# Impact of spectral balancing of seismic data on some unsupervised facies analysis

**SATINDER CHOPRA**
SAMIGEO, CALGARY, SATINDER.CHOPRA@SAMIGEO.COM
**RITESH KUMAR SHARMA**
SAMIGEO, CALGARY, RITESH.SHARMA@SAMIGEO.COM
**KURT J. MARFURT**
THE UNIVERSITY OF OKLAHOMA, NORMAN
**HEATHER BEDLE**
THE UNIVERSITY OF OKLAHOMA, NORMAN
**ALEXANDRO VERA-ARROYO**
THE UNIVERSITY OF OKLAHOMA, NORMAN

## Abstract

The interpretation of discrete stratigraphic features on seismic data is limited by its bandwidth and its signal-to-noise ratio. Unfortunately, well-resolved reflections from the top and base of subtle stratigraphic geologic boundaries occur only for features thick enough to be imaged by the bandlimited data. In contrast, seismically thin stratigraphic features approaching a quarter-wavelength in thickness give rise to composite, or "tuned," seismic reflections. Data conditioning to balance the seismic spectrum provides significant improvement not only in vertical, but also in lateral resolution of the seismic data. Spectral balancing enhances the frequency content of the seismic data and preserves the tuning features and amplitudes and leads to more accurate definition of the features of interest. We show that attributes computed from spectrally-balanced data better delineate the finer features in the zones of interest. Some attribute combinations can be effectively combined for seismic facies classification using unsupervised machine learning applications including $k_{means}$ clustering, self-organizing mapping (SOM) and generative topographic mapping (GTM). We find that improving the data bandwidth through spectral balancing improves not only the resolution of attributes by themselves, but also when combined using machine learning.

## Introduction

A common problem with seismic data is their relatively low bandwidth. Significant efforts are made during processing to enhance the frequency content of the data as much as possible to provide a spectral response that is consistent with the acquisition parameters. Traditional

seismic data are seen to preserve information with a frequency content going up to 60 or 70 Hz at the high end of the bandwidth. While such bandwidths may be acceptable for thicker conventional reservoirs, they could lack the needed resolution for thinner reservoirs or to map thinner architectural elements within a thicker reservoir. The advancements in seismic data acquisition and processing, coupled with computer capacities and speeds, provide cost effective solutions for such objectives.

There are several methods that are used during processing to enhance the frequency content of the input seismic data. Here we mention a few commonly used processes that can help the interpreter to extract meaningful information from seismic data. We follow that up with a method of choice and demonstrate its application to a vintage seismic dataset from the northern North Sea.

*Time-variant spectral whitening*: One of the oldest and most widely used methods is to use time-variant spectral whitening (TVSW). The method involves passing the input data through a number of narrow band-pass filters and determining the decay rates for each frequency band. The inverse of these decay functions for each frequency band is applied and the results are summed. In this way, the amplitude spectrum for the output data is whitened in a time-variant way. The number of filter bands, the width of each band and the overall bandwidth of application are the different parameters that are used and adjusted for an optimized result (Yilmaz, 2001). In this method, the high-frequency noise is usually amplified and so a band-pass filter must be applied to the resulting data. Since it is a trace-by-trace process, TVSW is not appropriate for AVO applications.

*Inverse Q-filtering*: If we had an analytic form for an attenuation function, it would then be easy to compensate for its effects. Thus, in practice, attempts are first made to estimate a Q-model for the subsurface. Inverse Q-filtering compensates for the time-variant absorption effects, thereby broadening the effective seismic bandwidth by correcting the loss of high-frequency signal. The physics of attenuation is complicated, with contributions from scattering from fractures and rugose surfaces, scattering from patchy saturation, the squirt mechanism of energy loss, and others. These attempts have met with a varying degree of success, depending on the assumptions used in the approach and how well they are met in practice.

*Frequency-split structurally oriented filtering*: Helmore (2009) introduced frequency split structurally oriented filtering wherein the input seismic data is divided into a number of frequency bands, followed by running structurally oriented filters separately to each of the bands and then recombining the results. This procedure reduces noise in selected frequency bands and results in a higher signal-to-noise ratio as well as enhanced resolution. Structurally oriented filters do not suffer from windowing artifacts and are precisely adapted to the local dip (Helmore, 2009).

In order to apply the spectral balancing method to seismic data described in the next section, we selected the 3D seismic data from Smeaheia area in offshore Norway. Smeaheia has been considered as one of the potential areas for $CO_2$ storage and evaluation.

The Smeahiea area lies about 30 km east of the Troll gas field (Figure 1), within the Norwegian continental shelf. The Smeahiea target is located in a fault block bounded by the Vette Fault to the west and the Øygarden Fault to the east and is raised about 300 m relative to the Troll field. The Late Jurassic Sognefjord, Fensfjord, and Krossfjord formations form the producing reservoir zones in the Troll gas field.

In the Smeaheia block, there are two four-way structural closures (Lauritsen et al., 2018), the Alpha structure to the west and the Beta structure to the east. Two exploration wells, namely 32/4-1 and 32/2-1 have been drilled into these structures, and although the reservoir is good, both wells turned out to be dry, indicating that the Smeaheia area is not charged with hydrocarbons.

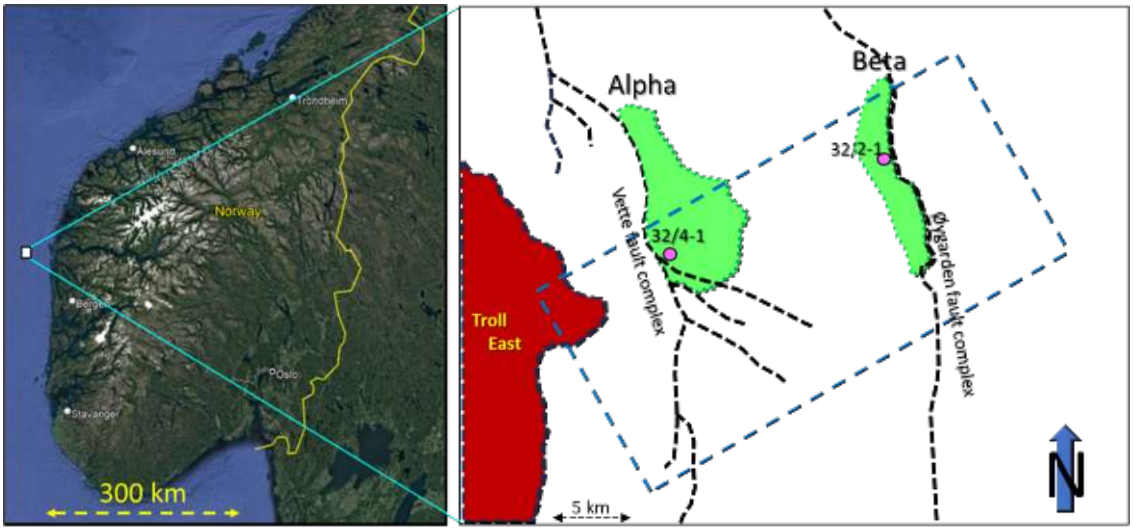



**Figure 1:** The Smeaheia area on the Norwegian continental shelf with Alpha and Beta structures indicated. The image to the left was prepared with the use of Google Earth Pro. *(Modified after Furre et al., 2017)*

In the Smeaheia area, the Sognefjord Formation is the primary reservoir, consisting of medium-to-coarse-grain, well-sorted, micaceous, and minor argillaceous sandstone. Below this formation lies the Fensfjord Formation, consisting of medium-grained, well-sorted sandstone with shale intercalations. Underlying the Fensfjord Formation is the Krossfjord Formation, with medium-to-coarse-grained, well-sorted sandstone.

Overlying the Sognefjord Formation are shales of the Heather and Draupne formations. While the Heather Formation is comprised of silty claystone with thin streaks of limestone interfingering the Sognefjord, Fensfjord and Krossfjord sandstones, the Draupne Formation

consists of dark grey to brown/black shale that is non-calcareous, carbonaceous, and fissile claystone. Both the Heather and Draupne formations serve as primary seals for the proposed $CO_2$ storage reservoir sandstones of the Sognefjord, Fensfjord and Krossfjord formations.

As the Sognefjord, Fensfjord and Krossfjord are sandstone reservoir formations, there is concern about evaluation of their properties for thicknesses that fall below seismic resolution, which could come from shale and carbonate stringers within these zones. Finally, the existence of faults/fractures that fall below seismic resolution could provide pathways for fluid losses. All these risks need to be evaluated and mitigated in the context of long-term $CO_2$ storage.

### Available seismic data

The available seismic data were the GN1101 3D survey covering the Smeaheia (blue dashed rectangle shown in Figure 1) acquired by Gassnova in 2011 and made publicly available by Gassnova and Equinor. The bin size for the data is 12.5 x 25 m, with a sample interval of 4 ms. Gassnova provided interpreted horizons and well log data for the two 32/4-1 and 32/2-1 wells along with well completion reports. These logs consisted of complete gamma ray curves, but with sonic and density logs that were not recorded for the shallower depths. The seismic data volume is of good quality.

## Spectral balancing of seismic data for attribute analysis

For a decade-old vintage seismic dataset, we need to apply an amplitude-friendly poststack spectral balancing procedure to enhance the vertical and lateral resolution. The spectral balancing procedure of choice was the method first discussed by Marfurt and Matos (2014), also demonstrated by Chopra and Marfurt (2016).

