CHAPTER 1. INTRODUCTION

Electromagnetic (EM) soundings, carried out with both controlled-source and natural-source techniques, are widely used in petroleum, mineral, groundwater, and geothermal exploration. Since controlled source time-domain electromagnetic (TEM) systems are broadband, they are affected by noise which arises from man-made and cultural sources (power lines, generators, radio-navigation, and maritime mobile transmitters) and natural sources such as sferics and wind. Therefore, sferics, power line, and very low frequency (VLF) noise provide major limitations to minimum detectable signals in ground and airborne EM surveys. Conventional brute force methods such as extended stacking times or dramatic increases in transmitter power are not a viable option in airborne electromagnetic (AEM) methods, and are undesirable for ground EM systems. It has been shown that a reduction in sferics noise by a factor of 5 to 10 in an area where the sferics noise predominates over variations in background geological signal will increase the detectable target depth by 50 to 80% (Buselli and Cameron, 1995).

The focus of this thesis is the development of new methods to reduce sferics noise. These methods are based on an artificial neural network. Both a local noise prediction filter (LNPF) and a remote noise prediction filter (RNPF) for reduction of VLF noise and high-frequency sferics noise in TEM measurements are discussed. In addition, a model training method (MTM) to map a noisy transient to a noise-free transient is investigated. A backpropagation method is used as a neural network algorithm to update the connection weights of the network.

The spatial correlation of sferics noise in both horizontal and vertical planes and the correlation between three orthogonal components of the noise are investigated to establish the efficacy of using multiple remote or local reference receivers. Also, the spatial correlation of sferics noise in the bandwidth of 1 Hz to 1 kHz has been studied.

Most EM noise reduction methods with a remote reference and an LNPF based on the tipper method (e.g., Spies, 1988) require an assumption about the relationship between a local field and a remote field (e.g., a linear point-to-point transfer function) or between two horizontal and vertical components (e.g., a linear convolution between two horizontal and vertical components in time-domain), and such methods focus on low-frequency EM noise reduction (e.g., Nichols and Morrison, 1988; San Filipo and Hohmann, 1983; Wilt et al., 1983). Some noise reduction techniques use fixed filter coefficients (e.g., Halverson, 1982 and 1990; Halverson et al., 1987; Spies, 1988). Such filters with fixed filter coefficients are not adaptive but typically process all training data before being used with new data. Such filters are effective in reducing statistically stationary EM noise but, when nonstationary noise is present, the performance of the filters is greatly reduced.

A multi-layer neural network can be considered as an universal approximator, because it is possible to approximate any multi-dimensional functional to any desirable degree of accuracy simply by superposition of a sigmoidal function (Cybenco, 1989; Hornik et al., 1989, Lippmann, 1987). Therefore, to develop a remote reference technique and an LNPF to reduce EM noise, a neural network can be found to approximate any transfer function without the need to state the function explicitly.

Adaptation or learning is a major benefit of a neural network. The ability to adapt and continue learning is essential in areas such as EM noise reduction when nonstationary noise is present, where the training data is limited, and when spatial characteristics of noise pattern change. Adaptation also provides a degree of robustness by compensating for minor variabilities in the characteristics of processing elements in the network.

The backpropagation training algorithm for neural networks is a nonlinear extension of the linear least mean square (LMS) algorithm commonly used in adaptive EM noise reduction (e.g., Kim and Hohmann, 1992) and adaptive signal processing (e.g., Widrow and Sterns, 1985; Widrow et al., 1975, 1976).

1.1 Natural and artificial EM sources

Natural EM fields in the frequency range of interest in exploration, 10^{-3} to 10^4 Hz, have their origin in the atmosphere and the magnetosphere. Their typical spectra (Macnae et al., 1984) are shown in Figure 1.1. The spectrum below 1 Hz originates predominantly in micropulsations. This geomagnetic micropulsations arise from interactions between radiation emitted by the sun, and the earth's ionosphere and magnetosphere. The spectrum above 1 Hz is primarily due to sferics, which are EM transients generated by lightning discharges, or by any subsidiary feature of the lightning that occurs in the lower atmosphere.

These natural EM fields are considered as a signal for the magnetotelluric (MT) method (Cagniard, 1953; Tikhonov, 1950; Vozoff, 1972; Vozoff and Ellis, 1966), while they are noise sources for frequency-domain electromagnetic (FEM) methods, TEM methods, and the Induced Polarisation (IP) method (Buselli, 1977; Buselli and Cameron, 1992, 1993 and 1995; Halverson, 1982 and 1990; Halverson et al., 1987; Kim and Hohmann, 1992; Macnae et al., 1984; Nichols and Morrison, 1988; San Filipo and Hohmann, 1983; Spies, 1988; Sumner, 1976; Vozoff, 1984).

