Spatial Interaction Modeling

Abstract The concept of spatial interaction (SI) encapsulates the domain of human activities that occur between a set of locations embedded within geographical space. Data about such processes are essential for studying a wide spectrum of geographic phenomena that are important to society, such as the accessibility of services, product demand, transportation trends, and demographic dynamics. In particular, SI models seek to explore, explain, and predict aggregate movements or flows that occur across an abstract or physical network, which can be useful on its own, as well as a factor within other regional models. As the number and nature of SI modeling applications have grown, the associated theory and tools have simultaneously evolved to consider more complex spatial relationships, resulting in numerous expansions of the modeling paradigm. In this chapter, some foundations of SI modeling are first laid out before presenting a simple demonstration and then describing several extensions to the core modeling methodology.

Keywords: spatial interaction; movement; flows; calibration; spatial effects

1. Introduction

The concept of spatial interaction (SI) encapsulates the domain of human activities that occur between a set of locations embedded within geographical space, such as the migration or commuting of people, the flow of trade goods or retail expenditures, or the diffusion of knowledge. Data about such processes are essential for studying a wide spectrum of geographic phenomena that are important to society, such as the accessibility of services, product demand, transportation trends, and demographic dynamics, making it a core concept within spatially-centered research (Haynes & Fotheringham, 1984; Fotheringham & O'Kelly, 1989; Sen & Smith, 1995; Roy, 2004; Batty, 2013). While the concept has a long history in the social sciences[1], it was in the mid-20th century that formal models of SI processes started to become widespread (Farmer & Oshan, 2017). Interest persisted and intensified throughout the second half of the century, along with a renewed focus on human movement under the banner of 'human mobility' at the turn of the century. This latter trend is primarily due to the widespread availability of spatially and temporally disaggregate mobility datasets from sources such as automated transportation systems, mobile phone records, GPS trajectories, and social media within the modern data ecosystem (Arribas-Bel, 2014; Manley & Dennett, 2019).

In particular, SI models seek to explore, explain, and predict aggregate movements or flows that occur across an abstract or physical network. Their aggregate nature provides an important tool that can avoid privacy issues associated with individual-based data while providing a link to techniques and theories for understanding individualistic behavior (Anas, 1983; Fotheringham, 1986b). SI modeling is also a central technique for understanding movement decision-making since it considers the attributes of places against the costs that must be overcome to travel between them. The generalizability of the framework has allowed it to be applied in diverse settings. Furthermore, knowledge of location choice preferences is important for policy development, which can be useful on its own, as well as a factor within other regional models, such as, land-use/land-change, market analysis, or location allocation. As the number and nature of SI modeling applications have grown, the associated theory and practical tools have simultaneously evolved to consider more complex spatial relationships, resulting in numerous expansions of the modeling paradigm.

In this chapter, some foundations of SI modeling are first laid out before describing available software options for carrying out SI modeling and demonstrating an application of modeling and predicting

[1] See, for example, (Carey, 1858).

bike-share trips in New York City. Several extensions to the core modeling methodology are then surveyed. Finally, a few concluding remarks are put forth that look to the future of SI modeling.

2. Foundations

2.1. From gravity to spatial interaction

The underlying hypothesis of SI models is that the volume of flows (edges) between origin locations and destination locations (nodes) is a function of the potential at each origin, the attractiveness of each destination, and the cost of overcoming the separation between each origin and destination (see figure 1). In particular, there is often special attention paid to the role of physical separation within these models, which is usually captured by inter-location distances. It is therefore unsurprising that early SI models were inspired by the physical law of gravitational attraction between two bodies (e.g., Ravenstein, 1885; Reilly, 1929; Stewart, 1941; Zipf, 1946; Dodd, 1950; Carrothers, 1956; Tinbergen, 1962), where the number of flows between two locations is given by the product of the masses of the origin and destination, divided by the distance between them. Farmer & Oshan (2017) further discuss some of the expansions upon this simple conceptual framework. Developments typically entailed specifying domain-specific variations of the model and/or defining additional components that accommodate the complexities of the subjects being modeled (see table 1). Regardless, models across domains, applications, and scales tend to share a common form, which can be generalized equivalently in either of the following manners,

$$T_{ij} = k\frac{V_i^{\mu} W_j^{\alpha}}{d_{ij}^{\beta}} \tag{1}$$

or equivalently,

$$T_{ij} = kV_i^{\mu} W_j^{\alpha} d_{ij}^{\beta} \tag{2}$$

where T_{ij} represents flows from origin i to destination j, V_i and W_j are origin and destination attributes, respectively, d_{ij} is the costs to overcome the physical separation between i and j (usually distance or time), k is a scaling factor that ensures the total number of modeled trips is equivalent to the total number of observed trips, and μ, α, and β are exponential parameters (Fotheringham & O'Kelly, 1989). In equation (1), β is in the denominator and in equation (2) it is typically expected that β takes on a negative value, in both cases capturing the deterrence of geographical separation and is therefore referred to as the 'distance-decay' or the 'friction-of-distance'. In some cases, one or more of the parameters are assumed to take on a fixed value (i.e., $\mu = 1$, $\alpha = 1$, and $\beta = 2$), harkening back to the simpler gravity analogies.

