

Stochastic Modelling of Urban Structure A Bayesian Perspective

Mark Girolami

Department of Mathematics, Imperial College London
The Alan Turing Institute
Lloyds Register Foundation Programme on Data-Centric Engineering

<https://iconicmath.org/>

Imperial College
London



The
Alan Turing
Institute



Bayes on the Beach 2017
14-10-2017



Louis Ellam
ICL & Turing



Mark Girolami
ICL & Turing



Greg Pavliotis
ICL



Sir Alan Wilson
UCL & Turing

Motivation



What are the workings of cities and regions, and how will they evolve over time?

Problem statement

- Over half of the world's population now live in a city.
- We should be interested in:
 - What is happening in the city;
 - How the city is evolving; and
 - How can we enable a better quality of life.
- Planning, policy and decision making (e.g. retail, health, crime, transport etc.).
- Requires an understanding of the underlying mechanisms and behaviours.
- On going task of matching socio-economic theories with empirical evidence.

This talk: Improving insights into urban and regional systems with the development of data-driven mathematical models.

What can mathematics and statistics offer?

- Mathematical models can represent socio-economic theories.
- Can help explain the behaviour of complex systems.
- Simulations may provide insights:
 - 'Flight simulators' for urban and regional planners.
 - 'What if' forecasting capabilities.
 - How does the system respond to change/initiatives?
 - Which parameters/mechanisms are most important?
 - What is the long term behaviour?
- Mathematical modelling long history (> 50 years) in urban and regional analysis
 - *Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models* (Harris and A. Wilson, 1978).

- Data sets are becoming routinely available:
 - Social media,
 - Location tracking,
 - Travel ticketing,
 - Census data,
 - Ad-hoc reports etc.
- The range of statistical models available is arguably somewhat limited.
- Ignoring either data or mathematical models seems unwise.
- A structured approach that's consistent with the available data is desirable.
- Build upon well-established mathematical formalisms with this new found data (rather than throwing the baby out with the bath water).

“All models are wrong; some models are useful.”

- Urban and regional systems are complex in nature.
- An emergent behaviour arises from the actions of many interacting individuals.
- Seamless integration of mathematical models with data.
- Quality of model vs. quality of data?
- Uncertainty should be addressed in the modelling process:
 - Model uncertainty and system fluctuations;
 - Parameter uncertainty; and
 - Observation error and bias.

Example: The London Retail System

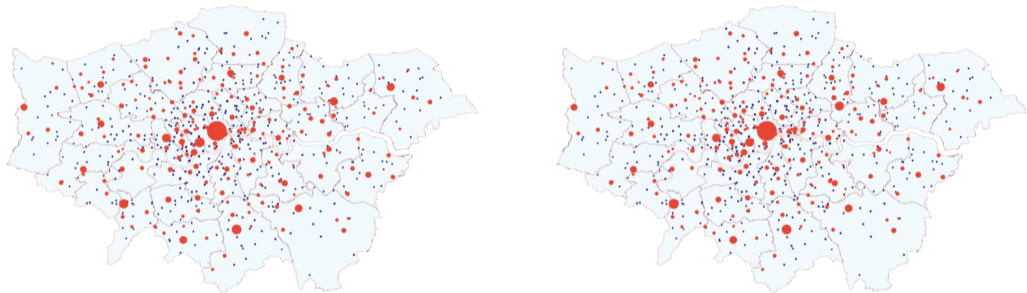
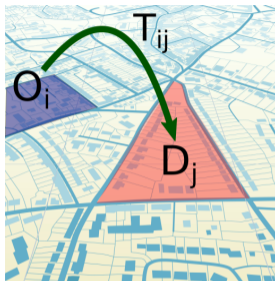


Figure: London retail structure for 2008 (left) and 2012 (right). The locations of retail zones and residential zones are red and blue, respectively. Sizes are in proportion to floorspace and spending power, respectively. $N = 625$ and $M = 201$.

The Forward Problem: Modelling Urban Structure



- Urban structure:
 - N origin zones and M destination zones.
 - Origin quantities $\{O_i\}_{i=1}^N$.
 - Destination quantities $\{D_j\}_{j=1}^M$.
- Spatial interaction:
 - Flows of activities at location are denoted $\{T_{ij}\}_{i,j}^{N,M}$, where T_{ij} is the flow from zone i to j .

