Data driven public transportation delay modelling

Stanimir Kabaivanov[1, a)] and Veneta Markovska[2, b)]

[1] *Plovdiv University "Paisii Hilendarski"*
24 Tzar Assen Str., 4000 Plovdiv, Bulgaria
[2] *University of Food Technology - Plovdiv*
26 Maritza Blvd, 4000 Plovdiv, Bulgaria

Author Emails
[a)] Corresponding author: stanimir.kabaivanov@uni-plovdiv.bg
[b)] v_markovska@uft-plovdiv.bg

Abstract. Modelling public transportation delays is crucial for building efficient and modern city infrastructure. It is directly related to passenger behavior and satisfaction, as well as to pollution, traffic jams and sustainable development. In this paper we focus on use of jump diffusion processes in analyzing delays. We believe that such an approach is better suited for handling all factors that effect transportation and can also be used to successfully model not only the absolute time spans of waiting times, but also disproportionally large economic impact of huge delays. Our model is tested with real world data and application of obtained results is discussed.

INTRODUCTION

Public transportation has turned from a very important service to a significant factor in ensuring sustainable economic city development. It has therefore been subject to a large number of studies and scientific papers. While it is hard to describe all of them, we have presented the most common approaches to analyzing public transportation in Table 1. Our study puts emphasis on delay modelling, due to the following reasons:

- Quality of service for public transportation services depends on it – both planning and actual execution characteristics are reflected in the way delays happen, how big they are and what impact they have on the users.
- Development of smart cities and intelligent transportation control depends on proper estimation of delays. To allow for flexible and efficient new means of transportation it is also necessary to understand and assess boundaries and limitations of existing public transportation.
- Economic impact of public transportation delays can be huge. Therefore, the negative effects need to be minimized and to do that it is necessary to be able to model delays in a way that also incorporates unexpected and random events.

Delay modelling is also special in terms of forecasting. For variables like number of passengers or number of transportation vehicles that need to serve a route it is necessary to be able to assess and forecast exact numbers. The more accurate they are, the better the model and the higher the benefits of using it. With delays it is sufficient to be able to predict either the possibility of an unexpectedly large delay, or the probability of surpassing a specific threshold value in order to act proactively and resolve problems before they have actually appeared.

To make sure that models are adequate, they have to be able to explain not only the most common characteristics of the delays, but also extraordinary and rare events. This can be achieved through calibration with real world data and selection of theoretical foundation in such a way, as to be able to also incorporate unusual observations and outliers.

TABLE 1. Common approaches for studying public transportation

Approach	Description
Technological point of view on city transportation and passenger behavior.	This approach puts emphasis on the potential of using new means of transportation or on-demand transport, while addressing mainly the available implementation technologies. Notable examples are autonomous vehicles ([1]), use of dedicated smart devices ([2]) or new recognition methods ([3]).
Algorithmic point of view on public transportation and passenger behavior.	The main focus in this approach is the way transportation is planned (as in [4]), (re)scheduled (as in [5]) and actually run (as in [6]).
Economic point of view, focusing on the funding and broad impact of public transportation on businesses.	Papers following this approach consider mainly the economic effects of public transportation and how business environment will be affected by changes in it (as in [7], [8], [9]).
Marketing centered point of view, that focuses on use of transportation services and how they are seen by customers.	Marketing-oriented approach is important for achieving a sustainable growth and city development. In [10] an interesting approach toward measuring the sustainability is suggested, while other research studies as for example [11] focus mainly on evaluation.

Technological solutions can be very flexible and innovative, but they are also limited in nature, due to the fact that it is not always possible to introduce the same solution in all real-world scenarios. The major reasons for this may be budget restrictions, lack of trained staff and the need of huge investments. If we also consider that some options need to undergo a long and quite expensive process of certification, this further limits the benefits of such high-tech options. In this paper we consider only solutions that are not really intrusive in terms of introducing a new technology and install special devices. Therefore, modelling options discussed here rely only on common data collected for public transportation and do not assume that more detailed or custom inputs are available.

