ANALYZING RATE DATA FROM PERMANENT DOWNHOLE GAUGES

A REPORT SUBMITTED TO THE DEPARTMENT OF

PETROLEUM ENGINEERING

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

By

Himansu Rai

June, 2005

I certify that I have read this report and that in my opinion it is fully adequate, in scope and in quality, as partial fulfillment of the degree of Master of Science in Petroleum Engineering.

_____

Prof. Roland N. Horne

(Principal Advisor)

Abstract

The wavelet processing algorithms developed for processing permanent downhole pressure data by Athichanagorn et al. (2002) and further improved by Khong (2001) were investigated for their suitability in processing permanent downhole rate data. An important difference was observed in the processing of step outliers between pressure and rate data. Four new approaches for accurate and reliable identification of transient breakpoints were investigated. The possibility of rate data aiding in improved transient identification was also investigated in one of the new methods. A significant improvement in screening out small or false breakpoints while retaining the true breakpoints was observed in some of the methods. Further, the impact of continuous rate data measurements on well test interpretation was investigated.

Acknowledgments

I would like to express my gratitude toward Professor Roland N. Horne for his continuous guidance, advice and encouragement throughout the course of the study.

I would also like to thank member companies of the SUPRI-D Research Consortium for Innovation in Well Test Analysis for providing financial support for this work.

Thank also to my colleagues, especially Masahiko, Pallav, Ritesh and Birendra for their valuable suggestions in this project.

Contents

List of Tables

List of Figures

Chapter 1

1. Introduction

Over time, permanent downhole gauges (PDG) have gained popularity in the oil and gas industry, due to their effectiveness in providing real time data about the well condition. This is reflected in the increased number of wells in which permanent downhole gauges are being installed. The installation of permanent downhole gauges has the capacity to provide us with information whose use is limited only by our interpretation capabilities. PDG data can be used for determining reservoir parameters, appropriate recovery schemes for enhanced oil recovery and better reservoir management. The data can also be used for day-to-day monitoring of well conditions like the development of skin, completion performance and evaluating the performance of stimulation or workover jobs.

Previously, permanent downhole gauges were mostly used for well pressure and temperature monitoring. Nowadays, with the advances in technologies, PDG can record other parameters like total downhole flow rate, phase flow rate and phase fractions. One of the most important parameters among these is the downhole flow rate.

Earlier with the unavailability of continuous rate data, an average estimation of rate was made from the available pressure data but now the continuous sampling of rate together with pressure can help in help in eliminating this step and hence any possible ambiguity in the flow rate estimation. The combined availability of downhole rates and pressure has the potential to provide a very powerful framework for extracting more information about the well and the reservoir. This information is certainly more valuable and can contain information about the reservoir at a considerable larger radius of investigation than that obtained by a single transient test performed on the well over a limited period of time.

1.1. Problem statement

Permanent downhole gauge data are recorded under dynamic changes occurring in the well and the reservoir. Athichanagorn et al. (2002) developed a wavelet based methodology for the preprocessing of pressure data before using it for interpretation purpose. Khong (2001) introduced some important improvements in the wavelet processing methodology. The first step in this research project was to investigate the applicability of the current approach of wavelet processing to downhole rate data and note the differences with processing of pressure data.

Accurate and reliable identification of transient break points is very important for further analysis of the data. Limitations of the current time-invariant spline wavelet-based approach of break point identification were studied and four alternate algorithms were proposed for improving the accuracy and reliability of break point estimation. The proposed methods were applied to real field data and the results compared to the time-invariant spline wavelet based approach. The proposed Haar wavelet based approach did not show any tangible improvement but the other three approaches showed considerable improvement over the spline wavelet-based approach.

1.2. Report outline

Chapter 2 briefly summarizes the processing steps of the wavelet program from Athichanagorn et al. (2002) and the addition to it by Khong (2001). The chapter further discussed differences in using these steps for rate data processing and their modification to process the rate data correctly.

Chapter 3 discusses the investigation of the spline wavelet transform based break point detection algorithm leading to the understanding of how difficult it is to predict the slope threshold for detecting a transient at various levels of detail signal decomposition.

Chapter 4 discusses four alternate algorithms for identification of transient break points. The algorithms used are stationary wavelet transform using Haar wavelet, Savitzky-Golay polynomial smoothing filters, segmentation method and a variant of segmentation

method. These algorithms were applied on real field data and the results were compared with the spline wavelet based approach.

Chapter 5 presents a preliminary investigation on the impact of continuous downhole flow rate data measurements on the interpretation results of pressure transient analysis. An infinite-acting reservoir model was used to illustrate this.

Chapter 6 summarizes on the results of this research. Conclusions are drawn from the implementation of the various algorithms for break point identification in data acquired from permanent downhole gauges. Further, possible areas for additional work in the future are suggested.

Chapter 2

2. Application of wavelets to rate data processing

Permanent downhole gauges record data over a long period of time. During this period, the data acquires noise from different sources, may contain outliers or even may fail to record data at several occasions. These issues can adversely affect the data interpretation process. Hence it becomes imperative to design a data processing system that can correctly process the data for further interpretation.

To address these issues, Athichanagorn et al. (2002) developed a step by step procedure to process the pressure data. Further, Khong (2001) made some improvements and did some additions of some more steps. In a combined form, their preinterpretation data-processing steps can be summarized as follows:

Prewavelet processing

- Data overlap removal
- Step outlier removal (graphically)
- Noise threshold estimation

Wavelet processing

- Spike and step outlier removal
- Denoising
- Transient identification
- Data reduction

These steps were applied for the processing of rate data but it was observed that the zero rates which correspond to a buildup were removed as step outliers. Figure 2-1 shows a real field rate data with zero rates. Figure 2-2 shows the processed rate data. It was observed that the zero rates were removed as step outliers during wavelet processing, which is undesirable. This warranted a slight modification in the processing steps. Since we do not want to remove the zero rates which otherwise in pressure data are considered as outliers, we can skip the prewavelet processing step of outlier removal. Further, an additional constraint was imposed in the wavelet step outlier program not to remove if rate data is zero. The processed rate data with the modified algorithm is shown in Figure 2-3.

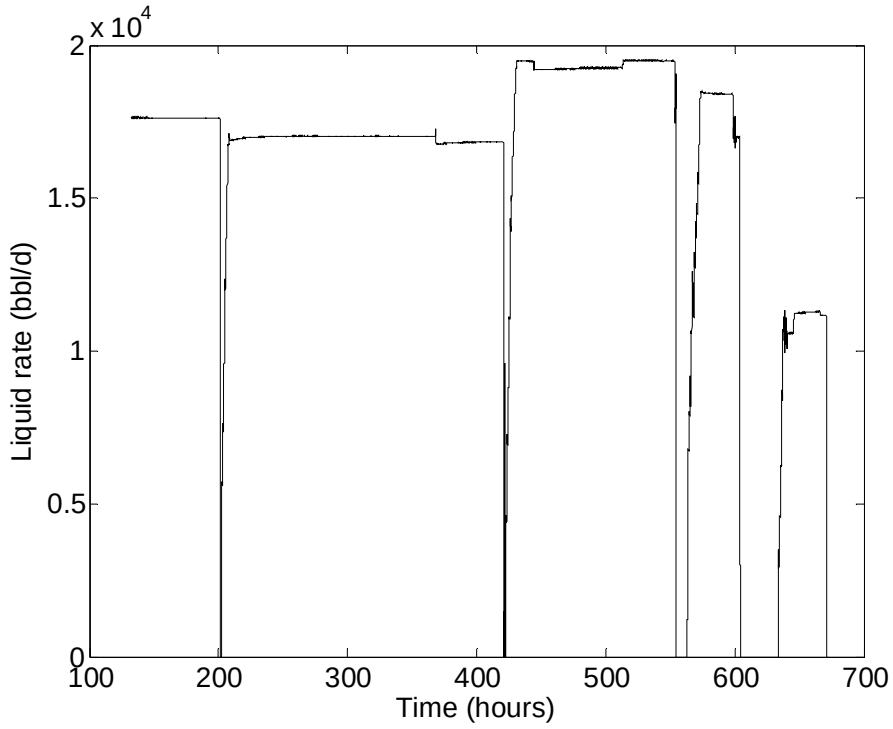

Figure 2-1: Original liquid rate data

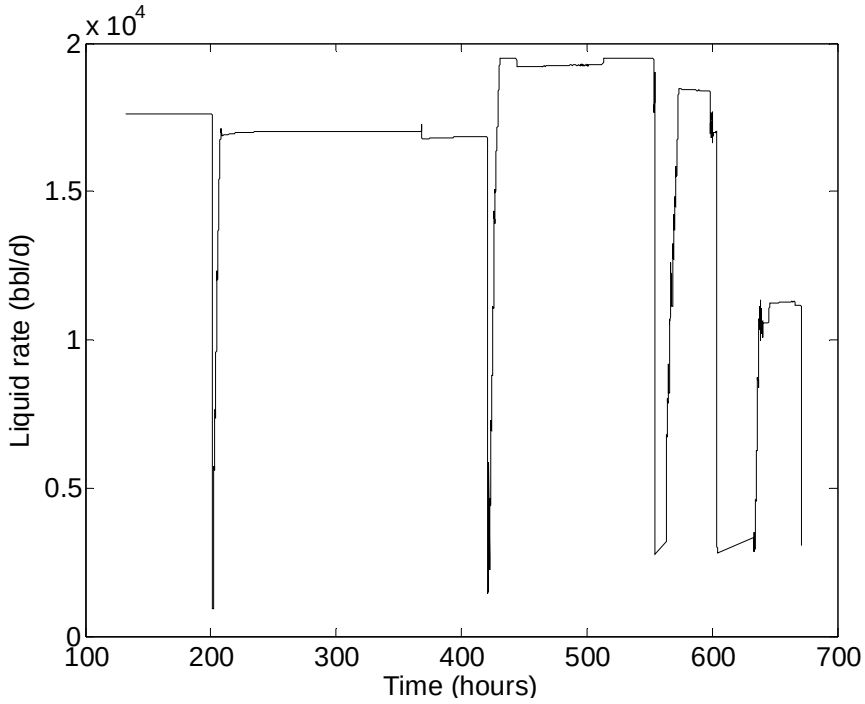

Figure 2-2: Liquid rate data processed with existing wavelet method

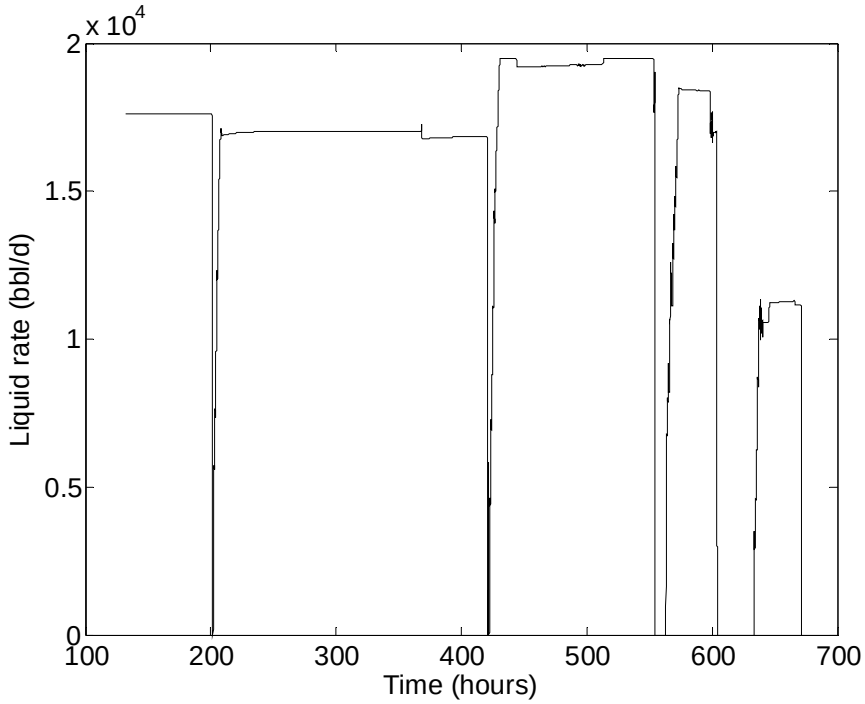

Figure 2-3: Liquid rate data processed with modified wavelet method

Chapter 3

3. Issues with break point identification

Accurate and reliable identification of transient break points in a permanent downhole gauge data is crucial for subsequent interpretation. The spline wavelet algorithm developed for finding break points is based on the definition of wavelet modulus maxima, which is basically a high or low point in a detailed signal. A decomposition level and a slope threshold for the detail signal have to be chosen for identifying the transient starting point. The slope threshold is defined as the pressure difference between two adjacent data points for the first level of detail signal. After deciding the parameters, possible locations for starting positions of new transients is given by the wavelet modulus maxima whose detail signals are higher than the slope threshold at a certain decomposition level.

Figure 3-1 and Figure 3-2 show an application of this algorithm (Athichanagorn et al., 2002) on a real field pressure data. It was found that for a certain low slope threshold setting of 30 psi, many smaller transients were identified along with the bigger transients and for a higher slope threshold setting of 40 psi, a few bigger transients were missed. Figure 3-3 shows a magnified plot of a section of the data which shows that a bigger transient has been missed and some smaller transient contained within that bigger transient has been identified.

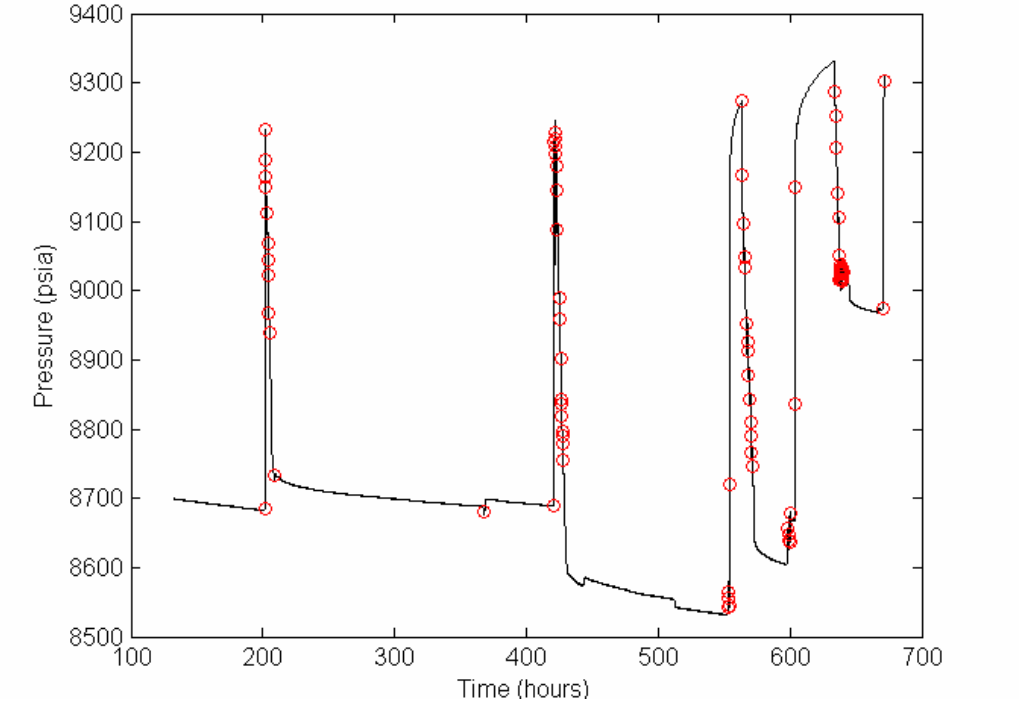

Figure 3-1: Transients identified by spline wavelet method for slope threshold of 30 psi

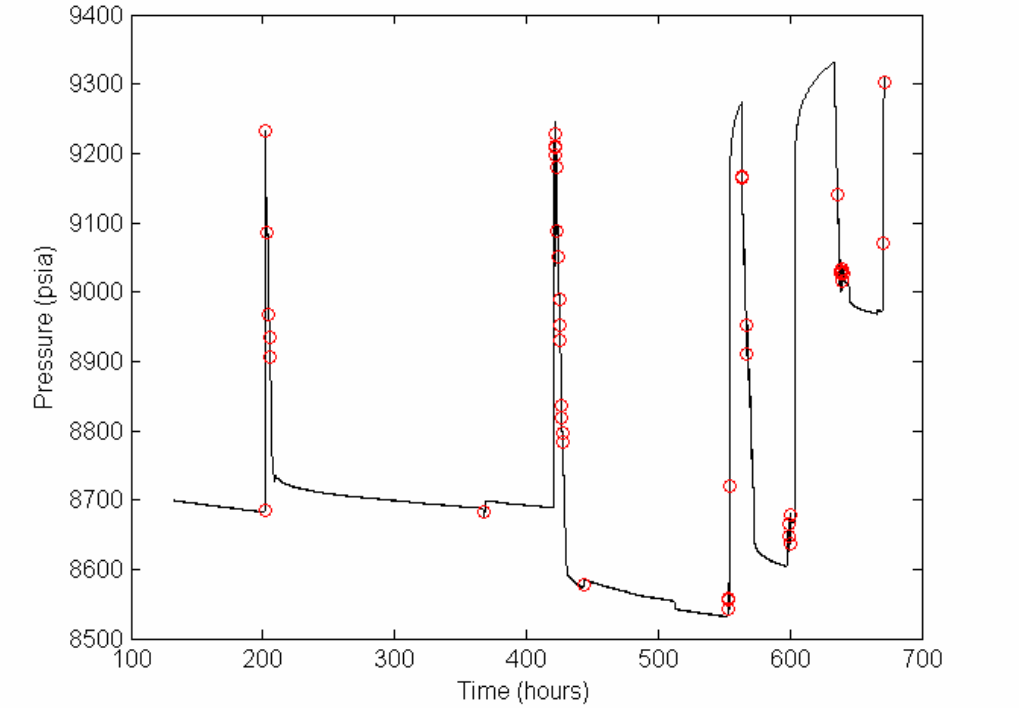

Figure 3-2: Transients identified by spline wavelet method for slope threshold of 40 psi

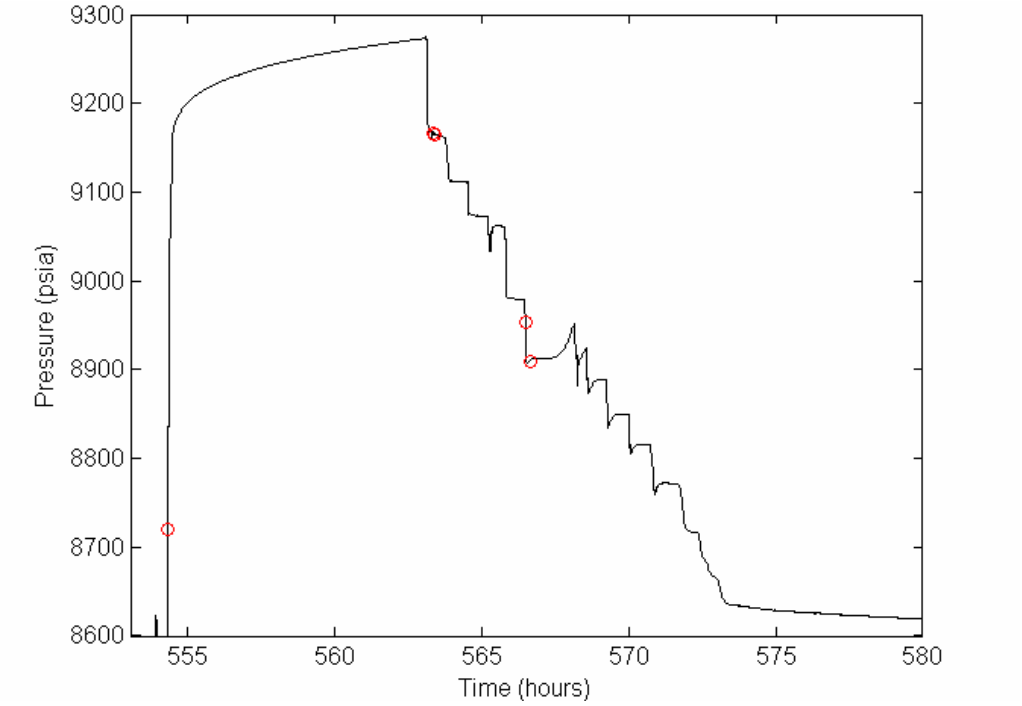

Figure 3-3: Close up view with bigger transient missed and smaller transients detected

The most key parameter in picking a transient is the slope threshold. Further investigation was carried out in the process of calculation of slope threshold. After deciding on the slope threshold for finding the transient points at the first level of detail decomposition, it is required to predict the corresponding slope threshold at higher level of detail decompositions where the wavelet modulus maxima above those thresholds are located as break points. The equation used for predicting the slope threshold at decomposition level j using *à trous* algorithm and spline wavelets is as follows:

$$(threshold)_j = (threshold)_1 * \left(2\sqrt{2}\right)^{j-1} \qquad (3\text{-}1)$$

This equation was tested against various signals to test its predictive capabilities. The results for various types of signal application are shown in Figure 3-4, Figure 3-5 and Figure 3-6. Figure 3-4 depicts a straight line signal with its slope constant and it can be seen that the slope threshold from actual wavelet decomposition and as predicted by the

equation for various levels of decomposition are the same. Figure 3-5 shows a synthetic drawdown pressure signal. From the comparison of slope threshold values at different levels of decomposition, significant differences in slope threshold can be seen at decomposition levels 5 and 6 between actual wavelet decomposition and that predicted from the equation. Further testing was done on a real buildup pressure signal shown in Figure 3-6. In this case the difference between the actual wavelet decomposition and that predicted from the equation is much more pronounced than the previous two cases. The differences grow significantly at wavelet decomposition levels of 5 and 6.

