

# An innovative approach to automation for velocity model building

Tony Martin<sup>1\*</sup> and Marcus Bell<sup>1</sup> describe the use of a Monte Carlo simulation, enabling multiple realizations of the solution in order to derive estimates of the uncertainty of an individual model, as well as drive velocity model building in an automated fashion.

## Introduction

While some efforts have been made to quantify structural uncertainty on seismic images (Osypov et al., 2011; Letki et al., 2013), the seismic data processing industry has found it challenging to measure the effectiveness of processing algorithms on seismic data. Understanding the success of a single algorithm may be time consuming, require significant work, including research and development, and is therefore also costly. Despite this, the demand for ‘error bars’ on processes is growing, while expectations are that projects should be completed faster. At the same time seismic projects are getting bigger. It is not uncommon to see projects recording up to 20,000,000,000,000 samples. For now, each project will typically have 15 to 20 major processing components, which are managed by intermediate data outputs, each having unique characteristics. More than any other industry, the seismic acquisition and processing companies should be at the forefront of big data analysis. However, there has been little progress from the industry in advanced analytics or artificial intelligence, and there has been no major effort in companies metamorphosing from geophysics to geophysical data science. Therefore, the innovation must come from within, by the geophysicists who use processing workflows on a daily basis. Modifications and manipulations to established systems enable advanced analytics, reductions in turnaround and the opportunity to provide confidence levels on output data volumes.

An example of where this can be demonstrated is building depth imaging velocity models. A model is used to provide an image of the subsurface, from which a range of probabilities and volumetric estimates may be made, drilling campaigns planned and then actioned. Seismic data underpins all of this, and the data’s spatial veracity is dependent on the model used to reposition the data, but the entire acquisition and processing flow is refined into one single model representation. Although other factors are important, the seismic processing project and its deliverables are all about the earth model.

While there is one ‘true’ answer, many models will give an equally acceptable solution based on the method used to derive the model, its convergence criteria and measures defining the success of the resulting model. The model can be verified, or even constrained by auxiliary data, but in most cases this data is sparse,

third-party processed, or subjective. Therefore, in isolation, a single model provides little useful evidence of the reliability of any individual model, or the seismic data.

In this study we describe the use of a Monte Carlo simulation, enabling multiple realizations of the solution in order to derive estimates of the uncertainty of an individual model. The method performs multiple random perturbations of the starting model followed by tomographic inversion. In this exercise the starting model is a final velocity model built during a seismic processing project. A resulting solution model population is then selected based on analysis of the volumetric residual move-out; all models exhibit a similar level of residual move-out as the starting model. This set of realizations is then used to derive a model population variance attribute. It is also used for error envelope analysis at key targets, giving an indication of the spatial reliability of the seismic data.

We then introduce two forms of automation associated with the Monte Carlo simulation. Both reduce project turnaround. The first is a feed-back loop to constrain the size of the model population used in the simulation. This mechanization defines a threshold of what is an appropriate sample set size, and can reduce the turnaround of the model uncertainty project by understanding where extra data provides no extra information. The second example is enabled through a modification to the workflow, what nowadays is referred to as a ‘hack’. Rather than look at the uncertainty of a single model and image project, we 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 results show an effective solution, achieved with more than an order of magnitude saving in turnaround.

## Method

### *Model uncertainty workflows*

Most velocity model building practices use some variant of an inversion scheme, whether velocity tomography, which is based on observed data recovered from an initial pre-stack depth migration, or Full Waveform Inversion (FWI), which attempts to minimize the residual of a modelled data set and an acquired one. The tomographic approach comprises the following three steps, illustrated in Figure 1:

<sup>1</sup> PGS

\* Corresponding author, E-mail: tony.martin@pgs.com

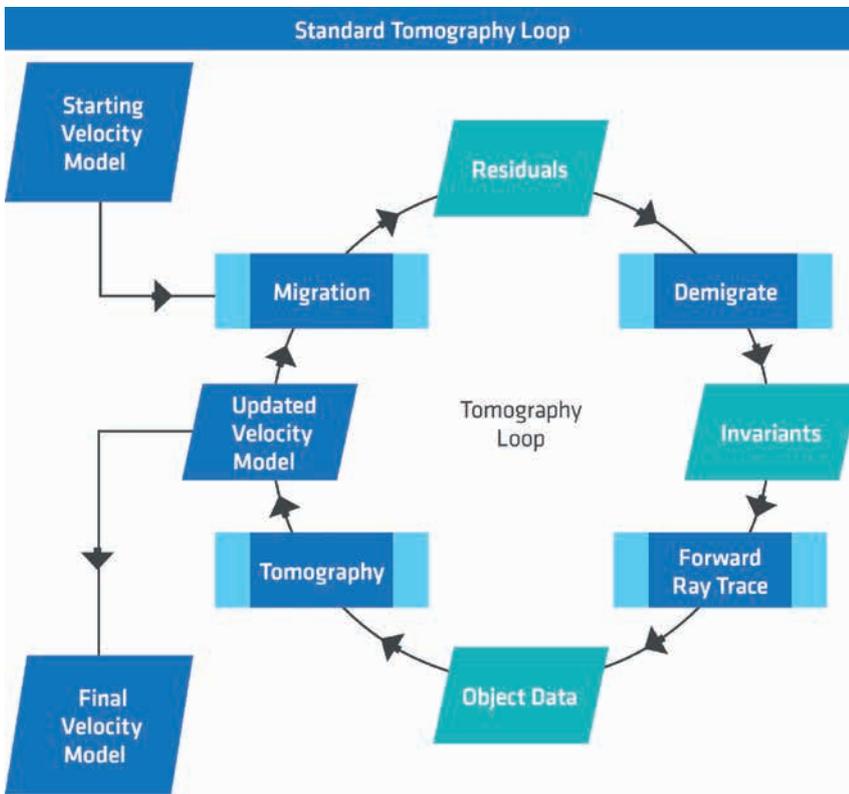


Figure 1 Schematic showing a standard tomographic loop.

