2nd International Engineering Conference (IEC 2017)
Federal University of Technology, Minna, Nigeria

Municipal Solid Waste Data Quality on Artificial Neural Network Performance

S. O. Masebinu[1,*], **E. T. Akinlabi**[2], **E. Muzenda**[3], **A. O. Aboyade**[4], **C. Mbohwa**[5], **M. Manyuchi**[6], **P. Naidoo**[7]
[1,2,3,4,5]Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa
[*]Corresponding author (masebinity@gmail.com)

ABSTRACT
Short and long-term municipal solid waste (MSW) management requires adequate planning. Understanding the relationship among variables that affect MSW generation and predicting MSW based on them is needed for an effective planning. Methodologies to forecast MSW are numerous and have been implemented at different level of data granularity. Lack of data in many African cities and countries has hampered effective waste management plan. The lack of data has mainly been attributed to insufficient budget and lack of capacity to implement such management structure. In this study, we investigated the impact of data quality on forecasting efficiency using advanced prediction techniques. It was observed that the quality of waste related data variables determines the extent of model reliability and prediction accuracy.

Keywords: *Municipal Solid Waste; Neural network; Forecasting; Waste Management.*

1 INTRODUCTION

Municipal solid waste (MSW) generation forecasting is crucial for an integrated waste minimisation strategy (Ordonez-Ponce 2004, Grazhdani 2016, Kolekar, Hazra et al. 2016, Marandi and Ghomi 2016). For example, the capacity of waste storage facilities distributed to households, transportation trucks required, capacity of the material resource recovery, composting and energy recovery plant, are directly related to the present and future quantity of waste generated (Bandara, Hettiaratchi et al. 2007, Abdoli, Nezhad et al. 2012, Abbasi and El Hanandeh 2016). The level of confidence associated with a forecast depends on the reliability of the MSW generation data set as well as other independent objective variables (Garrod and Willis 1998, Patel and Meka 2013). These variables can be classified into seven global variables as presented in **Error! Reference source not found.** (Ordonez-Ponce 2004, Patel and Meka 2013, Intharathirat, Abdul Salam et al. 2015, Sukholthaman, Chanvarasuth et al. 2015). Different research results considering these variables have often reported different factors mainly due to the quality of data available and location of research, however, population and income are the most hypothesized and investigated (Adamovic, Antanasijevic et al. 2016).

On data availability and quality, several studies highlight deficiency of quality data necessary for effective organisation of MSW management systems (Dyson and Chang 2005, Sukholthaman, Chanvarasuth et al. 2015, Younes, Nopiah et al. 2015, Azadi and Karimi-Jashni 2016). Both the planning and design of MSW management system require accurate prediction of generation rate (Dyson and Chang 2005). It is quite a challenge to achieve the desired level of prediction with high level of confidence in many African cities due to lack of data (Dyson and Chang 2005).The forecasting of MSW generation rate can be of three groups in terms of duration; a short term group with days to few months forecasting; mid-term group with few months to 3-5 years forecasting; and long term group with accurate forecasting capability greater than five years (Abbasi and El Hanandeh 2016). The forecasting models applied can be broadly classified into five groups based on the underlying mathematical methodology. They are descriptive statistical method, time series analysis, regression, input-output models and artificial intelligence model (Beigl, Lebersorger et al. 2008, Abbasi and El Hanandeh 2016). The five groups of models have their merits and demerits due to fluctuation and nonlinearity of quantity of MSW generated. This problem is a huge obstacle for accurate forecasting especially with traditional statistical model such as linear regression (Jiang and Liu 2016). Conventional forecasting methods such as time series, regression and input-output methods are designed based on semi-empirical mathematical methodology which requires long historical data. The non availability of detailed and granular historical data relating to the quantity and composition of MSW generated has limited the long term planning and/or short term expansion programmes intangible. This limitation have often been attributed to insufficient budget allocation for MSW management, and lack of skills on technology appropriation (Dyson and Chang 2005, Intharathirat, Abdul Salam et al. 2015, Azadi and Karimi-Jashni 2016). We review in brief, different forecasting methodologies applied to MSW management highlighting their pros and cons for accurate prediction of MSW and management decision making process. We further attempt to highlight impact of data quality on forecasting by applying artificial neural network.

