DOE/BC/14894--8

APPLICATION OF ARTIFICIAL INTELLIGENCE TO RESERVOIR CHARACTERIZATION: AN INTERDISCIPLINARY APPROACH

(DOE Contract No. DE-AC22-93BC14894)

Submitted by

The University of Tulsa
Tulsa, OK 74104

Contract Date:	October 1, 1993
Anticipated Completion Date:	September 30, 1996
Government Award:	$240,540
Program Manager:	B.G. Kelkar
	R.F. Gamble
Principal Investigators:	D.R. Kerr
	L.G. Thompson
	S. Shenoi
Reporting Period:	April 1 - June 30, 1995

Contracting Officer's Representative

Mr. Robert E. Lemmon
Pittsburgh Energy Technology Center
P.O. Box 10940, M/S 141-L
Pittsburgh, PA 15236-0940

Objectives

This basis of this research is to apply novel techniques from Artificial Intelligence and Expert Systems in capturing, integrating and articulating key knowledge from geology, geostatistics, and petroleum engineering to develop accurate descriptions of petroleum reservoirs. The ultimate goal is to design and implement a single powerful expert system for use by small producers and independents to efficiently exploit reservoirs.

The main challenge of the proposed research is to automate the generation of detailed reservoir descriptions honoring all the available "soft" and "hard" data that ranges from qualitative and semi-quantitative geological interpretations to numeric data obtained from cores, well tests, well logs and production statistics. In this sense, the proposed research project is truly multi-disciplinary. It involves significant amount of information exchange between researchers in geology, geostatistics, and petroleum engineering. Computer science (and artificial intelligence) provides the means to effectively acquire, integrate and automate the key expertise in the various disciplines in a reservoir characterization expert system. Additional challenges are the verification and validation of the expert system, since much of the interpretation of the experts is based on extended experience in reservoir characterization.

The overall project plan to design the system to create integrated reservoir descriptions begins by initially developing an AI-based methodology for producing large-scale reservoir descriptions generated interactively from geology and well test data. Parallel to this task is a second task that develops an AI-based methodology that uses facies-biased information to generate small-scale descriptions of reservoir properties such as permeability and porosity. The third task involves consolidation and integration of the large-scale and small-scale methodologies to produce reservoir descriptions honoring all the available data. The final task will be technology transfer. With this plan, we have carefully allocated and sequenced the activities involved in each of the tasks to promote concurrent progress towards the research objectives. Moreover, the project duties are divided among the faculty member participants. Graduate students will work in teams with faculty members.

The results of the integration are not merely limited to obtaining better characterizations of individual reservoirs. They have the potential to significantly impact and advance the discipline of reservoir characterization itself.

Summary of Technical Progress

1. Decomposition of System

We have decomposed the overall system development into smaller component parts to allow us to focus on the expert knowledge required for that component. In addition, the decomposition will facilitate the implementation of the system and its validation and verification. The three component systems will be representative of how each of the experts in geology, geostatistics, and engineering characterizes the reservoir. Figure 1 describes a model for this breakdown. The concurrent development of these component systems fits into the development of the large and small scale aspects of the system as originally stated in the proposal. In Figure 1, each component system in the model is depicted as interfacing (through the bi-directional links) with a central repository of reservoir descriptions. Though, portions of these description will essentially be passed from component to component as more information is gathered (as shown by the bi-directional links), the model of a central repository is an accurate account of how the components are integrated, i.e., the final descriptions in the repository are consistent will all of the information given by the components systems. This system model allows us to develop the system using an Artificial Intelligence technique called a *blackboard system,* in which information is centrally located, i.e., on a blackboard, and experts take their turn to update, change, and correct the information on the blackboard.

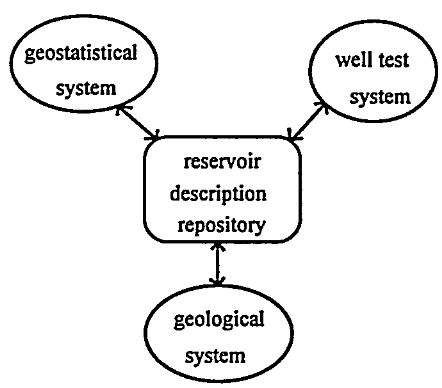


Figure 1: Expert System Decomposition

The geostatistical system continues to be tested and updated. We are currently converting all implemented code into C/C++ for integration. This includes both Fortran and Kappa-PC implementations of the various component systems that have been developed to this point. New work reported for this quarter includes a neural network implementation for log facies recognition that overcomes the difficulties faces by the previous system. We update the research on marker bed identification. Expert rules have been developed in more detail for correlating the sand bodies among the wells. We present these rules and the experiments to indicate their effectiveness. In addition, we have been performing wavelet transforms to determine the effect of compression to some part of the original data on the overall performance of the reservoir. The use of wavelets is believed to be needed for integrating the large-scale and small-scale data.

2. Geostatistical System

The two-scale, variogram pre-check approach was further studied during this period. This approach relies on getting an upscaled grid system which matches the "true" scale BHP's. We have been unsuccessful in this so far. Futher analysis of the effects of upscaling were therefore also carried out. We also began to investigate the concepts of local grid refinement and radius of investigation in the search for an appropriate upscaling methodology. The constant rate approach was the primary focus.

2.1 Two-Scale, Variogram Pre-Check Approach

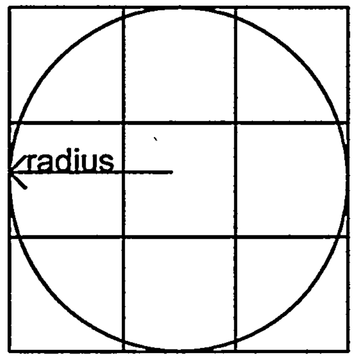
This approach, which was mentioned in the last report, uses the variogram-based change in the simulated annealing (SA) objective function value to determine whether we flow simulate. In this way, we remove unnecessary expensive flow simulations runs unless they are beneficial to our objective function. Further, by flow simulating on an upscale grid, we also reduce the execution time for the SA algorithm.

We have been experimenting with the use of different time ranges for matching the flow simulation part of the SA objective function. This was prompted by the observation that the upscaled grid BHP results do not capture the "true" scale BHP's until the start of pseudo-steady-state flow, and that even then the match is not ideal. We had also started to experiment with the weightings used for the variogram and the flow simulation parts of the objective function and the objective function tolerance, but decided to hold these constant at 0.5 and 0.05 respectively so that any variations in the results due to the time range variations may be captured.

2.2 Upscaling Analysis

We found that the errors from upscaling when we used the tensor method[1] were of the same order as those for the geometric averaging and renormalization methods. This led us to begin to consider local grid refinement and radius of investigation approaches.

With respect to local grid refinement, this modification does not seem to be a worthwhile consideration since it will allow matching with the fine-scale BHP's only up to the time when the radius of investigation is equivalent to the upscaled gridblock size.



upscaled gridblock

We are still looking at the radius of investigation considerations, but one idea is to use this concept to "fix" the near-well permeability values using the transient inflow performance relationship.

