Received: 22 March 2021 | Accepted: 27 October 2021

DOI: 10.1111/rssc.12532

ORIGINAL ARTICLE

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Series C Applied Statistics

The Bayesian Spatial Bradley–Terry model: Urban deprivation modelling in Tanzania

Rowland G. Seymour[1] | **David Sirl**[2] | **Simon P. Preston**[2] | **Ian L. Dryden**[2] | **Madeleine J. A. Ellis**[3] | **Bertrand Perrat**[3] | **James Goulding**[3]

[1]Rights Lab, University of Nottingham, Nottingham, UK

[2]School of Mathematical Sciences, University of Nottingham, Nottingham, Nottinghamshire, UK

[3]N/LAB, University of Nottingham, Nottingham, Nottinghamshire, UK

Correspondence
Rowland G. Seymour, Rights Lab, University of Nottingham, Nottingham, Nottinghamshire, UK.
Email: rowland.seymour@nottingham.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Numbers: EP/T003928/1, EP/R513283/1

Abstract

Identifying the most deprived regions of any country or city is key if policy makers are to design successful interventions. However, locating areas with the greatest need is often surprisingly challenging in developing countries. Due to the logistical challenges of traditional household surveying, official statistics can be slow to be updated; estimates that exist can be coarse, a consequence of prohibitive costs and poor infrastructures; and mass urbanization can render manually surveyed figures rapidly out-of-date. Comparative judgement models, such as the Bradley–Terry model, offer a promising solution. Leveraging local knowledge, elicited via comparisons of different areas' affluence, such models can both simplify logistics and circumvent biases inherent to household surveys. Yet widespread adoption remains limited, due to the large amount of data existing approaches still require. We address this via development of a novel Bayesian Spatial Bradley–Terry model, which substantially decreases the number of comparisons required for effective inference. This model integrates a network representation of the city or country, along with assumptions of spatial smoothness that allow deprivation in one area to be informed by neighbouring areas. We demonstrate the practical effectiveness of this method, through a novel comparative judgement data set collected in Dar es Salaam, Tanzania.

KEYWORDS
comparative judgement, networks, preference learning

1 INTRODUCTION

Deprivation statistics are used by governmental and non-governmental organizations to describe the standard of living in a small administrative areas (McLennan et al., 2019). Yet assessment of deprivation depends not only on the financial situation of those living in an area, but also factors such as health, housing, commercial activity and access to education. If correctly estimated, such statistics can be central to the design of successful policy interventions (see, e.g. USAID, 2019; Williams et al., 2021), supporting citizens and guiding decision makers in local government, non-governmental organizations and the business sector alike. However, obtaining deprivation estimates is often a surprisingly challenging task, particularly in developing countries. In such contexts traditional household surveys are often prohibitively expensive or logistically intractable. Data collection efforts are impaired by poor physical infrastructures restricting sample sizes. Mass urbanization can render estimates rapidly out-of-date; and a lack of financial transparency in the face of vast informal economies exacerbates the well-established response biases inherent to household surveying (Lynn & Clarke, 2002; Randall & Coast, 2015).

In Africa, according to the World Bank's chief economist, such issues have generated a 'statistical tragedy' (Devarajan, 2013). Dar es Salaam, the largest city in Tanzania, is a case in point. With a population of over 6 million the city has doubled in size in just a decade, leaving official statistics generated but 5 years ago broadly inapplicable. The United Nations has estimated that the annual growth rate of the city will continue to be 4.8%, and by 2030 Dar es Salaam will be home to at least 10 million people (United Nations Department of Economic and Social Affairs, 2019). Such rapid growth means citizens lack resources, with poor physical infrastructures and absent public services resulting in a low quality of living. Over 70% of citizens in Dar es Salaam live in unplanned settlements and (Limbumba & Ngware, 2016), water sources in the city are polluted (Napacho & Manyele, 2010) and outbreaks of diseases are common (McCrickard et al., 2017). Determining the level of deprivation in each part of this rapidly changing city is key to designing policies and strategies to alleviate these issues, especially in the face of limited resources, yet traditional household surveys are simply not viable (Randall & Coast, 2015).

Citizen science and comparative judgement offer a way to address the lack of official data and the rapid changes in the city, providing access to informed and up-to-date opinions from local citizens. Comparative judgement methods contrast sharply with traditional surveying approaches, in which a respondent might be asked to indicate the affluence level of an area, or their own household income, based upon some arbitrary scale. Instead, individuals are shown pairs of areas and asked which is the more affluent of the two. Making pairwise comparisons is preferable to making absolute judgements, which are well-evidenced as subject to strong biases and inconsistencies (Kalton & Schuman, 1982). With household income levels often being highly volatile in developing world contexts, and respondents often reticent to provide accurate responses due to the scale of the informal economy (Randall & Coast, 2015), this also provides scope to reduce response bias and logistical costs.

To achieve this one might fit a Bradley–Terry (BT) model (Bradley & Terry, 1952) to pairwise comparative judgement data. This allows not only areas to be ranked, but deprivation levels in each neighbourhood or region to be estimated. However, existing models still require an obstructively large number of individual comparisons to be provided in order to produce accurate estimates. With data collection infrastructures remaining poor in developing countries (see, e.g. Engelmann et al., 2018; van Etten et al., 2019), comparative judgement solutions can only become viable in practice if the amount of data required can be reduced. We address this key issue via development of a novel Bayesian Spatial Bradley–Terry (BSBT) model, which substantially decreases the amount of data required for reliable estimates of the parameters of interest. This model integrates a network representation of the city or country, along with assumptions of spatial smoothness that allow deprivation in one area to be informed by neighbouring areas.

Adding structure by including covariates in the standard BT model has only generally been achieved in a parametric framework with linear predictors (see, e.g. Springall, 1973; Stern, 2011). Nonparametric methods have received comparatively little attention. For example, a more flexible spline-based approach for explanatory variables has been proposed by de Soete and Winsberg (1993). A semi-parametric approach, which allows for subgroups within the set of objects being compared, has also been developed (Strobl et al., 2011). However, these methods are unsuitable for spatial explanatory variables, as it is difficult to propose covariates which can describe complex spatial structures. We instead avoid specifying any parametric functions and use a multivariate normal prior distribution to model the spatial structure. This novel treatment allows for far more flexibility as we do not need to propose strict parametric models, which often do not describe well the latent structure. We also extend the BSBT model to include ways to examine if different groups of judges (participants in the study who make the comparative judgements) hold different opinions. In developing countries, we are particularly interested in the differing opinions of men and women, as women can face starkly different health, social and economic difficulties to men. The BSBT model with judge information allows us to locate areas where men and women hold notably different opinions about the deprivation level.

1.1 | Empirical background

To demonstrate the practical effectiveness of this new method, we have additionally collected a large, novel comparative judgement data set to infer deprivation in Dar es Salaam. Ethical approval for this part of the study was obtained from the Nottingham University Business School ethical review committee, application reference No. 201819072. We include the resulting data set in the `BSBT` R package that accompanies the paper, as well as in the supplementary material. The Dar es Salaam comparative judgement data set contains 75,078 comparisons made by 224 local participants, whom we refer to henceforth as judges, as well as the gender of each judge. Dar es Salaam is divided into 452 administrative areas called *subwards*, which are the lowest level of administrative division in the city.

