Compositional Interpretation of Well-logs

by

J.J. Egozcue (*), R. Tolosana-Delgado (*) and V. Pawlowsky-Glahn (†)

e-mail: juan.jose.egozcue@upc.es; raimon.tolosana@upc.es

e-mail: vera.pawlowsky@udg.es

Abstract

Wireline logging have become one of the most important data sources used by geologists, because they are relatively cheap to obtain, and allow the implementation of quantitative models. These techniques arose in the oil industry, where they are widely used to infer oil content of porous rock formations: the widely-known Archie model relates logs with rock composition, through a set of functional parameters related to mineralogical composition and geological history of the rock formation. The main point is that, for given functional parameters (e.g. cementation coefficient) well-log data can produce results violating the nature of compositions. This approach avoids this problem by assuming these parameters to be random. Then, a statistical treatment is thought to be necessary because: (1) we can assess estimation error and risk, and (2) we will honor both the nature of compositions (through Aitchison’s compositional statistics) and the uncertainty inherent to the parameters values (through the Bayesian approach). Bayes’ Theorem is the tool to treat all this available empirical information and the wireline data. Thus, a compositional bayesian method (BCM) is built and tested against classical interpretation methods. BCM estimations are better, because they are generally statistically reliable and offer a measure of error estimation, although they are affected by the functional parameters’ uncertainty in poor oil rock formations.

1 Introduction

Borehole logging is the most expanded geological data source: since 1950, when Archie published its widely accepted equation, based upon some electric well logs, the logging techniques have become more varied and important. This fact can be explained considering two factors: (1) the new need for quantitative geological models, and (2) the fact that lab data acquisition is usually more expensive than logging data. The urge for more quantitative models is becoming so important (mainly in two fields, resource exploration and environmental monitoring) because they condition the viability of most projects. Furthermore, this economical condition implies also that the estimation errors must be controlled, in order to better assess risk. So, a statistical framework should be used: compositional data analysis techniques (Aitchison, 1986) arise as the best option to give a statistical basis to log interpretation, because they honour the intrinsic geometry of the compositional support, the simplex. Thereafter, we have constructed a new approach to the classic Archie problem, which includes the innermost geometry of compositional space, and a systematic control of uncertainty through a Bayesian approach. Also, we have tested its goodness in two ways: (1) a test of coherency and convergency with some simulated logs, (2) a test of general adjustment of simulations to real data.
2 Archie models

The well-known Wyllie (1) and Archie (2) equations

\[
\phi = \frac{\delta_i - \delta_m}{\delta_f - \delta_m}
\]

(1)

\[
S_w^n = \frac{a \cdot R_w}{\phi^m \cdot R_t}
\]

(2)

relate some logs with the most basic compositional system \(z\), constituted with solid particles \((s)\), brines \((w)\) and hydrocarbon \((o)\). The basic logs needed are porosity (sonic, neutron and/or density logs, expressed by \(\phi\)) and resistivity logs (some of an extensive variety, represented by \(R_t\)).

The link between logs and compositions (in the form of porosity, denoted as \(\phi\), and water saturation, expressed by \(S_w\)) is a set of parameters, which we have called functional parameters. They are represented by the vector \(\psi\). These parameters are: brine resistivity \((R_w)\), scale factor \((a)\), saturation coefficient \((n)\), cementation coefficient \((m)\), brine porosity-device response \((\delta_w)\) and solid porosity-device response \((\delta_m)\). They are believed to describe somehow the formation, and so, we can estimate them if we know the exact type of rock we have: as can be found in Doveton, 1999), cementation and saturation coefficients are related to the geometry of the porosity network and the consolidation history of the formation; Phase responses \((R_w, \delta_m, \delta_f)\) express extreme responses of pure phase components to the log stimuli. Actually, this relationship is not well-established, which means that most times basic log interpretation needs a calibration process to fix the values of these parameters. In a more compact form, these relationships and equations (1) and (2) are summarised as

\[
z = (s, w, o) = g(R_t, \delta_i; \psi) \quad \psi = (a, m, n, \delta_m, \delta_f, R_w)
\]

(3)

Anyway, when these values are known, rock composition can be computed point to point by deep. However, the results goodness cannot be evaluated anyway, because the process avoids purposely a statistical approach.