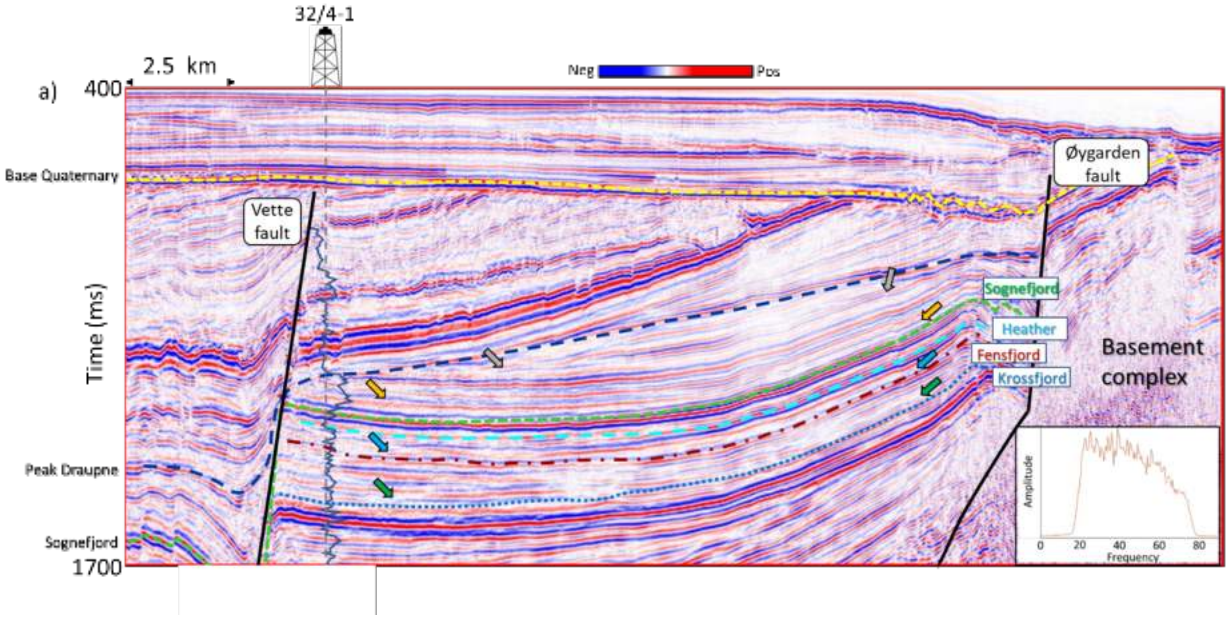
In this method, data are first decomposed into time-frequency spectral components. Then the power of the spectral magnitude, $P(t, f) = [m(t, f)]^2$ is averaged over all the traces ($j = 1, \ldots K$) in the data volume spatially and in the given time window, which yields a smoothed average power spectrum, $P_{avg}(t, f)$. Next, we compute the peak of the average power spectrum at each time sample, $P_{peak}(t) = MAX[P_{avg}(t, f)]$. Both the average power spectrum and the peak of the average power spectrum are used to design a single time-varying spectral-balancing operator that is applied to every trace in the data:

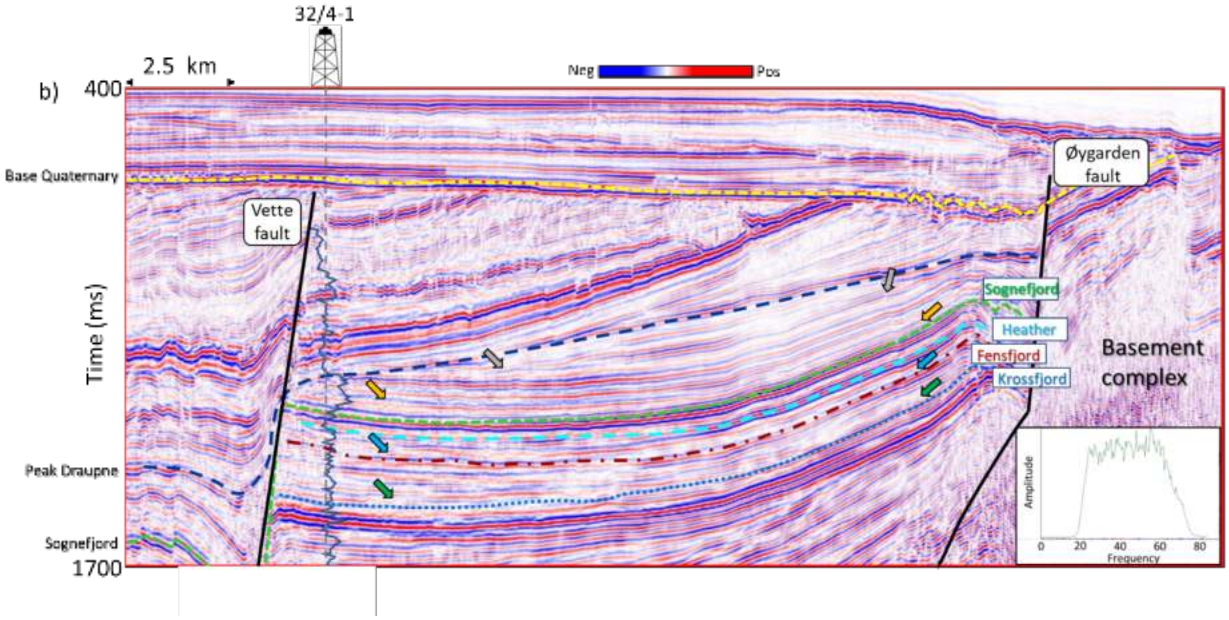
$$m_j^{bal}(t, f) = \left[\frac{P_{peak}(t)}{P_{avg}(t, f) + \varepsilon P_{peak}(t)}\right]^{\frac{1}{2}} m(t, f) \qquad (1)$$

where $\varepsilon$ is the prewhitening parameter. A conservative value would be $\varepsilon = 0.01$ which would constrain amplification of the spectral magnitude to range between 1 and 10. However, as with any filter, the interpreter needs to determine whether such aggressive spectral balancing

introduces ringing in the data. The spectral balancing given by equation 1 is amplitude friendly, as it applies the same time-varying filter to each trace of entire data volume.

Figure 2a and 2b show a comparison of segments of seismic sections traversing the 3D seismic volume before and after spectral balancing, along with their frequency spectra. Notice the well-defined appearance of the reflections in Figure 2b as well as the flattened appearance of the frequency spectra as compared with the input seismic volume.



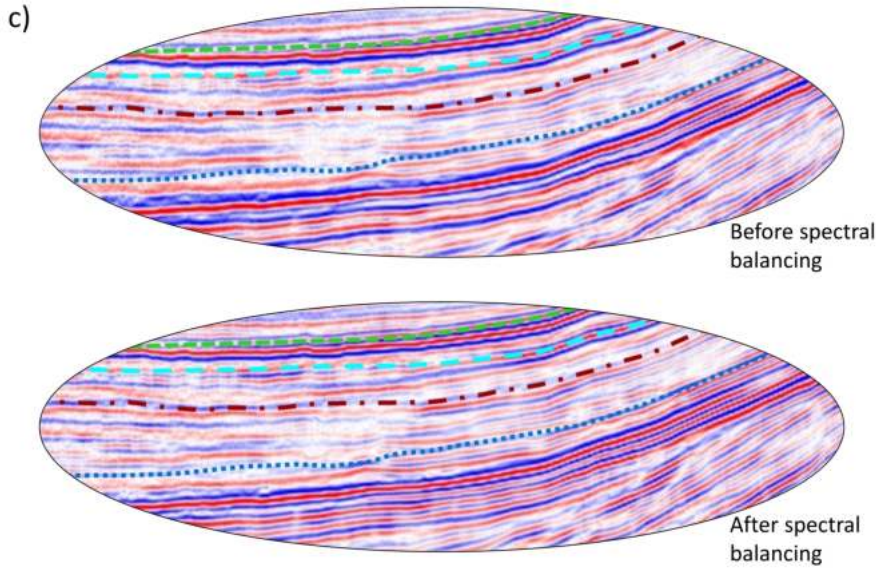



**Figure 2:** Segment of an inline extracted from (a) the input seismic data volume, and (b) the same line after spectral balancing. Some relevant markers as well as the gamma ray curve for well 32/4-1 are overlaid on the two sections. (c) Zoom of the two sections cropped to the area enclosed by the dashed grey oblong shapes on the two images shows clearly the higher frequency content after frequency balancing. The frequency spectra for the two data volumes in the indicated zones are also shown to the right. Notice the enhancement in the resolution of the reflections after spectral balancing, especially as indicated by coloured block arrows in the broad zone of interest.

This data volume was then put through structure-oriented filtering and attribute computation. To bring out the advantage of spectral balancing and structure-oriented filtering, we compute relevant attributes on the data before and after the two data-conditioning processes and compare the results in the next section.

## Generation of a suite of attributes

With the improved vertical resolution seen in Figure 2b, our next task is to determine if the spectrally balanced seismic data also enhances the lateral resolution as measured by seismic attributes. The following attributes were computed on spectrally balanced seismic data.

*Relative acoustic impedance:* Relative acoustic impedance is computed by continuous integration of the original seismic trace with the subsequent application of a low-cut filter. Because it assumes a zero-phase wavelet that is as close to a spike as possible, the improved resolution of spectral balancing will provide improved results over the original data. The impedance transformation of seismic amplitudes enables the transition from reflection interface to interval properties of the data, without the requirement of a low-frequency model.