2.3 Laplace Transform Modifications

A preliminary study of the use of the Chen et al approach[2] as a way of correcting the variable rate problem mentioned in the last report seems to suggest that this methodology *may* introduce a significant "extra load" on the execution time of the algorithm. The equation for the "corrected" BHP in Laplace space is given by:

$$\bar{p}_{wD}(s)=\frac{q_{D_1}}{s_1^2\left[C_D+\frac{1}{s_1 G(s_1)}\right]}+\sum_{j=2}^{n}\frac{t_D}{t_D-t_{D_{j-1}}}\frac{q_{D_j}-q_{D_{j-1}}}{s_j^2\left[C_D+\frac{1}{s_j G(s_j)}\right]} \tag{1}$$

where

$\bar{p}_{wD}(s)$ is a Laplace space approximation for a variable rate p_{wD}

s_j is the Laplace variable based on $t_D - t_{D_{j-1}}$

C_D is the wellbore storage coefficient

t_D is dimensionless time

q_D is dimensionless rate

$G(s)$ is defined by the equation: $\bar{p}_{wD} = \frac{1}{sG(s)} = p(s)$.

The first problem is that this equation may be incorrect, so we will have to verify its accuracy before any further consideration of its usage. Nevertheless, the structure of the equation suggests that, in order to determine the variable rate Laplace space BHP for a particular value of s, we have to determine the constant rate Laplace space solutions for a number of s values equal to the number of times the rates are changed. For this reason, we have suspended further consideration of this approach and for the moment we are concentrating on the constant rate approach.

3. Geological System: Sand Body Identification

In order to analyze well log data, we solve the following two problems sequentially:

- *Well log segmentation problem*
- *Log facies identification problem*

Well log segmentation. Given a well log data file the system determines the endpoints of every sand body present in the log file. This is needed to divide the well log (gamma ray) into discrete stratigraphic units. Such segmentation is for log facies identification and well-to-well correlation. A rule based system is applied to the original data file to determine the cuts or segments. The resulting file is then fed to the system in charge of solving the log facies identification problem.

Log facies identification. Given a well log data file and the predetermined cuts, the system determines which kind of facie or sand body is between any two cuts. A neural network is used to solve this problem. The input to the network is an intermediate file generated by the rule-based system.

Our neural network was previously trained with expert-classified well logs to recognize the following set of fundamental shapes:

bell, funnel, blocky, symmetrical, linear

The log files used for training must undergo a scaling and normalization process to compensate for well-to-well variations in gamma ray values as well as variations in thickness. Also, the same shape may have different sizes which, again, requires some kind of normalization.

3.1 Advantages of Neural Networks

Neural networks are particularly well suited to solve the sand body recognition problem for several reasons. Among its most important features are:

- Human -brain-like processing
- Abilily to recognize patterns for systems such as:
    1. pattern recognition systems
    2. handwriting recognition systems
    3. classification systems

Models for neurons and neural networks are available extensively in related literature as well as several techniques to train them which have been thoroughly tested.

3.2 Neural Network Architecture

Despite all the knowledge available on neural networks, no standard procedure exists to determine an optimal network configuration to solve a given problem. Such procedures exist only for the most simple kind of problems which are known as linearly-separable problems requiring very simple structures (single layer configurations).

However some guidelines exist to solve more complex problems (non linear-separable problems) and they require multiple layer structures. Tests must be made to ensure that a network structure will be capable to solve the problem up to some level of confidence.

In order to solve the log facies identification problem the following three-layer structure showed to give acceptable results:

- Input Layer (16 neurons): every shape is discretized so it will consist of 16 points. Each one of these points is fed to a neuron in this layer. In consequence, the network is looking at a whole shape at any given time.
- Hidden Layer (100 neurons): this layer is used by the network to store what is known as the internal representation.
- Output Layer (5 neurons): this layer is the output of the network which is designed and trained to recognized five fundamental shapes. Each neuron in this layer correspond to a different shape (bell, funnel, blocky, symmetrical and linear)

The neural network is fully connected, that is every neuron in a layer is connected to every neuron in the following layer.

3.2.1 Supervised learning algorithm

A supervised training algorithm is applied to train the neural network. The algorithm is known as *error-back propagation algorithm.* The output of the network is compared against that of an expert and errors at the output layer are computed. Errors are then propagated from the output layer to the input layer and weights are updated. It is an iterative process. The network is considered to be trained when the maximum number of iterations is exceeded or when the error is below some value. The algorithm is based on the well known gradient-descent technique and it tries to minimize a quadratic error measure.

3.3 Well Log Segmentation

Well logs have to be scaled and normalized in order to set a common ground on which the problem can be solved. In consequence every log file is scaled in such a way that:

- *maximum gamma ray value maps to 1*
- *minimum gamma ray value maps to 0*

As a result of this process all the gamma ray values will be within this range (0-1). This is done before attempting to solve either the log segmentation or the facies identification problems.

To segment the well log, we use a rule based system that is structured as a two-step system:

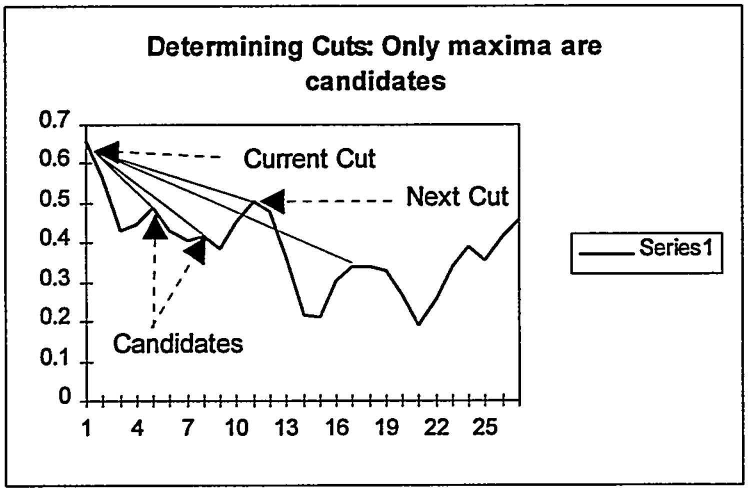
1. *Identify local maxima*
2. *Apply rules to obtain breaks*

The following rules are used by the system:

- Rule 1: a local maximum is a candidate break. This rule requires local maxima to be previously identified and is the reason for step 1 in our system.
- Rule 2: the local maximum following a candidate must not be seen from the previous identified break. By *not seen* we mean that the *candidate* is below the straight line joining the *previous identified break* and *the local maximum following the candidate*.

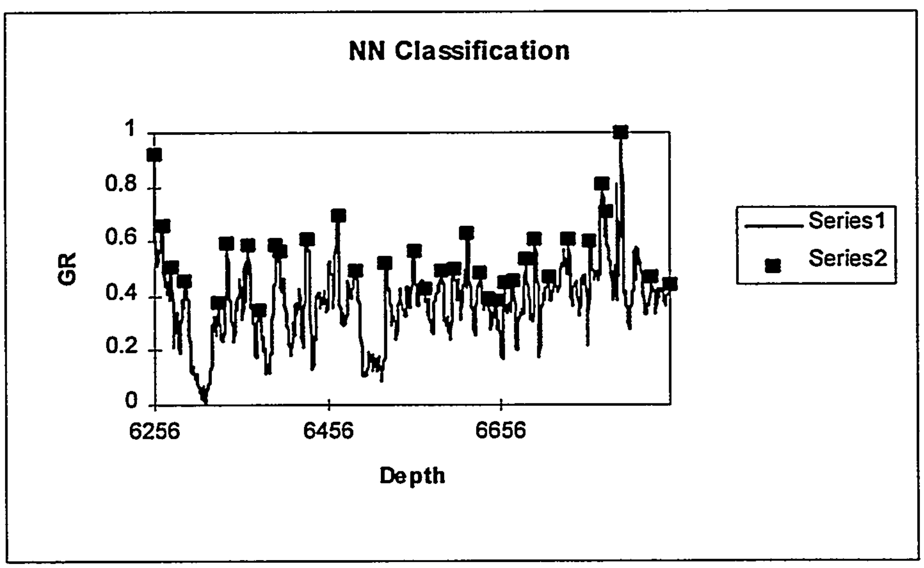
3.3.1 Applying Well Log Segmentation Rules

