Mathematical Biosciences and Engineering

MBE, 21(3): 3816–3837.
DOI: 10.3934/mbe.2024170
Received: 12 August 2023
Revised: 24 October 2023
Accepted: 06 November 2023
Published: 19 February 2024

http://www.aimspress.com/journal/MBE

Research article

Methodological approaches to exploring the spatial variation in social impacts of protected areas: An intercomparison of Bayesian regression modeling approaches and potential implications

Chrysovalantis Malesios*, Nikoleta Jones, Alfie Begley and James McGinlay

Institute for Global Sustainable Development, The Ramphal Building, University of Warwick, Coventry CV4 7AL, UK

* **Correspondence:** Email: Chrysovalantis.Malesios@warwick.ac.uk.

Abstract: Protected Areas (PAs) are widely used to conserve biodiversity by protecting and restoring ecosystems while also contributing to socio-economic priorities. An increasing number of studies aim to examine the social impacts of PAs on aspects of people's well-being, such as, quality of life, livelihoods, and connectedness to nature. Despite the increase in literature on this topic, there are still few studies that explore possible robust methodological approaches to capturing and assessing the spatial distribution of impacts in a PA. This study aims to contribute to this research gap by comparing Bayesian spatial regression models that explore links between perceived social impacts and the relative location of local residents and communities in a PA. We use primary data collected from 227 individuals, via structured questionnaires, living in or near the Peak District National Park, United Kingdom. By comparing different models we were able to show that the location of respondents influences their perception of social impacts and that neighboring communities within the national park can have similar perceptions regarding social impacts. Simulation based on existing data using the Bootstrap sub-sampling was also conducted to validate the association between social impacts and mutual proximity of residents. Our findings suggest that this type of data is better treated, in terms of accounting for potential spatial effects, using models that allow for proximity effects to be stronger between people living nearby, e.g. between neighbors in the same community and have minimum effects otherwise. Understanding the spatial clustering of perceived social impacts in and around PA, is key to understanding their causes and to managing and mitigating them. Our findings highlight therefore the need to develop new methodological approaches to assessing and predicting accurately the spatial distribution of social impacts when designating PAs. The findings in this paper will assist practitioners in this regard by proposing approaches to the consideration of the distribution of social impacts when designing the boundaries of PAs alongside typical ecological and socio-economic criteria.

Keywords: spatial; Bayesian modeling; social impacts; benefits; costs; designation; perception of social impacts; clustering; mapping

1. Introduction

A large number of countries around the world have committed to protect 30% of land and 30% of water by 2030 [1]. This ambitious target has been set in order to both halt biodiversity loss and improve climate change adaptation via nature-based solution. Protected Areas (PAs) can be defined as specific geographical area, recognised, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values [2]. This definition reveals two key characteristics of PAs. First, that PAs have specific boundaries within which certain regulations are enforced and second, that PAs have multidimensional targets focusing both on environmental and socio-economic aspects. Regarding the latter point scholars have highlighted the need to integrate local values into decision-making processes and put actions in place in order to maintain these values within PAs [3].

Due to the challenge of meeting these multidimensional targets for PAs, it is often the case that socio-economic aspects are not taken into consideration during PA designations or at least not to the same extent as ecological criteria. However, a growing body of literature has shown that PAs can have significant socio-economic benefits, such as the improvement of wellbeing, involvement in recreational activities and increase in personal income through tourism [4–8]. PAs can also have diverse negative impacts for local people [9]. This is because the designation of PAs often introduces key restrictions on human activity within certain geographical boundaries such as restrictions on fishing, logging and recreation [10,11]. Public perception of the social impacts of PAs, both positive and negative are important as they have been shown to influence public support for the PA [7,8,12]. It is therefore important to be able to identify and measure the type, extent and spatial distribution of such impacts in order to maximise PA management effectiveness both for nature and for people.

Despite the increase in research studies on the social impacts of PAs, there is very limited evidence regarding the distribution of social impacts across space within PAs, and relatedly, limited guidance on best practice for spatial assessment of impacts. Naidoo et al. [13] was one of the first studies that explored spatial aspects of social impacts and found that living near a PA has a positive effect on wellbeing levels in countries in the Global South. Jones et al. [9] also found that an individual’s location in the PA is an explanatory parameter for people’s wellbeing level, whilst McGinlay et al. [14] found that perceived impacts in the Eifel National Park, Germany influenced public support for the Park, and that impacts varied spatially in relation to settlement patterns and road infrastructure. Despite these contributions, our knowledge of how and why social impacts are distributed across space in PAs is very limited, with very few studies exploring whether social impacts differ between locations within a PA and whether the location of communities influence their perceptions.

The spatial aspect of social impacts is an important area of research, as defining a protected area is a spatial approach to biodiversity and landscape protection and enhancement. The fundamental aim is to set a specific boundary within which a new regulatory and governance regime will alter biophysical and social processes in a way that will benefit ecosystems, but which will also impact on people, potentially both negatively and positively. However, spatial aspects are often ignored when examining the factors that contribute to social impacts. Additionally, ignoring spatial autocorrelation, which refers to the correlation between nearby observations in space, can invalidate statistical modeling results if not accounted for [15].

Overall, given the scarcity of studies on the spatial distribution of social impacts in PAs, knowledge gaps are not only theoretical and conceptual, but also methodological, and so there is a need to consider what approaches to measuring, mapping and modeling impacts are likely to prove most helpful in supporting theory-building.

In this paper we compare different modeling approaches to explore links between perceived social impacts and the relative location of local communities in a PA. We additionally make recommendations on which approaches to take that can be used by researchers and practitioners when exploring the social impacts of new and existing PAs from a spatial perspective.

Our framework builds upon previous studies [13], while our approach is more comprehensive as we incorporate spatial information through kernel functions that consider proximity effects: i.e., the proximity of respondents to each other, as well as less detailed spatial information that only considers the neighboring proximity between municipalities of local residents.

2. Materials and methods

2.1. Data

Primary data were collected using structured questionnaires distributed in the Peak District National Park (United Kingdom) during the Summer of 2020 as part of the project FIDELIO (www.warwick.ac.uk/fidelio). The Peak District national park is located in central England and was established in 1951 (See Figure 1). The area is terrestrial (1,438 km^2) and socio-economic activities in the region are mainly tourism and recreation as well as agricultural activities. The specific research area was selected based on the following criteria that were suitable for our analysis: a) an area that has local communities living inside its boundaries; b) area that is large enough to allow the exploration of spatial distribution; c) area that is designated as National Park.

The National Park is designated as an IUCN category V acknowledging the interaction of local communities with the landscape and the contribution of this interaction to a number of aspects including the natural environment, cultural values and the local economy.

