Olympic Games passenger flow analysis and forecasting

Vincent Bories[1], Rémi Coulaud*[1], Yanan Xin[2], Oded Cats[2]

[1] Transilien SNCF Voyageurs, France

[2] Department of Transport & Planning, Delft University of Technology, the Netherlands

SHORT SUMMARY

Using smart card data collected during the 19-day Paris Olympic Games, this study offers unique insights into the travel dynamics during this mega-event and the event's impact on the public transport system during peak episodes. We identify six groups of departure flow profiles after game events that show distinct patterns in terms of the timing and volume of departure peak flows. A set of potential contributing factors including temporal, spatial, and event-related features are used to explain these clusters. The results show that difference among the departure profiles can be well characterized by event capacity, event duration, distance from the station to the stadium, distance from the station to Paris center, and event ending time. In addition, we applied forecasting models to predict passenger flow departure volumes at 30-minute intervals. Our results can support crowd management, optimize service capacity, and resources allocation effectively during large-scale events.

Keywords: Olympic games, passenger flow, departure time, travel behavior

1. INTRODUCTION

Organizing a mega-event such as the Olympic Games poses significant challenges to the city's public transport system. The 33rd Olympic Games hosted in Paris in 2024 relied primarily on the city's public transport system to move its spectators, similarly to the London Olympic Game in 2012 (Currie and Shalaby, 2015). Every single one of the stadiums as well as a large part of the Olympic Games venues were accessible by public transport, and no dedicated parking was offered. To organize public transport to accommodate more than the 300 events which take place during the Olympic Games, it is of utmost importance to understand the magnitude of the induced travel demand and the influence on the peak travel flows following the game events. Kassens-Noor (2013) pointed out that the information shared between an Olympic Game organization to another is surprisingly limited. Using the smart card data collected during the 19-day Paris Olympic Games, this study offers unique insights into the travel dynamics during this mega-event and the event's impact on the public transport system during peak episodes.

To accommodate the exceptional travel demand during the games, additional public transport services are often put into place in advance (Currie and Delbosc 2011; Currie and Shalaby, 2015; Robbins et al., 2007). Yukun (2008) provided high-level findings on the magnitude of passenger flows who made use of different service lines during the Beijing 2008 Olympic Games. Analysing data from the same games, Liu et al. (2008) investigated the link between Olympic events and travel demand for both public transport and car traffic. While these studies offer some insights into the transport impacts of Olympic Games on the hosting cities, there is still lack of knowledge on how those are related to the time, location and type of sport events which take place. Furthermore, the excessive demand often occurs during short durations, mainly shortly before and after the games, causing significant pressure on the transit system to absorb and diffuse passenger

Interne

flows. Therefore, accurate short-term forecasting of the spatial-temporal passenger demand patterns is paramount to avoid overcrowding at public transport stations.

Several studies have investigated travel demand during special events using disaggregate data. Pereira et al. (2014) measured the impact of events at exhibition centres and music halls by separating regular passenger flow from event-based passenger flow. Li et al. (2017) developed a deep neural network to forecast the next 30-minute passenger flow based on smart card data during pop-star concert events. Chen et al. (2019) proposed a generic framework based on ARIMA and GARCH models to forecast arrival and departure passenger flows profiles of 10 pop star and China Football League events. The ARIMA models were used to forecast regular flow while the GARCH model the volatility from event-based passenger flow.

In this study, we analyse passenger flow departure profiles for more than 200 Olympic Game events. We characterized the departure profiles into different clusters and examined how the difference in departure profiles associates with the spatial, temporal, and game event characteristics. In addition, we applied forecasting models to predict passenger flow departure volumes at 30-minute intervals using smart card data. Our results can support crowd management, optimize service capacity, and resources allocation effectively during large-scale events.

2. METHODOLOGY

We study passenger flow departure volume of each Olympic event linked to a Transilien SNCF Voyageurs station. We limit our analysis to the 13 so-called “Olympic stations” which directly served an Olympic stadium. Passenger flow volumes are observed through smart card data records collected between 4:30 am and 01:00 am of the following day with a sampling rate of 30 min. We link smart card data with Olympic events characteristics such as event duration, ending time and capacity. We focus on 207 Olympic events which correspond to 309 departure profiles (Olympic events × stations) during the 19 days of Paris 2024 Olympic Games (26/07/2024 – 11/08/2024).

