**Caset F, Blainey SP, Derudder B, Boussauw K & Witlox F**

**‘Integrating Node-Place and Trip End Models to Explore Drivers of Rail Ridership in Flanders, Belgium’**

Accepted version of paper published in Journal of Transport Geography, 2020.

**1. Introduction**

*1.1 Background*

Railway stations and their surroundings are a major focus of attention in scholarly work focusing on the integration of transport and land use development. One of the ways in which this integration can be realized in practice is through ‘transit oriented development’ (TOD). This planning paradigm refers to several mechanisms to intensify the density and diversity of housing and other activities near transit stops, with the overall objective of promoting transit ridership and active travel over the use of private cars (Cervero 2009).

One element of TOD research focuses on identifying the development potential of railway stations as an outcome of transport and land use dimensions. The ‘node-place model’ (NPM) (Bertolini 1999) is an analytical framework frequently used to make sense of these differentiated development opportunities of station(s) areas(s) on a regional scale. It was primarily developed in the Netherlands, and forms the basis for an extended range of *‘knooppuntmodellen’* or NPMs in the Dutch context (Peek 2006). During the last two decades, the model has been applied in other geographical contexts (mainly in Europe, North America, East Asia and Australia) and on different geographical scales (the railway corridor, regional and national scales), both in academic and policy settings. Some recent NPM applications include the work of Zhang et al. (2019), Pucci and Vecchio (2019), Jeffrey et al. (2019), Li et al. (2019) and Nigro et al. (2019). The model aims to operationalize and analyze the double function of stations as (1) points of access to trains and other modes of transport (the ‘node’ function), and as (2) urbanized places with a particular intensity and diversity of human activities present (the ‘place’ function). Both dimensions are conceptualized and operationalized on the basis of a series of underlying processes and indicators. The underlying rationale is that a systematic comparison of these characteristics for a set of transit hubs allows for empirical identification of potential station development opportunities as an objective starting point for policy and planning discussions.

Although different NPM applications vary to some extent there is a common objective in that they aim to support a transition to increased ridership and therefore, allegedly, a transition to more sustainable travel behavior. Surprisingly however, within the NPM literature, analyses of the importance of node and place indicators in explaining ridership remain thin on the ground. A thorough literature review of academic papers that directly draw on the NPM demonstrated that only few of them elaborate on the correlations between ridership and the node-place indicators (Zemp et al. 2011, Falconer et al. 2016 and Caset et al. 2019). Other studies incorporate ridership as one of the node indicators to arrive at classifications of stations (Reusser et al. 2008, Monajem and Nosratian 2015, Singh et al. 2017 and Kim et al. 2018) or as a means to validate station classifications (Higgins and Kanaroglou 2016). To our knowledge there is only one study that has used node-place indicators as a means to explain ridership determinants (Olaru et al. 2019). The latter study used regression analysis to ascertain the associations between train patronage (AM peak, full day weekday and weekend) and the node and place indicators.

This relative lack of analytical cross-fertilization between the NPM literature and the one explaining passenger numbers and characteristics is somewhat surprising given the substantial body of literature addressing the challenge of explaining ridership and forecasting at railway stations. The next subsection will elaborate on this literature in more detail.

*1.2 Understanding rail ridership: trip end models*

Quantifying the benefits of TOD in terms of ridership has traditionally been assessed using regional four-stage travel demand models (for example McNally 2000), an approach that entails several potential problems (Marshall and Grady 2006, Cervero 2006, Gutiérrez et al. 2011). In most cases rail travel only accounts for a small percentage of trips made and regional models tend to be primarily designed to analyse road-based modes. Moreover, as argued by Gutiérrez et al. (2011) regional models are generally insensitive to land use and four-step travel models are cumbersome and expensive. Their performance thus tends to be relatively poor when assessing rail travel disaggregated to the station level.

This is particularly the case when attempting to forecast the demand for new stations or services where there is no existing level of rail demand to use as a basis for simulation. The majority of these rail-specific models can be placed into one of three categories, known as ‘trip rate’, ‘trip end’, and ‘direct demand models’. The latter category is less relevant here as it forecasts the number of trips made on a given flow rather than total ridership at a station. Trip rate models forecast the number of trips made from a station as a function of the resident population in the station’s expected catchment area. Trip end models differ in that they include additional explanatory variables alongside population to provide a more comprehensive representation of the processes which may determine rail usage levels. For example, some studies have aimed to quantify the link with land-use patterns (Sung and Oh 2011, Frei and Mahmassani 2013, Sun et al. 2016, Li et al., 2016).

The use of trip rate and trip end models has a long history (Preston 1991, Lane et al. 2006). Ongoing research is still improving our knowledge on the matter, often by means of the adoption of enhanced techniques to improve model results. Examples include the use of geographically weighted regression to account for spatial variations in the demand impact of explanatory variables (Páez 2006, Chow et al. 2006, Kobayashi and Lane 2007, Blainey 2010, Cardozo et al. 2012, Jun et al. 2015), the combination of machine learning techniques with regression models (Chiang et al. 2011), and integration with station choice models to provide a more realistic representation of station catchment areas (Young and Blainey 2019). While such models are most commonly developed for use in a specific urban or regional context (for example Zhao et al. 2013), there are also examples of more general and transferable models which are capable of predicting rail usage to a high degree of accuracy (for example Blainey 2010).

*1.3 Integrating node-place and rail ridership models*

The relative lack of integration between the NPM and rail demand forecasting literatures is particularly unfortunate given the limitations of most rail ridership models when used to predict demand. While, as noted above, it is possible to generate transferable models that predict existing levels of rail usage to a high degree of accuracy, these predictions tend to be based on a relatively small number of explanatory variables. When forecasting demand at a new station (or the demand change resulting from an alteration in exogenous conditions at an existing station), stakeholders will often be interested in the likely impact on ridership of a particular local factor. This may however not be easy to capture using the limited range of explanatory variables in the ridership model. We therefore argue that the incorporation of node-place variables in demand forecasting models has the potential to (at least partially) overcome this problem by allowing the effect of a much wider range of exogenous impacts on rail trips to be captured and predicted. Likewise, transferring techniques from rail ridership models to NPM applications may improve the analytical strength of the latter framework, as it introduces knowledge about the likely success of particular node or place interventions in terms of impacting rail ridership.

*1.4 Research objectives and structure*

Against this backdrop, this paper has a double objective. First, there is a methodological objective in that we aim to add to the body of literature in which the explanatory power of node-place modeling indicators in terms of ridership is examined. To this end, we draw on the trip end modeling literature and apply regression analyses to determine the most important explanatory factors. The data that is used here draws on an earlier extensive node-place analysis for all railway stations in Flanders (Caset et al. 2019). The data obtained from Belgian national railway company NMBS allow us to explain ridership at a finer temporal level than is usually the case in similar analyses (see Lane et al. 2006, Blainey 2010). Second, there is an empirical and related policy-support objective in that we apply the model to the case of Flanders. The models developed should allow policy makers to conduct a robust assessment of likely usage at potential new stations in the region, and to investigate the possible demand impacts of changes to existing stations and the area surrounding them.

The remainder of this paper is organized as follows. The next section elaborates on the case of Flanders, the data and methodology used and their shortcomings. Afterwards, the results are discussed in section three followed by a discussion and conclusions in section four.

**2. Case, data and methodology**

*2.1 Empirical case*

The geographical scope of our research comprises the region of Flanders, which is operationalized as the Dutch speaking part of Belgium (Figure 1a). The 253 railway stations that are part of the analysis are indicated on the map. For each of these stations, we build on the range of node and place indicators that were developed and measured as part of an earlier analysis that is extensively discussed in Caset et al. (2019). We briefly summarize these indicators below to make this paper self-standing. The full list of independent indicators along with acronyms, brief descriptions and data sources are provided in Appendix A. Appendix B provides summary statistics of the main independent variables that are included in the regression models.

