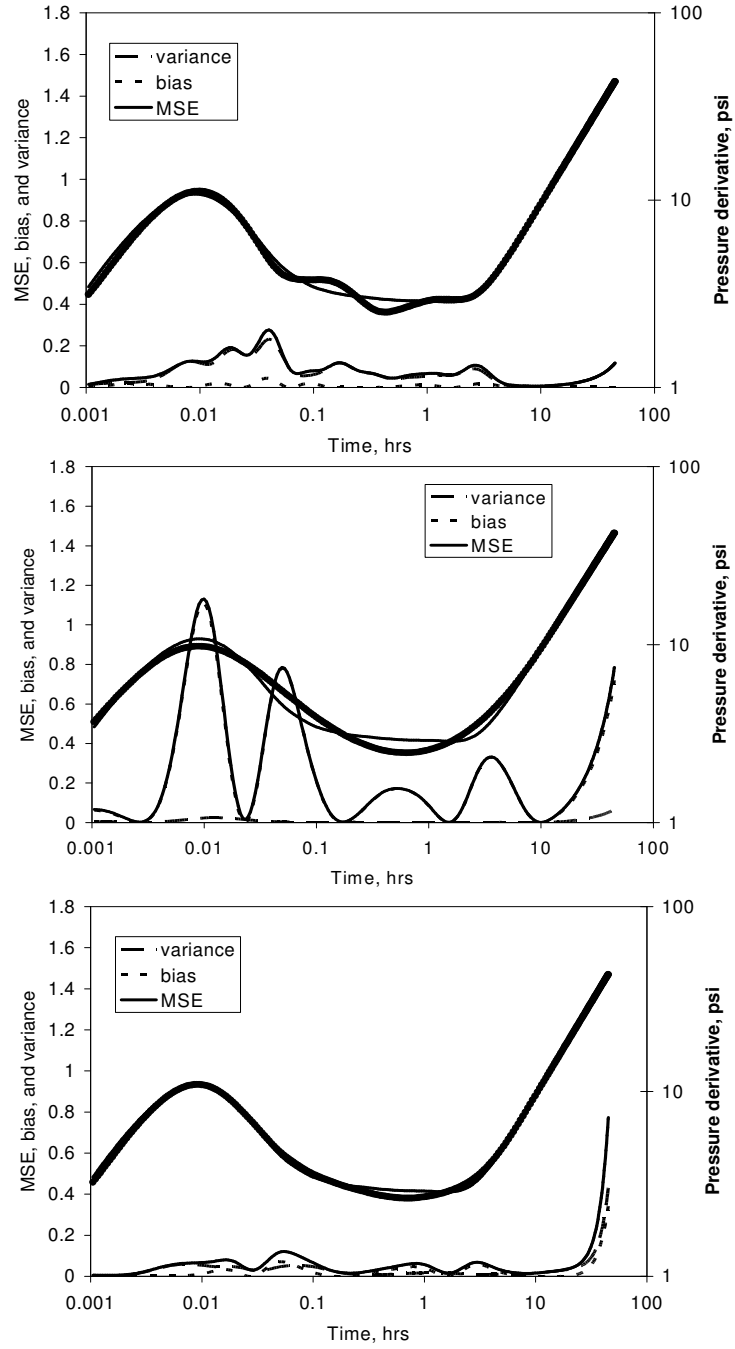


Figure 3.23: Local MSE, bias, and variance (ψ^2) for various smoothing control parameter (closed boundary model). Thin line: true solution and bold line: the estimate.

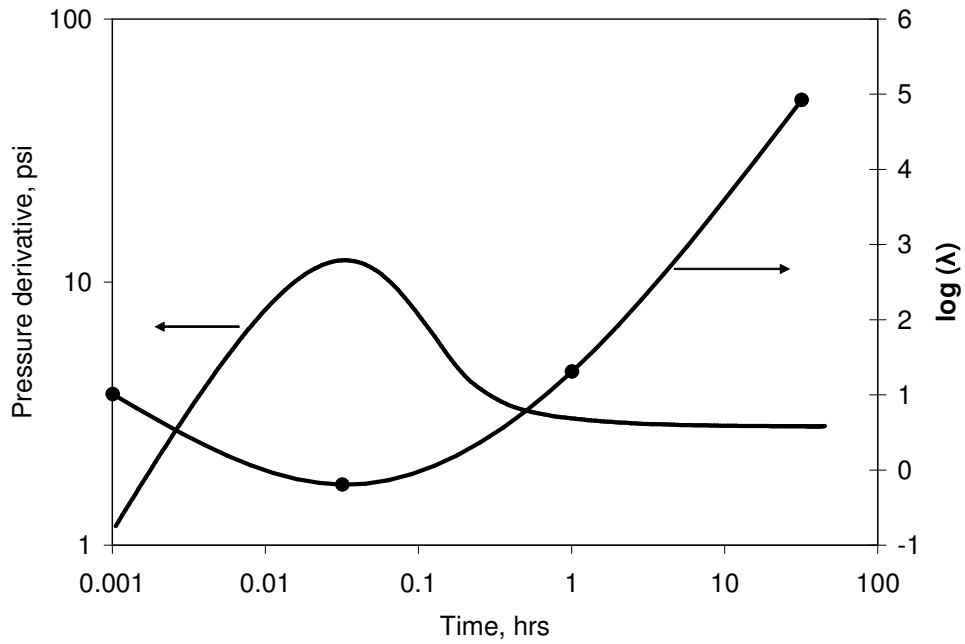


Figure 3.24: The concept of multiple smoothing parameter control.

The procedure to determine the λ values is the same as the single parameter control cases:

- Set initial guess of the spline coefficients for λ profile (C_p).

- Do the cross validation to determine the best values of the λ 's.

(2-1) Fit the model to pressure data under the specified λ profile.

(2-2) Calculate GCV score.

(2-3) Repeat (2-1) and (2-2) by changing λ values (spline coefficients C_p) and select the best values of λ 's which give the minimum GCV score.

- Fit the model to pressure data using the best λ values.

To investigate the effects of multiple smoothing control parameters on the estimates, we tested the method to the same test data used in the last section. In this investigation, we used several knot insertion schemes.

Figure 3.25 shows the estimated λ profile using 50 realizations for the infinite-acting radial flow model case. In this example, the knots for smoothing control were inserted at every 10 control points.

In the figure, the λ profile estimated from the MSE minimization is also plotted. As can be seen, λ from MSE minimization gives a reasonable profile, which means higher values in early and later region and lower values in the storage region. On the other hand, the λ profile estimated from GCV minimization has large scattering especially in the early and later region. This indicates a lack in the information needed to estimate the λ profile. Figure 3.26 shows the local MSE, bias and variance based on this λ profile. Due to the variance of λ profile itself, the resulting MSE (or variance) increased especially at the end point compared with the single parameter case. This tendency was enhanced if we increase the degrees of freedom in the λ profile representation (Figure 3.27). The parameter selection issue could not be overcome during this study.

3.4 Summary

A single-transient pressure data smoothing technique was investigated based on its characteristics. Through the development process and several numerical experiments, the following conclusions can be drawn.

1. The derivative constraints had partial success in removing oscillations. The smoothness is achieved by reducing the degrees of freedom of the model. These constraints decrease the MSE while suppressing the bias. Oscillation still becomes conspicuous for the higher noise level cases.
2. A roughness penalty was imposed to remove oscillation. A smoothing parameter control was carried out by the extended use of generalized cross validation (GCV). GCV criteria gave a reasonable value for the smoothing parameter in all cases.
3. By this regularization, the smoothing algorithm gave smoother results and effectively reduces MSE. The oscillation was removed, making the interpretation easier.

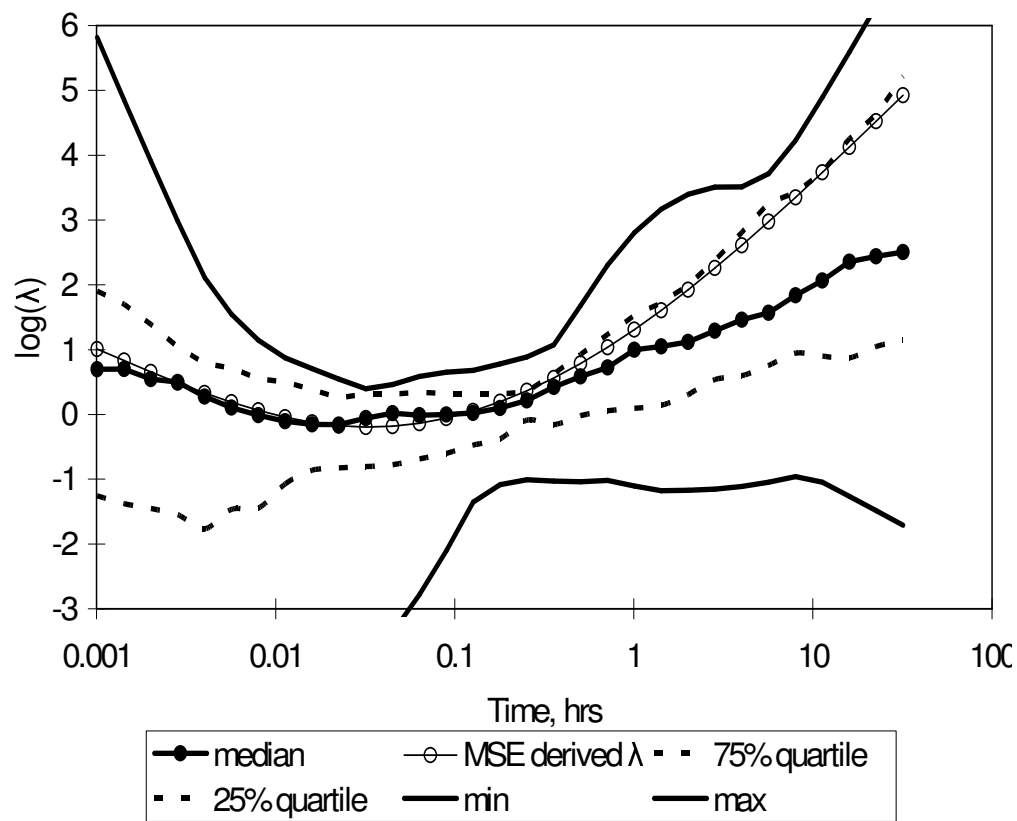


Figure 3.25: The estimated λ profile (knots are placed at every 10 control points).

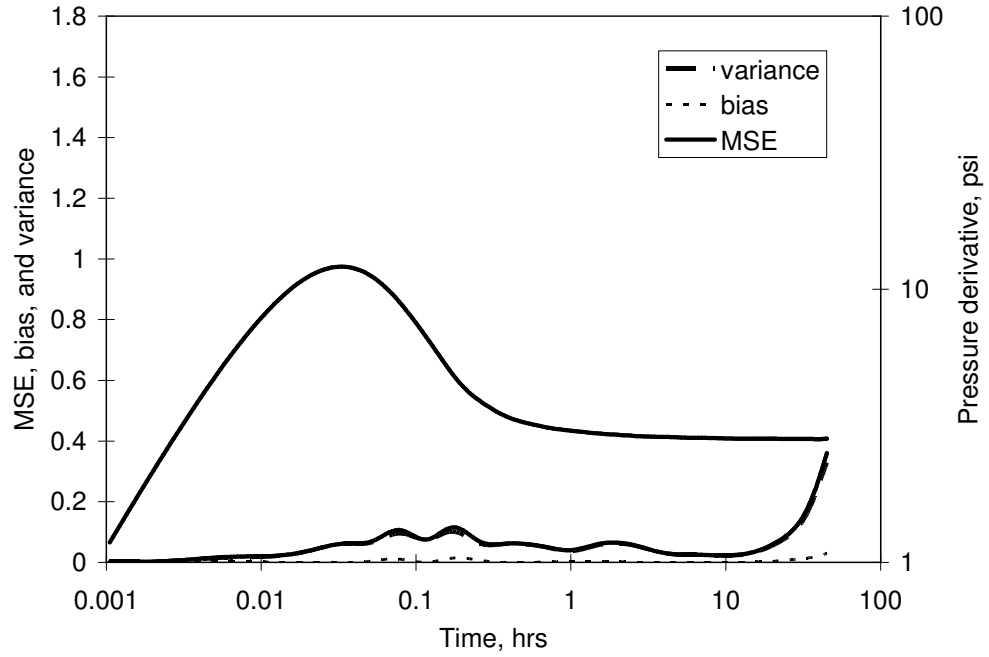


Figure 3.26: Local MSE, bias, and variance (psi^2).

4. The local balance between bias and variance is different region by region. For the higher noise case, the single parameter control results in the loss of local feature such as cutting mountains, filling valleys, and aggravated end performance, which is an important region in the model recognition process.

5. To account for local balance between bias and variance, multiple smoothing controls were considered. The estimated smoothing parameters vary quite widely and resulted in aggravated estimates. This indicates a lack of information to estimate the smoothing parameters from a finite number of data. The reasonable selection of multiple smoothing parameters is a further research issue.

6. The single parameter smoother can be utilized for the multitransient problem described in the later chapters.

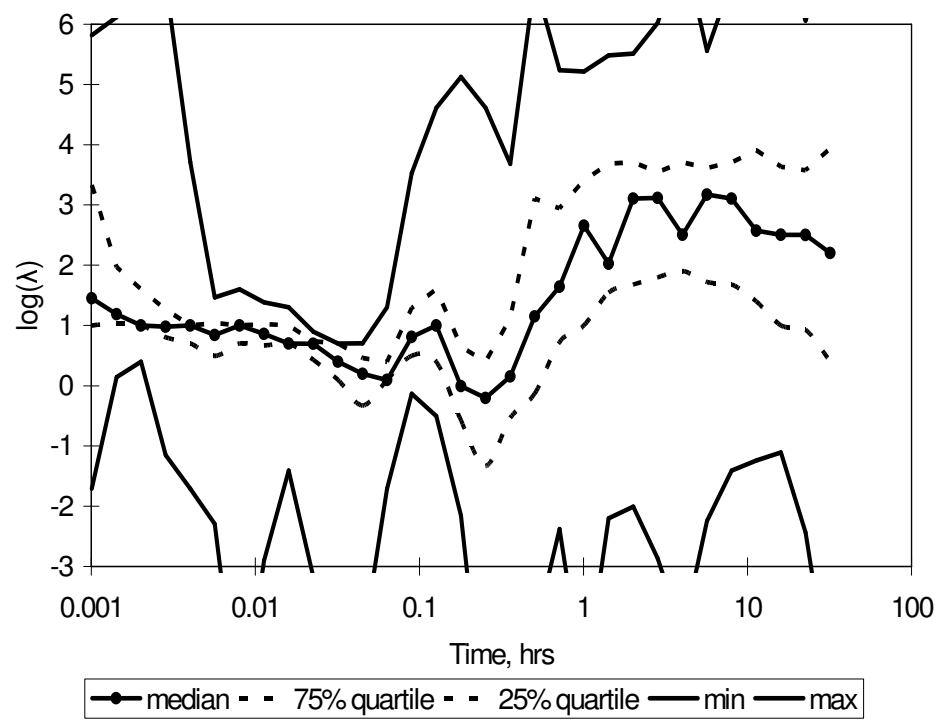


Figure 3.27: The estimated λ profile (knots are placed at every 2 control points).

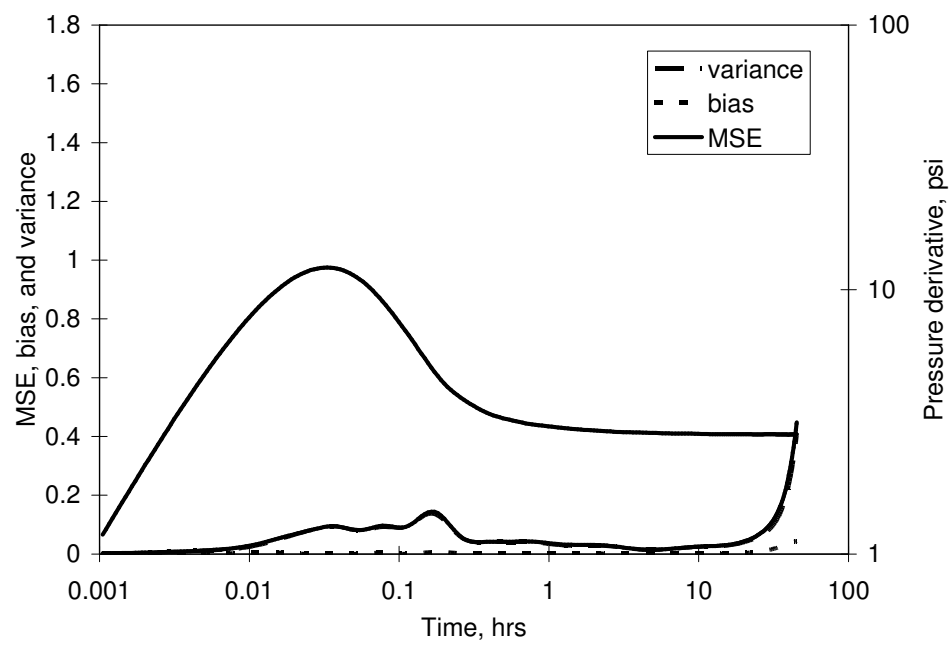


Figure 3.28: Local MSE, bias, and variance (ψ^2).

Chapter 4

Multitransient data analysis

In Chapter 3, a smoother algorithm for the single pressure transient data was described. This smoother is a nonparametric representation of the reservoir model and it has been found that it yields reasonable estimates of pressure derivative when the appropriate smoothing control is made.

A convolution equation describes the pressure drop at time t as follows:

$$\Delta P(t) = \int_0^t Q'(u)K(t-u)du \quad (4.1)$$

Here $K(t)$ and $Q'(t)$ are the response function and the derivative of flow rate at time t .

In a discrete form, the pressure drop at time t_i is:

$$\Delta P(t_i) = \sum_1^{n_t} a_j K(t_i - tb_j) \quad (4.2)$$

Here $a_j (= Q_j - Q_{j-1})$ and tb_j are the effective flow rate and the break point time for j th transient.

If we know the reservoir model *a priori*, we can enter the analytical expression for $K(t)$. However, in many situations, the reservoir model is unknown prior to the pressure data analysis.

The nonparametric representation of the reservoir model is a linear combination of the basis function:

$$K(t) = \sum_{p=1}^M b_p S_p(t) \quad (4.3)$$

Then entering the equation into a convolution equation gives:

$$\Delta P(t_i) = \sum_{j=1}^{n_t} \sum_{p=1}^M a_j b_p S_p(t_i - tb_j) \quad (4.4)$$

This equation is a bilinear formulation of the multitransient pressure behavior and can be viewed as a nonlinear smoother. This chapter describes the multitransient pressure fitting problem from a data smoothing aspect. The smoothing algorithm described here enables us to address the three technical issues. (1) model identification ($\vec{b}$), (2) flow rate recovery ($\vec{a}$), and (3) transient identification (n_t and tb_j). Then, all these parameters become the control parameters for the nonparametric model.

4.1 Model identification and flow rate recovery

Figure 4.1 shows a synthetic multitransient pressure data for the case with 1 % noise. Again, in this study the pressure error is defined as $\frac{\|\vec{\varepsilon}_p\|}{\|\vec{\Delta P}\|}$. Here, $\vec{\varepsilon}_p$ and $\vec{\Delta P}$ are the pressure error and drawdown vectors. In this example, the standard deviation of pressure error is 0.45 *psi* for the maximum pressure drop of 75 *psi*.

For the pressure analysis we need to define the transient (or break point) by partitioning the signal into appropriate segments as carried out in the common practice. In this section, we assume that all the break points (tb_j) are known exactly *a priori*.

The pressure value at time t_i is given by:

$$\begin{aligned} P(t_i) &= \sum_{j=1}^{n_t} \sum_{p=1}^M a_j b_p S_p(t_i - tb_j) + c \\ &= \sum_{j=1}^{n_t} a_j K(t_i - tb_j) + c \end{aligned} \quad (4.5)$$

Here c is the initial pressure. We fit the model to the data to extract the reservoir model and the flow rate profile. This fitting problem is similar to the regression spline problem with a fixed set of knots, although in our case the basis function (reservoir model) is also

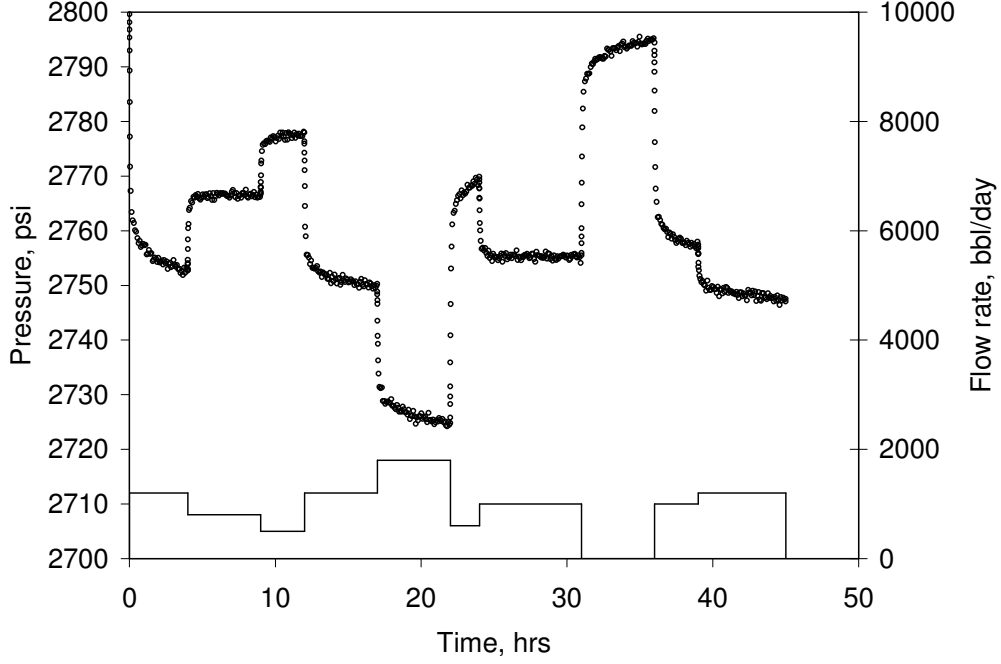


Figure 4.1: An example synthetic data with 1 % noise.

unknown. The formulation becomes:

$$\begin{aligned}
 Obj(\vec{\alpha}) &= \frac{1}{2} \|\vec{d} - \vec{P}\|^2 + \lambda \sum_{r=1}^{N_c} R_r^2(\vec{b}) + \mu \|\underline{F}(\underline{D}\vec{\alpha} - \vec{Q})\|^2 \\
 \underline{A}\vec{b} &\geq 0
 \end{aligned} \tag{4.6}$$

$$\begin{aligned}
 R_r(\vec{b}) &= \log(t_{r-1} \sum_{p=1}^M b_p S'_p(t_{r-1})) - 2\log(t_r \sum_{p=1}^M b_p S'_p(t_r)) \\
 &\quad + \log(t_{r+1} \sum_{p=1}^M b_p S'_p(t_{r+1}))
 \end{aligned} \tag{4.7}$$

Here $\vec{\alpha}$ is a parameter vector ($\vec{\alpha}^T = (\vec{b}^T, \vec{a}^T, c)$). The first term is the error variance. $\vec{P}$ is a model vector represented by Equation 4.5. The second term is the roughness penalty as defined in the single transient case. The third term is another regularization term to

account for the measurement uncertainties in flow rate. μ is the control parameter, which determines how much weight is imposed on the rate measurement. $\underline{D}$ is a lower triangular matrix whose components are all 1. $\vec{Q}$ is the measured flow rate profile (not effective flow rate). $\underline{F}$ is the diagonal matrix where the j th entry is 0 if the j th flow rate data is missing and the remaining entries are 1. If all the flow rate data are available, $\underline{F}$ becomes an identity matrix. $\underline{A}$ is a constraint matrix (identity matrix). This formulation becomes a bilinear least square problem with nonnegativity constraints.

