Short-Term Multi-Step Ahead Forecasting of Railway Passenger Flows During Special Events With Machine Learning Methods

Florian Toque, Etienne Come, Latifa Oukhellou, Martin Trepanier

CASPT 2018 Extended Abstract

Short-Term Multi-Step Ahead Forecasting of Railway Passenger Flows During Special Events With Machine Learning Methods

Florian Toqué · Etienne Côme · Latifa Oukhellou · Martin Trépanier

Abstract Forecasting of public travel demand is of great importance to public transport management. It is a very challenging task that relies on many kinds of dependencies, such as temporal, spatial or exogenous factors (e.g., weather, event, service breakdown, ...). This paper investigates the short-term multi-step ahead forecasting ($t+1$, ..., $t+8$) of passenger demand aggregated by time step of 15 minutes. The forecasting is performed with smart card data on a railway public transport network. Predicted flows could permit to optimize resource allocation, propose the best trip planning to passengers and better understand passenger flows during special events. We propose a state of the art deep learning approach, namely the gated recurrent unit (GRU), recurrent neural network, to tackle the short-term forecasting problem. We compared it to a well-known machine learning model namely Random Forest and long-term forecasting models. The experiments are conducted on a real 2-year smart card dataset provided by the transport organization authority of Ile-de-France (Ile-de-France Mobilités). The dataset depicts the passenger demand of 30 stations of the main Paris business district named La Défense, which corresponds to different transportation modes such as train (suburban railway service), metro, RER (Regional Express Network) and tramway. The evaluation of the models focuses on their performances in the presence of specific events through two subsets of data extracted from the whole dataset. These special periods correspond to transport network service anomaly periods such as service breakdown and special days period in term of passenger flow patterns such as public holiday.

Keywords Big Data · Machine Learning · Smart Card Data · Passenger Demand Forecasting · Public Transport · Forecasting Under Special Events

Florian Toqué, Etienne Côme, Latifa Oukhellou
Université Paris-Est, IFSTTAR, COSYS-GRETTIA
Champs-sur-Marne, France
E-mail: florian.toque@ifsttar.fr, etienne.come@ifsttar.fr, latifa.oukhellou@ifsttar.fr

Martin Trépanier
Ecole Polytechnique de Montréal, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (CIRRELT)
Montréal, Canada
E-mail: mtrepanier@polymtl.ca

1 Introduction

Every day a large amount of digital footprints is generated by citizens during their trips (GPS, WiFi, Bluetooth, smart card, social networks logs, etc.). These data can hold a high descriptive value of the urban mobility that could be helpful to develop innovative tools. Such tools can be useful to improve quality of services and to better organize the mobility of citizens in a city. Public authorities wish to develop sustainable urban mobility practices that aim to increase the use of public transport and reduce the pollution involved by private car use. In order to achieve such challenge, the understanding of urban mobility across public transport network is essential. We focus the study on one of the mainly used digital footprints which is the smart-card data collected by automated fare collection systems.

As early as 2004, Bagchi and White (2) studied the potential role that smart card data can play for travel behavior analysis and considered their potential to complement existing data collection such as survey methods. Since 2005, a multitude of studies highlighted the potential uses of smart card data in transit planing and transit uses (3; 24; 18; 19). One of the major disadvantage of this type of data, is the lack of information about the destination of passenger travels. In this respect, researches have been oriented to improve this source of data with studies like data completion and enrichment with the aim to estimate origin-destination (16) or to validate travel behavior estimation (15) and to perform trip purpose inference (10).

In order to encourage the use of public transport, transit planners need to correctly adapt the fare proposed to the passengers. To achieve such challenge, an in-depth understanding of the passenger behavior seems essential. In this context, passenger behavior analysis is a field that has been widely studied, which concerns the clustering of passengers depending on their transportation network activity (1; 13; 28; 21; 9; 14), or the characterization of station pattern (21).

Another parameter that directly impacts the satisfaction and the increase of public transport passengers is a good equilibrium between the offer and the demand of transport services. In this respect, urban mobility researchers recently studied the passenger demand with a focus on the short-term forecasting of railway passenger flows in (7; 8; 22; 23). Thereafter, researchers have oriented their work on the spatio-temporal aspect of the forecasting problem. For this purpose, researchers have worked on traffic forecasting with deep neural network and more precisely with combined convolutional and recurrent neural network in (25; 5) or graph convolutional neural network in (26). In the same line of work, authors of (27) have predicted the Citywide crowd flows with spatio-temporal convolutional neural network, and authors of (29) have designed spatio-temporal neural network to predict different sources of related space-time series.

Recently, studies involving passenger flows forecasting in public transport network have been oriented under atypical event. In (11), the authors use a multiscale radial basis function networks to perform a short-term multi-step ahead forecasting (t+15min, t+30min) subway passenger flow under special events scenarios without exogenous data about special event. In (17), short-term prediction approaches were developed to forecast subway passenger flows for the next 4 hours using social media data. The authors focused on the prediction of the total number of passengers (sum of entry and exit) of one subway station of the New York City network. They proposed a two-step methodology: hashtag-based event detection followed by the combined use of linear regression and a seasonal autoregressive moving average model using the social media event as exogenous features in input of the model.