Typical lightning has a total path length on the order of kilometres. Most lightning consists of two to four strokes and is produced by thunderclouds. Sferics generally propagate directly to a receiver near the source. Therefore, the effect of local sferics is to produce large spikes. Waveforms of sferics observed at some distance from the origin are controlled both by the characteristics of the lightning channel (electric current distribution, channel configuration, and orientation) and by

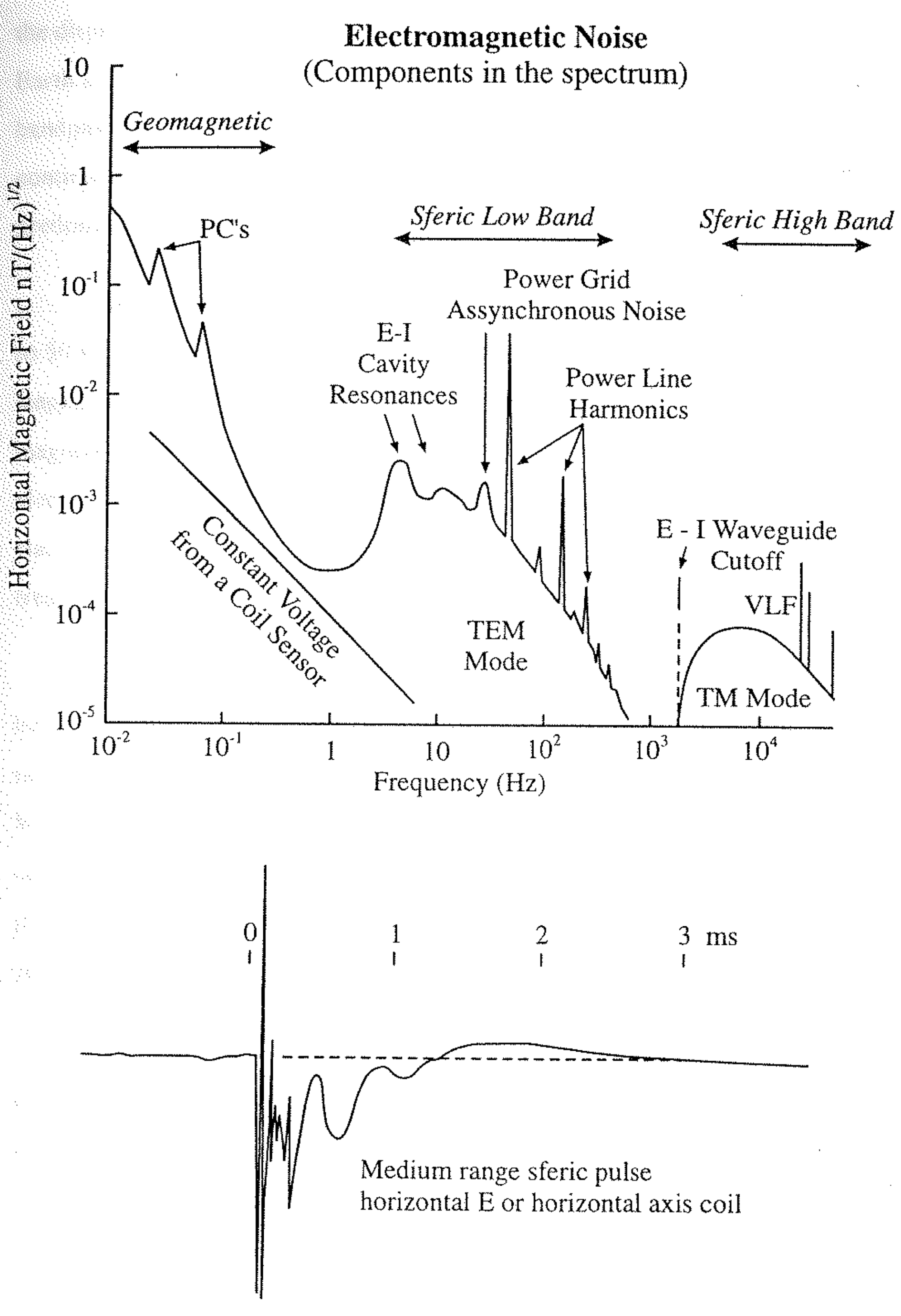

Figure 1.1 Components in the electromagnetic noise spectrum (schematic), plus detail on one sferics pulse. The amplitude and shape shown are based on typical published values, and will vary with location, time of day, year, etc (from Macnae et al., 1984).

the dispersion characteristics of the propagation medium between the surface of the Earth and ionosphere (Pierce, 1977). An essential feature of these waveforms is the bandpass filtering effect of the wave guide. The EM noise from sferics is quasi-continuous at the lower end of sferics low band shown in Figure 1.1 but is predominantly impulsive in nature at high frequencies.