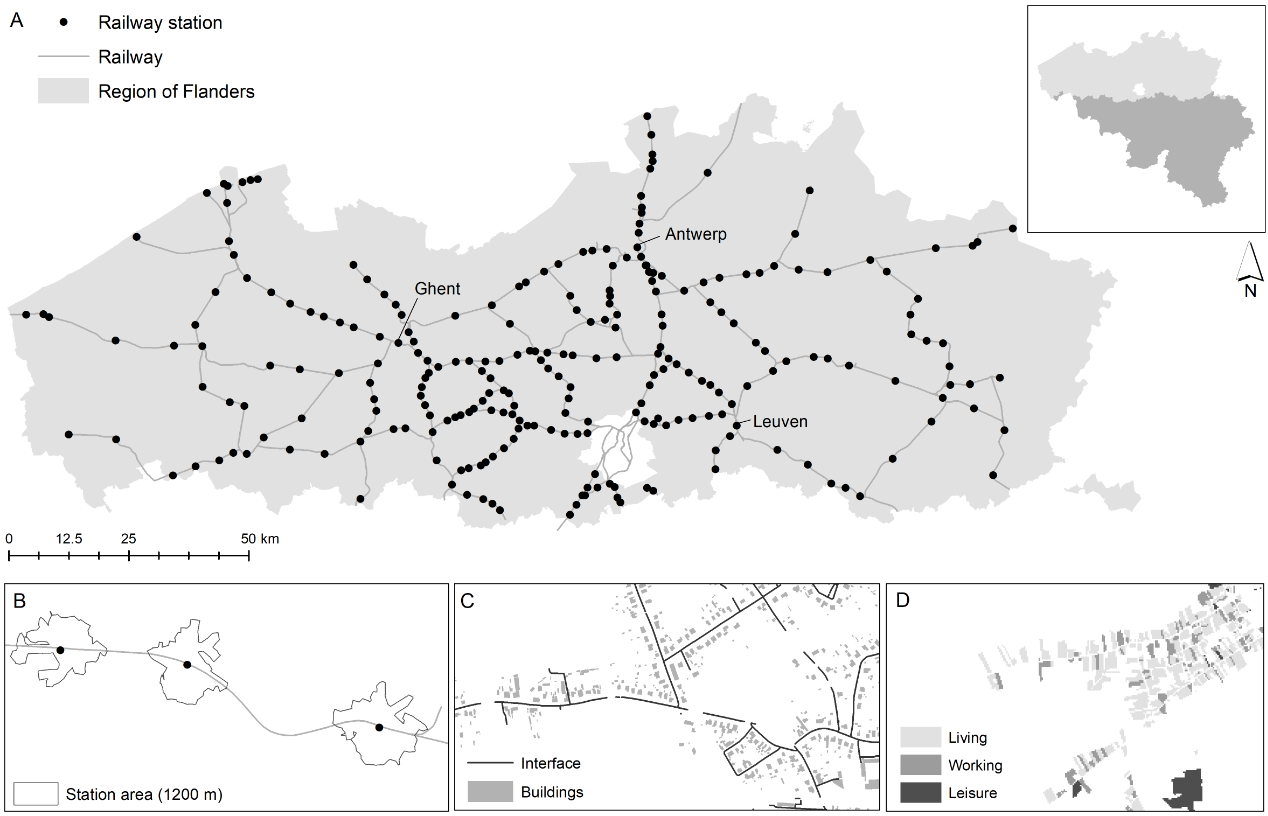


Figure 1: a) The Flemish railway network, b) illustration of walkable station areas,   
c) illustration of ‘interface catchment’ data layer, d) illustration of land use data layer

*2.2 Data*

2.2.1 Independent variables

In terms of node indicators two groups are distinguished: those measuring (1) feeder mode accessibility of the station (by car, bike and public transport), and (2) rail accessibility. As for the former, the car and bike indicators both reflect parking data capacity. The feeder public transport indicators include the number of bus, tram and metro (BTM) routes available at the station and the total frequency of BTM departures (both for a Tuesday). As for the second group, the indicators include frequencies on a Tuesday, distinguishing between the daily number of departing trains and the number of off-peak departures (between 9AM and 4PM), amplitude (the proportion of the day in which train services are offered), and two centrality measures coined ‘transfer centrality’ and ‘travel time centrality’. The latter two are inspired by the ‘degree’ and ‘closeness’ centrality indicators developed by Curtis and Scheurer (2010, 2016). Both measures capture the ease of reaching all other stations in the networks. Whereas the former does so in terms of the number of transfers needed, the latter takes into account the travel time and service frequencies between all pairs of stations in the network.

In terms of place indicators, three indicator groups are distinguished: those measuring aspects of ‘density’, ‘diversity’ and ‘design’ of the station area. The ‘station area’ is defined as the area that is reachable within a walkable distance of 1200 meter, which corresponds to approximately 15 minutes of walking and 5 minutes of cycling (see Figure 1b)[[1]](#footnote-1).

As demonstrated in Appendix A, the density dimension is composed of five indicators measuring the total number of inhabitants and jobs in the station area and the density of ‘basic’, ‘regional’ and ‘metropolitan’ amenities in the station area. The difference between the three types of amenities is specified in Verachtert et al. (2016). In general terms, basic amenities are considered necessary to organize daily life (e.g. a kindergarten). Regional amenities are assumed to have a larger catchment area serving different urban areas in the region (e.g. a shopping mall), whereas metropolitan amenities have the largest catchment area (e.g. tourist attractions).

Diversity of land use is measured both functionally (by means of a Shannon diversity index[[2]](#footnote-2)) and spatially (by means of an Interspersion and Juxtaposition index[[3]](#footnote-3)) and draws on a raster dataset containing three land use types: ‘living’, ‘working’ and ‘leisure’. The Fragstats software was used to calculate both indicators (see McGarigal and Marks 1995).

The design dimension is composed of three indicators and aims to reflect the extent to which the station area is ‘walkable’. In line with Pafka and Dovey (2018), the first indicator focuses on the extent of the public/private interface within the station area as a proxy measure for ‘how much’ is ‘caught’ within the area. More specifically, this interface catchment (IC) is calculated by summarizing the length of all walkable street segments (as a proxy for public space) that are also flanked by buildings (as a proxy for private space) (Figure 1c). The second indicator ‘permeability’ measures the extent to which the urban morphology is permeated by publicly accessible space (Marshall 2005) by mapping the total number of street crossings per station area (Ryan and Frank 2009). This measure relates to the ease of movement through an urban area due to the multiplicity of route choices between any pair of points. A third indicator measures the total length of walkable and bikeable street networks within the station area.

Competition between stations was not considered in defining station catchments and hence station areas may overlap. In practice, demand at local stations may however be affected by the proximity of other and larger stations. Therefore, station spacing indicators were calculated to represent travel times (by car and by train) from the local station to (1) the nearest station and (2) the nearest station of a higher order. This ‘order’ was determined by classifying the stations according to a Jenks classification method based on their service frequency on a Tuesday.

