sciendo

Transport and Telecommunication, 2020, volume 21, no. 4, 255–264
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2020-0020

A DEEP LEARNING APPROACH FOR PREDICTING BUS PASSENGER DEMAND BASED ON WEATHER CONDITIONS

Tânia Fontes[1], Ricardo Correia[2], Joel Ribeiro[3], José Luís Borges[4]

[1, 3, 4] INESC TEC - Institute for Systems and Computer Engineering, Technology and Science, Porto, Portugal, Rua Dr. Roberto Frias, 4200-465
[1]tania.d.fontes@inesctec.pt, [3]joel.ribeiro@inesctec.pt, [4]jlborges@fe.up.pt
[2] INESC TEC and Faculty of Engineering, University of Porto, Porto, Portugal, Rua Dr. Roberto Frias, 4200-465
ricardo.f.correia@inesctec.pt

This work apply a deep learning artificial neural network model – the Multilayer Perceptron – as a regression model to estimate the demand of bus passengers. Transit bus ridership and weather conditions were collected over a year from a medium-size European metropolitan area and linked under the assumption: individuals choose the travel mode based on the weather conditions that are observed during (a) the departure hour, (b) the hour before or (c) two hours prior to the travel start. The transit ridership data were also labelled according to the hour of the day, day of the week, month, and whether there was a strike and/or holiday or not. The results show that the prediction error of the model decrease by ~9% when the weather conditions observed two hours before travel start is taken into account. The model sensitivity analyses reveals that the worst performance is obtained for a strike day of a weekday in spring (typically Wednesdays or Thursdays).

Keywords: Predition; Urban public transport; Bus passenger demand; Weather conditions; Artificial neural networks.

1. Introduction

Weather conditions have a major impact on citizens' daily mobility. Depending on weather conditions, trips may be delayed or demand may change. Also, the modal shift can be significantly transformed, since some groups of passengers, such as the elderly, the sick and children, tend to avoid to travel under certain conditions. Moreover, with climate change, the influence of weather conditions on transport demand has been raising a growing interest. Weather variations may have a major impact on the use and operation of public transport, particularly in transport systems that operate close to its capacity.

Several authors have been studied the impact of weather conditions of transport demand in several countries as Portugal (Costa *et al*., 2017), Netherlands (Creemers *et al*., 2015), Spain (Arana *et al*., 2014), Australia (Tao *et al*., 2016), China (Li *et al*., 2018; Zhou *et al*., 2017), and United States (Kalkstein *et al*., 2009; Guo *et al*., 2008; Stover and McCormack, 2015). Buses (Tao *et al*., 2016; Stover and McCormack, 2015) and metro (Li *et al*., 2018; Zhou *et al*., 2017) have been the means of transport more studied. Table 1 presents an overview of the existing studies.

Many techniques have been applied to study the impact of weather conditions on public transport demand such as multivariate regressions (Arana *et al*., 2014; Zhou *et al*., 2017), linear regression (Creemers *et al*., 2015), Quasi-Poisson regression (de Montigny *et al*., 2011), negative binomial distribution (Schmiedeskamp and Zhao, 2016), minimum square method (Singhal *et al*., 2014; Stover and McCormack, 2015) and moved average (Li *et al*., 2018). These studies explain that the use of public transport is highly influenced by variations on weather conditions (e.g. Guo *et al*., 2008). Temperature is one of the variables that most affect passengers, unlike atmospheric pressure that does not appear to cause any negative effect (Costa *et al*., 2017).

A direct correlation between certain weather conditions and the demand generated in transport services was obtained by several authors (Guo *et al*., 2008; Singhal *et al*., 2014). Other studies highlight that precipitation and wind speed contribute negatively to the demand for public transport (Cools *et al*., 2010; Guo *et al*., 2008; Guo *et al*., 2007; Stover and McCormack, 2015; Singhal *et al*., 2014; Zhou *et al*., 2017; Kalkstein *et al*., 2009). Other variables as snow (Singhal *et al*., 2014) and relative humidity (Guo *et al*., 2008, Zhou *et al*., 2017) can also play an import role on the public transport demand.

The effects of weather conditions tend to vary over time. However, such impact is related with several factors as the transport mode (e.g. buses *vs.* metro) or the travel purpose (e.g. working *vs.* leisure). The season only plays a small role in the demand for public transport. Kalkstein *et al*., (2009) show that a given weather event causes similar responses in different seasons. Pendular movements are the least affected by weather conditions variations, mainly when it comes to cancelling the trip (Cools *et al*., 2010). Adverse weather conditions induce a greater variation in the demand for public transport on weekends (Li *et al*., 2018) which can be due to the purpose of the trips (mostly for leisure purpose).

