Seismic characterization of a Triassic-Jurassic deep geothermal sandstone reservoir, onshore Denmark, using unsupervised machine learning techniques

Satinder Chopra[1], Ritesh Kumar Sharma[1], Kenneth Bredesen[2], Thang Ha[3], and Kurt J. Marfurt[3]

Abstract

The Triassic-Jurassic deep sandstone reservoirs in onshore Denmark are known geothermal targets that can be exploited for sustainable and green energy for the next several decades. The economic development of such resources requires accurate characterization of the sandstone reservoir properties, namely, volume of clay, porosity, and permeability. The classic approach to achieving such objectives has been to integrate well-log and prestack seismic data with geologic information to obtain facies and reservoir property predictions in a Bayesian framework. Using this prestack inversion approach, we can obtain superior spatial and temporal variations within the target formation. We then examined whether unsupervised facies classification in the target units can provide additional information. We evaluated several machine learning techniques and found that generative topographic mapping further subdivided intervals mapped by the Bayesian framework into additional subunits.

Introduction

The estimated hydrocarbon reserves around the world when produced can keep us going for the next several decades, but scientific records and our own experiences are enough evidence that climate change is indeed happening. Attempts at addressing it are suggestive of energy extraction from nonfossil fuels. One such resource is the natural heat of the earth or geothermal energy.

There are different ways in which the heat of the earth can be used. We hear of natural hot springs at certain places, where somehow, groundwater is emerging through the porous and fractured rocks after making contact with the deeper and hotter layers of the earth's crust. Hot water geysers spout columns of hot water and steam through vents in earth's surface. Under suitable conditions, the geothermal system in place can be enhanced to our advantage. For instance, a fluid circulation cycle could be set up by injecting (pumping) cold water through a well to the depth of say a hot sandstone reservoir rock and be drawn up as hot water through another well a certain distance away. Of course, such an initiative requires the right kind of rocks through which a steady water-flow rate can be established. Such geothermally heated water (usually >75°C) is being used for heating buildings constructed in their proximity, in which hot water from the producing well transfers the heat to the housing heating grid.

The feasibility and success of such a geothermal reservoir are dependent on finding the candidate reservoir rock that will allow the water to percolate through. This would need good porosity and permeability, the presence or absence of faults and fractures, high enough temperature, and knowledge of the structural component of the target reservoir.

The Danish subsurface hosts low-enthalpy reservoirs (40°C–80°C) of Jurassic, Triassic, and Cretaceous age. The geothermal energy has the potential of supplying district heating for hundreds of years into the future, and three geothermal plants have been set up in Denmark. The Thisted geothermal plant commissioned in 1984 (Figure 1) supplies saline water at 43°C from the Gassum reservoir at a depth of 1.25 km to 2000 residential units with district heating. The Margretheholm plant, in operation since 2005, supplies saline water at 48°C from the Bunter sandstone reservoir at a depth of 2.6 km and has the capacity for heating 4500 residential units. The third plant, Sønderborg, began operation in 2013 and pumps water from the Gassum sandstone at a depth of 1.2 km. Each of the plants has one injection well and one production well producing heat from the sandstone reservoirs through heat exchanger pumps.

Bredesen et al. (2020) and Feng et al. (2020) discuss the seismic characterization of a Triassic-Jurassic deep geothermal sandstone reservoir, north of Copenhagen, in an area near Hillerød in northeast Zealand, onshore

[1]SamiGeo, Calgary, Alberta T3L 1W3, Canada. E-mail: satichop@gmail.com (corresponding author); riteshpes@gmail.com.
[2]Geological Survey of Denmark and Greenland (GEUS), Copenhagen 1350, Denmark. E-mail: kenb@geus.dk.
[3]The University of Oklahoma, Norman, Oklahoma 73109-1009, USA. E-mail: ha7675@ou.edu; kmarfurt@ou.edu.

Manuscript received by the Editor 22 April 2021; revised manuscript received 26 May 2021; published ahead of production 17 June 2021.
This paper appears in *Interpretation*, Vol. 9, No. 4 (November 2021); p. 1–10, 15 FIGS.
http://dx.doi.org/10.1190/INT-2021-0091.1.

Denmark (Figure 1). The data available for this study were a 2D seismic survey from 2013 (comprising five profiles with 3 km offset, designed for structural mapping, and outlining potential geothermal reservoirs), a local well (Karlebo-1A), and a few regional wells that penetrate the reservoir of interest. The more recent geologic and petrophysical analysis were also collated and made use of for impedance inversion and rock-physics modeling (Nielsen et al., 2004; Weibel et al., 2017).

The target geothermal interval is the sandstone-dominated Upper Triassic-Lower Jurassic Gassum Formation, which is being exploited for geothermal production and storage in Denmark (Kristensen et al., 2016; Vosgerau et al., 2017). Figure 2 displays seismic profile 5 with geologic interpretation with the Karlebo-1A well projection overlaid on it. The Lower Jurassic sandstone unit that overlies the Gassum Formation is a secondary geothermal target. Both these units lie at a depth of approximately 2 km below the ground level. Above the Lower Jurassic sandstone unit is the Fjerritslev Formation that is dominated by marine mudstones and shales, which is the regional cap rock. The Lower Cretaceous sandstone unit sits on top of the Fjerritslev Formation, and, in turn, it is overlaid by the high-velocity chalk formation that generates interfering multiples and converted waves, which makes processing of the seismic data challenging. Below the Gassum Formation is the impermeable mudstones of the Vinding, Oddesund, and other older formations (Weibel et al., 2017). The observations from Karlebo-1A well indicate that although the Lower Jurassic reservoir unit (LJRU) is a homogeneous unit, the Gassum sandstone contains interlacing of thinly bedded shale. The reservoir temperature ranges between 50°C and 65°C in the target interval.

The exploration well Karlebo-1A drilled for hydrocarbon exploration is located approximately 140 m from profile 5. It has a limited number of log curves (gamma ray, sonic, and porosity), whereas a nearby well, Margretheholm-1A, penetrates the same set of formations as the Karlebo-1A well and contains a complete set of log curves. Thus, the latter well was used to derive empirical relations between pairs of variables and was used to determine additional curves such as the density, shear sonic, and shale volume for the Karlebo-1A well.

Bredesen et al. (2020) demonstrate the seismic reservoir characterization of these different lithounits by carrying out prestack simultaneous impedance inversion and predicting facies and reservoir properties in a Bayesian framework. Their results demonstrate that several porous and clean water-bearing sandstones are potential high-quality geothermal reservoirs within the two target layers, namely, the LJRU and the Gassum Formation. Feng et al. (2020) obtain similar reservoir quality predictions using the same seismic inversion data as input to a new system of artificial neural networks-hidden Markov models. In general, the LJRU exhibits a more promising reservoir quality in terms of high porosity, permeability, and low shale content compared to the Gassum Formation. Given the availability and quality of the data and geologic complexities, the results are influenced by high uncertainty, but they highlighted the possible target layers.

