Using machine learning as an aid to seismic geomorphology, which attributes are the best input?

Lennon Infante-Paez and Kurt J. Marfurt

Abstract

Volcanic rocks with intermediate magma composition indicate distinctive patterns in seismic amplitude data. Depending on the processes by which they were extruded to the surface, these patterns may be chaotic, moderate-amplitude reflectors (indicative of pyroclastic flows) or continuous high-amplitude reflectors (indicative of lava flows). We have identified appropriate seismic attributes that highlight the characteristics of such patterns and use them as input to self-organizing maps to isolate these volcanic facies from their clastic counterpart. Our analysis indicates that such clustering is possible when the patterns are approximately self-similar, such that the appearance of objects does not change at different scales of observation. We adopt a workflow that can help interpreters to decide what methods and what attributes to use as an input for machine learning algorithms, depending on the nature of the target pattern of interest, and we apply it to the Kora 3D seismic survey acquired offshore in the Taranaki Basin, New Zealand. The resulting clusters are then interpreted using the limited well control and principles of seismic geomorphology.

Introduction

In today’s modern era, the most effective way to gain knowledge on how a certain geologic feature such as a lava flow appears in seismic data is to do a Google search and type a few key words such as “lava flow seismic” then go to the images section and even go through a couple of scientific publications, until we reach an “aha moment” when we find patterns that appear similar to those in our data set. This type of pattern recognition is easy for a human interpreter but is quite difficult for computers. The advantage of computers is that once such a task is well-defined, they can apply the analysis to every voxel in a large 3D seismic data volume, perhaps identifying subtle features that may have been overlooked by an overworked interpreter. Machine learning pattern recognition of seismic data goes beyond automation of time-consuming analysis tasks. Specifically, each prediction can be weighted by a confidence value, which can be used in subsequent risk analysis.

Machine learning was first used by Alan Turing to decipher the Nazi “enigma” code (Gunderson, 1964). Since then, it has branched out to nearly all forms of language analysis, including voice recognition and translators, and it has expanded into visual communication, marketing, and social media. Early machine learning applications to seismic facies analysis include work by Meldahl et al. (2001) and West et al. (2002), who use multilinear feed-forward neural networks with seismic attributes to produce a probability volume of gas chimneys. Linari et al. (2003), Coleou et al. (2003), and Popupon et al. (2004) use seismic amplitude waveform and self-organizing maps (SOMs) to define zones of interest. Similarly, Verma et al., (2012), Roy et al. (2013), Roden et al. (2015), and Zhao et al. (2016) use volumetric seismic attributes fed into SOM algorithms to find different facies in shale resources plays. Qi et al. (2016) and Olorunsola et al. (2016) use generative topographic mapping (GTM) to try to separate salt from clastic, mass transport deposits (MTDs) from layered sediments in the Gulf of Mexico, and producing from tight facies in the Granite Wash in the Texas Panhandle, respectively. Lubo-Robles (2018) applies independent component analysis of spectral components to try to predict sandy facies in the Miocene Moki A Formation in the Taranaki Basin, New Zealand.

Machine learning techniques are relatively simple mathematical algorithms that can learn from and generate clusters/classes based on patterns in (or interrelationships between) the data. Depending upon data availability, we can use either supervised or unsupervised algorithms. In supervised classification, the interpreter defines facies of interest, either by selecting specific voxels (Meldahl et al., 2001) or by drawing polygons around facies of interest (West et al., 2002; Qi et al., 2016), which serve as “training data” that are used to establish the relationship between input and output. Once trained, the algorithm is then applied to another subset of “validation data” (interpreted facies not used
in the training step) to determine if the algorithm is sufficiently accurate. If the validation is successful, the algorithm is then applied to the entire seismic data volume.

In principle, unsupervised classification requires no interpreter input. In practice, the interpreter strongly biases the results of the algorithm by choosing a suite of attributes that best differentiate facies of interest. In a seismic interpretation context, this machine learning technique extracts patterns that exhibit a similar attribute expression for similar geologic facies, displaying these similar expressions, or clusters, using a 2D color-coded palette to allow subtle patterns to be identified by the interpreter (Qi et al., 2016; Zhao et al., 2016, 2017).

Depending on the objective, supervised and unsupervised techniques use seismic attributes as input, in which the impedance and anisotropy attributes provide critical information for geomechanical clustering. In the absence of sufficient well control, instantaneous, geometric, spectral, and texture attributes provide critical information for interpreting seismic geomorphology from clustering (Zhao et al., 2016; Infante-Paez and Marfurt, 2017; Infante-Paez, 2018).