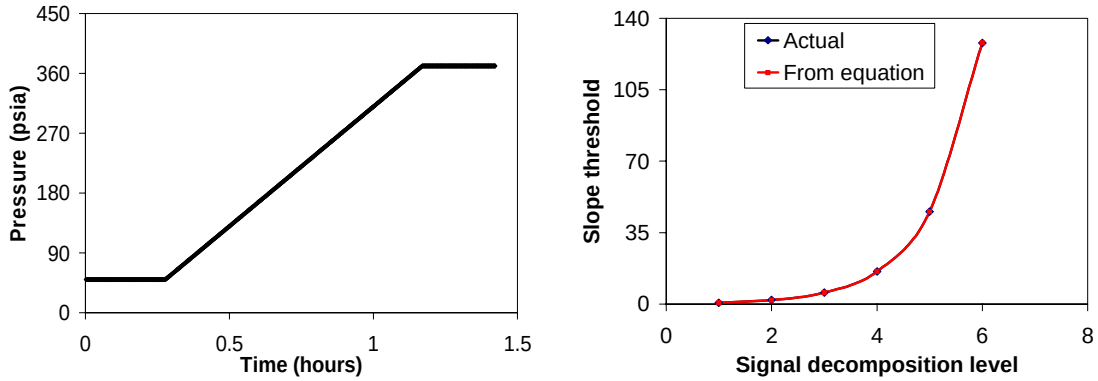

Figure 3-4: Straight line signal and its slope threshold at different levels of decomposition

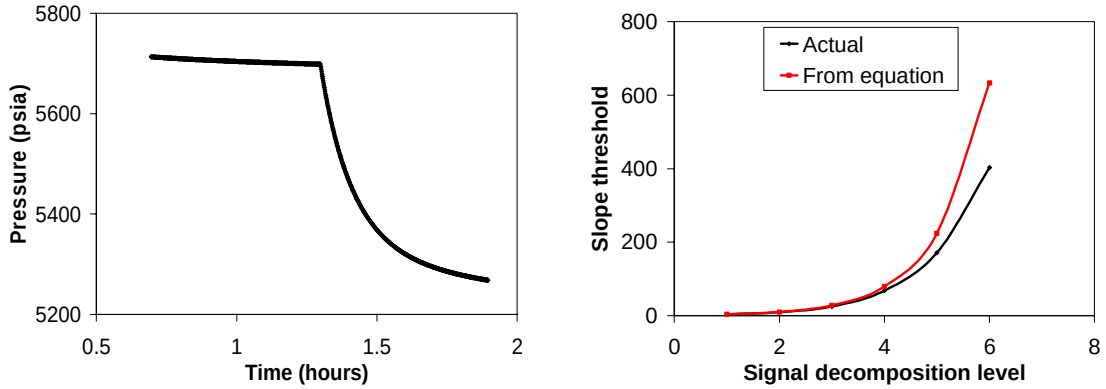

Figure 3-5: Drawdown signal and its slope threshold at different levels of decomposition

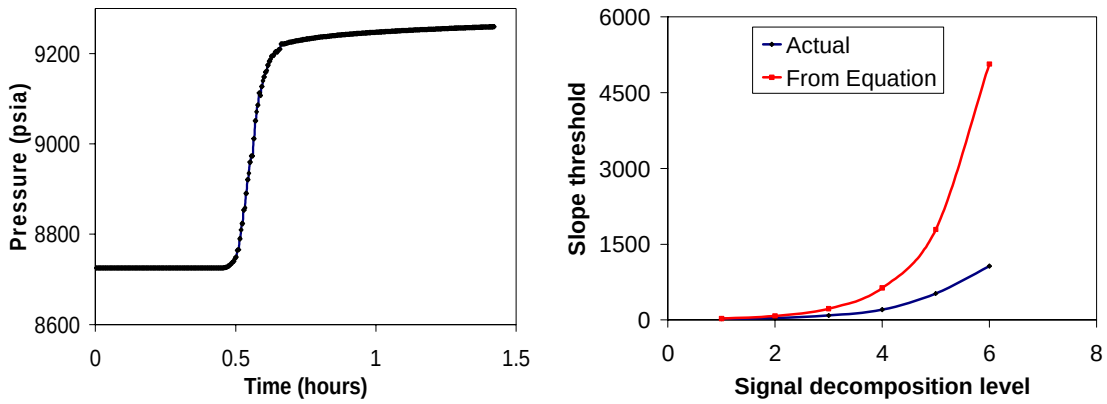

Figure 3-6: Build up signal and its slope threshold at different levels of decomposition

To define the wavelet modulus maxima of the detailed signal in order to obtain the location of break points, normally levels 5 or 6 are used as the effect of noise on the detail signal is suppressed at these levels. However, from Figure 3-5 and Figure 3-6 we can see how difficult it is to predict the slope threshold at various levels of decomposition as it seems to be governed by the shape or regularity of the signal at the transient break point as well as by the type of wavelet being used.

We can see from this investigation that the slope threshold alone is probably not a very effective criterion in discriminating between big and small transients in the permanent downhole gauge data. Also in using the spline wavelet approach it is very difficult to predict the slope threshold at higher levels of decomposition from the slope threshold at the first level of decomposition.

Chapter 4

4. Alternate algorithms for break point identification

4.1. Stationary wavelet transform using Haar wavelet

Wavelets have only recently been applied to a variety of petroleum engineering problems like upscaling of rock and reservoir properties and in the solution of two-phase flow problems. Wavelet methods for processing permanent downhole gauges were first described by Kikani et al. (1998) and Athichanagorn et al. (2002). A compactly supported nonorthogonal wavelet, namely the spline wavelet, was used by them for data processing and transient identification. Chapter 2 discussed the limitations of the spline wavelet based algorithm in detecting the bigger transients only while screening out the smaller and false break points.

As an alternative to the spline wavelet based algorithm, application of Haar wavelet is explored in this section. Haar wavelets are the most compactly supported wavelets of all the orthogonal family of wavelets. Another property which makes them attractive for identifying break points is that they are discontinuous and thus will tend to locate the break points more accurately as compared to other less compacted regular and orthogonal wavelets. The Haar wavelet function ψ, also interpreted as the impulse response of a high pass filter is defined as:

$$\psi(t) = \begin{cases} 1 & -1/2 \leq t < 0 \\ -1 & 0 \leq t < 1/2 \\ 0 & \textit{otherwise} \end{cases} \qquad (4\text{-}1)$$

The Haar wavelet function is also known as mother wavelet from which a family of wavelets can be constructed through translation and dilation. Also to recover the function from the wavelet transform a complementary function called scaling function φ is defined

which is interpreted as the impulse response of a low pass filter. Figure 4-1 illustrates the Haar wavelet function ψ and scaling function φ.

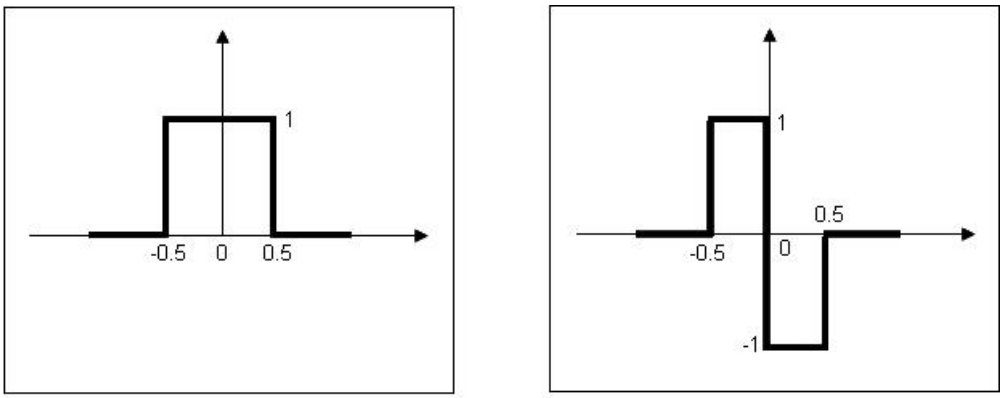

Figure 4-1: Haar wavelet function ψ and scaling function φ

The concept of multiresolution analysis (MRA), which establishes the relationship between wavelet and its complementary transform, is used here. This concept provides a framework to approximate a signal at different resolutions and determine the difference of information between two successive approximations. Hence at each resolution, the large scale features (averaged values) are kept in the approximation signal and the small scale features (the differences) are kept in the detail signal. For the purpose of identifying break points in the signal we make use of the coefficients in the detail signal. Further detailed discussion on wavelets and multiresolution analysis as applied to permanent downhole gauge data can be found in papers by Kikani et al. (1998) and Athichanagorn et al. (2002).

The classical algorithm for discrete wavelet transform (DWT) for computation of the approximation and detail signal from the original discrete signal and then reconstruction of the signal back from approximation and detail signal, utilizing their multiscale relationship through the wavelet and scaling functions, is known as pyramidal algorithm. The basic computation steps in this algorithm for signal decomposition are to convolve the approximation signal with the high and low pass filters and then downsample

(decimate) the detail and approximation signal by choosing the even indexed elements at each level of decomposition. This type of algorithm is used for data compression, upscaling and approximation problems where the number of data is to be reduced, but this algorithm lacks a property known as translation invariance which means that this algorithm cannot evaluate the approximation and detail signals at the same location across scales. Hence we need to have a different algorithm for signal decomposition and reconstruction which is translation invariant.

To develop the concept of an efficient translation invariant algorithm, we will start with the concept of translation invariance. Let $f_\tau(t) = f(t-\tau)$ be a translation of $f(t)$ by τ then the condition of translation invariance is satisfied if $Wf_\tau(u,a^j) = Wf_\tau(u-\tau,a^j)$, where W is an operator for the wavelet transform. Continuous wavelet transforms preserve the translation invariant property but in the discrete wavelet transform (DWT) a restriction on the sampling interval is imposed to preserve this property.

One of the approaches to correct these misalignments between features in the signal and features in a basis is to apply a shift in the signals so that their features change position. In the classical algorithm for discrete wavelet transform (DWT), this shift is applied by choosing every even indexed element during the decimation process, by defining a parameter ε which takes the value '0' for even and '1' for odd indexed elements. But translation invariance for the entire signal can not be guaranteed by one shift as there may be several discontinuities in the signal and the best shift for one discontinuity in the signal may also be the worse shift for another discontinuity. Figure 4-3 shows the detail signal at four different levels of decomposition of a test pressure data shown in Figure 4-2. It can be seen clearly that the relative proportion of the maxima corresponding to each peak in different levels of decomposition is not preserved. This makes the prediction of slope threshold at various levels of decomposition impossible or unreliable and thus this cannot be used for detecting break points.

Now the ε-decimation can also be carried out by choosing the odd indexed elements of the detail signal instead of the even indexed elements, i.e. setting ε = 1. This choice

concerns every step of the decomposition process, so at every level we choose odd or even indexed elements. So if we perform all the different possible decompositions of the original signal, we will have 2^J different decompositions, for a given maximum level J. The idea of the stationary wavelet transform is to average all the slightly different ε-decimated DWT at a given level and use it instead of using a DWT obtained from a single shift operation. So for a signal f (t: $0 \le t \le n$) let us define S_h as the circulant shift by h, $(S_h x)_t = x_{(t+h) \bmod n}$. then an stationary wavelet transform (SWT), considering a range H of shifts, will be given by (Coifman, Donoho 1995),

$$\overline{W}\left(x;(S_h)_{h\in H}\right) = Ave_{h\in H} S_{-h}\left(W\left(S_h(x)\right)\right) \qquad (4\text{-}2)$$

Figure 4-4 shows the detail signals at various levels obtained for a test signal. It can be seen that the relative proportion of the maxima corresponding to each peak in different levels of decomposition is preserved. This makes SWT an obvious choice over the classical DWT for transient break point identification.

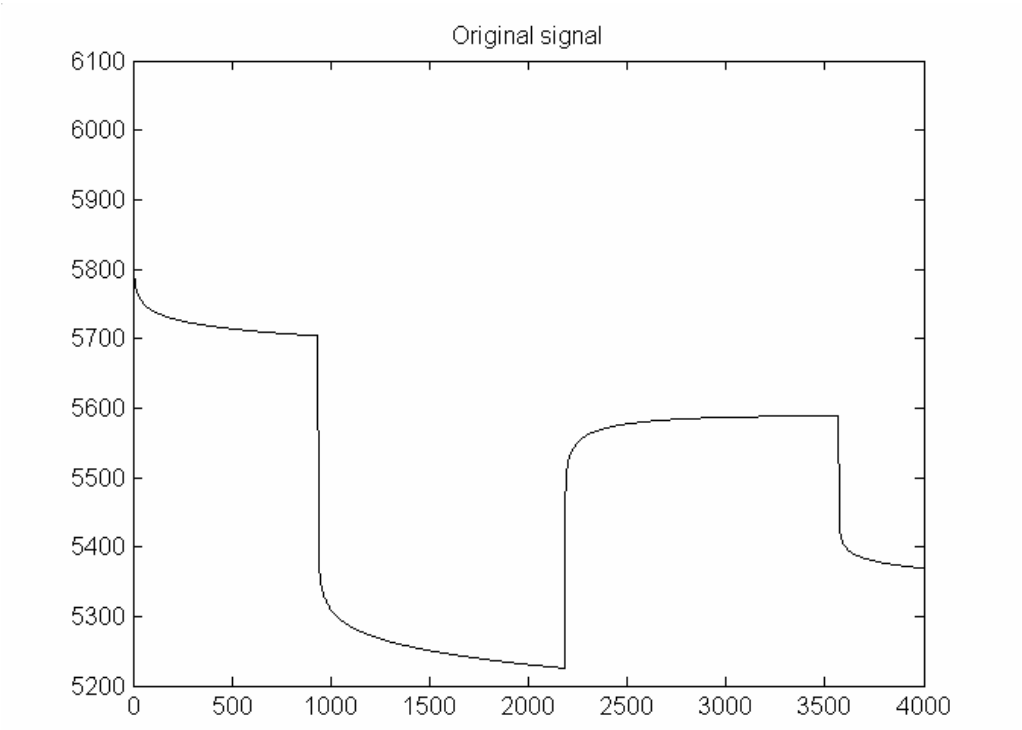

Figure 4-2: Test signal

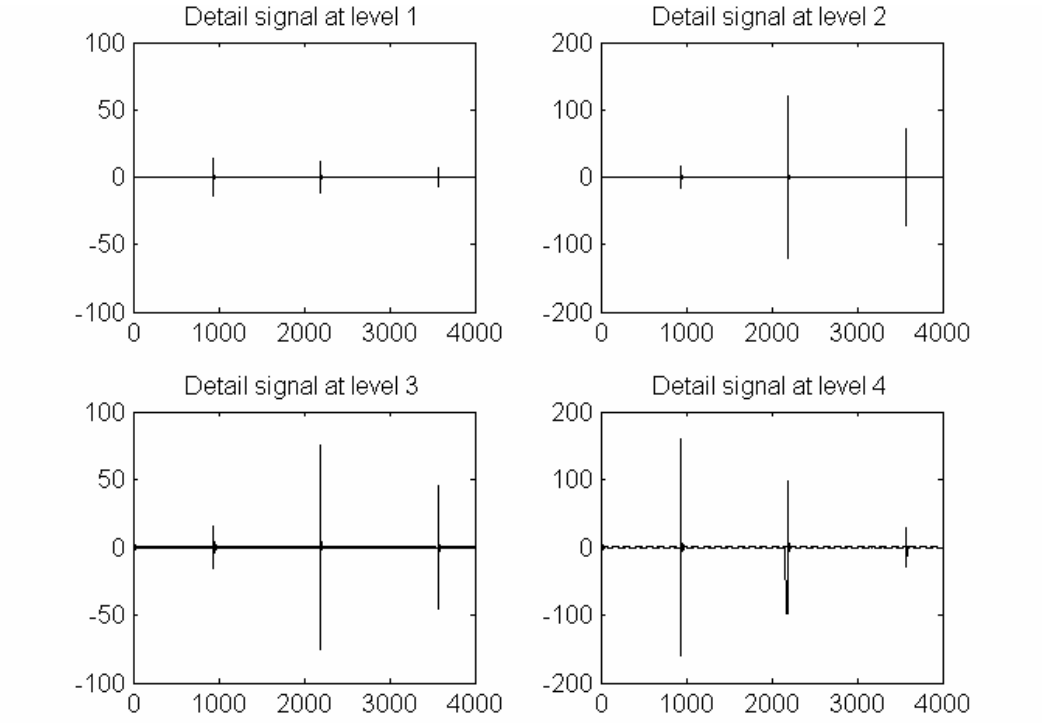

Figure 4-3: Detail signal at various levels by classical wavelet decomposition

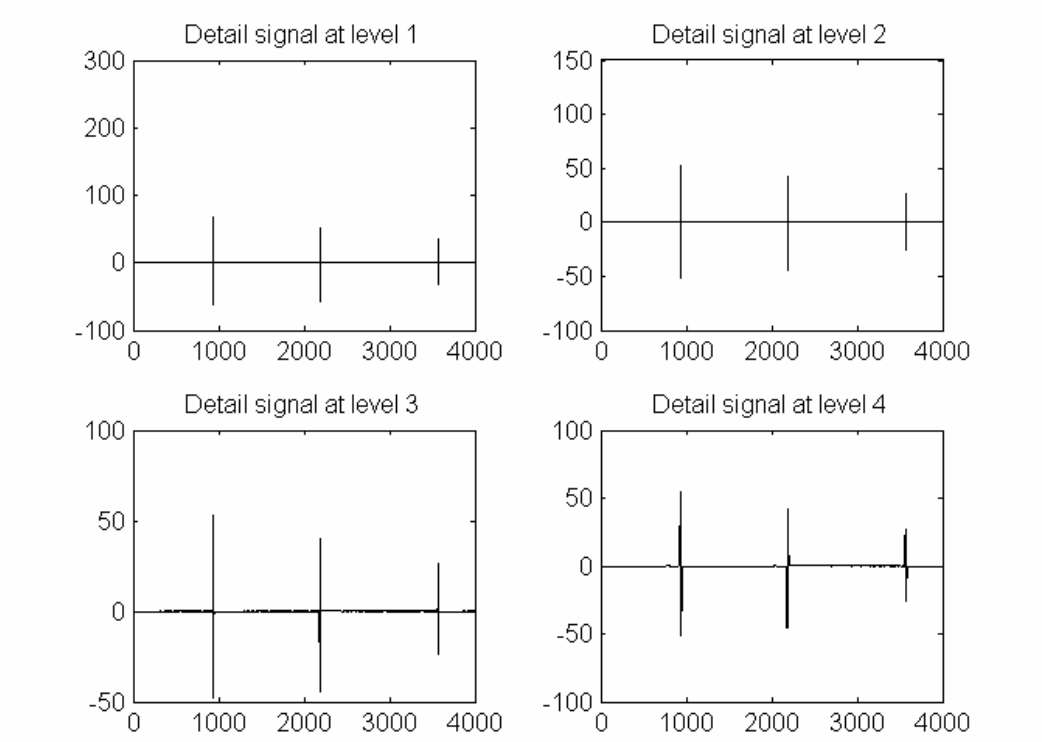

Figure 4-4: Detail signal at various levels by stationary wavelet decomposition

4.1.1. Application on field pressure data

The concept of wavelet modulus maxima i.e. to look for points whose absolute value is maximum within its neighborhood was used for detecting break points. Since it was observed in the previous chapter that defining a slope threshold at the first level of detail decomposition and predicting the corresponding slope at higher levels like 5 or 6 of detail signal is difficult, hence a different way of using the wavelet modulus maxima was taken. After choosing a level of decomposition, a fraction of the largest wavelet modulus maximum of all the maxima in the signal was taken as the threshold above which the break points were to be detected. Figure 4-5 shows an application of the Haar wavelet method on a field data set of pressure signal. In all, a total of 42 transients were identified. It was observed that out of a total of nine significant transients, two were missed whereas several other smaller transients were identified along with the other significant transients. This further emphasizes that slope threshold alone is not a sufficient criterion for differentiating between smaller and bigger transients.

