

Research



Cite this article: Ellam L, Girolami M, Pavliotis GA, Wilson A. 2018 Stochastic modelling of urban structure. *Proc. R. Soc. A* **474**: 20170700.
<http://dx.doi.org/10.1098/rspa.2017.0700>

Received: 2 October 2017

Accepted: 11 April 2018

Subject Areas:

statistics, complexity

Keywords:

urban modelling, urban structure, Bayesian inference, Bayesian statistics, Markov chain Monte Carlo, complexity

Author for correspondence:

L. Ellam

e-mail: l.ellam16@imperial.ac.uk

L. Ellam^{1,2}, M. Girolami^{1,2}, G. A. Pavliotis¹ and A. Wilson^{2,3}

¹Department of Mathematics, Imperial College London, London SW7 2AZ, UK

²The Alan Turing Institute, The British Library, London NW1 2DB, UK

³Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, UK

LE, 0000-0002-1542-811X

The building of mathematical and computer models of cities has a long history. The core elements are models of flows (spatial interaction) and the dynamics of structural evolution. In this article, we develop a stochastic model of urban structure to formally account for uncertainty arising from less predictable events. Standard practice has been to calibrate the spatial interaction models independently and to explore the dynamics through simulation. We present two significant results that will be transformative for both elements. First, we represent the structural variables through a single potential function and develop stochastic differential equations to model the evolution. Second, we show that the parameters of the spatial interaction model can be estimated from the structure alone, independently of flow data, using the Bayesian inferential framework. The posterior distribution is doubly intractable and poses significant computational challenges that we overcome using Markov chain Monte Carlo methods. We demonstrate our methodology with a case study on the London, UK, retail system.

1. Introduction

The task of understanding the inner workings of cities and regions is a major challenge for contemporary science. The key features of cities and regions are activities at locations, flows between locations and the structure that facilitates these activities [1]. It is well understood that cities and regions are complex systems,

© 2018 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

and that an emergent structure arises from the actions of many interacting individuals. The flows between locations arise from the choices of individuals. An understanding of the underlying choice mechanism is therefore advantageous for planning and decision-making. Economists have long supported the idea that consumer choices are derived from utility, a measure of net benefit, although preferences can only be measured indirectly by the phenomena they give rise to [2].

Random utility models, such as the multinomial logit model [3], provide a discrete choice mechanism based on a utility function. These models have received considerable attention in the econometrics literature [4]. The more conventional random utility models assume that choices are conditionally independent and require large volumes of flow data to calibrate. It is generally difficult to ascertain the flow data for a large number of individuals residing in a country or city, and this may require an extensive survey that suffers from sampling biases. On the other hand, the structure facilitating activities can be more straight-forward to measure.

It turns out that the flows between locations concern a vast number of individuals and are well represented by statistical averaging procedures [5]. It also turns out that the evolution of urban structure can be described by a system of coupled first-order ordinary differential equations that are related to the competitive Lotka–Volterra models in ecology [6]. The conventional Harris and Wilson model in [6] is obtained by combining Lotka–Volterra models with statistical averaging procedures, after having expressed the flows in terms of the evolving structure and spatial interaction. As it tends to be more feasible to observe the emergent structure, for example configurations of floorspace dedicated to retail activity, our work is largely motivated by the existing models of urban structure [1,6–9]. By adopting a similar approach, we view the flows between locations as ‘missing data’.

We note, however, that there is an urgent need to provide an improved modelling capability that captures the stochastic nature and uncertainty associated with the evolution of urban structure. The key shortcoming of the Harris and Wilson model is that it is deterministic and converges to one of multiple equilibria as determined by the initial conditions. In reality, the behaviour provided by the Harris and Wilson model would be accompanied by fluctuations arising from less predictable events. We instead introduce mathematically well-posed systems of stochastic differential equations (SDEs) to address this shortcoming, and provide an associated Bayesian inference methodology for parameter estimation and model calibration.

To this end, we take a novel approach and construct a probability distribution to represent the uncertainty in equilibrium structures for urban and regional systems. The probability distribution is a Boltzmann–Gibbs measure that is the invariant distribution of a related SDE model [10], and is defined in terms of a potential function whose gradient describes how we expect urban structure to evolve forward in time. The potential function may be interpreted as constraints on consumer welfare and running costs from a maximum entropy argument [8,11]. For the purposes of parameter estimation, the Boltzmann–Gibbs measure forms an integral part of the assumed data-generating process in a Bayesian model of urban structure [12,13]. A computational statistical challenge arises as there is an intractable term in the density of the Boltzmann–Gibbs measure that is parameter dependent. The intractable term must be taken into consideration when using Markov chain Monte Carlo (MCMC) to explore the probability distributions of interest [14,15]. Our approach is applicable to a wide range of applications in urban and regional modelling; we demonstrate our approach by inferring the full distribution over the model parameters and latent structure for the London, UK, retail system.

2. Modelling urban systems

In this section, we construct a probability distribution for urban and regional systems. We work in the setting of the Harris and Wilson model [6] and use consumer behaviour as an archetype; however, the methodology is general and has wider applications such as archaeology, logistics, healthcare and crime to name a few [1]. We are interested in the sizes of M destination zones where consumer-led activities take place, for example shopping. Similarly, there are N origin zones from where consumers create demands for each of the destination zones. We define urban structure

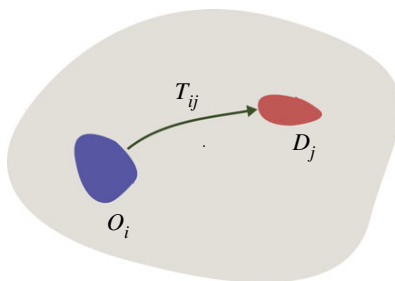


Figure 1. Illustration of a flow in an urban or regional system. It is assumed that there are N origin zones (e.g. left) and M destination zones (e.g. right). The flow T_{ij} denotes the flow of quantities from origin zone i to destination zone j . In an urban or regional system, there are NM flows similar to the one depicted.

as the vector of sizes $W = \{W_1, \dots, W_M\} \in \mathbb{R}_{>0}^M$. In what proceeds it is more natural to work in terms of log-sizes $X = \{X_1, \dots, X_M\} \in \mathbb{R}^M$, where each $W_j = \exp(X_j)$. We refer to log-size as the attractiveness, which is an unscaled measure of benefit, and by working in terms of attractiveness we avoid positivity issues when developing a stochastic model. We first describe a stochastic generalization of the Harris and Wilson model and then consider the equilibrium distribution as a probability distribution of urban structure.

(a) A stochastic reformulation of the Harris and Wilson model

The flow between destination zone j and origin zone i is denoted T_{ij} . We illustrate a component of an urban or regional system in figure 1. For a singly constrained urban system, the demands made by the N origin zones are

$$O_i = \sum_{j=1}^M T_{ij}, \quad i = 1, \dots, N, \quad (2.1)$$

and are known. The demands made for the M destination zones are

$$D_j = \sum_{i=1}^N T_{ij}, \quad j = 1, \dots, M, \quad (2.2)$$

and are to be determined. The demands for the destination zones depend on urban or regional structure. It is assumed that larger zones provide more benefits for their use and that local zones are more convenient and cost less to use. A suitable model of the flows is obtained by maximizing an entropy function subject to the constraint in (2.1) in addition to fixed benefit and cost constraints [5,8]. The resulting flows are

$$T_{ij} = O_i \frac{W_j^\alpha \exp(-\beta c_{ij})}{\sum_{k=1}^M W_k^\alpha \exp(-\beta c_{ik})}, \quad (2.3)$$

where $\alpha, \beta > 0$ are scaling parameters and each $c_{ij} \geq 0$ represents the cost or inconvenience of carrying out an activity at zone j from i .

We expect that zones with unfulfilled demand will grow, whereas zones that do not fulfil their capacity will reduce to a more sustainable size. It is therefore reasonable to expect a degree of stability in the sizes of the destination zones. A suitable model of the dynamics is given by the Harris and Wilson model [6], which is described by a system of ordinary differential equations (ODEs)¹

$$\frac{dW_j}{dt} = \epsilon W_j (D_j - \kappa W_j), \quad W(0) = w_0, \quad (2.4)$$

¹For simplicity, we are not explicit about time dependence when describing differential equations.

where $\epsilon > 0$ is the responsiveness parameter and $\kappa > 0$ is the cost per unit floor size. The assumption that zones aim to maximize their size until an equilibrium is reached is justified by including the cost of capital in the running costs. A natural generalization of the Harris and Wilson model is the following SDE with multiplicative noise that we interpret in the Stratonovich² sense:

$$dW_j = \epsilon W_j (D_j - \kappa W_j) dt + \sigma W_j \circ dB_j, \quad W(0) = w_0, \quad (2.5)$$

for a standard M -dimensional Brownian motion B and volatility parameter $\sigma > 0$. A heuristic interpretation of the SDE is that, over a short time δt , the net capacity term ' $D_j - \kappa W_j$ ' in (2.4) is randomly perturbed by centred Gaussian noise with a standard deviation of $\sigma \sqrt{\delta t}$. The noise term represents fluctuations in the growth rates arising from less predictable events that are not captured by the original model. The specification of multiplicative noise preserves the positivity of each W_j .

