Routing and Forecasting in the Collection of Recyclables

Decision-aid Methodologies in Transportation

Iliya Markov

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
École Polytechnique Fédérale de Lausanne

Joint work with
Matthieu de Lapparent, Michel Bierlaire, Sacha Varone

May 26, 2015

Contents

1 Introduction

2 Routing

3 Forecasting

4 Conclusion

Contents

1 Introduction

2 Routing

3 Forecasting

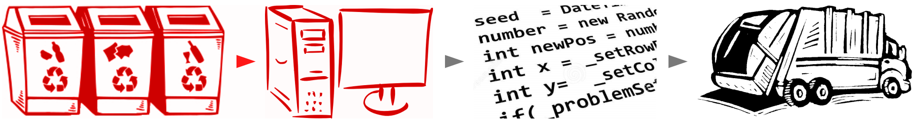
4 Conclusion

Ecological Waste Management

*ecopoint in Rue de Neuchâtel, Geneva; photo source: self

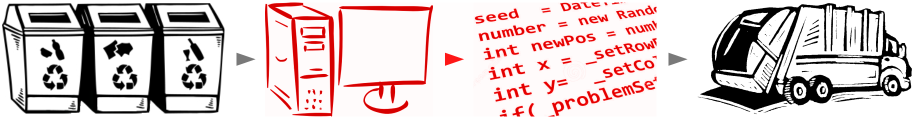
How it works...

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



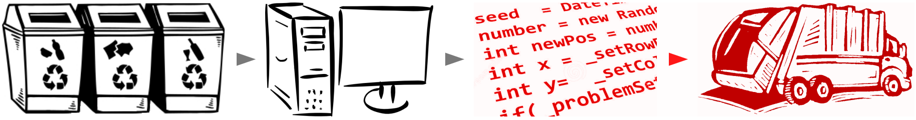
How it works...

- 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*



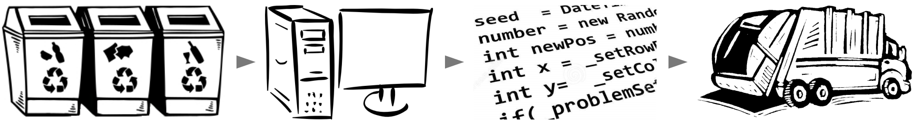
How it works...

- 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



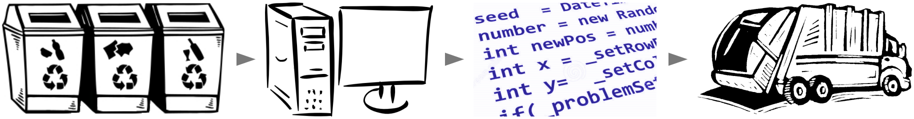
How it works...

- 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



How it works...

- 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



Contents

1 Introduction

2 Routing

3 Forecasting

4 Conclusion

Problem description

- A heterogeneous fixed fleet with different:
  - volume capacities
  - weight capacities
  - fixed costs
  - unit-distance running costs
  - unit-time driver wage rates
  - speeds
  - site dependencies (accessibility constraints)

Problem description

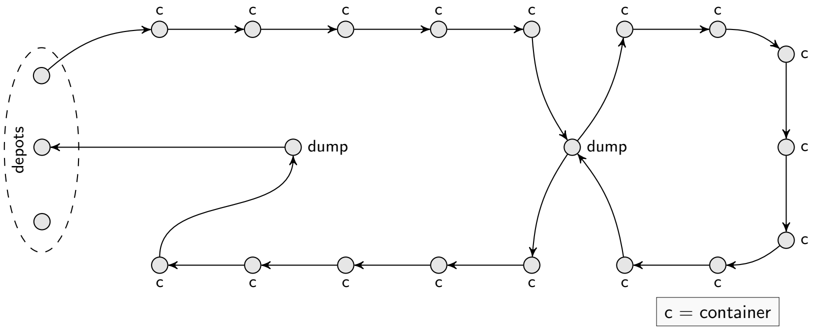
- A heterogeneous fixed fleet with different:
  - volume capacities
  - weight capacities
  - fixed costs
  - unit-distance running costs
  - unit-time driver wage rates
  - speeds
  - site dependencies (accessibility constraints)
- A set of depots
- A set of containers placed at collection points with time windows
- A set of dumps (recycling plants) with time windows

