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

[b]Haute École de Gestion de Genève
University of Applied Sciences Western Switzerland (HES-SO)

15th Swiss Transport Research Conference (STRC)
Monte Verità / Ascona, April 15–17, 2015

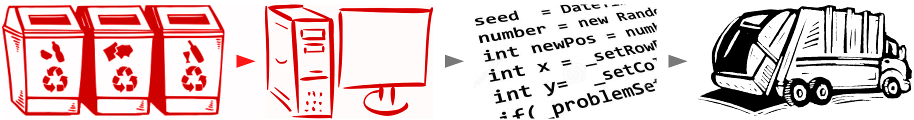
Overview

Overview

Efficient collection of recyclables in Geneva

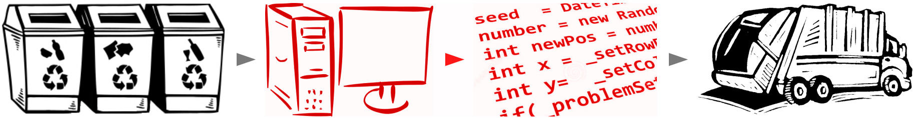
In more detail...

- Sensorized containers periodically send waste level data to a centralized database



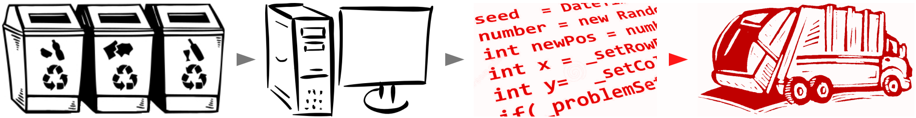
In more detail...

- Sensorized containers periodically send waste level data to a centralized database
- Level data is used for container selection and vehicle routing, with tours *often planned several days in advance*



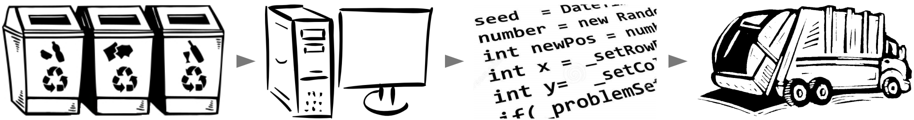
In more detail...

- Sensorized containers periodically send waste level data to a centralized database
- Level data is used for container selection and vehicle routing, with tours *often planned several days in advance*
- Vehicles are dispatched to carry out the daily schedules produced by the routing algorithm



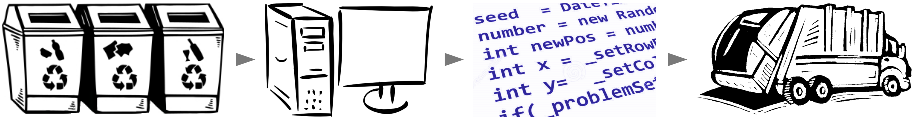
In more detail...

- Sensorized containers periodically send waste level data to a centralized database
- Level data is used for container selection and vehicle routing, with tours *often planned several days in advance*
- Vehicles are dispatched to carry out the daily schedules produced by the routing algorithm
- Efficient waste collection thus depends on the ability to:
    - make good forecasts of the container levels at the time of collection
    - and optimally route the vehicles to service the selected containers
- In this talk we will focus on the first part, i.e. *short-term* operational container level forecasting



In more detail...

- Sensorized containers periodically send waste level data to a centralized database
- Level data is used for container selection and vehicle routing, with tours *often planned several days in advance*
- Vehicles are dispatched to carry out the daily schedules produced by the routing algorithm
- Efficient waste collection thus depends on the ability to:
  - make good forecasts of the container levels at the time of collection
  - and optimally route the vehicles to service the selected containers
- In this talk we will focus on the first part, i.e. *short-term* operational container level forecasting

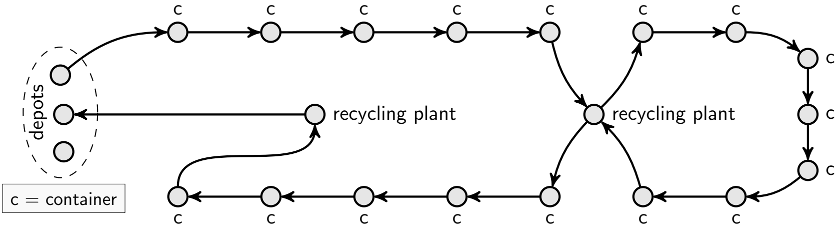


Overview

Routing problem illustration

- The routing problem was presented at STRC 2014 (Markov et al., 2014)
- It is a rich VRP with intermediate facilities, which integrates:
    - a heterogeneous fixed fleet with fixed and vairable costs
    - a flexible assignment of start and end depot
- The constraints and features are inspired by practical applications to collectors in Switzerland and France

