A Guide to the Practical Use of Neural Networks

Brouwer, Friso C. G.
dGB Earth Sciences
1 Sugar Creek Center BLVD STE 935
Sugar Land, Texas 77478
USA
email: friso.brouwer@dgbes.com

Tingahl, Kristofer
dGB Earth Sciences
1 Sugar Creek Center BLVD STE 935
Sugar Land, Texas 77478
USA
email: kristofer.tingdahl@dgbes.com

Connolly, David
dGB Earth Sciences
1 Sugar Creek Center BLVD STE 935
Sugar Land, Texas 77478
USA
email: david.connolly@dgbes.com

Abstract

Seismic attribute analysis is a great tool to enhance and isolate features related to seismic acquisition, processing, and geology. However, single or primary attributes have two drawbacks that can be addressed by more intelligent work flows. First, seismic attributes may not uniquely identify the seismic feature that is the target of the analysis. For example, assuming that faults are the target of our attribute analysis, a discontinuity attribute highlights any lateral change in the signal, including both incised sedimentary features, and faulting. Second, seismic attributes may reveal some of the target features, but not all. For example, discontinuity attributes will not highlight faults that have small fault offset compared with the seismic resolution. Methods that recombine two or more primary attributes can be used to improve a complete and unique isolation of a target feature in the seismic data. For example, fault detection can be performed by recombining discontinuity and curvature attributes, such that discontinuities attributed to sedimentary structures are suppressed, while low offset faults, represented by seismic flexures are highlighted.

Among other methods, neural networks are one of the most efficient methods to recombine multiple input attributes and achieve a high quality extraction of a target feature or rock property from seismic data. The method is complex and difficult to analyze, and often, the black-box character is cited as a reason to stay clear of this method. However, in specific cases the benefits of using neural networks compared to baseline methods is so large that by far they outweigh any (perceived) negatives, and the neural network work flow is the correct tool to enhance a geological interpretation of seismic data.

The aim of this paper is to "translate" the neural network method from a specialized tool that can only be used in the hand of an expert user, to a general interpretation work flow that can be used by informed (but not specialized) general interpreters. To achieve that, this paper will address the following issues:

- When to choose a neural network work flow (and when not).
- Basic theory of neural networks
- General, but practical, guidelines for designing and training a neural network.
- Methods for quality control and validation of neural network results.
- How to use the neural network results as part of the larger interpretation work flow.

These talking points will be illustrated with actual data examples.

Introduction

How to assess seismic attributes

The approach of using neural networks is very similar to seismic attributes. Therefore, it is useful to first discuss objectives and quality assessment of seismic attributes in general and expand our analysis of neural network work flows within these concepts.

The general objective of using seismic attributes is to isolate and map certain geology-related features in the seismic data. The quality of a seismic attribute can be quantified in terms of completeness and uniqueness of the geological target feature.

As an example, seismic discontinuity attributes (*e.g.*, similarity, coherence, or variance) are often erroneously equated to fault (mapping) attributes. However, these attributes map seismic discontinuities, which besides faults, can be associated with erosional incisions, gas chimneys, mass transport systems, and other geological features, making them geologically non-unique (Fig. 1). From geophysical knowledge, we know that not all faults visible in the seismic data are associated with reflector discontinuities. Faults having low dip-slip relative to the dominant seismic wave length may for instance appear to be seismic flexures, not seismic discontinuities. In that sense, the discontinuity attribute does not map all the faults that a human interpreter sees in the seismic data (Fig. 2). It should be noted that the impact of the geological nonuniqueness and geological noncompleteness of an attribute varies according to the application and is interpretation-dependent. It may be that some artifacts may not be harmful as they are easily recognized by the interpreter or corrected with a simple filter operation. However, some artifacts that may be harmless for a manual interpretation of the attribute may be very harmful to automated work flows.

The obvious solution to the problem of noncompleteness and nonuniqueness is to design better seismic attributes. One can focus on perfecting one single attribute; however, this is often difficult and often provides only partial solutions. A more productive work flow is to recombine several primary attributes into a more complex meta-attribute (Aminzadeh and de Groot, 2006). In this context, primary attributes are 'standard' attributes that can be readily derived from the seismic data. There are many methods for recombining seismic attributes, including neural network methods.

Meta-attributes

Meta-attribute methods come in a wide variety, from simple arithmetic operations to complex neural networks. The amount of improvement towards a robust two-way link of an attribute with a target feature varies with the chosen method. The investment of interpreter’s time and computational resources will also vary with methodology. The desired amount of refinement and complexity should be determined on a case-to-case basis; the choice of the best method depends on the character of the interpretation problem and the desired accuracy of the output. A number of methods have been suggested in the literature, such as common sense arithmetic combinations, cross-plotting, linear regression, nonlinear regression, neural networks, self organizing maps, and fuzzy logic, to name a few. Tables 1 and 2 list guidelines when different methods may be suitable.

Of the methods listed in Tables 1 and 2, the neural network option is very flexible for building an accurate meta-attribute, without too much user involvement. The main advantages of the neural network methods are:

1. The choice of input attributes is not critical, as several (but not all) input attributes may be correlated, meaningless, noisy, or inconsistent without affecting the neural network’s performance. This makes it suitable for geological interpretation of seismic data, in which one has to deal with noise, ambiguities, and lack of a clear model.

2. It is a versatile tool as it can be trained to mimic human interpretation.
3. It is a good method for capturing multivariable nonlinear relationships.

However, the neural network method does have disadvantages, most importantly:

1. The relationship between specific input and output within a neural network is hard to verify due to its inherent nonlinearity.
2. A neural network may find a local solution and not the best global solution. However, most neural networks do have algorithms that are able to avoid local solutions. Several validation methods exist to blind-test the neural network for accuracy, implicitly checking for local optimization problems.
3. Although a neural network gives great results within the domain that is sampled by training, it is not very good at extrapolation. So the application of the neural network is limited by its training experience.

In the cases that one wants or needs to combine many attributes to accurately map a geological feature, the neural network method is an attractive alternative to its competing methods.

Neural Network Methodology

History and introduction

Artificial neural networks have been inspired by what is known as the "brain metaphor." This means that these models try to copy the capabilities of the human brain into computer hardware or software. The human brain has a number of properties that are desirable for artificial systems (*e.g.*, Schmidt, 1994):

- It is robust and fault tolerant. Even if neurons die and are not replaced, the performance of the brain does not deteriorate immediately.
- It is flexible. This means that the human brain can adjust itself to new situations and can learn by experience.
- It can deal with information that is inconsistent or contaminated with noise.
- It can handle unforeseen situations by applying knowledge from other domains and extrapolating this knowledge to new circumstances.
- It can deal with large amounts of input data and quickly extract the relevant properties from that data.
- It is highly parallel; hence it has a high performance.

Neural network research started in the 1940s with McCulloch and Pitts' (1943) description of the logical function of a biological neuron. Specifically, they described the transmission of neural signals as an all-or-nothing situation. A neuron fired only if the cell had been stimulated above a certain threshold. The output signal would, in general, have a constant strength. In their paper, McCulloch and Pitts described that networks consisting of many neurons might be used to develop the universal Turing machine (a kind of computer described by Turing (1937) that could, in principle, solve all mathematical problems). Research in neural networks suddenly stopped following a publication by Minsky and Papert (1969). This paper showed that a relatively simple problem (the so-called XOR-problem) could not be solved by the linear algorithms used at the time. The major breakthrough, which relaunched the interest in neural networks, was the discovery in the eighties of a nonlinear optimization algorithm overcoming the previous limitations (Rumelhart *et. al.*, 1986).

Since then, neural networks have emerged as a promising computing technique that enables computer systems to exhibit some of the desirable brain properties. Various types of networks have been applied successfully in a variety of scientific and technological fields. Examples are applications in industrial process modeling and control, ecological and biological modeling, sociological and economical sciences, as well as medicine (Kavli, 1992). Within the exploration and production world, neural network technology is now being applied to geologic log analysis (Doveton, 1994) and seismic attribute analysis (Schultz *et al.*, 1994).

Numerical details

Multilayer perceptrons (MLP)

The most general and most widely used neural network model is the “multilayer perceptron” (MLP). The basic building block of this model is the perceptron (Fig. 3), a mathematical analog of the biological neuron, first described by Rosenblatt (1962).

The mathematical expression of the biological neuron can be written as an activation function, A, applied to a weighting function, W, defined as:

$$W(\mathbf{y}) = \sum_{i=0}^{L} w_i y_i \text{ , ... (1)}$$

where y is the neural network input vector written as y_i with $i = 1,\ldots,L$ and w is the weighting vector w_i with $i = 1,\ldots,L$.

$$A(W) = \begin{cases} 1 & W > 0 \\ 0 & W \leq 0 \end{cases} \quad (2)$$

In MLPs, the binary activation function is often replaced by a continuous function. The most widely used activation function is the sigmoid function (Fig. 4), defined as:

$$A(W) = \frac{2}{1+\exp(-W)} - 1 \quad (3)$$

In a MLP, the perceptrons are organized in layers (Fig. 5). In its simplest form, there are three layers; an input layer, a hidden layer, and an output layer, and there are no connections between neurons belonging to the same layer. MLPs are trained on a representative dataset, a form of supervised learning. Known examples, consisting of input patterns and corresponding output patterns, are repeatedly fed to the network during the training phase. The learning algorithm (back-propagation), which is widely used to train this type of network, attempts to minimize the error between the predicted network result and the known output by adjusting the weights of the connections. The algorithm has been derived independently by a number of researchers: Werbos (1974), LeCun (1985), Parker (1985), Rumelhart *et al*. (1986), and Fahlman (1988).

