A hierarchical systems modelling approach based on neural networks for forecasting global waste generation: A case study of Chile

[1]Samarasinghe, S. and Ordonez, E.

[1]Centre for Advanced Computational Solutions, Lincoln University, New Zealand e-mail: sandhya@lincoln.ac.nz

Keywords: *Waste Forecasting, Neural Networks*

EXTENDED ABSTRACT

In this-first every study for Chile, a neural network based hierarchical modelling approach is proposed for forecasting domestic waste generation for the whole country. Over 30 global variables from the 342 communes (municipalities) in the country were analysed extensively using statistical tools that led to 5 significant explanatory variables: population, percentage of urban population, years of education, number of libraries and number of indigents. The five explanatory variables were used to develop a feedforward neural network for predicting volume of global waste generation for a particular year (2002 in this case) in Chile and assessing the contribution of variables. The model had validation R^2 of 0.82.

Effective waste management programmes must be implemented at local level. This requires estimates of waste generated by individual communes. To simplify local modelling process, 342 communes (municipalities) were clustered into 3 groups (Low, Medium and High waste), using self-organising Feature Map (SOM) based on the 5 selected explanatory variables. Next, a search method was developed and implemented to identify the best representative commune for each group. Then, data for the explanatory variables were sourced for the representative communes for a range of years to develop forecast models. The best models for the Low and Medium group representatives were multilayer preceptrons and for the High group was Jordan recurrent network, with validation R^2 of 0.81, 0.91 and 0.98, respectively. Each of the three models was further validated with leave-few-out cross-validation using 3 randomly selected subsets of data for each model.

The developed local representative models were used to forecast total waste generation for the whole country. Specifically, a forecast for each representative commune was first obtained for the 2002 to 2010 period from the above models. Then, using the relationship between waste generation of the representative commune to that of the communes in the group it represents, a waste generation for each represented commune was obtained. These were aggregated to get an estimate for the groups. By combining group forecasts an estimated forecast for the country was obtained. Forecasts up to 2007 were deemed more reliable and on this basis, waste generation in the whole country will increase to 500,000 tonnes/month from 467 000 tonnes per month in 2002.

1. INTRODUCTION

The amount of waste generated in Chile has had a dramatic increase over the last decade. In the period 1996-2002 the total amount of waste generation rose an order of magnitude faster than the population growth which is widely considered to be an indicator of waste growth. Predicting waste generation has been considered important by environmental authorities developing the environmental agenda for the country [1].

2. AIM

The main aim of the research described in this paper is to develop an efficient method to forecast amount of waste generation in a country from the variables affecting waste generation, with Chile as a case study. The aim is to be achieved by designing a hierarchical communal analysis tool based on neural networks and statistical methods to study waste generating factors and to forecast waste generation levels.

3. BACKGROUND

Past researchers have studied the influence of a variety of variables, but population and income are the most widely studied, although with inconclusive results as to the relevance to waste generation. Among some of the variables considered are: household size, residency type, age groups, employment, electricity consumption, tipping fees, CPI, GDP, education, culture, geography and climate.

Population has been considered to be one of the most important variables affecting waste generation [2,3]. However, it has also been found to be of little statistical significance predicting average waste generation rate [4]. Income has also been found to be one of the most influential factors affecting waste generation [4,5,6]. Nevertheless, others have observed no influence of income on waste [7].

In developing relationships between variables and waste generation, most of the researchers have used regression analysis and time-series models for predicting waste generation. In 1974, Grossman et al.'s regression model neither explained nor forecasted waste generation, concluding that waste production occurred independently of the analysed variables and that these were not significant for the assessed community [8]. Ali Khan and Burney (1989) mixed different cities around the world to generate a single explanatory model; however, this approach cannot be justified due to contrasting waste generating conditions in different countries. They concluded that income, temperature and dwelling occupancy rate affected the percentage of waste components [5]. Buenrostro et al. [6] worked with monthly income and number of dwellers per household, but the study concluded that these variables were of limited value in explaining solid waste generation. Bagby et al. [9] developed models as part of Seattle's Solid Waste Plan. They found little growth in waste generation over the forecasted period due to Seattle's characteristics such as a continuing decline in the average household size and trends in the housing markets [9].

