Modelling the effect of spatial determinants on freight (trip) attraction
A spatially autoregressive geographically weighted regression approach

Reda, Abel Kebede; Tavasszy, Lori; Gebresenbet, Girma; Ljungberg, David

DOI
10.1016/j.retrec.2023.101296

Publication date
2023

Document Version
Final published version

Published in
Research in Transportation Economics

Research in Transportation Economics 99 (2023) 101296

Contents lists available at ScienceDirect

Research in Transportation Economics

journal homepage: www.elsevier.com/locate/retrec

Research paper

Modelling the effect of spatial determinants on freight (trip) attraction: A spatially autoregressive geographically weighted regression approach

Abel Kebede Reda [a,*], Lori Tavasszy [b], Girma Gebresenbet [a], David Ljungberg [a]

[a] Department of Energy and Technology, Swedish University of Agricultural Sciences, P.O. Box 7032, 750 07, Uppsala, Sweden
[b] Faculty of Technology, Policy and Management (TPM), Delft University of Technology, Jaffalaan 5, 2628BX, Delft, the Netherlands

ARTICLE INFO

Keywords:
Freight generation
Trip attraction
Spatial autocorrelation
Spatial heterogeneity
Geographically weighted regression
Multiscale GWR

ABSTRACT

This paper investigates the effect of spatial and locational characteristics of establishments on freight (trip) attraction (FA/FTA) models. The authors estimated econometric models of FA and FTA as a function of the establishment attributes as well as the spatial and locational determinant variables, using establishment-level data collected from Addis Ababa City, Ethiopia. The interconnected issues of spatial dependency and spatial heterogeneity, together with nonlinear specifications, were incorporated with the application of spatial techniques, including spatial error models (SEM), spatial autoregressive model (SAR), geographically weighted regression (GWR), multiscale-GWR (MGWR), and the combination GWR-SAR/MGWR-SAR. Regarding the explanatory variables, the empirical results revealed that firms in the manufacturing, wholesale and retail sectors located on the wider streets tend to receive more FA and FTA. The closeness to the primary road network and the city entry gate influences the FTA of manufacturing and construction firms. Moreover, retail establishments near the major market tend to receive more tonnage. The models also confirm that FA and FTA are the results of two different processes. Overall, the use of spatial regression techniques improves the accuracy of both FA and FTA models. MGWR-SAR exhibits superior performance by jointly addressing spatial dependency and heterogeneity. The MGWR-SAR model also uncovers the local variability of the variables representing the spatial and locational effects on freight attraction. The methodological analysis and empirical findings of the study could provide useful insights to support urban freight modelling, planning, and decision-making.

1. Introduction

Accurately estimating and predicting freight demand is critical to freight transportation planning. Freight demand models comprise two key variants; the amount of cargo in tonnage, i.e., freight generation (FG), which is attracted (FA) or produced, and the number of truck trips, i.e., freight trip generation (FTG) that is attracted (FTA) or produced (FTP) (Holguín-Veras et al., 2011). These models are subdivided into three spatial levels (Gonzalez-Feliu & Sánchez-Díaz, 2019): macroscopic (regional or city-level), mesoscopic (neighborhood or street level), and microscopic level (establishment level). Among these, establishment-level models are consistent with logistics processes and thus possess a greater potential to explain the variability within the data (Ortúzar & Willumsen, 2011). However, a lack of establishment-level data on firm characteristics and actual flows has hindered the development of such models, especially in developing countries (Sahu & Pani, 2019).

Freight demand modelling studies have utilised various parameters and methods to develop FG/FTG models. These include ordinary least squares (OLS) regression, generalised linear regression (GLM), ordered logit, negative binomial, multiple classification analysis (MCA), and artificial neural networks (ANN). Models typically build using explanatory variables such as employment, industry category, gross floor area, commodity type, and years in business (Balla et al., 2022). Several studies incorporate spatial variables representing the geographic location and the spatial effects when modelling FG and FTG. Locational effects in the analysis of the spatial data and modelling of the spatial processes manifest themselves in three ways: spatial dependence (spatial autocorrelation), spatial heterogeneity (spatial non-stationarity), and the "modifiable area unit problem" (MAUP) (Miller, 1999). The spatial autocorrelation effect arises due to the systematic spatial variation in the variables with high values near other high values and vice versa (Páez & Scott, 2004). Spatial non-stationarity refers to the parameters or their relationship behaving differently over space due to locational

* Corresponding author.
E-mail address: abel.kebede.reda@slu.se (A.K. Reda).

https://doi.org/10.1016/j.retrec.2023.101296
Received 17 September 2022; Received in revised form 7 March 2023; Accepted 4 May 2023
Available online 16 May 2023

uniqueness within the spatial area (Anselin et al., 1993). According to recent developments in spatial econometrics, the effects of nonlinearity, spatial autocorrelation, and spatial heterogeneity are inseparable when calibrating spatial models. The misspecification of one can cause adverse effects such as inducing, magnifying or eliminating the others (Geniaux & Martinetti, 2018). The spatial regression technique that incorporates the effects of nonlinearity and spatial autocorrelation or spatial autoregressive model (SAR) has been used to model FG at the establishment level, city/regional level, or national level (Novak et al., 2011) and to model the FTG of different industry sectors (Middela & Ramadurai, 2021; Sánchez-Díaz et al., 2016). However, no research has been done on modelling FG and FTG accounting for the effects of spatial heterogeneity along with nonlinearity and spatial autocorrelation. Our study aims to fill this void.

Therefore, this research investigated the effects of spatial and locational characteristics of establishments on FA and FTA using various spatial econometric techniques. The application of spatial regression methods aims to handle different spatial issues and enhance the predictive power of FG and FTG models. Methodologically we propose spatial regression techniques, i.e. geographically weighted regression (GWR) or multi-scale GWR (MGWR), which capture spatial non-stationarity effects, and the recently developed GWR-SAR/MGWR-SAR, capable of addressing the joint effects of spatial non-stationarity and spatial autocorrelation. To the best of the authors' knowledge, these models have not yet been empirically tested for freight transport. The research encompasses operationalising the various models, their implementation with appropriate data, and their interpretation. Several spatial regression techniques are utilised to model the relationship between the freight patterns of establishments and the built environment or the urban geography. Their accuracy is compared and discussed.

The structure of the paper is built up as follows: Chapter 2 provides a review of the relevant literature and positions the research in more detail. Chapter 3 introduces the methods and data. Chapter 4 presents and discusses the results. Chapter 5 summarises the main findings and concludes the paper.

2. Literature review

Freight demand generation models are usually separated into FG and FTG models. FG indicates the output of the production process at the establishments, and FTG results from logistical decisions. The correlation of establishment size variables (e.g., employment) with FG is relatively stronger than FTG, where FG is the direct output of the economic process of commodity production and attraction. FTG, on the other hand, is influenced by shipment size and other logistic considerations (Holguín-Veras et al., 2011). The output from the two model types also serves different purposes in freight planning. FG models are applicable for planning consolidation centres (Sánchez-Díaz, 2017), assessing regional development, port improvements, and land use ordinances (Sahu & Pani, 2019). FTG models are applied in freight planning for the design of freight infrastructure and policy (Sánchez-Díaz, 2017), logistics facilities of large manufacturers (Institute-Of-Transportation-Engineers, 2008), parking places (Jaller et al., 2014; Alho & Silva, 2014), and also feed traffic simulation models (Alho, De Abreu, De Sousa, & Blanco, 2018).

The factors of predominant importance in modelling FG/FTG are (i) the choice of the aggregation level, (ii) the type of aggregation procedures or techniques, and (iii) the selected classification system for the economic sector (Gonzalez-Feliu & Sánchez-Díaz, 2019; Holguín-Veras et al., 2014; Pani et al., 2018). The business size of establishments is a significant predictor for FTG, but most empirical models exhibit low explanatory power (Holguín-Veras et al., 2013). A small set of variables used in the freight models can be one of the reasons for their relatively low prediction power (Sánchez-Díaz et al., 2016). On the other hand, adding more explanatory variables in those models can complicate the analysis due to the correlation between variables. In the freight literature, numerous attempts have been made to expand the modelling methodology of FG/FTG models beyond the traditional OLS regression model (Balla et al., 2022), but spatial effects are rarely incorporated. Therefore, addressing the spatial issues inherent in freight patterns is essential to advance these models (Novak et al., 2011; Pani et al., 2018).

The analysis of freight demand is sensitive to the typical statistical characteristics of the spatial data (Sahu et al., 2020). In spatial analysis, the locational interdependencies are mainly expressed in three different ways: spatial dependence (spatial autocorrelation), spatial heterogeneity (spatial non-stationarity), and the MAUP effect (Miller, 1999). The spatial autocorrelation effect systematically clusters the variables with high values near other high values and vice versa. This correlation of variables may violate the linearity assumptions in the model (Novak et al., 2011) and result in biased parameter estimates (Sánchez-Díaz et al., 2016). Spatial non-stationarity refers to the parameters or their relationship behaving differently over space due to locational uniqueness within the spatial system. This might result in the inability of the spatial model to describe the overall process (Anselin et al., 1993), manifesting in biased parameter estimates or predictions (Griffith et al., 1999). The MAUP effect arises due to the mismatch between the data availability with the pre-defined areal units (such as administrative, census tracts, and districts) and the choice of zoning system for spatial analysis, which leads to inconsistent estimates and erroneous statistical inferences (Openshaw et al., 1979). Regarding freight demand modelling, only a handful of studies explicitly consider these spatial indicators to model the FG and FTG patterns.