We decided to repeat the prestack simultaneous impedance inversion on the same data by following a somewhat different workflow, and instead of the Bayesian classification for facies prediction, we made use of the available unsupervised machine learning (ML) techniques for facies classification and assess their comparison. Our results indicate a superior impedance inversion product

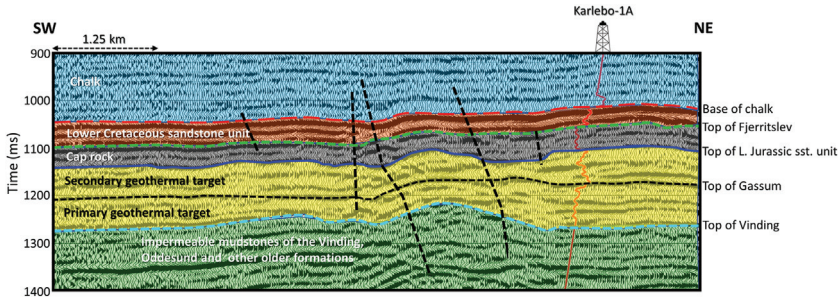

Figure 2. Seismic line 5 shown overlain by the main geologic units with the projected location of the Karlebo-1A well. Yellow indicates the primary and secondary geothermal targets.

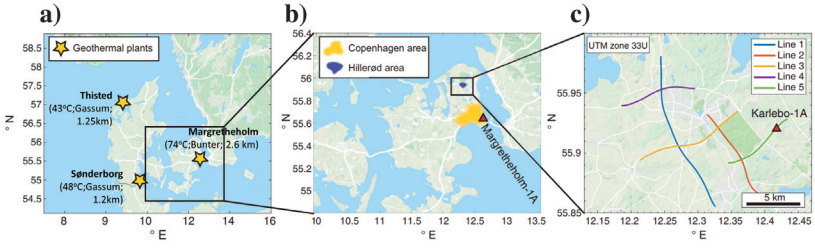

Figure 1. (a) The location of the three geothermal plants that are operational in Denmark where the map indicates that the temperature and depth of the saline water filled the sandstone reservoir for each plant. (b) The location of the study area magnified in northeastern Zealand. (c) The location of five seismic 2D lines and the Karlebo-1A well used in the study. Additional regional well data are available at Margretheholm, located approximately 30 km away from the prospect area. Courtesy of Google maps. Modified from Bredesen et al. (2020).

such as P- and S-impedance, which, along with other seismic attributes, were used for ML facies prediction.

We begin with a description of the workflow followed for the prestack simultaneous impedance inversion, interspersed with examples, and then we go over to the description of the individual unsupervised ML techniques as well as their applications.

Well-log correlation

Figure 3 illustrates the correlation of P-velocity, gamma-ray curves, and the synthetic seismogram with stacked seismic data. A zero-phase wavelet (shown on the top) was estimated from the seismic data using a statistical process. A reasonably good correlation is noticed. Four horizons corresponding to the base of chalk, top of Fjerritslev, top of Lower Jurassic sandstone unit, and top of Gassum are shown in Figure 3, although the horizon top of Vinding (below the Gassum Formation) was also picked and is shown in some of the subsequent figures.

Prediction of shear curve

A frequently encountered situation is when not many wells have shear sonic log curves available.

It is desirable to have a full suite of curves for a well falling on the 2D seismic profile being inverted because it can help in the generation of an accurate low-frequency model for impedance inversion, as well as for carrying out any neural network analysis for computation of a reservoir property. As stated above, a shear log curve was available in Margretheholm-1A but not in well Karlebo-1A. The linear relationship between the P- and S-impedance for the Margretheholm-1A well was used to predict the shear curve for the Karlebo-1A well, but it was not found to be satisfactory. It was then generated by using multiattribute analysis and checked for its accuracy by plotting the P-impedance against V_P/V_S. Figure 4 shows the crossplot in which the data points from both wells

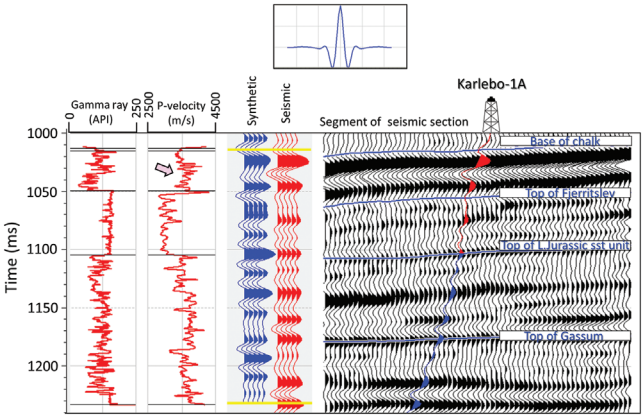

Figure 3. Correlation of the Karlebo-1A well-log curves with the seismic data. The blue traces represent the synthetics (generated with the wavelet shown above), whereas the red traces represent the seismic data. There is a reasonably good correlation coefficient of ρ = 0.66 between the synthetic and red traces in the time window indicated by the yellow bars.

are seen to overlap, an indication that the shear wave (S-wave) prediction can be considered reliable.

Low-frequency trend determination for impedance inversion

Because the basic information on subsurface geologic structure is contained in the low-frequency impedance component, any lack in its determination can degrade the quantitative prediction of properties based on seismic impedance inversion (Lindseth, 1979; Chopra and Sharma, 2017). It therefore becomes essential to build an accurate low-frequency model to have confidence in the impedance inversion and subsequently make quantitative predictions therefrom (Sams and Saussus, 2013). Generation of reliable low-frequency models over large 3D seismic volumes in which spatial variations in rock properties exists could be challenging. But for the exercise at hand, because we were going to invert a single 2D seismic profile, we kept the method simple. The Margretheholm-1A well was projected on this line, and well-to-seismic tie analysis was carried out to establish a time-depth relationship for this well. The low-frequency trends from this well were extrapolated over the complete 2D profile, constrained with the available horizons. By following this approach, the Karlebo-1A well could now be treated as a blind well.

Having determined the low-frequency models for compressional wave (P-wave) and S-wave impedance, the next step is to carry out preconditioning of the prestack data for enhancing its signal-to-noise ratio.