In a second development generation of methods, graphical treatments have been built up in order to avoid the calibration process: Hingle and Pickett methods and Hough Transform give graphical solutions that are well-exposed in Doveton, 1998). All of them work in the same way: some parameters are a priori needed to plot the log in a graphic, and thus proceed to the interpretation process. As a secondary result, the remaining parameters (those not previously needed) are a posteriori estimated as a function of the observed data logs. However, all methods have neither statistical assumption nor the estimations can be statistically interpreted. Furthermore, the uncertainty derived from fixing a priori some parameters affects enormously the estimation of low-content components: sometimes, negative contents will be obtained, thus falling outside the simplex, which is obviously a gross error.

The third generation methods are based upon the same equations, but are intended to avoid calibration through a statistical approach: they are the GLOBAL method (Mayer i Sibbit, 1980) and the multiple solutions method (King i Quirein, 1987). Others, as the CARBOLOG method
(Carpentier i cols, 1991), propose an indirect method of calibration, based on comparing Total Organic Carbon (TOC) content directly measured in the lab with TOC obtained from log computations.

3 Bayesian Compositional Method

Those problems and misbehaviours explained in the previous section can be successfully avoided when some statistical assumptions are taken:

- the inherent uncertainty of parameters should be interpreted as a probability distribution,
- and a reliable distribution model for the composition \( z \) should be imposed, since we want to know something about it.

There are plenty of possible probability distributions available to model uncertainty of compositions, but Compositional Data Statistics (Aitchison, 1986) must be considered if we want to do it properly: the probability distribution of \( z \) should be defined on the simplex. In this context, a well-known and versatile model is the Additive Logistic Normal distribution \( \text{aln} \), defined by Aitchison, 1986). The \( \text{aln} \) distribution is the result of applying to a multivariate normal distribution the so-called additive logratio transform, which is a non-singular transformation especially useful to analyse low-level content compositions. The \( \text{aln} \) distribution depends on a set of distributional parameters (a vector of means and a matrix of covariances, as any normal distribution) denoted by \( \theta \). For the sake of simplicity, we have considered the covariance matrix equal to identity.

Distributions of \( \psi \) (functional) and \( \theta \) (distributional) parameters are treated in a Bayesian approach as random variables with prior probability distributions, which are updated through Bayes’ formula by the whole wireline log data. In this way, we obtain the posterior distribution of those parameters, that takes into account both prior and data information. This development deserves the name of Bayesian Compositional Method (BCM).

A well-log observed at a given deep is denoted by \( x_j \), where \( j = 1, \ldots, n \) spanning a homogeneous layer. The components of \( x_j \) are \( R_t \) and \( \delta_t \). When \( x_j, j = 1, \ldots, n \) are treated as a matrix it will be denoted by \( x \). In the same way, \( z_j \) denotes a composition (solid, brine and oil contents) at a given \( j \) deep level, and \( z \) is the whole composition treated as a matrix.

Thus, given a set of observations \( x \) and once the functional parameters \( \psi \) are known, compositions \( z \) can be calculated through the Archie equations (1), (2) and (3). With them, for a given value of the \( \theta \) parameters, the likelihood can be calculated as

\[
L(\psi, \theta|x) = \prod_{j=1}^{n} f_z \left( g^{-1}(x_j; \psi)|\theta \right) = \Pr[x|\psi, \theta].
\] (4)

To extract information about the parameters from the sample, we should calculate \( \Pr[\psi, \theta|x] \), the posterior distribution function that can be obtained by using Bayes Theorem with the likelihood (4),

\[
\Pr[\psi, \theta|x] = C \cdot \Pr[x|\psi, \theta] \cdot \Pr[\psi, \theta] = C \cdot \Pr[x|\psi, \theta] \cdot \Pr[\psi] \cdot \Pr[\theta].
\] (5)
As is shown in the last equation, it is necessary to assume that $\psi$-parameters and $\theta$-parameters are independent. This is a hard but unavoidable hypothesis: though a link exists, it is badly known. In the equation (5), $C$ represents a normalizing constant.

