DEVELOPMENT OF INFILL DRILLING RECOVERY MODELS FOR CARBONATE RESERVOIRS USING NEURAL NETWORKS AND MULTIVARIATE STATISTICAL AS A NOVEL METHOD

R. SOTO*, CH. H., WU‡, and A. M. BUBELA ‡

‡ Texas A&M University, College Station, Texas 77843-3116, USA
Ecopetrol - Instituto Colombiano del Petróleo, A.A. 4185 Bucaramanga, Santander, Colombia
E-mail: rsoto@ecopetrol.com.co

This work introduces a novel methodology to improve reservoir characterization models. In this methodology we integrated multivariate statistical analyses, and neural network models for forecasting the infill drilling ultimate oil recovery from reservoirs in San Andres and Clearfork carbonate formations in West Texas. Development of the oil recovery forecast models help us to understand the relative importance of dominant reservoir characteristics and operational variables, reproduce recoveries for units included in the database, forecast recoveries for possible new units in similar geological setting, and make operational (infill drilling) decisions. The variety of applications demands the creation of multiple recovery forecast models. We have developed intelligent software (Soto, 1998), Oilfield Intelligence (OI), as an engineering tool to improve the characterization of oil and gas reservoirs. OI integrates neural networks and multivariate statistical analysis. It is composed of five main subsystems: data input, preprocessing, architecture design, graphic design, and inference engine modules. One of the challenges in this research was to identify the dominant and the optimum number of independent variables. The variables include porosity, permeability, water saturation, depth, area, net thickness, gross thickness, formation volume factor, pressure, viscosity, API gravity, number of wells in initial waterflooding, number of wells for primary recovery, number of infill wells over the initial waterflooding, PRUR, IWUR, and IDUR. Multivariate principal component analysis is used to identify the dominant and the optimum number of independent variables. We compared the results from neural network models with the non-parametric approach. The advantage of the non-parametric regression is that it is easy to use. The disadvantage is that it retains a large variance of forecast results for a particular data set. We also used neural network concepts to develop recovery models. The neural network infill drilling recovery model is capable of forecasting the oil recovery with less error variance compared with non-parametric, fuzzy logic and regression models.

Keywords: *neural networks, carbonate reservoirs, recovery models, infill drilling*

* *A quien debe ser enviada la correspondencia*

Este trabajo introduce una metodología novedosa para mejorar los modelos de caracterización de yacimientos. En esta investigación se usaron técnicas de estadística multivariada y redes neuronales para desarrollar modelos de predicción de los recobros primarios de aceite (PRUR), recobros de aceite al inicio de la inyección de agua (IWUR) y recobros de aceite debido a la perforación de pozos de relleno (IDUR) en yacimientos de carbonatos localizados en el este de Texas. Los modelos desarrollados fueron comparados con los modelos de regresión no-lineal y con los de regresión no-paramétrica. Uno de los desafíos en esta investigación fue identificar las variables independientes dominantes y el número óptimo de estas. Para ello se desarrolló un sistema inteligente (Soto, 1998), Oilfield Intelligence (OI), que integra conceptos de componentes principales, análisis de factores y redes neuronales. OI está compuesto por cinco subsistemas: carga y preprocesamiento de los datos, diseño de la arquitectura de la red neuronal, diseño gráfico y una máquina de inferencia. El análisis multivariado de componentes principales permite resolver el problema de dimensionalidad. Cuántas y cuáles variables deberían usarse en la obtención de cada modelo. Después se utilizaron las redes neuronales para desarrollar modelos capaces de predecir los recobros primarios, de inyección de agua y debido a la perforación de pozos de relleno en las formaciones de carbonato de San Andrés y Clearfork en el este de Texas. Los coeficientes de correlación son del orden del 99% con errores absolutos no mayores del 3% comparados con coeficientes de correlación del orden de 0.91 y errores absolutos alrededor del 27% de otros modelos publicados internacionalmente en los últimos 15 años. Las variables consideradas en esta investigación fueron porosidad, permeabilidad, saturación de agua, profundidad, área, espesor total, espesor neto, factor volumétrico de formación, presión, viscosidad, gravedad API, número de pozos al inicio de la inyección de agua, número de pozos para la recuperación primaria, número de pozos de relleno al inicio de la inyección de agua, PRUR, IWUR, e IDUR. Obviamente el desarrollo de un modelo en redes neuronales que represente con alta precisión los datos requiere experiencia del ingeniero para realizar un control de calidad de los datos, determinar las variables dominantes y optimizar la estructura o topología de la red neuronal.

NOMENCLATURE

ACE	alternating conditional expectations	*OOIP*	original oil in place (MSTB)
API	API gravity of crude oil	*p*	index of training set
AREA	productive area (acres)	*PERM*	permeability (mD)
BASIN	basin index: 1 for San Andres, 2 for Clearfork	*POR*	porosity (fraction)
E	performance index	*PRESS*	initial reservoir pressure (Pa)
f	transformed function	*PRUR*	primary ultimate recovery (MSTB)
FVF	oil formation volume factor (RB/STB)	*RB*	reservoir barrel
GROSS	gross pay (m)	*STB*	stock tank barrel, oil production unit
IDUR	infill drilling ultimate oil recovery (MSTB)	*SW*	initial water saturation (%)
IWUR	initial waterflood ultimate oil Recovery (MSTB)	*VIS*	oil viscosity (Pas)
k	node index of output layer	*W*	weights
n	index of independent variables	*x*	independent variable
NET	net pay (m)	*y*	dependent variable
NOIW	number of wells after infill drilling	*z*	value of individual transform
NOPW	number of wells for primary Recovery	δ	error signal
NOWW	number of wells for initial Waterflooding	η	learning rate
PRUR	primary ultimate oil recovery (MSTB)	Δ*IWUR*	IWUR - PRUR (MSTB)
o	output value from hidden nodes	Δ*IDUR*	IDUR - IWUR (MSTB)

INTRODUCTION

The amount of oil that can be recovered from an oil reservoir is dependent on the reservoir characteristics, recovery method, number of wells, and operations efficiency. Waterflooding is often used as a secondary recovery process after the reservoir has been produced during primary recovery. Since most reservoirs are heterogeneous, infill drilling after an initial water flood permits production of oil from parts of the reservoir that might otherwise have been bypassed (Wu *et al.* 1992; Wu *et al.,* 1988; Shao *et al.,* 1994a; French *et al.,* 1991). Researchers and engineers have shown that the infill drilling in the West Texas carbonate reservoirs has indeed accelerated and increased oil recovery (Lu *et al.,* 1993, 1994). However, the dominant mechanisms and parameters (independent variables) that affect the oil recovery are not fully identified or may be subjected to varying degrees of uncertainty. To forecast or predict infill drilling recovery efficiency of an individual unit or a reservoir is a difficult task for most reservoir-engineering professionals.

Many approaches are used to evaluate infill drilling recovery efficiency. One is based on unit recovery efficiency and reservoir analysis, plus the reservoir engineer's experience and intuition. A more elaborate approach is the use of a "reservoir simulator", which requires detailed and yet uncertain input reservoir and production data. The approach we take here is statistical. It is based on an oil recovery efficiency database developed for specific producing formations in a particular geological basin.

The field data from waterflood units in the Permian Basin were gathered and used to develop the oil recovery forecast models. The database includes reservoir and production data of 21 units in the San Andres formation, and 23 units in the Clearfork formation. Estimation of ultimate and incremental infill drilling recovery has been a difficult task because of the limited data and the uncertainties in the independent variables (Malik *et al.,* 1993; Shao *et al.,* 1994b; Wu *et al.,* 1993, Wu *et al.,* 1997). Many attempts have been made to develop the infill drilling oil recovery forecast models with reference to reservoir rock and fluid properties. Non-linear regression, statistical analysis, and fuzzy logic were used to analyze the oil recovery data. While these reported models have improved over the years, they are not entirely satisfactory due to the inexact nature of the data set and the inherent limitations in the models themselves.

With the development of linear regression (Wu *et al.,* 1998; Shao *et al.,* 1994a) or statistical analysis (French *et al.*, 1991) models of infill drilling ultimate recovery (IDUR), a substantial effort was put into determining which independent variables are the most important and what is the optimum number of the independent variables. Some of these independent variables go beyond the basic reservoir and field properties. Additional parameters such as the primary ultimate recovery (PRUR) and the initial waterflood ultimate recovery (IWUR) were needed to succesfully forecast IDUR (Lu, 1993).

In this work non-parametric regression analysis and neural network modeling are used to develop forecast models that improve the degree of consistency and accuracy. The non-parametric approach is based on the work of Breiman and Friedman (1985). Xue and Datta-Gupta (1996) applied this approach to integrate seismic data in reservoir characterization. While the non-parametric regression analysis provided better identification of the dominant independent variables and more consistent forecast, it did not improve the variance of forecast results.

The application of neural networks for modeling nonlinear systems has been improved substantially in recent years (Nikravesh *et al.,* 1996; Bomberger *et al.,* 1996; Rogers and Dowla, 1994; Al-Kaabi *et al.,* 1990; Azimi-Sajadi and Liou, 1989; Johnson and Wichern, 1998). However, one of the problems that still had to be addressed was the determination of the dominant variables, and the optimum number of independent variables. Some intuitive insight and functional knowledge of the physical behavior of the system turned out to be helpful for identifying the dominant independent variables. A non-parametric regression analysis and the principal components and factor analysis of multivariate statistical analysis were used to identify the dominant independent variables so we could develop more efficient and realistic neural-networks models.

