# Estimation of petrophysical properties using multiple attributes: Generalizing linear regressions

*Reinaldo J. Michelena and Juan R. Jiménez*, PDVSA Intevep*



## Summary

In this paper, we introduce a fast and simple method to estimate petrophysical properties from well information and two dimensional seismic attributes. The method generalizes the well known linear regression techniques where the reservoir properties are obtained as linear combination of seismic attributes. The estimates are obtained after solving a nonlinear optimization problem that attempts to: a) preserve seismic-property correlations estimated at wells locations; b) minimize the difference between estimated and measured properties; c) minimize the difference between real and estimated correlations among different properties measured at the same locations; d) minimize crossvalidation errors. A genetic algorithm is used to solve this nonlinear optimization problem. The method may uses all selected seismic attributes and all measured properties simultaneously. Final petrophysical property estimates are a linear combination of orthonormal seismic attributes that form a basis for the vector space of the original attributes extracted from the seismic data. By using a synthetic data set, we show petrophysical estimates obtained by using the new method yield mean crossvalidation errors of 10% and the estimated property maps are comparable to those obtained by using a conventional geostatistical technique such as colocated cokriging.

## Introduction

The estimation of petrophysical properties from measurements at selected well locations and low resolution seismic data is a problem that geologists, engineers, and geophysicists face everyday when characterizing the heterogeneities of a reservoir. Since the classic paper of Doyen (1988), geostatistical techniques have become the ultimate tool to solve this problem. These powerful techniques provide not only maps of estimates of reservoir properties but also estimates of the uncertainties in such maps. Unfortunately, as Hirsche and Davis (1997) point out, the use of geostatistics is often restricted to a few "experts" that master all the details in the proper use of the tools required to perform the estimations.

With the use of modern seismic interpretation systems, the extraction of seismic attributes from 3D seismic data is straightforward. Once seismic attributes have been extracted and petrophysical measurements have been incorporated into the data bases, interpreters tend to use conventional linear regression methods rather than geostatistical methods even though linear regression methods are known to have serious limitations. In many cases, petrophysical estimates from linear regressions become the final estimates that are used to make decisions about the reservoir, even thought they are meant to be fast, first order estimates that will be eventually replaced by more accurate geostatistical estimates.

We present in this paper a fast and simple method that overcome some of the limitations of conventional linear regression methods by producing estimates of petrophysical properties that attempt to honor all well information, preserve seismic-property correlations, preserve correlations between different properties measured in the same well locations, and minimize crossvalidation errors.

Even though the method relies on good estimates of the correlation between the seismic attributes and the petrophysical properties we want to estimate, it does not require such correlations to be necessarily high.

We start by introducing the basic theory of property estimation in Hilbert spaces. Then, we show how a property can be estimated as a linear combination of orthonormal seismic attributes weighted by the correlation coefficient between them and the given property. Since this estimate does not honor well information, we show how to modify it to honor such information as closely as possible. The modification consists of the addition of a small perturbation to the correlation coefficient. Such perturbation is estimated by solving a nonlinear optimization problem. Finally, by using a synthetic data set, we show the method yields reliable estimates that are comparable to those obtained by using a conventional geostatistical technique such as colocated cokriging.

## Estimation in Hilbert spaces

If we want to use seismic attributes to estimate a petrophysical property $P_j$ in areas of the reservoir were no information about such property is available, we can assume that both $P_j(x,y)$ and the seismic attributes are functions that belong to a Hilbert space $H$. If the vector subspace of the property is called $P$, and the vector subspace of the seismic attributes is called $E_A$, we can always express the vector subspace $P$ of the property as the sum of $E_A$ plus another subspace $E_{\perp A}$ orthogonal to $E_A$ (the orthogonal complement), as follows (Michelena and Harris, 1991):

$$P = E_A \oplus E_{\perp A}. \tag{1}$$

From the projection theorem (Stackgold, 1979), we can always decompose the property $P_j(x,y)$ into $g(x,y) + f(x,y)$, where $g(x,y)$ belongs to $E_A$ and $f(x,y)$ belongs to $E_{\perp A}$. If the vector subspace of the seismic attributes has dimension $M$ and the functions $\{A_i\}_{i=1}^{M}$ form a basis for this subspace, we can write $P_j(x,y)$ as

$$P_j(x,y) = \sum_{i=1}^{M} \omega_i A_i(x,y) + f(x,y). \quad (2)$$

The first term of equation 2 represents the projection of the property in the basis of the seismic attributes. The orthogonal complement $f(x,y)$ contains information about the property $P_j(x,y)$ that is not captured by the seismic data. The values of the weights $w_i$ can be easily obtained by minimizing the norm of the orthogonal complement $f(x,y)$ as follows:

$$min \left\| f(x,y) \right\|^2 = min \left\| P_j(x,y) - \sum_{i=1}^{M} \omega_i A_i(x,y) \right\|^2 . \quad (3)$$

