A Dynamic Model for Short Term Forecasts of Passenger Flows

Evelien van der Hurk *

Department of Decision and Information Sciences,

Rotterdam School of Management Erasmus University, Rotterdam, the Netherlands

Gábor Maróti

Innovation and Process Quality,

Netherlands Railways, Utrecht, The Netherlands

Leo Kroon

Department of Decision and Information Sciences,

Rotterdam School of Management Erasmus University, Rotterdam, the Netherlands

Peter Vervest

Department of Decision and Information Sciences,

Rotterdam School of Management Erasmus University, Rotterdam, the Netherlands

*Email: ehurk@rsm.nl

1 The need for short term forecasts of passenger demand

In research on disruption management in public transport the attention for passenger service steadily increases. The service level depends on the balance between demand and remaining capacity. However, a good understanding of passengers' demand for capacity is generally lacking.

Forecasts of passenger demand traditionally focus on the long-term, like [3, 5]. Forecasts for disruption management on the contrary require detailed short-term predictions. The detail needed for disruption management cannot be obtained in long term forecasts.

Until recently, insufficient information was available for making short term demand forecasts. Since the introduction of smart card systems around the world, detailed information on passenger demand is generated on a daily basis.

In this paper we propose an algorithm that, based on the characteristics of the disruption, dynamically transforms detailed journey data, such as smart card data, to suitable data for forecasting real-time passenger demand as required for disruption management. This is a new way of estimating demand for disruption management as well as a novel application of smart card data.

2 Problem description

An overview of the computationally complex problems involving operating a railway system are described in [1]. A disruption causes a whole set of new problems to arise as the timetable needs to be adjusted, rolling stock needs to be rescheduled, and new crew duties need to be generated. At the same time, passengers are choosing their best travel path.

These two processes depend on each other. Therefore a public transport operator (PTO) can optimize service level by both adjusting the logistic rescheduling to the demand of passengers and influencing the passenger's path choice by providing route advice.

To this end, a PTO needs detailed knowledge of the passenger flows: the origin, destination, time of journey and route choice of the passengers. In this research, we focus on short term forecasts of the passenger flows based on origin, destination and time of journey. Future research will, building on this, focus on modelling and influencing route choice. We use smart card data of Netherlands Railways (NS).

Although smart card data contains detailed information on the time and location of a journey, it is in its raw form usually not directly suitable for detailed short term forecasting. We address the following research question: *How can we deduce information for route advice and rolling stock rescheduling from smart card data?*

3 Smart card data

Smart card systems are a way of ticketing. In the Dutch system, which is similar to the systems implemented in Tokyo, Seoul, Singapore and the London subway, passengers need to tap their card on entrance and exit of each journey. The data generated by Dutch smart cards stores the start location, end location, start time and end time of each journey per card - but not the specific route of a passenger. Each card is uniquely defined and linked to a specific product type. Data on passengers' journeys was never before available in this detail.

4 Methodology

By transforming smart card data to a set of time series, it becomes suitable for detailed short term forecasting. We propose an algorithm that given a specific disruption, dynamically transforms smart card data into time series. These time series form the basis for the prediction models that forecast time dependent origin-destination (OD) passenger flows.

The clustering algorithm selects those ODs that are affected by a disruption and groups them according to similarity in the first and last decision point. A decision point is a station at which a passenger needs to make a routing decision due to the disruption. Clustering ODs leads to more accurate predictions.

We use econometrics and time series analysis for the prediction modeling. We choose to use an autoregressive integrated moving average (ARIMA) model for regression see e.g. [4, 2], a widely accepted econometric forecasting method. We use a robust estimation method, comparing results for a cleaning procedure for the data with a robust model estimation using least absolute errors instead of least squared errors.

5 Results

That clustering of ODs is crucial to come to accurate forecasts of passenger flows is evident from the uneven distribution of the number of journeys per OD, as shown in Figure 1. This graph shows that 15 percent of ODs is responsible for 80 percent of journeys. Though the majority of ODs thus has a low number of journeys, the aggregate accounts for one fifth of the volume. Therefore it is important to group the ODs to increase accuracy of forecasts without loosing important geographical information.

