Automation of marine seismic data processing

Andrew Long[1] and Tony Martin[2]

https://doi.org/10.1190/tle39040264.1

Abstract

Marine seismic data sets contain highly redundant information. Data analytics and machine learning-based solutions should provide opportunities to reduce turnaround and improve confidence levels in output data volumes. A proof-of-concept (POC) thrust regime example from Indonesia illustrates that parameter testing can almost be eliminated if existing project parameter data can be mined from a database. Where quality control (QC) is required for complex challenges such as noise removal, supervised classifiers are a platform that can enable rapid global quantitative decisions based on relevant data attributes, moving behind the subjective art of observational QC. Finally, many early processing steps depend on reasonable knowledge of the velocity model in addition to the explicit dependence of imaging steps. A POC Monte Carlo-based model building exercise in West Africa used an efficient tomographic platform to demonstrate that turnaround can be reduced from 90 days to only a few days, even when the starting model was significantly wrong. These examples illustrate that a lot is already within our reach, and the development of embedded feedback loops will improve the level of automation further, particularly if humans can learn to let the data speak for itself.

Introduction

The proposed application of automated processing to towed-streamer marine seismic projects broadly follows three considerations: (1) parameterization with minimal testing, (2) accelerated quality control (QC), and (3) derivation of the velocity model. This sequence acknowledges that appropriately conditioned data are required to build any model. How much further can we progress to full automation? Sheridan and Verplank (1978) provide a relevant 10-stage hierarchy of automation levels in which level seven (the computer does the entire job and tells the human what it did) represents the highest level of automation, where manual decisions still outrank the computer. Complete delegation of decision making to algorithms will conceivably be as much of a psychological barrier as it will be a technological innovation. In our proof-of-concept (POC) examples, we advocate the use of pragmatic solutions that can exploit the redundancy of information recorded by modern marine seismic surveys. The machine learning-type QC described by Bekara and Day (2019) is placed in the context of rapidly validating the parameterization of processing modules with data analytics solutions. Strategic data compression, onboard and onshore teams working in concert with common big data platforms, and the use of deep learning, data analytics, and Monte Carlo methods for automated velocity model building are all demonstrated to be relevant when streamlining project complexity and reducing project turnaround.

Bridging the vessel-office distance

Seismic vessel operations are complex enterprises that depend on the seamless integration of many systems and platforms to control a vast array of data collection. Modern vessels routinely tow 16–18 multisensor streamers with 8–10 km length, representing a receiver array with up to 17 km^2 of sensors, and record 2–10 TB of seismic, navigation, and ancillary data each day. The size of the recorded data volumes are a direct function of the number of channels recorded and the sample rate. While real-time condition-based monitoring of vessel performance data is already streamed to virtual instrument rooms in office locations (Courtenay, 2019), enabling data analytics and proactive management of critical systems, it remains impractical to transmit all of the uncompressed seismic data recorded each day to the office in near real time using geosynchronous satellite networks. Seismic data processing during the acquisition stage of any project must either be: (1) pursued onboard using available human and computing resources, (2) pursued onshore as the frequency of physical data drops allow, or (3) pursued onshore with strategic data subsets transmitted by satellite (possibly with data compression to reduce file sizes) and processed onshore in parallel with onboard activities.

Most streamer vessels have onboard human and computer resources that enable some form of data processing during acquisition. Fast-track preliminary interpretation products are correspondingly delivered in interim form during acquisition and in final form soon after the completion of acquisition using abbreviated processing flows (e.g., Walker et al., 2019). Processing flows use either testing parameterization or production parameterization with the final choice of parameters in each step. Traditionally, production processing with the full-integrity workflow sequence does not begin until the physical data are received in the office via scheduled data drops.

Assuming that near real-time processing at the rate of acquisition is desired in an office, 2 TB of uncompressed seismic data representing one wavefield component from one day of towed multisensor streamer acquisition will take aproximately one year to transmit using a standard 512 Kbps geosynchronous satellite connection. This reduces to approximately three days using a 64 Mbps connection that represents the upper bandwidth limit typically used for projects seeking near real-time transmission. It is therefore evident that such data must be heavily compressed to enable complete transmission in less than one day, though this remains uncommon. Alternatively, we can transmit strategic subsets of data to the office each day (e.g., shot gathers from one streamer only). Critical onboard QC, such as line acceptance decisions and parameterization of noise removal procedures, only requires subsets of field data (representative combinations of shot gathers, common channel ensembles, or near-field hydrophone

[1]PGS, West Perth, Australia. E-mail: andrew.long@pgs.com.
[2]PGS, Surrey, United Kingdom. E-mail: tony.martin@pgs.com.

data). Such data can be robustly transmitted using low rates of data compression and modest satellite bandwidth connection. Office support of vessel personnel enables rapid and robust decisions for the production processing steps possible within the acquisition timeframe of a project, but the majority of production processing is completed after the acquisition stage.

The frequency of physical data drops from the vessel to the office is linked to the rotation of vessel crew using either large vessel or helicopter transfers (typically between every two and five weeks). This critical path drives the time lag between the acquisition of each sail line and the onset of production processing. The time taken to acquire enough sail lines within swaths with sufficient crossline aperture for full testing of 3D algorithms, such as surface-related multiple elimination and migration, is determined by the length of each line and the overall shooting plan. Hence, full-volume QC may not be possible before much of the physical data have been received in the office.

Alternatively, if high rates of data compression (probably 50–100) are acceptable, all of the daily data could be efficiently streamed to the office, and production processing of the decompressed data could commence without waiting for physical data drops. Perhaps it is time for the industry to accept that data compression/decompression using modern algorithms is as acceptable as the effective signal compression introduced by sparsity-promoting inversion solutions, multichannel transforms, and seismic migration.

