Geographically weighted regression models for ordinal categorical response variables: An application to geo-referenced life satisfaction data

Guanpeng Dong[a,*], Tomoki Nakaya[b], Chris Brunsdon[c]

[a] Department of Geography and Planning, University of Liverpool, Room 713, Roxby Building, Chatham St, Liverpool L69 7ZT, UK
[b] Department of Geography, Ritsumeikan University, Japan, 56-1 Tojiin-kita-machi, Kita-ku Kyoto, 603-8577, Japan
[c] The National Centre for Geocomputation, Maynooth University, Iontas Building, Maynooth University, Maynooth, Kildare, Ireland

ARTICLE INFO

ABSTRACT

Ordinal categorical responses are commonly seen in geo-referenced survey data while spatial statistics tools for modelling such type of outcome are rather limited. The paper extends the local spatial modelling framework to accommodate ordinal categorical response variables by proposing a Geographically Weighted Ordinal Regression (GWOR) model. The GWOR model offers a suitable statistical tool to analyse spatial data with ordinal categorical responses, allowing for the exploration of spatially varying relationships. Based on a geo-referenced life satisfaction survey data in Beijing, China, the proposed model is employed to explore the socio-spatial variations of life satisfaction and how air pollution is associated with life satisfaction. We find a negative association between air pollution and life satisfaction, which is both statistically significant and spatially varying. The economic valuation of air pollution results show that residents of Beijing are willing to pay about 2.6% of their annual income for per unit air pollution abatement, on average.

1. Introduction

Geographically weighted regression (GWR) has been established as a flexible framework for modelling spatially varying relationships between predictor variables and an outcome variable (Brunsdon, Fotheringham, & Charlton, 1996; Fotheringham, Brunsdon, & Charlton, 2003). Recent years have seen active methodological development of GWR models due to an increasing demand of applying localised spatial models to data with complex structures and non-Gaussian types of outcome variables. For instance, GWR models have been extended to explore spatiotemporal data by incorporating temporal correlations between an observation at time period *t* and spatially nearby observations at previous periods into the overall local weights matrix for estimation (Fotheringham, Crespo, & Yao, 2015; Huang, Wu, & Barry, 2010). Harris, Dong, and Zhang (2013) presented a contextualized GWR model, thereby the contextual similarities of observations (e.g. similarity in the attributes of neighbourhoods where houses are located, measured by certain distance metric) were incorporated into the local weights matrix for implementing each local regression model. The key idea underlying this line of GWR model extension is to achieve a better or more realistic representation of spatial relationships between observations. Other methodological elaborations of GWR include developing formal statistical tests of spatial heterogeneity (Leung, Mei, &

Zhang, 2000), and the use of different distance metrics in constructing the spatial weights matrix (Lu, Charlton, Brunsdon, & Harris, 2016).

This paper contributes to the ongoing GWR developments by extending a geographically weighted ordinal regression model (GWOR) for properly exploring spatial data with ordinal categorical response variables. Ordinal response variables are commonly seen in social science research, especially when the research focus is in relation to individual opinions and attitudes towards events, or subjective assessment of life experiences such as life satisfaction and happiness. Detailed descriptions of the application scopes of ordinal response variables in a variety of social science disciplines are provided in Agresti (2010) and Greene and Hensher (2010). The motivation of extending a GWOR model lies in two aspects. The first is to address the issue of limited methodological options to deal with increasingly available geo-referenced survey data in the local spatial modelling literature. Such data usually quantify important information via categorical variables. Secondly, we are interested in exploring the socio-spatial variation of life satisfaction in Beijing, China and examining potential spatial heterogeneity in the association between life satisfaction and air pollution. The GWOR model, demonstrated by examining a geo-referenced life satisfaction data, can be applied to other spatial data.

Life satisfaction data are often collected based on surveys, in which questions such as "Overall, how satisfied are you with your life?" are

* Corresponding author.
E-mail addresses: Guanpeng.dong@liverpool.ac.uk (G. Dong), nakaya@ritsumei.ac.jp (T. Nakaya), Christopher.Brunsdon@nuim.ie (C. Brunsdon).

asked (e.g. Welsch & Ferreira, 2014). Responses to the question are usually recorded on a Likert scale, ranging from one being very unsatisfied to five being very satisfied, for instance. Life satisfaction scores or ratings are quantitative in nature but the between-category distances are unknown—the distance between categories of one and two might be quite different from that between categories of four and five (Agresti, 2010). This differs substantially from a Gaussian variable for which per unit difference is comparable. It has been shown that applying a linear regression model to an ordinal categorical variable would cause issues to model estimation and statistical inferences for regression coefficients, likely producing misleading model results (Agresti, 2010).

To date, the development of GWR focuses on outcome variables following a Gaussian (or Normal) distribution, with few notable exceptions in Nakaya, Fotheringham, Brunsdon, and Charlton (2005) where a geographically weighted Poisson regression model has been developed for exploring disease outcomes, and in Fotheringham et al. (2003) for a geographically weighted Binomial regression models. Closely related to this study, McMillen and McDonald (2004) presented a preliminary extension of GWR to geo-referenced ordinal response variables by proposing a locally weighted ordered probit model. However, discussions on model specifications such as choices of different link functions for the cumulative probabilities of responses and choices of adaptive or fixed kernels, and on approaches to test statistical significance of spatial heterogeneity in regression coefficients are not provided.

This paper extends the work of McMillen and McDonald (2004) by offering flexible tools to explore the potential spatial variability in relationships between an ordinal response variable and predictor variables. The estimation of GWOR model draws upon the locally weighted likelihood approach via an iterative numerical optimisation procedure (detailed below). It allows for great flexibility in model specification, including different link functions (logit and probit) for the cumulative probabilities of the responses, and a mixed model specification in which regression coefficients of some variables are spatially varying while coefficients of other variables are kept spatially invariant (e.g. Mei, Xu, & Wang, 2016). The R code for implementing various GWOR models are provided in the *Supplementary Information* of the paper.

