

Geographically weighted Poisson regression for disease association mapping

T. Nakaya^{1,*}, A. S. Fotheringham², C. Brunson³ and M. Charlton²

¹*Department of Geography, Ritsumeikan University, Japan*

²*National Centre for Geocomputation, National University of Ireland, Maynooth, Ireland*

³*School of Computing, University of Glamorgan, U.K.*

SUMMARY

This paper describes geographically weighted Poisson regression (GWPR) and its semi-parametric variant as a new statistical tool for analysing disease maps arising from spatially non-stationary processes. The method is a type of conditional kernel regression which uses a spatial weighting function to estimate spatial variations in Poisson regression parameters. It enables us to draw surfaces of local parameter estimates which depict spatial variations in the relationships between disease rates and socio-economic characteristics. The method therefore can be used to test the general assumption made, often without question, in the global modelling of spatial data that the processes being modelled are stationary over space. Equally, it can be used to identify parts of the study region in which ‘interesting’ relationships might be occurring and where further investigation might be warranted. Such exceptions can easily be missed in traditional global modelling and therefore GWPR provides disease analysts with an important new set of statistical tools. We demonstrate the GWPR approach applied to a dataset of working-age deaths in the Tokyo metropolitan area, Japan. The results indicate that there are significant spatial variations (that is, variation beyond that expected from random sampling) in the relationships between working-age mortality and occupational segregation and between working-age mortality and unemployment throughout the Tokyo metropolitan area and that, consequently, the application of traditional ‘global’ models would yield misleading results. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Poisson regression; geographically weighted regression; kernel mapping; spatial analysis; Tokyo

*Correspondence to: Tomoki Nakaya, Department of Geography, Ritsumeikan University, 56-1 Tojiin-kita-machi, Kita-ku, Kyoto 603-8577, Japan.

† E-mail: nakaya@lt.ritsumei.ac.jp

Contract/grant sponsor: Grant-in-Aid; contract/grant number: 15700549

Contract/grant sponsor: Science Foundation Research; contract/grant number: 03/RP1/I382

1. INTRODUCTION

Ecological analysis of disease by regression modelling has been repeatedly employed in epidemiology and health geography for trying to uncover plausible explanations for persistent geographical variations in disease incidence rates [1–5]. In such ecological analyses, conventional ‘global’ models postulating a geographically invariant spatial process have typically been used for relating local variations in incidence rates to global association rules. However, ecological associations may also vary geographically, because the meaning of covariates may change depending on different geographical contexts [6, 7]. As an example, consider a simple non-medical situation in which car ownership rates are related to a series of possible explanatory variables. Rates of car-ownership in a region generally correlate positively with average income or capital stocks, though the strength of the relationship might not be spatially constant. For example, the relationship might be relatively weak in rural areas where most people need a car for daily living, or in city centres where public transportation is well developed and having a car can be an inconvenience in terms of parking. Although this problem might be solved by introducing interaction terms of contextual covariates, it is unlikely that we can identify and utilize all contextual factors as operational variables in ecological studies. Alternatively, some relationships might be intrinsically different across space. For instance, people’s preferences for housing styles might vary over space due to factors which are enormously complex and impossible to identify in the form of single explanatory variables. In either case, we need to take into account geographical variations in the ecological associations in our method of analysis [7, 8]. It is worth mentioning that the levels of ecological association between health and other conditions have frequently been found to be different across nations or across regions in health studies [9]. It seems reasonable therefore to allow the possibility that spatial variations in relationships might vary *within* regions.

To deal with this problem, we propose a new statistical tool, that of geographically weighted Poisson regression (GWPR), for analysing disease patterns resulting from spatially non-stationary processes. We also demonstrate the proposed methodology by presenting an application based on a regional working-age mortality dataset for the Tokyo metropolitan area. To examine regional variations in association, multi-level modelling is a popular method with which we can assess variations in parameters between groups of observation units, typically areal units [10–12]. A disadvantage of this method, however, is that the groups of observation units need to be defined before the analysis but it is often difficult to construct such predefined categories of samples, particularly for areal units where administrative boundaries might have little meaning in terms of the processes being examined [13, 14]. Moreover, multilevel models aim to capture variances of parameter distributions rather than revealing their geographical patterns. This characteristic of multilevel modelling becomes apparent in its alternative name of *random coefficient modelling*. The alternative approach of geographically weighted regression (GWR), a non-parametric methodology for the investigation of geographical drifts of regression parameters [8, 15], provides an exploratory methodology of spatial associations. By introducing the concept of spatial kernels, the approach enables us to map geographical variations in regression parameters without predefined areal categorisation (such as in the case of multilevel modelling) and without predefined parameterisation of parameter drifts (such as in the case of the expansion method [16, 17]).

To date, GWR has been developed assuming a Gaussian modelling framework. Fotheringham *et al.* [18] applied Gaussian GWR to a long-term illness data set in the northeast

of England and found complicated regional variations in regression parameters. However, Poisson regression provides a more appropriate basis for the ecological analysis of areal disease datasets than conventional Gaussian regression, especially in the case where observed counts include low numbers [19, 20]. It is well-known that if we can regard each death count as the sum of a series of binomial independent trials with a low mortality rate, the counts are expected to approximately follow a Poisson distribution. Thus, numerous models assuming a Poisson distribution have been proposed as disease analysis tools. These include Poisson probability mapping [21], Poisson regression [14, 19] and multiple-level Poisson regression [5, 22].

An important theoretical advancement proposed here is the introduction of a *semi-parametric* model specification [23, 24] to GWPR modelling so that we can combine geographically varying and geographically constant parameters in a model. The semi-parametric model specification enhances the practical usefulness of non-parametric regression models [23] reflecting *a priori* knowledge of the fixed nature of specific parameters in the model. It is also suggested that in some situations constant terms contribute to model simplicity and avoid local collinearity between covariates so that the interpretation of the estimated geographically varying parameters becomes easier.

The remainder of this paper is organized as follows. In Sections 2 and 3, we describe the framework of GWPR and its calibration procedure, respectively. The semi-parametric variant of GWPR is developed in Section 4. An empirical application of GWPR to the working-age death in the Tokyo metropolitan area in 1990 is described in Section 5 and conclusions are drawn in Section 6.

2. SPECIFICATION OF GEOGRAPHICALLY WEIGHTED POISSON REGRESSION

2.1. SMR and Poisson regression

Before modelling disease incidence rates, some definitions are necessary. The most common indicator used for comparing regional death rates is the standardized mortality ratio (SMR) in place i which is calculated by

$$\text{SMR}_i = O_i / E_i \quad (1)$$

where O_i is observed number of deaths in place i and E_i is expected number of deaths based on reference mortality rates applied to regional demographic structure. E_i is defined as

$$E_i = \sum_g r_g \times p_{g,i} \quad (2)$$

where r_g is a standard mortality rate (e.g. national mortality rate) of demographic group g and $p_{g,i}$ is regional population size specific to demographic group g in place i . The demographic group g is usually determined by age or sex-age attributes. From equation (1), SMR_i is clearly a type of incidence rate index. Our following arguments can therefore be applied to general mortality or morbidity rates defined by ratios of observed counts to population at risk.

Consider a conventional Poisson regression model written as

$$O_i \sim \text{Poisson} \left[E_i \exp \left(\sum_k \beta_k x_{k,i} \right) \right] \quad (3)$$

where, $x_{k,i}$ is k th explanatory variable in place i and the β_k s are parameters and Poisson $[\lambda]$ indicates a Poisson distribution with mean λ . The intercept term β_0 is included by defining $x_{0,i} = 1(\forall i)$.

By rewriting equation (1) as

$$O_i = E_i \cdot \text{SMR}_i \quad (4)$$

we can see that the exponential term with linear predictors in the Poisson regression (equation (3)) corresponds to the modelled regional variation of the standardized mortality ratio. It is worthwhile noting that E_i is given exogenously and is often called an *offset variable* in general linear modelling [25].

The parameters of the Poisson regression are estimated by maximum likelihood. Letting l_i be the log-likelihood of sample (place) i , we can define the estimation problem as:

$$\begin{aligned} \max_{\boldsymbol{\beta}} L &= \sum_i l_i(\boldsymbol{\beta}|O_i) \\ &= \sum_i (-\hat{O}_i(\boldsymbol{\beta}) + O_i \log \hat{O}_i(\boldsymbol{\beta})) \end{aligned} \quad (5)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots)$ is the vector of the regression parameters. The conditional equation of the k th estimator is obtained from setting the first derivative of the likelihood function with respect to the parameter β_k to zero:

$$\frac{\partial L}{\partial \beta_k} = 0 \quad \forall k \quad (6)$$

Generally we need to employ an iterative process to find the solution of this set of equations.

