Transportation Research Part C 120 (2020) 102817

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

On demand forecasting of demand-responsive paratransit services with prior reservations

Ektoras Chandakas

Laboratoire Ville Mobilité Transport, Université Paris Est, 6 et 8 avenue Blaise Pascal, Cité Descartes, 77455 Marne-la-Vallée Cedex, Paris, France
Transamo, Transdev Group, 12 Rouget de l'Isle Street, 92789 Issy-les-Moulineaux Cedex, Paris, France

ARTICLE INFO

Keywords:
Paratransit services
Short-term forecasting
Demand-responsive transit
Stastistical modelling
Deep-learning

ABSTRACT

The present study investigates the determinants of the volatility of passenger demand for paratransit services and explores the feasibility of a data-driven model for medium-term forecast of the daily demand. Medium-term demand forecasting is a significant insight to optimise resource allocation (staff and vehicles) and reduce operations costs. Using operational data from the reservation platform of the paratransit services in Toulouse, France, and enriching them with exogenous information, the study derives statistical and deep learning models for medium-term forecast. These models include a seasonal ARIMAX model with rolling forecast, a Random Forest Regressor, a LSTM neural network with exogenous information and a CNN neural network with independent variables. The seasonal ARIMAX model yields the best performance, suggesting that when linear relationships are considered, econometric models and deep learning models do not have significant differences in their performance. All the models show limited ability to grasp unique events with multi-day impacts such as strikes. Albeit a highly volatile demand and limited knowledge ahead of the forecast, these models suggest the volume of early reservations is a good proxy for the daily demand.

1. Introduction

The United Nations (UN, 2018) project that by 2050 the world population living in urban areas will grow to 68% compared to 2018 where the urban population of 4,2 billion accounts to 55% of world population. This increase is believed to increase urban passenger transport by 60–70% by 2050 (OECD-ITF, 2018), while the car remains the main mode of transport. Various initiatives exist to develop mass transit systems, mainly in dense areas where high passenger flows can justify the investment and operation expenditures.

To provide an adequate alternative to private cars it is essential to complement the high performance transportation network with services which are better adapted to these particular cases: in low-density areas, during certain periods of the day and week (night hours, weekend, …) and for users with specific needs (seniors, reduced mobility, …). That can be addressed with demand-responsive transportation (DRT) services, where small- and medium- sized vehicles adapt the services to the specific demand (Alonzo-Gonzalez et al., 2018). The sanitary context in 2020 urged various transit networks replace the fixed-route services with demand-responsive ones. Changing mobility patterns and caution with high-occupancy vehicles lays ground for the emergence of Mobility as a Service and demand-responsive services.

In addition, current technology trends new information technologies and app-based reservation platforms increase the service reactivity with real-time reservation and vehicle location tracking. Attard et al. (2020) suggest the three-part IT architecture for the

E-mail address: ektoras.chandakas@transamo.com.

https://doi.org/10.1016/j.trc.2020.102817
Received 1 June 2020; Received in revised form 18 August 2020; Accepted 22 September 2020
Available online 8 October 2020

deployment of efficient demand responsive transit. These technologies, combined with Mobility as a Service (MaaS), drive the reinvention of transit services as a whole (Shaheen and Cohen, 2020). The combination of public policy towards shared mobility and technological development makes DRT service ideal to fill the gap of First-Mile-Last-Mile (FMLM) and paratransit services, providing an almost seamless inclusive urban transportation system. Coutinho et al. (2020) compare fixed routes and DRT services in a real case study to show that the latter offer a better operational performance as of vehicular kilometers, operational cost and greenhouse gas (GHG) emissions per passenger.

Demand Responsive Transportation services are subject by definition to a volatile demand, more so if they are focused on non-constraint journeys, such as the ones motivated by leisure activities or purchases. Rahimi et al. (2018) establish a link between the cost efficiency of large-scale DRT operations and the demand, the service area and the demand density. Palmer et al. (2008) suggest a positive impact on productivity when using automatic grouping of the requests. A DRT operator, in regards to the Level of Service Agreement with its clients (B2C) or public authorities (B2G), needs to anticipate the size of the operational fleet and the number of the employees needed to meet the demand. Overestimating them, leads to slack resources and increased production costs. An underestimation of the daily demand may lead to non-captured revenues, a loss of the customer image or specific penalties from the transportation authorities. Having a better knowledge on travel demand behavior of the potential customers and improving the forecast tools of passenger demand allows a near to real-time optimization of the vehicle fleet, along with the vehicle maintenance planning and the drivel and personnel scheduling.

The paper seeks to shed some light to the main variables affecting demand variability. In addition, it seeks to examine the feasibility of an operational model able to provide, via endogenous information, reservation trends and exogenous information (holidays, weather, etc.), an adequate forecasting of the number and type of reservations 7 days in advance or prior, when crew scheduling is done according to the transit network collective agreement. The direct and indirect impacts of an improved forecast are translated operationally to:

- A reduction of production costs through a reduction of the number of idle drivers
- An increase of commercial revenue, by reducing the refusal rate for late reservations
- A guarantee of a good image to the local transport authorities, therefore better chances to maintain the contract

The present paper is structured in 6 parts as follows: After a brief introduction, Section 2 reviews existing research on short-term forecast in traffic flow and passenger transportation networks. Section 3 describes the site and the demand data used for the modelling, as well as the pre-processing and feature engineering processes. Section 4 formalizes the modelling architecture, while Section 5 explores the specification and validation of the set of predictive models used for the short-term forecast. In conclusion, Section 6 interprets and compares the main outcomes of the models and sets out some areas of future research.

2. Background on short term forecasting models

The short-term forecasting in transportation is a vivid research area, especially in recent years. The main topics address flow and travel time forecast in road traffic and public transportation context. In Vlahogianni et al. (2016) the authors do an extensive review of neural network approaches in road traffic modelling and they set the main challenges in that area. The authors suggest the model interpretability, or lack of thereof, is one of the barriers for adopting more sophisticated deep learning models in traffic operations.

Vlahogianni and Karlaftis (2013) lay out a thorough comparison between statistical and neural networks in transportation, focusing mainly in road traffic. Karlaftis and Vlahogianni (2009) examine long-term processes for ARIMA models captured through fractional integration. Their study points out that over-differentiation tends to over-inflate the MA component of the model, thus weakening the quality of the results. Novel approaches in short-term forecasting of traffic flow speeds include Polson and Sokolov (2017) who develop a deep learning approach based on a regularization process to reduce the non-linearity of the traffic state transitions and Wang et al. (2014) who propose Bayesian combination approach factoring the prediction errors in order to improve the prediction. In the public transportation setting, Farid et al. (2016), among multiple data-driven models, suggest the Support Vector Regression (SVR) model is more accurate in the short-term forecast of bus travel times. Lin et al. (2018a) introduces a Particle Swarm Optimization and Extreme Learning Machine (PSO-ELM) neural network to quantify the uncertainty in short-term traffic prediction by examining the reliability and the sharpness of the Prediction Interval (PI).

Transit systems, TNC and micromobility operators pursue the digitalization of the production and client outreach processes, accumulating large amount of data. These are often used in the development of short term demand prediction models for mass transit services and micromobility schemes. Lin et al. (2018b) explore various Graph Convolutional Neural Networks with Data-driven Graph Filter (GCNN-DDGF) models to predict station-level hourly demand in a large-scale bike-sharing network. The graph approach of these models with the associate neural network architecture captures efficiently hidden correlations among bike-sharing stations.

The use of smart card data for demand forecasting is widely explored. Van Oort et al. (2015) propose a smartcard data-driven model for the analysis of transit service variations. Xiaoqing et al. (2018) build on smartcard tap-on and tap-off date to develop short-term metro passenger flow forecast through a combination of a probabilistic model selection and random forest classification model. In Gong et al. (2014) the authors develop a three step model based on seasonal ARIMAX model, an event-based method and a Kalman Filter based method to forecast short-term passenger flow at bus stops. Moreira-Matias and Cats (2016) propose a model through local constrained regression to infer passenger loads through automatic vehicle location data (AVL). Ma et al. (2014) specify an Interactive Multiple Model-based Pattern Hybrid approach for short-term demand forecast by developing pattern model based on historical demand data and dynamically optimize the interaction between them using real-time observations.

Noursalehi et al. (2018) propose a two-level approach for the short term prediction of passenger flows in transit stations. On the station level the authors opt for a univariate state-space model. A hierarchical clustering algorithm identifies similarities between stations and regroups them. Within each cluster, a dynamic factor model grasps the interdependencies between the stations. Toque et al. (2017) use LSTM and Random Forest models for short-term (15 and 30 min windows for rail and buses) and long-term (one-year) for demand forecasting in multimodal transportation facilities. Guo et al. (2019) use an approach based on the fusion of Support Vector Regression (SVR) and long short-term memory (LSTM) neural network to predict abnormal passenger flows in a metro station. Ke et al. (2017) propose a fusion convolutional long short-term memory architecture (FCL-Net), integrating spatio-temporary and non-spatial time-series variables for short-term forecasting of passenger demand of demand-responsive services.

Short-term prediction of demand-responsive transit schemes is a topic most frequently overlooked. However, the digitalization of the operations is a major driver for the creation of new categories of models in a real-time setting. Koffman and Lewis (1997) make the argument of the importance of adequately forecasting the demand for the optimal design of paratransit services. Benjamin et al. (1998) explore the determinants of choice of paratransit services and develop an econometric model to address mode choice. Furthermore, Deka and Gonzales (2014) provide a spatial structure of paratransit trips, by identifying potential trip generators.

The present study builds on previous scientific contributions both in short-term predictive models and in demand-responsive paratransit services. It expands the scope of the existing models, and provides a comparison of existing and novel models for short-term demand forecast. These contributions improve the understanding of the determinants of the demand variability and contribute on the understanding of theoretical and practical implications of machine learning and deep learning models on demand-responsive transit operations.