Once $\Pr[\psi, \theta|x]$ is obtained, we should marginalize with respect to $\psi$ parameters in order to eliminate their influence from the equation (5), thus obtaining an expression of $\Pr[\theta|x]$ as

$$
\Pr[\theta|x] = \int_\psi \Pr[\psi, \theta|x] \cdot d\psi = 
\int_\psi C \cdot \Pr[x|\psi, \theta] \cdot \Pr[\theta] \cdot \Pr[\psi] \cdot d\psi = 
C \cdot \int_\psi \prod_{j=1}^n \Pr[x_j|\psi, \theta] \cdot \Pr[\theta] \cdot \Pr[\psi] \cdot d\psi = 
C \cdot E_\psi \left[ \prod_{j=1}^n \Pr[x_j|\psi, \theta] \cdot \Pr[\theta] \right].
$$

Equation (6) can be used to extract information about the distributional parameters ($\theta$) describing composition ($z$) from the observed well-logs ($x$).

Computers allow to map the function of the equation (6) for as many values of $\theta$ as desired, effectively constructing discrete approximations to the marginal posterior distribution. Through a Monte Carlo simulation upon $\psi$ parameters we obtain

$$
\bar{L}(\theta|x) = \frac{1}{M} \sum_{m=1}^M \left( \prod_{j=1}^n \Pr[x_j|\psi_m, \theta] \cdot \Pr[\psi] \right),
$$

which approaches equation (6).

In the Archie model context -equations (1) and (2)-, the relationship between compositions ($z_j = (s, w, o)_j$) and well-logs ($x_j = (\delta_t, R_t)_j$) is well-known, and a priori distributions can be defined for all the functional parameters, attending to its intrinsic nature:

- the scale factor $a$ and $R_w$ is globally estimated from an auxiliary Spontaneous Potential logging,
- the cementation coefficient $m$ is supposed to follow a Generalized Pareto Distribution (Embrechts et al.,1997), that can be highly assymmetric,
- the saturation coefficient $n$ can be accurately approximated by a normal distribution,
- the solid response $\delta_m$ should follow also a GPD,
- finally, the fluid response $\delta_f$ is considered a constant, or a very low-variance normal.

The exact shape and scale of these distributions is controlled by their empirically reliable values and preliminary approximations obtained by using classical Pickett and Hingle methods.
4 Discussion

Some simulations have been conducted to test the goodness of estimations obtained with Bayesian Compositional Method (BCM), and with classical methods such as Hingle graphical analysis. In a first step, using a fixed value of \( \theta \)-parameters, we simulate a composition according to the chosen probability model (i.e.: the additive logistic normal, plotted as isoprobability regions in figure 1). The second step is the simulation of \( \psi \)-parameters values, according to their own probability distribution. The third step is simply computation of well-logs for each value of the composition sample and the given \( \psi \)-parameters.

![ALN distribution in the simplex for an oil-rich formation.](image)

From well-logs (plotted in figure 2 as a Hingle graph) to compositions, we can travel through the graphical classical methods (whose results are plotted in figure 3) or the BCM (figure 5). These figures show two main results:

- both estimations are compatible, in a general sense, since classic methods calculate punctual compositions while BCM estimates mean values,

- classic method results may (see figure 3) fall outside the simplex, thus becoming erroneous, if some of \( \psi \) parameters are badly chosen; in the other hand, BCM estimates may not lay outside the simplex.
These results lead to conclude that, although both methods are comparable, BCM estimations are more coherent with physical reality of compositions, as can be seen comparing the figures 6 and 4. Furthermore, BCM \textit{a posteriori} distributions can be interpreted also as a measure of uncertainty. Actually, what classic methods interpret as a series of \textit{different} compositions, is interpreted in the BCM as \textit{error} affecting a central estimation.

