# Determination of the porosity for the clastic sedimentary grain and the magmatic basement rocks of Cuulong basin from well log data using the artificial neural networks 

Ha Song Dang ${ }^{1, *}$, An Hai Le ${ }^{2}$<br>${ }^{1}$ Exploration Dept of PVEP POC, Graduate student Faculty of Geology - VNU University of Science, Vietnam<br>${ }^{2}$ Hanoi University of Mining and Geology, Vietnam

## ARTICLE INFO

## Article history:

Received 18 Oct. 2016
Accepted 18 Mar. 2017
Available online 30 June 2017

## Keywords:

Artificial neural network
Well log data
Magmatic basement rocks
Sedimentary clastic grain


#### Abstract

The porosity is the most important parameter in investigation of the oil bodies and during the production process. The existing softwares can only calculate the porosity with at least 6 accurately recorded well log curves. In fact, this requirement is very difficult to meet. This study presents a method to calculate the porosity of the magmatic basement rocks and the clastic sedimentary grain of Cuulong basin from well log data by using artificial neural network (ANN). Six curves must be collected and accurately recorded to meet the requirements of the existing softwares is very difficult. But each segment has 4 curves collect and accurately record to meet the requirement of this study is easy. The actual testing results on the wells are calculated by other softwares and the 15 drilled wells are used for calculating the porosity in this study. The Exploration Group Japan Vietnam Petroleum Company LTD (JVPC) has used the results of calculations of porosity by ANN of this study for 15 drilled wells in order to build the mining production technology diagrams. The other softwares are not able to calculate of 15 these drilled wells. It shows that the artificial neural network model of this research is a great tool for calculating porosity.


Copyright © 2017 Hanoi University of Mining and Geology. All rights reserved.

## 1. Introduction

The magmatic basement rocks and the sedimentary clastic grain (Oligocene, Miocene) are the large objects containing oil and gas in Cuulong basin. Many drilled wells above is sedimentary data, the lower is the magmatic

[^0]basement rock data. So investigations of both these objects are the necessary requirement.

The Cenozoic sediment unconformably covers up the weathering and eroded crystalline basement rocks. The oil body in the sedimentary clastic grain has many thin beds with the different oil-water boundaries. The oil body has small size (Hoàng Văn Quý, 2014). The mineral composition of the sedimentary rocks is consisted of: quatz, felspar kali, plagioclase, shale mineral kaolinite,
montmorilonite, chlorite, heavy mineral as muscovite, pirit.

The pre-Cenozoic fractured basement rocks consist of the old rock bodies and metamorphic sediment, carbonate rocks, intrusive formed before the formation of sedimentary basins. The oil body in the pre-Cenozoic fractured basement rocks has block shape, large size (Hoàng Văn Quý, 2014). The lower boundary is the rough surface, dependent on the development features of the fractured system.

The oil body has the complex geological structures, is the non traditional oil body. Mineral composition is consisted of albite (plagioclase), biotite (mica group), hornblend amphibol, orthoclase K-felds, quartz. These specific features created serious difficulties for investigation of the porosity (Hoàng Văn Quý, 2009). The existing softwares can be used for calculation of the porosity only when adequate and virtuous well $\log$ for 6 curves. This condition is very difficult to meet in practice. Many foreign contractors such as JVPC, etc... usually face to this difficulty. Therefore, figuring out a new method to calculate porosity is urgently needed.

This study has developed the ANN method to calculate the porosity. The purpose of this study is to calculate porosity for the sedimentary clastic grain and the pre-Cenozoic fractured basement rocks of Cuulong basin from well log data via using the ANN with adequate accuracy when the well log data are not complete and/or the material is of bad quality as usually found today.

## 2. Artificial neural network

### 2.1. Artificial neural network

The ANN is a mathematical model of the biological neural network to solve a specific problem. By connecting input and output of the neurals together, we would have a neural network (Bùi Công Cường, 2006). In the net, neurons are distinguished by their locations, specifically:

Input layer: The neurons receive information from outside the network. They are located outside the "left" and communicate with the neurons of hidden layer.

Output layer: Group of the neurons are connected to the other neurons through the
neurons of hidden layer. They stay in the position outside the "right" to translate the signal to the outside.