Urban mobility forecasting in public transport is a very challenging task, since it depends on plethora of external factors such as calendar, weather, events or service breakdown. These factors can be hardly taken into account because of their high temporal and spatial irregularity. In this paper, we mainly focus on the problem of short-term multi-step ahead passenger forecasting demand under special events, in order to be able to face some abnormal behavior in the data caused by temporal factors like days off, Christmas Eve, New Year's Eve and exogenous factors such as service breakdown. We can summarize the main contribution of this study in the three following points:

- Comparison of short-term forecasting model of the state of the art with multi-step ahead horizon.
- Comparison between short-term multi-step ahead forecasting models with long-term forecasting models, in order to evaluate until which step ahead, it is useful to perform short-term forecasting against long-term forecasting.
- Analyze short-term multi-step ahead forecasting results under special events. With special events corresponding to days with anomalies on the public transport network or specials periods such as public holiday.

2 Short-term forecasting

2.1 Context and objective

Passenger demand forecasting performed in a long-term horizon (until one year ahead) can be very useful to transit planners for strategic long-term planning of the network in order to reduce operating costs and optimize vehicle use over the network. Nevertheless, long-term prediction has inherent disadvantages in forecasting passenger demand under special event scenarios. Indeed, these special events can cause disruptive impacts on the transportation system, which are hard to predict because they correspond to abnormal behavior patterns. These special events could be categorized, into two groups (i) Anticipated events which may be known or expected in advance such as weather, cultural or sporting event, etc. Collect such event database could be difficult but nonetheless useful to incorporate in demand prediction models (20) (ii) Unexpected events particularly those corresponding to service breakdown, or events that are not often or never observed in the historical dataset. Taking into account such type of special event, demand prediction should be formalized as a short-term forecasting problem, requiring the use of last observations of the time series in the prediction model (11). We focus this study on the short-term forecasting of passenger demand by using only passenger demand historical database and calendar information such as Month, Day, Holiday, Public School Holiday, Extended Weekend, Christmas Eve and New Year's Eve. In this paper, a clear emphasis is given to the evaluation of the short-term multi-step ahead forecasting performances on an extracted subset of a 2 years smart card dataset, including special events. The main objective of this work is to achieve the more accurate passenger flows forecasting in the case of specific events, by means of dedicated forecasting models able to capture the behavior of passengers in such situations. Accurate forecasting results will help transit operators, to improve the re-routing of passengers and to adapt the supply transport the more precisely to fit the passenger demand, in case of atypical event.

We compare different type of short-term models that forecast the number of passengers entering each station during the next 8 time step of 15 minutes (2 hours). In order to evaluate the models in case of abnormal patterns, we select two subset of data that correspond to abnormal periods:

- Days off, Christmas Eve and New Year's Eve: this period can be known in advance and correspond to days with very specific behavior.

– Days with anomaly on the transportation network system: this period is selected with a database given by the transport organization authority. It corresponds to service breakdown due to different sources, such as, technical problems, fire incident or strike of transit operator.

2.2 Forecasting methodology

We work on multi-step short-term forecasting using past historical observation and exogenous features to predict the number of passengers entering the studied railway stations at the 8 next time step (data are aggregated with 15 minutes time step). Both univariate, multivariate and multivariate to univariate models are considered to forecast passenger flows. We compared six different models: a long-term model baseline, a long-term machine learning model, a naive short-term model, a statistical model, a machine learning model and a deep learning model with different variants. Each input and output of short-term models are described in Table 1.

Table 1 Inputs and outputs of passenger forecasting models at time step t on day d

Model	Input	Output
HA	Day (Monday, .., Sunday) and time step feature	$\hat{Y}_t^A$
RF LT	time information τ	$\hat{Y}_t^A$
LOCF	Y_{t-1}^A	Y_{t-1}^A
VAR	$Y_{t-n:t-1}^A$	$\hat{Y}_t^A$
RF ST UNI	$Y_{t-n:t-1}^s$ and time information τ	$\hat{Y}_t^s$
RF ST MULTI	$Y_{t-n:t-1}^A$ and time information τ	$\hat{Y}_t^A$
GRU UNI	$Y_{t-n:t-1}^s$ and time information τ	$\hat{Y}_t^s$
GRU MULTI	$Y_{t-n:t-1}^A$ and time information τ	$\hat{Y}_t^A$
GRU FUSION	$Y_{t-n:t-1}^A$, $Y_{t-n:t-1}^s$ and time information τ	$\hat{Y}_t^A$
GRU MULTI-UNI	$Y_{t-n:t-1}^A$ and time information τ	$\hat{Y}_t^s$

Where τ is the time information defined as time step t, encoded as integer value (1-96) and day encoded as day type dt with the vector of categorical variables described in Section 2.2.4. $\hat{Y}_t^A$ represents the predicted number of passengers in all the studied stations and $\hat{Y}_t^s$, the predicted number of passengers in the station s at time step t.

2.2.1 ***Historical average, HA***

This model model predicts the number of passenger in the corresponding time step by averaging the value of the historical observations by day (Monday, Tuesday, ..., Sunday). The prediction at 10:00 am on Monday corresponds to the average of all the historical values for Monday at 10:00 am. This model is the baseline of the long-term forecasting model.

2.2.2 ***Last observation carried forward, LOCF***

It is a naive method used in short-term time series forecasting. It returns the last observed value (observation at the previous time step).

2.2.3 *Vector autoregressive, VAR*

It is a well known statistical model (12) used in time series forecasting. It is a variant of the univariate autoregressive model (AR) which is a time-series regression model that predicts the next value by linearly taking into account a stochastic term and the previous observation. The VAR method is a multivariate linear forecasting approach that captures the linear interdependencies among the different time series to perform prediction. Each predicted variable can be forecasted by resolving an equation representing its evolution based on its own past observation and the past observation of the other variables.

