Use of neural networks in log's data processing: prediction and rebuilding of lithologic facies

Dominique Frayssinet
18 rue du docteur Plichon
94000 Créteil
France

Dominique Frayssinet[1,2], Sylvie Thiria[1], Fouad Badran[1,2] and Louis Briqueu[3]

[1] Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC)
case 100, Université Paris 6, 4 place Jussieu 75252 Paris cedex 05 France
[2] CEDRIC, Conservatoire National des Arts et Métiers, 292 rue Saint Martin 75003 Paris France
[3] Laboratoire de Géophysique, de Tectonique et de Sédimentologie, Université de Montpellier, UMR5576 CNRS – UMII case 066 Place Eugène Bataillon 34095 Montpellier Cedex 5 France.
thiria@lodyc.jussieu.fr badran@cnam.fr briqueu@dstu;univ-montp2.fr

Abstract

When a log is missing in a drilling hole, geologists hope to deduce it from others logs available in another part of the hole or in a neighbouring hole, in order to define the lithologic facies of the hole. This paper presents a neural network method to predict the missing log's measure from the other available log's measures. This method, based on Multi-Layer Perceptron (MLP) acts as a non linear regression method for the prediction task and as a probability density distribution approximation for the outlier rejection task. The result obtained when applied to actual log's data for prediction and rejection are presented in a separate section. The last section is dedicated to a non supervised neural method in order to reconstruct the lithologic facies of the concerned hole. This last experiment allows to validate and interpret the different results of the proposed methods.

1 Introduction

In order to investigate the subsoil's composition, geologists take sensors down in the drilling holes for measuring physical properties of the rocks. We call logs the result of these measures. The measure quality is highly disturbed by technical difficulties which may occur while a drilling process. So that, a log may be missing in a hole part of the well. The problem is, knowing p logs, how to deduce the $(p+1)^{th}$ missing log. Afterwards, $\mathbf{x} = (x_1,x_2,\ldots,x_p)$ is an observation of p log's values taken at a given depth level, and y is the value of the $(p+1)^{th}$ log we want to determine. In such a case, usually geologists use the correlation existing between the different available logs and the missing one, to deduce the value they need [références par Louis Briqueu]. We propose neural networks methods for helping geologists in determining the correspondence between several logs. In the following we use multilayer perceptrons (MLP) which allow nonlinear function approximation. The problem is thus to estimate the MLP's parameters, using a learning set that we call App = $\{(\mathbf{x}^1,y^1), (\mathbf{x}^2,y^2),\ldots(\mathbf{x}^N,y^N)\}$, where $\mathbf{x}^k$ =

$(x_1^k, x_2^k, ..., x_p^k)$ represents a measure of p logs and y^k the measure of the $(p+1)^{th}$ one. The information contained in this learning data set corresponds to a defined situation, which depends on geology, drilling tools and drilling conditions. This learning set implicitly determines the domain where MLP's predictions are valid. In practice, logs may present outlier values because of technical difficulties, or data corresponding to geological situations which are not represented in the learning set. Therefore, we have to throw out those data, while using the MLP for prediction. So, a major problem is to detect these measures **(x)** which are not in adequacy with the learning set. In this paper, we propose a neural network method which computes and interprets the probabilities' repartition of the random variable Y (the $p+1^{th}$ missing log), conditioned by an observation $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. The values provided by the neural network allows to build an accurate rejection method.

In the following we present the neural network method used for prediction and rejection and the results obtained when applied to actual log's data. In the last section we use a non supervised neural method in order to reconstruct the lithologic facies of the concerned hole. This last experiment allows to validate and interpret the different results of the proposed methods.

