Development, Testing and Validation of a Neural Model to Predict Porosity and Permeability from Well Logs, Grayburg Formation, McElroy Field, West Texas

Topical Report

September 30, 2002 – June 30, 2003

Jack W. Steen
Scott R. Reeves

Advanced Resources International
9801 Westheimer, Suite 805
Houston, TX 77042

July, 2003

U.S. Department of Energy
DE-FC26-01BC15357

U.S. Department Of Energy Disclaimer

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

Advanced Resources International Disclaimer

The material in this Report is intended for general information only. Any use of this material in relation to any specific application should be based on independent examination and verification of its unrestricted applicability for such use and on a determination of suitability for the application by professionally qualified personnel. No license under any Advanced Resources International, Inc., patents or other proprietary interest is implied by the publication of this Report. Those making use of or relying upon the material assume all risks and liability arising from such use or reliance.

Executive Summary

Accurate, high-resolution, three-dimensional (3D) reservoir characterization can provide substantial benefits for effective oilfield management. By doing so, the predictive reliability of reservoir flow models, which are routinely used as the basis for investment decisions involving hundreds of millions of dollars and designed to recover millions of barrels of oil, can be significantly improved. Even a small improvement in incremental recovery for high-value assets can result in important contributions to bottom-line profitability.

Today's standard practice for developing a 3D reservoir description is to use seismic inversion techniques. These techniques make use of geostatistics and other stochastic methods to solve the inverse problem, i.e., to iteratively construct a likely geologic model and then upscale and compare its acoustic response to that actually observed in the field. This method has several inherent flaws, such as:

- The resulting models are highly non-unique; multiple equiprobable realizations are produced, meaning
- The results define a distribution of *possible* outcomes; the best they can do is quantify the uncertainty inherent in the modeling process, and
- Each realization must be run through a flow simulator and history matched to assess it's appropriateness, and therefore
- The method is labor intensive and requires significant time to complete a field study; thus it is applied to only a small percentage of oil and gas producing assets.

Since the majority of fields do not warrant these efforts (today), the result is sub-optimal development for many fields. The industry therefore needs an improved approach to facilitate enhanced hydrocarbon recovery that is quicker, more accurate, and less expensive than today's standard methods. This will allow more reservoirs to be better characterized, allowing recoveries to be optimized and significantly adding to recoverable reserves.

A new approach to achieve this objective was first examined in a Department of Energy (DOE) study performed by Advanced Resources International (ARI) in 2000/2001. The goal of that study was to evaluate whether robust relationships between data at vastly different scales of measurement could be established using virtual intelligence (VI) methods. The proposed workflow required that three specific relationships be established through use of artificial neural networks (ANN's): core-to-log, log-to-crosswell seismic, and crosswell-to-surface seismic.

The results of that initial study showed that it is, in fact, feasible to establish the three critical relationships required, and that use of data at different scales of measurement to create high-resolution reservoir characterization is possible. Based on the results of this feasibility study, in September 2001, the DOE, again through ARI, launched a subsequent two-year government-industry R&D project to further develop and demonstrate the technology. The goals of this ongoing project are to:

- Make improvements to the initial methodology by incorporating additional VI technologies (such as clustering), using core measurements in place of magnetic resonance image (MRI) logs, and streamlining the workflow, among others.

- Demonstrate the approach in an integrated manner at a single field site, and validate it via reservoir modeling or other statistical methods.

The site selected for this project was the McElroy field in the Central Basin Platform of the Permian basin of West Texas, operated by ChevronTexaco. The producing horizon is the Permian Grayburg carbonate. This report describes the results of an important component of the overall project, the creation of an artificial neural network (ANN) model to relate well log response data to whole core measurements.

Based on the results of this study, the following conclusions have been drawn:

- An accurate ANN model has been developed to predict porosity and permeability from well logs in the complex Grayburg carbonate reservoir of west Texas. Correlation coefficients between actual and predicted values were 0.86 for porosity and 0.84 for permeability for the training/testing data, and 0.85 and 0.82 respectively for the validation data.

- The model can replicate vertical reservoir property variations very well, and predict discrete porosity and permeability streaks (for subsequent use in reservoir modeling) at a high vertical resolution. The main source in predictive error was extreme reservoir property variability over very short vertical intervals – the ANN model tended to "smooth" the high variability.

- The model was based on 6,000 data points with coincident well log and core measurement information; 60% of the data was used for training, 20% for testing, and 20% for validation.

- Considerable pre-conditioning of the data was employed to achieve these encouraging results. Some of the techniques utilized included duplication of extreme values to enhance their representation in the dataset, "depth windowing" of log data to reduce tool lag effects, and log clustering (unsupervised ANN modeling) to establish facies codes and mineralogical similarities as additional model input parameters.

This study has demonstrated that core porosity and permeability can be effectively predicted from geophysical log data using ANN's to provide a high (vertical) resolution reservoir characterization for subsequent flow modeling and field development optimization.

Table of Contents

List of Figures

List of Tables

1.0 Introduction

Accurate, high-resolution, three-dimensional (3D) reservoir characterization can provide substantial benefits for effective oilfield management. By doing so, the predictive reliability of reservoir flow models, which are routinely used as the basis for investment decisions involving hundreds of millions of dollars and designed to recover millions of barrels of oil, can be significantly improved. Even a small improvement in incremental recovery for high-value assets can result in important contributions to bottom-line profitability.

This is particularly true when Secondary Oil Recovery (SOR) or Enhanced Oil Recovery (EOR) operations are planned. If injectants such as water, hydrocarbon gasses, steam, CO_2, etc. are to be used, an understanding of fluid migration paths can spell the difference between economic success and failure. In these types of projects injectant costs can be a significant part of operating expenses, and hence their optimized utility is critical.

SOR/EOR projects will increasingly take place in heterogeneous reservoirs where interwell complexity is high and difficult to understand. Although reasonable reservoir characterization information often exists at the wellbore, the only economical way to sample the interwell region is with seismic methods. Surface reflection seismic has relatively low cost per unit volume of reservoir investigated, but the resolution of surface seismic data available today, particularly in the vertical dimension, is not sufficient to produce the kind of detailed reservoir description necessary for effective SOR/EOR optimization and planning.

Today's standard practice for developing a 3D reservoir description is to use seismic inversion techniques. These techniques make use of geostatistics and other stochastic methods to solve the inverse problem, i.e., to iteratively construct a likely geologic model and then upscale and compare its acoustic response to that actually observed in the field. This method has several inherent flaws, such as:

- The resulting models are highly non-unique; multiple equiprobable realizations are produced, meaning
- The results define a distribution of *possible* outcomes; the best they can do is quantify the uncertainty inherent in the modeling process, and
- Each realization must be run through a flow simulator and history matched to assess it's appropriateness, and therefore
- The method is labor intensive and requires significant time to complete a field study; thus it is applied to only a small percentage of oil and gas producing assets.

Since the majority of fields do not warrant these efforts (today), the result is sub-optimal development for many fields. The industry therefore needs an improved approach to facilitate enhanced hydrocarbon recovery that is quicker, more accurate, and less expensive than today's standard methods. This will allow more reservoirs to be better characterized, allowing recoveries to be optimized and significantly adding to recoverable reserves.

A new approach to achieve this objective was first examined in a Department of Energy (DOE) study performed by Advanced Resources International (ARI) in 2000/2001[1]. The goal of that study was to evaluate whether robust relationships between data at vastly different scales of measurement could be established using virtual intelligence (VI) methods. The proposed workflow required that three specific relationships be established through use of artificial neural networks (ANN's): core-to-log, log-to-crosswell seismic, and crosswell-to-surface seismic (Figure 1.1). One of the key attributes of the approach, which should result in the creation of high resolution reservoir characterization with greater accuracy and with less uncertainty than today's methods, is the inclusion of borehole seismic (such as crosswell and/or vertical seismic profiling – VSP) in the data collection scheme. As shown in Figure 1.2, borehole seismic fills a critical gap in the resolution spectrum of reservoir measurements between the well log and surface seismic scales, thus establishing important constraints on characterization outcomes.

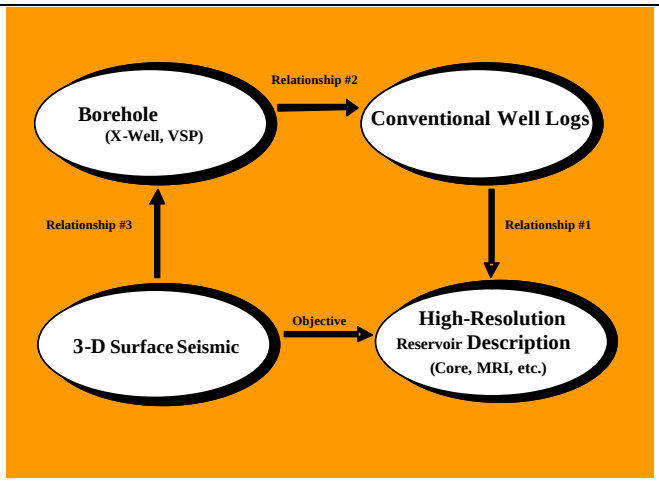

Figure 1.1: Pathway to 3D high-resolution reservoir description

The results of that initial study showed that it is, in fact, feasible to establish the three critical relationships required, and that use of data at different scales of measurement to create high-resolution reservoir characterization is possible. Based on the results of this feasibility study, in September 2001, the DOE, again through ARI, launched a subsequent two-year government-industry R&D project to further develop and demonstrate the technology.

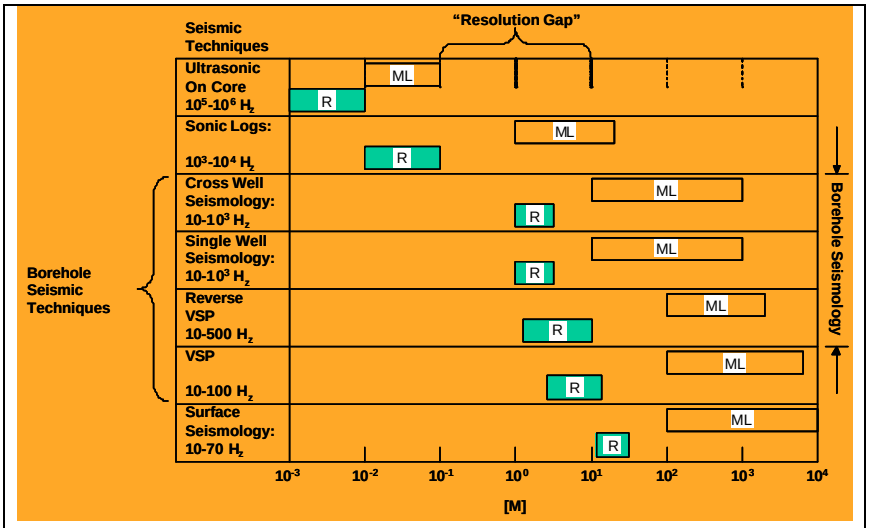

Figure 1.2: Different scales of measurement achieved by common clastic methods (after Montgomery[2])

The goals of this ongoing project are to:

- Make improvements to the initial methodology by incorporating additional VI technologies (such as clustering), using core measurements in place of magnetic resonance image (MRI) logs, and streamlining the workflow, among others.

- Demonstrate the approach in an integrated manner at a single field site, and validate it via reservoir modeling or other statistical methods.

This report describes the results of an important component of the overall project, the creation of an ANN model to relate well log response data to whole core measurements, i.e., Relationship #1 in Figure 1.1.

2.0 Field Test Site

The first step of the project was to locate a suitable test site. The merits of various potential sites, including data availability, resource size, and operator cost-share, were considered before ultimately deciding upon the McElroy field, a large oil field in the Permian Basin of west Texas, operated by ChevronTexaco (Figure 2.1).

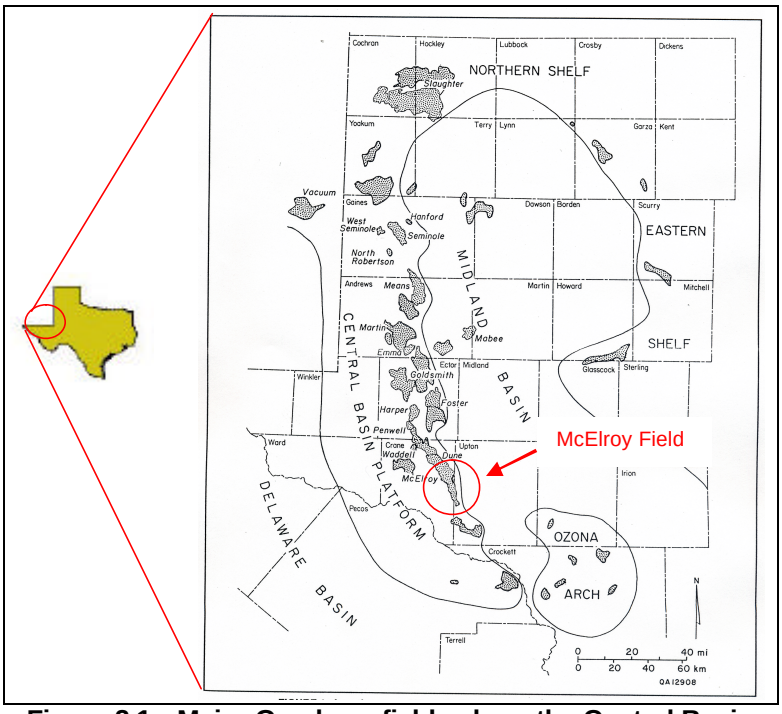

Figure 2.1: Major Grayburg fields along the Central Basin Platform trend (after Bebout & Harris[3])

McElroy was chosen because it met several important requirements:

- All data types required for the study were readily available, including surface (3D) seismic, crosswell seismic, modern logs and extensive core data.
- The operator (ChevronTexaco) was willing to work with the project team by providing access to proprietary data and technical personnel.
- Successful application of the technique could be immediately applied elsewhere in the field, as well as in nearby fields, to help boost domestic oil production.

On the other hand, the producing horizon, the Grayburg formation, is a complex carbonate reservoir. While it would have been preferred to test the new reservoir characterization approach in a less complex (e.g., clastic) reservoir, no such test site with all required information could be identified. Thus the McElroy field was selected as the test site for this study.

2.1 The Permian Basin

The Permian Basin of west Texas and southeast New Mexico is a prolific hydrocarbon province, second only to Alaska in terms of proved reserves[4]. It has produced billions of barrels of oil since the early

twentieth century and today produces 17% of the nation's crude oil and two thirds of Texas' crude oil. Annual production was approximately 340 million barrels of oil (MMBO) in 2001. Approximately 22,600 MMBO have been produced through 1998 and another 4,800 MMBO are classed as proved reserves[5].

Of note, there are 40+ CO_2 injection projects in the Permian Basin producing more than 20% of the area's total oil, or more than 140,000 barrels of oil per day (BOPD). These projects consume 1 billion cubic feet (BCF) of CO_2 daily. Recent studies indicate that there are more than 50 additional potentially economical CO_2 floodable reservoirs in the Permian Basin that represent incremental oil reserves of 500 to 1,000 MMBO[6]. Currently, the primary sources of CO_2 are natural deposits in Colorado, but an increasing amount is expected to be supplied from anthropogenic sources that would otherwise be vented to the atmosphere. By improving CO_2 project economics through technology such as that being developed in this project, this alternative CO_2 sourcing trend can be accelerated, leading to both increased domestic oil supply and reduced greenhouse gas emissions.

Geologically, the Permian Basin is divided into the Midland basin on the east and the Delaware basin on the west, separated by the Central Basin Platform (CBP), a high structural trend running roughly north-south (Figure 2.1). This broad, shallow uplift was alternately exposed to subtidal and supratidal environments as numerous sea level changes caused lowstand and highstand tracts to develop. Shallow shelf marine carbonate environments developed along the margins of the CBP and today form one of the largest oil producing complexes in the world.

The focus of early development efforts on the CBP was the Grayburg/San Andres formation. These Permian-age strata consist of a string of shallow structural and stratigraphic carbonate traps located on the margins of the CBP and extending northward onto the Northwest Shelf Margin. The map in Figure 2.1 highlights many of the major fields in this producing trend. Figure 2.2 shows a stratigraphic column in the area from surface through the Permian, with the Grayburg highlighted. The strata down through the lower Queen are interbedded siliclastics and carbonates. Below that the facies are dominated by carbonates, particularly crestal reef formations.

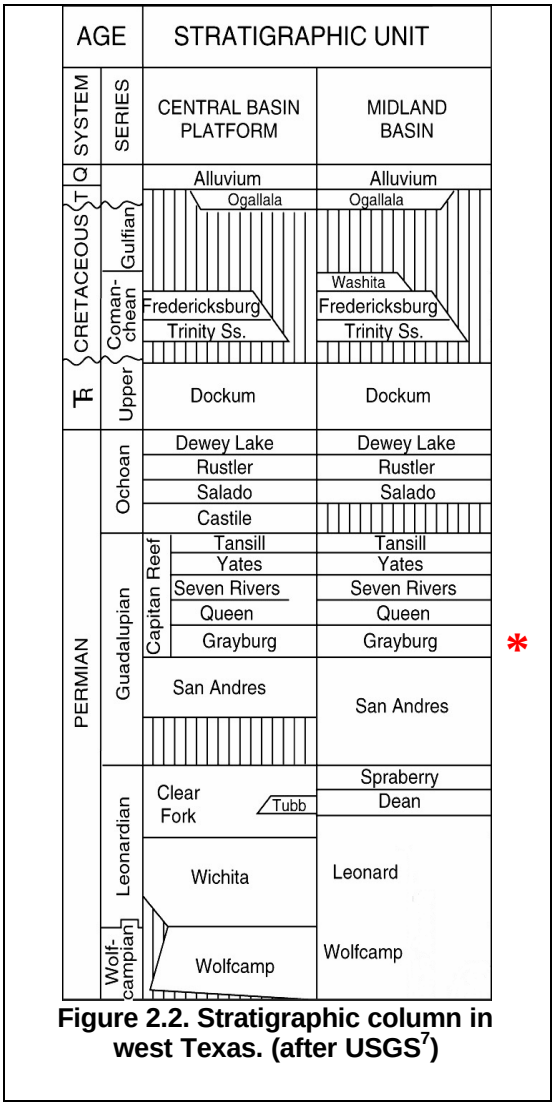

Figure 2.2. Stratigraphic column in west Texas. (after USGS[7])

2.2 The Grayburg Trend

The Grayburg/San Andres trend extends over a 105 mile long north-south fairway from the mighty Yates field in the south to Gaines and Yoakum Counties in the north. It is generally shallow, lying at an average depth of 3,780 ft., and produces light (32 degree API) low sulphur oil from rocks with an average porosity of 12 per cent and permeability in the 5-25 millidarcy (md) range. On the south CBP alone the Grayburg/San Andres complex is expected to recover approximately 2,712 MMBO of an estimated 10,286 MMBO original oil in place (OOIP)[8].

Figure 2.3 shows a type log through the Grayburg interval and indicates the various depositional environments common to the area. Figure 2.4 is a simplified depositional diagram, indicating the cyclicity evident in many of the carbonate reservoirs caused by numerous sea level changes.

