Development of a Prediction Model for Household Solid Waste Generation and Management in an Emerging Urban Area.

Oji Achuka Nwoke* and Wilfred Ifeanyi Okonkwo
Department of Agricultural and Bioresources Engineering, Faculty of Engineering
University of Nigeria Nsukka
*Corresponding Author: Email: nwoke.oji@unn.edu.ng, Phone No.: +2348038299006

Abstract- Accurate prediction of household solid waste (HSW) generation is essential for improved solid waste management system. Modeling techniques are useful for successful and accurate prediction of HSW generation in municipalities. Two different models for predicting generation rate of HSW and the Organic fraction of household solid waste (OFHSW) for individual households of the city of Nsukka, Enugu State Nigeria were developed using multiple linear regression technique. The predictor variables in the two models were household size, average income, and education. The determined R^2 values for the two models using the predictor variables were 0.949 and 0.734 for HSW and the OFHSW, respectively. The results of tests of accuracy of the two models were undoubted as the predicted values were very close to the observed values. Testing the validity of the two models using a new set of data showed that they are suitable for actual prediction purpose with predicted R^2 values of 0.704 and 0.688 for HSW and OFHSW, respectively.

Keywords: *Prediction Model; Household Solid Waste; Generation Rate; Composition; Housing Unit Types; Socioeconomic Variables*

1. Introduction

Household solid waste (HSW) is approximately 40 to 80% of municipal solid waste (MSW) generated in Nigeria (Ogwueleka, 2013; Sha'Ato et al.,2007). Like other cities in Nigeria, increasing population, urbanization and changing of consumption patterns are resulting in the generation of increasing amounts of HSW in Nsukka (Dangi et al., 2011 , 2008).Lack of sufficient data and speedy variation of socioeconomic conditions because of increase in gross domestic product (GDP) of a region, tourism, economic activity, etc. makes prediction of waste generation rate a difficult task (Beigl et al., 2008). Modelling methods for prediction of MSW generation rate is usually a panacea for such difficult conditions (Kumar and Samadder, 2017). Regression modelling is a statistical approach that is widely used for predicting solid waste generation rate based on socioeconomic factors and other explanatory variables (Kumar and Samadder, 2017). Regression models have been widely used at municipal or community level for daily or annual prediction of MSW generation rate due to its simple algorithm and well-developed theory (Lebersorger and Beigl, 2011; Ojeda-Benítez et al., 2008; Thanh et al., 2010; Gu et al, 2015; Kumar and Samadder, 2017). The relationships between the household solid waste generation rates and the corresponding socioeconomic variables have been discussed in many of the reported studies (Kumar and Samadder, 2017; Khan et al., 2016; Vivekananda and Nema, 2014; Ogwueleka, 2013; Lebersorger and Beigl, 2011; Thanh et al., 2010; Qu et al., 2009; Sujauddin et al., 2008; Buenrostro et al., 2001). The present study considered the socioeconomic parameters household size, average income of households, and educational level of household daily manager for development of two models for prediction of Household Solid Waste (HSW) generation rate and Organic Fraction of HSW (OFHSW) generation rate in Nsukka, Enugu State Nigeria.

2.0 MATERIALS AND METHODS

2.1 Description of Study Area

The present study was carried out in Nsukka Municipality, which is a part of Nsukka Local Government Area (LGA) Enugu State, South Eastern, Nigeria. Nsukka municipality has a total landmass of 45.38 km^2 on Latitude 6°51′24″N and Longitude 7°23′45″E. Nsukka Municipality consist of six administrative wards namely Nru, Nkpunano, Onuiyi, Ihe and Owerre, University of Nigeria Nsukka (UNN) and Government Reserved Area (GRA) with a total population of 309,633 according to Federal Government of Nigeria, Census 2006 **(National Population Commission [NPC], 2006)**. The number of households in the study area is approximately 63603(NPC, 2006).

