

Artificial Neural Network Based Model for Forecasting Sugar Cane Production

¹O.O. Obe and ²D.K. Shangodoyin

¹University Politehnica of Bucharest, Bucharest, Romania

²Department of Statistics, University of Botswana, Botswana

Abstract: Problem statement: The global need for alternative energy source has necessitated the exploration of vast organic agricultural products with a view of processing them for the production of ethanol in commercial quantity. To ascertain a sustainable production of ethanol from processed sugar cane, a predictive model based on non-linearity nature of its production is imperative. This is due to unavailability of sufficient reliable data and the wide yield fluctuation that was not well dispersed over time. **Approach:** This study employed heuristic technique to develop an Artificial Neural Network (ANN) model to forecast sugar cane production in Nigeria. The input data set used includes the socio-economic and agro-climatic factors affecting sugar cane production while the output is the actual sugar cane output for the period covering 1920-2005. Various hidden layers and processing elements were tested giving rise to different Artificial Neural Network (ANN) models. The performance of the ANN models were measured using the Mean Squared Error (MSE), Normalized Mean Squared Error (NMSE), correlation coefficient (r), Akaike's Information Criterion (AIC) and Minimum Description Length (MDL). The contributions of the inputs to the outputs were determined to know how variation of the input variables affected the output. **Results:** The 85.70% accuracy result of the best ANN model of 2-hidden layer network of 4 Processing Elements (PEs) indicated the efficacy of Artificial Neural Network in accurate prediction. **Conclusion:** The developed ANN based model fits well real data and can be used for predicting purpose with a high accuracy.

Key words: Artificial neural network, processing elements, performance measures, test, set data, training set

INTRODUCTION

Recent advances in the field of Artificial Intelligence (AI) have allowed the development of effective, newer time series (Time Series (TS) is collection of time ordered observations, each one being recorded at a specific time (period) (Cortez *et al.*, 2001)) forecasting techniques (different time series forecasting methods such as the Auto Regression Model have been developed and used over the years (Raczynski, 2003)) such as the Artificial Neural Networks (ANNs), which has been receiving greater attention lately.

An artificial neural network is an artificial representation of the human brain that tries to simulate its learning process. It makes use of collections of mathematical models that emulates the observed structure and dynamics of the brain in order to simulate its learning process (Pacific BNET Business Dictionary, 2009; Frohlich, 1996). The human brain consists of billion of neuron cells, each having limited

capabilities, but when connected together, these neurons forms the most intelligent system known. In the same way, artificial neural networks are made from hundreds or thousands of simulated neurons called Processing Elements (PEs) joined together the same way as the neurons in the human brain.

Hornik *et al.* (1989) and Stephen (1994) have shown that a neural network can approximate any given functional form to any desired accuracy level, if built with sufficient number of hidden layers. This is as a result of complex connections that exist between the neurons. Neural network learns from experience acquired from the past, by looking for patterns in the data being modeled or some form of relationship between the inputs and the result of each record.

Artificial neural network has the advantage over traditional linear models in that it can represent both linear and non-linear relationships and to learn these relationships directly from the data being modeled (NeuroDimension Inc., 2002). On the other hand,

traditional linear models are only limited to the modeling of linear relationship among data.

Neural Networks are being applied to an increasingly numbers of areas. They have been used in the task of scheduling airline flights, classification of radar clutter (Haykin and Deng, 1991), automatic target recognition (Roth, 1990) and medical diagnosis (Moein *et al.*, 2008), with considerable success. They have also been used in some forecasting tasks such as weather forecasting (California Scientific Software, 2003), power load forecasting (Benc, 1997), bond rating and a host of other uses in defense, health and business. They are known to have produced the best result to date in predicting secondary protein structure and some short-term time series prediction tasks. One of the main problems of ANN is that it requires a considerable number of data to learn from.

