

# Automatic QC of denoise processing using a machine learning classification

Maiza Bekara<sup>1\*</sup> and Anthony Day<sup>1</sup> describe a new approach to automate the quality control of denoise processes in seismic data processing using machine learning.

## Introduction

Noise attenuation is an important part of a typical seismic data processing sequence. The general purpose of noise attenuation is to improve the resolution of seismic images. It can also be used to pre-condition the data prior to the application of certain processes to avoid the generation of artefacts. For example, applying a de-bubbling filter to marine seismic data that contains moderate swell noise will create visible low-frequency artefacts in the output. Therefore, adequately removing the swell noise with techniques similar to the one proposed by Bekara and van der Baan (2010) is a prerequisite before applying such a process.

The first step in any denoising process is to optimize the key parameters to ensure that noise has been adequately attenuated without causing attenuation of signal. This is usually done by testing the filter using a range of parameters on a small subset of the data known as the test lines. At this stage, the optimality of the denoise process is visually assessed by the processor and there are generally no issues, as the volume of the test lines is small. Once the parameterization is approved, it will then be used to process data from the entire survey. However, there is no guarantee that the chosen parameterization will be optimal for the whole survey because the characteristics of the noise and the signal can change dramatically throughout the data. For example, in a typical narrow azimuth towed streamer marine survey (~ 9000 km<sup>2</sup>), which can last for up to four months, the swell noise characteristics (peak frequency, amplitude and spatial contamination) can vary significantly, as they depend on the weather conditions. This will make any fixed parameterization for the swell noise attenuation filter potentially sub-optimal for some parts of the survey. Therefore, Quality Control (QC) is mandatory to give people confidence in the denoise processing and to allow the processor to move to the next step in the processing sequence. In brief, the QC process is an attempt to provide a reliable answer to the following two questions:

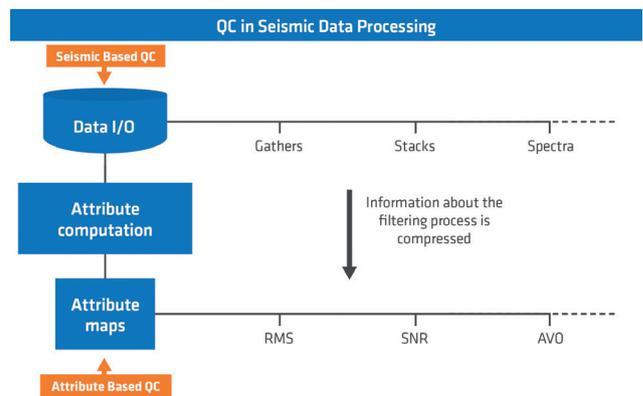
1. Is there any residual noise or signal attenuation in the data after the application of the filter?
2. If yes, where is it in the survey?

A reliable QC can take a considerable amount of time and resources in a typical seismic processing project. It is therefore advantageous to automate this process to improve project turnaround. This paper expands on previous work by Spanos and Bekara (2013), where an unsupervised outlier detection approach was used to answer the

above two questions. To improve the reliability of the automatic QC, we propose using a supervised learning approach to build an automatic classifier to predict the type of filtering as one of three classes (mild, optimal or harsh). The framework was tested on full-scale production to QC the results of noise attenuation prior to wavefield separation for towed streamer dual-sensor marine data.

## Evolution of QC in seismic processing

Early QC practices were based heavily on visual inspection of the data (e.g., gathers, 2D/3D stacks). This approach is the most reliable way to assess the quality of the seismic data. However, it is time consuming, impractical for large volumes of data, and requires the user to have a good geophysical understanding of both the data and the filtering process being applied. As the volume of data in a typical survey has increased over time, QC practice has moved towards assessing global attribute maps that are computed from the data such as RMS amplitude or signal-to-noise ratio maps (Figure 1). These attributes capture some aspects of the filtering process but in a highly compressed way. They provide the processor with a fast, global QC tool. However, it is not fully reliable or sufficient by itself as frequent cross-checks with the seismic data are required. When assessing these attribute maps, the focus is on detecting outliers and anomalies, hence the type of expertise needed is more statistical than geophysical. Attribute-based QC has managed to shift the QC from the data space to the attribute space making it streamlined and easy to use. The massive increase in the volume of processed seismic data led



**Figure 1** Types of QC in seismic data processing. The increase in volume of seismic data moved the QC point from the data space to the attribute space.