First results indicate that there is a large regularity in the number of journeys per day, per OD and over time. For example, Figure 2 shows the number of people traveling at any given time. Each line is a different weekday; the thickness of the line represents the variation in the number of passengers over a month time. Clearly, there is a regular pattern of the distribution of passengers over the day dependent on the day of the week. Another example is the number of journeys for a specific origin destination, as shown in

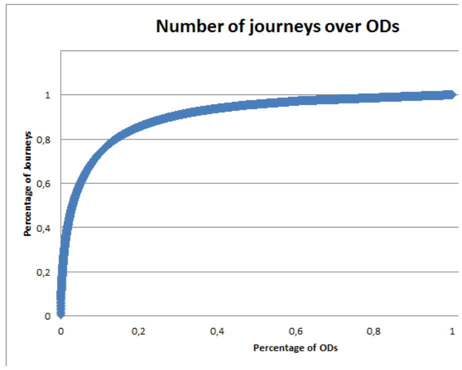

Figure 1: This figure shows that a very limited number of ODs are responsible for the large majority of journeys. 5 percent of ODs contains 60 percent of journeys, 15 percent of ODs contains 80 percent of journeys.

Figure 3. We see that the volume seems strongly dependent on the day of the week.

First results show that with simple regression models, taking into account autocorrelation, we are rather well able to predict the number of passengers traveling per day. Figure 4 shows preliminary results for forecasting the number of journeys per day for a specific OD inside (red dots) and outside (green dots) of sample. Using time series analysis with attention for multiple autocorrelation components we are confident that forecasts' precision of our final models will be more than sufficient for disruption management.

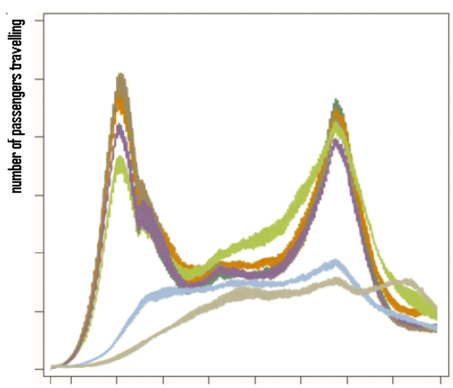

Figure 2: The number of passengers in the system per day over time. Each line represents a different weekday - the thickness of the line presents the variation in the number of people in the system on that weekday over a month time. Distribution of passengers in the system changes over the weekdays. The graph shows a limited part of the day (for data-sensitivity reasons)

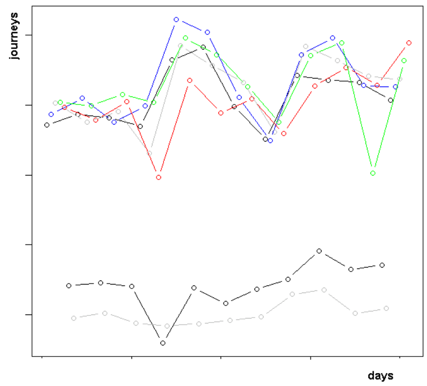

Figure 3: Here, for a specific Origin and Destination, the number of journeys per day are shown for 3 months time. All points belonging to a specific weekday are connected by a line. Again, the number of journeys seems related to the day of the week, but also shows some variation over time.

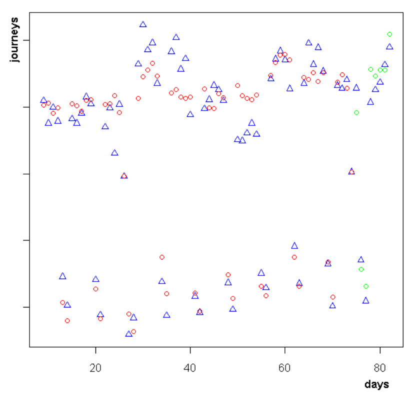

Figure 4: This figure represents some preliminary forecast results. The blue diamonds are the number of journeys per day for a single origin and destination over 3 months time. The dots are the fitted values, where red is inside the sample used for fitting the model, and the green dots are outside of the sample. In the model a trend, variables for the difference between weekdays and weekends, and a lag variable are included.

6 Summary

We define an online prediction model that forecasts real-time passenger flows based on a specific disruption, by first transferring data and then using this specific data for forecasting. To our knowledge this is the first research to focus on short term forecasts of passenger demand for disruption management.

References

[1] A. Caprara, L. G. Kroon, M. Monaci, M. Peeters, and P. Toth. Passenger railway optimization. In C. Barnhart and G. Laporte, editors, *Transportation. Handbooks in Operations Research and Management Science*, volume 14, pages 129–187. Elsevier, 2007.

[2] P. H. Franses. *Time series models for business and economic forecasting.* Cambridge University Press, 2004.

[3] M. Gaudry. An aggregate time-series analysis of urban transit demand: The montreal case. *Transportation Research*, 9(4):249 – 258, 1975.

[4] C. Heij, P. de Boer, P. H. Franses, T. Kloek, and H. K. van Dijk. *Econometric Methods with Applications in Business and Economics.* Oxford University Press, 2004.

[5] M. Wardman. Demand for rail travel and the effects of external factors. *Transportation Research Part E: Logistics and Transportation Review*, 42(3):129 – 148, 2006.