We focus on passenger flow entering station occurring between one hour prior to the event ending time and three hours after it, i.e. a time window of 240 minutes as display in Figure 2. We noted passenger flow during this departure time $\boldsymbol{y_e} = (y_{1,e}, \dots, y_{t,e}, \dots, y_{8,e})$ aggregate by a 30-min time step *t* as defined in Table 1.

Interne

Table 1: Main notations and variables

Variable	Notation	Domain	Description
Passenger flow	$\boldsymbol{y_e}=(y_{1,e}, \dots, y_{t,e}, \dots, y_{8,e})$	$[0, 1, \dots]^8$	Passengers' flow entering a station in 240 min
Time	t	1, …, 8	30-min time step
Departure profiles	e	0, 1, …, n	Departure profile is a unique key defined by an Olympic event and a station

We first construct passengers' flow groups using k-means with Euclidian distance on the vector of passenger flow ($\boldsymbol{y_e}$). We search for similitude between behavior of Olympic spectators leaving stadiums and select the number of clusters by applying the Elbow criterion.

We are interested in identifying a set of variables which explain the observed passenger flow departure profiles and thereby reveal their underlying determinants. A short description of the candidate variables compiled as part of this study is provided in Table 2. The set of variables considered includes temporal features (day number, ending time), spatial features (subway connection, station to Paris distance, stadium to station distance) and event characteristics (gold medal, event capacity, event duration). We deliberately refrain from including variables pertaining to individual stations of the Paris network since this will defeat the purpose of obtaining results which are useful and transferable for other contexts, such as the Los Angeles 2028 or the Brisbane 2032 Olympic Games and thereby support their (public) transport plans.

Table 2: Potential determinants of passenger flow departure profiles

	Variable	Unit	Description
Temporal features	Day number	0, 1,	Day number since the beginning of the Olympic Game
	Ending time	Hours	Ending hour of the event
Spatial features	Subway connection	Boolean	Is there a metro connection at the station?
	Station to Paris center distance	Meters	Geodesic distance between the center of Paris and the station
	Stadium to station distance	Meters	Geodesic distance between the stadium and the station
Event characteristics	Gold medal	Boolean	Is there a gold medal during the event?
	Event capacity	0, 1,	Capacity in terms of maximum admissible spectators to the event
	Event duration	Hours	Duration of the event

Interne

We estimate a random forest model which considers the temporal dependency of passenger flow modeling the impact of explanatory variables defined in Table 2 under the full vector $\mathbf{y}_\mathrm{e}$. The hyperparameters of the random forest (number of trees, number of explanatory variables per node) are optimized by 5-folds cross-validation on the training set (75% of all dataset).

We compare our random forest model to a benchmark model: the mean per time step defined as $\mathrm{y}_{\mathrm{t,e}} = \beta_t^0 + \varepsilon_{t,e}$ and a linear model enriched with explanatory variables of Table 2 per time step $\mathrm{y}_{\mathrm{t,e}} = \beta_t^0 + f_t(x_e) + \varepsilon_{t,e}$.

Finally, the performance of different statistical models in forecasting passenger flows is compared. We split thus the dataset randomly into a training set (75%) and a test set (25%). The metrics used to compare models are the Mean Absolute Error:

$$\mathrm{MAE} = \frac{1}{n \times 8} \sum_{e=1}^{n} \sum_{t=1}^{8} |y_{t,e} - \hat{y}_{t,e}|$$

and the Weighted Mean Absolute Percentage Error:

$$\mathrm{wMAPE} = \frac{1}{n} \sum_{e=1}^{n} \frac{\sum_{t=1}^{8} |y_{t,e} - \hat{y}_{t,e}|}{\sum_{t=1}^{8} y_{t,e}} \times 100$$

3. RESULTS AND DISCUSSION

Passenger flow analysis

We first conduct a detailed analysis of the passenger flow departure profiles after the completion of sport events to offer an overview of the induced demand on the public transport system, particularly at stations close to the event venues. Figure 1 displays how much more travel volume is induced by each of the events considered compared to normal times. This information can be helpful in planning additional transport services in advance.

