**ORIGINAL ARTICLE**

CrossMark

# Influence of population, income and electricity consumption on per capita municipal solid waste generation in São Paulo State, Brazil

Reinaldo Pisani Jr.[1] · Marcus César Avezum Alves de Castro[2] · Antonio Alvares da Costa[1]

## Abstract

Predicting municipal solid waste (MSW) generation is fundamental in choosing and scaling the processes involved in municipal management. The challenge for financial sustainability is to create indicators that enable MSW fees to be charged in proportion to the amount generated by each resident. Mathematical functions were tested to adjust the per capita waste generation rate (PCWG) in the municipalities of the state of São Paulo, based on population ($P$), per capita income (PCI) and per capita energy consumption (PCE). The dataset involved 238 municipalities in 2013 and 251 municipalities in 2014 that routinely weighed their wastes. The averaged PCWG increased from 0.65 to 0.90 kg inh.$^{-1}$ day$^{-1}$ (increment of 38%) when population enhanced from the range of 0–25,000 to 100,001–500,000 inh., mean per capita income grew from 10.1 to 13.6 USD inh.$^{-1}$ day$^{-1}$, and mean per capita electricity consumption expanded from 6.9 to 10.9 kWh inh.$^{-1}$ day$^{-1}$. The equation that best represented the data set resulted in r of 0.49, $R^2$ of 0.24, RMSE of 0.224 kg inh.$^{-1}$ day$^{-1}$ and $E_p$ of − 12.3%. Despite the relatively low $R^2$, it was demonstrated by Student's $t$ test that the proposed equation was able to represent mean values and result in the same variance with more than 99% probability.

## Introduction

Production processes and human consumption generate a great diversity and quantity of solid wastes that require environmentally adequate solid waste management. Municipal solid waste (MSW) represents a significant amount of solid waste that, because of health and environmental risks, needs integrated waste management in order to maintain environmental quality.

Predicting MSW generation rates is a basic requirement in choosing and planning the operations and processes involved in the waste management chain at the municipal level. Moreover, such forecasts are needed to estimate total MSW masses or volumes in order to specify the lifespan of the system or equipment employed in waste management. The per capita generation rate is an indicator widely used to estimate MSW production, because it represents the daily mass of waste produced per person in a given locality.

According to Article 7 of Brazilian Law No. 12305 [1], which establishes the National Solid Waste Policy, whose main target is "regularity, continuity, functionality and universalization of public space cleaning services and solid waste management, accompanied by managerial and economic mechanisms that ensure the recovery of the costs of services rendered, as a way of ensuring their operational and financial sustainability". In this regard, correlating the per capita MSW generation rate with the population's energy consumption and income, which are variables with time series available in the state of São Paulo, allows one to determine the individualized fee for the services proportional to the amount of waste generated by each citizen.

Municipal solid waste (MSW) is composed of household or residential wastes and of wastes resulting from urban

✉ Reinaldo Pisani Jr.
pisanijr@terra.com.br

1 Postgraduate Program in Environmental Technology, University of Ribeirão Preto, Avenida Costabile Romano, 2201, Ribeirânia, Ribeirão Preto, SP 14096-900, Brazil

2 Department of Applied Geology, Institute of Geosciences and Exact Sciences, Julio de Mesquita Filho University, Rio Claro Campus, Avenida 24-A, 1515, Bela Vista, Rio Claro, SP 13506-900, Brazil

cleaning services such as sweeping or cleaning of public spaces and thoroughfares (Fig. 1).

The per capita municipal solid waste generation rate is provenly dependent on demographic and socioeconomic conditions and on living standards such as income, gross national product (GNP), urban population, proportion of urban population, per capita consumption expenditure of urban households, and total energy consumption, although the influence of each parameter is not entirely clear [3–9].

At different locations in Turkey, Keser et al. [7] considered socioeconomic and climatic factors to determine the per capita MSW generation rate. In this research, spatial autoregression models and geographically weighted regression models were employed simultaneously to analyze spatial data, which showed a Pearson coefficient ($r$) ranging from 0.47 to 0.59, but the magnitude of the root-mean-square error (RMSE) was found to be 50% relative to the predicted values.

In the state of North Carolina, USA, Hockett et al. [10] correlated the per capita MSW generation rate as a function of ten variables involving socioeconomic aspects and urban population size of each county. The determination coefficient ($R^2$) reportedly fell within the range of 0.37–0.50.

Xu et al. [11] developed a hybrid model that combined the seasonal autologistic regression and a moving average to predict the generation of MSW in a specific city (in China) on multiple time scales without the need to consider other variables. Their results indicated that the model was very accurate, but was limited to the city under study.

For Mexico City, Benítez et al. [12] proposed models to relate the per capita household waste generation rate (dependent variable) to different (independent) variables such as education, household income and number of inhabitants. In order to analyze the model that best explained the prediction, the data set was analyzed based on normality, Kolmogorov–Smirnov, multicollinearity and heteroscedasticity tests. The maximum value of the determination coefficient ($R^2$) was 0.51.

Navarro-Esbrí et al. [13] employed models to predict the amounts of municipal solid waste generated in three Greek and Spanish cities. The authors compared two different techniques. The first technique was a seasonal AutoRegressive and Moving Average (sARIMA) model. The second model involved a non-linear systems analysis in which the MSW generation was assumed to be a discrete dynamical system whose dynamic behavior was extracted from a measured scalar variable. Despite deviations of up to 5%, equations that could be used in forecasting were not shown.

In Brazil, the relationship between demographic and socioeconomic factors has been studied mainly at the municipal level [14–16]. For example, it was demonstrated that, in the city of Belo Horizonte (state of Minas Gerais), population, urban life quality index and per capita income were the main parameters to correlate the MSW generation [17–19]. However, the equations, when described, are directly applicable only to this particular municipality.

The studies cited have had local or regional coverage, used many variables that are often not available for a wider region, which were associated with reasonable quality of adjustments, or results that cannot be extrapolated to other locations. Besides, the data fit had been frequently measured by the root-mean-square error and the determination coefficient. Based on these aspects, the objective of this study was to develop feasible equations that enable forecasting the per capita MSW generation rates (PCWG) in municipalities of the state of São Paulo.



