Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 47 (2020) 664–671

Transportation Research Procedia

www.elsevier.com/locate/procedia

22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18-20 September 2019, Barcelona, Spain

Spatiotemporal Variation of Taxi Demand

Pedro Rodrigues[a*], Ana Martins[b], Sofia Kalakou[c], Filipe Moura[b]

[a]*Instituto Superior Técnico, Universidade de Lisboa; Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal*
[b]*CERiS, Instituto Superior Técnico, Universidade de Lisboa; Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal.*
[c]*Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL); Av. Das Forças Armadas, 1649-026 Lisboa, Portugal.*

Abstract

The growth of urban areas has made taxi service become increasingly more popular due to its ubiquity and flexibility when compared with, more rigid, public transportation modes. However, in big cities taxi service is still unbalanced, resulting in inefficiencies such as long waiting times and excessive vacant trips. This paper presents an exploratory taxi fleet service analysis and compares two forecast models aimed at predicting the spatiotemporal variation of short-term taxi demand. For this paper, we used a large sample with more than 1 million trips between 2014 and 2017, representing roughly 10% of Lisbon's fleet. We analysed the spatiotemporal variation between pick-up and drop-off locations and how they are affected by weather conditions and points of interest. More, based on historic data, we built two models to predict the demand, ARIMA and Artificial Neural Network (ANN), and evaluated and compared the performance of both models. This study not only allows the direct comparison of a linear statistical model with a machine learning one, but also leads to a better comprehension of complex interactions surrounding different urban data sources using the taxi service as a probe to better understand urban mobility-on-demand and its needs.

Keywords: Taxi demand; ARIMA; Artificial Neural Network

1. Introduction

Different from other public transports like subways or buses which follow previously stipulated routes with waiting times, taxicabs have become increasingly popular in urban areas, due to its great mobility, plannable routes and uninterrupted door-to-door accessibility. However, in big cities taxi-service is imbalanced (Huang & Powell, 2012), often characterized by long waiting times and an excessive number of vacant vehicles (Zhao et al., 2016) wandering in the city increasing air pollution and traffic congestion, while decreasing customer satisfaction levels and companies' profit. Like many other fields, the taxi business is now undergoing a significative digital revolution, competing for

* Corresponding author. Tel.: +351 929 185 837.
E-mail address: geraldes.rodrigues@tecnico.ulisboa.pt

10.1016/j.trpro.2020.03.145

market shares with innovative new ride-hailing agents such as Uber, Lyft, etc. Therefore, the ability to predict where demand will emerge and minimize the vacant driving time is of the most importance to taxi drivers and companies, especially when there is no economic viability of adopting cruising strategies to find their next passenger (Moreira-Matias et al., 2013). Other authors (Yang et al., 2000) studied the asymmetry that characterizes taxi movements, showing that both supply and demand often cannot be met due to spatiotemporal variations in passenger demand and spatial preferences of the drivers. Following this thought, Nam et al. (2016) developed a geographical weight spatial regression model after decomposing the city of Seoul into grid cell structures and studying the spatial influences of urban density and traffic-related factors on taxi ridership. A framework to describe the spatiotemporal structure of passenger demand was presented by Lee et al (2008) applied on the Jeju Island, South Korea, whereas Yuan et al. (2011) presented a complete work with various methods where they focus the division of the urban area into pick-up zones using spatial clustering. In Portugal, Moreira-Matias (2013) combined three different time series models to the city of Porto using real-time streaming data to predict short-term demand using the city's taxi stands as clusters, achieving an accuracy of 23.97% with an Ensemble method. Phithakkitnukoon (2010) and Veloso (2011) both focused on modeling the city of Lisbon into a grid presenting an inference engine based on the naïve Bayesian Classifier to predict the number of vacant taxis in a given area based on time of the day, day of the week and weather conditions with a 56.6% accuracy. This paper uses the city of Lisbon as a case study to present an exploratory taxi fleet service analysis in which two forecast models (ARIMA and ANN) of short-term taxi demand are developed and compared. In Section 2 the case study is presented. Section 3 describes the methodology and the results of the analysis are presented in Section 4. Finally, Section 5 draws the main conclusions of this work.

