Automated facies classification in borehole log data

Roman Beloborodov
CSIRO Deep Earth Imaging Future Science Platform
Kensington, WA
Roman.Beloborodov@csiro.au

James Gunning
CSIRO Energy

Clayton, SA
James.Gunning@csiro.au

Marina Pervukhina
CSIRO Energy

Kensington, WA
Marina.Pervukhina@csiro.au

Michael Ben Clennell
CSIRO Energy

Kensington, WA
Ben.Clennell@csiro.au

Irina Emelyanova
CSIRO Energy

Kensington, WA
Irina.Emelyanova@csiro.au

Juerg Hauser
CSIRO Deep Earth Imaging Future Science Platform
Black Mountain, ACT
Juerg.Hauser@csiro.au

ABSTRACT

Seismic facies inversion relies on accurate facies–specific elastic parameter models determined from borehole log data. Typically, velocity depth trends and corresponding standard errors are manually determined for each of the facies by using cross-plots and a number of rock physics templates. The resulting facies classifications and models are often contentious due to loading effects and the subjective nature of the workflow. In complex scenarios, where multiple related rock types occur, it is even more challenging to identify all the rock types, select, and tune appropriate rock physics models (RPMs) objectively. It is highly desirable to perform this task automatically, using robust tools that require human intervention only at the quality control stage.

To this end, we introduce a modified Expectation Maximisation algorithm for simultaneous facies classification and fitting of RPMs. It operates in a semi-supervised fashion on multivariate well log data and is robust to outliers. The practical advantages of this approach are illustrated using data from the Opel-2 well, located in the Laverda Field of Carnarvon Basin, 58 km north-northwest from Exmouth, Western Australia. We compare the classification results to geological interpretation and furnish estimated parameters of the fitted RPMs.

THEORY

EM algorithm. The Expectation Maximisation algorithm (Dempster et al., 1977) is typically used for inference in Gaussian mixture models to identify the parameters of the mixture components. In this study we discuss its application to facies classification and simultaneous fitting of rock physics models for different rock types.

To begin with, we model observed well log properties P_q for a discrete class q (e.g. sandstone, shale) by the following Normal form:

$$P_q \sim \mathcal{N}(\mu_{P,q}, \sigma_{P,q}^2), \qquad (1)$$

where $\mu_{P,q}$, $\sigma_{P,q}^2$ are facies and property-specific means and variances. Most properties of sedimentary rocks depend on depth (d) due to lithogenesis and stress. For example, the elastic parameter pair {velocity, density} is ideally modelled by forms

$$V_q(d) \sim \mathcal{N}(\mu_{V,q}(\mu_{\rho,q}(d)), \sigma_{V,q}^2), \qquad (2)$$

$$\rho_q(d) \sim \mathcal{N}(\mu_{\rho,q}(d), \sigma_{\rho,q}^2), \qquad (3)$$

where the seismic velocity is linked to a density-depth trend through an RPM defining the functional trends $\mu_{V,q}, \mu_{\rho,q}$. Different rock types are parametrised by different RPMs, which allows predicting the seismic response of an altered (e.g., saturated, stressed) rock at different depths away from the well. Given the data tuples $\{P_q\}$, EM algorithm classifies them into

one of the pre-declared number of rock types N_c by objectively selecting and fitting the most likely model from a number of instances of models from the supplied rock physics model library for different lithologies. We entertain a number of plausible RPMs; each has typically a small handful of free parameters, but all are somewhat detailed to explain. The following simplest example is representative.

Example RPM. The *silty-shale* RPM relates seismic velocities V_P, V_S and porosity ϕ of a shale to its bulk density ρ_b and depth d from the well log, assuming that the porosity reduction is purely due to the water expulsion in the process of mechanical compaction. The model depends on a number of fixed parameters: the density of dry clay ρ_c, silt ρ_q, and water ρ_w, and the bulk K_q and shear μ_q moduli of silt. The remaining *free* parameters include the silt-to-clay ratio α and elastic moduli of clay minerals K_c and μ_c. First, the density trend is defined by fitting an exponential function of depth to a bulk density log:

$$\rho_b = m_1 - m_2 e^{-m_3 d}, \qquad (4)$$

where $m_i \geq 0$ are free parameters. The associated total porosity prediction is

$$\phi = \frac{\rho_c - \rho_b + \alpha(\rho_q - \rho_b)}{\rho_c - \rho_w + \alpha(\rho_q - \rho_w)} \qquad (5)$$

