Waste Management 39 (2015) 26–34

Contents lists available at ScienceDirect

Waste Management

journal homepage: www.elsevier.com/locate/wasman

Identification of influencing municipal characteristics regarding household waste generation and their forecasting ability in Biscay

Iraia Oribe-Garcia *, Oihane Kamara-Esteban, Cristina Martin, Ana M. Macarulla-Arenaza, Ainhoa Alonso-Vicario

Deusto Institute of Technology, DeustoTech Energy, University of Deusto, Bilbao, Bizkaia, Spain

ARTICLE INFO

Article history:
Received 23 October 2014
Accepted 9 February 2015
Available online 11 March 2015

Keywords:
Household waste generation
Relevant municipal characteristics
Forecasting
Regression
Biscay

ABSTRACT

The planning of waste management strategies needs tools to support decisions at all stages of the process. Accurate quantification of the waste to be generated is essential for both the daily management (short-term) and proper design of facilities (long-term). Designing without rigorous knowledge may have serious economic and environmental consequences. The present works aims at identifying relevant socio-economic features of municipalities regarding Household Waste (HW) generation by means of factor models. Factor models face two main drawbacks, data collection and identifying relevant explanatory variables within a heterogeneous group. Grouping similar characteristics observations within a group may favour the deduction of more robust models. The methodology followed has been tested with Biscay Province because it stands out for having very different municipalities ranging from very rural to urban ones. Two main models are developed, one for the overall province and a second one after clustering the municipalities. The results prove that relating municipalities with specific characteristics, improves the results in a very heterogeneous situation. The methodology has identified urban morphology, tourism activity, level of education and economic situation as the most influencing characteristics in HW generation.

1. Introduction

Sustainable development of cities requires an integral waste management strategy that takes into account all the stages from the generation to the final disposal (Aranda Usón et al., 2013). When taking decisions about the design of infrastructures or the implementation of management policies, experts worldwide have recognized the importance of considering the whole system in a holistic manner (Coffey and Coad, 2010). Waste treatment is usually pointed out as the most important stage of Urban Waste (UW) management. Nevertheless, this stage is directly linked to waste quantity and quality (composition). An inaccurate estimation of the amount of waste generated difficulties to optimise the design of the required infrastructures and facilities. Under- or over-estimation of the UW generation has therefore significant consequences in terms of additional costs and environmental impacts (Beigl et al., 2003). In general, the accurate design of the waste management strategies requires of meticulous analysis of the generation data.

The success of waste management planning either for short-term (daily municipal management) or long-term (design of processing facilities), lies in the knowledge of the problem as well as in the accuracy and reliability of the used data (Chen and Chang, 2000; Navarro-Esbrí et al., 2002; Zaman and Lehmann, 2013). Forecasting models are useful to estimate future UW generation profiles. However, forecasting or estimation of waste generation is not an easy issue, mainly due to the generally little amount of available data and to the rapid change of factors that may influence it, such as socio-economic factors like gross domestic product in developing countries or the impact of the tourism among others (Beigl et al., 2008; Mateu-Sbert et al., 2013).

There is a huge range of forecasting methodologies applied to UW generation, which are classified in two wide groups: qualitative and quantitative models. The first ones are based on expert knowledge and do not necessarily use quantitative data. The second ones are more comprehensive and can provide better results when accurate data about the influencing factors are available (Armstrong, 2001). The literature shows a wide range of quantitative models, but three main groups can be distinguished: time series models, data-driven models and factor models.

Time series models aim at deducing variation patterns with time and show great ability to determine data repeatability. These models only need historical data about the dependent

* Corresponding author at: Deusto Institute of Technology, DeustoTech Energy, University of Deusto, Av. de las Universidades 24, 48007 Bilbao, Bizkaia, Spain. Tel.: +34 944139000.

E-mail address: iraia.oribe@deusto.es (I. Oribe-Garcia).

http://dx.doi.org/10.1016/j.wasman.2015.02.017
0956-053X/

variable. The simplest models used simple autocorrelation functions for the detection of autocorrelation embedded in time series data (Chang and Lin, 1997). Time series models have also been used successfully in order to assess the seasonal variations of waste generation (Denafas et al., 2014). Other models combine autoregressive techniques with seasonal exponential smoothing (Rimaityte et al., 2012), grey system theory (Xu et al., 2013) or support vector machines (Pai et al., 2010). Data-driven models run input–output data being able to identify their relationships. In the UW generation several applications have been presented using neuronal networks (Kumar et al., 2011) or support vector machines (Abbasi et al., 2013).

The main drawback of these methods is that they do not allow empirical reasoning about the influencing factors, which makes it difficult to identify the most important aspects in UW generation and consequently, to implement measures to reduce or control its generation (Noori et al., 2009; Shan, 2010).

Factor models or regression models are statistical models that provide insights of the reasons behind the UW generation. They allow identifying the interrelationships among different socio-economic factors with UW generation. These methods have been widely used in order to explain UW generation due to their mature theory and simple algorithms in order to forecast daily or annual generation (Lebersorger and Beigl, 2011; Ojeda-Benítez et al., 2008), at household, municipal or regional level (Afon and Okewole, 2007). Being easily applied, the main difficulty lies in the preparation of the data.

UW generation is a direct consequence of human daily activities, and it is closely related to consumption patterns, which are very local and generally depend on social, cultural, economic, environmental and demographic factors (Li et al., 2011). Therefore, applying techniques that encourage grouping observations with similar characteristics should result on more robust and precise models. Beigl et al. (2004) and Bandara et al. (2007) stratified municipalities according to prosperity and income levels deducing a multivariable regression model for each stratum. Other authors have applied GIS-based estimating techniques in order to include the spatial dependency of socio-economic characteristic in their models (Keser et al., 2012; Purcell and Magette, 2009).

The aim of this work is to develop models based on relevant socio-economic features regarding Household Waste (HW) generation at municipality level, and assess their forecasting ability. The resulting models will be able to support decision making process in the short-term planning of HW management. The applicability of the methodology followed is shown using real data of socio-economic features such as unemployment rate or tourist activity, from the province of Biscay.

2. Materials and methods

2.1. Case study: the province of Biscay

Biscay is one of the three historical territories of the Autonomous Community of the Basque Country, along with Gipuzkoa and Araba. Biscay is a territory with a long industrial trajectory, but ever since the deep deindustrialization, the economy has come to rely on the services sector. Located in the north of the Iberian Peninsula, Biscay has 1.1 million inhabitants, an average population density of 523 people per square kilometre, and an area of 2217 km^2. The province has 112 municipalities, the vast majority of which, 79 out of 112, are classified as rural (Fig. 1). According to the classification used in Biscay, rural municipalities are those with less than 5,000 inhabitants, and urban municipalities those with more than 50,000 inhabitants. The remaining ones are, those between 5,000 and 50,000 inhabitants denoted semi-urban municipalities (Basque Government, 2010, 2009).