Once analysis have been conducted for a single rock model, the same process can be applied to as many simulated rock models as desired, each one characterised by a mean value of the solid, brine and oil contents. Applying BCM to each one, a punctual estimation for the same mean values of the composition can be obtained. Afterwards, the general robustness of the BCM is assessed comparing both \textit{simulating} and \textit{estimated} central values for each component. Results of this sensibility analysis are plotted in the figure 7, where data have been transformed through the additive logratioing in order to improve the resolution in low oil content rock models. In this figure, semaphore-coloured lines indicate the size of the error while the blue-rose map show the direction of error: it is easy to see that low oil content rocks (those plotted as points in the southeast corner of the graph) are affected by systematic overestimating errors; they are the rocks more sensible to calibration processes and to the risk that estimations fall outside the simplex. Thus, BCM estimations in poor reservoirs have low credibility, but better than classic estimations, usually physically impossible.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.pdf}
\caption{Hingle graph of a simulated oil-rich formation.}
\end{figure}

This results are conditioned to the goodness of those hypothesis made through the development process. Thus, it is necessary to analyse real well-log data, in order to test this goodness.
Some wireline logs have been obtained from a field, unknown due to prospecting secret. This secrecy policy is a common practice in the oil-prospecting industries, and means that most times analysts lack the necessary information to follow the calibration processes of traditional methods: thus, a non-calibrating method as the BCM is better in this secrecy environment.

An homogenous stratum have been selected and analysed through BCM: its Hingle graph is plotted in the figure 8, while the a posteriori mean distribution is represented in the figure 9. The hypothesis of aln distribution can be assessed comparing both figures 5 and 9: as can be seen, both a posteriori distributions are quite similar, although variability of the simulated data is greater than that of real logs. At the other hand both real (figure 8) and simulated (figure 2) data logs are also quite similar. Thus, the hypothesis of normality is credible.

Figure 3: Compositions obtained with the Hingle graphical analysis plotted in the simplex over the simulating ALN distribution. Notice a sample point outside the simplex triangle: compositional inconsistency.
Figure 4: Components obtained with the Hingle graphical analysis plotted against the real simulated values.

Figure 5: Mean distribution for compositions obtained with the BCM plotted in the simplex. Compare its center with center of ALN-distribution in figure 1
Figure 6: Components obtained with the BCM plotted against the real mean simulating values.

Figure 7: Map of the estimating error affecting BCM logratioed estimates. The axis show logratio transformations of the mean values used to simulate each rock model. The background colour show the direction of errors, measured as argument of the error vector. The colored lines stand for different quadratic error sizes: green ($0.5 \text{ units}^2$), yellow ($1 \text{ unit}^2$) and red ($5 \text{ units}^2$). Logratio space have been chosen to improve the area where BCM fails.
Figure 8: Hingle graph of real well-log from a homogeneous stratum.

Figure 9: Mean values distributions obtained for the composition with the BCM from real well-logs.
5 Conclusions

These are the main facts here exposed:

- Classic interpretation methods are affected by errors derived from the uncertainty of functional parameters ($\delta_m$-solid sonic response; $\delta_f$-fluid sonic response; $R_w$-brine resistivity; $a$-scale factor; $m$-cementation coeff.; $n$-saturation coeff.) and they may be inconsistent with the special geometry of compositions (positiveness and constant sum of components).

- Bayesian Compositional estimations are generally statistically reliable and offer a measure of error estimation. Furthermore, it does not need any calibration process (i.e. values of $\delta_m$, $\delta_f$, $m$, $n$ are not specified). Nevertheless, they are affected by the functional parameters’ uncertainty; this is important only in poor reservoirs.

- A mathematical conclusion: transformations of the data are sometimes necessary to highlight special characteristics of them. Thus, we have plotted our available data (i.e. the system formed with different proportions of solid particles, brine and oil) in alternate spaces: the simplex itself and some logratio transformation of it. In this latter case, physical sense is dropped to favor discriminability of different populations.

Finally, some words are worth to be said on future advancements. Essentially, the prior treatment of parameters should be refined. In fact, parameters as the cementation coefficient, saturation coefficient and the mean composition parameter could be a priori thought as being dependent. Also variance of logratio transformed variables should worthy be modelled.

References

J. Aitchison, 1986; The statistical analysis of compositional data; Chapman and hall, London.
J. Aitchison, 1997; The one-hour course in compositional data analysis or compositional data analysis is simple; Proceedings of the IAMG’97; CIMNE, Barcelona; pg 3-35.
