Introduction

MOL Norge AS, a subsidiary of the Hungarian state oil company MOL, owns a number of operated licenses on the Norwegian Continental Shelf. Explorationists in MOL Norge AS have recognized significant potential in the Upper Jurassic interval on these licenses and among others seeking for quantitative seismic methods to de-risk the identified prospects. However, to provide reliable results using traditional seismic inversion methods (like simultaneous pre-stack inversion) is sometimes more than challenging for the Jurassic sequences in this area.

To mitigate data related uncertainties (e.g. lack of usable wells within the area of interest with very limited elastic understanding of the expected sand properties) as well as to overcome the known limitations of the traditional, industry standard inversion methods, MOL Norge AS has been working together with dGB Earth Sciences. This involved applying a direct inversion technique called HitCube 'trace-matching' inversion, which utilizes Monte Carlo simulated pseudo-wells (Ayeni et al., 2007) to overcome lack of useful well data. In the first part of this cooperation, HitCube 'trace-matching' inversion provided significant improvement compared to simultaneous inversion results and the outputs acted as essential arguments for making a drill decision on one of the key licenses in MOL Norge AS` portfolio.

Additionally, it has been recognized that the nature of this inversion technique based on pseudo-well simulations, makes it a very good candidate to be upgraded to a machine learning supported approach. In the second part of the cooperation, numerous machine learning algorithms have been tested in order to find the optimal ones to predict elastic and reservoir properties from synthetic pre-stack seismic computed from the pseudo-wells. The ultimate goal is to apply the trained machine learning algorithms on the real pre-stack seismic data and come up with an alternative pseudo-well based machine learning inversion for reservoir characterization.

Theory and Background Work: HitCube 'trace-matching' Inversion

HitCube 'trace-matching' inversion requires a very detailed and comprehensive rock physics understanding of the expected facies in the reservoir and in the overburden. As such, we need to understand the elastic properties of the reservoir as a function of rock properties. In case of an expected shale-sand dominant Jurassic sequence we need to find out how velocity of the sand relates to porosity; how the velocity of sand and shale changes with depth etc… (Figure 1 – Step 1).

