3D lithofluid facies classification by machine learning algorithms

Lucas Gabriel Silva de Aguiar[1], Jadson Muniz de Oliveira[1], Maurício Laurindo de Matos[1], Rodrigo da Silva Canário[1] and Fernando Sérgio de Moraes[1], [1] Grupo de Inferência de Reservatório - GIR/LENEP

This paper was prepared for presentation during the 17th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 August 2021.

Abstract

There must be an adequate integration of well and seismic data to create a reliable facies model. However, heterogeneous rocks, lack of data and inaccurate information accumulate uncertainties for model characterization. Here we seek to simulate a scenario where there are few wells and only one seismic data to create the facies model. The main goal is to estimate the predominant lithology distribution with fluid information. This strategy can help in new wells' locations and reducing the risk of false discoveries. The Norne field located on the Norwegian Sea will be used as an example for this methodology. Machine learning techniques have been commonly used for automation and task optimization processes. We show that these methods can achieve high performance when receiving proper processed data. The summarized workflow can be described in two stages: electrofacies classification on eight different wells and use of these facies as labels for classification on the field's seismic volume. The results indicate accuracy close to 90% both in well and seismic scenarios. The facies model with lithofluid information is still compared with the reservoir simulator from the same period to show the high similarity.

Introduction

Seismic facies can be considered as classes that contain similar characteristics of seismic reflectors, varied attributes and spatial distribution (Conticini, 1984; John et al., 2008). The estimation of these facies can be difficult due to the low quality of seismic resolution, the lack of important seismic data for processing steps or even the geological complexity of the reservoir. For this reason, the aid of petrophysical data can help the creation of a more realistic facies model (Saussus and Sams, 2012).

Some studies use the aid of seismic attributes (e.g. P wave impedance, S wave impedance and density) from inversion to generate the facies model in the reservoir through classification (Roncarolo & Grana, 2010; Grana, 2016; Tellez et al., 2017). Since elastic inversions provide only information about the seismic bandwidth, some studies avoid this problem by combining the inversion with the low frequency model, in addition to petrophysical information, for the construction of the facies model (Sams & Saussus, 2013; Zabihi Naeini & Exley, 2017).

Our study addresses the estimation of the facies model with lithofluid distribution throughout the reservoir, with the aid of well data, low frequency seismic models and machine learning classification algorithms. The methodology developed in this work will be applied to the Norne field, from offshore Norway, between 2001 and 2006, when the field reached its oil production peak and water production increased considerably, making it crucial to identify unswept oil zones and oil-water contact (OWC).

The Norne field had many works published in the last decade regarding reservoir characterization (Rwechungura et al., 2010; Chen & Oliver, 2014; Møyner et al., 2015; Correia & Schiozer, 2016). One of the main problems related to reservoir characterization in the Norne Field is the discrimination of lithofacies since the reservoir presents a complex geological setting that goes from shallow-marine and deltaic sands from the Fangst Group (Middle Jurassic) to heterolithic tidal units of the Tilje Formation from the Båt Group (Early Jurassic) (Dalland et al., 1988; Hammer et al., 2010).

The reservoir quality in the Norne field can be affected by thin clay layers interbedded with the sandstones of interest. The correct identification of these facies would help to estimate zones of better reservoir quality. Like these clay layers, there are dolomitized limestone strata between the geological formations acting as stratigraphic barriers, restricting the vertical flow of fluids (Verlo & Hetland, 2008). The recognition of these seals is essential to determine better strategies for well location and defining perforation zones (injection and production).

More specifically, we propose using machine learning to provide accurate results to estimate the fluids contact and evaluate the potential distribution of relevant facies such as shales and limestones that might compromise the flow in the reservoir, all in 3D seismic scale level of detail. These results allow the engineers to strategically optimize the wells' locations to increase production while reducing the risk associated with false hydrocarbon discoveries and infill drilling. The idea is to emulate a situation with minimal dataset available such as a 3D seismic volume and a few wells, which might be the case for small companies and researchers.

Study Area

Norne field is situated in a horst block on the Norwegian Continental Shelf. It has dimensions of 9 x 3 km and its thickness ranges between 120m and 260 m. Two hydrocarbon compartments divide the field: Main Structure (segments C, D and E) and Northeast Structure (segment G) (Figure 1a).