Besides weather conditions, several demographic and socio-economic factors also contribute to the choice of the type of transport to be used, be it private or public. The quantification of the impacts from such factors is quite difficult though. While some of those factors can be predicted, as they have well-known relations, others have a random effect. As a result, weather conditions affect transport modes (Cools *et al*., 2010; Sabir *et al*., 2010; Singhal *et al*., 2014) and passenger types differently (Mahrsi, 2014; Hjorthol, 2013). The trip distance, the passenger gender and age, and whether or not have a driving license, are some factors that impact transport demand (Yagi and Mohammadian, 2008). The behaviour of passengers when choosing to travel or not, as well as when choosing their destinations, means of transport and routes also tend to be determined by weather conditions (Cools *et al*., 2010). This thinking - more focused on socio-economic variables – triggered the need to study the influence of weather conditions in different social strata and ages (Mahrsi, 2014). Adverse weather conditions have a greater impact on occasional travel (Cools *et al*., 2010) and a greater impact on senior passengers (over the age of 75 years) in winter compared to summer (Hjorthol, 2013). This occurs as passengers with reduced mobility have greater exposure to environmental conditions.

The literature review highlights that public transport demand can be influenced by particular weather conditions. Such influence may be higher in some transport modes and groups of the population. In a previous work, the authors of this research defined a model to analyse the impact of taking weather conditions into account in the prediction of the demand of urban bus passengers (Correia *et al*., 2020). The main goal of present work is to perform a sensitivity analysis of that model. Taking into account distinct traveller profiles, this analysis should identify the main factors that influence the prediction accuracy of the demand of bus passengers in an urban area.

Table 1. Summary of the studies that predicted the demand for public transport or/and soft modes using weather conditions

Ref.	City	Méthod	Fre-quency	Transport modes	Weather variables	Main outcomes
Tao *et al*., (2018)	Brisbane, Australia	Arimax and Sarimax	1H	B	T, R, RH, WS	Changes in weather conditions can generate fluctuations in the demand for buses
Li *et al*., (2018)	Nanjing, China	Movered average	24H	M	T, R, S	R seems to have more influence on demand than T
Tao *et al*., (2016)	Brisbane, Australia	ANOVA	24H	B	T, R, RH, WS	Metro demand is less influenced than buses demand
Zhou *et al*., (2017)	Shenzhen, China	Multivariate Regression	1H	B,M	T, R, RH, WS	The increase in RH, windshields a decrease in demand for public transport
Costa *et al*., (2017)	Porto, Portugal	Spearman correlation	1H	B,M	T, RH, WS, P	P does not seem to influence the demand for public transport against T
Creemers *et al*., (2015)	Netherland	Linear regression	1H		T, R, WS	T, R and fog seem to influence passengers decision making
Singhal *et al*., (2014)	New York, EUA	Least Square Method	1H	B,M	T, R, WS, S	R and S impact public transport demand, mainly during weekend
Stover and McCormack, (2015)	Pierce County, EUA	Least Square Method	24H	B	T, R, WS, S	Weather conditions have more influence in winter than in other periods
De Montigny *et al*., (2011)	Several cities	Quasi-Poisson regression	24H	F	T, R	T has the greatest influence on pedestrians flow
Arana *et al*., (2014)	Guipúscoa, Spain	Multivariate regression	24H	B	T, R, RH, WS	R and WS have the greatest influence on the demand for public transport at the end of the week, while mild T has an opposite impact
Schmiedeskamp and Zhao, (2016)	Seattle, EUA	Negative binomial distribution	24H	C	T, R	min(T) appears to have the greatest influence on the decision of the maximum than max(T)

Transport mode: bus (B), metro (M), bicycle (C) and on foot (F).
Weather variables: temperature (T), raining (R), relative humidity (RH), wind speed (WS), snow (S), atmospheric pressure (P).

2. Material and methods

This section describes the methodology to develop a forecasting model of public transport as function of temporal variables and weather conditions. The model should predict the total demand for urban buses, for an urban network. Several scenarios are analysed in order to select the most robust input variables of the model. Fig. 1 provides an overview of the methodology followed.

