



**HAL**  
open science

# Medium-term public transit route ridership forecasting: What, how and why? A case study in Lyon

Oscar Egu, Patrick Bonnel

## ► To cite this version:

Oscar Egu, Patrick Bonnel. Medium-term public transit route ridership forecasting: What, how and why? A case study in Lyon. *Transport Policy*, 2021, 105, pp.124-133. 10.1016/j.tranpol.2021.03.002 . halshs-04233578

**HAL Id: halshs-04233578**

**<https://shs.hal.science/halshs-04233578v1>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 Medium-term public transit route ridership forecasting:  
2 what, how and why ?  
3 A case study in Lyon

4 Oscar Egu<sup>1</sup> and Patrick Bonnel<sup>1,\*</sup>

5 <sup>1</sup>*LAET-ENTPE, Université de Lyon, CNRS, Rue Maurice Audin, 69518*  
6 *Vaux-en-Velin Cedex, France*

7 \* *Corresponding author: Patrick Bonnel, patrick.bonnel@entpe.fr*

8 **Abstract**

9 Demand forecasting is an essential task in many industries and the transportation sector  
10 is no exception. This is because accurate forecasts are a fundamental aspect of any rationale  
11 planning process and an essential component of intelligent transportation systems. In the  
12 context of public transit, forecasts are needed to support different level of planning and or-  
13 ganisational processes. Short-term forecast, typically a few hours in the future, are developed  
14 to support real-time operations. Long-term forecast, typically 5 years or more in the future,  
15 are essential for strategic planning. Those two forecast horizons have been widely studied  
16 by the academic community but surprisingly little research deal with forecast between those  
17 two ranges. The objective of this paper is therefore twofold. First, we proposed a generic  
18 modelling approach to forecast next 365 days ridership in a public transit network at differ-  
19 ent levels of spatiotemporal aggregation. Second, we illustrate how such models can assist  
20 public transit operators and transit agencies in monitoring ridership and supporting recurrent  
21 tactical planning tasks. The proposed formulation is based on a multiplicative decomposition  
22 that combines tree-based models with trend forecasting. The evaluation of models on unseen  
23 data proves that this approach generates coherent forecast. Different use cases are then de-  
24 picted. They demonstrate that the resulting forecast can support various recurrent tactical  
25 tasks such as setting future goals, monitoring ridership or supporting the definition of service  
26 provision. Overall, this study contributes to the growing literature on the use of automated  
27 data collection. It confirms that more sophisticated statistical methods can help to improve  
28 public transportation planning and enhance data-driven decision making.

29 **Keywords:** Public transit, Ridership forecasting, Machine Learning, Smart Card Data, Trans-  
30 port planning

## 31 **Highlights**

- 32 • Examines the need for medium-term forecasting in public transit systems
- 33 • Proposes an operational modelling framework to forecast ridership one year ahead at different  
34 levels of spatiotemporal aggregation
- 35 • Outlines different applications of the forecasts for tactical planning and ridership monitoring
- 36 • Exploitation of archived passive data can enhance data-driven decision making

## 37 **Author statement:**

38 **Oscar Egu:** Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original  
39 Draft, Visualization, Investigation. **Patrick Bonnel:** Supervision, Validation, Writing - Review  
40 and Editing.

## 41 **Acknowledgements**

42 This research was conducted as part of a research agreement between Keolis Lyon and the Urban  
43 Planning, Economics and Transport Laboratory (LAET). Their financial support is gratefully  
44 acknowledged. We would like to thank our colleagues from Keolis Lyon who provided the data  
45 and expertise that greatly assisted the research. The views expressed in this paper remain those  
46 of the authors.

1 Medium-term public transit route ridership forecasting:  
2 what, how and why ?  
3 A case study in Lyon  
4  
5

6 *Word Count: 7400*

7 **Abstract**

8 Demand forecasting is an essential task in many industries and the transportation sector  
9 is no exception. This is because accurate forecasts are a fundamental aspect of any rationale  
10 planning process and an essential component of intelligent transportation systems. In the  
11 context of public transit, forecasts are needed to support different level of planning and or-  
12 ganisational processes. Short-term forecast, typically a few hours in the future, are developed  
13 to support real-time operations. Long-term forecast, typically 5 years or more in the future,  
14 are essential for strategic planning. Those two forecast horizons have been widely studied  
15 by the academic community but surprisingly little research deal with forecast between those  
16 two ranges. The objective of this paper is therefore twofold. First, we proposed a generic  
17 modelling approach to forecast next 365 days ridership in a public transit network at differ-  
18 ent levels of spatiotemporal aggregation. Second, we illustrate how such models can assist  
19 public transit operators and transit agencies in monitoring ridership and supporting recurrent  
20 tactical planning tasks. The proposed formulation is based on a multiplicative decomposition  
21 that combines tree-based models with trend forecasting. The evaluation of models on unseen  
22 data proves that this approach generates coherent forecast. Different use cases are then de-  
23 picted. They demonstrate that the resulting forecast can support various recurrent tactical  
24 tasks such as setting future goals, monitoring ridership or supporting the definition of service  
25 provision. Overall, this study contributes to the growing literature on the use of automated  
26 data collection. It confirms that more sophisticated statistical methods can help to improve  
27 public transportation planning and enhance data-driven decision making.

28 **Keywords:** Public transit, Ridership forecasting, Machine Learning, Smart Card Data, Trans-  
29 port planning

# 1 Introduction

Demand forecasting is an essential task in many industries and the transportation sector is no exception. In fact, accurate prediction of future demand is an essential components of intelligent transportation systems [Vlahogianni et al., 2014, Koutsopoulos et al., 2019] and a fundamental aspect of any rationale planning process [Bonnell, 2002, Ortúzar and Willumsen, 2011]. Thus, it is not surprising that passenger demand forecasting is a widely studied subject. To summarize this area of research, we must take into account the domain of application and the type of planning issues the forecast intend to address. In many organisations including public transport, they are three commonly accepted level of planning and organisational control [van de Velde, 1999, Pelletier et al., 2011, Ortúzar and Willumsen, 2011]. Strategic level deal with long-term decisions and objectives. Tactical level focus on decisions that take place in the medium-term and aims to guarantee that the means to reach long-term goals are in place. Operational planning is concerned with short-term decisions that ensure the efficiency of the production. In agreement with this hierarchical order of decision-making activities, operators and transport agencies must generate different forecasts.

Short-term forecasting is a very active field [Vlahogianni et al., 2014] that deal with models that predict demand from few minutes to few hours into the future [Vlahogianni et al., 2014, Noursalehi et al., 2018]. **The wide adoptions of smart card fare collection systems has provided new opportunities for short-term forecasting.** Zhai et al. [2018] have discussed the literature about short-term bus passenger flow forecasting models and methods. Ma et al. [2014] proposes an Interactive Multiple Model-based Pattern Hybrid approach and evaluate it using empirical smart card data collected along one bus route over one year in Jinan, China. Ding et al. [2016] forecast short term subway ridership using gradient boosting decision trees. Using Beijing smart card data, Li et al. [2017] proposes a novel multiscale radial basis function to forecast the irregular fluctuation of subway passenger flows under special events. Using smart card data from London Underground, Noursalehi et al. [2018] proposed to predict short term passenger arrivals at transit stations using state-space models and considering station interactions. In the context of public transit (PT), authors argue that short-term forecasting can enable the design of better control strategies and improve passenger experience by proactively adjust services and customer information [Koutsopoulos et al., 2019, Noursalehi et al., 2018, Ma et al., 2014, Wei and Chen, 2012]. While appealing and surely helpful, short-term forecasting can be hard to implement in real-life settings because it requires real-time data, continuous actualization of the model and trust in the model output. Short-term forecasts are also developed to support near real-time operational decision making. However, PT systems cannot always react efficiently in real-time due to limited flexibility in resources (e.g vehicles, staff) and infrastructure constraints. Transport services and infrastructures are thus often designed and planned years in advance.