2.2.4 *Random forest, RF*

The random forest is a machine learning model that has been used for several real-world applications on problem of regression or classification. This model has been introduced by Breiman (4), it is an ensemble learning algorithm based on the average prediction of different decision trees (forest). The results obtained by the different trees make the RF more accurate and robust than a unique decision tree.

Random forest long-term, RF LT: This model forecast the number of passenger on each station until one year ahead. The inputs of the model are only the information of the time and date defined as day type, encoded with the following features:

- Day of the week (1-7): Monday, Tuesday, ..., Sunday.
- Month (1-12): January, February, ..., December.
- Public holiday (0-1): Public holiday in the studied region (for this study, Paris area).
- Extra day off (0-1): Working day of an extended weekend between public holiday and weekend, e.g., Friday is considered as an extra day off if the day before (Thursday) is a public holiday.
- School holiday (0-1): Day in period of school holiday.
- Christmas Eve (0-1): December 24.
- New Year's Eve (0-1): December 31.

Random forest short-term univariate, RF ST UNI: The RF short-term model is a univariate model, in order to forecast all the station one model per station is created. This RF ST UNI takes the n last observed values, often named the lag, the information of time and date defined as day type with the features detailed in Section 2.2.4.

Random forest short-term multivariate, RF ST MULTI: In order to take into account the transport network connection in the forecasting, we create a second type of RF model that is a multivariate variant. The principal objective is to learn information about incoming passenger flow from the predicted station and other stations of the transport network. The model takes the n last observed values of all the predicted stations, the day type defined with the features depicted in Section 2.2.4, and the time step in input.

2.2.5 *Gated recurrent unit, GRU*

This model is a variant of recurrent neural networks (RNNs) and has been introduced by (6). RNNs consider that outputs are dependent on previous computations in contrast with traditional neural networks where inputs (and outputs) are temporarily considered independent of each other. Recurrent neural networks keep a ”memory” of previous calculations under the form of a constantly updated hidden state. GRU, as well as long short-term memory (LSTM) [35], are a special type of RNN with a gate mechanism that prevents the vanishing gradient problem [36] associated with the base RNN model in order to be able to learn long-term range dependencies. Unlike the LSTM, GRU are composed of a simpler gate mechanism that allows a faster computation time during the learning and predicting step. Currently, GRU are state-of-the-art models for problems involving time series analysis (e.g., machine translation, music modeling). In this study we propose four different versions of the GRU model:

Gated recurrent unit univariate, GRU UNI The first version is an univariate model that uses the past observations of one time series in addition to time step and date information to predict the next value of the same time series. One model per station is created.

Gated recurrent unit multivariate, GRU MULTI In order to take into account the spatial information in the prediction model, we created a gated recurrent model with a multivariate shape. This model uses the past observations of all the time series to predict the next values of all the stations.

Gated recurrent unit fusion, GRU FUSION As multivariate and univariate models perform differently depending on the observed passenger flows, we created a model that combines the predictions of both multivariate and univariate models. This method is a combined model that links a weight with each of the multivariate and univariate model prediction. These weights are generated by an attention mechanism that captures the importance of each model (multivariate and univariate) at each time step of the prediction. The final output detailed in Equations 1 and 2 corresponds to the weighted sum of the model predictions.

$$\alpha_s(t) = \phi(\mathbf{H}_s^U(t)) \tag{1}$$

$$\hat{Y}_s^F(t) = \alpha_s(t)\hat{Y}_s^U + (1-\alpha_s(t))\hat{Y}_s^M \tag{2}$$

where $\hat{Y}_s^F(t)$ is the prediction of the number of passengers at station s at time step t of the GRU FUSION model (M corresponds to the GRU MULTI model and U to the GRU UNI model). The weights $\alpha_s(t)$ (one weight per station) are obtained by the attention mechanism represented by the sigmoid function ϕ over the hidden state ($\mathbf{H}_s^U(t)$) of the univariate GRU model at time step t.

Gated recurrent unit multivariate to univariate, GRU MULTI-UNI To account for the spatial information in the prediction model, we compute a multivariate-univariate variant of the gated recurrent model. This model uses the past observations of all the time series to predict the next values of one station. This variant involves the computation of one model per time series.

The univariate and multivariate-univariate approach builds as many models as there are stations, whereas the multivariate approach uses one model to forecast the passenger flows for all stations.

2.3 Multi-step ahead forecasting

Short-term multi-step ahead forecasting aims to predict the values of all stations A at time step $t+n$, with $n > 1$. In this study, we forecast until 8 step ahead, $\{Y_{t+n}^{A} | n \in [\![1,8]\!]\}$, with 15 minutes aggregated data. Multi-step ahead forecasting is a challenging task because of the lack of updated observation and the accumulated errors of the model. Different methods of multi-step ahead forecasting can be built. Here we have adopted the iterative method meaning that we iteratively use the prediction at time $t+1$ as input variables for the next step prediction. Thus, we can repeatedly predict step by step until obtaining the target horizon forecasting.

3 Experiments

3.1 Problem case study

To evaluate our models, a real 2 years smart card dataset (2014-2015) with passenger demand aggregated per 15 minutes is used. This dataset is provided by the transport organization authority of Ile-de-France (Ile-de-France Mobilités) and depicts the number of passengers entering the 30 stations of the multimodal transport network of the business district La Défense of Paris in France. The considered stations include different transport modes such as train (suburban railway service), metro, RER (Regional Express Network) and tramway. The considered district includes one tramway line with 13 stops and five metro, RER and train lines with 17 stations. Each day the studied railway stations handle more than 215k passengers while the tramway line transport 35k passengers. In order to highlight the analysis of passenger demand forecasting under special events, the evaluation of the proposed models is performed on the whole dataset as well as on a subset of data which contains only periods with non-regular behaviors.