With the change of variables $X_j = \ln W_j$, the Harris and Wilson model in (2.4) can be expressed as a gradient flow. The corresponding stochastic dynamics in (2.5) is an overdamped Langevin diffusion. To express this notion, we introduce a potential function $V: \mathbb{R}^M \rightarrow \mathbb{R}$, its gradient $\nabla V: \mathbb{R}^M \rightarrow \mathbb{R}^M$ and an 'inverse-temperature' parameter $\gamma = 2\sigma^{-2}$ and reformulate the stochastic dynamics as

$$dX = -\nabla V(X) dt + \sqrt{2\gamma^{-1}} dB, \quad X(0) = x_0, \quad (2.6)$$

where the potential function is

$$\epsilon^{-1} V(x) = -\alpha^{-1} \sum_{i=1}^N O_i \ln \sum_{j=1}^M \exp(\alpha x_j - \beta c_{ij}) + \kappa \sum_{j=1}^M \exp(x_j). \quad (2.7)$$

It is well understood that the density of $X(t)$, denoted $\rho(x, t)$, evolves in time according to the Fokker–Planck equation [10]. For the SDE in (2.6), the Fokker–Planck equation can be written as

$$\frac{d\rho(x, t)}{dt} = \nabla \cdot (\rho(x, t) \nabla V(x)) + \gamma^{-1} \Delta \rho(x, t), \quad \rho(x, 0) = \delta(x - x_0). \quad (2.8)$$

While (2.8) is very challenging to solve, especially in higher dimensions, its steady-state solution is available in closed form and is the density of a Boltzmann–Gibbs measure given by

$$\rho_\infty(x) = \frac{1}{Z} \exp(-\gamma V(x)), \quad Z := \int_{\mathbb{R}^M} \exp(-\gamma V(x)) dx. \quad (2.9)$$

The Boltzmann–Gibbs measure described by (2.9) forms the basis of our stochastic model of urban structure. The potential function given by (2.7) does not yield a well-defined probability distribution as the normalizing constant in (2.9) is not finite. In order to address the issue, we could restrict the dynamics to a bounded subset of \mathbb{R}^M , or introduce a confining term in the potential function. We adopt the latter approach and later argue that this approach amounts to an economically meaningful constraint.

(b) Boltzmann–Gibbs measures for urban structure

We model urban and regional structure as a single realization of the Boltzmann–Gibbs measure described by (2.9). The Boltzmann–Gibbs measure is the stationary distribution of the overdamped Langevin dynamics considered; however, we acknowledge that there are other stochastic processes that have the same stationary distribution [16]. It is desirable that the potential function satisfies the assumptions in appendix A. It suffices to say here that smooth potential functions that grow at least linearly but no faster than exponentially at infinity have the desired mathematical properties.

The Boltzmann–Gibbs measure can also be obtained from a maximum entropy argument [8, 11]. The advantage of this view is that the terms in the potential function can be interpreted

²The Stratonovich interpretation is obtained from a smooth approximation of white noise, which is appropriate for the dynamical system of interest. For further discussion, refer to [10].

as economic constraints. We consider a potential function with three components to develop a baseline model, although more comprehensive presentations are possible³

$$\epsilon^{-1}V(x) = V_{\text{Utility}}(x) + \kappa V_{\text{Cost}}(x) + \delta V_{\text{Additional}}(x), \quad (2.10)$$

where κ is as before and $\delta > 0$ is an additional parameter. The utility potential describes consumer welfare arising from utility-based choices; the cost potential enforces capacity limits in the system; and the additional potential is a confining term that represents government initiatives, continued investment or a background level of demand. If we consider a random variable $X \in \mathbb{R}^M$ that is subject to the following equality constraints:

$$\left. \begin{aligned} \mathbb{E}[V_{\text{Utility}}(X)] &= C_{\text{Utility}}, \\ \mathbb{E}[V_{\text{Cost}}(X)] &= C_{\text{Cost}} \\ \mathbb{E}[V_{\text{Additional}}(X)] &= C_{\text{Additional}}, \end{aligned} \right\} \quad (2.11)$$

and

with each $C_i \in \mathbb{R}$, then the maximum entropy distribution of X can be written as the Boltzmann–Gibbs measure whose density is given by (2.9) with reference to the potential function in (2.10). We now consider the meaning of each of these constraints in turn.

(i) Utility potential

A natural candidate for a utility potential is a measure of consumer welfare. For example, welfare may be taken to be the area under the demand curve [17], given by (2.2), that is equal to the path integral

$$\begin{aligned} V_{\text{Utility}}(x) &:= \int_{x_0}^x (D_1(x'), \dots, D_M(x')) \cdot dx' \\ &= \alpha^{-1} \sum_{i=1}^N O_i \ln \sum_{j=1}^M \exp(U_{ij}(x_j)) + \text{const.}, \end{aligned} \quad (2.12)$$

where we have defined the deterministic utility function

$$U_{ij}(x_j) = \alpha x_j - \beta c_{ij}. \quad (2.13)$$

In appendix B, we show that (2.13), but with α dependent on i , is also obtained by seeking a utility function consistent with a singly constrained model and the path integral in (2.12). The log-sum function is commonly used as a welfare measure in the economics literature [17–19]. To make the connection with random utility maximization explicit, we define the stochastic utility function for a choice being made from origin zone i as

$$\tilde{U}_{ij}(x_j) = U_{ij}(x_j) + \xi_{ij}, \quad (2.14)$$

where the ξ_{ij} are independent and identically distributed Gumbel random variables. Then under the utility maximization framework, the expected utility attained from a unit flow leaving origin zone i is

$$\mathbb{E} \left[\max_{1 \leq j \leq M} \tilde{U}_{ij}(x_j) \right] = \ln \sum_{j=1}^M \exp(U_{ij}(x_j)) + c, \quad (2.15)$$

where c is the Euler–Mascheroni constant [17]. The utility potential may then be expressed as the expected utility attained from all flows in units of α

$$V_{\text{Utility}}(x) = \alpha^{-1} \sum_{i=1}^N O_i \mathbb{E} \left[\max_{1 \leq j \leq M} \tilde{U}_{ij}(x_j) \right] + \text{const.} \quad (2.16)$$

³We present a simple aggregated model to illustrate the ideas. A refined model with disaggregation does not change the underlying arguments.

Two remarks are in order. First, the scaling factor of α^{-1} is necessary to ensure that the utility potential is non-constant in the limit $\alpha \rightarrow 0$. Second, the tight bounds [20]

$$-\alpha^{-1} \sum_{i=1}^N O_i \left\{ \max_{1 \leq j \leq M} U_{ij}(x) + \ln M \right\} \leq V_{\text{Utility}}(x) \leq -\alpha^{-1} \sum_{i=1}^N O_i \left\{ \max_{1 \leq j \leq M} U_{ij}(x) \right\} \quad (2.17)$$

show that $V_{\text{Utility}}(x)$ may be finite when any $x_j \rightarrow -\infty$, and so an additional potential is needed to prevent zones from collapsing from a lack of activity. The bounds again show that the utility potential is closely related to the best alternative available to each of the origin zones.

(ii) Cost potential

The cost potential prevents each zone from becoming too large, and is justified by the notion that running costs increase with size. We therefore require that $\lim_{x_j \rightarrow +\infty} V(x) = +\infty$. In view of the equality constraints in (2.11), an appropriate cost potential is the total size or capacity of the system

$$V_{\text{Cost}}(x) = \sum_{j=1}^M W_j(x_j), \quad (2.18)$$

in which $W_j(x_j) = \exp(x_j)$ is as before. Since we have

$$\frac{\partial V_{\text{Cost}}(x)}{\partial x_j} = W_j(x_j), \quad (2.19)$$

this choice of potential yields the linear cost term in the overdamped Langevin dynamics considered in (2.6).

(iii) Additional potential

The final potential term must satisfy $\lim_{x_j \rightarrow -\infty} V(x) = +\infty$ and must grow sufficiently fast at infinity in order for (2.9) to be well defined. The purpose of the additional potential is to prevent zones from collapsing from a lack of activity. Such mechanisms are commonplace in urban and regional systems, for example continued investment or government initiatives. In view of the equality constraints in (2.11), a suitable potential function is

$$V_{\text{Additional}}(x) = \sum_{j=1}^M x_j, \quad (2.20)$$

which ensures that the attractiveness of each zone is finite. The partial derivatives of the additional potential function are

$$\frac{\partial V_{\text{Additional}}(x)}{\partial x_j} = 1. \quad (2.21)$$

Therefore, the finiteness constraint requires that there is an additional positive constant term in the deterministic part of the SDE model, given by (2.6), to ensure that the SDE has a well-defined stationary distribution.

