Soft computing for qualitative and quantitative seismic object and reservoir property prediction

Part 3, Evolutionary computing and other aspects of soft computing

Fred Aminzadeh*

This is the third part of the series of review papers on soft computing applications in the petroleum industry. In this paper we will focus on evolutionary computing, with related topics such GA (genetic algorithms), genetic engineering, genome, DNA, artificial life and emergence intelligence. We will begin with a brief overview of evolutionary computing technology. We will also give an overview of GA in exploration and production (E&P). We will then highlight some recent applications of genetic algorithms in various aspects of hydrocarbon E&P. These include applications in production optimization, reservoir characterization, and permeability prediction. We will also propose a framework for the more effective use of GA (Genome) as well as likely applications of 'complexity theory' in seismic exploration.

Introduction

Evolutionary computing techniques cover a large spectrum of related technologies. Among them are: GA, genetic engineering, genome, DNA and emergence intelligence. These technologies are already having a profound impact on many areas. Most notably, human genome has already found practical applications in life sciences (e.g. medicine and pharmaceutical industry). Figure 1, from the US Department of Energy's human genome initiative shows the link between DNA and life.

Some of these methods have been used on their own or in conjunction with other soft computing methods in many aspects of geosciences and hydrocarbon exploration and production problems. Most GA algorithms have been used as a means of efficient optimization. They also have been used to discover and extract knowledge or rules, especially when a large body of information has to be searched. Limited applications have used genome or DNA type concepts to categorize rock formations, recognize seismic patterns or describe the sedimentation process. These are the most promising application of evolutionary computing for hydrocarbon exploration.

The rest of this introductory section gives an overview of evolutionary computing. It also gives an overview of some of these methods in the petroleum industry. The rest of the paper highlights some recent applications in E&P, and provides a brief description of complexity theory which is expected to have many applications in exploration.

What is evolutionary computing?

Evolutionary computing is based on similar principles to those in nature (survival of the fittest, Charles Darwin). GA and subsequently 'artificial life' concepts were originally presented by Holland (1975). He improved the understanding of natural adaptation process, designing artificial systems with properties similar to those of natural systems. Today, GA turns out to be one of the most promising approaches for dealing with complex systems in spite of their apparent simplicity. Purely analytical optimization methods, although very efficient, require many restrictive assumptions (e.g. differentiability). GA offers a solution and is applicable to multi-objectives optimization and handles conflicts. It is highly efficient, easy to use and robust where multiple solutions exist. The following, adapted from Rennard (2002), provides an introduction to GA.

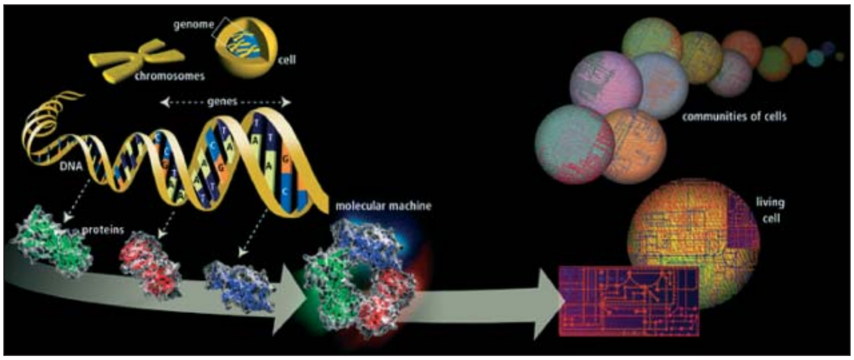

Figure 1 *From DNA to Life: US Department of Energy's Genome project (www.doegenome.gov)*

* *dGB-USA, 1 Sugar Creek Center Blvd., Suite 935, Sugar Land TX 77478, USA, fred.aminzadeh@dgb-group.com*

Evolution We go back 45 millions years examining a Basilosaurus, a 'prototype' whale. It had a quasi-independent head and posterior paws (Figure 2A). Its anterior members were reduced to small flippers requiring a big effort for movement in water. To swim effectively and be able to hunt with precision and speed, a double phenomenon must occur: the shortening of the 'arm' with the locking of the elbow articulation and the extension of the fingers which will constitute the base structure of the flipper. Through time, they appeared with longer fingers and short arms (Figure 2B). They could move faster and more precisely, living longer and having many descendants.

Other improvements in the general aerodynamic (e.g. integration of the head to the body and strengthening of the caudal fin) adapted it to the constraints of an aqueous environment. This process of adaptation is so perfect that nowadays the similarity between a shark, a dolphin or a submarine is striking. But the origin is a cartilaginous fish (Chondrichtyen) from the Devonian era (400 million years ago), long before the first mammal whose Cetacean descendants exist today. Thus, a Darwinian mechanism which is the basis of GA generates an optimization process of adaptation comprised of reproduction, selection and mutation.

A

B

Figure 2A *Basilosaurus, Pre-historic whale.* **2B** *Tursiops flipper, two fingers of the common dolphin is hypertrophied to the detriment of the rest of the member.*

Evolution and GA The genetic pool of a given population potentially contains the solution, or a better solution, to a given adaptive problem. This solution is not 'active' because the genetic combination on which it relies is split between several subjects. Only the association of different genomes can lead to the solution. For example, shortening of the paw and the extension of the fingers of our basilosaurus are controlled by two 'genes'. No subject has such a genome, but during reproduction and crossover, a new genetic combination occurs and a subject can inherit a 'good gene' from both parents. In Emmeche (1994) it is shown how mutation (which by itself does not lead to any optimal solution), in conjunction with John Holland's genetic recombination (crossover), greatly improves the capability of the algorithm and eventually finds the optimum.

