Seismic processing parameter mining – the past may be the key to the present

Tony Martin[1*], Andrew Long[1], Suhail Butt[1] and Khairom Baharom[1] describe a proof-of-concept test, mining an historical database to extract appropriate parameter choices for seismic processing. Using quantitative metrics, they compare the results of using the mined parameters with a full-integrity project, and draw conclusions on whether this approach can bypass testing and reduce turnaround times.

Introduction

Data volumes are increasing, seismic processing flows are becoming more complex, and pressure is growing to reduce project turnaround times. Automation may help with some of these contradictory challenges. Parameter testing, validation and administration are protracted and manually intensive. Testing optimizes the geophysical parameters for each step in the processing flow. Depending on the processing step and data challenges, it may require a lot of interaction with the data, which can be time consuming. Testing may also need many repeat-runs with minor adjustments to parameters, and as a consequence computer resources can be significant. We present an example from a Malaysian thrust belt data set, and demonstrate parameter testing can almost be eliminated if the project parameter data can be mined from an historical parameter database.

Method

The amount of seismic data processed each year is considerable and geographically diverse. An individual data set and processing sequence may have some unique characteristics, but will often share similarities with other data or flows, especially for seismic processing parameterization. This common ground may be exploited to help reduce the turnaround time in projects.

Determining or discovering useful information about data is often referred to as data analysis. Data mining is a field within data analysis and computer science which uses databases to extract knowledge that may be both useable and transferable to a variety of applications. It is not the extraction that is key, but the statistical analysis that leads to useful information, such as trends and similarities.

The historical activity of a seismic processing contractor may be a useful tool to help fine-tune parameter testing or bypass this process altogether. If all the parameters from all the major processing steps and historical projects could be collated into a parameter database, this may be mined for parameter trends relating to similarity based on key criteria such as the following: geological settings, processing challenges and objectives, acquisition geometries, environmental conditions and the processing sequence. An extraction of the useful information may then be carried forward to provide a focused starting point to a testing phase, or if the analysis shows a high similarity, exclude testing completely by using the mined parameters. The collective expertise and experience of contractor personnel stored in a database is an undeniably powerful tool for reducing turnaround.

To demonstrate whether the prototype is fit for purpose, we use a proof-of-concept project (POC) to verify that the underlying

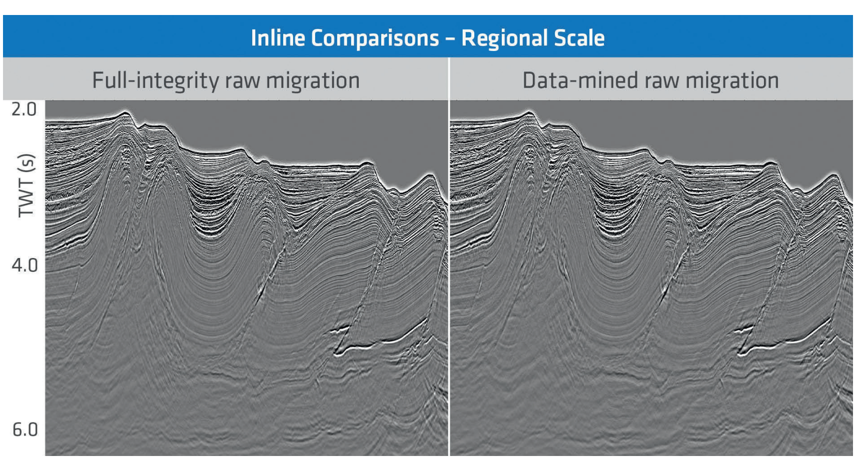

Figure 1 A regional scale raw migration stack response comparison of a full-integrity processing project (left); and a semi-automated approach using data-mining of a parameter database (right).

[1] PGS
[*] Corresponding author, E-mail: tony.martin@pgs.com
DOI: 10.3997/1365-2397.fb2020043