**Fig. 1** Solid waste classification by source, based on Federal Law No. 12305 [1] and Federal Decree No. 7404 [2]
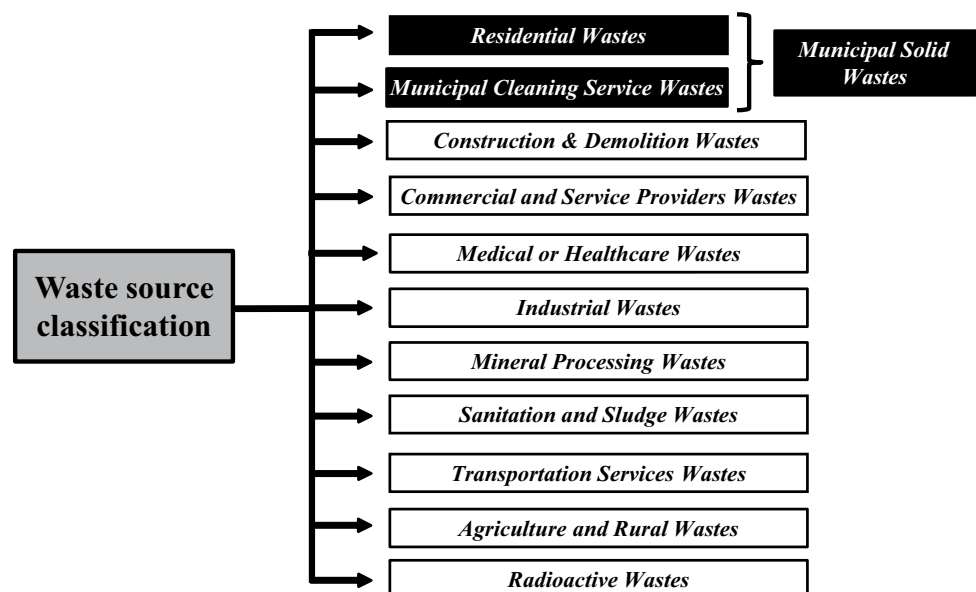
In 2016, the population of São Paulo state had 43,674,533 inhabitants, the population growth rate is around 0.83% per year. It has a population density of 175.95 inh. $km^{-2}$ and degree of urbanization of 96.37%. The illiteracy rate at the age of 15 years is 3.53%. The mean age of study of the population aged 15–64 years is 9.72 years. The gross domestic product was USD 573,031.72 million, and USD 12,777.62 per capita by year. The monthly per capita income in 2016 was USD $493.64 [20].

The São Paulo state has 645 municipalities. The household waste collection service covers 99.8% of the urban area [21], 94.1% of household waste is disposed in landfills, corresponding to 97.4% of the amount generated in the state of São Paulo [22]. The gravimetric composition of household waste is 2.3% of ferrous metals; 0.6% aluminum; 13.1% paper/cardboard; 13.5% plastic; 2.4% glass; 51.4% food waste and 16.7% non-recyclable materials. About 22.5% of the municipalities have selective collection programs for the recyclable materials. The average per capita recycled waste collection rate is only 5.8 kg inh.$^{-1}$ year$^{-1}$. Only 0.2% of the organic fraction of municipal solid wastes has been composted in São Paulo State [23].

## Materials and methods

Predicting the generation of MSW is essential for integrated solid waste management. In this context, the statistical methods and parameters used to verify the confidence level of samples, evaluate the dependence between variables and investigate the goodness-of-fit quality are described.

Total population and per capita income were used as independent variables because their influence is described in the literature [3, 5–10, 12, 14–19]. However, per capita electricity consumption was chosen because it is indirectly associated with the pattern of consumption, the available infrastructure and with access to goods and services, since it corresponds to the total of electric energy (residential, productive sector and public lighting) consumed in the municipalities in relation to the total resident population [9, 12].

The state of São Paulo is located in southeastern Brazil, bordering the state of Rio de Janeiro to the east, that of Minas Gerais to the north, Mato Grosso do Sul to the west and Paraná to the south (Fig. 2). In 2013, the state consisted of 645 municipalities, 533 of which participated in the 2013 Diagnosis of Solid Waste Management, and among them, 238 stated that they routinely weigh collected wastes. This group of 238 municipalities accounted for approximately 82% of the state's total population in 2013 (42,304,694 inhabitants). In 2014, 564 of the state's 645 municipalities participated in the 2014 Diagnosis of Solid Waste Management, and 252 reported routinely weighing their wastes. These municipalities had 36,366,233 inhabitants, or 85.2% of the state's total population in that year [21, 24].

## Data set

The initial data on solid waste generation (PCWG) in the municipalities that routinely weigh their MSW were obtained from the Brazilian National Information System on Water, Sanitation and Solid Waste—SNIS, which, in 2015 and 2016, published the Diagnosis of Urban Solid Waste Management documents corresponding to 2013 and
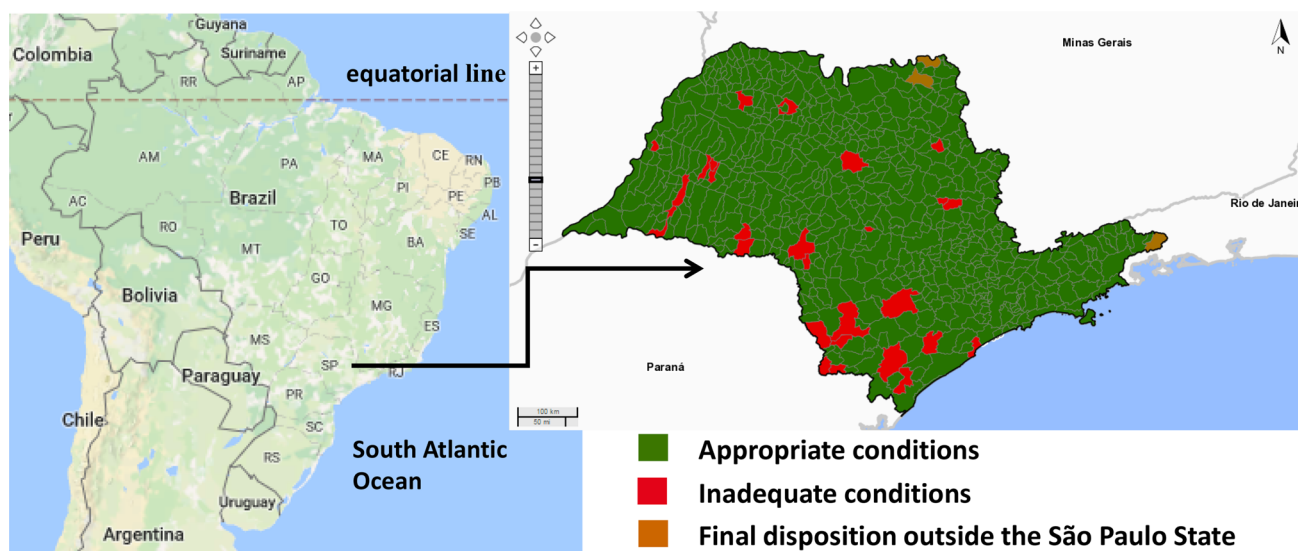


**Fig. 2** Geographical location of the state of São Paulo, highlighting the qualification of MSW final disposal in its municipalities in 2014, according to CETESB [22] and DataGeo [25]

2014 [21, 24]. Among other information, these documents provide separate descriptions of the amounts of household wastes, urban cleaning wastes and health care wastes, as well as the use of weighing machines to determine the total amount of wastes and the coverage of waste collection services. Therefore, it was hypothetically assumed that there were no portions of types of waste other than those included in the one classified in this survey. Per capita MSW generation rates that presented clear inconsistencies when the values reported to the SNIS were compared with those set forth in the Municipal Integrated Waste Management Plans (MIWMP) were substituted with the values reported in the MIWMP.

Information about the total population (*P*) and annual electricity consumption in each municipality in 2013 and 2014 was obtained from the State Data Analysis System Foundation website [20]. The annual per capita incomes in the municipalities of São Paulo were updated relative to the base year of 2010 and their values were converted into US dollars using this year's average exchange rate of R$ 1.76 per USD and these values were assumed constant in 2013 and 2014. In addition, the results were correlated in the form of per capita daily income (PCI) (USD inh.$^{-1}$ day$^{-1}$) and per capita daily energy consumption (PCE) in each municipality (kWh inh.$^{-1}$ day$^{-1}$).

