Waste Management 39 (2015) 15–25

Contents lists available at ScienceDirect

Waste Management

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

Waste container weighing data processing to create reliable information of household waste generation

Pirjo Korhonen *, Juha Kaila

Aalto University, School of Engineering, Department of Civil and Environmental Engineering, P.O. Box 12100, FI-00076 Aalto, Finland

ARTICLE INFO

Article history:
Received 10 October 2014
Accepted 16 February 2015
Available online 9 March 2015

Keywords:
Household waste
Data cleaning
Data mining
Dimensioning
Waste management

ABSTRACT

Household mixed waste container weighing data was processed by knowledge discovery and data mining techniques to create reliable information of household waste generation. The final data set included 27,865 weight measurements covering the whole year 2013 and it was selected from a database of Helsinki Region Environmental Services Authority, Finland. The data set contains mixed household waste arising in 6 m^3 containers and it was processed identifying missing values and inconsistently low and high values as errors. The share of missing values and errors in the data set was 0.6%. This provides evidence that the waste weighing data gives reliable information of mixed waste generation at collection point level. Characteristic of mixed household waste arising at the waste collection point level is a wide variation between pickups. The seasonal variation pattern as a result of collective similarities in behaviour of households was clearly detected by smoothed medians of waste weight time series. The evaluation of the collection time series against the defined distribution range of pickup weights on the waste collection point level shows that 65% of the pickups were from collection points with optimally dimensioned container capacity and the collection points with over- and under-dimensioned container capacities were noted in 9.5% and 3.4% of all pickups, respectively. Occasional extra waste in containers occurred in 21.2% of the pickups indicating the irregular behaviour of individual households. The results of this analysis show that processing waste weighing data using knowledge discovery and data mining techniques provides trustworthy information of household waste generation and its variations.

1. Introduction

The waste prevention objectives, set by the European Union (EU), require monitoring municipal solid waste generation in member states (EC, 2008). Monitoring based on waste weighing data provides closer insight into municipal solid waste generation at sources and offers more special and individual information about the quantities and variation of municipal solid waste arising. Reliable data of waste quantities and generation trends are important information e.g. for planning and modelling waste management and to estimate resource management and the workload of waste collection services (Beigl et al., 2008; Rimaityte et al., 2012; Shamshiry et al., 2011; Cherian and Jacob, 2012).

In general, municipal solid waste generation is expressed in quantity of waste generated per capita in different time frames, mostly kg per capita per day or per year, or kg per household per week, and is based on official statistics or sampling data. Additionally, in many studies numerous factors which influence the quantity and composition of solid waste at the household level are also identified (e.g. Beigl et al., 2004; Skumatz, 2008; Dahlén et al., 2009; Denafas et al., 2014). However, the commonly highlighted attributes of waste-related data are uncertainty and unreliability (e.g. Dahlén, 2008; Karadimas and Loumos, 2008; Dahlén and Lagerkvist, 2010; Rada et al., 2013). The lack of reliable and disaggregated waste data is recognised although the data is gathered daily at operational waste management level and modern traceability devices with Global Positioning System (GPS) and General Packet Radio Service (GPRS) technologies allow real-time data collection and transmission (Faccio et al., 2011).

This paper discusses a new aspect to utilise municipal solid waste weighing data. Mixed waste weighing data selected from a database of municipal waste management authority is processed to determine the household mixed waste generation characteristics in residential properties based on the mixed waste quantities related to container capacity. Knowledge discovery in databases and data mining are adopted to create property-based groups, and their waste generation profiles. The main aim of the whole

* Corresponding author. Gsm: +358 505417752.
E-mail addresses: pirjo.korhonen@aalto.fi (P. Korhonen), juha.kaila@kasui.fi (J. Kaila).

http://dx.doi.org/10.1016/j.wasman.2015.02.021
0956-053X/

research is to develop a method to identify trend changes in waste generation from households based on operational data. In the future the method will be employed to evaluate the effect of waste avoidance campaigns and recycling programmes on waste arising both at the property and at different spatial levels.

2. Waste collection data

In waste-related studies the development of municipal solid waste generation models and the evaluation of waste management systems are usually based on municipal solid waste collection data. Parfitt et al. (2001) used operational municipal solid waste collection data to compare regions with their waste management and recycling performances by a hierarchical cluster analysis method. Municipal solid waste generation was defined as kg per household per week. Dahlén et al. (2009) evaluated household waste generation impact factors and collection systems in Swedish municipalities based on waste collection data, and annual statistics of local authorities and waste management companies. Dahlén and Lagerkvist (2010) used official household waste data to evaluate the strengths and weaknesses of weight-based billing in household waste collection systems.

