Nonlinear estimation of missing logs from existing well log data

Rafael Banchs* and Juan R. Jiménez, PDVSA Intevep, Edgar Del Pino, Universidad Simón Bolívar.

Summary

Estimation of petrophysical properties of rocks from well log data plays a very important role in reservoir characterization and simulation. However, it is very common for petrophysicists to face the problem that well log data sets are not available or complete. In this work, we present a methodology that can be used to estimate missing well log data from the existing log data by using genetic algorithms.

Introduction

It is very common for petrophysicists to face the problem that well log data sets are not available or complete. In such cases, petrophysicists usually manage to work only with the available data. However, estimation of missing log data is actually done but, as far as we know, it does not constitutes a common practice yet.

Although some empirical relationships between different log data types has been developed, such as for example velocity and density (Gardner et al, 1974), they happen to be very particular and location dependent. On the other hand, linear regression methods and model based methods are also used for estimating missing well log data (Jain and deFigueiredo, 1982). However, in practice, relationships between different log data types are much more complex and, in most of the cases, nonlinear.

In this work, we present a method for estimating missing well log data from the existing log data at a given well location by using a parametric estimator that is adjusted using genetic algorithms. In the following section, a brief description of the proposed methodology is presented and finally, well log data estimation examples are presented.

Methodology and practical issues

The well log set completion method presented here constitutes a parametric estimation procedure which is based on genetic algorithms. The method is first trained at a well location where a complete log data set exists. In this first step, the genetic algorithm is used to adjust the parameters of a given functional relationship between the target log (which has to be known at the training location) and the source logs (which constitute the known well log data and have to be available at all the well locations of interest).

Then, in a second step, the target log can be estimated at all other well locations where it is missing by using the obtained parameters and the functional relationship along with the source logs.

When looking for functional relationships between data sets, an optimal mapping transformation between the given sets must be obtained. In general, the problem can be stated as follows:

$$g(z) = F\,[r_1(z),\ r_2(z),\ ...\ r_M(z)\,]\ ,\tag{1}$$

where $g(z)$ is the unknown log to be estimated, $r_n(z)$ are the known logs and F represents a functional relation.

In the case of parametric estimation, the mapping transform F is expressed in terms of known functions. Although almost any basis function can be considered, in the particular problem under consideration we have chosen polynomials as the functions to be used for representing the mapping function. This is because polynomial series have the ability of representing any continuous function in a given abscissa interval (examples of this are given by the Taylor series and the Legendre polynomials).

In this way, the target log can be expressed in terms of a polynomial series of the known well logs. Suppose we want to estimate an unknown log g, then the value for the estimated log at a given depth z can be expressed as:

$$\hat{g}(z) = \sum_{i=1}^{M} \sum_{k=-N}^{N} a_{ik}\,[r_i(z)]^{\,k}\ ,\tag{2}$$

where $g(z)$ is the unknown log to be estimated and $r_i(z)$ are the known logs. M corresponds to the total amount of known logs (r_i) to be used in the estimation, N is the maximum order of the polynomials to be used and a_{ik} are the unknown coefficients that must be determined.

In the methodology presented here, the unknown coefficients from equation (2) are determined via an optimization problem, which is reduced to maximize a fitness function of the form:

$$fitness = \exp(\text{-}error)\ ,\tag{3}$$

where the error might be any error measurement between the estimated data and the actual data at the training well, for example:

$$error = \frac{1}{nsamples} \sum_{n=1}^{nsample} |\,g(\,z_n)\text{ - }\hat{g}(\,z_n)\,|\ ,\tag{4}$$

where *nsamples* is the total number of samples in the interval under consideration, $g(z_n)$ is the actual target log sample at depth z_n ; and the values of $\hat{g}(z_n)$ are computed by means of equation (2).

Implementation of such an optimization problem by means of a genetic algorithm requires appropriate coding of the

unknown parameters a_{ik}. The most common way of dealing with this class of problem is by using the concatenated coding scheme, which has been successfully used in multivariable optimization (Goldberg, 1989).