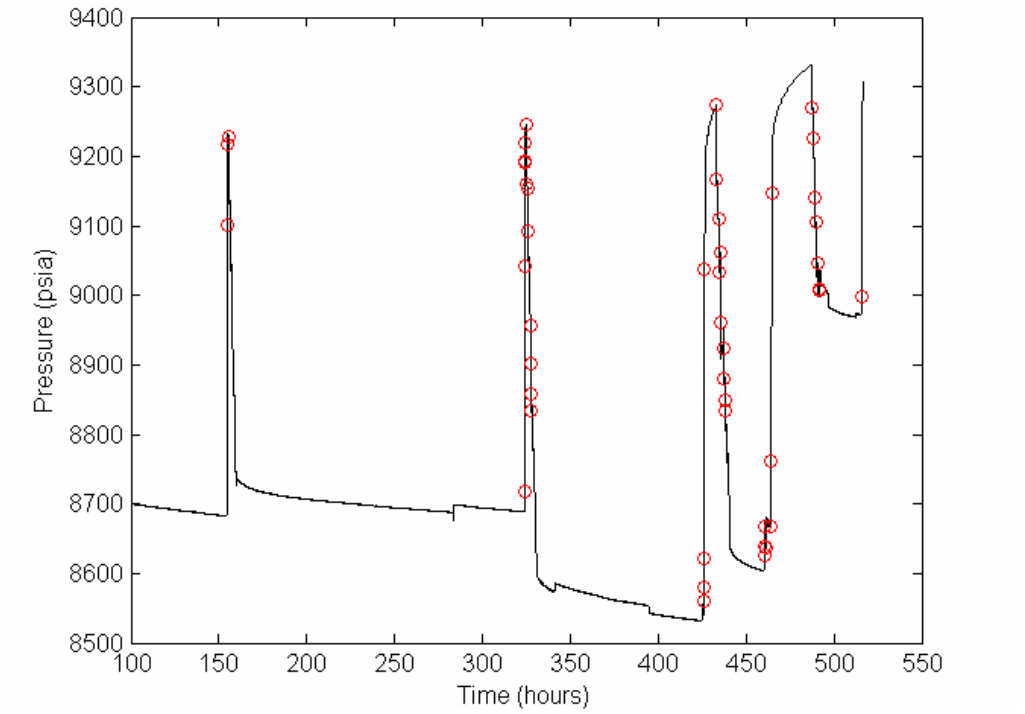

Figure 4-5: Transients identified by stationary wavelet method

4.2. Savitzky-Golay FIR smoothing filters

An FIR (finite impulse response) filter is a type of digital filter that replaces each data value f_i by a linear combination g_i of itself and some number of nearby neighbors,

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n} \tag{4-3}$$

Here n_L is the number of points used "to the left" of a data point i, i.e., earlier than it, while n_R is the number used "to the right", i.e., later. Also c_n are known as the filter coefficients. This process is also called moving window averaging and corresponds to Equation 4.3.

The Savitzky-Golay FIR smoothing filters, also known as polynomial smoothing or least squares smoothing filters, are generalizations of the FIR averager filter that can better preserve the high frequency content of the desired signal. They are filters that optimally fit a set of data points to polynomials of different degrees.

The idea of the Savitzky-Golay filtering is to find filter coefficients c_n that preserve higher moments. Equivalently, the idea is to approximate the underlying function within the moving window not by a constant (whose estimate is the average), but by a polynomial of higher order. For each point f_i, we least square fit a polynomial to all $n_L + n_R + 1$ points in the moving window, and then set g_i to be the value of that polynomial. No use of the value of the polynomial is made at any other point. Moving on to the next point f_{i+1}, a whole new set of least square fitting is done using a shifted window.

Carrying out a least square fitting in this manner as described above will be very time consuming. However, we can design particular sets of the filter coefficients c_n for which Equation 4.3 will automatically accomplish the process of polynomial least square fitting inside a moving window, as described in the following.

Let us consider a polynomial of degree M in i such that,

$$g_i = c_0 + c_1 i + ... + c_M i^M \tag{4-4}$$

which we want to fit to the values $f_{-nL},\ldots,f_{nR}$.

Considering a least square fit, we can define the performance index to be minimized as,

$$\Im = \sum_{i=-nL}^{nR} \left(f_i - \left(c_0 + c_1 i + ... + c_M i^M \right) \right)^2 \tag{4-5}$$

Now we can define the M-dimensional polynomial basis vectors which can be written in matrix form as:

$$S_{i,j} = i^j \tag{4-6}$$

where, $i = -n_L,\ldots,n_R,$ $j = 0,\ldots,M$

Thus in matrix vector notation we can write Equation 4.4 as,

$$\bar{g} = \bar{\bar{S}}\bar{c} \tag{4-7}$$

and the performance index as,

$$\Im = \left(\bar{f} - \bar{\bar{S}}\bar{c} \right)^T \left(\bar{f} - \bar{\bar{S}}\bar{c} \right) \tag{4-8}$$

From the minimization condition of setting the gradient of performance index with respect to $\boldsymbol{c}$'s to zero we get the normal equations:

$$\bar{\bar{S}}^T \bar{\bar{S}}\bar{c} = \bar{\bar{S}}^T \bar{f} \tag{4-9}$$

with optimal solution:

$$\bar{c} = \left(\bar{\bar{S}}^T \bar{\bar{S}} \right)^{-1} \bar{\bar{S}}^T \bar{f} \equiv \bar{\bar{G}}^T \bar{f} \tag{4-10}$$

where the $(n_L + n_R + 1) \times (M + 1)$ matrix $\boldsymbol{G}$ is defined as,

$$\overline{G} = \overline{S}\left(\overline{S}^T \overline{S}\right)^{-1} \tag{4-11}$$

Inserting the optimal coefficients $\boldsymbol{c}$ into Equation 4.7, we can find the smoothed values:

$$\overline{g} = \overline{S}\overline{G}^T \overline{f} \equiv \overline{B}\overline{f} \tag{4-12}$$

where the $(n_L + n_R + 1) \times (n_L + n_R + 1)$ matrix B is defined as,

$$\overline{B} = \overline{S}\overline{G}^T = \overline{S}\left(\overline{S}^T \overline{S}\right)^{-1} \overline{S}^T \tag{4-13}$$

The $(n_L + n_R + 1)$ columns or rows of matrix $\boldsymbol{B}$, as $\boldsymbol{B}$ is symmetric, are called the Savitzky-Golay (S-G) smoothing filters of length $(n_L + n_R + 1)$ and polynomial order M. Of all the columns of $\boldsymbol{B}$, the middle one, $\boldsymbol{b}_0$, is more important for our purpose because it smoothes the value f_0, which is symmetrically placed with respect to the other samples in $\boldsymbol{f}$.

Also it can be noticed that on setting $i = 0$, the middle smoothed value g_0 is equal to the polynomial coefficient c_0. Similarly the coefficient c_1 represents the first derivative of f_0 at $i = 0$ and so on. From Equation 4.10 we can see that these coefficients can be calculated from dot product of matrix $\boldsymbol{G}$ and matrix $\boldsymbol{f}$. Matrix G is therefore also interpreted as the differentiation filters.

Now it can be seen from these equations that in order to compute the polynomial coefficients beforehand, data window width and order of polynomial are required to be defined and then just one row of inverse matrix calculation is required (Equation 4.10). This can be easily carried out using any efficient matrix inversion technique like LU decomposition.

4.2.1. Determining window size and order of polynomial

As described above, filters can be generated for calculating the polynomial coefficients and smooth a given noisy data set and further calculate the first, second, third and fourth derivatives with the smoothed data. In practical application, the choice of order of polynomial depends on the order of derivatives we want to calculate for analysis purpose. Here, we make use of derivatives up to the fourth order; hence a polynomial of order four or higher would suffice.

Now, the size of the window that will smooth the noisy data sufficiently to have smooth derivative calculations has to be determined. A real noisy data set gathered at an interval of 5 seconds was used for determining the window size. Figure 4-6 shows the resulting derivative plots for a window size of 21. It can be clearly from the figure that although the first derivative is smooth, the higher derivatives i.e. second, third and fourth ones, which are increasingly sensitive to noise in the data, are still noisy and it would be difficult to carry out any analysis. Figure 4-7 shows the resulting derivative plots for a window size of 51. It can be seen from the figure that all the derivatives are sufficiently smooth and analysis is possible with them. Hence it was concluded that a window size of 51 or more can be used for smoothing the data and calculating smooth higher order derivatives using S-G filters.

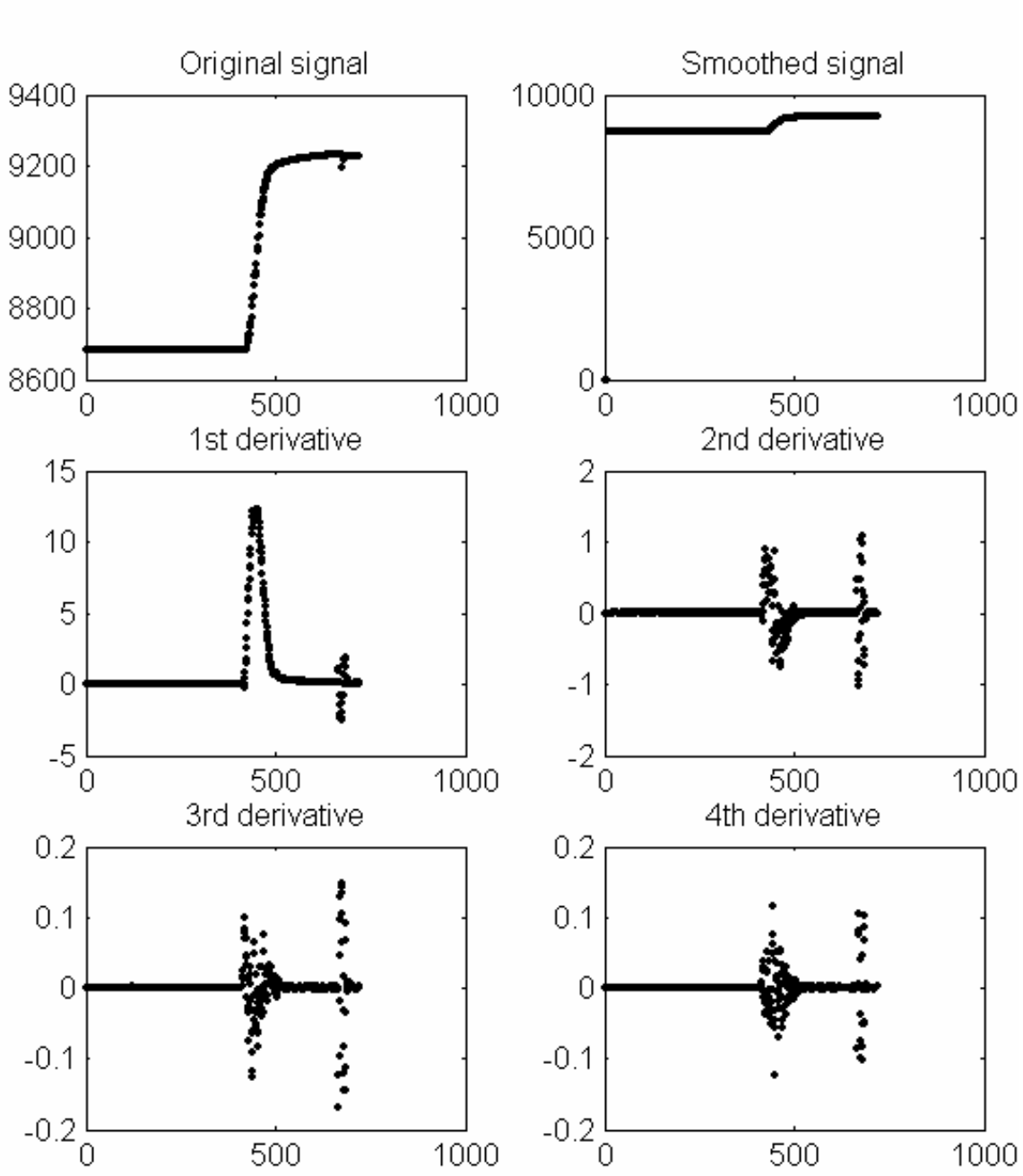

Figure 4-6: S-G smoothing and derivative calculation for window size of 21

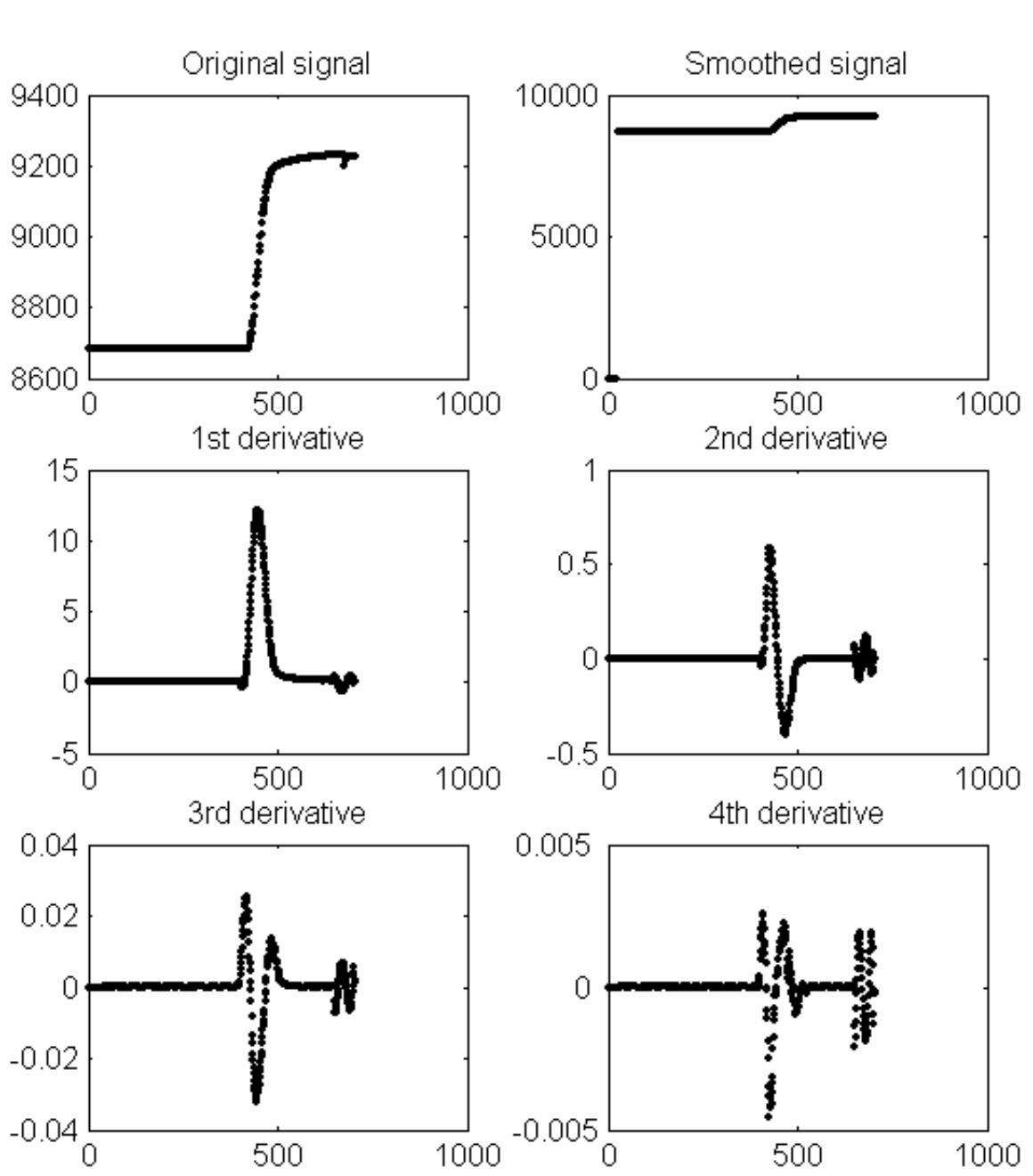

Figure 4-7: S-G smoothing and derivative calculation for window size of 51

4.2.2. S-G filters for break point identification

The derivatives of pressure or rate data exhibit a peak when there is a sudden change in the gradient of the data. Hence the peaks in first and other higher derivatives indicate the vicinity of the break point location. For an abrupt change in gradient of the data, the first and second derivative peaks will be closer to the actual break point location in the data, while for a smoothly changing data, the higher derivatives namely third and fourth will be closer to the break point location. The first step in identifying the break points was to find the peaks above a certain low threshold in the first derivative of pressure or rate data. After all the peaks in first derivatives are obtained, a backward search is carried out for identifying the corresponding peaks in the second, third and fourth derivatives of the data. All these derivative points corresponding to the break point location are mapped back on the original data. Then interpolation between these locations is carried out by checking the slopes between each pair of locations to be higher than a certain value which is determined from the data itself and is set as the average noise variance estimated from the data.

The area under a first derivative peak gives an indication about the net amount of gradient change in the rate or pressure data and the height of the peak gives an indication of the maximum gradient change in the data. The height of a peak and the area under it give an indication about the relative significance of a transient. Hence these two can become good criteria for screening out small transients identified in the data while retaining the bigger transients. An illustration of this is given in Figure 4-8, where the first transient is a bigger transient while the second one is a smaller transient and it can be effectively screened out with the peak height and area criteria. The area under the peak shown in Figure 4-9 is calculated numerically using the trapezoidal rule.

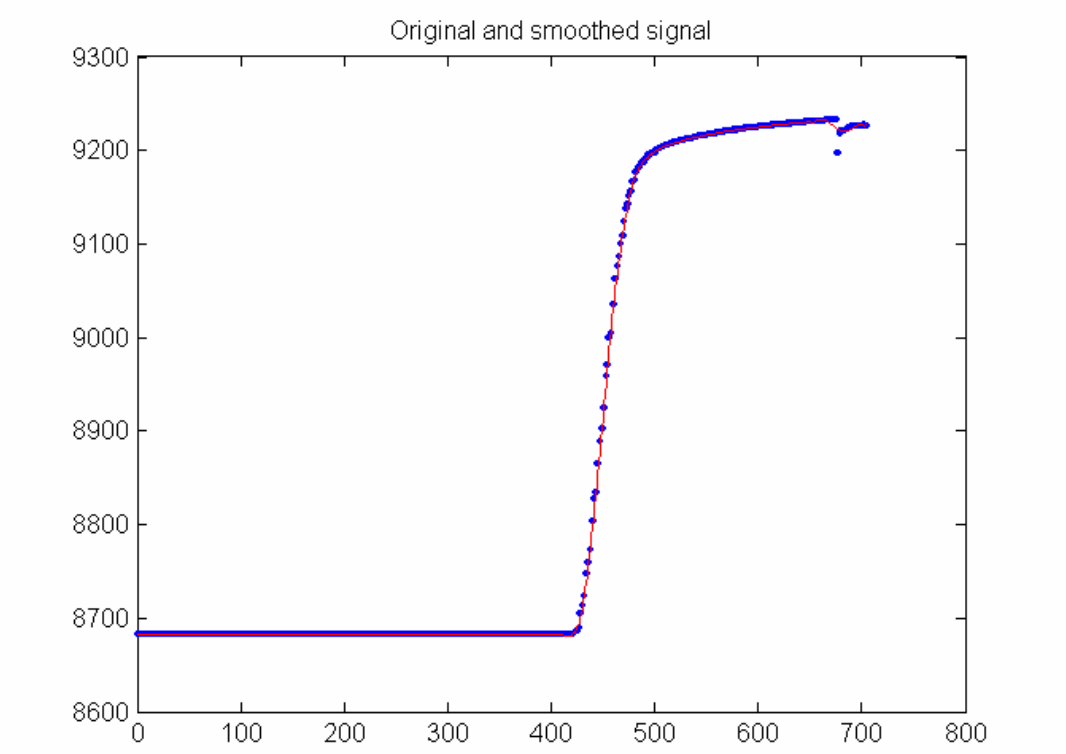

Figure 4-8: Original signal (blue) and smoothed signal (red) using S-G filters

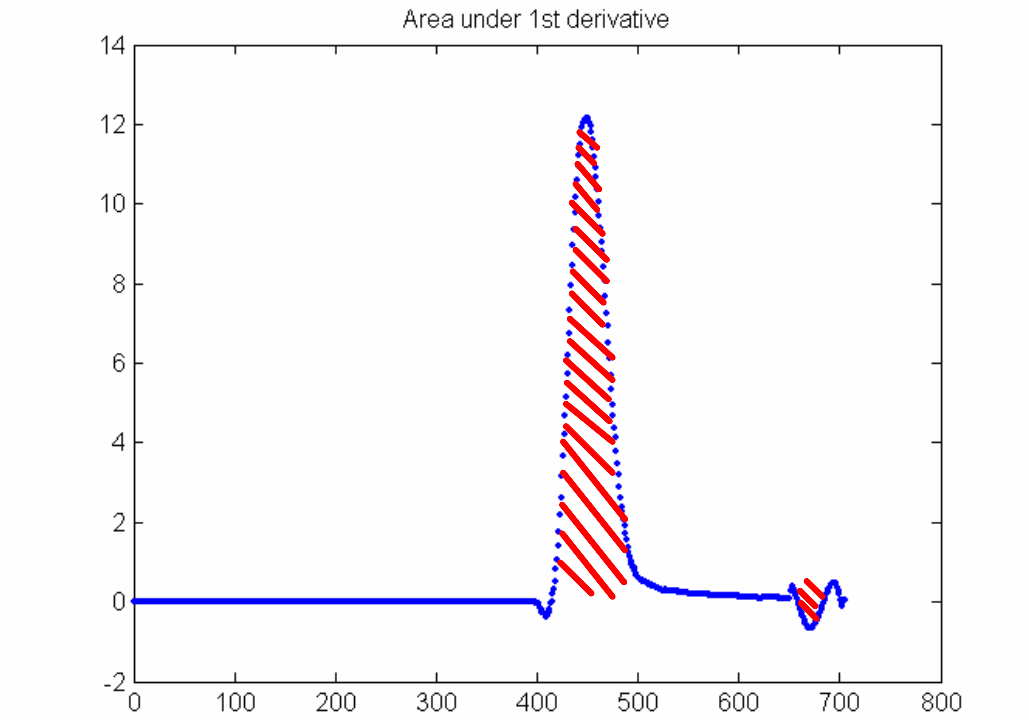

Figure 4-9: First derivative of smoothed signal and area calculation

To determine the effect of noise on the accuracy of peak identification of S-G filters, a test pressure data set was generated for an infinite-acting reservoir model with variable rate. Initially no noise was added into the data and the break points were estimated with the S-G filter method. True break point locations are at 1.2987, 3.0350, 4.9573 and 7.0408 hours. The corresponding break points estimated by the S-G filters method in the data without noise, Figure 4-10, are 1.2973, 3.0336, 4.9560 and 7.0395 hours. Next, the algorithm was applied after adding normally distributed random noise of zero mean and different variances. A noise histogram of one such distribution (0, 1.0) is shown in Figure 4-11. Figure 4-12 shows the break point location estimated by the S-G filters method on the pressure signal containing the above distribution of noise. The break point locations were found out to be the same as that estimated with the data without noise. This signifies that with the application of S-G filters having a window size sufficiently large to generate smooth derivatives, the locations of the derivative peaks of the signal are not affected by the noise in the signal. Hence this method can be applied directly on the signal without denoising it.