A weeklong example (Monday, February 23, 2015 to Sunday, March 1, 2015) of the most visited stations of each transportation mode (train and tramway) with data aggregated by 15 minutes is depicted in Figure 1. This figure shows that tram stations has noisy pattern, moreover it reveals for all transport modes, a familiar temporal trends in public transport usage such as the difference between week-end and weekday characterized by activity peaks in the morning the evening and by low passenger flows during the weekend. The greater number of passenger during the evening activity peak than the morning activity peak of the train station "La Défense Grande Arche" is due to the fact that this station is located in the center of the business district.

3.2 Evaluation methods

We evaluate the results obtained by the different forecasting models with several well-known metrics. In order to have a better understanding of the errors, two different measures of prediction accuracy were used, namely the Root Mean Square Error (RMSE) and the Mean Average Percentage Error at v (MAPE@v). The errors can be expressed as:

$$\text{RMSE} = \sqrt{\frac{\sum_{s=1}^{S} \sum_{t=1}^{T} (\hat{y}_s(t) - y_s(t))^2}{T \times S}} \tag{3}$$

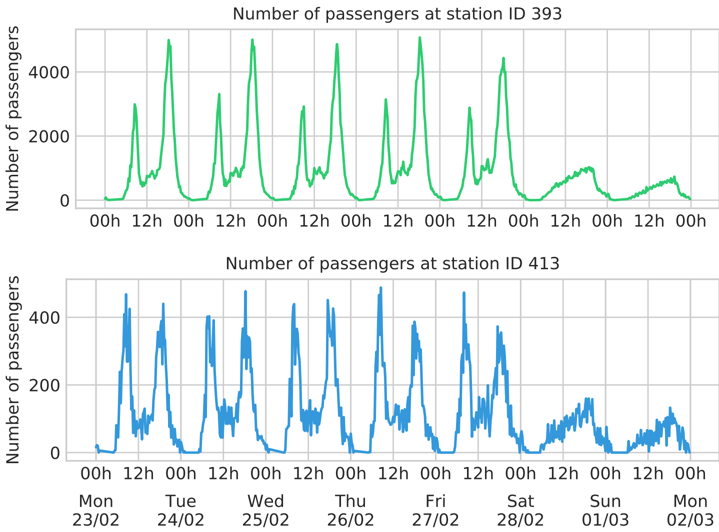

Fig. 1 Weekly pattern of the number of ticketing logs of the most visited train and tram stations aggregated by 15 minute intervals. The weeklong example extends from Monday, February 23, 2015 to Sunday, March 1, 2015.

$$\forall \mathrm{y_s(t)} > \mathrm{v},\ \mathrm{MAPE@v} = \frac{100}{T \times S} \times \sum_{s=1}^{S} \sum_{t=1}^{T} \left| \frac{y_s(t) - \hat{y}_s(t)}{y_s(t)} \right| \tag{4}$$

where $\hat{y}_s(t)$ is the forecast value of station s at time step t, $y_s(t)$ is the actual value, μ_s is the mean of the observed values y_s and S is the station number.

As explained in Figure 2, four training and test sets were used for the short term prediction. Short-term forecasting errors over the test set have been performed by concatenating the prediction of the 4 test sets. This kind of split better reflects the real situation where the model will be updated every quarter. Days with free transportation because of pollution peaks were removed from the dataset. There were four such days in 2014 and six days in 2015 (Friday 14 to Monday 17 March 2014, Saturday 21 to Monday 23 March 2015 and from Sunday 29 to Monday 30 November 2015). The transport was also free on Sunday 11 January 2015 because of the people's march through Paris after the terrorist attack. We also remove from the dataset the period between 23 July and 24 August 2015, because of the renovation work on the principal line of the study (RER A line).

4 Results

The results for the two long-term models and the four short-term prediction models detailed in Section 2.2 are summarized in terms of RMSE in Table 2 and MAPE@5 in Table 3. According to the evaluation protocol explained in Section 3.2, the performances for both the global training and test sets are given in this table. The high difference between the train set and the test set in terms of MAPE and RMSE error of the multivariate models (RF ST MULTI and GRU FUSION), demonstrate that these models are overfitted. The univariate models RF ST UNI and GRU UNI perform the best prediction in the test set in terms of

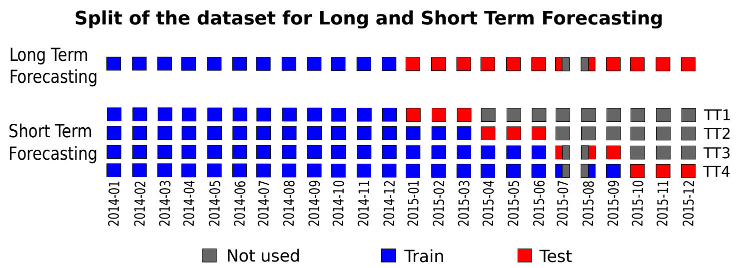

Fig. 2 Train sets are colored in blue, test sets are red and unused data is shown in gray. Four training and test sets (TT1, TT2, TT3, TT4) were used to capture real situations (the model is updated every three months) for the short-term forecasting.

