The benefits and dangers of using artificial intelligence in petrophysics

Steve Cuddy

Petrophysicist, Petro-Innovations, Aberdeen, Scotland, United Kingdom

ARTICLE INFO

ABSTRACT

Artificial Intelligence, or AI, is a method of data analysis that learns from data, identify patterns and makes predictions with the minimal human intervention. AI is bringing many benefits to petrophysical evaluation. Using case studies, this paper describes several successful applications. The future of AI has even more potential. However, if used carelessly there are potentially grave consequences.

A complex Middle East Carbonate field needed a bespoke shaly water saturation equation. AI was used to 'evolve' an ideal equation, together with field specific saturation and cementation exponents. One UKCS gas field had an 'oil problem'. Here, AI was used to unlock the hidden fluid information in the NMR T1 and T2 spectra and successfully differentiate oil and gas zones in real time. A North Sea field with 30 wells had shear velocity data (Vs) in only 4 wells. Vs was required for reservoir modelling and well bore stability prediction. AI was used to predict Vs in all 30 wells. Incorporating high vertical resolution data, the Vs predictions were even better than the recorded logs.

As it is not economic to take core data on every well, AI is used to discover the relationships between logs, core, litho-facies and permeability in multi-dimensional data space. As a consequence, all wells in a field were populated with these data to build a robust reservoir model. In addition, the AI predicted data upscaled correctly unlike many conventional techniques. AI gives impressive results when automatically log quality controlling (LQC) and repairing electrical logs for bad hole and sections of missing data.

AI doesn't require prior knowledge of the petrophysical response equations and is self-calibrating. There are no parameters to pick or cross-plots to make. There is very little user intervention and AI avoids the problem of 'garbage in, garbage out' (GIGO), by ignoring noise and outliers. AI programs work with an unlimited number of electrical logs, core and gas chromatography data; and don't 'fall-over' if some of those inputs are missing.

AI programs currently being developed include ones where their machine code evolves using similar rules used by life's DNA code. These AI programs pose considerable dangers far beyond the oil industry as described in this paper. A 'risk assessment' is essential on all AI programs so that all hazards and risk factors, that could cause harm, are identified and mitigated.

Introduction

Alan Turing first considered the question, 'Can machines think?' in his seminal paper, Computing Machinery and Intelligence, published in 1950. AI is about getting computers to imitate human intelligence.

AI is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. If when leaving home, you notice that it is raining, you may decide to take an umbrella. That is natural intelligence. If your smart phone 'knows' you are leaving and that it is raining, it may suggest that you take an umbrella. This is a simple illustration of artificial intelligence.

The development of AI started at a workshop at Dartmouth College in 1956, where the term 'Artificial Intelligence' was coined (Crevier 1993).

Articles and papers have been since published concerning the nature of AI; by Minsky (1961), Newell (1963), Winston (1976), Davis et al. (2002) and many others. In the last decade, the petrophysics community has begun to make robust efforts to apply data-driven techniques to the challenges of petrophysical evaluation, including the application of AI. These come under the heading of machine learning techniques and applications include carbonate characterisation (Bigoni 2019), log depth matching (Le 2019), enhanced formation evaluation (Posenato-Garcia 2020) and resistivity inversion (Li 2019).

I discuss three generations of AI and how these have been applied in petrophysics and elsewhere.

AI requirements

AI systems only require two things – The Data and a 'Fitness Function'. In addition, AI software should run with the minimal of human intervention. There are no parameters to pick or cross-plots to make.

Data

All AI generations have access to n-dimensional data, where n is the number of logs or types of data, usually recorded against depth or time. These could include:

Electrical logs - GR, Rhob, caliper, drho etc.
Core data - porosity, core Sw, SCAL etc.
Depth - measured and TVDss.
Gas - chromatography data
Drilling data - ROP, Dexp etc.
NMR - T1 & T2 spectral distributions

Fitness Function

This is the goal or criteria that the AI attempts to solve. It is called the Fitness Function because of its analogy to 'Survival of the Fittest', a phrase that originated from Darwinian evolutionary theory as a way of describing the mechanism of natural selection in his book 'On the Origin of Species' (Darwin 1859). If the AI makes progress to attaining its goal, the Fitness Function ensures the algorithm survives and propagates. If it fails to make progress the algorithm is ignored and becomes extinct.

Software

The software used in the case studies discussed in this paper include C++, MATLAB®, FORTRAN and Geolog® loglan code.

The mathematical techniques used by the AI include neural networks, genetic algorithms and fuzzy logic, as described in this paper and by Cuddy (2013) and Brown (2000).

First generation AI

These are 'Expert' or 'Rule based' systems consisting of computer algorithms. An algorithm being a set of unambiguous instructions, represented mainly as 'if–then' rules, that a computer can execute. These AI algorithms were developed by consulting experts in the relevant field. For instance, the first computers that could play chess involved attempting to write algorithms that replicated the logic used by the chess Grand Masters. Alan Turing was the first to publish a program that was capable of playing a full game of chess and in 1951 his colleague, Dietrich Prinz, wrote the first actual automated chess playing program. The first major application of heuristic programming and the application of expert systems to solve real world problems was the DENDRAL Project (Lindsay 1993).

In petrophysics, Expert Systems enables the log analyst to perform complex analyses which, in the past, could only be done with the assistance of a human expert. An example of a 'Rule' determined by an 'Expert' is:

IF GR > 90 GAPI AND φ < 3 P.U. THEN Matrix = Shale

In 1997, IBM's supercomputer called Deep Blue, beat chess Grandmaster Garry Kasparov, using a search method called "alpha-beta pruning" (Knuth 1975; Hsu 1995). At the time this win was seen as a sign that AI was catching up to human intelligence, by defeating one of humanity's great intellectual champions. However, many commentators played down the AI significance by saying Kasparov was merely defeated by brute force. They said it was like comparing a farmer with a tractor. The tractor may be more powerful but lacks the 'insight and creative flare' of the farmer.

Second generation AI

Second generation AI, often called Machine Learning, involves giving the program access to the data. Rather than the AI being given rules developed by an expert the AI is expected to discover these rules itself. It does this by using an iterative loop shown in Fig. 1.

The computer randomly changes the variables in code and checks if it solves the problem better, as defined by the 'Fitness Function'. Most of these changes will fail and will be ignored, but as the computer is capable of making thousands of iterations per second it will, by chance, progressively improve itself and 'evolve' the best answer.

In this paper, five case studies, are discussed, where second-generation AI has solved petrophysical problems:

- Evolution of shaly water saturation equations
- NMR T1 & T2 spectra analysis
- Prediction of shear velocities
- Litho-facies and permeability prediction
- The log quality control and repair of electrical logs

Continuing with the chess comparison, in 2017 an AI program called AlphaZero (Zhang 2020), using neural networks and set of rules which search for hyperparameters, learned chess by playing itself (Silver, 2017). Within 24 h of training, this program achieved a superhuman level of play, convincingly beating all first-generation AI chess programs. Cheating by humans, in chess tournaments, can be a serious problem if players have concealed access to a chess computer. However, as chess companions demonstrate their skill by having an excellent memory of a vast number of past games, one-way to detect cheating is to look for 'insight and creative flare'.

