Indian Journal of Pure & Applied Physics
Vol. 60, October 2022, pp. 849-854
DOI: 10.56042/ijpap.v60i10.63915

Investigation on the Classification Accuracy of Harmonic Signal Detection by Artificial Neural Network

Archana Bora*
Department of Applied Sciences, Gauhati University, Guwahati, Assam 781 014, India

Received 18 June 2022; accepted 16 September 2022

Harmonic signals are produced by many natural sources as well as man-made sources. The detection of the harmonic signal can reveal a lot about the physical system. Various Fourier harmonic analysis based numerical methods are commonly used to detect such signals. However, recently there has been considerable interest in other non-Fourier-based methods as well, to determine the harmonics. In the present work we have studied the application of artificial neural network based machine learning in frequency identification of sinusoidal signals. We considered training sets comprising harmonic signals with randomised phase or randomised amplitude or combination of the both. Based on such training sets the trained network is then applied to detect the frequency of unknown harmonics. Here we performed an investigation on the relative advantages of the network trained using sets of harmonic signals with different features.

Keywords: Artificial neural network (ANN); Harmonic signals; Detection accuracy

1 Introduction

Many signals associated with real life problems exhibit periodic components that repeat at fixed frequency or at an integral multiple of the main frequency, called the fundamental, throughout the signal duration. On the other hand, a function of time, say $f(t)$, corresponding to any arbitrary signal data, with a given periodic variation, can in general be expressed as a weighted sum of periodic sinusoidal components written as

$$f(t) = a_0 + \sum_{i=1}^{N} A_i \sin(\omega_i t + \epsilon_i) \qquad \ldots (1)$$

This is also known as the Fourier series expansion of the signal[1]. Here a_0 is the mean co-efficient, N is the total number of constituents, A_i, ω_i and ϵ_i are the amplitude, angular frequency and phase, respectively, of the i-th constituent. Each of these N sinusoidals can be termed as a harmonic with ω_i as fundamental harmonic. Sinusoidal signals, such as the gravitational continuous wave produced by pulsars[2], by unknown neutron stars in binary systems[3] or harmonic signals occurring in musical instruments or various machineries[4], irrespective of the sources, require their detection for any further analysis. Traditionally, Fourier methods that include Fourier transform[4] and fast Fourier transform[5], are used for detection of the harmonics. However, there exists other methods of detection such as use of Josephson Junction[6] and Hopf oscillators[7]. In the present work we have studied the application of artificial neural network (ANN) based machine learning for automated frequency identification, also referred to as classification, of sinusoidal signals. We have considered detection performance of the network trained using training data sets with randomised amplitude, phase and a combination of the two. Our investigation has revealed the optimal strategy for training to achieve the best classification accuracy of harmonics.

The paper begins with a brief discussion on ANN followed by a description on the generation of training and testing sets. We then perform harmonic detection on about 500 different signals. The paper concludes with a discussion on the results obtained.

2 Methodology

Artificial Neural Network Architecture

The Artificial Neural Network or (ANN) is a computational tool made up of number of interconnected processing units called *neurons.* The neurons are distributed in layers parallel to each other and process information from external inputs with their dynamic state response. There are three different types of layers, called *input layer*, *hidden layers* and *output layer.* The network receives information from the environment via the *input layer* and process the

*Corresponding authors: (Email: abora.80@gauhati.ac.in)

data in the *hidden layers*. The neurons of the hidden layers inter-connect the input and output layers through designated connections called *weights*. The strength of the interconnection between neurons determines the performance of the ANN. At the beginning each neuron of ANN is assigned some random weight value. However, during the training session these values change iteratively so that at the end of the session the output of the network becomes closer to the expected correct value for a particular set of input values. This is known as the *learning process* of the network[8]. Because of its ability to learn from the external environment, ANN has recently found widespread applications in many different disciplines such as for mathematical modelling[7], controlling systems[9], classification schemes [10-14], parameter finding[15], as decision making tools in the field of medical[16] as well as for many industrial application[17].

