

Automated interpretation of 3D seismic data using genetic algorithms



James K. Dirstein

James K. Dirstein^{1,3} and Gary N. Fallon²

¹Total Depth Pty Ltd, Perth, WA (www.totaldepth.com.au)

²Geophysical Resources & Services Pty Ltd, Newcastle, NSW

³Email: jim@td.iinet.net.au

This paper was presented at the Bowen Basin Symposium in October 2010. We thank the GSA Coal Geology Group for allowing this paper to be published in Preview.

Over the past twenty-five years geoscientists have acquired more than 550 000 square kilometres of 3D seismic data (APPEA statistics) over continental and offshore Australia in the pursuit of mineral and petroleum deposits. Whether the target is hydrocarbons of any phase (solid, liquid or gas) or minerals, the information extracted from the 3D seismic data when integrated with other geological and geophysical data helps form models of the subsurface. These models are the foundation upon which decisions are made, directing future exploration, appraisal and development activities. The success of these activities often depends upon the accuracy of these models.

Many advances in acquisition, processing and interpretation methods have been implemented since the first 3D seismic surveys were acquired in Australia during the 1980s. As a consequence of these advances, the geoscientist today is faced with dramatic increases in the volumes of high quality data available for analysis. However, the time available for thorough examination, analysis, extraction and integration of the information from these large, often multi-volume, datasets is always limited and is becoming more problematic. Typically, the geoscientist will spend most of their available time extracting information from small portions of these datasets with a disproportionate amount of time spent thinking about the significance of the results.

Fortunately, geoscientists are not the only, or the first scientists, to face challenges associated with the analysis of large amounts of data. Specifically, ideas developed during the course of the thirteen year Human Genome Project (HGP) have been adapted to help interpret seismic data by

automatically segmenting and identifying all surfaces within a 3D volume of data. The results are then stored in a visual database. Using this technology enables the geoscientist to analyse large amounts of data in an unbiased manner and thereby incorporate much more data into their models. The details of this patented technology are discussed and demonstrated on several examples.

Introduction

The use of 3D seismic data is fairly common practice for the evaluation of both coal and hydrocarbon exploration. While considerable effort and budget dollars are spent on the planning, collection, processing and interpretation of this data, the majority of seismic, in most cases, is underutilised. Table 1 shows how the typical interpreter might spend his time working with a 3D seismic volume. For 3D seismic volumes specifically acquired in coal operations the geological interpretation component is a little larger as the effort goes into detailing the characteristics of faults identified in the volume.

Table 1. Breakdown of 3D seismic interpretation time

Activity	Total project time (%)
Analysing data	10
Picking horizons	60
Creating geological interpretation	20
Significance of results	10

Table 1 suggests that the effort required to identify and map individual surfaces within each seismic volume is quite time consuming and limits the amount of data examined within any 3D seismic dataset. By automating the most time consuming element of the process and looking at all the data in an unbiased manner, more time should be available to develop a better understanding of the significance of the results. Given that many auto-tracking algorithms available in commercial workstations struggle to yield high quality surfaces for single horizons without constant corrections in erroneous event tracking, how will the automatic and simultaneous analysis of all surfaces provide a better solution?

Origins

The new technology outlined in this paper finds its inspiration, effectiveness and perhaps future refinements from the Human Genome Project (HGP). Therefore it is appropriate to begin with some background information about the HGP. The HGP was a very ambitious task undertaken by biologists and was perceived to be the last effort needed to conclude work in a field founded by Watson and Crick (1953) with their publication of the double helix model for deoxyribonucleic acid (DNA). Some of the objectives of the HGP were as follows:

- To identify all of the genes in the human DNA (initial expectation of as many as 150 000 genes),
- To determine the sequences of the 3 billion chemical base pairs,

- Store this information in databases and improve tools for analysis,
- Transfer related technologies to the private sector, and
- Address the ethical, legal and social issues (ELSI) that may arise from the project.

Thoroughly conducted scientific investigations generally yield data and insights that are beyond original expectations, and the HGP project was no exception (Baltimore, 2001). A summary of the HGP big surprises were:

- About 24 000 genes were identified in the Human Genome (still not 100% sure),
- Genes are much more complicated than originally imagined (before the HGP a two hour undergraduate lecture was adequate to describe the gene model; after the HGP three months of lectures are required to explain the concept of a Gene (Pearson, 2006)),
- The initial model developed pre-project used only the DNA (ignoring 50% of the mass which was comprised of the encasing protein),
- The ‘ignored’ protein plays a vital role in the Human Genome (now a new field called Epigenetics),
- The Human Genome has only 2000 more genes than the simple nematode,
- Cautionary insight: an accurate model could not be created using 50% of the data!

Given geoscientists in both the petroleum and coal industry use less than 10% of the surfaces available in 3D seismic data volumes in their analysis, it would be reasonable to assume that additional insights into the subsurface would be possible if all the data could be examined in a timely, accurate and cost effective manner.