DATA SET

The input data are shown in Tables 1, 2, 3, and 4. Tables 1 and 2 list the reservoir and operational properties of the San Andres units. The dependent variables are primary ultimate oil recovery (PRUR), initial waterflood ultimate recovery (IWUR) and infill drilling ultimate recovery (IDUR). The labels and units of each variable are referred to the Nomenclature. During the data analysis and model development, it is found that IWUR is strongly dependent on PRUR; and IDUR on PRUR and IWUR. The sequential dependency complicates the development of dependable IDUR models. Table 3 and 4 list the properties of Clearfork units.

A REVIEW OF THE INFILL DRILLING RECOVERY MODELS

The updated non-linear standard regression forecast models for primary ultimate oil recovery (PRUR), initial waterflooding ultimate oil recovery (IWUR), and infill drilling ultimate oil recovery (IDUR) of San Andres and Clearfork units are summarized in Table 5. For each model, we included the independent variables, the coefficient of determination and the average absolute error. Figure 1 shows a comparison of actual and predicted PRUR. Apparently, the correlation between predicted values and actual values is good but the average absolute error is about 23.46%. Figures 2, 3 and 4 show residual plots for the statistical model to predict PRUR, IWUR, and IDUR. The residual plots do not have a constant variance and the upward trend suggests that the models may need additional terms. From these plots and the average absolute errors, we could conclude that it was necessary to search other modeling techniques.

NON-PARAMETRIC REGRESSION APPROACH FOR ESTIMATING OPTIMAL TRANSFORMATIONS FOR MULTIPLE REGRESSIONS

Non-parametric regression is one of the novel approaches to constructing a suitable model description from available information. It is developed to alleviate the problem of parametric regression that often leads to erroneous results caused by the mismatch between assumed model structure and the physical relationships of the actual data. In non-parametric regression we do not fix a priori the form of the dependency of the dependent variable on the independent variables. In fact, one of the main results of non-parametric regression is the form of the relationship.

Non-parametric regression is intended to build a model in the form,

$$y = f_0^{-1}(z_0) \tag{1}$$

where, the inverse transformation, f_0^{-1} ,and the transform sum of the independent variables, z_0 ,are selected to maximize the correlation between the right-hand and left-hand sides of the relation:

$$z_0 = z_1 + z_2 + \ldots + z_n \tag{2}$$

Table 1. San Andres Data I

	Field / unit	OOIP	Area	Depth	NET	GROSS	POR	SW	API	Vis	Fvf	Press	Perm
		MSTB	km^2	m	m	m	%	%		Pas·10^{-3}	RB/STB	MPa	mD
1	ADAIR "SA"	169,439	21.60	1463.04	15.24	32.00	14.1	35.0	34	1.6	1.12	12.93	3.7
2	FUHRAMN M/BL10 "GBSA"	78,383	24.82	1310.64	12.50	76.20	7.7	40.0	31	3.5	1.15	11.03	2.4
3	FUHRAMN M/BL9 "GBSA"	55,939	15.98	1356.36	12.50	76.20	7.0	30.0	29	3.3	1.10	11.03	4.0
4	JOHNSON /"GB" "SA"	63,003	15.05	1264.92	15.24	39.62	6.7	21.8	33	3.6	1.20	11.00	5.3
5	JOHNSON /"AB" "SA"	18,247	3.40	1249.68	18.29	45.11	8.0	30.0	39	1.3	1.20	17.24	1.8
6	LEVELLA/N CEN UN "SA"	131,981	45.53	1447.80	9.45	21.34	8.0	25.0	31	2.5	1.23	11.65	1.8
7	MABEE/JE MABEE 'A' 'SA"	279,112	52.73	1432.56	12.19	15.24	10.5	29.0	32	2.4	1.08	13.13	1.5
8	MEANS "SA"	376,693	57.98	1310.64	16.76	91.44	9.0	28.8	29	6.2	1.04	12.76	29.0
9	OWNBY "SA"	47,508	11.98	1584.96	9.75	25.91	14.1	38.1	32	1.5	1.35	12.41	4.5
10	OWNBY/BL GILSTRAP "SA"	3,643	0.65	1595.63	12.19	25.91	14.2	38.0	31	2.0	1.20	12.41	4.5
11	SABLE "SA"	33,331	5.42	1584.96	17.37	24.08	9.0	25.0	32	2.2	1.20	10.69	1.5
12	SEMINLE "SA"	1`154,378	63.53	1615.44	38.40	46.94	12.0	16.0	35	1.1	1.34	13.93	31.2
13	SHAFTER "SA"	184,381	44.84	1310.64	16.76	60.96	6.5	25.0	32	1.3	1.25	12.86	5.0
14	SLAUGHTER/IGOE SM "SA"	63,155	8.60	1502.66	14.94	36.58	11.2	14.1	32	1.4	1.23	11.79	5.0
15	TRIPLE-N "GB"	18,683	8.26	1318.26	6.10	6.10	12.1	40.0	32	1.8	1.23	14.68	6.6
16	WASSON/BENNET "SA"	394,925	28.44	1554.48	39.62	263.65	10.0	27.0	33	1.6	1.31	12.44	1.7
17	WASSON/CORNELL "SA"	182,409	7.78	1493.52	67.06	92.96	8.5	15.0	33	1.3	1.30	12.76	3.7
18	WASSON/DENVER "SA"	2`172,316	103.21	1463.04	42.98	88.39	12.0	15.0	33	1.8	1.31	12.44	5.0
19	WASSON/REBORTS "SA"	394,971	54.94	1493.52	20.73	77.72	8.5	15.0	33	1.6	1.31	12.44	5.0
20	WASSON/WILLARD "SA"	699,419	54.07	1554.48	39.62	60.96	8.5	20.0	32	1.8	1.31	12.44	1.5
21	WEST SEMINOLE "SA"	196,021	14.73	1558.14	35.97	64.01	9.9	18.0	32	1.0	1.38	13.93	20.8

subject to some constraints. The data transforms is calculated:

$$z_1 = f_1(x_1), \quad z_2 = f_2(x_2), \; ..., \quad z_n = f_n(x_n) \text{ and}$$
$$z_0 = f_0(y) \tag{3}$$

In this case the symbol $f_n(x_n)$ does not necessarily mean a certain algebraic expression. It is rather a relationship defined point-wise. The method of alternating conditional expectations (ACE) (Breiman and Friedman 1985), constructs and modifies the individual transformations in order to maximize the correlation in the transformed space. Certain trivial constraints (zero mean and unit variance for the individual transformations) assure that the solution is almost unique. To make the ACE algorithm really work, however, one has to imply some kind of restriction on the smoothness of the individual transformations, and this is done somewhat hidden, the way a certain "smoother" is used to construct and improve the transformations.

One of the great advantages of non-parametric regression is that it provides an insight into the influence of the individual variables. The shape of the point-to-point transformation is very informative, and the range of the transformed variable z_i tells a lot about the relative significance of the independent variables.

We used a dummy variable called BASIN with values 1 for San Andres and 2 for Clearfork. The coefficients of determination and the average absolute errors

Table 2. San Andres Data II

	Field / unit	NOPW	PRUR	NOWW	IWUR	ΔIWUR	NOIW	IDUR	ΔIDUR
			MSTB		MSTB	MSTB		MSTB	MSTB
1	ADAIR "SA"	109	21,398	130	44,352	22,954	178	65,401	21,049
2	FUHRAMN MASHO/BL10 "GBSA"	108	7,733	118	9,957	2,224	133	10,513	556
3	FUHRAMN MASHO/BL9 "GBSA"	77	6,354	136	8,435	2,081	158	10,343	1,908
4	JOHNSON /"GB" "SA"	83	9,690	116	15,528	5,838	149	17,413	1,885
5	JOHNSON /"AB" "SA"	15	1,548	38	3,842	2,294	93	5,812	1,970
6	LEVELLAND/N CEN UN "SA"	268	18,112	363	33,947	15,835	489	57,198	23,251
7	MABEE/JE MABEE 'A' 'SA"	290	34,316	592	74,618	40,302	620	88,786	14,168
8	MEANS "SA"	299	64,477	398	127,512	63,035	754	151,695	24,183
9	OWNBY "SA"	49	6,224	59	10,886	4,662	72	16,175	5,289
10	OWNBY/BL GILSTRAP "SA"	4	389	5	1,257	868	8	1,593	336
11	SABLE "SA"	37	4,209	64	9,203	4,994	71	11,094	1,891
12	SEMINLE "SA"	327	196,265	523	448,771	252,506	604	537,711	88,940
13	SHAFTER "SA"	258	21,039	326	32,961	11,922	369	35,201	2,240
14	SLAUGHTER/IGOE SMITH "SA"	42	9,044	82	25,786	16,742	97	27,732	1,946
15	TRIPLE-N "GB"	23	2,096	40	4,966	2,870	73	6,876	1,910
16	WASSON/BENNET "SA"	213	36,325	293	97,279	60,954	468	119,502	22,223
17	WASSON/CORNELL "SA"	71	21,718	92	64,338	42,620	128	67,765	3,427
18	WASSON/DENVER "SA"	386	207,826	593	383,090	175,264	1,417	943,060	559,970
19	WASSON/REBORTS "SA"	194	45,491	377	101,171	55,680	424	111,600	10,429
20	WASSON/WILLARD "SA"	223	59,029	304	102,844	43,815	461	178,31	75,468
21	WEST SEMINOLE "SA"	65	10,073	93	28,537	18,464	152	240,421	11,884

for PRUR, IWUR, and IDUR are 0.9786, 0.933 and 0.9436, and 21.2%, 28.1% and 29.4 %, respectively. Figure 5 shows a comparison of actual and predicted PRUR for non-parametric model using GRACE. Figure 6 shows the residual plot for IDUR to check for adequacy of each model. We can see from this plot that the errors do not have zero mean, neither constant variances, and the model underestimates the IDUR as the actual IDUR increases.