The basic structure of an ANN is depicted in Fig. 1. The neural network Architecture considered in the present work is a *supervised* neural network with back propagation algorithm[11-14]. In supervised neural network, the training data of the network has both input data and the expected output of it. The input data is fed into the input layer of the network. The ANN then process the data in the hidden layers and estimates the outputs in output layer. In the next stage the algorithm compares the outputs thus formed with the expected outputs and minimises the error between them iteratively.

In the present work we have considered two hidden layers to train the network and the number of nodes in the hidden layers is kept as 64.

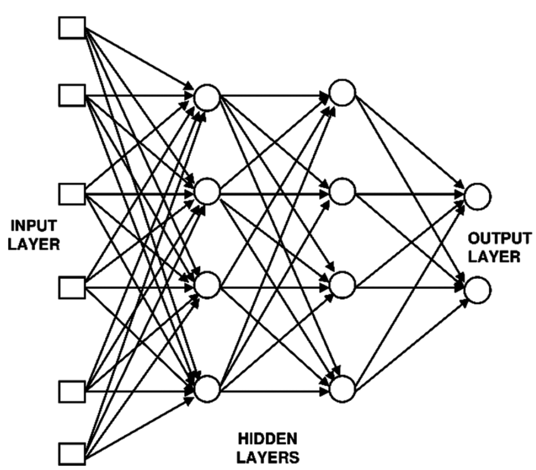

Fig. 1 — Schematic of a multilayer-layer feedforward ANN architecture.

3 Generation of Train and Test Data Set

We have used three independent sets of different harmonics as training data for the network. The fundamental harmonic is considered to be a sinusoid, $\sin(\omega t+\epsilon)$, with unit amplitude and phase zero (*i.e.* ϵ=0), while the other harmonics have frequency in the range of 2ω to 10ω. As each frequency of the harmonics can have arbitrary amplitude and phase in reality, so we can have separate cases of training data sets as described below:

Case I-Variation in Amplitude

In the first train set we have allowed the amplitude to have random values in the range of 0.001 to 1.0. We thus write

$$y(t) = A_j\, sin(\omega_i t + \epsilon) \quad \dots (2)$$

Here the amplitude A_j, for a particular value of frequency ω_i, takes 500 unique yet random values as $j = 1,2,\dots\dots,500$, keeping phase ϵ constant. On the other hand $\omega_1 = \omega$ represents the fundamental frequency and $\omega_{i=2,3,,\dots,10}$ represents the harmonics $2\times\omega$, $3\times\omega$ and so on, up to $10\times\omega$. Thus for $i = 1,2,\dots\dots,10$ we have a total of 5000 unique sinusoids with different amplitudes as train set.

Case II-Variation in Phase

In the second case, we have allowed the phase to have random values in the range of 0.05^0 to 58^0. Thus we write

$$y(t) = A\, sin(\omega_i t + \epsilon_j) \quad \dots (3)$$

where A is the constant amplitude and for a particular value of frequency ω_i, the phase ϵ_j takes 500 unique yet random values as $j = 1,2,\dots\dots,500$. Therefore for $i = 1,2,\dots\dots,10$ we have again a total of 5000 unique sinusoids with different phases as train set. However, in both the above two cases the number of classes is equal to the number of harmonics which is equal to 10. Moreover, each harmonic signal is generated in such a way that each has 200 data points individually.

Similarly, as the test sets we have generated 5000 other sinusoids with 500 different phase values as well as another 5000 sinusoids with 500 different amplitude values. However, in both the cases, although the amplitude and phase values of the test set lie in the same range as those of the training sets yet they are generated independently from those in the training set.

Case III-Variation in both Amplitude and Phase

Additionally, we have generated one more each of training and test set consisting of 500 number of signals where both the amplitude and phase are allowed to change simultaneously and randomly for each harmonic with the amplitude varying in the range of 0.001 to 1.0 and the phase varying in the range of 0.05° to 58°.

