Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 49 (2020) 107–118

Transportation Research Procedia

www.elsevier.com/locate/procedia

47th European Transport Conference 2019, ETC 2019, 9-11 October 2019, Dublin, Ireland

Assessing multimodal mobility trends using heterogeneous data sources: a case study for supporting sustainable policy goals within the region of Algarve

Elisabete Arsenio[a,*], Susana Silva[b], Helena Pereira Iglésias[b], Alexandre Domingues[c]

[a] *LNEC, Department of Transport, Lisbon, Portugal*
[b] *Universidade de Lisboa, Faculdade de Ciências, Departamento de Estatística e Investigação Operacional, Lisboa, Portugal*
[c] *Comissão de Coordenação e Desenvolvimento Regional do Algarve, Faro, Portugal*

Abstract

This paper is built around a research project developed with the support of the Regional Planning Authority of the Algarve Region in Portugal which assessed mobility patterns covering all modes of transport using heterogeneous data sources (time-series data). Data mining techniques helped to identify limitations in some data sets. The econometric analysis showed that integrated autoregressive models and moving averages for series with seasonality were successful in the prediction of passenger flows using time-series data gathered by the regional authority from transport operators and other entities.
Results from the analysis are useful to support a strategy to reverse current trends on continued car growth (along with public transport decrease) and to devise policy measures to enable a sustainable mobility path towards decarbonisation and social equity goals until 2030.

Keywords: sustainable mobility; passenger flow prediction models ; trend analysis; multimodal mobility

1. Introduction

Assessing regional mobility trends and predictions of transport passenger demand are essential tools for the planning and management of mobility and transport. Understanding long-term traffic trends for different travel modes

* Corresponding author.
Email address: earsenio@lnec.pt

10.1016/j.trpro.2020.09.010

(e.g. individual transport *versus* collective transport) helps to add knowledge on how each mode is expected to evolve and anticipate future states of the system. This will allow policy makers to tackle possible deviations from targets set to achieve sustainable mobility goals.

In some travel models such as aviation, accurate short-term passengers' flows predictions are needed for financial decisions such as those related to revenue management and regular adjustments to operational variables (e.g. ticket pricing policy). As earlier noted for the case of short-term rail passenger flow forecasting (Roos et al., 2017; Haworth, 2014) there's no consensus on the most suitable method to select and this is also influenced by the characteristics of the dataset in each territorial context.

Emerging developments in big data, sensor and wireless communications present nowadays new challenges and opportunities for travel demand forecasting methods, particularly for dealing with the dimensions of data acquisition, data processing, data analysis, and application of results (Zhao et al., 2018). However, advances in the applications of big data in the field of sustainable and multimodal mobility still require solving some constraints such as missing data (at the level of disaggregation aimed), discontinuities in traffic-related data acquisition, absence of comprehensive databases, lack of data integration and quality issues, which may apply in each specific case.

The research reported in this paper is considered opportune in the context of the region of the Algarve in Portugal. The region had developed successfully an "Action Plan for Sustainable Urban Mobility" engaging the regional network of sixteen cities and stakeholders with a focus on mobility decarbonisation goals (AMAL, 2017; Arsenio and Coelho, 2019).

This paper is built around of a joint research effort of the authors for the analysis of the passenger flows' data (time-series data) gathered by the Regional Coordination and Development Commission of the Algarve (CCDR Algarve) to assess mobility patterns covering all transport modes of transport. The main research questions can be described as follows:

- What mobility patterns are observed in the various transport modes in the Algarve region?
- What trends can be expected in the coming years?
- According to the passenger flow predicting models and their results, what recommendations can be useful for mobility and transport planning authorities for action until 2020 and beyond?

The remainder of this paper is organised as follows: Section 2 provides an overview of the main passenger flow predictions models: it sets the research scene background and outlines the various types of models found in the literature over the years. Section 3 presents the data and the selected modeling methodology. Section 4 addresses the prediction models estimated and presents some of the regional models estimated for collective transport modes (intercity bus and regional rail) and the region's main motorway. Finally, section 5 outlines conclusions and further research directions. In this paper, the terms "prediction" and "forecasting" are used interchangeably.

2. Overview of passenger flow forecasting models and modeling trends

Various approaches have been developed to predict passenger flows over the past few decades (Liu and Cheng, 2017; Qin et al., 2019). The most widely used forecasting models are regression models, smoothing models and general time series models (Montgomery et al., 2015).

Several model taxonomies were found in the literature review. Liu and Cheng (2017) classified passenger flow prediction models into main four categories:

- Models based on traditional classical algorithms (e.g. Kalman filtering based methods);
- Regression models;
- Machine learning based models (e.g. back propagation learning trained artificial neural networks);
- Hybrid models (combination of methods).

