Computers, Environment and Urban Systems 128 (2026) 102455

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

TransGM: Transferable gravity models for cross-city policy transfer

Adham Enaya, Chen Zhong [*], Michael Batty, Robin Morphet, Fulvio D. Lopane

Centre for Advanced Spatial Analysis, University College London, United Kingdom

ARTICLE INFO

Keywords:
Transfer learning
Policy transfer
Spatial interaction model

ABSTRACT

Urban policy transfer has the potential to enhance urban systems by applying successful strategies from one city to another. However, existing models for predicting policy impacts are often context-specific and lack generalisability across different urban settings. This study introduces TransGM, a novel framework that enables the transfer of spatial interaction models between cities via adaptive transfer learning. The framework employs: (1) spatial Kullback–Leibler divergence to quantify structural differences across urban contexts; (2) production-constrained gravity models for flow prediction, accounting for residues in urban features; and (3) feature-specific regularisation weights that adapt parameters based on the degree of spatial pattern similarity between source and target cities. TransGM is demonstrated through a case study of workplace attractiveness and its impact on commuting flows transferred from Birmingham to Coventry. The adapted transfer model replicates Birmingham's urban configuration in Coventry rather than learning Coventry's own priority structure. Transfer process is jointly governed by spatial divergence between the two cities and the data sufficiency: where spatial patterns align, local data dominate parameter estimation, while in data-sparse contexts, source-city regularisation guides model behaviour. This balance between place-specific urban structure and cross-city mobility and amenity transferability positions adaptive transfer learning as an effective tool for evaluating the feasibility of replicating one city's development model in another, offering simulation-based evidence to inform urban planning and policy decisions.

1. Introduction

Urban policy transfer refers to the process through which specific knowledge from one urban area is adopted in another, aiming to replicate its impact while often requiring adjustments to suit different local contexts (Dolowitz & Marsh, 2000). Urban mobility and transport planning are pivotal fields for policy transfer due to the urgent demand for sustainable solutions in rapidly growing cities, such as implementing Bus Rapid Transit (BRT) systems to reduce congestion and improve environmental outcomes (Thomas et al., 2018). The adaptation of Vancouver's sustainable urbanism model in Abu Dhabi was examined by Khirfan and Jaffer (2014), who highlighted the importance of context, authority, and iterative adaptation. Another example is provided by Canitez (2020), who studied the transfer of sustainable mobility policies from London to Istanbul, demonstrating how differences in local governance structures influenced the results.

While these studies demonstrate the importance of contextual adaptation, the policy transfer literature has largely relied on qualitative assessment to determine when and how policies should be adjusted. Urban modelling offers the potential to place these judgments on a more systematic, evidence-based footing (Batty, 2010; Lopane et al., 2023; Wilson, 2018). Yet current approaches have limitations that constrain this potential, as they calibrate parameters using single-city data, treating coefficients as entirely static and assuming universality across cities. These models are rooted in fundamental assumptions of underlying human behaviour, such as ‘utility maximisation’; for instance, individuals choose the travel option that offers the highest perceived benefit (e.g., shortest time). While the general mathematical form of this assumption can hold across different contexts, the specific model coefficients, which capture the strength of variable effects on behaviour, often do not (Hansen, 1981). Influenced by differences in socio-economic conditions and land-use configurations, the tendency to travel long distances for perceived opportunities can vary significantly across cities. Existing modelling processes lack systematic mechanisms to measure urban structural differences relevant to mobility modelling and to identify which urban features are transferable versus context dependent. This limits the potential for leveraging knowledge from data-rich cities while maintaining model performance in data-scarce

---

[*] Corresponding author at: University College London, Gower Street, WC1E 6BT, United Kingdom.
E-mail address: c.zhong@ucl.ac.uk (C. Zhong).

https://doi.org/10.1016/j.compenvurbsys.2026.102455
Received 12 August 2025; Received in revised form 24 April 2026; Accepted 4 May 2026

contexts.

This challenge is particularly pressing in many developing urban contexts, where the scarcity of flow and mobility data may necessitate borrowing models from other cities, rather than building one from scratch. However, doing so without accounting for structural differences risks poor performance or misleading policy insights. Similar analogies have been recognised in the field of machine learning, where models developed on a single dataset often struggle to generalise across different datasets (Feng et al., 2024; Gong et al., 2025), necessitating new data from the target domain. Crucially, such data can complement, rather than replace, the existing model by tuning parameters to the new domain's features thereby reducing data collection costs. These approaches are known as transfer learning techniques, which capture generic features across domains while tuning domain-specific features to a new domain through partial retraining of the same model structure (Jiang et al., 2022).

This paper proposes a novel methodological framework for an adaptive transfer of gravity models for flow prediction (TransGM), which employs transfer learning techniques in gravity models to support urban policy transfer applications. TransGM compromises three components: spatial and distributional difference measurement, baseline gravity model calibration, and divergence-weighted parameter tuning. We developed our methodology in line with multi-domain learning theory, which posits that greater similarity between source and target domains leads to lower generalisation error when transferring models (Ben-David et al., 2010). To operationalise these theoretical insights in the context of urban studies, it is necessary to explicitly measure the differences in urban systems and analyse their effects on model parameters. Specifically, measuring how differently amenities relate to employment patterns across cities allows us to adjust models of mobility from one city before applying it to another, determining which parameters require local recalibration and by how much.

TransGM achieves acceptable prediction performance; however, it is primarily conceived as a diagnostic tool. Our approach is a hybrid between the simple form of the gravity model and the transfer learning methods borrowed from deep learning, enabling transparency that is essential for planning deliberation, particularly in data-scarce contexts where it is hard for policymakers to anticipate the impact of new interventions. TransGM addresses this limitation by leveraging little data from the target city to fine-tune behavioural parameters borrowed from structurally similar, well-surveyed cities. This similarity is measured using freely available urban data such as Point-of-Interest (POI) data from OpenStreetMap to capture differences in land use.

We demonstrate this framework using Birmingham and Coventry, two UK cities with measurable differences in economic and spatial form. While the empirical scope is focused within a single national context, the methodological contributions extend beyond this case. The framework provides a generalisable approach for quantifying structural transferability requirements that can be applied to other city pairs, particularly when assessing whether development strategies from one urban context can be replicated in another.

2. Literature review

2.1. Urban policy transfer: The need for quantitative transferability assessment

Urban policy transfer (UPT) involves adapting successful policies from one city to another and is typically understood in three forms: policy transfer, which refers to the conscious adoption of policies; policy diffusion, the spontaneous spread of policies across cities; and policy mobility, the fluid circulation of urban ideas and models (Canitez, 2020; Glaser et al., 2022; Marsden and Stead, 2011a). In the case of policy transfer, policymakers and planners actively consider external models. Examples include transferring congestion charging policies from London to Istanbul (Canitez, 2020) and adopting Bus Rapid Transit systems from Bogotá in other Latin American cities (Dolowitz & Marsh, 2000; Thomas et al., 2018; Wood, 2014). These cases often inspired the development of conceptual frameworks to analyse the process, focusing primarily on the actors involved in the transfer.

However, to ensure effective policy outcomes, UPT research advocates for methodological foundations that allow for the generalisation of findings beyond idiosyncratic case studies (Marsden and Stead, 2011b). This demands rigorous analysis and testing of policy features against the specificities of the recipient context. The standard approach often compares cities that share similar levels of development; however, quantitatively capturing dissimilarity remains more challenging. Integrating technological and modelling approaches provides a significant methodological advance for measuring such divergence when simulating and monitoring policy implementation (Glaser et al., 2022). This integration supports empirically driven decision-making by rigorously accounting for local context and analysing how new policies may interact with existing urban systems once implemented.

2.2. Spatial interaction models and transferability constraints

Spatial Interaction Modelling (SIM) has long been employed to simulate policy outcomes under alternative scenarios. SIM tools, most notably the gravity model, are widely used to predict flows between urban zones and enable rapid analysis of UPT by accounting for factors such as population and travel cost (Kwon et al., 2023; Wilson, 1971). In the context of UK cities, Batty and Milton (2021) applied such models through the QUANT project to test scenarios assessing the impacts of transport infrastructure on the distribution of housing and employment. Owing to their simple structure and computational efficiency, the gravity model remains a popular and practical tool.

Despite these advantages, gravity models exhibit notable limitations. A core weakness lies in their limited contextual sensitivity, arising from the use of static regression coefficients (Ludwig et al., 2023; Sikder et al., 2013) and the assumption that interaction intensity decreases uniformly with distance. This simplification overlooks the complexity and heterogeneity of human travel behaviour (Capoani, 2023). Contemporary research in economic geography and policy analysis increasingly rejects the notion of universal spatial laws governing interaction patterns (Bathelt & Glückler, 2003). In response, Kim, Park, and Lee (Kim et al., 2018) extended the gravity model by incorporating additional variables to reflect heterogeneity in urban form. However, the influence of these variables remains highly context-dependent and is unlikely to be fully captured by a single, universal representation of travel behaviour. As a result, such models tend to perform best in the regions for which they were originally calibrated (Ludwig et al., 2023).

Addressing these limitations requires modelling approaches that recognise cities as dynamic systems and adopt a logic of situated knowledge. Although numerous empirical studies have attempted to improve model transferability through direct parameter adjustment, these efforts have often paid limited attention to urban structural and behavioural variability across cities. Neglecting them frequently leads to degraded performance and lack of contextual fit when models are transferred (Sikder et al., 2013). Seminal work by Ben-Akiva et al. (Ben-Akiva, 1979; Ben-Akiva et al., 1995; Ben-Akiva & Bolduc, 1987) introduced the concept of spatial transferability in travel demand modelling, typically operationalised through coefficient updating via averaging, scaling, or pooling data from source and target contexts (Sikder et al., 2013). While valuable, these approaches generally assume a broadly uniform coefficient update and would benefit from integration with systematic comparative frameworks that explicitly account for context differences.

2.3. Transfer learning and information-theoretic approaches to model fine-tuning

Recent advances in transfer learning have supported the transferability of spatial models (Kouw & Loog, 2019), driven in part by the growing availability of large-scale geospatial datasets (Klinkhardt et al., 2021; Zhong, 2024). Deep learning techniques, in particular, enable the encoding of spatial structure directly into model calibration and support multi-stage adaptation across geographical contexts (Chelba & Acero, 2006; Gong et al., 2025). For instance, Yeghikyan et al. (2020) employed neural network architectures to predict traffic flows using modular designs that explicitly distinguish between local and global spatial features (see also Jiang et al., 2022; Kouw & Loog, 2019). Comparable approaches have been applied to the prediction of commuting flows between urban areas, demonstrating strong predictive performance across diverse spatial systems (Murakami et al., 2025; Song et al., 2025).

Despite these advances, deep learning–based spatial models often operate as black boxes, providing limited insight into how explanatory variables shape model outputs (Luca et al., 2021). This lack of transparency constrains their applicability in policy-making contexts, where understanding the relationship between land-use patterns and travel demand is essential. Explainable artificial intelligence (XAI) techniques, such as SHAP, offer partial interpretability by attributing importance to individual features (Molnar, 2019; Simini et al., 2021). However, the assumptions underlying these methods can remain opaque to non-technical stakeholders, limiting their effectiveness as policy communication tools.