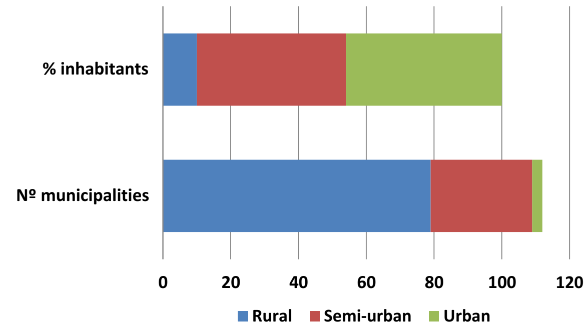


Fig. 1. Distribution of Biscay's municipalities by inhabitants according to waste management companies (Basque Government, 2010, 2009): rural (<5,000 inh), semi-urban (5,000 < inh < 50,000), urban (inh < 50,000).

Biscay towns are mainly grouped in communities (unions of services) in order to accomplish waste collection activities (Lozano Valencia and Lozano Valencia, 2008). All towns use separate collection systems for main recyclable materials. Glass, lightweight packages, paper and cardboard are collected mainly at drop-off points, and kerbside collection is used for mixed waste. Additionally, oil, textiles and batteries, as well as bulky and miscellaneous wastes, are collected separately by specific management services or at clean points. The 1999/31/CE European Directive (EU, 1999), transposed to Spanish legal framework by the Royal Decree 1481/2001 (MMA, 2001), established that in 2016 the amount of biodegradable waste sent to landfill must be reduced to 35% of the total generation in 1995, encouraging the separate collection of the organic fraction. Not only Biscay municipalities have started to implement totally or partially this collection system, 26 out of 112 municipalities, but also different treatments has been launched, such as composting or mechanical–biological treatment in order to accomplish that values (Biscay Provincial Council, 2013).

According to Biscay Council's terminology (Biscay Provincial Council, 2012a), Urban Wastes (UW) are classified according to Eq. (1). Industrial wastes are counted separately. Additionally, Household Wastes (HW) are divided into wastes strictly produced at home and similar waste produced at service establishments. Another difference is made with regard to the collection system. Wastes are distinguished between those collected separately, from those collected in a mixed way.

$$\begin{aligned} \textit{Urban Waste}(UW) = \textit{Household waste}(HW) \\ + \textit{Commercial Waste}(CW) \end{aligned} \tag{1}$$

In 2012 UW were composed of 75% of HW, and 25% of CW. The 79% of HW (without taking into account building waste) were not collected separately, that is as mixed waste. The remaining 21% was collected separately (Biscay Provincial Council, 2012b). The waste streams separately collected were sent to recycle programs. The mixed waste was partially incinerated and partially sent to landfill, 52% and 27%, respectively. Fig. 2 shows the evolution of the fractioning ratio, that is, the ratio between the gross amount of one waste fraction collected separately with the total amount of waste generated. Thus, for the 2012, the mixed waste was nearly 80% of the total generation, paper and cardboard 5%, glass 4%, lightweight packaging 3% and other separate fractions 8%.

2.2. Methodology

The methodology followed to identify relevant municipal characteristics regarding HW generation is depicted in Fig. 3. Hereinafter the "dependent variable" will refer to the HW per

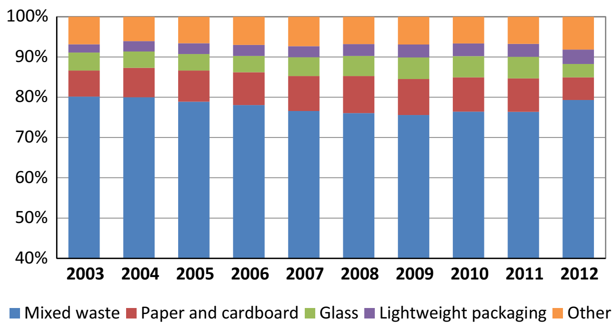


Fig. 2. Evolution of the fractioning rate of HW main waste streams.

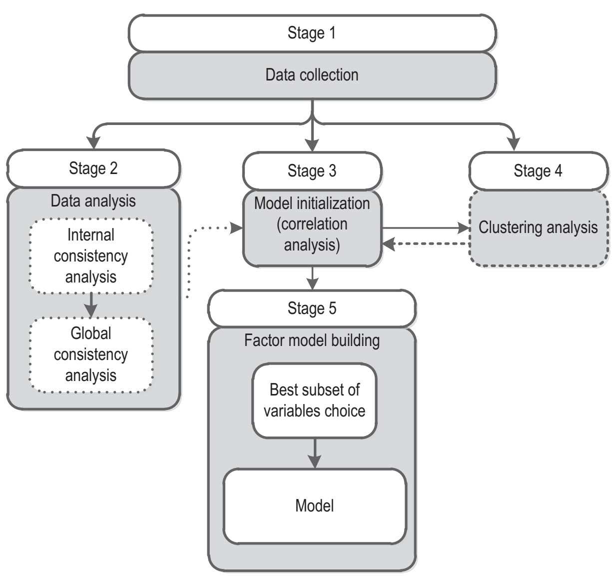


Fig. 3. The flowchart for factor models building in terms of sustainability indicators for the management of Household Waste (HW).

capita generation, and the "explanatory variables" to the data describing the population features, such as, unemployment rate, industrial activity, tourism relevance, etc. Firstly, data about the dependent and explanatory variables are collected (Stage 1). Subsequently, data about the dependent variable is analysed in order to find anomalous data, which will be corrected, and to identify municipalities with different behaviour (Stage 2). Previous to the model building, each model is initialized according to a correlation analysis (Stage 3). Additionally, and due to the aim of identifying influencing municipal characteristics regarding HW generation, municipalities are grouped by means of similar socio-economic characteristics (Stage 4), and the models are rebuilt. Finally, the factor model is built for each of the identified observation groups and its forecasting ability is assessed (Stage 5).

2.2.1. Step 1: Data collection

The first stage aims at collecting information about both the dependent variable and the explanatory variables. Data about dependent variable was gathered from Waste's Permanent Observatory of Biscay Provincial Council (Biscay Provincial Council, 2012b). The web portal contains data for each municipality of Biscay for the 1999–2013 period about the separate collected waste streams glass, paper and cardboard, lightweight packaging, bulky waste and battery waste, as well as data of not properly separated waste, that is mixed waste. In the 2013, data about separately collected biowaste was added. In order to check the data, complementary data was collected from waste management companies (Ecovidrio, 2013; Garbiker, 2013) and from municipalities or communities by personal communication.

The dependent variable is defined as the HW annual per capita generation at municipal level, according to the assertion concerning the dominant influence of the population. This study only applies to HW generation which is related to human daily activity and socio-economic characteristics. On the contrary, Commercial Wastes (CW) are more specific to their activity. Moreover, HW generation is significantly higher comparing to CW generation, 75% of HW generation against 25% CW generation (Biscay Provincial Council, 2012b). However, it should be taken into account that despite the division between household and commercial waste seems easy, nowadays it is nearly impossible the actual accounting for the main waste streams such as glass, paper and cardboard or mixed waste, due to the incorrect use of the service.