To carry out the judgements, we designed a web interface (see Figure 1) so that judges could be shown images of pairs of subwards and asked to compare the affluence. The interface relied on a Python back end alongside a relational database (PostgreSQL was used for the study) to collect and store comparative judgements. At the start of the study, judges were asked to identify areas of the city they were familiar with. Then, during the judging process, judges

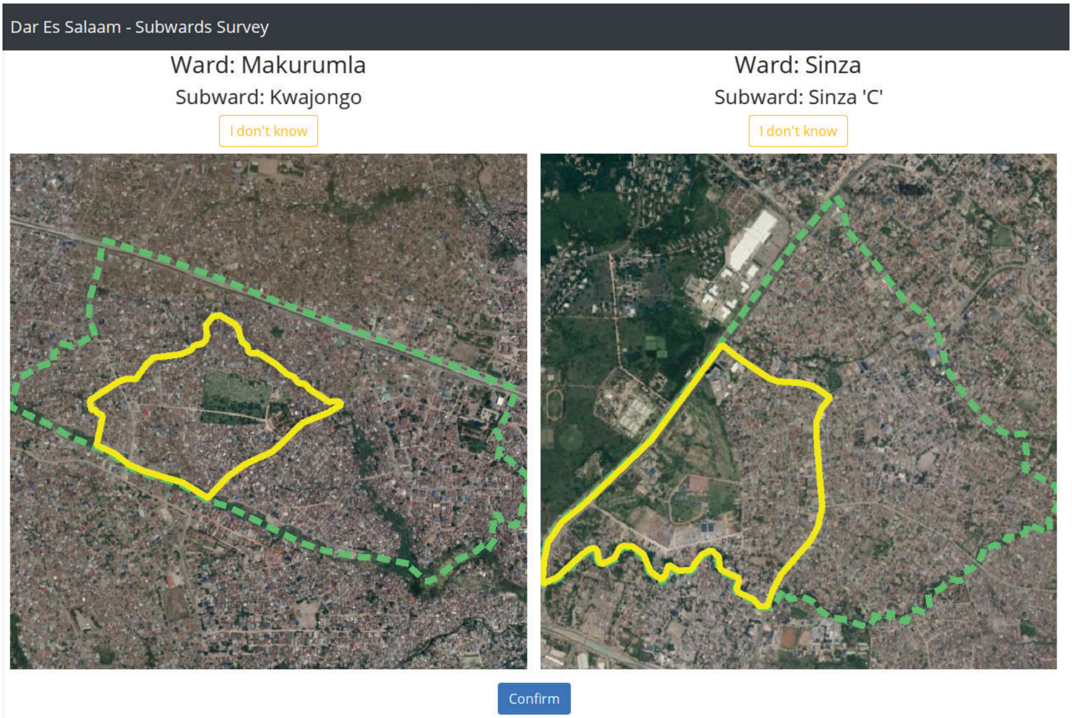

FIGURE 1 A screenshot of the software designed to carry out the comparative judgement study. In this example a user is asked to choose the most affluent between two subwards, *Kwajongo* and *Sinza C*. Images were zoomable, with both the subward and ward named directly in order to contextualize the user

had the option to indicate either (i) which of the two subwards they felt was more affluent, (ii) that the subwards were roughly equal in affluence or (iii) that they were unfamiliar with at least one of the two subwards. Comparisons corresponding to (ii) were recorded as a tie, and outcomes corresponding to (iii) were discarded and the judges were not asked about the subwards they were unfamiliar with again. Pairs of subwards for each judge were chosen uniformly at random from the list of all possible pairs of subwards which the judge was familiar with.

Judges were recruited through word of mouth by students at local universities, NGOs and also via a local taxi driver association. The rationale for recruiting these judges was that they were all citizens of Dar es Salaam with a wide working knowledge of different subwards in the city. Of the judges, by occupation 37% were students, 19% were unemployed, 13% had white-collar jobs (e.g. teacher, accountant), and the remaining 31% either had a job not in those preceding categories, or chose not to disclose this information. By gender, 40% of the judges were female, and 60% were male although male judges made 72% of the comparisons in the data set. This is because, on average, the women took longer to complete each comparison. Data were collected *in situ* over 2 weeks in August 2018 via 17 data collection sessions each lasting 2 h. Sessions were run in the morning, early and late afternoon and evening to ensure as many judges as possible could attend. Judges were only allowed to attend one session. At the start of each session, the judges received a 15 min training session in English and Swahili explaining how to make judgements and guidance on how to judge areas based on affluence and deprivation. Accompanying written instructions for the judges were provided in English and Swahili. One judge, who made over 2,000 comparisons,

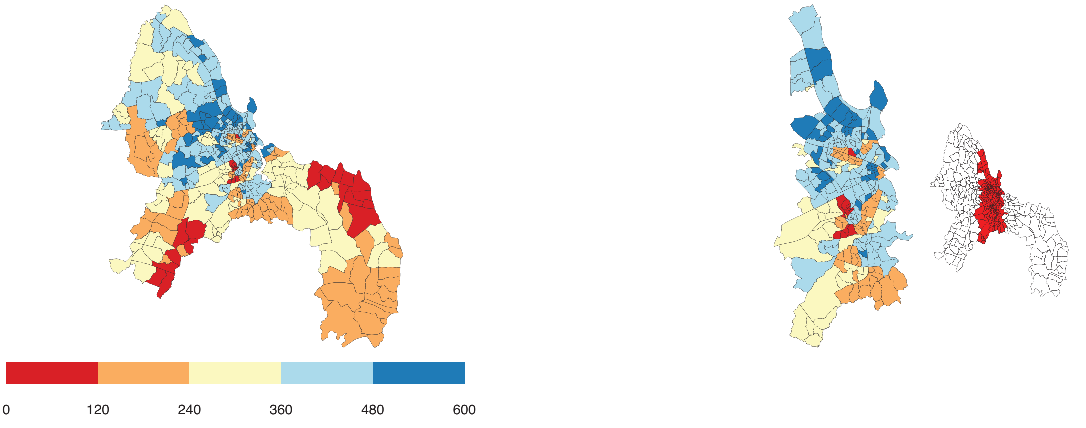

FIGURE 2 A map showing the number of times each subward featured in a comparison. The figure on the right shows a magnified section of the centre of the city, together with this central area highlighted on a map of the whole city

was excluded from the study as the comparisons seem spurious—they are not included in any of the data we report.

Figure 2 shows how many comparisons each subward was featured in, which ranged between 65 and 588 comparisons, with mean 150. The affluent areas in the city and central business district were the most well-known areas. A total of 14.6% of the comparisons made in the study were tied comparisons. There are several ways of dealing with tied comparisons (see, e.g. Davidson, 1970; Rao & Kupper, 1967; Turner & Firth, 2012) and we discuss these in Section 4.1. We chose to randomly assign one of the pair to be the more deprived subward.

An important aim of this work is to develop methodology that enables more efficient data collection, able to overcome the organizational challenges faced in the field. Two weeks were invested in collecting this large data set, and organization and recruitment of participants prior to the study took several months. A key aim of the scale of the data collection process undertaken was to conclusively evidence, with similar future efforts in mind, that the first 2 days could have been sufficient if improved modelling procedures are employed. This would save considerable time and resources in both collecting the data and reducing the number of participants needed to recruit, train and organize.

2 | A SPATIAL FRAMEWORK FOR THE BRADLEY–TERRY MODEL

2.1 | The standard Bradley–Terry model

Consider a comparative judgement data set consisting of K pairwise comparisons of N areas. We assign to each area what we call a *relative deprivation parameter* $\lambda_i \in \mathbb{R}$ $(i = 1, \ldots, N)$ and infer the value of each parameter using a comparative judgement model. We use the term ‘deprivation parameter’ because identifying deprivation is the primary focus of the paper, but note that, in keeping with most measures of this kind, larger (respectively, smaller) values are associated with more affluent (respectively, deprived) areas.

We begin by outlining the standard BT model, a commonly used comparative judgement model. If areas i and j are compared n_{ij} times, the number of times area i is judged to be more affluent than area j is modelled as

$$Y_{ij} \sim \mathrm{Bin}(n_{ij}, \pi_{ij}),$$

and we assume Y_{ij} are independent. Here the probability π_{ij} that area i is judged to be more affluent than area j depends on the difference in relative deprivation of i and j and is

$$\mathrm{logit}(\pi_{ij}) = \lambda_i - \lambda_j \Longleftrightarrow \pi_{ij} = \frac{\exp(\lambda_i)}{\exp(\lambda_i) + \exp(\lambda_j)} \qquad (i \neq j, 1 \leq i, j \leq N). \tag{1}$$

Model (1) is invariant to translations $\lambda_i \longrightarrow \lambda_i + c$ (for any $c \in \mathbb{R}$), so an identifiability constraint is needed. A common choice is $\sum_{i=1}^{N} \lambda_i = 0$, which means that relatively deprived areas will have negative parameters, relatively affluent areas will have positive parameters and areas with middling levels of relative deprivation will have parameters near zero.

We write y_{ij} for the number of times area i was judged to be more affluent than area j, and denote by $\boldsymbol{y}$ the set containing these outcomes for all pairs of areas. The likelihood function for the model is given by

$$\pi(\boldsymbol{y} \mid \lambda_1, \ldots, \lambda_N) = \prod_{i=1}^{N} \prod_{j<i} \binom{n_{ij}}{y_{ij}} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{n_{ij} - y_{ij}}. \tag{2}$$

We will compare our model to the standard BT model and the implementation provided in the `BradleyTerry2` R package (Turner & Firth, 2012), as this is a popular implementation of the method (see, e.g. Cattelan, 2012; Grinfeld et al., 2018; Varin et al., 2016). This package computes MLEs for the model parameters. We follow Turner and Firth (2012) and Firth and De Menezes (2004) and construct 95% confidence intervals using the quasi-variance for the estimates in the standard BT model. This is done using the `qvcalc` package (Firth, 2020), as this allows us to readily compare the relative deprivation levels.