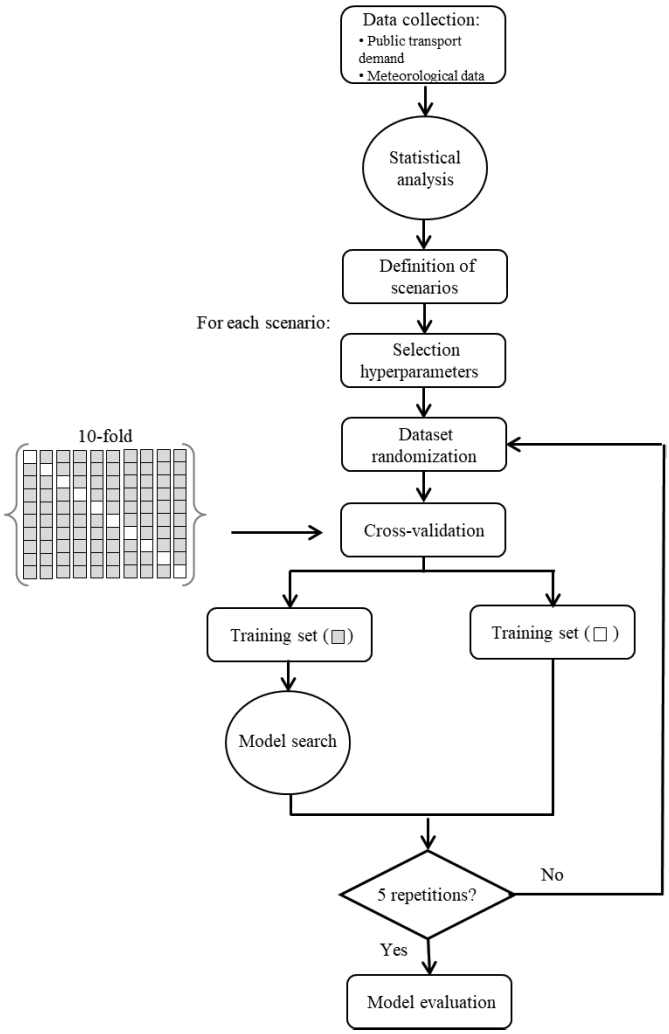


Figure 1. Methodology overview

2.1. Data collection

Data about demand for public transport and weather conditions were collected for the Metropolitan Area of Porto (AMP), for the period from 1 January 2013 to 31 December 2013.

The demand (D) for public transport was obtained from the ticket validations of the Andante system. Andante is an open zonal system, based on smartcards, that requires validation only when boarding. Each validated ticket allows to travel within a specified area and time period using different lines and transport modes, and each ticket allows for an unlimited number of trips. However, the data collected for this work refer only to trips made on buses using the main company operating in Porto (STCP, N = 77,213 million records). Each record contains information related to a trip, in particular the card ID, the stop and the service line, the service operator, the time of validation, the user's profile and the zones of access.

An automatic meteorological station managed by the Meteorological Institute (Pedras Rubras) was used to collect a set of variables that characterize weather conditions. The station is located at the airport

of Porto, which provide representative data for the whole study area. The collected weather variables were: temperature (*T*), wind speed (*WS*), relative humidity (*RH*), dew point (*DP*), atmospheric pressure (*P*), and cloudiness (*CL*). Data were collected each 30 minutes.

To obtain a uniform and a comparable set, the data were pre-processed and normalized. First, records with empty variables were removed, and then, the weather data were normalized using the z-score. This normalization is applied because the lower and upper limits are different for each of the variables, which can negatively influence the model's learning.

In addition to the bus passenger demand and weather conditions, four variables are also analysed: hour of the day (*H*), day of the week (*DS*), month (*M*), strike (*S*), and holiday (*O*). These variables are considered as they usually tend to influence the variation in demand for public transport. Table 2 summarizes the variables to be used in the analysis.

Table 2. Data characterization

Variables	Acronym	Description / Unit	Periodicity
Hour of the day	*H*	01h:00; 01h:30, ...	30 min
Day of the week	*DS*	Monday : 1; Tuesday: 2 ...	-
Month	*M*	January: 1; February: 2 ...	-
Strike	*S*	Yes: 1; No: 0	-
Holiday	*O*	Yes: 1; No: 0	-
Demand	*D*	nº of ticket validations	30 min
Temperature	*T*	ºC	30 min
Wind Speed	*WS*	km/h	30 min
Relative Humidity	*RH*	%	30 min
Dew Point	*DT*	ºC	30 min
Atmospheric Pressure	*P*	Mb	30 min
Cloudiness	*CL*	Clean; Some Clouds; Rain ...	30 min

2.2. Passengers types

Different types of passengers tend to react differently to changes in weather conditions, which may influence their predisposition to travel (Mahrsi, 2014; Hjorthol, 2013). Thus, to better understand the impacts of weather conditions on different types of passengers, they are segmented into four groups. For this purpose, only the data from travel passes validations are used (66.7% of total journeys). The different groups are created taking the age group, occupation and the existence of social supports into account:

- G1 (students): includes students aged between 4 and 23 years old (6%);
- G2 (normal): includes workers in general and validations carried out with single agents (62%);
- G3 (elderly): includes passengers over 65 years old (18%);
- G4 (social): includes passengers who are given social support (14%).

2.3. Input scenarios

In real world situations, it is usually difficult to collect a wide range of variables, as in the case study presented here. Thus, it is important to understand the impact on the model accuracy when different/reduced subset of variables are used. In order to select the variables with the greatest influence on the demand forecasting capacity, seven scenarios are defined considering different sets of inputs.

To define the scenarios, a statistical analysis of the input variables is performed. Correlation between bus passenger demand (*D*) and weather conditions variables (*T*, *WS*, *RH*, *DT*, *P* and *CL*) is analysed. The importance and direction of the association between these variables is analysed to check if the weather conditions influences the occurrence of a trip to be taken (a) immediately, (b) in the next hour or (c) in the next two hours. Variables of weather conditions without significance correlation ($p>0.05$) with demand are candidates to be discarded. Spearman's correlation (r_s) is used as the Kolmogorov-Smirnov test shows no evidence that the variables follow the normal distribution.

