Modeling a Waste Disposal Process via a Discrete Mixture of Count Data Models

Iliya Markov

Matthieu de Lapparent

Michel Bierlaire

Sacha Varone

April 2015

STRC | **15th Swiss Transport Research Conference**

Monte Verità / Ascona, April 15 – 17, 2015

Modeling a Waste Disposal Process via a Discrete Mixture of Count Data Models

Iliya Markov[a,b], Matthieu de Lapparent[a], Michel Bierlaire[a], Sacha Varone[b]

[a] Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne
phone: +41 (0) 21 693 25 32
fax: +41 (0) 21 693 80 60

[b] Haute École de Gestion de Genève
University of Applied Sciences Western Switzerland (HES-SO)
phone: +41 (0) 22 388 17 00
fax: +41 (0) 22 388 17 01

{iliya.markov, matthieu.delapparent, michel.bierlaire}@epfl.ch, sacha.varone@hesge.ch

April 2015

Abstract

We propose a modeling framework for the data generating process of waste disposal in recyclable waste containers. It is based on a discrete mixture of count data models representing populations depositing different quantities in the containers, thus reflecting a realistic underlying behavior. It is tested on real data coming from ultrasound sensors mounted inside the containers and exhibits better in- and out-of-sample performance compared to a simple count data model assuming only one deposit quantity. The purpose of the mixture model is to forecast container waste levels at a future date when collection will take place. It thus becomes the first-step ingredient in a framework for efficient waste collection optimization.

Keywords

waste generation, waste collection, discrete mixture model

1 Introduction

Waste collection is one of the most important transportation problems for any municipality. In recent years, increased competition, the entry of new technologies, data availability and the growth of off-the-shelf routing tools have exerted pressure on many collectors to optimize at least part of their processes in order to stay in business.

Waste collection problems differ depending on the type of waste and customer. Residential collection services private homes and is often solved as arc routing. Waste bags or bins are placed in front of buildings and vehicles pick from one or both sides of the street according to a predetermined schedule. Commercial collection handles waste from various businesses and institutions. It is modeled as vehicle routing due to the geographically distinct locations and often involves more intricate temporal and operational constraints imposed by the customers. Roll-on roll-off refers to the routing of large containers, such as those found at construction sites.

The problem discussed here is related to the collection of recyclable material, for example white and colored glass, PET[1], plastic, aluminum, and others. This waste is predominantly produced by private homes and restaurants, and deposited in large, often underground, containers. A single collection point usually accommodates containers for several types of material. This problem can be placed under the commercial collection category, since collection points are scatted geographically, and is commonly solved as a vehicle routing problem with intermediate facilities, with the latter representing the recycling plants that are periodically visited to empty a full vehicle.

In this context, efficient waste collection depends not only on the optimal or near-optimal routing of the collection vehicles, but also on the quality of the forecast of the container levels at the time of collection. The ratio of container volume to collection vehicle volume is much higher compared to residential collection, and vehicles must be emptied more often. Given that collection tours may be planned several days in advance, incorrect forecasts may thus lead to vehicle routes that are optimal with respect to the forecast demand, but far from optimal with respect to the realized one. This article focuses on forecasting recyclable waste generation using data on daily container levels coming from ultrasound sensors mounted inside the containers. The forecast levels can then then used as demand data for the solution of the subsequent vehicle routing problem, which is out of the scope of this article.

The contribution of this work is two-fold. First, most of the existing literature on waste generation

[1]Polyethylene terephthalate, a common polyester from which containers for food and liquid are made.

analysis and forecasting studies its determinants, such as socio-economic factors, at an aggregate city or neighborhood level. The analyses are used for policy purposes or for the planning of incineration plants and land fills. In this respect, we take a much more operational point of view, proposing a model that will be used in the planning of collection tours. Secondly, the literature facing similar operational problems is mainly focused on determining critical container waste levels, which when reached should trigger a collection operation. Therefore, this work goes one step further, by modeling the data generating process as a discrete mixture of count data models representing populations depositing different quantities in the containers, thus reflecting a realistic underlying behavior. The numerical experiments demonstrate the superior forecast qualities of the mixture model in comparison to simple count data models and the linear least square fits usually used in industry.

This remainder of this article is organized as follows. Section 2 is a review of related literature on forecasting, predominantly in the field of waste collection. Section 3 describes the discrete mixture model and how it is used for forecasting. Section 4 tests the proposed model on numerical data, and finally Section 5 concludes and outlines directions for future research.