THE OILFIELD INTELLIGENCE

The neural network simulator developed in this research is called Oilfield Intelligence (OI) (Soto, 1998). The neural network simulator was built in MATLAB, a high-performance language for technical computing that integrates computation, visualization, and programming in an easy-to-use environment.

OI is composed of five main subsystems: Loading data, preprocessing, architecture design, graphic design, and inference engine modules. We wrote more than 1,200 lines of programming as M-files using MATLAB as a platform. Figure 7 illustrates the architecture OI and Figure 8 shows the graphical user interface (GUI) with each of these modules.

THE DOMINANT INDEPENDENT VARIABLES

A multivariate statistical analysis was also performed to determine the dominant independent variables with

Table 3. Clearfork Data I

	Field / unit	OOIP	Area	Depth	NET	GROSS	POR	SW	API	Vis	Fvf	Press	Perm
		MSTB	km^2	m	m	m	%	%		Pas·10^{-3}	RB/STB	Mpa	mD
1	DIAMOND M/JACK	3,004	1.29	966.22	10.36	32.00	7.0	40.0	30.5	2.4	1.18	11.03	8.0
2	DIAMOND M/McLA AC 1	6,574	2.91	966.22	9.75	19.81	7.0	38.0	30.5	2.4	1.18	8.27	3.0
3	DOLLARHIDE 'AB'	72,873	10.65	1981.20	20.73	108.81	8.9	18.0	37.0	0.6	1.39	19.93	8.4
4	FLANAGAN/CLEARFORK	81,812	19.63	1944.62	9.75	142.65	11.4	24.9	32.2	1.7	1.26	12.93	5.2
5	FULLERTON	1`032,853	119.55	2042.16	26.52	152.40	10.0	22.3	42.0	0.5	1.50	20.27	3.0
6	GOLDSMITH 5600/CA	610,244	61.51	1706.88	22.86	106.68	15.0	31.0	38.0	0.7	1.50	16.06	28.0
7	GOLDSMITH/LANDRETH	119,967	31.62	1691.64	11.89	105.16	9.6	26.0	39.0	0.5	1.40	16.06	2.6
8	LEE HARRISON/WEST	20,698	3.72	1478.28	13.41	25.60	12.5	42.0	25.0	8.7	1.10	13.79	4.0
9	MONAHANS	111,620	19.02	1402.08	18.29	182.88	10.0	25.0	37.0	8.1	1.47	15.17	2.0
10	NORTH RILEY "CF"	140,362	28.17	1920.24	19.81	20.12	7.7	33.0	32.0	2.6	1.29	19.03	12.0
11	OWNBY/UCFU	39,283	8.63	1988.82	23.77	78.94	5.0	30.0	27.0	1.7	1.15	16.55	1.2
12	SMYER/EAST	80,445	12.63	2042.16	25.60	213.36	7.0	35.0	29.0	1.8	1.15	16.55	7.7
13	PRENTICE /6700/6700 CLF	161,577	27.63	2042.16	22.25	212.14	8.2	41.4	28.0	1.7	1.15	16.55	3.0
14	PRENTICE/NE	51,362	8.09	1965.96	30.48	112.78	6.2	38.6	28.0	1.7	1.15	16.55	3.0
15	ROBERTSON/NORTH	274,757	19.00	1767.84	71.93	396.24	6.3	30.0	31.0	1.2	1.38	20.34	0.7
16	RUSSELL/7000 CFU	209,836	34.44	2240.28	30.78	93.57	5.3	24.0	34.7	0.8	1.28	17.93	1.0
17	SMYER/EAST	63,419	17.85	1767.84	10.97	33.53	8.3	33.0	26.5	5.8	1.08	14.48	3.4
18	SMYER/ELLWOOD "A"	81,877	17.48	1825.75	11.89	53.04	8.3	20.0	25.0	5.1	1.06	12.81	5.0
19	WASSON 72/GAINES	108,446	17.81	1729.74	25.91	231.65	6.4	27.0	35.0	1.0	1.25	17.93	1.0
20	WASSON 72/GIBSON	151,836	15.22	2011.68	51.51	220.98	5.5	30.0	31.0	1.5	1.25	18.62	0.5
21	WASSON 72/SOUTH	240,354	20.08	1950.72	41.76	382.52	7.7	26.0	32.0	1.4	1.25	17.93	5.5
22	WASSON 72/YOAKUM	80,032	29.95	1729.74	11.28	102.41	6.4	27.0	35.0	1.0	1.24	17.93	0.5
23	WASSON NE CF/NORTH	69,224	17.48	1950.72	24.69	71.32	5.1	35.0	30.0	1.5	1.30	18.22	0.2

reference to that investigated by the non-parametric regression analysis. Table 6 shows the results of a principal component analysis for PRUR. As can be seen, seven principal components could explain about 90% of the total variance of the data. To describe the possible relationships among the variables and determine if there is any possibility for grouping variables, we used the concept of factor analysis (Johnson and Wichern 1998). Table 7 shows an output of the rotated factor loading. The first factor grouped the primary ultimate oil recovery (PRUR) with the productive area and the number of primary recovery wells (NOPW) as indicated by the loading values. The first factor shows the dependence of the primary ultimate oil recovery on the productive area and the number of wells for primary recovery. According to this factor analysis, we decided to use area, because of the highest coefficient in factor1, as independent variable to predict PRUR. The second factor grouped BASIN, DEPTH and initial reservoir pressure (PRESS). For this group, we selected BASIN as independent variable because it was easier for the neural network to recognize that pattern. The third factor grouped the API gravity and FVF. The fourth factor shows that gross thickness (GROSS) has the highest

Table 4. Clearfork Data II

	Field / unit	NOPW	PRUR	NOWW	IWUR	ΔIWUR	NOIW	IDUR	ΔIDUR
			MSTB		MSTB	MSTB		MSTB	MSTB
1	DIAMOND M/JACK	5	346	9	585	239	17	786	201
2	DIAMOND M/McLA AC 1	11	513	18	620	107	33	851	231
3	DOLLARHIDE 'AB'	77	13,663	80	25,511	11,848	175	37,070	11,560
4	FLANAGAN/CLEARFORK CONS	93	12,307	105	27,333	15,026	110	29,987	2,653
5	FULLERTON	739	119,055	821	217,634	98,579	1,136	325,222	107,588
6	GOLDSMITH 5600/CA GLDSMITH	461	64,109	661	119,520	55,411	800	121,038	1,518
7	GOLDSMITH/LANDRETH (2)	191	28,936	195	46,281	17,345	260	65,440	19,159
8	LEE HARRISON/WEST	12	2,212	19	2,861	649	26	3,513	652
9	MONAHANS	63	5,245	124	12,176	6,931	235	20,933	8,757
10	NORTH RILEY "CF"	131	18,378	139	23,718	5,340	232	38,688	14,970
11	OWNBY/UCFU	42	3,729	43	7,079	3,351	69	9,472	2,393
12	SMYER/EAST	73	18,861	74	24,068	5,208	95	36,601	12,533
13	PRENTICE /6700/6700 CLFK	129	33,200	139	77,712	44,512	273	99,519	21,807
14	PRENTICE/NE	50	6,554	69	16,322	9,768	125	23,696	7,374
15	ROBERTSON/NORTH	104	27,841	124	35,682	7,841	361	67,538	31,856
16	RUSSELL/7000 CFU	185	38,902	198	54,866	15,964	304	62,680	7,814
17	SMYER/EAST	52	4,949	100	14,507	9,557	134	14,592	86
18	SMYER/ELLWOOD "A"	108	8,191	135	20,365	12,174	154	26,402	6,037
19	WASSON 72/GAINES	105	17,563	107	21,869	4,305	138	23,655	1,787
20	WASSON 72/GIBSON	85	12,798	96	14,968	2,170	118	21,973	7,004
21	WASSON 72/SOUTH	121	41,098	171	59,401	18,303	184	73,063	13,662
22	WASSON 72/YOAKUM	91	13,697	130	15,621	1,924	145	17,662	2,040
23	WASSON NE CF/NORTH	82	9,986	96	13,400	3,414	117	17,212	3,812

loading factor. Factors five and six are explained by water saturation (SW) and porosity (POR) variables respectively. Similar principal component and factor analysis was performed for IWUR and IDUR.

Results of the principal component and factor analysis indicate that the dominant independent variables for PRUR are: productive area (AREA), BASIN, API gravity, FVF, gross thickness (GROSS), and initial water saturation (SW). The dominant independent variables for IWUR are: productive area, the number of initial waterflood wells (NOWW), DEPTH, the number of primary wells (NOPW), API, FVF, net pay (NET), porosity (POR), permeability (PERM), and viscosity (VIS). The dominant independent variables for IDUR are: BASIN, productive area (AREA), and the number of infill wells (NOIW), PRESS, API, FVF, NET, VIS, SW, and POR. Those dominant independent variables were used to develop neural network infill drilling recovery models.