Table 3 presents the Spearman's correlation coefficient (r_s) between bus passenger demand and the variables of weather conditions for each group of passengers (G1, G2, G3 and G4), for a trip taken immediately and in the next two hours. Statistical significance ($p < 0.05$ or $p < 0.001$) values of correlation are appropriately marked (*5% level; **1% level). In general, low correlation is found with high statistical significance ($p < 0.001$). Time (*H*, *DS*, *M, S* and *O*), temperature (*T*), wind speed (*WS*) and

relative humidity (*RH*) are the variables with the greatest significant correlation with the bus passengers demand. On the other hand, atmospheric pressure ($r_{s\,(G1)}$ = 0.026 and $p < 0.01$; $r_{s\,(G2)}$ = -0.003; $r_{s\,(G3)}$ = 0.012; $r_{s\,(G4)}$ = 0.002) does not appear to have a direct correlation with demand. With regard to results where there is a delay of one hour (results now show on Table 3) or a delay of two hours between bus passenger demand and weather conditions, it appears that the correlations of variables with demand tend to decrease, except for the cloudiness where the correlation is significant ($r_s < 0.01$) in the Normal (G2) and Social (G4) passengers.

Based on the results provided in Table 3, seven scenarios (C1-7) are defined based on different combinations of the input variables. C1, C2 and C3 use all the variables available, while C4, C5 and C6 use the meteorological variables that have a significant correlation with the demand for all the groups of passengers (G1-4) and all temporal variables expect the month. C7 uses only time variables. C1 and C4 use the weather conditions variables measured at the same time as the trips take place, while for C2 and C5 the demand estimates are made based on the weather conditions observed in the previous hour and for C3 and C6 the demand estimates are made based on the weather conditions observed in the previous two hours. Table 4 shows a summary of the variables considered by these scenarios.

Table 3. Spearman correlation

	Demand vs. Weather conditions				***Demand vs Weather conditions from previous 2h***			
Variable	***G1***	***G2***	***G3***	***G4***	***G1***	***G2***	***G3***	***G4***
T	0.206**	0.253**	0.324**	0.253**	0.162**	0.183**	0.228**	0.176**
WS	0.249**	0.221**	0.272**	0.225**	0.230**	0.177**	0.192**	0.173**
RH	-0.305**	0.275**	-0.372**	-0.276**	0.265**	-0.206**	-0.259**	-0.197**
DT	-0.004	0.056**	0.055**	0.053**	0.039**	0.015	0.020**	0.012
P	0.026**	-0.003	0.012	0.002	0.020*	-0.011	0.008	-0.007
CL	0.009	0.018*	0.007	0.018*	0.014	0.026**	0.008	0.025**

Notes: Dark values point out significant values of correlation at *5% level ; and ** 1% level .

Table 4. Set of input variables for each scenarios

Scenarios	*Temporal variables*					Weather conditions *variables*					
	H	*DS*	*M*	*S*	*O*	*T*	*WS*	*RH*	*DT*	*P*	*CL*
C1	x	x	x	x	x	x	x	x	x	x	x
C2 (With weather conditions of previous 1h)	x	x	x	x	x	x	x	x	x	x	x
C3 (With weather conditions of previous 2h)	x	x	x	x	x	x	x	x	x	x	x
C4	x	x		x	x	x	x	x			
C5 (With weather conditions of previous 1h)	x	x		x	x	x	x	x			
C6 (With weather conditions of previous 2h)	x	x		x	x	x	x	x			
C7	x	x	x	x	x						

2.4. Modelling

Artificial neural networks (ANN) are statistical models inspired by the biology of brain networks. These models have been used successfully in several areas of knowledge, from engineering to the social sciences (Zuoan and Gasteiger, 1993; Cochocki and Unbehauen, 1993; Rowland and Vrbka, 2016).

ANNs are used due its ability to generalize and to model complex nonlinear relationships. Unlike other forecasting techniques, ANNs do not impose restrictions on input variables and are able to model even with data with heteroscedasticity.

The model was implemented in Python using the KERAS library (Chollet, 2015). Due to its modular structure, this library allows for rapid development of ANNs for deep learning, with a greater number of iterations and intuitive model creation using modules that generate neuron layers, activation functions, cost functions, initialization schemes, regularization schemes and optimizers.

2.5. Model structure

We consider a Multilayer Percepron (MLP), a feedforward artificial neural network that allows mapping a series of input variables and obtaining a set of outputs through a composition of non-linear functions. The MLP architecture can be seen as a graph composed of a series of layers, which in turn are

composed of a series of processing units (neurons). The layers between the entrances and exits are called hidden layers. For each of the defined scenarios, an ANN architecture is defined based on a base architecture that can be represented in the following form: d: n_{hid}: n_{hid} : n_{hid} : 4, where d represents the number of input variables (which depends on the complexity of each scenario defined), n_{hid} is the number of neurons in each hidden layer, and 4 is the number of output neurons to estimate the demand for each user group defined previously (G1-4). A previous analysis shows that the problem requires a complex model structure in order to predict buses demand accurately.