(c) Model summary

In summary, we have specified the following potential function:

$$\epsilon^{-1} V(x) = \underbrace{-\alpha^{-1} \sum_{i=1}^N O_i \ln \sum_{j=1}^M \exp(\alpha x_j - \beta c_{ij})}_{\text{Utility}} + \underbrace{\kappa \sum_{j=1}^M \exp(x_j)}_{\text{Cost}} - \underbrace{\delta \sum_{j=1}^M x_j}_{\text{Additional}}, \quad (2.22)$$

which satisfies the assumptions in appendix A. The potential function is similar to the one obtained by reformulating the Harris and Wilson model in (2.6) and (2.7); however, it contains

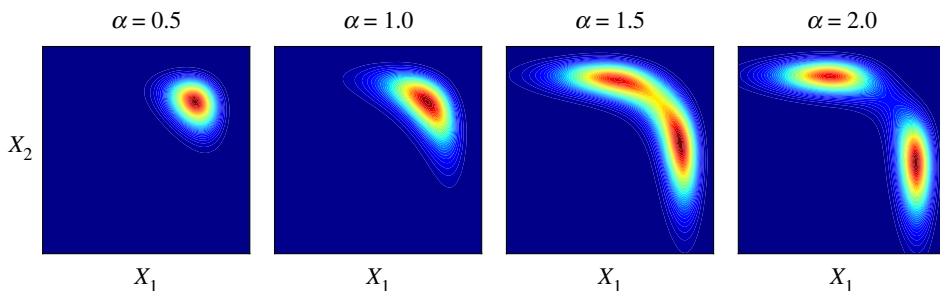


Figure 2. Illustration of the potential function for a small model comprising two competing zones. The profile of the potential function is largely determined by the α and β pairing; here, we have held β fixed and show $e^{-\gamma V(x)}$ for different values of α .

an additional term to prevent zones from collapsing. A stochastic generalization of the Harris and Wilson model is given by the overdamped Langevin diffusion in (2.6), for which the process converges at a fast rate to the well-defined Boltzmann–Gibbs measure described by (2.9). The corresponding size dynamics, in the form of (2.5), are given by the Stratonovich SDE

$$dW_j = \epsilon W_j (D_j - \kappa W_j + \delta) dt + \sigma W_j \circ dB_j, \quad W(0) = w_0, \quad (2.23)$$

which is a stochastic generalization of the original Harris and Wilson model that includes a positive shift to the multiplicative scale factor. In the limit $\delta, \sigma \rightarrow 0$, we obtain the original Harris and Wilson model in (2.4).

In the regime $\delta \rightarrow 0$, the potential function has stationary points coinciding with the fixed points of the original Harris and Wilson model. The stationary points for the potential function are given by M simultaneous equations

$$\sum_{i=1}^N O_i \frac{W_j^\alpha \exp(-\beta c_{ij})}{\sum_{k=1}^M W_k^\alpha \exp(-\beta c_{ik})} = \kappa W_j - \delta, \quad j = 1, \dots, M. \quad (2.24)$$

While the behaviour of the stochastic and deterministic models may be similar in low-noise regimes over finite time intervals, we emphasize that the asymptotic behaviour differs greatly between the two. Here, we consider a deterministic model to be given by (2.23) in the limit $\sigma \rightarrow 0$. For a deterministic model, the dynamics will converge to a stable fixed point satisfying (2.24), as determined by the initial condition. For a stochastic model, the system will converge to a statistical equilibrium that does not depend on the initial condition. As $t \rightarrow +\infty$, the stochastic model spends more time around the lower values of $V(x)$, which occur around stable stationary points, as summarized by the limiting stationary distribution given by (2.9).

We now comment on the Boltzmann–Gibbs measure described by (2.9). The Boltzmann–Gibbs measure is the equilibrium distribution of (2.6), but is also justified as a probability distribution for urban and regional structures with a maximum entropy argument. When considering the Boltzmann–Gibbs measure, we specify $\epsilon = 1$ to avoid over-parametrizing the model since the relative level of noise is controlled by the inverse temperature $\gamma = 2\sigma^{-2}$. As $\gamma \rightarrow +\infty$, the Boltzmann distribution collapses to a Dirac mass around the global minimum of $V(x)$, which is unlikely to provide a good fit to the observed urban structure. As $\gamma \rightarrow 0$, the distribution of sizes approaches an improper uniform distribution. The profile of $V(x)$ is largely influenced by the pair of α and β values, as illustrated in figure 2. A large α relative to β results in all activity taking place in one of the zones, whereas this regime is unlikely when α is low relative to β .

Lastly, we can use the deterministic model to specify appropriate values of the cost of floorspace κ and the additional parameter δ . By defining

$$\kappa = \frac{1}{K} \left(\sum_{i=1}^N O_i + \delta M \right), \quad (2.25)$$

the deterministic model converges to an equilibrium with a total size of K units. Setting κ as in (2.25) is justified with a supply and demand argument [21]. For simplicity, we use $K = 1$; the choice is arbitrary. We can then specify δ relative to the size of the smallest zone possible, since at equilibrium the size of a zone with no inward flows is δ/κ .

3. Parameter estimation

In this section, we consider the inverse problem; the task of determining α and β from observed urban structure. The value of α describes consumer preference towards more popular destinations and the value of β describes how much consumers are inconvenienced by travel. We use retail activity as an archetype; however, our methodology is general and can be applied to other singly constrained systems. While α and β can be estimated using discrete choice models [22–25], this approach requires large volumes of flow data and is impractical for large systems. We instead make use of the model described by the Boltzmann–Gibbs measure in §2.

We formulate the task of inversion as a statistical inference problem, as advocated in [12]. The Bayesian approach is based on the following principles: the unknown parameters are modelled as random variables; our degree of uncertainty is described by probability distributions; and the solution to the inverse problem is the posterior probability distribution. Unlike classical methods, a Bayesian approach is well posed and allows us to incorporate prior knowledge of the unknowns into the modelling process. A Bayesian approach yields a posterior probability over the model parameters, and the parameter values can be determined from the posterior mean or maximum *a posteriori* estimates.

(a) A Bayesian approach to parameter estimation

The Boltzmann distribution in §2 is assumed to form an integral part of the data-generating process; however, further uncertainty arises from measurement noise.⁴ To this end, we assume that an observed configuration of urban structure $Y \in \mathbb{R}_{>0}^M$ is related to some latent sizes $W \in \mathbb{R}_{>0}^M$ and multiplicative noise $E \in \mathbb{R}_{>0}^M$ by

$$\ln Y = \ln W + \ln E. \quad (3.1)$$

Multiplicative noise is appropriate as all measurements are positive, and there is more scope for error when measuring larger zones. As before, it is natural to work in terms of log-sizes $X = \ln W \in \mathbb{R}^M$, and we assume that $X \sim \rho_\infty$ is a realization of the Boltzmann–Gibbs measure given by (2.9) and (2.22). The latent variables X depend on the model parameters $\Theta = \{\alpha, \beta\} \in \mathbb{R}_{>0}^2$, which we summarize by a single variable for notational convenience. We assume that $\ln E \sim N(0, \Sigma)$ is a realization of Gaussian noise for some symmetric positive definite covariance matrix $\Sigma \in \mathbb{R}^{M \times M}$.

We specify a prior $\pi(\theta)$ on the model parameters. The prior distribution for the latent variables is denoted $\pi(x|\theta)$ and is given by (2.9), which we repeat here to make the θ -dependence explicit in our notation

$$\pi(x|\theta) = \frac{1}{z(\theta)} \exp(-\gamma V_\theta(x)), \quad z(\theta) = \int_{\mathbb{R}^M} \exp(-\gamma V_\theta(x)) dx. \quad (3.2)$$

We emphasize that $\pi(x|\theta)$ is only known up to a normalizing constant $z(\theta)$ that is a function of θ . The likelihood function $\pi(y|x)$ is the Gaussian density given by (3.1). The joint posterior density

⁴It may be necessary to include a model error term if the Boltzmann–Gibbs measure provides a poor fit to the data.

then has the form

$$\pi(x, \theta | y) \propto \pi(\theta) \frac{1}{z(\theta)} \exp(-\gamma V_\theta(x)) \pi(y | x), \quad (3.3)$$

and is ‘doubly intractable’ as both the normalization factor of (3.3) and the function $z(\theta)$ are unknown. The estimation of $z(\theta)$ is a notoriously challenging problem as it requires the integration of a complex function over a high-dimensional space [13,15]. The normalization constant $z(\theta)$ is a probability-weighted sum of all possible outcomes and is a necessary penalty against model complexity. The θ -dependence for the $z(\theta)$ -term poses significant computational challenges as the joint posterior density cannot be evaluated at all, not even up to an irrelevant multiplicative constant.