Most recent studies in seismic interpretation have been focused on applying and comparing different machine learning methods, such as the multilayer perceptron network, SOMs, the support vector machine, K-means, and GTMs (Meldahl et al., 2001; Roy et al., 2013; Qi et al., 2016; Snyder, 2016; Zhao et al., 2016).

We begin this study by defining the nature of the seismic patterns represented by volcanics in our seismic volume. We then propose a workflow that will allow interpreters to decide what machine learning algorithm to use, depending on the nature of the target pattern (TP). Next, we compute mathematically independent candidate attributes that highlight the continuity (such as gray-level co-occurrence matrix [GLCM] entropy), amplitude (peak spectral magnitude), and frequency (peak spectral frequency) of these TPs, with the goal of determining which input attributes best differentiate the volcanics from the surrounding clastic sediments. Finally, we input the GLCM entropy, peak spectral magnitude, and frequency attributes into the SOM, to interpret the seismic geomorphology of the internal elements of the Kora volcano.

**Methodology**

**Selection of the TPs**

The TPs in our study include some of the internal and external elements of the Kora volcano, as well as adjacent volcanics from the Mohakatino Volcanic Belt (MVB). These volcanics form potential analogs to the volcanics in the Songliao Basin, China (Figure 1) and andesites from the Jatibarang field in Java (Figure 2), which have produced more than 1.2 billion barrels of oil and >2.7 trillion cubic feet of gas between 1969 and 1990 (Kartanegara et al., 1996).

Figure 3 displays a vertical slice through the Kora 3D survey, where multiple TPs are highlighted by yellow boxes. Seismic-to-well ties indicate that these patterns have been drilled by exploration wells (Figure 4) validating the presence of volcanics.

**Nature of the TPs**

We define our human interpretation patterns as “monogenetic,” “composite,” and “intricate” patterns in which the goal is to examine relationships that can be evaluated by a machine.

**Monogenetic seismic patterns**

We define a monogenetic pattern as a facies that consists of a single seismic pattern. This pattern is statistically consistent, translational vertically and horizontally. The pattern is also consistent at different scales, such as conformal or chaotic reflectors within a 20 × 20 × 20 versus a 5 × 5 × 5 voxel window. Monogenetic seismic patterns are related to physical self-similarity, where the appearance of objects does not change at different scales of observation (Lam and Quattrochi, 1992; Dimri et al., 2011; Dasgupta, 2013; Herrera et al., 2017). Examples of monogenetic seismic patterns are shown in Figure 5.

**Composite seismic patterns**

Composite patterns are those facies that consist of two or more simpler patterns. Composite patterns do not entirely
preserve their character laterally, vertically, or at different scales, but they can still be distinguished from surrounding patterns (e.g., Figures 6 and 7).

**Intricate seismic patterns**

Intricate patterns are those facies that dramatically change their character with scale and location, and they

---

**Figure 2.** Comparison of the size and styles of (a) the Jatibarang field in Java that produced more than 1.2 Bbl. from fractured andesites (after Kalan et al., 1994) and (b) the Kora and adjacent volcanics. Even though the structural style in the Kora area is not exactly the same when compared with Jatibarang, the volcanics in Kora can potentially be fractured due to the radial and major rift faults.

**Figure 3.** (a) Composite vertical slice through the Kora 3D survey in which multiple patterns associated with the extrusive Kora volcano are highlighted by yellow boxes. (b) Expanded image of these four patterns and their descriptions. TP, target pattern. Seismic data courtesy of NZP&M.
are far from being self-similar; see, for example, Figure 8.

We hypothesize that using appropriate seismic attributes as input to a machine learning algorithm (e.g., SOM), monogenetic patterns will be represented by a single cluster. Composite patterns will be represented by two or more clusters, resulting in a machine learning classification that requires subsequent human "clumping." Intricate patterns are represented by multiple clusters, providing an image that may offer little value over the original seismic amplitude volume. Although composite and intricate patterns may be represented by more than one cluster, a given “cluster” may also represent more than one facies. For example, clustering based on reflector continuity and parallelism might result in marine shales and deepwater fans clumped together. To separate them, the interpreter might add energy or peak frequency as additional input data to break them apart. For this reason, if the multiple clusters representing an intricate pattern are unique, they can be subsequently clumped after clustering to form a single facies. Such clumping, which is desired (depending on the pattern of interest) is performed implicitly when computing the Bhattacharyya distance when using generative topologic mapping (Qi et al., 2016) where a probability density function, rather than a single prototype vector is computed for each voxel in the training data set. The sum of these PDFs can then represent more intricate patterns.

Convolutional neural network (CNN) may provide an alternative means to addressing intricate patterns. In the simplest workflow, the interpreter provides the original seismic amplitude data. Internally, the machine convolves adjacent voxels, computing its own attributes for evaluation. Alternatively, Qi (2018) uses CNN and a suite of input seismic attributes to predict fractures seen in image logs.