RMSE and MAPE errors. Besides these analysis we can also observe that recurrent neural network GRU models perform better multi-step prediction than the others short-term models and seems to be less sensible of error propagation between the different multi-step prediction.

Table 2 Results of the short-term multi-step ahead prediction models for passengers entering railway and tram stations in 15 min intervals on the global training and test set, in term of RMSE.

RMSE - Train set (2014) - Global set								
	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
HA	87.19	87.19	87.19	87.19	87.19	87.19	87.19	87.19
RF LT	39.18	39.18	39.18	39.18	39.18	39.18	39.18	39.18
LOCF	64.28	104.56	145.12	182.33	217.59	248.78	275.92	299.37
VAR	31.34	38.12	45.74	53.26	61.02	67.74	73.24	77.92
RF ST UNI	20.63	27.14	31.86	35.64	39.20	42.43	45.35	47.85
RF ST MULTI	**13.02**	**18.79**	**23.74**	**28.17**	32.61	36.76	40.68	44.20
GRU UNI	27.22	29.26	31.71	33.61	35.11	36.45	37.57	38.5
GRU MULTI	27.77	30.35	32.54	34.27	35.68	36.87	37.93	38.85
GRU FUSION	25.04	26.99	29.06	30.67	**31.95**	**32.98**	**33.84**	**34.55**
GRU MULTI-UNI	26.69	29.79	32.62	34.67	36.38	37.85	39.21	40.43
RMSE - Test set (2015) - Global set								
HA	88.03	88.03	88.03	88.03	88.03	88.03	88.03	88.03
RF LT	57.29	57.29	57.29	57.29	57.29	57.29	57.29	57.29
LOCF	65.72	107.52	149.86	188.23	224.74	256.85	285.08	309.38
VAR	34.60	42.41	51.35	59.60	68.42	76.18	82.82	88.58
RF ST UNI	31.34	34.42	38.08	40.84	44.12	46.78	49.50	51.51
RF ST MULTI	41.00	42.92	45.42	47.57	50.26	52.41	54.43	56.21
GRU UNI	**31.27**	**33.14**	**36.02**	**38.20**	**40.53**	**42.44**	**44.45**	**46.08**
GRU MULTI	37.74	41.54	44.64	46.83	48.67	50.23	51.61	52.79
GRU FUSION	32.33	34.74	37.74	40.23	42.52	44.30	45.92	47.28
GRU MULTI-UNI	34.98	38.70	42.56	45.09	47.08	48.77	50.39	51.91

HA, LOCF, VAR, RFs and GRUs are the models described in Section 2.2, and RMSE and MAPE@5 are the measures of prediction defined in Section 3.2.

In order to better understand how the models work, we focus the analysis of the forecasting results during two specific periods (i) Public holiday and special day period and (ii) transport network anomaly period. Indeed, such special periods are harder to predict and specifically at multi-step ahead.

Table 3 Results of the short-term prediction models for passengers entering railway and tram stations in 15 min intervals

MAPE@5 - Train set (2014) - Global set								
	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
HA	40.39	40.39	40.39	40.39	40.39	40.39	40.39	40.39
RF LT	26.32	26.32	26.32	26.32	26.32	26.32	26.32	26.32
LOCF	41.08	49.24	60.14	71.36	84.52	97.06	109.40	122.01
VAR	32.09	36.31	41.32	46.11	51.23	55.21	58.42	60.83
RF ST UNI	17.81	21.23	23.41	24.86	26.07	27.01	27.88	28.77
RF ST MULTI	**10.55**	**14.50**	**17.86**	**20.56**	**22.94**	24.97	26.72	28.24
GRU UNI	24.16	24.43	24.77	25.07	25.35	25.57	25.80	26.08
GRU MULTI	24.78	25.36	25.85	26.18	26.45	26.68	26.96	27.21
GRU FUSION	22.25	22.85	23.38	23.78	24.10	**24.36**	**24.59**	**24.86**
GRU MULTI-UNI	24.08	24.96	25.68	26.20	26.65	27.04	27.42	27.82
MAPE@5 - Test set (2015) - Global set								
HA	40.80	40.80	40.80	40.80	40.80	40.80	40.80	40.80
RF LT	33.89	33.89	33.89	33.89	33.89	33.89	33.89	33.89
LOCF	40.75	49.00	60.16	71.49	84.58	97.30	110.13	123.2
VAR	34.29	39.49	45.51	51.18	57.41	62.44	66.62	69.76
RF ST UNI	**27.49**	28.11	28.83	29.50	30.04	30.49	30.90	31.4
RF ST MULTI	30.29	30.82	31.36	31.85	32.40	32.78	33.11	33.43
GRU UNI	27.58	**27.94**	**28.52**	**28.98**	**29.43**	**29.81**	**30.19**	**30.52**
GRU MULTI	28.46	29.21	29.93	30.43	30.87	31.25	31.58	31.88
GRU FUSION	28.25	28.61	29.14	29.57	29.99	30.29	30.60	30.97
GRU MULTI-UNI	27.97	28.62	29.32	29.84	30.30	30.73	31.22	31.67

HA, LOCF, VAR, RFs and GRUs are the models described in Section 2.2, and RMSE and MAPE@5 are the measures of prediction defined in Section 3.2.

4.1 Public holiday, Christmas Eve and New Year's Eve

Public holidays, Christmas Eve and New Year's Eve, are special days during which passenger demand can be considerably different from normal. Here, we analyze the prediction of passengers flows aggregated by 15 minutes computed with the multi-step forecasting models at time step $t+n$ with n in the range $[1,8]$. In table 4 we can observe the RMSE and MAPE errors on the sub-test set corresponding to these special types of days (12 days in 2015).