(b) Computational strategies

To explore the posterior distribution, we resort to numerical simulation and use MCMC to estimate integrals of the form

$$\mathbb{E}[g(X, \Theta) | Y = y] := \int_{\mathbb{R}^2_0} \int_{\mathbb{R}^M} g(x, \theta) \pi(x, \theta | y) dx d\theta, \quad (3.4)$$

where $g(x, \theta)$ is an integrable function of interest. For example, (3.4) can be used to compute the mean, variance and density estimates of the posterior marginals. As suggested in [13], we can use an approximate method to estimate $z(\theta)$. We consider the quadratic approximation of $V_\theta(x)$ that is obtained from a second-order Taylor expansion around its global minima m_θ

$$\hat{V}_\theta(x) = V_\theta(m_\theta) + \frac{1}{2}(x - m_\theta)^T \Delta V_\theta(m_\theta)(x - m_\theta). \quad (3.5)$$

As the integral for $z(\theta)$ only has significant contributions in the neighbourhood of m_θ , where (3.5) is a good approximation, we estimate $z(\theta)$ as

$$\begin{aligned} z(\theta) &\approx \int_{\mathbb{R}^M} \exp(-\gamma \hat{V}_\theta(x)) dx \\ &= \exp(-\gamma V_\theta(m_\theta)) \int_{\mathbb{R}^M} \exp\left(-\frac{\gamma}{2}(x - m_\theta)^T \Delta V_\theta(m_\theta)(x - m_\theta)\right) dx \\ &= \exp(-\gamma V_\theta(m_\theta)) \frac{(2\pi \gamma^{-1})^{M/2}}{|\Delta V_\theta(m_\theta)|^{1/2}}. \end{aligned} \quad (3.6)$$

This is known as a saddle point approximation and is asymptotically accurate as $\gamma \rightarrow +\infty$ [26]. In all but special cases, the global minima of $V_\theta(x)$ is unique and can be found inexpensively using Newton-based optimization; for example, using the limited memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm with the right initial condition [27]. We run the optimization procedure for multiple initializations to provide a good coverage of the basins, although this is only necessary for $\alpha > 1$. The curvature term $\Delta V_\theta(m_\theta)$ is given by (A 5). With (3.6), we can proceed with the MCMC scheme in appendix C.

To obtain more accurate posterior summaries, especially in the case that the saddle point approximation performs poorly, we look towards a consistent estimator of (3.4). Despite the intractable $z(\theta)$ term, we are able to construct a Markov chain $\{X^{(i)}, \Theta^{(i)}, \Omega^{(i)}\}_{i=1}^n$ such that

$$\mathbb{E}[g(X, \Theta) | Y = y] = \lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n \Omega^{(i)} g(X^{(i)}, \Theta^{(i)})}{\sum_{k=1}^n \Omega^{(k)}}. \quad (3.7)$$

The estimator requires that we can obtain unbiased estimates of the reciprocal normalizing constant $1/z(\theta)$, which can be obtained by randomly truncating an infinite series involving importance sampling estimates of $z(\theta)$ [15,28]. The estimator given by (3.7) is an importance-sampling-style estimator, but with each weight $\Omega^{(i)} \in \{-1, 1\}$ equal to the sign of the unbiased estimate of $1/z(\theta)$ for that iteration. The suitability of the scheme is dependent on being able to obtain precise importance sampling estimates of $z(\theta)$, which is challenging for low-noise regimes due to the concentration of measure. Negative values of $\Omega^{(i)}$ arise from imprecision in the $z(\theta)$

estimates and have the effect of increasing the variance of the estimator given by (3.7). Further details of the scheme are in appendix C.

(c) Implementation details

We specify weakly informative uniform priors on α and β , restricted to the interval $[0, 2]$ with a suitable scaling of β determined by a preliminary study, as done in [1,21].⁵ In this setting, we are able to compare our inferred α and β values with the R^2 analysis performed for the deterministic Harris and Wilson model in [1,21]. While ideally we would place priors on all parameters that specify the Boltzmann–Gibbs measure, we acknowledge that in doing so we would encounter both identifiability issues and tuning difficulties with regards to the importance sampling scheme for $z(\theta)$. We are able to proceed by fixing the remaining hyperparameters to suitable values. We specify $\epsilon = 1$ to avoid over-parametrizing the model and specify γ to reflect a desired level of noise. We set δ to the size of the smallest zone. This is justified by considering the Gamma distribution of a zone with no inward flows. We normalize the origin quantities and total sizes to determine κ from (2.25). Lastly, for demonstration purposes, we specify independent and homogeneous observation noise by setting $\Sigma = \lambda^2 I$, where λ is the standard deviation of the noise.

To compute low-order summary statistics of the form in (3.4), we use the Monte Carlo scheme in appendix C, consisting of a block Gibbs scheme. We use a Metropolis–Hastings random walk with reflective boundaries for the Θ -updates and a Hamiltonian Monte Carlo (HMC) for the X -updates. We tune the step size parameter for the Θ -updates to obtain an acceptance rate in the range 30–70%, and we tune the step size and number of steps for the X -updates to obtain an acceptance rate of at least 90%. For the Θ -updates, we require either global minima of $V_\theta(x)$, for the saddle point approximation in (3.6), or unbiased estimates of $1/z(\theta)$, for the pseudo-marginal MCMC framework described in appendix C. When requiring global minima of $V_\theta(x)$, we perform multiple runs of the L-BFGS algorithm for M different initial conditions. When requiring consistent estimates of $1/z(\theta)$, we use annealed importance sampling (AIS) with HMC transition kernels. We initialize AIS with the log-gamma distribution that can be obtained from $\pi(x|\theta)$ by letting $\alpha, \beta \rightarrow 0$. We produce unbiased estimates by truncating an infinite series of importance sampling estimates with a random stopping time T with $\Pr(T \geq k) \propto k^{-1.1}$, and therefore requiring $T + 1$ runs of AIS. Running AIS a large number of times is a computationally intensive task; however, the estimates can be obtained in parallel.

4. Case study: the London retail system

In this section, we illustrate our proposed methodology with an aggregate retail model using London, UK, data similar to the example in [1,21]. While the model can be improved with disaggregation to capture further problem-specific characteristics, the underlying arguments would remain the same. We demonstrate how the Boltzmann–Gibbs measure can be used to simulate configurations of urban structure before setting out to infer the α and β values in the utility function. In the context of retail, the attractiveness term in (2.13) is justified by the benefit consumers gain from the improved range of options and economies of scale, and the cost term represents inconvenience of travel. The inverse problem is of particular interest in the context of retail as the flow data are difficult to obtain. On the other hand, urban structure is relatively straight-forward to measure and may be routinely available. While some attempts have been made in the literature to estimate the parameters of a similar spatial interaction model [1,21], these approaches are somewhat ad hoc but do provide a basis of comparison.

We obtain measurements of retail floorspace for London town centres for 2008 from a London Town Centre Health Check report prepared by the Greater London Authority [29]. We only include town centres classified as international, metropolitan and major town centre

⁵In our implementation, we normalize the cost matrix so that all elements sum to 7×10^5 .

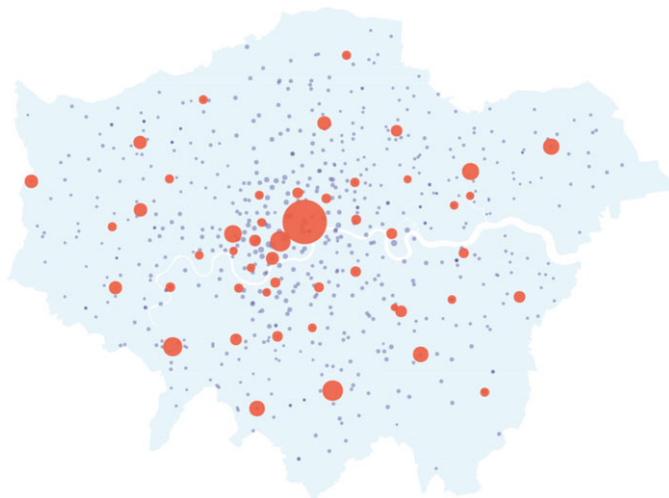


Figure 3. Visualization of the observation data Y (red) over the map of London. The red markers indicate the destination zones, which are the 49 town centres, and the blue markers indicate the origin zones, which are the 625 residential wards. The sizes of the markers are given by the respective Y and O values, and each zone is plotted at its longitude–latitude coordinate.

classifications in our study, giving $M = 49$ town centres. The remaining town centres are mostly district town centres that have a relatively high concentration of convenience goods and more localized catchment; we argue that these would be better modelled separately. We determine the origin quantities from ward-level household and income estimates, with $N = 625$, published by the Greater London Authority [30,31]. We take the origin quantities to be the spending powers as given by the population size multiplied by the average income. The floorspace measurements and residential data are presented in figure 3, over the map of London [32]. In our implementation, we calculate the cost matrix from Euclidean distance, although a better representation would use a transport network [1].

We first perform a preliminary study of our model in the limit of no observation noise $\lambda \rightarrow 0$, in which case the θ -marginal of (3.3) is

$$\pi(\theta|y) \propto \pi(\theta) \frac{1}{z(\theta)} \exp(-\gamma V_\theta(x)). \quad (4.1)$$

With this simplification, we are able to evaluate the posterior probabilities over a grid of α and β values. We evaluate the probabilities over a 100×100 grid for $\gamma = 10^2$ and $\gamma = 10^4$, representing high-noise and low-noise regimes, respectively. Using the justification given in the previous section, we specify $\delta = 0.006$ and $\kappa = 1.3$. We produce the grid by estimating $z(\theta)$ with the saddle point approximation in (3.6).⁶ The results are presented in figure 4, in which the scales indicate that the model with high noise provides the better explanation of the data. We find that the best fit for the high-noise regime is $\alpha = 0.90$ and $\beta = 0.46$ and the best fit for the low-noise regime is $\alpha = 1.18$ and $\beta = 0.28$. As expected, the low-noise regime suggests stronger attractiveness effects as the model with a higher level of noise is able to explain variation by stochastic growth. The α and β values are positively correlated; this can be seen in figure 4 and is due to the competing effects in the utility function in (2.13).