Some researchers have worked with Time-Series with better results. In 1986, Bridgwater made projections for up to fifty years, concluding that S-curves give the best results [10]. In 1993, Chang et al. used geometric lag time-series analysis for the period 1981-1990 and found a negative relationship between average waste generation per capita per day and total population, a relationship affected by a period of population mobilisation [4]. Chang and Lin (1997) applied an ARIMA (Auto Regressive Integrated Moving Average) model to time-series data for 12 districts of Taipei City, Taiwan, from 1990 to 1995, with predictions solely based on previous trends in waste generation [11]. Finally, Navarro-Esbrí et al. [11] analysed waste generation using sARIMA (seasonal ARIMA) and a non-linear technique and concluded that both methods give good results in terms of predictive accuracy and cumulative errors.

Koushki and Al-Khaleefi's [13]research on waste prediction in Kuwait related household's solid waste generation to monthly income, family size or to the number of persons employed per household. They concluded that any one of these three variables can forecast waste generation [13]. Chen and Chang [14] developed a grey fuzzy dynamic model for the prediction of solid waste generation in a city in Taiwan. The model depended on an extensive database [14].

4.0 METHODS

In this study, a hierarchical neural network modeling approach is used for forecasting waste generation in a country as presented in Figure 1. The approach takes advantage of the fact that countries are generally divided into several administrative regions. A brief description of the proposed method is as follows: In Chile, there are 342 such regions called communes. The volume of waste generation in each commune and data for a large number (about 30) of potential variables affecting it are collected from national databases. The data are preprocessed using statistical methods to select few significant variables affecting amount of total waste generation.

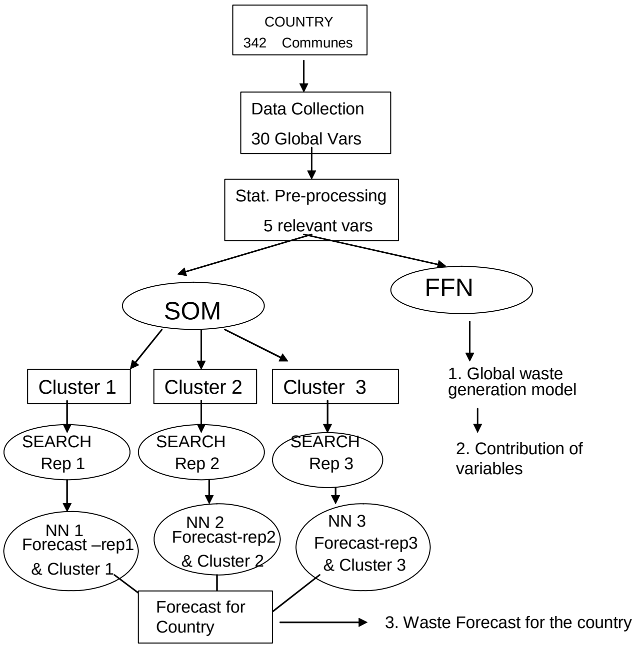

Figure 1. Proposed hierarchical neural network model

The selected variables are used to develop a feedforward neural network (FFN) model to assess their relationship to waste and ascertain the contribution of each variable to waste generation. The selected variables are then used to group the total number of communes into three clusters for efficient development of forecasting models at commune level. This is achieved by unsupervised clustering in SOM. Then a search method is used to select the most representative commune (Rep) from each cluster. Then, data are collected for yearly waste levels and the values for the selected variables for each of the 3 representative communes for developing forecast models. Feedforward and recurrent networks are developed to make 10 year extended forecasts of waste generation for the representative communes. An extended forecast for other communes in a group is obtained through a conversion factor that relates the waste levels of the representative commune in the cluster to that of the commune for which a forecast is sought. By combining forecasts for each cluster, an extended forecast for the country is obtained.

4.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are simplified computational models of the brain [14,15]. Capabilities of ANNs include: pattern classification, clustering, function approximation, forecasting and optimization. An ANN is formed by a large number of processing neurons interconnected by weights. ANNs are classified into feed forward and recurrent networks. In a feed forward network, neurons are grouped into layers and the signals flow from one layer to another in the forward direction. Multi Layer Perceptron (MLP) and radial basis function (RBF) networks are feed forward networks. A typical feed forward network consists of an input layer, a hidden neuron layer and an output layer of neurons. Input layer simply transmits inputs through weighted links to hidden neurons where weighted inputs are accumulated and processed by a transfer function to generate an output to be sent to the output layer. A similar process takes place in the neurons in the output layer where outputs are generated.