Preconditioning of seismic data

The prestack seismic data were conditioned carefully to make sure that amplitudes are preserved such that their variation with offset/angle could be used in a meaningful way. The specific processes used in the conditioning were supergathering (3 × 3), band-pass filtering, random noise attenuation, and trim statics, with difference plots taken at each step to ensure that no useful

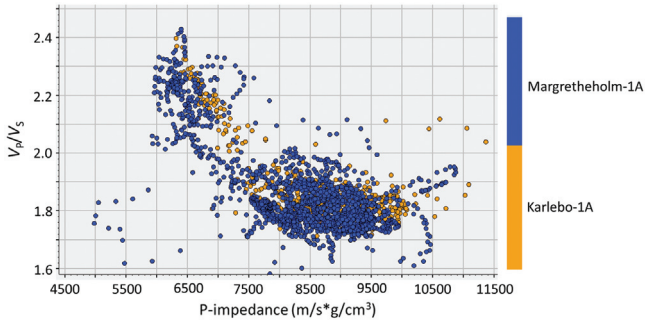

Figure 4. Crossplot of P-impedance versus V_P/V_S ratio from the Karlebo-1A and Margretheholm-1A wells. The S-wave velocity curve was predicted at the Karlebo-1A well using multiattribute analysis (Hampson et al., 2001; Pramanik et al., 2004). Note that the data points from the Karlebo-1A wells overlap those from the Margretheholm-1A well, indicating that the S-wave prediction is reliable.

signal was distorted or attenuated (Hunt et al., 2015; Chopra and Sharma, 2016).

Simultaneous impedance inversion

In simultaneous prestack inversion, multiple partial-offset or angle substacks are simultaneously inverted. For each angle stack, a unique wavelet is estimated. Subsurface low-frequency models for P-impedance, S-impedance, and density constrained with appropriate horizons in the broad zone of interest are constructed using the approach explained above. The models, wavelets, and partial stacks were used as the input in the inversion, and the output was the P-impedance and S-impedance. The angle range selected was 0°–32°.

Before impedance inversion is carried out on the whole seismic profile, generally, inversion is carried out at the location of the well, so that the inverted impedance traces can be compared with the P- and S-impedance well logs. This step is usually referred to as inversion analysis. In Figure 5a and 5b, the inversion results in red are shown overlaid on the equivalent log traces in blue, for both wells, as well as the equivalent low-frequency model traces in black. The high correlation between the individual traces indicates that the inversion is acceptable. One prominent deviation is indicated with the pink block arrows (Figure 5a), which is also seen in Figure 3, in which the seismic amplitude seems to be somewhat stronger than the synthetic amplitude at that location. The inverted impedance trace can be used to create a synthetic trace (red in Figure 6a), which again seems to match the real seismic traces well (black in Figure 6b), as seen by their computed difference, which is low as shown in Figure 6c. These encouraging results prompted carrying out the inversion over the complete seismic profile.

Figure 7 shows the seismic profile after stacking of the preconditioned gathers, with the overlay of the P-impedance curve from the Karlebo-1A well and the five horizons picked on it. The equivalent inverted P- and S-impedance sections are shown in Figure 7c and 7d, with the respective well logs overlaid as color strips as well as the horizons. The good correlation between the inversions and their respective logs lends confidence to their accuracy. Good lateral resolution and variation is noticed on both impedance sections. Between common mid point 2400–2500, there is a zone influenced by some vertical striping noise. This artifact also influences the subsequent results. It is suspected to be related to lateral variations in the fold of the seismic data because there were some areas along the seismic line in which the use of vibrators was not allowed close to local residents.

Once the P- and S-impedance are derived from prestack seismic data, other attributes can be derived therefrom. Such attributes can be useful for identifying different lithofacies. Bredesen et al. (2020) illustrate the use of P-impedance and V_P/V_S in classifying different facies. To follow a similar approach, we made a crossplot of P-impedance against V_P/V_S for both wells (Karlebo-1A and Margretheholm-1A) color coded with V_{clay} and porosity as shown in Figure 8. After examining the clusters of data points and the range of values of V_{clay} and porosity that they spanned, three ellipses were drawn to capture the data points as shown. The points within the three ellipses were then back projected onto the well logs to check where they were coming from.

The points enclosed by the gray color are associated with high V_{clay} and low porosity, so they can be considered as coming from a clay-rich formation, which is evident as per their back projection on well-log curves. Similarly, the points enclosed by the greenish and yellowish ellipses can be considered as coming from shaly sand and porous sand formations, respectively. With this information being established, these ellipses can be mapped on inverted attributes as shown in Figure 9. Figure 9a shows a crossplot of inverted P-impedance versus V_P/V_S over our zone of interest in which different ellipses correspond to the mapped lithologies. The resemblance of this crossplot with the crossplots shown in Figures 4 and 8 lends confidence in the inversion

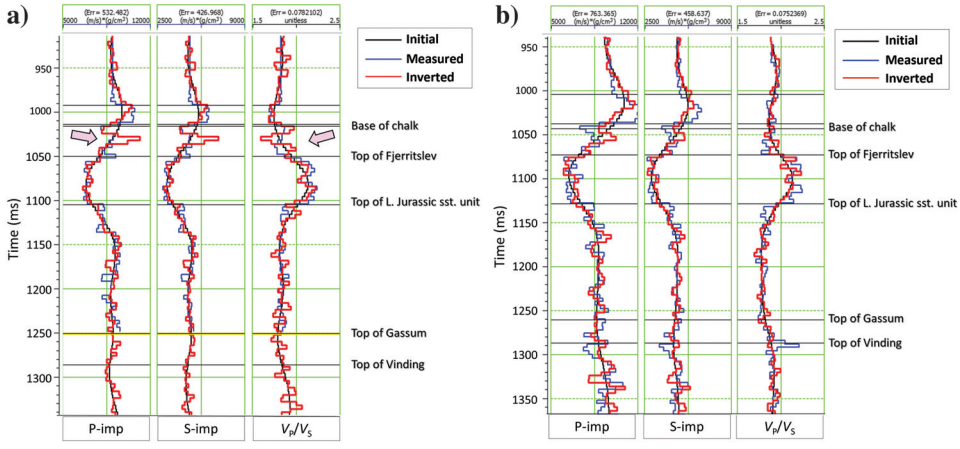

Figure 5. Inversion analysis carried out at the (a) Karlebo-1A and (b) Margretheholm-1A wells. The inverted trace in red has an overall good match with the measured well-log traces in blue. Note the prominent misfit indicated with the pink block arrows where in Figure 3 we see that the seismic amplitude seems to be somewhat stronger than that of the corresponding synthetic amplitude.

process. Figure 9b shows the spatial mapping of different lithofacies on the vertical section through back projection of the cluster points within the ellipses. Notice that although the cap rock is mainly clay-rich, the shaly sand and porous sand facies exist in the primary and secondary targets, which is expected as per the classification made on well-log curves. In addition, the results imply a higher proportion of porous sands toward the southwest. Although the analysis discussed so far seems to be a repetition of workflow followed by Bredesen et al. (2020), our inversion results and spatial distribution of different lithofacies appear to be more striking. As mentioned earlier, the purpose of this study was not only to repeat the previous exercise, but to seek additional value by way of ML applications for facies classification.

