New direction for regional reservoir quality prediction using machine learning - Example from the Stø Formation, SW Barents Sea, Norway

H.N. Hansen [*], B.G. Haile, R. Müller, J. Jahren

University of Oslo, Department of Geosciences, Sem Sælands Vei 1, 0371, Oslo, Norway

A B S T R A C T

Recently, the petroleum industry has focused on deeply buried reservoir discoveries and exploring potential CO_2 storage sites close to existing infrastructure to increase the life span of already operating installations to save time and cost. It is therefore essential for the petroleum industry to find an innovative approach that exploits the existing core- and well log data to be successful in their endeavor of effectively characterizing and predicting reservoir quality. Continuous data sources (e.g. wireline logs) have a huge potential compared with expensive, time inefficient and sporadic data from cores in determining reservoir quality for use in a regional context. However, whereas core analysis offers in-depth knowledge about rock properties and diagenetic processes, continuous data sources can be difficult to interpret without a formation-specific framework. Here, we demonstrated how the pre-existing core data could be effectively used by integrating petrographic- and facies data with a pure predictive machine learning (ML) based porosity predictor. The inclusion of detailed core analysis is important for determining which reservoir parameter(s) that should be modeled and for the inter-pretation of model outputs. By applying this methodology, a framework for deducing lithological and diagenetic attributes can be established to aid reservoir quality delineation from wireline logs that can be used in frontier areas. With the ML porosity model, a Random Forest Regressor, the square of the correlation was 0.84 between predicted- and helium porosity test data over a large dataset consisting of 38 wells within the Stø Formation across the SW Barents Sea. By integrating the continuous ML porosity logs and core data, it was possible to differentiate three distinct bed types on wireline log responses within the Stø Formation. Particularly, the relationship between Gamma ray (GR) and porosity was effective in separating high porosity clean sand-, low porosity cemented clean sand and more clay and silt rich intervals. Additionally, in the P-wave velocity (VP) - density domain, separation of high porosity clean sand- and heavily cemented low porosity clean sand intervals were possible. The results also show that the ML derived porosity curves coincide with previously published and independent facies data from a selection of the wells included in the study. This demonstrates the applicability of the model in the region, because the Stø Formation has been described to exhibit similar lithological- and mineralogical properties over large parts of the Western Barents Sea area. Even though, continuous porosity data could be estimated from other sources like VP, neutron or density logs, this would generally require matrix and fluid information. This study demonstrated the effectiveness of the ML model in generating continuous porosity logs that are useful for characterizing and predicting reservoir properties in new wells. This methodology offers a workflow for exploiting already acquired core and well log data for frontier exploration that can be adapted to other formations and exploration scenarios worldwide.

1. Introduction

Reservoir quality characterization and prediction outside cored intervals remains a key challenge in offshore subsurface exploration because reservoir properties cannot be accurately determined from any remote sensing tools. This makes in particular reservoir property assessments on a regional scale demanding because core data are expensive and time consuming to acquire and these data are sporadic rather than continuous measurements along the well track. Hence, various predictive models and workflows are constantly being established and refined to increase the success rate of accurate reservoir quality delineation e.g., (Ajdukiewicz and Lander, 2010). More recently,

machine learning, a pure predictive workflow, has been employed for this purpose e.g. (Ahmadi and Chen, 2019; Urang et al., 2020). ML can effectively generate continuous porosity profiles that can be used for reservoir quality assessment in a regional context, but the lack of geological understanding can make predictions ambiguous, particularly moving away from wells or intervals without core material.

Detailed core analysis is crucial for characterizing the depositional- and diagenetic history of a sedimentary unit, however, such a workflow is cumbersome and expensive for reservoir quality discrimination in a regional context. This approach can make it difficult to constrain the spatial and temporal distributions of intervals with varying reservoir quality. To address this problem, several studies have focused on the interpretation of lithological- and diagenetic facies-e.g. (Ozkan et al., 2011; Cui et al., 2017), and electrofacies analysis e.g. (Kiaei et al., 2015), from wireline log data, while other studies have focused on pure predictive workflows for estimating key reservoir parameters e.g., (Helle et al., 2001; Lim, 2005; Urang et al., 2020; Agbadze et al., 2022). Here we present a hybrid methodology, which integrates detailed core analysis with a pure predictive workflow to aid effective reservoir quality discrimination. This study demonstrate the potential of using historical core data to estimate reservoir properties using ML and how these results can be integrated with detailed petrographic- and lithological knowledge to collectively aid regional reservoir quality delineation in intervals without cores. The integration of detailed petrographic knowledge aid the interpretation of model results and forms the basis for generating formation-specific templates that can deduce lithological- and diagenetic characteristics from well log responses. This approach differs from conventional electrofacies analysis in that it uses predetermined diagenetic- and lithological information and a predicted reservoir parameter, in this case porosity, to aid the discrimination of diagenetic and lithological attributes from well log data. It also differs from pure predictive workflows because detailed core analysis from a selection of wells are integrated and fundamental to several key steps in the methodology (Fig. 2). The availability of well log data and routine core analysis within the most important reservoir sandstone units from the Norwegian Continental Shelf (NCS), and likely for equivalent settings elsewhere in the world, makes this hybrid methodology adaptable to several exploration scenarios. Exploiting the existing infrastructure with nearby field development can for example significantly increase the life span of an installation and reduce operating costs.

The Stø Formation was chosen to test this integrated methodology in predicting nonlinear heterogeneous reservoir properties because the sedimentary succession has proven to exhibit large porosity variations in otherwise similar sandstone intervals consisting of texturally- and mineralogical mature sedimentary units (Olaussen et al., 1984; Klausen et al., 2018) across larger parts of the SW Barents Sea. Moreover, a patchy illitic clay coating has been identified to be the most important factor controlling reservoir quality. The patchy nature of this clay coating ultimately dictates quartz cement volumes and thus porosity (Hansen et al., 2017; Løvstad et al., 2022). In the context of petroleum exploration, clay coated sandstone reservoirs have gained much attention because for their ability to retain excellent reservoir properties even at great burial (Heald and Larese, 1974; Ehrenberg, 1993; Storvoll et al., 2002; Berger et al., 2009; Taylor et al., 2010; Ajdukiewicz and Larese, 2012; Dowey et al., 2012; Haile et al., 2018; Line et al., 2018; Porten et al., 2019; Worden et al., 2020). However, despite their huge potential in preserving reservoir quality at depth, no attempts have been made to characterize these units using wireline log data, to increase their predictability. Up until now, reservoir quality assessment of the Stø Formation has relied on core data, like helium porosity measurements and thin section analysis, and where the extent of the patchy illite coating has proven difficult to quantify (Løvstad et al., 2022). Therefore, it is of particular interest in this study to establish a framework for separating these units based on simpler means of data, which are continuous in its nature and applicable on a regional scale. Successful identification of clay-coated sandstone intervals may have huge implications for

identifying hydrocarbon- and CO2 storage reservoir sites in frontier areas, without the need of additional core material.

This study intends to demonstrate the potential in using historical core data to aid effective reservoir quality delineation at a regional scale without the need of additional core data. This methodlogy will be exemplified with the use of Stø Formation in the SW Barents Sea as a case study. The research objectives are to: (1) establish a ML based porosity predictor that can serve the purpose of effectively generating continuous porosity profiles outside cored intervals, (2) demonstrate how the integration of detailed core analysis can be used to strategically sub group data and aid the interpretation of the modelling results and (3) exemplify how this integrated methodology can be applicable to construct formation-specific templates from well log responses and facies data to aid reservoir quality determination in intervals without core data.

2. Geological setting

The study area lies within the SW Barents Sea, which is part of an epicontinental sea situated at the north western corner of the Eurasian continental plate. The study includes wells situated in the Hammerfest Basin, Bjarmeland Platform, Fingerdjupet sub-basin, Bjørnøyrenna Fault Complex, Polheim sub-platform and Ringvassøy Fault Complex (Fig. 1), all of which comprise the Stø Formation. The Stø Formation is part of the Realgrunnen Subgroup and is a Jurassic sandstone that was deposited between the Pliensbachian and Bajocian times (Dalland et al., 1988). The sandstones of the Stø Formation comprises shallow marine to offshore deposits. The most reservoir prone clean sandstone intervals were deposited in a shallow water coastal environment with fluctuating energy levels (Olaussen et al., 1984) and with relative influence of tidal- and wave action at certain locations depending on sea-level fluctuations and local basin topography (Klausen et al., 2018). The Stø Formation have been interpreted to be deposited in low-accommodation basins over large parts of the SW Barents Sea region in an overall transgressive regime (Olaussen et al., 1984; Klausen et al., 2018) that was interrupted by several regressive cycles. The highly condensed nature of this succession testifies the co-acting of deposition, erosion and reworking over several million years, which resulted in the texturally - and mineralogical mature sandstones that is typical for this formation. The Stø Formation is currently not buried to its maximum burial depth because of extensive uplift that influenced the entire southwestern Barents Sea region sometime during the Oligocene or Eocene (Baig et al., 2016). For example, within the Hammerfest Basin, where most of the wells in this study are located, results of Baig et al. (2016) show that this area is uplifted from 800 – to 1400 m and where there is an increase in the magnitude from west towards the east. The burial history of the Stø Formation is of particular importance in areas where the formation has been subjected to large maximum burial depths (>2.5 km). Several studies (Olaussen et al., 1984; Bergan and Knarud, 1993) have shown that the Stø Formation is feldspar poor and consist predominantly of mature quartz arenites which have important implications for diagenetic processes that occur upon burial. Quartz cementation has been identified as the key controlling factor on reservoir quality heterogeneity in settings where the formation has been deeply buried (Olaussen et al., 1984). More detailed petrographic studies (Hansen et al., 2017; Løvstad et al., 2022) have revealed that a thin illitic clay coating is present in varying amounts within the Stø Formation, and those intervals with effective clay coats can limit the amount quartz cementation and thus preserve abnormally high porosities in certain units.