This study only takes into account HW's main waste streams: glass, paper and cardboard, lightweight packaging and mixed waste. Other waste streams are left aside due to missing data or unsteady generation. In the same manner, the data of 1999 year is not included in the study, due to missing data.

Literature mentions a wide range of socio-economic variables (explanatory variables) influencing UW generation and thus HW. In order to assure the applicability of the models in the future, information about explanatory variables from public open-sources is used, such as Udalmap (Udalmap, 2013) and Eustat (Eustat, 2013). An overall of 146 potential explanatory variables are gathered.

2.2.2. Step 2: Analysis of HW generation data

The aim of this step is to identify outlier municipalities with atypical or extreme behaviour over the entire time-spam. Data analysis is essential for the quality of the results. The process includes two steps to analyse the consistency of municipality data by means of Box and Whisker plot. Outliers are defined on the basis of the interquartile range (IQR). That is, a value which is 1.5 or 3 times the IQR distance from the first or third quartile is considered atypical or extreme value respectively.

The first step consists of internal data consistency analysis of each municipality, in order to correct occasional errors or deviations from the normal municipality behaviour. These occasional outliers, if possible, are corrected by other information sources, or alternatively by linear interpolation.

The second step consists of global consistency analysis. Municipalities which appear to be outliers with respect to the rest over the entire time-spam are excluded from the general group and analysed aside.

2.2.3. Step 3: Analysis of socio economic variables influencing HW generation

The aim is to model the HW generation by means of different municipal characteristics. Hence, different screening criteria are set according to the state of the art analysis. The first filtering process avoids duplicate explanatory variables, explanatory variables with insufficient data, explanatory variables for gender discrimination, etc. Subsequently, the explanatory variables are classified in different significant groups describing the main socio-economic characteristics of the different municipalities.

The model building process will be initialized by using one single explanatory variable for each group of the main socio-economic features considered regarding a bivariate explorative analysis. The selection of the pre-selected variables follows an iterative process. Initially the variable with the highest correlation within each group is included in the model, changing by another variable of the same group until the best one is found. Following this procedure, we ensure that only one variable within each group is included in the model.

2.2.4. Step 4: Clustering analysis

In a further step, and due to the aim of identifying relevant socio-economic characteristics regarding HW generation, municipalities with similar characteristics are grouped by a hierarchical clustering process. The hierarchical methods begin with N clusters consisting on the number of individuals (in this case municipalities), search the similarity matrix for the most similar pair of clusters and reduce the number of clusters until one through merging the most similar pair of clusters.

Criteria for the selection of cluster variables are methodological (partially correlated with the dependent variables and not highly correlated among then) and factual considerations. Dependent variable is not used as cluster criteria due to its nature as explained variable.

This research work use the squared Euclidean distance and the Ward's methods as clustering distance and algorithm respectively. The squared euclidean distance (Eq. (2)) measures the distance between individual, where x_{ic} and x_{jc} are the geometric centres of different individuals or clusters. The Ward method's objective is to find at each step a pair of clusters that would lead a minimum increase of variance by merging them (Hervada-Sala and Jarauta-Bragulat, 2004).

$$d(x_i, x_j) = \sum_{c=1}^{p} (x_{ic} - x_{jc})^2 \tag{2}$$

The number of final cluster is defined following the elbow criterion, that relates the number of clusters with the average of the distances between them (Mooi and Sarstedt, 2011).

2.2.5. Step 5: Model building

The model is developed in order to support the decision making process one year ahead ($n + 1$), assuming that waste management planners are taking decisions during the year n using data from $n-1$ year (Eq. (3)). In this manner, short-term waste management would be supported by the use of macro-economic variables (VAR_1, $VAR_2 \ldots VAR_m$).

$$HW_{n+1} = k + b_1 VAR_{1,n-1} + b_2 VAR_{2,n-1} + \ldots + b_m VAR_{m,n-1} \tag{3}$$

where k, b_1, $b_2 \ldots b_m$ are the parameter estimators of the models to be determined at each case study. Note that while data analysis process takes into account the whole time spam of available data (2000–2013), the models are built using data only from the last two years, the first one for training the model and another to validate it.

Best subset procedure (Frost, 2012) is used to build the model. It consists of evaluating different subsets of explanatory variables (one per significant group) in terms of R^2 and Mallows's Cp statistics. Using only R^2 could be misleading because it improves whenever a new variable is included. Therefore, the decision is supported by Mallows's Cp statistic, which represents an interesting selection criteria since it takes into account the number of factors included (Mallows, 1973), and besides, Minitab 16 software use it in order to execute Best subset Regression. Cp is defined by (Eq. (4)).

$$C_p = \frac{SSE_p}{\bar{\sigma}^2} - n + 2p \tag{4}$$

where SSEp is the residual sum of squares from a model containing p parameters, $\bar{\sigma}^2$ is an estimate of the error variance with all possible variables, and n is the number of observations. If $\bar{\sigma}^2$ is an unbiased estimator of total variance, the ratio $SSE_p/\bar{\sigma}^2$ has an expected value $n-p$ and hence a well fitted model is close to p (Siniksaran, 2008). Generally, models with low Cp values close to p are desirable. The Minitab 16 and SPSS 19 software have been used for statistical analysis.

Concurrently, and in order to ensure the validity of the model, the regression model's assumptions of multicollinearity, linearity, homoscedasticity and normality of the error distribution will be corroborated. Multicollinearity problems may appear when two or more explicative variables are highly correlated increasing the standard error of the model and the uncertainty of the estimated coefficients (high p values). Variance Inflation Factor (VIF) is

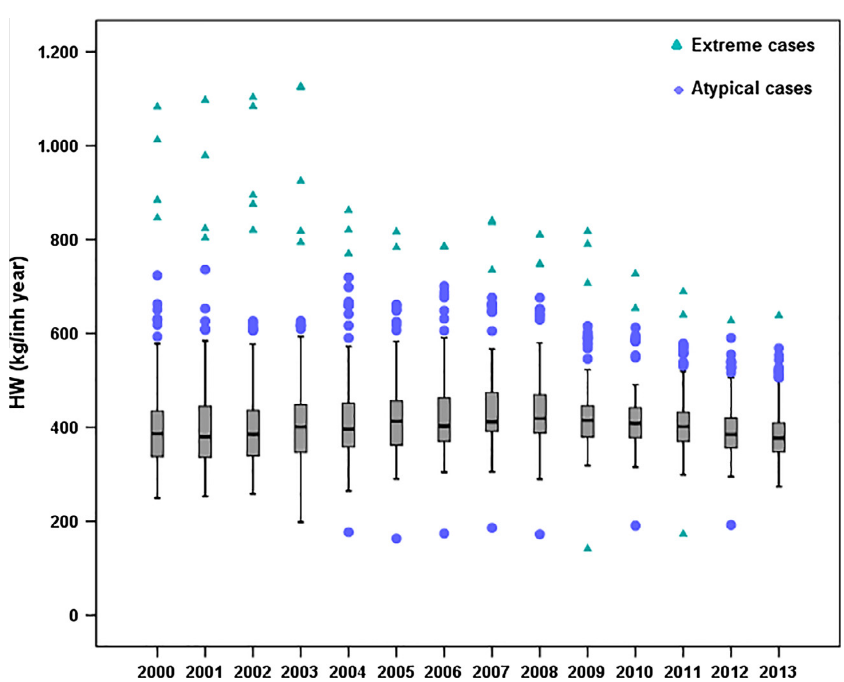


Fig. 4. HW per capita generation in Biscay municipalities.