## Sample representativeness

Since not all the municipalities of the state of São Paulo participated in the 2013–2015 Diagnosis of Solid Waste Management or reported routinely weighing their MSW, the sample space was established in the form of two main criteria, based on the number of municipalities or on the population covered in relation to the state's total population.

The sample's representativeness was evaluated based on the level of confidence associated with a margin of error. The margin of error corresponds to an interval that must comprise the estimated value within the scope of the survey. The confidence level expresses the degree of certainty that the value found actually lies within the stipulated margin of error. The relationship between these variables was expressed by (Eq. 1) [26]:

$$n = \frac{N\,Z^2\,p\,(1-p)}{(N-1)\,e^2 + Z^2\,p\,(1-p)}, \tag{1}$$

where *n* is the sample size, *N* is the size of the sample universe (645 municipalities or 42,304,694 inhabitants in 2013 and 42,673,386 in 2014), Z is the acceptable deviation from the average value of the adopted confidence interval (determined in the normal distribution curve) and is the stipulated margin of error, *p* is the expected proportion or heterogeneity of the sample space, usually 50% when information

on sample variability is not available [26]. Thus, margin of error and confidence interval values were adopted that would result in the number of municipalities or population that participated in the sample, based on the total number of municipalities or the population of the state.

## Evaluation of the model

Initially, the Pearson coefficients were used to evaluate the correlation between variables total population of each municipality (*P*), mean daily per capita income (PCI) and average daily energy consumption (PCE). Pearson coefficient below 0.30 led the variables were practically independent [27, 28]. After that, linear functions were adjusted to represent the PCWG rate as a function of the variables of total population, average daily per capita income, and average per capita energy consumption of each municipality in the sample space, combined in pairs independently or the three variables simultaneously. Later, the three independent variables were combined, but using logarithmic functions for total population (*P*) and per capita energy consumption (PCE) and linear dependence for per capita income (PCI) (Eqs. 2–7). This was done using free BioEstat 5.3 software for applications in Ecology, developed by the Mamirauá Institute for Sustainable Development [29], and Microsoft Excel, through the solver app for application of the least squares method.

$$PCWG = a + b\,P + c\,PCI, \tag{2}$$

$$PCWG = a + b\,P + c\,PCE, \tag{3}$$

$$PCWG = a + b\,PCI + c\,PCE, \tag{4}$$

$$PCWG = a + b\,P + c\,PCI + d\,PCE, \tag{5}$$

$$PCWG = a + b\,\ln(P), \tag{6}$$

$$PCWG = a + b\,\ln(P) + c\,PCI + d\,[\ln(PCE)]^e, \tag{7}$$

where *a, b, c, d* and *e* are fitting constants, *P* is the total population (inh.), PCE is the daily per capita energy consumption (kWh inh.$^{-1}$ day$^{-1}$) and PCI is the average daily per capita income (USD inh.$^{-1}$ day$^{-1}$) in each municipality in the state of São Paulo. The quality of the adjustment was evaluated through the determination coefficient ($R^2$), root-mean-square error (RMSE) and mean percentage error ($E_p$) values [27, 28, 30, 31].

The Akaike Information Criterion (Eq. 8, in the case of the least squares method) allowed to identify the foremost model among dissimilar ones with different degrees of freedom. A smaller AIC index (AIC) is implied in the best prediction. Thus, it was possible to rank the models by subtracting the AIC value from a given minimum value model (AIC$_{min}$). The difference was null for the best model [32].

$$\text{AIC} = n \, \ln \left\{ \sum_{i=1}^{n} \frac{\left[ \left( \text{PCWG}_{\text{observed}_i} - \text{PCWG}_{\text{adjusted}_i} \right)^2 \right]}{n} \right\} + 2\nu,$$

(8)

where is $\text{PCWG}_{\text{observed}_i}$ the observed dependent variable for each element of the data set, $\text{PCWG}_{\text{adjusted}_i}$ is the calculated dependent variable for each element of the data set, $n$ is the number of elements contained in the data set of each sample, and $\nu$ is the number of independent variables involved in the model.

Finding the best equation that fitted the PCWG in 2013 e 2014, the Student's $t$ test was applied to proposed equation to verify if was able to represent mean values and result in the same variance with more than 99% probability. This equation was also used to predict the observed PCWG for the year 2015 [20, 33]. The quality of the forecast was evaluated through the RMSE and $E_p$ values.

## Results and discussion

The original data from nine municipalities in 2013 and four in 2014 were disregarded because they presented excessively high PCWG rates; hence, the values reported to the National Information System on Water, Sanitation and Solid Waste (SNIS) were replaced with the generation rates found in the Integrated Solid Waste Management Plan (PCWG$_{\text{ISWMP}}$). Table 1 lists the information regarding these municipalities. Thus, only three municipalities were excluded from the data set because their plans could not be located.

The sample universe of 2013 comprised 238 municipalities, whose total population was 34,551,664 inhabitants (82%), compared to the number of municipalities that

responded to the data collection for the SNIS (533 municipalities) and to the total that made up the state of São Paulo, which was of 645 municipalities (42,304,694 inhabitants). Based on the criterion of number of cities, it can be stated that 238 municipalities represented the total of 645 with a margin of error of 5.1% and a confidence interval of 94.9%. On the other hand, based on the population, 34,551,664 inhabitants portrayed the total population of 42,304,694 inhabitants with a margin of error of only 0.015% and a confidence interval of 99.985%. In the data set for 2014 ($n = 251$), the sampling error for the number of participating municipalities ($e$) and the corresponding confidence intervals were, respectively, 4.9 and 95.1%. Conversely, based on the population covered, 35,229,824 inhabitants represented the total population of 42,673,386 inhabitants with a margin of error of only 0.000026% and a confidence interval of 99.999974%. Therefore, the samples used here proved to be representative of the state of São Paulo and the results obtained could be extrapolated to the other municipalities, including those that did not participate in the 2013 and 2014 Diagnosis of Solid Waste Management.

The Pearson coefficients were used to evaluate the correlation between variables total population of each municipality ($P$), mean daily per capita income (PCI) and average daily energy consumption (PCE) (Table 2). As extreme values ($-1$ and $1$) are rarely found in practice, it is important to qualify intervals for the coefficient $r$. For Cohen [30], values between 0.10 and 0.29 may be considered small; results between 0.30 and 0.49 can be evaluated as mean and from 0.50 to 1.00 as large. Dancey and Reidy [31] suggested a slightly different classification: $r$ from 0.10 to 0.30 for weak dependence between variables; 0.40–0.60 as moderate, and from 0.70 to 1.00 as a strong correlation. In consequence of Pearson coefficients had been less than 0.30, the degree of correlation between the

**Table 1** Municipalities for which information was disregarded

| Base year | Municipalities | Population (inh.) | PCWG$_{\text{observed}}$ (kg inh.$^{-1}$ day$^{-1}$) | PCWG$_{\text{ISWMP}}$ (kg inh.$^{-1}$ day$^{-1}$) |
|---|---|---|---|---|
| 2013 | Cruzália | 2222 | 1.24 | 0.84 |
| | Lutécia | 2682 | 1.56 | 0.86 |
| | Lavínia | 8666 | 1.99 | Not available |
| | Iacanga | 10,414 | 1.57 | 0.60 |
| | Ilhabela | 29,837 | 2.86 | Not available |
| | Paraguaçu Paulista | 42,858 | 3.20 | Not available |
| | Registro | 54,107 | 1.53 | 0.86 |
| | Atibaia | 130,606 | 1.58 | 0.86 |
| | Guarulhos | 1,260,840 | 1.53 | 0.84 |
| 2014 | Paraíso | 6060 | 3.09 | 0.83 |
| | Nhandeara | 10,780 | 2.98 | 0.93 |
| | Pedregulho | 15,940 | 2.43 | Not available |
| | Palmital | 21,362 | 2.07 | 0.71 |

**Table 2** Pearson coefficient (r) between variables total population of each municipality (*P*), mean daily per capita income (PCI) and average daily energy consumption (PCE) in state of São Paulo

| Compared variables (*n* = 489) | *r* (−) | Dependence |
|---|---|---|
| PCI and PCE | 0.20 | Weak |
| *P* and PCE | 0.23 | Weak |
| *P* and PCI | 0.01 | Weak |

variables was qualified as weak. Therefore, the variables were assumed independent.