Information-theoretic approaches can provide a complementary and more interpretable alternative for analysing model for urban policy applications. By representing variable distributions within their original feature space, these methods allow for the quantification of similarity or divergence between distributions (Lye et al., 2024). Metrics such as Kullback–Leibler divergence (Kullback & Leibler, 1951) and Rao's divergence (Rao & Nayak, 1985), along with extensions such as Batty's (Batty, 1974) spatial entropy for characterising urban structures, enable differences in spatial configurations to be quantified beyond simple summary statistics. In the context of policy transfer, these approaches provide transparent indicators of statistical coherence between source and target domains, facilitating the assessment of information flow during transfer processes (Schreiber, 2000).

2.4. *Workplace attractiveness and urban context: Birmingham–Coventry example*

The methodological challenges outlined above become concrete when exemplified by a case study of Birmingham and Coventry, which are neighbouring cities with markedly different economic structures. Workplace attractiveness offers a useful lens for examining these differences as workplace location attractiveness is a key consideration for both workers and investors (Overman & Xu, 2024). The immediate neighbourhood and its surrounding amenities are important determinants of workplace attractiveness (Tufts, 2003). Mixed-use environments that support daily activities positively influence office worker satisfaction, and employer priorities have increasingly shifted toward locations offering access to nearby services such as healthcare, dining, gyms, and well-being facilities, reflecting the growing importance of attracting and retaining talent (Ewing et al., 2011). By analysing the configuration of activities and neighbourhood vibrancy, attractiveness can be estimated and linked to employment patterns.

Evidence from the West Midlands Combined Authority (2025) indicates that office take-up in city centres such as Birmingham and parts of Coventry is strongly influenced by access to essential services and lifestyle amenities. High-value clusters, such as Birmingham's Snow Hill district (which hosts major professional services, e.g., law and accountancy firms), reinforce the city's role as a financial and service hub by creating high-quality employment environments. In contrast, specialised locations focused on advanced manufacturing and logistics require distinct site conditions, including high-power availability, logistics connectivity, and access to business parks. Data from the Annual Population Survey (Office for National Statistics, 2025) supports this divergence: Birmingham has a higher share of employment in professional and financial services, reflecting its service-oriented economy, whereas Coventry has a much larger share of manufacturing employment, highlighting its continued industrial base. This functional divergence underscores the challenges of applying universal models in urban policy analysis.

Building on this literature, urban policy transfer requires both an understanding of contextual differences and tools to quantify and adapt models across cities. Spatial interaction models, including gravity-based approaches, provide a basis for simulating policy outcomes, but their predictive power is constrained by static parameters and context-specific calibration. Advances in transfer learning offer ways to address these limitations by systematically comparing urban features and identifying transferable patterns. There is therefore a need for a framework that not only predicts flows but also evaluates the transferability of policy models across urban contexts. The following section introduces TransGM, a framework that operationalises these insights by quantifying structural similarities.

3. Methodology

As illustrated in Fig. 1, the framework consists of four modules: **(1) Feature Engineering:** amenity features (e.g., retail, transport) are residualised with respect to employment to isolate destination attractiveness beyond job concentration. Overall attractiveness combines employment elasticity with residualised amenity effects. **(2) Spatial Signature:** is computed within-city Kullback–Leibler (KL) divergence between employment and amenity distributions for both source and target cities. The absolute inter-city difference between these signatures is converted into an adaptive weight capturing structural similarity. **(3) Model Formation:** a production-constrained gravity model is trained on the source city using observed flows. Training uses a multitask objective combining Poisson deviance for OD flow accuracy with an inflow-alignment loss for destination totals. **(4) Adaptive Transfer Learning:** source parameters are transferred to the target city via divergence-weighted regularisation. The adaptive weight modulates the regularisation strength, anchoring parameters where spatial structures align and allowing adaptation where they diverge. The target model is then optimised using target features under this transfer constraint.

3.1. *Feature engineering*

3.1.1. *Attractiveness residualisation*

Classic gravity models use job counts as attractiveness factor for workplace locations (Wilson, 1971). We aim to capture more context-based features that contribute to attractiveness, particularly urban amenities (e.g., retail, services, transport). Drawing an analogy with Batty and March's (1976) method of residues in urban modelling, we construct attractiveness as a function of job counts as the primary driver and amenities as the secondary driver (see Section 3.3.2). However, the spatial distributions of urban amenities and jobs are often highly correlated due to agglomeration effects, and using them directly in the model risks multicollinearity and may amplify the influence of variables (e.g., job density). We therefore adopt a residue-based approach, which captures amenity provision beyond what is expected from employment, potentially reflecting strategic or community preferences, and enabling city-specific amenity-job spatial configurations. This residue-based measure is included as the secondary driver of workplace attractiveness. For each amenity type k, we estimate the relationship between jobs and the amenities across zones j, as shown in Eq. (1) (see Appendix A for the feature correlation test):

$$\mathrm{Amenity}_{jk} = \alpha_k \cdot \mathrm{Jobs}_j + \varepsilon_{jk} \tag{1}$$

and then we compute the residualisation of features $\boldsymbol{\xi}_{jk}$ using (García

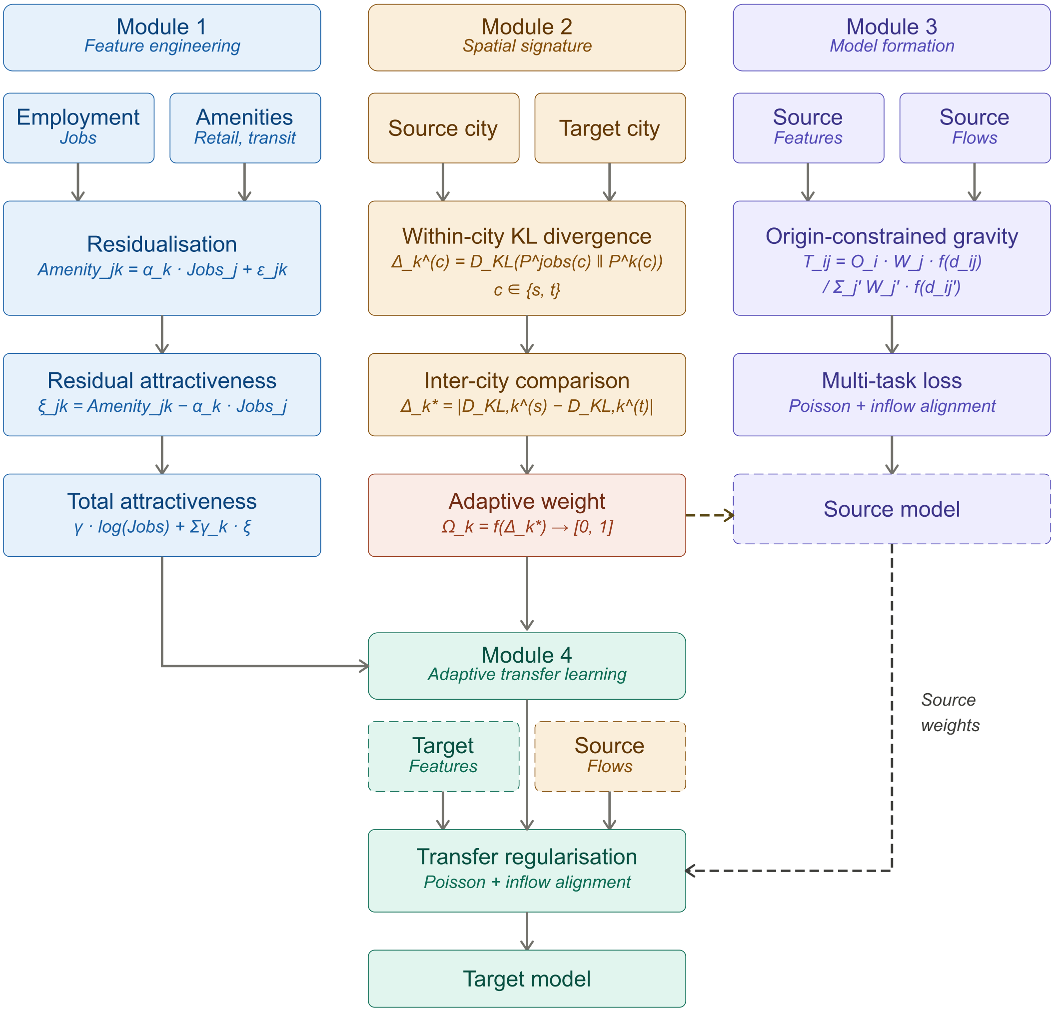


Fig. 1. The TransGM framework for transferring spatial interaction models between cities.

et al. (2019) to isolate the effect of jobs signal as in Eq. (2):

$$\xi_{jk} = \text{Amenity}_{jk} - \alpha_k \cdot \text{Jobs}_j \tag{2}$$

3.1.2. *Spatial distribution feature normalisation*

We then compute the spatial distribution of employment and residualised features within both the source (s) and target (t) cities separately, as shown in Eq. (3), residuals are truncated such that $\xi_j k^{(c)} \leftarrow \max\left(0, \xi_j k^{(c)}\right)$ prior to normalisation:

$$P_j^{\text{jobs}(c)} = \frac{\text{Jobs}_j^{(c)}}{\sum_j \text{Jobs}_j^{(c)}}, P_j^{\xi_k(c)} = \frac{\xi_{jk}^{(c)}}{\sum_j \xi_{jk}^{(c)}}, c \in \{s, t\} \tag{3}$$

3.2. *Spatial signature extraction*

To quantify the structural similarity between cities for adaptive transfer learning, we introduce a measure based on within-city land-use patterns. The key insight is that cities exhibiting a similar spatial organisation of economic activities, such as the co-location patterns of employment and amenities, should enable more robust knowledge transfer. While a structural signature refers to mathematical, feature-based representations of data, here, we characterise the internal structural signature of a city, ensuring that the transfer process is grounded in the functional spatial co-location patterns between employment and amenities.

3.2.1. *Within-city structural signature using spatial KL*

We characterise each city's internal spatial structure by measuring how employment and residualised amenities co-locate across destinations. To account for the spatial dependency between these variables, we compute the Kullback-Leibler divergence between the spatial distribution of jobs and each residualised amenities (ξ_k), as defined in Eq. (4):

$$\Delta_k^{(c)} = D_{KL}\left(P^{\text{jobs}(c)} \parallel P^{\xi_k(c)}\right) = \sum_j P_j^{\text{jobs}(c)} \log\left(\frac{P_j^{\text{jobs}(c)}}{P_j^{\xi_k(c)}}\right) \tag{4}$$

Where $\Delta_k^{(c)}$ measures the spatial divergence of residualised amenity k from the employment distribution within city c, interpreted as:

- **Low**: jobs and residualised amenities are highly co-located within the same zones, indicating a deeply integrated mixed-use urban environment.
- **Moderate**: characterised by moderately mixed activities with some specialised zoning.
- **High**: indicates strong zoning and clear spatial segregation of activities, such as isolated office parks versus disconnected retail malls.