usually used to identify this effect. VIF indicator measures the increase of the variance of a given factor due to the presence of multicollinearity. VIF $\leqslant 1$ indicates no multicollinearity, while predictors with VIF > 1 may be correlated. There is no clear threshold for this statistic, but the general acceptance is that VIF should not exceed $1/(1-R^2)$ (Kleinbaum et al., 1998). In addition, the decision of the existence of multicolinearity is supported with analysis of the condition index (CI) indicator and the variance proportion of each factor (Belsley et al., 1980). Note that VIF and CI analysis are complementary to support the same results.

Residuals should be normally distributed. This assumption is assessed by means of histogram and normal probability plot. Residuals should also hold homoscedasticity or the assumption of equal variances, which implies that the variation of residuals is uniform throughout the range of values; this will be assessed by scatterplot of standardized predicted dependent variable against the standardized residuals. Outliers are identified by means of standardized residuals and leverages. The first are deviations between observed values and fitted ones, while leverages measures abnormal values of the explanatory variables. The model is readjusted after excluding them.

2.2.6. Step 6: Forecasting ability

The model's forecasting ability is assessed by means of the Mean Absolute Percentage Error (MAPE, Eq. (5)), where y is the observed value, f is the forecasted value for i municipality, and n is the number of observations. MAPE is selected to evaluate the performance of the models since it is dimensionless.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - f_i}{y_i}\right| \tag{5}$$

Firstly, training-MAPE ($MAPE_t$) is analysed and then model's forecasting ability is evaluated ($MAPE_v$) by using data from 2011 to forecast HW generation of 2013.

3. Results and discussion

3.1. HW data generation in Biscay

The available information about dependent variable includes 14 datasets about glass, paper and cardboard, lightweight packaging and mixed waste generated by the 112 municipalities of Biscay for the 2000–2013 period, that is, exactly 6252 data (note that until 2004 in Biscay were 111). The two-phase correction process was performed for a total of 99 data, that is, 1.58%. Fig. 4 illustrates the municipalities identified by having an anomalous behaviour along the data range analysed, which comprised a total of 12 municipalities that host 2.4% of the total population. Some of these towns are characterized by a high tourist activity that implies high seasonal variability of the population along the year. Moreover, the management of 6 of them belong to a same community, which might be using a different quantification system. Fig. 4 shows that the average HW generation in Biscay has decreased from 418 kg inh^{-1} in 2008 to 377 kg inh^{-1} in 2013. Despite the population has increase in about 10,000 inhabitants, that is less than 1,5% of the current population, the waste generation has decreased by 10%, probably due to the financial crisis.

3.2. Data analysis, selection of main variables

In the Step 3, a deep analysis of the state of the art about the type of explanatory variables usually used in the literature was made. This was made in order to establish a set of screening criteria which would allow a previous reduction in the number of the explanatory variables. Thus, a set of 40 potential explanatory variables out of the initial 146 were identified. Table 1 resumes the seven most representative groups of variables in which the explicative variables with similar meaning were gathered. The tourist activity of the municipality will be featured in the models by at least one of the associated variables, the spaces for tourist accommodation (tACCO) and the ratio of hotel and catering

Table 1
Seven most representative groups of explicative variables (socio-economic indicators), the definition of the explicative variables and their acronym, and selected variables normality and the correlation with HW analysis.

Significant groups of explicative variables (main socio-economic features)	Explicative variables		Normality[1]	Correlation with HW[2] r_s
	Definition of the indicators	Acronym		
Economic structure	Population employed in the agriculture and fisheries sector (%)	popAGRO	No	0.362 [a]
Economic dynamism and resources of population	Unemployment rate aged 16 to 64 (%)	UNEM	Yes	−0.470 [a]
	Total personal income (€ Base CAE = 100)	INCOME	No	0.290 [a]
	Available personal income (€ Base CAE = 100)	INCOMEav	Yes	0.294 [a]
Tourist activity	Spaces for tourist accommodation (‰ inhabitants)	tACCO	No	0.225 [a]
	Hotel and catering establishments (‰ inhabitants)	HOCA	No	0.361 [a]
Commercial activity	Density of Retail outlets (‰ inhabitants)	RetD	No	0.447 [a]
	Density of occasional goods outlets (‰ inhabitants)	OCCA	No	−0.429 [a]
Demography	Overaging index	Over	No	0.204 [b]
Level of education	Population over 10 years that have completed at least secundary edu. (%)	SECUN	No	0.239 [b]
	Population over 10 years that have completed at least university education (%)	UNI	No	0.263 [a]
Urban morphology	Average surface of family dwelling (m^2)	DWE	No	0.438 [a]
	Density of housing on resident land (dwelling/ha)	denDWE	No	−0.372 [a]
	Municipal urban land (%)	URB	No	−0.327 [a]
	Population density (inh/km^2)	denPOP	No	−0.448 [a]

[1] Normality according to Anderson–Darling test.
[2] Correlation analysis according to Spearman coefficient.
[a] Significant level at 99%.
[b] Significant level at 95%.

Table 2
Factor models with their main characteristics.

Dependent variable and model	Model summary		Explicative variables			Significance test		Collinearity analysis	
				Non standardized coefficients	Beta coefficients	T	p value	VIF	Condition index
HW_{PM}	R^2	0.424	Constant	344.57		18.771	0.000		15.999
	S	27.512	URB	−1.01	−0.355	−3.964	0.000	1.325	
			HOCA	2.56	0.308	3.649	0.000	1.178	
			UNI	1.44	0.271	3.325	0.001	1.098	
			UNEM	−1.42	−0.191	−1.805	0.018	1.592	
HW_{CHH}	R^2	0.980	Constant	382		31.014	0.000		17.12
	S	4.016	HOCA	9.24	0.420	4.789	0.017	1.184	
			UNEM	−10.19	−1.077	−12.269	0.001	1.184	
HW_{CMH}	R^2	0.948	Constant	211		12.975	0.000		12.533
	S	9.346	HOCA	26.85	0.995	11.294	0.000	1.049	
			denPOP	−0.01	−0.278	−3.159	0.016	1.049	
HW_{CLL}	R^2	0.279	Constant	330		21.922	0.000		9.493
	S	32.99	HOCA	2.37	0.282	2.857	0.005	1.083	
			UNI	2.01	0.336	3.520	0.001	1.009	
			denPOP	−0.03	−0.191	−1.945	0.040	1.074	

establishments (HOCA). Similarly, the average surface of family dwelling, population and dwelling density and the ratio of municipal urban land describe the urban morphology of each town.