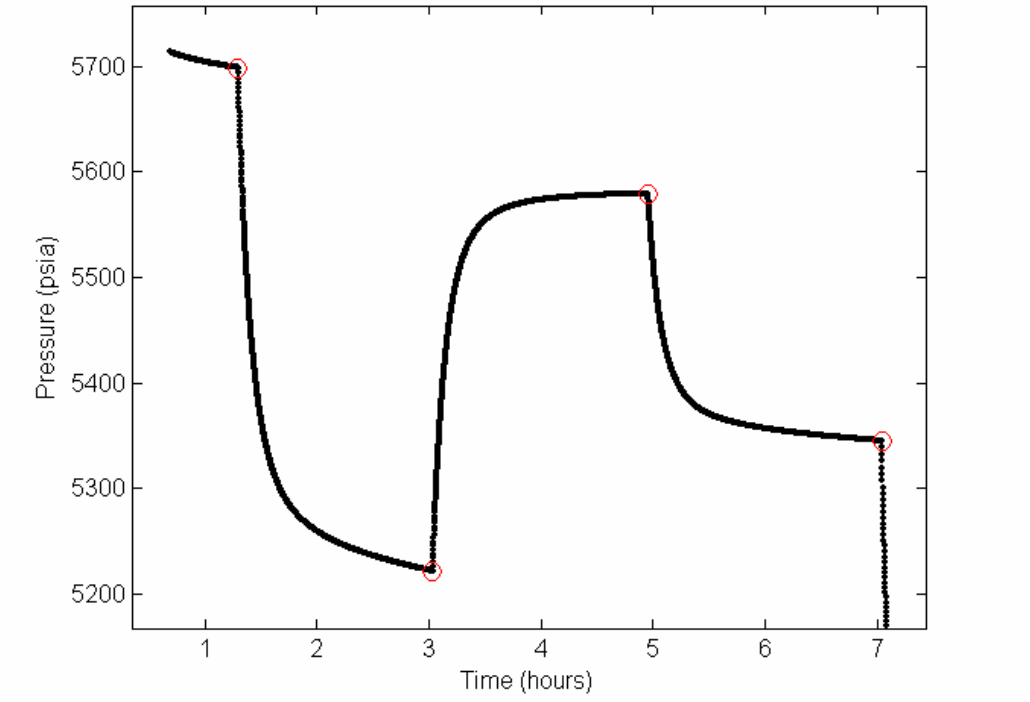

Figure 4-10: Test signal without any noise in the data

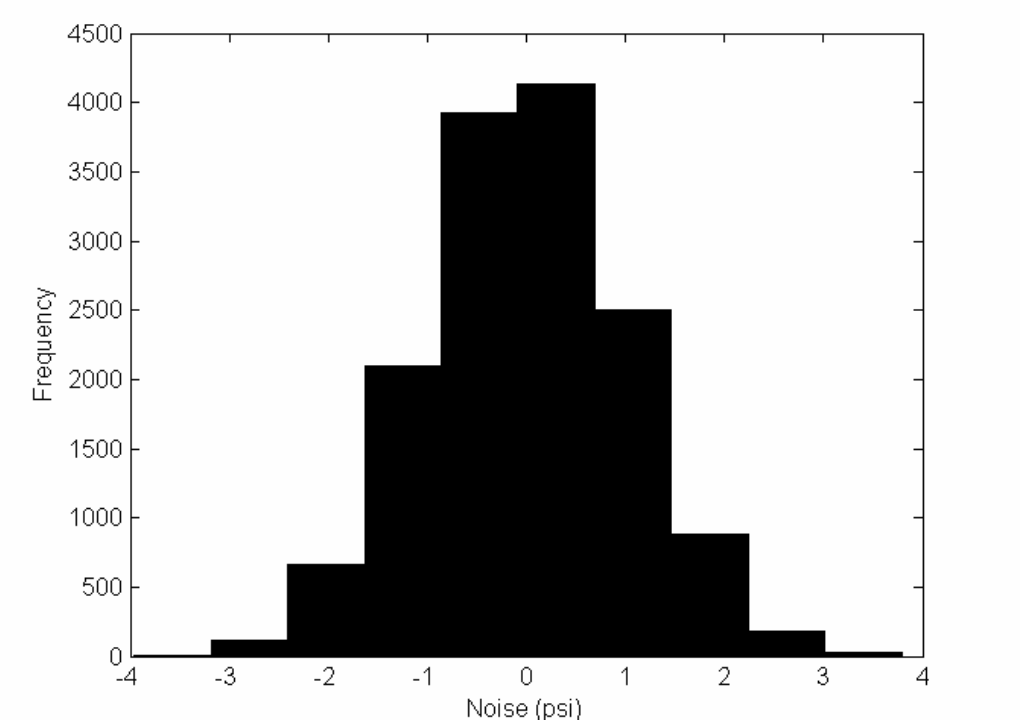

Figure 4-11: Noise histogram of the normally distributed noise added to the signal

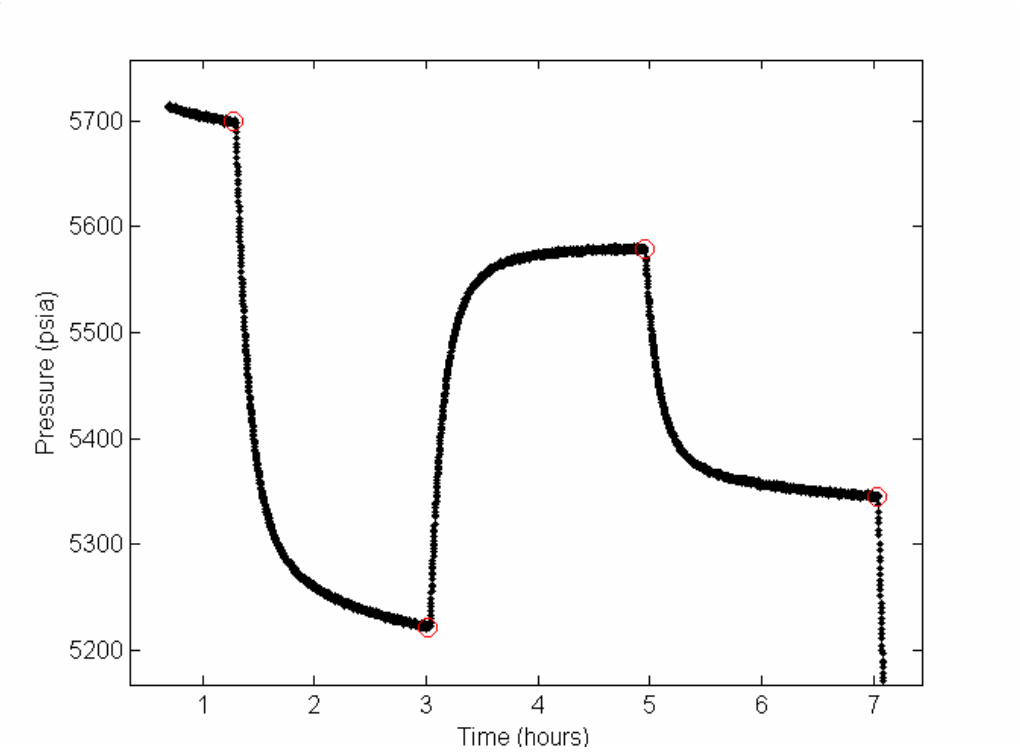

Figure 4-12: Signal with normally distributed noise added

4.2.3. Application and comparison with spline wavelet algorithm

The algorithm developed using S-G filters for break point identification was applied to real field pressure and rate data. The data was gathered at an interval of 5 seconds over a period of a month. It contains nine prominent buildups and drawdown. Figure 4-13 shows the break points identified in the pressure data from the application of S-G filters method. A total of 35 break points were identified in the data including all the bigger break points. Next, the spline wavelet based algorithm was applied on the same set of pressure data for different slope thresholds. Figure 4-14 shows the break points identified for a slope threshold of 30 psia and Figure 4-15 shows the break points identified with a slope threshold of 40 psia. For the lower slope threshold, 108 peaks were identified in the data whereas with a higher slope threshold of 40 psia, 42 break points were identified in the pressure data.

Further, liquid rate data which was measured simultaneously with the pressure data, was used for identifying break point by the two methods. Figure 4-16 shows the results of application of S-G filters method on rate data and Figure 4-17 and Figure 4-18 show the break points identified by the spline wavelet method for a slope threshold of 75 and 120 STB respectively. S-G filters method identified 34 break points, while the spline wavelet algorithm identified 102 and 77 break points respectively.

Comparing the break points identified by the two algorithms in pressure data, we see that S-G filters method identifies significantly fewer total break points than the spline wavelet method while identifying all the bigger break points. Also, for a higher slope threshold of 40 psia, the spline wavelet method identifies fewer smaller transients than a lower slope threshold but it misses some of the bigger break points. Similar observations can be made from the application of the two algorithms on liquid rate data.

Further, Figure 4-19 shows a zoomed plot of a section of the rate data with the break points, identified by the spline wavelet algorithm. It was observed that this section of the data contained noise with a higher variance than the universal noise threshold for denoising. It can be seen that several false transients were identified in this data section

by the spline wavelet algorithm; whereas the S-G filters method did not pick up the false break points as shown in Figure 4-20.

From the example applications it can be seen that a significant improvement over spline wavelet based algorithm in screening off small and false break points can be achieved by the application of S-G filters. Also the slope thresholds are determined directly from the first derivative in S-G filters method; hence it does not suffer from the disadvantage of difficulty in predicting slope threshold as in spline wavelet based method. Though this method can differentiate between smaller and bigger transients, but it still has problems differentiating between bigger transients and the smaller transients contained in the bigger transients as can be seen from the applications on the field data.

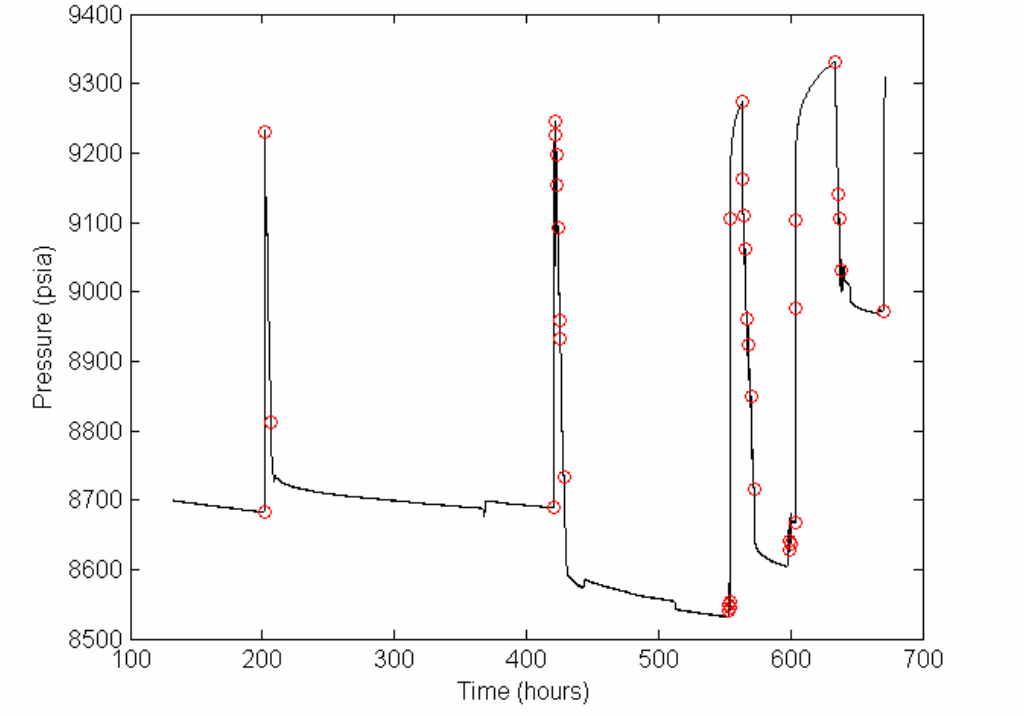

Figure 4-13: Transients identified by S-G filters method in pressure data

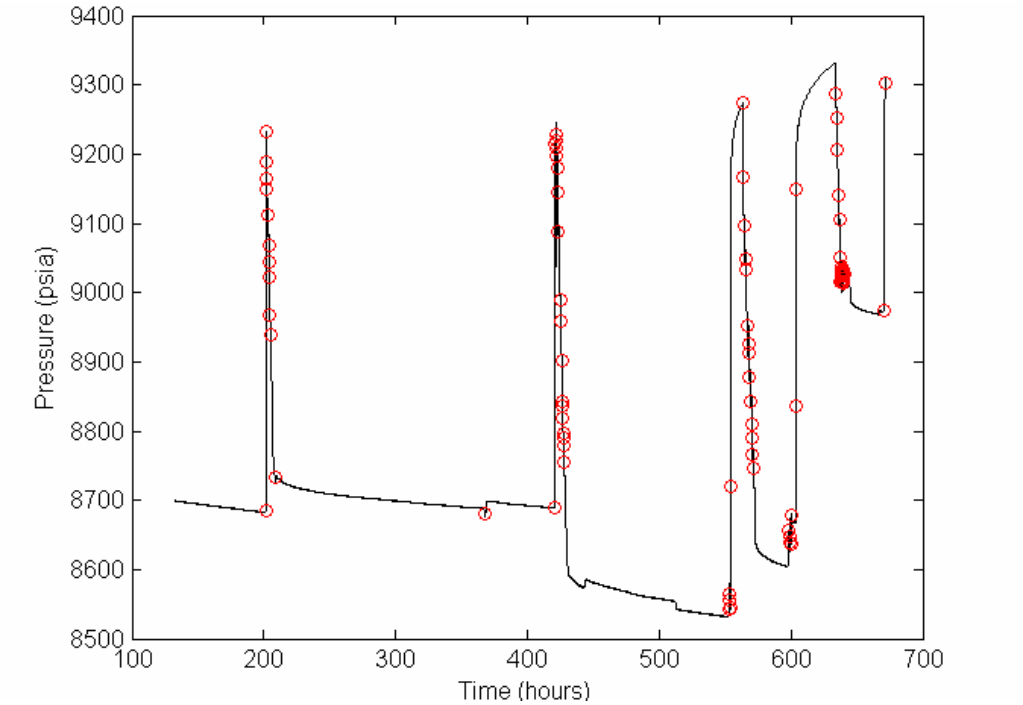

Figure 4-14: Transients identified by spline wavelet method for slope threshold of 30 psi

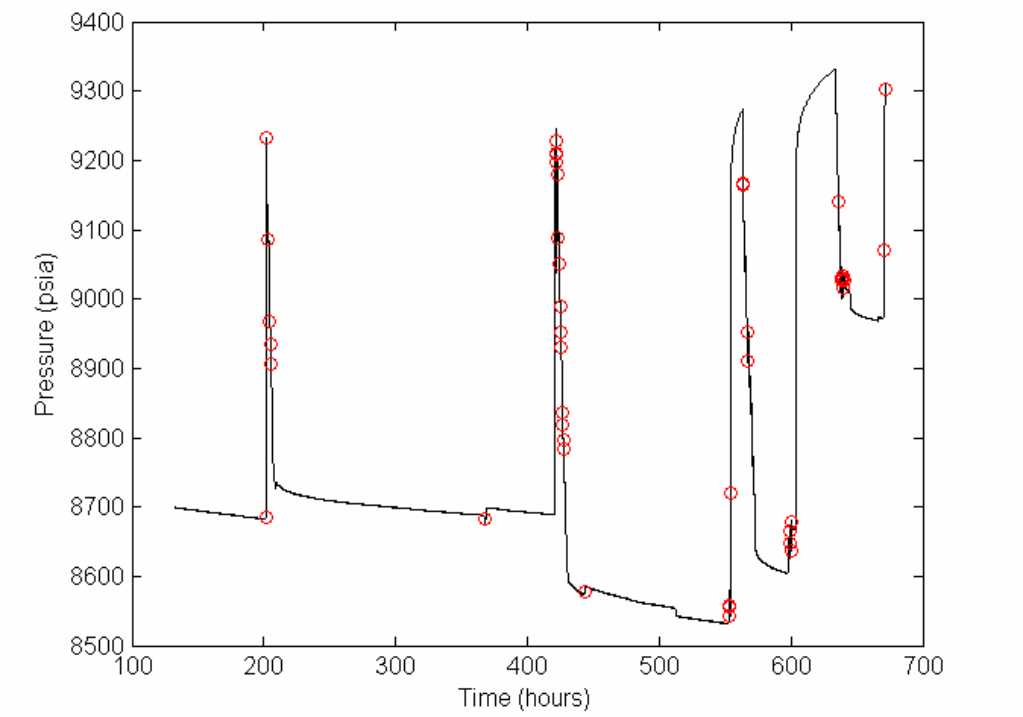

Figure 4-15: Transients identified by spline wavelet method for slope threshold of 40 psi

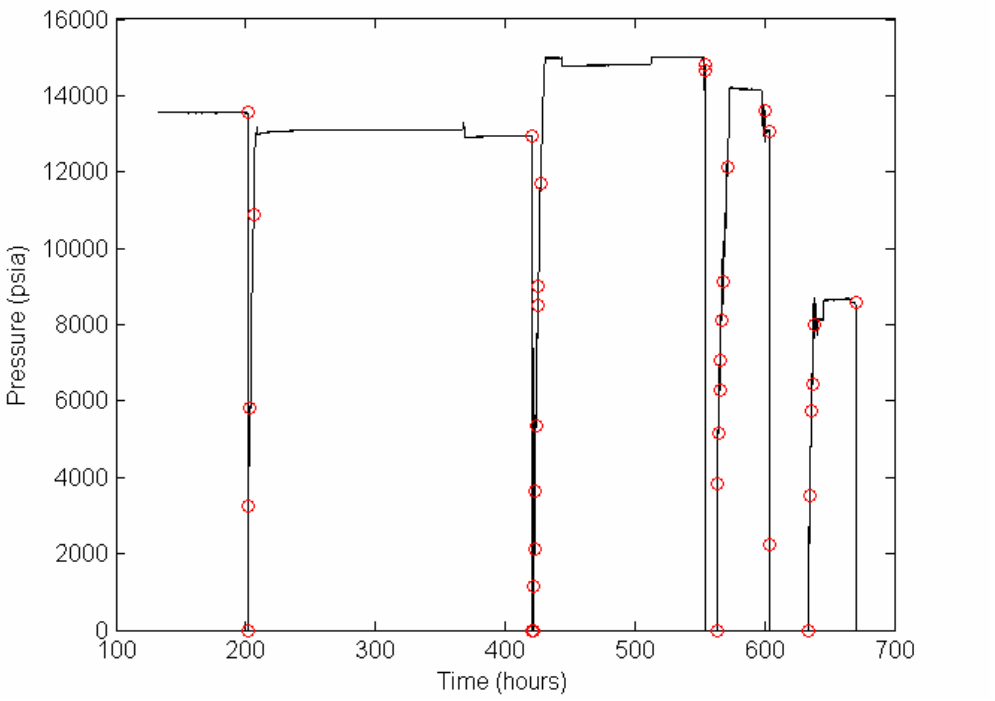

Figure 4-16: Transients identified by S-G filters method in downhole liquid rate data