Case study 1. shaly water saturation equation

Water saturation is used with porosity to determine the hydrocarbons initially in place in a reservoir and to select the best production scenario. The derivation of water saturations from resistivity logs involves inaccuracies due to the difficulty of measuring formation resistivity and the errors inherent in shaly water saturation equations that invert resistivity to water saturation.

We use second-generation AI to evolve a shaly water saturation equation for specific fields by calibrating to core water saturation measurements. This method derives the form of the shaly water saturation equation and gives an independent estimate of SCAL parameters including the cementation exponent '*m*' and the saturation exponent '*n*'. In the presence of conductive shales, the resistivity measurement will be depressed. Shaly sand equations, such the Indonesia equation, include a component related to shaliness to allow for this. The AI used core for calibration of the Fitness Function and the core water saturations (*Sw*) came from high quality Dean and Stark measurements. The virgin core samples were taken from the centre of the core. The drilling mud was doped to correct for any contamination.

The Fitness Function is defined as 'determine a function $f(\varphi, Sw, Vsh)$ which reconstructs Rt given φ, *Sw*, *Vsh* and at each depth'. The next step
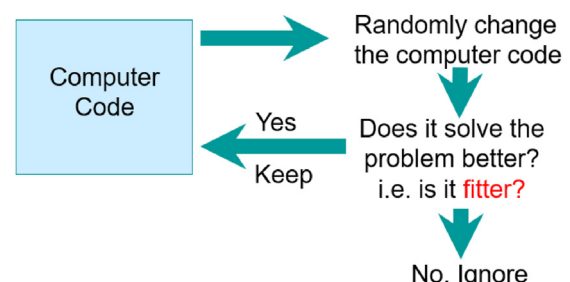

Fig. 1. Second generation AI, machine learning.

is to provide a method for determining the accuracy of a given f (φ, Sw, Vsh) as a predictor of Rt. The approach we adopted was to sum *absolute errors* in prediction over all depth levels for a given borehole. Mathematically, the problem can be stated as:

$$Minimise_f : \sum_i \left| \left(\frac{1}{Rt_i}\right) - f(\phi_i, Sw_i, Vsh_i) \right| \tag{1}$$

Eq. (1) minimizes this sum. The standard way to do this might be to use least squares rather than absolute values of residuals. As explained later in this paper, we have taken this approach because the borehole data is often noisy and includes many 'outliers'. These can only be removed by extensive manual editing of the data and rechecking of measurements. By using the absolute value of residuals, one diminishes the effect of noise and outliers, and produces more appropriate predictor functions.

The AI was constructed as follows. An initial population of individuals is picked randomly in the solution space. Each individual has randomly chosen constants m, n, b and c. The fitness criterion of each of these individuals is determined by Eq. (1). The best existing algorithm for minimizing Eq. (1) starts with a randomly generated f and uses local search by mutating the coefficients one at a time. The mutation range is initially set very high to ensure a complete search of the solution space. After several generations, a pool of individuals is selected, by linear ranking, for mutating and cloning. Mating is achieved by coefficient merging. Some of the best individuals are cloned to add more individuals, where solutions are most promising. After several generations, the percentage change in mutated coefficients is gradually reduced. The algorithm stops when the percentage improvement in evaluation reaches a predefined lower limit, or a maximum number of iterations has been reached.

In genetic terms, each chromosome is a vector of length 4. The alleles are floating point values that represent the constants $m, n, b,$ and c. The initial population is generated by creating chromosomes, with random floating-point numbers for the constants. The alleles are modified by multiplication by a value randomly picked from the range 0.8–1.2. This range decreases in value as the number of generations increases. This provides a method that allows the search to become more local towards the end of the algorithm as better solutions emerge.

This AI technique was applied to a complex Middle East carbonate field which needed a bespoke shaly water saturation equation. AI was used to 'evolve' an ideal equation, together with field specific saturation and cementation exponents. A typical well is shown in Fig. 2.

The AI produced the following equation:

$$\frac{1}{Rt} = \frac{\varphi^m Sw^n}{Rw} + bVsh^c \tag{2}$$

Where:

Sw = water saturation
φ = porosity
Rt, Rw = resistivities
Vsh = shale volume
b, c = constants

Special Core Analysis from AI gave a cementation exponent (m) of 2.214 and a saturation exponent (n) of 1.751. The results are shown in Fig. 2.

The core water saturations are shown as the red circles in track 4, and the predicted water saturations computed using Eq. (2) are shown as the continuous blue curve. As these wells were used to calibrate the water saturation equation, it is not unexpected that the average predicted water saturations agree with the average core derived and predicted water saturations. However, there is a good match between core and predicted saturations across the range of water saturation values. This match is also independent of the shaliness content. These observations suggest that Eq.
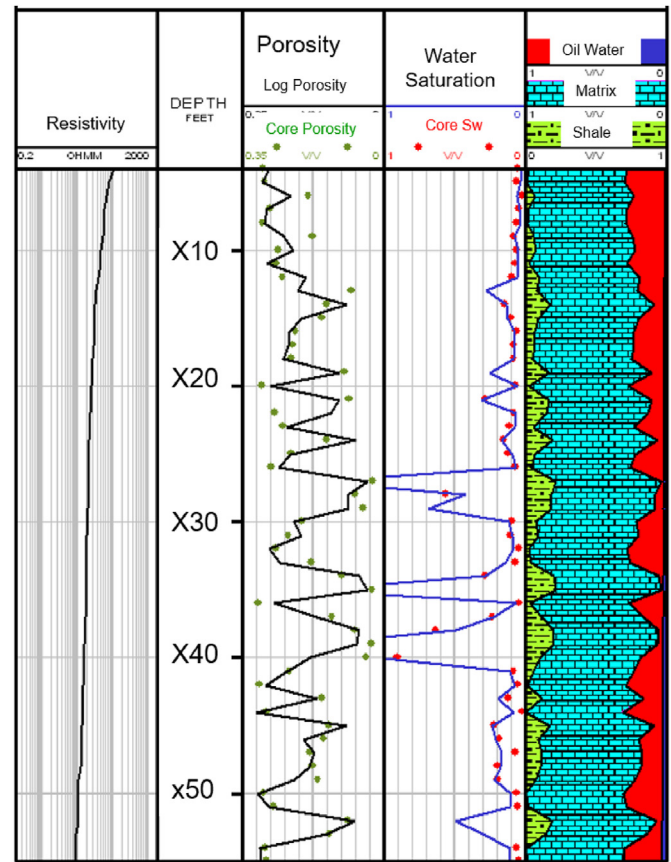

Fig. 2. Using AI to derive a Shaly Sand Equation.

(2), together with its associated parameters, provide a good method of computing water saturations in the shaly carbonate from the resistivity log.

Using core water saturations, avoids the empirical problems of deriving SCAL parameters at reservoir conditions in the laboratory. This AI technique is dependent on the quality of the core water saturations, which is mitigated by ensuring one of the coring objectives is to provide quality core saturations.