Long-term ridership forecasting deals with models that predict demand for time horizon ranging from 5 to 15 years ahead. In contrast to short-term forecasting, it is often a one-time exercise and forecasts are rarely generated continuously. Forecasts are produced to assess and evaluate future scenarios and support long-range strategical planning. The most common method for long-term ridership forecasting traditionally relies on four-step travel demand model [Boyle, 2006, Ortúzar and Willumsen, 2011, Bonnell, 2002]. Those models are data-intensive as they incorporate many elements such as land use, future population, choice model, network description etc. On one hand, it is a strength as this type of model can be used to forecast structural changes such as the introduction of a new route but also to evaluate different policy choice [Ortúzar and Willumsen, 2011, Bonnell, 2002]. On the other hand, it is a weakness because the resulting model can be costly, complex and hard to modify. Consequently, other forecasting approaches have been developed. For example, change in ridership in response to change in services or fares can be forecast using elasticity

78 coefficient [Boyle, 2006, Totten and Levinson, 2016, Paulley et al., 2006]. Other approaches include,  
 79 spatial regression method [Pulugurtha and Agurla, 2012], rules of thumb method based on similar  
 80 routes or professional judgement [Boyle, 2006, Diab et al., 2019]. In practice, those models and  
 81 approaches are often calibrated for a limited set of situation such as an average weekday, a typical  
 82 peak hour etc. They are, therefore, difficult to use in recurrent monitoring of the demand and  
 83 cannot support tactical and operational planning activities.

84 Surprisingly, little research deal with forecast between those two horizons i.e medium-term  
 85 forecasting. However, those kinds of forecasts are part of the toolbox needed for the effective  
 86 planning of public transit systems. They can be of great use for PT operators and agencies  
 87 that want not only to be able to adapt real-time operation (short-term forecasting) and evaluate  
 88 strategic plan (long-term forecasting) but also to monitor tactically and continuously the evolution  
 89 of demand. Toqué et al. [2017] have recently proposed to use machine learning models to predict  
 90 PT ridership one year ahead in a disaggregate and continuous manner. However, the authors  
 91 didn't introduce in their analysis a multiyear trend component neither a description of the level of  
 92 supply although they may surely be needed for medium-term forecasting. Moreover, they barely  
 93 discuss the practical aspect of the forecasting framework and motivate their work with use cases.  
 94 This paper seeks to address these research gaps and has two complementary objectives:

- 95 1. To develop a generic forecasting approach that combined **passenger flow trend analysis** with  
 96 machine learning model to predict one year in advance, at different levels of spatiotemporal  
 97 aggregation, ridership volume in a PT network. **The model should consider seasonal effects,**  
 98 **holiday period, bank holiday, day of the week, hours of the day and unusual effect to forecast**  
 99 **not only an average ridership one year in advance, but the variability of ridership across days,**  
 100 **hours and network elements.**
- 101 2. To illustrate how such forecasts can support tactical planning tasks and assist PT operators  
 102 and agencies in monitoring ridership. **The illustrations are based on real use cases from Lyon**  
 103 **public transit network.**

104 To this end, the paper has been organized in the following way. It begins with a presentation of the  
 105 modelling approach. After that, relevant data are identified, prepared and described. Then, the  
 106 forecasts are validated using unseen data and a bottom-up approach. Finally, use cases are provided  
 107 to illustrate the value of such data-driven framework. Results show that the proposed modelling  
 108 approach is valid. Transit agencies can leverage the resulting forecast to better monitor and  
 109 anticipate ridership in their network. Overall, this study contributes to the growing literature on  
 110 the use of automated data collection to improve public transportation planning. Findings confirm  
 111 that more sophisticated statistical methods can facilitate analysis that used to rely primarily on  
 112 professional judgement [Hanft et al., 2016, Pelletier et al., 2011, Koutsopoulos et al., 2019].

## 113 2 Modelling framework

114 Let's denote  $Y_i = (y_{i,1}, \dots, y_{i,t}, \dots, y_{i,m})$  as the vector of observed ridership volume on element  $i$  of  
 115 a PT network for each time step  $t \in 1, \dots, n$ . Let's denote  $X_{i,t}$  the set of features (explanatory  
 116 variables) whose values are known for  $t \in 1, \dots, m$  where  $m$  ( $m > n$ ) is the prediction horizon.  
 117 Our goal is to estimate a regression model  $f_i(\cdot)$  for each element  $i$  such that  $y_{i,t} = f_i(X_{i,t}) + \varepsilon$   
 118 where  $\varepsilon$  is the difference between the observed value and the predicted value  $\widehat{y}_{i,t} = f_i(X_{i,t})$ .  
 119 This formulation described a typical univariate regression problem where (1) each element of the  
 120 network  $i$  is considered independent of the rest of the network, (2) each observation  $y_{i,t}$  is considered  
 121 independent of the other's observations.

122 As proposed by Toqué et al. [2017], to estimate  $f_i(\cdot)$  we can use machine learning models such  
 123 as ensemble of decision trees. Those models have recently gained much popularities in various  
 124 fields due to their ability to learn complex non-linear relation between features and provide higher  
 125 accuracy than single machine learning models [Friedman et al., 2001, Opitz and Maclin, 1999].  
 126 Those models combine several decision trees into one single model that output the mean of all  
 127 decision tree. A single decision tree can be seen as a stratification of the ridership volume into  
 128 some simple regions using a set of splitting rules based on the available features. The underlying  
 129 principle of those models is that by combining single decision trees we can significantly improve the  
 130 predictive performance and reduce the variance of a single tree [Friedman et al., 2001, Breiman,  
 131 2001, Friedman, 2001].

132 To learn a set of decision trees effectively the two most common algorithms are random forest  
 133 (RF) and gradient boosting (GB). Random forest was introduced by Breiman [2001] and is based  
 134 on the concept of bagging. More precisely, each decision tree is built independently on a random  
 135 subset of the training data but also considering only a random sample of features at each split. In  
 136 doing so, the algorithm enforces diversity in the trees which when we average them may reduce the  
 137 variance of the model [Breiman, 2001]. Gradient boosting trees were formalized by Friedman [2001]  
 138 and differ from the random forest model in the sense that decision trees are built sequentially using  
 139 strategically resample training data. More specifically, each new tree is built to recover the errors  
 140 resulting from the model obtained with the ensemble of trees produced at the previous steps. In  
 141 this approach, each of the decision trees can remain rather small with performance slightly above  
 142 random guess. Fitting them sequentially can then help to improve the model in areas of the feature  
 143 space that were missed by previous ensemble [Friedman et al., 2001].

144 One aspect of those machine learning models is that they assume independence between the  $y_{i,t}$   
 145 and thus can't learn time-dependent structure and forecast potential trend. In fact, as explained  
 146 above the model output are computed by averaging decision trees which are themselves based on  
 147 historical ridership data. However, ridership volume may have statistical properties that evolve  
 148 through time and might also be view as a non-stationary time series. To deal with this aspect,  
 149 we can assume that ridership volume has two components a yearly trend-cycle component  $s_{i,T}$   
 150 and a yearly adjusted component  $a_{i,t}$ . Ridership level can then be expressed using multiplicative  
 151 decomposition:

$$y_{i,t} = s_{i,T} \times a_{i,t} \quad (1)$$

152 where  $s_{i,T}$  denotes the mean of  $y_{i,t}$  in year  $T$ ,  $t \in T$  and  $a_{i,t}$  is the yearly adjusted level ridership of  
 153 element  $i$  at time step  $t$ .

154 The forecasting task can then be divided into two subtasks where we have to forecast  $s_{i,T}$  and  
 155  $a_{i,t}$ . The rationale behind this process is twofold. First,  $a_{i,t}$  is assumed to be stationary and free  
 156 of trend and thus can be estimated with ensemble of decision trees ( $\widehat{a}_{i,t} = f_i(X_{i,t})$ ). Second,  $s_{i,T}$   
 157 is supposed to be non-stationary and evolve slowly over time and can be forecast by taking the  
 158 last value of the yearly trend-cycle component multiply by a growth factor  $\alpha_i$ .  $\widehat{y}_{i,t}$  can then be  
 159 obtained in the following way:

$$\widehat{y}_{i,t} = \widehat{a}_{i,t} \times (1 + \alpha_i) s_{i,T-1} \quad (2)$$

160 In other words, the forecasted ridership for element  $i$  is equal to the mean ridership of the previous  
 161 year  $s_{i,T-1}$  multiply by a growth factor while the deviation from the mean yearly ridership volume  
 162  $a_{i,t}$  is predicted using a regression model.  $\alpha_i$  can then be estimated using a weighting average of  
 163 the observed historical year-mean percentage evolution:

$$\alpha_i = \frac{\sum_{T=2}^N w_T \frac{s_{i,T}}{s_{i,T-1}}}{\sum_{T=2}^N w_T} \quad (3)$$

164 where  $N$  is the number of years used to train the model and  $w_T = 1/(N + 1 - T)$  is the weight  
 165 associated with year  $T$  defined with inverse time decay function. This weighted formulation is  
 166 proposed to estimate  $\alpha_i$  from multiyear behaviour of each element while giving higher weight to  
 167 more recent years.