The western margins of the Grayburg fields on the CBP were generally exposed to supratidal environments. Proceeding eastward the facies types pass through the reef itself and extend to fore-reef shoals and marine slope deposits as the formations spill into the Midland basin.

2.3 The McElroy Field

McElroy field still is one of the top producing fields in Texas, and is in the top 50 fields in the U.S. in terms of oil reserves. OOIP is estimated to be 2,544 MMBO of which 600 MMBO are expected to be recoverable. Since its discovery in 1926 it has made more than 532 million barrels of oil[10], and is still producing at the rate of 5.5 MMBO annually. The field contains 1,000 wells dispersed throughout a 31-square-mile area; thus nominal spacing is approximately 20 acres per well. Nominal expected ultimate recovery (EUR) per well is approximately 600,000 barrels of oil. Originally, Chevron operated a 22,400 acre area in the southern portion of the field, and Texaco operated the 11,600 acre northern portion. Since the merger of those two companies the entire 34,000 acre field has been under consolidated operatorship. Waterflooding operations began in the early 1960's [11,12].

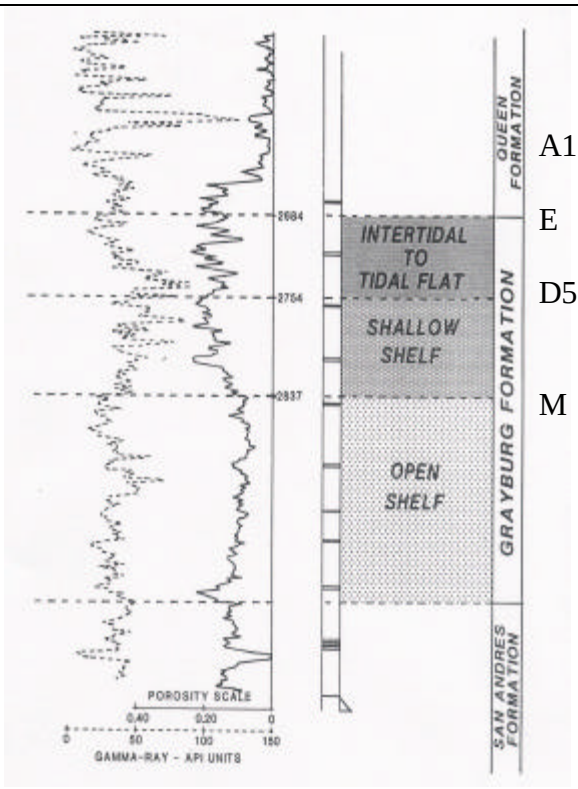

Figure 2.3: Type Log of shallow Permian Basin formations (after Harris & Walker [9])

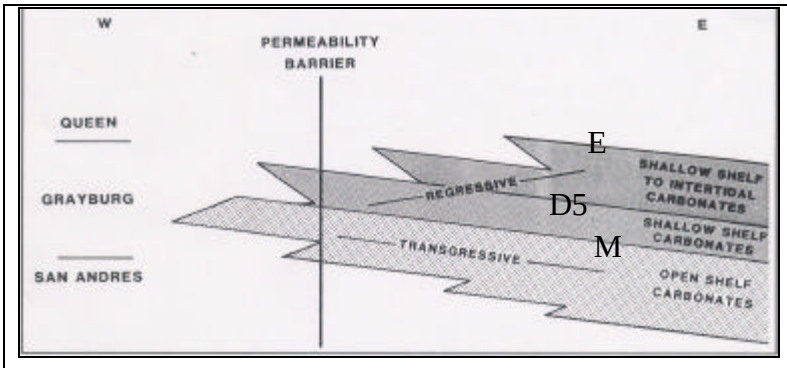

Figure 2.4: Simplified depositional history of shallow Permian Basin carbonates (after Harris & Walker[9])

Figure 2.5 shows a cross-section across the field from west to east and indicates the typical geology in the region. Note that dips are fairly shallow on the western side and increase in the fore-reef and marine environments on the east as the formation dips into the Midland basin.

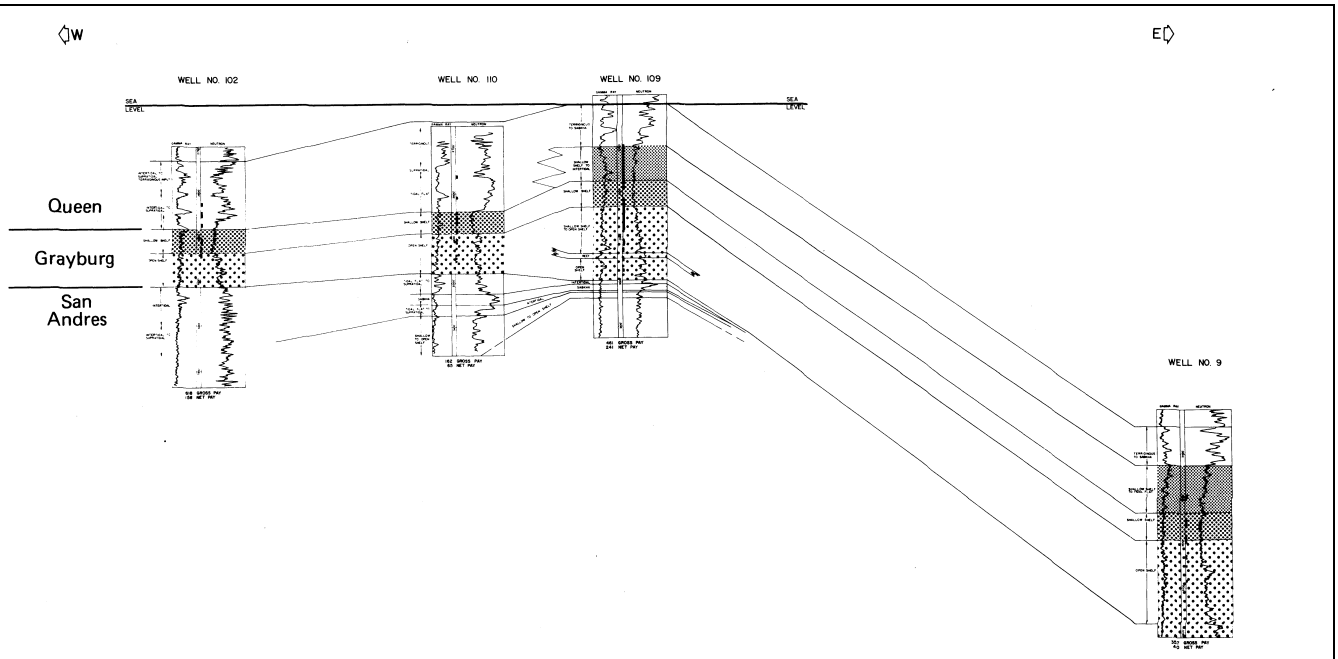

Figure 2.5: Regional W-E cross section (after Walker & Harris[3])

2.4 The Study Area

The study area for this project consists of approximately 2.5 square miles near the center of the McElroy field (Figure 2.6). This area has been the subject of intense investigation over the past ten years with substantial data being acquired. In part this is because the study area is centered on a single well pilot Light Oil Steam Flood (LOSF). The project was a pilot for the Huff-N-Puff steam flood method for enhanced recovery.

The study area is situated in a portion of the field that encompasses all the various depositional environments typical of the Grayburg formation in this region. The central area containing the LOSF and areas where cross-well seismic data was acquired is directly atop the crestal reef. To the east of this area the formation dips sharply down into the Midland basin. On the western side of the study area dips are more shallow and grade into back-reef and lagoonal facies. The square outline in Figure 2.6 denotes a "cutout" portion of a larger 3D survey acquired in 2000. This seismic "cutout" defines the extent of the study area.

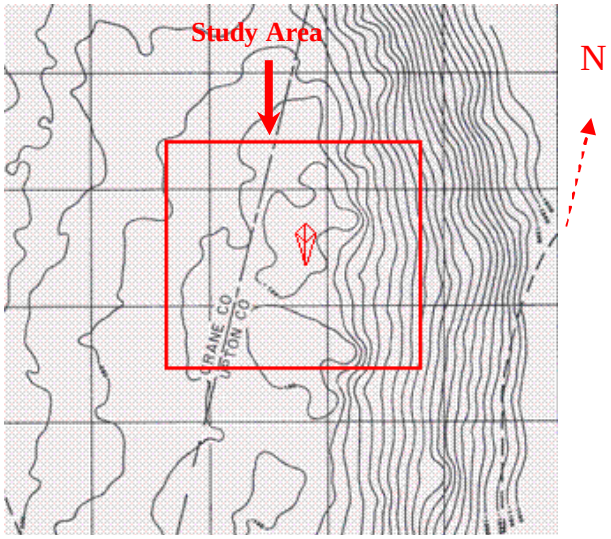

Figure 2.6: Grayburg structure near study area (after Harris & Walker[9])

The operator furnished data for 192 well locations within the survey area. The wells are generally drilled on nominal 20 acre spacing and fully penetrate the Grayburg interval. Of these, 60 are oil producers, 49 are water injectors, and 74 are either temporarily or permanently abandoned; there are also 7 observation wells. Two are classified as "other". Many of the older wells have substandard logs and log suites and were deemed unsuitable for use in this study.

Of the 192 wells, 59 within the study area are of modern vintage and have complete log suites. The surface locations of these wells are shown on Figure 2.7. Note that the wells are randomly scattered within the 3D seismic survey, but are generally situated on the western and southern sides of the survey. This is significant in light of the geologic setting. The axis of the crestal reef runs roughly north-south through the approximate center of the study area. Many of the modern vintage wells penetrate geologic strata that are in the back-reef region of the carbonate ramp complex. A comprehensive list of the wells, including well identifiers, is provided in Appendix A.

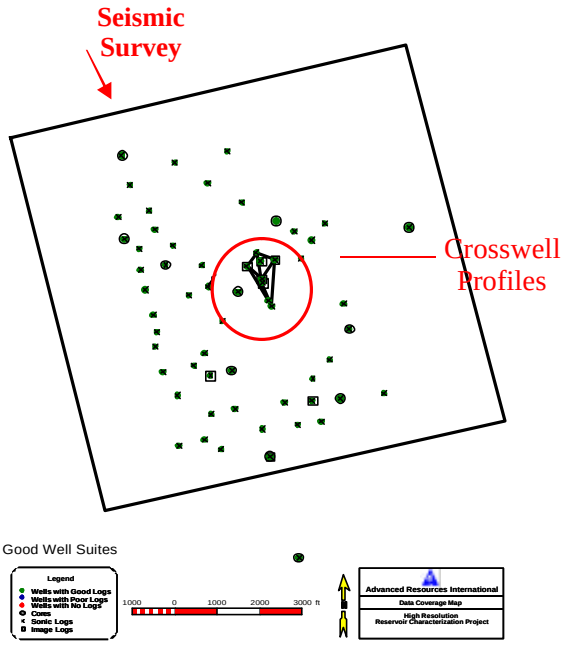

Figure 2.7: Wells within study area having complete modern log suites

2.5 Study Dataset

McElroy is a large and significant oilfield. Over the years, ChevronTexaco, the operator of the field, has collected extensive data in the study area; the data was collected by various investigators in an effort to better understand how to maximize oil recovery. Much of the data is of modern vintage, collected during the infill drilling program when the field was reduced to 20 acre well spacing, as well as during attempts to use CO_2 or steam injection enhanced oil recovery.

ChevronTexaco made available a rich dataset for the purposes of this study. Raw data were provided representing the four critical data types; surface seismic, crosswell seismic, well logs and core data. In addition, interpretive data such as formation tops and image log interpretations were provided to aid in the evaluation of the primary data. The data were delivered in digitized, electronic format. Table 2.1 is a summary of the data received for this project.

Table 2.1: Summary of Data Received for Study Area

Data Type	Amount Available
Well Locations	192
Seismic Survey, 3D P-P	2.5 sq. mi. (2000, post-steam) migrated stacked time
Crosswell Profiles	8 crosswell profile data files (1997, pre-steam)
Well Logs	Complete modern log suites for 59 wells (1984 – 2001)
Sonic Logs	84 sonic logs over survey area
Formation Tops	Interpreted formation tops (5) in 150 wells
Image Logs	8 image log files within the survey area
Core Logs	Core analysis logs for 13 cored wells in survey: approx. 325 ft. of whole core each with core porosity, saturation, and permeability measurements on ½ foot intervals.

Surface and crosswell seismic data was acquired in several programs before and after the initiation of a single –well LOSF pilot conducted in the central well of the cross-well seismic data area. This single well pilot used the Huff-N-Puff method with a period of steam injection followed by a period of oil production from the same well. In a true Huff-N-Puff project this cycle would be repeated over and over again. At McElroy only a single steam injection cycle was conducted.

Surface seismic data was acquired before and after steam injection. Only the 2000-vintage (post-steam) survey was available for this study. However, due to the small volume of steam injected and (very) limited overall effect, most of the area with coincident crosswell and surface seismic data was believed to be largely unaffected by the steam, and hence deemed suitable for correlation in this study.

The surface seismic consists of about 30,000 traces of time-migrated, stacked P-P amplitudes on 55 ft. by 55 ft. bin spacing. There are 175 inlines and 175 crosslines. Trace gathers were later obtained, but the quality is poor. The gathers were used, however, to pick first arrivals to aid in the seismic time-to-depth conversion process.

The crosswell profile data was obtained between six wells in the center of the study area, with the central well being the LOSF pilot well (Figure 2.7). Data quality varies between profiles, with the shorter profiles having the best quality. The source was piezeo-electric and the receiver array had 55 clamped geophones. The crosswell data used in this study was the set acquired before steam injection.

In terms of geophysical well logs, Table 2.2 summarizes the curves that comprise a complete log suite for the purposes of this study. Each of the 59 selected wells contain full coverage for all six log data types across the entire Grayburg interval, some 350 feet per well. A few have additional log data types, as indicated in Appendix A. These additional log data, such as the density porosity log (DPHI), were not used in this analysis, however, because they are not available for all the wells.

Table 2.2: Log Data Type Summary			
Data Type	**Abbreviation**	**Units**	**Indicator**
Gamma Ray	GR	API units	Shaliness
Compensated Neutron Log	CNL	Percent	Porosity
Bulk density Log	RHOB	gm/cc	Porosity
Sonic Travel Time	DT	usec/ft	Porosity
Photoelectric Effect	PE	barns/m2	Mineralogy
Latero-Log	LLD	Ohm-m	Resistivity

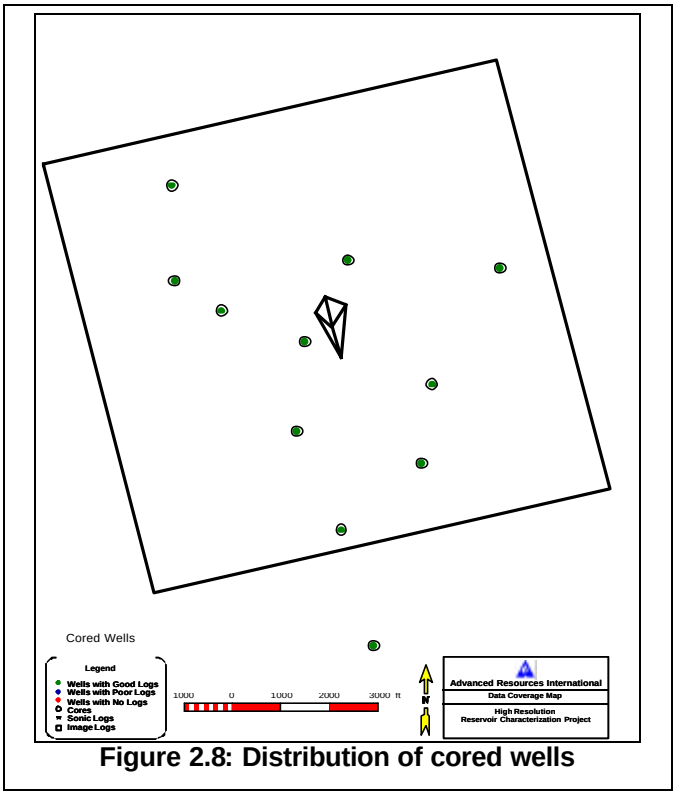

Figure 2.8: Distribution of cored wells

2.6 Whole Core Acquisition

Of particular importance to this study, in 13 wells the operator obtained whole core samples through the entire Grayburg interval. Of those, 10 had complete log suites (per Table 2.2) and were used in this study (Figure 2.8). One of the cored wells falls outside the bounds of the 3D survey, but was included in the analysis because it penetrates the same general structural and stratigraphic position as the majority of wells that do fall within the survey bounds.

Core lengths from each well range from 211 ft. to 421 ft., but on average are about 300 ft. per well. Cores were cut with conventional core barrels in water-based drilling fluid and transported to the laboratory for analysis. The cores were subjected to a full analysis (porosity, permeability, and water saturation) at one-foot intervals; data were summarized as a series of measurements versus depth, resulting in a core "log" that can be directly compared with conventional wireline log data. This data as originally received had measurements at half-foot intervals to put it on the same sampling frequency as the log data. This was achieved through linear interpolation between actual measurements. Analysis of the engineering model using both original data and data in which the interpolated values had been removed showed virtually no difference in ultimate results. Therefore it was decided to use the original half-foot data for this study, since it is consistent with the log data. Table 2.3 lists the important measured core parameters and their acronyms.

Table 2.3: Core Parameter Data		
Core Parameters	**Acronym**	**Units**
High Temperature Porosity	HTPOR	%
Low Temperature Porosity	LTPOR	%
High Temperature Permeability	PERM	md.
Low Temparature Permeability	PERMLT	md.
Oil Saturation	SO	%
Water Saturation	SW	%

2.7 Whole Core Data

The strong diagenetic overprint on the reservoir rock was of particular interest to the operator. Evaporite deposition during diagenesis can have a significant and negative affect on both porosity and permeability. In order to assess the effect of gypsum in particular on porosity and permeability, the cores from six wells were subject to two different analysis suites. The initial analysis was conducted on core samples at low temperature, below 104 degrees Fahrenheit. After this the cores were heated to drive off water of hydration from the gypsum and the measurements were repeated (after the cores had cooled)[13].

In this study only the values from low temperature core analyses were used because they are thought to more closely represent *in situ* rock characteristics. Figure 2.9 shows a crossplot of porosity values measured after heating the cores to dewater the gypsum versus low temperature porosity. Note that the values for high temperature porosity are generally shifted towards higher values, as would be expected.