2.2. Geographically varying parameters in Poisson regression

We now introduce the notion of allowing parameter values to vary with geographical location $\mathbf{u}_i = (u_{xi}, u_{yi})$ which is a vector of two dimensional co-ordinates describing the location of i (usually, x - y co-ordinates from a map). The Poisson model in equation (3) can then be rewritten as

$$O_i \sim \text{Poisson} \left[E_i \exp \left(\sum_k \beta_k(\mathbf{u}_i) x_{k,i} \right) \right] \quad (7)$$

This kind of model can generally be named a geographically varying coefficient Poisson regression model after Hastie and Tibshirani [26]. Following the technique of GWR, we can calibrate this model by a kernel regression methodology in which we estimate smoothed geographical variations of parameters with a spatial weighting kernel. In this methodology, parameter estimates are calibrated in a point-wise way. We then refer to this geographically varying coefficient Poisson regression model as a geographically weighted Poisson regression (GWPR).

To estimate GWPR parameters, we employ a variant of the local likelihood principle [27] which we refer to as a geographically weighted likelihood principle [28] and which is consistent with geographically weighted least squares for conventional Gaussian GWR

[8, pp. 90–91]. Following the local likelihood methodology, parameters at location i are estimated by solving the following maximisation problem of the geographically weighted log-likelihood.

$$\max L(\mathbf{u}_i) = \sum_j^N (-\hat{O}_j(\boldsymbol{\beta}(\mathbf{u}_i)) + O_j \log \hat{O}_j(\boldsymbol{\beta}(\mathbf{u}_i))) \cdot w_{ij}(\|\mathbf{u}_i - \mathbf{u}_j\|) \quad (8)$$

where $\hat{O}_j(\boldsymbol{\beta}(\mathbf{u}_i))$ is the predicted number of deaths at location j with parameters at regression point i ,

$$\hat{O}_j(\boldsymbol{\beta}(\mathbf{u}_i)) = E_j \exp\left(\sum_k \hat{\beta}_k(\mathbf{u}_i) x_{k,j}\right) \quad (9)$$

and w_{ij} is the geographical weight of the j th observation at the i th regression point. The weights of the observations gradually decrease as the distance between the regression point i and location at which the j th observation is recorded becomes larger. A classical choice of the weighting kernel is the Gaussian kernel [15] defined by

$$w_{ij} = \exp\left(-\frac{1}{2} \frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{G}\right) \quad (10)$$

where the parameter G (called the *bandwidth*) regulates the kernel size. It is useful to stress the spatial nature of the weighting scheme here. Data that are recorded at locations in close proximity to the regression point i are given higher weighting than data recorded at points further away. As the focal point of the regression moves to a different location, all the weights change. The bandwidth controls the rate at which the weight of a datum declines as the distance of the location at which it is recorded increases from the regression point. When the bandwidth is large, the weights decay slowly; when the bandwidth is small, the weights decay rapidly.

As an alternative to the weighting scheme in equation (10), we can use an adaptive approach where each kernel includes the same number of sample points (although with different weights). A popular adaptive kernel, the bi-square kernel, is defined by

$$w_{ij} = \begin{cases} [1 - (\|\mathbf{u}_i - \mathbf{u}_j\|/G_i)]^2 & \text{if } \|\mathbf{u}_i - \mathbf{u}_j\| < G_i \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where G_i is the distance to the M th nearest observation from regression point i and where the exogenously given number M controls the bandwidth size. Adaptive kernels are useful in the case that there is a large variation in the geographical density of the observed data. See Reference [8, pp. 56–59] for more detailed arguments concerning the choice of geographical weighting functions.

2.3. A note on the difference between GWPR and conventional local regression

It is important to note that what is being proposed is not local modelling in attribute space as is often indicated by the term ‘local regression’ [27] including GAM [24] and LOESS [29]. Here, the term ‘local’ refers to a spatial extent around a regression location. Consequently, what is being investigated is the extent to which processes might vary over space and hence

violate the assumption of spatial non-stationarity embedded in the use of traditional, global models of spatial processes.

Consider a simple Poisson regression model with one explanatory variable:

$$O_i \sim \text{Poisson}[E_i \exp(\beta_0 + \beta_1 x_i)] \quad (12)$$

Conventional local regression models estimate the form of regression functions non-parametrically.

$$O_i \sim \text{Poisson}[E_i \exp(f(x_i))] \quad (13)$$

Here, f is an unspecified function of x_i rather than a numerical coefficient. Note that the location in space is ignored in the two models (equations (12) and (13)).

GWPR estimates the geographical functional form of regression coefficients non-parametrically.

$$O_i \sim \text{Poisson}[E_i \exp(\beta_0(\mathbf{u}_i) + \beta_1(\mathbf{u}_i)x_i)] \quad (14)$$

where β_0 and β_1 are unspecified bi-variate functions of $\mathbf{u}_i (= (u_{xi}, u_{yi}))$. Although we now move back to a model with a linear predictor, we gain interaction between geographical location and functional relationships in the linear predictor: the model has potentially different coefficients for each location.

In conventional kernel regression modelling [27], one attempts to estimate a regression function f by approximating it by polynomials centred on specific values of x_i and maximising a weighted likelihood function, with the weights being a kernel function centred on x_i . In GWPR we estimate coefficients by calibrating a Poisson regression model whose likelihood is geographically weighted, with the weights being a kernel function centred on \mathbf{u}_i . The key difference is that in GWR the kernel is in geographical space and the regression model is in predictor-variable space, whereas in a standard kernel regression model both the regression model and the kernel are defined in predictor-variable space.

2.4. Kernel mapping and GWPR

Kernel mapping can be considered as a special case of GWPR. Consider the GWPR model without any explanatory variables and containing just a local intercept. Then,

$$O_i \sim \text{Poisson}[E_i \exp(\beta_0(\mathbf{u}_i))] \quad (15)$$

By solving

$$\begin{aligned} 0 &= \frac{\partial L(\mathbf{u}_i)}{\partial \hat{\beta}_0(\mathbf{u}_i)} \\ &= \sum_j^n (-E_j \exp(\hat{\beta}_0(\mathbf{u}_i)) + O_j) \cdot w_{ij}(\|\mathbf{u}_i - \mathbf{u}_j\|) \end{aligned} \quad (16)$$

we obtain the estimates of local mortality ratio with geographical weights as

$$\exp(\hat{\beta}_0(\mathbf{u}_i)) = \frac{\sum_j^N w_{ij} O_j}{\sum_j^N w_{ij} E_j} \quad (17)$$

In this case, local estimates of mortality ratio become a geographically weighted mortality ratio that corresponds to the ratio of the kernel density of observed deaths to the kernel density of expected deaths. The quantity is essentially the same as the relative risk surface for areal-based health data advocated by Bithell [30].

3. CALIBRATION PROCEDURE OF GWPR

3.1. Local scoring procedure for GWPR

The maximization problem of equation (8) can be solved by a modified local Fisher scoring procedure, a form of iteratively reweighted least squares (IRLS) [24]. The iterative procedure is necessary except for the special case of kernel mapping as shown in Section 2.4. In this local scoring procedure, the following matrix computation of weighted least squares should be repeated to update parameter estimates until they converge:

$$\boldsymbol{\beta}^{(l+1)}(\mathbf{u}_i) = (\mathbf{X}^t \mathbf{W}(\mathbf{u}_i) \mathbf{A}(\mathbf{u}_i)^{(l)} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(\mathbf{u}_i) \mathbf{A}(\mathbf{u}_i)^{(l)} \mathbf{z}(\mathbf{u}_i)^{(l)} \quad (18)$$

$\boldsymbol{\beta}^{(l+1)}(\mathbf{u}_i)$ is a vector of local parameter estimates specific to location i and superscript $(l+1)$ indicates the number of iterations.

$$\boldsymbol{\beta}^{(l)}(\mathbf{u}_i) = (\beta_0^{(l)}(\mathbf{u}_i), \beta_1^{(l)}(\mathbf{u}_i), \dots, \beta_k^{(l)}(\mathbf{u}_i))^t \quad (19)$$

\mathbf{X} is a design matrix and $(\mathbf{X})^t$ denotes the transpose of \mathbf{X} .