Based on a geo-referenced life satisfaction survey data in Beijing, this study explores how life satisfaction is spatially linked to air pollution and other factors. There has been a surge of using life satisfaction data to evaluate environmental amenities such as air quality, the economic value of which cannot be directly observed through market transactions (Ferreira & Moro, 2013; MacKerron & Mourato, 2009; Welsch, 2006). The theoretical underpinnings of life satisfaction based environmental evaluation approach are comprehensively reviewed in Welsch and Ferreira (2014). At its heart, the subjective life satisfaction is regarded as the experienced utility of individuals, and by estimating the life satisfaction equation with environmental quality indicators and income included, the (marginal) willingness to pay (WTP) for environmental quality can be estimated (e.g. Ferreira & Moro, 2010). The issue, however, is that the life satisfaction equation was predominantly estimated by using aspatial regression models, implicitly assuming WTP for environmental quality improvement to be constant across space. This is a rather restrictive assumption. It is likely that people living in different locations with varying socio-economic characteristics tend to have different preferences for air quality and thus varying WTP for air quality improvement or air pollution abatement (Bayer, Ferreira, & McMillan, 2007; Ferreira & Moro, 2013). The GWOR model enables us to estimate the spatially varying associations between life satisfaction and air pollution and income, taking locational heterogeneity into account.

The remainder of this paper is organised as follows. Section 2 provides an overview of non-spatial ordinal regression models. In Section 3, we describe the GWOR model and provide details of model estimation. Section 4 applies the GWOR model to explore the socio-spatial variation of life satisfaction and estimate the economic value of air

quality in Beijing. Conclusions are provided in Section 5.

2. A non-spatial ordinal regression model

Following Agresti (2010) and Greene and Hensher (2010), we use a latent variable approach to formulate an ordinal regression model due to its intuitive link to the simple linear regression models. Denote Y_i^* as a latent continuous outcome variable and x_i a set of predictor variables such as income, air pollution, and others. A linear regression model links Y_i^* to x_i,

$$Y_i^* = x_i\boldsymbol{\beta} + \epsilon_i; \ i = 1, \ ...,N \tag{1}$$

where i indexes each observation and N, the sample size. $x_i = [x_{i,1}, x_{i,2}, ..., x_{i,p}]$ is a row-vector of predictor variable values of observation i while $\boldsymbol{\beta}$ is a column-vector of regression coefficients to estimate. The mapping of the unobservable Y_i^* to the observed categorical response Y_i depends on a set of cut-off points or threshold values $[\alpha_0, \alpha_1,..., \alpha_J]$ on the scale of Y_i^*: $Y_i = j$ if $a_{j-1} < Y_i^* \le a_j$, $j = 1,..., J$ where J are the number of response categories. Ordinal regression models focus on the cumulative probability of an observation falling in category j or below, which is expressed as,

$$P(Y_i \le j) = P(Y_i^* \le a_j) = P(\epsilon_i \le a_j - x_i\boldsymbol{\beta}) = F(a_j - x_i\boldsymbol{\beta}). \tag{2}$$

Different specifications of the density function for ϵ leads to different forms of cumulative probabilities for $P(Y_i \le j)$: $1/(1 + exp(-a_j + x_i\boldsymbol{\beta}))$ if a logistic density was specified, and $\Phi(a_j - x_i\boldsymbol{\beta})$ if a Normal density was used where Φ is the cumulative distribution function of a standard Normal density. The logistic specification was favoured due to its simplicity in model parameter interpretation (Agresti, 2010). The probability of $(Y_i = j)$, conditioning on x_i, is $F(Y_i^* \le a_j) - F(Y_i^* \le a_{j-1})$. The GWOR models extended here allows for both Normal and logistic densities for ϵ.

The effect of a predictor variable, say x_1, on the cumulative probability of a response falling into category j is not linear because of the non-linear cumulative distribution function. This is seen from the partial derivative of the cumulative probability with respect to x_1, $\partial P(Y_i \le j)/\partial x_1 = f(a_j - x_i\boldsymbol{\beta})\beta_1$ where $f(.) = F'(.)$ is the density function and β_1 the regression coefficient of x_1. The interpretation of estimated coefficients can make use of the concept of odds ratios as in a simple binary logit model. Taking the log odds of the cumulative probability in (2) and inserting the cumulative logistic probability formula, we obtain,

$$log \frac{P(Y_i \le j)}{1 - P(Y_i \le j)} = a_j - x_i\boldsymbol{\beta}. \tag{3}$$

The equation shows that the effect of x_1 on the cumulative probability on the logit scale is simply β_1 regardless the response category. The maximum likelihood estimation approach is usually used for model estimation. For observation i, let $y_{i1}, ..., y_{i,J}$ be binary indicators of response categories, then we have $y_{ij} = 1$ and $y_{ik} = 0$ for $k \ne j$ if $Y_i = j$. The log-likelihood function of the model is,

$$l(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \log[F(Y_i^* \le a_j) - F(Y_i^* \le a_{j-1})] \tag{4}$$

where $\theta = [\boldsymbol{\beta}, \alpha_1, ..., \alpha_{J-1}]$. It's useful to note that only $J-1$ cut points are needed to divide the latent variable Y^* into J categories while α_0 and α_J are set to $-\infty$ and $+\infty$, respectively. Although there is not a tractable solution for the first-order conditions of the equation, it has been shown that the log-likelihood function has a unique global optimum so different types of iterative maximisation algorithms can be applied to estimate θ (Burridge, 1981; Pratt, 1981).

3. A geographically weighted ordinal regression model

We now describe the geographically weighted ordinal regression model that allows for regression coefficients varying across space.

Following the GWR notational convention (Fotheringham et al., 2003), let u_i ($u_{i,x}$, $u_{i,y}$) be the geographical coordinates of observation i. Eq. (1) can be re-written as,

$$Y_i^* = x_i \boldsymbol{\beta}(u_i) + \epsilon_i; \ i = 1, \ ..., N$$

$$Y_i = 1 \ if - \infty < Y_i^* \leq a_1(u_i)$$

$$Y_i = j \ if \ a_{j-1}(u_i) < Y_i^* \leq a_j(u_i), j = 2, \ ..., J - 1$$

$$Y_i = J \ if \ a_{J-1}(u_i) < Y_i^* \leq +\infty \tag{5}$$

Both the $J - 1$ cut points and the coefficient vector $\boldsymbol{\beta}$ to estimate are associated with a location indicator u_i, relaxing the restrictive assumption of spatially invariant covariate effects. Following previous studies (Loader, 2006; Nakaya et al., 2005; Páez, Uchida, & Miyamoto, 2002), a locally weighted local likelihood approach was employed to estimate parameters in GWOR.