3. Site and data description

The paper is focused in the paratransit services (Tisseo, Mobibus) in Toulouse, France. This service is operated through a Public Service Obligation Contract. The company has 58 vehicles of various sizes (from 3 to 12 passengers) and 75 employers (of which 60 drivers). It is operating during weekdays from 7:00 am to 0:30 am and weekends and bank holidays from 9:00 am to 0:30 am. The reservations open 28 days prior (D-28) and can be made at the latest 2 h (H-2) before the departure. No reservation is accepted if the period between the reservation and the trip is shorter than 2 h. Furthermore, once a reservation is made (from D-28 to H-2), the client can modify or cancel it for free up to 2 h (H-2) before the departure. These reservations are subject to grouping, in order to optimize the services and reduce production cost. The drivers' collective agreement stipulates the drivers' shifts are determined 7 days in advance (D-7). Further tuning of drivers schedules up to D-1 is based on negotiation and mutual agreement, but usually incur high marginal costs.

3.1. The initial database

The data acquired correspond to the data of reservation platform of the paratransit operations. This relational database stocks all data relevant to reservations management, staff and vehicle scheduling and service optimisation. The initial operational database may contain personal data subject to the General Data Protection Regulation (GDPR). However, the data extracted are anonymous. Further

Table 1
Data structure.

Column	Type	Description
REFCON	Int	*Reservation Unique ID*
JOUR (*DAY*)	Datetime	Trip date
IDCLIENT	Int	Client Id
IDSERVICE	Int	Id of the service
IDMAITRE (*IDPRIMARY*)	Int	Id of primary client (in case of grouped trips)
ADRDEB_ID (*ADRDEP_ID*)	Int	Id of the departure address (closest node of the route graph)
ADRDEB_CP (*ADRDEP_ZIP*)	Int	ZIP Code of departure.
ADRDEB_VILLE (*ADRDEP_CITY*)	Char	Departure borough name
ADRFIN_ID (*ADRARR_ID*)	Int	Id of the arrival address (closest node of the route graph)
ADRFIN_CP (*ADRARR_ZIP*)	Int	ZIP Code of arrival. Float transformed to Integer
ADRFIN_VILLE (*ADRARR_CITY*)	Char	Arrival borough name
HDEB_PLANIFIEE (*TOD_PLANNED*)	Datetime	Planned time of departure
HFIN_PLANIFIEE (*TOA_PLANNED*)	Datetime	Planned time of arrival
DATE_CREATION (*DATE_INIREQUEST*)	Datetime	Date the initial reservation was made
DATE_MODIFICATION	Datetime	Date of last modification of the request
IDAUTORISATION	Int	Tariff Id
IDCATEGORIE (*IDCATEGORY*)	Int	Trip purpose
KMS_MISSION_DIRECT	Float	Distance of a direct trip
KMS_MISSION_REEL (*KMS_MISSION_REAL*)	Float	Distance of planned trip (with grouping)
DUREE_MISSION_DIRECT (*DUR_MISSION_DIRECT*)	Float	Direct trip length in minutes
DUREE_MISSION_REEL (*DUR_MISSION_REAL*)	Float	Planned trip length in minutes
COMPLETED	Boolean	The reservation is completed
MODE	Char	Service category

analysis and treatment is grouped in such ways as to not be able to determine any personal data of the clients.

The relevant demand data are further extracted into a plain table, where each observation corresponds to a reservation, completed or cancelled. The data concern a period of roughly 4 years (1478 days from 25/05/2015 to 10/06/2019) and total 736 019 observations. Further transformation of the database includes the identification of the service category (mode), through the tariff rate. The structure of the reservation data is illustrated in Table 1.

3.2. The spatial distribution of the data

The density of the departures and arrivals per ZIP code is further explored. Furthermore, 22% of the trips are made within the same ZIP code. This ratio approaches 49% if the municipality boundary is considered as a level of aggregation. That illustrated the short distance of the trips which makes the trip aggregation harder. Fig. 1 illustrates the ratio of departure and arrival addresses per ZIP code. That highlights the volumes of the Toulouse municipality in the center, as it corresponds to 61% of departing trips. Furthermore, the data show some symmetry between departures and arrivals for each ZIP code area.

3.3. The temporality of the data

Due to the inherent volatility of the demand, the data are explored with respect to their temporal structure. The weekly demand, as illustrated in Fig. 2(a), suggests a positive linear trend of the demand, as well as some seasonality, accounting for Christmas, summer holidays and spring holidays in May, along with a significant effect of bank and school holidays. The day-to-day variation of the demand (Fig. 2(b)) during a 4 month period shows to a greater extent the variability during weekday and weekend. Since the initial data at the moment of the extraction included future dates of trips, the analysis is limited to the period until 30/04/2019.

Further analysis seeks to identify whether this profile is sensible to the type of service and type of day (weekday, weekend, …). The data analysis focuses on within-day dynamics of reservation. Fig. 3 illustrates reservation volumes per 15 min time-step per day of the week. It suggests significant difference with respect to the day of the week and an important variability per type of day throughout the years. Nevertheless, some within-day patterns of the demand per type of day are visible.

3.4. The qualitative attributes of the data

Addressing the qualitative characteristics of the demand, the acceptance ratio is roughly 71%, while a slight variation of this ratio is observed throughout the year and within the day, as illustrated in Fig. 4.

The mode variable accounts for the service type demanded are illustrated below. The “*Le Lien*” service has a very low number of observations, making their characteristics non-representative in statistical sense at that level of aggregation (see Table 2).

The journey duration with 15 min time steps is further described, with respect to the volume of departures. Travel times have an average value of 00:27:24 and standard deviation of 00:15:48. The travel time per reservation seems to be sensible to morning and evening peak hours, since vehicles are subject to general traffic. Nevertheless, the travel times do not seem to be impacted by the number of reservations. The average travel times seem to be lower during the noon off-peak period compared to the evening off-peak one. Differences in travel times may be attributed to longer route distances. Nonetheless, travel time is an appropriate proxy of service production per reservation, since the size of the service is marginal and it does not impact general car traffic.

The average travel time per commercial trip (reservation travel time) per service type and day of the week are illustrated in Table 3. The D2D service has a more important travel time than A2A type. The 4 min difference can be accounted to the additional time needed for accompanying the client from the door of the origin and destination to the vehicle. Furthermore, the commercial trip travel time during the weekend are lower than those from weekdays. Characteristics of the “Le Lien” service should be considered with extreme caution due to their limited sample.

The trips cover a variety of purposes. The main trip purposes for the use of the service, of the users that informed on the relevant trip

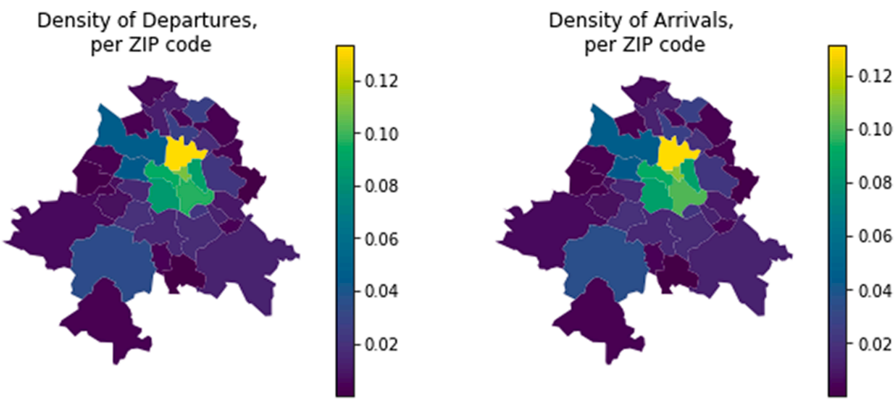

Fig. 1. Density of departures and arrivals per ZIP code.

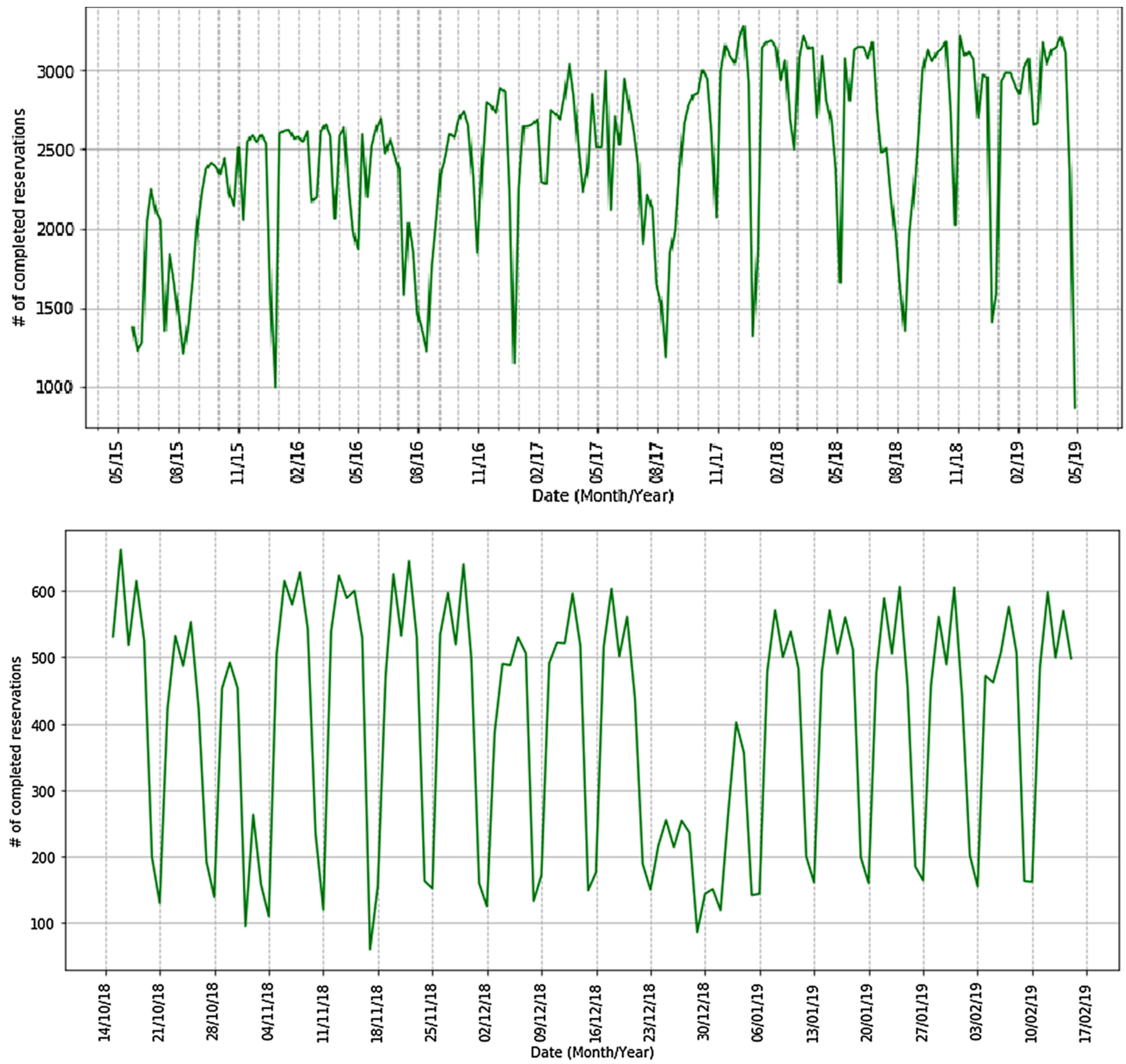

Fig. 2. Volume of (a) weekly completed reservations during the simulation period (top) and (b) daily completed reservations during October-2018 – February 2019 period (bottom).