Radiation from lightning discharges in a relatively low frequency range of ~0.3 to 50 kHz is the source of "whistlers" that propagate through the ionosphere and magnetosphere in the whistler mode. This mode of propagation is possible only in a magnetised plasma and at frequencies below both the plasma frequency and the electron gyrofrequency. Propagation in this mode is strongly affected by the earth's static magnetic field and is characterised by low propagation speeds that vary with frequency.

Some lightning currents, most probably subsequent return strokes, contain a long "continuous" component, which is the ELF component (< ~ 3 kHz). The terrestrial wave guide has a pronounced attenuation maximum around 1 to 3 kHz. When the free space wavelength is large, compared with the distance between two concentric spherical shells, only the lowest order transverse magnetic (TM) mode can propagate (Wait, 1970). This is clearly the case at ELF, with a wavelength of 30000 km at 10 Hz in the Earth-ionosphere wave guide (with a height of ~100 km).

The spherical Earth-ionosphere waveguide has resonant frequencies of approximately 8, 14, 20, 26, and 32 Hz. These Schumann resonances occur when the ELF component propagates and reinforces through the Earth-ionosphere wave guide (Polk, 1982; Balsfer and Wagner, 1960).

Many types of VLF (~10 to 50 kHz) waves in the magnetosphere are generated by naturally occurring plasma processes or by man-made devices on the ground rather than by sferics. These waves propagate in the whistler mode and interact with energetic electrons in the same way as do whistler. Naturally occurring

VLF signals, with a wide variety of tonal characteristics but distinctively different from whistlers, have been classified into two broad categories. One category, called "chorus", consists of a series of clearly discernible discrete tones, while the second category, called "hiss", sounds like amorphous broadband noise.

Several dozens of high-power transmitters (each up to a few megawatts radiated power) operate around the world in the ~10 to 50 kHz range for purposes of submarine communication and navigation. EM signals from these VLF transmitters enter the magnetosphere and propagate in the whistler mode through regions where the electron gyrofrequency exceeds the wave frequency. It has been discovered that radiation from electrical power lines is another source of whistler-mode waves in the magnetosphere (Park, 1982; Park and Helliwell, 1978).

Most of the world's power systems use either 50 Hz or 60 Hz. These power systems and their harmonics may be considered as local EM noise. Their high harmonics in the kilo-hertz range (produced by mechanical imperfections, nonlinear or unbalanced loads, etc.) can be radiated by transmission lines and leak into the magnetosphere with sufficient intensity to stimulate wave growth and emission generation, in the same manner as the VLF transmitter noise discussed above. However, since power line radiation is much weaker than VLF transmitter noise, observations of power line radiation effects in the frequency range of about 0.3 to 50 kHz depend strongly on amplification in the magnetosphere (Park, 1982).

1.2 Survey of EM noise reduction techniques

The general way to improve S/N ratios in TEM methods is exponential averaging or digital integration of successive transients (Macnae et al., 1984). The resultant reduction in noise is approximately proportional to the square root of the number of stacks. In quiet areas, very good data quality can be obtained in a relatively short time, whereas in very noisy environments, extensive stacking is needed and often it is virtually impossible to reduce noise to acceptable levels.

Buselli and Cameron (1995, 1993 and 1992) have developed a sferics rejection algorithm for SIROTEM (Buselli and O'Neill, 1977). In order to decrease the storage requirements, this sferics rejection method uses recursive algorithm for updating an M-estimator of the mean and its standard error each time that a new observation is collected. This method can reduce noise by a factor of about 5. In areas where sferics predominates over geological background signal, this method increases target detection depth by approximately 50%, without any increase in transmitter power.

An adaptive filter which uses the LMS method has been applied for reducing natural, nonstationary magnetic fields in TEM data (Kim and Hohmann, 1992). The LMS adaptive filter is formulated as a steepest descent method (Widrow et al., 1975; Widrow et al., 1976; Widrow and Stearns, 1985). Stacking of the adaptively filtered transient response achieves a mean-square noise reduction of 10 to 15 times greater than that obtained by the simple stacking procedure. The main advantage of the adaptive filter in EM noise cancellation is its ability to learn from the statistics of previous measurements and adjust its coefficients during the filtering process.