Computers have several advantages over humans: (1) They can perform repetitive analysis of billions of voxels without tiring, (2) they can be much more quantitative, and (3) they can easily compare similarities and differences among more than three attributes at the same voxel. In contrast, humans have advantages over

---

**Figure 4.** Wells that penetrate volcanoes and volcaniclastics and their corresponding seismic expression in the Kora 3D seismic survey. Track 1 displays the caliper log, track 2 displays the gamma ray, and track 3 displays the density porosity log. Note the low gamma-ray response in all of the wells consistent with intermediate versus felsic-magma composition (after Infante-Paez and Marfurt, 2017).
machines in that they can (1) see patterns on a much larger scale, thereby identifying a pattern in context, (2) compare patterns to those seen in others seismic surveys or in outcrop, and (3) integrate patterns as discrete components or elements that result from a geologic process.

Figure 5. Examples of approximately self-similar monogenetic seismic patterns TP 1 and TP 2. Note the scale bar in the expanded boxes. These patterns are relatively easy for a machine to identify. Seismic data are courtesy of NZP&M.

Figure 6. Example of a composite seismic pattern. Note the different scales in the expanded boxes show TP 3. The black dotted polygon in the time slice shows the extension of TP 3. TP, target pattern. Seismic data are courtesy of NZP&M.
Figure 7. Examples of composite seismic pattern. Note the different scales in the magnified boxes. The black dotted polygon in the time slice shows the extension of TP 4. TP, target pattern. Seismic data are courtesy of NZP&M.

Figure 8. Example of an intricate seismic pattern. Note that the character of the pattern is not the same at different scales. The pattern also changes laterally from flatter, converging to more steeply dipping, subparallel reflectors. This pattern is more difficult for a machine to identify. Seismic data are courtesy of NZP&M.
Seismic attribute selection

Figures 9, 10, and 11 demonstrate the importance of the input attributes to unsupervised machine learning. The goal is to try to organize the people in Figure 9a, who work at a university and determine which of them perform similar jobs. From the top and moving clockwise, we have the dean of the Mewbourne College of Earth and Energy Dr. J. Mike Stice, Ph.D. candidate in geophysics Lennon Infante, geophysics professor Dr. Kurt. J. Marfurt, and geology professor Dr. Roger Slatt. From their headshots, we can extract additional information, such as their hair length and smile (happiness). The fact that one of them is dean of the College of Earth and Energy, two of them are professors, and the last one is a student, suggest they have different incomes. These three attributes, the happiness, hair length, and income, provide the means to place the two professors in the same cluster (Figure 9b). Although the SOM put these four people into three different clusters, it cannot tell which cluster represents which job.

In reality, we cannot measure income from the input data (headshots). A more realistic scenario would be the one shown in Figure 10. Some of the attributes that can be extracted from the input data are gender, hair length, clothes type, and happiness (smile on their faces). When the gender, clothes, and hair length attributes are fed into the SOM, we obtain three clusters, which is the correct number of different jobs. However, one of the clusters is erroneous because it groups Professor Marfurt with Dean Stice, who have different jobs. Selecting the happiness attribute instead of hair length produces different clusters (Figure 11). In this case, the SOM outputs only one cluster. From this analogy, it is clear that the input attribute selection and not the SOM algorithm itself produces erroneous results.

Voxel-based approach for classifying monogenetic seismic patterns

Given that the approximately self-similar TPs of interest (Figure 4) preserve their character at different scales and distinct locations, we use the workflow described in Figure 12 to select attributes that differentiate the volcanics from each other and from the surrounding clastic sediments.

There are different approaches that a seismic interpreter can use to select the input seismic attributes for clustering of seismic facies. A simple but time-consuming and potentially dangerous approach is to apply all possible attributes and determine which combination best correlates with the desired facies. Kalkomey (1997) warns that this workflow may lead to false predictions. Principal component analysis reduces a suite of correlated attributes into a smaller number of composite attributes. Roden et al. (2015) use the first principal components to determine which attributes are most important in representing the seismic data volume. Unfortunately, such choices do not necessarily guarantee the differentiation of the pattern of interest, particularly if one or more of these patterns only represent 1%-5% of the data. Moreover, this approach may be suitable as a first pass for exploring the data, but it could fail when trying to isolate a specific pattern such as an MTD or pyroclastic flows. Thus, we recommend the calculation of seismic attributes based on the qualitative description of the patterns (Figure 13) (analogous to

Figure 9. Headshots of four people who work at a university. (a) The input data are three attributes that somehow distinguish them — hair length, happiness, and income. (b) Using this combination of attributes, the machine learning algorithm correctly clusters the two professors into the same group, and the student and the dean are in their own separate group. Note the three clusters in which the two professors form one cluster and the student and the dean form two different clusters.

