Environment and Planning A 1998, volume 30, pages 1905–1927

Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis

A S Fotheringham, M E Charlton
Department of Geography, University of Newcastle, Newcastle upon Tyne, NE1 7RU, England; e-mail: stewart.fotheringham@newcastle.ac.uk; martin.charlton@newcastle.ac.uk
C Brunsdon
Department of Town and Country Planning, University of Newcastle, Newcastle upon Tyne, NE1 7RU, England; e-mail: chris.brunsdon@ncl.ac.uk
Received 18 December 1996; in revised form 22 May 1997

Abstract. Geographically weighted regression and the expansion method are two statistical techniques which can be used to examine the spatial variability of regression results across a region and so inform on the presence of spatial nonstationarity. Rather than accept one set of 'global' regression results, both techniques allow the possibility of producing 'local' regression results from any point within the region so that the output from the analysis is a set of mappable statistics which denote local relationships. Within the paper, the application of each technique to a set of health data from northeast England is compared. Geographically weighted regression is shown to produce more informative results regarding parameter variation over space.

1 Spatial nonstationarity

A frequent aim of data analysis is to identify relationships between pairs of variables, often after negating the effects of other variables. By far the most common type of analysis used to achieve this aim is that of regression, in which relationships between one or more independent variables and a single dependent variable are estimated. In spatial analysis the data are drawn from geographical units and a single regression equation is estimated. This has the effect of producing 'average' or 'global' parameter estimates which are assumed to apply equally over the whole region. That is, the relationships being measured are assumed to be *stationary* over space. Relationships which are not stationary, and which are said to exhibit *spatial nonstationarity*, create problems for the interpretation of parameter estimates from a regression model. It is the intention of this paper to compare the results of two statistical techniques, Geographically weighted regression (GWR) and the expansion method (EM), which can be used both to account for and to examine the presence of spatial nonstationarity in relationships.

It would seem reasonable to assume that relationships might vary over space and that parameter estimates might exhibit significant spatial variation in some cases. Indeed, the assumption that such events do *not* occur, which until recently has been relatively unchallenged, seems rather suspect. There are three reasons why parameter estimates from a regression model might exhibit spatial variation: that is, why we might expect parameters to be different if we calibrated the same models from data drawn from different parts of the region (as shown by Fotheringham et al, 1996; 1997). The first and simplest is that parameter estimates will vary because of random sampling variations in the data used to calibrate the model. The contribution of this source of variation is not of interest here but needs to be eliminated by significance testing. In this paper we want to concentrate on large-scale, statistically significant variations in parameter estimates over space, the source of which cannot be attributed solely to sampling. The second explanation is that, for whatever reason, some relationships are intrinsically different across space. Perhaps, for example, there are spatial variations in people's tastes or attitudes or there are different administrative, political, or other contextual issues that produce differing responses to the same stimuli across space. In which case, it is clearly useful to have a technique that can identify the nature of these variations in relationships over space; without such a technique only a global or average relationship can be estimated and this may bear little resemblance to particular local relationships. This is a situation where we throw away a great deal of interesting spatial detail in relationships.

The third reason why some relationships might exhibit spatial variation is that the model from which the relationships are being estimated is a gross misspecification of reality and that one or more relevant variables have either been omitted from the model or represented by an incorrect functional form and are making their presence felt through the parameter estimates. Given that all models, by their nature, are likely to be misspecifications of reality, the potential for this misspecification to be sufficiently gross as to cause spatial nonstationarity in parameter estimates would seem quite high.

2 The expansion method approach to measuring spatial nonstationarity

Several techniques aimed at measuring and incorporating spatial nonstationarity already exist in the literature. Perhaps the best known is that of the expansion method (Casetti, 1972; Jones and Casetti, 1992) which is an attempt to measure parameter 'drift'. In this framework, parameters of a global model can be made functions of other attributes including geographic space so that *trends* in parameter estimates over space can be measured (Eldridge and Jones, 1991; Fotheringham and Pitts, 1995).

Initially, a global model is proposed such as:

$$y_i = \alpha + \beta x_{i1} + \ldots + \tau x_{in} + \varepsilon_i \,, \tag{1}$$

where y represents a dependent variable, the x are independent variables, α, β, ..., τ represent parameters to be estimated, ε represents an error term, and i represents a point in space at which observations on the y and x are recorded. The basic model can be expanded by allowing each of the parameters to be functions of other variables. Although most applications of the expansion method (see Jones and Casetti, 1992) have undertaken aspatial expansions, it is relatively straightforward to allow the parameters to vary over geographic space so that

$$\left.\begin{aligned} \alpha_i &= \alpha_0 + \alpha_1 u_i + \alpha_2 v_i \,, \\ \beta_i &= \beta_0 + \beta_1 u_i + \beta_2 v_i \,, \\ \tau_i &= \tau_0 + \tau_1 u_i + \tau_2 v_i \,, \end{aligned}\right\} \tag{2}$$

where u and v represent spatial coordinates with u being an easting and v being a northing. Equation (2) represents very simple linear expansions over space but more complex, nonlinear, expansions such as

$$\alpha_i = \alpha_0 + \alpha_1 u_i + \alpha_2 v_i + \alpha_3 u_i^2 + \alpha_4 v_i^2 + \alpha_5 u_i v_i \,, \tag{3}$$

can easily be accommodated although the interpretation of individual parameter estimates can then be difficult.

Whichever form of expansion is selected, once the expanded model has been calibrated, the estimated parameters are used to obtain spatially varying estimates of the parameters in equation (1) from either equation (2) or equation (3). The estimates of α_i, β_i, etc can then be mapped to display spatial variations in relationships. Quite clearly, though, the technique is limited to displaying trends in relationships over space, with the complexity of the trends measured dependent upon the complexity of the expansion equations. It should also be noted that the expansion equations are assumed to be deterministic to remove problems of estimation in the terminal model.

3 Geographically weighted regression

Consider the global regression model given by

$$y_i = a_0 + \sum_k a_k x_{ik} + \varepsilon_i \,. \tag{4}$$

GWR is a relatively simple technique that extends the traditional regression framework of equation (4) by allowing local rather than global parameters to be estimated so that the model is rewritten as

$$y_i = a_0(u_i,v_i) + \sum_k \mathrm{a}_k(u_i,v_i)x_{ik} + \varepsilon_i \,, \tag{5}$$

where (u_i,v_i) denotes the coordinates of the ith point in space and $\mathrm{a}_k(u_i,v_i)$ is a realisation of the continuous function $\mathrm{a}_k(u,v)$ at point i. That is, we allow there to be a continuous surface of parameter values and measurements of this surface are taken at certain points to denote the spatial variability of the surface. Note that equation (4) is a special case of equation (5) in which the parameter surface is assumed to be constant over space. Thus the GWR expression in equation (5) is a recognition that spatial variations in relationships might exist and provides a way in which they can be measured.

As it stands, though, there are problems in calibrating equation (5): there are more unknowns than observed variables. However, many models of this kind have been proposed before and they are reviewed by Rosenberg (1973) and Spjotvoll (1977) and more recent work has been carried out by Hastie and Tibshirani (1990). Our approach borrows from the latter particularly in the fact that we do not assume the coefficients to be random, but rather that they are deterministic functions of some other variables—in our case location in space. The general approach when handling such models is to note that although an *unbiased* estimate is not possible, estimates with a small amount of bias can be provided. We argue here that the estimation process in GWR can be thought of as a trade-off between bias and standard error. Assuming the parameters exhibit some degree of spatial consistency then values near to the one being estimated should have relatively similar magnitudes and signs. Thus, when estimating a parameter for a given point i, one can approximate equation (5) in the region of i by equation (4), and perform an ordinary least squares (OLS) regression with a subset of the points in the data set that are close to i. Thus, the $\mathrm{a}_k(u_i,v_i)$ are estimated for i in the usual way and for the next i, a new subset of 'nearby' points is used, and so on. These estimates will have some degree of bias, because the coefficients of equation (5) will exhibit some drift across the local calibration subset. However, if the local sample is large enough, this will allow a calibration to take place—albeit a biased one. The greater the size of the local calibration subset the lower the standard errors of the coefficient estimates; but this must be offset against the fact that enlarging this subset increases the chance that the coefficient 'drift' introduces bias. To reduce this effect one final adjustment to this approach may also be made. Assuming that points in the calibration subset further from i are more likely to have differing coefficients, a weighted OLS calibration is used, so that more influence in the calibration is attributable to the points closer to i.