2 Related Literature

Forecasting is an extensive scientific branch utilizing various mathematical and statistical techniques and its applications are numerous. A fair amount of literature is also explicitly concerned with forecasting in the field of waste management. Beigl *et al.* (2008) is a systematic review of 45 modeling approaches categorized in terms of the regional scale, the modeled waste flows, the hypothesized independent variables and the modeling method. The modeling approaches in the reviewed literature range from correlation and regression analysis to time series analysis and system dynamics, applied at the building, municipal, regional and country level.

At the municipal and regional level, Chen and Chang (2000) discuss a framework based on grey fuzzy dynamic modeling, which is appropriate in situations where there is a lack of historical record on waste generation and collection. It is implemented in Tainan, Taiwan and shows its better prediction accuracy compared to the conventional grey dynamic model, the least-square regression method, and the fuzzy goal regression technique. Limited sample data is also the topic of Dyson and Chang (2005) who use system dynamic modeling to forecast solid waste generation in San Antonio, US. They find that this approach covers a variety of possible causative models and outperforms traditional least-square methods. Navarro-Esbrí *et al.* (2002) describe a daily and monthly prediction technique based on non-linear dynamics and compare its performance to a seasonal ARIMA (SARIMA) methodology. The technique is assessed on the municipal solid waste (MSW) time series of three cities in Spain and Greece. The SARIMA model exhibits superior performance.

Benítez *et al.* (2008) use a multivariate regression with education level, income and number of residents to predict per capita production of residential waste in Mexicali, Mexico. Li *et al.* (2011) use the analytical hierarchy process to determine three sets of important parameters—waste generation per unit of consumer expenditure, consumer expenditure distribution to activities in unit time, and time assignment to activities by different resident groups. Their model is successful in forecasting MSW generation in Beijing. Xu *et al.* (2013) combine the SARIMA model and grey system theory to forecast MSW generation at multiple time scales without needing to consider demographics and socioeconomic factors. Their case study in Xiamen City, China, shows that the model is successful in forecasting seasonal and annual dynamics with various time scales with the desired accuracy.

The literature on waste generation at the container level is much more scarce, and it is usually combined with vehicle routing, in what is referred to as an inventory routing problem. Nolz *et al.* (2011) develop a tabu search algorithm for a stochastic collector-managed inventory routing problem for the collection of infectious waste from pharmacies. RFID allow the collector to keep track of the waste level at the pharmacies, which leads to a cost reduction of 50% compared to the current scenario where the pharmacists calls the collector when the inventory level reaches 85%. Nolz *et al.* (2014) propose a scenario sampling method and an adaptive large neighborhood search algorithm for the same problem. Bitsch (2012) develops a variable neighborhood search for the collection of recyclable waste in a Danish region. Waste level is stochastic and containers should be emptied so that the probability of overflow is six sigmas away. Visit schedules are determined by studying historical fill levels recorded at the time of emptying the containers. The obtained results considerably under-perform the state of practice, which is attributed to the conservative overflow acceptance criterion and the fact that, unlike the proposed algorithm, the driver is able to continuously track the waste level on the vehicle and adapt the route accordingly.

Johansson (2006) and Mes (2012) use simulation to confirm the benefits of migrating from static to dynamic collection policies in Malmö, Sweden and a study area in the Netherlands, respectively, where containers are equipped with level and motion sensors, respectively. Mes (2012) finds a positive added value of investing in level sensors compared to simple motion sensors that detect when a container was emptied. Mes *et al.* (2014) apply optimal learning techniques to tune the parameters related to inventory control (deciding which containers to select) assuming accurate container level information and find strong benefits from parameter tuning.

Krikke *et al.* (2008) describe a collection system for recyclable materials from the dismantling of old vehicles in the Netherlands. Sensors monitor the quantity of various types of liquids recovered from the cars, and report when a critical level is reached, so the point can be planned for collection in the next period. Although the collection periodicity of many sites is in the order of one year, costs savings can amount to 19%. McLeod *et al.* (2013) describe a similar system, where sensors are used to transmit the level at donation banks belonging to a UK charity. The level information is used to plan the daily routes. A more complex system, where both containers and vehicles are remotely monitored is presented in Faccio *et al.* (2011). It proposes a real-time dynamic routing model relying on simulated data and performs an economic feasibility study. Yet, none of the articles that either use or simulate container level data develops a forecasting model.