Bai et al. (2017) classified short-term transport forecasting approaches into two main categories: parametric and non-parametric methods. Parametric methods include models such as the autoregressive model, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA). The ARIMA model is indicated for non-seasonal and non-stationary data. Box and Jenkins (1970) have generalised the ARIMA model to deal with seasonality. Non-parametric models include alternatives such as artificial neural networks (ANN) and deep learning approaches such as deep belief networks. Focusing on short-term bus passenger flow prediction models, Zhai et al. (2018) classified applied methods into three main categories:

- Linear methods of time series analysis and regression;
- Nonlinear methods (e.g. artificial neural networks, support vector machine);
- Combination of linear and non-linear methods.

Non-linear models often require a large amount of traffic monitoring data preprocessing and show poor real-time nature (Zhao et al., 2018).

Ahmed and Cook (1979) conducted one of the former applications of Autoregressive Integrated Moving Average (ARIMA) model for short-term prediction of motorway traffic flows. The estimated ARIMA models were found to be more accurate to represent time-series data (166 data sets from three surveillance systems were used) in motorways of Los Angeles, Minneapolis, and Detroit in comparison to these alternative methods: moving-average, double exponential smoothing and trig and leach adaptive models, in terms of mean absolute error and mean square errors. Ghosh et al. (2005) found that seasonal ARIMA model was adequate to predict traffic flows during rush hours at a city junction. Moreira-Matias et al. (2013) used ARIMA models for predicting the spatial distribution of taxi passengers using streaming data. Guo and Yuan (2017) used ARIMA models and applied genetic optimization particle filter to improve the accuracy of short-term passenger flow predictions at each station for the case of urban rail.

Zhang and Song (2011) predicted passenger flows at key bus stops using the Kalman filter algorithm and data from automatic fare collection and video. As noted by Zhai et al. (2018), previous studies demonstrated that the Kalman filter method lead to better performance results (in terms of mean absolute deviation, mean square error, mean absolute percentage error, mean square percentage error) in comparison to the Back Propagation Artificial Neural Network. Hou and Ma (2011) used combined grey models (Deng, 1989) with linear regression to forecast railway passenger flows. Gutiérrez et al. (2011) used distance-decay functions and multiple regression models to forecast passengers' metro ridership at the station level. Distance-decay weighted regression means that demand data from the bands nearer the stations have a greater weighting than those located farther away.

Li et al. (2016) used the Back Propagation Neural Network model (BPNN) to predict urban rail passengers' traffic at the station level and used the population per distance band as new variable. These authors showed that the proposed BPNN had a better performance in comparison to the other three based (linear model based on population per distance band, BPMM model with total population, linear model based on total population) for these measures of effectiveness: maximum relative error, smallest relative error, average relative error, and mean square root of relative error. Regression forecasting models and time series prediction models have been much used to estimate future urban passenger traffic flows but also cover the interurban context. Lv et al. (2015) predicted short-term passenger traffic flows covering the case of long-distance coach as well. Ma et al. (2015) proposed the Interactive Multiple Model-based Pattern Hybrid (IMMPH) approach to predict short-term demand bus passenger and this method had a better forecast performance in comparison to an artificial neural network hybrid model.

Jian et al. (2014) used a hybrid approach involving the combination of two methods for short-term forecasting of high-speed rail demand: grey support vector machine and the ensemble empirical mode decomposition. This hybrid approach outperformed ARIMA and Support Vector Machine methods for the modelled direct origin-destination (OD) flow data but since interactions between different OD pairs had been ignored this may affect its forecasting ability. Xie et al. (2014) proposed a hybrid approach for the short-term forecast of air passenger flows which was based on the seasonal decomposition of time series and least squares support vector regression model.

Yu et al. (2011) compared several machine learning methods (Support Vector Machine, Artificial Neural Networks-ANN, k-Nearest Neighbor algorithms) and linear regression for prediction of bus time arrivals at bus stops with multiple routes. These authors found that the SVM model had the best performance of the four models.

Tsai et al. (2009) used multiple temporal units neural network models and parallel ensemble neural networks and four-year daily data to estimate short-term railway passenger demand. Sun et al. (2019) applied a mean impact value-based non-linear vector auto-regression neural network for forecasting air passenger flow. This method outperformed other hybrid approaches in terms of accuracy. Xu et al. (2019) proposed a hybrid approach of SARIMA-Support Vector Regression to forecast aviation demand indicators.

Deep learning was used by Liu and Chen (2017) to predict the hourly passenger flows for the BRT mode. Liu et al. (2019) used deep learning methods to forecast the inbound/outbound passenger flows for the metro. The model was trained using the Back propagation and Adaptative moment estimation (Adam) method (Kingma and Ba, 2014). Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. Bai et al. (2017) found that the multi-pattern deep fusion approach (by fusion of deep belief networks of multiple patterns) used for predicting short-term bus passenger flows had the best forecasting performance among the alternative methods. Table 1 summarizes the studies/methods above described.

