Global J. Environ. Sci. Manage., 3(4): 385-402, Autumn 2017
DOI: 10.22034/gjesm.2017.03.04.005

ORIGINAL RESEARCH PAPER

Predicting waste generation using Bayesian model averaging

M.G. Hoang [1,], T. Fujiwara [2], S.T. Pham Phu [1], K.T. Nguyen Thi [3]*

[1]*Graduate School of Environmental and Life Science, Department of Environmental Science, Okayama University, 3-1-1 Tsushima, Kita, Japan*

[2]*Waste Management Research Center, Okayama University, 3-1-1 Tsushima, Kita, Okayama 700-8530, Japan*

[3]*Faculty of Environmental Engineering, National University of Civil Engineering, 55 Giai Phong Road, Hai Ba Trung, Ha Noi, Viet Nam*

Received ; 23 April 2017 revised ; 24 July 2017 accepted ; 7 August 2017 available online 1 September 2017

ABSTRACT: A prognosis model has been developed for solid waste generation from households in Hoi An City, a famous tourist city in Viet Nam. Waste sampling, followed by a questionnaire survey, was carried out to gather data. The Bayesian model average method was used to identify factors significantly associated with waste generation. Multivariate linear regression analysis was then applied to evaluate the impacts of significant factors on household waste production. The model obtained from this study indicated that household location, household size, house area per person, and family economic activity are important determinants of the waste generation rate. The models could explain about 34% of the variation of the per capita daily waste generation rate. Diagnostic tests and model validation results showed that the regression model could provide reliable results of estimated household waste. The study revealed that per capita urban household waste generation is 70–80% higher compared to a rural household. The models also showed that if a family ran a business from home, the household waste generation rate would increase by about 35%. This result provides reliable information for better waste collection and management planning. Two other significant variables (family size and house area per capita) do not contribute much (less than 20%) to waste generation. Variables accounting for household income, presence of a garden, number of rooms in a house, and percentage of members of different ages were proven to be not significant. The study provides a reliable method for estimating household waste generation, providing decision makers useful information for waste management policy development.

KEYWORDS: *Bayesian model average (BMA); Multivariate linear regression; Municipal solid waste management (MSWM); Prognosis model; Waste generation.*

INTRODUCTION

Solid waste generation is a result of the production and consumption cycle. Rapid urbanisation and industrialisation in developing countries have led to a dramatic increase in the volumes of municipal solid waste (MSW) generated daily (Abdoli *et al.*, 2016). Currently, more than 24 million tonnes of solid waste are generated in Viet Nam annually, with the figure likely to reach 52 million tonnes by 2020. The increasing volume of MSW has become an emerging environmental issue for authorities in Viet Nam (Nguyen *et al.*, 2013). The growing amount of waste causes negative impacts on the environment and human health owing to inadequate disposal (Ngoc and Schnitzer, 2009). Eighty percent of MSW was disposed of in landfills without being recycled, reflecting material and energy losses to society (Ghinea *et al.*, 2016). Thus, integrated waste management, including recycling material and energy from MSW as well as

*Corresponding Author Email: *gianghm@nuce.edu.vn*
Tel.: +81 86 251 8994 Fax: +81 86 251 8994
Note: Discussion period for this manuscript open until December 1, 2017 on GJESM website at the "Show Article".

resource conversation, have become significant issues (van de Klundert *et al.*, 2001; Zurbrugg *et al.*, 2012). One of the challenges faced by local governments is predicting solid waste volumes reliably in order to devise appropriate actions and plans (Ghinea *et al.*, 2016). Predicting waste generation volumes is increasingly essential in waste collection planning and treatment strategies, and establishing policies toward a sustainable waste management system (Chen and Chang, 2000; Thanh and Matsui, 2011; Abbasi *et al.*, 2012). In the case of Viet Nam, National Technical Regulation QCVN 07:2010/BXD (MOC, 2010) provided a method to estimate waste generation for five types of urban areas, based on population and a waste generation rate, which can be determined as outlined in the document. However, these results are not reliable in terms of practical application. Solid waste generation is impacted not only by demographic factors, but also by social, economic, and other factors (e.g. family expenses or waste prevention policies). Therefore, a more recent edition of this regulation, QCVN 07:2016 (MOC, 2016), does not use either this method or any other model for waste estimation. The lack of research into and methods for estimating waste generation have led to considerable challenges in municipal waste management in Viet Nam. Various modelling techniques, such as time series analysis (Abbasi *et al.*, 2012; Kolekar *et al.*, 2016), artificial neural network (Noori *et al.*, 2010; Karpušenkaitė *et al.*, 2016; Memarianfard *et al.*, 2017), and fuzzy logic (Oumarou *et al.*, 2012; Vesely *et al.*, 2016), were applied to develop predictive models for solid waste generation and environmental management. According to Beigl *et al.* (2008), multivariate methods are very complex given the numerous interactions among the parameters and the difficulty of validating the models. Linear regression analysis, on the other hand, was popularly applied to estimate waste generation (Buenrostro *et al.*, 2001; Bach *et al.*, 2004; Beigl *et al.*, 2008; Thanh *et al.*, 2010; Ghinea *et al.*, 2016). The term 'linear' gives the casual observer the impression that linear models can only handle simple data sets; however, linear models can easily be expanded and modified to handle complex data sets, and they are used in empirical investigations and data prediction (Faraway, 2005). Previous studies modelled total municipal waste generation using various variables, and reported that the municipal waste generation rate was significantly correlated with economic factors. At the regional and national levels, Hockett *et al.* (1995) created a multiple linear regression model of per capita waste generation that is expressed by demographic, economic, and structural determinants. Their study found that the per capita purchase of goods and waste treatment fees are significant determinants of waste generation, and demographic factors are not significant as correlates of waste production. Thøgersen (1996) studied 18 member countries of the Organisation for Economic Co-Operation and Development (OECD), and showed that Gross Domestic Product (GDP) per capita explained 50% of the variations in per capita waste generation (R^2=0.5) based on simple linear regression analysis. Exponential and polynomial linear regression analysis in this research also proved that there was a significant correlation between GDP per capita and waste produced per capita. Daskalopoulos *et al.* (1998) provided predictive models for the European Union and the United States of America, showing the total amount of waste increases annually with gross GDP and population acting as predictor variables in polynomial equations. Another study estimated MSW generation using a multivariate linear regression model that considered the GDP per capita, infant mortality per 1,000 births, population of 15- to 59-year-olds, and average household size as predictor variables (Beigl *et al.*, 2004). Considering factors affecting household waste generation, Dennison *et al.* (1996) and Abu Qdais *et al.* (1997) estimated the household waste generation rate based on the number of members in a family, using linear regression analysis. Abu Qdais *et al.* (1997) indicated that the correlation between the amount of waste generated and the household size was weak (R=0.33), whereas the relationship between waste generated and property rental fee attributed to family income was strong (R=0.83). Lebersorger *et al.* (2003) found that for multi-family dwellings, significant linear correlation existed between the quantity of waste generated and the house type and age. A significant positive correlation between the number of rooms in a house and the waste production rate was uncovered by Monavari *et al.* (2012). Multivariate linear regression analysis was also applied widely in research on waste generation forecasting (Grazhdani, 2016). Stepwise analysis is normally utilised to ensure that the final regression model provides the best fit (Chang *et al.*, 2007; Shamshiry *et al.*, 2014; Boulet *et al.*, 2016; Akhtar *et al.*, 2017). However, using this analysis method was not recommended because it did not correctly determine the best set of variables and tended to yield irreproducible results (Derksen and Keselman, 1992; Hamby, 1994; Thompson, 1995). In addition, diagnostic checks to verify the statistical adequacy of the model were not carried out (Bdour *et al.*, 2007; Karpušenkaitė *et al.*, 2016). Not much attention was paid to model validation with a new data set (Lebersorger and Beigl, 2011; Ghinea *et al.*, 2016; Akhtar *et al.*, 2017). The two analyses noted above were essential to evaluate the reliability and performance of a linear regression model. Benítez *et al.* (2008) developed a prognosis model for residential solid waste generation using simple and multivariate linear regression, with household income, education levels, and household size as explanatory variables. The selected model could explain 51% of the variation in the waste generation rate but used all the predictor variables proposed. Apparently, in multivariate linear regression models, the higher the number of independent variables, the higher the value of the coefficient of determination (i.e. R-square). However, choosing a model based on the maximum R-square value is not a feasible option when dealing with many independent variables. This study aimed to provide a reliable model to help decision makers and stakeholders to forecast quantities of household waste. The Bayesian Model Average (BMA) method was used instead of stepwise regression to select predictor variables and prevent noise variables from gaining entry to the model (Derksen and Keselman, 1992) Multiple linear regression models were developed, with significant determinant variables being chosen by the BMA. Diagnostic tests for the hypothesis of the linear assumptions and a conventional validation method were conducted used to evaluate the performance of the model. The current study used data from Hoi An City, Viet Nam, carried out in 2015.