For a long offset transient electromagnetic (LOTEM) sounding, two selective stacking techniques which are termed the symmetric rejection and the area-defined rejection have been developed for reducing sporadic noise caused by many different culture sources (Strack et al., 1989). Sporadic noise may be caused by many different cultural sources such as water pumps, electric fences, trains, factories and/or vehicles passing by the receiver. An approach for eliminating this kind of noise with data processing is to consider the statistics of all signals, and to analyse their corresponding amplitude distributions. The first step in both selective stacking schemes is to sort in ascending order the data amplitudes at a given time sample for all transients. Then for symmetric rejection known as the alpha-trimmed mean method, a determined percentage (10 to 40%) of the total number of transients is symmetrically rejected from both ends of the sorted amplitudes. For the area-defined rejection method, amplitude-frequency distributions are calculated by sliding overlapping windows over the sorted amplitude curves for each time sample of all transients. A percentage of the area under each distribution curve, symmetric about the maximum, is calculated and all data within that area are kept.

For EM surveys in areas with coherent cultural EM noise (e.g., power line noise), a local noise compensation (LNC) technique has been developed for a multi-channel TEM acquisition system with dense station spacing (Stephan and Strack, 1990). This technique uses several mobile receivers referenced to a nearby (local) base station. LNC is applied as a prestack processing technique and it is intended for multi-channel systems with close receiver spacings. The main premise for LNC is that the regional noise is highly correlated over a region around the base station and that any other, more localised noise, is small.

Exploration for deep IP targets, particularly in conductive environments, can present severe temporal signal-to-noise problems. These problems are usually related to EM coupling, telluric noise (Sumner, 1976; Vozoff, 1984), and weak signal voltages. Halverson (1982 and 1990) and Halverson et al. (1987) have presented a method for real-time telluric noise cancellation in broad-band IP exploration. In one mode, an in-line telluric dipole, hard-wired to the receiver, provides cancellation for the pole-dipole array. In another mode, remote orthogonal telluric dipoles, which transmit telluric data to the IP receiver by radio telemetry, provide cancellation for the pole-dipole or dipole-dipole arrays. Berdichevskiy (1965) describes a method for telluric noise cancellation in IP data, whereby the telluric voltage on a random-oriented dipole can be predicted by a linear combination of telluric voltages measured with orthogonally-oriented dipoles located several or more kilometres away. In real-time cancellations, the potential-dipole and telluric-dipole voltages for successive segments of a wavetrain are averaged into half-period representations, labelled "stacks". Using a least-square algorithm, the telluric and potential-dipole stacks are combined to yield telluric-cancelled potential-dipole stacks.

Nichols and Morrison (1988) have claimed that the application of a remote reference receiver in a controlled source FEM survey reduces the natural magnetic fields by ~40 to 60 dB without increasing transmitter power and averaging time, because the fields are coherent over large distances. They use a least-square fit method to allow for correction of calibration and orientation errors of sensors as well as the removal of apparent fields originating from sensor movement.

A local noise prediction filter (LNPF) based on the tipper method (Vozoff, 1972) for noise reduction has been designed to predict the vertical magnetic field from simultaneous measurements of the horizontal magnetic fields (Spies, 1988). For an in-loop or central induction sounding over a horizontally-layered earth, the TEM signal exists wholly within the vertical component. Thus the predicted time series obtained from an LNPF is the predicted vertical-component EM noise which can be subtracted from the measured vertical component in subsequent processing. The magnetic transfer function in the time domain is solved by a least-square method using the approach of McMechan and Barrodale (1985). For low-frequency noise, the LNPF with a three-point prediction filter reduces the vertical magnetic field by a factor of 5 in amplitude.

The remote reference cancellation scheme with a simple, single component, analogue-bucking system in controlled source EM soundings has been shown to suppress natural magnetic fields by 20 dB (Wilt et al., 1983). The EM-60 (a frequency-domain system using three-component detection) was used for this survey. They applied ±65 A to a 100 m diameter four-turn horizontal loop, generating a dipole moment > 10^6 A-m^2 over the frequency range of 10^{-3} to 10^3 Hz. With such a transmitter and a remote magnetic reference, soundings were made at transmitter-receiver separations of up to 4 km over the frequency range of 0.05 to 500 Hz. A remote receiver (SQUID magnetometer) was placed at a location far enough (usually about 10 km) from the transmitter loop so that the measured remote fields would consist only of the geomagnetic variations. Measurements of the horizontal magnetic fields at the remote station were transmitted to the mobile receiver station from the remote station via FM radio telemetry. Before adding the transmitted remote fields to the local fields in order to reduce the geomagnetic variations, the remote fields were inverted and adjusted in amplitude.