Modelling Urban Structure

Urban structure is a vector of sizes:

$$w = \{w_1, \dots, w_N\} \in \mathbb{R}_{>0}^M.$$

Work in terms of attractiveness (log-size):

$$x = \{x_1, \dots, x_M\} \in \mathbb{R}^M, \quad w_j(x_j) = \exp(x_j).$$

Urban structure is realization of the Boltzmann distribution

$$\pi(x) = \frac{1}{Z} \exp\left(-\gamma V(x)\right), \quad Z := \int_{\mathbb{R}^M} \exp\left(-\gamma V(x)\right) dx,$$

specified by the potential $V : \mathbb{R}^M \rightarrow \mathbb{R}$ and 'inverse-temperature' $\gamma > 0$,

Assumptions for the Potential Function

Interpretation of gradient structure (in terms of net supply/demand Π_j):

$$-\partial_j V(x) = \epsilon \Pi_j, \quad j = 1, \dots, M.$$

$$V(x) = \underbrace{V_{\text{Utility}}(x)}_{\text{Demand}} + \underbrace{V_{\text{Cost}}(x) + V_{\text{Additional}}(x)}_{\text{Supply}}.$$

- 1 The potential function $V \in C^2(\mathbb{R}^M, \mathbb{R})$ is confining in that $\lim_{\|x\| \rightarrow +\infty} V(x) = +\infty$, and

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

- 2 The gradient ∇V satisfies the dissipativity condition: $\exists K_1, K_2 > 0$ s.t.

$$\langle x, -\nabla V(x) \rangle \leq K_1 + K_2 \|x\|^2, \quad \forall x \in \mathbb{R}^M.$$

Modelling Spatial Interaction

Flows from origin zones:

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

Flows to destination zones (demand function):

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

Nomenclature: A singly-constrained system since $\{O_i\}_{i=1}^N$ is fixed and $\{D_j\}_{j=1}^M$ is undetermined.

Utility Potential

- For a singly-constrained system we have that:

$$-\partial_j V_{\text{Utility}}(x) = \epsilon \sum_{i=1}^N v_{ij}(x) O_i, \quad \sum_{j=1}^M v_{ij}(x) \equiv 1.$$

- Can express weights v_{ij} in terms of utility functions u_{ij} :

$$-\partial_j V_{\text{Utility}}(x) = \epsilon \sum_{i=1}^N \frac{\varphi(u_{ij})}{\sum_{k=1}^M \varphi(u_{ik})} O_i. \quad (1)$$

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

$$V_{\text{Utility}}(x) = -\epsilon \sum_{i=1}^N O_i \left\{ f_i(x) \ln \sum_{j=1}^M \varphi(u_{ij}(x)) \right\}. \quad (2)$$

- Inserting Eq. (2) into Eq. (1), we obtain the requirement:

$$\frac{\varphi(u_{ij}(x))}{\sum_{k=1}^M \varphi(u_{ik}(x))} = \frac{df_i(x)}{dx_j} \ln \sum_{j=1}^M \varphi(u_{ij}(x)) + f_i(x) \frac{d\varphi(u_{ik}(x))}{dx_j} \left(\sum_{k=1}^M \varphi(u_{ik}(x)) \right)^{-1}. \quad (3)$$

- Eq. (3) is satisfied for $\varphi(x) = \exp(x)$,

$$u_{ij}(x) = \alpha_i x_j + \beta_{ij},$$

provided that each $\alpha_i \neq 0$ and that $f_i = \alpha_i^{-1}$.

We obtain:

$$V_{\text{Inflow}}(x) = -\epsilon \sum_{i=1}^N \left\{ \frac{O_i}{\alpha_i} \ln \sum_{j=1}^M \exp(u_{ij}(x)) \right\}.$$

Cost and Additional Potentials

Cost potential (linear cost):

$$-\partial_j V_{\text{Cost}}(x) = -\epsilon \kappa w_j(x_j), \quad j = 1, \dots, M,$$

$$V_{\text{Cost}}(x) = \epsilon \kappa \sum_{j=1}^M w_j(x_j).$$

Additional potential (constant support):

$$-\partial_j V_{\text{Additional}}(x) = -\epsilon \delta_j, \quad j = 1, \dots, M,$$

$$V_{\text{Additional}}(x) = \epsilon \sum_{j=1}^M \delta_j x_j.$$

End Result (Boltzmann Distribution)