Table 1 Factors affecting MSW generation rate

Global Variable	Sub-variables
Demographic	Population and population density, household count and household density, house hold type and household size, age, gender, occupation, expenditure on groceries, electric energy consumption and income level
Economic	Economic growth, gross domestic product, consumer price index, employment, unemployment and waste budget
Geographic	Climate and geographic conditions, natural resource
Technical/technology	Lack of manufacturing standards, engineering problems, inefficient facilities and equipment
Social	Awareness, level of literacy, public cooperation, religion and cultural practice, urbanisation, tourist attraction, political stability
Consumer behaviour	Consumption pattern, cooking activity, lifestyle, disposal pattern
Legislative and Administrative	Strategies, policies, laws, enforcement level and management institution efficiency. E.g. disposal fees, existence of recycling programmes and quantities recycled

2 Forecasting Methodologies

Several forecasting methodologies can be used in forecasting MSW generation and these include regression analysis, time series, input-output model and artificial intelligence.

2.1 Regression Analysis

The simplicity of regression analysis supported by known statistical theory makes it a good candidate for estimation of MSW generated requiring input such as economic and demographic data. However, the need for the input data to conform with the statistical theory such as normality of errors, constant variance and independence, place a limitation on its applicability to real dynamic problems (Hockett, Lober et al. 1995). Also, regression model does not consider socio-economic factors affecting MSW generation. Navarro-Esbrí, Diamadopoulos et al. (2002) investigated two forecasting techniques, seasonal autoregressive and moving average (sARIMA) and non-linear system analysis. The analysis was done for short term forecasting for three cities in Spain and Greece. Ghinea, Drăgoi et al. (2016) implemented a regression model using socio-economic factors to predict MSW generation. A S-curve model was reported to be most suitable for MSW forecasting. Weng, Fujiwara et al. (2009) developed a regression model to simulate MSW generation by fractional composition. The model is suitable for short term forecasting but require historical data. Aside historical data required, regression model have poor precision accuracy when fed with poor dataset and cannot learn or adapt to new events (Intharathirat, Abdul Salam et al. 2015).

2.2 Time Series

Time series approach depends on historical data and can factor the dynamic nature of the MSW data. However, the approach lack empirical justification and it ignore other factors such as socio-economic and demographic data that can impact of MSW generation rate (Younes, Nopiah et al. 2015). Mwenda, Kuznetsov et al. (2014) employed time series analysis for predicting of waste generation. The ARIMA (1,1,1) model was implemented was preferred to other model in terms of MAPE, MAD and RMSE measures. While Marandi and Ghomi (2016) found the ARIMA (2,1,0) to yield a higher performance in terms of MAPE, MAD and RMSE measures. Denafas, Ruzgas et al. (2014) investigated the effect of seasonal variation using time series model on waste composition in four European countries. Suggestion for long term data was made to test the accuracy of the model and increase robustness. Ferreira, Figueiredo et al. (2014) and Beigl, Lebersorger et al. (2008) highlighted forecasting of MSW considering the link between socio-economic conditions and the quantity of generated waste. They concluded that a combination of factors is more suitable than a single time series analyses.

2.3 Input-Output Model

Input-output model like material flow analysis can consider the dynamic properties of MSW generation. It is more suitable for MSW generation than collected waste forecasting as result of low consistency has been reported by Beigl, Lebersorger et al. (2008). Reynolds, Geschke et al. (2016) highlighted the low level of data, that is total waste generated, generally reported for MSW and developed an input-output model that expand the approach to help in providing data at a higher resolution of fractional composition. They concluded that other methodologies are needed to validate the result. Masebinu, Akinlabi et al.

(2017) applied the input-output model using material flow analysis for waste generated in the City of Johannesburg. They identified the role of informal recycler in waste recycling. They however, recommended increasing the robustness of the model.

