geographical analysis

*Geographical Analysis* (2015) ••, 1–22

# Geographical and Temporal Weighted Regression (GTWR)

A. Stewart Fotheringham[1], Ricardo Crespo[2], Jing Yao[3]

[1]GeoDa Center for Geospatial Analysis and Computation, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287, USA, [2]Facultad de Ingeniería y Administración, Universidad Bernardo O'Higgings, Santiago, Chile, [3]Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow G12 8RZ, UK

*Both space and time are fundamental in human activities as well as in various physical processes. Spatiotemporal analysis and modeling has long been a major concern of geographical information science (GIScience), environmental science, hydrology, epidemiology, and other research areas. Although the importance of incorporating the temporal dimension into spatial analysis and modeling has been well recognized, challenges still exist given the complexity of spatiotemporal models. Of particular interest in this article is the spatiotemporal modeling of local nonstationary processes. Specifically, an extension of geographically weighted regression (GWR), geographical and temporal weighted regression (GTWR), is developed in order to account for local effects in both space and time. An efficient model calibration approach is proposed for this statistical technique. Using a 19-year set of house price data in London from 1980 to 1998, empirical results from the application of GTWR to hedonic house price modeling demonstrate the effectiveness of the proposed method and its superiority to the traditional GWR approach, highlighting the importance of temporally explicit spatial modeling.*

## Introduction

Space and time are two fundamental dimensions providing the framework for all human activities, social events, and environmental processes. Spatiotemporal modeling has long been an important research focus in the field of geographical information science (GIScience) (Cressie 1993; Cressie and Wilkie 2011). An early example is the theoretical framework of time geography proposed by Hägerstrand (1970). With increasingly abundant spatiotemporal data becoming available, such as the trajectories collected by global positioning systems (GPS) and snapshots of remote-sensing images, there is an increasing interest in spatiotemporal modeling. Examples include exploring spatiotemporal patterns of human behavior (Chen et al. 2011; Kwan 2000, 2004), crime activities (Brunsdon, Corcoran, and Higgs 2007; Nakaya and Yano 2010), and

Correspondence: Stewart Fotheringham, GeoDa Center for Geospatial Analysis and Computation, School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85287, USA
e-mail: Stewart.Fotheringham@asu.edu

Submitted: October 24, 2014. Revised version accepted: October 27, 2014.

doi: 10.1111/gean.12071

disease outbreaks (Takahashi et al. 2008), as well as new methods to analyze and visualize space–time data (Andrienko et al. 2010; Demšar and Virrantaus 2010; Rey and Janikas 2010). However, enhancing the capability of spatiotemporal analysis and modeling in the current GIS environment still remains a major challenge, particularly in the era of big data (Goodchild 2013). Therefore, new methodologies need to be developed to encourage space–time thinking, to discover useful spatiotemporal information and knowledge in space–time data, and thereby better understand social and environmental dynamics. To this end, this article approaches the problem of spatiotemporal modeling from a local perspective by extending geographically weighted regression (GWR) to a temporal dimension.

GWR is a spatial statistical method for modeling spatially heterogeneous processes that allows the relationships between a response and a set of covariates to vary across geographic space (Brunsdon, Fotheringham, and Charlton 1996, 1998; Fotheringham, Brunsdon, and Charlton 1996, 2002; Fotheringham, Charlton, and Brunsdon 1997). Since its introduction, GWR has been a popular tool and widely applied in a variety of disciplines and areas, such as geology (Atkinson et al. 2003), environment science (Mennis and Jordan 2005; Harris, Fotheringham, and Juggins 2010), hedonic house price modeling (Bitter, Mulligan, and Dall'erba 2007; Huang, Wu, and Barry 2010), landscape ecology (Buyantuyev and Wu 2010), health research (Nakaya et al. 2005; Comber, Brunsdon, and Radburn 2011), and crime studies (Malczewskia and Poetzb 2005; Wheeler and Waller 2009). A fundamental component of GWR is the spatial weight matrix by which the local spatial relationships are constructed. Usually spatial weights are defined by spatial kernel functions such as Gaussian or bi-square functions (Fotheringham, Brunsdon, and Charlton 2002) in which larger weights are assigned to closer observations according to the well-known Tobler's First Law of Geography: "everything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236). Thus, localized regression models are fitted and calibrated by incorporating distance-decay effects in space: in essence GWR is a method of "borrowing" data from surrounding locations.