2.1 Data Collection

2.1.1. Determining the Number of Samples from Preliminary Survey

The optimal sample for Nsukka was approximated at 99% confidence interval with 10% sampling error of the means value. Since the standard deviation of the population is unknown, it is required to determine this parameter. The standard deviation SD was computed from one-day preliminary HSW survey using 25 households selected at random from the study area. The preliminary survey result showed that the average generation rate of HSW is 0.86kg/capita/day with standard deviation of 0.575 **Therefore,** the optimum sample size, n was determined using equation 1**(Kumar and Samadder, 2017; Ogwueleka, 2013; Gómez *et al.*, 2008; Abu Qdais *et al.*, 1997)**:

$$n = [\frac{z(SD)}{R}]^2 = [\frac{2.576(0.575)}{0.1}]^2 = 219\ samples \qquad (1)$$

where: n = minimum number of samples that will give the required precision; z = score determined from statistical tables of the percentage for standard normal distribution; SD=standard deviation of population which is equal to the standard deviation of the preliminary sample; and R = sampling error

2.1.2 Waste Sampling from Households in the Main Survey.

The main sampling survey was conducted consecutively for one week in the Month of April and May 2019 for 100 households making a total of 700 samples, which is greater than twice the required sample size of 219 determined in Equation (1). A survey team of 25 persons performed the samplings of the wastes. Waste sampling was carried out in each of the sampling points (Nru, Nkpunano, Onuiyi, Ihe/Owerre, and UNN/GRA) at one time for a complete 1 week for easy sampling and interpretation. The composition analysis of the waste samples was done using standard method **(ASTM D5231-92, 2016)**. The average percentage composition of the HSW in Nsukka was determined as shown in **Table 1.**

Table 1. Waste Composition in Nsukka Municipality

	Organic (%)	Metal (%)	Paper (%)	Plastic (%)	Glass (%)	Others (%)	Textile (%)
Weighted Average	59.27	5.48	22.42	6.06	4.09	2.40	0.28

2.2 Development of the Models for Prediction of HSW and OFHSW Generation Rates.

The developed prediction models were based on the principle of effect of socioeconomic factors on MSW generation rate **(Kumar and Samadder, 2017)**.

2.2.1 Dependent and Predictor Variables included in the Model

The description of the selected dependent and predictor variables is presented in **Table 2.**

Table 2. Predictor variables for the mathematical model and information analysis

Variable name	**Symbol**	**Type**	**Unit of measure**
Educational Level of Household Daily Manager	X_{EDU}	Independent discrete	Education /Household
Household Size	X_{HAB}	Independent discrete	Persons/household
Average monthly income per household	X_{INC}	Independent continuous	Average Monthly Income/household/(NGN)
Per capita production of Household solid waste (HSW) per day	Y_{HSW}	Dependent continuous	kg/person-daily
Per capita production of OFHSW Per day	Y_{OFHSW}	Dependent continuous	kg/person-daily

2.2.2 Multiple linear regression (MLR) analysis for prediction of HSW generation rate

A multiple linear regression model was chosen as the most excellent alternative due to its dependability and proof in earlier studies **(Ojeda-Benítez *et al.*, 2008; Thanh *et al.*, 2010; Gu *et al*, 2015; Kumar and Samadder, 2017)**. The model is as shown in **Equation (2)**:

$$\boldsymbol{Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{k-1} X_{k-1} + \beta_k X_k + \varepsilon} \tag{2}$$

Where, Y is the dependent variable (HSW generation, OFHSW generation); α is the intercept;

$\beta(1, 2, \ldots, \mathrm{k} - 1, \mathrm{k})$ are the slopes and signify the average change in the dependent variable; $X(1, 2, \ldots, \mathrm{k} - 1, \mathrm{k})$ are independent variables, such as household size **(Kumar and Samadder, 2017; Suthar and Singh,2015; Dangi et al., 2011; Purcell and Magette, 2009)**, household income **(Kumar and Samadder, 2017; Abdoli et al., 2012;Ojeda-Benítez et al., 2008)** and education level of household manager **(Keser et al., 2012)**; k is the number of independent variables included in the model; and ε is the average random error.

It is necessary to determine the statistically significant regression coefficients to develop the regression model **(Pires et al., 2008)**. The required statistical analysis was done in Statistical Package for the Social Sciences (SPSS), version 21.0 software for developing the model. The present study relied on the analysis of variance (ANOVA) output given by the statistical software, shown in ANOVA in **Table 3** for both the models. Another statistical test ‘‘t-test” was conducted at a = 0.05, to identify the correlations between dependent variables and coefficient parameters (β_k). The summary of both the model equations is given in **Table 4**. The coefficient of determination (R^2) was used to measure how well a multiple regression model fits a data set. As an alternative of R^2 as a measure of model competence, the adjusted multiple coefficients of determination (Adjusted R-Square or Ra^2) was used, because it takes into account both the sample size and the number of parameters included in the model **(Mendenhall and Sincich, 2012).**

Table 3. ANOVA Table for HSW and OFHSW Generation Rate.