The field has jurassic sub-arkosic sandstone reservoir that can be divided into four distinct formations: Not 2, Ile, Tofte and Tilje (Figure 1b). The gas is predominantly in the Not 2 Formation and 80 % of the oil is found in the Ile and Tofte formations (Rwechungura et al. 2010). Along the field, different stratigraphic barriers act as seals. The main one is the Not 1 Formation, a clay layer of approximately 10 m thickness that prevents communication between Not 2 and Ile. Dolomitized carbonate layers about 3 m thick are also identified and prevent the vertical flow of fluids between different units (Verlo & Hetland, 2008).

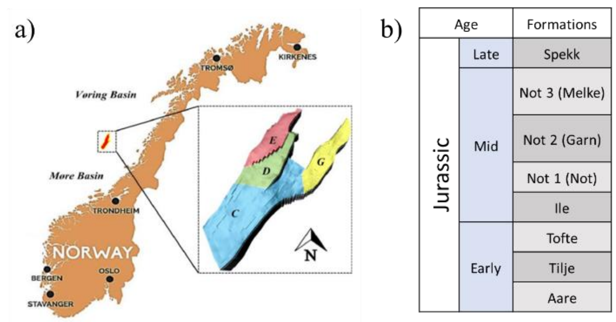

Figure 1 *– a) Norne field location and spatial representation of its segments (Maleki et al., 2018). b) Summarized stratigraphy.*

Method

The workflow integrates well information and post-stack seismic data to solve the problem of facies inference in the seismic domain, which results in a lithofluid facies model (Figure 2).

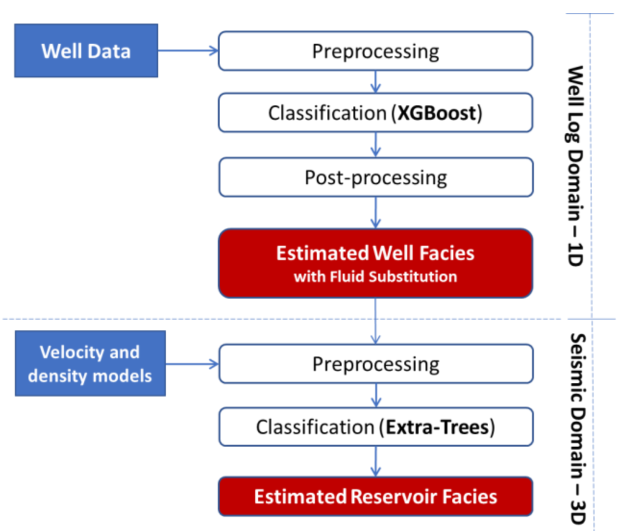

Figure 2 *- Workflow used to classify lithofluid facies present in the reservoir.*

Well data analysis and processing

Among 49 available wells, only eight were used due to the necessary correlation of their properties with seismic parameters (P and S wave velocities and density) (Figure 3). Only four of these eight wells have reliable electrofacies information from the composite logs present in the reports. Wells B-4H and D-4H compose the training data, where 15% represent the test set. The wells C-1H and E-3H are considered as the validation data to measure the algorithm performance. After a satisfying accuracy is achieved, the classification expands to the remaining four wells and the fluid contacts from the reports are compared to the predicted electrofacies. The facies used are Gas Sand, Oil Sand, Brine Sand, Shale (including Claystone and Siltstone) and Limestone.

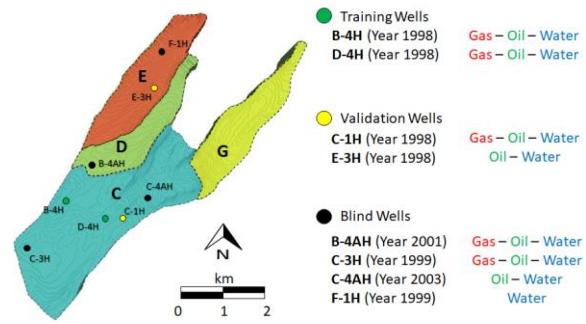

Figure 3 *- Well position along the field with their respective fluids and drilling year.*

From the available logs, the following were chosen to be used during the classification: bulk density (RHOB), neutron porosity (NPHI), water saturation (SW), permeability (KLOGH) and shale volume (VSH). Two more logs, created from S-wave and P-wave velocities, were also used and will be called relative Vp and relative Vs. These new logs indicate the variation of these two properties from a zero-mean and, thus, would better represent the real geological variations.