Methodology

The analysis system, inspired by the HGP, uses an entirely evolutionary process in the form of genetic algorithms to segment the seismic data. Genetic algorithms are mathematical processes (Grefenstette and Baker, 1989; Michalewicz *et al.*, 1992) that mimic the genetic process of biological evolution. The evaluation of a possible solution depends on the predetermined parameters associated with the “goodness of fit” criteria. The better the fit, the greater the chance of the solution surviving until the next generation of evaluations. Fang *et al.* (1996) and others have demonstrated the effective use of genetic algorithms in geoscience. By applying this approach, the analysis identifies unique waveform segments that relate to surfaces or horizons

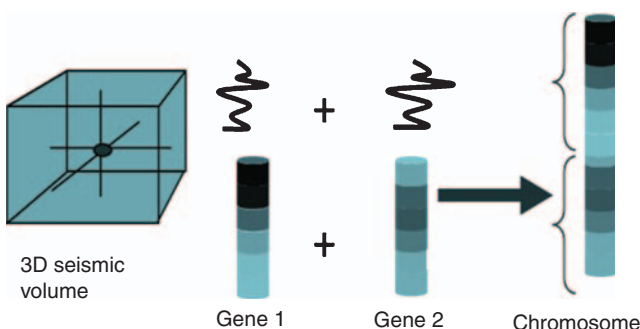


Fig. 1. Analogy between seismic and the chromosome. Image from Seisnetics LLC, unpublished material.

and are referred to as GeoPopulations™. These are automatically extracted quickly, accurately and in an unbiased manner.

To determine the extent of GeoPopulations™ these evolutionary algorithms apply the principles of natural selection and “survival of the fittest” to grow from disordered and random seed points to groups of genetically related individuals. A wide range of genetic algorithms have been used and proven to be both powerful and effective for a wide variety of optimization problems, such as medical, airline scheduling, stock market trading, adaptive control, military, and so forth.

The genetic analogy with the seismic volume (Figure 1) can be described as follows:

- A chromosome is analogous to a seismic trace.
- The seismic volume therefore, consists of many chromosomes.
- Each chromosome is made up of a group of genes just as each seismic trace consists of a group of waveforms. Therefore, seismic waveforms are considered equivalent to genes.
- Each gene (**waveform**) can be characterized by its own unique suite of attributes (i.e. location, amplitude value, neighbour trace shape, etc.).

Initially, the seismic volume is automatically segmented into a population of individual waveforms (Figure 2). Individuals within this collection of waveforms are randomly selected as new populations. This gene then looks both locally and globally for other genes with the most similar genetic characteristics (amplitude values, trace shapes, frequency or any combination of attributes that are associated to each sample).

As the populations grow (evolve), the common waveform or genotype changes as selection and reproduction continue according to criteria based on both local and global parameters. As the groups of waveforms grow they will eventually encounter other groups. If they are compatible both spatially and genetically they combine forming a new, larger subpopulation (offspring) that inherits the genotype (common waveform) of its two parents. The evolution continues throughout the entire volume until all GeoPopulations™ have been identified and categorised into a database of surfaces.

Like any evolutionary process, some elements evolve faster and others fail to evolve at all (see Figure 3). At the end of processing, each 3D seismic volume has hundreds, of identified GeoPopulations™. This database of surfaces needs to be reviewed, sorted and filtered. Based on the current requirements of the interpreter, a selection of these surfaces will be extracted

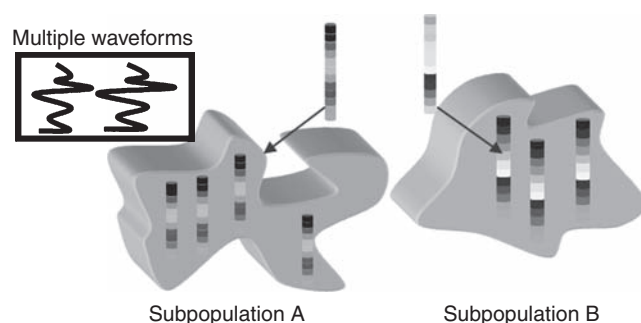


Fig. 2. The evolution of a GeoPopulation™. Image from Seisnetics LLC, unpublished material.

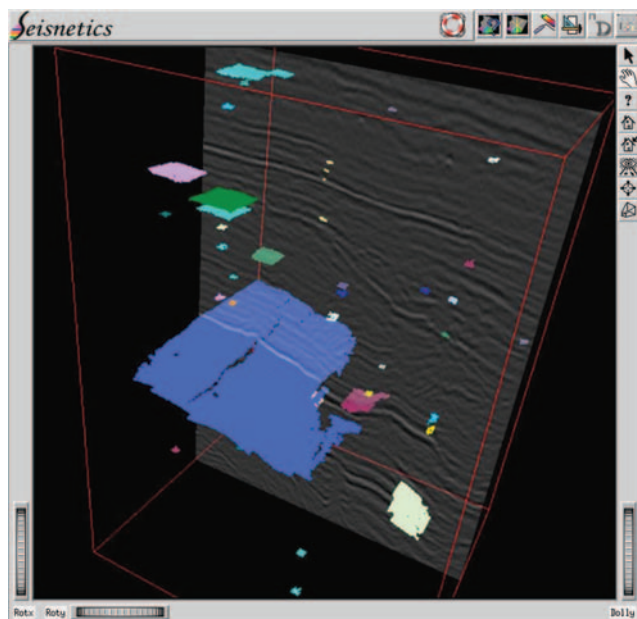


Fig. 3. Evolution of GeoPopulations™ (snapshot in time). Some populations evolve faster than others (blue horizon). Image from Seisnetics LLC.

for further analysis. The most effective means of reviewing all the results is by way of a visual database which enables subsets of the GeoPopulations™ to be reviewed and selected for extraction based on a number of statistical and visual criteria. For example, the interpreter might first select the largest surfaces to help develop an initial structural model (Figure 4). Later as objectives change, the visual database can be revisited and queried for other objectives such as a stratigraphic zone of interest, specific seam-roof-floor, shallow overburden assessment, or fault analysis.

The export of selected surfaces into an interpretation, GIS package or modelling software enables further analysis and leaves the integrity of the unbiased GeoPopulation™ database

intact. Within the visual database the identification of GeoPopulations™ which match specific criteria can be realised using a number of different filters and sorting techniques (population size, position, quality, etc.).