3.1. Model estimation

The log-likelihood function at the fit (or focal) location i is expressed as,

$$
\begin{aligned}
l(\theta(u_i)) &= \sum_{k=1}^{N} \sum_{j=1}^{J} w_{ik} y_{kj} \log[Y_k = j \mid x_k] \\
&= \sum_{k=1}^{N} \sum_{j=1}^{J} w_{ik} y_{kj} \log[F(a_j(u_i) - x_k \boldsymbol{\beta}(u_i)) \\
&\quad - F(a_{j-1}(u_i) - x_k \boldsymbol{\beta}(u_i))],
\end{aligned}
\tag{6}
$$

where w_{ik} is the geographical weight placed on the k-th observation when fitting the local regression model at location i. A weighting function is needed to determine the rate of decrease in weights assigned to observations when moving further away from the fit point. A commonly employed weighting kernel is the Gaussian kernel, in which $w_{ik} = exp.(-0.5 \times |u_i - u_k|/d^2)$ with $|u_i - u_k| = (u_{i,x} - u_{k,x})^2 + (u_{i,y} - u_{k,y})^2$. The parameter d is the kernel bandwidth, which controls how fast the weights decline with increasing distances away from the fit location. Alternative kernel functions such as bi-square kernel are also widely used (Fotheringham et al., 2003; Páez et al., 2002). A further decision is to choose between a fixed kernel approach where the distance bandwidth is fixed for all observations, and an adaptive kernel approach in which the bandwidth varies with each fit location while keeping the number of nearest observations from each fit point same.

An iterative maximisation algorithm is employed to estimate model parameters at location i. The maximum of each local log-likelihood function can be found by using the Newton-Raphson approach, as illustrated below (Franses & Paap, 2001, p.119),

$$\theta_m = \theta_{m-1} - H^{-1}(\theta_{m-1})G(\theta_{m-1}) \tag{7}$$

where $G(\theta)$ and $H(\theta)$ are the first-order and second-order derivatives of the local Log-likelihood function $l(\theta(u_i))$ with respect to model parameters $\boldsymbol{\theta}$, which are also known as the gradient and Hessian matrix of a likelihood function, respectively. From (6), the gradient vector can be obtained as,

$$G(\theta(u_i)) = \frac{\partial l(\theta(u_i))}{\partial \theta(u_i)} = \sum_{k=1}^{N} \sum_{j=1}^{J} \left[\frac{w_{ik} y_{kj}}{P(Y_k = j \mid x_k)} \frac{\partial P(Y_k = j \mid x_k)}{\partial \theta(u_i)} \right] \tag{8}$$

It becomes clear that only the conditional probability $P(Y_k = j|x_k)$ is involved in deriving the gradient vector. For notational simplicity, we denote $P(Y_k = j|x_k)$ as $P(Y_k = j)$ and the first-order derivative of the cumulative distribution $F(.)$ as $f(.)$. Then we have,

$$
\begin{aligned}
\partial P(Y_k = j)/\partial P(Y_k = j) &= [x_k(f(a_{j-1}(u_i) - x_k \boldsymbol{\beta}(u_i)) \\
&\quad - f(a_j(u_i) \\
&\quad - x_k \boldsymbol{\beta}(u_i))), 0, \\
&\quad ..., -f(a_{j-1}(u_i) - x_k \boldsymbol{\beta}(u_i)), f(a_j(u_i) - x_k \boldsymbol{\beta}(u_i)), ..., 0)',
\end{aligned}
\tag{9}
$$

which is a $(p + J - 1)$ column vector with p being the number of covariates. The Hessian matrix $H(\boldsymbol{\theta}(u_i))$ follows naturally as the derivative of (9) with respect to $\boldsymbol{\theta}(u_i)'$. To preserve the ordering of the cut points $\boldsymbol{a}$, a re-parameterisation approach was employed (Greene & Hensher, 2010). For example, instead of maximising each local log-likelihood function over the space of $\boldsymbol{a}$, we can re-parameterise $\boldsymbol{a}$ to a vector γ such that $a_j = a_{j-1} + exp.(\gamma_j), j = 1, 2, ..., J - 1$. Then, maximisation is applied to the local log-likelihood over $[\beta, \gamma]$.

3.2. Statistical inferences and tests of spatial heterogeneity

The statistical inference of model parameters at location i is drawn upon the Hessian matrix $H(\boldsymbol{\theta}(u_i))$ (Nakaya et al., 2005). The asymptomatic variance-covariance matrix of regression coefficients is produced by the negative inverse of the Hessian matrix, $-H(\boldsymbol{\theta}^{-1}(u_i))$. Local pseudo t statistics can be calculated consequently and used for inspection of statistical significances on local coefficients. As with other local spatial modelling techniques, GWR models also suffer from the issue of multiple hypothesis testing (Brunsdon & Charlton, 2011; da Silva & Fotheringham, 2016), the statistical significance of local parameters should be interpreted with caution.

In terms of the statistical significance test on the spatial non-stationarity of regression coefficients, a Monte Carlo simulation approach is used (Brunsdon, Fotheringham, & Charlton, 1999; Fotheringham et al., 2003). It proceeds as follows: (1) Calculate the variance of the estimated local coefficients based on the real spatial data configuration; (2) Randomly permute the spatial configuration of the real data, run the GWOR model and calculate the variance of local coefficients from each permutation. Repeat this step for n times, say 99; (3) Compute the rank of the variance calculated from Step (1) in the empirical distribution of all variances calculated from Step (2), R. Finally, a p-value can be approximated by $R/(1 + n)$.

To implement the GWOR model, the bandwidth parameter, either the distance threshold or the number of nearest neighbours needs to be found. We use a cross-validation approach to select the bandwidth in the GWOR model, as specified below,

$$CV = \sum_{i=1}^{N} (1 - [\widehat{P}(Y_i = j)]^{y_{ij}})^2. \tag{10}$$

An optimum bandwidth can be thought of a bandwidth that maximises the predicted probabilities for the chosen response categories for all observations (Agresti, 2010; Greene, 2003).