Interne

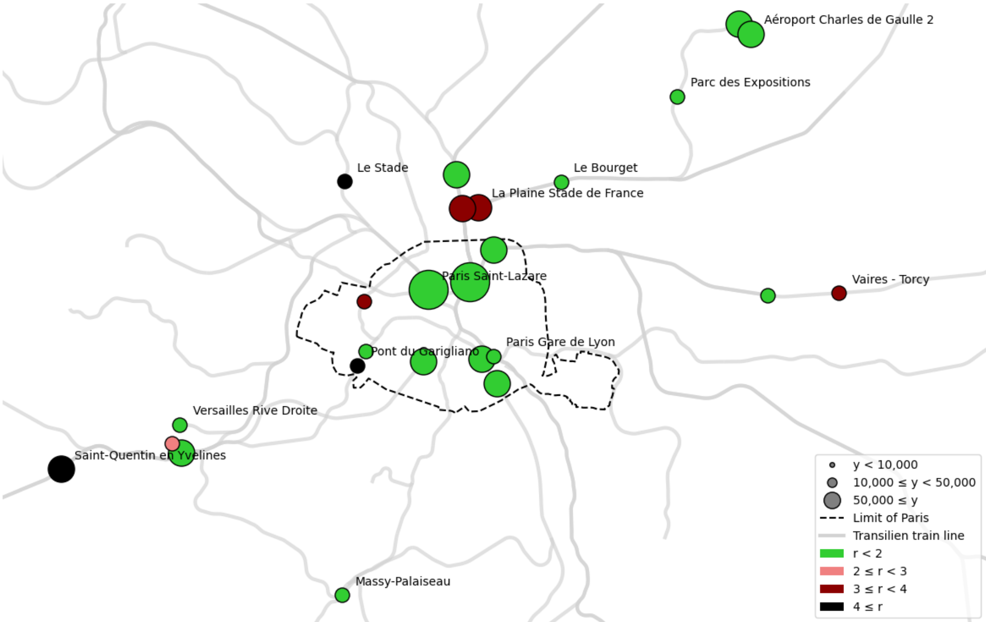

Figure 1: Map showing average absolute passenger flow volume between 8 and 9 PM during the Paris 2024 Olympic Games (circle size) and passenger flow increase compared to summer 2023 with $\mathbf{r} = \frac{\textbf{\textit{passenger flow during the Olympic Game}}}{\textbf{passenger flow } \textbf{\textit{during summer}} \textbf{ 2023}}$ **(color scheme) along the main commuter train stations**

We focus only on passenger flow induce by the ending of an event. An illustrative example is provided in Figure 2, i.e. the area highlighted in grey. We take the equivalent days smoothed as the sum of three gaussians in 2023 for reference. When several events could potentially impact the same station at the same time (i.e., having at least 30 minutes overlap between departure windows), we split the flows between the events according to the respective event capacities. This concerns three stations out of the 13 and impact a total of 90 departure profiles.