2.6. Learning optimization

To search for the optimal set of weights, the batch backpropagation algorithm is used.To define the ANN architectures for each scenario, first a set of preliminary experiments is carried out to define the number of training seasons (iterations of the optimization algorithm), the number of neurons in the hidden layers (n_{hid}), the number of samples per training lot (batch size) and the dropout value between layers of each ANN. Random search is used as it is more appropriate because not all hyperparameters are important for the proper functioning of ANN. This research is carried out with different ranges for each of the hyperparameters, namely:

- *Number of neurons*: ranging between 50 and 500 neurons in each hidden layer;
- *Epochs*: ranging between 100 and 600;
- *Batch size*: ranging between 8 and 100 records;
- *Dropout per layer*: ranging between 0 and 0.6 on each hidden layer.

For each set, 5 repetitions of stratified 10-fold cross-validation is perform. For each repetition, the whole dataset is randomized, and each training fold is normalized (Z-score), to have inputs with zero means and unit standard deviation.

The Leaky rectified linear activation function (Leaky ReLU) is applied. To increase, the model's ability to adjust or train from the data, a slope (a = 0.01) is considered to the negative part of the ReLU activation function domain as proposed by Maas and Ng (2013) avoiding in this way the conversion of negative values into zeros.

The optimization is carried using the algorithm HYPERAS (version 0.4.1) during 100 iterations. This algorithm evaluates several configurations at random, taking the provided parameters into account, in order to generate a good configuration of the hyperparameters. The model used 70% of the data set for training, 15% of the data set for validation and 15% of the data set for testing.

3. Results

3.1. Performance of the model

Imbalance can compromise the performance of learning algorithms that are not usually prepared for such imbalance class distribution. Also, traditionally accuracy measures are distribution dependent and do not provide a clear picture of the classifier's functionality. Several strategies have been assessed to solve (or at least minimize) the effects of class imbalance, either by adapting the traditional classifiers or by considering different measures of performance. We follow the latter approach. For that purpose, we consider three distinct measures of performance: the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the coefficient of determination (R2).

Table 5 show the hyperparameters used in each scenario, while Fig. 2 presents the obtained results. The analysis of these data shows that the results of the different performance measures for each of the models are quite similar for C1, C2 and C3, except for the RMSE value which is much lower in the C3 (when two-hour advance of the values of the weather variables are used). Observing the deviation patterns of the prediction errors, we can verify once again that the C3 is the one that achieves the best results. Also, an analysis of the scenarios using only the weather variables with the significant correlation (C4, C5 and C6), show worse results when compared to the first ones (C1, C2 and C3).

In general, the models that used all weather variables as input variables (C1, C2 and C3), performed better than when these variables are not used (C7) and when the variables used are only those that have a greater correlation with demand (C3, C4 and C5). The best results of the demand forecast are achieved when the models used the weather variables registered two hours before the trip took place (C3), as input variables. In this case, the forecast results improve, with a decrease of MAE and RMSE of 6% for total demand (comparison of C1 with C3). When all the weather conditions variables are used, the RMSE decreases by 8.2% and the MAE decreases by 8.8% (comparison of C3 with C7).

In comparison with the model for forecasting the demand for taxis developed by Mukai and Yoden (2012), it appears that the scenario C3 obtained better results, with an average percentage error of 7.29%, while the model created by Mukai and Yoden (2012) had an average percentage error of 13.59%.

The comparison of the average values of bus passenger demand of each traveller groups (G1-G4) throughout the year shows that, in percentage terms, the values of the average errors, when compared with the average demand, are: 14.72% for G1 (students), 5.49% for G2 (normal), 6.70% for G3 (elderly) and 6.30% for G4 (social). The differences between the value of the error percentage from passengers from group G1 and the remaining groups may be due, in part, to this group not having a well-defined routine, unlike the other groups that travel more at peak-hours (G2 and G4), or almost uniformly between 10h and 19h (in the case of group G3).

Table 5. Architecture, number of epochs, batch size and dropout values of each model

Scenario	Architecture	Epochs	Batch size	Dropout (1; 2; 3)
C1	11: 230 : 230 : 230 : 4	450	50	0.15; 0.2; 0.3
C2	11 : 300 : 300 : 300 : 4	350	70	0.1; 0.1; 0.1
C3	11 : 250 : 250 : 250 : 4	250	80	0.15; 0.15; 0.35
C4	7 : 240 : 240 : 240 : 4	200	80	0; 0.25; 0.2
C5	7 : 150 : 150 : 150 : 4	300	60	0.15; 0.27; 0.10
C6	6 : 200 : 200 : 200 : 4	250	70	0.05; 0.15; 0.10
C7	5 : 150 : 150 : 150 : 4	400	60	0.126; 0.12; 0.11