MLPs have two properties of interest: abstraction and generalization. Abstraction is the ability to extract the relevant features from the input pattern and discard the irrelevant ones. Generalization allows the network, once trained, to recognize input patterns that were not part of the training set.

Radial Basis Function (RBF) neural networks

Radial basis functions have been used for data modeling (curve fitting) by many researchers; *e.g.*, Powell (1987) and Poggio and Girossi (1989). Recently, these functions have been put in a neural network paradigm in what is called Radial Basis Function (RBF) neural networks (Broomhead and Lowe (1988); Moody and Darken (1988); Lee and Kil (1988); Platt (1991)). RBF networks have been applied in a seismic reservoir characterization study by Schultz *et al.* (1994).

RBF networks have the same feed-forward layered architecture as MLP networks (Fig. 6), but the weighting function W and the activation function A are different. With RBF networks, there are only weights between output layer and hidden layer (Fig. 6). Each node in the hidden layer has a unique function, called the basis function. For the simple network of Figure 4 having a single input, a single output, and two basis functions, the output is given by the sum of the two basis functions, each multiplied by its own weighting factor. In principle, any type of function can be used as a basis function. For example, Kavli (1992) uses splines as basis functions, but they are only denoted RBF networks if radial basis functions are used.

Radial basis functions give local support to data points. The output of the hidden nodes peaks when the input is near the centroid of the node, and then falls off symmetrically as the Euclidean distance between input and the centroid of a node increases (Fig. 7). The consequence of this behavior is that RBF networks are good for data interpolation but not good for data extrapolation.

Several different radial basis functions are in use, with the Gaussian function (Fig. 7a), being the most widely used. If the radial basis center R is defined as:

$$R = \sqrt{\sum_{i=1}^{L} \frac{(y_i - \mu_i)^2}{\sigma_i^2}} \text{ , } \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

where μ_i represents the center location of each basis and σ_i indicates a scaling of the width of each basis, then the Gaussian activation function is given by:

$$A(R) = \exp\left(-\frac{R^2}{2}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Multiplication of the activation function $A(R)$ with a weighting factor w then yields the output o (Fig. 6).

Another widely used RBF function is the so-called Inverse Multiquadratic Equation (IMQE, Fig. 7b), defined as:

$$A(R) = \frac{1}{\sqrt{R + k^2}} \text{ , } \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

where k is an empirically-determined smoothing factor.

Note, that the widths in RBF functions are specified independently from each input dimension, making the functions elliptical rather than spherical. Note as well, that unlike the activation functions for MLPs, no bias is included in the RBF functions. Center locations typically are determined by randomly selecting training examples from a large set of training data. The smoothing parameters and the number of nodes are typically adjusted empirically during training.

RBF neural networks and MLPs have been compared by many. Kavli (1992) report consistently better performance of RBF networks in five independent experiments. Another important aspect when comparing RBF networks and MLPs is the training speed. RBF networks can be trained within a fraction of the time that is required for training MLPs. RBF networks, however, generally require more nodes to obtain similar performances.

Neural network in the work flow of geological interpretation

When utilizing neural networks in geological interpretation and the mapping of a geological target, the process can be generalized by the following steps:

- Selection of the input attributes that will be input to the neural network,
- Selection of the training data for the neural network,

- Training of the neural network, including neural network validation by a random test set,
- Geological validation of the neural network, and
- Application of the neural network.

These steps are discussed in more detail using a few examples.

Examples of Neural Network Application in Geological Characterization

User-driven neural network mapping

Statement of the problem

Gas chimneys are normally visible in the seismic record. However, subtle gas clouds above hydrocarbon reservoirs, deep chimneys related to expulsion from source rocks, and fault-related hydrocarbon migration pathways are often difficult to distinguish in seismic data, as they have a diffuse character and weak expression in the seismic record; hence, they are difficult to map. They are often more visible on vertical seismic sections than on time slices, making it challenging to map their lateral extent. Therefore, a method for detection of gas chimneys in poststack 3D data is needed to improve the identification of gas chimneys in seismic data, to map their distribution, and to allow them to be visualized in three dimensions.

Chimneys are recognized as vertically-aligned, low-amplitude, chaotic zones in normally processed seismic data and will often cause a frequency washout or attenuation of high frequencies (Brouwer *et al.* 2008). Thus, individual trace-to-trace attributes such as similarity and dip variance will often highlight chimneys. Single trace attributes such as instantaneous amplitude, energy, and frequency will also show chimneys. However, individual attributes will also highlight features not related to chimneys; for example, low similarity will also result from faults or mass transport deposits.

A method combining multiple attributes through a neural network to isolate gas chimneys from other features has been developed by Heggland *et al.* (2000) and Meldahl *et al.* (2001). Figure 8 depicts the adopted supervised classification work flow. A number of locations are interpreted manually in the data, selecting locations where the chimneys can be seen in the data. Additionally, a number of counter examples are interpreted on locations that do not exhibit chimneys. A range of attributes are then calculated at these locations, and the neural network is trained to produce the value 1.0 at the chimneys and the value 0.0 at the nonchimney locations. Once the neural network is trained, the attributes are computed throughout the seismic volume, and the trained neural network will create a chimney volume from it, where high values indicate large probability of a chimney, while low values indicate low probability for a chimney to be present.

Selecting the input attributes

The practical approach for finding good input attributes is to systematically list those attributes that describe different aspects of the geologic target and that can be used to separate the geological target from potential false positives. The following list consists of examples of attributes that can detect gas chimneys:

- Verticality (implemented by setting the parameters of most other attributes to vertical features),
- Discontinuity,
- Chaos,
- Signal-to-Noise ratio,
- Frequency anomalies,
- Windowed RMS amplitude anomalies, and
- Two-Way-Time to calibrate the neural network for any depth related effects, for example frequency attenuation.

Selecting the training data

An interpreter provides examples locations (*x*,*y*,*z*) to a neural network to discriminate between chimney locations and nonchimney locations. In this case, one wants to select training points that allow the neural network to comprehend the full problem: that is, with emphasis on difficult zones where either the geological feature is imaged very subtly by only a few of the input attributes or areas where false positive are present in the image of one or more of the input attributes.

However, one should avoid ambiguous areas where the interpreter is unsure whether the geological feature is present or not. The procedure involves initially reviewing the seismic volume to select lines or cross-lines that display the suspected vertical hydrocarbon migration pathways or gas chimneys most clearly. Gas chimneys are often associated with shallow amplitude or AVO anomalies and these anomalies can be used to guide identification of chimneys. Surface seeps, either from synthetic aperture radar (SAR) or shallow geo-hazard surveys, can also guide finding locations of shallow chimneys. Chimney picks are made in the most obvious chimneys. Nonchimney picks are also made in chaotic or low amplitude areas that are suspected not to have chimneys. Nonchimney picks are also made along faults that show no evidence hydrocarbon migration (Fig. 9). A number of attributes that have been shown on numerous datasets to highlight chimneys are then evaluated on the key seismic lines. Attributes that show the chimneys most clearly are chosen as input to a neural network.

The type of neural network used in this methodology is a supervised neural network that learns to distinguish the chimney locations from the nonchimney locations and produce values of 1.0 at the chimney locations and 0.0 at nonchimney locations (Meldahl *et* *al.*, 1999). Once the interpreter is satisfied with the resultant neural network training, based on a reconnaissance of key lines, the neural network can be applied to the entire seismic volume. The resulting chimney probability volume will have values range between zero and one, based on each sample's similarity to the chimney picks. This volume can then be overlain on seismic sections or visualized in three dimensions.

Training the neural network, including neural network validation by a random test set

During neural network training, the example picks made by the interpreter are fed to the neural network, which is an input vector of attributes and the correct classification (here chimney or nonchimney). The neural network is then optimized by back-propagation of the erro, as described in the theory section. During the training phase, it is very important to find the proper stopping point, as overtraining may otherwise occur. Overtraining occurs when the neural network finds relations in the training examples that are not universal. One method to prevent this from happening is to randomly extract a number of the example points from the interpreter-provided training examples and exclude these from neural network training, while still evaluating the prediction error for these points, called the test set error. Generally, optimum training has been achieved once the test set error reaches a plateau or increases with increasing training iterations (Fig. 10).