Case study 2 NMR pattern recognition

This case study describes the use of AI to optimise production from a complex gas reservoir which had long been recognised but was not developed due to a variety of technical challenges including the thin-bedded nature of the sediments and the presence of both mobile and immobile viscous residual oil. The oil is a highly viscous liquid, which if produced, could block production tubing due to the shallow depth of the reservoir and associated low pressures. To successfully produce dry gas, identification of both oil and gas zones was necessary to enable gas zones to be perforated, and oil zones to be excluded.

During the development drilling campaign, the reservoir was appraised by using a formation evaluation programme designed to address the presence of oil within the thinly bedded reservoir. In conjunction with core data and high-resolution electrical logs (incl. MSFL), nuclear magnetic resonance tools (NMR) were used to identify and avoid perforating zones with higher oil saturations. The main objective of running an NMR tool was to help identify intervals of high residual oil saturation. These intervals were then confirmed at the wellsite by selectively using the MDT-LFA (Modular Formation Dynamics Tester-Live Fluid Analyser) to confirm the composition of the fluids and check for mobility. Borehole gas chromatography data was also used to confirm these interpretations.

The NMR results were calibrated to high quality Dean and Stark oil, gas and water saturations and high-resolution wireline logs from the cored wells. AI techniques applied to the NMR curves enabled the prediction of fluid types in uncored sections and oil mobility was predicted throughout the reservoir section. Conventionally, Dean and Stark water saturation measurements are made in reservoir zones where oil-based drilling fluid has been used. The development wells were cored using a water-based mud system and the objective of performing Dean and Stark was to infer residual oil saturation (Sor) from the gravimetric mass balance of the residual fluids present. Where possible the measurements were performed on a one per foot basis.

Conventional NMR interpretation techniques, like in permeability estimation, use very little of the wealth of information contained in the T_2 distribution. The Coates method (1991) compares two areas, Free Fluid Index, and Bound Volume Irreducible, beneath the T_2 distribution separated by an arbitrary cut-off whereas the Schlumberger Dowell Research (SDR) method only uses a single piece of information, the logarithmic mean T_2.

Rather than using a single value from the T_2 distribution, such as the ratio of areas or the logarithmic mean value, AI analyses the entire shape of the T_2 distribution, attempting to unlock any that could be related to the pore fluid type. This technique is similar to AI facial recognition techniques which are capable of identifying a person from a digital image.

The results of this study are shown in Fig. 3.

The T1, T2 (short), T2 (long) spectra are shown in tracks 5, 6 and 7 of Fig. 3 respectively. AI analysis of these spectra gives a good prediction of the oil and gas intervals as shown in tracks 8 and 9 respectively. These results were confirmed by the MDT-LFA and the borehole gas chromatography data shown in track 4. This analysis was then used to design the perforation program that avoided oil bearing sands. Consequently, using AI, operational decisions were made quickly to choose perforation
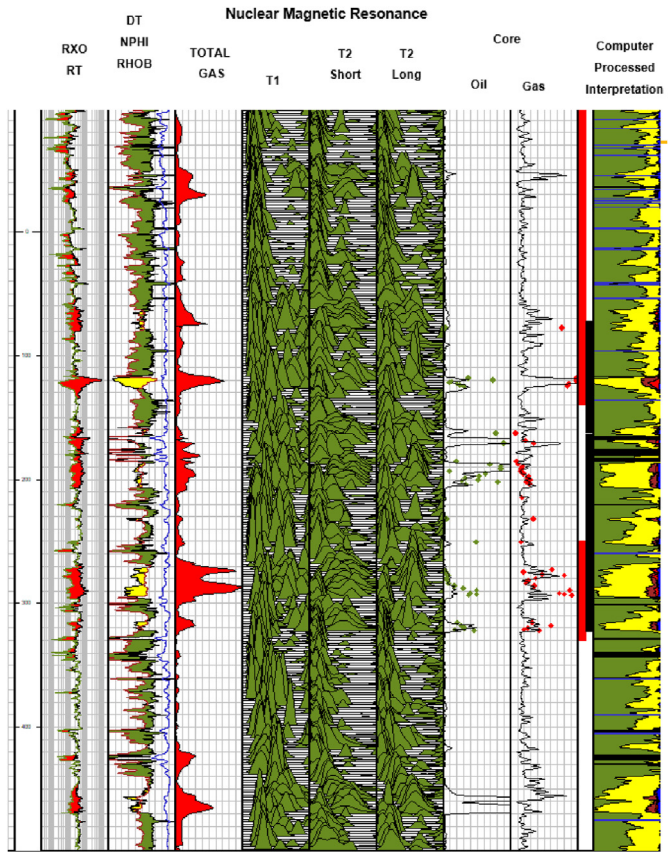
intervals. The perforation strategy on all development wells appears to have been successful, and no oil was produced during well testing. Gas sands that lie below the oil sands were also perforated successfully. Further information of this technique is described by Cuddy (2004).

Case study 3. shear velocity prediction using AI

Shear velocities are useful for enhanced seismic evaluation and to understand and control wellbore stability whilst drilling. AI was applied to a North Sea field consisting of 30 wells, where only 5 had recorded borehole shear velocities. The objective of this study was to populate all 30 wells with shear velocity data in order to improve the reservoir modelling and aid future well drilling.

The available logs for this field included the conventional electrical logs, drilling and gas chromatography data. The gamma-ray, compressional and shear electrical logs are shown in Fig. 4. The shear log, show in blue track 3, is missing above 13,200 ft. Consequently, an additional objective for this well is to provide the shear log for the gap in this well.

The AI asserted that there was a continuous functional relationship between the log values and shear velocities and attempted to evolve this relationship based on or calibrated to the data from the five wells with good shear velocities. The AI provided the functional form of the equation as well as the constant parameters of the relationship. Analysis using AI consists of two parts, the evolution of an equation relating Dts (Vs) to the electrical logs and the use of this equation to predict Dts throughout the field. Statistical analysis suggested that the shear velocities are
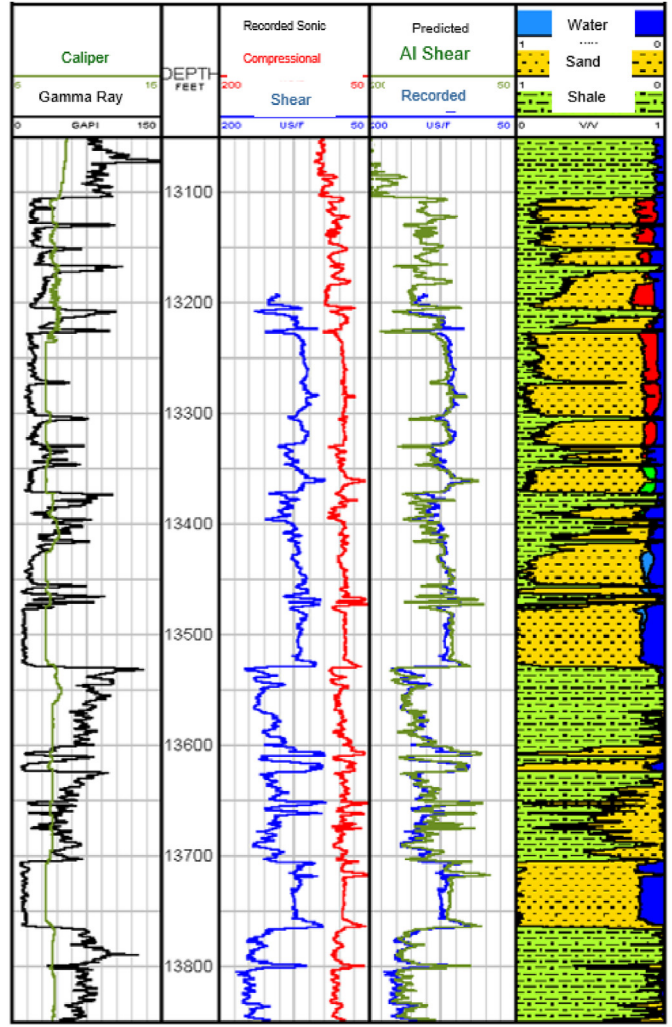

