Factors affecting bus bunching at the stop level: A geographically weighted regression approach

Journal Article

Author(s):
Chioni, Evangelia; Iliopoulou, Christina; Milioti, Christina; Kepapsoglou, Konstantinos

Publication date:
2020-09

Permanent link:
https://doi.org/10.3929/ethz-b-000587832

Rights / license:
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

Originally published in:
Factors affecting bus bunching at the stop level: A geographically weighted regression approach

Evangelia Chioni, Christina Iliopoulou, Christina Milioti, Konstantinos Kepaptsoglou

School of Rural and Surveying Engineering, National Technical University of Athens, 15770 Zografou Campus, Greece

ABSTRACT

Efficient operation of bus networks is vital for urban centers. Unfortunately, factors such as uneven passenger loads and congestion hinder the adherence to posted schedules, leading to reliability issues. Most notably, bus bunching has been identified as a significant reliability problem, impacting both users and operators. Bus bunching is treated as a route-level problem in the relevant literature, while spatial patterns in explanatory factors are overlooked. Diverging from the typically performed route-level analysis, this study exploits Automatic Vehicle Location data to investigate factors affecting bus bunching at the network level, while taking into account their spatial variability. For this purpose, a Geographically Weighted Regression Model is applied to model bus bunching, using bus stop and network attributes as explanatory variables. Results for approximately 360 bus stops in Athens, Greece underline the superiority of the proposed model to Ordinary Least Squares Regression and corroborate the presence of spatial variability in the factors affecting bus bunching. Indeed, the number of traffic lanes at the stop level is positively associated with bunching in heavy traffic segments, whereas a higher number of lanes is negatively linked to bunching in less congested regions. Further, the number of bunching occurrences generally increases with the number of routes serving each stop, as well as with the distance from subway stops in the outer parts of the city. Such findings highlight the need to consider spatial structures in relevant models and can help improve their reliability and accuracy.

1. Introduction

The advent of Intelligent Transportation Systems has created a new reality for public transport systems, allowing for monitoring public transport service quality, identifying performance issues and associated factors (Furth et al., 2006). Combining ITS and optimization in public transportation planning: state of the art and future research paths. In particular, the availability of precise Automatic Vehicle Location (AVL) data has enabled the investigation of reliability issues and most notably, the problem of bus bunching. The latter involves the simultaneous arrival at a stop of vehicles serving the same route as a result of successive deviations from the schedule (Verbich et al., 2016). Bus bunching is associated with a series...
of negative impacts for users, including increased waiting times and discomfort, while it naturally damages the operator’s public image (Moreira-Matias et al., 2014).

In recent years, the bus bunching problem has attracted a lot of research attention. A large body of literature has focused on the determination of optimal control and scheduling strategies to mitigate the problem (Andres and Nair, 2017; He et al., 2020; Iliopoulou and Kepaptsoglou, 2019; Liang et al., 2019; Wu et al., 2016, 2019a,b). Several studies also attempted to identify factors affecting bus bunching using statistical methods, typically focusing on a single bus line. Among such studies, Feng and Figliozzi (2011) investigated the causes of bus bunching using AVL and Automatic Passenger Count (APC) data, noting the effect of late departures from stops. Fonzone et al. (2015) analyzed the impact of passenger arrival distribution to bus bunching, creating a continuous logit model for emulating passenger behavior. Arriagada et al. (2019) underlined the role of other factors affecting bunching, such as the use of non-homogeneous bus fleets, the deployment of double-decker buses and the presence of on-street parking lanes adjacent to bus lanes. The authors also highlighted factors which can mitigate bus bunching, such as a traffic signal coordination, dispatch regularity and the existence of exclusive bus lanes. Moreira-Matias et al. (2012) identified stops with frequent bunching events and a systematic deviation between actual and planned headways. The study employed time-series analysis, showing that the occurrence of bus bunching at the beginning of the route amplifies the phenomenon at subsequent stops. Similar findings were reported in Nguyen et al. (2019) who employed binary logistic regression and survival analysis to model streetcar bunching, identifying stop location, departure headways and the combination of multiple vehicle types as significant factors for bunching formation. Along the same lines, Rashidi et al. (2017) applied Decision Tree Models and Logistic Regression for a single bus line and identified schedule deviation as the main cause for bunching. Recently, Enayatollahi et al. (2019) modeled bus bunching using cellular automata and reported that even a small decrease in passenger boarding/alighting times could significantly reduce bunching occurrences.