The research team sent 3100 postcards to a randomly selected sample of households in and immediately around the Peak District area inviting them to participate in the survey (for further details about this survey please see full report: [16]. This was estimated to be approximately 10% of the total population. The survey was also advertised online via social media and informal networks with the help of the Peak District National Park Authority. In total, 438 responses were received. Through the questionnaire, respondents were asked to provide information about their village of residence in order to capture their location which was essential for the spatial analysis in this paper. 227 respondents (51% of the sample) provided this information. The sampling frame included those who live inside the PA or within a 10km buffer around the PA boundary. The specific distance from the borders of the PAs was inspired by previous research [13,17]. A research protocol was designed and followed in order to ensure the validity and reliability of our analysis and derived results. Specifically, prior to the final collection of data a variety of proactive actions was performed, in terms of validity and respondent and non-respondent bias assessment. These actions followed include the pilot testing of the questionnaire to a small sample of respondents in all areas of the PA, to identify potential issues with question wording, terminology, response options etc. Potential selection bias was also considered prior to the data collection, by comparing the demographic characteristics of the target population of the PA in order to be aligned with the characteristics of survey respondents. Further, our sampling frame was distributed throughout all regions inside the PA and within the 10km outside buffer zone, in order to ensure the adequate regional representation of respondents.

Regarding perceptions of social impacts these were captured for 5 different aspects of personal well-being: personal income, quality of life, involvement in recreational activities, social relations and connectedness to nature (Table 1). Respondents were asked *"How has the designation of the PA impacted you regarding the following issues in the past years*?" and all questions were measured via a 5 point Likert Scale (1-Very negative impact, 2- Negative impact, 3-No Impact, 4-Positive Impact, 5-Very positive impact). These questions were part of a social capital and impact assessment questionnaire which was distributed in the area of the Peak District [18].

Table 1. Description of dependent variables (social impacts).

Dependent variables of social impacts	Question		Scale of measurement
Social impacts	How has the designation of the PA impacted you regarding the following issues in the past years?	Personal Income	5-point Likert scale: 1-very negative impact, 5- very positive impact, 3-neutral/no impact
		Your quality of life	
		Your involvement in recreational activities	
		Social relations with locals	
		Your connectedness to Nature	

Furthermore, spatial information was utilized as explanatory variables for the variation in perceived social impacts of PAs, aiming to capture the role of location, and of mutual spatial proximity between local residents on their views of perceived impacts.

Data collected included the location in the form of geographical coordinates, in particular, the spatial data for the analysis involves coordinates that relate to the respondent's location (specifically the village of residence of respondents). To collect the coordinates, the names for each location in our survey were entered into an Excel worksheet. Once opened in the browser, an extension called 'Geocode' was used. To create the polygons, a layer from ArcGIS Online with UK wards was added to the map. Then, the wards were arbitrarily divided between the number of villages in each municipality. To ascertain the average distance of each respondent from the remaining respondents we calculated the Euclidean distance between each location and all other locations in the Peak District dataset using the haversine formula. This calculates the distance between two sets of coordinates. As a summary measure of distance, the average distance was utilized.

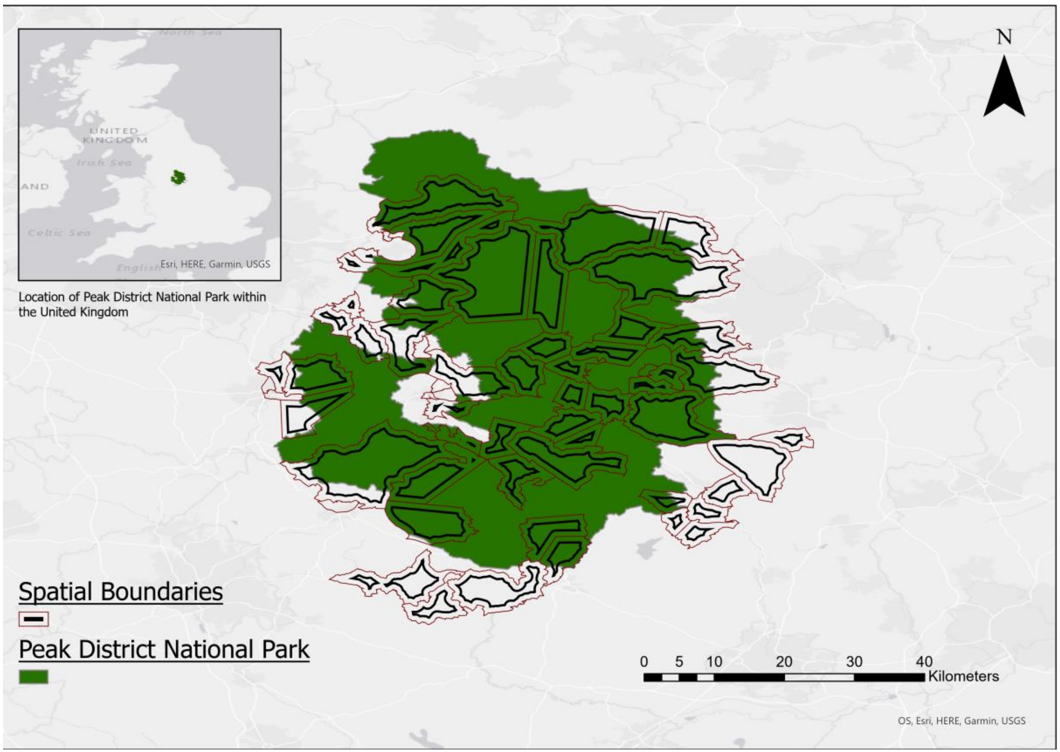


Figure 1. Location map of the case study area of Peak District, UK.

2.2. Statistical analysis

2.2.1. Bootstrap simulation

In order to understand the robustness of our analysis that uses a real dataset and validate the outcomes, we have simulated our data from Peak District PA using the Bootstrap simulation method [19–20]. Bootstrapping simulation approach involves resampling a single data set to create a multitude of simulated samples. Those samples are used to calculate standard errors, confidence intervals and for hypothesis testing. Bootstrap is also an appropriate way to control and check the stability of the results obtained by small datasets.

Simulating our research data prior to conducting the actual spatial analysis modeling can be an important step to help better understand the process, validate our methods, and gain more insights into potential outcomes. Specifically, we have simulated the main variables we are interested in for this study, such as the spatial proximity of local residents and the social impacts of respondents. We apply the non-parametric bootstrap technique for testing using the simulated samples if there are any spatial patterns present relating to the various social impacts. The procedure for bootstrapping essentially relies on resampling from an initial sample. Specifically, one creates B bootstrap samples by sampling with replacement from the original data.

The method of non-parametric resampling using bootstrap methodology for calculating estimates and confidence intervals for the parameter of interest is based on the following general scheme:

i. Sample n observations randomly with replacement from the initial sample of data, say vector $\mathbf{y_{obs}}$.

ii. Calculate the bootstrap version of the statistic of interest, say $\hat{\theta}$.

iii. Repeat steps i and ii a large number of times, say *B*, to obtain an estimate of the bootstrap distribution.

The *B* samples are called bootstrap samples. In practice, the number of *B* samples that is chosen for our study is 10,000.

2.2.2. Statistical modeling

Due to the nature of the response variables for social impacts of PAs, that is, a 1-5 Likert scale, we fit suitable spatial regression models, such as ordinal logistic spatio-temporal regression models to the data. Within this category, the selection of the appropriate data distribution was investigated by comparing the former to the Poisson [21] and the Negative Binomial (NB) [22] distributional families. Hence, we test alternative spatial regression models, in terms of distributional specification of the responses, for example Poisson distribution and alternatives, such as the negative binomial and the generalized Poisson [23], both being suitable for data that are not equi-dispersed (i.e., the value of mean is different from the value of variance). Especially the negative binomial model is frequently considered as an alternative to the Poisson distribution in cases of over dispersed data.