2.4 Artificial Intelligence

The fluctuation and nonlinearity of MSW is a huge obstacle for accurate forecasting especially with traditional statistical model such as linear regression (Jiang and Liu 2016). Artificial intelligence model is a modern forecasting approach which include expert system, fuzzy forecasting, evolutionary programming, support vector machine, grey dynamic modelling and artificial neural network (ANN). The artificial intelligence is the most widely investigated for MSW forecasting. It has the capacity to handle the dynamic nature of MSW as well as other socio-economic factors to forecast the future generation trend. ANN also has its downside such as overfitting training, local minimum, difficulty in determining network architecture and the need for redundant and noise in data to be excluded (Noori, Karbassi et al. 2010, Intharathirat, Abdul Salam et al. 2015). Stopping techniques used as a form of regularisation has been widely applied to overcome over fitting problems. Other demerits can be overcome through approach such as principal component analysis, gamma test and multicollinearity test (Noori, Abdoli et al. 2009, Noori, Karbassi et al. 2010).

Research has intensified on using artificial neural network (ANN) to forecast MSW generation due to its ability to predict non-linear systems and ease of implementation (Noori, Abdoli et al. 2009, Batinic, Vukmirovic et al. 2011). ANN architecture is composed of mainly three layers; input layer that feed the input into the model; the hidden layer where all the mathematical simulations occur on the input; and the output layer that return the result of the input. ANN can learn and construct a complex non-linear system through finding the relationship between a set of input independent variables and the output dependent variables through a process called training (Noori, Abdoli et al. 2009, Batinic, Vukmirovic et al. 2011, Abdoli, Nezhad et al. 2012, Younes, Nopiah et al. 2015). There are two broad classification of ANN, feedforward and back propagation. Batinic, Vukmirovic et al. (2011) applied ANN in forecasting MSW generation for ten municipalities in Serbia. The relationship between MSW quantity, its composition and socio-economic indicators were established. Four input variables (economic indicator, average age of population, level of education, and municipal sector), ten neurons in the hidden layer and six output dependent variables (organics, paper, plastic, glass, metal and others) of the MSW composition in kg/capita/day were considered. Thereafter, a mid-term forecast, 2010-2016, was made. The results showed good prediction accuracy despite the limited dataset used in the model training. Antanasijevic, Pocajt et al. (2013) compared back propagation (BP) and general regression neural network (GRNN), a feedforward non-linear regression method, in forecasting MSW generation for Serbia and Bulgaria using a limited sample data. GDP, domestic material consumption and resources productivity were considered as input variables. It was concluded that GRNN had a better predicting accuracy than BP with good stability and required less run time. Adamovic, Antanasijevic et al. (2016) investigated the performance of GRNN model with and without a structural break forecasting MSW generation at the national level for 44 countries. The model with structural break, that is the ability to consider and predict unforeseen event such as economic crisis, resulted in a much reliable result. Younes, Nopiah et al. (2015) employed a feedforward non-linear autoregressive network to predict MSW generation in Malaysia. Of the input variables considered, GDP, population and employment were the most correlated to MSW generation and were used for forecasting MSW generation. Noori, Abdoli et al. (2009) compared two forecasting techniques, a multivariate linear regression based on principal component analysis (PCA-MLR) and a feedforward ANN approach, for prediction of MSW generation. The ANN approach had a better result than the PCA-MLR. Weekly forecast of MSW generation using BP-ANN was performed by Noori, Karbassi et al. (2010). Too many input variables affected the performance of the ANN model implemented. Noori, Karbassi et al. (2010) suggested a pre-process approach of input variables using the principal component analysis (PCA) and gamma test. The PCA approach reduced the number of input variables from 13 to 5. Azadi and Karimi-Jashni (2016) compared ANN and multiple linear regression (MLR) performance on predicting mean seasonal MSW generation. Input variables used were the population, MSW collection frequency, maximum seasonal temperature and altitude. MLR showed poor prediction performance while ANN achieved a better result. Abdoli, Nezhad et al. (2012) implemented ANN for long term forecasting of MSW generation for Iran using a multilayer perception approach and showed that it has good prediction capability than traditional method. The combination of traditional and artificial intelligence models is also possible as reported by (Abbasi and El Hanandeh 2016, Singh and Satija 2016).