Bayesian classification

The characterization of a reservoir in terms of lithology variation from seismic data being an inversion problem could probably yield different models for the same input seismic response. Some of these outcomes could be more probable than others. The Bayesian facies classification is a probabilistic approach that allows combining a priori information about a model and data measurements. Thus, we can follow an approach that accounts for the uncertainties associated with reservoir characterization in the different lithounits (Chopra et al., 2019). This work follows the Bayesian classification approach (Grana, 2013) and provides a facies model reflecting the quality of the lithounits and a related uncertainty analysis. Bredesen et al. (2020) use this approach and selected weighted mean and standard deviation as statistical estimators to provide a reasonable representation of the reservoir predictions.

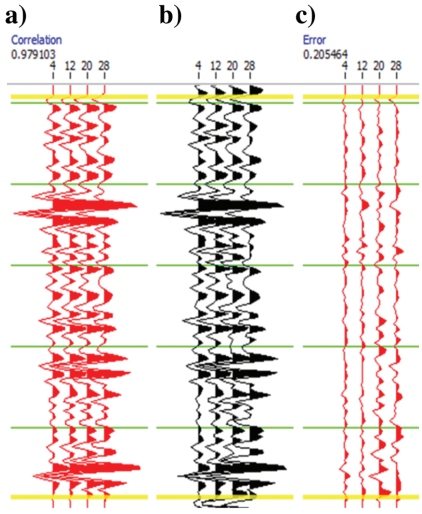

Figure 6. (a) Synthetic trace created by using the inverted P-impedance trace at the location of the Karlebo-1A well repeated four times shown in red, (b) the real seismic traces at the well location, and (c) the error or difference of (a and b). Because the error is small, the inversion can be taken as accurate.

When using Bayesian classification, Bredesen et al. (2020) define three facies based on the cutoff values on a crossplot of acoustic impedance against V_P/V_S well curves, and the probability distributions for each of these facies are represented by kernel-density-estimated probability density functions (PDFs) (Grana, 2018). Once the acoustic impedance and V_P/V_S attributes were derived from simultaneous impedance inversion, Bredesen et al. (2020) compute the facies using the PDFs.

Unsupervised ML facies classification

Deterministic techniques such as the Bayesian classification of rock-physics parameters directly relate the seismic response to known rock properties. In contrast, the petrophysical correlation of the seismic response to fluid flow units is in general unknown. Unsupervised learning provides a means to determine if the seismic response can be related to flow units or rock types that can be calibrated with additional well control, but for which we do not understand the underlying petrophysical or geologic theoretical support. Still, seismic interpreters

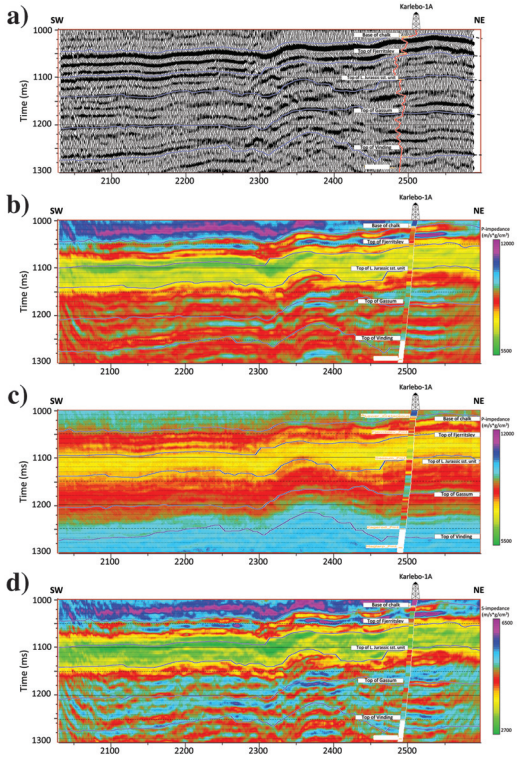

Figure 7. (a) Seismic line 5 with the P-impedance well log overlaid, (b) the equivalent P-impedance section with the P-impedance well log colored strip overlaid, (c) the equivalent P-impedance section from the inversion carried out by Bredesen et al. (2020) with the P-impedance well log colored strip overlaid, and (d) the equivalent S-impedance section with S-impedance well log colored strip overlaid. In addition to the reasonably good lateral variation seen on the sections, the correlation of the two impedance logs with the inverted sections shown in (b) seems to be exceptionally good. All five horizons picked on the seismic amplitude section are also shown overlaid.

face a perpetual challenge of extracting heterogeneous seismic facies on different generated attributes. The common analysis tools include corendering, crossplotting, and visualization, which can help to an extent in terms of simultaneous display of the input attributes. The data reduction approach resorted to at times applies mathematical techniques to reduce the number of attributes to a manageable subset. Clustering is another way to identify elements within the data that have similar expressions. In this study, we compare the application of some established ML techniques, namely, principal component analysis (PCA), independent component analysis (ICA), K_{means}, Gaussian mixture models (GMMs), self-organizing maps (SOMs), and generative topographic mapping (GTM). We find such an application promising because the facies results exhibit higher vertical and lateral resolution than the supervised Bayesian classification. Some applications of unsupervised ML applications for facies classification have been discussed earlier by Chopra and Marfurt (2018) and Chopra et al. (2019) for case studies from the Barents Sea and the Delaware Basin, US, respectively. Below, we briefly describe the ML techniques and their application to the geothermal sandstone reservoir in Denmark, which has been discussed previously.