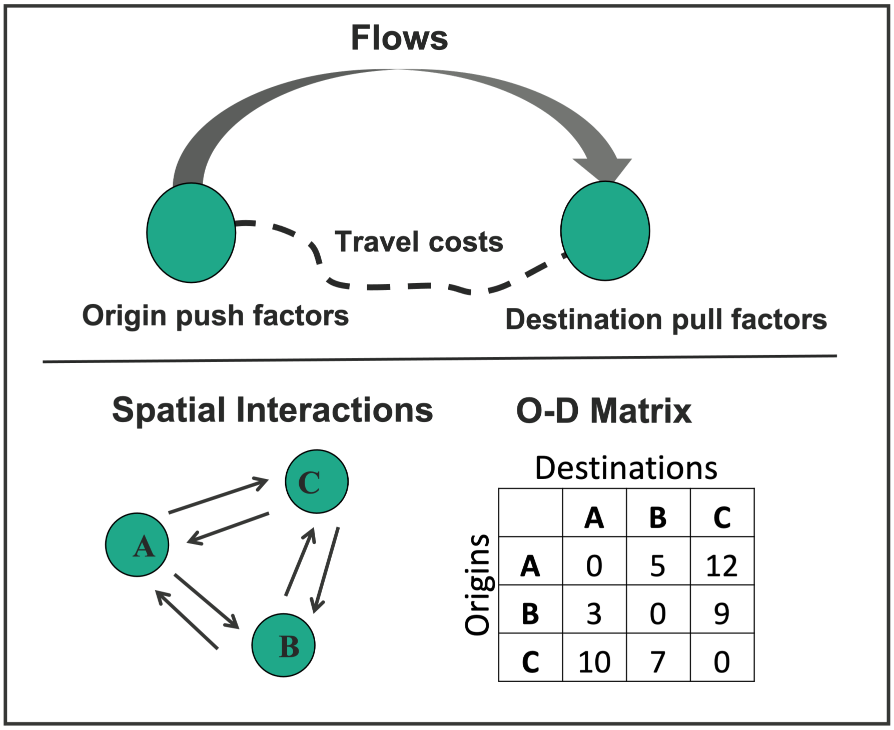

Figure 1: Origin-destination flow logic (top) & an illustrative SI system (bottom)

Table 1: Common spatial interaction phenomena and abstractions (expanded from Fotheringham, (2017)).

Theme	Flows (edges)	Locations (nodes)	Origin attributes	Destination attributes
Migration	Change of residence	Neighborhoods; Cities; States or provinces; Countries	Population; Crime; Jobs; Income	Population; Jobs; Income; Weather
Trade & macroeconomics	Commodity exchange; Investment ($)	Countries; Cities; Industrial sites	Supply; Population; GDP	Demand; Population; GDP
Retail & business intelligence	Shopping trips; Expenditure ($)	Residences/Stores	Population; Households; Income	Store size; Prices; Parking; Variety
Transport	Commuting; Traffic; Transit usage	Stations; Neighborhoods/ Offices; Airports	Population; Income; Car owners; Capacity	Jobs; Services; Capacity; Population

However, in the contemporary social sciences it is more common to recognize that nuances in context, populations, geography, and data sources imply substantial variation in the underlying factors generating flows within and across different SI systems. Therefore, when data for the individual model components are available, the parameters may be calibrated (also called estimation), summarizing the effect that each component contributes towards explaining the system of flows. The parameters can also be used in planning scenarios to predict unknown flows when there are deviations in model components or locations are added or deleted from the system. Due to the utility and flexibility of the parameters of SI models, a central focus of the framework is the estimation of them and their associated uncertainty.

By applying the statistical theory of entropy-maximization, Wilson (1967, 1971) analytically derived a "family" of gravity-type SI models. The essence of the derivation is to solve an optimization problem that finds the most probable configuration of flows between a set of locations out of all possible configurations, while making minimal assumptions about the system. Constraints can then be added to the optimization to obtain each member of the family of models. Ensuring that the total number of flows observed in a dataset is conserved in the modeled flows yields

$$T_{ij} = V_i^{\mu} W_j^{\alpha} exp(d_{ij}) \tag{3}$$

which is the gravity model introduced in equation (1) but with an exponential distance-decay function instead of a power distance-decay function and k is no longer included because it is implied in the nonlinear model derivation and calibration. This model is also referred to as the total flow constrained SI model or the unconstrained SI model, since no additional constraints are included beyond the total number of trips. If further constraints are introduced to also conserve the total inflows, outflows, or both, the production-constrained (destination-constrained), attraction-constrained (origin-constrained), or doubly-constrained models are obtained, respectively, as

Production-constrained

$$T_{ij} = A_i O_i W_j^{\alpha} exp(\beta d_{ij}) \tag{4}$$

$$A_i = \left(\sum_j W_j^{\alpha} exp(\beta d_{ij})\right)^{-1} \tag{4.1}$$

Attraction-constrained

$$T_{ij} = B_j D_j V_i^{\mu} exp(\beta d_{ij}) \tag{5}$$

$$B_j = \left(\sum_i V_i^{\mu} exp(\beta d_{ij})\right)^{-1} \tag{5.1}$$

Doubly-constrained

$$T_{ij} = A_i B_j O_i D_j exp(\beta d_{ij}) \tag{6}$$

$$A_i = \left(\sum_j B_j D_j exp(\beta d_{ij})\right)^{-1} \tag{6.1}$$

$$B_j = \left(\sum_i A_i O_i exp(\beta d_{ij})\right)^{-1} \tag{6.2}$$

where O_i and D_j are the total number of flows emanating or terminating at an origin or destination, A_i and B_j are balancing factors that ensure these totals are preserved in the predicted flows. Notice that the balancing factors in the doubly–constrained model (6.1-6.2) are interdependent and need to be computed iteratively when they are being calculated directly. It should also be noted that the power function arising from the physical gravity analogy can be simply substituted into equations (3-6) and represents a logarithmic evaluation of transport costs, which is the default form adopted throughout the remainder of the chapter. Alternative distance-decay functions have been explored (Vries et al., 2009; Martínez & Viegas, 2013), though they tend to be less common in practice.

The production-constrained and attraction-constrained models conserve either the number of total inflows or outflows at each location, making them particularly useful for building models that allocate flows either to a set of destinations or from a set of origins, respectively. The doubly-constrained model conserves both the inflows and the outflows at each location and is therefore sometimes referred to as the trip distribution model. An increase in constraints is typically associated with increased predictive power but is also associated with fewer estimated parameters and a decrease in the potential to gain insight into the SI system (Fotheringham & O'Kelly, 1989).

2.2. Intervening opportunities

Perhaps the second most ubiquitous hypothesis pertaining to human movement is that of intervening opportunities (IO), which posits that "the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities" (Stouffer, 1940). It was later refined such that IO were defined as those locations within a circle with a radius given by the distance between the origin and destination under consideration (Stouffer, 1960). This is in contrast to gravity-type SI models, which focus directly on continuous measures of the costs associated with physical separation between locations. There is much research comparing the two frameworks (Kaltenbach, 1972; Haynes et al., 1973; Dison & Hale, 1977; Smith, 1980; Elffers et al., 2008), several variations have been set forth which seek to include both gravity and IO effects (Wills, 1986; Ulyssea-Neto, 1993; Cascetta et al., 2007), and additional model forms and procedures have been proposed (Schmitt & Greene, 1978; Rogerson, 1986; Akwawua & Pooler, 2001; Afandizadeh & Hamedani, 2012; Nazem et al., 2015). While the IO model has its own rich literature, it is deployed less often compared to gravity-type SI models that will remain the focus of this chapter.