Functioning of a GA The 'chromosomes' encode a group of linked features. 'Genes' encode the activation or deactivation of a feature. Let us examine the global genetic pool of four basilosaurus belonging to this world. We will consider the 'chromosomes' which encode the length of anterior members ('paw' and 'fingers'.) They are encoded by four genes, the first two for the 'paw' and the others for the fingers. In our representation of the genome, the circle with blue background depicts the activation of a feature; the cross with green background is for deactivation. The ideal genome (short paws and long fingers) is: . The genetic pool of our population is comprised of A: , B: , C: and D: . Note that A and B are the closest to their ancestors and D is close to the optimum, needing a small lengthening of his fingers. The fitness measure or probability of reproduction is computed by giving one point to each gene corresponding to the ideal. The perfect genome will then get four points.

Consider a cycle of reproduction with four descendants. D will be selected four times, then getting four descendants. C will be selected twice, getting two descendants. A and B will only be selected once. The reproduction pattern is given in Table 1. During reproduction, crossovers occur at a random place (centre of the genome for A', B' and C', just after the first gene for D'). The link between the degree of adaptation and the probability of reproduction leads to a trend rising to the average fitness of the population. In our case, it jumps from seven to 10.

Subject	Received genes	Genome	Fitness	Reproduction probability
A'	A : , D :		2	2/10=0.2
B'	B : , D :		2	2/10=0.2
C'	D : , C :		3	3/10=0.3
D'	C : , D :		3	3/10=0.3
Total			10	10/10=1

Table 1 *Genetic pool of population*

During the following cycle of reproduction, C' and D' will have a common descendant, with the new subject inheriting the intended genome: D': + C': = . Crossover is the basis of GA: there are nevertheless other operators like mutation. In fact, the desired solution may happen not to be present inside a given genetic pool, even a large one. Mutations allow the emergence of new genetic configurations which, by widening the pool, improve the chances of finding the optimal solution.

Adaptation and selection: the scaling problem The conventional GA methods as described above have two problems. 1) A 'super-subject' being selected too often makes the population converge towards that genome and this reduces the diversity of the genetic pool, hampering the progress of GA. 2) With the progression of GA, the differences between fitness measures are reduced. This gives a similar selection probability to all, stopping the progress of the GA. To alleviate the problem, we transform the fitness values using one of the following methods: windowing, exponential, linear transformation, and linear normalization. The following summarizes a typical GA:

1. Encode of the problem in a binary string.
2. Generate a population (genetic pool): a group of possible solutions, randomly.

Advantages	Disadvantages
Only uses function evaluations.	Cannot use gradients
Easily modified for different problems	Not easy to incorporate problem specific info.
Handles noisy functions well	Not good at identifying local optima
Handles large, ill defined search spaces	Not effective terminator
Good for multi-modal problems,	Not effective for smooth uni-modal functions
Returns a suite of solutions	Needs coupled with a local search technique
Very robust to difficulties in the evaluation of the objective function	Needs to be coupled with a local search technique
Easily parallelized	

Table 2 *Advantages and disadvantages of Genetic Algorithms*

3. Calculate fitness measure (distance from the optimum) of each subject.
4. Select subjects that will mate based on their share in the global fitness population.
5. Genomes crossover and mutations.
6. Repeat the process from step 3.

Naturally, GA is not suitable for all types of problems. Table 2, from Carter (2003), highlights some of the advantages and disadvantages of GA. To address some of the short comings of GA, new coding methods based on biological DNA are introduced. Genome or DNA coding is suitable for knowledge extraction from a large data set. The DNA has many redundant parts which is important for extraction of knowledge. In addition, this technique allows overlapped representation of genes and it has no constraint on crossover points and the same type of mutation can be applied to every locus.

Here, the length of chromosome is variable and it is easy to insert and/or delete any part of DNA. The role of the genome is to gather the information needed to construct the phenotype of an individual. This should be done in a way that, when the crossover operator is applied, a viable offspring is produced with a high level of inheritance at the genotype level. The genome design should also try to preserve as much information about relationships between genes as possible, by collecting related genes together in the same chromosome, and even in the same part of the chromosome. The structure of a chromosome can be anything that helps retain important relationships between genes.

Applications of GA in geosciences, oil exploration and production

Most of the applications of GA in the area of petroleum reservoirs or in the area of geosciences are limited to inversion techniques or used as optimization techniques. The following are some of such applications:

Reservoir engineering: Sen et al. (1995) applied GA in a fairly standard way to the generation of stochastic permeability fields from outcrop and tracer flow data. The permeability values at grid points (matching parameters) were coded in a string of binary numbers (chromosomes). They worked with population sizes of typically more than 200 and used a random procedure to assign the initial values. Bush and Carter (1996) developed modified GA that included non-standard binary encoding and breeding strategies. The approach was tested on a synthetic, vertical cross section with only three parameters: sand and shale permeability, and fault throw. These parameters were encoded in a binary string of variable length. Guerreiro et al. (1998) successfully applied GA to the identification of properties of heterogeneous reservoirs through the matching of tracer breakthrough profiles using six parameters. These parameters were the geometry (position and sides dimensions) of a rectangular insertion and the porosities inside and outside the insertion, of a synthetic model built on a heterogeneous quarter of five-spot. Mohaghegh (2003) and Johnson and Rogers (2003) used GA for production optimization. The latter is highlighted below. See also Wong (2002) and Nikravesh et al (2003) for details of some of the GA based methods and their applications.