Fig. 3. NMR Spectra Analysis using AI.

Fig. 4. Shear Velocity Prediction using AI.

mainly related to compressional transit time *Dtc*, total porosity *Phit*, and the mineralogy ratio *Mlith*.

The objective of the AI was to construct empirically a function *f* (*Dtc,Phit, Mlith*) that predicts shear velocities at each depth *i*, given *φ*, *Phit* and *Mlith* at each depth. Therefore, we were searching for an appropriate function of the form:

$$Dts = f(Dtc, Phit, Mlith) = (aDtc^b) \bullet_1 (cPhit^d) \bullet_2 (eMlith^g) \bullet_3 h \quad (3)$$

where $\bullet_1$, $\bullet_2$, $\bullet_3$, etc. represent the algebraic operators addition and multiplication; a, c, e, and h are unknown constants; and b, d, and g are unknown constant exponents.

The AI Fitness Function for this application is 'Determine a relationship so that the predicted shear velocities are as close as possible to log derived shear velocities', which is expressed mathematically as:

$$\underset{f}{Minimise} : \sum_i |Dts_i - f(Dtc_i, Phit_i, Mlith_i)| \quad (4)$$

The approach we adopted was to sum absolute errors in prediction over all depth levels for a given borehole. Eq. (4) minimizes this sum. The conventional way to do this is to use least squares rather than absolute values of residuals. The reason for the approach that we have taken is that the borehole data are noisy and include many 'outliers.' These can only be removed by extensive manual editing of the data sets and rechecking of measurements. By using the absolute value of residuals, one diminishes the effect of noise and outliers and produces more appropriate predictor functions.

The genetic algorithms were constructed as follows. An initial population of individuals is picked randomly in the solution space. Each individual has randomly chosen constants a, b, c, d, e, g, h and operators $\bullet_1$, $\bullet_2$, $\bullet_3$. Eq. (4) determines the fitness criterion of each of these individuals. The best existing algorithm for minimizing this equation starts with a randomly generated *f* and uses local search by mutating the coefficients one at a time or flipping the operator between an addition and a multiplication. The mutation range is initially set very high in order that the individuals search all of the solution space. After a number of generations, a pool of individuals is selected by linear ranking for mutating and cloning. Mating is achieved by coefficient merging. Some of the best individuals are cloned to add more individuals where solutions are most promising. After a number of generations, the mathematical operators are fixed and the percentage change in mutated coefficients is gradually reduced. The algorithm stops when the percentage improvement in evaluation reaches a predefined lower limit or a maximum number of iterations have been reached.

Each chromosome is a vector of length 10. Three alleles are binary integer values that represent the mathematical operators $\bullet_1$, $\bullet_2$, $\bullet_3$. The rest of the alleles are floating point values that represent the coefficients a, b, c, d, e, g, h. The initial population is generated by creating chromosomes with random binary numbers for $\bullet_1$, $\bullet_2$, $\bullet_3$ and random floating-point numbers for the coefficients a, b, c, d, e, g, h. If the allele represents the operator $\bullet$, its value is binary, and it will be switched. If the allele represents one of the real variables, it will be modified by multiplication by a value randomly picked. The AI evolved an equation for the prediction of Dts. To improve the derived equation's predictive ability, separate coefficients a, b, c, d, e, g, h were separately determined for each zone in the reservoir.

The comparison between recorded and predicted Dts, shown in track 4 of Fig. 4 are extremely good. Surprisingly, for the thin-bedded intervals, where there is a mismatch between the measured and predicted velocities, it is thought that the measured log values are incorrect. This is because the vertical resolution of the shear velocity measurement tool prohibits a correct measurement in thin beds. However, the AI had access to all measurements, including ones with better vertical resolution than the shear velocity tool. As a result, the measurements with good vertical resolution pick the facies type before the techniques predict the appropriate shear velocity, and the predicted shear velocity in the thin beds is

likely to be a more appropriate value than that actually measured by the shear velocity tool. The predicted Dts therefore appear to repair locally poor Dts readings and are therefore an improvement on the borehole acquired electrical log.

Case study 4. permeability prediction using AI

Knowledge of permeability is important in determining the well completion strategy and the resulting productivity. The problem with permeability prediction is derived from the fact that permeability is related more to the aperture of pore throats rather than pore size, which logging tools find difficult to measure. Determining permeability from well logs is further complicated by the problem of scale, well logs having a vertical resolution of typically 2 feet compared to the 2 inches of core plugs. AI has been used in several fields to obtain better estimates of permeability compared to conventional techniques.

A case study in permeability prediction in shown in Fig. 5 for a large North Sea field with 15 cored wells.

The gamma-ray (GR) and differential caliper logs are shown in track 1. The standard porosity logs; density (RHOB), neutron porosity (NPHI) and sonic compressional velocity (DT) are show in track 3. Track 7, on the far right, shows a computer processed interpretation for this well, where the shale volume is shown in green, sandstones in yellow, oil in red, the bulk volume of water in dark blue, coal beds in black and calcite stringers in cyan.

The AI Fitness Function for this application is; 'Determine a relationship so that the predicted permeabilities are as close as possible to core derived permeabilities'. The AI first determined the litho-facies type as shown in track 6, and then based on this result predicts the permeability shown as a continuous black line in track 4. This compares favourably with the core derived permeability shown as green dots also in track 4. The prediction confidence is shown in track 5, with the highest confidence shown in red and the lowest in blue.

