GEOPHYSICS, VOL. 83, NO. 1 (JANUARY-FEBRUARY 2018); P. O15–O24, 10 FIGS., 4 TABLES.
10.1190/GEO2017-0013.1

A3Mark: Seismic attribute benchmarking

Liyuan Xing[1], Victor Aarre[2], Nader Salman[2], Theoharis Theoharis[1], and Egil Tjåland[3]

ABSTRACT

Seismic attribute computation is one of the most active research and engineering topics in the computational geophysics world. Although numerous algorithms for seismic attribute computation have been developed, relatively little work has been done on scientifically characterizing their quality and accuracy. Seismic attribute characterization effort is largely qualitative and subjective. We have developed a robust and reliable scientific process for functionally comparing similar attributes. Structural seismic attributes are addressed, including discontinuity, dip angle, dip azimuth, and curvature. To establish a common software platform, and collection of data sets for efficient fully automatic evaluation, a stand-alone flexible online web service is designed. It is based on a C#-MATLAB implementation, and it is called A3Mark. The publicly available web service enables the automatic evaluation of individual categories of seismic attributes using customized quality metrics. It can be easily extended to include new algorithms. Several new synthetic data sets covering various structural measurements are also created, including noise, with ground truth, and they are made publicly available through the web service. Finally, a comparative evaluation of some current seismic attribute algorithms is given with quantitative and qualitative results.

INTRODUCTION AND BACKGROUND

Seismic attributes are quantities, such as measurements, characteristics, or properties, extracted or derived from seismic data. They are analyzed to enhance information that might be hidden or found to be subtle in a traditional seismic image, leading to a better geologic or geophysical interpretation of the data, ultimately increasing the success rate by more optimal well positioning.

Seismic attributes were introduced in the early 1970s. Since then, many new seismic attributes have been derived and computed, following major advances in computer technology such as digital recording and modern visualization techniques (Chopra and Marfurt, 2005). Nowadays, different (or even the same) software platforms provide different calculations for a single seismic attribute, e.g., volumetric dip, discontinuity, and curvature. In the Petrel E&P software platform, for example, the chaos, variance, amplitude contrast, and residual consistent dip attributes can all be used for discontinuity calculation. With the dramatically increasing number of seismic attributes in the oil industry, one can relate to the decision-making paradox, as first identified by Triantaphyllou and Mann (1989). This stems from the observation that there are plentiful seismic attributes (Chopra and Marfurt, 2008), each claiming to be the "best," while implementing intrinsically the same functionality (Barnes, 2006). Furthermore, these attributes may often yield very different results when fed with exactly the same input data. In other words, what is clearly lacking in the highly competitive world of seismic attributes is not the attributes themselves, but a robust and reliable scientific process to compare functionally similar attributes.

Some experts use informal methods such as writing blogs (Hall, 2016) or giving lectures (Behzad, 2016) to explain the details of the seismic attributes, and to provide corresponding suggestions, based on their professional experience. However, the confusion for geoscientists in selecting the appropriate seismic attribute still remains. Gunther and Marfurt (2016) compare five different methods for volumetric dip computation. They begin by describing key details of five dip computation methods, and then they apply them to a 3D seismic survey acquired offshore New Zealand and examine the result with a human interpreter for differences in resolution, artifacts, and sensitivity to noise. Although going in the right direction, this approach remains qualitative and subjective.

There are some works (Clawson et al., 2003; Hart and Chen, 2004; Verma et al., 2015) that take the first steps toward the ana-

Manuscript received by the Editor 9 January 2017; revised manuscript received 6 June 2017; published ahead of production 11 September 2017; published online 01 December 2017.
[1]Norwegian University of Science and Technology, Department of Computer Science, Trondheim, Norway. E-mail: liyuan.xing@ntnu.no; theotheo@idi.ntnu.no.
[2]Schlumberger AS, Schlumberger Information Solutions, Stavanger, Norway. E-mail: vaarre@slb.com; nsalman@slb.com.
[3]Norwegian University of Science and Technology, Department of Geoscience and Petroleum, Trondheim, Norway. E-mail: egil.tjaland@ntnu.no.

lytical validation of attribute algorithms using synthetic seismic data. Hart and Chen (2004) use simple 1D convolutional models constructed from well control to validate the subsequent interpretation of seismic attribute anomalies. In Verma et al. (2015), 2D synthetic common-shot gathers are computed, processed, and migrated to quantify the response of coherence, curvature, and acoustic impedance through four case studies. Coherence, P-impedance, and other attributes computed from 3D convolution models (Clawson et al., 2003) were used to determine which attributes may help in the seismic prediction for improved hydrocarbon reserve estimation. Although these synthetic seismic data sets are sophisticated, they still contain only a few individually isolated complex configurations. Because these configurations, such as petrophysical properties, are all twisted together, it is hard to automatically separate the effect of an individual configuration, and thus validation remains qualitative and subjective.

In contrast, an objective approach should automatically evaluate attributes and algorithms quantitatively, but this is rather challenging. Generally speaking, an objective approach uses various quality metrics to assess if a technique fits better to the ground truth or reference data than another. Furthermore, in the field of computer vision, web services are built to share data, host participants' results, and compare these results using objective metrics with the aim of setting a standard through benchmarking of the algorithms. These objective approaches are widely used, for example, in stereo reconstruction (Scharstein and Szeliski, 2001; Seitz et al., 2006; Stretcha et al., 2008) and image quality assessment (Ponomarenko et al. 2009), but not in the world of seismic attributes.

A3Mark attempts to bridge this gap by objectively comparing seismic attributes and thus guiding consumer choices. Toward this end, we build a publicly available web service that hosts several synthetic data sets that are generated for this purpose. These data sets cover a wide range of structural features under various noise conditions, such as discontinuity, dip angle, dip azimuth, and curvature, and they have the associated ground truths. These structural seismic attributes are important in the accurate calculation of many geometric attributes and to some structure-oriented filters, and they are key components in the emerging seismic geochronology analysis software. The web service includes mechanisms for uploading and automatically evaluating the user submitted seismic attribute cubes and for presenting the results of the metrics numerically and visually. In particular, the evaluation methodology is supported by customized quality metrics in the server backend. The web service is available at A3Mark.idi.ntnu.no. So far, there are comparative test results from several software platforms.

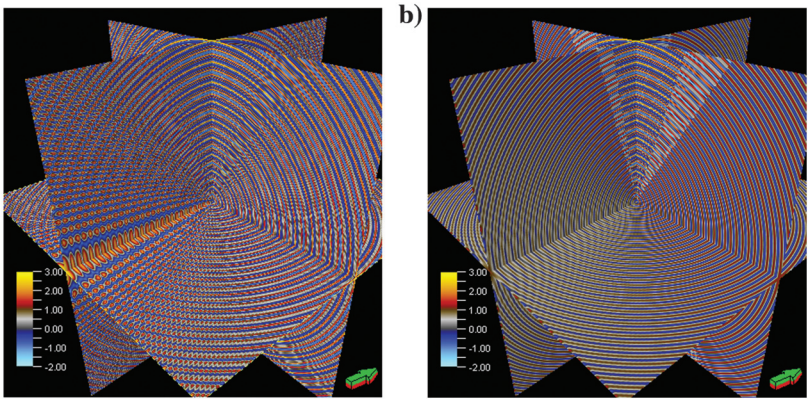

Figure 1. Base 3D seismic cube: (a) without antialias filter and (b) with antialias filter.