The data set (*n* = 489) was represented by histograms, in which the frequency distribution of the independent variables (*P*, PCI and PCE) and the dependent variable (PCWG$_{observed}$) could be verified. The daily per capita energy consumptions were highly concentrated and near the lower limit of its domain, especially in relation to a linear scale. This also occurred, but at a lesser level, for the population of municipalities (Fig. 3).

Table 3 shows the main descriptive statistical parameter for the data set that form the sample universe (*n* = 489) in 2013 and 2014. Regarding the variables population and per

capita consumption of electricity, the medians lay between modes and means, but the tendencies of these variables were not similar of normal distributions, since the kurtosis was 218 for *P* and 60.5 for PCE, values much higher than zero expected for normal distribution. Moreover, the skewnesses of P and PCE were 14 and 6.4, respectively, which indicated *P* and PCE were concentrated near at the lower limit of the data set (positive right-skewed). The transformations of *P* in ln*P* and PCE in lnPCE made these extreme behaviors attenuated, and had become similar to the per capita income variable, which may favor processing and analysis.

Table 4 shows the Pearson coefficients for the observed per capita MSW generation rates (PCWG$_{observed}$) in relation to the independent variables of population (*P*), mean daily per capita income (PCI) and average daily energy consumption (PCE) for the years 2013 and 2014.

In Table 4, note that correlations involving ln(P), PCI and ln(PCE) potentially better fit the PCWG$_{observed}$ rates, these parameters are significant at 1% level for dataset involved in 2013 and 2014 (*n* = 489). This will be demonstrated when Eqs. 2–7 were applied to adjust PCWG.

Table 5 describes the parameters of evaluation of the quality of fit. Better fits were achieved by using Eq. 7, since
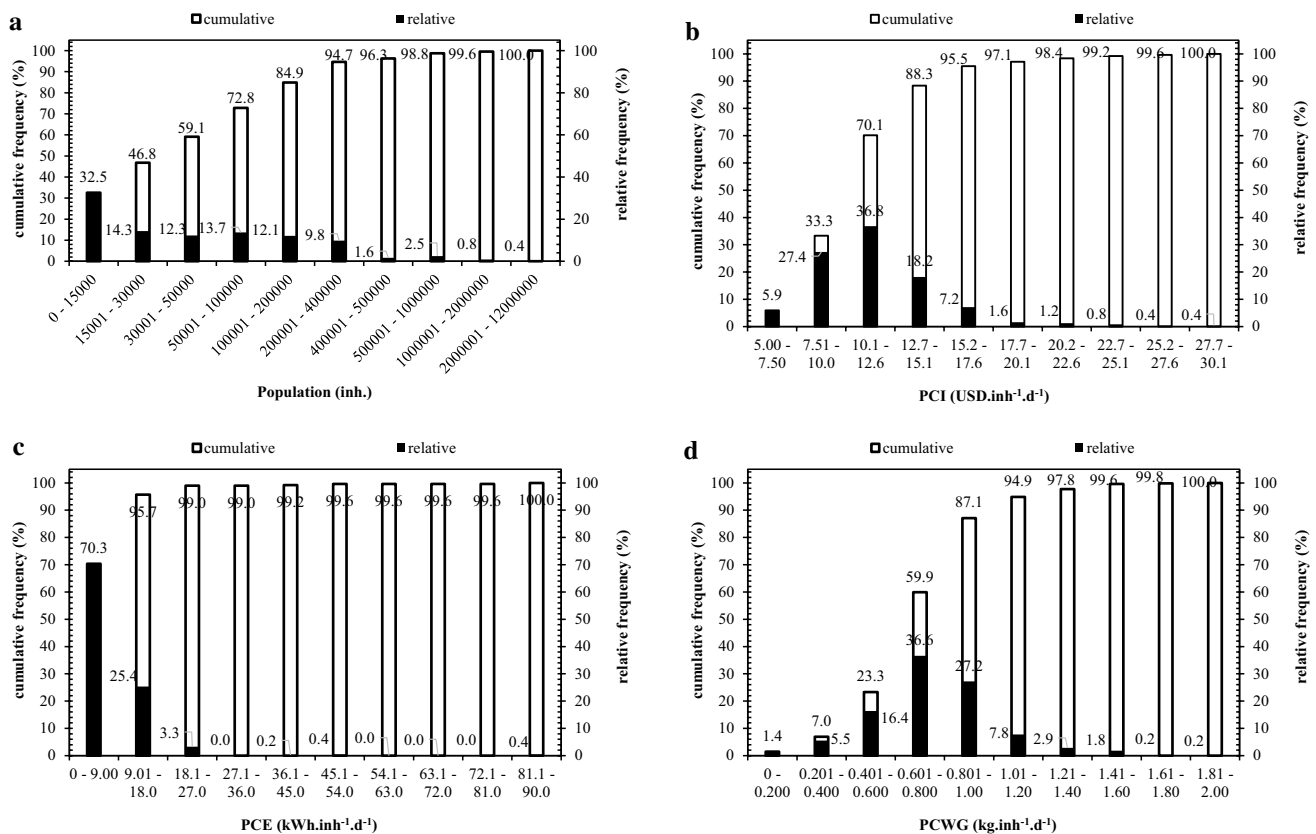


**Fig. 3** Frequency histograms for population of municipalities (**a**), mean daily per capita income (**b**), average daily energy consumption (**c**) and per capita municipal solid waste generation rates (**d**) of the state of São Paulo that reported weighing its MSW in 2013 and 2014 (*n* = 489)