MATERIALS AND METHODS

The general procedure of the methodology applied is briefly described as follows. First, household waste sampling and a questionnaire survey were carried out to gather data on waste generation and explanatory variables. Next, the BMA method was used to select significant predictor variables for the regression model. Then, the regression coefficients of the model were identified. Finally, the test for linear hypothetical assumptions and validation were performed to evaluate the model's performance.

Case study and data collection

The research was carried out in Hoi An City (HAC), the cultural and tourist centre of Quang Nam on the south central coast of Viet Nam. According to the Hoi An Statistical Year Book 2013, the city has a population of around 93,000 (HASD, 2013). Tourist activities characterise the municipality, which attracts about 1.5 million of tourists annually, and provides considerable employment. MSW in Hoi An is currently disposed of at open dump landfills without any precautions or any operational controls. One composting plant with a capacity of 55 tonnes per day has been operating inefficiently, and the product did not sell well in the market. The open dump landfill and the composting plant have caused huge adverse effects on the environment and public health. Hoi An is the first and only city in Viet Nam to successfully carry out waste separation (biodegradable and nondegradable waste) at source, and the awareness of residents, as well as the efficiency of waste separation, has been increasing (Chu, 2014). However, all waste treatment technologies in Hoi An have failed, including a new incinerator installed in 2015. Chu (2014) also showed that 30% of communities have been affected by environmental pollution related to solid waste treatment and recycling activities; 13 out of 44 communities reported that residents often complained about the collection system. Moreover, owing to the failure of new treatment plants, residents have lost faith in waste managers and authorities. The motivation of residents to separate waste at source might also be lost, since they do not see any improvements in the waste treatment practices and their surrounding environment. Citizens' lack of faith could negate the future efforts of the authorities to improve the waste management system. Therefore, the development of a sustainable solid waste management situation is an urgent need in HAC. The result of this study will provide an applied prognosis model with which to estimate the waste generation from households in Hoi An, which is an essential factor of MSW collection and management planning. The waste generated by families was assessed through door-to-door sampling of the whole city in 2015. To reduce variations in household waste generation, the stratified random sampling method was applied. The city was divided into two strata based on a rural–urban topology. The number of samples was estimated from the number of households in each stratum, with a sampling ratio of 13 households per 1000 (Hoang *et al*., 2017). The statistical sample size needed was 280; 321 households participated in the sampling program. Waste generated from households was sampled over 14 consecutive days by 25 students from Da Nang University and two authors. Every household participating in the program was given a code marker to put on their door to avoid collection mistakes. The sample collection and analysis procedures are presented in Fig. 1.

Then, a survey was conducted using face-to-face interviews at all households involved in the project. This was done to obtain information on the personal and socio-economic background of the family, such as house area, presence/absence of a garden, family size, ages of family members, and monthly income. The results obtained from this survey were interpreted to identify the explanatory variables for the mathematical models of estimates of household waste generation.

Variables used in modelling

Regression analysis was used to explain the relationship between the dependent variable Y (response or output variable), and one or more independent (predictor) variables X. Thus, analysis of data gathered from the questionnaire survey was followed by identifying the variables involved. The database comprised ten independent variables representing factors influencing the demographics, geography, and economics of household waste generation. The response variable is the mean of the per capita waste (kg/capita/ day) gathered from the household waste samples. Three types of independent variables, including categorical, discrete, and continuous variables, were employed. Table 1 explains the variables included and symbols assigned.

A variable matrix consisting of the information on households/families sampled was constructed and used as the input dataset for the analysis.

Selection of determinant variables

The data set was divided randomly into two: 70% of the data was used as the training set and the other 30% was used for testing. The training set was used to determine the predictor variables and identify the coefficient of the model, while the testing set was used to validate the model. The BMA method (Raftery *et al*., 1997, Hoeting *et al*., 1998) was utilised to identify the combination of significant independent variables that best explains MSW generation. The 'best' model can provide the most precise prediction with a reasonable number of variables or accurate estimations for new cases (Raftery *et al*., 1997). The BMA provides a consistent mechanism of accounting for model uncertainty; this is often ignored in model selection, leading to overfitting models and possibly causing over-confident inferences (Hoeting *et al*., 1999; Fernández *et al*., 2001). According to Hoeting *et al.* (1999), the BMA also improves the out-of-sample predictive performance of linear models. A BMA solution to this problem provides optimal predictive ability by averaging over all possible models (Madigan and Raftery, 1994). Quantities of interest and parameter estimates are provided via direct application of the principles described as follows:

The posterior distribution given data Z of the quantity of interest Δ, such as a model parameter or a future observable, is defined by Eq. 1.

$$p(\Delta \mid Z) = \sum_{k=1}^{K} pr(\Delta \mid M_k, Z)\, pr(M_k \mid Z) \quad (1)$$

Where, M_1, M_2, ..., M_k are the models under

Fig. 1: Waste sample collection procedures

Table 1: Types of variables in linear regression

Type of variable	Variable	Symbol	Unit/value/types	Determination of variables
Response variable Dependent continuous	Per capita waste generation per day	Y_{hhw}	kg per capita per day	Average daily waste generated from each family member in a household
Predictor variables				
Independent categorical	Household location	X_{plc}	Urban (=1) Rural (=0)	If the house is located in an urban area If the house is located in a rural area
Independent categorical	House garden	X_{gar}	Yes (=1) No (=0)	The house has a garden The house does not have a garden
Independent categorical	Home business	X_{bus}	Yes (=1)	The family members run a business (e.g. convenience store, restaurant, café bar, shop, mini hotel, vehicle rental) from home
			No (=0)	The family members do not run a business from home
Independent discrete	Family income	X_{inc}	1 2 3 4 5 6	Very low: less than 500 VND per person per month Low: 500–1,200 VND per person per month Lower-middle: 1,200–2,500 VND per person per month Upper-middle: 2,500–4,000 VND per person per month High: 4,000–6,000 VND per person per month Very high: more than 6,000 VND per person per month
Independent discrete	Household size	X_{siz}	Number	The number of individuals in the family
Independent discrete	Number of rooms	X_{rom}	Number	The number of rooms in the house
Independent continuous	House area	X_{are}	m^2	The total area of the house
Independent continuous	House area per person	X_{pa}	m^2 per person	The area of the house divided by the number of family members
Independent continuous	% of children	X_{chi}	Percentage	The percentage of people younger than 20 years in the family
Independent continuous	% of adults	X_{adu}	Percentage	The percentage of people aged 20-59 years in the family
Independent continuous	% of old people	X_{old}	Percentage	The percentage of people older than 59 years in the family

consideration. Eqs. 2 and 3 give the posterior probability for model M_k and the integrated likelihood of M_k respectively.

$$pr(M_k \mid Z) = \frac{pr(Z \mid M_k)pr(M_k)}{\sum_{k=1}^{K} pr(Z \mid M_k)pr(M_k)} \tag{2}$$

$$pr(Z \mid M_k) = \int pr(Z \mid \theta_k, M_k)pr(\theta_k \mid M_k)d\theta_k \tag{3}$$

Where, θ_k is the vector of parameters of model M_k, $pr(\theta_k \mid M_k)$ is the prior density of the parameters under the model, $pr(D \mid \theta_k, M_k)$ is the likelihood, and $pr(M_k)$ is the prior probability that M_k is the actual model. The BMA estimates a parameter θ using Eq. 4.

$$\hat{\theta}_{BMA} = \sum_{k=1}^{K} \theta_k pr(M_k \mid Z) \tag{4}$$

The Bayesian information criterion (BIC) approximation is formally defined in Eq. 5. The BIC is used as a criterion for model selection from the set of models, and the model with the lowest BIC approximation is preferred.

$$BIC = -2.\ log\ (RSS_p) + p.log\ n \tag{5}$$

where, RSS_p is the squared sum of residuals in the fitting sample data for the model with p independent variables, p is the number of regressors including the intercept, and n is the number of observations or the sample size.

Multivariate linear regression model

A multivariate linear regression model is described in Eq. 6.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \varepsilon \quad (6)$$

Where, α is the intercept term indicating the mean of the dependent variable Y in case all predictor variables X equal 0; and β is a vector of β_i, the slope of the model, that explains the average change in the dependent variable. The residual ε represents the difference between estimated and observed values. ε may include measurement error, although that is often due to the effect of variables that are unincluded or unmeasured (Faraway, 2005). In linear regressions, it is assumed that the errors are normally distributed, independent, and have equal variance σ^2 ($\varepsilon \sim N(0,\sigma^2 I)$). The correlation of residuals is vital for time series data because time series regression accounts for autocorrelations between times. Meanwhile, in non-time series regression, the independence of errors is presumed or at least minimised. Theoretically, the residuals from the model should not be correlated with either independent or dependent variables.