168 This forecasting formulation is well suited to our objectives for two reasons. First, it exploits  
 169 the property of ensemble models to learn the influence of various features on detrended ridership  
 170 data. Second, it aligned forecast with previous year observed ridership volume of element  $i$  to  
 171 ensure coherency of the forecast with the most recent element state. This formulation can be seen  
 172 as a conservative continuation of previous year’s ridership mean volume combined with a model  
 173 that learns historical behaviour of the network element under different conditions of the features  
 174 space. As noted by Hyndman and Athanasopoulos [2018], traditional time series models (such  
 175 as seasonal ARIMA) don’t allow irregular events and have difficulty to accommodate multiple  
 176 imbricated seasonality. The intuition behind the above formulation is that the decision tree model  
 177 is able to capture those effects while remaining robust to potential outliers.

178 At this step, we have now a modelling approach that can be used to forecast the future ridership  
 179 volume at time step  $t$  of each element  $i$  denoted as base forecast. To establish forecast for higher  
 180 levels of the network hierarchy such as a group of elements  $J$  (e.g group of stops, group of routes)  
 181 or larger temporal aggregation (e.g day, week, month, working day), we can employ the simple  
 182 and straightforward bottom-up approach. In this case, the forecasted ridership volume of higher  
 183 hierarchies element  $J$  for the period  $M = [t1, \dots, t2]$  denoted  $\widehat{R}_{J,M}$  is obtained by summing the  
 184 different base forecast:

$$\widehat{R}_{J,M} = \sum_{i \in J} \sum_{t=t1}^{t2} \widehat{y}_{i,t} \quad (4)$$

185 This approach is a classical way to deal with time-series collections that naturally presents hier-  
 186 archical properties and different seasonal pattern [Hyndman and Athanasopoulos, 2018, Kahn,  
 187 1998]. It has the advantage to preserve in the forecasting task the dynamic and characteristics  
 188 of individual elements while making it possible for the analyst to obtain forecast at the desired  
 189 spatiotemporal level of the PT network.

### 190 3 Data preparation and exploration

191 Since the introduction of automated data collection systems, it is now possible to access high-  
 192 resolution ridership data. Those new data sources have the potential to change the current state  
 193 of practice in ridership forecasting if they are accurate and of good quality [Boyle, 2006, Diab  
 194 et al., 2019]. To be used in medium-term forecasting those data need also to be correctly curated  
 195 and archived so there are enough observations to train models. The proposed modelling approach  
 196 requires the definition of a base forecast level. This level should not be too noisy to ensure models  
 197 can properly learn patterns and at the same time sufficiently detailed for most of the required  
 198 tactical analysis. This level should also be set from the perspective of the organisation managing  
 199 the network which can be different from the passenger perspective. We suggest that the most  
 200 appropriate spatiotemporal base level of aggregation for medium-term forecasting are:

- 201 • Temporal aggregation ( $t$ ): hourly data because it is needed to account for intra-day variation  
 202 and define supply level accordingly. However, it should be noted that the same methodology  
 203 could be applied to daily data if this resolution level is deemed sufficient.
- 204 • Spatial aggregation ( $i$ ): route level is most suitable than stop level because supply level is  
 205 more than anything else link to route elements.



206 In Lyon, the current fare transaction system is an entry only system that was introduced in  
207 2002. Passenger can validate using either smart card or magnetic paper. Validation is mandatory  
208 when boarding a bus or tramway and when entering the subway network. All transactions are  
209 then anonymized and stored with timestamp and localisation attributes. For this research, the  
210 fare transactions from 2014 to 2018 for the subway network (4 metro lines and 2 funicular lines),  
211 the tramway network (5 lines) and high-frequency buses known as line C (25 lines) were extracted  
212 from the Lyon PT operator data warehouse. A map of the extracted network is given in figure 1.  
213 Those routes were selected because they represent more than 80% of the total ridership of the  
214 network. This sample was also deemed to be enough to evaluate correctly our approach. Fare  
215 transactions were then summed by route and by hour<sup>1</sup>. To ensure, the completeness of the data,  
216 the following procedure was implemented recursively for each route:

- 217 • Generate the hourly time bin from the 1st of January 2014 00:00 to the 31st December 2018  
218 23:00.
- 219 • Lookup for the number of fare transactions summed by hour. If no transaction is found for  
220 a given hour it was assumed that no passenger board the route and the ridership value was  
221 set to zero. This procedure is put in place to ensure the consistency of the training data  
222 which must contain periods with and without ridership to correctly learn the route historical  
223 behaviour<sup>2</sup>.

224 For the rest of this paper, the dataset resulting from the above procedure will be referred to as the  
225 route ridership volume and will form the base level of our modelling framework.

226 Our modelling framework also required a set of features that may explain the route ridership  
227 volume variations while being available one year in advance. It is well recognized that PT ridership  
228 presents multiple temporal seasonalities and regular cycles that can be encoded with ordinary  
229 calendar attributes. Consequently, a calendar features table common for all routes was manually  
230 crafted. Features of this spreadsheet are described in table 1. Ridership is also influenced by the  
231 level of supply i.e the quantity of service produced. One of the most common and simple indicator  
232 to describe this aspect of transit is the number of commercial kilometres offered. This indicator  
233 was thus extracted from the operator data warehouse and choose as an explanatory variable.  
234 Unfortunately, for this research, this feature was available with enough historical depth only at the  
235 level of day and route. Noted also that in Lyon, commercial kilometres offered by routes are set  
236 one year in advance by the operator planning team for contractual and financial reasons. Thus,  
237 like calendar features, they can be determined sufficiently in advance and be used as predictors  
238 in medium-term forecasting (which is not the case of potential other features such as weather or  
239 unplanned events).

240 The calendar table, the supply feature and routes ridership volumes were then merged. To  
241 explore the characteristics of the resulting dataset we have aggregated the ridership volume of  
242 the 36 selected routes and perform exploratory visualizations in figure 2. In figure 2a, it can be  
243 seen that there are cyclical patterns that repeat every year such as a strong decrease of ridership  
244 volume during summer months (July and August). There are also long-term dynamics in the  
245 ridership volume with an observable increasing tendency. This is because Lyon network has been  
246 gaining popularities in the last few years. In figure 2c, the variation of daily ridership volume for  
247 the year 2018 are observed more finely. This figure indicated that they are multiple imbricated  
248 seasonalities and calendar effects. More precisely, it can be seen that school holiday periods (e.g  
249 in February, April, October and December) but also other events such as bank holidays or special

---

<sup>1</sup>Noted that for the subway fare transactions are linked to stations and were thus redistributed by route according to operator reporting procedure.

<sup>2</sup>Note that in the case of potential data loss this procedure will also result in ridership equal to zero.

