FORECASTING TRIP GENERATION FOR HIGH DENSITY RESIDENTIAL ZONES OF AKURE, NIGERIA: COMPARABILITY OF ARTIFICIAL NEURAL NETWORK AND REGRESSION MODELS

J.E. Etu¹, O. J. Oyedepo²
¹,²Department of Civil Engineering, Federal University of Technology, PMB 704, Akure Ondo State, Nigeria

Abstract — Evidence from literature has shown the absence of the use of Artificial Neural Network techniques in formulating trip generation forecasts in Nigeria, rather the practice has consisted more on use of regression techniques. Therefore, in this study, the accuracy of Radial Basis Function Neural Network (RBFNN) and Multiple Linear Regression model (MLR) in formulating home-based trips generation forecasts was assessed. Datasets for the study were acquired from a household travel survey in the high density zones of Akure, Nigeria and were analysed using SPSS 22 statistical software. Results of data analysis showed that the RBFNN model with higher Coefficient of Determination ($R^2$) value of 0.913 and lower Mean Absolute Percentage Error (MAPE) of 0.421 performed better than the MLR with lower $R^2$ value of 0.552 and higher MAPE of 0.810 in predicting the number of home-based trips generated in the study area. The study demonstrated the higher accuracy of the RBFNN in producing trip generation forecasts in the study area and is consequently recommended for researchers in executing such forecasts.

Keywords: home based trips; artificial neural network; radial basis function; regression; trip generation

1.0 INTRODUCTION

The travel demand forecasting process is a fundamental operation within any urban transport planning exercise. It is a process that is aimed at predicting travel demand for some time in the future with the aid of simplifying explanatory models which take cognizance of present travel behaviour mechanisms, growth potentials and changing socio-economic circumstances [1]. The traditional approach to travel demand forecasting is based on what is known as the classical four-step modelling procedure which comprises of:

- Trip Generation: where number of trips from given origins and destinations are estimated
- Trip Distribution: where destination for each trip from a given origin is determined
- Mode Choice: where mode choice for each trip is determined
- Route Assignment: where the specific route for each trip is determined

Trip generation forecasting as the first step of the four-step modelling procedure provides data for estimating the other three components of the procedure. It helps to estimate trip rates within a study location thereby serving as a precursor to provision of facilities to tackle anticipated transportation problems in an urban area. The development of accurate transportation plans for an urban area is dependent on the envisaged trip rates in the urban area. Therefore, conceptually sound and empirically accurate forecast of trip rates is required to maximize the benefits of an urban transportation system. Although many techniques have been deployed over the years for forecasting trip generation in Nigeria, these methods have been largely inadequate in meeting the demand of accuracy and are also not easily applicable to varying data types and sizes.
It is very important to state that the level of accuracy desired of any forecasting process is largely dependent on the technique/method used in carrying out such forecasts. Artificial Neural Network (ANN) techniques have been able to satisfy, to a great measure, this demand for accuracy required of forecasting in many fields of applications. ANN tools have come to be known as excellent tools for forecasting and hence its use in various fields has gained traction over the years in developed countries of the world. The ANN technique has been deployed in various studies across Nigeria such as: prediction of global solar radiation at Minna, Warri and Uyo cities of Nigeria [2], [3] and [4]; prediction of knowledge discovery [5]; electrical load demand forecasting [6]; cost optimization in a bottling and paint company in Nigeria [7]; cost optimization in a dual-source multi-destination outbound system [8]; forecasting Nigerian stock market returns [9]; path loss prediction at 900 MHz for Jos City, Nigeria [10]; estimation of atmospheric refractivity profile at Makurdi, Nigeria [11]; prediction of likely distress in Nigerian banking sector [12]; crude oil production forecasts [13]; insolvency prediction of some selected Nigerian banks [14]; modelling energy content of municipal solid waste [15]; and assessment of infant mortality rates [16].