Summarising on model suitability, the selection of a modelling method depends on the purpose of the model, data available, waste type and the degree of accuracy needed. Different model strength can be combined to achieved the desired result. Ferreira, Figueiredo et al. (2014) concluded that the application of time series analysis and input-output analyses is advantageous for insight into specific information such as impact of seasonal variation.

Towards understanding the limitation caused by poor dataset, the case of a City in South Africa was considered. Traditional method, time series, has been applied with population as the independent variable. The predicted waste generation will be 1.99 million by 2020 and 3.6 million by 2040 should population reach 9.2 million. Though population is an important factor, however, other factors such as income, employment, age group and education level have been reported to correlate strongly with MSW generation rate. Other independent variables have not been considered mainly due to lack of reliable data on them. Also, reported annual waste quantity by the City waste management company is not actually waste generated, it is rather waste disposed to landfill. The City has a long-term ambition of both waste minimisation, energy and resource recovery from MSW. Therefore, it is important to consider major factors aside population and income to effectively plan for the current challenges and develop frame work for the anticipated MSW. Based on the above argument, an attempt has been made to employ ANN considering socio-economic factors as well as population to train and validate an ANN model on the limited and available waste data.

3 METHODOLOGY

3.1 DATA COLLECTION

The waste generation factors that were considered are population, number of household, gross domestic product, employment and unemployment. The considerations were limited to these variables due to availability of data. The dataset retrieved were from 1997 to 2015. Complete dataset for this period was available for GDP, employment and unemployment. Population data and number of household for 2001 and 2011 were retrieved from 2001 and 2011 census data purchased from Statistic South Africa. Linear interpolation and published annual reports of the City was used to provide data for the missing interval. The annual waste reported by the City waste management company, from 2002 to 2015 was retrieved from online published reports. It is worth mentioning that the values reported in the annual reports were not the waste generated but the waste disposed at City's landfills. The actual data for the quantity of waste generated is not readily available in the published annual reports.

3.2 CORRELATION EVALUATION

The collected data were statistically analysed searching for correlation with MSW generation. Correlation among variables were investigated and result presented in a pair-wise correlation matrix. Due to the limited variables, multi-collinearity and heteroskedasticity tests were not conducted on the input variables.

3.3 DEVELOPMENT OF THE ANN

ANN was used to train and validate the model. The whole dataset was normalised to allow for fair comparison of paired variables. The model was implemented in the open source statistical tool R (Figure 1) based on the "neuralnet" package among other data preparation packages. In the model, only one hidden layer with 5 neurons was considered, as it is recommended to be sufficient for prediction. The output layer was the MSW generation. 80% of the data was used for training and 20% for validating. Due to insufficient data, model testing was not carried out.

