Call for abstract - HEART 2020 - Lyon

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

Oscar Egu[1,*] and Patrick Bonnel[1]

[1]*LAET-ENTPE, Université de Lyon, CNRS, Rue Maurice Audin, 69518 Vaux-en-Velin Cedex, France*
[*]*Corresponding author: Oscar Egu, oscar.egu@entpe.fr*

Word Count: 2921

1 Research goals

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 [Bonnel, 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 addressed [Hyndman and Athanasopoulos, 2018]. In many organisation including public transport, they are three commonly accepted level of planning and organisational control [van de Velde, 1999, Pelletier et al., 2011]. Strategic level deal with long-term decisions and objectives. Tactical level focus on decisions that take place in 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 need to 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]. In the context of public transit (PT), authors argue that it 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]. 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, Bonnel, 2002].

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

1. To develop a generic forecasting approach that combined trend analysis with machine learning model to predict one year in advance, at different levels of spatiotemporal aggregation, the ridership volume in a PT network.

2. To illustrate how such forecasts can support tactical planning tasks and assist PT operators and agencies in monitoring ridership.

2 Modelling approach

Let's denote $Y_i = (y_{i,1}, ..., y_{i,t}, ..., y_{i,n})$ as the vector of observed ridership volume on element i of a PT network for each time step $t \in 1, ., .., n$. Let's denote $X_{i,t}$ the set of features (explanatory variable) whose value are known for $t \in 1, ., .., m$ where m $(m > n)$ is the prediction horizon. Our goal is to estimate a regression model $f_i(.)$ for each element i such that $y_{i,t} = f_i(X_{i,t}) + \varepsilon$ where ε is the difference between the observed value and the predicted value $\widehat{y_{i,t}} = f_i(X_{i,t})$.

As proposed by Toqué et al. [2017], to estimate $f_i(.)$ we can use machine learning model such as ensemble of decision trees. Those models have recently gained much popularities in various fields due to their ability to learn complex non-linear relation between features and provide higher accuracy than single machine learning models [Friedman et al., 2001, Opitz and Maclin, 1999]. One aspect of those machine learning models is that they assume independence between the $y_{i,t}$ and thus can't learn time-dependent structure and forecast potential trend. However, ridership volume may have statistical properties that evolve through time and might also be view as a non-stationary time series. To deal with this aspect, we can assume that ridership volume posses two components a yearly trend-cycle component $s_{i,T}$ and a yearly adjusted component $a_{i,t}$. Ridership level can then be expressed using multiplicative decomposition:

$$y_{i,t} = s_{i,T} \times a_{i,t} \tag{1}$$

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

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

$$\widehat{y_{i,t}} = \widehat{a_{i,t}} \times (1 + \alpha_i) s_{i,T-1} \tag{2}$$

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

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

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

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

$$\widehat{R_{J,M}} = \sum_{i \in J} \sum_{t=t1}^{t2} \widehat{y_{i,t}} \tag{4}$$

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

3 Results

3.1 Models fitting and forecast error

For this research, the fare transaction from 2014 to 2018 for the subway network (4 metro lines and 2 funicular lines), the tramway network (5 lines) and high-frequency buses known as line C (25 lines) were extracted from the Lyon PT operator data warehouse. Those routes were selected because they represent more than 80% of the total ridership of the network. This sample was also deemed to be enough to evaluate correctly our approach. Fare transactions were then summed by route and by hour. Hence, route corresponds to the base elements i of the network and the smallest temporal aggregation step t corresponds to hourly data. Table 1, depicts the list of models that we have decided to evaluate. The first model formulation is a naive historical approach. In this case, the forecast is computed using the median of all past time step over the historical observation with the same characteristic as the forecasted time step. Table 1 also indicate if the models were trained with the raw ridership volume or with the decomposed ridership volume. In the second case, the equation 2 was used to forecast future ridership volume.

In this research, the hold out approach was selected to evaluate the accuracy of the forecast. The data was thus divided between a training set comprising all observations from 2014 to 2017 and a test set consisting of 2018 observations. To measure forecast errors the mean absolute error (MAE), the median absolute percentage error (MdAPE) and the mean absolute percentage error (MAPE) where used. In table 2, we have reported the error measures on the train and test dataset

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 1: List of implemented models, source: prepared by the authors

for different levels of aggregations. Several observations can be made from that table. First, as expected when the aggregation level increases the error decrease significantly. A second important observation is that ensemble models outperform baseline model in terms of absolute errors (MAE) but also in terms of percentage errors (MAPEv and MdAPE). What is also clear from this table is that the decomposition approach is essential when historical data with a certain depth is used to model future ridership. This is because the decomposition ensures the coherency of the forecast with the most recent observed ridership volume. Finally, the introduction of supply feature in ensemble models can reduce the errors especially large errors.