We also provide a tentative mix GWOR model to allow part of the regression coefficients spatially invariant (Fotheringham et al., 2003; Mei et al., 2016). This is useful in situations where there were not theoretical or empirical evidences suggesting the effects of certain predictors on the outcomes are spatially varying. This is the case in our life satisfaction study. For instance, it would be difficult to argue that the gender effect on life satisfaction is spatially varying. Alternatively, the spatial variability of relationships might not pass the statistical significance test discussed above. In a mixed GWOR model, Eq. (5) changes to,

$$Y_i^* = \sum_p \boldsymbol{\beta_p}(u_i)x_{i,p} + \sum_m \boldsymbol{\delta_m}(u_i)x_{i,m} + \epsilon_i, \tag{11}$$

where the coefficients associated with variables x_p are allowed spatially varying while the coefficients of variables x_m are kept constant. The estimation of a mixed GWR model is usually carried out by using an iterative procedure: the estimation for the fixed parameters $\boldsymbol{\delta_m}$ and for the varying parameters $\boldsymbol{\beta_p}(u_i)$ are iterated until convergence is achieved (Lu, Harris, Charlton, & Brunsdon, 2014; Nakaya et al., 2005). We are,

however, not pursuing such a complex and computationally intensive algorithm to estimate the mixed GWOR model here. Instead, we simply set the fixed coefficients to their estimates from a global ordinal response model. Local log-likelihood functions at each location are maximised to calibrate the spatially varying regression coefficients while keeping the coefficients δ_m invariant. Multi-collinearity has been presented a more serious issue in GWR modelling than in global models as global correlations among independent variables might be elevated in local samples (Wheeler, 2007; Wheeler & Waller, 2009). However, this argument has been challenged by recent evidences based on extensive Monte Carlo simulations. Fotheringham and Oshan (2016) shows that local parameter estimation in GWR is not sensitive to multi-collinearity if the level of multi-collinearity is not extremely high, especially for spatial data with large sample size. There are no theoretical reasons to expect multi-collinearity will present a bigger issue for GWOR than GWR, although a careful check on correlations among explanatory variables and excluding highly correlated variables are recommended when building statistical models. A Bayesian spatially varying coefficients model presents a useful alternative to GWR modelling if statistical inferences on local parameters and heterogeneity are of particular interest (Harris et al., 2013; Wheeler & Waller, 2009).

3.3. Implementation of GWOR

We coded the GWOR model using the R language with functions available in the *Supporting Information* accompanying the paper. There is great flexibility as to how users can specify their own GWOR models, including the use of different spatial weighting schemes (adaptive versus fixed kernels) and different kernel functions (Gaussian, exponential and bi-square kernels). In addition, both logit and probit link functions are available for the cumulative probability distributions. The maximisation of each local log-likelihood function is achieved by using the optimisation routines provided in the R *maxLik* package, a package dedicated to offering easy access to maximum likelihood estimation (Henningsen & Toomet, 2011).

4. Empirical analysis of life satisfaction

4.1. Data and variables

Our analysis mainly draws on a life satisfaction survey conducted in 2013 in Beijing, the key aim of which is to assess the socio-spatial variations in residents' life satisfaction. Only residents living in their current residences for at least six months were selected. A spatial stratified random sampling approach was enforced, with about 0.1% of the urban population in each of Beijing's six districts sampled. The total population in the study area is about 11 million in 2010 (Zhang, Yu, Li, & Dang, 2015). In total, about 7000 questionnaires were issued and about 6000 were returned (self-completion by post), of which about 5700 were valid. The samples are representative of Beijing's urban population for key socio-demographic indicators (Dong, Ma, Harris, & Pryce, 2016; Ma, Mitchell, Dong, & Zhang, 2017; Zhang et al., 2015). We further select residents who are not migrants, i.e. respondents whose household registration (*hukou*) addresses are in Beijing. Migrants tend to evaluate life satisfaction and environmental quality in a way that is quite different from local residents do (Chen, Chen, & Landry, 2013; Ma et al., 2017). A key feature of this survey data is its records of residence address for each respondent. Based on this information, the life satisfaction data have been geo-coded, which makes possible of extracting local air pollution levels and other locational variables for each respondent. Dropping samples with incomplete information on residential locations and key socio-demographic variables yields a final sample size of 2656 used in our analysis. The sub-district boundaries and spatial distributions of selected samples are shown in Fig. 1.

Life satisfaction is our outcome variable measured by the question: Overall, how are you satisfied with your life? The responses were

Fig. 1. The study area and spatial distribution of samples.

Table 1
Descriptive summaries of variables in the study.

Variable names	Description	Means (standard deviation)
Life satisfaction	1 = very unsatisfied	0.5%
	2 = unsatisfied	3.0%
	3 = fair	27.5%
	4 = satisfied	62.9%
	5 = very satisfied	6.2%
Log Income	Log of family's monthly income	9.1(0.7)
Pollution	Annual average concentration of PM$_{2.5}$ ($\mu g/m^3$)	83.5(8.2)
Age	< 30	35.6%
	30–39	35.1%
	40–49	21%
	50–59	8.3%
Female	Male as base category	49.2%
Marital status	Married	66.5%
Housing tenure	Owners	70.1%
Housing type	Commodity housing	8.0%
	Affordable housing	51.2%
	Danwei housing	21.5%
	Self-built housing	19.2
Education	Bachelor degree or above	75.9
Child presence	Household with child under 6	13.6
Distance to CBD	Log of distance to the city centre	2.2(0.6)
Distance to railway	Log of distance to the nearest subway station	0.2(0.9)
Log land price	The average land price around each residence	8.9(1.1)
N	Sample size	2656

quantified on a 5-point Likert scale ranging from 1 (very unsatisfied) to 5 (very satisfied). The majority (> 60%) of respondents were satisfied with their lives while about 27.5% of respondents rated life satisfaction level as fair (Table 1). Like many other life satisfaction surveys (e.g. Ferreira & Moro, 2010), only small proportions of respondents reported ratings of either very satisfied or very unsatisfied.

Our air pollution data is compiled from the real-time air pollution monitoring data, hosted by the Beijing Environmental Protection Bureau (BJEPB). There are 35 air pollution monitoring sites in Beijing, producing hourly readings for various air pollutants. As the particulate matter with a diameter of 2.5 μm or less (PM$_{2.5}$) is of particularly toxic

and more linked to health issues than other air pollutants in China (Zhang, Qi, Jiang, Zhou, & Wang, 2013), we use the concentration of PM$_{2.5}$, measured by μg/m^3, as a proxy variable for air pollution. Following Ferreira et al. (2013), an inverse distance weighted interpolation (IDW) approach was employed to extract the annual mean air pollution levels for each residential location in 2013.