DATA SETS WITH GROUND TRUTH

Real seismic data are often complicated because of complex geology and noise, and they are therefore not ideal for evaluation purpose. On the one hand, the interpreted data such as faults cannot be regarded as ground truth, because the interpreters often have different, and sometimes even incoherent, interpretation of the same data set. The real geologic ground truth of the subsurface is difficult to obtain; one needs to compare the geologic or physical models with other available information, such as aerial photographs, test borings, well logs, geologic maps, and field observations. On the other hand, even the complicated real data may not cover all the values for one attribute, such as dip angle, dip azimuth, and curvature, in a single data set. Synthetic data can to some extent avoid the aforementioned issues and offer more control over the data, and this is the type of data adopted for the proposed benchmark.

Synthetic data sets

To generate our synthetic data sets, we proceed in two steps: first to create a base 3D seismic cube and second to create associated tailored cubes covering various structural measurements and noise.

For the first step, a base 3D seismic cube is created algorithmically (see Figure 1a) composed of three spectral components (15, 25, and 40 Hz) emulating a perfect sphere, and hence containing all possible dip angles [0°, 90°], dip azimuths [0°, 360°], and a wide range of the curvature values [−0.25, 0.25]. The bin size is 12.5 m × 25.0 m × 4.0 ms; the reason behind the difference in the XYT dimensions is that it makes it possible to test whether a dip algorithm honors the actual bin size (known as the XYZ space), or simply assumes that all dimensions are equally spaced (known as the IJK space). The latter is not desirable.

Moreover, we implement a parameterized linear antialias filter for high frequencies, which offers control over the amount of aliasing in the synthetic data. In particular, this antialias filter is applied to cover steep dips (see Figure 1b). The "ripple pattern" shown in Figure 1b is a rendering effect only, due to the pixel renderer (shader) in the viewer. The base seismic cube is useful for testing their algorithms' robustness against borderline aliasing situations at steep dips as well as for testing the vertical resolution of various tools.

For the second step, three tailored data sets covering various structural measurements are generated. Data set 1 includes one normal fault (fault 1) and one reverse fault (fault 2), running parallel to the crossline axis, as shown in Figure 2a. Another fault, fault 3 that intersects fault 1 and fault 2, and it is parallel to the inline axis is added in data set 2, as shown in Figure 2b. Intersecting faults add more challenges for geometric attributes. Data set 3 has one more fault, fault 4, which is not parallel to any of the axes, but intersects faults 1, 2, and 3, as illustrated in Figure 2c; this further increases the complexity.

Furthermore, each data set contains a unique type of noise. Three levels of noise are incorporated: noise free (Figure 3a), 5 dB random noise (Figure 3b), and 5 dB coherent noise (Figure 3c).

In particular, random noise is added by a random scalar drawn from the standard normal distribution. Coherent noise is obtained by spatially shifting a scaled copy of the original seismic trace, which mimics multiples.

Ground truths

Because the synthetic data are created from a physical model, we know the correct answer for many different structural measurements, including discontinuity, dip angle, dip azimuth, and curvature. All these analytical (i.e., not measured or approximated) properties are stored as separate cubes, and they serve as the "ground truth," when validating structural attribute algorithms. For each of our three data sets, four new cubes in the XYZ space (not in the IJK space) are computed analytically and stored as the ground truths. Specifically, an implicit time-to-depth conversion from the XYT space to the XYZ space is done by using a default conversion velocity of 2000 m/s, which is quite a realistic value. The ground truths are defined as follows:

- Discontinuity: The discontinuities are labeled by discrete values, in which one is discontinuity, and zero is nondiscontinuity; these are shown in red and gray, respectively, in discontinuity inline 40 and 116 of Figure 2.
- Dip angle: 3D tangent plane dip angle (continuous values in the range [0°, 90°]; see Figure 4a). With the default conversion velocity, a slope of x ms/m is translated to x m/m dip (remember, TWT is the two-way time, so the distance traveled is twice the actual depth), and the angle is hence equal to atan(x).
- Dip azimuth: 3D tangent plane azimuth angle relative to true geographical north (continuous values in the range [0°, 360°]; see Figure 4b).
- Curvature k1: There are several curvature measures applicable to seismic interpretation (Roberts, 2001) that are applied purely qualitatively in practice, such as the most-positive and most-negative curvatures. However, qualitative estimates have no "correct" answer, and they are thus difficult to evaluate. However, the industry is moving toward workflows in which quantitative results matter, e.g., if one intends to incorporate rock physics/geomechanics. Therefore, principal curvature defined here as 1/radius in m is chosen (continuous values in the range [−0.25, 0.25]; see Figure 4c).

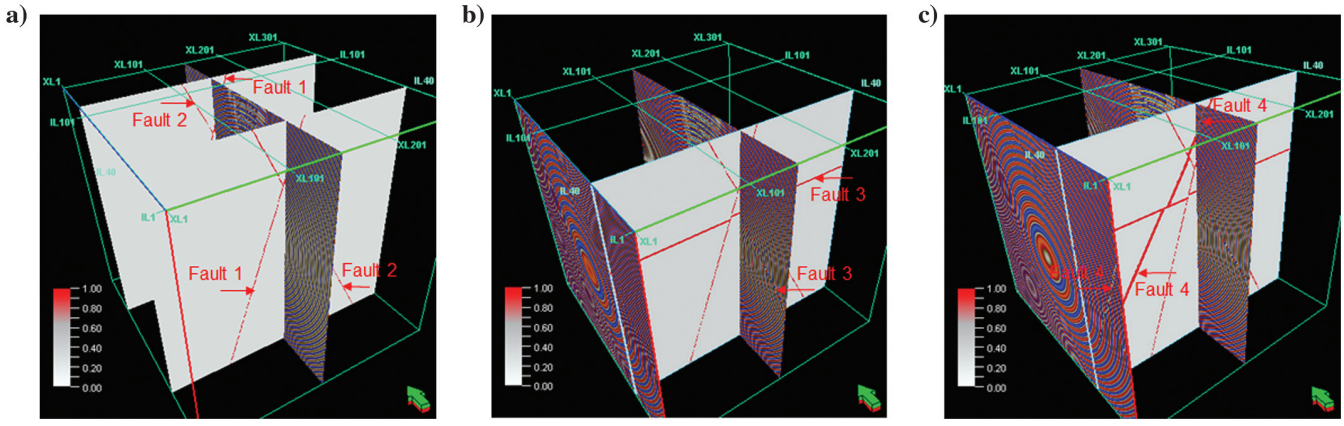

Figure 2. Seismic cubes for various structural measurements: (a) data set 1, discontinuity inlines 40 and 116, seismic crossline 151; (b) data set 2, discontinuity inline 40, seismic crossline 1 and 151; and (c) data set 3, discontinuity inline 40, seismic crossline 1 and 151.

