The Statistician (1998)
47, *Part* 3, *pp.* 431–443

Geographically weighted regression—modelling spatial non-stationarity

Chris Brunsdon†, Stewart Fotheringham and Martin Charlton

University of Newcastle, UK

[Received July 1996. Revised January 1998]

Summary. In regression models where the cases are geographical locations, sometimes regression coefficients do not remain fixed over space. A technique for exploring this phenomenon, *geographically weighted regression* is introduced. A related Monte Carlo significance test for spatial non-stationarity is also considered. Finally, an example of the method is given, using limiting long-term illness data from the 1991 UK census.

Keywords: Exploratory data analysis; Geographically weighted regression; Monte Carlo testing; Regression; Spatial non-stationarity

1. Introduction

The technique of linear regression has long resided in the analytical toolbox of the quantitative geographer. As a general technique for investigating the linkage between geographical variables, the method has featured in virtually countless publications. However, this is surprising in some respects, as the technique itself takes no account of location in its analysis of relationships between variables. The well-known components of a regression model are X a matrix containing a set of independent or predictor variables and y a vector of dependent or response variables. The relationship between these is modelled as

$$y_i = \sum_j X_{ij}\beta_j + \epsilon_i \tag{1}$$

where subscripts denote selecting individual elements of vectors or matrices, β is a vector of regression coefficients and ϵ is a random vector whose distribution is $N(0, \sigma^2 I)$. Some well-known theory gives, as a maximum likelihood estimate of β,

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y. \tag{2}$$

The understanding when this method is applied to geographical data is that each case corresponds to a geographical location, but beyond this implicit agreement space plays no role in the modelling process.

However, there may be situations when the nature of such models is not fixed over space. This is referred to here as spatial non-stationarity. Several more explicit ways of incorporating space have been considered. For example Cassetti (1972) proposed the *expansion method* where coefficients in the regression model were expressed as explicit functions of the spatial locations of the cases:

†*Address for correspondence*: Department of Town and Country Planning, Claremont Tower, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK.
E-mail: chris.brunsdon@newcastle.ac.uk

 0039–0526/98/47431

$$y_i = \sum_j X_{ij}\beta_j(p_i) + \epsilon_i. \quad (3)$$

Here p_i is the geographical location of the ith case. These $\beta_j(p_i)$ would themselves contain coefficients to be estimated. These expressions would be substituted into the original model, and the resulting expanded model—which may now be non-linear—would be calibrated. The advantage of a technique of this sort is that, once the model has been calibrated, it is possible to map the variation in the original regression parameters, and to gain some understanding of the spatial patterns in the association between the predictor and response variables.

Another more basic approach has been to include *dummy variables* for broad classes of spatial location—e.g. to indicate in which county a case is situated, or whether the location could be broadly classed as urban or rural. Although this approach can be useful, unlike Cassetti's method this does not allow other coefficients in the model to vary spatially. Gorr and Olligschlaeger (1994) have used spatial adaptive filtering to model spatial variations in coefficients, but this technique does not have any means of statistically testing the validity of the assumption of variation in the parameters. Other attempts, such as those of Jones (1991), make use of multilevel modelling techniques (Goldstein, 1987). Holt *et al.* (1996) modelled effects of aggregation and the 'ecological fallacy'—a phenomenon which occurs when aggregated variables for some area are used to represent individuals (or smaller areas) within that area. Here, within-area correlations observed in the dependent variable are at least partially explained by within-area correlations of auxiliary variables. Using this information, geographical variations in the regression model are considered in terms of local estimation bias.

A major difficulty with these approaches is their reliance on a predefined spatial hierarchy, such as the grouping of census wards into census districts. No work has been done to investigate the sensitivity of these models to changes in the hierarchical groupings of the zones.

Here, we attempt to build on the idea of Cassetti (1972). In the expansion method one is required to provide an explicit function to describe the variation of each of the elements of β over space. This can work well in some situations, e.g. when the study area is the core and periphery of a major city and some function of distance from the city centre may be intuitively specified, but at other times the specification of such functions is less intuitive. For example, when considering the northern region of England we encounter many villages, towns and cities having different characteristics—some are business centres, some rural market towns and so forth—and no obvious functional form for $\beta_j(p_i)$ suggests itself. Such a function could clearly have many peaks and troughs, and would most likely be highly non-linear, leading to many difficulties in the calibration of the expanded regression model. The method proposed here, *geographically weighted regression* (GWR), attempts to overcome this problem by providing a nonparametric estimate of $\beta_j(p_i)$.