4 Results

At the end of the training session, the ANN completes the *learning* process on the given data set. At this stage, the *weight values* of the connections get frozen and the ANN is now ready to be used on the test data sets. As mentioned in section (3), the present work uses test data sets generated by (i) either amplitude variation (when network is also trained with data comprising of harmonics with random variation in amplitude) or phase variation (when network gets trained with data comprising of harmonics with random variation in phase) (type-i) and (ii) simultaneous variation of amplitude and phase of the signals (type-ii). Accordingly, the harmonic frequency of the signals as measured by three differently trained network for different test sessions are presented in the following. Here, we have adopted 3D plots for better visual presentation of the classification or identification results of the harmonics by the ANN. In the plots among the two horizontal axes, the *input frequency* axis represents the frequency, say ω_l, of the generated test signal while the other *estimated frequency* axis represents the frequency, ω_m , of the same signal as predicted by the trained ANN. The horizontal axes labels are actually equal to (normalised frequency) × 1000. The vertical axis represents the number of signals (N), scaled as $N^{1/2}$, lying at a particular point *(l,m)* of the *input frequency- estimated frequency* plane in the plot. Accordingly, Fig. 2, Fig. 3 and Fig. 4 present the ANN test results in 3D form for Case I, II and III of trained ANN, respectively.

Figure 2(a) presents the classification result for test data type-i where the test data contains only amplitude variation while Fig 2(b) presents the classification result for test data type-ii having contribution to the signal from both amplitude as well as phase variation. Here as Case I, the network was trained with signals comprising of random variation of amplitude only. Similarly, Fig 3(a) presents the classification result for test data type-i where the test data contains only phase variation while Fig 3(b) presents the classification result for test data type-ii having contribution to the signal from both amplitude as well as from phase variation. Here as Case II, the network was trained with signals comprising of random variation of phase only.

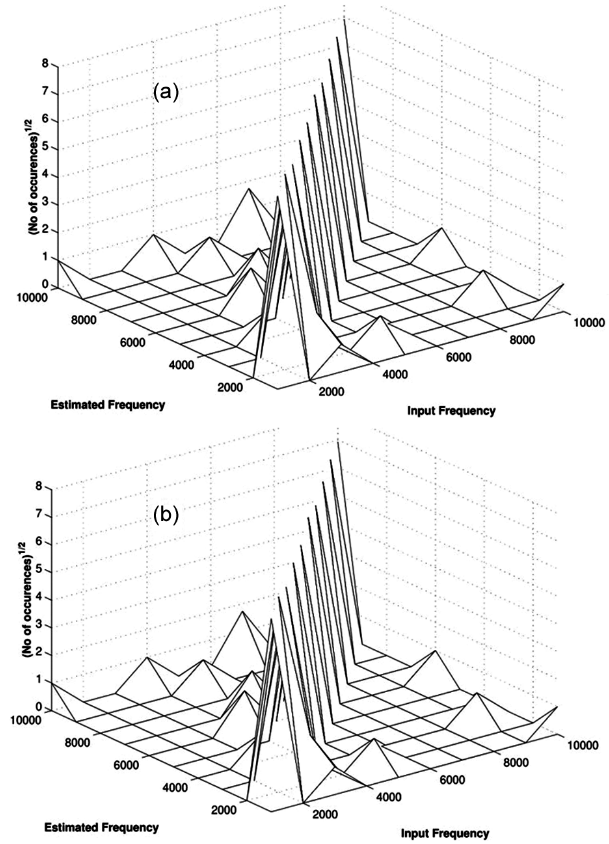

Fig. 2 — Scatter 3D plots of classification result for Case-I training session. (a) ANN tested with signals having only random variation in amplitude and (b) ANN tested with signals having only random variation in both amplitude and phase.

Lastly, Fig. 4 presents the classification result using Case III type training of ANN and the testing data set have signals with randomised amplitude as well as phase values.

