Journal of Cleaner Production 210 (2019) 401–409

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Artificial neural network modelling of the amount of separately-collected household packaging waste

Verónica Oliveira [a, b, *], Vitor Sousa [c], Celia Dias-Ferreira [a, b]

[a] Research Centre for Natural Resources, Environment and Society (CERNAS), College of Agriculture (ESAC), Polytechnic Institute of Coimbra, Bencanta, 3045-601 Coimbra, Portugal

[b] Materials and Ceramic Engineering Department, CICECO, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

[c] CERIS, Department of Civil Engineering, Architecture and GeoResources, Tecnico Lisboa - IST, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

A R T I C L E I N F O

Article history:
Received 5 July 2018
Received in revised form
3 November 2018
Accepted 6 November 2018
Available online 8 November 2018

Keywords:
Municipal solid waste
ANN
Genetic algorithm
Regression model
Urban waste
Recycling

A B S T R A C T

This work develops an artificial neural network (ANN) model using genetic algorithms to estimate the annual amount (kg/inhabitant/year) of separately-collected household packaging waste. The ANN model comprises one input layer, one hidden layer with seven neurons and one output layer. Ten variables affecting the amount of separately-collected packaging waste were identified and used in the ANN model. These variables are related to the level of education of the population, the size and level of urbanisation of the municipality, social aspects related to poverty and economic power and factors intrinsic to the waste collection service. A comparison between ANN and regression models for the estimation of packaging waste is also carried out. The performance of the proposed ANN model for a data set of 42 municipalities located in the centre of Portugal, measured by the R^2, is 0.98. This value is 34% higher than the best regression model applied to the same data set ($R^2 = 0.73$), indicating that ANN has a significantly higher explanatory power than traditional regression techniques. Another advantage is that ANN is not as sensitive to outliers as regression. However, ANN is more complex, has a higher number of variables, and the model development and interpretation of the results are more difficult. Nevertheless, the higher performance of ANN makes it a valuable tool in the definition of strategies to increase recycling and achieve circular economy goals.

1. Introduction

The circular economy package presented by the European Commission set targets for the re-use and recycling of 65% of municipal waste and for the recycling of 75% of packaging waste by 2030 (EPRS, 2016). A key aspect in reaching these targets is the ability to separate the waste into fractions to be sent for recycling. The Waste Framework Directive (European Parliament and of the Council, 2008) established that Member States shall set up separate collection schemes in order to promote high quality recycling of at least glass, paper, plastic and metal by 2015. As a result, separate collection of municipal waste is nowadays widely implemented in Europe. However, the percentage of the total waste that is separately collected varies significantly between Member States and within the various regions of each Member State. According to a report by the BiPRO/CRI (2015), an average of 19% of municipal waste is separately collected in 28 European capitals. The separate collection rates vary significantly amongst the 28 cities, with some cities achieving high rates (Ljubljana 55%; Tallinn 47%; Helsinki 39%; Dublin 37%; Vienna 29%), while others have lower performances (Zagreb 1.0%; Bucharest 2.9%; Sofia 4.0%; Warsaw 4.5%). This means that large percentages of recyclable materials, mostly packaging, are still discarded as unsorted waste, instead of separately collected. The consequences of a long-term non-appropriate management of the packaging materials are devastating for society and environmental sustainability (e.g. loss of valuable resources, increase of landfill space, scarcity of primary resources, overall higher costs with collection and treatment). The increase of separate collection of packaging waste enables the recovery and recycling of clean and high-value materials. This will reduce the consumption of raw materials in the manufacturing of new packaging products and promote a cleaner production, contributing to

* Corresponding author. Research Centre for Natural Resources, Environment and Society (CERNAS), College of Agriculture (ESAC), Polytechnic Institute of Coimbra, Bencanta, 3045-601 Coimbra, Portugal.
E-mail address: veronica.oliveira@esac.pt (V. Oliveira).

https://doi.org/10.1016/j.jclepro.2018.11.063
0959-6526/

achieving the "responsible production and consumption" goal set by the United Nations (United Nations, 2015). In addition, it will also enable the implementation of circular economy goals. A better understanding of the factors influencing the generation of packaging waste separately collected by the population is fundamental to increase recovery rates.

Several techniques have been used to model the municipal solid waste (MSW) generation, namely (Abbasi and Hanandeh, 2016): i) descriptive statistics; ii) regression analysis; iii) material flow models; iv) time series analysis and; v) artificial intelligence models. Artificial neural network (ANN) is an artificial intelligence tool widely used to model environmental complex systems. For instance, the amount of electricity generated and emissions from incineration and landfilling of MSW (Nabavi-Pelesaraei et al., 2017), or the determination of the lower heating value of MSW (Dong et al., 2003) or even the gasification characteristics of MSW (Xiao et al., 2009). ANN stands out for its ability to model increasingly nonlinear complex realities for which no adequate physical or mechanistic model exist.

There have been several researchers using ANN to explore the relation between social, economic and environmental factors and MSW production at local, regional, national and international scales. At a city wide scale, Sun and Chungpaibulpatana (2017) found that MSW generation in Bangkok was dependent on the *total number of residents*, *native people aged 15 to 59 years*, *total people aged 15 to 59 years*, *number of households*, *income per household*, and *number of tourists*. In another study, the long-term prediction of solid waste generation in the city of Mashhad was modelled using *population*, *household income* and *maximum temperature* as explanatory variables (Abdoli et al., 2012). At a regional scale, Kannangara et al. (2018) developed a model for the prediction of MSW generation ($R^2 = 0.72$) using as variables *the fraction of population over 45 years*, *median personal income* and *fraction of owned dwellings* for 220 municipalities in the region of Ontario, Canada. At a national scale, Chhay et al. (2018) reported *urban population growth* as the most influencing variable on MSW generation model ($R^2 = 0.931$) in China. The annual MSW generation in Malaysia was accurately predicted ($R^2 = 0.97$) using the *gross domestic product*, *population* and *employment* as inputs to an ANN model (Younes et al., 2015). Ordonez-Ponce et al. (2006) determined the *population*, *percentage of urban population*, *years of education*, *number of libraries*, and *number of indigents* as the most important variables affecting MSW generation in Chile ($R^2 = 0.819$). The prediction of MSW generation at an international scale (26 European countries) was studied by Antanasijevie et al. (2013) considering the *gross domestic product*, *domestic material consumption* and *resource productivity* as inputs in two ANN models with different architectures ($R^2 = 0.930$ and $R^2 = 0.981$). The model developed by Adamovi et al. (2017) for 44 countries used as variables the *urban population*, *population density*, *average household size*, *industry* (value added), *tourism expenditure in the country*, *population by age group 20–65*, *unemployment rates*, *alcohol final consumption expenditure* and *CO_2 emissions from residential buildings and companies and public services* ($R^2 = 0.96$). Some authors considered a waste collection utility scale and account for the waste collection service characteristics. Shamshiry et al. (2014) applied ANN to model the amount of solid waste weekly generated in Langkawi island in Malaysia ($R^2 = 0.98$), using as contributing variables *type of trucks* and their trips, *number of personnel in per trips* and *fuel cost*. Azadi and Karimi-Jashni (2016) mixed socio-economic factors with waste collection service characteristics, using the variables *population*, *solid waste collection frequency*, *maximum seasonal temperature* and *altitude* in an ANN model ($R^2 = 0.86$) to predict seasonal MSW generation for 20 cities located in Fars Province, Iran. The above literature shows that MSW generation is clearly affected by a wide range of factors which are different according to the scale (e.g., municipality, region or country) and scope (e.g., weekly, seasonal, annual) of the application of the model. Each municipality, region or country has its own characteristics and ANN models developed for a specific region may not be applicable to others.