In this paper we shall outline the methodology required for GWR and consider some associated issues, in particular methods of testing whether there is statistically significant spatial variation in the models, and methods of computation. In addition to this, an example of the technique will be given, involving data from the UK census relating to the incidence of limiting long-term illness and its tentative causes.

2. Methodology

The task of this section is to provide a brief outline of the GWR technique, but before doing this it is helpful to discuss some underlying ideas. Essentially the problem here is to provide estimates of $\beta_j(p_i)$, for each variable j and each geographical location i (for notational convenience, this

expression will subsequently be abbreviated to β_{ij}). This is achieved by considering data for places near location p_i. For example, if we drew a circle of some radius, say r, around one particular p_i, and calibrated an ordinary least squares regression model only on the basis of observations whose geographical location was within this circle, then the β_j obtained could be thought of as an estimate of the associations between the variables in and around p_j. In short, these are estimates of β_{ij}. By evaluating β_{ij} for each p_i it is possible to obtain a set of estimates of spatially varying parameters *without* specifying a functional form for the spatial variation. In a sense, this technique 'lets the data speak for themselves' when providing estimates of each β_{ij}. With some modifications, this is the underlying concept of GWR.

2.1. Kernel-weighted regression

An initial consideration of the above technique may raise issues relating to the notion of the 'circle of inclusion' of observations around each p_i. This circle has been specified to have radius r, but what value should r take? If r is very large, then data included in each β_{ij}-estimate will include most of the entire study area, so estimates for each case i will become very close to those obtained by applying the 'global' model of equation (1). If it is very small, few observations will be included in the calibration, which will lead to estimates of the β_{ij} with very large standard errors. Another issue which may be raised is the binary nature of inclusion of observations in the regression model calibration. An observation whose distance from p_i falls just below r will be included in the model, whereas an observation whose distance just exceeds this quantity will be excluded. It seems unnatural that the spatial association between the variables ends so abruptly.

We shall address the latter of these issues initially, and the former in Section 2.4. The regression model centred at each p_i could be thought of as a weighted ordinary least squares regression, with observations in the circle of inclusion weighted as 1 and other observations weighted as 0. Thus, for a given p_i, the weight α_{ik} given to observation k would be

$$\alpha_{ik} = \begin{cases} 1 & \text{if } d_{ik} < r, \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where d_{ik} is the distance between the locations of observations i and k. However, there is no reason to restrict the weighting function to a step function in this way. It is also possible to relate d_{ik} to α_{ik} with a continuous function. For example, a Gaussian distance-decay-based weighting would be achieved by

$$\alpha_{ik} = \exp(-d_{ik}^2/2h^2).$$

Here, the value of the weight would decay gradually with distance, to the extent that when $d_{ik} = h$ the weighting would be 0.05. Alternative weighting functions could be

$$\alpha_{ik} = \exp(-d_{ik}/h)$$

or

$$\alpha_{ik} = \begin{cases} \{1 - (d_{ik}/h)^2\}^2 & \text{if } d_{ik} < r, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

These functions will be referred to as *kernel functions* or *kernels* and denoted by the letter K as in $\alpha_{ik} = K(d_{ik})$. In each case, the constant h provides some control of the range of the 'circle of influence' of the geographical data, as r does in the basic step function example, but the degree of weighting decays with distance rather than suddenly dropping to 0 when a certain distance is reached. Generally, desirable features of a kernel function K are

(a) $K(0) = 1$,
(b) $\lim_{d\to\infty}\{K(d)\} = 0$ and
(c) K is a monotone decreasing function for positive real numbers.

It is also interesting that, although we have restricted our attention to estimating β_{ij}-values at points in space corresponding to the geographical locations of the observations, the same methodology can be applied to *any* point in space. If this is done using any of the three alternative weighting schemes suggested above, the β-estimates will be continuous functions on $\mathscr{R}^2$. This would not be the case with the initial crude estimator, however.