2. Case-study

2.1 Data

The taxi data used in this work was provided by a private company of taxis, AutoCoope, whose fleet represents 10% of Lisbon's active taxi fleet. Input data dates from January 1, 2014, to December 31, 2017. The information is categorized by departure and arrive location (latitude and longitude), departure and arrival time, taxi ID and type of call. A data enrichment step was performed on the collected data, allowing us to build a consistent database based on the initial sample, removing erroneous information. Due to the large number of registers, the database was set in MySQL. In the preliminary analysis we explored descriptive statistics of taxi data using the Jupyter notebook with Python, connected to the database. We discretized the city into a grid of 500m×500m cells, the time into intervals and we defined other needed variables such as distance travelled and duration of trip. Points of interest in the city (restaurants, hotels, transportation hubs, etc.) were identified. Weather phenomena were also considered in the analysis (temperature and precipitation). For data visualization we used the software Tableau and for the GIS QuantumGIS.

2.2 Spatiotemporal Analysis

An exploratory analysis is an effective tool to understand the system's variables and to detect emerging patterns and relationships between them. Our case-study is the council of Lisbon which encompasses an area of roughly 110 km^2 and a population of 800 000 habitants (585 000 in the city of Lisbon) with roughly 9 tourists per resident (CML, 2018). Lisbon's taxi demand is related to a myriad of factors, namely the high number of visiting tourists and working citizens that commute from the outskirts of the city, existence of points of interest (POIs), price, waiting time and other transport solutions (INE, 2018). Our dataset contains 632 794 unique trips and 254 distinct taxis over the course of 4 years. Since our sample size (*i.e.* number of operating taxicabs each year) is not constant over time, we selected the year of 2016 for our analysis and model training, and the first month of 2017 for model testing. We modelled Lisbon's city in a map with a grid of 500m×500m for our spatial analysis. Cell size was determined by the maximum average distance that passengers are willing to travel without using any other means of transport (Daniels & Mulley, 2011; Dunning & Ford, 2003; Thompson & Bae, 2014) and used by transport authorities in designing transportation hubs (Public Transport Authority, 2003). Fig. 1 shows the spatial distribution of pick-up (left) and drop-off (right) locations of taxi service in Lisbon. The cells with the highest density of taxi services are those that contain transportation hubs (*i.e.* airport, buses, trains, ferries) or commercial and business centres that enjoy a good public transportation network. The colour classes follow the model of data cluster optimization proposed by Jenks (1967).

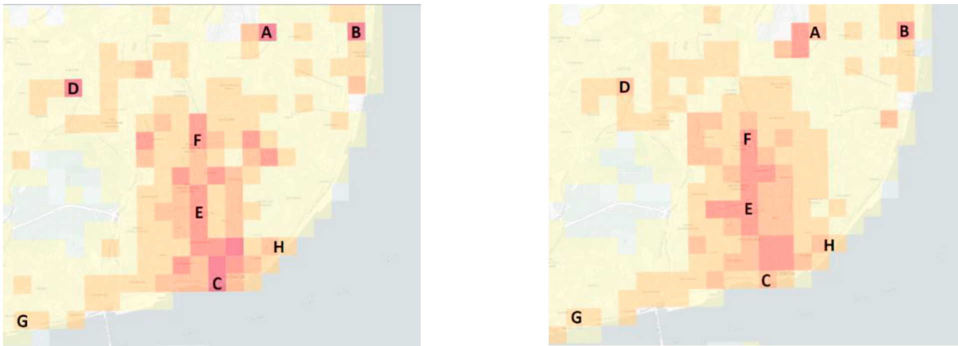


Fig. 1 - Taxi service pick-up (left) and drop-off (right) locations distribution in Lisbon, Portugal (legend in Figure 2).