Garrido and Mahmassani (2000) studied the effects of economic activity and spatial location on the freight transport demand of establishments based on the operational decisions of a single carrier. Kawamura and Miodonski (2012) estimated the FA of retail establishments using the explanatory variables that explain the socioeconomic and land-use characteristics of the geographic locations. Novak et al. (2011) used the US national data to model freight generation using spatial regression techniques. Sánchez-Díaz et al. (2016) assessed the impacts of spatial effects and locational characteristics on the FTG using the data set from New York City. Pani et al. (2018) analysed the impact of the interaction between the establishment characteristics and locational variables with freight generation. Pani et al. (2019) studied the effect of MAUP or zonal characteristics on the disaggregate freight (trip) generation patterns and demonstrated the sensitivity of the freight patterns to the choice of spatial scale. Sahu and Pani (2019) analysed the establishment-level FG models using OLS and MCA for businesses in Kerala, India, and revealed the geographic disparities in the models using ANCOVA to test the spatial transferability of these models (Pani et al., 2021). assessed the spatial transferability of FP and FTP models within regions and across states in India and indicated FP models are more transferable than FTP. Middela and Ramadurai (2021) modelled FTG with spatial autoregressive zero-inflated negative binomial techniques and showed the presence of sptial autocorrelation effect in almost all the models. The two main implications that can be extracted from the previous studies are: (1) the spatial effects and locational determinants do influence both the FG and FTG models, and incorporating those factors improves the predictive accuracy of the models, and (2) although the analysis considered the impact of spatial variables in the entire area but their effects are mostly different locally.

Spatial non-stationarity (heterogeneity) can be handled in two ways. The first is a compromise between the global and local models using switching regression, multi-level, or hierarchical models, and a local form of the spatial model using the GWR method (Páez & Scott, 2004). The inclusion of the spatial heterogeneity effect allows the relationship between the model parameters to vary over space and essentially captures the local impacts. The GWR technique has an advantage over the other methods by explicitly incorporating local relationships into the regression. As it assumes that scales of the spatial relationships are constant over space, it can be inadequate to analyse these relationships at different scales. Thus, different processes require varying spatial scales, and those assumptions are not usually valid. Multi-scale GWR (MGWR) relaxes the assumptions of GWR, which allows the relationship between the dependent variable and explanatory variables to vary locally. The flexibility of analysing the relationships at different spatial scales can minimise over-fitting, reduce bias in the parameter estimates, and mitigate collinearity (Fotheringham et al., 2003; Oshan et al., 2019).

Models that explicitly consider spatial autocorrelation and spatial heterogeneity are gaining the attention of spatial econometric studies. In the calibration of models, spatial autocorrelation, spatial heterogeneity, and nonlinearity are inseparable, and the misspecification of one can bias the others (Basile et al., 2014; Geniaux & Martinetti, 2018). Spatial regression methods such as spatial lag or autoregressive models assume the stationarity of spatial autocorrelation parameters throughout the analysis area. However, this assumption of the constant influence of the autoregressive term over space is invalid and biases the parameter estimates. The GWR-SAR or MGWR-SAR framework considers both spatial effects (spatial autocorrelation and spatial heterogeneity). This approach allows the spatial autoregressive parameter (explaining the spatial dependencies) to vary locally or with other parameters over different spatial scales. The spatial regression techniques used for modelling the local relationships, such as GWR/MGWR and GWR/MGWR-SAR, have not been previously applied for spatial analysis of FG/FTG patterns. The research contributes to modelling business establishments' freight (trip) attraction considering the above dependencies with local differentiation, using various methods. We operationalise, implement and interpret the findings. Data were obtained to create a new model for Addis Ababa, Ethiopia. We introduce our approach in the following sections.

3. Methodology

3.1. Overall approach

Below we describe the general approach taken in the study, organised into several steps (as shown in Fig. 1). The study starts with filtering and sorting the freight data from a 2019 survey of business establishments in Addis Ababa City, Ethiopia. The next step checks whether spatial dependency or autocorrelation is present in the freight patterns. Followed by the identification of different explanatory variables, these variables can explain the economic, spatial, and locational characteristics of business establishments responsible for freight activity. Next, modelling the relationship between the freight patterns and the relevant explanatory variables using the OLS regression to assess the correct model specification between linear and nonlinear forms and simultaneously checking the correlation and multicollinearity of these variables. The model specification, whether linear or nonlinear, was evaluated using the Ramsey regression equation specification error test (RESET). The correlation and multicollinearity of variables were analysed using Pearson correlation and the Variance Inflation Factor (VIF). After the model specifications and selection of essential explanatory variables, the next step is to model the OLS, SEM, SAR, GWR/MGWR, and GWR-SAR/MGWR-SAR. From the last two model types, Monte Carlo (MC) simulation test determines the choice of the model variant, which examines whether the explanatory variables are spatially variable or not. With the results of the MC test, if all the variables have significant spatial variability, then the variants of the local models will be the GWR and GWR-SAR. Otherwise, if any of the variables do not have spatial variability, then the local models of MGWR and MGWR-SAR will be used. Finally, the statistical indexes of adjusted R-square, AIC, and RMSE are used to evaluate and compare the model results.

3.2. Study area and description of the data

The primary data used in this research is based on establishment-level FG/FTG data collected from 451 businesses in Addis Ababa city, Ethiopia. The establishment–based freight survey (EBFS) was conducted in 2019. The survey targeted establishments in the freight-intensive industry sectors, including manufacturing and raw material production, construction, wholesale and retail trades. The data comprises information about the type of commodities, employment, land use, and industry sector. Specifically, the attributes related to the business size of establishments, i.e. the number of employees (E) and the freight measures, including the tonnage delivered to and shipped from the establishment with the number of deliveries and shipments. Many establishments in the retail trade and wholesale trade sectors fail to keep track of their outbound freight in tons/week and trips/week, and thus, the authors decided to focus the analysis on the FA/FTA part. The municipal wards of the city, locations of sampled establishments and other important attributes are presented in Fig. 2.

3.2.1. Explanatory variables

The independent variables for modelling FA and FTA are presented in Table 1. Based on previous studies and data availability, the variables are grouped into economic, land use, and network characteristics.

The explanatory value of these variables is used to model both FA

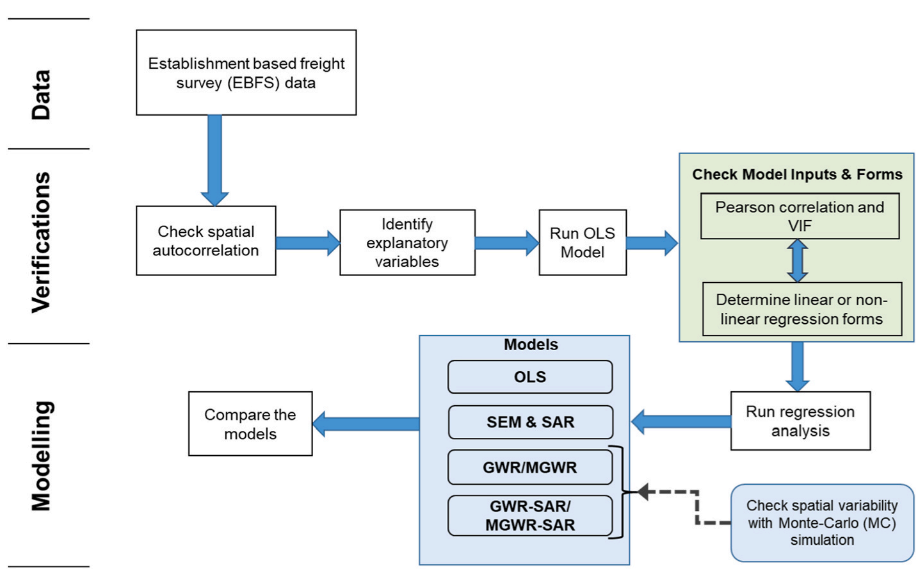

Fig. 1. Methodological framework of the study.

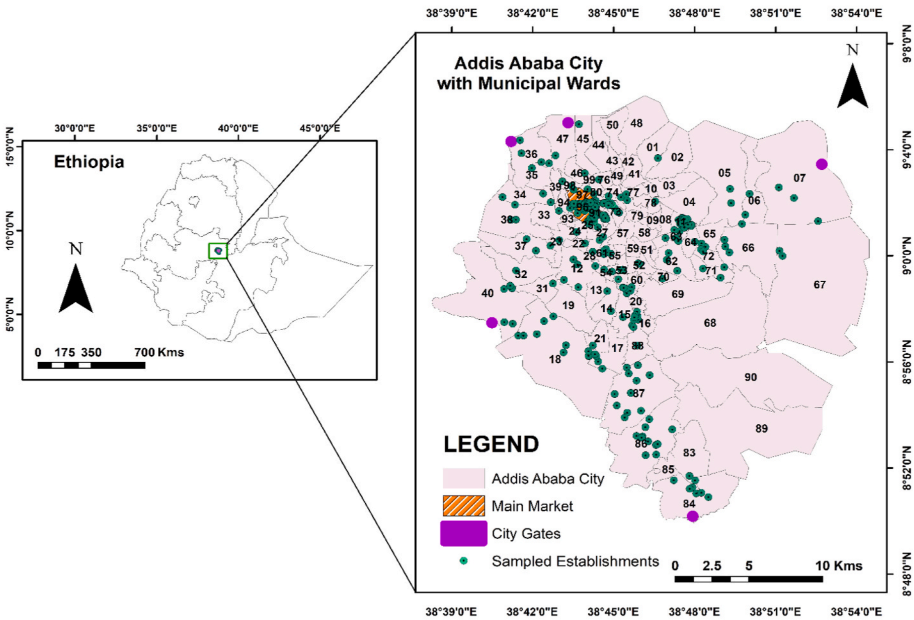

Fig. 2. Study area.

Table 1
Description of the variable.