As noted above, the calibration of equation (5) assumes implicitly that observed data near to point i have more of an influence in the estimation of the $\mathrm{a}_k(u_i,v_i)$ than data located farther from i. In essence, the equation measures the relationships inherent in the model *around each point i*. Hence weighted least squares provides a basis for understanding how GWR operates. In GWR an observation is weighted in accordance with its proximity to point i so that the weighting of an observation is no longer constant in the calibration but varies with i. Data from observations close to i are weighted more than data from observations farther away. That is,

$$\hat{\mathbf{a}}(u_i, v_i) = [\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}(u_i, v_i)\mathbf{y} \tag{6}$$

where the bold type denotes a matrix, $\hat{\mathbf{a}}$ represents an estimate of $\mathbf{a}$ and $\mathbf{W}(u_i, v_i)$ is an $n \times n$ matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data for point i. There are parallels between GWR and that of kernel regression and drift analysis of regression parameters (DARP) (Casetti, 1982; Cleveland, 1979; Cleveland and Devlin, 1988). In kernel regression and DARP, $\mathbf{y}$ is modelled as a nonlinear function of $\mathbf{X}$ by weighting data in attribute space rather than geographic space. That is, data points more similar to x_i are weighted more heavily than data points which are less similar and the output is a set of localized parameter estimates in x-space. However, Casetti and Jones (1983) do provide a limited spatial application of DARP which is very similar in intent to GWR although it lacks a formal calibration mechanism and significance-testing framework and so is treated by the authors as a rather limited heuristic method.

It should be noted that, as well as producing localised parameter estimates, the GWR technique described above will produce localised versions of all standard regression diagnostics including goodness-of-fit measures such as R^2. The latter can be particularly informative in understanding the application of the model being calibrated and for exploring the possibility of adding additional explanatory variables to the model.

4 Model specification

There are several other multivariate models that allow for spatial effects. One common example is the simultaneous autoregressive (SAR) model (for example, see Cliff and Ord, 1981). This model is identical to the OLS model in terms of the linkage between $E(y)$, the expected value of y, and the x variables, but differs in the model for the distribution of the residuals. In the SAR case the residuals are not independently distributed, but exhibit spatial autocorrelation. Essentially, then, SAR still provides a global prediction model. In fact, using OLS to calibrate an SAR process would still result in unbiased estimates, but would give less efficient coefficient estimators (the standard errors would be greater). This is not the case with GWR. Here the predictor model is not global in the observed variables, so that OLS (or indeed SAR) would not provide unbiased estimates. The basic GWR model agrees with OLS *only* in the form of the error term.

One could argue that some estimated nonstationarity in the regression coefficients may be the result of autocorrelation effects. That is, one may be attempting to calibrate a GWR model when in reality an SAR process is occurring—autocorrelation of residuals becomes displaced to variation in the regression parameters. However, by the same token, on other occasions one may be attempting to calibrate an SAR model when the reality is GWR—variation in regression parameters gets displaced to autocorrelation in residuals. Without divine intervention it is generally difficult to know with certainty which (if either) of the two above cases are true. However, looking at a GWR model estimation gives some insight into how localised effects affect coefficients attached to specific variables—something that SAR cannot do.

Similar arguments may also be applied to prewhitening procedures (Kendall and Ord, 1990). With this approach, spatial filters are applied to all variables (dependent and independent) before fitting a regression model, to remove autocorrelation effects. Again, however, the prediction model will be global. Thus, as before a global linkage is set up, although in this case variables on both sides of the regression equation are assumed to be realisations of random spatial processes. Perhaps a model having the 'best of both worlds' would be one with the E(y) predictor term of GWR but also having spatial autocorrelation in the error terms. Investigation into this kind of model is currently in progress (Brunsdon et al, 1998).

5 Choice of spatial weighting function

Until this point, it has merely been stated in GWR that $\mathbf{W}(u_i,v_i)$ is a weighting scheme based on the proximity of i to the sampling locations around i without an explicit relationship being stated. The choice of such a relationship will be considered here. First, consider the implicit weighting scheme of the OLS framework in equation (4). Here

$$w_{ij} = 1, \qquad \forall i,j, \tag{7}$$

where j represents a specific point in space at which data are observed and i represents any point in space for which parameters are estimated. That is, in the global model each observation has a weight of unity. An initial step towards weighting based on locality might be to exclude from the model calibration observations that are further than some distance d from the locality. This would be equivalent to setting their weights to zero, giving a weighting function of

$$w_{ij} = \begin{cases} 1, & \text{if } d_{ij} < d\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{8}$$

The use of equation (8) would simplify the calibration procedure because for every point for which coefficients are to be computed only a subset of the sample points need to be included in the regression model. However, the spatial weighting function in equation (8) suffers the problem of discontinuity. As i varies around the study area, the regression coefficients could change drastically as one sample point moves into or out of the circular buffer around i and which defines the data to be included in the calibration for location i. Although sudden changes in the parameters over space might genuinely occur, in this case changes in their estimates would be artifacts of the arrangement of sample points, rather than any underlying process in the phenomena under investigation. One way to combat this is to specify w_{ij} as a continuous function of d_{ij}, the distance between i and j. One obvious choice is

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{\beta^2}\right), \tag{9}$$

where β is referred to as the bandwidth. If i and j coincide (that is, i also happens to be a point in space at which data are observed), the weighting of data at that point will be unity and the weighting of other data will decrease according to a Gaussian curve as the distance between i and j increases. In the latter case the inclusion of data in the calibration procedure becomes 'fractional'. For example, in the calibration of a model for point i, if $w_{ij} = 0.5$ then data at point j contribute only half the weight in the calibration procedure as data at point i itself. For data a long way from i the weighting will fall to virtually zero, effectively excluding these observations from the estimation of parameters for location i.

Whatever the specific weighting function employed, the essential idea of GWR is that for each point i there is a 'bump of influence' around i corresponding to the weighting function such that sampled observations near to i have more influence in the estimation of the parameters of i than do sampled observations farther away.

6 Calibrating the spatial weighting function

One difficulty with GWR is that the estimated parameters are, in part, functions of the weighting function or kernel selected in the method. In equation (8), for example, as d becomes larger, the closer will be the model solution to that of OLS and when d is equal to the maximum distance between points in the system, the two models will be equal. Equivalently, in equation (9) as β tends to infinity, the weights tend to one for all pairs of points so that the estimated parameters become uniform and GWR becomes equivalent to OLS. Conversely, as the bandwidth becomes smaller, the parameter estimates will increasingly depend on observations in close proximity to i and hence will have increased variance. The problem is therefore how to select an appropriate bandwidth or decay function in GWR.

Consider the selection of β in equation (9). One possibility is to choose β on a 'least squares' criterion. One way to proceed would be to minimise the quantity

$$\sum_{i=1}^{n} [y_i - \hat{y}_i(\beta)]^2 , \tag{10}$$

where $\hat{y}_i(\beta)$ is the fitted value of y_i with a bandwidth of β. In order to find the fitted value of y_i it is necessary to estimate the $a_k(u_i,v_i)$ at each of the sample points and then combine these with the x-values at these points. However, when minimising the sum of squared errors suggested above, a problem is encountered. Suppose β is made very small so that the weightings of all points except for i itself become negligible. Then the fitted values at the sampled points will tend to the *actual* values, so that the value of expression (10) becomes zero. This suggests that under such an optimising criterion the value of β tends to zero but clearly this degenerate case is not helpful. First, the parameters of such a model are not defined in this limiting case and second, the estimates will fluctuate wildly throughout space in order to give locally good fitted values at each i.

A solution to this problem is a *cross-validation* (CV) approach suggested for local regression by Cleveland (1979) and for kernel density estimation by Bowman (1984). Here, a score of the form

$$\sum_{i=1}^{n} [y_i - \hat{y}_{\neq i}(\beta)]^2 \tag{11}$$

is used, where $\hat{y}_{\neq i}(\beta)$ is the fitted value of y_i with the observations for point i omitted from the calibration process. This approach has the desirable property of countering the 'wrap-around' effect, as when β becomes very small, the model is calibrated only on samples near to i and not at i itself.

Plotting the CV score against the required parameter of whatever weighting function is selected will therefore provide guidance on selecting an appropriate value of that parameter. If it is desired to automate this process, then the CV score could be maximised by using an optimisation technique such as a golden section search (Greig, 1980).