Problem description

- A heterogeneous fixed fleet with different:
  - volume capacities
  - weight capacities
  - fixed costs
  - unit-distance running costs
  - unit-time driver wage rates
  - speeds
  - site dependencies (accessibility constraints)
- A set of depots
- A set of containers placed at collection points with time windows
- A set of dumps (recycling plants) with time windows
- Maximum tour duration, interrupted by a break
- A tour is a sequence of collections and disposals at the available dumps, with a mandatory disposal before the end of the tour
- A tour need not finish at the depot it started from

Problem description

Figure 1: Tour illustration



Formulation

Sets

O' = set of origins
O'' = set of destinations
D = set of dumps
P = set of containers
$N = O' \cup O'' \cup D \cup P$
K = set of vehicles

Parameters

π_{ij} = length of edge (i, j)
α_{ijk} = 1 if edge (i, j) is accessible for vehicle k, 0 otherwise
τ_{ijk} = travel time of vehicle k on edge (i, j)
ϵ_i = service duration at point i
$[\lambda_i, \mu_i]$ = time window lower and upper bound at point i
H = maximum tour duration
η = maximum continuous work limit after which a break is due
δ = break duration
ρ_i^v, ρ_i^w = volume and weight pickup quantity at point i
Ω_k^v, Ω_k^w = volume and weight capacity of vehicle k
ϕ_k = fixed cost of vehicle k
β_k = unit-distance running cost of vehicle k
θ_k = unit-time wage rate of vehicle k

Formulation

Decision variables: binary

$$x_{ijk} = \begin{cases} 1 & \text{if vehicle } k \text{ traverses edge } (i,j) \\ 0 & \text{otherwise} \end{cases}$$

$$b_{ijk} = \begin{cases} 1 & \text{if vehicle } k \text{ takes a break on edge } (i,j) \\ 0 & \text{otherwise} \end{cases}$$

$$y_k = \begin{cases} 1 & \text{if vehicle } k \text{ is used} \\ 0 & \text{otherwise} \end{cases}$$

Decision variables: continuous

S_{ik} = start-of-service time of vehicle k at point i

Q^v_{ik} = cumulative volume on vehicle k at point i

Q^w_{ik} = cumulative weight on vehicle k at point i

Formulation

$$
\begin{aligned}
\text{Min} \quad & f = \sum_{k \in K} \left(\phi_k y_k + \beta_k \sum_{i \in N} \sum_{j \in N} \pi_{ij} x_{ijk} + \theta_k \left(\sum_{j \in O''} S_{jk} - \sum_{i \in O'} S_{ik} \right) \right) && && (1) \\
\text{s.t.} \quad & \sum_{k \in K} \sum_{j \in D \cup P} x_{ijk} = 1, && \forall i \in P && (2) \\
& \sum_{i \in O'} \sum_{j \in P} x_{ijk} = y_k, && \forall k \in K && (3) \\
& \sum_{i \in D} \sum_{j \in O''} x_{ijk} = y_k, && \forall k \in K && (4) \\
& \sum_{i \in N} x_{ijk} = 0, && \forall k \in K, j \in O' && (5) \\
& \sum_{j \in N} x_{ijk} = 0, && \forall k \in K, i \in O'' && (6) \\
& \sum_{i \in N \setminus O''} x_{ijk} = \sum_{i \in N \setminus O'} x_{jik}, && \forall k \in K, j \in D \cup P && (7) \\
& x_{ijk} \leqslant \alpha_{ijk}, && \forall k \in K, i \in N \setminus O'', j \in N \setminus O' && (8)
\end{aligned}
$$