Data analytics and processing automation

Testing, validation, and production administration are time consuming for any processing project. Testing is performed to optimize the parameters for each specific step in the processing sequence. Depending on the challenge the step is attempting to address and the complexity of the data, processing testing can require a lot of interactivity with the data, which can be both prolonged and computer-resource intensive.

As indicated, the amount of seismic data processed annually by a globally active contractor can be significant, especially when each step in the sequence has unique characteristics. If the contractor's historical activity can be used to construct a database of parameters applied to all data sets, it can be mined to extract the most appropriate parameters for the data processing. This is based on similarity criteria and considering geologic setting, processing challenges and objectives, acquisition geometry, environmental conditions, and specifics of the processing sequence. The collective expertise and experience of contractor personnel stored in a database is an undeniably powerful tool for reducing turnaround. The data could be mined to focus testing parameterization and reduce testing turnaround or to bypass testing altogether.

A 400 km^2 POC test was run with data from Indonesia, where key processing parameters for all steps in both the data domain preprocessing and migration were mined from a database. No testing was performed, and all workflows were actioned end on end. The resulting raw migration was then compared to the full-integrity processing project whose parameters were excluded from the database and which was run in advance of the testing. Figure 1 shows a comparison of the data from the (independent) full-integrity work compared to that where parameters have been determined in advance of the project and run without testing.

The migrated stacks look similar. However, quantitative comparison metrics were run, including correlation analysis, normalized root-mean-square difference (NRMSD), and signal-to-noise (S/N) content, to further analyze the two volumes. QCs were run after each key processing step, but at no point did they affect the original (mined) parameter choices, and for brevity, only the final comparisons are shown. Such metrics are common to 4D processing and are therefore a good indicator for comparing the full-integrity

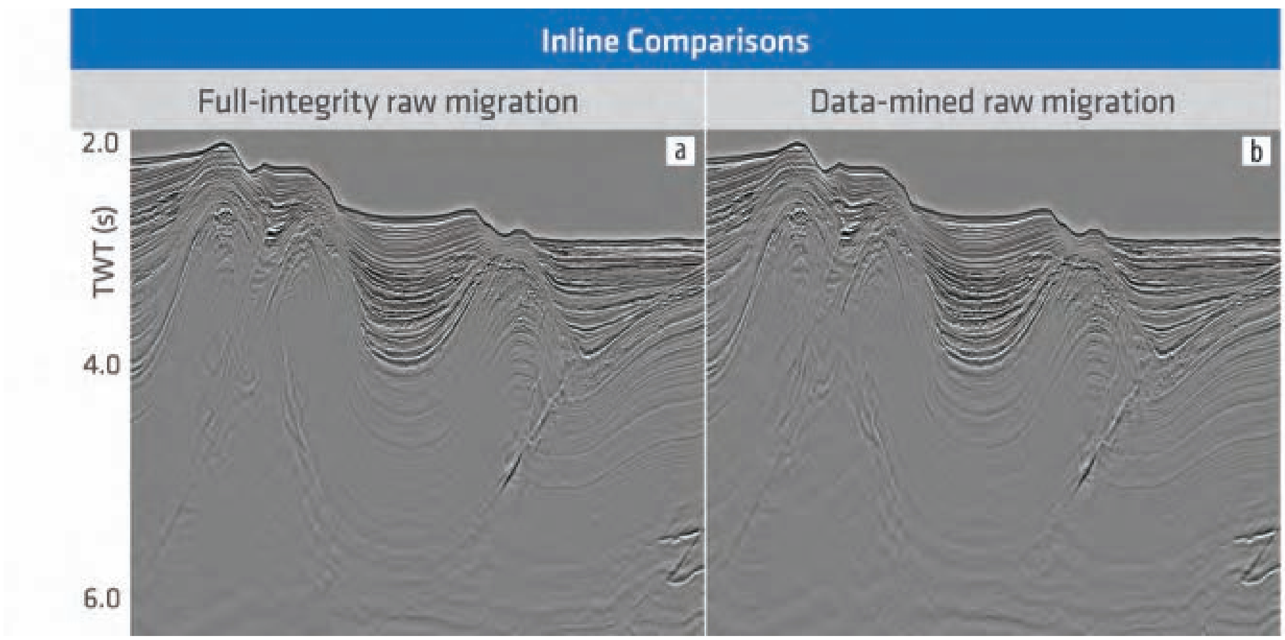

Figure 1. (a) A raw migration stack response comparison of a full-integrity processing project. (b) An automated approach using data mining of a parameter database.

volume and the automated equivalent. Correlation analysis between the two volumes (Figures 2a and 2b) and NRMSD (Figure 2c) highlight that deeper data are slightly noisier. Figure 2d suggests that the automated processing nevertheless preserved phase integrity. The S/N content in Figure 3b indicates that the full-integrity data have a slightly better response (notably 30–70 Hz), albeit marginal. Overall, the data quality from the database-mined processing automation is equivalent to the full-integrity process and was achieved in one-third of the time taken to create the full-integrity volume. As with all seismic processing projects, an equivalent level of success cannot always be expected. However, as such parameter databases become more sophisticated and better populated, the principles herein should be broadly applicable.

The only caveat in achieving comparable results in this processing automation POC work is the use of an a priori velocity model in the migration, which for comparison sake was taken from the full-integrity project. In a later section, we consider automation of the velocity model used for depth migrations, but first we address the obvious question of how the parameter selection can be efficiently validated.