purposes when making the reservation by phone of internet, are medical visits, professional trips (to and from work accounting to 25%) and leisure (45% of trips with known purpose). As of the concentration of the reservations to some clients, the volume of reservations per unique client is analyzed. It is observed that 30% of the most frequent clients, make for 50% of all requested reservations. On the other hand 40% of the least frequent clients make for only 20% of all requested reservations. That suggests a frequent use of the service by a given set of clients, although the frequent clients are not a dominant group. Hence, that should have a straightforward impact on the volatility of the demand.

This service accepts reservations up to 28 days in advance to the trip date. In addition, it is possible to pre-plan reservations that are repetitive for certain days per week or a period. These reservations are injected into the system 21 days before. Once the reservation is created, it can be subject to modifications, cancellation or rejection up to the trip day. Fig. 5(a) illustrates the part of all the reservations, with respect to the number of days in advance they were made. Indeed more than half the reservations are made 14 days prior to departure. Roughly 2/3 of the reservations are made 7 days before the departure.

Current cancellation policy accepts free cancellations up to 2 h before the departure, adding to the volatility of the final demand. Fig. 5(b) illustrates the cancellations per day prior to the trip. Whereas the cancellations amount to 30% of total reservations, roughly 70% of them occur during the last week, while 35% occur the last two days prior to the trip.

The reservation policy in place, lacking incentives for early reservation and for avoiding late cancelations, is a major source of demand volatility and cost inefficiency. However, this must be interpreted by the desire of the transport authority for a responsive service, thus facilitating mobility and inclusion for this particular client group.

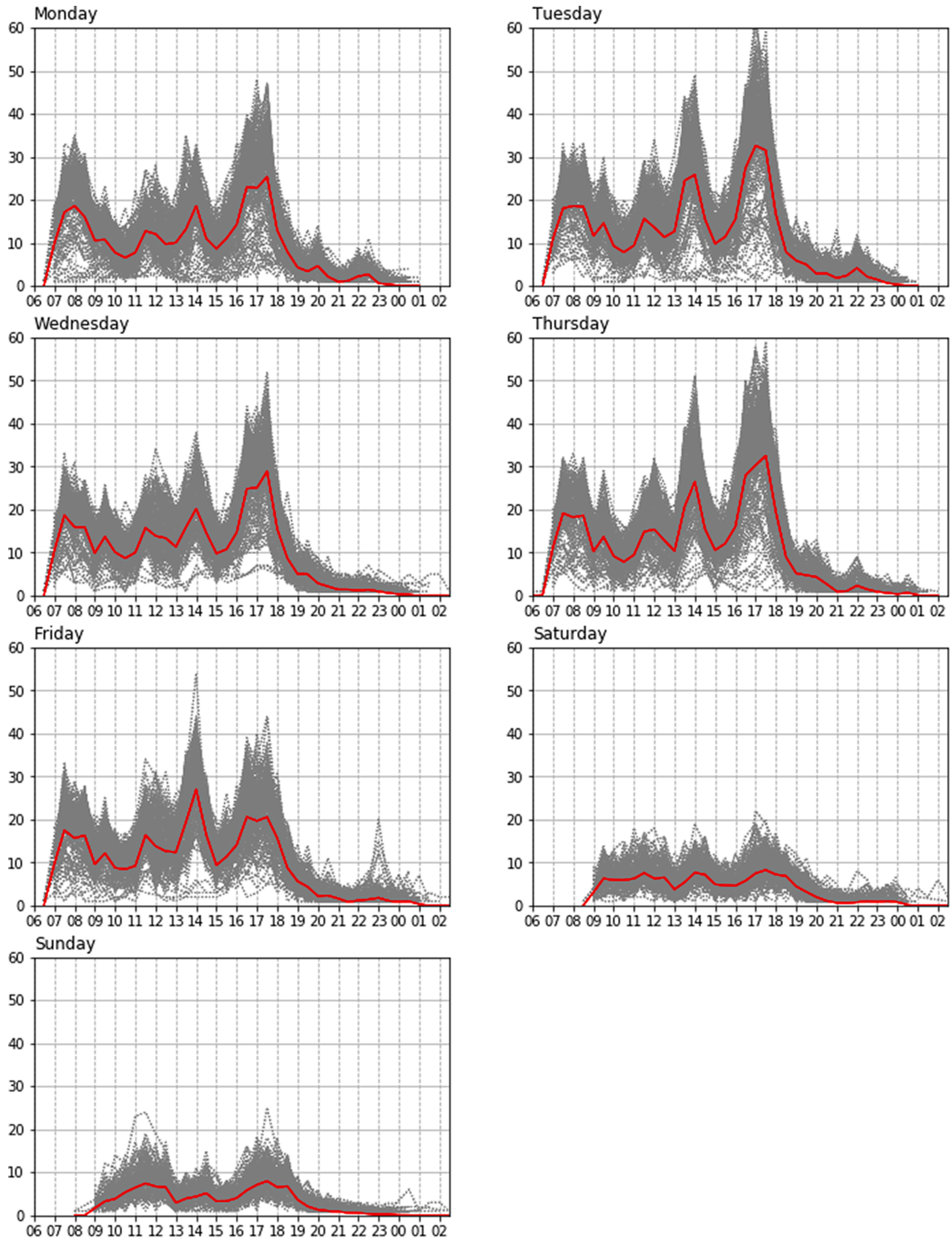

Fig. 3. Reservation volume per 15 min time-steps per day of the week for individual days (grey) and mean (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

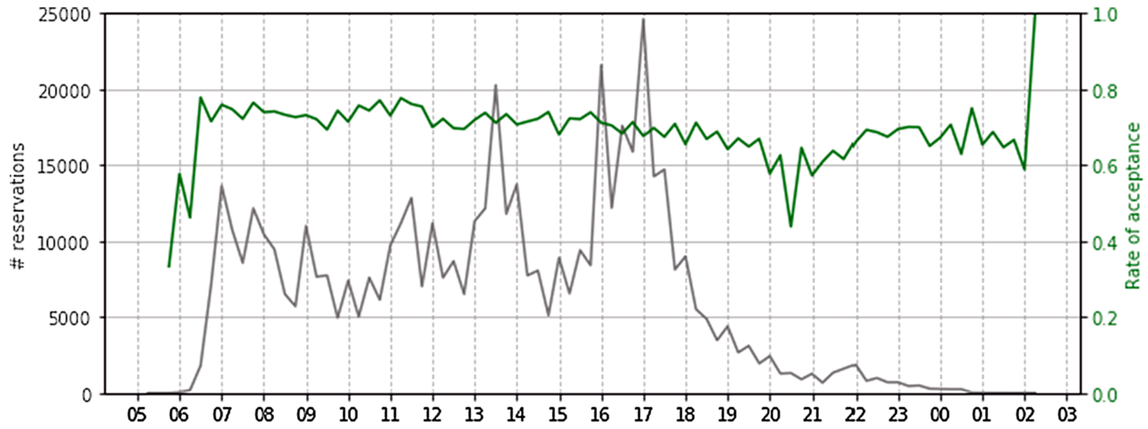

Fig. 4. Reservation acceptance rate per day.

3.5. The exogenous information

Additional features are added to the days, namely related to weather, bank and school holidays. In addition to bank and school holidays, the "*pont*", or extensive weekend, variable is determined: this dummy variable characterizes a Monday followed by a bank holiday, or a Friday following a bank holiday, where the demand is assumed to be closer to that of a bank holiday rather than of a weekday. As of the type of day, the following 5 categorical variables were added:

- TypeOfDay : categorical, corresponding to the day of the week: 1, …, 7 for Monday through Sunday, That is further transformed into three one-hot encoders:
  - o isMoWeFr
  - o isTuTh
  - o isSa
- IsWE: categorical, corresponding to weekend
- IsJF: categorical, corresponding to bank holidays
- isShoolHoliday, categorical, corresponding to school holidays
- isPont: categorical, corresponding to extended weekends

Historic hourly weather data of Toulouse are retrieved starting from 01/09/2016 from on-line weather archives. These data are transformed and the information on the mean, max and min temperature during each day and the number of hours with a rain fall is introduced as an exogenous information (see Fig. 6).

4. Modelling pipeline

The various models are compared within a standard modelling pipeline, as illustrated in Fig. 7. The following section details the data preparation process and the evaluation metrics used for the calibration and comparison of the models.

4.1. Data preparation

The final dataset is obtained from the combination of the endogenous and exogenous information for the period from 01/09/2016 to 30/04/2019. Each observation accounts to a specific date of that period. Two datasets are further specified into two distinct datasets:

- Y (972, 1) contains the time-series volumes to be predicted
- X_exogenous (972, 13) contains the exogenous data. These are described in the following table. The TypeOfDay is further transformed into a one-hot encoding (isMoWeFr, isTuTh, isSa) combined with the isWE variable (see Table 4).