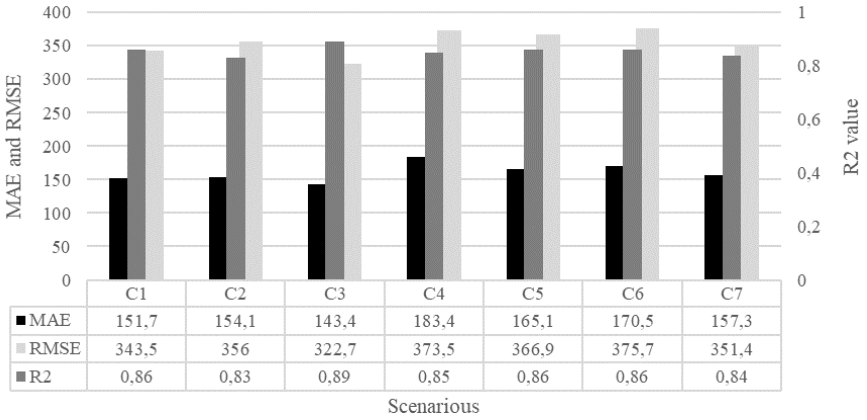


Figure 2. Performance results

3.2. Model sensitivity

To assess the sensitivity of the model, three test sets are performed in order to analyse the variability among (i) seasons (N = 200), (ii) days of the week (N = 200), and (iii) days with particular characteristics, such as holidays and strike days (N = 48). In the last test set, four days are analysed separately: (a) a random weekday, (b) a random weekend day, (c) a holiday day and (d) a strike day. For these evaluations, the ANNs model obtained for the scenario identified previously as having the better performance (scenario C3) is used. Table 6 presents an overview of the main characteristics of each test and the respective results.

Results may be quite good during the summer, as the student group (G1) travel less and the elderly group (G3) acquires more uniform mobility patterns at this period of the year.

Monday forecast got the better results on weekdays (MAE = 75.5; RMSE = 137.3). On weekends, Saturdays perform better results than Sundays. Although Saturdays have a high MAE and RMSE (MAE = 47.4; RMSE = 70.0) than Sundays (MAE = 41.8; RMSE = 64.7), when the MAE is compared with the average demand for each of those days, the average error is lower. The results for weekdays and weekends are quite good. For weekdays (Monday to Friday) we got an average percentage error of 6%, while for weekends the error is 10%. In a similar study, Mukai and Yoden (2012) obtained an error of 15.7% and 19.8% respectively for predicting taxi demand for weekdays and weekends.

In the last test set, the results show good performance, for both the weekday (Tuesday) and weekend day (Saturday) (Table 6). The analysis of dispersion charts (Fig. 3) show good results for all groups of passengers except for students (G1). Fig. 3a shows an example of demand dispersion for two

working days (Tuesday), one normal day with clean sky and a strike day with clouds. As one can observe, the lower performance of the model is obtained for the holiday and strike days. For those days, lower temperatures are observed when compared with the normal days. For the holidays, the performance is good for normal, elderly and social passengers. In these cases, an average percentage error of 9.6%, 12.7% and 10.4% are obtained respectively. The error for students is 50% higher than the one for other passengers (16.5%), which can be related with absence of data for the model learning for these day types. The worst case is obtained for the days with strikes (MAE: 164.8 and RMSE: 284.2). This may happen for three main reasons. First, during these days bus passenger's demand is directly related to the working companies. Second, the companies/drivers who work on days with strikes are very random since (not everyone joins the strike). Last, during the period in analysis (2013) only four days with strikes are recorded, which is not enough for the learning process of the model.

Table 6. Results obtained with the C3 model for (i) a random season, (ii) a random day of the week, and (iii) days with particular characteristics as holiday and strike

Periods	**Description**	**Weather conditions**			**N**	**MAE**	**RMSE**
		General	Temperature	Wind Speed			
Season	Spring	-	-	-	200	72.6	138.1
	Summer	-	-	-	200	63.4	111.3
	Autumn	-	-	-	200	68.1	121.6
	Winter	-	-	-	200	74.2	137.9
Day of the week	Mondays	-	-	-	200	75.5	137.3
	Tuesdays	-	-	-	200	78.1	137.8
	Wednesdays	-	-	-	200	80.6	154.1
	Thursdays	-	-	-	200	84.1	153.0
	Fridays	-	-	-	200	78.3	138.7
	Saturdays	-	-	-	200	47.4	70.0
	Sundays	-	-	-	200	41.8	64.7
Random days	09/07/13, Tuesday	Clean	21ºC	8 km/h	48	78.9	121.0
	28/09/13, Saturday	Rain	20ºC	26 km/h	48	44.9	66.2
	01/05/13, Wednesday (holiday)	Some clouds	12ºC	15 km/h	48	55.8	91.7
	05/03/13, Tuesday (strike)	Cloudiness	15ºC	11 km/h	48	164.8	284.2