Standard KL is sufficient when comparing feature distributions across identical spatial structures and uniform zone sizes. However, to enable comparisons across disparate regions or variable administrative boundaries, we must account for zone area by incorporating a spatial density correction into the divergence calculation by using including the spatial entropy (Batty, 1974), as shown in Eq. (5):

$$\Delta_k^{(c)} = \sum_j \mathrm{P}_j^{\mathrm{jobs}(c)} \ln\left(\begin{cases} \mathrm{P}_j^{\mathrm{jobs}(c)} \Big/ \mathrm{P}_j^{\xi_k(c)} & \text{uniform zones} \\ \left(\mathrm{P}_j^{\mathrm{jobs}(c)} \cdot \mathrm{A}_j^{\mathrm{jobs}(c)}\right) \Big/ \left(\mathrm{P}_j^{\xi_k(c)} \cdot \mathrm{A}_j^{\xi_k(c)}\right) & \text{variable zones} \end{cases}\right) \tag{5}$$

where:

- $\mathrm{P}_j^{\mathrm{jobs}(c)}$: The probability distribution of employment at destination j in city c.
- $\mathrm{P}_j^{\xi_k(c)}$: The probability distribution of the k-th amenity at destination j in city c.
- A_j: The physical area of zone j.

3.2.2. *Inter-city similarity metric*

Once each city has a structural signature, we measure similarity between source and target cities by computing the absolute difference as in Eq. (6):

$$\Delta_k^{*} = \left|\mathrm{D}_{\mathrm{KL},k}^{(s)} - \mathrm{D}_{\mathrm{KL},k}^{(t)}\right| \tag{6}$$

We interpret Δ_k^{*} denotes the inter-city structural difference for amenity k:

- **Small**: similar land-use organisation, e.g., both are moderately mixed-use cities.
- **Medium**: moderately different structures, e.g., a mixed-use city versus a single-use zones.
- **Large**: fundamentally different urban organisation, e.g., a compact mixed-use city versus a sprawling segregated suburban form.

3.3. *Model formation*

3.3.1. *Production-constrained gravity model*

We develop a production-constrained gravity model, denoted as TransGM, to predict trip flows T_{ij} between origin i and destination j. The model is architected to ensure the sum of all predicted outflows from a specific origin must exactly match the observed total outflows O_i for that location. The model follows the general functional form in Eq. (7):

$$\widehat{T}_{ij} = O_i \frac{W_j \cdot f(d_{ij})}{\sum_{j'} W_{j'} \cdot f(d_{ij'})} \tag{7}$$

where:

- $\widehat{\mathrm{T}}_{ij}$: Predicted flow from origin i to destination j.
- O_i :Observed total outflow from origin i.
- W_j: Attractiveness of destination j.
- $\mathrm{f}(\mathrm{d}_{ij})$: Distance decay function representing spatial friction.
- The production constraint ensuring conservation of flow.

By construction, this formulation satisfies the production constraint $\sum_j \widehat{\mathbf{T}}_{ij} = \mathbf{O}_i$, ensuring conservation of flow from each origin.

3.3.2. *Destination attractiveness*

Unlike traditional gravity models that rely on simple mass terms (e.g., population), our model learns a flexible, data-driven representation of destination attractiveness. By decomposing attractiveness into observed employment data and amenity residuals, we capture a more nuanced pull factor for each destination, as defined in Eq. (8):

$$log(W_j) = \gamma_{jobs} log(Jobs_j) + tanh\left(\sum_k \gamma_k log(\xi_{jk})\right) \tag{8}$$

where:

- **Jobs Component**: Captures the primary attractiveness derived from employment opportunities, where γ_{jobs} is the learned jobs elasticity parameter.
- **Residual Attractiveness**: We residualised amenity (ξ_{jk}) to capture context-specific amenity configurations, see Section 3.1.1.
- γ_k: Weights for residualised amenities.
- $\tanh(\cdot)$: Bounds the residual contribution to prevent overwhelming the primary jobs signal.

3.3.3. *Distance decay functions*

We adopt two functional forms to represent the friction of distance: Power Law Decay in Eq. (9) and using the Exponential Decay in Eq. (10):

$$f(d_{ij}) = (d_{ij})^{-\beta} \tag{9}$$

$$f(d_{ij}) = exp(-\beta d_{ij}) \tag{10}$$

where β is the decay parameter.

3.3.4. *Optimisation and gradient estimation*

To ensure numerical stability and efficiency, all components are implemented in log-space as shown in Eq. (11), which prevents underflow in large-scale spatial networks and enables an additive optimisation form. We utilise the L-BFGS-B algorithm (Zhu et al., 1997) for parameter estimation, with detailed settings provided in Appendix B. Gradients are computed via automatic differentiation using the PyTorch autograd engine, enabling precise estimation across the non-linear TransGM architecture:

$$\log(\widehat{\mathrm{T}}_{ij}) = \log(\mathrm{O}_i) + \log(\mathrm{W}_j) + \log(\mathrm{f}(\mathrm{d}_{ij})) - \log\left(\sum_{j'} \mathrm{W}_{j'} \mathrm{f}(\mathrm{d}_{ij'})\right) \tag{11}$$

3.3.5. *Multi-task loss function*

To estimate trip counts while mitigating parameter confounding between destination attractiveness (W_j) and distance decay (β), we employ a multi-component objective function as defined in Eq. (12). This multi-task approach ensures that the model separately accounts for individual flow accuracy and aggregate destination alignment:

$$\mathrm{L}_{\mathrm{total}} = \mathrm{L}_{\mathrm{OD}} + \mathrm{L}_{\mathrm{dest}} + \mathrm{L}_{\mathrm{reg},1} + \mathrm{L}_{\mathrm{reg\_transfer}} \tag{12}$$

where:

OD Pair Loss ($\mathbf{L}_{\mathbf{OD}}$): Poisson deviance for flow counts, defined in Eq. (13):

$$\mathrm{L}_{\mathrm{OD}} = \sum_{i,j} \left[\widehat{\mathrm{T}}_{ij} - \mathrm{T}_{ij} \log(\widehat{\mathrm{T}}_{ij})\right] \tag{13}$$

Destination Total Loss ($\mathbf{L}_{\mathbf{dest}}$): Reduces confounding between attractiveness and decay by ensuring the model estimates destination inflows correctly, while allowing the decay parameter to anchor to the value that describes the spatial distribution, as shown in Eq. (14):

$$\mathrm{L}_{\mathrm{dest}} = \mathrm{mean}_j \left[\widehat{\mathrm{T}}_j^{\mathrm{in}} - \mathrm{T}_j^{\mathrm{in}} \log\left(\widehat{\mathrm{T}}_j^{\mathrm{in}}\right)\right] \tag{14}$$

Standard L2 Regularisation ($\mathbf{L}_{\mathbf{reg},1}$) :An L_2 penalty applied to residual amenity parameters to prevent overfitting during initial training.

3.4. *Adaptive transfer learning*

3.4.1. *Adaptive similarity weight (Ω_k)*

For amenity weights, Ω_k varies based on the structural difference (Δ_k^{*}) between cities for feature k, and τ the sensitivity of difference in penalty calculation that controls how "strict" the similarity requirement is, as defined in Eq. (15):

$$\Omega_k = \begin{cases} \exp(-\tau \cdot \Delta_k^{*}) & \text{adaptive} \\ 1 & \text{fixed} \end{cases} \tag{15}$$

3.4.2. Adaptive transfer regularisation

A similarity-weighted penalty that anchors the target parameters (γ_k) to the source parameters ($\gamma_{\text{source}k}$) based on the calculated structural weight Ω_k, as defined in Eq. (15). The complete loss function combining data fidelity, standard regularisation, and adaptive transfer is summarised in Eq. (16):

$$L_{\text{total}} = \underbrace{L_{OD} + L_{\text{dest}}}_{\text{Data Fidelity}} + \underbrace{\lambda_1|\gamma|^2}_{\text{Standard Reg.}} + \underbrace{\lambda_{trans}\sum\nolimits_k \Omega_k\left(\gamma_k - \gamma_{\text{source},k}\right)^2}_{\text{Adaptive Transfer Loss}} \tag{16}$$

Rather than a direct parameter transfer, we adapt the target amenity weights γ_k based on a spatial divergence metric Ω_k. The adapted weight γ'_k is defined such that its influence is inversely related to the divergence between the source and target cities, As $\Delta^* \rightarrow 0, \Omega \rightarrow 1$ (Maximum anchor), and as $\Delta^* \rightarrow \infty, \Omega \rightarrow 0$ (Minimum anchor) to source, as shown in Eq. (17):

$$\gamma'_k \rightarrow \gamma_k \text{ as } \Omega_k \rightarrow 0 \tag{17}$$

4. Setup for the experiments

4.1. Study area

We selected Birmingham and Coventry as a methodological test case for evaluating transfer learning in spatial interaction models. Both cities are major urban centres in the West Midlands region of the United Kingdom (Fig. 2), yet they exhibit measurable structural differences that create an ideal natural experiment for assessing cross-city model transferability (Flynn & Taylor, 1986; Healey & Clark, 1984). Economic structure differs between the cities: Birmingham's economy is more service-oriented (70% services, 7% manufacturing), while Coventry retains a stronger manufacturing concentration (65% services, 12% manufacturing), reflecting its automotive and industrial heritage. Spatial form also varies considerably (Business Register and Employment Survey, 2022): Birmingham demonstrates higher employment concentration, with a spatial entropy of $H = 4.0$, indicating a more centralised CBD structure, compared to Coventry's $H = 3.2$, reflecting a more dispersed industrial employment pattern. These differences in economic composition and spatial organisation create observable variations in job geography and commuting patterns. These are the precise conditions needed to test whether models calibrated on one city's structure can adapt to another's. Full demographic and spatial characteristics of Birmingham and Coventry are provided in Appendix C, while Spatial entropy calculations for both cities' features are provided in Appendix D.

Importantly, both cities share institutional and policy contexts that control for confounding factors. They operate under the same UK planning frameworks, national transport policy structures, and statistical reporting standards, allowing us to isolate structural and behavioural differences from regulatory or institutional factors. Additionally, both cities have recently adopted strategic plans emphasising mixed-use development and economic growth in their city centres: Birmingham's Central Birmingham Framework 2045 (Birmingham City Council, 2023) targets CBD intensification and cultural investment, while Coventry's City Centre South Regeneration Framework (Coventry City Council, 2021) focuses on service sector enhancement and connectivity improvements. These parallel policy ambitions, despite different economic starting points, make the cities particularly relevant for testing whether development strategies can be transferred through data-driven modelling.

This pairing represents a proof-of-concept study for transfer learning in urban spatial interaction models. The cities exhibit sufficient structural differences to test transfer mechanisms while maintaining enough similarity to ensure meaningful comparison. However, we acknowledge that this constitutes a limited empirical scope: Birmingham and Coventry are both UK cities within the same region, sharing cultural context, regulatory environments, and data infrastructure. Future work must validate these findings across cities with greater morphological diversity (e.g., polycentric versus monocentric forms), cultural differences (e.g., European versus Asian urban systems), and regulatory variations (e.g., different planning regimes) to establish the generalisability of the transfer learning framework beyond the UK context.