Environmental issues in Nigeria: Environmental issues in Nigeria are most prevalent in the major oil producing area of Niger Delta (Shah, 2004). It is an area that comprises of 20,000 km², 70,000 km² of wetlands formed primarily by sediment deposition. It is also a home to about 20 million people and 40 different ethnic groups. This flood plain makes up 7.5% of Nigeria's total land mass. It is the largest wetland and maintains the third-largest drainage area in Africa. The Delta's environment can be broken down into four ecological zones: Coastal barrier islands, mangrove swamp forests, freshwater swamps and lowland rainforests. This incredibly well-endowed ecosystem, which contains one of the highest concentrations of biodiversity on the planet, in addition to supporting the abundant flora and fauna, arable terrain that can sustain a wide variety of crops, economic trees and more species of freshwater fish than any ecosystem in West Africa. The region could experience a loss of 40% of its inhabitable terrain in the next thirty years as a result of extensive dam construction in the region. The carelessness of the oil industry has also precipitated this situation, which can perhaps be best encapsulated by a 1983 report issued by the NNPC in 1983, long perform popular unrest surfaced. As at December 12, 2006, the oil price has risen to \$62 per barrel due to youth unrest in the Niger Delta area of Nigeria as reported by the OPEC news.

The continuum consequential destructive effect of oil exploration in the oil rich regions would far outweigh the advantages derivable from the proceeds of such exploit in no time. Hence, countries and companies around the world are aggressively looking at bio-fuels as a way to reduce their dependence upon traditional fuels. But potential government policies,

production and process technologies, costs, international trade issues, distribution and infrastructure issues and global oil prices are likely constraints to the boom of bio-fuels. Some of the alternative source of fuel includes coal, cassava, maize and sugar cane.

Sugar cane is a relatively hardy tropical and sub-tropical crop which has been adapted to grow both in high rainfall areas and in desert conditions in which it is entirely dependent on irrigation. It typically yields 50% more sucrose per hectare than a beet crop in temperate climates. It thrives greatly in the northern and many other part of Nigeria. The large land mass area cum population density of Nigeria promises the world a re-enactment of the global stand of the country in energy supply.

The alternative energy source production technologies are fully in place in countries like Brazil, India, China, Bundaberg, Australia, South Africa and many other parts of the world. The ECOWAS free trade zone and the EU declaration on sugar production has really challenged indigenous company like Dangote Sugar Refinery to see to the production of white sugar in Nigeria using Tate and Lyle's technology (the largest sugar refiner in Europe). With proper technology in place, Nigeria stands to out-wit the current global alternative energy supply giant-Brazil.

To meet the global need for alternative energy source through ethanol using sugar cane, proper planning, forecasting and budgetary allocations should be made by the concerned countries and companies. This research concerns itself with the forecasting of sugar cane production in Nigeria.

Over the years it has been the concern of Nigerian government to attain the level of sustainability in agriculture products by making projections into the future. Forecasting Sugarcane production is not an easy task due to unavailability of sufficient reliable data source, low level of awareness and wide yield fluctuation that was not well dispersed over time. The awareness has just been drawn lately in Nigeria and as such there exists few data to forecast sugar cane production in the country.

Thus a means of forecasting sugar cane production using the right forecasting model remains relevant. The non-linear nature of the data requires a predictive model such as Neural Network that is known to have recorded considerable success in predicting similar time series tasks.

When compared to most crops, sugar cane appears to be particularly sensitive to changes; therefore the contributions of the inputs to the production should also be determined to know which has the most effect on the production and to know the effect of having some set of input values on the production.

The objectives of this study are to know if an artificial neural network could perform very well in predicting sugar cane production output. If this is feasible then model an artificial neural network to forecast sugar cane production output.

MATERIALS AND METHODS

The data set: The data set consists of input factors/variables and an output variables. The input factors/variables represent the socioeconomic and agro-climatic factors that influence the production level of cocoa. The input variables used are: Rainfall, favorable market, government support, consumer and industrial demands, varieties and better farm practice. The output variable is the sugar cane production output. The data collection phase was carefully planned to ensure the adequacy of data, the basic principles at work are captured and that the data is noise free.

In order to get the best outcome from the network model, the data has to be expressed in a deterministic way. Therefore the domains of input variables shown in Table 1 were employed in the study.

The proposed neural network was trained using different set of data sets to get the best set of data that will be a good representation of the conditions that the network might encounter later.