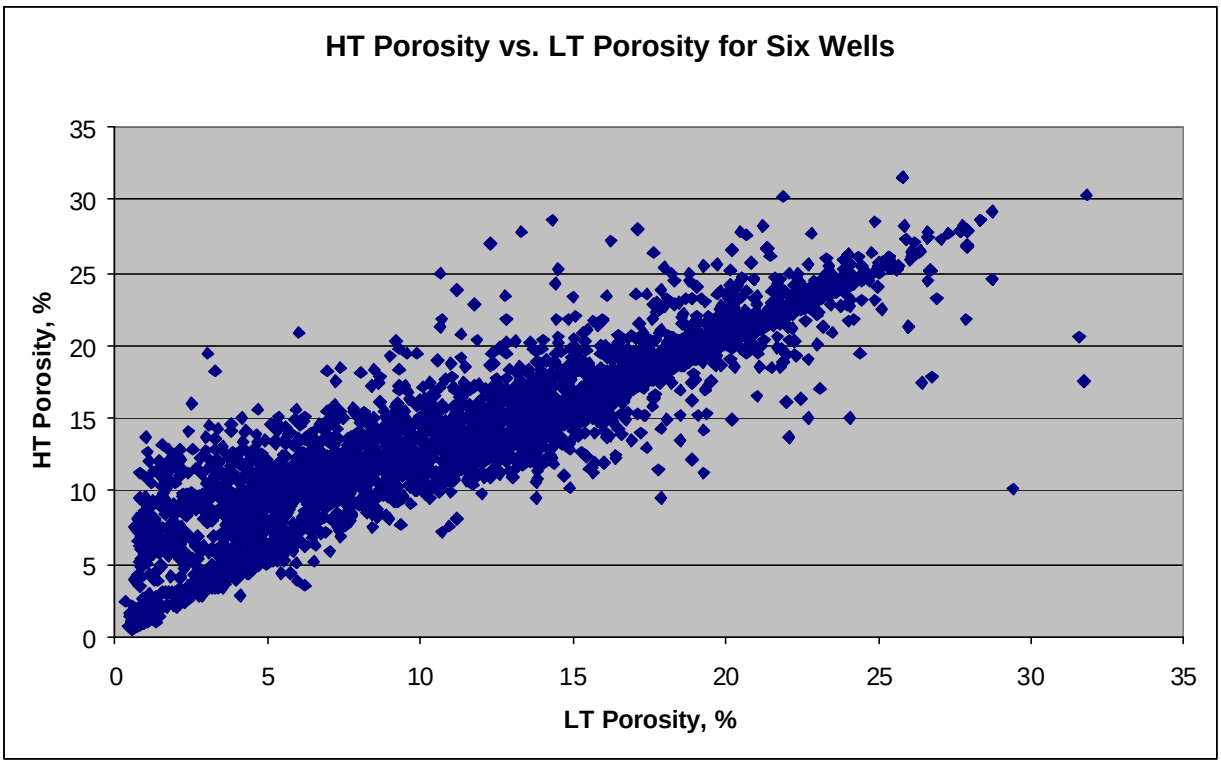

Figure 2.9: Low temperature vs. high temperature porosity measurements for six wells

Figure 2.10 is a similar crossplot for high and low temperature permeability values. The data exhibits more scatter (note the logarithmic scale), than the porosity crossplot, but generally the same pattern.

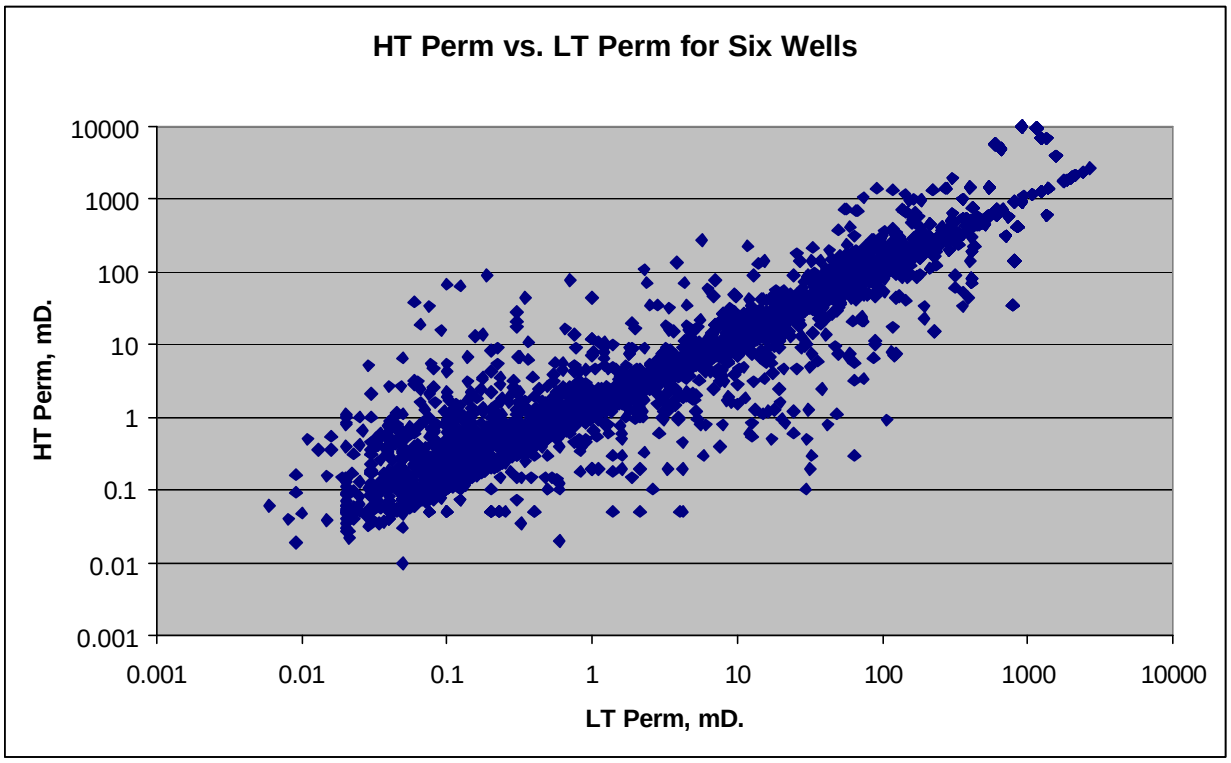

Figure 2.10: Low temperature vs. high temperature permeability measurements for six wells

The (low temperature) porosity distribution is shown in Figure 2.11. For this distribution the mode is 3.3%, the mean is 9.8%, the median is 10.3%, and the standard deviation is 6.6%. Of relevance to this study, the nature of ANN's is such that they tend to weight predictions towards portions of the data distribution where large numbers of samples exist, and perform somewhat more poorly at the extreme values where data density is less. In order to compensate for this tendency, it is necessary to artificially enhance the input data distribution to ensure that extreme values will not be neglected when the network models the data. This is done by repeating data values at the extreme ends of the distribution. Note that the data is not being altered, and no information is artificially added; rather, data values are simply repeated to form stronger patterns upon which the network can train.

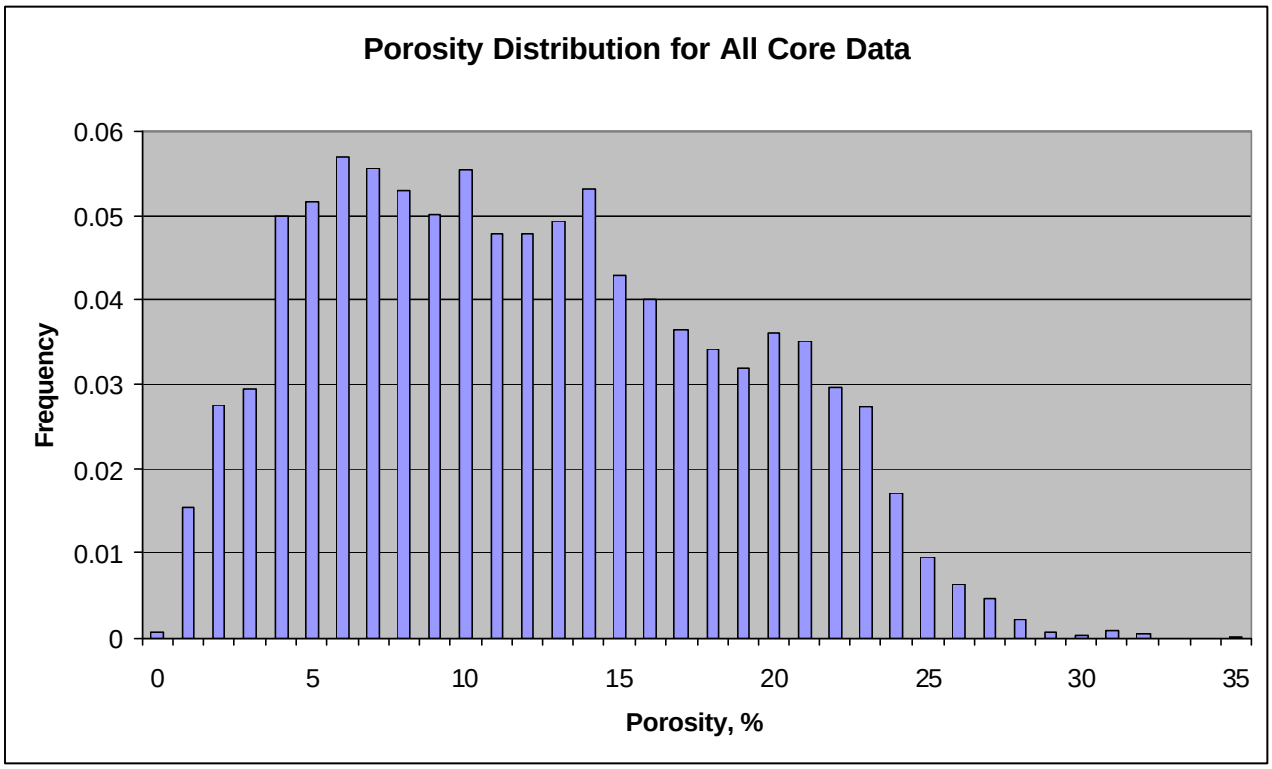

Figure 2.11: Low temperature porosity distribution from core measurements

Figure 2.12 shows the modified porosity distribution. Note the increased data density at the higher end of the distribution in the modified data set. Note also that, due to quirks of the various software programs used to perform the study, the actual data distributions presented in figures 2.11 and 2.12 are slightly different because of different data sampling procedures.

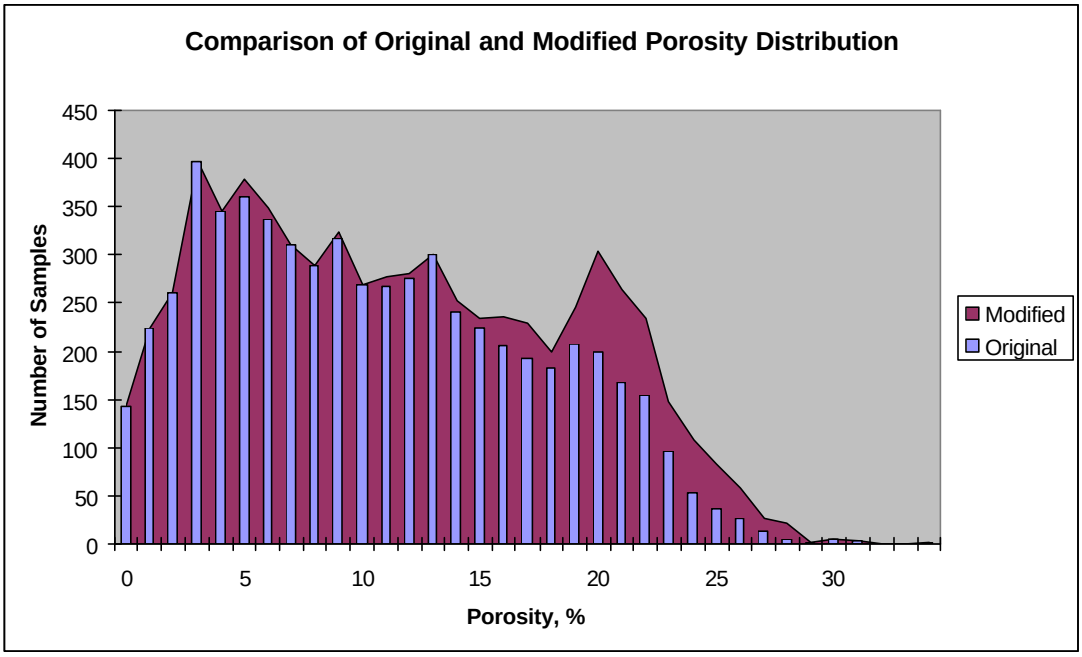

Figure 2.12: Core porosity distribution after modifications to enhance extreme values

The ability of ANN's to accurately model data is also affected by the data distribution of input parameters. Excessive variability in input values can slow network convergence and degrade the ability to generalize. Therefore, data distributions are often normalized to some range that allows the network to converge. For this study it was found that two datasets, specifically the LLD logs and the core permeability "logs" contained excessive variability that affected the ability of the network to converge. The LLD log and permeability values were therefore normalized using a log-base-10 transform. Both data types exhibited variations over five orders of magnitude, so transforming the data using the logarithm function limited the data to values between minus two and just over three.

The resulting log permeability distribution is shown in Figure 2.13. For this distribution the mode is -1.3 (0.05 md.), the mean is 0.56 (3.6 md.), and the median is 0.62 (4.2 md.). The standard deviation is 1.2 (15.9 md.). A strong second mode occurs between values of 1.5 (31.6 md.) and 1.8 (63.1 md.). Note that this mode occurs near the highest density of permeability values and is probably a more reliable measure of the permeability for the true reservoir rock. Permeability values of 0.01 md (log value of -2) and less were eliminated from the dataset as they were below the measurement resolution of the laboratory equipment used to estimate permeability.

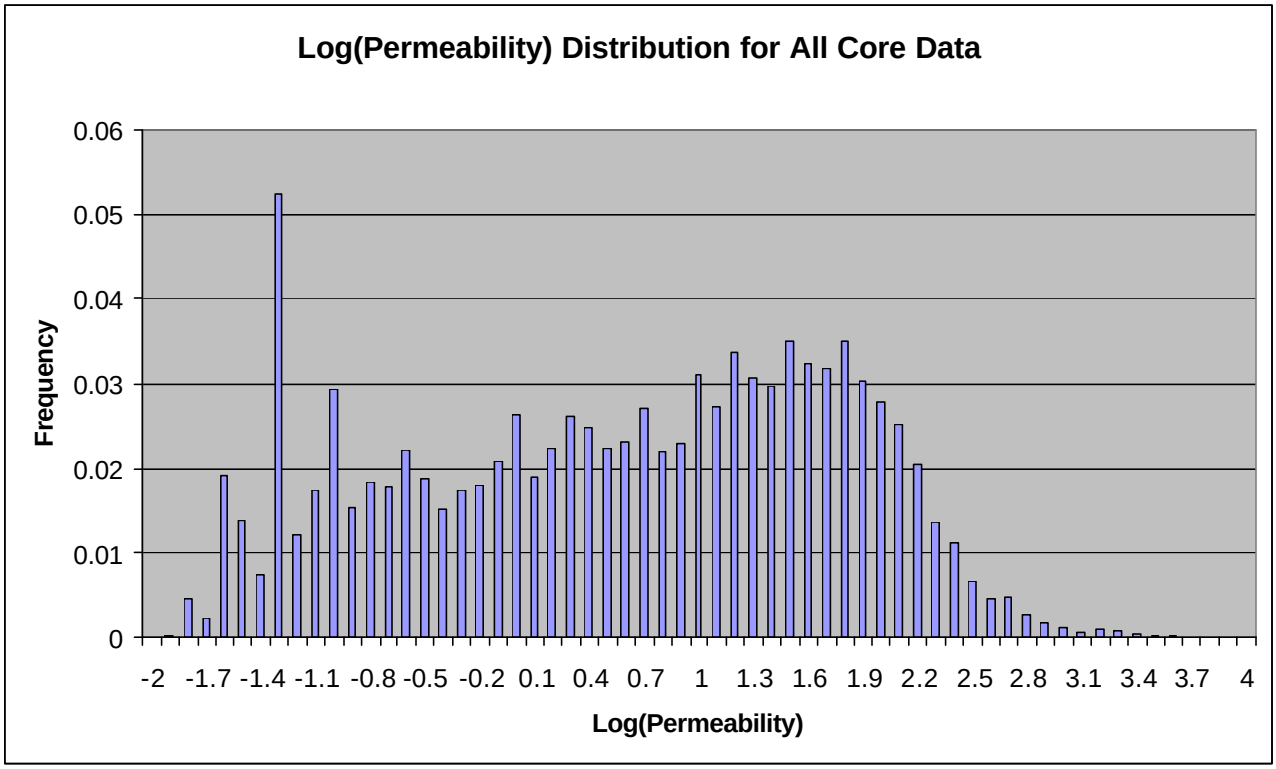

Figure 2.13: Low temperature permeability distribution from core measurements

Similar to the porosity distribution, the actual permeability distribution was modified to increase data intensity at the extreme ends. The modified log (permeability) distribution is shown in Figure 2.14. Again, the distribution of actual data differs slightly between Figures 2.13 and 2.14 due to different data sampling procedures.