The solutions $w_i$ of equation 3 are obtained after solving a linear system of equations whose matrix elements are the inner products among the different elements of the basis of the attributes subspace. If we assume that such basis function is orthogonal, the matrix becomes diagonal (Stackgold, 1979) and the estimate $P_j^*(x,y)$ of the property $P_j(x,y)$ can be expressed as

$$P_j^*(x,y) = \sum_{i=1}^{M} \langle P_j, A_i \rangle \, A_i(x,y), \quad (4)$$

where $\langle P_j, A_i \rangle$ is the inner product of the petrophysical property function and the seismic attribute $A_i(x,y)$. Notice that the estimate $P_j^*(x,y)$ given by equation 4 still depends on the unknown function $P_j(x,y)$ at every point in the area of interest. However, if we assume stationarity, we can estimate the inner product $\langle P_j, A_i \rangle$ from the expression

$$\langle P_j, A_i \rangle \approx r_{ij} \sqrt{\langle P_j, P_j \rangle \langle A_i, A_i \rangle}, \quad (5)$$

where $r_{ij}$ is an estimate of the correlation coefficient between $P_j(x,y)$ and $A_i(x,y)$ that is assumed to be equal to the correlation coefficient between $P_j(x,y)$ and the seismic attributes $A_i(x,y)$ measured at the well locations. Equation 5 assumes that both $P_j(x,y)$ and $A_i(x,y)$ have all zero mean. This equation can be further simplified if we normalize both the property and the seismic attributes ($\langle P_j, P_j \rangle = \langle A_i, A_i \rangle = 1$), resulting the following expression for the estimate $P_j^*(x,y)$ of the petrophysical property $P_j(x,y)$:

$$P_j^*(x,y) \approx m_j \sum_{i=1}^{M} r_{ij} A_i(x,y) + b_j . \quad (6)$$

The variables $m_j$ and $b_j$ are scaling and translation constants used to preserve the original range of variation of the function $P_j(x,y)$ measured at the wells.

For every point $(x,y)$, equation 6 provides an estimate of the property $P_j$ that is a linear combination of orthonormal seismic attributes weighted by the correlation coefficient. Attributes with high correlation coefficient contribute more to the estimate than attributes with low correlation coefficient. This estimate preserves the correlation between the seismic data and the petrophysical property but has the disadvantage that does not make any attempt to honor such properties at well locations. Next section explains how to deal with this issue.

## Nonlinear optimization problem

To overcome the limitation of the estimate $P_j^*$ not honoring the values of the property at the wells, we perturb each correlation coefficient $r_{ij}$ by adding a small parameters $e_{ij}$ such that the seismic-property correlation is still preserved, within a certain small error, and the well information is honored as closely as possible. The expression for the new estimate is

$$P_j^*(x,y) \approx m_j \sum_{i=1}^{M} (r_{ij} + e_{ij}) A_i(x,y) + b_j . \quad (7)$$

The parameters $e_{ij}$ can be obtained after minimizing the following objective function that includes information about more than one petrophysical property and other requirements besides honoring the well information:

$$\min \left\{ \begin{array}{c} \alpha \sum_{j=1}^{L} \sum_{k=1}^{N} \left| P_j(x_k,y_k) - P_j^*(x_k,y_k) \right| + \\ \beta \sum_{j=1}^{L} \sum_{h=j}^{L} \left| \langle P_j, P_h \rangle - \left\langle P_j^*, P_h^* \right\rangle \right| + \\ \gamma \sum_{j=1}^{L} \sum_{i=1}^{M} |e_{ij}| \end{array} \right\} . \quad (8)$$

In the previous equation, $L$ is the number of petrophysical properties we want to estimate and $N$ is the number of wells where those properties have been measured. The coordinates $(x_k, y_k)$ indicate the positions of the wells.

The objective function 8 is nonlinear in the perturbations $e_{ij}$. We find the set of perturbations that minimize this nonlinear equation by using a genetic algorithm. Genetic algorithms are easy to use, require only evaluations of the objective function, attempt to find the global minimum, and provide solutions that are independent on the choice of the initial model (Goldberg, 1989). Since the sum $(r_{ij} + e_{ij})$ is assumed to be a correlation coefficient, it is constrained to belong to the interval $[-1, 1]$.

The first term in the objective function tries to minimize the difference between properties $P_j(x_k,y_k)$ measured at the wells and properties $P_j^*(x_k,y_k)$ estimated at the same locations; the second term tries to preserve the correlations between properties measured at the wells where more than one property is available; the third term forces $e_{ij}$ to be small such that the seismic-property correlations are preserved as closely as possible across the area of interest. In essence, the addition of the small parameter $e_{ij}$ to the correlation coefficients $r_{ij}$ acknowledges the presence of errors in such coefficients and tries to compensate these errors by including additional well information from one or more properties.

