VNU Journal of Science: Earth and Environmental Sciences, Vol. 33, No. 1 (2017) 16-25

Correction and Supplementingation of the Well Log Curves for Cuu Long Oil Basin by Using the Artificial Neural Networks

Dang Song Ha[1,*], Le Hai An[2], Do Minh Duc[1]

[1]*Faculty of Geology, VNU University of Science, 334 Nguyen Trai, Hanoi, Vietnam*

[2]*Hanoi Mining and Geology University, 18 Vien, Duc Thang, Hanoi, Vietnam*

Received 06 February 2016

Revised 24 February 2016; Accepted 15 March 2017

Abstract: When drill well for the oil and gas exploration in Cuu Long basin usually measure and record seven curves (GR, DT, NPHI, RHOB, LLS, LLD, MSFL). To calculate the lithology physical parameters and evaluate the oil and gas reserves, the softwares (IP, BASROC...) require that all the seven curves must be recorded completely and accurately from the roof to the bottom of the wells. But many segments of the curves have been broken, and mostly only 4, 5 or 6 curves have could recorded. The cause of the curves being broken or not recorded is due to the heterogeneity of the environment and the lithological characteristics of the region. Until now the improvements of the measuring recording equipments (hardware) can not completely overcome this difficulty.

This study presents a method for correction and supplementing of the well log curves by using the Artificial Neural Networks.

Check by 2 ways: 1). Using the good recorded curves, we assume some segments are broken, then we corrected and supplemented these segments. Comparing the corrected and supplemented value with the good recorded value. These values coincide. 2). Japan Vietnam Petroleum Exploration Group company LTD (JVPC) measured and recorded nine driling wells. Data of these nine wells broken. This study corrected and supplemented the broken segments, then use the corrected and supplemented curves to calculate porosity. The porosity calculated in this study for 9 wells has been used by JVPC to build the mining production technology diagrams, whle the existing softwares can not calculate this parameter. The testing result proves that the Artificial Neural Network model (ANN) of this study is great tool for correction and supplementing of the well log curves.

Keywords: ANN (ArtificLal Neural Network), well log data, the lithology physical parameters, Cuu Long basin.

1. Introduction

The Cenozoic clastic grain sediments and the pre Cenozoic fractured basement rocks are the large objects contain oil and gas in Cuu Long basin. The Cenozoic sediment unconformably covers up the weathering and eroded fractured basement rocks. The oil body in the clastic grain sediments has many thin beds with the different oil- water boundaries. The oil body has small size [1]. The pre-Cenozoic basement rocks composed of the ancient rocks as sedimentary metamorphic,

* Corresponding author. Tel.: 84-938822216.
Email: songhadvl@gmail.com

carbonate rock, magma intrusion, formed before forming the sedimentary basins, has the block shape, large size [1]. 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. These characteristics trigger off the well log curves have the broken or not recorded segments. So the improvements of the measuring recording equipments (hardware) can not completely overcome.

1.1. Database

The following is a few lines of data in the 26500 lines of the DH3P well:

Depth GR DT NPHI RHOB LLD LLS MSFL

(M) (API) (s/fit) (dec) (g/cm^3) Ohm.m) (Ohm.m) (Ohm.m)

```
1989.9541  83.3086 -999.0000  0.4503  2.0891 -999.0000 -999.0000 -999.0000
..   ..  ..   ..   ..   ..   ..   ..
1994.3737  88.5760 -999.0000  0.3604  2.2282 -999.0000 -999.0000 -999.0000
1994.8309  77.1122  65.4558  0.3663  2.2742  0.5390  0.7460  0.7378
1994.9833  75.7523  65.0494  0.3346  2.3337  0.6042  0.7370  0.7923
..   ..   ..   ..   ..   ..   ..   ..
2337.2737 118.5451  87.2236  0.2207  2.5132  4.6080  3.0328  3.2493
2337.4261 121.1384  85.3440  0.2233  2.5135  3.6242  2.3838  2.3024
..   ..   ..   ..   ..   ..   ..   ..
3151.6993   72.4672   53.1495   -0.0010    2.6849 2749.8201  142.0989   13.0625
3151.8517   72.4670   53.1495   -0.0010    2.6816 2726.7100  142.0516   13.0625
```

GR (API): Gamma Ray log; DT (.uSec/ft): Sonic comprressional transit time; NPHI (dec): Neutron log; RHOB (gm/cc): bulk density log; LLD (ohm.m): laterolog deep; LLS (ohm.m): laterolog shallow; MSFL (ohm.m): microspherically.