3. Methods and data

3.1. Detailed petrographic study – a prerequisite for a successful modelling process

This methodology (Fig. 2) require an in-depth geological
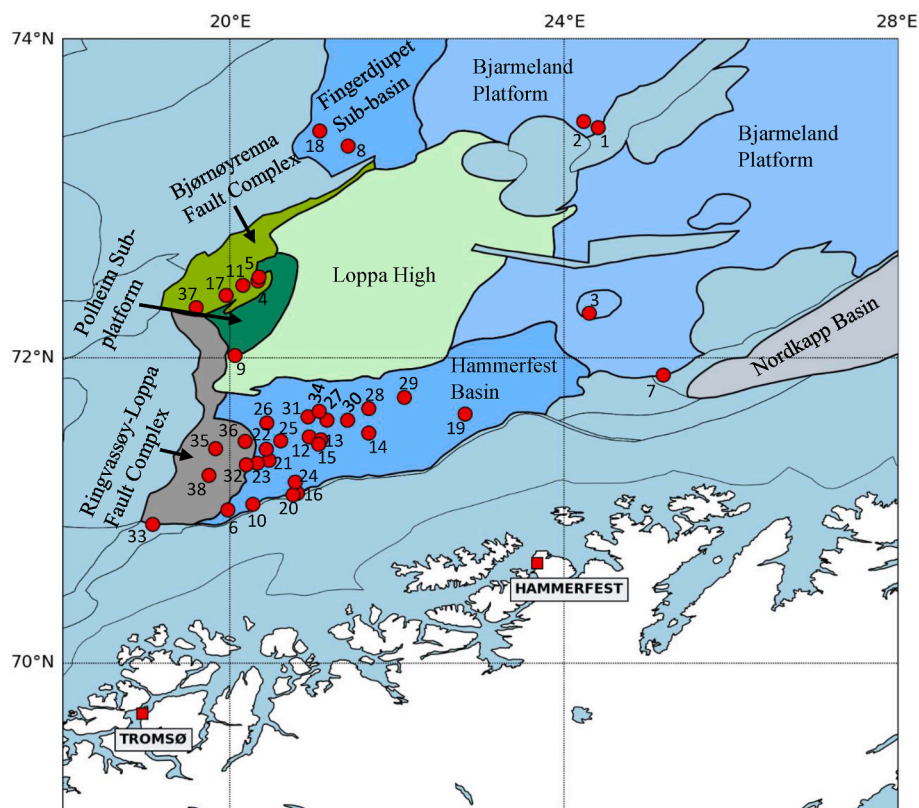
Fig. 1. Map showing the location of wells included in the study. Each well have a unique well id: **1:** 7324/8–1, **2:** 7324/7–2, **3:** 7224/7–1, **4:** 7220/8–1, **5:** 72220/5–1, **6:** 7119/12–2, **7:** 7125/1-1, **8:** 7321/8–1, **9:** 7220/10–1, **10:** 7120/10–1, **11:** 7220/7–1, **12:** 7120/9–1, **13:** 7121/7–1, **14:** 7121/5–3, **15:** 7121/7–2, **16:** 7120/12–2, **17:** 7219/9–1, **18:** 7321/7–1, **19:** 7122/6–1, **20:** 7120/12–1, **21:** 7120/8–2, **22:** 7120/8–1, **23:** 7120/7–2, **24:** 7120/12–3, **25:** 7120/8–3, **26:** 7120/5–1, **27:** 7121/4–1, **28:** 7121/5–2, **29:** 7122/4–1, **30:** 7121/5–1, **31:** 7120/6–1, **32:** 7120/7–1, **33:** 7019/1-1, **34:** 7121/4–2, **35:** 7119/9–1, **36:** 7120/7–3, **37:** 7219/8–2, **38:** 7119/12–3.

understanding of processes (e.g., diagenetic processes and lithological characteristics) that affect reservoir quality heterogeneity within the formation under consideration. This is important for a couple of reasons: (1) detailed core analysis will aid the interpreter deciding which reservoir quality parameters to model (e.g., porosity, permeability, water saturation, clay content) and (2) a comprehensive understanding of reservoir quality controlling factors is crucial for the interpretation of the model output and separating the modeled data into strategic subsets. The latter can aid the interpreter to successfully cluster data representing key lithological and/or diagenetic attributes.

In this study, porosity was chosen as the parameter to be modeled. This selection is based on detailed petrographic studies (Hansen et al., 2017; Løvstad et al., 2022) that concluded that quartz cementation is the predominant factor controlling reservoir quality heterogeneity. As a consequence, porosity and permeability tend to exhibit a linear relationship within these sandstones, which is further indicated by the study of Ogebule et al. (2020). Therefore, permeability was excluded from the modelling process to keep the model as simple as possible. However, permeability can be a crucial parameter to model in other scenarios and the parameter selection should be based on a solid understanding of reservoir quality controls.

3.2. Data preprocessing

The dataset consists of a collection of helium porosity- and wireline log data from 38 wells within the Stø Formation in the SW Barents Sea Area (Fig. 1 and Table 1). The included wireline logs were limited to the most basic well logs commonly available for all wells on the NCS, namely depth, GR, density, VP, neutron, medium- and deep resistivity (Table 2). This to ensure that the model can be relevant to most wells

with basic well log data. The initial dataset consisted of 20899 data points and included all selected wireline log- and helium porosity data from within the Stø Formation in the 38 wells. Several preprocessing steps were carried out before training the ML algorithms on the dataset. The first step was to remove all feature instances that did not contain an accompanied helium porosity value, meaning that the initial data set was reduced to 5915 data instances (Table 2). Next, the dataset was split into training- and test sets where 80% of the data was used in training and 20% was used for testing. This subdivision was performed using a random split of data instances, but with a seed that ensures reproducible results. Following the train-test split, the training set was standardized by removing the mean and scaling to unit variance.

3.3. The modelling process

In this study, we compared the performance of two machine-learning algorithms over a compiled dataset that has shown the ability to solve complex non-linear problems; a fully connected feed forward Neural Network (NN) and a Random Forest Regressor (RFG). The modelling process was defined as a supervised regression problem, using the wireline log- and helium porosity data as features and labels, respectively. The RFG and NN hyperparameter optimization was carried out using an exhaustive grid search in conjunction with 3-fold cross validation over the training set, where the exhaustive grid search was defined after some trial and error exploration. The chosen hyperparameter settings for the two models were the parameter combination showing the highest mean test score after cross validation. After searching for the optimal parameter settings, the models were trained using the training set and tested on the test set. A final evaluation and comparison of the two models was performed by assessing the coeffi-
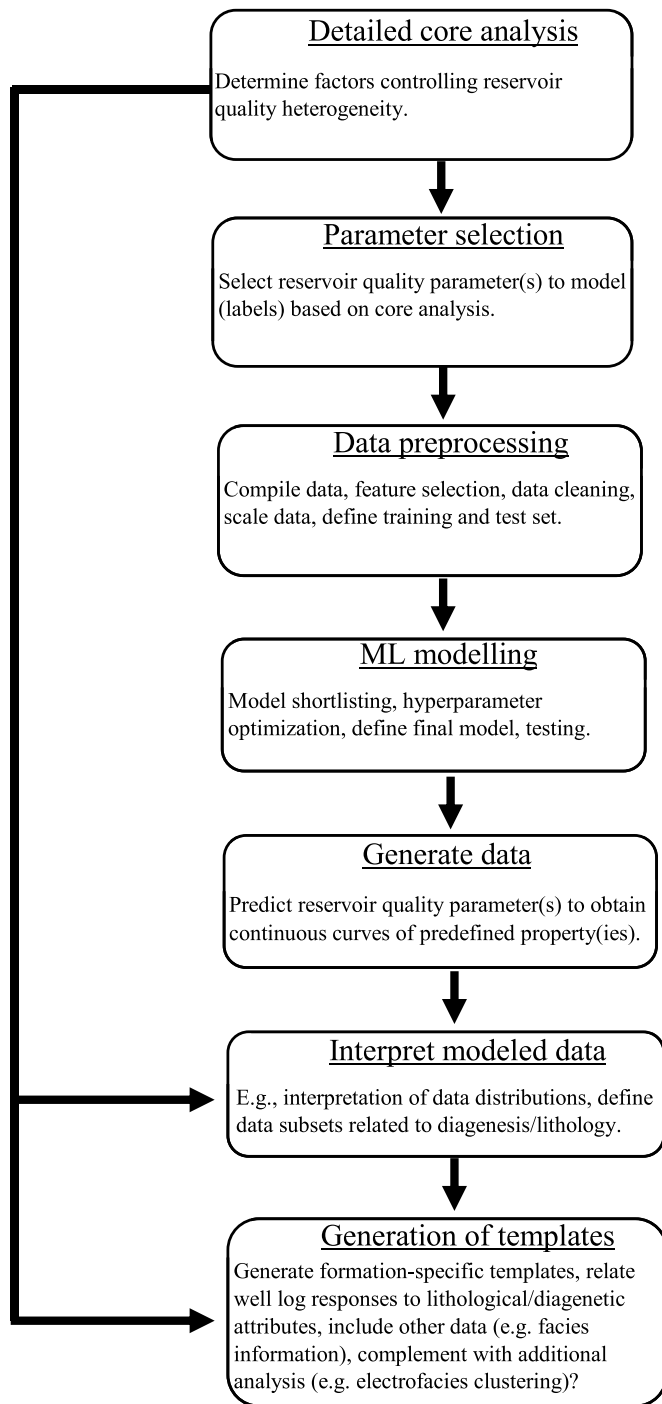
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n}}$$

where Y is the label, $\hat{Y}$ is the predicted value, $\bar{Y}$ is the mean value of all labels and n is the number of data points.

Based on the RFG model's hyperparameter settings, two specific porosity modelling approaches were tested and compared, (1) test the model's ability to predict porosity in wells where a subset of the core plug data was involved in the training set (random split - RS) and (2) test model performance in wells excluded from training (Blind test split- BS). The only difference between the generations of the models resulting from the two approaches is the train-test split. The former approach randomly splits data instances from the entire dataset into a train- and a test set, whereas the latter approach picks 33 random wells for training and uses the remaining 5 wells for testing. For both approaches, 25 unique iterations with varying random train-test splits were performed to assess the result, meaning that 25 models were generated in each case, all trained on a slightly different subset of the data. Additionally, the performances of the RS- and BS approaches were tested and compared for a fixed case based on data from three wells with good helium porosity coverage, namely 7121/5–1, 7120/6–1 and 7219/8–2. For the RS case, this meant simply using the RFG model to predict porosity in these wells, while a new model was established for the BS approach by excluding these three wells from the training process and only fitting the model based on the remaining 35 wells. The entire workflow from combing wireline log- and core plug data to cleaning- and plotting data and training the ML models was carried out using Python ([Van Rossum and Drake Jr, 1995](#)) and the third party libraries Pandas ([McKinney, 2010](#)), Matplotlib ([Hunter, 2007](#)) and Scikit-learn ([Pedregosa et al., 2011](#)), respectively. Specific details on how the NN and RFG models are trained and make predictions can be found in the literature ([Gardner and Dorling, 1998](#); [Breiman, 2001](#); [Pedregosa et al., 2011](#)).

3.4. Linking ML model results with key geological information

Continuous porosity logs were generated for all 38 wells within the Stø Formation based on the initial RFG model that was constructed based on the training set used to search for the optimal hyper-parameters. This enabled the inclusion of the entire dataset (20899 data points) with all wireline log measurements for further analysis. Shale volume (V-shale) was also estimated in all wells based on the GR log ([Asquith and Krygowski, 2004](#)) using the non-linear Larionov for older rocks correction ([Larionov, 1969](#)), where all data points with a V-shale > 20% were subsequently filtered out. Additionally, a secondary depth curve representing the maximum burial depth was generated on the basis of uplift estimates for the area presented by [Baig et al. (2016)](#). Based on previous detailed petrografic studies ([Hansen et al., 2017](#); [Løvstad et al., 2022](#)), the Stø Formation exhibits highly varying- and wide porosity distributions in deeply buried intervals depending on the effectiveness of a patchy illltic clay coating, which ultimately controls quartz cement volumes. This makes the porosity distribution of clean sand intervals within the Stø Formation interesting to study in a regional context and, in particular, the high-and low porosity range. To target the upper- and lower most part of the porosity distribution, the entire dataset was divided into three subsets from here on referred to as the Q1-, IQR- and Q3 data. For every 100-m depth interval, all data instances associated with that depth interval wered labeled Q1 if the accompanying predicted porosity value was lower than the 25th percentile, IQR if the porosity value was within the interquartile range or Q3 if the predicted porosity value was above the 75th percentile (e.g., [Fig. 6](#)A). Finally, Q1, IQR and Q3 labeled data points were merged with other similarly labeled data points from all other 100-m depth intervals, forming the full Q1, IQR and Q3 datasets along the entire porosity-depth profile (e.g., [Fig. 6](#)B–D). These datasets are fundamental for the

---

Detailed core analysis

Determine factors controlling reservoir quality heterogeneity.