In [1], the authors perform an R^2 analysis that we replicate here for our deterministic version of the Harris and Wilson model for a basis of comparison. The predicted value W_{Pred} is taken to be the equilibrium obtained from the ODE model given by (2.23) with $\sigma \rightarrow 0$ and the initial condition $w_0 = Y$. The R^2 value is defined as $R^2 = 1 - SS_{\text{res}}/SS_{\text{tot}}$, where $SS_{\text{res}}/SS_{\text{tot}}$ is the ratio of

⁶It is very difficult to estimate $z(\theta)$ for high values of γ using importance sampling techniques due to the concentration of measure.

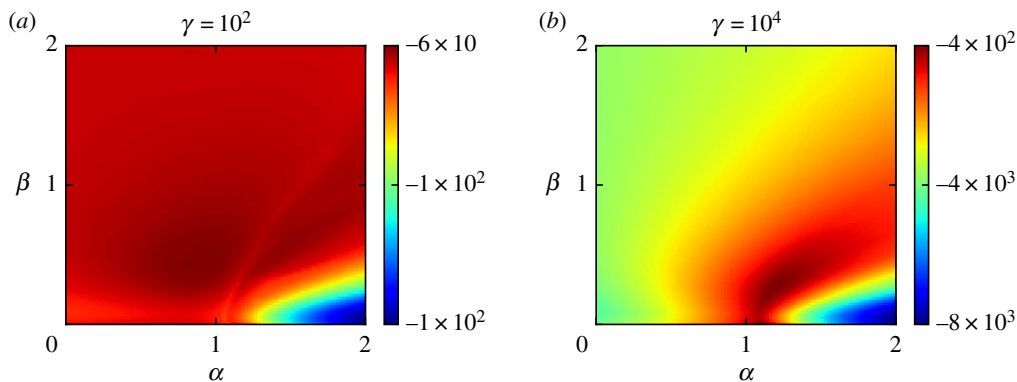


Figure 4. Evaluations of the logarithm of (4.1) over a grid of 100×100 values of α and β for a regime with high noise (a) and a regime with low noise (b).

the variance of the residuals $Y - W_{\text{Pred}}$ and the variance of the observed Y . While our Bayesian approach is fundamentally different, and we should not expect to obtain too similar results, the R^2 analysis yields a best fit of $\alpha = 1.36$ and $\beta = 0.42$. This is agreeable with the findings for the low-noise regime in figure 4. Furthermore, there are some strong similarities between the profile of posterior probabilities and the profile of the R^2 -values. First, both approaches find that the poorest fit is for a regime in which α is too high and β is too low; these values result in most activity taking place in a single zone. Second, both approaches agree in that a good fit can be found for $1 < \alpha < 2$.

Next, we draw the latent variables from the prior distribution $\pi(x|\theta)$ to verify the suitability of the modelling. For illustrative purposes, we consider a range of α values across $[0, 2]$, and hold $\beta = 0.5$ fixed. For the regime with high noise, the approximate draws are obtained by running a Markov chain of length 10 000 using HMC combined with parallel tempering for five different temperature levels [14]. For the regime with low noise, we plot configurations of the global minima of $V_\theta(x)$ obtained from numerical optimization as there is little variation between samples. The results are in figures 5 and 6, respectively. It can be seen that higher values of α and lower values of β create a sparse structure in that all activity takes place in very few zones. Conversely, lower values of α and higher values of β lead to a more dense structure.

We now return to the observation model in (3.1) to account for observation noise in the data. For illustrative purposes, we specify $\lambda = 0.1$ so that the relative noise for a zone of size $1/M^7$ is 3%. Although an improved specification of observation noise from a preliminary study would lead to more accurate inferences, the arguments and methodology we are presenting would remain the same. We run Markov chains of length 20 000. For the high-noise regime, we use the pseudo-marginal MCMC methodology in appendix C. Our importance sampling estimates comprised 10 particles and 50 equally spaced inverse temperatures. For the low-noise regime, we were unable to obtain precise importance sampling estimates due to the concentration of measure, so used the MCMC methodology in appendix C with the saddle point approximation in (3.6). For both examples, the empirical autocorrelation for α and β is below 0.2 after 25 steps. For the pseudo-marginal MCMC scheme, 88% of the signs are positive, which is acceptably high for the scheme to be used.

Plots of the smoothed density estimates for α and β for the high-noise regime are presented in figure 7. Plots of the latent sizes showing the posterior mean ± 3 s.d. are presented in figure 8 alongside plots of the expected residuals and observation data. The posterior marginals of α and β give mean ± 1 s.d. estimates of 0.35 ± 0.28 and 1.09 ± 0.46 , respectively, which appear reasonable in light of the analysis in figure 4. The plots of the expected residuals and observation

⁷Relative noise in this context is $\lambda/\log M$.

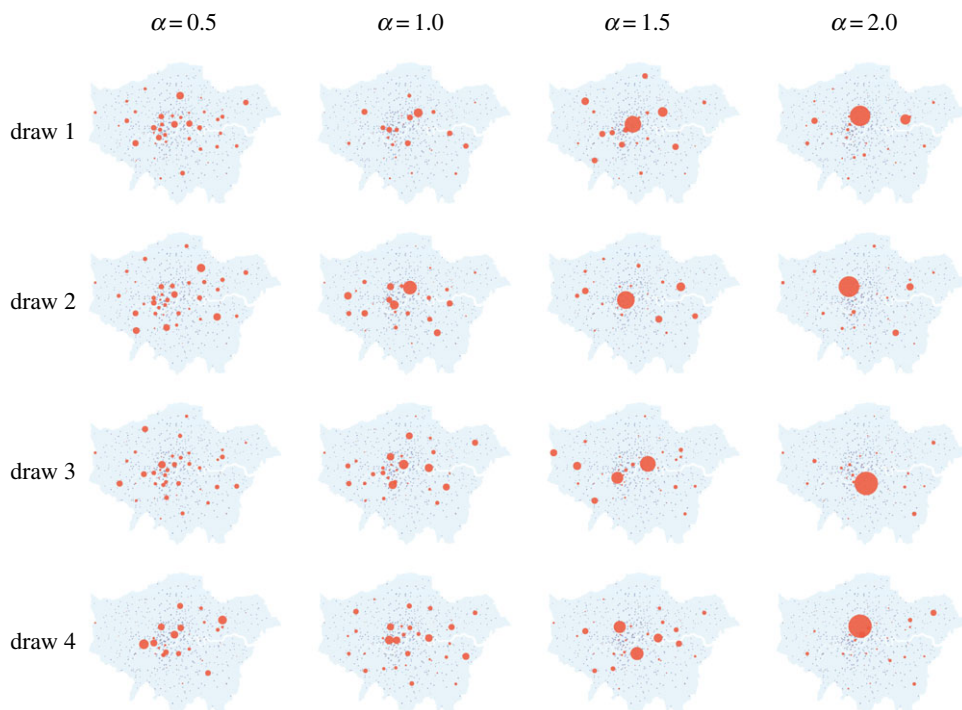


Figure 5. Approximate draws of the latent variables from $\pi(x|\theta)$ for a high-noise regime with $\gamma = 10^2$, obtained by running a Markov chain of length 10 000 using HMC combined with parallel tempering. Each row shows four randomly selected states from the Markov chain with α as specified and $\beta = 0.5$.

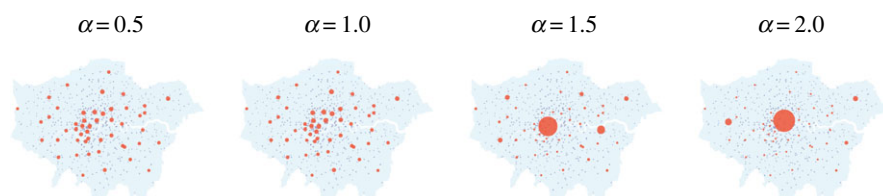


Figure 6. Global minima of the latent variables from $\pi(x|\theta)$, obtained by running the L-BFGS algorithm for M different initial conditions. These configurations are representative of draws from $\pi(x|\theta)$ in a low-noise regime with $\gamma \gg 1$.

data suggest that the model provides a reasonable fit to the data, and that the assumption of homogeneous observation noise is reasonable. This is to be expected as the high-noise model provides a flexible model. After taking into account the observation noise, a weaker attractiveness effect was observed.