1. Pick the residuals, which will then be used in the subsequent tomographic inversion to establish the misfit cost function. These are usually some form of residual depth error with respect to offset or angle measured in Common Image Gatherers (CIGs) generated by the initial migration.
2. Ray-traced demigration of these residuals is then performed using the migration model, creating observed data which are independent of a particular migration model. This ‘invariant’ data may be subsequently remigrated by ray-tracing performed in an initial inversion to create ‘observable’ data for tomography.
3. The final phase performs the linear inversion to update the model parameters by minimizing the cost function based on the observed data.

In step 2 the demigration and remigration removes the dependency of the initial residual observation on the migration model. This permits the last two steps of the process to form a mechanism that allows for numerous linear inversions. Each successive inversion recovers greater magnitude updates which overall can diverge from a linear update trend (Guillaume et al., 2008).

In the model uncertainty workflow, we use a Monte Carlo approach to generate a population of equiprobable outcomes prior to statistically understanding the behaviour of the products (Bell et al., 2016). In the case of velocity model building, we build a population of models that all equally explain the data, then we derive statistics that tell us about the variability of the models, prior to extracting information about the confidence one can have in the image associated with the traditional single model approach.

The method uses randomization to build the model population. However, this must be based on an understanding of the

variables that may affect the outcome of a tomographic inversion. These can be defined by two broad groups: how accurate are the measures or residuals we use to constrain the inversion, and how the tomographic inversion is parameterized? The first is dependent on the data, while the second is defined by how we want to sample the subsurface for the inversion.

To understand the impact of the variables for any given data set, we apply a perturbation to the model and compute the observed data from the invariants. The inversion is then run and the updated model recovered. Subsequent analysis is then performed to judge how well the inversion has recovered the reference model, given the data constraints implied by the residual observations.

We use this procedure to derive a simple model sensitivity check. We apply a checkerboard perturbation ( $P$ ) to the reference model ( $M_{mig}$ ) yielding model ( $M_o$ ), perform a subsequent inversion and then analyse the difference ( $dP$ ) between the initial and the final inversion ( $M_{inv}$ ) model to identify how well the perturbation has been recovered. This is summarized below:

$$M_{mig} + P = M_o \rightarrow invert \rightarrow M_{inv} - M_{mig} = dP$$

The closer  $dP$  is to zero the better the tomography has succeeded in recovering the reference model. We can use  $dP$  as an indication of the ability of the data to constrain the model. By varying both the magnitude and wavenumber of the perturbations we can determine the number of tomographic iterations to perform and the smoothing parameters to consider within a particular update.

In Figure 2 we see a cartoon representation of this approach. We observe that the inversion has resolved the wavelength and magnitude of the perturbation, which in turn provides information

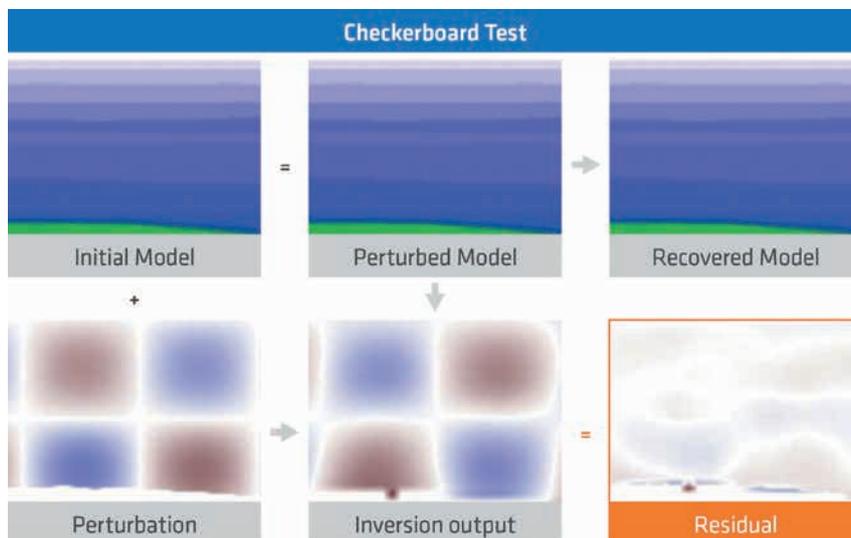
on the mechanism that will create our randomly perturbed equiprobable model population. In practice we use a sequence of automatically derived quantitative metrics that define thresholds for the suitability of the perturbations applied. This is done prior to their use in the creation of a randomly generated model set. Our first step fulfils our requirement to understand the sources of uncertainty prior to a full Monte Carlo simulation.