Interne

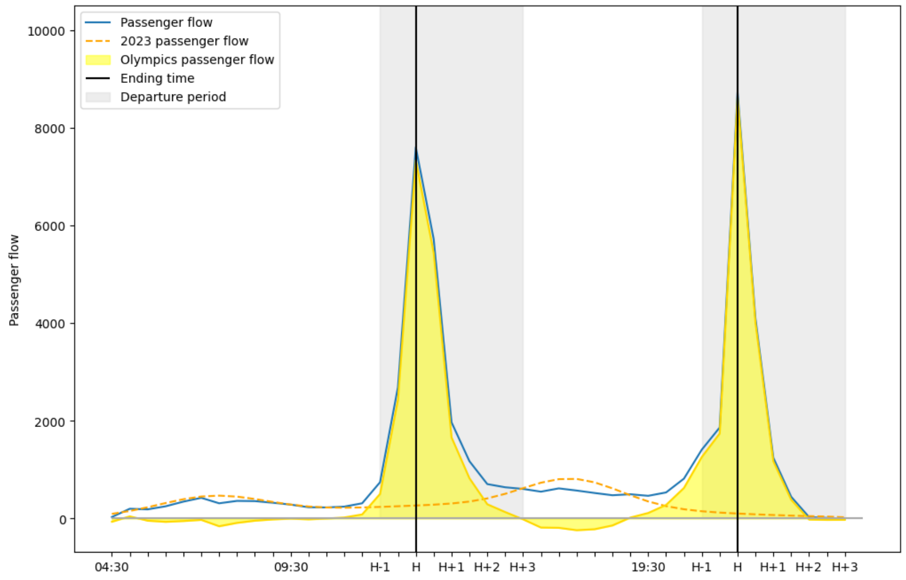

Figure 2: Passenger flows on August 2nd, 2024, at La Plaine Stade de France station. The blue line represents the passenger flow (entering station) observed with smart card data aggregated per 30 min. The orange dot line is the passenger flow volume of 2023. We study the Olympic Games-induced passenger flow, which is the difference between the total observed passenger flow and the 2023 passenger flow. Departure profiles are supposed to be the Olympic passenger flow occurring during the departure time window (grey area), which ranges from one hour before the event ending time (black vertical line) until three hours after the event

Interne

Analyzing passenger flows entering the station allows us to identify similar passenger flow profiles in terms of timing and volume. As can be observed in Figure 3, certain passenger flow profiles exhibit similar trends in terms of the shape and volume transported, e.g. when does the peak occur and how tall and wide is the peak.

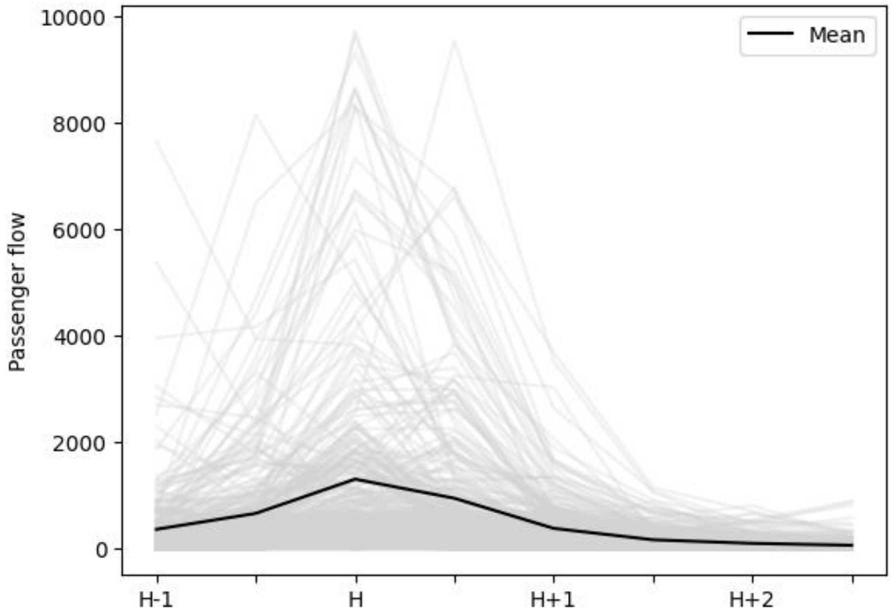

Figure 3: The 309 studied departure profiles (event × station) relative to the 207 distinct studied Olympic events

Next, we cluster the passenger flow profiles using k-means and identify 6 distinctive profiles using the Elbow criterion. To understand the factors contributing to the different passenger flow departure profiles, explanatory variables in Table 2 are linked with the clusters for further analysis. A comparison of the variable distributions across the six clusters is shown in Figure 4, while Figure 5 displays the distribution of the six clusters over Olympic stations.

In Figure 4, notably, clusters 3 (dark green), cluster 5 (light green), and cluster 6 (yellow) only associate with the "La Plaine Stade de France" station. As shown in Figure 4, these three clusters correspond to only high-capacity events. Since this station is next to the country's largest stadium, "Stade de France" stadium, which has a seating capacity until 81338 places, the cluster results clearly illustrate the significant influence of this stadium on departure flows. Furthermore, when looking at other features, cluster 5 (light green) only includes events ending late in the evening and cluster 6 (yellow) only includes short-duration events, while cluster 3 encompasses both. Therefore, we name the three clusters using the station name "La Plaine Stade de France" with additional suffixes "late" and "short" for cluster 5 and cluster 6 respectively.