3. Calibration, tools, and practical resources

Regression is perhaps the most frequently used calibration method due to its flexibility, although linear programming and non-linear optimization are also possible (i.e., Wilson & Senior, 1974; Batty & Mackie, 1972). Equation (2) can be linearized by taking the logarithm of both sides, yielding the so-called log-linear gravity model,

$$\ln T_{ij} = k + \mu \ln V_i + \alpha \ln W_j - \beta \ln d_{ij} \quad (7)$$

which can be expressed more generally as a log-normal regression specification and included within an ordinary least-squares regression specification as

$$\ln T_{ij} = k + \mu \ln V_i + \alpha \ln W_j + \beta \ln d_{ij} + \varepsilon_{ij} \quad (8)$$

where ε is a normally distributed error term with a mean of 0 and β is still expected to take on a negative value. However, a Poisson regression specification was proposed for the family of SI models over ordinary least squares regression because flows are often counts of people or objects and should be modeled as discrete entities and Poisson regression avoids the potential issue of taking the logarithm of zero-valued flows (Flowerdew & Aitkin, 1982; Flowerdew & Lovett, 1988; Fotheringham & O'Kelly, 1989; Silva & Tenreyro 2006; Farmer & Oshan, 2017). Another advantage of Poisson regression is that adding origin fixed effects and/or destination fixed effects (i.e., binary indicators or dummy variables) achieves the same objective as the balancing factors A_i and/or B_j in equations (4-6) to uphold the desired constraints (Tiefelsdorf & Boots, 1995). Suggestions have also been made to adopt a negative binomial distribution or zero-inflated extension when Poisson assumptions are not met (e.g., Burger et al., 2009; Metulini et al., 2018).

There are various resources and tools available for SI modeling. An exceptionally well-curated resource is the *IMAGE Studio* point-and-click tool (Stillwell et al., 2014); however, it is customized specifically for migration and is not cross-platform. Software and educational materials are available for SI modeling within the R statistical computing environment. Dennett (2012, 2018) provides informative tutorials in the context of migration modeling that walk-through of the deployment of the family of gravity-type SI models using the built-in Poisson GLM functionality for parameter estimation and custom scripts for iteratively computing balancing factors. Similarly, Crymble (2019) furnishes an additional example in R in the context of historical migration and Siah (2018) in the context of commuting. More formal SI modeling software packages in R include the *SpatialPosition* package (Giraud, 2015) that provides functions for some legacy models, the *simR* package (Brunsdon, 2018) for the family of gravity-type models, and the *gravity* package (Wölwer, 2018) for modeling within the context of trade and macroeconomics. The family of gravity-type SI models may also be calibrated in the Python scientific computing environment using the GLM functionality in the *statsmodels* module or using the *spint* module (Oshan, 2016). The latter has the advantage of being built upon sparse data structures that drastically decrease computation time for constrained members of the family and bypass memory bottlenecks associated with large SI systems. Oshan (2016) demonstrates basic *spint* functionality and highlights how to interpret parameter estimates and assess model fit based on several criteria. Moreover, new tools are emerging to efficiently process raw trajectory data into individual trips and to facilitate the construction of aggregate origin-destination flow matrices (Graser, 2019; Pappalardo et al., 2019). These resources contribute towards a rich SI modeling toolbox, encompassing the entire workflow to transform complex mobility data into SI information (e.g. Siła-Nowicka et al., 2016; Siła-Nowicka & Fotheringham, 2019).

4. Demonstrative example

To illustrate a SI system and the types of analytic actions that SI models support, a basic application using bike-share cycling trips in New York City is presented. The study area is displayed in figure 2 and consists of bike-sharing stations distributed across Manhattan and parts of Brooklyn and Queens. Stations have varying capacities (i.e., number of bike docks) with the largest stations being located in Midtown and Downtown Manhattan. Individual cycling trips are recorded between stations along with the duration of the trip, which can be aggregated to an origin-destination matrix of flows. Alternatively, distances between stations could be calculated 'as the crow flies' or more realistically along the road network rather than using the reported trip duration as a proxy for separation[2]. This data also includes limited information about the rider, such as gender and age, which can be used to filter and subset the trip data prior to aggregation into flows. After organizing the flow observations into tabular format, information regarding the capacity of each origin and destination station was merged into the dataset yielding columns={*origin_station, destination_station, trip_count, trip_duration, origin_capacity, destination capacity*}, allowing a series of SI models to be calibrated.

[2] This is more challenging to do when modeling flows between locations abstracted as areal units.