4.2. Datasets

We use Lower Layer Super Output Areas (LSOAs) as the spatial unit for our modelling and analysis. LSOAs are small geographic units in England and Wales developed by the Office for National Statistics for publishing small-area statistics, designed to maintain consistent population sizes of 1000–3000 residents (Office for National Statistics, 2025). This standardisation ensures spatial comparability and statistical robustness across our study regions. We extract and derive spatial features related to destination attractiveness for employment centres, drawing on the 'quality of place' literature (Glaeser, 2008; Tufts, 2003) to examine how amenity access influences employment location decisions in the West Midlands context. For each LSOA, we calculate the following variables:

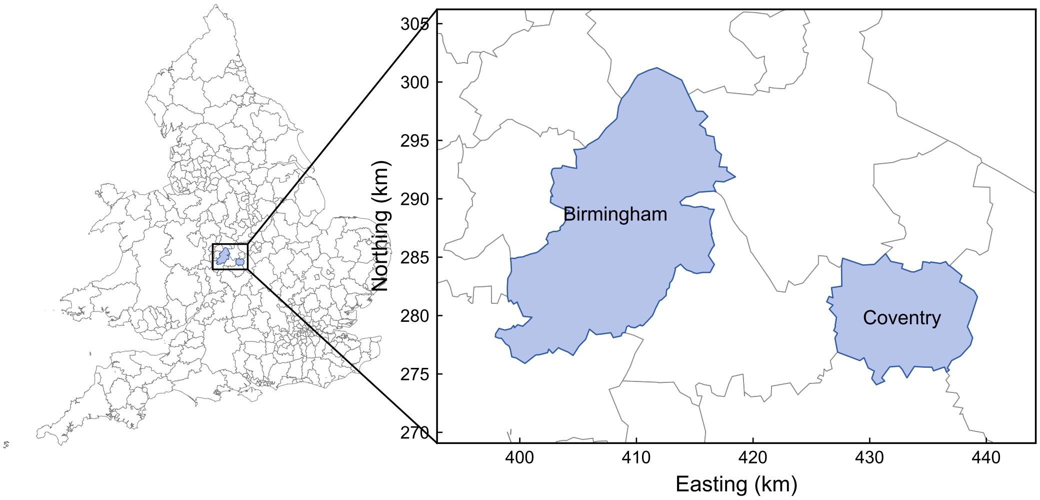


Fig. 2. A geographical representation of the Birmingham and Coventry (UK West Midlands).

1. **Employment Estimates:** We estimate total jobs by multiplying building Gross Floor Area (GFA) in each employment-related building category by the corresponding UK planning employment density factors (Homes and Communities Agency, 2013), then aggregating across categories to obtain workplace concentration at the LSOA level (see Table 1). The complete OSM tag-to-employment mapping and density factors are detailed in Appendix E. A summary of OSM building attributes across both cities is reported in Appendix F.
2. **Residualised Amenity Features**: We capture workplace attractiveness using residualised features that reflect influences beyond employment distribution alone. Buildings are grouped into six amenity types: Cultural Vibrancy, Essential Services, Logistics/Industrial, Office, Retail/Food, and Transport Accessibility (see Table 2). By residualising amenity GFA against employment, we isolate the strategic placement and quality of amenities that affect commuting decisions independently of job concentration. The full POI classification scheme for amenity influence calculation, and feature specifications and preprocessing procedures are documented Appendix G and feature correlation analysis results are presented in Appendix A.
3. **Observed Commuting Flows:** Origin–destination commuting flows are derived from anonymised, aggregated mobile app data at the LSOA level, capturing revealed workplace choice behaviour and providing a behavioural basis for model calibration and validation (Zhong et al., 2025).
4. **Spatial Distance Matrix:** We compute the Euclidean distance d_{ij} between LSOA centroids to represent spatial friction between origins and destinations, forming the basis for modelling distance-decay effects within the spatial interaction framework. We use Euclidean distance as a tractable proxy for travel impedance, simplifying complex road networks and actual travel times. For planning applications, it is recommended to test the model using actual accessibility measures.

5. Results

5.1. Model performance and transfer learning effectiveness

We evaluated five model configurations: within-city baselines for Birmingham and Coventry, direct parameter transfer, fixed transfer learning with uniform regularisation ($\lambda = 2.0$), and adaptive transfer learning with amenity-specific regularisation ($\tau = 17.0$). Fig. 3 and Table 3 present overall performance metrics across all models. We observe that the adaptive transfer outperforms both direct and fixed transfer. The full hyperparameter tuning process and results are documented in Appendix H.

Within-city models had R^2 values of 0.575 (Birmingham) and 0.472 (Coventry), with perfect production satisfaction (Origin $R^2 = 1.000$) and high destination attractiveness fit (Dest $R^2 = 0.973$ and 0.920, respectively). Direct transfer of Birmingham parameters to Coventry with 25% target data fine-tuning resulted in improved performance ($R^2 = 0.529$), representing a 12.1% improvement relative to Coventry's native baseline. Fixed transfer learning with uniform regularisation achieved $R^2 = 0.516$, while adaptive transfer learning with amenity-specific regularisation achieved the highest cross-city performance ($R^2 = 0.547$, +15.9% over baseline Coventry and + 3.4% over naive transfer with 25% target data). Notably, all models maintained perfect production constraints and high Common Part of Commuters (CPC > 0.86), indicating consistent spatial distribution predictions despite varying R^2 values. The destination R^2 metric achieved a high score across all models (0.933 ± 0.008), demonstrating that destination features (jobs and residualised amenities) achieve greater transferability than the complete origin-destination interaction patterns captured by R^2. The power law was chosen as the primary decay function due to superior empirical fit. Comparative results for exponential versus power-law decay functions are presented in Appendix I.

Table 2
Jobs Estimation in Coventry and Birmingham.

	Density (jobs/m^2)	Buildings Count		GFA-Based Jobs	
Category		BHM	COV	BHM	COV
High Density (Office)	0.10	799	377	24,814	19,001
Institutional Stable	0.08	2417	813	566,838	306,488
Medium Density (Retail)	0.07	5265	1561	90,891	29,719
Low Density (Industrial)	0.01	2805	1094	124,196	43,487
Total				**806,739**	**398,695**

5.2. Amenity-specific transfer weights

We chose the optimal sensitivity parameter (τ) for adaptive transfer learning. At the optimal value of $\tau = 17.0$, adaptive transfer learning computed amenity-specific regularisation (Ω) based on the structural similarity between Birmingham's and Coventry's spatial patterns. Table 4 presents the KL divergence (Δ_{KL}), omega weights, effective regularisation strengths (λ_{trans}), and interpretation for each amenity. $\lambda_{trans} = 2.0 \times \Omega$ combines the choice of transfer strength (λ) with data-driven evidence of transferability (Ω) to give each amenity the appropriate regularisation.

As shown in Fig. 4, Office_Amenity displays the largest spatial pattern difference ($\Delta_{KL} = 0.497$), indicating fundamentally distinct clustering of employment-office concentration in the two cities. Similarly, Essential_Service_Amenity achieves high divergence ($\Delta_{KL} = 0.340$) due to contrasting spatial configurations: Birmingham's essential services co-locate with Central Business District (CBD) employment, whereas Coventry centralises these services despite having more peripheral industrial employment.

The omega weights reveal three distinct transferability scenarios in the relationship between employment and amenities. Logistics_Industrial services achieve high similarity ($\Omega = 0.950$), indicating common location principles that transfer well between cities. Transport_Accessibility showed moderate similarity ($\Omega = 0.151$), suggesting partially transferable patterns. Cultural_Vibrancy ($\Omega = 0.041$), Essential_Service_Amenity (Ω=0.003), and Office_Amenity ($\Omega = 0.000$) demonstrate minimal similarity, indicating city-specific strategic placement patterns.

Table 1
Work amenities building classification, count, and Gross Floor Area (FAR) by city.

		Birmingham		Coventry	
Work Amenities	Influence	Count	FAR	Count	FAR
Cultural & Vibrancy	Entertainment, and social environment	588	0.16	226	0.19
Logistics & Industrial	Industrial environment indicator	5763	5.57	2168	4.15
Office	High-end professional	291	0.07	164	0.11
Retail & Food	Convenience and lifestyle services	28,236	6.63	1164	3.97
Transport	Mobility and connectivity factor	1935	1.02	836	1.17
Essential Services	Access to time-saving key services	1395	1.85	496	2.12

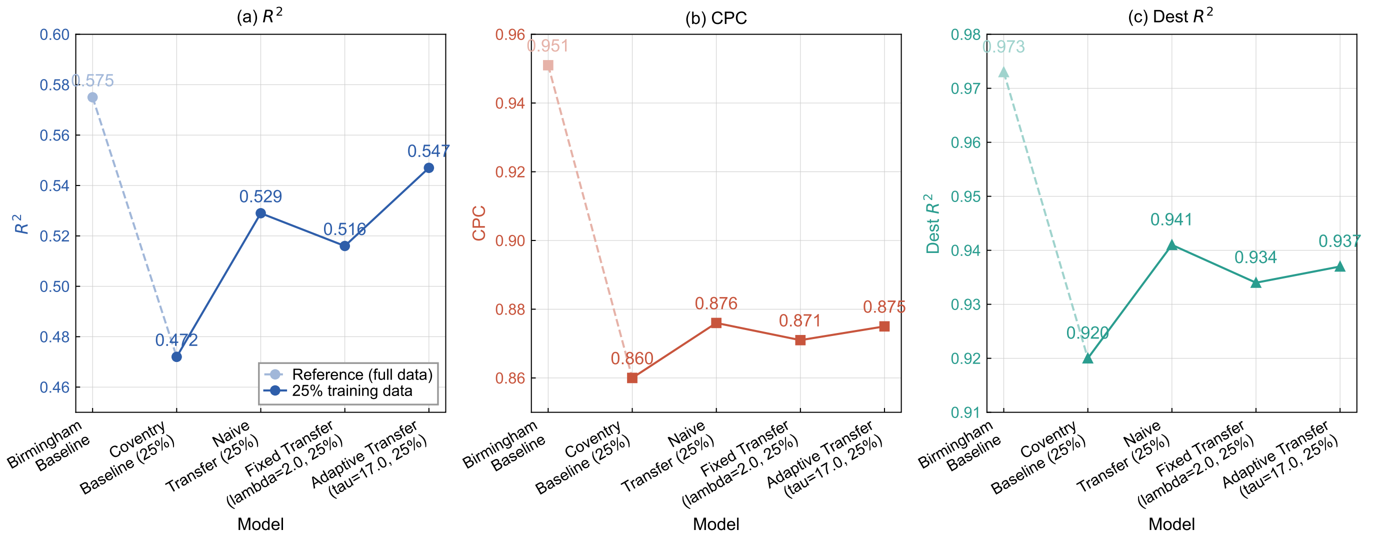


Fig. 3. Model performance (flows and inflows) for all models.

Table 3
Model performance for baseline and transferred models.

Model Type	Train City	Test City	R^2	CPC	Dest R^2
Baseline Models					
Within-CV	Birmingham	Birmingham	0.575	0.951	0.973
Within-CV (25%)	Coventry	Coventry	0.472	0.86	0.92
Transfer (No Adaptation)					
Naive Transfer (25%)	Birmingham	Coventry	0.529	0.876	0.941
Transfer (Adaptation)					
Fixed Transfer ($\lambda = 2.0$ 25%)	Birmingham	Coventry	0.516	0.871	0.934
Adaptive Transfer ($\tau = 17.0$ 25%)	Birmingham	Coventry	0.547	0.875	0.937

5.3. Learned and adapted attractiveness parameters

The learned and adapted attractiveness parameters reveal fundamentally different urban priorities between Birmingham and Coventry. Table 5 presents the parameter composition across models, as illustrated in Fig. 5. We observe diverse patterns in how parameters change across models.