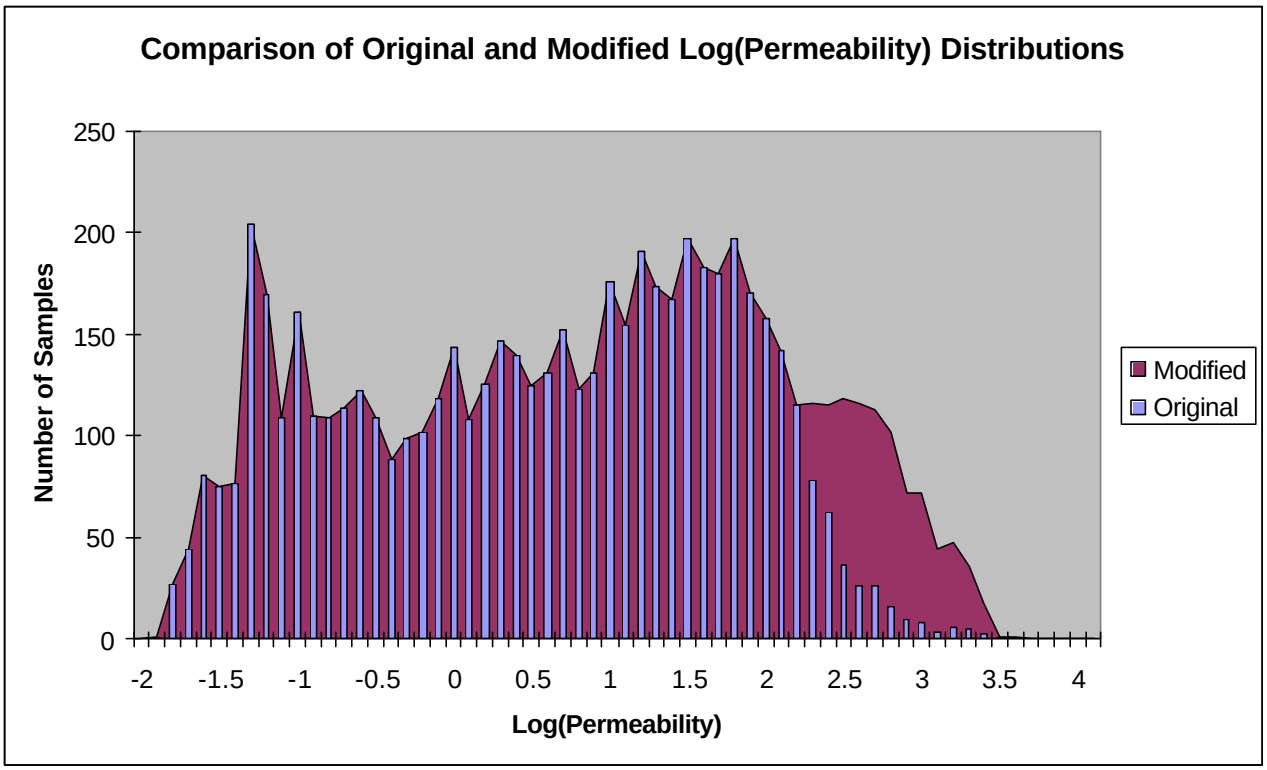

Figure 2.14: Core permeability distribution after modifications to enhance extreme values

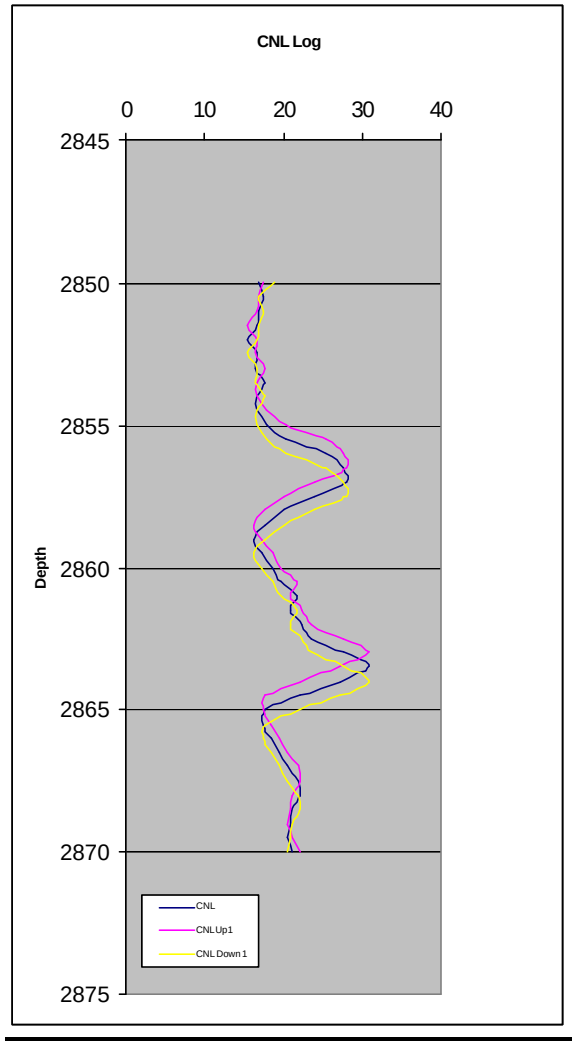

Figure 2.15: CNL depth windows

2.8 Wireline Log Data

Representative ANN models rely on having patterns that accurately represent the physical response of the system in question. The various well logs have tool response overprints that mask the true petrophysical values of the formations being measured. This is why a well log looks like a series of sine waves rather than a square, blocky type of response.

To compensate for the tool response overprint a technique known as "Depth Windowing" was used. This consists of duplicating the input logs at varying vertical offsets to the original. This not only removed the tool response effect, it also created a much richer data environment for network training. In this analysis each input log was copied and shifted up one foot and down one foot from the original matching the beam-width resolution of the LLD tool[14]. This effectively tripled the information available to the network for training.

Since there were six well logs used as input, creating depth windows resulted in eighteen inputs for the neural network representing six different petrophysical measurements. Figure 2.15 is an illustration of the "depth window" concept. In this illustration the compensated neutron log (CNL) along, with its two duplicates, is shown for a depth range across the Grayburg formation. Note that the log data has not been altered but simply shifted up one foot and down one foot. This shifting allows the log to "anticipate" excursions due to bed boundary effects and tool response to bed boundaries. In effect it removes some of the latency (lag) from tool response.

2.9 Other Data Types

The field dataset also includes interpreted tops picks based on geologic markers from well logs. The markers in the study area are chosen to be consistent with markers used on the west side of the field where they closely correlate to distinctive well log events. Figure 2.16 shows a type log from the area including curves for core porosity and permeability, as well as tops picks.

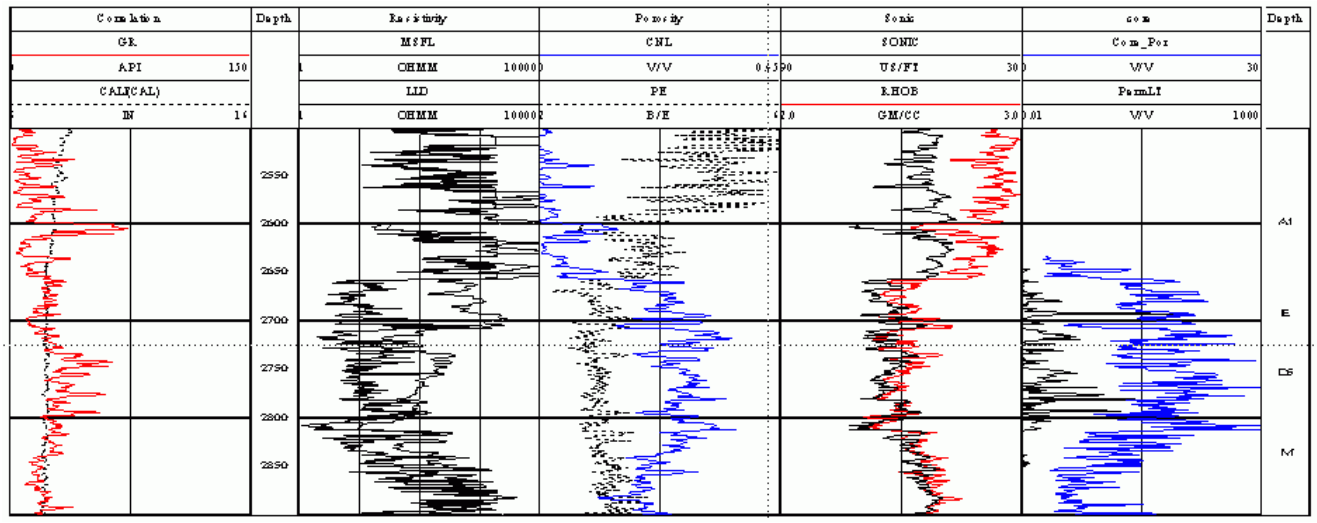

Figure 2.16: Type log for well EQ6777 in the study area

In the area of McElroy field the Grayburg markers are denoted A1, E, D5, and M. The next marker down the stratigraphic column is the San Andres, which is difficult to distinguish and has been picked in only a few of the study wells. On the type log for EQ6777 several important features are readily distinguishable. Note that good porosity development occurs between the A1 and E markers. The main reservoir is in the area of the D5 marker, and the facies near the M marker become increasingly tight. For this study the log and core data were restricted to the Grayburg, interpreted as the interval from the A1 marker to the base of the M zone. Based on available San Andres picks the base of the M was interpreted to be 150 ft. below the M marker.

3. Neural Network Construction and Training

The relationship between well logs and core parameters is strongly non-linear, particularly in non-clastic, carbonate rocks such as the Grayburg. With log and core values measured at half-foot intervals in ten wells, a large volume of data was available for analysis (approximately 6,000 data points – 10 wells x 300 ft/well x ½ ft measurement interval). These two attributes of the dataset (complex relationship and large volume of data) suggest ANN's can be an effective method to model the core-to-log relationship.

The strategy for creation of the model consisted of four steps:

1. Construct an ANN model and train it using data from cored wells.
2. Test the model during training to ensure that it is not memorizing test patterns.
3. Validate the trained network using data from cored wells that has not previously been presented to the network.
4. Use the trained network to predict core values in wells that do not have whole core samples available.

In keeping with standard practice described in the literature, the sample data set was broken into subsets consisting of 60% of data used for training, 20% of data used for testing, and 20% of data used for validation.

3.1 Neural Network Overview

For the reader not familiar with the technology of ANN's, a brief overview is given here to ensure a minimum level of understanding for the sections to follow. A more complete discussion of neural networks can be found in Appendix B.

An ANN is a simple computing machine. It consists of a series of *nodes* connected by *synaptic weights*, or simply "weights." Some nodes are for input, some for output, and some for calculation. Figure 3.1 shows a typical node[15].

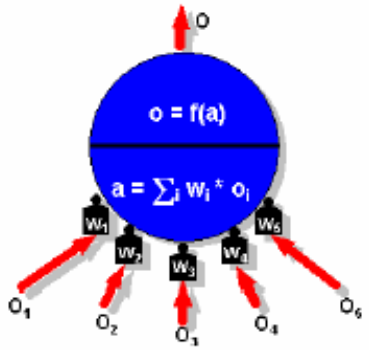

Figure 3.1: Single neural network node showing weights and activation function (after Bishop[15])

In this illustration the input values are denoted by $\mathbf{O_i}$ and the weights by $\mathbf{W_i}$. Also shown is the *activation function*, **a**, and the output, **O**. In this case the activation function is a simple linear summation. Other types of activation functions are sometimes used.

In practice each input value is multiplied by its associated weight. The activation function sums across all input products, and the result is an output value. If the node is an input node the weights are generally equal to one. If it is a calculation node the output value is passed to the output node(s). If it is an output node then this is the final result.

In one common type of network, the *feed-forward back-propagation* ANN, the final result is compared to a known value corresponding to the particular inputs.

An error function representing the degree of accuracy of the output is then calculated. This error is used to adjust the weights and another iteration is conducted. After many such iterations the weights converge to values that cause the node to produce accurate results.

Most neural networks consist of a large number of nodes and weights configured into *layers*. Figure 3.2 shows a typical neural network with an input layer consisting of three input nodes and an output layer with two nodes. The layers between these two are *hidden layers*, one consisting of two nodes and the other with four nodes. Note that each node is connected to every node in the next successive layer.

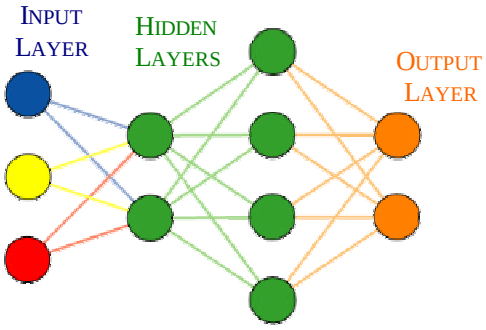

Figure 3.2: Typical neural network with two hidden layers

For this network the row vector containing the three input values and two output values are termed a single *pattern*. For data with complex relationships the network must be presented with a large number of such patterns in order to "learn."

Neural networks are *data driven* calculation machines. They work on the principal of pattern recognition. For the feed-forward, back-propagation network the training data must contain both input values and the known resulting output values corresponding to those inputs. Attempting to match these known output values is what allows the back-propagation algorithm to adjust the network weights until convergence.

ANN's store their knowledge in the synaptic weights. When a network is properly trained it can then process new input data (of the same type used in training) and will produce output values that correspond to the inputs *in the same way training outputs correspond to training inputs*. It is this learning process that gives the artificial neural network its name because it so closely mimics the learning process in the human brain, albeit at a much simpler level.

3.2 Neural Network Construction

Two constraints guided the architectural structure of the ANN used in this analysis:

1. The network must generate predictions of porosity and permeability simultaneously. This constraint was imposed by the requirements of the global process which calls for simple methods of reservoir characterization. (If we were to create individual models for each core parameter, then for each individual log (from crosswell data), then for each crosswell attribute (from 3D data), the ultimate workflow would be much more burdensome than desired).

2. The network must be able to reliably generalize from training data to the prediction mode. This constraint dictates that the internal structure of the neural network be relatively simple, with the fewest hidden node layers possible.

The imposition of these two constraints helps narrow the range of possible ANN architectures. The optimal architecture was determined to be a three layer model with one input layer, one hidden layer, and one output layer. As discussed in the previous section, the input well logs were "depth windowed," effectively tripling the apparent number of network inputs based on well logs. Although only six input log types were used the effect is to triple the number of log inputs to eighteen. While this makes the network

architecture appear more complicated there are still only six types of log data being used as input. As previously stated, the output layer consisted of both target values, core porosity and permeability.

The only remaining network variable was the number of nodes in the hidden layer. The optimum ANN configuration was determined through trial and error. A number of networks with varying numbers of nodes on the hidden layer were constructed and run on the same input dataset. The results were noted and compared. For completeness numerous alternate network architectures were constructed and tested including several with more than one hidden layer. The network which yielded the best results was found to be one with nine nodes on the hidden layer. As discussed previously, the input layer consisted of eighteen input nodes, giving a final network architecture of 18-9-2. An illustration of this network architecture is shown in Figure 3.3.

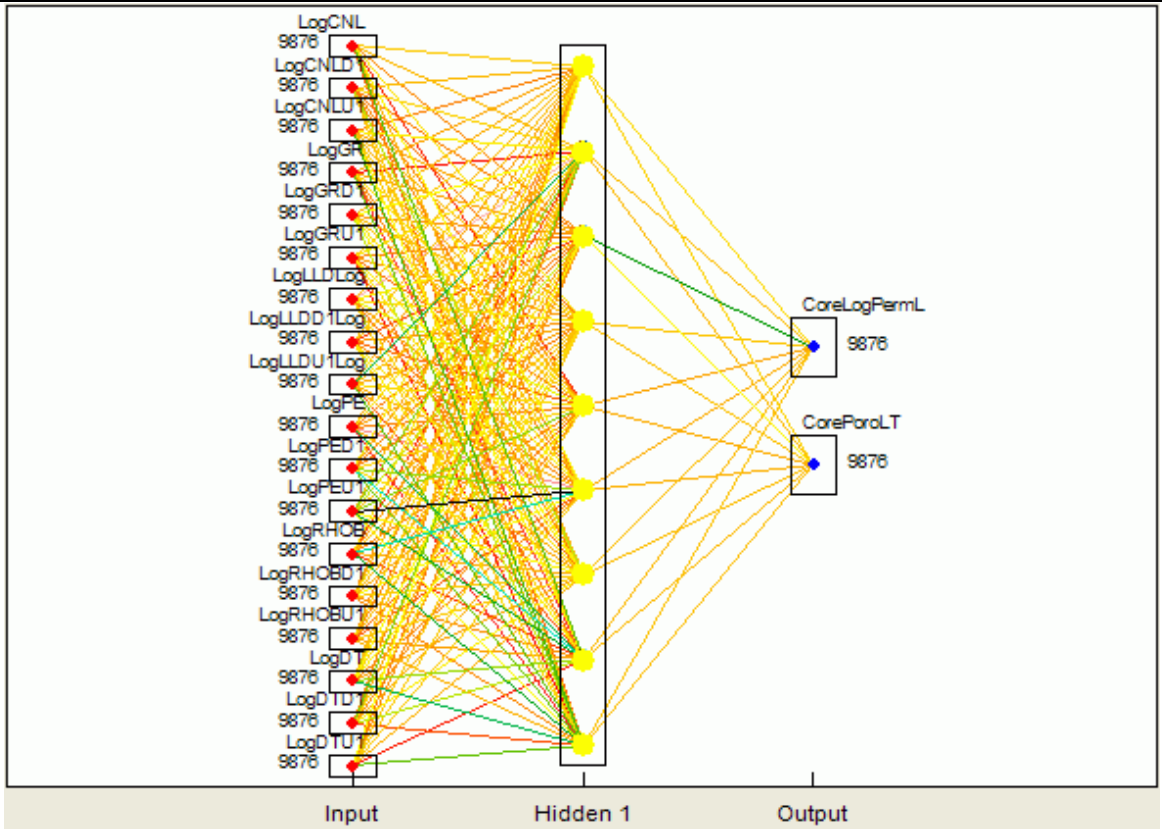

Figure 3.3: 18-9-2 Network architecture

The node activation functions in this network are linear for each input and output node. A sigmoidal activation function was used on the hidden layer nodes. The momentum was set to 0.1 and the learning factor was set to 0.2. These parameters govern the allowable adjustment of synaptic weights by the back-propagation algorithm during any given training epoch. They have a strong influence on rate of convergence and tendency to memorize training data. These values for momentum and learning factor can be considered conservative; i.e. they are fairly low and result in a network that learns more slowly, but is more stable.

3.3 Neural Network Training

ANN training is an iterative process. In the case of back-propagation networks the input and target vectors are fed through the network and the training algorithm is allowed to adjust the synaptic weights until the network can faithfully reproduce the desired output. After each iteration the adjusted weights are stored and made available for the next training pass.

For this study, training the neural network consists of feeding log data inputs and corresponding known values of core parameters to the network and allowing the back-propagation algorithm to iteratively adjust the synaptic weights until the network can accurately reproduce the core values in the testing dataset.

Each iteration is termed an *epoch*. The number of epochs, required to obtain a reliably trained network is not fixed. For this problem the rate of improvement fell to near zero after a few thousand epochs, although the network was allowed to continue to as many as 300,000 epochs on some runs to determine if more iterations materially improved the results.

As mentioned previously, the database was split into three components: one portion of data was held out of the training algorithm for later network validation, one portion was used for training purposes, and one portion was used for blind testing of the network during training. The relative sizes of the various portions are 60% training, 20% testing, and 20% validation. This split was made on an input dataset comprised of approximately 6,000 input patterns (10 wells x 300 ft of core per well x ½ ft core measurement intervals). The network therefore operated on 3,600 training patterns and 1,200 testing patterns. Another 1,200 patterns were available for network validation.

3.3.1 Training With and Without Clustering

Log data clustering finds similarities in data based in part on pattern recognition technologies. The clustering done in relation to this project is the subject of a separate report and will not be discussed in depth here. However, by using fuzzy logic to discover similarities and relationships in data, clustering is able to assign data to separate groups based on user defined criteria.

For the purposes of this study the log data was clustered according to probable rock type as defined by similar log response. Factors defining a particular rocks' log signature include mineralogy, morphology, diagenetic history, and fluid content; in short, it's facies type. The logs for the fifty-nine wells were clustered and the results expressed as a facies code at each measurement depth.

The result of log clustering is a continuous pseudo-log of type codes for the entire logged interval in each well. Based on *a priori* knowledge a geologist can determine the probable facies type associated with each cluster of data. In the case of McElroy the data was clustered into ten coded groups as described in Table 3.1.

TABLE 3.1: Clustering Facies Type Codes

Facies Code	Interpretation	Grain Density (g/cc)	Gamma Ray (API Units)	PE (barns/m^2)	Porosity (%)	Permeability (md)	% of all Samples	"Peer" Groups	Remarks
1	Sandstone	2.627	36	3.320	13.4	78	0.5		Poorly defined.
2	Anhy, w/Gypsum	2.908	12	2.827	3.6	38	7.9		Very low porosity.
3	Dolomite, w/Gypsum	2.813	20	3.277	9.6	29	26.4	3,7,8,9	High porosity.
4	Dolomite, w/Clay	2.830	53	3.165	7.0	8.9	3.8	4,5	Low porosity.
5	Dolomite, w/Clay ?	2.839	43	3.165	14.1	35.8	5.0	4,5	High CNL due to clay?
6	Dolomite	2.855	28	3.341	4.9	14.1	12.5	6,10	Intermediate porosity.
7	Dolomite w/Gypsum	2.806	13	3.102	7.1	57	7.2	3,7,8,9	Intermediate porosity.
8	Dolomite w/Gypsum	2.815	24	3.241	11.3	87	5.7	3,7,8,9	High porosity (poorly defined).
9	Dolomite w/Gypsum	2.762	28	3.106	17.7	62	18	3,7,8,9	Very high porosity.
10	Dolomite	2.869	21	3.642	5.5	35.2	13	6,10	Low porosity.

