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

Iliya Markov\textsuperscript{a,b}, Matthieu de Lapparent\textsuperscript{a}, Michel Bierlaire\textsuperscript{a}, Sacha Varone\textsuperscript{b}

\textsuperscript{a}Transport and Mobility Laboratory  
School of Architecture, Civil and Environmental Engineering  
École Polytechnique Fédérale de Lausanne

\textsuperscript{b}Haute École de Gestion de Genève  
University of Applied Sciences Western Switzerland (HES-SO)

15\textsuperscript{th} Swiss Transport Research Conference (STRC)  
Monte Verità / Ascona, April 15–17, 2015
Overview

1. Introduction
2. Background
3. Literature
4. Methodology
5. Numerical Experiments
6. Conclusion
Efficient collection of recyclables in Geneva
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*.
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

1. Introduction
2. Background
3. Literature
4. Methodology
5. Numerical Experiments
6. Conclusion
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 variable 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

1. Introduction
2. Background
3. Literature
4. Methodology
5. Numerical Experiments
6. Conclusion
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...
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)
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

1. Introduction
2. Background
3. Literature
4. Methodology
5. Numerical Experiments
6. Conclusion
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:
  \[ Q_{i,t} = C_i(L_{i,t} - L_{i,t-1}) \] (1)
- In case there was an emptying event at date $t$, we have:
  \[ Q_{i,t} = C_i L_{i,t} \] (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:

\[
Q^*_{i,t} = \sum_{k=1}^{K} n_{i,t,k} q_k
\]  
(3)

- Let \( n_{i,t,k} \overset{iid}{\to} P(\lambda_{i,t,k}) \) with probability \( \pi_{i,t,k} \). Then we obtain:

\[
\mathbb{E}(Q^*_{i,t}) = \sum_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k}
\]  
(4)

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

\[
\min_{\lambda, \pi} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( Q_{i,t} - \sum_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k} \right)^2
\]  
(5)

assuming strict exogeneity
Formulation

- Given vectors of covariates $x_{i,t}$ and $z_{i,t}$ and vectors of parameters $\beta_k$ and $\gamma_k$, we define Poisson rates and logit-type probabilities:

  \[ \lambda_{i,t,k}(\theta) = \exp\left(x_{i,t}^T \beta_k\right) \]  
  \[ \pi_{i,t,k}(\theta) = \frac{\exp\left(z_{i,t}^T \gamma_k\right)}{\sum_{j=1}^{K} \exp\left(z_{i,t}^T \gamma_j\right)} \]

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

  \[ \min_{\theta \in \Theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( Q_{i,t} - \sum_{k=1}^{K} \frac{\exp\left(x_{i,t}^T \beta_k + z_{i,t}^T \gamma_k + \ln(q_k)\right)}{\sum_{j=1}^{K} \exp\left(z_{i,t}^T \gamma_j\right)} \right)^2 \]  

  \[ \Theta := (\beta_k, \gamma_k : \forall k), \text{ and } \gamma_k^* = 0 \text{ for one arbitrarily chosen } k^* \]

- We will refer to this minimization problem as the mixture model
In case of only one deposit quantity, it degenerates to a pseudo-count data process:

\[
\min_{\theta \in \Theta} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( Q_{i,t} - \exp \left( x_{i,t}^T \beta + \ln(q) \right) \right)^2
\]

We will refer to this minimization problem as the \textit{simple model}.
Forecasting

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

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

(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

1. Introduction
2. Background
3. Literature
4. Methodology
5. Numerical Experiments
6. Conclusion
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 $x_{i,t}$ (rates) and $z_{i,t}$ (probabilities)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container fixed effect</td>
<td>dummy</td>
</tr>
<tr>
<td>Day of the week</td>
<td>dummy</td>
</tr>
<tr>
<td>Month</td>
<td>dummy</td>
</tr>
<tr>
<td>Minimum temperature in Celsius</td>
<td>continuous</td>
</tr>
<tr>
<td>Precipitation in mm</td>
<td>continuous</td>
</tr>
<tr>
<td>Pressure in hPa</td>
<td>continuous</td>
</tr>
<tr>
<td>Wind speed in kmph</td>
<td>continuous</td>
</tr>
</tbody>
</table>
Evaluating the fits

- Coefficient of determination

\[ R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \]  

(11)

with higher values for a better model

- Akaike information criterion (AIC):

\[ AIC = \left( \frac{SS_{res}}{N} \right) \exp\left(\frac{2K}{N}\right) \]  

(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

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_1$ (5L)***</th>
<th>$\hat{\beta}_2$ (15L)***</th>
<th>$\hat{\gamma}_2$***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum temperature in Celsius</td>
<td>1461.356</td>
<td>0.022</td>
<td>-0.037</td>
</tr>
<tr>
<td>Precipitation in mm</td>
<td>-0.821</td>
<td>-0.009</td>
<td>0.018</td>
</tr>
<tr>
<td>Pressure in hPa</td>
<td>-13.724</td>
<td>-0.001</td>
<td>0.010</td>
</tr>
<tr>
<td>Wind speed in kmph</td>
<td>7.580</td>
<td>-0.004</td>
<td>0.020</td>
</tr>
<tr>
<td>Monday</td>
<td>402.235</td>
<td>2.166</td>
<td>-9.693</td>
</tr>
<tr>
<td>Tuesday</td>
<td>1908.233</td>
<td>2.293</td>
<td>-9.977</td>
</tr>
<tr>
<td>Wednesday</td>
<td>-844.662</td>
<td>1.432</td>
<td>0.202</td>
</tr>
<tr>
<td>Thursday</td>
<td>1937.385</td>
<td>1.198</td>
<td>1.453</td>
</tr>
<tr>
<td>Friday</td>
<td>1876.162</td>
<td>1.239</td>
<td>4.419</td>
</tr>
<tr>
<td>Saturday</td>
<td>-6981.339</td>
<td>1.358</td>
<td>4.723</td>
</tr>
<tr>
<td>Sunday</td>
<td>1831.715</td>
<td>1.905</td>
<td>2.832</td>
</tr>
<tr>
<td>March</td>
<td>-27.136</td>
<td>2.955</td>
<td>-1.453</td>
</tr>
<tr>
<td>April</td>
<td>1071.406</td>
<td>2.746</td>
<td>-1.532</td>
</tr>
<tr>
<td>May</td>
<td>1689.979</td>
<td>2.988</td>
<td>-1.603</td>
</tr>
<tr>
<td>June</td>
<td>-2604.520</td>
<td>2.901</td>
<td>-1.452</td>
</tr>
</tbody>
</table>
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>
<thead>
<tr>
<th></th>
<th>Mixture model mean $R^2$</th>
<th>Simple model mean $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation</td>
<td>0.364 (AIC 51400)</td>
<td>0.302 (AIC 53600)</td>
</tr>
<tr>
<td>Validation</td>
<td>0.286</td>
<td>0.274</td>
</tr>
</tbody>
</table>

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

Figure 6: Histograms for estimation and validation samples

![Estimation and Validation Histograms](image-url)
Overview

1. Introduction
2. Background
3. Literature
4. Methodology
5. Numerical Experiments
6. Conclusion
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 of 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?