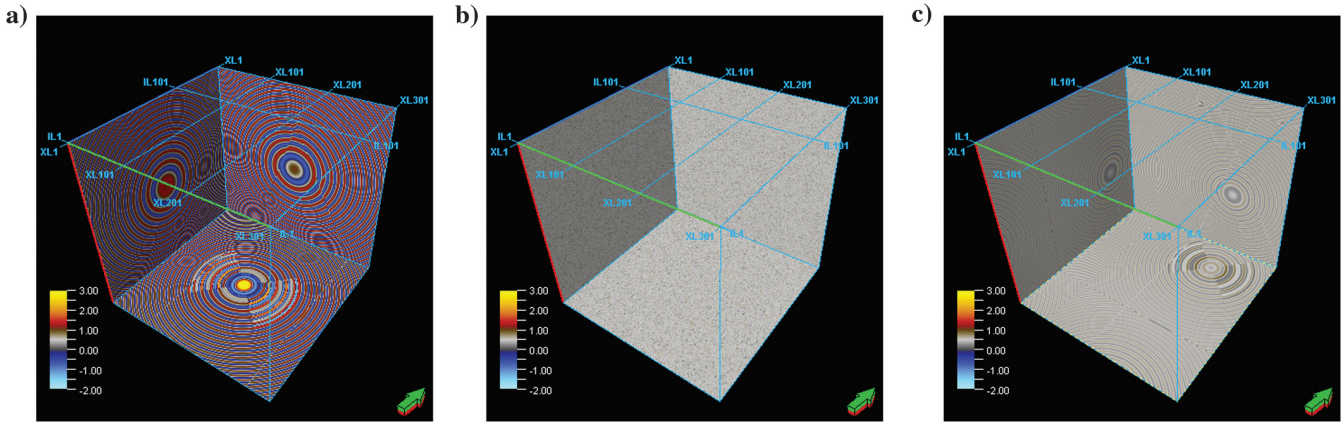

Figure 3. Seismic cubes for various noise types on data set 1: (a) noise free signal, (b) additive random noise at 5 dB, and (c) additive coherent noise at 5 dB.

EVALUATION METHODOLOGY

To evaluate the performance of a seismic attribute, we need a quantitative way to estimate its quality. We compute quality measures with respect to the four aforementioned ground truth seismic cubes, i.e., to their discrete or continuous values. We compute their variations, in 2D and 3D, as well as in different dip ranges: [0°, 90°] and [0°, 45°]. Precise explanations will be given below.

Basic quality metrics

A first metric arises out of regarding seismic attribute computation as a retrieval problem. In pattern recognition and information retrieval with binary classification, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, whereas recall (also known as sensitivity) is the fraction of relevant instances that are retrieved, as illustrated in Figure 5. Precision and recall are therefore based on a measure of relevance.

Seismic attribute computation can be assumed as a retrieval process with respect to the ground truth data. Precision and recall are then used to measure the retrieval effectiveness, based on the relevance between a calculated seismic attribute and its ground truth, as follows:

$$\text{precision} = \text{tp}/\text{Num}_{\text{sa}}; \text{recall} = \text{tp}/\text{Num}_{\text{gt}}, \tag{1}$$

$$\text{tp} = \sum_{i=1}^{i=\text{Num}_{\text{sa}}} \text{abs}(\text{val}_{\text{sa}}^{i} - \text{val}_{\text{gt}}^{i}) \le D, \tag{2}$$

where tp is the number of true positives of a specific seismic attribute. The terms Num_{sa} and Num_{gt} are the total numbers of specific seismic attributes and ground truths, respectively. In the discrete discontinuity case, Num is the number of voxels in which discontinuity equals to one, whereas in other continuous cases, Num is the whole volume. The function i is the index for those attributes counted in Num_{sa}, and $\text{val}_{\text{sa}}^{i}$ and $\text{val}_{\text{gt}}^{i}$ are the ith values in seismic attributes and ground truths, respectively. The function D is the threshold for relevance measurement, which is different for each ground truth category. For continuous categories, D is 20% of the ground truth range. For discrete categories, D is zero. If the submitted seismic attribute for discontinuity is continuous, i.e., not discrete as its ground truth, it is first discretized by an attribute value in the top 20%.

A second metric is root-mean-square error (rms error), a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. As defined in equation 3, the rms error of a seismic attribute measures the square root of the average of squared errors between the predicted values from specific seismic attribute and the observed values by ground truths:

$$\text{rms error} = \sqrt{\sum_{i=1}^{i=\text{Num}_{\text{sa}}} (\text{val}_{\text{sa}}^{i} - \text{val}_{\text{gt}}^{i})^2/\text{Num}_{\text{sa}}}, \tag{3}$$

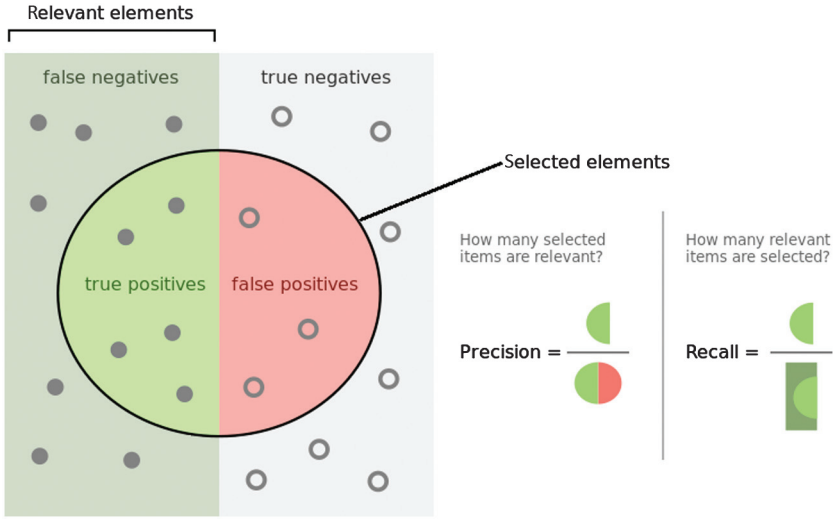

Figure 5. Illustration of precision and recall; image by Walber (2014).