<sup>1</sup> PGS

\* Corresponding author, E-mail: maiza.bekara@pgs.com

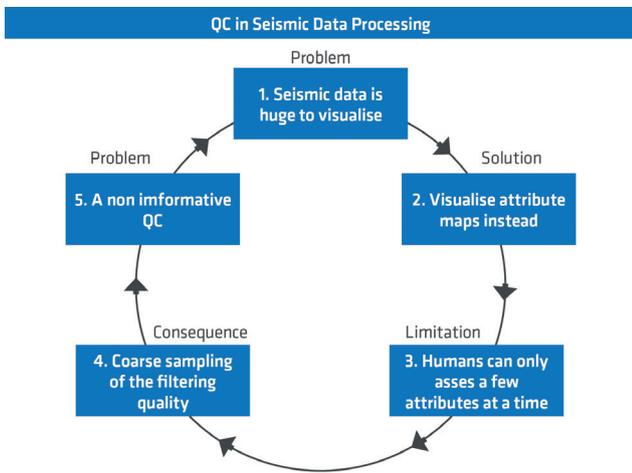


Figure 2 Limitations of attribute based QC.

to the evolution from seismic-based QC to attribute-based QC. However, humans cannot understand the visualization of more than two or three attributes at a time (Figure 2), let alone capture the intercorrelation between them. Therefore, the assessment is restricted to a few attributes to fit human visual understanding capability, leading to a coarse sampling of the filtering performance and hence to a non-informative QC. In order to overcome this limitation, we propose the following:

1. To bypass the visualization of attributes in order to break the chain.
2. To compute as many informative attributes as possible to give a better sampling of the filtering performance.
3. To use statistical data mining techniques to analyse the different attributes.

By overcoming these limitations we can naturally evolve the attribute-based QC to be an intelligent and automatic system.

### Proposed QC system

The main idea of the proposed QC system is to formulate the problem of automatic QC as a supervised classification problem

(Figure 3). The methodology consists of the following building blocks:

#### 1. Labelled training data

We assume that in addition to the optimal filtering results obtained on the test lines, we also have the results of a mild and a harsh filtering on the same lines, from the testing phase. These extra results are used to construct the training data for the cases of residual noise and signal attenuation. Figure 4 shows sample shot gathers from a test line with moderate contamination by swell noise. The test line is filtered with three types of filter (optimal, mild and harsh). The difference between optimal filtering (Figure 6) and mild filtering (Figure 5) is very subtle for these shots and also throughout the test lines. This corresponds to a typical mild classification as judged by an experienced processor. For the harsh filtering, the output looks cleaner and signal attenuation is visible in the difference section (Figure 7). It is very important that the mild and harsh filtering should correspond to typical mild and harsh filtering and not excessively mild and harsh in order to construct a reliable training data set.

#### 2. Attributes computation

We extend the type of computed attributes to include statistical ones that measure the level of similarity between the output and the difference of the filtering at an ensemble level (e.g., common shot gathers). These statistical attributes are computed from the seismic samples for a targeted time gate in each ensemble, as shown in Figure 4. In the case of optimal filtering, these attributes will not show any similarity between the output and the difference. This assumption comes from the fact that the signal (i.e., output) has nothing in common with the noise (residual) in the case of ideal filtering. When there is signal attenuation or residual noise, the level of similarity will increase as some noise is also present in the output or some signal present in the difference, and this will be picked up by the attributes. The attributes are multidimensional and form a cloud in the attribute space when computed for all the ensembles in the training lines.

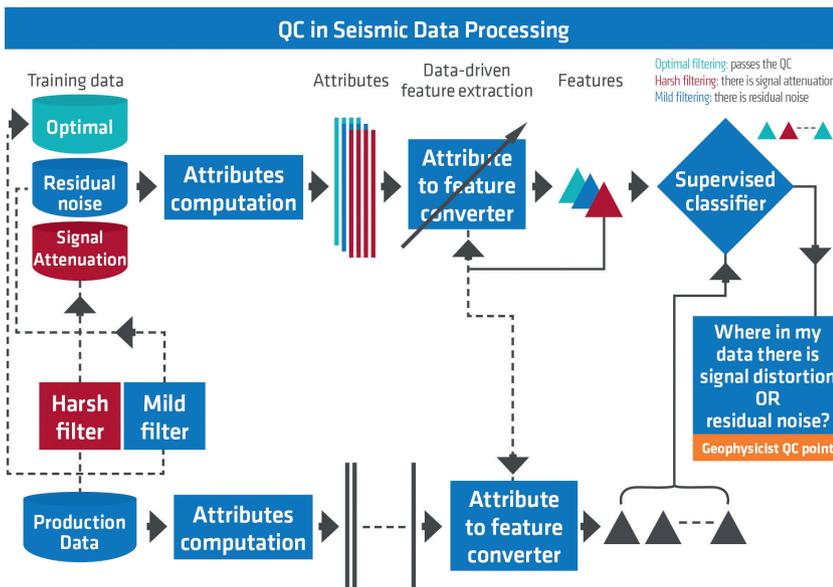
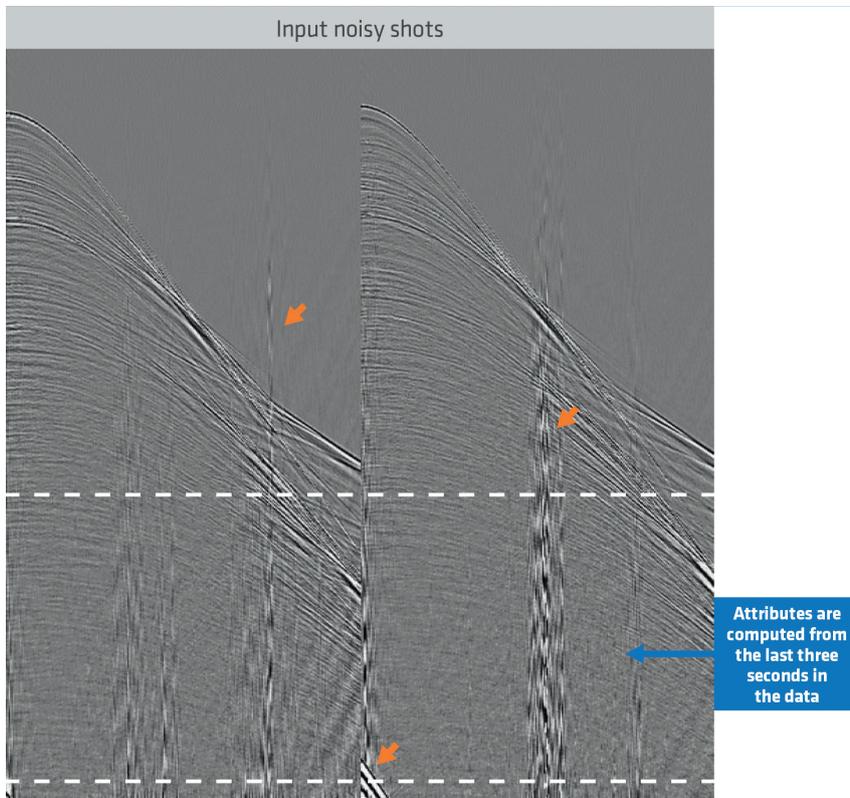