The linear approximation of the nonlinear function $P(t)$ and the quadratic approximation of the roughness penalty term yields:

$$\begin{aligned}
 Obj(\vec{\alpha}) &\approx \frac{1}{2} \|\vec{d} - \underline{J}\vec{\alpha} - \vec{f}_0\|^2 + \lambda \left(\frac{1}{2} \vec{\alpha}^T \underline{X} \vec{\alpha} + \vec{x}^T \vec{\alpha} \right) \\
 &+ \mu (\vec{\alpha}^T \underline{D}^T \underline{F} \underline{D} \vec{\alpha} - 2 \vec{Q}^T \underline{F} \underline{D} \vec{\alpha} + \vec{Q}^T \underline{F} \vec{Q}) \\
 &= \frac{1}{2} \vec{\alpha}^T (\underline{J}^T \underline{J} + \lambda \underline{X} + 2\mu \underline{D}^T \underline{F} \underline{D}) \vec{\alpha} - ((\vec{d} - \vec{f}_0)^T \underline{J} - \lambda \vec{x}^T + 2\vec{Q}^T \underline{F} \underline{D}) \vec{\alpha} \\
 &+ (\vec{d} - \vec{f}_0)^T (\vec{d} - \vec{f}_0) + \mu \vec{Q}^T \underline{F} \vec{Q} \\
 \underline{A} \vec{\alpha} &\geq 0
 \end{aligned} \tag{4.8}$$

The number of parameter (n_p) becomes $M + 1 + n_t$. $\underline{J}$ is a $(n \times n_p)$ Jacobian matrix and f_0 is the model value at the current parameter values. The representation of all the matrixes and vectors are adjusted following the $\vec{\alpha}$ representation. We used $\underline{F}^T = \underline{F} = \underline{F}^T \underline{F}$.

As in the single transient case, this problem can be solved by the Levenberg-Marquardt method combined with the active set method. Suppose we have k active constraints and the damping parameter w at the final iteration, the solution and the estimated function become:

$$\vec{\alpha} = \{ \underline{Z}_k^T \underline{G} \underline{Z}_k \}^{-1} \underline{Z}_k^T \{ \underline{J}^T (\vec{d} - \vec{f}_0) - \lambda \vec{x} + 2 \underline{D}^T \underline{F} \vec{Q} \} \tag{4.9}$$

$$\vec{f} = \underline{J} \underline{Z}_k \{ \underline{Z}_k^T \underline{G} \underline{Z}_k \}^{-1} \underline{Z}_k^T \{ \underline{J}^T (\vec{d} - \vec{f}_0) - \lambda \vec{x} + 2 \underline{D}^T \underline{F} \vec{Q} \} + \vec{f}_0 \tag{4.10}$$

where $\underline{G} = \underline{J}^T \underline{J} + \lambda \underline{X} + 2\mu \underline{D}^T \underline{F} \underline{D} + w \underline{I}$.

As for the initial guess, we generate the analytical solution for the infinite-acting radial flow model without skin and storage and then obtain the estimate for the response parameter $\vec{b}$ by fitting the single transient smoother. Before proceeding with the Levenberg-Marquardt method, we update the solution for the parameters by iterative minimization to obtain better starting values for $\vec{a}$ and c . Namely (1) update $\vec{a}$ and c for the fixed $\vec{b}$ and (2) update $\vec{b}$ for the fixed $\vec{a}$ and c . The process is repeated until the change of the objective function value becomes less than 10 % of the previous objective function value. Although we can solve the problem in this way until convergence, it has been found that this iterative minimization is quite time consuming requiring many more iterations compared with the Levenberg-Marquardt method under the same convergence criteria. Due to the bilinear nature, there are many equivalent solutions for $\vec{a}$ and $\vec{b}$ without the flow rate regularization. For example, doubling $\vec{a}$ and halving $\vec{b}$ yield the same fitting results. However this solution procedure always yielded one solution without any oscillation problem.

In this smoothing problem, the smoothness of each transient is achieved, while keeping the edge structure in the signal under the fixed break points. First we checked how each of the regularizations affects the response estimates.

We fitted the model with the roughness penalty to 50 realizations for the case of the infinite acting radial flow model. In this setting, we excluded the flow rate regularization term.

The smoothing parameter (λ) was similarly determined by the GCV score minimization. The GCV score is calculated by the trace of the linearized smoother matrix given by:

$$\underline{S} = \underline{J}\underline{Z}_k\{\underline{Z}_k^T \underline{G}\underline{Z}_k\}^{-1}\underline{Z}_k^T \underline{J}^T \quad (4.11)$$

Figure 4.2 shows an example of the GCV and the degrees of freedom for the various smoothing parameters. The increase of the smoothing parameter results in the decrease in the degrees of freedom of the bilinear model. The GCV profile has a similar shape to the one shown in the single transient case. The GCV increases sharply in the direction of the lower degrees of freedom and increases gradually in the opposite direction.

Figure 4.3 shows the mean squared error (MSE) of the pressure derivative estimates for the various noise levels. The white circles show the case without the smoothing control. As in the single transient cases, the MSE is effectively improved by imposing the roughness

penalty.

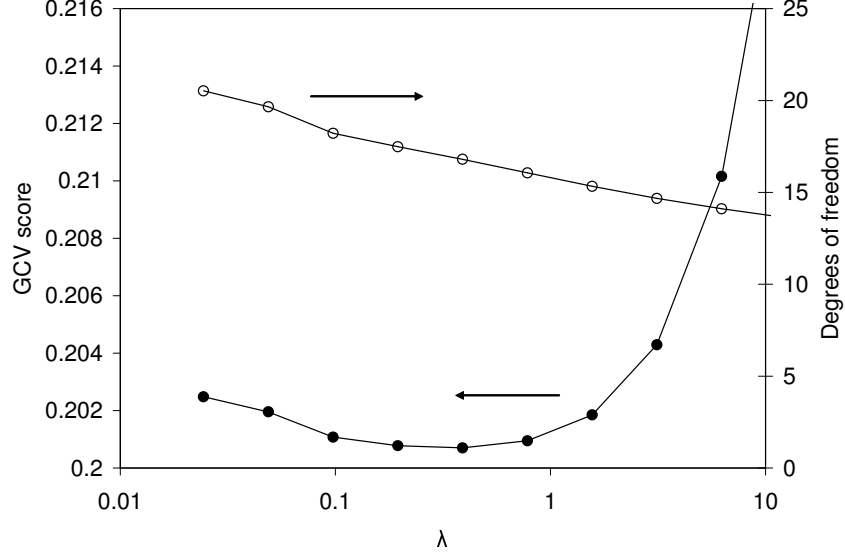


Figure 4.2: GCV score and degrees of freedom for the various smoothing parameters (pressure error 1%).

Similarly, we checked the effect of the flow rate regularization on the response estimates. For the large μ , the results exactly match the measured flow rate data even if the measurement error is large. For the zero μ , the flow rate information is completely neglected. Then the flow rate regularization also introduces bias of the estimate. Here we fitted the model to the same 50 realizations excluding the roughness penalty term. Figure 4.4 shows the GCV profile and the degrees of freedom for the various control parameters. As the μ value increases, the degrees of freedom decreases, since the increasing μ tries to fix the effective flow rate values to the measurement. The GCV profile has the similar concave shape around the minimum. We selected the this minimum point to select the best value of μ . Figure 4.5 shows the MSE of the pressure derivative estimates for the various measurement errors. The open circle shows the case without any control. In this study, the flow rate measurement error ε_q is defined as $\|\vec{Q}_{true} - \vec{Q}_{measured}\|/\|\vec{Q}_{true}\|$. We considered 0, 1, 5, and 10 % error for the flow rate.

For the accurate flow rate, the improvement of the MSE is drastic, the improvement of the MSE is moderate to small for the increasing error as expected.

We combined these effects to estimate the pressure derivative. To select the best values

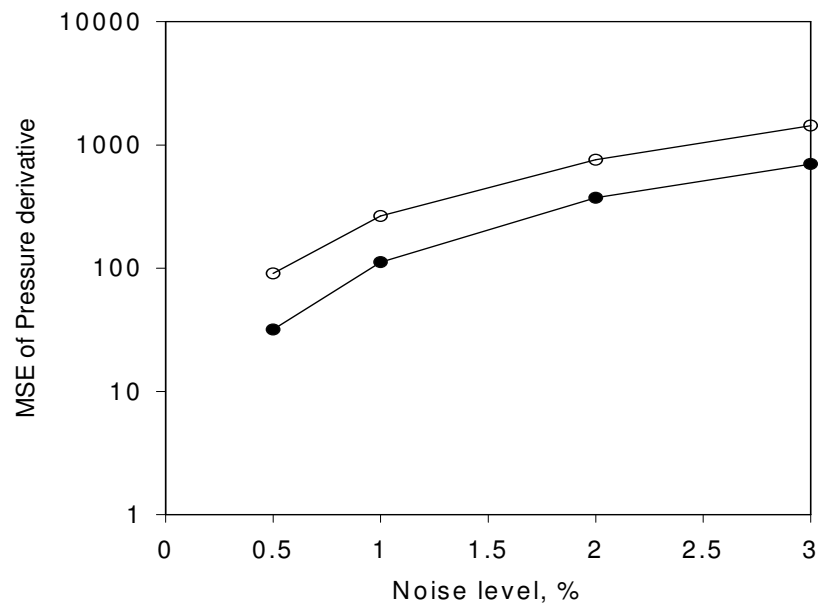


Figure 4.3: MSE of the pressure derivative (psi^2). Open circles: without smoothing control

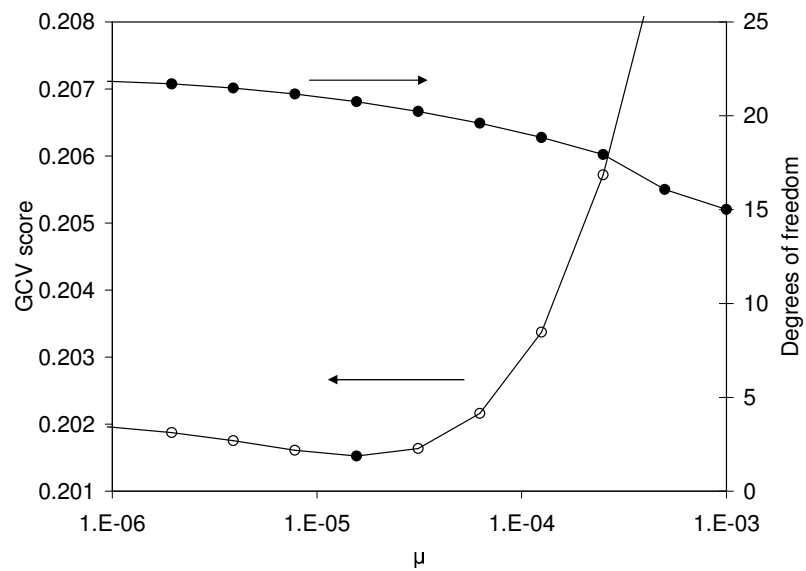


Figure 4.4: GCV and degrees of freedom for various control parameters.

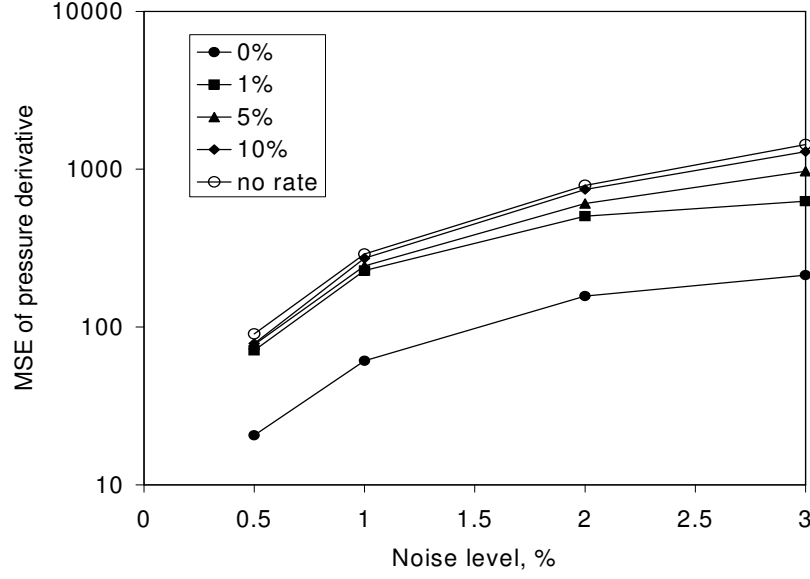


Figure 4.5: MSE of the pressure derivative (psi^2). Open circles: w/o regularization.

of λ and μ , we did the GCV minimization iteratively. Namely (1) minimize the GCV with respect to λ for a fixed μ and (2) minimize the GCV with respect to μ for a fixed λ . We employed the grid search. Usually a couple of iterations were sufficient to obtain the best values starting with $\mu = 0$ at the first step.

Figures 4.6 and 4.7 show the MSE of the pressure derivative estimates and the error of the estimated flow rate profile respectively. The MSE is well improved compared with the previous cases and the flow rate is estimated within 2 % error even for the severe measurement error.

Figures 4.8 to 4.10 show examples of the derivative estimates for the the various noise levels for the three reservoir models (infinite-acting, dual porosity, and closed boundary). Figure 4.11 shows the example of the flow rate estimation result for 10% rate error and 3 % pressure error for the infinite-acting radial flow model. Although the appearance of the derivative varies to some extent depending on the realization, considering the noise level expected in the actual data sets shown later, the effect of measurement error on the estimates is likely to be well suppressed by the regularization for the well-designed settings.

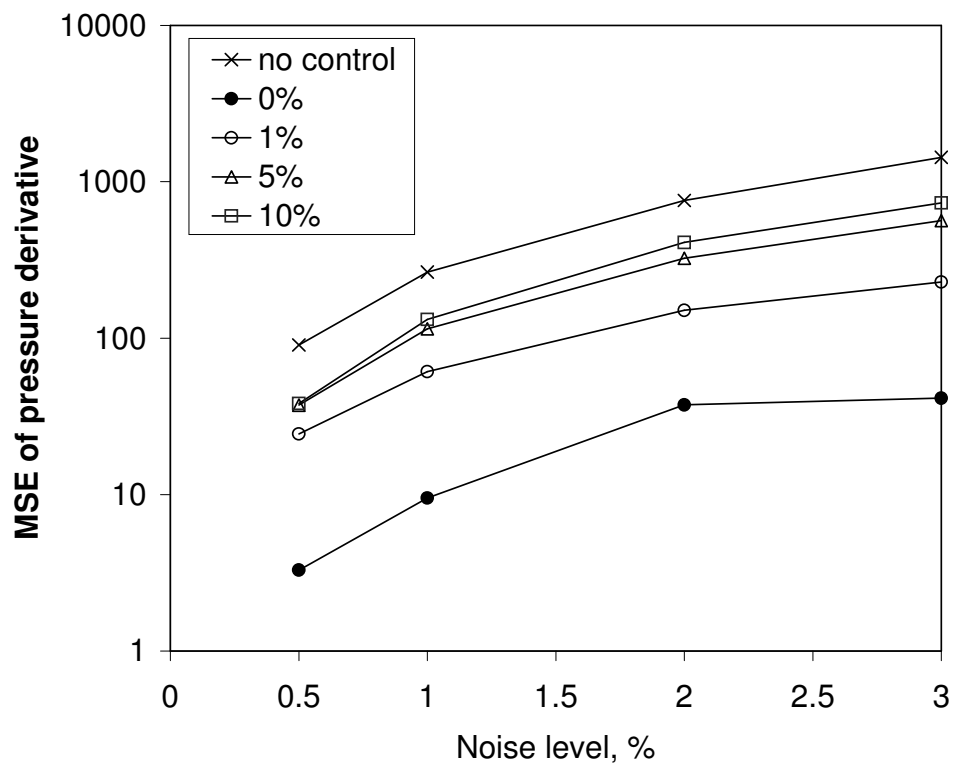


Figure 4.6: MSE of the pressure derivative (psi^2).

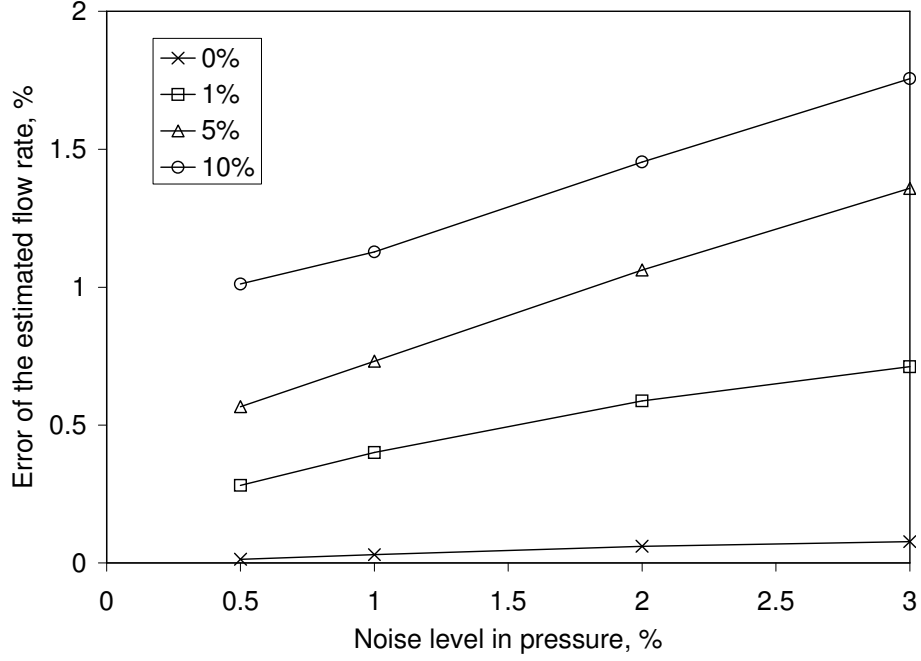


Figure 4.7: The error of the estimated flow rate.

4.2 Comparison with the existing method

In recent years Schroeter et al. [29, 28] have been developing a time-domain deconvolution method which account for pressure and flow rate measurement uncertainties. To our knowledge, the method is likely to be the most robust among the existing methods in terms of resistance to the measurement error [33, 20]. In this section we briefly describe their method and compare it with our method. For the details of their method, see the references [29, 28].

Their formulation is based on the parametrization of the derivative values on the log-log plot. Namely,

$$Z(\tau) = \ln\left(t \frac{dp}{dt}\right) \quad (4.12)$$

Here $\tau = \ln(t)$.

The convolution equation has the two equivalent forms:

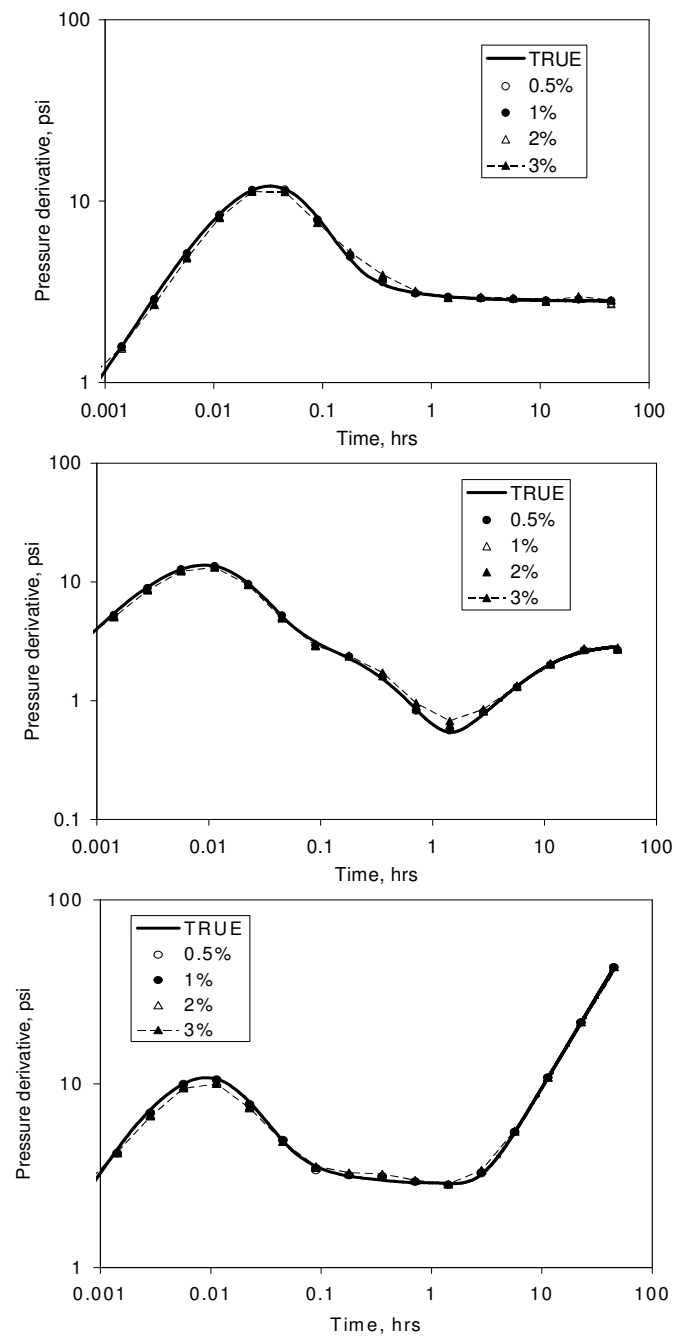


Figure 4.8: The deconvolved response with the exact flow rate profile. Solid line: true solution.

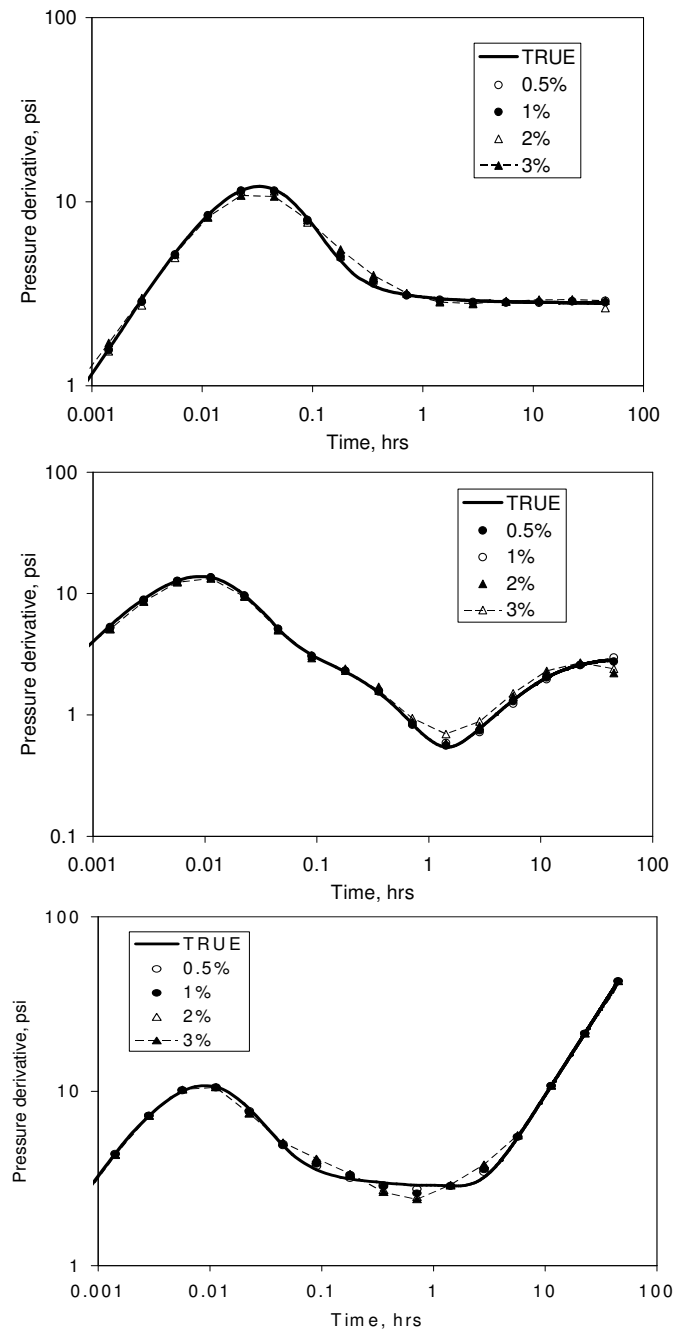


Figure 4.9: The deconvolved response with 1 % flow rate error. Solid line: true solution.