1.3 Neural networks

Neural networks are considered by some to be computer analogues to the brain's neural functions. They provide an approach which is designed to be closer to human perception and recognition than traditional computing and are intended to combine the efficiency of human perception with the precision and speed of a computer. Neural networks have shown capabilities for handwriting recognition (Denker et al., 1989; Huang and Lippmann, 1988), pattern recognition (Ahalt et al., 1989; Alvelda and Martin, 1989; Lee et al., 1990; Pao, 1989; Widrow et al., 1988), time series prediction (Jasic and Poh, 1995; Lowe and Webb, 1989; Weigend et al., 1990a, b), fault detection and diagnosis (Malkoff and Cohen, 1990), signal and speech processing (Malkoff, 1989; Lapedes and Farber, 1987; Lippmann and Beckman, 1989; Principe and Zahalka, 1994), dynamic control system (Gu et al., 1993; Pao et al., 1992), noise cancellation (Masters, 1993; Park, 1990; Tamura and Waibel, 1988), frequency-domain EM inversion problem (Poulton and Birken, 1995; Poulton et al., 1992a, b), inverse problem of downhole TEM data (Schmidt and Cull, 1995), inversion of seismic waveforms (Roth and Tarantola,1992), predicting lithology from P-wave and S-wave velocities, (Lorenzetti, 1992), trace editing of seismic data (Cary and Upham, 1992), picking peaks defining reflection events (Kemp et al., 1992) and feature recognition and lineaments analysis from potential fields (Guo et al., 1992; Fossati et al., 1992).

In this thesis, the focus on neural networks is system modelling to approximate any transfer function for EM noise reduction. Such neural networks attempt to model an arbitrary transfer function or mapping function, since an arbitrary decision surface can be formed in a multi-layer neural network with any continuous sigmoid nonlinearity (Cybenko, 1989; Lippmann, 1987).

The backpropagation algorithm, described in Section 1.3.4, as a supervised learning procedure is a generalisation of the least squares procedure to train multi-layer networks (Rumelhart et al., 1986a, b). Even though backpropagation for training networks has a strong mathematical foundation which is capable of producing very rich results, it has several disadvantages. First, there is no guarantee that the network will converge in a finite number of steps. Second, the phenomenon of paralysis, in which weights take on large negative values and learning ceases, can be a severe problem unless very small learning rates are used (Rumelhart et al., 1986a, b). However, small learning rates can greatly increase training times. Third, backpropagation networks can be trapped in local minima on highly convoluted error surfaces (Gori and Tesi, 1992; Sutton, 1986; Rumelhart et al., 1986a). Fourth, backpropagation networks are not well suited to temporal problems. i.e., if the network faces a continuously changing environment where it may never see the same input pattern twice, the training process may never converge (Wasserman, 1989). For example, if a network is learning to recognise the alphabet, it does no good to learn "B" if, in so doing, it forgets "A".

Many researchers have introduced improvements and extensions to the traditional backpropagation algorithm to solve the aforementioned disadvantages. Since the literature is far too extensive to explain here, the reader is referred to references for details (Ahmad and Tesauro, 1989; Bahl et al., 1987; Brent, 1991; Becker and Le Cun, 1988; Chan and Fallside, 1987; Fahlman, 1988; Fahlman and Lebiere, 1989; Jacobs, 1988; Moody, 1989; Moody and Darken, 1988 and 1989; Park et al. 1991; Pineda, 1988; Stornetta and Huberman, 1987; Sutton, 1986).

1.3.1 Biological neuron

Nerve cells, called neurons, are the fundamental elements of the central nervous system. The central nervous system is made up of about 5 billion neurons. Neurons possess a number of points in common with other cells in their general organisation and their biochemical systems, but they also possess a number of distinctive characteristics (Aleksander and Morton, 1989; Davalo and Naim, 1991).

A neuron is built up of three parts: the cell body, the dendrites, and the axon (Figure 1.2). The body of the cell contains the nucleus of the neuron and carries out the biochemical transformations necessary to synthesise enzymes and other molecules necessary to the life of the neuron. Its shape, in most cases, is a pyramid or a sphere. The shape often depends on its position in the brain, and most neurons in the neocortex have a pyramid shape. The cell body is some microns in diameter.

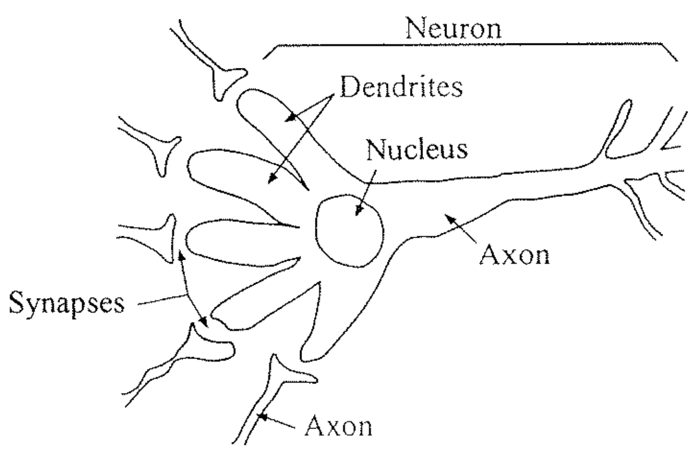

Figure 1.2 Biological neuron.