Figure 3 Proposed automatic QC system based on a machine learning classifier using training data.



**Figure 4** Shot gathers (single cable) from a towed-streamer marine seismic survey with visible contamination with swell noise (orange arrows).

Let us denote by  $(y_i, x_i), i = 1, 2, \dots, N$ , the  $N$  seismic samples from the input and output respectively in the targeted time gate and set the difference. Examples of statistical-based attributes include the following:

1. Pearson cross-correlation:

$$\rho_{xd} = \frac{\sum_{i=1}^N (x_i - \bar{x})(d_i - \bar{d})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (d_i - \bar{d})^2}} \quad (1)$$

where  $\bar{x}$  and  $\bar{d}$  are respectively the average sample value for the output and difference.

2. Mean Lambda (Spanos and Bekara, 2013) defined as

$$\bar{\lambda} = \sum_{i=1}^N \log\left(\frac{y_i^2}{x_i^2 + d_i^2}\right) \quad (2)$$

3. Kendall cross-correlation (Chen and Chu, 2005) defined as

- i. Randomly select two samples for X and D.,  $(x_p, d_p)$  and  $(x_q, d_q)$
- ii. Set  $a_1 = \text{sign}(x_p - x_q)$  and  $a_2 = \text{sign}(d_p - d_q)$
- iii. If  $(a_1 = a_2)$ , then the pairs are concordant else they are discordant
- iv. Kendall rank correlation is defined as:

$$K_{\text{cross}} = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{Total number of pairs}} \quad (3)$$

The cross-plots of five different attributes computed from three test lines are shown in Figure 8. These are only shown to validate

the attributes, which are overlaid for the optimal, harsh and mild filtering cases using a three-colour code (mild = blue, optimal = green and harsh = red). Some attributes (e.g., TLAMBDA an average of the trace based mean  $\lambda$ , Figure 8) show a good level of visual separation between the different types of filtering, particularly the harsh one. The clusters of attributes for the mild and the optimal filtering are close as they reflect the observation made earlier concerning the subtle differences between the two types of filtering. It is worthwhile highlighting that during the design of this framework we found that the attributes are the most important building block. The more informative the attributes, the more robust the outcome of the automatic classification.

### 3. Feature extraction

There will always be hidden correlations between the individual attributes due to their common origin. Their dimension can also be extremely large, making the subsequent classification problem harder. The task of de-correlating the attributes to extract useful structure in them is called feature extraction. It is a mapping process that transforms each vector of attributes into an optionally lower dimensional vector of features. Often the features tend to have a better cluster-discrimination power compared to the attributes. Key linear feature extraction procedures are principal component analysis (PCA) and independent component analysis (ICA) (Hyvärinen et al., 2001). To take the spatial consistency of the filtering outcome into consideration, attributes from adjacent shots are merged with the attributes of the central shot resulting in an augmentation of the total number of attributes for the central shot. Figure 9 shows the cluster of features obtained after applying a non-linear mapping (spatial augmentation with 20 shots followed by PCA) on the cluster of attributes in Figure 8.

One can see that the visual separation between the clusters of the different filtering has improved for the three dominant principal components (PCA01, PCA02 and PCA03) where most of the inter-dependency between the attributes resides. The higher order principal components are mixed up as they do not convey any information with regards to cluster discrimination. Therefore, they can be ignored, which will reduce the dimension of the input feature vector that will be fed to the classifier for the training phase, leading to a reduction in the training time.