Testing model assumptions

The validity of the assumptions underlying the chosen model should be verified. The residual ε was used to test the linear model assumptions. Formal diagnostic tests can ensure the accuracy of the results but may be powerless to detect unexpected problems, especially in data related to social and human activities. Graphical techniques are usually more efficient at revealing the overall structure of the data set. They tend to be more versatile and informative (Faraway, 2005). Moreover, graphical methods may be useful for describing and understanding the underlying structure of the data (Wilk and Gnanadesikan, 1968). Therefore, in this study, the graphical approach was applied as the diagnostic test for the hypothesis of the assumptions of the linear model. The normality of residual distribution is tested with a normal quantile plot of the residuals (Wang and Bushman, 1998), in which the ordered residuals from the fitted model (vertical axis) are plotted against the reference line of a normal distribution having the same mean and variance (horizontal axis). The model residual points should fall close to the reference line on such a plot if the errors are normally distributed. Violations of normality often occur because the distributions of either the predictor or the response variable are significantly not normal. We plotted the residuals versus the fitted values and the independent variables to find ways to improve the model. A useful method is to transform the predictor variable if the non-random shape occurs in only one plot. If it happens in more than one plot, we should transform the response variable to improve the model. The plot of residuals versus estimated values (fitted values) can also indicate constant variance if the scatter is symmetric vertically around zero. There are some approaches to dealing with non-constant variance violations in a linear regression model. Weighted least squares or transformations of the response variable can be used to achieve a constant variance of the outcome variable (Faraway, 2005). Likewise, to test for violations of independence, the distribution of the residuals should be random and symmetric around zero under all conditions. Outlier observations which do not fit the model, and influential observations that have large effects on the model, will be detected. The outlier test was carried out using the Bonferroni correction method (Faraway, 2005), and the Cook statistic was used for diagnostic tests of influence (Cook, 1977).

Model evaluation and validation

A conventional validation approach using an external validation method was applied to test the model to avoid over-fitting (Faber and Rajkó, 2007). This requires the validation samples to be entirely different from the training samples that constructed the model; this is necessary to properly assess the model's ability to forecast for unknown future samples (Bleeker *et al.*, 2003; Faber and Rajkó, 2007). To ensure the model can perform using a new data set, the authors divided the original data into two subsets including a training set (70%) and a testing set (30%). The former was used to construct the model, whereas the latter was for validation. A combination of statistical metrics, including coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), mean absolute error (MAE), root mean square error (RMSE), and normalised root mean square error (NRMSE), were applied to assess the model performance. R^2 is a useful property indicating the goodness of fit of the model. R^2_{adj} also indicates how well the model fits the data, but it adjusts for the number of independent variables in the model. MAE and RMSE are useful measures widely used to evaluate models. MAE assigns the same weight to all kinds of errors, which is appropriate to describe uniformly distributed errors, while RMSE favours errors with larger absolute values and is appropriate to explain normally distributed errors (Chai and Draxler, 2014). In this study, RMSE was used to assess the performance of the predictive model since the residuals of the linear regression model are expected to be normally distributed. If the RMSE of the test data set is significantly higher than that of the training data set, over-fitting occurs. If the two RMSEs are close, the model is valid and can be used to predict unknown data. However, the ranges of training and testing data differed; thus, to compare the RMSEs of the two data sets, NRMSE was used. NRMSE is the ratio of RMSE to the range of the data set; it ranges from 0 to 1. Eqs. 7 and 8 explain RMSE and NRMSE, respectively. MAE was calculated to describe the average magnitude of the errors.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{7}$$

$$NRMSE = \frac{RMSE}{y_{\max} - y_{\min}} \tag{8}$$

Where, $\hat{y}_i$ is the estimated value of the outcome variable of observation i, y_i is the observed value of the dependent variable observation i, $y_{\max}$ is the maximum observed value of the dependent variable, $y_{\min}$ is the minimum observed value of the dependent variable, and n is the sample size.

Bootstrapping with 10,000 replications on the training data set was carried out to calculate the 95% confidence interval of R^2 (95% CI). The coefficient of determination of the model run on the testing data were also calculated. The R^2 of the model run testing data was expected to fall in the 95% CI.

RESULTS AND DISCUSSION

Significant independent variables and selected models

The face-to-face interview questionnaire received a response from 286 out of 321 households, which was more than the statistically required sample size (281). Therefore, the nonresponse samples were removed in later analysis. Fig. 2 shows the estimates of the correlation coefficients of the variables, and indicates that the correlation coefficients of the outcome and explanatory variables are low (<0.33). In Fig. 3, the horizontal axis names options chosen by the BMA. Red indicates the predictor variables correlated with the outcome variable with a positive coefficient. Blue represents the negatively correlated variables, and the other colour shows that the variable is not present in the model. The result of BMA method indicates that the four independent variables, household location (X_{plc}), home business (X_{bus}), household size (X_{siz}), and house area per person (X_{pa}), proved to be significant for estimating daily per capita waste generation (Fig. 3). X_{plc} and X_{siz} were present in all groups of significant determinant variables selected by BMA (p=100), while the probability of X_{bus} and X_{pa} appearing in models chosen by the BMA is 96.4% and about 74%, respectively. Interestingly, household income, presence/absence of a garden, and percentage of members of the family of different age ranges were not significant, indicating that these factors do not explain variations in the waste generation rate.

X_{siz} is negatively correlated with daily per capita waste generation; an increase in the number of family members will lead to a decrease of daily per capita waste generation. Our finding that there is a qualitative relationship between household size and daily per capita waste generation agrees with those of previous studies; Benítez *et al.* (2008), Qu *et al.* (2009), and (Sukholthaman *et al.*, 2015) found the same negative influence of household size on waste generation rate per capita. The positive correlation between the regressor (X_{plc}) and the response variable indicated that the household waste generation rate is associated with the area the household is located in. People living in urban areas generated more waste than those living in rural areas. In contrast, Hockett *et al.* (1995) found that urbanisation was not a significant determinant of waste generation rate. In Viet Nam, homes commonly serve as bases for businesses such as convenience stores, restaurants, or a place for manufacturing goods. The presence of a business at home might affect the quantity of waste generated per capita (Parizeau *et al.*, 2006). X_{bus} was confirmed to be a significant determinant of household waste generation rate (Table 2). X_{bus} correlating positively with waste generation

