Short-term prediction of on-board passenger volumes in public transport vehicles based on non-exhaustive Automatic Passenger Counting data

Zoi Christoforou [1], Ektoras Chandakas [2], Ioannis Kaparias [3],

[1] Ecole des Ponts - ParisTech, France
[2] Transamo, France
[3] University of Southampton, UK

Introduction

On-board passenger volumes are critical for both public transport operators and passengers. They play an important role in schedule adherence and are also a key determinant of the quality of service. Up until recently, however, due to the absence of the relevant enabling technology, data on on-board passenger volumes was only available from the conduct of exhaustive manual passenger counts. These would be carried out infrequently and would provide data of questionable accuracy, as they would only give a snapshot of the condition at the time of the survey. As a result, the key issue of estimating on-board volumes in the short-term has received only little attention in the literature.

Recent technological advances have resulted in the development and implementation of Automatic Passenger Counting (APC) systems. These make use of a system architecture of sensors and actuators on board the public transport vehicle, thus providing much more accurate estimates of the vehicle on-board loads. This allows operators to gain an insight into the number of passengers boarding and alighting (B&A) at each station, and therefore assists them in both long-term planning, but also, more crucially, in real-time operations.

APC systems are therefore increasingly being installed on public transport vehicles in various cities around the world. However, the cost of each APC installation is high, and as a result not all vehicles are equipped. The common practice of public transport operators in France, for example, is that only 10-20% of the total fleet is equipped. This means that only partial knowledge on the loadings can be obtained, which has so far not allowed for short-term (real-time) loading predictions to be carried out.

The aim of the present study is, hence, to propose a novel methodology for real-time short-term prediction of on-board passengers, based on non-exhaustive APC data. The prediction is to be achieved per vehicle run, line stop, and 15-min time interval of a typical working day, in order to be suitable for use as input by operators in planning and real-time management processes. The proposed methodology is then tested and validated using an actual APC dataset covering a period of six months.

Study area and dataset

The present study focuses on the tramway system of the French city of Nantes, which consists of three lines (numbered 1, 2 and 3) running in 44 km of track and serving a total of 83 stations. The network is shown in Figure 1. Line 1, shown in green, has a length of 18.4 km and serves 34 stations. It consists of two branches at each end and a central trunk between the branches with 19 stations. Its frequency reaches 15 vehicles per hour during peak times, and it is the busiest line on the network (and with 120,000 passengers per day, it is also one of the busiest of the whole of France), serving several principal locations of the city, including the city's stadium and the main railway station. Line 2, shown in red, has a length of 11.7 km and serves 25 stations, including important educational (university) and health establishments. It has a frequency of 8 vehicles per hour during peak times, and its patronage approaches roughly 80,000 passengers per day. Lastly, Line 3, shown in blue, has a length of 14.1 km and serves 34 stations. It has a similar operation with Line 2, with which it shares the track for seven stations in the city centre. It serves several major commercial sites and is used by 75,000 passengers per day. The three lines run radially off the city centre but meet at the Commerce interchange station in the city centre.

The data used here have been collected from the Opthor APC system that detects passengers by an infrared system installed at each door of the vehicle. They were collected during a count conducted between 5 September 2013 and 10 April 2014. As part of the count, the operator collected 134,500 valid entries from 124 weekdays and a total of 4900 runs. This sample corresponds to roughly 10% of the number of daily runs. Each entry of the data reports the number of boarding and alighting passengers for a stop and for a given run. In addition, the system calculates the volume of on-board passengers at each station.