Each neuron has a hair-like structure of dendrites around it. These are fine tubular extensions some tenths of a micron across and tens of microns in length. They branch out into a tree-like form around the cell body. The dendrites are the principal receptors of the neuron and serve to connect its incoming signals.

The axon or nerve fibre is the outgoing connection for signals emitted by the neuron. It differs from dendrites in its shape and by the properties of its external membrane. In general, the axon is longer than dendrites, varying from a millimetre to more than a metre in length. It branches at its extremity where it communicates with other neurons, while the branching of dendrites takes place much closer to the cell body.

Neurons are connected one to another in a complex spatial arrangement to form the central nervous system. As shown in Figure 1.2, the connection between two neurons takes place at synapses, where they are separated by a synaptic gap of the order of one-hundredth of a micron.

1.3.2 Neuron operation and artificial neuron

A simple description of the operation of a neuron is that it processes the electric currents which arrive on its dendrites and transmits the resulting electric currents to other connected neurons using its axon. The classic biological explanation of this processing is that the cell carries out a summation of the incoming signals on its dendrites. If this summation exceeds a certain threshold, the neuron responds by issuing a new pulse which is propagated along its axon. If the summation is less than the threshold, the neuron remains inactive. The pulse which is propagated between different neurons is therefore an electrical phenomenon. An artificial neuron used in neural network methods is based on the biological neuron operation explained above. Figure 1.3 shows an artificial neuron. This artificial neuron or processing element (PE) has four important components:

- Input connections (synapses), through which the artificial neuron receives activation from other artificial neurons;
- Summation function that combines the various input activations into a single activation;

- Threshold function that converts this summation of input activations into output activation;
- Output connections (axonal paths) by which a neuron's output activation arrives as input activation at other neurons.

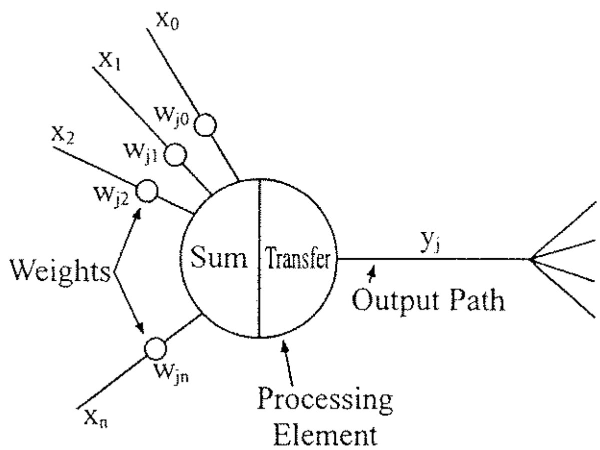

Figure 1.3 Artificial neuron.

1.3.3 Multi-layer neural network architecture

An artificial neural network consists of many PEs (artificial neurons) joined together. The PEs are usually organised into groups called layers. A typical network consists of a sequence of layers with full or random connections (called connection weight in Figure 1.4) between successive layers. There are typically two layers (like buffers) to be connected to the outside of a network: an input layer where data is presented to the network, and an output layer which holds the actual response of the network to a given input. Layers between the input and output layers are called hidden layers (Figure 1.4). The reader is referred to several references on neural networks for more details (Aleksander and Morton, 1989; Anderson and Rosenfeld, 1989; Caudill, 1988; Davalo and Naim, 1991; Dayhoff, 1989; Hinton, 1989; Lippmann, 1987; McClelland and Rumelhart, 1988; Rumelhart et al., 1986a, b; Wasserman, 1989).