Fig. 2 displays how the pick-up and drop-off areas relate with each other. The thickness of line represents the intensity of trips between two points. A few hotspots of demand appear to form: A (Airport); B (Train station with a commercial area); C (Train station and ferry dock); D (Commercial area); E (Business district); F (Commercial area); G and H (Train station).

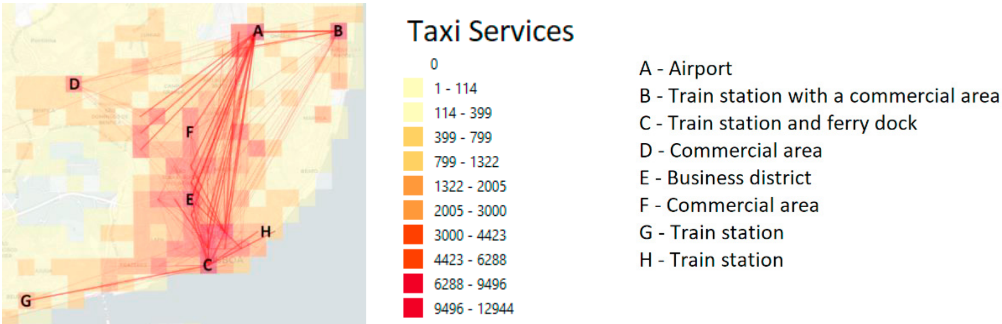


Fig. 2 – Hotspots of taxi demand in Lisbon, Portugal and how strongly connected they are, according to taxi services.

From this observation we estimate that the taxi service is often used as a bridge between public transportation modes, an answer to “first and last mile" problems, but mostly as a direct mobility solution between commercial and touristic centres. It is also important to point out that locations D, E, F (some of the most frequent pick-up or drop-off locations) give access to services and commercial areas, whereas locations C and G also encompass high night-life activities and serve as a major touristic hotspot, respectively. Taxi service demand varies not only in space, but also in time, adapting to the needs of citizens. Fig. 3 shows the temporal distribution of taxi services per day in 2016, where some cyclical patterns can be observed, especially weekend cycles. Some punctual disruptions are visible, coinciding with Christmas Eve, Christmas Day and New Year's Eve, where a reduction in the taxi service is present. Demand is influenced by the commuting behaviour associated with labour activity. An analysis to the weekly and weekend demand showed an increase in taxi services early in the morning (08h00 - 09h00) and in the afternoon (18h00 – 19h00) during the week. On weekends, services rise in the night (20h00 – 22h00) and dawn (00h00 – 01h00). To better infer on the patterns that characterize short term taxi demand, we studied the trips by distance and duration, as shown in Fig. 4. These graphs show that more than 90% of trips are greater than 1 km and 77% last between 5 and 20 minutes, revealing the use of the taxi for short trips within the city of Lisbon, despite the existing public transport network and other mobility alternatives.

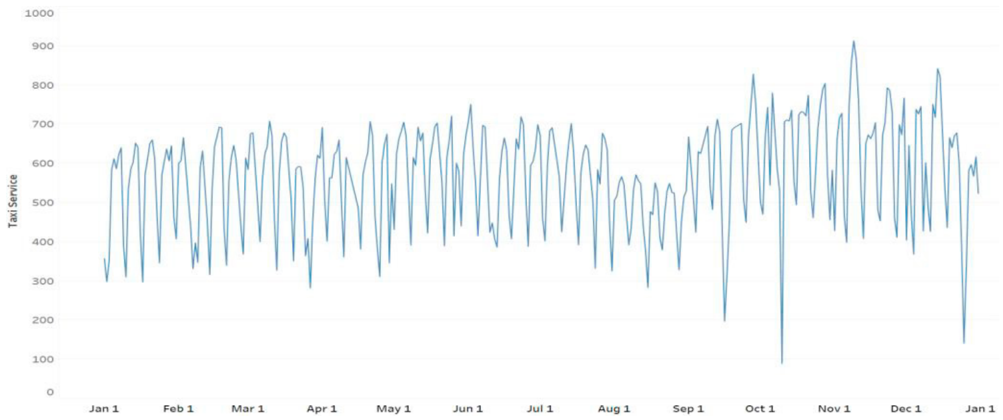


Fig. 3 - Temporal distribution of taxi services in 2016.