Optimization algorithms are very important in order to introduce public transportation services that are both economically sound and meet citizen needs. There are a lot of studies in this area, that present formally proven and fine-tuned specific solutions, provided that some specific assumptions are in place. While such solutions are very well described, they often lack flexibility and are bound to the initial model restrictions. Delay modelling discussed here suggests generic solutions that can be applied under various scenarios and with different assumptions.

Economic impact of new solutions and optimization of existing ones is very important when making decisions and comparing different options. There are some cases in which these effects cannot be exactly calculated – due to lack of historical data (for example when introducing new types of transportation). With this limitation in mind, we have decided to focus on measuring the immediate effects of delays. It is possible to extend our model in a way that accounts for the actual economic impact of delays, which will definitely differ based on day, time and other available means of transportation. However, this would require to make much more detailed and restrictive assumptions without actually adding value to suggested modelling approach. For example, savings in time may have different value in terms of money, when calculated against average salaries or cost levels across cities. If one has only the final amount of money saved by introducing a new transport or optimizing an existing schedule, this may be misleading and hard to compare – first because of differences in costs/salaries and second, due to the time span of a particular event.

In order to provide comparable and free-of-context information on the expected results, we also try to keep indicators measuring economic effects separate from those, that measure physical changes. To avoid further complication and the necessity to justify discounting of cash flows in time, we recommend to use complementary metrics – both physical and economic effects, even when they actually represent the same results.

APPLICATION OF JUMP DIFFUSION PROCESSES

There are several ways to measure public transportation delays, each of them offering important advantages for further analysis:

- absolute values (in time units) and their change over time;
- crossing a barrier, that would trigger a specific action (for example passenger would search and change to alternative transportation method);
- crossing a barrier, that would trigger a significant (and most of the time negative) economic impact.

Depending on the collected data, only some of these indicators may be usable – for example if there is no study on passenger behavior it would be impossible to judge how long does it take before an average person would look for another transportation alternative. To keep the analysis as generic as possible, we have decided to use the first type of input and rely on the absolute delays.

Stochastic models have been used for solving different problems related to transportation – like for example optimal placing of stops ([12]), routes ([13]) uncertainty dynamics ([13]) and capacity estimation ([14]) to name a few. In this paper we focus on use of jumps in the modelling, in order to be able to account for unexpected large delays and consider the probability of such events as additional quality metric for the transportation services.

Stochastic processes, that involve jumps and diffusion, are suitable for public transportation delay modelling due to the following reasons:

- due to accumulation of delays, the possibility to have additive jumps in delay times can match well situations in which one special event triggers a chain of other special situations, resulting in accumulating the delays. With a tight schedule and interconnected means of transportation this is a valuable feature, that reflects the actual cases.
- delays can have different impact, with some causes being so serious, that they can remain longer (for example a road construction or accident can have different impact and last longer than usual traffic jam).

Delay jumps can be compensated by buffers in the schedule. However, this is not always possible, as it depends on the root cause of the delays, and there are cases where major roads may be temporary closed. In order to model such sudden increases in the delays, we employ a model as the one described by equations (*1*) - *(3)*.

$$dV_t = \mu(V_t, t)dt + \sigma(V_t, t)dW_t + dP_t \tag{1}$$

$$dP_t = J(V_t, t, z_t)dN_t \tag{2}$$

$$J(V_t, t, z_t) = \begin{cases} z_t\ , \ for\ additive\ jumps \\ z_t V_t\ , \ for\ multiplicative\ jumps \end{cases} \tag{3}$$

Equation (1) includes all elements of the model, where V_t is the delay time, $\sigma(V_t, t)dW_t$ represents random delay changes and dP_t term indicates the jumps. Depending on $J(V_t, t, z_t)$, which is shown in equation *(3)*, both additive and multiplicative jumps can be modelled. As summarized in Table 2, delays, that accumulate at different rates can be handled by adapting the right side of *(2)*.