In addition, cluster 2 (red) and cluster 4 (turquoise) represent departure profiles from low-capacity events. The two clusters differ at the "station to stadium distance" feature in which cluster 4 (turquoise) has smaller distance. Indeed, Parc des expositions and Le Stade are stations near stadium, therefore we call it "Small volume, earlier departure". The first cluster (blue) is characterized by the relatively short duration of events, central location of stations, and proximity between stations and stadiums, therefore is names as "Short, earlier departure".

Interne

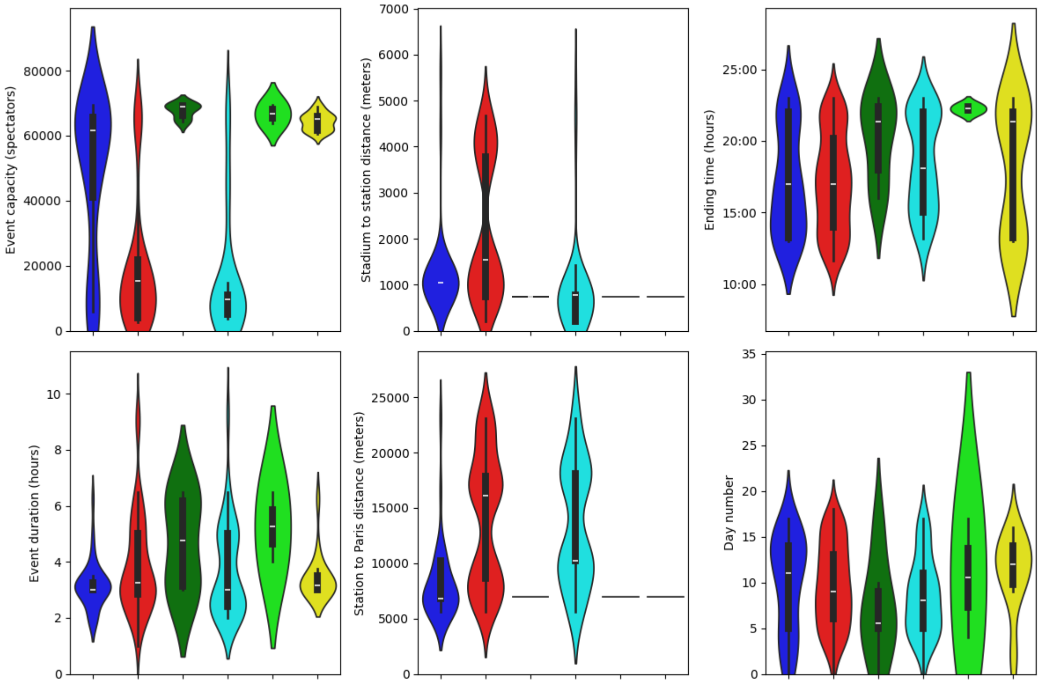

Figure 4: Distributions of event (left), spatial (middle) and temporal (right) features per cluster.

