Occupation Level of Buses

Jonas Meli
Institute for Transport Planning and Systems (IVT)
Swiss Federal Institute of Technology Zurich (ETH)
Zurich, Switzerland
jmeli@student.ethz.ch

***Abstract* — A steadily growing demand for increased public transportation in countries like Switzerland elevates the importance of both accurate predictions of occupancies and crowding information for passengers. The analysis in this thesis identifies several factors that influence the occupancies of buses and several factors that do not. Based on the influencing factors as well as the data provided by the VBZ, different linear regression models were built for two stops of various bus lines in Zurich. The results show that weather factors such as rain, temperature, or snow did not improve the quality of the models. In contrast, parameters that describe the time of day are shown to be predictive and should be included in future occupancy models. Future models can be improved by a better understanding of the correlations between the different factors and by further investigation into additional factors not identified here.**

Keywords— Occupation Level, Occupancy Prediction, Public Transportation

I. Introduction

Several train operators make occupancy predictions available for customers. For buses, this does not yet exist in Switzerland. Predicting occupancies is a challenging task due to various reasons. Among these reasons are the day-to-day variability, the headways, the many alternatives to buses in urban areas, and the weather. Occupancy predictions would not only be of value for the passengers but also for traffic operators. Knowledge regarding occupancy could help reduce crowding, as has been done for metro systems [1]. This knowledge could thus lead to improved health [2] and less dwell times [3]. Too long dwell times can lead to irregular arrival times and unevenly distributed occupancies, due to bus bunching [4]. Occupancy predictions can also help in route choice and trip planning.

The goal of this thesis is to find out whether it is possible to reliably predict the occupancy rate of buses. In order to do so, different factors possibly influencing the occupancy were investigated, and sequentially analyzed. Finally, several linear regression models were created and evaluated in MATLAB.

II. Approach

A. Tables

The VBZ provided the data of many public transport vehicles of the city of Zurich in 2018. Using MySQL, tables containing the needed information regarding the bus lines 46 and 72 were created. Among the information contained in these tables were: the stop ID, the next stop, the arrival / departure times, the date, and the occupancy. MeteoSwiss provided the following weather data: temperature, snow height, and precipitation.

B. Influencing Factors

The analyzed influencing factors were: sequence, direction, weekday (either Monday to Friday or Thursday), time of the day, rain, snow height, temperature, and month. The factors were analyzed regarding their influence on the two lines mentioned earlier.

C. Modelling

For modelling purposes, multiple linear regression models were used, trained in MATLAB. The explanatory variables were either 1 or 0. At first, a model was created for each of the two stops *Stampfenbachplatz* (line 46) and *Bucheggplatz* (line 72). The parameters described the direction, the weekday, the month, the weather (rain, temperature, and snow height), and the time. In total, 14 parameters were used. In a second step, a separate model was created for each direction of the two stops. These models had the same number of parameters describing the same factors but adapted to each direction's characteristics. Finally, based on the previously gained knowledge, one model for each stop was created again, but this time without the parameters for rain, snow height and high temperatures.

III. Results

A. One Model Per Stop

The model for *Stampfenbachplatz* had both a higher R^2-value and a lower RMSE compared to the model for *Bucheggplatz*. Table 1 contains the RMSE, R^2 and R^2_{adj} for the first set. Among the parameters negatively influencing the occupancy of the model for *Stampfenbachplatz* were the direction, the snow height, and temperatures above 21°C. For *Bucheggplatz*, the parameters with a negative influence were the rain, heavy rain fall, temperatures above 25°C, and the months of July and August. The biggest positive influences were exerted by both the parameters describing the morning and evening peaks. All five parameters describing the time were positive for each stop and were also the top five influencing parameters.