2.2 | The Bayesian Spatial Bradley–Terry model

In the BSBT model, we assume the relative deprivation parameters λ_i to be random and dependent on one another, with a higher level of dependence between nearby areas than areas which are further apart. To avoid making specific parametric assumptions about the level of deprivation in each area, we model the relative deprivation parameters using a multivariate normal prior distribution. We use a zero-mean multivariate normal prior distribution for the deprivation level parameters $\lambda = \{\lambda_1, \ldots, \lambda_N\}$ subject to the constraint $\mathbf{1}^T\lambda = 0$, where $\mathbf{1} = (1, \ldots, 1)^T$ is a vector of ones. This matches the condition in the standard BT model, that the sum of the deprivation levels is 0. Conditional on this constraint

$$(\lambda \mid \mathbf{1}^T\lambda = 0) \sim \mathrm{MVN}\left(\mathbf{0}, \Sigma - \Sigma\mathbf{1}(\mathbf{1}^{\mathrm{T}}\Sigma\mathbf{1})^{-1}\mathbf{1}^{\mathrm{T}}\Sigma\right). \tag{3}$$

2.2.1 | Modelling spatial covariance

The structure of the covariance matrix Σ is a modelling choice and there are number of options to choose from. In the simplest terms, we want to assign high covariance between deprivation levels in nearby subwards and low covariance between levels in distant subwards. A widely used option in Euclidean spatial domains is to use the squared-exponential covariance function (Rasmussen & Williams, 2006). Using this function, the covariance between the deprivation levels in subwards i and j is

$$\mathrm{cov}(\lambda_i, \lambda_j) = \Sigma_{i,j} = k(i, j; \alpha, l) = \alpha^2 \exp\left(-\frac{d_{ij}^2}{l^2}\right), \tag{4}$$

where d_{ij} is the Euclidean distance between areas i and j, α^2 is the prior variance hyperparameter and l is the characteristic length scale, which describes what is meant by nearby and distant. However, using a function which is stationary in Euclidean space may not capture the change in deprivation in all parts of the city. City centres are typically densely packed with small areas, with peri-urban and rural areas being larger. Modelling the spatial structure using a Euclidean metric is therefore unsuitable since, for example, two points 1km apart in a rural area are likely similar, but two points 1km apart near a city centre may be very different.

Urban regions are typically divided into sub-areas for administrative purposes, and these neighbourhoods often provide natural units over which to quantify deprivation. While spatially connected, such areas often vary greatly in size. In this paper, we model an urban region as a network, whereby these low-level areas are represented as nodes with edges joining neighbouring areas, such that we can use a network-based (i.e. non-Euclidean) distance to define spatial 'closeness' between pairs of areas when defining prior assumptions of spatial smoothness. Using a network metric allows us to model nonstationary structures. We therefore begin by transforming the set of areas into a network by treating each area as a node and placing edges between adjacent areas; some modelling choices are required when dealing with noncontiguous areas or islands. In the Dar es Salaam network, we add two additional edges over the Kurasini creek to reflect the high-volume road and ferry connections. Figure 3 shows a map of Dar es Salaam and the corresponding network.

We can adapt the squared-exponential covariance function in Equation (4) for use with a network by letting d_{ij} be the distance of the shortest path between subwards i and j. The shortest distance can be computed using Dijkstra's algorithm (see, e.g. Cormen et al., 2001). Although using a network reduces the issue of stationarity, specifying the value of the length scale still may be challenging or restrictive; instead, when using the rational quadratic covariance function, which is a mixture of squared-exponential covariance functions with different length scale values, we can specify the relative importance of long and short scale variation in deprivation. Another option is to use the Matérn covariance function, which would remove the assumption that the spatial structure is smooth. However, when using shortest-path distances in Equation (4), the resulting matrix is not guaranteed to be positive semi-definite and we may need to project the matrix into the space of covariance matrices. This can be done in a number of ways, including setting the negative eigenvalues to 0 or modifying the polar decomposition (Higham, 1988).

Instead of using a distance-based approach, we can construct the covariance matrix directly from the structure of the network. Estrada and Higham (2010) describe several options for quantifying the 'communicability' between two nodes of a network in terms of functions of the adjacency

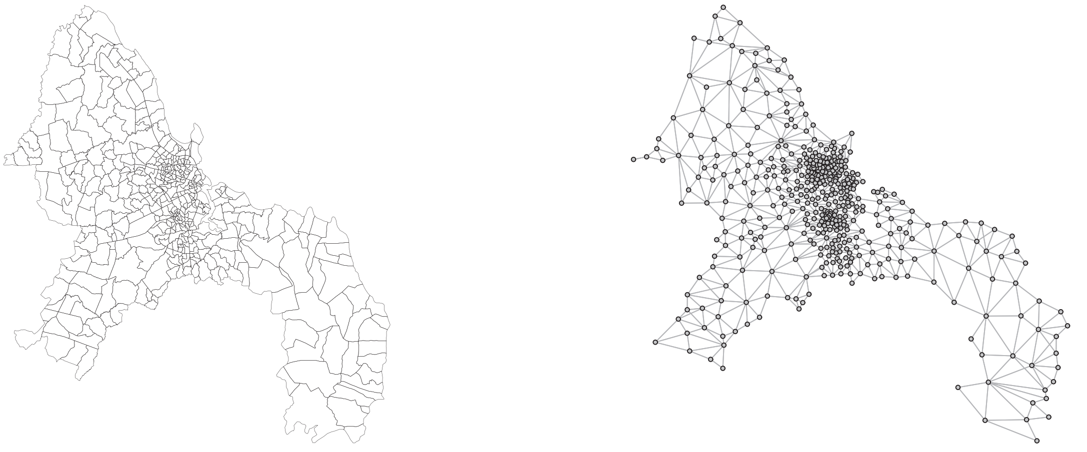

FIGURE 3 A map and the network representation of subwards in Dar es Salaam

matrix of the network. The option we choose is based on the matrix exponential of the adjacency matrix as this measure emphasizes connectedness over short distances rather than long distances to a greater extent than the alternatives described in Estrada and Higham (2010). Let $\Lambda = e^A$, where A is the network's adjacency matrix, and let D be a diagonal matrix containing the elements on the diagonal of Λ. The covariance matrix is given by

$$\Sigma = \alpha^2 D^{-\frac{1}{2}} \Lambda D^{-\frac{1}{2}}, \tag{5}$$

where α^2 is a hyperparameter describing the variance in the deprivation levels. The matrix Σ therefore has diagonal entries α^2 and off-diagonal entries proportional to the communicability of each pair of subwards in the network. We thus achieve our aim of assigning higher covariance between better-connected pairs of subwards, using a natural characterization of the network. Although we use the matrix exponential covariance matrix in the paper, we find no discernible difference in the results of Sections 3 and 4 when using the (network-adapted) squared-exponential covariance function.

2.3 | Incorporating judge information

We now incorporate judge covariates into the model as this avoids the assumption that the judges act homogeneously. Suppose there are G groups of judges and let $\boldsymbol{x}_g$ be the vector of length P containing the covariates for group g. We assume judges in the same group act homogeneously. The vector $\boldsymbol{x}_g$ may contain categorical, discrete or continuous covariates or a mixture of all three; for a categorical covariate we represent the q levels of the covariate by q indicator functions. If $\boldsymbol{x}_g$ contains categorical covariates the number of groups may be small, but if $\boldsymbol{x}_g$ contains a continuous covariate each judge may be its own group.

We model the deprivation in area i, as perceived by judges in group g, to be

$$\lambda_i^g = \lambda_i + \sum_{p=1}^{P} x_p^g \beta_{pi},$$

where x_p^g is the p^{th} element of $\boldsymbol{x}_g$ and β_{pi} is the parameter corresponding to x_p^g and area i, where $i = 1, \ldots, N$. Modifying the likelihood function in Equation (2) to take account of the contributions from each group of judges, we now have the likelihood function

$$\pi(\boldsymbol{y} \mid \lambda, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_P) = \prod_{g=1}^{G} \prod_{i=1}^{N} \prod_{j<i} \binom{n_{ijg}}{y_{ijg}} \pi_{ijg}^{y_{ijg}} (1 - \pi_{ijg})^{n_{ijg} - y_{ijg}}, \tag{6}$$

where n_{ijg} is the number of times judges in group g compared areas i and j, y_{ijg} is the number of times judges in group g judged area i to be more affluent than area j, π_{ijg} is the probability judges in group g judge area i to be more affluent than area j and is given by $\text{logit}(\pi_{ijg}) = \lambda_i^g - \lambda_j^g$, and $\boldsymbol{\beta}_p = \{\beta_{p1}, \ldots, \beta_{pN}\}$ is the set of parameters corresponding to p^{th} element of the set of judge covariates. We recover the model and likelihood of Section 2.2 by taking $G = 1$ and $P = 0$ in this formulation.