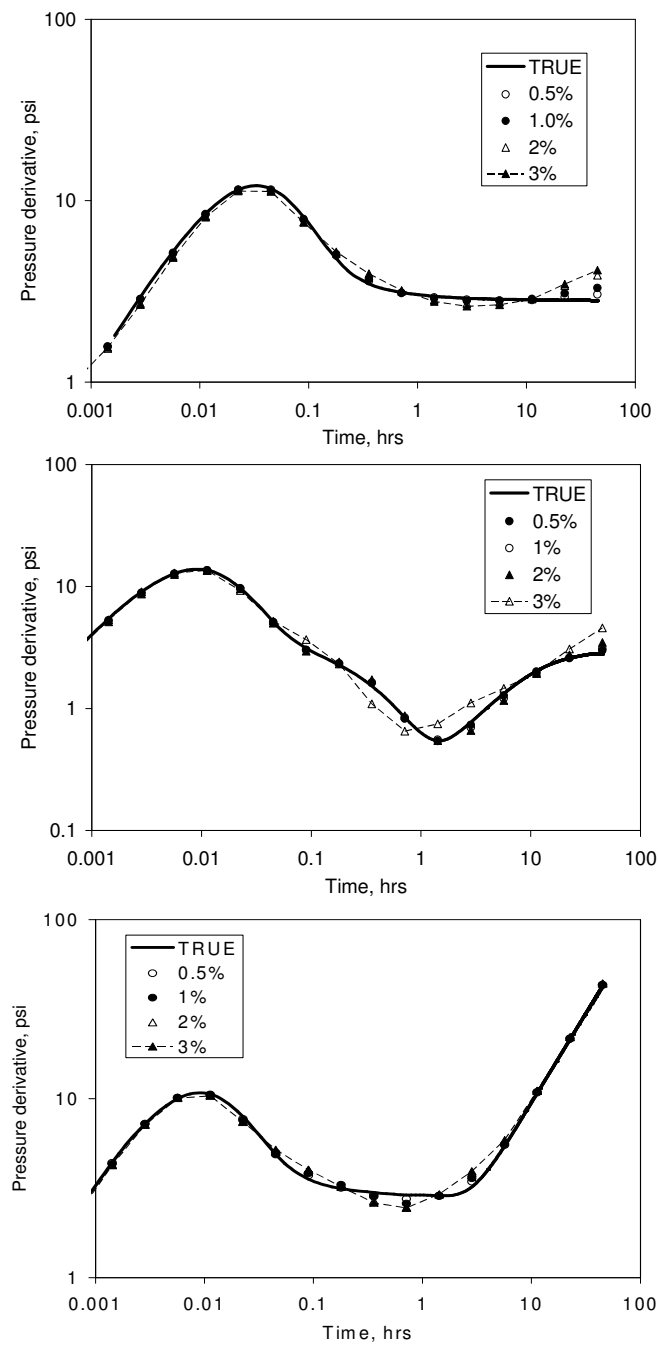


Figure 4.10: The deconvolved response with 10 % flow rate error. Solid line: true solution.

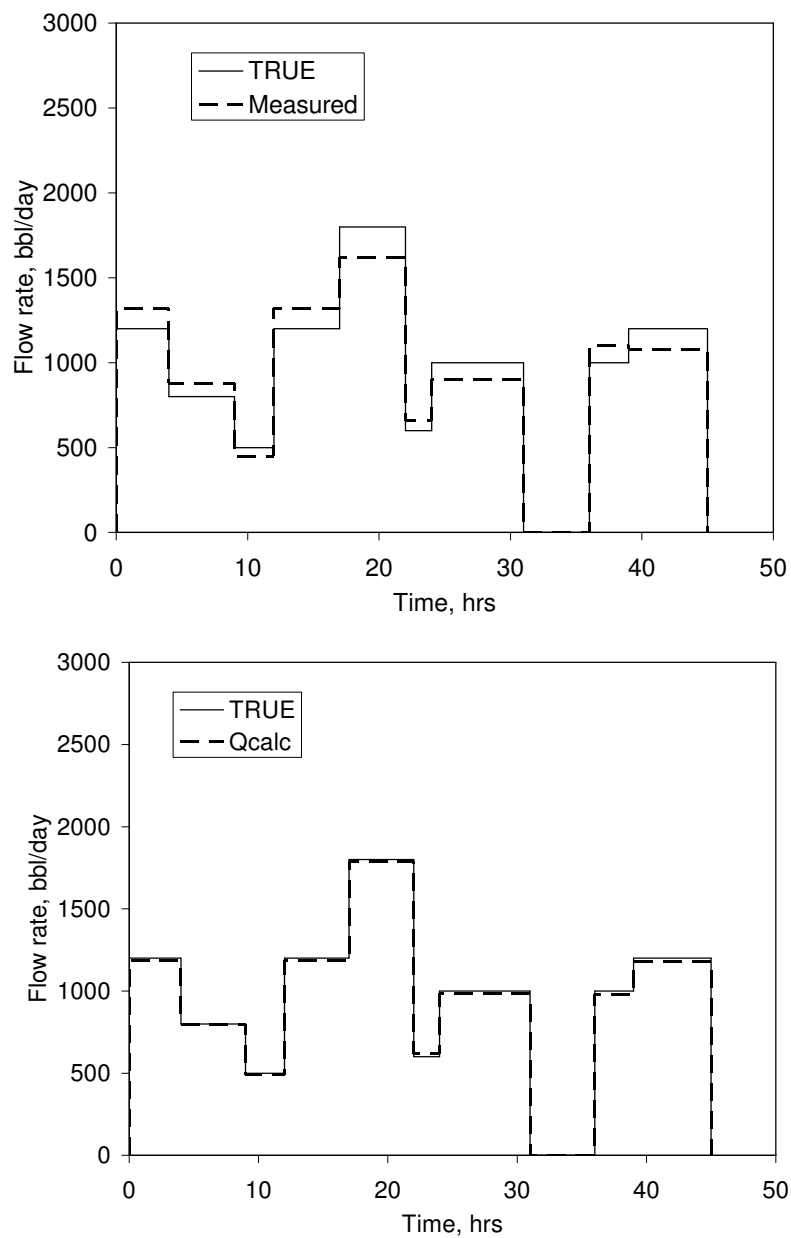


Figure 4.11: The estimated flow rate profile for pressure noise 3 % and rate noise 10 % (infinite-acting radial flow).

$$\begin{aligned}
\Delta P(t) &= \int_0^t Q'(u)K(t-u)du \\
&= \int_0^t Q(t-u)K'(u)du
\end{aligned} \tag{4.13}$$

This equation is satisfied if $K(0) = 0$, which follows from its definition, and $Q(0) = 0$. Using the logarithm of time (τ), the final expression in Equation 4.13 becomes:

$$\begin{aligned}
\Delta P(t) &= \int_0^t Q(t-u)uK'(u)\frac{du}{u} \\
&= \int_{-\infty}^{\ln(t)} Q(t-e^\tau)e^{Z(\tau)}d\tau
\end{aligned} \tag{4.14}$$

Here, $u = e^\tau$ and $Z(\tau) = \ln(uK'(u))$.

In a discrete form,

$$\vec{\Delta p} = \underline{C(\vec{z})}\vec{Q} \tag{4.15}$$

Here $\vec{\Delta p}$ and $\vec{Q}$ are the drawdown vector and flow rate vector respectively. $\underline{C(\vec{z})}$ is a matrix valued function of the response coefficients ($\vec{z}$) with components:

$$C_{ij}(\vec{z}) = \int_{-\infty}^{\ln(T)} \theta_j(t_i - e^\tau)e^{z(\tau)}d\tau \tag{4.16}$$

for $i = 1...m$ and $j = 1...N$. m and N are the number of pressure data points and the number of transients. T is the maximum observation time. $\theta_j(t)$ is defined as:

$$\theta_j(t) = \begin{cases} 1 & : t \in [tb_j, tb_{j+1}] \\ 0 & : otherwise \end{cases} \tag{4.17}$$

tb_j is the j th break point time.

In their formulation, the logarithm of pressure derivative ($Z(\tau)$) is represented using a linear interpolating function with the predefined knot sequence (Figure 4.12). The knot sequence is:

$$-\infty = \tau_0 < \tau_1 < \tau_1 < \dots < \tau_n = \ln(T) \quad (4.18)$$

In this pressure derivative representation, the unit slope behavior is assumed between $\tau_0 (= -\infty)$ and τ_1 [29, 28].

With the flow rate regularization term and the roughness penalty on the pressure derivative, the final objective function becomes:

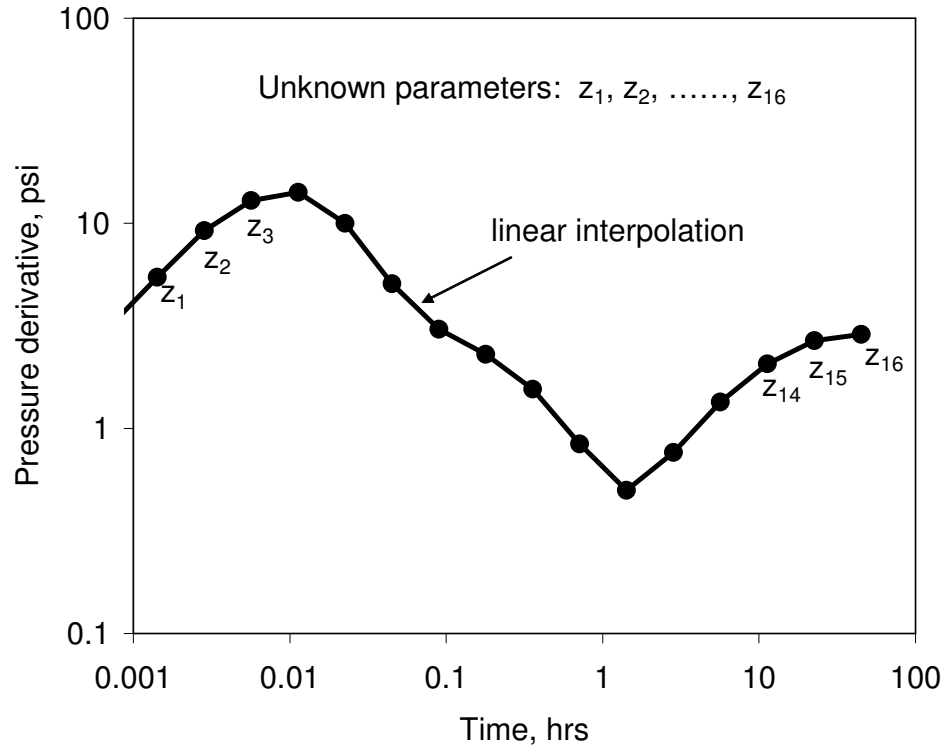


Figure 4.12: Schematic of Schroeter's formulation.

$$Obj = \|\vec{p} - \vec{p}_0 - \underline{C}(\vec{z})\vec{y}\|^2 + \nu\|\vec{y} - \vec{Q}\|^2 + \lambda\|\underline{D}\vec{z} - \vec{k}\| \quad (4.19)$$

The first term is the pressure error variance and the second and the third terms are the flow rate regularization and the roughness penalty terms. $\vec{p}$ and $\vec{p}_0$ are pressure data vector and initial pressure vector (each component has initial pressure value). $\vec{y}$ and $\vec{Q}$ are the flow rate vector (unknown vector) and measured flow rate data vector. $\vec{k}$ is a vector whose

Table 4.1: MSE of pressure derivative (psi^2) and flow rate estimation error.

MSE of pressure derivative			
	Infinite-acting	dual porosity	closed boundary
this study	1.86	0.35	0.31
Schroeter	2.08	4.46	5.63
Flow rate estimation error			
this study	0.59	0.61	0.58
Schroeter	0.83	0.83	1.11

first component is 1 and remaining component 0. $\underline{D}$ is the second derivative matrix.

Regularization parameter ν and smoothing parameter λ are set as:

$$\begin{aligned}
 \nu_{def} &= \frac{N \|\vec{\Delta p}\|^2}{m \|\vec{Q}\|^2} \\
 \lambda_{def} &= \frac{\|\vec{\Delta p}\|^2}{m}
 \end{aligned} \tag{4.20}$$

In their method, the monotone constraint (positive first derivative) is implicitly imposed on the formulation due to the representation on the log-log plot.

The main technical difference is that they used a more subjective default value for these two control parameters in the nonparametric model, while we select them jointly with the cross validation technique.

We compared this algorithm with ours with several reservoir models for pressure error 1 % and flow rate error 5 % case. For the comparison, we used the same knot insertion scheme in both methods.

Figure 4.13 shows an example derivative estimates for three reservoir models: infinite-acting radial flow, dual porosity model, and closed boundary model. Table 4.1 shows the mean squared error of pressure derivative at the exact knot location and flow rate estimation error. In the three reservoir models, our method consistently gives better results, while Schroeter's method tends to produce oversmoothed results. The flow rate estimation was better estimated in our method. Our better results are likely due to the control parameter selection.

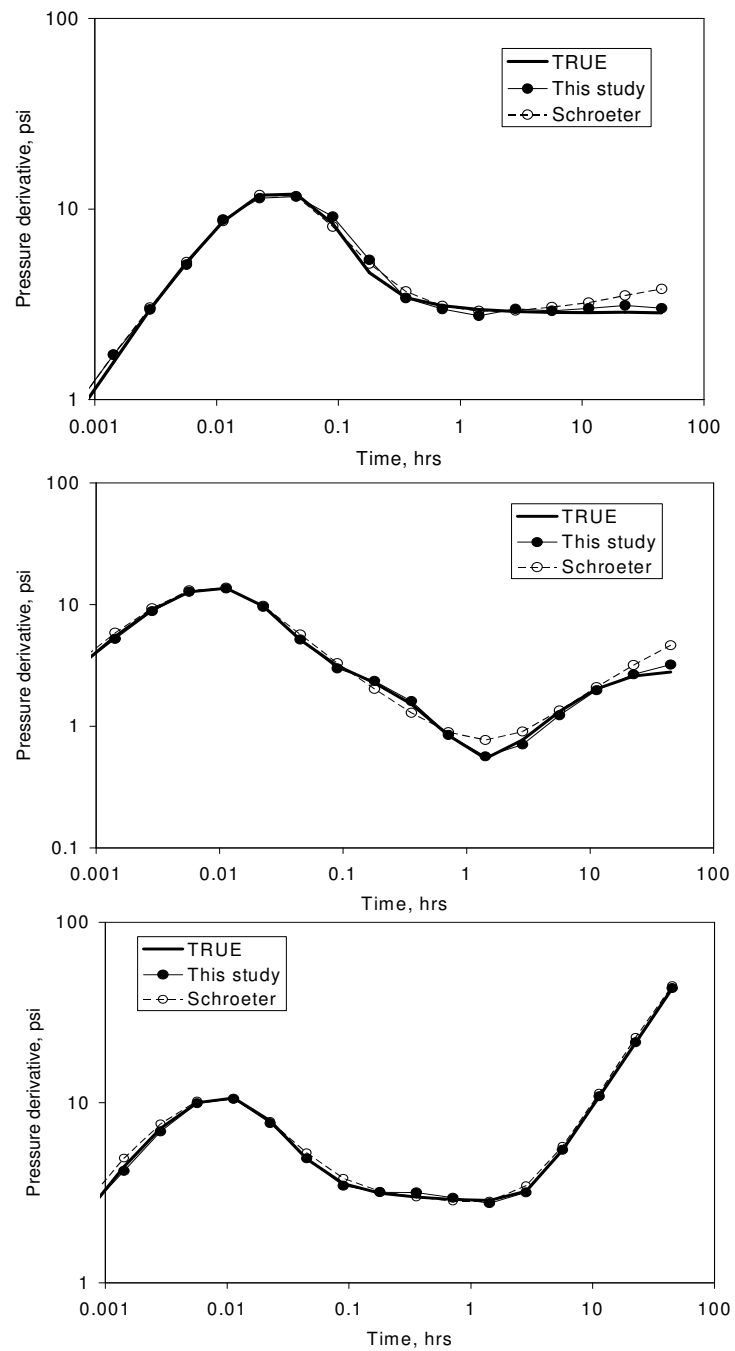


Figure 4.13: The derivative plot for 1 % pressure error and 5 % flow rate error.

4.3 Transient identification

So far we have been looking at the pressure fitting problem under well-defined conditions, namely under exact knowledge of the number and location of the break points. Figure 4.14 shows a synthetic multitransient pressure data set. In this example, the pressure data have many small transients as would be expected in the actual field data especially around the start of large transients. In order to analyze the data, it is necessary to know the flow rate profile behind the pressure signal, namely how the pressure signal is produced. However, continuous flow rate measurement is still uncommon in the industry, and thus usually the flow rate profile corresponding to every single pressure data point is not available. This is also true even in the conventional well testing schemes. Depending on the sampling interval, the break points may not be located at the exact data points. Under these circumstances, we need to define the transient to make the analysis possible by partitioning the data into individual segments. This problem may be referred to as *transient identification*. This section describes the transient identification problem.

4.3.1 Wavelet processing

The transient identification has been treated as an edge detection problem in signal processing by several authors [18, 26], since the pressure profile corresponding to the flow rate change tends to exhibit edge features (break points) in the signal. A common approach has been the local filtering (smoothing) method. The approach makes it easier to visualize and analyze the pressure data even for noisy cases. The authors attempted to detect the break points by defining the criteria from the visual characteristics of the processed signal.

Athichanagorn [1, 2] applied wavelet analysis for this particular issue. Wavelet analysis is very popular in signal processing and compression, since it is able to represent both smooth and/or locally bumpy functions in an efficient way. Wavelet analysis is commonly used to extract the information contained in different frequencies of the signal. The most attractive feature of the wavelet analysis is that it is able to separate high frequency contents which represent small-scaled information from low frequency components which represent large-scaled information. Here we briefly describe the method proposed by Athichanagorn as a representative example in this category.

Figures 4.15 and 4.16 show the wavelet transformation of the signal. Here the spline wavelet was employed [1, 2]. The wavelet transformation yields a view of the data from

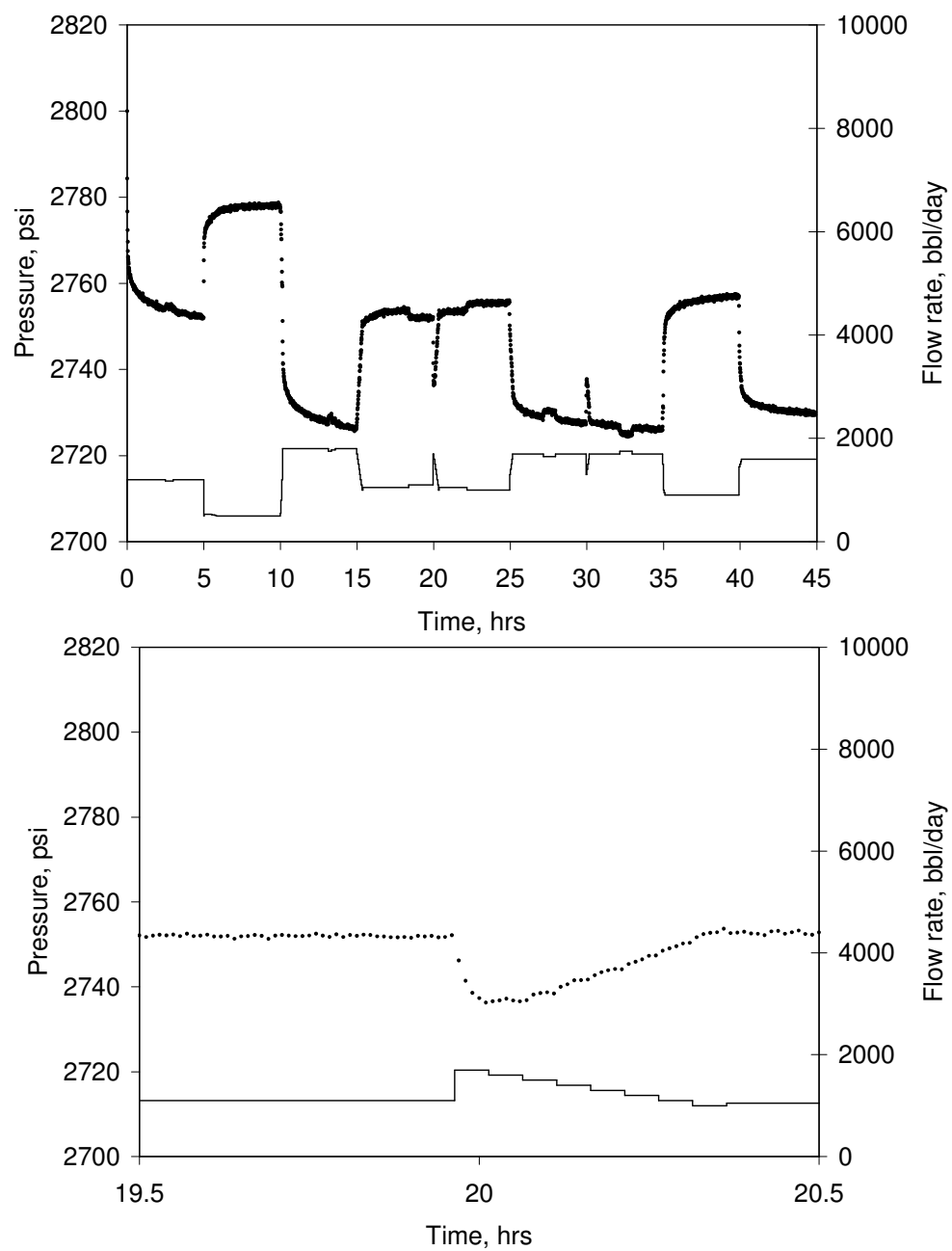


Figure 4.14: A synthetic pressure data and an expanded view (0.3 % noise).

multiple levels of resolution. Athichanagorn [1, 2] proposed the break point detection algorithm based on this representation. Specifically, two thresholding methods were adopted: (1) pressure thresholding and (2) slope thresholding. The method selects the peak which exceeds the user-specified threshold (pressure threshold) in the detailed signal at the selected level and maps the location back to the original resolution level. Then the redundant break points are removed by checking the slope difference (the magnitude of the edge) calculated from linear fitting with several data points after and before the detected break points.

Figures 4.17 and 4.18 show the processing results using three different pressure threshold values. The upper figure in Figure 4.17 shows the true break point locations. Here we used the fifth level of detailed signal and the pressure thresholding only. 2 *psi*, 0.5 *psi*, and 0.01 *psi* were selected as the pressure threshold values for the demonstration purpose. Figure 4.19 shows some of the expanded views of the processing results for the case with 0.5 *psi* threshold value.

Although the wavelet captures the large features of the signal, it misses visually identifiable break points depending on the threshold value. In this example, the small transients were missed while many false break points were detected even if we decreased the threshold value. Any filter-based approach undertaken so far has yielded similar results [18, 26]. Due to the inexact nature of the visual characteristics of the transient, it was a difficult problem to define universal criteria based on the processed information. Processing the pressure data in a discrete form may cause another issue, since the break point is not always located at the exact data locations especially for large sampling intervals.

At a glance, the pressure threshold of 0.01 *psi* seems to yield a good processing result compared to the others. We fitted the model with the roughness penalty using these break point locations to derive the response function. Here the smoothing parameter was determined manually such that the original feature of the response is well maintained while moderate smoothness is achieved. Figure 4.20 shows the deconvolved response using the wavelet-detected break points. In the figure the triangles show the result obtained by adjusting the estimated break points if they are placed at clearly wrong locations. Although the added pressure noise is 0.3 %, the derived response was totally different from the true one. The adjustment of the break points produces a substantial difference in the late time region.