Once the useable minimum spatial wavenumbers and maximum amplitude thresholds are established, the information is used to generate measures of the statistical reliability of the starting model. The underlying assumption is that all solution models in the population fit the data. Firstly, we generate a population of perturbed models that we apply to the starting model. After subsequent inversions we recover the solution models (see Figure 3). Secondly, migrations are performed for all the solution models and volumetric move-out attributes are generated from the resulting CIGs. We classify the realizations based on this residual move-out error, by using cumulative move-out functions

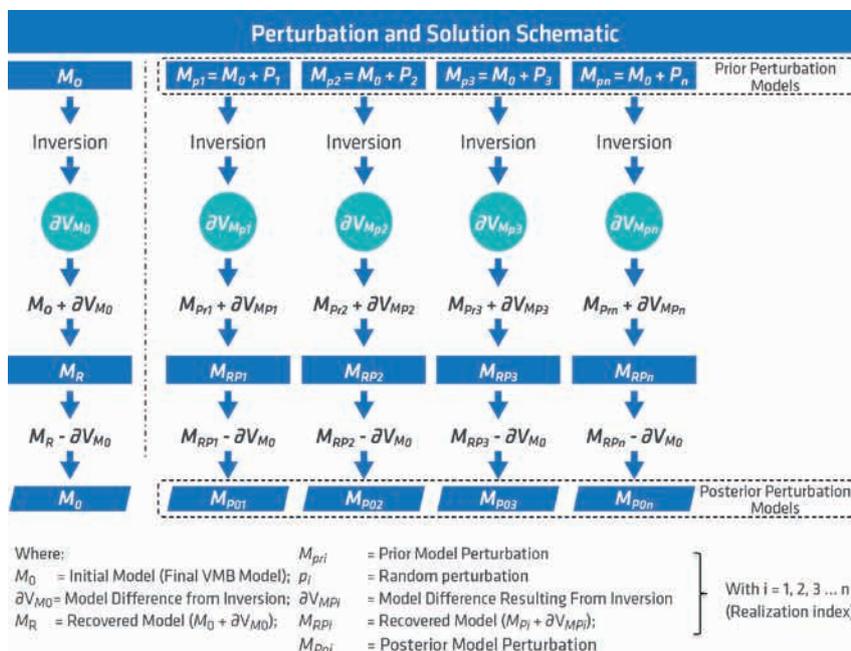
to select the subset of solution models for further statistical analysis. Given the earlier analysis that optimizes the model population, the proportion of excluded models is small.

The tomographic inversion platform uses a beam migration to establish the initial ray kinematics of the invariant data, which comprise wavelets extracted from the data through a multi-dimensional dip scanning process (Sherwood et al., 2008). This is performed within the migration model space generating the observed data. The process of model perturbation is performed in a residual migration. This applies the differential kinematic to the observed data consistent with the applied perturbation. The use of this approach is outlined in Figure 4, and enables the resolution of the model population in an accelerated timeframe, precluding the need to wait months for the results.

Once the acceptable solution model population is defined, we generate statistical attributes to constrain the uncertainty associated with the target model. The mean, variance and standard deviation parameters are computed from the population for each parameter



**Figure 2** Cartoon describing the checkerboard perturbation process used to determine how the data supports the wavelength and resolution tests.



**Figure 3** Schematic flow showing the perturbation and inversion process defining the simulation.

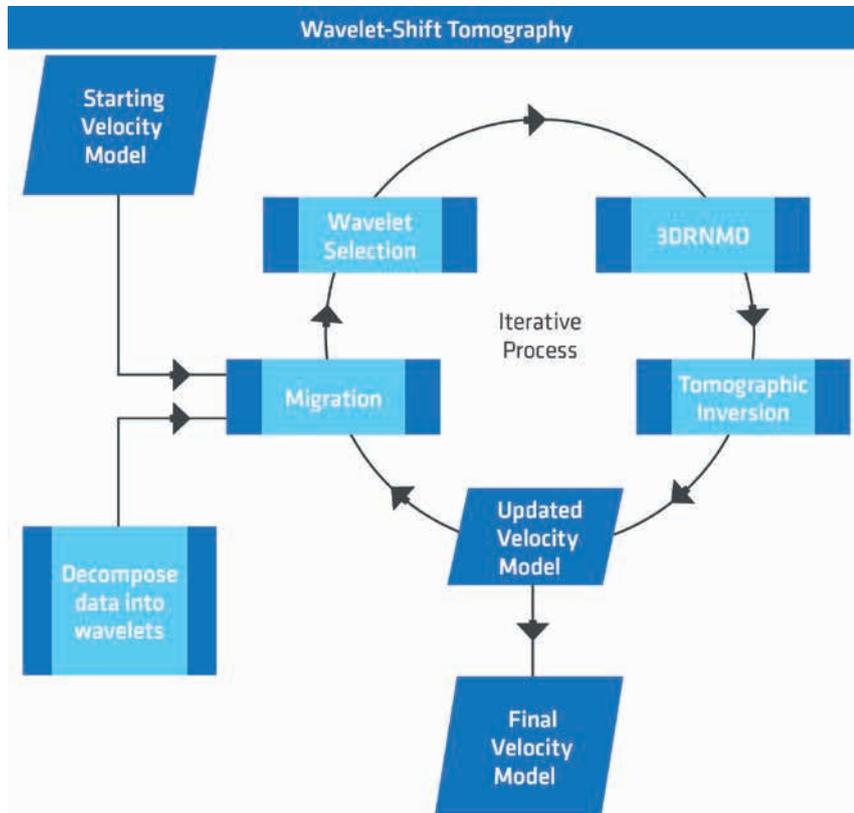
cell. The population of model realizations are used to explain positioning errors associated with a particular event. We create a Gaussian envelope mask based on the wavelength of the reflection coda about a specific event of interest. This mask is then applied to the migrated stacks generated from the model population. Using a sequence of cross-correlations we construct a mean vertical position, along with error envelopes defined by the standard deviation of the lags, modified to accommodate local dip deviations.