The prediction of packaging waste generation at waste collection service scale was addressed by Ferreira et al. (2014). The ANN model proposed ($R^2 = 0.672$) mixing socio-economic (population density, number of non-residential buildings) and waste collection service factors (number and type of bring-banks). In this study, the focus was waste production not on the factors influencing it. In another work (Oliveira et al., 2018), regression models were used to explain the relationships between socio-economic/waste collection service factors and the amount of separately-collected packaging waste. However only 73% of the variability could be explained. This means ordinary least squares regression is limited for determining the variables affecting the amount of separately-collected packaging waste.

The objective of the current study is to develop a model to predict the amount of separately-collected packaging waste that surpasses the performance of the previous models using ANN. The current work also aims to discuss the potential advantages of ANN over the conventional regression regarding the following parameters: i) accuracy and extrapolation; ii) handling of outliers and selection of relevant variables; and iii) interpreting the results and using the model.

2. Artificial neural network

Artificial neural network is an artificial intelligence tool that is able to identify patterns in data and "learn" from them. ANN mimics the brain's neural connections, where electrical impulses run through neurons and synaptic connections, as a response to the inputs perceived by sensory organs: vision, olfaction, audition, etc. On average, in a human brain there are 100 billion neurons, each with 1000–10000 connections with other neurons (Agatonovic-Kustrin and Beresfold, 2000). ANNs are a simplification of the brain's highly complex network of interconnecting neurons. The structure of a feed-forward multilayer ANN is formed by multiple layers of neurons, which are the individual processing units, receiving information, processing it and passing it forward to the following layer (Fig. 1a).

In a feed-forward multilayer ANN the neurons of a layer are connected to all the neurons in the next layer by links with different weights, similarly to the synaptic connections in a human brain. There are three types of neuron layers: i) the input layer, where the neurons represent the sensory receivers and the external information is fed into the system; ii) the hidden layers, which mimic the biological neural network the information travels in the brain; and iii) the output layer represents the decision output.

However, a feed-forward multilayer ANN encompasses a series of simplifications, when compared to the Human brain. First, there could exist multiple hidden layers, but in majority of the cases a single layer is used. The use of additional hidden layers doesn't improve the accuracy of results for small to medium samples and not very large number of input variables. Second, the neurons are only connected between layers, with no connections amongst the neurons in the same layer. Third, the transformation functions of all the neurons in a layer are the same. In theory, it would be possible to define different transformation functions for each neuron and use any transformation possible, but that would make the network mathematically difficult to build. Forth, there is a limited set of alternative transformation functions for the neurons. This limitation is mathematical, since the transformation functions need to compatible with the training algorithms. Fifth, the signal only

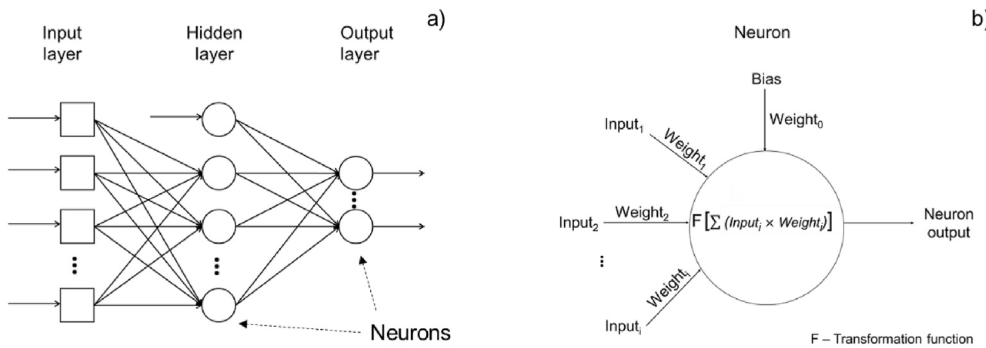

Fig. 1. a) Structure of a feedforward artificial neural network and b) processing model of the individual neuron.

travels one-way, from the input layer to the output layer.

All the neurons in a given layer are linked to all the neurons on the previous and on the following layer (except for the first and last layers). This allows to account for possible interactions between the independent variables. Fig. 1b represents the way a neuron processes information. Activation of a neuron is simply the weighted sum of the separate inputs from each neuron in the previous layer, plus a bias (or offset term). The activation of a neuron can be positive (an excitation) or negative (an inhibition). The neuron produces a response after being activated, which is modelled by a transformation function. The transformation function can be a hard-limiting threshold function (e.g. a sign function), a linear or semi-linear function (e.g. identity, exponential), or a smoothly limiting threshold (e.g. an s-shaped function such as sigmoid or hyperbolic tangent). If the transformation function is the identity function, this model resembles a multiple linear regression with interaction. The neural network must resort to an iterative process to account for the interaction, which makes it intrinsically different from multiple linear regression. The multiple linear regression has a closed mathematical solution for its regression coefficients, whereas the ANN do not by virtue of its multiple layers and potential combinations between the inputs.