The training phase also can be used to select or deselect attributes from the neural network input. Obviously, using too many attributes would lead to a heavy computational burden and an under-determination of the inverse problem. A method for reducing the number of attributes to the neural network, while at the same time optimizing the set of input attributes, is to test several sets of input attributes (that may be partially overlapping). For each test, the attributes that provide more important contributions to the output prediction are retained for the next test, while the less important attributes are discarded. By testing all possible input attributes, the final neural network is then created using only the best attributes.

Application and validation methods

Besides numerical validation using the test set, one would like to achieve secondary validation for the correctness of a neural network prediction. There are different methods to achieve this goal, which are very similar to methods to verify seismic attributes and well predictions in general:

- Geologic: The outcome of neural networks should fit in the standard geological understanding. That is if a neural network predicts hydrocarbon saturation, saturation should be found up-dip and have a down-dip structure conformable fluid contact. Gas chimneys mapped in 3D should generally terminate at a regional seal and be associated with active source rock.

- Correlation with independent data: For the example of gas chimneys, at locations where a borehole intersects a seismic gas chimney, elevated levels of C2 through C5 are often found in mud logs (Løseth *et al.*, 2002), which would be a positive confirmation of the neural network.

In the case of gas chimneys, there are several pitfalls related to both geological and geophysical phenomena that can occur in the seismic data. Vertically aligned zones of chaotic data can occur for other reasons than gas chimneys. Geologic features such as diapiric shale and salt, mass transport complexes, and volcanic pipes can look similar to chimneys. De-watering of mudstones related to burial compaction can result in polygonal faulting which can look chaotic on seismic section views. Poor seismic imaging related to surface statics, fault shadows, complex structuring (imbricate thrusts), or sub-salt or sub-thrust intervals also could be misleading. In the case of gas chimney prediction, the following five sources of validation can be systematically applied to the results (if available).

1. Presence of drilled fields or seismic evidence of hydrocarbon presence, such as direct hydrocarbon indicators (amplitude, AVO, or frequency attenuation anomalies), and chemosynthetic carbonate build-ups associated with the chimney.
2. Linking the gas chimney to shallow sea-bottom indications of hydrocarbon seepage detected through piston core data, side-scan sonar, and other scanning methods.
3. Matching of the chimneys with basin modeling, which can predict vertical hydrocarbon migration based on independent inputs.
4. Linking the chimneys within deep, thermally mature sediments that contain source rock intervals.
5. The chimney morphology should express a circular pockmark pattern that is characteristic of fault-related hydrocarbon migration. This is best evaluated on time and horizon slices. Wells drilled through gas chimneys in the North Sea often encounter elevated pore pressure, gas shows, and gas wetness (Loseth *et al.*, 2002).

If a neural network fails validation, one can improve the neural network by improving input attributes or design and include additional attributes that capture other aspects of the data. One can also include additional examples, especially by providing counter-examples to the neural network at locations of false prediction. Finally, as no prediction is perfect, one needs to understand the remaining uncertainty and take this into account during interpretation.

Application

Once the final neural network is tested and validated against key information, we can compute a 3D volume that optimally images gas chimneys or other geological features (such as faults, channels, salt, etc.) that we can recognize in the seismic data. The volume can now be displayed in 2D and 3D views, recombined with other information, and the interpreter can be confident he or she is looking at a specific geological feature (Figs. 11 and 12).

Example of quantitative neural network predictions

Statement of the problem

Well properties can be computed from seismic data using the result of seismic inversion such as acoustic impedance, elastic impedance, and shear impedance. The standard procedure to obtain these properties is to apply a regression of one or more inversion volumes with a downscaled version of the well log, representing the desired target property; *e.g.*, a rock or fluid property. Using a neural network has several advantages in this work flow. In comparison to multivariate linear regression, the neural network is better suited at handling highly correlated input data and incorporating nonlinear trends between input and output. Note that nonlinear relations are common in geological analysis and incorporating these correctly effect the output prediction materially. In addition, neural networks can more easily handle unrelated input volumes; for example, a neural network having as input seismic facies attribute can also input a two-way time (TWT) (or depth) attribute to detect and correct depth-dependent trends in the seismic facies attributes. In the example presented on Figures 13, 14, and 15, we use the neural network to predict a gamma ray response based on the seismic data.

Selecting the input attributes

For rock and fluid property prediction, we would typically use multiple types of attributes:

- Geophysical inversion volumes: These contain information about the elastic properties of the rock material that in turn may be correlated with target rock properties, which are valuable in predicting lithology and fluid type.
- TWT, depth, or stratigraphic attributes: TWT and depth attributes that can capture any compaction and pressure-related trends in the target property. Stratigraphic attributes are horizon-based and give the neural network an idea in which part of the vertical section the neural network is active.
- Velocity fields from processing and pressure prediction volumes: For detecting any relations between target property and pressure or rock velocity.
- Facies attributes: Facies attributes capture information related to the environment of deposition such as looking at seismic amplitude, frequency, statistics relating to reflector continuity, and dip relations (parallel, semi-parallel, variable). Facies

attributes are relevant as different geological facies may have nearly identical elastic properties but very different geological properties. For example, difference in grain size and grain size distribution that are related to type of sediment input and energy of the depositional processes and which can be inferred from facies attributes.

In our example, we use three kinds of attributes:

1. Acoustic impedance volume and a normalized acoustic impedance volume that is corrected for the correlation of acoustic impedance with depth.
2. Frequency. Typically, the frequency of the seismic data will change between rapid sand-shale successions and massive sand or shale formations. Thus, this facies attribute may help the neural network to determine the sedimentary environment.
3. Stratigraphic level: relative reference time to four horizons that define distinct system tracts is fed to the neural network. As each system tract will have different processes for (re)distribution of sediments and different sorting of grain size and minerals, the relationship between acoustic impedance and gamma ray will vary between system tracts. Feeding information about the stratigraphic level to the neural network allows it to fine-tune acoustic impedance to gamma ray relationships for every systems tract.

Selecting the training data

Neural networks for log property prediction do not require interpreters to pick training examples. Instead, example locations are extracted along the well path, where both seismic attributes and corresponding well properties are sampled. Care should be taken that data points having high noise level (such as locations in which there is borehole breakout) are excluded from the training data. In addition, statistical rebalancing of the input may be necessary. In many applications, the values of maximum interest for exploration and development (high porosity, high permeability, and high hydrocarbon saturation) may only be a small percentage of the input dataset, while also exhibiting an abnormal relation between input and output. If the training set contains mainly nonreservoir examples, then the neural network training will converge to a solution that predicts the large majority of the data as accurately as possible at the expense of less frequent examples (like those representative of reservoir zones). Rebalancing the input dataset puts extra emphasis on the limited amount of really interesting data. In our example, the bulk of the dataset has gamma ray levels around 40-60 API; however, there are values above or below this range that are more interesting as they represent clean reservoir or better seal. Therefore, the training dataset is rebalanced to remove points in the main range and add points in the extremes. The latter are copies of existing data points in which a small amount of noise is added. The rebalance will bias the neural network training towards predicting the extremes with more accuracy.

Training the neural network, including neural network validation by a random test set

The methodology for training a data-driven neural network is essentially the same as for the user-driven neural network. An example of the neural network training is shown in Figure 13.

Application and validation methods

Holding part of the well data back during training for blind testing can help validate data driven neural networks. This can be done on two scales:

First, well data may be withheld on the wells. The comparison between the neural network's prediction and the real values tells how well the neural network can predict within a well. The problem with this test is that the points in the test set will always be geographically near and thus relatively correlated.

The second way of holding back data from the training is to leave out entire wells from the training. The comparison between the neural network's prediction and the withheld well data tells how well the neural network can predict in the areas away from the wells.

Thus, only these blind tests truly test the predictive confidence of the neural network in areas away from the wells. If a large number of wells are available, the second method should be chosen. Finally, the output should be inspected visually to make sure it makes geological sense.

Application

Once the final neural network is tested and validated against key information, we can compute a 3D volume of the property of interest based on seismic data. The volume can now be displayed in 2D and 3D views, sliced along surfaces, recombined with other information, and the interpreter can be confident he or she is looking at a meaningful rock property. In addition, the results may be used in downstream applications such as reservoir simulation. For our example, we show the results between the two wells in Figure 14. One of the wells is used to blind test the results (Fig. 15). The cross-plots in Figure 15 show how the neural network can create a fairly accurate prediction of a rock property even if the correlation between elastic rock properties derived from the seismic data and the target property is weak. Though the final prediction certainly has an error spread, most trends are captured correctly, and the gamma ray volume is of great benefit for planning of tasks such as infill drilling.

Conclusions

Neural network technology is in many cases more robust than single attribute volumes to highlight geological features from seismic data and to predict rock properties from seismic attributes. They have been applied to seismic data for over 20 years, and are proven a stable, viable, and easy to use tool. Their shortcomings are relatively limited and can be monitored by blind-testing and other methods.

References

Aminzadeh, F. and de P.F.M. Groot, 2006, Neural Networks and Other Soft Computing Techniques with Applications in the Oil Industry: EAGE Publications BV, Houten, 161p.