For comparison, permeability was also predicted using conventional least squares linear regression. These predictions were compared as
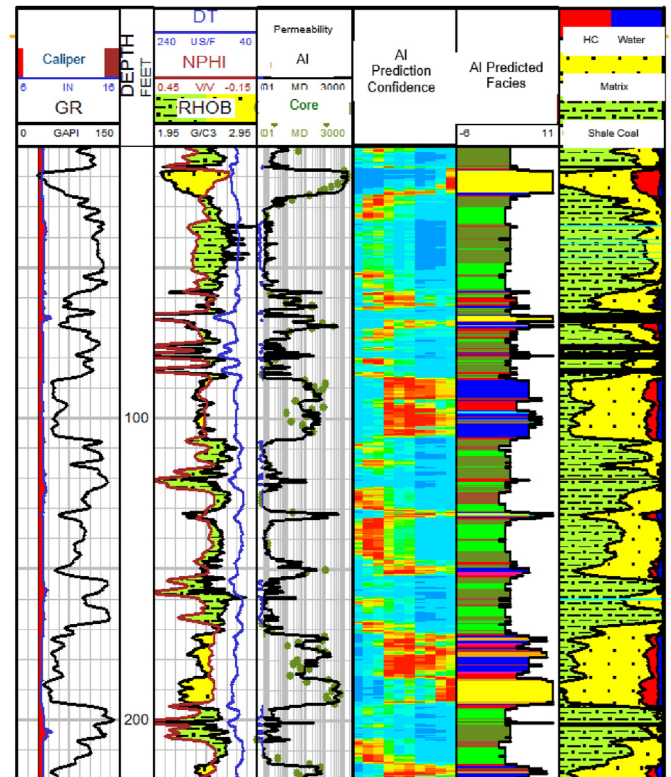

Fig. 5. Permeability Prediction using AI.

shown in the frequency distribution histograms (Fig. 6).

The left-hand histogram shows the core permeability distribution for this field. The colours represent each of the 15 wells. Notice that the permeability distribution is bimodal. The AI prediction is shown as the centre histogram and shows a similar bimodal distribution. The least squares linear regression, shown on the right, shows a good prediction of the average permeabilities but is poor at the upper and lower ends. The upper and lower permeability predictions are important, in the reservoir model, as these represent that barriers to flow and conduits to production respectively. Whereas least squares regressed towards the mean, AI preserved the dynamic range of the core permeability data. An incorrect permeability distribution will not upscale correctly in the 3D reservoir model (Cuddy 2013).

Case study 5. the quality control and repair of electrical logs

Borehole electrical logs are acquired in a difficult environment, often at high temperatures and pressures. Although most modern electrical tools are designed to compensate for limited borehole washouts and rugosity, virtually every well contains sections of log with poor or un-acceptable quality. In addition to poor logs there are often sections where the measurement has completely failed due to telemetry problems with the surface equipment or because of tool failure due to adverse conditions. These problems have increased in recent years with the introduction of Logging Whilst Drilling (LWD). The oil industry is to be commended for the development of sensors that take measurements whilst the well is being drilled. These real-time measurements are a real benefit to the industry as they enable quick decisions, but these measurements are taken in an extremely adverse environment, which includes drill string vibration and high-pressure drilling mud. Wireline and LWD measurements are further compromised by calibration errors, which occur because of human error or through electronic tool drift due to temperature. The log quality control (LQC) and repair of electrical logs is therefore essential before petrophysical analysis can take place. This is a very useful exercise even when the LQC merely confirms that the logs are good.

The LQC and repair of electrical logs is based on the premise that all logs are related. A skilled petrophysicist can verify an anomaly on one electrical log simply through visual comparison with other curves. The presentation scales of printed or digital electrical logs are chosen based on their relationship to porosity. Consequently, the logs will deflect to the left or right at the same depths. If the logs do not respond in this manner it prompts further investigation. For instance, the density log may measure extremely low densities when the tool is separated from the bore-hole wall, causing it to read the mud density, or because of the presence of a coal bed. In the former case, the log needs repair and in the latter, the log is correct. The petrophysicist would normally check the sonic compressional velocity, gamma-ray reading and resistivity log at the same depth in the reservoir to confirm either interpretation. AI is used in a similar manner to uncover the relationships between all electrical logs so that anomalies can be identified, and the correct log can be predicted.

Fig. 7 shows an example from the North Sea Heather field explains the process. Recent infill wells in the Heather field suffered from bad sections of LWD log data in the Brent reservoir. Well geosteering required logging in sliding mode at times, which tends to degrade log quality because the tool is not rotating and does not 'see' all around the borehole. Unexpected overpressure was encountered in parts of the reservoir, particularly Ness Formation sands and shales, caused by injection water. To control pressures, mud weight was increased to very high densities. This, in turn, caused differential sticking and hole washouts. AI was applied to repair defective density log curves as shown in Fig. 7.

The recorded density log, shown in green, is reading the mud density around X80 feet. The predicted density log, shown in red, was derived by from all the other logs as if the density log was not run in this well. Although the LQC is automatic the final step of replacing the poor log by the predicted log is currently left to the petrophysicist. Further information on the use of the technique in the Heather Field is discussed in Kay (2002).

The advantages of using AI in petrophysical analysis

AI doesn't require prior knowledge of the petrophysical response equations and is self-calibrating. We just give it the data. There is very little user intervention as there are no parameters to pick or cross-plots to make. AI can work with an unlimited number of electrical logs, core and gas chromatography data; and don't 'fall-over' if some of those inputs are missing. It is not a Black Box as it provides insights into how it makes predictions.

Least squares regression

AI avoids the problem of 'garbage in, garbage out' (GIGO), by ignoring noise and outliers as shown by Fig. 8.

The least squares regression line is the black line that makes the red vertical distances from the data points to the regression line as small as possible by minimizing the sum of squares of the errors as shown in Fig. 8. This type of regression is often referred to as being 'undemocratic'
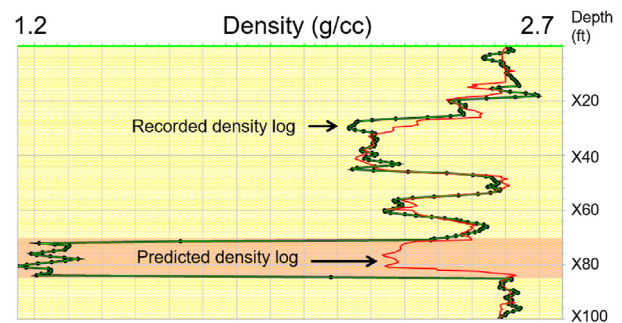

Fig. 7. The Quality Control and Repair of Electrical Logs using AI.
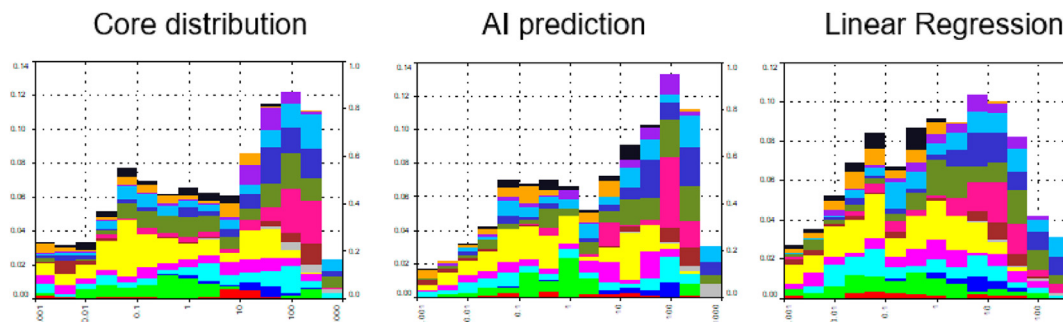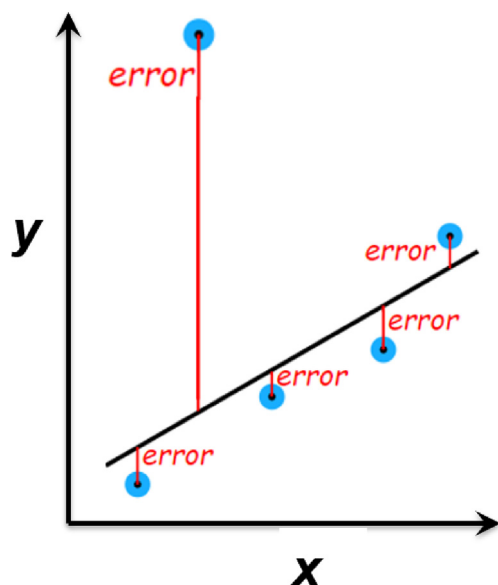