The porosity of a shale matrix without silt inclusions ϕ_{sh} is related to the total porosity through a coefficient C via

$$\phi_{sh} = \frac{\phi}{C}, \quad C = \phi + \frac{1 - \phi}{1 + \alpha}. \qquad (6)$$

The saturated clay matrix bulk K_{sh} and shear μ_{sh} moduli are estimated by

$$K_{sh} = K_c(0.5 - \phi_{sh}), \quad \mu_{sh} = \mu_c(0.5 - \phi_{sh}), \qquad (7)$$

and the effective moduli K_{eff} and μ_{eff} of a saturated silty-shale are approximaed by a Hashin-Shtrikman lower bound:

$$K_{eff} = \left[\frac{C}{K_{sh} + \frac{4}{3}\mu_{sh}} + \frac{1 - C}{K_q + \frac{4}{3}\mu_{sh}} \right]^{-1} + \frac{4}{3}\mu_{sh} \qquad (8)$$

$$\mu_{eff} = \left[\frac{C}{\mu_{sh} + Z_{sh}} + \frac{1 - C}{\mu_q + Z_{sh}} \right]^{-1} + Z_{sh}, \qquad (9)$$

$$Z_{sh} = \frac{\mu_{sh}}{6} \frac{9K_{sh} + 8\mu_{sh}}{K_{sh} + 2\mu_{sh}}. \qquad (10)$$

Finally, the velocities are calculated by:

$$V_P = \sqrt{\frac{K_{eff} + \frac{4}{3}\mu_{eff}}{\rho_b}}, \quad V_S = \sqrt{\frac{\mu_{eff}}{\rho_b}}. \qquad (11)$$

Initialisation. The EM algorithm requires fixing the number of classes in the data and supplying initial guesses for class probabilities for each data point, referred to as memberships. These can be assigned randomly or via any other clustering technique. Then the EM algorithm cycles through its Maximisation (M) and Expectation (E) steps until convergence. For the example in the following section we use random initialisation.

M-step. In the M-step, model parameters for each class are estimated assuming the memberships fixed at the current iterate of the algorithm. These model parameters include class proportions, location and scale for trend-less models, and coefficients of the trend function for depth dependent properties and the scale of the scatter around this trend.

E-step. In the E-step, memberships for each point are reevaluated using the model parameters determined in the preceding M-step.

Mitigating the effects of outliers in the data

Clustering and regression operations on multivariate petrophysical data is challenging due to the presence of outliers in the data. Sources of these outliers include poor tool coupling, the presence of rare lithologies with extreme properties, or ultra-thin beds in the measured formation.

Estimating the location and scale parameters of the property distribution using compact Gaussian models is not appropriate for data contaminated by strong outliers. Outlying values corrupt the statistics of different properties by skewing the mean values and trends of depth-dependent properties, and inflating the estimated variance of the residuals. Similarly, using Gaussian models to estimate the point memberships results in assigning the outliers to classes with a small number of observations. By modifying the M-step (parameter regression), and E-step (classification) parts of the algorithm separately, we have reduced the algorithm sensitivity to outliers as follows.

Robust M-step – Parameter estimation. At the M-step, the location parameter of the trend-less property is estimated by the weighted median. The scale parameters are estimated using the weighted median absolute deviation (MAD) calculated as a weighted median of the absolute deviations of the elements of a sequence from their weighted median. Both these estimators for the location and scale parameters are non-parametric, and have a high breakdown value of 50 % for symmetric distributions, and are formally consistent.

Robust M-step – Regression. For the depth dependent properties, instead of estimating the location parameter we use the value of the RPM at the corresponding depth. RPMs are fitted using a bound–constrained optimisation algorithm for non-linear regression. At each M-step, a number of predefined RPMs are fitted to membership–reweighted data. The fitted parameters of these models are then used in the E-step for the (re)estimation of memberships. To reduce the effect of outliers on regression, we minimise the Cauchy loss function and also use MAD to estimate the scale of a data scatter around the trend.

Robust E-step – Classification. Membership update is robustified by using a Student-t distribution PDF defined by the property distribution parameters estimated in the M-step. The Student-t distribution has heavier tails compared to the Normal distribution when its degree of freedom parameter is assigned a small value that reduces the classification error by modelling outliers as an inherent feature of the data being classified.

The simple example in Figure 1 illustrates the performance of standard and robust EM algorithm implementations applied to clustering of points generated from two uni-variate normal distributions with 20 % outliers. Here, the standard EM algorithm using Gaussian models fails, whereas its robust implementation recovers the original distributions securely.