Urban structure is realization of the Boltzmann distribution

$$\pi(x) = \frac{1}{Z} \exp\left(-\gamma V(x)\right), \quad Z := \int_{\mathbb{R}^M} \exp\left(-\gamma V(x)\right) dx,$$
$$\epsilon^{-1} V(x) = \underbrace{-\sum_{i=1}^N \alpha^{-1} O_i \ln \sum_{j=1}^M \exp(\alpha x_j - \beta c_{ij})}_{\text{Utility (Spatial Interaction)}} + \underbrace{\kappa \sum_{j=1}^M w_j(x_j)}_{\text{Cost}} - \underbrace{\sum_{j=1}^M \delta_j x_j}_{\text{Additional}}. \quad (4)$$

End Result (Chosen Utility Function)

The demand flows are

$$D_j = \sum_{i=1}^N O_i \frac{\exp(\alpha x_j - \beta c_{ij})}{\sum_{k=1}^M \exp(\alpha x_j - \beta c_{ik})}.$$

- $x_j := \ln w_j$ is the attractiveness of j .
- c_{ij} is the inconvenience of transporting from zone i to j .
- α is the attractiveness scaling parameter.
- β is the cost scaling parameter.
- $u_{ij}(x_j) = \alpha x_j - \beta c_{ij}$ is the net utility from transporting from zone i to j .

Random Utility Models (Alternative Derivation)

- Individuals make choices to maximize their (random) utility:

$$U_{ij} = u_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \text{Gumbel}(0, 1).$$

- The probability that individual from zone i makes choice j is:

$$P[Y_{ij} = 1] = P[\cap_{j \neq k} \{U_{ij} > U_{ik}\}] = \frac{\exp(u_{ij})}{\sum_{k=1}^M \exp(u_{ik})}.$$

- The expectation for all individuals gives the same demand flows as before:

$$D_j = \sum_{i=1}^N O_i P[Y_j = 1] = \sum_{i=1}^N O_i \frac{\exp(\alpha x_j - \beta c_{ij})}{\sum_{k=1}^M \exp(\alpha x_j - \beta c_{ik})}.$$

- The utility potential is the (unscaled) expected welfare:

$$\epsilon^{-1} V_{\text{Utility}}(x) = \alpha^{-1} \sum_{i=1}^N O_i \mathbb{E}[\max_j U_{ij}] + \text{const.}$$

Maximum Entropy Method (Alternative Derivation)

- Total welfare (consumer surplus):

$$\mathbb{E}_\pi \left[\alpha^{-1} \sum_{i=1}^N O_i \ln \sum_{j=1}^M \exp(u_{ij}(x_j)) \right] = S, \quad (5)$$

- Total size (capacity):

$$\mathbb{E}_\pi \left[\sum_{j=1}^M w_j(x_j) \right] = W, \quad (6)$$

- Expected attractiveness (benefit)

$$\mathbb{E}_\pi \left[\sum_{j=1}^M x_j \right] = X. \quad (7)$$

- Then the Boltzmann distribution $\pi(x) = Z^{-1} \exp(-\gamma V(x))$ is the maximum entropy distribution (maximal uncertainty) subject to these constraints.

Airports in England

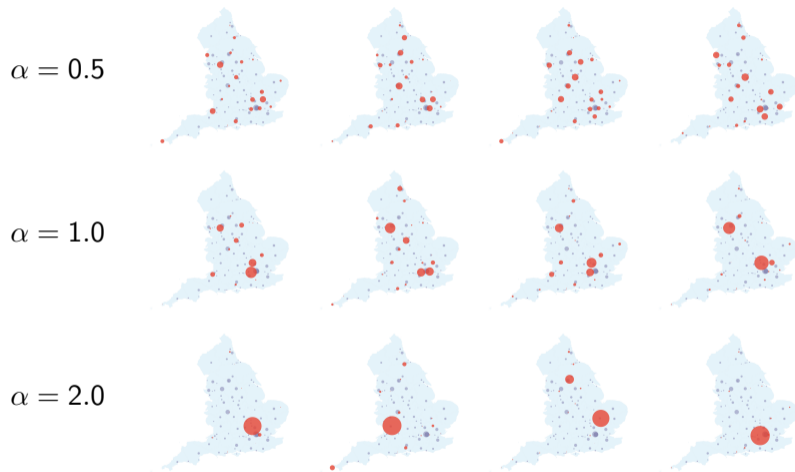


Figure: Approximate draws from $p(x|\theta)$ using HMC combined with parallel tempering.