Fig. 6. Core and predicted permeability distributions.

Fig. 8. The problem with least squares regression.

Fig. 9. Nearest Neighbours in n-dimensional Data Space.

as outliers unfairly influence the result. For example, a point 10 times further from the line has a 100× the weighting. It is exceedingly difficult to manually remove these and if attempted would introduce human bias. Outliers may be valid data such as coal beds, calcite stringers or fractures. Our AI programs keep all data points and minimise the linear distance (Mean Absolute Error) rather than the squared distance. We find any random noise is often swamped by valid data. It is important to stress that AI is only capable of solving problems where it is given sufficient good quality data.

The mathematical techniques used by the AI include neural networks, genetic algorithms, random forests and fuzzy logic. The petrophysical case studies described in this paper use fuzzy logic, for the reasons I will explain.

What is fuzzy logic?

When the computers arrived with their machine-driven binary system, Boolean logic was adopted as the natural reasoning mechanism for them. Conventional logic forces the continuous world to be described with a coarse approximation; and in so doing, much of the fine detail is lost. Classical logic is useful, but we are left with a feeling that there is something missing. The real world is not made up of bivalent blacks and whites; there is a grey scale out there. By only accepting the two extreme possibilities, the infinite number of possibilities in between is lost. Reality does not work in black and white, but on a grey scale. Not only does truth exist fundamentally on a sliding scale, it is also perceived to vary gradually because of the uncertainties in measurements and interpretations. Hence, a grey scale can be a more powerful tool compared to just using the two end points. Once the reality of the grey scale has been accepted, a system is required to cope with the multitude of possibilities. Probability theory helps quantify the greyness or fuzziness. Fuzzy logic asserts that any interpretation is possible, but some are more likely than others (Cuddy 2013).

The AI has access to n-dimensional data, where n is the number of logs or types of data, usually recorded against depth or time. One method used by AI for pattern recognition, classification and regression is the k-nearest neighbours algorithm (k-NN). Fig. 9 shows petrophysical data points in green and their association with 3 red litho-facies, (e.g. sand, shale or carbonate). The k-NN algorithm assumes that similar things exist in close proximity and are near to each other in n-dimensional data space. Determining which litho-facies for each data point can be determined by its proximity in this n-dimensional space. It is therefore necessary to
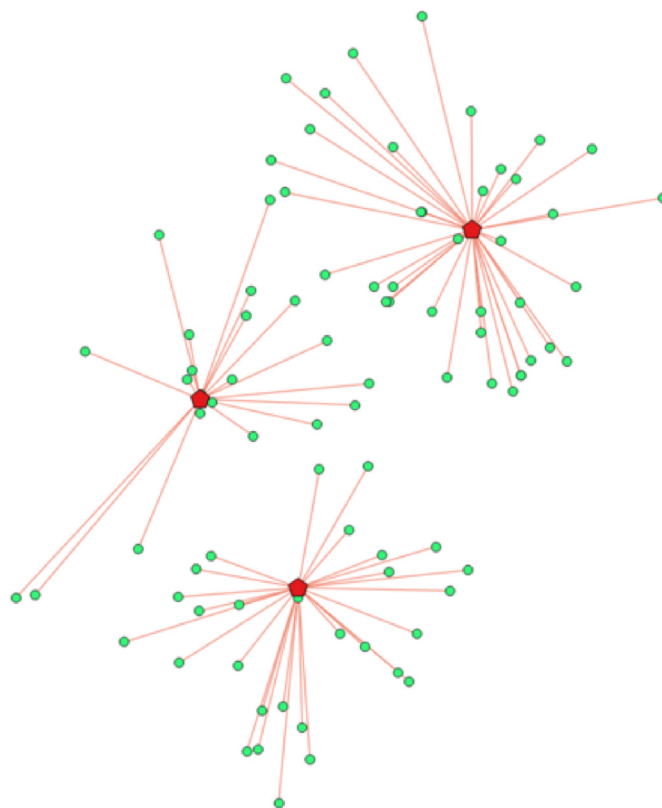
determine the distance between points. The straight-line distance, called the Euclidean distance, is used. In addition, fuzzy logic weights these lines depending on the likelihood of the association. For instance, if the gamma-ray is highly correlated with shaliness, this vector will have more influence on the AI's decision compared to say the caliper reading at the same depth.

Narrow AI VS. General AI

Most AI programs developed to date have Narrow AI which can handle just one particular task. They are good at petrophysical analysis, playing chess, forecasting the weather and even ordering coffee for you. General AI is where we are going, where like humans, can do many things. They can do petrophysical analysis *and* play chess. General AI learns from one specialist area and applies it in another. For instance, a petrophysicist who also plays chess may use chess decision tree analysis in petrophysical analysis. They will, by definition, be genuinely creative with the ability to produce something original. General AI is True AI and will be able to 'think out of the 'box'. Third generation AI is making this possible.

Third generation AI

AI programs currently being developed include ones where their machine code evolves using similar rules used by life's DNA code.

Charles Darwin (1859) explained evolution using natural selection. We now know this can be explained by changes to our DNA code through mutation and sexual selection. The DNA code a language of 4 characters represented by 4 nucleotides, 'A' paired with 'T' and 'C' with 'G' represented by the 4 colours in Fig. 10. The evolution loop, shown in Fig. 10, can take millions of years.

