9. Geographically Weighted Regression

MARIE-PIERRE DE BELLEFON, JEAN-MICHEL FLOCH
INSEE

Abstract

Geographically Weighted Regression (GWR) was developed in response to the finding that a regression model estimated over the entire area of interest may not adequately address local variations. The fairly simple principle on which it is based consists on estimating local models by least squares, each observation being weighted by a decreasing function of its distance to the estimation point. Combining these local models makes it possible to build a global model with specific properties. GWR can be used, in particular with the help of associated cartographic representations, to identify where local coefficients deviate the most from the overall coefficients, to build tests to assess whether the phenomenon is non-stationary and to characterise non-stationary. The method is presented using the example of a hedonic pricing model – prices of existing housing in Lyon. We show how to optimally determine the radius of the disk on which local regressions will be performed and we present the estimation results, the robust estimation methods and the tests of coefficients' non stationarity. In addition to this descriptive use, we present a more predictive approach, showing how taking non-stationarity into account makes it possible to improve an estimator over a spatial area. The example is based on a model linking the poor population and the number of beneficiaries of supplementary universal health coverage (CMU-C) in Rennes.

R Prior reading of chapter 3: "Spatial autocorrelation indices" is recommended.

9.1 Why use geographically weighted regression?

To identify the nature of relationships between variables, linear regression models the dependent variable y as a linear function of explanatory variables $x_1,...,x_p$. If you have n observations, the model is written:

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ik} + \varepsilon_i,$$

where β_0, β_1,..., β_p are the parameters and ε_1, ε_2,..., ε_n are the error terms. In this model, the coefficients β_k are considered identical across the study area. However, the hypothesis of spatial uniformity of the effect of explanatory variables on the dependent variable is often unrealistic (Brunsdon et al. 1996). If the parameters vary significantly in space, a global estimator will hide the geographical richness of the phenomenon.

Spatial heterogeneity corresponds to this spatial variability in the model's parameters or its functional form. When the territory of interest is well-known, it is often treated in empirical literature by adding dummy variables of geographical zones in the model – possibly crossed with each explanatory variable – by estimating the model for different zones or by conducting tests of geographical stability on the parameters (known as Chow tests). When the number of these geographic areas increases, this treatment nevertheless decreases the number of degrees of freedom and therefore the accuracy of the estimators.
Local regressions can also be used, the spatial application of which is referred to as GWR, Geographically Weighted Regression (Brunsdon et al. 1996). Through the example of the study of property prices in Lyon, we show the interest of performing a geographic regression (example 9.1) and how to implement it (example 9.2).

More complex methods coming from geographical researchers have been developed (Le Gallo 2004), but they remain largely descriptive and exploratory – in particular through graphical representations – as their theoretical behaviour is not fully known, in particular their convergence and their handling of geographic break.

■ **Example 9.1 — Use of a hedonic model to study real estate prices in Lyon.** Mapping changes in real estate prices makes it possible to generally deduct that prices tend to be higher in the centre than in the outskirts (Figure 9.1). However, these high prices may be explained by better quality in the housing which is sold in the centre. The hedonic model is aimed at **isolating the effect of localisation on prices.** The principle of this method is that the price of a property is a combination of the prices of its various attributes

$$y_i = \beta_0 + \sum_{k}^{p} \beta_k x_{ik} + \varepsilon_i \tag{9.1}$$

with x_{ik} the characteristic k of property i, β_k the coefficient associated with this characteristic and p the number of explanatory variables.

The assumptions underlying the hedonic model are that sellers and buyers are individual agents, without market power, and that this is a situation of perfect competition. The hedonic regression coefficient corresponding to a characteristic informs about the value which the purchasers **at equilibrium at a given time** would give to **an increase in the quantity of this characteristic.**

Figure 9.2 depicts the residuals of a hedonic regression of flats' prices on their physical characteristics. These residuals are not randomly distributed in space – the null hypothesis of the Moran test is rejected. The Moran's I of the distribution of residuals is positive, which is a

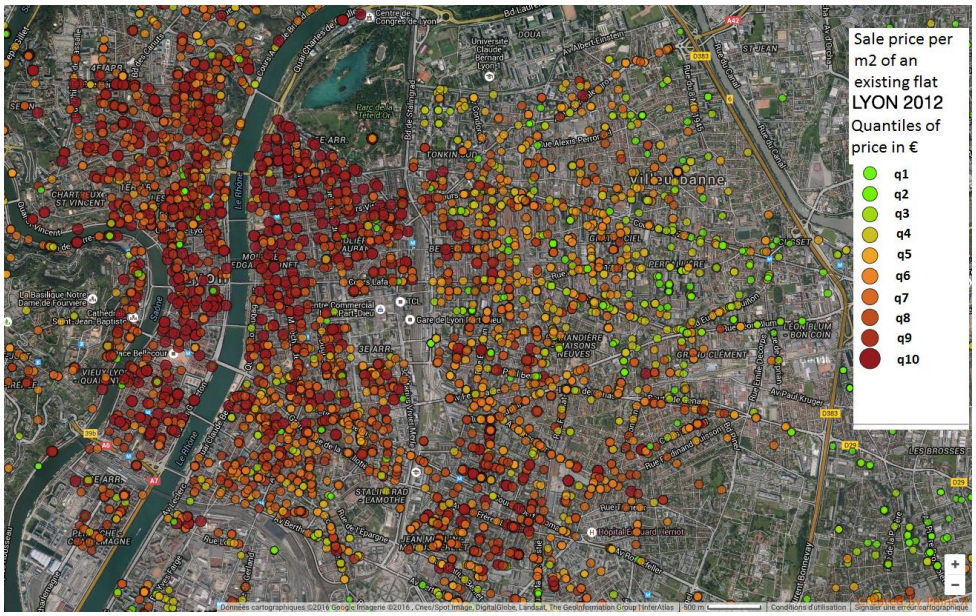

Figure 9.1 – Sale price per m² of an existing flat - 2012
Source: *PERVAL base*
Scope: Lyon conurbation

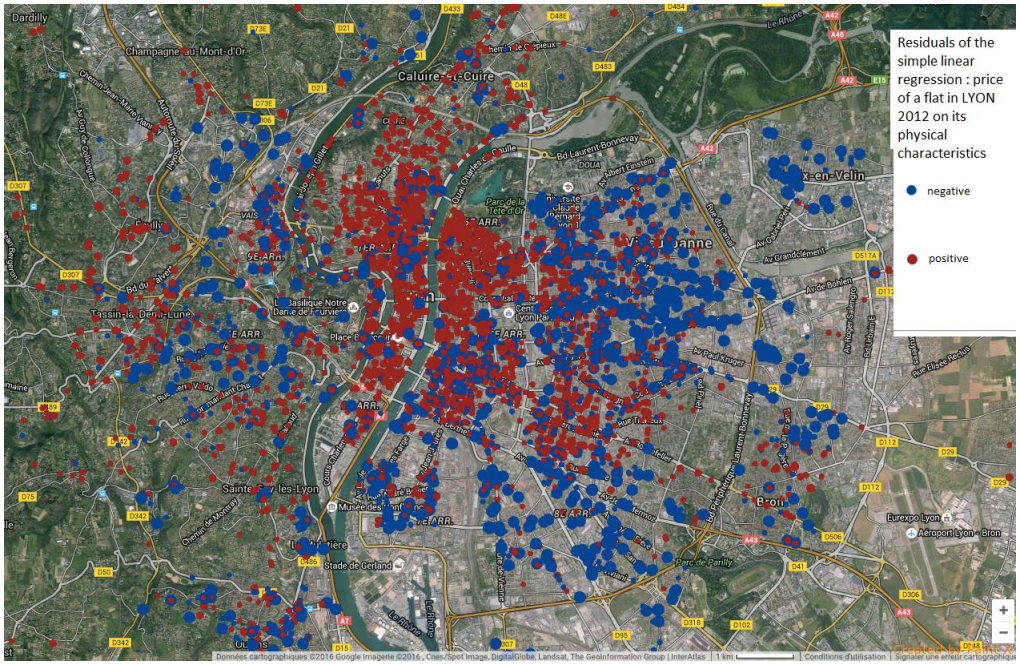

Figure 9.2 – Residuals from the hedonic regression of price on the characteristics of the property
Source: *PERVAL base*
Scope: Lyon conurbation

sign of positive spatial correlation in the residuals. **The hypothesis of spatial stationarity of the relationship between price and characteristic of the property is not valid**. We can therefore conclude in the existence of **spatial heterogeneity**.

As shown above, in order to take into account the variation in the model parameters with the location, a commonly used method consists in introducing geographical dummy variables as explanatory parameters. Let us examine the evolution of the influence on the price per square meter of an existing flat in Lyon in 2012 of a construction year between 1992 and 2000 as opposed to between 1948 and 1969, depending on the arrondissement in which the property is located.