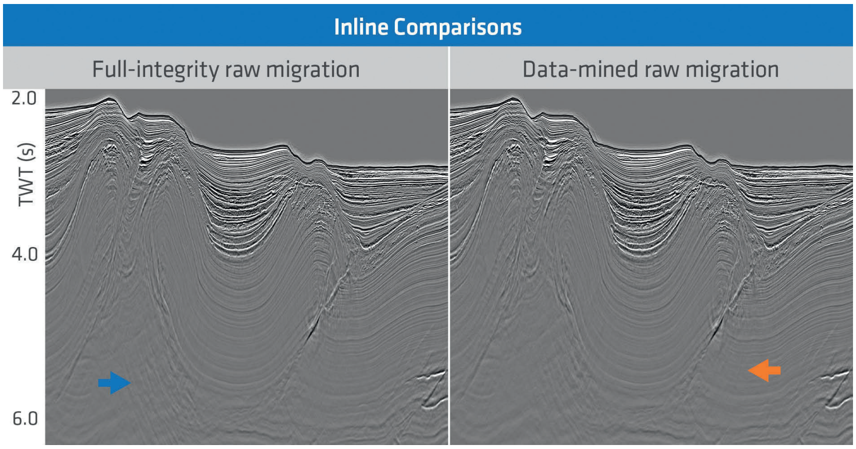

Figure 2 A zoom scale raw migration stack response comparison of a full-integrity processing project (left); and a semi-automated approach using data-mining of a parameter database (right). Blue arrow (left) highlights more coherent noise, whilst the orange arrow (left) indicates more scattered noise.

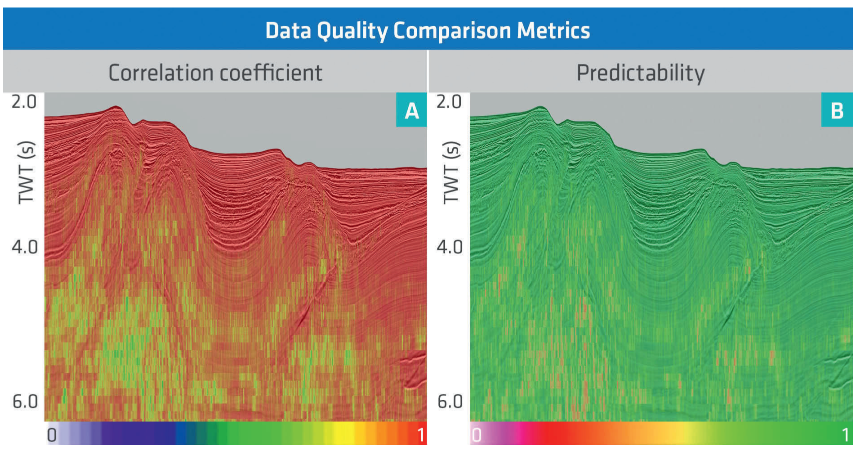

Figure 3 (A) Correlation coefficient and (B) predictability.

hypothesis can be met using a real-world example. An historical record of processing parameters was developed for the major processing steps of a typical marine seismic processing project. Information about the key parameters based on ranked similarity to the POC data set were then extracted. One key consideration is that we were not trying to predict parameters based on a trained system and labelled data, but determine statistical trends of similarity from existing labelled data. Following the parameter mining, workflows for all the major steps were generated and then run non-stop using the appropriate computer resources. No testing of parameter variations was performed.

The goal of the POC work was to establish whether this type of semi-automated data mining approach could produce meaningful or equivalent results. Quantitative metrics were used to understand the data quality of the database-mined work compared to a full-integrity processing project. Once established, statistics were produced to understand potential turnaround benefits. Conclusions were drawn on the validity of the process along with possible scenarios where this type of approach could be used. Finally, we evaluated where the method might fail.

Case study

The Malaysian Sabah 3D data set was acquired with 16 multi-sensor streamers, each 100 m apart, and in an environment where water depths vary from 100 m to 3000 m. The area contains a fold and thrust belt within the Miocene turbidites, which are a proven petroleum system. A 400 km^2 subset of the data was used as a POC test, where key parameters for all steps in the pre-processing and migration stages of the project were mined from a parameter database. No testing was undertaken for the data-mined solution, and all processing workflows were manually created and run non-stop.

The Sabah data were chosen because the data is complex, representative of many large-scale deep-water exploration projects, contained all typical noise types and had a standard signal-to-noise ratio (S2N). Additionally, the deep-water environment and complex geology simplified the evaluation of the demultiple effectiveness. Finally, the data had recently been processed through a full-integrity sequence, and was available for comparison.

After all the non-stop workflows had been run, the raw migration was then compared to the full-integrity processing project whose parameters were excluded from the database. Figure 1 compares the regional scale data from the full-integrity work and the data-mined approach, where parameters were determined in advance and run without testing. The migrated stacks look similar. There are subtle indications that the S2N and multiple content differs between the volumes.

The orange arrow in the zoomed section of Figure 2 indicates that there is more scattered noise in the data-mined version, whilst the blue arrow shows that there is also more coherent noise in the full-integrity data.