Table 1. Overview of passenger flow prediction models by transport mode.

Studies	Transport mode	Type of model
Li and Yang (2007)	Railway	Neural network
Tsai et al. (2009)	Railway	Neural network
	Railway	Grey models and linear regression
Hou and Ma (2011)	High-speed rail	Hybrid (SVM and EEM decomposition
Jian et al. (2014)	Railway	SARIMA
Milenković et al. (2016)	Railway	Neural network
Guo, L. and Yang, Y. (2017)	Urban rail	ARIMA &genetic optimization particle filter
Li, J. et al. (2016)	Urban rail/metro	Back Propagation Neural Network
Liu et al. (2019)	Metro	Deep Learning
Gutiérrez et al. (2011)	Metro	Multiple regression
Liu and Chen (2017)	BRT	Deep learning
Bai Y. et al. (2017)	Bus	Multi-pattern deep fusion/Deep Belief Networks
Zhang and Song (2014)	Bus	Kalman filter
Ma et al. (2014)	Bus	Interactive Multiple Model-based Pattern Hybrid
Yu et al. (2011)	Bus	Support Vector Machine, ANN, k-NN, linear regr.
Lv et al. (2015)	Long-distance Coach	Regression
Moreira-Matias et al. (2013)	Taxi	ARIMA
Ghosh et al. (2005)	Traffic flows/junction	Seasonal ARIMA, Holt-Winters
Sun et al. (2019)	Aviation	Mean impact value-based non-linear vector auto-regression neural network
Xie et al. (2014)	Aviation	Hybrid (seasonal decomposition and Least Squares Support Vector Regression)
Xu et al (2019)	Aviation	SARIMA-Support Vector Regression

Table I shows that parametric methods such as ARIMA and SARIMA have been applied across several transport modes in the context of passengers' flow predictions. On the other hand, deep learning approaches to artificial intelligence seem to be promising in big data, using the principles of learning multiple levels of decomposition, which can be applied in machine learning frameworks (Goodfellow et al., 2016). As already noted by Liu and Chen (2017), these methods are at the intersection of research into neural networks, artificial intelligence, graphical modeling, optimization, pattern recognition, and signal processing.

3. Data and modeling framework

3.1. Data description

The data supplied by CCDR Algarve is raw data of passenger's flows covering all transport modes and time periods (day, month) which was collected from heterogeneous sources and covered the period 2007 to 2017. Basically, it comprised datasets (time-series) of the transport operators in the region and the national manager of road and rail infrastructures for the case of daily road traffic flows in main roads.

As earlier noted by Rahm and Do (2000), major problems exist in raw traffic data such as referential integrity, uniqueness, misspelling, redundancy and contradictory values. Therefore, the preprocessing of data involved data cleaning for detecting and removing errors and inconsistencies from data to improve its quality. The approach involved data analysis (including data mining to detect specific patterns and outliers), mapping rules, verification, transformation and backflow of cleaned data.

In the research reported herein the focus is made on prediction models estimated for the case of regional collective transport (bus and rail) and road traffic only. Considering data privacy issues that apply to each transport operator in the region, the data is aggregated by trimester to be able to publicly disseminate prediction models for each mode.

3.2. Modelling framework

A time series comprises a set of observations statistically related, associated to a specific random event and that are sequentially collected over time (Brockwell and Davis, 2006). A time series function can be disaggregated into its trend, seasonal, cyclic and residuals' parts (Murteira, 2000).

The time series analysis aims to describe, model, predict and control future behaviour (Ehlers, 2009) as follows:

- Description: The graphical analysis of the series allows the identification of trend patterns, seasonal or cyclic variations and abrupt change in behaviour, outliers and turning points. The analysis of the time series requires the selection of a suitable probability model for the data.
- Modelling: After knowing the behaviour of the series, models need to be built to explain the data over the observed period.
- Prediction: Future values of a series can be predicted based on past values and models built to explain behaviour.
- Control: Future behaviour can be controlled through adjusting the values of some variables.

The methodology used for prediction analysis is based on Box-Jenkins (Box et al., 2015) and followed these main steps using the R[†] software (R, 2016):

- Graphical analysis of the time-series.
- Verify the stationarity (or non-stationarity) of the time-series, that is if these exhibits similar "statistical behavior" in time; the unit root test, test of stationarity by Dickey and Fuller (1979) is used.
- Parameter estimation of the identified models to fit the data; model identification involves checking the graphical analysis, autocorrelations functions (ACF) and partial autocorrelation functions (PACF) and Akaike information criteria (AIC); for example, the identification of ARIMA models would require that both ACF and PACF will exhibit exponential decay and/or damped sinusoid behaviour.
- Diagnosis assessment, that is to analysis the adequacy of the model through residual analysis.
- Model reformulation, if required as a result of the previous step.
- Predict of future observations with the selected model: predictions are envisaged to the year 2020 by the regional planning authority.