2.2. Calibration

Thus, having chosen a weighting function, equation (2) can be modified to give

$$\hat{\beta}_i = (X^{\mathrm{T}} W_i X)^{-1} X^{\mathrm{T}} W_i y \tag{6}$$

where

$$W_i = \begin{pmatrix} \alpha_{i1} & 0 & \dots & 0 \\ 0 & \alpha_{i2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_{iN} \end{pmatrix} \tag{7}$$

and N is the number of observations. Here the weight matrix W_i is a diagonal matrix, whose diagonal elements correspond to the weights when calibrating a weighted regression around point p_i. Thus expression (6) is not a single equation, but essentially an array of equations, with each β_i corresponding to a column of the matrix whose elements are β_{ij}. In fact, this matrix is the key output of GWR. By selecting individual rows, it is possible to see how a coefficient corresponding to a given predictor variable changes geographically. In addition to this, it is possible to compute the standard error of each coefficient estimate. This may be particularly helpful for investigating the precision of these spatially varying estimates. In particular any 'edge effects' could be investigated in this way.

Once each of the α_{ik} has been computed, the β-matrix can be computed column by column by repeated application of expression (6) for each i. It is worth noting the similarities between this approach and the technique of kernel regression as discussed in Wand and Jones (1995). In that approach, a similar weighting matrix is constructed. However, in that instance, weights were based on similarities between the predictor variables themselves, rather than their geographical location. This provided a nonparametric approach to regression modelling.

2.3. Computational issues

As demonstrated in Section 2.2, one way of generating β_{ij} for all i- and j-values is to apply the matrix formula (6) N times. Clearly, if N is very large, there will be a large computational cost in doing this. For this reason, methods of speeding up the computation should be considered. One of the simplest ways of doing this is to adopt a weighting scheme such as

$$\alpha_{ik} = \begin{cases} \{1 - (d_{ik}/h)^2\}^2 & \text{if } d_{ik} < r, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

noting that if d_{ik} exceeds h then observation k is zero weighted, and therefore effectively excluded from the calculation. This is particularly important in terms of computation. If the number of observations for which weighting is non-zero is significantly smaller than N, then the computa-

tional overhead of expression (6) can be reduced by treating the problem as though it were only applied to the non-zero weighted observation. There are parallels between this approach and that suggested by Silverman (1986) to reduce computational overheads for probability density estimates.

In very large data sets, a further efficiency gain might be achieved by using some form of spatial indexing, such as tesseral addressing, to organize the data. Techniques of this sort are designed to facilitate faster retrieval of information from a spatially referenced database when the querying of the database is itself spatial. In this way, observations which were non-zero weighted —those within a given distance of some point—could be retrieved from the database more rapidly.

2.4. Choosing h

Another important issue in GWR is the choice of h—sometimes referred to as the *kernel bandwidth*. As stated earlier, this can greatly affect the properties of the β-matrix. Again following the advice of Silverman (1986) when considering kernel density estimates, there are occasions when a subjective choice lends itself well to the problem in hand. If we have strong theoretically based prior beliefs about the value of h in a given situation, then it is reasonable to make use of them.

However, there are many situations in which no such theoretical understandings exist, and in these cases some form of automatic data-led choice of h may be more appropriate. One method suggested here is that of *least squares cross-validation*. A common calibration technique is that of *least squares*. Suppose that, for a prespecified kernel function, the predicted value of y_i from GWR is denoted (as a function of h) by $\hat{y}_i(h)$. The sum of squared errors may then be written as

$$\mathrm{SS}(h) = \sum_i \{y_i - \hat{y}_i(h)\}^2. \tag{9}$$

A logical choice may then be to find h minimizing equation (9). However, at this stage a problem is encountered. As $h \to 0$, $\hat{y}_i(h) \to y_i$, so equation (9) is minimized when $h = 0$. To see why, note that, for all K functions, $\alpha_{ii} = 1$, and that, if $i \neq k$, then $h \to 0 \Rightarrow \alpha_{ik} \to 0$, so the weighted regression is dominated by the term for observation i. This suggests that an unmodified least squares automatic choice of h would always suggest $h = 0$, or possibly result in computational errors. This problem can be avoided if, for each i, a GWR estimate of y_i is obtained by *omitting* the ith observation from the model. This is equivalent to replacing the kernel function K by a modified function K^* such that

$$K^*(0) = 0,$$
$$K^*(d) = K(d) \qquad \text{if } d \neq 0.$$

If the modified GWR estimate of y_i is denoted by $\check{y}_i(h)$ then the *cross-validated* sum of squared errors is denoted by

$$\mathrm{CVSS}(h) = \sum_i \{y_i - \check{y}_i(h)\}^2. \tag{10}$$

Choosing h to minimize equation (10) provides a method for choosing h automatically that does not suffer from the problems encountered by working with equation (9).