↓

Parameter selection

Select reservoir quality parameter(s) to model (labels) based on core analysis.

↓

Data preprocessing

Compile data, feature selection, data cleaning, scale data, define training and test set.

↓

ML modelling

Model shortlisting, hyperparameter optimization, define final model, testing.

↓

Generate data

Predict reservoir quality parameter(s) to obtain continuous curves of predefined property(ies).

↓

Interpret modeled data

E.g., interpretation of data distributions, define data subsets related to diagenesis/lithology.

↓

Generation of templates

Generate formation-specific templates, relate well log responses to lithological/diagenetic attributes, include other data (e.g. facies information), complement with additional analysis (e.g. electrofacies clustering)?

Fig. 2. Workflow diagram highlighting key steps of the methodology.

cient of determination (R^2), means squared error (MSE), mean absolute error (MAE) and root mean squared error (RMSE), which formed the basis for the final model selection:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \widehat{Y}_i|$$

Table 1
Well data summary. Data retrieved from the Norwegian Petroleum Directorate.

Well	Latitude	Longitude	Stø FM top [MD]	Stø FM base [MD]	Uplift [m]	Maxb. depth (Stø FM top) [*]	WD [m]	KB [m]	No. core plugs	Publication date
7324/8-1	73.45373600	24.40383900	662.03	678.94	1900	2139.03	398	25	55	September 17, 2015
7324/7-2	73.49287900	24.23357000	712.01	735.94	1950	2204.01	418	40	64	July 06, 2016
7224/7-1	72.28705600	24.30039700	894.03	918.87	1300	1901.53	269	23.5	67	December 02, 2004
7220/8-1	72.49332900	20.33481900	1276.09	1353.97	1250	2129.09	374	23	221	May 02, 2013
7220/5-1	72.51881600	20.342294	1337.05	1414.93	1250	2159.05	388	40	285	March 24, 2014
7119/12-2	71.01629300	19.97229500	1372.11	1516.89	950	2117.11	180	25	228	February 11, 2005
7125/1-1	71.89201600	25.18734500	1399.08	1520.85	1250	2373.38	252.2	23.5	83	January 11, 2005
7321/8-1	73.33866700	21.41496000	1437.03	1454.86	1850	2796.03	468	23	31	January 06, 2005
7220/10-1	72.01786800	20.05774100	1513.08	1644.90	1300	2438.08	341	34	230	January 13, 2015
7120/10-1	71.05364000	20.27483700	1568.09	1654.96	1100	2460.09	183	25	39	February 11, 2005
7220/7-1	72.46238700	20.15321600	1781.15	1856.89	1150	2526.15	365	40	284	April 11, 2014
7120/9-1	71.49263900	20.94706700	1840.13	1895.90	1000	2497.13	320	23	171	February 11, 2005
7121/7-1	71.47436300	21.08756300	1849.12	1908.71	1100	2598.12	329	22	154	February 11, 2005
7121/5-3	71.51735100	21.66193700	1880.06	1927.91	1000	2511.06	345	24	127	April 11, 2003
7121/7-2	71.44578700	21.05595800	1882.00	1938.00	1000	2535.00	325	22	138	February 11, 2005
7120/12-2	71.12699100	20.80524900	1892.09	1977.90	1000	2703.09	164	25	47	February 11, 2005
7219/9-1	72.40217600	19.95403900	1951.07	2061.87	1100	2672.07	356	23	373	December 03, 2004
7321/7-1	73.43410900	21.07447400	1999.08	2021.94	1900	3400.58	475	23.5	19	January 05, 2005
7122/6-1	71.64063100	22.81216800	2015.08	2037.94	1100	2691.08	401	23	72	May 18, 2004
7120/12-1	71.11543700	20.75555500	2047.09	2151.94	1000	2855.09	167	25	16	February 11, 2005
7120/8-2	71.33961800	20.46592200	2081.07	2189.88	950	2761.07	245	25	278	February 11, 2005
7120/8-1	71.41148500	20.43497000	2092.04	2189.88	950	2747.04	270	25	74	February 11, 2005
7120/7-2	71.32177500	20.32887400	2150.11	2255.87	850	2737.11	241	22	218	February 11, 2005
7120/12-3	71.19727500	20.77877300	2158.03	2219.91	850	2800.03	185	23	55	February 11, 2005
7120/8-3	71.46547100	20.60257000	2192.02	2276.90	900	2789.52	280.5	22	100	June 29, 2004
7120/5-1	71.58300200	20.43665200	2285.13	2426.87	1000	2967.13	296	22	368	May 18, 2004
7121/4-1	71.60184200	21.15635500	2318.05	2395.93	900	2861.05	335	22	207	August 03, 2021
7121/5-2	71.67437600	21.65688400	2323.00	2400.00	1100	3073.00	328	22	354	January 05, 2005
7122/4-1	71.74929000	22.08527600	2326.00	2386.00	1050	3008.00	344.5	23.5	195	May 18, 2004
7121/5-1	71.600508	21.406112	2369.11	2444.85	1000	3011.11	336	22	209	January 04, 2005
7120/6-1	71.621865	20.933245	2386.02	2470.00	1100	3149.02	314	23	183	February 11, 2005
7120/7-1	71.31199700	20.18938000	2408.12	2521.96	950	3099.62	233.5	25	173	February 11, 2005
7019/1-1	70.91997900	19.07261500	2447.14	2609.90	800	3033.14	190	24	51	February 11, 2003
7121/4-2	71.65918100	21.06277600	2480.05	2556.86	900	3041.05	317	22	215	February 11, 2005
7119/9-1	71.41669600	19.82850500	2748.13	2867.91	850	3372.13	201	25	69	January 04, 2005
7120/7-3	71.46395400	20.17944200	2889.10	2968.95	850	3458.10	259	22	67	May 18, 2004
7219/8-2	72.323499	19.589978	2898.04	2984.91	1100	3623.04	344	31	313	September 30, 2015
7119/12-3	71.24085100	19.74367800	3144.06	3298.90	750	3654.06	211	29	82	February 11, 2005

presented study. Additionally, facies data in four wells obtained from Klausen et al. (2018) offered the possibility to study four wells in more detail by relating the ML generated porosity and various wireline log parameters to facies and diagenetic fingerprints (Fig. 1, wells: 7219/8–2 (37), 7219/9–1 (17), 7220/7–1 (11) and 7220/5–1 (5)). These wells are particularly suited for this purpose because their spatial distribution is limited and facies interpretations have been correlated between them. In addition, well 7219/8–2 (37) have been buried about 1000 m deeper compared to the other three wells, which is ideal for comparing the diagenetic effect with respect to the wireline log responses and across the various facies. Analysis where facies data are included were not filtered based on V-shale, rather all data points along the well track were

Table 2
Summary of data used in porosity modelling.

	FEATURES							LABEL
	Depth [MD]	Gamma ray (GR) [API]	Neutron [v/v]	Resistivity Deep [ohm m]	Resistivity medium [ohm m]	P-sonic (VP) [μs/ft]	Density [g/cc]	Helium porosity [%]
Count	5915	5915	5915	5915	5915	5915	5915	5915
Mean	2073.1	35.2	0.1	471.8	88.1	80.7	2.4	17.4
Std.	500.0	18.2	0.1	5087.1	471.5	9.2	0.1	6.1
Min	663.2	7.2	0.0	0.2	0.3	50.3	2.0	0.1
25%	1829.6	20.7	0.1	1.2	1.6	75.5	2.3	13.7
50%	2168.7	32.4	0.2	10.2	10.6	80.1	2.4	18.0
75%	2387.2	46.1	0.2	50.0	32.7	85.9	2.4	21.8
Max	3270.6	143.7	0.4	100000.0	12721.9	112.8	2.9	33.7

included.

Aided by the ML-derived porosity data, the study will focus on identifying distinct log responses related to lithological- and diagenetic character using the GR, VP, density and the P-impedance, which is a product of VP and density.

4. Results

4.1. Machine learning – an effective method for generating porosity data

The data presented here show the efficiency of ML in generating accurate continuous porosity logs. The NN and RFG models show very similar performances in predicting porosity over the presented dataset (Table 3) and the porosity distributions obtained from both models seem consistent with the distribution of the helium porosity data (Fig. 3). However, the RFG model showed overall slightly better results for all reported metrics compared to the NN model (Table 3). However, the difference between the root mean squared error (RMSE) and mean absolute error (MAE), which can be used to diagnose the variance in individual errors, are shown to be similar. Based on the superior performance, the RFG model was used to estimate continuous porosity logs in all wells included in the study.

The results after 25 unique runs for the random split (RS)- and blind test (BT) approach indicate that the RFG model is capable of accurately predict porosity in the former approach (Fig. 4 A), while the RMSE- and MAE scores for the latter approach indicate that it is less accurately predicting porosity in blind wells (Fig. 4 B). Moreover, the results show that there is a noticeably higher variance in predictions from the BT-approach compared to the RS-approach. This can be exemplified by the e.g., the calculated 95% confidence intervals of RMSE, showing 2.66 ± 0.06 and 2.87 ± 0.11 for the RS and BT approach, respectively. For this study, additional wells without core plug data were not included in the study because the need for accurate porosity data is crucial to describe the porosity distribution in detail and link this property to wireline log responses. The results from the two different train-test split approaches in the fixed case (Table 3 and Fig. 5) shows that the random split approach outperforms the blind well test approach. Moreover, the two porosity curves deviate more from each other at certain intervals, while other intervals are similar. Both approaches show higher deviations in intervals where the helium porosity data fluctuate

Table 3
Summary of the metrics of the Neural Network (NN)- and Random Forest Regressor (RFG). * RS = random split, BT = blind test. Reported metrics for the fixed test case. Reference to Fig. 5. See method section for more details.

| Metrics – Performance test set				
	R^2	MSE	RMSE	MAE
Neural Network (NN)	0.826	6.478	2.545	1.584
Random Forest Regressor (RFG)	0.838	6.039	2.457	1.488
RFG – RS*	0.829	4.416	2.101	1.421
RFG – BT*	0.732	6.905	2.628	1.951
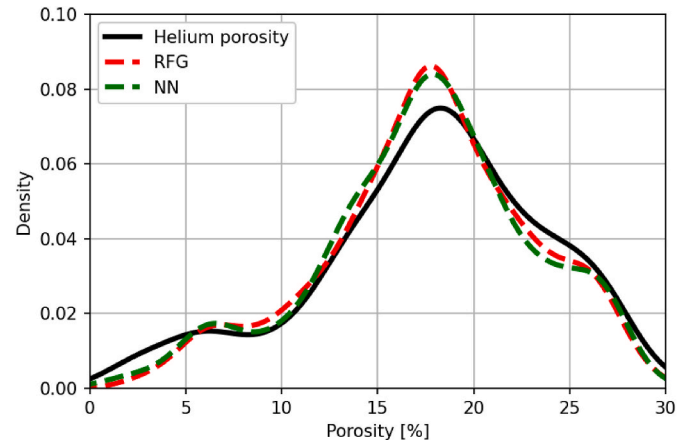
Fig. 3. Kernel density estimation (KDE) of the porosity distribution obtained on the test set for the NN and RFG model. Their distribution is compared to the helium porosity data.

considerably over short depth intervals (Fig. 5).