ANOVA Table for HSW and OFHSW Generation Rate

Model		Sum of Squares	df	Mean Square	F	Sig.
Y_{HSW}	Regression	3.837	3	1.279	287.948	.000[b]
	Residual	.204	46	.004		
	Total	4.041	49			
Y_{OFHSW}	Regression	1.528	3	.509	42.226	.000[b]
	Residual	.555	46	.012		
	Total	2.083	49			

b. Predictors: (Constant), Household Size (X_{hab}), Average Income of Household (X_{inc})(NGN), Educational Level of Household Daily Manger (X_{edu}),

Table 4. Model Summary for HSW and OFHSW Generation Rate

Model Summary for HSW and OFHSW Generation Rate

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
HSW	.974[a]	.949	.946	.066648	1.911
OFHSW	.857[a]	.734	.716	.109824	1.644

a. Predictors: (Constant), Household Size (XHAB), Average Income of Household (XINC)(NGN), Educational Level of Household Daily Manger (XEDU),

2.3. Prediction accuracy of the model and validation

To evaluate the accuracy of the proposed models of the current study, four popular measures of prediction accuracy namely mean absolute deviation (MAD), mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) **were used (**Dodge, 2008; **Mendenhall and Sincich, 2012; Kumar and Samadder, 2017) as shown in Equations (3)-(6).** For better performance of the model, the test results **[Equations (3)-(6)**] should be nearer to zero **(Azadi and Karimi-Jashni, 2016)**.

$$\boldsymbol{MAD = \frac{\sum_{i=1}^{n}|A_i - F_i|}{n}} \tag{3}$$

$$\boldsymbol{MSE = \frac{\sum_{i=1}^{n}(A_i - F_i)^2}{n}} \tag{4}$$

$$\boldsymbol{RMSE = \sqrt{\frac{\sum_{i=1}^{n}(A_i - F_i)^2}{n}}} \tag{5}$$

$$\boldsymbol{MAPE = \frac{\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i}}{n} \times 100} \tag{6}$$

Where A_t = Actual or Observed value, F_t = Forecasted or predicted value, n = sample size.

The most effective and widely used technique for regression model validation is the use of the developed model to a new set of collected data for prediction. This can be done by comparing the percentage of variability (R^2 pred) in the developed model using new data set to the coefficient of determination (R^2) of the developed model from sampled data. The expression of R^2 pred is given in **Equation. (7) (Kumar and Samadder, 2017):**

$$\boldsymbol{R^2_{pred.} = 1 - \left(\frac{\sum_{i=n+1}^{n+m}(Y_i-\hat{Y}_i)^2}{\sum_{i=n+1}^{n+m}(Y_i-\bar{Y})^2}\right)} \quad \textbf{(7)}$$

where $R^2_{pred.}$ is the percentage of variability in the new set of data, Y_i is the observed values of the dependent variables of the new data set, $\hat{Y}_i$ is the predicted values of the dependent variables from the fitted model, $\bar{Y}$ is the sample mean of the observed data, n is the number of samples used for the model development, and m is the number of samples used for the model validation. **Montgomery et al. (2006) recommended a minimum of 15–20 new observations for validation of a** regression model. For this study, 100 households were surveyed; the data was divided into two equal parts (50:50). 50 households were used for model development and the other 50 households were used for model validation (excluding 50 households that were used for the model development).

3.0 RESULTS AND DISCUSSION

3.1. Multiple Linear Regression (MLR) Equations

The regression equation for mean HSW generation rate is given by **Equation (8)**

$$\boldsymbol{Y_{HSW} = -0.036 - 0.030X_{edu} + 0.111X_{hab} + 0.0000008483X_{inc}} \quad (8)$$

The regression equation for mean OFHSW waste generation rate is given by **Equation (9).**

$$\boldsymbol{Y_{OFHSW} = 0.036 - 0.030X_{edu} + 0.068X_{hab} + 0.0000003493X_{inc}} \quad (9)$$

where

$Y_{HSW} = Household\ Solid\ Waste(kg/capita/day)$

$Y_{OFHSW} = Organic\ Fraction\ of\ Household\ Solid\ Waste(kg/capita/day)$

$X_{hab} = household\ size$(Persons)

$X_{inc} = household\ \ Income\ (Naira/capita/month$

$X_{edu} = Education\ level\ of\ \ Household\ daily\ manager.$

The results of the global F-test **(Table 3)** showed that the level of significance (α = 0.05, 95% confidence level) exceeded the observed level of significance (p = 0.000), which indicated at least one of the regression coefficients was non-zero, thus the null hypothesis is rejected, which signifies that the models are statistically valid.