In addition to the exogenous information, the input variables include the volume of reservations made 14, 10 and 7 days prior to the day of the trip, referring to the ResBef_14d, ResBef_10d and ResBef_7d respectively. These values do not take into consideration further modifications or cancellations.

The datasets (Y, X_exogenous) are separated into three distinct datasets. To ensure comparability of the results, the following datasets are used by all the algorithms:

- A training dataset, used for the training of the models

Table 2
Service characteristics.

ID	Name	Characteristics	Price	# Observations (Tot = 736 069)
A2A	Address-to-Address	Pick-up and drop-off are made at the desired address	1,10 – 1,35 whether it's peak or off-peak	239 116
D2D	Door-to-Door	Like A2A, where the driver accompanies the client up to the door	Like A2A with additional charge for accompany person	494 983
Le Lien	Metro connection	Connection to metro or BRT transit services	Free of charge	1 920

Table 3
Average travel time (in minutes) per commercial trip with respect to the service type and the day of the week.

Service Type	Monday(0)	Tuesday(1)	Wednesday(2)	Thursday(3)	Friday(4)	Saturday(5)	Sunday(6)	Mean
A2A	25.6	25.4	25.6	25.3	24.4	21.5	22.6	24.9
D2D	22.2	30.2	28.9	29.1	28.5	26.2	27.0	28.9
Le Lien	23.5	23.3	26.5	22.8	27.7	22.0	24.4	25.3
Mean	27.9	28.6	27.8	27.8	27.2	24.7	25.9	26.7

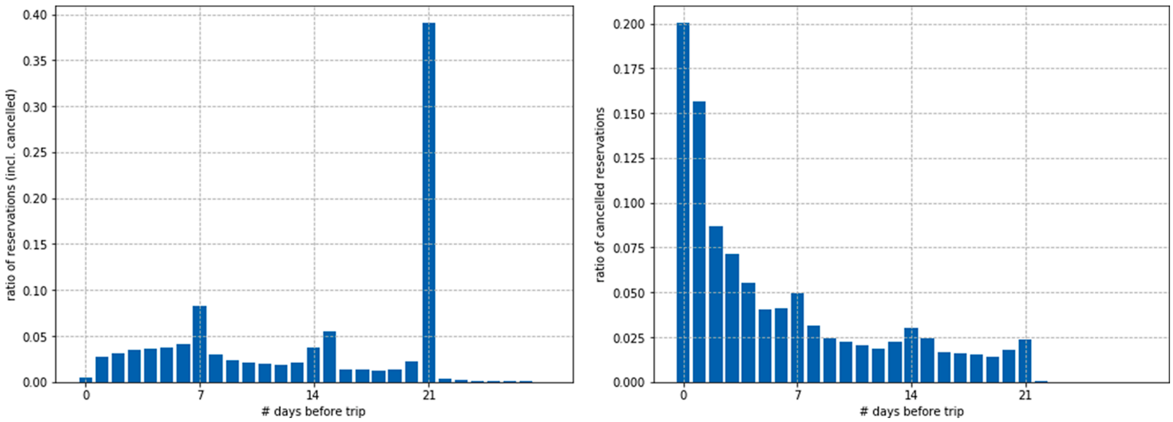

Fig. 5. (a) Volume of reservations based on the number of days reserved prior to trip – left (b) Volume of reservations cancelled with respect to the day before the trip a cancellation occurs.

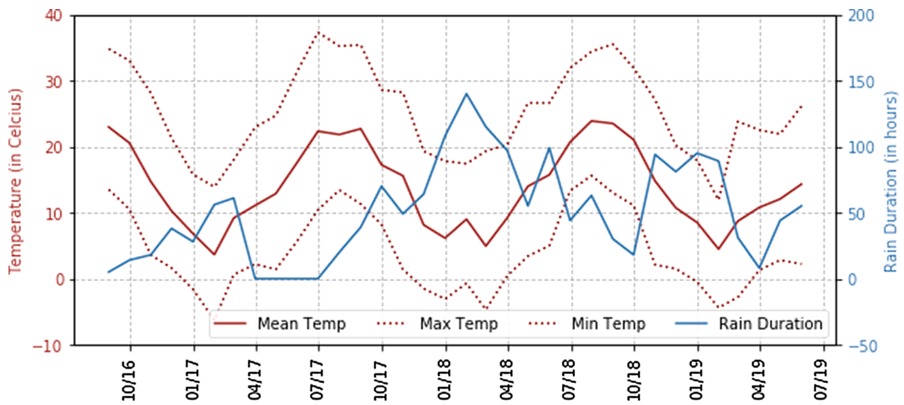

Fig. 6. Local weather data and main exogenous information.

- A validation dataset, used for the selection of the best model
- A test dataset used for the unbiased appreciation of the model precision

4.2. Evaluation metrics

The quality of the models is evaluated through it is proximity to observed data. The evaluation metrics, summarized in Table 5, allow for the comparison of the forecast accuracy. While the RMSE is the most usual metric, used when large errors are undesired, the MAE is most robust to the outliers. The MASE allows for the comparison of the forecast with the MAE produced by a naïve forecast. Finally, the MAPE metrics address the forecast error as a proportion of the actual values.

5. Modelling framework

Various families of models can be specified to deal with modelling time-series with exogenous data. The experiment platform uses Python 3.7 with scikit-learn (Pedregosa et al, 2011), statsmodel (Seabold & Perktold, 2010), tensorflow (Martín et al, 2015) and keras (Chollet, 2015) for comparing the models.

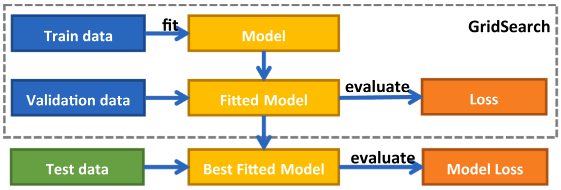

Fig. 7. The modelling pipeline used for the model specification and comparison.

Table 4
Descriptive statistics of the input variables.

Variable	Type	Descriptive statistics
ResBef_14d	Continuous	Mean = 226.2 stddev = 132.0
ResBef_10d	Continuous	Mean = 259.9 stddev = 141.1
ResBef_7d	Continuous	Mean = 299.1 stddev = 156.6
isMoWeFr	One-hot	F(1) = 0.43
isTuTh	One-hot	F(1) = 0.29
isSa	One-hot	F(1) = 0.14
isWE	One-hot	F(1) = 0.28
isJF&Pont	One-hot	F(1) = 0.03
isSchoolHoliday	One-hot	F(1) = 0.31
MeanTemp	Continuous	Mean = 13.6 stddev = 6.8
MaxTemp	Continuous	Mean = 18.3 stddev = 7.7
MinTemp	Continuous	Mean = 9.5 stddev = 6.4
RainDuration	Continuous	Mean = 1.7 stddev = 3.2

The present study explores different model families, related to forecasting univariate time-series of as a standard regression model. The exploration phase of the present study included a multitude of relevant machine learning and deep learning models. For simplicity, it was chosen to pursue with a limited number of representative models from each category, with the selection criteria covering both practical and theoretic aspects. In addition it provides a naïve model to serve as a baseline forecast. The following table resumes the main families examined.

	Univariate time series	Standard regression
Machine learning	Seasonal ARIMA model with exogenous information (SARIMAX)	Random Forest Regression (RFR)
Deep learning	Long Short Term Memory neural network (LSTM)	Convolutional Neural Network (CNN)

5.1. *A naïve model for the baseline forecast*

A naïve forecast model is built to serves as a baseline estimate. The naïve model assumes for each day the forecast of the reservations corresponds to the average values of the training set values, with respect to the type of day within the week. The evaluation metrics suggest that the forecast cannot grasp the difference on the level of reservations due to the within-year seasonality and the unique events.

5.2. *An autoregressive model to address a univariate time series*

Addressing the dataset as a univariate time series with exogenous data, an Autoregressive Integrated Moving Average (ARIMA)

Table 5
Description of the evaluation metrics.

Evaluation metrics	Loss function
Mean Absolut Error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\|y_i - \widehat{y}_i\right\|$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}$
Mean Average Scaled Error (MASE)	$MASE = \frac{MAE_{actualforecast}}{MAE_{naiveforecast}}$
Mean Average Percentage Error (MAPE)	$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left\|y_i - \widehat{y}_i\right\|}{y_i}.100\%$

model (Box and Pierce, 1970) is used. It is a model that is adapted to short-term real-time forecasting provided addressing issues such as complexity of model specification, estimation and maintenance (Williams, 2001). The analysis of the data and their seasonal decomposition suggest seasonality based on the period of the year and the day of the week (Durbin and Koopman, 2012). In addition exogenous data relevant to the type of day and the temperature are added to a seasonal ARIMA model.

As long as the time-series data are considered, the Dickey-Fuller Test suggests the time series is stationary at a 1%. The autocorrelation function suggests a strong autocorrelation for the autoregressive part for a small lag and some strong seasonality every 7 time-steps. The partial autocorrelation function shows a lag greater than a week.

The seasonal ARIMAX (SARIMAX) is characterized by non-seasonal (p, d, q) and seasonal (P, D, Q, s) hyper-parameters as well as the use of the exogenous data, The non-seasonal hyper-parameters are:

- The autoregression (p) denoting the past values included in the regression
- The differencing (d) defining the differencing transformations
- The moving average (q) denoting the lag of the error

The trend parameter can take the following values: none, constant, linear, linear & constant. The seasonal parameters (P, D, Q) represent the same hyper-parameters for the seasonal part of the time series. Finally, s corresponds to the number of timesteps in the season.

The calibration of the SARIMAX model needs the exploration of the space around 8 hyperparameters, with a walk forward validation, more adapted to a time-series context. The final dataset used for the gridsearch is limited to the 292 days, of which the forecast is limited to 73 days. For each time-step a forecast of 7 days ahead is made, to account for the operating lag between schedule optimization and operations.

An exhaustive Gridsearch of the hyper-parameters suggestions the parameter combination [(1, 0, 1), (1, 0, 0, 7), linear trend] yields the optimal score. The first results indicate that for the best fitted model the variables relative to the weather conditions are not significant. The non-significant variables are excluded from the final model. Hence the results of the final SARIMAX are summarized in Table 6.