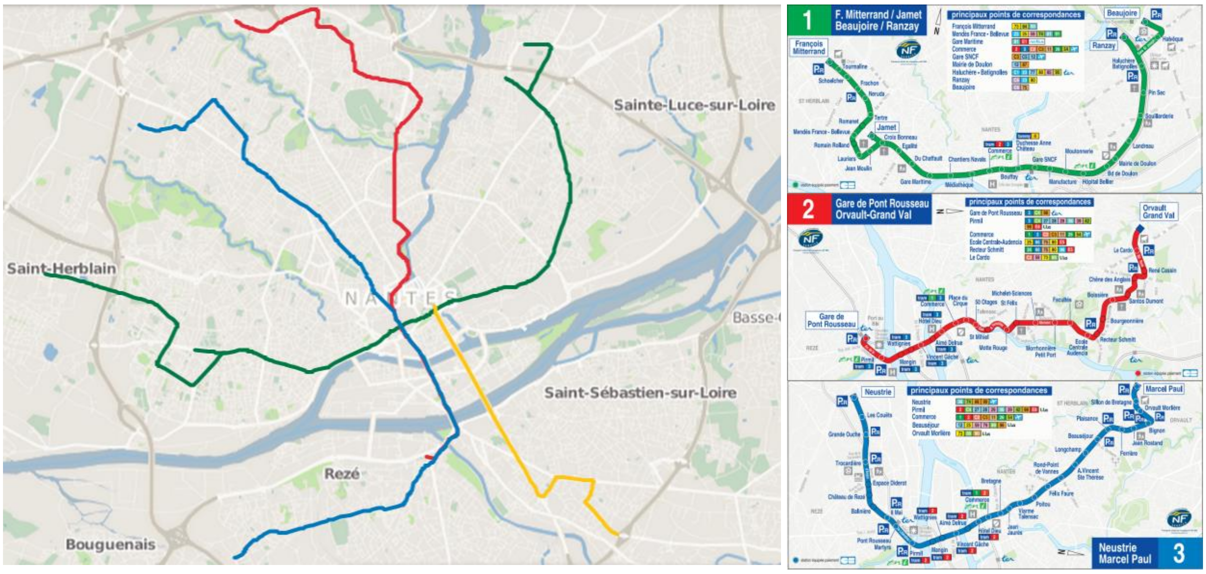

Figure 1: The Nantes tramway network (Source: www.tan.fr)

Methodology

The analysis proceeds in two stages, whereby the dataset is split into two halves. The first half is used to build a 'historic' B&A matrix, while the second is used to measure the prediction accuracy.

Initially, APC is combined with Automatic Vehicle Location (AVL) data in order to precisely reconstruct all runs and attribute to each stop a time of vehicle arrival, a time of departure, total number of boarding passengers, total number of alighting passengers, and to infer the total number of on-board passengers. The historic B&A matrix is then built following a clustering technique per 15-min interval and per day of week. The historic matrix is also represented graphically through a set of heat maps, allowing for the variation of the loading as a function of the location along the line and of the time of the day to be observed. This provides a long-term prediction of passenger loadings, but enables also the preliminary identification of causal relationships between on-board loading and other relevant parameters, preparing the ground of the second stage of the analysis, which involves the use of statistical modelling to more formally identify these relationships and derive a prediction method as a result.

Subsequently, the second half of the dataset is used in order to derive short-term predictions. Two problem settings are defined: (i) a run of an equipped APC vehicle, where upstream information is known and forecasting concerns downstream information; and (ii) a run of a non-equipped vehicle, where forecasting concerns all stops and is based on historical data and the latest run of equipped vehicle, whose headway (from the current run) is also a variable. In the first case, clustering techniques are used to identify similar patterns. In the latter case, autoregressive integrated moving average (ARIMA) modelling is used.

Results

Initial results of the first stage of the analysis are provided here. Six heat maps demonstrating the spatial and temporal variation of the on-board loading are shown in Figure 2, each one corresponding to each of the two directions of each line. The first column on the figure lists the stops of the line in their order of occurrence in any run from bottom to top, while the two columns of heat maps show the on-board passenger volume for each of the two directions of each line, expressed by means of crowding levels of service (LOS), as defined in TRB's Transit Capacity and Quality of Service Manual [1]. It should be noted that in the second column the heat maps are inverted, i.e. they should be read from top to bottom.