Automated QC: Supervised large volume noise removal

Most onboard line acceptance and QC activities during marine seismic acquisition are based on the assessment and removal of noise in many thousands of shot records. Once the field data are accepted, modern seismic data processing flows typically have 15 to 20 major components, each having unique characteristics managed by intermediate data outputs. Traditional QC has relied heavily on visual inspection of the prestack and poststack results of multiscenario parameter testing and attribute generation. However, the simultaneous assessment of many attributes is subjective, empirical, and challenging.

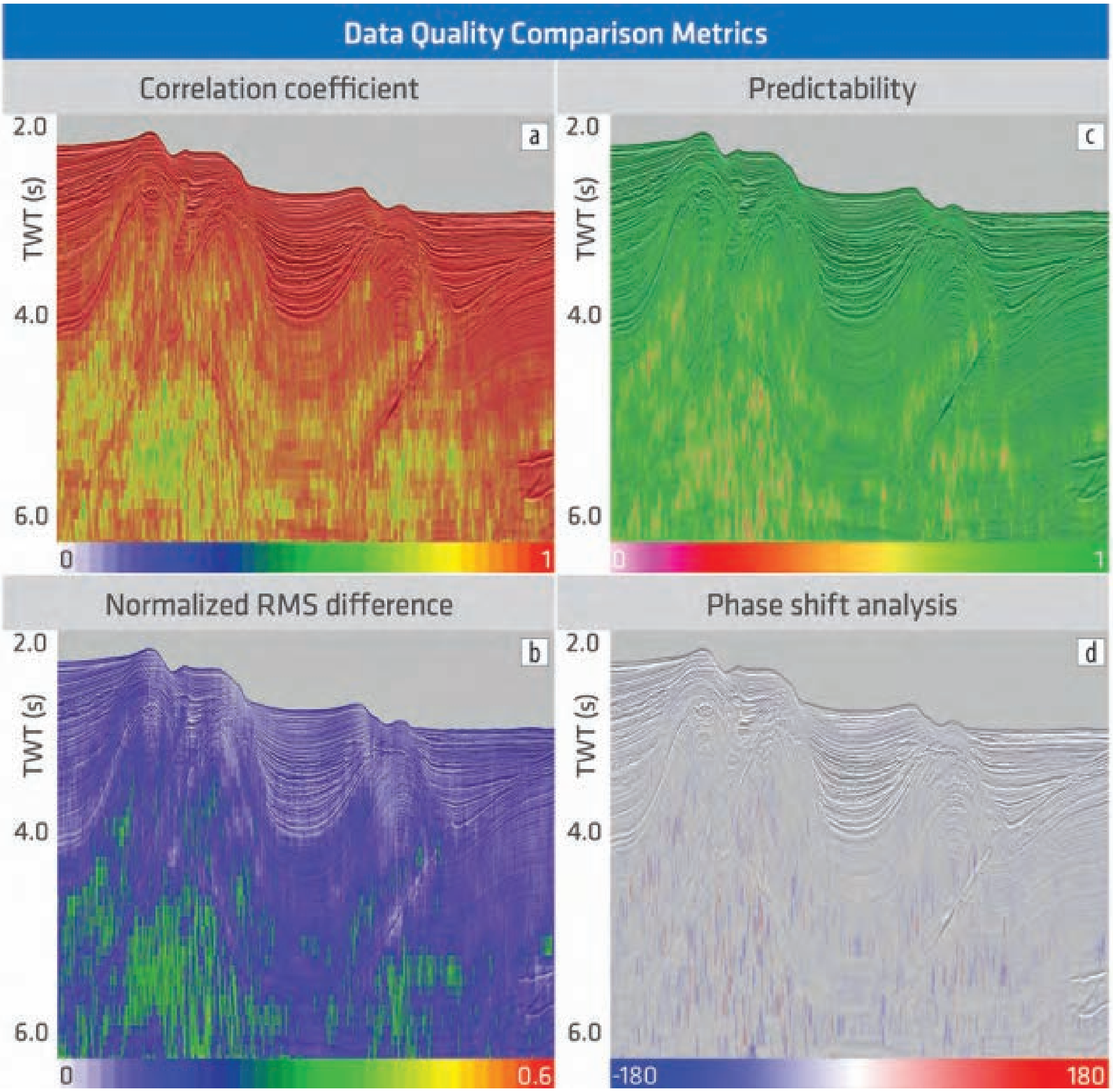

Figure 2. (a) Correlation coefficient. (b) Predictability. (c) NRMSD. (d) Phase.

Marine seismic data sets contain highly redundant information, so data analytics and deep learning-based solutions provide opportunities to reduce turnaround and improve confidence levels on output data volumes. As previously alluded to, early-stage processing QC occurs in concert between onboard and onshore resources, enabled by satellite transmission and data compression. As the volume of data in a typical survey has increased over time, QC practice has moved toward assessing global attribute maps that are computed from the data, such as root-mean-square amplitude or S/N maps. However, such simplistic tools require frequent cross-checks with the seismic data. The focus is on detecting outliers and anomalies, and humans cannot understand the visualization of more than two or three attributes at a time. Clearly, we want to compute as many informative attributes as possible to give a better sampling of the filtering performance. This can be facilitated by using statistical data mining techniques to analyze the different attributes. Correspondingly, Bekara and Day (2019) describe a relevant POC supervised learning framework for automatic denoise classification that expands on the unsupervised outlier detection methodology of Spanos and Bekara (2013). Their example applies to one step (denoise) of a processing flow, of which there will be several in practice. Six sail lines evenly dispersed throughout a semicompleted multisensor streamer survey were split into training and validation data sets of raw shot gathers. Shot gather-based multidimensional statistical attributes measuring the similarity between the output of various degrees of noise removal and the difference between input and output were computed within time-spatial windows. Similarity will increase with increasing signal leakage into the filtering.