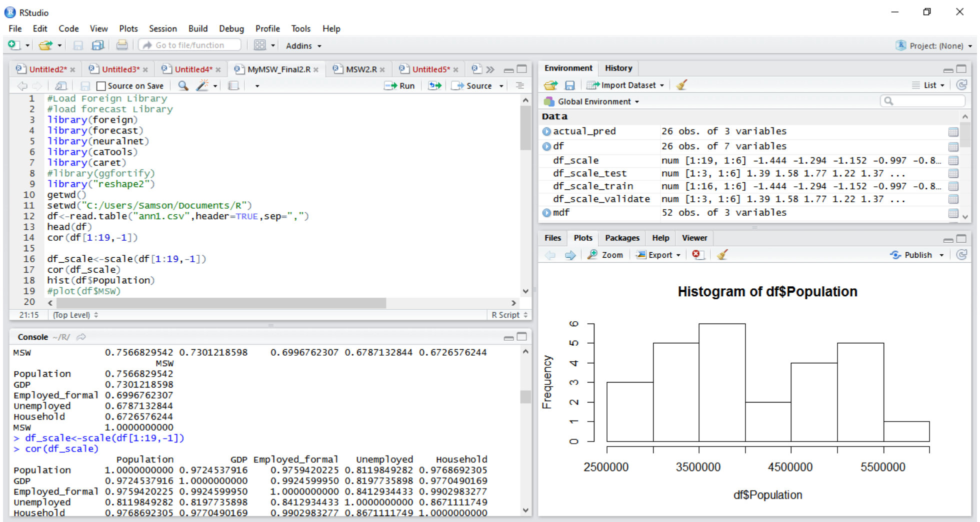

Figure 1 R Studio Modelling Environment

4 Results and Discussion

4.1 Pair-wise Correlation Matrix

The pair wise correlation matrix, shown in Table 2, indicated the relationship between the independent variables and the dependent variable (MSW). The data showed that GDP is most correlated to the quantity of MSW. The least correlated variable is number of dwellings. There is a high degree of multi-collinearity between the independent variables. Where multi-collinearity exists among two variables, one of such variable can be relaxed. However, for simplicity sake, such has not been considered due to limited data set.

4.2 Parameters

In Figure 2, Each layer and their respective weights are linked by the black line. The input layer consists of the independent variable (Population, GDP, the number of formally employed, the number of unemployed and number of households). The bias term, which can be related to a linear model intercept, is added to each step and is shown with the blue line. The weights are randomly generated with an attempt to fit the model. Once the model converges, it indicates it is ready to forecast. The accuracy of the forecasting is dependent on the precision achieved during training which is measured by the mean square error.

4.3 Assessing Impact of Data Quality

The model validation from 2011 to 2015 performs poorly in on the data analysed as shown in Figure 3. Though different tuning of the model was conducted inclusive of using 2 hidden layer at different numbers of neuron, the accuracy of the model did not improve.

Applying the same model but on another dataset of waste generation from an European country as downloaded from Eurostat (2017). The error obtained was minimal compared to that presented in Figure 3. The model performance presented in Figure 4 allows for accurate planning of future waste management strategy.

Table 2 Correlation matrix of variables

	Population	GDP	Employed formal	Unemployed	dwellings	MSW
Population	1					
GDP	0.97	1				
Employed formal	0.98	0.99	1			
Unemployed	0.83	0.84	0.86	1		
Dwellings	0.98	0.97	0.99	0.88	1	
MSW	0.60	0.61	0.57	0.58	0.51	1

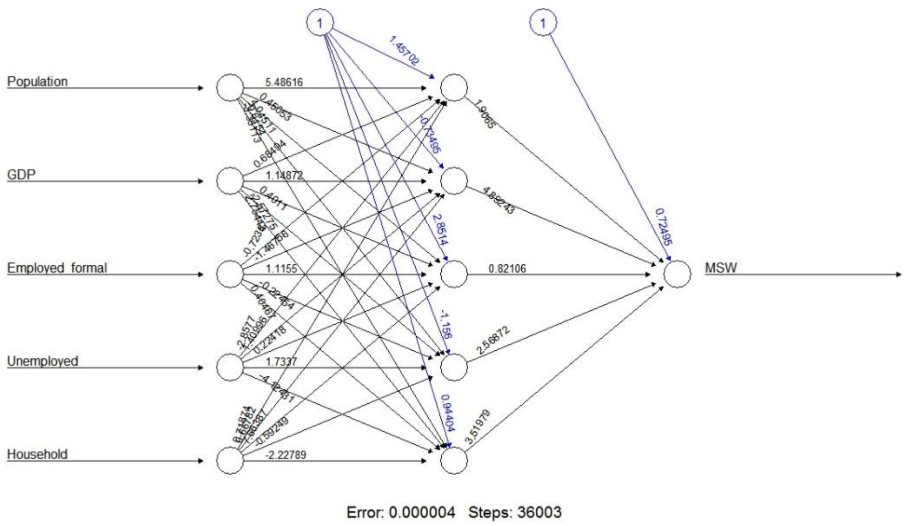

Figure 2 ANN Network (1 hidden layer with 5 neurons)