Beyond space, however, time is also an essential dimension pertaining to social activities and environmental processes as mentioned previously. Many variables of interest in geosciences are observed not only across space, but also over time. Temporal data can provide valuable information on the dynamics of the underlying spatial process and enable the forecasting of relevant variables, which is of interest in research areas such as the diffusion of contagious diseases, the spread of air or water pollution, and the expansion of urban sprawl. Not surprisingly, extensive efforts have been devoted to incorporating the temporal dimension into spatial regression (Pace et al. 1998, 2000; Anselin 1999; Elhorst 2003; Gelfand et al. 2004; Giacinto 2006; Crespo, Fotheringham, and Charlton 2007; Cressie and Wilkie 2011). Most of these, however, approach the problem from a global modeling perspective where temporal effects are assumed to be constant over space. For example, Pace et al. (2000) propose a spatiotemporal autoregressive model, and account for spatial and temporal dependence in the error terms in particular.

Exceptions are the work of Elhorst (2003) and Giacinto (2006) where parameter estimates are allowed to drift over space although these techniques are preferably utilized to model processes at an aggregated level such as states, counties or regions of a country. Crespo, Fotheringham, and Charlton (2007) extended GWR by developing spatiotemporal bandwidths that accounts for varying local spatial effects across time. Subsequently, Huang, Wu, and Barry (2010) attempt to incorporate temporal effects into GWR by integrating both temporal and spatial information in the weighting matrices, which has been improved and integrated with a spatial autogressive (SAR) model by Wu, Li, and Huang (2014). A similar concept was also proposed by

Yu (2014). A spatiotemporal weight matrix was constructed using spatiotemporal distances between observations. In other words, spacetime distance-decay functions were employed to measure spatiotemporal relationships among observations. Although addressed in the literature to some extent (Wu et al. 2013; Wrenn and Sam 2014), calculating distance in three dimensions for this method remains a challenge because a sole measure integrating spatial and temporal distances can be misleading as location and time are usually measured at different scales. Another issue is that it is unclear how to adjust bandwidths to account for variations in spatiotemporal processes over space and time. This article will therefore contribute to the spatiotemporal extension of GWR through developing the bandwidth concept in GWR by including a temporal dimension. Specifically, a spatiotemporal kernel function is proposed and a procedure for choosing the optimal spatiotemporal bandwidth is developed.

The remainder of the article is organized as follows. A temporal extension of GWR (geographical and temporal weighted regression, GTWR[1]) involving a spatiotemporal bandwidth is presented in the next section. Both the model formulation and the estimation of GTWR are detailed, focusing on spatiotemporal kernel function definition and spatiotemporal bandwidth optimization. This is followed by an empirical study of hedonic house price modeling in London from 1980 to 1998 using GTWR. Further, the performance of GTWR is examined by comparison with basic GWR models. The article ends with discussion and conclusions, highlighting the effectiveness and potential superiority of GTWR in local spatiotemporal modeling.

## GTWR

### Model formulation

To define GTWR, it is helpful to give the generic GWR formulation first, which is illustrated by equation (1) (Fotheringham, Brunsdon, and Charlton 2002).

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{1}$$

where $i$ is the index of a spatial point with $(u_i, v_i)$ denoting its coordinates. Accordingly, $y_i$, $x_{ik}$, $\varepsilon_i$ are dependent variable, $k$th independent variable and error term for the $i$th observation (point), respectively. The distinct character of GWR is that the parameters $\beta_k(u_i, v_i)$ are allowed to vary across space to measure spatially nonstationary relationships. If using matrix representation, the estimated parameters can be expressed by equation (2):

$$\hat{\beta}(u_i, v_i) = \left(\boldsymbol{X}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{W}(u_i, v_i)\boldsymbol{y} \tag{2}$$

where $\boldsymbol{W}$ is a diagonal matrix with elements representing the geographical weights of each observation of the $i$th point.

When the data points are collected across time and space at a set of locations $S_t = [s_1, s_2, \ldots, s_{n_t}]$, where $n_t$ is the number of locations where the data are observed at time period $t$, the GWR model in equation (1) can still be used to derive local estimates, but in this case by incorporating the data measured at prior time periods $t-1$, $t-2$, . . .. $t-q$, with $q$ being the number of time lags in addition to those from the same time period. The parameter estimation still can be obtained by equation (2).