The analysis of waste generation is rich and varied, and the operational element has become more pronounced in recent years, with the emergence of new and cheaper technologies for container waste level tracking. Yet, proprietary data for the development of forecasting models may not be easy to obtain, or a sufficient sample may not be available that will allow the development of an appropriate model. We thus contribute with a new operational model estimated and validated on real container-based data samples.

3 Methodology

For each container, we observe the total waste level, which is detected by an internal ultrasound sensor and periodically transmitted through the GSM network to a centralized database. The raw sensor data is preprocessed automatically on-site to remove noise and extrapolated at the end of each date. Let $L_{i,t}$ denote the waste level of container i at date t in percent, then $Q_{i,t} = C_i(L_{i,t} - L_{i,t-1})$ is the observed volume deposited in container i at date t, where C_i is the usable capacity of that container. Emptying events are detected as sharp decreases in the level, and then $Q_{i,t} = C_i L_{i,t}$, reflecting the fact that the observed volume deposited in the container after an emptying event is equal to the measured level times the capacity. For all practical purposes, the above calculations are correct because: (1) containers are usually emptied early in the morning to avoid heavy traffic, (2) it is forbidden to deposit during the night, (3) and container levels pertain to the end of each date.

We define the data generating process of the total volume of waste deposited in container i at date t as:

$$Q^{\star}_{i,t} = \sum\nolimits_{k=1}^{K} n_{i,t,k} q_k \tag{1}$$

where $n_{i,t,k}$ is the number of visits to container i at date t with disposal quantity q_k, where $k = 1, \ldots, K$, indexes the set of distinct disposal quantities, which is discrete and exogenously determined with the purpose of representing the reality reasonably well. We model the number of visits by disposal quantity as independent Poisson processes[2]:

$$n_{i,t,k} \xrightarrow{\text{iid}} \mathcal{P}\left(\lambda_{i,t,k}\right) \tag{2}$$

Because we do not observe how visits by disposal quantity are distributed in reality, we fit them to probability distributions. Let $\pi_{i,t,k}$ be the probability of having a disposal quantity of q_k in container i at date t. Then the expected value of the total volume of waste deposited in container i at date t becomes:

$$\mathbb{E}\left(Q^{\star}_{i,t}\right) = \sum\nolimits_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k} \tag{3}$$

As defined above, $Q_{i,t}$ denotes the observed volume of waste deposited in container i at date t. We minimize the sum of squared differences between observed and expected over N containers and T days, as expressed by:

$$\min_{\lambda,\pi} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \left(Q_{i,t} - \sum\nolimits_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k}\right)^2 \tag{4}$$

This objective function represents an implicit likelihood for the observed values. The errors are assumed to be independent and identically distributed. This strict exogeneity is a strong assumption, which we leave at this point for further research. We make additional parametric assumptions for the Poisson rates:

$$\lambda_{i,t,k}\left(\boldsymbol{\theta}\right) = \exp\left(\mathbf{x}_{i,t}^{\top} \boldsymbol{\beta}_k\right) \tag{5}$$

and the logit-type probabilities of the disposal quantities:

$$\pi_{i,t,k}\left(\boldsymbol{\theta}\right) = \frac{\exp\left(\mathbf{z}_{i,t}^{\top} \boldsymbol{\gamma}_k\right)}{\sum_{j=1}^{K} \exp\left(\mathbf{z}_{i,t}^{\top} \boldsymbol{\gamma}_j\right)} \tag{6}$$

where $\mathbf{x}_{i,t}$ and $\mathbf{z}_{i,t}$ represent the vectors of covariates, and $\boldsymbol{\beta}_k$ and $\boldsymbol{\gamma}_k$ the vectors of parameters of the Poisson processes and the probability distributions, respectively. The vectors of covariates $\mathbf{x}_{i,t}$ and $\mathbf{z}_{i,t}$ may or may not coincide. In compact form, the minimization problem then writes as:

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \left(Q_{i,t} - \sum\nolimits_{k=1}^{K} \frac{\exp\left(\mathbf{x}_{i,t}^{\top} \boldsymbol{\beta}_k + \mathbf{z}_{i,t}^{\top} \boldsymbol{\gamma}_k + \ln\left(q_k\right)\right)}{\sum_{j=1}^{K} \exp\left(\mathbf{z}_{i,t}^{\top} \boldsymbol{\gamma}_j\right)}\right)^2 \tag{7}$$

[2]We are aware that the negative binomial distribution models over-dispersion.