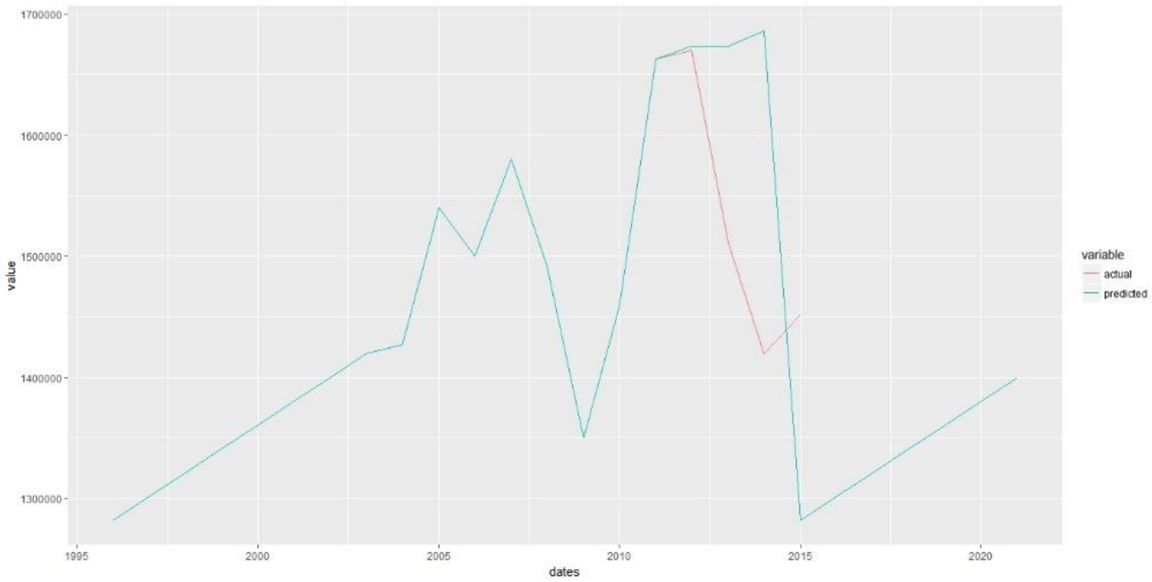

Figure 3 Waste quantity forecasting

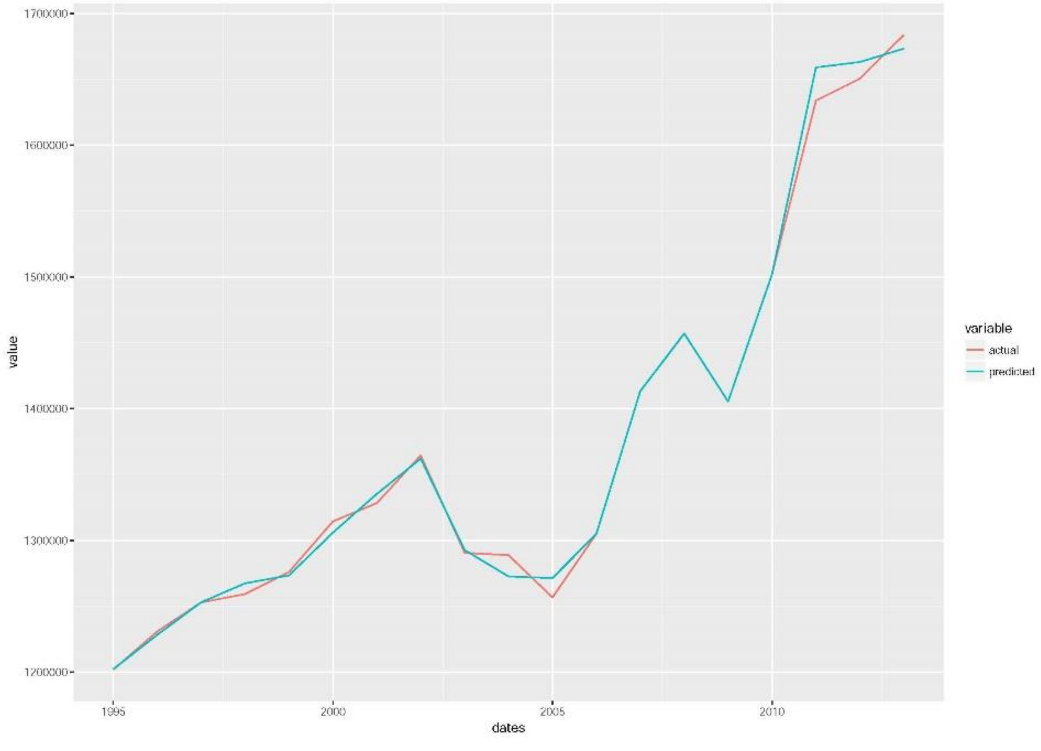

Figure 4 Waste quantity forecasting with clean dataset

5 CONCLUSION

In this study, we have reviewed forecasting methodologies for waste generation. We highlight the impact of poor waste generation data as well as other independent variables affecting waste generation. We compared data from a City in and African Country against that of the developed country in Europe. The results showed that the historical length of the data, its quality and preciseness are critical for a reliable neural network model performance. This study is part of an extended research work on Waste-to-Energy at the University of Johannesburg. The case study City in this study has installed weighing bridges at all their landfill sites to capture the actual tonnages of waste discharged. This will improve the quality of data available and assist in accurate forecasting of future generation/disposal rate. Thus, effective planning and management can be implemented.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the City of Johannesburg, Environment and Infrastructural Service Department of the City of Johannesburg and the University of Johannesburg for providing grants for this research work.

REFERENCES

Abbasi, M. and A. El Hanandeh (2016). "Forecasting municipal solid waste generation using artificial intelligence modelling approaches." Waste management (New York, N.Y.) **56**: 13-22.

Abdoli, M. A., M. F. Nezhad, R. S. Sede and B. Sadegh (2012). "Longterm forcasting of solid waste generation by the artificial neural networks." Environmental Progress and Sustainable Energy **33**: 1-9.