The best results in term of RMSE are obtained by the univariate GRU model until time step $t+3$, then it is the long-term model RF that performs the best prediction. These results could be explained by the fact that short-term models give more importance to past observation than exogenous calendar features during the prediction and that RMSE explodes after several multi-step forecast during such special periods. In term of MAPE, multi-step results are less impacted by the exploding error propagation. The best models are the multivariate models GRU, due to his high capability to generalize, the GRU FUSION model is the best prediction model in term of multi-step horizon forecasting.

4.2 Forecasting with Transport network anomaly periods

To quantify the robustness of multi-step ahead prediction models under special events, we evaluate their performances within special days periods created with an anomaly database provided by the transport authority of Ile-de-France (Ile-de-France Mobilités). The anomalies of this database can be categorized into the following types:

Table 4 Results of the long- and short-term prediction models for passengers entering train and tram stations during three special periods: public holidays, Christmas Eve and New Year's Eve, in term of RMSE and MAPE errors.

RMSE - Sub test set (2015) - Public holiday, Christmas Eve and New Year's Eve (12 days)								
	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
HA	294.64	294.64	294.64	294.64	294.64	294.64	294.64	294.64
RF LT	31.27	31.27	31.27	**31.27**	**31.27**	**31.27**	**31.27**	**31.27**
LOCF	26.69	36.32	45.81	53.70	61.95	68.28	74.00	78.80
VAR	24.57	36.08	47.25	56.83	65.86	74.00	80.93	86.62
RF ST UNI	26.57	37.29	45.97	53.66	62.26	70.56	78.21	84.61
RF ST MULTI	25.73	31.70	39.32	44.37	49.84	53.82	57.91	64.14
GRU UNI	**21.38**	**25.34**	**30.71**	35.54	39.53	42.59	45.16	46.90
GRU MULTI	23.57	29.20	34.67	38.77	41.35	43.07	44.26	45.04
GRU FUSION	21.52	26.58	31.65	35.77	38.44	40.40	42.14	44.18
GRU MULTI-UNI	23.15	30.79	39.14	46.04	51.65	56.16	59.70	62.14
MAPE@5 - Sub test set (2015) - Public holiday, Christmas Eve and New Year's Eve (12 days)								
	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
HA	253.59	253.59	253.59	253.59	253.59	253.59	253.59	253.59
RF LT	39.06	39.06	39.06	39.06	39.06	39.06	39.06	39.06
LOCF	43.62	46.73	51.60	56.23	61.59	67.66	73.23	78.58
VAR	38.18	44.76	53.83	63.21	74.04	84.56	94.23	102.47
RF ST UNI	35.53	38.40	41.33	44.49	47.89	50.94	53.91	56.66
RF ST MULTI	33.62	35.15	36.57	38.03	39.78	41.73	43.62	45.78
GRU UNI	33.82	34.54	35.76	36.94	37.92	38.95	39.96	40.71
GRU MULTI	32.64	33.82	35.25	**36.50**	37.68	39.00	40.20	40.91
GRU FUSION	33.16	34.22	35.45	**36.50**	**37.28**	**38.07**	**38.76**	**39.35**
GRU MULTI-UNI	**31.86**	**33.12**	**34.93**	36.85	38.65	40.78	43.50	45.51

HA, LOCF, VAR, RFs and GRUs are the models described in Section 2.2, and RMSE and MAPE@5 are the measures of prediction defined in Section 3.2.

- Breakdown or external problem, such as fire.
- Renovation of the station.
- Rail workers strike.
- Special days with an opening of the AFC systems.

We filtered this database in accordance with the station and period study, resulting in a subset of 45 days during the test set period (2015).

Table 5 depicts the multi-step ahead prediction results of the models, in terms of RMSE and MAPE during the transport anomaly period.

We can observe that long-term models have difficulty to forecast the number of passenger during anomaly periods in terms of RMSE and MAPE. In contrast, short-term univariate model GRU UNI performs equivalent result to the global performance in Table 2 and Table 3. The results in terms of MAPE error show that GRU models are better prediction models for multi-step ahead prediction. The short-term model results on this special period that contains not predictable events, demonstrate that short-term models perform better prediction at multi-step ahead than long-term models.

5 Conclusion

Passenger demand prediction could be useful for trip planning, transport operation and passenger information. From this perspective, we investigated the short-term multi-step prediction of passenger flows with real smart card dataset.

Table 5 Results of the long- and short-term prediction models for passengers entering train and tram stations during special periods: sub-dataset of days with anomalies.