The generation of continuous porosity logs within the Stø Formation in all 38 wells allow for detailed characterization of the porosity distribution as a function of depth (Fig. 6). Firstly, even though there are few data points above 2270 m, there seems to be a marked change in the porosity - maximum burial depth trend at this interval, where the rate of porosity loss is increasing abruptly as a function of depth (Fig. 6 A and D). Further, this porosity-depth trend can be viewed in two different ways; (1) characterization of various parts of the porosity distribution for all depths (Fig. 6A–D) and (2) characterization of the entire porosity distribution within a specific depth interval (Fig. 6E–H). In the first case, the cross plots of porosity vs maximum burial depths colored coded with Q1, IQR and Q3 distributions show that the Q1 (Fig. 6 B) and IQR data (Fig. 6 C) contain the whole range of porosity values, whereas the Q3 data (Fig. 6 D) are skewed toward higher porosities. Additionally, the Q3 data have a minimum porosity around 10%, while the Q1-and IQR data show porosities close to 0%. In the latter case, the shallowest porosity data (Fig. 6 F) (<2500 m) are normally distributed with a mean porosity of about 25%. The intermediate depth (2500 m–2800 m) porosity data (Fig. 6 G) show a similar pattern, but where the entire distribution is shifted toward lower porosities (mean porosity around 20%). In contrast, the porosity data that lie below 2800 m show a clear bimodal distribution with a noticeable subpopulation of higher porosities (Fig. 6 H).

To characterize the maximum and minimum rates of porosity loss within the Stø Formation, data points shallower than 2270 m and the IQR data were excluded, the Q1-and Q3 porosity data were plotted as a function of maximum burial depth (Fig. 7). The results also show a tendency for the Q1 distribution to become narrower with an increase in the burial depth, while the Q3 distribution show an opposite trend and
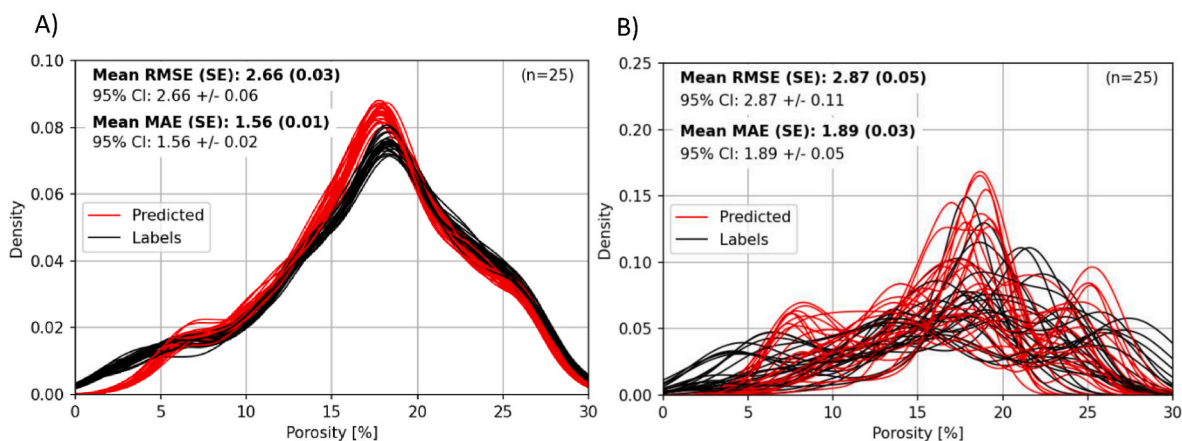
Fig. 4. KDE plots comparing predicted- (red) vs. helium (black) porosity distributions across 25 different train test splits. A) Random split approach across the entire data set with 20% used for testing and 80% of the data used for training. B) Blind test, 5 random wells used for testing in each run, while the remaining 33 wells was used for training. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
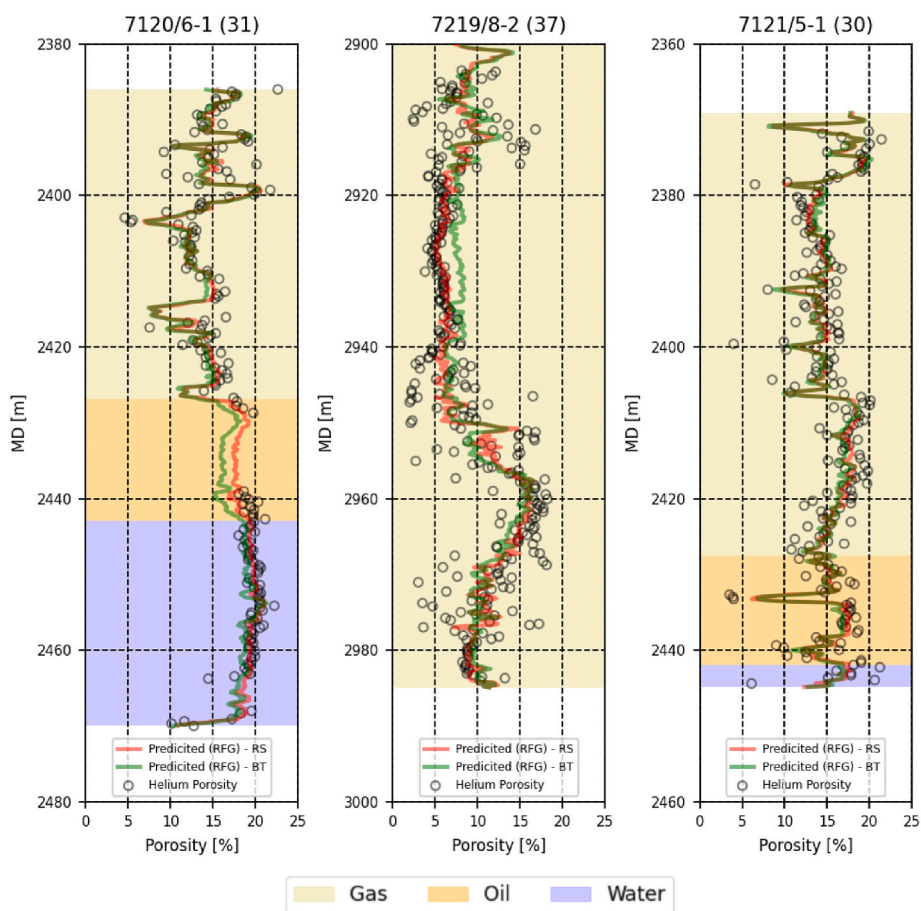
Fig. 5. Predicted porosity from RFG model compared to helium porosity measurement from the well 7120/-6-1, 7219/8–3 and 7121/5–1. Well 7219/8–2 is situated in the Bjørnøyrenna Fault complex, while the other two wells are located further south in the Hammerfest Basin. Log tracks are colorceded with fluid content and the depth represents the measured depth from the rotary table. Model metrics are listed in Table 3, referred to as RFG-RS and RFG-BT.

becomes wider as a function of depth (Fig. 7 A). There is a noticeable difference in the rate of porosity loss between the two distributions, where the fitted lines demonstrate a porosity loss of 8.1% and 6.4% per 500 m for the Q1 and Q3 distributions, respectively (Fig. 7 A). However, data below 3300 m in the Q1 distribution deviate slightly from the linear trend line and shows somewhat lower rates of porosity loss compared to the shallower data. Fig. 7B illustrates a new case representing a modified version of the Q3 distribution that is filtered on based on porosity greater or equal to 12%. This result indicates that the average rate of porosity loss is 5.6% per 500 m for the modified Q3 distribution.
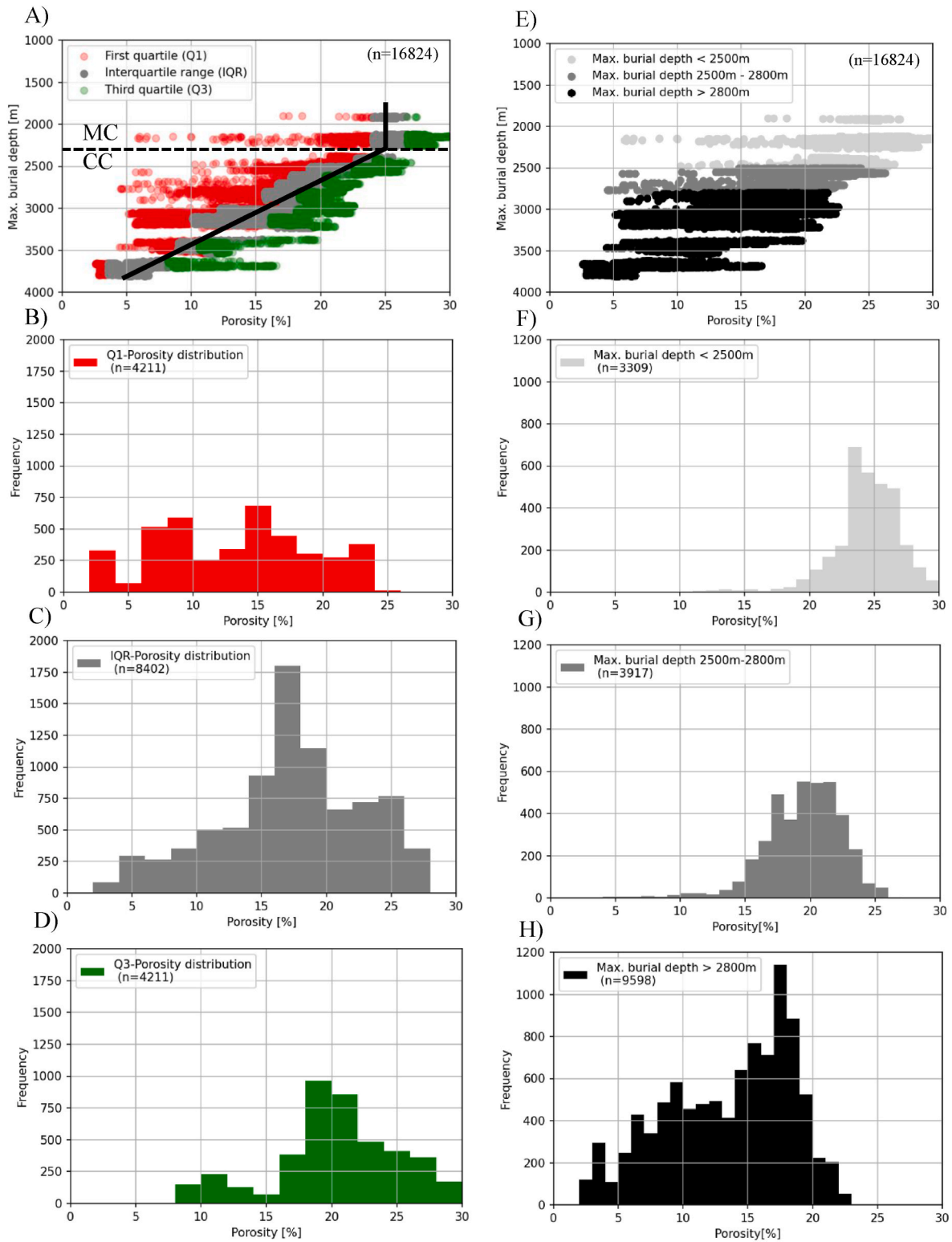
Fig. 6. RFG porosity data from all 38 wells with shale volume <0.2. A) Porosity-depth trend colored with the Q1, IQR and Q3 data. Depth represents the maximum burial depth from the seafloor. B-D) Distribution of the Q1, IQR and Q3 data. E) Porosity-depth trend colored with maximum burial depth. F–H) Distribution of the porosity data as function of depth.