Clustering assigns to each depth interval ten fuzzy probabilities based on the likelihood that the rock at that depth belongs to each group. The ten probabilities sum to one. To test the effect of facies type code clustering on ANN performance and the quality of modeling results, the ANN model was run on the same data both with and without cluster fuzzy probabilities as input. That is, a network was first constructed as described in Section 3.2 with six logs multiplied into eighteen inputs. Then the model was modified to include not only the eighteen log inputs but the ten fuzzy cluster probabilities as well. For the network using clustering fuzzy probabilities the optimum network configuration was one with 14 nodes on the hidden layer, yielding a 28-14-2 final architecture. This network is illustrated in Figure 3.4.

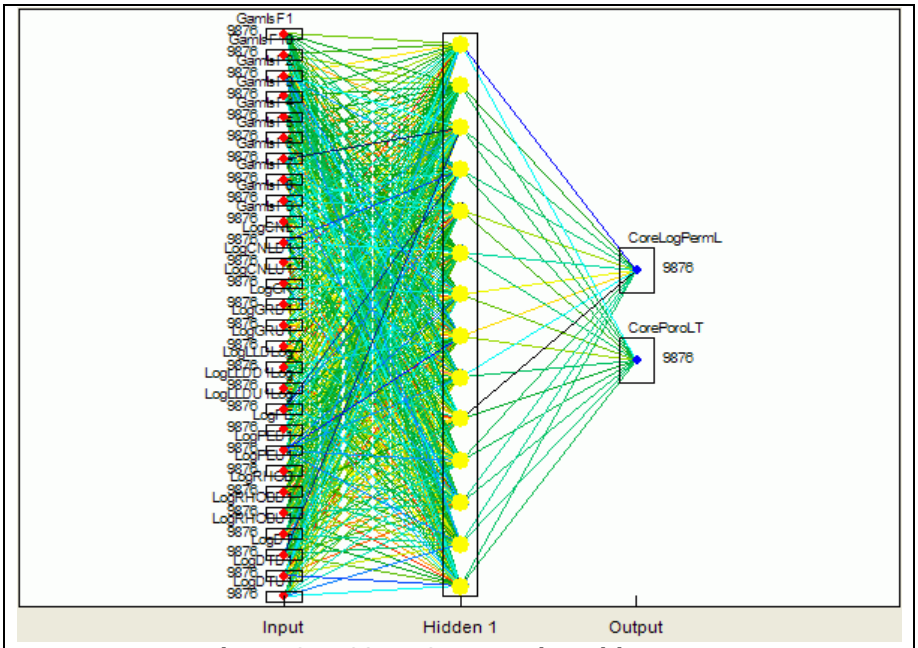

Figure 3.4: 28-14-2 Network architecture

The results from the two networks were then compared to determine whether clustering results improved the predictive capability of the model. To assess that, cross-correlation coefficients were computed between the measured core data and the values predicted by ANN model. Table 3.2 summarizes the results for the training and testing data.

Table 3.2: Summary of Correlation Coefficients Between Measured Data and Values Predicted by ANN's With and Without Fuzzy Clustering Probabilities as Input

Parameter	With Fuzzy Probabilities	Without Fuzzy Probabilities
Porosity	0.863	0.855
Permeability	0.843	0.822

As the table shows, using the results of clustering resulted in a slight improvement of the correlation between actual data and network predicted values. One negative impact of using fuzzy probabilities as input, however, was a less stable network. With log data alone, network performance as exhibited by the error function root-mean-square (RMS), continuously improved regardless of how many epochs the network was allowed to process. One model was allowed to run nearly one-half million epochs with steady, gradual improvement in fit of predicted values.

The network using fuzzy probabilities, on the other hand, converged very quickly but began to diverge after about 8,000 epochs. This did not present a significant obstacle to use of clustering results for network modeling however. Ultimately, due to it's superior predictive capability, the model that included the fuzzy facies codes was selected for use in this study. One benefit of using this network is that convergence in 8,000 epochs used only about 3 minutes of central processing unit time on a Pentium IV class computer.

3.3.2 Regression Fit to Porosity

The resulting porosity crossplot with predicted vs. actual values is shown in Figure 3.5. This plot includes output data for both the training and testing datasets. Note that a line-slope of less than unity and a positive Y-axis intercept indicates that the ANN model tends to weight predictions towards the center of the distribution (i.e., under-estimates extreme data values). This is a common outcome of ANN models, and results in a "smoothing" of highly variable data. A slight improvement in ANN predictive accuracy can be obtained by transforming the ANN predictions using the best-fit line formula.

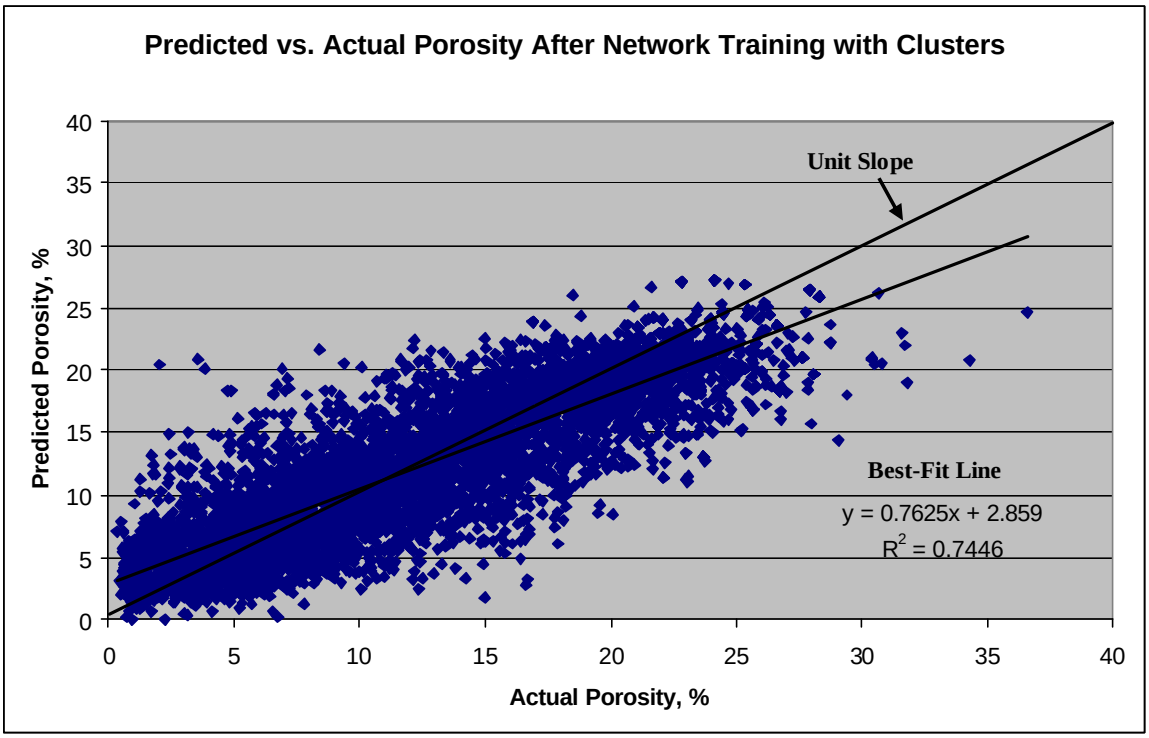

Figure 1

Figure 3.5: Training/testing scatterplot for porosity

Figure 3.6 shows the distribution of porosity values in the measured and predicted data sets. Note that the ANN has most difficulty (and has the greatest source of error) replicating values at the extreme ends of the distribution, but is reasonably accurate in the central portion of the distribution where data density is highest.

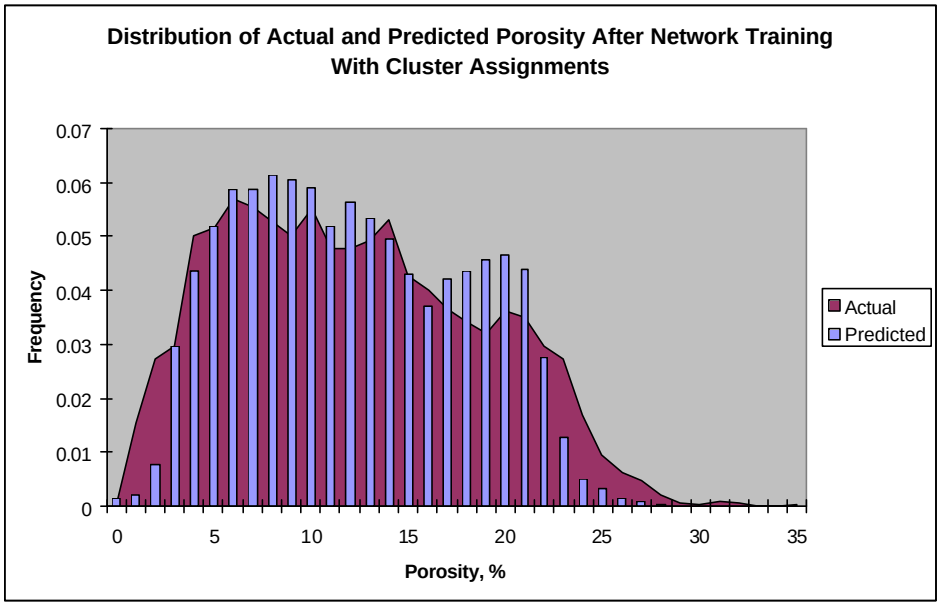

Figure 3.6: Porosity profile for training and testing data

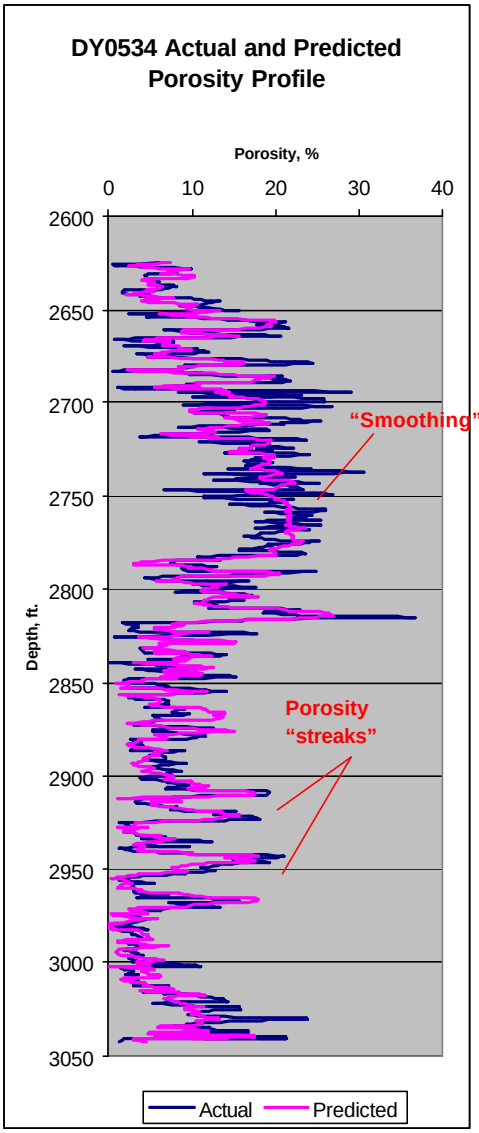

Figure 3.7: Actual vs. predicted porosity with depth, well DYO534

Just as important, and perhaps more insightful, is whether or not the ANN model captured the general *profile* of porosity values observed in the wellbore. Figure 3.7 is a plot of predicted and actual porosity versus depth for one of the wells in the dataset. Note that the neural network was able to closely match the porosity profile, including discrete high and low porosity "streaks" where variability is extreme. However, the model tends to "smooth" extreme variability such as the interval from 2725-75ft. This result was obtained without transforming the ANN predictions with the best-fit line formula from Figure 3.5; it could be improved somewhat by doing so. The vertical resolution that the ANN model provides is more than sufficient for use in reservoir flow modeling. Similar plots for other wells in the cored well dataset exhibit the same good fit to porosity profile.

3.3.3 Regression Fit to Permeability

The actual vs. predicted permeability crossplot is shown in Figure 3.8. This plot also includes output data for both the training and testing datasets. Similar to the results for porosity, the best-fit line slope and intercept suggest the ANN is underestimating extreme values of the distribution. Another feature to note on the plot is the behavior of data points at the very low end of the actual permeability scale. The data contains some vertical "columnar" features that make fitting a line more difficult. The effect is especially pronounced at a log(permeability) value of -1.3 corresponding to a permeability of 0.05 milli-darcies. In fact the actual input data contains a very large number of permeability measurements with exactly this value; an artifact of core

measurement resolution at such low values. Therefore the data appears to be "stacked" in a column on this type of plot.

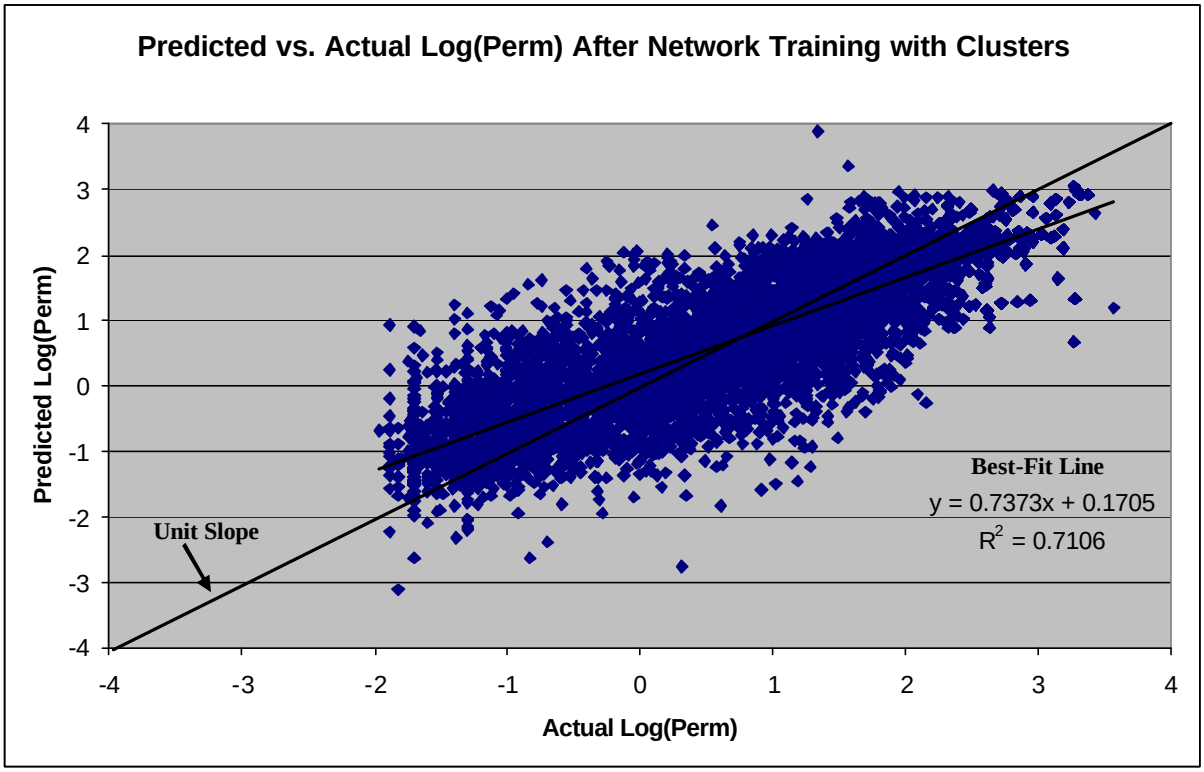

Figure 3.8: Training/Testing scatterplot for permeability

The large number of 0.05 md permeability values (log value of -1.3) can also clearly be seen in Figure 3.9, which is a distribution plot showing actual measured data and neural network predictions. A large spike is also evident in the low end of the original distribution. With the exception of this feature the network has reproduced the distribution reasonably well.

Of particular importance, the network has captured the upper end of the permeability distribution quite well. There is a very close match between the number of predicted and actual permeability values above about log(permeability)=2, or above 100 md. As mentioned, high permeability zones do exist at McElroy and are difficult to deal with operationally because of the need for injected fluid conformance control. It is vital that any model accurately identify these zones.

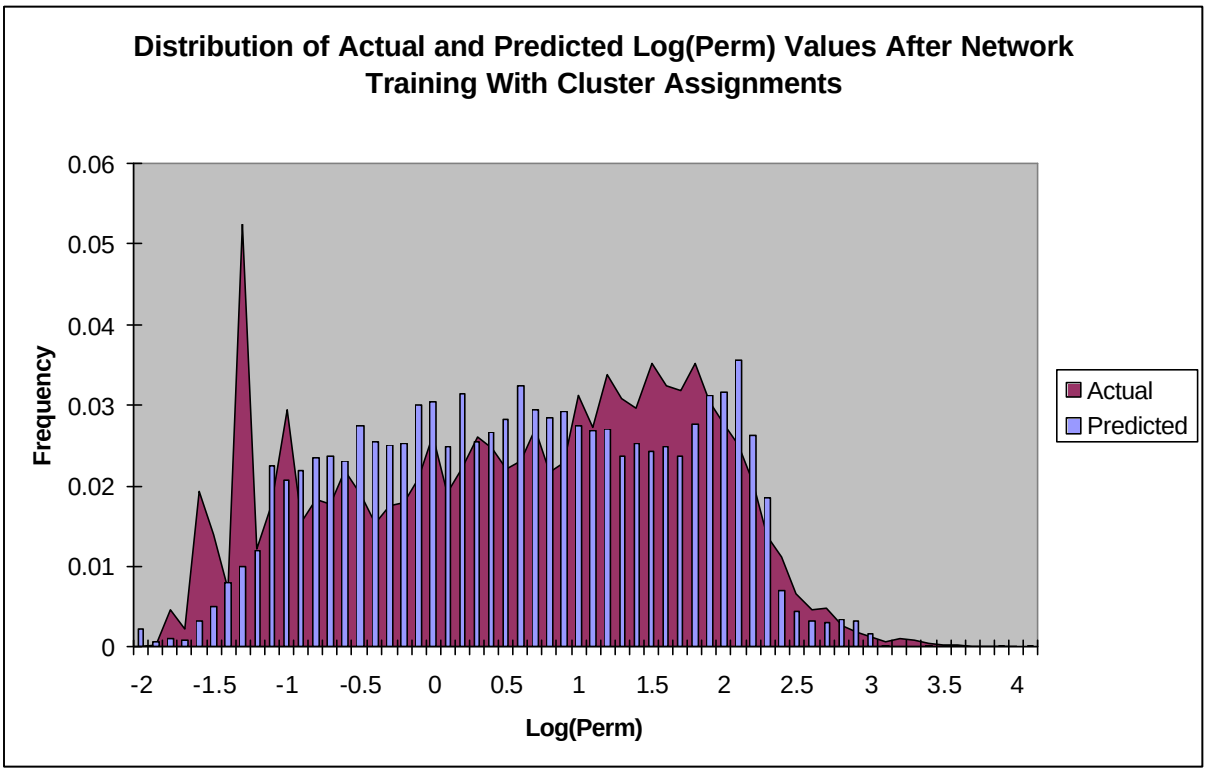

Figure 3.9: Permeability profile for training and testing data

Figure 3.10 is a semi-log plot of predicted and actual permeability versus depth, without transformation of the ANN predictions with the best-fit line formula. The permeability-axis has a logarithmic scale with permeability values in millidarcies. Note the close correspondence between the predicted and actual permeability profiles. Again, the network has successfully modeled high and low permeability "streaks" at the appropriate depths, with a high degree of accuracy. "Smoothing" of extreme variability is also present, as was the case with porosity.