Broomhead, D.S., and D. Lowe, 1988, Multivariable functional interpolation and adaptive networks: Complex systems, v. 2, p. 231-355.

Brouwer, F., A. Welsh, D.L. Connolly, D. Curia, C. Selva, and A. Huck, 2008, High frequency attenuation and low frequency shadows in seismic data caused by gas chimneys, Onshore Ecuador: EAGE 70th Conference & Exhibition.

Doveton, J.H., 1994, Geologic log analysis using computer methods: AAPG Computer Applications in Geology, No. 2, 169 p.

Fahlman, S.E., 1988, An empirical study of learning speed in back propagation networks: Technical Report CMU-CS-88-162, 19 p.

Heggland, R., P. Meldahl, P. de Groot, and F. Aminzadeh, 2000, Seismic chimney interpretation examples from the North Sea and the Gulf of Mexico: American Oil and Gas Reporter, February issue, p 78-83.

Kavli, T.Ø., 1992, Learning principles in dynamic cControl: University of Oslo, PhD dissertation.

LeCun, Y., 1985, Une procedure d'apprentissage pour réseau à seauil asymétrique (A learning procedure for asymmetric threshold networks): Proceedings of Cognitiva: v. 85, p. 599-604.

Lee, S. and R.M. Kil, 1988, Multilayer feedforward potential function network: IEEE International Conference on Neural Networks, p. I-161 – I-171.

Løseth, H., L. Wensaas, and B. Arntsen, 2002, Gas chimneys—indicating fractured cap rocks: AAPG Hedberg Conference.

McCulloch, W.S., and W. Pitts, 1943, A logical calculus of idea's immanent in nervous activity: Bulletin of Mathematical Biophysics, v. 5, p. 115-133.

Meldahl, P., R. Heggland, P. de Groot and A. Bril, 1999, The chimney cube, an example of semi-automated detection of seismic objects by directive attributes and neural networks: Part 1; Methodology: Extended Abstracts of 69th Annual SEG Meeting.

Meldahl, P., R. Heggland, B. Bril, and P. de Groot, 2001, An Iterative method for Identifying seismic objects by their texture, orientation and size: Extended Abstracts of 71st Annual SEG Meeting.

Minsky, M. and S. Papert, 1969, Perceptrons: An Introduction to Computational Geometry: MIT Press, 268p.

Moody, J., and C. J. Darken, 1988, Learning with localized receptive fields, *in* D.S. Touretsky, G.E. Hinton, and T.J. Sejnowski, eds, Proceedings of the 1988 Connectionist Models Summer School: Morgan-Kaufmann, p. 133-143.

Parker, D.B., 1985, Learning-Logic: Technical Report. TR-47, MIT Center for Computational Research in Economic and Management Science, Cambridge, MA.

Platt, J., 1991, A resource-allocating network for function interpolation: Neural Computation, v. 3, p. 213-225.

Poggio, T. and F. Girosi, 1989, A theory of networks for approximation and learning: Technical Report, MIT Artificial Intelligence Laboratory.

Powel, M.J.D., 1987, Radial basis functions for multivariable interpolation: A review, *in* Mason, J.C., and M.G. Cox, eds, Algorithms for Approximation: Clarendon Press.

Rosenblatt, F., 1962, Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms: Spartan Books, 616 p.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams, 1986. Learning internal representations by error propagation, *in* D.E. Rumelhart, J.L. McClelland, and the DPD Research Group, eds, Parallel Distributed Processing: MIT Press, p. 318-362.

Schmidt, W.F., 1994, Neural pattern classifying systems: theory and experiments with trainable pattern classifiers: PhD Dissertation Technical University Delft, 224 p.

Schultz, P.S., S. Ronen, M. Hattori, and C. Corbett, 1994, Seismic-guided estimation of log properties (Part 1: A data-driven interpretation methodology): The Leading Edge, v. 13, p. 305-320.

Turing A.M., 1937, On computable numbers, with an application to the Entscheidungsproblem. A correction: Proceedings of the London Mathematical Society, second series, v. 43, p 544 – 546.

Werbos, P.J., 1974, Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences: PhD Dissertaion Harvard University.

Table 1. Different geological features and corresponding best-practice methods for highlight the features using seismic attributes.

Criterion	Best method	Examples
The geological target is well-defined by a single attribute, or geologically precise definition of the attribute is not very important.	Use a single attribute.	Amplitude, frequency, discontinuity, AVO gradient.
An easy recombination of two or three attributes will enhance the geological imaging to acceptable levels.	Use arithmetic recombination of the attributes.	Sweetness attribute, (F-N)*F (AVO) attribute, Frequency Ratio.
A recombination of two or three attributes will enhance the geological imaging, but more accuracy or control is necessary.	Use cross-plotting to recombine the attributes.	Cross-plotting of AVO attributes.
Data for an exact quantification (calibration) is lacking; general interpretive rules are available to recombine a series of attributes towards a geological prediction.	Use fuzzy logic and expert systems.	AVO in absence of well control.
The expression of the geological target in the seismic data is highly variable, nonlinear or weak, and three or more attributes or attribute parameterizations are needed to adequately image the target.	Use multivariant linear and nonlinear methods including neural network and self-organizing maps.	See Table 2.

Table 2. Geologic features and their corresponding best-practice multiattribute methods.

Criterion	Best method	Example
There are multiple attributes that image a geologic target, but the relationship between the attributes is linear.	Use linear regression.	
The relationship between the attributes and geological target is complex, sometimes inconsistent, and may involve nonlinear features such as threshold values or altering sensitivity over the range.	Use neural networks or self-organizing maps.	Imaging of gas chimneys, facies differences, porosity prediction, including compaction curves and facies type attributes.

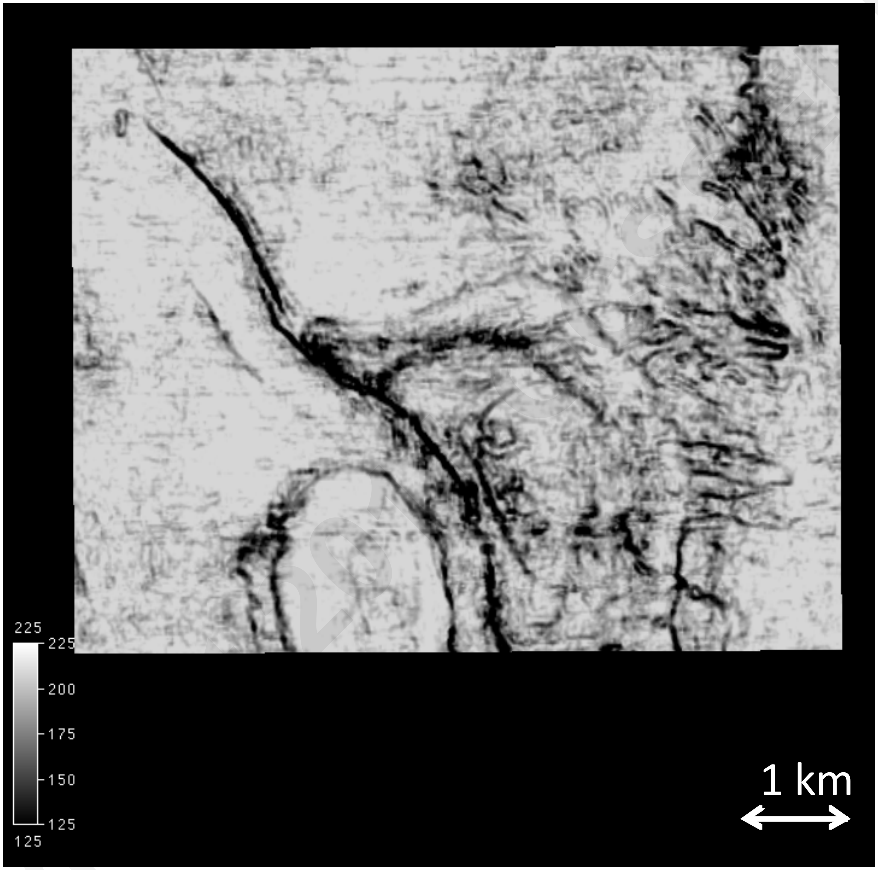

Figure 1. An example of the geological nonuniqueness of an attribute. The discontinuity attribute images both faults and incised sedimentary features. To illustrate the problem, we have not added any pointers to the image in which anomalies represent faults and incised features. Which features can be interpreted with high confidence?