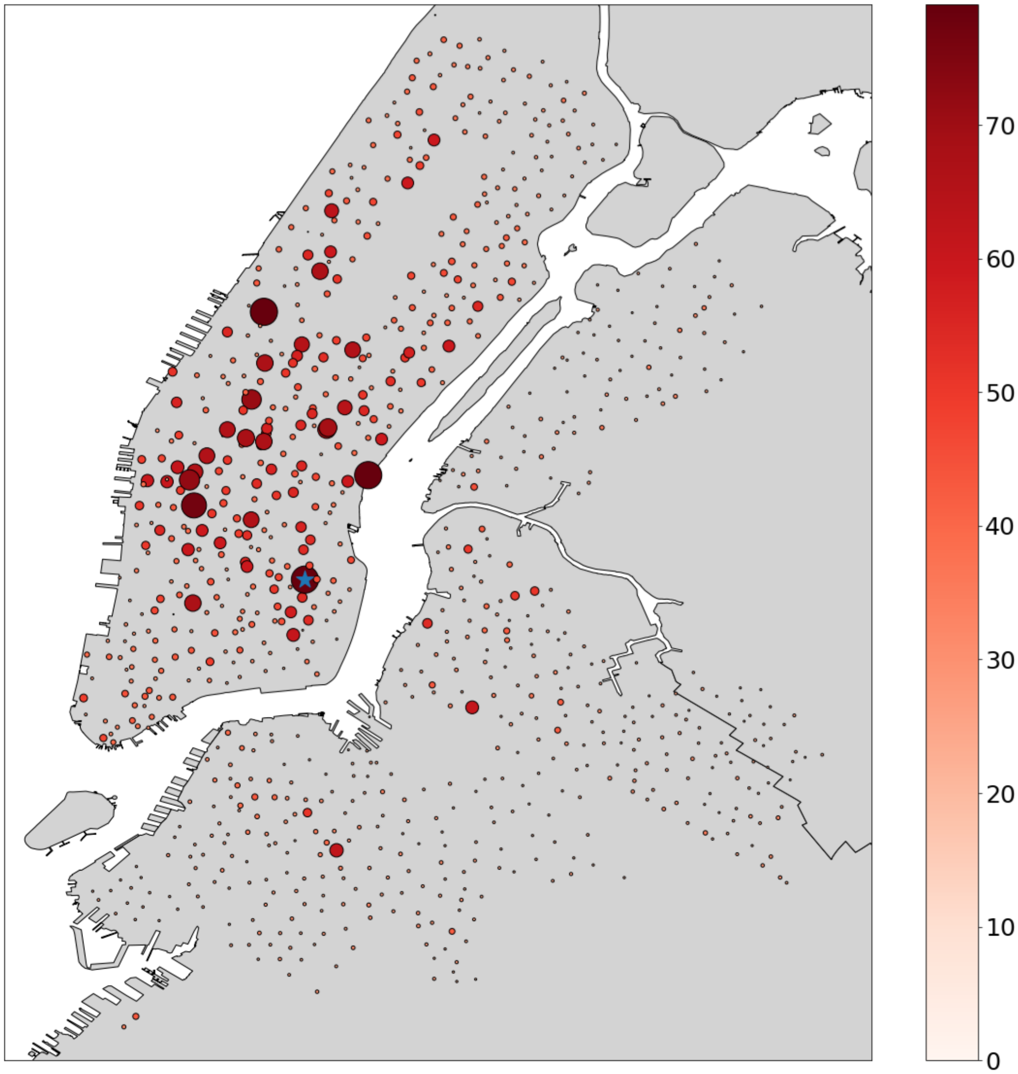

Figure 2: Distribution of bike-share stations in New York City. Larger, darker circles correspond to larger station capacity. The blue star denotes station 445.

Each member of the family of SI models were calibrated on the dataset described above using origin capacity and/or destination capacity as the locational attributes and trip duration as the cost variable. The logic here is that: origins with a larger capacity are likely to generate more trips as they have a larger supply of bikes; destinations with a larger capacity are more likely to attract more riders as they minimize the chance of not having an empty dock to park; and longer duration trips are less likely to occur because they require more effort, riders may have a limited time budget, and the system is priced to discourage excessively long rides. Various preprocessing steps are possible, such as filtering out extremely short or long duration trips or filtering out trips that begin and end at the same location. The need for these steps depends on the dataset being analyzed and the phenomena being modeled and though they may be appropriate in this context, no additional filtering was carried out here for simplicity.

Table 2: Comparison of parameter estimates and model fit from calibrating each member of the family of spatial interaction models. Standard errors are reported inparentheses. UC = unconstrained; PC = production-constrained; AC = attraction-constrained; DC = doubly-constrained.

	UC	PC	AC	DC
Origin capacity	0.78 (0.0022)	-	0.56 (0.0006)	-
Destination capacity	0.76 (0.0022)	0.54 (0.0023)	-	-
Trip duration	-1.25 (0.0011)	-1.40 (0.0011)	-1.41 (0.0011)	-1.53 (0.0012)
Pseudo R^2	0.33	0.44	0.45	0.54

Table 3: Comparison of production-constrained model results for trips only by maleriders or only by female riders. Standard errors are reported in parentheses.

	Male	Female
Destination capacity	0.52 (0.0030)	0.35 (0.0050)
Trip duration	-1.39 (0.0015)	-0.83 (0.0023)
Pseudo R^2	0.45	0.22

The resulting parameter estimates and model fit are reported in table 2 for each member of the family calibrated using Poisson regression and a power decay function. As described in section 2.1, it can be seen in table 2 how the highest model fit is achieved for the doubly-constrained model, followed by the two singly-constrained models, and finally the unconstrained model. However, more parameter estimates are obtained when using the unconstrained model and this is important when the goal is explanation and several location attributes under consideration[3]. For instance, urban environment attributes could be calculated using buffers around stations or socioeconomic attributes could be incorporated based on the municipal unit each station exists within. In this simple example with only a few relatively straightforward variables, the signs of each parameter estimate all match up with their expected values; larger station capacity is associated with larger flows while longer trip durations are associated with small flows. We can extend this analysis by investigating if there are differences in these relationships by gender. A production-constrained model was also calibrated on data that was first filtered by whether not bike share member identified as male or female[4] before aggregating trips into flows. The results in table 3 suggest that trips by those identifying as females are generally not as well explained as trips by those identifying as males by this simple model and that those identifying as females appear to be less constrained by longer trip durations than those identifying as males, after accounting for station capacity. Of course, further analysis is necessary to examine the validity of these results, but this demonstrates how SI models might be deployed to answer existing questions or to generate new hypotheses. A similar comparative strategy can be used to observe how relationships change over time (Batty, 2018; Oshan, 2020).

[3] These attributes need not be the same for both origins and destinations.

[4] There is also a category for trips where gender was not defined but it was not included here for brevity.

Another potential analytical task focuses on predicting changes in flows when a station is added, removed, or some attributes are changed. This entails essentially extrapolating flow values using previously modeled relationships. For example, we can treat station 445 (blue star in figure 2) as if it does not exist and then attempt to predict the flows to and from this station if it were added to the existing system. First, we calibrate a model, here production-constrained, for all of the data excluding flows associated with station 445 (table 3). Then we can use the estimated parameters and equation 3 to compute flows for all data including station 445. The only unknown value in this case is the total outflows O_i for station 445, which could be estimated using an unconstrained model or through other external information. Assuming the true value of O_i is known, the predicted outflows and inflows for station 445 attain Pearson's correlation coefficients with the observed values of 0.66 and 0.80, respectively. However, any inaccuracies in estimating O_i would decrease the accuracy of these predictions. If we instead are interested in predicting flows after an attribute changes, we do not need to estimate a new O_i if we are willing to assume the attribute change may only redistribute trips rather also increasing or decreasing overall trip generation[5].