7 Bias–variance trade-off

It is important at this stage to expand on the ideas discussed briefly in section 3 on the relationship between bias and variance both generally and within the context of GWR. To do this, one needs to discuss some properties of the estimator of y, $\hat{y}$. At any point in space (u,v), if we are given a set of predictors, $\mathbf{X}$, and a set of coefficient estimators $\hat{\mathbf{a}}$, then $\hat{y} = \mathbf{X}^{\mathrm{T}}\hat{\mathbf{a}}$ is an estimate of y at that point. However, $\hat{\mathbf{a}}$ is an estimate of $\mathbf{a}$ based on a sample of spatially diffuse $\mathbf{X}$ and y observations. Because of the randomness of the y term, $\hat{\mathbf{a}}$ is random, and therefore so is $\hat{\mathbf{y}}$. Two important properties of the distribution of $\hat{\mathbf{y}}$ are its standard deviation and its expected value,

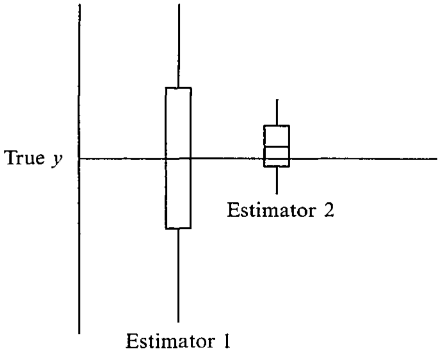

Figure 1. Prediction error distribution of y-estimators.

SD($\hat{y}$) and E($\hat{y}$), respectively. When, for all **X**, E($\hat{y}$) = E(y), the estimator is said to be unbiased. In this case, SD($\hat{y}$) is a useful measure of the quality of $\hat{y}$ as an estimator of y. However, zero bias does not in itself guarantee an optimal estimator.

Consider figure 1. The horizontal line indicates the true value of y and the two boxplots represent the probability distributions of the two estimates of y. Call these $\hat{y}_1$ and $\hat{y}_2$ respectively. Although $\hat{y}_2$ is a biased estimator, its overall variability is less than that of $\hat{y}_1$. Thus, the extreme values of potential errors in the prediction of y are less for $\hat{y}_2$—the only advantage $\hat{y}_1$ has to offer is that the error distribution is centred on zero. If one were to consider the distributions of the prediction squared error (PSE), that for the first estimator would have a much longer tail.

This is an example of *bias–variance trade off*—an important issue that occurs in many types of statistical modelling [for example, see work on multilevel modelling (Goldstein, 1987)]. It is certainly an issue in GWR. If regression coefficients vary continuously over space, then using weighted least-squares regression is unlikely to provide a completely unbiased estimate of $\mathbf{a}(u,v)$ at the given point (u,v) because for each observation there will be a different value of **a** but the regression requires that this value is the same for all observations. The best one can hope is that the values do not vary too much—and this is best achieved by only considering observations close to the point (u,v) at which we wish to estimate $\mathbf{a}(u,v)$. However, because this reduces the effective sample size for the estimate, the standard error of $\hat{\mathbf{a}}(u,v)$ will increase. Thus, the question arises as to *how close* to (u,v) should points be considered. Too close and variance becomes large but the bias is small; too far and the variance is small but the bias is large. At one extreme, if a global model is chosen, then $\hat{\mathbf{a}}(u,v)$ is assumed constant for all (u,v) and if there is much variability in the true $\mathbf{a}(u,v)$ then clearly bias will cause problems. It is the consideration of this bias–variance trade-off that should guide the choice of bandwidth.

This provides some justification for the use of cross-validation scores as a means of choosing bandwidth. A CV score is essentially the sum of estimated squared prediction errors—the quantity discussed earlier. PSEs can be thought of as a measure of the overall performance of a particular bias–variance combination. We cannot know the exact PSEs (if we did, we would know the true E(y) and **a** values and would need no statistical prediction!), but CV scores provide an estimate which can then be used as a basis for selection.

At this point it is worth making two observations on choice of bandwidth. These both depend on the fact that the bandwidth is essentially a measure of how close to (u,v) one has to use data to get a 'good' bias–variance combination. Note that bandwidth choice is a function of the spatial distribution of the observations—if a lot of observations are close to (u,v) then one may not need to look far to find enough observations to reduce the standard error satisfactorily. However, if observations are sparser, then one may well need to extend the search radius to achieve an optimal effect. If the spatial distribution of cases varies notably in density throughout the study area, then suitable bandwidths will not be constant over space. Currently GWR is calibrated by using a global bandwidth selection method but work is in progress on *localised* bandwidth choice where the bandwidth can vary not only across space but for each parameter estimate (Fotheringham et al, 1998).

The second point also follows from the linkage between sample geography and bandwidth choice. For a given set of variables—and therefore a given model—optimal bandwidth will change if the sampling strategy is altered. Thus bandwidth choice is *not* a parameter relating to the model itself, but is essentially part of the calibration strategy for a given sample. For example, if further sample points were added to the model, although one hopes to achieve better estimates of $\mathbf{a}(u,v)$, one would expect the optimal bandwidth to decrease. Ultimately, if the sample size is continually increased, $\hat{\mathbf{a}}(u,v)$ should tend to $\mathbf{a}(u,v)$ but the bandwidth should tend to zero. However, this does not mean that by altering the bandwidth, and observing the changes in $\hat{\mathbf{a}}(u,v)$ one cannot gain some insight into the different scales of variation in $\mathbf{a}(u,v)$. Another view of GWR is that of a spatial filter, which is capable of extracting different spatial frequency components of spatial nonstationarity in regression coefficients. In the light of this interpretation, optimal bandwidth can be regarded as creating a filter which allows nonstationarity trends of all frequencies to pass, but filters out random noise.

8 Testing for spatial nonstationarity

To this point the techniques associated with GWR have been predominantly descriptive. However, it is useful to assess the question: "Does the set of local parameter estimates exhibit significant spatial variation?" The variability of the local estimates can be used to examine the plausibility of the stationarity assumption held in traditional regression. In general terms, this could be thought of as a variance measure. For a given k suppose $\hat{a}_k(u_i,v_i)$ is the GWR estimate of $a_k(u_i,v_i)$. Suppose we take n values of this parameter estimate (one for each point i within the region), an estimate of variability in the parameter is given by the standard deviation of the n parameter estimates. This statistic will be referred to as s_k.

The next stage is to determine the sampling distribution of s_k under the null hypothesis that the global model in equation (1) holds. Although it is proposed to consider theoretical properties of this distribution in the future, for the time being a Monte Carlo approach will be adopted. Under the null hypothesis, any permutation of (u_i,v_i) pairs amongst the geographical sampling points i are equally likely to occur. Thus, the observed values of s_k could be compared with the values obtained from randomly rearranging the data in space and repeating the GWR procedure. The comparison between the observed s_k value and those obtained from a large number (99 in this case) of randomisation distributions forms the basis of the significance test. Making use of the Monte Carlo approach, it is also the case that selecting a subset of random permutations of (u_i,v_i) pairs amongst the i and computing s_k will also give a significance test when compared with the observed statistics.

9 An empirical comparison of the expansion method and GWR: the distribution of limiting long-term illness in the northeast of England

The application of both methods of measuring spatial nonstationarity in relationships is now described with data on the spatial distribution of limiting long-term illness (LLTI) which is a self-reported variable from the UK Census of Population. It encompasses a variety of severe illnesses such as respiratory diseases, multiple sclerosis, heart disease, severe arthritis, as well as physical disabilities which prevent people from being in the labour market. The study area encompasses 605 census wards in four administrative counties in northeast England: Tyne and Wear, Durham, Cleveland, and North Yorkshire. Tyne and Wear is a heavily populated service and industrial conurbation in the northern part of the study area and is centred on the city of Newcastle. To the south Durham has been heavily dependent on coal mining in the eastern half of the county with the western half being predominately rural. Cleveland, to the southeast, is a largely urban, industrial area with heavy petrochemical and engineering works clustered around the Tees estuary and centred on Middlesbrough. North Yorkshire, to the south, is a predominantly rural and fairly wealthy county with few urban areas. The distribution of urban areas throughout the region is shown in figure 2.

Figure 2. Urban areas in the study region.