		TEST SET (2018)							
		HM	**HMT**	**RF**	**RFT**	**RFTS**	**GB**	**GBT**	**GBTS**
Base level (route-hour)	MAE	131	115	118	99	95	123	104	**93**
	MAPE_50 (%)	21,7	21,7	19,9	19,8	**17,2**	21,1	21,4	17,4
	MdAPE (%)	14,5	12,8	13,7	11,9	**11,7**	14,4	12,6	11,7
Aggregated (route-day)	MAE	2662	2157	2339	1708	**1480**	2351	1720	1497
	MAPE (%)	36,3	36,9	26,8	26,7	**16,6**	28,1	27,8	17,7
	MdAPE (%)	10,7	7,7	9,9	6,8	**6,4**	10,0	6,9	6,5
Aggregated (hour)	MAE	3562	2878	2993	2140	2182	3226	2349	**1976**
	MAPE (%)	25,7	24,9	19,7	19,1	16,7	25,4	25,1	**16,4**
	MdAPE (%)	10,0	7,1	8,4	**5,7**	5,9	9,7	6,8	5,8
Aggregated (day)	MAE	75721	54068	63457	33658	31259	63643	35738	**30671**
	MAPE (%)	55,2	55,7	28,9	26,7	**9,0**	32,9	31,0	13,8
	MdAPE (%)	8,1	4,4	7,2	**2,8**	3,1	7,8	3,5	3,1

Table 2: 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 decomposition approach, S indicates the incorporation of supply feature), source: prepared by the authors

3.2 Use-case analysis

With the above modelling approach, it is possible to obtain coherent aggregated ridership forecasts. This type of forecast can be used for tactical decision making. For instance, they can assist senior executives in setting next year goals and monitoring their PT network with a more data-driven approach. Samely, every year, transit agencies need to prepared next year budget which obviously requires at one point some sort of ridership forecast. In practice, this process can be tedious and is often a matter of expert. In the case where different operators coexist in the same network, this task is even more complex and critical for the proper functioning of the system. Agencies could use the proposed approach to generate forecasts for all routes and then aggregated them to prepare budgets that match more closely the future variation of ridership and fit the financial/operational division of the network. To illustrate this first use case, the weekly forecast for the two random forest models obtained through decomposition (RFT and RFTS) by type of transportation mode is plotted in figure 1.

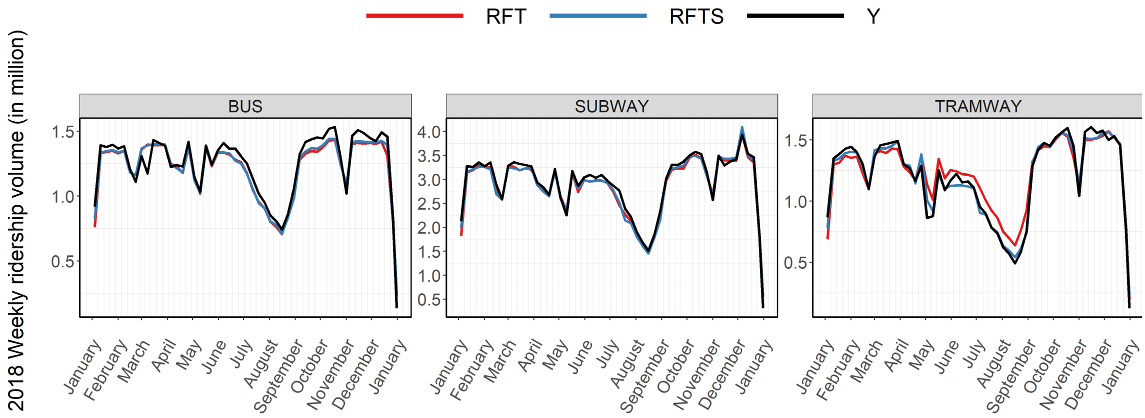
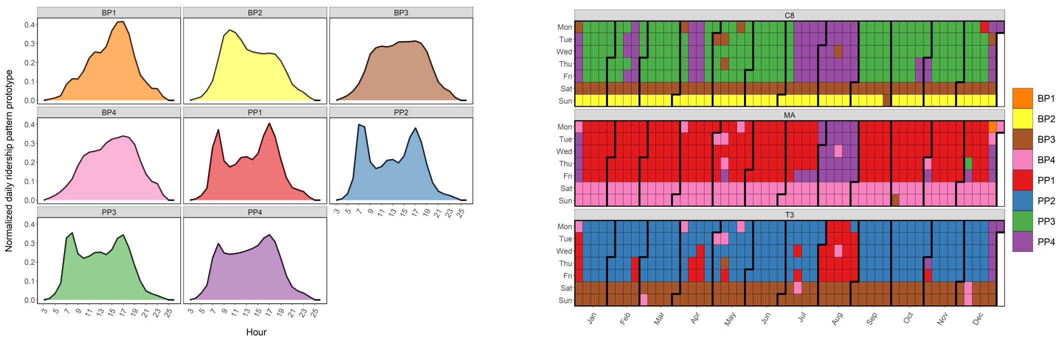


Figure 1: Weekly forecast one year in advance for the three transportation mode (RFT: decomposed random forest, RFTS: decompose random forest with supply features, Y: real ridership volume), source: prepared by the authors