Similar plots are presented in figures 9 and 10 for the low-noise regime, and the posterior marginals of α and β give a mean of 1.17 ± 0.01 and 0.26 ± 0.01 , respectively. The inferred values for the low-noise regime are in line with figure 4. Both sets of posterior summaries suggest that attractiveness and inconvenience effects are present in the data, though the inferred α values are considerably higher for the low-noise model. The plots of the expected residuals and observation data suggest that the model also provides a reasonable fit to the data; however, there is more dispersion in the plotted quantities and possibly a degree of heteroscedasticity due to the less flexible model. The model with more noise favours the simpler explanation that most variation is due to stochastic growth, whereas the low-noise model is more constrained. There is notably

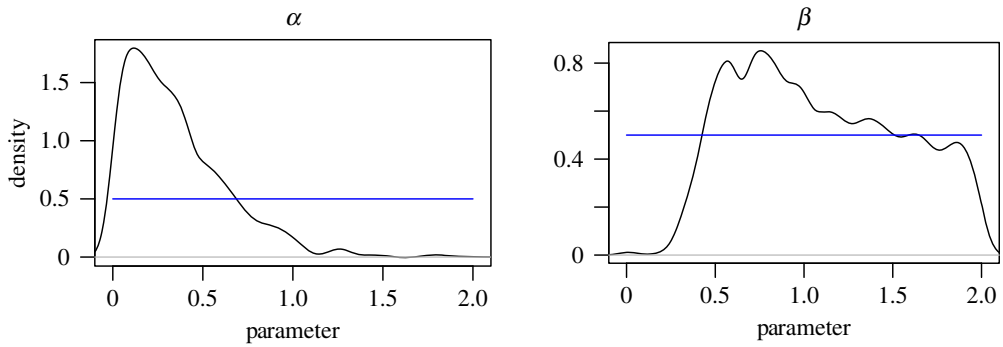


Figure 7. Posterior marginal density estimates for α and β for the high-noise regime ($\gamma = 10^2$). The smooth density estimates were obtained by applying (3.4) to a Gaussian kernel. The blue line indicates the uniform prior density.

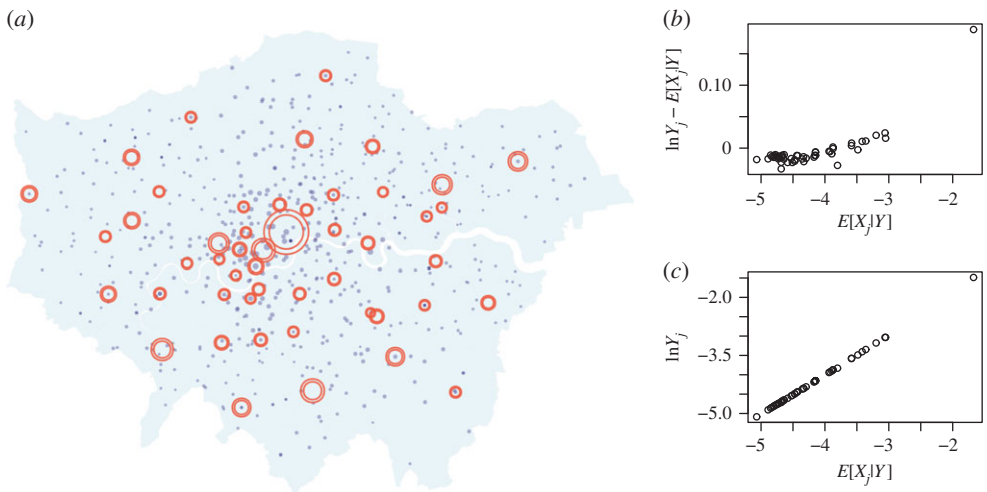


Figure 8. Visualization of the posterior latent variables X for the high-noise regime ($\gamma = 10^2$). (a) The outer and inner rings show the posterior mean ± 3 s.d., respectively. (b) Expected attractiveness against the expected residual. (c) Expected attractiveness against the observed value.

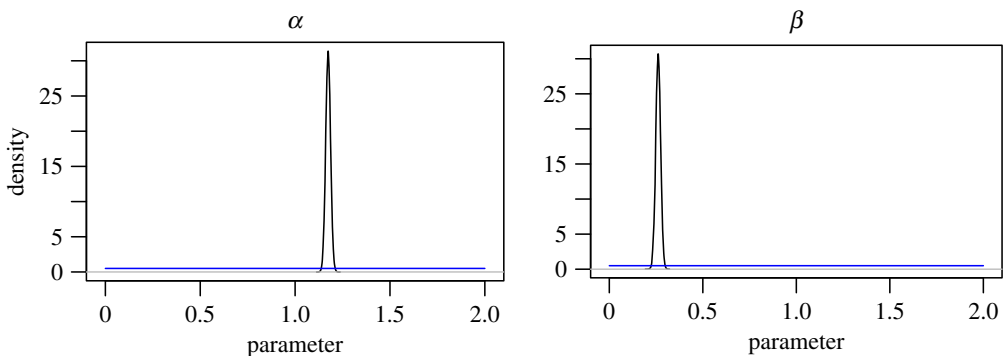


Figure 9. Posterior marginal density estimates for α and β for the low-noise regime ($\gamma = 10^4$). The smooth density estimates were obtained by applying (3.4) to a Gaussian kernel. The blue line indicates the uniform prior density.

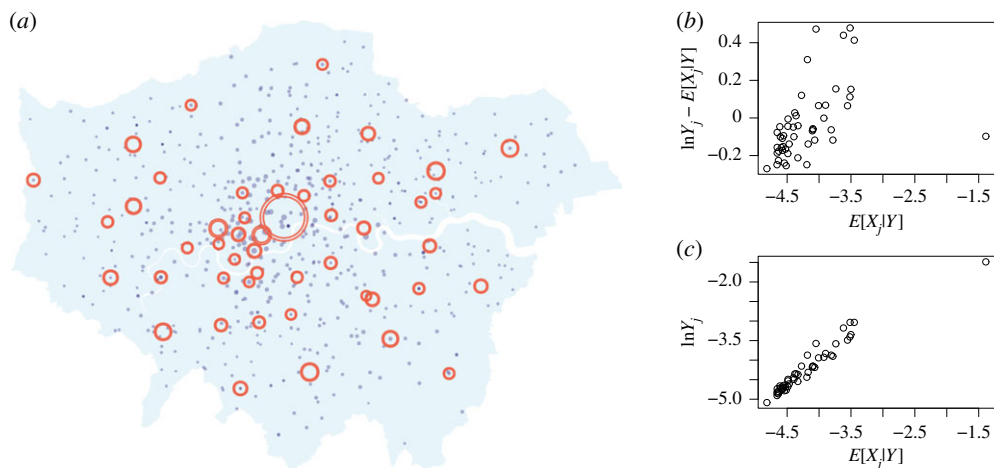


Figure 10. Visualization of the posterior latent variables X for the low-noise regime ($\gamma = 10^4$). (a) The outer and inner rings show the posterior mean ± 3 s.d., respectively. (b) Expected attractiveness against the expected residual. (c) Expected attractiveness against the observed value.

more uncertainty in the latent variables and the model parameters for the high-noise regime, as there are more possible explanations for the observation data. The uncertainty in the α and β estimates is so great for the high-noise regime that limited insights are gained for the purposes of model calibration. As a result, we conclude that strong assumptions are required in the prior modelling in order to be able to exploit known structure in the data-generating process. The required assumptions can be made, for example, through the prior modelling of α and β or by specifying a high value of γ . On the other hand, the low-noise regime results in very confident posteriors. Although the resulting inferences are consistent with previous studies, care must be taken to avoid being overconfident in a particular model by not adequately accounting for uncertainty in the modelling process.

5. Discussion

We have developed a novel stochastic model to simulate realistic configurations of urban and regional structure. Our model is a substantial improvement on existing deterministic models as it fully addresses the uncertainties arising in the modelling process. Unlike existing time-stepping schemes, our model can be used to simulate realistic configurations of urban structure using MCMC methods without recourse to numerical error. We have demonstrated that our model can be used to infer the components of a utility function from observed structure, thereby providing an alternative to the existing discrete choice models. The key advantage is that we avoid the need to collect vast amounts of flow data. While we have presented our methodology in the context of consumer-led behaviour, our approach is applicable to other urban and regional settings such as archaeology, logistics, healthcare and crime to suggest a few.

Our work has led to specific areas for further research. We are actively investigating the deployment of our methodology to large-scale urban systems, for which there are substantial computational challenges to overcome. The cost of a potential or gradient evaluation is $O(NM)$; however, increasing M means that the $z(\theta)$ estimates are more challenging to obtain owing to the curse of dimensionality. It is of interest to develop more tractable methods, for example optimization based, so that inference can be performed for international models on a practical time scale. We have presented an aggregate model that can be refined to better represent domain-specific characteristics as discussed in [1]. It remains to use the proposed methodology as part of a more realistic study with wider objectives. Lastly, we emphasize that our methodology is only

applicable to cross-sectional data. In practice, many applications of interest require processing time-series data that are highly correlated over time. In this setting, we would need to solve the filtering or smoothing problem for (2.23), and in doing so would also need to account for general trends and seasonality effects that are exogenous to our model. Our work continues to be part of ongoing efforts to draw insights from data by making use of the known mathematical structure [33].

Data accessibility. The code and data used for the case study in this manuscript can be found at the following repository: https://github.com/lellam/cities_and_regions.

Authors' contributions. L.E., M.G., G.A.P. and A.W. designed the research, performed the research and prepared the article. L.E. performed the numerical experiments.

Competing interests. We declare that we have no competing interests.