As previously mentioned, the model building will be initialized by using one single explicative variable within each significant group. A bivariate explorative analysis is conducted using the Spearman correlation (r_s) coefficient in order to assess the correlation among dependent and explanatory variables, due to most of the explanatory data do not follow a normal distribution according to the Anderson–Darling test. Additionally, r_s has the strength to evaluate monotonic relationship between variables, that is, the variables tend to change together but not necessarily at constant rate. The variable selected will be the one with highest correlation coefficient with the dependent analysis (Table 1). Initially included variables are popAGRO (Economic structure), UNEM (Economic dynamism and resources of population), HOCA (Tourist activity), RetD (commercial activity), UNI (level of education), Over (Demography) and denPOP (Urban morphology).

3.3. Factor models for the HW generation in Biscay

As mentioned in global data consistency analysis, there are 12 municipalities that do not follow the general group's generation profile (Fig. 4). For that reason, a partial model (*PM*) is developed without those municipalities (Eq. (6)).

$$HW_{PM} = 344.57 - 1.01\ URB + 2.56\ HOCA + 1.44\ UNI - 1.42\ UNEM \quad (6)$$

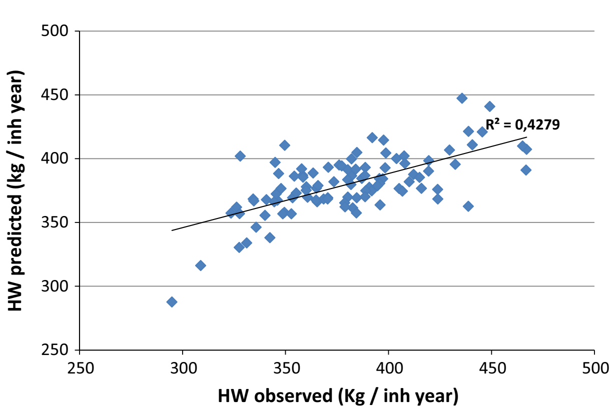


Fig. 5. PM model fit.

Table 2 shows the main characteristics of the PM model. The model with the explanatory variables HOCA, UNI, UNEM and URB shows the best performance explaining 42.4% of the variation of HW (Fig. 5). The typified beta coefficients show that municipal urban land ratio (URB) has the highest relative impact in the model.

The descriptive capability of the model is within the published results in literature. According to Lebersorger and Beigl (2011), R^2 rarely exceeds 50%, except for cases with small sample size or models with large amount of independent variables. Ojeda-Benítez et al. (2008) developed a model with three variables that explained 51% of the daily UW per capita generation in Mexicali (Baja California); Lebersorger and Beigl (2011) achieved to explain 74.3% of the variability with a four variable model applied to several municipalities in Styria (Austria) and Afon and Okewole (2007) explains 88% with a five variable model applied to the UW generation in Nigeria.

The significance test (T and p value) shows that all explanatory variables included in the model are significant. VIF values are close to 1 and CI is 16, smaller than 30, the threshold value set (Belsley et al., 1980; Kleinbaum et al., 1998), indicating non collinearity problems. Anderson–Darling normality test' shows that is not sufficient evidence to suggest that the standardized residuals do not follow a normal distribution. Regression residuals plots show also that the regression model assumptions of homoscedasticity, linearity of the relationship between dependent and explanatory variables are met.

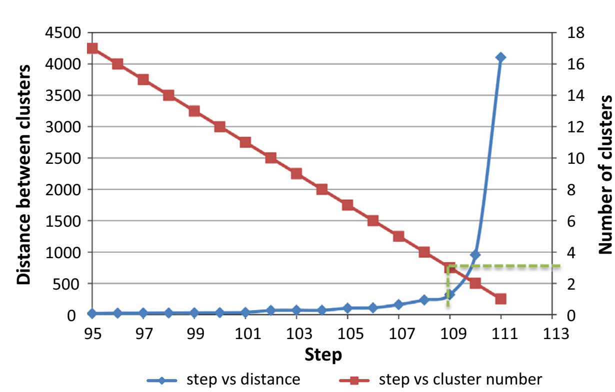


Fig. 6. Elbow-criterion for the decision making on the optimal cluster number for the 2010.

3.4. Factor models for the different clusters of municipalities identified

Although meeting regression model's assumptions, the model will be enhanced by grouping municipalities of Biscay with similar socio- economic characteristics into one. Hierarchical clustering is applied to group municipalities. Within all the potential variables those highly correlated with HW generation are chosen (Table 1). Clustering analysis is made regarding unemployment rate (UNEM) and density of retail outlets (RetD) because of their high degree of evidence shown concerning waste generation, as indicators of purchasing power of the citizens and potential to consume within the municipality. They are partially correlated with the dependent variable (Table 1) and within them ($r_s = 0.543$, significant at 0.001 level). Additionally, due to significant differences between municipalities regarding urban morphology (Fig. 1) this characteristic is included by the population density (denPOP).

The most significant "elbow" corresponds to 3 clusters (Fig. 6). The clusters are defined according to high (H), medium (M) or low (L) unemployment and commercial rate.

Fig. 7 sums up the cluster's characteristics for the year 2010. Cluster C_{HH} is formed by 6 municipalities, which comprises 52% of the population of Biscay. Municipalities in Cluster C_{HH} are featured by having *high* unemployment rate and *high* commercial activity (above the mean of Biscay). Moreover they are big towns featured by high population density. The municipalities of Cluster C_{LL} have *low* unemployment rate and *low* commercial activity, covering 23% of the total population. These municipalities are mainly little and disperse towns with low population density (Fig. 7). Municipalities from cluster C_{MH} (10 municipalities, 22% population) are featured by, with medium unemployment and high commercial activity.

From the various tested models, Eq. (7)–(9) shows the models selected as the best ones for each cluster. Table 2 the model's main statistics.

$$HW_{CHH} = 382 + 9.24\ HOCA - 10.19\ UNEM \quad (7)$$

$$HW_{CMH} = 211 + 26.85\ HOCA - 0.01\ denPOP \quad (8)$$

$$HW_{CLL} = 330 + 2.37\ HOCA + 2.01\ UNI - 0.03\ denPOP \quad (9)$$

While the model for cluster C_{LL} has little explicative capacity, the models deduced for cluster C_{HH} and C_{MH} have exceptional explicative capacity, describing up to 98.05% and 94.82% respectively (see Fig. 8). All the parameters included in the models are significant at the 5% error level and in all cases regression models assumptions are met. However, the explicative capacity of the model developed for the C_{LL} cluster is far away from the other two models. This might be due to the no inclusion in the study of the explicative variables behind the HW generation in that kind of municipalities.