2.5. Testing the geographically weighted regression hypothesis

At this stage, we have outlined a method for calibrating a nonparametric model of spatial drift in regression parameters, and we have outlined an automatic method for choosing the bandwidth

parameter h. However, we have yet to justify the use of a spatially varying regression model at all! In this section, we shall consider the development of a hypothesis test for

$$\begin{array}{lll} & H_0\colon \beta_{ij} = \beta_j, & \forall\, i, \\ \text{against} & H_1\colon \beta_{ij} \text{ not all the same} & \forall\, i. \end{array} \tag{11}$$

One useful statistic to measure the variability of β_{ij} as i varies for a fixed j is the variance of β_{ij} across i:

$$\upsilon_j = \sum_i (\beta_{ij} - \beta_{\cdot j})^2 / N \tag{12}$$

where a dot in a subscript position denotes averaging over that subscript.

Clearly, the lower the value of υ_j, the stronger the evidence that the coefficient corresponding to υ_j is fixed. For individual variables, hypothesis (11) could be tested if the null distribution of υ_j were known. Unfortunately this is not the case, but Monte Carlo techniques (Hope, 1968) offer an alternative approach. For this particular problem a randomization test may be carried out. Under hypothesis H_0 we assume that the β_{ij} do not vary with i for variable j. This suggests that, if the GWR model were to be calibrated with the locations of the observations randomly assigned to the predictor and response variables, there should be little difference in the patterns of β_{ij}—if the β_{ij} are fixed over space then spatial location should not greatly affect their calibration! Therefore, using Monte Carlo tests it should be possible to compare the distribution of the υ_j under the randomization hypothesis. The procedure, for a given j, is as follows.

(a) Make a note of υ_j for the correctly located observations.
(b) Randomly 'scramble' the locations p_i among the observations.
(c) Repeat the previous step $P - 1$ times, noting υ_j each time.
(d) Compute the rank of υ_j for the correctly located case, R.
(e) The p-value for the randomization hypothesis is R/P.

One method for reducing the computational time when carrying out this procedure is to compute υ_j for all j simultaneously—so that only P randomizations need to be computed, rather than PM, if M is the number of variables in the model. However, if this approach is taken, the hypothesis tests for each individual j will not be independent.

A related test is also possible, although computationally expensive. Assuming that the randomization process produces data from a population with little spatial pattern, the most suitable h-value would be ∞. As a test statistic for the overall model $s = 1/h$ may be appropriate. In the case of poor spatial pattern, this would take values close to 0. By carrying out a randomization test similar to that set out above, it is possible to compare simulated s-values, chosen to optimize equation (10), with those for the actual data. In this case the randomization step would require s rather than υ_j to be computed. However, this would require an optimization process to take place in the optimization process, so the overall computational requirements of this approach are expected to be very large.

2.6. Alternative methods for assessing the spatial variability of coefficients

Although the previous tests are useful, two points need to be made. Firstly, as they are Monte Carlo based, they will require relatively large amounts of computer time. The second point is perhaps more fundamental. There has been much recent discussion among statisticians about the use of significance tests (Chatfield, 1994). A major issue is that many significance tests are based on 'pinhead hypotheses' such as some parameter in a model being equal to 0. If a sufficiently large sample is taken, the hypothesis test will become sufficiently powerful to detect minute deviations

from this, such as the parameter being equal to 0.001. Thus a large sample is extremely likely to yield a highly significant result, but looking at the p-value on its own fails to identify the triviality of the effect that it has detected. It is suggested that perhaps more emphasis should be placed on the magnitude of effects, rather than on p-values associated with hypothesis tests. In particular, confidence intervals are perhaps more helpful than p-values, as they convey an idea not only of the magnitude of some parameter but also how precisely this has been determined.