Annual passengers' flows time series exhibited strong seasonal patterns. This was somehow expected as the region of Algarve is the main touristic destination at the national level (and internally known for its long coast and seaside) and, hence, demand for services are intrinsically seasonal having its peak over the summer period.

[1] The R package is available at: https://cran.r-project.org

3.3. Seasonal ARIMA and ARIMA models

As mentioned by Xu et al. (2019), the SARIMA model consists of six parts: Autoregressive (AR), Moving Average (MA), Seasonal Autoregressive (SAR), Seasonal Integration (SI) and Seasonal Moving Average (SMA). A time-series Y_t is a Autoregressive process with p-order, AR (p), if:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t = \mu + \sum_{I=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t \quad (1)$$

$$\varepsilon_t \sim WN\ (0, \sigma_\epsilon^2) \quad (2)$$

Where $\varphi_1,\ldots, \varphi_p$ are non-null fixed constants (autoregression parameters); p is the time lag and ε_t is the white noise (sequence of uncorrelated random observations from the same distribution, with mean μ and constant variance σ_ϵ^2), in time *t*. Considering the lag operator B, where $BY_t = Y_{t-1}$, equation (1) is equivalent to:

$$\varphi(B)Y_t = \mu + \varepsilon_t, \text{ where (B)} = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \quad (3)$$

Hence, (B) =0 defines an autoregressive model. B is, by convention, a real or complex number. The process AR (p) is stationary if and only if the roots of the characteristic equation are higher or equal to 1 (Cowpertwait & Metcalfe, 2009). An AR(p) model can be characterized by:

- an autocorrelation function, ACF, with infinite and exponential decay after p lags;
- a partial autocorrelation function, PACF, equal to zero for time lags greater to lag p.

A time-series Y_t is a Autoregressive process of p-order and moving averages, ARMA (p,q), if:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t+1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4)$$

$$Y_t = \mu + \sum_{I=1}^{q} \varphi_i Y_{t-i} + \sum_{I=o}^{q} \theta_i \varepsilon_{t-i} \quad (5)$$

$$\varphi(B)Y_t = \mu + \theta(B)\varepsilon_t \quad (6)$$

Where μ is the average of the process and $\theta_1, \ldots ,_q$ are fixed constants and different from zero. The process of Moving Averages MA (q) has:

- an autocorrelation function, ACF, equal to zero for time lags higher than lag q;
- a partial autocorrelation function, PACF, infinite and with exponential decay after lag q.

The autocorrelation function of the ARMA (p, q) model combines the behavior of models AR(p) and MA(q), but if time lags considered are below q, then the ACF is equal to the AR(p) model.

The Autoregressive Integrated of Moving Averages model, ARIMA (p,d,q) is a generalization of the ARMA models, to integrate non-stationary time-series (the parameter d represents the d^{th} differencing in process to achieve stationarity). SARIMA models are represented by ARIMA (p,d,q) $(P,Q,D)_m$ where m represents the number of seasons (e.g. four trimesters) and the parameters P,Q,D have the same meaning as p, d and q in the ARIMA model and correspond now to the seasonal part of the model. The SARIMA model uses season differencing of appropriate order to remove non-stationarity of the series (Xu et al., 2019).

For the selection of the forecasting model, it was used the Akaike information criterion (Akaike, 1974). The Ljung-Box test was applied to verify the existence of autocorrelations between the former B lags of the time-series (function Box.test if the R package). The Shapiro-Wilk test (Shapiro and Wilk, 1965) was used to test the normality of residuals (function shapiro.test in the R package).

4. Development of regional prediction models: collective transport versus car-based trends

4.1. Collective transport: intercity bus

Figure 1 shows a continuous decreasing of the total annual passenger traffic flows over the period 2007 to 2017. Between 2007 and 2014, 3.16 million of passengers were lost. After 2014, intercity annual bus demand remains approximately constant.

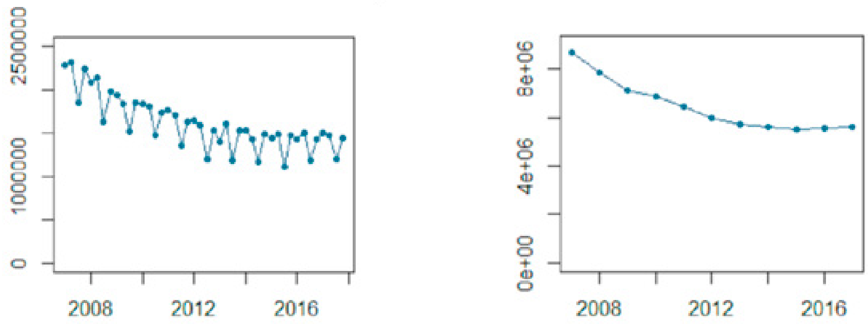

Fig. 1 Annual passenger traffic flows in intercity buses by trimester (left) and year (right).