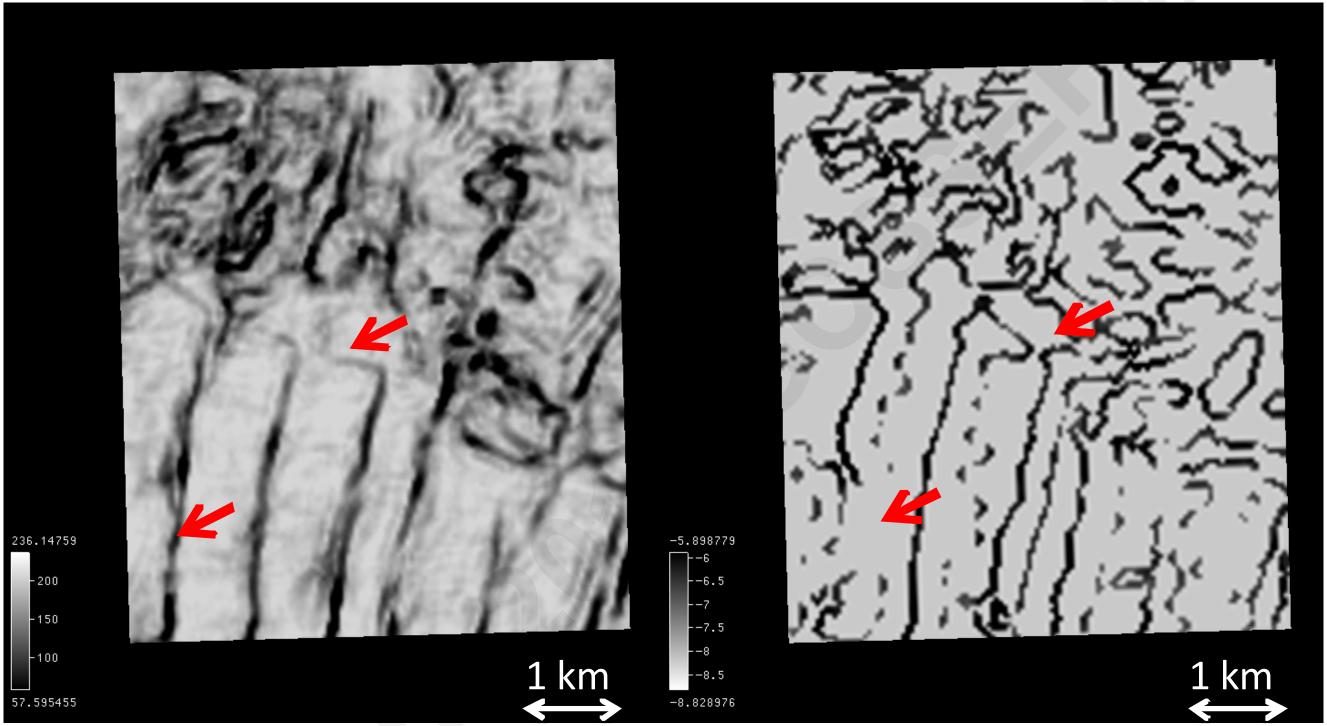

Figure 2. Side by side comparison of a discontinuity attribute (left) and a modified curvature attribute. Each attribute images a part of the complete fault system. The red arrows indicate some points of interest, where one attribute provides additional information. Combining the different attributes would work towards a more complete image of the fault system.

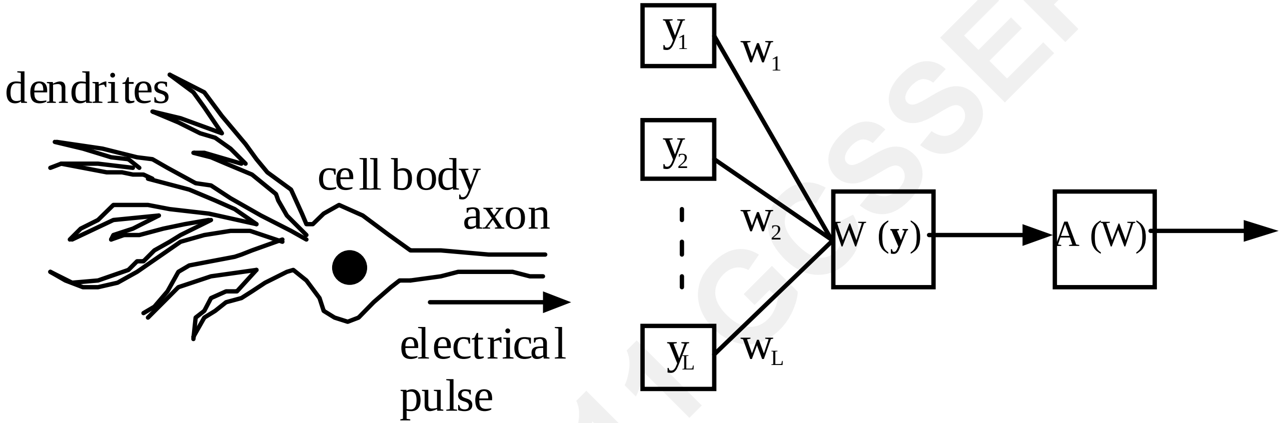

Figure 3. Left: A biological neuron, a basic building block for the biological neural networks. Right: Inspired on the biological neuron, the perceptron is the basic building block for artificial neural networks.

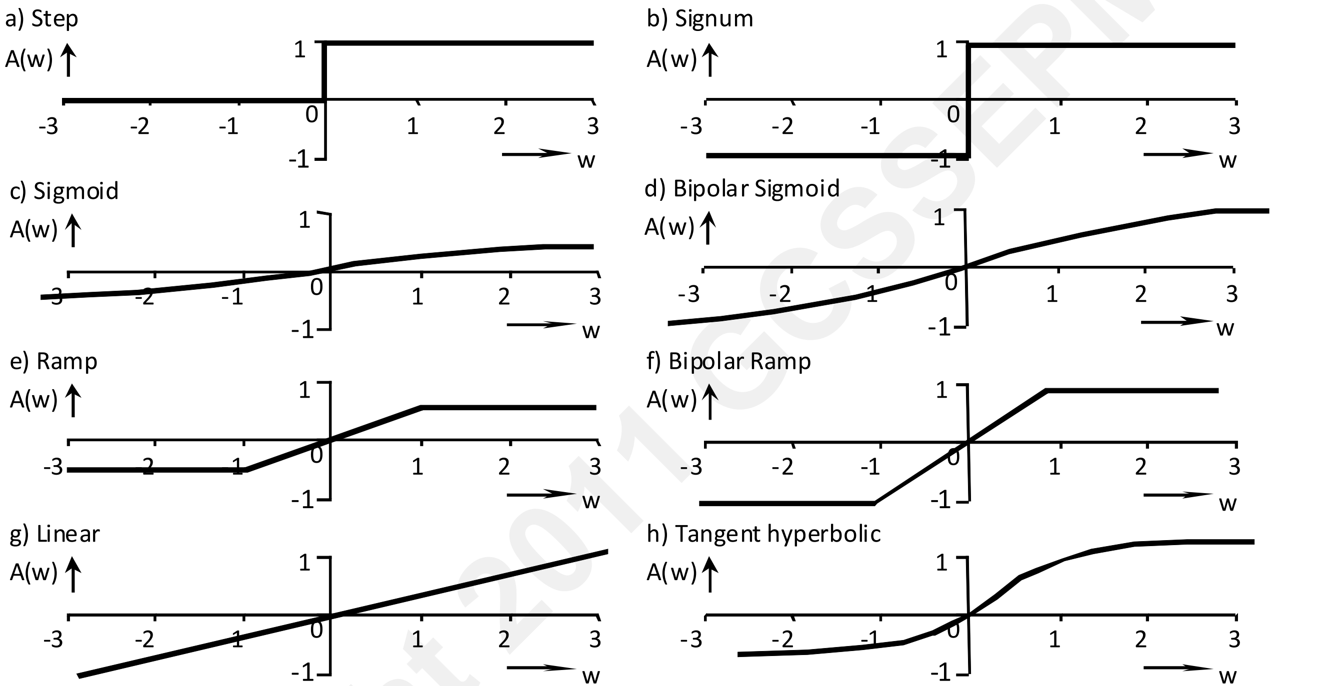

Figure 4. Typical activation functions used in neural networks. Note especially the sigmoid activation function (c) is widely used for perceptrons.