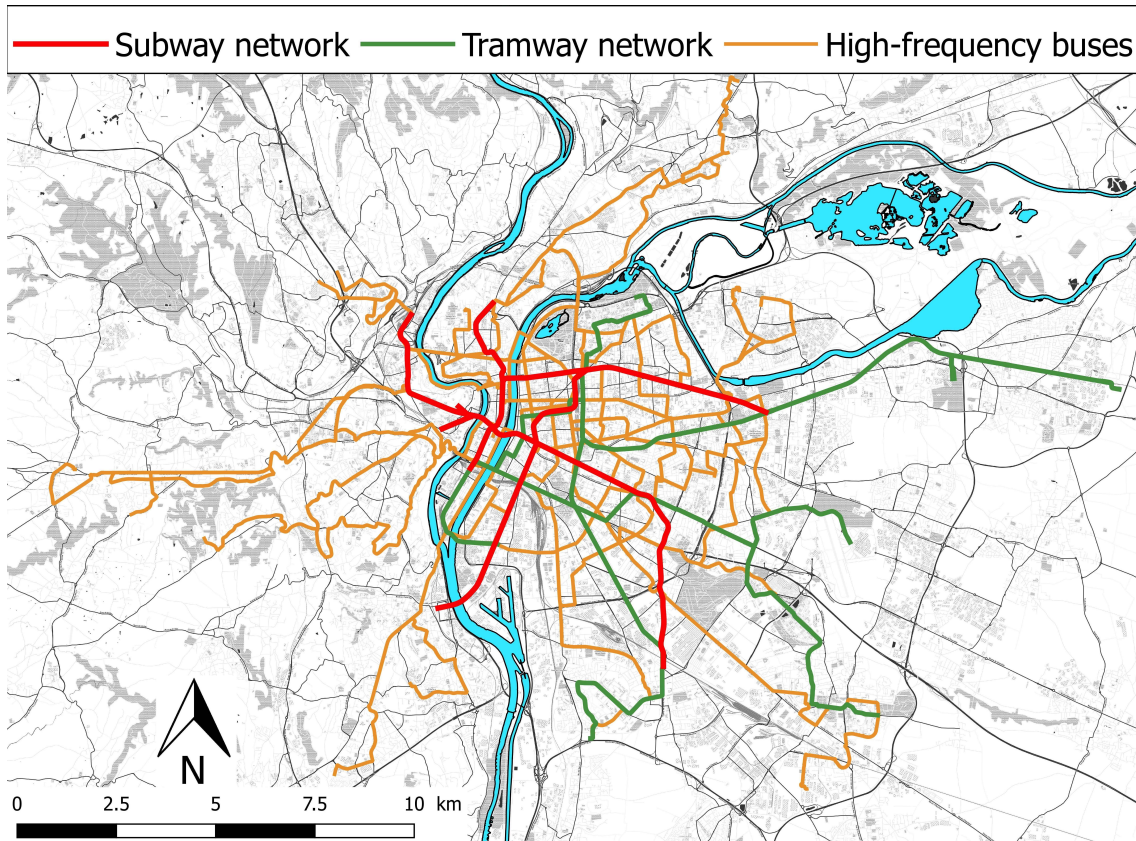


Figure 1: Map of the study network, source: prepared by the authors

Features	Description
Hour	24 categories
Month	12 categories
Week number	Numeric from 1 to 53
Day of the week	7 categories
School holiday	7 categories, one for each holiday period
Public holiday	14 categories, one for each French public holiday
Adjacent to public holiday	Binaries if weekdays and adjacent to public holiday
Light and music festivals	Binaries if major citywide events

Table 1: Hand crafted calendar features, source: prepared by the authors

250 events have an impact on the ridership volume. Similarly, we observe a strong decrease in ridership  
 251 during weekends and summer months. As indicated by figure 2a, those seasonalities seems to be  
 252 recurrent every year but calendar effects are moving in time because each year events and holidays  
 253 can fall in different days of the year. In figure 2b, the relation between the total daily number  
 254 of commercial kilometres produced and the ridership volume is depicted. This figure shows that  
 255 there are three main states in the network that correspond to three typical planned service output

256 (weekday, Saturday and Sunday). Figure 2b also demonstrates that there are more variations  
 257 in the ridership volume than in the service produced. For instance, 85,000 kilometres produced  
 258 corresponds to a typical weekday supply level but for this same amount of supply, the ridership  
 259 volume can vary between 0.8 to 1.1 million. This is because the supply is often fixed based on a  
 260 set of typical days while the demand is more volatile. Finally, in figure 2d the hourly ridership  
 261 from the 5th May 2018 to 19th May 2018 is given. This period exhibits distinct hourly patterns.  
 262 During the first week, there are two public holidays on the 8th of May and 10th of May and the  
 263 demand pattern on those day does not exhibit strong peak compare to the other weekday of the  
 264 week. During Saturday and Sunday, the hourly ridership curve has a bell shape. Finally, from the  
 265 14th of May to the 18th of May we observe a typical demand pattern with peak periods in the  
 266 morning and the evening but also higher hourly ridership volume than during the previous week.

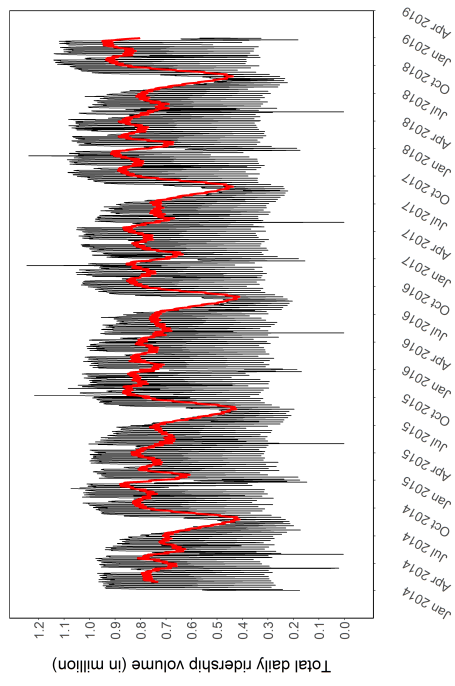
267 Overall, the previous plots confirm that ridership volume presents multiple moving imbricated  
 268 seasonality and irregular events. Changes appear from one year to another, from one month to  
 269 another, during the hour of the day, the days of the week etc. There is also a relation between  
 270 the level of supply and demand. Those observations are not per se surprising but we should  
 271 recall that those are the elements that each base model needs to capture independently for each  
 272 route. In fact, while the previous plots were obtained based on the aggregation of all route,  
 273 the model will be trained for each route separately and we can expect that routes have distinct  
 274 behaviour. For example, the influence of holiday may be less important for some part of the  
 275 network, some routes may show higher multi-year growth than other, some routes may have specific  
 276 characteristics on weekends because they serve specific areas of the city such as touristic places.  
 277 Hence, it is fundamental that the model can adapt correctly and in an automatic way to varying  
 278 route behaviours while capturing system-wide demand variation and structure.

## 279 4 Models fitting and forecast errors

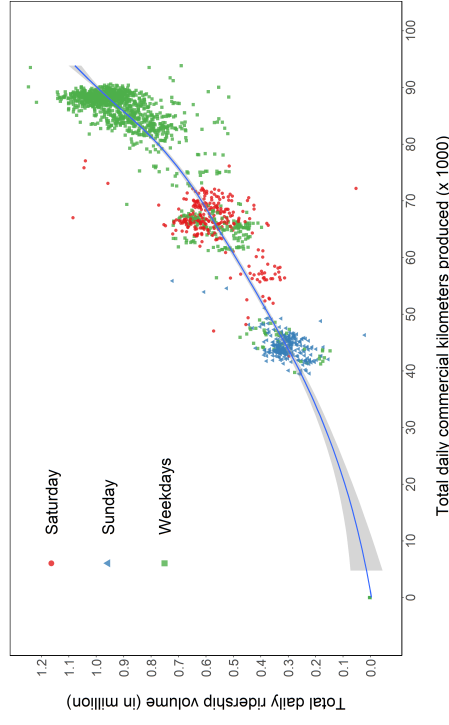
Name	Model	Features	Decomposition
HM	Historical median	Hour, Type of day, Month, Working day	
HMT	Historical median	Hour, Type of day, Month, Working day	x
RF	Random forest	Calendar features	
RFT	Random forest	Calendar features	x
RFTS	Random forest	Calendar features and supply feature	x
GB	Gradient boosting	Calendar features	
GBT	Gradient boosting	Calendar features	x
GBTS	Gradient boosting	Calendar features and supply feature	x

Table 2: List of implemented models, source: prepared by the authors

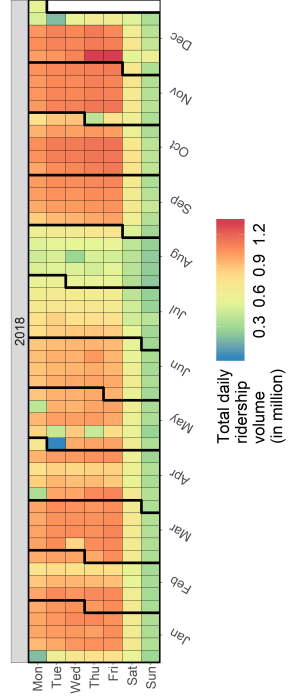
280 In the previous sections, we have defined the problem, proposed a modelling framework, gather  
 281 the data and perform some preliminary exploratory analysis, we now present models fit. Table  
 282 2, depicts the list of models that we have decided to evaluate. The first model formulation is a  
 283 naive historical approach. In this case, the forecast is computed using the median of all past time  
 284 step over the historical observation with the same characteristic as the forecasted time step. Time  
 285 step characteristics were in this case defined with the cartesian product between the hour, the  
 286 month, the type of day (weekday, Saturday, Sunday) and a boolean indicator for working day (no  
 287 public holiday, no school holiday, no special events). This formulation will serve as the baseline to  
 288 evaluate the two ensemble models described in section 2. Table 2 also indicates if the models were