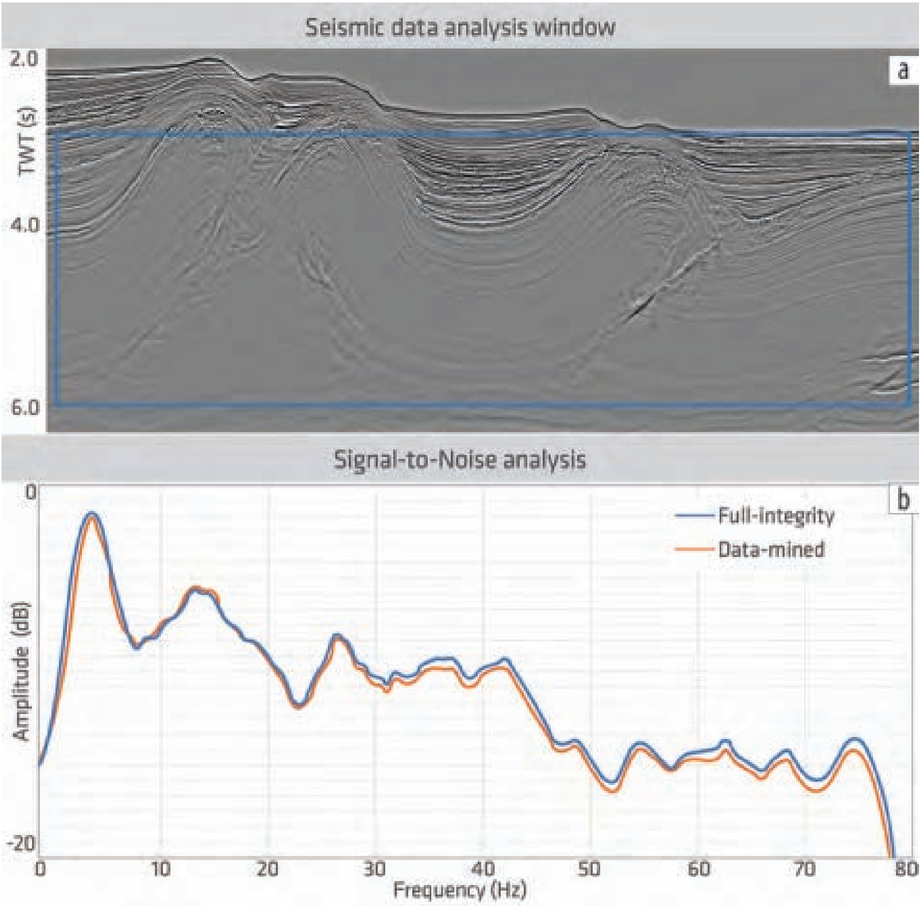

Figure 3. (a) Analysis window used to compute the S/N attribute. (b) S/N comparison of the full-integrity and data-mined results.

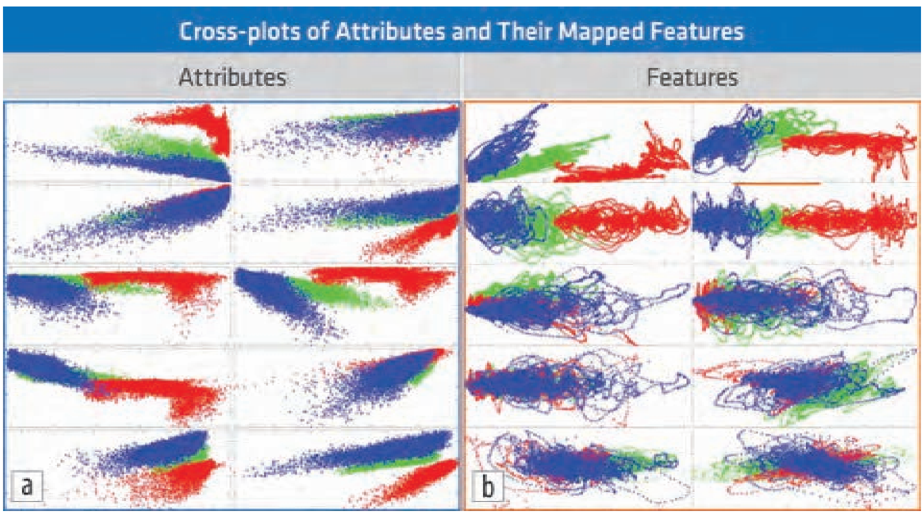

Figure 4. (a) Crossplot of five attributes and (b) the equivalent crossplots of five principal components computed after spatial augmentation of the attributes. Each dot within the crossplot distribution of the three colors of red (mild), green (optimal), and blue (harsh) represent one filtered shot gather. Note that visual separation between the different clusters has improved for the primary principal components, and the corresponding decision space yielded negligible false-positive results by comparison to the result based on attributes. From Bekara and Day (2019).

The crossplots of five different attributes computed from three test lines are shown in Figure 4. These are only shown to validate the attributes, which are overlaid for the optimal, harsh, and mild filtering cases using a three-color code (mild is blue, optimal is green, and harsh is red). 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 decorrelating the attributes to extract useful structure 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 4b shows the cluster of features obtained after applying a nonlinear mapping (spatial augmentation with 20 shots followed by PCA) on the cluster of attributes in Figure 4a. A supervised classification based on support vector machines (Cristianini and Shawe-Taylor, 2000) was constructed using the training data, yielding three decision spaces corresponding to optimal, mild, and harsh filtering. When using the attributes to train the machine learning classifier, those selected were informative, as the training error for all three scenarios was negligible (< 3%). The validation error for harsh and mild filtering was similarly small; however, about 20% of the optimal filtering points were initially misclassified as mild or harsh filtering. This error significantly decreased (from 20% to 1%) when the machine learning classifier was trained instead with the features. As noted in the previous section, the POC example may not necessarily be as successful elsewhere for this equivalent processing step. Other major processing components would need different attributes within the same learning framework. However, the strategy of making better-informed decisions with more data references should remain robust.