It is important to mention here that in all the above cases, the class of a given harmonic signal was encrypted with harmonic and amplitude or phase information using the following formula as

$$class - value = f_{harmonic} \times 1000 + \frac{phase\ or\ amplitude}{100} \quad \ldots (4)$$

In the above formula the highest weightage value has been assigned to the frequency than the random phase or amplitude added to the signal. This is because the work presented here is investigating only

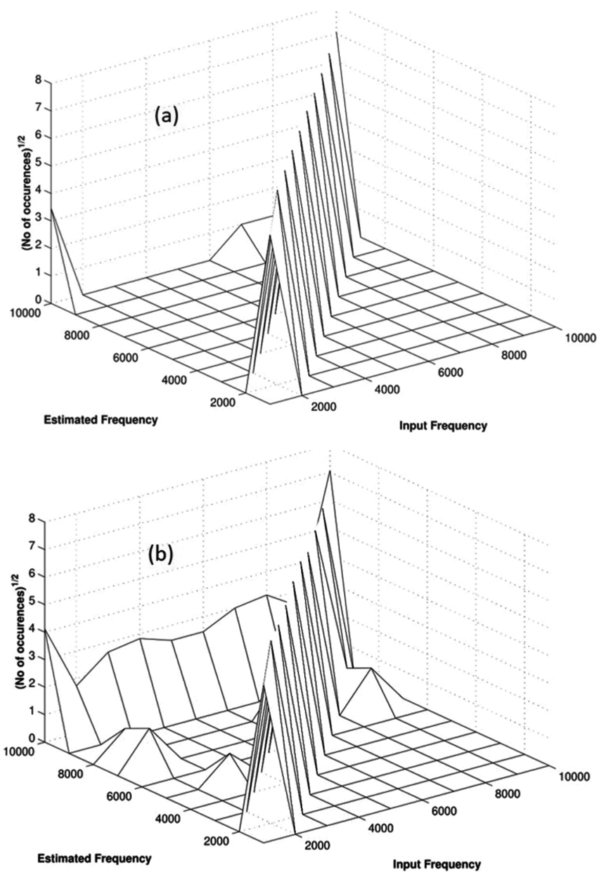

Fig. 3 — Scatter 3D plots of classification result for Case-II training session. (a) ANN tested with signals having only random variation in phase, and (b) ANN tested with signals having only random variation in both amplitude and phase.

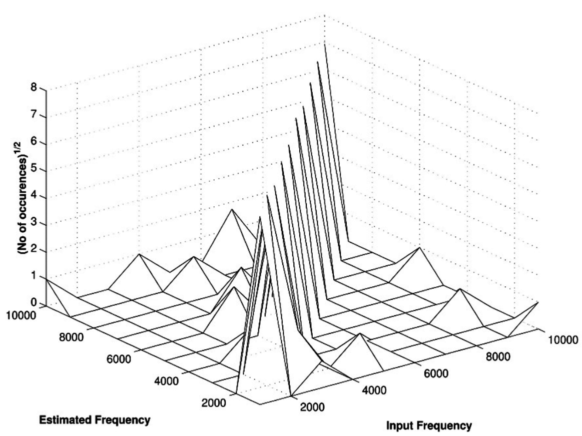

Fig. 4 — Scatter 3D plots of classification result for Case-III training session.

the ANN's ability to estimate the frequency of unknown sinusoidal signal rather than estimating the phase or the amplitude of the signals. Also, in the 3D plots, the vertical axes are scaled as $N^{1/2}$, where N is the number of signals present for a particular frequency. It is because taking the square root of the actual number N helps in better visualisation as otherwise in the cases where this number is relatively small, the corresponding points would appear too small in the plots. The numbers on the horizontal axes of the figures above in fact refer to the coded class values for frequency of the harmonics as adopted in the present work.

5 Summary and Discussion

The performance of the ANN classification of harmonic signals has been summarised in the form of confusion *matrices* and have been presented in the Table-1, 2 & 3 for the three different categories of network training as mentioned in section 4. In these tables, the horizontal rows represent the actual frequencies of the fundamental and different harmonics (normalised by the fundamental) while the vertical columns represent the respective frequencies as predicted by the ANN.