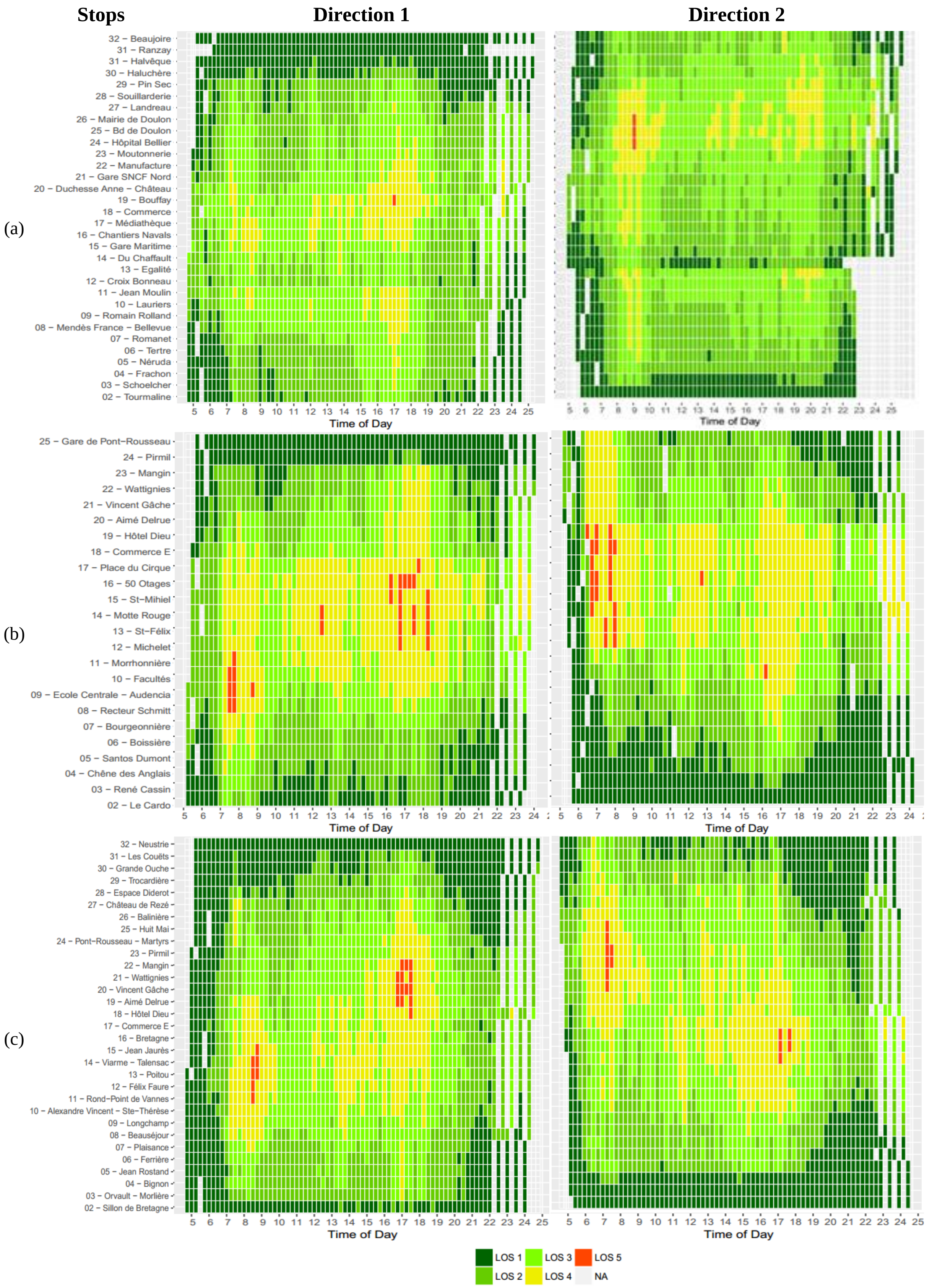

Figure 2: On-board volume heat maps for each line and direction: (a) Line 1; (b) Line 2; (c) Line 3

A key observation that can be made from the heat maps is that, as expected. the highest on-board loads are observed close to the central portions of all three lines and usually during the morning and afternoon peaks. It is notable, however, that the “yellow” state, which denotes standing passengers, is very frequently encountered, and in many cases it is irrespective of the time of the day or location of the vehicle. This is something that the further statistical analysis and the corresponding prediction model that will be derived will attempt to explain in more detail.

Concluding remarks

The results are promising and are useful to operators in many ways. Long-term prediction of passenger loading allows for better scheduling and fleet sizing. Short-term prediction allows for detailed quality of service estimation and could be communicated to passengers at stops so that they plan their journeys accordingly.

References

[1] Transportation Research Board (1999). Transit Capacity and Quality of Service Manual