In addition to the above, ridership may also be affected by socioeconomic characteristics of the station area residents such as income, age, race or ethnicity and car ownership (Stead 2001, Ewing and Cervero 2001). The latter may be of particular importance in the Flemish context given the large influence of car ownership, or in a broader sense car availability, on car use in Flanders (Van Acker 2010). However, recent car availability data for Flanders is limited to the municipal scale, which is too coarse for the purpose of this research. Furthermore, company cars, representing a large share of the car fleet in Flanders (see May 2019) are registered at the company’s main address, reducing the usefulness of this dataset. As a consequence, car availability data could not be included. We nonetheless collected other socioeconomic data[[4]](#footnote-4) at the scale of the census ward[[5]](#footnote-5) pertaining to (1) the labor market situation of the station area residents and (2) the economic sector in which the station area residents are employed (both based on the latest census data from 2011). These data were summarized for all statistical sectors that geographically intersect with the station area.

2.2.2 Dependent variables

The ridership data reflects the number of boardings on a regular weekday in 2018, based on NMBS ticket sales data. It takes into account both individual and season tickets. For 32 stations, tickets could not be attributed to a specific station because they are located in a ‘tariff zone’ (for example, tariff zone Ghent comprises three stations). For these stations, publicly available passenger count data for October 2018 were used.

Besides this ticket sales data, NMBS also provided the proportions of people boarding on a regular workday during the morning peak (6 to 9 AM), evening peak (4 to 7 PM) and off-peak (between 9 AM and 4 PM). This allowed us to approximate boardings for these different periods of the day. These data distributions are highly skewed, illustrating that the Flemish railway network has very few stations with very high ridership numbers for the different periods of the day. The largest stations in terms of ridership are Ghent, Antwerp and Leuven (indicated in Figure 1a) with 56.314, 34.998 and 34.267 boardings, respectively, on a regular workday. The small stations of Aalst-Kerrebroek, Aalter and Hambos each accommodate 26, 28 and 32 travelers.

It should be noted that some factors may to a certain extent impact the accuracy and completeness of this data. For example, train subscriptions do not reveal when commuters actually travel, electronic tickets can be purchased hours before the actual trip will be made, and some ticket formulas (‘Go Pass’ and ‘Rail Pass’) do not reveal any information at all about the trip made. The ridership data provided by NMBS is nonetheless robust in the sense that it was generated in a consistent way for all railway stations at the highest level of detail possible.

In summary, our calibration dataset is cross-sectional and consists of a 36 x 4 matrix, with 36 potential ridership determinants and four groups of ridership data (total, morning, evening and off-peak).

*2.3 Methodology*

2.3.1 Regression analysis

In order to estimate the importance of the independent variables in explaining ridership, we applied two different types of regression analysis.

First, a series of ordinary least square (OLS) regression equations were estimated to test the relative significance of the independent variables in predicting ridership. These analyses were conducted for two groups of data: (1) the full set of 253 railway stations and the different time windows, and (2) a geographical segmentation of the data according to an empirical station typology that was established in earlier work (see Caset et al. 2019). The four station types that resulted from this previous analysis are distinguished by different node and place characteristics[[6]](#footnote-6) as illustrated in Figure 2. The cluster sizes are: type 1 (122 stations), type 2 (68 stations), type 3 (8 stations) and type 4 (55 stations). Given the very small N for the type 3 stations, we will only interpret and elaborate on the results for station types 1, 2 and 4.

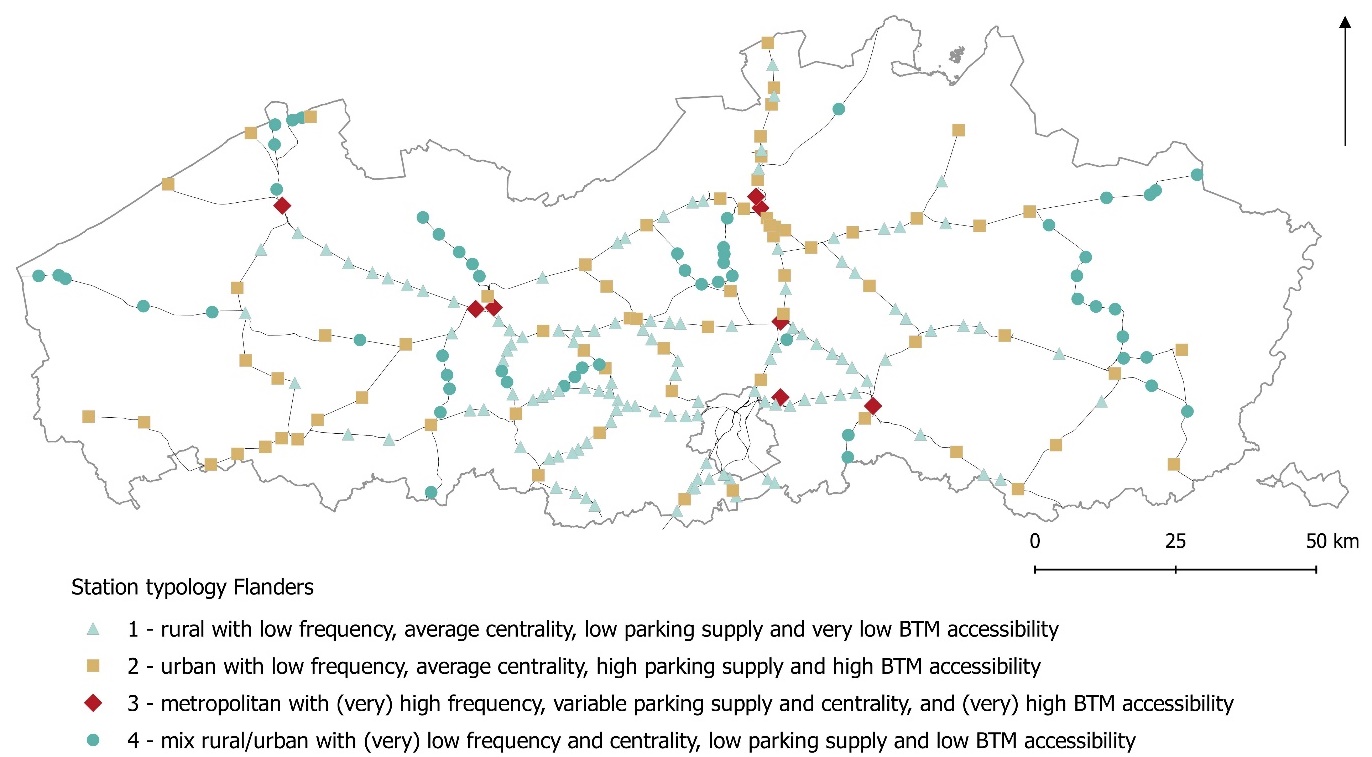


Figure 2: Station typology for Flanders (see Caset et al. 2019)

The independent variables were added in a stepwise manner using forward selection. More specifically, during each step in the process of adding variables, the variable that led to the highest model fit (in terms of the R2adj statistic) was included in the model. This procedure continued until no remaining candidate variables were left.

In order to limit multicollinearity we analyzed the correlations between all pairs of independent variables. The correlation matrix showed that there are indeed strong and significant correlations (>0.8), mainly within the group of place indicators and between the feeder public transport routes and frequency indicators. This has important implications for the co-presence of these indicators in the same model. We monitored this more closely by assuring that the variance inflation factors (VIF) did not exceed the commonly used value of 3. More specifically, when a variable was added that led to VIF values higher than 3, the variable that was last added was again removed.

Because of the skewed data distributions we log-transformed all indicators. This resulted in four log-log models given by:

ln (Y) = α + β1 ln (x1) + β2 ln (x2) + … + βi ln (xi) + ɛ (1)

with intercept α, coefficients β and error term ɛ. As all coefficients are fixed or ‘global’, this type of regression is called a ‘global regression model’.

