Bandwidth Extension of Seismic Data and its Impact on Seismic Unsupervised Facies Classification

SATINDER CHOPRA

SAMIGEO, CALGARY, CANADA, SATINDER.CHOPRA@SAMIGEO.COM

RITESH KUMAR SHARMA

SAMIGEO, CALGARY, CANADA

JOHN P. CASTAGNA

THE UNIVERSITY OF HOUSTON, USA

KURT J. MARFURT

THE UNIVERSITY OF OKLAHOMA, NORMAN

Abstract

Bandwidth extension of seismic data is a desirable goal when the available data has inadequate frequency content to delineate thin beds, facies changes, and subtle faults. Though significant efforts are expended to preserve frequency content during processing of seismic data, processes such as multi-window statistical deconvolution, inverse-Q filtering, as well as time-varying spectral whitening, are not effective enough when the intervals to be characterized are below tuning. Consequently, methods have been developed for spectral balancing as well as bandwidth extension of the input seismic data. This article describes and illustrates the performance of sparse-layer seismic reflectivity inversion, carried out using basis pursuit decomposition to extend the seismic bandwidth. This method yields a reflectivity series which can subsequently be filtered to a desirable bandwidth that exhibits optimum resolution and reasonably accurate synthetic ties to wells. This method can also be used to derive relative acoustic impedance. We also show the improved lateral resolution of bandwidth extension on attributes such as coherence and curvature. We use these and other attributes to classify seismic facies, using several unsupervised machine learning methods, and find they provide a higher level of detail, whether it is in the lineaments corresponding to faults or in the thin-layered litho-intervals. Examples are shown from a 3D and wells at Smeaheia, Norway.

Introduction

During the last decade, developments in broadband acquisition and processing for land and marine seismic data have demonstrated the importance of enhanced signal bandwidth. The processing of such broadband data comes with challenges, a salient one being noise attenuation at low frequencies for gaining signal bandwidth. Bandwidth extension processing, when carried out post-imaging, can facilitate improved seismic interpretation, attribute analysis

and reservoir characterization, and can benefit both legacy and modern broadband seismic data.

As broader band and higher frequency data result in better seismic resolution (Kallweit and Wood, 1982), typically, every effort is made during seismic data processing (post-imaging) to preserve the frequency content of the data. A general practice is to use a two- or three-window *statistical deconvolution* to correct for the dynamic loss of high frequencies with increasing travel time. These windows usually overlap to avoid artifacts, thereby introducing phase distortions. Furthermore, frequencies with poor signal-to-noise ratio may be amplified and there is consequently a trade-off between data quality and bandwidth.

Inverse-Q filtering, which requires the subsurface Q profile, can be used to boost high frequencies (Bickel and Natarajan, 1985; Tonn, 1991), but suffers the same signal-to-noise ratio limitation. Because of mixed elastic and inelastic effects on the seismic spectrum, Q is difficult to estimate from surface seismic data, and therefore most workers use the constant Q model (e.g., Kjartansson, 1979). In addition to compensating for absorption by boosting the amplitude of the higher frequency components, an accurate Q-filter will also compensate for the corresponding dispersion effects by rotating the phase components of the seismic data. This stabilization of the phase spectrum of the seismic wavelet may be more important than boosting high frequency amplitudes, which if overdone will magnify noise. Although a constant Q model may provide an accurate representation of intrinsic attenuation, it does not properly represent attenuation due to scattering, such as from rugose surfaces, periodic fractures, or patchy saturation.

Time-varying spectral whitening (TVSW) involves passing the input data through a number of bandpass filters and determining the decay rates for each band. The inverse of these decay functions for each frequency band is applied, and the results are summed so that the amplitude spectrum for the output data is whitened in a time-variant way. As reflectivity spectra are almost never white, and most commonly blue (Walden and Hoskin, 1985), this approach is inappropriate for data conditioning prior to inversion, as it distorts the reflectivity.

As it is a trace-by-trace process, TVSW is not appropriate for AVO applications (Yilmaz, 2001). Like the other processes discussed above, at a given frequency it multiplies signal and noise by the same scalar value and can reduce overall signal-to-noise ratio by boosting frequencies with less signal than noise. All these methods often prove to be ineffective when the target intervals to be characterized are seismically thin below the Widess (1973) tuning limit. For this reason, other advanced methods have been developed for spectral balancing (Marfurt and Matos, 2014; Chopra and Marfurt, 2016) as well as extending the bandwidth of the input seismic data (Puryear and Castagna, 2008; Chopra et al., 2006; Smith et al., 2008; Zhang and Castagna, 2011). Liang et al. (2017) provide a comparison of seismic bandwidth extension methods.

In this paper, we describe the sparse-layer seismic reflectivity inversion method (Zhang and Castagna, 2011) for bandwidth extension. This method makes use of the basis pursuit decomposition and is superior to the traditional way of broadening the seismic bandwidth with the application of spiking deconvolution. The inversion method determines a sparse number of reflectivity patterns which, when summed together, form the original seismic trace. The determined pattern, on convolution with a wavelet of extended bandwidth, yields the desired bandwidth extension. The details of the method are discussed in the following section.

Next, we demonstrate the advantages that accrue, by performing bandwidth extension on a 3D seismic dataset (along with data for two wells on the 3D survey) from the Smeaheia area offshore Norway. The first advantage is in terms of reasonably accurate synthetic ties to wells. The second advantage is that seismic attributes generated on the bandwidth extended data exhibit improved lateral resolution. We demonstrate this with the generation of attributes such as relative acoustic impedance, multispectral energy ratio coherence, and multispectral curvature.

Finally, we take the exercise forward by generating a number of attributes (relative acoustic impedance, instantaneous amplitude, instantaneous frequency, gray-level co-occurrence matrix (GLCM) energy and spectral magnitude at different frequencies) on both the original and bandwidth extended versions of the seismic data. The generated attributes are then used to perform unsupervised machine learning facies classification using two known methods, *self-organizing mapping* (SOM) and *generative topographic mapping* (GTM). On comparing the stratal displays extracted from the facies volumes so generated, we find that the facies generated on bandwidth extended data yield higher spatial resolution, in both the lineaments corresponding to faults and the thin-layered litho-intervals.

In view of such observations, it would not be incorrect to say that bandwidth extension is an enabler for the application to narrow band legacy seismic data. Modern broadband seismic data have a wider band of frequencies than legacy data but may still fall short of expectations for thin reservoir targets. In those cases, bandwidth extension could prove to be useful, as well.

Sparse-layer-based bandwidth extension versus spiking deconvolution

The classical way to broaden the seismic bandwidth is to apply spiking deconvolution to the data. Although the wavelet is unknown, it must be minimum phase, which requires a preprocessing step for vibroseis data. Then, through a suite of autocorrelations, a wavelet is estimated that best reduces the seismic data into a sequence of spikes. Because the effective source wavelet changes with time, the algorithm must be applied in overlapping windows or in a running-window manner. Also, the very low and high frequencies are not present in the input data, such that a bandpass filter needs to be applied to the output to reduce the noise. This method can improve resolution somewhat by shaping the spectrum of the wavelet embedded in

the seismic data but does not extend signal to frequencies where it is initially absent. It also makes the mathematically convenient, but unjustified, assumption that the reflectivity series is white, and therefore distorts the reflectivity spectrum.

In principle, spiking deconvolution should be able to resolve a thin bed that falls below the tuning frequency. In practice, because the wavelet is unknown, this resolution is poor, where phase effects introduced by tuning are partially incorporated in estimating the phase of the source wavelet.
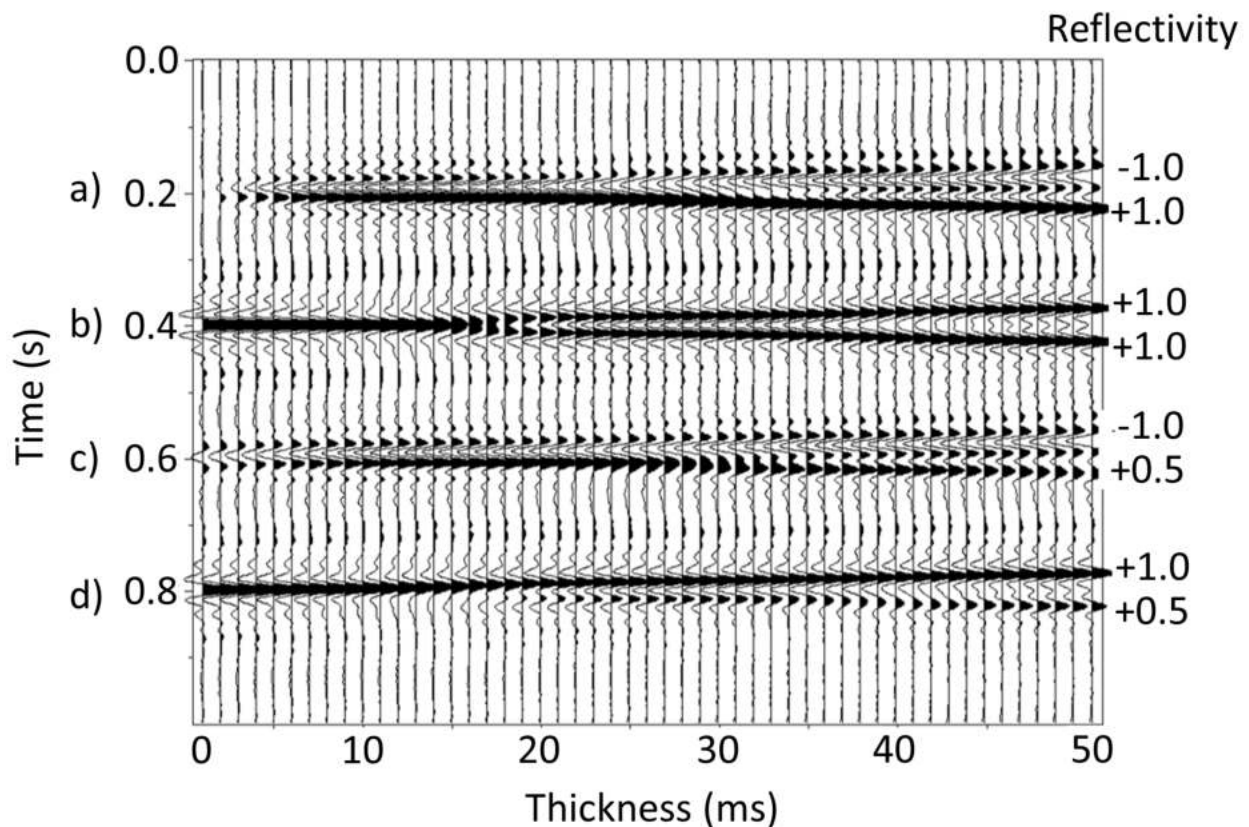
Figure 1: Four wedge models described by Chung and Lawton (1995) where the thickness ranges from 0 to 50 ms at 1 ms increments and where (a) the reflectors are equal and opposite polarity, (b) the reflectors are equal and the same polarity, (c) the reflectors are unequal and opposite polarity, and (d) the reflectors are unequal and the same polarity. In this example, each wedge has been convolved with a 5-10-40-50 Hz Ormsby wavelet. For thin bed inversion, the wedges in (a) and (b) are convolved with either a well tie source wavelet or with a statistical wavelet, thereby forming 101 basis functions that will be fit to the data at each sample.