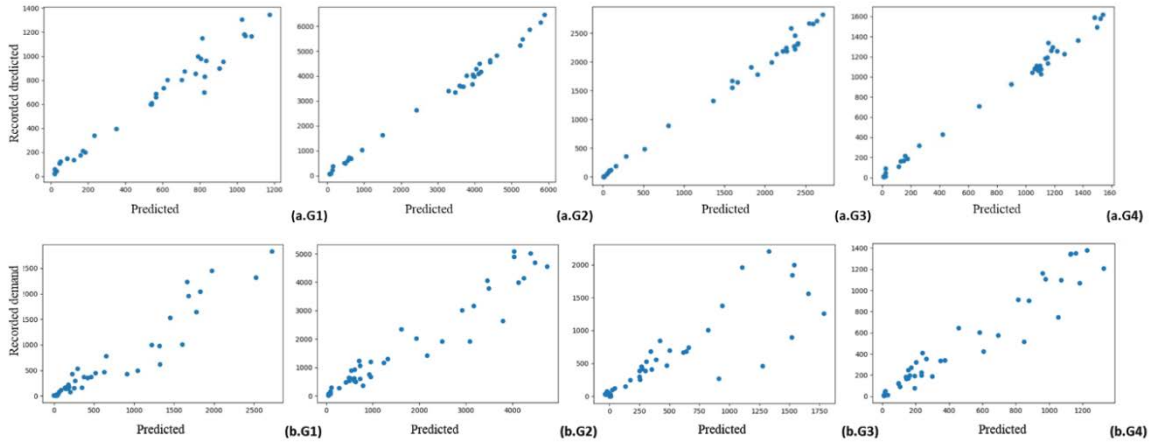


Figure 2. Recorded *vs.* predicted demand for buses by passenger group (G1-G4) using C3 model for a (a) Tuesdays normal day (09/07/13) and (b) a Tuesday with strike (05/03/13)

4. Conclusions

A multilayer perceptron with three hidden layers was applied to forecast the demand for buses in a metropolitan transport network taking the influence of weather factors into account. Demand was forecasted for four distinct groups of passengers: students (G1), workers (G2), elderly people (G3), and passengers who are given social support (G4). Seven scenarios were defined based on different combinations of the input variables. In these scenarios, the weather conditions observed in the same hour (C1, C4 and C7), in the previous hour (C2 and C4) and in the previous two hours (C3 and C6) of the demand were analysed.

It is found that the use of weather variables allows to substantially increase the forecasting capacity of the model. At the same time, it appears that when a two-hour delay in the demand is applied to the weather variables, the forecast results improve, with a decrease of MAE and RMSE of 6% for total demand (comparison of C1 with C3). However, a high complex model (with 150 to 300 hidden neurons) is required to obtain MAE ranging from 143 and 180.

The model sensitivity analysis shows that the best results of the model are achieved for normal days (i.e. days without holidays or strikes), particularly during summer weekends, while the worst performance is achieved for a weekday with strike in spring (typically Wednesdays or Thursdays).

The present work reveals that such a model can help decision makers to forecast buses demand, in particular in periods with fluctuations of weather conditions. This enables the transport providers to manage more efficiently the available resources, particularly vehicles and drivers. As the model learns from data, no impact on model accuracy is expected due to changes in the system operation.

Acknowledgements

This paper is a revised and expanded version of the paper "Forecasting of Urban Public Transport Demand Based on Weather Conditions" presented at the 5th Conference on Sustainable Urban Mobility, Greece, June 17-19, 2020. The work was financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - *Fundação para a Ciência e a Tecnologia* within project PTDC/ECI-TRA/32053/2017 - POCI-01-0145-FEDER-032053. Tânia Fontes also thanks FCT for the Post-Doctoral scholarship SFRH/BPD/109426/2015.