Hidden layer: The remaining neurons do not belong the two above layers.

The network is divided into layers. The neurons in the same layer have the same function. Although the neural network can consist of multiple hidden layers, LiminFu (Limin Fu et al., 1994) demonstrated that only one hidden layer is sufficient to model any function. So, the networks only need three layers (input layer, hidden layer and output layer) to operate. The following figure is an artificial neural model, which includes $n$ inputs of $x_{1}, x_{2} \ldots \ldots x_{n}$ and 1 output $y$ (Girish Kumar Jha, 2012)

### 2.2. Database



Figure 1. Model of one artificial neural.
These are 6 curves (GR, DT, NPHI, RHOB, LLD, LLS) of the sedimentary clastic grain in the drilled well KNT3X:

GR (API ): Gamma Ray log ; DT (.uSec/ft) : Compresional sonic transit time; NPHI (dec): Neutron log; RHOB (gm/cc): bulk density log; LLD (ohm.m) : Deep Resictivity; LLS (ohm.m ) : MedResictivity.

Nevertheless, the data collected from the roof to the bottom of the well, rarely are adequate and good enough to fully satisfy the calculating conditions of the existing softwares. From the top to bottom of the wells, many segments of recorded curves have been broken, and mostly only 4 to 5 curves have been recorded. The actual obtained data is difficult to meet the requirements of the existing softwares, however, these data easily to meet the requirements of the artificial neural network method of this study.

Table 1. Database: Well $\log$ of KNT3X

| Line | Depth | GR (API) | DT ( $\mu$ s/fit) | NPHI (dec) | RHOB (g/cm $\left.{ }^{3}\right)$ | LLD (Ohm.m) | LLS (Ohm.m) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  |  |  |  |  |
| 1 | 3864.2546 | 80.3320 | 69.3297 | 0.1878 | 2.4257 | 15.9866 | 4.5653 |
| 2 | 3864.4070 | 77.5129 | 67.8614 | 0.1724 | 2.4027 | 16.8112 | 2.6569 |
| 3 | 3864.5594 | 73.8561 | 68.4874 | 0.1700 | 2.3907 | 18.0029 | 1.3123 |
| 4 | 3864.7118 | 69.3158 | 71.0964 | 0.1808 | 2.3939 | 18.3438 | 1.1473 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| 1562. | 4290.0602 | 43.3562 | 66.1937 | 0.0528 | 2.4786 | 19.8068 | 17.2769 |
| 1563. | 4290.2126 | 50.8247 | 66.7832 | 0.0552 | 2.4818 | 18.3849 | 13.3551 |
| 1564. | 4290.3650 | 57.8918 | 67.2986 | 0.0576 | 2.4903 | 13.8037 | 13.0863 |

### 2.3. Network development:

The authors develop the artificial neural network that consists of 3 layers :

- Input layer consists of $n$ neurons: $x_{1}, x_{2}, \ldots x_{n}$.
- Hidden layer consists of $k$ neurons and the transfer functions $f_{j}(x)$ with $j=1,2 \ldots, k$.
- Output layer consists of $m$ neurons and the transfer functions $F_{l}(x)$ with $l=1,2, . ., m$.

Each neuron is a unit of account with many inputs and one output (Nguyễn Doãn Phước, Phan Xuân Minh, 2006). Each neuron has an energy of its own called its bias threshold, and it receives the energy from other neurons with different intensity as the corresponding weight. Neuron $j$ of the hidden layer has the bias threshold is $\omega_{H j}$, the value of neuron $j$ of the hidden layer receive from the input layer is $\sum_{i=1}^{n} \omega_{i j}^{1} \cdot x_{i}$ (Nguyễn Doãn Phước, Phan Xuân Minh, 2006). So it's value is $\omega_{H j}+\sum_{i=1}^{n} \omega_{i j}^{1} \cdot x_{i}$, where $\omega_{i j}^{1}$ are weight. With the transfer function $f_{j}(x)$, it's value will output is $f_{j}\left(\omega_{H j}+\sum_{i=1}^{n} \omega_{i j}^{1} \cdot x_{i}\right)$. This value is sent to the output neurons $l$ with $l=1,2, . ., m$ with weights $\omega_{j l}^{2}$, so the value of neuron $l$ of the output layer is
$b_{o l}+\sum_{j=1}^{k} \omega_{j l}^{2} . \mathrm{f}_{\mathrm{j}}\left(\omega_{H j}+\sum_{i=1}^{n} \omega_{i j}^{1} x_{i}\right)$, where $b_{o l}$ is the bias threshold of the output neuron $l$. With transfer function $F_{l}(x)$, the value of the neurons $l$ of the output layer will out of is:

$$
y_{l}=F_{l}\left(b_{o l}+\sum_{j=1}^{k} \omega_{j l}^{2} \cdot f_{j}\left(\omega_{H j}+\sum_{i=1}^{n} \omega_{i j}^{1} x_{i}\right)\right)
$$

with $l=1,2, \ldots m$ (1)
In this study, the transfer functions: $F_{l}(x)=f_{j}(x)=\tan \operatorname{sig}(x)$ with $x \in[0 ;+\infty)$,so the formula (1) takes the form:

$$
\begin{equation*}
y_{l}=f\left(b_{o l}+\sum_{j=1}^{k} \omega_{j l}^{2} \cdot f\left(\omega_{H j}+\sum_{i=1}^{n} \omega_{i j}^{1} x_{i}\right)\right) \tag{2}
\end{equation*}
$$

with $l=1,2, \ldots m$
In which: $f(x)=\tan \operatorname{sig}(x)$. This value in the training process is compared with the target value to calculate the error. In the calculation process, this value will be out.

Back-propagation algorithm (Pof. S. Sengupta, 2001) was used to train network.

Error function is calculated by using the formula (Girish Kumar Jha, 2012):

$$
\begin{equation*}
\text { Ero }=\frac{1}{p} \sum_{i=1}^{p}\left(O_{i}-t_{i}\right)^{2} \tag{3}
\end{equation*}
$$

### 2.3.1. Building the training set for calculating the porosity PHI

Well log data of a drilled well is from 6000 to 20000 lines. We choose 360 lines, which six curves are fully collected and accurately recorded
and this study offers a new method to calculate the porosity to make the training set. The input columns of the training set are transferred to Logs matrix, column PHI is transferred to the column matrix TARGET, we have the training set (Logs TARGET).

### 2.3.2. Standardization of data

GR, DT, RHOB are standardized by using the Div (X) coefficients (Lê Hải An, Đặng Song Hà, 2014) as
$\operatorname{Div}(X)=\frac{\max (X)}{k}$ with $k \in\left[\begin{array}{ll}0.70 & 0.95\end{array}\right]$
NPHI is standardized by the exponent coefficient:

$$
\begin{equation*}
N P H I_{S \tan d}=0.7 \cdot \frac{e^{N P H I}}{e^{\max (N P H I)}} \tag{5}
\end{equation*}
$$

LLD, LLS are standardized by the average formula. The standardized value $x_{S \tan d}$ of $x$ is

$$
x_{S \tan d}=\left\{\begin{array}{l}
\frac{x}{2 * \operatorname{mean}(X)} \quad \text { if } x \leq \operatorname{mean}(X) \\
\frac{1}{2}+\frac{x-\operatorname{mean}(X)}{2 *(\max (X)-\operatorname{mean}(X))} \text { if } x>\operatorname{mean}(X)
\end{array}\right.
$$

### 2.3.3. Matching principle

A training set can be used to calculate for many wells. But the calculated well must satisfy the matching principle. The content of the matching principle is that: the $\operatorname{Div}(\mathrm{X})$ coefficients and the parameters in the formulas of average values of the calculated well must coincide with these values of the training set.

### 2.3.4. Design the net to calculate porosity

The number of hidden layer neurals is difficult to determine and usually is determined by using the trial and error technique. Surveying the relationship between the values of the well log datas and the porosity, this study concludes that number of the hidden layer neurals increases rate with input number and the comllexity of the well. The comllexity of the well is function of mean(RHOB),mean(GR),mean(NPHI). Network has 4 input, the hidden layer has k neurals is designed as in Figure 2.


Figure 2. The network for calculation of porosity.