From the top to the bottom of the wells, many segments of the curves have been broken, and mostly only 4 to 6 curves have been recorded. The broken data is written by -999.000. The GR curve of the DH3P well has 4 segments have been broken, which need to correct and supplement:

Table 1. The broken segments of the DH3P well

Broken segment	From line.. to line	Number of broken lines
1	260 -	53
2	312	114
3	501 -	64
4	614	119
	753 -	
	816	
	1003 -	
	1121	

Such databases are all 7 curves. The good record segments are database for correction and supplementing of the broken segments.

1.2. Approach

This study uses the Artificial Neural Networks (ANN) to correct, supplement the broken segments of the well log curves in Cuu Long basin. Following presents the method of correction and supplementing of the GR curve. The other curves also do the same but with a few minor details need specific treatment.

To correct and supplement the GR curve, we choose Output is GR. Inputs are four curves are selected in the 6 remaining curves.

1.3. Purpose

From the curves have the broken segments, this study supplements to these broken segments for the curve with the complete data from the roof to the bottom of the well. The supplementary curves must meet the condition: The supplementary segments accurately reflect the geological nature of the corresponding depth. The scientific basis of the method will present in discussions.

2. Methods

Artificial neural networks

The ANN is the mathematical model of the biological neural network. LiminFu [2] (1994) demonstrated that just only one hidden layer is sufficient to model any function. So the net only need 3 layers (input layer, hidden layer and output layer) to operate. The processing information of the ANN different from the algorithmic calculations. That's the parallel processing and calculation is essentially the learning process. With access to nonlinear, the adaptive and self-organizing capability, the fault tolerance capability, the ANN have the ability to make inferences as humans. The soft computation has created a revolution in computer technology and information processing [3], solving the complex problems consistent with the geological environment heterogeneity.

3. Results

3.1. Development of the Cuu Long network

The supplementing GR Cuu Long network is developed as follows:

- Input layer consists of n neurals: $x_1, x_2, ...x_n$,

- Hidden layer consists of k neurals and the transfer functions $f_j(x)$ with $j = 1,2...k$

- Output layer consists of *one* neural and the transfer function $f(x) = \tan sig(x)$ with $x \in (0,.05 , 0.95)$

Each neural is a calculating unit with many inputs and one output [4]. Each neural has an energy of its own called it's bias threshold , and it receives the energy from other neurals with different intensity as the corresponding weight. Neurals of the hidden layer receive information from the input layer. It calculates then sent the results to the output neural. The computing results of the Output GR neural is:

$$y_o = f(b_o + \sum_{j=1}^{k} \omega_j^2 . f(b_{Hj} + \sum_{i=1}^{n} \omega_{ij}^1 x_i)) \quad (1)$$

the transfer functions $f(x) = \tan sig(x)$ with $x \in (0,.05 , 0.95)$

in which, b_o , b_{Hj} are the threshold bias of the Output GR neural and the j neural of Hidden layer ($j = 1,2,...k$)

ω_{ij}^1 is weight of the Intput neural i sent to the neural j of Hidden layer,

ω_j^2 is weight of the j neural of Hidden layer sent to the Output neural Gr.

k is the number of neurals of the Hidden layer, n is the number of neurals of the Input layer. Value y_o in the training process is compared with the target value to calculate the error. In the calculating process, it will be out.

The Back-propagation algorithm [5] was used to train the net.

Error function is calculated by using the formula [4]:

$$Ero = \frac{1}{p}\sum_{i=1}^{p}(O_i - t_i)^2 \quad (2)$$

3.2. Building the training set for the supplement of the GR curve

- With the broken segments (we want to supplement) we calculate: DT_{min}=min(DT), DT_{Max}= max(DT). Similarly with NPHI, RHOB, LLD, LLS, MSFL.