**Table 3** Summary of descriptive statistical parameter for population of each municipalities, daily per capita income, daily per capita electricity consumption, and per capita municipal solid waste generation rates in São Paulo State in 2013 and 2014 ($n = 489$)

| Statistical parameter | $P$ (inh.) | PCI (USD inh.$^{-1}$ day$^{-1}$) | PCE (kWh inh.$^{-1}$ day$^{-1}$) | PCWG$_{observed}$ (kg inh.$^{-1}$ day$^{-1}$) |
|---|---|---|---|---|
| Arithmetic mean | 142,702 | 11.6 | 8.51 | 0.760 |
| Standard error | 33,734 | 0.153 | 0.32 | 0.012 |
| Median | 36,286 | 11.0 | 6.76 | 0.740 |
| Mode | 11,540 | 8.76 | 2.74 | 0.860 |
| Standard deviation | 745,964 | 3.38 | 6.99 | 0.256 |
| Variance | 5.56E+11 | 11.4 | 48.8 | 0.065 |
| Kurtosis | 218 | 4.77 | 60.5 | 1.788 |
| Skewness | 14 | 1.65 | 6.40 | 0.644 |
| Range | 11,511,913 | 24.8 | 84.5 | 1.875 |
| Minimum value | 1923 | 5.10 | 2.51 | 0.078 |
| Maximum value | 11,513,836 | 29.9 | 87.0 | 1.95 |
| Confidence interval (95%) | 66,281 | 0.300 | 0.621 | 0.0227 |

**Table 4** Pearson coefficient ($r$) between observed per capita MSW generation rates (PCWG$_{observed}$) and independent variables in the municipalities of the state of São Paulo

| | Independent variables | | | | | |
|---|---|---|---|---|---|---|
| | $P$ (inh.) | $\ln(P)$ | PCI (USD inh.$^{-1}$ day$^{-1}$) | $\ln(PCI)$ | PCE (kWh inh.$^{-1}$ day$^{-1}$) | $\ln(PCE)$ |
| $r$ in 2013 ($n = 238$) | 0.12 ($p = 0.065$)[*] | 0.41 ($p < 0.00001$)[****] | 0.43 ($p < 0.00001$)[****] | 0.42 ($p < 0.00001$)[****] | 0.21 ($p = 0.0011$)[**] | 0.32 ($p < 0.00001$)[****] |
| $r$ in 2014 ($n = 251$) | 0.18 ($p = 0.0042$)[**] | 0.43 ($p < 0.00001$)[****] | 0.40 ($p < 0.00001$)[****] | 0.41 ($p < 0.00001$)[****] | 0.27 ($p = 0.000014$)[****] | 0.37 ($p < 0.00001$)[****] |
| $r$ in 2013 and 2014 ($n = 489$) | 0.15 ($p = 0.00088$)[***] | 0.42 ($p < 0.00001$)[****] | 0.36 ($p < 0.00001$)[****] | 0.36 ($p < 0.00001$)[****] | 0.24 ($p < 0.00001$)[****] | 0.34 ($p < 0.00001$)[****] |

Significance test: *represents ($p > 0.05$), **represents ($p < 0.01$), ***represents ($p < 0.001$), and ****($p < 0.0001$)

AIC–AIC$_{min}$ was null. Despite the low coefficients of determination (0.24–0.26), the percentage deviations associated with the adjustments were acceptable, around $-14$ to $-10\%$, and the negative sign indicated a tendency to estimate values above those observed. The RMSE values were 0.217 kg inh.$^{-1}$ day$^{-1}$ in 2013, 0.227 kg inh.$^{-1}$ day$^{-1}$ in 2014, and 0.224 kg inh.$^{-1}$ day$^{-1}$ in 2013–2014, which represented 29% of the mean of all values observed. Thus, based on the year 2013, the use of these three variables resulted in Eq. 9:

$$PCWG = -11.280 + 0.0356\ \ln(P) + 0.0179\ PCI + 11.347\ [\ln(PCE)]^{0.0103} \tag{9}$$

which is valid for 1923 inh. $\leq P \leq 11,446,275$ inh; 6.23 USD inh.$^{-1}$ day$^{-1}$ $\leq$ PCI $\leq 29.89$ USD inh.$^{-1}$ day$^{-1}$; and 2.51 kWh inh.$^{-1}$ day$^{-1}$ $\leq$ PCE $\leq 87.04$ kWh inh.$^{-1}$ day$^{-1}$. For the 2014 sample, the regression resulted in (Eq. 10):

$$PCWG = -14.162 + 0.0441\ \ln(P) + 0.0160\ PCI + 14.164\ [\ln(PCE)]^{0.0154} \tag{10}$$

which is valid for the following intervals of 2,044 inh. $\leq P \leq 11,513,836$ inh., 5.10 USD inh.$^{-1}$ day$^{-1}$ $\leq$ PCI $\leq 24.48$ USD inh.$^{-1}$ day$^{-1}$ and 2.69 kWh inh.$^{-1}$ day$^{-1}$ $\leq$ PCE $\leq 81.67$ kWh inh.$^{-1}$ day$^{-1}$. A regression involving the samples of the years 2013 and 2014 simultaneously, i.e., 489 data, was also performed, and very similar trends and fit qualities were obtained (Eq. 11).

$$PCWG = -13.476 + 0.0436\ \ln(P) + 0.0114\ PCI + 13.521\ [\ln(PCE)]^{0.0143} \tag{11}$$

indicated for 1923 inh. $\leq P \leq 11,513,836$ inh., 5.10 USD inh.$^{-1}$ day$^{-1}$ $\leq$ PCI $\leq 29.89$ USD inh.$^{-1}$ day$^{-1}$ and 2.51 kWh inh.$^{-1}$ day$^{-1}$ $\leq$ PCE $\leq 87.04$ kWh inh.$^{-1}$ day$^{-1}$.

Figures 4, 5 and 6 show the per capita generation rates of MSW as a function of the variables of population, per capita income and energy consumption of the municipalities. It was found that the results tended to improve when using a logarithmic scale on the axes of the variables $P$ and PCE. This was corroborated by the higher Pearson coefficients

**Table 5** Parameters for evaluating fit quality between observed and predicted PCWG rates, based on Eqs. (2–7)

| Mathematical function | $R^2$ (−) | RMSE (kg inh.$^{-1}$ day$^{-1}$) | $E_p$ (%) | AIC | AIC– AIC$_{min}$ | Model selection position |
|---|---|---|---|---|---|---|
| PCWG = 0.454 + 2.44 × 10$^{-8}$ $P$ + 0.0261 PCI Equation (2)—2013 and 2014 | 0.14 | 0.238 | − 14.6 | − 1402 | 58 | 5th |
| PCWG = 0.689 + 5.08 × 10$^{-8}$ $P$ + 0.00750 PCE Equation (3)—2013 and 2014 | 0.08 | 0.246 | − 15.5 | − 1371 | 89 | 6th |
| PCWG = 0.419 + 0.0247 PCI + 0.00640 PCE Equation (4)—2013 and 2014 | 0.16 | 0.235 | − 14.1 | − 1416 | 44 | 4th |
| PCWG = 0.433 + 2.70 × 10$^{-8}$ $P$ + 0.0233PCI + 0.00651PCE Equation (5)—2013 and 2014 | 0.17 | 0.234 | − 14.4 | − 1417 | 43 | 3th |
| PCWG = 0.0457 + 0.0684 ln($P$) Equation (6)—2013 and 2014 | 0.17 | 0.232 | − 13.5 | − 1427 | 33 | 2nd |
| PCWG = −11.280 + 0.0356 ln ($P$) + 0.0179 $PCI$ + 11.347 [ln (PCE)]$^{0.0103}$ Equation (7)—2013 | 0.24 | 0.217 | − 14.2 | − 725 | – | – |
| PCWG = −14.162 + 0.0441 ln ($P$) + 0.0160 PCI + 14.164 [ln (PCE)]$^{0.0154}$ Equation (7)—2014 | 0.26 | 0.227 | − 10.2 | − 742 | – | – |
| PCWG = −13.476 + 0.0436 ln ($P$) + 0.0114 PCI + 13.521 [ln (PCE)]$^{0.0143}$ Equation (7) – 2013 and 2014 | 0.24 | 0.224 | − 12.3 | − 1460[*] | 0 | 1st |