As in the BSBT model with no judge covariates, we place a constrained multivariate normal prior distribution on the spatial parameters λ, shown in Equation (3). We also place an independent, constrained, multivariate normal prior distribution on each $\boldsymbol{\beta}_p$ which allows us to model the effect of each covariate spatially. So that the deprivation parameters, λ, represent the grand mean of the deprivation for all judges, we enforce a second constraint among the set of parameters $\boldsymbol{\beta}_p$, which correspond to a given categorical covariate, as this allows us to treat each category symmetrically, that is, we avoid fixing one category as a reference category and then not having any uncertainty associated with it. For a group of q covariates representing the q categories of covariate p, the corresponding parameters $\beta_{p_1 i}, \ldots, \beta_{p_q i}$ need a constraint to ensure identifiability. We use $\beta_{p_1 i} + \cdots + \beta_{p_q i} = 0$ for each area $i = 1, \ldots, N$.

An example of including judge information is investigating how judges of different genders view different subwards. In less developed countries, women may be more vulnerable to different forms of exploitation than men (e.g. female genital mutilation, modern slavery and forced marriage) and finding areas women view as more deprived than men may indirectly give information about where these practices are happening. We sort the judges into two groups (i.e. $G = 2$), men and women. We let $\boldsymbol{x}_1^T = (1\ \ 0)$ for male judges and $\boldsymbol{x}_2^T = (0\ \ 1)$ for female judges (i.e. $P = 2$). The appropriate constraint to ensure identifiability is then $\beta_{1i} + \beta_{2i} = 0$ for each area i.

2.4 | Fitting the model

Now we have described the BSBT model, we develop a Markov chain Monte Carlo (MCMC) algorithm to infer the model parameters given the observed comparative judgements $\boldsymbol{y}$, and the judge covariates $\boldsymbol{x}$. The model parameters are: the deprivation parameters λ, any covariate parameters $\boldsymbol{\beta}_p$, and the covariance function variance hyperparameters α_λ^2 and $\alpha_1^2, \ldots, \alpha_P^2$. By Bayes' theorem, the posterior distribution is

$$\begin{aligned}\pi(\lambda, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_P, \alpha_\lambda^2, \alpha_1^2, \ldots, \alpha_P^2 | \boldsymbol{x}, \boldsymbol{y}) &\propto \pi(\boldsymbol{y} | \lambda, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_P) \pi(\lambda | \alpha_\lambda^2,\ \mathbf{1}^T \lambda = 0) \pi(\alpha_\lambda^2) \\ &\quad \times \prod_{p=1}^{P} \pi(\boldsymbol{\beta}_p | \alpha_p^2,\ \mathbf{1}^T \boldsymbol{\beta}_p = 0) \pi(\alpha_p^2).\end{aligned} \tag{7}$$

The first term on the right-hand side is the likelihood function (6) and the second term is the prior density for the spatial component λ, for which we use the constrained prior distribution (3). We place an independent prior distribution on the variance hyperparameter α_λ^2, which is the third term on the right-hand side. The product term contains the prior distributions for the covariate parameters $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P$ and the variance hyperparameters $\alpha_1^2, \dots, \alpha_P^2$ for these distributions.

The posterior density cannot be computed explicitly, but it can be sampled from using Algorithm 1. This MCMC algorithm involves iterating Gibbs updates for the variance hyperparameters, $\alpha_\lambda^2, \alpha_1^2, \dots, \alpha_P^2$, and Metropolis–Hastings updates for the spatial components, $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P$. For analytical convenience, we place a conjugate inverse-Gamma prior distribution on the variance hyperparameters, the density function of which is

$$\pi(x; \chi, \omega) = \frac{\omega^\chi}{\Gamma(\chi)} \frac{1}{x^{\chi+1}} \exp\left(-\frac{\omega}{x}\right) \qquad (x > 0; \chi > 0, \omega > 0).$$

The Gibbs updates are possible because the full conditional distribution for α_λ^2 has a closed form. It is given by

$$\alpha_\lambda^2 \,|\, \boldsymbol{\lambda} \sim \text{inv} - \Gamma\left(\chi + \frac{N}{2}, \omega + \frac{1}{2}\boldsymbol{\lambda}\overline{\Sigma}^{-1}\boldsymbol{\lambda}^T\right),$$

where $\overline{\Sigma}$ is the covariance matrix of the constrained prior with $\alpha^2 = 1$ in Equation (5). Analogously, the full conditional distribution for α_p^2 is

$$\alpha_p^2 \,|\, \boldsymbol{\beta}_p \sim \text{inv} - \Gamma\left(\chi + \frac{N}{2}, \omega + \frac{1}{2}\boldsymbol{\beta}_p\overline{\Sigma}^{-1}\boldsymbol{\beta}_p^T\right),$$

To update the deprivation parameters, $\boldsymbol{\lambda}$, we use a Metropolis–Hastings sampler with an underrelaxed proposal mechanism (Neal, 1998). This allows us to update the parameters as a block and reduces the computational complexity compared to updating each deprivation parameter individually. Given the current deprivation parameters $\boldsymbol{\lambda}$, we propose new values by

$$\boldsymbol{\lambda}' = \sqrt{1-\delta^2}\boldsymbol{\lambda} + \delta\boldsymbol{v},$$

where $\delta \in (0, 1]$ is a tuning parameter and $\boldsymbol{v}$ is a draw from the constrained prior distribution in Equation (3). We accept this proposal with probability

$$p_{\text{acc}} = \min\left(\frac{\pi(\mathbf{y}|\boldsymbol{\lambda}', \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P)}{\pi(\mathbf{y}|\boldsymbol{\lambda}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P)}, 1\right).$$

Algorithm 1 MCMC Algorithm for the BSBT Model

1: Choose initial values for $\boldsymbol{\lambda}$, $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P$ and $\alpha_\lambda^2, \alpha_1^2, \dots, \alpha_P^2$.
On iteration j of the MCMC algorithm do
2: Update $\boldsymbol{\lambda}_i$ using a Metropolis-Hastings step with underrelaxed proposal mechanism;
3: Update $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P$ using a Metropolis-Hastings step with underrelaxed proposal mechanism;
4: Update α_λ^2 using a Gibbs step;
5: **for** i in 1 to P **do**
6: Update α_i^2 using a Gibbs step;
7: **end for**

The proposal ratio using the underrelaxed proposal mechanism is the inverse of the prior ratio, meaning the acceptance probability is the ratio of the likelihood function with the proposed and current deprivation parameters. We follow an analogous process for the covariate parameters $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_P$.

2.5 | Implementing the model

We have developed an R package to allow any user to implement this method on a comparative judgement data set. The package `BSBT` is available on CRAN (Seymour & Briant, 2021). It includes the novel comparative judgement data set on deprivation in Dar es Salaam, Tanzania, shapefiles for the 452 subwards in the city and a vignette containing instructions on how to reproduce the analysis in Section 4. The package allows users to place a constrained multivariate normal prior distribution for deprivation parameters over a predetermined network (it also facilitates constructing the network) and fit the model using the MCMC algorithm in Algorithm 1. We provide a number of covariance functions, including the squared-exponential, Matérn and matrix exponential functions. The MCMC functions included in the package can be used to fit either a spatial model, or a spatial model with a single covariate for judge information. Due to our formulation of the likelihood function, the computational time for the BSBT implementation scales according to the number of areas, whereas the implementation provided in the `BradleyTerry2` package scales according to the number of comparisons.

3 | SIMULATION STUDY

To assess the model's ability to infer deprivation levels in a realistic scenario, we simulate deprivation levels for the subwards in Dar es Salaam by drawing a sample from the prior distribution, then seek to infer these from simulated comparative judgements. A map of the city and the corresponding network are shown in Figure 3. We simulate the comparisons according to the model in Equation (2) and choose pairs of areas uniformly at random to compare. We simulate data sets of various sizes to mimic real data collection. The sizes of simulated data sets used in this paper are shown in Table 1. We use 'judge hours' to quantify the number of comparisons by the total judging time required, assuming 20 s per comparison or 180 comparisons per judge hour.