The Zhang and Castagna's (2011) *sparse-layer seismic reflectivity inversion* is different in two ways. First, a wavelet estimate that can be temporally and spatially varying, is made. Unlike deconvolution, the method effectively shapes the spectrum of the wavelet rather than the data. Second, the basis function does not consist of an isolated spike (with an unknown wavelet) but

rather of a library of thin bed responses convolved with an externally estimated seismic wavelet field using *a priori* information and statistical assumptions. Figure 1a and 1b show 101 basis functions (the response for Figure 1a at thickness 0 is 0.0 and ignored) for a 5-10-40-50 Hz Ormsby wavelet. Zhang and Castagna (2011) note that the response of Figure 1c and 1d can be represented as a weighted sum of the two basis functions in Figure 1a and 1b.

Given these dipoles (2-layer or thin bed) basis functions (layer responses), there are multiple ways to fit the data, thereby giving a non-unique solution. For this reason, Zhang and Castagna (2011) use a *basis-pursuit decomposition* that not only least-squares fits the basis function to the data but favors simpler model fits using an L1-norm penalty function. As discussed by Liang et al. (2017), invoking sparsity to the layer functions that sum to form the seismic trace, rather than to the reflectivity series, does not discriminate against thin layers, as does sparse-spike inversion.

The sparse-layer inversion thus determines a sparse number of patterns which, when summed together, form the original seismic trace. To extend the bandwidth, we now use the same dipoles that were convolved with the original wavelet field, but now replace the original wavelet with one with an extended bandwidth. Explicitly stated, the algorithm replaces the original data with a model of dipoles convolved with the well-log generated synthetic wavelet or statistical wavelet, with a wavelet of our own choosing. In this manner, the unmeasured high and low frequencies in the new extended bandwidth wavelet are consistent with the same model used to represent the original data. This process is equivalent to harmonically extrapolating the sinusoidal layer response reflectivity spectrum outside the original frequency band. In this way, rather than boost signal and noise by the same amount outside the main frequency band, as is done in wavelet shaping processes such as deconvolution or inverse-Q filtering, the signal is boosted more than the noise, essentially stealing signal-to-noise ratio from the band where signal is strongest.

The three major limitations of the method are:

1. Like all inversion methods, it requires a correct seismic wavelet field,
2. it requires sufficient character and amplitude of a layer reflectivity spectrum within the original seismic band, and
3. it assumes a sparse reflectivity series

Therefore:

1. As there is no perfect wavelet extraction method, one should test the sensitivity of the output to possible variations in the extracted wavelet field.

2. The method cannot resolve a layer that is so thin that the shape of the reflectivity spectrum within the seismic band is not strongly influenced by that layer (the lowest resolvable thickness then becomes a matter of signal-to-noise ratio (S/N) within the seismic band). As this is a trade-off between character and S/N, there is no exact rule for the resolvable thickness, as it depends on S/N, original bandwidth, and geology. Typically, layers below $1/16^{th}$ of a wavelength are not well resolved (see Puryear and Castagna, 2008), and the broader the original band, the greater the extension that can be achieved. The bandwidth requirement is evident, as no extension can be performed on a sine wave.

3. As the method assumes sparsity, no extension can be achieved when reflectivity is random. As discussed by Liang et al. (2017), even in this worst-case scenario, the method can be used to amplify signal more than noise for low S/N frequencies within the original seismic band and can thus outperform deconvolution. There is the theoretical possibility that long high frequency periodic reflectivity signals outside the original seismic band may also be missed, although these are generally not important, and may in fact make contributions within the original seismic band because of spectral leakage due to windowing.

The sparse-layer inversion does not directly use well data in the inversion, though well data may be used in wavelet extraction and parameter selection (such as degree of sparsity), and of course for validation. By operating on a trace-by-trace basis, this inversion yields a reflectivity series, which is then filtered to a desirable bandwidth that exhibits an optimum combination of resolution and reasonably accurate synthetic ties to wells. The output reflectivity series can also be used to derive relative acoustic impedance.

Application of sparse layer inversion method to seismic data

For demonstrating the application of the sparse-layer inversion to seismic data, we selected the 3D seismic data from Smeaheia area offshore Norway. Smeaheia has been considered as a potential area for CO_2 storage and evaluation.

The Smeaheia area lies about 30 km east of the Troll gas field (Figure 2), on the Norwegian continental shelf. It 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 Jurassic Sognefjord, Fensfjord, and Krossfjord Formations form the producing reservoir zones in the Troll gas field.
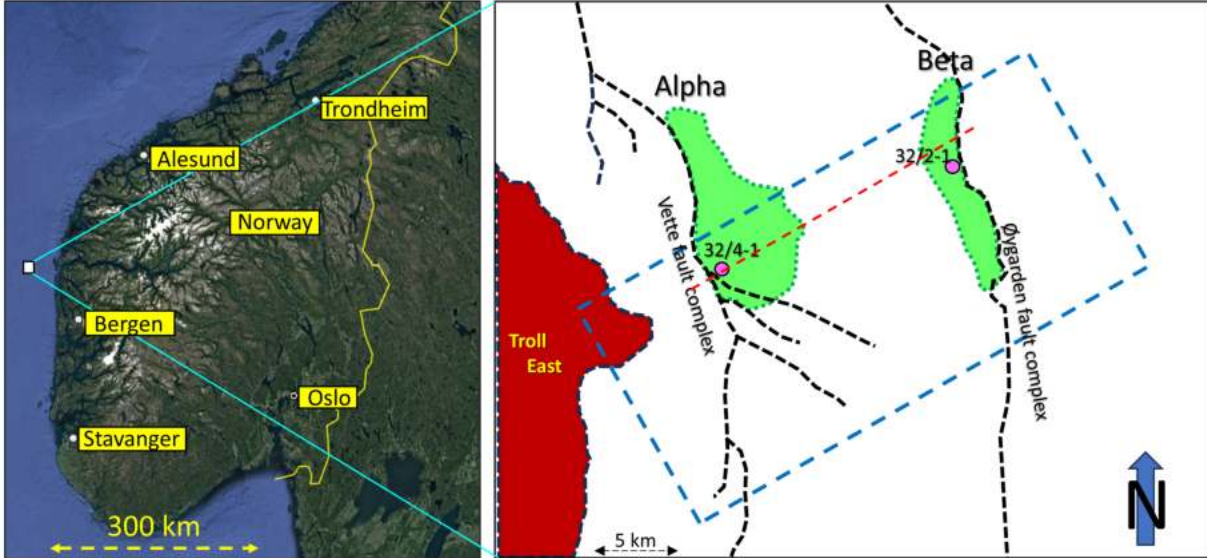
Figure 2: (Left) Location of the Smeaheia area on the Norwegian continental shelf. (Right) Zoom of the Smeaheia area indicating the two interpreted structures, 'Alpha' and 'Beta'. The two main fault complexes in the area are marked as dashed black lines. The dashed blue rectangle depicts the outline of the 3D seismic volume being used, and the dashed red line is the location of the seismic inline segments shown in Figures 4 and 7. The image to the left was prepared with the use of Google Earth Pro (modified after Furre et al., 2017).

In the Smeaheia block, there are two four-way dip 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 wells encountered reservoir of acceptable quality, both turned out to be dry.

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 the Heather and Draupne Formation shales. The Heather is comprised of silty claystone with thin streaks of limestone interfingering the Sognefjord, Fensfjord and Krossfjord sandstones. The Draupne consists of dark grey to brown/black shale that is non-calcareous, carbonaceous, and fissile claystone. Both the Heather and Draupne 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 data for demonstration of bandwidth extension

The available seismic data were the GN1101 3D survey covering the Smeaheia (blue dashed rectangle shown in Figure 2) acquired by Gassnova in 2011 and made publicly available by Gassnova and Equinor. The bin size for the data is 12.5 m 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. Logs included the complete gamma ray curves, but sonic and density logs were not recorded at the shallower depths. The seismic data volume is of good quality.

Figure 3 shows the correlation of well curves with seismic data. The blue traces represent the synthetic seismograms generated with the zero-phase wavelet shown above. The red traces represent the seismic data (a) before, and (b) after bandwidth extension. Windows of seismic data before and after bandwidth extension from where the red traces are extracted are also shown alongside. There is a reasonably good correlation between the synthetic in blue and the seismic traces in red in the time window indicated, but there are more reflection cycles showing good correlation after bandwidth extension.
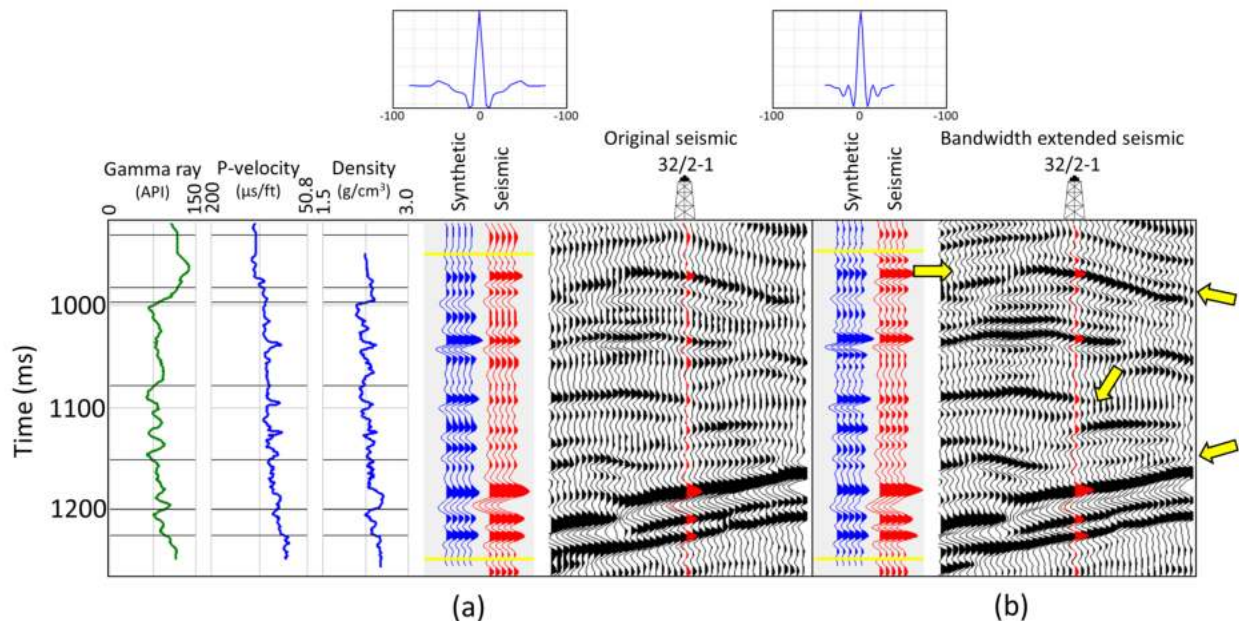