These arguments could reasonably be levelled at GWR. It is more important to know the *degree* to which a coefficient varies over space than simply to test whether it does. Again, v_i-statistics (or perhaps their square roots which are in the same units as the regression coefficients) are useful here, as they measure spatial variability. Unfortunately, since little is known about their distributions, it is difficult to compute confidence intervals for these quantities. This could be overcome with repeated observations of the data set, so that repeated observations of v_j existed, and nonparametric confidence intervals such as those outlined by Maritz (1985) could be computed; however, as geographical data over wide areas are expensive to obtain, such situations do not often arise.

A workable alternative to this is to compare the variability of the β_{ij}-coefficients with the standard errors of the β_j-coefficients in the global regression model (1). This can be done for each variable by tabulating $\sqrt{v_j}$ against $\mathrm{SE}(\beta_j)$ from model (1).

3. Limiting long-term illness—an example

In this section, a practical example of GWR will be given. This will be based on data taken from the 1991 UK census (Office of Population Censuses and Surveys, 1991), for a large region consisting of several counties in the north of England (Fig. 1). A series of predictor variables will be used to model levels of limiting long-term illness measured for people 45–65 years old at the census ward level.

3.1. Background

In the 1991 census, a new question relating to 'limiting long-term illness' was introduced. Unlike questions in previous censuses, which related more specifically to disability, this covered a broader range of illnesses (Dale and Marsh, 1993). Notably this variable now included illnesses which could have resulted from exposure to certain work-related risks. Also, some illness which may have resulted from poor housing conditions or from stress-related conditions were also included. This variable, together with several socioeconomic variables that are also compiled from the census, now makes it possible to investigate the links between the incidence of limiting long-term illness and other social and economic phenomena quantitatively. In particular, applying GWR here allows us to investigate whether any relationships that exist are stable over space, or whether they change to reflect the characteristics of different localities in the study area. To control for different age distributions between census wards, only limiting long-term illness in people 45–65 years old will be considered.

3.2. Variables chosen

A set of predictor variables was chosen to represent various hypothesized contributory factors to prevalence of limiting long-term illness. The variables are listed below.

(a) LLTI is the percentage of individuals in households in each ward where a member of the household has some limiting long-term illness. This is the response variable. To control for

Fig. 1. Study area for the limiting long-term illness model: urban areas in the study area

different age profiles in areas, this is only computed for people 45–65 years old—an age category that is perhaps most likely to suffer limiting long-term illness as a result of working in the extractive industries.

(b) CROWDING is the proportion of households in each census ward having an average of more than one person per room. This is an attempt to measure the level of cramped housing conditions in each ward.

(c) DENSITY is the housing density of each ward, measured in millions per square kilometre. This is intended to measure 'ruralness' of areas. Note the differences between this and the previous variable—a remote village with poor housing conditions may score low in this variable, but high in the previous variable.

(d) UNEMP is the proportion of male unemployment in an area. This is generally regarded as a measure of economic well-being for an area.

(e) SC-1 is the proportion of heads of households whose jobs are classed in social class I

in the census. These are professional and managerial occupations. Whereas the previous variable measures general well-being, this measures affluence.

(f) SP-FAM is the proportion of single-parent families in an area. This is an attempt to measure the nature of household composition in areas.

3.3. Results of global regression

A table of correlations between the variables is shown in Table 1. The largest correlation is between limiting long-term illness and male unemployment—perhaps this is hardly surprising, as several studies have already highlighted linkages between deprivation and health problems. For example see Townsend *et al.* (1988).

Next, the regression model itself will be considered. Results for this are given in Table 2. From this it can be seen that, at least globally, every variable has a statistically significant coefficient except SP-FAM. The general fit of the model, as measured by the R^2-statistic, is good. Perhaps the most surprising results are the coefficients for DENSITY and CROWD, which are both negative. However, although it is possible that cramped housing and urban environments contribute to ill health, it is also the case that once people have become ill they may tend to move away from the worst environments. Also, considering the relationship demographically, many elderly people are likely to suffer from limiting long-term illness, but these people are less likely to live in crowded households, as any children they have are likely to have left home. They are also likely to retire to rural areas in some cases, and so this demographic factor is likely to influence the coefficients for the CROWD and DENSITY variables. Another possible explanation for these surprising results is the effect of omitting one or more influential variables from the model.