It is likely that the sampling of the wireline logs, 12.5 cm, does not correctly identify transition regions from one lithology to another. Then, two samples between lithological limits are removed from the training wells to decrease this uncertainty in the model.

After prediction, if all samples have a unique type of facies (e.g. Shale) and there is only one different between them (e.g. Limestone), it could represent an error in the classification. It might be acceptable to believe that this is a punctual outlier and that the correct prediction would be of the dominant facies (Bestagini et al., 2017). Therefore, it was decided to use the prevailing type of facies around to replace this outlier, when there is just one sample of it.

Extreme Gradient Boosting (XGBoost)

This study performs electrofacies classification by the XGBoost algorithm. It uses ensembles of decision trees (boosted trees) that create weak models and each model tries to correct the errors present in the previous one (Dietterich, 1999). This process continues until the prediction is made correctly or a maximum number of models are added. Developed by Chen & Guestrin (2016), XGBoost has stood out in researches that address classification problems. (Torlay et al., 2017; Zhang et al., 2018).

XGBoost performs gradient boosting and it is essential to avoid data overfitting. The new model is updated with the latest prediction, minimizing the objective function using gradient descent (Friedman, 2001; Chen & Guestrin, 2016).

Fluid Substitution

As production progresses, the pattern of fluid distribution in the reservoir is modified and the properties originally captured by the well logs may not exactly represent the current characteristics of the reservoir. In order for the information extracted from the wells to be used, Gassmann's (1951) fluid substitution routine must be applied so that the logs can be updated and effectively represent the reservoir properties for the desired time. Information on the distribution of saturation and pressure in the wells for the time of interest is obtained with the help of a reservoir simulator software.

This study performs fluid substitution only in the sandstone samples since shale and carbonate are seals and their fluid content does not change. So before making the classification in the seismic data, the RHOB, SW, Vp and Vs logs of each well were updated to the time of interest, established in this work as of August 2006. Then, the same electrofacies classification model was used, with the two training wells (B-4H and D-4H) in their original configurations, to estimate the electrofacies in the eight updated wells for the new fluid distribution scenario.

Facies classification in the seismic domain

To implement the facies classification on the reservoir volume, velocity and density models are needed. In this study, the AVO inversion did not show reliable results, so it was preferred to build these models using a well log interpolator that can follow the interpreted seismic geometry. This approach uses a seismic interpretation software to combine well log data (Vp, Vs and density) and a horizon cube, which generates automatic interpretation of seismic horizons. The process results in low frequency velocity and density models.

These new properties (Vp, Vs and density) become the features of the classification model, in the position of the well tracks in the seismic cube. And the electrofacies from the classification of the wells after fluid substitution are considered as the label. Still, 30% of this data is separated as a test set to evaluate the model accuracy. The classification uses the Extra-Trees method and is applied over low frequency Vp, Vs and density of the whole seismic cube.

Extremely Randomized Trees (Extra-Trees)

The Extra-Trees method also uses ensembles of decision trees as the XGBoost algorithm described before. Its main differences from other ensembles methods are the addition of total randomness to define the features and its respective cut-points in node splitting, as it uses all the samples during the training to build the trees (Geurts et al., 2006). These conditions help to reduce the variance of the model, in exchange for an increase of bias (Géron, 2017). Therefore, the randomness of trees tends to increase the accuracy, since the error in the prediction that each tree provides tends to be uncorrelated to each other (Breiman, 2001).

Performance metrics

- Accuracy: prediction evaluation criterion where precision is estimated by the number of samples correctly classified divided by the total number of samples for each well.

- Confusion matrix: also used to evaluate the performance of the electrofacies classification. In a matrix C, the notation $\mathbf{C_{i,j}}$ represents the number of real class **i** and predicted class **j** observations. In the case of binary classification, for example, the count of true negatives is $\mathbf{C_{0,0}}$, false negatives $\mathbf{C_{1,0}}$, true positives $\mathbf{C_{1,1}}$ and false positives $\mathbf{C_{0,1}}$ (Figure 4) (Pedregosa et al., 2011).

