geoconvention
Calgary • Canada • May 11-13 2026

Logistic Regression in Facies Analysis

John Pendrel
Consultant

Summary

Logistic Regression (LR) and Bayesian inference were used to estimate facies in the Volve data set. The inputs were P Impedance and Vp/Vs from a seismic inversion. LR provided the likelihoods required for Bayesian inference. The results were compared to a more traditional approach, estimating likelihoods with probability density functions (PDFs). The results of the two methods were similar except deeper in the reservoir where stationarity was an issue. Then, the addition of time as an additional input in the LR method was a clear advantage. However, the LR method was also sensitive to errors in the well logs used in its training phase.

Introduction & Method

Bayesian inference has been a popular approach to the determination of reservoir facies for some time (Pendrel et al. 2006, Pendrel and Schouten, 2020). It requires as inputs, likelihoods and prior probabilities corresponding to each member in the facies set. Traditionally, likelihoods have been determined with the aid of probability density functions (PDFs). Logistic Regression provides an alternate approach, obviating the need for PDFs. LR has been widely used in a variety of disciplines to model categorical variables. It is a machine learning method which requires training prior to the estimation of facies probabilities. The training data can come from well logs or models. When inversion outcomes are used to determine facies, the well logs used in training are usually a combination of P impedance, Vp/Vs and density or their derivatives. To complete the training data set, facies logs are included. They can be derived from either elastic or petrophysical properties and need not mirror the inversion output types. However, there needs to exist a relation between them and the elastic properties. This is an important part of any feasibility study. Logistic Regression creates labels for the facies: 1 for existence and 0 for non-existence. The final LR estimates range from 0 to 1 and correspond to facies probabilities.

Once a training set has been identified, facies probabilities are predicted by a linear weighting of the input properties. Each property is first scaled to zero mean and unit standard deviation to prevent it from disproportionally impacting the analysis. Squares and cross-products of the input properties are added to the inputs to facilitate modelling of non-linearities. Depth or time can also be modellable inputs. Rather than modelling the facies probabilities explicitly, the logit (log odds) function is modelled instead, since it spans the possible property space, both positive and negative. For the i-th facies with probability p_i, the logit function is

$$L_i = \ln\left(\frac{p_i}{1-p_i}\right) = \sum_{n=1}^{N} w_n v_n$$ where the v_n and w_n are the input variables and weights, respectively

The logit is further transformed by a sigmoid function to ensure that the estimate is in the range, 0 - 1. Then, the predicted facies probability becomes

$$\hat{p}_i = \frac{1}{1+e^{-L_i}}$$

For further details, see, for example, Elio et al., 2017.

Initial weights are chosen and adjusted iteratively in successive *epochs* to minimize the mean square error via a steepest descent algorithm. Training data is taken from a random subset of the log information; the remaining being saved for a validation phase. Several LR training parameters are available and include learning rate, sample data size and training percentage. The prediction error is monitored and when it becomes stable, the iteration is terminated. Finally, the weighted prediction is validated with the input data not used in training. After satisfactory training, the model is applied to the inputs from inversion across the project and facies probabilities determined.

Results

The method was applied to data from the Volve project, kindly provided by Equinor. The Volve field produced from the Hugin formation structural high from 2008 to 2016 with a cumulative production of 63 million barrels of oil (Weijermars, 2024). The Hugin is set in a shallow marine environment with transgressive sequences from 5 m to 126 m thick. Estuarian channels are present in the middle Hugin where the highest porosities (20%) and permeabilities (890 mD) occur. Variable porosities and permeabilities can lead to elastic non-stationarity in the Hugin.

The input data consisted of three partial-angle stacks: 3-18 deg., 18-32 deg. and 32-41 deg. The seismic band extends from approximately 10 Hz to 60 Hz. Logs were available from five wells although each required significant corrections, often involving shear sonic modeling. We regard the well log modelling and horizon interpretations as works in progress.

Facies were designed from gamma ray and effective porosity logs, high-cut filtered at the seismic band (Figure 1, left). They were:

- High Quality sandstone (HQ SS)
- Low Quality sandstone (LQ SS)
- Silty Sandstone
- Silty
- Shale 1
- Shale 2
- Shale 3

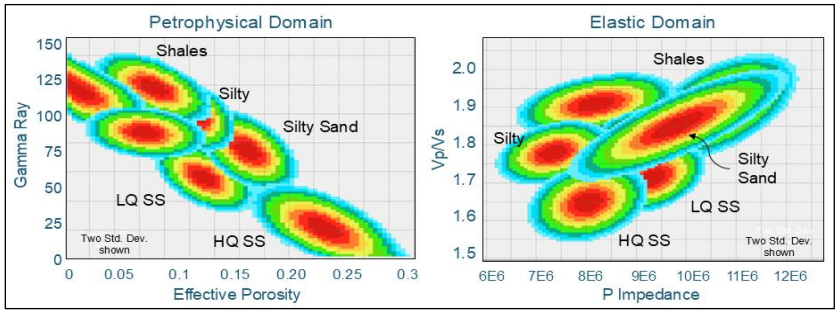


Figure 1: Representations of the facies set in the petrophysical (left) and elastic (right) domains. HQ sands will be resolvable in the elastic domain.