The two main challenges when developing an ANN are: i) selecting the network structure (input variables, number of hidden layers and number of neurons in each layer); and (ii) defining the network parameters (the transformation functions in the hidden and output layers and the connection weights). ANN have no embedded method to address these challenges, except for the connection weights. The initial values for the connection weights are randomly assigned at the beginning and then re-calculated to minimize the error between the output of the ANN (predicted value) and the known values. This is usually done resorting to the backpropagation algorithm. The backpropagation algorithm redistributes the error (observed at the output) by the connections, enabling estimating the gradient (direction in which errors are reduced), combined with an algorithm to update the connection weights that minimizes an error function (e.g. sum of the squared errors). Their high flexibility makes ANNs prone to overfitting, resulting in high explanatory capability within the observed results used to develop the model, but low for new cases. To address this issue, the data is split into training and testing sets to enable its evaluation with cases that were not used in the development. The other challenges have to be approached separately. For instance, if for regression models algorithms such as the forward stepwise assist on selecting the parameters, there are no equivalent techniques for ANN. One possibility is to test all possible combinations of network structures and network parameters. This rapidly becomes computationally intensive when the number of explanatory variables increases and there is no warranty that the best or a good solution was chosen. The same applies to the number of neurons in the hidden layer and the transformation functions.

3. Development of the ANN to model the amount of separately-collected packaging waste

3.1. Initial data set

The selection of variables that influence the amount of separately-collected packaging waste is essential to build an accurate model. Previously, 14 independent variables were pinpointed as relevant (Oliveira et al., 2018). These variables fall into two groups: i) socio-economic/demographic; and ii) those related to the waste collection service. In the current work, the same explanatory variables were considered as a starting point for building the ANN model, as listed in Table 1. More detailed explanation on the meaning of each variable and corresponding calculation formula can be found in Oliveira et al. (2018).

The initial data set was built with the values of each explanatory variable in 42 municipalities (in this case, an area of 8848 km^2 in the centre of Portugal). The registered (known) values of the amount of packaging waste (kg/inhabitant/year) were retrieved from official databases and reports (ERSAR, 2015; ERSUC, 2016; Statistics Portugal, 2016, 2014a; 2014b, 2013a; 2013b, 2011a; 2011b, 2007; VALORLIS, 2016). The full dataset comprises the year 2015 and is included as supplementary material to this work (Table S1).

3.2. Development of the ANN

In the present work a fully connected, feed-forward ANN comprising three layers (input, hidden and output) is used. The inputs were normalized for the interval [-1, 1] to ensure scale

Table 1
Set of explanatory variables used in the development of the ANN model.

Variable	Unit
Population	inhabitants
Area	km^2
Degree of urbanisation	%
Purchase power per capita	%
Purchase power index	–
Deprivation index	–
Population over 65 years old	%
Population over 15 with first education cycle or less	%
Number of school years attended	–
Bring-banks per area	number/km^2
Inhabitants per bring-brank	–
Accessibility to separate collection services	%
Relative accessibility to bring-banks	–
Civic amenity drop-off sites per area	number/km^2

similarity. The transformation functions were set to logistic in the hidden layer and identity in the output layer. Other functions (e.g. hyperbolic tangent) were tested, but as long as non-linear function was used in the hidden layer the performance was similar. The backpropagation algorithm was combined with gradient descent for estimating the connection weights. Other algorithms were also tested (e.g. Levenberg-Marquardt), achieving similar results.

The best set of explanatory variables and the best parameters of the ANN were selected using genetic algorithms. In addition, the cross validation process was used to minimize the bias due to dataset split and initial connection weights. Similarly to an ANN, the genetic algorithm (GA) also mimics a biological process, this case the process of natural selection. In nature, the process of natural selection results that only the fittest survive and get to reproduce. Each new generation combines the genetic features of its predecessors along with natural random mutations to better adjust to the environment. The general idea underlying the method is generating a pseudo random initial set of solutions (initial population) and then carry on mixing the best ones, eventually introducing random changes (mutations), to search for a better solution (higher accuracy).

The selection of the most adequate set of explanatory variables to use in a model is a complex challenge in any regression or classification problem. Whereas for ordinary linear regression approaches such as stepwise, forward or backward are available to aid in the selection of the most suitable set of variable to use in the model, the same does not exist for ANN. One option is to use a “brute force” approach, which implies testing all possible combinations of the potential variables. This option rapidly becomes very computationally demanding or even impractical in terms of the computational time required. Another option is to use optimization algorithms that enable this task to be carried out with less computational effort. The nature of the optimization algorithms will depend on the model to be developed. For instance, linear or non-linear programming problems can be optimized with algorithms that attain an absolute optimum solution (e.g., simplex). On the other hand, for highly non-linear models such as ANN the optimization algorithms available may not provide the absolute optimum solution. Genetic algorithms are tools that fall into the last category and was used to select the number and the variables to use in the ANN model (and tune the ANN parameters also). It was considered that the number of variables in the model could vary between 4 and 14 (inclusion of all variables), a population size of 5 and maximum of 200 generations. An early stop criterion of 20 generations without improvement was also used. The selection scheme used was the tournament, with a size of 0.25. The probability for mutation was set to 1/n, where n is the total number of variables. The initial set of solutions may be totally random or may be selected using the set of variables that, individually, present the highest correlation with the output (option used herein).

In addition to the variables selection, a similar GA approach was also used to optimize ANN parameters (number of neurons in the hidden layer, number of training cycles, learning rate, momentum and error epsilon). The number of neurons was set to vary between (number of attribute + number of classes)/2−3 and (number of attribute + number of classes)/2 + 3 (with a minimum of 4 since the regression model had 5 variables). The number of training cycles ranged between 500 and 5000 and the learning rate between 0.05 and 1. The momentum varied between 0.05 and 0.5 and the error epsilon between 10^{-7} and 10^{-3}.

Table 2 presents the features and respective value ranges considered in the genetic algorithms.

Cross-validation was used in order to minimize the influence from how the data is split and the initial values of the connection weights in the ANN performance. To perform cross-validation, the dataset was divided into 5 subsamples. For each run, 4 subsamples were used for training the network and the remaining subsample for testing. The process was repeated until all 5 subsamples were used once for testing. The neural networks performance was evaluated by the average performance of the 5 ANN models. The best ANN model was selected using different data splits and initial connection weights.

Table 2
Structure and parameters used in genetic algorithms.