In order to increase prediction accuracy and more effectively estimate parameters, it may be necessary to account for spatial effects, such as dependence or heterogeneity, using one of the extensions described briefly below but discussed in more detail in (Oshan, 2021).

5. Extensions to capture spatial effects

5.1. Local models and spatial nonstationarity

A popular extension of gravity-type SI models is to calibrate separate parameters for flows disaggregated from each origin to all destinations, which is often called an origin-specific local model (Fotheringham, 1981; Stewart Fotheringham & Brunsdon, 1999). The set of parameter estimates obtained for each origin can then be mapped in order to explore any spatial variation in the relationships generating the flows. Applying a traditional "global" SI model when there are local relationships could lead to misleading interpretations and poor model accuracy (Farmer & Oshan, 2017). It can also be useful for identifying misspecification or anomalous behavior (e.g., Fotheringham, 1983a). One way to achieve localization is to filter the overall dataset into subsets that only include flows from a single origin and then to calibrate an entirely separate model for each subset. Functions to carry out this technique are available within the Python *spint* module. Alternatively, it is possible to specify a single regression model that appropriately categorizes the data based on each origin. This is achieved by introducing interaction terms into the regression so that each explanatory variable interacts with a categorical variable indicating each flow's origin. The main difference between these methods is that the latter assumes a common variance amongst the individual subsets of the data, which can result in slightly different parameter estimates. Another proposed strategy to investigate heterogeneity amongst SI processes is to incorporate the geographically weighted regression methodology (Fotheringham et al., 2002) into SI models (Nakaya, 2001; Nissi & Sarra, 2011; Kalogirou, 2015; Kordi & Fotheringham, 2016; Zhang, Cheng, & Jin, 2019; Zhang, Cheng, Jin, et al., 2019).

5.2. The theory of Competing Destinations

Traditional gravity-type SI models assume that the volume of flows between two locations is independent of the other locations in the system. That is, individuals consider all locations in the system and evaluate them independently when selecting a destination. The competing destinations model (Fotheringham, 1983a, 1983b, 1984, 1986a; Fotheringham & O'Kelly, 1989) considers that in

[5] Code examples of these prediction scenarios are also available in the demonstration notebook.

reality spatial decision-making often arises from a hierarchical two-stage or multi-stage decision-making process. Since individuals may not be able evaluate all alternative locations, it is likely that they instead first select a region or cluster of locations and then subsequently choose an individual destination from within that cluster. The effect is that as the accessibility of a particular destination, j, to all other potential destinations increases, j may experience greater competition from other destinations, and therefore the volume of flows to $j$would be smaller than predicted by a traditional SI model. Practically, this effect is captured by introducing a new variable into SI models to measure destination accessibility, A_{ij} , and can be thought of as the likelihood that other destinations are also considered along with destination j (Fotheringham & O'Kelly, 1989). A failure to account for spatial structure can result in biased parameter estimates. Furthermore, the bias can be categorized in terms of the perception of destination clusters where there are competition effects (i.e., negative exponent on A_{ij}) or agglomeration effects (i.e., positive exponent on A_{ij}). The type of effect that arises is typically dependent upon the type of SI process being modeled; however, competition is observed more often in empirical settings, which may be driven by the behavioral tendency of individuals to underestimate the overall attractiveness of large clusters (Fotheringham et al., 2007).

5.3. An econometric spatial autoregressive specification

Recent work has proposed an extension to the SAR model (introduced in chapter 6) to accommodate flow data (LeSage & Pace, 2008; LeSage & Fischer, 2014; LeSage & Thomas-Agnan, 2015). To include a spatially dependent process in the unconstrained SI model, Lesage & Pace (2008) suggest spatial weight matrices that define neighborhoods from the perspective of origins, destinations, or origin-destination pairs, respectively. This specification is motivated by the theory that the movements that people decide to make are based upon their knowledge of neighboring flows in a previous time period. Including the spatial lag of the dependent variable accounts for endogenous effects that may arise due to changes in shared resources such as transportation infrastructure that can cause reactions to defuse through an entire system, which is also known as global spillovers. It is also suggested to include a spatial lag of the origin and destination explanatory variables within this specification to account for exogenous effects or local spillovers that are not caused by changes in shared resources so that feedback is not expected to propagate beyond immediate neighbors. The spatial lag of explanatory variables is also justified on practical grounds to protect against omitted variable bias (LeSage & Fischer, 2014).

Importantly, Lesage & Fischer (2014) raise an important issue that occurs when locations are both origins and destinations and the same variable is used to represent both of them. In this scenario, it is not possible to interpret how a change in a single origin attribute (destination attribute) would affect flows originating (terminating) from that origin (destination) without also considering how that change would also affect flows that terminate at that origin (originate at that destination). That is, a change in a single locational attribute can cause changes across many flows in the system, which can further diffuse across the system if observations are spatially dependent. Therefore, scalar summary measures that capture these multiple changes, which are known as *effects estimates*, should be used rather than directly interpreting the regression parameter estimates (LeSage & Fischer, 2014; LeSage & Thomas-Agnan, 2015). In the case where there is no spatial dependence, interpretations based on the parameter estimates and the effects estimates should be approximately the same.

5.4. A spatial Eigenvector filtering approach

Eigenvector spatial filtering (ESF) is a technique that accounts for spatial autocorrelation based on the interpretation that the eigenvectors of a projected connectivity matrix are the set of possible orthogonal and uncorrelated map patterns given a particular definition of connectivity (Griffith, 1996, 2011). By selecting a subset of the eigenvectors and creating a linear combination, it is possible to produce a synthetic variable that serves as a proxy for omitted spatially autocorrelated exogenous variables (Griffith, 2004). This specification has been shown to produce results where the error term does not violate independence assumptions (Griffith, 2000) and provides a mechanism to account for positive spatial autocorrelation in discrete probability models, where it remains challenging to directly specify and estimate autoregressive extensions (Griffith, 2002, 2004).