If the weighting factors $\alpha$, $\beta$, and $\gamma$ are set before solving the optimization problem 8, their relative sizes will determine the relative contributions of the different terms in the final estimate $P_j^*(x, y)$. If the second term is included in the optimization problem (i.e. $\beta \neq 0$), all properties are estimated simultaneously. Otherwise, the estimation of one property is performed independently of the rest. If the coefficients $\beta$ and $\gamma$ are both zero, the estimate will attempt to honor only the well information. The parameters $\alpha$, $\beta$, and $\gamma$ are all positive and satisfy the constraint $\alpha + \beta + \gamma = 1$

As we will show in the examples below, the weighting factors can be estimated such that the mean crossvalidation error is minimum. The search of the optimum weighting factors can be seen as an optimization problem by itself that we can be solved efficiently by using gradient based methods.

## Synthetic examples

To test how the algorithm performs when estimating a single property ($\beta = 0$) from well information and seismic attributes, we used a synthetic data set that consisted of the map of the property (Figure 1), the value of such property at 42 locations (which is assumed to be the "given" well information), and 27 attributes generated by doing a non conditional simulation turning band that preserved the variogram of the property. As Figure 1 shows, the property presents North-South alignments with lower values towards the North.

From the original 27 attributes we generated, we ended up using only six for the estimation. We did not used those whose correlation with the well information was less than 0.1, those whose trend was not the expected one (which means that, in practice, we should examine all attributes and keep only those that make more sense from the geological point of view), and those whose correlation with any other attribute was above 0.9. The remaining attributes were transformed into orthonormal vector set by using the process of Gram-Schmidt orthogonalization.

The precision selected for the parameters $e_{ij}$ was 0.0002, a population of 50 individuals was used, and the iterations of the genetic algorithm were large enough (typically 4000) to assure convergence was reached.

Figure 2 shows the map of the estimated property using all 42 wells. The general trend of lower values towards the North is as expected from Figure 1 (the "real" property). The correlation between real and estimated property maps is 0.64 and the mean relative error is 8% at the well location. Figure 3 shows the absolute relative error between real and estimated maps. Ideally, this kind of map should look similar everywhere, which means the algorithm performs similarly for different locations and different property values. However, we see higher property values to the South are estimated more accurately than lower property values to the North. The reasons for these systematic variations in the estimation errors are not clear yet.

Figure 4 shows the crossvalidation crossplot obtained for the value of $\alpha$ (0.95) that minimizes the mean crossvalidation error. Property values at well locations estimated without including such information explicitly in the optimization problem are estimated with a mean error of 10%.

Figure 5 shows a map estimated using colocated cokriging with the attribute that showed the highest correlation (0.8) with the petrophysical property at the 42 wells. Since the seismic-property correlation at the wells is so high in this case, the final property estimate shown in Figure 6 is almost identical to the attribute map that was used for the cokriging. In this case, the correlation between the real map (Figure 1) and the map estimated by using cokriging is 0.56.

When comparing Figures 2 and 6, we observe the correlations between real and estimated maps is higher for the map of Figure 2 obtained by using the procedure proposed in this paper. Therefore, at least for this particular example, property estimates obtained after minimizing expression 8 are at least comparable to estimates obtained by using geostatistical techniques in the way they are most commonly applied in the industry. However, we must keep in mind that property estimates obtained by using cokriging may still be improved if we include more attributes in the estimation procedure. In our example, if we wanted to perform cokriging estimation using the same 6 attributes we used to generate Figure 2, we needed to generate 7 autovariogram models 21 crossvariogram models. Such task was beyond the scope of this paper.

## Discussion and final remarks

We have presented a new method to estimate petrophysical properties using seismic attributes and well information that attempts to honor the given well information, seismic-property correlations, property-property correlations (in case of more than one petrophysical property), and minimizes crossvalidation errors. The estimate of the petrophysical property is a linear combination of orthonormal seismic attributes.

Even though the results presented in this paper with synthetic data are encouraging, there are still many issues related to the new method that need more research. The selection of the proper attributes is a key issue. Even if the method may include any attribute in the estimate, we should be careful in understanding which of them are more related to the property we want to estimate to make sure we use them. Including attributes that have negligible correlation with the property, or behave in a way that is against our geological intuition about the study area, may prevent us to obtain good estimates. The relation between number of attributes to use and the number of wells available requires also further attention.

As with other estimation methods, this one may be in trouble when the property we want to estimate is non stationary. In may also present difficulties when trying to estimate abrupt changes in the property whose seismic expression is smoother. The method provides no estimates of uncertainty. More research needs to be done to understand and solve all these issues.