- The training set consists of 360 data lines, selecte in the well and has to satisfy the condition: 7 data are good record. The values DT, NPHI, RHOB, LLD, LLS, MSFL must satisfy conditions: $DT_{\min} \le DT \le DT_{Max}$, $NPHI_{\min} \le NPHI \le NPHI_{Max}$. Similarly with RHOB, LLD, LLS, MSFL. The input columns of the training set are sent to the LOG matrix, column GR is sent to the column matrix TARGET, we have the training set (LOG TARGET), consists of 360 lines.

3.3. Standardization of data

GR,DT,RHOB are standardized by using the Div (X) coefficients [6] as

$Div(X) = \frac{\max(X)}{k}$ with $k \in [0.70 \quad 0.95]$.

Value $x_{S\tan d}$ of x is:

$$x_{S\tan d} = \frac{x}{Div(x)} \quad (3)$$

NPHI is standardized by the exponent coefficient. Value $NPHI_{S\tan d}$ of $NPHI$ is:

$$NPHI_{s\tan d} = 0.80.\frac{e^{NPHI}}{e^{\max(NPHI)}} \quad (4)$$

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

$$x_{S\tan d} = \begin{cases} \dfrac{x}{2*mean(X)} & if \quad x \le mean(X) \\ \dfrac{1}{2} + \dfrac{x - mean(X)}{2*(\max(X) - mean(X))} & if \quad x > mean(X) \end{cases} \quad (5)$$

3.4. Design the network. Training the network

The number of the 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, this study concludes that the number of the hidden layer neurals increases e with the number of the input and the comllexity of the well. The comllexity of the well is function of *mean(RHOB), mean(GR), mean(NPHI).* The net consists of 4 input, the hidden layer has from 6 to 9 neurals.

Training the network is to adjust the values of the weights so that the net has the capable of creating the desired output response, by minimum the value of the error function via using the gradient descent method. Function *newff* creates the untrained net $net0$ (read: *net zero*) in the big rectangle below; 4 column LOG in the training set (LOG TARGET) are sent into 4 rows of 360 columns in 4 rectangles on the left (DT, Nphi, Rhob, LLD). The TARGET column was sent into 1 line 360 columns is the rectangular on the right as figure 1.

Phase 1:

Step 1: Values $DT_1, Nphi_1, Rhob_1, LLD_1$ are sent to 4 Input neurals :DT, Nphi, Rhob, LLD (4 red circles on the left). Value Gr_1 is sent to the Output neural Gr (red circle on the right). Four neurons DT, Nphi, Rhob, LLD receive and transfer the values $DT_1, Nphi_1, Rhob_1, LLD_1$ to the hidden layer neurons (which multiplied by the weight).

The hidden layer neurons H_1, H_2... H_k aggregated information, calculated by their transfer functions then sent the results (weights multiplied) to the Output neural Gr.

The Neural Gr receives information, uses it's transfer function to calculate the Output value by formula (1). The Output value was compared with the value Gr_1 on the right. Calculate the error E. E is greater. Phase 1 ended. Switched to phase 2.