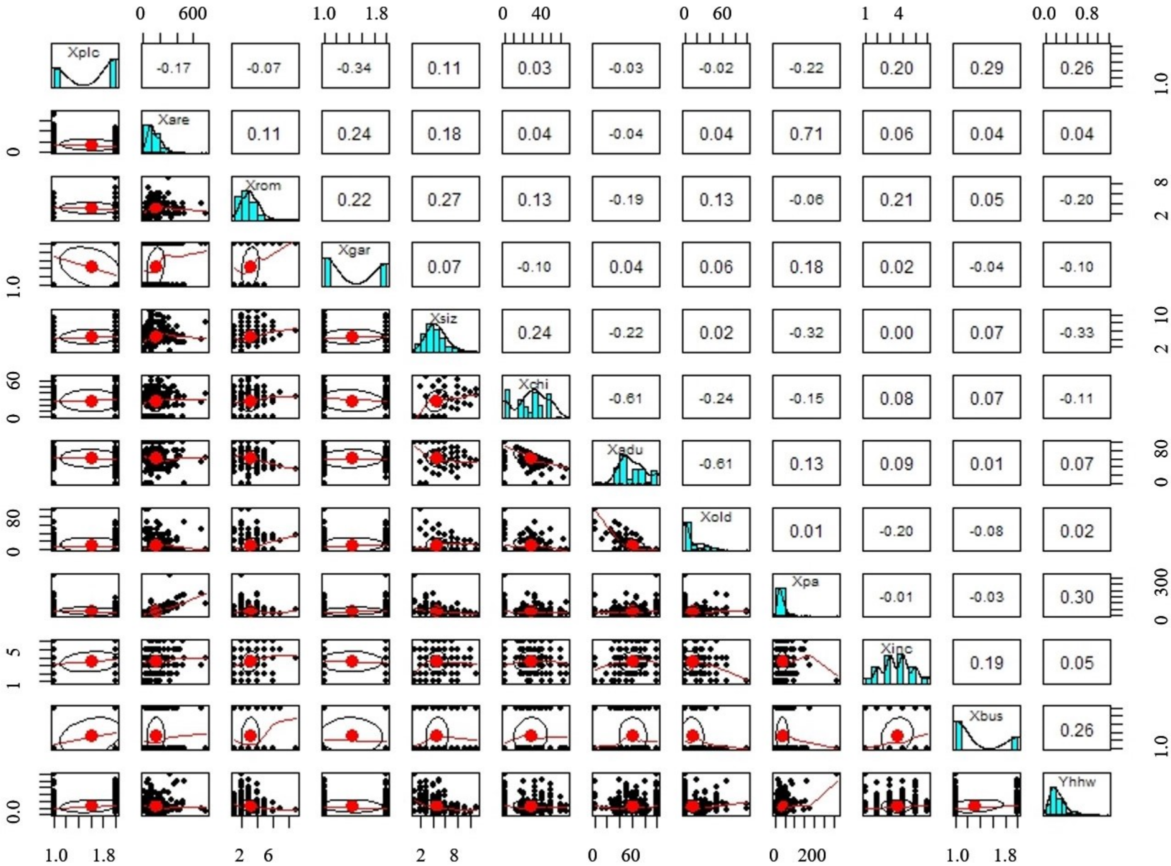

Fig. 2: Correlation coefficients of the variables

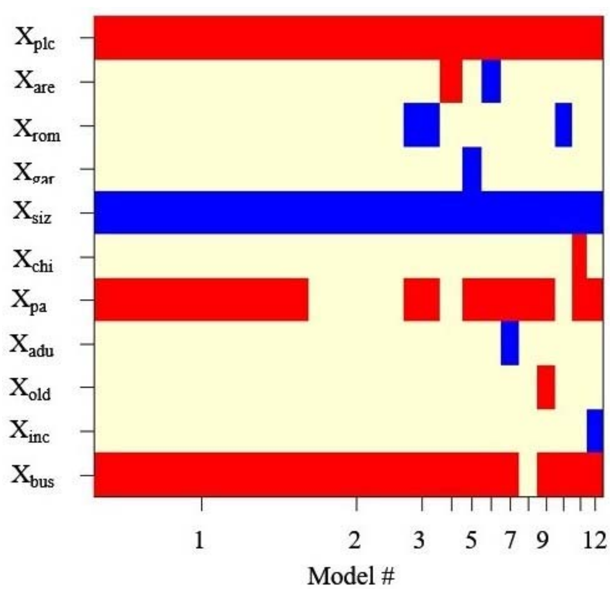

Fig. 3: Predictors chosen for the most reliable models by BMA

means that families running a home business have a higher per capita daily waste generation rate than those without a home business. Higher income leads to the consumption of more goods and therefore to the production of more waste (Buenrostro *et al.*, 2001). Nevertheless, other researchers found that the household income is not related to waste generation by measuring different types of income, such as continuous income (Bernache-Pérez *et al.*, 2001; Benítez *et al.*, 2008; Grazhdani, 2016), categorical income (Bolaane and Ali, 2004, Gomez *et al.*, 2008), or proxy variables (Mbande, 2003; Gomez *et al.*, 2008; Prades *et al.*, 2014). The result of this study indicates that direct income (X_{inc}) is not a significant determinant of waste generation. Investigating this relationship is complex because accurate income data are difficult to solicit from households, especially in developing countries (Parizeau *et al.,* 2006). In HAC, people might consider their income to be a private matter, and business households try to conceal their real income to avoid paying more taxes. A proxy variable of income, the total area of the house (X_{are}), was not significant in the estimation of per capita waste generation, while X_{pa} did prove to be a determinant of the same. This means that the amount of waste produced is correlated to the average space in the house per family member. The number of the rooms in the house was not an explanatory variable for waste generation estimation, according to the BMA. This result is inconsistent with a previous study in which the production of household waste was found to be positively correlated with the number of rooms (Monavari *et al.,* 2012). The presence or absence of a garden and the age ranges of household members were not significantly correlated with the quantity of waste produced. The BMA method not only detected the best model for predicting household waste generation, but also suggested other reliable models based on BIC approximation. Thus, we had different models with which to identify the amount of waste generated. Table 2 shows the five best waste generation prognosis models suggested by the BMA. Model 1, with four predictors, has the lowest BIC approximation (-62.3), which means that the linear regression model using four variables (X_{plc}, X_{bus}, X_{siz}, and X_{pa}) is the best multivariate model among all the possibilities.

Posterior probability represents the likelihood that a model will explain the observed data correctly. The posterior probabilities of Models 1 and 2 are higher than those of the other models, approximately 42% (0.422) and 19% (0.187), respectively. This indicates that Models 1 and 2 explain observations on waste generation more accurately than the other models, with posterior probabilities around 5%. Model 2, with three independent variables, has lower R^2_{adj} (about 30%), and models with four and five regressors have R^2_{adj} values of about 33%, which are similar. This means that adding more than four independent variables to the model will not improve its fit. Table 2 also shows the significance level of every variable in each model. Models 3, 4, and 5 each have more than four regressors, but not all the explanatory variables are significant. The variables X_{rom} (Model 3) and X_{gar} and X_{are} (Models 4 and 5) proved to be negligible determinants of waste generation. Thus, we choose

Table 2: Best models as selected by the Bayesian Model Average method

Independent variables	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	-1.539 (***)	-1.293 (***)	-1.402 (***)	-1.894 (***)	-1.517 (***)
X_{plc}(Urban)	0.578 (***)	0.535 (***)	0.560 (***)	0.457 (***)	0.557 (***)
X_{are}	-	-	-	0.0005 ( )	-
X_{rom}	-	-	-0.055 ( )	-	-
X_{gar}(YES)	-	-	-	-0.08 ( )	-0.061 ( )
X_{siz}	-0.128 (***)	-0.147 (***)	-0.120 (***)	-	-0.125 (***)
X_{pa}	0.004 (**)	-	0.004 (**)	-	0.004 (**)
X_{inc}	-	-	-	-	-
X_{bus}(YES)	0.302 (**)	0.317 (***)	0.308 (**)	0.292 (**)	0.304 (**)
Number of variables used	4	3	5	4	5
BIC	-62.3	-60.7	-58.8	-57.8	-57.5
Posterior probability	0.422	0.187	0.071	0.043	0.038
R^2	0.343	0.319	0.348	0.328	0.344
R^2_{adj}	0.329	0.309	0.332	0.318	0.327
F-statistic	25.43	30.64	20.76	10.61	20.38

Note: p ~ 0 (***); p < 0.001 (**); p < 0.05 (*); p < 0.1 (.); p < 1 ( )

Models 1 and 2, in which each predictor variable is significant, BIC approximations are small, and posterior probabilities are high, to test the assumptions and predictive performance of the linear model.

Multivariate linear regression models for household waste generation

Eqs. 9 and 10 show the parameter estimates for the selected Models 1 and 2, respectively:

$$\text{Log}\,(Y_{HHW}) = -1.539 + 0.578X_{plc}(\text{Urban}) - 0.128X_{siz} + 0.004X_{pa} + 0.302X_{bus}(\text{YES}) \quad (9)$$

$$\text{Log}\,(Y_{HHW}) = -1.293 + 0.535X_{plc}(\text{Urban}) - 0.147X_{siz} + 0.317X_{bus}(\text{YES}) \quad (10)$$

The intercept -1.539 (in Eq. 9) is the unconditional expected mean of the logarithm of the waste generation rate. Therefore, 0.215 (kg/capita/day), which is the exponential value of the intercept, is the geometric mean of the waste generation rate. The exponential value of the coefficient for X_{plc} is 1.78 ($e^{0.578}$=1.78), indicating that average per capita waste generation of households in urban areas is 78% higher than that of households in rural areas when other independent variables are held constant. Similarly, a person in a family running a home business generated 35% more waste than one living in a family that does not ($e^{0.578}$=1.35). Household size is the only significant predictor variable negatively correlated with waste generation, indicating that an increase in the number of family members is associated with a decrease in per capita waste generation. The coefficient of X_{siz} in the model is -0.128, meaning that an increase of one person in a family leads to a 12% decrease in waste