Input/ANN structure and parameters	Min	Max
Number of inputs	4	14
Number of neurons in the hidden layer	3	10
Learning rate	0.05	1
Momentum	0.05	0.5
Training cycles	500	5000
Error	1E-7	1E-3

The process of developing of ANN model is illustrated in Fig. 2. It consists in: i) selection of a set of explanatory variables; and ii) optimize the ANN parameters for each set of explanatory variables using cross validation for each solution of parameters. Repeat the process until the best or better ANN model is found.

The first genetic algorithm (GA1) starts by generating a sample of pseudo-random possible solutions of inputs. For each of those solutions, the second genetic algorithm (GA2) generates a sample of random parameters and builds the ANN, applying the cross-validation process to each set of parameters. From the ANNs tested, a selection is made of those presenting the best performances and their parameters are mixed to create the new generation ANN. Mutations are simulated by random changes to the network parameters from the ANN models being mixed. The process is repeated to optimize the network parameters for a given set of inputs and then for each set of inputs to find the best

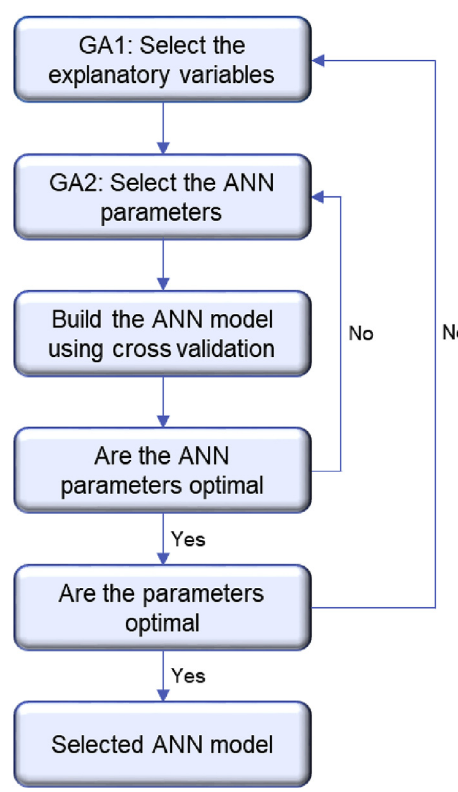

Fig. 2. Process used to develop the ANN model.

combination of inputs and parameters. This enables a search for possible solutions driven by maximum performance. The process could be repeated indefinitely. In the current work, the limit to the number of generations was set to 200 for each GA, and the limit of number of generations without performance improvement was 20. These limits were set by a trial and error process with the goal of ensuring that the solution was independent. It should be noticed that a GA does not ensure that the overall best ANN model is found, but assists in finding a good ANN model with less computational effort. In the limit, and considering cross-validation, GA1(Max Generations = 200 × Population size = 5) x GA2(Max Generations = 200 × Population size = 5) x Cross-Validation (number of folds = 5) ANN models could have been tested during the development of the ANN model.

3.3. Evaluation of model performance

Several accuracy measures can be used to evaluate the performances of predictive models (Abdoli et al., 2012; Azadi and Karimi-Jashni, 2016). These accuracy measures represent the difference between the known and the predicted value obtained by the model. In the current work, the prediction performance of the proposed models is evaluated through the correlation coefficient (R^2) that is calculated according to the following equation:

$$R^2 = \left[\frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} \right]^2$$

where x is the known value for the amount of packaging waste, y is the value of the amount of packaging waste predicted by the model and $\bar{x}$ and $\bar{y}$ are the known and predicted means of amount of packaging waste, respectively. The best predictive model will have an R^2 closest to one. This accuracy measure provides an easy and intuitive way of judging the correlation between predicted and known values (Azadi and Karimi-Jashni, 2016).

4. Results and discussion

4.1. Artificial neural network model developed

An ANN was developed to model the impact of 14 variables (listed in Table 1) on the amount of separately-collected packaging waste in 42 municipalities. The fit between the values predicted by the best neural network and the values known for the dataset is shown in Fig. 3. The ANN model has a correlation (R^2) of 0.98 between predicted and known values. Therefore, the ANN can very accurately predict the amount of packaging waste that is separately collected for the dataset.

The structure of the best ANN model obtained is shown in Fig. 4. Ten variables (out of the initial 14) were selected by the GA as inputs to the model. The remaining 4 variables (*population*, *population over 65 years old*, *purchase power* per capita and *bring-banks per area*) were excluded because they did not significantly influence the amount of packaging waste that is separately collected.

The sensitivity analysis showed that fitness of the model did not significantly improve when more than 7 neurons (+the bias) were used in the hidden layer. Networks with 1 neuron in the hidden network consider input variables to be independent and resemble an ordinary least squares regression model. In the current ANN, with 7 neurons in the hidden layer, possible combinations of input variables are considered.

As described in section 2, the network is trained to estimate the value of the output variable (amount of separately-collected

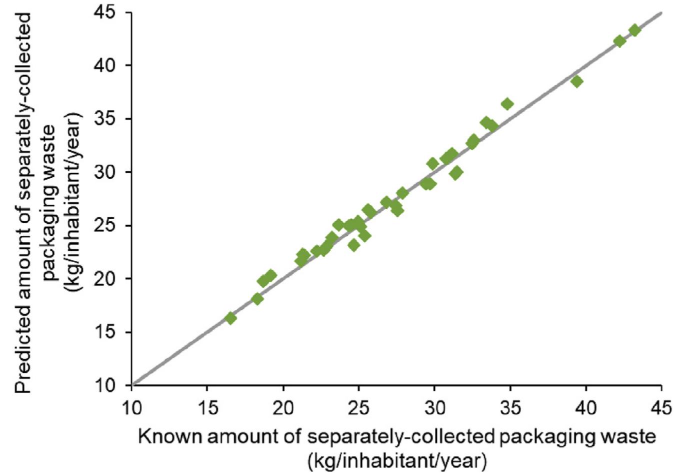

Fig. 3. Scatter plot of known versus predicted amount of separately-collected packaging waste by ANN ($R^2 = 0.98$).

packaging waste) as a function of input variables. The calculation uses the connection weights between each neuron, which are calculated by the network during the training process. The connections weights reflect the contribution of each input variable to the output. The final values obtained for the connection weights at the end of training process are shown in Table 3.