*$AIC_{min}$ minimum of Akaike Information Criterion among all the models

calculated for ln($P$, ln(PCE) and PCI when compared with those obtained for $P$ and PCE (Table 4). In addition, the observed and predicted PCWG rates were highly dispersed, which implies a relatively low coefficient of determination ($R^2$). The PCWG rates in some municipalities are extremely low and in others unexpectedly high. For example, there are municipalities with populations of less than 10,000 whose PCWG rates exceed 1.00 kg inh.$^{-1}$ day$^{-1}$, which may be due to improper final disposal of construction waste in conjunction with urban solid waste. At the same time, some cities with a population ranging from 6000 to 60,000 have PCWG rates in the order of 0.10 kg inh.$^{-1}$ day$^{-1}$, possibly due to inconsistencies in the data reported by the municipalities

to the SNIS. It was also found that the PCWG rates showed an upward trend in relation to the independent variables of population, per capita income and daily per capita energy consumption in the municipalities.

The plots in Fig. 7 show the observed PCWG rates and those adjusted by Eq. 11 to evaluate the quality of fit; the closer the points are to the diagonal line, the more accurate the adjustment. As can be seen, therefore, most of the points obey this condition. However, note that there are conditions in which PCWG rates were clearly underestimated and overestimated, as shown by the dashed regions in the figure. Nevertheless, most of the points fall within a good prediction interval, i.e., between the
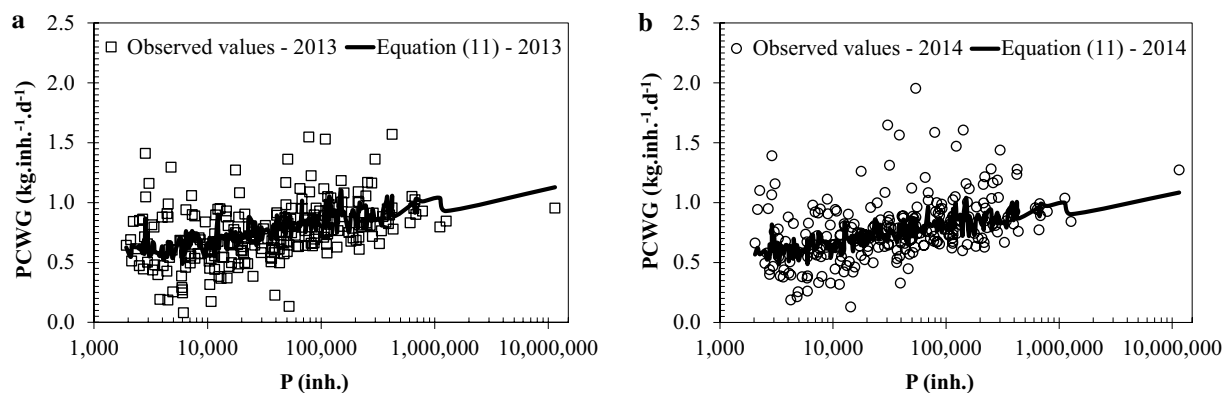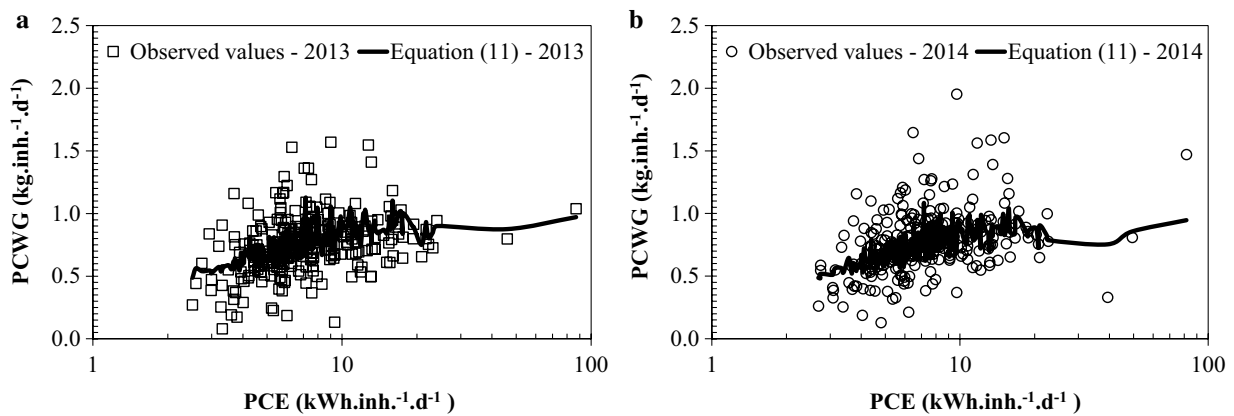


**Fig. 4** PCWG rates as a function of total population in each municipality of the state of São Paulo that reported weighing its MSW: **a** in 2013 and **b** in 2014
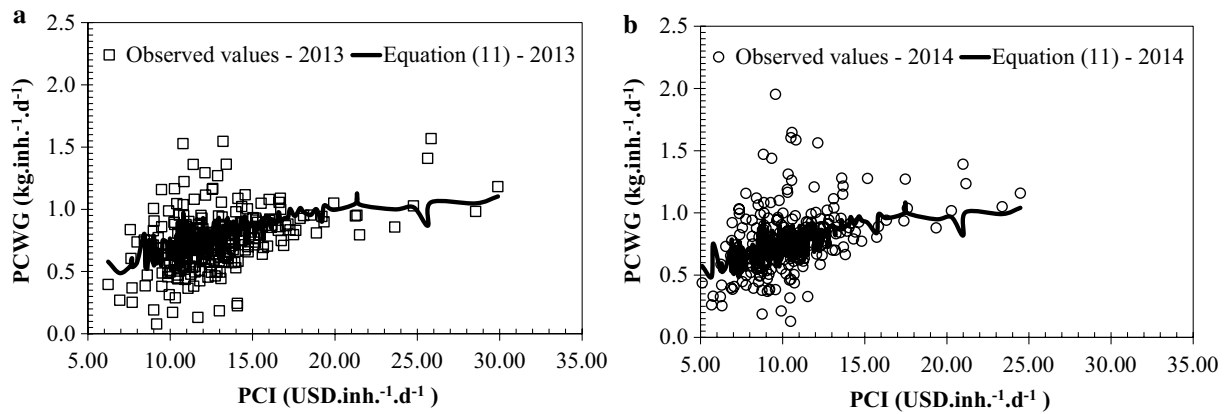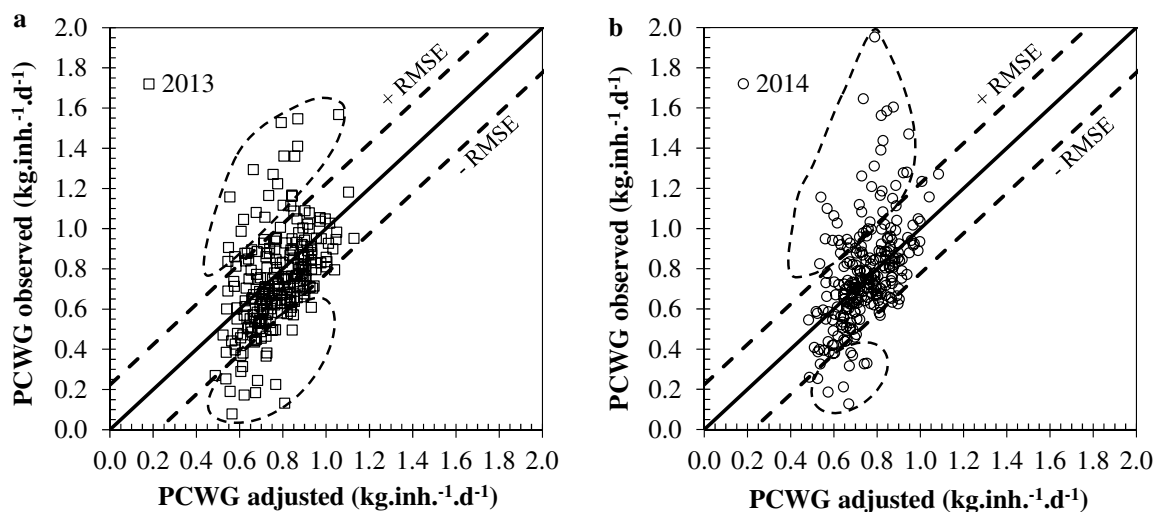
**Fig. 5** PCWG rates as a function of daily per capita income in each municipality of the state of São Paulo that reported weighing its MSW: **a** in 2013 and **b** in 2014