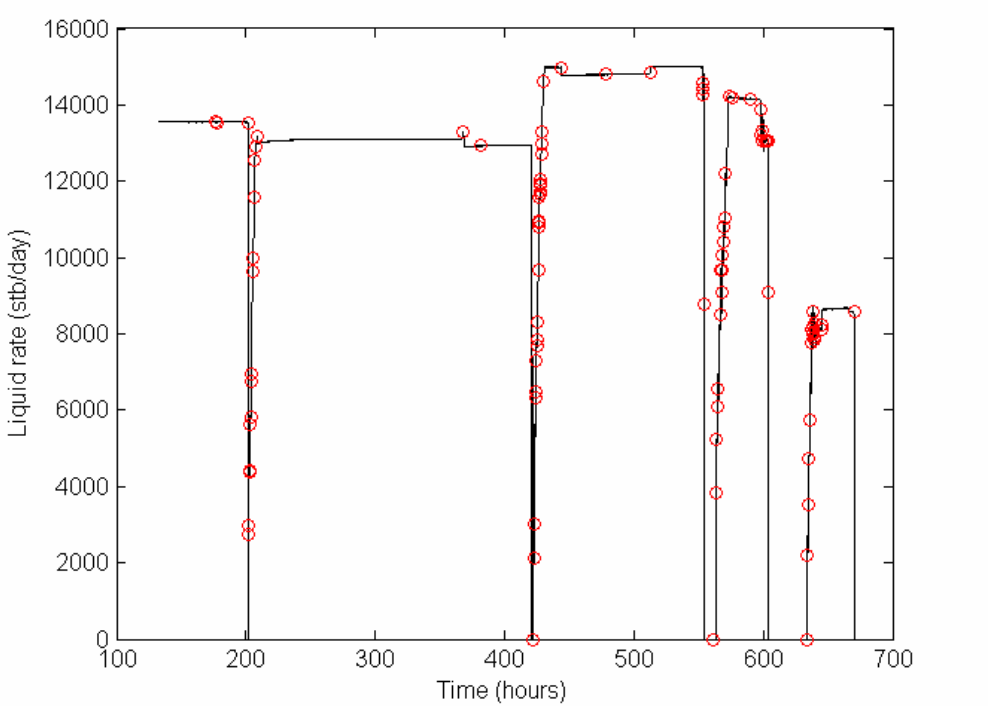

Figure 4-17: Transients identified by spline wavelet method for slope threshold of 75 stb/day

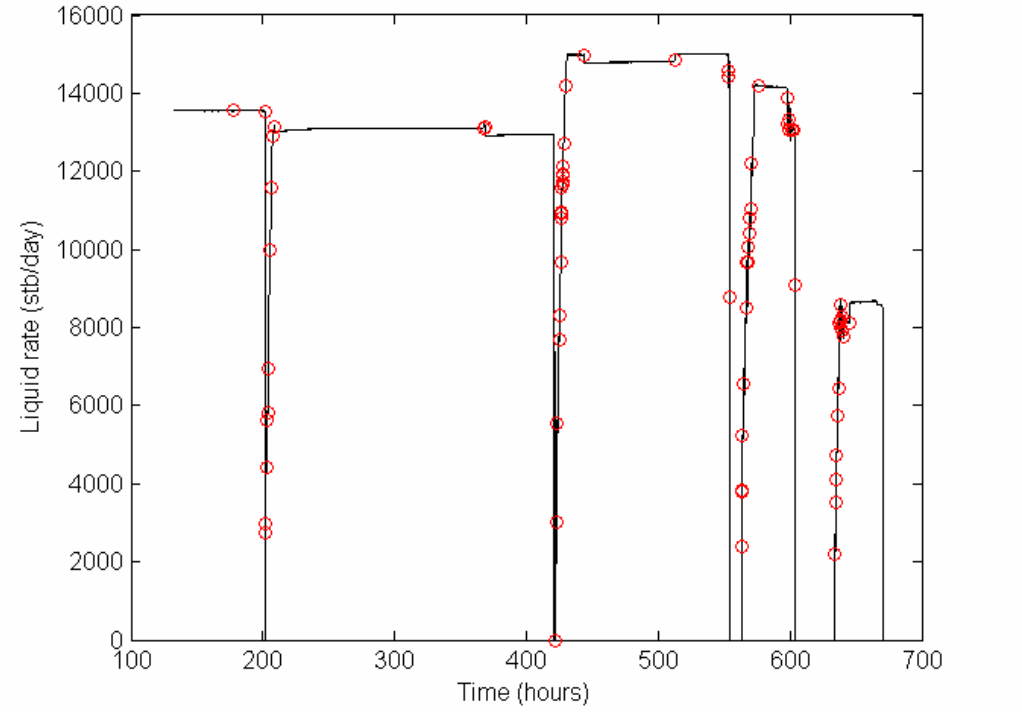

Figure 4-18: Transients identified by spline wavelet method for slope threshold of 120 stb/day

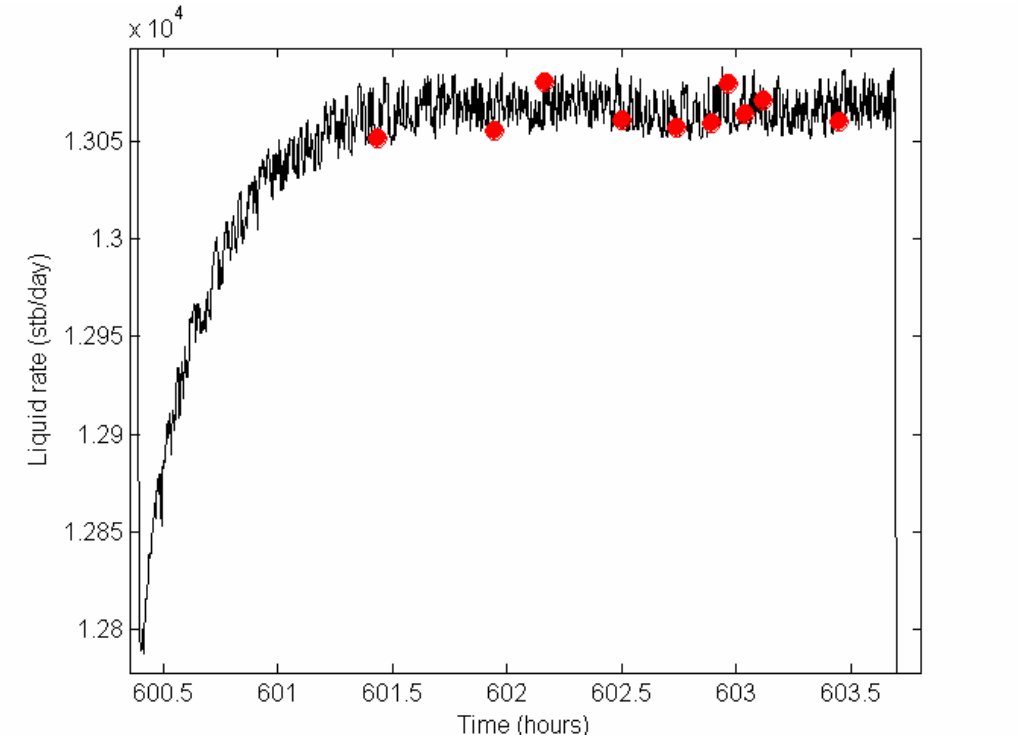

Figure 4-19: Close up view showing false transients picked by spline method

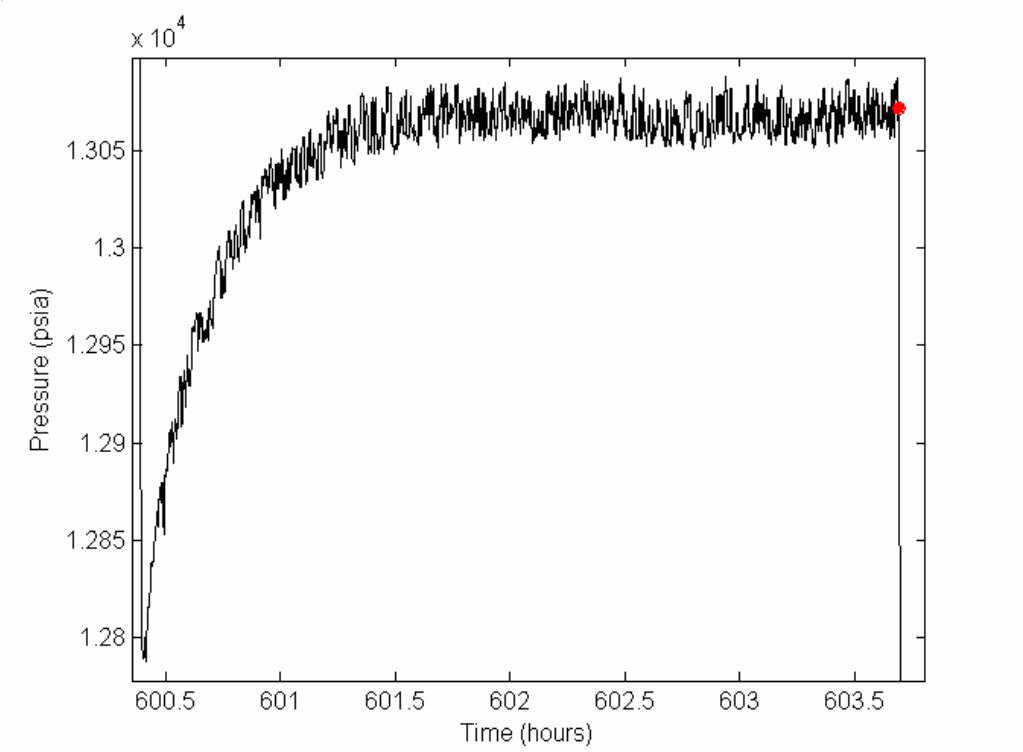

Figure 4-20: Close up view showing no false transients picked by S-G filters method

4.3. Segmentation method

The permanent downhole gauges produce an astonishing number of data. This underlines the need of an algorithm that must work with minimal user input so as to achieve a full automation. In this section an algorithm is described that locates the significant break points in the signal, screening out the small and false transients and uses only one parameter as input. This parameter's value can be bounded from below by the knowledge of the resolution of the instrument or from a measure of the noise in the data. Hence this method has the potential of becoming fully automated. Further this algorithm does not make any assumption about the peak shape and requires no smoothing of data. The algorithm is based on a time series segmentation method that reduces the entire dataset into strategic points.

This algorithm can be outlined in three steps. The first step is to identify what can be called strategic points, by solving a sequence of maximum orthogonal (Euclidean) distance problems. To begin with, the first and the last points in the dataset are marked as two strategic points. Then another point is selected whose orthogonal distance from the line segment joining the two strategic points is greatest. Once a point with greatest orthogonal distance is identified, it joins the collection of strategic points and, in turn, becomes an end point for two new line segments from which a point with the greatest orthogonal distance is chosen. This iterative numerical scheme is performed until the greatest orthogonal distance of data point from the associated line segments falls below a prescribed tolerance τ. This tolerance can be estimated statistically from the dataset itself. Also the selection of these points does not require equally spaced data in time.

Once a set of strategic points are collected, the next step is to calculate the area under each strategic point by considering the two other strategic points to its either side. The area can be approximated numerically through a trapezoidal rule or polygon method. The small or false break points are then deflated using a very small area cutoff which is taken as a very small fraction of the highest area in the dataset. After this step of deflation, further screening is done by checking the forward and backward slope. As a rule of

thumb, if the forward slope is less then half the backward slope, the point is discarded. The slopes were obtained by least square fitting.

Figure 4-21 to Figure 4-23 show a graphical representation of the segmentation step of the method using a simple signal containing three peaks. The first two strategic points chosen are always the last two points in the dataset. A line is drawn connecting these points and the data point with the maximum orthogonal distance from the line is selected as a new strategic point. This process is then iterated until the maximum orthogonal distance falls below a threshold parameter.

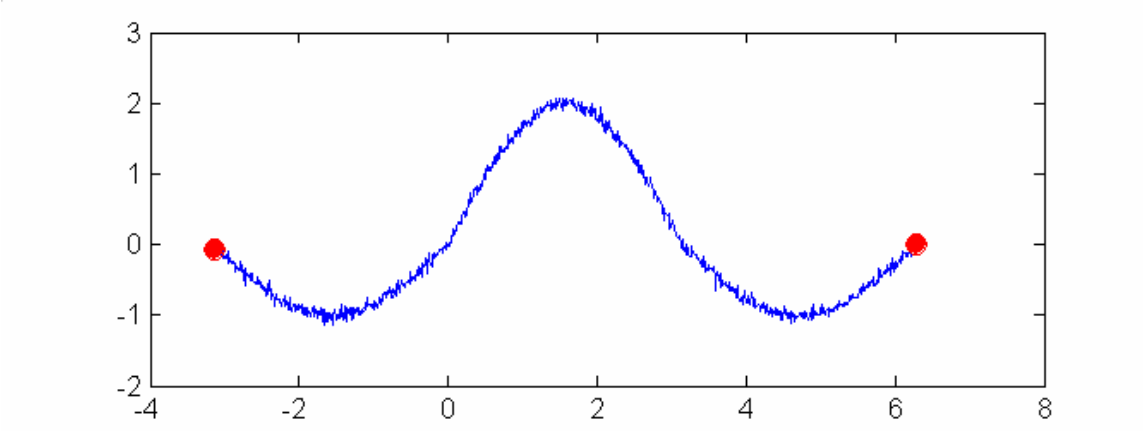

Figure 4-21: The first and the last point in the dataset marked as two strategic points

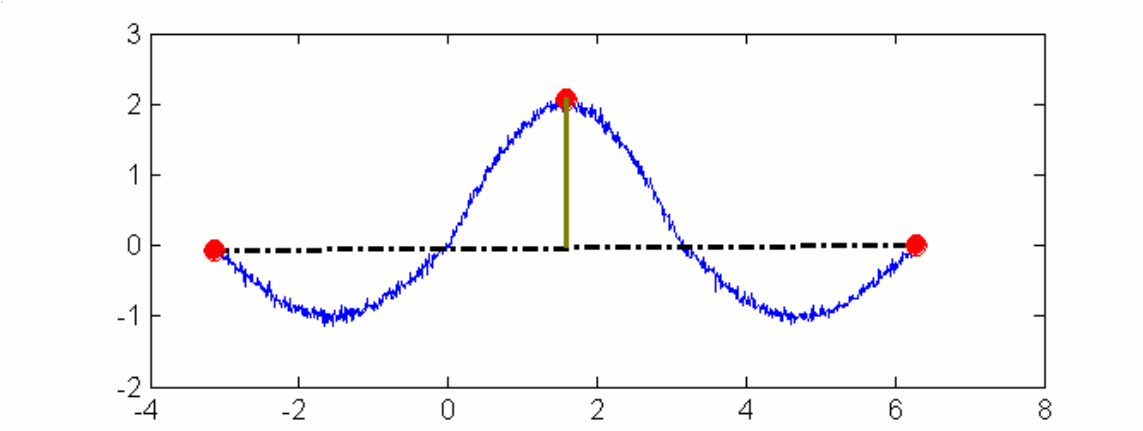

Figure 4-22: Point with maximum orthogonal distance from the line segment joining the first two strategic points selected

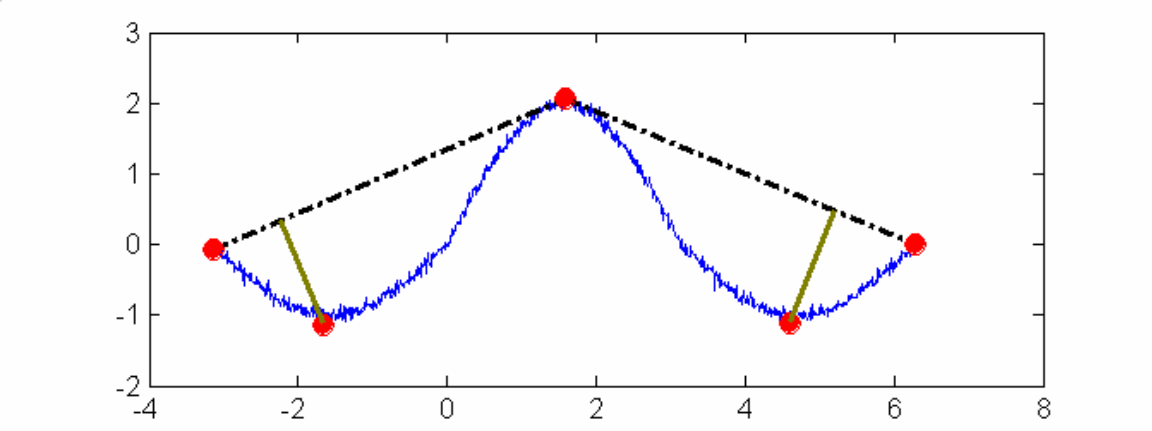

Figure 4-23: Points with maximum orthogonal distance from the line segment joining each consecutive pair of strategic points selected

The most logical lower bound for the tolerance τ can be thought of as an estimate of the noise in the data. Hence we estimate the denoising threshold in the data and denote it by σ. Now in order to test the sensitivity of this parameter to the number of break points identified a field data containing nine bigger transients spanning over a period of approximately one month was taken. The data were sampled uniformly at an interval of five seconds. The denoising threshold was estimated as 0.27 psi, using the noise determination algorithm from Khong (2001). Next the algorithm was used for various values of the tolerance τ ranging from 4σ psi to 40σ psi. The results from the runs are summarized in Table 4-1. It can be seen that even for a large range of the tolerance τ, the number of break points are virtually identical, though a few smaller break points tend to remain when τ is kept very close to the noise threshold.

Figure 4-24 and Figure 4-25 show the break points identified for a lower tolerance 8σ and a higher tolerance 40σ. It can be seen that all the nine significant break points have been identified for both values of τ and only a couple of small break points are left along with them. This shows that the smaller and false break points have been vastly reduced by this method in comparison to the previous methods in Sections 4.1 and 4.2. Also the reliability of this method can be seen in that all the significant break points have been identified for a wide range of the tolerance τ.

Table 4-1: Number of break points identified for different values of **τ** in data set 1

Value of τ (psi)	Number of strategic points	Number of break points
1	109	13
2	72	12
4	61	12
6	52	11
8	47	11
10	38	11

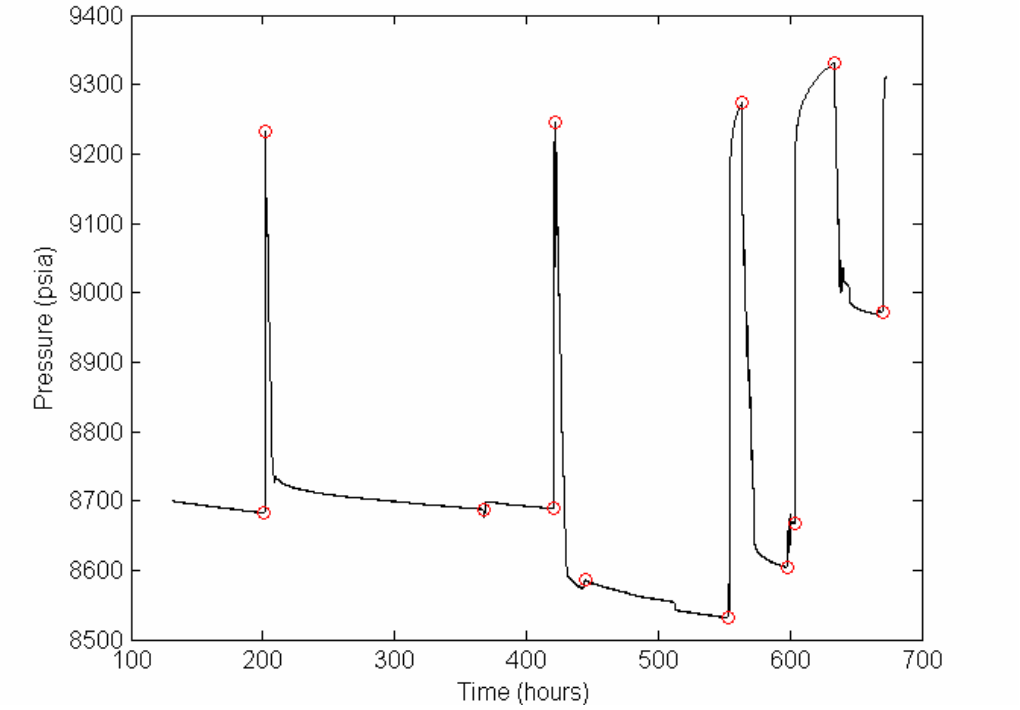

Figure 4-24: Transients identified for a tolerance τ of 2.0 psi

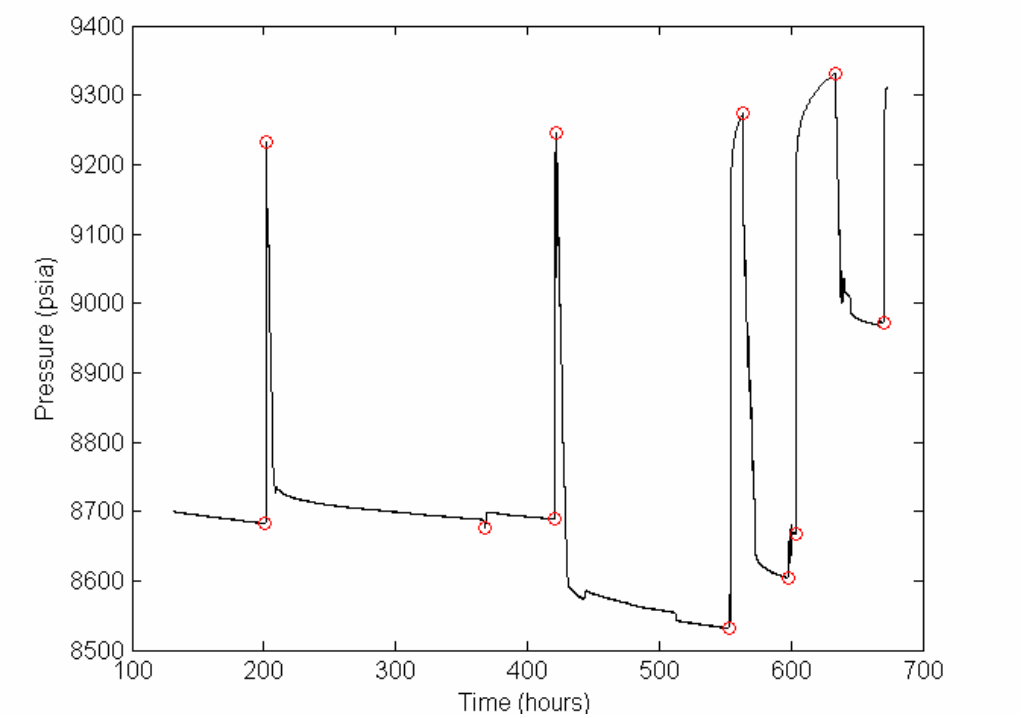

Figure 4-25: Transients identified for a tolerance τ of 10.0 psi

To test the segmentation algorithm further, another set of field data from a different well was considered. The data were sampled uniformly at an interval of five seconds and the data set spanned an interval of over two months. A total of 12 bigger transients are contained in this data along with numerous small transients. The denoising threshold σ was estimated to be 0.34 psi. Again the segmentation algorithm was used for a wide range of the tolerance value from 4σ to 40σ. For a few values of τ the results are summarized in Table 4-2. It can be observed that for these range of tolerance τ the number of break points found do not vary significantly. For values of τ more than 20σ only one lesser break point was selected along with all the 12 significant break points (Figure 4-27). As the tolerance is reduced towards the denoising threshold the number of small break points selected increases by two or three (Figure 4-26).