Birmingham's within-city model identified Essential_Service as the dominant attractiveness driver ($\gamma = 0.394$), while Cultural_Vibrancy contributed moderately ($\gamma = 0.157$). In contrast, Coventry's within-city model showed Cultural_Vibrancy ($\gamma = 0.266$) and Transport_Accessibility ($\gamma = 0.218$) as primary drivers, with Essential_Service playing only a minor role ($\gamma = 0.022$). When applying transfer learning, the models inherited Birmingham's Essential_Service dominance: both fixed and adaptive transfer mechanisms maintained Essential_Service above 0.41. Notably, the adaptive mechanism also constrained Cultural_Vibrancy to a moderate value ($\gamma = 0.184$). A similar pattern also emerged for the other parameters, where their values remained close to Birmingham's initial settings, indicating that the adaptive transfer's amenity-specific regularisation restricted substantial deviation due to low differences in spatial structure.

5.4. Urban priorities: transfer learning from Birmingham to Coventry city centres

To evaluate model quality beyond predictive accuracy, we assess the effectiveness of transfer by examining whether the transferred model preserves meaningful behavioural responses. Specifically, we measure how predicted flows (e.g., city-centre inflows) respond to controlled changes in input features related to urban context. Sensitivity is quantified as the relative change in predicted inflow resulting from a unit perturbation in residual attractiveness, which captures the influence of nearby amenities not explained by employment. Formally, sensitivity is defined as:

$$\epsilon = \frac{\Delta \text{Inflow}/\text{BaselineInflow}}{\delta},$$

where ΔInflow denotes the change in predicted city-centre inflow relative to the baseline prediction, and δ represents the magnitude of the applied perturbation. Residual attractiveness at the city centre is systematically perturbed over the range $\delta \in [0, 600]$, and the resulting changes in inflow predictions are recorded. Normalising by baseline inflow allows sensitivity values to be compared across cities and models.

Table 4
Spatial divergence of jobs-amenities locations and transfer weight.

Residualised Amenity	KL_{BHM}	KL_{COV}	Δ_{KL}	Ω	$\lambda_{trans}{}^*$	Interpretation
Cultural_Vibrancy	1.447	1.258	0.188	0.041	0.081	Partial transfer
Essential_Service_Amenity	0.780	1.120	**0.340**	0.003	0.006	No transfer
Logistics_Industrial	0.867	0.863	0.003	**0.950**	**1.900**	Full transfer
Office_Amenity	1.154	1.651	**0.497**	0.000	0.000	No transfer
Retail_Food_Amenity	0.987	1.156	0.169	0.056	0.113	Partial transfer
Transport_Accessibility	0.622	0.732	0.111	**0.151**	**0.303**	Partial transfer

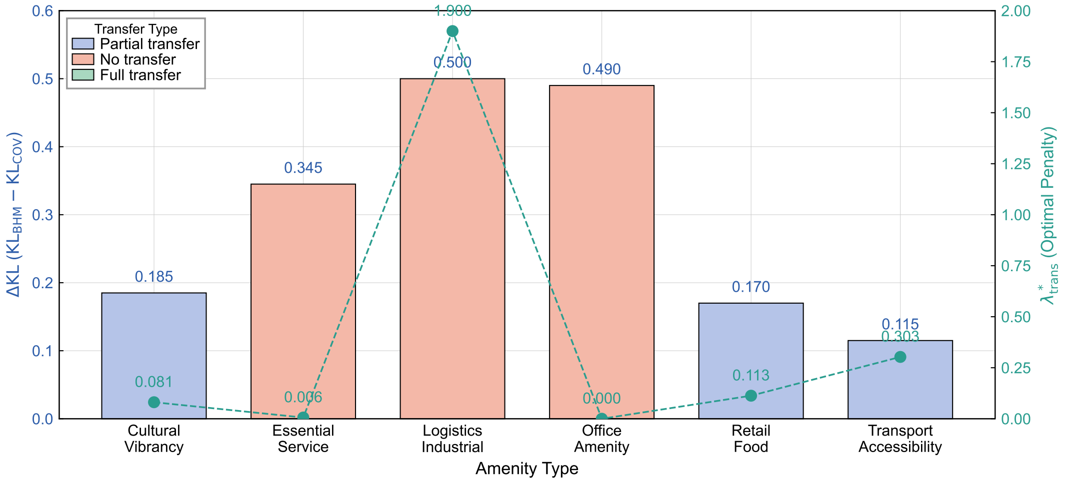


Fig. 4. Residualised job-amenity co-location differences and transfer term across cities.

Table 5
Parameters weights of amenities residuals across models.

City	Distance	Jobs	Cultural	Essential	Logistics	Office	Retail	Transport
Birmingham	0.11	0.17	0.157	**0.394**	0.036	0.162	0.05	0.031
Coventry	0.21	0.43	**0.266**	0.022	0.016	0.009	0.037	**0.218**
Fixed Transfer	0.11	0.093	0.18	**0.418**	0.042	0.173	0.057	0.036
Adaptive Transfer	0.11	0.095	0.184	**0.415**	0.043	0.168	0.058	0.037

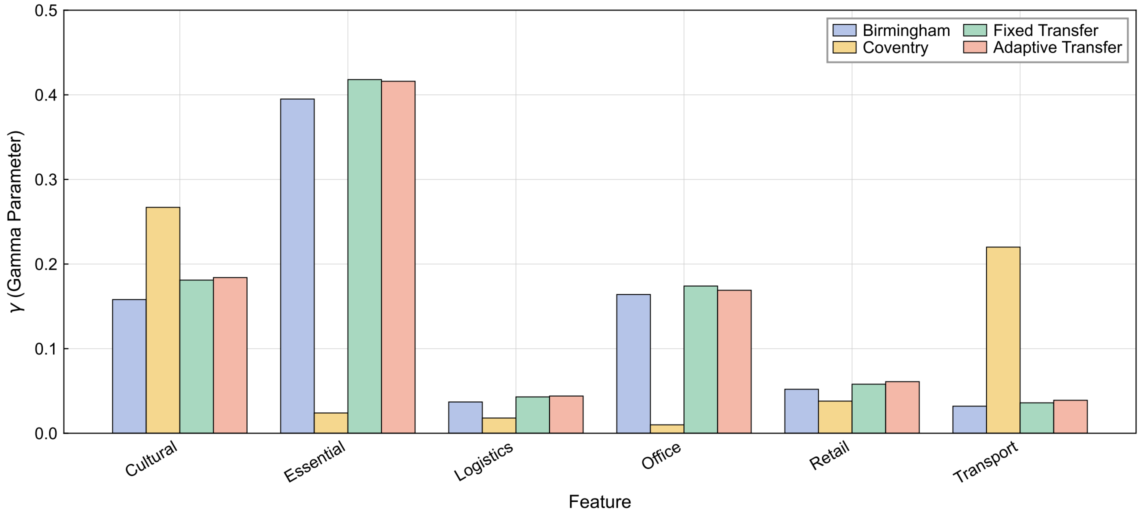


Fig. 5. Parameters weights of amenities residuals across models.

The resulting sensitivity estimates in Fig. 6 and Table 6 reflect how strongly city-centre inflows respond to amenity-driven attractiveness. Parameters exhibiting notable differences in sensitivity across the two cities indicate structural differences in mobility responses and provide evidence on where transfer learning succeeds or fails to preserve underlying urban interaction mechanisms.

The Birmingham model exhibited low residual sensitivity to Cultural amenities ($\varepsilon = 0.0113$) but higher dependence on Essential Services ($\varepsilon = 0.0285$). In contrast, the Coventry model showed the highest sensitivity to Cultural residuals ($\varepsilon = 0.0372$) and Transport ($\varepsilon = 0.0304$), while demonstrating minimal dependence on Essential Services ($\varepsilon = 0.0031$). The transfer learning models demonstrated increased sensitivity to Cultural residuals relative to Birmingham ($\varepsilon \approx 0.017$), shifting toward Coventry's baseline but not fully reaching it. For Essential Services, the Fixed Transfer substantially overshot both baselines ($\varepsilon = 0.0417$), while the Adaptive Transfer reduced this sensitivity ($\varepsilon = 0.0228$) but remained well above Coventry's near-zero level. Office amenities showed notable variation: the Fixed Transfer retained Birmingham-like sensitivity ($\varepsilon = 0.0172$), whereas the Adaptive Transfer achieved near-zero dependence ($\varepsilon = 0.0004$). Both transfer models maintained low Transport sensitivity ($\varepsilon < 0.004$), failing to capture Coventry's elevated baseline ($\varepsilon = 0.0304$).

These results indicate that transfer learning largely replicates Birmingham's urban configuration into Coventry rather than learning Coventry's own priority structure. Both the Fixed and Adaptive Transfer retain Birmingham-like sensitivity patterns, particularly the strong dependence on Essential Services. Although Cultural sensitivity increases after transfer, it remains closer to Birmingham's baseline than to Coventry's. This suggests that the transferred models reproduce

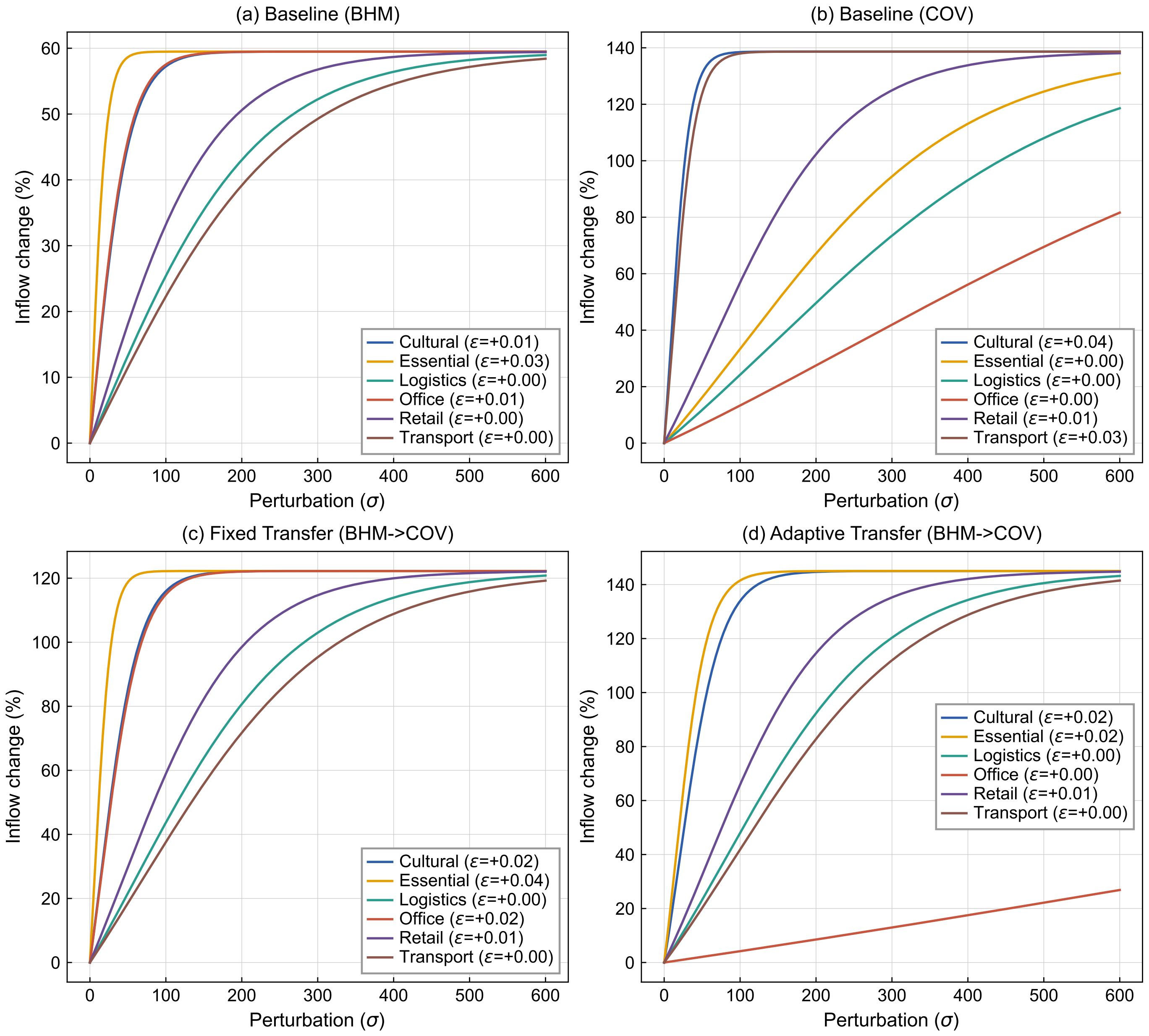


Fig. 6. Model sensitivity analysis results (changes in residual attractiveness and centre inflows).