Another important aspect is the normalization of the data sets. Normalization of the data is required in order to improve the methodology performance. After some experimentation, we found the following normalization to provide a reasonable performance:

$$h_i = \frac{r_i - \text{mean}(r_i)}{\max(|r_i|)} + 1 \quad , \tag{5}$$

where h_i is the normalized well log, $mean(r_i)$ is the average value of the log in the interval of interest and $max(|r_i|)$ is the absolute value of the largest deviation. The normalization described in equation (5) guaranties that the average value of the normalized well log h_i is *1* an that its largest deviation will be in between *0.5* and *1.5*. This avoids large power terms in the polynomial series to produce large peaks in the estimated data at those locations where the input data exhibit large deviations.
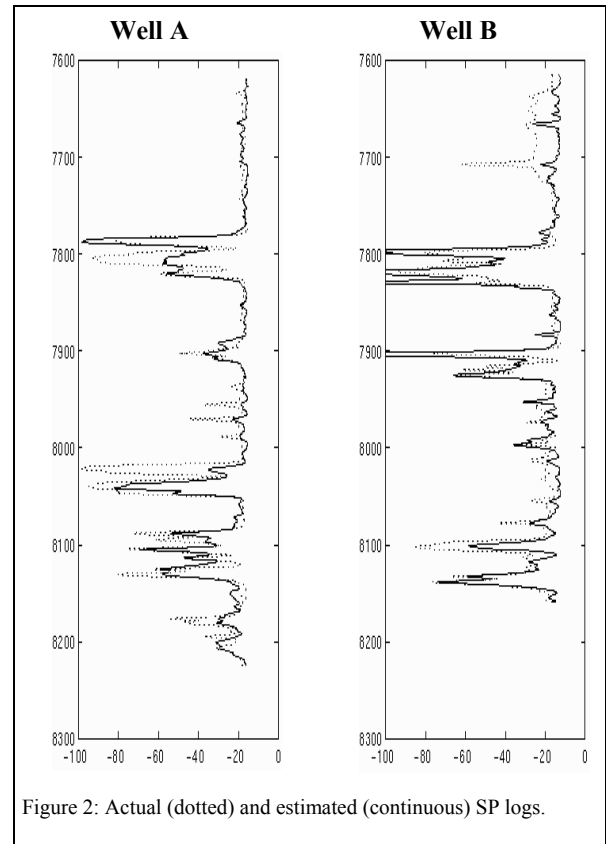
Field data examples

In this section, the methodology proposed above is used to estimate well log data at some well locations in an field from eastern Venezuela. The idea of the presented results is actually to test the proposed technique and for this reason well locations with a complete set of log data were selected, so the estimated and the actual log values can be compared at both, the training and the target locations.

Figure 1: SP vs. GR and estimated mapping function
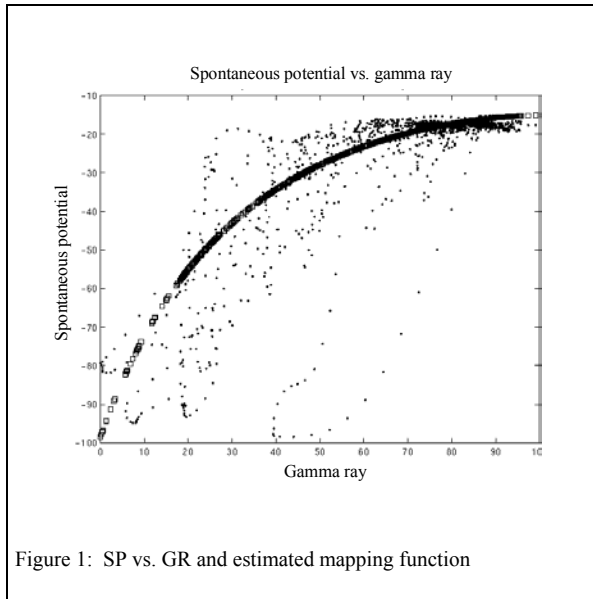
In the first example, the nonlinear capabilities of the method are illustrated by training an estimator for the spontaneous potential log from the gamma ray log. Figure 1 illustrates the crossplot between the actual logs along with the polynomial relationship obtained by the genetic algorithm at an specific well location A.

The mapping function presented in Figure 1 was obtained by using a polynomial of seven terms (*N=3*). The value of N should be selected by trial and error, since there is not a theoretical criterion for such a selection. In fact, small values of N lead to low order polynomials which cannot provide a good mapping function; but, on the other hand, large values of N result in too many degrees of freedom which makes more difficult the solution of the optimization problem.