4.2. Generating formation-specific templates

4.2.1. VP, density, P-impedance and ML porosity

Fig. 8 shows cross plots of VP vs density (A and B) and P-Impedance and porosity (C and D) where each parameter combination is shown for the Q1 and Q3 distribution in each plot but with a depth constrain. Fig. 8 A and C demonstrate the responses from VP - density and P-impedance-porosity in shallow buried intervals (<2700 m), respectively, whereas Fig. 8 B and C show the same parameter combinations for deeply buried units (>3300 m). The results show that the Q1 and Q3 distributions are
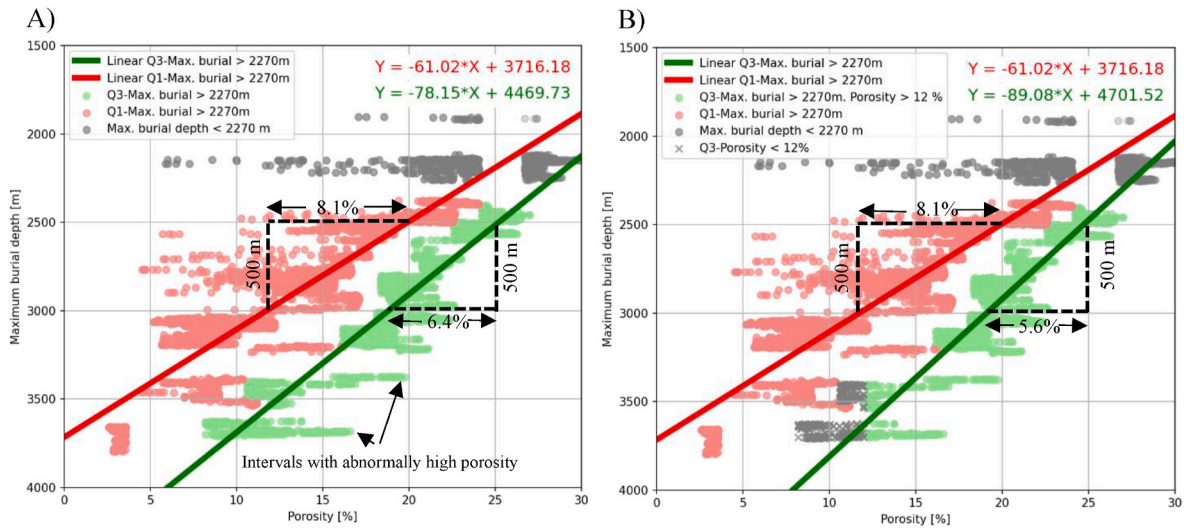
Fig. 7. Maximum burial depth vs porosity trends for the Q1 and Q3 distributions for data points with maximum burial greater than 2270 m and shale volume <0.2. The fitted lines illustrate the differences in rate of porosity loss for the Q1-and Q3 data for two cases: A) all data included. B) Porosity less than 12% excluded from the Q3 distribution.

difficult to separate from the VP - density responses and with respect to P-impedance-porosity in shallow buried intervals. Even though it is possible to distinguish certain parts of the two distributions from each other, this is especially true for the P-impedance-porosity case, large part of the two distributions is clustered (Fig. 8 A and C). In contrast, data points from deeply buried intervals (>3300 m) indicate that the Q1 and Q3 distributions are easily distinguished (Fig. 8 B and D).

A more detailed characterization which include facies data show a similar pattern in four wells (Fig. 1, well id: 37, 17, 11, 5) as

demonstrated in Fig. 9. The results show that the P-impedance-porosity- and Vp-density signatures from the three wells with maximum burial between about 2100 m and 2600 m have an unordered structure where most data points are clustered (Fig. 9 A, E, B, F, C and G). However, all these wells have a tail of data points deviating from the overall cluster, which mainly concerns the offshore-embayment facies that consist of finer grained- and silty material, and likely parts of the cleaner sandstone intervals with abundant carbonate cement (see Table 1 in Klausen et al. (2018) for complete facies description). These deviating lithologies
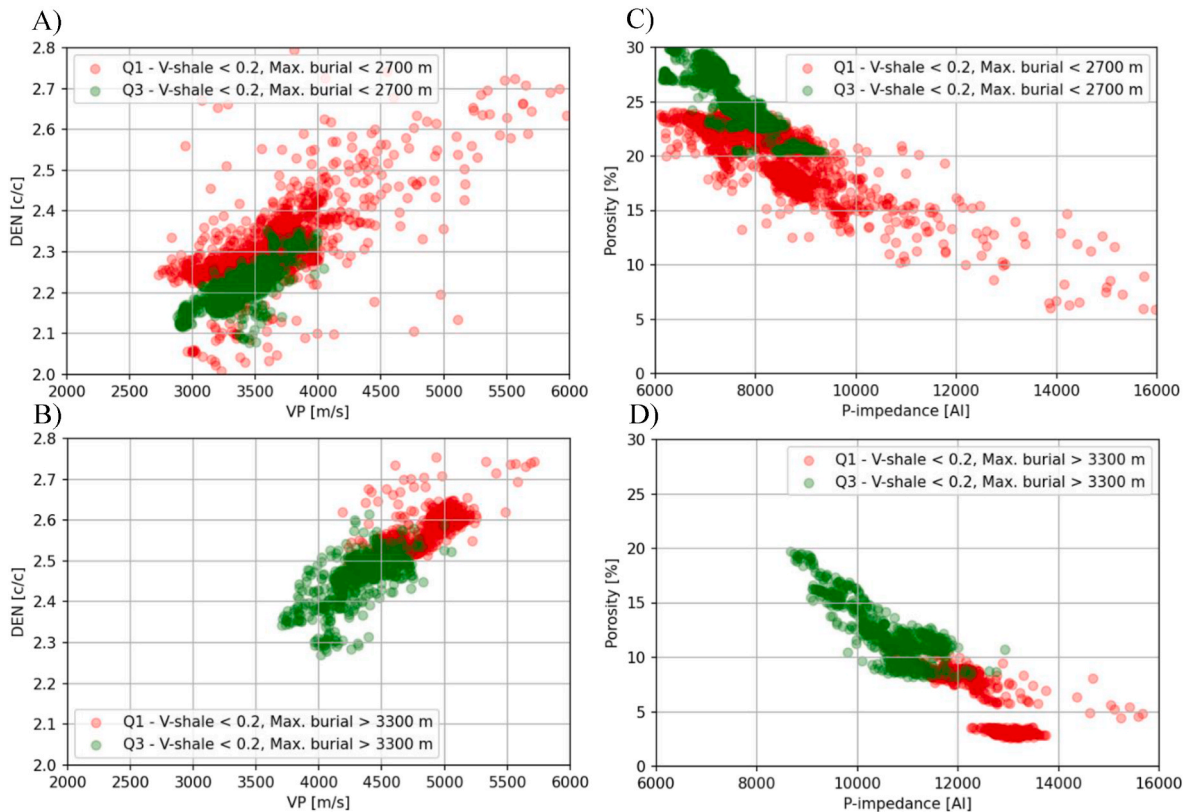
Fig. 8. Comparison of V_p, density and P-Impedance of the Q1 and Q3 distributions at shallow (<2700 m) and deep (>3300 m) maximum burial. A-B) V_p-density plot with shallow data (A) and deep data (B). C-D) P-impedance – porosity plots with shallow data (C) and deep data (D).

are characterized by an elevated Vp and density (and P-impedance) and corresponding decrease in porosity. For well 7219/8–2, the deeply buried well (Fig. 9 D and H), the results show that recorded facies are more easily separated both with respect to the P-impedance-porosity and V_p – density parameter combinations. The clean sand of the upper- and lower shoreface facies are clustered and exibit a wide range of values with respect to the various parameters. This elongated distribution has been interpreted to represent the cement trend. The offshore transition/inner shelf facies follow a similar pattern but are characterized by slightly elevated density readings compared to the shoreface facies in the V_p-density domain and slightly lower porosities in the porosity-P-impedance parameter space.

4.2.2. GR log vs. ML porosity

The four wells with available facies data were also characterized with the use of GR and porosity (Fig. 10A–D). The results show that this parameter combination can be useful for discriminating lithological characteristics in wells with shallow maximum burial (Fig. 10A–C), where the more clay rich inner shelf deposits are separated from shoreface facies associated with an increase in GR. Moreover, a clear negative correlation exits between the GR log and porosity where higher gamma coincides with a decrease in porosity (Fig. 10A–C). This trend is particularly dominant in well 7220/7–1 and 7220/8–1, where even a separation of the upper- and lower shoreface facies is evident based on the GR and porosity response. In the deeply buried well, 7219/8–2 (Fig. 10 D), the shoreface facies span predominantly over a large range of porosities but with consistently low GR readings. However, some data points within the lower shoreface facies show elevated gamma readings, which contribute to an overall "L-shaped" trend within lower shoreface facies in deeply buried intervals. It should be mentioned that the elevated gamma signals are likely not caused by any variation in K-feldspars content because the sediment is feldspar depleted (Bergan and Knarud, 1993), which could be a cause for slightly elevated GR readings. On the contrary, the upper shoreface facies shows in general lower GR readings (Fig. 10D). The facies associated with lower depositional energy, like the more distal inner shelf facies, show even higher gamma readings, consistent with the trend observed in wells with shallow burial. Based on this plot, three distinct endmember bed types can be differentiated: 1: clean sandstone with high porosity, 2: clean sandstone with low porosity, 3: low porosity fine-grained sandstone with silty- and clay rich material. The same parameter combination is plotted in two other deeply buried wells from the study area, well 7120/5–1 from the Hammerfest Basin and well 7119/9–1 from the Ringvassøy Fault Complex (Figs. 1 and 10 E). The results demonstrate a similar trend where the three distinct bed types can be distinguished from one another. Fig. 10 F, show four more examples from wells with different maximum burial depths but without facies data. The results show that the three bed types could be recognized and it demonstrates that the Stø Formation not necessarily contain all the endmember lithologies in each well.

5. Discussion

5.1. Machine learning – an effective porosity prediction method

Being able to accurately predict reservoir quality from continuous data sources, like wireline logs, has a huge potential compared to expensive and sporadic data obtained from cores. Several studies have focused on the interpretation of diagenetic-and lithological characteristics (Avseth et al., 2001; Ozkan et al., 2011; Cui et al., 2017) from wireline log data to determine reservoir quality, while other studies have focused on using a pure predictive workflows for estimating reservoir parameters e.g., (Helle et al., 2001; Urang et al., 2020). A pure predictive workflow can effectively generate large amounts of porosity data for use in a regional exploration context, but the lack of a geological understanding, can make predictions away from wells or intervals without core material ambiguous. This study has employed an integrated methodology that combines core analysis with a ML based porosity predictor, to establish a framework that can deduce diagenetic and lithological characteristics from distinct well log responses to aid reservoir quality determination. In this way, historical data can be effectively used to aid detailed interpretations in blind wells. The consistent and good performance of the RFG-RS porosity model provide reliability in the model's capability for generating accurate continuous porosity logs in wells where some of the helium porosity data were involved in training (Table 3, Fig. 4). The RFG-BT model (Table 3) is slightly less robust for making accurate porosity estimates, compared to the RFG-RS model (Fig. 4), but the results are still adequate to make porosity predictions in blind wells (Fig. 5). Consequently, the RFG-BT modelling approach can still be useful in the exploration of frontier areas. The findings in this study exemplify that once a predictive framework for determining lithological- and diagenetic characteristics from wireline log responses has been established; the inclusion of RFG-BT porosity estimator can complement these interpretations in new wells that lack core information (Fig. 5). However, in the process of establishing a formation-specific framework we propose using the RFG-RS modelling approach, where continuous porosity logs are predicted in wells where helium porosity data were involved in training.