Figure 5 shows a tricolor decision map for every available shot in the POC study. Subsequent evaluation of the shot locations,

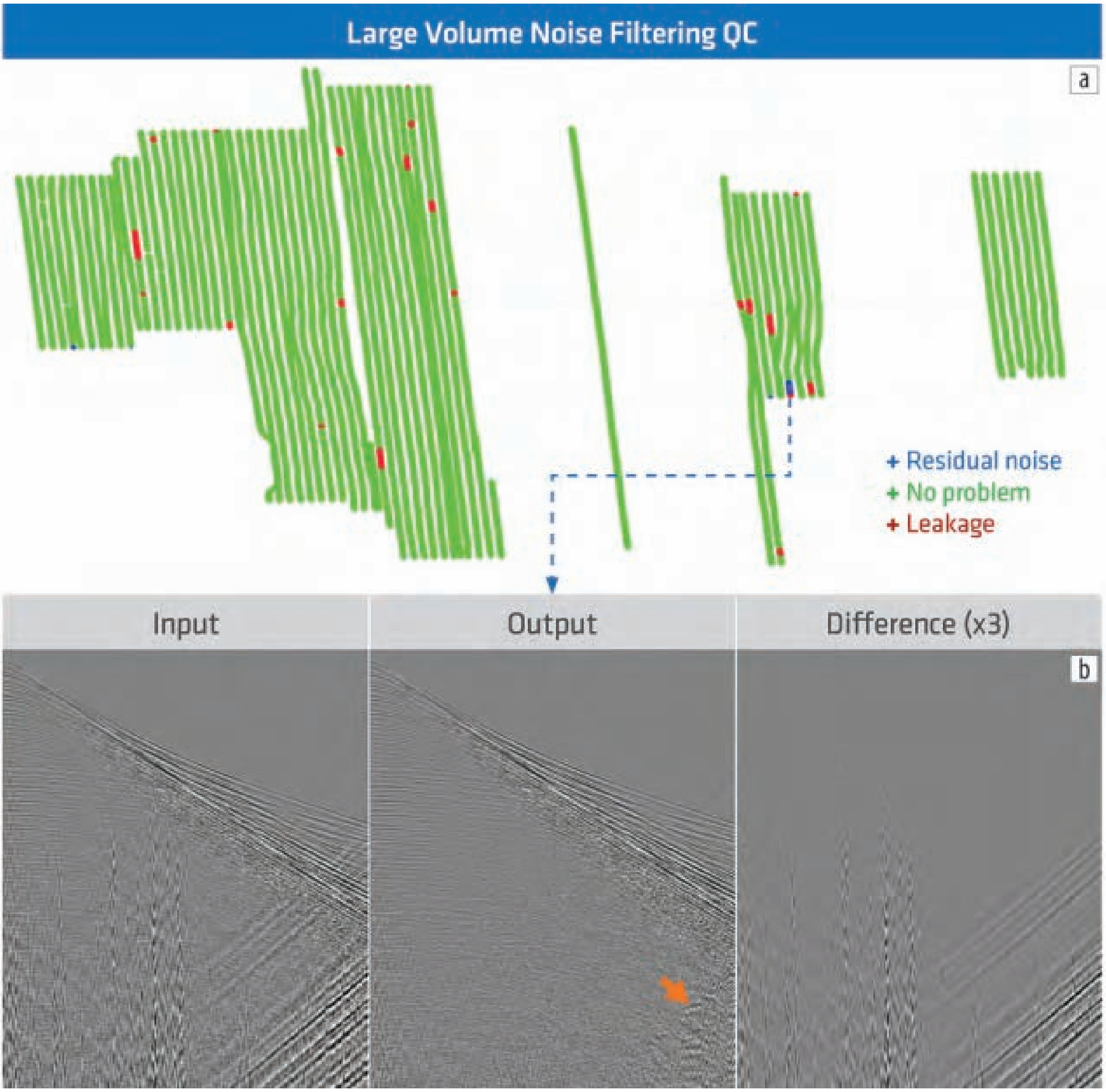

Figure 5. (a) Classification of all shot locations. (b) An example of a shot gather identified as requiring residual noise removal. The decision map contains one point for each shot gather location. The colors follow the same scheme used in Figure 4.

identified by blue points, would result in additional residual noise removal. Red points correspond to false positives produced when training the solution with attributes.

In the dynamic offshore environment, the described approach would help focus attention on any priority areas with potential problems, thereby optimizing the use of resources working within challenging timeframes. More generally, supervised classifiers should enable global quantitative decisions based on many relevant data attributes, moving behind the subjective art of observational QC. While the POC example shown is for validating and classifying denoise, the philosophy could be extended to other major processing steps. Looking forward, the development of feedback loops will enable processing flows with even higher levels of automation. For example, level eight in the hierarchy of Sheridan and Verplank (1978) is "computer does whole job and tells human what it did only if human explicitly asks."