TABLE 2. **Inputs used to model bus delays**

Inputs	Remarks
Time delays in time units (minutes, seconds)	Delays in accordance to predefined schedule for public transportation.
Accumulated delays through single course of the bus on a specific lane.	Accumulation of delays (and potential compensation of delays) throughout courses of public transportation vehicles.
Breakdowns and incidents	Breakdowns and incident rates (frequency over a defined period – typically one month).

Data collected by Automatic Vehicle Location (AVL) units, includes position of each public transportation vehicle with high precision and high frequency. When compared the predefined schedules, it is trivial to calculate delays at each point and at any given time. Since these are pretty basic operations, they are universally applicable and information required to assess parameters of equations (*1*) - *(3)* is easy to find. It is quite often the case, that authorities monitor public transportation vehicles for security, safety and economic reasons (for example to calculate funding and transfers based on distance travelled and people using the service). Thus, the necessary inputs, which are presented in more details in Table 2, are already available for every city with connected and monitored transportation network. Even if there is no real-time observation on the public transportation vehicles, it is still sufficient to have a limited number of points, where they are tracked (like stops, crossroads, etc.) in order to be able to calculate delays and use the results for model calibration.

INPUT DATA AND PARAMETER ESTIMATION

To estimate the parameters of the model, we use MLE ([14], [15]) over a set of preprocessed inputs, that represent absolute delays in minutes. Table 3 provides a summary of inputs being used, which consist of minute delays over a single route/line. While the same approach can be used to model data, that comes from multiple routes, it is easier to judge on the applicability of jump-diffusion processes, when they are applied over data, that covers the same route.

TABLE 3. **Input data summary**

Parameter	Value
Number of observations (n)	Total of 346 observations with noted delay of a single bus line. Delays noted in minutes.
Delays summary	Min 1.00, Q1: 5.00, Median: 8.00, Mean: 15.23, Q3: 10, Max: 1000
Standard deviation and Shapiro-Wilk normality test	$\sigma = 59.52$; W = 0.13837, p-value < 2.2e-16

Parameters estimated by the MLE procedure are summarized in Table 4, where two simple models have been estimated – those described by equation (4).

$$dV_t = \mu V_t dt + \sigma V_t dW_t + dN_t \tag{4}$$

Despite being rather simple, this model can provide useful information on observed delays and help in improve planning of the public transportation services. In order to check performance of the model and estimates under different conditions, the MLE has been run with two subsets of the original data set – one that contains all observations and another one, that as all delays larger, than 400 filtered out.

TABLE 4. **Numerical results**

Parameter	Value	
	All	Delays < 400
Number of estimated jumps	31	27
Average inter-arrival times	0.021032	0.022131
Average jump size	0.235484	0.196296
Standard Dev. of jump size	278.782811	35.676886
Jump Threshold	15.874508	15.874508

The results of sample calibration for model (4) are shown on Figure 1, where the left panel presents raw times, while the right one has both absolute delays and forecasted output in red color (also indicated with a small dot).