Further, in the present paper a comparison is conducted, between candidate spatial approaches, in order to examine potential spatial effects of proximity between the local residents of the protected areas. Specifically, we compare stochastic spatial models (see, e.g., [24] that utilize precise spatial information in terms of exact co-ordinates of local residents to calculate the distances between respondents, to spatial regression models that use less precise information of proximity, based upon the conditional autoregressive spatial models (CAR) of Besag and Kooperberg [25].

This comparison is useful, since it can provide important insights into the spatial patterns of the social impacts of PAs and how spatial proximity between local residents can affect their views regarding these impacts.

We take a model comparison approach to assess the relative significance of the proposed spatial models, hence we chose to fit a series of four candidate distributional specifications for the response variables of the five impacts (Table 1), along with a series of candidate spatial models. For comparison reasons, we also fit a simple baseline regression model that includes only an intercept in order to assess the overall spatial effects by comparing this model to the spatial ones.

In the next sub-sections, we describe in detail the various specifications in terms of distributional assumptions and spatial approaches of the fitted regression models.

2.2.2.1. Kernel-based spatial regression model(s)

As a measure of spatial proximity between the local residents, kernel-based spatial regression models utilized for our analysis use the average distance (in km) of each resident from all other local residents in the PA. The former measure of spatial proximity between local residents aims to capture patterns of spatial similarity (or dissimilarity) between local residents relative to one another regarding their views on perceived impacts as a result of their proximity to the PA.

The analytical expression of the Bayesian kernel-based statistical models used for the current study is mathematically formulated as follows:

Ordinal Logit model:

$$\begin{cases} y_i = 1 \text{ if } Y_i^* \le \mu_1 \\ y_i = 2 \text{ if } \mu_1 < Y_i^* \le \mu_2 \\ y_i = 3 \text{ if } \mu_2 < Y_i^* \le \mu_3 \\ y_i = 4 \text{ if } \mu_3 < Y_i^* \le \mu_4 \\ y_i = 5 \text{ if } \mu_4 < Y_i^* \le \mu_5 \end{cases} \tag{1}$$

$$Y_i^* = \beta_0 + f(d_i) + \varepsilon$$

Poisson model:

$$y_i \sim Poisson(\lambda_i) \tag{2}$$

$$\log(\lambda_i) = \beta_0 + f(d_i) + \varepsilon$$

Generalized Poisson (GP) model:

$$\begin{aligned} & y_i \sim Poisson(\lambda_i^*) \\ & \lambda_i^* = (1-\omega)\cdot\lambda_i + \omega\cdot y_i \\ & \log(\lambda_i) = \beta_0 + f(d_i) + \varepsilon \end{aligned} \tag{3}$$

Negative binomial (NB) model:

$$\begin{aligned} & y_i \sim \mathrm{NB}\left(\mu_i = \frac{r(1-q_i)}{q_i}, q_i\right) \\ & \frac{r(1-q_i)}{q_i} = \beta_0 + f(d_i) + \varepsilon \end{aligned} \tag{4}$$

where y_i denotes the *i*th response value of 1-5 Likert scale $(i = 1,2,...,227)$, taking the values of 1,2,3,4,5, and in the ordinal logistic model, Y^* denotes a continuous, unmeasured latent variable which is assumed to give rise to the observed categories 1,2,3,4,5 [26]. Also, β_0 is the intercept. In addition, λ_i and λ_i^* are the parameters of the Poisson and generalized Poisson distribution, respectively and r, q_i are the parameters of the negative binomial distribution. Finally, ε is the error term.

The generalized Poisson regression model is a generalization of the standard Poisson regression model. When the dispersion parameter ω = 0, the probability function in Eq (3) reduces to the Poisson model (1). When ω>0, the generalized Poisson model represents count data with overdispersion and when ω<0, the generalized Poisson model represents count data with underdispersion.

The information collected on the perceived impacts of protected areas (PAs) on local residents is spatial in nature and varies across different locations. To address this, we have incorporated a set of distance measures in Eqs (1)–(4) as additional factors that account for the spatial differences in the prediction of perceived impacts of PAs on local residents. These distance measures will be used as covariates in the four distributional specifications, denoted as d_i. The average distances, d_i, were included in the regression models via a typical decaying function of distance.

A commonly used distance kernel function is the inverse distance power function, with:

$$f(d_i) = \left(\frac{1}{d_i}\right)^a \tag{5}$$

for positive integer α, with α often taken to be 1 [27,28]. Another typical function used is the negative exponential, given by:

$$f(d_i) = \exp(-\alpha \cdot d_i) \tag{6}$$

with $\alpha > 0$, the parameter controlling for the rate of decay with distance [27,29].

In the current paper we utilize both kernel functions and compare their performance. In this way, we seek to capture the spatial pattern of density decay with average distance, d_i, of each local resident respondent from the remaining respondents of the Peak District protected area. The inverse distance power function has a fatter tail over long distances [30]. This is also true when compared to the negative exponential function.

2.2.2.2. CAR spatial panel model(s)

Conditional autoregressive (CAR) models are regularly used to analyze data in a large range of disciplines, such as in demography, economy, epidemiology and geography [31]. A conditional autoregressive modeling approach following the Bayesian paradigm is employed in order to examine the potential spatial effects on the response variables of social impacts.

Let y_{is} denote the *i*th response for the s^{th} municipality of Peak District region, UK. Hence, the CAR Bayesian regression model is subsequently written in its general form as:

$$\begin{gathered} y_{is} \sim DIST(\mu_{is}) \\ \mu_{is} = \beta_0 + \eta_s \end{gathered} \tag{7}$$

for the CAR model with mean μ_{is}, which is suitably adjusted for the various distributional specifications of ordinal logistic, Poisson and generalizations of the latter, as presented already in Eqs (1)–(4). In the model formulation of Eq (7), β_0 is the intercept.

CAR is a spatial regression method that incorporates spatial dependency into the data analysis through the inclusion of a spatial component, say, η_s that is utilized for the modelling of the space random effects, and is represented as:

$$\eta_s | \eta_{\setminus s} \sim N\left(\overline{\eta}_s, \frac{\sigma_\eta^2}{N_s}\right) = N\left(\frac{1}{\sum_j w_{sj}} \cdot \sum_{j=1}^{N_s} w_{sj} \cdot n_j, \frac{\sigma_n^2}{\sum_j w_{sj}}\right) \tag{8}$$

In the previous equation, $\setminus s$ denotes all municipalities excluding s, $\overline{\eta}_s = \frac{1}{\sum_j w_{sj}} \cdot \sum_{j=1}^{N_s} w_{sj} \cdot n_j$ denotes the means of the neighbouring random effects for municipality s, and σ_n^2 is the corresponding variance.

A weight matrix (also called a spatial proximity matrix) is then utilized with its elements w_{ij} being some type of measure of the spatial relationship between the *i*th and *j*th municipalities. A common definition for distance-based weights is the weights given by:

$$w_{ij} = \begin{Bmatrix} 1, \text{if areas i and j are neighbours} \\ 0, \text{otherwise} \end{Bmatrix} \tag{9}$$

The assumption of "neighbours" typically refers to areas that share a common boundary [32,33]. We must also note that the property of symmetric matrices of weights $\mathbf{w}$, $\mathbf{W}'$ must hold for deriving valid results.