We fit the model to each data set, running the MCMC algorithm for 1,500,000 iterations and removing the first 500,000 iterations as a burn-in period. We fix the tuning parameter $\delta = 0.01$, based on initial runs of the algorithm. For the prior distribution on α_λ^2, we fix $\chi = \omega = 0.1$ which results in a somewhat noninformative distribution (Gelman, 2006). To assess the model fit, we compute the mean absolute error (MAE) for the result of each set of comparisons, which is given by

TABLE 1 Data set sizes used in the simulation studies, using 180 comparisons per judge hour

Judge hours	1	2	5	10	25	50	100	250	500	1,000
Comparisons	180	360	900	1,800	4,500	9,000	18,000	45,000	90,000	180,000

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\lambda_i - \hat{\lambda}_i|,$$

where $\hat{\lambda}_i$ is the estimate corresponding to the MLE or posterior mean for area i.

Figure 4 shows the log MAE for each data set. The BSBT model outperforms the standard model for all sizes of data set used. For a fixed number of comparisons, the BSBT model has smaller error than the standard model. For example, when using 1,800 comparisons (10 judge hours), the MAE using the BSBT model (0.260) is less than a third of the error in the standard model (0.907). Figure 4 also shows that we can substantially reduce the number of comparisons required to achieve a given level of error by using the BSBT instead of the standard BT model. For example, MAE in the BSBT with 5 judge hours is similar to that in the standard model with 50 judge hours, a decrease in judge hours of 90%; and 250 judge hours with the standard model yields similar MAE to 100 judge hours with BSBT, a still substantial reduction of 60% in terms of the data required to give a similar level of performance. For small data sets we are unable to compute the MLE for all areas and so the corresponding MAE is undefined for the standard BT model. Here we see one of the main advantages of the BSBT model: including weak prior assumptions about spatial correlations allows it to learn about areas featured in very few, or even no, comparisons from information about nearby areas.

We observe that the performance of the BT and BSBT models is very similar when the number of judgements is large. This is to be expected from the Bernstein–von Mises theorem (Kleijn & van der Vaart, 2012) whereby the posterior distribution of finite dimensional parameters and the MLEs tend to the same asymptotic multivariate normal distribution for large samples, subject to smoothness and identifiability conditions on the prior distribution and a positivity condition on the prior at the true value.

We also present a simulation study on a synthetic 1-d 'city' in Section 1 of the supplementary material. Although less realistic than the 2-d study above, it has the significant advantage

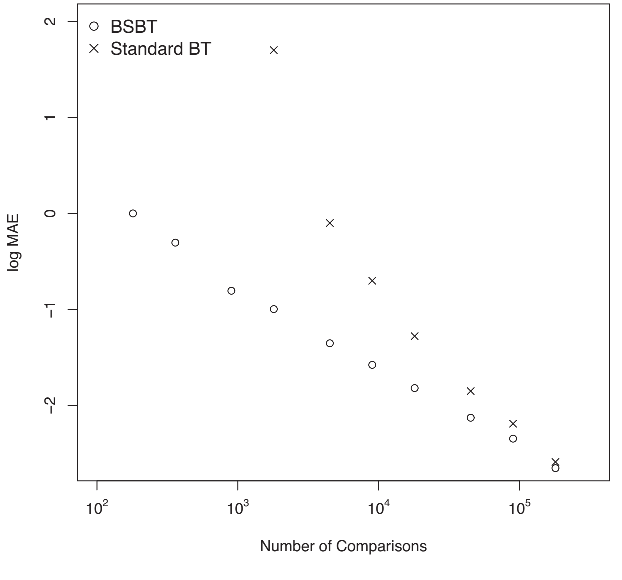

FIGURE 4 Log MAE for the simulation study comparing performance of the standard BT and the BSBT models in terms of error as a function of the number of comparisons

of allowing much easier visualization of the synthetic ground truth, the simulated data and the results of fitting our model; aiding interpretation of what our methods achieve.

4 | DEPRIVATION IN DAR ES SALAAM, TANZANIA

4.1 | Bayesian Spatial Bradley–Terry model

We fit the BSBT model to the data and run the MCMC algorithm shown in Algorithm 1 for 1,500,000 iterations, removing the first 500,000 iteration as a burn-in period. This took around 3 h on a 2019 iMac with a 3 GHz CPU. We examined trace plots to ensure adequate mixing of the Markov chain and to choose the length of the burn-in period. These are given in Section 2 of the supplementary material. We fix the tuning parameter $\delta = 0.01$, based on initial runs of the algorithm, and the inverse gamma prior distribution parameters $\chi = \omega = 0.1$.

The resulting estimates for the level of deprivation in each subward in the city are shown in Figure 5. We see a north–south trend, whereby the level of deprivation increases further south in the city. We find several sharp changes in deprivation in the city centre, where slums neighbour affluent subwards. The most affluent subward is Masaki, and the 10 most affluent areas are all concentrated around the Masaki peninsula directly north of the city centre and home to most of the affluent expatriate communities. The 10 most deprived subwards are geographically spread out, with one, Mpakani, being located in the centre of Dar es Salaam and the others spread across the outer regions of the city. Four of the 10 most deprived subwards are in the Somangila ward, a coastal ward in the east of the city.

The uncertainty in the estimates for the level of deprivation in each subward differs considerably, as shown in Figure 6. We see a correlation between the level of uncertainty in our estimate and the estimated level of deprivation. As the most affluent areas tend to also be well known areas, such as tourist resorts or areas with government buildings, we were able to collect more

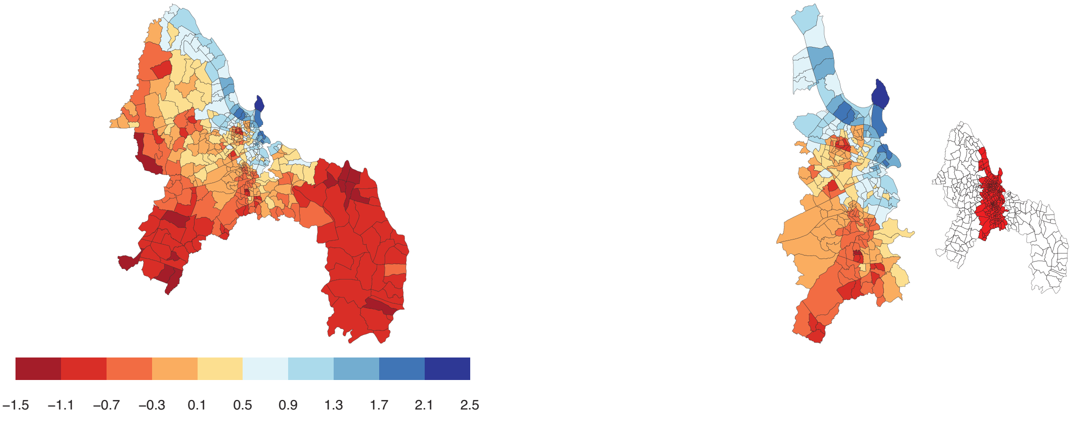

FIGURE 5 The posterior mean values for the BSBT model applied to the Dar es Salaam data set. The figure on the right shows a magnified section of the centre of the city, together with this central area highlighted on a map of the whole city

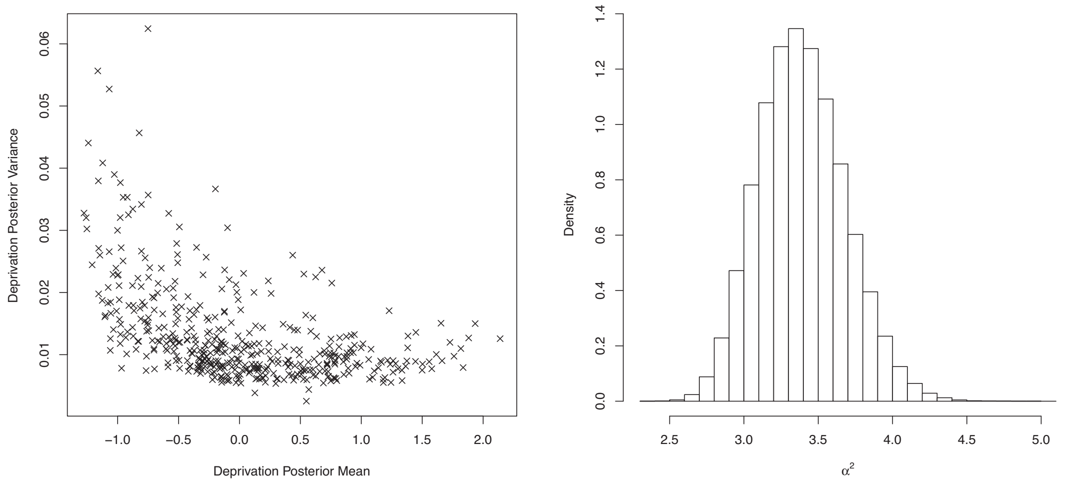

FIGURE 6 Uncertainty in the fitted BSBT model for the Dar es Salaam data set. Left: The estimated level of deprivation in each subward against the uncertainty in the estimate. Right: The posterior distribution for the variance parameter α_λ^2

comparisons involving these subwards and therefore there is less uncertainty in our estimates for the deprivation in these areas. We also estimate the variance parameter α_λ^2; its posterior mean is 3.378 with 95% CI (credible interval) (2.868, 3.993) and the posterior distribution is shown in Figure 6. Section 2 of the supplementary material gives more diagnostic information and a short investigation of judge reliability which concludes that no judges provide judgements which are notably out of line with the fitted model.