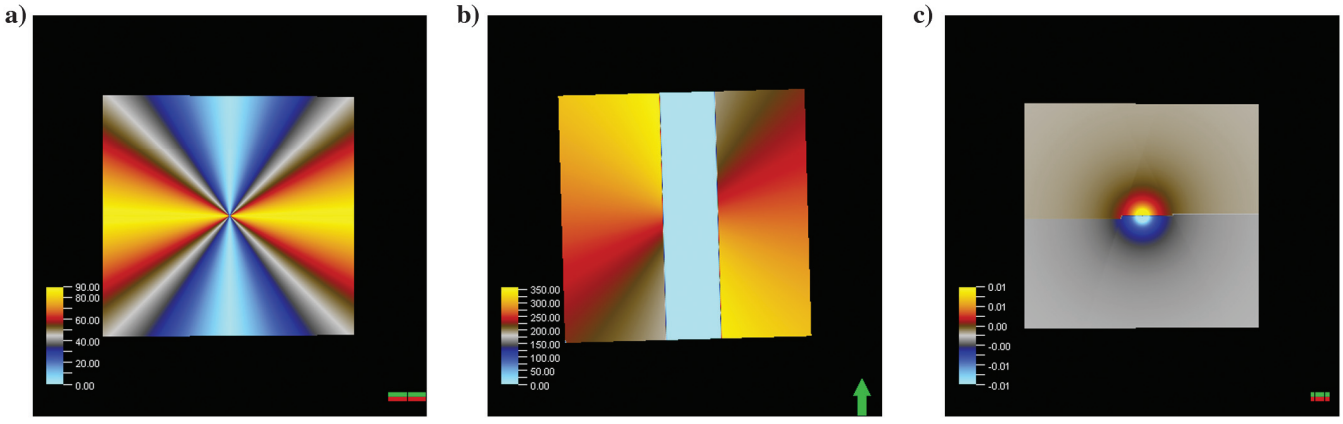

Figure 4. Ground truths on data set 1: (a) dip angle, crossline 151; (b) dip azimuth, time −1800; and (c) curvature k1, inline 76.

Moreover, rms error variations, rms error distance, and rms error discontinuity are especially designed for the discrete and continuous categories, respectively. The definitions are

$$\text{rms error distance} = \sqrt{\sum_{i=1}^{i=\text{Num}_{\text{sa}}} ((\text{val}_{\text{sa}}^{i} - \text{val}_{\text{gt}}^{i}) \times \text{Distance}_{\text{gt}}^{i})^2 / \text{Num}_{\text{sa}}}, \quad (4)$$

$$\text{Distance}_{\text{gt}}^{i\pm\text{range1}} = \text{sigmf}(0.5, 10), \quad \text{if } \text{val}_{\text{gt Discontinuity}}^{i} == 1, \quad (5)$$

$$\text{rms error discontinuity} = \sqrt{\sum_{i=1}^{i=\text{Num}'_{\text{sa}}} (\text{val}_{\text{sa}}^{i} - \text{val}_{\text{gt}}^{i})^2 / \text{Num}'_{\text{sa}}}, \quad (6)$$

$$\text{Num}'_{\text{sa}} = \text{Num}_{\text{sa}} - \sum_{i=1}^{i=\text{Num}_{\text{gt}}} \text{Discontinuity}_{\text{gt}}^{i} == 1, \quad (7)$$

$$\text{Discontinuity}_{\text{gt}}^{i\pm\text{range 2}} = 1, \quad \text{if } \text{val}_{\text{gt Discontinuity}}^{i} == 1, \quad (8)$$

where rms error distance is the rms error weighted by the nonlinear sigmoidal distance sigmf in equation 5. In particular, $i \pm \text{range 1}$ considers neighbors within range 1; range 1 = 30 is set empirically; see Figure 6a. A sigmf instance in 3D is shown in Figure 6c, which is applied on the discontinuity ground truth shown in Figure 6b. Thus, the discontinuities extracted by the seismic attributes are treated differently depending on the distance to the ground truth discontinuities.

rms error discontinuity is the rms error between the attribute values and the ground truth without including the discontinuities area defined in equation 8. Specifically, $i \pm \text{range 2}$ means neighbors within range 2, and range 2 = 5 is empirically set. This is because most structural attributes, except discontinuity attributes, are undefined at or very near discontinuities, because derivatives are undefined at discontinuities.

When calculating the difference between the seismic attributes $\text{val}_{\text{sa}}^{i}$ and the ground truths $\text{val}_{\text{gt}}^{i}$ in formulas 2, 3, and 6, a special case that needs to be considered is the dip azimuth, because it is cyclic angular data, meaning that 0° equals to 360°. There are different conventions for specifying a direction, e.g., [−180°, +180°] and [0°, 360°]. The difference between the angles is wrapped on the interval [0, 180°] without considering the direction. So if $\text{abs}(\text{val}_{\text{sa}}^{i} - \text{val}_{\text{gt}}^{i})$ is greater than 180, the difference is replaced by $360\text{-abs}(\text{val}_{\text{sa}}^{i} - \text{val}_{\text{gt}}^{i})$.

The precision and recall values are in the range [0, 1], the larger the better. An rms error-related quality metrics are in the range [0, inf), the smaller the better.

Variations of basic metrics

All already mentioned, basic metrics come in 3D and 2D varieties. In the 3D case, the quality measures are computed over the entire seismic cube, whereas in the 2D case, they are first computed on every inline, crossline, and time section of the cube, followed by a mean operation.

A caveat of the above metrics is that they do not discriminate between attributes at different dip angle ranges. However, in practice, it is computationally expensive to image dips beyond 45° because it requires compute-intensive migration algorithms to properly image those steeply dipping structures. Hence, this option is often skipped. Therefore, more quality metrics (called *_LowDip), which account only for the dip angle in the range [0°, 45°], are developed.

Thus, for the discrete categories (e.g., discontinuity), the metrics are 3D versions precision 3D, recall 3D, rms error distance 3D, and their 2D versions precision 2D, recall 2D, rms error distance 2D, as well as LowDip versions precesion 3D_LowDip, recall 3D_LowDip, rms error distance 3D_LowDip, precision 2D_LowDip, recall 2D_LowDip, and rms error distance 2D_LowDip.

For the continuous categories (e.g., dip angle, dip azimuth, curvature), the metrics include 3D versions recall 3D, rms error 3D, rms error discontinuity 3D, and their 2D versions rms error 2D, rms error discontinuity 2D, as well as LowDip versions recall 3D_LowDip, rms error 3D_LowDip, rms error discontinuity 3D_LowDip, rms error 2D_LowDip, and rms error discontinuity 2D_LowDip.