In a recurrent network the flow is forward and backwards. In recurrent nets for time series forecasting, outputs of some neurons are fed back to the same or other neurons in preceding layers. The Elman and the Jordan nets are examples of recurrent networks. In Elman networks, hidden layer outputs are fed back to the input layer for processing in the next time step and in Jordan network, output layer output is fed back to the input layer. This feedback helps incorporate temporal effects into recurrent networks.

ANNs are modelled via a learning process which can be supervised or unsupervised. In supervised learning, the network is presented with the inputs and target outputs and the network iteratively adjusts its weights using efficient learning methods such as steepest descent and second order methods. The aim is to minimise the error by generating outputs as close as possible to the targets. Feedforward and recurrent networks are examples of supervised networks. Conversely, unsupervised learning uses no external supervision and clusters the data presented to the network based on the properties of the data in a self-organising manner. An example where unsupervised learning is used is self organizing map (SOM). In an SOM multidimensional data are projected onto a 2-dimensional grid of neurons where similar input vectors form clusters in the course of learning.

In this research, the relationship of the selected variables to waste generation is developed using an MLP, an SOM is used for clustering the communes, and waste forecasting models are developed using MLP and recurrent networks. The software used is NeuroShell 2 by Ward Systems Group®, Inc.[16]. The dataset is divided into three distinct sets. The training set is used to train the network, validation set is used to assess the model at various stages of training and the test set is used to test the model predictions on unseen data (generalisation). In the evaluation process, the best networks are selected based on the highest coefficient of multiple determination (R^2) and the lowest mean squares error (MSE).

4.2 Model development and variable contribution

Global Variables: Based on an extensive literature review, possible waste generating factors were evaluated and a preliminary set of potential global variables selected for data collection from the 342 communes. These variables were grouped according to population, economic, education, dwelling, geographic and waste related indicators as follows: *Population indicators*: Population, Percentage of Urban Population, Population Density, Gender, Age Groups and Native Population. *Economic indicators*: Poverty Level and Income per Household, Economic Activities, Regional GDP, Foreign Investment, Exports, Construction Rate, Vehicles, Employment, Labour Force and Unemployment. *Education indicators*: Years of Education, Cultural Activities, Number of Public Libraries and Illiteracy Rate. *Dwelling indicators*: Number of Houses and Households and Number of People per Household. *Geographic indicators*: Geography and Climate, and *Waste-related indicators*: Waste Generation (tonnes/month), Waste Generation Rate, Per Capita Waste Generation and Existence of Disposal Sites.

Data Collection: Data on global variables was sourced from a number of locations in Chile such as the Central Bank of Chile, the National Institute of Statistics, National Commission for the Environment, and relevant Ministries. Data was available for 2002 only.

Data Pre-Processing: Data was processed searching for multicollinearity and heteroskedasticity. A correlation analysis showed that Waste Generation is highly correlated with Urban Population, Gender, Population, Non-Poor Population, Number of Houses, Age Group and Number of Vehicles. These variables are also highly correlated with each other and therefore, variables with high correlation with the dependent variable and low correlation with the other independent variables were selected. The Breusch and Pagan test detected heteroskedasticity in the data. However, its effect was reduced using the Two-Step Weighted Least Square method. The final selected explanatory variables and their correlation to waste are: Population (POP): 0.875, Percentage of Urban Population (PUP): 0.502, Years of Education (EDU): 0.519, Number of Libraries (LIB): 0.522 and Indigent Population (IND): 0.503.

Relationships Establishment: MLP networks were used to determine the relationship between waste and the selected generating factors as well as their contribution to the variable Waste. The aim was to analyse how the variables impact waste generation across the country. The data were normalized so that they had similar ranges. A three layer MLP network modelled the relationship between the explanatory variables and waste generation with an R^2 = 0.819 and a correlation coefficient equal to 0.915 based on the validation dataset. The architecture of the MLP had five input units, twenty hidden units and one output unit, both hidden and output units using logistic functions. Training was back propagation with momentum. Figure 2 shows actual and the network predicted waste generation indicating an accurate model.

The relative contribution of a variable in predicting waste generation was obtained by assessing its effect on error when its value is held fixed at the mean compared to the error from the original model. These contributions were: 0.413 for POP, 0.169 for LIB, 0.154 for IND, 0.138 for PUP and 0.125 for EDU. All the variables contribute positively to Waste Generation. Results showed a remarkable nonlinear influence of POP. The influence of other variables were only slightly nonlinear [18].