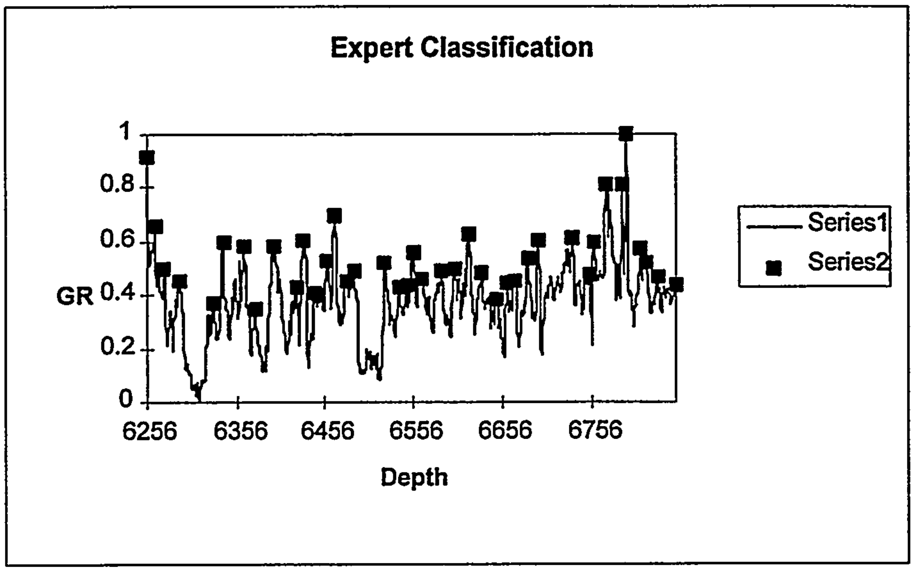
This figure shows how this rules are applied to a section of a log:



3.3.2 Segmentation Results

Figures comparing the neural network performance against expert classification for one testing log (gamma ray) is presented here. Squares are cuts.





The output of this system is then applied to the neural network.

3.4 Log Facies Identification

A neural network is also used to identify the fundamental shapes of sand bodies. The log facies must be normalized and discretized scaled so they can be fed to the network.

- Scale/normalize individual segments: a facie occurs between two consecutive breaks. All points between two consecutive breaks are scaled so:
  1. maximum value maps to 1
  2. minimum value maps to 0
- Scale/discretize depth values for each facie: the original data is linearly interpolated at 16 equally-spaced depth values and discretized accordingly. Points 1 and 16 are breaks.

3.4.1 Identification Results

Table 3.1 illustrates the interpretations obtained from the neural network as compared with the human expert for identifying the well log segments or cuts.

	Breaks (NN)	Breaks (Expert)	Percentage difference	Matching Breaks	Matching Breaks (%)
well6l	34	40	15.0	29	72.5
well6u	31	29	6.9	22	75.9
well7l	32	39	17.9	24	61.5
well7u	27	23	17.4	17	73.9
well8l	39	43	9.3	29	67.4
well8u	32	30	6.7	22	73.3
Total	**195**	**204**	**4.4**	**143**	**70.1**

Table 3.1

3.5 Neural Network Training and Testing

To identify the sand bodies, the neural network was trained using 11 different well logs with a 97.2% recall rate or correct classification when compared to that of the expert. The performance is described in Table 3.2 below.

Facies	Actua l	Identifi ed	Percenta ge
Bell	49	49	100.0
Funnel	54	54	100.0
Blocky	81	79	97.5
Symmetrical	80	80	100.0
Linear	25	19	76.0
Total	**289**	**281**	**97.2**

4. Geological System: Marker Bed Identification

In order to identify log facies and correlate stratigraphic units within wells across a reservoir, a reference point must be available that identifies the interval of observation for such analysis to take place. Marker bed identification provides the beginning and ending interval depths for this analysis.

A marker bed is a specific unit of formation that is widely distributed and laterally stable across an area. Marker beds can be traced universally between different continents, regionally across a whole basin, locally in a field-scale area, or for a very limited area of interest of some formation interval. The scope of study extends to marker bed identification in a local field-scale area.

The focus of our study is to identify the main marker beds that are common across all the wells within a field and their corresponding beginning and ending depths. While a marker bed can be easily identified in the logging curves by the naked eye of an expert, the prototypical expert system requires the processes, rules, and experiences be captured in order to arrive at the same conclusions as the expert.

4.1 Background

Different approaches have been attempted in order to identify the main marker beds in a field. The first approach involved applying a set of heuristics that characterized marker beds to the gamma ray, resistivity, and sp logs of wells in a field. This approach identified a set of potential marker beds which included the real marker beds in each well of the field. The second approach involved applying a cross-correlation algorithm to the gamma ray or resistivity log of one well with the gamma ray or resistivity log of the potential marker bed of another well. Positions of high correlation between these two logs should show areas where the real marker beds lie across these two wells. The cross-correlation algorithm did reduce the search of main marker beds from a series of wells by identifying the potential marker beds within a single well. However, the overall results showed that it does not guarantee main marker bed identification across a set of wells in all cases. More details of this research and the results are available in the last quarterly reports.

Because the approaches described above for identifying the main marker beds across a set of wells were not sufficient, we have devised an alternate approach using a sum-difference algorithm. We describe this work and the results further.

$$\frac{\sum_{i=0}^{n} (L_i - TL_i)}{\sum_{i=0}^{n} (TL_i)}$$

where

L =	gamma log for well under study
TL =	type log for field under study
n =	number of feet of marker bed in type log

Figure 4.1 Sum-Difference Equation

4.2 The Sum-Difference Approach

One of the characteristics of main marker beds is that they have high gamma ray values and low resistivity values. The sum-difference algorithm exploits these characteristics by identifying areas in a log where such characteristics exist. The areas identified by the algorithm are then interpreted as being marker beds for that particular well.

A type log is provided to the algorithm that generalizes the marker bed characteristics for that particular field. Two type logs are required. Open-hole wells and cased-hole wells often have different gamma ray characteristics. As a result, one type log generalizes the marker bed characteristics for open-hole wells, and the second generalizes the marker beds characteristics for cased-hole wells for the field under study.

The type logs are then used along with the scaled gamma ray log for a particular well to find marker bed areas in that well. For example, if the type log indicates that a marker bed for the wells in the field under study are 3 feet in depth on average, the sum-difference algorithm calculates a ratio for every 3 feet increment of the gamma ray log for a particular well based on the sum-difference equation shown in Figure 4.1. The sum-difference equation calculates a ratio by comparing the gamma ray characteristics of the well log to that of the type log for the 3 feet interval. A ratio of zero indicates that the area under scrutiny (also referred to as the match position) in the well log is an identical match with the marker bed of the type log. An exact match is indicated by match position 67 in Figure 4.2. As a result, a ratio close to zero points to a marker bed in the well under study.