All the explanatory variables present both positive and negative connection weights, depending on the neuron on the hidden layer they are connecting to. However, two variables, *inhabitants per bring-bank* and *degree of urbanisation*, present mostly negative connection weights, while the remaining variables present mostly positive values. This is an indication of how the former mostly contribute to increase the amount of packaging waste that is separately collected and the latter to its decrease, considering that the connection weights with the output neuron are mostly negative. However, there are cancelation effects between variables, since some weights have negative signs and others are positive. The sign of the weight *per se* is not enough to show the effect of an input variable. The interaction among variables results in some cases in one specific input increasing the amount of packaging waste that is separately collected and in other situations decreasing it.

The relative connection weights of each input variable are represented in Fig. 5, with the values normalized for the interval [0, 1] to facilitate the analysis. The variable *relative accessibility to bring-banks* is the one with the highest influence on the amount of packaging waste that is separately collected (0.15) followed by the variables *number of school years attended* (0.13) and *inhabitants per bring-bank* (0.12). The variables with less influence on the amount of packaging waste that is separately collected are related to the economic levels and also to waste collection services.

4.2. Comparison of ANN and regression techniques for modelling the amount of packaging waste that is separately collected

The comparison between the ANN technique and the regression technique comprises the following topics: i) accuracy and extrapolation; ii) handling of outliers and selection of relevant variables; and iii) interpreting the results and using the model.

4.2.1. Accuracy and extrapolation

Fig. 6 compares the performances of the ANN model developed in the current work and a regression model that uses the same dataset and the same input variables (Oliveira et al., 2018). This

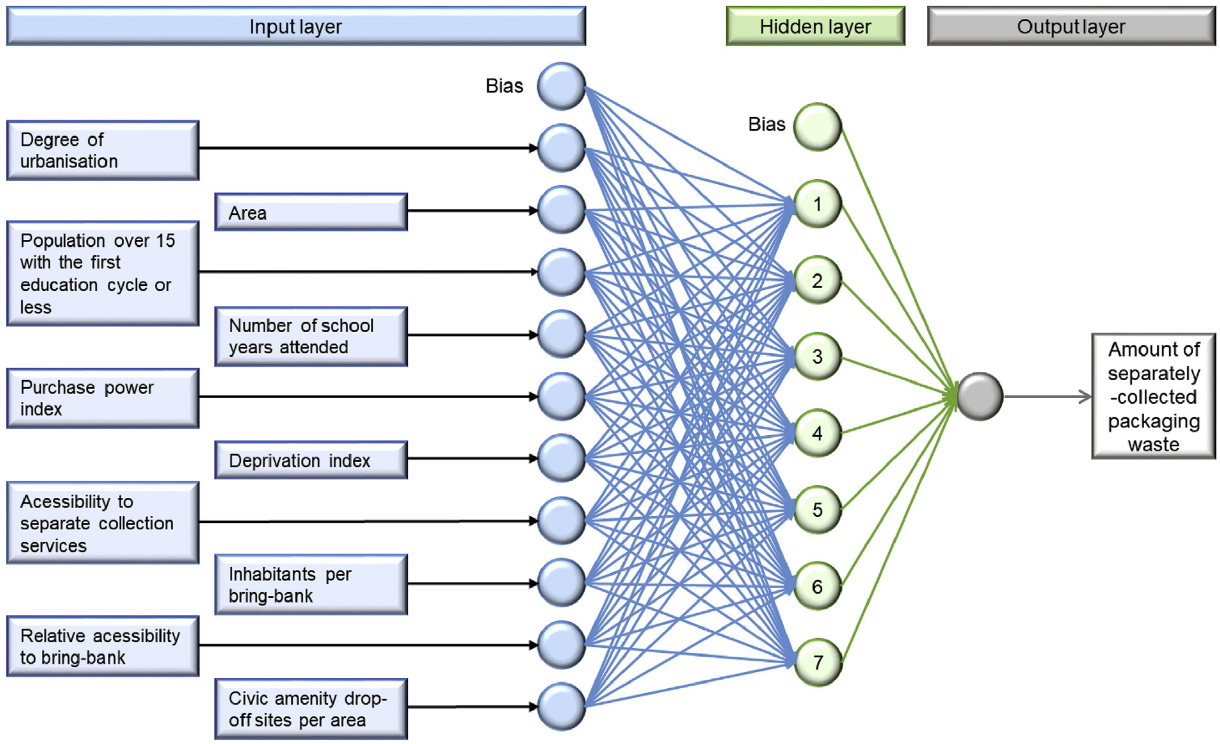

Fig. 4. ANN model for the prediction of amount of household packaging waste that is separately collected (kg/inhabitant/year).

Table 3
Connection weights between the neurons for ANN model for prediction of amount of packaging waste that is separately collected.

Variable	Hidden layer						
	Neuron1	Neuron 2	Neuron 3	Neuron 4	Neuron 5	Neuron 6	Neuron 7
Degree of urbanisation	−0.255	−0.063	3.02	−0.662	−0.639	−0.012	−1.838
Area	1.268	0.288	0.042	−3.109	1.596	0.3	0.577
Population over 15 with first education cycle or less	−3.047	0.228	1.228	1.615	0.977	0.264	−0.339
Number of school years attended	1.128	0.021	0.847	−3.016	1.298	0.012	2.224
Purchase power index	1.809	0.349	0.04	−2.262	0.385	0.331	0.846
Deprivation index	0.857	0.115	−0.691	−2.052	0.469	0.077	0.133
Accessibility to separate collection services	0.962	0.106	0.354	1.273	0.004	0.058	−1.385
Inhabitants per bring-bank	3.142	−0.011	−0.431	−1.606	−0.598	−0.072	−1.947
Relative accessibility to bring-banks	1.231	0.352	−0.859	0.998	−1.455	0.301	4.565
Civic amenity drop-off sites per area	−1.208	0.684	−0.516	1.047	1.559	0.735	0.366
Bias	−0.303	−0.956	−1.407	−1.641	−0.228	−0.992	0.972
To output layer (threshold = 0.355)	*−2.476*	*−0.068*	*1.859*	*−2.56*	*2.067*	*−0.048*	2.806

scatter plot presents the known value of the amount of packaging waste that is separately collected against the values estimated both by the ANN and by the regression model. The relationship between the actual and predicted values is higher for the ANN model ($R^2 = 0.98$) than for the multiple non-linear regression model ($R^2 = 0.73$). This means that the ANN model has a significantly higher explanatory power. The performance of the proposed ANN model was in the upper range of several works predicting MSW generation using ANN technique (0.72–0.98). Furthermore, the inclusion of a wide range of demographic and socio-economic factors affecting the production of packaging waste resulted in a more reliable predictive model for this waste flux, when compared to the model proposed by Ferreira et al. (2014). The explanatory power of the proposed ANN model is about 46% higher than the value found in Ferreira et al. (2014) ($R^2 = 0.672$).