Another group of studies focused on the link between bunching and service network characteristics, investigating the phenomenon at shared corridors. In this direction, Verbič et al. (2016) concluded that bus bunching increases both dwell time at stops and bus travel time, while all routes running along a shared corridor are affected by bunching events. Under a similar scope, Diab et al. (2016) also investigated the impact of overlapping lines on bus bunching, concluding that these increased headway delays and service times. Schmöcker et al. (2016), argued that overlapping lines negatively impact service regularity unless overtaking between successive buses is allowed. Under the same scope, the study by Wu et al. (2017) corroborated the positive effect of bus overtaking and distributed passenger boardings on service regularity through simulation, underlining that proposed benefits would be higher for high-frequency services.

Overall, existing work mostly treats bus bunching as a route-level or corridor-level problem, ignoring any potential variability in explanatory variables over space. Although there is a large body of literature investigating the relationship between route characteristics and bus bunching, the existence of spatial relationships and patterns linked to bus bunching has been so far overlooked, while no study has developed econometric models at the network level. To the authors’ knowledge, the study by Iliopoulou et al. (2018) was the only to consider the existence of spatial autocorrelation among stops with frequent bus bunching events for a high-frequency bus route. In a different approach, Iliopoulou et al. (2020) employed clustering algorithms to identify spatio-temporal patterns of bus bunching in urban networks, detecting distinct types of bus bunching with different underlying characteristics. However, these studies presented empirical findings based on statistical tests and density-based clustering algorithms, with no econometric model developed.

In the same direction, diverging from the typically performed route-level analysis, this study develops econometric models, which analyze bus bunching as a network-wide problem, exhibiting spatial patterns. To investigate the factors affecting bus bunching, while considering spatial variability, a Geographically Weighted Regression (GWR) model is developed and applied for the first time, to the authors’ knowledge, to model bus bunching. GWR models can improve model accuracy in the presence of spatial structures and enhance the understanding of a problem, shifting from a global to local perspective (Cardozo et al., 2012). In recent years, GWR models have been applied to investigate the spatial variability of factors affecting public transport accessibility and ridership (Blairey and Mulley, 2013; Cardozo et al., 2012; Chiou et al., 2015; Chow et al., 2006; Du and Mulley, 2006; Tu et al., 2018). Interestingly, so far such analysis has been only applied for the existence of spatial structures in transit demand, despite the fact that the level of service provided by public transport significantly varies by region (Cardozo et al., 2012; Chiou et al., 2015; Mesbah et al., 2012). In this context, this study takes into account the presence of spatial patterns in public transport supply, aiming to identify spatial variations in variables linked to bus bunching. In order to understand and model how different factors affect bunching, data regarding stop attributes were collected and analyzed for a network of approximately 360 stops in central Athens. Subsequently, a GWR model was estimated, including three local parameters, while a comparison with a traditional global regression model was performed, demonstrating the suitability of the proposed model.

This remainder of this paper is organized as follows: the next section presents the data set used, the investigation of spatial patterns in the variables and the development of the GWR Model. Model results are subsequently presented and visualized. Last, conclusions are summarized and suggestions for further research provided.
2. Methodology

2.1. Data analysis

In this study, AVL data for 40 bus routes in Athens were analyzed for a single-month period (October 2017). The AVL data corresponded to stop-level records and include the arrival time at a stop for each vehicle, the vehicle code, the vehicle registration plate, the stop code and name, the stop coordinates, the route and direction. In order to retain data reflecting typical operations, problematic entries, records corresponding to holidays/special occasions and public transport strikes were removed from the dataset. As such, only records for typical weekday operations were retained, resulting in a total of approximately 2 million observations. To identify bunching occurrences, headways, i.e. the time difference between consecutive arrivals of same-line buses per stop, were subsequently computed. A threshold of 1 minute was employed to distinguish these events, as extreme cases of bunching were of interest. Subsequently, bunching events per stop were aggregated for the period considered and this figure was used as the model’s dependent variable. A set of 359 stops was then selected in order to capture locations in both the central business district of Athens as well as points of interest and attractions, such as universities, museums and major transport hubs. These stops as well as the total number of bunching events per stop are shown in Fig. 1.