5.2. Machine learning derived porosity profile is consistent with petrographic analysis

It is essential to have a good understanding of diagenetic- and/or depositional processes that control reservoir quality variations when trying to deduce lithological characteristics from well log data within a specific formation. In the Stø Formation, the main reservoir intervals are found within shallow marine shoreface facies consisting mainly of texturally- and mineralogically mature sedimentary units (Olaussen et al., 1984; Klausen et al., 2018; Ogebule et al., 2020). In deeply buried parts of the Stø Formation (about >3000 m), quartz cement is the main factor controlling porosity, which has subsequently been interpreted to be controlled by the presence or absence of an illitic clay coating (Hansen et al., 2017; Løvstad et al., 2022). With the use of helium porosity data from 14 wells, mainly within the Hammerfest Basin, Løvstad et al. (2022) also found that the rate of porosity loss between coated- and negligible coated intervals becomes increasingly larger as a function of burial depth.

In the Stø Formation case, it is therefore essential to evaluate the ML generated porosity profile's ability to capture this trend, if present, in a regional context. The Q1 and Q3 datasets, which are meant to represent negligible coated and coated intervals, respectively, exhibit expected porosity distributions as a function of maximum burial depth (Fig. 6B and D). When all porosity data are included, the porosity distributions as a function of varying depth also exhibited expected patterns that reflect the gradual influence of diagenesis (Fig. 6F–H). Here, shallow and intermediate buried intervals are normally distributed (Fig. 6F and G), whereas the deeply buried intervals show a bimodal distribution with a clear subpopulation of abnormally high porosity (Fig. 6H). This subpopulation of higher porosities can be a clear sign of a porosity preserving mechanism (Bloch et al., 2002), in this case the clay-coated intervals of the Stø Formation. When examining the Q1 and Q3 data separately, it is evident that the Q1 porosity distribution becomes narrower- and the Q3 distribution becomes wider as a function of increasing burial depth (Fig. 7). This observation explains the tendency of negligible coated intervals to become increasingly more quartz cement with an increase in the time-temperature integral (TTI) (Walderhaug, 1994, 1996), whereas the quartz cement volumes in the coated intervals are dictated by the clay coating coverage and thus exhibit a wider range of porosity (Ajdukiewicz and Larese, 2012). The presented results imply that the ML model is capable of representing the petrographic observations and interpretations made from core data by Løvstad et al. (2022), which shows a greater difference in porosity loss as a function of burial depth between negligible coated- and coated intervals (Fig. 7).
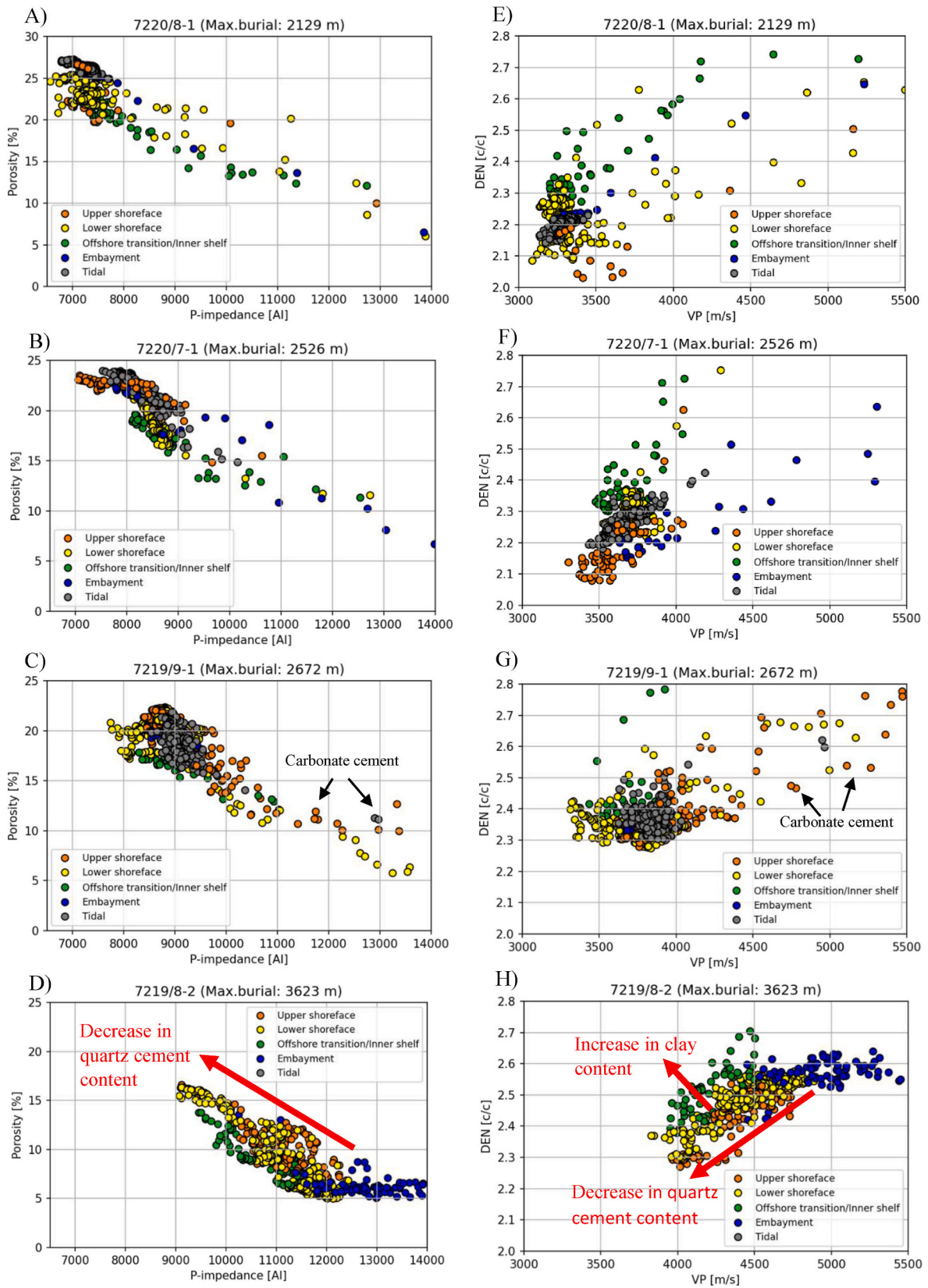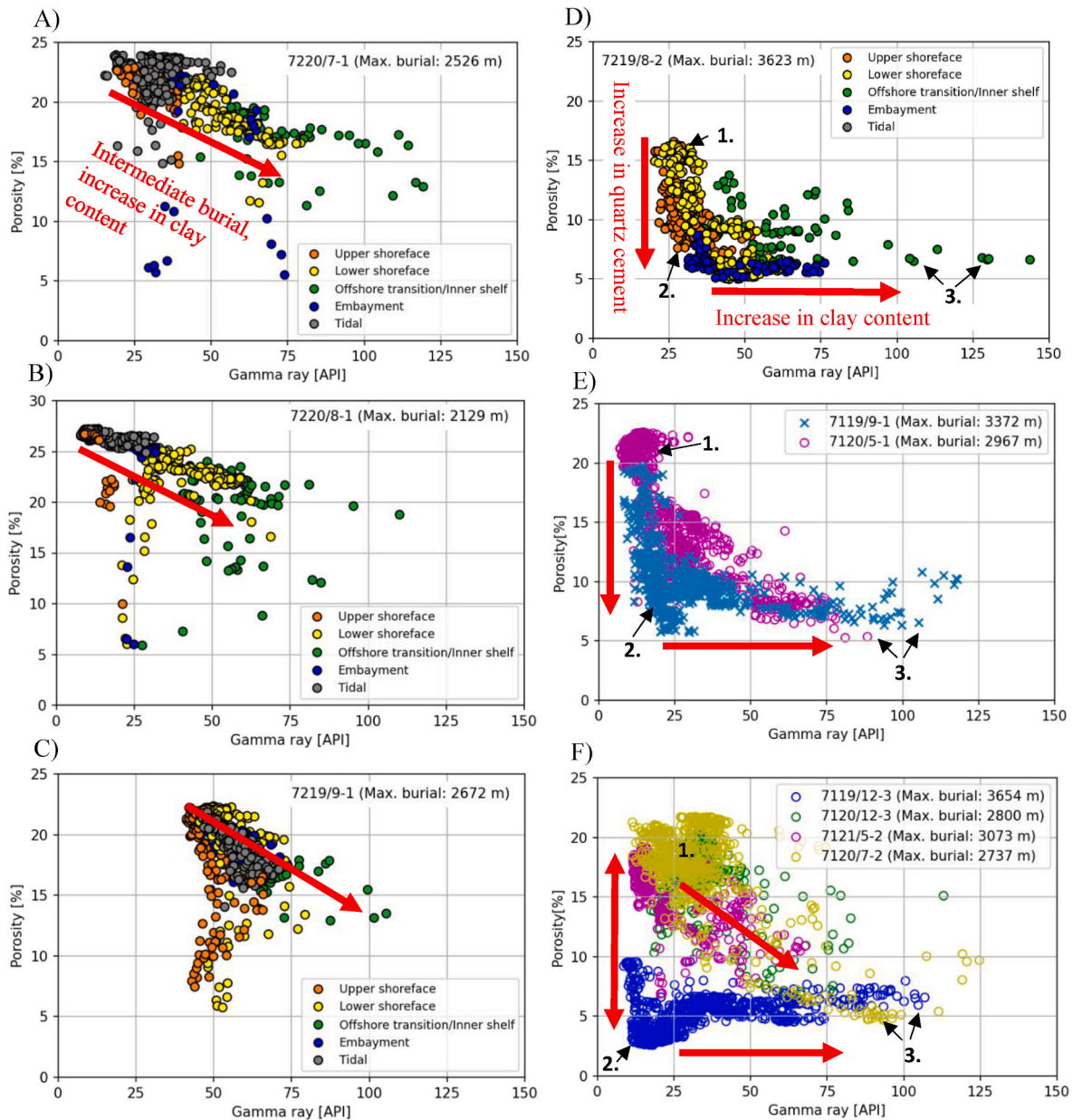
Fig. 9. Characterization of elastic parameters in four wells colored with facies data from Klausen et al. (2018). A-D) P-impedance vs porosity and E-F) Vp-density plots. D and H plots are from the more deeply buried well 721978–2, while the other plots are from wells with maximum burial of 1000 m shallower or more.