Table 3.2 presents the correlation matrix between each individual well log responses and the ANN model, and actual (core) porosity and permeability. One can observe that while sonic travel time is a reasonable predictor of porosity, none of logs can predict permeability to any meaningful degree. Further, even when examining the various geophysical log traces against the permeability "log", identifying permeability streaks with confidence was not possible. Hence the ability of the ANN model to derive useful information (permeability) from what otherwise might be viewed as unrelated and complex data (well logs) has been demonstrated.

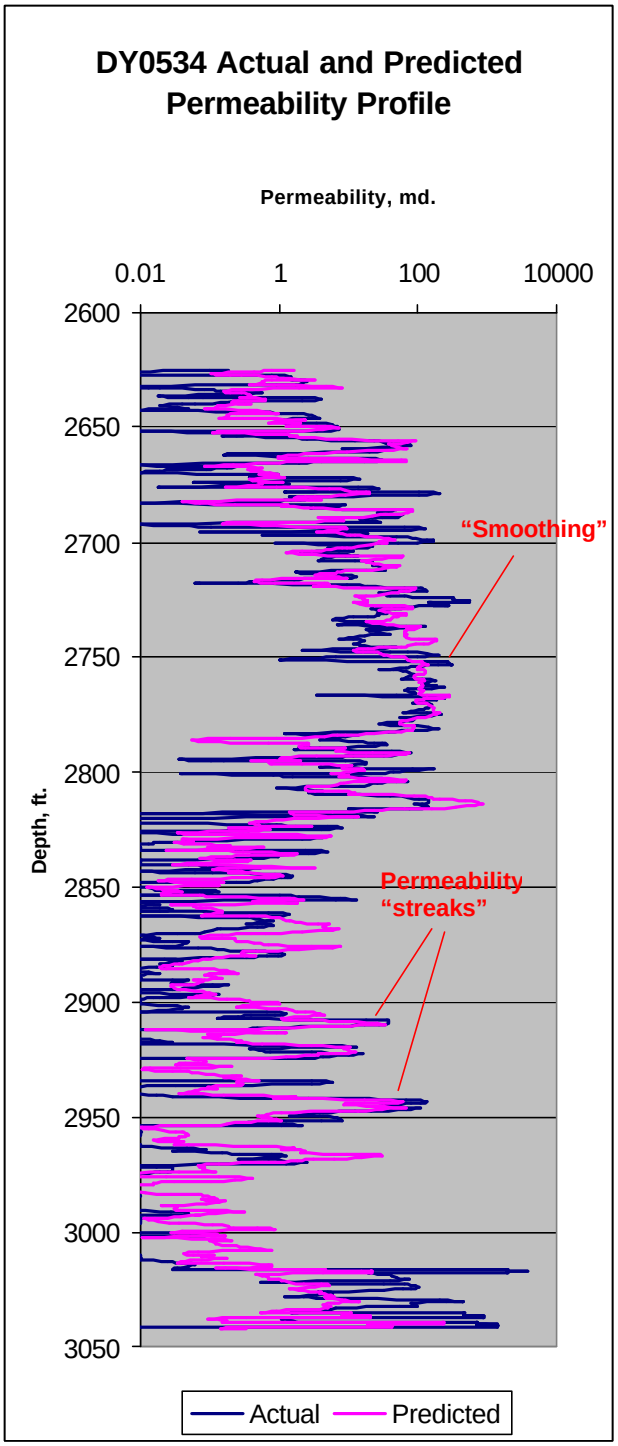

Figure 3.10: Actual vs. predicted permeability with depth, well DY0534

Table 3.3: Correlation Matrix (R^2) for Training & Testing Data

	CNL	GR	LLD	PE	RHOB	DT	ANN Model
Permeability	0.089	0.002	0.009	0.024	0.126	0.173	0.71
Porosity	0.536	0.021	0.100	0.236	0.581	0.673	0.74

4. Neural Network Validation

A trained ANN model should then be validated against fresh data that it has not seen before. The strategy is to use the logs as input to the neural network, let the network calculate predicted output values, and compare those to actual measured core values. The expectation is that a properly trained network will be able to match the measured values very well; preferably as well as the training/testing results.

As stated previously, the entire training data set was broken into sets consisting of 60% training, 20% testing, and 20% validation. The validation data set consisting of about 1,200 data patterns was extracted from the entire data set using a random selection process. Thus the validation data should accurately represent the entire population of data values.

The log data for the validation set was fed into the trained ANN model. During training the network must proceed through iterations so that the back-propagation algorithm can tune the weighting values, but after training this is not necessary. Weighting values are applied to each input value and the data passes more or less straight through the model, producing output very quickly. Because of this, creating output values during validation takes only a few seconds.

Figure 4.1 is a crossplot of predicted porosity versus actual measured core porosity values for the validation data set. For this data the correlation coefficient is 0.852, and is consistent with the training/testing results. This suggests that the ANN model yields strongly reproducible results, a desired trait.

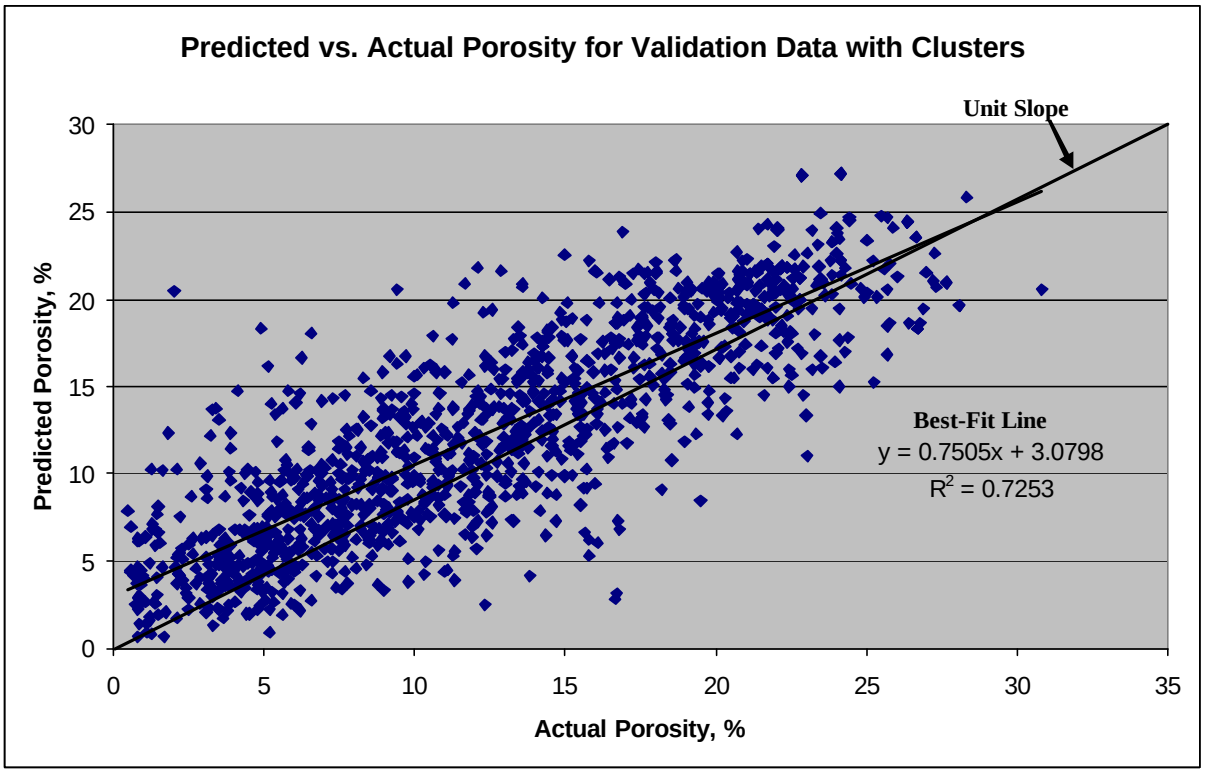

Figure 4.1: Predicted vs. actual porosity values for validation data

Figure 4.2 shows the distribution of predicted and actual porosity for comparison. Again the network has done a reasonably good job of matching the distribution seen in the measured data. The distribution of actual values closely resembles that of the entire data set (shown earlier), increasing confidence that the randomly chosen validation data accurately represents the entire data population. The correspondence between the predicted distributions for higher porosity values is similar to that seen in network training data.

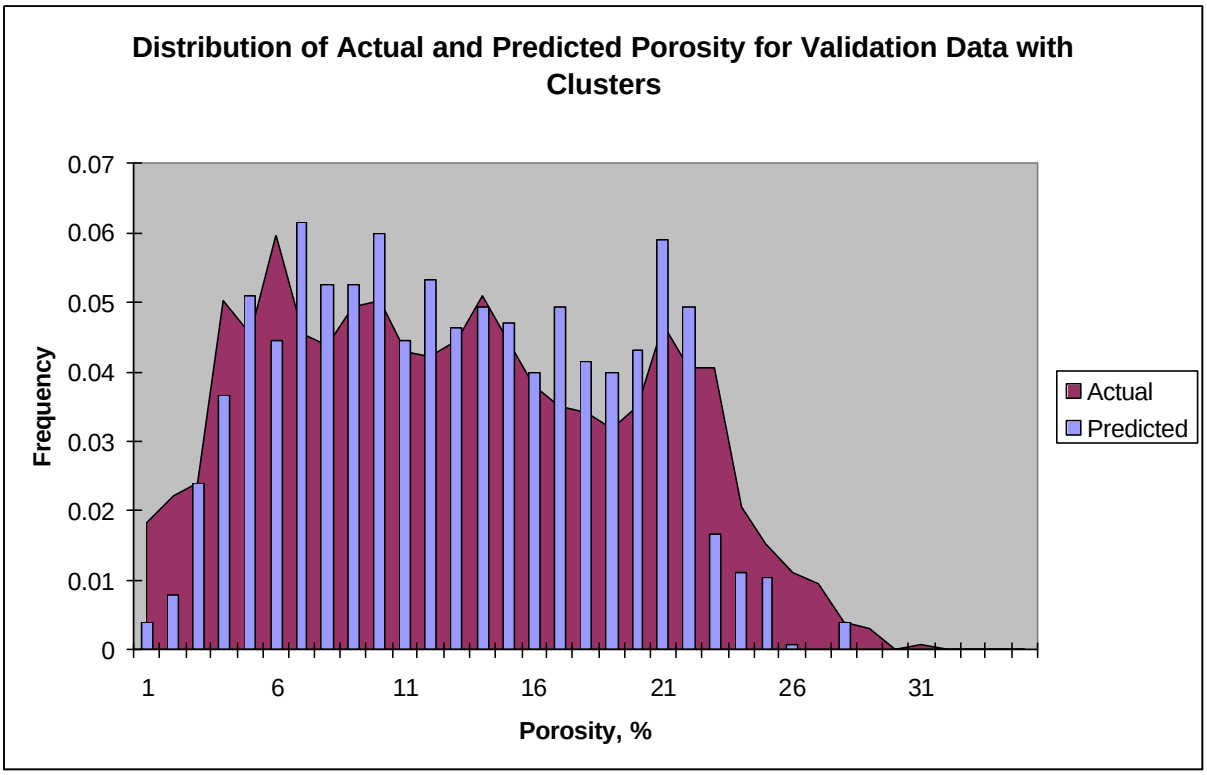

Figure 4.2: Actual vs. predicted porosity distribution, validation data

The model also predicted permeability values. Figure 4.3 shows a scatterplot of predicted versus actual permeability data. A least-squares line has been fit to the data; the line fit parameters are shown on the plot. For this validation data the correlation coefficient is 0.820, again in reasonable agreement with the training/testing data.

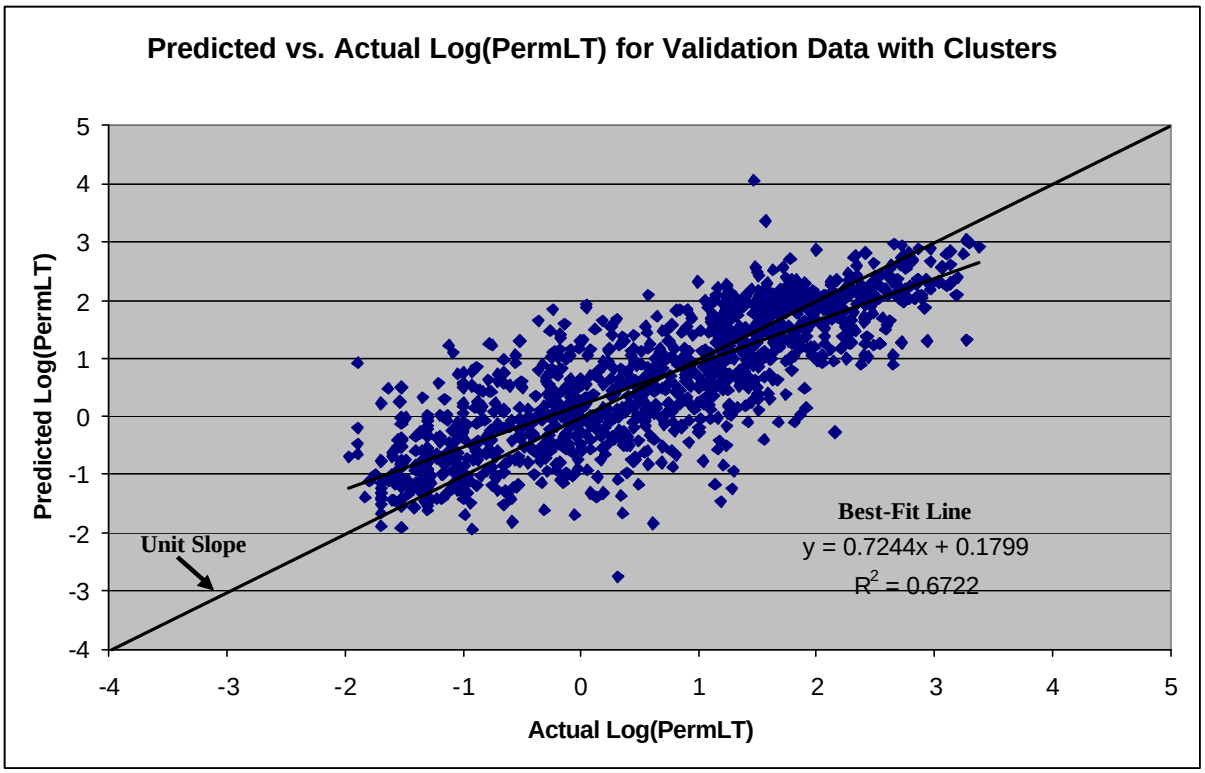

Figure 4.3: Predicted vs. actual permeability values for validation data

The predicted permeability distribution is shown in Figure 4.4, along with the distribution of actual measured values. There is good correspondence between predicted and actual values. Again, this profile is consistent with that of the training/testing data.

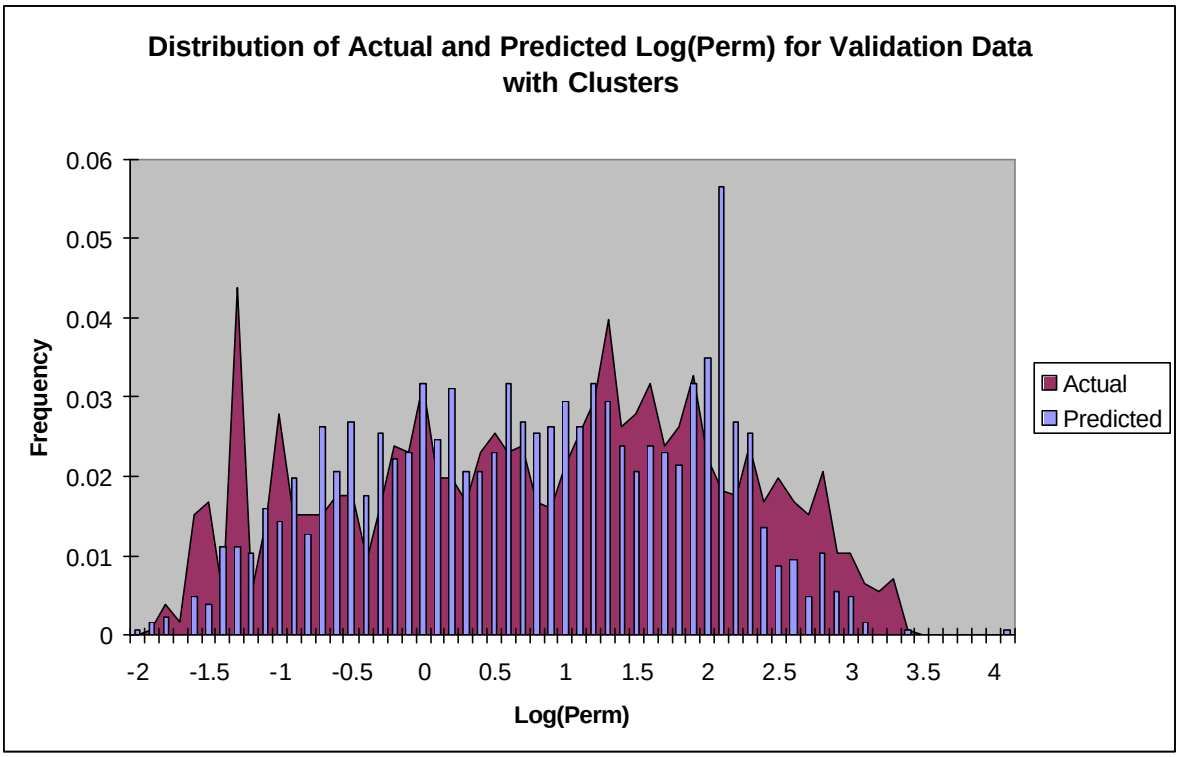

Figure 4.4: Actual vs. predicted permeability distribution, validation data

5. Neural Network Generalization

The final ANN model can be used to predict porosity and permeability values at any location where appropriate (log) input data exists, even though actual core measurements are not available. This process is called network generalization. In this study the trained network was used to predict porosity and permeability for a number of wells in the study area where no core samples had been obtained.

During testing and validation the predicted output values from the ANN model could be compared to measured values taken from core to evaluate their accuracy. In the generalization phase no such values exist. How, then, are we to gauge the validity of values predicted by the network?