The spatial distribution of a standardised measure of LLTI (defined as the percentage of individuals aged 45 – 65 living in a household where LLTI is reported) throughout the study region is shown in figure 3 (over). As one might expect, the LLTI variable tends to be higher in the industrial regions of Tyne and Wear, east Durham, and Cleveland and lower in the rural areas of west Durham and North Yorkshire. To model this distribution, the following global regression model was constructed:

$$\mathsf{LLTI}_i = a_0 + a_1\,\mathsf{UNEM}_i + a_2\,\mathsf{CROW}_i + a_3\,\mathsf{SPF}_i + a_4\,\mathsf{SC1}_i + a_5\,\mathsf{DENS}_i\,, \tag{12}$$

where LLTI is the age-standardised measure of LLTI described above; UNEM is the proportion of economically active males and females who are unemployed (the denominator in this variable does not include those with LLTI who are not classed as being

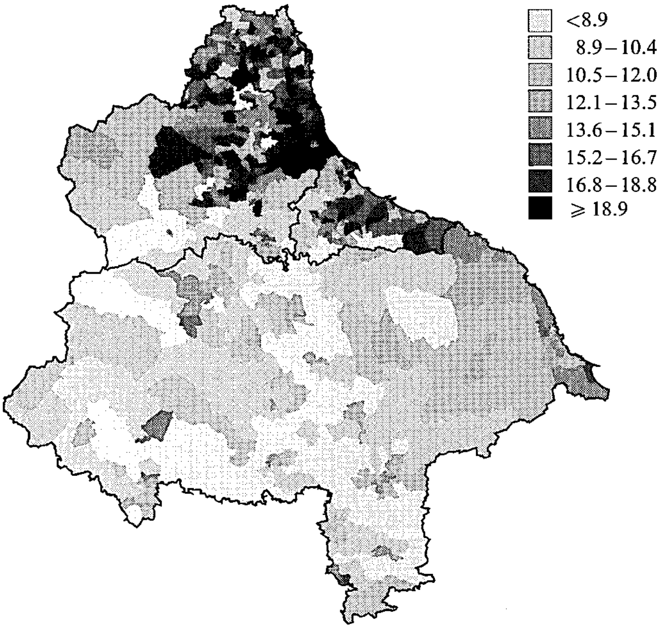

Figure 3. Spatial distribution of standardised limiting long-term illness variable.

economically active); CROW is the proportion of households whose inhabitants are living at a density of over 1 person per room; SPF is the proportion of households with single parents and children under 5; SC1 is the proportion of residents living in households with the head of household in social class I (employed in professional nonmanagerial occupations); and DENS is the density of population in millions per square kilometre. This last variable discriminates particularly well between urban and rural areas. The model is guided by the findings of Rees (1995) in his examination of LLTI at a much coarser spatial resolution (English and Welsh counties and Scottish regions). The data are extracted from the 1991 UK Census of Population Local Base Statistics. The areal units used are census wards which contain on average approximately 200 households per ward. With these data, the calibrated form of the global model is:

$$\mathrm{LLTI}_i = \underset{(1.3)}{3.8} + \underset{(3.4)}{96.6}\mathrm{UNEM}_i + \underset{(3.9)}{31.1}\mathrm{CROW}_i - \underset{(2.3)}{3.5}\mathrm{SPF}_i - \underset{(4.1)}{22.5}\mathrm{SC1}_i - \underset{(2.5)}{5.6}\mathrm{DENS}_i \qquad (13)$$

where the numbers in parentheses represent t-statistics and the R^2 value associated with the regression is 0.76. The results suggest that across the study region LLTI is positively related to unemployment levels and crowding. The relationship with unemployment reflects perhaps that the incidence of LLTI is related to both social and employment conditions in that the areas of higher unemployment tend to be the poorer wards with declining heavy industries. The relationship with crowding suggests a link between LLTI and social conditions, with levels of LLTI being higher in areas with high levels of overcrowding. LLTI is negatively related to the proportion of professionally employed people in a ward and to population density. The negative relationship between LLTI and SC1 supports the hypothesis that LLTI is more prevalent in poorer areas which have fewer people in professional occupations. It also reflects the fact that industrial hazards which are a factor in the incidence of LLTI are less likely to occur to people in professional occupations. The negative relationship between LLTI and DENS is somewhat counterintuitive in that it suggests that LLTI is greater in less densely populated areas, ceteris paribus. The nature of this relationship is explored in greater detail below in the discussion of the GWR results. Only the single-parent family variable is not significant (at 95%) in the global model.

To this point, the empirical results and their interpretations are typical of those found in standard regression applications: parameter estimates are obtained which are assumed to describe relationships that are invariant across the study region. We now describe the application of both the expansion method and GWR to examine the validity of this assumption and explore in greater detail spatial variations in the relationships described above.

9.1 Expansion method results

Each of the six parameters in the global model given in equation (12) was expanded in terms of both linear and quadratic functions of space. The results are given in tables 1 and 2, below and page 1918, along with the parameter estimates from the global model.

The linear results are perhaps the easier to interpret (table 1). They suggest that significant spatial variation in three parameters exists: the unemployment parameter, the crowding parameter, and the density parameter all appear to become less positive in the eastern part of the region. The other parameters apparently exhibit no significant spatial variation. To show this more clearly, values of the coordinates (u_i, v_i) for each spatial unit (the 605 census wards) are input into the calibrated expansion equations for each parameter and the resulting locally varying parameter estimates are mapped. These spatial distributions for the linear expansion equations are shown below for the three parameters discussed above plus the intercept term.

The spatially varying intercept is shown in figure 4. It depicts a trend in which higher values of the constant are found in the northern part of the region suggesting that, once spatial variations in the five variables in the model have been accounted for,

Table 1. Linear expansion method (EM) results.

Variable	Global	Linear EM
Constant	3.8***	52
Constant_u		6.4×10^{-5}
Constant_v		-6.7×10^{-5}
UNEM	93***	430***
UNEM_u		-4.8×10^{-4}***
UNEM_v		-2.3×10^{-4}
CROW	31***	−220
CROW_u		1.7×10^{-4}
CROW_v		3.4×10^{-4}
SC1	−23***	210
SC1_u		-2.8×10^{-4}***
SC1_v		-2.2×10^{-4}
DENS	−5.6***	140
DENS_u		-1.6×10^{-4}***
DENS_v		-1.3×10^{-4}
SPF	−3.5	55
SPF_u		-3.8×10^{-5}
SPF_v		-8.3×10^{-5}
R^2	0.76	0.80
Degrees of freedom	589	577

*** Significant at the 95% confidence level

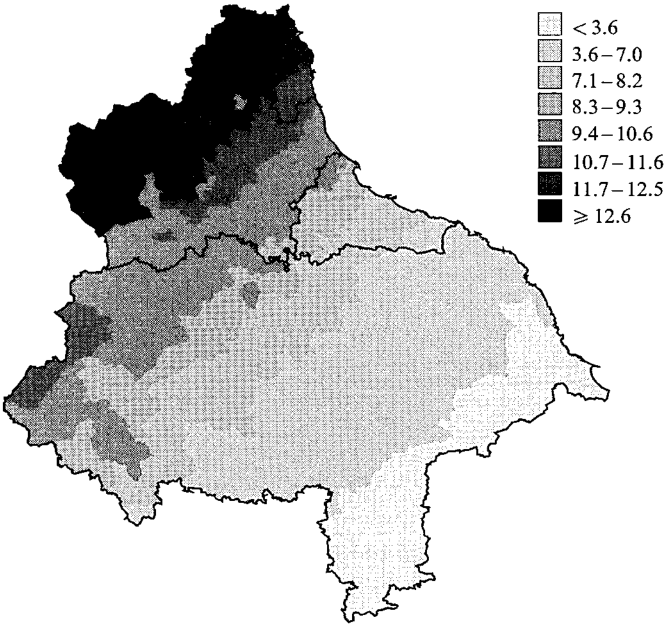

Figure 4. Linear expansion of the constant.

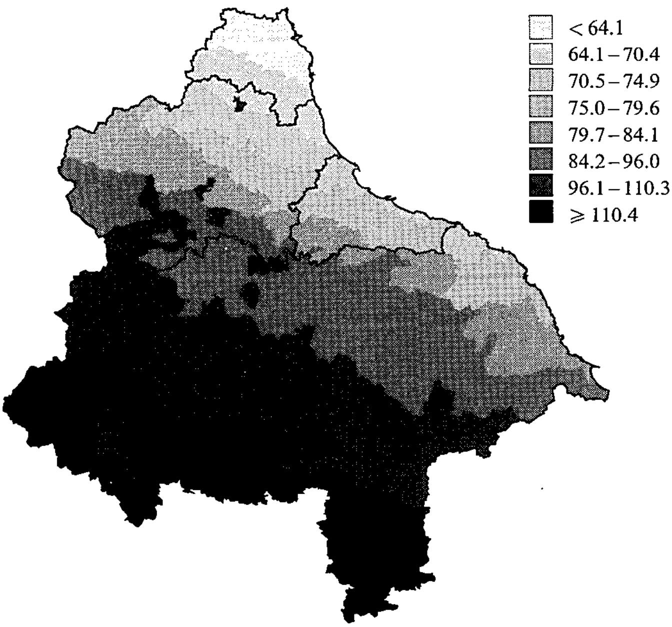

Figure 5. Linear expansion of the unemployment parameter.

standardised rates of LLTI still appear to be higher in the northern part of the region than in the south. The linear trend in the unemployment parameter estimate is shown in figure 5 in which it can be seen that the estimate is larger in the south than in the north, suggesting that LLTI rates are more sensitive to unemployment variations in the south. The linear trends in both the social class parameter in figure 6 and the density