*Model uncertainty – field example*

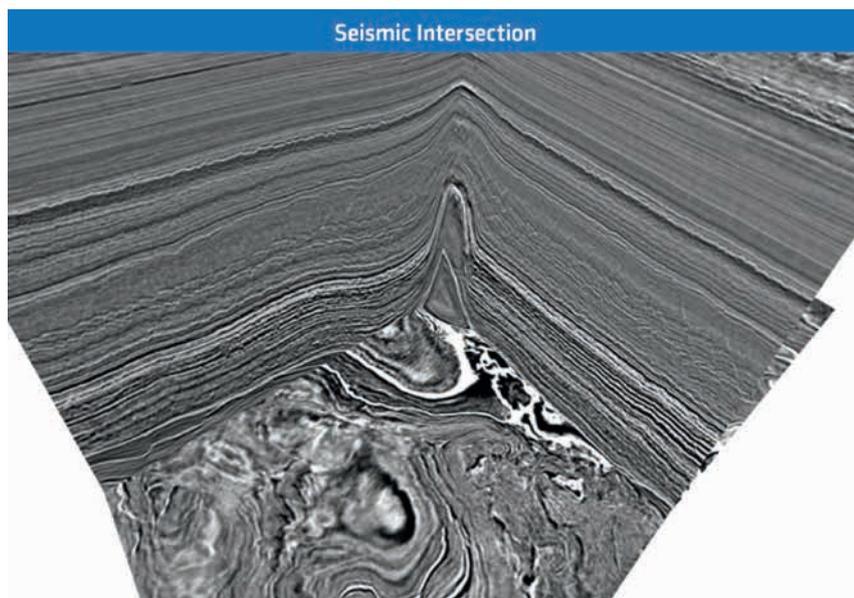
A field example is demonstrated in Figure 5. The data is from the Central Graben in the North Sea. The intersection of data

in x, y and z shows a complex geological setting affected by a salt diapir. The overburden contains a sequence of localized fault blocks set in a shallow water environment. Deeper thick chalk is pierced by salt, which has then receded. Pre-chalk data contains structurally complex and poorly illuminated reflectivity.

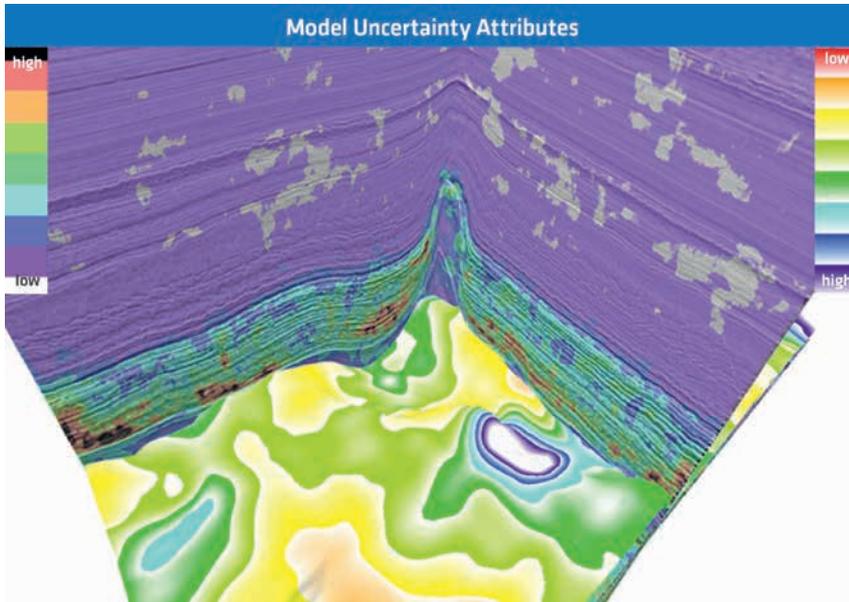
Figure 6 shows the same data with model uncertainty attributes overlain and co-rendered. The variance of the 120 models used for the population is co-rendered on the inline and crossline sections, and shows that the majority of variation in the model population is within the chalk interval, especially in small gentle synclinal structures in the lee of the salt intrusions.



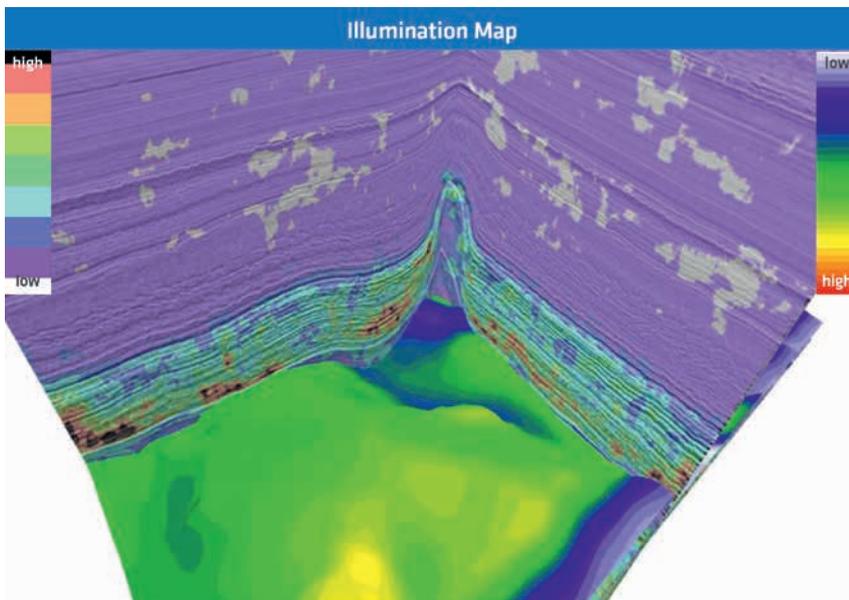
**Figure 4** Schematic showing the hyperTomo engine used for the Monte Carlo simulation.