Xu et al. (2013) based their hybrid model on historical municipal solid waste generation time series data (from 2000 to 2009) without demographics and socio-economic factors. They combined grey system theory with seasonal autoregressive integrated moving average (sARIMA) model to forecast seasonal and annual municipal solid waste generation. Navarro-Esbrí et al. (2002) utilised daily and monthly municipal solid waste collection data in their study to predict waste generation by a non-linear dynamics technique producing a result which was comparable to that of the sARIMA methodology. Benitez et al. (2008) collected residential waste samples of households for the waste generation analysis and modelling. In addition to the total weight of waste sampling bags, education level, household size, and income of participating households were included but seasonal variation was not taken into account in the analysis.

Waste collection data is typically discrete data and the variation of individual container weight values is quite large. Dahlén (2008) has pointed out that the precise metadata and the awareness of uncertainty sources of waste collection data increases the quality of input data and provides better research results. The general data problems presented by Dahlén are the result from (a) factors affecting waste generation, (b) technical devices, (c) variation of data registration level, (d) inadequate waste generator-related additional data, and (e) gaps in waste flows which mean that all wastes from households do not end up in the waste flows of municipal waste management. The variation of waste arising from households is also a result of different waste practises in households (Beigl et al., 2008).

Waste management information systems consist of several subsystems with databases such as collection and transfer systems, waste reception systems (e.g. weight bridge systems), and invoicing systems. The collection and transport systems feature logistic systems, such as transport control systems (TCS), which have applications for mobile terminal, software for a map and navigation, data transferring, and office software designed for the driver of a waste collection vehicle (Rada et al., 2010, 2013; Faccio et al., 2011). The data of the waste collection and transfer system are related to waste management services provided for the customers and offer special information to customers. The services are governed by municipal waste management regulations that define for instance separate collections of recyclables and maximum collection frequencies.

3. Material and methods

3.1. Data

3.1.1. Study area

The study area is the operational area of Helsinki Region Environmental Services Authority (HSY) covering five municipalities with an area of 1136 km^2 on the south coast of Finland. At the end of 2013, the population in the study area was in total 1,128,515 inhabitants. The characteristic of the area is mostly urban and sub-urban. Two thirds of the population lives in blocks of flats and only one tenth in single-family houses. From the households in the study area, 350,480 tonnes of municipal solid wastes were collected in the year 2013. Of this 189,488 tonnes (54%) were collected as mixed waste, 38,846 tonnes (11%) as biowaste, 115,526 tonnes (33%) as other recyclables, and 6620 tonnes (2%) as other waste (HSY, 2014a,b).

In Finland, according to the Waste Act (646/2011) the municipalities are responsible for offering municipal solid waste collection services to residential properties, public services, and private health and educational services. Mixed waste from every property is collected by on-site or property-close collection systems. There is also separate collection of the waste fractions for material recovery from multi-family properties, and public and private services. The number of fractions to be separated at a source is defined in the municipal waste management regulations according to the number of housing units in a property or the weekly generation of a waste fraction at the service properties. In Helsinki Metropolitan Area residential properties with ten or more housing units are required to organise the separate collection of paper, biowaste, and cardboard, and the properties with 20 or more housing units have to organise the separate collection of paper, biowaste, cardboard, glass and metal in an addition of mixed waste collection. Properties with less than ten households are recommended to take voluntary source-separated waste fractions to drop-off points and compost their biowaste. The authority (HSY) organises the collection of recyclables except for recyclable paper, which is under producer liability and it is forbidden to put recyclable paper into mixed waste containers by regulation (HSY, 2012).

The property owner signs a waste service agreement with HSY for household waste collection and transportation. The contract is saved in a waste management database which includes the identifiable information about the customers and their waste collection points, separate collected waste fractions, number and capacity of bins and containers, type of bins and containers, and emptying schedules. Collection task lists for the drivers of waste collection vehicles are generated from the customer database. The task lists contain all necessary information required to empty the right bins or containers at the property, such as the collection address, and the number of bins or containers at that location. After emptying the driver signs the task and, when a container is weighed, the weight is updated to the task row manually or it is transferred from the scale to the computer in the vehicle by wireless data transfer technology. After the collection route is completed, the emptying data is transferred to the waste management database for billing and other utilisation.

In 2013 the total number of pickups of mixed waste bins and containers was 5,609,466 in the service area of HSY. Table 1 shows

Table 1
Mixed waste pickups from different bin types at the study area. (HSY, 2014b).

Bin/container	Nr of pickups 2013
140–300 L bins and bags	1,352,224
600–660 L bins	4,189,731
Others	67,511
Total	5,609,466