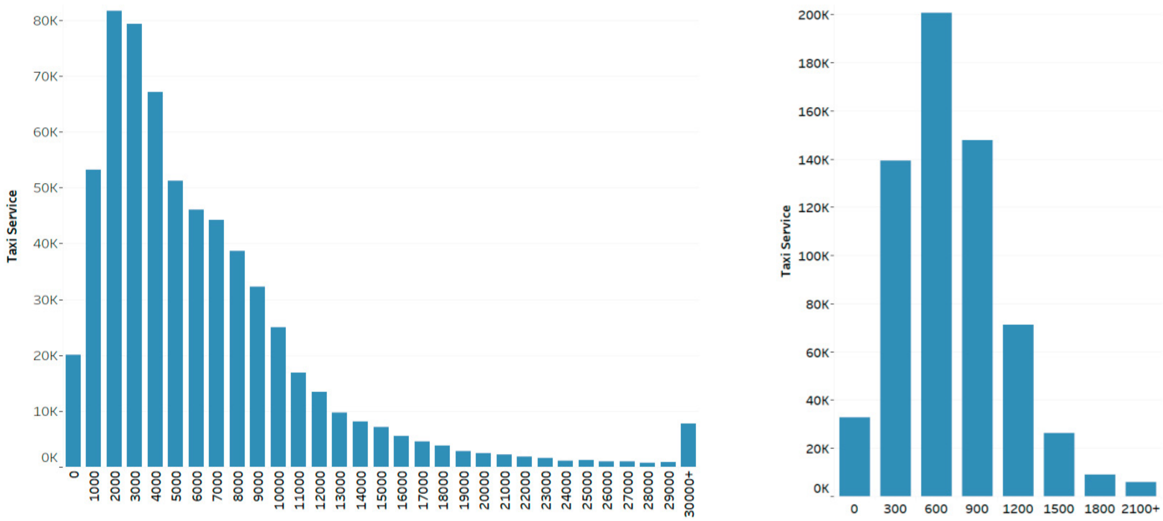


Fig. 4 - Distribution of trips distances per 1000 m (left) and trips duration in 5 min intervals (300 seconds, right) in Lisbon, 2016.

3. Methodology

3.1 ARIMA

The AutoRegressive Integrated Moving average (ARIMA) is a linear method widely used to model and predict traffic flow data and speeds (Min & Wynter, 2011), electricity prices (Contreras et al., 2003), as well as other short-term prediction problems like taxi demand. It is mostly used with univariate time series data, assuming that traffic conditions can be modelled as a stationary process, with properties like mean and variance constant over time. The model combines the most recent observed data to generate a forecast with the ability to update itself according to possible changes in the model. One of ARIMA's main advantages is its flexibility in representing different types of time series such as the AutoRegressive (AR), Moving Average (MA) and a combination of both (ARMA). As proposed by Box et al. (1976), the model parameters and weights are obtained by analysing the autocorrelation (ACF) and partial autocorrelation functions (PACF) of the time series. Despite ARIMA models' high popularity in modelling time series data, they fail to directly handle several seasonal patterns (Washington et al., 2003). To tackle multiple seasonality, external regressors were added to the ARIMA model in the form of Fourier terms, to model each seasonal cycle independently. Our model's equation shifts from the ARIMA generic one to equation 1 presented below. By introducing Fourier regressors to model seasonality, we had to prevent the ARIMA to also model seasonality by creating a condition where that particular property of the model was restrained to prevent redundancy.

$$y_t = a + \sum_{i=1}^{M} \sum_{w=1}^{W} \left[\alpha_w \sin\left(\frac{2\pi wt}{m}\right) + \beta_w \cos\left(\frac{2\pi wt}{m}\right) \right] + A_t \quad \textbf{(1)}$$