Automated velocity model building

Any fast-track products or progressive interpretation deliverables, such as angle-range gathers and stacks, explicitly depend on the early availability of an accurate velocity model for the entire data set. Simple velocity picking by onboard personnel or by office personnel using remote sessions to the onboard computers is robust during acquisition. A reasonable starting model can be produced rapidly with a short time lag after the receipt of data in the office. If data compression is acceptable to the client, there is no technical reason why highly compressed (and possibly subsampled in time) shot gathers could be transmitted to the office in near real time for input to full-waveform inversion (FWI), especially given that irreversible signal distortion from high compression rates is generally prevalent at higher frequencies of negligible relevance to FWI. Therefore, an FWI-based velocity model could in principle be ready when the physical data drop is received by the office, enabling zero wait to progress to demultiple, assuming that all shot domain denoise pursued on the vessel met the project technical ambitions. Furthermore, if elements of the demultiple workflow have also been completed on the vessel and/or in the office before the physical data are received, the time between data receipt and the commencement of imaging will be further reduced (e.g., Saint Andre et al., 2010).

More generally, model building for depth imaging is one of the largest bottlenecks in processing workflows as well as one of the most critical steps. Such models are used to provide an image of the subsurface, from which a range of probabilities and volumetric estimates may be made and drilling campaigns planned and then actioned. Although FWI represents the pinnacle of velocity model building (VMB) for many practitioners, its high computational cost makes it impractical for scenario testing of different model realizations or uncertainty. Deep model building is often challenging for standard streamer lengths, even if cycle-skipping-mitigated full-wavefield FWI is achievable (e.g., Ramos-Martínez et al., 2019). Considerable scope still exists for pragmatic non-FWI solutions to augment faster processing workflows.

Bell et al. (2016) describe the use of a Monte Carlo simulation that enables multiple realizations in order to derive estimates of the uncertainty of an individual velocity model. The method performs multiple random perturbations of a starting model followed by tomographic inversion. This platform uses an efficient

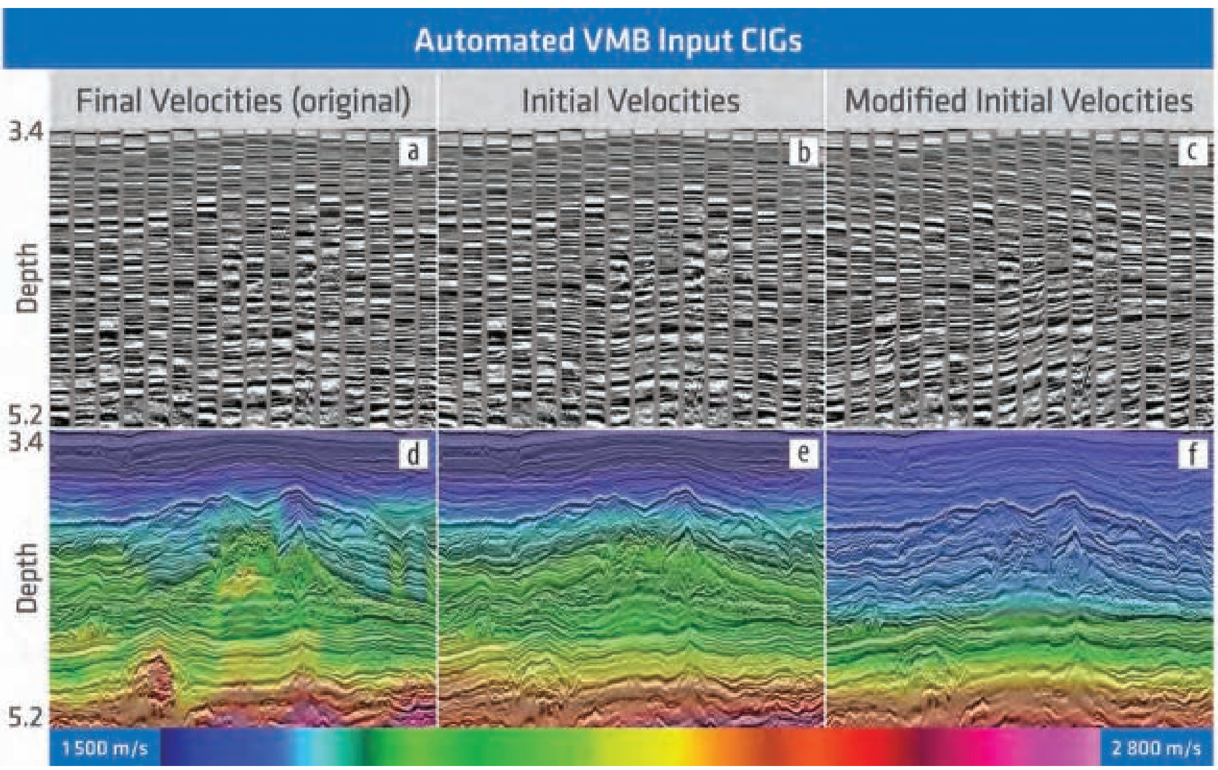

Figure 6. CIGs for the (a) final tomographic model, (b) initial model for (a), (c) modified and locally erroneous initial model, and (d)–(f) migrated stacks with corendered velocities corresponding to CIGs in the upper row.

beam migration to establish the initial ray kinematics of the invariant data, which comprise wavelets extracted from the data through a multidimensional dip scanning process (Sherwood et al., 2008), performed within the migration model space generating the observed data. The process of model perturbation is performed in a residual migration and applies the differential kinematic to the observed data, consistent with the applied perturbation. Rather than look at the uncertainty of a single model and the imaging products, the methodology can also be adapted to create a depth imaging velocity model from scratch using either a benign or incorrect starting point through the same Monte Carlo simulation of the model space.