Each GeoPopulation™ has a set of attributes associated with each member of the population. One of these attributes is called “Fitness” which provides a measure of “genetic likeness” for each member in the population when compared to the common waveform (Genotype) of the same population. This fitness criteria shows individuals that might still be related but are best described as first or second cousins. The best way to assess the genetic variability within a population is to view the fitness values as a map. The fitness map shows areas of high fitness (green) with lower fitness values as blues and reds (Figure 5). Investigation of the lower fitness values which form linear and curvilinear features on the map are predominately caused by subtle faulting with some subtle stratigraphic elements as well. The waveform located in the lower left hand corner of Figure 5 is the common waveform for this GeoPopulation™ and is referred to as the Genotype (shown in red).

A 3D seismic volume is reported to contain many attributes (>150), however most of these attributes are derived from, and dependant on, other attributes: e.g. the gradient is from the Two-Way-Time (TWT) horizon. Consequently, some seismic attributes are useful while others will be redundant or useless and can confuse seismic interpretation more than they help (Barnes, 2007). Using attributes which have a greater degree of orthogonality (or independence) provides better discriminatory power and produces more reliable results. This all clearly assumes the seismic volume has been correctly processed in the first place to minimise artefacts and truly represent the signal and image characteristics at each reflective horizon. Other independent attributes identified for each individual in the GeoPopulation™ are Amplitude and TWT (Figure 6). High quality surfaces will result in more meaningful horizon amplitudes and TWT structure. TWT with Fitness and Amplitude enables a rapid assessment of the volume surfaces.

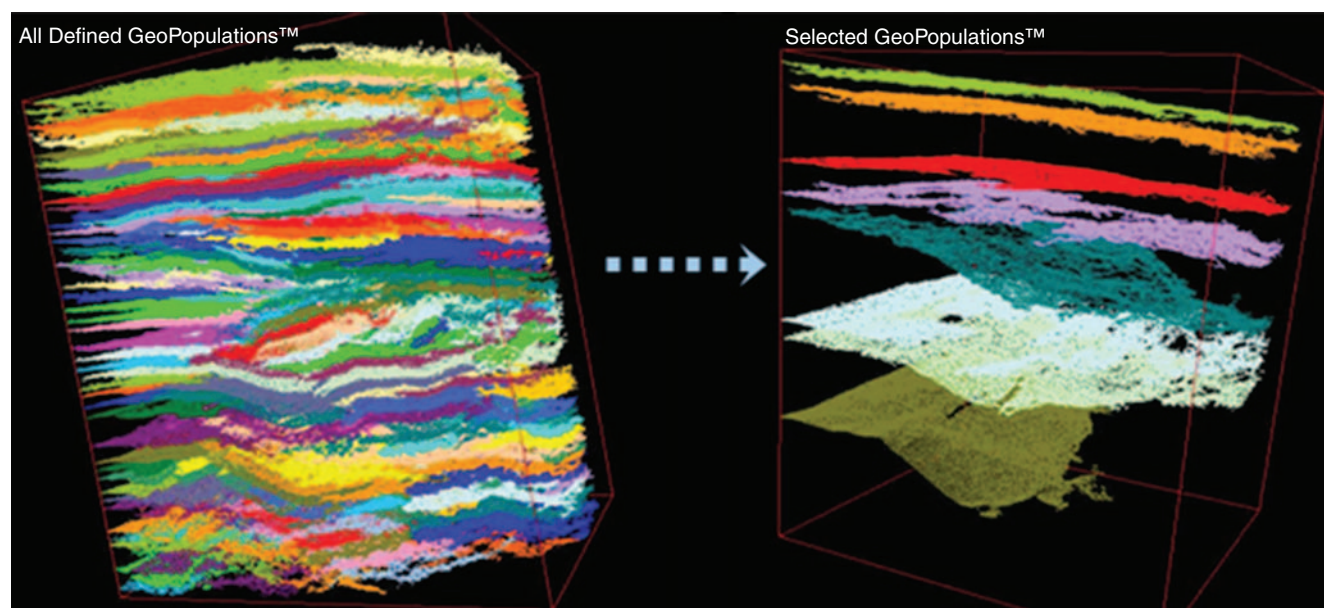


Fig. 4. Horizons which address the current task are identified and then exported into GIS, interpretation or modelling software. Image from unpublished Seisnetics presentation.