	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$	$t+6$	$t+7$	$t+8$
RMSE - Test set (2015) - Sub-dataset transport anomaly								
HA	114.76	114.76	114.76	114.76	114.76	114.76	114.76	114.76
RF LT	97.19	97.19	97.19	97.19	97.19	97.19	97.19	97.19
LOCF	56.65	90.87	125.85	157.37	187.52	213.98	237.14	257.15
VAR	37.13	48.66	61.02	71.91	82.76	92.71	100.36	107.20
RF ST UNI	32.32	38.39	44.93	50.53	55.76	59.54	63.21	66.06
RF ST MULTI	62.86	66.00	71.52	76.52	82.22	86.15	89.07	91.24
GRU UNI	**32.29**	**35.63**	**40.88**	**45.65**	**50.67**	**55.44**	**60.09**	**64.10**
GRU MULTI	52.55	59.83	66.37	71.13	75.19	78.83	81.94	84.56
GRU FUSION	37.60	42.22	48.54	54.17	59.29	63.56	67.49	70.81
GRU MULTI-UNI	44.22	51.09	58.57	63.72	67.92	71.47	74.82	78.00
MAPE@5 - Test set (2015) - Sub-dataset transport anomaly								
HA	56.96	56.96	56.96	56.96	56.96	56.96	56.96	56.96
RF LT	47.71	47.71	47.71	47.71	47.71	47.71	47.71	47.71
LOCF	41.55	48.24	58.4	68.21	79.17	90.2	101.54	112.44
VAR	37.83	44.68	52.93	60.70	69.17	76.73	83.27	88.26
RF ST UNI	**29.91**	**31.21**	**32.60**	34.37	35.60	36.92	38.05	39.18
RF ST MULTI	39.02	40.55	42.15	43.63	45.14	46.34	47.23	48.46
GRU UNI	30.67	31.34	**32.60**	**33.81**	**35.08**	**36.25**	**37.46**	**38.37**
GRU MULTI	35.00	36.63	38.52	39.84	41.15	42.27	43.16	43.83
GRU FUSION	32.35	33.19	34.76	36.01	37.24	38.05	38.99	40.00
GRU MULTI-UNI	33.07	34.40	35.99	37.21	38.39	39.63	41.22	42.28

LOCF, VAR, RF ST UNI and RF ST MULTI are the short-term models described in Section 2.2, and RMSE, MAE and MAPE@5 are the measures of prediction defined in Section 3.2.

The case study considered in this paper involves several stations of different modes in a major business district in the Paris metropolitan area (La Défense). We compare basic and machine learning long-term models with short-term models under two special periods, in order to understand the results of short-term multi-step forecasting. We considered different variants of short-term models including statistical, machine learning and deep learning models both in univariate and multivariate case. The results have demonstrated that deep learning models performed better forecasting during the special periods containing public holidays and special days. On the other special periods defined by days with transport network anomalies, the univariate models RF and GRU obtained the best results. Besides these results, this study has demonstrated the effectiveness of recurrent neural networks for multi-step prediction tasks. These models have shown to be less sensitive to error propagation along the different time step contrary to the other models such as vector autoregressive or random forest.

Future research should investigate different methodologies of multi-step ahead forecasting, in order to better incorporate the spatiotemporal links between the different time series. From this perspective, it could be relevant to investigate other architectures of neural network that could better capture these links within a transport network and thus make better prediction during atypical periods. Future work should also investigate the application of prediction model on other cities and new datasets that involve more information such as event database.

Acknowledgements The authors wish to thank the transport organization authority of Ile-de-France (Ile-de-France Mobilités) for graciously providing the smart card data.

References

1. Agard, B., Morency, C., Trépanier, M.: Mining public transport user behaviour from smart card data. In: The 12th IFAC Symposium on Information Control Problems in Manufacturing (INCOM), pp. 17–19 (2006)
2. Bagchi, M., White, P.R.: What role for smart card data from bus systems. In: Proceedings of the Institution of Civil Engineers. Municipal Engineer. March Issue ME1, vol. 157, pp. 39–46 (2004)
3. Bagchi, M., White, P.R.: The potential of public transport smart card data. Transport Policy **12(5)**, 464–474 (2005)
4. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
5. Cheng, X., Zhang, R., Zhou, J., Xu, W.: Deeptransport: Learning spatial-temporal dependency for traffic condition forecasting. arXiv preprint arXiv:1709.09585 (2017)
6. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
7. Cui, C., Jia, H., Huang, L., Zhang, X.: Fuzzy multivariate narx model for passenger entrance flow prediction in the shanghai subway system. Journal of Intelligent & Fuzzy Systems **31**(6), 3047–3054 (2016)
8. Ding, C., Wang, D., Ma, X., Li, H.: Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. Sustainability **8**(11), 1100 (2016)
9. Kieu, L.M., Bhaskar, A., Chung, E.: Passenger segmentation using smart card data. IEEE Transactions on Intelligent Transportation Systems **16**(3), 1537–1548 (2015). DOI 10.1109/TITS.2014.2368998
10. Lee, S.G., Hickman, M.: Trip purpose inference using automated fare collection data. Public Transport **6**(1), 1–20 (2013)
11. Li, Y., Wang, X., Sun, S., Ma, X., Lu, G.: Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. Transportation Research Part C: Emerging Technologies **77**, 306–328 (2017)
12. Lütkepohl, H.: Vector autoregressive models. In: International Encyclopedia of Statistical Science, pp. 1645–1647. Springer (2011)
13. Ma, X.l., Wu, Y.J., Wang, Y.h., Chen, F., Liu, J.f.: Mining smart card data for transit riders' travel patterns. Transportation Research Part C: Emerging Technologies **36**(0), 1 – 12 (2013)
14. Mahrsi, M.K.E., Côme, E., Oukhellou, L., Verleysen, M.: Clustering smart card data for urban mobility analysis. IEEE Transactions on Intelligent Transportation Systems **18**(3), 712–728 (2017). DOI 10.1109/TITS.2016.2600515
15. Munizaga, M., Devillaine, F., Navarrete, C., Silva, D.: Validating travel behavior estimated from smartcard data. Transportation Research Part C: Emerging Technologies **44**, 70 – 79 (2014)
16. Munizaga, M.A., Palma, C.: Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smart card data from santiago, chile. Transportation Research Part C: Emerging Technologies **24**, 9–18 (2012)
17. Ni, M., He, Q., Gao, J.: Forecasting the subway passenger flow under event occurrences with social media. IEEE Transactions on Intelligent Transportation Systems **18**(6), 1623–1632 (2017)
18. Park, J.Y., Kim, D.J., Lim, Y.: Use of smart card data to define public transit use in seoul, south korea. Transportation Research Record: Journal of the Transportation Research