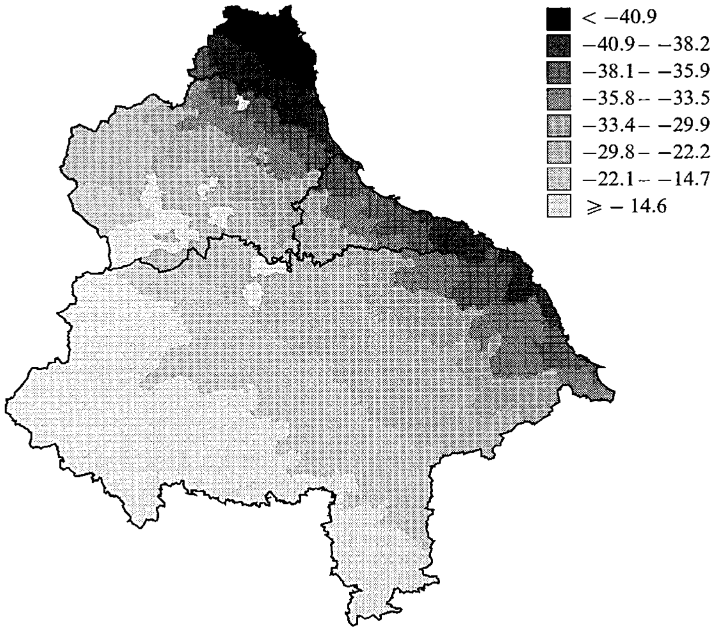

Figure 6. Linear expansion of the social class parameter.

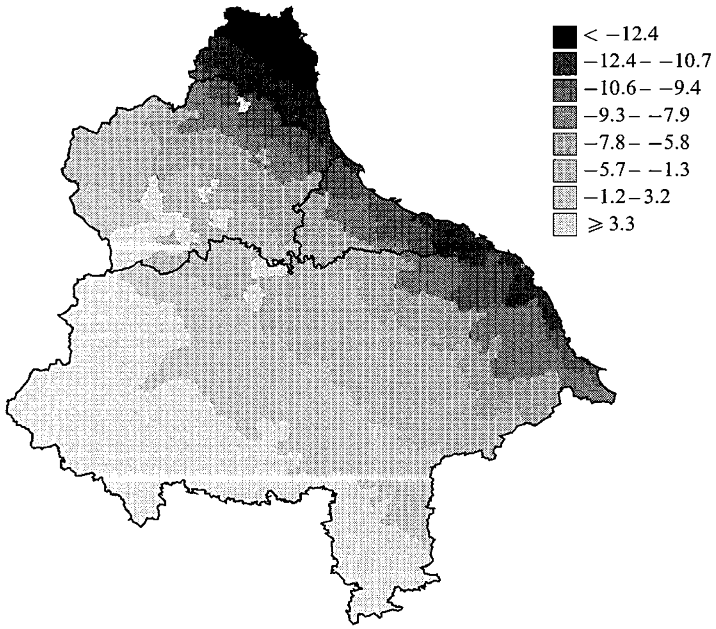

Figure 7. Linear expansion of the density parameter.

parameter in figure 7 suggest that the relationships between LLTI and social class and between LLTI and housing density both become more negative towards the coast. Of course, these are very simple trends which need to be explored in more detail.

The quadratic expansion results are depicted in table 2 and the spatial distributions of the intercept, unemployment, social class, and density parameters, equivalent to

Table 2. Quadratic expansion method (EM) results.

Variable	Global	Quadratic EM
Constant	38***	480
Constant_u		-2.0×10^{-3}
Constant_{uu}		1.6×10^{-9}
Constant_v		-7.3×10^{-5}
Constant_{vv}		1.2×10^{-9}
Constant_{uv}		-6.2×10^{-10}
UNEM	93***	−8400***
UNEM_u		2.0×10^{-3} ***
UNEM_{uu}		-1.1×10^{-8}
UNEM_v		1.7×10^{-2}
UNEM_{vv}		-8.3×10^{-9}
UNEM_{uv}		-2.2×10^{-8}
CROW	31	2400
CROW_u		-6.8×10^{-3}
CROW_{uu}		1.4×10^{-9}
CROW_v		-3.4×10^{-3}
CROW_{vv}		-2.7×10^{-9}
CROW_{uv}		1.2×10^{-8}
SC1	−23	−4300
SC1_u		1.9×10^{-2} ***
SC1_{uu}		-6.2×10^{-9} ***
SC1_v		-6.6×10^{-4}
SC1_{vv}		1.7×10^{-8} ***
SC1_{uv}		-3.0×10^{-8} ***
DENS	−5.6***	−110
DENS_u		1.7×10^{-3}
DENS_{uu}		-9.8×10^{-10}
DENS_v		-1.3×10^{-3}
DENS_{vv}		2.8×10^{-9}
DENS_{uv}		-2.1×10^{-9}
SPF	−3.5	−250
SPF_u		3.0×10^{-3}
SPF_{uu}		-2.0×10^{-9}
SPF_v		-2.2×10^{-3}
SPF_{vv}		3.9×10^{-9}
SPF_{uv}		-2.5×10^{-9}
R^2	0.76	0.83
Degrees of freedom	589	559

***Significant at the 95% confidence level.

figures 4–7, are shown in figures 8–11, respectively. Because of the increased complexity of the parameter surfaces being estimated, it is more difficult to interpret particular parameter estimates in table 2. However, two parameters, unemployment and social class, appear to exhibit significant spatial variation. The distribution of the expanded intercept shown in figure 8 is puzzling being the opposite of that of the linear trend and having positive values in the southern part of the region and negative values in the northern part. Presumably, this unexpected result is caused by the complexities of the other parameters varying spatially in sometimes complicated ways although this highlights a problem of the expansion method in that more complex expansions often produce parameter estimates which can be difficult to interpret.

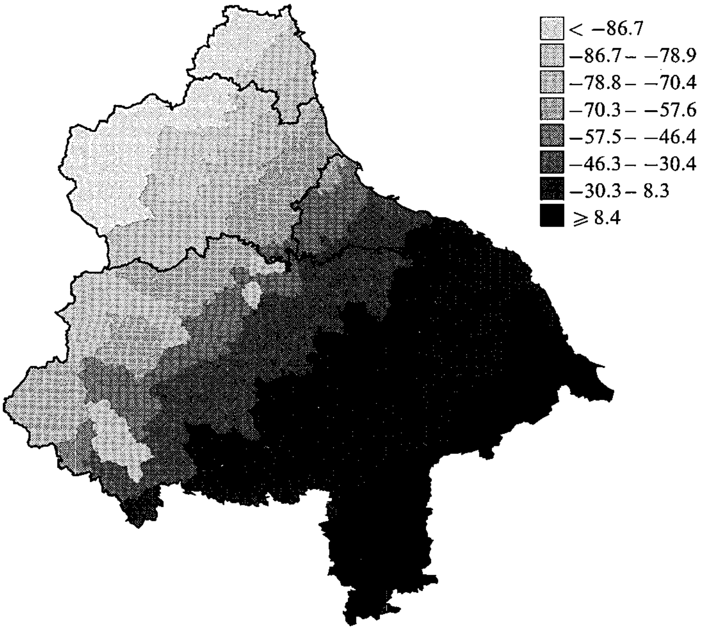

Figure 8. Quadratic expansion of the constant.

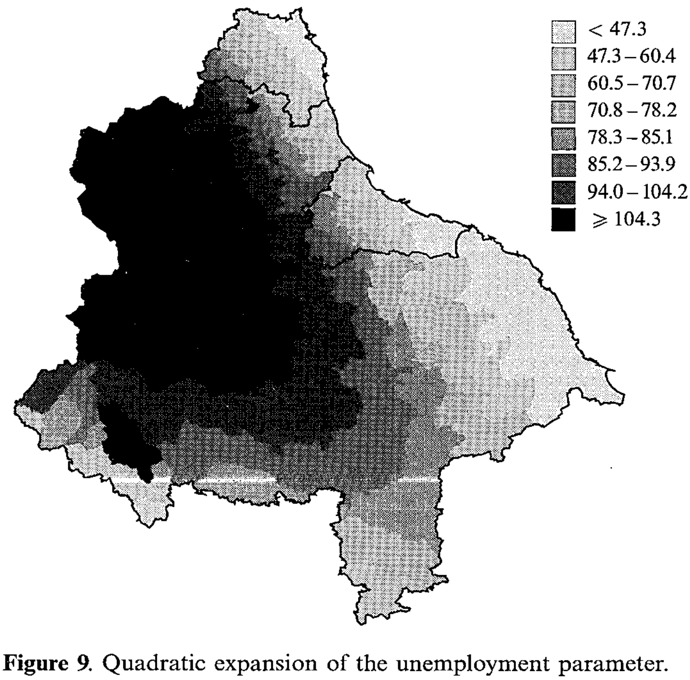

Figure 9. Quadratic expansion of the unemployment parameter.