NEURAL NETWORKS

A neural network is a series of layers with nodes and weights that represent complex relationships among

Table 5. Current statistical oil recovery models of San Andres and Clearfork units (Soto, 1998)

Parameter	Model equation	R square	Average absolute error (%)
$PRUR_{\text{San Andres}}$	$PRUR = 10^{-3.012258}(OOIP)^{0.357853}(NOPW)^{0.721305}(NET)^{0.371451}(100-SW)^{1.120312}(VIS)^{0.238351}$	0.9823	25.71
$PRUR_{\text{Clearfork}}$	$PRUR = 10^{-3.21592}(OOIP)^{0.860267}(DEPTH)^{0.731893}(PERM)^{-0.091822}(POR)^{0.800064}(FVF)^{-3.29715}(VIS)^{-0.624600}$	0.9784	22.49
$IWUR_{\text{San Andres}}$	$IWUR = 10^{-0.989581}(PRUR)^{0.357853}(WSW)^{-0.386933}(100-SW)^{0.751919}$	0.9912	32.73
$IWUR_{\text{Clearfork}}$	$IWUR = 10^{-3.7981}(PRUR)^{0.497429}(WSW)^{-0.457180}(DEPTH)^{0.751919}(PERM)^{0.11601}(RGN)^{0.18648}$	0.9860	17.78
$IDUR_{\text{San Andres}}$	$IDUR = 10^{-4.882035}(IWUR)^{0.877953}(INOIW)^{0.232663}(DEPTH)^{1.461133}(GROSS)^{-0.095902}$	0.9929	22.27
$IDUR_{\text{Clearfork}}$	$IDUR = 10^{0.208079}(IWUR)^{0.943158}(INOIW)^{0.099273}$	0.9906	26.30

input and output variables. The first layer has input nodes representing the input variables (independent variables) specified by the problem. The node of a hidden layer uses the sum of the weighted outputs of previous layer and sigmoid function to provide output for the nodes in the subsequent hidden layer. The number of nodes in each layer and the weights are determined by trials and by optimization. The objective of the neural network is to obtain optimal weights to give a best value for the nodes (the dependent variable) of the output layer.

The advantages of the neural network approach are several. It does not require an explicit functional relationship between the input and output variables. It can be trained from past available data to learn and approximate the nonlinear relationships to any degree of accuracy. It is applicable to multivariate systems. One of the significant drawbacks of the neural network approach is that the input nodes must be specified a priori. The type of input variables and the optimal number of input variables can not easily be determined from neural network analysis.

Performance learning is one of the most important steps where network parameters are adjusted in an effort to optimize the performance of the network. Two steps are involved in this optimization process. The first step is to define a quantitative measure of network performance, performance index. It represents a global error *(E)* of the neural network defined as,

$$E = \sum_p \sum_k (y_{pk}^{real} - y_{pk}^{real}) \qquad (4)$$

Where the inner summation is over all nodes in the output layer, and the outer sum is over the number of the training set. During training, the size of error generally decreases until it reaches a threshold level.

The second step of the optimization process is to search the parameter space to reduce the performance index. A steepest descent algorithm is often used for the optimization. Mathematically, the weights *(W)* are adjusted as follows:

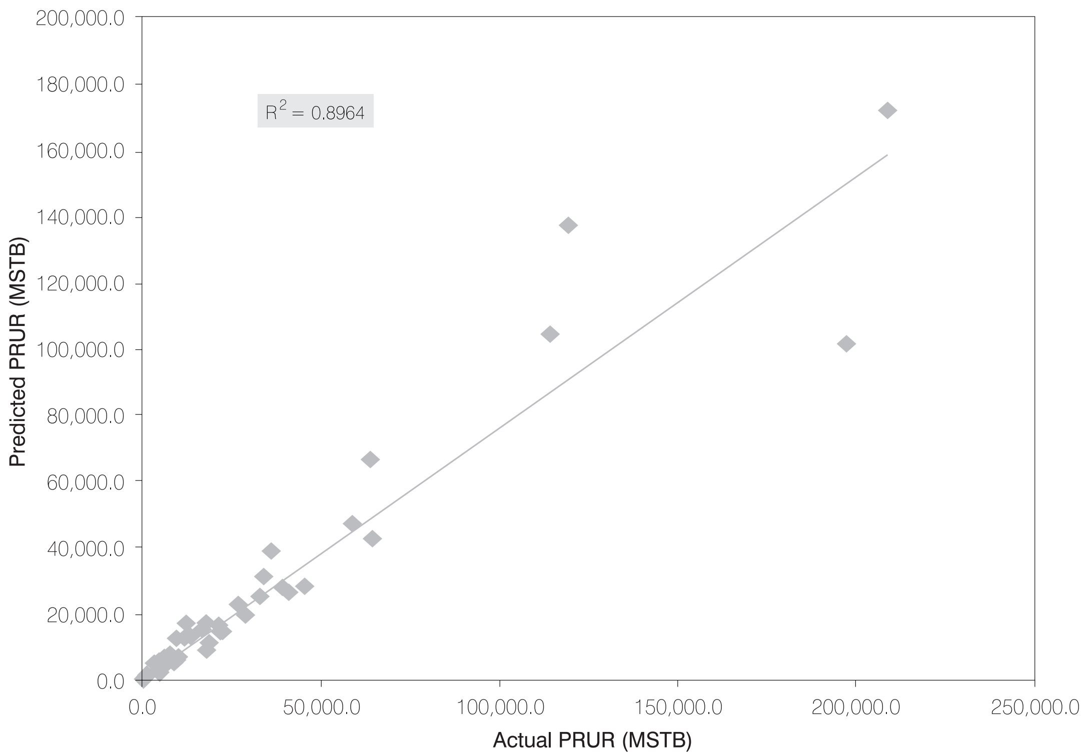

Figure 1. Comparison of calculated and measured PRUR for non-linear regression model.

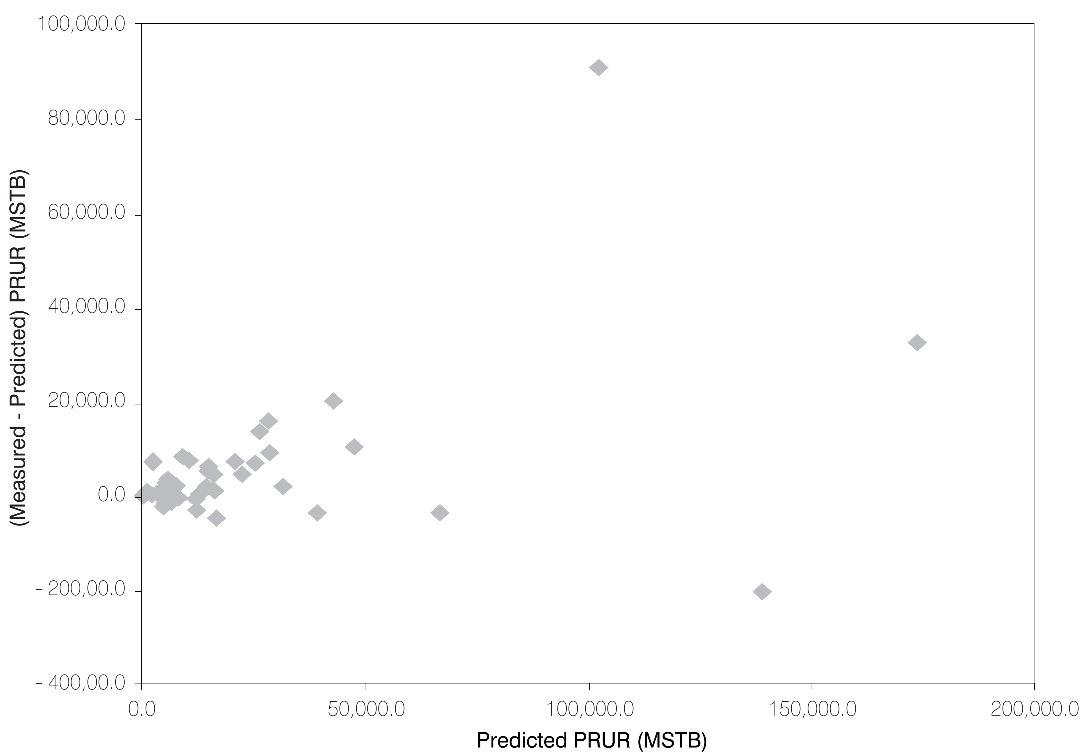

Figure 2. Residual plot for the non-linear regression model to predict PRUR

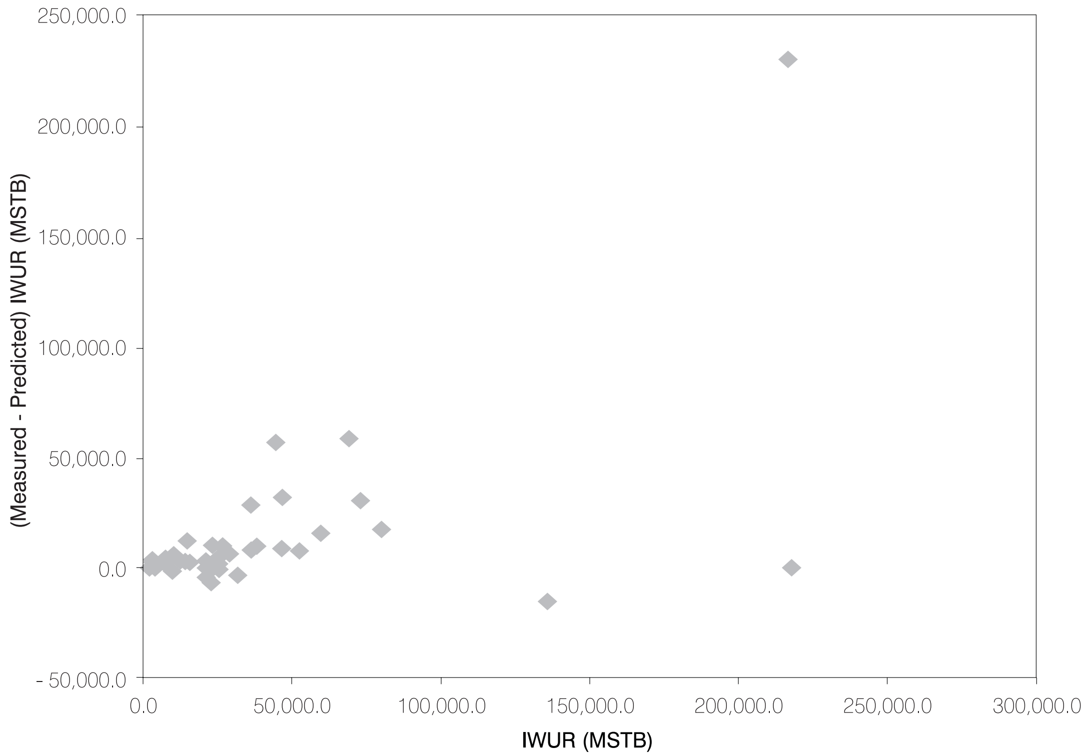

Figure 3. Residual plot for non-linear regression model to predict IWUR.