Category	Variable	Description	Abbrev.
EconomicCharacteristics	SIC	A binary variable representing the industry sectors: Manufacturing, Wholesale, Retail, Construction	
	Employment	Number of employees working in the establishment	E
Land-use variables	Land-market value	The unit price of land (in thousands of US$ per sq. meters)	LV
Network Characteristics	Steet width	Width of the street (in meters) in front of the establishments	SW
	Distance to the primary network	Euclidian distance (in meters) to the closest street that is part of the primary road network	DPN
	Minimum distance to the city gate	Network distance (in Kilometers) to the closest city gate (among the five gates to the city)	DCG
	Distance to the main market	Network distance (in Kilometers) to the main market located at the centre of the city	DMM

and FTA. The data can be obtained directly from different public data sources or estimated based on the geo-locations of establishments.

- Economic attributes include the industry sector and employment of the establishment. These variables measure the size of the establishment and distinguish the type of business activity taking place in the establishment. Based on the economic activity at the establishments, the standard industrial classification (SIC) was used to create relevant categories of business establishments with similar freight patterns.
- Land use characteristics are represented with the variable land value at the establishment's location and extracted based on the land lease information obtained from the city's land development and management office. The land value is a proxy for explaining the establishment's revenues and associated inventory costs that potentially affect freight patterns.
- Network characteristics connect the freight patterns of establishments with the locational attributes, mainly related to the transport infrastructure. The network characteristics include distance to the primary network, distance to a major market, distance to the nearest city gate, and width of a street in front of the establishments. The distance to the primary network measures the establishment's location relative to the main streets. Distance to a major market measures the closeness of the establishment with a big market located at the centre of the city, which is the city's main commercial hub. Distance to the nearest city gate represents the establishment's location relative to the closest one among the five city gates located in four different directions around the city. The street width in front of the establishment accounts for the locational vitality and is a proxy variable to enhance the locational representation on top of other network characteristics attributes.

3.2.2. Descriptive statistics

The descriptive statistics of the dataset are summarised in Table 2. The breakdown by industry sector using the standard industrial classification (SIC) shows that most of the sampled establishments are in the retail trade (53%) and manufacturing (34%). The FA value varies between 0.3 and 1150 tons per week. The FTA value ranges between 1 and 56 weekly deliveries. The retail establishments attract the smallest average number of trips (2 weekly deliveries) and shipment weight (4.45 tons weekly), while construction establishments attract the biggest shipment weight (228 tons weekly) and wholesale establishments receive the largest number of deliveries (12 weekly deliveries). The number of employees ranges from 1 to 1260, where retail establishments have the lowest average employment (3 employees), and construction establishments have the highest average employment (166 employees). The land values indicate retail establishments tend to be mostly located in the most expensive area ($10 thousand per square meter) with narrow street widths (9.3 m), and the construction establishments tend to be located in the cheapest area (2.7 thousand $ per square meter) with wider street widths (15.8 m). In relation to the locational variables, the wholesale establishments are mostly located nearest to the primary road network, the main market, and the city gates (19.3 m, 1.7 km, and 9.5

Table 2
Descriptive statistics.

Variable	Unit	SIC 20-39: Manufacturing					SIC 15-17: Construction				
		Obs.[a]	Mean	Std. Dev.[a]	Min	Max	Obs.	Mean	Std. Dev.	Min	Max
FA	tons/week	151	43.9	97	0.8	500	18	228	356.3	5	1150
FTA	deliveries/week	151	5	5	1	27	18	10	9	1	43
Employment	employees	151	150	204	5	1260	18	166	118	31	402
Land value	thousand US$/Sq.m	151	2.9	3.6	0.6	17.2	18	2.7	1.9	1.1	8.7
Street width	meters	151	12.4	5.8	4.5	25	18	15.8	6.3	5.2	30
DPN	meters	151	76	71	3	336	18	49	80.1	5.4	274
DCG	kilometers	151	9.6	3.54	0.3	15.5	18	10	2.6	5.6	13.8
DMM	kilometers	151	8.7	6.0	0.0	25	18	8.2	6.5	0.3	21.7
Variable	**Unit**	**SIC 50–51: Wholesale Trade**					**SIC 52–59: Retail Trade**				
FA	tons/week	34	60.5	71	2	280	243	4.45	5.8	0.3	35
FTA	deliveries/week	34	12	15	1	56	243	2	1	1	12
Employment	employees	34	12	22	2	126	243	3	2	1	15
Land value	thousand US$/Sq.m	34	9.7	5.6	2.3	17.2	243	10.0	6.6	1.3	17.2
Street width	meters	34	15.3	5	4.5	20	243	9.3	5.5	3.8	27.5
DPN	meters	34	19.3	30.2	0.0	143.8	243	25.1	35	0.0	210.3
DCG	kilometers	34	9.5	1.2	7.9	12.8	243	9.4	2.4	0.8	16.5
DMM	kilometers	34	1.7	2.3	0.14	8.4	243	5.74	4.9	0.0	19

[a] Obs. – Observations; Std. Dev. - Standard Deviation.

km, respectively), while manufacturing establishments are located further away from the primary road network, the main market and the city gates (with an average of 76 m, 8.7 km and 9.6 km respectively). These locational results are plausible that retail establishments are placed at prime locations to attract consumers and manufacturing establishments located away from the main activity centres due to larger space requirements to carry out their activities.

3.3. Econometric models

3.3.1. FA/FTA models only with employment

The FTG modelling is described here using the case of FTA and will, in abstract terms, apply equally to FTP. The main predictor variable is the total employment offered by establishments in the respective industry sectors. For estimation, the common and extensively used OLS regression can be utilised. The general form of the models is given as:

$$y_i = \beta_0 + \beta_1 (E_i) + \epsilon_i \tag{1}$$

Where:

y_i is the FA in tons/week or FTA in number of delivery/week,

E_i is the total employment as business size variable of the ith establishment in the industry sector

β_1 is the regression coefficient of E_i; β_0 is a constant or intercept term; ϵ_i is a random error.

The FA/FTA model formulation above can result in three cases: (i) model only with the coefficient term, (ii) model only with the constant term, and (iii) model with both the constant and coefficient terms (Gonzalez-Feliu & Sánchez-Díaz, 2019; Holguín-Veras et al., 2011; Pani et al., 2018).

3.3.2. Spatial indicators

The spatial location of establishments has vital implications on their FA/FTA and possesses the potential to capture the interaction between establishments and the urban environment. The freight needs are in correlation to the underlying spatial processes. Spatial indicators quantify the dependency of establishments on the surrounding environment because of the effects related to their spatial locations. The most widely used technique to express the underlying spatial relation is spatial autocorrelation (Anselin & Bera, 1998), and Moran's I measure its magnitude. The specification of Moran's I:

$$I = \left(\frac{n}{\sum_i\sum_j w_{ij}}\right)\left(\frac{\sum_i\sum_j w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_i (x_i-\bar{x})^2}\right) \tag{2}$$

where n is the number of observations, x_i is the value at location i, $\bar{x}$ is the mean value of x, w_{ij} is the spatial weight defined with the inverse of the distance between establishments located with i and j.

The expected value of Moran's I when no spatial autocorrelation is:

$$E(I) = \frac{-1}{n-1} \tag{3}$$

The global tendency of spatial autocorrelation is given by the difference between the value of Moran's I and the expected value. When Moran's I value is greater than E(I) indicates a global tendency toward clustering, and when the value is less than E(I) indicates a tendency towards dispersion or uniformity (Moran, 1950). Moran's I is sensitive only to large spatial autocorrelation, but autocorrelation may exist only in some zones of a given spatial analysis area. In these cases, the analysis area is divided using Thiessen polygons, and the interaction between the neighbouring polygons can be calculated with the Local Indicators of the Spatial Association (LISA). The specific location of the significant interactions with LISA can be mapped using statistical tests (Anselin, 1988).

The presence of spatial autocorrelation indicates the need to incorporate spatial interactions when modelling the freight (trip) generation. The systematic handling of the spatial autocorrelation between the establishments in freight (trips) generation and their relation with the geographic location enhances the resulting FA (Novak et al., 2011) and FTA models (Sánchez-Díaz et al., 2016).

This study examines the relation of FA/FTA with the economic characteristics and environmental effects of establishments using five different models at two levels. The global models include the traditional ordinary least squares (OLS) regression, spatial lag model (SLM), and spatial error model (SEM). The local models are geographically weighted regression (GWR), and GWR/Multiscale GWR–Spatial autoregessive model (SAR).

3.3.3. Global models

The starting point to model FA/FTA in this study is the traditional OLS regression. The general form of the models is given (Ward & Gleditsch, 2018):

$$y_i = \beta_0+\beta_i\, x_i+\varepsilon_i \tag{4}$$

Where y_i is the amount of freight both in FA in tons/week or FTA in deliveries/week, β_0 is the intercept, x_i is the vector of explanatory variables representing the characteristics of the ith business establishment in the respective industry sector, β_i is the vector of regression coefficients, and ε_i is a random error term.

The regression coefficients in the OLS regression are optimised by minimising the sum of squared prediction errors. The main assumptions in OLS regression are independence between observations and uncorrelated error terms (Anselin & Arribas-Bel, 2013). These models relate the freight amount with the establishments' economic characteristics and environment. The OLS assumes that the observations (establishments for this case) are independent and do not consider the spatial interaction between establishments. However, spatial interactions between establishments play an essential role in modelling FA/FTA, as demonstrated by the results from previous freight studies (Novak et al., 2011; Sánchez-Díaz et al., 2016). The variants of OLS that take into account the spatial dependence between observations are the spatial lag model (SLM) and the spatial error model (SEM) (Anselin, 2003; Ward & Gleditsch, 2018). The SLM and SEM models exploit the spatial correlation between the establishments to model FA/FTA, which considers spatial weights and lags in regression models.