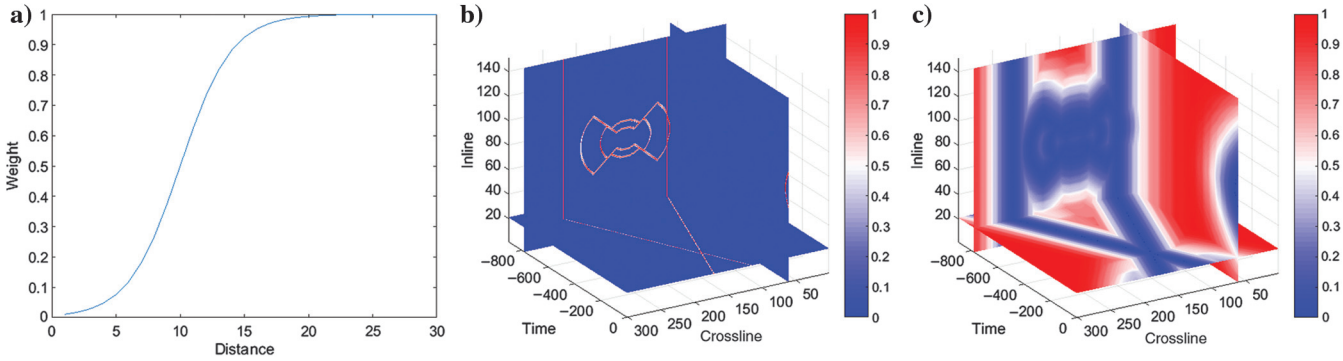

Figure 6. Illustration of nonlinear sigmoidal distance: (a) sigmf in one dimension, (b) discontinuity ground truth in 3D, and (c) sigmf instance in 3D applied on the discontinuity ground truth.

The web service: A3Mark

A3Mark offers the following main functionality:

- training and test *data sets* available for download
- online *submission* script for uploading seismic attribute cubes
- online *evaluation* tools for uploaded seismic attributes.

Data sets

Three tailored data sets are generated with different structural complexities, as described in the "Synthetic data sets" section. The simplest structural data set, with fault 1 and fault 2 and three noise levels, is the training data set. The other two data sets, with fault 3 and fault 4 and the associated noisy versions, are the test data sets.

Submission

Users can submit and evaluate their seismic attribute cubes on the training data set an indefinite number of times. They can see how their attributes compare against other published attributes, per ground truth category, and decide whether to publish their results or not. Results can be published multiple times on the same training data set; these are shown in a temporary table on the submission page, but only the latest results are present in the permanent table in the evaluation page.

Users' results on test data sets can be uploaded and published once they have gone through the training data set, and the relevant results are published. To prevent fitting to the test data sets, results can only be uploaded once, and cannot be viewed until after they are published.

Additional information per attribute such as reference and parameters can be provided so long as the user has published at least one result for that attribute.

Evaluation

Once a user chooses to publish his/her results, these appear on the evaluation page, as numerical metrics and images. The metrics represent the quantitative results that are calculated on the back end right after the seismic attribute cube is uploaded. The images are the qualitative snapshots of the uploaded seismic cube and the corresponding ground truth, from two different view angles. Detailed information on the attribute computation is also shown, e.g., reference, parameters, user, and experience.

RESULTS AND DISCUSSION

Several seismic attributes from several E&P software platforms are uploaded to our web service to gain a sense of how their relative performance vary against the four ground truth categories. The results from these attributes are reported in numerical metrics and images. In particular, the average numeric results for a specific category of seismic attribute on all data sets are shown in the following tables, and the images from a specific viewing angle on a noise-free training data set are illustrated in the following figure parts. These metrics are compared with the images (observed seismic attribute quality) and appear to be consistent.

Discontinuity

The following attributes are uploaded in A3Mark for discontinuity; their results appear in Table 1 and Figure 7:

- Petrel: amplitude contrast, chaos, variance, residual consistent dip
- OpendTect: fault similarity, instantaneous amplitude.

According to the metrics in Table 1, chaos outperforms (bold italic cells) in six metrics (4 3D/2D + 2 LowDip) and variance in four metrics (1 3D/2D + 3 LowDip). Amplitude contrast underperforms (bold cells) in six metrics (4 3D/2D + 2 LowDip) and fault similarity in four metrics (0 3D/2D + 4 LowDip). Others (residual consistent dip and instantaneous amplitude) have mid-level performance. In particular, the maximum precision of 3D/2D is 0.0999 (for the variance attribute), and the minimum precision of 3D/2D is 0.0085 (for the amplitude contrast attribute). Precision is consistently rather low, whereas recall is relatively high and with significant variation among the attributes. Specifically, the maximum recall of 3D/2D is 0.8771 for the chaos attribute, and the minimum recall of 3D/2D is 0.0809 for the amplitude contrast attribute. Overall, this means that there exists many false positives when most of the true positives are extracted. As for rms error distance, the maximum and minimum of 3D are 0.3976 and 0.1817, and those of 2D

Table 1. Discontinuity attributes and their metrics averaged over all data sets. Metric values are given for 3D and 2D versions in separate columns and their corresponding LowDip subversions also in separate columns.

Attributes	Metrics											
	precision		precision		recall		recall		rms error distance		rms error distance	
	3D	LowDip	2D	LowDip	3D	LowDip	2D	LowDip	3D	LowDip	2D	LowDip
Amplitude contrast	**0.0085**	0.2839	**0.0196**	0.3929	**0.0809**	**0.0082**	**0.1020**	**0.0113**	0.2841	0.0111	0.3703	***0.0101***
Chaos	***0.0870***	0.1339	0.0870	0.1544	***0.8771***	***0.8862***	***0.8735***	***0.9056***	0.1880	0.0638	***0.2815***	0.2073
Variance	0.0554	***0.4523***	***0.0999***	***0.4840***	0.5541	0.4621	0.5758	0.4987	0.2494	***0.0077***	0.3443	0.0173
Residual consistent dip	0.0313	0.1408	0.0487	0.1808	0.6524	0.5367	0.6886	0.5734	**0.3976**	0.0995	**0.5500**	0.2251
Fault similarity	0.0311	**0.0470**	0.0279	**0.0763**	0.3225	0.4934	0.2761	0.4649	***0.1817***	**0.1128**	0.3036	**0.3226**
Instantaneous amplitude	0.0369	0.0796	0.0397	0.1207	0.3739	0.4745	0.3435	0.4504	0.2132	0.0943	0.3403	0.2380

are 0.5500 and 0.2815. In other words, those false positives have average distance to ground truths approximately from 7 to 9 voxels in 3D and from 8 to 10 pixels in 2D.

A similar conclusion can be drawn from Figure 7. Chaos (Figure 7c) and variance (Figure 7d) appear to be closest to the ground truth (Figure 7a), although chaos has some false positives in the red bowknot (discontinuity in this case) in the low dip angle region (<45°) and variance in the high dip angle region (>45°), refer to the dip angle region in Figure 8a. The amplitude contrast (Figure 7b) extracts no true positives, but there are many false positives in the high dip angle region. Fault similarity (Figure 7f) extracts some true positives but many false positives at the low dip angle region. The true positives of residual consistent dip (Figure 7e) and instantaneous amplitude (Figure 7g) include almost all the relevant elements (discontinuity) in the ground truth. However, there are many false positives, in particular, the false positives of residual consistent dip are mainly at the high dip angle region, whereas those of instantaneous amplitude are mainly at the middle dip angle region (≈ 45°).