The accuracy of the ANN, measured by its fit to the known data, has to be balanced with its generalization capability. This is done to some extent implicitly by using training, testing and, eventually, validation datasets. Herein a comparison of the extrapolation estimates of the ANN models with multiple non-linear regression model was performed by selecting 6 random municipalities. The *inhabitants per bring-bank* and *relative accessibility to bring-banks* were varied by +10% and −10%, one at a time, creating a hypothetical dataset of 24 cases. For this dataset, the amount of packaging waste that is separately collected was calculated using the regression and ANN models. It is possible to observe in Fig. 7 that the predicted amount of packaging waste that is separately collected from both tools (ANN and regression) are similar and within the margin of error that differentiates the models in the original dataset. This indicates that the ANN model extrapolation is similar to the regression model extrapolation to some extent. If this was not the case, and the ANN model was overfitted, the results from the regression and ANN models for new unseen data would tend to diverge. Overfitting occurs when the ANN captures very well the cases used for the development of ANN, but performs poorly in the estimation of the response for new cases. When

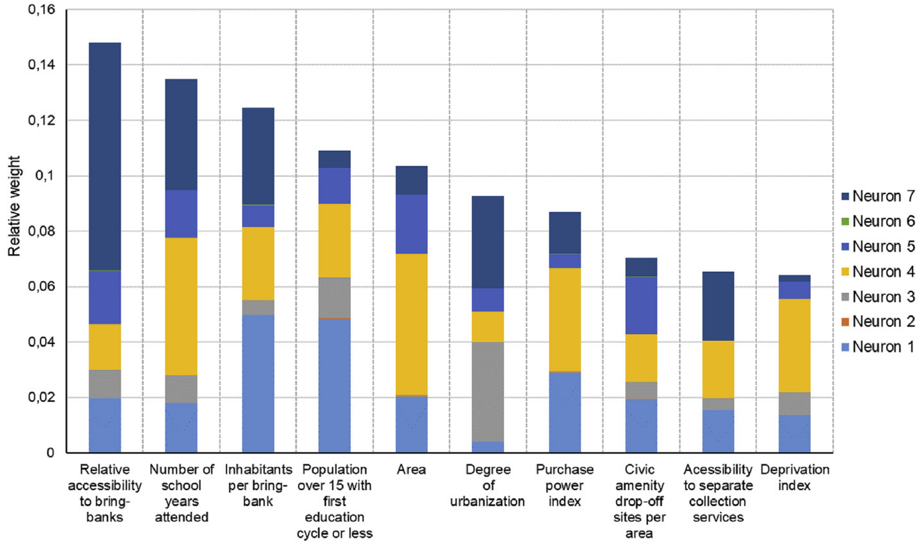

Fig. 5. Relative connection weights for each input variable (normalized for the interval [0, 1]).

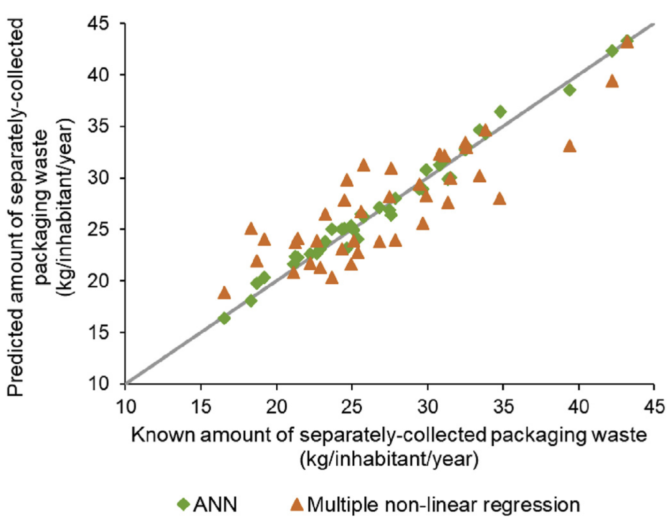

Fig. 6. Scatter plot of known versus predicted values of the amount of packaging waste that is separately collected by ANN ($R^2 = 0.98$) and by multiple non-linear regression model ($R^2 = 0.73$).

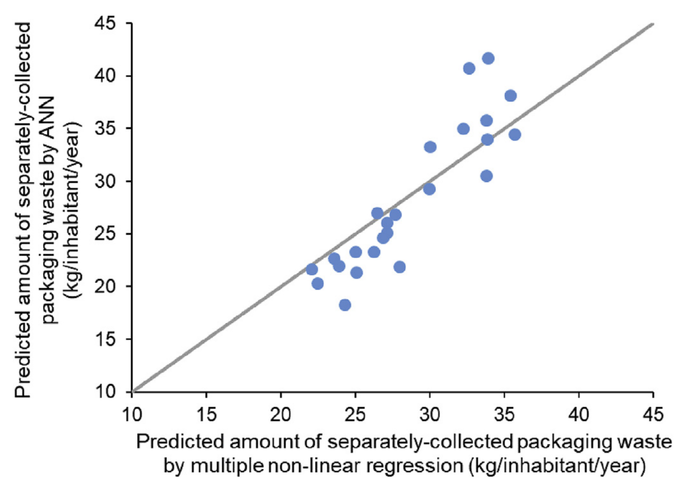

Fig. 7. Scatter plot of predicted amount of separately-collected packaging waste by multiple non-linear regression model (without outliers; $R^2 = 0.73$ - in Oliveira et al. (2018)) versus ANN model.

extrapolating, the response values are unknown, and cannot be compared with known values to assess the performance of the estimation. Since similar extrapolation results were obtained between ANN and regression, there is no indication of that overfitting is occurring in the ANN model.

4.2.2. Handling of outliers and selection of relevant variables

The initial dataset consisted of a double entry table, in which the municipalities are the lines (42 entries) and the variables are the columns (15 entries). This results in an initial dataset consisting of 630 data points (42 lines x 15 columns). Even though both models (regression and ANN) start with the same initial dataset, the final dataset used in each model consisted of a different subset of the initial one. This is due to the handling of outliers and to the selection of relevant variables being different within each modelling technique.