Figure 10. Headshots of the same four people shown in Figure 9 in which the input attributes are hair length, gender, and dress code. All four samples are male and have similar dress code (wearing a tie). For this reason, the clustering is driven by hair length alone, misclustering one of the professors with the dean.
how a geologist describes facies in outcrops or in core). Because most commercial and research software that implement machine learning techniques do what they are supposed to do, e.g., organize the data into clusters, the challenge for interpreters in applying SOM and similar algorithms to seismic data is the attribute selection. For example, TP 3 in Figure 6 is characterized by semichaotic, low- to moderate-amplitude reflectors with occasional isolated continuous, moderate-amplitude reflectors that are parallel. During the description process, key words such as “chaotic reflectors” can help us think of seismic attributes that best highlight such features (such as the GLCM texture entropy attribute). TP 1 is characterized by continuous high-amplitude reflectors. TP 2 is characterized by chaotic, moderate-amplitude reflectors, whereas TP 3 and TP 4 are characterized by semicontinuous to semichaotic, moderate- to high-amplitude reflectors with few isolated parallel reflectors. Attributes that measure such observations include the amplitude attributes (envelope, root-mean square [rms] amplitude, energy, and colored inversion), continuity attributes (GLCM entropy, chaos, and coherence), frequency attributes (the peak frequency, average frequency, and bandwidth), and conformity attributes (reflector convergence and parallelism).

Figure 11. The same “samples” as in Figures 9 and 10 in which the attributes are clothing, gender, and happiness. In this example, the SOM results in only one cluster, suggesting a happy conformity in this university, but no indication of the role occupied by each person.

Figure 12. Proposed workflow to decide clustering of the patterns approach in seismic data.
Spectral components are also helpful, but they are more difficult for a human interpreter to assign to a given seismic pattern. Seismic patterns exist at zones (geologic age), so one can constrain the attribute analysis within a zone of interest bounded by seismic horizons to minimize variability not only to geology but also to seismic resolution.

**Seismic attributes that assist interpreters versus seismic attributes that assist machine learning**

For monogenetic and composite seismic patterns to be successfully clustered using the voxel approach, all of the input attributes should highlight the same sample (voxel) in the seismic volume. Therefore, we must differentiate among attributes that assist the interpreter in highlighting key geologic features (Figure 14) and attributes that assist machine learning algorithms to isolate specific geologic features (Figure 15). That is, if the target seismic features to map are faults, one should avoid using input attributes to machine learning such as the most-positive and most-negative curvatures and coherency together. Though they do highlight the fault in a visual way that is clear to a human interpreter, they do not highlight the fault at the same seismic sample (voxel). The same principle applies to different facies such as sinkholes. If one is trying to isolate sinkholes using a clustering technique such as SOM, feeding complementary attributes such as the most-positive and most-negative curvatures and coherency would not produce an accurate result because these attributes highlight various parts of the sinkhole but not at the same voxel. The most obvious example is combinations of spectral components, which differentiate thicknesses and lithologies within a channel system, and coherency, which highlights the channel edges. Seismic noise also results in patterns that may be mentally “filtered out” by a human interpreter but form (ideally, its own) noise clusters.

**Seismic attribute expression of volcanic rocks that assist machine learning clusters.**

Based on the descriptions (e.g., continuous versus chaotic, low-amplitude versus high-amplitude) of the TP, the input seismic attributes for clustering of

**Figure 13.** Proposed workflow to decide which seismic attributes to select for machine learning.

**Figure 14.** Cartoon showing a normal fault and its attribute response. Such red-yellow and blue patterns are easily recognized by a human interpreter. However, because they occur at laterally shifted voxels, they are more difficult to interpret for a machine. The most-positive curvature $k_1$ (in red) illuminates the footwall, the most-negative curvature $k_2$ (in blue) illuminates the hanging wall, whereas coherency (in yellow) illuminates the fault discontinuity (after Qi, 2018).

**Figure 15.** Cartoon showing the same normal fault as in the previous figure. The attributes are coherence, dip magnitude, and aberrancy (Qi, 2018). These three attributes image the fault at the exact same location (voxel) and are therefore amenable to machine learning for clustering (after Qi, 2018).
these patterns would be three types of attributes that highlight the amplitude, continuity, and frequency content. Using the same composite section from Figure 3 as a reference, we compute a suite of candidate attributes to visually evaluate the attribute response of the TPs (Figures 16, 17, 18, 19, 20, 21, and 22). Figure 17 shows the peak spectral magnitude calculated from the continuous wavelet transform (CWT). This attribute gives a response similar to the rms amplitude and highlights the strength of the reflectors. The peak spectral magnitude shows that there are differences in all TPs. TP 1 is characterized by high magnitude, whereas TP 2 is highlighted by low magnitudes. However, TP 3 and TP 4 consist of low to moderate magnitude.