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{K,1} \\ 1 & x_{1,2} & \cdots & x_{K,2} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1,N} & \cdots & x_{K,N} \end{pmatrix} \quad (20)$$

$\mathbf{W}(\mathbf{u}_i)$ denotes the diagonal spatial weights matrix for location i :

$$\mathbf{W}(\mathbf{u}_i) = \begin{pmatrix} w_{i1} & & & 0 \\ & w_{i2} & & \\ & & \ddots & \\ 0 & & & w_{iN} \end{pmatrix} \quad (21)$$

and $\mathbf{A}(\mathbf{u}_i)^{(l)}$ denotes the variance weights matrix associated with the Fisher scoring for each location i

$$\mathbf{A}(\mathbf{u}_i)^{(l)} = \begin{pmatrix} \hat{O}_1(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) & & & 0 \\ & \hat{O}_2(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) & & \\ & & \ddots & \\ 0 & & & \hat{O}_N(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) \end{pmatrix} \quad (22)$$

Finally, $\mathbf{z}^{(l)}(\mathbf{u}_i)$ is a vector of adjusted dependent variables defined as

$$\mathbf{z}^{(l)}(\mathbf{u}_i) = (z_1^{(l)}(\mathbf{u}_i), z_2^{(l)}(\mathbf{u}_i), \dots, z_N^{(l)}(\mathbf{u}_i))^t \tag{23}$$

where

$$\begin{aligned} z_j^{(l)}(\mathbf{u}_i) &= \left(\beta_0^{(l)}(\mathbf{u}_i) + \sum_k^K \beta_k^{(l)}(\mathbf{u}_i)x_{k,j} \right) + \frac{O_j - \hat{O}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))}{\hat{O}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))} \\ &= \eta_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i)) + \frac{O_j - \hat{O}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))}{\hat{O}_j(\boldsymbol{\beta}^{(l)}(\mathbf{u}_i))} \end{aligned} \tag{24}$$

where η_j is a linear predictor of j th observation. By repeating the iterative procedure for every regression point i , sets of local parameter estimates are obtained.

At convergence, we can omit the subscripts (l) or ($l + 1$) and then rewrite equation (18) as

$$\boldsymbol{\beta}(\mathbf{u}_i) = (\mathbf{X}^t \mathbf{W}(\mathbf{u}_i) \mathbf{A}(\mathbf{u}_i) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(\mathbf{u}_i) \mathbf{A}(\mathbf{u}_i) \mathbf{z}(\mathbf{u}_i) \tag{25}$$

Since the linear predictor for each observation j is estimated for each regression point i , the mapping from adjusted dependent variables to linear predictors is as follows:

$$\begin{pmatrix} \hat{\eta}_1(\mathbf{u}_1) & \hat{\eta}_1(\mathbf{u}_2) & \cdots & \hat{\eta}_1(\mathbf{u}_N) \\ \hat{\eta}_2(\mathbf{u}_1) & \hat{\eta}_2(\mathbf{u}_2) & \cdots & \hat{\eta}_2(\mathbf{u}_N) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\eta}_N(\mathbf{u}_1) & \hat{\eta}_N(\mathbf{u}_2) & \cdots & \hat{\eta}_N(\mathbf{u}_N) \end{pmatrix} = \mathbf{R} \begin{pmatrix} z_1(\mathbf{u}_1) & z_1(\mathbf{u}_2) & \cdots & z_1(\mathbf{u}_N) \\ z_2(\mathbf{u}_1) & z_2(\mathbf{u}_2) & \cdots & z_2(\mathbf{u}_N) \\ \vdots & \vdots & \cdots & \vdots \\ z_N(\mathbf{u}_1) & z_N(\mathbf{u}_2) & \cdots & z_N(\mathbf{u}_N) \end{pmatrix} \tag{26}$$

where the i th row of the matrix \mathbf{R} , \mathbf{r}_i , is

$$\mathbf{r}_i = \mathbf{x}_i (\mathbf{X}^t \mathbf{W}(\mathbf{u}_i) \mathbf{A}(\mathbf{u}_i) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(\mathbf{u}_i) \mathbf{A}(\mathbf{u}_i) \tag{27}$$

As shown in equation (7), only the diagonal elements (i.e. $i = j$) of the matrix on the left-hand side of equation (26) are used for predicting the numbers of deaths at data locations. By focusing on these elements, the mapping from adjusted dependent variable to linear predictor is defined as

$$\hat{\boldsymbol{\eta}} = \mathbf{S} \mathbf{z} \tag{28}$$

where

$$\hat{\boldsymbol{\eta}} = (\hat{\eta}_1(\mathbf{u}_1), \hat{\eta}_2(\mathbf{u}_2), \dots, \hat{\eta}_N(\mathbf{u}_N))^t \quad \text{and} \quad \mathbf{z} = (z_1(\mathbf{u}_1), z_2(\mathbf{u}_2), \dots, z_N(\mathbf{u}_N))^t \tag{29}$$

The ij th element of the matrix \mathbf{S} , S_{ij} , is

$$S_{ij} = R_{ij} \frac{z_i(\mathbf{u}_j)}{z_j(\mathbf{u}_j)} \tag{30}$$

where R_{ij} is the ij th element of the matrix \mathbf{R} . This matrix \mathbf{S} corresponds with the so-called hat matrix for Gaussian local regression modelling [27].

3.2. Standard error of parameters in GWPR

At convergence for each regression point i , the vector of local parameter estimates is

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\mathbf{u}_i) &= \mathbf{C}(\mathbf{u}_i)\mathbf{z}(\mathbf{u}_i) \\ &= (\mathbf{X}^t\mathbf{W}(\mathbf{u}_i)\mathbf{A}(\mathbf{u}_i)\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}(\mathbf{u}_i)\mathbf{A}(\mathbf{u}_i)\mathbf{z}(\mathbf{u}_i)\end{aligned}\quad (31)$$

where $\mathbf{A}(\mathbf{u}_i)$ and $\mathbf{z}(\mathbf{u}_i)$ in the right-hand side of equation (31) are calculated by the converged estimates of parameters $\hat{\boldsymbol{\beta}}(\mathbf{u}_i)$. Since $\mathbf{z}(\mathbf{u}_i)$ approximately follows the normal distribution with zero mean and variance–covariance $\mathbf{A}(\mathbf{u}_i)^{-1}$, the asymptomatic variance–covariance matrix of regression parameters, $\text{cov}(\hat{\boldsymbol{\beta}}(\mathbf{u}_i))$, is given by

$$\text{cov}(\hat{\boldsymbol{\beta}}(\mathbf{u}_i)) = \mathbf{C}(\mathbf{u}_i)\mathbf{A}(\mathbf{u}_i)^{-1}\mathbf{C}(\mathbf{u}_i)^t \quad (32)$$

The standard error of the k th parameter estimate is given by

$$\text{Se}(\beta_k(\mathbf{u}_i)) = \sqrt{\text{Cov}(\hat{\boldsymbol{\beta}}(\mathbf{u}_i))_k} \quad (33)$$

where $\text{Cov}(\hat{\boldsymbol{\beta}}(\mathbf{u}_i))_k$ is the k th diagonal element of the variance–covariance matrix. The local pseudo t statistic for the local version of the k th parameter estimate is then computed by

$$t_k(\mathbf{u}_i) = \beta_k(\mathbf{u}_i)/\text{Se}(\beta_k(\mathbf{u}_i)) \quad (34)$$

This quantity approximately follows the standard normal distribution if the true regression parameter is zero. This can be used for local inspection of parameter significance. Although the usual threshold of p values for a significance test (effectively $|t| > 1.96$ for tests at the 5 per cent level with large samples) suffers from the multiple comparison problem in this approach, mapping the pseudo t surface still can yield useful insights into spatial variations in relationships, as we demonstrate below.

3.3. Degrees of freedom and deviance of GWPR

If we use a very small bandwidth for the weighting kernel, we may find seemingly strong geographical variation in the local parameter estimates but the estimates are not very reliable, being based on relatively few observations. On the other hand, although a large bandwidth provides more reliable local parameter estimates, these estimates may contain an unacceptably large amount of bias, being based as they are on data at increasing distances from the regression point. This trade-off between spatial scale and statistical stability or between bias and variance can be considered by model selection indicators such as AIC [31–33].

The AIC of the model with bandwidth G is defined as

$$\text{AIC}(G) = D(G) + 2K(G) \quad (35)$$

where D and K denote the deviance and the effective number of parameters in the model with bandwidth G , respectively. The model with the smallest AIC should be selected as an optimal model called MAICE (minimum AIC estimator). A common rule-of-thumb in the use of AICc is that if the difference in AICc values between two models is less than or equal to

2, there is no substantive difference in the performance of the two models [32]. In the case of local regression, since the degrees-of-freedom are likely to be small, small sample bias adjustment (second order adjustment) in the AIC definition [34, 35], particularly the use of AICc, is highly recommended [33, 36]. AICc derived by [35] is written as follows:

$$\begin{aligned} \text{AICc}(G) &= D(G) + 2K(G) + 2 \frac{K(G)(K(G) + 1)}{N - K(G) - 1} \\ &= \text{AIC}(G) + 2 \frac{K(G)(K(G) + 1)}{N - K(G) - 1} \end{aligned} \quad (36)$$

If the effective number of parameters, K , is small relative to the number of observations, N , then the difference between AIC and AICc is negligible. The derivation of AICc is based on Gaussian linear models. However, applying AICc to non-Gaussian modelling is generally acceptable unless the underlying probability distribution is extremely non-normal [33].