2 Neural methods: prediction and rejection

Prediction by multilayer perceptron

A MLP is made of 3 kinds of neuron's layers : an input layer, one or several hidden layers and an output layer. In our case, every neuron on a layer is fully connected to every neurons on the next layer. The MLP can be decribed by a direct graph where the vertices are the neurons and the edges are the connections. A real value is affected to each connection and is denoted by its weight. We call W the set of connections' weights. The MLP takes as input the vector **x**, propagates this vector's value through hidden layers, and computes the activation of the output neuron. This activation, that we note F(**x**,W) depends on **x** and W. If the input layer has p neurons and the output layer has a single neuron, then the MLP will appear as a nonlinear function from R^p into R. A brief presentation of MLP can be found in [Thiria 1993], and a complete presentation in [Bishop 1995]. The function F(**x**,W) is a non linear differentiable function, and we propose to use it to predict the missing log, knowing the vector $\mathbf{x} = (x^1, \ldots, x^p)$ which represents the p logs' values. Computing nonlinear regression is fundamental, because usually, the underlying relation between the available logs and the missing one is nonlinear. The adaptation of MLP's parameters to the particular problem we want to resolve is made with the learning set App, using the retropropagation algorithm [Bishop 1995]. Learning process consists to minimize a cost function *C* which is the quadratic error between the predicted value and the observed one:

$$C = \sum_{k=1}^{N} \left(y^k - F(\mathbf{x}^k, W) \right)^2 .$$

We know that minimizing this function implies that F(**x**,W) ≅ E(Y/**x**) where E(Y/**x**) corresponds to the statistical average of the missing log y, knowing p observed others logs **x**. [Bishop 1995]. Considering the prediction of the missing log as a regression problem supposes that the distribution of the random variable Y/**x** is an unimodal distribution (normal distribution). So, for the model F(**x**,W) can predict the missing log, the unimodal distribution of Y must be checked. It is this idea that we develop in the next paragraph, with the aim of detecting the values which are not in adequacy with the whole learning set.

Outlier detection by multilayer perceptron

The method is based on the methodology developed in [Mejia 1992] and [Thiria and al 1993], it uses MLP's which approximate the conditional distribution Y/**x**. Any input vector **x** for which this distribution is significantly multimodal has to be eliminated from the MLP prediction process. Thus, in our application's framework, whereas we predict the p+1th missing log as a real value, we subdivide it in K intervals. If the p+1th missing log value belongs to the ith interval, we associate to it the vector **z** of R^K, **z** = (0,0,…,1,…,0,0), where the single number 1 appears at the ith position. The learning process consists in modeling the association (**x** , **z**) by minimizing the quadratic error C. The MLP (denoted MLP classifier hereinafter), which has K output neurons, computes K distinct values for a given **x** = (x_1,x_2,x_3), and we can show that the ith one represents the conditional probability that the p+1th missing log value belongs to the ith interval. So, for a given vector **x**, the activation of the output layer corresponds to the histogram of the conditional Y/**x** random variable, knowing a given observation **x**. For each vector of p log's measures, the network provides an output state curve. If a single activation pick appears then the underlying conditional distribution of Y/**x**. is unimodal. Several activation picks point out that the distribution is multimodal and the p+1th missing log value may belong to several different intervals, thus several values can be proposed from the p measures. Due to the uncertainty, we decide to eliminate these input's logs from the MLP prediction process.

3. Experimental results

In the following we present the results we have, decoding actual data with the method presented above. Our study turns to two drilling holes named MAR203 and MAR402, which have been drilled by ANDRA [], in the Marcoule place, inside the French South-East basin. Subsoil is made of clayey and sandy sedimental series, which have been deposited at cretaceous. Our aim was to reconstruct the NPHI (neutronic porosity measured in percentage of porosity) from three others logs: PEF (photoelectric effect), RHOB (relative density in gr/cm^3), and GR (gamma ray in API numbers). So we have four log's data : PEF, RHOB, GR and NPHI. Those measures have been taken every half-foot (15.24 cm). In the drilling hole MAR203, we have 5590 measures' levels, and 9962 measures' levels in MAR402. The vertical resolution is 50 centimeters, for MAR402 these data are presented in figure 4-2, 4-3, 4-4, 4-5.