The solution comes from classical petrophysics. In conventional well log analysis certain geophysical measurements are directly indicative of near-wellbore rock properties. After visual inspection of the data used for this study it is apparent that the sonic log has an excellent correlation with formation porosity. A crossplot of core porosity versus sonic log travel times is displayed in Figure 5.1. Note that the R^2 for this dataset is different than that between the same two parameters in Table 3.3 because Figure 5.1 encompasses a slightly larger dataset (inclusive of validation data).

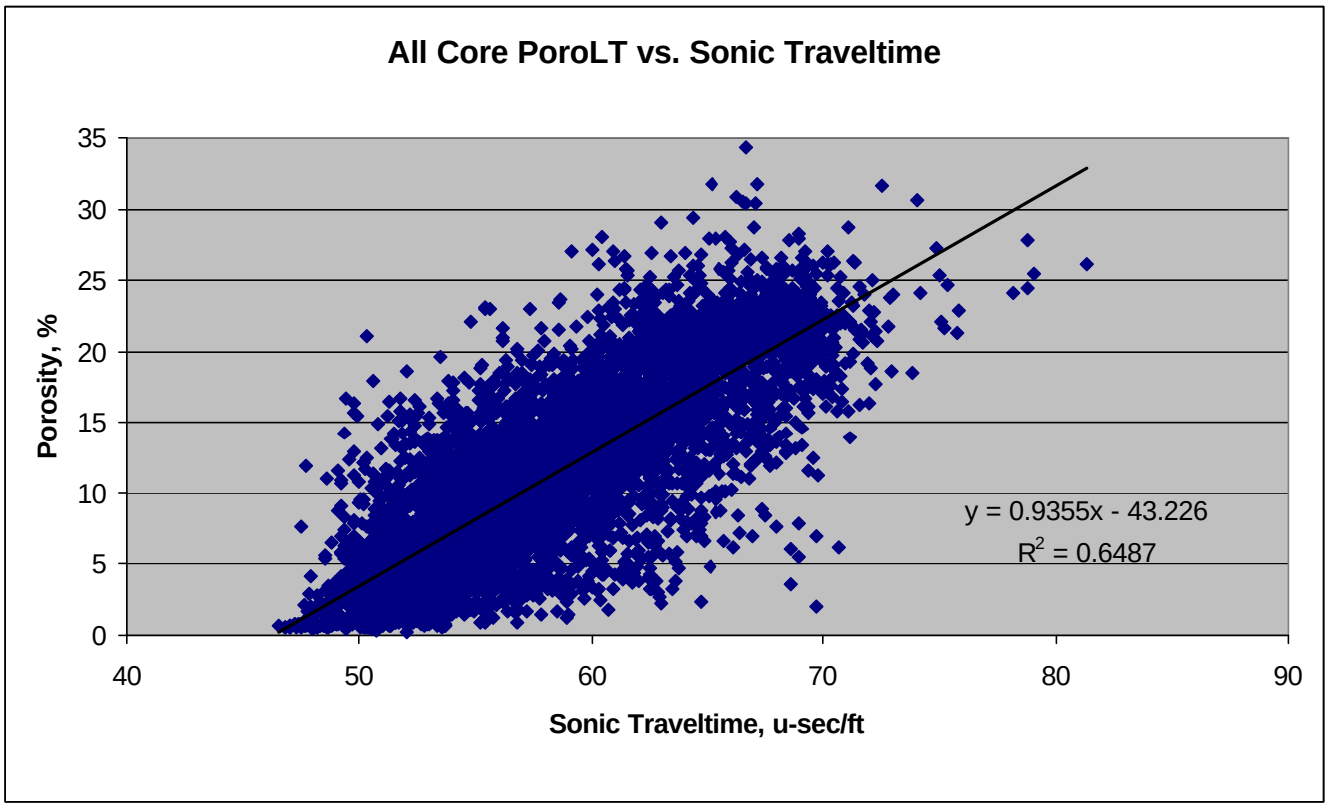

Figure 5.1: Correlation between core porosity and sonic traveltime

The high correlation between core porosity and sonic travel time can be used to validate the output from network generalization. By plotting the network predicted porosity values against sonic travel time logs, which are commonly acquired in McElroy wells, a similar correlation pattern should emerge. If the correlation parameters show a strong relationship between porosity and sonic travel time then confidence is improved in the ANN model.

Likewise, a reliable check on permeability values is needed. Prediction of permeability from well logs has long been a goal of numerous investigators with generally mixed results. No single well log available for this study exhibits a simple, reliable relationship with permeability. Indeed, this is the motivation for using the non-linear prediction capabilities of neural networks to establish a relationship between logs and permeability. However, as Figure 5.2 shows, a reasonable relationship does exist between core porosity and the logarithm of core permeability.

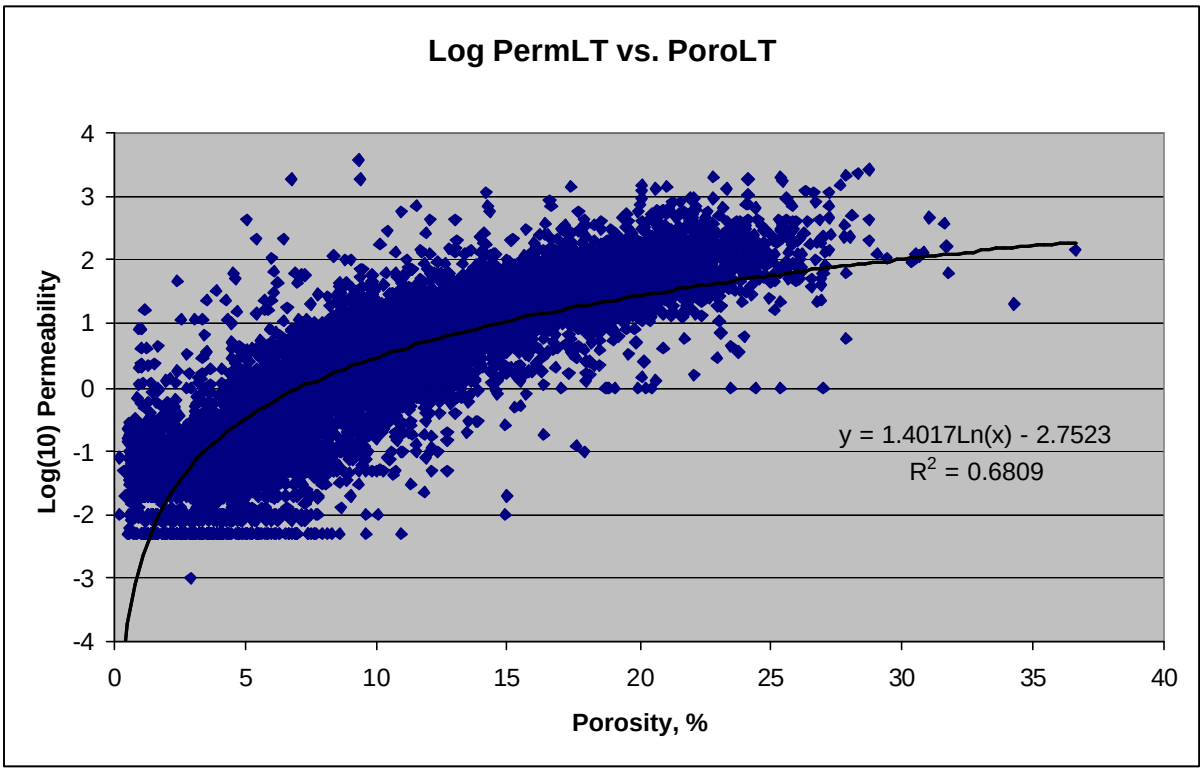

Figure 5.2: Logarithm of permeability versus porosity for all cored wells

Although having a good relationship between permeability and some directly measured parameter would be an ideal situation for validating predicted permeability values, using cross plots of network-predicted log permeability versus porosity is acceptable. Those cross plots should show a relationship similar to the one displayed in Figure 5.2.

The trained network was used to predict the two output values for a number of wells within the study area. Figure 5.3 shows porosity results for one such well, IZ1247. This well is in the north-west corner of the field. The trained neural network was used to predict porosity values for this well and the results displayed as a pseudo-log along with measured values from the CNL and sonic travel time log. Visually, the agreement between predicted porosity and log values appears to be quite good, particularly the correlation with the sonic travel time.

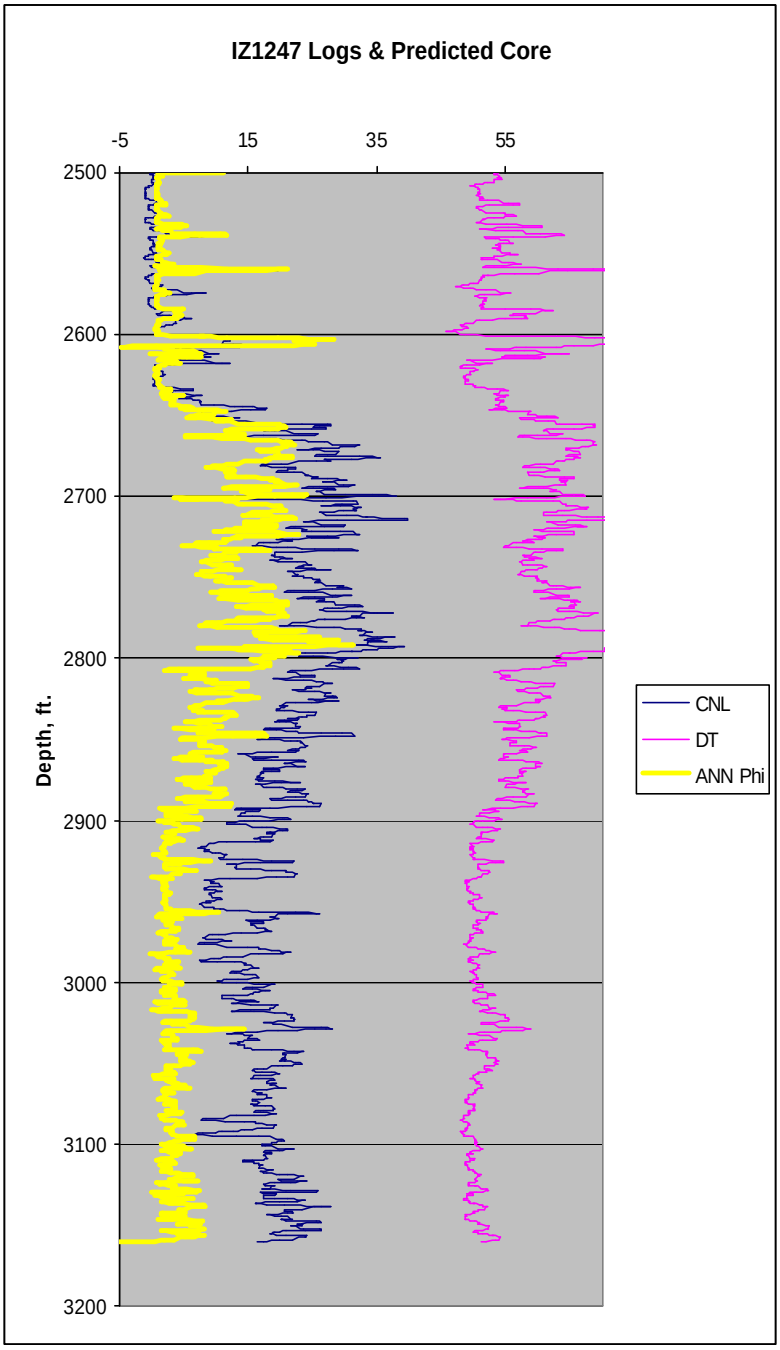

Figure 5.3: Pseudo-log for well IZ1247 showing CNL and DT logs and network predicted porosity

To further investigate the correlation between predicted porosity and sonic travel time Figure 5.4 is a crossplot similar to Figure 5.1. It shows the data from predictions in well IZ1247 with a least squares line fit through the data. The correlation parameters exhibit similarities to the core porosity/sonic travel time crossplot in Figure 5.1.

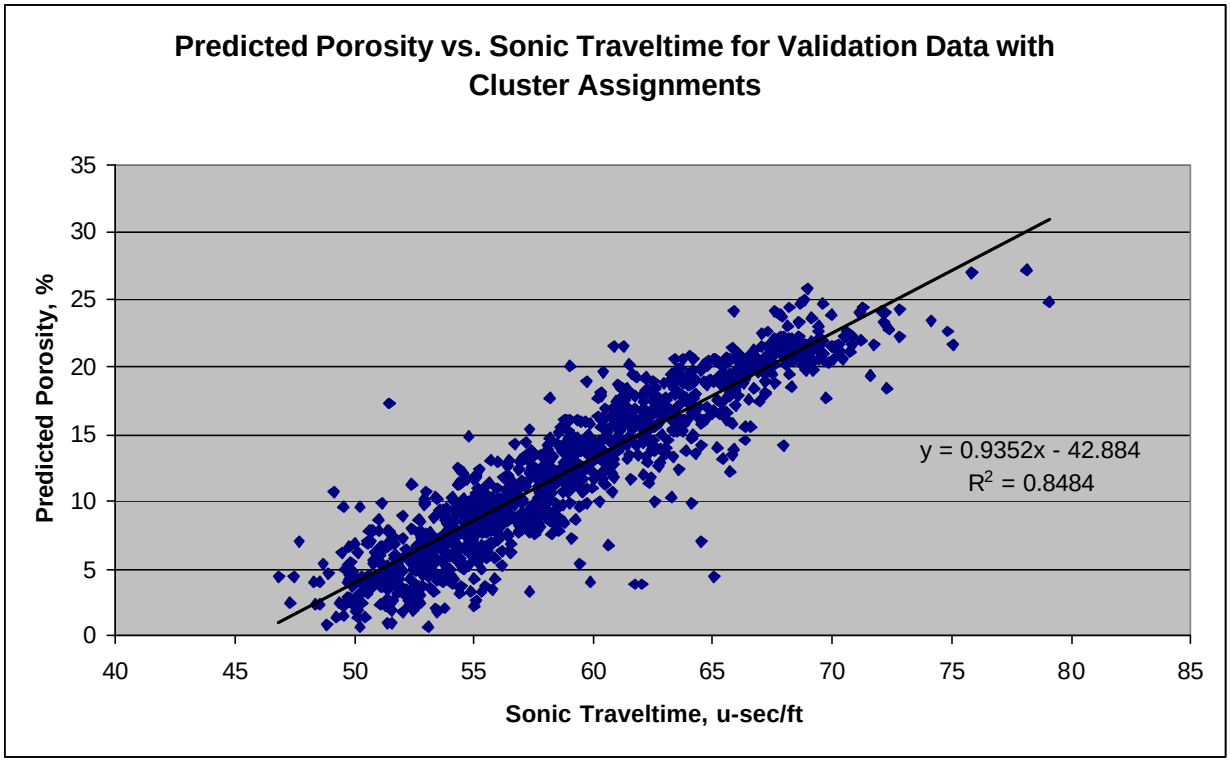

Figure 5.4: Predicted porosity vs. sonic travel time for well IZ1247

Figure 5.5 shows network-predicted permeability versus predicted porosity. The data in this crossplot exhibits similar behavior to the core data in Figure 5.2.

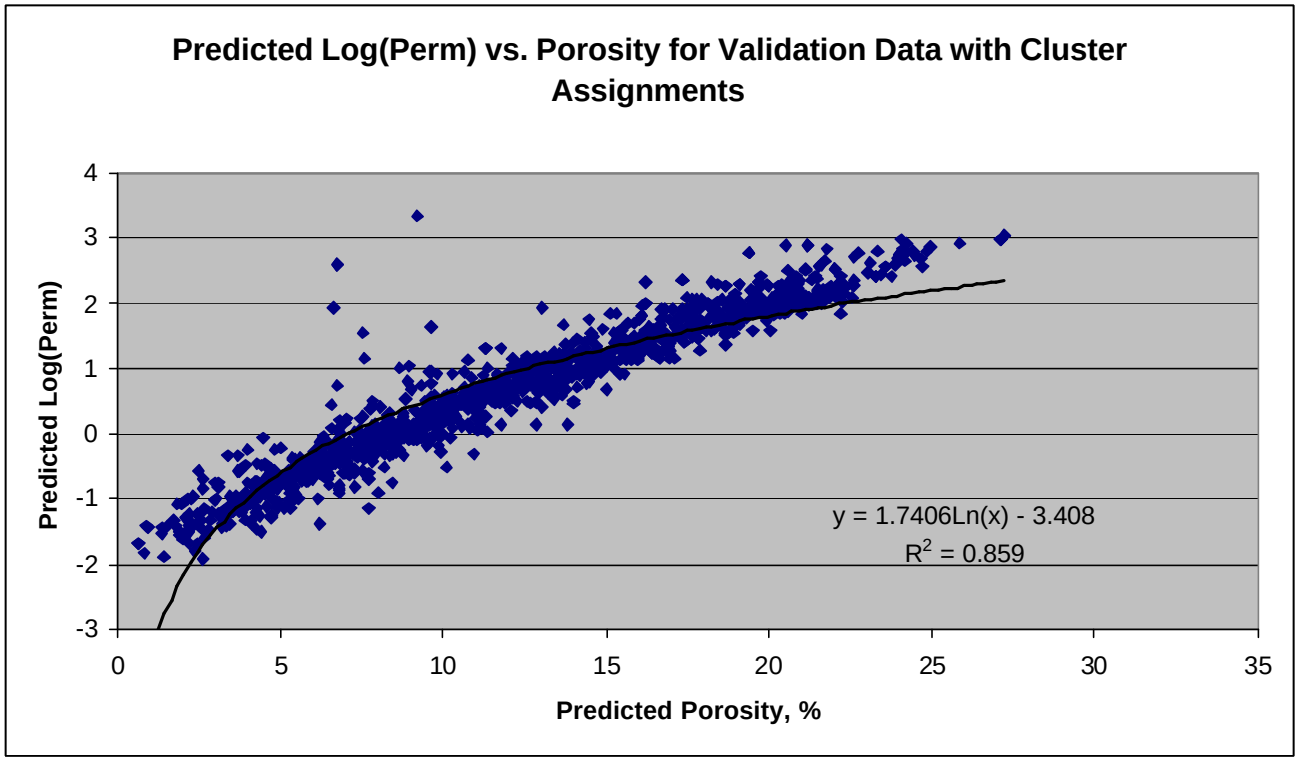

Figure 5.5: Log Perm vs. porosity values for well IZ1247

These results indicate that the porosity and permeability predictions generated during network generalization are reasonable.

The attentive reader might also note with interest that the capability of sonic travel time to predict porosity is not unreasonable and that porosity is not an unreasonable predictor of permeability for this dataset. What then, is the advantage of using ANN's to predict porosity and permeability?

The answer is twofold:

- First, the ANN model is more accurate in predicting both porosity and permeability. Further, if one considers the cumulative error in predicting porosity from sonic travel time, then permeability from porosity, or essentially that of predicting permeability directly from sonic travel time, the R^2 is 0.17 (Table 3.3), much less than the ANN's 0.71 and 0.67 for training/testing and validation data respectively. Hence the ANN model provides a considerable accuracy advantage over simple correlations with logs.

- Second, the ANN model can predict both parameters in one step. In the context of the overall workflow, this represents a significant time and cost savings.