It is therefore alarming and mind boggling, that despite the application of the ANN predictive capabilities in other fields of endeavour in Nigeria, evidence from literature does not show its use and adoption in forecasting trip generation rates in the country. Rather, studies on trip generation in Nigeria have dwelt on the use of traditional regression modelling techniques, some of which include: estimation of factors influencing travel demand using stepwise regression models [17]; calibration of trip generation models for varying residential density zones of Akure, Nigeria using multivariable regression analysis [18]; investigation of socio-economic factors influencing trip generation in Ifo Local Government Area, Ogun State, Nigeria using simple Pearson Product Moment Correlation [19]; development of trip generation models for land uses in Nigeria using regression analysis [20]; analysis of socio-economic factors influencing intra-urban trip generation in Abeokuta, Nigeria using multiple linear regression analysis [21] and modelling of household trip generation rates in Ado-Ekiti, Nigeria using regression analysis [22]. In summary, none of the aforementioned studies have considered the use of ANN in estimating trip generation rates in Nigeria. They have dwelled mostly on use of regression techniques which are limited in their ability to provide accurate forecast for non-linear relationships, large and noisy data sets. The ANN techniques offer superior capabilities such as ability to provide solutions to non-algorithmic problems and ability to learn how to deal with new and unexpected situations by the help of experience. Furthermore, they can relate input with output, allow large number of variables and are error tolerant [23].

In view of the outlined advantages offered by ANN models, and their successful use for forecasting in other areas of application in Nigeria, this study applied the radial basis function neural network (RBFNN) model to estimating trip generation rates in high density areas of Akure metropolis. The Radial Basis Function Neural Network (RBFNN) is an ANN technique which has a feed-forward structure consisting of three layers: input layer, nonlinear hidden layer and linear output layer. Their simple structure enables learning in stages, and gives a reduction in the training time. The major advantages of RBFNN over other ANN techniques is that RBFNN requires less time to reach the convergent condition and ensures global convergent conditions; in other words, it requires minimal training time.

The purpose of this study is to assess the suitability of RBFNN and multiple linear regression models (MLR) in forecasting trip generation rates in high density zones of Akure, Nigeria. To the best of our knowledge, this study makes it the first time an ANN technique is being used for trip generation modelling in Nigeria. The results of the RBFNN model will be compared to those of multiple linear regression. It is expected that the results of the study will uphold the superior accuracy of ANN tools in trip generation forecasting in Nigeria.

2.0 METHODOLOGY

2.1 THE RADIAL BASIS FUNCTION NEURAL NETWORK ALGORITHM

Artificial neural networks help to estimate a function from some example input-output pairs with little or no knowledge of the form of the function through supervised learning.
The function is learned from the supplied input-output pairs called the training set. The training set contains elements which consist of paired values of the independent input variable and the dependent output variable. For example, the independent variable in the functional relation

\[ y = f(x) \]  

is \( x \) a vector and the dependent variable is \( y \) a scalar. The value of the variable \( y \) depends through the function \( f \) on each of the components of the variable \( x \).

\[ x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \]

The training set for \( n \) pairs (from 1……….. \( p \)) is given by

\[ T = \{(x_i, y_i)\}_{i=1}^{p} \]  

Where \( y_i \) for this study is the number of home-based trips and the \( x_i \) includes, the household characteristics selected as independent variables for the forecasting. So that \( T \) is pairs of number of home-based trips and household characteristics.

For linear models in a neural network, the function \( f(x) \) for \( y \) is given by:

\[ f(x) = \sum_{j=1}^{\varnothing} w_j \varnothing_j(x) \]  

The model \( f \) is expressed as a linear combination of a set of \( \varnothing \) fixed functions often called basis functions. \( w \) represents the coefficients of the linear combinations and letter \( \varnothing \) represents the basis functions and are referred to as weights and hidden units respectively. The hidden units provide a set of functions that constitute an arbitrary basis for the input patterns.

For radial basis function neural networks (typical architecture shown in Figure 1):

- hidden units are known as radial centres and represented by the vectors \( c_1; c_2; \ldots \ldots; c_n \)
- transformation from input space to hidden unit space is nonlinear whereas transformation from hidden unit space to output space is linear
- dimension of each centre for a \( k \) input network is \( k \times 1 \)
The radial basis functions in the hidden layer produces a significant non-zero response only when the input falls within a small localized region of the input space around the mean. Each hidden unit has its own receptive field in input space. An input vector $x_i$ which lies in the receptive field for centre $c_j$, would activate $c_j$ and by proper choice of weights the target output is obtained as:

$$y = \sum_{j=1}^{h} w_j \varphi_j(x)$$  \hspace{1cm} (4)

where $w_j$ is weight of $j^{th}$ centre, $\varphi_j$ is the radial function

There are different radial functions used in practice, but the most commonly used is the Gaussian activation function and it has been adopted in this study. The Gaussian activation function is of the form:

$$\varphi_j(z) = e^{-z^2/2\sigma^2}$$  \hspace{1cm} (5)

where $z = (\|x - c_j\|)$

where $x - c_j$ is the Euclidean distance between input vector ($x$) and the radial basis function centre ($c_j$) $\sigma$ is the width of radial basis function.