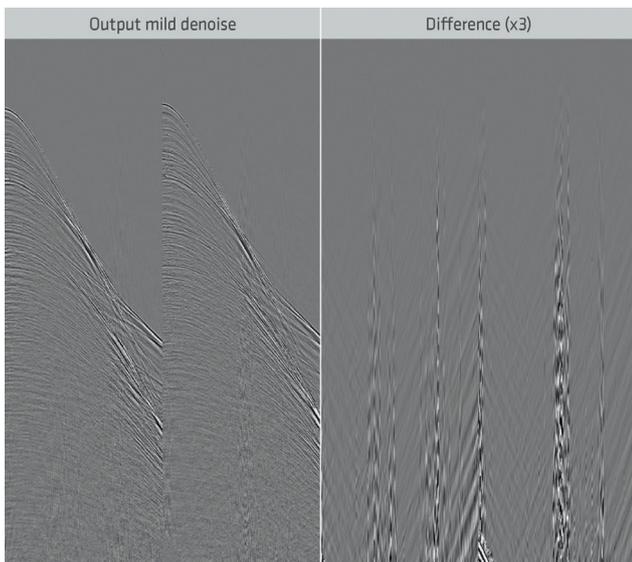
#### 4. Training the classifier

A supervised classification algorithm based on support vector machines (SVM) (Cristianini and Shawe-Taylor, 2000) is constructed using the training data (features plus filtering type). Building a machine classifier results in partitioning the feature space into regions called decision subspaces that correspond to

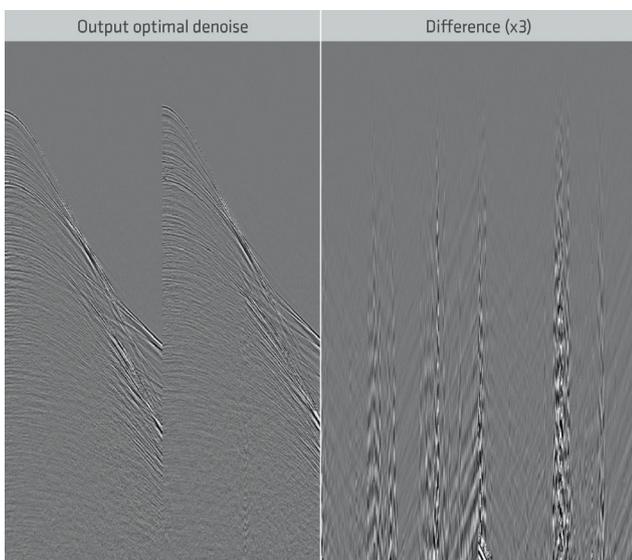
optimal, mild and harsh filtering. To use the machine classifier to predict the class-type of any new single vector of attributes is simply a matter of first mapping the attributes into the feature space and then identifying in which decision subspace they fall. Other classifiers such as neural networks (Goodfellow et al., 2016) have been tried, but it was found that the resultant classifier is not robust unless the architecture of the neural network (number of layers and the number of nodes per layer) is optimized using validation data to avoid overfitting. Even when calibrated on some validation lines, it was found that neural network architecture does not generalize well from one project to another, and retraining with a different architecture is always needed for a new project. Thus, we have lost a key advantage of neural networks, which is the ability to transfer the learning across projects. Moreover, the optimized learning process with neural networks requires more computer resource time compared to SVM learning. Finally, from a user's perspective, a better knowledge of data science is needed to run and optimize a neural network classifier compared to a SVM classifier.

#### Illustrative example

To show an example of how automatic classification with SVM works in practice, we consider the case of using only two attributes ( $\bar{\lambda}$  and Kendall cross-correlation) computed from a subset of the training lines (three lines) as indicated by arrows in Figure 10b. No feature mapping was applied in this example. Figure 11b shows the partition of the feature (attribute) space into different decision subspaces using SVM with polynomial kernels. The boundaries of the decision subspaces carve the clusters of the attributes for the different filters (Figure 11a). They are optimally constructed to minimize the misclassification rate for the training data, also known as the training error (Figure 11c). The training error for each filtering type is the percentage of shots from the training lines that have been misclassified. The training error is low for the harsh denoise (~ 0.6%) as its corresponding attributes stand out