Figure 2: Actual (dotted) and estimated (continuous) SP logs.

The mapping funcion presented in Figure 1 was used to estimate spontaneous potential logs at the training site (well location A) and a testing site (well location B) which was separated approximately 500 meters from the former. Figure 2 shows the actual (dotted line) and the estimated (continuous line) SP logs at both well locations.

Nonlinear estimation of missing logs from existing well log data

The correlation coefficients between the actual and the estimated SP logs at both locations were 0.83 and 0.77, respectively.

In the second example a sonic log was estimated from other four well logs: GR, LD, LM and SP. The algorithm was trained at well location X and the performance of the estimation was evaluated at well location Y, which is approximately 560 meters away from the former. In this example, a polynomial of seven terms was also used.

Figure 3 shows the actual (dotted line) and the estimated (continuous line) DT logs at both sites, well X and well Y. In this case the correlation coefficients were 0.75 and 0.73 respectively.

Figure 3 shows the high frequency content of the sonic log is lost in the estimation. This is due to the fact that sonic logs have a richer frequency content which cannot be totally recovered from the other logs. However, as we can see from Figure 3, the estimated logs do follow very well the low frequency trends of the actual sonic logs.
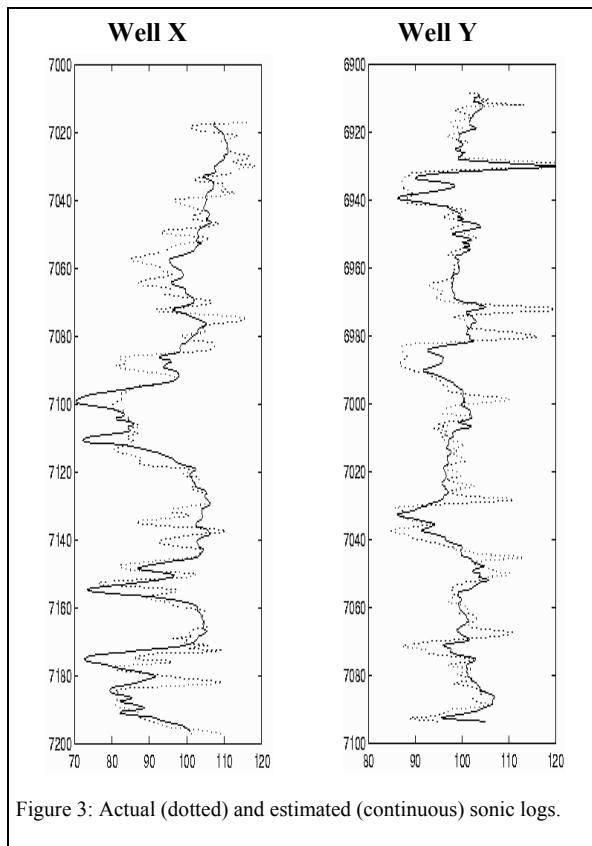

Figure 3: Actual (dotted) and estimated (continuous) sonic logs.

Conclusions and future work

As illustrated in the previous section, the presented methodology provides a good approach for well log data estimation. Different from conventional linear regression techniques, this method allows to represent nonlinear relationships existent among different log data sets.

Nevertheless, some important considerations have to be taken into account in order to improve the performance of the technique:

- Since none spatial information is used during the training process, the quality of the estimation diminishes as the distance from the training location is increased.
- Since relationships among log data also vary with depth, the performance of the technique is affected when large intervals are used during the training step. Then, it is recommended to restrict the analysis to intervals of interest.
- The quality of the estimation also depends on the frequency content of the log data under consideration. We have noticed that high frequency content can be lost or partially recovered.

As a final remark about future developments in this area, we can mention the idea of incorporating information from more than one training location into the estimation model. This can be achieved either by incorporating multiple well log data into the parametric model defined in equation (2), or by including, into the error function defined in equation (4), additional terms for minimizing some crossvalidation measurements.

References

Gardner, G., Gardner, L. and Gregory, A., 1974, Formation velocity and density – The diagnostic basis for stratigraphic traps: Geophysics, v.39, p. 770-780.

Goldberg, D. 1989, Genetic algorithms, Addison Wesley.

Jian, K. and deFigueiredo, R., (editors), 1982, Concepts and techniques in oil and gas exploration, Society of Exploration Geophysicists.

Acknowledgments