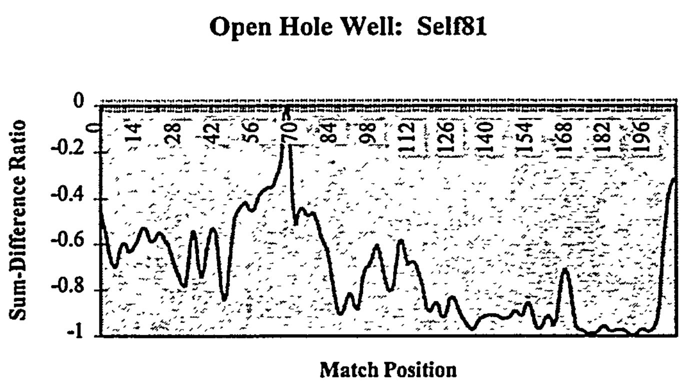


Figure 4.2

4.3 Results

The sum-difference algorithm is first applied to the open-hole type log for the field under study. For those wells whose marker beds cannot be identified by the open-hole type log, the sum-difference algorithm is applied once again to the cased-hole type log for that field. A marker bed for a well is identified if the ratio at a particular interval falls within the designated threshold specified by the user prior to the commencement of the testing phase.

The sum-difference algorithm was applied to fifteen wells dispersed in the Glenpool field. Nine of these wells were open-hole wells, and the other 6 were cased-hole wells. Marker beds of nine out of the fifteen wells (seven open-hole and two cased-hole) were identified successfully. Examples of the results of an open hole as well as a cased hole well is shown in Figures 4.3 and 4.4. Marker beds in these graphs lie at the match positions whose ratios approach zero.

Open Hole Well: Self56

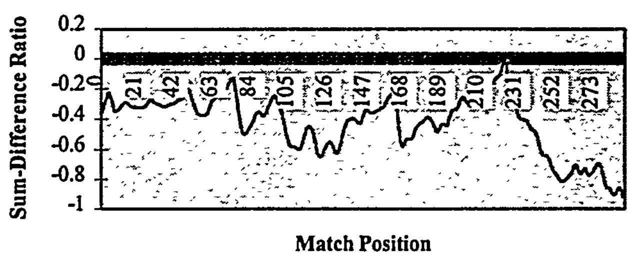


Figure 4.3

Cased Hole Log: Self69

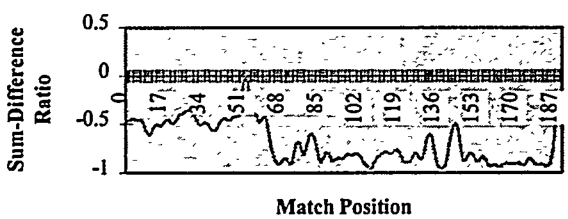


Figure 4.4

Four out of the fifteen wells tested (two cased-hole and two open-hole) indicated marker beds in the wells where none existed. Finally, the sum-difference algorithm was not successful at identifying marker beds in the two remaining wells (both cased-hole). For these six wells in which marker beds were not clearly identified, it was decided that heuristics would be constructed to isolate the real marker beds. These heuristics are discussed further in the upcoming section. It was also suggested that upon application of the heuristics, those wells in which marker beds could not be clearly identified would be returned to the user. The user would then select marker beds for a well from the potential marker beds based on knowledge of the reservoir under study.

4.4 Conclusions

One of the heuristics under study is to apply the algorithm to the resistivity logs of all wells under study. By finding common marker beds identified by both the resistivity logs as well as the gamma ray logs, we narrow the number of potential marker beds while increasing confidence that the marker beds identified are indeed real ones for that particular well. By constructing and testing similar heuristics, we hope to refine the sum-difference approach.

The results of the sum-difference approach indicate that it alone will not suffice in identifying the main marker beds. As with the cross correlation approach mentioned earlier, our approach does reduce the search of main marker beds from a series of wells by identifying the potential marker beds within a single well. While our approach will identify potential marker beds within a single well, it still needs to be modified to identify the main marker beds across a set of wells. We are currently considering extensions to this approach as well as other alternatives.

5. Geological System Components: Correlation Program

Well-to-well correlation is a complicated process that involves several aspects of depositional units. The most important ones are the position with respect to the marker beds, log facies shape and thickness of the beds being compared. However, certain factors are considered more important than others to a geology expert. We discussed this problem with the associated geology experts and used their opinion about the relative weights that can be assigned to various factors. The weighting scheme provided by the geologist is shown in Table 5.1.

factors	same	close	different
mbed_dist	3	2	1
thickness_diff	6	5	4
log shape	3	na	2

Table 5.1

Three factors are considered in Table 5.1: (1) the distance of the bed from the marker bed (*mbed_dist*), (2) thickness difference between the two beds being compared (*thickness_diff*) and (3) log shape. The distance of the bed from the marker bed is computed by taking the difference in depth of marker bed and the depth of the mid-point of top and the bottom boundaries of the bed. Thickness difference is the absolute difference in thickness between the two beds. Log shape is the facies as identified by the facies identification program.

Information about the top and bottom boundaries and the log shape are provided by the respective programs. The weights are computed for the two units based on the above scheme. The pairs with the maximum weights are considered correlatable. It may be noted that the parameters "same", "close" and "different" in Table 4.1 are defined by certain tolerance. *mbed_dist* is considered same for two beds if the difference is less than *x*tan(theta)*, where *x* is the horizontal distance between the two wells and *theta* is the stratigraphic dip of the bed. In absence of a stratigraphic dip, a value of 5 degrees is used. At this time, there is no usable definition for "close" but we have included it in the table as it relates to the weights. Any value that is not "same" is considered "different". Thickness difference between the two beds is considered "same" if the difference is less than 4 feet. Otherwise they are considered "different". The log shape can be either "same" or "different" as well. The program has been written and tested on sample data provided by the geologist.

The sample data files for three wells are shown in Tables 5.2A - 5.2C. Distinct numbers are given to each sand body in a well for the correlation for identification. The facies types are numbered from 1 to 5 for bell, funnel, blocky, symmetrical, and linear, respectively.

Well No. 78 (Markerbed depth: 1460-1462 ft)			
sand no.	top depth (ft)	bottom depth (ft)	facies
11	1463	1469	1
12	1470	1488	1
13	1489	1493	1

Table 5.2A

Well No. BG 18-33 (Markerbed depth: 1434-1436 ft)			
sand no.	top depth (ft)	bottom depth (ft)	facies
7	1437	1461	1
8	1462	1481	3
9	1482	1515	3

Table 5.2B

Well No. self 81 (Markerbed depth: 1429-1431ft)			
sand no.	top depth (ft)	bottom depth (ft)	facies
8	1432	1440	1
9	1441	1452	1
10	1453	1462	1

Table 5.2C

The results of the program run for computing the weights are given in Tables 5.3A - 5.3F. The "c" in computed weight indicates the well to which sand bodies are being compared.

Well No.	78	BG18-33	BG18-33	BG18-33
Unit id	11	7	8	9
top depth (ft)	1463.0	1437.0	1462.0	1482.0
bottom dep. (ft)	1469.0	1461.0	1481.0	1515.0
facies	1	1	3	3
computed wgt.	c	8	7	7

Table 5.3A

Well No.	78	BG18-33	BG18-33	BG18-33
Unit id	12	7	8	9
top depth (ft)	1470.0	1437.0	1462.0	1482.0
bottom dep. (ft)	1488.0	1461.0	1481.0	1515.0
facies	1	1	3	3
computed wgt.	c	9	9	7

Table 5.3B

Well No.	78	BG18-33	BG18-33	BG18-33
Unit id	13	7	8	9
top depth (ft)	1489.0	1437.0	1462.0	1482.0
bottom dep. (ft)	1493.0	1461.0	1481.0	1515.0
facies	1	1	3	3
computed wgt.	c	8	7	7

Table 5.3C

Well No.	BG18-33	self 81	self 81	self 81
Unit id	7	8	9	10
top depth (ft)	1437.0	1432.0	1441.0	1453.0
bottom dep. (ft)	1461.0	1440.0	1452.0	1462.0
facies	1	1	1	1
computed wgt.	c	10	10	10

Table 5.3D

Well No.	BG18-33	self 81	self 81	self 81
Unit id	8	8	9	10
top depth (ft)	1462.0	1432.0	1441.0	1453.0
bottom dep. (ft)	1481.0	1440.0	1452.0	1462.0
facies	3	1	1	1
computed wgt.	c	9	9	9

Table 5.3E

Well No.	BG18-33	self 81	self 81	self 81
Unit id	9	8	9	10
top depth (ft)	1482.0	1432.0	1441.0	1453.0
bottom dep. (ft)	1515.0	1440.0	1452.0	1462.0
facies	3	1	1	1
computed wgt.	c	9	9	9

Table 5.3F

It may be noted that in some cases the correlation weights shown in the above tables may be same for different pairs of units under consideration. Under such circumstances *mbed_dist* is given priority in order to decide the actual correlation.