These results suggest that weather conditions are not significant to the prediction of the volume of reservations. That may be attributed to the presence of constrained trip purposes (work, medical visits …). Furthermore, the volume of reservations 10 (Res_Bef10d) and 14 (Res_Bef14d) days before is not significant. It shows that the presence of the variable Res_Bef7d, for the volume of reservation 7 days prior to the trip, is probably a sufficient proxy for this evaluation. Finally, the non-significance of the isHolidays variable, integrating school holidays periods, may be attributed to the characteristics of the service's clients, with a large proportion of seniors, and the presence of variables as the volume of reservations 7 and 14 days.

The exploration of the residuals, plotted in Fig. 8, shows non-auto-correlated and normally distributed residuals. The plot of the standardized residual seems to take random values around 0. However, it may be showing some cyclical structure of the residuals. The forecast errors seem to be normally distributed around a zero mean. The Q-Q plot with residual quantiles vs the theoretical normal quantiles shows a slight tail, Nevertheless, it is deemed satisfactory. Finally the autocorrelation of the residual error does not show any significant correlation between the residuals.

The test process of the final model is made through an adaptation of the previous algorithm and the use of a different set of data. The test set covers an entire year prior to the data used for the selection of the best model. The model forecasts 73 days, corresponding to

Table 6
Summary of the results of the final SARIMAX model.

Dep. Variable	y		No. Observations	292
Model:	SARIMAX(1, 0, 1)x(1, 0, 0, 7)		Log Likelihood	−1224.286
			AIC	2468.572
Sample:	0–292		BIC	2505.340
Covariance Type:	opg		HQIC	2483.300

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	−0.0113	0.023	−0.499	0.617	−0.056	0.033
Res_Bef7d	1.0854	0.015	73.326	0.000	1.056	1.114
isWE	53.4241	4.307	12.404	0.000	44.982	61.866
isJF&Pont	−19.2174	5.233	−3.673	0.000	−29.473	−8.962
isMoWeFr	63.3974	6.812	9.307	0.000	50.046	76.749
isTuTh	59.0542	7.663	7.707	0.000	44.035	74.073
ar.L1	0.8252	0.348	2.373	0.018	0.144	1.507
ma.L1	−0.7859	0.379	−2.072	0.038	−1.529	−0.042
ar.S.L7	0.2894	0.065	4.464	0.000	0.162	0.416
sigma2	252.3721	22.014	11.464	0.000	209.225	295.520

Ljung-Box (Q):	99.44	Jarque-Bera (JB):	0.23
Prob(Q):	0.00	Prob(JB):	0.89
Heteroskedasticity (H):	0.85	Skew:	−0.03
Prob(H) (two-sided):	0.43	Kurtosis:	2.88

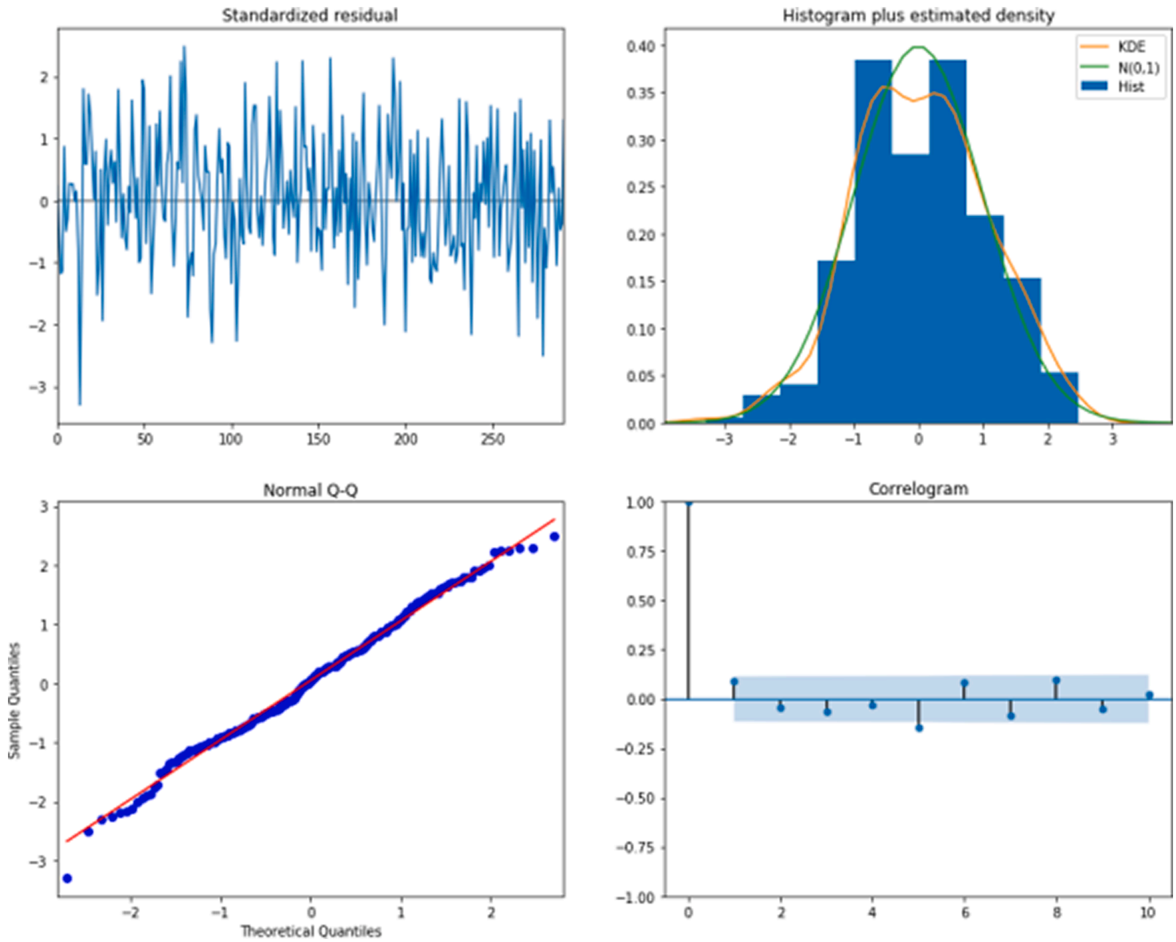

Fig. 8. Final SARIMAX model residual evaluation.

20% of a year. A rolling forecast of the 7th day ahead for each timestep is made on the basis of the previous reservations and the exogenous variables.

The analysis of the residuals of the fitted model on the test dataset allows to show non-correlated and fairly normal distribution of the residuals (see Fig. 9). The Q-Q plot with the residual vs the theoretical quartiles shows a slight trail. However, the autocorrelation of the residual error seems to show some correlation with 7 day lag of the residuals. That could be attributed to the weekly periodicity of the initial demand.

5.3. A Random Forest regressor

A Random Forest regressor (Ho, 1995) is designed to address the demand forecast as a regression problem addressing in an independent way each daily prediction. In such case, the input variables are limited to the exogenous information of the dataset. The regression model is based on the scikit-learn algorithm (Pedregosa et al, 2011). An exhaustive Gridsearch algorithm explores the main parameters of the Random Forest Regression: number of decision trees and max depth. The training step suggests 1000 decision trees with up to 5 splits.

The analysis of the residuals of the fitted model of the test dataset shows non-auto-correlated and normally distributed residuals, corresponding to a white noise. The plot of residuals vs fitted values suggests the residuals take random values (see Fig. 10).

5.4. A long short term memory neural network model, adapted to time-series

A Long Short-Term Memory neural network (Hochreiter and Schmidhuber, 1997) is specified for the regression problem of a time-series with exogenous information. This architecture includes a pre-processing phase necessary for the adaptation to the data structure of the LSTM model. The data are transformed into 3D ndarrays with the following dimensions [nb_of_samples, look_back timesteps, nb of features], as well as the output data which are transformed into 3D ndarrays with the following dimensions [nb_of_samples, look_back values] (Fig. 11). The target data, as well as input data relevant to reservations, are scaled to a (0,1) range by a MinMax