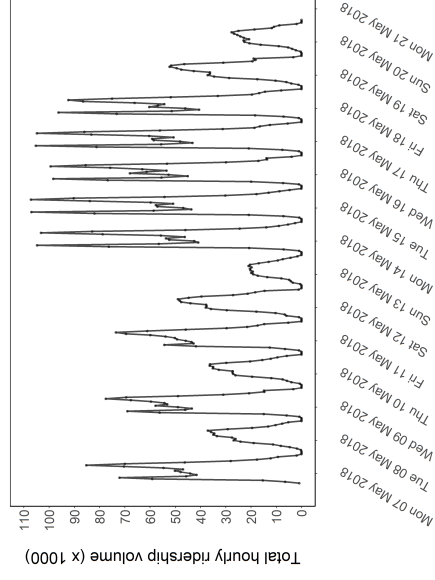
(a) Dayli ridership volume from 2014 to 2018, red curve correspond to 30 days moving average



(b) Relation between the supply feature and the daily ridership volume, points are colored by day type



(c) Calendar heatmap of daily ridership volume in 2018



(d) Hourly ridership volume from the 7th of May 2018 to the 20th of May 2018

Figure 2: Exploration plot based on the aggregated ridership volume of the 36 selected routes, source: prepared by the authors

289 trained with the raw ridership volume or with the decomposed ridership volume. In the second  
 290 case, the equation 2 was used to forecast future ridership volume.

291 The accuracy of a forecast procedure can only be truly determined using data that were not  
 292 used to fit the model [Hyndman and Athanasopoulos, 2018]. To do so, there are two common  
 293 approaches: the cross-validation where several sets of tests are sequentially constructed and the  
 294 hold-out approach where the data is divided into two portions. In this research, the hold-out ap-  
 295 proach was selected. The data was thus divided between a training set comprising all observations  
 296 from 2014 to 2017 and a test set consisting of 2018 observations. To measure forecast errors the  
 297 following standard metrics were chosen:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_{i,t} - \widehat{y}_{i,t}| \quad (5)$$

298

$$\text{MAPE}_v = \frac{1}{n} \sum_{i=1}^n \frac{|y_{i,t} - \widehat{y}_{i,t}|}{y_{i,t}}, \quad \forall y_{i,t} > v \quad (6)$$

299

$$\text{MdAPE} = \text{median}\left(\frac{|y_{i,t} - \widehat{y}_{i,t}|}{y_{i,t}}\right) \quad (7)$$

300 In table 3, we have reported the error measures on the train and test dataset for different levels of  
 301 aggregations. First, as expected, when the aggregation level increases, errors decrease significantly.  
 302 For instance, the best model at the base forecast has a MdAPE of 11.7 % almost four times bigger  
 303 than the best model MdAPE once forecasts are aggregated by day (2.8 %). This indicates that  
 304 the base levels are challenging to forecast but the bottom-up approach allows to generate coherent  
 305 results. In other words, base forecast errors tend to compensate when forecasts are aggregated by  
 306 route or by time. A second important observation is that ensemble models outperform baseline  
 307 model in terms of absolute errors (MAE) but also in terms of percentage errors (MAPE<sub>v</sub> and  
 308 MdAPE). Ensemble models consider a more complex feature space but still can generalize to  
 309 previously unseen data more comprehensively and without overfitting. However, there are no  
 310 important differences in terms of accuracy between the gradient boosting and the random forest  
 311 which is not surprising since both methods have strong similarities. What is also clear from this  
 312 table is that the decomposition approach is essential when historical data with a certain depth is  
 313 used to model future ridership. This is because the decomposition ensures the coherency of the  
 314 forecast with the most recent observed ridership volume. Finally, the introduction of supply feature  
 315 in ensemble models can reduce the errors especially large errors. If we compare the RFT (random  
 316 forest with decomposition) and RFTS (random forest with decomposition and supply feature), it  
 317 can be seen that median absolute percentage errors are quite similar but mean absolute percentage  
 318 error can be considerably higher when the supply feature is omitted. This is because, when there  
 319 is a significant reduction in the kilometres produced the model based on calendar features will  
 320 assume no change in the ridership volume.

## 321 5 Use case analysis

322 In the previous section, we have fitted different models and generate forecasts of ridership volume  
 323 one year in advance. We have shown that it was possible to obtain rather low errors at higher  
 324 aggregated levels but the forecast quality decreases at lower levels of aggregation. While forecast  
 325 accuracy is in itself an important matter it is also important to recall that PT networks are complex  
 326 systems that are impacted by all sort of unexpected and exogenous events. Therefore, the above  
 327 models should not be seen solely for their prediction capability but also for their ability to assist PT

		TRAIN SET (2014-2017)							
		HM	HMT	RF	RFT	RFTS	GB	GBT	GBTS
Base level (route-hour)	MAE	97	93	79	74	69	81	77	<b>67</b>
	MAPE.50 (%)	17,4	16,8	15,1	14,5	<b>12,9</b>	16,6	16,1	13,4
	MdAPE (%)	10,5	10,0	10,3	9,6	<b>9,0</b>	10,7	10,1	9,1
Aggregated (route-day)	MAE	1795	1670	1236	1063	<b>712</b>	1230	1088	859
	MAPE (%)	25,6	24,8	11,2	10,3	<b>5,5</b>	13,1	12,1	8,2
	MdAPE (%)	6,3	5,6	5,2	4,3	<b>3,0</b>	5,4	4,6	3,8
Aggregated (hour)	MAE	2682	2618	1913	1816	1741	2040	1987	<b>1564</b>
	MAPE (%)	16,0	15,7	17,1	16,9	18,0	20,2	20,1	<b>14,8</b>
	MdAPE (%)	6,4	6,0	5,5	5,0	5,0	6,4	6,2	<b>4,9</b>
Aggregated (day)	MAE	51725	49869	30793	26737	<b>18264</b>	31063	28508	21007
	MAPE (%)	67,0	67,1	14,7	14,1	<b>6,3</b>	18,5	18,4	11,5
	MdAPE (%)	4,4	4,0	3,1	2,4	<b>1,8</b>	3,3	2,8	2,2
		TEST SET (2018)							
		HM	HMT	RF	RFT	RFTS	GB	GBT	GBTS
Base level (route-hour)	MAE	131	115	118	99	95	123	104	<b>93</b>
	MAPE.50 (%)	21,7	21,7	19,9	19,8	<b>17,2</b>	21,1	21,4	17,4
	MdAPE (%)	14,5	12,8	13,7	11,9	<b>11,7</b>	14,4	12,6	11,7
Aggregated (route-day)	MAE	2662	2157	2339	1708	<b>1480</b>	2351	1720	1497
	MAPE (%)	36,3	36,9	26,8	26,7	<b>16,6</b>	28,1	27,8	17,7
	MdAPE (%)	10,7	7,7	9,9	6,8	<b>6,4</b>	10,0	6,9	6,5
Aggregated (hour)	MAE	3562	2878	2993	2140	2182	3226	2349	<b>1976</b>
	MAPE (%)	25,7	24,9	19,7	19,1	16,7	25,4	25,1	<b>16,4</b>
	MdAPE (%)	10,0	7,1	8,4	<b>5,7</b>	5,9	9,7	6,8	5,8
Aggregated (day)	MAE	75721	54068	63457	33658	31259	63643	35738	<b>30671</b>
	MAPE (%)	55,2	55,7	28,9	26,7	<b>9,0</b>	32,9	31,0	13,8
	MdAPE (%)	8,1	4,4	7,2	<b>2,8</b>	3,1	7,8	3,5	3,1

Table 3: Model results on the train and test dataset for different level of aggregation (HM: historical median, RF: random forest, GB: gradient boosting, T indicates the use of the proposed decomposition approach, S indicates the incorporation of supply feature), source: prepared by the authors

328 organisations in monitoring ridership and in improving tactical decision-making activities. This is  
329 what we intend to demonstrate in this section by introducing a few use cases.