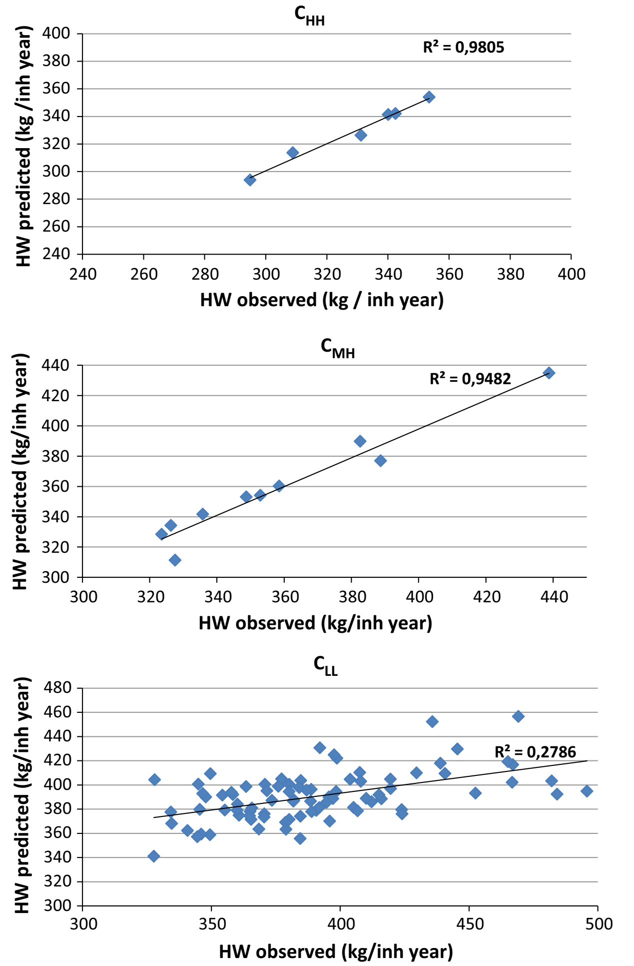
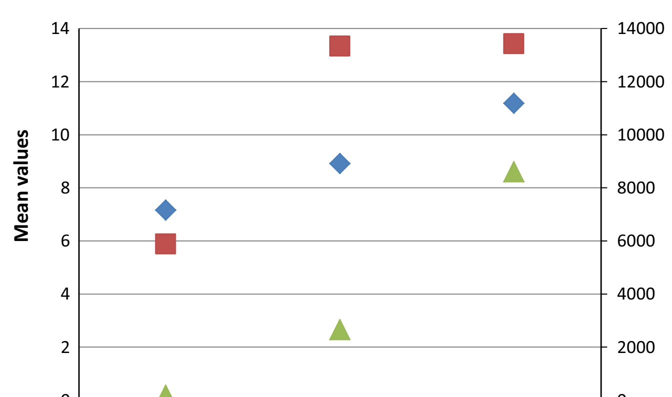


Fig. 8. Cluster models fit: C_{HH} (6 municipalities, 52% population), C_{MH} (10 municipalities, 21% population) and C_{LL} (84 municipalities, 23% population).



	CLL	CMH	CHH
UNEM (%)	7,16	8,91	11,18
RetD (‰ inh)	5,88	13,33	13,42
denPOP (inh/km2)	187,28	2658,68	8598,93

Fig. 7. The main characteristics (mean values) for the clusters on the basis of high (H), medium (M) or low (L) levels of UNEM, RetD and denPOP for the 2010. (UNEM and RetD left axis, denPOP right axis).

According to beta coefficients, the most significant explicative variables (the maximum absolute value of beta) in each model are different, being UNEM, HOCA and UNI respectively. This endorses the necessity of discrimination between municipalities.

3.5. Forecasting ability of each model

Despite having less explicative capacity in one of the models deduced, the $MAPE_t$ of 6.28 achieved with PM model has been improved until a $MAPE_t$ of 5.05 by aggregating each municipality to its model deduced. Additionally, it is analysed whether the models developed could be extrapolated in the time by validating each model with socio-economic data of 2011 in order to forecast 2013 HW generation ($MAPE_v$). In this case also, the models deduced after the clustering analysis give better results than the PM model improving the $MAPE_v$ of 8.93 until 6.15. However, it should be noticed that the forecasting ability of the models worsen a bit. This is due to the rapid change of the explicative variables value compared to the HW generation.

3.6. Meaningful explanatory variables

The identification of relevant explanatory variables is not straightforward, more in a very heterogeneous observation group as shown in the cluster analysis step. From the different models tested it is shown how the relevant variables are different. Nonetheless, the characteristic groups are quite similar.

Tourism has a huge positive impact on the economy of the municipalities and territories. However, the negative impact of tourism is the waste generation. Mateu-Sbert et al. (2013) and Ranieri et al. (2014) do not hesitate to say that one of the main impacts of tourism is waste generation. Lorena et al. (2013) estimate that a tourist can represent 0.3–0.6 kg of sorted HW per day. In this work tourism activities of municipalities are described by hotel and catering establishment present in the municipality and the spaces available for tourist accommodation, which are partially correlated ($r_s = 0.433$, significant at 0.001 level). This effect has also be studied by other authors in terms of the overnight rate in touristic establishments (Bach et al., 2004; Ibáñez et al., 2011) which would be really interesting for further research.

Urban morphology appears relevant in nearly all models deduced. It includes variables describing municipalities such as population density, dwelling density, municipal urban land and average surface of family dwelling all of then strongly correlated. These parameters present a decreasing effect on HW generation. This may be due to the quantification methodology used in order to assign waste generation to each municipality, which may be favouring large municipalities. Other studies have also reported its decreasing effect on HW generation describing its impact by household size (persons per household) (Bandara et al., 2007; Lebersorger and Beigl, 2011).

Not surprisingly, educational level appears to have significant importance on HW generation. UNI refers to population older than 10 that have completed university studies. However, while high educational level is expected to be related with high level of awareness on environmental issues, the model reflects the opposite (with a negative coefficient for this factor). This may be due to the fact that the variable is cumulative. That is, every year the variable increases with the new number of graduates, but the environmental awareness does not increase at the same rhythm.

Finally, the economic situation of inhabitants is crucial for HW generation. The unemployment rate (UNEM) is an indirect indicator of the economic situation of the municipalities. It refers to the number of unemployed people in the municipality. When the unemployment rate increases, available income of the families decreases ($r_s = -0.498$), and therefore HW generation. Obviously, the purchasing power of households with unemployed members is undermined, and thus the consumption. This result is consistent with other studies (Beigl et al., 2004; Keser et al., 2012).