Table 4-2: Number of break points identified for different values of $\boldsymbol{\tau}$ in data set 2

Value of τ (psi)	Number of strategic points	Number of break points
3	48	16
5	38	16
7	32	15
9	31	14
11	27	13
13	27	13

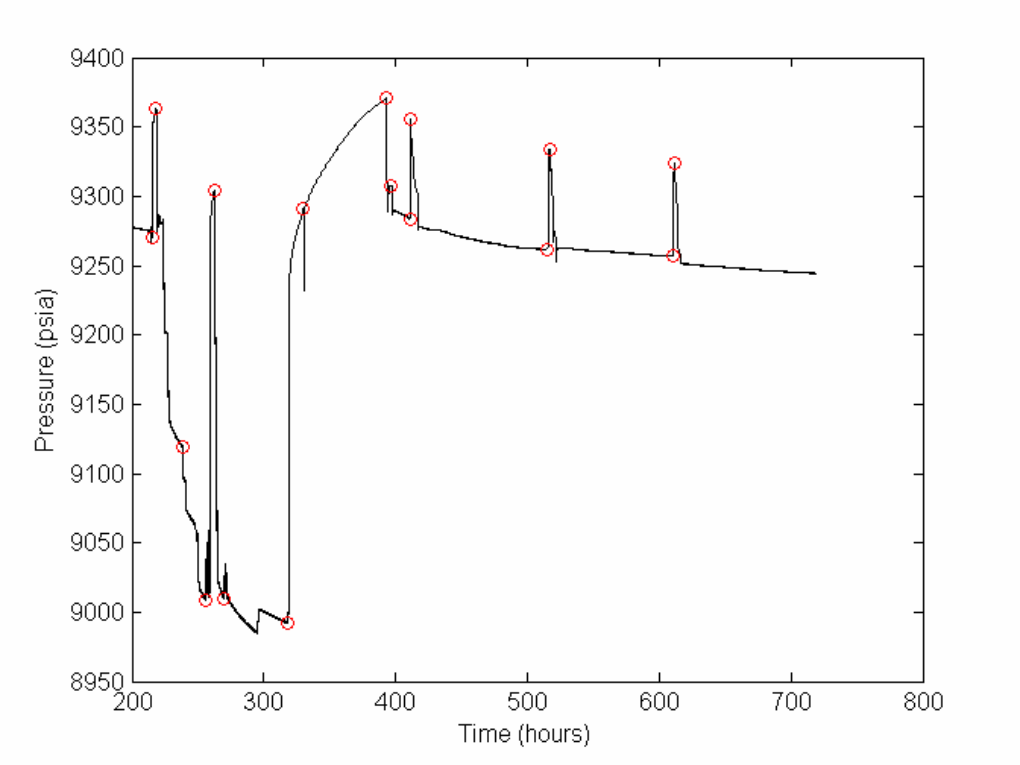

Figure 4-26: Transients identified for a tolerance τ of 3.0 psi

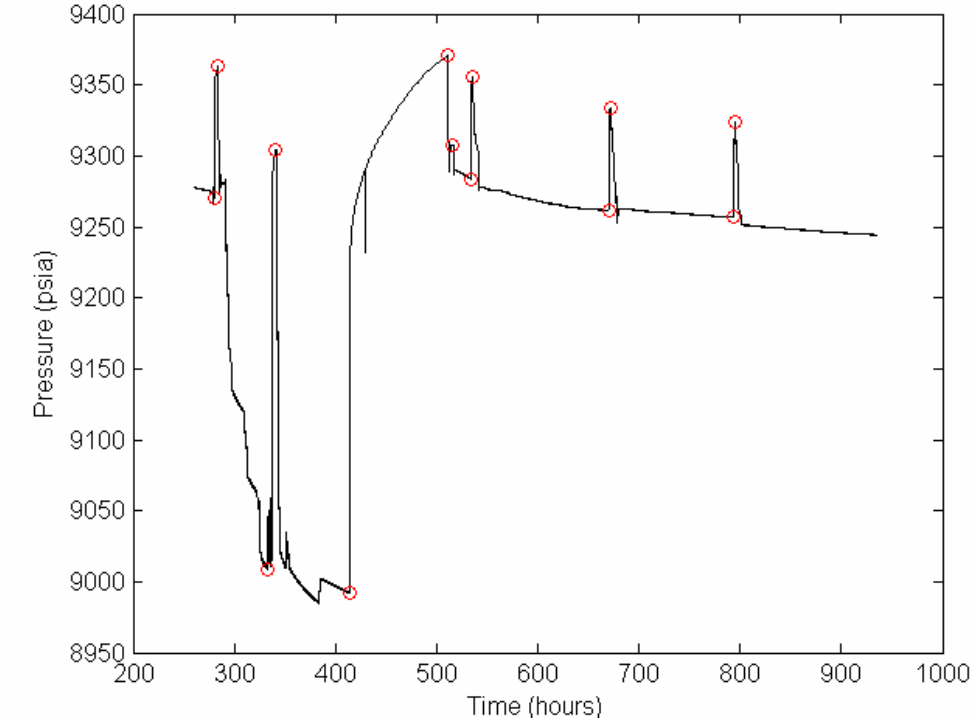

Figure 4-27: Transients identified for a tolerance τ of 13.0 psi

4.4. Variant of segmentation method

Most of the time only pressure data are measured from permanent downhole gauges. However, if simultaneously measured downhole flow rate is also available along with the pressure data, then it is possible to use this additional information for determining break points. This section describes an algorithm which uses a concept similar to the segmentation method and can be called a variant of the segmentation method. The variant method can be used either independently, to determine the significant break points, or in conjunction with the methods described previously in this chapter, to validate and screen smaller or false break points.

Figure 4-28 shows a magnified section of downhole pressure and flow rate plotted on the same time axis. It was observed that the response in the rate data lags that of the pressure data by around a minute. Our algorithm needs to accommodate the fact that the response in downhole flow rate either lags or is at the same time as the pressure data. Now, when the well is closed for build up, the liquid rate will go to zero after some time and when the well is again opened, the liquid rate will increase from zero to some finite value. Hence, the first step in the variant algorithm is to identify the beginning and end of the time period when the liquid rate data is zero. Since the pressure and liquid rate data have been obtained at the same time, the time axis values of the strategic points obtained in the liquid rate data can be transferred onto the pressure data. Now, the next step is to use the concept of the segmentation method and look for points in the pressure data having maximum or minimum orthogonal distance from the time axis, between a pair of strategic points. The point with minimum orthogonal distance will be the break point corresponding to the beginning of a drawdown, while the point with maximum orthogonal distance will be the break point corresponding to the beginning of a buildup.

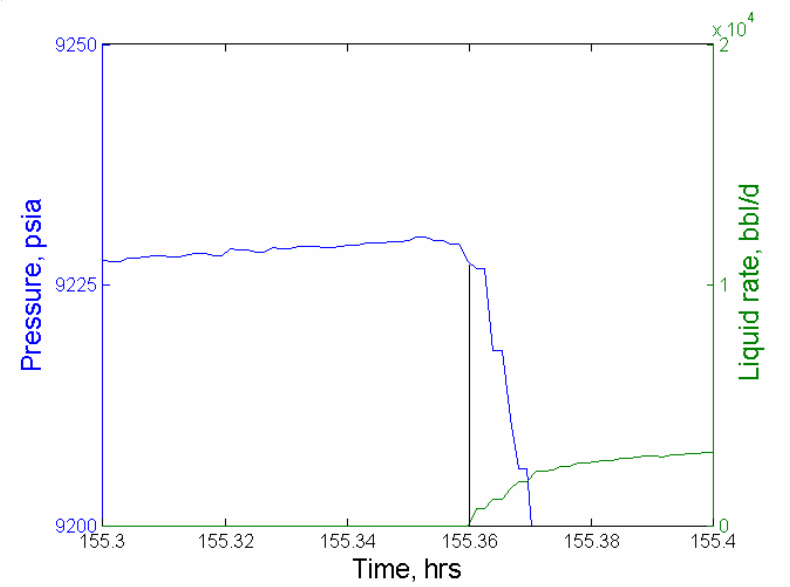

Figure 4-28: Magnified section of pressure (blue color) and flow rate (green color) plotted on same time axis

To test the variant algorithm a field dataset with simultaneous record of downhole flow rate and pressure, with data sampled evenly at an interval of 5 seconds, was considered. Figure 4-29 and Figure 4-30 show the downhole liquid rate and pressure data respectively. The resulting break points identified by the application of the variant algorithm to the liquid rate and pressure data are shown in Figure 4-30. It was observed that all the significant break points were identified except for one drawdown break point at around 420 hours. In this case, the point at 368 hours was identified by the variant algorithm instead because this point was having the minimum orthogonal distance between the strategic points at around 202 hours and 421 hours. Another field data set of simultaneously measured flow rate (Figure 4-31) and pressure (Figure 4-32) was considered. Figure 4-32 shows the break points identified by the variant algorithm, superimposed on the downhole pressure data. In this case too, all of the significant transients, except for one at 794.2 hours, were identified.

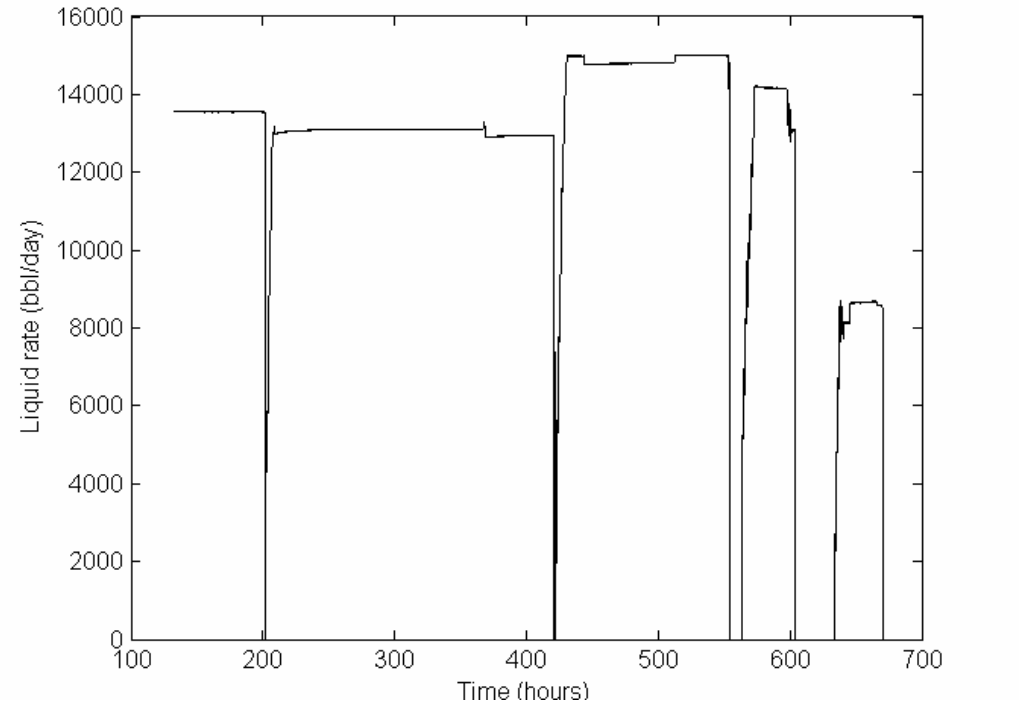

Figure 4-29: Continuously measured flow rate used for identifying transients

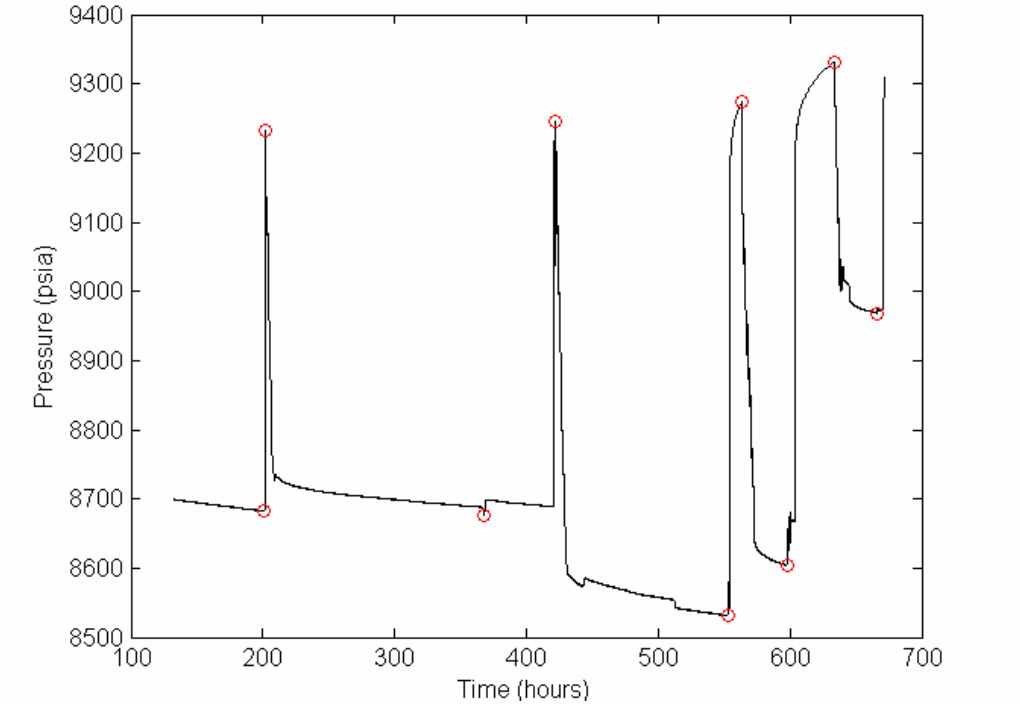

Figure 4-30: Transients identified using both pressure and flow rate information

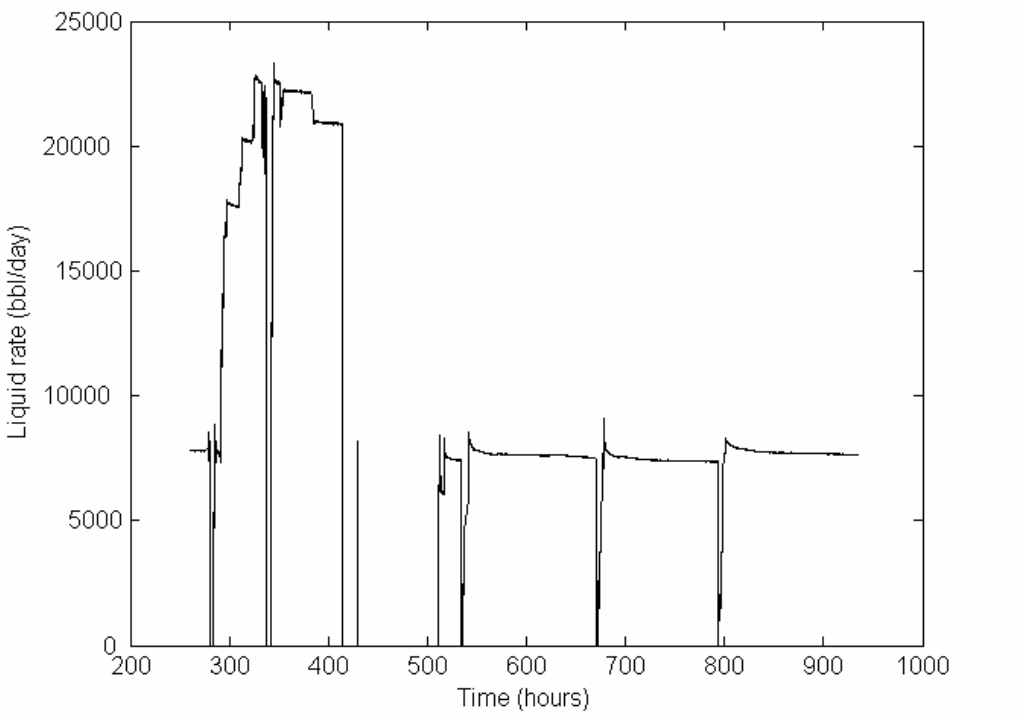

Figure 4-31: Continuously measured flow rate used for identifying transients

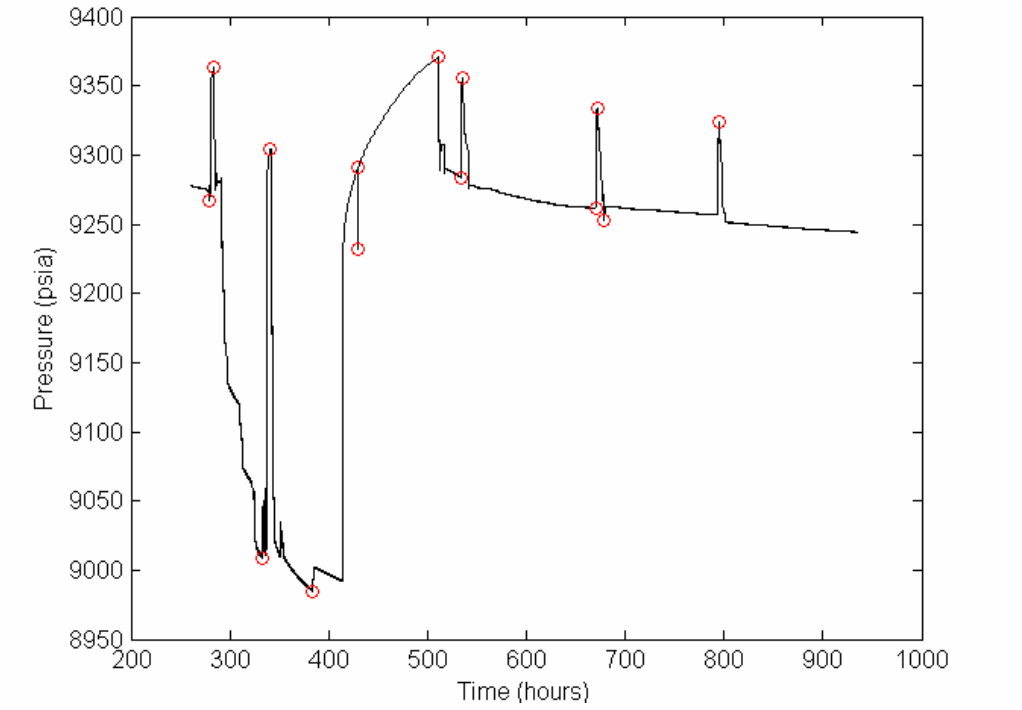

Figure 4-32: Transients identified using both pressure and flow rate information

Chapter 5

5. Impact of continuous rate data on interpretation

Accurate measurements of flow rate is a very important ingredient in obtaining various well and reservoir parameters from pressure transient analysis. Unfortunately, rates are very rarely monitored continuously and the problem of accurate flow rate measurement is worse in cases when many wells are flowing in a commingled fashion from different layers. This often leads us to assume a constant rate over an entire period of buildup or drawdown test. However analysis with constant rate assumptions is hindered greatly by several factors like afterflow effects. Guillot and Horne (1986) emphasized the use of simultaneous measurements of pressure and downhole flow rates to overcome many problems like afterflow among others. In this chapter we will discuss the possible impact of continuous rate measurements on pressure transient analysis.