Direct station-level ridership forecasting models using OLS regression have a significant drawback in that they do not account for the possibility that parameters may not be constant (or ‘non-stationary’) across different points in space (see Blainey 2010 for a fuller discussion of this issue). In order to examine whether such spatial variation exists we therefore recalibrated the four best fit OLS models for the full set of railway stations (one for each of the four time windows) using geographically weighted regression (GWR, Brunsdon et al. 1996, 1999, Fotheringham et al. 1998) with the software package GWR, version 4 (National Centre for Geocomputation 2009). GWR explicitly considers the spatial component of the data by incorporating the value of the geographical coordinates of observations in the model equation (Cardozo et al., 2012). The βj coefficients (j = 0, 1, …, p) of the j variables xj (j = 1, …, p) may thus vary for each location. In other words, instead of calibrating a single regression equation, GWR generates a separate regression equation for each observation with coordinates (ui, vi). The value of the dependent variable yi is estimated as follows:

ln (Yi) = β0 (ui, vi) + β1 (ui, vi) ln (x1) + … + βp (ui, vi) ln (xp) + ɛi

For each of the best fit models, a geographical variability test was conducted to determine if and which variables were fixed (‘global’) or varying (‘local’). We then estimated these global and local parameters using a Gaussian model with an adaptive bi-square kernel function, and a ‘golden-section search’ method to automatically search for the optimal bandwidth size.

2.3.3 Methodological limitations

We conclude this methodological section by appraising a number of methodological limitations, some of which emanate from interpreting results from cross-sectional models as causal mechanisms.

Due to the snapshot of data points at a single moment in time, cross-sectional models (in contrast to panel data models) are not suited to pinpoint the extent to which the associations found are causal or not. Therefore, they serve to provide rough estimates of ridership determinants at best (Liu et al. 2014). The associations described below thus warrant careful interpretation, as some of the variables included may be confounding in terms of the causal relationships at play. For example, the inclusion of train service variables may produce endogeneity problems, since service supply not only affects transit demand, but it may equally be a function of demand (Taylor and Fink 2003, Gutiérrez et al. 2011). Additionally, due to the generally high levels of correlation among (mainly) spatial and socio-economic variables, it is hard to untangle the relations between these various factors on the one hand and on transit ridership on the other hand (Crane 2000, Taylor and Fink 2003).

Also, some potentially relevant variables were not included in our analyses. Some of these were either hard to operationalize for all 253 stations, or data was not available at the required level of detail. Examples of the former issue include variables measuring transit service quality in terms of comfort, reliability and convenience or car driver friendliness and traffic congestion levels of the station area. In terms of the latter solid data capturing car availability of households at a sufficient geographical scale was not available. Additionally, we did not explore specifications of the model including interaction variables, given that these would have to be drawn from earlier conceptual and empirical research that suggested or showed specific interactions to be germane.

Lastly, the analyses conducted in this section represent ecological correlations, meaning that the findings can only be interpreted at the level of station areas and may not be extrapolated to the level of station area residents. These ecological correlations are in turn influenced by the geographical scales at which the point-based data is aggregated into statistical sectors (the socio-economic data) and raster pixels (the ‘place’ data) and into station areas with a particular size (both pertain to the ‘modifiable areal urban unit problem’).

**3. Results**

*3.1 The scale of Flanders and different time windows*

Table 1 summarizes the four best fit models in terms of R2adj values, the unstandardized variable coefficient β, the t-statistic, the level of significance and the VIF. All variables included are statistically significant at the 95% confidence level. The high R2adj values demonstrate that the model fits well with the observed data.

The total number of boardings (model 1) is best explained by six determinants. According to the β coefficients, *transfer centrality* has by far the largest impact on total ridership in Flanders. The model indicates that a 1% increase in transfer centrality will result in a 2% increase in ridership. Likewise, *train frequency* and *interface catchment* exert an impact of 0.6% and 0.4%, respectively, while the remaining variables (especially *BTM routes*) all have lower β values. Interestingly, when running the total ridership model with just *train frequency* and *total car parking capacity* included, an R2adj value of .758 is obtained, which is remarkably high for just two explanatory factors.

Morning peak (model 2) has largely similar determinants as total ridership with some exceptions: *BTM routes* is not selected, whereas *amplitude of train services* does play a role. The negative coefficient for *density of metropolitan amenities* seems to indicate that a sizeable proportion of the stations with high levels of morning ridership are located outside of the largest metropolitan areas. The highest elasticities are noted for *interface catchment* and *train frequency* (both 0.6).

The determinants included for the evening peak (model 3) are largely similar to those of the total ridership model, except for *job density*. The largest elasticity values are nonetheless (again) recorded for *transfer centrality* (a 3% increase in evening ridership when this variable increases by 1%) and – to a lesser extent – train frequency (0.7%).

Lastly, the off-peak model (model 4) has the highest overall fit and the largest number of variables included. The variables are similar to those of the total model, but a number of additional variables were selected: *amplitude of train services* and the station spacing variable measuring the *travel time by train to stations of a higher order*. The positive (and rather low) coefficient value of the latter implies that off-peak ridership generally increases when a station is connected poorly to stations of a higher order. Besides the very high coefficient value for transfer centrality (3.6%), amplitude of train services also has a high value (1.4%).

Some more general observations can be made. First, the strongest determinants for ridership in Flanders seem to be situated within both domains of ‘node’ and ‘place’, but with the former being more important: the group of node variables generally exhibits most of the statistically significant variables with higher coefficient values. The train variables in particular seem to exert the strongest impacts on ridership. Further node-related findings are that including *total parking capacities* results in higher model fits than the disaggregate *free* and *toll parking* variables (for both the bike and car modes). Also, feeder public transport accessibility in terms of unique routes available (*BTM routes*) is more frequently significant and leads to higher model fits than the daily *frequency* of departures, suggesting that an increase in the number of available feeder transit routes will have a higher impact on ridership than an increase in daily frequency. Next, the only train variable that did not feature in any of the models is the *travel time centrality* variable. In fact, this indicator is not significant in any of the model runs, while *transfer centrality* does feature prominently. This leads us to suggest that rail travel demand in Flanders is more strongly impacted by the number of transfers required than the required travel times and service frequencies of rail trips.

Place variables are scarcer than node variables in the best fit models. Given the strong correlations within this group of variables, most models only include one place variable in order to avoid multicollinearity. An interesting observation here is that none of the two diversity variables (the *Shannon diversity index* and the *IJI*), feature in the best fit models. Although land use diversity is considered an important ingredient of successful TOD planning, it does not seem to impact ridership in Flanders to a great extent. However, some model runs do result in significant outcomes and have model fits that are similar to those with density or design variables included. This is especially the case for some of the evening and off-peak model runs, where both diversity variables are significant and lead to high model fits. The morning peak model runs on the other hand, do not include any significant diversity variables. Interestingly, the *Shannon diversity index* always leads to higher fits than the *IJI*, which suggests that the functional land use mix measure appears to be a stronger explanatory factor for ridership in Flanders than the spatial land use mix.

Another notable finding is the strong significance of the design variable *interface catchment.* Although most other place variables also render statistically significant outcomes, *interface catchment* always leads to the highest R2adj values (except for model 3 where *job density* features more strongly). Of course, the interpretation of this land use variable in terms of its practical demand impact is less straightforward than that of an indicator measuring aspects of density (e.g. residential or job density). After all, increasing the public/private interface within a station area translates into a rather abstract planning task, compared to increasing the residential density of a station area by a certain percentage or number of housing units. It could therefore be questioned whether this variable may be instrumental in planning discussions. Furthermore, in line with the methodological limitations discussed above, and given the strong correlations between *interface catchment* and *residential density* (.917 at the .01 level) or *basic amenity density* (.905 at the .01 level), it is questionable whether this design variable is indeed a direct ridership determinant or rather a feature of high density station areas. As for the other two design variables, *permeability* alsorenders model fits that are close to those of *interface catchment*, while the indicator measuring the total length of bikeable and walkable infrastructure (*network length*) clearly performs weaker.