Where y_t is the predictive variable, m the seasonality frequency, M the number of Fourier terms necessary to model that frequency, W the selected cluster of the time series, t is time interval and A_t is an ARIMA process; a, α_w and β_w are Fourier constants. The addition of Fourier terms allows for the seasonal patterns to be modelled using them as external regressors. This approach is flexible enough to allow the incorporation of multiple terms, and each period, M, is obtained by minimizing the corresponding AIC. A unique Fourier equation is used for each seasonal pattern, where an iteration process computes multiple combinations of M periods, storing only the forecast values with the minimum AIC value. According to Hyndman & Athanasopoulos (2013) this approach has several advantages, namely the ability to incorporate any seasonality despite its length and the seasonal pattern is smooth for small values of M and more unstable seasonality can be handled by increasing M. Due to ARIMA's lack of spatial correlation, the model was set by learning the independent historical time-series of pick-up taxi services in each separate cluster for the year of 2016. To do so, we computed periodograms using R programming language, extracting the main power frequencies that were then introduced into the ARIMA as Fourier regressors. We used the *auto.arima*, an automatic function in the [forecast] R package (Hyndman & Khandakar, 2008), restricting its default seasonality capability.

3.2 Artificial Neural Network: Multi-Layer Perceptron

An artificial neural network attempts to emulate the human brain, featuring an architecture that simulates the nonlinear and complex processing of biological units. Any ANN is made of a number of neurons organized in layers, consisting of at least two layers, the input layer and the output layer. A third category of layers, called hidden layers, is often added to increase complexity and problem-solving capability by allowing more connections between neurons. Our approach started by normalizing our inputs through a robust scaler before feeding them to the MLP. The emerging demand was identified as our dependent variable whereas our independent input variables were those whose influence we wanted to test: number of taxis *per day*, period of the day, day of the week, day of the month, month of the year, lag variables as well as precipitation and temperature. These new variables account for relevant weather conditions as shown in (Pele & Morency, 2014; Veloso, 2016) and are parameters ideally suitable for a ANN solution (Mitchell, 1997). The training set was the whole year of 2016, with 10% of its data used also for validation, and the first month of the year of 2017 used as test set. After randomly generating its architecture, the model trains itself against the separate dataset from each cluster of 2016, performing up to 1000 backpropagation iterations to adjust the weights of every connection between neurons. Afterwards, it tests its predictions in every cluster for the whole month of January 2017, calculating the sMAPE error for each one. This is done 100 different times, totaling in 100 different neural networks, each one performing up to 1000 iterations per cluster. The final parameters for the ANN are shown below.

Table 1 - Neural Network parameters

Final network implementation parameters	
Hidden layers and neurons	[15, 15, 10, 10, 5, 5, 5, 5, 5, 10, 5, 5, 5]
Loss function	Mean squared error
Optimization algorithm	ADAM optimizer
Activation function	Rectified linear unit function, ReLU
Overfitting prevention	Early stopping after ten training epochs with an improvement inferior to 1×10^{-4} in the loss function of the validation set.

The selected optimization algorithm was the ADAM, whose strengths are shown by Kingma and Ba (2014). The chosen activation function was the ReLU (Rectified Linear Unit). Although there is no universal optimal activation function, the ReLU has been increasing in popularity, and it doesn't saturate the gradients (like the sigmoid or tanh do) proving to often accelerate the convergence of the gradient descent, due to its linear and non-saturating form (Krizhevsky et al., 2012).