4.3. Clustering of Communes

SOM was used to cluster the 342 communes into three clusters based on the 5 influential variables selected. The SOM net clustered the communes into groups with 91, 156 and 95 communes. The three groups can be seen in the bi-dimensional plot of the 342 communes shown

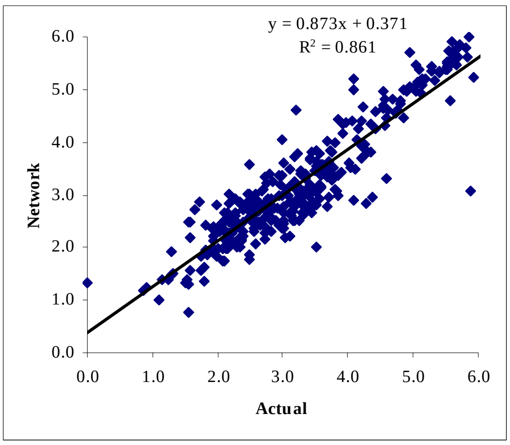

Figure 2: Actual and predicted waste generation for the 342 communes of Chile

in Figure 3, where weighted population is plotted against the weighted sum of the other four variables. Weights represent the percentage contribution of variables to waste generation as found by the MLP network.

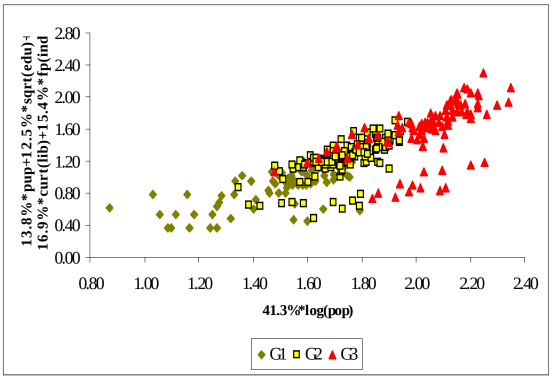

Figure 3. Communes Clustered into Groups

Determining Group Representatives: The most representative communes are those that are close to the largest number of communes in the cluster. In this study, the one that embodies the largest number of communes within a radius of 15% perturbations to its values for explanatory variables was selected using a search process. The selected representative commune of Group 1 covers 44% of the communes of the group (40 communes); the representative of Group 2 covers 75% (117) and the representative from Group 3 covers 76.8% of the communes of the group (73 communes). The representatives together cover 67.3% of the total number of communes (230 communes) within 15% perturbation range of their values for explanatory variables.

Data Collection: The selected representative communes were visited to collect data but all the expected data could not be collected due to lack of information gathering in the country. A new representative from Cluster 1 and 2 had to be selected due to lack of data. Thus real coverage range decreased to 39.6% for Group 1 (36 communes), 38.5% for Group 2 (60) and 74.7% for Group 3 (71). This means decreasing the total number of communes covered from 67.3% to 48.8%, i.e., from 230 to 167 communes.

4..4. Forecasting Waste Generation

Forecasting Waste Generation for the representatives: Several MLPs and recurrent networks were trained to forecast waste generation for the three groups, with the aim of forecasting amounts (and trends) in waste generation for the period up to 2010 from past and current data. In modelling terms, this specifically involves forecasting next year generation from previous year explanatory variables. In a time series, next outcome can be highly correlated with the current outcome (e.g. WG next year may be correlated to WG this year). This is possible because this year's outcome may capture substantially the effects of explanatory variables on the next outcome. However, time series models can be further improved if the explanatory variables are also included in order to capture the aspects that are not accounted for by this year's waste alone (exogenous variables).

Unfortunately, the data for all the explanatory variables was not available for all the past years due to the lack of data collection in Chile. For example, Groups 1 and 2 only had POP and LIB and Group 3 only and data on POP and PUP. Data for EDU and IND could not be obtained for any of the communes.

Time series were analysed using MLP and recurrent networks. Networks were trained using the explanatory variables for which data was available and current per capita waste generation (PCWG). These inputs for the current year were used to forecast waste for the next year. Many networks were tested and the best nets were: Group 1 (MLP) with R^2 0.81, Group 2 (MLP) with R^2 0.91, and Group 3 (Jordan) with R^2 0.98. All the models showed a higher forecasting accuracy for the period for which actual data.