Formulation

$$
\begin{aligned}
\text{s.t.}\quad & Q_{ik}^{v} \leqslant \Omega_{k}^{v}, && \forall k \in K, i \in P && (9)\\
& Q_{ik}^{w} \leqslant \Omega_{k}^{w}, && \forall k \in K, i \in P && (10)\\
& Q_{ik}^{v} = 0, && \forall k \in K, i \in N \setminus P && (11)\\
& Q_{ik}^{w} = 0, && \forall k \in K, i \in N \setminus P && (12)\\
& Q_{ik}^{v} + \rho_{j}^{v} \leqslant Q_{jk}^{v} + \left(1 - x_{ijk}\right) M, && \forall k \in K, i \in N \setminus O'', j \in P && (13)\\
& Q_{ik}^{w} + \rho_{j}^{w} \leqslant Q_{jk}^{w} + \left(1 - x_{ijk}\right) M, && \forall k \in K, i \in N \setminus O'', j \in P && (14)\\
& S_{ik} + \epsilon_i + \delta b_{ijk} + \tau_{ijk} \leqslant S_{jk} + \left(1 - x_{ijk}\right) M, && \forall k \in K, i \in N \setminus O'', j \in N \setminus O' && (15)\\
& \left(S_{ik} - \sum_{m \in O'} S_{mk}\right) + \epsilon_i - \eta \leqslant \left(1 - b_{ijk}\right) M, && \forall k \in K, i \in N \setminus O'', j \in N \setminus O' && (16)\\
& \eta - \left(S_{jk} - \sum_{m \in O'} S_{mk}\right) \leqslant \left(1 - b_{ijk}\right) M, && \forall k \in K, i \in N \setminus O'', j \in N \setminus O' && (17)\\
& b_{ijk} \leqslant x_{ijk}, && \forall k \in K, i, j \in N && (18)\\
& \left(\sum_{j \in O''} S_{jk} - \sum_{i \in O'} S_{ik}\right) - \eta \leqslant \left(\sum_{\substack{i \in N \setminus O'' \\ j \in N \setminus O'}} b_{ijk}\right) M, && \forall k \in K && (19)
\end{aligned}
$$

Formulation

$$\text{s.t.} \quad \lambda_i \sum_{j \in N \setminus O'} x_{ijk} \leqslant S_{ik} \leqslant \mu_i \sum_{j \in N \setminus O'} x_{ijk}, \qquad \forall k \in K, i \in N \setminus O'' \tag{20}$$

$$\sum_{j \in O''} S_{jk} - \sum_{i \in O'} S_{ik} \leqslant \mathrm{H}, \qquad \forall k \in K \tag{21}$$

$$x_{ijk}, y_k, b_{ijk} \in \{0, 1\}, \qquad \forall k \in K, i, j \in N \tag{22}$$

$$Q_{ik}^v, Q_{ik}^w, S_{ik} \geqslant 0, \qquad \forall k \in K, i \in N \tag{23}$$

Formulation: Relocation decisions

$$z_{ijk} = \begin{cases} 1 & \text{if } i \text{ is the origin and } j \text{ the destination of vehicle } k \\ 0 & \text{otherwise} \end{cases}$$

Ψ = weight of relocation term

$$\text{Min} \quad f = \text{Objective (1)} + \Psi \sum_{k \in K} \sum_{i \in O'} \sum_{j \in O''} \left(\beta_k \pi_{ji} + \theta_k \tau_{jik} \right) z_{ijk} \tag{24}$$

s.t. Constraints (2) to (23)

$$\sum_{m \in P} x_{imk} + \sum_{m \in D} x_{mjk} - 1 \leqslant z_{ijk}, \qquad \forall k \in K, i \in O', j \in O'' \tag{25}$$