3.3 Evaluation Metrics

We used the real demand values to measure the performance of both models. Two error metrics were also used to measure the algorithms' accuracy. The first used was the root mean squared error (RMSE), which is mostly used to measure the model's relative forecast accuracy, shown in the left pane of equation 2 below. The second metric is the mean symmetric error percentage (sMAPE). This metric is a widely used alternative when there are demand values close to or equal to zero. It reduces the influence of these small volume data by limiting the error rate to 200%. Low volume observations are problematic because they could have infinitely high error rates that would distort the overall error rate.

$$RMSE = \sqrt{\frac{1}{D}\sum_{d=1}^{D}(\hat{y}_w^d - y_w^d)^2} \qquad sMAPE_w = \frac{1}{D}\sum_{d=1}^{D}\frac{|\hat{y}_w - y_w|}{Q_w} \qquad Q_w = \begin{cases} \hat{y}_w + y_w & \text{if} \quad (\hat{y}_w > 0 \vee y_w > 0) \\ 1 & \text{if} \quad (\hat{y}_w = 0 \wedge y_w = 0) \end{cases} \qquad \textbf{(2)}$$

Where D is the predicted sample size (number of predicted 8h intervals) of cluster w, $\hat{y}$ is the predicted demand and y is the actual demand.

4. Experimental Results

4.1 Experimental Setup

Both models were applied to the following clusters from Fig. 2: **A** (Airport); **B** (Train station with a commercial area - Oriente); **C** (Train station and ferry dock - Downtown); **D** (Commercial area - Colombo). Clustering was essential in the analysis due to the inherent operation characteristics of ARIMA which does not allow spatial correlation. An aggregation period of 8 hours was defined, meaning a new forecast was generated every 8 hours. This aggregation period was created to ensure that there were significant data points in the time-intervals and to mimic the working shifts of taxi drivers. A sliding window of one year was used, where both models generated 3 forecasts being updated with the real demand values every 24h.

4.2 Results

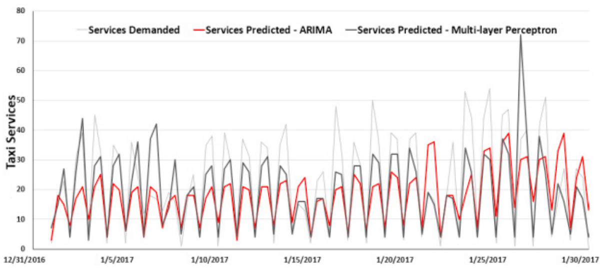
We present our results in two distinct perspectives. First, we showcase the error of both models; Second, we plot a comparative analysis where both model's forecasts are measured against the actual demand. Table 2 shows the accuracy of both algorithms in the selected clusters and scenarios.

Table 2 - Error measured on both models using sMAPE and RMSE.

	ARIMA with Fourier Regressors				
	Cluster	**00h-08h**	**08h-16h**	**16h-00h**	**24h**
sMAPE	**A – Airport**	43.5%	25.6%	22.9%	30.7%
	B – Oriente	45.2%	30.6%	21.5%	32.4%
	C – Downtown	29.1%	23.0%	18.7%	23.6%
	D – Colombo	50.9%	35.1%	30.5%	38.8%
RMSE	**A – Airport**	6.2	16.7	13.8	13.0
	B – Oriente	2.2	4.1	4.1	3.6
	C – Downtown	8.0	10.9	5.0	8.3
	D – Colombo	1.4	3.6	4.8	3.6

Multi-Layer Perceptron					
	Cluster	**00h-08h**	**08h-16h**	**16h-00h**	**24h**
sMAPE	**A – Airport**	30.1%	15.3%	17.2%	20.9%
	B – Oriente	32.1%	22.8%	22.9%	26.0%
	C – Downtown	20.6%	22.5%	26.7%	23.3%
	D – Colombo	31.1%	26.2%	20.6%	26.0%
RMSE	**A – Airport**	6.2	11.8	11.8	10.3
	B – Oriente	3.9	3.4	4.7	4.0
	C – Downtown	9.1	21.2	46.4	29.3
	D – Colombo	2.6	2.5	3.3	2.8