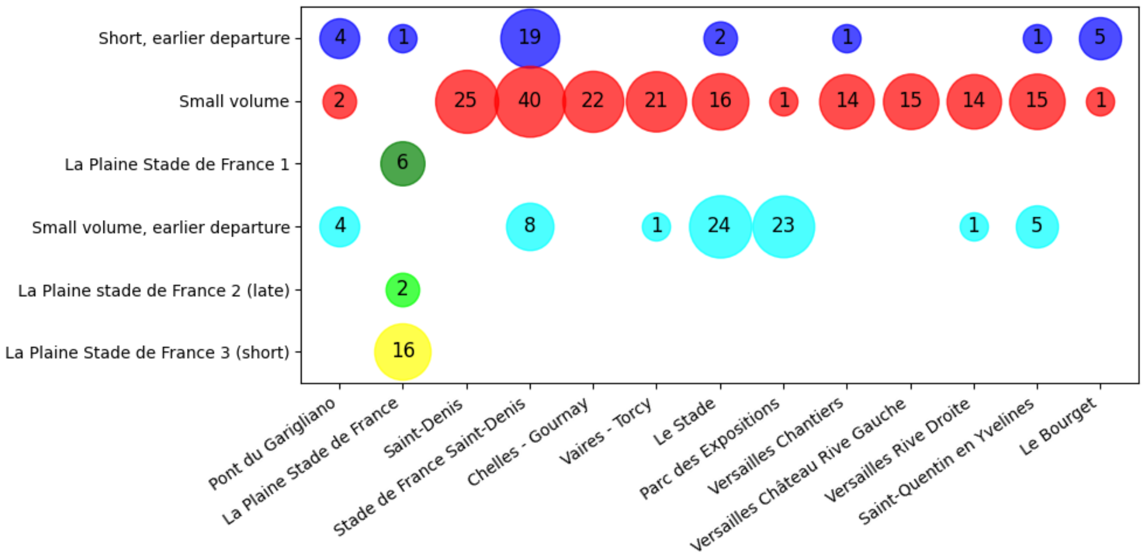

Figure 5: Heatmap of departure profiles per clusters (y-axis) and stations (x-axis) (same color per cluster as in Figure 4). The number inside the circle denotes the number of event-station departure profiles.

Our findings show that passenger flow departure profiles from various stations and events differ and take distinctive forms depending on the spatial, temporal, and event characteristics context. This means that these variables influence passenger flow profiles in terms of both shape and

Interne

intensity, which suggest that these variables can be useful in forecasting passenger flow during the Olympic Games.

Passenger flow profiles forecasting

Following the analysis of passenger flow by grouping departure profiles together and giving general insights on which variables are correlated with passenger flow profiles, we now turn to compare forecasting performance of different models in relation to passenger flow departure profile.

In Table 3, we compare the prediction results of three models. The first model estimates each departure profile simply by taking the mean per time step. The average passenger flow per half hour is about 400, therefore a mean absolute error of 464 passengers indicates very low performance when simply using the mean estimate. The second model is a linear model predicting volume per time step, using Table 2 explanatory variables. We observe that both MAE and wMAPE errors are lower than the first model. However, this linear model is limited in capturing the non-linear associations between the explanatory variables and the target variable. On the contrary, the third model, random forest model, better models the complex associations between variables. It offers a significant performance improvement and halves the prediction error compared to the linear model. Main parameters of the random forest model were found with 5-folds cross-validation. We ended up using three random variables per node and 400 trees.

Table 3: Forecasting performance of passenger flow models in terms of mean absolute error and weight mean absolute percentage error on the test set with 95% confidence interval (±)

	Mean estimate	Linear model	Random forest model
MAE	433 ± 60	362 ± 50	195±29
wMAPE (%)	247±110	129±35	60±13

In addition to the forecasting accuracies reported in Table 3, we also want to identify the most important features contributing to the random forest forecasting (Figure 6). This importance score shows how frequently a variable is used in constructing a tree in the random forest model. We observe that the most important variables are event capacity, the distance from the stadium to the station, and the distance between the stadium location and the center of Paris. This is in line with expectations since the shapes of departure profiles are mostly characterized by the timing and volume of peaks of spectator flows. The peak volume is directly associated with event capacity. While the peak departure time is closely related to the travel distance from the stadium to the station and from the station to the center of Paris which decides whether walking or taking a shuttle bus is needed to reach a stadium from the station.