In figure 1, it can be seen that the different series are overall quite close and the aggregation of models captures typical calendar patterns such as the impact of school holidays, public holidays, or citywide events (i.e peak at week 50 for subway network). It can also be seen that the ridership volume of high-frequency bus network has increased more quickly than anticipated by the two models, especially after September. Finally, figure 1 indicates that for the tramway routes both model yield different forecasts. This is because in 2018 the tramway network was impacted by important construction work of a new tramway line (T6) that leads to change in the service provision. Those type of elements can be further inspected by an analyst using the model output in a retrospective manner. Especially, the model can be used to identify routes whose behaviour deviate from historical behaviour as captured by the models. In doing so, operators could identify and prioritize which route they need to review first in a more efficient manner than doing it based on "anecdotal knowledge" or fixed cycle [Coleman et al., 2018]. Moreover, they can use these "business as usual" forecast to estimate in advance the number of trips impacted by a planned disruption. Once the period of disruption and the routes impacted are known, an analyst can interrogate models to retrieve the forecasted number of impacted trips. This information can then be used in the designed of replacement services or mitigation plan.

The base model i.e by route and hour can also be of value to apprehend in-depth ridership pattern. To this aim, forecast needs to be synthesized in such a way that the information can directly be helpful for tactical planning purposes. One way to do so would be to cluster the model output to identify a set of typical daily ridership profiles. Spherical kmean is a popular text mining prototype-based clustering algorithm [Buchta et al., 2012] and for the sake of illustration 8 clusters were derived with this method accounting for almost 80% of the dissimilarity. Those 8 clusters synthesize the high-resolution data contained in the base model. They facilitate the analysis and provide more manageable information to transportation planner 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 analysis on the ridership volume a calendar of future daily ridership profile. The output of this procedure is given in figure 2b 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 differents daily ridership profile for each route. For instance, bus route C8 is characterized by a change from prototype PP3 to prototype PP4 during holiday weekdays.



(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 2: Clustering analysis of model forecast, source: prepared by the authors

4 Conclusions

This research was undertaken with two objectives. First, to design a modelling approach suitable for medium-term forecasting of route ridership volume. Second, to evaluate the potential benefit of such forecast for transit agencies and operators. The evaluation of different models has shown that the proposed decomposition formulation can learn complex patterns from historical data that can then be satisfactorily projected into the future. The resulting forecasts have proven to be quite appropriate to facilitate recurrent tactical decision making and help planning better service. The use-case section outlined different applications such as setting future goals, monitoring ridership volume at different aggregations level of the network, estimating the impact of future disruption and supporting the definition of future service provision. It is therefore recommended that transit agencies complement their traditional reporting tools with these type of predictive approach. By doing so, they can value at their full potential their historical data and enhance data-driven decision making. Moreover, the cost of implementation of such models is limited and not too challenging. The model should simply be trained automatically based on the above procedure. Then the provision of modern dashboarding tools with drill-down capability will make the data available through a web application for anyone interested in the organisation.

References

P. Bonnel. *Prévision de la demande de transport.* Presses de l'École Nationale des Ponts et Chaussées, Paris, 425p, 2002.

D. K. Boyle. *Fixed-route transit ridership forecasting and service planning methods*, volume 66. Transportation Research Board, 2006.

C. Buchta, M. Kober, I. Feinerer, and K. Hornik. Spherical k-means clustering. *Journal of Statistical Software*, 50(10):1–22, 2012.

M. Coleman, L. Tarte, S. Chau, B. Levine, and A. Reddy. A data-driven approach to prioritizing bus schedule revisions at new york city transit. *Transportation Research Record*, 2672(8):86–95, 2018.

J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

K. B. Kahn. Revisiting top-down versus bottom-up forecasting. *The Journal of Business Forecasting*, 17(2):14, 1998.

H. N. Koutsopoulos, Z. Ma, P. Noursalehi, and Y. Zhu. Transit data analytics for planning, monitoring, control, and information. In *Mobility Patterns, Big Data and Transport Analytics*, pages 229–261. Elsevier, 2019.

Z. Ma, J. Xing, M. Mesbah, and L. Ferreira. Predicting short-term bus passenger demand using a pattern hybrid approach. *Transportation Research Part C: Emerging Technologies*, 39:148–163, 2014.

P. Noursalehi, H. N. Koutsopoulos, and J. Zhao. Real time transit demand prediction capturing station interactions and impact of special events. *Transportation Research Part C: Emerging Technologies*, 97:277–300, 2018.

D. Opitz and R. Maclin. Popular ensemble methods: An empirical study. *Journal of artificial intelligence research*, 11:169–198, 1999.

J. d. D. Ortúzar and L. Willumsen. *Modelling transport*. John wiley & sons, 2011.

M.-P. Pelletier, M. Trépanier, and C. Morency. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568, 2011.

F. Toqué, M. Khouadjia, E. Come, M. Trepanier, and L. Oukhellou. Short & long term forecasting of multimodal transport passenger flows with machine learning methods. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 560–566. IEEE, 2017.

D. M. van de Velde. Organisational forms and entrepreneurship in public transport: classifying organisational forms. *Transport policy*, 6(3):147–157, 1999.

E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19, 2014.

Y. Wei and M.-C. Chen. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies*, 21(1):148–162, 2012.