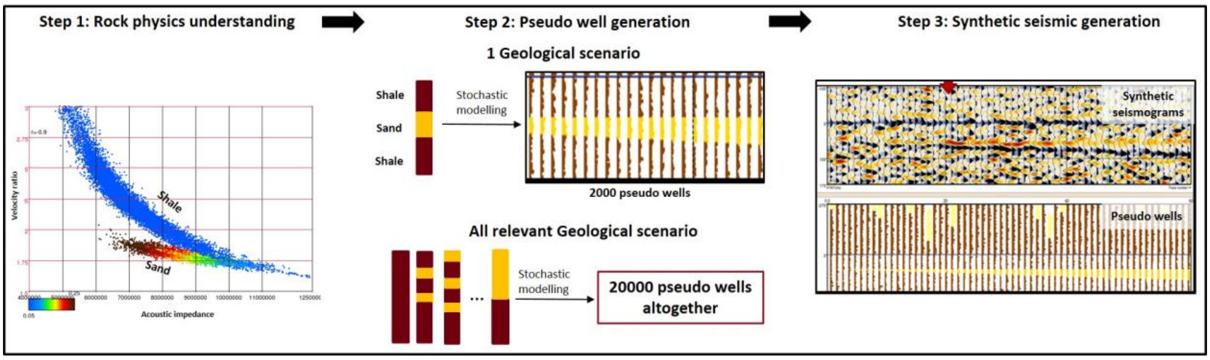

***Figure 1** Rock physics background of HitCube 'trace-matching' inversion.*

The second step is the real special part of this inversion technique called Monte Carlo pseudo-well generation. In this step, we create a possible geological scenario for the reservoir and fill the sands and shales with the most likely elastic parameters. After this we perform stochastic modelling by perturbing the sand/shale parameters using the observed rock physics trends – this will give us 2000 pseudo-wells for this particular geological scenario. Similarly, we simulate all other possible and relevant geological scenarios by again using the rock physics trends in the stochastic modelling, and generate more pseudo-wells. For this project, we generated a total of 20000 pseudo-wells that describe all possible geological scenarios (Figure 1 – Step 2). The elastic parameters are embedded in the pseudo-wells, so we can create synthetic seismogram for each pseudo-well. As we modelled 20000 pseudo-wells from various geological scenarios, therefore, we have 20000 synthetic seismograms (Figure 1 – Step 3). The inversion process is to simply correlate each synthetic seismogram with each and every real seismic trace – 20000 * 800000 correlations altogether in this case. The best correlations are accepted and the associated pseudo-wells are selected at every trace location. Consequently, all the elastic and reservoir parameters that were included in the stochastic pseudo-well modelling can be direct inversion outputs.

Results: HitCube 'trace-matching' Inversion

The main purpose of running this inversion was to delineate the Upper Jurassic reservoir identified by our geological understanding and also to try to evaluate the expected quality of the reservoir. For this purpose, we chose porosity as primary reservoir quality indicator and the most useful inversion output (AI and Vp/Vs volumes were also generated).