Local stop attributes of the area were also collected, representing physical and operational characteristics and used as explanatory variables. These are presented in Table 1.

As seen in Table 1, three types of bus stops with respect to proximity to intersections were considered, using a 60-meter (200-foot) threshold (Pulugurtha and Vanapalli, 2008; Pulugurtha, and Sambhara, 2011). In terms of stops located at signalized intersections, a 30-meter (100-foot) value was used for defining a more confined influence area, due to the high density of the city’s road network (Pulugurtha and Vanapalli, 2008). The number of routes and the distance to the nearest transit/metro station were also used as variables to capture the connectivity and activity level of stops (Cardozo et al., 2012; Chiou et al., 2015). Furthermore, the proximity to a university was selected as an explanatory variable based on empirical findings from Iliopoulou et al. (2020), with more bus bunching cases observed near universities. Last, variables indicating the existence of a non-exclusive bus lane, the number of lanes at the stop level and the inclusion in the traffic ring area were
used to reflect traffic conditions in the vicinity of the stops. The traffic ring area also coincides with the central business district of the city, encompassing many points of interest.

2.2. Geographically weighted regression

Ordinary Least Squares Regression (OLS) assumes independence between observations (Fotheringham and Rogerson, 2008). However, this assumption does not typically hold in the case of spatial data (Rogerson, 2014), as the value of one observation may depend on the value of a neighboring observation. In this context, the application of GWR is proposed, which considers the presence of spatial dependencies between observations (Rogerson, 2014). To test whether when spatial dependence is present in this case and establish the need for applying GWR, a spatial autocorrelation test is carried out for the dependent and explanatory variables with relevant results shown in Fig. 2 and Table 2, respectively. Subsequently, the GWR model is presented.

2.3. Spatial autocorrelation

Spatial autocorrelation refers to the presence of systematic spatial patterns in explanatory variables, measuring their similarity within an area, the nature and the level of interdependence between the observations. The most common spatial autocorrelation indicator is Moran’s $I$ index, which recognizes spatial patterns and may be interpreted similarly to a correlation coefficient when the spatial weights matrix is row-standardized (Lee, 2001; Rogerson, 2014). In the latter case, the values

<table>
<thead>
<tr>
<th>Variables</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic signal</td>
<td>Binary</td>
<td>Stop located 30 m or less behind signalized intersection.</td>
</tr>
<tr>
<td>Traffic ring</td>
<td>Binary</td>
<td>Area with special traffic regulations in the center of Athens.</td>
</tr>
<tr>
<td>Bus lane</td>
<td>Binary</td>
<td>Existence of non-exclusive bus lane at stop.</td>
</tr>
<tr>
<td>Mid - block</td>
<td>Binary</td>
<td>Stop located 60 m or more from intersection.</td>
</tr>
<tr>
<td>Near - side</td>
<td>Binary</td>
<td>Stop located less than 60 m before intersection.</td>
</tr>
<tr>
<td>Far - side</td>
<td>Binary</td>
<td>Stop located more than 60 m after intersection.</td>
</tr>
<tr>
<td>Number of lanes</td>
<td>Continuous</td>
<td>Number of lanes at the stop level.</td>
</tr>
<tr>
<td>Number of routes</td>
<td>Continuous</td>
<td>Number of bus routes serving the stop.</td>
</tr>
<tr>
<td>University</td>
<td>Continuous</td>
<td>Distance to the nearest university.</td>
</tr>
<tr>
<td>Transit</td>
<td>Continuous</td>
<td>Distance to the nearest railway or metro station (lines 1, 2 &amp;3)</td>
</tr>
<tr>
<td>Metro lines 2 &amp; 3</td>
<td>Continuous</td>
<td>Distance to the nearest metro stations (lines 2 &amp; 3).</td>
</tr>
</tbody>
</table>