**Figure 5** Inline, crossline and depth slice intersection showing the CGR data set's subsurface complexity.



**Figure 6** CGR data with model uncertainty attributes. Inline and crossline sections show the co-rendered model population variance, which shows biggest variations in the chalk. Depth deviation is shown for the Base Cretaceous Unconformity, whose largest errors are correlated with the areas of biggest model variance.



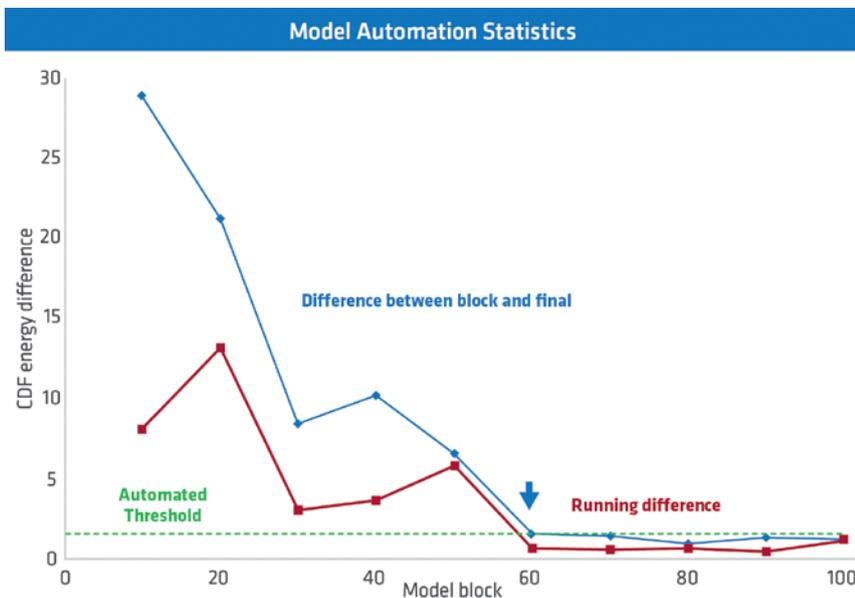
**Figure 7** CGR data with illumination shown for the Base Cretaceous Unconformity. The zone of blue (poor illumination) correlates with the depth deviation in Figure 6.

Thick chalk is challenging for tomography, as the extraction of residual measurements from the data are controlled by reflectivity and data diversity; both of which are limited in chalk.

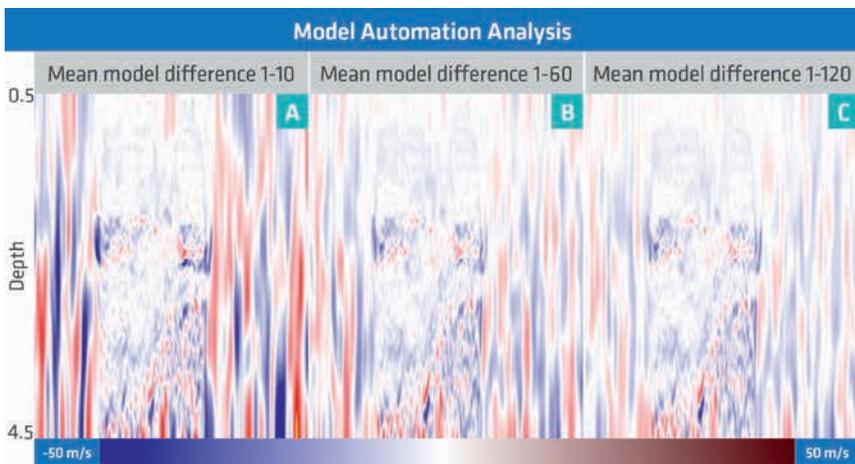
The depth deviation map for the Base Cretaceous Unconformity (BCU), which sits in the pre-chalk section, is also shown in Figure 6. There is a strong correlation of depth deviation on the BCU directly below the areas of high variance in the model population. An illumination map (Figure 7) for the BCU also illustrates that in the lee of the receded salt there is an illumination ‘hole’. We can build a story about where the single model and image uncertainty exists in this data, based on the statistical attributes that come from the Monte Carlo simulation. Poor illumination and challenging geology lead to localized areas of poorly constrained tomography. This is defined by large variance in the model population. Confidence in the structural positioning in the image may be weaker beneath thick chalk synclines, where salt diapirs pierce the overburden.