Athichanagorn et al. (2002) and Khong (2001) developed a methodology for flow rate history reconstruction where flow rate information is inadequate. In order to solve the problem of lack of flow rate measurements, Athichanagorn included the unknown flow rates as model parameters in a nonlinear regression model to match pressure response. Further, to increase the uniqueness of estimation of flow rates, Athichanagorn sought to constraint the regression model to some known flow rates or cumulative production data. An issue associated with this method is that the quality of the result could be affected by the noise in the pressure data. Another issue with the window method of determining unknown flow rates is that for a given transient it seeks to determine one average flow rate, for a given time step, usually in hours, over the entire transient.

Most of the data obtained from permanent downhole gauges are different from that in a planned buildup or drawdown test monitored under strict control. Mostly the shut-ins or drawdowns are unplanned, occurring due to shutdown of facilities, due to operational

conditions in the flowlines and the well. Generally, well shut-ins or openings are performed by either operating a valve at the facilities away from the well or by operating the flowline wing valves/chokes at the wellhead tree. Shutting a valve at the facilities will have afterflow effects depending on the length of the flow pipelines on the surface. Operating the valves/chokes could also take time, as closing the valve quickly against the flow is not advisable, especially against high flow rates, to avoid erosion effects. Due to all these reasons, most of the tests that we normally obtain from permanent downhole gauges do not have a flow rate that goes instantly to zero during build up or comes up instantly to a near plateau rate during drawdown. Figure 5-1 shows continuous flow rate data recorded by a permanent downhole flow meter for a period of around a month. The flow rate data contains a total of nine buildup and drawdown tests and in all of them the flow rate does not go instantly to zero or come up instantly to a plateau rate during drawdown. Figure 5-2 and Figure 5-3 show a close up view of the permanent downhole flow rate corresponding to a buildup and drawdown test respectively. In the build up test, the downhole flow rate goes to zero after an interval of around 10 minutes after the beginning of the test, as a result of afterflow effects, whereas, in the drawdown test, it takes around 8 hours for the downhole flow rate to go to a near plateau rate of 23,000 stb/day.

Coludrovich et al. (2004) also carried out interpretation work with simultaneously measured pressure and rate data, available from permanent downhole gauges. Coludrovich et al. called these tests, where the downhole rate does not go to zero instantaneously, as “soft” tests and the ones where the rate goes to zero instantaneously as “hard” tests. Coludrovich et al. found that when they interpreted these “soft” tests assuming them to be “hard”, i.e. the rate going to zero instantaneously, it yielded unrealistically low estimates of permeability and skin factor, as compared to the results from a few “hard” data.

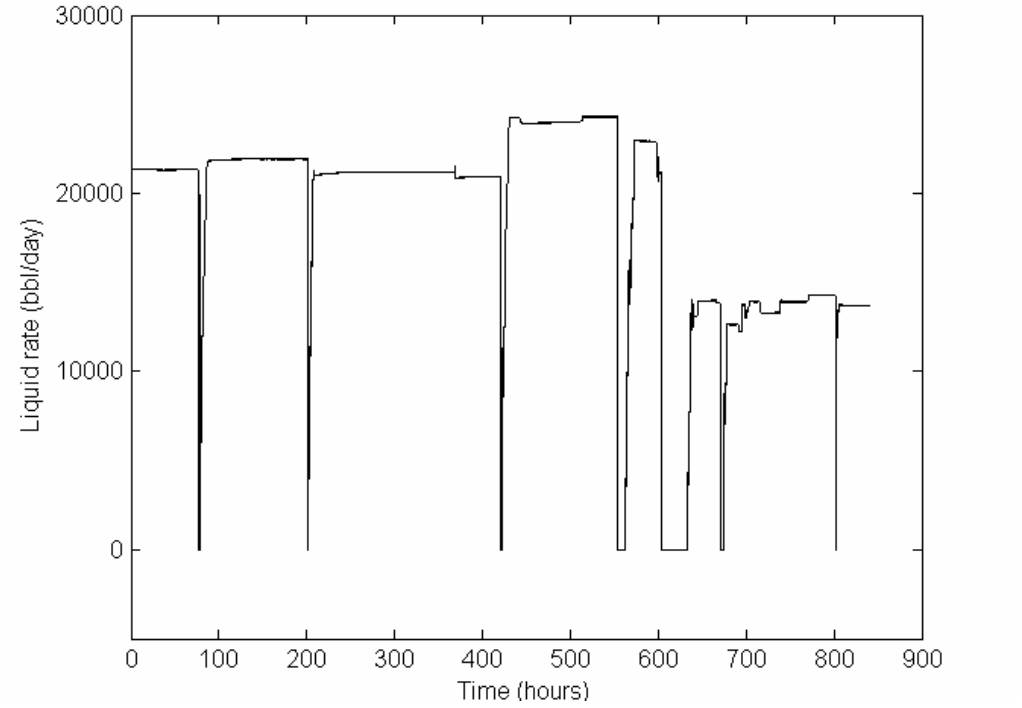

Figure 5-1: Flow rate gathered from PDG over a period of a month

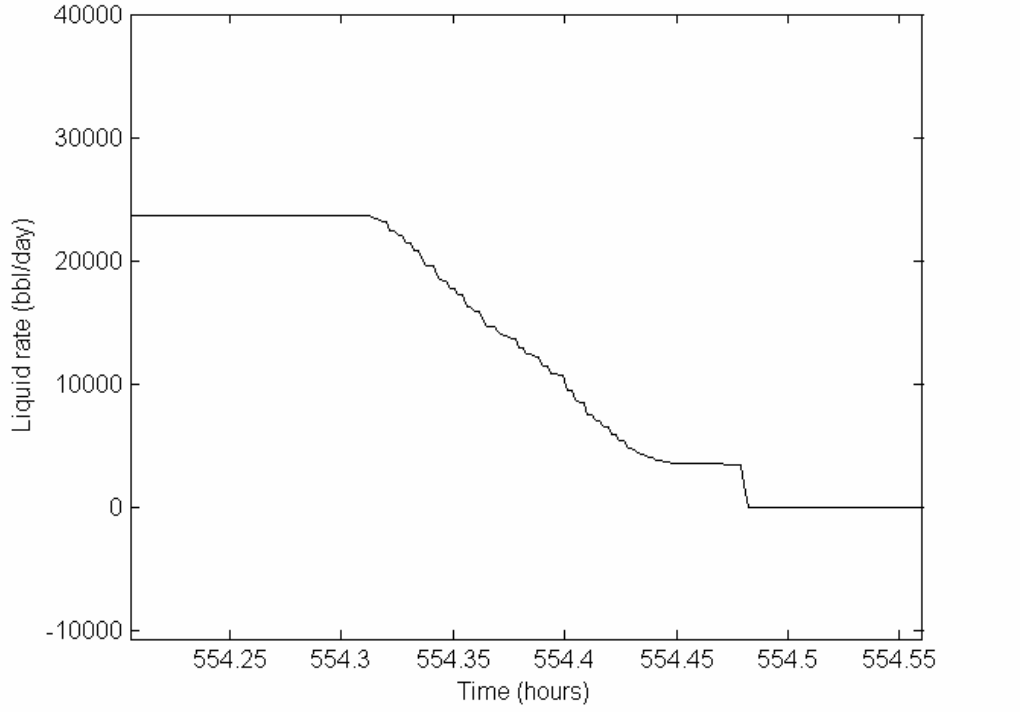

Figure 5-2: Magnified section of flow rate during a build up

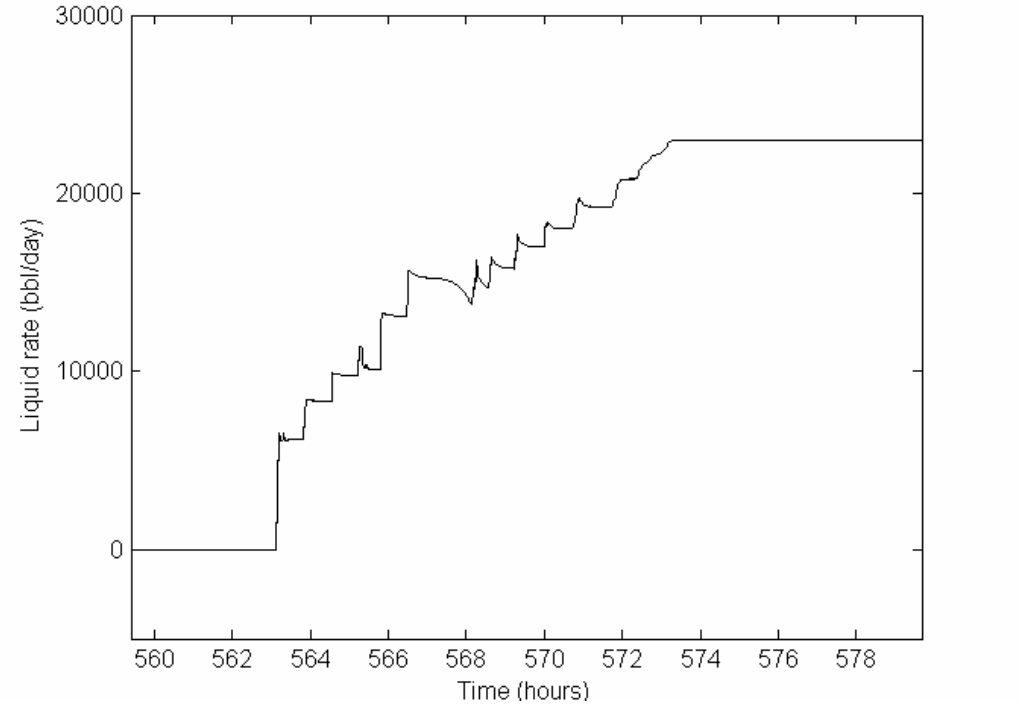

Figure 5-3: Magnified section of flow rate during a drawdown

An observation made during this work was that many times the bigger transients contained many smaller transients especially during a drawdown period when rate begins from zero and goes to a near plateau value. The effect of using an average rate, for the periods of smaller transients, instead of the entire rate history is investigated here using the flow rate shown in Figure 5-3. For simplification, a test pressure was generated using an infinite-acting reservoir with wellbore skin as the model, using the actual flow rate (blue color) history shown in Figure 5-4. Figure 5-5 shows the generated pressure data (blue color) over a period of around 8 hours with sampling interval of five seconds. Table 5-1 shows the permeability and skin values, which were used for generating the pressure response. Next the rate data (red color) was averaged volumetrically over the smaller transients contained within the bigger transients and is shown in Figure 5-4, plotted over the actual flow rate. A regression match was carried on pressure with the volumetrically averaged flow rate, using Saphir (Kappa software). The pressure match (red color) obtained from regression is shown in Figure 5-5. Permeability and skin values (Table 5-1) of 390 md and -2.5 respectively were obtained from the regression match. It can be seen

from the results that lower values of permeability and skin are obtained by using the averaged flow rates as compared to using the continuously measured flow rate. Hence in the absence of continuously measured flow rate data, there could be considerable error in the results obtained from normal analysis using an averaged flow rate, which could render the correct interpretation of these “soft” tests, obtained from permanent downhole gauges, meaningless.

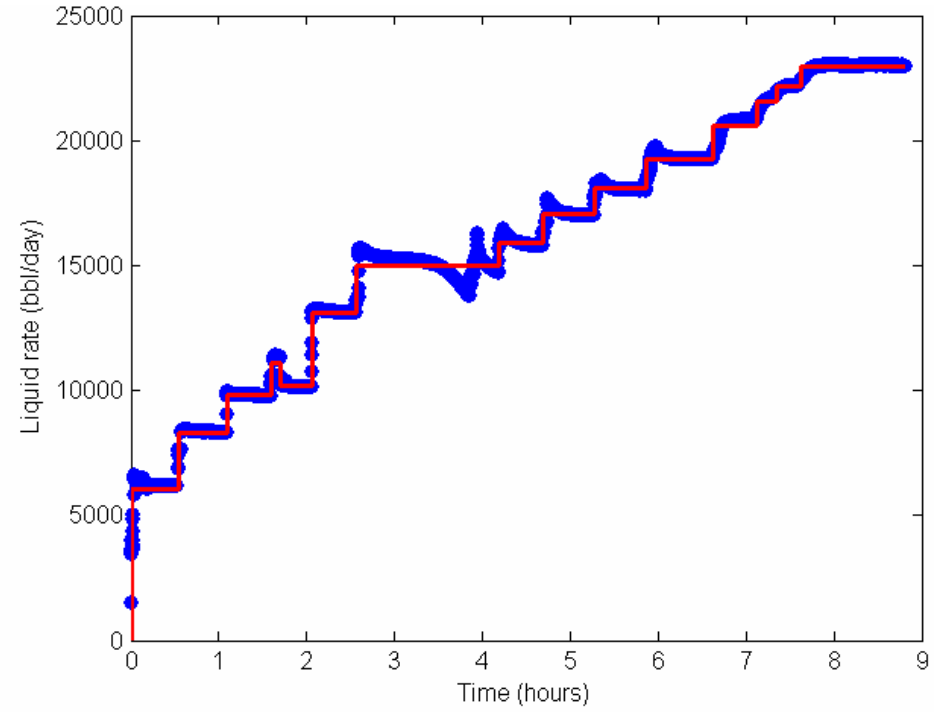

Figure 5-4: Actual flow rate (blue color) and volumetrically averaged flow rate (red color)

Table 5-1: Permeability and skin obtained from actual flow rate and averaged flow rate

	Actual flowrate data	Averaged flowrate data
Permeability (md)	700	390
Skin	2	-2.25

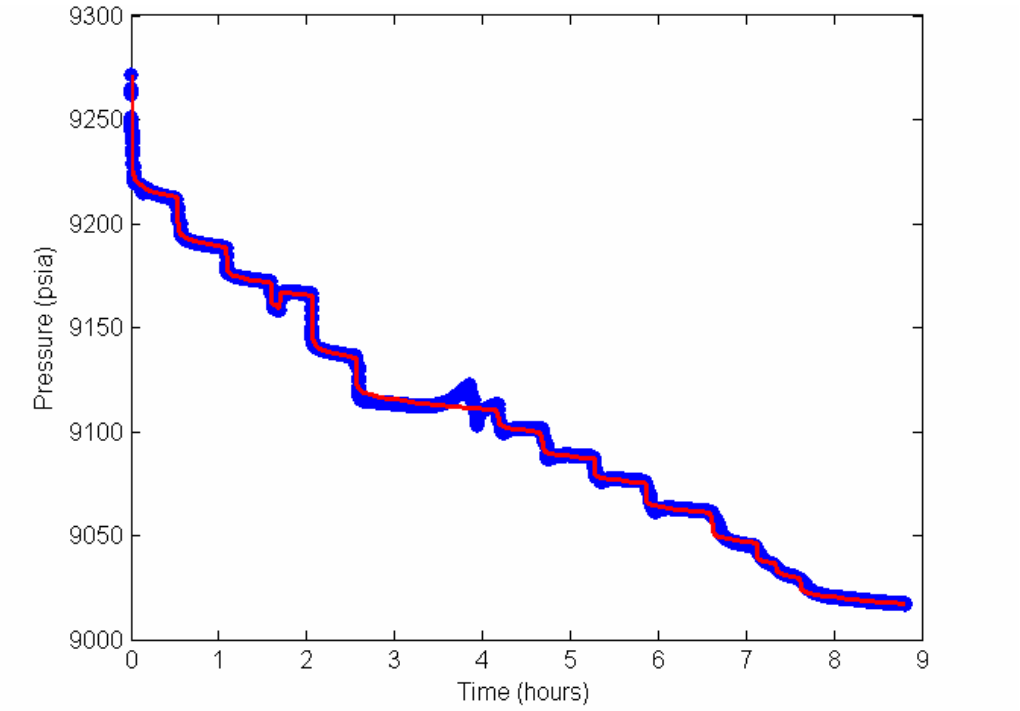

Figure 5-5: Actual pressure response (blue color) and pressure response from regression match (red color)

Chapter 6

6. Conclusions and recommendations

Some aspects of permanent downhole flow rate and pressure data have been investigated in this work. First of all the wavelet processing algorithm developed by Athichanagorn et al. (2002) was modified slightly to correctly process the downhole flow rate. Next, an important issue of accurate and reliable identification of transient break point was addressed. It was found that the spline wavelet algorithm (Athichanagorn et al., 2002) had some limitations in terms of identifying the significant transients while screening out smaller and false break points. A host of new methods were investigated, which sought to improve the effectiveness and reliability of break point identification. Further the impact of using continuously measured downhole flow rate on pressure transient interpretation was investigated. It was observed that in cases of "soft" transients, the interpretation based on average flow rate could provide erroneous results.

Some important conclusions, summarizing the work described in the previous chapter, are as follows:

1. The wavelet method, developed by Athichanagorn et al. (2002) for processing pressure data, can be used for effectively processing flow rate data too, after a slight modification to ensure nonremoval of zero rates corresponding to buildups, as step outliers.

2. The spline wavelet algorithm for transient identification was found to have limitations in screening out smaller and false break points effectively while retaining the significant transients. The calculation of slope threshold at a higher level of decomposition j was found to be highly dependent on the signal shape as well as the type of wavelet being used. Hence, it is difficult to predict the slope threshold at a higher level of signal decomposition j from the slope threshold at

the first level of decomposition. Further, it was found that slope threshold alone is not a very effective criterion for screening smaller transients contained in bigger transients.

3. As an alternate algorithm, the stationary wavelet transform (SWT) using the Haar wavelet, which is the most compactly supported orthogonal wavelet, was applied to identify significant break points. This approach too, with only its slope threshold criterion, was not very effective in the screening of smaller break points.

4. As an alternative, a Savitzky-Golay (S-G) polynomial smoothing filter based algorithm was used as another algorithm for transient identification. In this method, first and higher derivatives of the signal, calculated in an efficient way using the S-G filters, were used to identify the accurate location of break points. Further, the area under the first derivative, which gives the net change in pressure or flow rate, was used as another criterion for screening of smaller break points. It was found to perform better than the wavelet based approaches, both on pressure and flow rate data, by identifying most of the true break points while retaining fewer small and false break points. Also, this method could screen out false break points in regions of high noise variance, which otherwise were picked by the spline wavelet algorithm as break points. Still, the S-G filters method had some limitations in that it had difficulty in screening some smaller transients that are contained in the significant transients.

5. Another algorithm based on a time series segmentation method that reduces the entire dataset into strategic points was applied for identifying significant breakpoints. Smaller and false break points were further screened out by using the area, calculated under a transient using a polygon method, as a screening criterion and by checking the forward and backward slope. The method was applied to real field pressure datasets and was found to be very effective in identifying all the significant break points while retaining very few smaller transients, for a wide range of the stopping criterion of segmentation, τ.

6. Another variant of the segmentation method was developed, which uses both flow rate and pressure data to identify the true transients. The algorithm first identifies the zeros in the flow rate and then between two consecutive zeros finds the peak using the segmentation approach. It was found that this method performs as well as the segmentation method, but requires both pressure and flow rate data.

7. Continuously measured flow rate was found to affect the interpretation of pressure transient tests especially in cases where the tests are “soft” i.e. the flow rate does not go instantaneously to zero or a plateau value. Using averaged flow rate instead of the actual flow rate tends to give a low estimate of permeability and skin.

The following points can be considered from the viewpoint of carrying this research work further:

1. Flow rate data combined with pressure data has a lot of potential for providing additional information about the well and reservoir properties. Further ways can be explored to utilize the continuously measured flow rate and pressure data for better pressure transient analysis (PTA) and production analysis (PA).

2. The permanent downhole gauge data, gathered at a high frequency over a long period of time and from different channels simultaneously, presents a huge problem of data handling and storage. Hence, the development of efficient dynamic memory handling capabilities, to manage loading and reloading of data during processing and interpretation, can be looked into.

3. Flow rate data can play an important role in another aspect of data processing, i.e. behavioral filtering. Athichanagorn et al. (2002) used variance between the regression match and the actual data as a criterion to eliminate transients with unusually high variance. Using average rate data in the regression match could result in unwanted removal of transients that are actually a result of variation in flow rate. Figure 6-1 illustrates this possibility where the transient between 2.5 and 4 hours will be removed using regression with average flow rate, even though

it is actually a response of the flow rate change (Figure 6-2). Hence, continuously measured flow rate data can be used to make the process of behavioral filtering more robust and unambiguous.

4. S-G filters and segmentation can be applied to a wide variety of real field data to determine possible pitfalls in the methods and further make the methods more robust. One potential area for improvement in both the methods is the calculation of area under the transient and its use in effectively screening smaller and false transients.