The station spacing variables do not feature in the majority of best fit models, but are nonetheless positively associated and statistically significant in most of the other model runs with a lower fit. Importantly, this implies that ridership is not only influenced by characteristics of the station itself (captured by the node and place variables), but also by inter-station characteristics such as their proximity in space and time.

To conclude, unlike the groups of variables discussed above, none of the socio-economic characteristics feature in the best fit models. Although most of these variables feature significantly and positively in model runs together with node variables, they are not significant when the place variables are added.

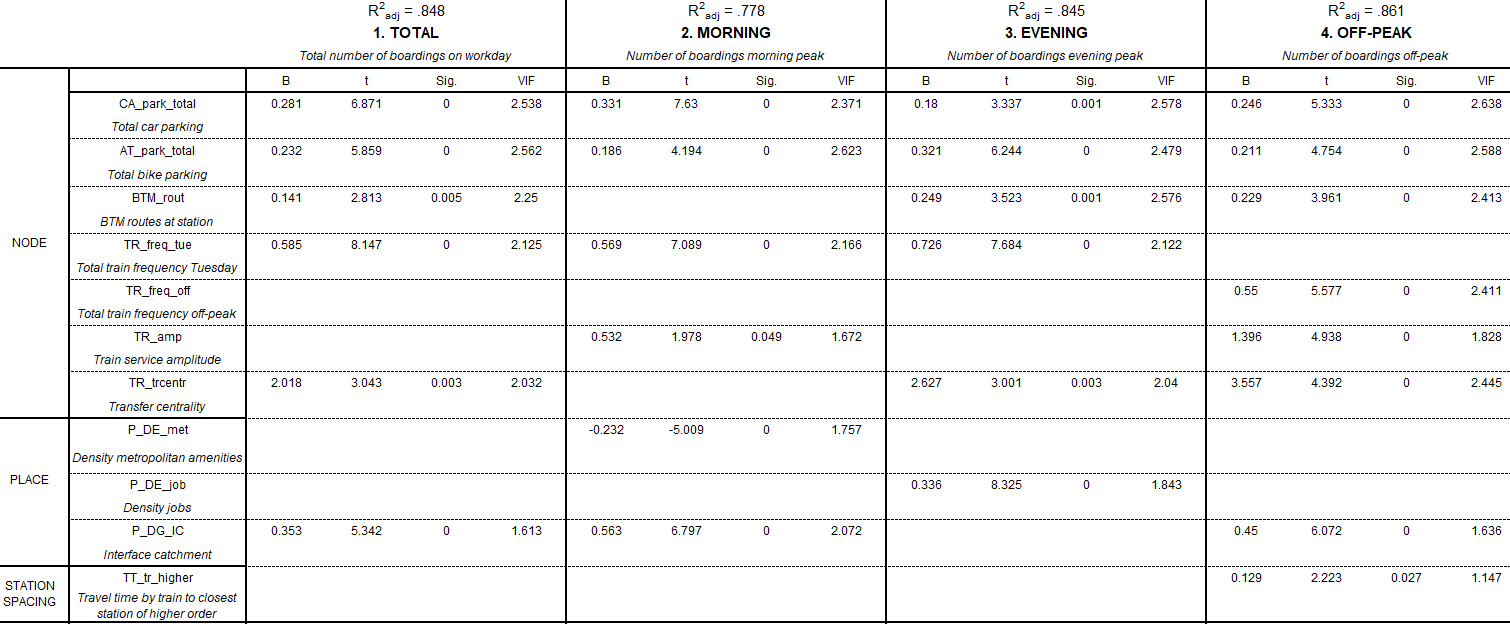


Table 1: Summary statistics of the best fit models

*3.2 Different types of railway stations, different determinants?*

Table 2 summarizes the results of the OLS regression analyses for the different station types. As indicated earlier, given the very small N for the type 3 stations we will only interpret and elaborate on the results for station types 1, 2 and 4. In line with Table 1, the statistically significant variables of the best fit models are provided with some key statistics. Since none of the socio-economic variables were included in the best fit models, we removed those from the table. The discussion below is structured around the outcomes per station type, followed by a summary of the main trends.

Type 1. This group is composed of the most rural stations in Flanders, and is further characterized by low train frequency, average railway network centrality, low parking supply and very low BTM accessibility. The type 1 models demonstrate more or less consistent fits across the four time windows with R2adj values fluctuating around .650. The best fits are found for the morning and the off-peak models. In general, ridership for this group seems most influenced by car and train accessibility, by all three place dimensions and by station spacing. The total and morning peak model are very similar and indicate that ridership is by far most influenced by the station’s *transfer centrality*, followed by *travel time by car to the closest station of a higher order.* The positive coefficient for the latter implies that stations will generally perform better when there is no convenient alternative to drive by car to a nearby station with a higher train service frequency. *Residential density* also plays a significant role, followed by *total car parking capacity*. For the case of evening peak ridership, *train frequency* and *job density* also play a role, as does the design variable *permeability*. Ridership during the off-peak hours seems in turn most strongly influenced by *transfer centrality* and *interface catchment*.

Type 2. These stations are predominantly located in urban areas and are characterized by an average centrality in the railway network, high parking supply and high BTM accessibility. Judging from Table 2, the model fits are generally (much) higher than the group 1 stations, with the highest model fits for the evening and off-peak data. The determinants between both groups also clearly differ. Instead of comprising the different node, place and station spacing groups, the determinants of type 2 are predominantly clustered within the node dimension, except for the evening peak model which includes high B values for the *Shannon diversity index* and *interface catchment*. In contrast to type 1, besides the car, feeder mode accessibility by bike and BTM are also important. Furthermore, train frequencies seem more important than the station’s centrality in the network.

Type 4. The main feature that distinguishes this group of stations from the others is their very low performance on all train variables. A large share of these stations is located at the periphery of the network. Model fits are rather poor compared to the other types, except for the off-peak model which has a R2adj value of .846. In general, *amplitude* seems to play a very large role, which may not surprise given the low network centrality positions of these stations hence the need to be able to reach the station early in the morning and/or late in the evening. *Car parking capacity* does not seem to play a large role while *bike parking capacity* seems rather important, especially for the evening peak model.



Table 2: Summary of the best fit models for the different station types

Some general observations can be made. First, in line with the findings for the full set of railway stations it seems that the place variables play a less dominant role in explaining ridership in Flanders than is the case for the node variables. Judging from Table 2, place variables are only relevant in explaining ridership for the rural stations (type 1) and for the evening peak models. Moreover, the B coefficients of the place variables are relatively small compared to those of other variables (some exceptions aside for the evening peak model).

Second, some of the variables that did not show up in any of the generic models do appear in Table 2 (some of these with high B values): *residential density*, *Shannon diversity index* and *permeability*. Some of the findings for the generic models are nonetheless corroborated by the geographically segmented analyses. For example, the explanatory power of the following variables is very limited in the Flemish context: *density of basic, regional and metropolitan amenities*, *toll and free parking supply*, *BTM frequency*, and *total street network length of walk- and bike infrastructure*.

Third, it seems that the model fits are generally lowest for the morning peak. This seems to imply that there are important determinants that were not factored in in our research design, or that were not measured in an appropriate way. This remark might especially pertain to the socio-economic variables, which capture a fraction of what could possibly be relevant judging from earlier travel behavior research for the case of Flanders (Van Acker 2010).

Fourth, although the geographical segmentation of data provides more insight into the relative importance of determinants for different types of stations, the model fits of the generic models are all higher (which may be explained by the larger calibration datasets of the generic models). This is most notable for the total and the morning peak models.