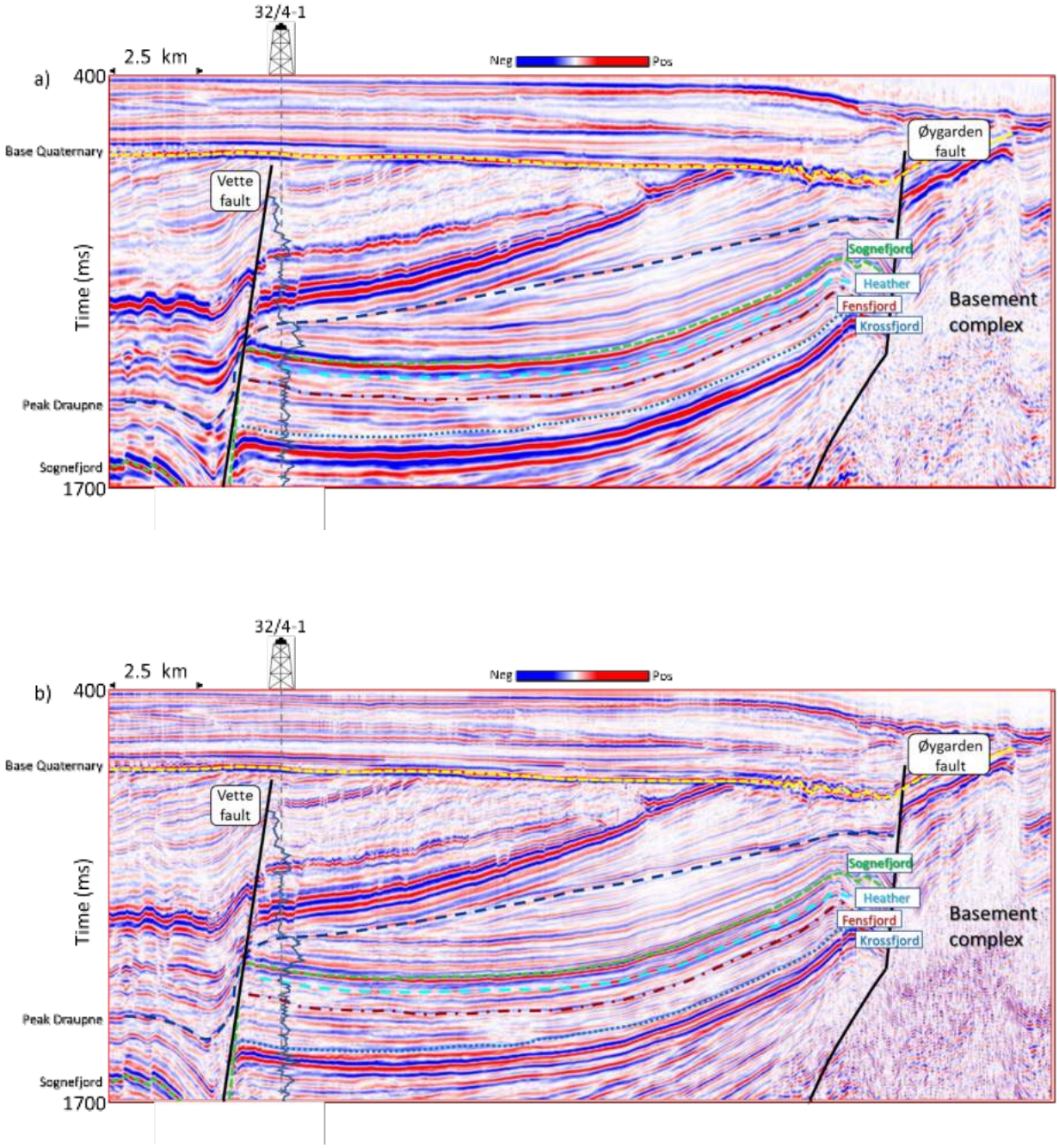



**Figure 3:** Segment of an inline extracted from relative acoustic impedance attribute computed on the (a) input seismic data volume, and (b) input seismic data volume after spectral balancing. Some relevant markers and the gamma ray curve for well 32/4-1 are overlaid on the two sections. Notice the enhancement in the resolution of the reflections after spectral balancing.

Figure 3 shows a comparison of inline sections from relative impedance computed on the input seismic volume (Figure 3a) and the equivalent section from relative impedance computed on spectrally-balanced seismic volume (Figure 3b). Notice the enhanced resolution and better relative impedance definition on the section in Figure 3b. A similar comparison of stratal slices 32 ms above the Sognefjord marker from the relative acoustic impedance attributes computed from input seismic data and input seismic data after spectral balancing is shown in Figure 4. Notice the crisp definition of the faults as indicated by the highlighted areas in dashed purple outlines.

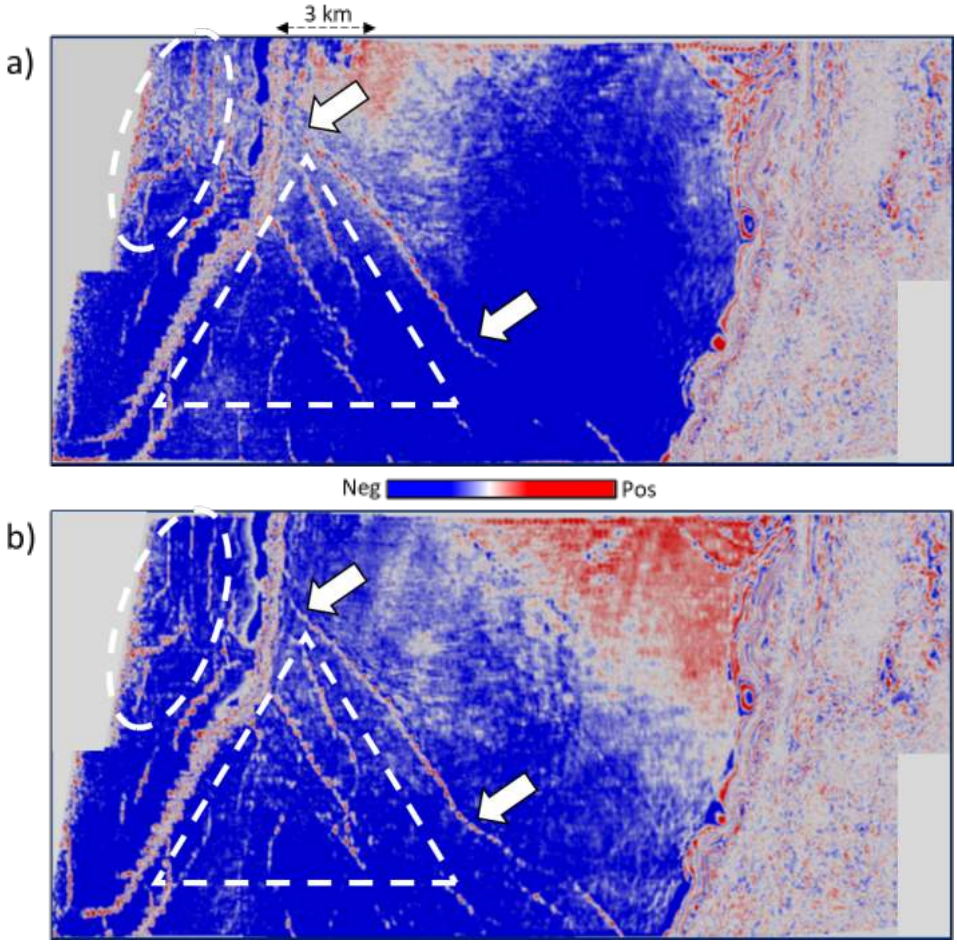



**Figure 4:** Stratal slice 32 ms above the Sognefjord marker through the relative acoustic impedance attributes computed from (a) the original seismic data, and (b) the seismic data after spectral balancing. Notice the crisp definition of the faults as indicated by the white block arrows as well as the two highlighted areas in dashed white outlines and block arrows.

Likewise, the other attributes computed on the two seismic volumes are listed below along with their brief descriptions.

*Instantaneous envelope/frequency*: Instantaneous envelope is a measure of the instantaneous energy of the analytic seismic trace, independent of phase, and provides information on intensity of reflections. Similarly, instantaneous frequency provides information on attenuation and layer thickness. We use a smoother, more stable version of the instantaneous frequency usually obtained by weighting it by the envelope.

*Sweetness*: Sweetness is a "meta-attribute" or one computed from others, which in this case is the ratio of the envelope to the square root of the instantaneous frequency. A clean sand embedded in a shale will exhibit high envelope and lower instantaneous frequency, and thus higher sweetness, than the surrounding shale-on-shale reflections.

*GLCM Energy*: GLCM or grey-level co-occurrence matrix energy is a measure of textural uniformity in the data. If the reflectivity along a horizon is nearly constant, it will exhibit high GLCM energy.

*Spectral magnitude*: The magnitude of spectral components ranging from 20 Hz to 70 Hz, which is the effective bandwidth of the input seismic data.

Specifically, the attributes used for the computation of seismic facies classification using some of the unsupervised machine learning methods were the relative acoustic impedance, envelope, sweetness, GLCM energy and spectral magnitudes at 25 Hz, 40 Hz and 55 Hz.

Two sets of these different attributes were generated – one from the original data and one from the data preconditioned using spectral balancing and structure-oriented filtering. Each dataset will serve as input to unsupervised seismic facies classification using machine learning techniques described in the next section.