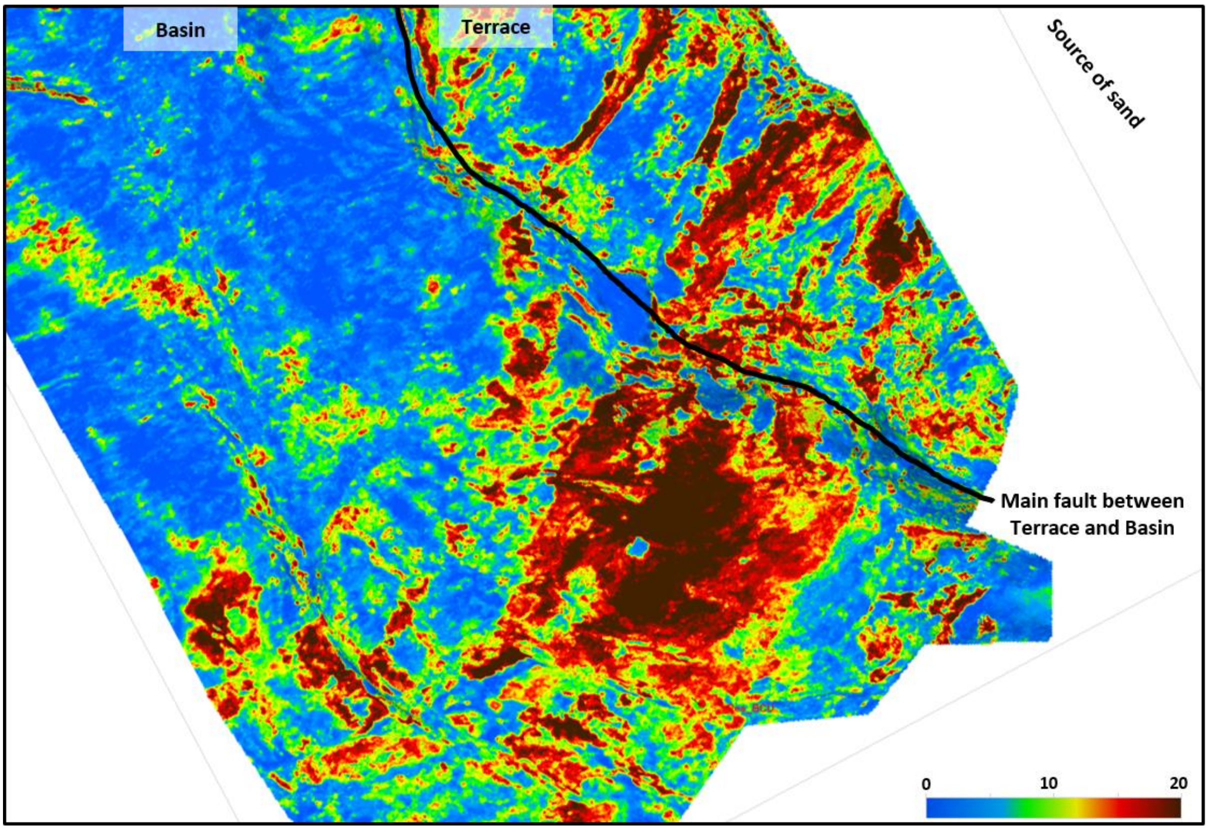

Figure 2 *Inverted porosity map – average values from the reservoir interval.*

The porosity map (Figure 2) suggests the possibility of moderate-to-good quality sand is coming from the source direction transported by channels through the terrace and also deposited in the basin along the major fault. The results of this unique inversion technique seem to support the geological model of the Upper Jurassic reservoirs and significantly contributed to de-risk prospectivity. Lack of reliable sand analogy is still considered as the main risk of this inversion workflow. It has also been recognized that this inversion technique with 20000 Monte Carlo simulated pseudo-wells as the core of the process is a very good candidate to be upgraded to a machine learning approach.

Theory and Tested Algorithms: Machine Learning Inversion

In order to predict elastic and reservoir properties directly from available pre-stack seismic, in addition to HitCube 'trace-matching' inversion, a supervised machine learning inversion approach is utilized. Supervised machine learning is essentially a process of teaching a mathematical model to identify relationships between a set of input and output data, by feeding it input data as well as correctly labelled output data. This is done in the so-called ‘training’ step. Once the machine learning model is trained, it can be applied on a new set of input data to predict a set of output data. In this study, 75% of the available set of pseudo-wells (i.e. 15000 pseudo-wells) is used for training various machine learning algorithms to recognize relationships between synthetic pre-stack seismic and pseudo-well logs (e.g. porosity, AI and Vp/Vs ratio). The remaining 25% (i.e. 5000 pseudo-wells) essentially acts as blind ‘test’ set – never seen by machine learning algorithms – to evaluate the quality of the prediction. The prediction quality is measured by computing the coefficient of determination, R2 (R-squared), between predicted and available pseudo-well logs.

1. **Linear Regression**
2. **Ridge and Lasso Regression**
3. **K-neighbor Regression**
4. **Gaussian Process Regression**
5. **Naïve Bayes Regression**
6. **Support Vector Regression**
7. **Decision Tree Regression**

Weak regressors

8. **Random Forest Regression**
9. **Deep Neural Network (DNN)**
10. **Gradient Boosting Regression on Decision Trees**
11. **AdaBoost Regression on Decision Trees**
12. **XGBoost Regression on: Decision Trees, Linear Model and Random Forests**

These involve boosting, i.e. sequentially fit one model after another to improve the prediction.

Figure 3 *Various algorithms tested for Machine Learning based inversion.*

A number of available supervised machine learning algorithms in the Python’s open source scikit-learn library are utilized for this purpose (Figure 3). During the training process, none of the twelve algorithms tested here were able to reliably predict porosity directly from synthetic pre-stack gathers. This was confirmed by low R2 values (< 0.4) between predicted and available porosity logs of the test pseudo-well set. Afterwards, all these machine learning algorithms were instead trained to directly predict elastic properties (AI and Vp/Vs ratio) from synthetic pre-stack gathers. The first seven algorithms tested were still found to be weak regressors (R2 values of prediction were around 0.5 to 0.6). However, the rest of the algorithms were indeed able to learn and reliably predict AI and Vp/Vs ratio from synthetic pre-stack gathers. The last three algorithms of Figure 3 involve boosting (i.e. sequentially fitting one model after another to improve the prediction), and thus can be very resource heavy. Therefore, Deep Neural Network (DNN) and Random Forest which was first introduced by Breiman (2001) are considered to be the optimum algorithms.

Results: Machine Learning Inversion

Between DNN and Random Forest, the Random Forest is chosen as the main machine learning algorithm in this study for direct elastic property prediction from pre-stack seismic data. This is because the algorithm is easier to understand with better result interpretability. Furthermore, it is also relatively easy to tune the algorithm’s hyperparameters; mainly the number of trees, which should be set high (Probst et al., 2019). Figure 4 shows the validation cross-plots between Random Forest predicted (Y-axis) and available AI and Vp/Vs ratio logs of the 5000 blind test pseudo-wells (X-axis), within the upper Jurassic reservoir interval. The R2 values between predicted and available AI and Vp/Vs logs are fairly high, 0.8 and 0.7, respectively. Finally, this trained Random Forest model is applied on the available pre-stack seismic data to predict AI and Vp/Vs ratio within the Upper Jurassic reservoir. The resulting volumes do match nicely with the log data of an available well inside the seismic survey. Also, they are in agreement with the corresponding HitCube ‘trace-matching’ inversion products in some parts of the survey. However, there are still discrepancies between them and only by acquiring more well log data, better estimation and comparison of the predictive power of the two methods can be carried out.