FIELD EXAMPLE – OPEL 2 WELL

The algorithm was tested on data from the Opel-2 well in North West Shelf of Australia. The well intersects a number of lithologies of different geological age. Going from the top to bottom of the well, we encounter the Tertairy Cardabia calcilutite followed by Cretaceous Korojon and Toolonga calcilutites. The Cretaceous proceeds into the thin layer of Windalia calcilutite, followed by Muderong shale, and is concluded by poorly sorted fine-to-medium grained sandstone of the Lower Barrow formation interbedded with claystone. Jurassic section below is represented by shales of Lower Barrow formation that transitions into reservoir hosting Macedon mudstone. Macedon formation sandstones constituting hydrocarbon reservoir with a clear gas-to-oil contact transition from poorly-sorted medium-to-large grained sand on top to fine well sorted clean sand at the bottom.

Figure 2 illustrates the classified logs from the well. All the formations are well matched by the identified facies. For this example, three custom theoretical models were considered in this well each representing different lithology: calcilutite, silty shale, and soft sand. Seven instances of the aforementioned models completely describe the litho-fluid types given the borehole data, which is illustrated by colour-coding on the figure and trends through the classes. Some rock physics model parameters of interest for different formations are summarised in Tables 1 and 2 and compared to geological field description and well interpretation respectively. Here, shale models slightly underpredict the clay volume fraction in the three reported intervals, but are consistent with the decrease of clay fraction with depth. Sandstone models show very close porosity levels to that of petrophysical interpretation in a ballpark of 30 %. Water saturation levels,however, show somewhat pessimistic prediction by overstating the water saturation levels by 20 to 30 percent in gas and oil sands, respectively. Nevertheless, this is an excellent result for completely automated rock physics and petrophysical workflow which is subject to further interpretation and adjustment by an expert petrophysicist. For example, significant improvements can be achieved by explicitly specifying some fitting parameters in the model such as gas and oil densities, bulk moduli of clay and silt and etc. given laboratory and field test results.

CONCLUSION

We have developed a semi-supervised algorithm for facies classification using multivariate log data coupled with rock–physics models for loading. The algorithm identifies a predefined number of facies and fits appropriate rock physics models for each of them in a robust fashion. The suggested method allows fast and unbiased facies classification that results in radical time reduction between exploration and production. The performance of the algorithm is illustrated on real data. Outputs of the algorithm include facies labels and free parameters in the corresponding RPMs, including but not limited to porosity, sorting and cement volume fraction. These parameters can be easily interpreted and directly used in downstream workflows such as facies-based seismic inversion.