## Seismic facies classification using machine learning techniques on input seismic and spectral balanced seismic data

We apply three different unsupervised seismic facies classification machine learning methods to the two attribute data sets: $k_{means}$, *self-organizing mapping* (SOM), and *generative topographic mapping* (GTM). Machine learning uses mathematical operations to learn from the similarities and differences in the provided data and make decisions or predictions. There are two broad families of machine learning algorithms. The first algorithm family includes dimensionality reduction algorithms such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA). When plotted against a 2D color bar, the interpreter may "see" clusters, but the algorithm output is a continuum of data in a lower dimensional space. The second, unsupervised classification algorithm family attempts to explicitly cluster the data into a finite

number of groups that in some metric "best represent" the data provided. $k_{means}$ clustering is one such process. Before the analysis, there is no interpretation assigned to any given group; rather, "the data speak for themselves". However, the choice of input attributes biases the clustering to features of interpretation interest. Biasing the training data to favor geologic features of interest (e.g., by more heavily weighting a bright-spot anomaly) also provides interpreter control of the output. We also show the application of *self-organizing mapping* (SOM) and *generative topographic mapping* (GTM) to the Smeaheia data volume. All the attribute data should first be scaled to account for different units of measurement, mean, and standard deviation. For a non-Gaussian distribution of attribute values, we follow Ha et al. (2021) to further scale the data.

## $k_{means}$ clustering

$k_{means}$ clustering is one of the simplest clustering algorithms and is available in most seismic interpretation software. $k_{means}$ organizes a given distribution of length-$N$ attribute vectors at $R$ voxels, $x_r$, where $r = 1, 2, \ldots R$, into a desired number of $K$ clusters. The clustering process begins by assigning at random $K$ centroids which can serve as centers of the groups we wish to form, where each centroid defines one cluster. Next, the distance between each data point and the centroid of that cluster is calculated. A point may be within a cluster if it is closer to the centroid in that cluster than any other centroid. As some reorganization of the points in different clusters has taken place, the centroids are recalculated for each cluster. These two steps are carried out iteratively until there is no more shifting of the centroids and the process has converged. The calculation of distance between the centroid and the data points referred to above is the traditional Euclidean distance computed from the scaled data, which assumes there is no correlation between the classification variables.

Figure 5 shows a stratal slice comparison 80 ms below the Sognefjord marker, which is within the Sognefjord Formation, for $k_{means}$ clustering generated with five clusters for both the input and spectrally-balanced seismic data. The number five is no magic number, but the maximum number of facies expected at that level. We see a better distribution of coloured patches on the display in Figure 5b, which are a representation of the different facies in the data at that level.

A similar comparison of stratal slices from the $k_{means}$ volume is shown in Figures 6 and 7 at levels within the Fensfjord and Krossfjord formations, and at each level we see a superior distribution of the seismic facies corresponding to the different colours for the $k_{means}$ seismic facies generated on the spectrally-balanced version.

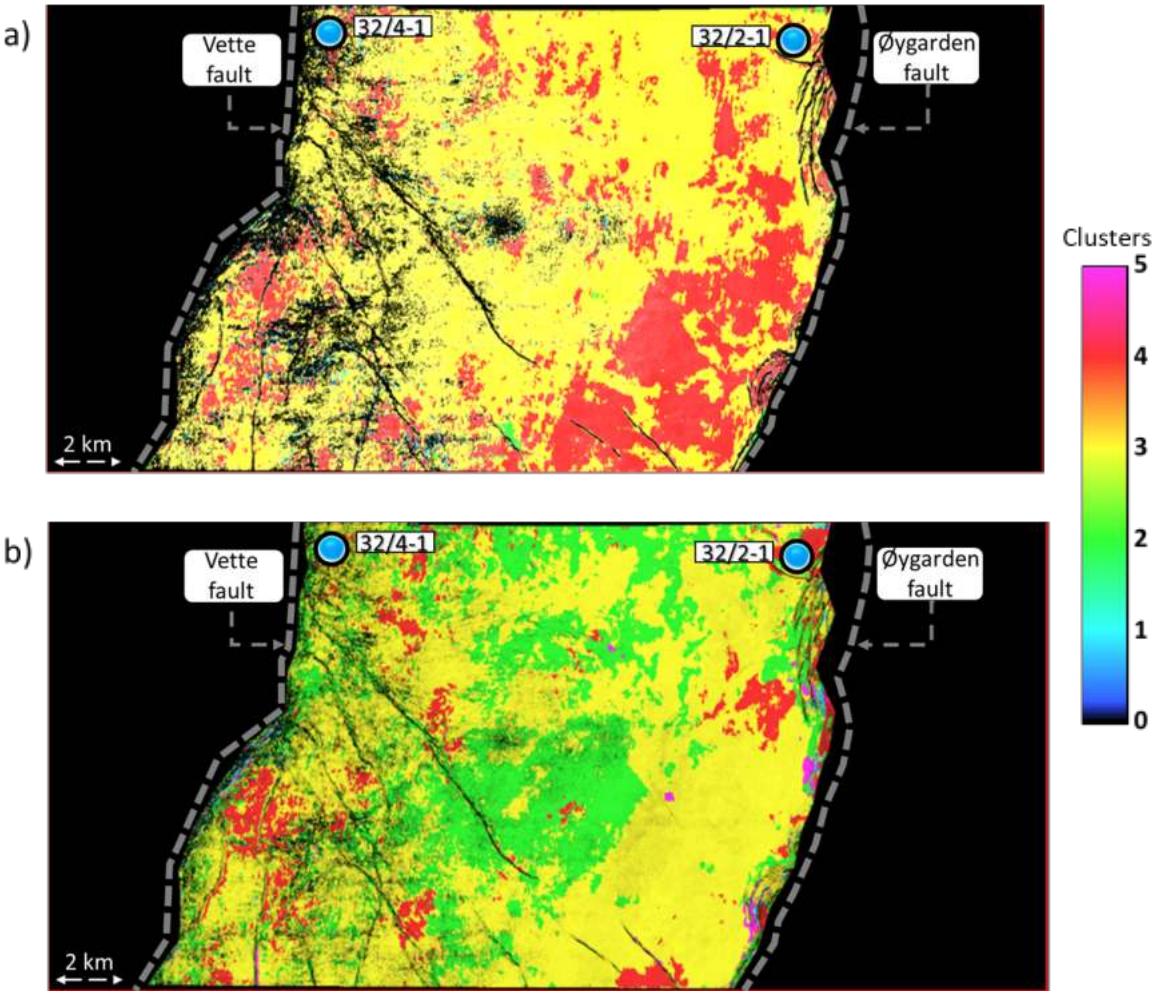



**Figure 5:** Stratal slices 80 ms below the Sognefjord marker (within the Sognefjord Fm) through the $k_{means}$ volume generated from attributes computed from (a) the original seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes (not used in the clustering) to delineate edges. While only two seismic facies (yellow and red) are seen on the display in (a), the display in (b) exhibits three facies (green, yellow, and red) and thus provide better spatial facies resolution. Only the region of interest between the Vette and Øygarden faults was used in the classification in order to minimize the variability in the data and provide maximum resolution in the target zone.

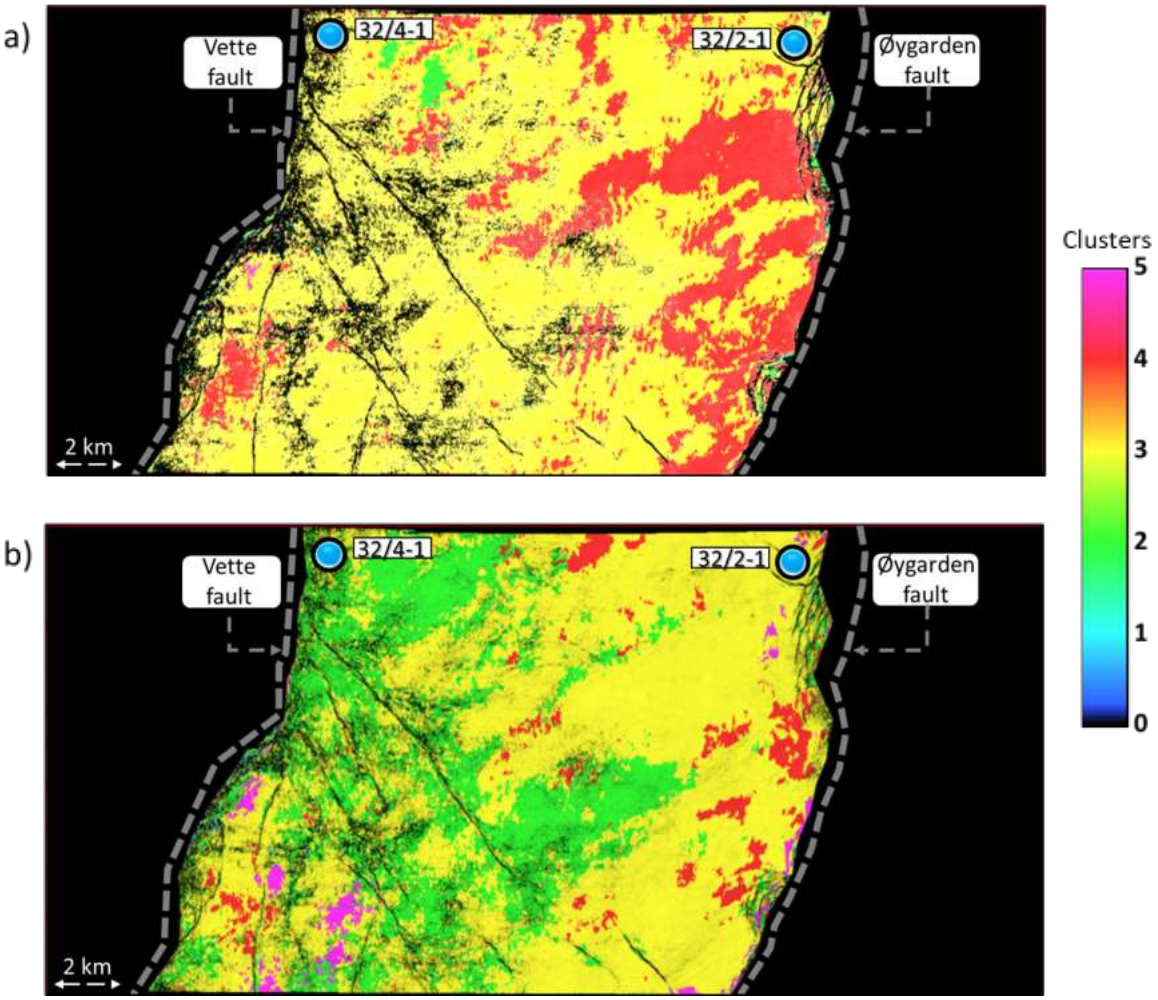



**Figure 6:** Stratal slices at 136 ms below the Sognefjord marker (within the Fensfjord Fm) extracted from the $k_{means}$ volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. While only two seismic facies (yellow and red) are seen on the display in (a), the display in (b) exhibits four facies (green, yellow, purple and red) and thus better spatial facies resolution. Only the region of interest between the Vette and Øygarden faults was used in the classification in order to minimize the variability in the data and provide maximum resolution in the target zone.