The spatial dependence can be introduced into the model specification in two ways: spatial lag dependence or spatial error dependence (Anselin, 1988). The SLM or spatial autoregressive (SAR) model assumes spatial dependency between the dependent and explanatory variables and includes an additional spatially-lagged dependent variable to account for spatial dependence (Anselin, 2003). With the mixed regressive form, the spatial autoregressive (SAR) model is denoted by:

$$y_i = \beta_0 + \beta x_i + \rho W_i\, y_i + \varepsilon_i \tag{5}$$

Where ρ is the spatial lag (spatial autoregressive) parameter, and W_i is the spatial weights matrix. The spatial lag for variable $\mathbf{y}$ at i is expressed as $\boldsymbol{W_i}\,\boldsymbol{y_i}$. The spatial weight matrix W is built with the distances between every pair of observations and then row-standardising W to interpret $\boldsymbol{W}$. $\boldsymbol{y}$ as the average FA/FTA of the neighbours.

The spatial dependence in the SEM assumes the error term of the OLS model decomposes into two terms ($\lambda \mathbf{W_i}\, \xi_i$ and ε_i). The SEM model is denoted as (Ward & Gleditsch, 2018):

$$\boldsymbol{y_i} = \beta_0 + \beta \boldsymbol{x_i} + \lambda \boldsymbol{W_i} \xi_i + \varepsilon_i \tag{6}$$

Where ξ_i is the spatial component of the error at i, λ indicates the level of correlation between these components, and ε_i is a spatially uncorrelated term.

The correlation and multicollinearity between the candidate explanatory variables are evaluated using the variance inflation factor (VIF) and Pearson correlation coefficients. A VIF value of 5 (moderate multicollinearity) and a Pearson's correlation coefficient value of 0.75 were considered the threshold for strong correlation. The variable(s) with higher multicollinearity or correlation than the threshold values were eliminated from the subsequent analysis. The analysis of the global models (OLS, SAR, and SEM) used the same variables to compare the model estimates consistently. These models are analysed and implemented in GeoDa 1.8 (Anselin et al., 2010).

3.3.4. Local models

Global models such as OLS, SAR, and SEM assume spatial stationarity when modelling the relationship between the dependent and explanatory variables, where these relationships do not vary over space. The GWR model extends the general regression model and relaxes the spatial stationarity assumption to allow the variables to vary over space (Brunsdon et al., 2002). The mathematical notation of the GWR model is extended from **Eqn. (1)** and adopted as (Fotheringham et al., 2003):

$$\boldsymbol{y_i} = \sum_{j=0}^{m} \boldsymbol{\beta_j}(\boldsymbol{u_i}, \boldsymbol{v_i})\, \boldsymbol{x_{ij}} + \boldsymbol{\varepsilon_i} \tag{7}$$

Where y_i denotes the amount of freight (FA/FTA) at point i, and $\beta_j\,(u_i, v_i)$ is a varying jth coefficient in the estimation of continuous explanatory variables, x_{ij}, at any point i within the given spatial analysis area, $\boldsymbol{\varepsilon_i}$ is a random error term.

Multi-scale GWR (MGWR) is an extension of GWR and allows the conditional relationships between the dependent variable and different predictor variables to vary locally and/or not at all. Mainly the extension allows each variable to have a distinct bandwidth, where the data-borrowing range (bandwidth) can vary across parameter surfaces. The MGWR model takes the form (Fotheringham et al., 2017):

$$\mathbf{y}_i = \sum_{j=0}^{m} \boldsymbol{\beta_{bwj}}(\boldsymbol{u_i}, \boldsymbol{v_i})\, \boldsymbol{x_{ij}} + \boldsymbol{\varepsilon_i} \tag{8}$$

where (u_i, v_i) denotes the x-y coordinates of the i^{th} point, y_i is the amount of freight (FA/FTA), bwj in β_{bwj} indicates the bandwidth used for calibration of the jth conditional relationship.

The spatial variability of all parameters, both the selected explanatory variables and the spatial autocorrelation term, should be tested with the Monte Carlo (MC) test for spatial variability. The MC test indicates whether all the parameters are local and if there are any global variables. Based on the results of the MC tests, both the spatial heterogeneity or the joint model of spatial heterogeneity and spatial autocorrelation takes two forms. When all the parameters are local or have spatial variability, then the type of model will be GWR and GWR-SAR. Otherwise, the model type becomes MGWR and MGWR-SAR. The MGWR-SAR model specified by (Geniaux & Martinetti, 2018) takes the possibility of having both global and local variables, and the general form of this model is:

$$\boldsymbol{y} = \boldsymbol{\rho}(\boldsymbol{u_i}, \boldsymbol{v_i})\boldsymbol{W}\boldsymbol{y} + \boldsymbol{\beta_c}\boldsymbol{X_c} + \boldsymbol{\beta_v}(\boldsymbol{u_i}, \boldsymbol{v_i})\, \boldsymbol{X_v} + \boldsymbol{\varepsilon} \tag{9}$$

where (u_i, v_i) denotes the x-y coordinates of the i^{th} point, y_i is the amount of freight (FA/FTA), X_c are K_c independent variables with constant coefficients (β_c), and X_v are K_v independent variables with spatially varying coefficients (β_v), Wy is the spatial lag variable with the spatially varying coefficient $\rho(u_i, v_i)$.

This study calibrates the local models based on GWR or MGWR with a gaussian kernel function and fixed bandwidth. Local models executed in MGWR 2.2 (Oshan et al., 2019). The optimal bandwidth for the kernels is selected based on minimising the corrected Akaike Information Criterion (AICc). The GWR-SAR/MGWR-SAR models include a spatially-lagged dependent variable that can introduce endogeneity by correlating with the error term ε. The spatial two-stage least square (S2SLS) technique is used to get rid of the endogenous part (Anselin, 2003). Moreover, another problem when estimating the GWR is the artificial increase in the t-value results because the local GWR estimates often overlap. The significant level of the estimates is adjusted to solve this problem and can take the form (Da Silva & Fotheringham, 2016):

$$\alpha = \frac{\xi_m}{p_e/p} \tag{10}$$

where ξ_m is the usual α, p_e is the effective number of parameters, and p is the number of parameters

3.3.5. Evaluation of models

The measures used to compare the performance of the global and local models are the adjusted R-square, AIC, and root mean square error (RMSE). These are defined as:

$$\text{AIC} = 2k - 2\ln(\widehat{L})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y - \widehat{y})^2}{n}} \tag{11}$$

Where, $\widehat{y}$ and y are the estimated and observed values of the dependent variable, respectively, n is the number of observations, the number of estimated parameters k, and $\widehat{L}$ is the maximum value of the likelihood function for the model.

4. Results and discussion

4.1. Aggregate/disaggregate FA/FTA models

The linear FA/FTA models use the OLS technique, with the number of employees representing the business size of the establishments in the industry sector. The freight models resulting from these simple formulations possess many potential applications. The model results can take three separate cases/forms: (i) constant only, (ii) coefficient only, and (iii) both constant and coefficient. The final FA/FTA model results for all three cases are shown in Table 3. The manufacturing and whole sector models for both variants and the FTA part of the retail sector have the case (iii) types model forms, with a constant term and FA/FTA rate per employment. Both freight variants of the construction sector and the FA part of the retail sector have the case (ii) types model forms, with only a constant FA/FTA rate per employment. The number of employees measures the size of business establishments, where the unit increase tends to marginally increase the FA and FTA of the establishments, except for FA of wholesale, construction, and retail shows a parallel increase.

The employment-based FA/FTA models with the OLS techniques for our industry sectors have R-square values ranging from 11% to 84%. These R-square values are comparable with the results in other similar freight studies. For instance, the R-square value of up to 73% (Iding et al., 2002), ranges from 1.1% - 65.4% (Jaller et al., 2014), 15%–82% (Gonzalez-Feliu & Sánchez-Díaz, 2019), 58%–72% (Pani et al., 2018), and 15%–75% (Sahu & Pani, 2019).

4.2. Spatial effect indicators

The spatial autocorrelation effect is explained with the global Moran's I and LISA indicators for each industry sector. The summary of the analysis results on the spatial effect indicators is presented in Table 4 Table 3 and 4. The Moran's I value is close to zero for most of the cases and different from the expected value indicating some degree of spatial autocorrelation. The global tendency of spatial autocorrelation was conclusive (at the 5% confidence level) for construction and retail sectors in both FA and FTA. In addition, the local analysis reveals the presence of several positive associations or clusters between a polygon and its neighbours in the case of construction. The local analysis for the retail sector also reveals several positive and negative associations between neighbouring polygons. In this case, a polygon with a high amount of tonnage received or delivered is surrounded by polygons with a high tonnage received or delivered, and vice versa.

On the other hand, the manufacturing and wholesale sectors exhibit a global tendency toward dispersion, but the local analysis reveals several clusters and outliers. These two sectors show positive and negative associations between neighbouring polygons, where polygons with a high level of tonnage received or delivered are surrounded by polygons with a similar level of freight received or delivered. Overall, the LISA indicators imply the presence of spatial autocorrelation in almost all sectors. The findings from previous studies revealed that the systematic handling of spatial effects enhances FA models (Novak et al., 2011) and FTA models (Sánchez-Díaz et al., 2016). The next step is to evaluate these spatial effects on FA and FTA using variables explaining the built environment or the effects related to the shared neighborhood and locational effects.