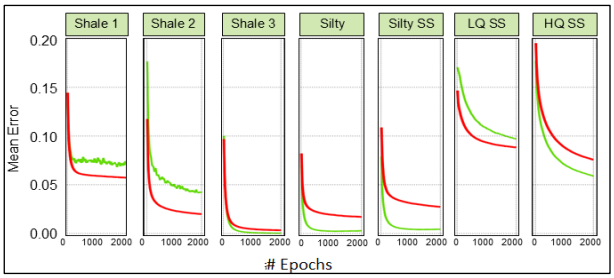


Figure 2: Training (red) and validation (green) loss curves. Prediction errors are 0.1 or less.

With this definition, facies logs were created using Bayesian inference. The facies were also viewed in the elastic domain (Figure 1, right). The HQ sand is resolvable, indicating that inversions should be effective in imaging it. The elastic domain PDFs can also be employed to derive facies likelihoods for Bayesian inference in a more traditional approach.

Four of the most-consistent log sets were used to complete LR training. Four geologic layers were recognized and training done individually for each. They were the Hugin, two reservoir bounding

layers and another shallower layer. The properties used in the training were P Impedance and Vp/Vs. The training set comprised 85% of the available data. It was chosen randomly and re-selected at each iteration. The learning rate was set low enough to ensure steady minimization of the prediction error while avoiding false minima. The learning was validated by applying the model to the remaining samples. It was found that 2000 iterations were sufficient to achieve convergence. This procedure was then repeated separately for all seven facies. The results for the Hugin layer are shown in Figure 2. They were satisfactory with mean predicted errors below 0.1.

A sparse spike inversion algorithm (Pendrel et al., 2000) was used to simultaneously invert the three partial stacks. The low frequency models for inversion were obtained by the PDF Transform method described in Pendrel, 2025.

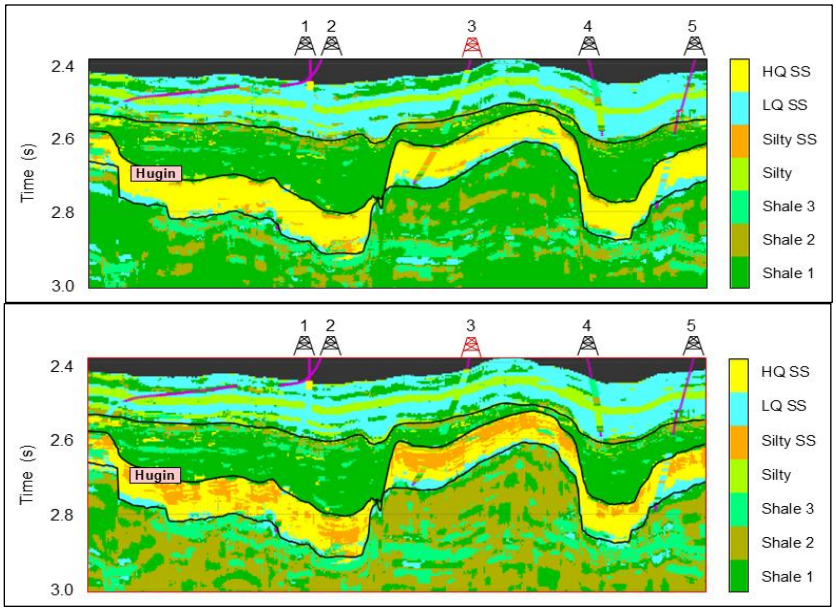


Figure 3: Most-probable facies from Bayesian inference with LR likelihoods. LR was trained with four wells (upper) and five wells (lower).

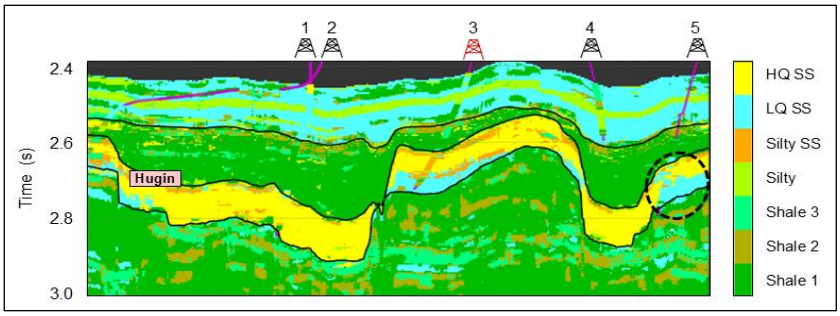


Figure 4: Most-probable facies with time added as an input. Wells 1,2,4 and 5 were used in training. Note the improved agreement at well 5 compared to Figure 3 (upper).

