REGION
The Journal of ERSA
Powered by WU

Volume 1, Number 1, 2020, 999–999
DOI: 10.18335/region.v??i??.???
journal homepage: region.ersa.org
ISSN: 2409-5370

A practical guide to spatial interaction modelling

Blinded[1]

1

Received: /Accepted:

Abstract. This document is only a demo explaining how to use the template.

1 Introduction

Spatial interaction models (SIMs) is a core tool to simulate flows between different locations in physical space. They are a valuable resource through which the geographic structure between locations encoded in aggregate flows of people, information and goods can be represented and understood. Intuitively, SIMs seek to capture the spatial interaction between places as a function of three components: origin attributes, destination attributes and their separation. Inspired by Newtonian concepts developed in physics, spatial flows between locations are conceived as the result of their proportional gravitational force and inverse association with spatial separation. Attributes of origin and destination locations are employed to represent gravitational forces pushing and pulling people, information and goods between specific locations. Various forms of distance and costs are used to represent the deterring effects of geographical separation on spatial flows.

SIMs are widely used for prediction and inference. They are used to make inference about the factors contributing to influence spatial flows. They have been used to understand the magnitude and direction of influence of individual and place-level attributes on geographic flows. Understanding the effect of these factors offers valuable evidence to inform the development of appropriate plans, strategies and interventions (Fotheringham, O'Kelly 1989). SIMs are also used to make predictions of the size of spatial flows. These predictions are normally used to assess the impact of interventions and creation of "what-if" scenarios, providing guidance for the identification of optimal locations and size for potential new service units [REF]. In this context, SIMs are often used to evaluate the impact of new bus stops, shopping stores, schools or housing units on their potential demand and traffic changes (Fotheringham, O'Kelly 1989). To these ends, SIMs have been used to address questions in a variety of settings, including retail, migration, transport, trade, commuting, school travel and more broadly urban planning.

Yet, the implementation of SIMs remains a challenge. Algorithms to calibrate the parameters of SIMs have remained locked away, either behind dense algebraic notation in dusty papers from the 1970s, or behind paywalls of commercial software (Rowe et al. 2024). Additionally, Rowe et al. (2024) noted a dearth of knowledge within geographical education as SIMs are not widely taught in undergraduate programmes in the same way as, for instance, regression models are taught in economics or social psychology. This situation is argued to have occurred despite the availability of effective routines to calibrate SIMs via popular linear and general linear modelling frameworks, and as practical expediency is sacrificed at the expense of theoretical or technical prowess (Rowe et al. 2024). The ways in which calibration procedures are presented as lengthy mathematical derivation or passing reference to ordinary least square tend to hamper accessibility for the easy implementation of SIMs.

This computational notebook contributes to redressing these issues. It aims to provide an intuitive, understandable and practical guide to estimate SIMs in a variety of modelling frameworks. It will include the necessary code to calibrate SIMs, using origin-destination travel-to-work data for the United Kingdom in R programming language. The code provided is generisable and can be adapted to different origin-destination flow data and contexts, including migration, student, transport, trade, currency, data transfer, vessel, shipment and freight flows.

The notebook is structured as follows. The next section sets out some fundamental concepts and definitions relating to SIMs. Section 3 identifies the libraries used before Section 4 describes the data. Section 5 illustrates key techniques to visualise complex spatial interaction data, and Section 6 shows and explains how to estimate SIMs using a range of modelling frameworks. It start with traditional mathematical and Ordinary Least Squares (OLS) approaches to more advanced statistical frameworks, such as Generalised Linear Mixed Models (GLMMs) and machine learning algorithms.

2 Context

SIMs take various forms. Newtonian gravity models are probably the most widely known and used form of SIMs. Inspired by Newton's law of gravity, the basic gravity version of these models assumes that the spatial flows or interactions between an origin (i) and a destination j is proportional to their masses (M_i and M_j) and inversely proportional to their separation (D_{ij}). Locations are expected to interact in a positively reinforcing manner that is multiplicative of their masses, but to diminish with the intervening role of their separation. The separation between locations is often represented by a distance decay function and is measured in terms of the distance, cost or time involved in the interaction. Generally, the model includes a constant (G) ensuring that the expected flows do not exceed their respective observed counts, and a parameter (k) representing the deterring effect of geographical separation. The task is to estimate these parameters. Formally a gravity model can be expressed as:

$$T_{ij} = G\frac{M_i M_j}{D_{ij}^k}$$

SIMs have three key inputs: (1) a matrix of flows between a set of origins and destinations; (2) a measure of separation between origins and destinations; and, (3) measures of masses at origin and destination locations. The literature usually considers a family of SIMs taking four forms which refers to various constraints placed on parameters of the model (Wilson 1971). There is an *unconstrained* version which is actually constrained to ensure that the total sum of the predicted flows from a gravity model be equal the total sum of the observed flows across all origins and destinations. Constrained versions are used to ensure that specific origin or destination observations are met. Three general formulations of constrained models are used: *production-constrained*, *attraction-constrained* and *doubly-constrained* models. *Production-constrained* versions are used to constrain a model so that the predicted flows emanating from individual origins is equal to their respective observed numbers. *Attraction-constrained* versions do the same but at individual destinations. *Doubly-constrained* versions combine these two sets of constraints to ensure predicted flows are equal to observed flows at both origins and destinations.

3 Computation environment

4 Data

Commuting or migration?

5 Visualising spatial interaction data

6 Estimating spatial interaction models

6.1 Mathemathical gravity models

6.2 Statistical gravity models

6.3 Extensions

6.3.1 Generalised linear mixed gravity models

6.3.2 Machine learning gravity models

References

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

Rowe F, Lovelace R, Dennett A (2024, 05) *Spatial interaction modelling: a manifesto*, 177–196. Edward Elgar Publishing

Wilson AG (1971, 03) A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space* 3: 1–32. CrossRef