*3.3 Analyzing spatial nonstationarity*

The outcomes of the GWR analyses are summarized in Table 3, which displays some key statistics. Judging from these results, it seems that spatial nonstationarity needs to be considered. The GWR versions of the best fit log-log models all exhibit higher R2adj values than the global model calibrations. Furthermore, the results of the ANOVA testing the null hypothesis that the GWR model represents no improvement over the global model are statistically significant, demonstrating that the GWR models provide a better fit. The drop in the Akaike Information Criterion (AIC) with more than 3 units for the GWR models compared to the OLS models furthermore demonstrates that the GWR models significantly improve the model fits (Fotheringham et al. 2002). Similarly, the reduction in the Moran’s I statistics with all the GWR calibrations shows that the level of spatial autocorrelation in the model residuals has been substantially reduced. Additionally, the best bandwith sizes for each of the models are satisfactory, as the minimum number of data points considered in each local calibration of the model is no less than 76 (morning peak model).

The geographical variability tests indicate that the coefficients of all variables vary significantly over space, except for two variables. For the total ridership model, the coefficients of the *number of bus routes* variable (BTM\_routes) do not vary locally. The same is true for the off-peak model, which includes a second variable with globally fixed coefficients, i.e. the *travel time by train to stations of a higher order* (TT\_tr\_higher). The *number of bus routes* variable is very close to being categorized as one with global coefficients in the evening peak model, and is therefore noted between brackets in Table 3.

Judging from the table, spatial nonstationarity is most important for the morning peak model, since the GWR model exhibits a very high fit (R2ad = .863) compared to its log-log counterpart (.778). This difference in fit is much more pronounced compared to the other models. For this reason, we focus on the results of this GWR model in the remainder of this subsection.



Table 3: Summary statistics of the best fit log-log models and GWR models

3.3.1 The morning peak model

Figure 3 visualizes[[7]](#footnote-7) the results of the GWR analyses: the parameter estimates and t-statistics for all local variables included (a to f), the local R2 values (g) and the residuals (h).

Figures 3 a to f reflect the importance of the variables in explaining the variance in morning peak ridership. This importance can be expressed in terms of both the unstandardized B coefficients and in terms of the t-values. As demonstrated by Mennis (2006), maps of both key statistics are important if spatial nonstationarity is to be interpreted effectively.

A first observation is that there is only one variable for which the local coefficients are statistically significant for all stations: *interface catchment*. Judging from map f, this variable is most important and significant for the central area of Flanders, around and in between Antwerp and the Brussels Capital Region (BCR). This seems to indicate that the walkability of the station area plays a larger role in explaining morning peak ridership for these stations. Second, the largest B values are nonetheless found for *amplitude*. For some stations, especially those located around and just west of the BCR, B coefficients of 2 or higher are found. This implies that a one percent change in amplitude will impact morning peak ridership by two percent. The geographical distribution of these high coefficient values west of the BCR and in the east of Flanders could be explained by the mostly low performance of these stations in terms of train services. Judging from Figure 2, a sizeable share of these stations belong to type 4, which are characterized by very low train frequencies and very low network centrality performances. The subsequent high coefficient values for these adjacent stations might indicate that there is a lot of trading going on with respect to this variable. Similar assumptions[[8]](#footnote-8) may be deduced for the other variables. For example, *train frequency* seems most important and most significant for the western part of Flanders. Especially the stations in between the regional cities of Bruges and Kortrijk exhibit the highest t values. This observation seems to indicate that in this area, trading of between stations mainly occurs based on service frequency. Amplitude plays a non-significant role in explaining morning peak ridership.

The spatial variation of the model’s explanatory power can be determined by examining the spatial distribution of local fits produced with GWR. Judging from map g, the model has a higher predictive capacity in the east, in the central area of Flanders located just west of the Brussels Capital Region (BCR) and the small group of stations just north of the BCR. A high number of stations have local R2 values that are higher than .83 with maximal values of .89. These model fit values are comparable with the best-performing models developed in other analyses using similar calibration techniques. This indicates that the model does a good job of capturing the variation in rail demand across the study region.

|  |  |
| --- | --- |
|  |  |
| b) |  |
|  |  |
| a) |  |
|  |  |
| d) |  |
|  |  |
| c) |  |
|  |  |
| f) |  |
|  |  |
| e) |  |
|  |  |
| g) | h) |

Figure 3: Parameter estimates of local variables (a to f), local R2 values (g) and residuals (h)

A comparison of the residuals between the GWR and OLS models indicates that the former has smaller residual values. Whereas the GWR residuals vary between -1.3 and 1.3, the OLS residuals vary between -2.2 and 1.7. Map h illustrates the geographical distribution of the GWR residuals. Their seemingly random distribution over space indicates that there is no major dynamic that was not captured by the model (otherwise a spatially clustered pattern would appear).

**4. Discussion and conclusions**

This paper aimed to identify the relative importance of a wide range of node-place variables in explaining rail ridership in Flanders. To this end, we conducted a series of regression analyses based on temporal and geographical segmentations of the data. We also verified the importance of spatial nonstationarity. Below, we will discuss the main conclusions and indicate room for improvement and further reflection.

First, there is a clear differentiation between the determinants of ridership at the station level at different times of day. As demonstrated for both the full scale (253 stations) and the segmented OLS regression analyses, the explanatory power of variables may differ drastically. Evening peak data arguably allows to better explain the dynamics at destination stations, while morning peak data likely explains ridership dynamics best for origin stations. Off-peak ridership data provides an additional perspective and likely allows to better explain leisure motivations and non-regular trips. Besides this temporal differentiation, the regression analyses based on the NTP station typology demonstrated that geographical segmentation according to similarity in station characteristics also reveals differences in determinants. For example, ridership at ‘rural’ stations compared to ‘urban’ stations in Flanders seems impacted by rather different factors.

Whereas these analyses at finer temporal and geographical levels provide useful clues for exploring ridership in Flanders, an additional examination verifying spatial nonstationarity was conducted. Due to the time-intensive operation of the stepwise regression procedure described above, we opted to use the four best fit log-log models as an input for the GWR4 software. This resulted in the four models that are summarized in Table 3. While carrying out additional GWR analyses that build on different global regression models would bolster the robustness of the results, this should not substantially affect the validity of the conclusions drawn here.

In terms of the relative importance of node-place variables, the OLS models indicate that (as for the unstandardized coefficients), the train variables exert the largest impact on ridership with transfer centrality clearly dominating. In some of the models amplitude plays a large role as well. The relative dominance of these train-related variables in the models is perhaps unsurprising, given that this has been a common feature of trip end models developed previously in other contexts (Blainey, 2010). Feeder mode accessibility in terms of the car and bike parking supply and public transport provision is also important, but generally less so than other included variables. The selected place variables are much less numerous in the different models, which may in part be due to the high levels of multicollinearity for these variables. We nonetheless found empirical confirmation that some less commonly used place variables, such as interface catchment, explain a significant and sizeable degree of variation in station usage. Finally, besides the socio-economic variables, the explanatory power of the following variables is very limited in the Flemish context: density of basic, regional and metropolitan amenities, toll and free parking supply, BTM frequency, and total street network length of walk- and bike infrastructure. This generally reflects previous findings from similar models developed elsewhere (see section 1), which are seldom capable of directly capturing the impact of these variables (Young & Blainey, 2019). These observations call for future validation across different empirical cases in order to see to what extent these findings are reproduced.