		Predicted Class	
		Positive	Negative
True Class	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Figure 4 *- Demonstration of how a confusion matrix works. Blue colors represent correct predictions and red ones represent errors.*

Results

First, the electrofacies classification on the test data (15% of the samples from the training wells) achieved an accuracy of 87%. After that, the model reproduced the classification on the two validation wells (C-1H and E-3H) and the result was compared with their true facies (Figure 5). The XGBoost algorithm showed similar performance for both wells, achieving an accuracy of 87% and 89% in C-1H and E-3H, respectively. Similar precision values in the test and validation data indicate a good generalization of model prediction, which is ideal to avoid overfitting.

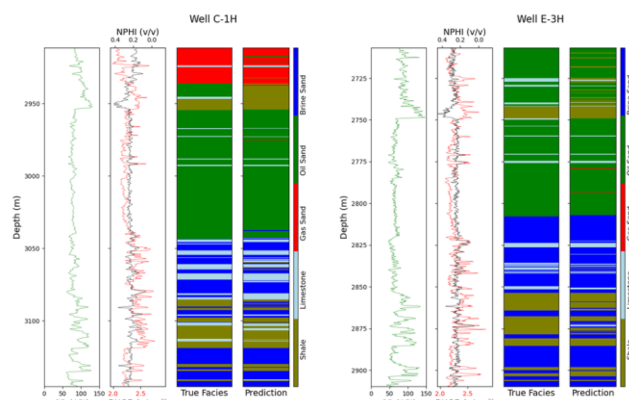

Figure 5 *- Electrofacies classification on the two validation wells, C-1H and E-3H.*

In general, the process was quite robust for most of the facies, although the model showed some difficulty in correctly predicting Limestone (Figure 6). Similar values of RHOB, KLOGH, and SW can explain the tendency to confuse these facies with Shale. On the other hand, there is high accuracy for Sandstone variations, which is relevant since it is considered as the reservoir and zone of greatest interest.

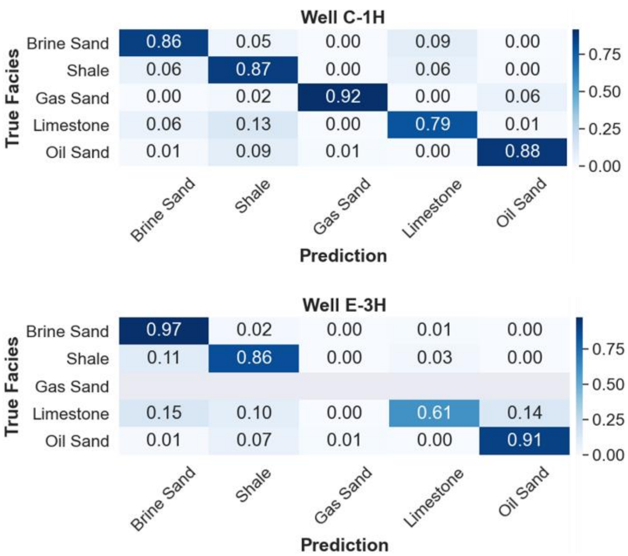

Figure 6 *- Normalized confusion matrices of wells C-1H and E-3H, respectively.*

After concluding that the testing and validation steps were satisfactory, the classification was made on the blind wells (B-4AH, C-3H, C-4AH and F-1H), which have no lithology information. The results were then compared to the fluid contacts obtained in the reports as a performance metric (Figure 7). It is noticed that the position of these contacts agrees with the inference of facies with fluid information.

It is also worth noting that the facies estimated for wells C-4AH, E-3H and F-1H, which do not have all three types of fluid, maintained the predominance of this fluid information even though the model was trained with gas, oil and water sands. For example, the well report shows that F-1H has only water in its rock pores; therefore, the fact that there is no predicted sand with gas or oil in this well proves that the classification has preserved these relevant characteristics.

Eight wells underwent fluid substitution to conform these to the seismic period, 2006, after the quality and robustness of the model had been proven. The classification on the seismic volume also shows high accuracy on the test data (30% of the original data), reaching about 94% of correctness. There is a high similarity in the distribution of fluids when compared with the reservoir model of the same year from the simulator (Figure 8).

There are overestimated gas and oil when looking at the lithofluid estimation in the G segment. This answer is not surprising, as the hydrocarbon in the G segment has a higher viscosity and density than the rest of the reservoir due to the absence of communication between these compartments (Verlo & Hetland, 2008; Maheshwari, 2011). As none of the wells used in the model is part of the G segment, there is high uncertainty in that region due to the lack of petrophysical information. Thus, the properties captured by the wells in the rest of the reservoir may not represent the fluid properties of that segment.