Thus, an ANN model to predict porosity and permeability from well logs for the Grayburg carbonate at the McElroy field in west Texas has been successfully developed, trained, tested, validated and generalized. Performance of the model confirms that ANN's can be used to integrate data on differential scales for high-resolution reservoir characterization.

6. Conclusions

Based on the results of this study, the following conclusions have been drawn:

- An accurate ANN model has been developed to predict porosity and permeability from well logs in the complex Grayburg carbonate reservoir of west Texas. Correlation coefficients between actual and predicted values were 0.86 for porosity and 0.84 for permeability for the training/testing data, and 0.85 and 0.82 respectively for the validation data.

- The model can replicate vertical reservoir property variations very well, and predict discrete porosity and permeability streaks (for subsequent use in reservoir modeling) at a high vertical resolution. The main source in predictive error was extreme reservoir property variability over very short vertical intervals – the ANN model tended to "smooth" the high variability.

- The model was based on 6,000 data points with coincident well log and core measurement information; 60% of the data was used for training, 20% for testing, and 20% for validation.

- Considerable pre-conditioning of the data was employed to achieve these encouraging results. Some of the techniques utilized included duplication of extreme values to enhance their representation in the dataset, "depth windowing" of log data to reduce tool lag effects, and log clustering (unsupervised ANN modeling) to establish facies codes and mineralogical similarities as additional model input parameters.

This study has therefore demonstrated that core porosity and permeability can be effectively predicted from geophysical log data using ANN's to provide a high (vertical) resolution reservoir characterization for subsequent flow modeling and field development optimization.

Acknowledgement

The authors express their gratitude to the U.S. Department of Energy, which provided funding for the research and preparation of this publication (Agreement No. DE-FC26-01BC15357). We also wish to acknowledge the many contributions of the Bureau of Economic Geology at the University of Texas at Austin, as well as ChevronTexaco, the field operator, who so generously contributed the field data for the project. Finally, we wish to acknowledge Decision Team Software for providing the ANN software (DECIDE! FOR OIL & GAS) to perform the study.

References

1. Reeves, S. R., Mohaghegh, S. D., Fairborn, J. W., and Luca, G.: "Feasibility Assessment of a New Approach for Integrating Multiscale Data for High-Resolution Reservoir Characterization," SPE 77759, presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 29 September – 2 October, 2002.
2. Montgomery, S.: "Cross-Well Seismology: A "Frontier" Technology", Petroleum Frontiers, Vol. 12 No. 4, Quarterly 1996.
3. Walker, S. D. and Harris, P. M.: "McElroy field: Development of a Dolomite Reservoir, Permian Basin of West Texas," in *Hydrocarbon Reservoir Studies San Andres/Grayburg Formations, Permian Basin*, Bebout, D. G. and Harris, P. M., eds. Permian Basin Section of SEPM, Publication No. 86-26. (1986). p. 127-132.
4. Energy Information Administration: U.S. Crude Oil, Natural Gas, and Natural Gas Liquids Reserves 2001 Annual Report. http://www.eia.doe.gov/pub/oil_gas/natural_gas/data_publications/crude_oil_natural_gas_reserves/current/pdf/appd.pdg
5. The University of Texas, Permian Basin (UTPB) website, http://www.utpb.edu/PBDPL/Statistics/main_permian_basin_oil.htm
6. PTTC West Coast Workshop Documents found at http://www.westcoastpttc.org/presentations/00-01/092001/03-Don%20Clarke.pdf
7. USGS Website: http://energy.cr.usgs.gov/1995OGData/REG5.htm, Region 5 – West Texas And Eastern New Mexico, Province 44: Permian Basin Province
8. Galloway, W. E., Ewing, T. E., Garrett, C. M., Tyler, N., and Bebout, D. G.: "Atlas of Major Texas Oil Reservoirs." Bureau of Economic Geology, Austin, Texas. 1983.
9. Harris, P. M., and Walker, S. D.: "McElroy Field: Development Geology of a Dolostone Reservoir, Permian Basin, West Texas," in *Geologic and Engineering Approaches in Evaluation of San Andres/Grayburg Hydrocarbon Reservoirs – Permian Basin*, Bebout, D. G. and Harris, P. M., eds. Bureau of Economic Geology, Austin, Texas (1990). P. 275-296.
10. Texas Railroad Commission, ACTI database, www.rrc.state.tx.us
11. Tucker, K. E., Harris, P. M., and Nolen-Hoeksema, R. C.: "Geologic Investigation of Cross-Well Seismic Response in a Carbonate Reservoir, McElroy Field, West Texas." AAPG Bulletin, V. 82, No. 8 (August 1998). P. 1463-1503.
12. Williams, P.: "New Life In Old Basins," Oil and Gas Investor, A Hart Publication, Houston, Texas. October , 2002. p. 28.
13. Nolen-Hoeksema, R. C., Wang, Z., Harris, J. M., and Langan, R. T.: „High-resolution crosswell imaging of a west Texas carbonate reservoir: Part 5-Core analysis," Geophysics, Society of Exploration Geophysicists, Tulsa, OK. Vol. 60, No. 3 (May-June 1995). P. 712-726
14. Dewan, J.T.: "Essentials of Modern Open-Hole Log Interpretation. "Penwell Publishing Co., Tulsa, OK (1983). P.83.
15. Bishop, C. M.: *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York City (1995)

List of Acronyms and Abbreviations

ANN – Artificial Neural Network
ARI – Advanced Resources International
BCF – Billion Cubic Feet of gas
BOPD –Barrels of Oil per Day
CBP – Central Basin Platform of the Permian Basin of west Texas
DOE – Department of Energy
EOR – Enhanced Oil Recovery.
EUR – Expected Ultimate Recovery
LOSF – Light Oil Steam Flood
md – millidarcies
MMBO – Millions of Barrels of Oil
MRI – Magnetic Resonance Imaging
OOIP – Original Oil In Place
RMS – Root Mean Square
SOR – Secondary Oil Recovery
3D – Three Dimensional
USGS-United States Geological Survey
VI – Virtual Intelligence
VSP – Vertical Seismic Profile

Appendix A

List of Study Wells

	Comp Date	UWI	SYM	Status	ELEV_KB	TD	========== Log curves ==========										Other	Core
1	1/16/1996	BC3188	OIL	SP	2628	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		Yes
2	1/15/1996	BC3187	OIL	T/A	2614	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Image	Yes
3	9/25/1996	BI1895	INJ		2617	3005	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		Yes
4	8/30/1997	BN7574	INJ		2584	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			Yes
5	11/5/1984	DY0534	INJ		2609	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic		NEUT		Yes
6	12/7/1984	DY4439	OIL	SP	2605	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic		NEUT		Yes
7	7/19/1981	EQ6777	OIL	SP	2580	3125	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	NEUT, SNP		Yes
8	2/28/1990	IU4476	OIL	RP	2600	3316	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		Yes
9	2/27/1990	IU4716	OIL	T/A	2611	3208	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		Yes
10	2/25/1993	OU5347	OIL	RP	2583	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		Yes
1	12/27/1995	BC5358	INJ		2608	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		
2	9/21/1996	BI1832	OIL	P/A	2618	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		
3	9/17/1996	BI1833	WI	S/I	2619	2998	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			
4	10/1/1996	BI1834	OIL	T/A	2616	3030	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI	Image	
5	9/12/1996	BI1835	INJ		2613	2900	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		
6	9/10/1997	BN4693	OIL	RP	2582	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			
7	9/17/1997	BN4694	OIL	SP	2588	3000	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			
8	4/21/1998	BN7653	OIL	SP	2593	2817	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			
9	3/28/1998	BO3825	OIL	RP	2589	2950	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Image	
10	9/13/1997	BO3826	OBSERV		2590	2950	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Xwell, Image	
11	8/26/1997	BO3827	OBSERV		2589	2950	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Image	
12	8/30/1997	BO3828	OBSERV		2586	2950	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Xwell, Image	
13	9/3/1997	BO3829	OBSERV		2596	2950	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Xwell, Image	
14	9/7/1997	BO3830	OBSERV		2592	2950	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			
15	2/7/2000	BX9023	INJ		2597	3000	Dept		CNL	GR	LLD	PEF	RHOB	Sonic		DPHI	Image	
16	12/6/1984	DY0243	OIL	SP	2610	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
17	11/9/1984	DY0374	INJ		2595	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
18	11/7/1984	DY0386	OIL	T/A	2600	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Xwell	
19	9/25/1984	DY0434	INJ		2585	3185	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				

	Comp Date	UWI	SYM	Status	ELEV_KB	TD	========== Log curves ==========										Other	Core
20	9/30/1984	DY4431	OIL	T/A	2590	3180	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
21	12/12/1984	DY4438	INJ		2607	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
22	11/12/1984	DY4440	INJ		2593	3198	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
23	10/18/1984	DY4441	OIL	T/A	2595	3220	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic	MSFL		Xwell	
24	12/28/1984	DY4443	OIL	SP	2601	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
25	11/1/1984	DY7214	OIL	RP	2577	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
26	3/4/1985	DY7215	OBSERV		2576	3200	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
27	11/1/1985	DY7286	INJ		2603	3202	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
28	11/2/1984	DY7289	INJ		2582	3239	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic				
29	8/10/1988	IM1897	OIL	T/A	2599	3150	Dept	Cal	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
30	11/30/1988	IN7299	OIL	SP	2610	3657	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		
31	3/16/1990	IU4714	OIL	T/A	2626	3202	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
32	4/2/1990	IU4715	OIL	P/A	2595.2	3204	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		
33	4/30/1990	IU4814	INJ		2593.3	3203	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic	MSFL	DPHI		
34	4/11/1990	IU4815	OIL	T/A	2598.9	3206	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
35	5/3/1990	IU4817	INJ		2605.2	3210	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
36	12/18/1990	IZ1247	INJ		2577.8	3200	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic	MSFL			
37	5/21/1992	ON8760	INJ		2608	3087	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
38	2/8/1994	ON8762	INJ		2612	3100	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
39	5/21/1992	OO7035	INJ		2603	3078	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
40	7/15/1992	OQ2908	OIL	RP	2606	3102	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
41	7/20/1992	OQ2916	OIL	T/A	2605	3098	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
42	3/21/1993	OU5334	INJ		2601	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
43	6/10/1993	OU5339	INJ		2579.9	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
44	1/30/1993	OU5342	OIL	RP	2605	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				
45	3/16/1993	OU5343	OIL	SP	2601.6	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
46	1/14/1993	OU5345	INJ		2592	3020	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
47	3/16/1993	OU5346	OIL	SP	2580.8	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
48	1/5/1993	OU5348	INJ		2579.3	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic		DPHI		
49	3/20/1993	OU5354	OIL	RP	2598	3000	Dept	CAL	CNL	GR	LLD	PE	RHOB	Sonic				

Appendix B

General Discussion on Artificial Neural Networks

A neural network performs mappings from an input space to an output space. The class of problems to which neural networks are usually applied are exercises in numerical optimization of nonlinear objective function(s).

When discussing a neural network as a mathematical construct we should more properly say Artificial Neural Network, to distinguish it from biological neural networks. One of the best sources for information on this topic comes from the FAQ of the Usenet Internet group comp.ai-nn which states:

"There is no universally accepted definition of an Artificial Neural Network (ANN). Most people in the field would agree that an ANN is a network of many simple processors ("units"), each possibly having a small amount of local memory. The units are connected by communication channels ("connections") which carry numeric (as opposed to symbolic) data, encoded by any of various means. The units operate only on their local data and on the inputs they receive via the connections. The restriction to local operations is often relaxed during training.

Most ANNs have some sort of "training" rule whereby the weights of connections are adjusted on the basis of data. In other words, ANNs "learn" from examples. If trained carefully, ANNs may exhibit some capability for generalization beyond the training data, that is, to produce approximately correct results for new cases that were not used for training."

According to Haykin, "A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process
2. Inter-neuron connection strengths, known as synaptic weights, are used to store the required knowledge

The process used to perform the learning process is called a *learning algorithm*, the function of which is to modify the synaptic weights of the network in an orderly fashion to attain a desired design objective."

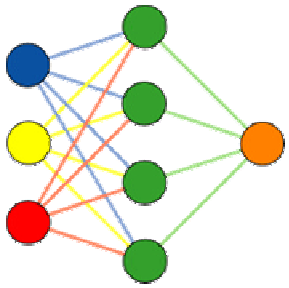

Figure B-1.: Structure of a typical Artificial NeurNeural Network.

Figure B-1 shows the structure of a typical neural network. There are numerous neurons in the input layer, a hidden layer with additional neurons, and an output layer with one or more neurons.

The lines represent synaptic weights. During training of the neural network these weights are adjusted using a learning algorithm. The actual “knowledge” is stored in these synaptic weights, in combination with the node activation functions.

“The two main kinds of learning algorithms are supervised and unsupervised.

- In supervised learning, the correct results (target values, desired outputs) are known and are given to the ANN during training so that the ANN can adjust its weights to try to match its outputs to the target values. After training, the ANN is tested by giving it only input values, not target values, and seeing how close it comes to outputting the correct target values.
- In unsupervised learning, the ANN is not provided with the correct results during training. Unsupervised ANNs usually perform some type of data compression, such as clustering.”

Unsupervised ANNs generally are classification systems, while supervised ANNs are regression systems. Both are excellent tools for analysis of non-linear relationships embodied in physical system response. All ANNs can be considered to be data driven analysis engines.

In the earth sciences both types of learning algorithms have found widespread use. Unsupervised learning algorithms are known by various names: simply as “clustering” or “Kohonen Self-Organizing Maps,” although this latter is not strictly correct because Kohonen investigated many different types of ANNs. These types of networks are widely used to cluster well log or seismic data for the purpose of lithology identification and geobody classification. Early in this study Self Organizing Maps were used to assign facies type codes to depth intervals based on SOM results. This effort overlaps the work of other team members and was done for data investigation only.

Supervised learning algorithms use either feedforward or feedback network topologies. The most prevalent type are feedforward networks. In a feedforward network the connections between units do not form cycles. Most feedforward ANNs can be trained using a wide variety of efficient conventional numerical methods and usually produce a response to an input very quickly.

In feedback, or recurrent, ANNs there are cycles in the connections. In some feedback ANNs, each time an input is presented the ANN must iterate for a potentially long time before it produces a response. Feedback ANNs are usually more difficult to train than feedforward ANNs.

In this study, for the supervised learning algorithm using feedforward topology, the type of network learning method used is back-propagation. This refers to an iterative methodology of using the gradient of the case-wise error function computed on the output to adjust the synaptic weights in a feedforward network before the next iteration through the system. Mathematically this can be considered an application of the chain rule of elementary calculus.

According to the comp.ai-nn FAQ, “standard backprop can be used for both batch training (in which the weights are updated after processing the entire training set) and incremental training (in which the weights are updated after processing each case). For batch training standard backprop usually converges (eventually) to a local minimum, if one exists. For incremental training, standard backprop does not converge to a stationary point of the error surface. To obtain convergence, the learning rate must be slowly reduced. This methodology is called “stochastic approximation” or “annealing.” In this study batch training is used to analyze log data.

For any given problem, when a suitable network architecture has been devised and a network is trained the goal is often to use the network in a predictive mode to generalize to data not present during training. As Bishop puts it, “the goal of network training is not to learn an exact representation of the training data itself, but rather to build a statistical model of the process which generates the data. This is important if the network is to exhibit good generalization, that is, to make good prediction for new inputs.” In this study a neural network is trained on data from wells where the reservoir has been sampled by taking whole core sections. Then the trained network is used to predict core values in wells where no core exists.

Certain parameters associated with network architecture and the backpropagation method can cause a network to be ill behaved. Most notably, a network can either fix on local minimi/maxima and ignore the global min/max, or the network can converge rapidly to some minimum error function value and then begin to diverge, or it can “memorize” the training data set, making it able to faithfully reproduce the training data but completely unable to generalize.

For these reasons the usual practice is to validate a trained neural network using several methods. Commonly, data will be with held from the training data set and then used in a blind test to verify the networks ability to calculate the desired target data. In this method the testing data values are calculated after each iteration through the network and the user compares the resulting correlation coefficient to similar values for the testing data. A network that is converging well should exhibit similar correlation values in the training and testing data.

In addition, investigators will frequently hold out large blocks of data containing target values and use the trained network to calculate predicted target function values in batch mode. In contrast to the test data set calculated during network training, this test is a final step to show network ability to generalize before applying the trained network to data where target values do not exist. Divergence from actual measured data is used to judge the ability of the network to generalize.

Empirical rules-of-thumb exist to guide the division of data between training, testing, and verification sets. Of course, the division of data depends to some extent on the quantity of data available. In some smaller data sets it may not be possible to omit too much data from the training set. In this study, however, ample data exists, allowing a data split

that includes 60% of data patterns for network training, 20% for testing, and 20% for later validation.

For years engineers and scientists have struggled to model the physical universe using closed form, analytic expressions. The idea behind ANN technology, in contrast, is data driven. That is, given a body of observed data gathered from any given physical system the ANN seeks to find patterns in the data and to store knowledge about those patterns for later use. Given the same types of input data the ANN will be able to produce the output values that should be associated with the inputs.

For example, Table B-1 shows a simple data set consisting of seven patterns of well log data. For five of the patterns physical core samples are available. The mineralology interpretation of these samples is shown in the last columns. No cores were obtained for two lower depths. What is the likely composition at these depths?

By observing the data we see that GR values are higher in shale zones, and Resistivity values are lower. Likewise, GR values are relatively low in sand sections. Resistivity values depend to some extent on the fluids contained in the sands. If the fluid is oil the resistivity is high; water wet sand exhibits lower resistivity, but still higher than any values observed in shale zones.

By inspection, it seems likely that the zone at 2700 is an oil sand, and the zone at 2800 is a water wet sand. Although the values do not exactly correspond to the known data at 2200 and 2300, the *pattern* of low GR and higher resistivity intuitively leads us to these conclusions.

Table B-1 Example data set.

Index	**Log Data**		**Sample Data**
Depth	GR	Res.	Core Obs.
2000	100	50	Sh.
2100	90	70	Sh.
2200	20	5000	Oil Sand
2300	40	150	Wet Sand
2400	100	50	Sh.
2700	22	4700	?
2800	50	135	?

Most observers bring *a priori* knowledge to the table. As geoscientists we know that clean sands typically exhibit low gamma ray readings and that shales contain radio-active materials that cause high GR readings. The neural network does not have the

benefit of this experience and so demands a greater data volume than this simple example to learn the pattern of input and expected output.

On the other hand, the neural network can easily work with problems that are far too complex for most humans. If we increase the number of log types in the example above to, say, fifteen then most human observers would be confused by the similarity of many of the patterns. The problem has now been transformed from a two dimensional problem (GR & Res.) to a fifteen dimensional problem. The neural network can solve problems in this domain with ease.

REFERENCES

1. Bishop, C. M.: *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York City (1995)

2. Haykin, S.: *Neural Networks, A Comprehensive Foundation*, second edition, Prentice-Hall, Inc. Upper Saddle River, NJ. (1999)

3. Comp.ai-neuralnet newsgroup Frequently Asked Questions (FAQ). Available on the Internet.

Appendix C

ANN Input Dataset