#### *Automation one – reducing turnaround in model uncertainty*

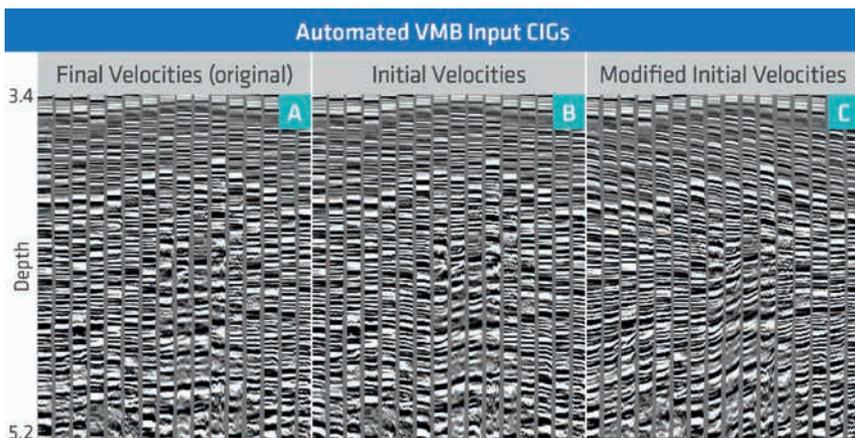
In the example shown in Figures 5, 6 and 7, 120 models were used to determine the model metrics (variance) and the spatial reliability indicators (depth deviation). As previously described, these models are created randomly, and are constrained by an understanding of how the data supports the model. The analysis is derived from the migrated data that result from the 120 models. The total number of volumes created through the process to create the metrics is in excess of 2000 volumes, half of which are pre-stack. Resolving hundreds of models tomographically prior to imaging, high-grading the results, and performing statistical analysis is a time consuming process, even when using a ‘super-fast’ inversion and migration scheme, and dedicated computing resources. It may not be necessary. Understanding when, for a given data set, we determine an appropriate model sample set that accurately describes an ideal model population, could save time by reducing the number of models needed to generate meaningful statistics.



**Figure 8** Statistical analysis of Cumulative Distribution Functions (CDFs) for both running differences (red) and differences with the initial model (blue). Both converge to show no change in relative energy after 60 models are used (blue arrow).



**Figure 9** a) Mean model of the models 1-10, b) Mean model of the models 1-60, c) Mean model from the models 1-120. Little difference can be seen in b and c.



**Figure 10** Common Image Gatherers: a) Final tomographic model (project); b) Initial model (for a); c) modified and locally erroneous initial model.

Using a statistical feedback loop of the variance in the running model population mean and deviation, we can determine, in real-time, when both of these statistics start to stabilize globally. Running analysis of Cumulative Distribution Functions (CDFs) of blocks of models show, for this example, that progressive addition of models slowly decreases the relative energy in the CDFs, which in turn adds little to the statistical relevance of the outcome. Figure 8 shows a graphical representation of the running statistical analysis, while Figure 9 shows the mean of the first group of ten

models, and then the group containing 60 models and 120 models. Both figures show that at 60 models the amount of meaningful information achieved by adding more data does not necessarily result in more relevant statistics. In this example, 60 or 70 models would appear to be appropriate, based on the method used to create the model population. The total energy in the CDF representation does not increase, and the mean model does not change. The implementation of this automation through a feedback loop could have saved time for both the workflows and outcomes of the simulation.

### Automation two – reducing turnaround in velocity model building

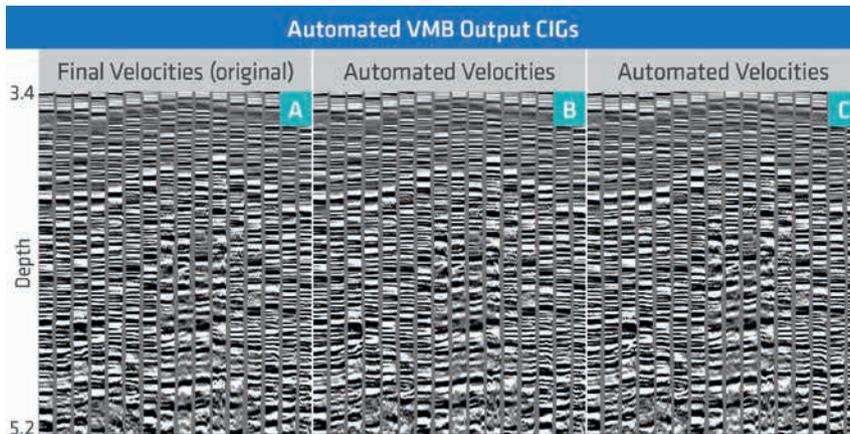
Determining design features from one process to apply in another way for a different process is the cornerstone of reverse engineering. When generating analytical information from the model uncertainty flow, we are attempting to provide confidence level metrics of a velocity model. We have previously referred to this as a final velocity model, or the starting model for the uncertainty analysis. Equally, it could also be used to check the reliability of a velocity model building exercise, if it is being built in a layered approach. However, if we modify the workflows using some simple changes, can we generate a usable velocity model using the same underlying Monte Carlo simulation criteria, rather than provide statics on a final one? How close does the starting model need to be to the ideal final one? Are there implications if the model is significantly wrong? Would this type of automation without testing, QC and manual intervention help to reduce turnaround? Would the result be equivalent to a ‘normal’ velocity model building exercise?