The lowest passenger flows are registered in the 3rd trimester of the year (summer period). Unlike the other collective transport modes (railway, inland waterways transport), intercity bus is mainly used by commuters and to home-school travel. The Dickey-Fuller stationarity test (p-value = 0.7477) does not reject the null hypothesis of non-stationarity of the time-series and it's necessary to apply a non-seasonal differencing, d=1, to achieve a stationary series (Figure 2).

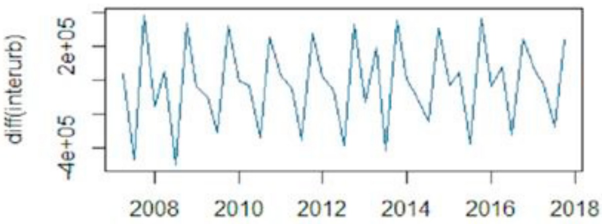

Fig. 2 Time-series differencing (d=1).

The ACF and PACF functions of the time-series free of seasonality and trends are represented in Figure 3. These are used to determine the other parameters of the model.

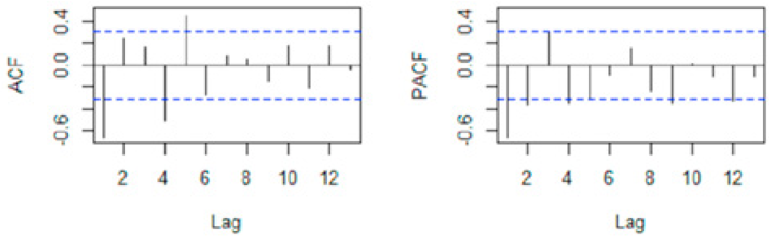

Fig. 3 ACF (left) and PACF (right).

The peaks in lags 1 and 4 of the ACF (Fig.3) suggest a non-seasonal MA component of order 1 (q=1), along with a seasonal of order 1 as well, Q=1. The peaks in lags 1 and 2 of the PACF (Fig. 3) suggest a non-seasonal AR of order 2 (p=2), whereas the peak in lag 4 suggest a seasonal component of order 1 (P=1). Therefore, two possible best fit models are compared as follows:

- ARIMA (0,1,1) $(0,11)_4$ with AIC = 983.45
- ARIMA (2,1,0) $(1,1,0)_4$ with AIC= 980.73 (model selected)

For the selected model, the non-seasonal AR parameter of order 2 (p=2) is not significant and it's eliminated which lead to the following model which is used in Figure 4:

- ARIMA (1,1,0) $(1,1,0)_4$ with AIC = 979.53

The residuals of the final model are not correlated (Box-Ljung test, p-value = 0.334) and these follow approximately a normal distribution (Shapiro-Wilk, -p=0.6339).

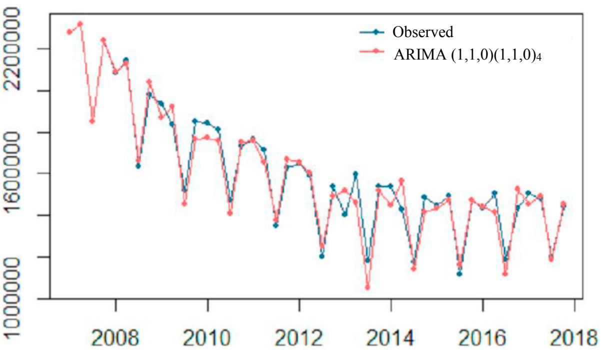

Fig. 4 Observed versus adjusted observations with the selected prediction model.

The predicted annual passenger flows that are expected to use the intercity bus will remain approximately constant until 2020 (Figure 5). In 2020, the highest passenger flow is registered in the 1st trimester, with a total of 1.522554 passengers and the lowest in the 3rd trimester, 1.230955 passengers.

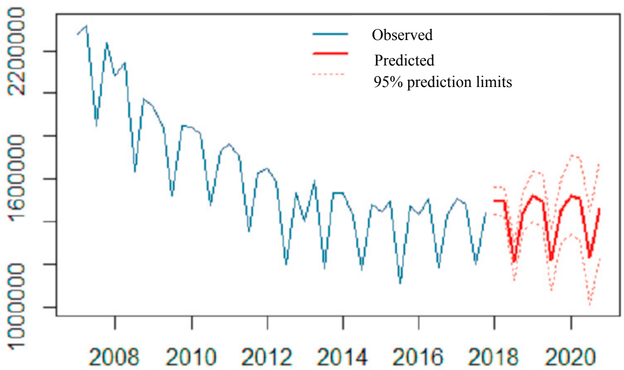

Fig. 5 Intercity bus passenger flows' predictions until 2020.