The result of the quadratic expansion of the unemployment parameter shown in figure 9 shows a more complex pattern than that of the linear expansion with a ridge of higher values running from the northwest to the southeast. Similarly, the spatial distribution of the social class parameter in figure 10 suggests greater sensitivity in the relationship between LLTI and social class in areas towards the northeast coast

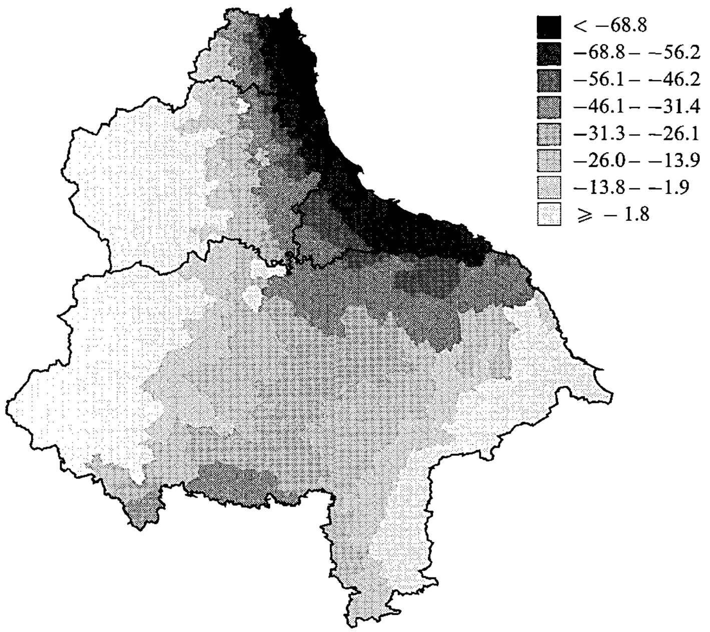

Figure 10. Quadratic expansion of the social class parameter.

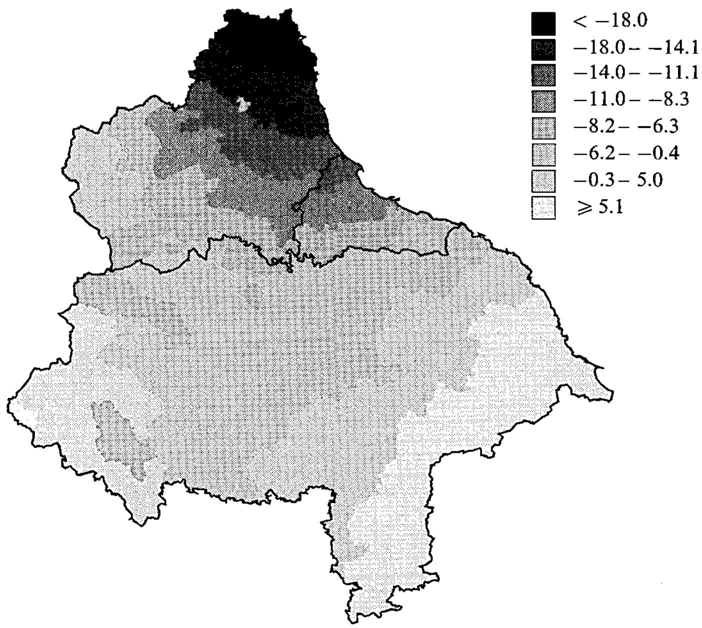

Figure 11. Quadratic expansion of the density parameter.

and running through a saddle point towards the southwest. The spatial distribution of the quadratic expansion of the density parameter in figure 11 now suggests a north–south trend in the relationship between LLTI and density instead of the northeast-to-southwest trend suggested by the linear expansion.

Whereas the quadratic expansions tend to highlight the simplistic nature of the linear expansion results, the quadratic results themselves are potentially simplistic representations of complex spatial patterns. In order to examine this complexity in more detail, it is necessary to move away from the expansion method and to GWR.

9.2 Geographically weighted regression results

Prior to calibration of the LLTI model by GWR, the Gaussian weighting function in equation (9) was calibrated by the cross-validation technique described in equation (11) and the estimated value of $\beta = 13.5$ resulted in a weighting function that tends to zero at a distance of approximately 19 km from a point at which parameter estimates are obtained (for a Gaussian kernel, data at a distance of greater than twice the bandwidth are virtually zero weighted). A plot of the CV scores against bandwidth is shown in figure 12.

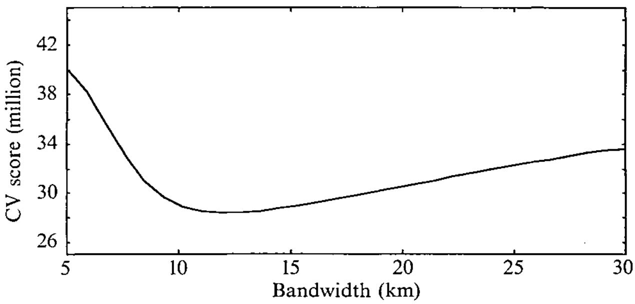

Figure 12. Cross-validation (CV) score as a function of bandwidth.

Each local parameter estimate is then obtained by weighting the data according to this function around each point and using the estimator in equation (6). The spatial distributions of the localised estimates of the intercept, unemployment, social class, and density parameters are shown in figures 13 – 16, respectively. The interpretation of each of the spatial estimates depicted in these figures is that it reflects a particular relationship, ceteris paribus, *in the vicinity of that point in space.*

Figure 13 (see over) shows the spatial variation in the estimated constant term obtained from GWR and it clearly exhibits much greater detail than either of the intercept maps derived from the expansion method. The estimates in figure 13 show the extent of LLTI after the spatial variations in the explanatory variables have been taken into account. The high values which occur primarily in the industrial areas of Cleveland and east Durham suggest a raised incidence of LLTI in these areas even when the relatively high levels of unemployment and low levels of employment in professional occupations are accounted for. The Monte Carlo test of significance for the spatial variation in these estimates described above indicates that the spatial variation is significant. This test might form a useful basis for testing for model specification. Presumably there are other attributes that might be added to the model that would reduce the spatial variation in the constant term. One could be satisfied with the specification of the model once the spatial variation in the constant term fails to be significant. The map in figure 13 also acts as a useful guide as to what these attributes should be. The model apparently still does not adequately account for the raised incidence of LLTI in the mainly industrial areas of the northeast and perhaps other employment or social factors would account for this.

The spatial variation in the unemployment parameter shown in figure 14 (see over) depicts the differing effects of unemployment on LLTI across the study area. All the

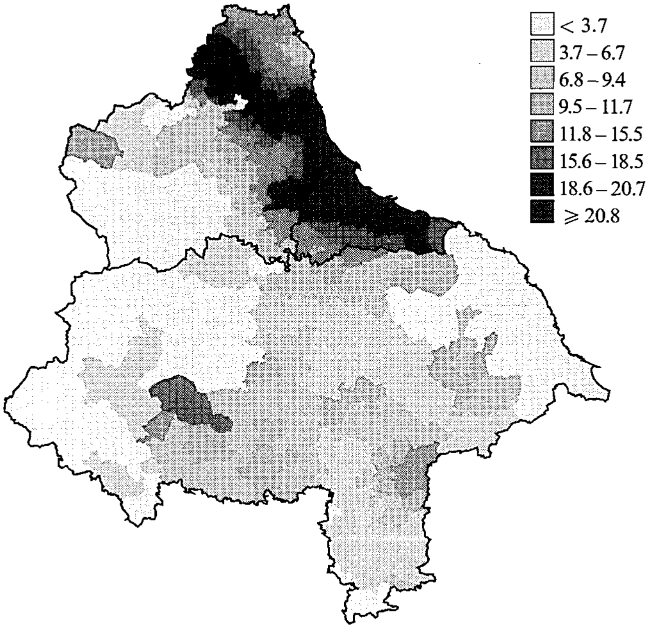

Figure 13. Geographically weighted regression distribution of the constant.