3.4. Results of geographically weighted regression

In this section, the results of calibrating a GWR model for the limiting long-term illness data are discussed. Firstly, spatial referencing must be considered. Data observations here correspond to census wards, the second smallest aggregational unit used in the 1991 census. There are

Table 1. Correlation matrix for the limiting long-term illness data

Variable	*DENSITY*	*LLTI*	*SC-1*	*SP-FAM*	*UNEMP*
CROWD	0.249	0.371	−0.501	0.382	0.557
DENSITY		0.271	−0.185	0.149	0.480
LLTI			−0.458	0.237	0.711
SC-1				−0.249	−0.438
SP-FAM					0.374

Table 2. Results of the global regression model†

Variable	*β-coefficient*	*SE(β)*	*Significance*
CONSTANT	13.52	0.841	$p < 0.01$
CROWD	−7.592	2.48	$p < 0.01$
DENSITY	−4.506	1.62	$p < 0.01$
SC-1	−17.40	2.634	$p < 0.01$
SP-FAM	−1.700	1.452	Not significant
UNEMP	46.07	2.411	$p < 0.01$

†$R^2 = 0.547$.

approximately 500 of these in the study area. To give a point location value to each observation (i.e. a p_i-value), the centroid of each ward was used. Next, the choice of h and the kernel is considered. Experience has shown that the method is relatively insensitive to the choice of the kernel function, provided that it is smooth and exhibits a distance decay property. In this case, a Gaussian kernel was used. After using Newton's method to minimize equation (10) with respect to h, a value of 17.1 km was chosen.

Using the Monte Carlo technique described earlier, the results of randomization tests on each of the coefficients is given in Table 3. From this it can be seen that the coefficients CONSTANT, DENSITY and CROWD vary significantly over space. However, bearing in mind the warnings about interpreting significance levels in isolation, the degree of variability for each coefficient is compared with the global standard error in two further columns in equation (3). From this it may be seen that in *all* cases $\sqrt{v_j}$ exceeds $\mathrm{SE}(\beta_j)$, although by less spectacular amounts in the cases of the non-significant variables. This suggests that there is some justification in considering the patterns of spatial variation in all coefficients in the model.

To give a brief indication of the nature of the variation, consider Fig. 2. This is particularly interesting, as the coefficient takes both positive and negative values. In particular, in an area to the north-east, there is a strong negative relationship (so that high density implies low limiting long-term illness) whereas in most other areas the relationship is positive. The latter is perhaps what one would intuitively expect to see—urban areas having a greater prevalence of long-term illness. But, we must consider that, in the Durham area, historically there is a strong tradition of employment in the coal-mining industry. Coal-mining communities exist typically in pit villages—which are relatively sparsely populated compared with urban areas. It is also well known that various illnesses are associated with working in coal-mines. In contrast, one city in this area is Durham City, and much employment within the city is associated with its university, cathedral, shops and tourism industry. Considerably fewer illnesses are associated with occupations of this sort! In the context of this locality a negative association between population density and limiting long-term illness is not surprising. However, when we look further south on the map to the county of North Yorkshire, the areas having lower housing density are associated with more traditional rural communities, and so the more usual positive association between housing density and illness is seen.

4. Conclusions

In this paper we have outlined a technique for producing local 'mappable' regression coefficients. This can be thought of as one response to Fotheringham's (1992, 1994) and Openshaw's (1993) calls for more geographically localized statistical techniques, and a more critical look at existing 'whole map' statistics. In many situations, the smoothing process of computing coefficients for a global regression can overlook interesting geographical features in the relationships between

Table 3. Results of GWR

Variable	$\sqrt{v}$	*SE(β)*	*Monte Carlo significance*
CONSTANT	3.82	0.841	$p = 0.03$
CROWD	4.51	2.48	$p = 0.01$
DENSITY	7.67	1.62	$p = 0.01$
SC-1	14.77	2.634	$p = 0.28$
SP-FAM	3.36	1.452	$p = 0.66$
UNEMP	14.71	2.411	$p = 0.40$