The data collected were divided into three parts namely:

- Training set: Used to build the model. 40 cases, which represent about 66% of the whole data, were used for the training
- Verification set: Used for the verification of the network. 7 cases, which represent about 11% of the whole data, were used
- Test set: Used for the prediction, from where the prediction accuracy is determined. 14 cases, which represent about 23% of the whole data, were used

During the training phase, the system adjusts its connection/weight strengths in favor of the inputs that are most effective in determining a specific output.

The use of verification set in the study is an important guard against overtraining/over-fitting the network. An over-trained network means the network has lost its ability to generalize its outcomes. Generalization is used to connote the ability of a neural network to perform very well on data it has not been trained with. The generalization errors of the models were observed and used to estimate the performance of the model.

Table 1: Input’s domain table

Input variable	Domain	
Rainfall	Below 1000 mm	1
	1000<1400 mm	2
	1400-1700 mm	3
	Above 1700 mm	4
Favorable market	Favorable	2
	Not favorable	1
Government support	Very low	1
	Mild	2
	Severe	3
Consumer and industrial demands	Very low	1
	Mild	2
	High	3
Varieties		5
	1920-1930	1
	1931-1940	2
	1941-1950	6
	1951-1960	8
	1961-1970	12
	1971-1980	16
	1981-1990	32
1991-2000	45	
Better farm practices	2001-2010	76
	Very low	1
	Mild	2
	High	3

Missing values: 0

The neural network topology: The network topology describes the arrangement/structure of the neural network. Choosing the topology of a neural network is a difficult decision (Principe *et al.*, 2000). An understanding of the topology as a whole is needed before the number of hidden layers and the number of Processing Elements (PEs) in each layer could be determined.

Time-lagged recurrent network was used for the work. The reason being not far fetched: the data being modeled contains information in its time structure, i.e., how the data changes with time (time series). Time-lagged recurrent networks are very good in nonlinear time series prediction, system identification and temporal pattern classification. Time Lagged Recurrent Networks (TLRNs) are Multilayer perceptions that has their static Processing Elements (PEs) being replaced by PEs with short term memories, such as the gamma, the Laguarre or the tap delay line. Multilayer perceptrons have been proven to be universal function approximators (Principe *et al.*, 2000). This means that they can approximate any input/output map. In order to build and train the network, the input is delayed by L samples before being fed into the network and the input signal without delays becomes the desired response. Six input nodes were used in the chosen network’s topology, which represent the input factors being considered in the study. The block structure is shown in Fig. 1.

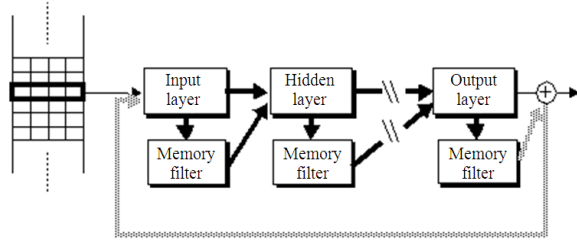


Fig. 1: Block diagram of a time lagged recurrent network

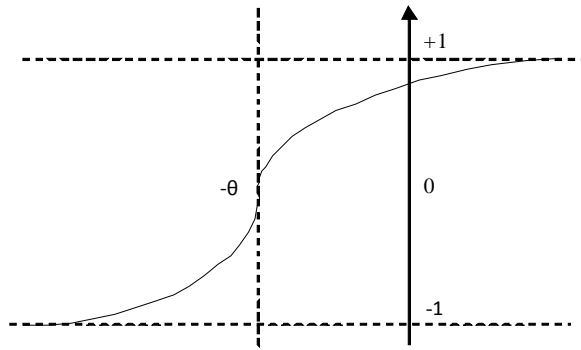


Fig. 2: A TanhAxon-with-bias type of transfer characteristic

Varied number of hidden layers and processing elements of the neural network topology were tested as well, to find out the right combination of PEs and hidden layers to solve the problem with acceptable training times and performance. These varied numbers of hidden layers and PEs give rise to different models.

The experiment was started with small number of hidden layer (0-hidden layer) and then increased gradually (growing method). The number of PEs was also varied in the course of the study from 1-5 nodes, to get the best performance network.