The GLCM provides a group of texture attributes: homogeneity, entropy, dissimilarity, contrast, mean, energy, correlation, and variance. Hall-Beyer (2007) defines texture as “an everyday term relating to touch that includes such concepts as rough, silky, and bumpy. When a texture is rough to the touch, the surface exhibits sharp differences in elevation within the space of your fingertip. In contrast, silky textures exhibit very small differences in elevation.” Seismic textures work in an analogous manner with elevation replaced by amplitude, and the probing of a finger replaced by a rectangular or elliptical analysis window oriented along the structure. From these eight attributes, the most useful are entropy and homogeneity (see, Gao, 2003; Qi et al., 2016; Gao et al., 2017; Zhao et al., 2017; Marfurt, 2018), although they are somewhat coupled. Detailed examination of the entropy and homogeneity of TP 1 and TP 2 (Figures 18 and 19) shows that TP 1 displays moderately low values of entropy and moderately high values of homogeneity. The opposite is true for the TP 2, suggesting that we can separate these two patterns using these texture attributes. TP 3 and TP 4 display zones where the entropy is high to moderate. High values of entropy mean that the reflectors are chaotic (not layer cake). Figure 20 shows the peak spectral frequency attribute, which displays the dominant frequency for the entire section. In the volcanic sequence, the peak spectral frequencies range between 40 and 50 Hz. Although TP 1, TP 3, and

![Figure 16. The same vertical amplitude slice as in Figure 3. The color scale has been changed to black and white to facilitate corendering with seismic attributes. The yellow boxes represent extrusive volcanics and volcaniclastics. Seismic data are courtesy of NZP&M.](image1)

![Figure 17. Vertical amplitude slice corendered with peak spectral magnitude. The same seismic section as in Figure 3. The yellow boxes represent extrusive volcanics and volcaniclastics. Seismic data are courtesy of NZP&M.](image2)

![Figure 18. Vertical amplitude slice corendered with GLCM entropy. The same seismic section as in Figure 3. The yellow boxes represent extrusive volcanics and volcaniclastics. Seismic data are courtesy of NZP&M.](image3)
TP 4 exhibit a similar range of frequencies, TP 2 is characterized by low to mid frequencies (15–25 Hz).

In the same way, Figures 21–23 show the magnitude of reflector convergence, the dip magnitude, and the coherence attribute, all corendered with the seismic amplitude, highlighting different aspects of the seismic patterns of interest, but not at the same voxel. Therefore, for the Kora 3D survey, we conclude that the attributes that would assist machine learning are (1) the texture attributes, which are a measure of continuity, (2) peak spectral magnitude, which measure the strength of the reflectors, and (3) peak spectral frequency, which measures the dominant frequency. These attributes are mathematically independent, but coupled through the geology, making them candidates for SOM.

**SOMs and seismic geomorphology**

In seismic interpretation, the SOM is a clustering technique that extracts similar patterns across multiple seismic attribute volumes and displays those similarities as a color-coded map, with similar colors representing clusters that a human interpreter can visualize as similar facies (Zhao et al., 2016). Because several of the attributes that we use (GLCM entropy and homogeneity, peak magnitude, and peak frequency) measure spatial patterns, SOM will be able to cluster spatial patterns as well. The SOM workflow used in this study is shown in Figure 24. The input attributes to feed the SOM are of three types: attributes that highlight the continuity — how layer cake the reflectors are (homogeneity and entropy), the amplitude (peak magnitude), and the frequency (peak frequency) of the TPs. These attributes are extracted from the raw amplitude data using software developed at the University of Oklahoma (Matos et al., 2011; Qi et al., 2016).

We input the previously computed seismic attributes into the SOM algorithm. Because we are using four attributes at each voxel, the analysis is in the 4D attribute space. In this case, the objective of SOM is to fit a deformed 2D surface (called a manifold) to the distribution of the data points living in 4D space. Each data point is projected onto the nearest part of a manifold, which is then mapped to a 2D color bar. In this manner, voxels that exhibit a similar response (they lie next to each other in 4D space) project onto nearby locations on the manifold and are displayed as similar colors. In contrast, voxels that exhibit...
a very distinct attribute behavior (they lie far from each other in 4D space) project onto different parts of the manifold and appear as different colors. Details of SOMs applied to seismic data can be found in Roy et al. (2013), Roden et al. (2015), and Zhao et al. (2016).