2.2.3. Prior Specification and inference for the Bayesian regression models

Model implementation and inference in this paper is based upon a fully Bayesian approach. For inference by following the Bayesian paradigm, suitably vague priors for the parameters of the non-spatial, CAR and kernel-based spatial regression models are specified. In particular, the prior distributions for the intercept (β_0) and the spatial parameters are specified through a Gaussian distribution, with zero-mean vector and a non-informative variance, i.e. $N(0,1000)$. Accordingly, the variance parameters σ^2, were assigned weakly informative half-normal prior distributions. Finally, in function specifications 5 & 6, we do not assign a specific value on parameter α, instead allowing α to vary by assigning a weakly informative prior distribution.

In order to adopt the Bayesian paradigm for parameter estimation, the WinBUGS software has been utilized [34]. The model's parameters were estimated via Markov chain Monte Carlo (McMC) simulation. The convergence of the McMC chains was assessed through visual inspection of the posterior distributions. For selecting the optimal model among the candidate models fitted, suitable goodness-of-fit criteria are utilized, both from the Bayesian and frequentist perspective. The comparison of the models is done through the *DIC* criterion (Deviance Information Criterion, which is a generalization of the criterion *AIC* (Akaike Information Criterion) and which was proposed by

Spiegenhalter et al. [35] and measures the variability of likelihood. Thus, the deviance information criterion (*DIC*) values reported, with smaller values of these measures indicating towards selection of the corresponding model. The *DIC* is based on the posterior mean deviance and the effective number of parameters, p_D, expressed as: $DIC = p_D + \overline{D}$ ($\overline{D}$ denotes the posterior mean deviance). The smaller the *DIC* value we have, the better fit the model makes to the data. The WinBUGS code of all spatial and non-spatial fitted models is available upon request by the corresponding author.

3. Results

3.1. Descriptive statistics

The following table (Table 2) shows descriptive statistics for the five variables of perceived social impacts reported in the survey at Peak District National Park. On average, local residents of the PA evaluate highly positively the effects of the national park on their lives, with the least positive effects observed on their personal income (average score 3.665 on the 5 point Likert scale). This is also the variable with the highest diversity in the responses of local residents (standard deviation: 1.052) and so the greatest variation from household to household across the Park.

Table 2. Descriptive statistics for the variables of social impacts.

Impact	Min	Mean	Max	Standard deviation
Personal income	1	3.665	5	1.052
Quality of life	2	4.731	5	0.619
Recreational activities	1	4.661	5	0.725
Social relations	1	4.788	5	0.571
Connectedness to nature	2	4.493	5	0.766

The spatial distribution of the results depended on who responded to the survey out of the residents invited to participate, and who also gave their geographical location. Of greater interest is whether the magnitude of the reported impact scores of a household were related to, or independent of the geographical location in or around the Park, and in particular whether the score given by one household at one location was related to or independent of the scores given by other households nearby. In other words, firstly, did scores vary across the landscape and was location in the Park a significant factor affecting scores given, and secondly, might one household's perception influence that of others nearby. To address these questions, a range of models were used to predict how responses might vary across the Park, and to assess which modelling approach was most informative and useful.

Overall, the modeling results indicate that impact scores from a household were related to where they were in the Park and were not independent of the scores of other nearby households. For the 5 impacts studied, it appears that the impact of the PA on the social relations of local residents is the least affected by the proximity of other residents, and also varied least across the landscape (had the lowest standard deviation) (Table 2).

3.2. Results of Bootstrap simulation

The Bootstrap subsampling simulation has been performed prior to the statistical spatial regression modeling by the utilization of a total sample of 10,000 iterations to collect the Bootstrap sub-samples from the five samples of social impacts and the variable of average distances of each resident from all other local residents in the PA.

In Figure A1 in the Supplementary, the histograms and normal probability plots for the overall bootstrap samples, based upon the 10,000 bootstrap sub-samples, are presented. As we observe, the bootstrap samples follow the normal distribution.

Upon the generation of the bootstrap samples, subsequently we examine potential spatial patterns in the data by calculating the Pearson's correlation coefficients between the five social impacts bootstrap samples and the average distance of respondents (Table A1).

As is seen by the correlations, the simulated values of the average distance between local residents are significantly correlated with all five simulated social impacts variables. This result is an indication that spatial dependence exists regarding the perceived social impacts, and subsequently that these variables are not randomly distributed but instead exhibit patterns or relationships based on their physical locations. Hence, these findings suggest that the spatial dependencies can be examined through techniques like spatial regression, which help uncover how the proximity or spatial arrangement influences social outcomes or impacts in the particular region.

3.3. Results of spatial regression modelling and comparisons

In this sub-section, the results in terms of the deviance information criterion for all fitted regression-type Bayesian models are reported. Hence, Tables 3–6 show the *DIC* values for the simple baseline regression model, the CAR panel spatial model and the two alternative spatial regression kernel function models, for each one of the four alternative distributional specifications for the response variables of social impacts (ordinal logistic, Poisson, generalized Poisson and negative binomial).

The results clearly show that the fit of all models based on the ordinal logistic distribution (Table 3) outperformed all other distributional assumptions for the specific dataset, since the deviance information criterion values for the spatial ordinal logistic regression model are the lowest in comparison to all other modeling specifications.

The next best fit is observed for the generalized Poisson specification, as revealed by the *DIC* estimates in Table 5.

Table 3. Posterior mean deviance for the baseline and spatial logistic regression models.

Impact	Baseline model	CAR spatial model	Kernel-distance spatial model (inverse distance power)	Kernel-distance spatial model (negative exponential)
Personal income	598.6	593.6	596	595.93
Quality of life	299.3	292.6	296.9	296.77
Recreational activities	343.9	336.7	341.2	341.22
Social relations	250.1	243.05	245.7	243.97
Connectedness to nature	429.4	426.4	427.8	427.41

Table 4. Posterior mean deviance for the baseline and spatial Poisson regression models.

Impact	Baseline model	CAR spatial model	Kernel-distance spatial model (inverse distance power)	Kernel-distance spatial model (negative exponential)
Personal income	796.94	789.58	788.42	788.33
Quality of life	808.63	803.13	799.99	800.03
Recreational activities	812.41	807.49	804.84	804.82
Social relations	809.68	803.48	800.08	800.09
Connectedness to nature	808.36	801.89	798.75	798.75

On the other hand, the worst model fit is seen for the negative binomial regression models (Table 6), with similar results observed for the Poisson models (Table 4).

After the selection of the best distributional specification (ordinal logistic, Table 3), we subsequently compare the various alternative spatial model formulations in comparison to the baseline regression model. When considering the alternative spatial specifications and their comparison to the non-spatial baseline regression model, it is observed that the best fit is achieved for the conditional autoregressive panel spatial model under an ordinal logistic distribution for the response variables (Table 3). CAR model outperforms the baseline regression model and the two alternative distance kernel function specifications in terms of model fit according to the *DIC* estimations (smallest *DIC* values in Table 3). On the other hand, the kernel-based spatial models have slightly better fit compared to CAR in the case of the Poisson and negative binomial specifications.

Table 5. Posterior mean deviance for the baseline and spatial generalized Poisson regression models.