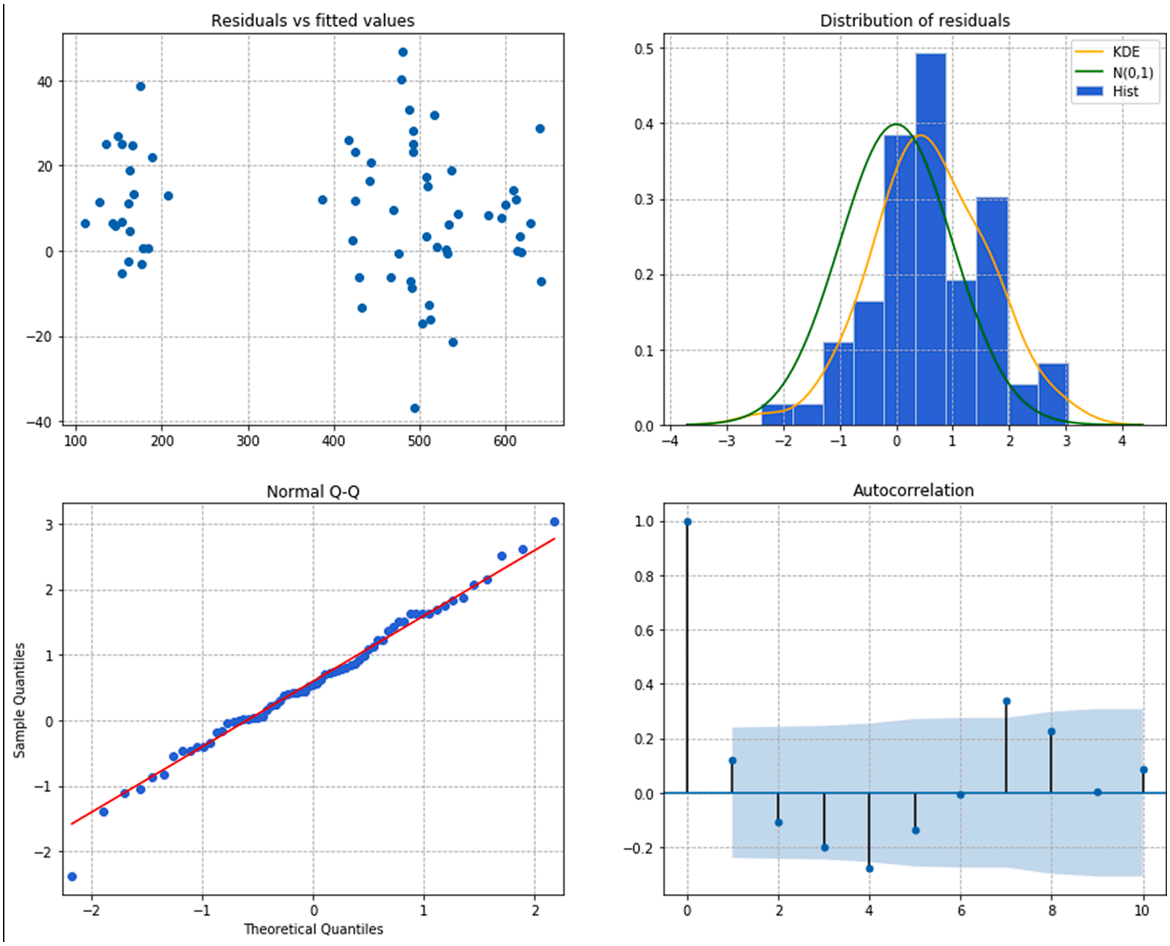

Fig. 9. Evaluation of the residuals of the test dataset of the SARIMAX model.

scaler.

The design of the model follows a standard architecture. In addition to the LSTM layer, dense layers with a time distributed wrapper are considered. Table 7 resumes the design of the model with 500 epochs.

A rolling forecast is used to predict the reservation values. For each step throughout the validation and test dataset, the historic data of current forecast are added to the input data and used for the forecast of the following time-step.

The analysis of the residuals on the test datasets suggests non-auto correlated and normally distributed residuals. The residuals vs fitted values plot suggests constant variance and the presence of some outliers. The Q-Q plot shows a slight tail on both extremities. Overall, it shouldn't impact the quality of the model (see Fig. 12).

5.5. A convolutional neural network model

A convolutional neural network (Krizhevsky et al., 2012) is designed for the forecast of reservation values on the basis of a particular day's features. The particularity of this model is that it is not modelled as a time-series; rather each observation (day) is considered independently from the others. The CNN model is built with:

- An input layer: a 1D convolutional layer
- Four Hidden layers, a 1D convolutional layer, followed by 1D average pooling and flattening, and 3 dense layers,
- An output dense layer with one unit

In addition a dropout ratio of 0.25 is chosen between the dense layers to avoid overfitting during the training phase. The activation layers correspond to relu, sigmoid and linear functions. The CNN model uses the Adam optimization algorithm (Kingma and Ba, 2015). A Gridsearch suggests the use of 500 epochs for the final CNN model. The characteristics of the layers are as follows (see Table 8):

The input data is reshaped with respect to the 2D ndarary [sample, nb of features] needed and a MinMax scaler is used to scale target date to a (0,1) range. The futures included correspond to the input variables used in all the models. With respect to the insight of

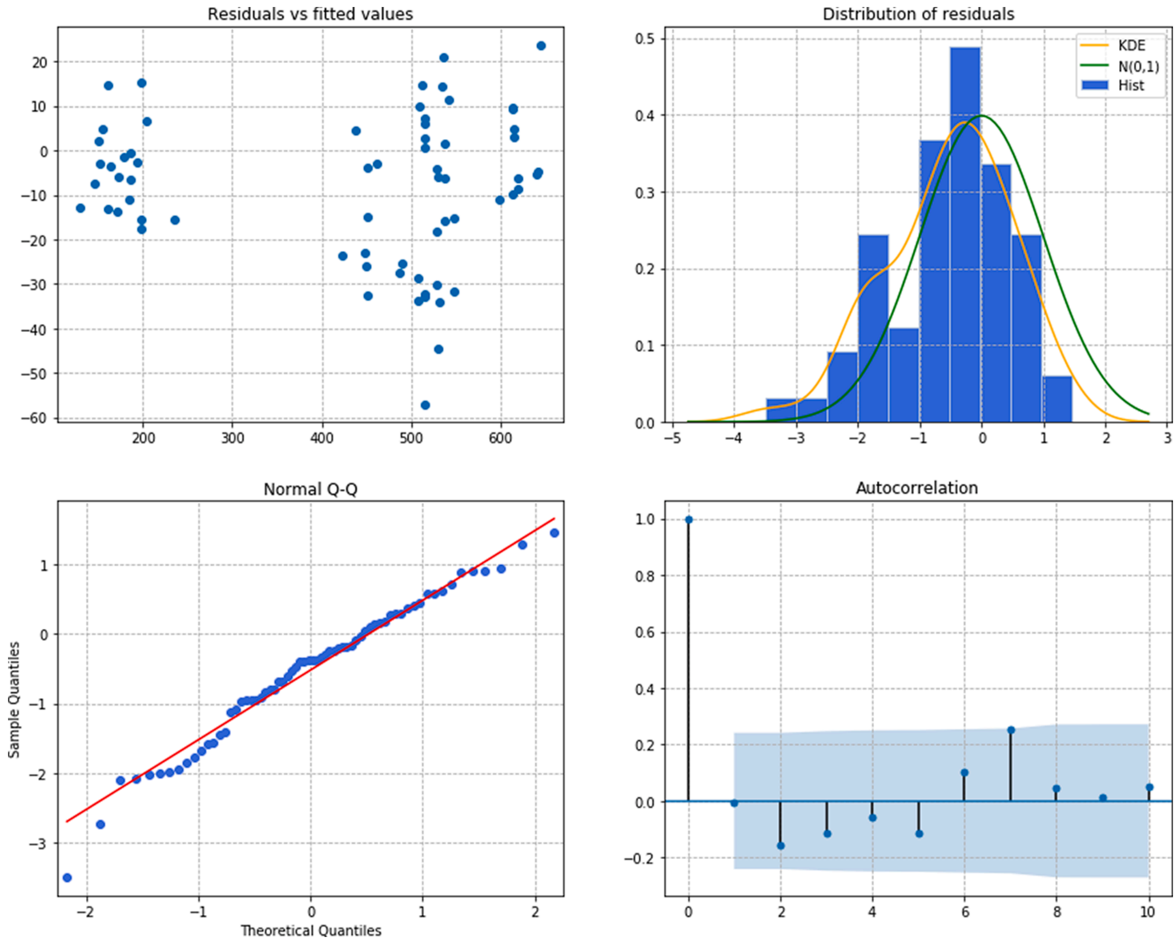

Fig. 10. Evaluation of the residuals of the test dataset of the Random Forest Regressor.

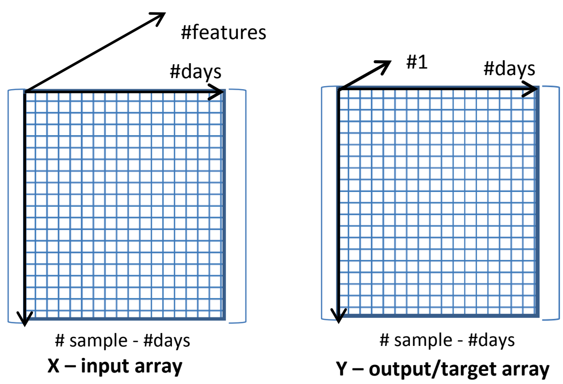

Fig. 11. Representation of the shapes of the X and Y arrays used with keras.

the SARIMAX model, the input data omit the weather-related features (min, average and max temperature and rain duration).

The analysis of the residuals of the test dataset suggests non-auto-correlated and normally distributed residuals. The plot of the residuals vs the fitted values, suggests some outliers. Nevertheless, they seem to have a zero mean and a constant variance. The Q-Q plot of the residuals vs theoretical normal quantiles shows a slight tail (see Fig. 13).

Table 7
Description of the LSTM model.

Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 100)	43200
repeat_vector_1	(RepeatVecto	(None, 7, 100)0
lstm_2 (LSTM)	(None, 7, 100)	80400
time_distributed_1	(TimeDist (None, 7, 200)	20200
time_distributed_2	(TimeDist (None, 7, 50)	10050
time_distributed_3	(TimeDist (None, 7, 1)	51

Total params: 153 901.
Trainable params: 153 901.
Non-trainable params: 0.