The London Retail System

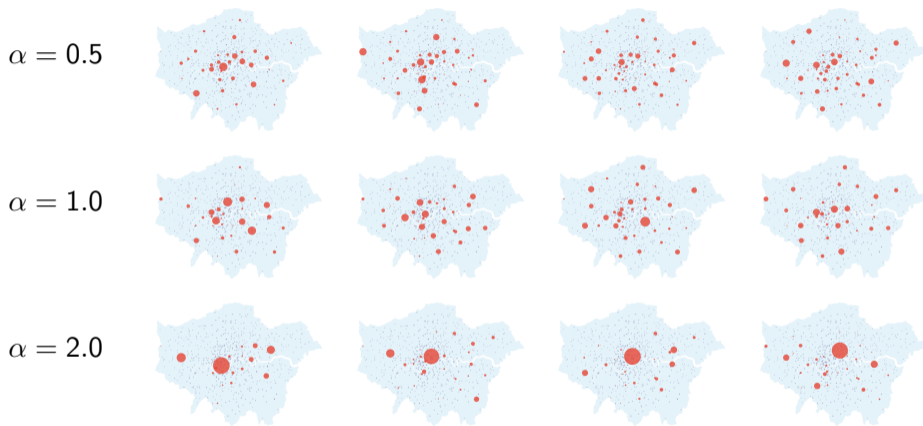


Figure: Approximate draws from $p(x|\theta)$ using HMC combined with parallel tempering.

Overdamped Langevin Dynamics

- With the specification of $V(X)$, overdamped Langevin dynamics give the Harris and Wilson model, plus multiplicative noise.

SDE Urban Retail Model

Floorspace dynamics is a stochastic process that satisfies the following Stratonovich SDE.

$$\frac{dW_j}{dt} = \epsilon W_j \left(D_j - \kappa W_j + \delta_j \right) + \sigma W_j \circ \frac{dB_j}{dt},$$

where $(B_1, \dots, B_M)^T$ is standard M -dimensional Brownian motion and $\sigma = \sqrt{2/\epsilon\gamma} > 0$ is the volatility parameter.

- Fluctuations (missing dynamics) are modelled as Stratonovich noise.
- The extra parameter δ to represents local economic stimulus to prevent zones from collapsing (needed for stability).
- The Markov process is well-defined and converges geometrically fast to $\pi(x) = Z^{-1} \exp(-\gamma V(x))$.

The Inverse Problem: Inferring the Utility Function

A Hierarchical Model of Urban Structure

Given observation data $Y = (Y_1, \dots, Y_M)$, of log-sizes, infer the parameter values $\theta = (\alpha, \beta)^T \in \mathbb{R}_+^2$ and corresponding latent variables $X \in \mathbb{R}^M$.

Assumption (Data generating process)

Assume that each observation Y_1, \dots, Y_M is an independent and identical realization of the following hierarchical model:

$$Y_1, \dots, Y_M | X, \sigma \sim \mathcal{N}(X, \sigma^2 I),$$

$$X | \theta \sim \pi(X | \theta) \propto \exp(-U(X; \theta)),$$

$$\theta \sim \pi(\theta).$$

Joint Posterior Distribution

The joint posterior is given by

$$\pi(X, \theta | Y) = \frac{\pi(Y|X, \theta)\pi(X, \theta)}{\pi(Y)}, \quad \pi(Y) = \int \pi(Y|X, \theta)\pi(X, \theta)dXd\theta.$$

We have a hierarchical prior given by

$$\pi(X, \theta) = \frac{\pi(\theta) \exp(-U(X; \theta))}{Z(\theta)}, \quad Z(\theta) = \int \exp(-U(X; \theta))dX.$$

The joint posterior is doubly-intractable

$$\pi(X, \theta | Y) = \frac{\pi(Y|X, \theta) \exp(-U(X; \theta))\pi(\theta)}{\pi(Y)Z(\theta)}.$$

Russian Roulette

Figure: On Russian Roulette Estimates for Bayesian Inference with Doubly-Intractable Likelihoods. AM Lyne, M Girolami, Y Atchade, H Strathmann, D Simpson. *Statistical Science*, 30 (4), 443-467