If we exclude the regions of high dip angle, performance improves significantly, especially for those attributes that have false positives mainly at a high dip angle, such as amplitude contrast, variance, and residual consistent dip, as verified by the LowDip metrics in Table 1. In particular, there is much improvement in precision and rms error distance. For examples, the precision of variance reaches 0.4840, amplitude contrast arrives at 0.3929, and residual consistent dip becomes 0.1808, all from less than 0.1 in 3D/2D. Most of the rms error distance values are less than 0.1 (≈ 5 voxels/pixels) in LowDip compared with a residual consistent dip of 0.3976 in 3D and 0.5500 in 2D, amplitude contrast of 0.2841 in 3D and 0.3703 in 2D, as well as variance of 0.2494 in 3D and 0.3443 in 2D. Meanwhile, the performance changes in recall are approximately 0.1, among which amplitude contrast, variance, and residual consistent dip decrease, whereas others increase. These observations indicate the reliability of the proposed metrics.

Dip angle

Two attributes are uploaded in A3Mark for dip angle and their results appear in Table 2 and Figure 8:

- Petrel: consistent dip, local structural dip. Both are converted from IJK to XYZ because ground truth is in the XYZ space.

The 3D/2D metrics in Table 2 show that consistent dip is better than the local structural dip in recall, but not rms error or rms error

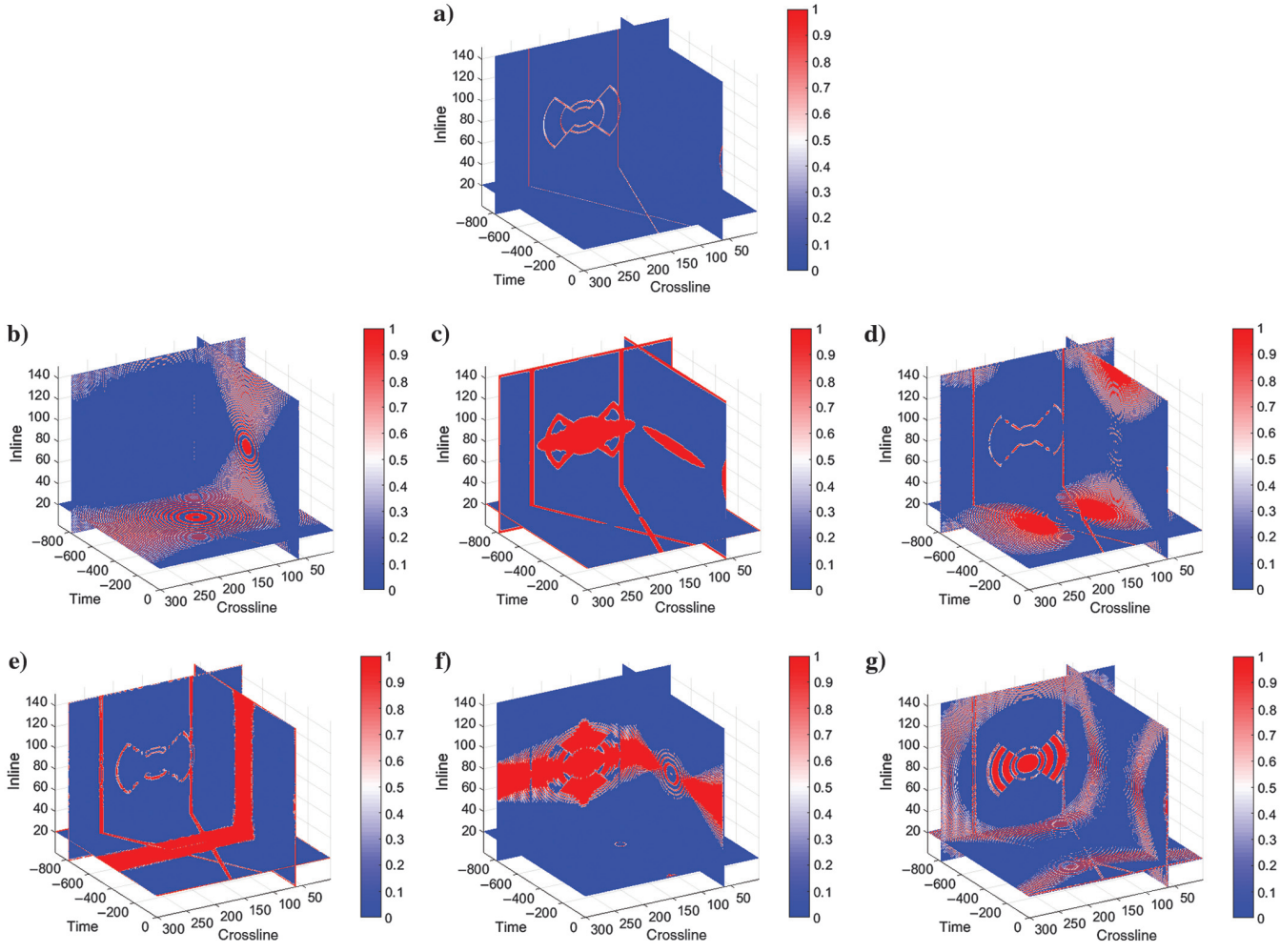

Figure 7. Discontinuity attributes on noise-free training data set and related ground truth. Ground truth: (a) discontinuity. Attributes: (b) amplitude contrast, (c) chaos, (d) variance, (e) residual consistent dip, (f) fault similarity, and (g) instantaneous amplitude.

discontinuity. Specifically, the recall of consistent dip is greater than the local structural dip by approximately 0.32. The rms error and rms error discontinuity of consistent dip are above 14, whereas that of local structural dip is approximately eight. This is because consistent dip has many outliers at approximately 90°, as shown in Figure 8b, when compared with the ground truth in Figure 8a and the local structural dip in Figure 8c. Therefore, without including the outliers in the high dip angle region, all the LowDip metrics of consistent dip outperform that of local structural dip. In particular, the recall of the consistent dip remains greater than the local consistent dip by approximately 0.3. Although the rms error and rms error discontinuity of consistent dip becomes approximately four, which is better than the local structural dip value of greater than 5.

Dip azimuth

Table 3 and Figure 9 give the results for the dip azimuth from the following two attributes:

- Petrel: consistent dip azimuth, local structural azimuth. Both are converted from IJK to XYZ because ground truth is in the XYZ space.