where $\boldsymbol{\theta} := (\boldsymbol{\beta}_k, \boldsymbol{\gamma}_k : \forall k)$. We need to set $\boldsymbol{\gamma}_{k^\star} = \mathbf{0}$ for one arbitrarily chosen $k^\star$ to ensure a one-to-one mapping between the vector of parameters to estimate and the objective function to minimize. This model will be referred to later in the text as the mixture model. If only one disposal quantity is considered, it degenerates to a pseudo-count data model:

$$\min_{\theta\in\Theta} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \left(Q_{i,t} - \exp\left(\mathbf{x}_{i,t}^\top \boldsymbol{\beta} + \ln(q)\right)\right)^2 \tag{8}$$

where $\boldsymbol{\beta}$ is the only parameter vector to estimate, and q the only disposal quantity used. We use the qualifier pseudo to underline that the observed volumes do not necessarily take integer values. This model will be referred to as the simple model.

From a conceptual point of view, the appeal of the mixture approach is that it models a realistic behavior, although simplified, without applying any data transformations. The only explicit assumption is on the q_k. Since our data generating process does not produce count data, the application of classical count data models, such as Poisson or negative binomial regression, would require a data transformation. However, this data transformation is based on additional assumptions and introduces additional biases on top of those already present in the data. The above approach is thus a more elegant way to directly model the observed variable. We tested other approaches as well, including dynamic panel regression, machine learning, classical count data and hurdle models, and they produced inferior results on our data sets.

The purpose of this model is to make short term predictions on waste level in a set of containers for recyclable waste by considering time varying and container fixed effects. We are not interested in explaining container fixed effects through population density and socio-economic factors, but only in estimating them. Once the model is estimated, using new sets of covariates $\dot{\mathbf{x}}_{i,t}$ and $\dot{\mathbf{z}}_{i,t}$, and the estimates $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\gamma}}_k$, we can generate a forecast as follows:

$$\dot{Q}_{i,t} = \sum\nolimits_{k=1}^{K} \frac{\exp\left(\dot{\mathbf{x}}_{i,t}^\top \hat{\boldsymbol{\beta}}_k + \dot{\mathbf{z}}_{i,t}^\top \hat{\boldsymbol{\gamma}}_k + \ln(q_k)\right)}{\sum_{j=1}^{K} \exp\left(\dot{\mathbf{z}}_{i,t}^\top \hat{\boldsymbol{\gamma}}_j\right)} \tag{9}$$

A brief discussion is due on the appropriate covariates to use. Beside the container specific effects, we need time varying explanatory factors, that are easy to obtain both for estimation (past) and forecasting (future). Waste disposal volumes exhibit strong weekly seasonality, with peaks observed during the weekends. They may also be affected by annual seasonality, and not least by weather. For operational purposes, much of this data can be automatically generated or obtained from web APIs. The produced forecast is used in the first level of the waste collection planning, namely the selection of the appropriate containers to empty at a particular future date.

Waste disposal is an inherently stochastic process. Therefore, however realistic, the proposed model will only be able to partially explain the observed in-data variance. Moreover, what is reported by the ultrasound sensors may differ from what happened in reality due to sensor malfunctioning, which is not rare. In addition, throwing an item in a container may disturb the surface of the waste inside and thus lead to a reported decrease rather than an increase in the level. The automatic post-processing of the data also introduces a bias. The last two effects, however, are less consequential. Thus, another benefit of modeling the underlying data generating process is the fact that poor fits can actually be used to detect malfunctioning container sensors.

4 Numerical Experiments

The numerical experiments are performed on a data set consisting of the daily waste levels of 36 containers for PET in Geneva, Switzerland, with capacity of 3040 or 3100 liters. They form a balanced panel covering the four months from March to June, 2014, or 122 days, which brings the total number of observations to 4392. Originally, the data set consisted of more containers. After visual inspection, containers were removed that had unreliable waste level data, high number of missing values, or insufficient sample length. The common sample period of the final set is such that it has the least amount of missing data. Any missing data in the common sample periods is linearly interpolated for the values of $Q_{i,t}$ for each container.

4.1 Estimating the models

Figure 1 plots the residuals normalized for standard deviation of the mixture model estimated on the complete data set. The vertical breaks represent the separate containers. It can be observed that the more than 99% of the values fall between -2 and 2, with no pronounced auto-correlation or heteroskedasticity effects. The underlying model assumptions are thus correct. The residual plot for the simple model is very similar, as shown in Fig. 2

Figure 3 plots the mean daily volume deposited in the containers. The vertical breaks represent the weeks in the sample. It can be observed that waste disposal exhibits a very strong weekly seasonality. A slight drift is also visible in the series. Very often in industry, when producing forecasts, an average past value over a rolling window is used, for example, one produced by a least squares linear fit to the data. It can easily be seen that such a forecast will fail to capture the much more complex dynamics present in the series.