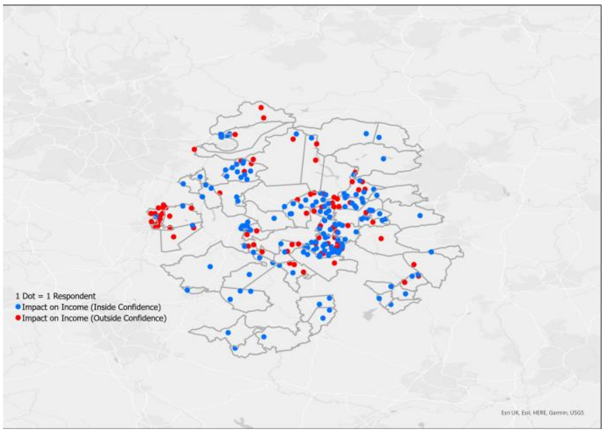
Impact	Baseline model	CAR spatial model	Kernel-distance spatial model (inverse distance power)	Kernel-distance spatial model (negative exponential)
Personal income	718.55	713.36	714.95	714.97
Quality of life	781.56	776.55	777.94	777.78
Recreational activities	778.97	771.91	773.33	773.35
Social relations	783.31	779.44	780.82	780.73
Connectedness to nature	769.06	764.01	765.51	765.33

Table 6. Posterior mean deviance for the baseline and spatial Negative Binomial regression models.

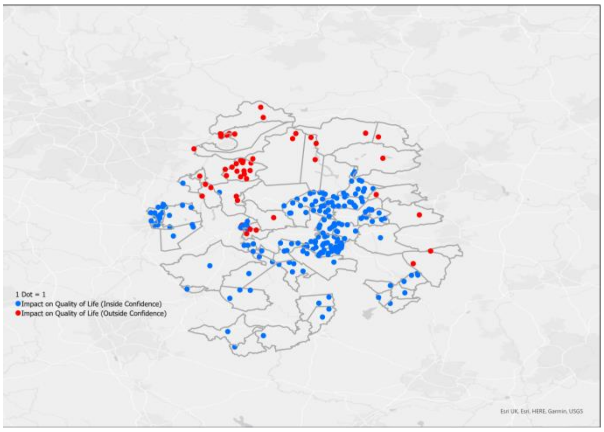
Impact	Baseline model	CAR spatial model	Kernel-distance spatial model (inverse distance power)	Kernel-distance spatial model (negative exponential)
Personal income	796.25	791.18	790.46	790.66
Quality of life	808.27	804.87	802.67	802.50
Recreational activities	813.02	809.40	807.38	807.44
Social relations	809.85	805.49	802.83	802.76
Connectedness to nature	808.93	803.55	801.38	801.35

Overall, social impacts are best predicted by a spatial CAR regression model that takes into consideration the similarity of responses of residents in neighboring areas, and by considering an ordinal logistic distribution to link the responses to the spatial information.

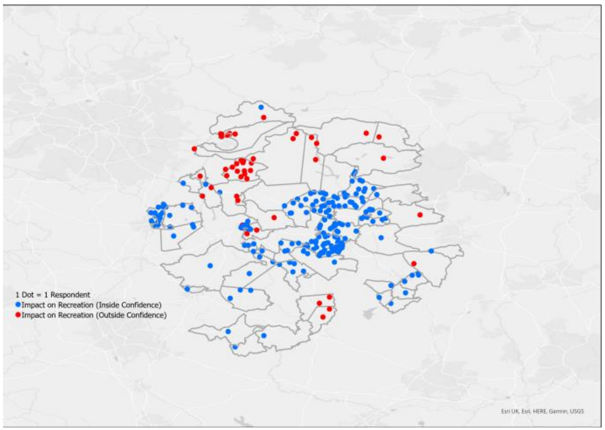
As a final assessment of the best model performance, Figure 2 compares observed with predicted values of the best fitted spatial regression model (CAR spatial ordinal logistic model). Specifically, blue dots in the five impact graphs indicate a good predictive performance (observed social impact value falls within the 95% credible intervals of prediction), whereas red dots indicate poor performance (observed value falls outside the 95% credible intervals). Results indicate a good fit in most impacts, but with poorer fit observed for the impact on “Social relations” (Figure 2d).



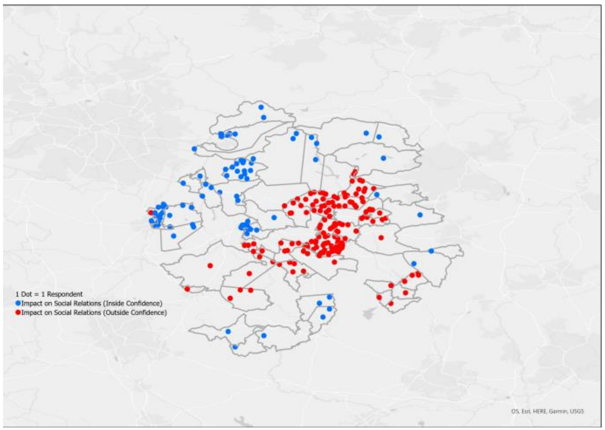
(a)



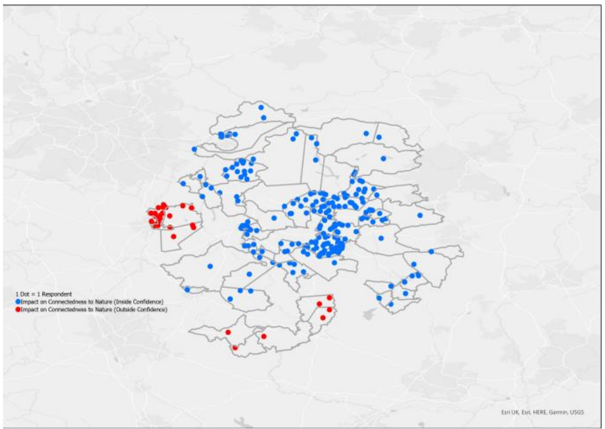
(b)



(c)



(d)



(e)

Figure 2. Predictive spatial maps for the Peak District PA for a. impact on "Income", b. impact on "Quality of life", c. impact on "Recreational activities", d. impact on "Social relations" and e. impact on "Connectedness to Nature".

4. Discussion

Describing, mapping and explaining the spatial clustering of social impacts in and around PA, and of how they are perceived, is key to understanding their causes and to managing and mitigating them. In this paper we have examined the spatial variation in perceived social impacts and associated clustering effects in a large European Protected Area using survey data from local households. In particular, we have implemented generic Bayesian methodology, which offers flexibility in the ability to fit various distributional specifications, both typical and more complex as well as a variety of spatial regression models that are able to capture effectively the potential effects of spatial proximity between local residents of a protected area on their impacts assessment.

Our findings suggest that spatial proximity of local residents to each other in the Peak District National Park plays an important role in how people perceive social impacts, since all spatial Bayesian regression models showed a relatively improved fit compared to the baseline regression model that did not include any type of spatial information in the form of covariate.

Among the spatial regression models tested, the conditional autoregressive model that assigns a weight for adjacent municipalities and ignores spatial effects of non-adjacent areas gave the best fit. There are a few possible reasons why the conditional autoregressive (CAR) spatial model with binary weights outperformed the more detailed spatial regression model that uses exact coordinates to measure distance when analyzing data on local residents of a protected area.