The overall performance of both models is good, with the MLP overshadowing the ARIMA in every cluster and working shift in both sMAPE and RMSE. Although the RMSE is a relative error measure, the MLP model performance is in line with the real predicted taxi demand, which is noisy non-stationary, and was able to observe non-linear relations that the ARIMA couldn't. Both models don't depend on how the data are spatially aggregated, depending only on the aggregation period by the user. The ARIMA model's errors are higher during the first working shift (00h – 08h), probably because it's a period that encompasses a heterogeneous group, such as late-night activities and early morning commuting. The highest sMAPE error (50.9%) belongs to the smaller cluster containing a commercial area that closes at 23h00 and opens at 09h00, causing a reduction in available taxis in the area, proving that cluster area and number of available taxis in a given cell are important variables. The averaged error of the ARIMA was 31.4% and the MLP was 23.4%. Other works in Portugal achieved a sMAPE of 27% (Moreira-Matias et al., 2013) for the ARIMA in Porto, but they used taxi stands with a 50m radius and not clusters for the emerging demand. In Lisbon, Ant & Veloso (2016) used a Bayesian inference engine but his model also accounted for mobile usage and vehicles emissions with a 56.6% accuracy. Fig. 5 shows the forecasted demand of both models against the real demand in two distinct scenarios. In cluster **A**, we have a major transportation hub (Airport) with regular patterns of emerging services (caused by the arrivals of airplanes), whereas in cluster **D** we have one of the biggest commercial centres of Lisbon, causing the demand to be irregular and without a usual pattern. Both examples allow us to illustrate the forecast accuracy of both models in distinct periods and scenarios, proving the non-linearity of taxi demand.



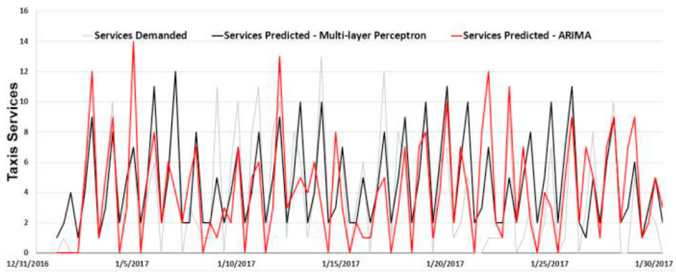


Fig. 5 - Services Demanded and Services Predicted by both models in periods of 8h during the month of January 2017 in cluster **A** (left) and cluster **D** (right). **Red** represents the ARIMA, **Black** the MLP and **Grey** the real demand.

5. Conclusions and Future Work

In this paper, we compared the performance of two short-term traffic demand forecasting models based on the analysis of taxi data from a leading taxi company in Lisbon, Portugal. Both the ARIMA and the MLP model aimed to predict the short-term taxi demand in various hotspots of emerging services. We presented the obtained results in four different clusters with data of 213 distinct vehicles retrieved over 1 year that. The data was converted into a time series and used as a training base for our models and one month was used as a test set. Both models exhibit satisfactory results, with averaged errors of 31.4% for the ARIMA and 23.4% for the Neural Network. The introduction of Fourier terms to the ARIMA to model seasonal patterns allowed the model to mine both the periodicity and seasonality of passenger demand, while updating itself using historical data as a learning base. We were able to observe spatiotemporal patterns of taxi demand, identify the most common origin and destination service areas and showed how the taxi exhibits strong connections between areas with large transport interfaces, appearing to be a connecting link between the different transport alternatives available. As future works, the sliding window of the ARIMA model could be reduced to reduce noisy data. Other Neural Network models should be tested with this dataset, namely LSTMs and RNNs that have been widely used recently in predicting short-term taxi demand, to try and identify long-term dependencies.

Acknowledgments

The Portuguese Funding Agency for Science and Technology (FCT, I.P.) is gratefully acknowledged for financing the second author with grant No. PD/BD/128045/2016. The authors would also like to thank AutoCoope and its team for the data supplied.

References

Ant, M.; Veloso, M. (2016) Analysis of Taxi Data for Understanding Urban., (February).

Box, G.; Jenkins, G.; Reinsel, G. (1976) Time Series Analysis. San Francisco.

CML (2018) ESTUDO URBANÍSTICO DO TURISMO EM LISBOA.