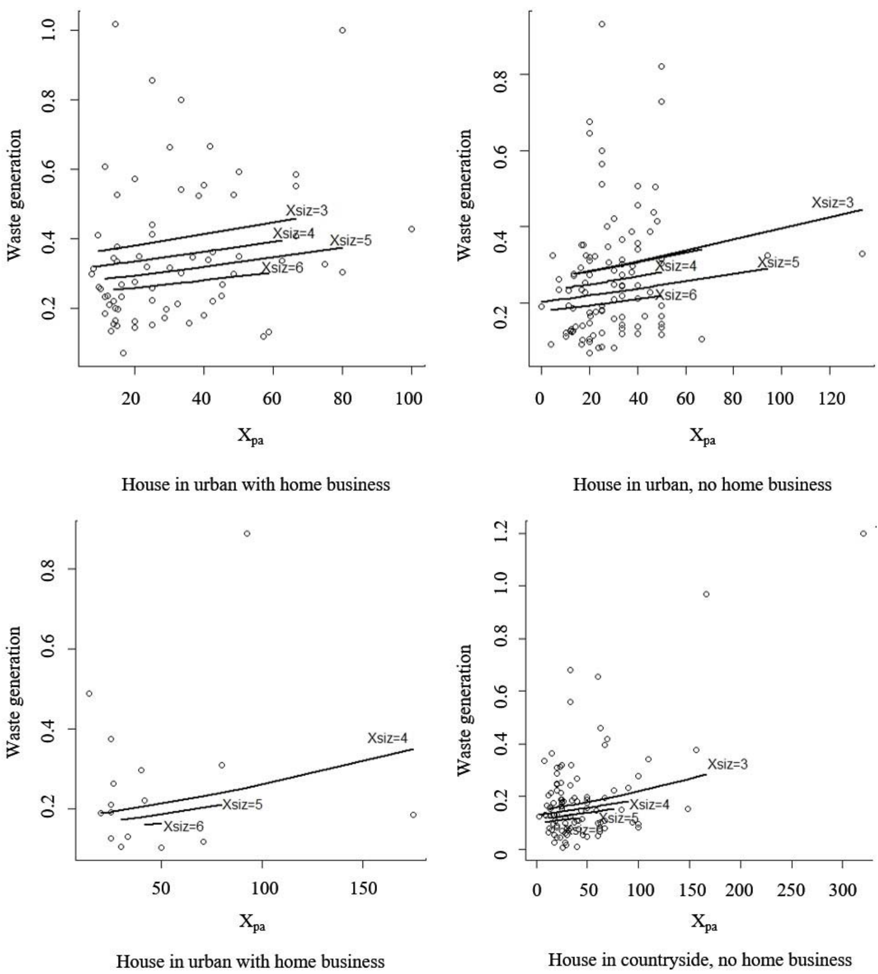

Fig. 4: Fitted line plot of a model with four regressors: Xplc, Xsiz, Xpa, and Xbus

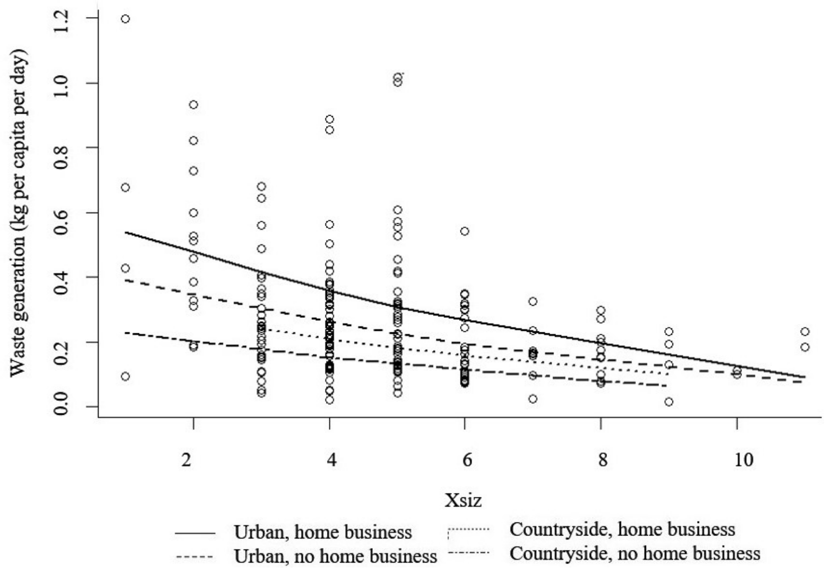

Fig. 5: Fitted line plot of model with three regressors Xplc, Xsiz, and Xbus

generation rate ($e^{-0.128}$=0.88) when other variables are held constant. The parameter explanation is similar to that of Model 2. Fig. 4 describes the predicted value of per capita waste generation from Model 1, which has four predictor variables. The different lines in the plot represent the estimated waste generation rate for various values of X_{siz}, and four values of X_{siz} (X_{siz}=3, 4, 5, and 6) are shown. Fig. 5 also describes the predicted waste generation rates based on Model 2, which has three independent variables.

The purpose of this study was to find a simple and reliable model to estimate waste generation, in order to contribute to improving waste management. Models can provide reliable information to support current waste collection and transportation methods. Exact estimation of waste generation in rural and urban areas results in better design and arrangement of vehicles, labour, and collection routes. This in turn will improve the current collection system, which has so far been inefficient owing to poor calculation and design. Moreover, results from the model estimation show that a decentralised management approach could benefit waste collection, as the waste generation rate varies in urban and rural areas. On-site or small-scale treatments might reduce the cost of collecting the low amount of waste generated in faraway areas, especially since biodegradable waste has great potential be composted at home or recycled into feed for animals in agricultural areas.

The result of the study also suggested that home businesses contributed considerably to the total waste required for collection from a household. Therefore, estimates of waste generated from households running home businesses could provide a basis for the decision-makers to improve the waste management system, by assigning more importance to the commercial and tourist sectors in the city. For instance, increasing the waste collection fee for the home and business sectors can improve waste management, since the number of households involved in business activities are increasing quickly.

Analysis of models

Diagnostic test for linear model assumption

Fig. 6 presents the results of tests for normality, constant variance, and autocorrelation. Two quantile-quantile plots (Figs. 6.1 and 6.7) compare the residuals (the points on the graph) to 'ideal' normal observations (the line). The residuals follow the line approximately, indicating that the errors of both models are normal. The plot of the residuals versus fitted values (Figs. 6.2 and 6.8) are used for a test of non-constant variance. The scatter is symmetric vertically around zero, demonstrating that there is no evidence of non-constant variance. Moreover, Figs. 6.2, 6.3, 6.4, 6.5, and 6.6 (Model 1) and Figs.