Another relevant aspect is the identification of the Not 1 clay layer over a large part of the reservoir, between the units with gas and oil. When consulting the literature, it is known that this layer has high importance for the field, as it acts as a stratigraphic barrier to the vertical flow of fluid between formations. It is also worth noting the recognition of facies heterogeneity in the lower portion of the reservoir, represented by the well-known Tilje heterolytic formation (Figure 9). We note that the classification on the low-frequency model has some similarity with the model resulting from the inversion. However, the latter presents more noise and highlights less the continuity of facies along the field, which may add imprecision to a future interpretation.

Conclusions

The facies classification for the Norne field proved to be quite relevant due to the automation of the facies interpretation process and the estimation of lithofluid contacts. We also showed that stratigraphic barriers can be identified, although thinner ones tend to be more challenging due to the lower resolution of the 3D model in seismic.

The accuracy close to 90%, both in well and seismic scenario, proved that these tools can be quite powerful when aligned with the proper processing of your data and the choice of powerful algorithms. The similarity between the final reservoir classification and the reservoir simulator in the same year shows that this process can be of great help. This application in a larger data set with more relevant information available, as pre-stack seismic and more wells, could assist the interpreter's work even more.

Acknowledgments

The authors would like to thank Petrobras for funding the R&D project registered as “Desenvolvimento de fluxos de trabalho e metodologias para processamento e interpretação quantitativa de dados sísmicos 3D e 4D” and Equinor for providing the Norne dataset used during this study.

We would like to thank dGB Earth Sciences for the open-source seismic interpretation software OpendTect, and also the Open Porous Media initiative contributors for making their software available for reservoir simulator visualization.

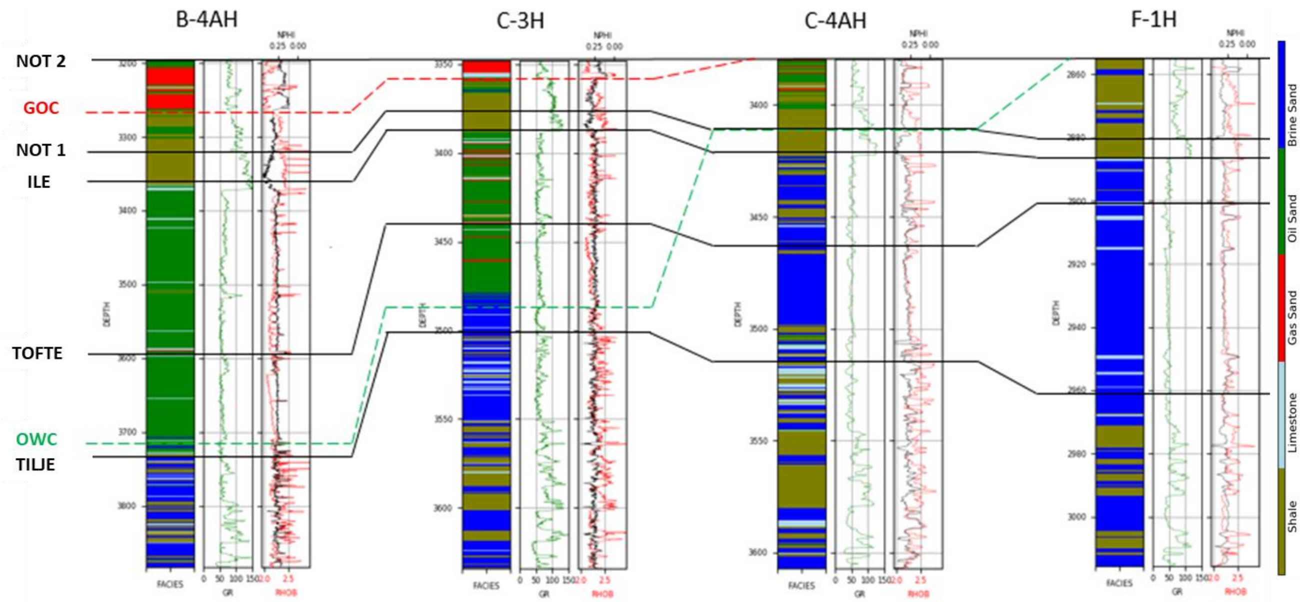

***Figure 7** - Electrofacies classification on the four blind wells, that had no previous label information. The solid lines represent stratigraphic tops and the dashed lines show the original lithofluid contacts depths from the reports.*