**Fig. 6** PCWG rates as a function of total daily per capita energy consumption in each municipality of the state of São Paulo that weighed its MSW: **a** in 2013 and **b** in 2014



**Fig. 7** Adjusted and observed values of per capita municipal solid waste generation rates in relation to the bisector of the first quadrant: **a** related to 2013 and **b** related to 2014
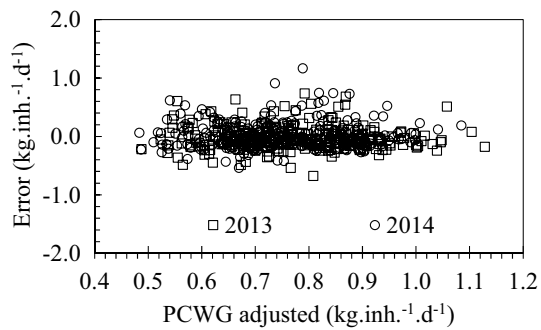
**Fig. 8** Error between observed and adjusted per capita municipal solid waste generation rates as a function of adjusted value
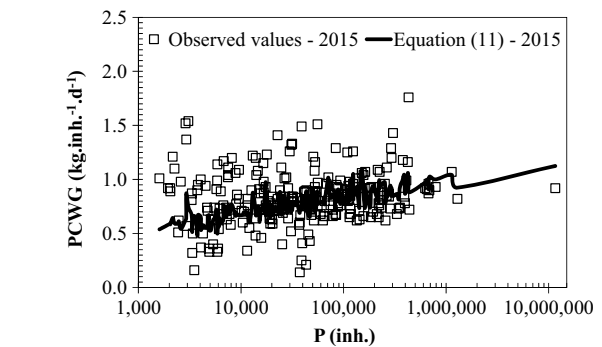


**Fig. 9** PCWG rates predicted for 2015 as a function of total population in each municipality of the state of São Paulo that reported weighing its MSW



**Fig. 10** PCWG rates predicted for 2015 as a function of daily per capita income in each municipality of the state of São Paulo that reported weighing its MSW



**Fig. 11** PCWG rates predicted for 2015 as a function of total daily per capita energy consumption in each municipality of the state of São Paulo that reported weighing its MSW

mean value and a root-mean-squared error of plus and of minus ($0.224$ kg inh.$^{-1}$ day$^{-1}$). As can be seen from the randomness of the distribution of errors compared to the PCWG rates adjusted by Eq. 11, there was no systematic error in the prediction (Fig. 8), despite a relatively low $R^2$ equal to 0.24 (Table 4).

Student's $t$ test proved that Eq. 11 was able to represent mean PCWG values with more than 99% probability. The two-tailed critical $t$ value was 2.59 for the same means and 2.58 for the same variances, while the $t$ value of the data set (489 elements) was 1.00 in the two tests, confirming that the means and variances of the observed and calculated values for 2013 and 2014 were practically equal. Therefore, these equations were considered adequate to represent the PCWG rates in the municipalities of the state of São Paulo.

The proposed equation (Eq. 11) was applied to predict PCWG rates for the next year (data released in 2017 relative to 2015). Equation (11) was able to describe the dependence of the PCWG on the variables $\ln P$, PCI and $\ln$PCE (Figs. 9, 10, 11). The mean percentage deviations associated with the prediction were substantially small, $-3.1\%$. The RMSE value was $0.255$ kg inh.$^{-1}$ day$^{-1}$, which represented 31% of the average observed rates in 2015, which were similar to the values found in the data adjustment (2013–2014). Therefore, Eq. 11 was shown to be adequate for predicting the per capita municipal solid waste generation in State of São Paulo. However, RMSE reduction should be pursued in future studies.
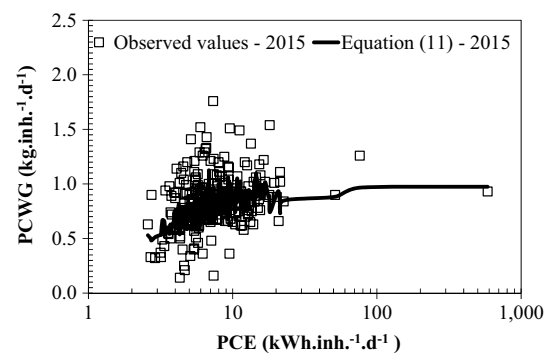
## Conclusions

In the present study, the per capita MSW generation rates in the state of São Paulo, Brazil, were investigated as a function of population, per capita income and energy consumption, which enabled us to draw the following conclusions:
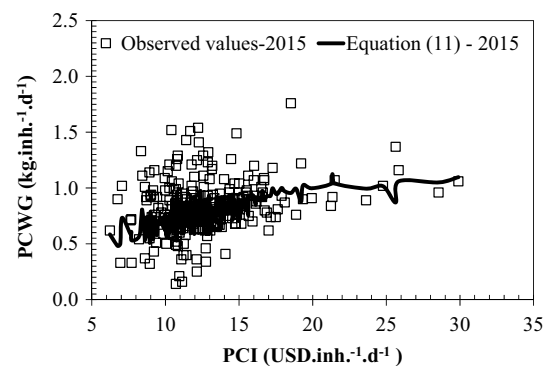
- The municipalities belonging to the sample proved to be representative of the state of São Paulo, in accordance with a minimum confidence interval of 94.1%, and the results can be extrapolated to the entire state.

- The independent variables of total population ($P$), daily per capita income (PCI) and daily per capita energy consumption (PCE) in each municipality interfered in the per capita municipal solid waste generation rates (PCWG), especially considering the logarithmic dependence of $P$ and PCE, and the linear role of the PCI. The Pearson correlation coefficients obtained here were compatible with a moderate correlation between the dependent and independent variables.