After the ESF framework was established, it was subsequently extended from spatial data aggregated to n areal units to SI flow data that occurs between n^2 pairs of origins and destinations (Griffith, 2007; Chun, 2008; Fischer & Griffith, 2008a; Fischer & Griffith, 2008b; Griffith, 2009). Several variations have been proposed based on whether eigenvectors are introduced to account for aspects of spatial autocorrelation that pertain to origins, destinations, or origin-destination flows. This framework has also been motivated based on links to mixed models that use random effects. Here the selected eigenvectors can be thought of as spatially correlated random effects that account for latent spatial variables. It is demonstrated that when ESFs are used to account for residual spatial variation, that also including also a normally distributed (i.e., non-spatial) random effect becomes approximately equal to using the fixed effects or balancing factors to achieve constrained variants of the family of gravity type models (Griffith & Fischer, 2013).

6. Concluding remarks

The focus of this chapter was threefold. First, it introduced SI models, covering the basic historical, conceptual, and technical ideas. Second, a demonstrative application was presented to highlight some practical aspects of the framework. Third, it reviewed several extensions to account for more complex spatial relationships. The choice of analytical actions and subsequent extension depends upon the application context, motivation of the analytical method, underlying theory, and the nature of the available data. Together, these efforts were meant to paint a broad picture of the SI modeling landscape and offer only an initial entry point into the otherwise vast literature and diverse thinking on the topic.

SI is itself an idea and a method that has interacted between disciplines, diffusing across the academy and developing over time as the nature of the social sciences has evolved. At the same time, new data and tools are enhancing and reinvigorating this classic spatial analysis framework, providing the impetus for a new era of spatial interaction. Some examples include: (1) the availability of more representative samples and time-aware flows (Waddington et al., 2019; Oshan, 2020; Park et al., 2021); (2) the use of machine learning techniques to capture spatial, temporal, and network dependencies (Ren et al., 2019; Yeghikyan et al., 2020; Yang et al., 2020); (3) connecting SI models and agent-based modeling methods (Kowalski, 2019; Makarov, 2019; Ge et al., 2021); and (4) incorporating multiple scales of analysis (Thompson et al., 2019; Zhang et al., 2019; McCulloch et al., 2021). These nascent trends hold promise for building upon the rich traditions of SI modeling presented here by combining data-driven and theory-driven modeling approaches. Defining potential and overcoming pitfalls at this nexus will be an important task in this new era of spatial interaction.

Biography Taylor M. Oshan is an Assistant Professor in the Center for Geospatial Information Science within the Department of Geographical Sciences at the University of Maryland, College Park. His research interests are centered on developing methods to analyze spatial and temporal processes and applying them in the context of urban health and transportation, as well as building open source tools. In particular, his work has focused on spatial interaction models and local multivariate statistical models.

References

Afandizadeh, S., & Hamedani, S. M. Y. (2012). A fuzzy intervening opportunity modelto predict home-based shopping trips. *Canadian Journal of Civil Engineering*, *39*(2), 203–222.

Akwawua, S., & Pooler, J. A. (2001). The development of an intervening opportunities model with spatial dominance effects. *Journal of Geographical Systems*, *3*(1), 69–86.

Anas, A. (1983). Discrete choice theory, information-theory and the multinomiallogit and gravity models. *Transportation Research Part B: Methodological*,*17*(1), 13–23.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, *49*, 45–53.

Batty, M., & Mackie, S. (1972). The Calibration of Gravity, Entropy, and Related Models of Spatial Interaction. *Environment and Planning A*, *4*(2), 205–233.

Batty, Michael. (2013). *The New Science of Cities*. The MIT Press.

Batty, M. (2018). Visualizing aggregate movement in cities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1753), 20170236.

Brunsdon, C. (2018). *Chrisbrunsdon/simR* [R]. https://github.com/chrisbrunsdon/simR (Original work published 2018)

Burger, M., Van Oort, F., & Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, *4*(2), 167–190.

Carey, H. C. (1858). *Principles of Social Science*. J. B. Lippincott & co.

Carrothers, G. A. P. (1956). An Historical Review of the Gravity and Potential Concepts of Human Interaction. *Journal of the American Institute of Planners*, *22*(2), 94–102.

Cascetta, E., Pagliara, F., & Papola, A. (2007). Alternative approaches to trip distribution modelling: A retrospective review and suggestions for combining different approaches. *Papers in Regional Science*, *86*(4), 597–620.

Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, *10*(4), 317–344.

Crymble, A. (2019). Introduction to Gravity Models of Migration & Trade.

Programming Historian. https://programminghistorian.org/en/lessons/gravity-model

Dennett, A. (2012). *Estimating flows between geographical locations:'get me started in'spatial interaction modelling* (Working Paper No. 184; CASA Working Paper Series). Citeseer.

Dennett, A. (2018). *Modelling population flows using spatial interaction models*. http://minerva-access.unimelb.edu.au/handle/11343/233564

Dison, D., W., & Hale, C., W. (1977). Gravity versus Intervening Opportunity Models in Explanation of Spatial Trade Flows. *Growth and Change*, *8*(4), 15–22.

Dodd, S. C. (1950). The Interactance Hypothesis: A Gravity Model Fitting Physical Masses and Human Groups. *American Sociological Review*, *15*(2), 245–256.

Elffers, H., Reynald, D., Averdijk, M., Bernasco, W., & Block, R. (2008). Modelling Crime Flow between Neighbourhoods in Terms of Distance and of Intervening Opportunities. *Crime Prevention and Community Safety: An International Journal*, *10*(2), 85–96.

Farmer, C., & Oshan, T. (2017). Spatial Interaction. *The Geographic Information Science & Technology Body of Knowledge*, *(4th Quarter 2017 Edition)*.

Fischer, M., & Griffith, D. A. (2008a). Modeling spatial autocorrelation in spatial interaction data: A comparison of spatial econometric and spatial filtering specifications. *J Reg Sci*, *48*(5), 969–989.

Fischer, M. M., & Griffith, D. A. (2008b). Modeling Spatial Autocorrelation in Spatial Interaction Data: An Application to Patent Citation Data in the European Union*. *Journal of Regional Science*, *48*(5), 969–989.

Flowerdew, R., & Aitkin, M. (1982). A Method of Fitting the Gravity Model Based on the Poisson Distribution. *Journal of Regional Science*, *22*(2), 191–202.

Flowerdew, R., & Lovett, A. (1988). Fitting Constrained Poisson Regression Models to Interurban Migration Flows. *Geographical Analysis*, *20*(4), 297–307.