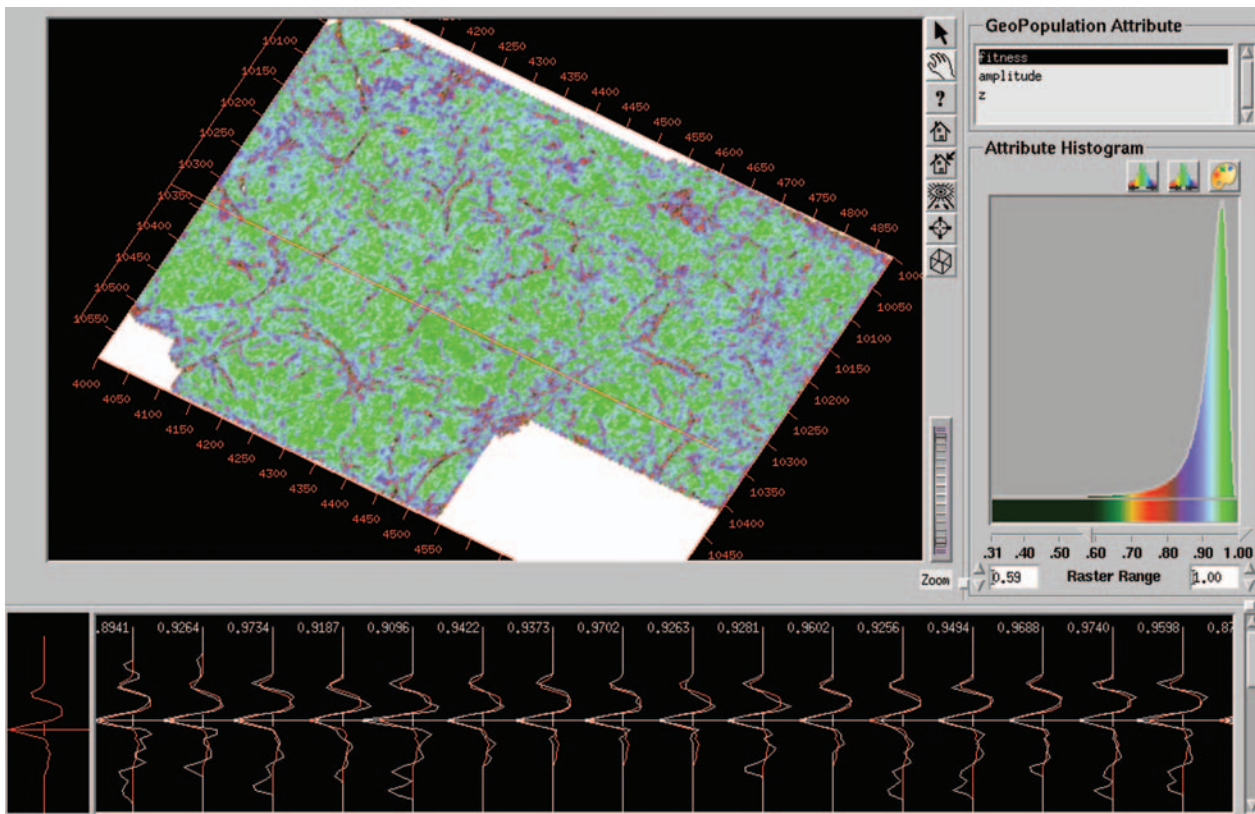


Fig. 5. The map shows the 'fitness' of a single GeoPopulation™. The common waveform or genotype is shown as the red trace in the lower left hand corner of the image. High fitness values are shown as green on the map meaning these traces have the highest degree of similarity with the genotype. The low fitness values shown are not necessarily an indication of a poor pick. In this case, the areas of lower fitness values identify subtle structural and stratigraphic features.

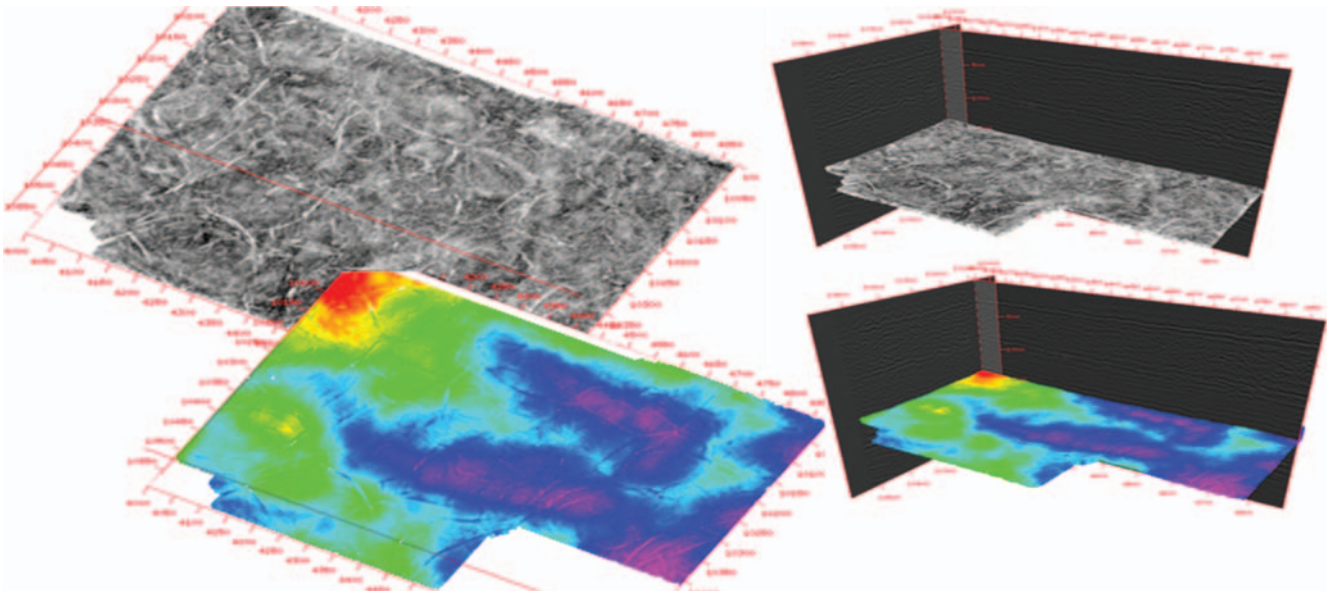


Fig. 6. Amplitude and TWT when combined with Fitness provide a rapid visual assessment of the GeoPopulation™.

Results

We have applied the processing algorithm described above to thousands of square kilometres of 3D seismic surveys over both onshore and offshore Australia. In addition to final processed TWT volumes, some other data types processed in this manner include but are not limited to:

- Time, Frequency or Depth Domain.
- Post-Stack (angle stacks, AVO and Inversion attributes, Reflectivity, most seismic attributes).
- Spectral attributes volumes such as Spectral Decomposition and Spectral Attenuation.
- Pre-Stack (gathers, shots for first break or refraction analysis).