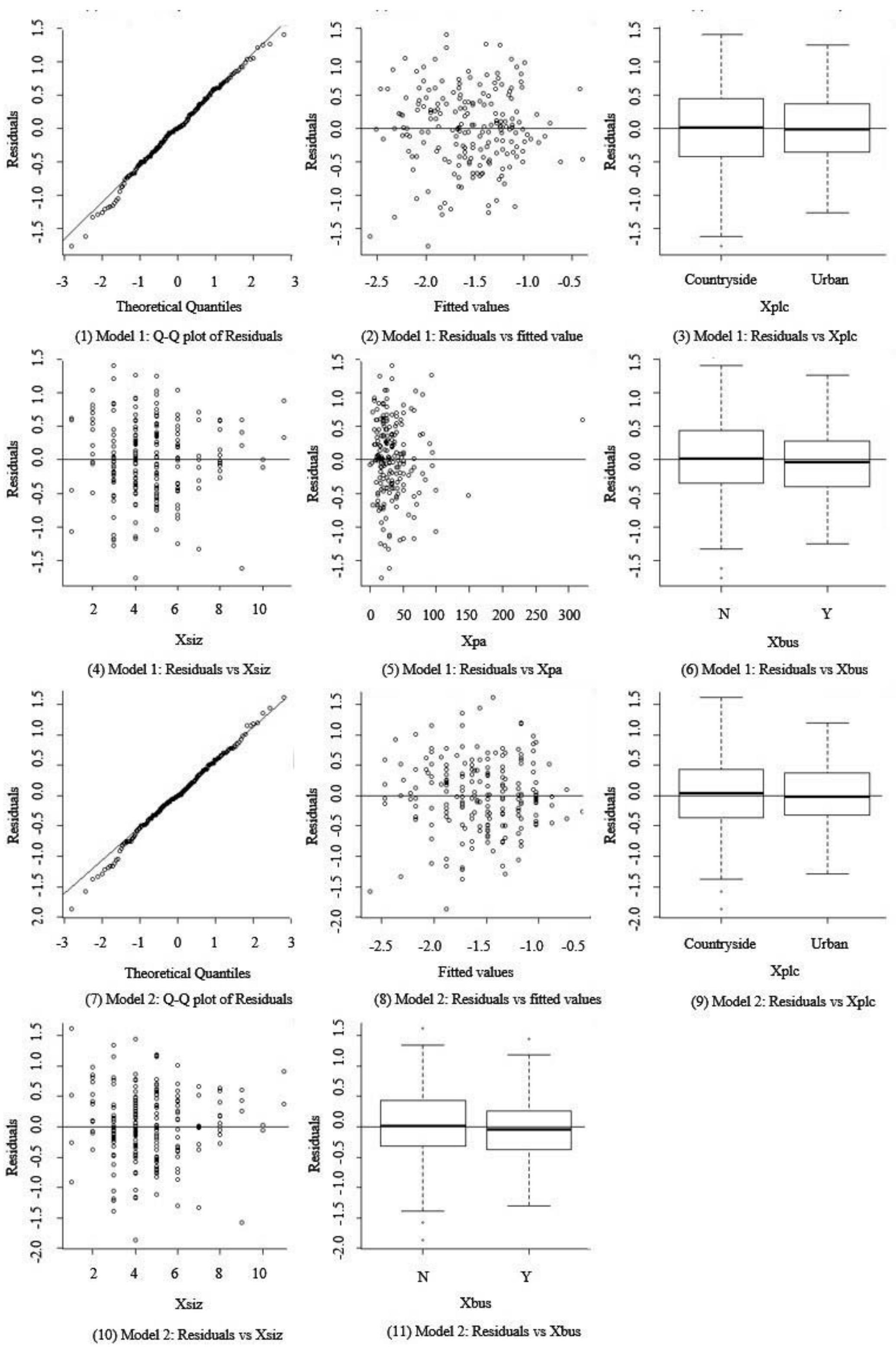

Fig. 6: Tests for the linear assumption of Models 1 and 2

6.9, 6.10, and 6.11 (Model 2) show that the scatter of residuals is symmetric approximately around zero in a plot with all independent variables. This means that there is no problem with the correlation of the residuals. The results from the graphical test indicate that all linear assumptions were satisfied.

Influence of variables and observations

Fig. 7 shows the relative importance (with 95% confidence interval) of the regressors for the two models, as determined by the Local Matching Gabor method (Johnson and LeBreton; 2004, Grömping, 2006). Household location (X_{plc}) acts as the primary predictor variable in both models because its percentage of contribution is about 40%, followed by household size (X_{siz}) at around 30%. The independent variables that contribute less to waste generation are household area per person (X_{pa}) and home business (X_{bus}), with values of 10% and 20%, respectively.

Fig. 8 indicates that the observations have a large impact on the predicted values, measured by Cook's distance (Cook, 1977, Cook, 1979). Observations numbered 13, 70, and 184 significantly influence the fit of the models compared to the other observations, but none of them has too much influence (Cook's distance less than 1.0) (David, 2007). Outlier tests show that observation 47 was an outlier in both models.

Model validation

The R^2 values of the models are 0.34 (model 1) and 0.32 (model 2), meaning that they explain about 34% and 32%, respectively, of the variation in daily

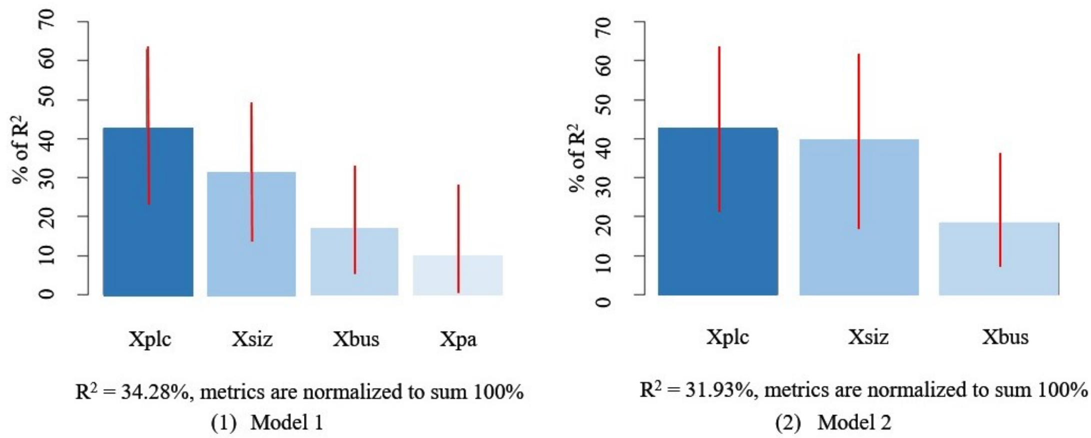

Fig. 7: Relative importance of regressors for waste generation rate

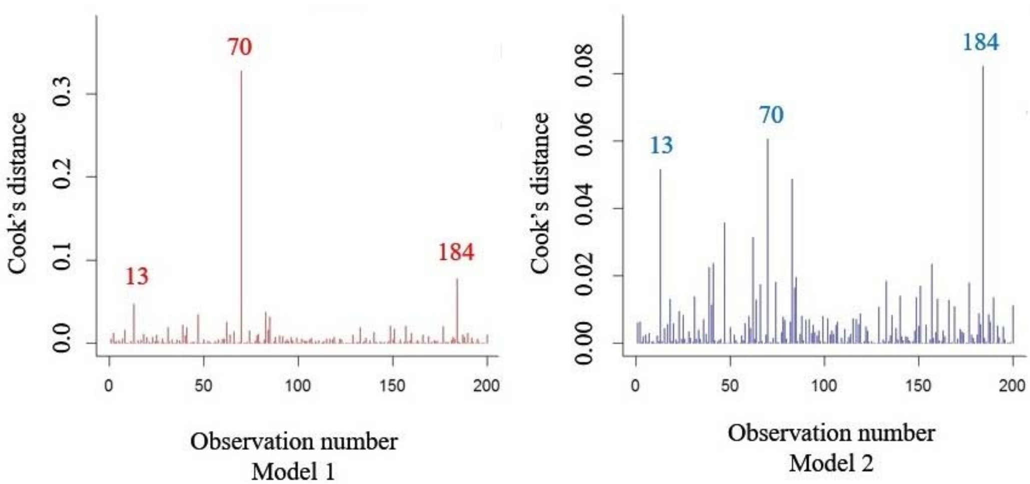

Fig. 8: Influence of observations on Models 1 and 2. Observations with a particularly high level of influence are numbered in red and blue.

Fig. 6: Tests for the linear assumption of Models 1 and 2

Model	Datasets	Standard deviation	R^2	Mean R^2 (bootstrap)	95% CI of R^2	RMSE	NRMSE	MAE
Model 1	Train set	0.712	0.343	0.354	0.235 – 0.428	0.576	0.131	0.451
	Test set		0.281			0.678	0.137	0.488
Model 2	Train set	0.712	0.319	0.327	0.220 – 0.406	0.586	0.134	0.453
	Test set		0.256			0.689	0.139	0.492

per capita waste generation rate. Other multivariate linear regression studies had low R^2 values; 51% in Benítez *et al.* (2008), 36% in Grossman *et al.* (1974), and 48.7% in the study by Bach *et al.* (2004). The weak coefficient of determination could be explained by the fact that waste generation studies, which attempt to predict human behaviours such as habit and lifestyle, normally have R^2 values lower than 50%. Human behaviours are simply harder to predict than physical processes. However, the goal is not to maximise the coefficient of determination, because obtaining more predictor variables may cause over fitting. In other words, a low R^2 value does not mean the model is useless, and a significant R^2 value cannot indicate that the model is useful (Brown and Berthouex, 2002). A good model can also maximise the percentage of variations explained but limit the ability of the results to be generalised (Beigl *et al.*, 2008). If a model satisfies all the assumptions of the linear regression, it is the correct one to use to estimate waste generation. Moreover, it can help one draw meaningful conclusions about how changes in the predictor variables are associated with variations in the response variables.

The multivariate linear regression model created by the training data set was run on the testing data set and the statistical performance metrics were calculated. Table 3 explains the model validation results. In both models, the R^2 values of the testing data set are in the 95% CI, and the NRMSEs and MAEs of both data sets are very close, which means the two models perform well with the new data. On the other hand, the RMSEs are smaller than the standard deviation of the response variable (0.712), indicating that the models produce less variation than the observations. Lastly, low NRMSEs (about 0.13 out of a possible range of 0 to 1) demonstrate that the fitted values are quite close to the observations. Thus, both models show good performance in predicting household per capita waste generation. Model 1 performed better than Model 2 because it has a higher R^2 value, a higher posterior probability, and smaller errors (RMSE and MAE).