Fotheringham, A. Stewart. (1981). Spatial Structure and Distance-Decay Parameters. *Annals of the Association of American Geographers*, *71*(3), 425–436.

Fotheringham, A. S. (1983a). A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning A*, *15*(1), 15–36.

Fotheringham, A. S. (1983b). Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning A*, *15*(8), 1121–1132.

Fotheringham, A. S. (1984). Spatial flows and spatial patterns. *Environment and Planning A*, *16*(4), 529–543.

Fotheringham, A. S. (1986a). Further Discussion on distance-deterrence parameters and the competing destinations model. *Environment and Planning A*, *18*(4), 553–556.

Fotheringham, A. S. (1986b). Modelling hierarchical destination choice. *Environmentand Planning A*, *18*(3), 401–418.

Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial Interaction Models:Formulations and Applications*. Kluwer Academic Publishers.

Fotheringham, A. Stewart, & Brunsdon, C. (1999). Local Forms of Spatial Analysis. *Geographical Analysis*, *31*(4), 340–358.

Fotheringham, A. Stewart, Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Fotheringham, A. S., Brundson, C., & Charlton, M. (2007). Spatial Modelling and the Evolution of Spatial Theory. In *Quantitative Geography*. SAGE Publications Ltd.

Fotheringham, A. S. (2017). Spatial Interaction. In *International Encyclopedia of Geography* (pp. 1–9). American Cancer Society.

Ge, J., Polhill, J. G., Macdiarmid, J. I., Fitton, N., Smith, P., Clark, H., Dawson, T., & Aphale, M. (2021). Food and nutrition security under global trade: A relation-driven agent-based global trade model. *Royal Society Open Science*, *8*(1), 201587.

Giraud, T. (2021). *Riatelab/SpatialPosition* [R]. riatelab. https://github.com/riatelab/SpatialPosition (Original work published 2015)

Graser, A. (2019). MovingPandas: Efficient Structures for Movement Data in Python. *GI_Forum*, *1*, 54–68. https://doi.org/10.1553/giscience2019_01_s54

Griffith, D. A. (1996). SPATIAL AUTOCORRELATION and EIGENFUNCTIONS OF THE GEOGRAPHIC WEIGHTS MATRIX ACCOMPANYING GEO-REFERENCED DATA. *Canadian Geographer / Le Géographe Canadien*, *40*(4), 351–367.

Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, *2*(2), 141–156.

Griffith, D. A. (2002). A spatial filtering specification for the auto-Poisson model. *Statistics & Probability Letters*, *58*(3), 245–251.

Griffith, D. A. (2004). A spatial filtering specification for the autologistic model. *Environment and Planning A*, *36*(10), 1791 – 1811.

Griffith, D. A. (2007). Spatial structure and spatial interaction: 25 years later. *The Review of Regional Studies*, *37*(1), 28.

Griffith, D. A. (2009). Modeling spatial autocorrelation in spatial interaction data: Empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical Systems*, *11*(2), 117–140.

Griffith, D. A. (2011). Visualizing analytical spatial autocorrelation components latent in spatial interaction data: An eigenvector spatial filter approach. *Computers, Environment and Urban Systems*, *35*(2), 140–149.

Griffith, D. A., & Fischer, M. M. (2013). Constrained variants of the gravity model and spatial dependence: Model specification and estimation issues. *Journal of Geographical Systems*, *15*(3), 291–317.

Haynes, K. E., & Fotheringham, A. S. (1984). *Gravity and spatial interaction models* (Vol. 2). Sage publications.

Haynes, K. E., Poston, D. L., & Schnirring, P. (1973). Intermetropolitan Migration in High and Low Opportunity Areas: Indirect Tests of the Distance and Intervening Opportunities Hypotheses. *Economic Geography*, *49*(1), 68.

Kalogirou, S. (2015). Destination Choice of Athenians: An Application of Geographically Weighted Versions of Standard and Zero Inflated Poisson Spatial Interaction Models: Destination Choice of Athenians. *Geographical Analysis*.

Kaltenbach, K., D. (1972). *Application of Gravity and Intervening Opportunities Modelsto Recreational Travel in Kentucky* (No. 336). Department of Highways Division of Research.

Kordi, M., & Fotheringham, A. S. (2016). Spatially Weighted Interaction Models (SWIM). *Annals of the American Association of Geographers*, *106*(5), 990–1012.

Kowalski, L. (2019). Comparing Spatial-Interaction and Hybrid Agent-Based Modelling Approaches: An Application to Location Analysis of Services. *Journal of Artificial Societies and Social Simulation*, *22*(1), 1.

LeSage, J. P., & Fischer, M. M. (2014). Spatial regression-based model specifications for exogenous and endogenous spatial interaction. *Available at SSRN 2420746*. http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2420746

LeSage, J. P., & Pace, R. K. (2008). Spatial Econometric Modeling Of Origin-Destination Flows. *Journal of Regional Science*, *48*(5), 941–967.

LeSage, J. P., & Thomas-Agnan, C. (2015). Interpreting Spatial Econometric Origin-Destination Flow Models. *Journal of Regional Science*, *55*(2), 188–208.

Makarov, V., Bakhtizin, A., Beklaryan, G., Akopov, A., Rovenskaya, E., & Strelkovskii, N. (2019). Aggregated Agent-Based Simulation Model of Migration Flows of the European Union Countries. *Ekonomika i Matematicheskie Metody*, *55*(1),3–15.

Manley, E., & Dennett, A. (2019). New Forms of Data for Understanding Urban Activity in Developing Countries. *Applied Spatial Analysis and Policy*, *12*(1),45–70.

Martínez, L. M., & Viegas, J. M. (2013). A new approach to modelling distance-decay functions for accessibility assessment in transport studies. *Journal of Transport Geography*, *26*, 87–96.

McCulloch, K., Golding, N., McVernon, J., Goodwin, S., & Tomko, M. (2021). Ensemble model for estimating continental-scale patterns of human movement: A case study of Australia. *Scientific Reports*, *11*(1), 4806.

Metulini, R., Patuelli, R., & Griffith, D. (2018). A Spatial-Filtering Zero-Inflated Approach to the Estimation of the Gravity Model of Trade. *Econometrics*, *6*(1), 9.