Figure 3: Correlation of well curves with seismic data. The blue traces represent the synthetics (generated with the wavelet shown above), whereas the red traces represent the seismic data (a) before, and (b) after bandwidth extension. The correlation coefficients for the ties were 0.745 for (a) and 0.756 for (b). Thus, there is a good correlation between the synthetic and red traces in the time window indicated for both data volumes but note the resolution of additional events after bandwidth extension, indicated by the yellow arrows.
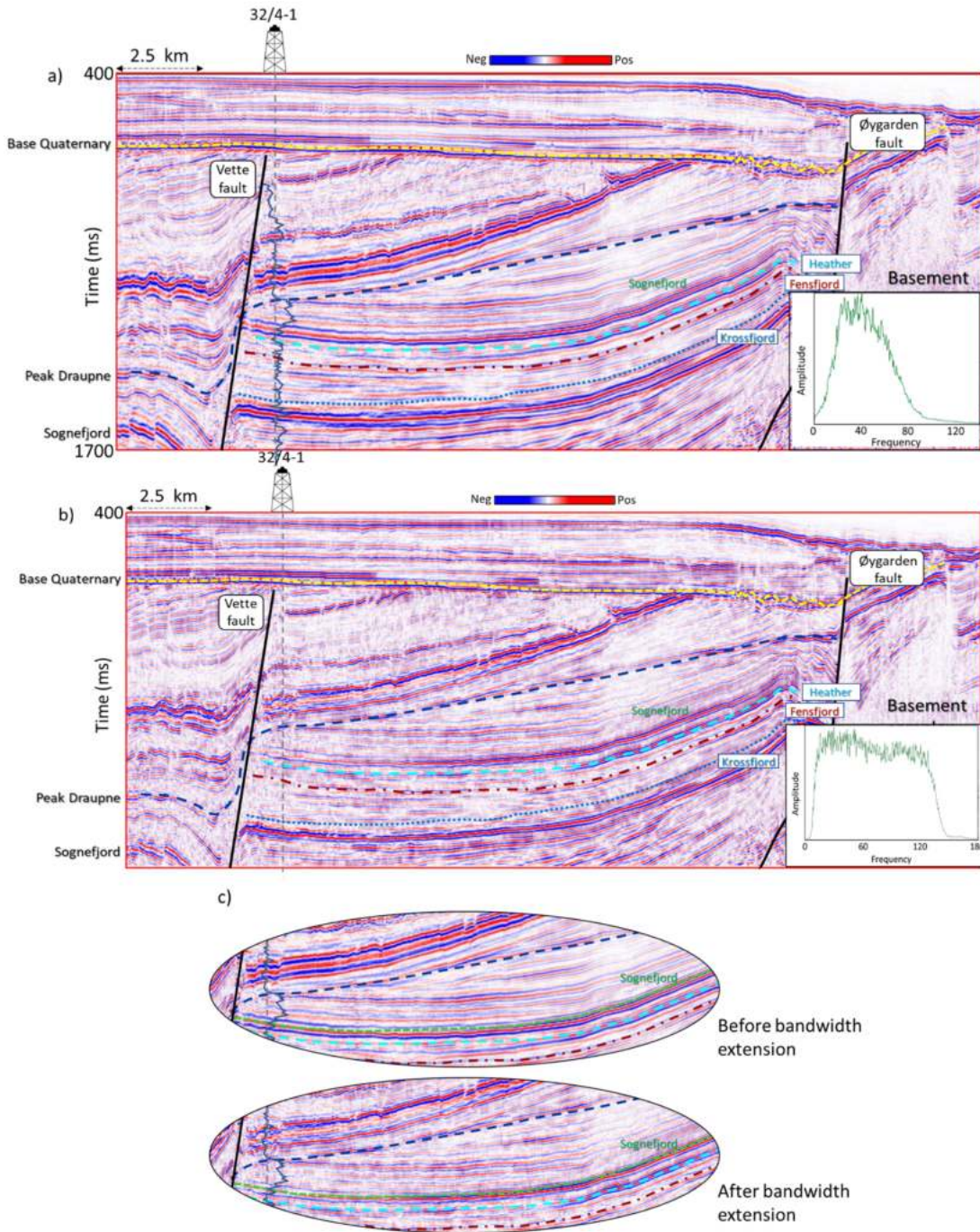
Figure 4: Segment of an inline extracted from (a) input seismic data volume, and (b) input seismic data volume after bandwidth extension. Some relevant markers and gamma ray curve for well 32/4-1 are overlain on the two sections. (c) Zoom of the two sections cropped to the area enclosed by the dashed grey oblong shapes on the two images clearly shows the higher frequency content after bandwidth extension. The frequency spectra (computed over the time window exhibited and over the whole seismic volume) for the two data volumes are also shown to the right. Notice the enhancement in resolution of the reflections after bandwidth extension.

Figure 4a shows a segment of an inline seismic section with key horizons extracted from the available seismic data volume (which has a bandpass filter applied), which passes through well 32/4-1. The gamma ray curve for well 32/4-1 is shown overlain on the section. The 'Base Quaternary' (in yellow) represents an angular unconformity, and the 'Peak Draupne' (in dark blue) the top of the shale formation. The 'Sognefjord' (in green), 'Fensfjord (in cyan) and 'Krossfjord' (in bluish green) are the markers of interest representing sandstone formations as described above. Two prominent faults in solid black (Vette Fault to the west and Øygarden Fault to the east) are also shown overlain on the section. The frequency spectrum computed for the time interval of the seismic data exhibited is shown to the right, which indicates a roll-off of frequencies beyond 40 Hz. The bandwidth extension method employed here uses the spectrum ranging from 5-80 Hz to define a model (the weights in the fit are the spectral magnitudes in Figure 4a, so the 40 Hz is weighted more than 80 Hz). Then this model extends the spectrum from 80 to 140 Hz where no data were measured. The number 140 Hz is no magic number and is arrived at by checking a range of numbers for the high end of the frequency spectrum and making an optimum choice, where the bandwidth extended data does not exhibit ringing in the seismic reflections.

Generation of different attributes

With the improved vertical resolution seen in Figure 4b, our next task is to determine if bandwidth extension also enhances the lateral resolution as measured by seismic attributes such as coherence and curvature. Specifically, we anticipate that some lateral discontinuities fall near the limits of seismic resolution in the original data but will be better resolved in the bandwidth extension data.

Energy ratio coherence: Encouraged with the higher-frequency content of the seismic data, we first generated the coherence attribute. Much has been written about this attribute and the usefulness of its application. We make use of the energy ratio-based coherence algorithm (Chopra and Marfurt, 2008). It is a generalization of Gersztenkorn and Marfurt's (1999) eigenstructure-based coherence computation, where the original trace and its Hilbert transform are used to construct the covariance matrix, thereby minimizing artifacts at zero crossings, and thereby providing the ability to use shorter analysis windows resulting in reduced stratigraphic mixing for short window computations. Coherence run on band-pass filtered or spectral voice components often delineate edges at or near the tuning frequency of a given formation (Chopra and Marfurt, 2018a, 2018b). In general, shorter, more vertically limited faults and channel edges are often better delineated at higher frequencies, while through-going faults are often better delineated at lower frequencies.

In addition to computing a covariance matrix from each bandpass filtered version of the data and then computing coherence for each one, we can also add the covariance matrices and

compute coherence from its sum, giving rise to *multispectral coherence* computation (Li and Lu, 2014; Chopra and Marfurt, 2019).

We will show the corendered displays of different seismic facies with energy-ratio coherence computed both on the conditioned input as well as bandwidth extended seismic data. The reader will notice crisper, focused, and clearer definition of the fault lineaments (in black) seen on the facies displays in colour which are shown in the next section.

Figure 5 shows a comparison of time slices at 1.66 s extracted from multispectral energy ratio coherence computed on the input seismic data with structure-oriented filtering (Figure 5a) and input seismic data with bandwidth extension and structure-oriented filtering (Figure 5b). Structure-oriented filtering is an effective incoherence noise suppression process which follows the dipping planes in the seismic data volume but does not smear across the discontinuities (Höcker and Fehmers, 2002). Notice the superior definition of the lineaments in terms of their continuity and intensity seen on the display in Figure 5b. Some of the weak lineaments seen in Figure 5a appear nice and bright in Figure 5b. The fault damage zone to the left is also much better defined in Figure 5b.
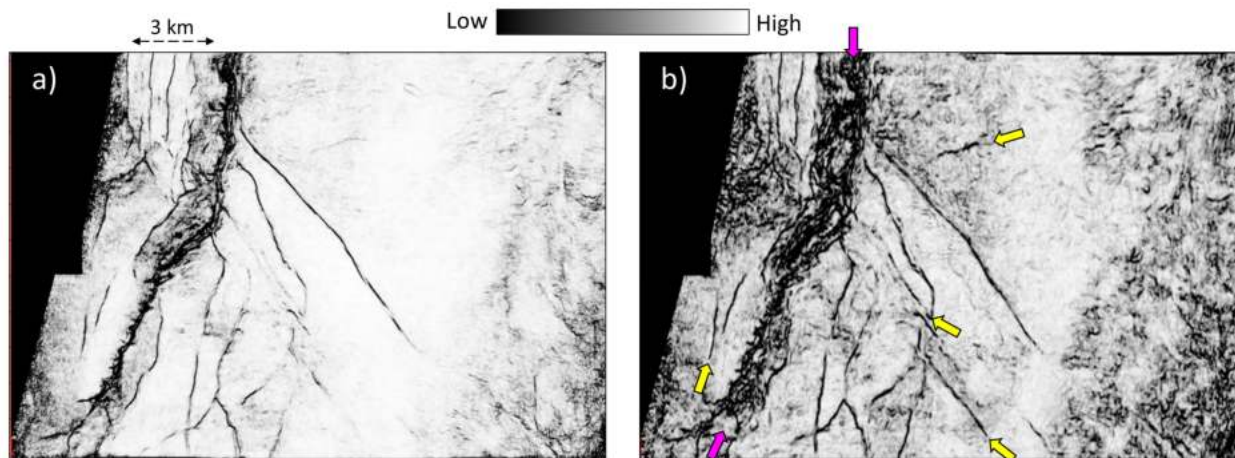

Figure 5: Time slice at 1.66 s through multispectral energy ratio coherence volumes computed on (a) input seismic data with structure-oriented filtering, and (b) input seismic data with bandwidth extension and structure-oriented filtering. Notice the superior definition of the lineaments in terms of their continuity and intensity seen on the display in (b). Some of the weak lineaments seen in (a) are better delineated in (b) indicated by yellow arrows. The fault damage zone indicated by magenta arrows is better defined in (b).