Arrondissement	Parameter estimate	Significance
1st	1,1511	.
2nd	1,1499	.
3rd	**1,1481**	***
4th	1,360	**
5th	1,4909	***
6th	**1,3085**	***
7th	1,1897	***
8th	1,1487	***
9th	1,1981	***

Table 9.1 – Significance of regression coefficients associated with the time of construction
*, **, *** are the significance thresholds at 10, 5 and 1%
Source: *PERVAL base*
Scope: Lyon conurbation

The value and significance of the coefficients change with the arrondissements (table 9.1). Buyers therefore value the time of construction differently, depending on their location. However, why would the boundaries that define the changes in model match the administrative boundaries? **Geographically Weighted Regression allows study a model that varies spatially in a continuous way.**

■

9.2 Geographically Weighted Regression

9.2.1 A model with variable coefficients

Geographically Weighted Regression belongs to the category of models with variable coefficients. The regression coefficients are not fixed, they depend on the geographical coordinates of observations. In other words, **the coefficients of the explanatory parameters form continuous surfaces that are assessed at certain points in space**,

$$y_i = \beta_0(u_i, v_i) + \sum_k^p \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{9.2}$$

where (u_i,v_i) are the geographical coordinates.

9.2.2 How to estimate the model?

To estimate the model, the following hypothesis is used: **the closer two observations are geographically, the more similar the influence of the explanatory variables on the dependent variable, *i.e.* the closer the coefficients of the explanatory parameters of the regression.**

Therefore, to estimate the model with variable coefficients at point i, we want to use the fixed-coefficients model and include in the regression only the observations close to i. However, the more points are included in the sample, the lower the variance, but the higher the bias. The solution is therefore to **reduce the importance of the most remote observations by giving each observation a decreasing weight with the distance to the point of interest.**

The model to be estimated is as follows:

$$\mathbf{Y} = (\beta \otimes \mathbf{X})\mathbf{1} + \varepsilon \tag{9.3}$$

$\mathbf{Y}$: vector $n \times 1$ of the dependent variable.
$\mathbf{X}$: matrix $n \times (p+1)$ of p explanatory variables + constant
$\mathbf{1}$: vector $(p+1) \times 1$ of 1

The coefficients β of the model can be expressed in matrix form:

$$\beta = \begin{bmatrix} \beta_0(u_1, v_1) & \dots & \beta_p(u_1, v_1) \\ \beta_0(u_j, v_j) & \dots & \beta_p(u_j, v_j) \\ \beta_0(u_n, v_n) & \dots & \beta_p(u_n, v_n) \end{bmatrix} \tag{9.4}$$

The $\otimes$ operator multiplies each component of the coefficients matrix β by the corresponding element of matrix X which contains the characteristics of the observations.

In order to give a weight to observations decreasing with their distance to the point of interest, an estimate is performed using weighted least squares, the weighting being governed by weight matrix $W_{(u_i, v_i)}$. The parameters governing the construction of this matrix are detailed in section 9.2.3.

In accordance with the principle of weighted least squares, coefficients $\hat{\beta}(u_i, v_i)$ at the point of geographic coordinates (u_i, v_i) minimize sum 9.5:

$$\sum_{j=1}^{n} w_j(i)(y_j - \beta_0(u_i, v_i) - \beta_1(u_i, v_i)x_{j1} - \dots - \beta_p(u_i, v_i)x_{jp})^2 \tag{9.5}$$

$$\hat{\beta}(u_i, v_i) = (\mathbf{X^T W_{(u_i, v_i)} X})^{-1} \mathbf{X^T W_{(u_i, v_i)} Y} \tag{9.6}$$

$\hat{\mathbf{Y}} = \mathbf{SY}$,where S is the "hat matrix" defined by Equation 9.7. Let's note $\mathbf{x_i^T} = (1 \quad x_{i1} \quad x_{i2} \quad \dots \quad x_{ip})$ column i of the explanatory variable matrix $\mathbf{X}$. Then

$$\mathbf{S} = \begin{bmatrix} (\mathbf{x_1^T X^T W_{(u_1, v_1)} X})^{-1} \mathbf{X^T W_{(u_1, v_1)}} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x_n^T X^T W_{(u_n, v_n)} X})^{-1} \mathbf{X^T W_{(u_n, v_n)}} \end{bmatrix} \tag{9.7}$$

Reminder: ordinary least squares estimation

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{9.8}$$

$$\hat{\beta} = (\mathbf{X^T X})^{-1} \mathbf{X^T Y} \tag{9.9}$$

$\mathbf{Y}$: vector $n \times 1$ of the dependent variable.
$\mathbf{X}$: matrix $n \times (p+1)$ of the p explanatory variables + the constant.

9.2.3 Choosing the estimation parameters

Matrix $W_{(u_i,v_i)}$ contains the weight of each observation according to its distance to the point i of coordinates (u_i, v_i) (Figure 9.3). We assume that observations close to point i have more influence over the estimated parameters at place i than more remote observations. The weight of observations therefore decreases with the distance to the point i. There are several ways of specifying this decrease. Here we show the main parameters governing the decrease function.

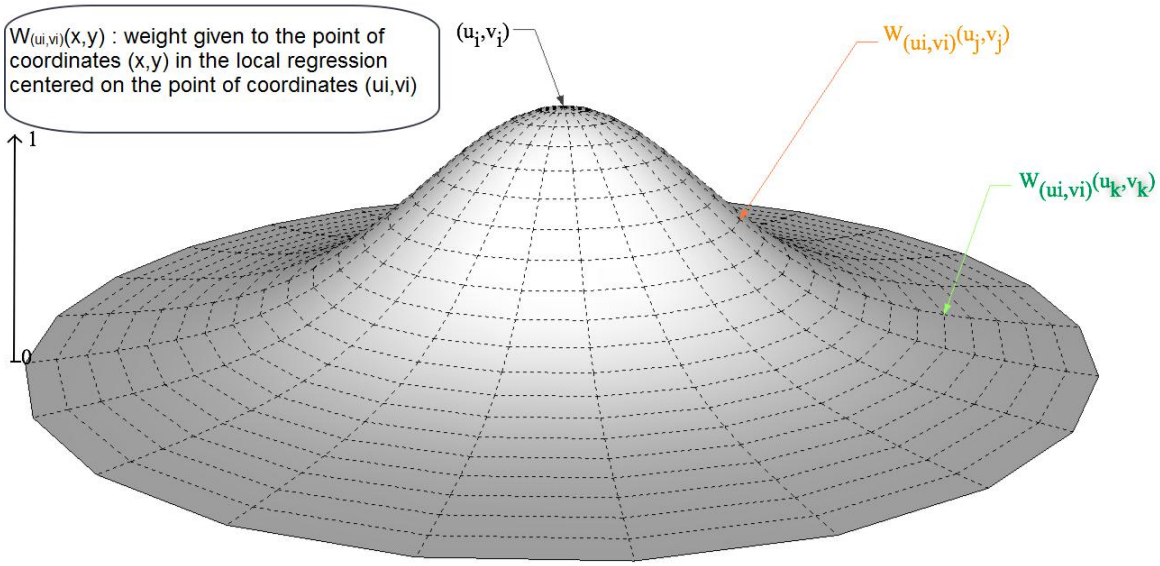

Figure 9.3 – Graphical representation of matrix W

The decrease in the weight of each observation with distance to the point of origin is determined by a **kernel function**. The key parameters of the kernel function are:

— the shape of the kernel ;
— fixed kernel versus adaptive kernel ;
— bandwidth size.

Shape of the kernel

We can distinguish the continuous kernel that weights all the observations (Figure 9.4; table 9.2) of the kernel with compact support (Figure 9.5; table 9.3) for which the weight of observations is zero beyond a certain distance. However, **the shape of the kernel only changes the results slightly** (Brunsdon et al. 1998).

Uniform kernel	$w(d_{ij}) = 1$
Gaussian kernel	$w(d_{ij}) = exp(-\frac{1}{2}(\frac{d_{ij}}{h})^2)$
Exponential kernel	$w(d_{ij}) = exp(-\frac{1}{2}(\frac{\lvert d_{ij}\rvert}{h}))$

Table 9.2 – Continuous kernel

— Choosing a uniform kernel means doing an ordinary least squares regression at each point.
— The Box-Car kernel handles a continuous phenomenon in a discontinuous way.
— Gaussian and exponential kernels weight all the observations, with a weight that tends towards zero with the distance to the estimated point.
— The bisquare and tricube kernels also give observations a decreasing weight with distance, but this weight is zero beyond a certain distance h called bandwidth).