$$z_{ijk} = \{0, 1\}, \qquad \forall k \in K, i \in O', j \in O'' \tag{26}$$

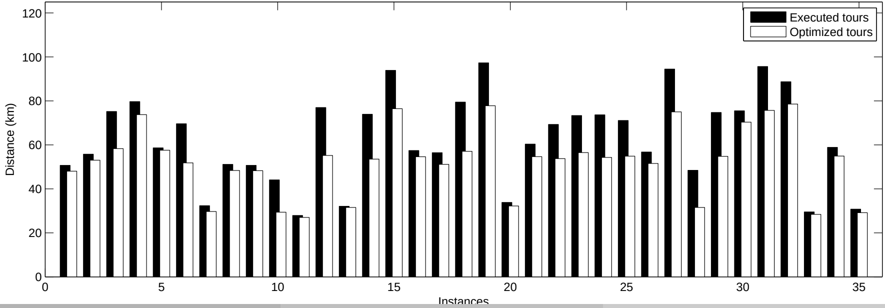
Solution methodology

- Strengthened with valid inequalities, the formulation is capable of solving problems with 10-15 containers, 4-5 dumps, and several vehicles, often with excessive computation time.
- For larger instances, we developed a local search heuristic which will not be presented here.
- It has been integrated and is currently being tested by a Swiss software-as-a-service provider for collectors.

Results: Comparison to the state of practice

- 35 tours planned by specialized software for the canton of Geneva
- 7 to 38 containers per tour, up to 4 dump visits per tour
- LS heuristic improvements range from 1.73% to 34.91%, on average 14.75%, with running time in the order of a few seconds.

Figure 2: Comparison to the state of practice (average of 10 runs per instance)



Contents

1 Introduction

2 Routing

3 Forecasting

4 Conclusion

Methodology

- 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:

$$Q^\star_{i,t} = \sum_{k=1}^{K} n_{i,t,k} q_k \tag{27}$$

- 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:

$$\mathbb{E}\left(Q^\star_{i,t}\right) = \sum_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k} \tag{28}$$

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

$$\min_{\boldsymbol{\lambda},\boldsymbol{\pi}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Q_{i,t} - \sum_{k=1}^{K} q_k \lambda_{i,t,k} \pi_{i,t,k}\right)^2 \tag{29}$$

assuming strict exogeneity

Methodology

- 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:

$$\lambda_{i,t,k}\left(\boldsymbol{\theta}\right) = \exp\left(\mathbf{x}_{i,t}^{\mathsf{T}}\boldsymbol{\beta}_k\right) \tag{30}$$

$$\pi_{i,t,k}\left(\boldsymbol{\theta}\right) = \frac{\exp\left(\mathbf{z}_{i,t}^{\mathsf{T}}\gamma_k\right)}{\sum_{j=1}^{K}\exp\left(\mathbf{z}_{i,t}^{\mathsf{T}}\gamma_j\right)} \tag{31}$$

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

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Q_{i,t} - \sum_{k=1}^{K}\frac{\exp\left(\mathbf{x}_{i,t}^{\mathsf{T}}\boldsymbol{\beta}_k + \mathbf{z}_{i,t}^{\mathsf{T}}\gamma_k + \ln\left(q_k\right)\right)}{\sum_{j=1}^{K}\exp\left(\mathbf{z}_{i,t}^{\mathsf{T}}\gamma_j\right)}\right)^2 \tag{32}$$

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

Methodology

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

$$\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Q_{i,t} - \exp\left(\mathbf{x}_{i,t}^{\mathsf{T}}\boldsymbol{\beta} + \ln(q)\right)\right)^2 \tag{33}$$

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

Methodology

- 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:

$$\dot{Q}_{i,t} = \sum_{k=1}^{K} \frac{\exp\left(\dot{\mathbf{x}}_{i,t}^{\top}\hat{\boldsymbol{\beta}}_k + \dot{\mathbf{z}}_{i,t}^{\top}\hat{\gamma}_k + \ln\left(q_k\right)\right)}{\sum_{j=1}^{K} \exp\left(\dot{\mathbf{z}}_{i,t}^{\top}\hat{\gamma}_j\right)} \tag{34}$$

- 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...