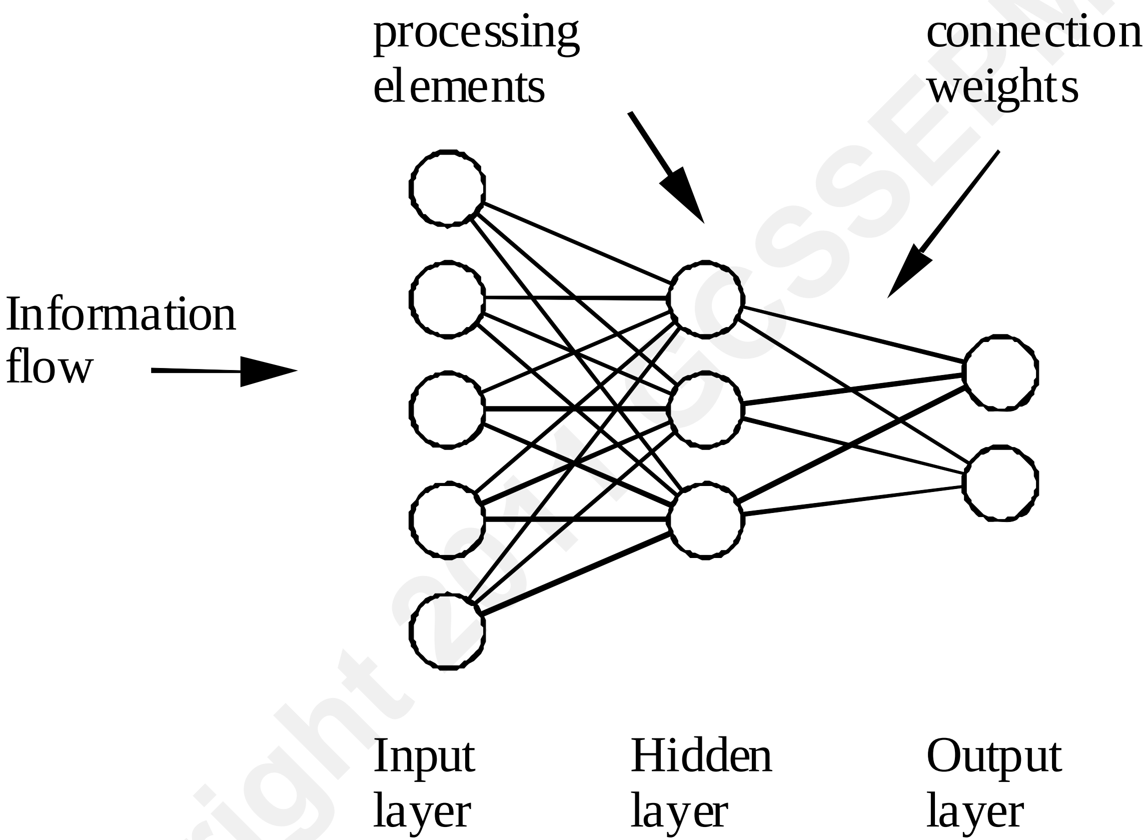

Figure 5. Typical architecture of feed-forward artificial Neural Networks, such as multilayered perceptrons.

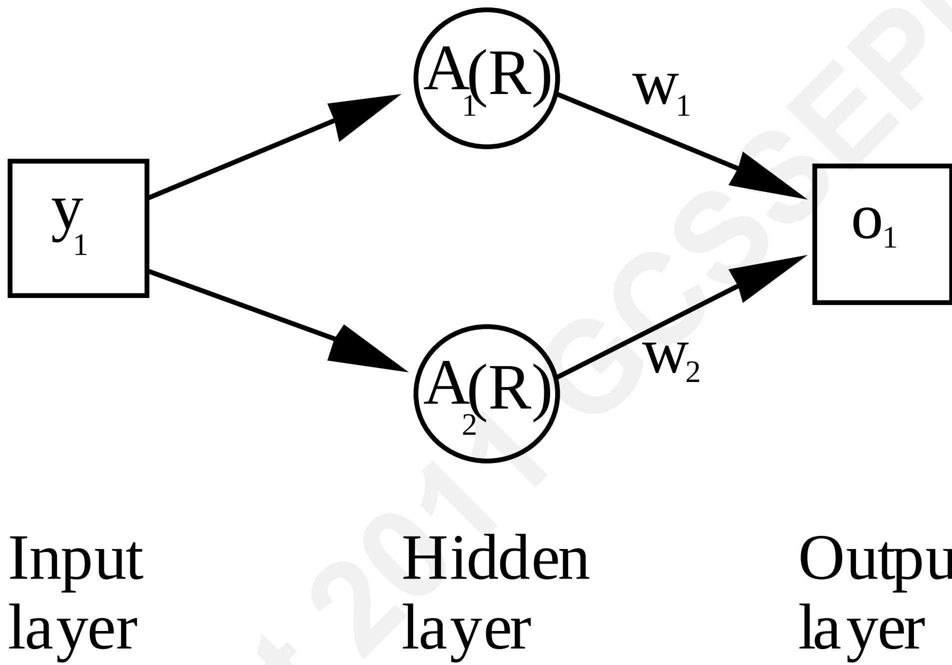

Figure 6. Architecture of radial basis function neural network; note that compared with Figure 5 the weighting is limited between the hidden and output layer only.

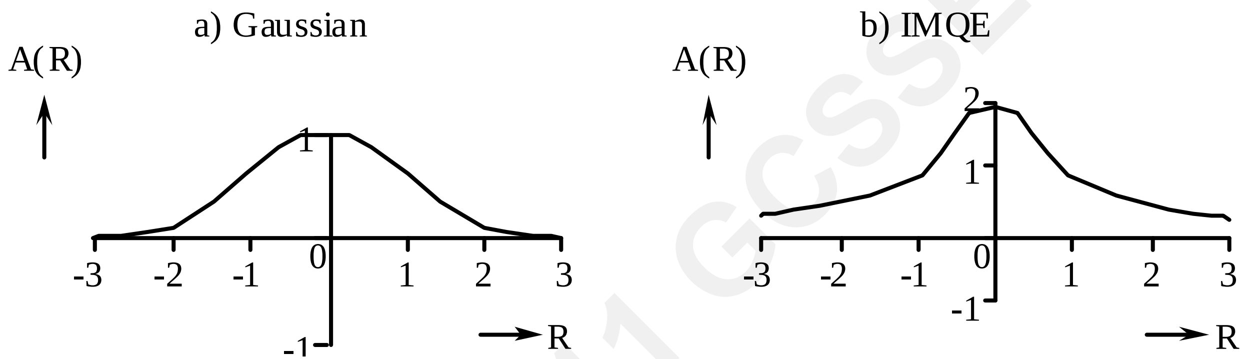

Figure 7. Main activation functions of a radial basis function neural network.

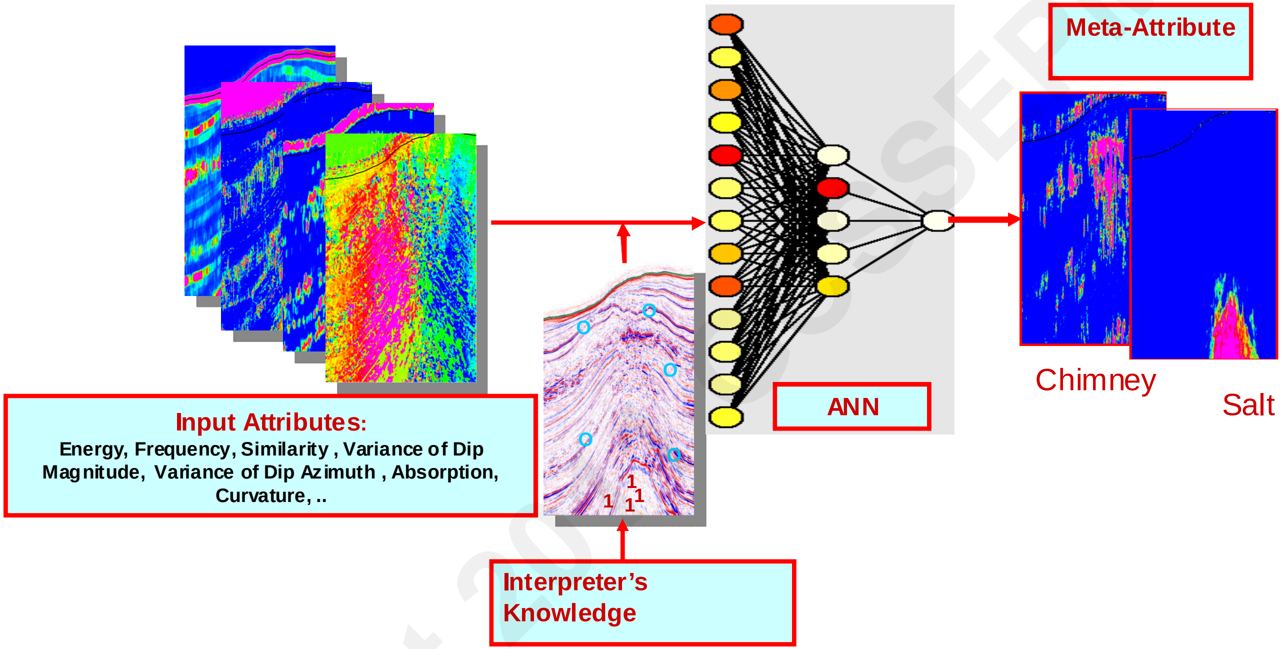

Figure 8. Concept of the use of neural network s for the use of geological characterization of seismic data. The sections are 8.3 km wide.