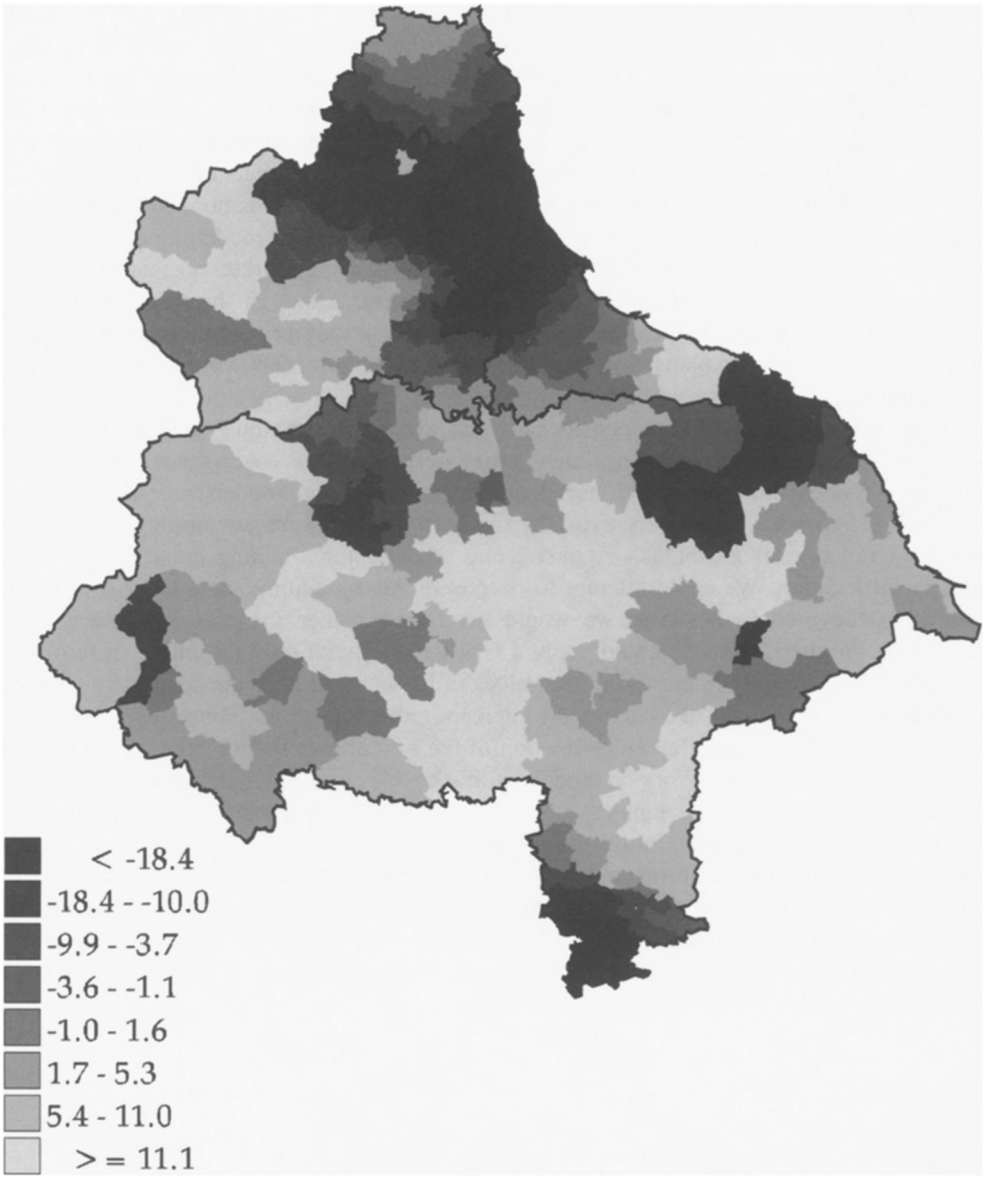

Fig. 2. Geographical variation of the density coefficient

variables. It is hoped that the technique presented here has provided a direct and flexible approach to the problem, and that the ability to carry out statistical testing on hypotheses relating to spatial stability in regression parameters is a useful feature.