The control of the learning parameters is an unsolved problem in ANN research as well as in optimization theory. The goal is to reach the optimum performance in small training time. The learning parameters of the chosen network topology that fit into the research study were studied to determine the best parameters' setting. The conventional approach, which was employed in the study, is to select the learning rate and a momentum term. Momentum learning is an enhancement over the straight gradient descent search, by imposing a "memory factor" on the adaptation. This had the advantage of fast adaptation, at the same time reducing the probability of getting hooked at local minima. Thus, we have the learning equation:

$$\Delta W_{ij}(K) = \gamma \Delta W_{ij}(K-1) - \mu \partial E(k)$$

Where:

μ = The learning rate

γ = A constant (normally set between 0.5 and 0.9)

For the study, different learning rates ranging between 1 and 0.001 with the momentum term of 0.7 were used for the hidden layers and the output layer.

The neural network uses the TanhAxon transfer function (Fig. 2). The TanhAxon is normally employed as hidden and output layers in Multilayer perceptron topologies. The TanhAxon applies a bias and tanh() function to each neuron in the layer, thus outputting values within the range -1 and 1 for each neuron. This hyperbolic tangent (tanh) function is expressed as:

$$F(\text{net}) = \tanh(\alpha \text{ net})$$

The initial weights were started at random values. The use of varied random starting weights on each run could generate different outcomes; therefore five independent runs were made on each topological model in order to get the best result.

In order to get a feel of the performances of the models and how difficult the problem is, the ANN models were initially trained without the use of the cross-verification stopping criterion, until a well defined number of training epochs (set to 1000 in this study) has been reached. The learning then proceed by switching to cross-verification and observed until the Mean Square Error (MSE) for the verification set records no improvement after 100 epochs or the number of the preset training epochs has been reached. An increase in the MSE suggests that the network has begun to overtrained which can leads to poor generalization.

Performance measures: The overall performances of the models were evaluated by some forecasting accuracy measures. Five performance measures were used in the study.

Mean Squared Error (MSE): The mean squared error is two times the average cost defined by the formula:

$$MSE = \frac{\sum_{j=0}^P \sum_{i=0}^N (d_{ij} - y_{ij})^2}{NP}$$

Where:

P = Number of output processing elements

N = Number of exemplars in the data set

Y_{ij} = Network output for exemplar i at processing element j
 d_{ij} = Desired output for exemplar i at processing element j

Normalized Mean Squared Error (NMSE): This is defined by the formula:

$$NMSE = \frac{P \sum_{j=0}^N \sum_{i=0}^N d_{ij}^2 - (\sum_{i=0}^N d_{ij})^2}{N \sum_{j=0}^N \sum_{i=0}^N d_{ij}^2}$$

Where:

P = Number of output processing elements
 N = Number of exemplars in the data set, MSE. Mean Squared Error
 d_{ij} = desired output for exemplar i at processing element j

The correlation coefficient r : The performance of the network output to the desired output can be measured with Mean Square Error (MSE) value, but it doesn't necessarily tell us the direction of movement of the two set of data, hence the need for the correlation coefficient r . The correlation coefficient between a network output x and a desired output d is given by:

$$r = \frac{\sum_i (x_i - \bar{x})(d_i - \bar{d})}{\sqrt{\sum_i (d_i - \bar{d})^2} \sqrt{\sum_i (x_i - \bar{x})^2}}$$

It is the ratio of the covariance between the input and desired data over the product of their standard deviations.

The correlation coefficient ranges between $[-1, 1]$. The goal is to have the value of r close to 1 as possible.

Akaike's Information Criterion (AIC): It measures the tradeoff between training performance and the size of the network. The aim is to reduce this term to get a network having the best generalization. The term is given below:

$$AIC(k) = N \ln(MSE) + 2k$$

Where:

K = Number of network weights
 N = Number of exemplars in the training set
MSE = Mean Squared Error

Minimum Description Length (MDL): Rissanen's Minimum Description Length (MDL) criterion is similar to the AIC in that tries to combine the model's error with the number of degree of freedom to determine the level of generalization. The goal is to minimize this term.