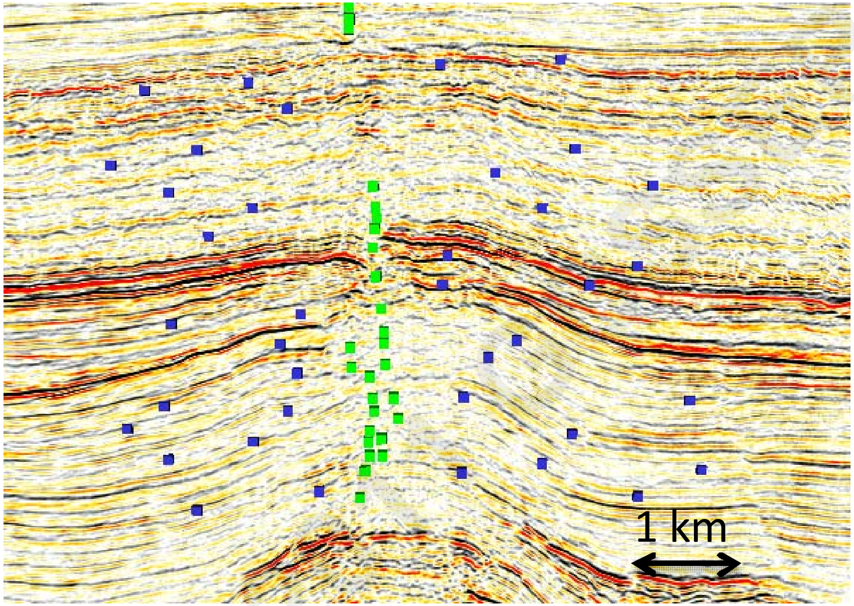

Figure 9. Example picks of a neural network. The green points are locations the interpreter has classified as part of a gas chimney; the blue points are locations the interpreter has classified as not being part of the gas chimney. Guided by these interpreted examples, the neural network will be trained to recognize combinations of seismic attributes that are predictive for chimney and nonchimney.

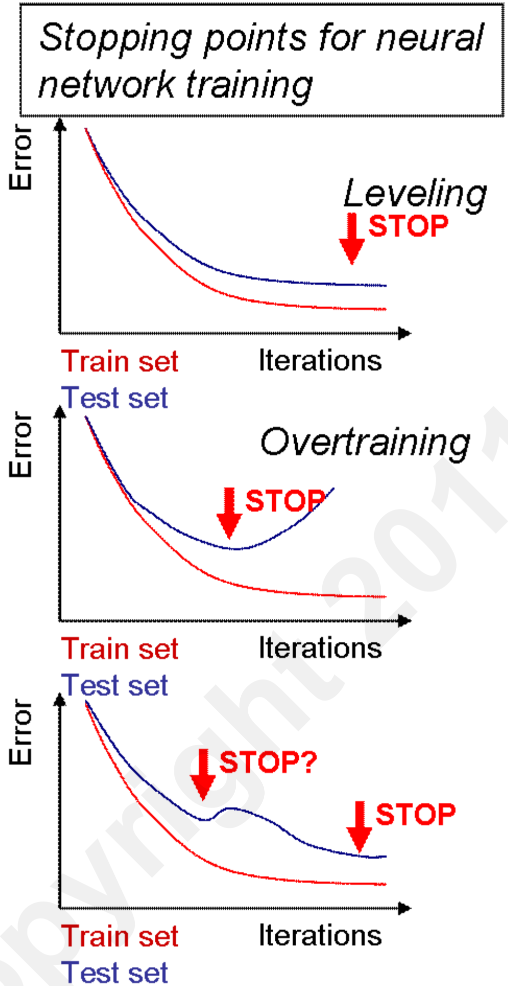

Figure 10. Stopping criteria for a neural network using a test set not used to update the neural network nodes. Careful monitoring of test set performance will avoid overtraining, one of the main pitfalls in the use of neural networks.

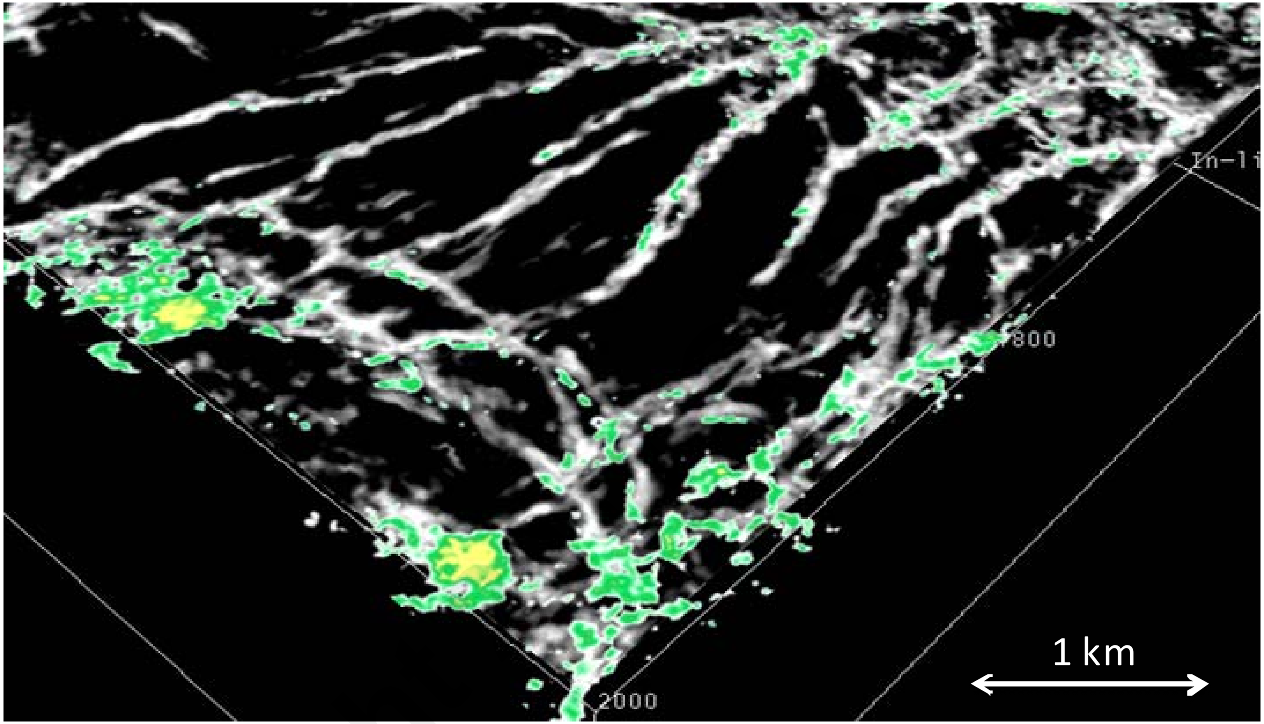

Figure 11. White tones: neural network map of fault system. Green-Yellow tones: neural network map of gas chimney. In one view, the interpreter can access the vertical leakage of hydro-carbons along a fault plane and assess vertical charge efficiency and/or vertical fault seal risk at traps associated with the fault system.

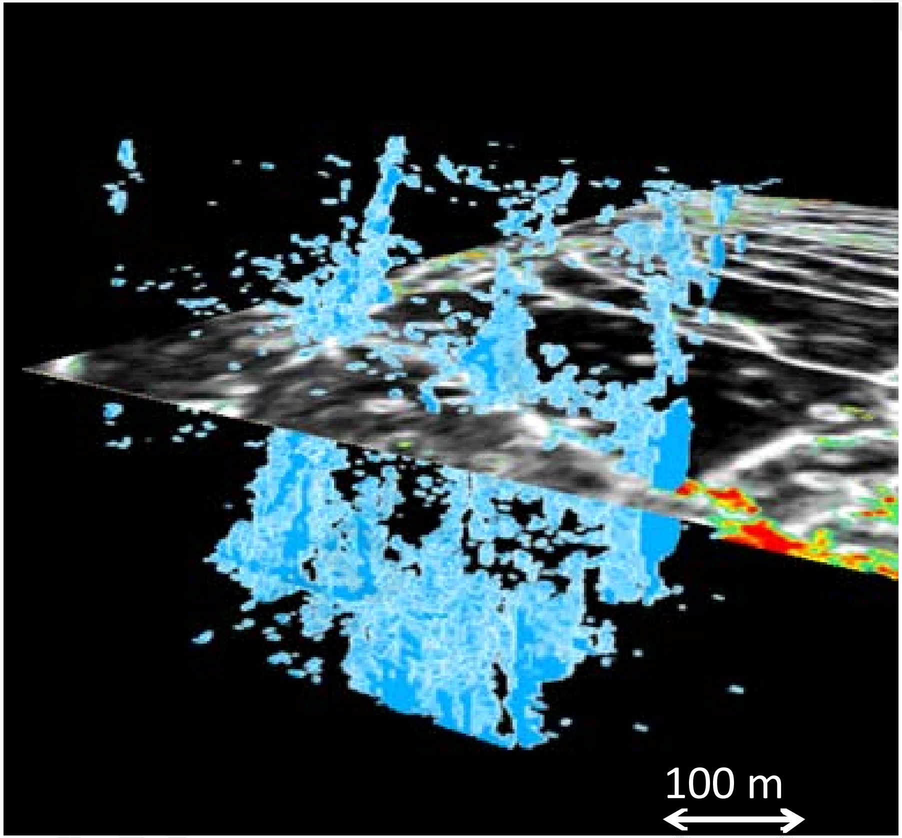

Figure 12. After accurate 3D mapping of geological features, they can be extracted in 3D.