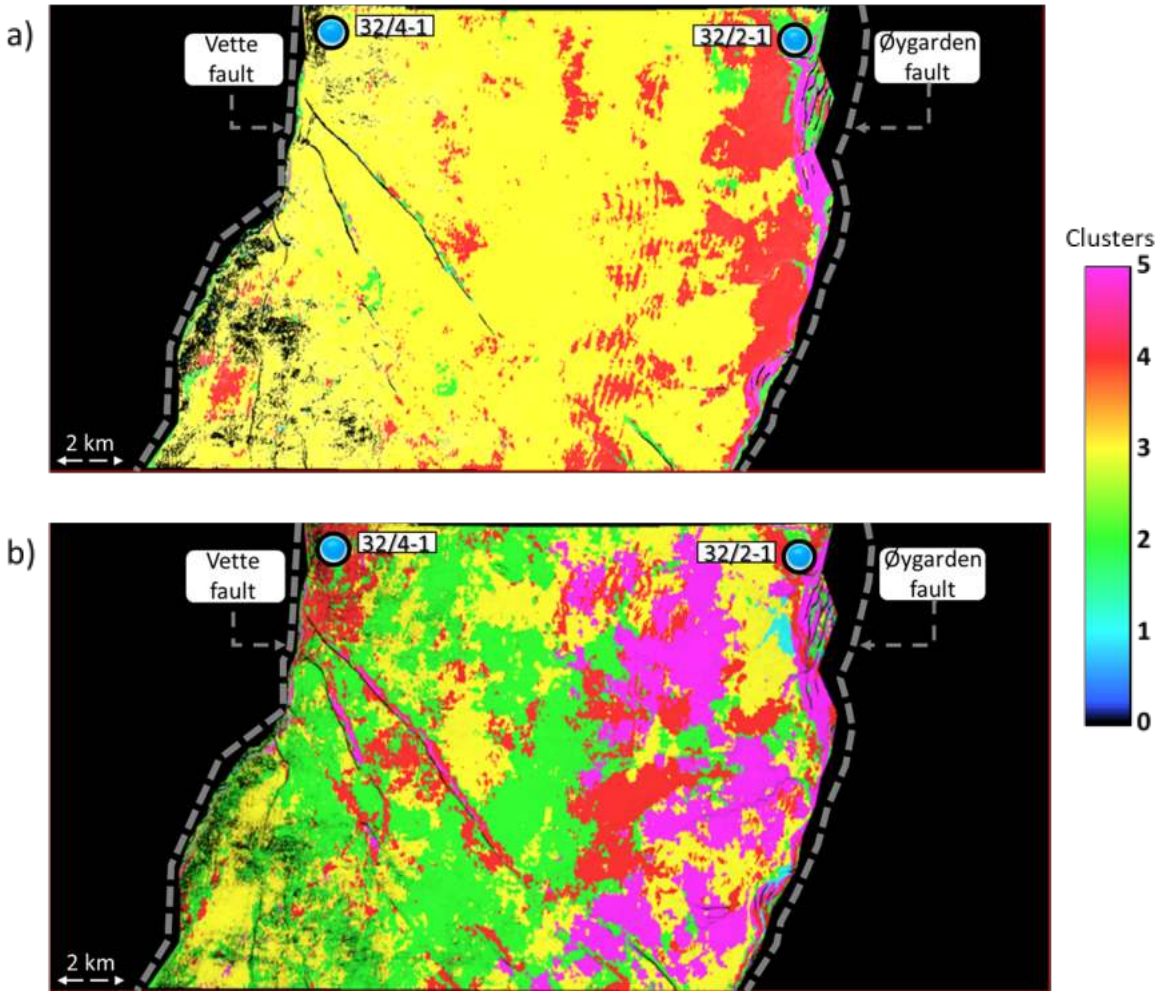



**Figure 7:** Stratal slices at 228 ms below the Sognefjord marker (within the Krossfjord Fm) extracted from the $k_{means}$ volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. While only two seismic facies (yellow and red) are seen on the display in (a), the display in (b) exhibits four facies (green, yellow, purple and red) and thus better spatial facies resolution. Only the region of interest between the Vette and Øygarden faults was used in the classification in order to minimize the variability in the data and provide maximum resolution in the target zone.

## Self-organizing maps

Like $k_{means}$, *self-organizing mapping* (SOM) is a technique that generates a seismic facies map from multiple seismic attributes, again in an unsupervised manner. In contrast to $k_{means}$, SOM defines its initial cluster centroids in an *N*-dimensional attribute data space by least-squares fitting the data with a plane that best fits the data defined by the first two eigenvectors of the covariance matrix (Kohonen, 1982, 2001). This plane with centroids locked to it is then

iteratively deformed into a 2D surface called a manifold that better fits the data. After convergence, the *N*- dimensional data are projected onto this 2D surface, which in turn are mapped against a 2D plane or "latent" (hidden) space defined by axes SOM-1 and SOM-2, onto which the interpreter either explicitly defines clusters by drawing polygons, or implicitly defines clusters by plotting the results against a 2D color bar.

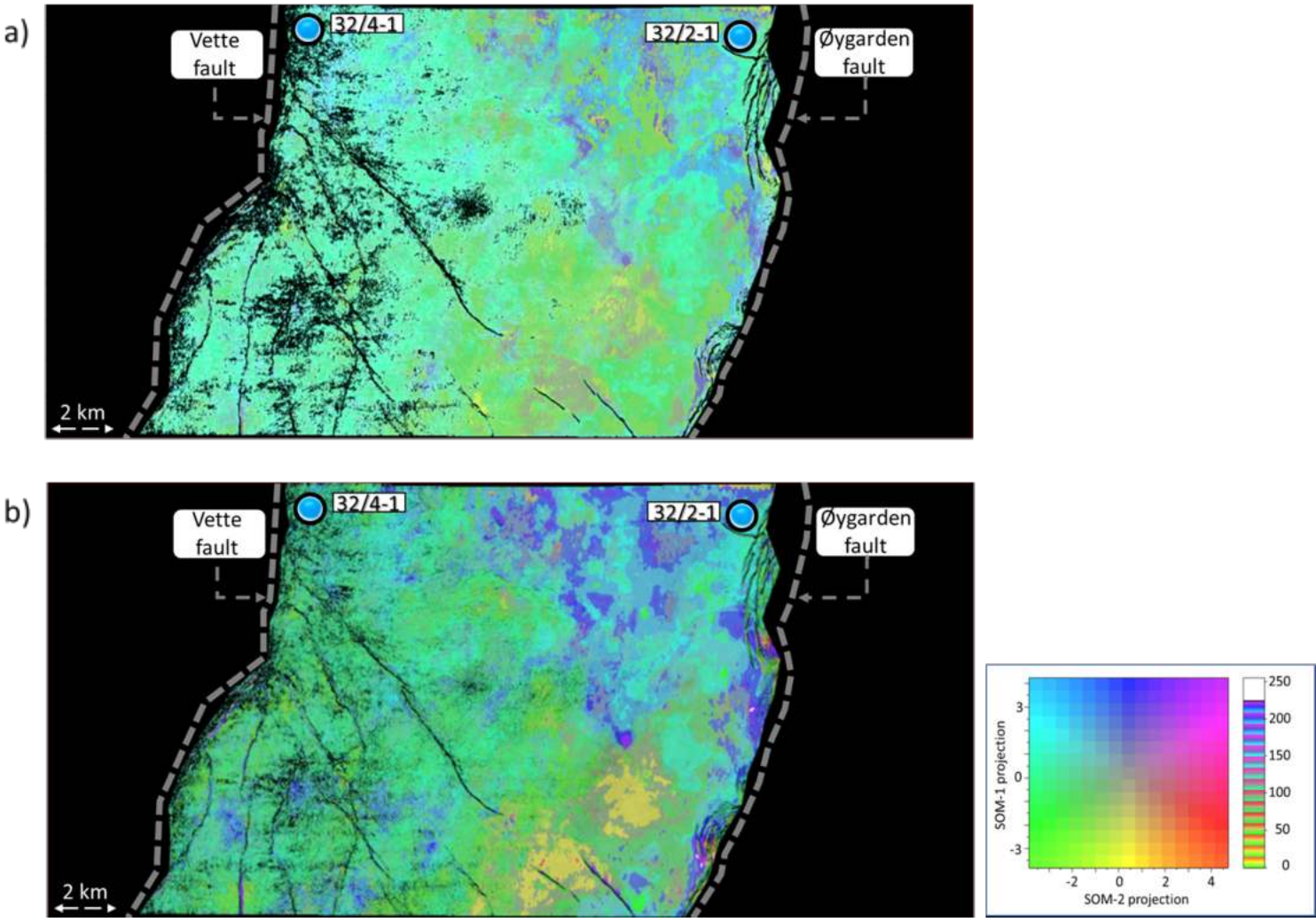



**Figure 8:** Stratal slice at a level within the Sognefjord Formation (80 ms below the Sognefjord marker) extracted from the SOM crossplot volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). As with $k_{means}$, only the target area between the Vette and Øygarden was classified.