Funding. L.E. was supported by the EPSRC (EP/P020720/1). M.G. was supported by the EPSRC (EP/J016934/3, EP/K034154/1, EP/P020720/1, EP/R018413/1), an EPSRC Established Career Fellowship, the Alan Turing Institute — Lloyd's Register Foundation Programme on Data-Centric Engineering and a Royal Academy of Engineering Research Chair in Data Centric Engineering. G.A.P. was supported by the EPSRC (EP/P031587, EP/L020564, EP/L024926, EP/L025159). A.W. was supported by the EPSRC (EP/M023583/1). This work was supported by the Alan Turing Institute under EPSRC grant no. EP/N510129/1.

Acknowledgements. We thank the anonymous reviewers for their helpful comments and suggestions.

Appendix A. Assumptions for the potential function

We make the following assumptions for the potential function $V(x)$ with reference to the overdamped Langevin diffusion described by (2.6) and Boltzmann–Gibbs measure defined by (2.9):

- (i) $V(x)$ is C^2 and is confining in that $\lim_{|x| \rightarrow +\infty} V(x) = +\infty$ and

$$e^{-\gamma V(x)} \in L^1(\mathbb{R}^M), \quad \forall \gamma > 0. \quad (\text{A } 1)$$

- (ii) $V(x)$ satisfies the following inequality for some $0 < d < 1$:

$$\liminf_{|x| \rightarrow \infty} \{(1-d)|\nabla V(x)|^2 - \gamma^{-1} \Delta V(x)\} > 0. \quad (\text{A } 2)$$

Assumption (i) is necessary to ensure that the Boltzmann–Gibbs measure is well defined. Assumptions (i) and (ii) are sufficient to show that the distribution of $X(t)$, described by (2.6), converges exponentially fast to its Boltzmann–Gibbs measure. The reader is referred to [34] for further details.

The integrability condition in assumption (i) is satisfied by lemma 3.14 in [35]. To show that assumption (ii) holds, we define

$$\Lambda_{ij} = \frac{\exp(\alpha x_j - \beta c_{ij})}{\sum_{k=1}^M \exp(\alpha x_k - \beta c_{ik})}, \quad (\text{A } 3)$$

then

$$|\nabla V(x)|^2 = \sum_{j=1}^M \left| \sum_{i=1}^N O_i \Lambda_{ij} - \kappa \exp(x_j) + \delta \right|^2 \quad (\text{A } 4)$$

and

$$\Delta V(x) = \sum_{j=1}^M \left\{ \kappa \exp(x_j) - \alpha \sum_{i=1}^N O_i \Lambda_{ij} (1 - \Lambda_{ij}) \right\}. \quad (\text{A } 5)$$

Then for $0 < d < 1$ we have

$$\lim_{x_j \rightarrow +\infty} (1-d)|\nabla V(x)|^2 - \gamma^{-1} \Delta V(x) = +\infty \quad (\text{A } 6)$$

and

$$\lim_{x_j \rightarrow -\infty} (1-d)|\nabla V(x)|^2 - \gamma^{-1} \Delta V(x) \geq \delta^2 > 0, \quad (\text{A } 7)$$

as claimed.

Appendix B. Utility function for a singly constrained potential

In this appendix, we present an alternative argument to obtain the utility potential V_{Utility} as defined by (2.12). We rewrite the destination quantities in (2.2) as a gradient flow

$$D_j = -\frac{\partial V_{\text{Utility}}(x)}{\partial x_j}, \quad (\text{B } 1)$$

and look towards specifications of V_{Utility} that satisfy the constraint in (2.1). The constraint is satisfied whenever the flows leaving the origin zones are convex sums in that

$$T_{ij}(x) = O_i v_{ij}(x), \quad \sum_{j=1}^M v_{ij} = 1, \quad v_{ij} \geq 0. \quad (\text{B } 2)$$

Instead we can express (B 2) in terms of utility functions U_{ij} and some positive function φ so that

$$T_{ij} = O_i \frac{\varphi(U_{ij})}{\sum_{k=1}^M \varphi(U_{ik})}. \quad (\text{B } 3)$$

By inspection, we look for a potential function of the form

$$V_{\text{Utility}} = -\sum_{i=1}^N O_i \left\{ f_i \ln \sum_{j=1}^M \varphi(U_{ij}) \right\}, \quad (\text{B } 4)$$

for some functions f_i . Then by taking the gradient and substituting into (B 1), we obtain the requirements

$$\frac{\varphi(U_{ij})}{\sum_{k=1}^M \varphi(U_{ik})} = \frac{df_i}{dx_j} \ln \sum_{j=1}^M \varphi(U_{ij}) + f_i \frac{d\varphi(U_{ij})}{dx_j} \left(\sum_{k=1}^M \varphi(U_{ik}) \right)^{-1}, \quad (\text{B } 5)$$

for $i = 1, \dots, N$. The requirements are satisfied for $\varphi(\cdot) = \exp(\cdot)$ when each utility function is linear with respect to the attractiveness of the destination zone in question,

$$U_{ij} = \alpha_i x_j + \beta_{ij}, \quad (\text{B } 6)$$

provided that each $\alpha_i \neq 0$ and that $f_i = \alpha_i^{-1}$. The resulting potential function is

$$V_{\text{Utility}}(x) = -\sum_{i=1}^N \alpha_i^{-1} O_i \ln \sum_{j=1}^M \exp(U_{ij}), \quad (\text{B } 7)$$

which is slightly more general than the potential considered in (2.12).

Appendix C. Markov chain Monte Carlo for doubly intractable distributions

The idea behind MCMC is to construct an ergodic Markov chain whose time average can be used to estimate integrals of interest [14,36]. We use a block Metropolis-within-Gibbs scheme and alternate between θ -updates and X -updates. The following steps can be repeated in succession to obtain a Markov-chain chain that is $\pi(x, \theta|y)$ -invariant.

Latent variable update. Hamiltonian Monte Carlo can be used for the X -updates to suppress random walk behaviour [37]. We propose momentum variables $P \sim N(0, I)$ and update (X, P) by simulating Hamiltonian dynamics with a volume-preserving integrator to obtain (X', P') . We accept/reject

according to the Metropolis–Hastings acceptance probability

$$a_X(x', p' | x, p) = \min \left\{ 1, \frac{\pi(y|x', \theta) \exp(-\gamma V_\theta(x') - (1/2)|p'|^2)}{\pi(y|x, \theta) \exp(-\gamma V_\theta(x) - (1/2)|p|^2)} \right\}, \quad (\text{C } 1)$$

to correct for numerical error from numerically simulating Hamiltonian dynamics.

Model parameter update. Random walk Metropolis with reflective boundaries can be used for the Θ -updates. The Metropolis–Hastings acceptance probability is given by

$$a_\Theta(\theta' | \theta) = \min \left\{ 1, \frac{\pi(y|x', \theta') z(\theta) \exp(-\gamma V_{\theta'}(x)) \pi(\theta')}{\pi(y|x, \theta) z(\theta') \exp(-\gamma V_\theta(x)) \pi(\theta)} \right\}. \quad (\text{C } 2)$$

The key challenge arises from proposing new Θ -values, in which case the acceptance probability contains an intractable ratio $z(\theta)/z(\theta')$. We can either proceed with a deterministic estimate of $z(\theta)$, at the expense of a bias, or we can obtain a consistent estimator of (3.4) with pseudo-marginal MCMC [38] provided that unbiased and reasonably precise estimates of $1/z(\theta)$ are available.

(a) Unbiased estimates of the reciprocal of the normalizing constant

An unbiased estimate of $z(\theta)$ is given by averaging over a batch of importance weights. The importance weights are evaluations of

$$w(x) = \frac{\exp(-\gamma V_\theta(x))}{q(x)}, \quad (\text{C } 3)$$

at locations drawn from a proposal distribution with density $q(x)$. By Jensen's inequality, the reciprocal of an importance sampling estimate of $z(\theta)$ is a biased estimate of $1/z(\theta)$. Instead, unbiased estimates of $1/z(\theta)$ can be obtained by randomly truncating an infinite series: for a sequence $\{\mathcal{V}_i\}$ satisfying $\lim_{i \rightarrow +\infty} \mathbb{E}[\mathcal{V}_i] = 1/z(\theta)$, and a random stopping time T , the estimator

$$S = \mathcal{V}_0 + \sum_{i=1}^K \frac{\mathcal{V}_i - \mathcal{V}_{i-1}}{\Pr(T \geq i)} \quad (\text{C } 4)$$

gives an unbiased estimate of $1/z(\theta)$ [28,39–41]. We follow [15] and use the increasing averages estimator

$$\mathcal{V}_i = \frac{i+1}{\sum_{k=0}^i w(\hat{X}^{(k)})}, \quad \hat{X}^{(k)} \sim q, \quad (\text{C } 5)$$

with reference to the importance sampling weights described by (C 3). The unbiased estimates of $1/z(\theta)$ can have high variance when the importance weights are highly variable. Fortunately, importance sampling may be carried out on an augmented state space, for example using AIS [42].