Petrophysics and geologic applications: - Huang, et al (1998) showed the use of GA in combination with neural-fuzzy-analysis to interpolate and analyze log data. Nikravesh et al (1999) proposed use of a neuro-fuzzy-genetic model for data mining and fusion in the area of geosciences and petroleum reservoirs. In addition, use of a neuro-fuzzy-DNA model has been proposed for extraction of knowledge from seismic data and mapping the wireline logs into seismic data and reconstruction of porosity (and permeability) using multi-attributes seismic mapping. Saftic, and Velic (2000), established"gnetic" stratigraphic sequences in upper Miocene sediments. Potter, and Corbett, (2000,) used genetic petrophysics and data integration in to improve prediction of key parameters.

Seismic applications: Stoffa and Sen (1991) applied GA for the inversion of plane-wave seismograms. Sen and Stoffa, (1992) showed examples of seismic waveform inversion with rapid sampling of model space based on the GA. Stork, and Kusuma, (1992) used a genetic method to correct for large amplitude statics in noisy data. Mallick and Frazer (1995) used a GA for model-based inversion of pre-stack data. Roth, and Hollinger (1998) used GA to derive near surface P- and S-wave velocity information from dispersion characteristics of Rayleigh and guided waves. The method is based on the assumption that the near surface Poisson's ratio is high, allowing the guided waves to be approximated as acoustic signals. Snieder, (1997) used GA for reconditioning inverse problems. Barnes (1997) applied genetic classification for complex seismic trace attributes analysis. Ma (1998) used simulated annealing for generating acoustic impedance from post stack seismic data and the respective interfaces.

McCormack et al (1999) overviewed applications of GA in exploration and production. Mallick (1999, 2001) introduced a refined pre-stack inversion method using a GA. Casting the GA into a Bayesian framework, a priori information of the model parameters and the physics of the forward problem were used to compute synthetic data. The synthetic data were then matched with observations to obtain approximate estimates of the marginal a posteriori probability density (PPD) functions in the model space. Examining these PPD, the interpreter can choose models which best describe the specific geologic setting and lead to an accurate prediction of seismic lithology. Ji et al (2000), quantified inversion sensitivity of amplitude variations with slowness. Porsani and Ursin, (2000,) used GA for deconvolution and wavelet estimation. Duncan and Latkiewicz, (2002) discusssed use of genome code to unravel the intrinsic characteristics of the subsurface from 'decoding' the seismic data. We will elaborate on this topic below. Qin et al (2003) describe a velocity analysis method based on CDP mapping using a GA.

Highlights of some recent GA applications in E&P

In what follows we will highlight some of the recent applications of GA in production optimization, reservoir characterization and permeability prediction.

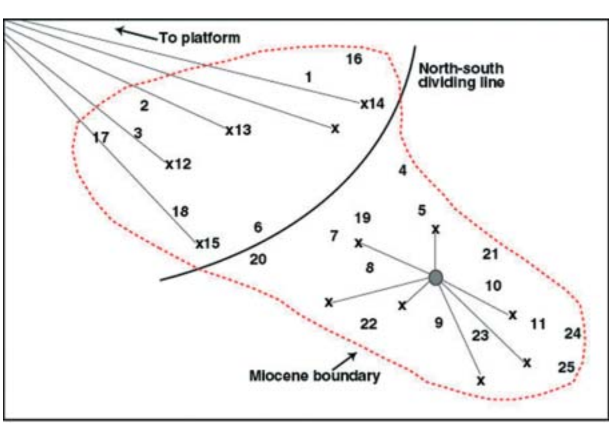

Figure 3 *Production (x) and injunction (numbered) or planned injection (#/x). The numbering is arbitrary, as is the location in the bit-string.*

Production optimization

In this section GA method is applied to optimize water flooding for production optimization by searching for best well combination. It is adapted from Johnson and Rogers (2003) where GA was used as search engines to train a neural network to be used for an optimization / water flooding project in a test field. The goal was to locate the combination of 1-4 injection wells which would maximize the field's simple net profit over seven years. Using a standard reservoir simulator, a knowledge base of 550 simulations sampling different combinations of 25 potential injection locations was created. The knowledge base was first queried to answer questions concerning optimal scenarios for maximizing simple net profit over three and seven years. A discussion on the use of GA to address this problem is given below.

Use of GA The 25 well locations which form the decision variables are represented in the GA as a string of 25 bits, each of which can be on or off. The spatial location of each well is fixed and implicit in the representation. The order of the well locations in the string is indicated by their identification numbers in Figure 3. The search is initialized with a set of 100 well combinations each evaluated according to an objective function. A new generation of 100 combinations is created from the old one by means of three mechanisms: selection, reproduction, and mutation.