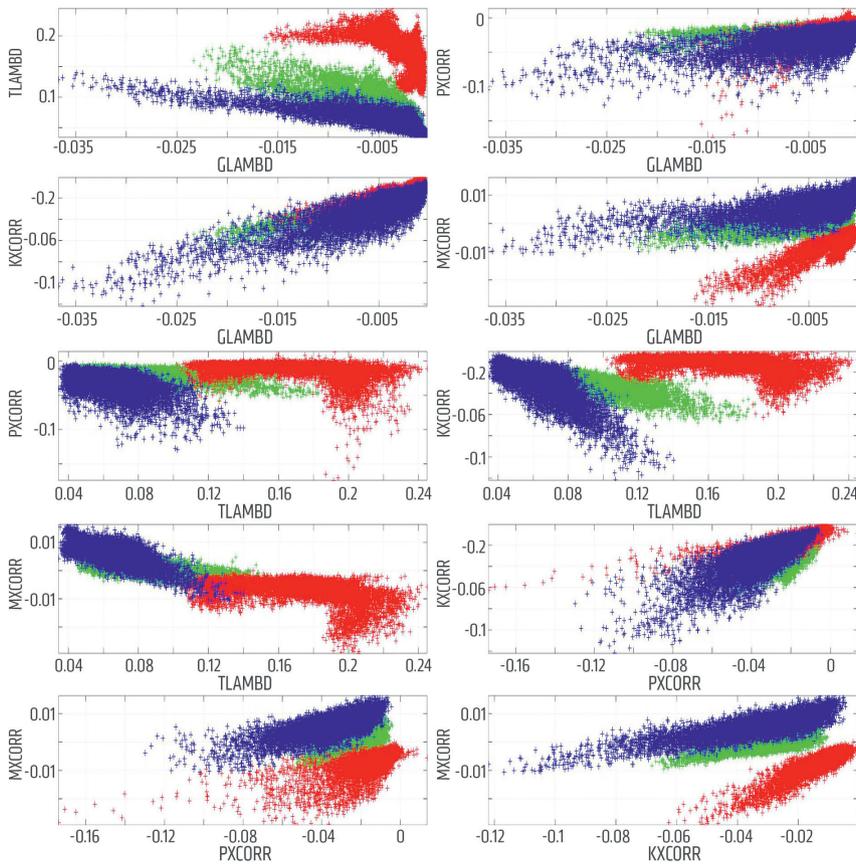
**Figure 5** The output of mild denoise (left) and the corresponding difference (right) scaled by a factor of 3. Some residual noise is still observed in the output.



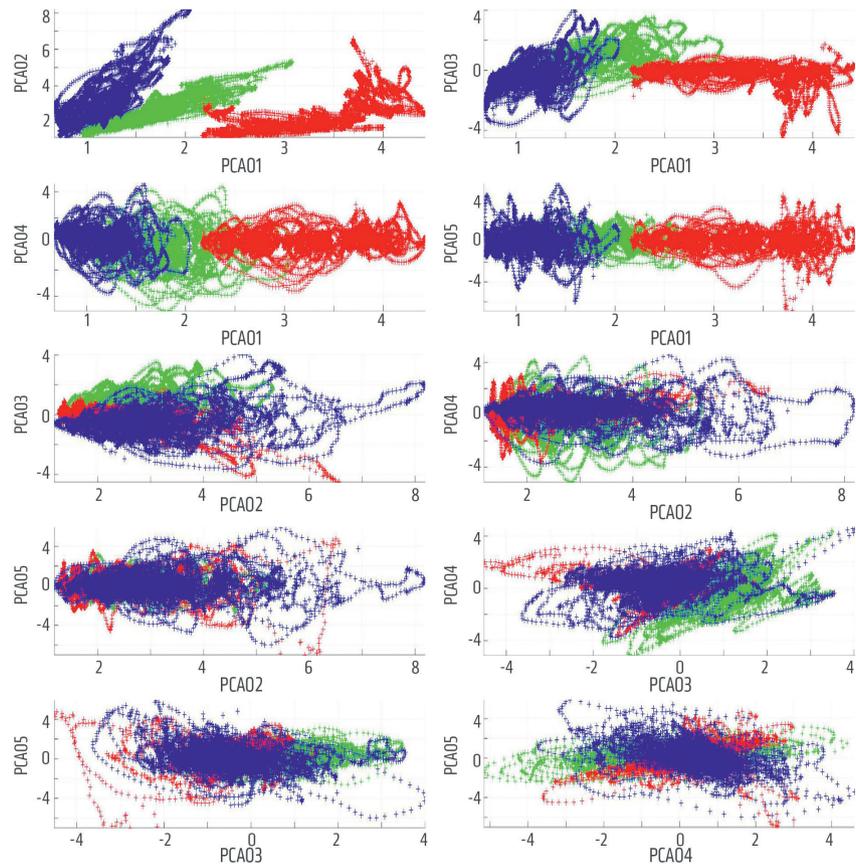
**Figure 6** The output of optimal denoise (left) and the corresponding difference (right) scaled by a factor of 3. Very little difference when compared to the mild denoise in Figure 5.



**Figure 7** The output of harsh denoise (left) and the corresponding difference (right) scaled by a factor of 3. The output is clear, but signal leakage is visible as indicated by arrows.



**Figure 8** Cross-plots of five attributes computed from training lines for harsh denoise (red), mild denoise (blue) and optimal denoise (green).