Quantitative comparison metrics were generated, including correlation analysis, normalized RMS difference (NRMSD), and S2N, to further analyse the two volumes. Quality control (QC) checks were run after each key processing step, but at no point did they affect the mined parameter choices. For brevity only the final comparisons are shown. These metrics are common to 4D processing, and are therefore a good indicator of equivalence between the full-integrity volume and the semi-automated result.

Correlation analysis between the two volumes (Figures 3A and 3B), and NRMSD statistics (Figure 4A) highlight that the deeper data-mined example is slightly different, indicating a subtle difference in noise content between the two volumes. Figure 4B suggests that the automated processing nevertheless preserved phase integrity. The S2N content in Figure 5B suggests the full-integrity data has a slightly better response (notably 30-70 Hz), albeit marginal.

Parameter comparisons for the two different methods show the most significant variation between the POC test and the full-integrity project were in the level of incoherent noise attenuation, harshness of model adaption for the demultiple step and operator limitations in the migration. These reinforce the conclusions drawn from the quantitative metrics, and the visual inspection of the data.

Considerations

To achieve comparable imaging results in this POC work we used a pre-existing velocity model for the migration, which for comparison sake was taken from the full-integrity project. Modern velocity model building methods use inversion schemes, such as tomography or full waveform inversion (FWI). The accuracy of the result depends on many factors. With tomography the

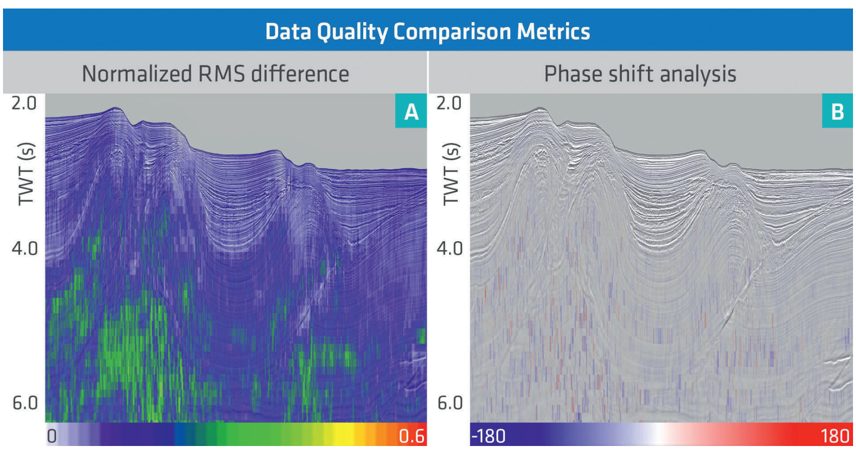

Figure 4 (A) NRMSD and (B) phase shift.

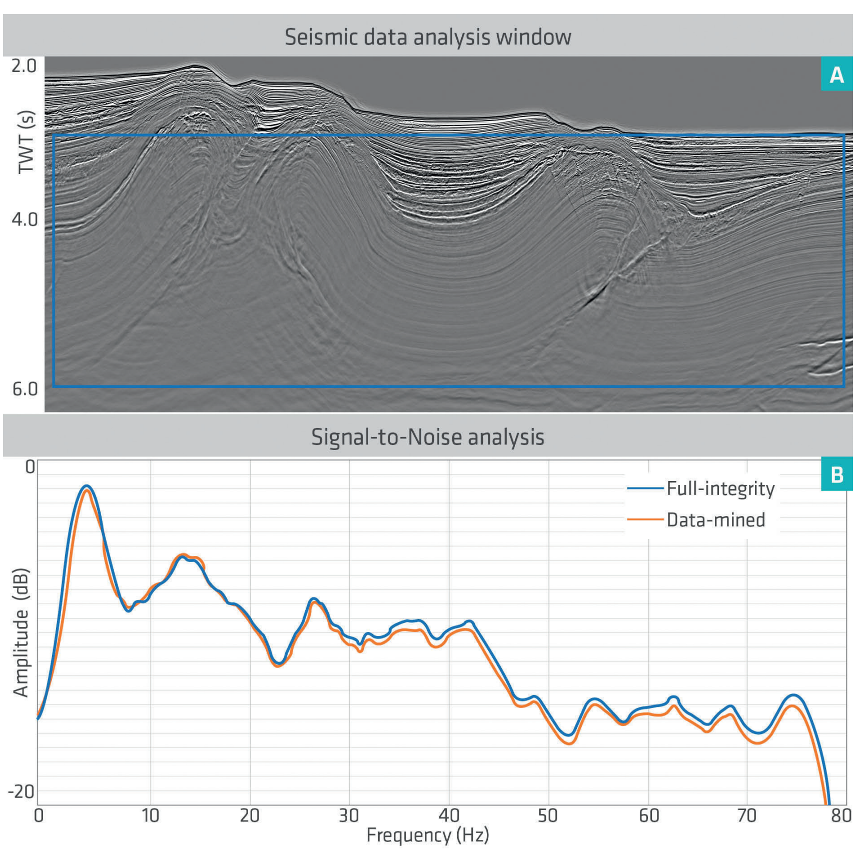

Figure 5 (upper) Analysis window used to compute the Signal-to-Noise attribute; and (lower) Signal-to-Noise comparison of the full-integrity and data-mined results.