Curvature: Volumetric computation of curvature was introduced by Al-Dossary and Marfurt (2006). By first estimating the volumetric reflector dip and azimuth that represents the best single dip for each sample in the volume, followed by computation of curvature from adjacent measures of dip and azimuth, a full 3D volume of curvature values is produced. There are many curvature measures that can be computed, but the most-positive and most-negative curvature measures are the most useful in that they tend to be most easily related to geologic structures.

Volumetric curvature attributes are valuable in mapping subtle flexures and folds associated with fractures in deformed strata (Chopra and Marfurt, 2007; 2010). In addition to faults and fractures, stratigraphic features such as levees and bars and diagenetic features such as karst collapse and hydrothermally altered dolomites also appear to be well-defined on curvature displays. Channels appear when differential compaction has taken place. An interesting feature of this attribute is that the curvature computation that can yield both long wavelength and short wavelength estimates which enhance geologic features having different scales. Curvature images having different wavelengths provide different perspectives of the same geology (Bergbauer et al., 2003).
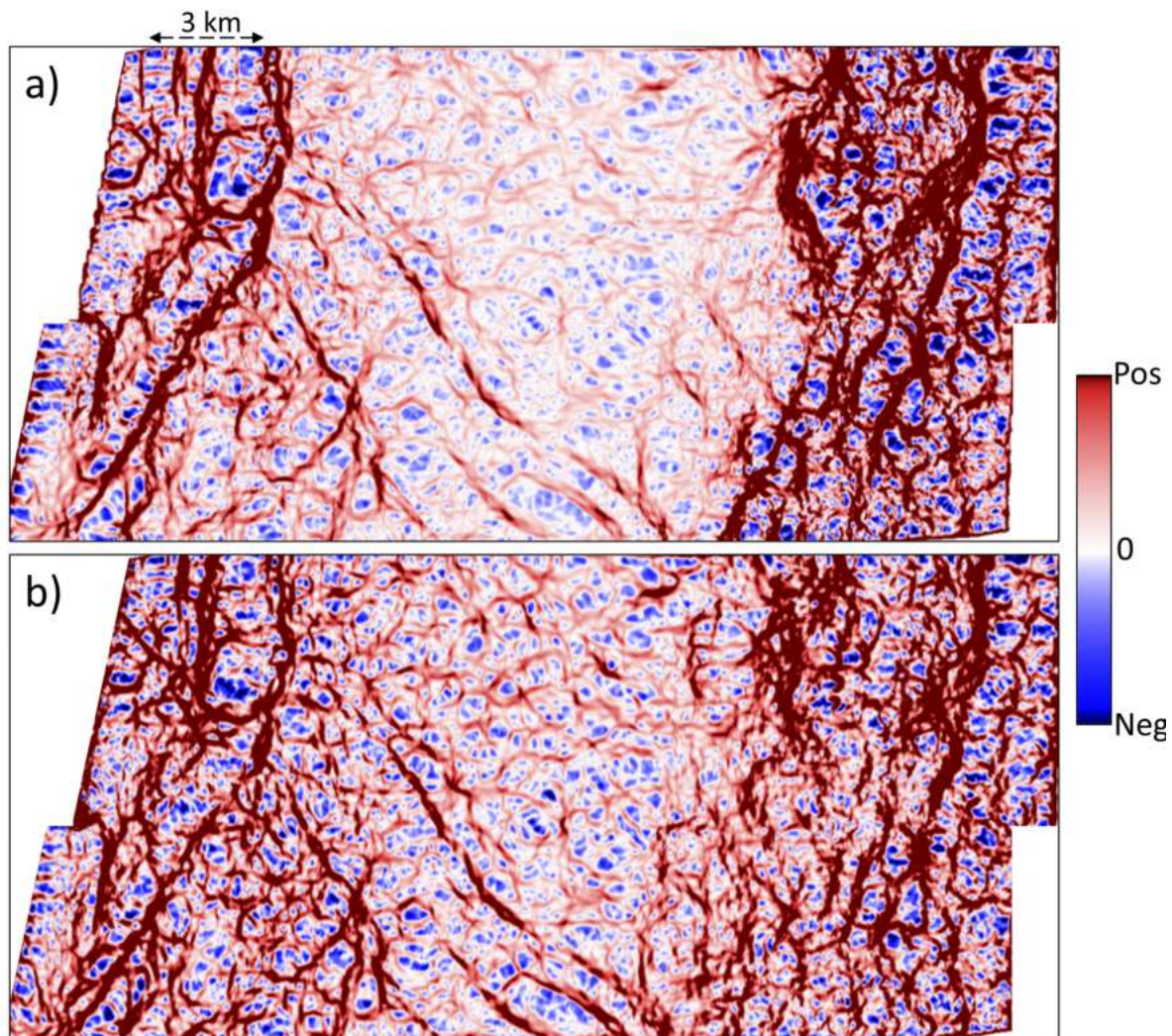

Figure 6: Time slices at 1.66 s through most-positive curvature (long-wavelength) volumes computed on (a) input seismic data with structure-oriented filtering, and (b) input seismic data with bandwidth extension and structure-oriented filtering. Notice the superior definition of the lineaments in terms of their continuity, intensity and resolution seen on the display in (b). Some of the weak lineaments seen in (a) appear as nice and bright in (b), especially in the middle part of the displays.

Figure 6 shows a comparison of time slices at 1.66 s from most-positive curvature (long-wavelength) volumes computed on the input seismic data with structure-oriented filtering and input seismic data with bandwidth extension and structure-oriented filtering. Notice the superior definition of the lineaments in terms of their continuity, intensity and resolution seen on the display in Figure 6b. Some of the weak lineaments seen in Figure 6a appear as nice and bright in Figure 6b, especially in the middle part of the displays.
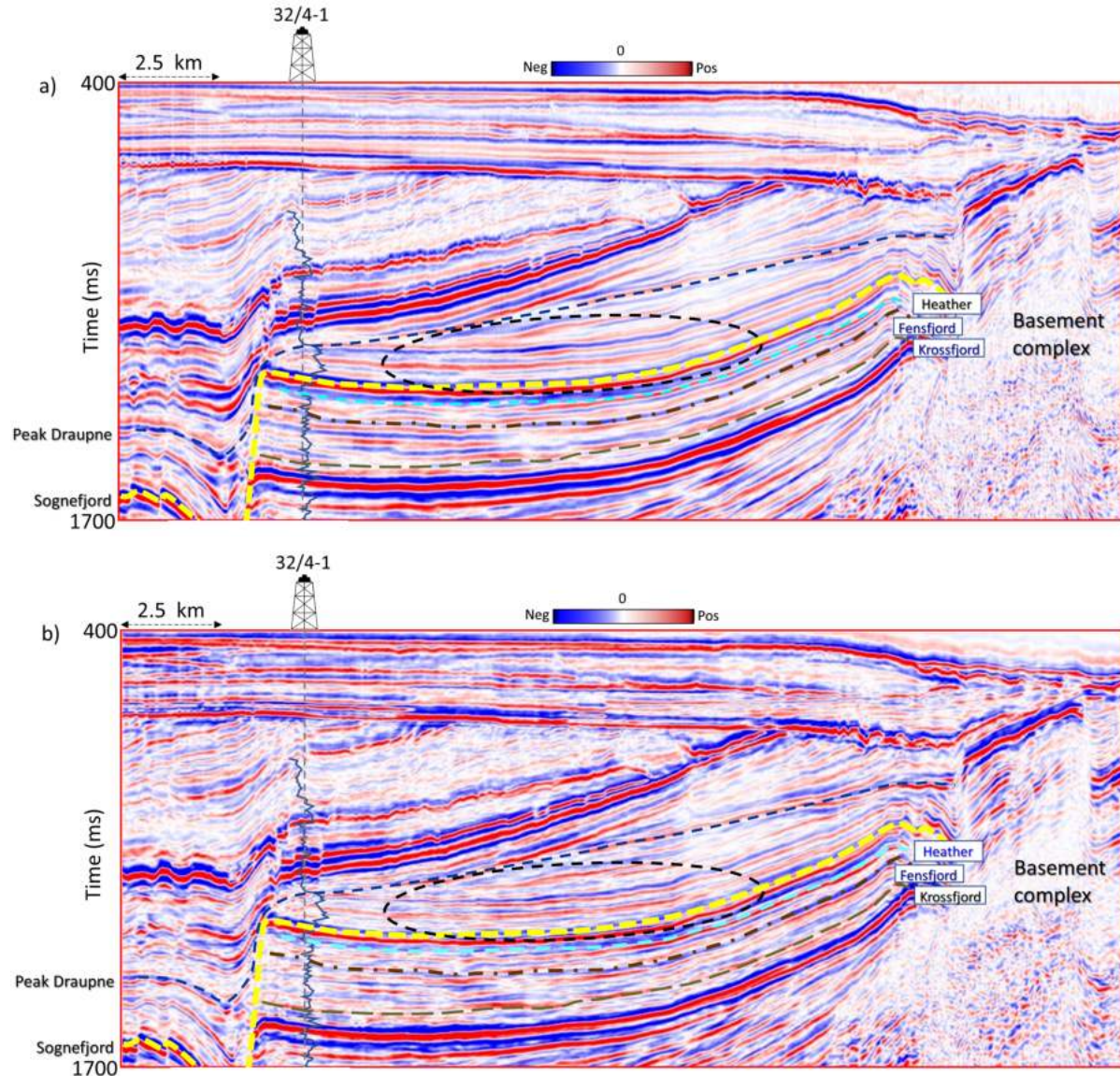
Figure 7: Segment of an inline extracted from relative acoustic impedance attribute computed on the (a) input seismic data volume, and (b) input seismic data volume after bandwidth extension. Some relevant markers and gamma ray curve for well 32/4-1 are overlaid on the two sections. Notice the enhancement in the resolution of the reflections after bandwidth extension. The tuning thickness at the levels of Sognefjord, Fensfjord, and Krossfjord formations was 16 m, 19 m, and 24 m, respectively, before and these improved to 13 m, 15 m, and 17 m, respectively, after bandwidth extension.

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 bandwidth extension 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 7 shows a comparison of inline sections from relative impedance computed on the input seismic volume (Figure 7a) and the equivalent section from relative impedance computed on frequency-balanced seismic volume (Figure 7b). Notice the enhanced resolution and better relative impedance definition on the section in Figure 7b. 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 bandwidth extension is shown in Figure 8. Notice the crisp definition of the faults as indicated by the highlighted areas in dashed purple outlines.

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

Instantaneous envelope/instantaneous 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.