Table-1 corresponds to the Case-I of ANN training with random amplitude data. Any entry *p(q)* to (*l,m*)-th cell of this table, indicates that *p* number of signals with actual frequency ω_l got classified as ω_m when the ANN was tested with random amplitude data. On the other hand *q* indicates the number of signals with actual frequency ω_l getting classified as ω_m when the ANN was tested with data generated by simultaneous random variation of both amplitude and phase values of the signal. Similarly, Table-2 corresponds to the case-II of ANN training with random phase data. In this table, any entry *p(q)* to (*l,m*)-th cell of Table-2, indicates that *p* number of signals with actual frequency ω_l got classified as ω_m when the ANN was tested with random phase data. While *q* indicates the number of signals with actual frequency ω_l getting classified as ω_m when the ANN was tested with data generated by simultaneous random variation of both amplitude and phase values of the signal.

Lastly, Table-3 presents the prediction of the network for case-III training, when the ANN was trained with simultaneous random variation of phase and amplitude data set and unlike the other two cases, this third category of network was tested with only one test data set. The test data set, in this case, contains simultaneous random variation in both amplitude and phase like the training data set.

The entire classification accuracy of the network for various harmonics has been summarised in Table 4. The classification accuracy for each of the

Table 1 — Confusion matrix when ANN trained with random Amplitude data set. The number of occurrences appear in the form of p(q), where p is the number of classified events when tested with random Amplitude data and q is the same when tested with random Amplitude and Phase data.

Predicted $\omega_j \Rightarrow$ (N= 500) Actual ω_i $\Downarrow$	1	2	3	4	5	6	7	8	9	10
1	49(49)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	1 (1)
2	0 (0)	50 (50)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
3	1 (1)	3 (3)	45 (45)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
4	0 (0)	0 (0)	0 (0)	47(47)	0 (0)	2 (2)	0 (0)	0 (0)	0 (0)	1 (1)
5	0 (0)	0 (0)	1 (1)	0 (0)	46(46)	0 (0)	2 (2)	0 (0)	1 (1)	0 (0)
6	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	50 (50)	0 (0)	0 (0)	0 (0)	0 (0)
7	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	46 (46)	1 (1)	0 (0)	3 (3)
8	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	48 (48)	2 (2)	0 (0)
9	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	49 (49)	0 (0)
10	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	48 (48)

Table 2 — Confusion matrix when trained with random phase data set. The number of occurrences appear in the form of **p(q)**, where **p** is the number of classified events when tested with random Phase data and **q** is the same when tested with random Amplitude and Phase data.

Predicted $\omega_m \Rightarrow$ (N= 500) Actual ω_l $\Downarrow$	1	2	3	4	5	6	7	8	9	10
1	38(33)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)	12 (17)
2	0 (0)	50 (44)	0 (0)	0 (0)	0 (0)	0 (0)	0 (2)	0 (1)	0 (0)	0 (3)
3	0 (0)	0 (0)	50 (42)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (7)
4	0 (0)	0 (0)	0 (0)	50 (42)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (8)
5	0 (0)	0 (0)	0 (0)	0 (0)	50 (44)	0 (0)	0 (0)	0 (0)	0 (0)	0 (6)
6	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	50 (44)	0 (0)	0 (0)	0 (0)	0 (6)
7	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	49 (39)	0 (2)	0 (0)	1 (9)
8	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	49 (40)	0 (0)	1 (10)
9	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (2)	0 (1)	0 (0)	1 (6)
10	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	50 (50)

Table 3 — Confusion matrix when trained with random Amplitude and Phase data set.

Predicted $\omega_m \Rightarrow$ (N= 500) Actual ω_l $\Downarrow$	1	2	3	4	5	6	7	8	9	10
1	49	0	0	0	0	0	0	0	0	1
2	0	50	0	0	0	0	0	0	0	0
3	1	3	45	1	0	0	0	0	0	0
4	0	0	0	47	0	2	0	0	0	1
5	0	1	0	0	46	0	2	0	1	0
6	0	0	0	0	0	50	0	0	0	0
7	0	0	0	0	0	0	46	1	0	3
8	0	0	0	0	0	0	0	48	2	0
9	0	0	1	0	0	0	0	0	49	0
10	1	0	0	0	0	1	0	0	0	48

cases of training and testing are considered in terms of the root mean square (RMS) error. A classification RMS error of 1000 indicates that a harmonic of frequency, say 2ω can at worse, be classified either as 3ω or 1ω as suggested by Eq. (4). As such, from the first column of the Table 4, we can say that when ANN training set is uniquely identified by the amplitude value of the sinusoids, it shows better

Table 4 — Summary of Classification Results.