Selection This mechanism determines which members of the current generation will be selected for carry-over, in one form or another, to the new generation. To ensure the highest-ranking combinations are not lost through accidents of selection and crossover, the top three combinations are copied over to the new generation intact. The remaining 97 slots are filled through a process where parents must be selected. Selections for reproduction are made, two at a time, to obtain parent combinations from which a child combination will be formed. This process is repeated until all children have been generated. The same combination may constitute both members of the pair, in which case the child is simply a clone of the parent.

Reproduction (Crossover) The most common form of reproduction is single-point crossover. Child combinations are constructed by breaking the parent combinations apart at some randomly selected crossover position in the bit-string and joining segments from each parent. We create new combinations from 'chunks' of old ones when proximity in the bit-string is important. That is, the proximity of wells in the bit-string should reflect one or more dimensions of relatedness in the physical problem it represents. To break up spurious associations, a different reproductive mechanism, uniform crossover, is used. In this method, the value of each bit in the child string is set independently of every other bit. The exchange probability can be biased to favor the fitter parent, if any; but in this study the exchange probability is kept at an impartial 0.5.

Mutation - Mutation is a way to maintain diversity in a population by arbitrarily changing the values of bits in the child combinations according to some rate, often the inverse of the population size. A high mutation rate can undermine the effects of crossover; a low one limits the introduction of 'novelty' into the population.

Typically, optimization of termination criteria is usually based on some notion of convergence to a single best solution. Here, we are interested in generating sets of near-optimal solutions rather than a single best solution. This goal is achieved by tying termination criteria to the performance score of the population rather than to the highest-ranking individual combination. The search terminates when either the mean population score fails to improve over five consecutive generations or some maximum number of generations (25) have elapsed. The maximum number at the end of every generation is saved. The top-ranked unique combinations become the set of near-optimal solutions. The outcome of search is influenced by the random choices that are made. To improve the stability of the outcome, the results of each search consist of combined results from 10 searches, each with a different seed initializing the pseudo-random number generator.

Emphasis was placed on the best performing scenario located by each method. This consisted of sets of near-optimal scenarios which can be analyzed in an effort to better understand the underlying physical reasons why these scenarios are optimal answers to a particular management question. For example, an examination of the top 25 well combinations from the ANN-GA search found that well 7 figured in 100% of the combinations, followed at a distance by well 24 (32%), well 11 (28%), and well 16 (24%). One might speculate that well 7 has a larger sweep of neighboring producers that are important to production over the seven year time-frame. The other popular wells may be reflective of more conventional wisdom regarding the desirability of raising pressures near the boundaries of the reservoir.

Another important issue involved uncertainties associated with the reservoir simulator itself. Given the fact that there are many alternative 'nearly optimal', it is important to have some indication of how great a variation is introduced by considering these alternatives. Although this is generally a major challenge, there may be some incremental strategies for incorporating aspects of uncertainty analysis into the ANN-GA methodology at different stages of the optimization process. A very simple strategy is to rank each well location by the relative certainty of the physical properties in its vicinity. The objective function being optimized would contain a penalty term based on that rank, which will reflect the informational-risk associated with including that well in the scenario.

Reservoir modelling and characterization

This section is adapted from Romero and Carter (2003) where a GA for reservoir characterization is introduced. In a typical reservoir modelling/characterization problem, the objective is to assign initial values for various reservoir parameters in every grid block (usually about 20,000). The size of the model is often determined by three factors: the perceived geological complexity of the reservoir, the computational resources and the time available for a study to be completed.

Some of the reservoir properties to identify are:

- Location in space
- Size
- Porosity or the fraction of the total volume that is occupied by fluid
- Saturation or the fraction of the pore volume occupied by each type of fluid
- Pressure
- Permeability or the ease with which fluids can flow horizontally and vertically
- V-shale or fraction of non-reservoir rock

These values must be consistent with some physical properties, e.g. the grid blocks must form a continuous, non-overlapping, space; oil/water mixtures are found above water only regions; and pressures must be consistent with principles of fluid hydrostatics. The state-of-the-art at the moment usually assumes that reservoir characterization is a process that determines four variables for each grid block, these being: horizontal permeability; vertical permeability, porosity and the volume of shale fraction (V-shale or net-to-gross ratio). Other variables may be included in the reservoir characterization process are: properties of geological faults, connate water saturation, irreducible oil saturations, and well skin factors.

The daunting task is to determine the values of approximately 100,000 data values, which determine the initial conditions, with very small number of data measurements. The reservoir used in this study was 11 km x 3 km x 200 m, and contained 11 production wells and six injection wells. We therefore have an aerial coverage of one measurement of each

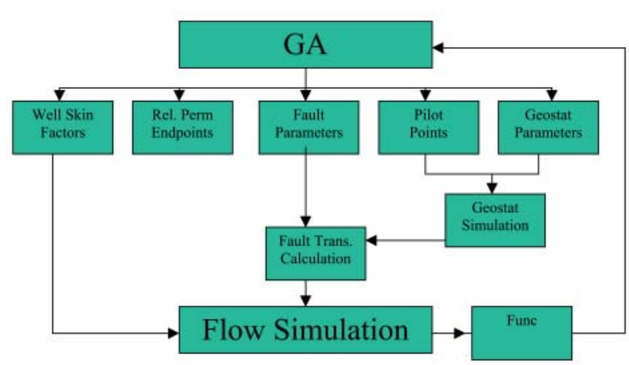

Figure 4 *Schematic of the reservoir characterization process with 'key variables'*

property for every 2 km^2. Each measurement is valid for a volume of about $1m^2$ and 20 m in height, so we have a total of 510 measurements. We also have measurements of water/oil flow rates through each well every month for four years (a total of not more than 1344 well measurements). Soft information from geological and geophysical data helps alleviate the problems associated with sparse hard measurements. Spatial distribution of properties must be consistent with the statistical properties of the geological processes that create the reservoirs.