It is also used to establish the generalization capability of a model by combining the model's error with the number of degrees of freedom. The aim is to have the least value for the term:

$$MDL(k) = N \ln(MSE) + 0.5k \ln(N)$$

Where:

K = Number of network weights
 N = Number of exemplars in the training set
MSE = Mean Squared Error

RESULTS AND DISCUSSION

Five different runs were made on each model on a Pentium 4 Intel MMX/2.5 GHz personal computer with the learning curves observed and the mean of the results taken. These were done to get a statistical sound model since the initial weights on each run are generated randomly, producing different result in each run.

Figure 4-6 and Table 3 and 4 gives the summary of the model selection criteria results for the training and the cross-verification of each of the different models considered.

Result of prediction and accuracy of result: When the performance of the network was gauged using the test data sample, the result of Table 2 was obtained.

By using the Sum Squared Error (SSE), Root Mean Squared Error (RMSE) and percentage accuracy formulae:

$$SSE = \sum_{i=1}^L e_i^2, \quad RMSE = \sqrt{\frac{SSE}{L}} \quad \text{and}$$

$$Accuracy(\%) = \frac{(RMSE \times 100)}{X}$$

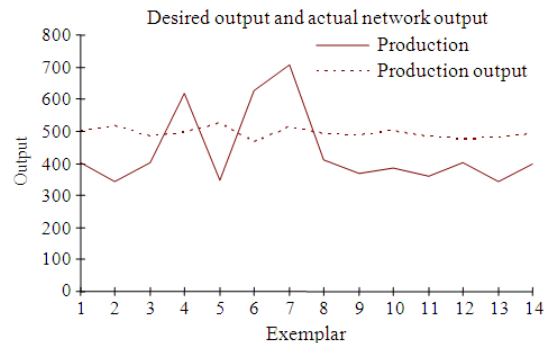


Fig. 3: Inputs contributions to the output

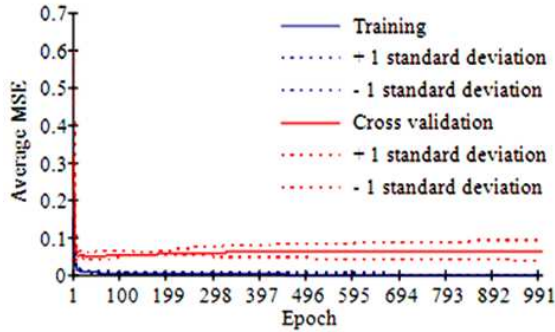


Fig. 4: Average MSE with standard deviation boundaries for 3 Runs

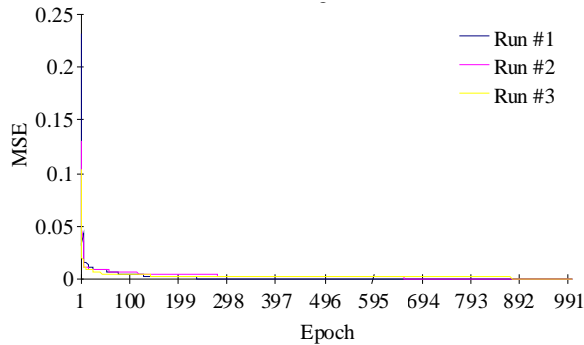


Fig. 5: Training MSE

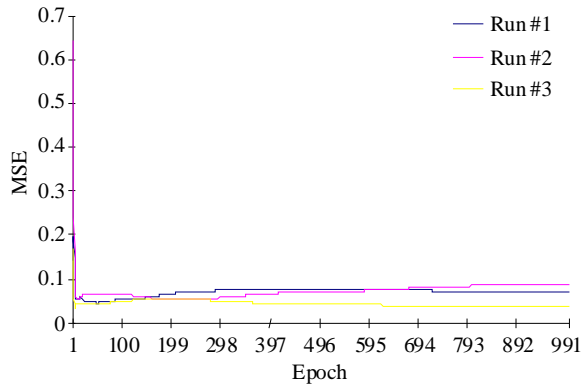


Fig. 6: Cross validation MSE

Where the error of a forecast is given by the difference between actual value and the predicted value, L is the number of forecast and X is the mean of actual values, the percentage accuracy 83.7% was archived.

The result gives trend accuracy of 94.81% which is an indicator of how the predicted productions move with the actual productions. The trend is depicted in the graph of Fig. 3.