Using these simple observations, with respect to the covariates, we decided to use dummy variables for the days of the week and the months of the sample period. To capture other

Figure 1: Residual plot of the mixture model on the complete data set

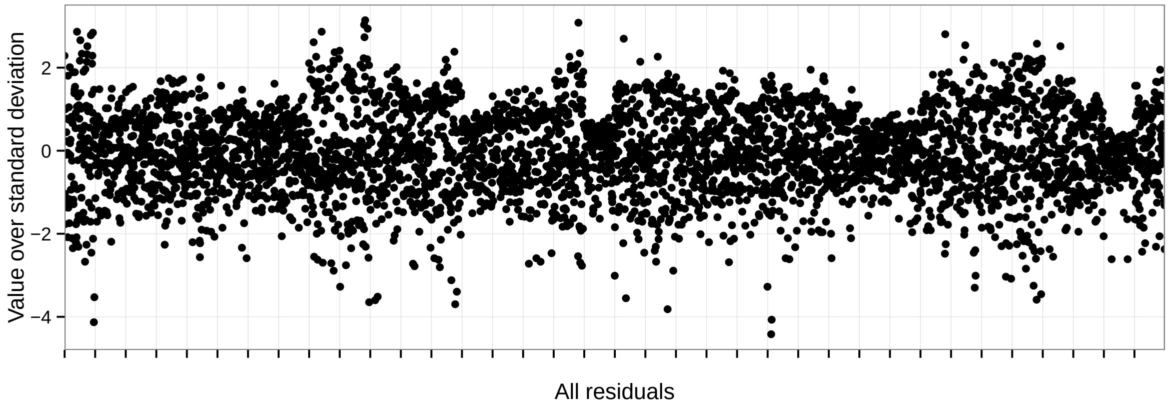
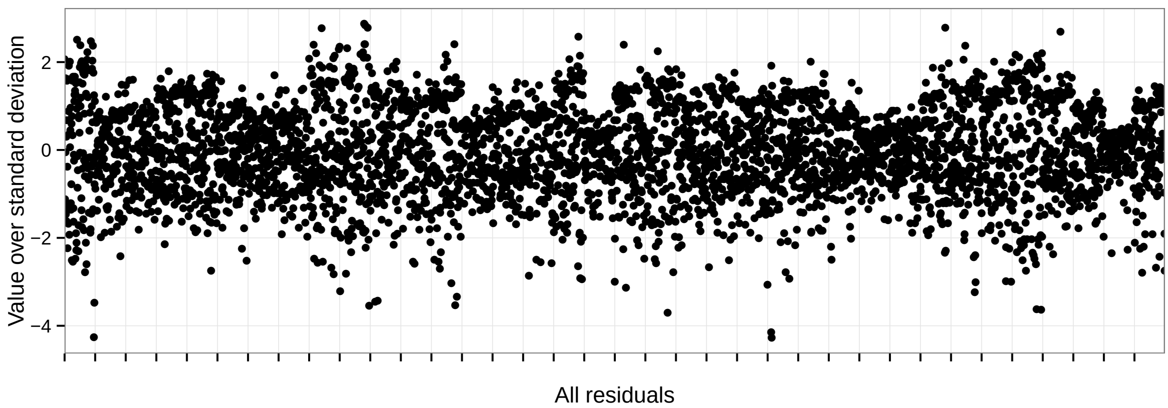


Figure 2: Residual plot of the simple model on the complete data set



behavioral aspects of waste generation, we also included several weather variables, namely the minimum daily temperature in Celsius, the precipitation in mm, the atmospheric pressure in hPa, and the wind speed in kmph. The latter were obtained from the World Weather Online Historical Weather API[3] and are the least correlated of the available weather variables. Finally, we also included dummy variables that characterize the container fixed effects. Table 1 summarizes the covariates, all of which are used for both $\mathbf{x}_{i,t}$ and $\mathbf{z}_{i,t}$.