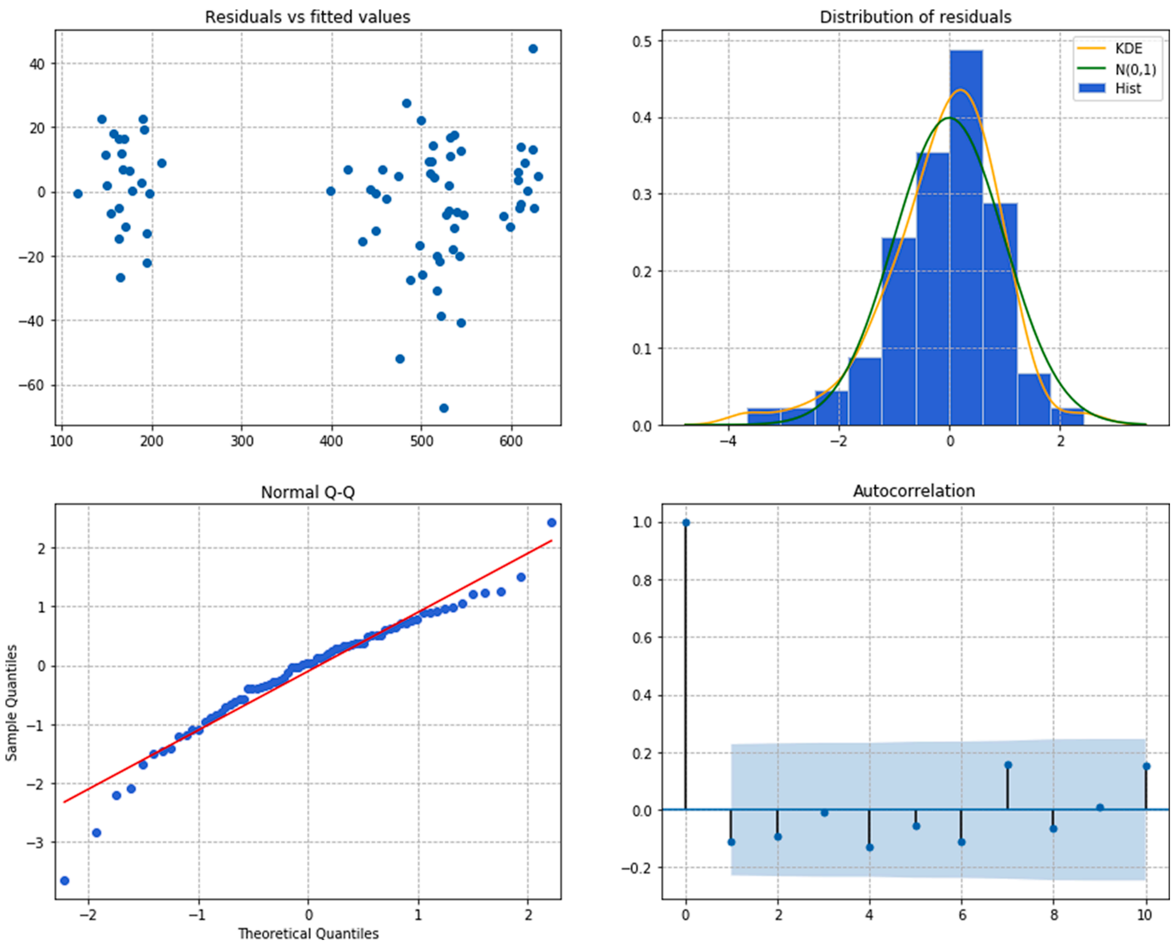

Fig. 12. Evaluation of the residuals of the test dataset of the LSTM model.

6. Model comparison and concluding remarks

The paper explores the determinants of the volatility of passenger demand of paratransit services and the feasibility of a data-driven model to accurate forecast future passenger demand and to improve operations scheduling. In addition to a naïve model, four models are developed: a seasonal ARIMAX model with a rolling forecast, a Random Forest Regressor with independent variables, a LSTM neural network with exogenous variables and a rolling forecast and a CNN model with independent variables. The results on the test dataset are illustrated in Fig. 14.

The evaluation metrics, summarized in Table 9, suggest the SARIMAX model performs well on most metrics, suggesting it is more robust to outliers, while large errors are limited. The SARIMAX underperforms on the MAPE metric, where the Random Forest Regressor performs best. This may be attributed to less accuracy when forecasting days with lower volumes. SARIMAX slightly outperforms the Deep Learning models, although the precision level is similar. This comes in line with Karlaftis and Vlahogianni (2011)

Table 8
Description of the CNN model.

Layer (type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 2, 64)	448
average_pooling1d_1	(Average(None, 1, 64)	0
flatten_1 (Flatten)	(None, 64)	0
dense_1 (Dense)	(None, 400)	26000
dropout_1 (Dropout)	(None, 400)	0
dense_2 (Dense)	(None, 80)	32080
dense_3 (Dense)	(None, 1)	81

Total params: 58 609.
Trainable params: 58 609.
Non-trainable params: 0.

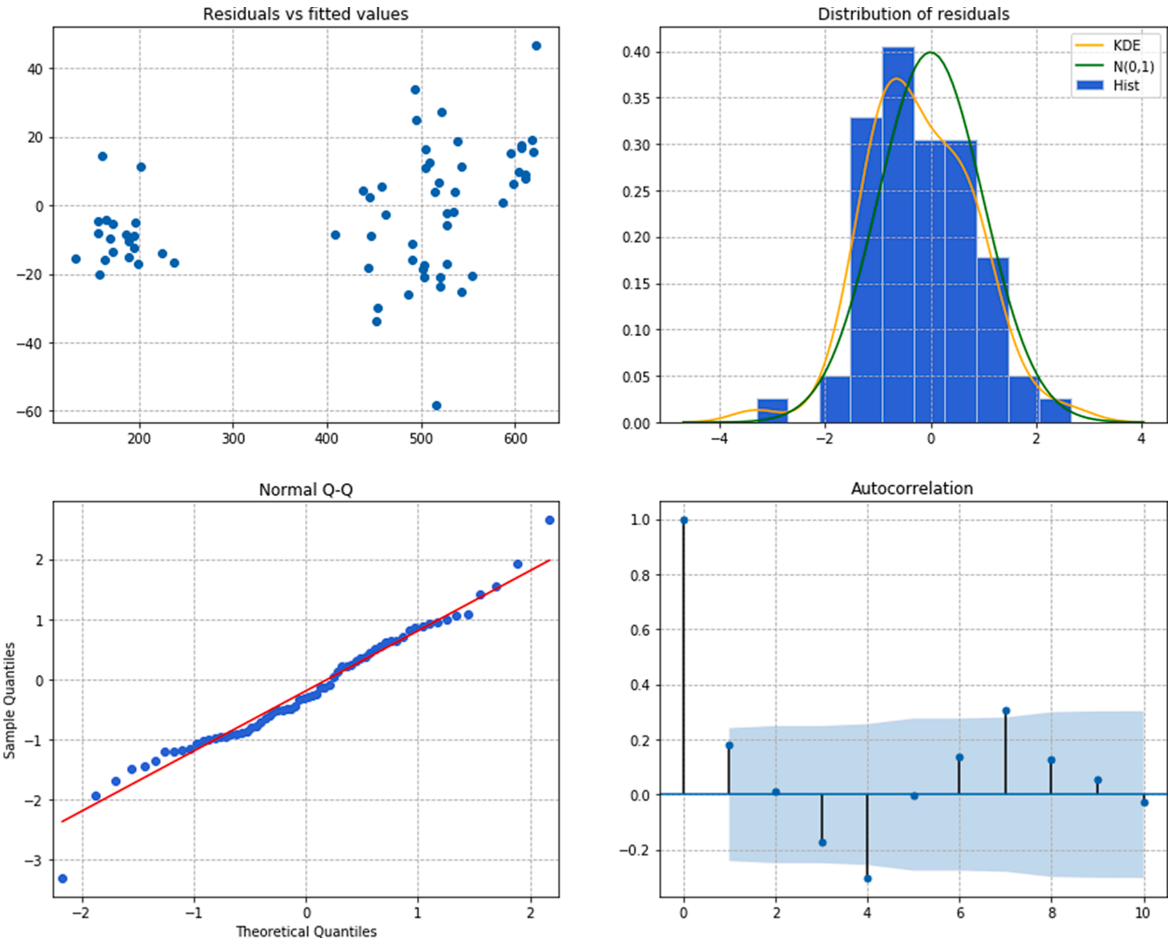

Fig. 13. Evaluation of the residuals of the test dataset of the CNN model.

who suggest that statistical and deep learning models have similar performance when addressing phenomena with linear relationships.

A qualitative interpretation of the results of the fitted models suggests they adequately forecast the volume of demand for typical days and predictable events, such as weather episodes. However, all the models present a significant error of roughly 10% on March 19th, which was the day of an interprofessional strike in Toulouse.

The time-series based simulations, as well as the other models, suggest that only a limited number of variables are necessary for an accurate forecast of daily demand. For a DRT service with prior reservation, these variables can be limited to the number of reservations seven days ahead of the departure, as well as the type of the day (day of the week and holidays). Additional variables, such as specific events (strikes etc.) could improve the forecast. These variables do not include weather related variables, such as the temperature, and the rain fall, as they do not seem significant for forecasting the demand. However, they may be significant for measuring the operational performance of the service.