These results also have a broader significance in light of the purported ‘dynamic’ character of the node-place model. As stated by Bertolini (1999: 203): “The starting point is the assumption that in the long term (…) and in the measure that demand and supply mechanisms are free to operate, the demand for transportation services from the activity place and the demand for activities from the transportation node will find a (temporary) balance”. However, as demonstrated in our analyses, the interaction potential between factors of supply (‘node’ and ‘place’) and demand (ridership) is not evenly distributed for all railway stations. For example, interventions in terms of a station’s node value will likely have a higher effect on daily boardings for some stations compared to others. Therefore, besides the ‘disturbing’ and exogenous factors mentioned in the quote above, we would argue that existing demand dynamics should be recognized more explicitly as an additional limiting factor when aiming to realize the hoped-for development paths towards a more balanced arrangement of node and place performance.

**Appendix A**



**Appendix B**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Minimum** | **Maximum** | **Mean** | **Std. Deviation** |
| CA\_park\_total | 0 | 2056 | 209 | 287 |
| AT\_park\_total | 0 | 7499 | 338 | 761 |
| BTM\_routes | 0 | 52 | 7 | 9 |
| TR\_freq\_tue | 6 | 1093 | 81 | 99 |
| TR\_freq\_off | 1 | 58 | 5 | 5 |
| TR\_amp | 0.3 | 0.8 | 0.6 | 0.1 |
| TR\_ttcentr | 19.4 | 83.7 | 70.7 | 10.4 |
| TR\_trcentr | 2.6 | 4.1 | 3.4 | 0.2 |
| P\_DE\_bas | 12.1 | 291.8 | 152.9 | 56.7 |
| P\_DE\_reg | 14.3 | 282.1 | 127.5 | 53.9 |
| P\_DE\_met | 0.9 | 256.1 | 90.7 | 58.1 |
| P\_DE\_job | 14 | 69719 | 2607 | 5512 |
| P\_DE\_res | 47 | 48575 | 5581 | 5296 |
| P\_DI\_shan | 0.0 | 8.9 | 5.3 | 2.0 |
| P\_DI\_IJI | 40.6 | 92.3 | 71.1 | 9.6 |
| P\_DG\_IC | 38.0 | 1184.0 | 382.8 | 216.7 |
| P\_DG\_perm | 31 | 805 | 299 | 157 |
| TT\_tr\_near | 2 | 15 | 4 | 2 |
| TT\_tr\_higher | 0 | 41 | 10 | 7 |
| TT\_car\_higher | 0 | 47 | 16 | 8 |

**References**

Bertolini, L. (1999) ‘Spatial development patterns and public transport: The application of an analytical model in the Netherlands’, *Planning practice and research* 14 (2), 199 – 210.

Blainey, S.P. (2010) ‘Trip end models of local rail demand in England and Wales’, *Journal of Transport Geography* 18(1), pp. 153-165.

Brunsdon, C., Fotheringham, A.S., Charlton, M. (1996) ‘Geographically weighted regression: a method for exploring spatial nonstationarity’, *Geographical Analysis* 28, 281 – 298.

Brunsdon, C., Fotheringham, A.S., Charlton, M. (1999) ‘Some notes on parametric significance tests for geographically weighted regression’, *Journal of Regional Science* 39, 497 – 524.

Cardozo, O.D., Garcia-Palomares, J.C., Gutierrez, J. (2012) ‘Application of geographically weighted regression to the direct forecasting of transit ridership at station-level’, *Applied Geography* 34, 548-558.

Caset, F., Teixeira, F., Derudder, B., Boussauw, K., Witlox, F. (2019) ‘Planning for nodes, places, and people in Flanders and Brussels: An empirical railway station assessment tool for strategic decision-making', *Journal of Transport and Land Use* 12 (1), 811-837.

Cervero, R. (2002) ‘Built environments and mode choice: Toward a normative framework’, *Transport Research Part D*, 7, 265 – 284.

Cervero, R. (2006) ‘Alternative approaches to modeling the travel-demand impacts of smart growth’, *Journal of the American Planning Association* 72 (3), 285 – 295.

Cervero, R. (2009) ‘Public Transport and Sustainable Urbanism: Global Lessons’, in Curtis, C., Renne, J. L., Bertolini, L. (2009) (Eds.) *Transit Oriented Development: Making it Happen*. Routledge.

Chiang, W-C., Russell, R. A., Urban, T.L. (2011) ‘Forecasting ridership for a metropolitan transit authority’, *Transportation Research Part A* 45, 696-705.

Chow, L-F., Zhao, F., Liu, X., Li, M-T., Ubaka, I. (2006) ‘Transit Ridership Model Based on Geographically Weighted Regression’, *Transportation Research Record* 1972, 105-114.

Crane, R. (2000) ‘The impacts of urban form on travel: An interpretative review’, *Journal of Planning Literature* 15, 3 – 23.

Curtis, C., Scheurer, J. (2010) ‘Planning for sustainable accessibility: Developing tools to aid discussion and decision-making'*, Progress in Planning* 74, 53 – 106.

Curtis, C., Scheurer, J. (2016) *Planning for public transport accessibility: An international sourcebook*. Routledge.

Ewing, R., Cervero, R. (2001) ‘Travel and the built environment: A synthesis’, *Transportation Research Record* 1780, 87 – 114.

Ewing, R., Cervero, R. (2010) ‘Travel and the built environment: A meta-analysis', *Journal of the American Planning Association* 76 (3), 265 – 294.

Falconer, R., Babb, C., Olaru, D. (2016) ‘Cities as systems: Node and place conflict across a rail transit network’, In: Biermann, S., Olaru, D., Paul, V. (Eds.) *Planning boomtown and beyond*, 460 – 489.

Fotheringham, A.S., Charlton, M.E., Brunsdon, C. (1998) ‘Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis’, *Environment and Planning A* 30, 1905 – 1927.

Fotheringham, A.S., Brunsdon, C., Charlton, ME. (2002) *Geographically weighted regression: The analysis of spatially varying relationship*. New York, NY: Wiley.

Frei, C., Mahmassani, H. S. (2013) ‘Riding more frequently: Estimating disaggregate ridership elasticity for a large urban bus transit network’, *Transportation Research Record: Journal of the Transportation Research Board* 2350, 65 – 71.

Gutiérrez, J., Cardozo, O.D., García-Palomares, J.C. (2011) ‘Transit ridership forecasting at station level: an approach based on distance-decay weighted regression’, *Journal of Transport Geography* 19 ,1081 – 1092.

Higgins, C.D., Kanaroglou, P.S. (2016) ‘A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region’, *Journal of Transport Geography* 52, 61 – 72.

Jeffrey, D., Boulangé, C., Giles-Corti, B., Washington, S., Gunn, L. (2019) ‘Using walkability measures to identify train stations with the potential to become transit oriented developments located in walkable neighbourhoods’, *Journal of Transport Geography* 76, 221 – 231.

Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., Kim, H.-J. (2015) ‘Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul’, *Journal of Transport Geography* 48, 30 – 40.

Kim, H., Sultana, S., Weber, J. (2018) ‘A geographic assessment of the economic development impact of Korean high-speed rail stations’, *Transport Policy* 66, 127 – 137.

Kobayashi, T. and Lane, B. (2007) ‘Spatial heterogeneity and transit use’, 11th World Conference on Transportation Research, Berkeley, June 24th-28th 2007

Lane, C., DiCarlantonio, M. and Usvyat, L. (2006) ‘Sketch models to forecast commuter and light rail ridership: Update to TCRP report 16’, *Transportation Research Record* 1986, 198 - 210.

Li, X., Liu, Y., Gao, Z., Liu, D. (2016) ‘Linkage between passenger demand and surrounding land-use patterns at urban rail transit stations: A canonical correlation analysis method and case study in Chongqing’, *International Journal of Transportation Science and Technology* 5, 10-16.