Fig. 10. Gamma ray-porosity plots. A-D) The four wells with facies information obtained from Klausen et al. (2018). Note the three distinct lithologies that can be recognized in D: 1: Clean sand with high porosity, 2: clean sand with low porosity and 3: sility/shaley sand with low porosity. E) Well 7119/9–1 and 7120/5–1 with maximum burial of about 3300 m and 2900, respectively. F) Four wells form different parts of the study area with varying maximum burial depths and with only certain distinct lithologies present. Arrows in figures B, C, E and F indicate the relative amount of cement and clay content according to figures A and D.

However, the porosity data from the deepest part of the Q1 data deviate slightly from the fitted line. A similar trend was observed by Marcussen et al. (2010) in the Etive formation in the northern North Sea, where the porosity-depth gradient is steeper than for shallower buried intervals. The reason for this is that the surface area available for quartz nucleation is reduced as the pore volumes are filled with significant amounts of quartz cement (Walderhaug, 1996). The above results indicate that the petrographic observations of Løvstad et al. (2022) are applicable in a regional context within the Stø Formation. More importantly, the separation of the Q1 and Q3 data exemplifies the potential for scaling up petrographic analysis and integrating with a pure predictive workflow to assess reservoir quality in frontier areas within the same formation. The successful characterization of these distinct diagenetic features within the Stø Formation via the ML based porosity data makes it possible to link these diagenetic attributes to typical well log responses.

5.3. Machine learning porosity data can distinctly identify well log responses of lithological- and diagenetic characters

As discussed above, petrographic results concerning reservoir quality controls were crucial for scaling up the diagenetic variation observed within the Stø Formation to the ML generated porosity data, which in this case resulted in the Q1 and Q3 data subdivision. These distinct diagenetic attributes can be characterized from well log responses aided by the ML porosity profile that are color coded by this diagenetic property, i.e. clay coated high porosity sand (Q3) and negligible coated heavily cemented sand (Q1). For other formations, diagenetic alterations that control reservoir quality may differ and should be adapted accordingly. For intermediate and shallow buried intervals, facies data could also be effective, if available, because clay and silt content tends to control reservoir quality more frequently than quartz cementation. In this way, formation-specific frameworks that can deduce lithological-

and diagenetic characteristics related to primary reservoir controls can be established for use in frontier regions.

VP and density are interesting parameters to investigate for several reasons. Firstly, they are usually recorded along most boreholes in their entirety, which makes them applicable to use for interpretations in wells on a regional scale. Secondly, they have seismic properties, which mean that they can be linked to seismic amplitude information e.g., (Avseth et al., 2001). Thirdly, VP and density are particularly sensitive to diagenetic alterations because of their strong correlation with the amount of quartz cement volume and hence porosity (Marcussen et al., 2010). The clustering behavior of the Q1 and Q3 data for VP and density (Fig. 8A) indicates that intervals with intermediate maximum burial (<2700 m) have very similar acoustic impedances (Fig. 8C). This means that the reasoning for the Q1 and Q3 data labeling may not hold for intermediately buried intervals. According to Løvstad et al. (2022), the significant porosity variation across negligible- and coated intervals within the Stø Formation was solely investigated in deeply buried intervals. As mentioned in the previous section, the ML based porosity data show only a bimodal distribution with a subpopulation of higher porosity in deeply buried intervals, meaning that the differentiation between intervals affected by the presence or absence of clay coats seems only applicable to units with larger TTI's. The successful separation of the Q1 and Q3 data with the use of VP and density at intervals with significant burial depths (>3300 m) agrees with this interpretation (Fig. 8B). However, the boundary between Q1 and Q3 labeled data could be challenging to depict from raw well log data, which emphasizes the potential in the scaling up interpretations of core analysis to the ML generated porosity profile. Additionally, note that Q1 and Q3 data are filtered on shale volume, which could be necessary to avoid overlap from more silt and clay rich intervals that are common in certain parts of the Stø Formation (Olaussen et al., 1984; Klausen et al., 2018). This can be particularly important for multi-well analysis because there will be a higher risk of masking small but important variations in VP and density response compared to single well analysis. The results from unfiltered data in single well analysis, that includes facies information, show the potential for separating distinct lithological characteristics, both in terms of cement- and matrix content variations, from the P-impedance-ML porosity and VP - density signatures (Fig. 9 D and H, respectively). Still, the separation seems ambiguous for intervals with intermediate burial depths (Fig. 9 A-C and 9 E-G). Consequently, the need for a parameter combination that can handle variations in cement- and matrix content irrespective of burial depth is needed for truly being able to delineate reservoir quality variations in blind wells on a regional scale.

This study have shown that the GR - ML porosity combination can be well suited for this purpose (Fig. 10). This is also where the integration of a ML based porosity predictor will truly shows its potential. This is because, (A) the ML model enables porosity to be used directly, which means that we do not need to infer this key property via some other parameter. (B) It does not require any known fluid- or rock properties to predict porosity from well logs in blind wells (Helle et al., 2001) in appose to density- or the sonic derived porosity. (C) It is computationally time efficient to make continuous porosity profiles in new wells once a pipeline has been established. Alternatively, in contrast to P-wave- and density parameters discussed earlier, the GR nor porosity can be directly tied to seismic amplitude information, which could be a limiting factor if results are to be integrated with seismic data. The GR – porosity relationship had earlier been investigated in one well from the Stø Formation (Ramm, 1991). Ramm (1991) discovered an interesting relationship between these parameters, but the relationship was not studied in detail with the inclusion of facies data nor the applicability on a regional scale. This study have shown that the GR – ML porosity plots enable the separation of three distinct bed types irrespective of burial depth; (1) high porosity clean shoreface sands with a varying degree of clay coating, (2) heavily cemented clean shoreface sands with negligible clay coating and (3) silt- and clay rich intervals (Fig. 10A–D). Furthermore, adding Q1 and Q3 data labels to the cleaner shoreface facies in this parameter

domain could further facilitate a simple way of mapping out these units in deeply buried intervals. This could for example be useful for linking clay-coated intervals between wells in regional studies within the Stø Formation. The test of the GR – porosity combination in wells without facies information at various locations in the SW Barents Sea (e.g., Hammerfest Basin and Bjørnøyrenna Fault Complex, see Fig. 10E and F and map in Fig. 1) shows the potential of this parameter domain for use in reservoir quality delineation on a regional scale within the Stø Formation.

Additionally, we could speculate that the elongated and "L-shaped" trends observed within the upper- and lower shoreface facies for intermediate and deeply buried wells respectively, could reflect varying amount- and different modes of clay within the Stø Formation. From this, we could interpret the lower shoreface facies to have a higher total clay content compared to the upper shoreface facies. Moreover, the occurrences of clay in the lower shoreface facies is dominated as either clay coats or pore-filling, i.e. the low GR-high porosity- and slightly higher GR-lower porosity responses, respectively. The GR-ML porosity response for the upper shoreface facies reflect in general a cleaner sandstone, where also the extent of effective clay coats is lower, leading to more heavily quartz cemented units. Based on this result we can speculate that effective clay coats are most prone to develop in the lower shoreface facies. This interpretation is comparable with the findings of Hansen et al. (2017) and Løvstad et al. (2022), which linked the amount of post depositional reworking to clay coat coverage. The amount and modes of occurrences of clay have also been shown to vary significantly within juxtaposed coastal sub-environments in other studies, that ultimately can be a key factor controlling diagenetic signatures (Haile et al., 2018). Moreover, as indicated by Wooldridge et al. (2017) there seems to be an optimum range of total clay content within the sediment that can aid the development of effective clay coats at depth.

5.4. General implications

The Stø Formation and time-equivalent formations have been studied in the context of depositional environment and mineralogical composition from several locations in the greater Western Barents Sea area and include, but not limited to, rock and core data from the Hammerfest Basin, Ringvassøy-Loppa Fault Complex, Bjørnøyrenna Fault Complex, Wilhelmøya at Svalbard and the Bjarmeland Platform (Olaussen et al., 1984; Hansen et al., 2017; Klausen et al., 2018, 2019; Haile et al., 2019; Løvstad et al., 2022). These studies show that the Stø Formation is predominantly consisting of mineralogically- and textural mature quartz arenitic sandstone beds representing wave dominated shallow marine deposits that originated in an overall transgressive development.

The results show the potential in effective use of historical core data and how the presented integrated methodology can be used to construct formation specific templates that can display lithological- and diagenetic attributes from distinct well log responses. Due to the widespread and consistent composition of the Stø Formation and its time-equivalents in the greater Barents Sea area, the presented results can have important implications for effective reservoir quality delineation in intervals or wells without core data in this region or in other similar settings worldwide.

6. Conclusion

The petroleum industry is increasingly seeking new reservoir discoveries and potential CO_2 storage sites close to existing infrastructure to increase the life span of already operating installations to save time and cost. After several tens of years of exploration on the NCS, an extensive database consisting of wireline log and core data is available. This valuable dataset has a huge potential for being exploited to establish formation-specific predictive frameworks for use in already mature provinces. ML has enabled an effective (both and time and cost) and

accurate method for estimating reservoir properties from existing core data. This study demonstrate that effective use of historical core data in conjunction with a pure predictive ML-based workflow can be used to establish formation-specific frameworks for deducing distinct lithological- and diagenetic attributes from well log data. The study also emphasis the importance of conducting detailed core analysis prior to utilizing data-driven methods for predicting reservoir quality parameters, because: (1) detailed geological information can aid the geologist to decide on which reservoir quality parameters to model and (2) lithological and diagenetic information will assist the interpretation of data derived from the model. The latter can be crucial for making strategic data subsets that can be used to link key lithological and diagenetic attributes to well log responses. The results show that high porosity clean sand-, cemented clean sand- and clay/silt rich intervals can be distinguished within Stø Formation. These distinct bed types can be recognized from basic well log data in new wells without core material and thus serve as a framework for effectively delineate reservoir quality variations on a regional scale. Particularly, the relationship between GR and ML porosity shows promising results for reservoir quality delineation because this domain can handle the effect of varying silt/clay- and quartz cement content. Moreover, the results from this parameter combination could indicate that effective clay coats are most prone to develop in lower shoreface facies within the Stø Formation. The distinct VP and density response for the high porosity clean sand- and cemented sand intervals show a potential for linking these parameters to seismic amplitude information, which could have huge implications for connecting high porosity zones between wells. Integrating historical core data with a ML-based reservoir property predictor can aid reservoir quality determination in new un-cored wells or intervals. By using already acquired data in mature provinces, the presented methodology can be employed to establish similar formation-specific frameworks elsewhere.

Author contribution

Henrik Nygaard Hansen: Concept development, Data processing and interpretations, Text, Figures, Editing, Beyene Girma Haile: Concept development, Data processing and interpretations, Text, Editing, Reidar Müller: Data processing and interpretations, Text, Editing, Jens Jahren: Concept development, Data processing and interpretation, Text, Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

Agbadze, O.K., Qiang, C., Jiaren, Y., 2022. Acoustic impedance and lithology-based reservoir porosity analysis using predictive machine learning algorithms. J. Petrol. Sci. Eng. 208, 109656 https://doi.org/10.1016/j.petrol.2021.109656.

Ahmadi, M.A., Chen, Z., 2019. Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. Petroleum 5, 271–284. https://doi.org/10.1016/j.petlm.2018.06.002.

Ajdukiewicz, J.M., Lander, R.H., 2010. Sandstone reservoir quality prediction; the state of the art. AAPG (Am. Assoc. Pet. Geol.) Bull. 94, 1083–1091. https://doi.org/10.1306/intro060110.