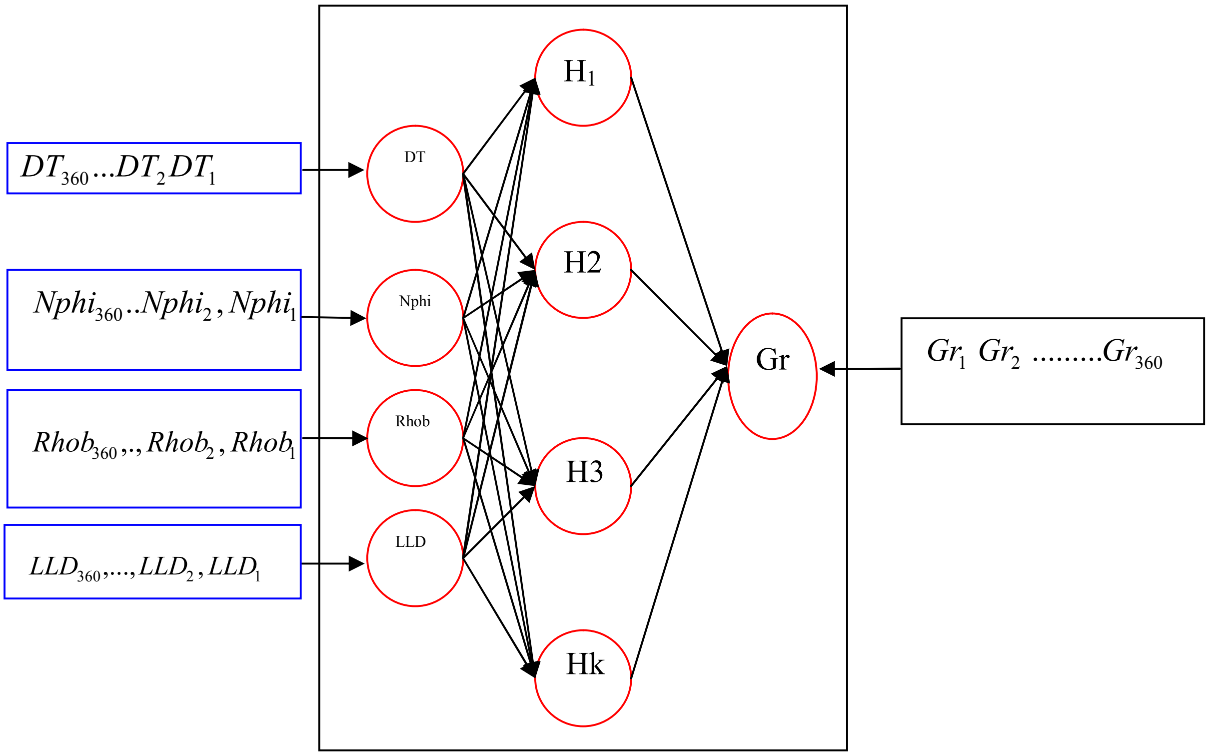


Figure 1. The training net.

Phase 2:

Step 2: From Output Neural return Hidden layer. Calculate $\frac{\partial E}{\partial \omega_{ij}^2}$.

Step 3: From the Hidden layer return Input layer. Calculate $\frac{\partial E}{\partial \omega_{ij}^1}$

Step 4: At Input layer: The weights are adjusted by solving the system of the partial differential equations [4] :

$$\begin{cases} \frac{\partial E}{\partial \varpi_{ij}^1} = 0 \\ \frac{\partial E}{\partial \varpi_{ij}^2} = 0 \end{cases} \quad (6)$$

These weights satisfied conditions minimizing of the error function, so better the weights in the loop of the previous step. Step 4 ends. The cycle repeated thousands of times to make the weights as the later the better [4]. When the error is small enough, the first training shift ended. The second training shift starts and over 360 shifts of such training, the untrained net $net0$ becomes the trained net net .

The calculating net consists of 4 Input, Hidden layer k neurals is designed:

In the big rectangle is the trained net net . The calculating net received Input from the need supplement segments. The Gr neural calculates and sends the results out.

Programming by using functions: Function *newff* creates $net0$. Function *train* traines $net0$ become net . Function *sim* uses *net* to model.