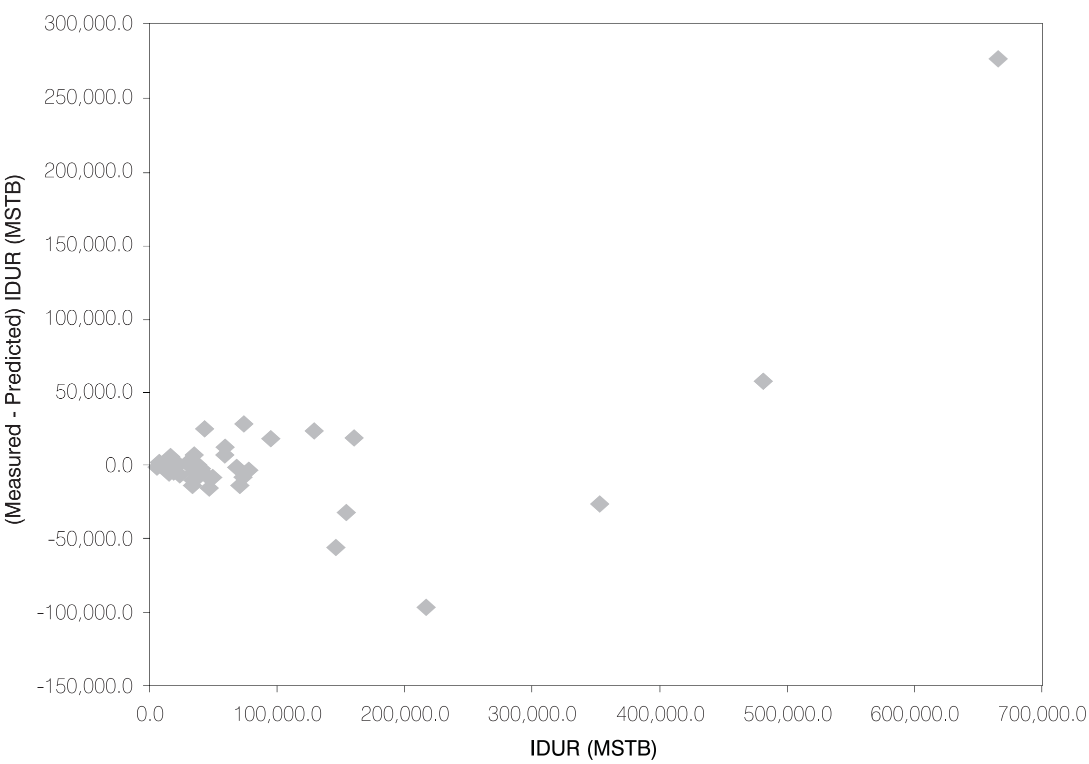

Figure 4. Residual plot for the non-linear regression model to predict IDUR

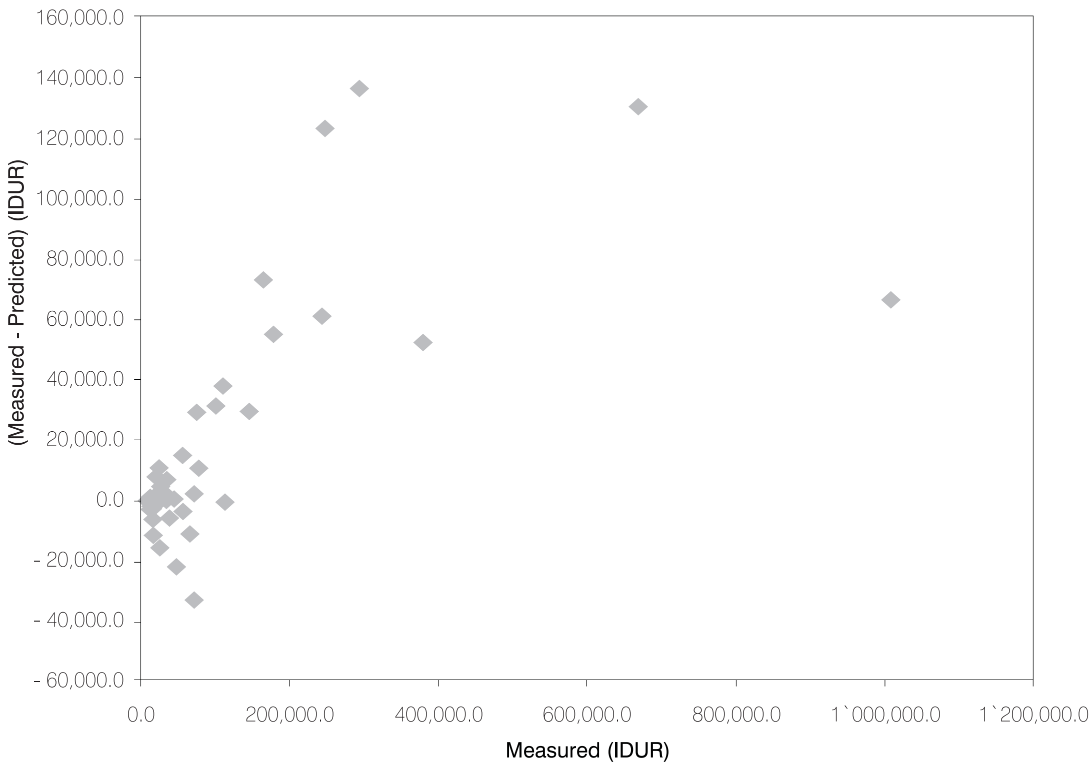

Figure 5. Comparison of calculated and measured PRUR for non-parametric pegression model

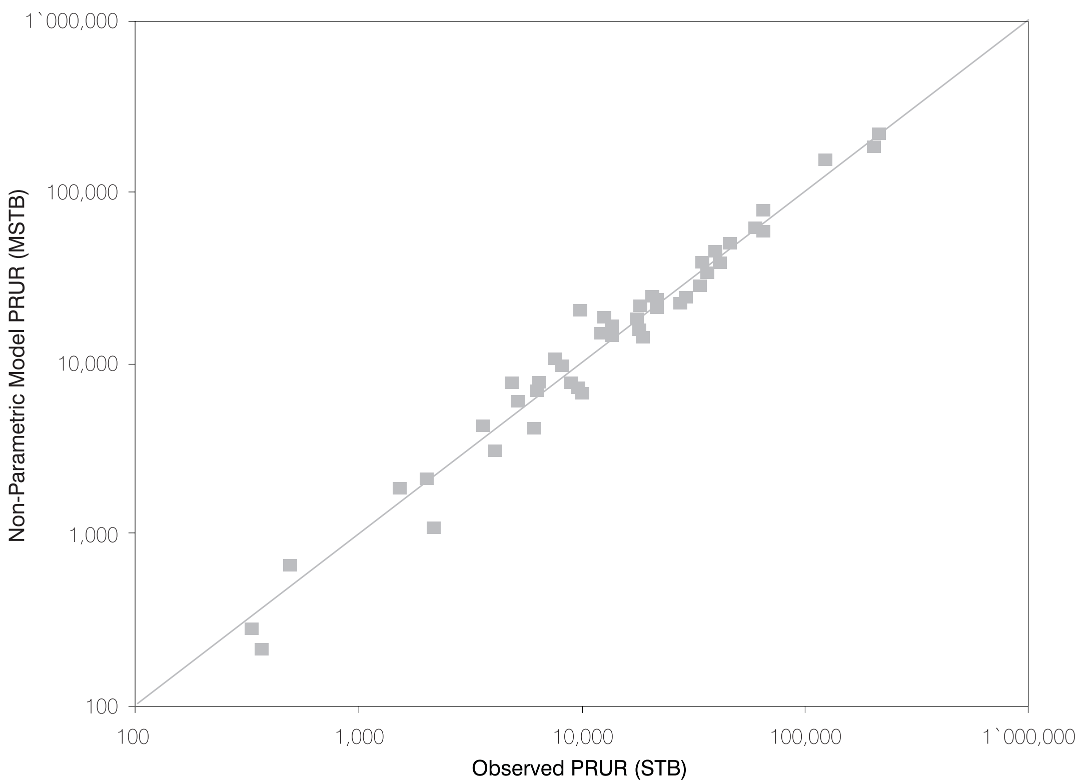

Figure 6. Residual plot for the non-parametric pegression model using GRACE to predict IDUR