Many of the distributions of reservoir properties (e.g. permeability and porosity) that are consistent with the measurements of oil/water flow rates are either inconsistent, or geologically improbable. Therefore we need to determine which distributions of initial conditions are consistent with the expected geological statistics and which lead the numerical model (simulator) to predict measurements that are sufficiently close to the measured data. We have moved from trying to find descriptions that match the measurements, to finding descriptions that have a high probability of being consistent with the available knowledge. In many cases we satisfy ourselves with trying to identify the description with the highest probability.

In the reservoir characterization process shown in Figure 4, the following are key variables:

- Variables that describe the distribution of three field properties (porosity, permeability and V-shale) throughout the reservoir
- Some variables that describe the flow of fluids across faults.
- Variables for the mechanical skin factors in each of the wells.
- Relative permeability end-points.

As shown in Figure 4, field properties are used in two ways, directly in the flow simulation and as an input to the calculation of the fault properties. Geological faults are common features in hydrocarbon reservoirs, and numerical models therefore must include geological faults. Sealing capacity of major faults is one of the most influential parameters contributing to the overall model uncertainty. Faults influence flow in a reservoir simulation model in two ways. They modify the connectivity of sedimentological flow units, because displacements across faults cause partial or total juxtaposition of different flow units. They also affect flow because of the petrophysical properties of the fault-rock. Both thickness and fault permeability are physically observable properties of fault zones. Well skin factor usually lowers the well-bore pressure.

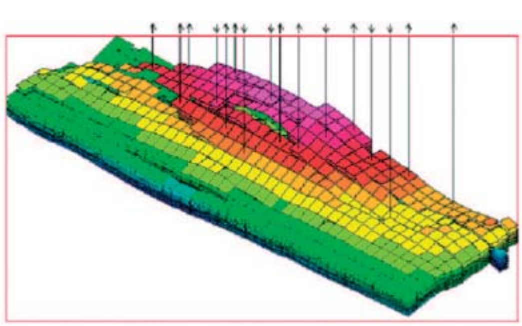

Figure 5 *Reservoir structure with well locations from Romero and Carter (2003)*

Reservoir description and production plan

In this work a synthetic reservoir model was used to avoid problems with scarcity of data and structural uncertainty. The reservoir model structure is fairly typical of North Sea fault-bounded trap reservoirs. It is bounded by a large reservoir normal fault. As shown in Figure 5, intra-reservoir normal faults have down-throw directions which are both synthetic and antithetic to those of the main fault. To generate the `truth' case of our reservoir model, the methodology for generating models within the reservoir characterization process and associated key variables was used. The production plan was made such that it was realistic and provided adequate data for history matching. Among many simplifying assumptions made were: vertical wells, one drilling phase identical operational constraints each year. Production profiles were generated using the true models. Gaussian errors were added to the measurements.

Design of the GA Figure 4 shows a GA to control the inputs to the reservoir description stage. Among the reasons for using the GA instead of conventional gradient-based algorithms are:

- The large size of the search space
- Difficulty in pre-determining characterization sensitivity to different variables
- Lack of uniqueness in mapping variables to a given reservoir description
- Problems with non-linearity, multi-modality and being highly underdetermined
- Gradient information is not readily available for every variable.

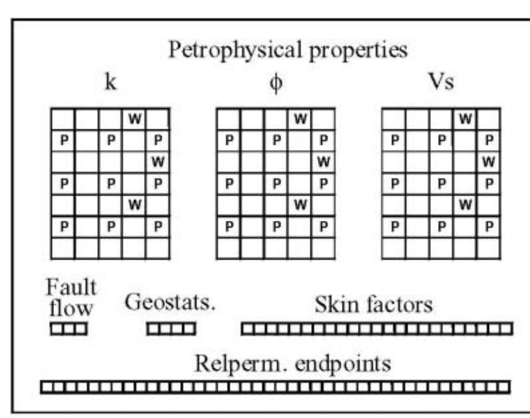

Figure 6 *Non-standard structure for the genome design*

Figure 6 shows the structure of genomes for different parameters. The first three chromosomes (related to reservoir properties) share the same complex 3D structure. The fault parameters are encoded as three real numbers, with known upper and lower bounds. The geostatistical parameters are encoded as four real numbers, again with known upper and lower bounds. The skin factors and the relative permeability endpoints are encoded as 17 and 36 real numbers, respectively. Two crossover operators were used. For the one dimensional chromosomes, a conventional k-point crossover was the choice, where the probability of switching to extracting information from the other parent is constant for each chromosome. For the three dimensional chromosomes a special bit-flip crossover operator was designed.

The optimization progress in GA usually involves a very steep decline of the objective function followed by less steep decline. The error from the water production component of the function has the largest contribution. The good history matching results were validated at the well locations of Figure 5. Among main advantages of the method are:

- The algorithms returns a suite of solutions from which the best one(s) are chosen.
- Multiple realizations could be obtained without repeating history matching.
- Easy to implement and robust with respect to lost or corrupted solutions
- Easily parallelized because it is computationally cost effective
- Requires only a modest # of forward simulations to obtain good solutions.
- Not very sensitive to the parameter settings, thus suitable for characterization.