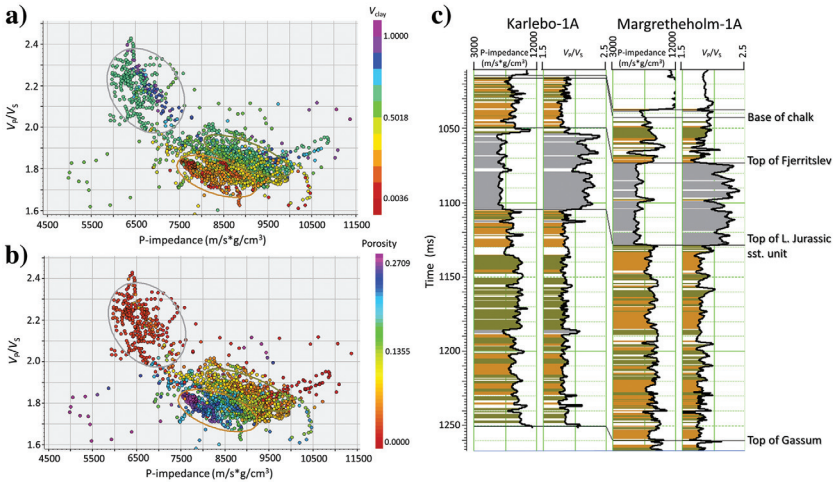

Figure 8. Crossplot of P-impedance versus V_P/V_Scolor coded with (a) V_{clay} and (b) porosity using data for both wells, Karelbo-1A and Margretheholm-1A. The three ellipses in gray, orange, and olive are defined by the different values of V_{clay} and porosity indicating different rock types. (c) Back projection of three colored facies enclosed within the ellipses onto the well curves for the two wells.

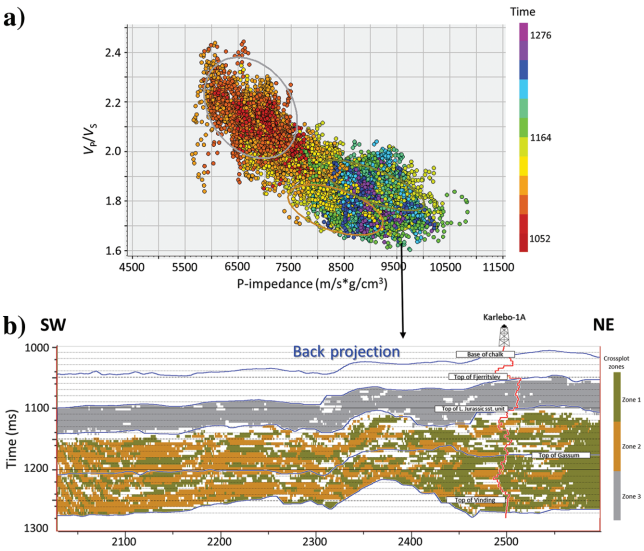

Figure 9. Crossplot of (a) P-impedance versus V_P/V_Susing well data from both wells and (b) inverted P-impedance versus V_P/V_S. (c) Back projection of cluster points enclosed within the ellipses seen on the inverted data crossplot onto the vertical section. The gamma-ray log curve has been overlaid on the section.

PCA

PCA aims to identify patterns in the input attributes by detecting correlation between them. If a strong correlation exists between some of them, then those attributes can be lumped together. Thus, PCA is a useful dimensionality reduction tool and assumes that the input seismic attributes exhibit a Gaussian distribution.

Many of our seismic attributes are coupled through the underlying geology, for example, a fault could give rise to lateral changes in the waveform, dip, peak frequency, and amplitude. Less desirably, there are attributes such as the alternative measure of coherence (Barnes, 2007) or a suite of closely spaced spectral components, which may be coupled mathematically. While using the seismic attributes, the amount of redundancy can be measured with the covariance matrix. Before we compute the covariance matrix, because different attributes may have different units or are unitless (V_P/V_S, Poisson's ratio, and coherence), they need to be normalized, for which Z-score normalization is usually adopted. In an $N \times N$ covariance matrix, the element C_{mn} is the crosscorrelation between the mth and nth scaled attributes over the interval or volume of interest. Once the covariance matrix is formed, to be able to gauge information on the magnitude and direction of variation, it needs to be decomposed into eigenvectors and eigenvalues. The attributes used in the present exercise are the instantaneous amplitude, spectral magnitude (40 Hz), P-impedance, V_P/V_S, lambda-rho, and porosity. All of these attributes are seismic amplitude-derived through prestack simultaneous impedance inversion or otherwise, which expectedly should furnish information on the rock types better than some of the other attributes.

By convention, the first step is to order the eigenvalues from the highest to the lowest. The eigenvector with the highest eigenvalue is the first principal component of the data set (PCA1); it represents the vector representing the maximum variance in the data and thereby the bulk of the information that would be common in the attributes

used. The eigenvector with the second-highest eigenvalue, called the second principal component (PC2), exhibits lower variance and is orthogonal to PCA1. PCA1 and PCA2 will lie in the plane that represents the plane of the data points. Similarly, the third principal component (PC3) will lie in a plane orthogonal to the plane of the first two principal components. Because seismic attributes are correlated through the underlying geology and the band limitations of the source wavelet, the first two or three principal components will almost always represent much of the data variability. The input seismic attributes used for the present exercise comprised the P-impedance, S-impedance, instantaneous amplitude, porosity, lambda-rho, and spectral component at 45 Hz.

In Figure 10a, 10b, and 10c, we show section displays for seismic profile 5 for the PCA-1, PCA-2, and PCA-3 components, respectively. Overlaid in black are the five picked horizons on the section as well as the P-impedance well log for the Karlebo-1A well. We see different-colored patches on the display, which are a representation of the different facies in the data. Figure 10d exhibits the three individual displays corendered together using the red, green, blue (RGB) color scheme. This display is particularly useful in that the facies information contained in each of the three principal components can be conveniently interpreted on a single display.

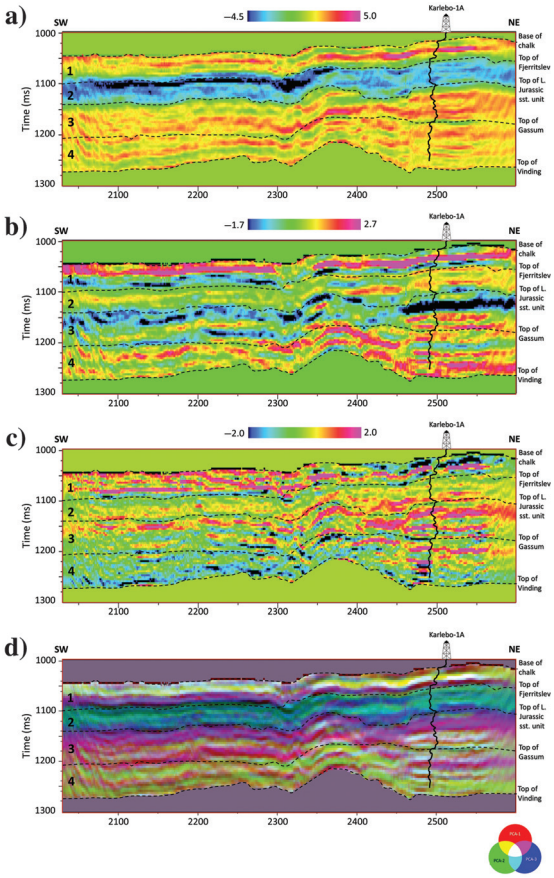

Figure 10. Section display for seismic profile 5 for the (a) PCA1, (b) PCA2, and (c) PCA3 components. (d) Equivalent display from the PCA1, PCA2, and PCA3 corendered using RGB blending. The P-impedance well log for well Karlebo-1A is shown overlaid on each section. Of the four rock types marked to the left of the composite section, not only does rock type 2 stand out in a different color pertaining to the shale facies, but the variation in facies can also be seen in units 3 and 4 as well.