Fig. 2. Moran’s $I$ Index for the Dependent variable.
Table 2
Moran’s I index for independent variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Moran’s I</th>
<th>Z-score</th>
<th>P-value</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lanes</td>
<td>1.309</td>
<td>8.947</td>
<td>0.000</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>Number of routes</td>
<td>2.044</td>
<td>13.992</td>
<td>0.000</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>Metro lines 2 &amp; 3</td>
<td>1.178</td>
<td>8.081</td>
<td>0.000</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>Traffic ring</td>
<td>1.423</td>
<td>9.723</td>
<td>0.000</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>Bus lane</td>
<td>1.223</td>
<td>8.373</td>
<td>0.000</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>Traffic signal</td>
<td>0.306</td>
<td>2.120</td>
<td>0.034</td>
<td>Clustered</td>
</tr>
<tr>
<td>Mid-block</td>
<td>0.442</td>
<td>3.033</td>
<td>0.002</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>Far-side</td>
<td>0.377</td>
<td>2.599</td>
<td>0.009</td>
<td>Highly clustered</td>
</tr>
<tr>
<td>University</td>
<td>1.127</td>
<td>7.728</td>
<td>0.000</td>
<td>Highly clustered</td>
</tr>
</tbody>
</table>

obtained range between −1 and 1, otherwise extreme values of the indicator can be analytically derived (De Jong et al., 1984). High negative values correspond to strong negative spatial autocorrelation, while high positive values correspond to a strong positive spatial autocorrelation. The index is based on the acceptance or rejection of the null hypothesis, which assumes the lack of a spatial pattern, i.e. randomly distributed values of the variable in space. In fact, the combination of statistically significant p-values and positive z-values indicate a clustered spatial pattern, while significant p-values and negative z-values indicate a scattered pattern (Pan et al., 2019). The sign of Moran’s I Index reveals the type of pattern, while the former values show its significance.

The Moran’s I index, for the dependent variable, i.e. the logarithm of bunching events for the study period, is presented in Fig. 2. It should be noted that we applied a log-normal transformation to the bunching event data to support the normality assumption and improve model fit (Tu et al., 2018).

Evidently, there is a significantly clustered pattern in the data, while there is less than 5% probability that this may be attributed to random factors according to the p-value. Similarly, the Moran’s I index for the independent variables was calculated, revealing spatial patterns and establishing the suitability of a geographically weighted model (Table 2). As seen in Table 2, variables exhibit varying degrees of spatial autocorrelation, while some appear to affect bunching equally inside the given area. These observations establish the need to investigate both global and local variables during model development.

2.4. Model description

A GWR Model relates the coefficients of the independent variables of a linear equation to the observation coordinates (Rogerson, 2014). So, a GWR model, such as the one employed herein using the GWR 4.0 application (Nakaya et al., 2009), adopts an equation which includes parameters with variable coefficients. Thus, the following Equation applies:

\[ Y_i = b_1(u_i, v_i) * x_1 + b_2(u_i, v_i) * x_2 + \ldots + e_i \]  

(1)

where \( Y_i \) is the dependent variable, \( b_i(u_i, v_i) \) denotes the coefficient value for the independent variable \( x_i \) at a stop with coordinates \((u_i, v_i)\). The specification of local coefficients for the independent parameters requires the composition of a weight matrix \( W(u_i, v_i) \) based on proximity rules, which reflects the definition of a neighborhood around each point (Rogerson, 2014). Algebraically, the GWR estimates are obtained as follows (Fotheringham and Rogerson, 2008):

\[ \hat{b}(u_i, v_i) = \left[ X^T W(u_i, v_i) X \right]^{-1} X^T W(u_i, v_i) Y \]  

(2)

In the present study, the bi-square nearest neighbor function (Fotheringham et al., 2003) is selected to calculate the observation weight, based on an adjustable bandwidth this reflects the distance beyond which an observation is assumed to have zero influence on another observation.