Box-Car Kernel	$w(d_{ij}) = 1$ if $\|d_{ij}\| < h$, 0 otherwise
Bi-Square	$w(d_{ij}) = (1 - (\frac{d_{ij}}{h})^2)^2$ if $\|d_{ij}\| < h$, 0 otherwise
Tri-Cube Kernel	$w(d_{ij}) = (1 - (\frac{\|d_{ij}\|}{h})^3)^3$ if $\|d_{ij}\| < h$, 0 otherwise

Table 9.3 – Kernel with compact support

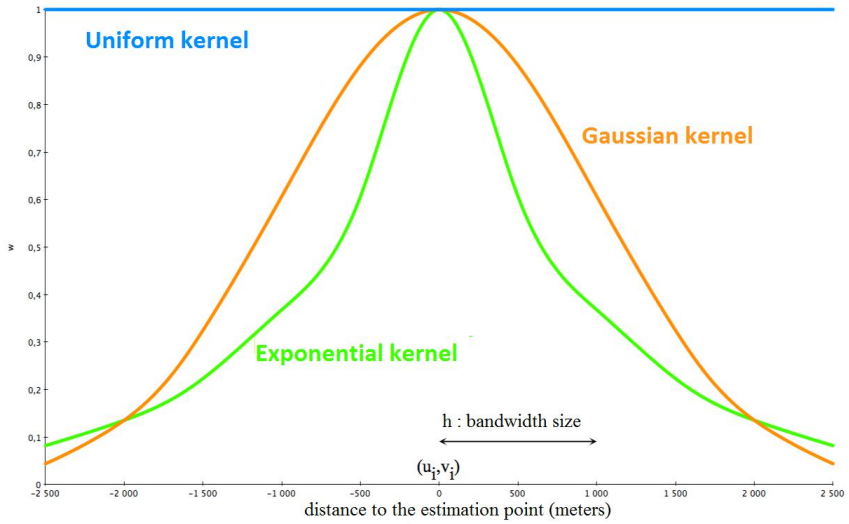

Figure 9.4 – Continuous kernel

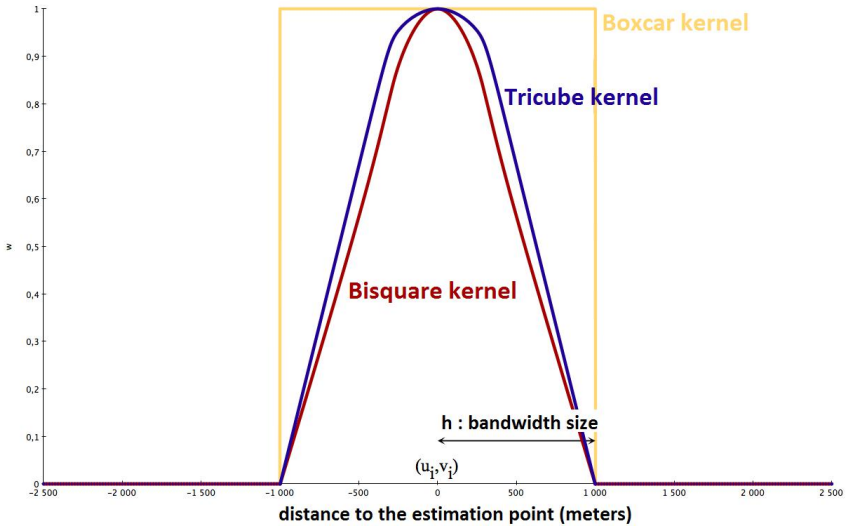

Figure 9.5 – Kernel with compact support

⇒ **The bisquare kernel should be preferred in order to optimise calculation time.**

Fixed kernel versus adaptive kernel

Definition 9.2.1 — Fixed kernel. The extent of the kernel is determined by the **distance** to the point of interest. The kernel is identical at any point in space (Figure 9.6).

Definition 9.2.2 — Adaptive kernel. The extent of the kernel is determined by the **number of neighbours** of the point of interest. The lower the density of the observations, the smaller the kernel (Figure 9.7).

— A fixed kernel is suited to a uniform spatial distribution of data but not very effective in the case of a non-homogeneous distribution. Its radius must be at least equal to the distance between the most isolated point and its first neighbour, which may cause the number of points included in the regression to vary significantly.
— In low-density areas, a fixed kernel that is too small will include too few points in the regression. The variance will be higher.
— In very dense areas, a fixed kernel that is too big will overlook variations on a fine scale. The bias will be higher.

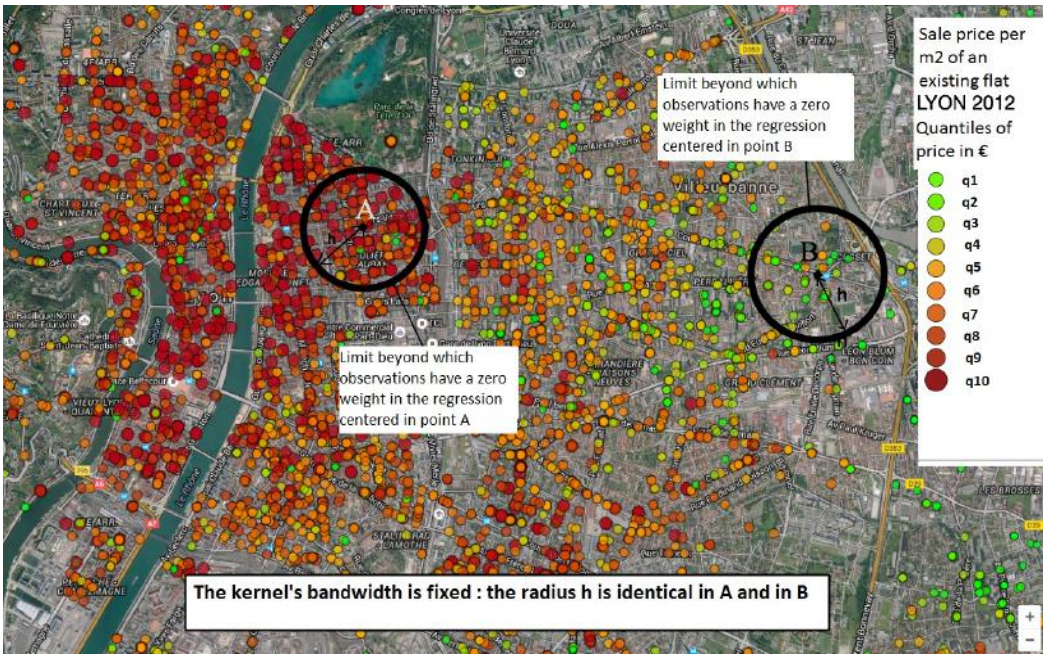

Figure 9.6 – Fixed kernel
Source: *PERVAL base*

Definition and choice of bandwidth

The bandwidth is a distance beyond which the weight of the observations is assigned the value 0. **The value of bandwidth h is the parameter the choice of which has the strongest influence on results.** The larger the bandwidth, the higher the number of observations to which the kernel gives a non-zero weight. The local regression will then include more observations and the results will be smoother than with a small bandwidth. When the bandwidth tends towards infinity, the results of the local regression are similar to those of ordinary least squares regression.