Adamovic, V. M., D. Z. Antanasijevic, M. Ristic, A. A. Peric-Grujic and V. V. Pocajt (2016). "Prediction of municipal solid waste generation using artificial neural network approach enhanced by structural break analysis." Environmental Science and Pollution Research: 1-13.

Antanasijevic, D., V. Pocajt, I. Popovic, N. Redzic and M. Ristic (2013). "The forecasting of municipal waste generation using artificial neural networks and sustainability indicators." Sustainable Science **8**: 37-46.

Azadi, S. and A. Karimi-Jashni (2016). "Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran." Waste Management **48**: 14-23.

Bandara, N. J. G. J., J. P. A. Hettiaratchi, S. C. Wirasinghe and S. Pilapiiya (2007). "Relation of waste generation and composition to socio-economic factors: A case study." Environmental Monitoring and Assessment **135**: 31-39.

Batinic, B., S. Vukmirovic, G. Vujic, N. Stanisavljevic, D. Ubavin and G. Vukmirovic (2011). "Using ANN model to determine future waste characteristics in order to achieve specific waste management targets -case study of Serbia." Journal of Scientific and Industrial Research **70**: 513-518.

Beigl, P., S. Lebersorger and S. Salhofer (2008). "Modelling municipal solid waste generation: A review." Waste Management **28**: 200-214.

Denafas, G., T. Ruzgas, D. Martuzevi??ius, S. Shmarin, M. Hoffmann, V. Mykhaylenko, S. Ogorodnik, M. Romanov, E. Neguliaeva, A. Chusov, T. Turkadze, I. Bochoidze and C. Ludwig (2014). "Seasonal variation of municipal solid waste generation and composition in four East European cities." Resources, Conservation and Recycling **89**: 22-30.

Dyson, B. and N.-b. Chang (2005). "Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling." Waste Management **25**: 669-679.

Eurostat. (2017). "Municipal waste by waste operation." Eurobase Data Explorer Retrieved 02/02/2017, 2017, from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=env_wasmun&lang=en.