All these different attributes have been generated on both the input seismic and the bandwidth extended versions to use them as input for unsupervised seismic facies classification using machine learning techniques, which are described in the next section.
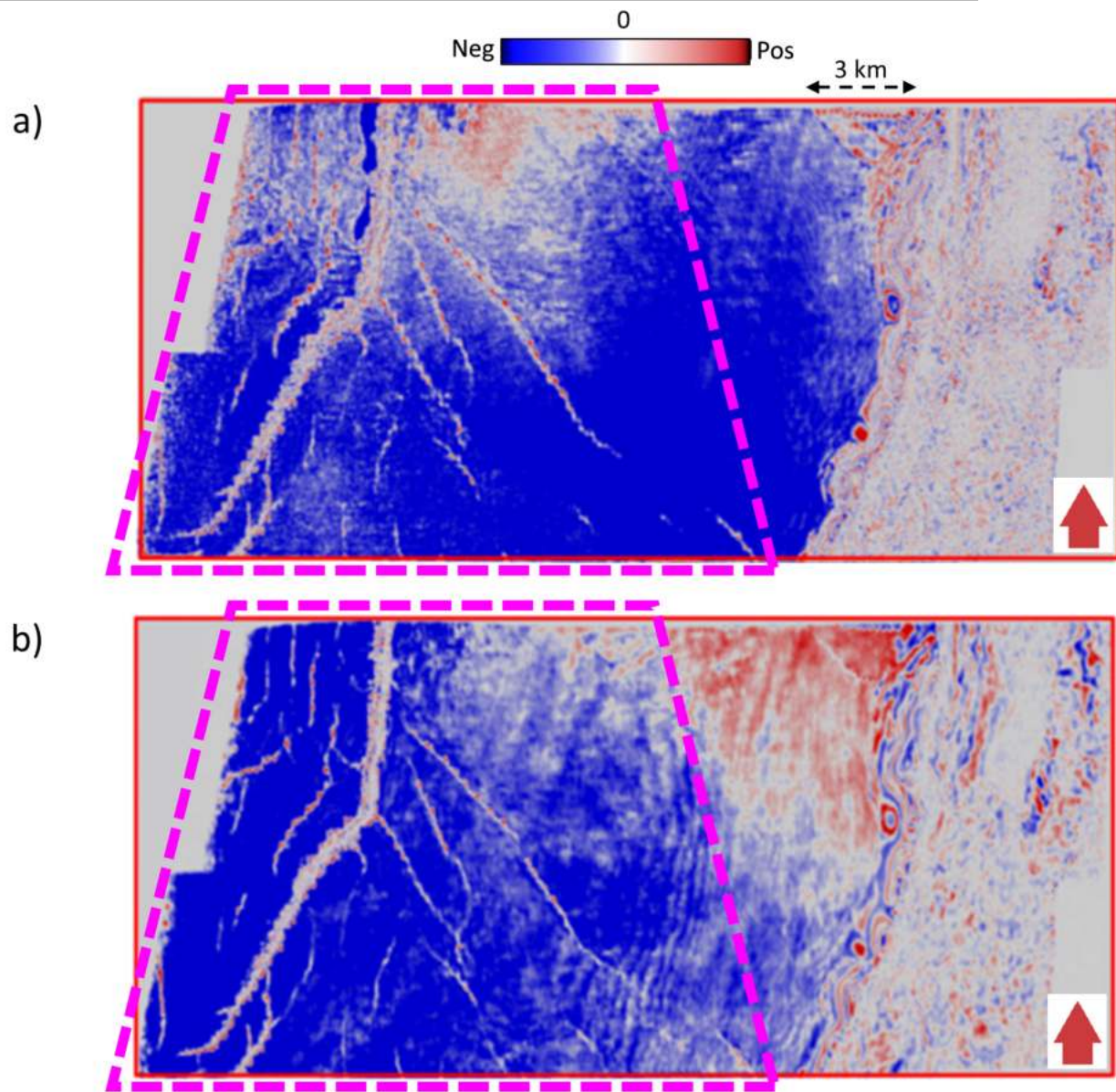
Figure 8: Stratal slice 32 ms above the Sognefjord marker from the relative acoustic impedance attributes computed from (a) input seismic data, and (b) input seismic data after bandwidth extension. Notice the crisp definition of the faults highlighted areas in dashed purple outline in (b).

Seismic facies classification using machine learning techniques on input seismic, and bandwidth extended seismic data

We carry out seismic facies classification with the application of two unsupervised machine learning methods, namely self-organizing mapping (SOM), and generative topographic mapping (GTM). A comparison of the two versions reveals the superior spatial facies resolution as well as crisp definition of faults/fractures as seen on the bandwidth extended seismic data.

We describe the details of this exercise, methods, and results in the following sections.

Machine learning uses mathematical operations to learn from the similarities and differences in the provided data, in order to make decisions or predictions. There are two broad families of machine learning algorithms. The first algorithm family includes dimensionality reduction algorithms such as PCA and 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. 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 show the application of *self-organizing mapping* (SOM) and *generative topographic mapping* (GTM) on the Smeaheia data volume.

Self-organizing mapping

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 9 shows the equivalent stratal displays (within the Sognefjord formation) extracted from the SOM crossplot volume computed for the input and frequency-balanced versions of the seismic data, using a 2D color bar. Some of the clusters seen on the display in Figure 9b are better defined than the ones shown in Figure 9a.