Table 1: *R^2, R^2 adjusted and RMSE for stop models, first set*

	Stampfenbachplatz	**Bucheggplatz**
R^2	0.62	0.50
R^2 adjusted	0.62	0.50
RMSE	9.4	9.7

B. Two Models Per Stop

Since there is a model for each direction, the models of this set are only based on half of the data compared to the previously discussed set of models. As before, the top influencing factors are all time-describing parameters. Nevertheless, for direction two at *Stampfenbachplatz*, the parameter for the morning peak, as well as another time parameter have negative values. This is likely due to the high value of the intercept (-24.5 wherease before it was 5.3 for *Stampfenbachplatz*).

Table 2 shows the results of the models for each direction. For all models of the second set, the value of R^2 decreased, especially for direction 2 of Bucheggplatz. However, the RMSE for this model decreased, as well as for direction 1 of Stampfenbachplatz. For the other models, the RMSE increased. This is due to the fact that the models which have a lower RMSE in Table 2, model the directions with a generally lower occupancy.

Table 2: *R^2, R^2 adjusted and RMSE for stop models, second set*

	Stampfenbachplatz		**Bucheggplatz**	
	Direction 1	**Direction 2**	**Direction 1**	**Direction 2**
R^2	0.52	0.59	0.49	0.15
R^2 adjusted	0.52	0.59	0.48	0.15
RMSE	9.0	10.9	10.5	8.5

C. Models without weather factors

The third set of models does not include weather factors such as rain, snow height, or temperatures above a certain value. The idea behind this was that for these factor,s the number of measurements was particularly low, leading to possibly questionable results. The results of these models are similar to the first set of models. The top five influencing factors are again the five time-describing parameters. The values for both the morning and the evening peaks are similar to the first set of models. This set of models is based on less data than the first set of models. Nevertheless, the values for R^2 and the RMSE are similar to the first ones, as can be seen in Table 3. This shows that the influence of the weather is, for this data, negligible.

Table 3: *R^2, R^2 adjusted and RMSE for stop models, third set*

	Stampfenbachplatz	**Bucheggplatz**
R^2	0.63	0.50
R^2 adjusted	0.63	0.50
RMSE	9.3	9.8

IV. Conclusion

After studying the influencing factors and after creating several models, it becomes clear that the factors influencing the occupancy of buses are multiple and partially dependent on each other (e.g. temperature at certain times of the day). Furthermore, weather factors are among the influencing factors, although their influence is small compared to the time. In fact, the exclusion of weather factors can produce a slightly improved model. Nevertheless, it must be said that there was too little data backing the weather parameters. With more data behind those parameters, there might be a different conclusion regarding their importance.

In addition, only two stops were analyzed. Hence, the results may not be representative for bus occupancy prediction models. Moreover, this study is based solely on the city of Zurich. For regions with strongly differing climates, the weather factors could be of much more importance.

Furthermore, the connection between the different influencing factors remains unclear. The time also influences the temperature. Therefore, a linear relationship may not be the most accurate approach.

All these findings indicate that the occupation level of buses can be predicted with an accuracy of 50% or more for the city of Zurich under the following conditions: normal weekday, no precipitation, and temperature below a certain mark. Furthermore, the time, the weekday, and the direction are among the influencing factors. It also remains unclear in which relationship the different factors stand. Future studies should consider applying different model types in order to more comprehensively understand the interaction of the various factors influencing the occupancy.

References

[1] E. Jenelius, “Car-Specific Metro Train Crowding Prediction Based on Real-Time Load Data,” in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 78–83.

[2] A. Tirachini, D. A. Hensher, and J. M. Rose, “Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand,” *Transp. Res. Part Policy Pract.*, vol. 53, pp. 36–52, Jul. 2013.

[3] M. Yap, O. Cats, and B. van Arem, “Crowding valuation in urban tram and bus transportation based on smart card data,” *Transp. Transp. Sci.*, pp. 1–20, Oct. 2018.

[4] M. Simeunovic, M. Lekovic, Z. Papic, and P. Pavle, “Influence of vehicle headway irregularity in public transport on in-vehicle passenger comfort,” vol. 7(32), pp. 2874–2881, Aug. 2012.