Table 3 shows that 3D/2D metrics of consistent dip azimuth are much better than local structural azimuth in recall, but slightly better in rms error and rms error discontinuity. Specifically, consistent dip azimuth is greater than local structural azimuth by approximately 0.44 in recall, whereas it is smaller by approximately three in

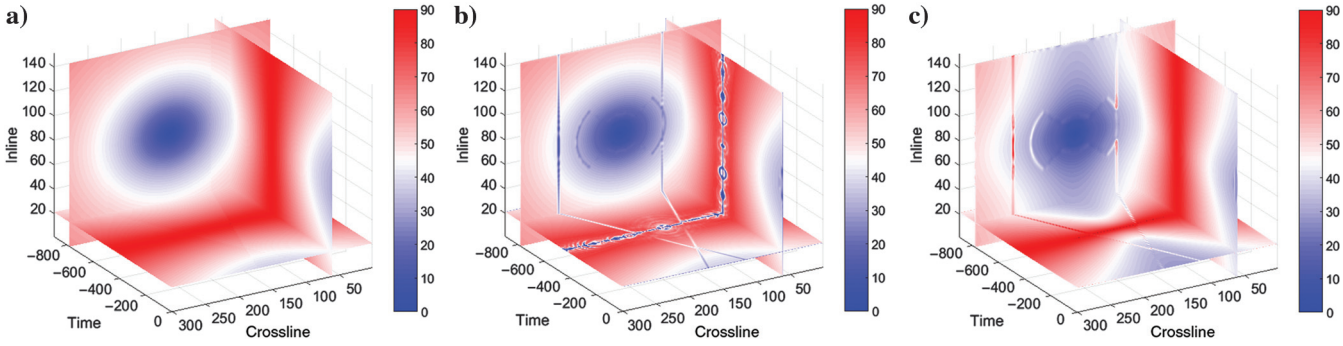

Figure 8. Dip angle attributes on noise free training data set and related ground truth. Ground truth: (a) dip angle. Attributes: (b) consistent dip and (c) local structural dip.

Table 2. Dip angle attributes and their metrics averaged over all data sets. Metric values are given for 3D and 2D versions in separate columns and their corresponding LowDip subversions also in separate columns. In the case of recall for continuous categories, the values are the same for 3D and 2D versions.

Attributes	Metrics									
	recall		rms error		rms error		rms error discontinuity		rms error discontinuity	
	3D	LowDip	3D	LowDip	2D	LowDip	3D	LowDip	2D	LowDip
Consistent dip	***0.4984***	***0.6382***	19.2649	***4.1535***	15.0059	***4.1068***	19.2291	***3.7730***	14.4338	***3.1330***
Local structural dip	0.1787	0.3488	***8.1654***	5.3779	***7.8223***	6.5384	***8.1009***	5.3191	***7.7580***	5.4314

Table 3. Dip azimuth attributes and their metrics averaged over all data sets. Metric values are given for 3D and 2D versions in separate columns and their corresponding LowDip subversions also in separate columns. In the case of recall for continuous categories, the values are the same for 3D and 2D versions.

Attributes	Metrics									
	recall		rms error		rms error		rms error discontinuity		rms error discontinuity	
	3D	LowDip	3D	LowDip	2D	LowDip	3D	LowDip	2D	LowDip
Consistent dip azimuth	***0.7164***	***0.8700***	***15.7835***	***4.1670***	***11.2497***	***4.7243***	15.5727	***1.8328***	***10.5203***	***1.6199***
Local structural azimuth	0.2772	0.3747	16.8037	9.1200	14.6162	10.4066	***15.2005***	6.8599	13.2706	7.0120

rms error and rms error discontinuity of 2D and almost the same in rms error and rms error discontinuity of 3D. As shown in Figure 9b and 9c, they are quite similar to each other and close to ground truth in Figure 9a, except some clear dip azimuth outliers at approximately 90° dip angle in the former. If the high dip angle region is excluded, all the LowDip metrics of consistent dip azimuth are much better than that of local structural azimuth, as listed in Table 3. Correspondingly, consistent dip azimuth remains larger than local structural azimuth by approximately 0.5 in recall, whereas smaller by approximately five in rms error and rms error discontinuity of 3D/2D.

Curvature K1

The following attributes are uploaded in A3Mark for curvature k1; their results are shown in Table 4 and Figure 10:

- Petrel: consistent curvature, 3D curvature (impossible to convert from IJK to XYZ)
- OpendTect: curvature gradient.

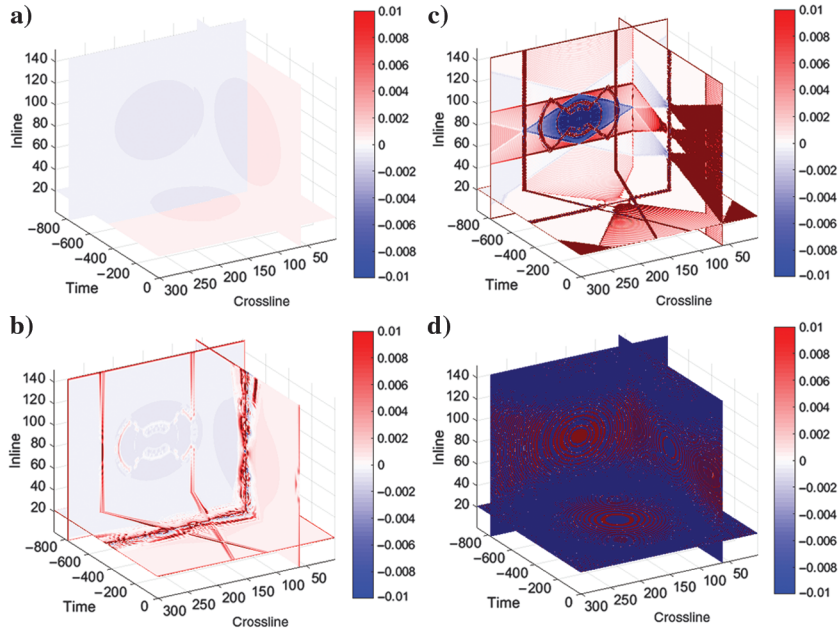

Figure 10. Curvature k1 attributes on noise free training data set and related ground truth. Ground truth: (a) curvature k1. Attributes: (b) consistent curvature, (c) 3D curvature, and (d) curvature gradient.

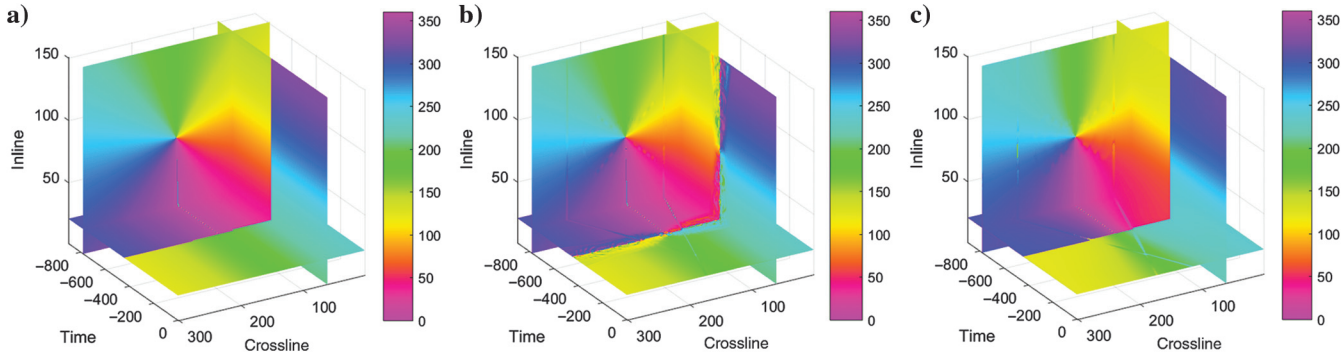

Figure 9. Dip azimuth attributes on noise-free training data set and related ground truth. Ground truth: (a) dip azimuth. Attributes: (b) consistent dip azimuth and (c) local structural azimuth.