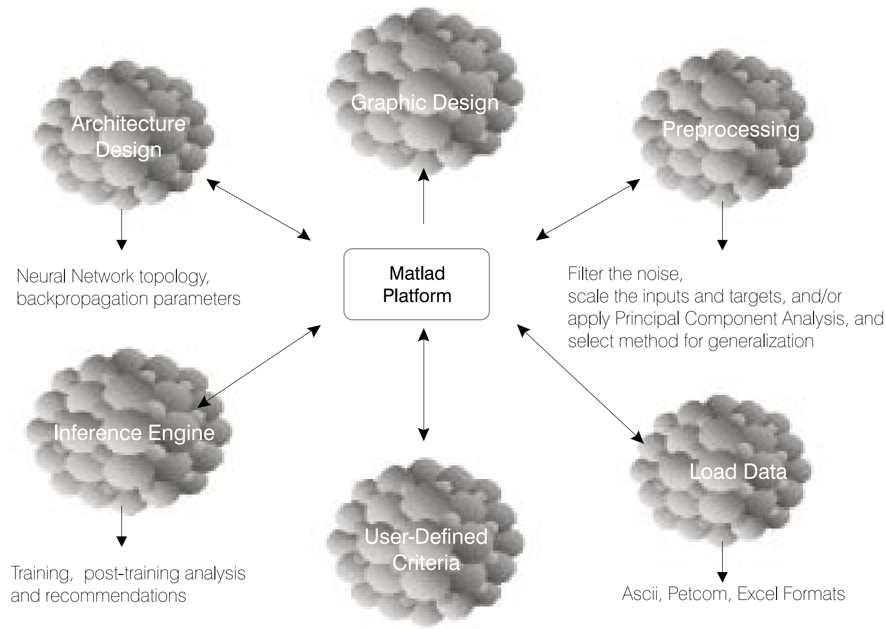

Figure 7. Architecture of oilfield intelligence

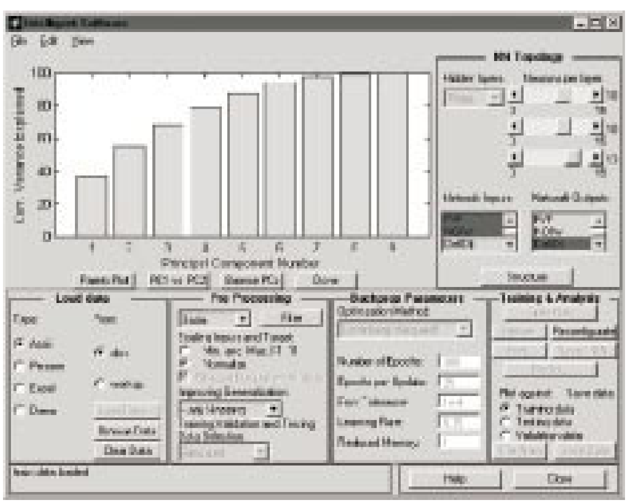

Figure 8. Graphical user interface of Oilfield Intelligence

$$W_{new} = W_{old} + \eta \cdot \delta \cdot o \qquad (5)$$

Where η is learning rate and o is the output value from hidden nodes and δ is the error signal term produced by the nodes in the hidden or output layer,

$$\delta = -\frac{\partial E}{\partial(\sum W_{old} \cdot o)} \qquad (6)$$

Before training the network, the learning rate (η) for the network must be specified. The new weights are then used to calculate the new output. The procedure is repeated until a tolerance is satisfied.

We used a backpropagation algorithm with the Levenberg-Marquardt procedure as an optimization method for convergence was used. The 44-sample data set was divided into three subsets for training set (72%), validation (14%) and testing (14%). A post-training evaluation of the performance of the trained neural networks is carried out by calculation of the errors for the training, validation, and testing data sets.

The final topology of the neural network for prediction of PRUR has 7 neurons (independent variables) in the input layer and two hidden layers with 12 and 10 neurons, respectively. For IWUR, the topology has 9 neurons in the input layer and two hidden layers with 10 and 6 neurons respectively. The topology of the neural network for IDUR has 9 neurons in the input layer, and two hidden layers with 10 and 8 neurons, respectively. After the neural networks were "trained," the weight and bias vectors were incorporated into Fortran-90 and Visual Basic interfaces so that the results could be used in a practical manner.

The neural network model showed very good performance for prediction of PRUR, IWUR, and IDUR.

Table 6. PRUR principal components

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	1.37501	0.48509	0.35456	0.35456
PRIN2	0.88991	0.49477	0.22947	0.58404
PRIN3	0.39514	0.12847	0.10189	0.68593
PRIN4	0.26667	0.05136	0.06876	0.75469
PRIN5	0.21531	0.02799	0.05552	0.81021
PRIN6	0.18732	0.02899	0.04830	0.85851
PRIN7	0.15832	0.02338	0.04083	0.89934
PRIN8	0.13494	0.04149	0.03480	**0.93414**
PRIN9	0.09346	0.02555	0.02410	0.95824
PRIN10	0.06791	0.02022	0.01751	0.97575
PRIN11	0.04769	0.02118	0.01230	0.98804
PRIN12	0.02651	0.00981	0.00684	0.99488
PRIN13	0.01669	0.01352	0.00431	0.99918
PRIN14	0.00317	0.00000	0.00082	1.00000

Table 7. PRUR rotated factor loadings

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7	FACTOR 8
BASIN	-0.01042	**0.56157**	0.13817	-0.11926	0.31755	0.09918	0.11858	-0.18657
AREA	**0.46452**	-0.00983	-0.06421	-0.07218	0.06722	0.08984	-0.17765	-0.09534
NET	-0.05270	-0.14476	-0.11316	**0.61773**	-0.05406	-0.08267	0.06656	0.05867
POR	-0.06108	0.08698	-0.00592	0.06310	0.05034	**1.02195**	-0.13180	0.01320
SW	0.10311	-0.07040	-0.03856	0.11391	**0.97638**	-0.18867	0.02639	0.06323
API	-0.03532	-0.08297	**0.58774**	-0.12251	0.00345	0.08644	-0.04098	-0.13480
VIS	0.03059	0.04800	0.05336	0.06179	0.00214	-0.17756	-0.06160	**0.49233**
FVF	-0.14050	0.01949	**0.51102**	0.08040	0.17243	-0.08363	0.01365	0.12214
DEPTH	0.00448	**0.56556**	-0.34269	-0.15889	-0.20299	-0.36244	-0.12011	0.40166
GROSS	0.01756	-0.07490	0.07738	**0.79419**	0.18190	0.28622	-0.09721	0.03603
PRESS	-0.01853	**0.32704**	0.08920	-0.03770	-0.05751	0.06433	0.09882	-0.04928
PERM	-0.10414	0.04018	-0.00424	-0.01917	-0.04296	0.04682	**0.32101**	-0.17247
NOPW	**0.42865**	0.00465	0.00249	-0.07670	0.02886	0.18458	-0.09693	-0.10831
PRUR	**0.31515**	-0.03005	-0.18314	0.12197	-0.03446	0.00673	0.07818	0.05876

The comparison of calculated and measured PRUR, IWUR and IDUR are presented in Figures 9 to 11. The coefficients of determination and the average absolute errors for PRUR, IWUR, and IDUR are 0.998,