In the regression technique, Cook's distance criterion was adopted to identify data outliers (Oliveira et al., 2018). Outliers have the tendency to deviate the least squares fit in their direction by receiving more weight. In the process of minimizing the residual errors, the least squares estimates of the regression coefficients are distorted by this deviation (Huber, 2011). Following the application of the Cook's distance criterion, three municipalities out of the initial 42 were previously identified (Oliveira et al., 2018) as statistical outliers. These were manually removed from the dataset used for the regression model, thus improving the accuracy of the regression model from $R^2 = 0.65$ to $R^2 = 0.73$. This represented an improvement of 12% of the explanatory power of the regression model.

ANN is not as sensitive to outliers as the regression technique, because of its higher fitting capability. It is not necessary to exclude outliers when ANN is used, but these may contribute for overfitting when the network tries to correctly estimate their value. Furthermore, the influence of outliers will depend on how they are distributed by the training or testing datasets. When similar performance and weights are obtained in ANN models developed from datasets with and without outliers, it becomes complicated to conclude anything regarding their influence. The variability due to the data split and initial random initialization remains, even using techniques such as cross-validation, making it difficult to conclude about the effect of the outliers.

The selection of variables is also different in the regression and in the ANN techniques. Therefore, the best set of attributes to include in an ANN are not necessarily the same that for a regression. The intrinsic capability of traditional regression technique does not explore automatically the interactions between variables when the model is built, as in ANN technique. This means the effect of each explanatory variable is assumed to be independent of the value of the remaining variables. Consequently, this entails that possible interactions between the independent variables cannot be accounted for in the regression models, unless transformation of variables is used. However, as the interactions between the variables are unknown, the transformation of original variables and the exploration of all possible combinations is a cumbersome task. In this context, ANN is a more powerful technique than regression, because it intrinsically explores all possible interactions between explanatory variables. The regression model (Oliveira et al., 2018) assumed the independence of the explanatory variables and 5 statistically significant variables were identified as influencing the amount of packaging waste that is separately collected (using the Akaike information criterion to select the best subset of variables). Whereas, in the ANN model, the GA selected a set of 10 relevant variables (Fig. 4).

As mentioned in section 4.1, in the ANN model, the variable with the highest influence on the amount of separately-collected packaging waste is *relative accessibility to bring-banks* (relative weight of 0.15) followed by *number of school years attended* (relative weight of 0.13) and *inhabitants per bring-bank* (relative weight of 0.12) (Fig. 5). It is noted that the most influential variables identified in the regression model (*inhabitants per bring-bank*—0.30; *number of school years attended*—0.22; *relative accessibility to bring-banks*—0.18; *degree of urbanisation*—0.17 and *area*—0.12) (Oliveira et al., 2018) are also present in the group of relevant variables found in the ANN model. However, a smaller impact was found, since in the latter more variables were identified as statistically significant, and the individual relative weight of each is lower.

Overall, the performance of ANN models for predicting the amount of separately-collected packaging waste is much higher than the best regression model (multiple non-linear regression) reported previously (Oliveira et al., 2018), regardless of the inclusion or elimination of outliers of the dataset. This is a strong indication that the interaction between the variables (assumed independent by the regression model) does not only occur but contributes to explain the amount of separately-collected packaging waste.

4.2.3. Interpreting the results and using the model

Regression models, especially when there is no interaction between the independent variables, are easy to interpret. The contribution of each input to the output is simply given by the signal and value of the corresponding regression coefficient. If the inputs are normalized beforehand, the value of the regression coefficient will reveal its relative importance. When the inputs aren't normalized, the assessment of the relative importance of each input is not direct since they may not be in the same scale. It can be computed or evaluated performing a sensibility analysis on the model's inputs. In other words, for the regression technique it is very simple to change the input of a given parameter and see how this affects the output. This leads to another point, that is the easiness of using the model either by hand (with a calculator) or implementing it in a spreadsheet.

On the other hand, interpreting the results of ANN using connections weights is not as straightforward. Aggregating all the connection weights from each input to the output layer may enable some interpretation of the results, but it will always be limited due to the inputs interaction. Moreover, when the number of neurons on the hidden layers and the number of hidden layers increase, this rapidly ceases to be valid. This difficulty on interpreting the results is reflected by the black-box classification usually attributed to ANN models. Additionally, for the same dataset and network topology, there may be more than one ANN model with the same result but different weights. From a traditional statistics viewpoint, an ANN is a non-identifiable model. A hands-on practical application of an ANN model is a very tricky task, even for small models, because a scientific calculator is required for computing the transformation functions. The implementation of the resulting ANN in a spreadsheet is possible, but its use is typically done resorting to special purpose software, using some of the various packages freely available (e.g. NeuralNet for R, scikitlearn for Phyton).

Regression models allow for insight and use by any user with basic mathematical skills. ANN models have potentially higher accuracy than regression models, but its use with a spreadsheet is more difficult. Its implementation and interpretation of the results are also extremely difficult due to the interaction and different paths from the input to the output neurons.

5. Conclusions

This work describes development of a model for predicting the amount of packaging waste that is separately collected at the municipality level. The model is based on the artificial neural network (ANN) technique and was developed using a dataset of 42 municipalities of differing characteristics (size, population, age, economic level, education, waste collection service, etc).

The GA used together with the ANN model allowed the identification of ten explanatory variables as the most influential to the amount of packaging waste that is separately collected. The performance of the ANN model demonstrated to be superior ($R^2 = 0.98$) to the best regression model ($R^2 = 0.73$, multiple non-linear regression).

However, model development and the practical interpretation of the results is more difficult. The ANN model proposed in the current study is the most powerful explanatory model describing how the amount of separately-collected packaging waste is influenced by various variables at a municipality level. It is thus a valuable tool in the definition of strategies to increase municipal packaging recycling.

Acknowledgments

Celia Dias-Ferreira and Verónica Oliveira have been funded through FCT “Fundação para a Ciência e para a Tecnologia” by POCH – Programa Operacional Capital Humano within ESF – European Social Fund and by national funds from MCTES (SFRH/BPD/100717/2014; SFRH/BD/115312/2016). Vitor Sousa acknowledge the support from FCT and CERIS. The authors also gratefully acknowledge project Life PAYT – “Tool to Reduce Waste in South Europe” (LIFE15 ENV/PT/000609) for technical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2018.11.063.