## Acknowledgements

Many thanks to Robert Porjesz for generating the synthetic data used to test the algorithm and to Ezequiel Gonzalez for performing the estimation using cokriging. Thanks to PDVSA Intevep for permission to publish these results.

## References

Doyen, P. M., 1988, Porosity from seismic data: A geostatistical approach: Geophysics, **53**, 1263–1275.

Goldberg, D., 1989, Genetic algorithms in search, optimization and machine learning: Addison-Wesley.

Hirsche, K., Porter-Hirsche, J, Mewhort, L., and Davis, R., 1997, The use and abuse of geostatistics: The Leading Edge, **16**, 253–260.

Michelena, R. J., and Harris, J. M., 1991, Tomographic traveltime inversion using natural pixels: Geophysics, **56**, 635–644.

Stackgold, I., 1979, Green's functions and boundary value problems: John Wiley & Sons, Inc.

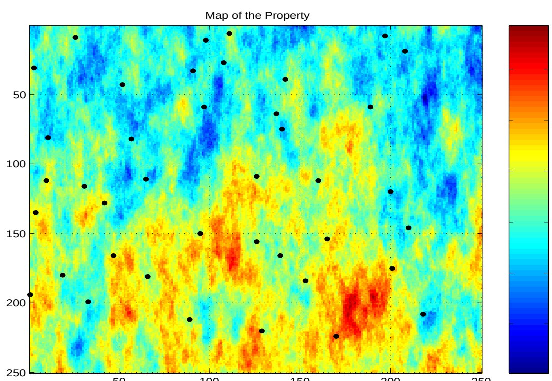


Fig. 1: Original map of the property

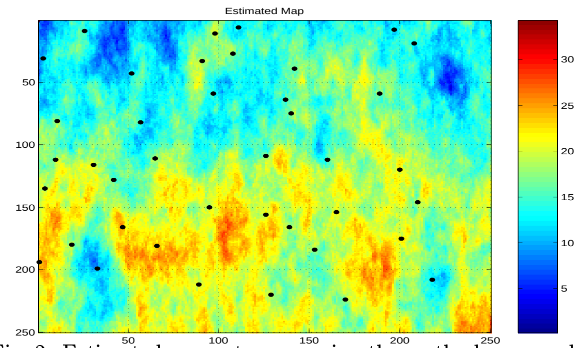


Fig. 2: Estimated property map using the method proposed in this paper. The correlation coefficient between this map and Fig. 1 is 0.65

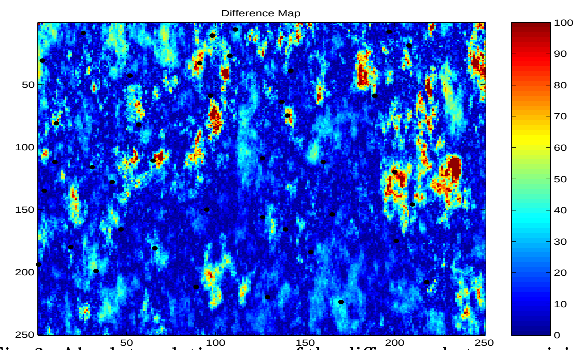


Fig. 3: Absolute relative error of the difference between original and estimated map

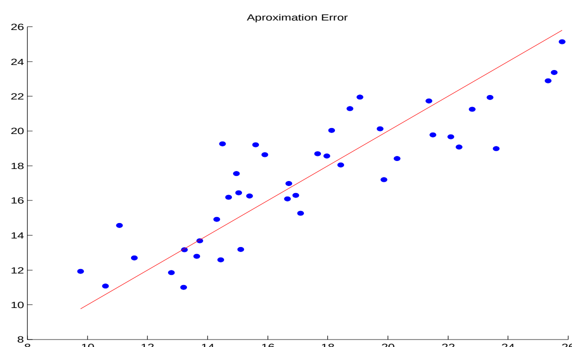


Fig. 4: Crossvalidation error. The mean crossvalidation error is 10 %

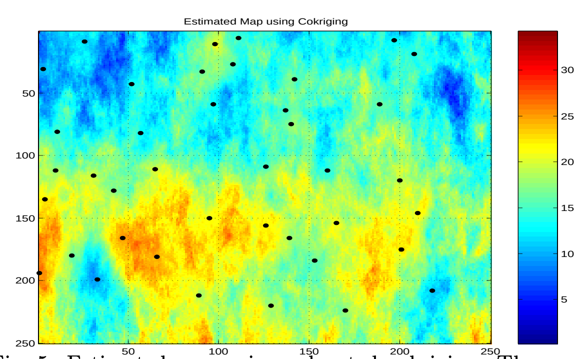


Fig. 5: Estimated map using colocated cokriging. The correlation coefficient between this map and Fig. 1 is 0.56