330 With the above modelling approach, it is possible to obtain coherent aggregated ridership  
331 forecasts. These types of forecasts could be used for tactical decision making. For instance, they  
332 can assist senior executives in setting next year goals and monitoring their PT network with a more  
333 data-driven approach. Similarly, every year, transit agencies need to prepared next year budget  
334 which obviously requires at one point some sort of ridership forecast. In practice, this process can  
335 be tedious and is often a matter of expert. In the case where different operators coexist in the same  
336 network, this task is even more complex and critical for the proper functioning and financing of  
337 the system. Agencies could use the proposed approach to generate forecasts for all route and then  
338 aggregated them to prepare budgets that match more closely the future variation of ridership and  
339 fit the financial/operational division of the network<sup>3</sup>. To illustrate this first use case, the weekly  
340 forecast for the two random forest models obtained through decomposition (RFT and RFTS) by  
341 type of transportation mode is plotted in figure 3.

342 In figure 3, it can be seen that the different series are overall quite close and the aggregation of

<sup>3</sup> Those same forecasts could also be used to schedule workforces such as field information officers or fare officers

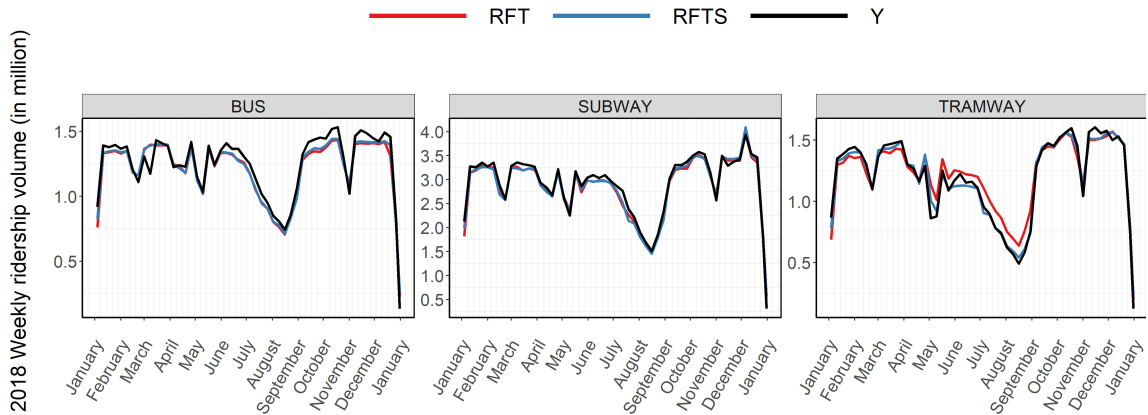


Figure 3: Weekly forecast one year in advance for the three transportation mode (RFT: decomposed random forest, RFTS: decompose random forest with supply features, Y: real ridership volume), source: prepared by the authors

343 models captures typical calendar patterns such as the impact of school holidays, public holidays,  
 344 or citywide events (i.e peak at week 50 for subway network). It can also be seen that the ridership  
 345 volume of high-frequency bus network has increased more quickly than anticipated by the two  
 346 models, especially after September. Finally, figure 3 indicates that for the tramway routes both  
 347 model yield different forecast. This is because in 2018 the tramway network was impacted by  
 348 important construction work of a new tramway route (T6) that leads to change in the service  
 349 provision. Those type of elements can be further inspected by an analyst using models outputs in  
 350 a retrospective manner. Especially, models can be used to identify routes whose behaviour deviate  
 351 from historical behaviour as captured by the models. In doing so, analysts could identify and  
 352 prioritize which route they need to review first in a more efficient manner than doing it based on  
 353 "anecdotal knowledge" or fixed cycle [Coleman et al., 2018]. In figure 4, we have plotted each route  
 354 as a point in a two-dimension space where the x-axis is the annual percentage error between the  
 355 RFTS model and the real 2018 ridership and the y-axis is the total real ridership volume. Using  
 356 figure 4 it is easy to identify routes that require further analysis, balancing the importance of the  
 357 route (y-axis) and the importance of the deviation from the model (x-axis). Figure 4 indicates  
 358 that most routes errors are under 5% but some route exhibit higher discrepancies between model  
 359 and real observations such as C3, C14, C15, C9 or T5. This route level analysis can be further  
 360 extended to go deeper into the available data. For instance, real and model route ridership volume  
 361 were aggregated by month and inspected in figure 5.

362 In figure 5, we can observe that there is indeed a large diversity of route ridership historical  
 363 pattern. For instance, route C3 is the busiest bus route in Lyon and important road work were  
 364 carry on between mid-2016 to mid-2019 to provide exclusive right-of-way. The route was divided  
 365 into two sections and part of the stops were relocated. As indicated by figure 5, it had a great  
 366 impact on the 2017 ridership volume and the model forecast further decreased of the demand  
 367 in 2018. However, by the end of 2018, as the route progressively returns to better operational  
 368 condition, the ridership starts to increase significantly in an unforeseen manner for the model.  
 369 In 2016, a strong decrease of ridership is observed on route C14<sup>4</sup> follow by a return at historical  
 370 ridership volume during 2017. These disruptions were not well assimilated by the models, and

<sup>4</sup>Route C14 share a portion of C3 layout and was also impacted by road work

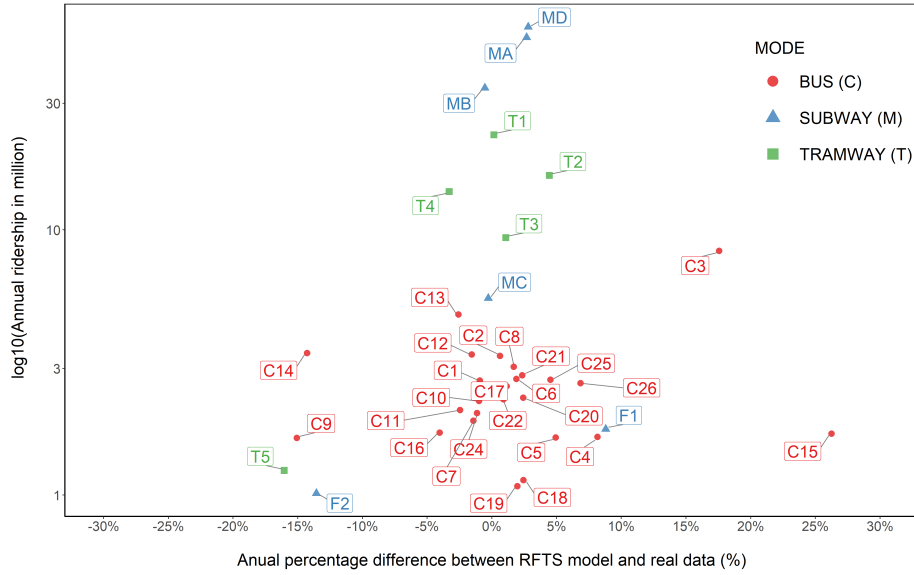


Figure 4: Annual ridership volume forecast error for decomposed random forest with supply feature model (RFTS), source: prepared by the authors

371 further growth of the route was forecast. As mentioned earlier, in 2018 the tramway network  
 372 was impacted by the construction of a new tramway line (T6). In figure 4, it can be seen that  
 373 while tramway routes T1 and T3 are forecast correctly the model underestimates/overestimates the  
 374 traffic on routes T2/T4. Figure 5 indicates that those routes exhibit a strong decrease in summer  
 375 month (July/August). Models with supply feature partially anticipate this decrease (see figure 5  
 376 and figure 3) but models with only calendar features couldn't predict those changes (see figure 3).  
 377 The impacts of those kinds of exogenous elements are in practice very difficult to understand and  
 378 measure clearly without confronting the real ridership with a "business as usual" scenario. With  
 379 the above modelling approach, transit agencies can now more easily do so. They can monitor  
 380 route ridership with a data-driven approach using a predictive model that is easy to re-actualize  
 381 but also to implement. Moreover, they can use these "business as usual" forecast to estimate in  
 382 advance the number of trips impacted by a planned disruption. Once the period of disruption  
 383 and the routes impacted are known, an analyst can interrogate models to retrieve the forecasted  
 384 number of impacted trips. This information can then be used in the design of replacement services  
 385 or mitigation plan. In a similar vein, if different periods of disruption are being considered, models  
 386 could also be used to identify the period that minimizes the impacted volume of trips. For this  
 387 kind of use case, one may prefer to leverage on model using calendar features so the forecast is a  
 388 pure "business as usual" prediction.