GP classifiers for modelling reservoir permeability

In Part 2 of this series (Aminzadeh and Wilkinson, 2004), we showed how fuzzy logic was used to predict permeability, the most difficult reservoir property. The modelling task in that study was carried out using Genetic Programming (GP), Koza (1992). Wilkinson et al (2003) describe the modeling task using GA classifiers as follows:

In each lithology group, we used GP to train the classifiers that predict permeability ranges. Each classifier is specified with different thresholds for different permeability ranges. For example, the two thresholds to distinguish high, medium and low permeability in sand are 3.0 and 2.0. For shaly sand, sandy shale and shale, those threshold pairs are (2.5, 1), (2.5, 1), and (1, 0) respectively. The GP software used is only able to train binary classifiers. For lithology groups that have three possible classes, two classifiers are required. The GP parameters used to train these classifiers are listed in Table 3. The fitness is based on hit rate: the percentage of the training data that is correctly classified.

During tournament selection, four candidates are randomly selected to compete for two winners. If two candidates are 'tied' in their hit rates, the squared error measurement is used to select the winners. The 'tied threshold' for squared error measurement is 0.01% in this work. If two classifiers are tied in both their hit rates and squared error measurements, one of them is selected as the winner randomly. The best classifiers are assembled into teams. A team may consist of any odd number of classifiers. The decision of a team is according to the majority votes. Since the number of classifiers in a team is odd, there is always a winner (i.e. no tie). A team example is given here:

$$f_1(\phi, v, \rho) = K\ \rho * v - \phi^2 / \rho K$$

$$f_2(\phi, v, \rho) = K\ \rho^v - \phi^2 / \rho + v + 1.2K$$

$$f_3(\phi, v, \rho) = K\ v * v - \phi^2 / \rho - (\phi + 2.3)K$$

threshold = 2.5;

Majority ((f1 >threshold), (f2 > threshold), (f3 > threshold))

Objective	Generate a classifier that distinguishes high, medium and low permeability.
Functions	addition; subtraction; multiplication; division; abs; data transformation
Terminals	Porosity (ϕ), density (ρ), velocity (1/v) and constants.
Fitness	Hit rate then squared error fitness
Hit rate	The percentage of the training data that are correctly classified.
Selection	Tournament (4 candidates/2 winners)
Pop size	100,000
Max generation	9,000,000
Max length	256 nodes
Genetic operators	50% crossover (Homologous 95%), 95% mutation (Block mutation 30%, Instruction mutation rate 65%)

Table 3 *GP parameters*

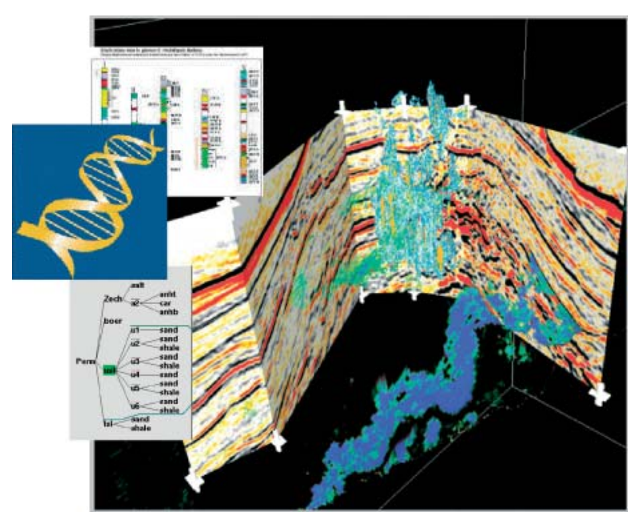

Figure 7 *An illustration of how seismic patterns could be linked to DNA (seismic visualization, courtesy of www.opendtect.org)*

Future applications of evolutionary computing and related techniques

In this section human genome (DNA) related methods and complexity, two emerging technologies with possible profound future impact in the petroleum industry, are discussed.

Genome for 'fingerprinting' outcrops, geochemical and seismic data

The ultimate dream of an explorer is to be able to use a small sample from direct measurements such as outcrops and thin sections from cores and deduce something concrete about the geologic history and prospectivity of the subsurface structures. The second best thing is to be able to glean similar information from indirect measurements such as seismic data, well data or surface geochemistry data. With the advances made in human genome technology and DNA matching, such work is becoming more practical.

The idea is to be able to link patterns from seismic data, well log data or bio marker data to specific 'finger prints' of different geologic regimes or sedimentation process that make it likely to have hydrocarbon source rock and economic reservoirs. The possibility of success in such efforts will be strengthened if one is able to create models that include the required 'building blocks' allowing necessary perturbation and yielding easily identifiable 'genes' that are prospective. Building blocks may explain how each gene works in a team. Similar to a sedimentation process, the agent plays against the environment and if it wins the process goes on.

That is what the business of 'emergence' or adaptation is all about: building blocks at one level combining into new building blocks at a higher level. A 'unified theorem' promised by complexity, as described in the next section, may offer some hope of deriving such DNAs from different types of data, irrespective of the scale and noise/uncertainty issues. For examples, recognizing different patterns or texture within seismic data and relating them to a carefully designed hierarchical geologic model may be made possible by linking them to a set of DNA or genomes (Figure 7).