Interne

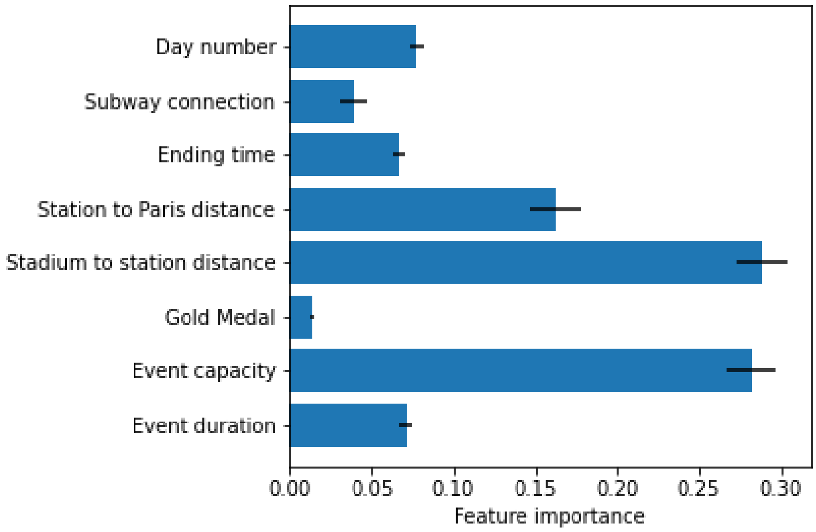

Figure 6: Features importances of the variables included in the estimated Random Forest model, along with their respective 95% confidence intervals.

4. CONCLUSION

This short paper presents the initial analytical results on the travel demand and flow dynamics associated with public transport use during the 2024 Paris Summer Olympic Games. Leveraging a rich set of smart card data, we identified six groups of departure flow profiles after game events that show distinct patterns in terms of the timing and volume of departure peak flows. A set of potential contributing factors including temporal, spatial, and event-related features are used to explain these clusters. The results show that difference among the departure profiles can be well characterized by event capacity, event duration, distance from the station to the stadium, distance from the station to Paris center, and event ending time. Furthermore, based on the insights gained from the descriptive analysis, three forecasting models, mean estimate, linear model, and random forest, are used to forecast departure volume at the 30-minutes intervals for all Olympic Games. The random forest produced the best prediction performance. Further analysis on the feature importance shows that the event capacity, distance from the station to stadium, and the distance from the station to the city center features are most important for the random forest forecasting.

The infrequency and high relevance of Olympic Games pose unique challenges to study and forecast their impacts on travel demand during the event. This study provides valuable addition to our understanding on the travel dynamics pertaining to the use of public transport as a major mode for event travel. Meanwhile, the results could facilitate designing crowd management, information provision, and public transport operations measures.

Future research can be extended in a few directions. In the current study, the analysis focused on the departure profiles, however, the same methods can also be applied to analyze travel flows related to event arrival. In addition, other forecasting methods combined with additional feature set can be further explored to improve the prediction accuracy and confidence, such as the neural network approach or probabilistic forecasting.

Interne

REFERENCES

Chen, E., Ye, Z., Wang, C., Xu, M. 2019. Subway passenger flow prediction for special events using smart card data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 21, No. 3, pp. 1109-1120.

Currie, G., Delbosc, A. 2011. Assessing travel demand management for the summer Olympic games. *Transportation research record*, Vol. 2245, No. 1, pp. 36-48.

Currie, G., Shalaby, A. 2012. Synthesis of transport planning approaches for the world's largest events. *Transport Reviews*, Vol. 32, No. 1, pp. 113-136

Jiang, Y. 2008. Analysis on Beijing subway flows during the 29th Olympics. *Journal of Transportation Systems Engineering and Information Technology*, Vol. 8, No. 6, pp. 46-51.

Kassens-Noor, E. 2013. Transport legacy of the Olympic Games, 1992–2012. *Journal of Urban Affairs*, Vol. 35, No. 4, pp. 393-416.

Li, Y., Wang, X., Sun, S., Ma, X., Lu, G. 2017. Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies*, Vol. 77, pp. 306-328.

Pereira, F. C., Rodrigues, F., Polisciuc, E., Ben-Akiva, M. 2015. Why so many people? explaining nonhabitual transport overcrowding with internet data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 3, pp. 1370-1379.

Robbins, D., Dickinson, J., Calver, S. 2007. Planning transport for special events: A conceptual framework and future agenda for research. *International Journal of Tourism Research*, Vol. 9, No. 5, pp. 303-314.

Xiaoming, L. I. U., Jifu, G. U. O., Zhuangzhi, S. U. N. 2008. Traffic operation with comments during Beijing Olympic Games. *Journal of Transportation Systems Engineering and Information Technology*, Vol. 8, No. 6, pp. 16-24.

Interne