ICA

ICA is another ML technique that classifies the different input seismic attributes into independent components, but it does not require them to have a Gaussian distribution. Besides this, the other difference between ICA and PCA is that the independent components are not orthogonal (Chopra et al., 2018; Lubo-Robles, 2018). Thus, from the given set of seismic attributes used as input for the analysis, ICA finds the independent components from that mixture. The problem is cast as a matrix equation, which is then solved using higher order statistics. We show the ICA application to a set of input seismic attributes, in which the generated independent components exhibit better resolution and separation of the geologic features (Chopra and Marfurt, 2019).

Figure 11 shows a section display for seismic profile 5 from the ICA-1, ICA-2, and ICA-3 RGB coblended data. Notice the appearance of the clusters in different colors resembles the cluster patterns obtained from the PCA coblended data display in Figure 10d, except they appear to be somewhat better defined and exhibit better spatial resolution.

K_{means} clustering

The K_{means} clustering approach clusters a given distribution of unlabeled points into a desired number of groups, such that the points within each cluster have greater similarity with one another than the points in another group. The process begins by assigning at random some centroid seed points that serve as centers of the groups that we wish them to form. Each centroid defines a cluster. The distance between each data point

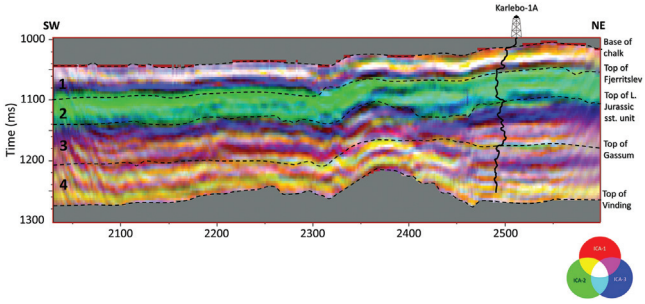

Figure 11. Section display for seismic profile 5 for the ICA1, ICA2, and ICA3 components corendered using RGB blending. Compared to the composite display in Figure 10d, the lateral resolution and contrast appears to be better on this display in each of the four intervals defined by the five horizons shown overlaid. The colors representing the different facies in the intervals also seem to jibe well with the overlaid P-impedance log. The variation in the individual facies in the four intervals seems to be defined somewhat better on this display than the PCA composite display in Figure 10d.

and cluster centroid is calculated. Usually, the Euclidean distance between the points is calculated and a point is considered within a cluster if it is close to the centroid. This way, the points within each cluster get reorganized when the centroids are recalculated based on the reorganized points within each cluster. The last two steps are carried out iteratively until there is no more movement of the centroids when convergence is achieved. The projection of the clusters onto the three principal axes is output from the computation (K_{means}-1, K_{means}-2, andK_{means}-3) rather than the clusters themselves, which can then be combined into a single display with RGB corendering.

Figure 12 shows a section display for seismic profile 5 for K_{means}-1, K_{means}-2, and K_{means}-3 components corendered using RGB blending. Compared with the composite displays in Figures 10d and 11, especially focusing on intervals 3 and 4, the lateral resolution and contrast on this display does not appear to be better. Figure 11 representing the corendering of ICA attributes still takes the lead in terms of clarity.

GMM

GMM is a technique for carrying out semiparametric density estimation, in which the PDF is represented as a weighted sum of normal or Gaussian distributions (Wallet et al., 2014). It could also be viewed as a probabilistic approach to clustering, in which each Gaussian term in the model may represent a class and the weights representing the a priori probabilities of the different classes.

Figure 13 shows a section display for seismic profile 5 for unsupervised facies classification generated using GMM application, in which the data were classified into 12 clusters. The display looks like the K_{means} clustering display shown in Figure 12, with some enhanced lateral resolution in some pockets in different intervals.

SOMs

The SOM is another unsupervised ML technique that generates a seismic facies map from multiple seismic attributes, similar to the K_{means} clustering approach. But it differs from the K_{means} in that it first defines the initial cluster centroids in an N-dimensional attribute data space by fitting a plane defined by the first two eigenvectors of the covariance matrix to the data in a least-squares sense (Kohonen, 1982, 2001). With the centroid still locked to this plane, it is iteratively deformed into a 2D surface that fits the data even better. Once convergence is reached, the N-dimensional data are projected onto this 2D surface. Thus, SOM may be considered as projection from a multidimensional attribute space to a 2D space. Usually, the output from SOM computation is obtained in the form of two projections on the two SOM axes, which can then be directly cross plotted and displayed using a 2D RGB color bar (Chopra and Marfurt, 2019).

Figure 14 shows a section display for seismic profile 5 for the SOM-1 and SOM-2 cross plotted together using a 2D color bar as shown to the lower right. Some of the clusters seen on this display are better defined than the ones shown earlier from PCA and ICA analysis in Figures 10 and 11 or the K_{means} or GMM clustering displays in Figures 12 and 13.

The 2D color bar visualization may not be available in the interactive interpretation workstations used by interpreters. A convenient way is to use the multiplexed 2D color bar shown alongside. The white space above the color bar deserves an explanation.

The colors that we see on the workstation are stored in the computer in the form of "bit" sizes. A single bit would have two values, 0 and 1, and for a pixel, this

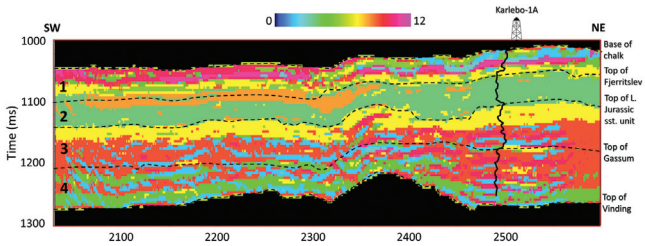

Figure 13. Section display for seismic profile 5 from unsupervised facies classification generated using the Gaussian mixture models application.