Based on the above results, the correlation in Table 5.4 is presented for the three wells namely well 78, well BG18-33 and self 81. The table shows the correlation results for the three wells. Each row contains results about a pair of depositional units belonging to two different wells. For example the first row indicates that unit 11 and unit 7 of wells 78 and BG18-33 are correlatable. "Topx" and "Botx" are the top and bottom depths of unit 11 of well 78 and "Topy" and "Boty" are the top and bottom depths unit 7 of well BG18-33.

Wellx	Welly	Unitxid	Unityid	Topx(ft)	Topy(ft)	Botx(ft)	Boty(ft)
78	18	11	7	1463	1437	1469	1461
78	18	12	8	1470	1462	1488	1481
78	18	13	7	1489	1437	1493	1461
18	81	7	8	1437	1432	1461	1440
18	81	8	9	1462	1441	1481	1452
18	81	9	10	1482	1453	1515	1462

Table 5.4

Table 5.4 correlation results shows discrepancy for unit 7 of well BG18-32 . Unit 7 of well BG18-32 is correlatable with both unit 11 and 13 of well 78 as per the computed weights. This kind of anomaly occurs due to the rigid assignment of weights to several factors. In practice the geologists change the weights assigned to various factors such as distance from marker beds, log facies, thickness difference etc. This issue needs further consideration by geology experts and is, in fact, the biggest challenge in automating the correlation process.

6. Well-Test Interpretation

Transient testing is a major source of vital information about reservoir parameters like permeability, reservoir pressure, wellbore conditions, reservoir discontinuities and other information that is essential for reservoir studies. The accuracy of these properties estimated from well tests depends on prior identification of a model that describes the reservoir accurately. This model is known as the 'well test interpretation model'.

The procedure for finding an appropriate model can be quite complex. It usually resides deep in the expert's mind. The failure of mathematical models to solve the problem can be attributed to the nature of the problem itself since the procedure is not completely quantitative and relies on experience. Computerized well testing involves interpreting the various forms of the time vs. pressure data and other well data and to determine the well model and calculate the various parameters. Basically this can be grouped into two parts: the qualitative analysis, and the quantitative analysis. Qualitative analysis deals with selecting the appropriate well model from the input data. Quantitative analysis involves calculating various well parameters like permeability, skin factor (if applicable) etc. Estimation of well properties depends upon the selection of the right model and hence qualitative analysis is a crucial part of the problem.

The expert system component developed for the well test interpretation consists of rules and facts for buildup test analysis gathered from an expert in this field. These carefully extracted rules simulate the reasoning process used by an expert to identify the appropriate interpretation model for a well test. In this approach, the system is designed to use a description of the shapes of the derivative plot. When needed, the system seeks information in addition to test data from relevant sources such as known reservoir and fluid properties, production statistics, well logs and geological data.

This section presents the well test interpretation package. The different modules that make up the system, and the steps to use the system are discussed in the following sections.

6.1 System Overview

The well test interpretation system can be divided into two main modules. The model identification model deals with analyzing the input data to characterize the well. The other module is the parameter estimation model that used the model information to calculate the parameters required. Figure 6.1 shows the components in the system.

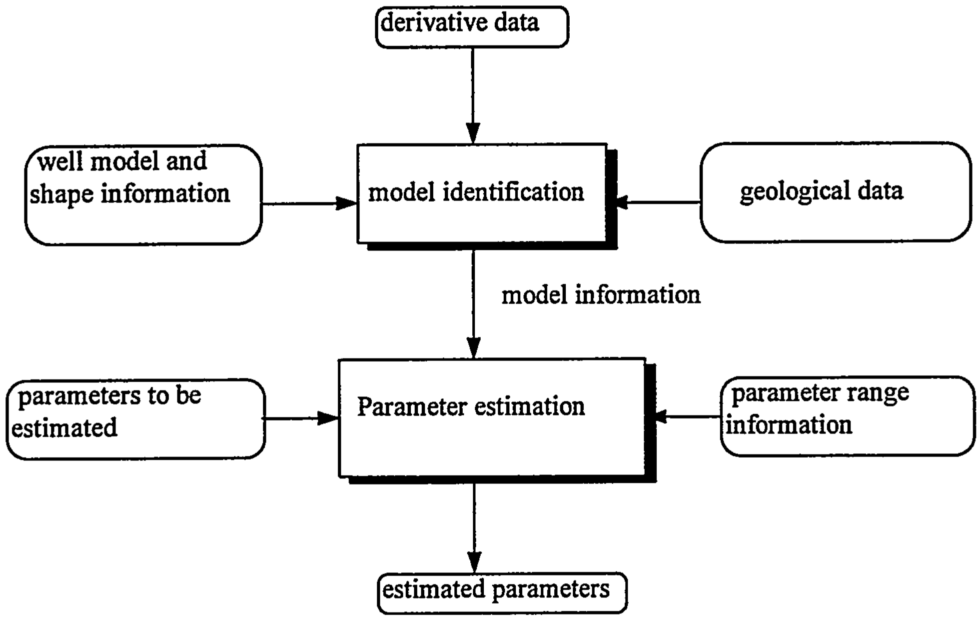


Figure 6.1 : System modules

The input to the system is the time and pressure derivative. This information is provided to the system in an input file. The model identification module analyses this input and creates a description of the shape of the plot in terms of primitives. This shape is then compared with the available model and shape information to select a model. In case of multiple choices, the selection is narrowed down using geological information provided by the user. Once the module is able to classify the input, the model information is then used to estimate the parameters. The parameters that are to be estimated is specified by the user. The user may also specify the range of values that each parameter can have. This helps in making the estimation process more efficient. The output from the parameter estimation module is the final output required from the system. The following section describes each module in greater detail.

6.2 System components

The following are the main modules used in the system.

6.2.1 Model identification

The model identification process is made up of two distinct steps. These are: creation of the internal representation of the plot using symbols, and matching this representation with the shape information stored in the system. Each of these steps are discussed below.

The input data is analyzed and converted into an internal representation for further analysis. This provides a simple and efficient method of matching shapes. This representation is done in terms of the following symbols:

up, down, flat, maximum, minimum, plateau, valley

A flat segment is a segment with slope smaller than 0.1 in absolute value. An *up* segment is one which has slope greater than 0.1. *Plateau, valley, maximum and minimum* are determined from the primitives *up, down* and *flat*. This convention is similar to that of Startzman and Kuo[4] who first observed the usefulness of the symbolic representation of log data.