In all calculations below, we use a deposit quantity of 10 liters for the simple model—the value of the latter was actually found not to affect the estimated model. For the mixture model, the deposit quantity values were found to be relatively important for the quality of the fit. We chose to use 5 and 15 liters, which are equidistant from 10, are deemed reasonable and were found to produce good results. As a data preprocessing step, we equated to zero all negative $Q_{i,t}$ and

[3] `http://www.worldweatheronline.com/`

Figure 3: Mean daily volume deposited in the containers

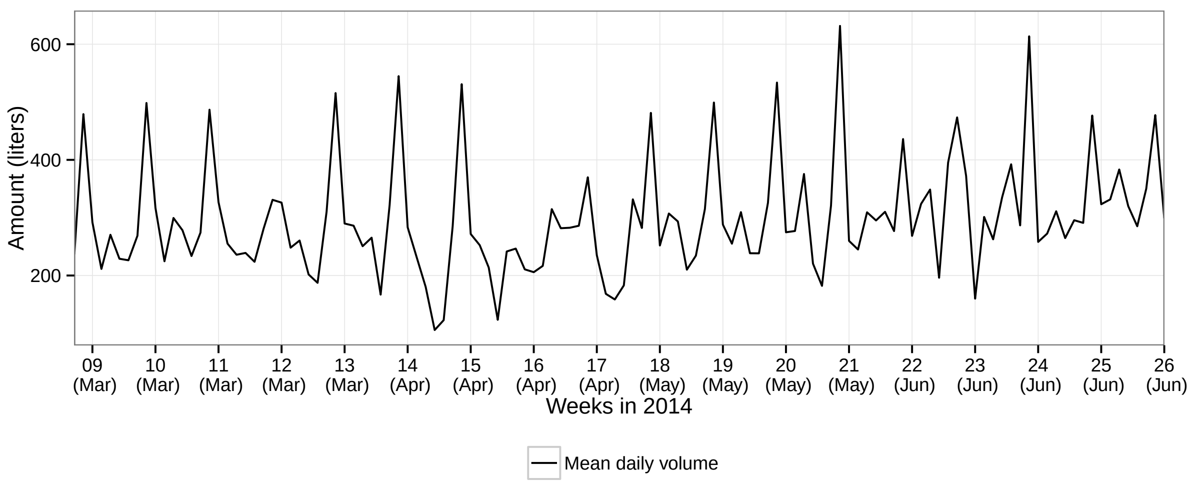


Table 1: Table of covariates

Variable	Type
Container fixed effect	dummy
Day of the week	dummy
Month	dummy
Minimum temperature in Celsius	continuous
Precipitation in mm	continuous
Pressure in hPa	continuous
Wind speed in kmph	continuous

removed the 99% outliers from the series. To evaluate the quality of the fits, we use the Akaike Information Criterion (AIC) and the coefficient of determination R^2. The AIC is defined as:

$$\text{AIC} = \left(\frac{SS_{\text{res}}}{N}\right)\exp(2K/N) \tag{10}$$

where the exponential penalizes model complexity. The lower the value of the AIC, the better the model is. The coefficient of determination R^2 in its most general form is defined as:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{11}$$

where a higher value corresponds to a better model, able to explain more of the observed in-data variance. Given the distribution of the residuals plotted in figures 1 and 2, we consider this R^2

Table 2: Estimated coefficients for the mixture model

	$\hat{\boldsymbol{\beta}}_1$ (5L)***	$\hat{\boldsymbol{\beta}}_2$ (15L)***	$\hat{\boldsymbol{\gamma}}_2$***
Minimum temperature in Celsius	1461.356	0.022	-0.037
Precipitation in mm	-0.821	-0.009	0.018
Pressure in hPa	-13.724	-0.001	0.010
Wind speed in kmph	7.580	-0.004	0.020
Monday	402.235	2.166	-9.693
Tuesday	1908.233	2.293	-9.977
Wednesday	-844.662	1.432	0.202
Thursday	1937.385	1.198	1.453
Friday	1876.162	1.239	4.419
Saturday	-6981.339	1.358	4.723
Sunday	1831.715	1.905	2.832
March	-27.136	2.955	-1.453
April	1071.406	2.746	-1.532
May	1689.979	2.988	-1.603
June	-2604.520	2.901	-1.452

The triple asterisk next to the parameter vector name designates that the presented parameter values in the corresponding column are significant at the 99% level.

as an appropriate indicator of the quality of the fit. In the above expressions, K is the number of estimated parameters, N is the number of observations, SS_{res} is the residual sum of squares $\sum_i (o_i - f_i)^2$ and SS_{tot} is the total sum of squares $\sum_i (o_i - \bar{o})^2$, with o_i and f_i representing the observed and fitted values, respectively. The objective function, its gradient, and Hessian were coded in R and optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method.