Figure 8 shows the equivalent stratal displays (within the Sognefjord formation) extracted from the SOM crossplot volume computed for the input and spectrally-balanced versions of the seismic data, using a 2D color bar or multiplexed into a 1D color bar as shown alongside. Some

of the clusters seen on the display in Figure 8b are better defined than the ones shown in Figure 8a, or the equivalent $k_{means}$ clustering displays shown in Figure 5.

A similar comparison of stratal slices from the SOM crossplot volume is shown in Figures 9 and 10 at levels within the Fensfjord and Krossfjord formations respectively, and at each level we see a superior distribution of the seismic facies corresponding to the different colours for the SOM seismic facies generated on the spectrally-enhanced version.

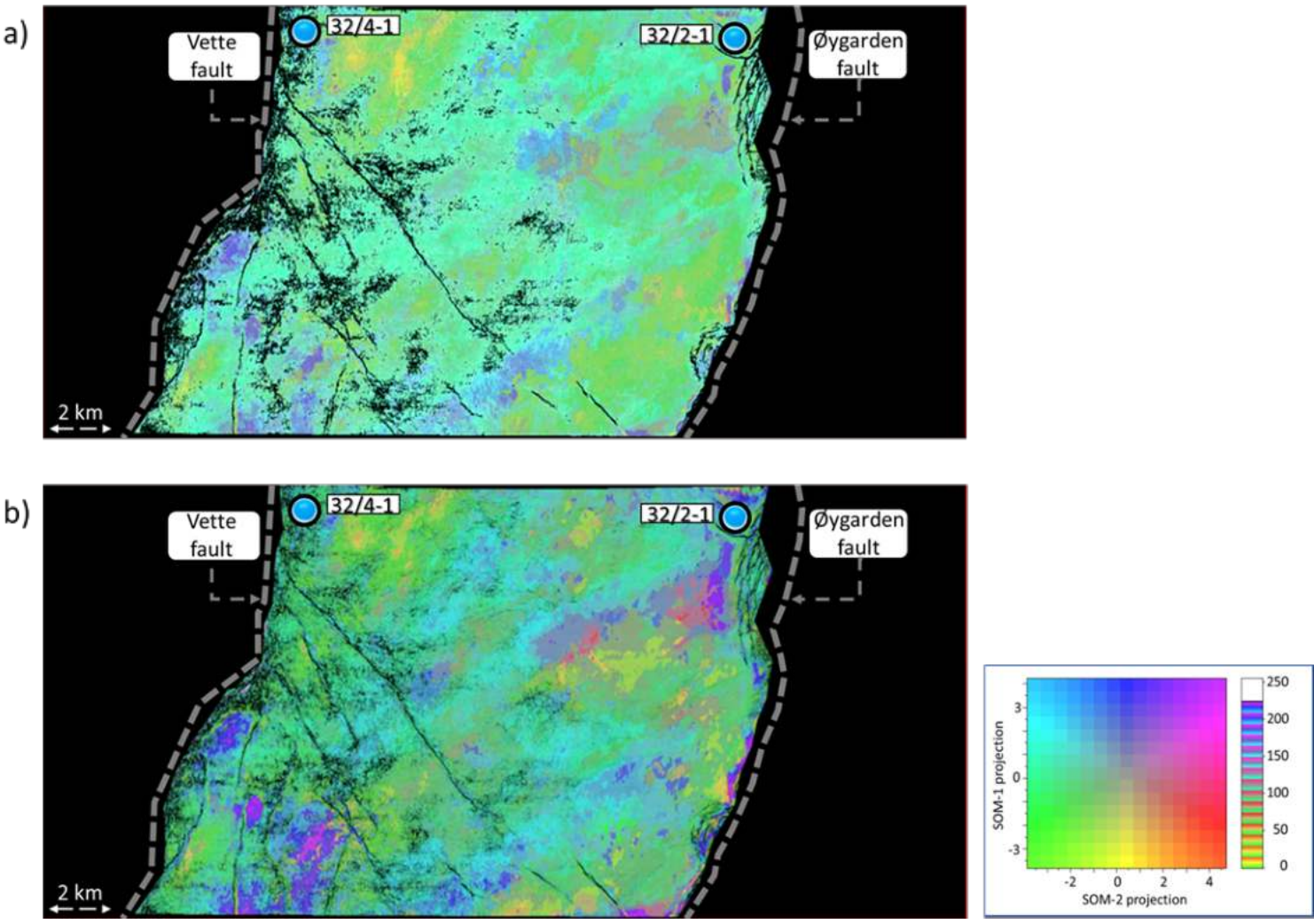



**Figure 9:** Stratal slice at 136 ms below the Sognefjord marker (within the Fensfjord Fm) extracted from the SOM crossplot volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). As with $k_{means}$, only the target area between the Vette and Øygarden was classified.

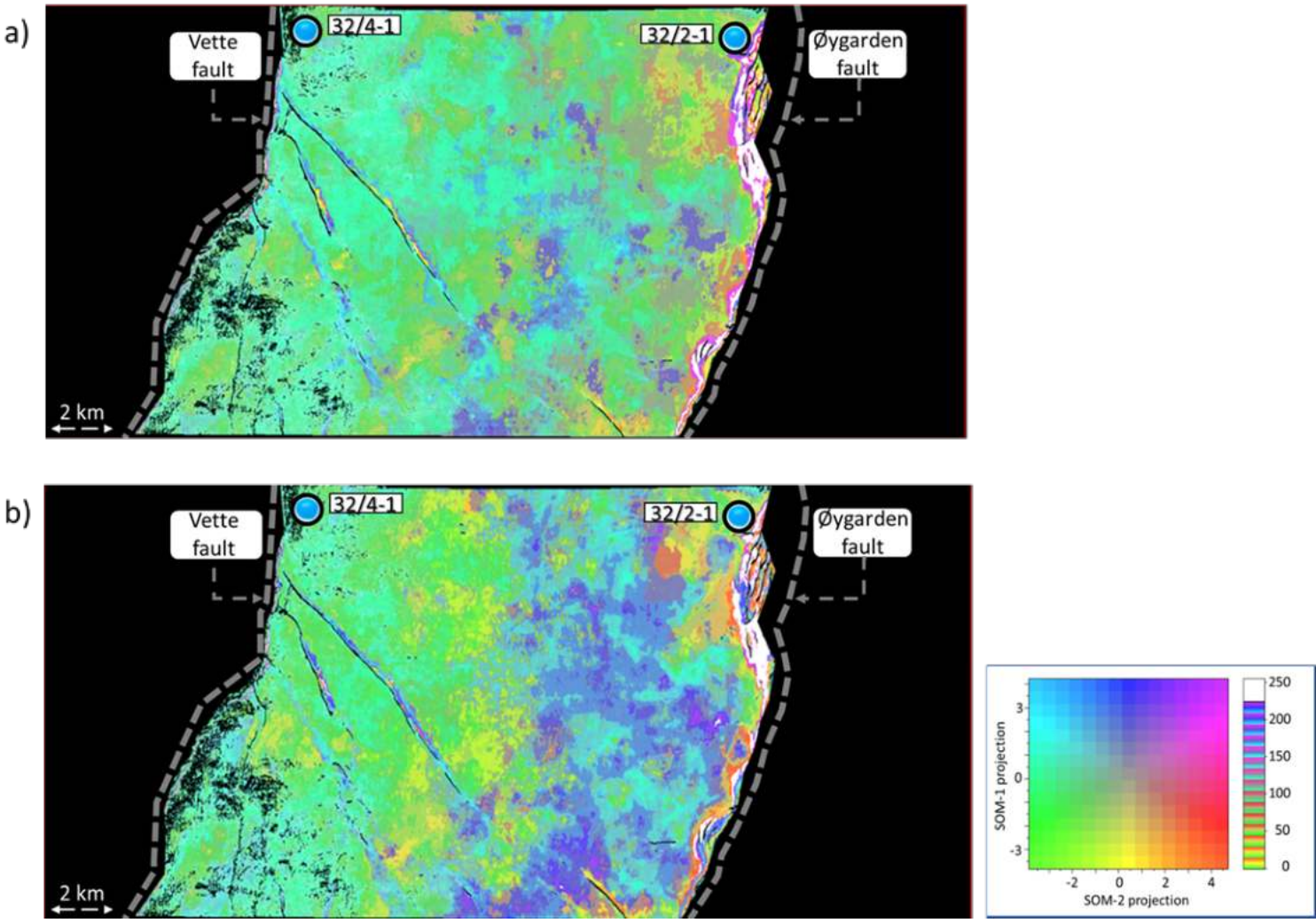



**Figure 10:** Stratal slice at 228 ms below the Sognefjord marker (within the Krossfjord Fm) extracted from the SOM crossplot volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). As with $k_{means}$, only the target area between the Vette and Øygarden was classified.