Train Data Parameter	Test Data Parameter	R.M.S.E	Detection accuracy in %
Amplitude	Amplitude	811.5	95.6
	Amplitude +Phase	811.5	95.6
Phase	Phase	1387	97
	Amplitude +Phase	2175	83.6
Amplitude +Phase	Amplitude +Phase	811.5	95

performance in estimating the frequencies of unknown harmonics than when the training set is identified by the phase part of the sinusoids. However, it is seen that when the test data set is allowed to have phase variation in addition to the amplitude variation the classification performance of the ANN trained with phase dependent data set deteriorates. Yet another way of representing the classification accuracy is in terms of detection accuracy in percentage-

(defined as $\frac{No.of\ successfull\ case}{Total\ number\ of\ cases} \times 100$).

Interestingly, when we consider the performance of ANN in terms of the detection accuracy in percentage, it is seen from the second column of the Table 4 that the network trained with unique phase dependent data set performs better than the other cases. This behaviour is contrary to the ANN performance when we consider the performance in terms of RMS error. The overall success percentage is highest when both train and test data sets are determined by phase only.

However, the performance of the network trained with randomized phase data has a significantly reduced accuracy when the test data set has both amplitude and phase variation, which may very likely occur in a realistic scenario. On the other hand, the network trained with random amplitude data set provides good performance for both types of test data set, *i.e*, test data set with random phase and test data set with random phase and amplitude. The performance of this network, is even better compared to the network trained with randomized amplitude and phase when the test data set has both amplitude and phase variation.

From our investigation, it is thus observed that the optimal strategy for detection of harmonic signals, comprising signals with randomised amplitude or both randomised amplitude and phase for a given unknown harmonic, is by using a network trained with a random amplitude data set.

References

1 Chatfield C, The analysis of time series: an introduction, CRC press, 2016.

2 Bruce A, Papa M A & Schutz B F, *Phys Rev D*, 66 (2002) 102003.

3 Covas P B & Sintes A M, *Phys Rev D*, 99 (2019) 124019.

4 Piper J E, Reese S S & Reese W G, *Oceans*, (2012) 1.

5 Marino C S & Chau P M, *IEEE Trans Instr Meas*, 61 (2012) 2413.

6 Filatrella, Giovanni & Pierro V, *Phys Rev E*, 82 (2010) 046712.

7 Maoiléidigh, Dáibhid Ó & Hudspeth A J, *Phys D: Nonlinear Phenomena*, 378 (2018) 33.

8 Haykin S S, *et al*., Simon Haykin, New York, Prentice Hall, 2009.

9 Gu W, Valavanis K P, Rutherford M J & Rizzo A, *J Intel Robotic Syst*, 100 (2020) 1469.

10 Storrie-Lombardi M C, Lahav O, Sodre L & Storrie-Lombardi L J, *Month Not Royal Astronom Soc*, 259 (1992) 8P.

11 Bora A, Gupta R, Singh H P, Murthy J, Mohan R & Duorah K, *Month Not Royal Astronom Soc*, 384 (2008) 827.

12 Bora A, Gupta R, Singh H P & Duorah K, *New Astronomy*, 14 (2009) 649.

13 Gulati R K, Gupta R, Gothoskar P & Khobragade S, *The Astrophys J*, 426 (1994) 340.

14 Gulati R K, Gupta R & Rao N K, *Astron Astrophys*, 322 (1997) 933.

15 Malinov S, Sha W & McKeown J J, *Computat Materials Sci*, 21 (2001) 375.

16 Lisboa P J & Taktak A F G, *Neural Networks,* 19 (2006) 408.

17 Meireles M R G, Almeida P E M & Simoes M G, *IEEE Trans Indus Electron*, 50 (2003) 585.