A similar comparison of stratal slices from the SOM crossplot volume is shown in Figures 10 and 11 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 frequency-enhanced version.
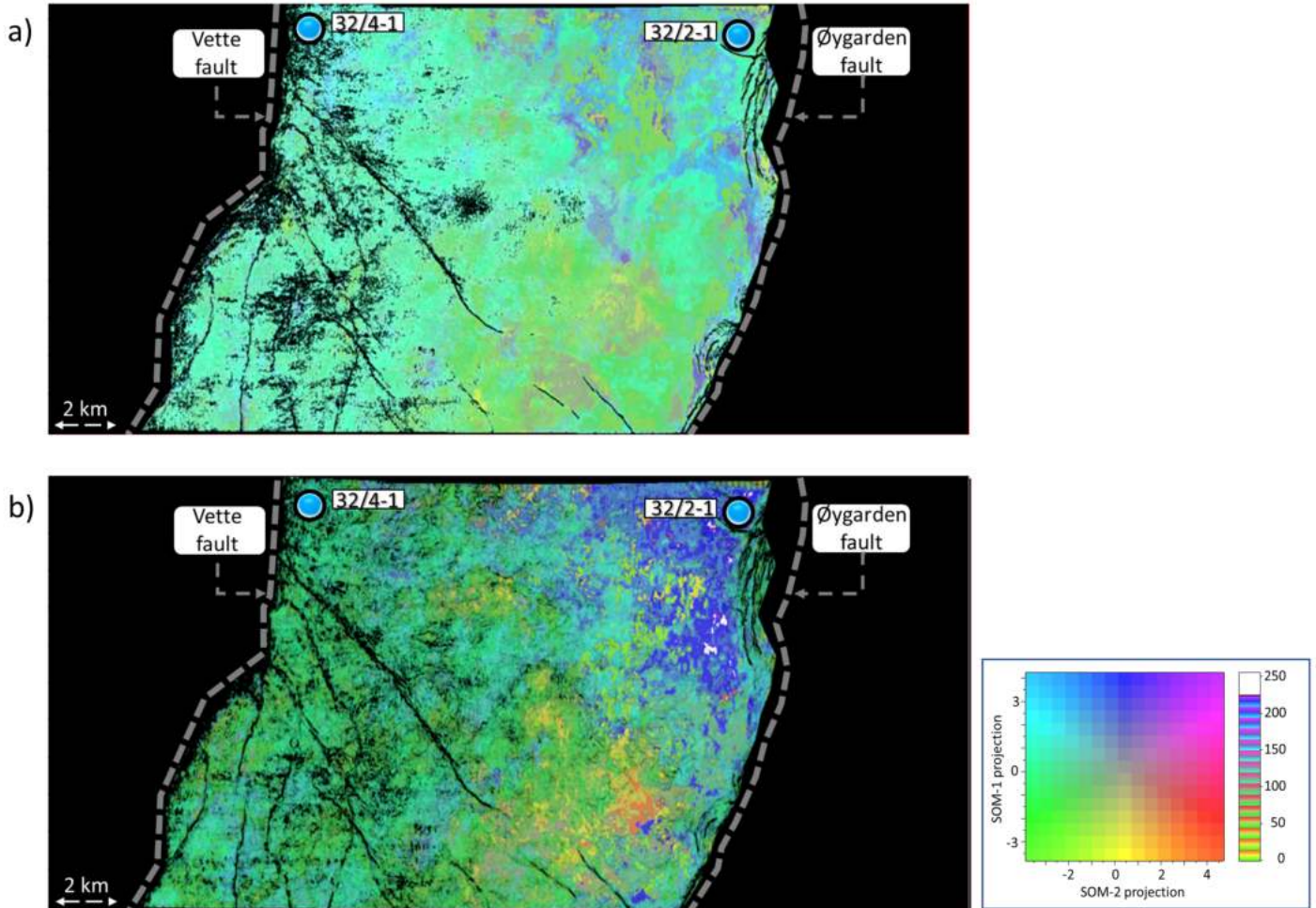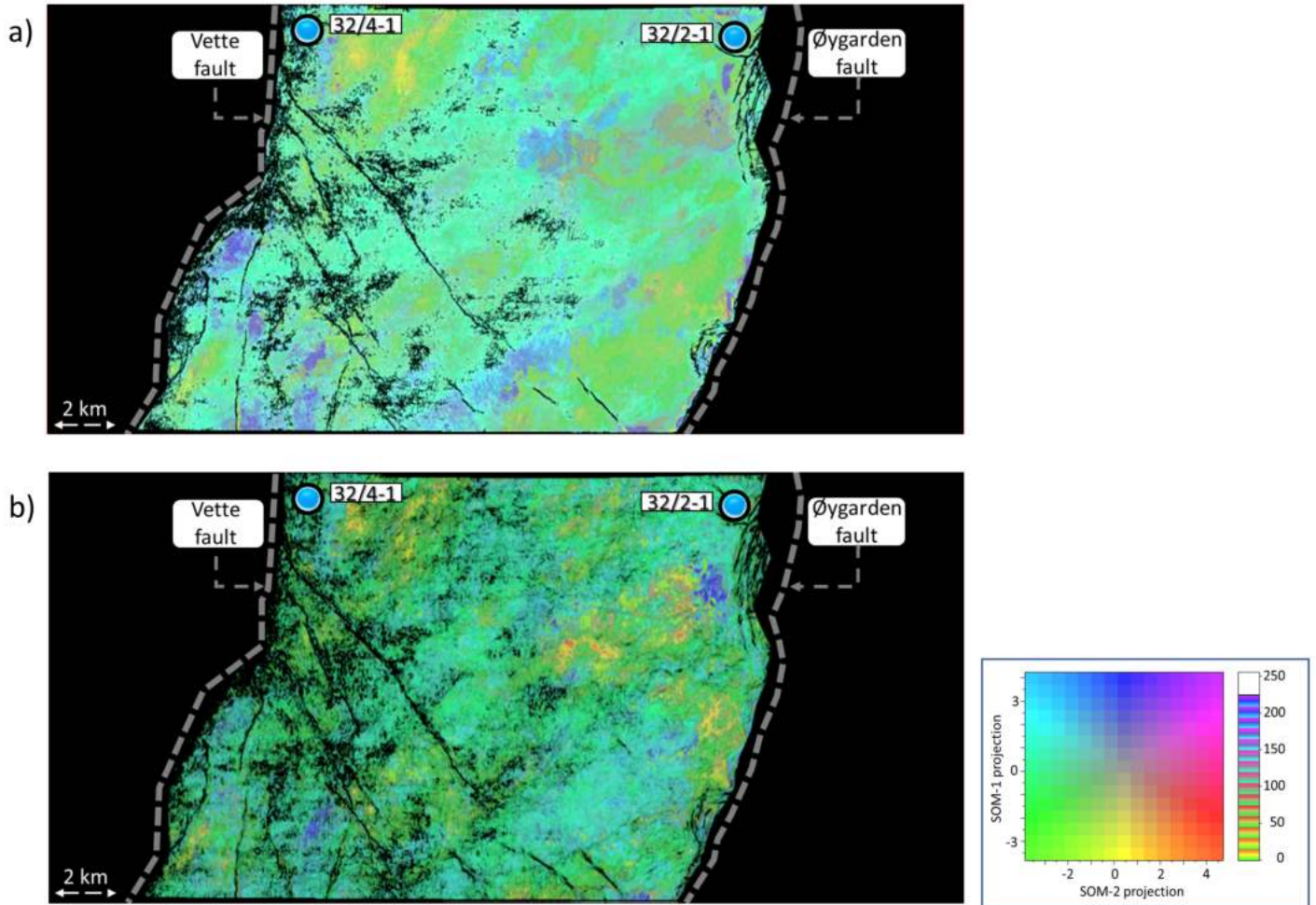
Figure 9: Stratal slice at 80 ms below the Sognefjord marker (within Sognefjord Fm) extracted from the SOM cross-plot volume computed on attributes generated on (a) input seismic data volume, and (b) bandwidth extended input seismic data volume. The two displays have been co-rendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). Only the target area between the Vette and Øygarden faults was classified and has been shown.
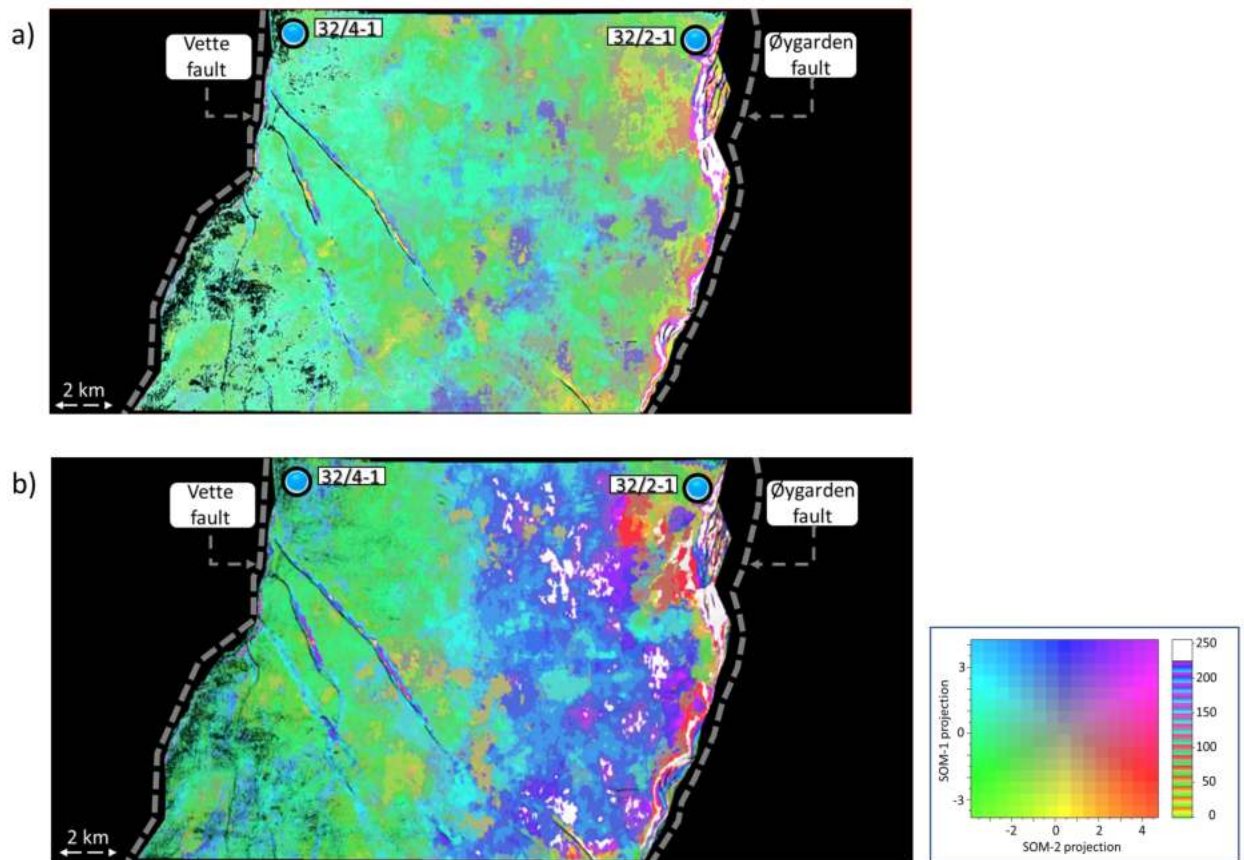
Figure 10: Stratal slice at 136 ms below the Sognefjord marker (within Fensfjord Fm) extracted from the SOM cross-plot volume computed on attributes generated on (a) input seismic data volume, and (b) bandwidth extended input seismic data volume. The two displays have been co-rendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). Only the target area between the Vette and Øygarden faults was classified and has been shown.

Figure 11: Stratal slice at 228 ms below the Sognefjord marker (within Krossfjord Fm) extracted from the SOM cross-plot volume computed on attributes generated on (a) input seismic data volume, and (b) bandwidth extended input seismic data volume. The two displays have been co-rendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). Only the target area between the Vette and Øygarden faults was classified and has been shown.

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. 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 a similar dimensional

non-Euclidean curved surface as a corresponding vector (m_k) embedded into different dimensional data space in GTM. Each data vector (x_k) mapped into this space is modeled as a suite of Gaussian probability density functions centered on these reference vectors (m_k). The components of the Gaussian model are then iteratively made to move toward the data vector that it best represents. 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.
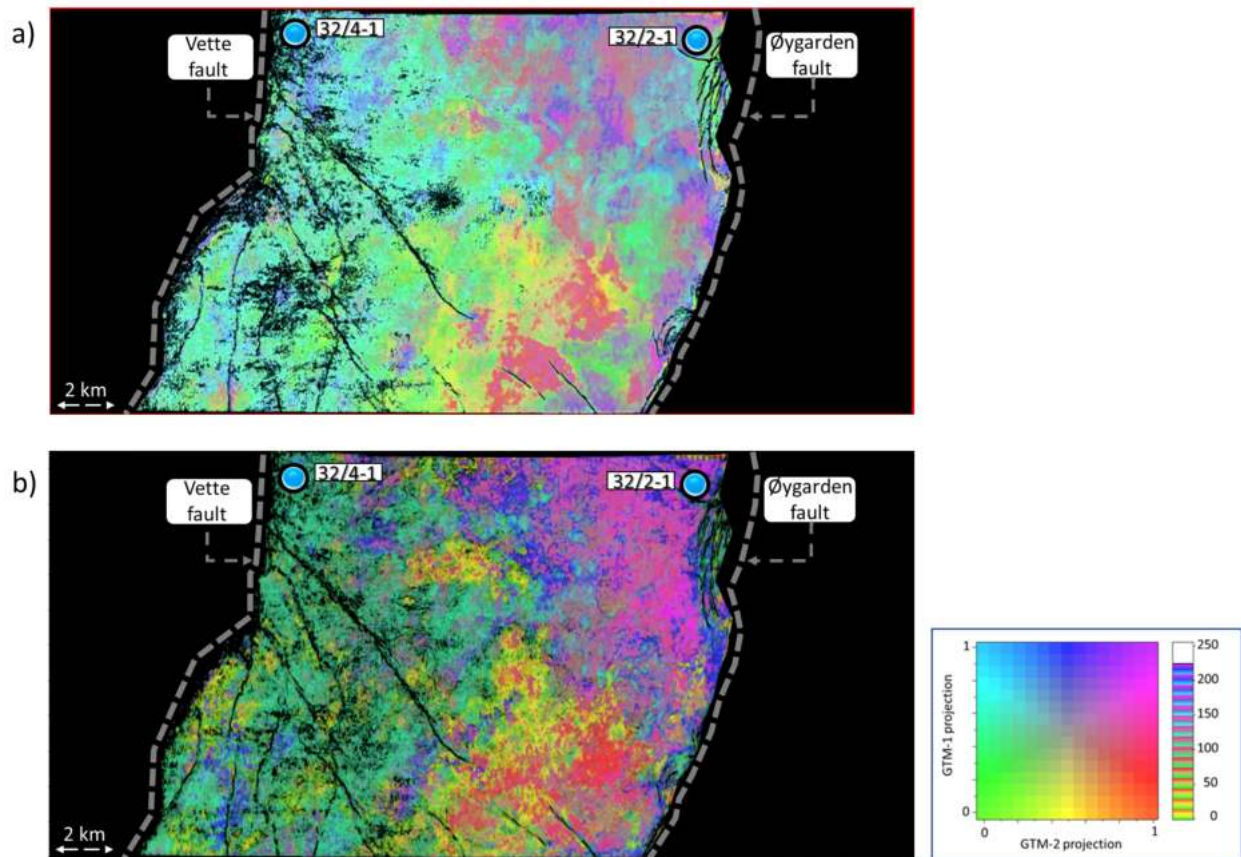