In the following we, first construct a MLP to predict the NPHI value knowing the three other data (PEF, RHOB and GR) for MAR203. The MLP that we consider has two hidden layers with 12 neurons in each hidden layer. In order to estimate the MLP's parameters we construct a learning set by using the 2/3 of MAR203's data. The last 1/3 of the MAR203's data serves as the test set. The network evaluates the error between the NPHI's value that it predicted and the measured one. Then, using retropagation algorithm, the network adjusts all the connection's weights, for minimizing the quadratic error function C. We stop the learning process when the quadratic error on the test set reaches it's minimum value. Then, we freeze the weights, and we use the current MLP for predicting the NPHI's values in MAR402 which was not used in the constitution of the learning and the test sets. So for every point's measure of the MAR402 drilling hole we use the three logs values (PEF, RHOB and GR) as input of the MLP to predict the NPHI log's value. The Root Mean Square error for the test set of MAR402 is equal to *** which prove an accurate mean prediction, Figure 1 shows the reconstructed NPHI in the well MAR402, between 600 and 800 meters versus the observed NPHI. Nevertheless, in some areas, the MLP gives a series of incoherent values, very different from the observation.

We applied now the rejection process training a MLP classifier, as for the preceding MLP the learning and the test sets are the 2/3 and 1/3 of the drilling hole MAR203. The NPHI's measures we are disposing of, are varying from 0.023 to 0.823 % of porosity. We subdivide the range of the NPHI's value in 16 discrete and regular intervals, and we consider a MLP classifier with p = 3 input neurons (PEF, RHOB and GR) and K=16 output neurons. Each one of those 16 neurons represents one of the 16 NPHI's intervals. The activation of the output layer corresponds to the histogram of the conditional NPHI's random variable, knowing the three measures (PEF, RHOB and GR). Figures 2 shows two possible situations; Figure 2-a shows the curve we get when the distribution is unimodal, which is the general case, a single activation pick appears and then NPHI's value is clearly defined; Figure 2-b shows the curve we get when the distribution is multimodal. Several activation picks point out that NPHI's value may belong to several different intervals, and several values of the missing NPHI's log can be proposed from the three measures (PEF, RHOB and GR). So these input's logs have to be eliminated from the MLP prediction process.

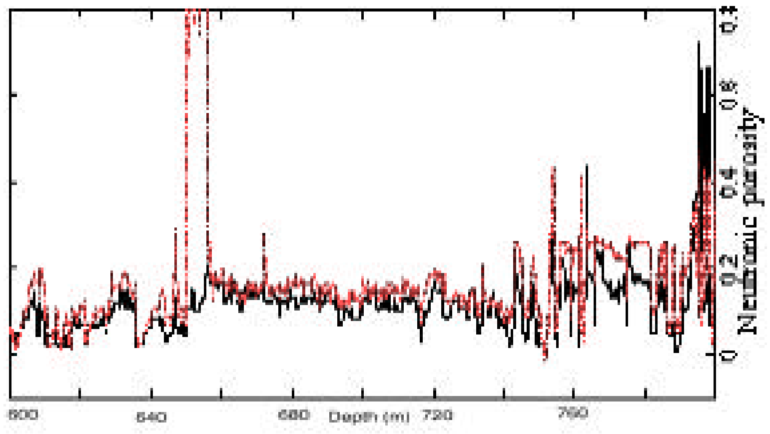
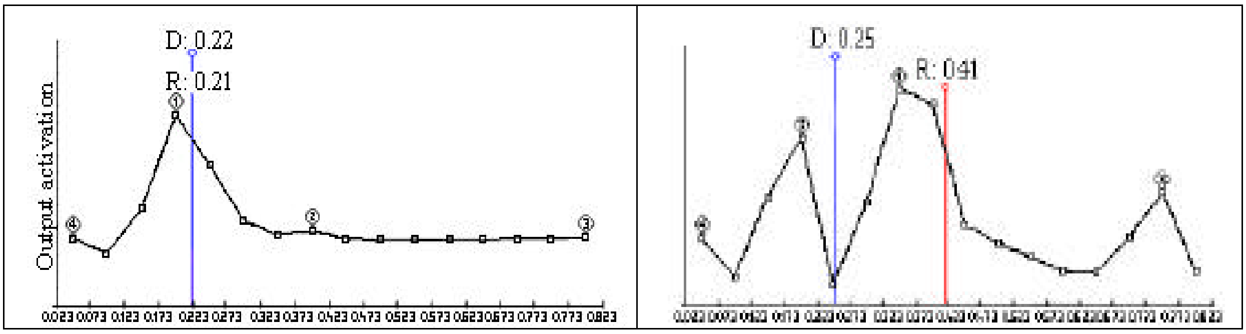