Table 4. Curvature k1 attributes and their metrics averaged over all data sets. Metric values are given for 3D and 2D versions in separate columns and their corresponding LowDip subversions also in separate columns. In the case of recall for continuous categories, the values are the same for 3D and 2D versions.

Attributes	Metrics									
	recall		rms error		rms error		rms error discontinuity		rms error discontinuity	
	3D	LowDip	3D	LowDip	2D	LowDip	3D	LowDip	2D	LowDip
Consistent curvature	***0.9547***	***0.9884***	***0.0031***	***0.0014***	***0.0027***	***0.0021***	***0.0030***	***0.0010***	***0.0025***	***0.0009***
3D curvature	0.4052	0.3471	0.4534	0.2083	0.3898	0.2235	0.4174	0.0587	0.3350	0.0724
Curvature gradient	**0.0015**	**0.0017**	**0.7476**	**0.6882**	**0.7469**	**0.6759**	**0.7515**	**0.6780**	**0.7482**	**0.6729**

When compared with the ground truth in Figure 10a, it is obvious that the performance order of curvature seismic attributes, from highest to lowest, is consistent curvature (Figure 10b), 3D curvature (Figure 10c), and curvature gradient (Figure 10d). Therefore, all the 3D/2D metrics and LowDip metrics in Table 4 have the same order as previously mentioned, in which the highest are marked in the bold italic table cells and the lowest are marked in bold table cells. In particular, the recall of consistent dip is close to one and rms error and rms error discontinuity are almost equal to zero. On the contrary, the recall of the curvature gradient is almost equal to zero and rms error and rms error discontinuity are approximately 0.7. Moreover, a slightly better performance in LowDip can be concluded from Table 4 when compared with 3D/2D. As expected, some curvature outliers approximately 90° dip angle can be observed from Figure 10.

CONCLUSION

Although many algorithms for seismic attribute computation have been developed, relatively little work has been done on scientifically characterizing their quality and accuracy. The A3Mark web service addresses this gap by establishing the first publicly available seismic attributes benchmark, and it is available at A3Mark.idi.ntnu.no. A3Mark comes with a set of synthetic data sets of different complexities including discontinuity, dip angle, dip azimuth, and curvature ground truth as well as a set of quality metrics to automatically evaluate seismic attributes. The experiments reported here demonstrate the limitations of some existing attributes and show their sensitivity to key parameters. The results presented are a starting point for comparison of other attributes. The users and developers of commercial or open-source seismic attributes are invited to run their algorithms on our web service and report their results for the benefit of the seismic attribute world. In the near future, we plan to continue improving A3Mark, for example, on data set and results presentation. In particular, noise such as steeply dipping noise that cuts through reflections at all times on marine surveys, will be considered for inclusion in our synthetic data sets. The quality metric results with color coding are not intuitive enough to tell at a glance what is "good" or "bad." Perhaps one single metric that reflects the overall performance needs be designed or all quality metrics need be normalized in the range [0,1].

Moreover, we hope to add more data sets to our web service, including synthetic salt domes and real seismic data sets. It is also intended to include more categories of attributes, such as instantaneous frequency that is regarded as one of the important physical attributes, in addition to structural attributes, and their respective ground truths and quality metrics.

ACKNOWLEDGMENTS

Many thanks to J. Xu for helping out with the web service and D. Harishidayat for uploading the seismic attributes' results from OpendTect. This research is supported in part by NTNU and Schlumberger AS.

REFERENCES

Barnes, A., 2006, Too many seismic attributes?: CSEG Recorder, **31**, 41–45.
Behzad, A., 2016, Seismic attributes and their applications in seismic interpretation, https://www.eage.org/nl/education/short-course-catalogue/geology/alaei-seismic-attributes-and-their-applications, accessed 10 October 2017.
Chopra, S., and K. J. Marfurt, 2005, Seismic attributes — A historical perspective: Geophysics, **70**, no. 5,3SO–28SO, doi: 10.1190/1.2098670.
Chopra, S., and K. J. Marfurt, 2008, Emerging and future trends in seismic attributes: Geophysics, **27**, 298–318, doi: 10.1190/1.2896620.
Clawson, S., H. Meng, M. Sonnenfeld, M. Uland, S. Atan, M. Batzle, M. Gardner, and M. S. Uman, 2003, The value of 3D seismic attributes for illuminating deep water deposits by seismic forward modeling of the Brushy Canyon formation: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2431–2434, doi: 10.1190/1.1817880.
Gunther, M. J., and K. J. Marfurt, 2016, A comparison of alternative volumetric dip computations: 86th Annual International Meeting, SEG, Expanded Abstracts, 2123–2128, doi: 10.1190/segam2016-13964446.1.
Hall, M., 2016, White magic: Calibrating seismic attributes, http://www.agilegeoscience.com/blog/2016/1/25/white-magic-calibrating-seismic-attributes, accessed 10 October 2017.
Hart, B., and M. Chen, 2004, Understanding seismic attributes through forward modeling: The Leading Edge, **23**, 834–841, doi: 10.1190/1.1803492.
Ponomarenko, N., V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, 2009, TID2008 — A database for evaluation of full-reference visual quality assessment metrics: Advances of Modern Radioelectronics, **10**, 30–45.
Roberts, A., 2001, Curvature attributes and their application to 3D interpreted horizons: First Break, **19**, 85–100, doi: 10.1046/j.0263-5046.2001.00142.x.
Scharstein, D., and R. Szeliski, 2001, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms: International Journal of Computer Vision, **47**, 7–42, doi: 10.1023/A:1014573219977.
Seitz, S., B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, 2006, A comparison and evaluation of multi-view stereo reconstruction algorithms: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, 519–526, doi: 10.1109/CVPR.2006.19.
Strecha, C., W. V. Hansen, L. V. Gool, P. Fua, and U. Thoennessen, 2008, On benchmarking camera calibration and multi-view stereo for high resolution imagery: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, doi: 10.1109/CVPR.2008.4587706.
Triantaphyllou, E., and S. H. Mann, 1989, An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox: Journal Decision Support Systems, **5**, 303–312, doi: 10.1016/0167-9236(89)90037-7.
Verma, S., O. Mutlu, T. Ha, W. Bailey, and K. J. Marfurt, 2015, Calibration of attribute anomalies through prestack seismic modeling: Interpretation, **3**, no. 4, SAC55–SAC70, doi: 10.1190/INT-2015-0072.1.
Walber, , 2014, Precision recall. Own work. Licensed under CC BY-SA 4.0 via Wikimedia Commons, https://commons.wikimedia.org/wiki/File%3APrecisionrecall.svg, accessed 10 October 2017.