### 2.3.5. Programming

Program to train and calculate porosity $\phi$ consists of 3 blocks:

Block 1: The training set consists of 360 rows, 8 colomns: Depth, GR, DT, NPHI, RHOB, LLD, LLS and PHI. The Input set consists of 6 columns GR, DT, NPHI, RHOB, LLD, LLS, the TARGET set is column PHI. Establish matrix A :

A $=$ [GR DT NPHI ROHB LLD LLS PHI].
Standardize A by formulas (4), (5) and (6) we receive the standardized matrix C :

C=[C1 C2 C3 C4 C5 C6 C7].

From C choose 4 colomns in the first 6 columns, send to matrix LOG.The $7^{\text {th }}$ column is sent to the column matrix TARGET, we receive the training set (LOG,TARGET).

Building the calculating set also do the same. Pick up 4 columns in the first 6 columns, send to matrix LOGtt ( 4 columns, $n$ rows).

Block 2: Function newff creates the untrained net.

Function train trains the untrained net, create the trained net.

Block 3: Function sim will calculate and print the results.


Figure 3. Programme to train and calculate porosity $\phi$


Figure 4. Comparison porosity Basroc and porosity ANN. Well RD_1P_Ha.txt.


Figure 5. The distribution of porosity in depth of well RD_1P

### 2.4. Verification of the accuracy of the method

The Mean square error after training the network is: 0.00001327
2.4.1 . Comparison the accuracy with the other softwares

Choose wells that porosity was calculated by
the existing softwares (BASROC, IP), then we calculate porosity by ANN with 4 curves. The test is done in many wells and each well include many thousands of lines of data.
Comparison for the magmatic basement rocks
The RD1P well has 5860 lines of data. X label presents porosity BASROC (dec); Y label is porosity ANN (dec). These two values distribute


Figure 6. Well KNT3X (the sedimentary clastic grain).


Figure 7. Compare porosity IP and ANN: Well KNT3X (the sedimentary clastic grain).
on the diagonal of the square (Figure 4). Distribution of porosity in depth shown in Figure 5.

0 x is depth, 0 y is porosity
Red curve : porosity calculate using ANN
Blue curve: porosity calculate using BASROC
The mean square error $=0.00002415$
Average value of porosity calculates by BASROC $=0.094179$

Average value of porosity calculates by ANN $=0.093847$

## Comparison for the sedimentary clastic grain

X label is porosity BASROC (dec); Y label is porosity ANN (dec). These two values distribute on the diagonal of the square.
(A) Number present of porosity

The Mean square error $=0.00007000$.

Average value of porosity calculates by IP = 0.07973376

Average value of porosity calculates by ANN $=0.07987584$

### 2.4.2. Compare the results calculated by different input's combinations

There are 15 combinations of 4 from 6 curves. The different ways of the calculations by ANN are the same results confirm the accuracy of the method. Only by 2 or 3 ways is enough. The square error after commas are 4 zero (see appendix 1).

### 2.4.3. Compare the results calculated in interaction with other parameters

Using the results of this study to develop the mining production technology diagrams, JVPC concludes: The results of ANN calculation of porosity PHI of this study have very high accuracy. This accuracy was confirmed when considered in relationship to the other parameters of the mining production technology diagrams.

## 3. Results and discussions

From the basic research and practical experience of handling 19 wells, this study has developed a system of programs (MATLAB language) and offer the rule of processing for the problem calculate porosity.

Calculating 19 wells of JVPC. The results are as follow:

The BASROC software only can be used to calculate for 4 wells

The ANN of this study is applied to calculate 15 wells. The results are good for 15 wells, like the well RD_5P, give in the appendix 1 (Table 1), and JVPC used this result to develop the mining production technology diagrams.

Four curves obtained good record are the condition for this study to calculate porosity, while the existing softwares require 6 curves have to receive and record completely and exactly. The accuracy of the two methods are the same. Therefore this study meet the actual needs, while the existing softwares are unable to meet. Obviously if the 6 curves obtained good record, this study also meet more easily.

Comparison the values of porosity calculated by the existing softwares and the values of
porosity calculated by the ANN for many wells and each well include many thousands of lines of data. It was found that these two values are equal. The square error after commas are 4 zero.

Figure 4 presents 5860 values of porosity calculated by the BASROC and by the ANN software for magma basement rock. These two values distribute on the diagonal of the square. It so that the porosity BASROC equals the porosity ANN.