389 The above use cases were based on spatial and/or temporal aggregation of model outputs but the  
 390 base model i.e by route and hour can also be of value to apprehend in-depth ridership pattern. To  
 391 this aim, forecasts need to be synthesized in such a way that the information can directly be helpful  
 392 for tactical planning purposes. One way to do so would be to cluster the model output to identify  
 393 a set of typical daily ridership profiles. Spherical kmean is a popular text mining prototype-based  
 394 clustering algorithm [Buchta et al., 2012]. It relies on the cosine distance to measure dissimilarity  
 395 between objects which has interesting properties for our problem as the observations are cluster



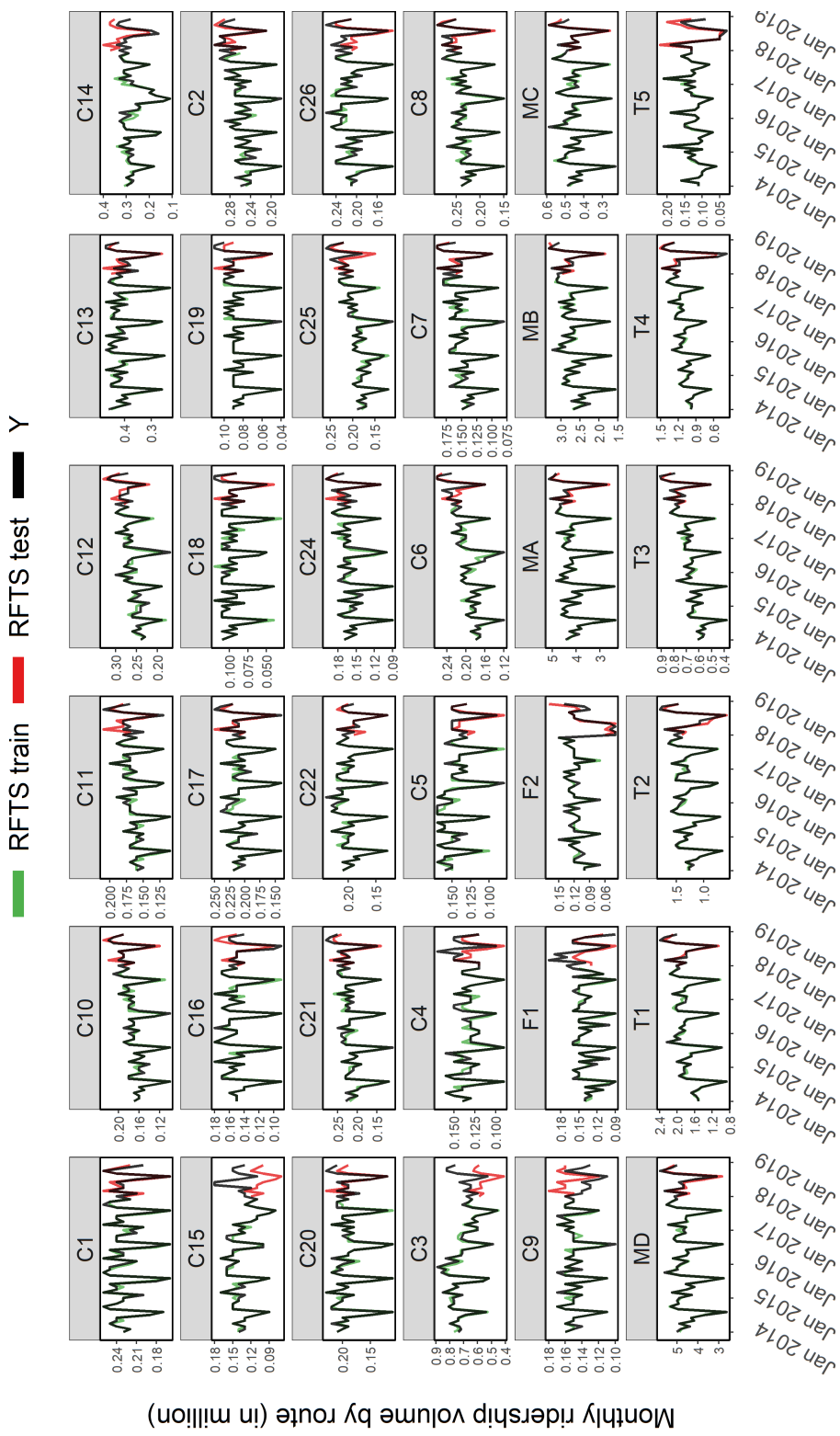
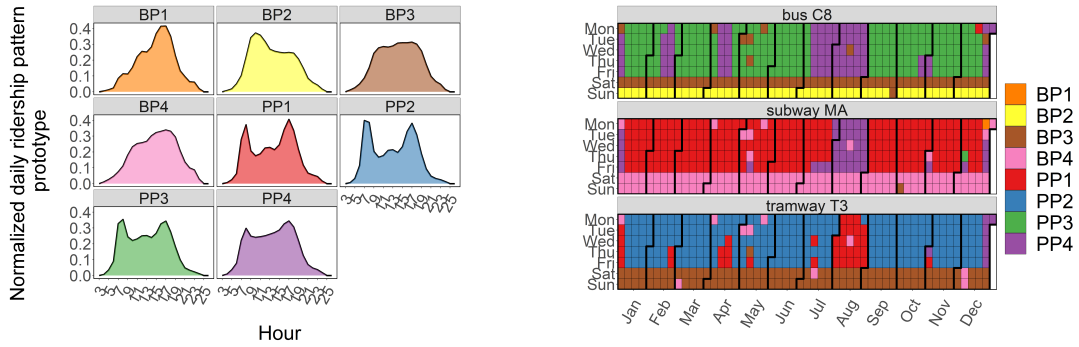


Figure 5: Monthly time series of ridership volume by route (RFTS train: training of decomposed random forest with supply feature (2014-2017), RFTS test: testing of decomposed random forest with supply feature (2018), Y: real monthly ridership by route), source: prepared by the authors

irrespectively of their magnitude i.e the ridership volume. It can be mathematically express as a minimization of the cosine distance of all possible allocation of objects  $m$  to clusters  $c(m) \in 1, \dots, k$  and over all prototypes  $p_1, \dots, p_k$  in the same feature space as object  $m$  denoted  $x_m$ :

$$\min \sum_m 1 - \frac{x_m \cdot p_{c(m)}}{\|x_m\| \cdot \|p_{c(m)}\|} \quad (8)$$

In our problem, the observation  $x_m$  is the vector of hourly forecasted route ridership volume on a given day. The features space dimension is 24 and the number of observations is equal to the number of days ahead we forecast multiply by the number of routes<sup>5</sup>. In prototype-based clustering, the number of clusters  $k$  need to be defined first. Then, a fixed-point algorithm can iterate between optimal allocation of observations to cluster and optimal prototypes to attempt to determine a partition that minimizes equation 8. For the sake of illustration, we have decided to retain 8 clusters accounting for almost 80% of the dissimilarity. The fixed-point algorithm was then run 20 times and the best solution was retained. The resulting clusters prototypes are given in figure 6a and were named according to their shape: bell shape clusters (BP) from 1 to 4 are daily ridership profiles that don't exhibit two peak periods while peak profiles clusters (PP) numbered from 1 to 4 are daily ridership profiles that exhibit a morning and evening peak pattern.



(a) Clustering prototype derived from RFTS model (random forest decomposed with supply)

(b) Calendar of forecasted daily prototype for three selected routes: bus route C8, subway route MA and tramway route T3)

Figure 6: Clustering analysis of model forecast, source: prepared by the authors

Those 8 clusters synthesize the high-resolution data contained in the base model. They facilitate the analysis and provide more manageable information to transportation planners and analysts. They can be used to further explore the future demand of each route. A typical use case would be to produce in addition to the previous analysis on the ridership volume a calendar of future daily ridership profile. The output of this procedure is given in figure 6b for three selected routes (bus route C8, subway route MA and tramway route T3). In this figure, it can be seen that the models have learned different daily ridership profiles for each route. For instance, bus route C8 is characterized by a change from prototype PP3 to prototype PP4 during holiday weekdays. It means that the model forecast that during school holidays the peak period will be less important with regard to the daily volume. Moreover, the model has learned that a typical Saturday will exhibit a BP3 profile while Sunday will exhibit a higher concentration of trip in the morning (prototype BP2). Subway route A daily profiles are overall more stable. Unlike route C8, Saturday and Sunday

<sup>5</sup>For 2018,  $N = 13140 = 36 \times 365$

422 prototype are equal. During weekday the assigned prototype remains the same (PP1) except during  
423 the month of august (PP4). Finally, the tramway route T3 is characterized by weekday with a high  
424 concentration of trip during morning peaks periods (prototype PP2). Compared to the subway  
425 route the weekend prototype is BP3 meaning that in proportion less ridership is forecasted in the  
426 late evening. These observations demonstrate that the models have learned and forecast distinct  
427 daily ridership profiles. Once cluster those forecasts can be examined more easily and help analysts  
428 to adjust service plans to intra-day demand variation. Altogether, this use case section offers  
429 some insights into how the proposed modelling framework could be used in real-world applications  
430 and answer common tactical issues. One of the main advantages of the proposed solution is the  
431 flexibility to explore forecasts at different levels of aggregation.