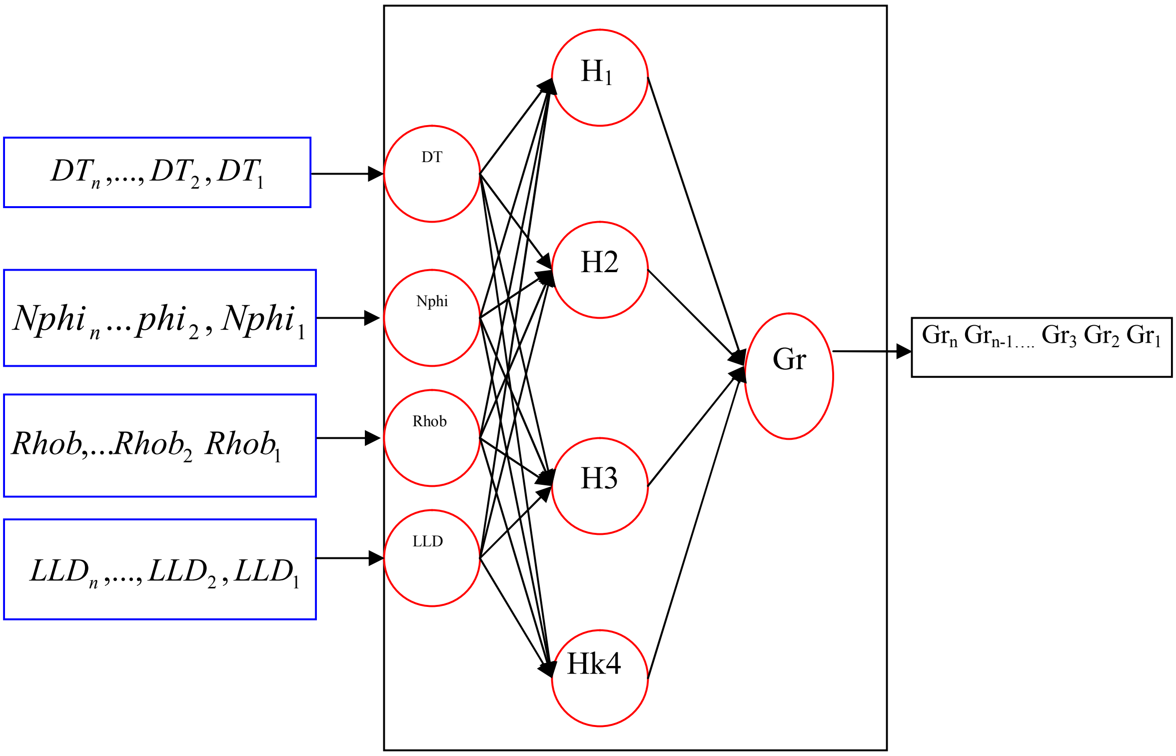


Figure 2. The ANN net for supplementing of the GR curve.

3.5. Create the GR curve from the top to the bottom of the well by ANN

From 5 curves DT, NPHI, RHOB, LLD, LLS, the ANN can create the GR curve from the top to the bottom of the well coincides with the curve obtained when drill well, by using the net as above but the calculating set is the 5 curves DT, NPHI, RHOB, LLD, LLS from the top to the bottom of the well.

Figure 3 below is the GR curve obtained by POC record when drill well (red) and the GR curve created by ANN of this study (blue). Two these curves overlap.

Figure 4 below: Ox presents GR recorded by the POC, Oy presents GR created by the ANN of this study. They are distributed on the diagonal of the square. So the two curves overlap.

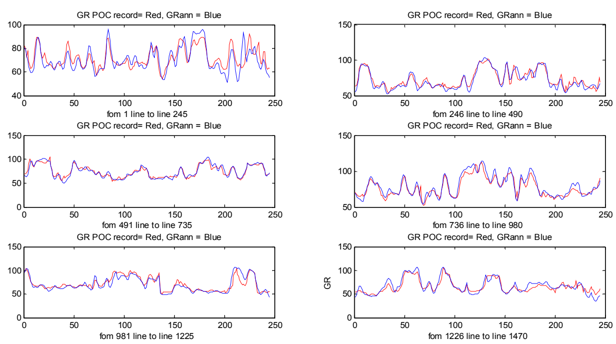


Figure 3. Curve GR recorded by the POC (red), and GR created by ANN of this study (blue) from the top to the bottom of the well DH5P.