Training of RBFNN requires optimal selection of the parameters vectors centres ($c$) and weight ($w$). Both layers are optimized using different techniques and in different time scales. Gradient Descent Learning (GDL) technique was used to update the weights and centres of the RBFNN used in this study. The GDL technique carries out supervised training by error correcting term and the update rule for learning is derived below [24]:

For centre learning:

$$c_{ij}(t + 1) = c_{ij}(t) - \eta_1 \frac{\partial E}{\partial c_{ij}} , \text{for } i = 1 \text{ to } p, j = 1 \text{ to } h$$  \hspace{1cm} (6)

For weight learning:
\[ w_i = (t + 1) = w_i(t) - \eta_2 \frac{\partial E}{\partial w_i} \]  

where \( E = \frac{1}{2} \sum (y^2 - y)^2 \) is the cost function \( \emptyset \)

differentiating \( E \) with respect to \( w_i \), we get

\[ \frac{\partial E}{\partial w_i} \frac{\partial E}{\partial y} \frac{\partial y}{\partial w_i} = -(y^d - y) \emptyset_i \]  

(8)

differentiating \( E \) with respect to \( c_{ij} \), we get

\[ \frac{\partial E}{\partial c_{ij}} \frac{\partial E}{\partial y} \frac{\partial y}{\partial \emptyset_i} \frac{\partial \emptyset_i}{\partial c_{ij}} = -(y^d - y) \times w_i \times \frac{\partial \emptyset_i}{\partial z_i} \times \frac{\partial z_i}{\partial c_{ij}} \]  

(9)

Therefore,

\[ \frac{\partial \emptyset_i}{\partial z_i} = \frac{z_i}{\sigma^2} \emptyset_i \quad \text{and} \quad \frac{\partial z_i}{\partial c_{ij}} = \frac{\partial}{\partial c_{ij}} \left( \sum (x_j - c_{ij})^2 \right)^{1/2} = -(x_j - c_{ij}) / z_i \]  

(10)

Further simplification gives the update rule for centre learning as:

\[ c_{ij}(t+1) = c_{ij}(t) + \eta_1 (y^d - y) w_i \frac{\emptyset_i}{\sigma^2} (x_j - c_{ij}) \]  

(11)

While that for the linear weights becomes:

\[ w_i (t+1) = w_i(t) + \eta_2 (y^d - y) \emptyset_i \]  

(12)

Equations 11 and 12 therefore, represents the criteria for optimal selection of centres and weights for the network for each pairs of number of home-based trips and household characteristics.

2.2 MULTIPLE LINEAR REGRESSION MODEL (MLR)

The Multiple Linear Regression Model (MLR) was also used in developing trip generation models to determine the daily home-based trip rates from the study area.

The MLR takes the form

\[ T_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n \]  

(13)

where:

\( T_{ij} \) is number of home-based trips generated by households daily.

\( \beta_0 \) is the model constant and

\( \beta_1 - \beta_n \) are regression coefficients associated with the household characteristics \( X_1 - X_n \).

2.3 CRITERIA FOR PERFORMANCE ASSESSMENT OF THE MODELS

Performance assessment and comparisons of the two models was carried out using the numerical values of their Coefficient of determination \( (R^2) \) and Mean Absolute Percentage Error (MAPE). The \( R^2 \) is a measure of the strength of the relationship between the dependent and independent variables. It ranges between 0 and unity. However, values of 0.5 were the standard adopted in this research.
The magnitude of Mean Absolute Percentage Error (MAPE) refers to the percentage difference between the mean observed values and mean predicted values of the models. This error value between the observed survey results and the predicted results with the RBFNN and MLR models is expressed by:

$$\text{MAPE} = \frac{M_O - M_P}{M_P} \times 100$$  \hspace{1cm} (14)

where

$M_O$ is the observed mean value of home-based trips and

$M_P$ is the predicted mean value of home-based trips.

Given that both models fit on the same dataset, the model with higher $R^2$ and lower MAPE is considered to outperform the other one.