Ajdukiewicz, J.M., Larese, R.E., 2012. How clay grain coats inhibit quartz cement and preserve porosity in deeply buried sandstones: observations and experiments. AAPG (Am. Assoc. Pet. Geol.) Bull. 96, 2091–2119. https://doi.org/10.1306/02211211075.

Asquith, G.B., Krygowski, D., 2004. Gamma ray (chapter 3). In: MANCINI, E.A. (Ed.), Basic Well Log Analysis, 2 ed. AAPG, Tulsa, Oklahoma.

Avseth, P., Mukerji, T., Jørstad, A., Mavko, G., Veggeland, T., 2001. Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system. Geophysics 66, 1157–1176. https://doi.org/10.1190/1.1487063.

Baig, I., Faleide, J.I., Jahren, J., Mondol, N.H., 2016. Cenozoic exhumation on the southwestern Barents Shelf: estimates and uncertainties constrained from compaction and thermal maturity analyses. Mar. Petrol. Geol. 73, 105–130. https://doi.org/10.1016/j.marpetgeo.2016.02.024.

Bergan, M., Knarud, R., 1993. Apparent Changes in Clastic Mineralogy of the Triassic–Jurassic Succession, Norwegian Barents Sea: Possible Implications for Palaeodrainage and Subsidence, vol. 2. Norwegian Petroleum Society Special Publications, pp. 481–493. https://doi.org/10.1016/B978-0-444-88943-0.50034-4.

Berger, A., Gier, S., Krois, P., 2009. Porosity-preserving chlorite cements in shallow-marine volcaniclastic sandstones: evidence from Cretaceous sandstones of the Sawan gas field, Pakistan. AAPG Bull. 93, 595–615. https://doi.org/10.1306/01300908096.

Bloch, S., Lander, R.H., Bonnell, L.M., 2002. Anomalously high porosity and permeability in deeply buried sandstone reservoirs: origin and predictability. AAPG (Am. Assoc. Pet. Geol.) Bull. 86, 301–328. https://doi.org/10.1306/61EEDABC-173E-11D7-8645000102C1865D.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Cui, Y., Wang, G., Jones, S.J., Zhou, Z., Ran, Y., Lai, J., Li, R., Deng, L., 2017. Prediction of diagenetic facies using well logs–A case study from the upper Triassic Yanchang Formation, Ordos Basin, China. Mar. Petrol. Geol. 81, 50–65. https://doi.org/10.1016/j.marpetgeo.2017.01.001.

Dalland, A., Worsley, D., Ofstad, K., 1988. A litostratigraphic scheme for the Mesozoic and Cenozoic succession offshore mid- and northern Norway. NPD Bulletin No 4, 42–51.

Dowey, P.J., Hodgson, D.M., Worden, R.H., 2012. Pre-requisites, processes, and prediction of chlorite grain coatings in petroleum reservoirs: a review of subsurface examples. Mar. Petrol. Geol. 32, 63–75. https://doi.org/10.1016/j.marpetgeo.2011.11.007.

Ehrenberg, S.N., 1993. Preservation of anomalously high porosity in deeply buried sandstones by grain-coating chlorite: examples from the Norwegian Continental Shelf. AAPG (Am. Assoc. Pet. Geol.) Bull. 77, 1260–1286. https://doi.org/10.1306/BDFF8E5C-1718-11D7-8645000102C1865D.

Gardner, M.W., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32, 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0.

Haile, B.G., Czarniecka, U., Xi, K., Smyrak-Sikora, A., Jahren, J., Braathen, A., Hellevang, H., 2019. Hydrothermally induced diagenesis: evidence from shallow marine-deltaic sediments, Wilhelmøya, Svalbard. Geosci. Front. 10, 629–649. https://doi.org/10.1016/j.gsf.2018.02.015.

Haile, B.G., Klausen, T.G., Czarniecka, U., Xi, K., Jahren, J., Hellevang, H., 2018. How are diagenesis and reservoir quality linked to depositional facies? A deltaic succession, Edgeøya, Svalbard. Mar. Petrol. Geol. 92, 519–546. https://doi.org/10.1016/j.marpetgeo.2017.11.019.

Hansen, H.N., Løvstad, K., Müller, R., Jahren, J., 2017. Clay coating preserving high porosities in deeply buried intervals of the Stø Formation. Mar. Petrol. Geol. 88, 648–658. https://doi.org/10.1016/j.marpetgeo.2017.09.011.

Heald, M., Larese, R., 1974. Influence of coatings on quartz cementation. J. Sediment. Petrol. 44, 1269–1274. https://doi.org/10.1306/212F6C94-2B24-11D7-8648000102C1865D.

Helle, H.B., Bhatt, A., Ursin, B., 2001. Porosity and permeability prediction from wireline logs using artificial neural networks: a North Sea case study. Geophys. Prospect. 49, 431–444. https://doi.org/10.1046/j.1365-2478.2001.00271.x.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9, 90–95.

Kiaei, H., Sharghi, Y., Ilkhchi, A.K., Naderi, M., 2015. 3D modeling of reservoir electrofacies using integration clustering and geostatistic method in central field of Persian Gulf. J. Petrol. Sci. Eng. 135, 152–160. https://doi.org/10.1016/j.petrol.2015.08.019.

Klausen, T.G., Müller, R., Poyatos-Moré, M., Olaussen, S., Stueland, E., 2019. Tectonic, provenance and sedimentological controls on reservoir characteristics in the upper triassic–middle jurassic realgrunnen Subgroup, SW Barents Sea. Geological Society, London, Special Publications 495, SP495–2018. https://doi.org/10.1144/SP495-2018-165. -2165.

Klausen, T.G., Müller, R., Sláma, J., Olaussen, S., Rismyhr, B., Helland-Hansen, W., 2018. Depositional history of a condensed shallow marine reservoir succession: stratigraphy and detrital zircon geochronology of the Jurassic Stø Formation, Barents Sea. J. Geol. Soc. 175, 130–145. https://doi.org/10.1144/jgs2017-024.

Larionov, V., 1969. Radiometry of Boreholes. Nedra, Moscow, p. 127.

Lim, J.-S., 2005. Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. J. Petrol. Sci. Eng. 49, 182–192. https://doi.org/10.1016/j.petrol.2005.05.005.

Line, L.H., Jahren, J., Hellevang, H., 2018. Mechanical compaction in chlorite-coated sandstone reservoirs – examples from middle – late triassic channels in the southwestern Barents Sea. Mar. Petrol. Geol. 96, 348–370. https://doi.org/10.1016/j.marpetgeo.2018.05.025.

Løvstad, K., Hansen, H.N., Jahren, J., 2022. The porosity preserving effect of basin wide illitic coating in deeply buried sandstone intervals of the lower Jurassic Stø Formation, Barents Sea. Mar. Petrol. Geol. 137, 105498 https://doi.org/10.1016/j.marpetgeo.2021.105498.

Marcussen, Ø., Maast, T.E., Mondol, N.H., Jahren, J., Bjørlykke, K., 2010. Changes in physical properties of a reservoir sandstone as a function of burial depth – the Etive Formation, northern North Sea. Mar. Petrol. Geol. 27, 1725–1735. https://doi.org/10.1016/j.marpetgeo.2009.11.007.

Mckinney, W., 2010. Data structures for statistical computing in python. Proc. 9th Python Sci. Conf. 51–56. Austin, TX.

Ogebule, O.Y., Jahren, J., Mondol, N.H., 2020. Compaction, rock physics and rock properties of sandstones of the Stø Formation: case study of five wells from the southwestern Barents Sea, Norway. Mar. Petrol. Geol. 119, 104448 https://doi.org/10.1016/j.marpetgeo.2020.104448.

Olaussen, S., Dalland, A., Gloppen, T., Johannessen, E., 1984. Depositional environment and diagenesis of Jurassic reservoir sandstones in the eastern part of Troms I area. In: SPENCER, A.M.E.A. (Ed.), Petroleum Geology of the North European Margin. Springer, Dordrecht.

Ozkan, A., Cumella, S.P., Milliken, K.L., Laubach, S.E., 2011. Prediction of lithofacies and reservoir quality using well logs, late cretaceous williams fork formation, mamm creek field, piceance basin, Colorado. AAPG Bull. 95, 1699–1723. https://doi.org/10.1306/01191109143.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Porten, K.W., Warchoł, M.J., Kane, I.A., 2019. Formation of detrital clay grain coats by dewatering of deep-water sands and significance for reservoir quality. J. Sediment. Res. 89, 1231–1249. https://doi.org/10.2110/jsr.2019.65.

Ramm, M., 1991. Diagenesis and Porosity Evolution in Lower and Middle Jurassic Reservoir Sandstone in Troms-I Area, Barents Sea. In: Porosity-depth Trends in Reservoir Sandstones : a Quantitative Study on Effects of Variations in Pore Pressure Temperature and Sandstone Petrography on Reservoir Quality. Dr.Scient Degree, Department of Geology, University of Oslo. *Paper 5.*.

Storvoll, V., Bjørlykke, K., Karlsen, D., Saigal, G., 2002. Porosity preservation in reservoir sandstones due to grain-coating illite: a study of the Jurassic Garn Formation from the Kristin and Lavrans fields, offshore Mid-Norway. Mar. Petrol. Geol. 19, 767–781. https://doi.org/10.1016/S0264-8172(02)00035-1.

Taylor, T.R., Giles, M.R., Hathon, L.A., Diggs, T.N., Braunsdorf, N.R., Birbiglia, G.V., Kittridge, M.G., Macaulay, C.I., Espejo, I.S., 2010. Sandstone diagenesis and reservoir quality prediction: models, myths, and reality. AAPG Bull. 94, 1093–1132. https://doi.org/10.1306/04211009123.

Urang, J.G., Ebong, E.D., Akpan, A.E., Akaerue, E.I., 2020. A new approach for porosity and permeability prediction from well logs using artificial neural network and curve fitting techniques: a case study of Niger Delta, Nigeria. J. Appl. Geophys. 183, 104207.

Van Rossum, G., Drake Jr., F.L., 1995. Python Reference Manual. Centrum voor Wiskunde en Informatica, Amsterdam.

Walderhaug, O., 1994. Temperatures of quartz cementation in Jurassic sandstones from the Norwegian continental shelf; evidence from fluid inclusions. J. Sediment. Res. B Stratigr. Global Stud. 64, 311–323. https://doi.org/10.1306/D4267D89-2B26-11D7-8648000102C1865D.

Walderhaug, O., 1996. Kinetic modeling of quartz cementation and porosity loss in deeply buried sandstone reservoirs. AAPG (Am. Assoc. Pet. Geol.) Bull. 80, 731–745. https://doi.org/10.1306/64ED88A4-1724-11D7-8645000102C1865D.

Wooldridge, L.J., Worden, R.H., Griffiths, J., Utley, J.E., 2017. Clay-coated sand grains in petroleum reservoirs: understanding their distribution via A modern AnalogueL. J. wooldridge et al. CLAY-COATED sand grains in petroleum reservoirs. J. Sediment. Res. 87, 338–352. https://doi.org/10.2110/jsr.2017.20.

Worden, R., Griffiths, J., Wooldridge, L., Utley, J., Lawan, A., Muhammed, D., Simon, N., Armitage, P., 2020. Chlorite in sandstones. Earth Sci. Rev. 204, 103105 https://doi.org/10.1016/j.earscirev.2020.103105.