Estimating the mixture model on the complete data set produces an AIC of 52900 and an R^2 of 0.341, compared to 53700 and 0.300, respectively, for the simple model. Table 2 reports the estimated values of the parameters of the mixture model. We set $\boldsymbol{\gamma}_1 = \mathbf{0}$, and thus it is not estimated. There is no intercept in the model, and the table also omits the values of the parameters for the container fixed effects as they cannot be explained. It can be observed that the estimates $\hat{\boldsymbol{\beta}}_1$ are larger in absolute value. Therefore, they capture most of the variance which is consistent with the assumption that the small disposal quantities are produced by individuals, while the larger ones by restaurants and businesses and are thus more stable. Considering that the inverse of the Hessian matrix at the optimum is the variance covariance matrix, we

Table 3: Mean R^2 for estimation and validation sets

	Mixture model mean R^2	Simple model mean R^2
Estimation	0.364 (AIC 51400)	0.302 (AIC 53600)
Validation	0.286	0.274

are also able to examine the statistical significance of the estimated parameters. Essentially, all parameters, with rare exceptions for the unreported container fixed effect parameters, are statistically significant at the 99% level.

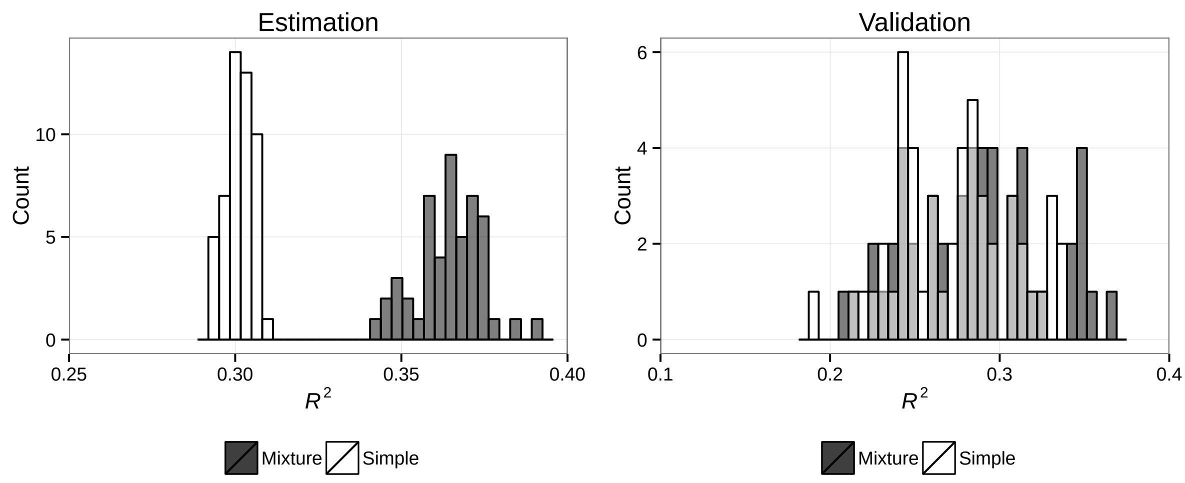
As expected, the minimum temperature is has a positive effect on disposals. Precipitation seems to have the expected negative effect, while the effect of pressure and wind speed is less clear. Regarding the days of the week, we can observe the gradually increasing importance of the larger disposal quantities as we approach the end of the week, both in terms of visit rate and probability, as determined by $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\gamma}}_2$, respectively. Yet, Sunday is very important for the small disposal quantities as well. The values for the months of the sample period indicate an increased importance of small disposal quantities relative to large ones for the months of April and May. Figure 3 shows the depression in mean volumes for April and May. The April vacation period could be responsible, by for example reducing the number of large disposal amounts because people are on vacation and thus accumulate less PET at home. A similar argument may hold for restaurants as well.

4.2 Validating the models

To validate the models, we performed 50 experiments, in each of which a random sample of 90% of the observations was drawn and both the mixture and the simple models were estimated on it. We wanted to make sure that all containers appeared in this sample, and all months appeared. If this was not the case, another sample would be redrawn until it conformed to this requirement. However, this was not necessary during the experiments. As mentioned above, the sample period covers 122 days per container. Thus the estimation is carried out on average on 109 days per container.