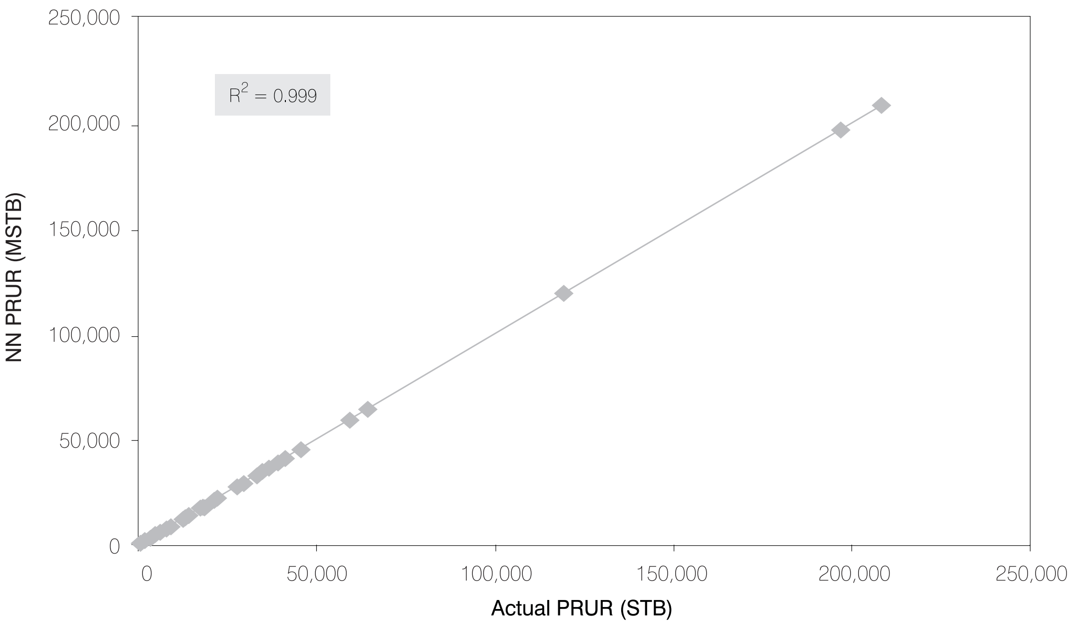

Figure 9. Calculated PRUR from neural network model versus actual PRUR

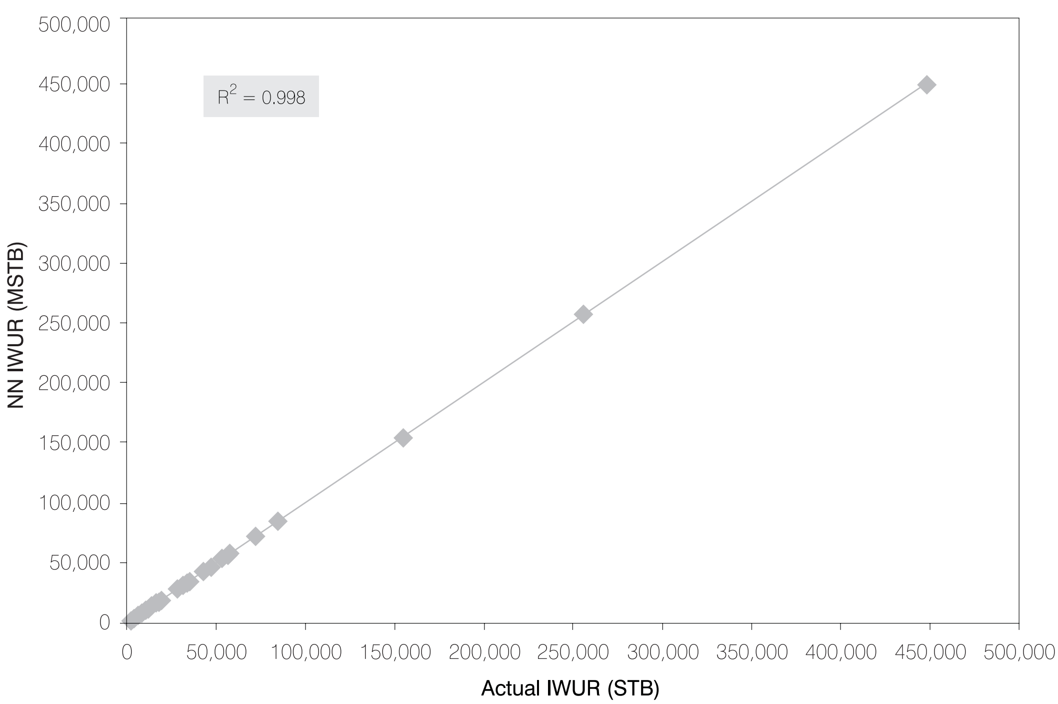

Figure 10. Calculated IWUR from neural network model versus actual IWUR

0.992 and 0.9995, and 1.0%, 2.1% and 3.3%, respectively. Using the same scales of the residual plots for non-parametric regression models, we plotted residual plots for the neural network models of PRUR, IWUR, and IDUR (Figure 10 shows an example). Checking the adequacy of each model we can see from this plot that the errors do have zero mean and constant variance.

The dominant independent variables identified for each model are used to develop the neural network oil recovery models. A series of sensitivity analysis with

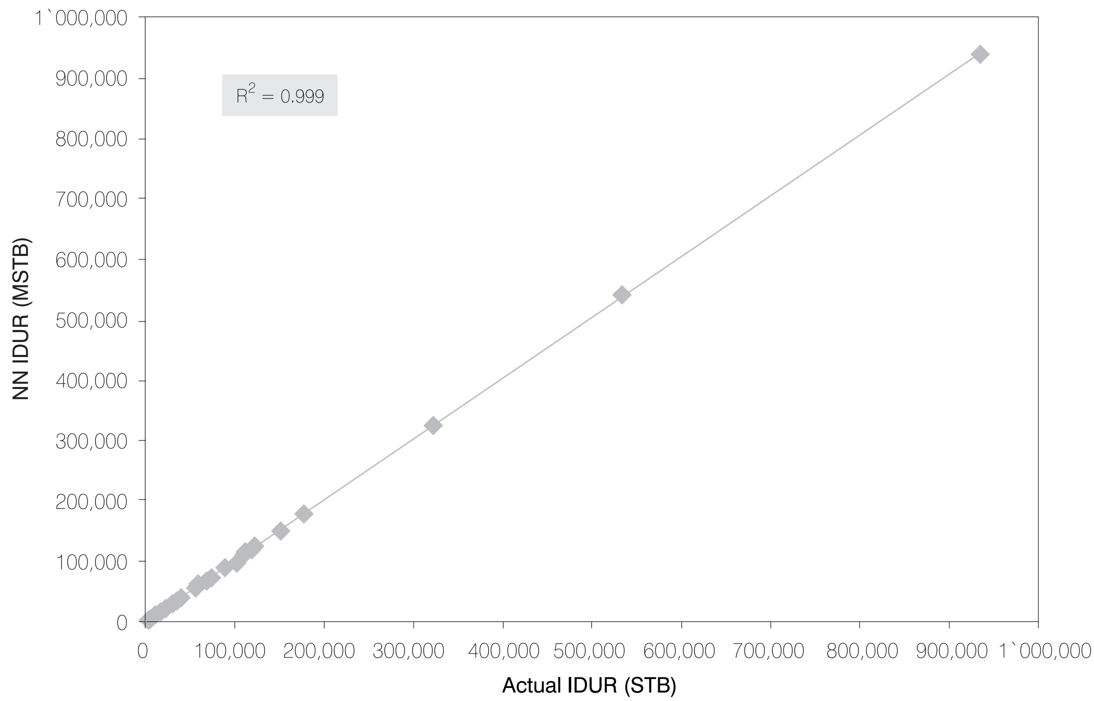

Figure 11. Calculated IDUR from neural network model versus actual IDUR

different neural network topologies is performed to develop the best neural network models. The approach helped eliminate over-fitting and meaningless dependency of certain independent variables. For the sensitivity analysis of variable dependency, we made a series of runs for each basin (San Andres and Clearfork) by varying the value of each independent variable while keeping other independent variables at individual mean. As an example, Figure 12 shows the dependency of calculated IDUR on the productive area. The monotonely increasing relationship indicates physically meaningful dependency of the variables. Figure 13 shows the dependency of the calculated IDUR on the number of infill wells. Similar sensitivity analysis was made for other independent variables. Table 8 shows a summary of the oil recovery forecast model performance.

CONCLUSIONS

- The correlation coefficients of the non-linear regression models for predicting the infill ultimate oil recovery for both San Andres and Clearfork carbonate formations in West Texas apparently are good but the average absolute error is about 23.46%

- One of the significant constraints for the model development is the limited number of field data that are inexact and often exhibit uncertain relationships. Principal components and factor analysis help understand the relative importance of dominant reservoir characteristics and operational variables to improve the modeling.

- The advantage of the non-parametric regression is that it is easy to use and can quickly provide results that reveal the dominant independent variables and relative characteristics of the relationships. The disadvantage is retaining a large variance of forecast results for a particular data set. The average absolute errors for PRUR, IWUR, and IDUR are 21.2%, 28.1% and 29.4%, respectively. The residual plots showed that the errors do not have zero mean, nor constant variances.

- Multivariate principal component and factor analyses were employed to develop an effective neural network. The neural network infill drilling recovery model is capable of forecasting the oil recovery with less error variance. The average absolute errors for PRUR, IWUR, and IDUR are 1.0%, 2.1% and 3.3% respectively. The residual plots showed that

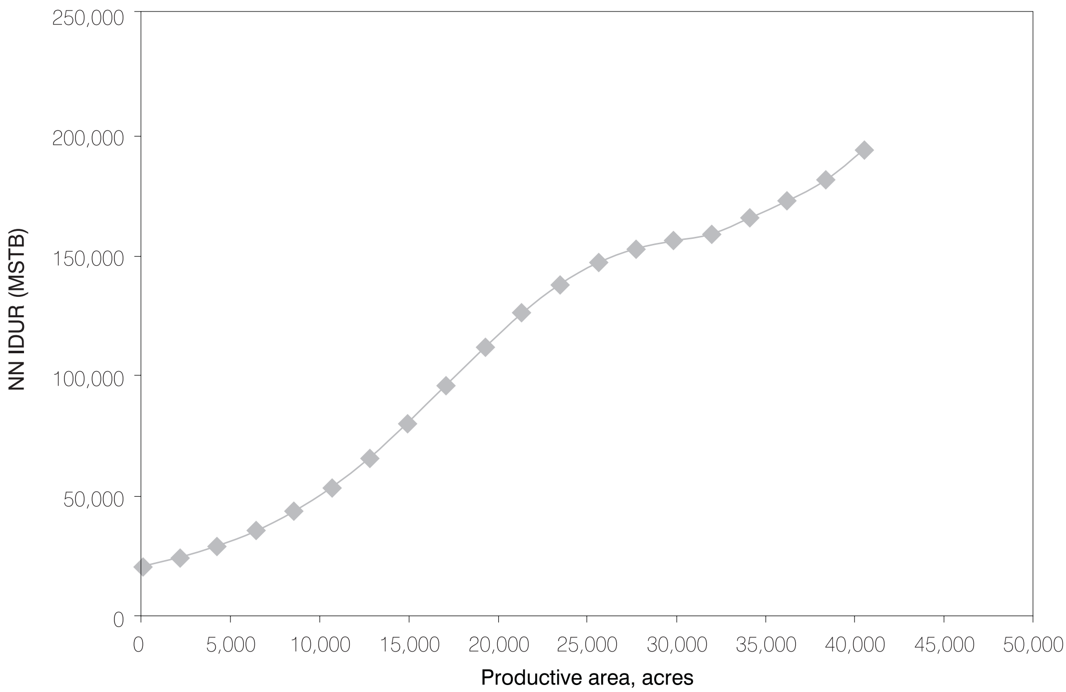

Figure 12. Effect of productive area on IDUR predicted by the neural network model for San Andres units