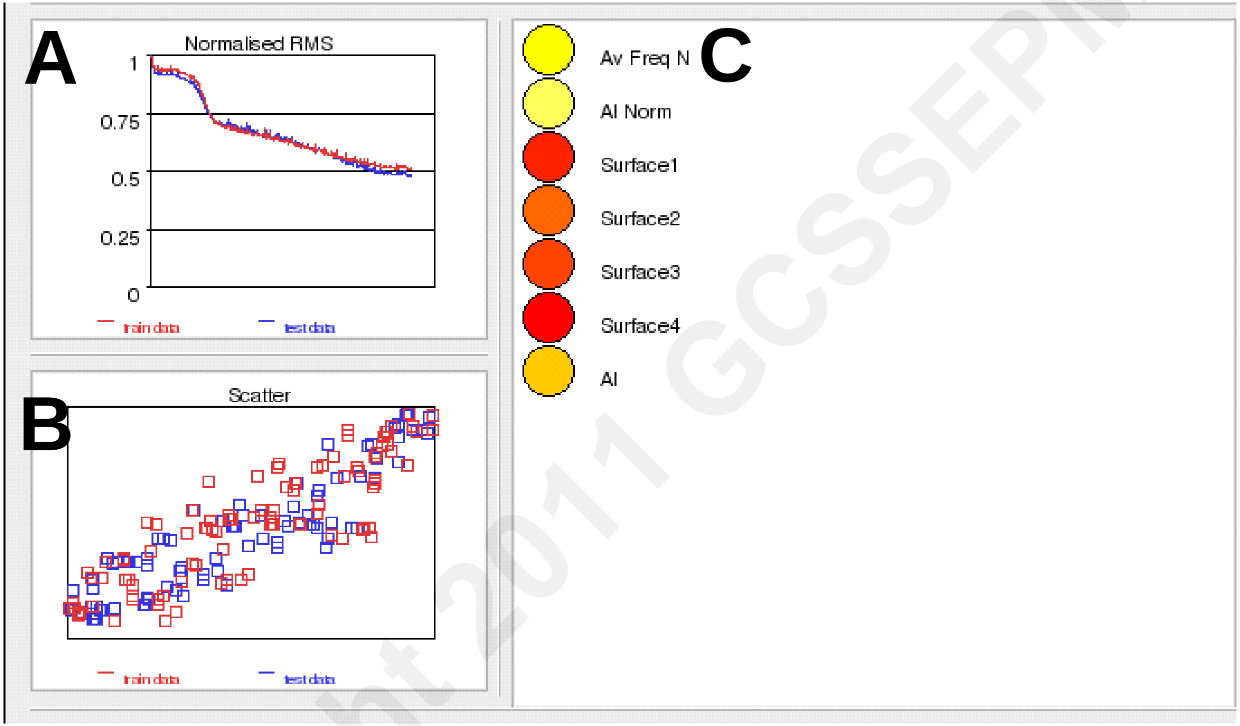

Figure 13. Neural network training window. In this window we find three measures used to control the training. The error of training set and test set is monitored (A), a scatter plot displaying the predicted value versus actual value is given (B) and the input nodes of the neural network together with a "temperature type" color coding representing relative weighting of the input is given (C). Based on these controls, the interpreter decides when neural network training is completed.

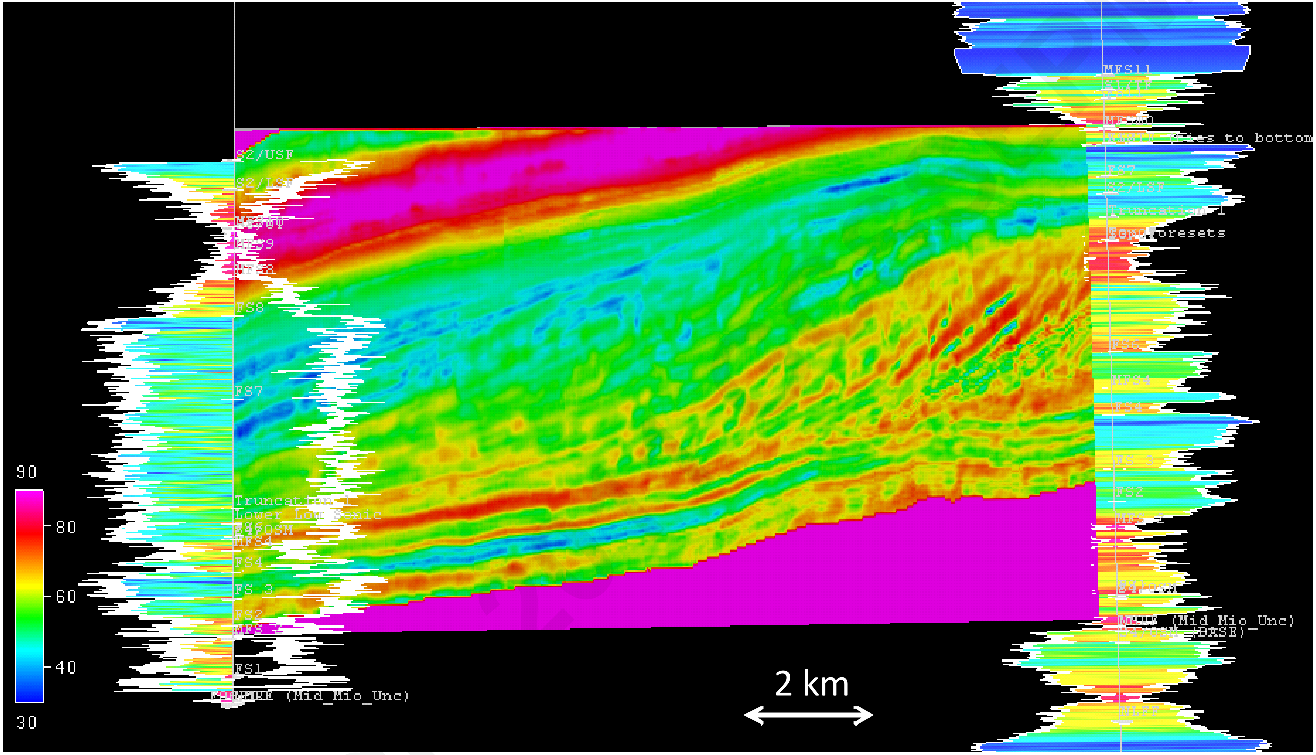

Figure 14. Predicted gamma ray and neural network together with gamma ray well logs. The right log has been used to train the neural network (see Fig. 13). The left log has not been used in neural network training and functions as independent validation (blind test).

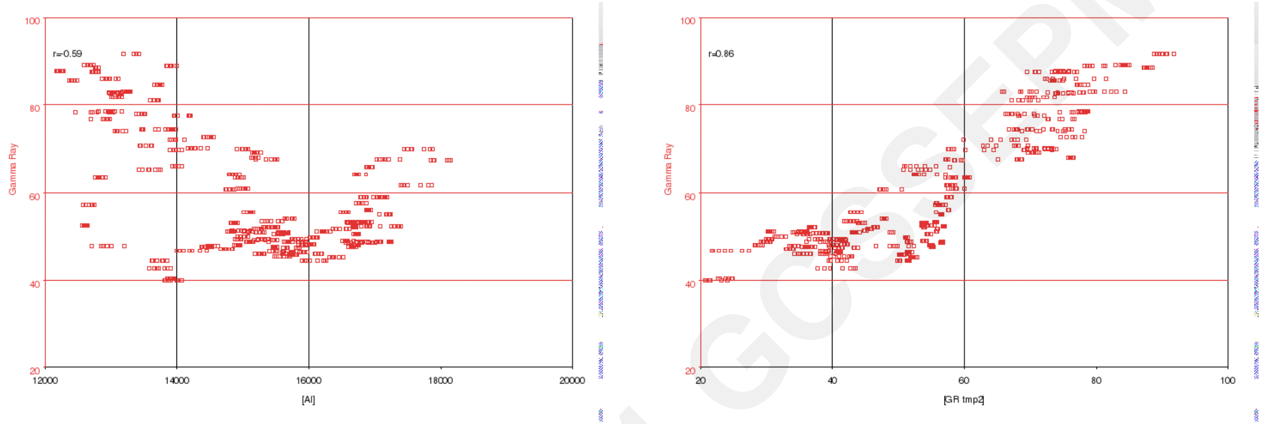

Figure 15. Cross-plots of gamma ray vs. acoustic impedance (left) and gamma ray vs. predicted gamma ray (right). Extraction is made from the well track of the blind test location mentioned in Figure 14. It is obvious that the neural network performs a better prediction of gamma ray than a linear correlation using acoustic impedance data alone.