CONCLUSION

The models constructed in the current study are valuable in estimating waste generation, as they provide observational evidence of the influence of multiple factors. They indicate that the impacts of socio-demographic and geographic variables and family economic activities are highly significant for waste generation rates. The models cannot predict waste generation in the future, but they can provide reliable information needed to improve current waste management systems. Household location is the predictor that most affects the daily waste generated per capita. Both models showed that a person in an urban household produced much more solid waste (70–80%) than one in a rural household. This information provides an exact estimate of waste generation in rural and urban areas, and can be used to improve calculation and arrangement of vehicles, labour, and collection routes. It also suggests that a decentralised treatment approach could reduce the collection cost for the low amount of waste generated in areas located further away from landfills. However, the implications of this need to be studied carefully, and appropriate legislation would need to be passed to encourage decentralised waste treatment. Education and awareness on waste generation were successfully carried out in HAC, and these should be maintained and improved. Another important factor influencing household waste generation in HAC is family economic activity. The two models showed that if a family runs a business from home, their household waste generation rate will increase by about 35%. Waste fees for the business sector might be an important factor to consider in waste collection and management planning, since the number of households running businesses, such as small restaurants, homestays, shops, convenience stores, and vehicle rentals to provide services for tourists and locals, have been rising gradually. This study found that household size and area of the house are also significant determinants of per capita waste generation, while other variables, particularly income per person, proved not to be significant as correlates of waste production. The results from this study demonstrate that the Bayesian Model Average (BMA) method is a robust one to determine firm options for multiple linear regression models, especially those dealing with a large number of independent variables. The result of the BMA method indicated that a linear regression model with four independent variables (X_{plc}, X_{siz}, X_{pa}, and X_{bus}) was the best model to estimate waste generation in HAC because it had the lowest BIC approximation and was of a reasonable size. The model with three regressors (X_{plc}, X_{siz}, and X_{bus}) had a slightly lower performance, but is still very useful to quickly predict household waste generation because information on the predictor variables is available in the census database. This study attempted to increase the understanding of waste generation to support waste management planning for the city; thus, the two models are useful not only for analysing the key factors influencing waste generation but also for providing waste managers a way to estimate waste generation volumes in order to improve waste reduction and management efforts. Thus, the results and methodology are expected to be informative for authorities, decision makers, stakeholders, and planners to develop waste management plans. Note, however, that the model developed in this paper is not reliable for predicting future waste generation; a lack of historical data caused difficulties in the development of a predictive waste model. Thus, future studies should concentrate on devising a municipal waste generation model that can forecast future waste volumes.

ACKNOWLEDGMENTS

The authors acknowledge with gratitude the efforts of students in the survey team and field assistants from Hoi An public work LTD. Co. They are also grateful to the Hoi An solid waste treatment plant as well as the College of Technology, the University of Danang, for the use of space and facilities. The financial support of a Research Grant for Encouragement of Students of Okayama University is greatly acknowledged.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

ABBREVIATIONS

BMA	Bayesian model average
BIC	Bayesian information criterion
EU	European Union
Eq.	Equation
GDP	Gross domestic product
HAC	Hoi An city
MAE	Mean absolute error
m^2	Cubic meter
MSW	Municipal solid waste
NRMSE	Normalized root mean square error
MSWM	Municipal solid waste management
OECD	Organization for Economic Co-operation and Development
P	Probability value
%	Percentage
R^2	Coefficient of determination
RMSE	Root mean square error
USA	United States of America

REFERENCES

Abbasi, M.; Abduli, M.; Omidvar, B.; Baghvand, A., (2013). Forecasting municipal solid waste generation by hybrid support vector machine and partial least square model. Int. J. Environ. Res., 7(1): 27-38 (**12 pages**).

Abdoli, M.A.; Rezaei, M.; Hasanian, H., (2016). Integrated solid waste management in megacities. Global J. Environ. Sci. Manage., 2(3): 289-298 (**10 pages**).

Abu Qdais, H.A.; Hamoda, M.F.; Newham, J., (1997). Analysis of Residential Solid Waste At Generation Sites. Waste. Manage. Res., 15(4): 395-406 (**11 pages**).

Akhtar, S.; Saleem, W.; Nadeem, V.M.; Shahid, I.; Ikram, A., (2017). Assessment of willingness to pay for improved air quality using contingent valuation method. Global J. Environ. Sci. Manage., 3(3): 279-286 (**17 pages**).

Bach, H.; Mild, A.; Natter, M.; Weber, A., (2004). Combining socio-demographic and logistic factors to explain the generation and collection of waste paper. Resour. Conserv. Recyc., 41(1): 65–73 (**9 pages**).

Bdour, A.; Altrabsheh, B.; Hadadin, N.; Al-Shareif, M., (2007). Assessment of medical wastes management practice: a case study of the northern part of Jordan. Waste. Manage., 27(6): 746–759 (**14 pages**).

Beigl, P.; Lebersorger, S.; Salhofer, S., (2008). Modelling municipal solid waste generation: A review. Waste Manage, 28(1): 200-214 (**15 pages**).

Benítez, S.O.; Lozano-Olvera, G.; Morelos, R.A.; Vega, C.A.d., (2008). Mathematical modeling to predict residential solid waste generation. Waste Manage., 28, Supplement 1: S7-S13 (**7 pages**).

Bernache-Pérez, G.; Sánchez-Colón, S.; Garmendia, A.M.; Dávila-Villarreal, A.; Sánchez-Salazar, M.E., (2001). Solid waste characterisation study in the Guadalajara Metropolitan Zone, Mexico. Waste Manage. Res., 19(5): 413– 424 (**15 pages**).

Bleeker, S.E.; Moll, H.A.; Steyerberg, E.W.; Donders, A.R.T.; Derksen-Lubsen, G.; Grobbee, D.E.; Moons, K.G.M., (2003). External validation is necessary in prediction research:: A clinical example. J. Clin Epidemiol., 56(9): 826-832 (**7 pages**).

Bolaane, B.; Ali, M., (2004). Sampling Household Waste at Source: Lessons Learnt in Gaborone. Waste Manage. Res., 22(3): 142–148 (**7 pages**).

Boulet, S.; Boudot, E.; Houel, N., (2016). Relationships between each part of the spinal curves and upright posture using Multiple stepwise linear regression analysis. J. Biomec., 49(7): (**7 pages**).

Brown, L.C.; Mac Berthouex, P., (2002). Statistics for environmental engineers. CRC press.

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 (**8 pages**).

Chai, T.; Draxler, R.R., (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geosci. Model Dev., 7(3): 1247-1250 (**14 pages**).

Chang, Y.F.; Lin, C.J.; Chyan, J.M.; Chen, I.M.; Chang, J.E., (2007). Multiple regression models for the lower heating value of municipal solid waste in Taiwan. J. Environ. Manage., 85(4): 891-899 (**9 pages**).

Chen, H.W.; Chang, N.-B., (2000). Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling. Resour. Conserv. Recyc., 29(1-2): 1-18 (**18 pages**).

Chu, M.T., (2014). Báo cáo tổng kết đề tài: Xây dựng mô hình đồng quản lý rác thải tại hai xã, phường Cẩm Hà và Cẩm Phô, thành phố Hội An (Project report: building co-operative waste management model in two wards, Cam Ha and Cam Pho, Hoi An city). Management board of Cham islands MPA, Hoi An, Viet Nam.

Cook, R.D., (1977). Detection of Influential Observation in Linear Regression. Technometrics, 19(1): 15-18 (**4 pages**).

Cook, R.D., (1979). Influential Observations in Linear Regression. J. Am. Stat. Assoc., 74(365): 169-174 (**6 pages**).

Daskalopoulos, E.; Badr, O.; Probert, S.D., (1998). Municipal solid waste: a prediction methodology for the generation rate and composition in the European Union countries and the United States of America. Res. Conserv Recyc., 24(2): 155-166 (**12 pages**).

David, M.L., (2007). Influential Observations, Online Statistics Education: A Multimedia Course of Study, Rice University Publication.

Dennison, G.J.; Dodd, V.A.; Whelan, B., (1996). A socio-economic based survey of household waste characteristics in the city of Dublin, Ireland — II. Waste quantities. Resour. Conserv. Recyc., 17(3): 245-257 (**13 pages**).

Derksen, S.; Keselman, H.J., (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. Brit. J. Math. Stat. Psychol., 45(2): 265–282 (**18 pages**).

Faber, N.M.; Rajkó, R., (2007). How to avoid over-fitting in multivariate calibration—The conventional validation approach and an alternative. Anal. Chim. Acta., 595(1–2): 98-106 (**9 pages**).

Fernández, C.; Ley, E.; Steel, M.F.J., (2001). Model uncertainty in cross-country growth regressions. J. Appl. Econ., 16(5): 563-576 (**14 pages**).

Ghinea, C.; Drăgoi, E.N.; Comăniţă, E.-D.; Gavrilescu, M.; Câmpean, T.; Curteanu, S.; Gavrilescu, M., (2016). Forecasting municipal solid waste generation using prognostic tools and regression analysis. J. Environ. Manag., 182: 80-93 (**14 pages**).

Gomez, G.; Meneses, M.; Ballinas, L.; Castells, F., (2008). Characterization of urban solid waste in Chihuahua, Mexico. Waste Manage., 28(12): 2465-2471 (**7 pages**).

Grömping, U., (2006). Relative importance for linear regression in R: the package relaimpo. J. Stat. Software, 17(1): 1-27 (**27 pages**).