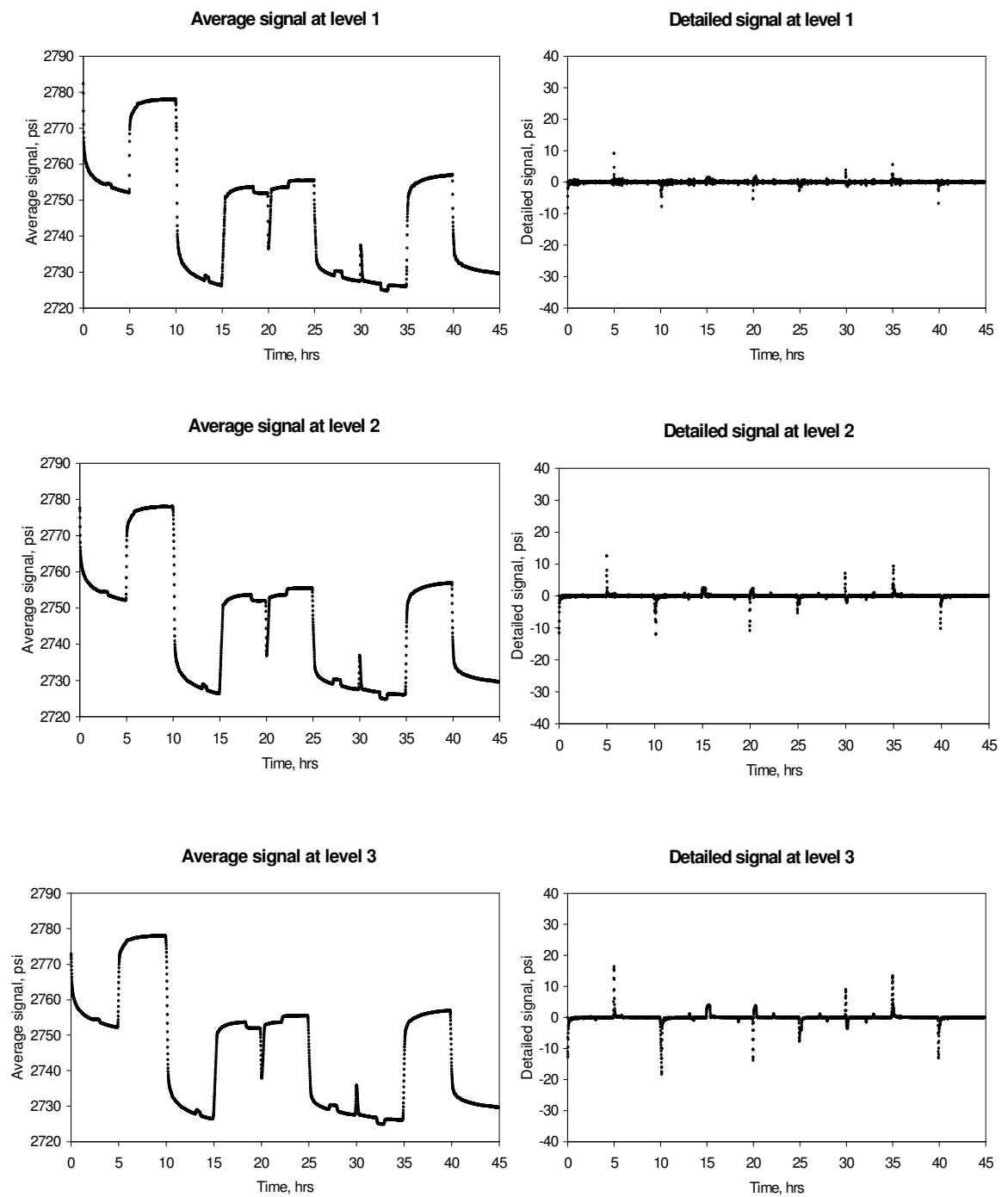


Figure 4.15: Wavelet transformation of the original signal (1). Left: the approximated signal. Right: the detailed signal.

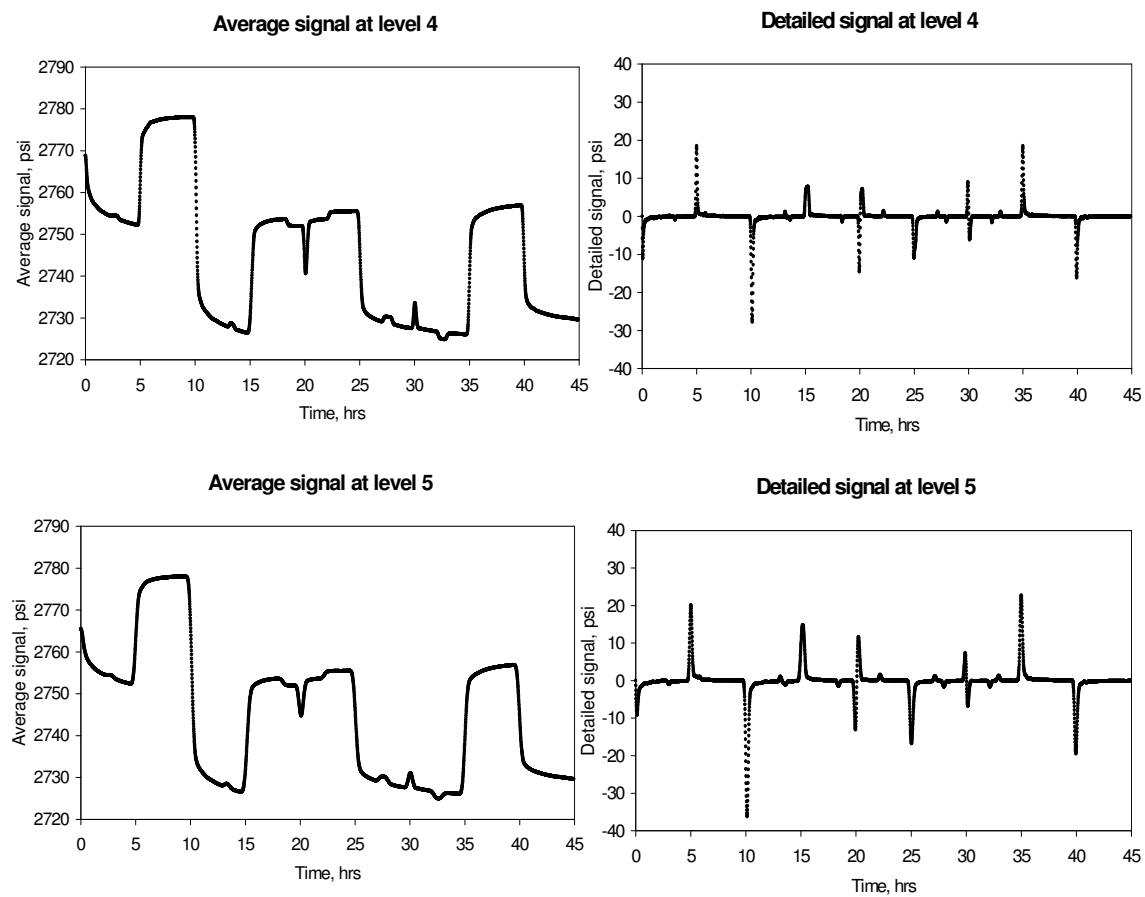


Figure 4.16: Wavelet transformation of the original signal (2). Left: the approximated signal. Right: the detailed signal.

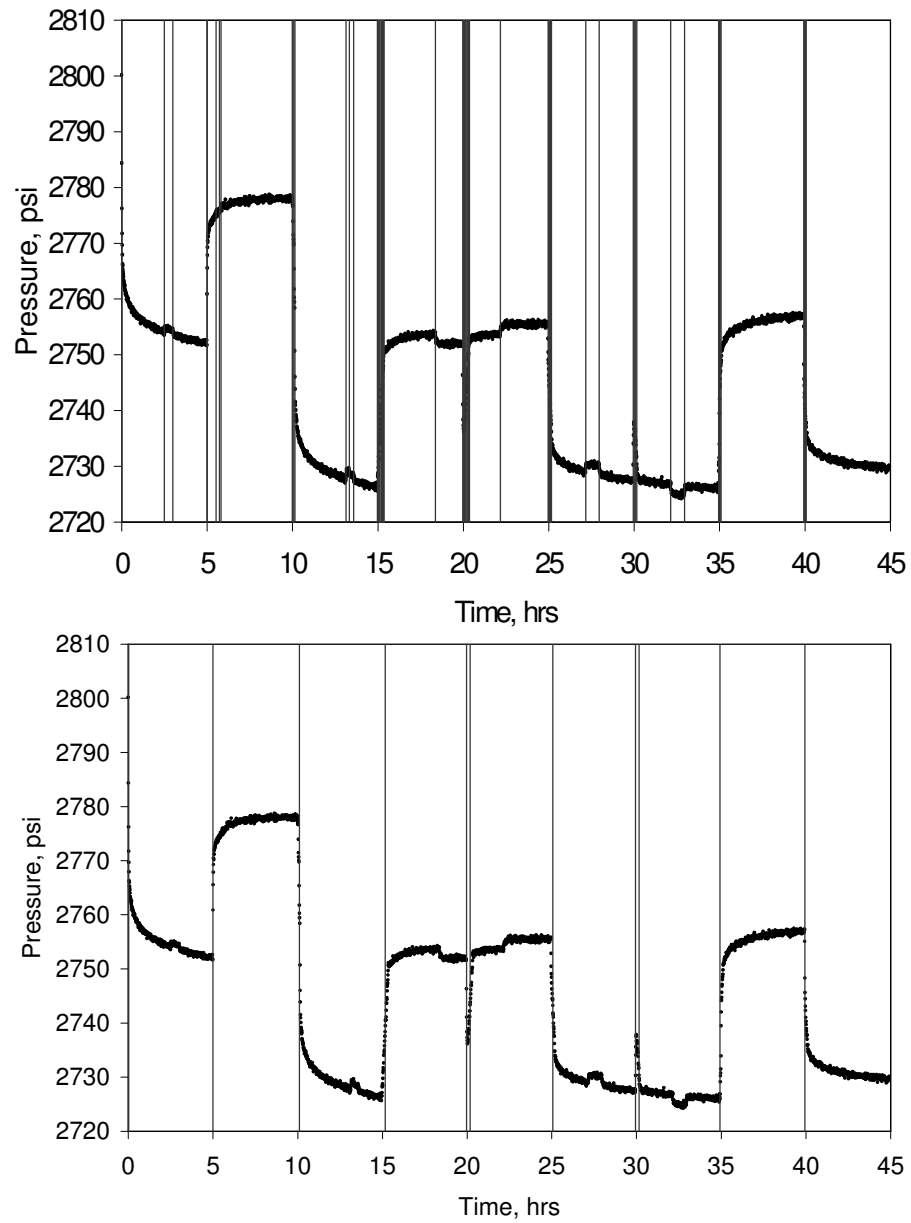


Figure 4.17: Wavelet processing results (1). Upper: true break point locations. Lower: the detected break point locations with 2 *psi* threshold.

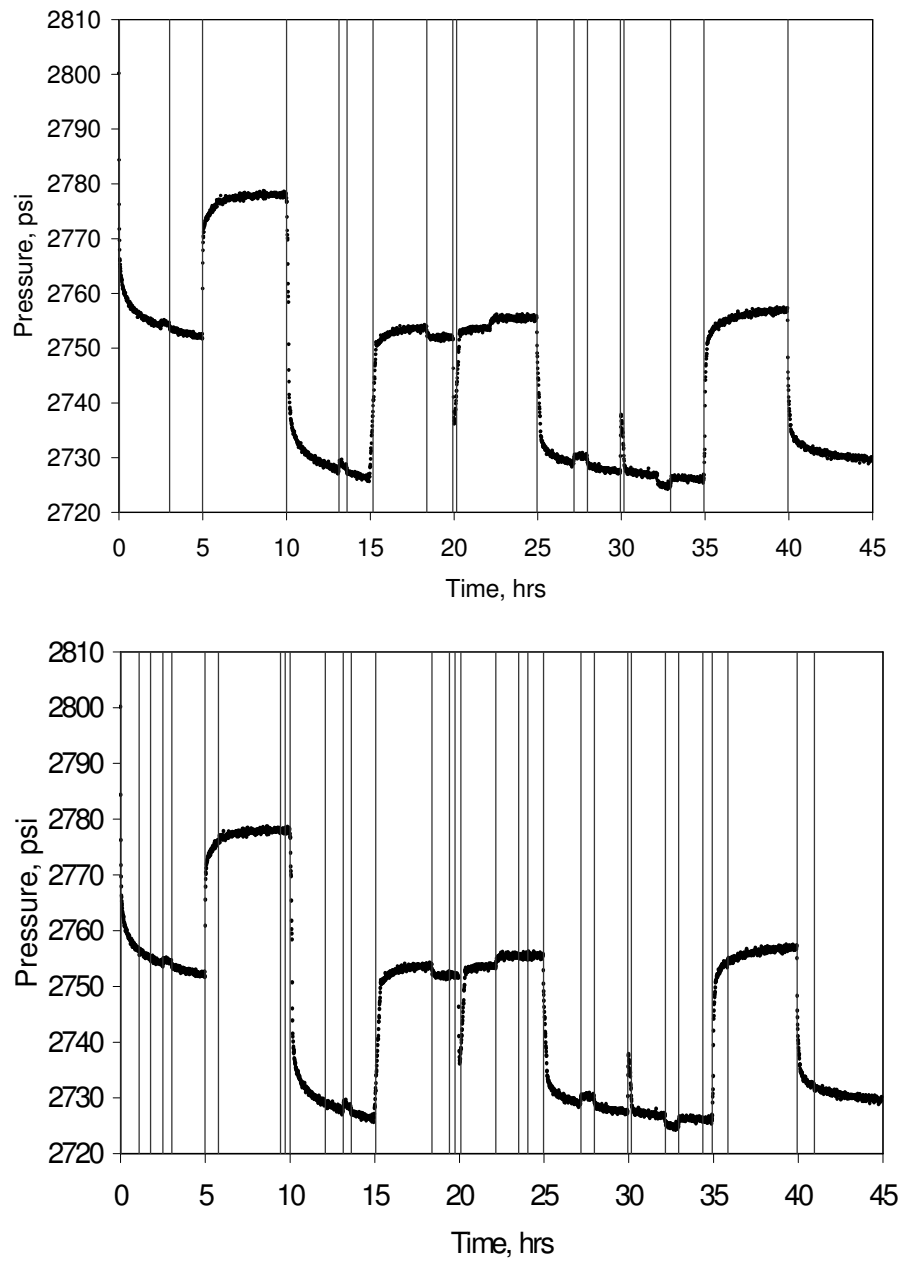


Figure 4.18: Wavelet processing results (2). Upper: the detected break point locations with 0.5 *psi* threshold. Lower: the detected break point locations with 0.01 *psi* threshold.

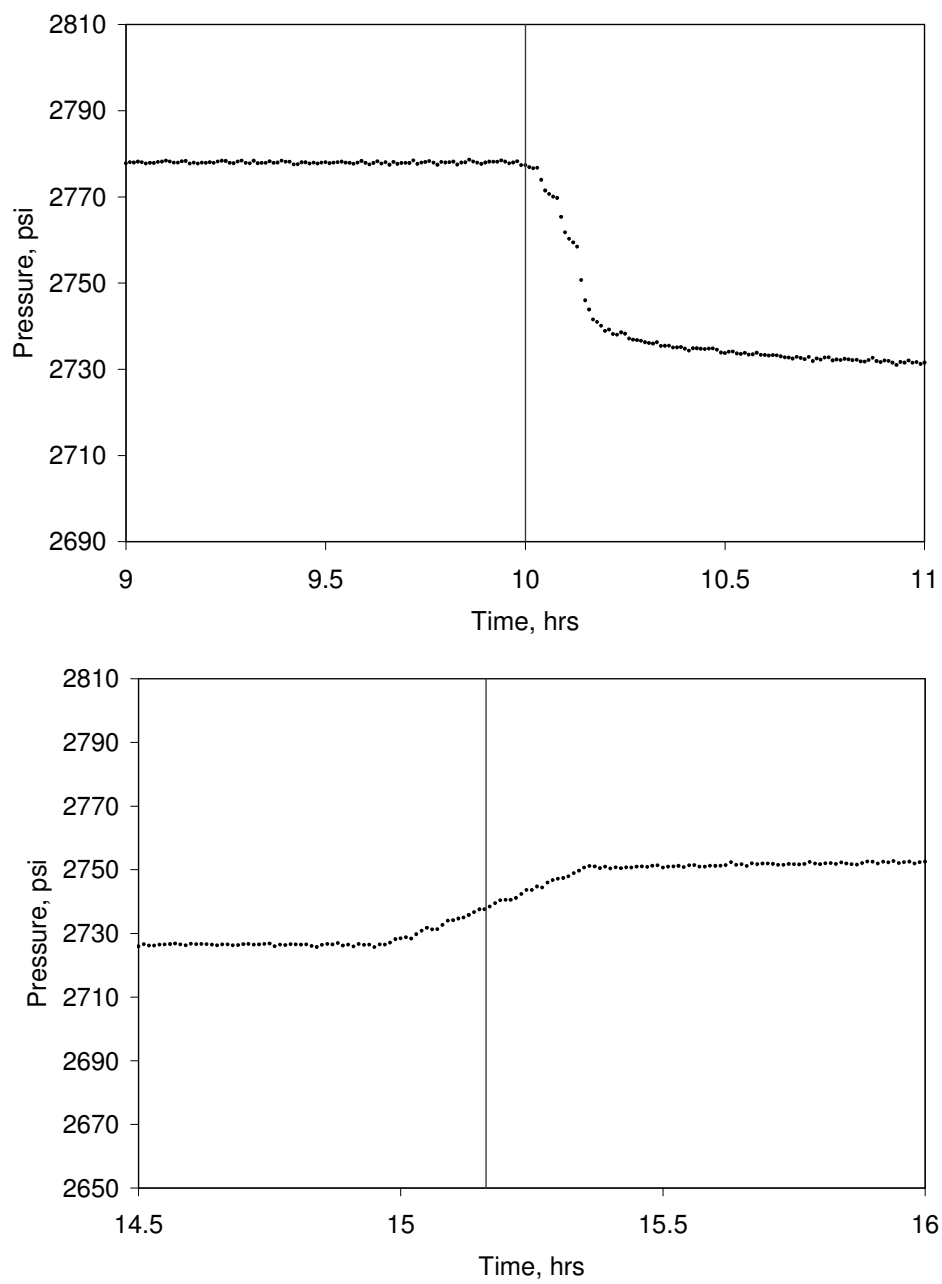


Figure 4.19: The expanded view of wavelet processing results with 0.5 *psi* threshold.

As can be seen, the effect of missing break points (insufficient processing) on the deconvolved response is much more significant compared with the flow rate or pressure measurement uncertainty (as seen earlier, for example, in Figure 4.10).

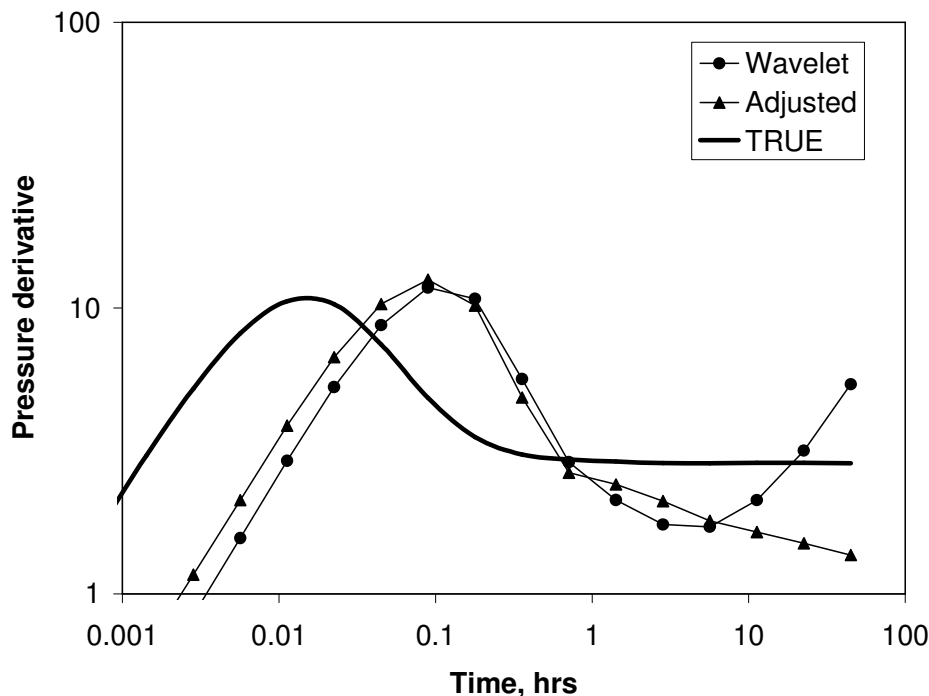


Figure 4.20: The deconvolved response using the break points with wavelet. Circle: w/ wavelet-detected break points and triangle: w/ adjusted break points.

This example suggests that any local filter-based approach has one obvious limitation. It is impossible for us to judge whether the processing results are sufficient for the subsequent data analysis by itself. In that sense, the data processing is an inseparable issue from the subsequent data analysis. However, the advantage of the wavelet-based or any other filter-based approach is that they promptly yield feasible locations of the important break points even if some of them may be missed.

4.3.2 Detection algorithm

We have been looking at the impact of the data processing results on the parameter estimation.

In this study, we treated the transient identification problem as a pressure fitting problem

which accounts for the entire process producing the given pressure signal. Again the model representation is:

$$\begin{aligned}
 P(t_i) &= \sum_{j=1}^{n_t} \sum_{p=1}^M a_j b_p S_p(t_i - tb_j) + c \\
 &= \sum_{j=1}^{n_t} a_j K(t_i - tb_j) + c
 \end{aligned} \tag{4.21}$$

We fit the convolution equation to the pressure data by adjusting the break points (tb_j) as well. This fitting problem is similar to the knot insertion problem in the spline fitting [8], although, in our case, the basis function $K(t)$ is also unknown.

Then, the objective function becomes:

$$\begin{aligned}
 Obj(\vec{b}) &= \frac{1}{2} \|\vec{d} - \vec{P}\|^2 \\
 &= \frac{1}{2} \sum_{i=1}^n \{d_i - \sum_{j=1}^{n_t} \sum_{p=1}^M a_j b_p S_p(t_i - tb_j) - c\}^2 \\
 \underline{A}\vec{b} &\geq 0
 \end{aligned} \tag{4.22}$$

The linear approximation of the nonlinear function $P(t)$ yields:

$$\begin{aligned}
 Obj(\vec{\alpha}) &\approx \frac{1}{2} \|\vec{d} - \underline{J}\vec{\alpha} - \vec{f}_0\|^2 \\
 \underline{A}\vec{\alpha} &\geq 0
 \end{aligned} \tag{4.23}$$

Here $\vec{\alpha}$ is a parameter vector ($\vec{\alpha}^T = (\vec{b}^T, \vec{a}^T, c, \vec{tb}^T)$). The number of parameter (n_p) becomes $M + 1 + 2n_t$. $\underline{J}$ is a $(n \times n_p)$ jacobian matrix and $\vec{f}_0$ is the model value at the current parameter values.

Similarly, for the k active constraints and the damping parameter w the solution and the approximated signal are given by:

$$\vec{\alpha} = \{\underline{Z}_k^T \underline{G} \underline{Z}_k\}^{-1} \underline{Z}_k^T \underline{J}^T (\vec{d} - \vec{f}_0) \tag{4.24}$$

$$\vec{f} = \underline{J}\underline{Z}_k\{\underline{Z}_k^T\underline{G}\underline{Z}_k\}^{-1}\underline{Z}_k^T\underline{J}^T(\vec{d} - \vec{f}_0) + \vec{f}_0 \quad (4.25)$$

Here, $\underline{G} = \underline{J}^T\underline{J} + w\underline{I}$.

Then the problem becomes finding the break points in the pressure signal with the nonlinear smoother. In this problem the number of break points is another unknown parameter. Minimizing the objective function over all possible locations for a given number (n_t) of break points is a fairly difficult computational task. Moreover, optimizing the number of break points is an even more formidable problem. Here we describe the procedure to seek an approximate solution in order to enable the pressure analysis from the data smoothing aspect.

We start with the subset of locations defined by the distinct values realized by the data set. Namely, we employ the wavelet processing result as the initial guess for the number and location of the break points. The wavelet processing gives potential break point locations and is expected to capture the global nature of the data. Then we try to capture the local nature of the data through iterative pressure fitting.