The starting point for the full automation of VMB in Martin and Bell (2019) begins with the same steps of determining what the data support in the model space prior to creating a randomly generated model population. Once generated, the population is tomographically inverted, and statistical analysis is performed on the model updates prior to reintroducing a pass of random model generation. The process is repeated with the goal to produce a model that explains the data by producing flat common-image gathers (CIGs) that have a zero residual for tomographic inversion. This is quantified by determining moveout-related metrics after each pass of the simulation. Convergence of the solution determines how many iterations are used.

A 500 km^2 data set from West Africa was used in a POC test to reduce the time taken to produce a model by removing human intervention. Two initial models were tested: the starting model used for the actual tomographic model building project and one where the initial model was modified to incorporate a locally varying error up to 10% in the starting model. Once randomly perturbed, the secondary starting model could be locally up to 15% too fast or slow. The results were checked against the final tomographic model, which was built using the same data and generated in 90 days.

Figure 6 shows three sets of CIGs and three stacks with their associated velocities corendered on the seismic sections. Figures 6a and 6d are the result of the 90-day model building exercise. The central image shows the starting point for the automated Monte Carlo model building process. The starting CIGs in Figures 6b and 6c show a significant level of moveout, as the model was up to 15% wrong. The results in Figures 7b and 7c show the product of the automated model building. Gather flatness is equivalent to the conventional approach (Figure 7a), and the corendered velocity models closely resemble the model built in 90 days.

Progressive analysis of metrics on moveout show an equivalent level of convergence in the resulting models, irrespective of the starting point (Figure 8). The workflows were initiated by a geophysicist who had no prior knowledge of the data or models, and no well constraints were available to confirm the accuracy of any of the resulting models. The implications of this approach are considerable. While the original model building project took 90 days, both automated models were achieved in less than an order of magnitude of that time.

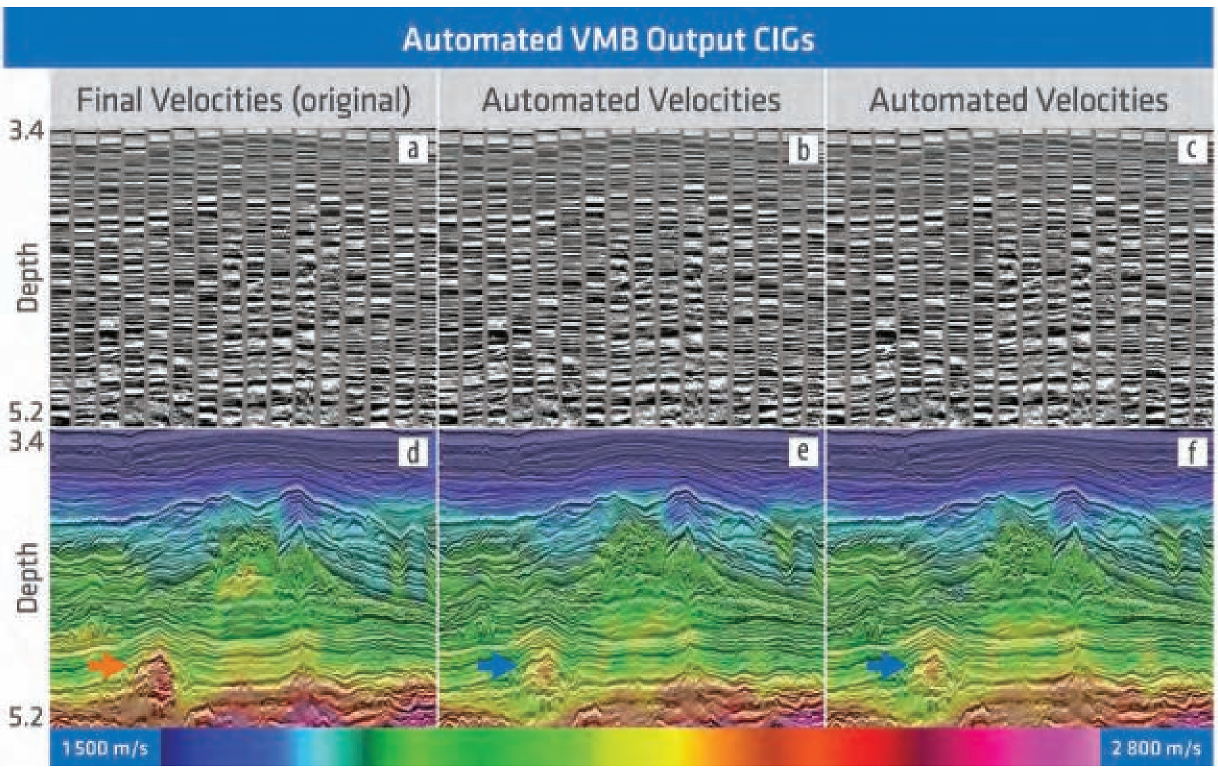

Figure 7. CIGs for the (a) final tomographic model, (b) final automated model starting with Figure 6b, (c) final automated model starting with Figure 6c, and (d)–(f) migrated stacks with corendered velocity models corresponding to CIGs in the upper row. The orange arrow in (d) shows the location of the masked and updated geobody (channel). Blue arrows in (e) and (f) show the channels captured with the automated approach. The automated models in (e) and (f) otherwise show a strong correlation with the model built during a conventional velocity model workflow.