2.4 DATA COLLECTION AND VARIABLE SELECTION

Data for this study were collected through a household interview travel survey. Data were collected from 1005 households in the high-density area of Akure using systematic random sampling technique. The questionnaire used for the survey contained questions on household characteristics such as gender, age, economic status, number of household members, number of cars available for use by household members, the number and type of driving licences owned by household members and other household attributes. Collection of data also involved administration of a travel diary to household members above 15 years of age. The travel diary was required to get information on trips made the previous day by such residents. The travel diary section of the questionnaire included questions concerning travelling (e.g. origin and destination address, mode of transport, trip purpose etc.).

From the survey data obtained, the independent variables were selected using the Pearson correlation analysis. This helped in selecting household factors with high correlation with the number of home-based trips for use in the models. The selected household factors were further subjected to multi-collinearity test using their person correlation coefficients. Household factors which passed the correlation and multi-collinearity tests served as the independent variables used in this study. These variables were then subject to descriptive statistical analysis and were used in calibrating the MLR and RBFNN models. Table 1 shows the variables used in generating the models.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Abbreviation</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of home-based trips</td>
<td>NHB</td>
<td>Dependent variable</td>
</tr>
<tr>
<td>Number of household members</td>
<td>NHM</td>
<td></td>
</tr>
<tr>
<td>Number of employed household members</td>
<td>NEHM</td>
<td></td>
</tr>
<tr>
<td>Number of students in household</td>
<td>NSH</td>
<td>Independent variables</td>
</tr>
<tr>
<td>Number of Household members greater than age 15</td>
<td>NHM15</td>
<td></td>
</tr>
<tr>
<td>Number of Driver’s license holders in the household</td>
<td>NDLMH</td>
<td></td>
</tr>
</tbody>
</table>

Descriptive statistics of the data used in developing the trip generation models are shown in Table 2. Further correction of the data collected from the 1005 households yielded 966 valid datasets and thereby considered in the modelling process.
Table 2 Descriptive statistics of variables

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHB</td>
<td>6.17</td>
<td>2.852</td>
<td>966</td>
</tr>
<tr>
<td>NHM</td>
<td>4.47</td>
<td>2.193</td>
<td>966</td>
</tr>
<tr>
<td>NEHM</td>
<td>2.1</td>
<td>1.089</td>
<td>966</td>
</tr>
<tr>
<td>NHM15</td>
<td>2.63</td>
<td>1.530</td>
<td>966</td>
</tr>
<tr>
<td>NSH</td>
<td>1.61</td>
<td>1.611</td>
<td>966</td>
</tr>
<tr>
<td>NDLH</td>
<td>0.46</td>
<td>0.763</td>
<td>966</td>
</tr>
</tbody>
</table>

Statistical Package for Social Sciences Version 22 (SPSS 22) was the analytical tool used in carrying out analysis of data for this study. The software was used in carrying out descriptive statistical analysis and calibration of the regression (MLR) and radial basis function neural network (RBFNN) models.

3.0 RESULTS AND DISCUSSION

3.1 MULTIPLE LINEAR REGRESSION MODEL (MLR) RESULTS

Table 3 shows the coefficients of the variables in the MLR model developed to predict the mean number of home-based trips generated in the study area. The coefficients showed that all the independent variables were positively correlated with the dependent variable and were all significant at the 95% confidence level (with Sig. <0.05). This shows that an increase in any of the household variables considered will bring about a subsequent increase in the number of home-based trips. Equation 15 is the model equation established from the model coefficients.

\[ NHB = 1.930 + 0.442N_{HM} + 1.38N_{HM15} + 0.599N_{DLH} + 0.456N_{SH} + 0.401N_{EHM} \]  

(15)

Table 3 Coefficients of MLR model

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Standard Error</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>1.930</td>
<td>0.158</td>
<td>12.209</td>
<td>0.000</td>
</tr>
<tr>
<td>NHM</td>
<td>0.442</td>
<td>0.047</td>
<td>9.480</td>
<td>0.000</td>
</tr>
<tr>
<td>NHM15</td>
<td>0.138</td>
<td>0.067</td>
<td>2.058</td>
<td>0.040</td>
</tr>
<tr>
<td>NDLH</td>
<td>0.599</td>
<td>0.086</td>
<td>6.967</td>
<td>0.000</td>
</tr>
<tr>
<td>NSH</td>
<td>0.456</td>
<td>0.054</td>
<td>8.527</td>
<td>0.000</td>
</tr>
<tr>
<td>NEHM</td>
<td>0.401</td>
<td>0.080</td>
<td>5.002</td>
<td>0.000</td>
</tr>
</tbody>
</table>