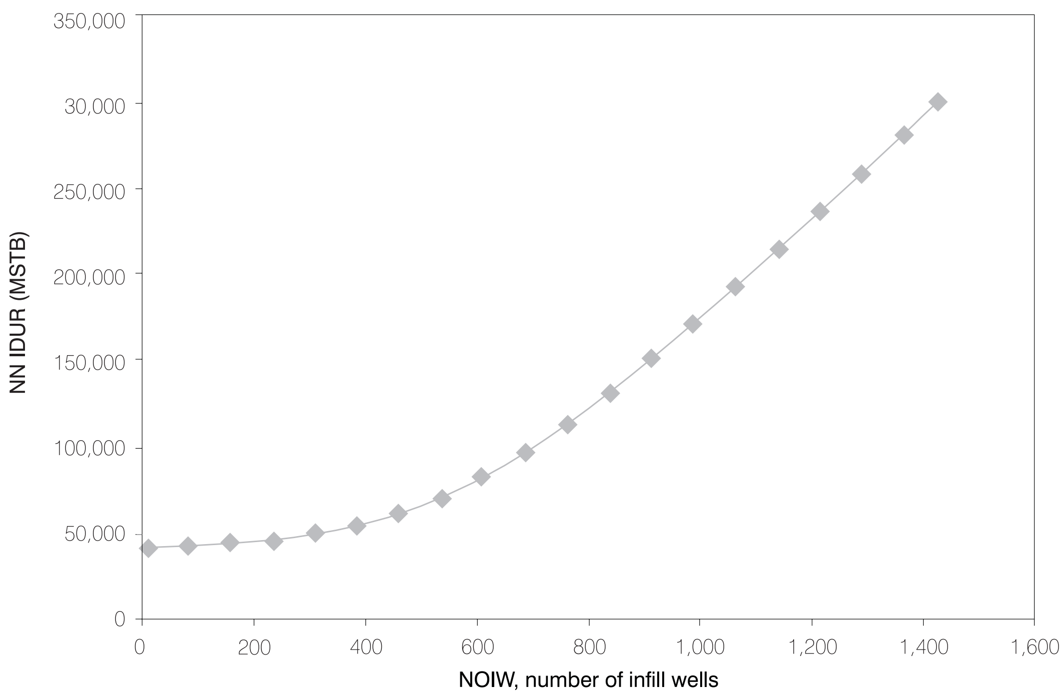

Figure 13. Effect of number of infill wells over that of initial waterflooding on IDUR predicted by the neural network model

Table 8. Summary of oil recovery forecast model performance (Wu, 1997)

Method	Parameter	Independent variables	Correlation coeff between measured and calc parameter	Average absolute Error%
Standard Regression	$PRUR_{San\ Andres}$	Area, Porosity, Sw, h_{net}, NOPW, Viscosity	0.9226	24.49
Standard Regression	$PRUR_{Clearfork}$	Area, Porosity, Sw, h, Depth, Viscosity, Permeability, FVF,	0.9472	21.51
Standard Regression	$IWUR_{San\ Andres}$	PRUR, WSW, Sw	0.9056	34.17
Standard Regression	$IWUR_{Clearfork}$	PRUR, NOWW, Depth, h_{gross}, Permeability, h_{net}	0.9682	17.01
Standard Regression	$IDUR_{San\ Andres}$	IWUR, NOIW, Depth, Gross	0.9649	21.20
Standard Regression	$IDUR_{Clearfork}$	IWUR, NOIW,	0.9221	26.30
Neural Network	$PRUR_{San\ Andresand\ Clearfork}$	Area, Basin, API FVF, Sw, Porosity	0.999	1.0
Neural Network	$IWUR_{San\ Andresand\ Clearfork}$	Area, NOWW, PRUR, API, FVF, hnet, Porosity, Sw, Viscosity, Basin	0.998	2.1
Neural Network	$IDUR_{San\ Andresand\ Clearfork}$	Area, NOIW, IWUR, Basin, API, FVF, h_{net}, Viscosity, Sw, Porosity	0.999	3.3

the errors do have zero mean and constant variances.

- The novel methodology applied in this research can be used to get better models for a reservoir characterization.

ACKNOWLEDGMENT

This material is based in part upon work supported by the Texas Advanced Technology Program under Grant No. ATP-036327-066 1998. The Instituto Colombiano del Petroleo support for part of the works is also appreciated. We would like to acknowledge Peter Valko's contribution to the development and application of multiple forecast models.

REFERENCES

Al-Kaabi, A.U., McVay, D.A. and Lee, W. J., 1990. "Using an Expert System to Identify a Well-Test Interpretation Model", *J. Pet. Tech.* (May): 654 - 661.

Azimi-Sajadi, M. and Liou, R. J., 1989. "Fast Learning Process of Multilayer Neural Networks Using Recursive Least Squares Method", *IEEE Trans. On Sig. Proc.*, 40 (2): 446 - 449.

Bomberger, J. D., Seborg, D. E., Lightbody, G., and Irwin, G. W., 1996. "Experimental Evaluation of Neural Nonlinear Modeling", *presented at the Chemical Process Control V*, Tahoe City, CA. (26 January).

Breiman, L. and Friedman, J. H., 1985. "Estimating Optimal Transformations for Multiple Regression and Correlation", *J. of the American Statistical Associations* (September): 580.

French, R. L., Brimhall, R. M., and Wu, C. H., 1991. "A Statistical and Economic Analysis of Incremental Waterflood Infill Drilling Recoveries in West Texas Carbonate Reservoirs," *SPE 22624 presented at the SPE Annual Technical Conferences and Exhibition*, Dallas (Oct. 6 -9).

Johnson, R. A. and Wichern, D. W. 1998. *Applied Multivariate Statistical Analysis*, Fourth Edition, Printice-Hall, Inc., Upper Saddle River, NJ.

Lu, G. F., Brimhall, R. M. and Wu, C. H., 1993. "Geographical Distribution and Forecast Models of Infill Drilling Oil Recovery for Permian Basin Carbonate Reservoir", *SPE 16503 poster session at the nnual Technical Conference and Exhibition, Houston* (Oct. 3 - 6).

Lu, G. F., Jagoe, B. L. and Wu, C. H., 1994. "Technical Factors Useful for Screening Reservoirs for Water-flood Infill Drilling", *SPE 27660 (poster) presented at the Permian Basin Oil & Gas Conference*, Midland (March 16 - 18)

Malik, Z. A., Silva, B. A., Brimhall, R. M. and Wu, C. H., 1993. "An Integrated Approach to Characterize Low Permeability Reservoirs Connectivity to Optimize Waterflood Infill Drilling", *SPE 25853 presented at the Joint Rocky Mountain Regional and Low Permeability Reservoirs Symposium,* Denver (April 26 - 28).

Nikravesh, M., Kovscek, A. R., Johnston, R. M. and Patsek, T. W., 1996. "Prediction of Formation Damage During Fluid Injection Into Fractured, Low Permeability Reservoirs via Neural Networks", *paper SPE 31103 presented at the SPE Formation Damage Symposium*, Lafayette, LA. (February 16 - 18)

Nikravesh, M., Kovscek, A. R., Patzek, T.W. and Soroush, M., 1996. "Identification and Control of Industrial Scale Processes via Neural Networks", *presented at the Chemical Process Control V, Tahoe City*, CA. (January 26).

Rogers, L.L. and Dowla, F. U., 1994. "Optimization of Groundwater Remediation Using Artificial Neural Networks with Parallel Solute Transport Modeling", *Water Res. Res.*, 30 (2): 457 - 481.

Shao, H., Brimhall, R. M., Ahr, W. M. and Wu, C. H., 1994a. "Integrated Recovery Efficiency Forecast Models for San Andres Reservoirs of Central Basin Platform and Northern Shelf, West Texas", *SPE 27697 (poster) presented at the Permian Basin Oil & Gas Conference*, Midland (March 16 - 18).

Shao, H., Brimhall, R. M., Ahr, W. M. and Wu, C. H., 1994b. "An Application of Idealized Carbonate Depositional Sequences for Reservoir Characterization", *SPE 28458 (poster) presented at the Annual Technical Conference and Exhibition*, New Orleans (Sept. 25-28).

Soto B., R., 1998. "Reservoir Characterization Using Core, Well Log, and Seismic Data and Intelligent Software", *Dissertation*, Texas A&M University, College Station, Texas.

Wu, C. H., Lu, G. F., Xue, G. P. and Brimhall, R. M., 1993. "Empirical Oil Recovery Forecast Models for Waterflood Infill Drilling in West Texas Carbonate Reservoirs", *presented at the Southwestern Petroleum Short Course*, Texas Tech. University, Lubbock, Texas (April 21 - 22).

Wu, C. H., Lu, G. F., Gillespie, W. and Yen, J., 1997. " Statistical and Fuzzy Infill Drilling Models for Carbonate Reservoirs", *paper SPE 37728 presented at the SPE Middle East Oil Show & Conference*, Bahrain (March 15 - 18)

Wu, C.H., 1992. "Infill Drilling Use Improves Waterflood Oil Recovery", *The American Oil & Gas Reporter* (61 March).

Wu, C. H., Laughlin, B. A. and Jardon, M., 1988. "Impact of Infill Drilling on Waterflood Recovery: West Texas Carbonate Reservoirs", *SPE 17286 presented at the SPE Permian Basin Oil and Gas Recovery Conference*, Midland (March 10-11).

Xue, G. and Datta-Gupta, A., 1996. "A New Approach to Seismic Data Integration During Reservoir Characterization Using Optimal Non-Parametric Transformations", *paper SPE 36500 presented at the SPE Annual Technical Conference*, Denver (October 6 - 9).

Xue, G., Datta-Gupta, A., Valkó, P. and Blasingame, T., 1996. "Optimal Transformations for Multiple Regression: Application to Permeability Estimation From Well Logs", *paper SPE 36500 presented at SPE Improved Oil Recovery Symposium*, Tulsa (April 21).