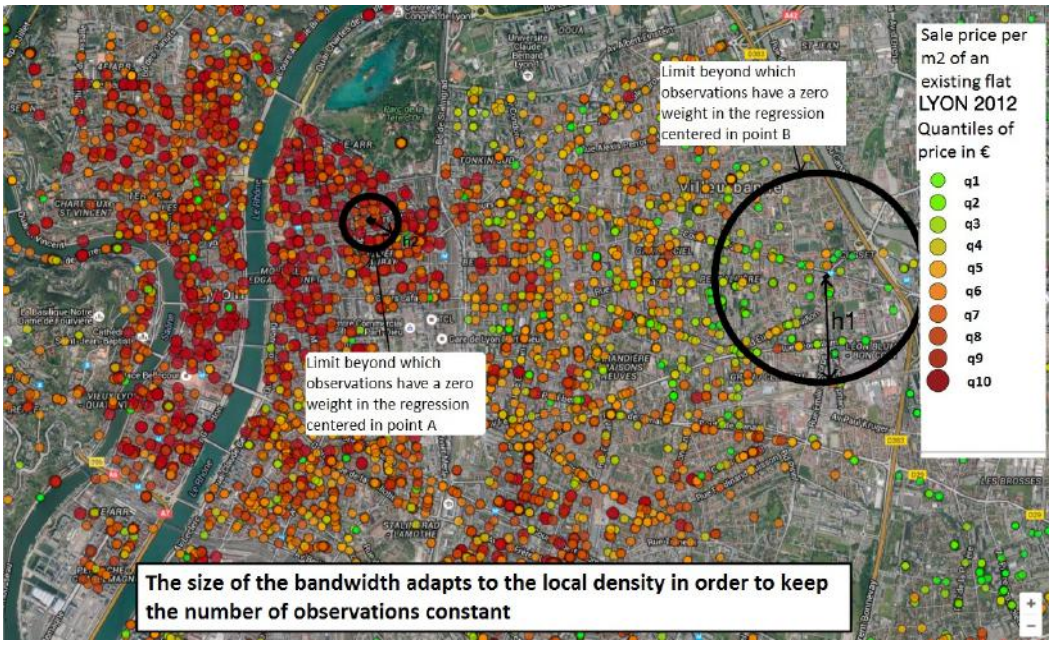

Figure 9.7 – Adaptive kernel
Source: *PERVAL base*

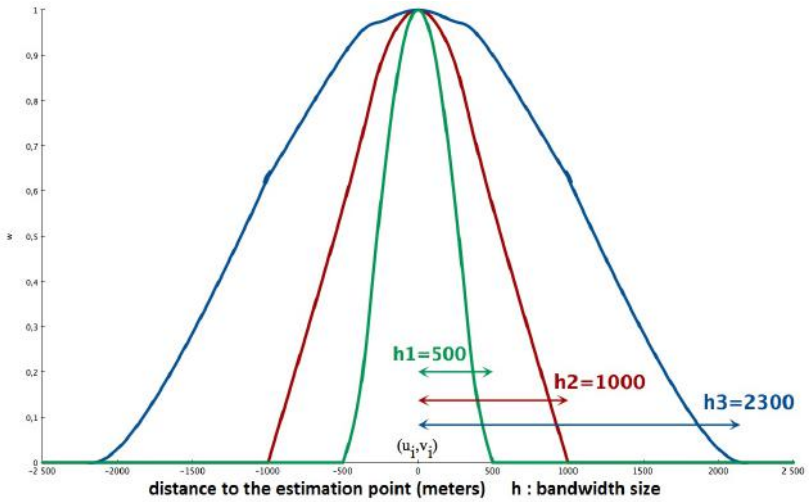

Figure 9.8 – Influence of the choice of bandwidth on the kernel

The choice of the bandwidth is not linked to the model itself, but to the calibration strategy. If the kernel includes points that are too far away, the variance will be low but the bias high. If the kernel only covers the closest points, the bias will be low but the variance high. Several statistical criteria can help choose the most suitable bandwidth. The *GW.model* R package makes it possible to determine the bandwidth that minimises either of the two criteria — the cross-validation criterion and the adjusted Akaike criterion (see boxes 9.2.1 and 9.2.1).

The value of the bandwidth minimising these criteria is also a valuable indication of the relevance of a Geographically Weighted Regression modelling. If the bandwidth tends to the maximum possible – the entire extent of the study area, or all the points – then the local heterogeneity is probably not significant and the GWR is not necessary. Conversely, an extremely small bandwidth should be seen as an alert to the risk that the underlying process could be random (Gollini et al. 2013). It should also be remembered that the bandwidth that minimises the statistical criteria is based on the prediction of the dependent variable, and not of the regression coefficients – which are however those used later to test the validity of the non-stationarity hypothesis.

Box 9.2.1 — Cross-validation criteria.

$$CV = \sum_{i=1}^{n} \left[y_i - \hat{y}_{\neq i}(h) \right]^2$$

$\hat{y}_{\neq i}(h)$ is the value of y at point i predicted when calibrating the model with all the observations except y_i. If the model were estimated with all its observations, the optimum bandwidth would indeed be 0 given that, when $h = 0$ there is no other point than y_i in the regression; hence $\hat{y}_i = y_i$ which is the attainable optimum.
Bandwidth h that minimises CV – the cross-validation score – **maximises the model's predictive power**.

Box 9.2.2 — Adjusted Akaike criterion.

$$AIC_c(h) = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\left\{ \frac{n + tr(S)}{n - 2 - tr(S)} \right\}$$

n is the sample size; $\hat{\sigma}$ is the estimate of the standard deviation of the error term; $tr(S)$ is the trace of the projection matrix (hat matrix) of observed variable y on estimated variable $\hat{y}$.
The AIC criterion favours a **compromise between the predictive power of the model and its complexity.** The lower the bandwidth, the more complex the global model. The AIC criterion generally favours larger bandwidths than the CV criterion.

9.3 Robust Geographically Weighted Regression

Just like standard linear regression, Geographically Weighted Regression is sensitive to outliers. These points distort local parameters surface estimates (Brunsdon et al. 1996). Since Geographically Weighted Regression takes into account a different model at each point of the space, it is sufficient that one point be unusual **relative to the local context** for the estimate to be distorted. There is, however, more chance for a point to be unusual in relation to the local context rather than the global. By looking for outliers at global level, one thus may overlook points that are unusual locally, but not globally. Two methods have been developed to remedy this problem.

Method 1: filter according to standardised residuals

The aim of method 1 is to detect observations with very high residuals and exclude them from the regression.

Let $e_i = y_i - \hat{y}_i$ be the residual of the estimate at point i.If y_i is an outlier, e_i should have a very high value. However, the residuals do not all have the same variance, so they must be standardised so that they can be compared and a decision made as to which need to be removed from the regression.

Note $\hat{y} = \mathbf{Sy}$ where S is the hat matrix defined above. $\mathbf{e} = \mathbf{y} - \mathbf{Sy} = (\mathbf{I} - \mathbf{S})\mathbf{y}$ with $\mathbf{e}$ the vector of the residuals.
$var(\mathbf{e}) = (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^{\mathbf{T}} var(\mathbf{y}) = (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^{\mathbf{T}}\sigma^2$ with σ the standard deviation of y
The variances of the e_is are therefore the leading diagonal elements on the matrix $(\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^{\mathbf{T}}\sigma^2$, which in general are not equal (Brunsdon et al. 1996).

Consider $\mathbf{Q} = (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^{\mathbf{T}}$ and q_{ii} the ith element of the diagonal of $\mathbf{Q}$

$r_i = \frac{e_i}{\hat{\sigma}\sqrt{q_{ii}}}$ is called the *internally standardised residual*.
If point i is unusual, including it in the estimate of $\hat{\sigma}^2$ may produce a bias. The value of σ is thus estimated by excluding the potentially outlier observation i, σ_{-i}

$r_i^* = \frac{e_i}{\hat{\sigma_{-i}}\sqrt{q_{ii}}}$ is called *externally standardised residual*.

With method 1, the observations for which $|\mathbf{r_i^*}| > 3$ are filtered (the threshold of 3 is proposed by Chatfield 2006).

Disadvantage: $\mathbf{Q}$ is a matrix $n * n$ of which **the calculation time is prohibitive for large databases**, at this time and with a machine with conventional calculation power. For example: Brunsdon et al. 1996 deem that this method cannot be used beyond 10 000 observations.

Method 2: reducing the weight of observations with high residuals

The objective of method 2 is to lower the weight of the observations with high residuals (Huber 1981). After an initial estimation of the model, weight $w_r(e_i)$ is ascribed to each observation i. This weight must be multiplied with the weight which varies according to the distance to the point i. A new matrix W is thus created, which is the term by term product between the old matrix W and a matrix W_r of the residual weights, defined as:

$$w_r(e_i) = \left\{ \begin{array}{c} 1 \quad if \quad |e_i| \leq 2\hat{\sigma} \\ \left[1 - (|e_i| - 2)^2\right]^2 \quad if \quad 2\hat{\sigma} < |e_i| < 3\hat{\sigma} \\ 0 \quad otherwise \end{array} \right\} \tag{9.10}$$

If none of the residuals from the first regression is higher than two standard deviations, the second model is identical to the first. The observations of which residuals are between two and three standard deviations have their weight reduced in the second regression, while the observations whose residuals exceed three standard deviations are excluded.

Discussion

Method 2 is much quicker to calculate than method 1 since each cycle only requires the calculation of the n residuals rather than a matrix $n * n$. However, it does not take into account differences in variance between residuals and eliminates more points than does method 1.

Application with R

Package *GW model* is used to implement Geographically Weighted Regression. The first step consist in calculating the distances between all the observations thanks to function `gw.dist`. Then function `bw.gwr` is used to calculate the bandwidth of the kernel function, optimal with respect to a given statistical criterion. Lastly, the local coefficients of the Geographically Weighted Regression are derived thanks to function `gwr.robust`. The results are contained in an object of class `gwrm`, containing in particular an object of type `SpatialPointsDataFrame`, the contents of which are detailed below.

Options of function `gw.dist`

— `dp.locat`: coordinates of observations;
— `rp.locat`: coordinates of points at which to calibrate the model (*e.g.* points on a regular grid);
— `p`: governs the choice of distance (p=1: Manhattan - p=2: Euclidean);
— `theta`: angle to rotate the coordinates system (useful for Manhattan distance).