As in the case of conventional Poisson regression, deviance as defined by McCullagh and Nelder [25], is given by

$$D(G) = \sum_i^N (O_i \log \hat{O}_i(\boldsymbol{\beta}(\mathbf{u}_i), G) / O_i + (O_i - \hat{O}_i(\boldsymbol{\beta}(\mathbf{u}_i), G))) \quad (37)$$

However, unlike the case of global Poisson regression, in GWPR the second term in the right-hand side of equation (37) is not necessarily zero.

Since the trace of the hat matrix corresponds to the number of regression parameters in a general linear model (including a global Poisson model), it is straightforward to define the effective number of parameters of GWPR, K , by the trace of \mathbf{S} :

$$K(G) = \text{trace}(\mathbf{S}(G)) \quad (38)$$

Obviously, the following equation is obtained from equation (30):

$$\text{trace}(\mathbf{S}(G)) = \text{trace}(\mathbf{R}(G)) \quad (39)$$

Therefore, the matrix \mathbf{S} is not necessary for calculation of the effective number of parameters in GWPR.

The information criterion AICc and the idea of MAICE can be used for not only bandwidth selection but also for model comparisons between kernel mapping, global Poisson regression models, GWPR models and regionalisation models [14].

4. SEMI-PARAMETRIC EXTENSION OF GWPR

4.1. Semi-parametric specification of GWPR

In some empirical applications, there might be reasons, either generated from theory or past experience, not to allow certain parameters to vary geographically. Alternatively, there may be no significant geographical variation in some local parameter estimates. In such cases, we

should consider the following variant of the GWPR model which allows some local parameters to vary spatially and some to be held constant.

$$O_i \sim \text{Poisson} \left[E_i \exp \left(\sum_k \beta_k(\mathbf{u}_i) x_{k,i} + \sum_m \gamma_m x_{m,i} \right) \right] \quad (40)$$

where the coefficient with m th variable, γ_m , is not assumed to depend on geographical locations. This kind of model is called a semi-parametric model [23,24] or a mixed model [8].[‡]

4.2. Estimation of semi-parametric GWPR

Semi-parametric GWPR can be considered as a ‘two-smoother’ model [24]: one smoothing mechanism is parametric while the other is non-parametric. Applying the adjusted dependent variable approach, we solve two simultaneous matrix equations iteratively.

$$\boldsymbol{\gamma}^{(l+1)} = (\mathbf{X}_{\text{par}}^t \mathbf{A}^{(l)} \mathbf{X}_{\text{par}})^{-1} \mathbf{X}_{\text{par}}^t \mathbf{A}^{(l)} (\mathbf{z}^{(l)} - \boldsymbol{\eta}^{(l)}) \quad (41)$$

$$\boldsymbol{\eta}^{(l)} = \mathbf{S}^{(l)} (\mathbf{z}^{(l)} - \mathbf{X}_{\text{par}} \boldsymbol{\gamma}^{(l)}) \quad (42)$$

where \mathbf{X}_{par} is the design matrix of explanatory variable whose parameters are assumed not to depend on geographical locations. Equation (42) means that linear predictors of geographically varying parameters are estimated by applying GWPR to the adjusted residuals of parametric predictors.

At convergence, if we use equation (42) to eliminate $\boldsymbol{\eta}$ in equation (41) by ignoring the iteration superscript (l), then we obtain the following equation:

$$\boldsymbol{\gamma} = (\mathbf{X}_{\text{par}}^t \mathbf{A} (\mathbf{I} - \mathbf{S}) \mathbf{X}_{\text{par}})^{-1} \mathbf{X}_{\text{par}}^t \mathbf{A} (\mathbf{I} - \mathbf{S}) \mathbf{z} \quad (43)$$

This equation can be used for fast calculation of the iterative procedure:

$$\boldsymbol{\gamma}^{(l+1)} = (\mathbf{X}_{\text{par}}^t \mathbf{A}^{(l)} (\mathbf{I} - \mathbf{S}^{(l)}) \mathbf{X}_{\text{par}})^{-1} \mathbf{X}_{\text{par}}^t \mathbf{A}^{(l)} (\mathbf{I} - \mathbf{S}^{(l)}) \mathbf{z}^{(l)} \quad (44)$$

The algorithm of the iterative procedure can be summarised as follows.

- (1) Assign the initial values of $\boldsymbol{\gamma}$ (e.g. using a global Poisson regression model).
- (2) Apply GWPR to the adjusted residual to estimate geographically varying parameters and calculate the non-parametric smoother \mathbf{S} (equation (42)).
- (3) Use equation (41) or (44) to update the fixed parameters, $\boldsymbol{\gamma}^{(l+1)}$.
- (4) Iterate step (2) and (3) until convergence.

It should be noted that step (2) needs iteration as explained in Section 3.1 so that this algorithm has two loops: an implicit inner loop of the non-parametric part within step (2) and an outer loop of the parametric part shown as steps (2) and (3).

[‡]The term ‘mixed model’ is used here to denote a model in which some of the parameters are held constant over space and others are allowed to vary spatially. This is in contradistinction to the more established use to denote a model in which there is a stochastic/fixed parameter split.

4.3. Inference and model comparison of semi-parametric GWPR

Standard errors and the effective number of parameters for model comparisons can be defined for semi-parametric GWPR based on the same arguments for GWPR in Section 3. From equation (43), the asymptotic variance–covariance matrix of $\hat{\gamma}$, $\text{cov}(\hat{\gamma})$ is obtained as

$$\text{cov}(\hat{\gamma}) = \mathbf{C}_\gamma \mathbf{A}^{-1} \mathbf{C}_\gamma^t \quad (45)$$

where

$$\mathbf{C}_\gamma = (\mathbf{X}_{\text{par}}^t \mathbf{A} (\mathbf{I} - \mathbf{S}) \mathbf{X}_{\text{par}})^{-1} \mathbf{X}_{\text{par}}^t \mathbf{A} (\mathbf{I} - \mathbf{S}) \quad (46)$$

The standard error of the m th estimates are given by

$$\text{se}(\gamma_m) = \sqrt{\text{Cov}(\hat{\gamma})_m} \quad (47)$$

where $\text{Cov}(\hat{\gamma})_m$ is the m th diagonal element of the asymptotic variance–covariance matrix.

By equations (42) and (43), estimated linear predictors can be written as a product of the smoother matrix with the vector of the adjusted dependent variable:

$$\hat{\boldsymbol{\eta}} + \mathbf{X}_{\text{par}} \hat{\boldsymbol{\gamma}} = \mathbf{T} \mathbf{z} \quad (48)$$

where

$$\mathbf{T} = \mathbf{S} + (\mathbf{I} - \mathbf{S}) \mathbf{X}_{\text{par}} (\mathbf{X}_{\text{par}}^t \mathbf{A} (\mathbf{I} - \mathbf{S}) \mathbf{X}_{\text{par}})^{-1} \mathbf{X}_{\text{par}}^t \mathbf{A} (\mathbf{I} - \mathbf{S}) \mathbf{z} \quad (49)$$

Recalling the argument in Section 3.3, a natural definition of the effective number of parameters for model comparisons, K_s , is

$$K_s = \text{trace}(\mathbf{T}) \quad (50)$$

By using AICc, we can obtain the best (i.e. MAICE) bandwidth of the non-parametric part in semi-parametric GWPR. We can also compare other possible models with AICc. One interesting comparison is between a fully geographically varying parameter model (full GWPR) and a semi-parametric GWPR in which only one parameter (associated with a particular variable) is assumed to be geographically fixed but the remaining parameters are allowed to vary geographically. Through this comparison, we can judge the significance of the spatial variation in each set of local estimates in the full GWPR model. If AICc of the full GWPR model is higher (i.e. worse) than that of the semi-parametric GWPR model, with a particular parameter held constant, there is little evidence to suggest this particular relationship varies over space. However, if the AICc of the full model is lower (i.e. the model fit is better), then there is evidence that this particular relationship varies spatially. This comparison can be done for each relationship in the model in turn.