The starting point for the full automation of velocity model building begins with the same steps of determining what the data supports in the model space prior to creating a randomly generated model population. Once generated, the population is tomographically inverted. Rather than trying to understand whether we remove the perturbation applied to the initial model, we simply perform a statistical analysis on the resulting 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 CIGs which have

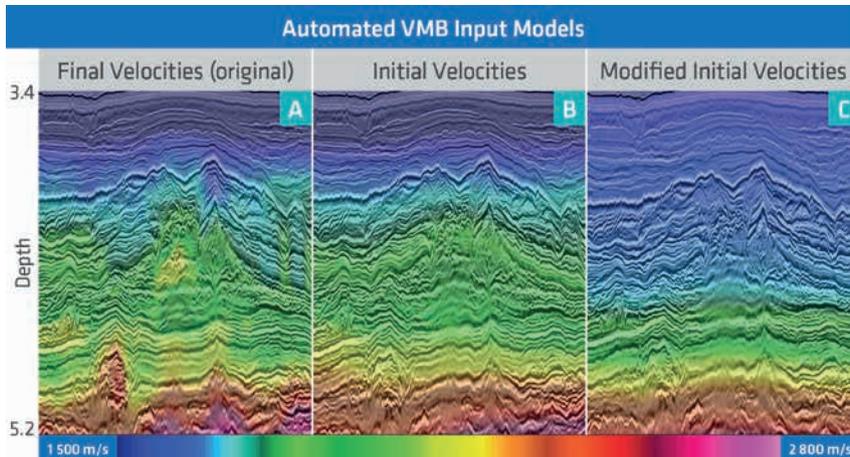
a zero residual for tomographic inversion. This is quantified by determining move-out related metrics after each pass of the simulation. We map the progressive convergence of the solution using these metrics.

The results from the second automation process were run on a 500 km<sup>2</sup> data set from West Africa. The test was performed to reduce the time taken to produce a model by removing any 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 of 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.

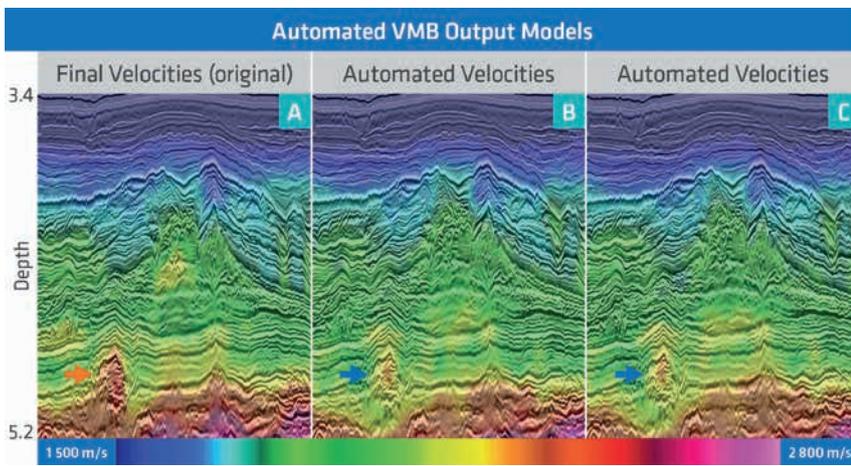
Figure 10 shows three sets of CIGs, where Figure 10b is the initial model used for the full tomographic model building, and Figure 10c is the modified version, which is locally 10% different to Figure 10b. Figure 10a is the result of the 90 day model building exercise. Figure 11a is the same as 10a, while Figures 11b and 11c show the results from the automated Monte Carlo model building process for starting points 10b and 10c respectively. CIGs in Figure 11a, 11b and 11c all have a similar level of move-out. Progressive analysis of metrics on move-out show an equivalent level of convergence in the resulting models irrespective of the starting point (Figure 14). Figures 12 and 13 show the same sequence with velocity models co-rendered on the vertical seismic stack section. The models that result from the



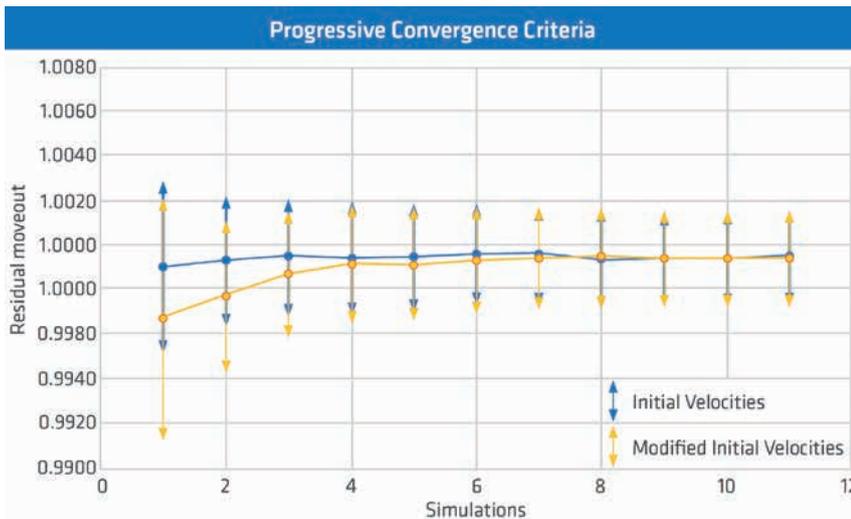
**Figure 11** Common Image Gathers; a) Final tomographic model (project); b) Final automated model starting with 10b; c) Final automated model starting with 10c.



**Figure 12** Migrated stacks with co-rendered velocities; a) Final tomographic model (project); b) Initial model (for a); c) modified and locally erroneous initial model.



**Figure 13** Migrated stacks with co-rendered velocities; a) Final tomographic model (project); b) Final automated model starting with 12b; c) Final automated model starting with 12c. Orange arrow (12a) shows the location of the geobody masked and update channel. Blue arrow shows the channels capture with the automated approach.