The most-probable facies from Bayesian inference using LR likelihoods are shown in Figure 3 (upper) along an arbitrary line through the wells. Uniform priors were used in this work. Focus is on the Hugin layer. Much of the information in the other layers was missing or not adequate. In the Hugin, agreement with the well facies was good at wells 1 and 2 and fair at well 5. However, it was poor at well 3. This well was less consistent with the others and not used in the training. Analysis showed that the likely cause was poor borehole competence in the Hugin. Nevertheless, its effects on training when it was included were instructive. Figure 3 (lower) shows the most-probable facies when all five wells were used to train the LR. Agreement at well 3 is now good but HQ sand has been replaced by Silty Sand across the project. Clearly, well log quality control prior to LR training is very important. One more experiment was conducted to address the only-fair match at well 5. To model any non-stationarities, ‘time’ was used as an additional input to the

LR training. The results for the most-probable facies with four-well training are shown in Figure 4. The match at well 5 is now much better. Stratigraphic slices (Figure 5) showed much more geologic HQ Sand distributions deeper in the Hugin compared to Bayesian inference using PDF likelihoods. The PDF approach would be improved by subdividing the Hugin into two layers but this has not been investigated here.

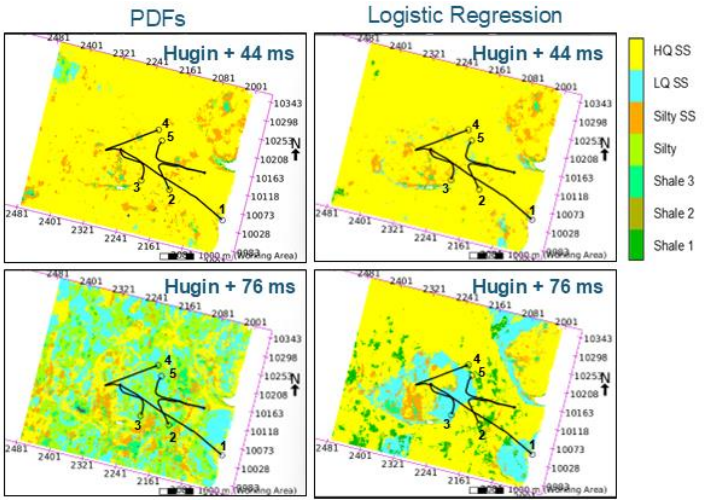


Figure 5: Most-probable facies stratigraphic slices in the middle and deeper in the Hugin for the PDF (left) and LR (right) methods. The LR and PDF results are similar in the middle but much different deeper where the LR method is better equipped to address non-stationarity.

Conclusions

Logistic Regression is a viable approach to estimating facies likelihoods for Bayesian inference. Care must be taken to quality control the training data if reliable results are to be obtained. Logistic Regression is also a good tool in non-stationary situations since time or depth can be included as an input property. In this respect, it has outperformed Bayesian inference with probabilistic inputs from PDFs.

Acknowledgements

Thanks to Equinor for kindly providing the data set. Special thanks to Benoit Luquet and Harry Debeye for their contributions to the LR and inversion algorithms respectively when they were with GeoSoftware. Thanks also to the Jason team at GeoSoftware for their support.

References

Elio, J., Crowley, Q., Scanlon, R., Hodgson, J., Long, S., 2017, Logistic regression model for detecting radon prone areas in Ireland, Science of the Total Environment, v 599-600, p.1317-1329 https://doi.org/10.1016/j.scitotenv.2017.05.071

Pendrel, J., Debeye, H., Pedersen-Tatalovic, R., Goodway, B., Dufour, J., Bogaards, M., and Stewart, R., (2000), "Estimation and interpretation of P and S impedance volumes from simultaneous inversion of P-wave offset seismic data," SEG Technical Program Expanded Abstracts: 146-149. https://doi.org/10.1190/1.1815683

Pendrel, J., Mangat, C., Feroci, M., 2006, Using Bayesian inference to compute facies-fluids probabilities: CSEG GeoConvention Abstracts

Pendrel, J., Schouten, H., 2020, Facies – the drivers for modern inversions, The Leading Edge, 39, 102-109 https://doi.org/10.1190/tle39020102.1

Pendrel, J., 2025, Improving seismic inversions with PDF Transforms. SEG/AAPG Int. Mtg. for Applied Geoscience & Energy, Abstracts

Weijermars R., Fast production and water-breakthrough analysis methods demonstrated using Volve Field data, Petroleum Research 2024, v9 #3, p327-346 https://doi.org/10.1016/j.ptlrs.2024.03.001