Figure 5 presents the porosity in the depth (the horizontal axis presents the depth, the vertical axis presents porosity). Two curves coincide. It so that the porosity BASROC equals the porosity ANN.

Figure 6 is similar Figure 4 but for the clastic sedimentary grain rock.

Figure 7 compares the porosity ANN and porosity IP for each piece of 245 lines of data (corresponding to 37 meters in depth) to see the overlap of two porosity.

The ANN model to calculate porosity has great accuracy is due to:

The training set has been built to ensure the representativeness and completeness, compatible with each calculating well. With 360 trainning units, the net is trained all parameters to achieve the best.

Offering the standardized data method accuracy. The standardizing formulas maintain the original nature of the input and does not alter the relationship between input and output. Formulas (4), (5) and (6) respond fully and accurately the very heterogeneous environment of Cuulong basin.

This study find out the matching principle and comply this principle strictly.

The results of this study can be used both in the basic research and in practical calculations to develop the mining production technology diagrams.

## 4. Conclusion

- ANN model is a pratical tool for calculating porosity to study basins of South Vietnam. Calculating porosity is good with 4 inputs. 4 input are picked up in 6 curves: GR, DT, NPHI, RHOB, LLD, LLS. The training set should select from 300 to 400 as well. Do not choose more.
- The sedimentary basins of South Vietnam have the very high heterogeneity, the well log curves complicated. To calculate porosity exactly have to perform preprocessing scientifically and objectively.
- ANN model of this study can be applied to the other research in basins of South Vietnam.


## 5. Acknowledgments

The authors would like to thank: JVPC has used the results of this study to develop the mining production technology diagrams.

## References

Bùi Công Cường, 2006. Mathematical Institute of Vietnam. Publishing scientific and technical 2006. Artificial Neural Networks and fuzzy systems (in Vietnamese).
Girish Kumar Jha I A.R.I NewDelhi-110012 (girish iarsi@rediffmail.com).

Artifical Neuralnetworks and its applications.

Hoàng văn Quý PVEP, 2009. Ho Chi Minh city. Research fractured basement rock by using the BASROC software (in Vietnamese).
Hoàng văn Quý PVEP, 2014. Ho Chi Minh city. Lecture interpretation theory well log data (in Vietnamese).
Lê Hải An, Đặng Song Hà, 2014. Determination of the Mineral Volumes for The Pre-Cenozoic Magmatic. basement rocks of Cuu Long basin from Well log data via using the. Artificial Neural Networks. VNU, Jurnal of Earth and Environmental Sciences 30, No, 1, 1-12.
Limin Fu 460 McGraw-Hill,Inc- 1994. Neural networks in computer intelligence.
Nguyễn Doãn Phước, Phan xuân Minh, 2006. Hanoi Polytechnic University. Publishing scientific and technical 2006. Introduction to the Artificial neural networks. (in Vietnamese).
Pof. S. Sengupta Departmen of Electionl Communication Engineering IIT.2001. The Backpropagation.
(MATLAB/toolbox/nnet/backpr52.html)..