From the model summary in Table 4, the MLR gave \( R^2 \) value of 0.552. This showed that the model could explain 55.2% of the variation between the variables used in the model and showed that the model is a good fit for the variables considered in the model. The MLR model predicted mean number of home-based trips of about 6.12 as against 6.17 observed from field results. This shows that the model had a
mean average percentage error (MAPE) of 0.810% in predicting the number of home-based trips generated in the study area.

Table 4 Summary of the MLR model results

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Error</th>
<th>MAPE</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.17</td>
<td>6.12</td>
<td>0.05</td>
<td>0.810</td>
<td>0.552</td>
</tr>
</tbody>
</table>
3.2 RADIAL BASIS FUNCTION NEURAL NETWORK MODEL (RBFNN) RESULTS

In developing the RBFNN, Table 5 shows that the proportion of the sample used in training, testing and Holdout/validation for the network were 60.4%, 29.3% and 10.4% respectively.

Table 5 Case processing summary for RBFNN model

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>583</td>
<td>60.4%</td>
</tr>
<tr>
<td>Testing</td>
<td>283</td>
<td>29.3%</td>
</tr>
<tr>
<td>Holdout/validation</td>
<td>100</td>
<td>10.4%</td>
</tr>
<tr>
<td>Total</td>
<td>966</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

The network used in developing the RBFNN model comprised of 1 hidden layer with 8 neurons. The architecture for the network is shown in Figure 2. The RBFNN summary data as shown in Table 6 indicated that the data used in the testing stage predicted the mean number of home-based trips in the study area as 6.144.

![Figure 2 RBFNN model architecture](image)

As seen in Table 6, the RBFNN model predicted 6.144 mean number of home-based trips. The prediction was carried out with an $R^2$ value of 0.913 which means that the model was able to explain 91.3% of the relationship between the output and input variables. Also, the RBFNN model prediction of 6.144 home-based trips as against 6.17 observed from field results, thereby giving a mean absolute percentage error (MAPE) of 0.421.
Table 6 Summary of RBFNN model

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Error</th>
<th>MAPE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.17</td>
<td>6.144</td>
<td>0.026</td>
<td>0.421</td>
<td>0.913</td>
</tr>
</tbody>
</table>

3.3 PERFORMANCE ASSESSMENT OF RBFNN AND MLR MODELS

The accuracy of the RBFNN and MLR models was assessed based on their respective $R^2$ and MAPE values. From Table 7, the RBFNN model with higher $R^2$ value of 0.913 and lower MAPE of 0.421 performed better than the MLR with lower $R^2$ value of 0.552 and higher MAPE of 0.810.

Table 7 Performance measures of RBFNN and MLR models

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>0.552</td>
<td>0.810</td>
</tr>
<tr>
<td>RBFNN</td>
<td>0.913</td>
<td>0.421</td>
</tr>
</tbody>
</table>

4.0 CONCLUSION

This study assessed the accuracy of the radial basis neural network and multiple linear regression models in forecasting trip generation rates in the high density zone of Akure, Nigeria. This was to fill the gap in the application of neural network models to trip generation forecasting in Nigeria. Data for the model analysis were acquired from a travel survey of households in the study area. A total of about 966 datasets acquired from the survey were used in model analysis. Data were analysed using SPSS 22. The accuracy of the MLR and RBFNN models was assessed using their respective $R^2$ and MAPE values. Results of the model showed that the RBFNN model had $R^2$ and MAPE values of 0.913 and 0.421 respectively while the MLR model had $R^2$ and MAPE values of 0.552 and 0.810 respectively. These results showed that the RBFNN model with $R^2$ and lower MAPE values was more accurate than the MLR model with lower $R^2$ and higher MAPE. These results thereby uphold the superior capacity of RBFNN models to forecast trip generation models in Nigeria. The study recommends a deviation from the use of regression techniques to the use of ANN techniques in formulating trip generation forecasts across urban centres of Nigeria as it has displayed higher forecasting accuracy.
REFERENCES