The algorithm begins by calculating the slopes between the data points. This data is stored in a list. Then the algorithm proceeds by scanning this list and replacing each slope with a symbolic representation. Heuristic rules are applied to create the correct representation. The algorithm does a second scan through the list to come up with a final representation of the whole plot. Here the algorithm uses rules which have been developed to eliminate redundant symbols, or identify new ones based on the primitive symbols (*up, down, flat*). Small disturbances due to noise which slightly distort the plot is also eliminated here. Typical rules are as follows:

- *up* followed by an *up* is *up*
- *down* followed by *down* is *down*
- *up* followed by *flat* followed by *down* is a *maximum* if the number of *flats* in between is sufficiently small, otherwise it is a *plateau.*

Using such rules, the algorithm produces a final list of representative symbols which describe the whole plot. Though it depends upon the particular data used, usually four to five symbols describe one complete graph. Consecutive identical symbols (e.g. *up, up*) are compressed into a single symbol representing several segments.

An interpretation model is usually obtained by combining several components which produce the observed shapes on different parts of the derivative. The expert system is provided with the models and a representation of their derivative.

The initial implementation includes the description of two models which are:

1. Finite wellbore radius well; Infinite isotropic reservoir; Single porosity system.
2. Fully penetrating vertically fractured well; Anisotropic Single porosity system.
3. Finite wellbore radius well; Infinite isotropic reservoir; Dual porosity system.
4. Fully penetrating vertically fractured well; Anisotropic Dual porosity system.
5. Fully penetrating line source well; Anisotropic Dual porosity system

The characteristic derivative shapes associated with these models were determined and are as follows:

1. model1: [*up, maximum, down, flat*]
2. model2: [*up, flat*]
3. model3: [*up*]

Model3 is an alternate representation of model2. All the above models are a representation of the plots that show the time vs. derivative curve.

The matching algorithm is designed along the lines of the work done by Allain and Horne[3]. The aim is to find the model that qualitatively matches the data. The algorithm starts with selecting all the models that match with the input data at the first symbol. It then proceeds to the next symbol and eliminates all the models which do not match with the input symbols at any stage. Finally at the end of the process, the algorithm ends up with the correct model, or no model in case there is no match. In certain cases, the system cannot narrow down the search to a single model. This may be due to the nature of the input where it matches more than one model. The geological information may be used to proceed further.

Our matching algorithm has been modified to work correctly with the symbolic representation of models in our system. A particular input data matches a model if its sequence of symbols completely match with those of a particular model. Since the model information is complete and has not been broken down to different regimes, a complete match of all the symbols of the input data with those of a model is considered. This approach is simpler than considering the different regimes (as done by Allain and Horne[3]), and works well when complete model information has been coded in the system.

The fact that one can hardly select an appropriate interpretation model from the pressure transient data alone emphasizes the need for incorporation of external data into well test analysis. This information can help in narrowing down the choice of models. The system presently uses 5 parameters to make its decisions. The parameters and their possible values are shown in Table 6.1.

Parameter	Possible values
Geometry	Vertically Radial Well (VRW) Horizontal Well (HW) Vertically Fractured Well (VFW) Radially Heterogeneous Well (RHW) Layered System (LS)
Penetration	Fully Penetrating Partially Penetrating
Porosity	Single porosity Dual porosity
Conductivity	Finite Conductivity Infinite Conductivity
Group	This parameter can be any of the possible values in the above groups or any of their possible combinations. Eg: Vertically Radial Well-Fully Penetrating

Table 6.1

The parameter *Group* is used by the system to select the set of interpretation models. All the other parameters are taken as input from the user. Information about *Geometry* and *Penetration* is usually available. *Porosity* and *Conductivity* on the other hand may not always be known.

6.2.2 Parameter estimation

Parameter estimation is carried out after a model has been selected by the system. Though they are not intended to be the final result of the interpretation, these estimates constitute the starting point for an automated type curve matching analysis. The system was thus extended to include a very powerful parameter estimation algorithm. Figure 6.2 shows the parameter estimation module in relation to the other modules in the system:

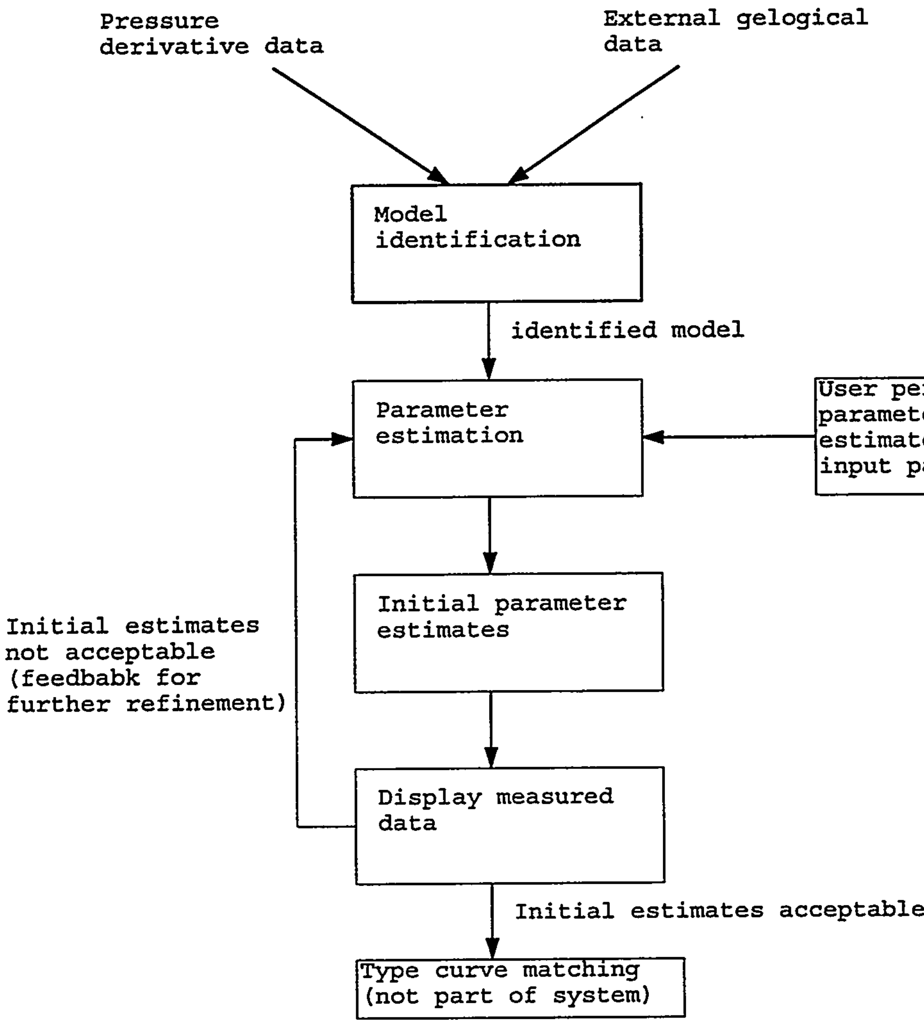


Figure. 6.2: Parameter estimation module

Presently, the system carries out the estimation process only after a single model is identified. In case multiple models are selected, the user needs to narrow down the choice to only one model. The user then selects the parameters that are to be estimated. The user is provided with an interface for this purpose. Based on the model identified and the parameters specified, the system builds up an input to the estimation program. The estimation module is then executed. The output from this module is obtained in an output file which gives the values of the required parameters.

The parameter estimation module also produces the calculated pressure and derivative values based on the model and the parameter values. This data is displayed to the user and helps in judging the correctness and the accuracy of the estimated values. The user may decide to accept these values, or run the estimation procedure again. Refining the estimated values may require changing certain parameters used by the estimation program.