Birmingham's city-centre configuration in Coventry.

Table 6
Sensitivity analysis of residual of amenities and centre inflows.

	Sensitivity			
Model	Cultural	Essential	Office)	Transport
Birmingham Baseline (BHM)	0.012	0.029	0.012	0.002
Coventry Baseline (COV)	0.037	0.003	0.001	0.030
Fixed Transfer (BHM → COV)	0.018	0.042	0.017	0.004
Adaptive Transfer (BHM → COV)	0.017	0.023	0.000	0.004

6. Discussion

Our adaptive transfer learning analysis reveals nuanced dynamics between spatial structures and learned commuting behavioural parameters. When cities share similar spatial co-location patterns for a particular amenity, parameters tend to be regularised toward source values during transfer, particularly under limited target data. When patterns differ, the model gains greater freedom to adapt locally. However, parameter variation is not driven solely by spatial divergence to control the adaptation, but also by the extent to which source and target flow data exhibit consistent behavioural signals for a given amenity within the spatial interaction model.

For example, the importance of nearby office amenities to employment shows the most divergent spatial patterns between Birmingham and Coventry, yet parameters barely change because both cities learned that office amenities contribute relatively modestly beyond the main jobs signal. Conversely, essential services (e.g., healthcare) show significant differences in spatial structures across cities; however, the transfer models retained Birmingham's high weighting, suggesting the optimisation landscape favours the source city's configuration when the target city has insufficient evidence in the training data to determine employment behaviour with respect to this amenity type. These parameter dynamics across cities confirm that travel behaviour is highly context-dependent and influenced by factors beyond universal factors such as jobs.

The adaptive mechanism also reveals common patterns: industrial and logistics services follow similar spatial logic relating to jobs across Birmingham and Coventry, given their economic geography as two adjacent agglomerations in the English Midlands. This results in minimal reweighting of their related services during the transfer learning process. This occurs despite the unique nature of economic activities across the two cities: Coventry as a key player in the automotive industry, and Birmingham as a service-oriented economy. This suggests the model captures planned co-location of these services near employment locations, but may not detect more subtle local signals in how jobs are spatially distributed and the type of supporting amenities in shaping job attractiveness.

In contrast, transport accessibility shows moderate spatial differences between cities, with Coventry placing substantially higher importance on this factor compared to Birmingham, yet the transfer models retained Birmingham's low weighting. This suggests that Coventry's distributed employment pattern, with peripheral factories and business parks, creates stronger dependencies on transport infrastructure. Local workers in Coventry rely more heavily on transport infrastructure to overcome the city's distributed economy, a pattern that Birmingham's transferred model underweights. Similarly, the retention of Birmingham's essential services weighting in the adapted models despite Coventry's within-city model placing minimal importance on them. These observations may present critical insights requiring policy deliberation, but they also mask local priorities that have emerged organically in how jobs, essential services, and accessibility are spatially aligned in Coventry that transfer learning fails to capture.

A similar pattern emerges in the distance decay parameters. Birmingham's model exhibits a flatter distance decay, reflecting a highly centralised CBD where concentrated jobs pull commuters across greater distances. In contrast, Coventry naturally exhibits a steeper decay, where the friction of distance is higher due to peripheral industrial estates separated from a centralised service hub. When Birmingham's model is transferred, it imposes an artificially flat distance penalty on a city that naturally relies on shorter, more localised interactions. This technical mismatch highlights a spatial incompatibility between the two urban systems.

A deeper sensitivity analysis reveals how the Birmingham model imposes its urban configuration on Coventry through the optimisation process. Testing how city-centre flows respond to changes in model parameters demonstrated that Birmingham's structural relationships, particularly the strong dependence on essential services, are transferred into the Coventry context, even where Coventry's own baseline model shows minimal sensitivity to these factors. This finding encourages reframing transfer learning as a policy transfer simulation rather than a technical generalisability test. Such a perspective shifts policy questions from “can we use the Birmingham model in Coventry?” to “what is the feasibility cost of copying Birmingham versus optimising Coventry's own model?” or “what structural assumptions and trade-offs arise from applying the Birmingham model in Coventry?”

This interpretability highlights a key advantage of our framework over “black-box” deep learning methods, which are often challenged to explain the underlying structural dynamics between cities. By quantifying the transfer process through divergence and regularisation weights integrated with the simple gravity form, we help operationalise urban policy transfer from qualitative judgment into measurable simulation. This allows planners to see exactly where a source city's logic is being artificially imposed on a target context, providing the transparency that is vital for accountable policy-making.

This same interpretability also serves a diagnostic function, even though the framework still requires some data that under-resourced cities often lack. As we have seen, the model performs well with as little as a quarter of target flow data, and even this small amount is itself informative: the framework can diagnose structural incompatibility by testing whether the spatial profile of a data-rich source city is fundamentally at odds with that of a data-poor target. By identifying these specific areas of incompatibility, the framework can be used to guide targeted data collection efforts toward parameters that require local context, such as specific amenity attractions, rather than exhausting resources on a broad data acquisition strategy that may only confirm universal patterns already captured by the source model.

By surfacing where a source city's logic fails to fit the target, the framework encourages planners to engage with the target city's own patterns rather than imposing an external template. This diagnostic nature positions transfer learning mechanisms as instruments to enable more bottom-up planning approaches that consider local community priorities when studying employment attractiveness. Recognising that Coventry's model emphasises transport accessibility over essential services, contrary to Birmingham's pattern, suggests that improving transport connectivity in Coventry may yield superior returns. The distance decay mismatch further suggests that policy interventions should address spatial fragmentation directly, such as transport connectivity between peripheral employment zones and the central service hub rather than attempting to replicate Birmingham's compact CBD model. Investing in Coventry's existing strengths would be more effective and less risky than assuming successful cities must converge on the same development model.

These observations remain preliminary. We recommend deeper empirical analysis of employment and amenity clustering, such as spatial autocorrelation tests, to determine whether these results reflect data noise or meaningful spatial patterns arising from planning decisions. Such analysis could inform interventions to adjust travel patterns through transport investments that increase accessibility to target services from employment locations and incentivise adjustments in activity distributions. Another potential area of future research is to examine interaction features such as the transport network and incorporate them into the transfer process, to investigate whether transfer learning can capture network variabilities and yield more contextualised insights into spatial interactions.

7. Conclusion

This research presents TransGM, a methodological framework that supports the transfer of urban policy models across cities. By integrating information-theoretic divergence measures with adaptive regularisation, the framework provides a systematic tool for measuring structural compatibility and predicting transfer costs before committing resources to large-scale interventions.

We found that model transfer operates along two independent dimensions: structural similarity sets the bounds on how freely parameters can deviate from the source model, while observed behavioural patterns determine whether deviation occurs. Applied to Birmingham–Coventry commuting patterns, this analysis uncovered distinct urban configurations: Birmingham's model, in which essential services support the attractiveness of CBD employment, versus Coventry's model, in which cultural vibrancy and transport accessibility anchor distributed industrial employment. These typologies emerged organically from the data rather than from theoretical imposition, demonstrating the framework's capacity to reveal structural differences that may affect policy transfer. This is particularly useful for urban planning practice because TransGM enables evidence-based assessment of urban policy transfer. Rather than assuming successful cities follow universal templates, planners can quantify structural alignment with external models and determine whether enhancing local strengths or adopting external strategies better serves their context, enabling more bottom-up planning approaches that respect local community priorities instead of imposing external templates.

This analysis is subject to several methodological limitations, including the use of OSM building footprints, inconsistent classification tags, and spatially variable data coverage, suggesting that employment estimates should be interpreted as indicative rather than definitive.

Future development should validate generalisability across cities with greater morphological and cultural diversity, incorporate deeper industry composition analysis, and pursue methodological refinements such as alternative divergence metrics, robustness-aware training, and the inclusion of city network structures and residual travel cost analysis within the model design. Extending the framework to temporal transfer learning and additional policy domains would further demonstrate broader applicability, positioning TransGM as a flexible platform for evidence-based urban policy transfer.

CRediT authorship contribution statement

Adham Enaya: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Chen Zhong:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Michael Batty:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Robin Morphet:** Writing – review & editing, Supervision, Conceptualization. **Fulvio D. Lopane:** Writing – review & editing, Conceptualization.

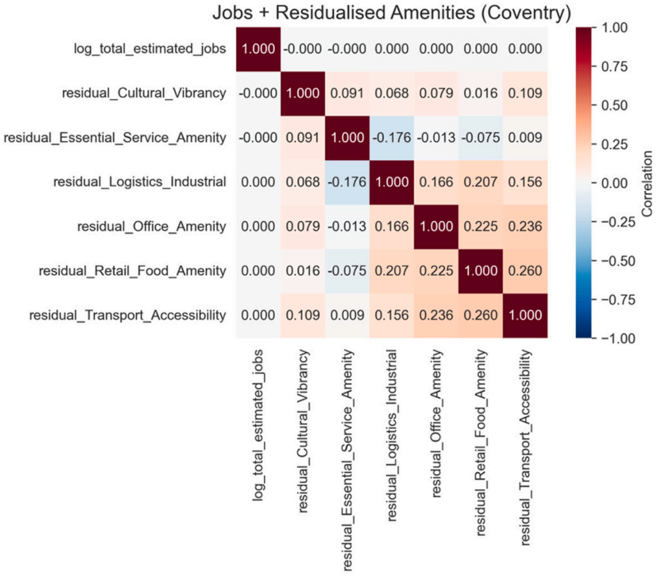
Declaration of competing interest

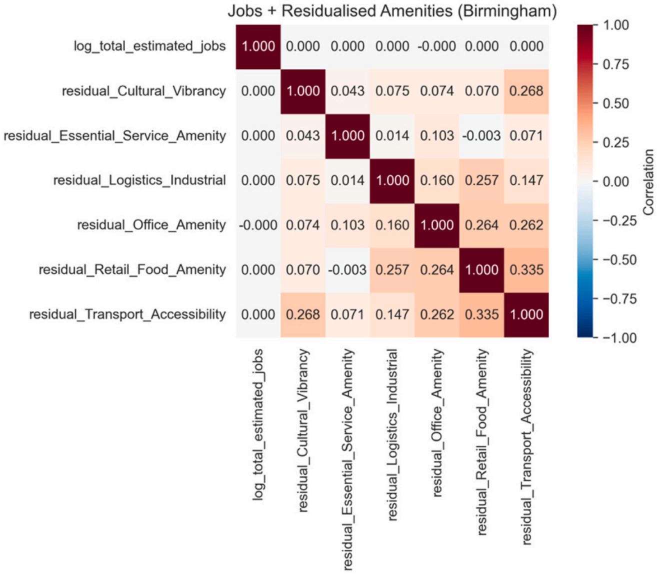
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by the European Research Council (ERC) under the European Union‘s Horizon 2020 Research and Innovation Programme (Grant agreement No. 949670). The authors are grateful to Dr. Nathan Wycoff (University of Massachusetts Amherst) for his generous assistance and to Dr. Beatrice Taylor (University College London) for her insightful comments and constructive feedback, which substantially improved the quality of this work.