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

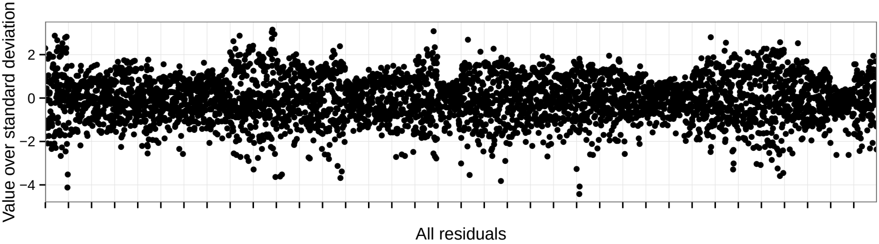
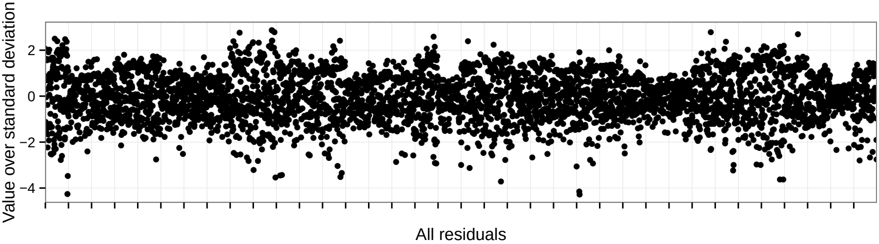


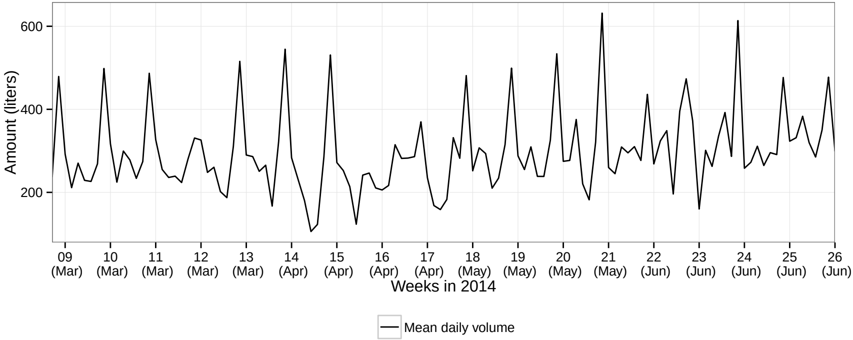
Figure 4: Residual plot of the simple model



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

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{35}$$

  with higher values for a better model

- Akaike information criterion (AIC):

$$\text{AIC} = \left(\frac{SS_{\text{res}}}{N}\right) \exp(2K/N) \tag{36}$$

  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) with 5L and 15L
- Simple model: R^2 of 0.300 (AIC 53700) with 10L

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

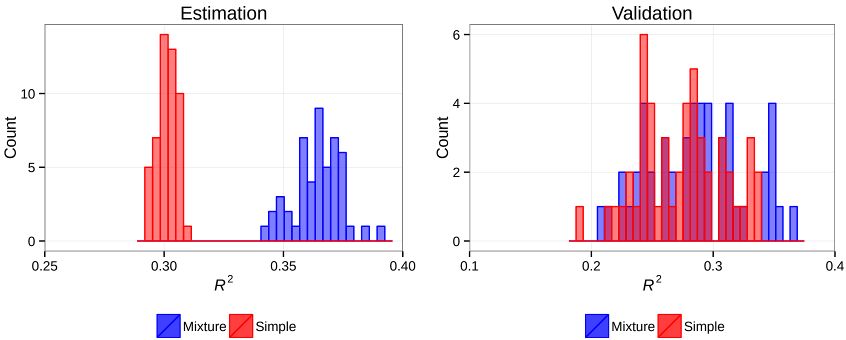
- We performed 50 experiments
- Both the mixture and the simple model are estimated on a random sample of 90% of the panel
- They are validated on the remaining 10%
- It was made sure that all containers and all months appeared in the random samples

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



Contents

1 Introduction

2 Routing

3 Forecasting

4 Conclusion

Conclusion

- At the moment, the forecasting model can produce future levels, for which the routing problem is solved.
- Future research will focus on integrating the forecasting model and the routing algorithm into an inventory routing problem (IRP).
- IRP solves simultaneously the container selection problem based on forecast levels and the routing problem in a periodic framework.
- The increasing amount of available data will allow for more extensive testing and results.

Thank you.
Questions?