For the model developed for prediction of HSW generation rate (**Equation (8)**), this indicated that, all the selected independent variables except ''X_{EDU}'' (p = 0.197) were statistically significant and were expected to influence the prediction of HSW generation rate. For the model developed for prediction of OFHSW generation rate **(Equation (9))**, this indicated that only the variable "X_{HAB}" was statistically significant and is expected to influence the prediction of OFHSW generation rate.

Although t-test showed that the p-values of the variable ''X_{EDU}'' was higher than 0.05 (a = 0.05) in both models and the p-value for "X_{INC}" was higher than 0.05 (α = 0.05) for the developed model for prediction of OFHSW generation rate, it did not necessarily mean that the variables were not contributors to the HSW and OFHSW generation rates (which is supported by **Hockett et al., 1995**). **Ogwueleka (201**3) also concluded that income and education are the major influencing factors of HSW generation rate. Thus, the independent variables, ‘‘X_{EDU}” and “X_{INC}” have been included in the proposed prediction models.

3.2. Assessment of MLR model for prediction of the HSW and OFHSW generation rate

The R^2 values of the two model equations were 0.949 and 0.734 respectively. This indicated that 94.9% of HSW generation rate and 73.4% of OFHSW generation rate were influenced by the selected socioeconomic parameters.

3.3. Model validation and assessment of prediction accuracy

The model efficacy was checked by comparing the R^2pred values of the new set of data (50 households) and the values of the coefficient of variation (R^2) of the developed MLR models. The R^2pred values of both the models were 0.704 (HSW) and 0.688 (OFHSW) respectively, which were slightly smaller than the R^2 values [0.949 (HSW and 0.734(OFHSW)] of the original fit models. The values of the accuracy tests for MAD, MSE, RMSE, and MAPE of the model for prediction of HSW generation rate were 0.038,0.007,0.085, and 5.28 respectively, whereas the test results of the model for prediction of OFHSW generation rate were 0.037, 0.007, 0.086, and 10.52, respectively. The accuracy tests conducted on the developed models showed very appreciable results and thus the present models can be used successfully for prediction of HSW and OFHSW generation rates in actual practice.

4.0 CONCLUSIONS

In the present study, two models have been developed using MLR technique, one for prediction of HSW generation rate and the other for prediction of OFHSW generation rate. This study revealed that the variable household size had the maximum influence on the generation rate of HSW and OFHSW. The R^2 values of the prediction models for HSW generation rate and OFHSW generation rate were 0.949 and 0.734, respectively. The developed models will help the decision makers to estimate the MSW generation rate accurately for effective planning of MSWM facilities for similar cities. Identification and inclusion of other influencing explanatory variables may further enhance the prediction accuracy of the model.

References

Abdoli, M.A., Nezhad, M.F., Sede, R.S., Behboudian, S., 2012. Longterm forecasting of solid waste generation by the artificial neural networks. Environ. Progress Sustain. Energy 31 (4), 628–636.

AbuQdais, H.A., Hamoda, M.F., Newham, J., (1997). Analysis of residential waste at generation sites. Waste Management and Research 15 (4), 395–406.

ASTM D5231-92, 2016. Standard Test Method for Determination of the Composition of Unprocessed Municipal Solid Waste. ASTM International, West Conshohocken, PA.

Azadi, S., Karimi-Jashni, A., 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 Manage. 48, 14–23.

Beigl, P., Lebersorger, S., Salhofer, S., (2008). Modelling municipal solid waste generation: a review. Waste Manage. 28, 200-214.

Buenrostro, O., Bocco, G., Vence, J., 2001. Forecasting generation of urban solid waste in developing countries—a case study in Mexico. J. Air Waste Manage. Assoc. 51 (1), 86–93.