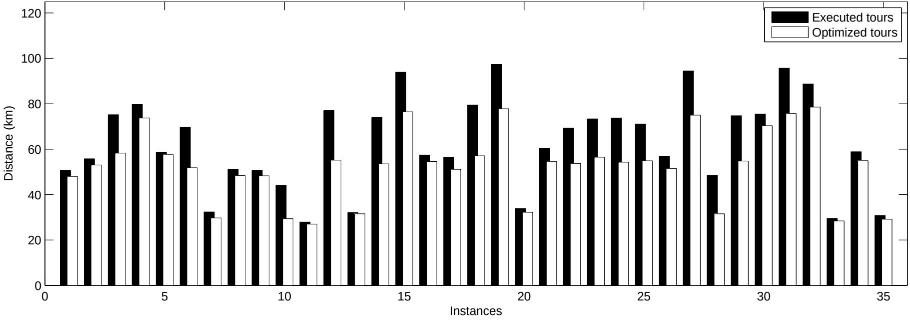
Figure 1: Example of a tour



Solution and results

- The problem was modeled as a MILP
- It was solved using a local search algorithm
- Applied to a set of executed tours for collecting white glass and PET in Geneva, it reduced travel distance by 15% on average

Figure 2: Executed vs. optimized tours in Geneva



Overview

Literature

- The literature on waste generation forecasting is abundant and varied (for a survey see Beigl et al., 2008)
- Much of it is focused on city and regional level: Tainan, Taiwan (Chen and Chang, 2000); San Antonio, US (Dyson and Chang, 2005); Beijing, China (Li et al., 2011), etc...

Literature

- The literature on waste generation forecasting is abundant and varied (for a survey see Beigl et al., 2008)
- Much of it is focused on city and regional level: Tainan, Taiwan (Chen and Chang, 2000); San Antonio, US (Dyson and Chang, 2005); Beijing, China (Li et al., 2011), etc...
- And a fairly small amount on the container (micro) level, e.g.:
  - Inventory levels in pharmacies (Nolz et al., 2011, 2014)
  - Recyclable materials from old cars (Krikke et al., 2008)
  - Charity donation banks (McLeod et al., 2013)
  - Waste container levels (Johansson, 2006; Faccio et al., 2011; Mes, 2012; Mes et al., 2014)

Literature

- The literature on waste generation forecasting is abundant and varied (for a survey see Beigl et al., 2008)
- Much of it is focused on city and regional level: Tainan, Taiwan (Chen and Chang, 2000); San Antonio, US (Dyson and Chang, 2005); Beijing, China (Li et al., 2011), etc...
- And a fairly small amount on the container (micro) level, e.g.:
    - Inventory levels in pharmacies (Nolz et al., 2011, 2014)
    - Recyclable materials from old cars (Krikke et al., 2008)
    - Charity donation banks (McLeod et al., 2013)
    - Waste container levels (Johansson, 2006; Faccio et al., 2011; Mes, 2012; Mes et al., 2014)
- Contribution:
    - Operational container level forecasting
    - We develop a forecasting model estimated and validated on real data, whereas most of the container level literature is focused on critical levels. Moreover, much of it uses simulated data.

Overview

Data preparation

- Container levels are:
  - detected by internal ultrasound sensors
  - periodically transmitted to a central database
  - post-processed for noise removal
  - extrapolated at the end of each date
- Let $L_{i,t}$ denote the level of container i at the end of date t
- Let C_i denote the usable capacity of container i
- Then the observed quantity deposited in container i at date t is:

$$Q_{i,t} = C_i(L_{i,t} - L_{i,t-1}) \tag{1}$$

- In case there was an emptying event at date t, we have:

$$Q_{i,t} = C_i L_{i,t} \tag{2}$$

Formulation

- Let $n_{i,t,k}$ denote the number of deposits in container i at date t of size q_k. We define the data generating process as follows:

$$Q^{\star}_{i,t} = \sum\nolimits_{k=1}^{K} n_{i,t,k} q_k \tag{3}$$

- Let $n_{i,t,k} \xrightarrow{\text{iid}} \mathcal{P}\left(\lambda_{i,t,k}\right)$ with probability $\pi_{i,t,k}$. Then we obtain:

$$\mathbb{E}\left(Q^{\star}_{i,t}\right) = \sum\nolimits_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k} \tag{4}$$