Board **2063**, 3–9 (2008)
19. Pelletier, M.P., Trépanier, M., Morency, C.: Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies **19**(4), 557–568 (2011)
20. Pereira, F.C., Rodrigues, F., Ben-Akiva, M.: Using data from the web to predict public transport arrivals under special events scenarios. Journal of Intelligent Transportation Systems **19**(3), 273–288 (2015)
21. Poussevin, M., Baskiotis, N., Guigue, V., Gallinari, P.: Mining ticketing logs for usage characterization with nonnegative matrix factorization. In: SenseML 2014 – ECML Workshop. Nancy, France (2014)
22. Roos, J., Bonnevay, S., Gavin, G.: Short-term urban rail passenger flow forecasting: A dynamic bayesian network approach. In: Machine Learning and Applications (ICMLA), 2016 15th IEEE International Conference on, pp. 1034–1039. IEEE (2016)
23. Toqué, F., Khouadjia, M., Côme, E., Trepanier, M., Oukhellou, L.: Short & long term forecasting of multimodal transport passenger flows with machine learning methods. In: Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on, pp. 560–566. IEEE (2017)
24. Utsunomiya, M., Attanucci, J., Wilson, N.: Potential uses of transit smart card registration and transaction data to improve transit planning. Transportation Research Record **1971**, 119–126 (2006)
25. Wu, Y., Tan, H.: Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint arXiv:1612.01022 (2016)
26. Yu, B., Yin, H., Zhu, Z.: Spatio-temporal graph convolutional neural network: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017)
27. Zhang, J., Zheng, Y., Qi, D.: Deep spatio-temporal residual networks for citywide crowd flows prediction. In: AAAI, pp. 1655–1661 (2017)
28. Zhao, J., Tian, C., Zhang, F., Xu, C., Feng, S.: Understanding temporal and spatial travel patterns of individual passengers by mining smart card data. In: Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on, pp. 2991–2997. IEEE (2014)
29. Ziat, A., Delasalles, E., Denoyer, L., Gallinari, P.: Spatio-temporal neural networks for space-time series forecasting and relations discovery. In: 2017 IEEE International Conference on Data Mining (ICDM), pp. 705–714 (2017). DOI 10.1109/ICDM.2017.80

A Appendix

Table 6 Informations about the studied stations

ID	Name	Line	Type	Latitude	Longitude	Agency
393	LA DEFENSE-GRANDE ARCHE	A	RER	48.892187	2.237018	RATP
394	LA DEFENSE-GRANDE ARCHE	L	Transilien	48.892187	2.237018	SNCF
269	ESPLANADE DE LA DEFENSE	1	Metro	48.888631	2.247932	RATP
151	CHARLES DE GAULLE ETOILE	A	RER	48.874238	2.294491	RATP
577	NANTERRE-PREFECTURE	A	RER	48.895745	2.223213	RATP
414	LA DEFENSE-GRANDE ARCHE	1	Metro	48.892187	2.237018	RATP
578	NANTERRE-UNIVERSITE	A	RER	48.901550	2.215232	RATP
357	HOUILLES-CARRIERES-SUR-SEINE	A	RER	48.920379	2.185353	SNCF
580	NANTERRE-VILLE	A	RER	48.895126	2.195364	RATP
100	BOULOGNE-PONT DE SAINT CLOUD	10	Metro	48.840745	2.228537	RATP
812	SAINT-CLOUD	L	Transilien	48.846103	2.217621	SNCF
843	SURESNES-MONT VALERIEN	L	Transilien	48.871714	2.221030	SNCF
29	ASNIERES	J	Transilien	48.905706	2.283629	SNCF
436	LE VAL-D’OR	L	Transilien	48.856420	2.216572	SNCF
217	COURBEVOIE	L	Transilien	48.898119	2.247943	SNCF
579	NANTERRE-UNIVERSITE	L	Transilien	48.901746	2.215111	SNCF
712	PUTEAUX	L	Transilien	48.883384	2.233692	SNCF
413	LA DEFENSE-GRANDE ARCHE	T2	Tramway	48.892123	2.237214	RATP
1085	PONT DE BEZONS	T2	Tramway	48.923268	2.217590	RATP
844	SURESNES-LONGCHAMP	T2	Tramway	48.868239	2.221411	RATP
1090	VICTOR BASCH	T2	Tramway	48.914048	2.229572	RATP
1088	CHARLEBOURG	T2	Tramway	48.907665	2.238309	RATP
623	PARC DE SAINT-CLOUD	T2	Tramway	48.843124	2.221851	RATP
713	PUTEAUX	T2	Tramway	48.883335	2.233871	RATP
1086	FAUBOURG DE L ARCHE	T2	Tramway	48.896664	2.240079	RATP
1087	LES FAUVELLES	T2	Tramway	48.902768	2.239542	RATP
1089	JACQUELINE AURIOL	T2	Tramway	48.910803	2.233989	RATP
457	LES COTEAUX	T2	Tramway	48.857471	2.220604	RATP
1091	PARC PIERRE LAGRAVERE	T2	Tramway	48.917515	2.224812	RATP
461	LES MILONS	T2	Tramway	48.849850	2.221233	RATP