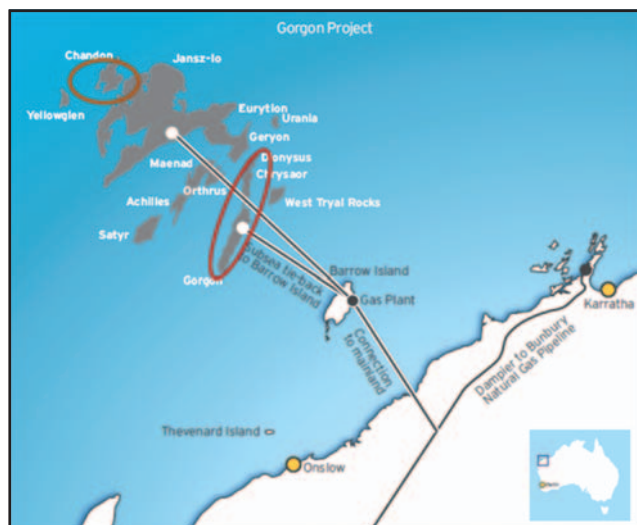


Fig. 7. Location map for Chandon and Gorgon gas fields. Extract from Chevron Publication 'The Power of Human Energy': <http://www.chevronaustralia.com/Libraries/Publications>.

Several examples are shown here from across Australia using datasets collected over the Gorgon Project area, the South Australia portion of the Cooper/Eromanga basin and data collected from a coal minesite in Queensland.

Example 1: the Chandon Gas field (Chevron Operation) reservoir level

The Chandon gas discovery was drilled in 2006 in 1,200 metres of water and is located in the North Carnarvon Basin (Figure 7). This field is one of a number of large gas accumulations which form the Gorgon project area. Chevron reports that these fields contain approximately 40 Trillion cubic feet (TCF) of natural gas and Chevron cites this as Australia's largest natural gas resource (Chevron Australia, 2010). The Seisnetics genetic algorithm was applied to a sub-volume extract from the Chandon 3D seismic volume. The subset processed consists of just under 500,000 traces or about two gigabytes of data. After approximately eight hours of processing more than 700,000 generations of "evolution" identified about 120 million individuals which were assigned to GeoPopulations™. Figure 8 shows a surface near the top of the reservoir section contained in tilted fault blocks. This surface is one of hundreds of high quality surfaces automatically extracted by the Seisnetics processing algorithm. The interpreter then reviews the resulting surfaces to decide which of those surfaces provide the most meaningful geological insight.

Example 2: the Gorgon Gas field (Chevron operation) – outgassing and geohazard

The Gorgon field was discovered in 1981 and is located in the southeast corner of the Gorgon project area. The production lifespan of the project may approach 60 years. In this example, the entire sixteen gigabyte dataset was processed for GeoPopulations™ of both peaks and troughs. As with the Chandon example hundreds of high quality surfaces were automatically extracted after several days of processing. Figure 9 shows an extract from the surface associated with the sea floor. The round circular patterns are pockmarks which are geomorphologic features which are often indicative of upward

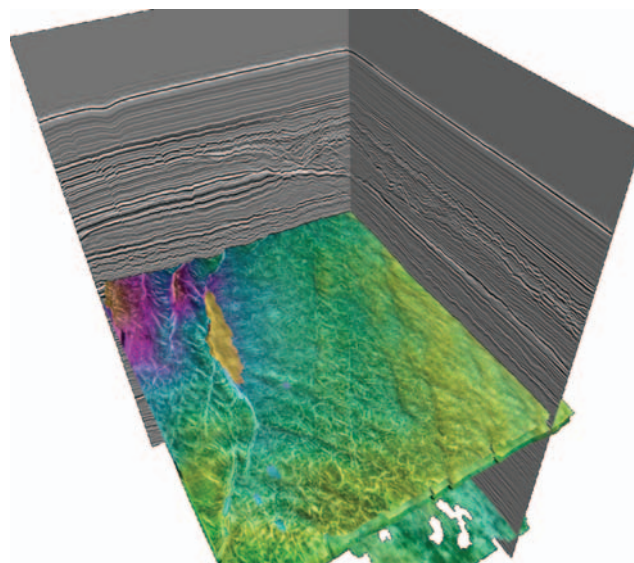


Fig. 8. Image taken from visual database of GeoPopulations™ from the Chandon 3D volume. The GeoPopulation™ shown is located just above the top of the reservoir section.

fluid flow and the venting of gas. The Gorgon field, along with many other gas accumulations offshore from Western Australia, show evidence of outgassing and upward fluid flow. This outgassing has, in places, resulted in both small scale and large scale depressions in the sea floor. While some pockmarks can be small and below the imaging resolution of conventional exploration 3D surveys, many (like those shown here) are much larger and can measure hundreds of metres in diameter. Regionally, areas of higher density of pockmarks have contributed to slumping and sea-floor instability over large areas during the course of geological time.

Example 3: the Cooper/Eromanga basin (South Australia and Queensland)

The Cooper and Eromanga basins, which span northeast South Australia and southwest Queensland, form Australia's largest onshore petroleum province (Fig. 10). Currently, more than sixty 3D seismic data volumes comprising of about 13 000 square kilometres of seismic data have been processed from this area using the automated genetic algorithm from Seisnetics. The initial phase of the project which processed forty volumes was completed within four calendar months. The integration into GIS applications of these high quality GeoPopulations™ with open-file well control, production data, and zones of interest enables both regional and very detailed models to be developed (using much more of the available data). Moreover, these models are entirely data driven and can provide an effective means of extending constrained models into areas which have less data coverage. This type of integration is underway offshore and onshore in every State in the Commonwealth as all seismic data collected eventually becomes open-file.