4.2. Collective transport: regional railway

Figure 6 shows a decreasing trend in the use of regional railways between 2010 and 2012, with a loss of 358.7 thousand passengers. After 2012, passenger annual flows started to increase reaching in 2017 the total of around 1.9 million passengers, a similar value to the one observed in 2010.

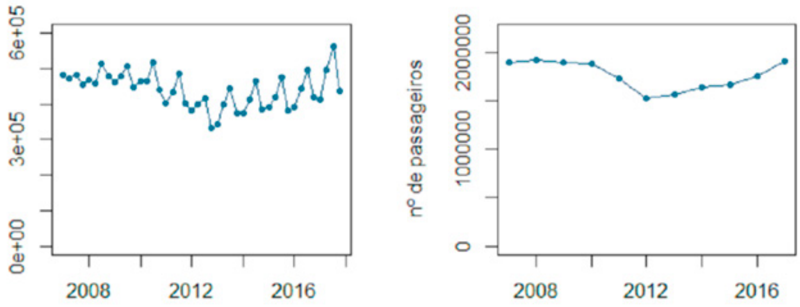

Fig. 6 Regional railway passenger flows per trimester (left) and year (right).

Following the same modelling framework as described in section 4.1, the best fit to the data model can be represented as ARIMA (0,1,0) $(0,1,1)_4$. The forecasted railway passenger flows until 2020 are shown in Figure 7.

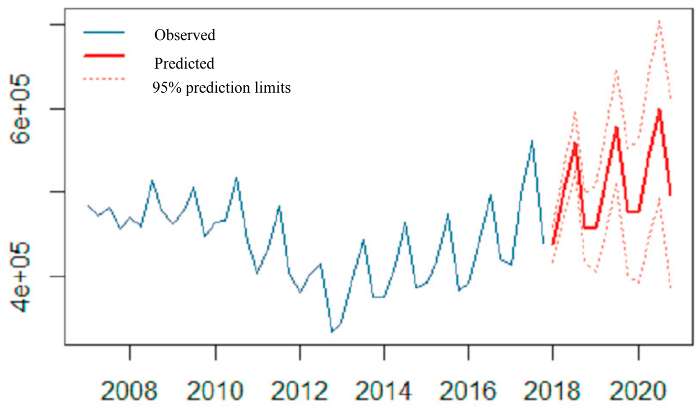

Fig. 7 Regional railways passenger flows' predictions for 2018 – 2020.

In the period up to 2020, the total number of passengers that will use the regional railway is expected to increase, registering in the 3rd semester its maximum with 598578 passengers which coincides with the period of higher touristic flows in the region. The amplitude of the prediction limits over time influences the accuracy of the forecasts (Fig. 7).

4.3. Traffic flows in the main regional motorway A22

Figure 8 shows the Annual Daily Traffic (ADT) flows in the A22 regional motorway: between 2007 and 2012 it registered a decreasing trend; after the 2014 value (6000 vehicles/day), the ADT continued to increase until 2017 but totals are inferior to the ones registered in 2008.

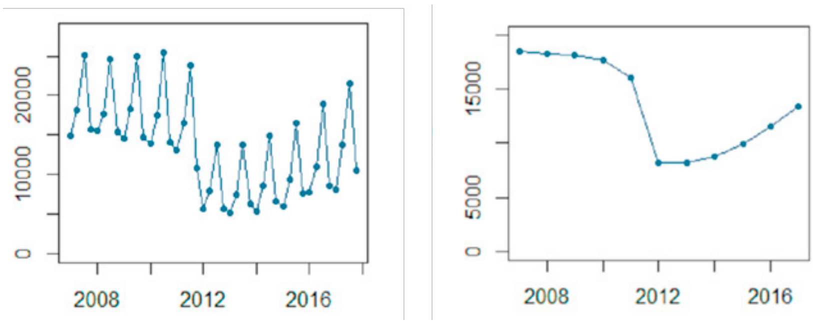

Fig. 8 ADT flows in the A22 regional motorway, per trimester (left) and year (right).

Following the same modelling framework as described in section 4.1, the best fit model to the data was ARIMA (0,1,0) $(0,1,0)_4$. The predicted ADT flows in the A22 motorway until 2020 are shown in Figure 9.

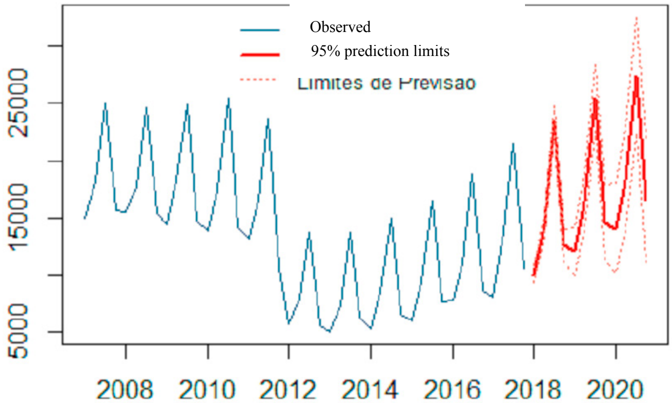

Fig.9 Predicted ADT flows in the A22 regional motorway until 2020.