## Generative Topographic Mapping

The Kohonen self-organizing map described above, while the most popular unsupervised clustering technique, being easy to implement and computationally inexpensive, has limitations. There is no theoretical basis for selecting the training radius, neighborhood function and learning rate as these parameters are data dependent (Bishop et al., 1998; Roy, 2013). No cost function is defined that could be iteratively minimized and would indicate the convergence of the iterations during the training process, and finally no probability density is defined that could yield a confidence measure in the final clustering results. Bishop et al. (1998) developed an alternative dimension reduction technique called a *generative topographic mapping* (GTM) algorithm that provides a probabilistic representation of the data vectors in latent space.

The GTM method begins with an initial array of grid points arranged on a lower dimensional latent space. Each of the grid points are then nonlinearly mapped onto the lower dimensional non-Euclidean curved surface defined by $K$ centroids $\mathbf{m}_k$ of the N-dimensional Gaussians with a fixed variance 1/β that best represent the $R$ data vectors. At each iteration, the variance 1/β is decreased and the Gaussian centroids $\mathbf{m}_k$ moved until we reach convergence. Roy (2013) and Roy et al. (2014) describe the details of the method and demonstrate its application for mapping of seismic facies to the Veracruz Basin, Mexico.

As it may have become apparent from the descriptions above, the SOM and GTM techniques project data from a higher dimensional space (8D when 8 attributes are used) to a lower dimensional space which may be a 2D plane or a 2D deformed surface. Once they are projected onto a lower dimensional space, the data can be clustered in that space, or interactively clustered with the use of polygons.

In Figures 11 to 13 we show the displays equivalent to those shown for $k_{means}$ or SOM analysis, where some of the clusters can be interpreted with ease with less background clutter and confusion.

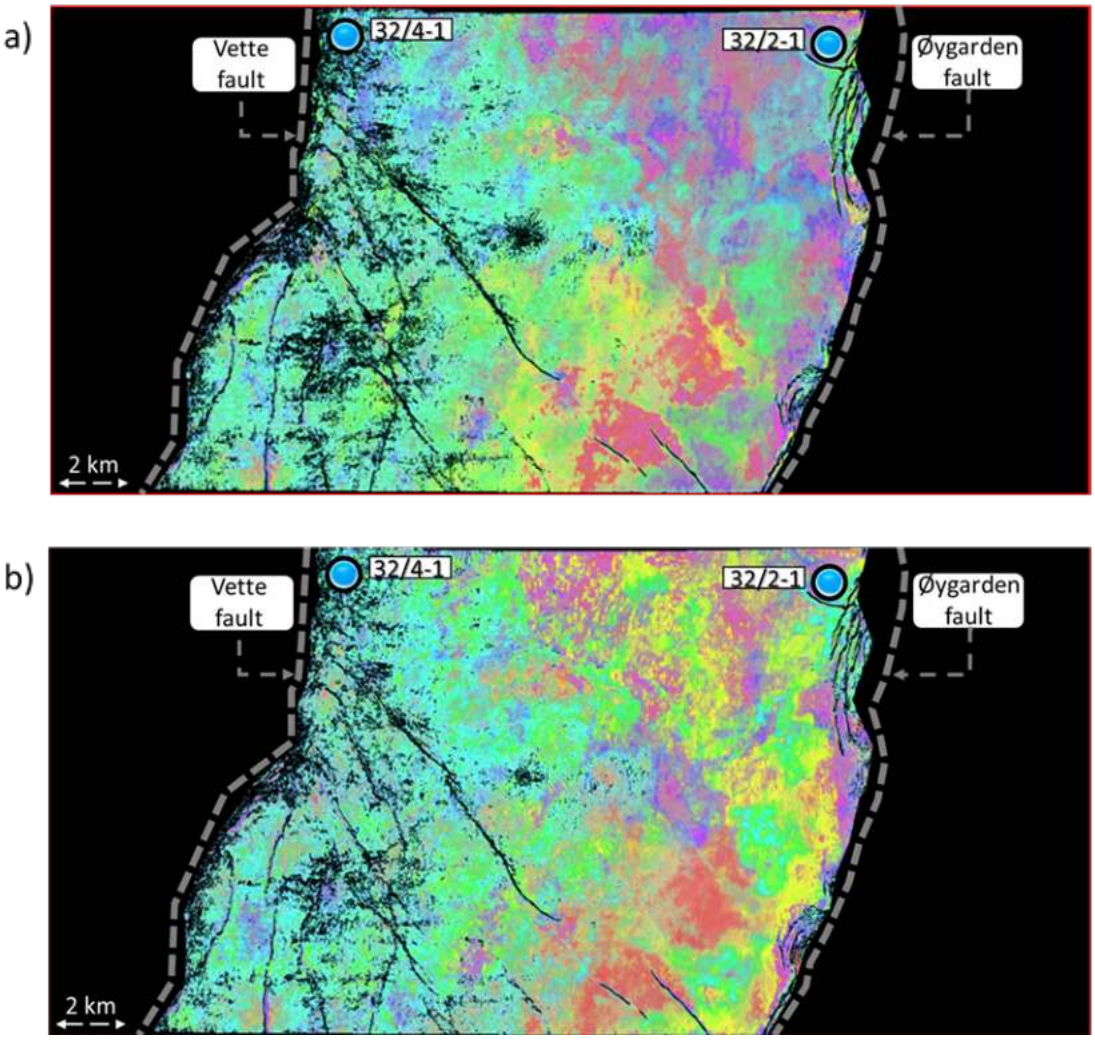



**Figure 11:** Stratal slice at 80 ms below the Sognefjord marker (within the Sognefjord Fm) extracted from the GTM crossplot volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). As with $k_{means}$, only the target area between the Vette and Øygarden was classified.

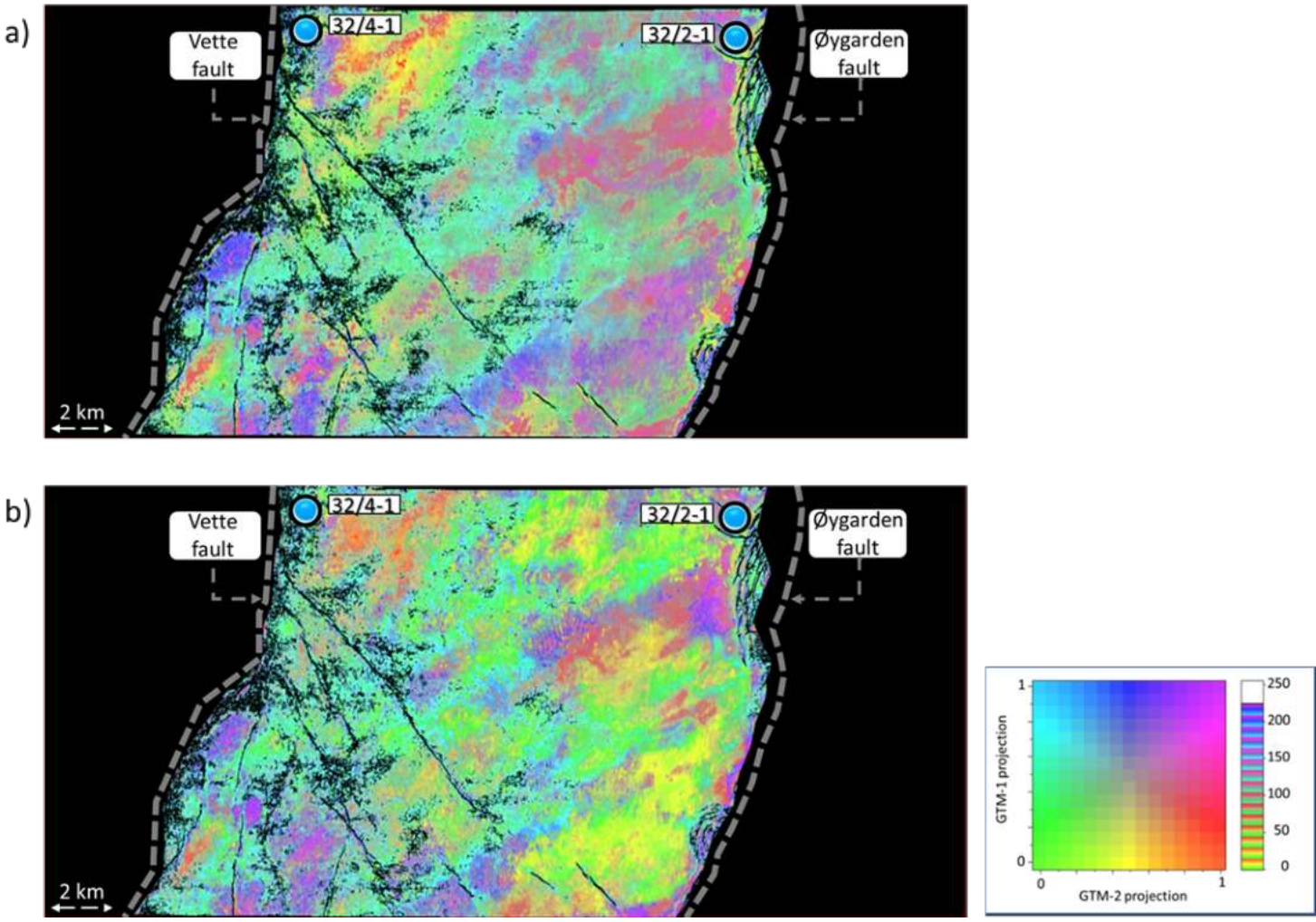



**Figure 12:** Stratal slice at 136 ms below the Sognefjord marker (within the Fensfjord Fm) extracted from the GTM crossplot volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). As with $k_{means}$, only the target area between the Vette and Øygarden was classified.