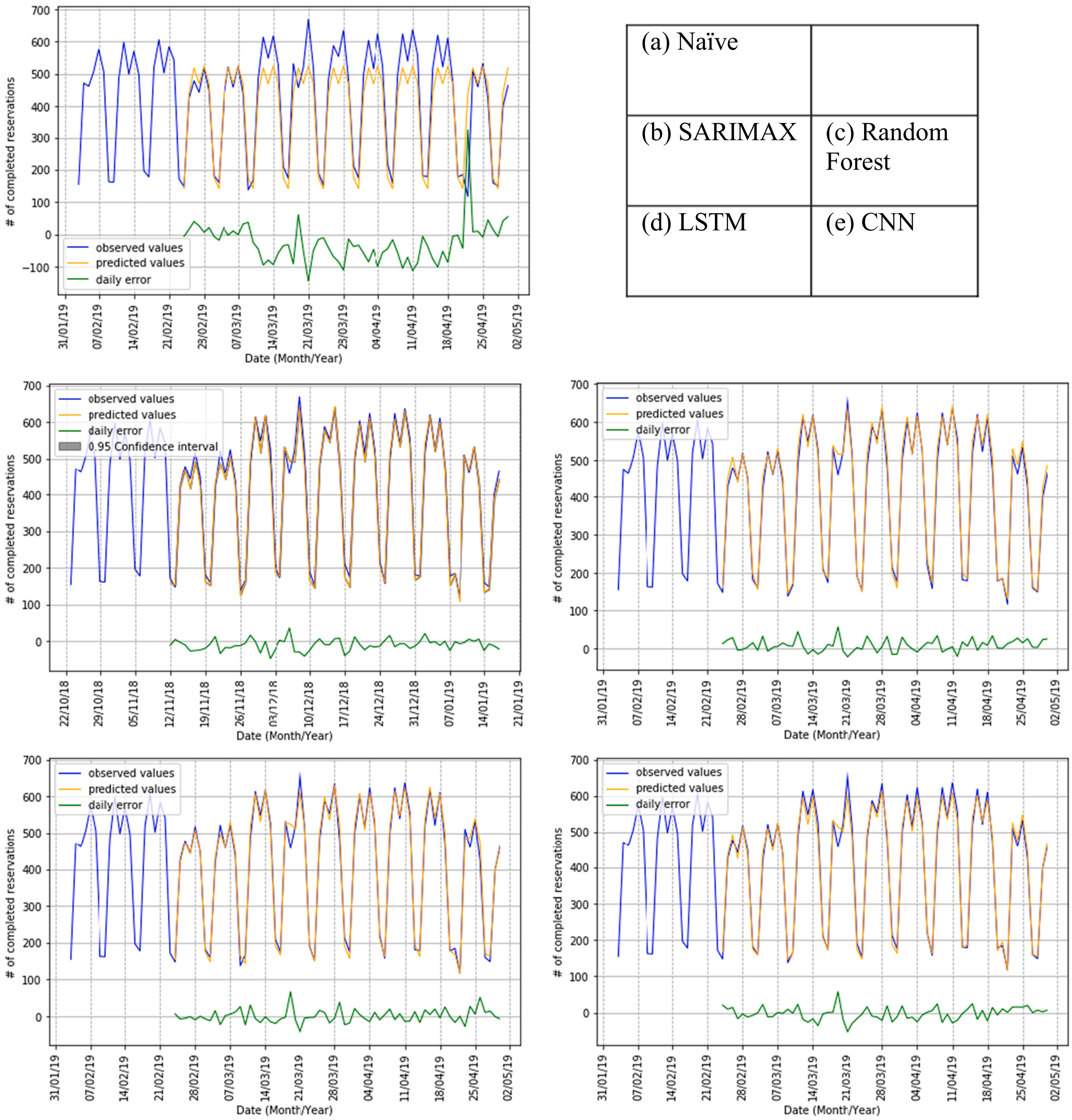

Fig. 14. Predicted vs, observed value of reservation with (a) top-left: naïve model (b) center-left: SARIMAX model, (c) center-right: Random Forest regression model (d) bottom-left: LSTM model (e) bottom-right: LSTM model.

Table 9
Evaluation metrics (Bold values indicate the best metrics).

Model type	MAE	RMSE	MASE	MAPE
Naïve Model	49.80	69.11	0.3537	14.87%
SARIMAX model	14.03	**17.85**	0.1034	4.24%
Random Forest Regressor	14.16	18.36	0.1006	**3.89%**
LSTM model	**13.73**	18.44	**0.0974**	4.08%
CNN model	14.63	17.87	0.1039	4.27%

Future work will address the specification and maintenance issues involved in including this SARIMAX architecture to the production chain. It will include model variation for D-10 and D-4 in order to provide some insight for resources planning. Future research will include the exploration of additional features, with respect to exceptional events, such as strikes, and will focus on the forecasting of multivariate time series, in order to address the within-the-day temporality of the demand. It will also seek to expand the medium-term forecasting to the human and vehicle resources needed. Additional model will include CNN-LSTM model architectures and a variation of the SARIMAX model with Kalman filters in order to grasp the hidden Markov chains of the time-series.

CRediT authorship contribution statement

Ektoras Chandakas: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing - original draft, Writing - review & editing, Visualization.

Acknowledgement

The author would like to thank Transdev for supplying the data used in this study, and in particular Thierry CHIOCCA for his valuable remarks in the analysis of the T-PRM dataset and the inspiration behind this paper.

References

Alonzo-Gonzalez, M., Liu, T., Cats, O., Van Oort, N., Hoogendoorn, S., 2018. The potential of demand-responsive transport as a complement to public transport: an assessment framework and an empirical evaluation. Transp. Res. Rec. 2672 (8), 879–889.

Attard, M., Camilleri, M., Muscat, A., 2020. The technology behind a shared demand responsive transport system for a university campus. Res. Transport. Bus. Manage. https://doi.org/10.1016/j.rtbm.2020.100463. In press.

Benjamin, J., Kurauchi, S., Moriwaka, T., Polydoropoulou, A., Sasaki, K., Ben-Akiva, M., 1998. Forecasting Paratransit ridership using discrete choice models with explicit consideration of availability. Transp. Res. Rec. 1618 (1), 60–65.

Box, G.E., Pierce, D.A., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series. J. Am. Stat. Assoc. 65, 1509–1526.

Chollet, F., 2015. keras, GitHub, https://github.com/fchollet/keras.

Coutinho F.M., van Oort N. Christoforou Z., Alonso-González M.J., Cats O. and Hoogendoorn S., 2020. Impacts of replacing a fixed public transport line by a demand responsive transport system: case study of a rural area in Amsterdam. Res. Transport. Econ. https://doi.org/10.1016/j.retrec.2020.100910. In Press.

Deka, D., Gonzales, E., 2014. The generators of paratransit trips by persons with disabilities. Transp. Res. Part A 70, 181–193.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods, second ed. Oxford University Press.

Farid, Y.Z., Christofa, E., Paget-Seekins, L., 2016. Evaluation of short-term bus travel time by using low-resolution automated vehicle location data. Transp. Res. Rec. 2539 (1), 113–118.

Gong, M., Fei, X., Wang, Z.H., Qiu, Y.J., 2014. Sequential framework for short-term passenger flow prediction at bus stop. Transp. Res. Rec. 2417 (1), 58–66.

Guo, J., Xie, Z., Qin, Y., Jia, L., Wang, Y., 2019. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. IEEE Access.

Ho, Tin Kam, 1995. Random decision forests, in: Proceeding of the 3rd International Conference on Document Analysis and Recognition, Montreal, pp. 278–282.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780.

Karlaftis, M.G., Vlahogianni, E., 2009. Memory properties and fractional integration in transportation time-series. Transport. Res. Part C: Emerg. Technol. 17, 444–453.

Karlaftis, M.G., Vlahogianni, E., 2011. Statistics versus neural networks in transportation research: differences, similarities and some insights. Transport. Res. Part C Emerg. Technol. 19 (3), 387–399.

Ke, J., Zheng, H., Yang, H., Chen, X., 2017. Short-term forecasting of passenger demand under on-demand ride services: a spatio-temporal deep learning approach. Transp. Res. Part C 85, 591–608.

Kingma, D., Ba, J., 2015. Adam: a method for stochastic optimization. in: Published as a Conference Paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Koffman, D., Lewis, D., 1997. Forecasting demand for Paratransit required by the Americans with disabilities act. Transp. Res. Rec. 1571 (1), 67–74.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, In: Advances in Neural Information Processing Systems, pp. 1097–1105.

Lin, L., Handley, J., Gu, Y., Zhu, L., Wen, X., Sadek, A., 2018a. Quantifying uncertainty in short-term traffic prediction and its application to optimal staffing plan development. Transport. Res. Part C: Emerg. Technol. 92, 323–348.

Lin, L., He, Z., Peeta, S., 2018b. Predicting station-level hourly demand in large-scale bike-sharing network: a graph convolutional neural network approach. Transport. Res. Part C: Emerg. Technol. 97, 258–276.

Ma, Z., Xing, J., Mesbah, M., Ferreira, L., 2014. Predicting short-term bus passenger demand using a pattern hybrid approach. Transport. Res. Part C: Emerg. Technol. 39, 148–163.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, 2015. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, Software available from tensorflow.org.

Moreira-Matias, L, Cats, O, 2016. Toward a Demand Estimation Model Based on Automated Vehicle Location. Transportation Research Record 2544 (1), 141–149.

Noursalehi, P., Koutsopoulos, H., Zhao, J., 2018. Real time transit demand prediction capturing station interactions and impact of special events. Transport. Res. Part C: Emerg. Technol. 97, 277–300.

OECD, International Transportation Forum, 2018. How to make urban mobility clean and green, Policy Brief, COP24, Available from https://www.itf-oecd.org/urban-mobility-clean-green.

Palmer, K., Dessouky, M., Zhou, Z., 2008. Factors influencing productivity and operating cost of demand responsive transit. Transp. Res. Part A 42, 503–523.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thririon, B., Grisel, O., Blondel, M., Prettehnhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Mach. Learn. Python JMLR 12, 2825–2830.

Polson, N.G., Sokolov, V.O., 2017. Deep Learning for short-term traffic flow prediction. Transp. Res. Part C 79, 1–17.

Rahimi, M., Amirgholy, M., Gonzales, E., 2018. System modelling of demand responsive transportation services: evaluation cost efficiency of service and coordinated taxi usage. Transp. Res. Part E 112, 66–83.

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and statistical modeling with python, in: Proceedings of the 9th Python in Science Conference, 2010.

Shaheen, S., Cohen, A., 2020. Chapter 3 – Mobility on demand (MOD) and mobility as a service (MaaS): early understanding of shared mobility impacts and public transit partnerships. In: Antoniou, C., Efthymiou, D., Chaniotakis, E. (Eds.), Demand for Emerging Transportation Systems – Modeling Adoption, Satisfaction, and Mobility Patterns, 2020.

Toque, F., Khouadjia, M., Come, E., Trepanier, M., Oukhellou, L., 2017. Short & Long term forecasting of multimodal transport passenger flows with machine learning methods. In: Paper Presented in IEEE 20th International Conference on Intelligent Transportation Systems (ITSC).

United Nations, Department of Economic and Social Affairs, Population Division, 2018. World Urbanization Prospects: The 2018 Revision, Online Edition, Available from https://esa.un.org/unpd/wup/Publications.

Van Oort, N., Brands, T., de Romph, E., 2015. Short-term prediction of ridership on public transport with smart card data. Transp. Res. Rec. 2535, 105–111.
Vlahogianni, E., Karlaftis, M.G., 2013. Testing and comparing neural network and statistical approached for predicting transportation time series. Transp. Res. Rec. 2399 (1), 9–22.
Vlahogianni, E., Karlaftis, M.G., Gollias, J., 2016. Short-term traffic forecasting: where and where we are going. Transport. Res. Part C: Emerg. Technol. 43, 3–19.
Wang, J., Deng, W., Guo, Y., 2014. New Bayesian combination method for short term flow forecasting. Transportation Research Part C 43, 79–94.
Williams, B., 2001. Multivariate vehicular traffic flow prediction. Transp. Res. Rec. 1776 (1), 194–200.
Xiaoqing, Dai, Lijun, Sun, Yanyan, Xu, 2018. Short-term origin-destination based metro flow prediction with probabilistic model selection approach. J. Adv. Transport. 2018.