The ADT flows in the regional motorwayA22 are expected to increase until 2020. Traffic flows are expected to reach the previous values registered in 2007, before the economic and financial crisis. ADT flows maximum are expected to occur in the 3rd trimester of 2020 with 27451 vehicles/day.

4.4. Discussion of results: policy implications

Road traffic trends and the estimated model predictions of regional collective transport are useful for policy makers. The Algarve territory is aligned with the concept of a polycentric region (Schuthof, 2013). It is characterized by the presence of several centres in which goods and services are dispersed in different cities of the region and there exists important shares of daily mobility flows that occur outside the boundaries of a single city, due to daily commuting patterns and additional seasonal interactions generated by intense tourism flows during summer (Arsenio and Coelho, 2019). Around 65% of the population uses car as main mode of daily transport in the region and it's important for planning authorities to reverse trends. Traffic growth can be explained by several deterministic factors such as the absence of public investment in the region of Algarve for creating effective alternatives to car use (intercity bus and regional railway supply remain unchangeable since 2007), significant internal and external touristic demand, specially over the summer period and high supply levels of rent-a-car services in the Faro region (including at Faro airport).

Considering that predictions of railway passenger flows are expected to increase until 2020 with demand peaks over the 3rd trimester, more investment seems to be required to meet future demand. Regarding the intercity bus, this mode is mainly used for commuting (e.g. home to school) and the demand levels are expected to remain almost unchangeable until 2020. Therefore, it seems important to improve quality of service and its connectivity to other modes such a cycling and walking to make this mode more attractive than car. These actions would also help to meet the decarbonisation goals set in the Action Plan for Sustainable Urban Mobility.

5. Conclusions and further research

The research reported in this paper represents the former collaborative effort to transform the raw data of the CCDR Algarve concerning passenger flows (time-series) in the region of Algarve, in Portugal, into useful information for mobility planning purposes. Results from the analysis are useful for decision makers to reverse current trends on continued road traffic growth and to devise policy measures to enable a sustainable mobility path towards decarbonisation and social equity goals until 2030, as envisaged by the "Action Plan for Sustainable Mobility" in the Region.

The parametric passenger flow prediction models estimated represent main mobility patterns at the regional level. However, these models are unable to react to varying conditions and strategies. Once real-time data becomes available for other modes such as cycling, future research shall proceed with the development of passengers' flow prediction models at the local (city) level aiming to use more disaggregated historical data by time of day and explore emerging methods such as deep learning.

References

Ahmed, M.S., Cook, A.R. (1979). Analysis of freeway traffic time-series data by using Box-Jenkins techniques, *Transp. Research Record: Journal of the Transp. Res. Board* 722, 1–9.

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Trans. Autom. Control* 19, 716-723.

AMAL (2017) Plano de Ação para a Mobilidade Urbana no Algarve/PAMUS Algarve (Action Plan for Sustainable Urban Mobility in the Algarve), VAMUS Compromisso para a Mobilidade Urbana Sustentável no Algarve, Comunidade Intermunicipal do Algarve, Faro.

Arsenio, E. and Coelho, J. (2019) The role of sustainable urban mobility plans and ICT to improve public space for social inclusion, *in* Menezes, M.&Smaniotto, C. *Neighbourhood & City. Between digital and analogue perspectives*, COST European Cooperation in Science and Technology and Edições Universitárias Lusófonas, ISBN 978-989-757-077-3.

Bai, Y., Sun, Z., Zeng, B., Deng, J., Li,C. (2017). A multi-pattern deep fusion model for short-ter, bus passenger flow forecasting, *Applied Soft Computing 5*8, 669-680.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Brockwell, P.J. and Davis, R.A. (2006). Introduction to Time Series and Forecasting. 3rd edition, Springer.

CCDR Algarve - Comissão de Coordenação e Desenvolvimento Regional do Algarve (2019). [Online:] https://www.ccdr-alg.pt/site/ [Accessed: 8th July 2019].

Copertwait, P., Metcalfe, A. (2009). Introductory Time Series with R, Springer New York.

Deng, J. (1982). Control problems of grey system. *Systems & Control Letters*, 1, 288–294.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimates for autoregressive time series with a unit root, *Journal of the American Stat. Association* 74, 427-431.

Ehlers, R.S. (2009). Análise de Séries Temporais, Universidade de São Paulo.