One possible objection to this technique, or indeed to any of those discussed in Section 1, is that much of the spatial variation in parameters could be removed by the addition of further explanatory variables. In response to this, we would argue that, although the spatial variation may indeed be caused by factors that are not accounted for in our models, there are three reasons for adopting this method. The first is essentially for exploration. If we have a good subjective

knowledge of the geography of an area, then it is possible that factors causing the spatial patterns might be suggested. We could then go on to add further variables to the model. Secondly, it is possible that the factors affecting the spatial variations in the model may not be easily measured; for example when examining spatial variations in models predicting voting patterns there could be cultural attitudes that are peculiar to certain localities which cause unusual voting behaviour for given social and economic conditions. At least in this approach the effect of this can be seen, even if it is difficult to quantify the phenomenon itself. Finally, there are situations in which the desired output of the model is geographical. A good example of this is in the modelling of house prices. It may be useful from a viewpoint of understanding local housing markets to map things like the contribution of a second bathroom to the cost of a house. Although the value of this may be determined by such things as local demographic patterns, it may be that this quantity itself is of interest rather than other explanatory variables. If that is the case, GWR provides a direct method of obtaining it.

In conclusion, we would like to consider ways in which the technique could be improved. One possibility is to consider models in which some coefficients vary over space, while others are fixed. Such models may be useful, for example, in regression models where some possibly unmeasurable nuisance parameters exist. If, for example, we were attempting to model house prices as a function of attributes of houses, one other factor affecting house prices might be geographical location. We could attempt to overcome this by adding some local socioeconomic variables to the model—however, we would need to consider the choice of such variables carefully. As an alternative, we could apply a GWR model with only the intercept term varying over space. If we are only interested in the values of the housing attribute coefficients, the rest of the model could be thought of as a varying intercept term. We are not interested in the causes of the variation in this instance, but an estimation of the spatial patterns in variation is necessary to model the other coefficients. Work on models of this kind is currently in progress by the authors.

Linked with this is the development of asymptotic significance tests comparing models with different parameters varying. Unlike the Monte Carlo tests, these would be based on known (approximate) distributional forms and would not therefore incur the computational costs of the techniques described in this paper. Such tests will be based on those applying to the locally weighted regression techniques of Cleveland and Devlin (1988). It is hoped that with the development of these, and other related techniques, a coherent theory of spatially non-stationary regression models can be developed.

References

Cassetti, E. (1972) The expansion method. *Geogr. Anal.*, **4**, 432–449.
Chatfield, C. (1994) *A Guide to Statistical Problem Solving*. London: Chapman and Hall.
Cleveland, W. S. and Devlin, S. J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Statist. Ass.*, **83**, 596–610.
Dale, A. and Marsh, C. (1993) *The 1991 Census User's Guide*. London: Her Majesty's Stationery Office.
Fotheringham, A. S. (1992) Exploratory spatial data analysis and GIS. *Environ. Planng* A, **25**, 156–158.
———(1994) On the future of spatial analysis: the rôle of GIS. *Environ. Planng* A, **23**, 34–36.
Goldstein, H. (1987) *Multilevel Models in Educational and Social Research*. Oxford: Oxford University Press.
Gorr, W. L. and Olligschlaeger, A. M. (1994) Weighted spatial adaptive filtering: Monte Carlo studies and application to illicit drug market modelling. *Geogr. Anal.*, **26**, 67–87.
Holt, D., Steel, D. G., Tranmer, M. and Wrigley, N. (1996) Aggregation and ecological effects in geographically based data. *Geogr. Anal.*, **28**, 244–261.
Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *J. R. Statist. Soc.* B, **30**, 582–598.
Jones, K. (1991) Specifying and estimating multilevel models for geographical research. *Trans. Inst. Br. Geogr.*, **16**, 148–159.
Maritz, S. J. (1985) *Distribution-free Statistical Methods*. London: Chapman and Hall.
Office of Population Censuses and Surveys (1991) *Census Definitions*. London: Her Majesty's Stationery Office.

Openshaw, S. (1993) Exploratory space-time-attribute pattern analyzers. In *Spatial Analysis and GIS* (eds A. S. Fotheringham and P. A. Rogerson), pp. 147–163. London: Taylor and Francis.
Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation, Inequality and the North*. London: Croom Helm.
Wand, C. and Jones, B. (1995) *Kernel Smoothing*. London: Chapman and Hall.