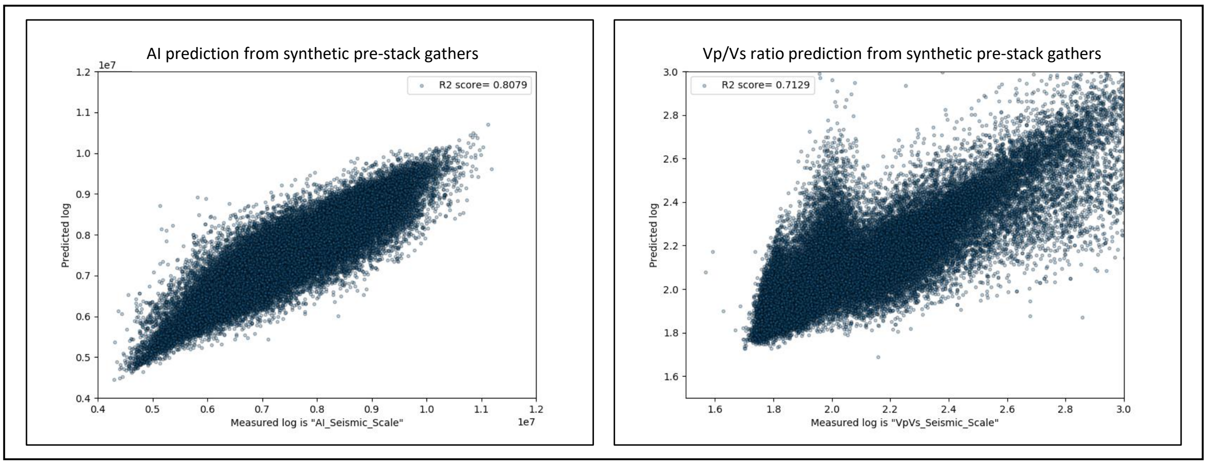

Figure 4 *Cross-plot of AI and Vp/Vs ratio logs of blind test pseudo-wells with the ones predicted by Random Forest from synthetic pre-stack gathers.*

Conclusions

MOL Norge AS and dGB Earth Sciences composed a detailed workflow to try and understand Upper Jurassic prospectivity on the Norwegian Continental Shelf. HitCube 'trace-matching' inversion – with a very comprehensive, stochastic rock physics modelling – resulted in geologically reasonable outputs in an area where the lack of analogous wells poses the main challenge. The inversion results significantly contributed to the delineation of the expected shale/sand system in the Upper Jurassic sequences.

Random Forest was found to be an optimum machine learning algorithm, which was able to accurately train on and thereafter predict elastic properties of simulated pseudo-wells from synthetic pre-stack seismic. The application of the trained Random Forest model on the available pre-stack seismic data resulted in volumes of AI and Vp/Vs, which can be used as 'second opinion' along with HitCube products to make more informed drilling decisions.

Acknowledgments

We thank MOL Norge AS for allowing to publish the results of this work.

References

Ayeni, G.O., Huck, A. and de Groot, P. [2007] The HIT Cube; matching Monte Carlo simulated pseudo-wells to seismic data for predictions with uncertainties. 69th EAGE Conference & Exhibition, London.

Breiman, L. (2001) Random Forests. Machine Learning, 45, 5-32, http://dx.doi.org/10.1023/A:1010933404324

Probst, P., Boulesteix, A.L. and Wright, M. (2019) Hyperparameters and Tuning Strategies for Random Forest. WIREs Data Mining And Knowledge Discovery, 9 (3), https://doi.org/10.1002/widm.1301