4.3. OLS, SEM, and SAR model estimates

The estimated OLS and spatial models, SEM and SAR, are summarised in Table 5 for the FA part and Table 6 for the FTA part. The modelling starts with estimating OLS regression with multiple linear and nonlinear specifications for all industry sectors. The model misspecification was checked with the Ramsey regression equation specification error test (RESET) to confirm whether the nonlinear model specifications were significant. The RESET test was conclusive for the FA and FTA models of almost all the industry sectors, whereby rejecting the null hypothesis H_0 (all the nonlinear coefficients are zero) at a 5% level of confidence. The nonlinear models are log-transformed or in log-scale and estimated with an ordinary least squares regression technique that minimises the residuals sum of squares. In addition, the Pearson correlation and VIF values were checked in regression analysis, and the explanatory variables above the threshold values were eliminated from the analysis. The land-value (LV) variable has a high VIF value above the threshold in the FA and FTA of wholesale, retail, and construction sectors and is consequently eliminated from the subsequent analysis.

The nonlinear models were selected after extensive testing of various functional forms using employment as the main variable for all the industry sectors. In the case of FA models, the locational variables, including the width of the front street, the distance to the city gate, and the distance to the major market, play a significant role. However, land value and distance to the primary road network are not significant predictors of the freight attracted in all industry sectors. On the other hand, the locational variables significantly predicted the FTA models, including the width of the front street, the distance to the primary road network, and the major market. All the locational variables have positive and negative signs, as expected in both FA and FTA models of the industry sectors. Important locations in the city usually have wider streets, and establishments in these streets are likely to attract more tonnage and number of deliveries which may relate to their customer orientations and smaller space for inventory. Regarding network characteristics, the establishments near the primary road network, the city gate, and the major market location tend to attract more tonnage and frequent deliveries. Therefore, these variables have a negative sign in both FA and FTA models for all the industry sectors.

The FA models presented in Table 5, have a nonlinear relationship with employment, and the ln-coefficients range from 0.47 (wholesale) to 1.77 (construction). The coefficients reveal that larger establishments have higher FA than smaller establishments, but the magnitude varies largely over the industry sectors. For wholesale and retail establishments, the value of FA increases at a diminishing marginal rate where

Table 3
Employment-based freight attraction models (weekly - FA/FTA).

Industry Sectors (with SIC classification[a])	Freight Model	Obs.	Employment-based model			RMSE
			Constant/ Employment		Adjusted R-Square	
			Const.	Est.		
Manufacturing	FA	151	1.98 (1.25)[a]	0.28 (8.95)	0.56	72.17
	FTA	151	2.63 (6.3)	0.016 (8.38)	0.38	5.47
Construction	FA	19	–	1.78 (6.47)	0.69	363.3
	FTA	19	–	0.071 (10.06)	0.84	14.35
Wholesale	FA	33	41.4 (3.35)	1.61 (3.25)	0.25	70.08
	FTA	33	6.84 (2.82)	0.42 (4.54)	0.39	15.01
Retail	FA	239	–	1.13 (10.23)	0.31	3.6
	FTA	239	1.26 (7.46)	0.21 (5.23)	0.11	2.1

Note.
[a] SIC refers to an international standard for industrial classification; t-values are given with the bracket.

Table 4
Spatial effects indicators.

Description (SIC)	Variable	Moran's I		LISA		Remarks
		I	E(I)	Clusters	Outliers	
Manufacturing (SIC 20-39)	FA	0.02	−0.006	YES/HIGH & LOW	YES	LISA reveals several clusters and outliers
	FTA	0.021	−0.006	YES/HIGH & LOW	YES	LISA reveals several clusters and some outliers
Construction (SIC 15-17)	FA	0.94	−0.059	YES/HIGH	No	Moran I shows a global tendency toward clustering
	FTA	0.769	−0.059	YES/HIGH	No	Moran I shows a global tendency toward clustering
Wholesale (SIC 50-51)	FA	0.147	−0.031	YES/HIGH & LOW	YES	LISA reveals some clusters and one outlier
	FTA	0.13	−0.031	YES/HIGH & LOW	YES	LISA reveals some clusters and outliers
Retail (SIC 52-59)	FA	0.142	−0.004	YES/HIGH & LOW	YES	Moran I shows a global tendency toward clustering
	FTA	0.046	−0.004	YES/HIGH & LOW	YES	Moran I shows a global tendency toward clustering

Table 5
Results for econometric models, global models of freight attraction FA.

Dependent variable: ln(tons delivered/week)	Manufacturing	Construction		Wholesale	Retail		
	OLS	OLS	SEM	OLS	OLS	SEM	SAR
Spatial lag coefficient, ρ							0.37 (3.10)
Spatial error coefficient, λ			−0.55 (−2.13)			0.34 (2.48)	
Constant	−5.53 (−5.9)						
Economic characteristics							
Ln(Employment)	1.19 (11.57)	1.77 (9.39)	1.78 (14.08)	0.43 (2.28)	0.47 (2.3)	0.49 (2.47)	0.49 (2.49)
Land-use variable							
Ln(land market value)							
Network characteristics							
Ln(SW)	0.78 (2.93)			0.98 (2.19)	0.89 (3.85)	0.82 (3.57)	0.78 (3.46)
Ln(DPN)							
Ln(DCG)		−1.39 (−2.65)	−1.61 (−5.16)				
Ln(DMM)					−0.56 (−6.95)	−0.47 (−4.34)	−0.3 (−2.89)
n	151	18	18	34	243	243	243
Adj. R^2	0.56	0.86	0.91	0.26	0.29	0.33	0.34
F-Stat	39.09	28.12	n.a.	4.85	21.42	n.a.	n.a.
AIC	532	36.95	34.89	102.8	957.23	952.3	949.92
Log-likelihood at convergence	−260.04	−13.47	−12.44	−47.4	−472.61	−470.15	−467.96

Note: t-statistics are shown in round brackets. All the variables given are significant at 5% and 10% levels, (*) indicates significance at 10% level.

Table 6
Global models of freight trips attraction FTA.

Dependent variable:ln(number of deliveries/week)	Manufacturing	Construction		Wholesale		Retail
	OLS	OLS	SEM	OLS	SEM	OLS
Spatial lag coefficient, ρ						
Spatial error coefficient, λ			−0.48 (−1.72)*		−2.95 (−2.86)	
Constant	−2.03 (−3.89)					
Economic characteristics						
Ln(Employment)	0.58 (10.1)	0.91 (7.06)	0.87 (9.48)	0.53 (2.50)	0.53 (3.31)	0.28 (4.46)
Land-use variable						
Ln(land market value)						
Network characteristics						
Ln(SW)	0.37 (2.52)			1.13 (2.21)	1.4 (3.66)	0.12 (1.7)*
Ln(DPN)	−0.12 (−2.33)		0.12 (2.14)			
Ln(DCG)		−0.7 (−1.93)*	−0.77 (−3.35)			
Ln(DMM)						
n	151	18	18	34	34	243
Adj. R^2	0.51	0.80	0.86	0.26	0.45	0.13
F-Stat	31.78	18.29	n.a.	4.96	n.a.	7.99
AIC	349.36	23.25	22.42	111.75	107.96	379.93
Log-likelihood at convergence	−168.68	−6.62	−6.21	−51.87	−49.98	−183.96

Note: t-statistics are shown in round brackets. All the variables given are significant at 5% and 10% levels, (*) indicates significance at 10% level.

the ln-coefficients of employment are less than 1. According to Eqn. (12), the back-transformed equation of FA for the retail sector (similar to the wholesale sector), a smaller marginal increase in the FA resulted as the employment increases. However, the ln-coefficients of employment for the manufacturing and construction sectors are greater than 1, where the value of FA increases parallel with employment. In the case of FTA, the ln-coefficients of employment are less than 1 for all the industry sectors, and the FTA values marginally increase with the increase in the number of employees. Sánchez-Díaz et al. (2016) also found similar results regarding the relationship between the FTA and employment in the trade sector establishments.

$$FA = E^{0.47} \times SW^{0.98} \times DMM^{-0.56} \tag{12}$$

The variables characterising the locational effects relating the FA/FTA with the urban environment significantly improve the model fits. The front street width is a significant predictor similar to the FA and FTA of manufacturing, wholesale and retail establishments.

The ln-coefficient of street width is smaller than 1 for both FA and FTA of all the industry sectors, except the FTA of wholesale has a value greater than 1. The nearness to the closest city gate is statistically significant on both the FA and FTA of the construction establishments, which makes more sense due to construction materials production sites mainly located outside the city. The distance to the primary road network significantly affects the FTA of firms in the construction sector. In addition, nearness to the major market is a significant predictor, in which retailers near these locations are likely to receive more tonnage.

The SEM and SAR models account for spatial dependence and enhance OLS models' performance for FA and FTA across almost all the industry sectors, as depicted in Tables 5 and 6 The SEM model is a significant predictor of FA for the construction and retail sectors and the FTA models for the construction and wholesale industry sectors. In addition, the variable DPN is not significant in the OLS model of FTA for the construction sector and became significant in the corresponding SEM model. The SAR model is only significant for the FA model of the retail sector. The spatial models (SEM and SAR) are jointly significant for FA of the retail sector, and especially the SAR model exhibits a better fit to address the spatial autocorrelation. Generally, the coefficient estimates have changed from the OLS models to the spatial models, i.e., the resulting coefficient estimates became less biased when addressing the effects of the spatial autocorrelation. The SEM and SAR models show slight improvements over the OLS models with higher adjusted R-square and lower AIC values. Still, these models can be improved by incorporating other spatial effects, such as spatial heterogeneity, using local regression models, as presented in the next sections.

4.4. GWR/MGWR and GWR/MGWR-SAR models: a spatial interpretation

The estimation of the local regression models was applied only to the FA model of the retail industry sector, where the SAR model exhibited a significant improvement over the OLS model. Using the SAR model makes possible the calibration of the GWR/GWR-SAR model to simultaneously deal with the spatial autocorrelation and spatial heterogeneity problems. The GWR calibrates the regression model separately at each location but with a similar scale (bandwidth) for each relationship in a given model. Applying GWR can result in biased estimates when the data has multiple spatial scales. The MGWR is an extension of GWR that can eliminate the scale problem by allowing each local relationship to occur at a different scale.