Because approximately one in seven of the comparisons in the data set are tied, which is a substantial proportion, we must take care that our approach to treating ties does not substantially affect the inferred deprivation levels. For the results in this paper, wherever a comparison was tied we randomly allocated a winner. In Section 3.1 of the supplementary material, we carry out a sensitivity analysis of these random allocations, examining 20 data sets generated via different random seeds, and confirm the robustness of our results. In Section 3.2 of the supplementary material, we consider two alternative treatments for the tied comparisons (treating a tie as 'half a win' for both subwards involved, and discarding the ties altogether). We found the posterior means were largely unaffected by the treatment of ties. Discarding the ties increases the uncertainty as we are discarding a considerable amount of data, and treating the ties as half a win yields estimates that have the lowest variance of any treatment we considered. We have favoured the treatment of allocating winners of the tied comparisons at random. This is on the basis that the results appear insensitive to the specific random allocation used, it makes use of all the available data, and it is conservative in terms of the resulting uncertainty in parameter estimates.

Results for the standard BT and BSBT models are very similar; we see very similar inferred deprivation levels and uncertainties. (See Section 4 of the supplementary material.) However, the data set that we have is quite large, so this is likely a data saturation effect (cf. Figure 4). An important aim of our work is to investigate if many fewer comparative judgements could have been collected, at a much reduced cost, with little loss of information.

4.2 | Efficiency of the BSBT model

To investigate the effectiveness of the model when we have a smaller number of comparisons, we also fit both the standard BT and BSBT models to the comparisons collected on the first 2 days of the field work. This subset includes 13,361 comparisons (around 18% of the original data set). All subwards feature in this partial data set and the number of comparisons each subward was featured in ranged from 2 to 233, with mean 60. Five subwards 'lost' every comparison they were featured in, making it difficult to estimate their deprivation level using the standard BT model. We compute the MAE taking the true values to be the inferred deprivation levels using the full data set. Using the BSBT model on this partial data set roughly halves the MAE compared to the standard BT model, reducing it from 0.523 to 0.267. We are still able to identify sharp changes in deprivation levels, for example where slums neighbour affluent areas in the city centre.

In Figure 7, we report the posterior mean and variance for the deprivation in each subward. There is some shrinkage in the estimates for the most deprived subwards, but no consistent change elsewhere. There is strong linear correlation between the estimated deprivation levels using the full and partial data set ($\rho = 0.832$), showing that in terms of identifying subwards as, for example, somewhat affluent or very deprived, very little is lost by using the partial data set. As is expected, using less data results in higher uncertainty, however, the uncertainty is generally small with respect to the deprivation parameter values and the additional uncertainty does not appear to apply to subwards in any systematic way. Alongside the analysis shown in Figure 4, this shows that by using the BSBT model, in future we can collect far fewer comparisons yet attain similar levels of error in the results. This will reduce the time and cost associated with data collection in similar future fieldwork.

4.3 | Judge information in Dar es Salaam

First, we investigate if the men and women in the study perceived subwards differently. For the Dar es Salaam data, there were 91 female judges and 133 male judges. For reasons outlined

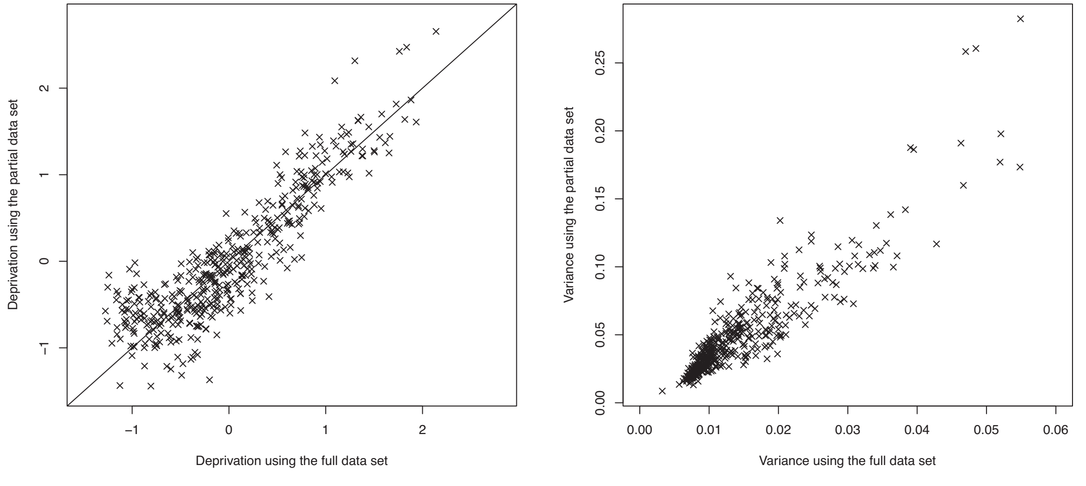

FIGURE 7 Left: The posterior mean estimates for deprivation using the full and partial data sets. Right: The posterior variance using the full and partial data sets

in the introduction we are interested in determining whether different genders have different perceptions of some parts of the city. Our first observation is that each male judge did on average 328 comparisons, whereas the average among female judges was 200. This is because the women took longer to carry out individual comparisons than the men. Another difference is that the women tended to be familiar with fewer subwards than the men, perhaps suggesting they are less mobile in the city. We fit the BSBT model with gender effect to the data, here $G = 2$ as we sort comparisons into two groups (men and women) and $P = 2$ as we model the effect of being male or female. We run the MCMC algorithm for 5,000,000 iterations, which took 1 day on a 2019 iMac with 3 GHz CPU. Diagnostic plots can be found in Section 5 of the supplementary material. We fix $\delta = 0.01$ based on initial runs of the algorithm. We estimate the variance α_λ^2 (for $\boldsymbol{\lambda}$) to be 3.846 (95% CI: (3.073, 3.694)) and α_1^2 (for $\boldsymbol{\beta}_1$) 0.026 (95% CI: (0.002, 0.034)). Such a small value of α_1^2 suggests the men's and women's perceptions are highly correlated.

Figure 8 shows the distribution of the posterior mean deprivation levels perceived by men and women. We see that the distribution of the levels of deprivation perceived by men and women is largely the same. We also show posterior density estimates for men's and women's perceptions of two subwards. In Kibonde Maji A, a somewhat deprived subward in the south of the city on a trunk road, there is no perceptible difference in how men and women perceive the subward. In Hananasif, an inner city subward near the business district, women perceive the subward to be considerably more deprived than men do. In Figure 9 we show the spatial structure in the difference between how men and women view the subwards, based on whether or not CIs for the discrepancies $\beta_{0,i}$ (for each subward i) contain zero. The subwards women view as more deprived than men are mostly concentrated in the centre of the city, and the majority of the subwards which women view as less deprived are in the outer regions of the city. We suggest two reasons for the difference in perceptions: the first is personal safety, as women may perhaps feel less safe in the city centre; the second is because the centre is the location of both the central business district and many nightlife venues, which may offer better opportunities to men.