References

1. Tao, S., Corcoran, J., Hickman, M., Stimson, R. (2016) The influence of weather on local geographical patterns of bus usage. *Journal of Transport Geography,* 54, 66–80. DOI: 10.1016/j.jtrangeo.2016.05.009.
2. Stover, V., McCormack, E. (2015) The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation,* 15(1), 95–110. DOI: 10.5038/2375-0901.15.1.6.
3. Li, J., Li, X., Chen, D., Godding, L. (2018) Assessment of metro ridership fluctuation caused by weather conditions in Asian context: Using archived weather and ridership data in Nanjing. *Journal of Transport Geography,* 66(35), 356–368. DOI: 10.1016/j.jtrangeo.2017.10.023.
4. Zhou, M., Wang, D., Li, Q., Yue, Y., Tu, W., Cao, R. (2017) Impacts of weather on public transport ridership: Results from mining data from different sources. *Transportation Research Part C: Emerging Technologies,* 75, 17–29. DOI: 10.1016/j.trc.2016.12.001.
5. Singhal, A., Kamga, C., Yazici, A. (2014) Impact of weather on urban transit ridership. *Transportation Research Part A: Policy and Practice,* 69, 379–391. DOI:
6. Guo, Z., Wilson, N., Rahbee, A. (2008) Impact of Weather on Transit Ridership in Chicago, Illinois. Transportation Research Record: *Journal of the Transportation Research Board*, 2034, 3–10.
7. Guo, Z., Wilson, N. H., Rahbee, A. (2007) The Impact of Weather on Transit Ridership Chicago. *TRB Annual Meeting,* pages 3–10.
8. Kalkstein, A. J., Kuby, M., Gerrity, D., Clancy, J. J. (2009) An analysis of air mass effects on rail ridership in three US cities. *Journal of Transport Geography*, 17(3), 198–207. DOI: 10.1016/j.jtrangeo.2008.07.003.
9. Cools, M., Moons, E., Creemers, L., Wets, G. (2010) Changes in Travel Behavior in Response to Weather Conditions. *Transportation Research Record: Journal of the Transportation Research Board* 2157, 22–28.
10. Sabir, M., van Ommeren, J., Koetse, M. J., Rietveld, P. (2010) *Impact of weather on daily travel demand*, p.25, ID: 2060163.
11. Costa, V., Fontes, T., Borges, J., Dias, T. (2017) Impacts of Weather Conditions in Urban Public Transport: Understanding the Effects of Climatic Changes using Big Data. *Transportation Research Board*, Washington D.C.
12. Arana, P., Cabezudo, S., Peñalba, M. (2014) Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transportation Research Part A: Policy and Practice* 59, 1–12. DOI: 10.1016/j.tra.2013.10.019.
13. Kalkstein, A. J., Kuby, M., Gerrity, D., Clancy, J. J. (2009) An analysis of air mass effects on rail ridership in three US cities. *Journal of Transport Geography* 17(3), 198–207. DOI: 10.1016/j.jtrangeo.2008.07.003.
14. Yagi, S., Mohammadian, A. (2008) Policy simulation for New BRT and area pricing alternatives using an opinion survey in Jakarta. *Transportation Planning and Technology* 31(5), 589–612. DOI:10.1080/03081060802087676.

15. Mahrsi, M. K. E. (2014) A Novel Method of Multi-Information Acquisition for Electromagnetic Flow Meters. *Sensors*, 16(1), 25. DOI:10.3390/s16010025.
16. Hjorthol, R. (2013) Winter weather – an obstacle to older people's activities? *Journal of Transport Geography* 28, 186–191. DOI: 10.1016/j.jtrangeo.2012.09.003.
17. Zupan, J., Gasteiger, J. (1993) Neural Networks for Chemists: An Introduction. John Wiley Sons, Inc., USA.
18. Cochocki, A., Unbehauen, R. (1993) Neural Networks for Optimization and Signal Processing. John Wiley Sons, Inc., USA, 1st edition.
19. Rowland, Z., Vrbka, J. (2016) Using artificial neural networks for prediction of key indicators of a company in global world. In T. Klieštik, editor, Globalization and its socio-economic consequences, 16th international scientific conference proceedings, PTS I-V, pages 1896–1903, Žilina, Slovensko.
20. Chollet, F. (2015) keras. https://github.com/fchollet/keras.
21. Creemers, L., Wets, G., Cools, M. (2015) Meteorological variation in daily travel behaviour: evidence from revealed preference data from the Netherlands. *Theoretical and Applied Climatology*, 120(1-2), 183–194. DOI: 10.1007/s00704-014-1169-0.
22. de Montigny, L., Ling, R., Zacharias. J. (2011) The Effects of Weather on Walking Rates in Nine Cities. *Environment and Behavior*, 44(6), 821–840. DOI: 10.1177/ 0013916511409033.
23. Schmiedeskamp, P., Zhao, W. (2016) Estimating Daily Bicycle Counts in Seattle, Washington, from Seasonal and Weather Factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2593(1), 94–102. DOI: 10.3141/2593-12.
24. Maas, A.L., Ng. A.Y. (2013) Rectifier non linearities improve neural network acoustic models. 28. *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, JMLR: W&CP volume 28.
25. Li, J., Li, X., Chen, D. and Godding L. (2018) Assessment of metro ridership fluctuation caused by weather conditions in Asian context: Using archived weather and ridership data in Nanjing. *Journal of Transport Geography*, 66(35):356–368. DOI:10.1016/j. jtrangeo.2017.10.023.
26. Correia, R., Fontes, T., Luís Borges, J., (2020). Forecasting of urban public transport demand based on weather conditions. Advances in Mobility as a Service Systems, Springer, *Book Series Advances in Intelligent Systems and Computing*, Edited by Nathanail E.G., Adamos, G., Karakikes, I., in press.
27. Mukai N. and Yoden N. (2012) Taxi Demand Forecasting Based on Taxi Probe Data by Neural Network. In: Watanabe T., Watada J., Takahashi N., Howlett R., Jain L. (eds) *Intelligent Interactive Multimedia: Systems and Services. Smart Innovation, Systems and Technologies*, vol 14. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29934-6_57.