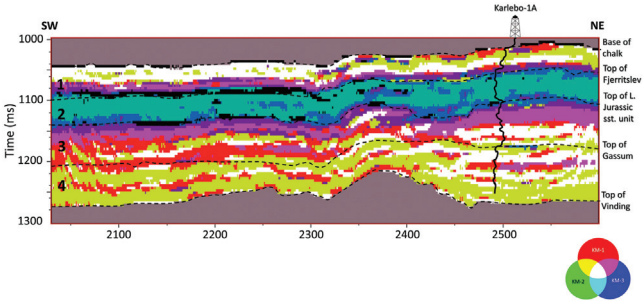

Figure 12. Section display for seismic profile 5 for K_{means}-1, K_{means}-2, and K_{means}-3 components corendered using RGB blending. Compared with the composite displays in Figures 10d and 11, the lateral resolution and contrast does not appear to be better on this display, especially focusing on intervals 3 and 4.

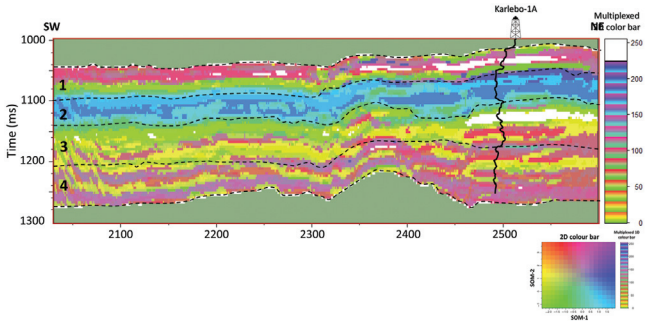

Figure 14. Section display for seismic profile 5 from the SOM1 and SOM2 cross plotted using a 2D color bar. As compared with the earlier methods, besides the better spatial resolution in intervals 3 and 4, even in interval 2, the resolution is seen to be superior.

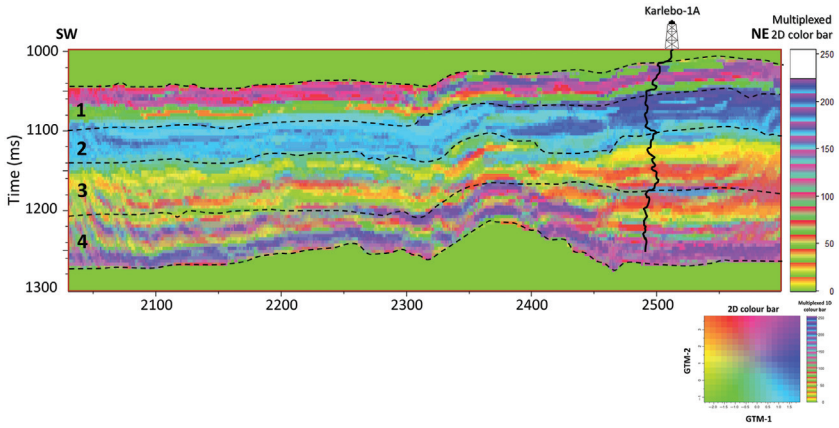

Figure 15. Section display for seismic profile 5 from corendered GTM1 and GTM-2 using a 2D color bar. Multiplexed color bar shown in the above display. This display exhibits the best spatial resolution in all four intervals 1–4 compared with all the other methods discussed in this exercise.

would represent just black and white. Two bits imply that there are four possible combinations (00, 01, 10, and 11), and three bits suggest eight possible combinations (000, 001, 010, 011, 100, 101, 110, and 111). In general, we can say that the number of combinations is two raised to the power of the number of bits. Thus, an eight-bit pixel (or one byte) would represent $2^8 = 256$ possible integer values, which are represented as integers between 0 and 255.

For RGB color blending, we assign six levels of red, green, and blue to three different volumes, resulting in $6^3 = 216$ color levels. Similarly, 3D color bars use six levels for the color on each axis, and when multiplexed into a 2D color bar appears as shown in Figure 14. Thus, there is a mismatch between the two and anything above 216–256 is assigned the white color. Such a mismatch does not result in any loss of resolution on the display.

GTM

Although the Kohonen SOM method described above is easy to implement and is computationally inexpensive, and thus a popular unsupervised clustering approach, it does have limitations. First, there is no theoretical basis for the selection of parameters such as the training radius, neighborhood function, and learning radius because all of these are data-dependent (Bishop et al., 1998; Roy, 2013). Second, no cost function is defined in the method that could be iteratively minimized indicating convergence during the training process. Finally, as a measure of confidence in the final clustering results, no probability density is defined. Thus, an alternative approach to the Kohonen SOM method, called GTM, was developed by Bishop et al. (1998), which overcomes the above-stated limitations. It is a nonlinear dimension-reduction technique that provides a probabilistic representation of the data vectors in latent space (Chopra and Marfurt, 2019).

The GTM method begins with an initial array of grid points arranged on an N-dimensional space or a lower dimensional latent space, for example, the first three principal components or the ICA components. These grid points are then nonlinearly mapped onto an equivalent dimensional non-Euclidean curved surface as a corresponding vector ($\boldsymbol{m}_k$) embedded into different dimensional data space in GTM (Chopra et al., 2019). Each data vector ($\boldsymbol{x}_k$) mapped into this space is modeled as a suite of Gaussian PDFs centered on these reference vectors ($\boldsymbol{m}_k$). The components of the Gaussian model are then iteratively made to move toward the data vector that it best represents (Chopra et al., 2019). The mathematical details of the method as well as its applications have been described by Roy (2013) and Roy et al. (2014).

Thus, as the above descriptions suggest, the PCA, ICA, SOM, and GTM methods project data from a higher dimensional space (8D when eight attributes are used as the input) to a lower dimensional space, which may be a 2D plane or a 2D deformed surface. Once projected onto these planes, the data can be clustered in that space, corendered with RGB or cross plotted using a 2D color bar.

In Figure 15, we show a section display for seismic profile 5 for the GTM-1 and GTM-2 cross plotted together using a 2D color bar as shown to the lower right. This display exhibits the best spatial resolution in all four intervals 1–4 compared with all of the other methods discussed in this exercise. The individually colored patches or facies are crisper and could lead to more accurate interpretations.

Conclusion

We have shown a comparison of seismic facies classification using ML methods such as PCA, ICA, K_{means}, GMM, SOM, and GTM to a seismic profile from Denmark.

In summary, we find that the machine methods hold promise as each of them exhibits more vertical and spatial resolution than a supervised Bayesian classification approach as applied in a previous study on the same dataset. Among ML methods, ICA furnishes more detail than PCA. The SOM and GTM methods provide promising results, with the latter yielding a more accurate definition as seen on the displays.