Other independent variables are broadly divided into several categories. The first is a range of individual and household socio-economic and demographic characteristics, including monthly income, age, gender and family structure (the presence of children). It is useful to note that the original income variable is coded as a categorical variable with seven bands. For the purpose of valuing air pollution, income has been converted to a continuous variable by using the midpoint of each income band; and for the open-ended top category by using an extrapolation approach (Ferreira & Moro, 2013). Housing characteristics of tenure (owning or renting a home) and type (commodity, affordable, *Danwei* or self-built) are also included in our model (Table 1) as they are related to the socio-spatial variations of life satisfaction (Ma et al., 2017). In addition, two locational variables are also included in the model: the Euclidian distances of each residence to the city centre and to the nearest railway station. As suggested in Ferreira and Moro (2010), the average per-square meters price of residential land parcels leased from 2003 to 2013 within a radius of two kilometers of each residence was also included in the analysis to control for the capitalisation effect of air pollution on land prices. Land prices have been adjusted to the 2003 constant price by using the consumer price indices of Beijing (Harris et al., 2013). Finally, district dummy variables are included in the model to control for the possible contextual effect on life satisfaction arising from the socio-economic variations across sub-districts.

4.2. Life satisfaction model specification

The empirical life satisfaction model is specified as,

$$satisfaction_i^* = \gamma_i Log\ Income_i + \beta_i Pollution_i + L_i \delta_i + X_i \phi + \epsilon_i \quad (12)$$

where $satisfaction_i^*$ is the latent life satisfaction of respondent *i*. The key predictor variables of interest are air pollution (*Pollution*) and income (*Log Income*), and the regression coefficients of them are allowed to vary spatially. It has been shown that the associations between locational factors (here L_i) and subjective travel satisfaction are spatially varying (Dong et al., 2016). We test whether similar arguments hold in the case of life satisfaction. X_i includes other predictor variables, the effects of which are assumed to be spatially constant.

Following the environmental evaluation literature (e.g. Welsch & Ferreira, 2014), the latent variable $satisfaction_i^*$ represents the experienced utility of individual *i*, which leads to the formula derived for economic valuation of air pollution (or the marginal WTP for air quality improvement),

$$WTP_i = \frac{\partial satisfaction_i^*}{\partial Pollution_i} \bigg/ \frac{\partial satisfaction_i^*}{\partial Income_i} = \frac{\beta_i}{\gamma_i} \times Income_i \quad (13)$$

Essentially, the ratio of the two first-order partial derivatives in the equation produces the marginal rate of substitution between income and air pollution, which measures the amount of money placed on a marginal change in air pollution levels (Ferreira & Moro, 2010). With respect to the marginal WTP, the interpretation in this study is how much money are people willing to pay for a one unit improvement in air quality, holding the level of latent life satisfaction constant.

5. Model estimation results

5.1. Global ordinal regression model results

Two global ordinal regression models with a logit link function were

Table 2
Model estimation results from global ordinal regression models.

Variables	Coefficients	Std. error	Coefficients	Std. error
Log Income	0.43**	0.061	0.291**	0.066
Pollution	−0.037**	0.014	−0.05**	0.015
Age				
30–39			−0.246*	0.12
40–49			−0.184	0.148
50–59			−0.177	0.188
Female			−0.139	0.081
Married			0.242*	0.123
College degree			0.344**	0.103
Child presence			−0.374**	0.128
Housing tenure			0.341**	0.096
Work-unit housing			−0.454**	0.114
Affordable housing			−0.387**	0.105
Self-built housing			−0.119	0.162
Log distance to city centre			0.324**	0.115
Log distance to railway station			−0.226**	0.058
Average land price			−0.035	0.037
District fixed effects	Yes		Yes	
β/γ	−0.087*	0.034	−0.171**	0.064
Willingness to pay	979 [220, 1739]		1926 [519, 3334]	
Log likelihood	−2485		−2440	
AIC	4996		4933	

* represent statistical significance levels of 0.05.
** represent statistical significance levels of 0.01.

first estimated to explore the socio-spatial variation in life satisfaction in Beijing. We started with estimating a baseline model only with the income and air pollution variables included. Then, a range of socio-economic, demographic and locational variables are incorporated into the second model. Following Brown, Oueslati, and Silva (2016), the regression coefficient of a variable is interpreted as an income-equivalent marginal effect on life satisfaction, calculated by dividing it by the income coefficient.

Table 2 present estimation results. We also calculated the marginal WTP using (13). The standard errors of the WTP estimates are obtained by using the delta method. The estimated income and air pollution effects (Column 2) on life satisfaction are both statistically significant at the 1% significance level. Air pollution is negatively associated with life satisfaction, with a one unit decrease in air pollution having an equivalent average effect on life satisfaction as an about 8.7% increase of monthly income or about 0.7% increase of annual income, holding other variables constant. Translating into marginal WTP, people are willing to pay about 979 RMB (*Chinese yuan*) with a 95% confidence interval of [220, 1739] for a marginal decrease in air pollution, evaluated at the mean monthly income level.

In the next model with a range of socio-economic, demographic and locational predictors included, the model fit was significantly improved, as indicated by the substantial increase in the log likelihood value and the decrease of AIC comparing to the first model. The associations of income and air pollution to life satisfaction remain statistically significant. In fact, the ratio of two estimated coefficients almost doubled from −0.087 to −0.171, indicating that the equivalent effect of a one unit decrease in air pollution on life satisfaction amounts to the effect of a 17.1% increase in monthly income (or 1.4% increase of annual income), everything else equal. The marginal WTP for air quality is subsequently increased to about 1926 RMB with a 95% confidence interval of [519, 3334].

Comparing to baseline age group (below 30), statistically significant differences in life satisfaction were found for the age group of 30–39 but not for others. As is often found in the literature, higher educational achievement is associated with higher levels of life satisfaction, ceteris paribus. Married people is also positively associated with higher levels

of life satisfaction, having an equivalent effect of an 83.1% increase in monthly income or 6.9% increase of annual income. Child presence is, however, negatively associated with life satisfaction. Our findings on how housing tenure and types are associated with life satisfaction are largely in line with previous life satisfaction studies in China: home owners tend to report higher levels of life satisfaction; residents living in work-unit and affordable housing tend to report statistically significantly lower levels of life satisfaction than those living in commodity housing (e.g. Bian, Zhang, Yang, Guo, & Lei, 2015; Liu, Wang, & Chen, 2017; Ma et al., 2017).