Options the function `bw.gwr`

— `formula`: the model $y \sim x_1 + x_2 + ... + x_p$
— `approach`: optimal bandwidth calculation method: CV (Cross Validation) or AIC (Akaike Information Criterion).
— `kernel`: type of kernel: "gaussian", "exponential", "bisquare", "tricube", "boxcar"
— `adaptive` if TRUE, then the bandwidth is a number of neighbours, and the kernel is adaptive. if FALSE, the bandwidth is a distance, and the kernel is fixed.
— `dMat`: pre-calculated distance matrix.

Options of function `gwr.robust`

— `regression points`: geographical coordinates of points where the model will be evaluated.
— `bw`: size of the bandwidth.
— `filtered`: if TRUE, filter the observations according to the value of standardised residuals robust regression method 1) and if FALSE, estimate the model a second time by weighting the observations according to the value of their residuals (robust regression method 2)
— `F123.test`: calculates the Fischer statistics (FALSE by default)
— `maxiter`: maximum number of iterations of the automatic approach (Method 2). It is equal to 20 by default.
— `cut1`: $\sigma cut1$ is the threshold value of the residuals beyond which the observations have a weight < 1 (set to 2 by default).
— `cut2`: $\sigma cut2$ is the threshold value of the residuals beyond which the observations have a null weight (set to 3 by default).
— `delta`: tolerance threshold of the iterative algorithm (set to $1.0e^{-5}$ by default).

Interpreting the results: the contents of file $SDF

— The $SDF file is of "SpatialPointsDataFrame" type, which contains attributes associated with geographic coordinates.
— *c_x*: estimation of the coefficient associated with characteristic x at each point.
— *yhat*: predicted value of y.
— *intrinsic*, *Stud_residual*: residual and standardised residual
— *CV_score*: cross validation score
— *x_SE*: standard error of the estimate of the coefficient associated with characteristic x.
— *x_TV*: t-value of the estimate of the coefficient associated with characteristic x.
— *E_weight*: weight of the observations in the robust regression (to be multiplied by the weight obtained with the kernel function).

■ **Example 9.2 — Application to the study of Lyon real estate prices.** Geographically Weighted Regression makes it possible to study the influence of a property's location on its price, while taking into account spatial heterogeneity — the fact that the influence of the characteristics of a property on its price depends on its location. The coefficient associated with the constant of the geographic regression is the price of a reference apartment — the price of an apartment, once the influence of its physical characteristics has been taken into account.

Meaning of the variables in the example below:
f_lgpx: logarithm of the price per square metre.
c_epoqueA: dummy variable of construction before 1850
c_epoqueF: dummy variable of construction between 1981 and 1991
c_epoqueG: dummy variable of construction between 1992 and 2000
c_mmut1, 2, 3: dummy variable of a transfer in January, February, March, etc.
c_sdbn_2: dummy variable of the existence of two bathrooms.
c_cave1: dummy variable of the existence of a cellar.

```
library(GWmodel)
dm.calib <- gw.dist(dp.locat=coordinates(lyon2012))

#Calculation of a distance matrix between the points
bw0 <- bw.gwr(f_lgpx~c_epoqueG+c_mmut_1+c_mmut_2+
                c_mmut_3+c_epoqueA+c_epoqueF+c_sdbn_2+c_cave1,
              data=lyon2012, approach="AIC", kernel="bisquare",
              adaptive=TRUE,dMat=dm.calib)

gwr.robust.lyon2012 <- gwr.robust(f_lgpx~c_epoqueG+c_mmut_1+c_mmut_2+
                c_mmut_3+c_epoqueA+c_epoqueF+c_sdbn_2+c_cave1,
                bw=bw0, kernel="bisquare", filtered=FALSE, adaptive=TRUE,
                                                dMat=dm.calib)

#Extraction of the constant: price of the reference property (Figure 9.9)

lyon2012.intercept.robust <- gwr.robust.lyon2012$SDF[,c(1)]
# 1 is the position of the constant in the file containing the regression
    results.
lyon2012.intercept.robust$Intercept <- exp(lyon2012.intercept.robust$
    Intercept)

#Extraction of the coefficient linked to building before 1850 rather than
    between 1948 and 1969 (reference period) - Figure 9.10
lyon2012.epoqueA.robust <- gwr.robust.lyon2012$SDF[,c(15)]
lyon2012.epoqueA.robust$c_epoqueA <- exp(lyon2012.intercept.robust$c_
    epoqueA)

#Estimate of the (non robust) model on a grid of 100$*$100 metres (Figure
    9.11)
#Let "quadrillage" be a file of type "SpatialGridDataFrame" covering the
    area to be studied
dm.calib.quadrillage <-coordinates(quadrillage) gw.dist(dp.locat=
    coordinates(lyon2012),rp.locat=coordinates(quadrillage))
gwr.lyon2012<-gwr.basic(f_lgpx~c_epoqueG+c_mmut_1+c_mmut_2+c_mmut_3+c_
```

```
epoqueA+c_epoqueF+c_sdbn_2+c_cave1,regression.point=quadrillage,bw=bw0,
 kernel="bisquare", filtered=FALSE, adaptive=TRUE,    dMat=dm.calib.
quadrillage)
```

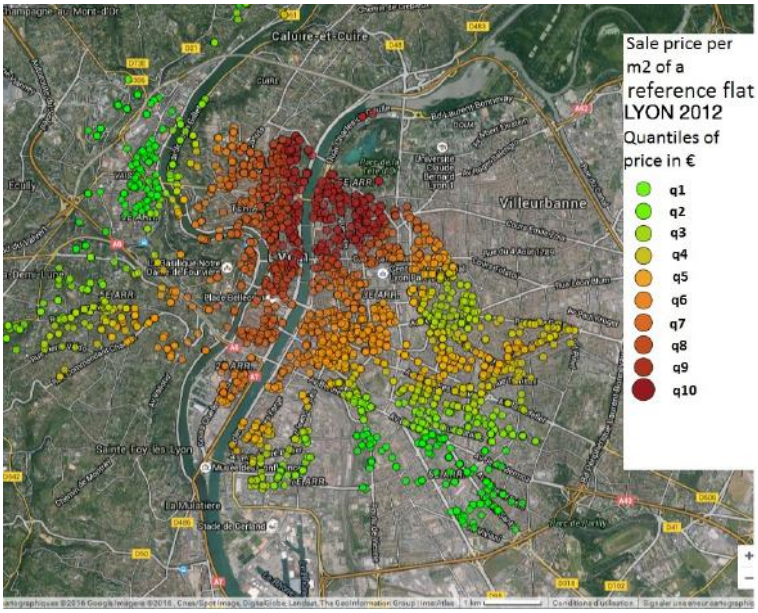

Figure 9.9 – Local constant: price of the reference property
Source: *PERVAL base*

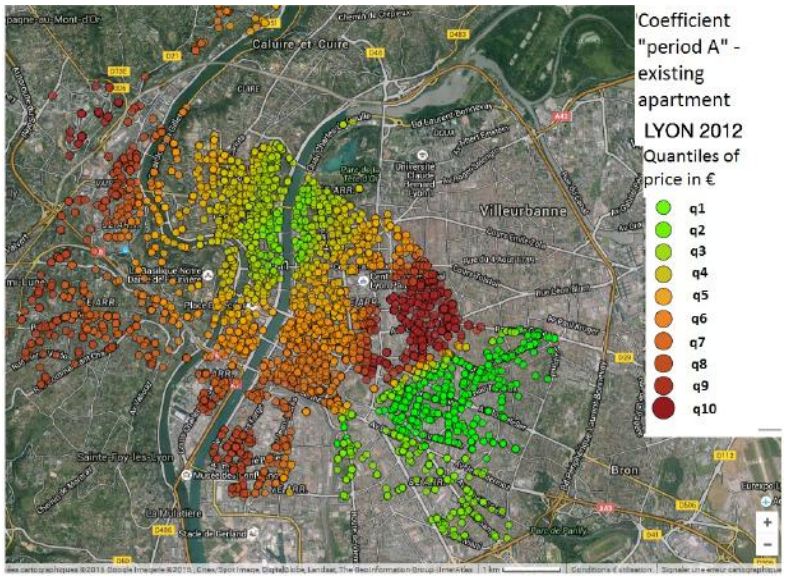

Figure 9.10 – Coefficient associated with building before 1850 rather than between 1948 and 1969 (reference period)
Source: *PERVAL base*

Hedonic regression coefficients vary in space (Table 9.4). Geographically Weighted Regression has made it possible to better understand the spatial richness of changes in the explanatory parameters of real estate prices since the estimates are independent of the arrondissements' administrative boundaries. On Figures 9.9 and 9.10, the points at which the coefficients have been estimated