Figure 1: Reconstruction of the NPHI in the hole MAR402, between 600 and 800 meters. Around 650 meters, we show the biggest error made by the MLP. Black line represents measured NPHI and the gray line the predicted one.



(a) (b)

Figure 2: The x-axis corresponds to the 16 intervals of NPHI. The y-axis corresponds to the activation value of the neuron associated to each interval. D is the observed NPHI's value, R is the NPHI's computed by the MLP predictor . (a): The output state curve produced when probabilities' distribution of Y/**x** is unimocal. (b): The output state curve produced when this distribution is unimocal.

In order to decide if the conditional random variable Y/**x** is unimodal or not, we analyse the differences between two picks of the output state curve of the MLP classifier. If this difference is less than a fixed threshold value we then decide that the distribution is multimodal and we eliminate the current vector log's values **x** from the predicting process. On figure 4, in the fifth column, we have plotted the measured NPHI, and in the sixth column, we plotted the computed NPHI (by the MLP predictor). We didn't draw the line of the predicted value when the MLP classifier indicates a multimodal distribution. We can see, next to 650 meters, that the biggest error made by the MLP predictor is thrown out. We have plotted together the predicted NPHI's curves (column 6) and the diametric log (caliper) (column 1). We know that caliper's variations reflect drilling hole's deformations. We find out that rejections coincide with caliper's disturbances, therefore with the hole's deformations. The maximum error is setting just in front of a caliper's break, characteristic of a considerable excavation. In such conditions, logging tools can't give calibrated measures, and the values taken in this place are unavailable. As the input data has no coherence, it is normal that the regression neural treatment provides an aberrant answer. For understanding better what happens, we wished to know subsoil's composition in Marcoule's sector. This is the main goal of the next section where we trained a topological map on the same MAR203 data, as we have four distinct measures for each level ((PEF, RHOB, GR, NPHI) we use all of them in order to get the more accurate reconstruction.

4 Lithologic facies reconstruction using a Self-Organizing Map

To represent the facies of the drilling hole MAR402, we used a Self-Organizing Map (SOM). The map is a discrete set (C) of formal neurons. Each neuron of the map is associated to a referent vector in the data space. The map has a discrete topology defined by an undirect graph, usually a regular grid in one or two dimensions. For each pair of neurons (c,r), the distance $\delta(c,r)$ is defined as being the shortest path between c and r on the graph. The SOM algorithm makes use of a neighborhood system of which the size, controlled by the parameter T decreases as learning proceeds. At the end of the learning algorithm two neighboring neurons on the map have close referent vectors in the data space. Each referent vector defines a particular subset of the data space (usually its Voronoï domain). So the data space is divided in several subsets, each one being represented by a particular neuron of the map. To two neighboring neurons on the map will correspond two close subsets in the data space [Kohonen 1994], [Yacoub and al, 2000].
This partition of the data space permits to affect each vector of the data space to a particular neuron on the map. A generalization of the SOM model, the Probabilistic Self-Organizing Map (PRSOM) [Anouar et al. 1997] uses a probabilistic formalism. PRSOM is a probabilistic model which associates to a neuron c of the map a spherical Gaussian density function f_c., it approximates the data's density distribution using a mixture of normal distributions. This algorithm improves the map's spreading out over the data space.