In this study age structure and economic structure do not show significant effect on HW generation. However, other studies have shown that this aspects are critical (Afon and Okewole, 2007; Beigl et al., 2004).

4. Conclusions

Following the methodology described in this paper, different regression models were deduced in order to forecast HW per capita generation at municipality level regarding influential municipality characteristics. The methodology has been tested in the Province of Biscay that has 112 municipalities hosting 1.1 million of inhabitants. The study about the nature of the dependent variable (HW per capita) proves the existence of 12 municipalities with anomalous behaviour in the HW generation. While some of them have great tourism activity which may not be measure by the explanatory variables included in this work, other may be using different quantification system.

Due to the high heterogeneity between Biscay's municipalities hierarchical clustering process has been used. Three well differentiated clusters of municipalities have been deduced according to high, medium and low unemployment rate (UNEM), density of retail outlets (RetD) and population density (denPOP) (Fig. 7).

Altogether, four models have been developed, one for the overall group without anomalous municipalities (PM model), and one model per each cluster as shown in Table 2. The PM model has a descriptive capacity of 42.2%. The biggest cluster models, regarding hosted population, have significantly better descriptive capacity, with 98% and 94% for C_{HH} and C_{MH} clusters respectively. Nonetheless, the model developed for little and disperses municipalities, cluster C_{LL}, do have little explicative capacity, only 27%. Despite this, to relate each municipality with a specific model, regarding different socio-economic characteristics, improves the overall result. In this study this has been proved by the improved $MAPE_t$ and $MAPE_v$ get by the clusters models.

Thereby, urban morphology, tourism activity, educational level and economic dynamism and resources of population, have stood out over other characteristics showing significant effects in different models. While some of them describe the purchasing power of the families, other are related to the impact generated by tourism activities.

The methodology highlights the importance of having accurate data about both dependent and explanatory variables in order to deduce robust models for forecasting household waste generation. Relating municipalities with similar socio-economic characteristics allows inferring models with higher forecasting ability. At this point, the biggest challenge is being able to identify and include the influential explanatory information in the model.

In the same manner that there are significant differences between municipalities in terms of socio-economic characteristics, the same happens within big municipalities. The assessment of more deep levels of generation such as neighbourhoods or streets, will promote the optimization of service within municipalities.

Acknowledgements

The authors wish to thanks the Waste section from the Environment Department of the Provincial Council of Biscay, the different companies responsible for the management of the urban wastes, as well as the municipalities of the province of Biscay for supplying the waste related data necessary for accomplish this work.

References

Abbasi, M., Abduli, M.A., 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, 27–38.

Afon, A.O., Okewole, A., 2007. Estimating the quantity of solid waste generation in Oyo, Nigeria. Waste Manage. Res. 25, 371–379. http://dx.doi.org/10.1177/0734242X07078286.

Aranda Usón, A., Ferreira, G., Zambrana Vásquez, D., Zabalza Bribián, I., Llera Sastresa, E., 2013. Environmental-benefit analysis of two urban waste collection systems. Sci. Total Environ. 463–464, 72–77. http://dx.doi.org/10.1016/j.scitotenv.2013.05.053.

Armstrong, J.S., 2001. Selecting forecasting methods. In: Armstrong, J.S. (Ed.), Principles of Forecasting. A Handbook Fof Researches and Practitioners. Kluwer academis Publishers, pp. 365–386.

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. Recycl. 41, 65–73. http://dx.doi.org/10.1016/j.resconrec.2003.08.004.

Bandara, N.J.G.J., Hettiaratchi, J.P.a., Wirasinghe, S.C., Pilapiiya, S., 2007. Relation of waste generation and composition to socio-economic factors: a case study. Environ. Monit. Assess. 135, 31–39. http://dx.doi.org/10.1007/s10661-007-9705-3.

Basque Government, 2009. Convenio Marco entre la Comunidad Autónoma Vasca del País Vasco y el sistema integrado de gestión promovido por Ecoembalajes España, S.A.

Basque Government, 2010. Convenio entre la administración general del País Vasco (Departamento de Medio Ambiente, planificación territorial, agricultura y pesca) y la sociedad Ecológica para el reciclado de los envases de vidrio (ECOVIDRIO) para la gestión de envases usados y residuos de envases de vidrio en la CAPV.

Beigl, P., Wassermann, G., Schneider, F., Salhofer, S., Mackow, I., Mrowinski, P., Sebastian, M., 2003. LCA- IWM Report D2.1: Waste Generation Prognostic Model., The Use of Life Cycle Assessment Tool for the Development of Integrated Waste Management Strategies for Cities and Regions with Rapid Growing Economies LCA- IWM.

Beigl, P., Wassermann, G., Schneider, F., Salhofer, S., 2004. Forecasting Municipal Solid Waste Generation in Major European Cities. In: PahlWostl, C., Schmidt, S., Jakeman, T. (Eds.), IEMSs 2004 International Congress: "Complexity and Integrated Resources Management". Osnabrueck, Germany.

Beigl, P., Lebersorger, S., Salhofer, S., 2008. Modelling municipal solid waste generation: a review. Waste Manage. 28, 200–214. http://dx.doi.org/10.1016/j.wasman.2006.12.011.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Detecting and assessing collinearity. In: Wiley (Ed.), Regression Diagnostic: Identifying Influential Data and Sources of Collnearity. pp. 85–191.

Biscay Provincial Council, 2012a. Segunda revisión ordinaria del II Plan Integral de Gestión de Residuos Urbanos de Bizkaia 2005.

Biscay Provincial Council, 2012b. Observatorio permanente de residuos urbanos [WWW Document]. URL http://www.bizkaia.net/home2/Temas/DetalleTema.asp?Tem_Codigo=7709&idioma=CA&bnetmobile=0&dpto_biz=9&codpath_biz=9|351|7709.

Biscay Provincial Council, 2013. 5th container for organic matter [WWW Document]. URL http://www.bizkaia.net/home2/Temas/DetalleTema.asp?Tem_Codigo=8668&idioma=CA&bnetmobile=0&dpto_biz=9&codpath_biz=9|351|8668 (accessed 6.23.14).

Chang, N.-B., Lin, Y.T., 1997. An analysis of recycling impacts on solid waste generation by time series intervention modeling. Resour. Conserv. Recycl. 19, 165–186. http://dx.doi.org/10.1016/S0921-3449(96)01187-1.

Chen, H.W., Chang, N.-B., 2000. Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling. Resour. Conserv. Recycl. 29, 1–18. http://dx.doi.org/10.1016/S0921-3449(99)00052-X.

Coffey, M., Coad, A., 2010. Collection of Municipal Solid Waste in Developing Countries. Gutenberg Press, Malta.

Denafas, G., Ruzgas, T., Martuzevičius, D., Shmarin, S., Hoffmann, M., Mykhaylenko, V., Ogorodnik, S., Romanov, M., Neguliaeva, E., Chusov, A., Turkadze, T., Bochoidze, I., Ludwig, C., 2014. Seasonal variation of municipal solid waste generation and composition in four East European cities. Resour. Conserv. Recycl. 89, 22–30. http://dx.doi.org/10.1016/j.resconrec.2014.06.001.