Table 2: Table of desired output and actual network output

Production desired output	Production actual output
401.50	497.122468
341.00	504.488244
403.30	490.960426
620.00	499.524353
348.90	482.699990
628.00	497.699111
705.00	466.404457
409.45	510.186588
370.00	501.281548
387.00	503.039333
358.00	504.087936
400.30	496.702056
343.70	495.206139
398.00	498.650450

Table 3: Table showing training and cross validation of models

All runs	Training minimum	Training standard deviation	Cross validation minimum	Cross validation standard deviation
Average of minimum MSEs	0.001835786	0.000203261	0.044595765	0.01032658
Average of final MSEs	0.001835786	0.000203261	0.064429906	0.025930384
Best networks		Training	Cross validation	
Run #	1		3	
Epoch #	1000		7	
Minimum MSE	0.001699405		0.033653597	
Final MSE	0.001699405		0.035839224	

Table 4: Performance measures

Performance	Production
MSE	18174.777360000
NMSE	1.362658573
MAE	129.474804100
Min Abs error	87.660425720
Max Abs error	238.595543100
r	-0.490917260

More of the output graphs and results are shown in the Fig. 4-6 and Table 3 and 4.

CONCLUSION

Artificial neural network was used to get predicted output from the actual output data used to train the network. Different result was derived using artificial neural network by training the network with the data and the network learn from the data trained with bring the desired result of sugar cane production. So the result that has the nearest margin between the actual and desired output was selected. The contribution of the input variables to the output was determined to know the contribution of each input to the output.

The ANN result was also compared with the statistical based model based on their percentage accuracy. These results confirms the result from earlier study that shows that no simple structures without

enough hidden layer and processing elements will give inadequate capabilities without enough degree of freedom while too many hidden layers and processing elements lead to overspecialization

REFERENCES

- Benc, R., 1997. Application of artificial neural networks to power systems. <http://members.tripod.com/ann2ps/toc.html>
- California Scientific Software, 2003. Weather forecasting with neural networks. <http://www.calsci.com/Weather.html>
- Cortez, P., M. Rocha and J. Neves, 2001. Evolving time series forecasting neural network models. Proceeding of International Symposium on Adaptive Systems: Evolutionary Computation and Probabilistic Graphical Models, July, 2-2, Havana, Cuba, pp: 84-91. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.60.4494>
- Frohlich, J., 1996. What are Neural Networks? <http://fbim.fh-regensburg.de/~saj39122/jfroehl/diplom/e-11-text.html>
- Haykin, S. and C. Deng, 1991. Classification of radar clutter using neural networks. *IEEE Trans. Neural Networks*, 2: 589-600. DOI: 10.1109/72.97936
- Hornik, K., M. Stinchcombe and H. White, 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2: 359-366. DOI: 10.1016/0893-6080(89)90020-8
- Moein, S., S.A. Monadjemi, and P. Moallem, 2008. A novel fuzzy-neural based medical diagnosis system. *Proc. World Acad. Sci. Eng. Technol.*, 27: 157-161.
- NeuroDimension Inc., 2002. NeuroSolutions: What is a neural network? <http://www.nd.com/neurosolutions/products/ns/whatisNN.html>
- Pacific BNET Business Dictionary, 2009. Business definition for: Artificial neural network. <http://dictionary.bnet.com/definition/artificial+neural+network.html>
- Principe, J.C., N.R. Eliano and W.C. Lefebvre, 2000. *Neural Networks and Adaptive Systems: Fundamentals through Simulation*. John Wiley and Sons, USA., ISBN: 10: 0471351679, pp: 672.
- Raczynski, S., 2003. PredictHit: A great small prediction tool. <http://www.raczynski.com/pn/predict.htm>
- Roth, M.W., 1990. Survey of neural network technology for automatic target recognition. *IEEE Trans. Neural Networks*, 1: 28-43. DOI: 10.1109/72.80203
- Shah, A., 2004. Global issues. <http://www.globalissues.org/article/86/nigeria-and-oil>
- Stephen, G.H., 1994. *Applied Economic Forecasting Techniques*. 2nd Edn., Harvester Wheatsheaf, ISBN: 9780745013923, pp: 15-111.