The developed models for each group were validated with leave-few-out cross validation due to limited data. For this, original yearly data sets were randomly divided into 3 sets with 2 years left out from each set. Separate models were trained with each data set and validated with the data for 2 years left out. The validation R^2 was consistent across all models for the three groups. The predictions from the best models were extremely accurate for the period for which actual data was available.

For space limitation, results for only Group 3 are presented in Figure 4 which shows that the model predictions for Group 3 for the period 1992-2002 for which data was available are extremely accurate. In the figure, bars represent actual observed waste. The yearly rate of change of WG peaks at 6% by 2006-2007.

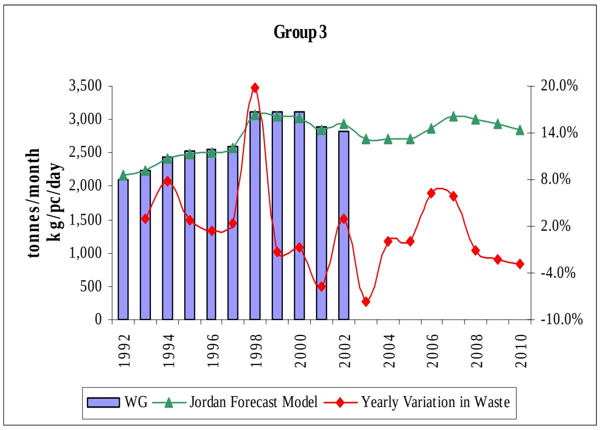

Figure 4: Waste Forecasts up to 2010 for Group 3

The model for the represented commune of Group 1 (not shown) indicates that its waste generation will reach 100 tonnes/month by 2010 from 70 tonnes/month in year 1998. The best MLP model predictions for Group 2 forecasts that the level of waste generation will reach around 240 tonnes/month by 2010 from 150 tonnes/month in 1997.

Estimating waste forecasts for communes and Groups: In order to forecast waste generation for every one of the communes represented in each group, an annual conversion factor (CF_i) is created based on the representative commune's waste generation forecasts (WG_{RC}) and its level of waste generation in the previous year (starting from 2002). The conversion factor is determined by the following equation:

$$CF_i = \frac{WG_{RC,i} - WG_{RC,i-1}}{WG_{RC,i-1}} \qquad [1]$$

where i is year from 2002 to 2010 and RC is Representative Commune.

The estimated level of waste generation for a represented commune j in a year i ($WG_{j,i}$) is determined from its waste generation from the previous year (starting from 2002) and the CF_i for the representative commune for the respective year I as

$$WG_{j,i} = WG_{j,i-1} * (1 + CF_i) \qquad [2]$$

The total waste forecasts for the represented communes in the three groups are shown in Figure 5. In aggregated terms, the three groups of represented communes will behave in different ways for the projected period. From the figure it can be seen that total waste from Group 1 will remain at a similar level up to 2010, total waste from Group 2 will increase steadily and total waste from Group 3 will peak in 2007 and then drop.

In detail terms, the 36 represented communes from Group 1 will increase their total waste generation from 3,400 tones/month to a peak of over 3,800 tonnes/month by 2010. The 60 communes from Group 2 will increase their total waste generation up to over 18,500 tonnes/month by 2010, with an average per commune level of around 308 tonnes/month by 2010. This group will increase its waste generation by 16% from 2002 to 2010. For the 71 communes in Group 3, the average waste generation level per commune will vary from around 4,100 to more than 4,600 tonnes/month. The represented communes cover only a portion of the communes in a group. Using the ratio of the combined waste for the represented communes to the total waste in a group, obtained from the years for which the whole group waste generation was known, an estimate for the total waste in a group was found. By adding these, projected estimates for the country was obtained and presented in Table 1 which shows forecasts up to 2007 predicts a 7.6% increase for the country from 2002.

The forecasts in this study were made with very limited data for the individual communes. With more comprehensive data, the approach proposed in this paper can be expected to produce much more improved waste forecasts.