Figure 4.21 shows the first fitting result and its residual profile. Figure 4.22 shows its expanded view. Throughout this study, we used the break points obtained with the pressure threshold of 0.5 *psi* at the fifth level of decomposition as the initial guess. In several regions, especially around the start of large transients, the fitting error is significant, creating the higher residual.

Powell [25] considered a criterion for testing the trend of the fitting error for the spline fitting problem. The test is based on the value of the expression:

$$R = \sum_{k=p+1}^{p+q} r_{k-1}r_k \quad (4.26)$$

Here r_k is the residual for the k th data point and $q + 1$ is the number of data points between the break points (knots). Powell [25] defined the criterion for trend testing such that there is a trend in the fitting error if the calculated R is greater than its statistical mean:

$$R \geq \sqrt{q} \sum_{k=p}^{p+q} r_k^2 / (q + 1) \quad (4.27)$$

By rearranging:

$$R(q+1) \bigg/ \sum_{k=p}^{p+q} r_k^2 \sqrt{q} \geq 1 \quad (4.28)$$

Figure 4.21 shows the R values for each region partitioned by the break points. Intuitively the R value may be a good indication of the missing break points, since missing break points produce a large residual trend due to the larger pressure change around the break points.

Then we try to improve the fitting quality through break point insertion based on this information. The break point insertion procedure is summarized as follows:

- Fit the model to the pressure data by adjusting the break points (tb_j) for a given number of break points (n_t).
- Calculate the R value of each region.
- Insert the additional break points in the middle of the potential region.
- Repeat (1)-(3) until the satisfactory condition is met.

At the third step, we tested the trend based on Equation 4.27, then selected the ten regions with the higher R values. In Figure 4.21, the arrows show the selected regions. Due to the global nature of the superposition function, the missing break points in one region affect the fitting quality in the remote regions. Therefore, we limited the number of the added break points. If the number of potential regions is less than ten, we added the break points in the regions with the higher residuals among the remaining regions. Then repeating the process improves the fitting quality in terms of the residual.

It has been found that, during the fitting, some of the break points overlap at the same location (converge to the same location). At each iteration, we checked such redundancy and then removed one of them to avoid the unnecessary breakpoints and the computational instability. Figures 4.23 and 4.24 show the pressure fitting results and the expanded view after the second iteration. As can be seen, the overall fitting quality is improved by increasing the number and adjusting the location of the break points. As the procedure proceeds, the degrees of freedom increase by inserting the additional break points, especially in the regions where the pressure signal requires more degrees of freedom to be represented. This process was repeated until the satisfactory condition was met.

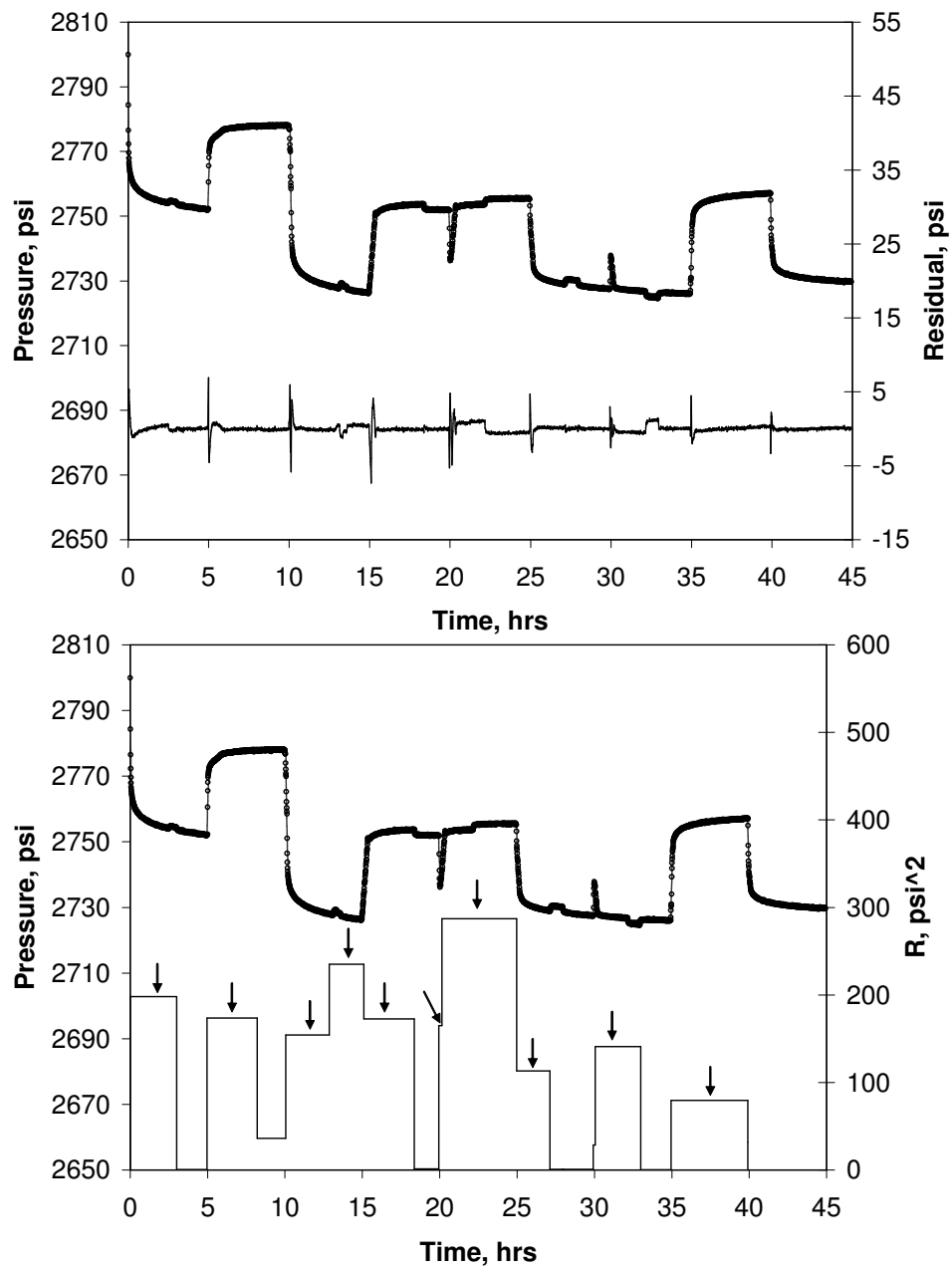


Figure 4.21: Pressure fitting result after the first iteration and the selected regions for break point insertion.

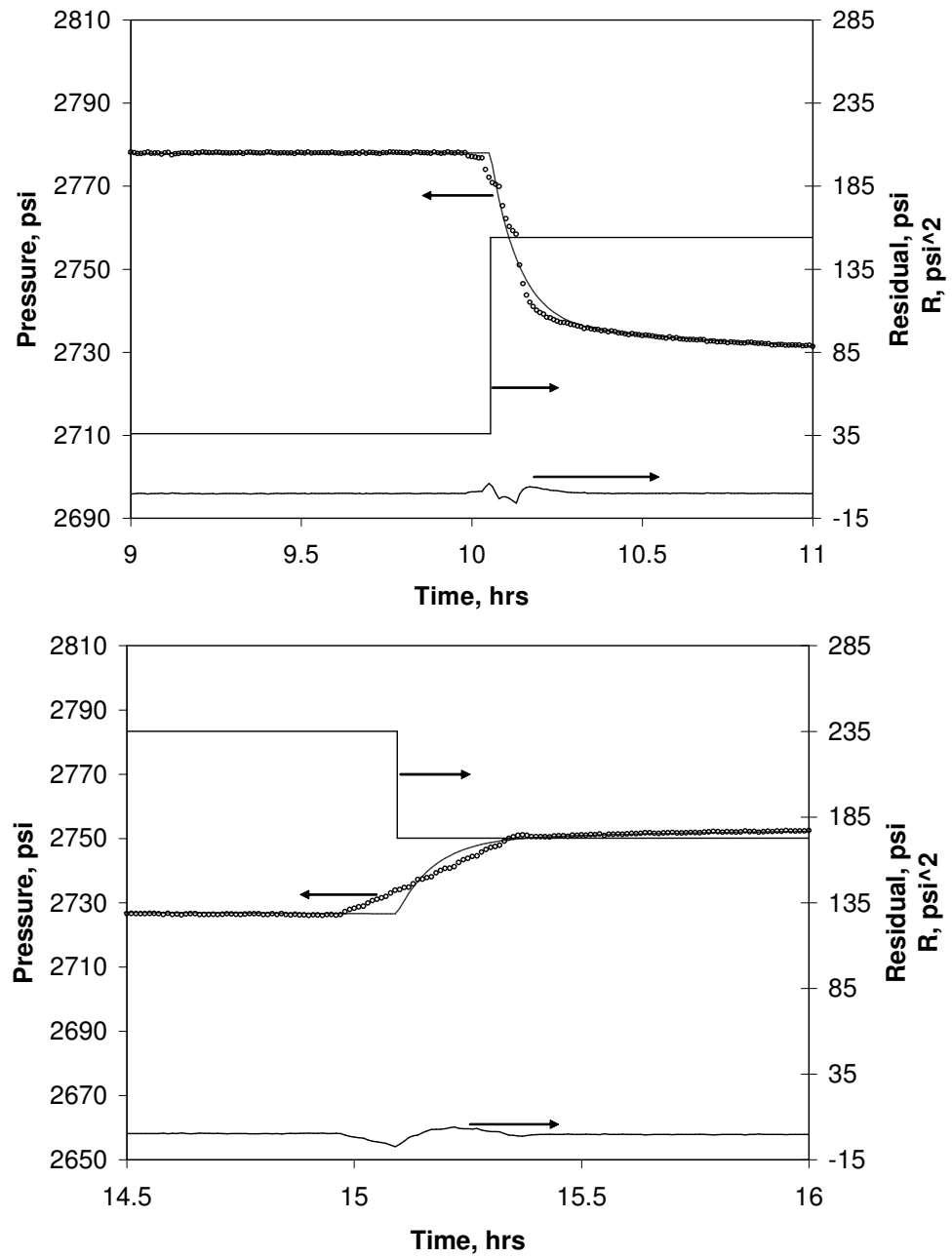


Figure 4.22: The expanded view of the pressure fitting results after first iteration.

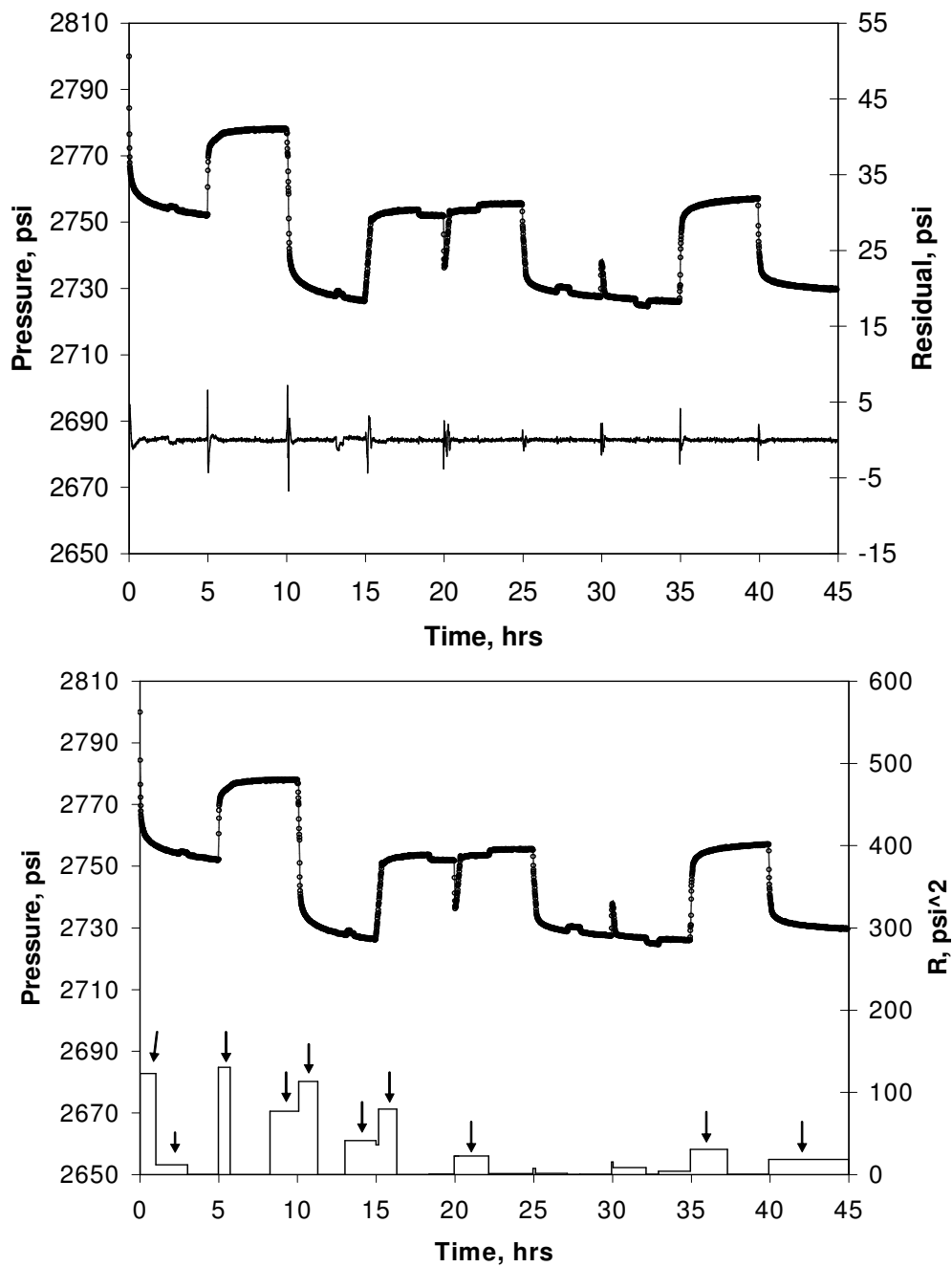


Figure 4.23: Pressure fitting result after the second iteration and the selected regions for break point insertion.

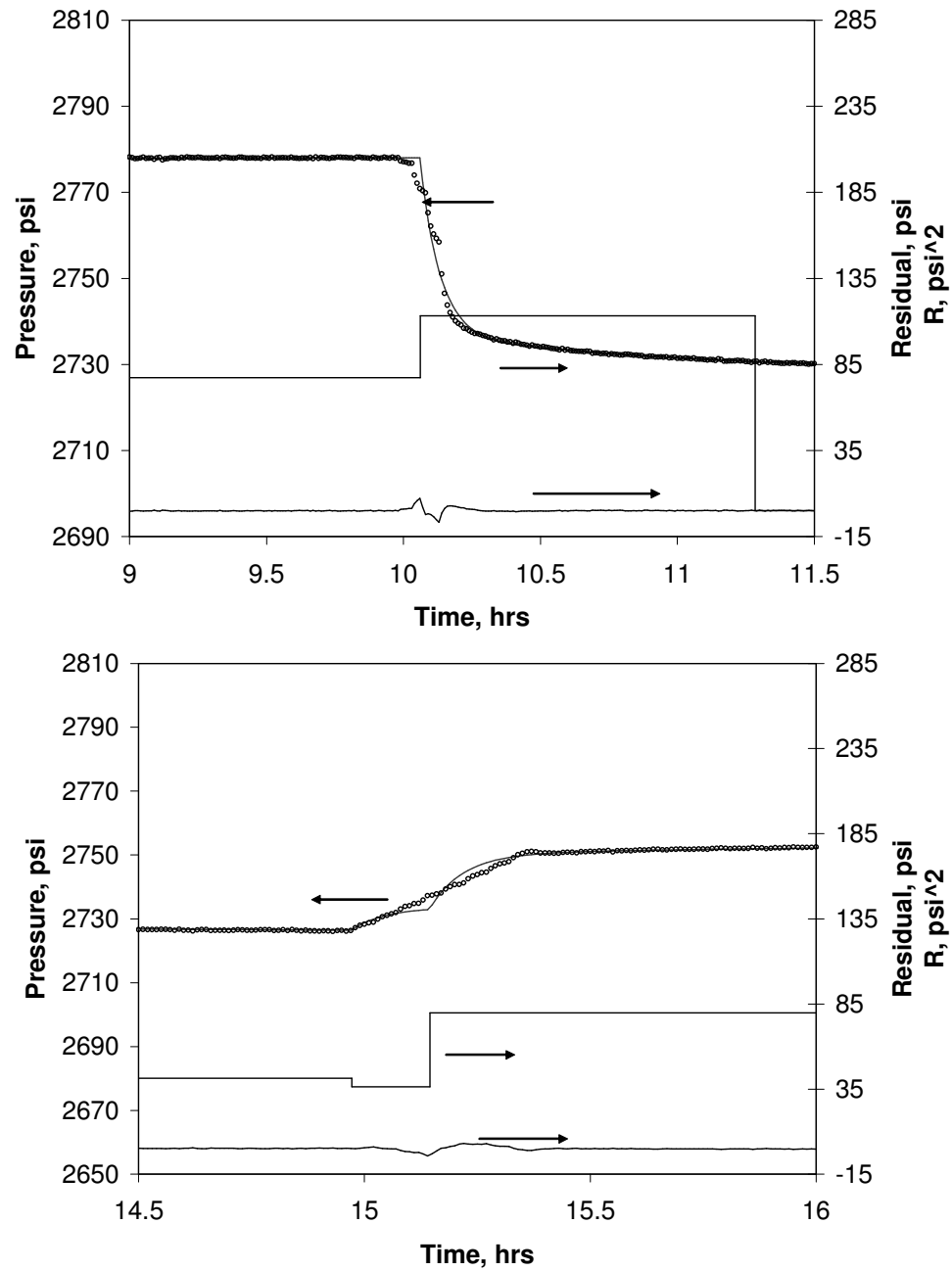


Figure 4.24: The expanded view of the pressure fitting results after the second iteration.

The procedure was designed to decrease the residual in terms of the magnitude and trend. Then continuing this procedure will produce an interpolating function. Therefore, it is important to consider the stopping criteria to avoid missing break points as well as the introduction of too many break points for the subsequent parameter estimation. Powell [25] proposed a stopping criteria using the overall trend test. Namely:

$$Rn / \sum_{k=1}^n r_k^2 \sqrt{n-1} \geq 1 \quad (4.29)$$

Then, the iterative fitting procedure is stopped if the value of the left hand side is less than 1. Here n is the total number of data.

In this study, we traced the generalized cross validation score (GCV score) calculated from the trace of the linearized smoother matrix given by:

$$\underline{S} = \underline{J} \underline{Z}_k \{ \underline{Z}_k^T \underline{G} \underline{Z}_k \}^{-1} \underline{Z}_k^T \underline{J}^T \quad (4.30)$$

The GCV may be one of the good candidates for the stopping criteria, since the cross validation generally teaches us the model complexities (overfitting or underfitting).

Figure 4.25 shows the number of break points, the degrees of freedom, the GCV score, and the error variance plotted versus the number of iterations. As the number of break points increases, the GCV score decreases drastically. At a certain iteration, the GCV score fails to decrease further and gradually increases exhibiting fluctuation while the error variance continues to decrease. The increase of GCV score may suggest that we start to overfit the model to the data globally.

Figure 4.26 shows the final fitting results. Figure 4.27 shows the estimated and true break point locations. The number of detected break points was 136 compared to the true number 57. The residual profile is well flattened indicating the goodness of the overall fitting. Figures 4.28 shows expanded views of the fitting results and the estimated break point locations. In the figures, the circles show the true break point locations. As can be seen, the estimates approximate the location of the true break points reasonably well, although some redundant break points tend to be detected.

The same procedure was applied to the same data set but using different initial guesses obtained from different thresholds. Table 4.2 shows the number of the detected break points

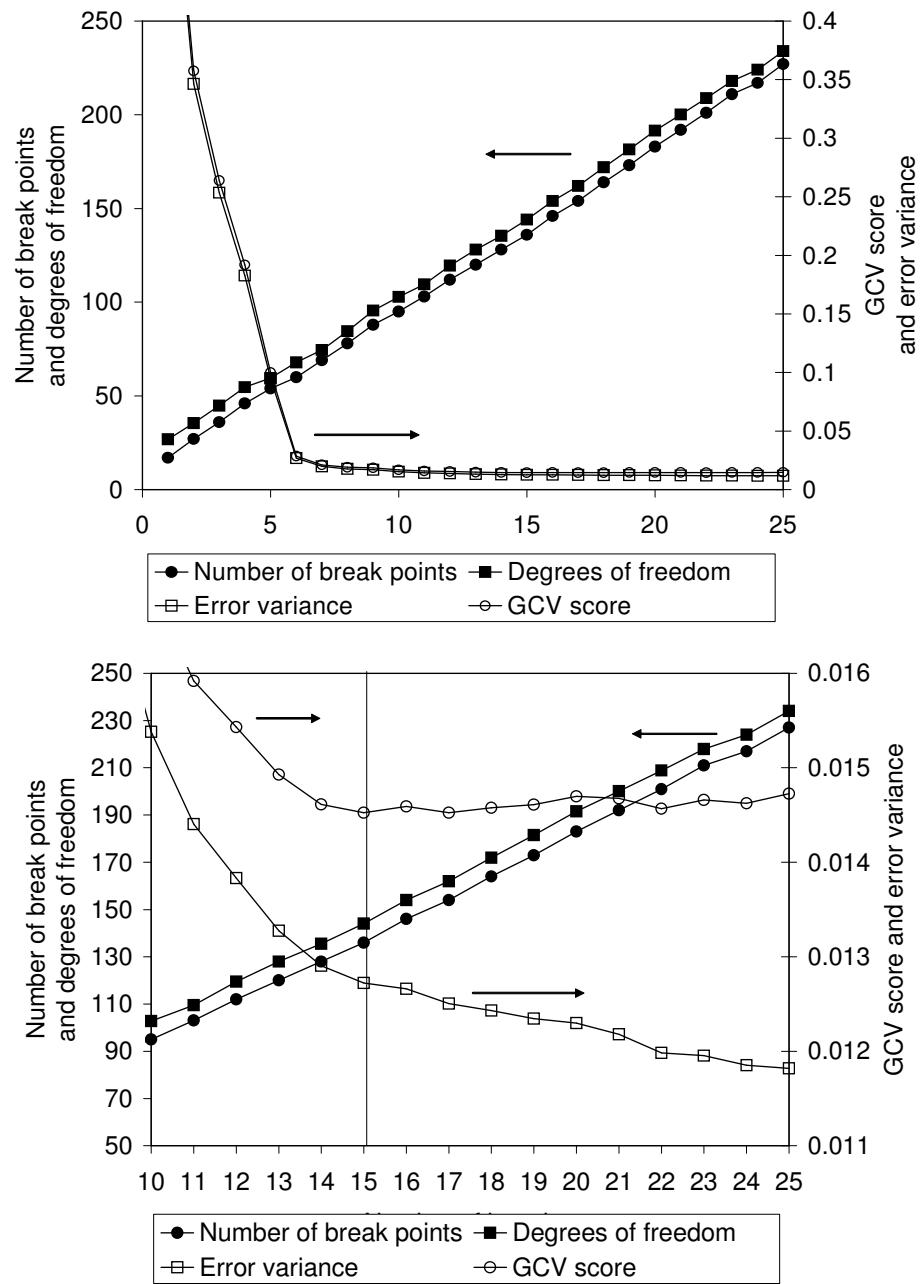


Figure 4.25: GCV score during the break point insertion.