Acknowledgments

We thank the Geological Survey of Denmark and Greenland (GEUS) for making the seismic data available for the study presented in this paper. The first author would also like to thank the Geomodeling Technology Corporation for making the Attribute Studio software available, which has been used for some visualization displays shown in this paper, the Hampson Russell software for carrying out the impedance inversion presented in this study, as well as the Attribute-Assisted Seismic

Processing and Interpretation Consortium, University of Oklahoma, for access to their software, which has been used for all attribute computation.

Data and materials availability

Data associated with this research are confidential and cannot be released.

References

Barnes, A. E., 2007, Redundant and useless seismic attributes: Geophysics, **72**, no. 3, P33–P38, doi: 10.1190/1.2716717.

Bishop, C. M., M. Svensen, and C. K. I. Williams, 1998, The generative topographic mapping: Neural Computation, **10**, 215–234, doi: 10.1162/089976698300017953.

Bredesen, K., E. Dalgaard, A. Mathiesen, R. Rasmussen, and N. Balling, 2020, Seismic characterization of geothermal sedimentary reservoirs: A field example from the Copenhagen area, Denmark: Interpretation, **8**, no. 2, T275–T291, doi: 10.1190/INT-2019-0184.1.

Chopra, S., D. Lubo-Robles, and K. J. Marfurt, 2018, Some machine learning applications in seismic interpretation: AAPG Explorer, **39**, 22–24.

Chopra, S., and K. J. Marfurt, 2018, Seismic facies classification using some unsupervised machine learning methods: 88th Annual International Meeting, SEG, Expanded Abstracts, 2056–2060, doi: 10.1190/segam2018-2997356.1.

Chopra, S., and R. K. Sharma, 2012, A solid step toward accurate interpretations: AAPG Explorer, **33**, 36–37.

Chopra, S., and R. K. Sharma, 2016, Preconditioning of seismic data prior to impedance inversion: AAPG Explorer, **37**, 30–33.

Chopra, S., R. K. Sharma, and K. J. Marfurt, 2019, Unsupervised machine learning facies classification in the Delaware Basin and its comparison with supervised Bayesian facies classification: 89th Annual International Meeting, SEG, Expanded Abstracts, 2619–2623, doi: 10.1190/segam2019-3214088.1.

Feng, R., N. Balling, and D. Grana, 2020, Lithofacies classification of a geothermal reservoir in Denmark and its facies-dependent porosity estimation from seismic inversion: Geothermics, **87**, 101854, doi: 10.1016/j.geothermics.2020.101854.

Grana, D., 2013, Bayesian inversion methods for seismic reservoir characterization and time-lapse studies: Ph.D. thesis, Stanford University.

Grana, D., 2018, Joint facies and reservoir properties inversion: Geophysics, **83**, no. 3, M15–M24, doi: 10.1190/geo2017-0670.1.

Hampson, D., J. Schuelke, and J. Quirein, 2001, Use of multiattribute transforms to predict log properties from seismic data: Geophysics, **66**, 220–236, doi: 10.1190/1.1444899.

Hunt, L., P. Cary, D. Tican, M. Perz, R. K. Sharma, N. Nagarajappa, X. Li, and S. Chopra, 2015, VOQC: Value oriented quality control in seismic processing: CSEG Recorder, **40**, 42–51.

Kohonen, T., 1982, Self-organized formation of topologically correct feature maps: Biological Cybernetics, **43**, 59–69, doi: 10.1007/BF00337288.

Kohonen, T., 2001, Self-organizing maps: Springer-Verlag.

Kristensen, L., M. L. Hjuler, P. Frykman, M. Olivarius, R. Weibel, L. H. Nielsen, and A. Mathiesen, 2016, Pre-drilling assessments of average porosity and permeability in the geothermal reservoirs of the Danish area: Geothermal Energy, **4**, 1–27, doi: 10.1186/s40517-016-0048-6.

Lindseth, R. O., 1979, Synthetic sonic logs — A process for stratigraphic interpretation: Geophysics, **44**, 3–26, doi: 10.1190/1.1440922.

Lubo-Robles, D., 2018, Development of independent component analysis for reservoir geomorphology and unsupervised seismic facies classification in the Taranaki Basin, New Zealand: M.S. thesis, University of Oklahoma, Norman.

Nielsen, L. H., A. Mathiesen, and T. Bidstrup, 2004, Geothermal energy in Denmark: Geological Survey of Denmark and Greenland Bulletin, **4**, 17–20, doi: 10.34194/geusb.v4.4771.

Pramanik, A. G., V. Singh, R. Vig, A. K. Srivastava, and D. N. Tiwari, 2004, Estimation of effective porosity using geostatistics and multiattribute transforms: A case study: Geophysics, **69**, 352–372, doi: 10.1190/1.1707054.

Roy, A., 2013, Latent space classification of seismic facies: Ph.D. dissertation, The University of Oklahoma, 212.

Roy, A., A. S. Romero-Pelaez, T. J. Kwiatkowski, and K. Marfurt, 2014, Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico: Interpretation, **2**, no. 1, SA31–SA47, doi: 10.1190/INT-2013-0077.1.

Sams, M., and D. Saussus, 2013, Practical implications of low frequency model selection on quantitative interpretation results: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3118–3122, doi: 10.1190/segam2013-0660.1.

Vosgerau, H., U. Gregersen, L. Kristensen, S. Lindström, A. Mathiesen, C. M. Nielsen, M. Olivarius, and L. H. Nielsen, 2017, Towards a geothermal exploration well in the Gassum Formation in Copenhagen: Geological Survey of Denmark and Greenland Bulletin, **38**, 29–32, doi: 10.34194/geusb.v38.4393.

Wallet, B. C., P. Roderick, R. C. Altimar, and R. M. Slatt, 2014, Unsupervised classification of $\lambda\rho$-$\mu\rho$ attributes derived from well log data in the Barnett Shale: 84th Annual International Meeting, SEG, Expanded Abstracts, 1594–1598, doi: 10.1190/segam2014-1586.1.

Weibel, R., M. Olivarius, C. Kjøller, L. Kristensen, M. L. Hjuler, H. Friis, P. K. Pedersen, A. Boyce, M. S. Andersen, E. Kamla, L. O. Boldreel, A. Mathiesen, and L. H. Neilsen, 2017, The influence of climate on early and burial diagenesis of Triassic and Jurassic sandstones from the Norwegian–Danish Basin: The Depositional Record, **3**, 60–91, doi: 10.1002/dep2.27.