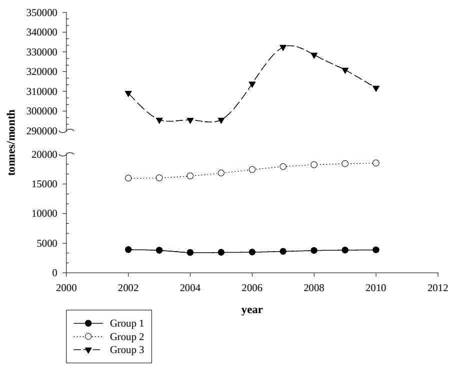

Figure 5. Total Forecasted Levels of Waste Generation for the Represented Communes up to year 2010

Table 1. Total forecasted waste (tonnes/month) up to year 2007 (* indicates model year is 2002)

waste	2002*	2003	2004	2005	2006	2007
Total Rep.	329,009	315,352	315,256	315,852	334,740	353,978
Total Chile	466,769	447,392	447,256	448,102	474,899	502,191

5. CONCLUSIONS

This research aims to design a communal analysis tool to study waste generation factors and forecast waste generation in a country. It shows the development of a systematic process where factors affecting waste generation are determined to forecast waste generation. Three groups of communes were classified based on the relevant waste related variables. Representative communes, one per group, were selected and their results were used for estimating future generation for the groups they represent. These are used to obtain an estimate for the country. Despite the limited data available for the case of Chile, satisfactory results were obtained from the models validating the proposed approach.

Reference

1. Comisión Nacional del Medio Ambiente – CONAMA, *Gestión Integral de Residuos Sólidos Domiciliarios: Política* ,_Santiago de Chile, CONAMA, 1997. http://www.conama.cl

2. McBean E and Fortin M, A Forecast Model of Refuse Tonnage with Recapture and Uncertainty Bounds, *Waste Management & Research*, *11*, pp. 373-385, 1993.

3. United States Environmental Protection Agency – USEPA, Adjusting Waste Generation. *Appendix H: Methodology to calculate waste generation based on previous years*, USEPA, 1997.

4.. Chang N-B, Pan YC and Huang S, Time Series Forecasting of Solid Waste Generation, *Journal of Resource Management and Technology*, *21*(1), pp. 1-10, 1993.

5. Ali Khan M and Burney F, Forecasting Solid Waste Composition - An Important Consideration in Resource Recovery and Recycling, *Resources, Conservation and Recycling*, *3*, pp. 1-17, 1989.

6. Buenrostro O, Bocco G and Vence J, Forecasting Generation of Urban Solid Waste in Developing Countries - A Case Study of Mexico, *Journal of Air & Waste Management Association*, *51*, pp. 86-93, 2001.

7. Hockett D, Lober D and Pilgrim K, Determinants of Per Capita Municipal Solid Waste Generation in the Southeastern United States, *Journal of Environmental Management*, *45*, pp. 205-217, 1995.

8. Grossman D, Hudson J and Marks D, Waste Generation Models for Solid Waste Collection, *Journal of The Environmental Engineering Division*, *100*(EE6), pp. 1219-1230, 1974.

9. Bagby J, Ernsdorff S, Kipperberg G and Perrin L, *Seattle's Solid Waste Plan: On the Path to Sustainability*, 2001.

10. Bridgwater AV, Refuse Compositions Projections and Recycling Technology, *Resources and Conservation*, *12*, pp. 159-174, 1986.

11. Chang N-B and Lin Y, An Analysis of Recycling Impacts on Solid Waste Generation by Time Series Intervention Modeling, *Resources, Conservation and Recycling*, *19*, pp. 165-186, 1997.

12. Navarro-Esbrí J, Diamadopoulos E and Ginestar D, Time series analysis and forecasting techniques for municipal solid waste management, *Resources, Conservation and Recycling*, *35* (3), pp. 201-214, 2002.

13. Koushki PA and Al-Khaleefi AL, An Analysis of Household Solid Waste in Kuwait: Magnitude, Type, and Forecasting Models, *Journal of Air & Waste Management Association*, *48*, pp. 256-263, 1998.

14. Smith M, *Neural Networks for Statistical Modelling*. London: International Thomson Computer Press, 1996.

15. Haykin, S. Foundation of Neural Networks, McMillan Publishers, USA., 1999.

16. NeuroShell 2.0. Ward Systems, USA., 1997.

17. Ordóñez-Ponce E, Samarasinghe S and Torgerson L, *Relations and Recovery of Domiciliary Solid Waste Using Artificial Neural Networks: A Case Study of Chile*, Proc. 19th International Conference on Solid Waste Technology and Management, Philadelphia, PA U.S.A., March 21-24, 2004.

18. Ordóñez-Ponce E, Samarasinghe S and Torgerson L, *A Method for Assessing*

Waste Generation Factors and Forecasting Waste Generation Using Artificial Neural Networks: A Case Study of Chile, Unpublished masters thesis, Lincoln University, Lincoln, New Zealand, 2004.