Firstly, the CAR model is designed specifically to account for spatial dependence in the data [36]. This type of model assumes that the correlation between observations decreases as distance increases, but does not explicitly model the distance between observations. This can be appropriate when the spatial resolution of the data is low or when the research question is focused on identifying spatial patterns rather than estimating the magnitude of spatial effects. In the case of local residents of a protected area, the binary weights in the CAR model may have been able to capture the underlying spatial structure of the data, resulting in a better fit to the data compared to the more detailed spatial regression model.

Another reason why the CAR model may have outperformed the more detailed spatial regression model is that the latter may have suffered from overfitting [37]. When the spatial resolution of the data is high and the spatial regression model includes a large number of covariates or a complex functional form, it can be prone to overfitting. This means that the model fits the noise in the data rather than the true underlying patterns, which can result in poor out-of-sample performance. The binary weights in the CAR model may have provided a more parsimonious and robust approach for modelling the spatial dependence in the data, leading to better out-of-sample predictive performance.

It is worth noting that the choice of spatial model may also depend on the level of spatial resolution of the data [38]. If the data is highly aggregated (e.g., at the level of counties or postcodes/zip codes), then a CAR model with binary weights may be sufficient. However, if the data is more finely resolved (e.g., at the level of individual addresses), then a more detailed spatial regression model may be necessary to accurately capture the spatial variation in the data.

Overall, the choice of model will depend on the specific research question, the nature of the data being analyzed, and the level of spatial resolution of the data. Both types of models have their own strengths and weaknesses, and the choice between them should be based on careful consideration of these factors.

The key contribution of this paper lies in the implementation and intercomparison of different modeling strategies concerning the distributional specification of the dependent variable, as well as the focused implementation of spatial modeling in terms of accounting for spatial proximity effects in the estimation of social impacts in PAs.

In doing this, various alternative distributions have been utilized for the most accurate modeling of social impacts in terms of spatial effects assessment. The results of an ordinal logistic regression based on assigning a logistic distribution on the dependent variable have been compared with other suitable distributions for positive count data, such as the Poisson and the negative binomial, as well as to the generalized Poisson regression model, proposed as an alternative to the typical Poisson model when the response data suffers from underdispersion.

The Poisson regression model is typically used for count data, where the response variable represents the number of occurrences of an event in a fixed period of time or space. However, the Poisson distribution assumes that the mean and variance of the response variable are equal, which may not always be the case in real-world data. When the variance is greater than the mean, a negative binomial regression model may be more appropriate [39]. The generalized Poisson regression model, on the other hand, allows for the variance to be greater than or less than the mean, making it a more flexible option for count data. It achieves this by including an additional parameter, called the dispersion parameter, which allows for more flexibility in modeling the variance of the response variable [40].

The results indicated that the most suitable distribution for such type of response data, being a 5-point Likert scale, is the ordinal logistic regression modeling approach. A useful alternative has been shown to be the Generalized Poisson spatial regression model, which can be attributed to the fact that the specific dataset is an example of under-dispersed data. On the contrary, the negative binomial regression model did not performed well when compared to all other distributional specifications. The finding that the generalized Poisson regression model outperformed the typical Poisson and negative binomial regression models suggests that this particular model may be the most appropriate choice for analyzing the data at hand.

The modelling analysis of our survey findings suggests that the perceived social impacts reported by residents of a protected area are not independent of those reported by people living nearby. That is, their perceptions of impacts are influenced by the perceptions of others living in proximity, the effect diminishing with distance between neighbours. This may be due to a number of factors related to physical distance and social interactions.

Firstly, perceptions of impacts may be clustered because the causes of impacts themselves are clustered spatially, for example, in relation to settlements, infrastructure and processes (e.g., at traffic corridors, hotspots etc.). In the Eifel National Park, Germany, McGinlay et al. [14] found that many reported negative impacts were related to noise and disturbance from visitor traffic in the Park. Such disturbance (noise, air pollution, severance, traffic congestion) was spatially concentrated along road access corridors and around local honeypot sites with tourist attractions, local services and parking facilities. So positive perceptions are likely to cluster around sources of benefits, such as attractive areas for recreation and negative impacts around sources of disturbance.

Further to this, however, people's perceptions of the impacts may be amplified or dampened by social relational influences such as the norms, attitudes, opinions and experiences of their friends and neighbours [41,42]. People living closer together may have more frequent interactions with each other, leading to a greater sense of community and shared experiences related to living in the protected area. They may also have more opportunities to discuss and share their opinions about the social impacts of residing in the area, which could contribute to more nuanced and informed attitudes, modulating their individual impressions with a further relational aspect.

In contrast, residents living further apart from each other may have fewer opportunities to interact with their neighbours, and may be or feel more isolated and disconnected from the community in general. They may therefore be less likely to discuss and share their opinions about the social impacts of living in the protected area, which could contribute to more individualised perceptions and attitudes. Alternatively a given household may simply be more remote from the communities and locations most affected by a particular social impact, which will therefore have less relevance for them.

5. Conclusions

We mapped the perceptions of a range of social impacts as reported by local people living in and around the Peak District National Park in an online survey, in order to investigate the spatial patterns of distribution of impacts. The resultant spatial patterns were modeled using a range of statistical modeling techniques in order to assess which approaches may prove most useful in assessing the spatial distribution of impacts.

Our statistical modeling and mapping work indicates that local people's perceptions of the social impacts of the Park tend to be clustered, that is to say, the reported impacts at one location are not independent of impacts at another location.

Among the tested spatial modeling techniques to capture clustering effects, the conditional autoregressive model (CAR) that assigns a weight for adjacent municipalities and ignores spatial effects of non-adjacent areas gave the best fit for the specific data. This could be attributed to technical reasons such as low spatial resolution or overfitting. This finding is also a strong indication that the effect of spatial dependence in this particular PA diminishes in a very rapid way, or in other words, the proximity effect is much stronger between residents living near to each other (e.g., between friends and close neighbors in same community) and has minimum effects between non-adjacent municipalities. The current analysis has generally shown that the choice of best modelling approach for assessing the spatial effects on social impacts may depend on the specific dataset, research question, and the level of spatial resolution of the data. The alternative types of models presented have their own strengths and weaknesses, and the choice between them should be based on careful consideration of the above factors. The findings of this study and the proposed modeling approach may assist in identifying localities inside or close to protected areas with similar social impacts on local residents.

Understanding the spatial distribution of perceived social impacts is key to understanding their causes, and to finding solutions to manage and mitigate them. Indeed, perceptions of social impacts may be clustered because impacts themselves are clustered in relation to settlements, infrastructure and processes (e.g., at traffic corridors, hotspots etc.), but perceptions may also be clustered because of social factors reinforcing perceptions.

Finally, in order to predict in advance the likely future impacts of a new PA, modeling and prediction tools will be needed to assist practitioners in designing PAs. This work should be of utility to policy makers and practitioners in understanding which modeling and prediction tools will be of most use in this regard.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research programme (Project FIDELIO, grant agreement no. 802605).

Conflict of interest

The authors have no conflict of interest to declare.