Both models are validated on the rest of the observations. i.e. on average on 13 observations per container, which is reasonable and even longer than what would be needed in reality. Table 3 reports the mean R^2 for the estimation and validation over the 50 experiments. The mixture model exhibits better fit at the 99% confidence level on the estimation samples, and at the 90% confidence level on the validation samples. Figure 4 plots the histograms of the fits from the

Figure 4: Histograms for estimation and validation sets



mixture and the simple model, and shows that in both the estimation and the validation sets, the dark grey bars, representing the mixture model, are shifted to the right with the respect to the white ones, representing the simple model. Therefore, representing a realistic, though simplified behavior, the mixture model appears to better explain the data generating process of waste disposal.

5 Conclusion

This article develops a discrete mixture framework for the modeling of waste disposal by mimicking the underlying data generating process. The proposed mixture model is significantly better at explaining the observed in-data variance compared to a simple count data model. It maintains modeling simplicity, is based on a solid and intuitive approach, and has few exogenous assumptions. Future research will focus on reformulating the objective function as a maximum likelihood function and analyzing the model's performance for an larger set of disposal quantities and for a continuous distribution of the disposal quantities.

Acknowledgement

This work is partially funded by the Commission for Technology and Innovation (CTI) under grant number CTI 15781.1 PFES-ES. The authors would also like to thank EcoWaste SA (http://www.ecowaste.ch/) for their expert contribution and the problem data.

6 References

Beigl, P., S. Lebersorger and S. Salhofer (2008) Modelling municipal solid waste generation: A review, *Waste Management*, **28** (1) 200–214.

Benítez, S. O., G. Lozano-Olvera, R. A. Morelos and C. A. de Vega (2008) Mathematical modeling to predict residential solid waste generation, *Waste Management*, **28** (Supplement 1) S7 – S13.

Bitsch, B. (2012) Inventory routing with stochastic demand, Master Thesis, Aarhus School of Business and Social Sciences, Aarhus University, Aarhus, Denmark.

Chen, H. and N.-B. Chang (2000) Prediction analysis of solid waste generation based on grey fuzzy dynamic modeling, *Resources, Conservation and Recycling*, **29** (1–2) 1–18.

Dyson, B. and N.-B. Chang (2005) Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling, *Waste Management*, **25** (7) 669–679.

Faccio, M., A. Persona and G. Zanin (2011) Waste collection multi objective model with real time traceability data, *Waste Management*, **31** (12) 2391 – 2405.

Johansson, O. M. (2006) The effect of dynamic scheduling and routing in a solid waste management system, *Waste Management*, **26** (8) 875–885.

Krikke, H., I. le Blanc, M. van Krieken and H. Fleuren (2008) Low-frequency collection of materials disassembled from end-of-life vehicles: On the value of on-line monitoring in optimizing route planning, *International Journal of Production Economics*, **111** (2) 209 – 228. Special Section on Sustainable Supply Chain.

Li, Z.-s., H.-z. Fu and X.-y. Qu (2011) Estimating municipal solid waste generation by different activities and various resident groups: A case study of Beijing, *Science of The Total Environment*, **409** (20) 4406–4414.

McLeod, F., G. Erdoğan, T. Cherrett, T. Bektaş, N. Davies, C. Speed, J. Dickinson and S. Norgate (2013) Dynamic collection scheduling using remote asset monitoring, *Transportation Research Record*, **2378**, 65–72.

Mes, M. (2012) Using simulation to assess the opportunities of dynamic waste collection, in S. Bangsow (ed.) *Use Cases of Discrete Event Simulation*, 277–307, Springer Berlin Heidelberg.

Mes, M., M. Schutten and A. P. Rivera (2014) Inventory routing for dynamic waste collection, *Waste Management*, **34** (9) 1564–1576.

Navarro-Esbrí, J., E. Diamadopoulos and D. Ginestar (2002) Time series analysis and forecasting techniques for municipal solid waste management, *Resources, Conservation and Recycling*, **35** (3) 201–214.

Nolz, P. C., N. Absi and D. Feillet (2011) Optimization of infectious medical waste collection using RFID, in J. W. Böse, H. Hu, C. Jahn, X. Shi, R. Stahlbock and S. Voß (eds.) *Computational Logistics*, vol. 6971 of *Lecture Notes in Computer Science*, 86–100, Springer Berlin Heidelberg.

Nolz, P. C., N. Absi and D. Feillet (2014) A stochastic inventory routing problem for infectious medical waste collection, *Networks*, **63** (1) 82–95.

Xu, L., P. Gao, S. Cui and C. Liu (2013) A hybrid procedure for MSW generation forecasting at multiple time scales in Xiamen City, China, *Waste Management*, **33** (6) 1324–1331.