In geosciences and exploration, we may have some of the ingredients for this type of analysis and Although some preliminary attempts have been made, we Duncan and Lat 2002 are far from making full use of such technology in exploration problems.

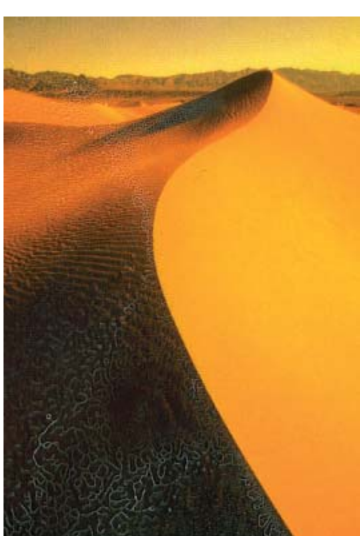

Figure 8 *Edge of chaos from Waldrop (1992)*

Complexity

One of the likely areas with potentially major impact in the way we explore in the years to come is use of complexity theory. Complexity is the science of understanding how independent agents interact with each other to influence each other and the whole. Organisms constantly adapt to each other through evolution, thereby organizing themselves into an exquisitely tuned ecosystem. Self-organizing systems are adaptive. They actively try to turn whatever happens to their advantage. They possess a kind of dynamism that makes them qualitatively different from static objects. Complex systems are more spontaneous, more disorderly, and more alive than the chaotic systems.

Chaos theory has shaken science to its foundations with the realization that very simple dynamical rules can give rise to extraordinary intricate behaviour - river and fractals. Complex systems have somehow found a way to acquire the ability to bring order and chaos into balance. Complexity theory can best be described as where order and chaos meet. The edge of chaos is where new ideas and innovative genotypes are forever nibbling away at the edges of status quo. The edge of chaos is the constantly shifting battle zone between stagnation and anarchy, the one place where a complex system can be spontaneous, adaptive, and alive.

As Waldrop (1992) states in a fascinating book on complexity, '…he could sense that the old categories of science were beginning to dissolve. Somehow, a new, unified science was out there waiting to be born. It would be a rigorous science. He was convinced, just as 'hard' as physics ever was, and just as thoroughly grounded in natural law. But instead of being a quest for the particles, it would be about flux, change and the forming and dissolving of patterns. Instead of ignoring everything that wasn't uniform and predictable, it would have a place for individuality and the accidents of history. Instead of being about simplicity, it would about, well, complexity.'

There are some related concepts in complexity theory. For example, the illusive 'unified theory' that has been the focus of research at the Santa Fe Institute, as described by

Waldrop (1992), has begun to yield some results by establishing some correlation between otherwise seemingly unrelated fields and concepts as follows:

1. Cellular Automata Classes: I & II => 'IV' => III
2. Dynamic Systems: Order => 'Complexity => Chaos
3. Matter: Solid => 'Phase Transition' => Fluid
4. Computation: Halting => 'Undecidable' => Non-halting

As it was suggested in Aminzadeh (2000), chaos and complexity theory are considered to be among the possible frameworks for new generation of seismic signal processing and analysis tools. They offer a robust and stable behaviour in a highly complex and multi-dimensional system.

The question is whether we can exploit these ideas to better understand some of the problems we deal with in oil exploration and production:

- How much more can we learn from seismic response at the fault location?
- Are salt boundaries, or elsewhere with significantly different texture (edge of chaos), important physically and how we can establish the morphology?
- Is there a theory that can describe 'super-permeability'?
- How we can better exploit high frequency loss information usually present under major HC bearing reservoirs?
- Are source rock generation and HC expulsion related to the 'edge of chaos'?

It is hoped future research will provide some concrete answers to these questions, allowing us to take a leap forward in our technology applications

Conclusions

We have given a brief overview of evolutionary computing and GA. We have reviewed the use of GA in the petroleum industry. We have concluded that such applications are only scratching the surface regarding potential impacts of evolutionary computing in hydrocarbon exploration. Some future potential applications of DNA and complexity are also discussed, illustrating their potential use in different aspects of exploration, field development and production.

References

Aminzadeh, F. [2000] Challenges - Direct Future of Geophysics, *American Oil and Gas Reporter,* January, 123-132.

Barnes, A. E., [1997] Genetic classification of complex seismic trace attributes. *67th Annual Meeting, EAGE.* Geophys., 1151-1154.

Bush M.D. and Carter, J.N. [1996] Application of a Modified Genetic Algorithm to Parameter Estimation in the Petroleum Industry, In Dagli et al. (eds.), *Intelligent Engineering Systems through Artificial Neural Networks* 6, ASME, Press, York, 397-402.

Carter, J. N. [2003] Genetic computing, in *Soft Computing and Intelligent Data Analysis* Eds., Nikravesh, M., Aminzadeh, F., Zadeh, L. A., Elsevier.

Curtis, A. and Snieder, R. [1997] Reconditioning inverse problems using the genetic algorithm and revised parameterization. *Geophysics,* **62**, 1524-1532.

Dorrington , K. P. et al. [2004] Genetic-algorithm/neural-network approach to seismic attribute selection for well-log prediction. *Geophysics* **69**, 212.