Internal elements of the Kora volcano

Integrating well reports from Kora-1, -2, and -3 (where core data were acquired) and their seismic patterns, provide geologic control to the chaotic moderate-amplitude seismic pattern. Internally to the Kora volcano, the seismic data exhibit two main patterns: strong continuous (TP 1) and moderate chaotic reflectors (TP 2). According to the core data and seismic-to-well ties from wells Kora-1 through 3, we interpret that the penetrated chaotic, moderate-amplitude reflectors correspond to pyroclastic flows whereas the geologic process corresponding to the strong, continuous reflectors remains unknown.

Figure 25 shows a vertical slice connecting the four Kora wells illustrating the distribution of TP 1 and TP 2.

External elements of the Kora volcano and adjacent volcanoes

Subaqueous flows

Volcanic eruptions allow the volcano to grow. A coned geometry such as the one observed in the Kora and nearby volcanoes suggests that the volcanoes grew by preferential addition of material to the summit area (Magee et al., 2013). Furthermore, a discontinuous to semi-chaotic region with short (100 m) continuous reflectors, creating a distinctive seismic pattern (TP 3) can be seen as far as 20 km to the northwest from the Kora volcano. Fortunately, this and other similar seismic patterns have been penetrated by several wells offshore Taranaki Basin, including the Ariki-1, Arawa-1, Kanuka-1, Moana-1, and other wells (well report series). Well-to-seismic ties coupled with completion reports indicate that this seismic pattern is representative of sediments with significant volcaniclastic content called the Mohakatino Formation named by Hansen and Kamp (2004). The Ariki-1 well drilled through semicontinuous, semichaotic seismic patterns (Figure 4) similar to the MTDs previously documented by Posamentier and Kolla (2003), Lee et al. (2004), Dallas et al. (2013), and Qi et al., (2016) in steep slope clastic environments. In our seismic survey, the MTDs seem at least partially originated from the west flank of the Kora volcano (Figure 28a and 28b). The Ariki-1...
well completion report defines cuttings from the interval 2256 to 2556 m as volcaniclastic deposits, in which “Tuffaceous material is most abundant in the lower part of the formation and decreases upwards, reflecting the waning of volcanic material.” The lithology varies from gray mudstones with a tuffaceous matrix including biotite, chlorite, pyrite, and at the base, light- to moderate-gray sandy tuffs. These tuffs contain abundant angular to subangular, fine- to medium-grained, poorly sorted clasts of biotite, garnet, olivine, hornblende, and aphanitic material together with quartz and feldspar of sedimentary and volcanic origin. Comparable descriptions are also given for the cuttings of the other wells that drilled through similar seismic patterns in the Arawa-1, Kanuka-1, and Moana-1 well series reports. Given the morphology of TP 3 (Figure 28a), it is possible to infer a depositional process. Figure 28c shows a scour-like base of approximately 3 km width that spreads out in a fan-like geometry beyond 20 km, the limits of the 3D survey reaching Ariki-1. This geometry is highlighted by SOM purple/dark blue facies where the chosen input seismic attributes (entropy/homogeneity, peak magnitude, and peak frequency) highlight the characteristics of TP 3. Interestingly, although the nature of TP 3 is considered to be composite, it is still possible to isolate and map TP 3 because this pattern is very different from the background clastic sediments. The fact that the reflectors in TP 3 are far from being parallel (indicative of tuff clouds settling in volcanic facies) and that they form a fan-like geometry, indicates that the process that deposited the volcaniclastic material was a subaqueous flow. We use the term volcanic MTD to describe these volcanics.

In addition, Figure 28b shows evidence of sill junctions that appear to have erupted lava onto the west flank of the Kora volcano, which could explain the spatial distribution of these flows in the western flank of Kora (Figure 27).

**Pyroclastic flows from volcanoes adjacent to Kora**

Given the proximity of TP 4 to the Kora volcano, it is reasonable to attribute TP 4 to a younger eruption in the history of this volcano. Nevertheless, the distribution of the submarine volcanoes of the MBV mapped by Giba et al. (2010) (Figure 2b), depicts younger (8–4 Ma compared with Kora’s 16–12 Ma) volcanoes to the east of Kora. According to the mapped geometry of these andesitic volcanoes, they appear to coalesce instead of forming a single volcanic cone like Kora. Furthermore, their aerial extent appears to be at least five times larger than Kora. Detailed examination of the Kora 3D seismic survey indicates that TP 4 is found only on the eastern section onlapping onto the Kora volcanic edifice (Infante-Paez and Marfurt, 2017; Infante-Paez, 2018). Therefore, we interpret TP 4 to be related to the activity of the volcanoes located east of Kora, which is also confirmed by the SOM clustering results (Figure 29), in which the pur-
ple/dark-blue facies appear to dominate the entire area, even though isolated blocks of orange/yellow facies can be observed within the purple/dark-blue facies. The presence of these two facies is because the nature of TP 4 is “composite,” consisting of two or more patterns. In this scenario, the orange/yellow facies represents a greater content of clastic material being deposited within the volcanics in the basin. Due to the similarities of TP 2 and TP 4 and the fact that other andesitic volcanoes exist adjacent to Kora, we interpret TP 4 to be pyroclastic flows originated from the previously mentioned younger andesitic volcanoes from MVB.