Appendix A. Destination features correlation test





Appendix B. Model optimisation settings

B.1. L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Bounds)

Parameter	Value	Description
History size	10	Number of previous Hessian approximations stored
Max iterations	1000	Maximum Optimisation steps
Convergence tolerance	1e-6	Gradient norm threshold for convergence
Line search	Strong Wolfe	Ensures robust convergence
Typical convergence	200–500 iterations	Actual iterations needed

Parameter Bounds:

- Distance decay: $\beta \in [0.5, 5.0]$
- Jobs elasticity: $\gamma\_jobs \in [0.01, 0.50]$
- Utility weights: γ_k

Numerical Stability:

- All computations performed in log-space
- LogSumExp trick to prevent overflow/underflow
- Automatic differentiation via PyTorch autograd

Regularisation: $\lambda 1 \in \{0.001, 0.01, 0.1, 1.0\}$ (selected via cross-validation).

Appendix C. Study area characteristics

Birmingham (Source Domain)	
Attribute	Value
Population	1,141,816
LSOA zones	639
OD pairs (training)	4355
Total commute trips	298,450
Mean trip distance	8.2 km
Urban profile	High concentration of office and service employment

Coventry (Target Domain)	
Attribute	Value
Population	369,127
LSOA zones	174
OD pairs (total)	1278
Training (30%)	383 OD pairs
Testing (70%)	895 OD pairs
Total commute trips	98,250
Mean trip distance	6.8 km
Urban profile	Industry focused economy

Appendix D. Spatial entropy for Coventry and Birmingham

D.1. Jobs

- Birmingham: Spatial entropy $H = 4.0$ (centralised)
- Coventry: Spatial entropy $H = 3.2$ (distributed)

Complete spatial entropy values for Birmingham and Coventry POI categories:

POI Category	Birmingham H	Coventry H	ΔH
Leisure_Recreation	0.722	0.884	0.162
Community_Cultural	0.850	0.897	0.046
Professional_Office	0.700	0.648	0.052
Essential_Services	0.954	0.935	0.019
Industrial_Commercial	0.904	0.860	0.044
Transport_Accessibility	0.949	0.938	0.012

Appendix E. OSM-based job estimation

This table provides the exhaustive list of OpenStreetMap tags mapped to each employment category and their associated job density factors: Job counts = GFA × density factor, based on the Based on the UK standards.

Job Category	Focus	Density (Jobs/m^2)
High Density Office	White-Collar, Finance	0.10
Institutional and Stable	Health, Education, Public	0.08
Medium Density Retail	Food, Shops, Service	0.07
Low Density Industrial	Warehouse, Manufacturing	0.01

Category	OSM Key	Included Tags / Values
High Density Office	office	company, corporate, coworking, administrative, lawyer, notary, architect, engineer, surveyor, consulting, it, research, telecommunication, financial, insurance, accountant, tax_advisor, estate_agent, government, diplomatic
	amenity	townhall, courthouse, embassy, parliament, public_building, bank, bureau_de_change, money_transfer, coworking_space
	government	administrative, archive, city, customs, legislative, military, ministry, tax
	building	government, civic, public, office, commercial
Institutional Stable	amenity	doctors, clinic, hospital, dentist, pharmacy, veterinary, nursing_home, social_facility, school, college, university, kindergarten, childcare, language_school, music_school, driving_school, prep_school, library, research_institute, post_office, police, fire_station, ambulance_station, place_of_worship, monastery
	healthcare	doctor, dentist, hospital, clinic, physiotherapist, optometrist, psychotherapist, midwife, nurse, alternative, laboratory, rehabilitation
	social_facility	group_home, nursing_home, assisted_living, daycare, shelter, food_bank
	building	hospital, school, college, university, kindergarten, church, mosque, temple, synagogue, chapel
Medium Density Retail	amenity	restaurant, cafe, fast_food, food_court, ice_cream, pub, bar, biergarten, marketplace, fuel, car_wash, car_rental, vehicle_inspection, storage_rental
	shop	ALL (Wildcard)
	leisure	sports_centre, fitness_centre, swimming_pool, amusement_arcade, bowling_alley, escape_game, cinema, theatre, nightclub, casino, dance
	building	retail, supermarket, commercial, kiosk
Low Density Industrial	landuse	industrial, brownfield, depot, quarry, port
	industrial	ALL (Wildcard)
	man_made	works, wastewater_plant, water_works, gasometer, storage_tank, silo
	craft	builder, electrician, plumber, roofer, carpenter, painter, plasterer, hvac, stonemason, welder, sawmill, metal_construction, caterer, distillery, brewery, winery, bakery, car_repair, electronics_repair
	building	warehouse, industrial, factory, hangar
	power	plant, substation

Appendix F. Amenities gross floor area

		GFA Ratio		Buildings/LSOA		Building Counts	
Category	Influence	BHM	COV	BHM	COV	BHM	COV
Cultural & Vibrancy	Entertainment, and social environment	56,420	37,060	39.8	2.3	26,237	473
Logistics & Industrial	Industrial environment indicator	32,471	37,704	1.6	2	1601	852
Office	High-end professional	375	619	2.4	4.2	313	188
Retail & Food	Convenience and lifestyle services	3429	3366	0.5	0.9	4042	1385
Transport	Mobility and connectivity factor	4631	7488	6.1	6.8	1665	812
Essential Services	Access to time-saving key services	3869	4868	2.5	4	1065	415

Appendix G. OSM-based amenities influence

This table outlines the categorisation of Points of Interest (POI) used to calculate environmental "influence" factors. Also, influence is calculated based on Gross Floor Area (GFA).

Category	Calculation	OSM Key	Included Tags / Values
Office Amenity	GFA	office	company, corporate, coworking, lawyer, architect, consulting, engineer, it, financial, insurance, accountant, tax_advisor, estate_agent
		amenity	townhall, courthouse, embassy, parliament, coworking_space, conference_centre

(continued on next page)

(*continued*)

Category	Calculation	OSM Key	Included Tags / Values
		government	administrative, ministry, tax
		building	office, government, civic
Retail & Food Amenity	GFA	amenity	restaurant, cafe, fast_food, food_court, ice_cream, pub, bar, biergarten, marketplace, atm, vending_machine
		shop	ALL (Wildcard)
		building	retail, supermarket, kiosk
Essential Service Amenity	GFA	amenity	doctors, clinic, hospital, dentist, pharmacy, veterinary, bank, atm, post_office, post_box, childcare, kindergarten, fuel, charging_station, car_wash
		healthcare	doctor, dentist, hospital, clinic, physiotherapist, pharmacy
		building	hospital, school, university
Cultural Vibrancy	GFA	amenity	cinema, theatre, nightclub, casino, arts_centre, library, archive, community_centre, events_venue, exhibition_centre, conference_centre
		tourism	museum, gallery, zoo, aquarium, theme_park, attraction
		leisure	community_centre, stadium, sports_centre, park, garden, playground
		historic	monument, memorial, castle, archaeological_site
Logistics & Industrial	GFA	landuse	industrial, quarry, port
		industrial	ALL (Wildcard)
		man_made	works, wastewater_plant, water_works
		building	warehouse, industrial, factory
Transport Accessibility	Count	amenity	bus_station, ferry_terminal, taxi, parking, bicycle_parking, bicycle_rental, charging_station
		public_transport	station, platform, stop_position
		railway	station, halt, tram_stop, subway_entrance
		aeroway	aerodrome, terminal
		highway	bus_stop

Feature Engineering: Destination Features (7 features).
Primary:

- Jobs: Total employment at destination (ONS Business Register)

Residual Utilities (6 features):

- Office_Amenity: Office floor space (m^2)
- Retail_Food_Amenity: Retail + food establishments count
- Transport_Accessibility: Bus stops + rail stations density
- Essential_Service_Amenity: Schools + healthcare facilities count
- Cultural_Vibrancy: Leisure + recreation POI count
- Logistics_Industrial: Industrial floor space (m^2)

Preprocessing:

- Log transformation: log(feature +1) to handle zeros
- No standardization (raw log-features perform better)
- All features aggregated to LSOA zone level

Distance Matrix.
Metric: Euclidean distance between LSOA centroids

- Birmingham: 639 × 639 matrix
- Coventry: 174 × 174 matrix
- Units: kilometers

Nugget term: $\delta = 0.1 \times \text{mean(distance)}$ added to prevent $d = 0$ issues

Appendix H. Hyperparameter tuning

H.1. Lambda Sweep (transfer regularisation strength)

Objective: Identify optimal transfer regularisation strength.
Configuration:

- λ_transfer $\in \{0.5, 1.0, 2.0, 5.0\}$
- $\tau = 1.0$ (baseline, fixed)
- Models: Fixed transfer, Adaptive transfer

Selected: λ_transfer = 2.0 (balanced regularisation).
Tau Sweep (Adaptive Sensitivity Parameter).
Objective: Optimize utility-specific adaptation sensitivity.

Configuration:

- $\tau \in \{1.0, 2.0, 5.0, 10.0, 17.0, 26.0\}$
- λ_transfer = 2.0 (fixed from lambda sweep)
- Training: Birmingham → Coventry

Key finding: Performance peaks at $\tau = 17.0$ then declines → clear optimum.

Appendix I. Exponential gravity model results

Performance:

- Generally, 1–3% lower R^2 than power law
- Less commonly used in spatial interaction models

Conclusion: Power law selected as primary decay function.
Distance Decay Function Comparison.

Decay Function	Formula	Birmingham CV R^2	Birmingham Self-test R^2	Coventry Cross-city R^2
Power Law (selected)	d^(−β)	**0.575**	**0.973**	**0.211**
Exponential	exp(−β × d)	0.560	0.945	0.205
Difference	–	**+0.015**	**+0.028**	**+0.006**

I.1. Key findings

1. **Power law consistently outperforms exponential** across all metrics (within-city: +2.6%, cross-city: +2.9%)
2. **Better transferability:** Power law maintains higher performance in cross-city prediction
3. **Theoretical justification:** Power law decay is standard in UK transport modelling (Batty & Milton, 2021)
4. **Interpretation**: β_power = 1.85 indicates moderate distance sensitivity, β_exp = 0.023 harder to interpret directly

Selection rationale: Power law chosen as primary decay function due to superior empirical fit and consistency with established transport literature.