Li, Z., Han, Z., Luo, X., Su, S., Weng, M. (2019) ‘Transit oriented development among metro station areas in Shanghai, China: Variations, typology, optimization and implications for land use planning’*, Land Use Policy* 82, 269 – 282.

Liu, C., Erdogan, S., Ma, T., Ducca, F.W. (2014) ‘How to increase ridership in Maryland? Direct ridership models (DRM) for policy guidance’. In: TRB 93rd Annual meeting compendium for papers, Washington D.C., 3414 – 3417.

Marshall, S. (2005) *Streets & Patterns*. London: Spon Press.

Marshall, N., Grady, B. (2006) ‘Sketch transit modeling based on 2000 census data’, *Transportation Research Record: Journal of the Transportation Research Board* 1986 (1), 182 – 189.

May, X., Ermans, T., Hooftman, N. (2019) ‘Company cars: Identifying the problems and challenges of a tax system’, *Brussels Studies*. Retrieved from (August 19th 2019): https://journals.openedition.org/brussels/2499.

McGarigal, K., Marks, B. J. (1995). *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*. (Technical Report PNW-GTR-351). Portland, OR: US. Department of Agriculture, Forest Service, Pacific Northwest Research Station.

McNally, M. G. (2000) *The four step model*. Institute for Transportation Studies, University of California, Irvine.

Mennis, J. (2006) ‘Mapping the results of geographically weighted regression’, *The Cartographic Journal* 43 (2), 171 – 179.

Monajem, S., Nosratian, F. E. (2015) ‘The evaluation of the spatial integration of station areas via the node place model: an application to subway station areas in Teheran’, *Transportation Research Part D: Transport and Environment* 40, 14 – 27.

National Center for Geocomputation (2009). Maynooth, Ireland: National University of Ireland; 2009. [last accessed June 20th, 2019]. gwr.maynoothuniversity.ie/gwr4-software/

Nigro, A., Bertolini, L., Moccia, F.D. (2019) ‘Land use and public transport integration in small cities and towns: Assessment methodology and application’, *Journal of Transport Geography* 74, 110 – 124.

NMBS Holding (2013) *Modal split van de klanten in de Belgische stations: Resultaten en duiding*. Internal report.

Olaru, D., Moncrieff, S., McCarney, G., Sun, Y., Reed, T., Pattison, C., Smith, B., Biermann, S. (2019) ‘Place vs. Node transit: Planning policies revisited’, *Sustainability* 11 (2), 1 – 14.

Páez, A. (2006) ‘Exploring contextual variations in land use and transport analysis using a probit model with geographical weights’, *Journal of Transport Geography* 14, 167 – 176.

Pafka, E., Dovey, K. (2018). ‘Walkable access’. In Dovey, K., Pafka, A. Ristic, M. (Eds.), *Mapping Urbanities: Morphology, Flows, Possibilities*, 41 - 61. London: Routledge.

Peek, G-J. (2006) *Locatiesynergie. Een participatieve start van de herontwikkeling van binnenstedelijke stationslocaties*. Eburon, Delft.

Preston, J.M. (1991) ‘Demand forecasting for new local rail stations and services’, *Journal of Transport Economics and Policy* 25 (2), 183-202.

Pucci, P., Vecchio, G. (2019) ‘Stations: Nodes and places of everyday life’. In: Pucci, P., Vecchio, G. (Eds.) *Enabling mobilities: Planning tools for people and their mobilities*. Springer International, 59 – 79.

Reusser, D. E., Loukopoulos, P., Stauffacher, M., Scholz, R. W. (2008) ‘Classifying railway stations for sustainable transitions – balancing node and place functions’, *Journal of Transport Geography* 16 (3), 191 – 202.

Ryan, S., Frank, L.F. (2009). ‘Pedestrian environments and transit ridership’, *Journal of Public Transportation* 12 (1), 39 – 57.

Singh, Y.J., Lukman, A., Flacke, J., Zuidgeest, M., van Maarseveen M. (2017) ‘Measuring TOD around transit nodes – Towards TOD policy’, *Transport Policy* 56, 96 – 111.

Stead, D. (2001) ‘Relationships between land use, socio-economic factors, and travel patterns in Britain’, *Environment and Planning B* 28 (4), 499 – 528.

Sun, G., Zacharias, J., Ma, B., Oreskovic, N. M. (2016) ‘How do metro stations integrate with walking environments? Results from walking access within three types of built environment in Beijing’, *Cities* 56, 91 – 98.

Sung, H., Oh, J. (2011) ‘Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea’, *Cities* 28, 70 – 82.

Taylor, B. D., Fink, C. N. Y. (2003) ‘The factors influencing transit ridership: A review and analysis of the ridership literature’. UC Berkeley. Retrieved from (August 19th 2019): https://escholarship.org/uc/item/3xk9j8m2.

Van Acker, V. (2010) *Spatial and Social Variations in Travel Behaviour: Incorporating Lifestyles and Attitudes into Travel Behaviour-Land Use Interaction Research*. Doctoral dissertation, Ghent University.

Young MA & Blainey SP (2019) ‘An Automated Online Tool to Forecast Demand for New Railway Stations and Analyse Potential Abstraction Effects’, Transport Practitioners’ Meeting, 10-11 July 2019, Oxford.

Zemp, S., Stuaffacher, M., Lang, D. J., Scholz, R. W. (2011) ‘Classifying railway stations for strategic transport and land use planning: Context matters’, *Journal of Transport Geography* 19 (4), 670 – 679.

Zhang, Y., Marshall, S., Manley, E. (2019) ‘Network criticality and the node-place-design model: Classifying metro station areas in Greater London’, *Journal of Transport Geography*. https://doi.org/10.1016/j.jtrangeo.2019.102485.

Zhao, J., Deng, W., Song, Y., Zhu, Y. (2013) ‘What influences Metro station ridership in China? Insights from Nanjing’, *Cities* 35, 114-124.

1. According to a study by the NMBS Holding (2013) in which the modal split of Belgian train stations in terms of feeder mode travel was examined, walking seems to represent the most important mode for more than half of all (N=83) examined stations. A 1200 m distance therefore seems justifiable. [↑](#footnote-ref-1)
2. The Shannon diversity index increases as the number of different land use types increases, and as their proportions in terms of space occupied become more balanced. [↑](#footnote-ref-2)
3. The Interspersion and Juxtaposition Index (IJI) (see McGarigal and Marks 1995) measures the interspersion of different land use types in an area. Well-interspersed station areas have land use ‘patches’ or types that are equally adjacent to each other. [↑](#footnote-ref-3)
4. Car availability may nonetheless be correlated to these socioeconomic data, given that it has frequently been demonstrated (see Van Acker and Witlox 2010 for an overview) that variables such as age, employment status, educational level and other socioeconomic and demographic characteristics impact car ownership and car use. [↑](#footnote-ref-4)
5. The ‘statistical sector’ is the most detailed territorial level used by Statistics Belgium (Statbel). [↑](#footnote-ref-5)
6. These clusters were validated by running one-way analyses of variance (ANOVA) for the different key characteristics (as revealed by the factor loadings and cluster centers) that shaped the station groupings. [↑](#footnote-ref-6)
7. The classification method for the legend items is based on standard deviations as the data is normally distributed (see Mennis 2006 for more cartographic support with respect to GWR analyses). [↑](#footnote-ref-7)
8. Arguably, these assumptions could be verified more effectively when the station demand model would be integrated with a station choice model (see Young and Blainey 2019). Unfortunately, to the best of our knowledge there is no data available in Flanders at the scale required to calibrate such a model. [↑](#footnote-ref-8)