Nakaya, T. (2001). Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal*, *53*(4), 347–358.

Nazem, M., Trépanier, M., & Morency, C. (2015). Revisiting the destination ranking procedure in development of an Intervening Opportunities Model for public transit trip distribution. *Journal of Geographical Systems*, *17*(1), 61–81.

Nissi, E., & Sarra, A. (2011). Detecting Local Variations in Spatial Interaction Models by Means of Geographically Weighted Regression. *Journal of Applied Sciences*, *11*(4), 630–638.

Oshan, T. M. (2016). A primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library (PySAL). *REGION*, *3*(2), 11.

Oshan, T. M. (2020). Potential and Pitfalls of Big Transport Data for Spatial Interaction Models of Urban Mobility. *The Professional Geographer*, *72*(4), 468–480.

Oshan, T. M. (2021). The spatial structure debate in spatial interaction modeling: 50 years on. *Progress in Human Geography*, 0309132520968134.

Pappalardo, L., Simini, F., Barlacchi, G., & Pellungrini, R. (2019). scikit-mobility: A Python library for the analysis, generation and risk assessment of mobility data. *ArXiv:1907.07062 [Physics]*. http://arxiv.org/abs/1907.07062

Park, S., Oshan, T. M., El Ali, A., & Finamore, A. (2021). Are we breaking bubbles as we move? Using a large sample to explore the relationship between urban mobility and segregation. *Computers, Environment and Urban Systems*, *86*, 101585.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, 167–235.

Reilly, W. J. (1929). *Methods For The Study of Retail Relationships*. University of Texas, Austin.

Ren, Y., Chen, H., Han, Y., Cheng, T., Zhang, Y., & Chen, G. (2019). A hybrid integrated deep learning model for the prediction of citywide spatio-temporal flow volumes. *International Journal of Geographical Information Science*, 1–22.

Rogerson, P. A. (1986). Parameter estimation in the intervening opportunities model. *Geographical Analysis*, *18*(4), 357–360.

Roy, J. R. (2004). *Spatial Interaction Modelling: A Regional Science Context*. Springer Berlin Heidelberg.

Sen, A., & Smith, T. (1995). *Gravity Models of Spatial Interaction Behavior*. Springer.

Schmitt, R. R., & Greene, D. L. (1978). An alternative derivation of the intervening opportunities model. *Geographical Analysis*, *10*(1), 73–77.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, *88*(4), 641–658.

Siah, J. (2018). *RPubs—Lab 10: Spatial Interaction Modeling in R*. https://rpubs.com/Striker/makingmaps_lab10

Siła-Nowicka, K., & Fotheringham, A. S. (2019). Calibrating spatial interaction models from GPS tracking data: An example of retail behaviour. *Computers, Environment and Urban Systems*, *74*, 136–150.

Siła-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., & Fotheringham, A. S. (2016). Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, *30*(5), 881–906.

Smith, S., L. J. (1980). Intervening opportunities and travel to urban recreational centers. *Journal of Leisure Research*, *12*(4), 296–308.

Stewart, J. Q. (1941). An Inverse Distance Variation for Certain Social Influences. *Science*, *93*(2404), 89–90.

Stillwell, J., Daras, K., Bell, M., & Lomax, N. (2014). The IMAGE Studio: A Tool for Internal Migration Analysis and Modelling. *Applied Spatial Analysis and Policy*,*7*(1), 5–23.

Stouffer, S. A. (1940). Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review*, *5*(6), 845–867.

Stouffer, S., A. (1960). Intervening opportunities and competing migrants. *The Journal of Regional Science*, *2*(1), 1–26.

Thompson, C. A., Saxberg, K., Lega, J., Tong, D., & Brown, H. E. (2019). A cumulative gravity model for inter-urban spatial interaction at different scales. *Journal of Transport Geography*, *79*, 102461.

Tiefelsdorf, M, & Boots, B. (1995). The specification of constrained interaction models using the SPSS loglinear procedure. *Geographical Systems*, *2*, 21–38.

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. Twentieth Century Fund.

Ulyssea-Neto, I. (1993). The development of a new gravity-opportunity model fortrip distribution. *Environment and Planning A*, *25*, 817–826.

Vries, J. J. D., Nijkamp, P., & Rietveld, P. (2009). Exponential or power distance-decay for commuting? An alternative specification. *Environment and Planning A*, *41*(2), 461 – 480.

Waddington, T., Clarke, G., Clarke, M. C., Hood, N., & Newing, A. (2019). Accounting for Temporal Demand Variations in Retail Location Models. *Geographical Analysis*, *51*(4), 426–447.

Wills, M. J. (1986). A flexible gravity-opportunity model for trip distribution. *Transportation Research*, *20B*(2), 89–111.

Wilson, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, *1*, 253–269.

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A*, *3*, 1–32.

Wilson, A. G., & Senior, M. L. (1974). Some relationships between entropy maximizing models, mathematical programming models, and their duals. *Journal of Regional Science*, *14*(2), 207–215.

Wölwer, A.-L., Burgard, J., Kunst, J., & Vargas, M. (2018). Gravity: Estimation Methodsfor Gravity Models in R. *Journal of Open Source Software*, *3*(31), 1038.

Yang, Y., Heppenstall, A., Turner, A., & Comber, A. (2020). Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems. *Computers, Environment and Urban Systems*, *83*, 101521.

Yeghikyan, G., Opolka, F. L., Nanni, M., Lepri, B., & Lio, P. (2020). *Learning Mobility Flows from Urban Features with Spatial Interaction Models and Neural Networks**To appear in the Proceedings of 2020 IEEE International Conference on Smart Computing (SMARTCOMP 2020)*. 57–64.

Zhang, L., Cheng, J., & Jin, C. (2019). Spatial Interaction Modeling of OD Flow Data: Comparing Geographically Weighted Negative Binomial Regression (GWNBR) and OLS (GWOLSR). *ISPRS International Journal of Geo-Information*, *8*(5), 220.

Zhang, L., Cheng, J., Jin, C., & Zhou, H. (2019). A Multiscale Flow-Focused Geographically Weighted Regression Modelling Approach and Its Applicationfor Transport Flows on Expressways. *Applied Sciences*, *9*(21), 4673.

Zipf, G. K. (1946). The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, *11*(6), 677–686.