With respect to locational factors, the proximity to city centre is negatively associated with life satisfaction, suggesting people who lives further away from the centre tend to be more satisfied than those living closer to the city centre, holding other variables constant. This might point to the net negative externalities associated with living around the city centre. Although being close to the city centre implies greater access to employment opportunities and other facilities, the housing pressure (high prices and overcrowded spaces), as well as other negative externalities such as noise, traffic congestion and high-intensity land development are also greater. Being close to subway stations is significantly related to higher levels of life satisfaction. There is an insignificant association between land prices and life satisfaction, controlling for the capitalisation effects of locational variables and air pollution on land prices. However, as mentioned above, the uneven spatial distributions of locational facilities and the process of residential sorting primarily driven by household income might lead to spatially varying preferences for air pollution, which will be tested in the GWOR models.

5.2. GWOR model estimation results

This section explores the spatial variations in the associations between air pollution, income, locational variables and life satisfaction while keeping other regression coefficients constant across spaces. The results reported in Table 3 are based on a fixed Gaussian kernel approach with a distance bandwidth of 15.8 km obtained by using the cross-validation procedure outlined above. We also estimated an adaptive GWOR model with an optimum bandwidth of 956 nearest neighbours, which accounts for roughly 36% of the total sample. It takes about 35 min and 41 min for implementing the fixed and adaptive GWOR models on a MacBook with a 3.3 GHz Intel Core i5 processor, respectively. The obtained spatial patterns of local coefficients are similar. Ideally, we would want to develop a formal statistical test for choosing between models with adaptive and fixed kernels, as in GWR models for continuous response variables. However, due to the
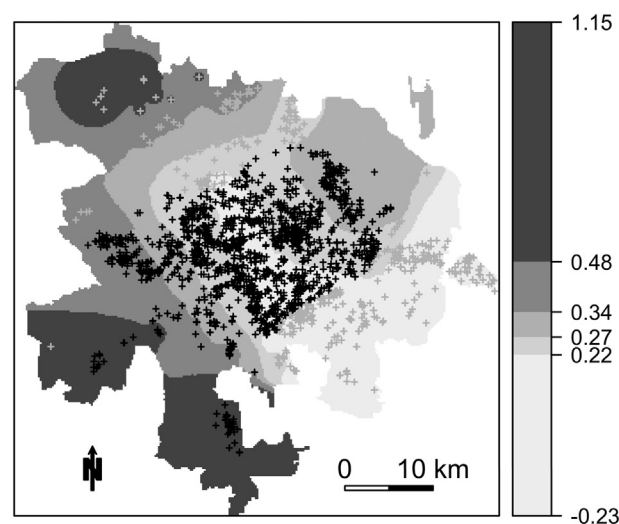
Table 3
Estimation results from the GWOR model.

Summaries on coefficients	Log Income	Pollution	Log distance to city centre	Log distance to railway station	WTP
2.5%	0.073	− 0.1	− 0.249	− 0.367	− 0.169
25%	0.214	− 0.091	0.337	− 0.276	0.2
50%	0.234	− 0.077	0.429	− 0.241	0.312
75%	0.285	− 0.054	0.474	− 0.149	0.417
97.5%	0.74	0.037	0.631	− 0.08	0.457
Standard deviations	0.133	0.034	0.462	0.104	0.186
Monte Carlo non-stationary (p-values)	0.09	0.01	0.03	0.06	

Note: The GWOR model is estimated by using a fixed kernel approach with an optimum bandwidth of about 15.8 km. WTP measures the marginal willingness to pay for air quality improvement, presented as percentages of the average sample income.
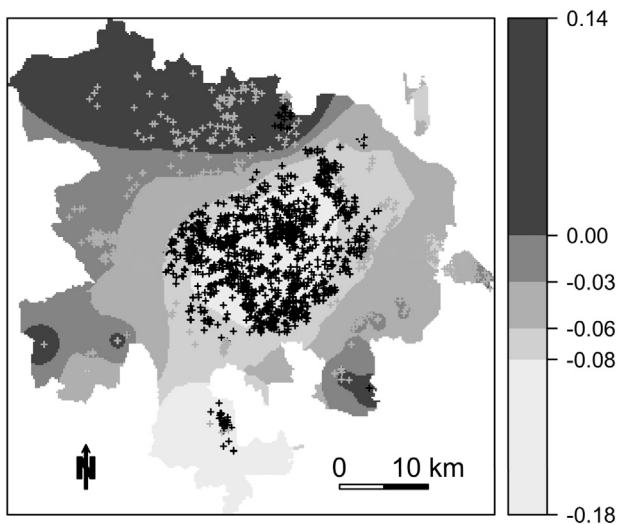

Fig. 2. The interpolated surface on associations between income and life satisfaction. Sample locations are presented by cross symbols: black colour indicating significant coefficients and grey colour insignificant coefficients, depending on whether the absolute t-values of local coefficients are above 1.96 or not.

complexity of ordinal regression models, it is not clear how to form a hat matrix for a GWOR model, a key element for such type of model comparison (Brunsdon et al., 1996; Fotheringham et al., 2003). A Monte Carlo simulation approach, outlined above, was used to test whether the spatial variations in local coefficients are statistically significant, or due to random sampling uncertainty. The non-stationary test results based on 99 permutations are also reported in the last row of Table 3. Overall, the non-stationary test results suggest that the spatial heterogeneity in the associations of income, air pollution and locational factors to life satisfaction is unlikely due to random sampling uncertainties.

Looking at the distribution of the local coefficients of income (Log Income), there is relatively large variability, shown by the difference between the 2.5-th and 97.5-th percentiles of the distribution and the large standard deviation (Table 3). Fig. 2 depicts the interpolated surface of the associations between income and life satisfaction with breaking points being the quintiles of the local coefficients. Approximate significance inferences of local coefficients are also presented in the map. Residential locations are presented by cross symbols and the statistical significances of local coefficients are indicated by different colours: black for significant coefficients and grey for insignificant coefficients, depending on whether the absolute t-values of local coefficients are above 1.96 or not. From Fig. 2, we see that the income-life satisfaction associations are increasing from the central urban area of Beijing to the suburban areas of Beijing. The local coefficients of air pollution also exhibit large spatial variability, indicated by a roughly 41% increase in the local coefficients from the lower quantile to the upper quantile of the distribution. The Monte Carlo non-stationary test for air pollution yields a *p*-value of 0.01, producing strong evidence on the spatial variation in how air pollution is related to life satisfaction. Fig. 3 shows an interesting spatial pattern of the local pollution coefficients. Overall, the central urban areas of Beijing present a stronger negative relationship between air pollution and life satisfaction than the suburban areas do, indicating a stronger preference for air quality for residents living in central urban areas. For most of our sampling locations, proximity to the city centre is negatively associated with life satisfaction while public transport accessibility tends to increase life satisfaction (Table 3).