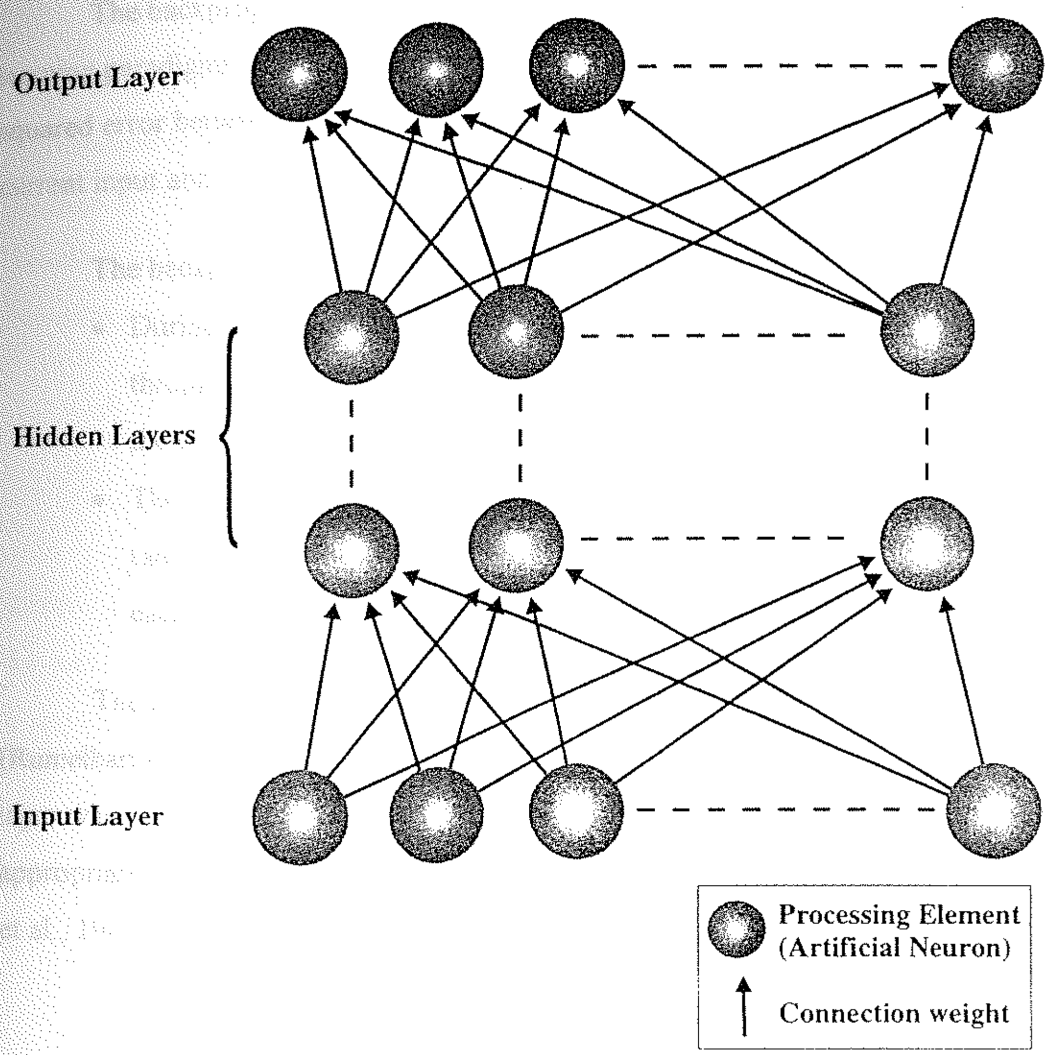

Figure 1.4 Configuration of a typical multi-layer neural network.

1.3.4 Learning by backpropagation

The backpropagation learning procedure has become one of the most popular learning paradigms. It is a supervised learning paradigm that seeks to minimise the squared error between a desired output and the actual network's output given by its current input and network connection strengths.

The backpropagation learning procedure involves two phases:

- During the first phase the input is presented and propagated forward through the network, and is compared with the desired output to produce the error for each output PE;
- The second phase involves a backward pass through the network (analogous to the initial forward pass) during which the error is passed to each PE in the network and the appropriate weight changes are made.

The error at the output layer can be easily calculated using the delta rule (Rumelhart et al, 1986a, b) but the error cannot be so easily computed at the hidden layer where the desired output is unknown. For a network with hidden layers, the appropriate weight changes are calculated from the generalised delta rule (Rumelhart et al., 1986a, b; Werbos, 1974).

The net output of PE j for a pattern p (which is one input set of training or testing data) can be calculated as follows:

$$a_{pj} = \sum_i \omega_{ji} O_{pi} \ , \qquad (1.1)$$

where ω_{ji} is the connection weight from PE i to PE j and O_{pi} is the activation value of the net output from PE i for pattern p. The activation value from PE j for pattern p, $f(a_{pj})$, is given by:

$$O_{pj} = f(a_{pj}) \ , \qquad (1.2)$$

where f is a non-linear transfer function such as a sigmoid function. The weight changes at the output and hidden layers are made as follows:

$$\Delta\omega_{pji}(n+1) = \eta\left(\delta_{pj} O_{pi}\right) + \alpha\Delta\omega_{pji}(n) \ , \tag{1.3}$$

where $\Delta\omega_{pji}$ is the change to be made to the weight from PE i to PE j following presentation of a training pattern p, n is the presentation number, δ_{pj} is the error gradient at PE j, η is the learning rate, and α is the momentum rate which determines the effect of the past weight change on the current direction of convergence movement in error weight space.