The spatial variability of the estimated parameters was checked with the Monte Carlo (MC) test, which clearly indicates whether to use GWR or MGWR. The MC test indicated that not all the variables could be characterised as local. The explanatory variables *E* and *SW* are spatially stationary and only have a global effect on the model. Therefore, the final models are MGWR and MGWR-SAR, as presented in Table 7. The average values of the parameter of both MGWR and MGWR-SAR are not similar to the estimates of OLS and SAR models due to the calibration of the MGWR model with multiple distinct spatial scales.

The problem of local multicollinearity occurs for some network characteristic variables in both MGWR and MGWR-SAR models, but its effect reduces from MGWR to MGWR-SAR. As expected, the model with the MGWR-SAR has a better model fit in terms of adjuster R-square, AIC, AICc, and log-likelihood estimation. The non-stationary spatial variables with the MC test have larger differences between the minimum and maximum values in both variants of the local models. The most important explanatory variable of FA is the number of employees at the establishment, as confirmed with 100% and 82.72% significance for MGWR and MGWR-SAR models, respectively.

Another variable affecting the tonnage attracted to retail establishments is the street width (*SW*), with more than 63% significance for MGWR-SAR but not significant in the MGWR model. Apart from having better overall significance, the MGWR-SAR model has improved the significance of individual variables, especially the poorly fitted ones.

Table 7
MGWR and MGWR-SAR estimates of freight attracted (FA) for Retail Trade.

Dependent variable: ln (tons/week)	Mean	SD	Min	Median	Max	Adj. critical t-value (95%)	Percentage of Significance at 95% level	Percentage of Cases with Local VIF >10	MC
	MGWR estimates								
Intercept	0.029	0.253	−0.386	0.074	0.312	2.531	40.74%	–	
Ln(E)	0.147	0.001	0.145	0.147	0.148	1.981	100.00%	0.00%	
Ln(SW)	0.020	0.109	−0.307	0.024	0.127	2.004	0.00%	0.00%	
Ln(DPN)	0.199	0.003	0.190	0.199	0.204	2.392	100.00%	0.00%	
Ln(DCG)	−0.084	0.117	−0.240	−0.123	0.203	2.295	35.39%	33.33%	
Ln(DMM)	−0.180	0.000	−0.180	−0.180	−0.180	1.971	0.00%	55.14%	
Kernel function	Fixed Gaussian		N	243		AICc	585.61	Log-likelihood	−278.63
Optimal bandwidth Criteria	AICc		Adj. R^2	0.42		AIC	583.94		
	MGWR-SAR estimates								
Intercept	−9.605	13.528	−31.450	−7.171	3.003	2.789	100.00%	–	SN
Ln(E)	0.066	0.001	0.065	0.066	0.066	1.978	82.72%	0.00%	SS
Ln(SW)	0.133	0.016	0.092	0.140	0.156	2.089	63.37%	0.00%	SS
Ln(DPN)	0.007	0.138	−0.382	0.043	0.217	2.577	57.79%	0.00%	SN
Ln(DCG)	0.052	0.920	−2.477	−0.102	2.078	2.688	26.75%	8.00%	SN
Ln(DMM)	3.090	4.415	−0.063	0.903	12.584	2.607	25.10%	51.00%	SN
Wy*ln(FA)	−9.194	12.384	−31.749	−2.456	0.078	2.937	46.91%	9.00%	SN
Kernel function	Fixed Gaussian		N	243		AICc	443.87	Log-likelihood	−170.54
Optimal bandwidth Criteria	AICc		Adj. R^2	0.71		AIC	425.58		

Notes: MGWR 2.2 software was used to calibrate MGWR and MGWR-SAR models. MC denotes the Monte Carlo test for spatial variability. SN denotes spatial non-stationarity, and SS denotes spatial stationarity at a 5% level of significance.

Specifically, the network characteristics variables *SW* and *DMM* are not statistically significant in MGWR and largely improved in MGWR-SAR. Generally, the local models (both MGWR and MGWR-SAR) have a general advantage over the OLS and SAR models, where a variable does not require to be significant throughout the spatial analysis area.

The proximity to the primary road network (DPN) is the next important variable affecting the freight attracted (FA) to retail establishments. The spatial analysis results of the estimated parameters for the DPN variable are presented in Fig. 3. The variable DPN has a stronger and more significant impact on the FA of retail establishments near the city centre due to the higher density of the primary road network in those locations. Moreover, the distance to the closest city gate *DCG* (as shown in Figure A2) predominantly impacts the amount of fright attracted to the retail establishments located in the southern part of the city. The nearest city gate to this location (locally named Kality gate) is often regarded as the main freight entry/exit gate to the city serving more than half of the daily freight traffic (Kebede & Gebresenbet, 2017). Retail establishments closer to the city gate can interact more with freight generators outside of the city due to their locational proximity and easily attract more freight.

Addis Ababa city has a major market located in the northwest central part and is referred to as one of the biggest open markets in the African continent. The distance to the major market location, *DMM*, is another important variable characterising the locational attributes in the FA of retail firms. The variable *DMM* is a proxy for the relative location of a retail establishment that can relate to the other locational variable *DCG*. The retail establishment near the major market tends to have more agglomeration with neighbouring establishments. The highest impact of this variable was exhibited for retail establishments located in the eastern and southern parts of the city, where becoming closer to this major market has the key importance of receiving more freight tonnage.

The MGWR-SAR model estimates also include the spatial autoregressive parameter and allow the estimation of this parameter to change significantly over the spatial analysis area, as presented in Fig. 4. The spatial autoregressive parameter can be better explained in correlation with the above DCG and DMM results (Figure A2 and Figure A3). The spatial autoregressive parameter is significant at mid-way locations between the major marker and the city gates in the eastern and southern parts. Moreover, retail establishments at those locations have higher spatial dependence, implying their relative advantage to interact with other establishments at the major market within and outside the city administrative boundary. Conversely, for the establishments located in an area with low or insignificant spatial dependence, the freight analysis better employs attributes of the retail establishment itself and its close neighbours other than the locational characteristics.

4.5. Compare the model fit

The performance of the models developed with OLS, SEM, SAR, MGWR, and MGWR-SAR are compared using three indicators: adjusted R-square, AIC, and RMSE, as presented in Table 8. From those indicators, the MGWR-SAR is the best method to model freight generation at a retail establishment and its interaction with the surrounding environment. The MGWR-SAR model explains the variance of the freight attraction pattern more than twice when compared to the global models (OLS, SEM, SAR) and shows a vast improvement over the local model with the MGWR. The MGWR-SAR model explains more than two-thirds of the variances in the FA pattern of retail establishments, and the remaining variance might be due to the effect of other unobserved factors not included in our analysis. This model simultaneously addresses the problems of nonlinearity, spatial dependency, and spatial heterogeneity.

5. Conclusions

In this study, the objective was to enhance the prediction of freight patterns of business establishments accounting for their spatial and locational characteristics. The accuracy of estimating freight (trip)

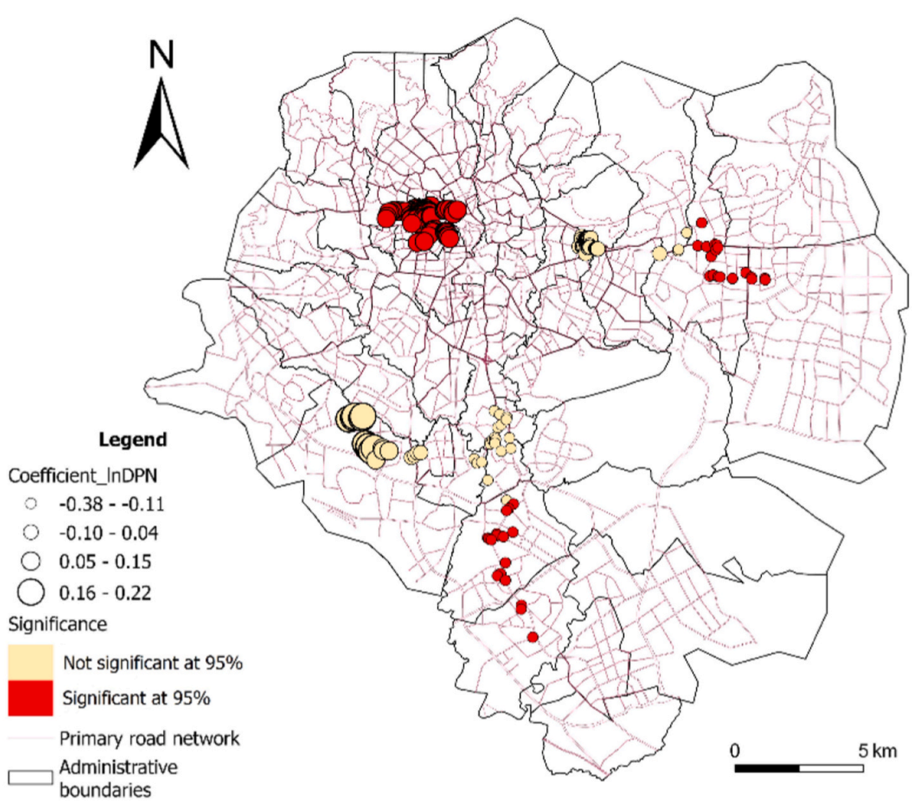

Fig. 3. DPN, local coefficients and statistical significance of estimates.