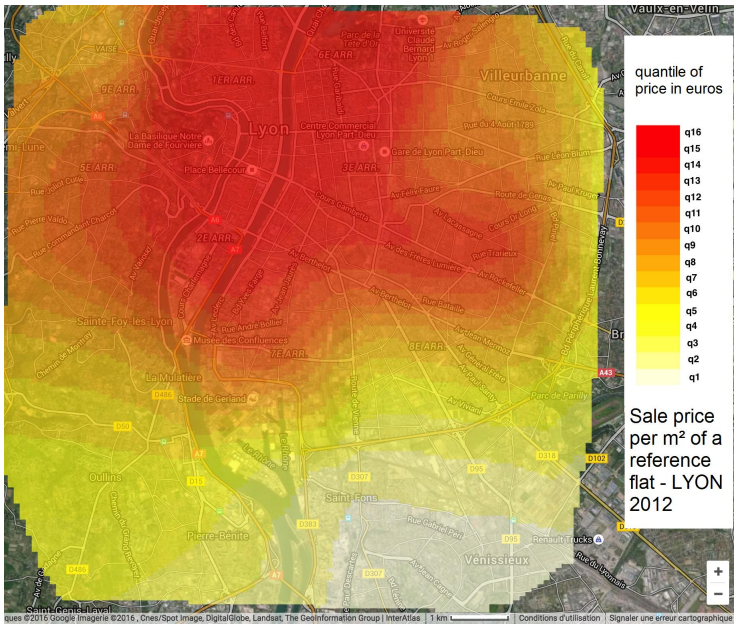

Figure 9.11 – Estimate of real estate prices on a square grid of 100m*100m
Source: *PERVAL base*

Coefficient	Min	1st quartile	Median	Average	3rd quartile	Max
constant	1666	2220	2668	2705	3088	4030
period A	0.6250	0.9533	1.1480	1.1070	1.2470	1.8190

Table 9.4 – Descriptive statistics of the hedonic GWR parameter estimates
Source: *PERVAL base*

are those where transactions have taken place. However, one of the interests of GWR also lies in the ability to estimate the values of coefficients continuously. Figure 9.11 presents an estimate of parameters on a grid of 100∗100 meters. Section 9.4 shows a method to assess the significance of the spatial variation of parameters. ■

9.4 Quality of estimates

9.4.1 Accuracy of parameter estimation

When we estimate a GWR with an adaptive kernel in an area where the observations are not very dense, the points used to calibrate the model may have a very low weight – they are located at a long distance from the point of estimation.

Let $\mathbf{C}$ be the matrix such that:

$$\hat{\beta}(u_i, v_i) = (\mathbf{X^T W_{(u_i,v_i)} X})^{-1} \mathbf{X^T W_{(u_i,v_i)} y} = \mathbf{CY} \tag{9.11}$$

The variance of the estimated parameter is:

$$Var\left[\hat{\beta}(u_i, v_i)\right] = \mathbf{CC^T}\sigma^2 \tag{9.12}$$

With σ^2 the sum of the standardised residuals of the local regression:

$$\sigma^2 = \sum_i (y_i - \hat{y}_i)/(n - 2\nu_1 + \nu_2) \tag{9.13}$$

$$\nu_1 = tr(\mathbf{S}) \tag{9.14}$$

$$\nu_2 = tr(\mathbf{S^T S}) \tag{9.15}$$

$$\hat{Y} = \mathbf{SY} \tag{9.16}$$

Once the variance of each parameter has been estimated, the standard errors are calculated using Equation 9.17

$$SE(\hat{\beta}(u_i, v_i)) = \sqrt{V}ar\left[\hat{\beta}(u_i, v_i)\right] \tag{9.17}$$

One can thus calculate confidence intervals for the coefficients.

Application with R

The $SDF file containing the results of the Geographically Weighted Regression makes it possible to access the standard errors associated with the various coefficients. For example, in the case of the example of Lyon real estate prices developed previously:

— `y`: selling price.
— `yhat`: estimated selling price.
— `Intercept_SE`: standard error of the coefficient associated with the constant.
— `Intercept_TV`: variation rate of the coefficient associated with the constant.

9.4.2 Testing non-stationarity of coefficients

The GWR does not take into account the hypothesis that the coefficients are stationary in a certain geographic area. To verify the relevance of the model, it is interesting to test the non-stationarity of the coefficients. **Do the coefficients vary enough in space to reject the hypothesis that they are constant throughout the study surface?**

In statistical terms, the question can be stated:
— H_0: $\forall k, \beta_k(u_1, v_1) = \beta_k(u_2, v_2) = ... = \beta_k(u_n, v_n)$
— H_1: $\exists k$, all $\beta_k(u_i, v_i)$ are not equal.
To answer this question, a simulation method of the Monte Carlo simulation type can be used.

Principle: If there were no underlying spatial phenomena, the geographical coordinates of the observations could be permuted randomly, and the variance would remain unchanged. In a Monte Carlo simulation, the geographic coordinates of the observations can be permuted n times. This results in n estimates of the spatial variance of coefficients. The next step consist in estimating the $p-value$ of the coefficients' spatial variability and rejecting - or maintaining - the null hypothesis that they are stable in space.
It should be remembered, however, that the methods simulating a spatial distribution of the observations depend upon the initial dataset. Leung et al. 2000 describe a more robust and less time-consuming calculation method for testing the coefficients' non-stationarity.

Application with R

Function `montecarlo.gwr`
— same parameters as function `gwr.robust`
— `nsims`: number of simulations
— `sortie`: vector containing p-values of all GWR parameters

9.5 A predictive application

Geographically Weighted Regression has been used primarily to highlight spatial heterogeneity. Like other regression methods, it can also be used for predictive purposes, for example to allocate values to unsampled units in a survey. This section of the article is based on work carried out by E.Lesage and J-M. Floch for the 2015 JMS [1], and presented at the 2016 workshop dedicated to Advanced Methods for the Analysis of Complex Samplings. In the small area estimation methods, approaches based on models using BLUP (Best Linear Unbiased Predictors) estimators (Chambers et al. 2012) are more and more frequent. The values of unsampled units are replaced by the predicted values from a model whose parameters are estimated using the values of the sampled units. An extension of these methods has been proposed (Chandra et al. 2012) in a non-stationary framework, using geographically weighted regression. The use of Geographically Weighted Regression in small-area estimation methods appears to be favoured in recent literature over methods derived from spatial econometrics, notably using spatial autoregressive models (SAR). Geographically Weighted Regression provides a more flexible way of taking into account phenomena of spatial variability. This consideration of spatial heterogeneity must theoretically improve the accuracy of estimators.

9.5.1 Problem overview

At INSEE, empirical research has used GWR to build estimators of population census data on priority neighbourhoods. In these neighbourhoods, 40% of housing is surveyed (over a five-year

1. Journée de Méthodologie Statistique – Workshop of Statistical Methodology, organised by INSEE every three years

period), but the sampling design is not optimal, because belonging to a priority neighbourhood is not one of the balancing variables. As there is a high demand for precise information on these neighbourhoods, we have sought to mobilise comprehensive or nearly-comprehensive administrative sources (tax data, health insurance data) to improve the accuracy of the estimators. To do so, a model was estimated over the housing units of the population census (RP), in which the variable of interest was a variable of the census, and the auxiliary variables were variables derived from administrative sources, well correlated to the variable of interest. The estimators made it possible to predict a value for the housing units that had not been sampled. **The estimator of the total of the variable of interest is the sum of the values observed for the sampled units and of the predicted values for the non-sampled units, the sampling weight no longer being taken into account in this calculation**.

This empirical work used Geographically Weighted Regression to model and take into account the significant heterogeneity found in urban data. However, the gain in precision compared to a non-spatial model had not been studied. This is why a comparison of three estimators is proposed from an experimental system based on actual administrative data — the Filosofi source, which makes it possible to calculate the population of low-income households, and the CNAM source (data from the National health insurance fund) , which provides the number of CMUC beneficiaries (Universal Supplementary Health Coverage). The Filosofi source is almost comprehensive and provides access to "real" figures on the number of people with incomes below the poverty rate.

Both sources are localised and can theoretically be matched based on their geographic coordinates. For reasons of confidentiality, it was not possible to do so, and we calculated the number of low-income individuals and the number of CMUC beneficiaries on a grid made up of 100 m $*$ 100 m squares, a compromise with the use of individual data deemed acceptable. These 100m squares play the role of statistical individuals on which we will carry out measurements.

The territory of interest is the municipality of Rennes. Inside the database, a sample of 40% of the squares is drawn, such as what is done in the population census. These squares will serve as a basis for estimating the number of low-income people (Figure 9.12). We have all the information, but for the model, low incomes are only known for sampled squares, while the beneficiaries of the CMUC are known for all squares. We select a sample of size $n = 856$, referred to as s, by simple random sampling without replacement (less complex than the sampling made for the population census). The sampling rate is $n/N = 40\%$. In addition, we note r the complement of sample s in U (the set of the inhabited squares in the Rennes region). The calculations made at square level allow calculations on IRIS level, with each square being assigned to an IRIS [2].