References

1. *Convention on Biological Diversity:* Report of the expert workshop on the monitoring framework for the post-2020 global biodiversity framework, 2022. Available from: https://www.cbd.int/doc/c/3190/c3f4/1d9fe2d2dedc8c8b97023750/id-om-2022-01-02-en.pdf.
2. J. Day, N. Dudley, M. Hockings, G. Holmes, D. A. Laffoley, S. Stolton, et al., *Guidelines for applying the IUCN Protected Area Management Categories to Marine Protected Areas*, IUCN, Switzerland, 2012.
3. S. Hernandez, M. D. Barnes, S. Duce, V. M. Adams, The impact of strictly protected areas in a deforestation hotspot. *Conserv. Sci. Pract.*, **3** (2021), e479. https://doi.org/10.1111/csp2.479
4. F. Romagosa, Physical health in green spaces: Visitors' perceptions and activities in protected areas around Barcelone, *J. Outdoor Recreation Tourism*, **23** (2018), 26–32. https://doi.org/10.1016/j.jort.2018.07.002
5. D. Burdon, T. Potts, E. McKinley, S. Lew, R. Shilland, K. Gormley, et al., Expanding the role of participatory mapping to assess ecosystem service provision in local coastal environments. *Ecosyst. Serv.*, **39** (2019), 101009. https://doi.org/10.1016/j.ecoser.2019.101009
6. G. J. Rodrigues, S. Villasante, I. S. Pinto, Non-material nature's contributions to people from a marine protected area support multiple dimensions of human well-being. *Sustainab. Sci.*, **17** (2022), 793–808. https://doi.org/10.1007/s11625-021-01021-x
7. N. Jones, Using perceived impacts, governance and social indicators to explain support for protected areas, *Environ. Res. Lett.*, **18** (2023). https://doi.org/10.1088/1748-9326/acc95b
8. J. McGinlay, N. Jones, C. Malesios, P. G. Dimitrakopoulos, A. Begley, S. Berzborn, et al., Exploring local public support for protected areas: What social factors influence stated and active support among local people?, *Environ. Sci. Policy*, **145** (2023), 250–261. https://doi.org/10.1016/j.envsci.2023.04.003
9. N. Jones, C. Malesios, A. Kantartzis, P. Dimitrakopoulos, The role of location and social impacts of protected areas on subjective wellbeing, *Environ. Res. Lett.*, **15** (2020). https://doi.org/10.1088/1748-9326/abb96e

10. S. J. Campbell, G. J. Edgar, R. D. Stuart-Smith, G. Soler, A. E. Bates, Fishing-gear restrictions and biomass gains for coral reef fishes in marine protected areas, *Conserv. Biol.*, **32** (2017), 401–410. https://doi.org/10.1111/cobi.12996
11. J. S. Brandt, V. Butsic, B. Schwab, T. Kuemmerle, V. C. Radeloff, The relative effectiveness of protected areas, a logging ban and sacred areas for old growth forest protection in southwest China, *Biol. Conserv.*, **181** (2015), 1–8. https://doi.org/10.1016/j.biocon.2014.09.043
12. N. J. Bennett, A. Di Franco, A. Calò, E. Nethery, F. Niccolini, M. Milazzo, et al., Local support for conservation is associated with perceptions of good governance, social impacts, and ecological effectiveness, *Conserv. Lett.*, **12** (2019), e12640. https://doi.org/10.1111/conl.12640
13. R. Naidoo, D. Gerkey, D. Hole, A. Pfaff, A. M. Ellis, C. D. Golden, et al., Evaluating the impacts of protected areas on human well-being across the developing world, *Sci. Adv.*, **5** (2019), eaav3006. https://doi.org/10.1126/sciadv.aav3006
14. J. McGinlay, J. Holtvoeth, A. Begley, J. Dörstel, A. Kockelmann, M. Lammertz, et al., Perceived social impacts of protected areas, their influence on local public support and their distribution across social groups: Evidence from the Eifel national park, Germany, during the COVID-19 pandemic, *Sustainability*, **15** (2023), 10848. https://doi.org/10.3390/su151410848 .
15. A. D. Cliff, J. K. Ord, *Spatial processes: models and applications*, Taylor & Francis, 1981.
16. N. Jones, J. McGinlay, *The impact of COVID-19 restrictions on local communities of Peak District National Park and management options during the Pandemic*, 2021. Report available at: https://www.fidelio.landecon.cam.ac.uk/files/peak_district_report_final.pdf.
17. J. A. Oldekop, G. Holmes,W. E. Harris, K. L. Evans, A global assessment of the social and conservation outcomes of protected areas, *Conserv. Biol.*, **30** (2016), 133–141. https://doi.org/10.1111/cobi.12568
18. N. Jones, J. McGinlay, V. Gkoumas, C. Malesios, V. Maguire-Rajpaul, L. Walker, SOCIAT: Social Impact Assessment Tool for Protected Areas, 2023. Available from: https://warwick.ac.uk/fac/arts/schoolforcross-facultystudies/research/projects/fidelio/sociat/about/.
19. B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia, 1982.
20. B. Efron, D. Rogosa, R. Tibshirani, Resampling methods of estimation, in *International Encyclopedia of the Social & Behavioral Sciences,* Elsevier, New York, (2004), 13216–13220.
21. W. H. Greene, *Models for Event Counts and Duration, Econometric Analysis*, Upper Saddle River: Prentice Hall, (2008), 906–944.
22. P. McCullagh, J. Nelder, *Generalized Linear Models*, 2nd edition, Boca Raton: Chapman and Hall/CRC, 1989.
23. F. Famoye, Restricted generalized Poisson regression model. Communications in statistics, *Theory Methods*, **22** (1991), 1335–1354. https://doi.org/10.1080/03610929308831089
24. C. Malesios, N. Demiris, K. Kalogeropoulos, I. Ntzoufras, Bayesian epidemic models for spatially aggregated count data, *Stat. Med.*, **36** (2017), 3216–3230. https://doi.org/10.1002/sim.7364
25. J. Besag, C. Kooperberg, Biometrika trust on conditional and intrinsic autoregression, *Biometrika*, **82** (1995), 733–746. https://doi.org/10.1093/biomet/82.4.733
26. P. McCullagh, Regression models for ordinal data, *J. Royal Stat. Soc. Series B*, **42** (1980), 109–127. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x