6.3 Using the system

The system provides the user with an interface that makes it easy to provide input and analyze the results. Using the well test interpretation system is described using the following example. The interpretation process is started by the user by clicking on the 'Interpret' button. The user is then prompted for the input file name. This file is loaded by the system and the plot is displayed to the user. The system then matches the plot to the stored model information and prints out the model identified. As shown in Figures 6.3, 6.4, and 6.5, the system comes up with two candidate models. In this case, the plot information was not enough to select a single model.

Insert figures here

A single model needs to be selected before the parameter estimation algorithm can be executed. The user can provide the model to the system using the *Enter model* button. External information can also be to possibly narrow down the search. This information is entered into the system using the *External Data* button. In this case, the knowledge that this is a fractured well was is entered by choosing the corresponding entry in the *Geometry* field in external data entry screen.

Running the interpretation after providing this information resulted in the system discarding the first option based on the external data, and the second model was selected (Figure. 6.5).

The parameter estimation module is executed using the *Estimate Parameters* button. The options that need to be specified, like the parameters to be estimated, and their range, is entered by clicking on the *Set Parameters* button. This invokes a screen to enter these values.

6.4 Further extensions

The program as shown in the previous section was coded in KAPPA. For greater efficiency and scalability, this is being converted to C++ code. Though the internal representation and some of the interfaces have been enhanced for better performance, the basic functionality provided by the system remains the same. The following extensions are required in the system:

1. Extending the model information in the system: The system needs to be extended to handle all possible well models. This process does not require any major change to the system. Only the shape information for new models need to be provided. Entering the shape information into the system is all that is required to extend the system. This can be done by the following steps:
2. Eliminating noise: The system needs to be extended to handle noisy data. This can be achieved by simply adding a smoothing routine to smooth the plot before analyzing it.

6.5 Conclusion

The well test interpretation system is an efficient tool to analyze well data. The system starts with the minimum information available (derivative data), and tries to classify the well. Other information is used if required. Once the well is classified, the system can then be used to estimate the parameters for that well. This system has been developed as a stand-alone tool to be used by an expert. It frees the user of the routine and tedious task of classifying the wells and estimation of parameters. The information obtained from the system is also useful, and can be used by other groups in the project.

7. Wavelet Transform Base Simulation

Wavelet transformations are being used to take cross sectional data of the well site and reduce the number of data points it in such as manner that analysis computations can be performed more quickly and without loss of critical information. This section presents the results of a single phase flow simulation where the permeability distribution has been transformed into frequency domain space using a wavelet transform, compressed with different compression factors, and back-transformed into its original space using the inverse wavelet transform. The objective of the flow simulation is to see the effect of compression to some part of the original data on the overall performance of the reservoir. Early results were obtained on a Sun workstation using MatLab in conjunction with a public domain wavelet package, WaveLab. All testing was done with the Coiflet-2 wavelet of order 3. A MatLab script has been written to run the 3-D wavelet transformation and inverse transformation on a 2-D matrix of data. Specific results were optimistic. After more data is analyzed, different wavelets functions can be tried and the order of the wavelet can varied.

Several compression factors have been used in this process, i.e., 25%, 50%, 60%, 75%, 80%, 85%, and 95%. Each compression factor signifies the amount of data removed from the frequency domain. For example the 95% compression factor means 95% of the data has been removed and only the remaining 5% were kept. Intuitively, the smaller the compression factor, the closer it would be to the character of the original distribution. In all cases the overall features of the original distribution are captured by the transformed data even though it is quite clear that as the compression factor increases the distribution becomes smoother. The flow simulation is then required to see whether this transformation and compression process did or did not change the flow performance.

A single phase flow simulation model for this purpose has been developed using 90x90 grid blocks. Six wells are defined in the model. The difference between the initial pressure and the bottom hole pressure (the pressure drop) of each well for each compression factor is calculated. Then, the relative error of this pressure drop as a function of time, as defined in the following equation, is calculated for each case.

$$Rel.\,Err = \frac{\left(P_{in} - P_{wf}\right)_{original\ distribution} - \left(P_{in} - P_{wf}\right)_{compressed\ distribution}}{\left(P_{in} - P_{wf}\right)_{original\ distribution}} x100\%$$

Comparisons of the relative errors for 25%, 50%, 75%, and 95% compression factors are presented in Figures 7.1A - 7.1D.