3. Application

In this study, the logarithm of the aggregated bus bunching events for the study period was used as the dependent variable, while stop attributes were used as explanatory variables. The stand-alone GWR 4.0 application (Nakaya et al., 2009) was used for developing the GWR model, while ARCGIS was used for visualization and Python 2.7 for data analysis. The model parameters were estimated using 90% of the data points, randomly selected. The remaining 10% of the data were used for model validation. Initially, a global model was built using the OLS method, with corresponding results shown in Table 3. Subsequently, a GWR model was developed, under the same explanatory variables. Results show the suitability of the latter, as established through the comparison between the two models using the adjusted R-squared and Akaike Information Criterion (AIC) figures. As can be seen in Table 3, the improvement in the model’s explanatory power becomes clear from the value of the adjusted R-squared, confirming the need to apply a methodology which considers the spatial dependence of variables. The local R-squared values of the stop-level GWR models range from approximately 0.1–0.87, while the overall mean local R-squared value was 0.57.
Another advantage of GWR is the generation of a set of local estimates for the coefficients at each point. A summary of key values for these estimates, namely the median, upper and lower quadrant values, as well as the minimum and maximum values is given in Table 4.

Based on these results, conclusions may be drawn for the degree of spatial variability in the explanatory variables. As seen in Table 4, the values of the variables considered extend over a wide range, a fact which underlines the need to consider spatial variability. Typically, if the range of local estimates between the quadrant regions is greater than twice the standard deviation, the relationship may be considered as non-static spatial. Table 5 shows that this condition is met for all variables, except for the number of routes. However, following further analysis, it was concluded that the treatment of this variable as a local one, improved model results, since the significance of the variable also varied significantly over space.

4. Results and discussion

In the following section, model results are mapped, including the values of R-squared, coefficient estimates as well as t – statistic values for each point in the area. Subsequently, the model is validated in out-of-sample data.

4.1. Results’ visualization

It is worth pointing out that a weighted model essentially generates a regression equation for each pair of coordinates. Thus, the value of R-squared at each point indicates the explanatory power of the model per observation. In the present study, the model yielded an average value of adjusted R-squared equal to 0.57 as shown in Table 3. The values of the indicator per observation are shown on Fig. 3, presenting the spatial variation in the model’s explanatory power.

As shown in Fig. 3, the R-squared value is greater than 0.15 for most of the study area. Higher values are concentrated in the northern–western regions (0.6–0.87) featuring main transport corridors and bus lines linking the suburbs to the center of the city. Values ranging between 0.3 and 0.6 are noted in the downtown area, while lower values (0.1–0.3) are observed in

Table 3
Diagnostic information for OLS Regression and GWR.

<table>
<thead>
<tr>
<th>Index</th>
<th>OLS Regression</th>
<th>GWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual sum of squares:</td>
<td>265.535</td>
<td>50.495</td>
</tr>
<tr>
<td>Number of parameters:</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Classic AIC:</td>
<td>863.357</td>
<td>654.228</td>
</tr>
<tr>
<td>AICc:</td>
<td>863.546</td>
<td>674.215</td>
</tr>
<tr>
<td>CV:</td>
<td>0.841</td>
<td>0.451</td>
</tr>
<tr>
<td>R-squared:</td>
<td>0.119</td>
<td>0.654</td>
</tr>
<tr>
<td>Adjusted R-squared:</td>
<td>0.108</td>
<td>0.570</td>
</tr>
</tbody>
</table>

Table 4
Summary statistics for local coefficients.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Lanes</th>
<th>Routes</th>
<th>Metro lines 2 &amp; 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>Mean</td>
<td>0.646320</td>
<td>0.278889</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>1.187707</td>
<td>0.391820</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>−1.662861</td>
<td>−0.348334</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>3.784550</td>
<td>1.615645</td>
</tr>
<tr>
<td></td>
<td>Range</td>
<td>5.447411</td>
<td>1.963979</td>
</tr>
<tr>
<td></td>
<td>Lower Quartile</td>
<td>−0.237333</td>
<td>−0.065448</td>
</tr>
<tr>
<td></td>
<td>Upper Quartile</td>
<td>1.593696</td>
<td>0.602254</td>
</tr>
<tr>
<td></td>
<td>Interquartile R</td>
<td>1.831029</td>
<td>0.667703</td>
</tr>
<tr>
<td></td>
<td>Robust STD</td>
<td>1.357323</td>
<td>0.494961</td>
</tr>
</tbody>
</table>

Table 5
Spatial variability test for explanatory variables and intercept.