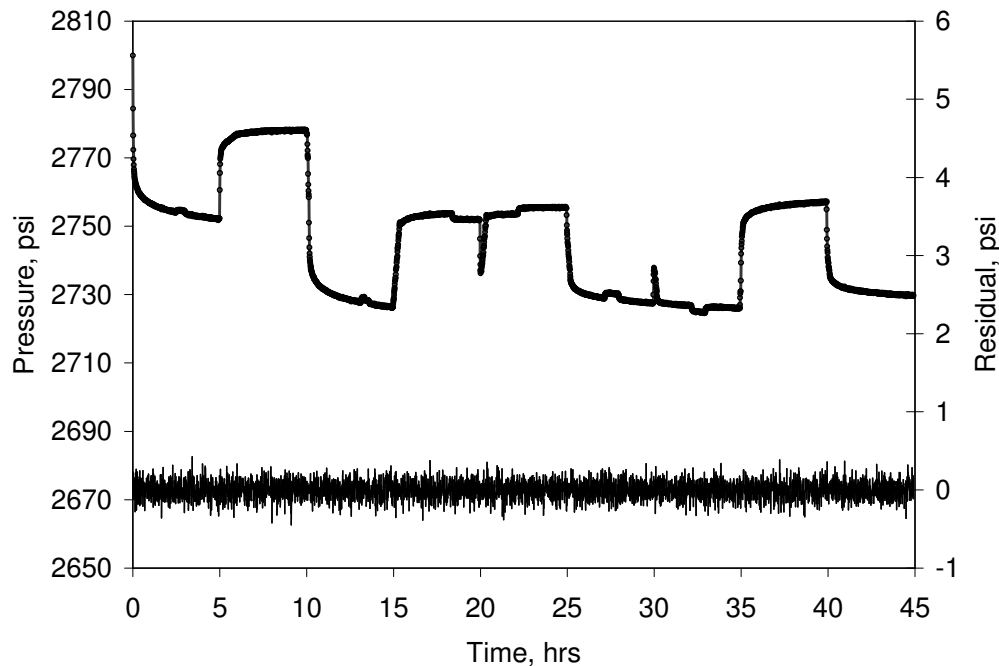


Figure 4.26: Final fitting results.

Table 4.2: The number of break points and GCV score (insertion scheme).

case	threshold	initial guess	estimation	GCV score
1	0.5	18	136	0.01452
2	0.01	36	167	0.01423
3	2	11	151	0.01455

and the corresponding GCV scores. Depending on the initial guess, the number of break points varies from 136 to 167. The procedure yields a larger number of break points than the true one in all cases.

Due to the forward strategy in the break point insertion scheme, the fitting procedure tends to yield redundant break points (or local overfitting) as shown in the previous results. This indicates the potential decrease in GCV score through the backward deletion. It is also desirable to reduce the number of break points for the ease of computation in the subsequent data analysis. This consideration lead us to introduce the backward deletion scheme.

Figure 4.28 also shows the estimated effective flow rate profile. Physically speaking, the

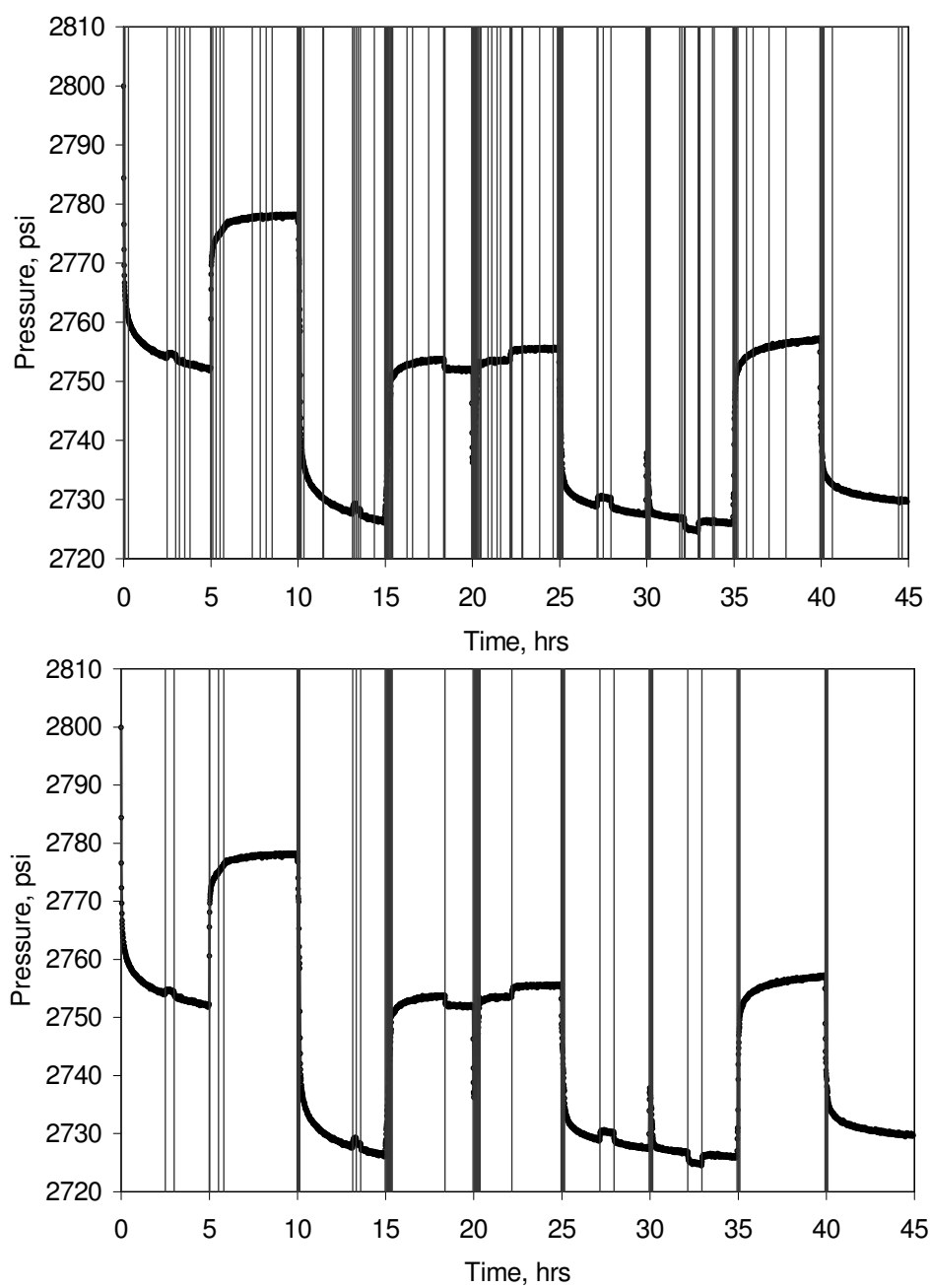


Figure 4.27: The estimated break point location. Upper: the estimated location of the break points. Lower: the location of the true break points

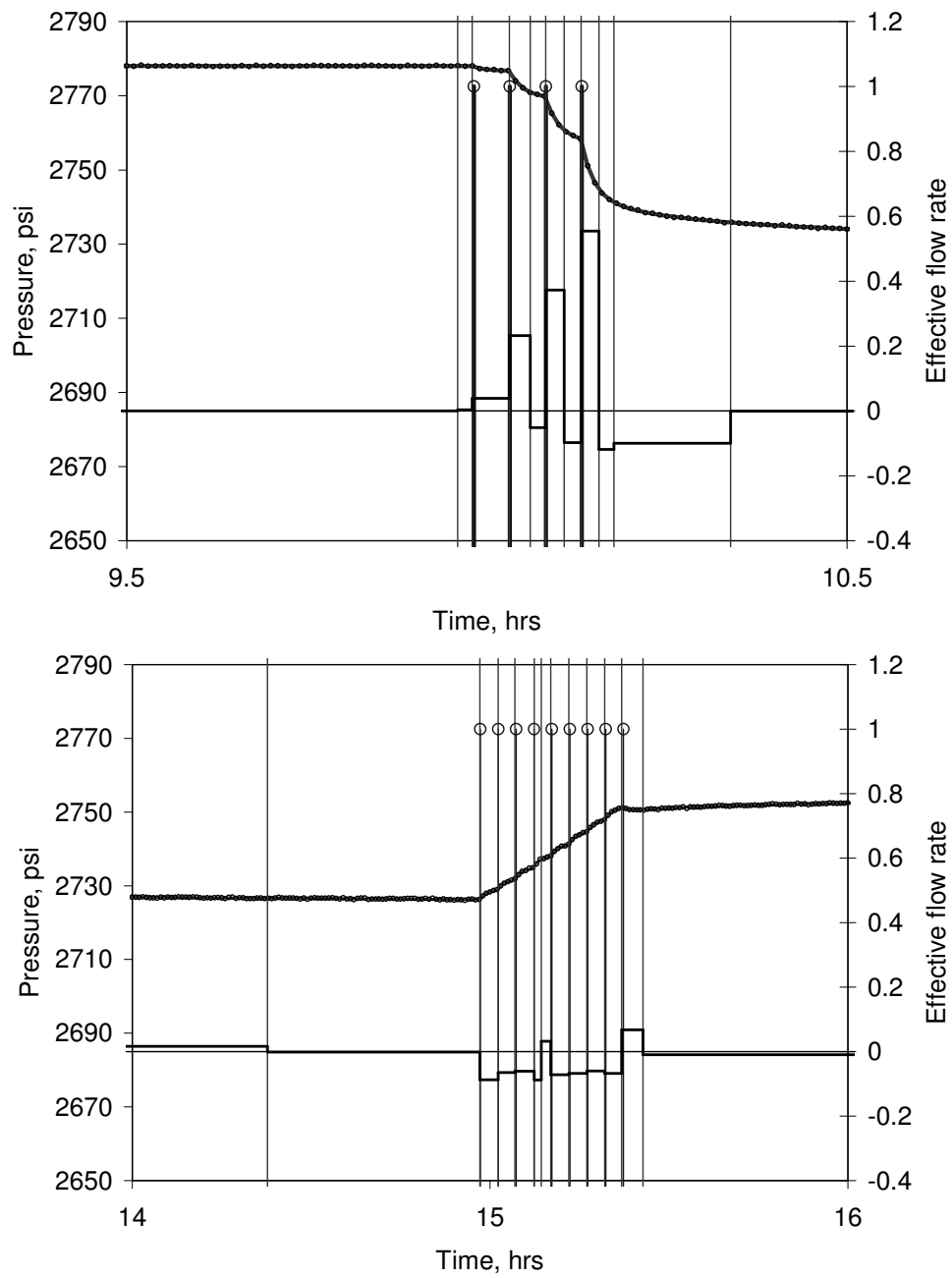


Figure 4.28: The expanded view of the estimated break point location. Open circles: true location.

magnitude of the effective flow rate reflects the importance of the transients. Therefore, during the deletion stage, we checked the magnitude of the absolute effective flow rate value, then deleted the five break points with the lowest values at one iteration. The deletion procedure is summarized as follows:

- Fit the model to the data
- Calculate the absolute value of the effective flow rate
- Remove the five break points with the lowest values

Figure 4.29 shows the GCV score profile during the whole sequence. As can be seen, GCV score decreases by removing the overfitting component effectively. We repeat this deletion until GCV score starts to increase sharply and select the solution with the minimum GCV. Figure 4.30 shows the resulting break points after the deletion scheme. The redundant break points are effectively removed compared with the previous results.

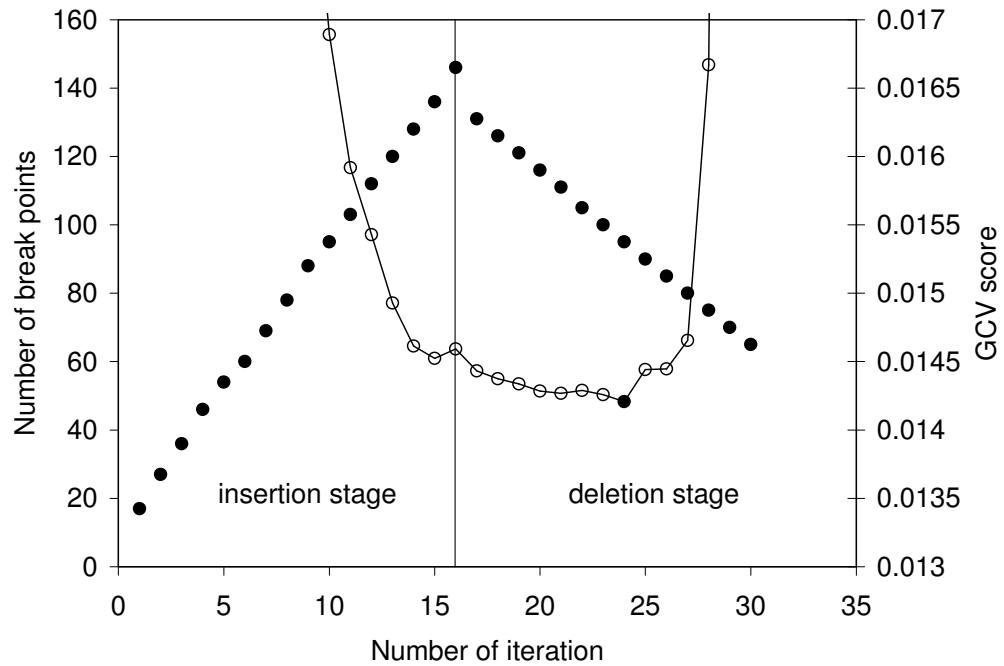


Figure 4.29: GCV score during the whole sequence.

Table 4.3 shows the resulting number of break points estimated using the different initial guesses. The deletion strategy decreases the GCV, while the number of break points are

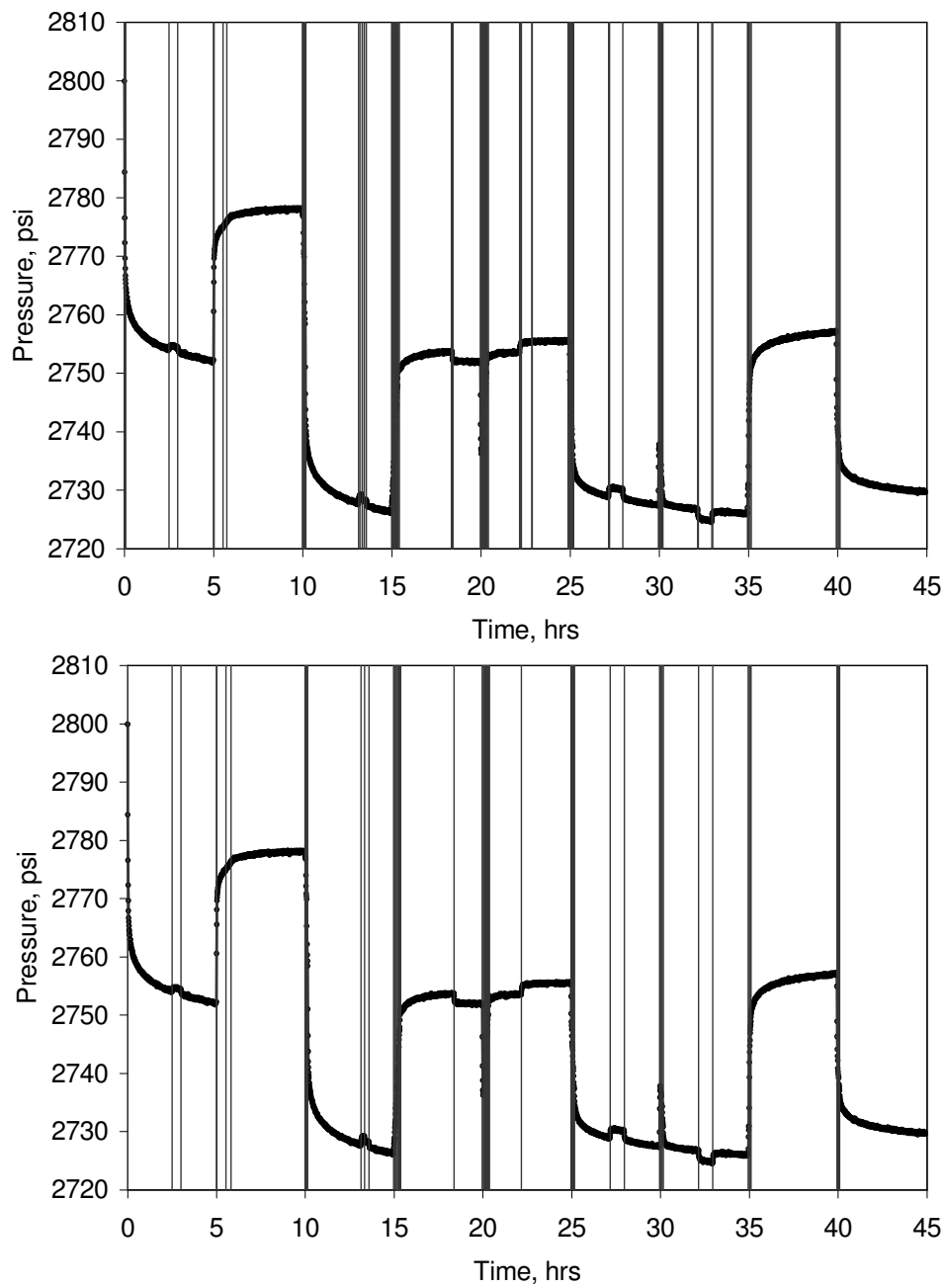


Figure 4.30: The estimated break point location. Upper: the estimated location of the break points. Lower: location of the true break points.

Table 4.3: The number of break points and GCV score (after deletion scheme).

case	initial guess	insertion	GCV score	deletion	GCV score
1	18	136	0.01452	95	0.01421
2	36	167	0.01423	81	0.01390
3	11	151	0.01455	101	0.01421

effectively reduced. The initial guess determines the forward and backward sequence and therefore the resulting number of the break points and the GCV scores are different. In terms of the GCV, case 2 gives the best solution. Using the break points in these three cases, we estimated the response function with the roughness penalty. Figure 4.31 shows the GCV score for the various smoothing parameters for case 1. As can be seen, the GCV score is further improved by the smoothing control. Figure 4.32 shows the resulting derivative estimate for three cases. In terms of the pressure derivative case 1 gave the best results despite the higher GCV value.

The difficulty in this smoothing problem is that the insertion and deletion scheme is not guaranteed to produce the best result in terms of the GCV score with respect to the number and location of the break points, since at each iteration, we try to minimize the residual of the fit for a given number of break points placed within limited locations.

Ideally the GCV score should be minimized jointly with the other control parameters such as the smoothing parameter and flow rate regularization, however the GCV minimization is formidable, since we have to check all the possible locations and numbers of the break points for each different value of the other control parameters. Moreover, another difficulty is that in this multitransient smoothing problem, we try to capture the local nature by inserting the break points just as we tried to estimate the local smoothing parameter values in the single transient smoothing problem from the finite data sample. Then as we encountered, the GCV minimization is not guaranteed to always produce a reasonable model due to the lack of local information.

Based on this consideration and lacking a better estimator, we adopted the sequential GCV minimization demonstrated in this study. However, the resulting estimates were all reasonable in practice.

The transient identification algorithm and the entire procedure are summarized in Figure 4.33 and 4.34.

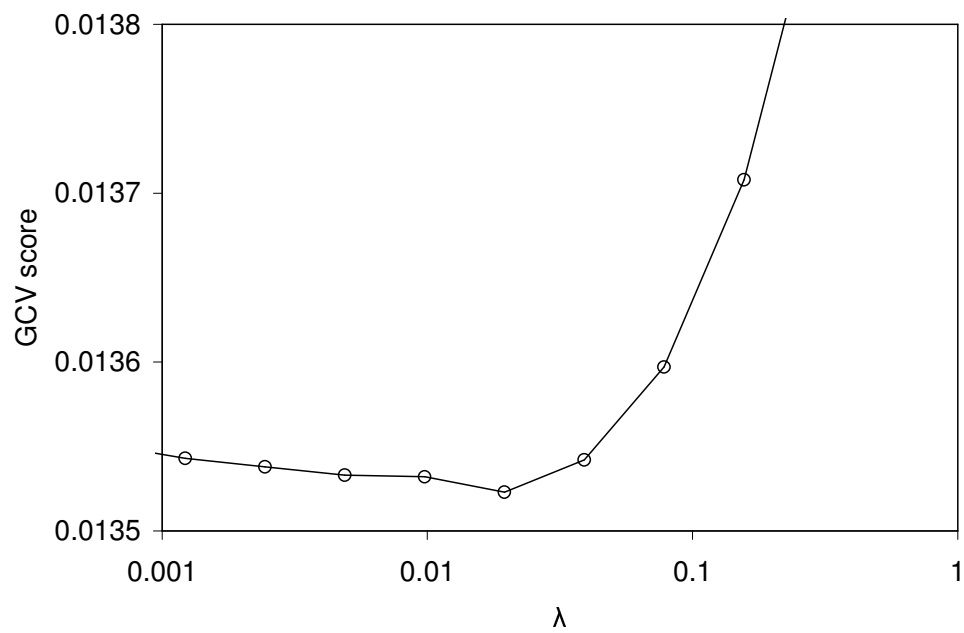


Figure 4.31: GCV score plotted versus the smoothing parameter in case 1.

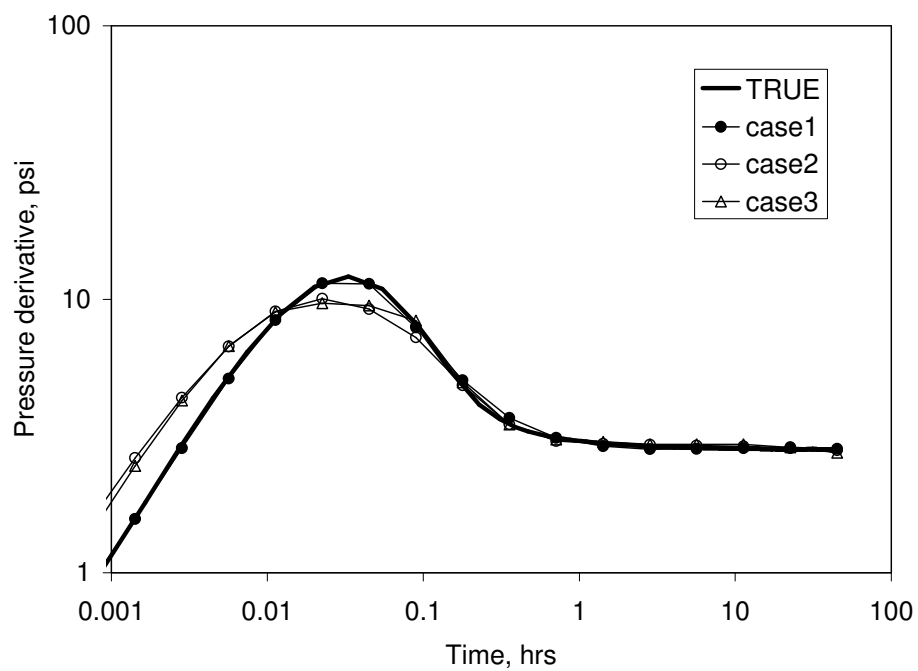


Figure 4.32: The estimated pressure derivative

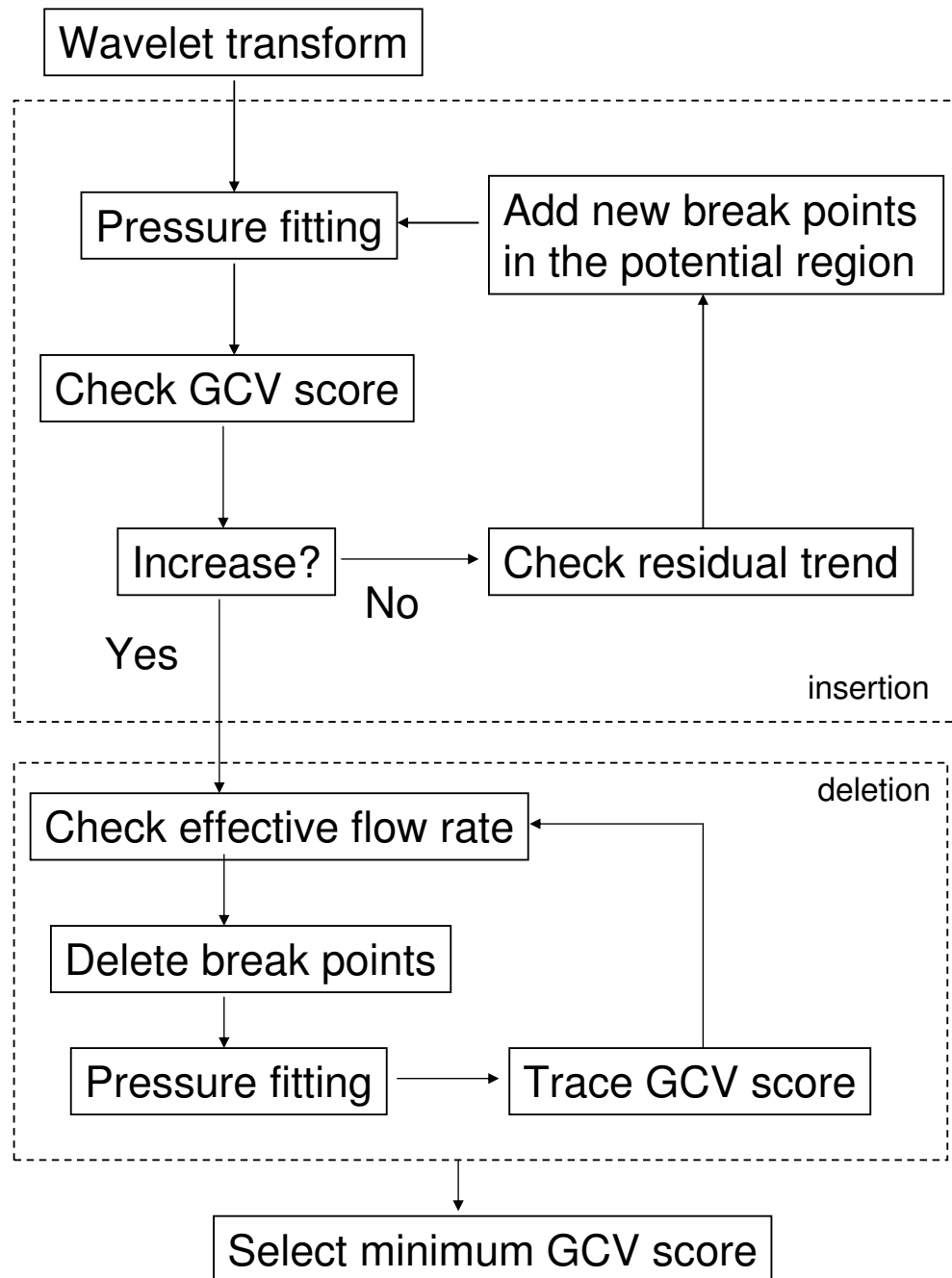


Figure 4.33: A procedure for the transient identification.