Grazhdani, D., (2016). Assessing the variables affecting on the rate of solid waste generation and recycling: An empirical analysis in Prespa Park. Waste Manage., 48: 3-13 (**11 pages**).

Grossman, D.; Hudson, J.; Marks, D., (1974). Waste generation models for solid waste collection. J. San. Eng. Division, 100(6): 1219-1230 (**12 pages**).

Hamby, D.M., (1994). A review of techniques for parameter sensitivity analysis of environmental models. Environ. Monit. Assessment, 32(2): 135–154 (**10 pages**).

HASD, 2013. Statistical Yearbook - Hoi An city, Hoi An Statistical Department, Viet Nam.

Hoang, M.G.; Fujiwara, T.; Pham Phu, S.T., (2017). Municipal waste generation and composition in a tourist city - hoi an, vietnam. J. SCE, 5(1): 123-132 (**10 pages**).

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 (**13 pages**).

Hoeting, J.A.; Madigan, D.; Raftery, A.E.; Volinsky, C.T., (1999). Bayesian model averaging: A Tutorial. Stat. Sci., 14(4): 382-401 (**20 pages**).

Johnson, J.W.; LeBreton, J.M., (2004). History and use of relative importance indices in organizational research. Organ. Res. Method., 7(3): 238-257 (**20 pages**).

Karpušenkaitė, A.; Denafas, G.; Ruzgas, T., (2016). Forecasting hazardous waste generation using short data sets: Case study of Lithuania. Science–future of Lithuania/Mokslas–lietuvos ateitis, 8(4): 357–364 (**8 pages**).

Kolekar, K.A.; Hazra, T.; Chakrabarty, S.N., (2016). A review on prediction of municipal solid waste generation models. Procedia Environ. Sci., 35: 238-244 (**7 pages**).

Lebersorger, S.; Beigl, P., (2011). Municipal solid waste generation in municipalities: quantifying impacts of household structure, commercial waste and domestic fuel. Waste Manage., 31(9-10): 1907-1915 (**9 pages**).

Lebersorger, S.; Schneider, F.; Hauer, W., (2003). Waste generation in households–models in theory and practical experience from a case study of multifamily dwellings in Vienna. Proceedings in Sardinia, (**4 pages**).

Madigan, D.; Raftery, A.E., (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. J. Am. Stat. Assoc., 89(428): 1535-1546 (**12 pages**).

Mbande, C., (2003). Appropriate approach in measuring waste generation, composition and density in developing areas. J. S. Afr. Inst. Civil Eng., 45(3): 2-10 (**9 pages**).

Memarianfard, M.; Hatami, A.M.; Memarianfard, M., (2017). Artificial neural network forecast application for fine particulate matter concentration using meteorological data. Global J. Environ. Sci. Manage., 3(3): 333-340 (**8 pages**).

MOC, (2010). National Technical Regulation Chapter 9: Solid waste collection, separation, transportation, treatment system and public toilet, Hanoi, Vietnam, QCVN 07:2010/BXD

MOC, (2016). National Technical Regulation, Chapter 9: Technical Infrastructure Works, Solid Waste Treatment and Public Toilet, QCVN 07:2016/BXD.

Monavari, S.; Omrani, G.; Karbassi, A.; Raof, F., (2012). The effects of socioeconomic parameters on household solid-waste generation and composition in developing countries (a case study: Ahvaz, Iran). Environ. Monit. Asses., 184(4): 1841–1846 (**6 pages**).

Ngoc, U.N.; Schnitzer, H., (2009). Sustainable solutions for solid waste management in Southeast Asian countries. Waste Manage, 29(6): 1982-1995 (**14 pages**).

Nguyen, D.L.; Hoang, M.G.; Bui, X.T., (2013). Challenges for municipal solid waste management practices in Vietnam. Waste Technol., 1(1): 17-21 (**5 pages**).

Noori, R.; Karbassi, A.; Salman Sabahi, M., (2010). Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste prediction. J. Environ. Manage., 91(3): 767-771 **(5 pages).**

Oumarou, M.B.; Dauda, M.T.; Abdulrahim, A.; Abubakar, A.B., (2012). Municipal solid waste generation, recovery and recycling: A case study. World J. Eng. Pure Appl. Sci., 2(5): 143-147 (**11 pages**).

Parizeau, K.; Maclaren, V.; Chanthy, L., (2006). Waste characterization as an element of waste management planning: Lessons learned from a study in Siem Reap, Cambodia. Resour. Conserv. Recyc., 49(2): 110-128 (**19 pages**).

Prades, M.; Gallardo, A.; Ibàñez, M.V., (2014). Factors determining waste generation in Spanish towns and cities. Environ. Monit. Assess., 187(1): 4098-4105 (**8 pages**).

Qu, X.-Y.; Li, Z.-s.; Xie, X.-y.; Sui, Y.-m.; Yang, L.; Chen, Y., (2009). Survey of composition and generation rate of household wastes in Beijing, China. Waste Manage., 29(10): 2618-2624 (**7 pages**).

Raftery, A.E.; Madigan, D.; Hoeting, J.A., (1997). Bayesian model averaging for linear regression models. J. Am. Stat. Assoc., 92(437): 179-191 (**3 pages**).

Shamshiry, E.; Bin Mokhtar, M.; Abdulai, A., (2014). Comparison of artificial neural network (ANN) and multiple regression analysis for predicting the amount of solid waste generation in a tourist and tropical area—Langkawi Island. Pproceeding of International Conference on Biological, Civil, Environmental Engineering (BCEE): 161-166 (**6 pages**).

Sukholthaman, P.; Chanvarasuth, P.; Sharp, A., (2015). Analysis of waste generation variables and people's attitudes towards waste management system: a case of Bangkok, Thailand. J. Mater. Cycle. Manage.: 645–656 (**12 pages**).

Thøgersen, J., (1996). Wasteful food consumption: Trends in food and packaging waste. Scand. J. Mgmt., 12(3): 291-304 (**15 pages**).

Thanh, N.P.; Matsui, Y., (2011). Municipal solid waste management in Vietnam: Status and the strategic actions. Int. J. Environ. Res., 5(2): 285-296 (**12 pages**).

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 (**5 pages**).

Thompson, B., (1995). Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial. Educ. Psychol. Meas., 55(4):525-534 **(10 pages).**

van de Klundert, A.; Anschütz, J.; Scheinberg, A., (2001). Integrated sustainable waste management: the concept. Tools for decision-makers. experiences from the urban waste expertise programme (1995-2001). Waste.

Vesely, S.; Klöckner, C.A.; Dohnal, M., (2016). Predicting recycling behaviour: Comparison of a linear regression model and a fuzzy logic model. Waste Manage., 49: 530-536 (**7 pages**).

Wang, M.C.; Bushman, B.J., (1998). Using the normal quantile plot to explore meta-analytic data sets. Psychol. Method., 3(1): 46-54 (**9 pages**).

Wilk, M.B.; Gnanadesikan, R., (1968). Probability plotting methods for the analysis for the analysis of data. Biometrika, 55(1): 1-17 (**17 pages**).

Zurbrugg, C.; Gfrerer, M.; Ashadi, H.; Brenner, W.; Kuper, D., (2012). Determinants of sustainability in solid waste management--the Gianyar Waste Recovery Project in Indonesia. Waste Manage., 32(11): 2126-2133 (**8 pages**).

AUTHOR (S) BIOSKETCHES

Hoang, M.G., Ph.D. Candidate, Instructor, Okayama University, Graduate school of Environmental and Life Science, Department of Environmental Science 3-1-1 Tsushima, Kita, Japan. Email: *gianghm@nuce.edu.vn*

Fujiwara, T., Ph.D., Professor, Waste Management Research Center Okayama University, 3-1-1 Tsushima, Kita, Okayama 700-8530, Japan. Email: *takeshi@cc.okayama-u.ac.jp*

Pham Phu, S.T., Ph.D. Candidate, Instructor, Okayama University, Graduate school of Environmental and Life Science, Department of Environmental Science 3-1-1 Tsushima, Kita, Japan. Email: *ppstoan@gmail.com*

Nguyen Thi, K.T., Ph.D., Professor, Faculty of Environmental Envineering National University of Civil Engineering, 55 Giai Phong Road, Hai Ba Trung, Ha Noi, Viet Nam. Email: *thaicanh1949@gmail.com*

HOW TO CITE THIS ARTICLE

Hoang, M.G.; Fujiwara, T.; Pham Phu, S.T.; Nguyen Thi, K.T., (2017). Predicting waste generation using Bayesian model averaging. Global. J. Environ. Sci. Manage., 3(4): 385-402.

DOI: 10.22034/gjesm.2017.03.04.005

url: http://gjesm.net/article_26611.html