Ecovidrio, 2013. Vidrio aportación ciudadana [WWW Document]. URL <http://www.ecovidrio.es/estadisticas_espana.aspx>.

EU, 1999. Council Directive 1999/31/EC of 26 April on the landfill of waste.

Eustat, 2013. Basque statistic office [WWW Document]. URL <www.eustat.es>.

Frost, J., 2012. Regression Smackdown: Stepwise versus Best Subsets! [WWW Document]. Minitab Blog. URL <http://blog.minitab.com/blog/adventures-in-statistics/regression-smackdown-stepwise-versus-best-subsets> (accessed 6.23.14).

Garbiker, 2013. Datos recogida del contenedor amarillo [WWW Document]. URL <http://garbiker.bizkaia.net/esp/ca_Cantid_Recogida_Municipios.asp?General=1&Pred_1=Municipios>.

Hervada-Sala, C., Jarauta-Bragulat, E., 2004. A program to perform Ward's clustering method on several regionalized variables. Comput. Geosci. 30, 881–886. http://dx.doi.org/10.1016/j.cageo.2004.07.003.

Ibáñez, M.V., Prades, M., Simó, A., 2011. Modelling municipal waste separation rates using generalized linear models and beta regression. Resour. Conserv. Recycl. 55, 1129–1138. http://dx.doi.org/10.1016/j.resconrec.2011.07.002.

Keser, S., Duzgun, S., Aksoy, A., 2012. Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey. Waste Manage. 32, 359–371. http://dx.doi.org/10.1016/j.wasman.2011.10.017.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A., 1998. Applied Regression Analysis and Other Multivariable Methods, Third Edit. ed.

Kumar, J.S., Subbaiah, K.V., Rao, P.V.V.P., 2011. Prediction of municipal solid waste with RBF net work- a case study of Eluru, A. P, India. IJIMT 2, 2–7.

Lebersorger, S., Beigl, P., 2011. Municipal solid waste generation in municipalities: quantifying impacts of household structure, commercial waste and domestic fuel. Waste Manage. 31, 1907–1915. http://dx.doi.org/10.1016/j.wasman.2011.05.016.

Li, Z., Fu, H., Qu, X., 2011. Estimating municipal solid waste generation by different activities and various resident groups: a case study of Beijing. Sci. Total Environ. 409, 4406–4414. http://dx.doi.org/10.1016/j.scitotenv.2011.07.018.

Lorena, A., Niza, S., Ferrão, P., 2013. Forecasting Municipal Solid Waste Collection and Benchmarking Infrastructure Provision. In: 7th International Conference of the International Society for Industrial Ecology. University of Ulsan June 25–28.

Lozano Valencia, M.A., Lozano Valencia, P.J., 2008. Las mancomunidades de servicios, un ejemplo de vertebración territorial para Guipúzcoa. Caracterización de los residuos domésticos en dicho territorio a partir de la gestión mancomunda. Bolentín AGE 48, 155–174.

Mallows, C.L., 1973. Some comments on Cp. Technometrics 15, 661–675.

Mateu-Sbert, J., Ricci-Cabello, I., Villalonga-Olives, E., Cabeza-Irigoyen, E., 2013. The impact of tourism on municipal solid waste generation: The case of Menorca Island (Spain). Waste Manage. 8–12. http://dx.doi.org/10.1016/j.wasman.2013.08.007.

MMA, 2001. Real Decreto 1481/2001, de 27 de diciembre, por el que se regula la eliminación de residuso mediante depósito en vertedero.

Mooi, E., Sarstedt, M., 2011. Cluster analysis, in: Springer, Berlin Heidelberg (Ed.), A Concise Guide to Market Research. The Process, Data, and Methods Using IBM SPSS Statistics. pp. 237–284. http://dx.doi.org/10.1007/978-3-642-12541-6.

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. http://dx.doi.org/10.1016/S0921-3449(02)00002-2.

Noori, R., Abdoli, M.A., Farokhnia, A., Abbasi, M., 2009. Results uncertainty of solid waste generation forecasting by hybrid of wavelet transform-ANFIS and wavelet transform-neural network. Expert Syst. Appl. 36, 9991–9999. http://dx.doi.org/10.1016/j.eswa.2008.12.035.

Ojeda-Benítez, S., Lozano-Olvera, G., Morelos, R.A., De Vega, C.A., 2008. Mathematical modeling to predict residential solid waste generation. Waste Manage. 28 (Suppl 1), S7–S13. http://dx.doi.org/10.1016/j.wasman.2008.03.020.

Pai, P.-F., Lin, K.-P., Lin, C.-S., Chang, P.-T., 2010. Time series forecasting by a seasonal support vector regression model. Expert Syst. Appl. 37, 4261–4265. http://dx.doi.org/10.1016/j.eswa.2009.11.076.

Purcell, M., Magette, W.L., 2009. Prediction of household and commercial BMW generation according to socio-economic and other factors for the Dublin region. Waste Manage. 29, 1237–1250. http://dx.doi.org/10.1016/j.wasman.2008.10.011.

Ranieri, E., Rada, E.C., Ragazzi, M., Masi, S., Montanaro, C., 2014. Critical analysis of the integration of residual municipal solid waste incineration and selective collection in two Italian tourist areas. Waste Manage. Res. http://dx.doi.org/10.1177/0734242X14533605.

Rimaityte, I., Ruzgas, T., Denafas, G., Racys, V., Martuzevicius, D., 2012. Application and evaluation of forecasting methods for municipal solid waste generation in an Eastern-European city. Waste Manage. Res. 30, 89–98. http://dx.doi.org/10.1177/0734242X10396754.

Shan, C.S., 2010. Projecting municipal solid waste: the case of Hong Kong SAR. Resour. Conserv. Recycl. 54, 759–768. http://dx.doi.org/10.1016/j.resconrec.2009.11.012.

Siniksaran, E., 2008. A geometric interpretation of Mallows' Cp statistic and an alternative plot in variable selection. Comput. Stat. Data Anal. 52, 3459–3467. http://dx.doi.org/10.1016/j.csda.2007.10.023.

Udalmap (2013) Municipal-Level Indicators of Sustainability [WWW Document], n.d. URL <http://www.ogasun.ejgv.euskadi.net/r51-udalmap/es/contenidos/informacion/udalmap/es_udalmap/udalmap.html> (accessed 10.8.13).

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 Manage. 33, 1324–1331. http://dx.doi.org/10.1016/j.wasman.2013.02.012.

Zaman, A.U., Lehmann, S., 2013. The zero waste index: a performance measurement tool for waste management systems in a "zero waste city". J. Clean. Prod. 50, 123–132. http://dx.doi.org/10.1016/j.jclepro.2012.11.041.