5. A GWPR ANALYSIS OF WORKING-AGE MORTALITY IN THE TOKYO METROPOLITAN AREA

5.1. The Problem and the Data

It is recognized that within all societies there are persistent spatial variations in health statistics: the worst health records are generally associated with the most deprived areas. The strength of the relationship between ‘health’ and ‘wealth/deprivation’ should be a concern to all governments that espouse ideals of equality. Consequently, there are many studies which try to

quantify aspects of the link between ‘health’ and ‘wealth/deprivation’ [3, 5, 9, 37, 38]. However, the vast majority of studies to date have attempted to link some measurement of health to various socio-economic and environmental characteristics using a ‘global’ model which may disguise some interesting spatial variations in relationships. Here, we use a local model, that of GWPR, to examine the relationship between mortality rates and socio-economic factors across the metropolitan area of Tokyo. We demonstrate how the use of local modelling provides a great deal more information to health analysts than does traditional global modelling.

The study area, that of the Tokyo metropolitan area, is enclosed by an approximate 70 km radius from the centroid of the Chiyoda ward of Tokyo where the Imperial Palace is located. The area consists of 262 municipality zones and approximately corresponds to the functional region of Tokyo City, the core part of the Tokyo metropolitan area. The data to be analysed in this study are the working-age death rates of those aged 25–64 years in 1990. Age–sex specific populations and deaths are taken from the 1990 national census and the 1990 vital statistics in Japan, respectively. Age groups are defined by five-year age intervals. The expected number of deaths for each zone i is calculated by

$$E_i = \sum_{\text{sex}} \sum_{\text{age}} r_{\text{sex, age}} \times p_{\text{sex, age, } i}$$

$$\text{sex} \in \{\text{male, female}\}$$

$$\text{age} \in \{25-29, 30-34, \dots, 60-64\}$$
(51)

where r and p are the national sex–age specific mortality rate and regional population size, respectively.

Nakaya [14, 39] analysed mortality rates of infant and elderly people in the same area. These mortality patterns, particularly those of elderly people, show clear geographical patterns characterised by concentric and sectoral variations that match well aspects of the socio-economic structure of the Tokyo metropolitan area. Figure 1(a) shows the distribution of SMR across the 262 municipality zones. Summary statistics across the 262 zones for observed death counts and standard mortality rates for the 25–64 age group are given in Table I.

5.2. Kernel mapping

In Figure 1(a) the geographical pattern of working-age deaths is not clear, mainly because the number of deaths tends to be small, particularly in peripheral zones. This problem is often called the small-number problem. To overcome this problem, geographical aggregation of areal units is a straightforward way to obtain statistically more reliable estimates of regional parameters [14]. Kernel mapping can be used as an analogy to geographical aggregation here. To determine an appropriate bandwidth size, we compared the values of the AICc statistic associated with kernel mapping at many different bandwidths. Recall that kernel mapping is a special case of a GWPR model which contains only the intercept term (Section 2.4). We tested bandwidth sizes from 3 to 70 km in steps of 1 km with a Gaussian fixed kernel.[§]

[§]A range of bandwidths between 3 and 70 kms was examined. The smallest bandwidth is dictated by the amount of data in each local regression—any smaller bandwidth would not allow reliable local parameters to be estimated. The largest value is an approximation to the radius of the region. Any larger values would create overly smooth parameter surfaces which would hide any local variations.

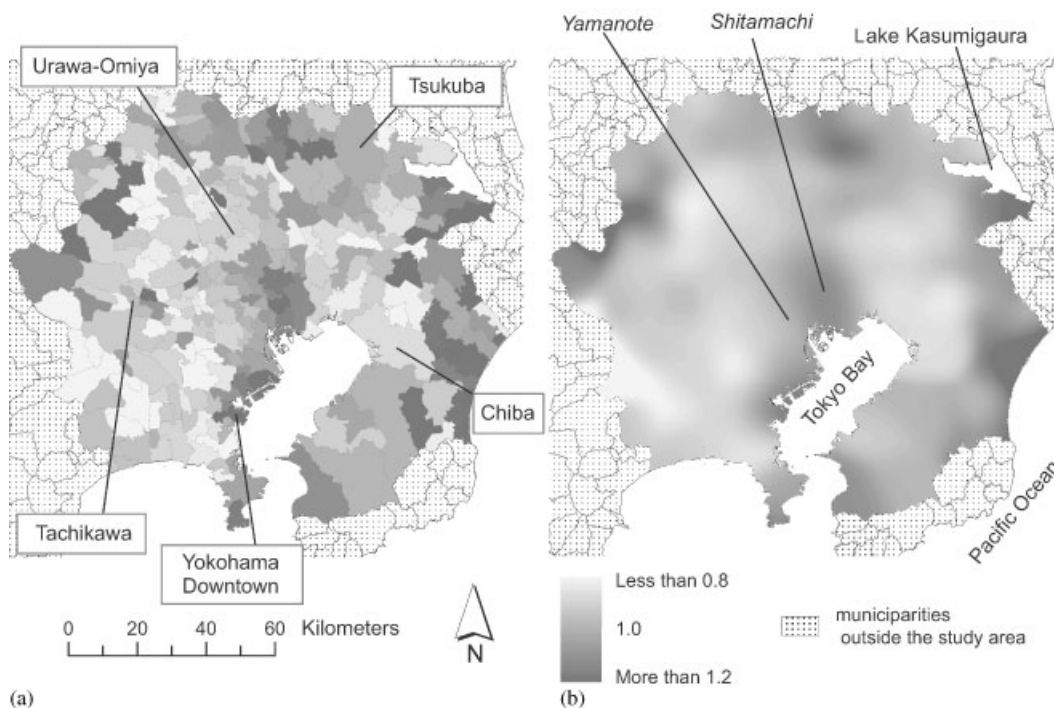


Figure 1. Geographical distribution of working-age mortality in the Tokyo metropolitan area as of 1990; (a) SMR based on 262 municipalities; and (b) kernel map of SMR.

Table I. Summary statistics from the 262 zones.

	Observed death count	SMR 25–64
Max.	1215	2.41
Min.	4	0.57
Mean	176.2	0.99
S.D.	216.5	0.22

The result of this procedure suggested an appropriate bandwidth for kernel mapping of 5 km.

The resultant kernel mapping is shown in Figure 1(b) which presents a smoothed picture of the geographical distribution of working-age deaths. Low rates are mainly found in the ring of suburbs surrounding the city's core. The inner and peripheral zones tend to have higher working-age mortality rates. We can also see some sectoral divisions of mortality. For instance, the industrial zones in the so-called *Shitamachi* area are characterised by higher mortality rates, while the western sector of Tokyo City, the so-called *Yamanote* area, has lower mortality rates.

5.3. Global Poisson regression

To examine possible determinants of the geographical patterns of working-age mortality rates described above, we first calibrated a traditional global Poisson regression model relating counts of working-age mortality to an offset variable, which is the expected number of working-age deaths based on the demographic composition of each zone and national mortality rates for these demographic sectors, and to four socio-economic covariates as shown below.

$$\hat{O}_i = E_i \exp\left(\underset{(-4.6)}{-0.032} - \underset{(-14.1)}{0.092} \text{PRO} + \underset{(11.1)}{0.077} \text{OLD} - \underset{(-5.5)}{0.050} \text{OWNH} + \underset{(5.8)}{0.035} \text{UNEMP}\right) \quad (52)$$

The AICc associated with this model is 399.5 which is meaningless on its own but which we will later compare to other models. The values in parenthesis are t values which indicate that all the global parameter estimates are significant at the 1 per cent level. Each explanatory variable is standardized to have zero mean and one standard deviation so that the values of regression parameters are comparable for evaluating the degree-of-contribution to the model.

The first explanatory variable (PRO) is the proportion of professional and technical workers in each zone. As the proportion of such workers in a zone increases, the working-age mortality decreases. This variable acts as a surrogate for high socio-economic status and the ability to afford better access to medical resources. The explanatory power of this variable is the highest of the four explanatory variables.

The second most important explanatory variable is the proportion of elderly people (aged over 64) within each zone (OLD). Higher counts of working-age mortality are found in areas with higher proportions of elderly people. Such areas having high proportions of elderly tend to be those which are relatively deprived in inner city or rural parts of the metropolitan area. The younger population have moved away from such relatively deprived areas in search of better opportunities and/or living conditions.

The third explanatory variable is the rate of house-ownership (OWNH) that reflects household affluence and living conditions. The index is also inversely related to the degree of urbanisation: the rate tends to be higher in rural and suburban zones due to the land price decline away from the core metropolitan zones. The mortality counts tend to be low where the rate of home ownership is high.

The fourth variable is unemployment rate (UNEMP). The mortality counts tend to be higher where unemployment rates are high, although this is the weakest of the four explanatory variables.