Figure 12: Stratal slice at 80 ms below the Sognefjord marker (within Sognefjord Fm) extracted from the GTM cross-plot volume computed on attributes generated on (a) input seismic data volume, and (b) bandwidth extended input seismic data volume. The two displays have been co-rendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). Only the target area between the Vette and Øygarden faults was classified and has been shown.
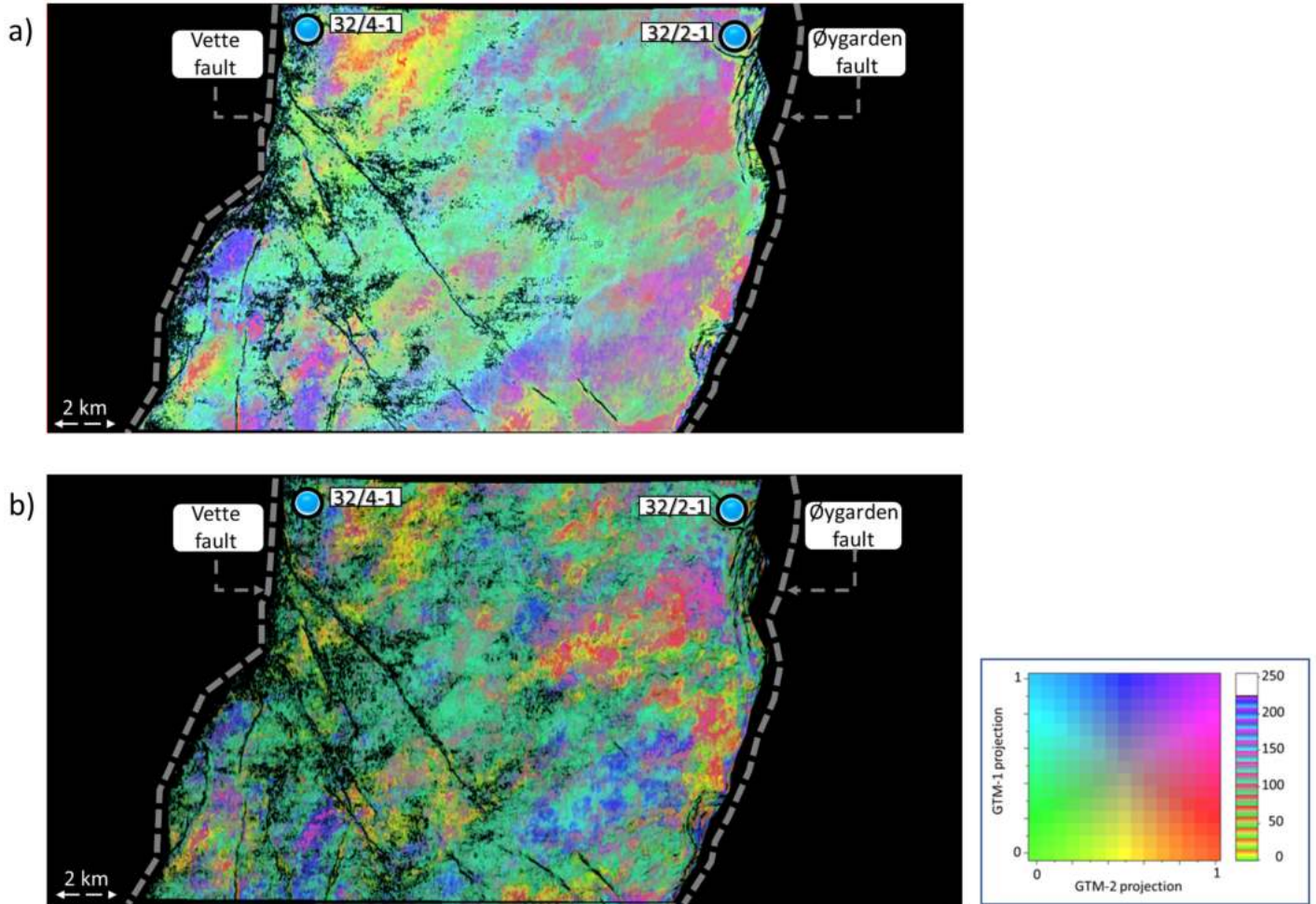
Figure 13: Stratal slice at 136 ms below the Sognefjord marker (within Fensfjord Fm) extracted from the GTM cross-plot volume computed on attributes generated on (a) input seismic data volume, and (b) bandwidth extended input seismic data volume. The two displays have been co-rendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). Only the target area between the Vette and Øygarden faults was classified and has been shown.
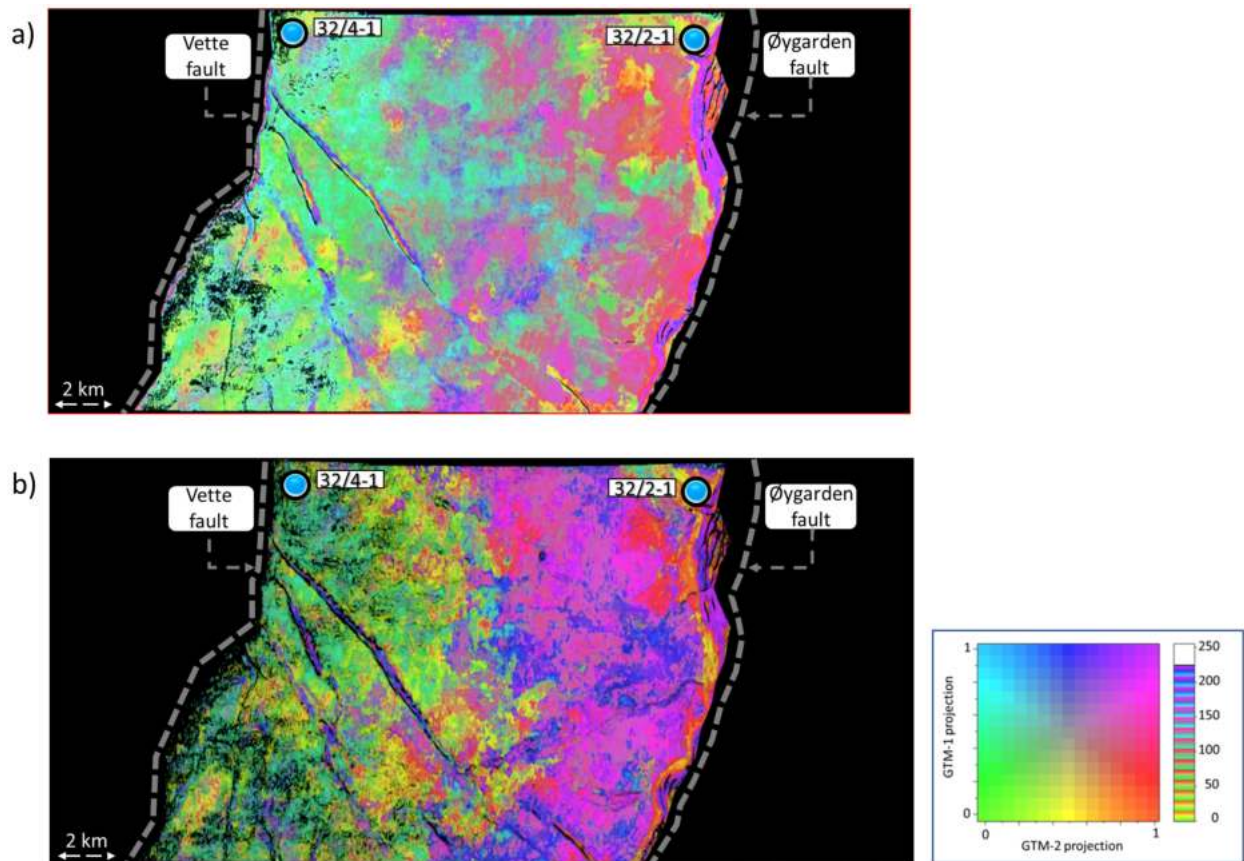
Figure 14: Stratal slice at 228 ms below the Sognefjord marker (within Krossfjord Fm) extracted from the GTM cross-plot volume computed on attributes generated on (a) input seismic data volume, and (b) bandwidth extended input seismic data volume. The two displays have been co-rendered with the respective multispectral energy ratio coherence attribute volumes. Better spatial resolution of the seismic facies is seen in (b) than in (a). Only the target area between the Vette and Øygarden faults was classified and has been shown.

In Figures 12 to 14 we show the displays equivalent to those shown for SOM analysis, where some of the clusters can be interpreted with ease with less background clutter and confusion.

In the comparison of the seismic facies displays generated for the original seismic data as well as the bandwidth extended version and shown above at different levels, we see almost no variation of facies carried out on the former, but a definite variation of seismic facies carried out on the latter from well 32/4-1 to the left to well 32/2-1 to the right. To ascertain such a variation, we picked up the available well data at the two wells, and carried out Bayesian facies classification (Grana, 2013). Figure 15a shows a cross-plot of porosity against gamma-ray well curves. The data points in blue come from well 32/4-1 and the yellow ones from well 32/2-1. Based on the cutoff values for porosity and gamma-ray, five coloured ellipses enclose clusters of points assigned based on Bayesian classification. When back-projected on the curves for the two wells, they highlight zones based on the colour of the cluster they represent. We notice that the red facies in well 32/4-1 appear as green in the same zone in well 32/2-1 in the Sognefjord,

Heather, Fensfjord and Krossfjord intervals, which suggests different facies in the two wells at these levels. This lends support to the facies variation we see after bandwidth extension of seismic data.
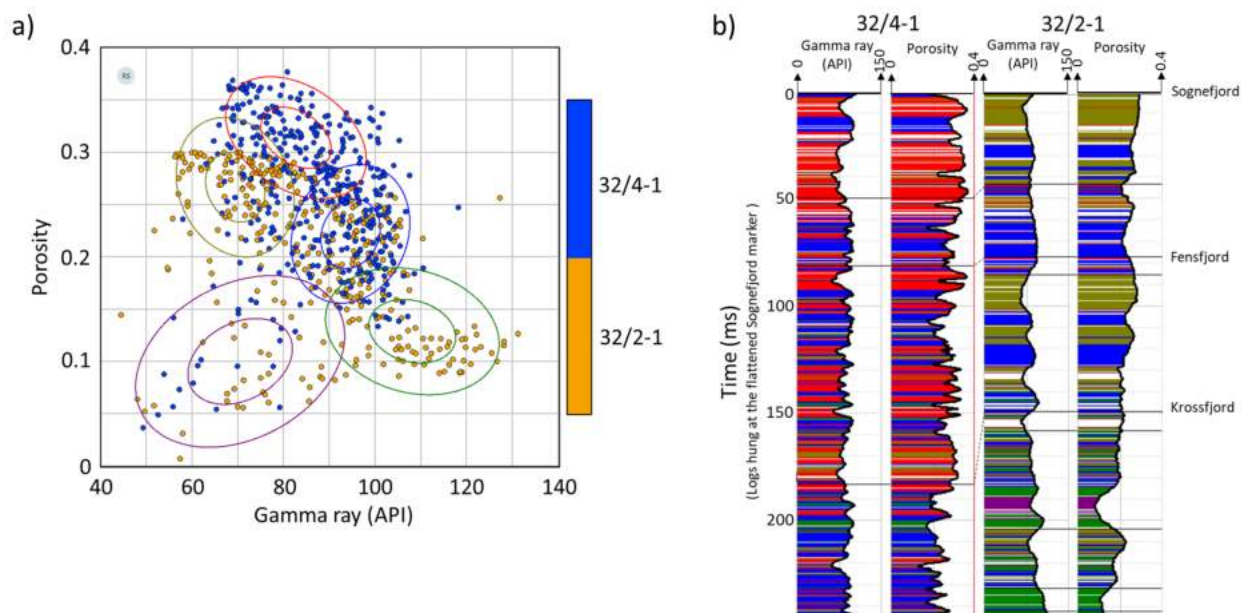

Figure 15: (a) Cross-plot between gamma ray and porosity for the interval between the flattened Sognefjord marker and about 240 ms below for wells 32/4-1 and 32/2-1. (b) The data points in blue come from well 32/4-1 and the yellow ones from well 32/2-1. Five coloured ellipses enclose clusters of points assigned based on Bayesian classification. When back-projected on the curves for the two wells they highlight zones based on the colour of the cluster they represent. We notice that the red facies in well 32/4-1 is seen more as green in the same zone in well 32/2-1 in the Sognefjord, Heather, Fensfjord and Krossfjord intervals.