**Figure 9** Cross-plots of five principal components computed after spatial augmentation of the attributes in Figure 8. Note that visual separation between the different clusters have improved for the primary principal components.

relative to the attributes computed from the mild and the optimal filtering. The training error for mild denoise is however relatively

large (~ 19%) and the training error for optimal denoise is small (~ 11%). This is expected as the attributes for mild denoise

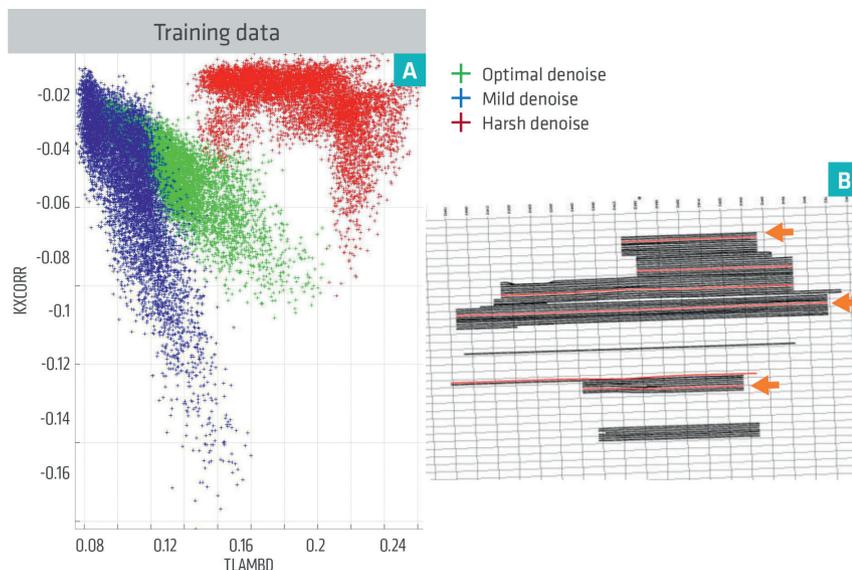
overlap with the attributes of the optimal denoise. We can say that 11% of shots with no issues (i.e., optimally filtered) have been misclassified as shots with residual noise and this is called, in the nomenclature of classification, the false positive or false alarm rate. Likewise, 19% of shots with residual noise have been classified as shots with no problem and this is known as the false negative rate. This finding indicates that the two attributes considered in Figure 10a are not informative enough to reliably detect shots with residual noise. These misclassification errors are subjective and biased as the classification was performed on the same data that is being used to train the classifier.

We now try to classify a new line which has not been used in the training process, to assess its robustness. The line is indicated by the black arrow in Figure 12a. The classification is visually easy to understand and is based on plotting the attributes and looking into which decision subspace they fall. We can compute the misclassification rate for each filtering type (Figure 12a). The

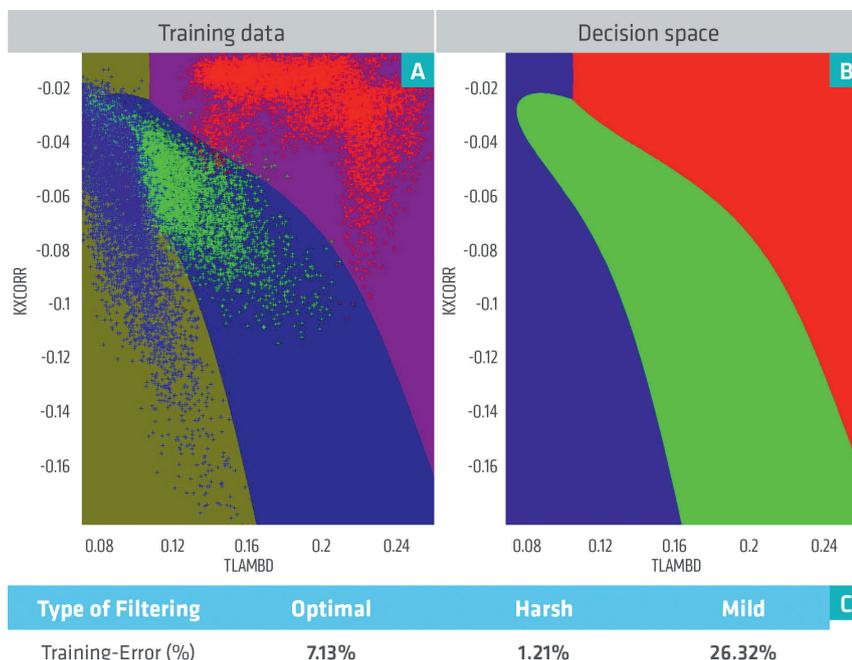
harsh denoise was nicely classified with an error rate of less than 31%. The majority of the attribute falls in the harsh denoise decision subspace. A slightly larger classification error is observed for the optimal denoise, where about 3% of the shots that fall near the boundaries of the decision subspaces are classified as shots with signal residual noise. The mild denoise, however, has the largest misclassification error (~19%) and this is not surprising as this filter is very close to the optimal filtering. In the next section, we show that by computing multiple attributes and using feature mapping we can significantly reduce the misclassification error and improve the robustness of the classification.