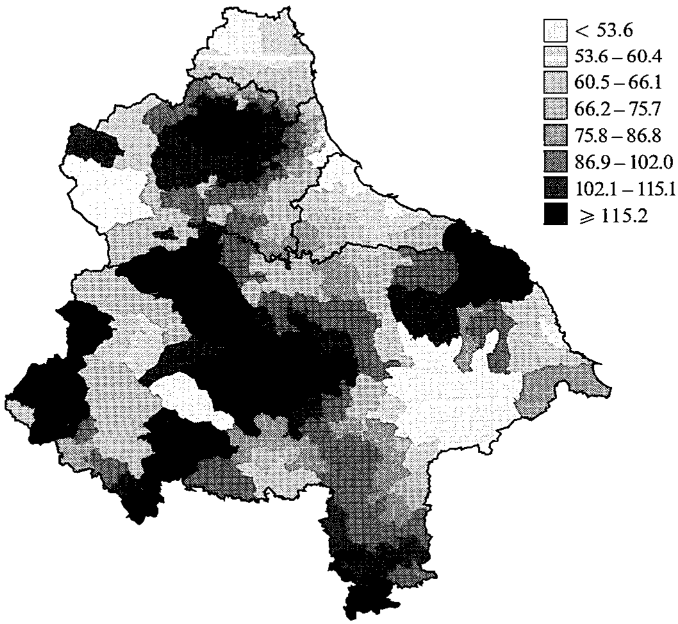

Figure 14. Geographically weighted regression distribution of unemployment parameter.

parameters are significantly positive but are smaller in magnitude in the urbanised wards centred on Cleveland and Tyne and Wear. Again, the spatial variation in these parameter estimates is significant. The results suggest a possible link to environmental causes of LLTI, with levels of LLTI being high regardless of employment status in Cleveland which has large concentrations of chemical processing plants and, until recently, steelworks.

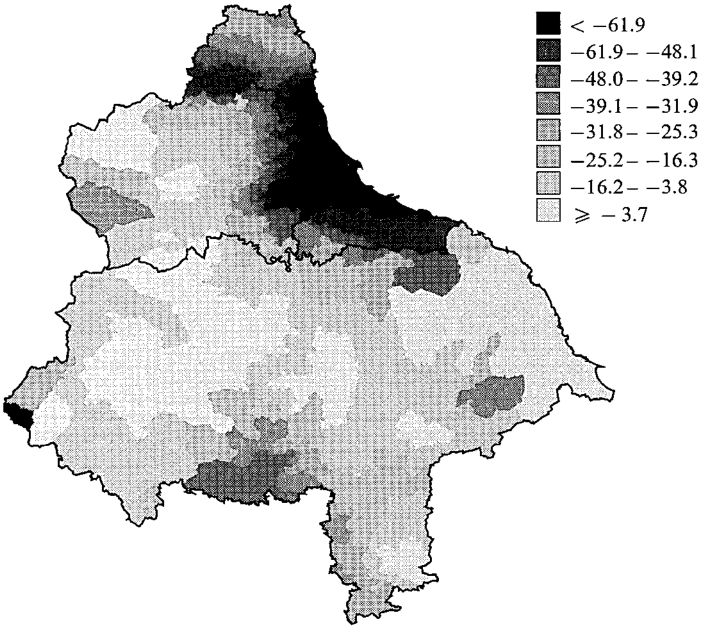

Figure 15. Geographically weighted regression distribution of social class parameter.

Another possibility is that levels of LLTI are high regardless of employment status in these areas because employment is concentrated in declining heavy industries and a large proportion of the unemployed were probably formerly employed in such industries which are associated with high levels of LLTI.

The global estimate of the social class 1 variable is significantly negative and all the spatial estimates, shown in figure 15, are negative and exhibit significant spatial variation. The more negative estimates are concentrated along the industrial parts of Cleveland, east Durham, and Tyne and Wear indicating that levels of LLTI are more sensitive to variations in social class in urban areas than in rural areas. Within urban areas LLTI is presumably linked to blue-collar occupations whereas in rural areas, the incidence of LLTI is more evenly distributed across types of employment.

Perhaps the best example of the use of GWR is provided in the spatial pattern of the estimates for the density variable given in figure 16. The global estimate for population density is significantly negative which is somewhat counterintuitive: we might expect that LLTI would be higher in more densely populated urban wards than in sparsely populated rural wards, ceteris paribus. The spatial variation of this parameter estimate indicates that the most negative parameter estimates are those for wards centred on the coalfields of Eastern Durham. The probable explanation for this is that LLTI is closely linked to employment in coal mining (pneumoconiosis and other respiratory diseases being particularly prevalent in miners) but that settlements based on coal mining do not have particularly high densities of population in this area—the area is characterised by many small pit villages. However, population density rises rapidly in the urbanised areas both immediately south and north of the coalfields where employment is less prone to LLTI. Hence, within the locality of east Durham it is clear that the relationship between LLTI and density is significantly negative. In the more rural parts of the study area, particularly in west Durham and North Yorkshire, the relationship is positive with t-values in excess of 2 in many places. Hence the more intuitive relationship, where LLTI increases in more densely populated areas, does exist in much of the study region but this information is completely hidden in the global

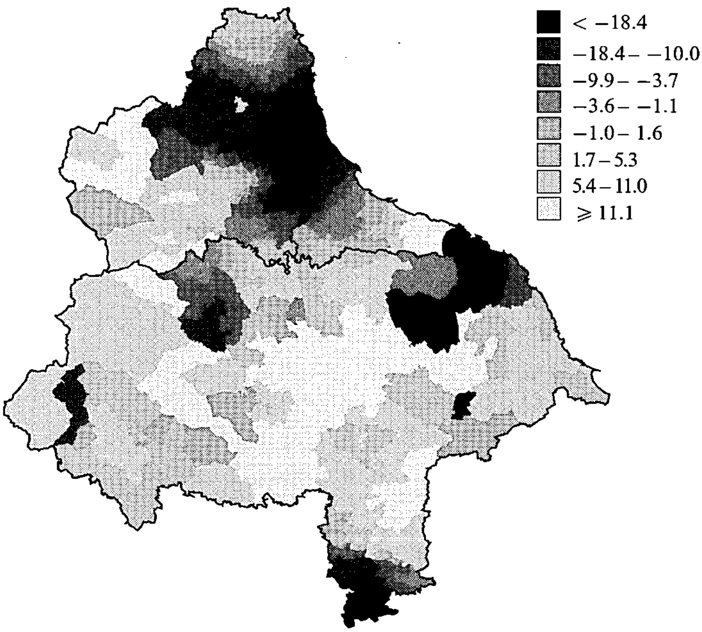

Figure 16. Geographically weighted regression distribution of density parameter.

estimation of the model and is only seen through GWR. The different relationships between LLTI and population density that exist across the region and which are depicted in figure 16 highlight the value of GWR as an analytical tool.

Although they are not reported here because of space limitations, it is quite easy to produce maps of t-values for each parameter estimate from GWR. These maps depict each spatially weighted parameter estimate divided by its spatially weighted standard error. Generally, the patterns in these maps are very similar to those depicted in the maps of the parameter estimates but occasionally some minor differences can occur because of spatial variations in standard errors which are based on data points weighted by their proximity to each point for which the model is calibrated. The statistics are useful, however, for assessing variations in the strengths of relationships across space.

One further spatial distribution from the GWR analysis is that of the spatially varying goodness-of-fit statistic, R^2, shown in figure 17. These values depict the accuracy with which the model replicates the observed values of LLTI in the vicinity of the point for which the model is calibrated. The global value of this goodness-of-fit statistic is 0.75 but it can be seen that there are large variations in the performance of the model across space ranging from a low of 0.23 to a high of 0.99 for the local model. In particular, the model explains observed values of LLTI well in a large group of wards in south Cleveland and the northern extremity of North Yorkshire and also in a group of wards in the southern and westerly extremes of the study region. The model appears to replicate the observed values of LLTI less well in parts of North Yorkshire and parts of Durham. The distribution of R^2 values in figure 17 can also be used to develop the model framework if the areas of poorer replication suggest the addition of a variable that is well represented in such areas and less well represented in areas where the model already works well. For instance, there is evidently a coalfield effect missing from the model and the low values of R^2 in North Yorkshire suggest the model still fails to account adequately for rural variations in LLTI.

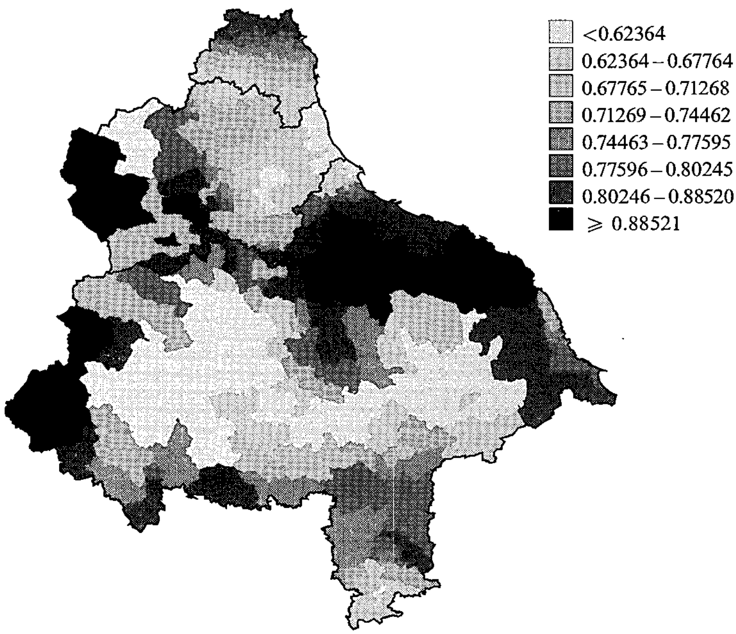

Figure 17. Geographically weighted regression R^2 distribution.