There is a strong linear link between the number of people with low incomes and the number of CMUC beneficiaries. The intercepts vary little from one square to the other. The slopes vary significantly from 1.6 to 3.3. The gradient of local situations is depicted in Figure 9.13. In the approach referred to as “model based”, the values y_i of the non-sampled squares are predicted using the model estimated with all the sampled data and with the auxiliary information x available for non-sampled squares. We build three estimators, j representing the IRIS:

Definition 9.5.1 — Horvitz-Thompson Estimator.

$$\widehat{t}_y(j) = \frac{N}{n} \sum_{i \in s_j} y_i \tag{9.18}$$

2. IRIS are the smallest French administrative delineations

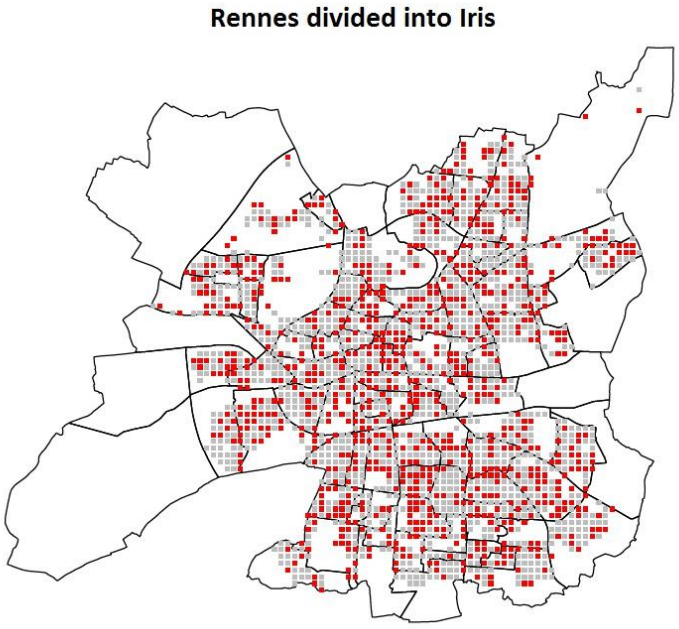

Figure 9.12 – 100 m squares in Rennes, sampled (in red) or not (in gray)

Definition 9.5.2 — Estimator based on "classic"' regression, without taking spatial heterogeneity into account.

$$\widehat{t}_{y,reg}(j) = \sum_{i \in s_j} y_i + \sum_{i \in r_l} \tilde{y}_l \tag{9.19}$$

where $\tilde{y}_l = \beta^T x_l$

Definition 9.5.3 — The estimator based on Geographically Weighted Regression.

$$\widehat{t}_{y,RGP}(j) = \sum_{i \in s_j} y_i + \sum_{i \in r_l} \breve{y}_l \tag{9.20}$$

where $\tilde{y}_l = \hat{\beta}_l^T x_l$ and $\widehat{\beta}_l$ is the vector of the coefficients of the Geographically Weighted Regression for square l.

9.5.2Results

This process is repeated $K = 1000$ times. For each IRIS, 1 000 values are derived for each of the three estimators. From these 1 000 values, Monte Carlo estimates of the biases and the root mean squared errors of estimators are developed.

If we note $\widehat{t}_y(j)^{(k)}$ the estimator of the total of variable y for IRIS j and for simulation k, the "Monte Carlo" root mean squared error can be calculated with the following equation:

$$EQM(\hat{t}_y(j)) = K^{-1} \sum_{k=1}^{K} (\hat{t}_y(j)^{(k)} - t_y(j))^2 \tag{9.21}$$

as the exact total $t_y(j)$is known.

The indicator deduced will be used to compare the results of the three estimates, the square

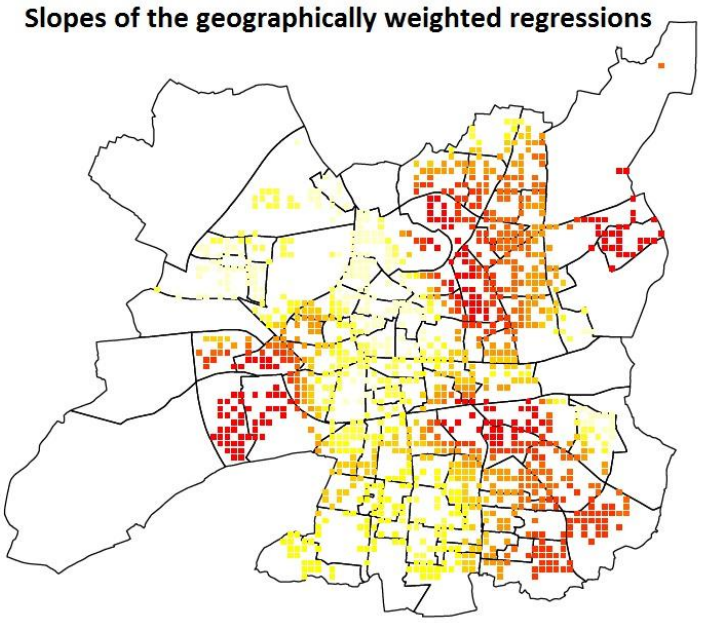

Figure 9.13 – Graphic representation of the slopes of the Geographically Weighted Regressions
Note: The scale used is a "heat scale" which ranges from the colour yellow (the highest values) to the colour red (the lowest values).

root of the mean squared error:

$$RCEQMR(\hat{t}_y(j)) = \frac{\sqrt{EQM(\hat{t}_y(j))}}{t_y(j)} \tag{9.22}$$

The IRISes of the municipality of Rennes are ranked in order of increasing population size and the RCEQMs are represented on Figure 9.14 for each of the IRIS.

The first message is the improvement of the accuracy in both model-based approaches, due to the good linear relationship between variable y (low-income individuals) and variable x (the CMUC beneficiaries). The RCEQMR is approximately 0.4 for the Horwitz-Thompson estimator, in the order of 0.12 for the models. The difference between the regression and the GWR is not very visible on Figure 9.14. The results are very close. The box-plots of Figure 9.15 make it possible to take the comparison slightly further.

In view of Figure 9.14 the GWR estimator nevertheless proves better than the regression estimator — for 75% of the IRIS, the RCEQMR of the GWR estimator is less than 0.156, the corresponding value for the regression estimator being 0.178.

9.6 Precautions to take

9.6.1 Multicolinearity and correlation between coefficients

Detecting colinearity

In order to estimate the numerous coefficients of a Geographically Weighted Regression, the weighted least squares technique imposes many constraints on the regression parameters (Leung et al. 2000). These constraints can link the GWR coefficients and create multicolinearity problems. The multicolinearity between the coefficients may be responsible for great instability in the coefficients (change of sign when adding a new variable into regression), the counter-intuitive sign of

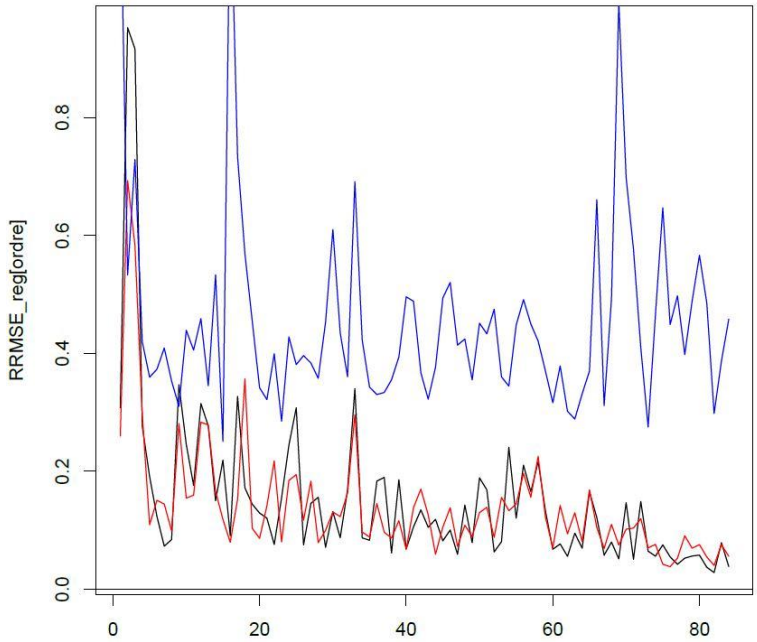

Figure 9.14 – RCEQMR (RRMSE on the figure) of the Horwitz-Thompson estimator (in blue), of the regression estimator (in black) and the estimator by GWR (in red), according to the IRISes, ranked by increasing size