5.4. Model selection using GWPR

The global parameter estimates from the Poisson model appear to make intuitive sense in terms of the association between working-age mortality and the four ecological variables describing the Tokyo metropolitan area. However, in accepting this model we would be implicitly assuming that these relationships are the same everywhere within the study area. This might turn out to be a reasonable assumption but we should not simply accept it without further

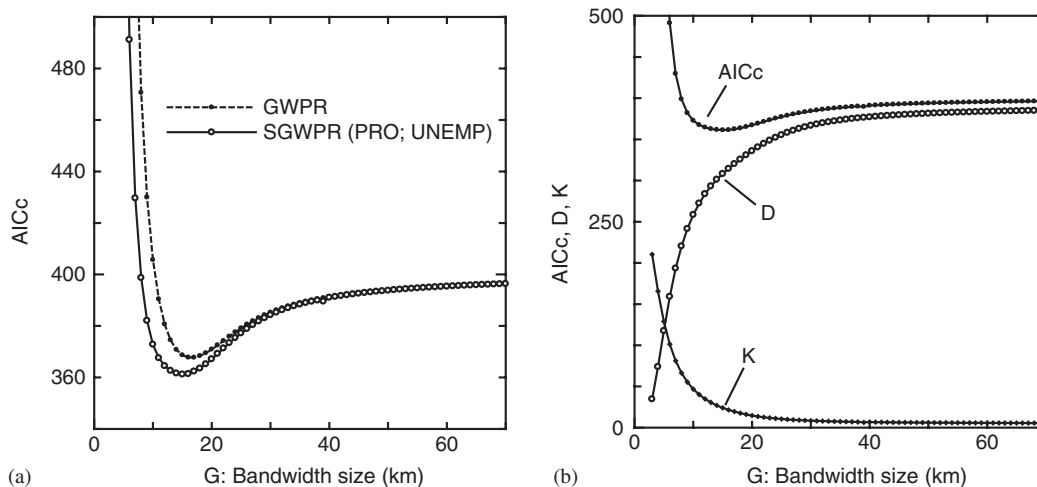


Figure 2. Bandwidth selection by AICc for GWPR models; (a) relationship between bandwidth size and AICc; (b) relationship between bandwidth size, deviance and effective number of parameters.

examination. It could be, for instance, that these global averages reported in the calibration of the traditional model hide some interesting local relationships.

The research question is then to examine whether there are any meaningful spatial variations in the relationships between working-age mortality and the four explanatory variables in the model. To address this issue, we calibrated GWPR and semi-parametric GWPR models using software written by the authors.[¶] Bandwidth size for each model is separately determined by the same method as that used for kernel mapping: applying bandwidth sizes ranging from 5 to 70 km in steps of 1 km to find the lowest AICc as the indicator of optimal bandwidth size.^{||} Figure 2(a) shows the relationship between AICc and bandwidth size for two selected models. One of these is a full GWPR model in which all the parameters were allowed to vary spatially and one is a semi-parametric model in which the parameters associated with the variables PRO and UNEMP are allowed to vary spatially while the parameters associated with the variables OLD and OWNH are held constant. In both cases the optimal bandwidth was around 15 km. Figure 2(b) shows the increase in deviance of the model (D) and the decrease in the effective number of parameters in the model (K) as the bandwidth increases. As the bandwidth increases, the model becomes less accurate and tends towards the global

[¶]Details of Windows-based software for GWR (GWR 3.5) are available at www.may.ie/NCG/GWR or by emailing Stewart.Fotheringham@may.ie. This software is written in FORTRAN but with a user-friendly Windows-based front-end that allows GWR models to be calibrated by simple point and click operations. The calibrations of GWPR and its semi-parametric variants described in this paper were undertaken with a code written in C. This code will be incorporated into the next version of the GWR software to be released by early 2005.

^{||}The authors' software for GWR contains the option of using either AICc or cross-validation (CV) for bandwidth selection. We chose to use AICc here mainly because this function contains a more proper penalty for model complexity in a local modelling context and so is less likely to lead to an over-parameterized model than CV and conventional AIC. However, we recognize that more research is needed into the properties of the various statistics that can be used for bandwidth selection in GWR and its variants.

Table II. Summary statistics of model comparisons.

	Bandwidth	Deviance	K	AICc	Diff_AICc
Kernel map	5 km	343.2	66.5	522.3	161.1
Global PR	NA	389.3	5.0	399.5	38.3
GWPR	17 km	304.5	28.1	367.7	6.5
SGWPR(OWNH;OLD;UNEMP)	17 km	318.8	24.7	373.5	12.3
SGWPR(PRO;OLD;UNEMP)	16 km	304.0	26.9	364.2	3.0
SGWPR(PRO;OWNH;UNEMP)	15 km	296.2	29.8	363.6	2.4
SGWPR(PRO;OLD;OWNH)	16 km	316.9	25.8	374.4	13.2
SGWPR(PRO;UNEMP)	15 km	308.0	24.0	361.2	0.0

Notes: SGWPR($X; Y; Z$) indicates a semi-parametric geographically weighted Poisson regression model in which only the parameters of the variables X , Y and Z are allowed to vary spatially. Diff_AICc is the difference of AICc of a model from that of the best model, SGWPR(PRO;UNEMP).

model both in terms of model fit and number of parameters. Table II contains a summary of the results of the full GWPR model and the semi-parametric model along with several other models that are discussed below.

The results in Table II highlight several interesting features of the relationships being investigated. A comparison of the AICc values suggests that the worst ‘model’ is that of the kernel map which contains no explanatory variables and simply allows the intercept term to vary locally. The global Poisson model with four explanatory variables is the next worst performing model and all the GWPR models outperform the global model suggesting that there is spatial non-stationarity in at least some of the relationships being examined.

The optimal model is a semi-parametric one, SGWPR(PRO;UNEMP), in which the parameters associated with the variables OLD and OWNH are held constant and the parameters associated with the variables PRO and UNEMP are allowed to vary spatially. This suggests that the relationships between working-age mortality and these latter two variables are not stationary but vary over the study region for some reason. The calibrated version of this model is

$$\hat{O}_i = E_i \exp(\hat{\beta}_0(\mathbf{u}_i) + \hat{\beta}_{\text{PRO}}(\mathbf{u}_i)\text{PRO} + 0.069 \underset{(9.2)}{\text{OLD}} - 0.063 \underset{(-4.7)}{\text{OWNH}} + \hat{\beta}_{\text{UNEMP}}(\mathbf{u}_i)\text{UNEMP}) \quad (53)$$

where the constant term and two of the four regression parameters are geographically varying. The estimated fixed parameters and their associated t values are shown and are similar to those in the global model.

Through comparisons of the AICc value associated with the full GWPR with those of the semi-parametric GWPRs in which only one regression parameter is fixed and the other three are allowed to vary spatially, we can also examine whether the spatial variation of a local regression parameter is meaningful. As explained in Section 3.3, better models are indicated by lower AICc values although a common rule-of-thumb is that if the difference in AICc values between two models is less than or equal to 2 [32], there is no substantive difference

in the performance of the two models. By comparing the AICc values in this way, it is clear that when the parameters of the variables OWNH, OLD and UNEMP are allowed to vary spatially but the parameters of the variable PRO are held constant over space, this produces a poorer model than the full GWPR model (the AICc for the former is 373.5; the AICc for the latter is 367.7). Similarly, when the other three sets of parameters are allowed to vary spatially but the parameter for UNEMP is fixed (AICc = 374.4), the model is worse than the full GWPR model. In comparison the optimal model, in which only the parameters for the variables PRO and UNEMP are allowed to vary, a semi-parametric model which also allows the parameters of OLD to vary performs marginally worse. Similarly, allowing the parameters of the variable OWNH to vary as well as those for PRO and UNEMP produces a marginally worse performance.

These results strongly suggest that the relationships between working-age mortality and the two variables, OLD and OWNH, are stationary over space but that the relationships between working-age mortality and PRO and UNEMP are not stationary but vary over space. We now examine these two relationships more closely.

5.5. Interpretation of geographical parameter variations

The above results suggest that the relationships between working-age mortality and the percentage of employment in professional occupations and between working-age mortality and unemployment rates vary spatially. The spatial variations in these local parameters can be mapped as shown in Figures 3 and 4. The local parameter surfaces depicted in these two maps are taken from the semi-parametric model in which only these two parameters were allowed to vary spatially. We define a standard odds ratio showing the sensitivity of working-age death to a change of one standard deviation of each explanatory variable, $SD(x_k)$, as

$$\text{Odds}_k(\mathbf{u}_i) = \exp(\beta_k(\mathbf{u}_i) SD(x_k)) \quad (54)$$

where $SD(x_k)$ is the standard deviation of k th variable (k is PRO or UNEMP here). It should be noted again that every explanatory variable was standardized in this study so that $SD(x_k)$ is equal to one for all k . In Figures 3 and 4, these odds ratios are shown by monochrome gradation with contour lines of pseudo t values.