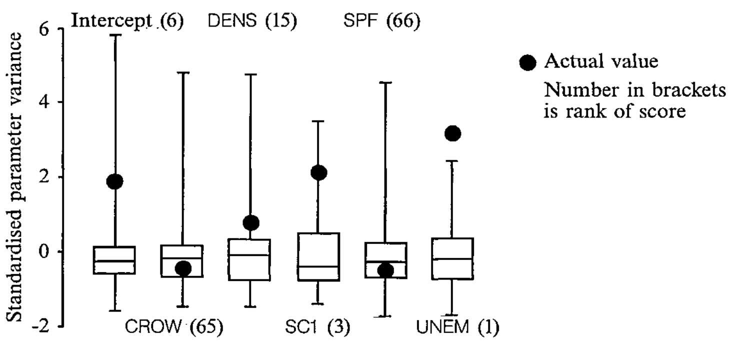

Figure 18. Results of Monte Carlo simulation in box-plot form.

Finally, it is possible to comment on the spatial variability of each parameter estimate from the results from the Monte Carlo procedure. Here, 99 random mixings of the data were undertaken and the GWR procedure was run for each data set. Thus, for each parameter, there are 100 estimates: 1 from the actual spatial arrangement of the data and 99 from the random arrangements of data. The variability of each set of parameter estimates is shown in the box-and-whisker plots of figure 18. The number in brackets for each parameter is the ranking of the spatial variability of each parameter estimate within the 100 sets of estimates with a ranking of 1 indicating the greatest amount of spatial variability and a ranking of 100 the least amount of spatial variability. The parameter estimates obtained from the actual spatial arrangement of the data are shown as dots on the plots. It can be seen that those sets of parameter estimates which exhibit large amounts of spatial variability from the original data are those for the intercept, social class 1 and unemployment with that for density being of marginal significance. The parameter estimates for crowding and single-parent families do not exhibit any significant spatial variation.

10 Discussion

Both the expansion method and GWR are analytical techniques that can be used to produce local or 'mappable' statistics from a regression framework. As such, they are both responses to calls such as those of Fotheringham (1992; 1994), Fotheringham and Rogerson (1993), and Openshaw (1993) for a move away from 'whole-map' or global statistics which merely present averages across space and which therefore discard large amounts of potentially interesting information on spatial variations of relationships and model performance. The output from GWR, as shown in an empirical example using the spatial distribution of limiting long-term illness in four counties in the United Kingdom, is much more realistic than that of the expansion method with the latter showing very simple trends in parameter variation over space. GWR produces localised parameter estimates which can exhibit a high degree of variability over space and demonstrate highly complex spatial patterns. These patterns inform on the spatial nature of relationships and on the spatial consequences of modelling such relationships which can be used in an exploratory mode to assist in model building. The spatial distributions of the regression constant and the goodness-of-fit statistic are seen as particularly important in this context.

The concept of measuring spatial variation in parameter estimates within a GWR framework lends itself to some obvious developments beyond those described above. It would be interesting and useful for example to show what, if any, relationships exist between GWR results and those from more traditional regression diagnostics such as the battery of leverage statistics available. Perhaps the most obvious development of the GWR is to experiment with various ways of making the spatial weighting function adapt to local environments. As defined above, the weighting function is a global one. Local versions could be produced with the aim of producing larger kernels in more sparsely populated rural areas and smaller kernels in more densely populated urban areas. Also it would be possible to examine whether different sizes of kernels were appropriate for different parameters and if so, it would be interesting to speculate on what this meant in terms of different spatial processes operating at different spatial scales. Another avenue of research that is being pursued is to examine the statistical estimation of 'mixed' models in which some parameter estimates are allowed to vary spatially and others remain fixed over space. One could imagine, for example, a hedonic house-price model in which all parameters were fixed except for the constant which is allowed to vary spatially. In this way, housing attributes might have a constant effect on house price and the geographical components of house-price variation are exhibited in the spatially varying constant. Alternatively, certain house-price determinants might vary spatially in their effect on house prices: the determination of which variables have a fixed effect and which have a spatially varying effect will probably be a matter for empirical investigation in most cases.

It is felt that GWR provides a significant advance in spatial analysis and that global regression applications with spatial data will be seen as rather limited. The application of GWR is not difficult and code will soon be available from the authors. It is hoped that GWR will promote interest in more genuinely geographical approaches to spatial analysis and can be used as a diagnostic to improved spatial understanding or as a means of allowing unknown spatial effects to enter the regression framework.

References

Bowman A W, 1984, "An alternative method of cross-validation for the smoothing of density estimates" *Biometrika* **71** 353 – 360

Brunsdon C, Fotheringham A S, Charlton M E, 1998, "Spatial nonstationarity and autoregressive models" *Environment and Planning A* **30** 957 – 973

Casetti E, 1972, "Generating models by the expansion method: applications to geographic research" *Geographical Analysis* **4** 81 – 91

Casetti E, 1982, "Drift analysis of regression parameters: an application to the investigation of fertility development relations" *Modeling and Simulation* **13** 961 – 966

Casetti E, Jones III J P, 1983, "Regional shifts in the manufacturing productivity response to output growth: sunbelt versus snowbelt" *Urban Geography* **4** 286 – 301

Cleveland W S, 1979, "Robust locally weighted regression and smoothing scatterplots" *Journal of the American Statistical Association* **74** 829 – 836

Cleveland W S, Devlin S J, 1988, "Locally weighted regression: an approach to regression analysis by local fitting" *Journal of the American Statistical Association* **83** 596 – 610

Cliff A D, Ord J K, 1981 *Spatial Processes: Models and Applications* (Pion, London)

Eldridge J D, Jones III J P, 1991, "Warped space: a geography of distance decay" *Professional Geographer* **43** 500 – 511

Fotheringham A S, 1992, "Exploratory spatial data analysis and GIS" *Environment and Planning A* **25** 156 – 158

Fotheringham A S, 1994, "On the future of spatial analysis: the role of GIS" *Environment and Planning A* **25** S30 – S34

Fotheringham A S, Pitts T C, 1995, "Directional variation in distance decay" *Environment and Planning A* **27** 715 – 729

Fotheringham A S, Rogerson, P A, 1993, "GIS and spatial analytical problems" *International Journal of Geographic Information Systems* **7** 3 – 19

Fotheringham A S, Brunsdon C, Charlton M, 1998, "Scale issues and GWR" *Scale Issues and GIS* Ed. N Tate (John Wiley, Chichester, Sussex) in the press

Fotheringham A S, Charlton M, Brunsdon C F, 1996, "The geography of parameter space: an investigation into spatial non-stationarity" *International Journal of Geographical Information Systems* **10** 605 – 627

Fotheringham A S, Charlton M, Brunsdon C F, 1997, "Two techniques for exploring non-stationarity in geographical data" *Geographical Systems* **4** 59 – 82

Goldstein H, 1987 *Multilevel Models in Educational and Social Research* (Oxford University Press, Oxford)

Greig D M, 1980 *Optimisation* (Longman, Harlow, Essex)

Hastie T, Tibshirani R, 1990 *Generalized Additive Models* (Chapman & Hall, London)

Jones III J P, Casetti E, 1992 *Applications of the Expansion Method* (Routledge, London)

Kendall M, Ord J K, 1990 *Time Series* (Edward Arnold, London)

Openshaw S, 1993, "Exploratory space-time attribute pattern analysers", in *Spatial Analysis and GIS* Eds A S Fotheringham, P A Rogerson (Taylor and Francis, London) pp 147 – 163

Rees P, 1995, "Putting the census on the researcher's desk" *The Census Users' Handbook* Ed. S Openshaw (GeoInformation International, Cambridge) pp 27 – 81

Rosenberg B, 1973, "A survey of stochastic parameter regression" *Annals of Economic and Social Measurement* **1** 381 – 397

Spjotvoll E, 1977, "Random coefficients regression models, a review" *Mathematische Operationsforschung und Statistik* **8** 69 – 93