27. E. W. Duncan, N. M. White, K. Mengersen, Spatial smoothing in Bayesian models: a comparison of weights matrix specifications and their impact on inference, *Int. J. Health Geographics*, **16** (2017), 47. https://doi.org/10.1186/s12942-017-0120-x
28. A. Getis, J. Aldstadt, Constructing the spatial weights matrix using a local statistic, *Geographical Anal.*, **36** (2004), 90–104. https://doi.org/10.1111/j.1538-4632.2004.tb01127.x
29. P. Congdon, Random-effects models for migration attractivity and retentivity: a Bayesian methodology, *J. Royal Stat. Soc. Series A*, **173** (2010), 755–774. https://doi.org/10.1111/j.1467-985X.2009.00625.x
30. Z. Fric, V. Hula, M. Klimova, K. Zimmermann, M. Konvicka, Dispersal of four fritillary butterflies within identical landscape, *Ecol. Res.*, **25** (2010), 543–552. https://doi.org/10.1007/s11284-009-0684-4
31. D. Obaromi, Spatial modelling of some conditional autoregressive priors in a disease mapping model the Bayesian approach, *Biomed. J. Sci. Techn. Res.*, **14** (2019), 10681–10686. http://dx.doi.org/10.26717/BJSTR.2019.14.002555
32. H. Xia, B. P. Carlin, Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality, *Stat. Med.*, **17** (1998), 2025–2043.
33. A. Earnest, G. Morgan, K. Mengersen, L. Ryan, R. Summerhayes, J. Beard, Evaluating the effect of neighbourhood weight matrices on smoothing properties of conditional autoregressive (CAR) models, *Int. J. Health Geographics*, **6** (2007), 1–12. https://doi.org/10.1186/1476-072x-6-54
34. D. J. Lunn, A. Thomas, N. Best, D. Spiegelhalter, WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility, *Stat. Comput.*, **10** (2000), 325–337. https://doi.org/10.1023/A:1008929526011
35. D. J. Spiegelhalter, N. G. Best, B. P. Carlin, A. Van Der Linde, Bayesian measures of model complexity and fit, *J. Royal Stat. Soc. Series B*, **64** (2002), 583–639. https://doi.org/10.1111/1467-9868.00353
36. S. E. Ahmed, A. Hussein, M. Al-Momani, Efficient estimation for the conditional autoregressive model, *J. Stat. Comput. Simul.*, **85** (2015), 2569–2581. https://doi.org/10.1080/00949655.2014.893346
37. C. Merow, M. J. Smith, T. C. Edwards Jr, A. Guisan, S. M. McMahon, S. Normand, et al., What do we gain from simplicity versus complexity in species distribution models?, *Ecography*, **37** (2014), 1267–1281. https://doi.org/10.1111/ecog.00845
38. M. Pogson, P. Smith, Effect of spatial data resolution on uncertainty, *Environ. Model. Software*, **63** (2015), 87–96. https://doi.org/10.1016/j.envsoft.2014.09.021
39. J. O. Lloyd-Smith, Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases, *PLoS ONE*, **2** (2007), e180. https://doi.org/10.1371/journal.pone.0000180
40. P. C. Consul, F. Famoye, Generalized poisson regression model, *Commun. Stat. Theory Methods*, **21** (1990), 89–109. https://doi.org/10.1080/03610929208830766
41. D. Pennington, K. Gillen, P. Hill, *Social Psychology*, London: Hodder Education, (1999), 244–278.
42. K. Keizer, P. W. Schultz, Social norms and pro-environmental behaviour, in *Environmental Psychology – An Introduction*, (eds. L. Steg, A.E. Van den Berg and J.I.M. de Groot), Chichester: BPS Blackwell, (2013), 153–163.

Appendix

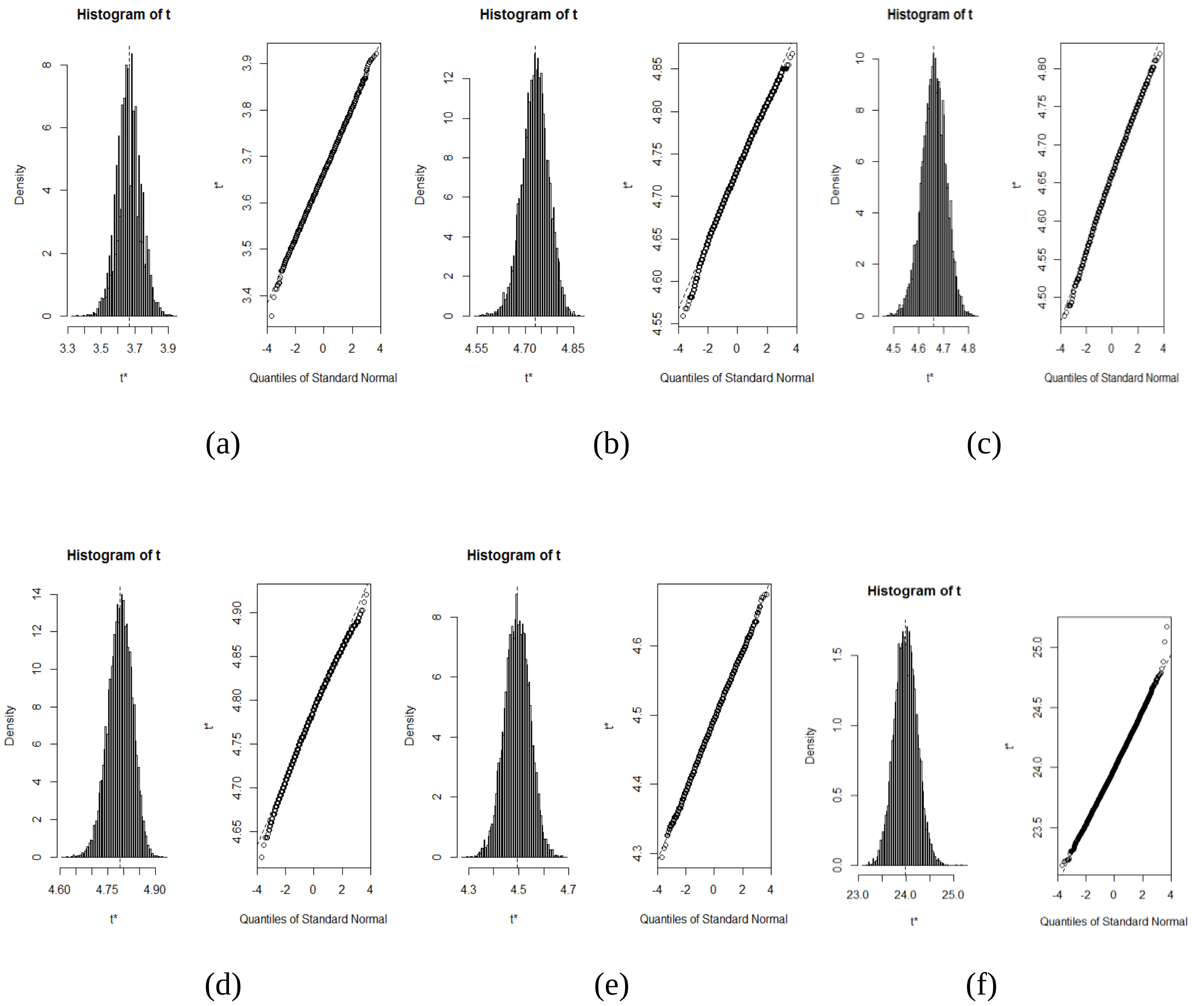


Figure A1. Histograms and normal probability plots for the bootstrap samples based upon the 10,000 iterations (a. Personal income; b. Quality of life; c. Recreational activities; d. Connectedness to Nature; e. Social relations; f. Average distance (km)).

Table A1. Pearson's correlation coefficients for the bootstrap samples.

	Personal income	Quality of life	Recreational activities	Connectedness to Nature	Social relations	Average distance d (km)
Personal income	1					
Quality of life	0.450*	1				
Recreational activities	0.338*	0.716*	1			
Connectedness to nature	0.363*	0.749*	0.725*	1		
Social relations	0.381*	0.545*	0.465*	0.523*	1	
Average distance (km)	0.229*	0.430*	0.329*	0.377*	0.494*	1

*. Correlation is significant at the 0.01 level.