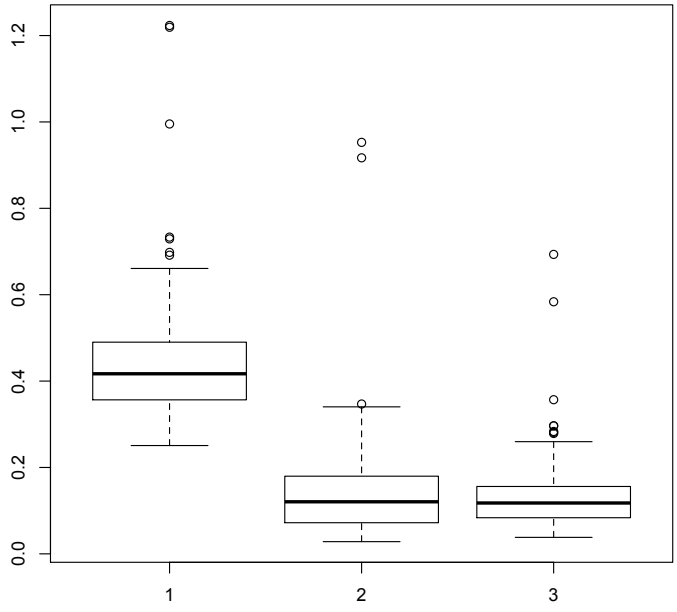

Figure 9.15 – Box-plot of the RCEQMR of the Horwitz-Thompson Estimator (1), of the regression estimator(2) and of the estimator by GWR (3)

one of the coefficients of regression, or high standard errors of the parameters (Wheeler et al. 2005). If the data correlation structure is heterogeneous in space, some regions may show colinearity between their variables, while others will not.

Function `gwr.collin.diagno` of package *GW model* allows to implement several types of colinearity detection, in particular local correlations between pairs of coefficients and the variance inflation factors (VIF) for each coefficient. These elements are detailed in Gollini et al. 2013 where examples of application with R are presented.

Box 9.6.1 — Variance inflation factor: VIF. Let R_j^2 be the coefficient of determination of the regression of variable X_j on the $p-1$ other variables.

$$VIF_j = \frac{1}{1-R_j^2}$$

If R_j^2 tends towards 1, VIF_j tends toward $+\infty$ hence the term "variance inflation". In general, the literature considers that there is a problem of multicolinearity when a VIF is greater than 10, or when the average of the VIFs is greater than 2 (Chatterjee et al. 2015).

Taking colinearity into account

One method for reducing the colinearity problems implemented in package *GW Model* is ridge regression. The principle is to increase the weight of the diagonal elements of the variance-covariance matrix to reduce the weight of the non-diagonal elements (which contain the terms of colinearity). In the general case, it can be written that:

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T Y \tag{9.23}$$

The disadvantage of this method is that $\hat{\beta}$ is biased and standard errors are no longer available.
In the case of Geographically Weighted Regression, a local ridge regression can be defined, such that:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X + \lambda I(u_i, v_i))^{-1} X^T W(u_i, v_i) Y \tag{9.24}$$

$\lambda I(u_i, v_i)$ is the value of λ at location (u_i, v_i). It is also possible to use a statistical criterion such as the cross validation score to choose the bandwidth of the local ridge regression.

9.6.2 Interpreting the parameters

The multiple testing problem

When estimating a Geographically Weighted Regression, the result is, at each point, an evaluation of the significance of each coefficient thanks to the calculated t-values. For each coefficient, there are as many t-values as points at which they were estimated. We then come up against the problem of multiple testing presented in Chapter 3, in the case of local spatial autocorrelation indicators.

If we estimate the significance of a coefficient in 100 locations with a significance threshold defined at 95%, we expect to estimate the coefficient as significative in at least 5 locations, simply because of the statistical principle of the the test, regardless of any actual correlation between the dependent variable and the explanatory variables. To remedy this problem, a Bonferroni adjustment method can be used, which will increase the value of the threshold beyond which the result of the local test will be judged non-significant - at a constant global significance level. However, adjustment methods have the disadvantage of being often too restrictive, which may lead to some coefficients being judged non-significant when they are actually significant.

Brunsdon et al. 1998 advise caution when using the t-values produced when estimating a GWR. They consider that an area with a large proportion of locally very different coefficients is a better indicator of local non-stationarity than a surface where only a small proportion of coefficients exceeds a significant value.

Effect of the local context or incorrect specification

Before interpreting local coefficient values as characteristics of the local context, it is important to explore the possibility of a poor model specification. For example, the fact that the influence of having a garage on a property price depends on the location may be due to the fact that the density of public parking spaces varies in space, or that the hedonic model is poorly specified.

Interpreting the local constant

In a Geographically Weighted Regression, the constant can vary locally. It may therefore capture all the explanatory power of the exogenous variables, particularly when they have a much more marked influence in certain locations (phenomenon of spatial clustering). In this case, the explanatory variables will appear to be non-significant. If such a phenomenon is suspected, a so-called "mixed" Geographically Weighted Regression can be used, in which the constant does not vary.

Conclusion

GWR which was first suggested in 1998 (Brunsdon et al. 1998) has been the subject of numerous practical applications, in particular in geographic and epidemiological studies. The theoretical foundations have been significantly developed. If some authors have highlighted certain limitations of the method, particularly colinearity problems (Wheeler et al. 2005, Griffith 2008), GWR is now an integral part of spatial analysis tools. It is presented in general works (Waller et al. 2004, Schabenberger et al. 2017, Lloyd 2010, Fischer et al. 2009) as well as in spatial econometrics textbooks (Arbia 2014). Extensions to the method – generalised linear models – have also been proposed.

GWR can be used in two different ways. First of all, it can be used as an exploratory method to detect areas where specific spatial phenomena occur and subject them to a comprehensive study. Secondly, it can help in building a relevant model — the detection of spatial non-stationarity then becomes symptomatic of a problem in the definition of the global model. Brunsdon et al. 1998 consider that most assertions made at a global level about the spatial relationship between objects deserve to be reviewed locally using GWR to test their validity.

Spatial dependency between the error terms decreases when GWR is used, since spatial autocorrelation is sometimes the result of a non modelled instability in parameters (Le Gallo 2004). In addition, GWR makes it possible to calculate spatial autocorrelation indicators for a variable, conditional on the spatial distribution of other variables, which is not possible with the univariate spatial autocorrelation indicators presented in Chapter 3. We therefore encourage the joint study of spatial dependency - with spatial autocorrelation indicators - and spatial heterogeneity - with Geographically Weighted Regression.

References - Chapter 9

Arbia, Giuseppe (2014). *A primer for spatial econometrics: with applications in R*. Springer.

Brunsdon, Chris, A Stewart Fotheringham, and Martin E Charlton (1996). « Geographically weighted regression: a method for exploring spatial nonstationarity ». *Geographical analysis* 28.4, pp. 281–298.

Brunsdon, Chris, Stewart Fotheringham, and Martin Charlton (1998). « Geographically weighted regression ». *Journal of the Royal Statistical Society: Series D (The Statistician)* 47.3, pp. 431–443.

Chambers, Ray and Robert Clark (2012). *An introduction to model-based survey sampling with applications*. Vol. 37. OUP Oxford.

Chandra, Hukum, Ray Chambers, and Nicola Salvati (2012). « Small area estimation of proportions in business surveys ». *Journal of Statistical Computation and Simulation* 82.6, pp. 783–795.

Chatfield, Chris (2006). « Model uncertainty ». *Encyclopedia of Environmetrics*.

Chatterjee, Samprit and Ali S Hadi (2015). *Regression analysis by example*. John Wiley & Sons.

Fischer, Manfred M and Arthur Getis (2009). *Handbook of applied spatial analysis: software tools, methods and applications*. Springer Science & Business Media.

Gollini, Isabella et al. (2013). « GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models ». *arXiv preprint arXiv:1306.0413*.

Griffith, Daniel A (2008). « Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR) ». *Environment and Planning A* 40.11, pp. 2751–2769.

Huber, Peter (1981). « J. 1981. Robust Statistics ». *New York: John Wiley*.

Le Gallo, Julie (2004). « Hétérogénéité spatiale ». *Économie & prévision* 1, pp. 151–172.

Leung, Yee, Chang-Lin Mei, and Wen-Xiu Zhang (2000). « Statistical tests for spatial nonstationarity based on the geographically weighted regression model ». *Environment and Planning A* 32.1, pp. 9–32.

Lloyd, Christopher D (2010). *Local models for spatial analysis*. CRC press.

Schabenberger, Oliver and Carol A Gotway (2017). *Statistical methods for spatial data analysis*. CRC press.

Waller, Lance A and Carol A Gotway (2004). *Applied spatial statistics for public health data*. Vol. 368. John Wiley & Sons.

Wheeler, David and Michael Tiefelsdorf (2005). « Multicollinearity and correlation among local regression coefficients in geographically weighted regression ». *Journal of Geographical Systems* 7.2, pp. 161–187.