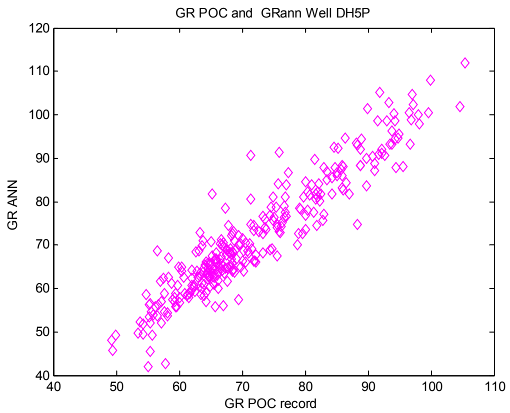


Figure 4. Values GR recorded by the POC (Ox), and GR created by ANN of this study (Oy) from the top to the bottom of the well DH5P.

The absolute error and the square error of the different ways of calculation as follows:

Input is DT, NPHI, RHOB, LLD

Neural of Hiddenlay	Absolute error	Square error
6	0.04632	0.001579
7	0.04187	0.001557
8	0.04110	0.001447
9	0.04023	0.001946

Input is DT, NPHI, RHOB, LLS

Neural of Hiddenlay	Absolute error	Square error
6	0.042039	0.001749
7	0.041518	0.001652
8	0.044010	0.001912
9	0.042713	0.001894

From the table we see: The error very small and quite stable. Great precision.

3.6. Supplement of the GR curve

Only use the GR curve created by the ANN to supplement into the broken segments. The good recorded segments are not change.

The broken segments of the GR curve of the DH3P well are supplemented. The first broken segment consist of 53 lines, from the 260th line to the 312th line (table1).

Figure 5a: The good recorded lines are presented by red colour. The broken lines are presented by black colour.

Figure 5b presents the curve after supplementing by the ANN of this study.

Figure 5c: The red curve is the supplemented curve, the blue curve is the curve is created by the ANN of this study. The two curves overlap.

The orther broken segments are presented in Appendix

3.7. Application of Cuu Long net for correction and supplementation for well log curves

1. Just 360 lines of data that the 7 curves are recorded completely and accurately we can supplement the broken segments. The current measuring and recording always meet this requirement easily.

- The GR curve, the DT curve can supplement very good. The ANN can be used to create two curves from the top to the bottom of the well. Use 2 curve created by ANN to calculate porosity. This porosity coincides with the porosity calculates by use the two good record curves.

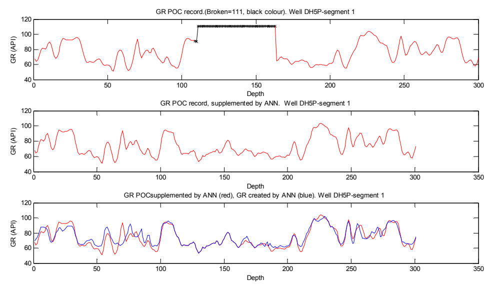


Figure 5. The segment consist of 301 lines has the first broken segment(53 lines).

a) Red colour are the good recorded lines, black colour are the broken lines; b) The curve after supplementing by ANN ; c) The red curve is the supplemented curve, the blue curve is the curve created by the ANN of this study. The two curves overlap.

- The NPHI curve and the RHOB curve can supplement the broken segments. The accuracy acceptable.

- The resistivity curves (LLD, LLS, MSFL) can not supplement.

2. With the 4 supplementary curves are sufficient to calculate the porosity by using the ANN that the other softwares are not able to calculate of porosity.

3. The Exploration Group Japan Vietnam Petroleum Company LTD (JVPC) drilled, recorded 9 wells. The curves of the these 9 wells were broken. This study supplemented the broken segments, then use the supplemented curves to calculate porosity. The JVPC has used the results of calculations of porosity of this study for the these 9 drilling wells in order to build the mining production technology diagrams. JVPC evaluated the porosity calculated by this study has very high accuracy. The other softwares can not calculate the porosity for the these nine drilling wells.

4. All the drilling wells always have the broken segments, and are able to use this study to supplement. The results of this study are put to use in preprocessing of the well log data.

4. Discussion

Corection and supplementing of the GR curve may use 3 Input curves. Preferably selecs 4 Input curves in 6 curves: DT, NPHI, RHOB, LLD, LLS, MSFL.

The training set consisting of 360 lines is good. Do not select more.