References

Abbasi, M., Hanandeh, A. El, 2016. Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste Manag. 56, 13–22.

https://doi.org/10.1016/j.wasman.2016.05.018.
Abdoli, M.A., Nezhad, M.F., Sede, R.S., Behboudian, S., 2012. Longterm forecasting of solid waste generation by the artificial neural networks. Environ. Prog. Sustain. Energy 31, 628–636. https://doi.org/10.1002/ep.10591.
Adamovi, V.M., Antanasijevi, D.Z., Risti, M.Đ., Peri, A.A., Pocajt, V.V., 2017. Prediction of municipal solid waste generation using artificial neural network approach enhanced by structural break analysis. Environ. Sci. Pollut. Res. 24, 299–311. https://doi.org/10.1007/s11356-016-7767-x.
Agatonovic-Kustrin, S., Beresfold, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharmaceut. Biomed. Anal. 22, 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1.
Antanasijević, D., Pocajt, V., Popović, I., Redžić, N., Ristić, M., 2013. The forecasting of municipal waste generation using artificial neural networks and sustainability indicators. Sustain. Sci. 8, 37–46. https://doi.org/10.1007/s11625-012-0161-9.
Azadi, S., Karimi-Jashni, A., 2016. Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: a case study of Fars province, Iran. Waste Manag. 48, 14–23. https://doi.org/10.1016/j.wasman.2015.09.034.
BiPRO/CRI, 2015. Assessment of Separate Collection Schemes in the 28 Capitals of the EU. Final report.
Chhay, L., Reyad, M.A.H., Suy, R., Islam, M.R., Mian, M.M., 2018. Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. J. Mater. Cycles Waste Manag. 20, 1761–1770. https://doi.org/10.1007/s10163-018-0743-4.
Dong, C., Jin, B., Li, D., 2003. Predicting the Heating Value of MSW with a Feed Forward Neural Network 23, pp. 103–106. https://doi.org/10.1016/S0956-053X(02)00162-9.
EPRS, 2016. Closing the Loop - New Circular Economy Package. Briefing.
ERSAR, 2015. Relatório Anual dos Serviços de Águas e Resíduos em Portugal, vol. 1. Caracterização do setor de águas e resíduos - relatório dados indicadores (In Portuguese), 2015.
ERSUC, 2016. ERSUC - Resíduos Sólidos do Centro. S.A. Relatório e Contas 2015 117 (In Portuguese).
European Parliament and of the Council, 2008. Directive 2008/98/EC of the European parliament and the Council of 19 November 2008 on waste and repealing certain directives (text with EEA relevance). Offic. J. Euro. Communities L, 312/3-30.
Ferreira, J.A., Figueiredo, M.C., Oliveira, J.A., 2014. Forecasting household packaging waste generation : a case tudy. Lect. Notes Comput. Sci. 14th Int. Conf. Comput. Sci. Its Appl. LNCS 8581, 523–538.
Huber, P.J., 2011. Robust statistics. In: Lovric, M. (Ed.), International Encyclopedia of Statistical Science. Springer, pp. 1248–1251.
Kannangara, M., Dua, R., Ahmadi, L., Bensebaa, F., 2018. Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste Manag. 74, 3–15. https://doi.org/10.1016/j.wasman.2017.11.057.
Nabavi-Pelesaraei, A., Bayat, R., Hosseinzadeh-bandbafha, H., 2017. Modeling of energy consumption and environmental life cycle assessment for incineration and landfill systems of municipal solid waste management - a case study in Tehran Metropolis of Iran. J. Clean. Prod. 148, 427–440. https://doi.org/10.1016/j.jclepro.2017.01.172.
Oliveira, V., Sousa, V., Vaz, J.M., Dias-Ferreira, C., 2018. Model for the separate collection of packaging waste in Portuguese low-performing recycling regions. J. Environ. Manag. 216, 13–24. https://doi.org/10.1016/j.jenvman.2017.04.065.
Ordónez-Ponce, E., Samarasinghe, S., Torgerson, L., 2006. Artificial neural networks for assessing waste generation factors and forecasting waste generation: a case study of Chile. J. Solid Waste Technol. Manag. 32, 167–184.
Shamshiry, E., Mokhtar, M. Bin, Abdulai, A., 2014. Comparison of artificial neural network (ANN) and multiple regression analysis for predicting the amount of solid waste generation in a tourist and tropical area - Langkawi Island. In: International Conference on Biological, Civil and Environmental Enginnering, pp. 161–166. https://doi.org/10.15242/IICBE.C0314099.
Statistics Portugal, 2016. Território. https://ine.pt/. (Accessed 23 January 2017).
Statistics Portugal, 2014a. População. https://ine.pt/. (Accessed 23 January 2017).
Statistics Portugal, 2014b. Tipologia de Áreas Urbanas - Relatório Técnico, p. 20 (In Portuguese).
Statistics Portugal, 2013a. Crime e justiça criminal. https://ine.pt/. (Accessed 23 January 2017).
Statistics Portugal, 2013b. Condições de vida e cidadania. https://ine.pt/. (Accessed 23 March 2016).
Statistics Portugal, 2011a. Censos 2011 Resultados Definitivos - Portugal, p. 560 (In Portuguese).
Statistics Portugal, 2011b. Mercado de Trabalho. https://ine.pt/. (Accessed 23 January 2017).
Statistics Portugal, 2007. Saúde. https://ine.pt/. (Accessed 23 January 2017).
Sun, N., Chungpaibulpatana, S., 2017. Development of an appropriate model for forecasting municipal solid saste beneration in Bangkok. Energy Procedia 138, 907–912. https://doi.org/10.1016/j.egypro.2017.10.134.
United Nations, 2015. The Sustainable Development Agenda. https://www.un.org/sustainabledevelopment/development-agenda/. (Accessed 19 September 2018).
VALORLIS, 2016. Valorlis - Valorização e Tratamento de Resíduos Sólidos, S.A. Relatório & Contas 2015, p. 126 (In Portuguese).
Xiao, G., Ni, M., Chi, Y., Jin, B., Xiao, R., Zhong, Z., Huang, Y., 2009. Gasification characteristics of MSW and an ANN prediction model. Waste Manag. 29, 240–244. https://doi.org/10.1016/j.wasman.2008.02.022.
Younes, M.K., Nopiah, Z.M., Basri, N.E.A., Basri, H., Abushammala, M.F.M., Maulud, K.N.A., 2015. Prediction of municipal solid waste generation using nonlinear autoregressive network. Environ. Monit. Assess. 187, 753. https://doi.org/10.1007/s10661-015-4977-5.