Example 4: coal mine from the Bowen Basin

This example shows 3D data collected over a Bowen Basin coal mine. The actual location and orientation of the data is withheld at the request of the company operating the mine.

The Bowen Basin in central Queensland is subject to a significant amount of open cut and underground coal extraction.

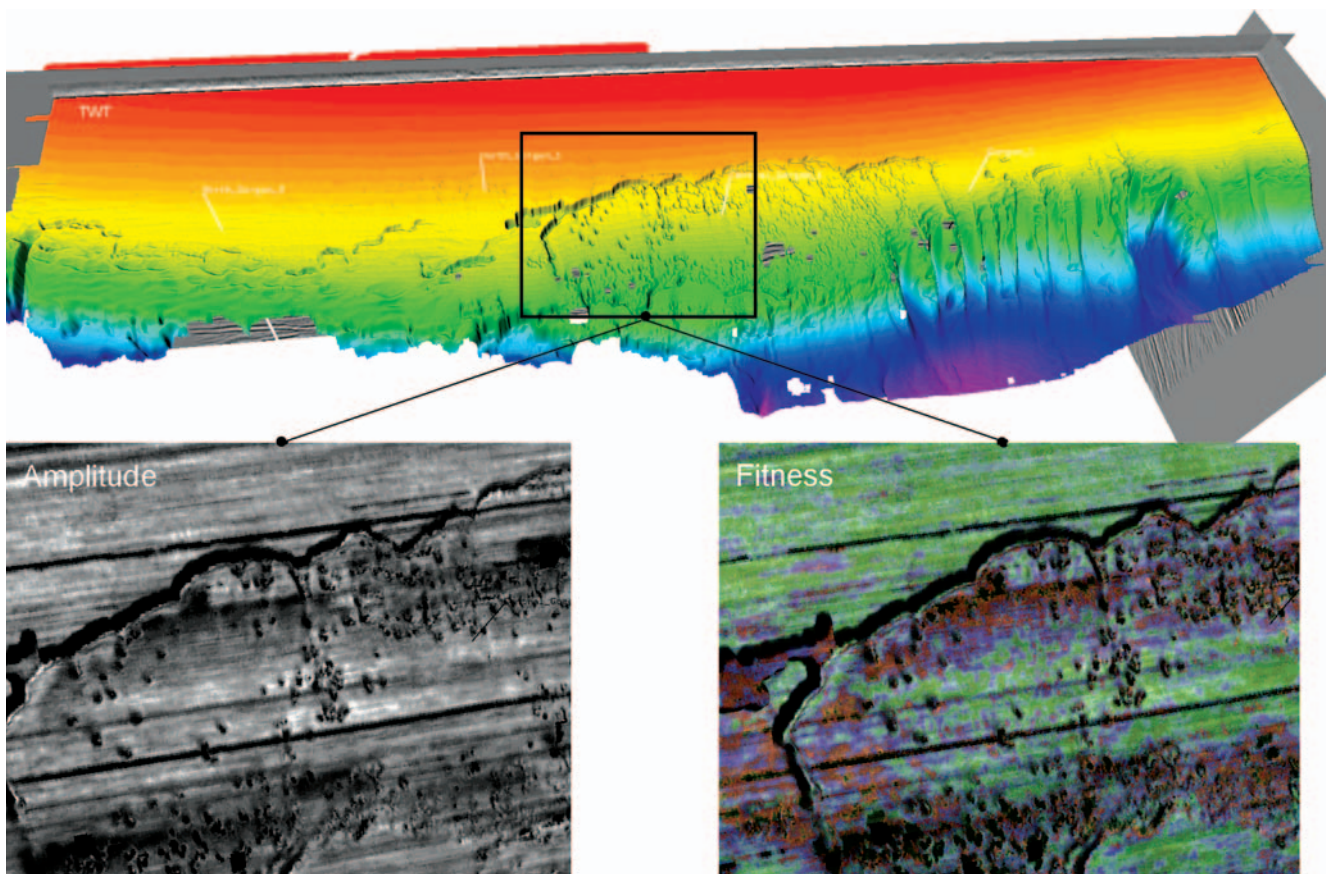


Fig. 9. Evidence of outgassing over the Gorgon Gas Field offshore Western Australia. While zones of outgassing can reduce the exploration risk and demonstrate areas of active hydrocarbon migration, they can also provide an indication of possible drilling hazards.

At this site the target coal seam is approximately 210m below the surface. The seam has an average thickness obtained from the 60 core samples of 2.1m and the survey area is 7.6 square km. The coal seam of economic interest is the German Creek seam within the Permian Moranbah Coal Measures. There are several much thinner seams existing above the German creek, however these are not of underground economic interest within the project area. The seismic data are derived from a 3D dynamite source survey acquired on a brick acquisition geometry using six geophones grouped into a 2m array length, spaced 15m apart, along lines separated by 32m. A 150 gm PETN booster was used for the charge placed 2metres below the base of weathering. The dominate wavelength for the final processed signal is approximately 18 to 20metres. Higher resolution will generate greater detail at or near the target horizon, but this does not necessarily bring greater clarity or certainty to the interpretation. Figure 11 illustrates one horizon of many that can be used by the interpreter to provide an objective starting point to providing a meaningful geological interpretation.