**Potentially enhanced volcanic reservoirs**

As stated in the previous sections, the submarine volcanoes in the Taranaki Basin represent potential reservoirs as indicated by the DST in Kora-1A and the good log and core porosity from the Kora and other wells. The uncertainty in this type of reservoir is how well connected those pores are. For that reason, an area of potentially enhanced flow capacity is that of the interpreted pyroclastic flows (TP 4) adjacent to the east flank of Kora. In this area, the volcanics are probably fractured due to the faults associated with the Kora and/or the adjacent andesitic volcanoes (Figure 29b) perhaps similar (even though it is not the same structural style) to the case of the fractured andesites in the Jatibarang field in Indonesia where permeability is up to 10 D (S. Schutter, personal communication, 2018).

**Limitations**

Different authors (Meldahl et al., 2001; Roy et al., 2013; Qi et al., 2016; Sinha et al., 2016, 2017; Kumar and Mandal, 2017; Zhao et al., 2017; Qi, 2018) have used different methods, including MLFN, SOM, and CNN to predict well production performance and to cluster different patterns (gas chimneys, faults, and MTDs) in seismic data. However, these patterns have different natures (e.g., monogenetic, composite, and intricate) that present different levels of difficulty for machine learning. Thus, we propose that monogenetic and composite patterns can be mapped by feeding appropriate geometric, instantaneous, spectral, and seismic inversion-derived attributes to an unsupervised machine algorithm such as SOM. However, intricate patterns (Figure 8) may need a different method, such as CNN, where the algorithm convolves adjacent amplitude values to generate its own “attributes.” At present, there is not a single method (SOM, GTM, MLFN, SVM, and CNN) that is best to map seismic patterns. The clustering method chosen depends on the nature of the pattern that represents the facies of interest (Figure 12).

As seen in Figure 26, the voxel-based approach is most useful in monogenetic patterns in which we can easily differentiate the interpreted lava flow facies from the pyroclastic flows (TP 1 and TP 2, respectively). Similarly, Figures 28 and 29 show that the voxel-based approach is also useful in isolating composite seismic patterns, although they are represented by more than one cluster. The key to a

---

**Figure 26.** Vertical slice connecting the four Kora wells through the SOM clusters showing the two distinctive colors (purplish and yellowish) indicating two different facies. The green facies represent clastic sediments. Facies are colored based on the latent space projection (after Infante-Paez and Marfurt 2017). Seismic data are courtesy of NZP&M.

**Figure 27.** SOM clusters extracted along the top of the Kora structure map. Core descriptions provided in the well completion reports for wells Kora-1, -2, and -3 indicate the purple facies to be pyroclastic flows. In contrast, based on their geomorphology and structural relation to the volcanic cone, the yellow/orange facies are interpreted to be lava flows such as those reported by Klarner and Klarner (2012) and Holford et al. (2012). Clusters are colored based on the latent space axes (after Infante-Paez and Marfurt, 2017).
Figure 28. (a) SOM clusters extracted on a slice horizon close to the base of the Kora volcano indicated by the purple pick in (b). The colors in the horizon slice indicate similar facies (similar colors = similar facies). Note the purplish colors in a fan-like geometry with an approximately 2 km scour suggesting that TP 3 is associated with a landslide from the west flank of the Kora volcano. (b) A vertical slice through the seismic amplitude volume inside the subaqueous flow showing the extension of TP 3. (c) A vertical slice perpendicular to that in (b) through corendered amplitude and SOM clusters. Note the distinct purple facies associated with the landslide or volcanic MTD.

Figure 29. (a) The SOM clusters extracted along a stratal slice inside the volcanics onlapping the eastern flanks of the Kora volcano. (b) The same image corendered with the dip magnitude attribute highlighting the normal faults. The area is completely dominated by the purple facies, which, according to Albacore-1, are andesitic detritus probably derived from pyroclastic flows from adjacent younger volcanoes to the east of Kora.
successful clustering of a specific seismic facies is to determine the nature of that pattern. The limitation in this proposed workflow is twofold. First, a human interpreter needs to quantitatively define the nature of the TP. The illustrations shown here provide an example appropriate for volcanics in the Taranaki Basin. Second, once the pattern is recognized, the interpreter has to decide which attributes provide a quantitative measurement that serve as input to a machine learning algorithm. Our task as interpreters is to construct a dynamic library of the attributes expressions of different geologic facies in seismic data that can be updated as they are encountered with new facies and new attributes as they are developed.