Contreras, J.; Espinola, R.; Nogales, F. J. (2003) Arima models to predict next-day electricity prices. In *IEEE Transactions on Power Systems, vol. 18, no. 3* (pp. 1014–1020).

Daniels, R.; Mulley, C. (2011) Explaining walking distance to public transport: the dominance of public transport supply.

Dunning, B.; Ford, I. (2003) Personal automated transportation: status and potential of personal rapid transit. *Advanced Transit Association*.

Huang, Y.; Powell, J. (2012) Detecting Regions of Disequilibrium in Taxi Services. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems* (pp. 139–148).

Hyndman, R. J.; Athanasopoulos, G. (2013) Forecasting: Principles and Practice.

Hyndman, R. J.; Khandakar, Y. (2008) Automatic Time Series Forecasting: the forecast Package for R. *Journal of Statistical Software*, *27* (3).

INE (2018) Inquérito à Mobilidade nas Áreas Metropolitanas do Porto e de Lisboa Predomínio do automóvel nas deslocações dos residentes das Áreas Metropolitanas de Porto e Lisboa., pp 1–22.

Jenks, G. (1967) International Yearbook of Cartography 7: The Data Model Concept in Statistical Mapping.

Kingma, D. P.; Ba, J. (2014) Adam: A Method for Stochastic Optimization., pp 1–15.

Lee, J.; Shin, I.; Park, G. L. (2008) Analysis of the passenger pick-up pattern for taxi location recommendation. *Proceedings - 4th International Conference on Networked Computing and Advanced Information Management, NCM 2008*, pp 199–204.

Min, W.; Wynter, L. (2011) Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, *19* (4), pp 606–616.

Mitchell, T. M. (1997) Machine Learning. New York: McGraw-Hill.

Moreira-Matias, L.; Gama, J.; Ferreira, M.; Mendes-Moreira, J.; Damas, L. (2013) Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, *14* (3), pp 1393–1402.

Nam, D.; Hyun, K.; Kim, H.; Ahn, K.; Jayakrishnan, R. (2016) Grid Cell Based Taxi Ridership Analysis Using Large Scale GPS Data. *Transportation Research Board*.

Pele, N.; Morency, C. (2014) Where and How Taxis Are Used in Montreal.

Phithakkitnukoon, S.; Veloso, M.; Bento, C.; Biderman, A.; Ratti, C. (2010) Taxi-aware map: Identifying and predicting vacant taxis in the city. In *Lecture Notes in Computer Science book series (volume 6439)* (pp. 86–95). Berlin: Springer.

Public Transport Authority (2003) Design and Planning Guidelines for Public Transport Infrastructure: Bus Route Planning and Transit Streets. Western Australia: Public Transport Authority.

Thompson, M.; Bae, H. (2014) A Functional Thinking Approachto the Design of Future Transportation Systems: Taxis as a Proxy for Personal Rapid Transit in South Korea. (pp. 47–63). Denmark.

Veloso, M. (2016) Analysis of taxi data for understanding urban dynamics. Universidade de Coimbra.

Veloso, M.; Phithakkitnukoon, S.; Bento, C. (2011) Sensing urban mobility with taxi flow. *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks - LBSN '11*, pp 1.

Washington, S. P.; Karlaftis, M. G.; Mannering, F. L. (2003) Statistical and Econometric Methods for Transportation Data Analysis (Second Edi). CRC Press.

Yang, H.; Lau, Y. W.; Wong, S. C.; Lo, H. K. (2000) A macroscopic taxi model for passenger demand, taxi utilization and level of services. *Transportation*, *27* (3), pp 317–340.

Yuan, J.; Zheng, Y.; Zhang, L.; Xie, Xi.; Sun, G. (2011) Where to find my next passenger? In *13th international conference on Ubiquitous computing - UbiComp '11* (pp. 109–118). Beijing.

Zhao, K.; Khryashchev, D.; Freire, J.; Silva, C.; Vo, H. (2016) Predicting taxi demand at high spatial resolution: Approaching the limit of predictability. In *IEEE International Conference on Big Data (Big Data)*.