## 432 6 Conclusions

433 This research was undertaken with two objectives. First, to design a modelling framework suitable  
434 for medium-term forecasting of route ridership volume. Second to evaluate the potential benefit  
435 of such forecasts for transit agencies and operators. **The proposed model can identify trends in  
436 ridership over the past years in order to consider them for forecasting. It is also able to consider  
437 variability within each year due to seasonal effects, holiday, bank holiday, weekday effects and hour  
438 of the day. Finally, the proposed model can consider measures of supply evolution over time for  
439 each PT route.** The evaluation of different models has shown that the proposed decomposition for-  
440 mulation can learn complex patterns from historical data that can then be satisfactorily projected  
441 into the future. The resulting forecasts have proven to be quite appropriate to facilitate recurrent  
442 tactical decision making and help planning better services. The use case section outlined different  
443 applications **to demonstrate the potential of such a modelling approach for tactical planning to  
444 support PT operator or authority needs. Examples of applications include setting future goals,  
445 monitoring ridership volume at different levels of aggregations, estimating the impact of future  
446 disruptions and supporting the definition of future service provision.** It is therefore recommended  
447 that transit agencies complement their traditional reporting tools with these types of predictive  
448 approaches. By doing so, they can value at their full potential their historical data and enhance  
449 data-driven decision making. Moreover, the cost of implementation of such a tool is reduced and  
450 not too challenging. Models should simply be trained automatically by batch (e.g. every month)  
451 based on the above data preparation procedure. Then the provision of modern dashboarding  
452 tools with drill-down capability will make the data available through a web application for anyone  
453 interested in the organisation.

454 Although the study has demonstrated that the proposed modelling formulation generate co-  
455 herent forecasts and is operational, it also has several limitations that need to be considered and  
456 further investigated. First, the ensemble models are trained using the same weight for all historical  
457 observations. Even though this approach is simple and generic it would be interesting to elabor-  
458 ate methods to identify non-representative historical data that should not be part of the training  
459 set. We should also investigate mechanisms that put more weight on recent observations such as  
460 what is proposed for the estimation of trend component. Second, while we did incorporate one  
461 supply feature in the analysis, more research is required to establish additional evidence on which  
462 variables are needed and how they should be treated in the modelling process. In this regard, one  
463 direction would be to incorporate previously calculated elasticities factors in the model. Third, for  
464 this research, we assumed that higher hierarchies forecast can be obtained simply by summing the  
465 base forecast but other approaches need to be considered such as optimal forecast reconciliation  
466 or middle-out approach.

## 467 Highlights

- 468 • Examines the need for medium-term forecasting in public transit systems
- 469 • Proposes an operational modelling framework to forecast ridership one year ahead at different  
470 levels of spatiotemporal aggregation
- 471 • Outlines different applications of the forecasts for tactical planning and ridership monitoring
- 472 • Exploitation of archived passive data can enhance data-driven decision making

## 473 Author statement:

474 Empty for double-blind review policy.

## 475 Acknowledgements

476 Empty for double-blind review policy.

## 477 References

- 478 P. Bonnel. *Prévision de la demande de transport*. Presses de l'École Nationale des Ponts et  
479 Chaussées, Paris, 425p, 2002.
- 480 D. K. Boyle. *Fixed-route transit ridership forecasting and service planning methods*, volume 66.  
481 Transportation Research Board, 2006.
- 482 L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- 483 C. Buchta, M. Kober, I. Feinerer, and K. Hornik. Spherical k-means clustering. *Journal of*  
484 *Statistical Software*, 50(10):1–22, 2012.
- 485 M. Coleman, L. Tarte, S. Chau, B. Levine, and A. Reddy. A data-driven approach to prioritizing  
486 bus schedule revisions at new york city transit. *Transportation Research Record*, 2672(8):86–95,  
487 2018.
- 488 E. Diab, D. Kasraian, E. J. Miller, and A. Shalaby. Current state of practice in transit ridership  
489 prediction: Results from a survey of canadian transit agencies. *Transportation Research Record*,  
490 page 0361198119841858, 2019.
- 491 C. Ding, D. Wang, X. Ma, and H. Li. Predicting short-term subway ridership and prioritizing its  
492 influential factors using gradient boosting decision trees. *Sustainability*, 8(11):1100, 2016.
- 493 J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer  
494 series in statistics New York, 2001.
- 495 J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*,  
496 pages 1189–1232, 2001.
- 497 J. Hanft, S. Iyer, B. Levine, and A. Reddy. Transforming bus service planning using integrated  
498 electronic data sources at nyc transit. *Journal of Public Transportation*, 19(2):6, 2016.

- 499 R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- 500 K. B. Kahn. Revisiting top-down versus bottom-up forecasting. *The Journal of Business Fore-*  
501 *casting*, 17(2):14, 1998.
- 502 H. N. Koutsopoulos, Z. Ma, P. Noursalehi, and Y. Zhu. Transit data analytics for planning,  
503 monitoring, control, and information. In *Mobility Patterns, Big Data and Transport Analytics*,  
504 pages 229–261. Elsevier, 2019.
- 505 Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu. Forecasting short-term subway passenger flow under  
506 special events scenarios using multiscale radial basis function networks. *Transportation Research*  
507 *Part C: Emerging Technologies*, 77:306–328, 2017.
- 508 Z. Ma, J. Xing, M. Mesbah, and L. Ferreira. Predicting short-term bus passenger demand using a  
509 pattern hybrid approach. *Transportation Research Part C: Emerging Technologies*, 39:148–163,  
510 2014.
- 511 P. Noursalehi, H. N. Koutsopoulos, and J. Zhao. Real time transit demand prediction capturing  
512 station interactions and impact of special events. *Transportation Research Part C: Emerging*  
513 *Technologies*, 97:277–300, 2018.
- 514 D. Opitz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of artificial*  
515 *intelligence research*, 11:169–198, 1999.
- 516 J. d. D. Ortúzar and L. Willumsen. *Modelling transport*. John wiley & sons, 2011.
- 517 N. Paulley, R. Balcombe, R. Mackett, H. Titheridge, J. Preston, M. Wardman, J. Shires, and  
518 P. White. The demand for public transport: The effects of fares, quality of service, income and  
519 car ownership. *Transport policy*, 13(4):295–306, 2006.
- 520 M.-P. Pelletier, M. Trépanier, and C. Morency. Smart card data use in public transit: A literature  
521 review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568, 2011.
- 522 S. S. Pulugurtha and M. Agurla. Assessment of models to estimate bus-stop level transit ridership  
523 using spatial modeling methods. *Journal of Public Transportation*, 15(1):3, 2012.
- 524 F. Toqué, M. Khouadjia, E. Come, M. Trepanier, and L. Oukhellou. Short & long term forecasting  
525 of multimodal transport passenger flows with machine learning methods. In *2017 IEEE 20th*  
526 *International Conference on Intelligent Transportation Systems (ITSC)*, pages 560–566. IEEE,  
527 2017.
- 528 J. C. Totten and D. M. Levinson. Cross-elasticities in frequencies and ridership for urban local  
529 routes. 2016.
- 530 D. M. van de Velde. Organisational forms and entrepreneurship in public transport: classifying  
531 organisational forms. *Transport policy*, 6(3):147–157, 1999.
- 532 E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias. Short-term traffic forecasting: Where we are  
533 and where we’re going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.
- 534 Y. Wei and M.-C. Chen. Forecasting the short-term metro passenger flow with empirical mode  
535 decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*,  
536 21(1):148–162, 2012.
- 537 H. Zhai, L. Cui, Y. Nie, X. Xu, and W. Zhang. A comprehensive comparative analysis of the basic  
538 theory of the short term bus passenger flow prediction. *Symmetry*, 10(9):369, 2018.