- The dependence of the PCWG on the variables $P$, PCI and PCE proved to have only one domain of validity; therefore, it was not necessary to subdivide the tested equations into different intervals.

- The PCWG rates varied significantly in the municipalities that reported routinely weighing the MSW produced, with maximum recorded values of 1.57 kg inh.$^{-1}$ day$^{-1}$ in 2013 and of 1.95 kg inh.$^{-1}$ day$^{-1}$ in 2014, in contrast to minimum values of 0.080 kg inh.$^{-1}$ day$^{-1}$ in 2013 and 0.13 kg inh.$^{-1}$ day$^{-1}$ in 2014, which were reflected in a relatively low coefficient of determination for the suggested equation ($R^2$ of 0.24).

- Linear correlations involving total population, per capita income, and per capita energy consumption were not satisfactory to predict the observed PCWG rates.

- However, considering the function involving variables in the form of $\ln(P)$, $\ln(PCE)$ and PCI, the mean percentage errors were $-14.1\%$ for 2013 and $-10.2\%$ for 2014. In addition, it was demonstrated that the highest predicted PCWG rates fell approximately within the range of the root-mean-square error (RMSE), rendering the prediction pertinent in most cases.

- The Student's $t$ tests performed to check the hypotheses of the equality of means in the paired case and of the same variances for the observed and predicted data (corresponding to 2013 and 2014), proved the validity of the proposed equation (Eq. 11) with a higher than 99% confidence interval.

- Equation 11 was associated with the RMSE value of 0.255 kg inh.$^{-1}$ day$^{-1}$, which represented 31% of the average observed rates in 2015, and to the mean percentage deviations of $-3.1\%$. Therefore, the proposed equation was appropriate to predict the per capita municipal solid waste generation in State of São Paulo.

## References

1. Brazilian Federal Law No. 12,305 of 02 Aug (2010) Establishes the National Solid Waste Policy; amends Law No. 9,605 of 12 Feb 1998, and makes other provisions (in Portuguese)

2. Brazilian Federal Decree No. 7,404/2010 of 23 Dec (2010) Regulates Law No. 12,305 of 02 Aug 2010, which establishes the national solid waste policy, creates the inter-ministerial committee of the national solid waste policy, the steering committee for the implementation of reverse logistics systems, and makes other provisions (in Portuguese)

3. Daskalopoulos E, Badr O, Probert SD (1998) Municipal solid waste: a prediction methodology for the generation rate and composition in the European Union countries and the United States of America. Resour Conserv Recycl 24:155–166

4. Chowdhury M (2009) Searching quality data for municipal solid waste planning. Waste Manag 29:2240–2247

5. Purcell M, Magette WL (2009) Prediction of household and commercial BMW generation according to socio-economic and other factors for the Dublin region. Waste Manag 29:1237–1250

6. Liu C, Wu X (2010) Factors influencing municipal solid waste generation in China: A multiple statistical analysis study. Waste Manag Res 29:371–378

7. Keser S, Duzgunb S, Aksoy A (2012) Application of spatial and non-spatial data analysis indetermination of the factors that impact municipal solid waste generation rates in turkey. Waste Manag 32:359–371

8. Ghinea C, Dragoi EN, Comanita ED, Gavrilescu M, Câmpean T, Curteanu S, Gavrilescu M (2016) Forecasting municipal solid waste generation using prognostic tools and regression analysis. J Environ Manage 182:80–93

9. Kawai K, Tasaki T (2016) Revisiting estimates of municipal solid waste generation per capita and their reliability. J Mater Cycles Waste Manag 18:1–13

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

11. Xu L, Gao P, Cui S, Liu C (2013) A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City China. Waste Manag 33:1324–1331

12. Benítez SO, Lozano-Olvera G, Morelos RA, Veja CA (2008) Mathematical modeling to predict residential solid waste generation. Waste Manag 28:S7 – S13

13. Navarro-Esbrí J, Diamadopoulos E, Ginestar D (2002) Time series analysis and forecasting techniques for municipal solid waste management. Resour Conserv Recycl 35:201–214

14. Campos HKT (2012) Evolution of income and per capita generation of solid wastes in Brazil (in Portuguese). Eng Sanit Ambient 17:171–180

15. Dias DM, Martinez CB, Barros RTV (2014) Generation estimate of municipal solid waste as subsidy actions aimed to environmental sustainability (in Portuguese). R Bras Cienc Ambient 33:13–20

16. Melo LA, Sautter KD, Janissek PR (2009) Scenario studies for the management of urban solid waste in Curitiba (in Portuguese). Eng Sanit e Ambient 14:551–558

17. Dias DM, Martinez CB, Barros PTV, Libânio M (2012) Model to domestic solid waste generation estimative in urban areas based on socioeconomic conjuncture variables (in Portuguese). Eng Sanit Ambient 17:325–332

18. Pinto MR, Pereira DRM, Freitas RC (2012) Factors social, economic and demographic associated with household waste generation in the city of Belo Horizonte (in Portuguese). Reuna 17:27–44

19. Silva H, Barbieri AF, Mór RLM (2012) Demography of urban consumption: a study on the generation of solid waste in the city of Belo Horizonte (in Portuguese). R Bras Est Pop 29(2): 421–449

20. Foundation state system of data analysis—SEADE (2016). http://www.seade.gov.br/. Accessed 20 May 2016

21. National Secretary of Environmental Sanitation – SNIS (2016). National Information System on Sanitation: Diagnosis of municipal solid waste management—2014 (in Portuguese). Ministry of Cities, Brasilia

22. Environmental Company of the São Paulo State–CETESB (2015) State inventory of household solid waste 2014 (in Portuguese). http://cetesb.sp.gov.br/residuossolidos/wpcontent/uploads/sites/26/2013/11/residuosSolidos2014.pdf. Accessed 29 Nov 2017

23. São Paulo State Secretariat of the Environment (2014). Solid waste plan of State of São Paulo. Environmental Company of the São Paulo State. http://www.ambiente.sp.gov.br. Accessed 10 Sep 2017

24. National Secretary of Environmental Sanitation – SNIS (2015). National Information System on Sanitation: Diagnosis of municipal solid waste management—2013 (in Portuguese). Ministry of Cities, Brasilia

25. Data Geo (2017) Infrastructure of environmental spatial data of the State of São Paulo—IDEA—SP. http://datageo.ambiente.sp.gov.br. Accessed 20 Ma 2017

26. Israel GD (1992) Determining sample size. Florida Cooperative Extension Service, University of Florida, Fact sheet PEOD-6, November, pp 1–5

27. Legates DR, McCabe GJ Jr (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res 35:233–241

28. Moriasi DN, Arnold JG, Liew MWV, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Am Soc Agric Biol Eng 50:885–890

29. Ayres M (2007) BioEstat Statistical applications in the areas of biomedical sciences (in Portuguese), 5 edn. Federal University of Pará, Belém

30. Cohen J (1988) Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale

31. Dancey C, Reidy J (2006) Statistics without math for psychology: using SPSS for Windows (in Portuguese). Artmed, Porto Alegre

32. Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York

33. National Secretary of Environmental Sanitation–SNIS (2017). National Information System on Sanitation: Diagnosis of municipal solid waste management—2015 (in Portuguese). Ministry of Cities, Brasilia