The ANN to complement the well log curves of this study has the great precision because:

- Has built the training set to ensure the representativeness and completeness, suitable for each broken segments. With 360 trainning units, the net is trained all parameters to achieve the best

- The matching principle is: The coefficient in the formula (3), the coefficient in the formula (4) and the parameter in the formula (5) of the calculating well and the training well must be the same. The training set is built from the data of the supplement well it's self, so the matching principle was self-fulfilling.

- Find out the data standardized method accuracy. The average contribution of input variable i is [4]:

$$C_i = \frac{\sum_{j=1}^{k} \varpi_{ij} . x_i}{\sum_{i=}^{n} \sum_{j=1}^{k} \varpi_{ij} . x_i} \quad with \quad i = 1,2,...n \qquad (7)$$

From (7) we see the contribution dependent on x_i . In Cuu Long basin, GR, DT, RHOB have the Normal distribution (Gauss distribution). NPHI has the Normal loga distribution. LLD, LLS, MSFL have the χ^2 distribution with many the different free degrees, dependent on the value of mean(X) with X is LLD, LLS, MSFL. Formula (3), (4), (5) retain the nature of the input values, does not change the relationship of the input to the Output, meet the very heterogeneous environment of the Cuu Long basin.

- Base on the analysis of the characteristics of the resistivity curves (LLD, LLS, MSFL), the NPHI curve and the geological nature of the Cuu Long basin, this study selects the transfer function is $f(x) = \tan sig(x)$ with $x \in (0,.05 , 0.95)$ is suitable. Select $x \in (0,.05 , 0.95)$ makes the net does not give the extreme value.

The very heterogeneous environment of the Cuu Long basin creates condition for the ANN can from the values of DT, NPHI, RHOB, LLD, LLS, easily infers the value of GR. This is the scientific basis of the method. Because the environment is a unified whole that all the phenomena are in a relationship of mutual binding.

5. Conclusions

Cuu Long net for correction and supplementation for well log curves is a good tool for preprocessing of the well log data.

ANN is a good tool for redicting the lithology physical parameters.

The training set ensures the representativeness, remove anomalous data and standardization of data accuracy are important factors to use ANN.

Acknowledgments

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

References

[1] Hoàng văn Quý PVEP 2014. Ho Chi Minh city. Lecture interpretation theory well log data (in Vietnamese).

[2] LiminFu McGraw-Hill, NewYork (1994). Neural networks in computer intelligence.

[3] Bùi Công Cường: Mathematical Institute of Vietnam.Publishing scientific and technical 2006. Artificial Neural Networks and fuzzy systems (in Vietnamese).

[4] Girish Kumar Jha I A.R.I NewDelhi-110012. Artifical Neuralnetworks and its applications.

[5] Pof. S. Sengupta, Department of Electronics and Electrial Communication Engineering IIT. The Backpropagation (neural network toolbox-MATLAB).

[6] Lê Hải An, Đặng Song Hà. 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 Vol 30, No, 1, 2014, 1-12.

Sửa chữa, bổ sung các đường cong địa vật lý giếng khoan bể Cửu long bằng mạng Nural nhân tạo

Đặng Song Hà[1], Lê Hải An[2] , Đỗ Minh Đức[1]

[1]*Khoa Địa chất Đại học Khoa học Tự nhiên-ĐHQGHN, 334 Nguyễn Trãi, Hà Nội, Việt Nam*