Using the estimated local coefficients of air pollution and income, local marginal WTP was estimated and presented as the percentages of average income in the last column of Table 3. It is useful to note that the marginal WTP is estimated as 42% and 20% of the average monthly

Fig. 3. The interpolated surface on associations between pollution and life satisfaction.
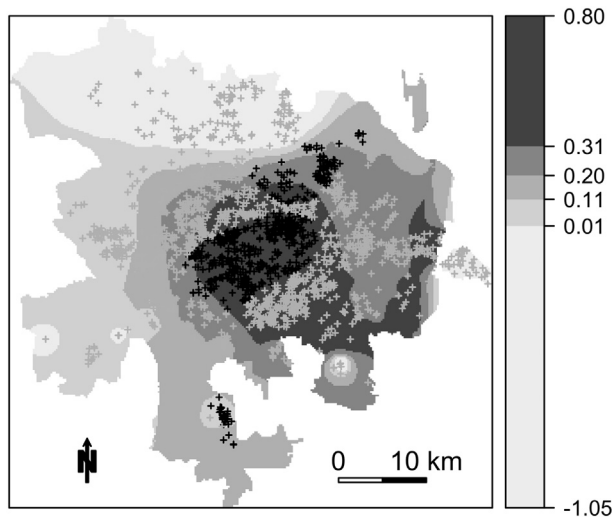

Fig. 4. The interpolated surface on estimated WTP for air quality improvement, presented as percentages of the average monthly income.

income at the upper and lower quantiles of its distribution in the GWOR model. This is in contrast with the estimated WTP as 17.1% of the average income in the global ordinal regression model (Table 2). This suggests that people living in different locations are willing to pay for air pollution abatement differently. The median marginal WTP for per unit air pollution abatement was about 31.2% of the average monthly income (or 2.6% of the average annual income). This represents a small share of income that residents are willing to pay for improvement of air quality, comparing to the usual WTP estimates reported in the European environmental valuation studies (e.g. Ferreira et al., 2013). The spatial variation of WTP is presented in Fig. 4 with breaking points being the quintiles of the distribution. A noticeable feature is the decrease in WTP when moving away from the central urban area to the more suburban areas. A plausible explanation is that residents living in the central and southern areas of Beijing are on average more exposed to air pollution than whose living in the northern areas of Beijing (Zhang et al., 2013), and thus a higher WTP for air pollution abatement.

Finally, we calculate the monetary welfare effect of discrete changes (rather than the per unit or marginal change) in air pollution. Compensating Surplus (CS) is a popular measure for policy analysis in the environmental economics literature (e.g. Welsch & Ferreira, 2014). Following Ferreira and Moro (2010), CS is calculated as:
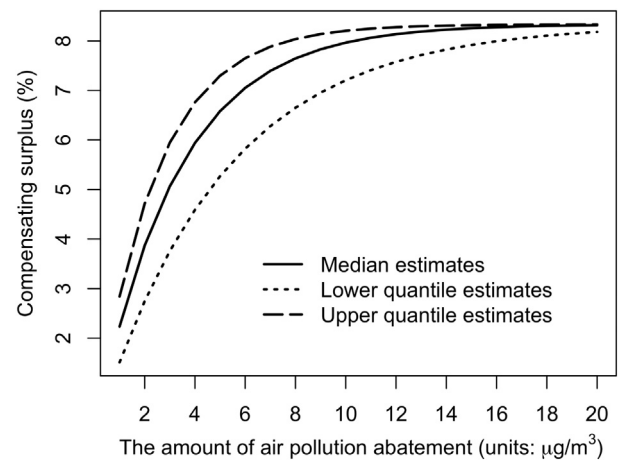

Fig. 5. The compensating surplus for discrete amounts of air pollution abatement.

$\overline{Income}\,[1 - exp(\hat{\beta}/\hat{\gamma} \times \Delta Pollution)]$, in which $\overline{Income}$ is the sample mean income and $\Delta Pollution$ measures the discrete amount of air pollution abatement. In short, CS measures the amount of money that would keep the (latent) life satisfaction of an individual at the original level when a change of air pollution has occurred. We use the lower, median and upper quantiles of the estimated marginal WTP from the GWOR model (Table 3) to calculate the CS for discrete amounts of air pollution abatement. Fig. 5 shows the calculated CS, presented as percentages of the average annual income. The CS curves show an interesting feature: they increase rapidly first, and then tend to be stable. At the currently high levels of air pollution in Beijing, the amount of money people are willing to scarce for air pollution abatement increases quickly, reflecting a great desire for a drop of air pollution. However, after a decrease of about $15\,\mu g/m^3$ in the air pollution concentration, the amount of money people would like to give up stays stable, roughly equal to 8.3% of the average annual income. In other words, further reduction of air pollution levels won't contribute much to the increase in individual life satisfaction. The issue facing Beijing and Chinese cities in general, however, is that a moderate reduction of air pollution levels is far from reaching the air quality standard deemed to be not harmful to human health (Zhang et al., 2013).

6. Conclusion

In this paper, we extend the GWR modelling framework to accommodate ordinal categorical responses. Ordinal response variables are increasingly explored in a wide range of social science disciplines, although in a way that the spatiality of the data under study is not often taken into account. The GWOR model offers a flexible exploratory tool to explore the spatial aspects of data and address the issue of spatial heterogeneity. Estimation methods and statistical tests on spatial heterogeneity in covariate effects for GWOR models are elaborated in the study. The usefulness of the model is demonstrated by exploring the socio-spatial variations of resident's life satisfaction in urban Beijing and the spatially varying relationships between life satisfaction and income, air pollution and other locational factors.