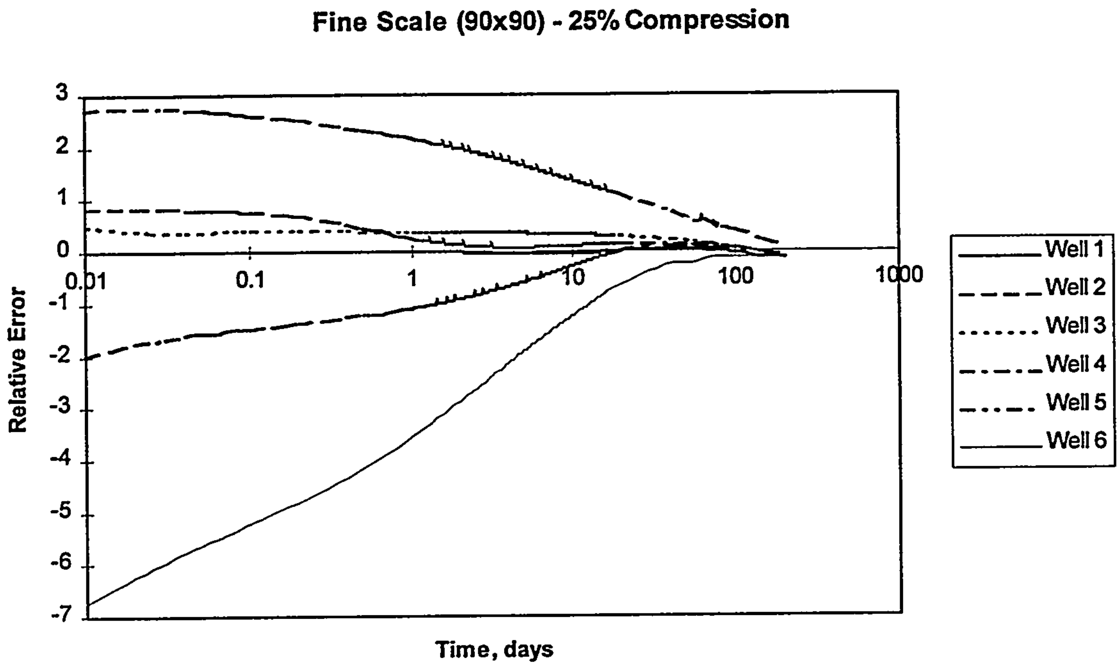


Figure 7.1A

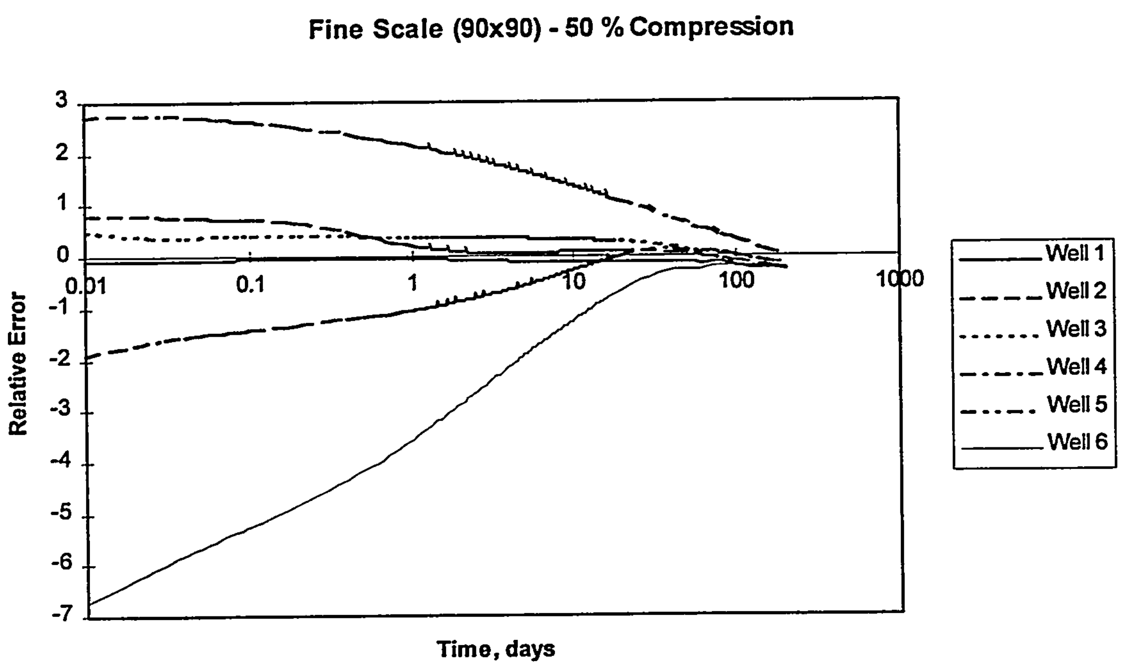


Figure 7.1B

Fine Scale (90x90) - 75% Compression

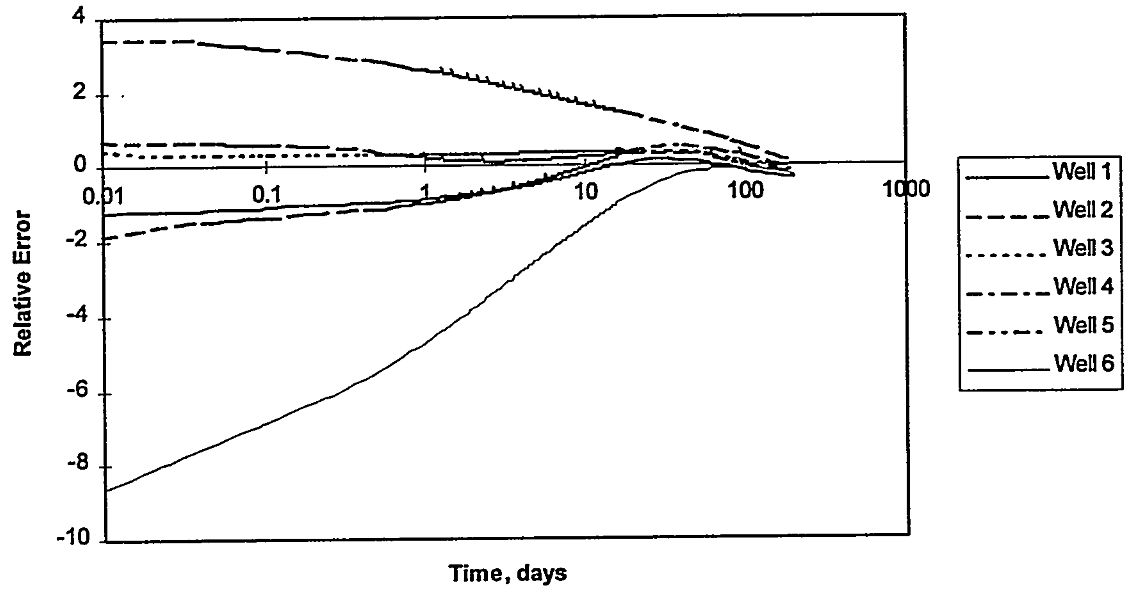


Figure 7.1C

Fine Scale (90x90) - 95% Compression

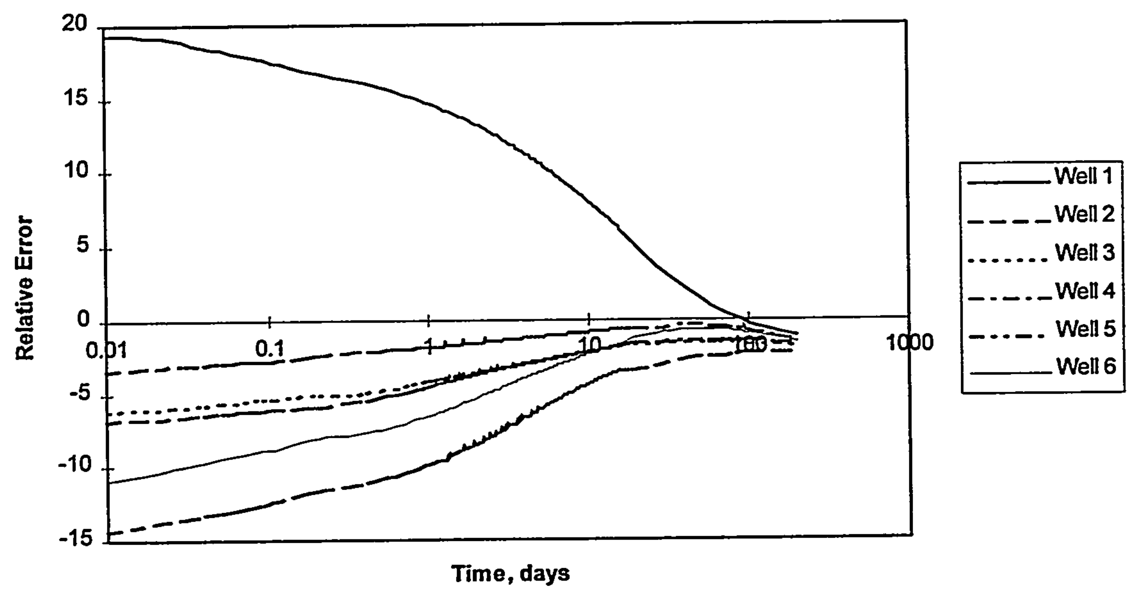


Figure 7.1D

Figures 7.1A and 7.1B present the results of 25% and 50 % compression. There is no significant difference between them. Figure 7.1C and 7.1D present the results for 75% and 95% compression factors. With 75% compression the error start to increase. This increase is clearer for the compression factor of 95%. Another aspect that can also be observed from the result of the 95% compression is the fact that at late time (close to 200 days) the simulation result does not converge to zero as in the other cases. Thus, this result indicates that there is a certain cut-off value of compression factor, which is 75% in this particular case, that needs to be used if we want to compress the original data through the wavelet transform while keeping the overall performance of the reservoir.

One problem that was encountered in using this transformation is the presence of negative permeability. This occurs when a very low value is transformed. At this time, this problem is temporarily solved by assigning a very low positive value to replace the negative number. A two phase flow model will be tested for these data in the near future.

8. References

1. Aasum, Yngve: "Effective Properties of Reservoir Simulator Grid Blocks", PhD Dissertation, The University of Tulsa, Tulsa, OK (1992)
2. Chen, Chih-Cheng and Raghavan, Rajagopal: "An Approach To Handle Discontinuities by the Stehfest Algorithm", SPE 28419 presented at the SPE 69th Annual Technical Conference and Exhibition held in New Orleans, LA, U.S.A., 25-28 September 1993
3. Allain, O. F. and Horne, R. N.: "Use of Artificial Intelligence for Model Identification and Parameter Estimation in Well Test Interpretation", paper SPE18160, 1988.
4. Startzman, R. A. and Kuo, T. B.: "A Rule-Based System for Well Log Correlation", paper SPE15295, 1986.