Ghosh, B., Basu, B. and O'Mahony, M. (2005). "Time-series modelling for forecasting vehicular traffic flow in Dublin." 84th Annual Meeting of the Transportation Research Board, Washington, DC.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

Guo,L. and Yuan, Y. (2017). Forecast method of short-term passenger flow on urban rail transit, Association for Computing Machinery ISBN 978-1-4503-5366-3/17/12. https://doi.org/10.1145/3171592.3171633

Gutiérrez, J., Cardozo, O.D., García-Palomares, J.C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression, *Journal of Transport Geography*, vol. 19 (6), pp 1081-1092.

Haworth, J. (2014) Spatio-temporal forecasting of network data. Ph.D. thesis, University College London.

Hou, L. M., & Ma, G. F. (2011). Forecast of Railway Passenger Traffic Based on a Grey Linear Regression Combined Model [J]. *Computer Simulation*, 7.

Jiang, X., Zhang, L., Chen, X. (2014). Short-term forecasting of high-speed rail demand: a hybrid approach combining ensemble empirical model decomposition and gray support vector machine with real-world application in China, *Transportation Research Part C* 44, 110-127.

Kingma, D.P., Ba, J. (2014). Adam: A Method for Stochastic Optimization. Published as a conference paper at ICLR 2015. https://arxiv.org/pdf/1412.6980.pdf

Li, J., Yao, M., Fu, Q. (2016). Forecasting method for urban rail transit ridership at station level using back propagation neural network, *Discrete Dynamics in Nature and Society*, vol 2016. http://dx.doi.org/10.1155/2016/9527584

Liu,Y., Liu, Z., Jia, R. (2019). DeepPF: A deep learning based architecture for metro passenger flow prediction, *Transportation Research Part C* 101 (2019) 18–34

Liu, L. and Chen, R. (2017), A novel passenger flow prediction model using deep learning methods, *Transportation Research Part C* 84, 74-91.

Lv, Z., Li, X., Hu, J., Yin, L., Zhang, B., Feng, S. (2015). Virtual geographic environment based coach passenger flow forecasting, In: Proc. IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, pp. 1–6.

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.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh,C, Byers, A.H. (2011). Big data: the next frontier for innovation, competition and productivity.

Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., & Avramović, Z. (2018). SARIMA modelling approach for railway passenger flow forecasting. Transport, 33(5), 1113-1120.

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting. John Wiley & Sons.

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), 1393-1402.

Murteira, B., Muller, D., & Turkman, K. (2000). Análise de successões cronológicas, McGrawHill.

R Core Team (2016) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna.

Rahm, E. and Do, H. H. (2000). Data cleaning: Problems and current approaches, IEEE Data Eng. Bull., vol. 23, no. 4, pp. 3–13.

Roos, J. Gavin, G. and Bonnevay, S. (2017) A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. *Transportation Research Procedia* 26, 53-61.

Shapiro, S.S., Wilk, M.B. (1965). An analysis of variance test for normality, *Biometrika*.

Qin, L., Li, W., Li, S. (2019). Effective passenger flow forecasting using STL and ESN based on two improvement strategies, Neurocomputing 356, 244-256.

Sun, S., Lu, H., Tsui, K., Wang, S. (2019). Nonlinear vector auto-regression neural network for forecasting air passenger flow, *Journal of Air Transport Management* 78, 54-62.

Schuthof, R. (2013). Benefit from the Poly-SUMP Methodology. A proven methodology for sustainable mobility in complex regions, Freiburg: ICLEI Europe.

Tsai, T., Lee, C., Wei,C. (2009). Neural network based temporal feature models for short-term railway passenger demand forecasting, *Expert Systems and Application* 36 3728-3736.

Wilhelm, E., Rodgers, L., Bornatico, R., and Soh, G. (2013). Towards Real-Time Identification of Electric Vehicle Mass, SAE Technical Paper 2013-01-0063, https://doi.org/10.4271/2013-01-0063.

Xie,G., Wang, S., Lai, K.K. (2014). Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches, *Journal of Air Transport Management* 37, 20-26.

Xu, S., Chan, H. K., & Zhang, T. (2019). Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. Transportation Research Part E: Logistics and Transportation Review, 122, 169-180.

Zhai, H., Cui, L., Nie, Y., Xu, X. and Zhang, W. (2018). A comprehensive comparative analysis of the basic theory of the short term bus passenger flow prediction, *Symmetry*, 10, 369.

Zhao, Y., Zhang, H., An, L., Liu, Q. (2018). Improving the approaches of traffic demand forecasting in the big data era, Cities 82, 19-26.

Zheng, Y., Capra, L., Wolfson,O., Yang, Y. (2014). Urban Computing: concepts, methodologies and applications, *ACM Transactions on Intelligent Systems and Technology* (TIST) 5, no. 3 (2014): 38.

Yu, B., Lam, W.H.K., Tam, M. L. (2011). Bus arrival time prediction at bus stop with multiple routes, *Transportation Research part C* 19, 1157-1170.