**Figure 14** Move-out convergence criteria quality control. Blue curve using starting model shown in Figures 10b and 12b, Orange curve using starting model shown in Figures 10c and 12c. Both blue and orange curves converge to the same level of move-out.

automated model build (Figures 13b and 13c) are very similar to each other, and closely resemble the model built in 90 days (Figure 12a and 13a), which used a hybrid-layered approach with local high velocity channels benefiting from geobody interpretation and masked updates (orange arrow, Figure 13a). During the model building process, this was deemed the only way to focus the update. However, the same channel is partially updated using the automated approach (blue arrows, Figures 13b and 13c).

The automated approaches to building a velocity model using a Monte Carlo simulation of the model space appear to converge on solutions of near equal quality. In this field example, the starting point was up to 15% in error, but still resolved. The workflows were initiated by a geophysicist who had no prior knowledge of the data or models. 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. Additional nuanced automation steps have been devised to further reduce the turnaround, the aim being to create an automated model in days rather than months.

### Conclusions

Data sets in seismic acquisition and processing are getting larger. Expectations are growing for confidence levels on seismic

deliverables, and project turnaround is assumed to be decreasing. We have introduced a method that provides both data reliability metrics and automation for depth imaging velocity model building. Using a randomly generated model population, created after understanding the constraints on the model by the data, we derive statistics that can be used to mitigate risk by providing ‘error bars’ on the spatial reliability of a typical single model and image project.

Modifications to the same workflow enable the creation of an automated depth imaging velocity model building scheme. Depending on the data, the method has shown that the starting model may not be critical, and turnaround can be reduced by greater than an order of magnitude over traditional methods, while producing near-equivalent results. Additional automation steps have been identified within the workflows which, when used in conjunction with the products of continuing work in learning algorithms, will enable further reductions in turnaround for this critical step in a processing project. Automation in seismic processing is looming on the horizon, but is the industry ready to embrace change?

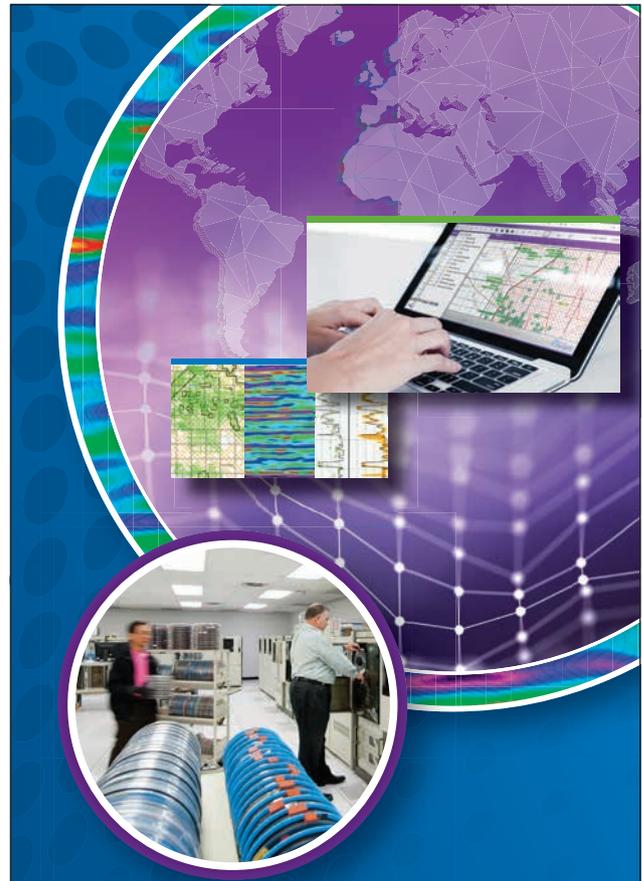
### Acknowledgements

The authors wish to thank PGS for authorization to show examples from the different surveys acquired by PGS. We would like to give special thanks to Tony Bell, Lorenzo Russo and John Wil-

lis for their significant contribution to this work. We also thank Andrew Long and Nizar Chemingui for fruitful discussions, and Aoife O'Mongain and Vinnie Papenfus for their contributions.

## References

- 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.
- Guillaume, P., Lambaré, G., Leblanc, O., Mitouard, P., Le Moigne, J., Montel, J. P., Prescott, T., Siliqi, R., Vidal, N., Zhang, X. and Zimine, S. [2008]. Kinematic invariants: an efficient and flexible approach for velocity model building. *78th SEG Annual International Meeting, Expanded Abstracts*.
- Letki, L.P., Ben-Hadj Ali, H. and Desegaulx, P. [2013]. Quantifying uncertainty in final seismic depth image using structural uncertainty analysis – Case study offshore Nigeria. *75th EAGE Conference & Exhibition, Extended Abstracts*, Tu 07 14.
- Osyopov, K., O'Briain, M., Whitfield, P., Nichols, D., Douillard, A., Sexton, P. and Jousselin, P. [2011]. Quantifying Structural Uncertainty in Anisotropic Model Building and Depth Imaging - Hild Case Study. *73rd EAGE Conference & Exhibition, Extended Abstracts*, F010.
- Sherwood, J.W.C., Sherwood, K., Tieman, H. and Schleicher, K. [2008]. 3D beam prestack depth migration with examples from around the world. *78th SEG Annual International Meeting, Expanded Abstracts*, 438-442.



Digitise The World<sup>SM</sup>  
at EAGE #1060

## Subsurface Data Management

Access your seismic, well data and interpretation volumes for analysis and analytics. Katalyst manages the complete subsurface data life cycle for over 40 petabytes of data.

Contact Katalyst Data Management:

Angus Craig | 44 (0) 7770 872739  
David Norburn | 44 (0) 7946 387139  
sales@katalystdm.com

[katalystdm.com](http://katalystdm.com)