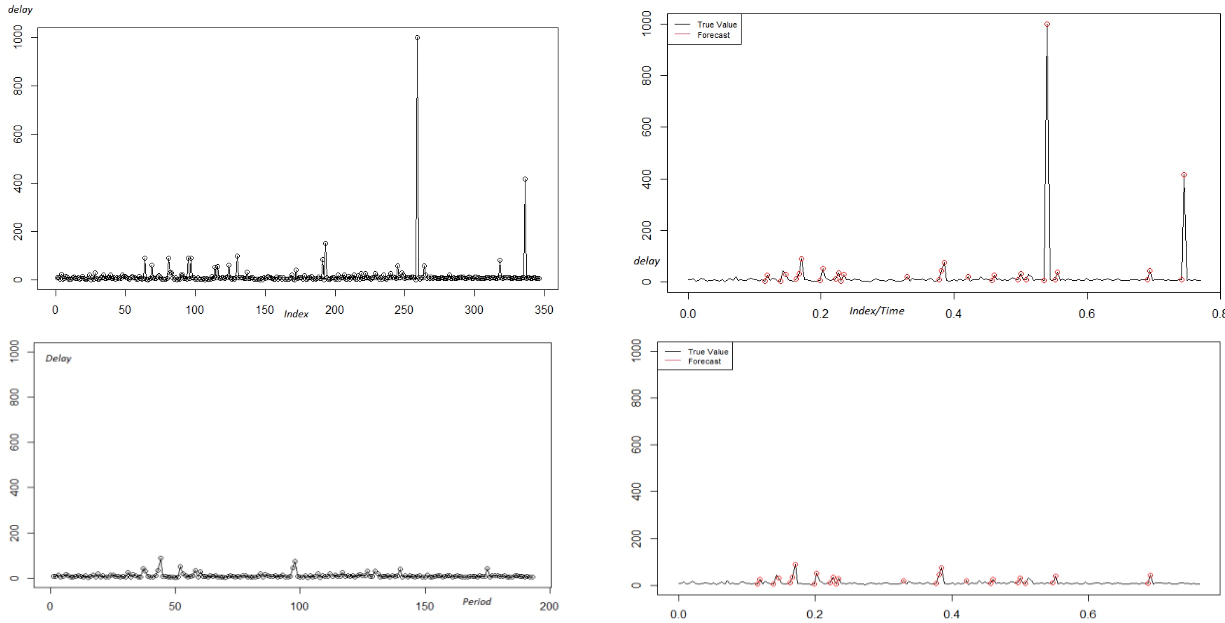

FIGURE 1. Actual and modelled values of data sample for bus delays for all cases (top) and delays < 400 (bottom).

Use of actual data indicates, that it should be possible to improve handling of special situations and reacting to accidents when in real use. There are several areas, where it may be especially valuable:

- In identifying areas, where jumps are more likely to appear. The quantification of jumps provided by this method can also be used to justify spending more resources and time on trying to address issues in these particular areas.
- In identifying areas, where delays vary (based on estimates of the standard deviation). This would help in trying to assess and address the core reasons, that lead to variability in the reliability of public transportation schedules.
- In optimizing existing schedules in a way, that creates larger buffers whenever necessary and introduces extra capacity to compensate for large unexpected delays.

It is also possible to apply the model with several different goals in mind – for example schedule optimization and support for better resource allocation. It should be noted though, that this model is not well suited for tracking down individual passenger behavior, as it focuses on transportation delays, that are the same for all users, regardless of their individual characteristics and requirements.

CONCLUSION

Our tests have indicated that delay modelling can be used to analyze special events in city transportation. The results can be used to support planning and scheduling. While this requires to take into consideration a number of factors that are not available for the inputs used here, the algorithms and approaches used are still applicable. Therefore, we can outline three major benefits from the suggested model:

- Jump diffusion processes can be used to model delays, that are caused by a various number of reasons.

Extreme delays and accidents can lead to situations that are hard to model with other analytical tools. Jump diffusion processes offer a well-known approach that matches the observed obstructions in the real world.

- They can be used as part of the techniques used to optimize transportation schedules and analyze customer satisfaction from public transportation system.

Customer satisfaction is heavily influenced by the delays. If we assume there is a “tolerable slowdown” which causes inconvenience but no significant negative impact on users, surpassing it leads to notable negative effects.

- Average jump sizes and jump thresholds can be used to measure the throughput of the transportation system and analyze different scenarios.

To sum up, use of jump diffusion processes can improve traffic simulation and predict emergency cases strategies in transportation. If used in a proper way, it can provide new metrics on user satisfaction, economic effects and efficiency of existing transportation schedules.

REFERENCES

[1] Ł. Owczarzak and J. Żak, "Design of passenger public transportation solutions based on autonomous vehicles and their multiple criteria comparison with traditional forms of passenger transportation," *Transportation Research Procedia 1,* vol. 10, no. 1, p. 472–482, 2015.