## Appendix 1

\% Phil= NDLR Phi2 =GDLR phi3= GDNR \% Phi Average value File: RD_5P.txt \%Depth Phil Phi2 Phi3 Phichung Phi Average value 3719.0173000 .1429050 .1375470 .1379870 .1307000 .137395 3719.1697000 .1425760 .1364230 .1379330 .1305300 .137243 3719.3220000 .1450950 .1389010 .1381080 .1310700 .138095 3719.4746000 .1450810 .1324150 .1363360 .1298600 .136903 3719.6270000 .1450510 .1368120 .1367700 .1303900 .137245 3719.7793000 .1437510 .1354980 .1387880 .1319000 .138307 3719.9316000 .1442240 .1286880 .1390680 .1322300 .138647 3720.0840000 .1447950 .1370120 .1376010 .1310400 .137759 $3720.2366000 .145318 \quad 0.1393010 .1398430 .1329200 .139481$ $3720.3889000 .144048 \quad 0.1406300 .1409340 .1333200 .139657$ 3720.5413000 .1426260 .1404750 .1445170 .1327700 .139087 3720.6936000 .1427320 .1381960 .1382310 .1306400 .137441 3720.8462000 .1426790 .1365620 .1345670 .1294500 .136313 3720.9985000 .1432750 .1368910 .1364160 .1299100 .136742 3721.1509000 .1436200 .1336490 .1334750 .1278300 .134687 3721.3032000 .1454130 .1327690 .1323120 .1277200 .134668 3721.4558000 .1483050 .1339750 .1336570 .1291800 .136359 3721.6082000 .1486030 .1347200 .1344650 .1297900 .136958 3721.7605000 .1487580 .1357010 .1359870 .1297300 .137473 3721.9128000 .1479320 .1367110 .1364320 .1324300 .138446 3722.0652000 .1480900 .1403300 .1404560 .1355600 .141078 3722.2178000 .1488620 .1435760 .1438760 .1407700 .144196 3722.3701000 .1485940 .1435420 .1438760 .1396700 .143837 3722.5225000 .1487510 .1446950 .1444360 .1401200 .144565 3722.6748000 .1480770 .1419400 .1417680 .1387600 .142679 3722.8274000 .1473620 .1338130 .1339040 .1377500 .138184 3722.9797000 .1468030 .1340950 .1344760 .1379000 .138223 3723.1321000 .1462510 .1261480 .1265720 .1354500 .133499 3723.2844000 .1477350 .1105390 .1108620 .1325900 .125351 3723.4370000 .1478010 .1099030 .1094560 .1326300 .125059 3731.9712000 .1284640 .1279830 .1278920 .1187200 .125788 $3732.1238000 .1304940 .1242590 .124468 \quad 0.1186300 .124411$ 3732.2761000 .1266100 .1186140 .1187810 .1105900 .118607 3732.4285000 .1244130 .1159770 .1153450 .1065500 .115729 3732.5808000 .1056230 .1045540 .1047650 .0908500 .101395 3732.7334000 .0882160 .0975710 .0943110 .0769700 .090082 3732.8857000 .0618160 .1043720 .1020910 .0736900 .086063 3733.0381000 .0393370 .0766310 .0757210 .0450000 .059400 3733.1904000 .0313010 .0573580 .0579180 .0283600 .043594 3733.3430000 .0281090 .0510620 .0556720 .0252700 .038876 3733.4954000 .0238230 .0405060 .0442130 .0179400 .030694 3733.6477000 .0224730 .0393850 .0374250 .0174200 .029666 3733.8000000 .0203530 .0290410 .0276110 .0139200 .023089 3733.9524000 .0200320 .0274420 .0277650 .0178200 .023184 3734.1050000 .0164100 .0140980 .0144360 .0126600 .014316 3734.2573000 .0135050 .0105200 .0107890 .0104500 .011249 3734.4097000 .0225600 .0220840 .0224370 .0130600 .019947 3734.5620000 .0291700 .0383660 .0387810 .0199300 .031458 3734.7146000 .0361230 .0424490 .0427920 .0260500 .036768 3734.8669000 .0534970 .0688690 .0685600 .0478100 .059761 3735.0193000 .0481010 .0597310 .0594610 .0395500 .051778 3735.1716000 .0469570 .0670530 .0679740 .0423600 .055856 $3735.3242000 .047108 \quad 0.072188 \quad 0.072870 \quad 0.044980 \quad 0.059116$ 3735.4766000 .0476590 .0747310 .0743420 .0458100 .060733 3735.6289000 .0486760 .0866730 .0869210 .0534900 .068878 3769.9189000 .0229070 .0211900 .0234560 .0166300 .020479 3770.0713000 .0222990 .0193410 .0173460 .0151900 .019043 3770.2236000 .0194260 .0152250 .0167810 .0112200 .015274 3770.3762000 .0190990 .0140020 .0144720 .0108000 .014476 3770.5286000 .0193350 .0121370 .0125890 .0104600 .013517 3770.6809000 .0200190 .0113580 .0117890 .0110600 .013449 3770.8333000 .0227090 .0166430 .0162310 .0156800 .017919 $3770.9856000 .0253840 .0226690 .022489 \quad 0.021520 \quad 0.023061$

## Appendix 2




[^0]:    *Corresponding author
    E-mail: songhadvl@gmail.com