### Application

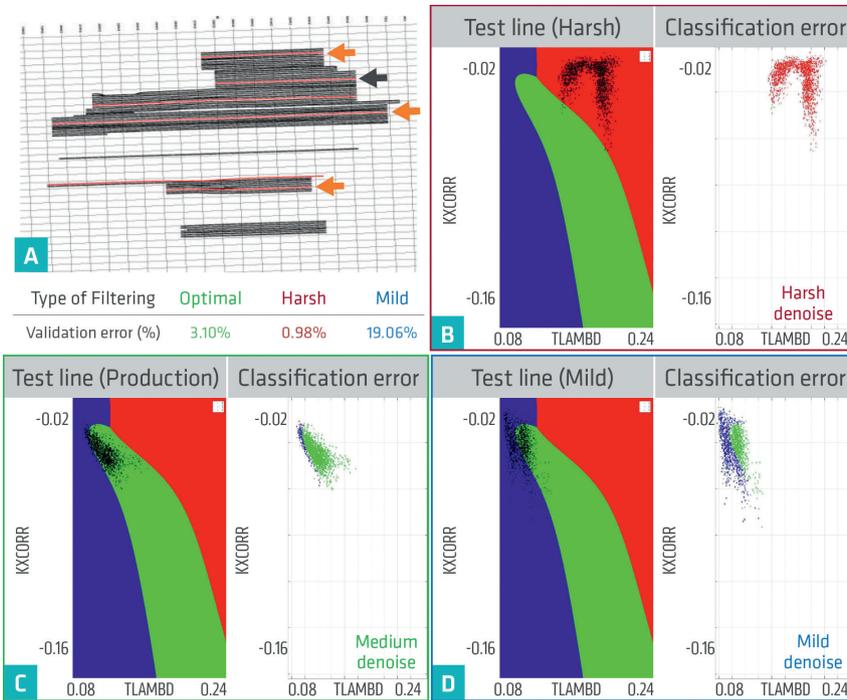
The proposed framework was tested on a full production data set to QC the denoise process prior to wavefield separation using dual-sensor marine data. For this test, we had access to six lines, each with three types of filtering (Figure 13a). In



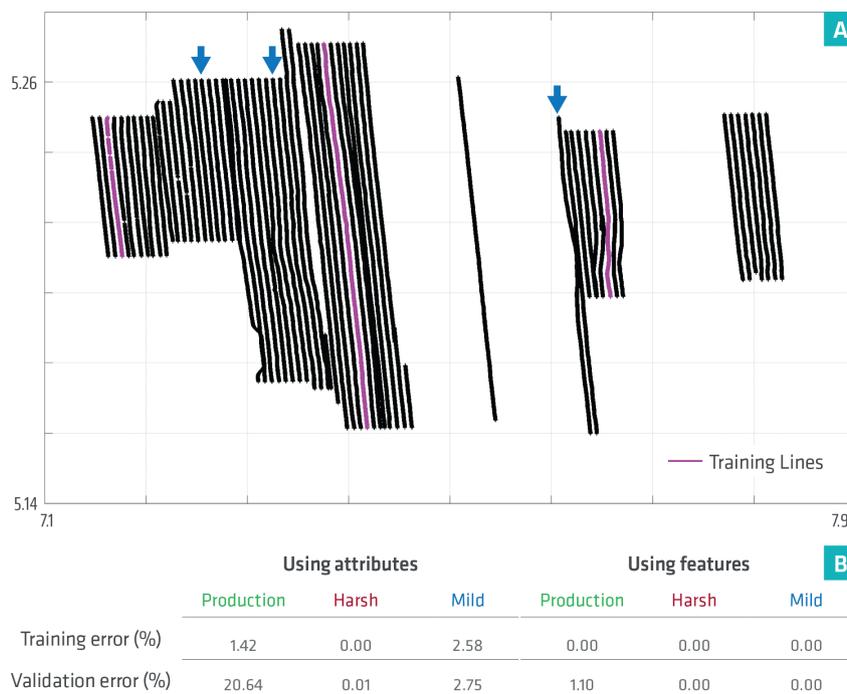
**Figure 10** a) Cross-plots of two attributes for the three types of filtering computed from three training lines (indicated by arrows), (b) location of training lines in the survey.



**Figure 11** a) Cross-plot of attributes (optimal=green, harsh=red, mild=blue); b) Decision subspace corresponding to each filtering type; c) training error.



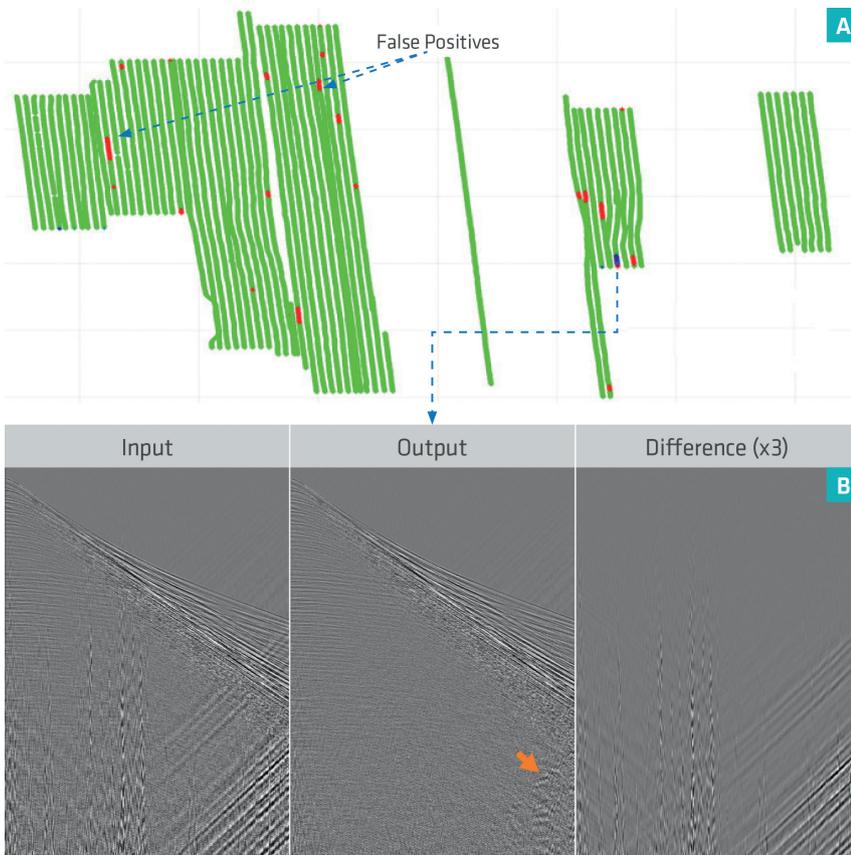
**Figure 12** a) Location of the training (black arrow) and validation lines (black arrow). Cross-plots of attributes for b) harsh noise, c) optimal noise and d) mild noise overlaid with the decision subspaces for each filtering class.