The learning algorithms of SOM and PRSOM are unsupervised algorithms, which adapt the map to a set of learning samples. Those algorithms allow realizing a partition of the data space, with each subset associated to a one neuron of the map. If we label each neuron of the map with particular rock classes, the map becomes a classifier. Labeling process may be made by an expert or by automatic methods, as hierarchical classification, for more details see [Yacoub and al, 2000], [Frayssinet, 2000], [Gottlib-Zeh and al, 1999].

We trained a rectangular map of size (13x7) with PRSOM algorithm, using the whole set of logging measures we dispose of in the drilling hole MAR203. We take advantage of the knowledge provided by the coring of this hole for labelling the map. Figure 3 shows the topologic map we get. Graphics where made using the package of Kohonen, according to its representations, Grey level between two neighboring neurons represents the distance between them. We can see that rock classes are represented by neurons gathered together. We notify a gradient's presence between the different sediments. The sediments which come from the mainland, as sandstone, stand on the left side of the map, and the sediments which come from lakes or oceans, as limestone, stand on the right side of the map. On the other hand, the lignite is dispersed on different places on the map. It seems to be difficult to recognize.

The labeled map, constructed with MAR203's data, is used as an automatic classification tool for MAR402's data. As such, we draw MAR402's facies (column 6, figure 4). We notice that the biggest MLP prediction error is made in lignite class rock. If we consider the whole facies build by labeled map, we notify that most of ambiguous rejections are concomitant with lignite. Then we can establish a strong correspondence between caliper's disturbances, ambiguous rejections and lignite's presence. Lignite is a delicate rock, which doesn't stand drilling tools. They may product big excavations in lignite's banks, and then the log's measures are unavailable. That's why lignite's situation is not hole defined on the map.

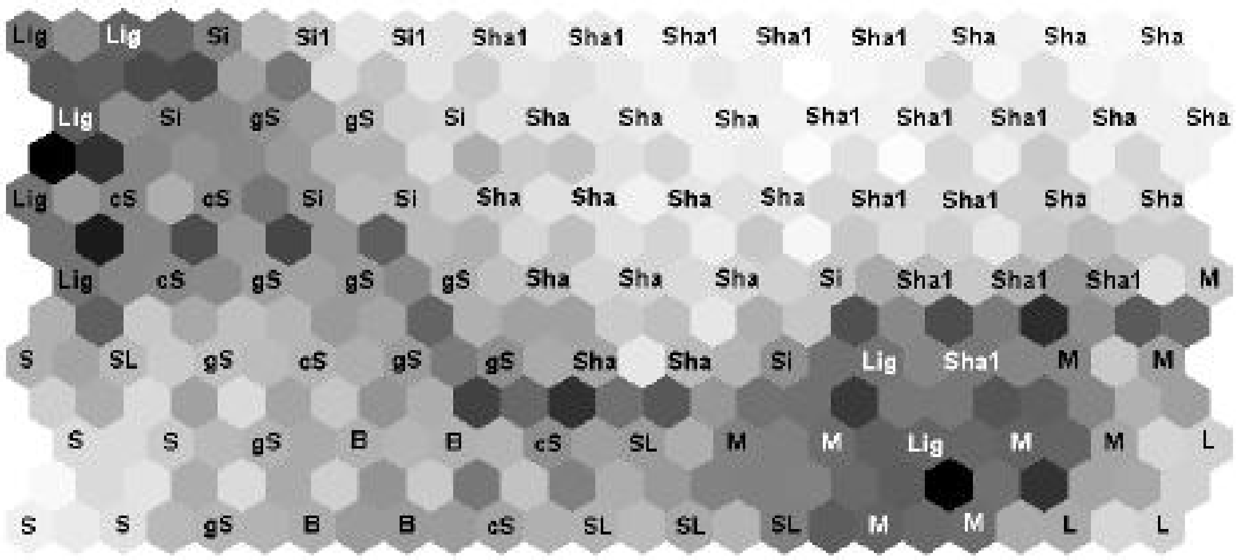


Figure 3: The topologic map established with the values of 4 log's data, PEF, RHOB, GR and NPHI measured in the drilling hole MAR203.
L = limestone, M = marls, gS = glauconitic sandstone, Sha = shales, Sha1 = others shales, Si = silts, Si1 = others silts, cS = coarse grained sandstone, S = strict sandstone, SL = sandy limestone, B = sandy breccia, Lig = lignite.