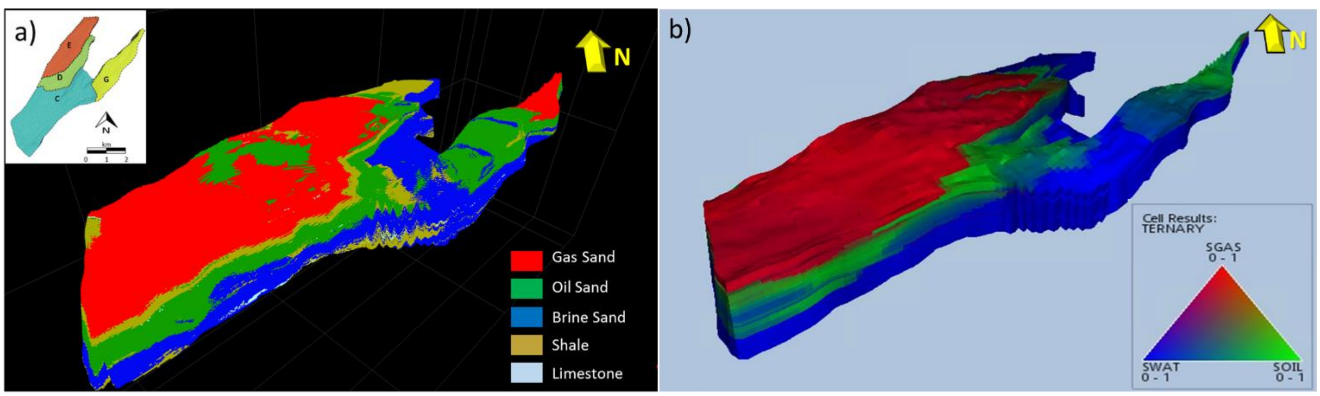

***Figure 8 -** a) Low-frequency 3D model with lithofluid facies classification for Norne field in 2006. b) Fluids representation according to their respective saturations from the simulator in 2006.*

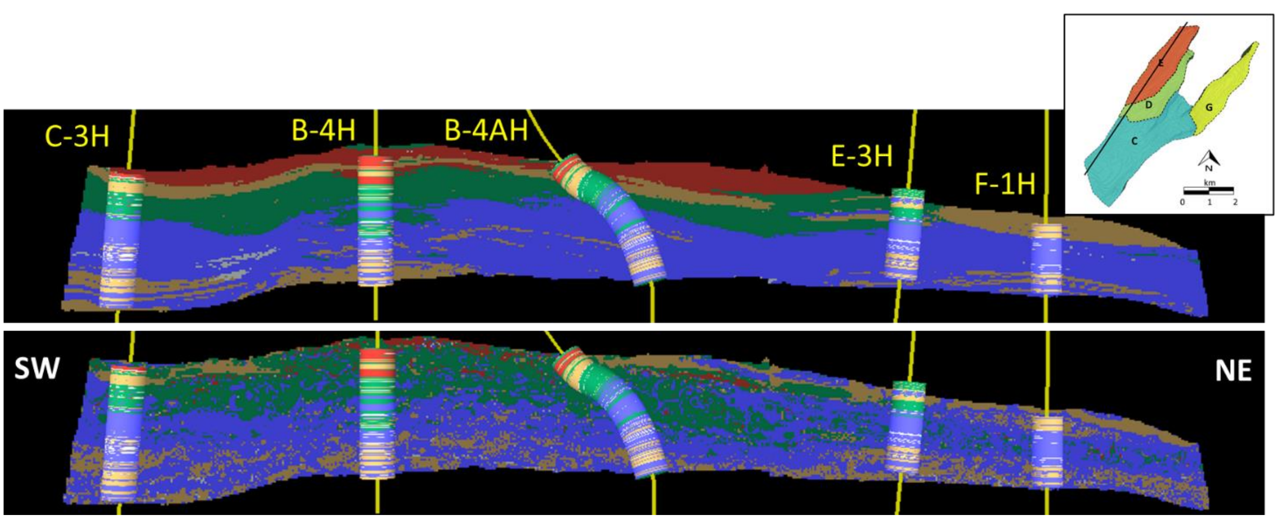

***Figure 9** - Section along the reservoir with facies classification on low-frequency model (top) and inversion model (bottom).*

References

Bestagini, P., Lipari, V., & Tubaro, S. (2017). A machine learning approach to facies classification using well logs. SEG Technical Program Expanded Abstracts 2017.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Chen, Y., & Oliver, D. S. (2014). History matching of the Norne full-field model with an iterative ensemble smoother. SPE Reservoir Evaluation & Engineering, 17(02), 244-256.