**Figure 13** a) Locations of training lines (pink) and validation lines (arrows); b) training and validation errors using attributes (left) and features (right).

order to assess the quality of the prediction, the six lines were split into three lines for training the machine learning classifier and the other three lines were used to validate the prediction (validation lines indicated by arrows in Figure 13a). The training lines were spread throughout the survey to improve the robustness of the learning. The validation error for a given filtering type is the percentage of shots, from the validation lines, that were not correctly classified. Similarly, the training error is the same as the misclassification rate but computed for the shots that belong to the three training lines. The training error is a measure of how separable the training features are, and the validation error is a measure of how robust the QC

system is when predicting the filtering type for a new shot that was not used in the training. Figure 13b shows the training and validation error rates using the raw attributes (i.e. no feature mapping). One can see that the training error for all the types of filtering is small, indicating that the attributes are informative. There is a massive reduction in the training error compared to the example in Figure 11c where only two attributes were used (Figure 11). This is evidence that multi-dimensional attributes can help to improve the construction of the classification (i.e., more data is used in the fitting). The validation errors are also small, except for the optimal filtering, where they are moderately large. Approximately 20% of the optimally filtered



**Figure 14** a) Decision QC map that indicates the locations of shots with potential signal leakage (red) and potential residual noise (blue). b) Cross-check with the seismic data.

shots are misclassified as harsh or mild filtering, with a bias towards mild. This misclassification error diminishes when the features, obtained through spatial augmentation and PCA, are used instead of the raw attributes to train the machine learning classifier (Figure 13b). We found that only the combination of both the spatial augmentation and PCA leads to this reduction in the validation error; using either one alone was not sufficient.

The classifier was then used to predict the outcome of the filtering for 68 lines that constitute a large portion of the survey. The result of the automatic QC is shown as a three-colour decision map for every shot in the survey (Figure 14a). This map captures all the information that the user needs to make the decision about the filtering process. It indicates where the filtering is suboptimal in terms of signal attenuation (red flag) or residual noise (blue flag). Inspection of the seismic data at the highlighted blue locations shows some residual turn noise left in the data (Figure 14b). An extra denoise pass was applied on this line to address the problem. The majority of localized small red anomalies were false positives. They were related to the attenuation of coherent linear noise that was not present in the training lines, but removed by the denoise process.

### Conclusions

Machine learning can be used to automate and assist the user in the quality control of denoise processing. The level of interaction from the user is minimal and is restricted to the training phase. The main challenge is to improve the robustness of the automatic QC by reducing the rate of false positives. This can be achieved by using more informative attributes that can incorporate some physical properties about the noise being attenuated.

Alternatively, training the classifier directly with raw seismic data is potentially appealing and can improve the robustness of the classifier but it is computationally expensive and would not be economically feasible if the trained classifier is not transferable across projects. In this paper, the QC of denoise process was considered, but the same concept can be generalized to QC in other processes such as multiple attenuation.

### Acknowledgements

We would like to thank the crew of *Ramform Sterling* who acquired the data used in this test, and PGS for the permission to publish these results.

### References

Bekara, M. and Van der Baan, M. [2010]. High-amplitude noise detection by the expectation-maximization algorithm with application to swell-noise attenuation. *Geophysics*, **75**, V39-V49.

Chen, C. and Chu, T.H. [2005]. Similarity Measurement Between Images. *IEEE Conference on Computer Software and Applications*, 41-42.

Cristianini, N. and Shawe-Taylor, J. [2000]. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.

Goodfellow, I, Bengio, Y. and Courville, A. [2016]. *Deep Learning*. MIT press.

Hyvärinen, A., Karhunen, J. and Oja, E. [2001]. *Independent component analysis*. John Wiley & Sons.

Spanos, A. and Bekara, M. [2013]. Using statistical techniques to improve the QC process of swell noise filtering. *75th EAGE Conference & Exhibition*, Extended Abstracts.