5 Conclusion

The conjoined use of several kinds of neural networks allows us to settle several tools: reconstruction of missing data, trusting of the results, probabilistic classification of the lithologic facies, detection of measures taken in degrading conditions, detection of input data out of the learning space.

Some points may be improved, as log's selection, probabilities computing before the labeling process, tools convolution's modelisation.

The results described here open the field of neural research for log's data: permeability's study, generalization of neural networks to a drilling field, construction of fictitious drilling hole made of the best log's measures of the drilling field, and, why not, construction of an universal classification tool.

Acknowledgments

This work was supported by Elf Production (Pau, France). The simulations were run by using SNNS provided by Stuttgart University, SN2 software provided by the Neuristic Company, and the Spoutnik software of the LODYC, written by Philippe Daigremont. Figure 3 was made using Kohonen software **This application was made in collaboration with Philippe Rabiller, geological adviser at ELF Production and Stephanie Gottlib-Zeh from ELF Exploration.

References

Anouar F., Badran F. and Thiria S., (1997): Self Organized Map, A probabilistic approach, proceedings of the workshop on Self Organized Maps. Helsinki University of Technology, Espoo, Finland, June 4-6.

Bishop C. (1995): Neural Networks for Pattern Recognition. Oxford University Press.

Frayssinet D. (2000): Utilisation des réseaux de neurones en traitement des données de diagraphies: prédiction et reconstitution de faciès lithologiques. Mémoire d'ingénieur C.N.A.M.

Gottlib-Zeh S., Briqueu L. and Veillerette A. (1999): Indexed Self-Organization Map: a new calibration system for a geological interpretation of logs. In Proceedings of IAMG'99, VI, pp 183-188,. Lippard, Naess and Sinding-Larsen Eds Norway.

Kohonen T. (1984): Self organization and associative memory. Springer Series in Information Sciences, 8, Sprinter Verlag, Berlin (2^{nd} ed 1988).

Kohonen T. (1995): SOM-PAK. The Self-Organizing Map Program Package. ftp cochlea.hut.fi (or 130.233.168.48).

Mejia C.: Architectures Neuronales pour l'Approximation des Fonctions de Transfert: application à la télédétection. Thèse de doctorat, Université de Paris Sud, centre d'Orsay.

Rabaute A. (1999): Obtenir une représentation en continu de la lithologie et de la minéralogie. Exemples d'applications du traitement statistique de données de diagraphie aux structures sédimentaires en régime de convergence de plaques (Log ODP 134, 156 et 160). Thèse de doctorat, Université de Montpellier II. Mémoires géosciences – Montpellier.

Thiria S., Mejia C., Badran F. (1993): A neural network approach for modeling nonlinear transfer functions: Application for wind retrieval from spaceborne scatterometer data. JGR, vol. 98, N° C12, Pages 22,827-22,841, december 15, 1993.

Thiria S., Lechevallier Y., Gascuel O. et Canu S. (1997): Statistiques et méthodes neuronales. Dunod.

Actes des journées scientifiques CNRS/ANDRA. Bagnols-sur-Cèze, 20 et 21 octobre 1997. Etude du Gard Rhodanien. Editions EDP sciences.