(b) Pseudo-marginal Markov chain Monte Carlo

The unbiased estimators of $1/z(\theta)$ given by (C 4) may be a negative estimate, which prohibits the use of the pseudo-marginal MCMC. Fortunately, the so-called 'sign problem' has been addressed in [28]. We can use the following importance sampling style of estimator that gives a consistent estimator of (3.4) in that:

$$\mathbb{E}_{x,\theta|y}[g] = \lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n \Omega^{(i)} g(X^{(i)}, \Theta^{(i)})}{\sum_{k=1}^n \Omega^{(k)}}, \quad (\text{C } 6)$$

where $\{X^{(i)}, \Theta^{(i)}, \Omega^{(i)}\}_{i=1}^n$ is a Markov chain obtained using the Metropolis-within-Gibbs scheme described at the start of this section but with the following acceptance probability for Θ -updates:

$$a_\Theta(\theta' | \theta) = \min \left\{ 1, \frac{\pi(y|x', \theta') |S'| \exp(-\gamma V_{\theta'}(x)) \pi(\theta')}{\pi(y|x, \theta) |S| \exp(-\gamma V_\theta(x)) \pi(\theta)} \right\} \quad (\text{C } 7)$$

and $\Omega^{(i)} = \text{sgn}(S^{(i)})$. It is necessary to cache the value of $S^{(i)}$ at each iteration as part of the pseudo-marginal MCMC scheme.

Appendix D. Table of key parameters and variables

For convenience, we provide a table of the key parameters and variables with brief explanations below.

parameter	explanation	reference
α	attractiveness scaling parameter	(2.3)
β	cost scaling parameter	(2.3)
δ	additional parameter	(2.10)
ϵ	responsiveness parameter	(2.4)
γ	inverse temperature	(2.6)
κ	cost per unit size	(2.4)
λ	standard deviation of observation noise	(3.1)
σ	noise parameter, equal to $\sqrt{2\gamma^{-1}}$	(2.23)
O_i	origin quantity for origin zone i	(2.1)
D_j	destination quantity for destination zone j	(2.2)
T_{ij}	flow from origin zone i to destination zone j	(2.1)
U_{ij}	utility function for a flow from origin zone i to destination zone j	(2.13)
c_{ij}	cost of a flow from origin zone i to destination zone j	(2.3)
W_j	size of destination zone j	(2.4)
X_j	attractiveness of destination zone j , given by $\ln W_j$	(2.6)
Y_j	observed size of destination zone j	(3.1)
N	number of origin zones	(2.1)
M	number of destination zones	(2.2)

References

1. Dearden J, Wilson AG. 2015 *Explorations in urban and regional dynamics: a case study in complexity science*, vol. 7. London, UK: Routledge.
2. Marshall A. 1920 *Principles of economics: an introductory volume*. London, UK: Royal Economic Society (Great Britain).
3. McFadden D. 1973 Conditional logit analysis of qualitative choice behaviour. In *Frontiers in econometrics* (ed. P Zarembka), pp. 105–142. New York, NY: Academic Press.
4. Baltas G, Doyle P. 2001 Random utility models in marketing research: a survey. *J. Bus. Res.* **51**, 115–125. (doi:10.1016/S0148-2963(99)00058-2)
5. Wilson AG. 1967 A statistical theory of spatial distribution models. *Transp. Res.* **1**, 253–269. (doi:10.1016/0041-1647(67)90035-4)
6. Harris B, Wilson AG. 1978 Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models. *Environ. Plan. A* **10**, 371–388. (doi:10.1068/a100371)
7. Wilson AG. 2000 *Complex spatial systems: the modelling foundations of urban and regional analysis*. London, UK: Pearson Education.
8. Wilson AG. 2011 *Entropy in urban and regional modelling*. London, UK: Routledge.
9. Wilson AG. 2012 *The science of cities and regions: lectures on mathematical model design*. Berlin, Germany: Springer Science & Business Media.
10. Pavliotis GA. 2016 *Stochastic processes and applications*. Berlin, Germany: Springer.
11. Lasota A, Mackey MC. 2013 *Chaos, fractals, and noise: stochastic aspects of dynamics*, vol. 97. Berlin, Germany: Springer Science & Business Media.
12. Kaipio J, Somersalo E. 2006 *Statistical and computational inverse problems*, vol. 160. Berlin, Germany: Springer Science & Business Media.
13. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. 2014 *Bayesian data analysis*, vol. 2. Boca Raton, FL: CRC press.
14. Liu JS. 2008 *Monte Carlo strategies in scientific computing*. Berlin, Germany: Springer Science & Business Media.
15. Murray I, Ghahramani Z, MacKay DJC. 2006 MCMC for doubly-intractable distributions. In *Proc. of the 22nd Annu. Conf. on Uncertainty in Artificial Intelligence (UAI-06)*, Cambridge, MA, 13–16 July 2006, pp. 359–366. AUAI Press.

16. Duncan AB, Lelievre T, Pavliotis G. 2016 Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.* **163**, 457–491. (doi:10.1007/s10955-016-1491-2)
17. Williams HC. 1977 On the formation of travel demand models and economic evaluation measures of user benefit. *Environ. Plan. A* **9**, 285–344. (doi:10.1068/a090285)
18. Small KA, Rosen HS. 1981 Applied welfare economics with discrete choice models. *Econometrica: J. Econ. Soc.* **49**, 105–130. (doi:10.2307/1911129)
19. De Jong G, Daly A, Pieters M, Van der Hoorn T. 2007 The logsum as an evaluation measure: review of the literature and new results. *Transp. Res. A: Policy Pract.* **41**, 874–889. (doi:10.1016/j.tra.2006.10.002)
20. Nielsen F, Sun K. 2016 Guaranteed bounds on information-theoretic measures of univariate mixtures using piecewise log-sum-exp inequalities. *Entropy* **18**, 442. (doi:10.3390/e18120442)
21. Dearden J, Wilson A. 2011 A framework for exploring urban retail discontinuities. *Geogr. Anal.* **43**, 172–187. (doi:10.1111/j.1538-4632.2011.00812.x)
22. McFadden D. 1978 Modeling the choice of residential location. In *Spatial interaction theory and planning models* (eds A Karlqvist, F Snickars, J Weibull), pp. 75–96. Amsterdam, The Netherlands: North-Holland.
23. McFadden D. 1981 Econometric models of probabilistic choice. In *Structural analysis of discrete data with econometric applications* (eds CF Manski, D McFadden), pp. 198–272. Cambridge, MA: MIT Press.
24. McFadden D, Train K. 2000 Mixed MNL models for discrete response. *J. Appl. Econ.* **15**, 447–470.
25. Anas A. 1983 Discrete choice theory, information theory and the multinomial logit and gravity models. *Transp. Res. B: Methodol.* **17**, 13–23. (doi:10.1016/0191-2615(83)90023-1)
26. Butler RW. 2007 *Saddlepoint approximations with applications*, vol. 22. Cambridge, UK: Cambridge University Press.
27. Nocedal J, Wright S. 2006 *Numerical optimization*. Berlin, Germany: Springer Science & Business Media.
28. Lyne AM, Girolami M, Atchadé Y, Strathmann H, Simpson D. 2015 On Russian roulette estimates for Bayesian inference with doubly-intractable likelihoods. *Stat. Sci.* **30**, 443–467. (doi:10.1214/15-STS523)
29. GL Authority. 2009 2009 London Town Centre Health Check Analysis Report. Technical report.
30. GL Authority. 2016 Ward profiles and atlas. See data.london.gov.uk.
31. GL Authority. 2015 Household income estimates for small areas. See data.london.gov.uk.
32. GL Authority. 2014 Statistical GIS boundary files for London. See data.london.gov.uk.
33. Apte A, Jones CK, Stuart A, Voss J. 2008 Data assimilation: mathematical and statistical perspectives. *Int. J. Numer. Methods Fluids* **56**, 1033–1046. (doi:10.1002/flid.1698)
34. Roberts GO, Tweedie RL. 1996 Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, 341–363. (doi:10.2307/3318418)
35. Menz G, Schlichting A. 2014 Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape. *Ann. Probab.* **42**, 1809–1884. (doi:10.1214/14-AOP908)
36. Robert CP. 2004 *Monte Carlo methods*. Wiley Online Library.
37. Neal RM. 2011 MCMC using Hamiltonian dynamics. In *Handbook of Markov chain Monte Carlo*, (eds S Brooks, A Gelman, GL Jones, X-L Meng), vol. 2, pp. 113–162. Boca Raton, FL: Chapman & Hall/CRC.
38. Andrieu C, Roberts GO. 2009 The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Stat.* **37**, 697–725. (doi:10.1214/07-AOS574)
39. McLeish D. 2010 A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods Appl.* **17**, 301–315. (doi:10.1515/mcma.2011.013)
40. Glynn PW, Rhee Ch. 2014 Exact estimation for Markov chain equilibrium expectations. *J. Appl. Probab.* **51**, 377–389. (doi:10.1239/jap/1417528487)
41. Wei C, Murray I. 2017 Markov chain truncation for doubly-intractable inference. *Proc. Mach. Learn. Res.* **54**, 776–784.
42. Neal RM. 2001 Annealed importance sampling. *Stat. Comput.* **11**, 125–139. (doi:10.1023/A:1008923215028)