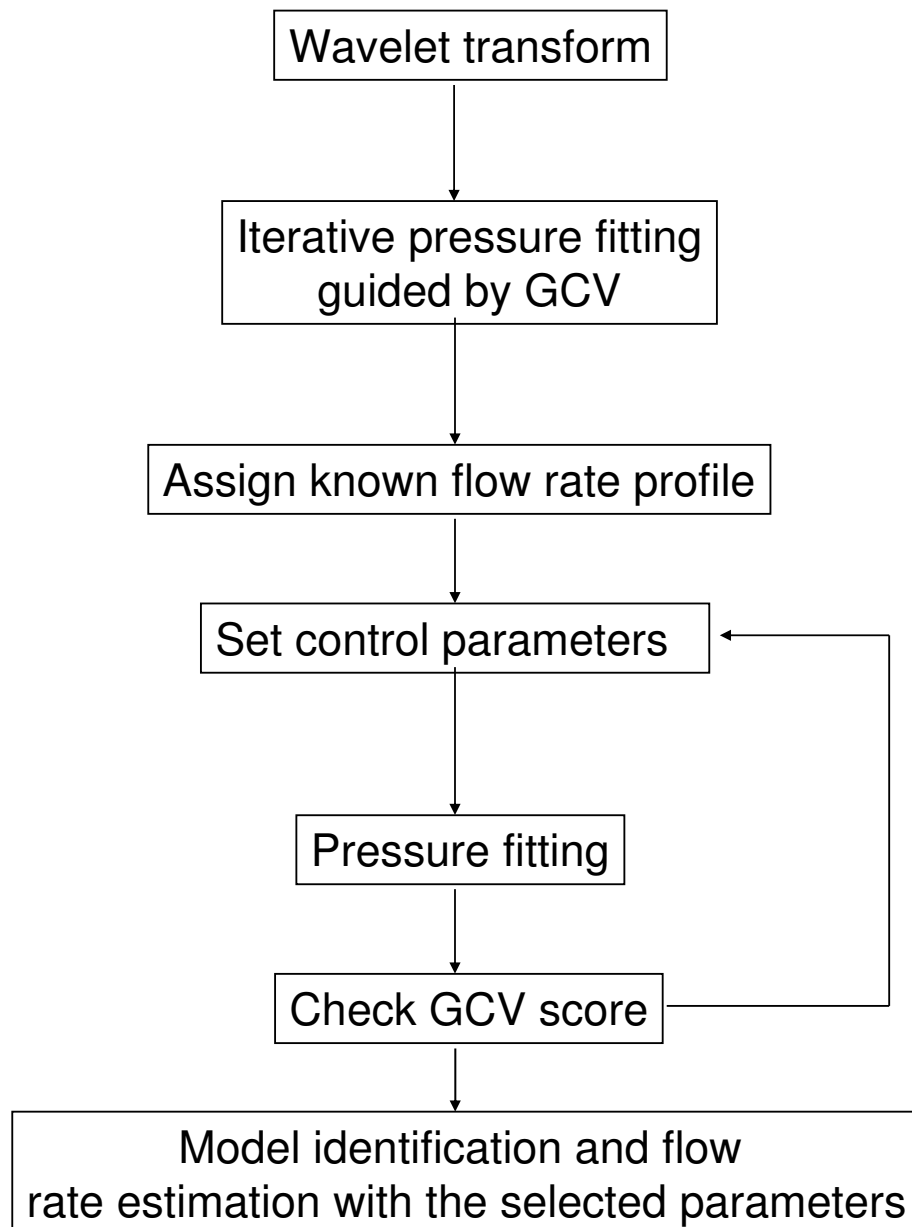


Figure 4.34: The procedure for the data analysis.

4.4 Field application

In this section, we demonstrate the application to the two field data sets. One set is conventional wireline well test data (but still multitransient) and another is permanent downhole gauge data.

Figure 4.35 shows the fall-off test data, which last 186 *hrs*. The number of pressure data is 988, and the flow rate data are not available in this data set. The wavelet processing detected 33 break points as an initial guess. Starting with this number and location of break points, we applied the insertion and deletion scheme. Figure 4.36 shows the GCV profile during the whole sequence. In this example, the detected number of break points was 120. Using these break point locations, the smoothing parameter was selected with GCV technique (Figure 4.38). Figures 4.37 and 4.39 show the final fitting result and the estimated pressure derivative. The pressure data were well fitted and the rough estimate of pressure data error was 0.25 %. Although there is no information available for this reservoir, a smooth pressure derivative was obtained exhibiting the wellbore storage behavior in the early time region.

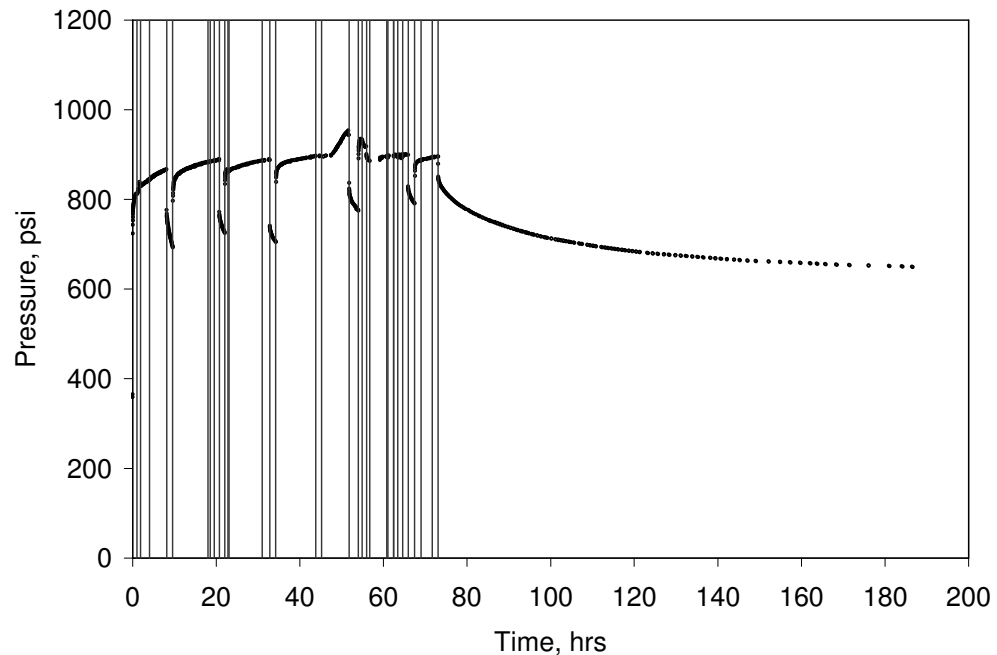


Figure 4.35: The original data and the wavelet processing results (field data 1).

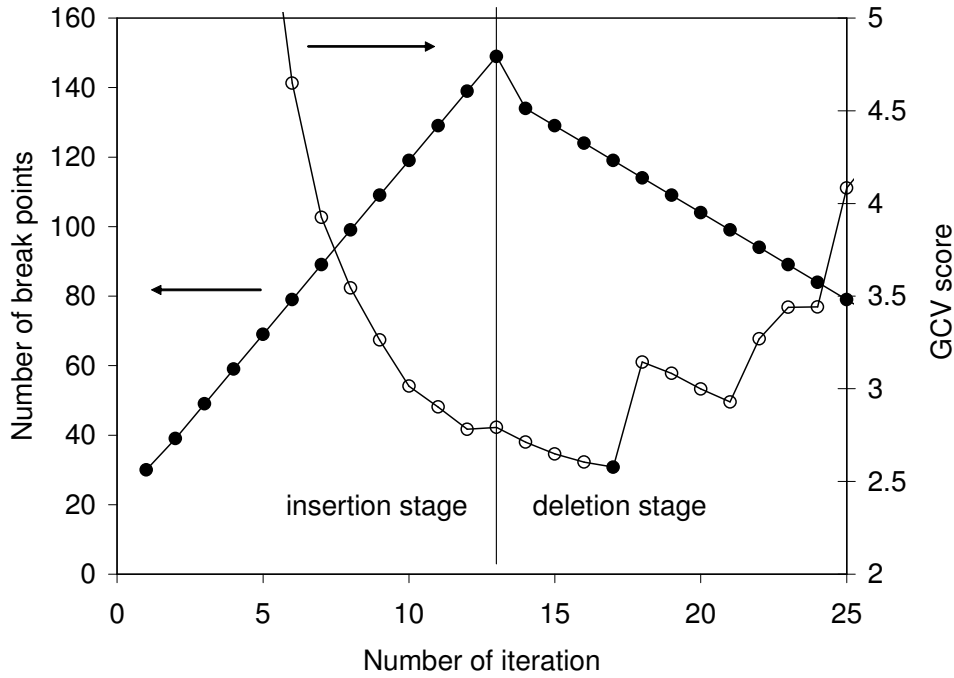


Figure 4.36: The GCV score and the number of break points.

Figure 4.40 shows the pressure data acquired from a permanent downhole gauge, which last 1167 *hrs*. The number of pressure data is 21640, and the flow rate data are not available in this data set. The wavelet processing detected 54 break points as an initial guess. Starting with this number and location of break points, we similarly applied the break point detection algorithm. Figure 4.41 shows the GCV profile during the whole sequence. In this example, the detected number of break points was 193. Using these break point locations, the smoothing parameter was selected with the GCV technique (Figure 4.43). Figures 4.42 and 4.44 show the final fitting result and the estimated pressure derivative. The pressure data was well fitted and the rough estimate of pressure data error was 0.2 %. As in the previous example, the entire procedure was applied successfully and a smooth pressure derivative was obtained.

In this data set, the previous history was not available, although the data start from the middle of the history. The missing previous history is also a further research issue.

The CPU time required for each field data analysis was approximately 3 *hrs* and 49 *hrs* on a Pentium-4 3 *GHz* processor. For the practical use, the improvement of computational efficiency will be necessary.

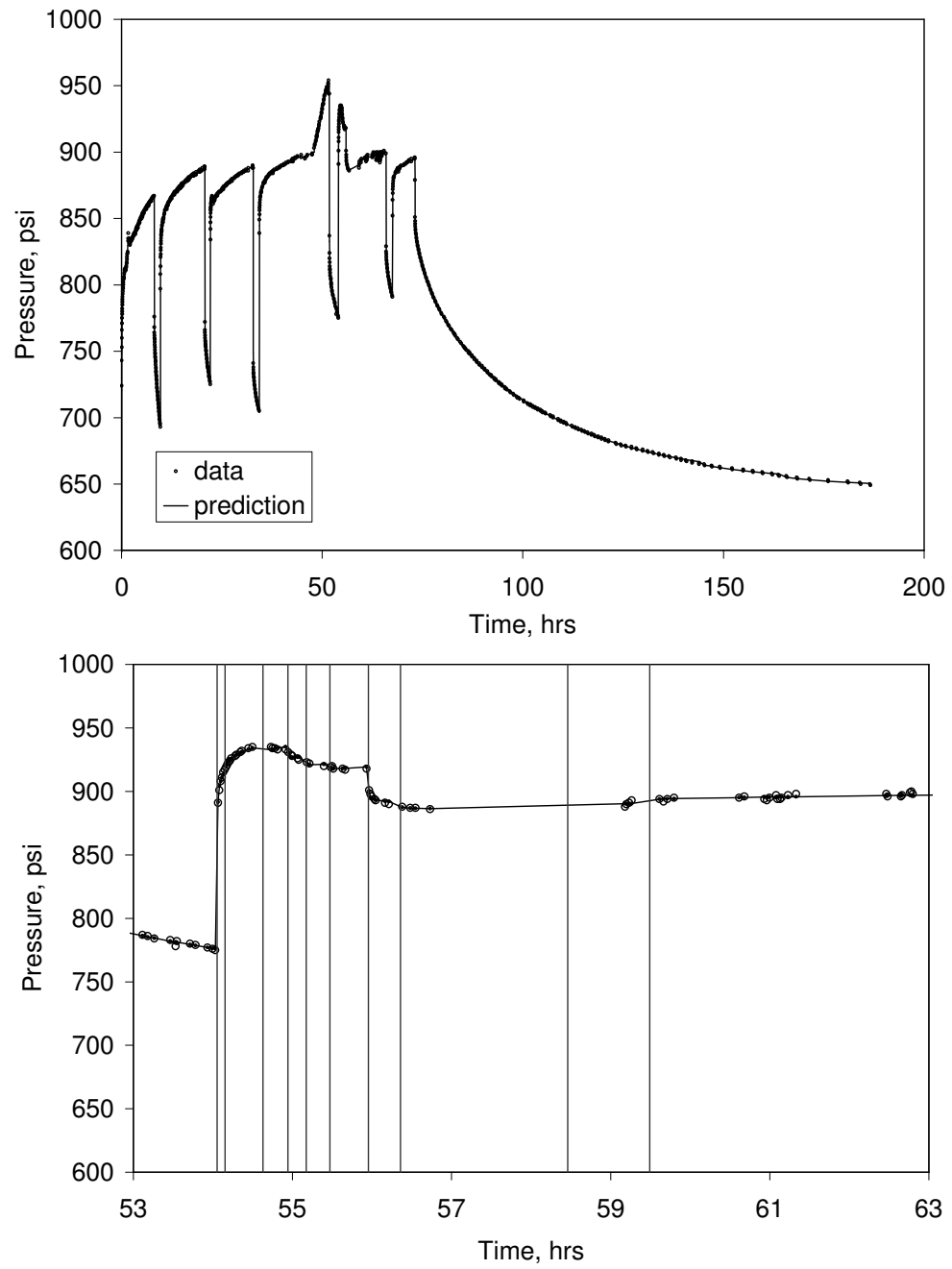


Figure 4.37: The final fitting results.

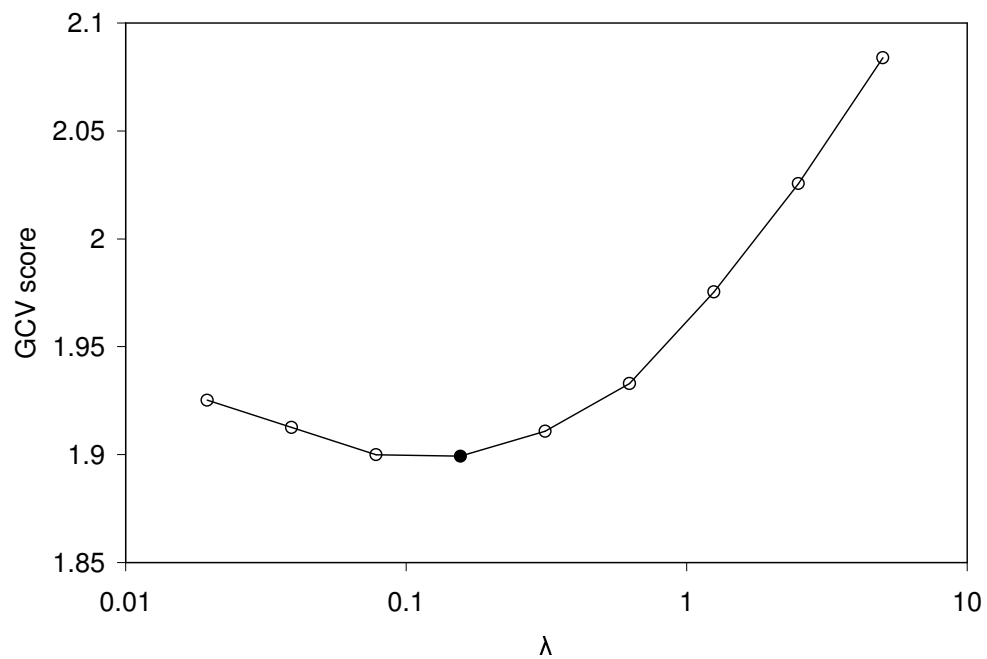


Figure 4.38: The GCV score for the smoothing parameters.

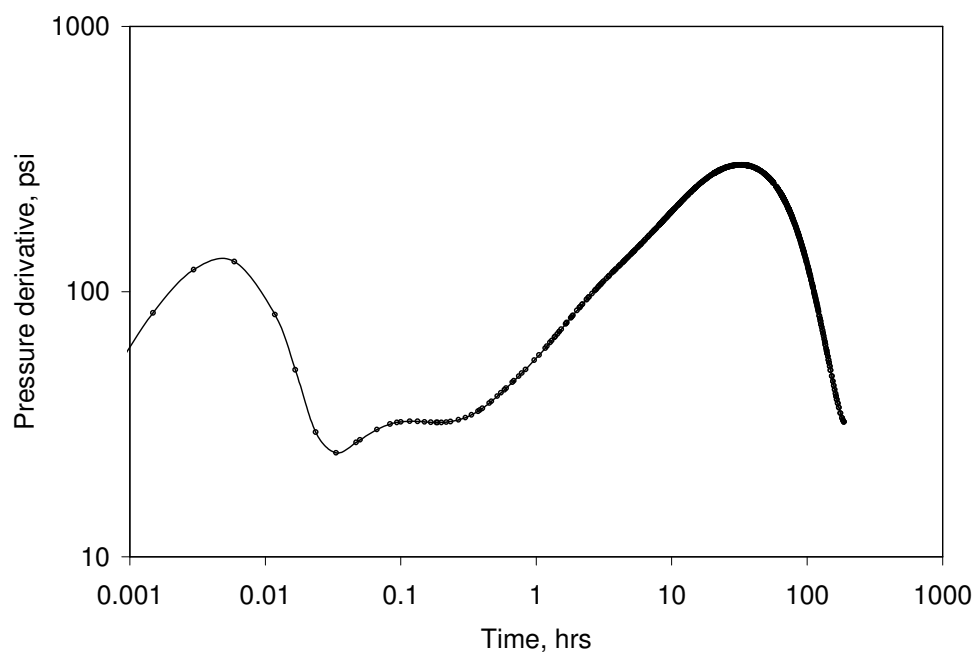


Figure 4.39: The derivative estimates with the selected smoothing parameter.

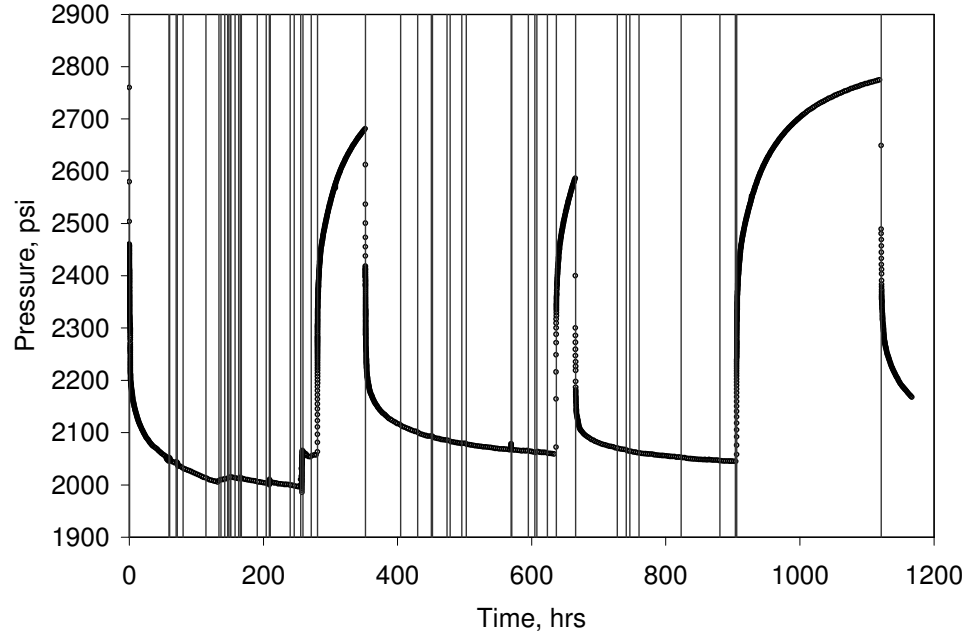


Figure 4.40: The original data and the wavelet processing results (field data 2).

4.5 Summary

In this chapter, we described the data processing and analysis method to accommodate the complexities in the permanent downhole gauge data. Although some of the points in the method may be improved, the method enabled the pressure analysis by fully utilizing the information from the measurements. Based on the numerical experiments, the following conclusions can be drawn.

- The roughness penalty and flow rate regularization lower the degrees of freedom of the nonlinear smoother represented by the convolution equation. Both control terms contribute to the improvement of the pressure derivative estimate.
- The GCV minimization enabled the joint estimation of control parameters λ and μ .
- For the expected measurement error in the actual field data, the method is likely to be sufficiently resistant. The response function and the flow rate are expected to be estimated well.
- The transient identification (break point detection) problem was described. The effect

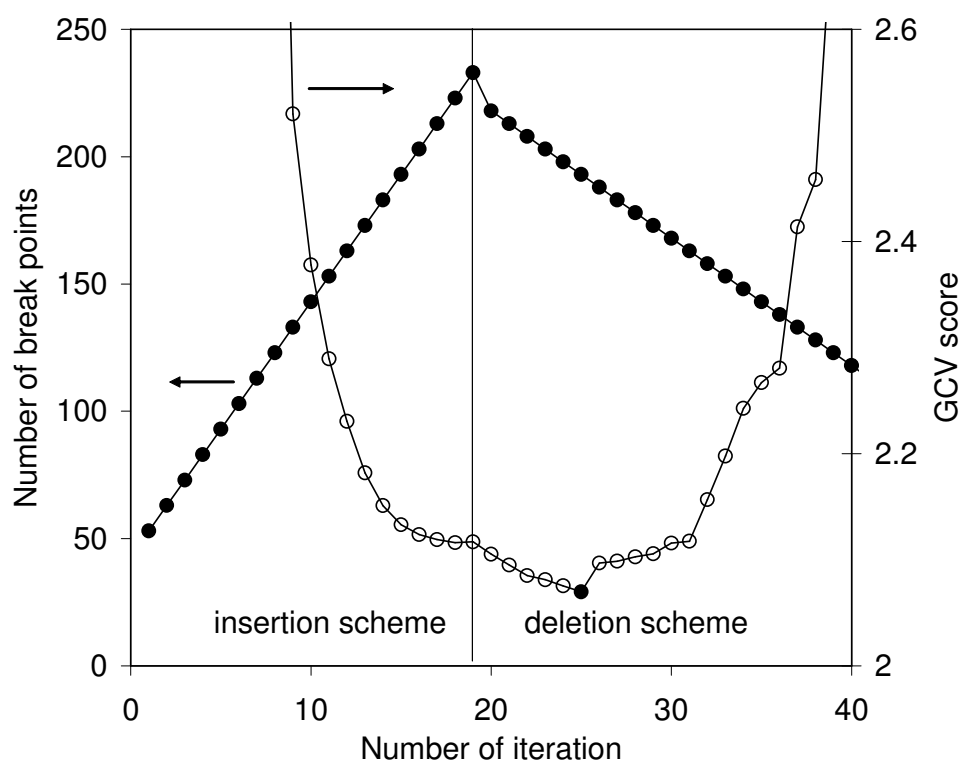


Figure 4.41: The GCV score and the number of break points.