Chen, T., & Guestrin, C. (2016). XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16.

Conticini, F. (1984). Seismic facies quantitative analysis: New tool in stratigraphic interpretation. SEG Technical Program Expanded Abstracts 1984.

Correia, G. G., & Schiozer, D. J. (2016). Reservoir characterization using electrofacies analysis in the sandstone reservoir of the Norne Field (offshore Norway). Petroleum Geoscience, 22 (2), 165-176.

Dalland, A., Worsley, D., & Ofstad, K. (1988). A lithostratigraphic scheme for the Mesozoic and Cenozoic succession offshore mid and northern Norway. Norwegian Petroleum Directorate Bulletin. 4, 1–65.

Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Machine Learning, 40(2), 139–157.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Gassmann, F. (1951). Uber die elastizitat poroser medien. Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich, 96, 1-23.

Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

Grana, D., Lang, X., & Wu, W. (2016). Statistical facies classification from multiple seismic attributes: comparison between Bayesian classification and expectation-maximization method and application in petrophysical inversion. Geophysical Prospecting, 65(2), 544–562.

Hammer, E., Mørk, M.B.E. & Næss, A. (2010). Facies controls on the distribution of diagenesis and compaction in fluvial-deltaic deposits. Marine and Petroleum Geology, 27, 1737–1751.

John, A. K., Lake, L. W., Torres-Verdin, C., & Srinivasan, S. (2008). Seismic Facies Identification and Classification Using Simple Statistics. SPE Reservoir Evaluation & Engineering, 11(06), 984–990.

Maheshwari, Y. K. (2011). A Comparative Simulation Study of Chemical EOR Methodologies (Alkaline, Surfactant and/or Polymer) Applied to Norne Field E-Segment. Master Thesis. Department of Petroleum Engineering and Applied Geophysics. Norwegian University of Science and Technology, Norway, 2011.

Maleki, M., Davolio, A., & Schiozer, D. J. (2018). Quantitative integration of 3D and 4D seismic impedance into reservoir simulation model updating in the Norne Field. Geophysical Prospecting.

Møyner, O., Krogstad, S., & Lie, K. A. (2015). The application of flow diagnostics for reservoir management. SPE Journal, 20(02), 306-323.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830.

Roncarolo, F., & Grana, D. (2010). Improved Reservoir Characterization Integrating Seismic Inversion, Rock Physics Model, and Petroelastic Log Facies Classification: A Real Case Application. SPE Annual Technical Conference and Exhibition.

Rwechungura, R., Suwartadi, E., Dadashpour, M., Kleppe, J., & Foss, B. (2010). The Norne Field Case - A Unique Comparative Case Study. Proceedings of SPE Intelligent Energy Conference and Exhibition. Utrecht, Netherlands, 23–25.

Sams, M., & Saussus, D. (2013). Practical implications of low frequency model selection on quantitative interpretation results: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3118–3122.

Saussus, D., & Sams, M. (2012). Facies as the key to using seismic inversion for modelling reservoir properties: First Break, 30, July, p45-52.

Statoil. (2001).PL128-Norne Field Reservoir Management Plan.

Tellez, J., & Slatt., R. (2016). Seismic geomorphology and characterization of deep water architectural elements and its applications in 3D modeling. A case of study, North Carnarvon Basin Australia. AAPG International Conference & Exhibition – 2016. Cancun, México.

Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciu, M., (2017). Machine learning - XGBoost analysis of language networks to classify patients with epilepsy. Brain Inform. 4 (3), 159–169.

Verlo, S. B., & Hetland, M. (2008). Development of a field case with real production and 4D data from the Norne Field as a benchmark case for future reservoir simulation model testing. Master's degree, NTNU.

Zabihi Naeini, E., & Exley, R. (2017). Quantitative interpretation using facies-based seismic inversion. Interpretation, 5(3), SL1–SL8.

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., Si, Y., (2018). A data-driven design for fault detection of wind turbines using random forests and XGboost. IEEE Access 6, 21020–21031.