Ferreira, J. A., M. C. Figueiredo and J. A. Oliveira (2014). "Forecasting household packaging waste generation: A case study." Springer International Publishing Switzerland: 523-538.

Garrod, G. and K. Willis (1998). "Estimating lost amenity due to landfill waste disposal." Resources, Conservation and Recycling **22**: 83-95.

Ghinea, C., E. N. Drăgoi, E.-D. Comăniţă, M. Gavrilescu, T. Câmpean, S. Curteanu and M. Gavrilescu (2016). "Forecasting municipal solid waste generation using prognostic tools and regression analysis." Journal of Environmental Management **182**: 80-93.

Grazhdani, D. (2016). "Assessing the variables affecting on the rate of solid waste generation and recycling: An empirical analysis in Prespa Park." Waste Management **48**: 3-13.

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

Intharathirat, R., P. Abdul Salam, S. Kumar and A. Untong (2015). "Forecasting of municipal solid waste quantity in a developing country using multivariate grey models." Waste Management **39**: 3-14.

Jiang, P. and X. Liu (2016). "Hidden Markov model for municipal waste generation forecasting under uncertainties." European Journal of Operational Research **250**: 639-651.

Kolekar, K. A., T. Hazra and S. N. Chakrabarty (2016). "A Review on Prediction of Municipal Solid Waste Generation Models." Procedia Environmental Sciences **35**: 238-244.

Marandi, F. and S. M. T. F. Ghomi (2016). Time series forecasting and analysis of municipal solid waste generation in Tehran city. 12th International Conference on Industrial Engineering**:** 4-8.

Masebinu, S. O., E. T. Akinlabi, E. Muzenda, A. O. Aboyade and C. Mbohwa (2017). Evaluating the municipal solid waste management approach of a city using material flow analysis. International conference on inductrial engineering and operation management. Rabat, Morocco, IEOM.

Mwenda, A., D. Kuznetsov and S. Mirau (2014). "Time Series Forecasting of Solid Waste Generation in Arusha City." Mathematical Theory and Modelling **4**: 29-40.

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

Noori, R., M. A. Abdoli, M. J. Ghazizade and R. Samieifard (2009). "Comparison of Neural Network and Principal Component- Regression Analysis to Predict the Solid Waste Generation in Tehran." **38**: 74-84.

Noori, R., A. Karbassi and M. Salman Sabahi (2010). "Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction." Journal of Environmental Management **91**: 767-771.

Ordonez-Ponce, E. A. (2004). A model for assessing waste generation factors and forecasting waste generation using artifical neural network: A case study of Chile**:** 214.

Patel, V. and S. Meka (2013). "Forecasting of Municipal Solid Waste Generation for Medium Scale Towns Located in the State of Gujarat , India." International Journal of Innovative Research in Science, Engineering and Technology **2**: 4707-4716.

Reynolds, C., A. Geschke, J. Piantadosi and J. Boland (2016). "Estimating industrial solid waste and municipal solid waste data at high resolution using economic accounts: an input???output approach with Australian case study." Journal of Material Cycles and Waste Management **18**: 677-686.

Singh, D. and A. Satija (2016). "Prediction of municipal solid waste generation for optimum planning and management with artificial neural network—case study: Faridabad City in Haryana State (India)." International Journal of System Assurance Engineering and Management: 1-7.

Sukholthaman, P., P. Chanvarasuth and A. Sharp (2015). "Analysis of waste generation variables and people's attitudes towards waste management system: a case of Bangkok, Thailand." Journal of Material Cycles and Waste Management: 1-12.

Weng, Y. C., T. Fujiwara and Y. Matsuoka (2009). "Municipal solid waste management and short-term projection of the waste discard levels in Taiwan." Journal of Material Cycles and Waste Management **11**: 110-122.

Younes, M. K., Z. M. Nopiah, N. E. A. Basri, H. Basri, M. F. M. Abushammala and K. N. A. Maulud (2015). "Prediction of municipal solid waste generation using nonlinear autoregressive network." Environmental monitoring and assessment **187**: 753.