Discussion and interpretation

The ability of an interpreter to provide a geological interpretation that is close to reality depends on a numbers of factors (e.g. experience, survey design, sampling, data quality, noise levels, etc.) If the interpreter focus is only on one or two horizons then the 3D volume has almost certainly been underutilised. Geological events both syn and post depositional combined with lithification can create a complex environment which requires a thorough assessment of the likely hazards

associated with placing people and machinery within that world. For the biologists, the initial model of the Human Genome was very wrong, because collectively they had chosen to ignore half of the data when they stripped away the encasing protein to study the DNA molecule. Since geoscientists create models from seismic 3D volumes often using less than 10% of the available data, these models are also likely to be incorrect or heavily biased. To use an often paraphrased quote; 'Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful' (Box and Draper 1987). While the genetic processing technique described above is capable of objectively extracting all the surfaces in a data volume, it is unlikely whether anyone would want to use all data for model construction. However, being able to review all the surfaces in a volume, and to develop an understanding of the variability within the dataset, should enable the geoscientist to develop models that are sufficiently detailed to capture that variability and not be so overly simplified as to render them useless.

From the examples shown, examination of the GeoPopulations™ provided insight into structure, geomorphology, fluid-flow, out-gassing and sea-floor stability at both regional and local scales. These insights can only be made when the volume surfaces can be reviewed in detail (preferably by a multi-disciplinary team). Often different disciplines are able to extract different types of meaningful information from the results. Therefore, by automating the surface extraction process and providing one or two orders of magnitude higher quality surfaces than conventional interpretation techniques, more time can be spent developing an understanding of the results instead

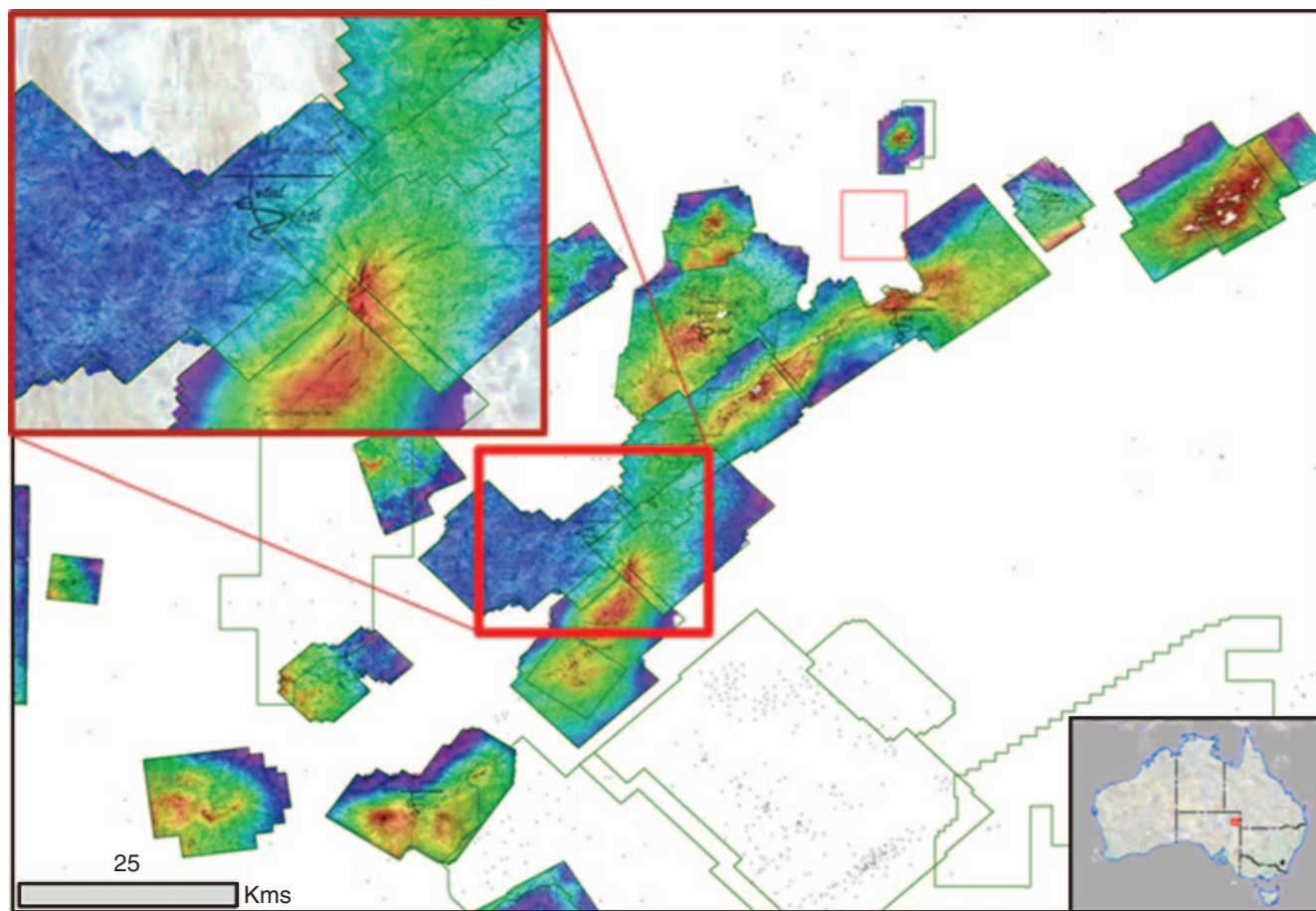


Fig. 10. The image shows the location of some of the 3D datasets available from the Cooper Eromanga basin. The coloured images show the TWT attribute for the same horizon on all 3D datasets. From the visual database of GeoPopulations™ similar composites could be made at virtually any surface. Colour bars are scaled independently.