<table>
<thead>
<tr>
<th>Variable</th>
<th>S.E.</th>
<th>2 * S.E.</th>
<th>Interquartile R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.565</td>
<td>1.130</td>
<td>1.831029</td>
</tr>
<tr>
<td>Lanes</td>
<td>0.188</td>
<td>0.376</td>
<td>0.667703</td>
</tr>
<tr>
<td>Routes</td>
<td>0.061</td>
<td>0.122</td>
<td>0.110221</td>
</tr>
<tr>
<td>Metro lines 2 &amp; 3</td>
<td>0.0004</td>
<td>0.0007</td>
<td>0.000710</td>
</tr>
</tbody>
</table>
the southern regions. The following maps focus on the local effect of each factor visualizing the spatial variability in the corresponding coefficient estimates.

The two maps of Fig. 4 provide information about the estimate of the number of lanes at the stop level, as well as its statistical significance over the study area. In both cases, there is a large region in the northwest (red area on the right map) and a smaller southeastern area (green area on the right map), where the variable’s significance may be distinguished these regions feature the highest and lowest coefficient estimates respectively. Both sub-areas include university campuses attracting thousands of passengers daily (students and employees). The western region contains the city center and university campuses, where heavy passenger loads are observed, while transferring to several metro stations is also possible. In this case, high positive values for the coefficient underline the fact that a higher number of lanes increases the number of bunching events this may be explained by increased traffic flows in the region. On the contrary, the coefficient for the number of lanes is negative for the small region located east, reflecting a negative association between the number of lanes and bunching, which is perhaps expected in less congested conditions.

Next, the spatial variability of the coefficient for the number of different bus routes serving each stop is presented in Fig. 5 (left), along with its significance (right).

As shown above, the coefficient is significant in most of the city’s eastern region of the study area (red area on the right map of Fig. 5), as well as in some parts of the northwestern and southern regions (green areas on the right map of Fig. 5). Positive values suggest that bus bunching is more frequent at stops served by multiple bus lines. This finding implies that the interaction between overlapping routes may lead to delays and schedule deviations. Similar findings were reported by Diab et al. (2016) and Schmöcker et al. (2016) who concluded that simultaneous arrivals of different-line buses increased dwell times and resulted in delays. Indeed, the specific region features corridors with several overlapping lines, traversing the city center and serving major university campuses in Greece. On the other hand, in areas with low service frequencies and high passenger loads, a higher number of routes can prevent crowding at stops and thus alleviate bus bunching. This complex relationship between headway delay and overlapping services has indeed been reported in the literature (Diab
et al., 2016; Verbich et al., 2016). In fact, Diab et al. (2016) suggested an optimal headway of 20 minutes between overlapping services to minimize delays.

Finally, Fig. 6 illustrates the effect of bus stop proximity to metro stations (lines 2 and 3) on bus bunching. More particularly, it depicts how the distance between the nearest metro station for a bus stop is linked to bunching. Apparently, the coefficient is statistically significant at the outer regions of the study area. The positive coefficient values (red/orange areas on the right map of Fig. 5) indicate higher bunching rates in regions located far from metro stations. This finding suggests that the increase in bunching events may be attributed to the reduced public transport accessibility in the suburbs of Athens in conjunction to the high demand for transportation towards and from the city center. Indeed, in these areas there are relatively few bus lines linking residential neighborhoods to the closest metro stations, while passenger demand levels are high during peak hours, leading to delays and deviations from schedule. In contrast, in some regions located near the city center and served by multiple bus lines (green areas on the right map of Fig. 5) the negative coefficient values suggest that bunching decreases for distances to the subway stations up to a certain value. This finding may be attributed to the increased number of passengers transferring to or from the subway, as well as the congested conditions in the vicinity of subway stations located in the central area. In general, a complex relationship between subway and bus service characteristics has been reported in relevant studies applying GWR for forecasting transit ridership (Blainey and Mulley, 2013; Tu et al., 2018).

Fig. 4. Coefficient values (left) and t-statistic values (right) for the number of lanes variable.