problem is under/mixed-determined; there can be many unconstrained but equiprobable models generated by the inversion scheme. This may result in a level of uncertainty in the velocity model, and the migrated image using this model, which can be measured using uncertainty analysis (Bell et al., 2016).

If this kind of parameter mining and extraction were used to improve turnaround, automation of the velocity model building should also be considered. Monte Carlo simulations with crude reinforcement mechanisms are one way to achieve this (Martin and Bell, 2019). If this or equivalent methods were used in conjunction with mechanized QC, near full automation may be achievable, where the computer does the entire job and tells the human what it did (Sheridan and Verplank, 1978).

Overall, the data quality from the semi-automated database-mined processing is equivalent to the full-integrity work. Scaled for volume size, the result was achieved in one-third of the time. As with all seismic processing projects, an equivalent level of success cannot always be expected, but as parameter databases evolve and are better populated, and mining approaches become more sophisticated, the principles herein should be broadly applicable.

Discussion

In our example, we show a pragmatic solution using the historical record to reduce testing turnaround through parameter extraction. If this kind of approach was implemented to reduce the time taken for seismic processing, the entire workflow would also need quick validation of the extracted parameter's effectiveness. Machine learning-type QC is a way in which supervised and unsupervised learning can be used to quickly corroborate the semi-automated parameterization of processing modules (Bekara and Day, 2019; Martin et al, 2015).

How much further can we progress to full automation? For example, semi-automated parameterization validated with efficient and robust QC may be more relevant for automated velocity model building (VMB), as data conditioning is inevitably required before VMB, including Full Waveform Inversion (FWI). Equally, mined parameterization might also provide a more robust fast-track seismic processing data set, especially if used with automated QC and model-building techniques.

The evolution of a database to incorporate new technology and developments would need attention. The inclusion of seismic processing algorithms that may improve data quality will be absent from the database until an appropriate usage enables its addition to the statistical record.

In this example, turnaround improvements appear to be achievable with comparable seismic data quality, but to do this it requires a more stand-back approach. Unlike other automation techniques, this methodology does not involve a transfer of resources and therefore computer usage may also decrease as parameter testing is bypassed. However, to achieve the reduction in turnaround it requires minimal interference; it will be necessary to decrease the participation by both the contractor and customer. This is no different to other forms of automation that are being enabled by machine learning approaches. The big question is, if reducing turnaround time is critical, are individuals and companies in the seismic processing industry ready to step back and allow automation to evolve?

Conclusions

Our POC example demonstrates that a collectivized digital experience database can be mined to extract the processing parameters for several consecutive processing steps without human intervention. The data quality metrics generated from the raw prestack depth migration show that, for this data, the seismic quality and similarity are comparable to a full-integrity project. Auxiliary data, including the velocity model, were created during the full-integrity work, and most supplementary data are generated in parallel with data processing. Nevertheless, the turnaround benefits from automation for the testing and production phases of the seismic processing were significant, as the database-mined sequence was completed in significantly less time than the full-integrity work.

Acknowledgements

The authors wish to thank PGS for permission to publish the paper.

References

Bekara, M. and Day, A. [2019]. Automatic QC of denoise processing using a machine learning classification. *First Break*, **37**(9), 51-58.

Bell, A. C, Russo, R., Martin, T., van der Burg, D. and Caselitz, B.P. [2016]. A workflow to quantify velocity model uncertainty. *78th EAGE Conference & Exhibition,* Extended Abstracts, We P7 11.

Martin T., Saturni, C. and Ashby, P. [2015]. Using machine learning to produce a global automated quantitative QC for noise attenuation. *85th Annual International Meeting,* SEG, Expanded Abstracts, 4790-4794.

Martin, T. and Bell, B. [2019]. An innovative approach to automation for velocity model building. *First Break,* **37**(6), 57-66.

Sheridan, T. and Verplank, W. [1978]. Human and computer control of undersea teleoperators. *MIT Technical Report*, 187.