Duncan, P. M. and Latkiewicz, C. [2002] Technology Unravels 'Genome Code' of 3D Data to Improve Quality, Speed of Seismic Exploration. *World Oil*, 99-107

Emmeche C. [1994] *Garden in the Machine. The Emerging Science of Artificial Life.* Princeton University Press, 114

Goldberg, D. E. [1989] *Genetic algorithms in search optimization and machine learning.* Addison-Wesley Publ. Co.

Guerreiro, J.N.C., Barbosa, H.J.C., Garcia, E.L.M., Loula A.F.D, and Malta, S.M.C. [1998] Identification of Reservoir Heterogeneities Using Tracer Breakthrough Profiles and Genetic Algorithm. *SPE Reservoir Evaluation & Engineering,* 218-223.

Holland J.H. [1975] *Adaptation in natural and artificial system.* Ann Arbor, University of Michigan Press.

Huang, Y., Wong, P. M. and Gedeon, T. D. [1998] Neural-Fuzzy-Genetic-Algorithm Interpolator in Log Analysis. *60th Annual Meeting, EAGE,* Session: P106.

Ji, Y., Singh, S. and Hornby, B. [2000] Sensitivity study using a genetic algorithm: inversion of amplitude variations with slowness. *Geophysical Prospecting* **48**, 1053-1074.

Johnson, V. M. and Rogers, L. L. [2003] Applying soft computing methods to improve the computational tractability of subsurface simulation-optimization problem, in *Soft Computing and Intelligent Data Analysis*, Eds., Nikravesh, M., Aminzadeh, F., Zadeh, L. A. Elsevier.

Koza, J. [1992] *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, Cambridge, MA.

Ma X. Q. [1998] An Accurate Seismic Inversion Approach Using Exact and Over-Parameterization methods. *SEG Extended Abstracts.*

Mallick, S. [1995] Model-based inversion of amplitude-variation-with-offset data using a genetic algorithm. *Geophysics,* **60**, 939–954.

Mallick, S. [1999] Some Practical Aspects of Prestack Waveform Inversion using a Genetic Algorithm: An example from the East Texas Woodbine Gas Sand. *Geophysics,* **64**, 2, 326-336.

Mallick, S. [2001] Prestack waveform inversion using a genetic algorithm - The present and the future. *Recorder*, **26**, 6, 78-84

McCormack, M. D., Stoisits, R. F., Macallister, D. J. and

Crawford, K. D. [1999] Applications of genetic algorithms in exploration and production. *The Leading Edge* **18**, 6, 716-718.
Mohaghegh, S. D. [2003] Virtual Magnetic resonance logs, a low cost Reservoir Description tool, in *Soft Computing and Intelligent Data Analysis,* Eds., Nikravesh, M., Aminzadeh, F., Zadeh, L. A., Elsevier.
Nikravesh, M., Aminzadeh, F., Zadeh, L. A. [2003] *Soft Computing and Intelligent Data Analysis*. Elsevier.
Porsani, M. and Ursin, B. [2000] Deconvolution and wavelet estimation by using a genetic algorithm. *70th Annual Meeting, SEG,* 2185-2188.
Potter, D. and Corbett, P. [2000] Genetic petrophysics and data integration in pegasus - improved prediction of key parameters. *62nd Annual Meeting, EAGE,* Session:X0012.
Qin, Y., Zhang, Z. and Xu, S. [2003] CDP mapping in tilted transversely isotropic (TTI) media. Part II: Velocity analysis by combining CDP mapping with a genetic algorithm. *Geophysical Prospecting,* 51, pp. 325-332.
Rennard, J. P. [2000] www.rennard.org/alif
Romero C.E. and Carter J.N. [2003] Using genetic algorithms Genetic Algorithm for Reservoir Characterization, in *Soft Computing and Intelligent Data Analysis,* Eds., Nikravesh, M., Aminzadeh, F., Zadeh, L. A. Elsevier.
Roth, M. and Hollinger, K. [1998] Joint Inversion of Rayleigh and Guided Waves in High Resolution Seismic Data Using Genetic Algorithms. *SEG Extended Abstracts.*
Saftic, B. and Velic, J. [2000] Genetic stratigraphic sequences in the upper Miocene sediments of the Sava basin. *62nd Annual Meeting, EAGE,* Session: P0033.
Sen, M. K., and Stoffa, P. L. [1992] Rapid sampling of model space using genetic algorithms: Examples from seismic waveform inversion. *Geophys. J. International,* **108**, 281–292
Sen, M.K., Datta-Gupta, A. K., Stoffa, P.L., Lake L.W., and Pope, G.A. [1995] Stochastic Reservoir Modeling Using Simulated Annealing and Genetic Algorithms. *SPE Formation Evaluation*, 49-55.
Stork, C. and Kusuma, T. [1992] Hybrid genetic autostatics: A new approach for large amplitude statics with noisy data. *62nd Annual Meeting, SEG,* 1127-1131.
Waldrop, M. M. [1992] *Complexity, the Emerging Science at the Edge of Order and Chaos*. Simon and Schuster, NY.
Wilkinson, D., Yu, T., and Xie, D. [2003] *Genetic Programming Theory and Practice*. Kluwer Academic Publishers, 271 – 289.
Wong, P. M., Aminzadeh, F. and Nikravesh, M. [2002] Soft Computing for Reservoir Characterization and Modeling. *Series Studies in Fuzziness and Soft Computing,* **80**, Physical Verlag, Springer, 2002.