- We minimize the sum of squared differences between observed and expected over all containers and dates:

$$\min_{\boldsymbol{\lambda},\boldsymbol{\pi}} \sum\nolimits_{i=1}^{N} \sum\nolimits_{t=1}^{T} \left(Q_{i,t} - \sum\nolimits_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k}\right)^2 \tag{5}$$

assuming strict exogeneity

Formulation

- Given vectors of covariates $\mathbf{x}_{i,t}$ and $\mathbf{z}_{i,t}$ and vectors of parameters $\boldsymbol{\beta}_k$ and γ_k, we define Poisson rates and logit-type probabilities:

$$\lambda_{i,t,k}(\boldsymbol{\theta}) = \exp\left(\mathbf{x}_{i,t}^{\mathsf{T}}\boldsymbol{\beta}_k\right) \tag{6}$$

$$\pi_{i,t,k}(\boldsymbol{\theta}) = \frac{\exp\left(\mathbf{z}_{i,t}^{\mathsf{T}}\gamma_k\right)}{\sum_{j=1}^{K}\exp\left(\mathbf{z}_{i,t}^{\mathsf{T}}\gamma_j\right)} \tag{7}$$

- Then, in compact form, the minimization problem writes as:

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{N}\sum_{t=1}^{T}\left(Q_{i,t} - \sum_{k=1}^{K}\frac{\exp\left(\mathbf{x}_{i,t}^{\mathsf{T}}\boldsymbol{\beta}_k + \mathbf{z}_{i,t}^{\mathsf{T}}\gamma_k + \ln\left(q_k\right)\right)}{\sum_{j=1}^{K}\exp\left(\mathbf{z}_{i,t}^{\mathsf{T}}\gamma_j\right)}\right)^2 \tag{8}$$

- $\boldsymbol{\Theta} := (\boldsymbol{\beta}_k, \gamma_k : \forall k)$, and $\gamma_{k^\star} = \mathbf{0}$ for one arbitrarily chosen $k^\star$
- We will refer to this minimization problem as the *mixture model*

Formulation

- In case of only one deposit quantity, it degenerates to a pseudo-count data process:

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Q_{i,t} - \exp\left(\mathbf{x}_{i,t}^{\mathsf{T}}\boldsymbol{\beta} + \ln(q)\right)\right)^2 \tag{9}$$

- We will refer to this minimization problem as the *simple model*

Forecasting

- Using new sets of covariates $\dot{\mathbf{x}}_{i,t}$ and $\dot{\mathbf{z}}_{i,t}$, and the estimates $\hat{\boldsymbol{\beta}}_k$ and $\hat{\gamma}_k$, we can generate a forecast as follows:

$$\dot{Q}_{i,t} = \sum_{k=1}^{K} \frac{\exp\left(\dot{\mathbf{x}}_{i,t}^{\top}\hat{\boldsymbol{\beta}}_k + \dot{\mathbf{z}}_{i,t}^{\top}\hat{\gamma}_k + \ln\left(q_k\right)\right)}{\sum_{j=1}^{K}\exp\left(\dot{\mathbf{z}}_{i,t}^{\top}\hat{\gamma}_j\right)} \tag{10}$$

- Given the operational nature of the problem, the covariates should be quick and easy to obtain
- Examples include days of the week, months, weather data, holidays, etc...

Overview

Data

- 36 containers for PET in the canton of Geneva with capacity of 3040 or 3100 liters
- Balanced panel covering March to June, 2014 (122 days), which brings the total number of observations to 4392
- The final sample excludes unreliable level data (removed after visual inspection)
- Missing data is linearly interpolated for the values of $Q_{i,t}$

Residual plots

Figure 3: Residual plot of the mixture model

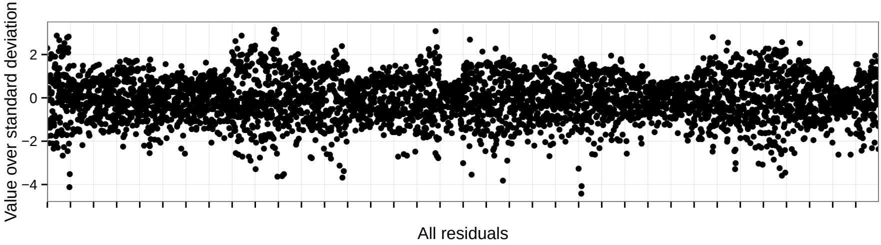
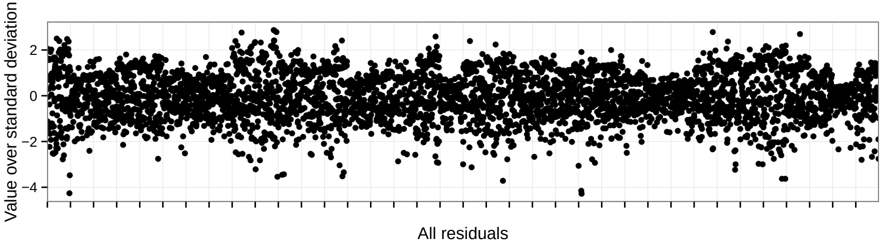