δ_{pj} at the output layer is calculated by:

$$\delta_{pj} = \left(t_{pj} - O_{pj}\right)\frac{\partial f(a_{pj})}{\partial\omega_{ji}} \ , \tag{1.4}$$

where t_{pj} is the desired output for PE j for pattern p, O_{pj} is the actual output from PE j for pattern p, and $\partial f(a_{pj})/\partial\omega_{ji}$ is the derivative of an activation function with respect to a connection weight from the hidden PE i to the output PE j.

δ_{pj} at the hidden layer is calculated by a recursive equation that propagates the error from the output layer to the hidden layer:

$$\delta_{pj} = \frac{\partial f(a_{pj})}{\partial\omega_{ji}}\sum_{k}\delta_{pk}\,\omega_{kj} \ , \tag{1.5}$$

where ω_{kj} is the weight connecting PE k back to PE j and $\partial f(a_{pj})/\partial\omega_{ji}$ is the derivative of an activation function with respect to a connection weight from the previous layer's hidden PE i to the present hidden layer's PE j.

1.4 Outline of thesis

The contents of this thesis are summarised as:

where f is a non-linear transfer function such as a sigmoid function. The weight changes at the output and hidden layers are made as follows:

$$\Delta \omega_{pji}(n+1) = \eta\left(\delta_{pj} O_{pi}\right) + \alpha \Delta \omega_{pji}(n) \ , \tag{1.3}$$

where $\Delta\omega_{pji}$ is the change to be made to the weight from PE i to PE j following presentation of a training pattern p, n is the presentation number, δ_{pj} is the error gradient at PE j, η is the learning rate, and α is the momentum rate which determines the effect of the past weight change on the current direction of convergence movement in error weight space.

δ_{pj} at the output layer is calculated by:

$$\delta_{pj} = \left(t_{pj} - O_{pj}\right) \frac{\partial f(a_{pj})}{\partial \omega_{ji}} \ , \tag{1.4}$$

where t_{pj} is the desired output for PE j for pattern p, O_{pj} is the actual output from PE j for pattern p, and $\partial f(a_{pj})/\partial \omega_{ji}$ is the derivative of an activation function with respect to a connection weight from the hidden PE i to the output PE j.

δ_{pj} at the hidden layer is calculated by a recursive equation that propagates the error from the output layer to the hidden layer:

$$\delta_{pj} = \frac{\partial f(a_{pj})}{\partial \omega_{ji}} \sum_k \delta_{pk} \omega_{kj} \ , \tag{1.5}$$

where ω_{kj} is the weight connecting PE k back to PE j and $\partial f(a_{pj})/\partial \omega_{ji}$ is the derivative of an activation function with respect to a connection weight from the previous layer's hidden PE i to the present hidden layer's PE j.

1.4 Outline of thesis

The contents of this thesis are summarised as:

Chapter I describes general EM noise sources, conventional EM noise reduction methods, the general description of a multi-layer neural network, and the backpropagation learning algorithm.

Chapter II describes instrumentation for sferics measurements. The correlation relationship between the measured X and Y components, the Z and X components, and the Z and Y components of sferics pulses recorded at a single station, and the spatial correlation of the X, Y, and Z components of sferics pulses recorded simultaneously at two stations are discussed.

Chapter III presents a model training method (MTM) to realise a mapping from a set of noisy transients to a set of noise-free transients. The advantages and disadvantages of this approach to improve S/N ratio in TEM and AEM methods without increasing source power and intensely stacking successive transients are discussed.

Chapter IV explains a neural network-based local noise prediction filter (LNPF) to reduce the vertical component of EM noise fields (high-frequency sferics, VLF EM noise, and geomagnetic field variations below 1 Hz) in an in-loop TEM geometry.

Chapter V describes a remote noise prediction filter (RNPF) for TEM and AEM methods using a remote reference receiver for minimising high-frequency sferics.

Chapter VI presents comparisons of neural-network-based EM noise reduction methods with other methods as follows:

- Comparison of an MTM with least-square fitting and simple stacking methods;
- Comparison of a neural network-based LNPF with the tipper method;

- Comparison of an RNPF with simple subtraction, autoregressive moving average, and interpolation methods.

Chapter VII shows examples of the applications of neural network filters to EM field data as follows:

- Application of noise prediction filters to AEM data (SALTMAP data);
- Application of noise prediction filters to ground-based TEM data obtained in exploration conditions at a mineral prospect near Parkes, NSW.

Finally, Chapter VIII presents a general discussion of the results and conclusions obtained in earlier chapters.