In Figure 3, the local estimates when exponentiated are all less than 1.0 which suggests that as the proportion of professional people in a zone increases, working-age mortality tends to decrease. Apart from the eastern part of the region, the general spatial pattern of the local estimates suggests that the sensitivity of working-age mortality to variations in the proportion of professional people in each zone increases toward the central part of the area with ridges along the major railway lines. This general pattern of sensitivity indicates that the socio-economic gradient of health would be clearer in densely inhabited areas in and around Tokyo City. A larger socio-economic inequality of health is likely to emerge through larger inequalities in living conditions and levels of social support across different income groups within urban areas than within rural areas (cf. References [37, 38]).

An exception to this general pattern is found in the eastern part of the region between Chiba City and the Pacific Ocean where the local odds are much lower. This is an area where suburban development has not been as intense as in the west (note the low density of railway lines) but where new residential developments of white-collar workers commuting to Chiba or Tokyo were developed in the 1980s primarily near to the railway lines for ease

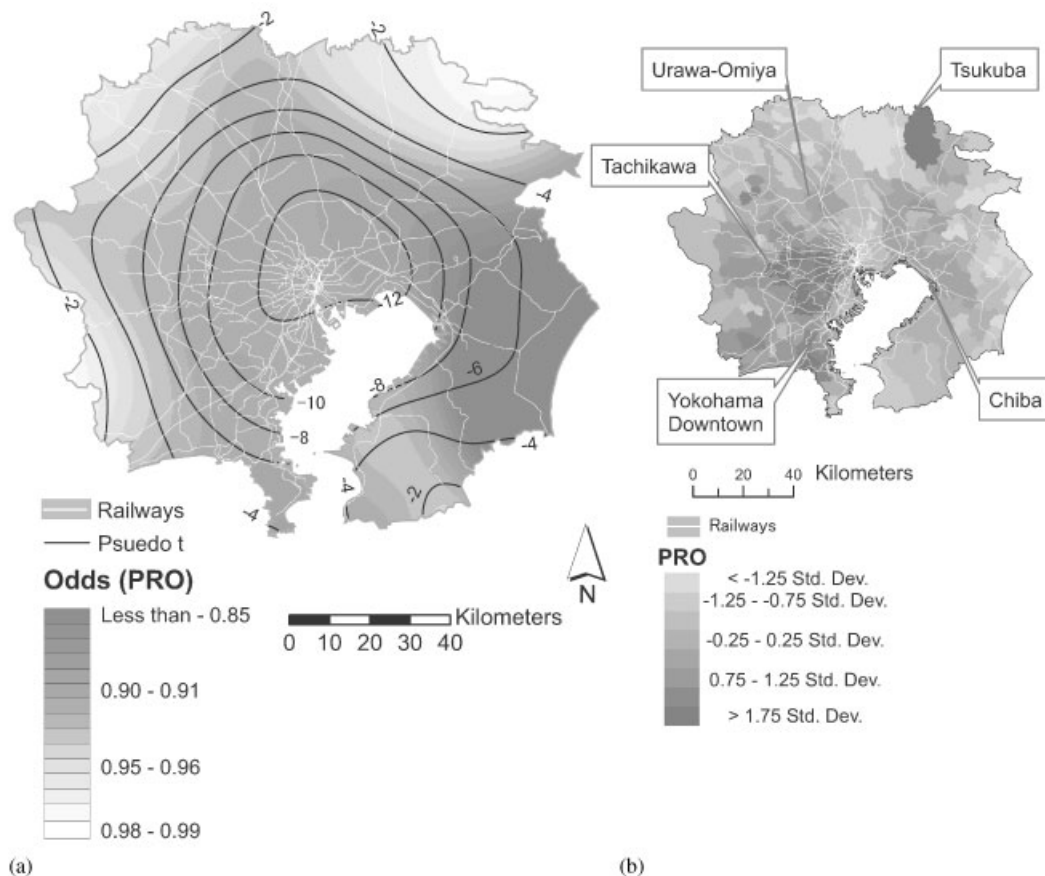


Figure 3. Spatial distributions of PRO and its sensitivity to working-age SMR: (a) map of odds ratio of PRO; (b) map of PRO distribution.

of commuting. These new developments have created a strong local contrast between the relatively healthy commuter suburbs occupied by newcomers and the relatively unhealthy rural areas in the region.

As is evident on the right-hand map in Figure 3 showing the distribution of professional workers throughout the region, Tsukuba City has a much higher proportion of professional people than its surrounding rural areas. However, there is no strong local contrast in working-age mortality patterns and the relationship between the proportion of professional people and working-age mortality is almost negligible. Tsukuba Science City has been developed since around 1980 by moving governmental research institutes and universities from Tokyo. A 'Tsukuba Syndrome', a pseudonym for mental disorders caused by a feeling of being trapped in an isolated artificial environment, has been reported among professional people in the city during 1980s [40]. Thus the health gap between newly resident professional workers and the original rural inhabitants is much smaller around Tsukuba City than in other areas such as Chiba City and consequently the local effect of variations in professional employment on SMR around Tsukuba is very small.

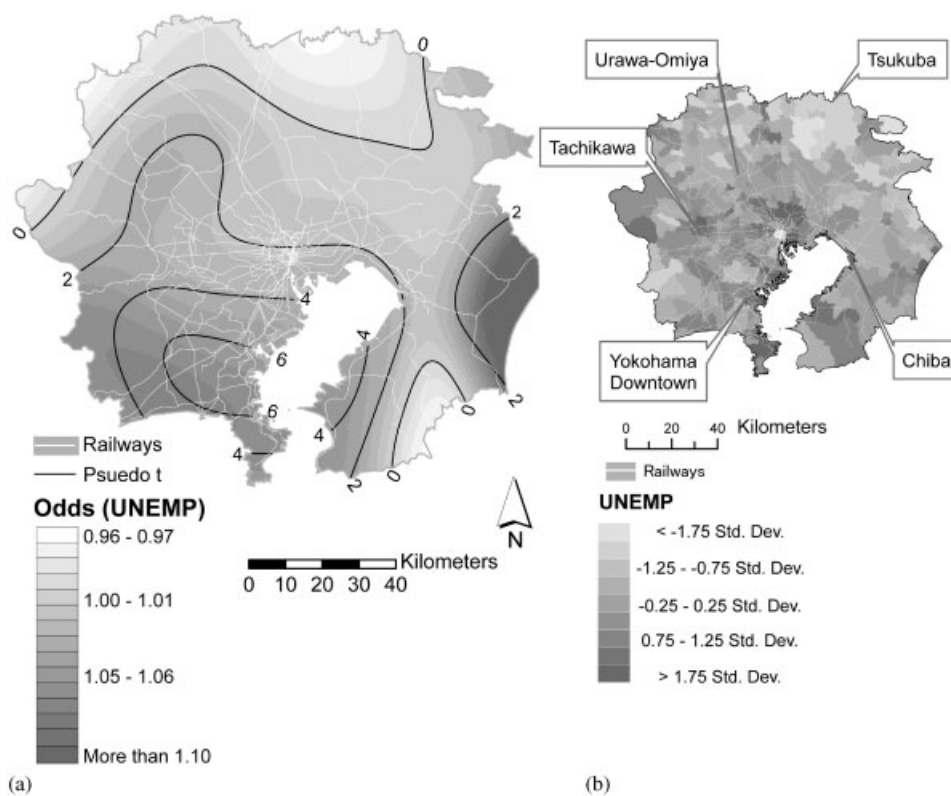


Figure 4. Spatial distributions of UNEMP and its sensitivity to working-age SMR: (a) map of odds ratio of UNEMP; (b) map of UNEMP distribution.

The spatial variation of the local parameter estimates associated with the unemployment variable is shown in Figure 4. The local estimates are positive in some areas (odds > 1) and negative in others (odds < 1). The areas where significant positive relationships occur are generally urbanized areas in the south-west and industrialized areas surrounding Tokyo Bay. In these areas, as unemployment rates increase, working-age mortality increases. However, the area having the strongest positive relationship is a rural coastal zone in the far eastern edge of the region. This area is dominated by the fishing and tourism industries. Declining trends in employment in these sectors might explain the strong local relationship between unemployment and working-age mortality.

In most other parts of the study area, the relationship between working-age mortality and unemployment is very weak, as shown by the pseudo t values. These zones are largely agricultural areas where unemployment tends to be masked by temporary employment and by underemployment on family farms.