Second, we investigate if students perceived deprivation differently to the other judges. Students made up 37% of the judges and made 41% of the comparisons. We fit the BSBT model with $P = 2$, as there are two groups of judges (students and non-students). As in the gender differences model, we run the MCMC algorithm for 5,000,000 iterations. We run the MCMC algorithm for 5,000,000 iterations, which took 1 day on a 2019 iMac with 3 GHz CPU. Diagnostic plots can be found in Section 5 of the supplementary material. We find there is no difference between how the students and non-students perceive deprivation in the city; all 95% CIs for the discrepancy between students and non-students contain 0. The mean absolute discrepancy is 0.016 and the maximum absolute discrepancy is 0.035; indicating very little difference between the two groups. We estimate the variance α_λ^2 (for λ) to be 4.953 (95% CI: (3.982, 6.052)) and α_1^2 (for $\boldsymbol{\beta}_1$) 0.005 (95% CI: (0.004, 0.007)). We note that the posterior mean estimate for the variance for the discrepancy parameter is an order of magnitude smaller than in the gender discrepancy results, further suggesting the students and non-students have very highly correlated responses.

5 | DISCUSSION

In this paper we have developed a nonparametric spatial version of the Bradley–Terry model and fitted it to a novel data set to infer deprivation levels in Dar es Salaam, Tanzania. Our methods also allow us to incorporate judge information into the model, for example, judge gender or occupation, to understand the perceptions of different groups of judges.

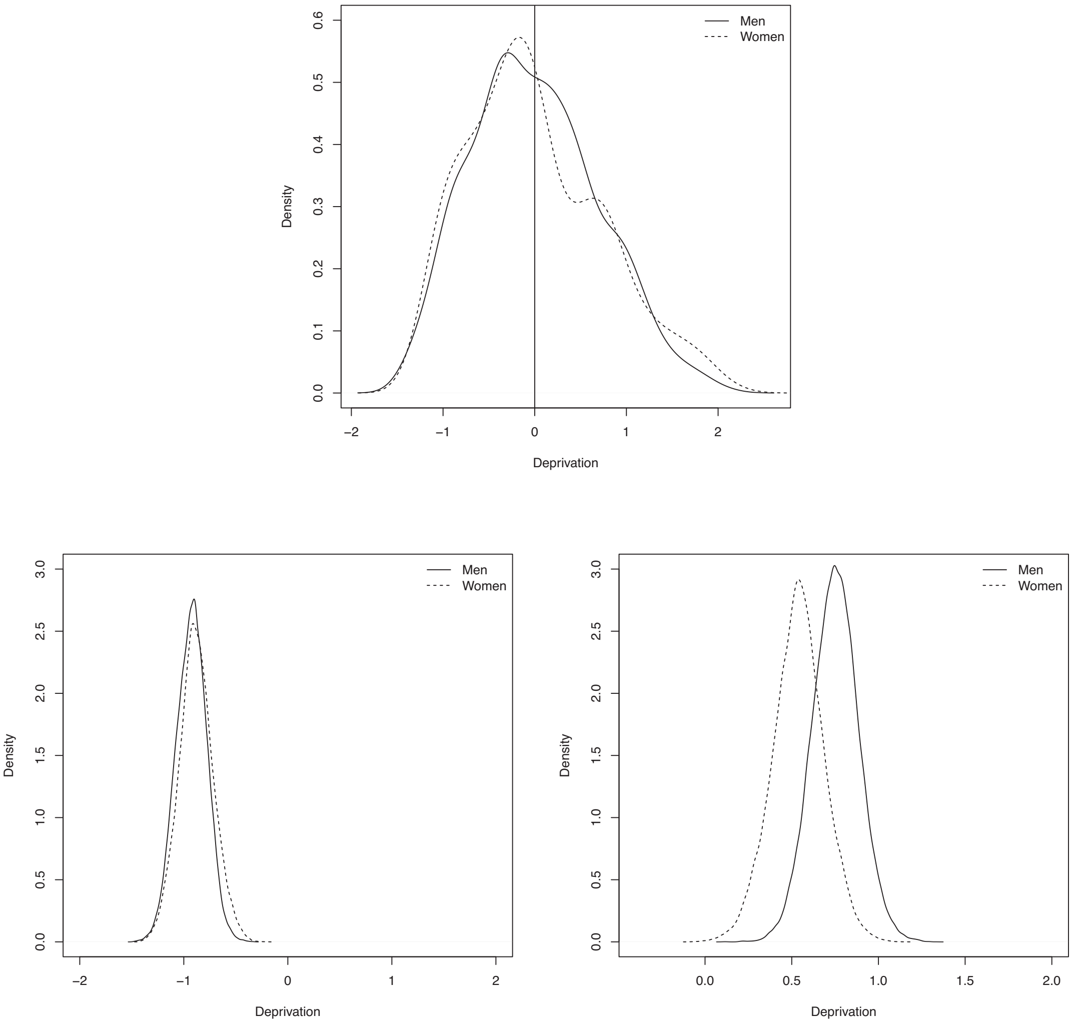

FIGURE 8 Top: Kernel density estimation of posterior means for the levels of deprivation given by men and women. Bottom: The posterior distributions for the men's and women's perceptions of Kibonde Maji A (left) and Hananasif (right). We do not infer a difference between how men and women perceive Kibonde Maji A, but we do for Hananasif

We analysed a novel data set on deprivation in Dar es Salaam, not only estimating the level of deprivation in the city's 452 subwards, but also demonstrating the effectiveness of the BSBT model in significantly reducing data requirements by incorporating spatial correlations in the prior distribution for deprivation levels. As far as we are aware, no estimates for deprivation on such a fine scale are currently available. We were able to identify slums in the centre of the city and estimate the level of deprivation in the peri-urban outer regions of the city. Our findings show that there are several sharp changes in the level of deprivation in the centre of the city where very affluent areas neighbour slums. There is also a difference in how men and women view some areas; specifically we find that women view some parts of the centre of Dar es Salaam as more deprived than men do, but tend to view some parts of the outer regions of the city as less deprived than men do. Our data collection, modelling and analysis provides up-to-date

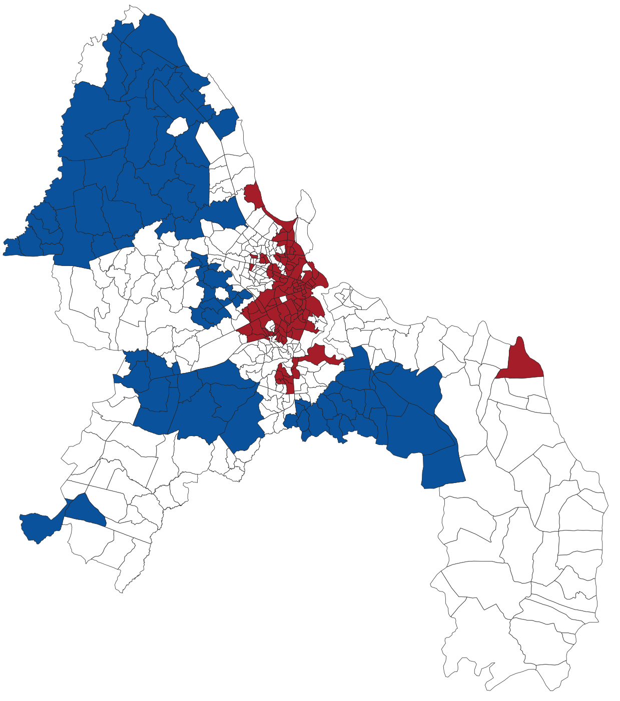

FIGURE 9 The difference between men and women's perceived deprivation levels, coloured by the 95% credible interval. Blank subwards have a credible interval which contains 0. Red and blue subwards have a credible interval which does not contain 0, with blue showing positive and red showing negative

estimates of deprivation levels in Dar es Salaam via the involvement of over 200 of the citizens of the city.

There is scope for agencies in developing countries to use the BSBT model to design interventions based on a quantitative analysis of social issues. This is advantageous to agencies working in environments where official statistics are low quality or not available. This is not limited to deprivation but any social issue that citizens can compare areas on, for example estimating prevalence of Female Genital Mutilation, or prevalence of black market trading. Similarly, such studies need not be limited to cities, but any context which has a spatial or network component; for example a group of villages spread out across a large area or a network of individuals linked by telecommunications data.

There are a number of possible directions in which the BSBT model may be fruitfully extended and further explored. The BSBT model has a large computational cost and there is scope to reduce the computational time required by developing a more efficient MCMC algorithm, for example by adaptive updating of the under-relaxed tuning parameter δ. We could further reduce the amount of data required by optimizing the experimental design and identifying pairs of areas which should be asked about or adaptively identifying areas which need to be compared (see, e.g. Pfeiffer et al., 2012; Pollit, 2012).