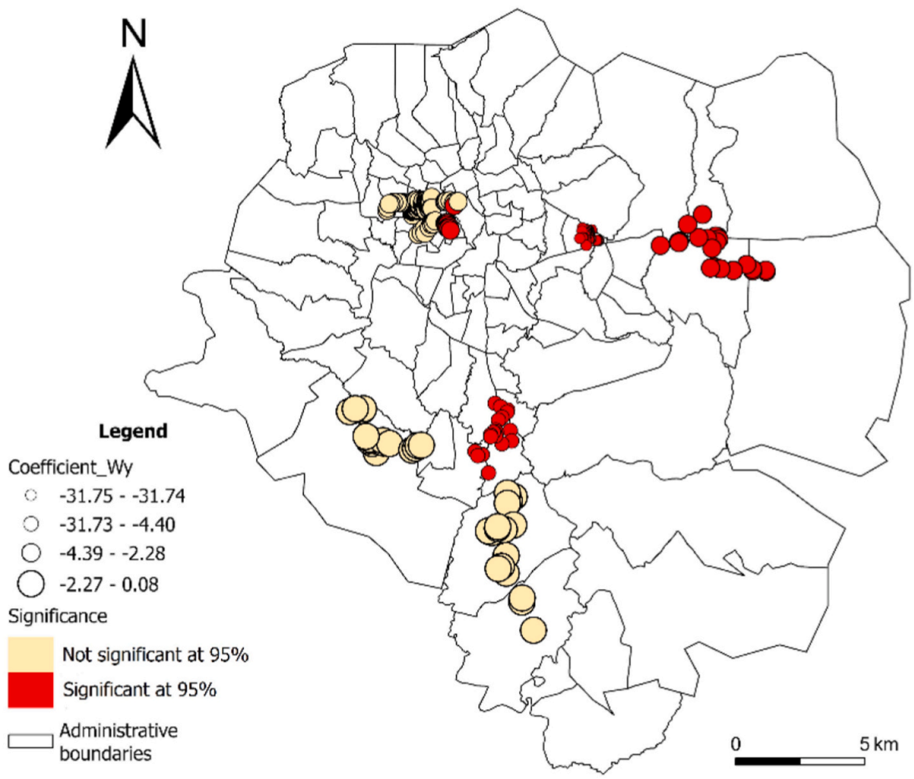

Fig. 4. Spatial autoregressive term Wy*ln(FA), local coefficients and statistical significance of estimates.

Table 8
Comparison of goodness-of-fit measures for FA model, retail trade sector.

Models	Adj. R^2	AIC	RMSE
OLS	0.29	957.2	1.69
SEM	0.33	952.3	1.67
SAR	0.34	949.9	1.65
MGWR	0.42	583.9	0.76
MGWR-SAR	0.71	425.6	0.49

attraction (FA/FTA) has low explanatory power when using global modelling techniques, such as ordinary least squares (OLS) regression. The spatial econometric techniques considered here have the capacity to consider local spatial variability and to incorporate spatial and locational determinants, improving the predictive accuracy of freight models. To demonstrate this, the study used establishment-based freight survey data collected from Addis Ababa City, Ethiopia. The relationship between variables explaining the establishment with its location and freight (trip) attraction is modelled with several spatial techniques, including spatial error model (SEM), spatial autoregressive model (SAR), geographically weighted regression (GWR), multiscale-GWR (MGWR), and the combination GWR-SAR/MGWR-SAR.

The variables used to predict FA/FTA are employment at the business establishment level, together with other locational variables. The nonlinear functional specification is better suited to model the FA/FTA with predictor variables. A finding in line with previous studies is that the coefficient of employment ranges between zero and one for all FTA models across industry sectors, indicating that FTA is increasing with employment at a diminishing marginal rate. However, the FA also shows a parallel increase with employment. These results are conceptually sound and have the substantial implication of supporting the notion that FA and FTA are outputs of different decisions at the establishment level. On the other hand, the spatial and locational variables show different influences depending on the industry sector and the metrics for measuring freight activity.

The spatial and locational variables indicating street width, primary road network and nearness to the city gate affect the FA/FTA of establishments. For the construction sector, establishments located near the city gate tend to have high FA and FTA. For the manufacturing industry, establishments closer to the primary road network tend to attract more freight. The manufacturing, wholesale and retail establishments tend to have higher FA and FTA when located along wider streets. In addition, retail establishments near the major market tend to have a higher FA. The FA models of retail establishments using the MGWR and MGWR-SAR account for the local variability and autocorrelation effects. Compared to OLS, SEM and SAR models, these models show superior performance on important metrics including R-square, AIC and RMSE.

The research addresses inherent spatial and locational influences in business establishments' freight patterns (FA and FTA). Regarding its applicability, the complexity of incorporating spatial processes can be reduced considerably by using geographic information systems (GIS) for data collection and spatial representation, together with use of open-source tools (such as MGWR 2.2) for spatial analysis. The local spatial GIS-based models allow a high accuracy in explaining the local area's unique characteristics and create a better understanding of spatial patterns in freight activity. Additionally, the models offer a visual representation of the results that facilitate communication with decision-makers. With the presented improvements in predicting freight demand, these models will be useful for making better informed decisions regarding infrastructure planning, operational strategies and designing new or refining existing urban freight policies. Further research could investigate the effect of spatial determinants on freight patterns of establishments in non-freight intensive industry sectors, such as service sectors. Additionally, work is needed on the transferability of the proposed spatial models to a different location and their applicability to other samples, to compare with the improvements found here and strengthen the generalizability of these results.

CRediT authorship contribution statement

Abel Kebede Reda: Conceptualization, Methodology, Data curation, Software, Validation, Formal analysis, Writing – original draft. **Lori Tavasszy:** Conceptualization, Methodology, Writing – review & editing. **Girma Gebresenbet:** Conceptualization, Methodology, Writing – review & editing, Project administration. **David Ljungberg:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

No conflict of interest exists in the submission of this manuscript, and that is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and is not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

Acknowledgement

The authors acknowledge the support and funding for this research by EP-NUFFIC, the Dutch Organization for Internationalization of Education through the CBP-Ethiolog project (NICHE-ETH-285). The authors are grateful for the statistical inputs from Claudia von Brömssen (Ph.D.) in the data collection plans and the analysis. Also, they appreciate the insights from Jose Holguin-Veras (Professor) to improve the analysis part.

Appendix

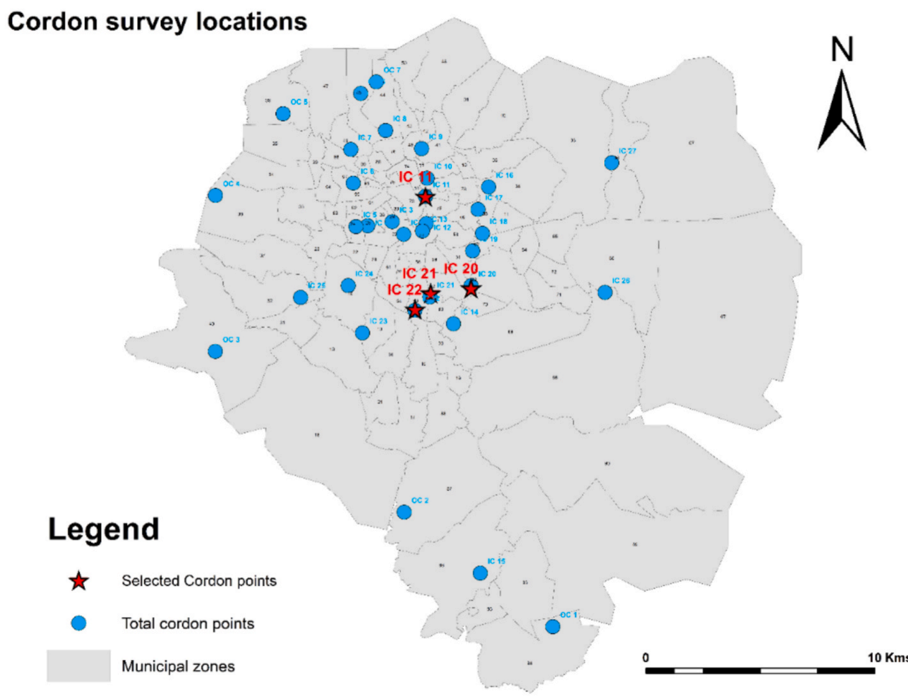

Fig. A1. Cordon survey locations in Addis Ababa (Addis-Ababa-City-Transport-Authority, 2018)