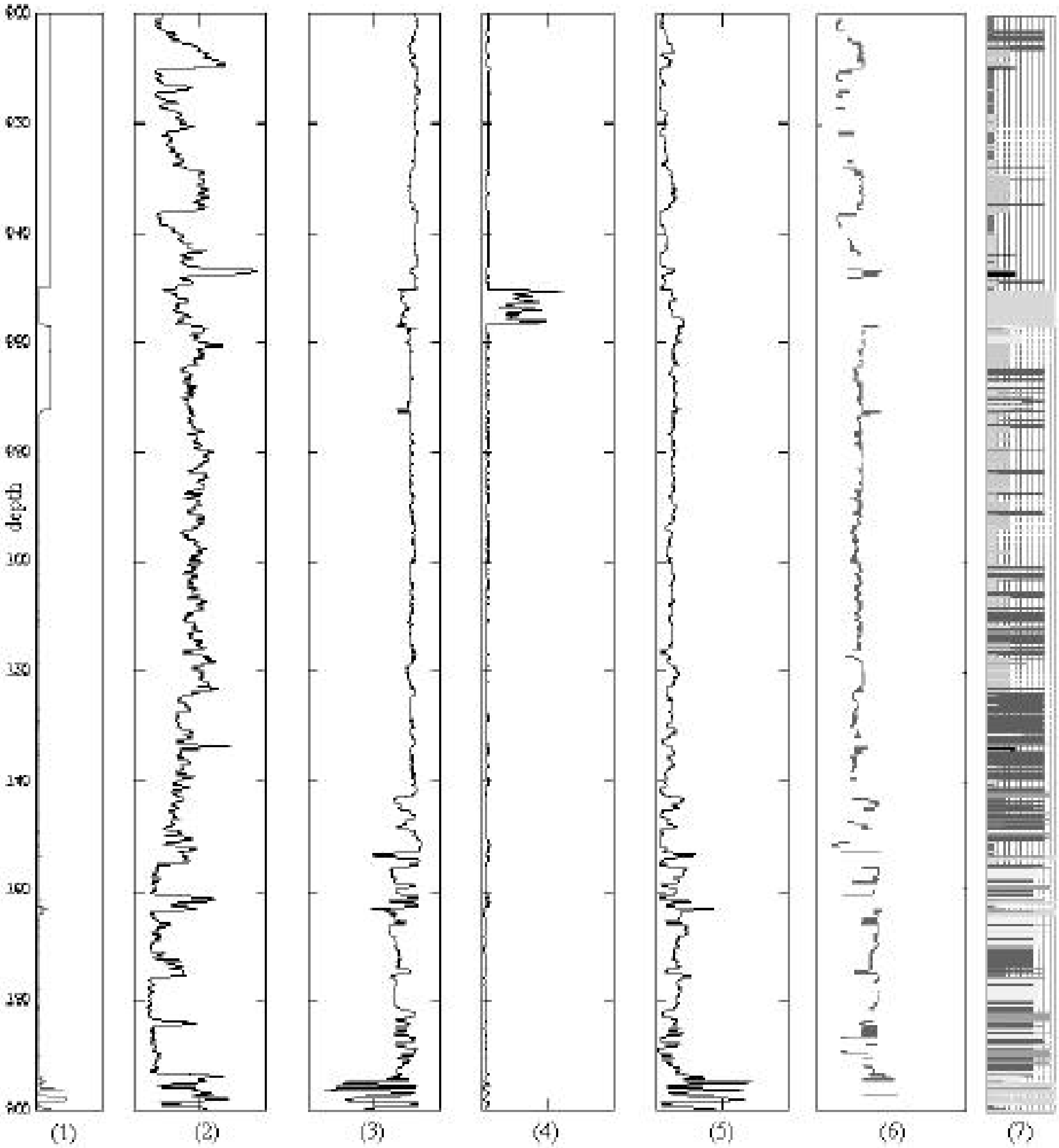


Figure 4: We show the part of MAR402 drilled between 600 and 800 m. On column 1 is the caliper. On column 2, 3, 4, 5 are plotted PEF, RHOB, GR and NPHI. On column 6, the predicted NPHI comports gaps which correspond to ambiguous values which are thrown out by the detection of ouliers. On column 7 is the lithologic facies of MAR402. Around 650 meters, in front of the excavation revealed by the caliper (column 1), there is a big gap in the predicted NPHI (column 6), concomitant with a bank of lignite (column 7).