Dangi, M. B., Urynowicz, M. A., Gerow, K. G., & Thapa, R. B. (2008). Use of stratified cluster sampling for efficient estimation of solid waste generation at household level. Waste Management & Research, 26, 493-499.

Dangi, M.B., Pretz, C.R., Urynowicz, M.A., Gerow, K.G., Reddy, J.M., 2011. Municipal solid waste generation in Kathmandu, Nepal. J. Environ. Manage. 92 (1), 240– 249.

Dodge, Y. (2008). Coefficient of Skewness; Coefficient of Kurtosis In: The Concise Encyclopedia of Statistics (Springer Reference). Springer, New York, NY. https://doi.org/10.1007/978-0-387-32833-1_64. pp. 91-95.

Gomez, G., Menesesa, M., Ballinas, L., Castells, F., 2008. Characterization of urban solid waste in Chihuahua, Mexico. Waste Manage. 28 (12), 2465–2471. http:// dx.doi.org/10.1016/j.wasman.2007.10.023.

Gu, B., Wang, H., Chen, Z., Jiang, S., Zhu, W., Liu, M., … Bi, J. (2015). Characterization, quantification and management of household solid waste : A case study in China, Resources, Conservation and Recycling *98*, 67–75.

Hockett, D., Lober, D.J., Pilgrim, K., 1995. Determinants of per capita municipal solid waste generation in the Southeastern United States. J. Environ. Manage. 45 (3), 205–217.

Keser, S., Duzgun, S., Aksoy, A., 2012. Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey. Waste Manage. 32 (3), 359–371.

Khan, D., Kumar, A., Samadder, S.R., 2016. Impact of socioeconomic status on municipal solid waste generation rate. Waste Manage. 49, 15–25.

Kumar, A., Samadder, S.R., 2017. An empirical model for prediction of household solid waste generation rate–A case study of Dhanbad, India. Waste Manage. 68, 3–15.

Lebersorger, S., Beigl, P., 2011. Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel. Waste Manage. 31 (9), 1907–1915

Mendenhall, W., Sincich, T., 2012. A Second Course in Statistics: Regression Analysis. Prentice Hall, Boston, USA.

Montgomery, D., Peck, E., Vining, G., 2006. Introduction to Linear Regression Analysis. Wiley, New York, USA.

NPC (2006). population and housing census of the Federal Republic of Nigeria : Housing Characteristics and amenity table (priority Tables LGA), Vol.II. www.population.gov.ng

Ogwueleka,T.C.(2013).Survey of household waste composition and quantities in Abuja, Nigeria. Resources, Conservation and Recycling 77 (2013) 52– 60.

Ojeda-Benítez, S., Lozano-Olvera, G., Morelos, R.A., Armijo de Vega, C., 2008. Mathematical modeling to predict residential solid waste generation. Waste Manage. 28, S7–S13.

Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C., 2008. Selection and validation of parameters in multiple linear and principal component regressions. Environ. Modell. Software 23 (1), 50–55.

Purcell, M., Magette, W.L., 2009. Prediction of household and commercial BMW generation according tsocio-economic and other factors for the Dublin region Waste Manage. 29 (4), 1237–1250.

Sha'Ato, R., Aboho , S.Y., Oketunde, F.O., Eneji , I.S., Unazi , G., Agwa, S. (2007). Survey of solid waste generation and composition in a rapidly growing urban area in Central Nigeria Waste Management 27,352–358

Sujauddin, M.S., Huda, S.M.S., Rafiqul Hoque, A.T.M., (2008). Household solid waste characteristics and management in Chittagong, Bangladesh. Waste Manage. doi: 10.1016/j.wasman.2007.06.013.

Suthar, S., & Singh, P. (2015). Household solid waste generation and composition in different family size and socio-economic groups: A case study. *Sustainable Cities and Society*, *14*, 56–63. http://doi.org/10.1016/j.scs.2014.07.004

Thanh, N.P., Matsui, Y., Fujiwara, T., 2010. Household solid waste generation and characteristic in a Mekong Delta city, Vietnam. J. Environ. Manage. 91 (11), 2307–2321.

Vivekananda, B., Nema, A.K., 2014. Forecasting of solid waste quantity and composition: a multilinear regression and system dynamics approach. Int. J. Environ. Waste Manage. 13 (2), 179–198.