There is further information to be extracted from the data collected in Dar es Salaam. For example, in addition to our analysis of the effect of gender and occupation, it may be of interest to local agencies to understand whether other covariates (or combinations of covariates) are associated with different perceptions of deprivation. We can also investigate the tied comparisons using a multinomial model, (see, e.g. Davidson, 1970; Rao & Kupper, 1967), to investigate the effect of comparing subwards which had similar deprivation levels.

We have developed new models for efficiently estimating the level of deprivation in urban areas based on comparative judgement data. Existing comparative judgement models require a large amount of data to produce high-quality results and collecting such quantities of data is often difficult or infeasible when working in developing countries, where data collection can be prohibitively expensive and time-consuming. Using the Bayesian Spatial Bradley–Terry model, we could have collected considerably fewer comparisons without affecting the quality of our results. When using the data collected only on the first 2 days on the fieldwork, the error in the BSBT model is small, and substantially smaller than when using the standard model. We achieved this by including a spatial element in the model, where the level of deprivation in one subward is correlated with the level in nearby subwards. We modelled the spatial structure using a multivariate normal prior distribution with a covariance matrix based on the network structure of the city, which avoids making rigid parametric assumptions. We also showed how our method can be used to analyse how different genders perceive the level of deprivation in different areas, and how different their perceptions are. This can help researchers identify areas where one gender may be facing specific problems.

ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council [grant references EP/T003928/1 and EP/R513283/1]. We also thank the Humanitarian OpenStreetMap Team (HOT) for their support in data collection. We are grateful to the two reviewers and associate editor for helpful and constructive comments that have improved this article.

ORCID

Rowland G. Seymour https://orcid.org/0000-0002-8739-3921
David Sirl https://orcid.org/0000-0002-2639-1666
Simon P. Preston https://orcid.org/0000-0002-1910-4227
Ian L. Dryden https://orcid.org/0000-0003-4900-3571

REFERENCES

Bradley, R.A. & Terry, M.E. (1952) Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39, 324–345.

Cattelan, M. (2012) Models for paired comparison data: a review with emphasis on dependent data. *Statistical Science*, 27(3). Available from: https://doi.org/10.1214/12-sts396

Cormen, T.H., Stein, C., Rivest, R.L. & Leiserson, C. E. (2001) *Introduction to algorithms*. 2nd edn. Cambridge, MA: MIT Press.

Davidson, R.R. (1970) On extending the Bradley-Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65, 317–328. Available from: http://www.jstor.org/stable/2283595

Devarajan, S. (2013) Africa's statistical tragedy. *Review of Income and Wealth*, 59, S9–S15. Available from: https://doi.org/10.1111/roiw.12013

Engelmann, G., Smith, G. & Goulding, J. (2018) The unbanked and poverty: Predicting area-level socio-economic vulnerability from m-money transactions. In 2018 IEEE International Conference on Big Data (Big Data), 1357–1366. IEEE.

Estrada, E. & Higham, D.J. (2010) Network properties revealed through matrix functions. *SIAM Review*, 52(4), 696–714. Available from: https://doi.org/10.1137/090761070

van Etten, J., de Sousa, K., Aguilar, A., Barrios, M., Coto, A., Dell'Acqua, M. et al. (2019) Crop variety management for climate adaptation supported by citizen science. *Proceedings of the National Academy of Sciences*, 116, 4194–4199. Available from: https://doi.org/10.1073/pnas.1813720116

Firth, D. (2020) qvcalc: quasi Variances for Factor Effects in Statistical Models. R package version 1.0.2.

Firth, D. & De Menezes, R.X. (2004) Quasi-variances. *Biometrika*, 91(1), 65–80. Available from: https://doi.org/10.1093/biomet/91.1.65

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian Analysis*, 1, 515–534. Available from: https://doi.org/10.1214/06-BA117A

Grinfeld, J., Nangalia, J., Baxter, E.J., Wedge, D.C., Angelopoulos, N., Cantrill, R. et al. (2018) Classification and personalized prognosis in myeloproliferative neoplasms. *New England Journal of Medicine*, 379, 1416–1430.

Higham, N.J. (1988) Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103, 103–118. Available from: https://doi.org/10.1016/0024-3795(88)90223-6

Kalton, G. & Schuman, H. (1982) The effect of the question on survey responses: a review. *Journal of the Royal Statistical Society Series A (General)*, 145, 42–73. Available from: http://www.jstor.org/stable/2981421

Kleijn, B. & van der Vaart, A. (2012) The Bernstein-von-Mises theorem under mis-specification. *Electronic Journal of Statistics*, 6, 354–381. Available from: https://doi.org/10.1214/12-ejs675

Limbumba, T.M. & Ngware, N. (2016) Informal Housing Options and Locations for Poor Urban Dwellers in Dar es Salaam City. *The Journal of Social Sciences Research*, 2, 93–99.

Lynn, P. & Clarke, P. (2002) Separating refusal bias and non-contact bias: evidence from UK national surveys. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51, 319–333.

McCrickard, L.S., Massay, A.E., Narra, R., Mghamba, J., Mohamed, A.A., Kishimba, R.S. et al. (2017) Cholera mortality during urban epidemic, Dar es Salaam, Tanzania, August 16, 2015–January 16, 2016. *Emerging Infectious Diseases*, 23(13), 154–157.

McLennan, D., Noble, S., Noble, M., Plunkett, E., Wright, G. & Gutacker, N. (2019) The English indices of deprivation 2019. Technical report, Ministry of Housing, Communities and Local Government, London, UK. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833951/IoD2019_Technical_Report.pdf

Napacho, Z.A. & Manyele, S.V. (2010) Quality assessment of drinking water in Temeke District (part II): Characterization of chemical parameters. *The Journal of Social Sciences Research*, 4, 775–789.

Neal, R.M. (1998) Suppressing random walks in Markov Chain Monte Carlo using ordered overrelaxation. In *Learning in graphical models*. Dordrecht: Springer Netherlands, pp. 205–228.

Pfeiffer, T., Gao, X., Chen, Y., Mao, A. & Rand, D. (2012) Adaptive polling for information aggregation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 26. Available from: https://ojs.aaai.org/index.php/AAAI/article/view/8099

Pollit, A. (2012) The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 19, 281–300.

Randall, S. & Coast, E. (2015) Poverty in african households: the limits of survey and census representations. *The Journal of Development Studies*, 51, 162–177.

Rao, P.V. & Kupper, L.L. (1967) Ties in paired-comparison experiments: a generalization of the Bradley-Terry model. *Journal of the American Statistical Association*, 62, 194–204.

Rasmussen, C.E. & Williams, C.K.I. (2006) *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts: MIT Press.

Seymour, R.G. & Briant, J. (2021) BSBT. Available from: https://doi.org/10.5281/zenodo.5466044. R package version 1.1.0.

de Soete, G. & Winsberg, S. (1993) A Thurstonian pairwise choice model with univariate and multivariate spline transformations. *Psychometrika*, 58, 233–256.

Springall, A. (1973) Response surface fitting using a generalization of the Bradley-Terry paired comparison model. *Journal of the Royal Statistical Society Series C*, 22, 59–68.

Stern, S.E. (2011) Moderated paired comparisons: a generalized Bradley-Terry model for continuous data using a discontinuous penalized likelihood function. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60, 397–415.

Strobl, C., Wickelmaier, F. & Zeileis, A. (2011) Accounting for individual differences in Bradley-Terry models by means of recursive partitioning. *Journal of Educational and Behavioral Statistics*, 36, 135–153.

Turner, H. & Firth, D. (2012) Bradley-Terry models in R: the BradleyTerry2 Package. *Journal of Statistical Software*, 48(9), 1–21. Available from: https://doi.org/10.18637/jss.v048.i09

United Nations Department of Economic and Social Affairs. (2019) World urbanization prospects: the 2018 revision. New York: United Nations.

USAID. (2019) Artificial intelligence in global health: Defining a collective path forward. Report, USAID Center for Innovation and Impact.

Varin, C., Cattelan, M. & Firth, D. (2016) Statistical modelling of citation exchange between statistics journals. *Journal of the Royal Statistical Society: Series A*, 179, 1.

Williams, D., Hornung, H., Nadimpalli, A. & Peery, A. (2021) Deep learning and its application for healthcare delivery in low and middle income countries. *Frontiers in Artificial Intelligence*, 4, 30–35. Available from: https://doi.org/10.3389/frai.2021.553987

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Seymour, R.G., Sirl, D., Preston, S.P., Dryden, I.L., Ellis, M.J.A., Perrat, B. et al. (2022) The Bayesian Spatial Bradley–Terry model: Urban deprivation modelling in Tanzania. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1–21. Available from: https://doi.org/10.1111/rssc.12532