REFERENCES

Dempster, A.P., Laird, N.M. and Rubin, D.B. [1977] Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.
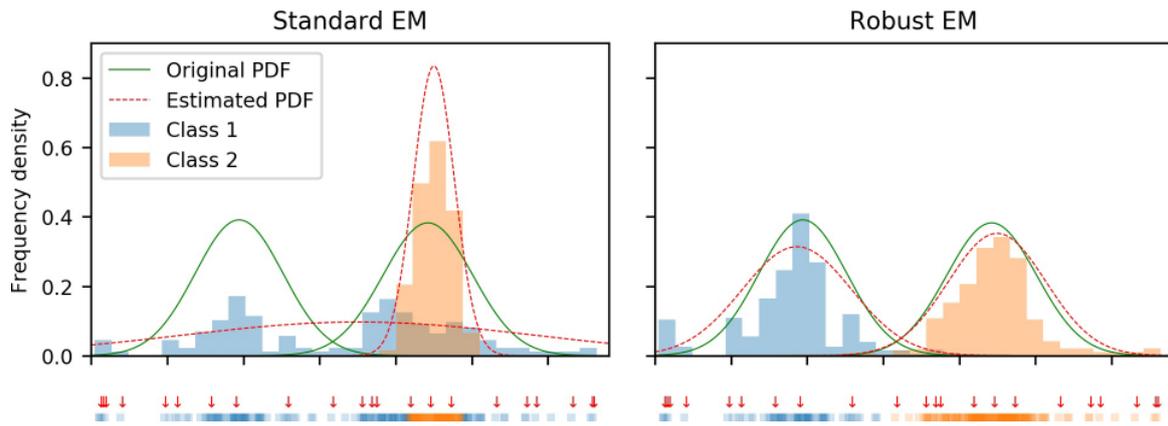
Figure 1: Classification example illustrating the performance of standard and robust EM algorithm implementations. Data points at the bottom of each panel are sampled from two uni-variate normal distributions (green PDFs) and contaminated with random noise (red arrows). Histograms represent the determined class distributions weighted by the estimated point memberships.
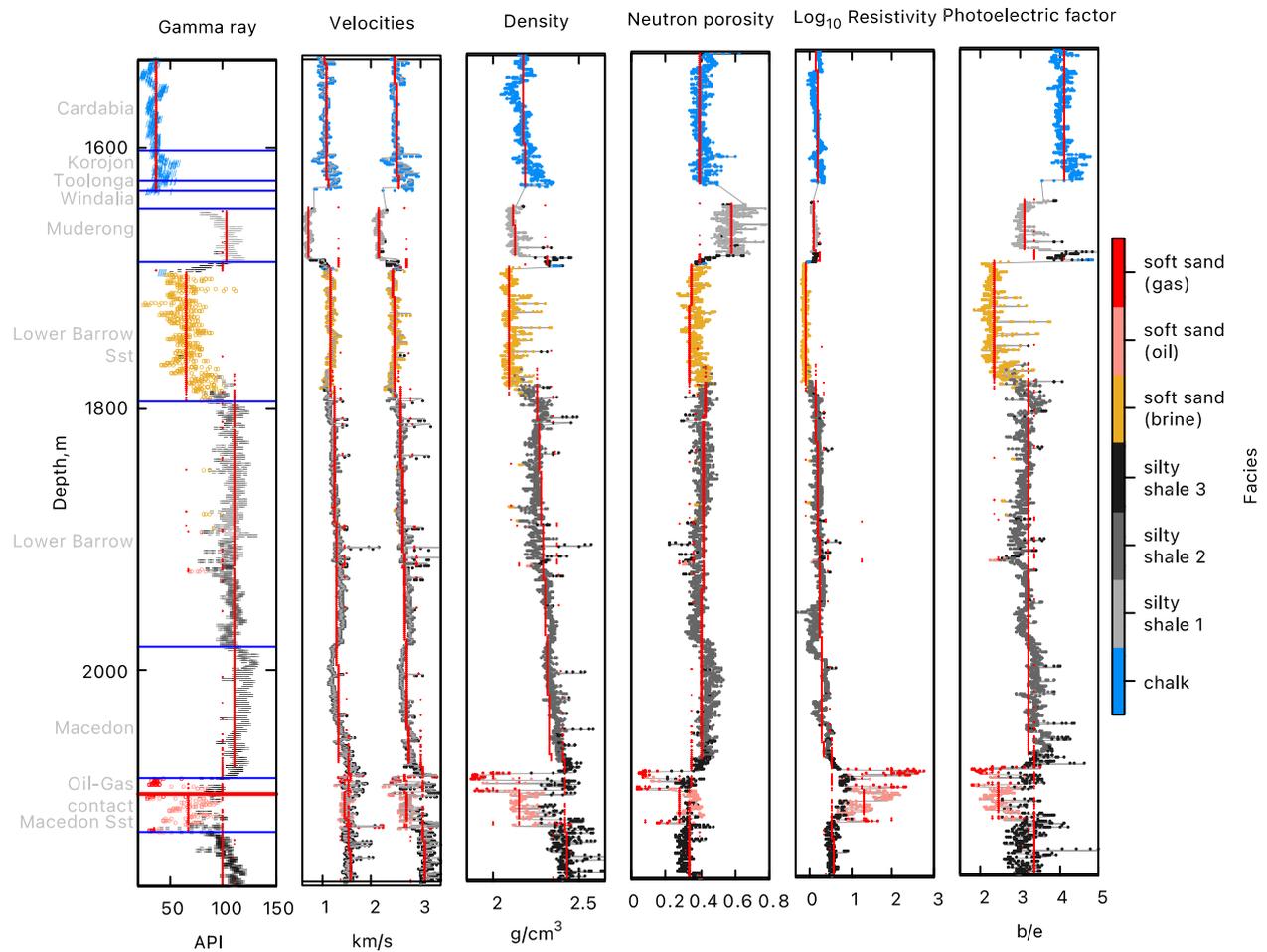

Figure 2: Opel-2 well classification example using rock physics models. In the Gamma ray panel, solid blue lines indicate formation tops from geological interpretation and horizontal red line shows gas-oil contact determined in the well. Sub-vertical red lines in all the panels illustrate RPM trends identified by the algorithm.

Table 1: Parameters of the fitted RPMs for shales

Facies	Clay fraction, %	
	model	field description
Shale 1	96	98
Shale 2	73	80
Shale 3	52	70

Table 2: Parameters of the fitted RPMs for reservoir sandstones

Facies	Porosity, %		Water saturation, %	
	Model	Petrophysical interpretation	Model	Petrophysical interpretation
Soft sand (gas)	28	30	35	14
Soft sand (oil)	30	31	52	23