[2]*Đại học Mỏ Địa chất, 18 Phố Viên, Đức Thắng, Hà Nội, Việt Nam*

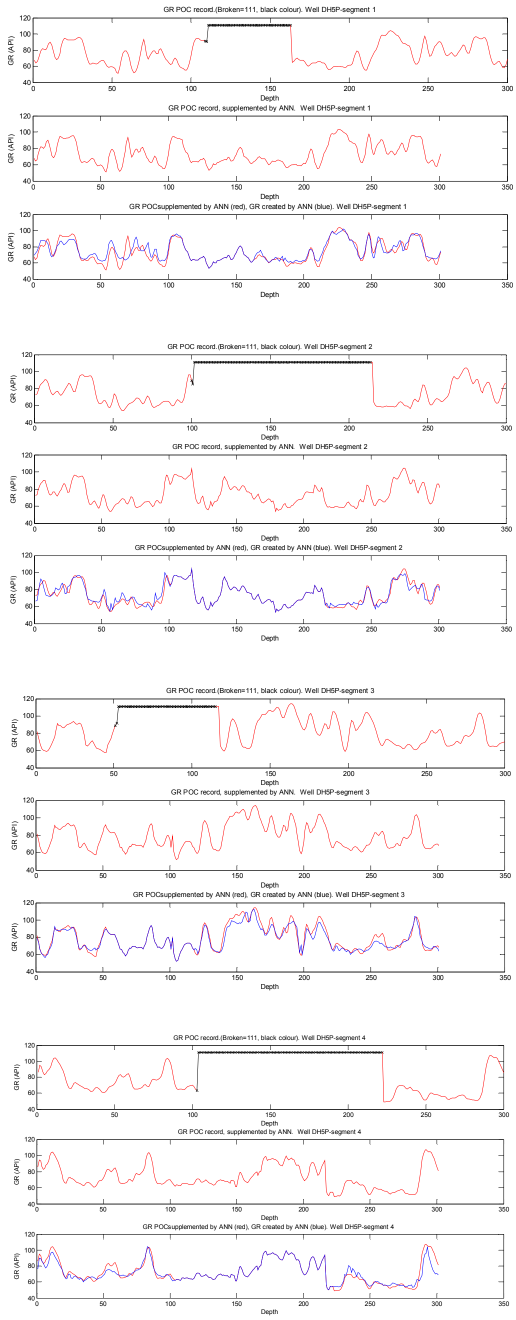
Tóm tắt: Khoan giếng thăm giò khai thác dầu khí bể Cửu long thường thu 7 đường cong (GR DT NPHI RHOB LLD LLS, MSFL). Để tính các tham số vật lý thạch học và dánh giá trữ lượng dầu khí thi 7 đường cong phải thu được đầy đủ và tốt từ nóc móng đến đáy giếng. Nhưng có những khúc chỉ thu ghi tốt được 4, 5 hoặc 6 đường cong. Nguyên nhân thu ghi bị hỏng là do sự bất đồng nhất của môi trường và đăc điểm vật lý thạch học của khu vực gây nên. Vì vậy cải tiến thiết bị thu ghi (phần cứng) không thể khắc phục được hoàn toàn.

Nghiên cứu này đưa ra phương pháp sửa chữa, bổ sung từng đường cong từ tài liệu ĐVLGK bằng mạng nơron nhân tạo (ANN).

Kiểm tra bằng 2 cách: 1). Dùng các đường cong thu ghi tốt, ta giả sử một số đoạn là thu ghi hỏng rồi bổ sung các đoạn này. So sành giá trị ta bổ sung với giá trị thu ghi tốt ta thấy giống nhau. 2). Exploration Group Japan Vietnam Petroleum Co.. LTD (JVPC) thu ghi 9 giếng khoan bị hỏng, các phần mềm hiện có không tính được độ rỗng. Nghiên cứu này đã bổ sung các đoạn thu ghi hỏng rồi sử dụng các đường cong đã bổ sung để tính độ rỗng. Kết quả tính độ rỗng này đã dược JVPC sử dụng để xây dựng sơ đồ công nghệ khai thác mỏ. Kiểm tra này chứng tỏ : Mô hình mạng Nơron nhân tạo (ANN) của nghiên cứu này là công cụ tốt để sửa chữa, bổ sung các đường cong từ tài liệu ĐVLGK

Từ khóa: Mạng Neural nhân tạo (ANN) , đường cong địa vật lý giếng khoan (LGK), tham số vật lý thạch học, Bể Cửu long.

Appendix



a) Red colour are the good recorded lines, black colour are the broken lines;
b) The curve after supplementing by ANN; c) Red is thecurve after supplementing by ANN of this sudy;
blue is the GR created by ANN of this study. The two curces overlap.