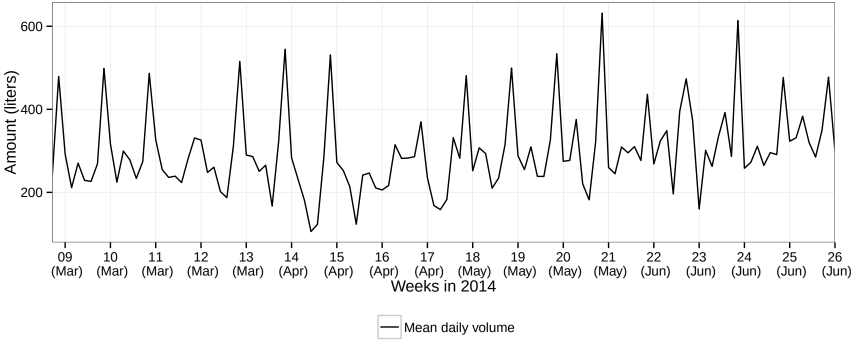
Figure 4: Residual plot of the simple model



Seasonality pattern

- Waste generation exhibits strong weekly seasonality
- Peaks are observed during the weekends
- There also appear to be longer-term effects for months

Figure 5: Mean daily volume deposited in the containers



Covariates

- Based on the above observations, we use the following covariates
- They are all used both for $\mathbf{x}_{i,t}$ (rates) and $\mathbf{z}_{i,t}$ (probabilities)

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

Evaluating the fits

- Coefficient of determination

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

  with higher values for a better model
- Akaike information criterion (AIC):

$$\text{AIC} = \left(\frac{SS_{\text{res}}}{N}\right) \exp(2K/N) \tag{12}$$

  with lower values for a better model. The exponential penalizes model complexity
- SS_{res} is the residual sum of squares
- SS_{tot} is the total sum of squares
- K is the number of estimated parameters
- N is the number of observations

Estimation on full sample

- Mixture model: R^2 of 0.341 (AIC 52900)
- Simple model: R^2 of 0.300 (AIC 53700)

Table 2: Estimated coefficients of mixture model

	$\hat{\beta}_1$ (5L)***	$\hat{\beta}_2$ (15L)***	$\hat{\gamma}_2$***
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

Validation

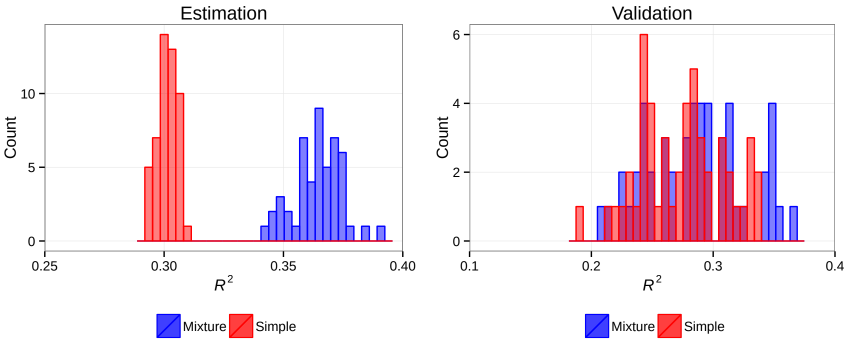
- We performed 50 experiments
- Both the mixture and the simple model are estimated on a random sample of 90% of the panel
- They are validated on the remaining 10%
- It was made sure that all containers and all months appeared in the random samples

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

Validation

Figure 6: Histograms for estimation and validation samples



Overview

Conclusion

- Mixture model representing the data generating process of a realistic underlying behavior
- Preliminary testing shows its better in- and out-of-sample performance
- Future research will focus on:
  - reformulating the objective function as a likelihood function
  - testing a higher number if discrete deposit sizes
  - and a continuous distribution of the deposit size
  - integrating the forecasting approach and the vehicle routing algorithm into an inventory routing platform

Thank you for your attention!
Questions?

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

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

Faccio, M., Persona, A., and Zanin, G. (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., le Blanc, I., van Krieken, M., and Fleuren, H. (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., Fu, H.-z., and Qu, X.-y. (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.

Markov, I., Varone, S., and Bierlaire, M. (2014). Vehicle routing for a complex waste collection problem. In *Proceedings of the 14th Swiss Transport Research Conference (STRC)*, Ascona, Switzerland.

McLeod, F., Erdoğan, G., Cherrett, T., Bektaş, T., Davies, N., Speed, C., Dickinson, J., and Norgate, S. (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 Bangsow, S., editor, *Use Cases of Discrete Event Simulation*, pages 277–307. Springer Berlin Heidelberg.

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

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

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