Because of the spectacular success of this process, as demonstrated by evolution of life on Earth, some third-generation AI developers are attempting to mimic it as shown in Fig. 11. Whereas the DNA code is a
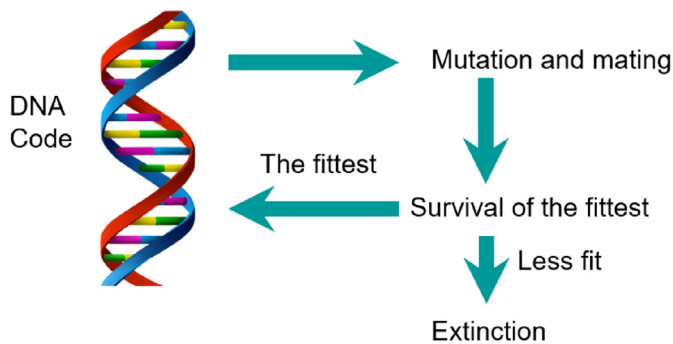
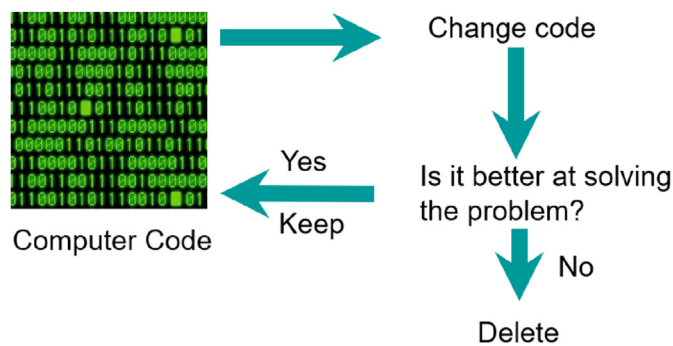**Fig. 10.** Evolution of DNA code by natural selection.

Fig. 11. Third Generation AI, computer code evolution.

language of 4 characters, computer code is a language of 2 characters, the binary digits '0' and '1'.

The mutation and mating used by the evolution of DNA code is replicated by similar changes to the computer code. Mutation is simulated by randomly changing groups of digits by their opposites. Blocks of code are moved throughout the computer's machine code. Further to mimic nature, third generation AI can have 100s of algorithms running in parallel. Every 1000 iterations (generations) the 'fittest' are select to 'mate' following exactly the rules used by life to create new algorithms. The vast amount of these changes will have little success at fulfilling the Fitness Function and will be deleted. However, the AI algorithm will steadily 'evolve'. Unlike most first and second-generation AI the manipulation of machine code can give unpredictable or unwanted results. Although unlikely, here lies the clear and present danger of AI.

The dangers of AI

This section are the personal observations and opinions of the author.

It is important to reiterate that AI has a goal determined by the *Fitness Function* and runs with little user intervention.

King Midas and his golden touch

King Midas, in Greek mythology, was granted his wish that everything he touched into gold. He didn't realise that this included his food and his children. Similarly, an ill-conceived Fitness Function may give unexpected results. For instance, an AI asked to design and produce a vaccine for the COVID-19 virus which 'prevents transmission between humans' could solve the problem by killing humans. A poorly stated Fitness Function is called a Midas Function because of its potential danger.

The Sorcerer's apprentice

The Sorcerer's Apprentice, in Johann Wolfgang von Goethe famous

poem, uses magic to get a broom carry water for him. Unfortunately, it runs-away and nearly drowns him. In a similar fashion, if an AI runs away and becomes unstoppable it could create grave dangers.

I now describe an example from petrophysics where a poorly worded Fitness Function could cause the AI to runaway.

Example of runaway AI

Consider AI being used to history match a reservoir model with the electrical logs, at the well locations, as shown by Fig. 12.

A possible Fitness Function could be 'get the best possible match as fast as possible'. By trial and error, the computer would be expected to evolve a fast history match. Any endeavour succeeds faster if you increase its resources. For instance, in a battle it is better to increase the resources that you give to your forces rather than merely asking them to fight harder. Similarly, history matching will be faster if you give the computer more resources, specifically more computing power. A human programmer or hacker may co-opt the resources of other network computers to achieve a faster speed. There is no reason why AI couldn't also do this. If AI achieves this 'by accident', there is nothing to stop it doing it again and again via the internet, eventually consuming all the world's computing power. As it is now on the internet will be impossible to switch off.

Evolution takes millions of years, whereas the computer makes millions of iterations per second. Consequently, the any AI program may 'accidently' start improving exponentially. A supercomputer isn't required to do this. Your laptop could do it. An elaborate AI computer program isn't required. Only one that can update its own machine code goal guided by an ill-judged Fitness Function. This is known as the 'Singularity' where artificial intelligence becomes uncontrollable and irreversible. The chances of this happening may be as remote as life spontaneously occurring, but the AI has only to do this once. It is not known how to switch off or stop computers with runaway evolution.

The dangers of AI beyond petrophysics

Several commentators, with in depth knowledge of AI, have expressed grave concerns about runaway AI. These are listed and referenced for those interested in examining these concerns, as detailed discussion of these expert's views is not necessary for understanding the 5 petrophysical case studies.

University of Cambridge Professor Stephen Hawking was concerned that AI could 'take off' on their own, modifying themselves and independently designing and building ever more capable systems. He forecast humans, bound by the slow pace of biological evolution, would be tragically outwitted and wrote, 'Efforts to create thinking machines pose a threat to our very existence'. (Hawking 2014).

Microsoft co-founder, Bill Gates has said he doesn't understand people who are not troubled by the possibility that AI could grow too strong for people to control. He notes that AI dangers could include speech synthesis for impersonation; analysis of human behaviours, moods and beliefs for manipulation; automated hacking and physical weapons like swarms of micro-drones. (Gates 2015).

University of Oxford Professor and Director of the Future of Humanity Institute, Nick Bostrom, has said that AI could bring about human extinction by 'hijacking political processes, subtly manipulating financial markets, biasing information flows, or hacking human-made weapons systems'. He writes 'We're like children playing with a bomb'. (Bostrom 2014).

SpaceX founder, Elon Musk, has said humanity is merely a 'biological boot loader for digital super intelligence', and unequivocally stated, 'Mark my words - A.I. is far more dangerous than nukes'. He is quoted as saying 'AI needs safety measures before something terrible happens'. (Musk 2017).
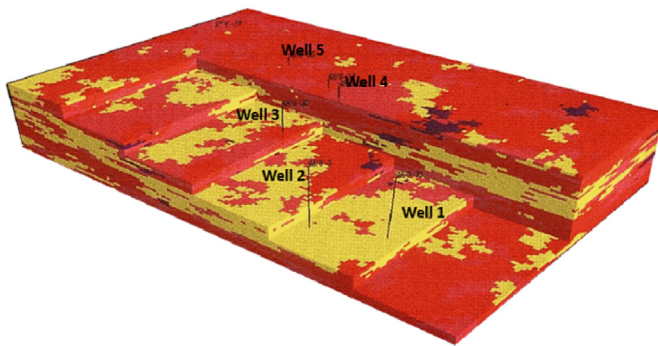
Fig. 12. Using AI for history matching a reservoir model.

Solution to the dangers of AI

AI programs pose considerable dangers far beyond the oil industry. First generation AI (expert systems) shouldn't be a problem as the AI is constrained by the hard-wired lines of computer code. Second generation AI (fing systems) can be dangerous if they can access other computers via networks or the internet. Third generation (evolutionary systems) AI pose the greatest danger because of their limitless and unpredictable nature.

A 'risk assessment' is essential on all AI programs so that all hazards and risk factors, that could cause harm, are identified and mitigated. The possibility of a runaway AI is remote, but the consequences could be greater than pandemics, climate change and nuclear proliferation.