Fig. 5. Coefficient values (left) and t-statistic values (right) for the number of routes variable.
Indeed, in some regions bus lines act as feeders for the subway station, while in other regions the two modes compete for patronage.

4.2. Model validation

The model was validated using 10% of the sample and predicting the number of bunching events at these points. A regression with zero intercept was performed to fit the predicted against the observed values, and the correlation coefficient and the slope were calculated to evaluate the predictive power of the GWR model. Corresponding plots are shown in Fig. 7.

Results show that the model does a fair job at predicting bus bunching events at out-of-sample data points, with a calculated R-squared value of 0.84 and a slope of 0.78. Still, the model underestimates the expected events in some cases with very high counts, a fact which may perhaps be explained by the relatively small share of points with very high counts in the training set. We expect that the expansion of the analysis to a larger sample would improve the model’s predictive ability.

5. Conclusion

This study investigated bus bunching by applying GWR and considering spatial variability in dependent and independent variables of the model. For the purposes of this work, bus bunching events were aggregated over a one-month period, for approximately 360 stops in Athens, Greece. The data set also included bus stop attributes and area features such as the number of lanes at the stop level, the presence of exclusive bus lanes, the number of bus routes serving the stop, as well as the distance to the nearest university campus and metro station among others. In order to analyze spatial variability in the data set, spatial patterns were examined by computing the global Moran’s I Index for each variable. The identification of highly clustered patterns established the need to consider local parameters within a GWR model framework. The model outcome was compared to a classic OLS regression model, clearly indicating the superior performance of the GWR approach. The latter resulted in an adjusted R-squared mean value equal to 0.57 considering three local variables, suggesting that the GWR model
could better explain the spatial variations in the data. The model’s predictive ability was also validated using 10% of the data points. Results showed that in areas around the city center and the western region, where heavy traffic flows occur, multiple lanes appear associated to higher congestion levels and thus an increase in bunching. On the contrary, in the eastern area, bus bunching is reduced in the presence of a higher number of lanes due to the possibility of vehicle overtaking in less congested conditions. Furthermore, the number of routes per stop appears to be linked to an increase in bunching events in stops served by several bus lines, fact which may be attributed to delays as a result of interactions between vehicles of overlapping lines. The distance to the nearest metro station was found to increase bunching in the suburbs, fact which is apparently caused by the absence of metro stations in these regions in combination to the high demand for transportation.

This application of the GWR model has quantified precisely the extent to which these factors play a role in spatial variations. The results can facilitate the estimation of bus bunching events at stops in urban networks with varying characteristics. Better estimation of bus bunching can support planning decisions for the improvement in the public transport level of service, through both infrastructure enhancements and operational/real-time strategies. Localized information regarding the effects of the different variables can be exploited to determine the best course of action. For instance, multi-line control schemes could be employed for lines running on shared corridors in an effort to mitigate reliability issues created by overlapping services. Similarly, scheduling improvements and capacity increases can be implemented for feeder routes serving regions with reduced transit accessibility.

Overall, this study investigated bus bunching through an innovative perspective. Indeed, on the one hand this study introduces GWR as a methodological tool for analyzing bus bunching and its causes, while on the other hand, the problem is decoupled from bus line operation and investigated at the network level, focusing on individual bus stops. The GWR model offers a flexible exploratory tool to explore the spatial aspects of data and address the issue of spatial heterogeneity. Evidently, there are various directions for expanding this research. For instance, expanding the data set to cover a larger set of stops would enhance the reliability of results. Most importantly, the use of GWR should be further investigated by adding more explanatory variables, and particularly passenger demand data. Indeed, the availability of data from APC or smart cards could certainly improve the explanatory power of the model and provide insights on the link between passenger demand and bunching patterns.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was carried out within the framework of “ESPA (2014-2020) Fellowships of Excellence for Post-Doctoral Research in Greece by IKY, State Scholarships Foundation (Recipient Christina Milioti) and of the Action “PhD Scholarships” by the General Secretariat of Ministry and Education and the Hellenic Institute for Research and Innovation (Grant No 1824-Recipient Christina Iliopoulou). The authors would like to thank the Athens Urban Transport Organization – OASA for providing all the necessary data for the analysis.

References