Drawing on a recent large-scale survey data in Beijing, the study assessed potential spatial variability in the WTP for air quality. The median marginal WTP for per unit air pollution abatement was estimated as about 31.2% of the average monthly income (or 2.6% of the average annual income) under the GWOR model. This demonstrates a low willing to pay for marginal improvement of air quality. Although having experienced rapid economic growth for decades, China remains a developing country and the preference for clean air is not expected to be as strong as that in developed countries, on average. We also identify a ceiling point in the amount of money people are willing to pay for air pollution reduction, beyond which people become reluctant to sacrifice

more money for further air pollution abatement. This might reflect the currently weak preference for air quality of residents in Beijing and likely in other Chinese cities.

Some limitations remain. First, this study has not devised an appropriate approach for model comparisons due to the difficulty of applying the usual analysis of variance approach to models with ordinal response variables. Another issue is in relation to our tentative development of the mixed GWOR model. The development of an iterative estimation algorithm for the mixed GWOR model is our next step. Despite these limitations, GWOR offers a suitable tool for local spatial analysis and modelling of ordinal categorical responses.

Acknowledgements

Funding

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compenvurbsys.2018.01.012.

References

Agresti, A. (2010). *Analysis of ordinal categorical data.* John Wiley & Sons.
Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighbourhoods. *Journal of Political Economy, 115*, 588–638.
Bian, Y., Zhang, L., Yang, J., Guo, X., & Lei, M. (2015). Subjective wellbeing of chinese people: A multifaceted view. *Social Indicators Research, 121*, 75–92.
Brown, Z. S., Oueslati, W., & Silva, J. (2016). Links between urban structure and life satisfaction in a cross-section of OECD metro areas. *Ecological Economics, 129*, 112–121.
Brunsdon, C., & Charlton, M. E. (2011). An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection. *Environment and Planning. B, Planning & Design, 38*, 216–230.
Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial non-stationarity. *Geographical Analysis, 28*, 281–298.
Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science, 39*, 497–524.
Burridge, J. (1981). A note on maximum likelihood estimation for regression models using grouped data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 43*, 41–45.
Chen, J., Chen, S., & Landry, P. F. (2013). Migration, environmental hazards, and health outcomes in China. *Social Science & Medicine, 80*, 85–95.
da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis, 48*, 233–247.
Dong, G., Ma, J., Harris, R., & Pryce, G. (2016). Spatial random slope multilevel modelling using multivariate conditional autoregressive models: A case study of subjective travel satisfaction in Beijing. *Annals of the American Association of Geographers,*

106, 19–35.
Ferreira, S., Akay, A., Brereton, F., Cuñado, J., Martinsson, P., Moro, M., & Ningal, T. F. (2013). Life satisfaction and air quality in Europe. *Ecological Economics, 88*, 1–10.
Ferreira, S., & Moro, M. (2010). On the use of subjective well-being data for environmental valuation. *Environmental and Resource Economics, 46*(3), 249–273.
Ferreira, S., & Moro, M. (2013). Income and preferences for the environment: Evidence from subjective well-being data. *Environment and Planning A, 45*, 650–667.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. E. (2003). *Geographically weighted regression: The analysis of spatially varying relationships.* John Wiley & Sons.
Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis, 47*, 431–452.
Fotheringham, A. S., & Oshan, T. M. (2016). Geographically weighted regression and multicollinearity: Dispelling the myth. *Journal of Geographical Systems, 18*, 303–329.
Franses, P. H., & Paap, R. (2001). *Quantitative models in marketing research.* Cambridge University Press.
Greene, W. H. (2003). *Econometric analysis.* Pearson Education India.
Greene, W. H., & Hensher, D. A. (2010). *Modelling ordered choices: A primer.* Cambridge University Press.
Harris, R., Dong, G., & Zhang, W. (2013). Using contextualized geographically weighted regression to model the spatial heterogeneity of land prices in Beijing, China. *Transactions in GIS, 17*, 901–919.
Henningsen, A., & Toomet, O. (2011). Maxlik: A package for maximum likelihood estimation in r. *Computational Statistics, 26*, 443–458.
Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science, 24*, 383–401.
Leung, Y., Mei, C.-L., & Zhang, W.-X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A, 32*, 9–32.
Liu, Z., Wang, Y., & Chen, S. (2017). Does formal housing encourage settlement intention of rural migrants in Chinese cities? A structural equation model analysis. *Urban Studies, 54*, 1834–1850.
Loader, C. (2006). *Local regression and likelihood.* Springer Science & Business Media.
Lu, B., Charlton, M. E., Brunsdon, C., & Harris, P. (2016). The Minkowski approach for choosing the distance metric in geographically weighted regression. *International Journal of Geographical Information Science, 30*, 351–368.
Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014). The gwmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science, 17*, 85–101.
Ma, J., Mitchell, G., Dong, G., & Zhang, W. (2017). Inequality in Beijing: A spatial multilevel analysis of perceived environmental hazard and self-rated health. *Annals of the American Association of Geographers, 107*, 109–129.
MacKerron, G., & Mourato, S. (2009). Life satisfaction and air quality in London. *Ecological Economics, 68*, 1441–1453.
McMillen, D. P., & McDonald, J. F. (2004). Locally weighted maximum likelihood estimation: Monte Carlo evidence and an application. *Advances in spatial econometrics* (pp. 225–239). Springer.
Mei, C. L., Xu, M., & Wang, N. (2016). A bootstrap test for constant coefficients in geographically weighted regression models. *International Journal of Geographical Information Science, 30*, 1622–1643.
Nakaya, T., Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2005). Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine, 24*, 2695–2717.
Páez, A., Uchida, T., & Miyamoto, K. (2002). A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A, 34*, 733–754.
Pratt, J. W. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association, 76*, 103–106.
Welsch, H. (2006). Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics, 58*, 801–813.
Welsch, H., & Ferreira, S. (2014). Environment, well-being, and experienced preference. *International Review of Environmental and Resource Economics, 7*, 205–239.
Wheeler, D. C. (2007). Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A, 39*, 2464–2481.
Wheeler, D. C., & Waller, L. A. (2009). Comparing spatially varying coefficient models: A case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. *Journal of Geographical Systems, 11*, 1–22.
Zhang, A., Qi, Q., Jiang, L., Zhou, F., & Wang, J. (2013). Population exposure to pm 2.5 in the urban area of Beijing. *PLoS One.* http://dx.doi.org/10.1371/journal.pone.0063486.
Zhang, W., Yu, J., Li, Y., & Dang, Y. (2015). *Urban settlement and spatial behaviour of residents.* Beijing Science Press.