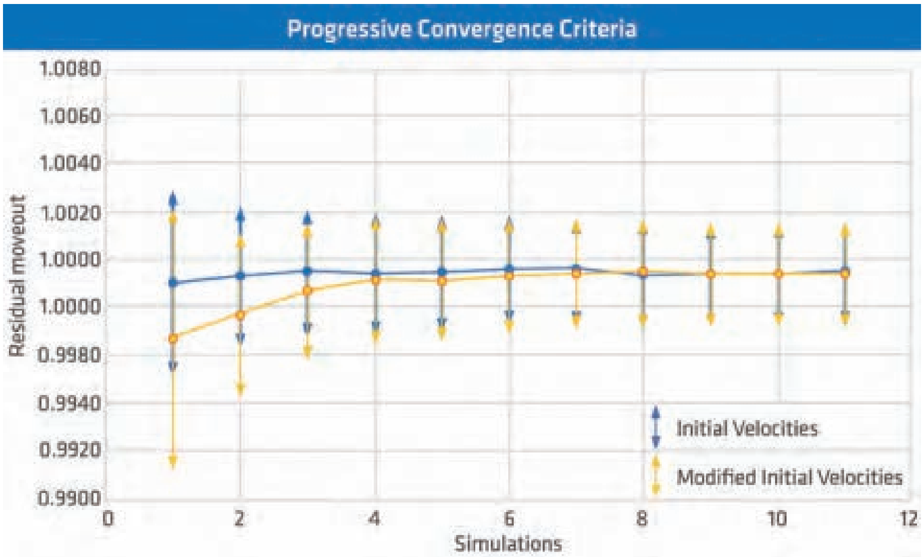

Figure 8. Moveout convergence criteria QC. The blue curve using a starting model was derived from semblance-based velocity picking. The orange curve using a starting model is shown in Figure 6b. Both blue and orange curves converge to the same level of moveout.

Summary

The progress from the sequential series (with many steps and interactive QC events in legacy seismic processing flows) to full automation will occur in a piecemeal fashion as the industry learns to embrace what will essentially be a hands-off paradigm. Towed-streamer marine seismic surveys can acquire vast data volumes each day, presenting an early-stage project challenge to cost-effective near real-time streaming of the data to onshore supercomputer facilities using geosynchronous satellite networks. An acceptance of high rates of data compression and/or the sharing of strategic subsets of data with onshore resources is the pragmatic solution to initiate production processing early during the acquisition stage.

Our POC example demonstrated that a collectivized digital experience database can be mined to fully parameterize several consecutive processing steps without human intervention. An efficient QC system is correspondingly necessary to validate such an approach. A supervised learning example of efficient denoise QC is demonstrated as being a potentially efficient platform for using all of the data acquired to augment better acquisition QC decisions in less time. It presumably heralds the way to similarly augment more efficient QC for other steps in a typical processing flow.

Automated parameterization validated with efficient and robust QC platforms is also particularly relevant for automated VMB, as data conditioning is inevitably required before VMB, including FWI. Although FWI represents the pinnacle of model building VMB for many practitioners, considerable scope still exists for pragmatic non-FWI inversion solutions to augment faster processing workflows. Correspondingly, an efficient wavelet-based beam migration platform was shown in a large POC study to accurately recover depth velocity models using Monte Carlo-based tomographic inversion of moveout residuals, even when the starting model was highly inaccurate. Overall, a pragmatic combination of supervised deep learning, data analytics, and efficient imaging solutions can deliver substantial reductions in project turnaround while balancing human interaction and full automation. Further iterations of this workflow with embedded feedback loops would improve the level of automation. TLE

Acknowledgments

We thank Suhail Butt and Khairom Baharom for help with the Indonesian data test, Marcus Bell for help with the automated velocity model building, and PGS for permission to publish this work.

Data and materials availability

Data associated with this research are confidential and cannot be released.

Corresponding author: andrew.long@pgs.com

References

Bekara, M., and A. Day, 2019, Automatic QC of denoise processing using a machine learning classification: First Break, **37**, no. 9, 51–58.

Bell, A. C., L. Russo, T. Martin, D. van der Burg, and B. P. Caselitz, 2016, A workflow to quantify velocity model uncertainty: 78th Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.201601098.

Courtenay, P., 2019, Building a dream team: Oilfield Technology, **12**, no. 5, 13–15.

Cristianini, N., and J. Shawe-Taylor, 2000, An introduction to support vector machines and other kernel-based learning methods: Cambridge University Press, https://doi.org/10.1017/CBO9780511801389.

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

Martin, T., and M. Bell, 2019, An innovative approach to automation for velocity model building: First Break, **37**, no. 6, 57–66.

Ramos-Martínez, J., L. Qiu, A. A. Valenciano, X. Jiang, and N. Chemingui, 2019, Long-wavelength FWI updates in the presence of cycle skipping: The Leading Edge, **38**, no. 3, 193–196, https://doi.org/10.1190/tle38030193.1.

Saint Andre, C., N. Khaled, L. Bovet, J. Brittan, M. Bekara, T. Martin, and M. Phillips, 2010, Angola wide azimuth towed streamer (WATS): 3-D SRME whilst shooting: 80th Annual International Meeting, SEG, Expanded Abstracts, 3416–3420, https://doi.org/10.1190/1.3513558.

Sheridan, T., and W. Verplank, 1978, Human and computer control of undersea teleoperators: MIT Technical Report.

Sherwood, J. W. C., K. Sherwood, H. Tieman, and K. Schleicher, 2008, 3D beam prestack depth migration with examples from around the world: 78th Annual International Meeting, SEG, Expanded Abstracts, 438–442, https://doi.org/10.1190/1.3054840.

Spanos, A., and M. Bekara, 2013, Using statistical techniques to improve the QC process of swell noise filtering: 75th Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.20130884.

Walker, M., D. Davies, C. Hill, C. Page, P. Smith, and A. Irving, 2019, Full-cycle iterative processing: When is "good", good enough? A 4D North Sea case study: 81st Conference and Exhibition, EAGE, Extended Abstracts, https://doi.org/10.3997/2214-4609.201901578.