Metropolis-within-Gibbs with Block Updates

- We are interested in low-order summary statistics of the form

$$\mathbb{E}_{X, \theta | Y} [h(X, \theta)] = \int h(X, \theta) \pi(X, \theta | Y) dX d\theta.$$

- X and θ are highly coupled, so we use Metropolis-within-Gibbs with block updates.
- X -update. Propose $X' \sim Q_X$ and accept with probability

$$\min \left\{ 1, \frac{\pi(Y|X', \theta) \exp(-U(X'; \theta)) q(X|X')}{\pi(Y|X, \theta) \exp(-U(X; \theta)) q(X'|X)} \right\}.$$

- θ -update. Propose $\theta' \sim Q_\theta$ and accept with probability

$$\min \left\{ 1, \frac{\pi(Y|X, \theta') Z(\theta) \exp(-U(X; \theta')) q(\theta|\theta')}{\pi(Y|X, \theta) Z(\theta') \exp(-U(X; \theta)) q(\theta'|\theta)} \right\}.$$

- Unfortunately the ratio $Z(\theta)/Z(\theta')$ ratio is intractable!

Unbiased Estimates of the Partition Function

- We can use the Pseudo-Marginal MCMC framework if we have an unbiased, positive estimate of $\pi(X|\theta)$, denoted $\hat{\pi}(X|\theta, u)$, satisfying

$$\pi(X|\theta) = \int \hat{\pi}(X|\theta, u)\pi(u|\theta)du.$$

- The Forward Coupling estimator (FCE)¹ gives an unbiased estimate of $1/Z$:

$$\mathbb{E}[S] = 1/Z.$$

- The idea is to find two sequences of consistent estimators $\{\mathcal{V}^{(i)}\}$, $\{\tilde{\mathcal{V}}^{(i)}\}$, each with the same distribution, such that $\mathcal{V}^{(i)}$ and $\tilde{\mathcal{V}}^{(i-1)}$ are “coupled”.

¹Markov Chain Truncation for Doubly-Intractable Inference, C. Wei, and I. Murray, AISTATS (2016)

Unbiased Estimates of the Partition Function

- Requires N estimates of $1/Z$ using path sampling e.g. annealed importance sampling or thermodynamic integration, for a random stopping time N .
- Coupling between $\mathcal{V}^{(i)}$ and $\tilde{\mathcal{V}}^{(i-1)}$ is introduced with a Markov chain that shares random numbers.
- Variance reduction technique.
- Then the unbiased estimate is given by

$$S := \mathcal{V}^{(0)} + \sum_{i=1}^N \frac{\mathcal{V}^{(i)} - \tilde{\mathcal{V}}^{(i-1)}}{\Pr(N \geq i)}.$$

The Signed Measure Problem

- S can be negative when $\mathcal{V}^{(i)} < \tilde{\mathcal{V}}^{(i-1)}$ for many i . This is known as the ‘sign problem’.
- Rejecting when S is negative would introduce a bias.
- Instead, we note that

$$\begin{aligned}\mathbb{E}[h(X, \theta)] &= \frac{1}{\pi(Y)} \int h(X, \theta) \pi(Y|X, \theta) \hat{\pi}(X|\theta, u) \pi(\theta) \pi(u) du d\theta dX, \\ &= \frac{\int h(X, \theta) \sigma(X|\theta, u) \check{\pi}(X, \theta, u|Y) du d\theta dX}{\int \sigma(X|\theta, u) \check{\pi}(X, \theta, u|Y) du d\theta dX},\end{aligned}$$

where σ is the sign function and we have defined

$$\check{\pi}(X, \theta, u|Y) = \frac{\pi(Y|X, \theta) |\hat{\pi}(X|\theta, u)| \pi(\theta) \pi(u)}{\int \pi(Y|X, \theta) |\hat{\pi}(X|\theta, u)| \pi(\theta) \pi(u) du d\theta dX}.$$

Pseudo-Marginal Markov Chain

- We can sample from $\tilde{\pi}(X, \theta, u|Y)$ using Metropolis-within-Gibbs with block updates.
- X -update. Propose $X' \sim Q_X$ and accept with probability

$$\min \left\{ 1, \frac{\pi(Y|X', \theta) \exp(-U(X'; \theta)) q(X|X')}{\pi(Y|X, \theta) \exp(-U(X; \theta)) q(X'|X)} \right\}.$$