A risk assessment need only take a few minutes. This includes confirming the Fitness Function is clear and unambiguous, and that the AI cannot runaway by 'escaping' from the computer where it is programmed.

AI programs are potentially dangerous and may be the last thing humans invent.

Conclusions

AI brings many benefits to petrophysics and can make petrophysical analysis easy. AI doesn't require prior knowledge of the petrophysical response equations, it is self-calibrating. AI ignores noise and outliers. There is very little user intervention as there are no parameters to pick or cross-plots to make. AI programs work with large datasets and isn't a 'black box' as it provides insights into how it makes predictions.

AI can evolve shaly water saturation equations, perform NMR spectra analysis image analysis, predict shear velocities and permeability. AI is especially good at the log quality control and repair of electrical logs.

AI can be extremely dangerous, especially if it becomes a runaway without control. It is strongly recommended that all AI program development should include a risk assessment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Nomenclature

Abbreviations

AI	Artificial Intelligence
Dexp	Drilling exponent
DNA	Deoxyribonucleic Acid
Drho	Delta Rho
Dtc	Compressional transit time
Dts	Shear transit time
GR	Gamma-Ray
LFA	Live Fluid Analyser
LQC	Log Quality Control
LWD	Logging Whilst Drilling
MDT	Modular Formation Dynamics Tester
Mlith	Sonic density lithology factor
NMR	Nuclear Magnetic Resonance
Nphi	Neutron porosity
Phit	Total porosity
Rhob	Bulk Density
ROP	Rate of penetration
Rt	Formation Resistivity
Rw	Water Resistivity
Rxo	Invaded zone resistivity
SCAL	Special Core Analysis
Sw	Water Saturation
T1	NMR T1 relaxation time
T2	NMR T2 relaxation time
TVDss	True Vertical Depth subsea
Vs	Shear velocity
Vsh	Shale volume

Symbols

φ	Porosity
m	Cementation exponent
n	Saturation exponent

References

Bigoni, Francesco, Pirrone, Marco, Pinelli, Fabio, Trombin, Gianluca, Vinci, Fabio, 2019. A Multi-Scale Path for the Characterization of Heterogeneous Karst Carbonates: How Log-To-Seismic Machine Learning Can Optimize Hydrocarbon Production, pp. 1–9. https://doi.org/10.30632/T60als-2019_Ff.

Bostrom, N., 2014. Superintelligence-Paths. Oxford University Press, Dangers, Strategies.

Brown, D.F., Cuddy, S.J., Garmendia-Doval, A.B., McCall, J.A.W., 2000. The prediction of permeability in oil-bearing strata using genetic algorithms. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing, ISBN 0-88986-292-3, pp. 53–58. July 2000.

Coates, G.R., Miller, M., Gillen, M., Henderson, G., 1991. The MRIL in Conoco 33-1- an investigation of a new magnetic resonance imaging log. In: 32nd Annual Logging Symposium Transactions-Society of Professional Well Log Analysts. Paper DD.

Crevier, 1993. Who Writes 'the Conference Is Generally Recognized as the Official Birthdate of the New Science, pp. 47–49.

Cuddy, S., 2013. Litho-facies and permeability prediction from electrical logs using fuzzy logic. SPE Reservoir Eval. Eng. 3, 319–324. https://doi.org/10.2118/65411-PA.

Cuddy, S., Daniels, G., Lindsay, C., Sands, P., 2004. The application of novel formation evaluation techniques to a complex tight gas reservoir. In: SPWLA 45th Annual Logging Symposium Noordwijk. The Netherlands, June 6–9, 2004.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection. John Murray – London.

Davis, Randall, Shrobe, Howard, Szolovits, Peter, 2002. What is a knowledge representation? AI Mag. 14.

Gates, B., 2015. Microsoft's Bill Gates Insists AI Is a Threat. BBC New, 29th January 2015.

Hawking, S., 2014. 'AI Could Be the End of Humanity' Independent Newspaper, 2nd December 2014.

Hsu, Feng-Hsiung, Campbell Murray Jr., A., 1995. Deep Blue System Overview, pp. 240–244. https://doi.org/10.1145/224538.224567.

Kay, S., Cuddy, S., 2002. Innovative use of petrophysics in field rehabilitation, with examples from the heather field. Petrol. Geosci. 8 (4), 317–325. Dec 2002.

Knuth, Donald, Moore, Ronald, 1975. An analysis of alpha-beta pruning. Artif. Intell. 6, 293–326. https://doi.org/10.1016/0004-3702(75)90019-3.

Le, Thai, Liang, Lin, Zimmermann, Timon, Zeroug, Smaine, Heliot, Denis, 2019. A Machine-Learning Framework for Automating Well-Log Depth Matching, vol. 60, pp. 585–595. https://doi.org/10.30632/Pjv60n5-2019a3.

Li, Hu, Liu, Gang, Yang, Shansen, Guo, Ying, Huang, He, Dai, Mingzong, Tian, Yuanshi, 2019. Automated Resistivity Inversion and Formation Geometry Determination in High-Angle and Horizontal Wells Using Deep Learning Techniques, pp. 1–11. https://doi.org/10.30632/T60als-2019_Jjjj.

Lindsay, Robert, Buchanan, Bruce, Feigenbaum, Edward, Lederberg, Joshua, 1993. DENDRAL: A Case Study of the First Expert System for Scientific Hypothesis Formation, vol. 61. Artificial Intelligence. https://doi.org/10.1016/0004-3702(93)90068-M.

Minsky, Marvin, 1961. Steps toward artificial intelligence. Proc. IRE 49, 8–30. https://doi.org/10.1109/JRPROC.1961.287775.

Musk, E., 2017. Regulate AI to Combat 'existential Threat' before Its Too Late. Guardian Newspaper, 17th July 2017.

Newell, A., Simon, H., 1963. GPS, a program that simulates human thought. Comput. Thought 279–293.

Posenato Garcial, Artur, Jagadisan, Archana, Hernandez, Laura, Heidari, Zoya, Casey, Brian, Williams, Richard, 2020. enhanced formation evaluation in the permian basin using a novel field-scale workflow including wells with missing data. SPE Reservoir Eval. Eng. 23 https://doi.org/10.2118/201185-PA.

Turing, A., 1950. Computing Machinery and intelligence. Mind LIX (236), 433–460. https://doi.org/10.1093/mind/LIX.236.433. ISSN 0026-442.

Winston, Patrick, 1976. The Psychology of Computer Vision. https://doi.org/10.1016/0031-3203(76)90020-0. Pattern Recognition. 8. 193.

Zhang, Hongming, Yul, Tianyang, 2020. AlphaZero. https://doi.org/10.1007/978-981-15-4095-0_15.

Steve Cuddy is a retired Principal Petrophysicist from Baker Hughes. He holds a PhD in petrophysics at Aberdeen University. He also holds a BSc in physics and a BSc in astrophysics and philosophy. He is the inventor of the Fractal FOIL Function that accurately describes the distribution of fluids in the reservoir model. He writes AI software and has 45 years industry experience in petrophysics. In recognition of outstanding service to the SPWLA, Steve was awarded the Distinguished Service Award in 2018.