**Conclusion**

From our experience in trying to isolate the extrusive volcanics related to the Miocene volcanism in Taranaki Basin, New Zealand, we realized that when trying to isolate a TP, interpreters usually describe it regarding their continuity, parallelism, amplitude, and frequency. Therefore, a good rule of thumb as to what attributes to isolate a TP, interpreters usually describe it regarding their continuity, parallelism, amplitude, and frequency. Thus, there is a need for a new set of seismic attributes that assist machine learning, so that they can identify patterns would be three types of attributes: attributes that highlight the amplitude (such as envelope, energy, rms amplitude, and relative acoustic impedance), attributes that highlight the continuity (such as GLCM entropy chaos), and attributes that highlight the frequency (such as peak spectral frequency, average spectral frequency, and bandwidth). Furthermore, there is a need of a seismic attribute that measures the self-similarity of the different patterns in the seismic section (with a change in the lateral and vertical location, and scale). Ideally, the interpreter draws a polygon around the TP to be mapped, and this new seismic attribute (self-similarity) would quantitatively evaluate whether the TP is monogenic or more complex (intricate).

The SOM and similar clustering algorithms do what they are supposed to do: cluster the attributes they are fed. Attributes that are good for 3D interactive interpretation may not be appropriate for machine learning. Thus, there is a need for a new set of seismic attributes that assist machine learning, so that they can identify more complex facies such as composite and intricate seismic patterns.

**Acknowledgments**

We would like to thank New Zealand Petroleum and Minerals and specifically H. Cameron for his friendly support in providing the Kora 3D seismic survey to the University of Oklahoma (OU) for use in research and education. Financial support for this effort is provided by the sponsors of the Attribute-Assisted Seismic Processing and Interpretation Consortium at OU. We thank our colleagues R. P.-D. Lima, G. Machado, S. Sinha, J. Qi, and M. Matos for their valuable input and suggestions. We would also like to thank T. Clee and L. Huang for reviewing our paper and for their insightful comments. All of the 3D seismic displays were made using licenses of Petrel provided to the University of Oklahoma for research and education courtesy of Schlumberger. We also would like to thank E. D. Silva for her time and dedication with the interpretation of the Kora horizon.

**DATA AND MATERIALS AVAILABILITY**

Data associated with this research are available and can be accessed via the following URL: https://data.nzpam.govt.nz/GOLD/system/mainframe.asp.

**References**


Hansen, R. J., and P. J. J. Kamp, 2004, Late Miocene to early Pliocene stratigraphic record in northern Taranaki Basin: Condensed sedimentation ahead of northern graben extension and progradation of the modern continental


Kalan, T., H. P. Sitorus, and M. Eman, 1994, Jatibarang field, geologic study of volcanic reservoir for horizontal well proposal: 23rd Annual Convention Proceedings, 1, 229–244.


Marfurt, K. J., 2018, Seismic attributes as the framework for data integration throughout the oilfield life cycle: SEG, Distinguished Instructor Series 21.


Lennon Infante-Paez received a bachelor’s degree in geophysical engineering from Simon Bolivar University, Venezuela. Given his passion for geology, Lennon decided to pursue a degree in geology. Therefore, he received a master’s degree in geology from the University of Oklahoma under Roger Slatt. He then received a doctorate degree (2018) in geophysics from the University of Oklahoma. He was mentored by SEG distinguished lecturer Kurt Marfurt. His interests include interpretation of seismic amplitude data, reservoir characterization, and seismic expression of igneous bodies and their implications for hydrocarbon exploration.

Kurt J. Marfurt began his geophysical career teaching geophysics and contributing to an industry-supported consortium on migration, inversion, and scattering (project MIDAS) at Columbia University’s Henry Krumb School of Mines in New York City. In 1981, he joined Amoco’s Tulsa Research Center and spent the next 18 years leading research efforts in modeling, migration, signal analysis, basin analysis, seismic attribute analysis, reflection tomography, seismic inversion, and multicomponent data analysis. In 1999, he joined the University of Houston as a professor in the department of geosciences and as a director of the Allied Geophysics Laboratories. He is currently a member of the Geophysical Societies of Tulsa and Houston, SEG, EAGE, AAPG, AGU, and SIAM, and he serves as the Deputy Editor-in-Chief for Interpretation. His current research activity includes prestack imaging, velocity analysis and inversion of converted waves, computer-assisted pattern recognition of geologic features on 3D seismic data, and interpreter-driven seismic processing. His research interests include seismic signal analysis, 3D seismic attributes, seismic velocity analysis, subsurface imaging, and multicomponent data analysis.