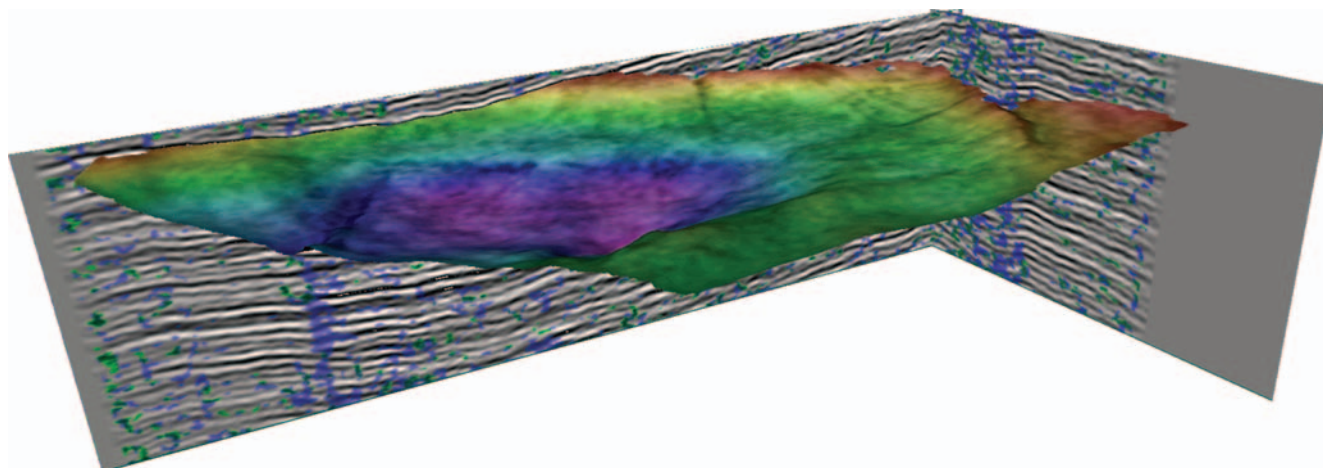


Fig. 11. TWT GeoPopulation™ with fault zones.

of losing time in the mechanics of the extraction process. In areas where a horizon is noisy or subject to coherent interference, the ability to rapidly have an objective horizon for critical review by the interpreter can significantly improve the reliability of the interpretation. Moreover, when this analysis is incorporated during the processing of the data, additional information can be used to optimise the processing of the data

and get useful data to the interpreter at a much earlier stage. Finally, older legacy data volumes, with the incorporation of geological and engineering data from the sub-surface team into the visual database, will form a knowledge base and provide teaching opportunities for the next project and the next wave of geoscientists. One might also speculate on how this technology would apply to other sets of waveform data collected by the

minesite (i.e. analysis of the radar guidance waveforms from the long-wall shearer).

Conclusions

This paper has described and illustrated a mathematical process for objectively providing a series of automatically picked horizons within a 3D seismic volume. The mathematical process emulates biological evolutionary stages whereby an initial population of individuals are randomly identified and given the opportunity to evolve. At the end of each generation, individuals which match the selection criteria, combine with an existing population forming offspring which inherit the genotype of two parents. Through the generations, the fittest have more chance to be selected and to reproduce, which enable them to grow faster than less fit individuals. Typically, the more continuous surfaces evolve first with the more complicated surfaces evolving last. Using this process we have illustrated examples where large multi-horizon 3D datasets can be assessed at either a micro or macro scale for horizon characteristics. The aim of the process is to provide a method whereby the interpreter can rapidly examine all the data, assess the significant aspects of the data then create a meaningful geological model which has been created based on a review of all the data.

Acknowledgements

The authors gratefully acknowledge the efforts of the team at Seisnetics (patent holder of this technology). Specifically, the guidance from David Jetelina in the use of the 3D visualization

tool used to view the visual database of GeoPopulations™. Finally, the authors would like to thank and congratulate Nabil Tnacheri on the development of this beautiful technology and encourage the Seisnetics team to keep up the good work.

References

- Baltimore, D., 2001, Our genome unveiled. *Nature*, **409**, 814–816. doi:10.1038/35057267
- Barnes, A. E., 2007, Redundant and useless seismic attributes. *Geophysics*, **72**, 33. doi:10.1190/1.2716717
- Box, G. E. P., and Draper, N. R., 1987, *Empirical Model-Building and Response Surfaces*, Wiley, p. 424.
- Chevron Australia, 2010, *The Power of Human Energy*, http://www.chevronaustralia.com/Libraries/Publications/100510_Chevron_Australia_Corporate_Brochure_2010.sflb.ashx
- Fang, J. H., Karr, C. L., and Stanley, D. A., 1996, Transformation of geochemical log data to mineralogy using genetic algorithms. *The Log Analyst*, **37**(2), 26–31.
- Grefenstette, J. J., and Baker, J. E., 1989, How genetic algorithms work: a critical look at implicit parallelism. In: *The Proceedings of the Third International Conference on Genetic Algorithms*, Kaufmann, M., San Mateo, California, pp. 12–19.
- Michalewicz, Z., 1992, *Genetic Algorithms + Data Structure = Evolution Programs*, 3rd edition, Springer-Verlag, p. 387.
- Pearson, H., 2006, What is a gene. *Nature*, **441**, 398–401. doi:10.1038/441398a
- Watson, J. D., and Crick, F., 1953, Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738. doi:10.1038/171737a0



**From rugged landscapes
to stringent environmental
standards, whatever
Australia dishes out,
the GSR can take.**

The cable-free Geospace Seismic Recorder (GSR) is transforming the seismic industry from Tierra del Fuego to Turkey. It's allowing contractors to work quickly, safely and nondisruptively in terrain they never thought possible – with the smallest environmental footprint ever.

OYO GEOSPACE

www.oyogeospace.com

0011.1.713.986.4444 USA

sales2@oyogeospace.com