6. CONCLUDING REMARKS

The approach of geographically weighted Poisson regression (GWPR) described here can be considered to be a natural extension or integration of two conventional statistical tools for

disease map analysis: kernel mapping [30, 41] and Poisson regression modelling [14, 19, 20]. While the traditional kernel methodology in spatial epidemiology is used for disease mapping to reveal geographical variations of disease/health *levels*, GWPR is used to reveal geographical variations in *relationships*. Through the use of GWPR we can look for meaningful geographical drifts and/or spatial anomalies (clusters) in the associations between health outcomes and various socio-economic and environmental variables.

Within a geographical information systems environment, this new method of analysis can be easily used as a new visualisation and explorative spatial analysis technique. It would also be possible to extend our approach of geographically weighted regression to a general linear modelling framework by following similar arguments to those set out here. Particularly, binomial logistic modelling of health data with geographically varying coefficient models may be useful for analysing case-control data at the individual level with detailed geographical referencing in order to avoid ecological fallacy problem [42].

We should mention that the idea to combine kernel mapping and regression models for disease map analysis is not new. Originally, spatial kernel weighting was introduced to estimate smoothed surface of disease rates [41, 43]. Such kernel mapping has been extended to take into account geographically invariant effects of explanatory variables (i.e. invariant regression parameters with a spatial varying intercept term) [44, 45]. Rather than assuming invariant effects on explanatory variables, however, GWPR postulates that unknown (contextual) factors regulating disease distribution can alter the association of disease with covariates. We expect that the information on the spatial variation of the association of disease rate with covariates will provide important clues to understanding geographical contextual effects regulating disease distributions.

Most importantly, GWPR is proposed here to increase attention on the local nature of ecological associations in health studies. As discussed in Reference [6], the effort to measure local properties has a close relationship with the mode of geographical thought stressing geographical contextuality/locality. It is hoped that this approach will add value to spatial ecological studies of disease and health and will lead to further advances of understanding in this area.

ACKNOWLEDGEMENTS

We thank Dr Yuzuru Isoda for his helpful comment on the earlier draft of this manuscript. This paper is supported by a Grant-in-Aid for Encouragement of Young Scientists of Japan Society for the Promotion of Science (No. 15700549). The second author would like to acknowledge the generous support of a Science Foundation Research Professorship Award (03/RP1/I382).

REFERENCES

1. Cliff AD, Haggett P. *Atlas of Disease Distributions: Analytic Approaches to Epidemiological Data*. Blackwell: Oxford, 1988.
2. Douven W, Scholten HJ. Spatial analysis in health research. In *The Added Value of Geographical Information Systems in Public and Environmental Health*, de Lepper MJC, Scholten HJ, Stern RM (eds). Kluwer: Dordrecht, 1995; 117–133.
3. Pyle GF, Rees PH. Modeling patterns of death and disease in Chicago. *Economic Geography* 1971; **47**: 475–488.
4. Marshall RJ. A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society, Series A* 1991; **154**:421–441.

5. Lawson A, Biggeri A, Bohning D, Lesaffre E, Viel J-F, Bertollini R (eds). *Disease Mapping and Risk Assessment for Public Health*. Wiley: Chichester, 1999.
6. Jones III JP, Hanham RQ. Contingency, realism, and the expansion method. *Geographical Analysis* 1995; **27**: 185–207.
7. Fotheringham AS, Brunsdon C. Local forms of spatial analysis. *Geographical Analysis* 1999; **31**:340–358.
8. Fotheringham AS, Brunsdon C, Charlton M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley: Chichester, 2002.
9. Macintyre S, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In *Social Epidemiology*, Berkman LF, Kawachi I (eds). Oxford University Press: Oxford, 2000; 332–348.
10. Leyland AH, Goldstein H. *Multilevel Modelling of Health Statistics*. Wiley: Chichester, 2001.
11. Jones K. *Multi-level Models for Geographical Research*. CATMOG 54, Geo Books: Norwich, 1991.
12. Gould MI, Jones K. Analysing perceived limiting long-term illness using UK Census Microdata. *Social Science and Medicine* 1996; **42**:857–869.
13. Openshaw S. *The Modifiable Areal Unit Problem*. CATMOG 38, Geo Books: Norwich, 1984.
14. Nakaya T. An information statistical approach to the modifiable areal unit problem in incidence rate maps. *Environment and Planning A* 2000; **32**:91–109.
15. Brunsdon C, Fotheringham AS, Charlton M. Geographically weighted regression: a method for exploring spatial non-stationarity. *Geographical Analysis* 1996; **28**:281–298.
16. Casetti E. Generating models by the expansion method: applications to geographic research. *Geographical Analysis* 1972; **4**:81–91.
17. Jones JP, Casetti E. *Applications of the Expansion Method*. Longman: London, 1992.
18. Fotheringham AS, Brunsdon C, Charlton M. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and Planning A* 1998; **30**:1905–1927.
19. Lovett A, Bentham CG, Flowerdew R. Analysing geographic variations in mortality using Poisson regression: the example of ischaemic heart disease in England and Wales, 1969–1973. *Social Science and Medicine* 1986; **23**:935–943.
20. Lovett A, Flowerdew R. Analysis of count data using Poisson regression. *Professional Geographer* 1989; **41**: 190–198.
21. Choynowski M. Maps based on probabilities. *American Statistical Association Journal* 1959; **54**:385–388.
22. Elliott P, Wakefield JC, Best NG, Briggs DJ (eds). *Spatial Epidemiology: Methods and Applications*. Oxford University Press: Oxford, 2000.
23. Green PJ, Silverman BW. *Nonparametric Regression and Generalised Linear Models: A Roughness Penalty Approach*. Chapman & Hall: London, 1994.
24. Hastie TJ, Tibshirani RJ. *Generalized Additive Models*. Chapman & Hall: London, 1990.
25. McCullagh P, Nelder JA. *Generalized Linear Models*. Chapman & Hall: London, 1989.
26. Hastie TJ, Tibshirani RJ. Varying-coefficient models. *Journal of the Royal Statistical Society B* 1993; **55**: 757–796.
27. Loader C. *Local Regression and Likelihood*. Springer: New York, 1999.
28. Nakaya T. Local spatial interaction modelling based on the geographically weighted regression approach. *GeoJournal* 2001; **53**:347–358.
29. Cleveland WS, Devlin SJ. Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association* 1988; **83**:596–610.
30. Bithell JF. Disease mapping using the relative risk function estimated from areal data. In *Disease Mapping and Risk Assessment for Public Health*, Lawson A, Biggeri A, Bohning D, Lesaffre E, Viel J-F, Bertollini R (eds). Wiley: Chichester, 1999; 247–255.
31. Akaike H. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Petrov BN, Csaki F (eds). Akademiai Kiado: Budapest, 1999; 267–281.
32. Sakamoto Y, Ishiguro M, Kitagawa G. *Jyohoryo Tokeigaku* (Information Statistics). Kyoristu-shuppansha: Tokyo, 1983 (in Japanese).
33. Burnham K, Anderson DR. *Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach* (2nd edn). Springer: New York, 2002.
34. Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods A* 1978; **7**:13–26.
35. Hurvich CM, Tsai C-L. Regression and time series model selection in small samples. *Biometrika* 1989; **76**: 297–307.
36. Hurvich CM, Simonoff JF, Tsai C-L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society B* 1998; **60**:271–293.
37. Wilkinson RG. *Unhealthy Societies: The Afflictions of Inequality*. Routledge: London, 1996.
38. Marmot M, Wilkinson RG. *Social Determinants of Health*. Oxford University Press: Oxford, 1999.
39. Nakaya T. An information statistical evaluation of areal units for mortality mapping: toward constructing an analyzing system of disease maps on geographical information system. *Theory and Application of GIS* 1996; **4**:53–60 (in Japanese with English Abstract).

40. Ota S. Tsukuba Gakuen toshi no seishin-hoken: iwayuru 'Tsukuba syokogun' no rekishi (Mental health in Tsukuba Science City: history of the 'Tsukuba Syndrome'). *Tsukuba Forum* 1991; **30**:26–35 (in Japanese).
41. Bithell JF. An application of density estimation to geographical epidemiology. *Statistics in Medicine* 1990; **9**:691–701.
42. Robinson WS. Ecological correlations and the behavior of individuals. *American Sociological Review* 1950; **15**:351–357.
43. Kelsall JE, Diggle PJ. Kernel estimation of relative risk. *Bernoulli* 1995; **1**:3–16.
44. Kelsall JE, Diggle PJ. Spatial variation in risk: a nonparametric binary regression approach. *Applied Statistics* 1998; **47**:559–573.
45. Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics (with discussion). *Applied Statistics* 1998; **47**: 299–350.