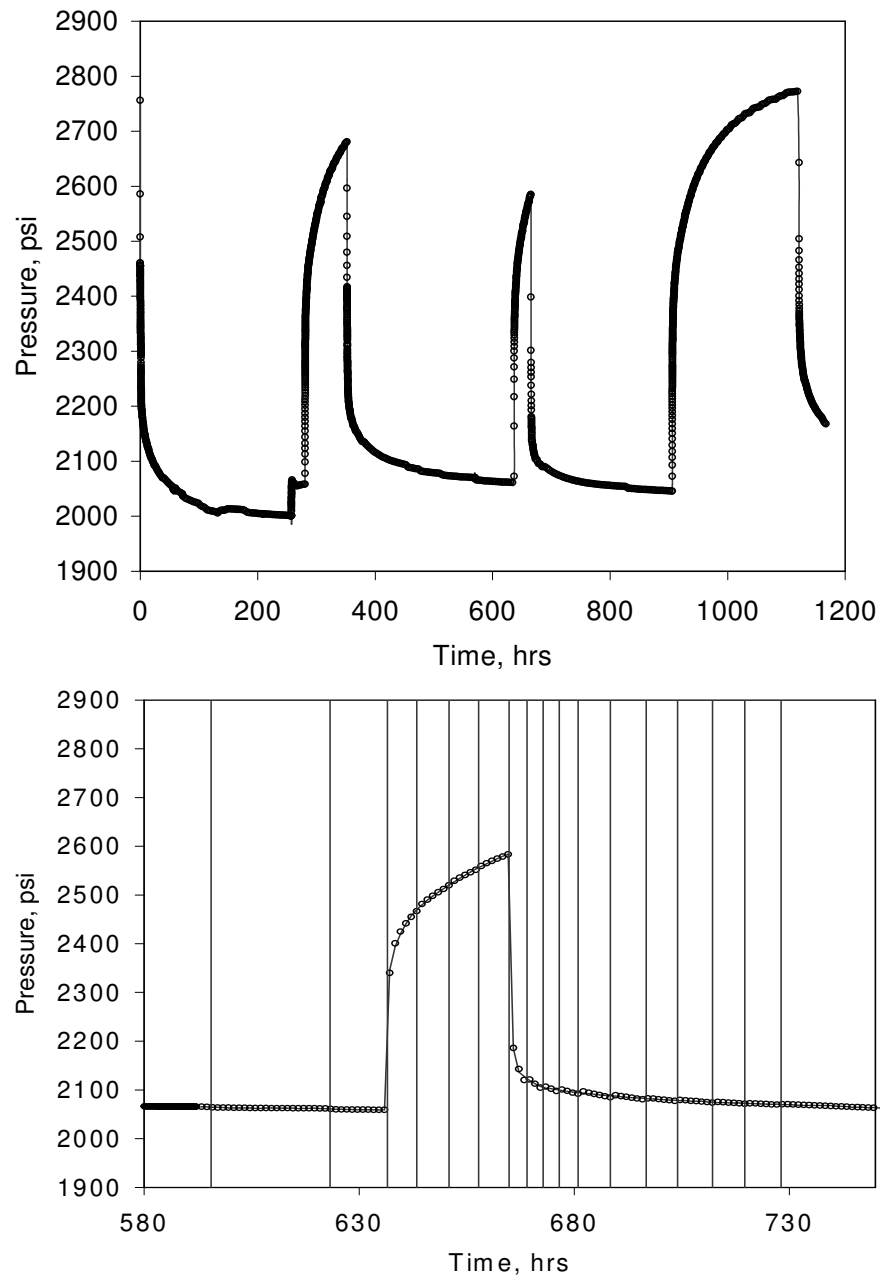


Figure 4.42: The final fitting results.

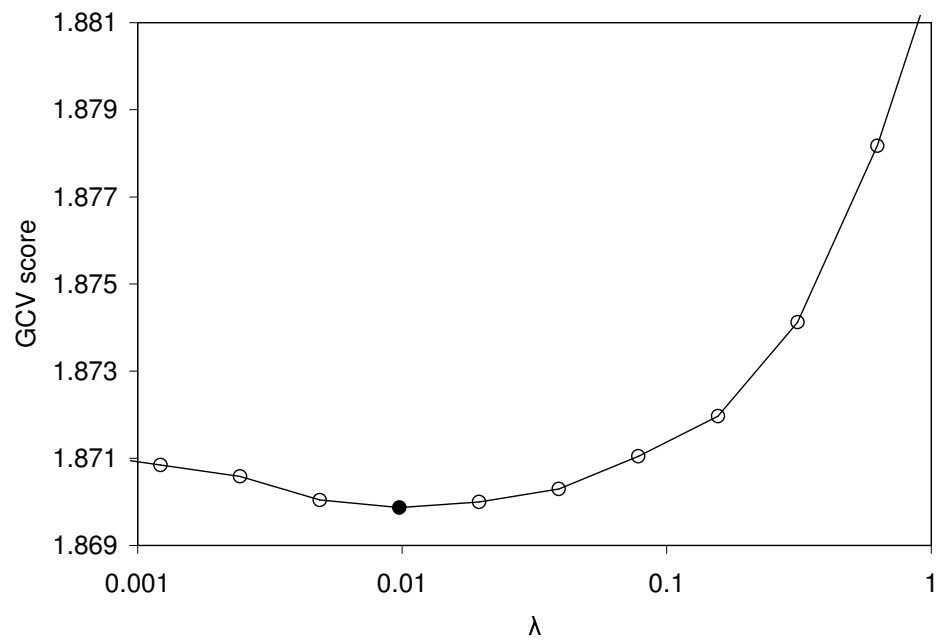


Figure 4.43: The GCV score for the smoothing parameters.

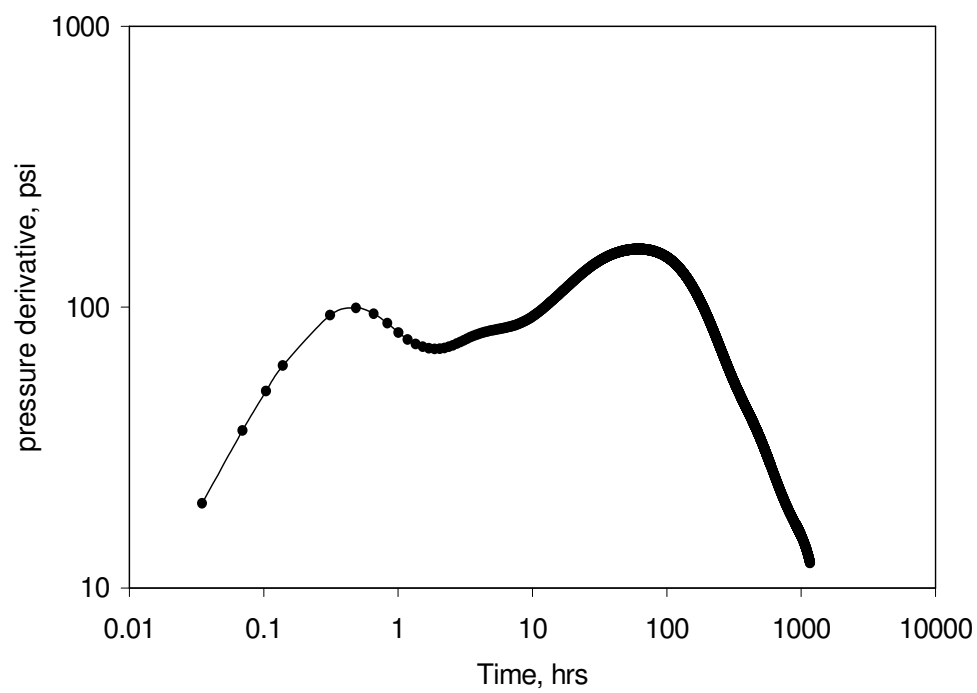


Figure 4.44: The derivative estimates with the selected smoothing parameter.

of missing break points on the estimates was much more significant than the effect of the measurement error.

- The break point detection algorithm was developed combined with the global information obtained from the wavelet representation of the signal. The local nature of the signal was well captured by the algorithm, while detecting the break points successfully.
- In the developed procedure, the model improvement was carried out based on the sequential GCV minimization in a consistent manner. The joint selection of all the control parameters may be a potential improvement. This joint selection is a further research issue.
- The developed analysis method is not subjective and is therefore suitable for automated analysis, as is our intention for the development. Through the application to the synthetic and field data sets, the effectiveness of the method was confirmed.
- Although not addressed during this study, if the pressure data has correlated noise or noise with trend (aberrant data sections), the approach will probably fail to reproduce the realistic solution. The separation of such noises from the given data is a further research issue.
- For the practical use, the improvement of computational efficiency is another issue.

Chapter 5

Conclusions and Future work

In this study, a methodology to analyze the data acquired from permanent downhole pressure gauges has been investigated. The complexities associated with the data brought a data mining aspect to the problem. The approach undertaken was nonparametric regression. During the work, we have been seeking a less subjective and more reliable method. The developed method is data-driven and enabled us to address the three technical issues with the connection to the physical behavior of the reservoir from a data smoothing aspect: (1) model identification, (2) flow rate estimation, and (3) break point identification. The applicability of the approach was demonstrated on synthetic data and on real field data.

Important remarks on the single transient smoother algorithm can be summarized as follows:

1. An effective single transient smoother was developed to obtain the interpretable estimates. In the smoother algorithm, the smoothness of the pressure derivative was achieved by the two steps: (1) pressure derivative constraints and (2) direct control with the roughness penalty. Both steps decrease the degrees of freedom of the nonparametric model, while reducing the MSE effectively. The oscillation in the derivative estimates was removed, making the interpretation easier.
2. The derivative constraints are based on the physical characteristics of the pressure transient behavior and thus reduce the MSE without increasing the bias.
3. The roughness penalty introduces bias to the estimates. The magnitude was controlled within the data allowance through the cross validation technique. The generalized

cross validation (GCV) technique was extended to the developed nonlinear smoother. The GCV technique selected the reasonable level of the smoothing parameter in all cases.

The multitransient pressure data analysis method was investigated and developed combined with a single transient smoothing algorithm. The method and results can be summarized as follows:

1. The single transient data smoother was entered into the convolution equation. The resulting formulation can be viewed as a nonlinear smoother accounting for the physical process that produces the signal.
2. To account for the measurement uncertainty and accommodate the missing flow rates, an additional regularization term on the flow rate was imposed. The smoothing control of the model response and the flow rate regularization decrease the degrees of freedom of the model while effectively reducing the MSE of the pressure derivative estimates.
3. For the moderate error cases considered in this study, the pressure derivative and the flow rate were estimated well. The method is likely to accommodate the expected range of measurement error in practice.
4. Missing break points drastically aggravate the estimate of the pressure derivative. The uncertainty of the break points was the most significant factor among the other uncertainties.
5. The multitransient smoothing algorithm was developed to identify the transients, combined with wavelet processing. Based on the global nature of the data the wavelet processing derives, the local nature was well captured by increasing the number of break points in the regions with the lower fitting quality. The GCV criterion was employed for the stopping criterion of the additional break point insertion.
6. The break point deletion scheme was introduced to reduce the redundant break points for the ease of computation in the subsequent data analysis and to decrease the GCV score. The redundant break points were removed effectively, while the GCV score was improved.

7. The multitransient smoothing algorithm has four parameters to control the model complexities: (1) smoothing parameter, (2) rate regularization parameter, (3) the number of break points, (4) the location of the break points. In this study, all the control parameters were selected based on the GCV in a sequential manner. In practice, the procedure yielded reasonable results without being involved in the difficulties in the joint GCV minimization.
8. The method was applied successfully to field data sets and its applicability was confirmed.

Although some of the points in the algorithm should be improved, the strength of this method is that it enables the pressure transient analysis in a consistent manner with less subjectivity. The following points are recommended for further improvement and future study.

1. The developed smoother may be an aid to several data smoothing applications in addition to application to the multivariate data analysis. The smoother can accommodate the monotonic and concave functions by simply changing the basis function. The development procedure of the smoother may be extended for other derivative constraints.
2. The insertion and deletion scheme in the pressure fitting algorithm may be further improved to decrease the GCV score effectively.
3. The joint GCV minimization or other selection method for model control parameters may be a further research issue including the multiple smoothing parameter selection in the single transient case.
4. In the actual situation, temporal changes of reservoir properties are expected during the long-term production. The limitation of the developed method may be investigated under such circumstances.
5. The window analysis, which restricts the fitting data sections, may be one potential approach to tackle the data aberrancy problem. The consistency check among the estimates obtained from each window analysis may suggest the potential location of the aberrant data sections.

6. The current approach treats the pressure as a function only of time. By utilizing other auxillary time series data such as temperature and flow rate (if available), the multivariate data analysis such as the ACE algorithm may be useful to detect the aberrant data sections.

Nomenclature

α coefficient

$\bar{\phi}(\underline{X})$ function in Hilbert space H_{add}

β coefficient

$\hat{f}$ estimate with fixed control parameter

λ smoothing parameter

μ regularization parameter for flow rate estimation

ν regularization parameter for flow rate estimation (in the Schroeter's method)

ν_{def}, λ_{def} default values for ν and λ

$\phi_j(X_j)$ transformation of predictor variable X_j

$\rho_k(\lambda)$ k th eigenvalue of smoother matrix $\underline{S}_\lambda$

σ^2 error variance

τ logarithm of time

$\theta(Y)$ transformation of response variable Y

$\theta_j(t)$ rate interpolants

$\triangle P(t)$ pressure drop at time t

$\underline{\Omega}$ penalty matrix

$\underline{A}$ design matrix in parametric regression

$\underline{A}_k$	k linearity constraint matrix
$\underline{C}(\vec{z})$	design matrix for pressure match
$\underline{D}$	lower triangular matrix
$\underline{F}$	diagonal matrix
$\underline{G}$	Hessian matrix
$\underline{G}'$	updated Hessian matrix with damping parameter w
$\underline{H}$	hat matrix
$\underline{J}$	Jacobian matrix
$\underline{K}$	penalty matrix
$\underline{P}$	permutation matrix
$\underline{S}^\star$	centered smoother matrix
$\underline{S}_c$	centering operator
$\underline{S}$	smoother matrix
$\underline{S}_\lambda$	smoother matrix with λ
$\underline{T}$	threshold matrix
$\underline{W}$	wavelet matrix
$\underline{Z}_k$	projection matrix in active set method
ε	error component
$\vec{1}$	constant vector
$\vec{\alpha}$	parameter vector
$\vec{\beta}$	spline coefficient vector
$\vec{b}$	unknown parameter vector for response function

$\vec{d}$	data vector
$\vec{f}$	estimated vector
$\vec{g}$	gradient
$\vec{p}$	descent direction
$\vec{Q}$	measured flow rate vector
$\vec{u}_k$	k th eigenvector of smoother matrix $\underline{S}_\lambda$
$\vec{y}$	data vector
a_j	effective flow rate at j th transient
b_p, c_q	spline coefficient
$Bias(X)$	squared bias of X
c	initial pressure
$C_{ij}(\vec{z})$	component of the matrix $\underline{C}(\vec{z})$
d_k	k th eigenvalue of penalty matrix $\underline{K}$
DF	degrees of freedom
$E(X)$	expectation of X
e^2	error variance
$f(X)$	true function to be estimated
GCV	generalized cross validation score
H_{add}	Hilbert function space $(\sum_{j=1}^p \phi_j(X_j))$
H_j	Hilbert function space $(\phi_j(X_j))$
H_Y	Hilbert function space $(\theta(Y))$
k	span for local smoother

$K(t)$	response function at time t
L	span for cross validation
m	number of pressure data points
MSE	mean squared error
N	number of transients
$N(x_i)$	nearest neighborhood of x_i
N, n	number of data point
N_λ	number of basis function to represent λ profile
N_c	number of control points for smoothing control
Obj	objective function
P_{add}	projection operator onto Hilbert space H_{add}
P_j	projection operator onto Hilbert space H_j
P_Y	projection operator onto Hilbert space H_Y
PSE	prediction squared error
$Q(t)$	flow rate at time t
R	Powell's criterion for trend testing
r_k	residual for kth data point
R_r	roughness defined at rth control point
RSS	residual sum of squares
$S_p(t)$	spline basis
t_{max}	maximum observation time
T_q	power basis of degree $q - 1$

tb_j break point time at j th transient

$trace(\underline{X})$ trace of matrix $\underline{X}$

$U_p(t)$ hat function

$Var(X)$ variance of X

w dumping parameter

$W(u)$ tricube weighting function

X predictor variable

x_i i th observation of X

$X_k(x)$ k th basis function of x

Y response variable

$Y^\star$ estimated response variable with optimal transformation

y_i i th observation of Y

$Y_i^\star$ realization at x_i

Bibliography

- [1] S. Athichanagorn. *Development of an Interpretation Methodology for Long-term Pressure Data from Permanent Downhole Gauges*. PhD thesis, Stanford University, 1999.
- [2] S. Athichanagorn, R. N. Horne, and J. Kikani. Processing and interpretation of long-term data acquired from permanent pressure gauges. *SPE Reservoir Evaluation and Engineering*, 2002.
- [3] M. Bourgeois and R. N. Horne. Well test model recognition with Laplace space type curves. *SPE Formation Evaluation*, 1993.
- [4] L. Breiman and J.H. Friedman. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, Vol.80 No.391, 1985.
- [5] E. Burg and J. Leeuw. Non-linear canonical correlation. *British J. Math. Statist. Psych.*, 1983.
- [6] Yueming Cheng, W. John Lee, and Duane A. McVay. A deconvolution technique using fast-Fourier transforms. *paper SPE 84471 presented at the SPE ATCE, Denver, Colorado, 5-8 Oct., 2003*.
- [7] W.S. Cleveland. Robust locally-weighted regression and smoothing scatterplots. *J. Am. Statist. Assoc.* 74, 829-36, 1979.
- [8] P. Dierckx. Curve and surface fitting with splines. *Monographs on Numerical Analysis, Oxford Science Publications*, 1993.
- [9] V.A. Epanechnikov. Nonparametric estimation of a multivariate probability density. *Theor. Prob. Appl.* 14, 153-8, 1969.

- [10] J.H. Friedman and W. Stuetzle. Smoothing of scatterplots. *PROJECT ORION REPORT 003*.
- [11] R.J. Gajdica, R.A. Wattenbarger, and R.A. Startzman. A new method of matching aquifer performance and determining original gas in place. *SPE Reservoir Evaluation and Engineering*, 1988.
- [12] G.H. Golub, C.F. Van Loan, and G. Wahba. Generalized crossvalidation as a method for choosing a good ridge parameter. *Technometrics*, 1979.
- [13] T. Hastie and R. Tibshirani. Generalized additive models (with discussions). *Statist. Sci.*, 1986.
- [14] T. Hastie, R. Tibshirani, and J. H. Friedman. The elements of statistical learning. Data mining, inference, and prediction. *Springer Series in Statistics*, 2001.
- [15] T.S. Hutchinson and V.J. Sikora. A generalized water-drive analysis. *Trans., AIME*, 1959.
- [16] D.L. Katz, M.R. Tek, and S.C. Jones. A generalized model for predicting the performance of gas reservoirs subject to water drive. *paper SPE428 presented at the 1962 SPE Annual Meeting, Los Angeles, 7-10 October, 1962*.
- [17] J.R. Maccord K.H. Coats, L.A. Rapoport and W.P. Drews. Determination of aquifer influence functions from field data. *Trans., AIME*, 1964.
- [18] K. C. Khong. Permanent downhole gauge data interpretation. MS report, Stanford University, 2001.
- [19] J. Kikani, P. S. Fair, and R. H. Hite. Pitfalls in pressure-gauge performance. *SPE Formation Evaluation*, 1997.
- [20] F.J. Kuchuk, R.G. Carter, and L. Ayestaran. Deconvolution of wellbore pressure and flow rate. *SPE Formation Evaluation*, 1990.
- [21] M. M. Levitan, G. E. Crawford, and A. Hardwick. Practical considerations of pressure-rate deconvolution of well test data. *paper SPE 90680 presented at the SPE ATCE, Houston, Texas, 26-29 Sept., 2004*.

- [22] F. O'Sullivan and G. Wahba. A cross validated Bayesian retrieval algorithm for non-linear remote sensing experiments. *Journal of Computational Physics*, 1985.
- [23] L. Ouyang and J. Kikani. Improving permanent downhole gauge (PDG) data processing via wavelet analysis. *paper SPE 78290 presented at the SPE 13th European Petroleum Conference, Aberdeen, 29-31 Oct. , 2002.*
- [24] L. Ouyang and R. Sawiris. Production and injection profiling: A novel application of permanent downhole pressure gauges. *paper SPE 84399 presented at the SPE ATCE, Denver, Colorado, 5-8 Oct. , 2004.*
- [25] M.J.D. Powell. Curve fitting by splines in one variable. *Numerical Approximation to Functions and Data (ed. J.G. Hayes)*, 1970.
- [26] H. Rai. Analyzing rate data from permanent downhole gauges. MS report, Stanford University, 2005.
- [27] K. Sato. Productivity correlation for horizontal sinks completed in fractured reservoirs. *SPE Reservoir Evaluation and Engineering*, 2000.
- [28] T. Schroeter, F. Hollaender, and A. C. Gringarten. Analysis of well test data from permanent downhole gauges by deconvolution. *paper SPE 77688 presented at the SPE ATCE, San Antonio, Texas, 29 Sept. - 2 Oct., 2002.*
- [29] T. Schroeter, F. Hollaender, and A. C. Gringarten. Deconvolution of well test data as a nonlinear total least squares problem. *paper SPE 71574 presented at the SPE ATCE, New Orleans, Louisiana, 30 Sept. - 3 Oct., 2001.*
- [30] O. Thomas. The data as the model: Interpreting permanent downhole gauge data without knowing the reservoir model. MS report, Stanford University, 2002.
- [31] T. Unneland and T. Haugland. Permanent downhole gauges used in reservoir management of complex north sea oil fields. *SPE Production and Facilities*, 1994.
- [32] T. Unneland, Y. Manin, and F. Kuchuk. Permanent gauge pressure and rate measurements for reservoir description and well monitoring. *paper SPE 38658 presented at the SPE ATCE, San Antonio, Texas, 5-8 Oct., 1997.*

- [33] S. J. C. H. M. van Gisbergen and A.A.H. Vandeweyer. Reliability analysis of permanent downhole monitoring systems. *SPE Drilling and Completion*, 2001.
- [34] A. F. Veneruso, C. A. Economides, and A. M. Akmansoy. Computer based downhole data acquisition and transmission in well testing. *paper SPE 24728 presented at the SPE ATCE, Washington D.C., 4-7 Oct., 1992.*
- [35] A. F. Veneruso, S. Hiron, R. Bhavsar, and L. Bernard. Reliability qualification testing for permanently installed wellbore equipment. *paper SPE 62955 presented at the SPE ATCE, Dallas, TX, 1-4 Oct., 2000.*
- [36] G. Xue, A. Datta-Gupta, P. Valko, and T. Blasingame. Optimal transformations for multiple regression: Application to permeability estimation from well logs. *SPE Formation Evaluation*, 1997.