Conclusions

We have found that bandwidth extension of the input seismic data when used for attribute generation and further used in some of the multi-attribute processes discussed in this paper can significantly aid in accurate interpretation of seismic facies. Results obtained for the unsupervised machine learning applications employing both the input seismic as well as its bandwidth extended version depict superior performance of the latter in terms of clarity of clusters as well as color variations within them. Such an observation seems to agree with the observation about variation in facies from the well to the left of the seismic volume to the well to the right.

Applications of 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 with GTM having an edge over SOM in terms of the detailed distribution of seismic facies with better resolution and distinct definition of the geologic features seen on the displays.

Usually, 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 bandwidth extension 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 the motivation for the work described in this paper.

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

Acknowledgements

References

Al-Dossary, S. and Marfurt, K. J., 2006, 3-D volumetric multispectral estimates of reflector curvature and rotation, Geophysics, **71**(5), 41–51. https://doi.org/10.1190/1.2242449

Bergbauer, S., Mukerji, T., and Hennings, P., 2003, Improving curvature analyses of deformed horizons using scale-dependent filtering techniques, AAPG Bulletin, **87**(8), 1255-1272. https://doi.org/10.1306/0319032001101

Bickel, S.H. and Natarajan, R. R., 1985, Plane-wave Q deconvolution, Geophysics, **50**(9), 1426-1439. https://doi.org/10.1190/1.1442011

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

Chopra, S., Castagna, J. P. and Portniaguine, O, 2006, Seismic resolution and thin-bed reflectivity inversion, CSEG RECORDER, **31**(1), 19-25.

Chopra, S., and K. J. Marfurt, K. J., 2007, Seismic Attributes for Prospect Identification and Reservoir Characterization, SEG.

Chopra, S. and Marfurt, K. J., 2010, Integration of coherence and volumetric curvature images. The Leading Edge, (9), 1092-1107, https://doi.org/10.1190/1.3485770

Chopra, S., and Marfurt, K. J., 2008, Gleaning meaningful information from seismic attributes, *First Break*, **26**, (9), 43–53, http://dx.doi.org/10.3997/1365-2397.2008012

Chopra, S., and Marfurt, K. J, 2019, Multispectral, multiazimuth, and multioffset coherence attribute applications: *Interpretation*, **7**(2), SC21-SC32. http://dx.doi.org/10.1190/INT-2018-0090.1.

Chopra, S., and Marfurt, K. J., 2018a, Coherence attribute applications on seismic data in various guises — Part 1: Interpretation, **6**(3), T521-T529. http://dx.doi.org/10.1190/INT-2018-0006.1.

Chopra, S., and Marfurt, K. J., 2018b, Coherence attribute applications on seismic data in various guises — Part 2: Interpretation, **6**(3), T531-T541. http://dx.doi.org/10.1190/INT-2018-0007.1.

Chung, H., and Lawton, D. C. 1995, Frequency characteristics of seismic reflections from thin beds, Canadian Journal of Exploration Geophysics, **31**(1-2), 32-37.

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.

Gersztenkorn, A., and Marfurt, K. J., 1999, Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping, Geophysics, **64**(5), 1468–1479. https://dx.doi.org/10.1190/1.1444651

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

Höcker, C. and G. Fehmers, 2002, Fast structural interpretation with structure-oriented filtering, The Leading Edge, **21**(3), 225-320. https://doi.org/10.1190/1.1463775

Kallweit, R. S., and Wood, L. C., 1982, The limits of resolution of zero-phase wavelets, Geophysics, **47**(7), 1035-1046. https://doi.org/10.1190/1.1441367

Kjartansson, E, 1979, Constant Q-wave propagation and attenuation, J. Geophys. Res., **84**(B9), 4737–4748. https://doi.org/10.1029/JB084iB09p04737

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

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

Lauritsen, H., Kassold, S., Menegholo, R. and Furre, A., 2018, Assessing potential influence of nearby hydrocarbon production on CO_2 storage at Smeaheia, Presented at EAGE Fifth CO_2 Geological Storage Workshop, 21-23 November 2018, Utrecht, The Netherlands Th CO2 05.

Li, F. Y., and Lu, W. K., 2014, Coherence attribute at different spectral scales, Interpretation, **2**, (1), SA99–SA10. http://dx.doi.org/10.1190/INT-2013-0089.1.

Liang, C., Castagna, J. P. and Torres,R. Z., 2017, Spectral bandwidth extension – invention versus harmonic extrapolation, Geophysics, **82**(4), W1-W16. https://doi.org/10.1190/geo2015-0572.1

Marfurt, K. J., and Matos, M., 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.

Puryear, C. I., and Castagna, J. P., 2008, Layer thickness determination and stratigraphic interpretation using spectral inversion: Theory and application, Geophysics, **73**(2), R37-R48. https://doi.org/10.1190/1.2838274

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**(1), SA31-SA47. https://doi.org/10.1190/INT-2013-0077.1

Smith, M., Perry, G., Stein, J., Bertrand, A., and Yu, G., 2008, Extending seismic bandwidth using the continuous wavelet transform, First Break, **26**(6), 97-102. https://doi.org/10.3997/1365-2397.26.1288.28410

Tonn, R., 1991, The determination of the seismic quality factor Q from VSP data: A comparison of different computational methods, Geophy. Prosp., 39, 1-27. http://doi:10.1111/j.1365-2478.1991.tb00298.x

Walden, A. T., and Hosken, J. W. J., 1985, An investigation of the spectral properties of primary reflection coefficients, Geophys. Prosp., **33**(3), 400–435. https://doi.org/10.1111/j.1365-2478.1985.tb00443.x

Widess, M. B., 1973, How thin is a thin bed? Geophysics, **38**(6), 1176-1180. https://doi.org/10.1190/1.1440403

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

Zhang, R., and Castagna, J. P., 2011, Seismic sparse-layer reflectivity inversion using basis pursuit decomposition, Geophysics, **76**(6), R147-R158. https://doi.org/10.1190/geo2011-0103.1

About the Authors

Satinder Chopra is the founder and President of SamiGeo Consulting Ltd., based in Calgary. He has 38 years of experience as a geophysicist specializing in processing, special processing, and interactive interpretation of seismic data. His research interests focus on techniques aimed at characterization of reservoirs. He has published eight books and more than 500 papers and abstracts. His work and presentations have won several awards from international professional societies the most notable ones being the 2021 Roy O. Lindseth CSEG Medal Award (2021), AAPG Distinguished Service Award (2019), EAGE Honorary Membership (2017), CSEG Honorary Membership (2014), Meritorious Service (2005) Award, 2014 APEGA Frank Spragins Award, the 2010 AAPG George Matson Award, and the 2013 AAPG Jules Braunstein Award. He has been the 2010/11 CSEG Distinguished Lecturer, the 2011/12 AAPG/SEG Distinguished Lecturer, and the 2014/15 EAGE e-Distinguished Lecturer.

Ritesh Kumar Sharma received his master's and Ph.D. degrees in geophysics from the University of Calgary. He has worked as senior reservoir geoscientist at TGS, Canada till 2020, and now works for SamiGeo Consulting Ltd. at Calgary. Ritesh has vast experience in working with applications such as AVO analysis (including azimuthal AVO), rock-physics analysis, frequency enhancement of seismic data, simultaneous inversion, extended elastic impedance inversion as well as geostatistical inversion. He has delivered many oral and poster presentations and published several papers in renowned journals. He has won the best poster award for his presentation at the 2012 GeoConvention held at Calgary, the Jules Braunstein Memorial Award for the best AAPG poster presentation at the 2013 AAPG Convention held at Pittsburgh, CSEG Honorable Mention for the Best Recorder Paper award in 2013 and is the recipient of Honorable Mention Best Poster Paper in SEG 2017. He is an active member of SEG and CSEG.

John P. Castagna specializes in exploration geophysics research and development. He is widely known for his work in direct hydrocarbon detection and reservoir characterization. John is a graduate of Brooklyn College, where he earned a B.S. in geology in 1976 and an M.S. in high-temperature geochemistry in 1981. He completed his doctoral degree in exploration geophysics at the University of Texas at Austin in 1983. During his career, he has held several positions at ARCO research and Vastar resources. In 2000 and 2010, he founded Fusion Geophysical and Lumina Geophysical, respectively. He was named a distinguished lecturer for SEG, delivering the fall lecture on "Applied AVO analysis: use and abuse of amplitude variation with offset." He has served SEG in various other capacities including chairman of The Leading Edge editorial board, first vice-president, and technical program chairman for the 2003 Annual Convention in Dallas, and associate editor for Geophysics. He has written two book, Offset-dependent-reflectivity: Theory and practice of AVO analysis, and AVO. He currently holds the Robert Sheriff Chair of Geophysics at the University of Houston.

Kurt J. Marfurt received a Ph.D. (1978) in applied geophysics from Columbia University's Henry Krumb School of Mines in New York, where he also taught as an assistant professor for four years. He joined the University of Oklahoma (OU) in 2007, where he served as the Frank and Henrietta Schultz professor of geophysics within the ConocoPhillips School of Geology and Geophysics and is now Emeritus Professor there. He worked for 18 years in a wide range of research projects at Amoco's Tulsa Research Center, after which he joined the University of Houston for eight years as a professor of geophysics and the director of the Allied Geophysics Lab. He has received the following recognitions: SEG best paper (for coherence), SEG best presentation (for seismic modeling), as a coauthor with Satinder Chopra for best SEG poster (for curvature) and best AAPG technical presentation, and as a coauthor with Roderick Perez-Altamar for best paper in Interpretation (on a resource play case study). He also served as the SEG/EAGE Distinguished Short Course Instructor for 2006 and 2018.