[2] *Understanding Passenger Patterns in Public TransitThrough Smart Card and Socioeconomic Data: A case study in Rennes, France,* 2014.

[3] C. Wang and M. Tian, "Passenger flow direction detection for public transportation based on video," in *2010 International conference on multimedia communications*, IEEE, 2018.

[4] M. Rodrigues, A. Queirós and C. Rodrigues, "mart Mobility: A Systematic Literature Review of Mobility Assistants to Support Multi-modal Transportation Situations in Smart Cities," in *Integrated Science in Digital Age 2020*, 2020.

[5] A. Schöbel, "An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation," *Transportation Research Part C: Emerging Technologies,* vol. 74, no. 1, pp. 348-365, 2017.

[6] D. Esztergár-Kiss, Z. Rózsa and T. Tettamanti, "An activity chain optimization method with comparison of test cases for different transportation modes," *Transportmetrica A: transport science,* vol. 16, no. 2, pp. 293-315, 2020.

[7] Z. Zhao, H. Koutsopoulos and J. Zhao, "Discovering Latent Activity Patterns from Transit Smart Card Data: A Spatiotemporal Topic Model," Urban Mobility Lab at MIT, 2020.

[8] A. Halvorsen, H. Koutsopoulos, Z. Ma and J. Zhao, "Demand Management of Congested Public Transport Systems: A Conceptual Framework and Application Using Smart Card Data," Urban Mobility Lab at MIT, 2019.

[9] K. Turoń and A. Kubik, "Economic aspects of driving various types of vehicles in intelligent urban transport systems, including car-sharing services and autonomous vehicles," *Applied Sciences,* vol. 10, no. 16, p. 5580, 2020.

[10] A. Alkharabsheh, S. Moslem, L. Oubahman and S. Duleba, "An integrated approach of multi-criteria decision-making and grey theory for evaluating urban public transportation systems," *Sustainability,* vol. 13, no. 5, p. 2740, 2021.

[11] X. Lai, J. Teng and L. Ling, "Evaluating Public Transportation Service in a Transit Hub based on Passengers Energy Cost," in *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020.

[12] W. Klumpenhouwer and S. C. Wirasinghe, "Optimal time point configuration of a bus route - A Markovian approach," *Transportation Research Part B: Methodological,* vol. 117, no. A, pp. 209-227, 2018.

[13] L.-M. Kieu, D. Ngoduy, N. Malleson and E. Chung, "A stochastic schedule-following simulation model of bus routes," *Transportmetrica B: Transport Dynamics,* vol. 7, no. 1, pp. 1588-1610, 2019.

[14] A. Trivella and F. Corman, "Modeling uncertainty dynamics in public transport optimization," in *19th Swiss Transport Research Conference*, Monte Verità / Ascona, 2019.

[15] C. Wang, Z. Ye, J. D. Fricker, Y. Zhang and S. V. Ukkusuri, "Bus Capacity Estimation using Stochastic Queuing Models for Isolated Bus Stops in China," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2672, no. 8, pp. 108-120, 2018.

[16] S. C. Nixon and M. (. Chen, "Maximum likelihood estimation of stock volatility using jump-diffusion models," *Cogent Economics & Finance,* vol. 7, no. 1, p. 1582318, 2019.

[17] S. Beckers, "A Note on Estimating the Parameters of the Diffusion-Jump Model of Stock Returns," *The Journal of Financial and Quantitative Analysis ,* vol. 16, no. 1, pp. 127-140, 1981.

[18] S. Kabaivanov, V. Markovska and M. Milev, "Using real options analysis to support strategic management decisions," in *AIP Conf. Proc. 1570*, Sozopol, 2013.

ACKNOWLEDGMENTS

Authors wish to acknowledge the assistance and funding of this research as part of project UIA05-202 "INNOAIR - Innovative demand responsive green public transportation for cleaner air in urban environment", funded by the European Union initiative - Urban Innovative Actions. (UIA).