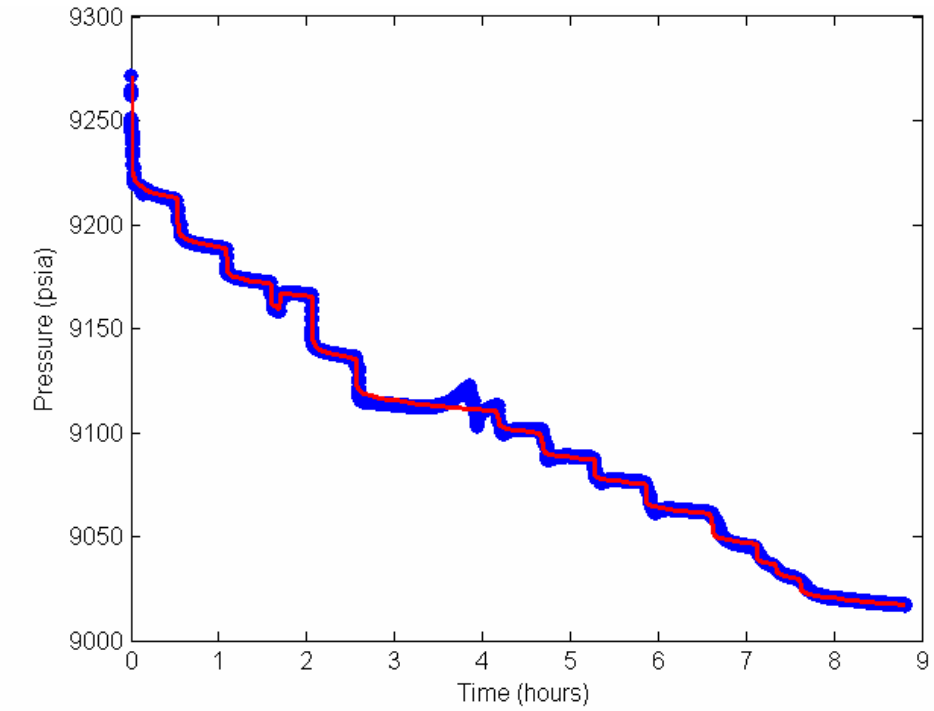

Figure 6-1: actual flow rate (blue color) and volumetrically averaged flow rate (red color)

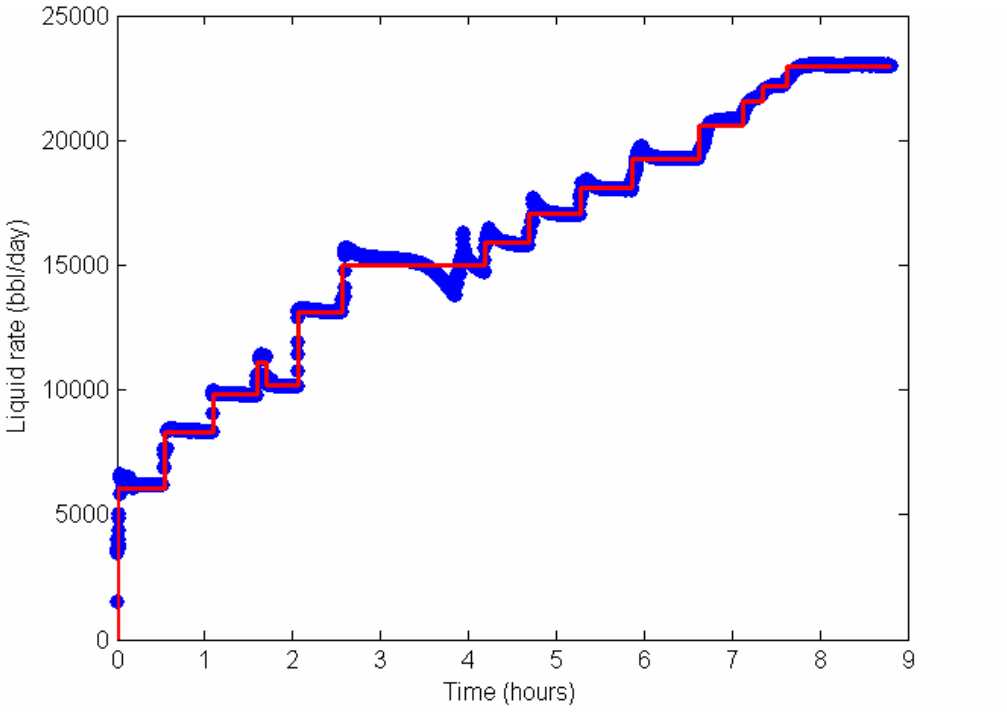

Figure 6-2: Actual pressure response (blue color) and pressure response from regression match (red color)

Nomenclature

a = Dilation step
c_n = Savitzky-Golay filter coefficients
g_i = polynomial of degree M in i
G = Differentiation filters
M = degree of polynomial
n_L = number of points to left of a data point i
n_R = number of points to right of a data point i
S_h = Circulant shift by h
$S_{i,j}$ = Matrix of Polynomial basis vectors
t = time, hours
u = Time Translation parameter
W = Wavelet transform operator
$\overline{W}$ = Stationary wavelet transform operator

GREEK LETTERS

Ψ = Wavelet function
Φ = Scaling function
$\Im$ = Performance index

SUBSCRIPTS

i, j, k, L, n = Integer indices

SUPERSCRIPTS

T = Transpose of matrix or vector

References

Athichanagorn, S.: *Development of an Interpretation Methodology for Long-Term Pressure Data from Permanent Downhole Gauges*, PhD dissertation, Stanford University, June 1999.

Athichanagorn, S., Horne, R.N., and Kikani, J.: "Processing and Interpretation of Long-Term Data Acquired from Permanent Pressure Gauges", SPE Reservoir Evaluation & Engineering, 5(5), October 2002, 384-391.

Bromba, M.U.A. and Ziegler, H.: "Application Hints for Savitzky-Golay Smoothing Filters," *Analytical Chemistry*, vol. 53, pp. 1583-1586, 1981.

Coifman, R.R. and Donoho, D.L.: "Translation Invariant De-noising," *Lecture Notes in Statistics*, vol. 103, pp. 125-150, 1995.

Coludrovich, E.J., McFadden, J.D., Palke, M.R., Roberts, W.R. and Robson, L.J.: "The Boris Field Well management Philosophy – The Application of Permanent Downhole Flowmeters to Pressure Transient Analysis: An Integrated Approach," paper *SPE* 90316 presented at the 2004 SPE Annual Technical Conference and Exhibition, Houston, TX, September.

Douglas, D.H. and Peucker, T.K.: "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature," *The Canadian Cartographer*, vol. 10, issue 2, pp. 112-122, 1973.

Guillot, A.Y. and Horne, R.N.: "Using Simultaneous Downhole Flow-Rate and Pressure Measurements to Improve Analysis of Well Tests," *SPE Formation Evaluation*, (June 1986), 217.

Horne, R.N.: *Modern Well Test Analysis, A Computer-Aided Approach*, Petroway, Inc., Palo Alto, 2nd edition (1995)

Khong, C.K.: *Permanent Downhole Gauge Data Interpretation*, Master's thesis, Stanford University, June 2001.

Kikani, J. and He, M.: "Multi-Resolution Analysis of Pressure Transient Data Using Wavelet Methods," paper SPE 48966 presented at the 1998 SPE Annual Technical Conference and Exhibition, New Orleans, LA, September.

Mallat, S.: *A Wavelet Tour of Signal Processing*, Academic Press (1998).

Nason, G.P. and Silverman, B.W.: "The stationary wavelet transform and some statistical applications," *Lecture Notes in Statistics*, vol. 103, pp. 281-299, 1995.

Orfanidis, S.J.: *Introduction to Signal Processing*, Prentice Hall, Upper Saddle River, (1996).

Pesquet, J.C., Krim, H. and Carfatan, H.: "Time-invariant orthogonal wavelet representations," *IEEE Transaction on Signnal Processing*, vol. 44, issue 8, pp. 1964-1970, 1996.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P.: *Numerical Recipes in Fortran*, Cambridge University Press, 2nd edition (1992)

Savitzky A. and Golay, M.J.E.: "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, pp. 1627-1639, 1964.

Walker, J.S.: *A Primer on Wavelets and their Scientific Applications*, Chapman & Hall/CRC (1999).

Wallace, W.E., Kearsley, A.J. and Guttman, C.M.: "An Operator-Independent Approach to Mass Spectral Peak Identification and Integration", *Analytical Chemistry*, vol. 76, pp. 2446-2452, 2004.

Appendix A

A. Program Guide

This section is intended as a quick guide for the execution of the four break point identification programs. Description of the input files needed and output files produced after executing the programs are provided here. A listing and explanation of the various user input parameters required to run the programs is also included.

The input and output file formats are common to all the four break point identification programs. The input file should be a text file with any file extension (*.txt, *.tpr etc.). The input file should have only two rows of data, without any text in any part of the file. The first row should contain the time data in hours and the second row should have the pressure or liquid rate data. An example of the input file format is shown below:

Filename.txt
0.0013889 9200.67
0.0027778 9168.78
0.0041667 9140.89
.
.
.
200.0 7698.98

The output text file with *.tpr as the filename extension, contains the break points identified in the permanent downhole gauge data. The format of data in the output file is similar to the input file with two rows of time and pressure/liquid rate data. The input file and the programs should be in the same directory and the output file will also be generated in the same directory. The break points identified by all the break point identification algorithms are written in the output file called break_point.tpr. Further, we

will discuss the various user input parameters required in different break point identification algorithms.

A.1. SWTHaar.m

Stationary wavelet transform using Haar wavelet provides translation-invariant signal decomposition. The input data should be sampled uniformly and the data size should be a multiple of 2^4, where the power four is the level of decomposition. A threshold value has to chosen in the detail signal for identifying the transients. A value for calculating the fraction of the largest wavelet modulus maximum of all the maxima in the signal has to be given as user input. The fraction of the largest wavelet modulus maximum of all the maxima is taken as the threshold, above which all the break points are detected. A plot (Figure A-1) of the detail signal is generated by the program, as a reference for the user to choose the value for calculating the threshold.

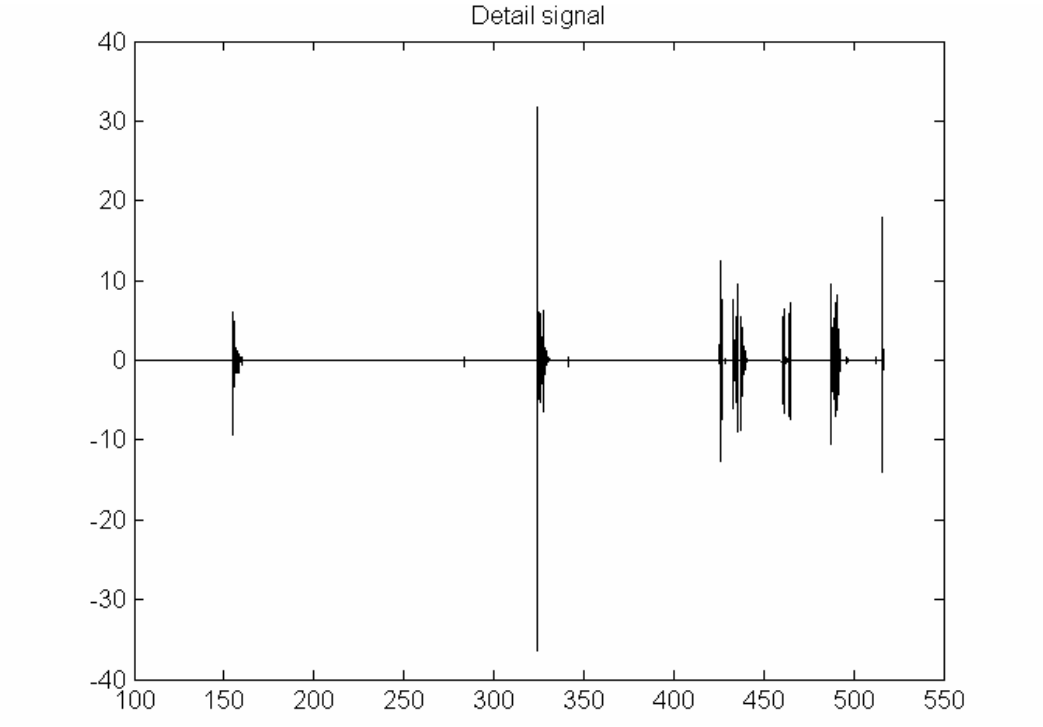

Figure A-1: Detail signal

An example of the Matlab console input for the Haar wavelet algorithm is presented below:

Filename (with filename extension: .txt .tpr .dat): pressdat.tpr
Program output: largest wavelet modulous maxima: ---
Max_WMM = 105.0806
Input a value for calculating the fraction of the largest wavelet modulous maxima: 10

A.2. SGfilter.m

Savitzky-Golay (S-G) filters are used for smoothing the signal and calculating its derivatives. Here, a numeric value is required as user input for calculating the fraction of the largest first derivative peak, to be used as a threshold for identifying break points. The numeric value should be sufficiently high; so as to identify all the smaller and larger break points and can be chosen from a range of 30-60. A plot (Figure A-2) of the first derivative is generated by the program, as a reference for the user to choose the value for calculating the threshold. After obtaining all the break points, the area under the first derivative peaks are calculated for each break point. Now an area threshold and a peak height threshold needs to specified for screening out the smaller and false break points. Again a numeric value for calculating fraction of the maximum area and the maximum first derivative peak has to be specified by the user as input. Typically, these values can range from 5 to 10.

An example of the Matlab console input for the S-G filters algorithm is presented below:

Filename (with filename extension: .tpr .dat .denoise): pressdat_comp.tpr
Input the order of polynomial: 4
Input the window size of data for least square fitting: 51
Input a value for calculating fraction of largest first derivative peak: 30
Input a value for calculating fraction of largest area for screening: 10
Input a value for calculating a fraction of largest first derivative peak for screening: 10

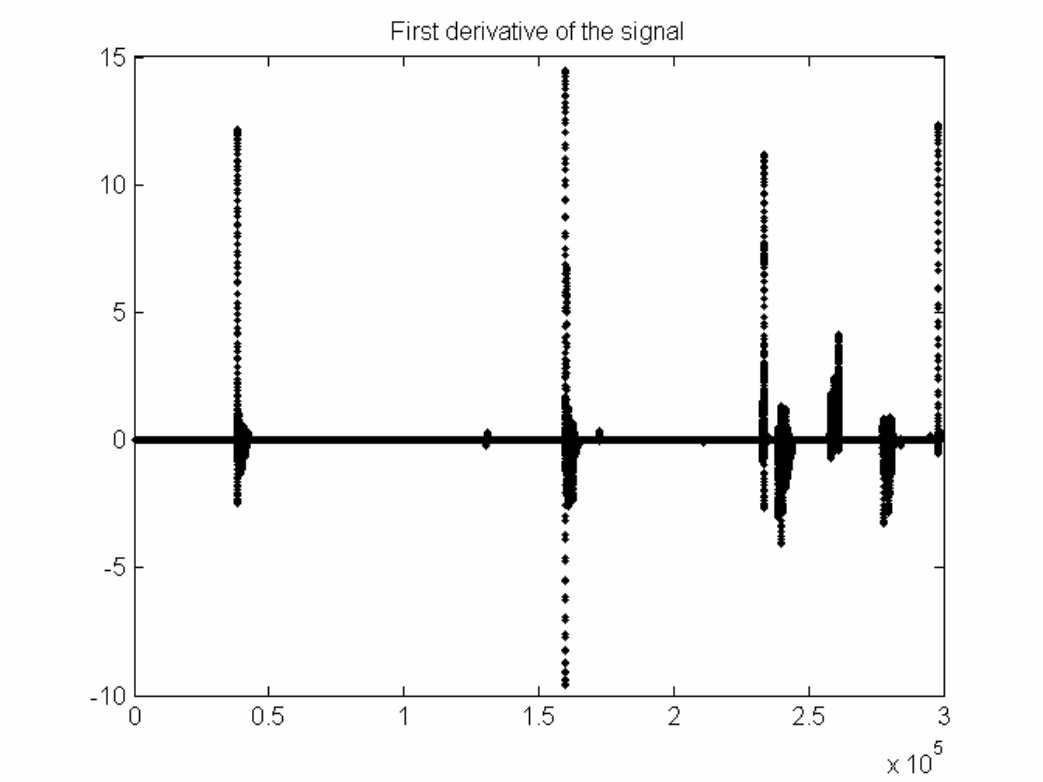

Figure A-2: First derivative calculated from S-G filters

A.3. segment.m

This method does not require that the data be sampled uniformly. In this method, we iteratively solve a sequence of maximum orthogonal (Euclidean) distance problem. The iterative numerical scheme is performed until the greatest orthogonal distance of data point from the associated line segments falls below a prescribed tolerance τ. The lower bound of τ can be derived from an estimate of the denoising threshold σ in the data. Hence, first an estimate of the noise threshold σ in pressure units (psi) is made from the data, by calling a program called noise.m developed by Khong (2001). Next a multiplier

of the denoising threshold σ has to be given as the value of tolerance of τ. For all practical purposes, it was found that the range of τ can be taken as 15σ-30σ. The lower bound of 15σ is to prevent the noise peaks from being identified as break points. Also, the height of the smallest peak to be identified should be considered while deciding the value of τ. For example, let us consider the permanent downhole pressure data in Figure A-3. The smallest peak i.e. the difference between a drawdown point (at 515 hours) and a buildup point (at 517 hours) is around 70 psi. It was observed that the value of τ should be less that 5-10 times of the smallest peak (70 psi), for preventing the exclusion of any significant break point. Further, after identifying the strategic points, the area under the transient is calculated and a fraction of the maximum area is to be given by user as input to screen strategic points with very small area. Typically, the value for determining the fraction of the area cutoff can be chosen from the range of 300-600.

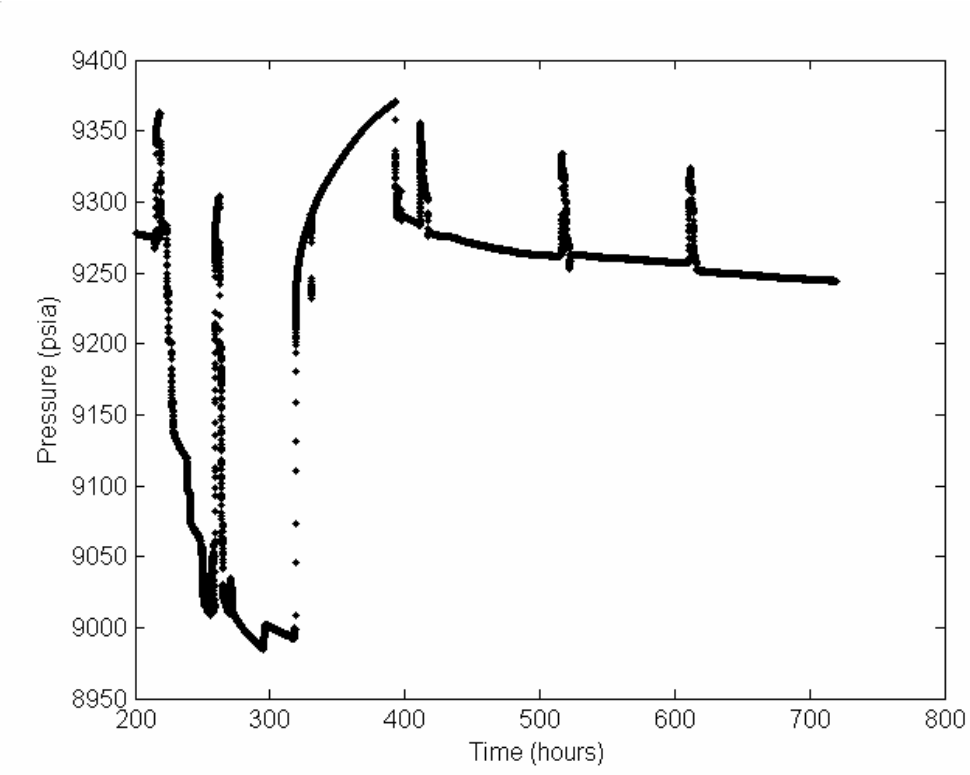

Figure A-3: A field data set of permanent downhole pressure

An example of the Matlab console input for the segmentation method is presented below:

Filename (with filename extension: .txt .tpr .dat): well2_comp.tpr
Program output: denoising threshold (psi) ---
noise_thres = 0.1222
Input the multiplier to denoising threshold for obtaining the tolerance parameter: 15
Program output: value of the largest area under a strategic point (psi^2) ---
Maximum area = 5000
Input a value for calculating the fraction of the maximum area, for screening: 400

A.4. variant_segment.m

In the variant algorithm, simultaneously measured pressure and rate data is required. First we identify the beginning and end of zeros in the rate data and then transfer those points on to the pressure data for finding the drawdown and buildup peaks. The variant method does not require any other user input parameter, apart from the pressure and liquid rate files.

An example of the Matlab console input for the variant algorithm is presented below:

Rate data filename (with filename extension: .txt .tpr .dat): well2_rate.tpr

Pressure data filename (with filename extension: .txt .tpr .dat): well2_pres.tpr