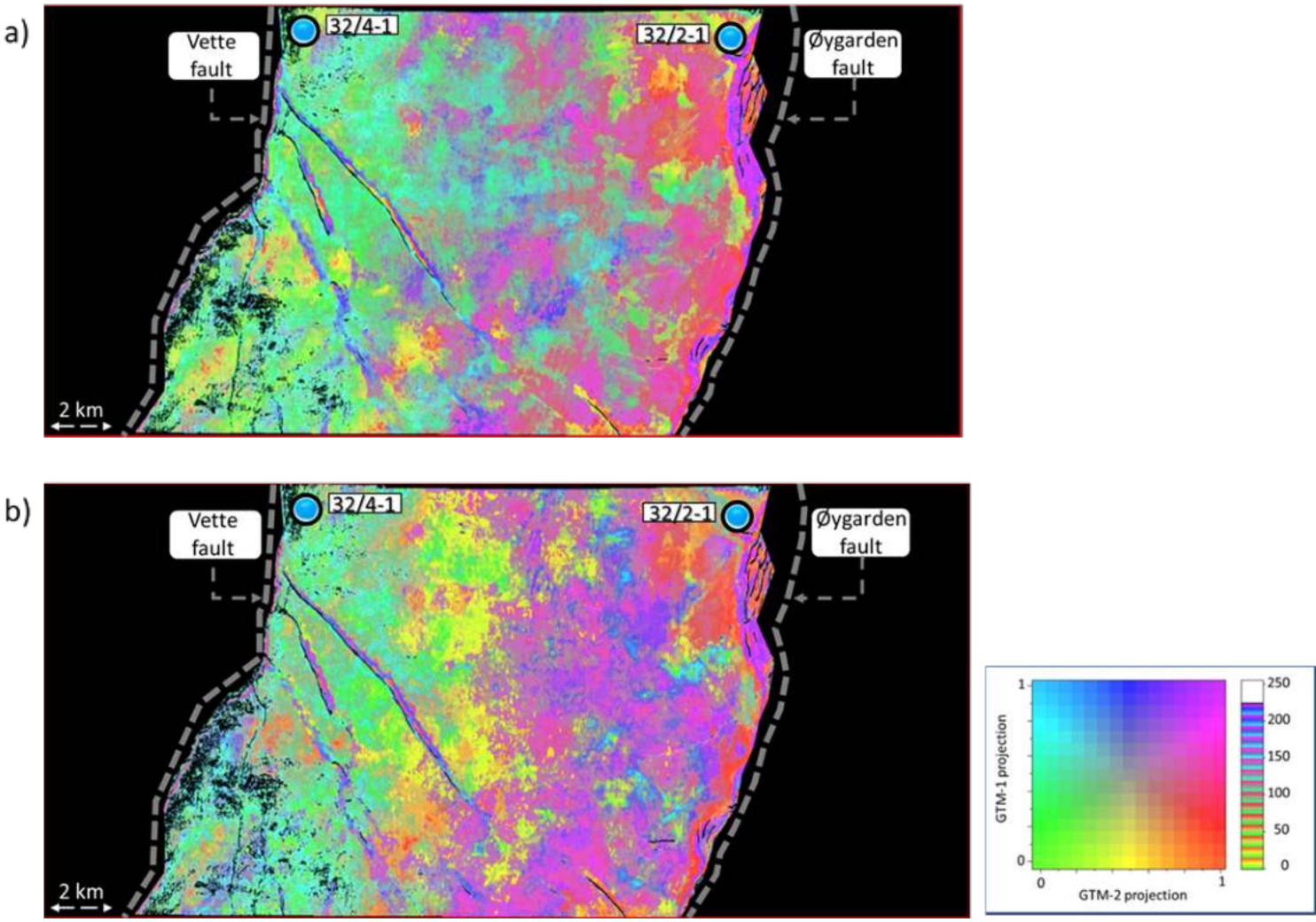



**Figure 13:** Stratal slice at 228 ms below the Sognefjord marker (within the Krossfjord Fm) extracted from the GTM crossplot volume computed on attributes generated on (a) the input seismic data volume, and (b) the spectrally balanced input seismic data volume. The two volumes have been corendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). As with $k_{means}$, only the target area between the Vette and Øygarden was classified.

## Conclusions

We have found that spectral balancing of the input seismic data when used for attribute generation and further used in some of the multiattribute processes discussed in this paper can significantly aid in accurate interpretation. Results obtained for the unsupervised machine learning applications employing both the input seismic as well as its spectrally-balanced version depict superior performance of the latter in terms of clarity of clusters as well as color variations within them, probably in response to the expected geologic variations as mentioned in the introduction.

Applications of $k_{means}$, SOM and GTM techniques to the same data allowed us to assess their relative strengths as well as their suitability. We found that both GTM and SOM show more promising results than $k_{means}$, with GTM having an edge over SOM in terms of the detailed distribution of seismic facies and in terms of better resolution and distinct definition of the geologic features seen on the displays. The actual cluster number assigned to a $k_{means}$ cluster changes after data conditioning which causes challenges when trying to compare the two results. In contrast, both SOM and GTM are computed on 2D manifolds where the clusters are ordered such that small-to-moderate changes in the input through data conditioning gives rise to correspondingly small-to-moderate changes in the clustering.

Usually, the seismic facies maps in the zones of interest are calibrated with the lithofacies information obtained from well cores and cuttings. As there is appreciable difference in resolution between the two types of data, it is advisable to enhance the resolution of seismic data by adopting a spectral balancing workflow. Such a workflow can narrow down the resolution gap between the facies data types (seismic and geologic) as well as help perform a better correlation/calibration between the two. This is thus the motivation for the work described in this paper.

Although we can neither resolve such heterogeneities nor define their cause without additional well control, the seismic data will often vary in subtle ways through lateral changes in the amplitude, spectrum, and continuity which we can quantitatively measure with seismic attributes. The N attributes define an N-dimensional vector in attribute space. By clustering like vectors together, we implicitly cluster reservoir facies that exhibit the same seismic response. Once clustered, the a posteriori analysis can take two forms:

(1) the human interpreter looks at each cluster and describes a pattern commonly used in interactive interpretation with words describing the cluster such as "smooth", "blocky", "wormy", "railroad track", "rugose", and so forth.

(2) The human interpreter looks at the values of the cluster attribute centroid $mu_k$ (for $k_{means}$), prototype vector/neuron $p_k$ (for SOM), or mean, $m_k$, (for GTM), unscales them, and then compares the unscaled values to each attribute histogram for the entire survey. The resulting description will then be more quantitative than the previous form with words like "high bandwidth", "low coherence", "high entropy", "moderate energy" and so forth, where the values of high, moderate, and low refer to how the cluster attribute vector values relate to the median of each of the unscaled (world coordinate) attributes used interactively.

Once we have ground truth provided by well control, simulator history matching, microseismic events, or other measurements, the interpreter can then assign a more geologic or petrophysical name to the cluster such as "limestone stringers", "poor flow unit/baffle", "brittle" as appropriate.

Though the analysis is qualitative at present, it paves the way for more detailed work as more well and other data become available.

## Acknowledgements

The first author would also like to thank the Attribute-Assisted Seismic Processing and Interpretation (AASPI) Consortium, University of Oklahoma, for access to their software, which has been used for all attribute computation as well as Subsurface AI (formerly Geomodeling Technology Corp.) Calgary, for making the Attribute-Studio™ software available. We wish to thank Gassnova and Equinor for access to the Smeaheia 3D seismic and other associated data used in this exercise.

### References

Bishop, C. M., Svensen, M. and Williams, C. K. I. ,1998, The generative topographic mapping: Neural Computation, **10**(1), 215-234.

Chopra, S., and K. J. Marfurt, 2016, Spectral decomposition and spectral balancing of seismic data: *The Leading Edge*, **26**(9), 936–939, http:// dx.doi.org/10.1190/tle35020xxx.1.

Furre, A., Bussat, S., Ringrose, P., and Thorsen, R., 2017, Optimizing monitoring strategies for contrasting offshore CO2 storage sites, in EAGE/SEG workshop on $CO_2$ storage monitoring.

Ha, T. N., D. Lubo-Robles, K. J. Marfurt and B. C. Wallet, 2021, An in-depth analysis of logarithmic data transformation and per-class normalization in machine learning: Application to unsupervised classification of a turbidite system in the Canterbury Basin, New Zealand, and supervised classification of salt in the Eugene Island minibasin, Gulf of Mexico, Interpretation, 9(3), T685 – T710. http://dx.doi.org/10.1190/INT-2021-0008.1.

Helmore, S., 2009, Dealing with the noise – Improving seismic whitening and seismic inversion workflows using frequency split structurally oriented filters: 78th Annual International Meeting of the SEG, Expanded Abstracts, 3367-3371.

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

Kohonen, T, 2001, *Self-organizing Maps*: Springer-Verlag.

Marfurt, K., and M. Matos, 2014, Am I blue? Finding the right (spectral) balance: AAPG Explorer, http://www.aapg.org/publications/news/explorer/column/articleid/9522/am-i-blue-finding-the-right-spectral-balance, accessed 12 March 2015.

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

Roy, A., Romero-Peleaz, A. S., Kwiatkowski, T. J. and Marfurt, K. J., 2014, Generative topographic mapping for seismic facies estimation of a carbonate wash, Veracruz Basin, southern Mexico, Interpretation. 2, SA31-SA47.

Yilmaz, O., 2001, Seismic data processing, SEG.