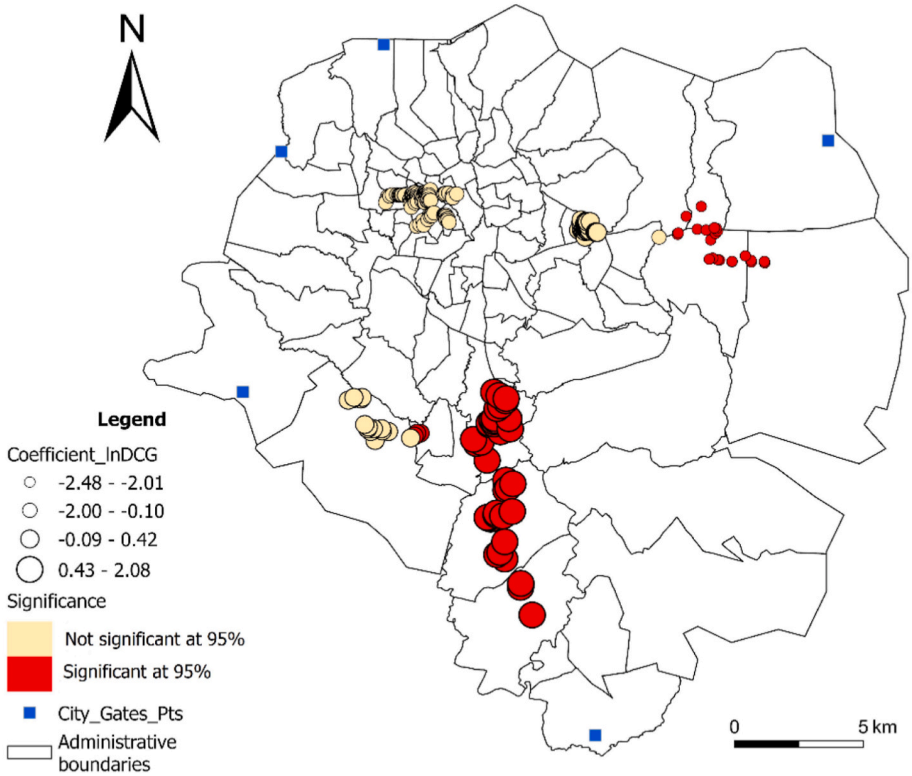

Fig. A2. DCG, local coefficients and statistical significance of estimates

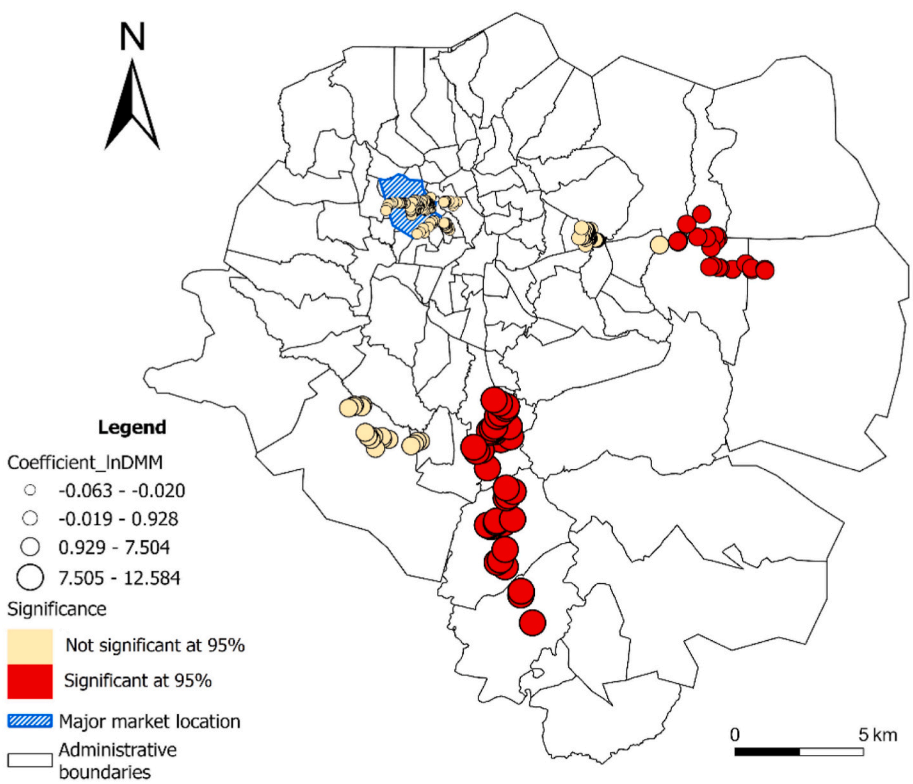

Fig. A3. DMM, local coefficients and statistical significance of estimates

References

Addis-Ababa-City-Transport-Authority. (2018). *Urban freight transport management of Addis Ababa. Addis Ababa, Ethiopia*.

Alho, A. R., & Silva, J. D. A. E. (2014). Freight-trip generation model: Predicting urban freight weekly parking demand from retail establishment characteristics. *Transportation Research Record, 2411*, 45–54.

Alho, A. R., De Abreu E Silva, J., De Sousa, J. P., & Blanco, E. (2018). Improving mobility by optimising the number, location and usage of loading/unloading bays for urban freight vehicles. *Transportation Research Part D: Transport and Environment, 61*, 3–18. https://doi.org/10.1016/j.trd.2017.05.014.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. (Boston, MA): kluwer academic.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review, 26*, 153–166.

Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science, 92*, 3–17.

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs, 155*, 237–290.

Anselin, L., Dodson, R. F., & Hudak, S. (1993). Linking GIS and spatial data analysis in practice. *Geographical Systems*.

Anselin, L., Syabri, I., & Kho, Y. (2010). *GeoDa: An introduction to spatial data analysis. Handbook of applied spatial analysis*. Springer.

Balla, B. S., Sahu, P. K., Pani, A., Sharma, S., & Majumdar, B. B. (2022). Comparison of parametric and non-parametric methods for modeling establishment-level freight generation. *Transportation Research Record*, Article 03611981221116369, 0.

Basile, R., Durbán, M., Mínguez, R., María Montero, J., & Mur, J. (2014). Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control, 48*, 229–245.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (2002). Geographically weighted summary statistics — a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems, 26*, 501–524.

Da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis, 48*, 233–247.

Institute-Of-Transportation-Engineers. (2008). In I. O. T. ENGINEERS (Ed.), *Trip GenerationGeneration: An ITE informational report* (Washington, D.C).

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the Association of American Geographers, 107*, 1247–1265.

Garrido, R. A., & Mahmassani, H. S. (2000). Forecasting freight transportation demand with the space–time multinomial probit model. *Transportation Research Part B: Methodological, 34*, 403–418.

Geniaux, G., & Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics, 72*, 74–85.

Gonzalez-Feliu, J., & Sánchez-Díaz, I. (2019). The influence of aggregation level and category construction on estimation quality for freight trip generation models. *Transportation Research Part E: Logistics and Transportation Review, 121*, 134–148.

Griffith, D. A., Layne, L. J., Layne, L. J., Ord, J., & Sone, A. (1999). *A casebook for spatial statistical data analysis: A compilation of analyses of different thematic data sets*. Oxford University Press on Demand.

Holguín-Veras, J., Jaller, M., Destro, L., Ban, X., Lawson, C., & Levinson, H. S. (2011). Freight generation, freight trip generation, and perils of using constant trip rates. *Transportation Research Record, 2224*, 68–81.

Holguín-Veras, J., Jaller, M., Sánchez-Díaz, I., Campbell, S., & Lawson, C. T. (2014). 3 - freight generation and freight trip generation models. In L. TAVASSZY, & G. DE JONG (Eds.), *Modelling freight transport*. Oxford: Elsevier.

Holguín-Veras, J., Sánchez-Díaz, I., Lawson, C. T., Jaller, M., Campbell, S., Levinson, H. S., & Shin, H.-S. (2013). Transferability of freight trip generation models. *Transportation Research Record, 2379*, 1–8.

Iding, M. H., Meester, W. J., & Tavasszy, L. (2002). *Freight trip generation by firms*.

Jaller, M., Sanchez-Diaz, I., Holguín-Veras, J., & Lawson, C. T. (2014). *Area based freight trip generation models*. Washington DC: Transportation Research Board 93rd Annual Meeting.

Kawamura, K., & Miodonski, D. (2012). *Examination of the relationship between built environment characteristics and retail freight delivery*.

Kebede, A., & Gebresenbet, G. (2017). Mapping out goods flow to Addis Ababa city, Ethiopia, and its impact on environment. *Transportation Research Procedia, 25*, 1008–1020.

Middela, M. S., & Ramadurai, G. (2021). Incorporating spatial interactions in zero-inflated negative binomial models for freight trip generation. *Transportation, 48*, 2335–2356.

Miller, H. J. (1999). Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis, 31*, 373–399.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*, 17–23.

Novak, D. C., Hodgdon, C., Guo, F., & Aultman-Hall, L. (2011). Nationwide freight generation models: A spatial regression approach. *Networks and Spatial Economics, 11*, 23–41.

Openshaw, S., Taylor, P. J., & Wrigley, N. (1979). *Statistical applications in the spatial sciences* (pp. 127–144). London: N. Wrigley. Pion.

Ortúzar, J., & Willumsen, L. G. (2011). *Modelling transport*. John wiley & sons.

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Fotheringham, A. S. (2019). Mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information, 8*, 269.

Páez, A., & Scott, D. M. (2004). Spatial statistics for urban analysis: A review of techniques with examples. *Geojournal, 61*, 53–67.

Pani, A., Sahu, P. K., & Bhat, F. A. (2021). Assessing the spatial transferability of freight (trip) generation models across and within states of India: Empirical evidence and implications for benefit transfer. *Networks and Spatial Economics, 21*, 465–493.

Pani, A., Sahu, P. K., Chandra, A., & Sarkar, A. K. (2019). Assessing the extent of modifiable areal unit problem in modelling freight (trip) generation: Relationship between zone design and model estimation results. *Journal of Transport Geography, 80*, Article 102524.

Pani, A., Sahu, P. K., Patil, G. R., & Sarkar, A. K. (2018). Modelling urban freight generation: A case study of seven cities in Kerala, India. *Transport Policy, 69*, 49–64.

Sahu, P. K., Chandra, A., Pani, A., & Majumdar, B. B. (2020). Designing freight traffic analysis zones for metropolitan areas: Identification of optimal scale for macro-level freight travel analysis. *Transportation Planning and Technology, 43*, 620–637.

Sahu, P. K., & Pani, A. (2019). *Freight generation and geographical effects: Modelling freight needs of establishments in developing economies and analysing their geographical disparities*. Transportation.

Sánchez-Díaz, I. (2017). Modeling urban freight generation: A study of commercial establishments' freight needs. *Transportation Research Part A: Policy and Practice, 102*, 3–17.

Sánchez-Díaz, I., Holguín-Veras, J., & Wang, X. (2016). An exploratory analysis of spatial effects on freight trip attraction. *Transportation, 43*, 177–196.

Ward, M. D., & Gleditsch, K. S. (2018). *Spatial regression models*. Sage Publications.