Data availability

data used for this study is in public domain

References

Bathelt, H., & Glückler, J. (2003). Toward a relational economic geography. *Journal of Economic Geography, 3*(2), 117–144. Available at: https://doi.org/10.1093/jeg/3.2.117.

Batty, M. (1974). Spatial entropy. *Geographical Analysis, 6*(1), 1–31. Available at: https://doi.org/10.1111/j.1538-4632.1974.tb01014.x.

Batty, M. (2010). *Urban Modelling: Algorithms, Calibrations, Predictions*. in. Available at: https://www.semanticscholar.org/paper/Urban-Modelling%3A-Algorithms%2C-Calibrations%2C-Batty/ee16cf2a1099cc02f20611c3560fdfd8867d01b3 (Accessed: 5 May 2025).

Batty, M., & March, L. (1976). The method of residues in urban modelling. *Environment and Planning A: Economy and Space, 8*(2), 189–214. Available at: https://doi.org/10.1068/a080189.

Batty, M., & Milton, R. (2021). A new framework for very large-scale urban modelling. *Urban Studies, 58*(15), 3071–3094. Available at: https://doi.org/10.1177/0042098020982252.

Ben-Akiva, M. (1979). *Issues in transferring and updating travel behavior models. 79–27.* Available at: https://trid.trb.org/View/1153691 (Accessed: 30 April 2024).

Ben-Akiva, M., Bolduc, D., & Pene, K. (1995). The combined estimator approach to model transferability and updating. *Empirical Economics, 20*(1), 167–175. Available at: https://doi.org/10.1007/BF01235164.

Ben-David, S., et al. (2010). A theory of learning from different domains. *Machine Learning, 79*(1), 151–175. https://doi.org/10.1007/s10994-009-5152-4

Birmingham City Council. (2023). *Central Birmingham Framework 2045*. Available at: https://www.birmingham.gov.uk/info/50273/our_future_city_plan_ofcp/2303/our_future_city_-_central_birmingham_framework_2045 (Accessed: 21 May 2025).

Ben-Akiva, M., & Bolduc, D. (1987). Approaches to model transferability and updating: The combined transfer estimator. *Transportation Research Record*, (1139), 1–7.

Business Register and Employment Survey (2022). Available at: https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/businessregisterandemploymentsurvey (Accessed: 19 January 2026).

Canitez, F. (2020). Transferring sustainable urban mobility policies: An institutional perspective. *Transport Policy, 90*, 1–12. Available at: https://doi.org/10.1016/j.tranpol.2020.02.005.

Capoani, L. (2023). Review of the gravity model: Origins and critical analysis of its theoretical development. *SN Business & Economics, 3*(5), 95. Available at: https://doi.org/10.1007/s43546-023-00461-0.

Chelba, C., & Acero, A. (2006). Adaptation of maximum entropy capitalizer: Little data can help a lot. *Computer Speech & Language, 20*(4), 382–399. Available at: https://doi.org/10.1016/j.csl.2005.05.005.

Coventry City Council. (2021). *City Centre south*. Coventry City Council. Available at: https://www.coventry.gov.uk/regeneration-1/city-centre-south (Accessed: 21 May 2025).

Dolowitz, D. P., & Marsh, D. (2000). Learning from abroad: The role of policy transfer in contemporary policy-making. *Governance, 13*(1), 5–23. Available at: https://doi.org/10.1111/0952-1895.00121.

Ewing, R., et al. (2011). Traffic generated by mixed-use developments—Six-region study using consistent built environmental measures. *Journal of Urban Planning and Development, 137*(3), 248–261. Available at: https://doi.org/10.1061/(ASCE)UP.1943-5444.0000068.

Feng, S., et al. (2024). Deep neural networks for choice analysis: Enhancing behavioral regularity with gradient regularization. *Transportation Research Part C: Emerging Technologies, 166*, Article 104767. Available at: https://doi.org/10.1016/j.trc.2024.104767.

Flynn, N., & Taylor, A. P. (1986). Inside the Rust Belt: An analysis of the decline of the west midlands economy. 1: International and National Economic Conditions. *Environment and Planning A: Economy and Space, 18*(7), 865–900. Available at: https://doi.org/10.1068/a180865.

García, C. B., et al. (2019). Residualization: Justification, properties and application. *Journal of Applied Statistics, 47*(11), 1990–2010. Available at https://doi.org/10.1080/02664763.2019.1701638.

Glaeser, E. L. (2008). The economic approach to cities. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1080294

Glaser, M., et al. (2022). Learning through policy transfer? Reviewing a decade of scholarship for the field of transport. *Transport Reviews, 42*(5), 626–644. Available at: https://doi.org/10.1080/01441647.2021.2003472.

Gong, Z., et al. (2025). Multi-spatial urban function modeling: A multi-modal deep network approach for transfer and multi-task learning. *International Journal of*

Applied Earth Observation and Geoinformation, 136, Article 104397. Available at: https://doi.org/10.1016/j.jag.2025.104397.

Hansen, S. (1981). In favor of cross-cultural transferability of travel-demand models. In *New Horizons in Travel-Behavior Research', in. This paper was presented at the Fourth International Conference on Behavioral Travel Modeling*. Available at: https://trid.trb.org/View/699205 (Accessed: 29 September 2024).

Healey, M., & Clark, D. (1984). Industrial decline and government response in the west midlands: The case of Coventry. *Regional Studies, 18*(4), 303–318. Available at: https://doi.org/10.1080/09595238400185301.

Homes and Communities Agency. (2013). *GOV.UK*. Available at: https://www.gov.uk/government/organisations/homes-and-communities-agency (Accessed: 19 January 2026).

Jiang, J., Shu, Y., Wang, J., & Long, M. (2022). Transferability in deep learning: A survey. *Journal of Machine Learning Research, 23*(340), 1–64.

Khirfan, L., & Jaffer, Z. (2014). Sustainable urbanism in Abu Dhabi: Transferring the Vancouver model. *Journal of Urban Affairs, 36*(3), 482–502. Available at: https://doi.org/10.1111/juaf.12050.

Kim, J., Park, J., & Lee, W. (2018). Why do people move? Enhancing human mobility prediction using local functions based on public records and SNS data. *PLoS One, 13*(2), Article e0192698. Available at: https://doi.org/10.1371/journal.pone.0192698.

Klinkhardt, C., et al. (2021). Using OpenStreetMap as a data source for attractiveness in travel demand models. *Transportation Research Record, 2675*(8), 294–303. Available at: https://doi.org/10.1177/0361198121997415.

Kouw, W. M., & Loog, M. (2019). A review of domain adaptation without target labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(3), 766–785. https://doi.org/10.1109/TPAMI.2019.2945942

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79–86.

Kwon, O.-H., et al. (2023). Multiple gravity laws for human mobility within cities. *EPJ Data Science, 12*(1), Article 57. Available at https://doi.org/10.1140/epjds/s13688-023-00438-x.

Lopane, F. D., et al. (2023). A land-use transport-interaction framework for large scale strategic urban modeling. *Computers, Environment and Urban Systems, 104*, Article 102007. Available at: https://doi.org/10.1016/j.compenvurbsys.2023.102007.

Luca, M., et al. (2021). A survey on deep learning for human mobility. *ACM Computing Surveys, 55*(1), 7:1–7:44. Available at: https://doi.org/10.1145/3485125.

Ludwig, M., et al. (2023). Assessing and improving the transferability of current global spatial prediction models. *Global Ecology and Biogeography, 32*(3), 356–368. Available at: https://doi.org/10.1111/geb.13635.

Lye, A., Ferson, S., & Xiao, S. (2024). *Distribution-free stochastic model updating for the physics-guided reliability analysis of a material thermal property under limited data*.

Marsden, G., & Stead, D. (2011). Policy transfer and learning in the field of transport: A review of concepts and evidence. *Transport Policy, 18*(3), 492–500. Available at: https://doi.org/10.1016/j.tranpol.2010.10.007.

Molnar, C. (2019). *Interpretable machine learning*.

Murakami, D., Kajita, M., & Kajita, S. (2025). Spatial process-based transfer learning for prediction problems. *Journal of Geographical Systems, 27*(1), 147–166. Available at: https://doi.org/10.1007/s10109-024-00455-y.

Overman, H. G., & Xu, X. (2024). Spatial disparities across labour markets. *Oxford Open Economics, 3*(Supplement_1), i585–i610. https://doi.org/10.1093/ooec/odae005

Rao, C., & Nayak, T. (1985). Cross entropy, dissimilarity measures, and characterizations of quadratic entropy. *IEEE Transactions on Information Theory, 31*(5), 589–593. Available at: https://doi.org/10.1109/TIT.1985.1057082.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters, 85*(2), 461–464. Available at: https://doi.org/10.1103/PhysRevLett.85.461.

Sikder, S., et al. (2013). Spatial transferability of travel forecasting models: A review and synthesis. *International Journal of Advances in Engineering Sciences and Applied Mathematics, 5*(2–3), 104–128. Available at: https://doi.org/10.1007/s12572-013-0090-6.

Simini, F., et al. (2021). A deep gravity model for mobility flows generation. *Nature Communications, 12*(1), 6576. https://doi.org/10.1038/s41467-021-26752-4

Song, Y., et al. (2025). Transfer learning for high dimensional spatial autoregressive model. *Spatial Statistics*. , Article 100908. https://doi.org/10.1016/j.spasta.2025.100908

Thomas, R., et al. (2018). Is transit-oriented development (TOD) an internationally transferable policy concept? *Regional Studies, 52*(9), 1201–1213. Available at: https://doi.org/10.1080/00343404.2018.1428740.

Tufts, S. (2003). Florida, Richard 2002. The rise of the creative class: And how it's transforming work, leisure, community and everyday life. New York: Basic Books, 404p. *Just Labour*. https://doi.org/10.25071/1705-1436.180

Office for National Statistics (2025) *Annual Population Survey: April 2024 – March 2025*. Available at: https://www.nomisweb.co.uk/datasets/apsnew (Accessed: 15 August 2025).

West Midlands Combined Authority (2025). Available at: https://www.wmca.org.uk/media/ondaaslh/economic-geographies-of-the-west-midlands.pdf (Accessed: 14 December 2025).

Wilson, A. (2018). The future of urban modelling. *Applied Spatial Analysis and Policy, 11*(4), 647–655. Available at: https://doi.org/10.1007/s12061-018-9258-6.

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A: Economy and Space, 3*(1), 1–32. Available at: https://doi.org/10.1068/a030001.

Wood, A. (2014). Moving policy: Global and local characters circulating bus rapid transit through South African cities. *Urban Geography, 35*(8), 1238–1254. Available at: https://doi.org/10.1080/02723638.2014.954459.

Yeghikyan, G., Opolka, F. L., Nanni, M., Lepri, B., & Liò, P. (2020). Learning Mobility Flows from Urban Features with Spatial Interaction Models and Neural Networks. *2020 IEEE International Conference on Smart Computing (SMARTCOMP)* (pp. 57–64). https://doi.org/10.1109/SMARTCOMP50058.2020.00028.

Zhong, C., et al. (2024). *Anonymised human location data for urban mobility research*. https://doi.org/10.13140/RG.2.2.33103.29609

Zhong, C., et al. (2025). Anonymised human location data in England for urban mobility research. *Scientific Data, 12*(1), 2040. https://doi.org/10.1038/s41597-025-06323-8

Zhu, C., et al. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software, 23*(4), 550–560. Available at: https://doi.org/10.1145/279232.279236.