- (θ, u) -update. Propose $(\theta', u') \sim Q_{\theta, u}$ and accept with probability

$$\min \left\{ 1, \frac{\pi(Y|X, \theta') |S(\theta, u)| \exp(-U(X; \theta')) \pi(\theta') \pi(u') q(\theta|\theta')}{\pi(Y|X, \theta) |S(\theta', u')| \exp(-U(X; \theta)) \pi(\theta) \pi(u) q(\theta'|\theta)} \right\}.$$

- Posterior expectations are estimated using

$$\mathbb{E}_{X, \theta|Y}[h(X, \theta)] = \lim_{N \rightarrow \infty} \frac{\sum_{i=1}^N h(X_i, \theta_i) \sigma(X_i|\theta_i, u_i)}{\sum_{i=1}^N \sigma(X_i|\theta_i, u_i)}.$$

Airports in England

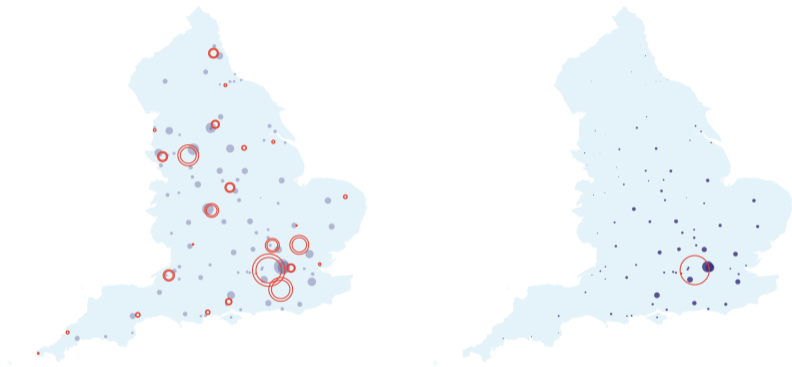


Figure: Left: Visualization of the posterior-marginals for the latent variables $\{x_j\}_{j=1}^M$ over a map of England. Right: Visualization of the posterior demands.

The London Retail System

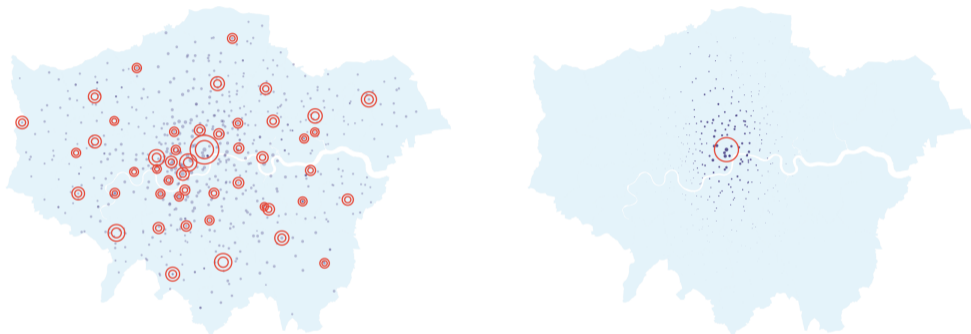


Figure: Left: Visualization of the posterior-marginals for the latent variables $\{x_j\}_{j=1}^M$ over a map of London Right: Visualization of the posterior demands. Observation data obtained

Conclusions and Outlook

Further work and extensions

- Investigation of new data assimilation methodologies to calibrate models to data available at different scales. For example:
 - Population data;
 - Cost matrix; or
 - Time dependent parameters.
- Deployment of new methodology to a global problem to provide new insights into urban retail structure.
- Extension of discrete-choice approach to other socio-economic phenomena e.g. crime.
- Melding of data and models takes us beyond data analytics.

- We have developed a novel stochastic model to simulate realistic configurations of urban and regional structure.
- Our model is an improvement on existing deterministic models in the literature, as we account for uncertainties arising in the modelling process.
- We presented a Bayesian hierarchical model for urban and regional systems.
- Our model can be used to infer the components of a utility function from observed structure, rather than flow data.
- We have demonstrated our approach using an example of airports in England and retail in London.

Acknowledgment

- EPSRC Programme Grant - EP/P020720/1
- Inference, Computation, and Numerics for Insights into Cities - ICONIC
- <https://iconicmath.org/>