The distinction, however, lies in the weight matrix $\boldsymbol{W}$, which is constructed in a different way in order to capture both spatial and temporal effects from observations nearby both in space and

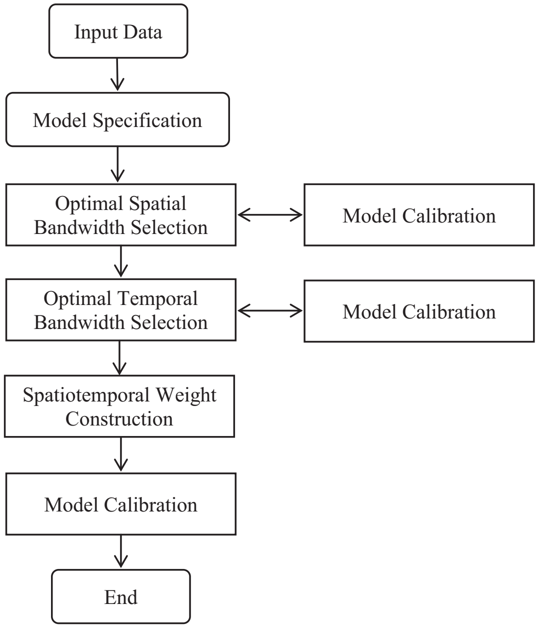


**Figure 1.** Scheme of geographical and temporal weighted regression (GTWR) model.

time. Fig. 1 summarizes the general routine to estimate local parameters using equation (2) for GTWR. First, a traditional GWR model is specified using the input data according to equation (1). Then an optimal spatial bandwidth is specified for each time period based on a goodness-of-fit criterion such as cross-validation (CV) or Akaike information criterion (AIC). Using the optimal spatial bandwidth, the optimal temporal bandwidth is determined, again based on CV or AIC. Once both optimal spatial and temporal bandwidths are derived, they can be used to construct the spatiotemporal weight matrix $\boldsymbol{W}$, which allows local parameters to be estimated using equation (2). It should be noted that both spatial and temporal bandwidth optimization involves heavy computation as those steps require repeated temporary model calibrations. The remainder of this section will focus on the structure of the spatiotemporal weight matrix $\boldsymbol{W}$ and the section entitled *Bandwith optimization* will detail the procedures for optimally selecting spatial and temporal bandwidths.

As mentioned, to solve equation (2), it is essential to define a kernel function to obtain a geographical weight matrix $\boldsymbol{W}$ for each observation. Of course, traditional kernel functions can still be employed to spatially weight data points from previous time periods, but in this way, temporal relationships among points are assumed to be spatially stationary, which is not necessarily true. Thus, new kernel functions are needed to account for both spatial and temporal relationships among observations. In this study, a spatiotemporal kernel function is proposed, which consists of mixed spatial and time-decay bandwidths. In this type of mixed kernel function, weights given to data points are calculated not only based on the distance between the regression point[2] and each data point, but also based on the separation in time between them. The use of a temporal bandwidth assumes that local estimates are not constant over time at a given location $i$ because if they were, local estimates could be derived using only a spatial bandwidth in the traditional way to calibrate a model by GWR without the inclusion of a temporal bandwidth. As

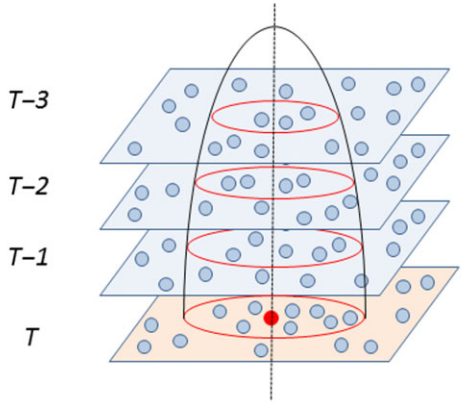


**Figure 2.** An example of time-decay spatiotemporal bandwidth.

a result, a spatiotemporal bandwidth can be thought as an extension of the traditional GWR method for exploring spatial and temporal nonstationary relationships.

Equation (3) shows an example of a spatiotemporal weighting function ($w^t_{ijS,T}$) specific for data points located at time $t$ according to a general form of a spatiotemporal kernel function where weights are given by a spatial kernel function ($k_S$) with $d_{s_{ij}}$ being the Euclidean distance between the regression point $i$ and a data point $j$. Note that the data point $j$ can be located at any set $S_t$, $S_{t-1}$, . . . , $S_{t-q}$. The spatial bandwidth is given by $b_S$ while the temporal kernel is given by $k_T$ where $d_{t_{ij}}$ is the distance in time between the regression point $i$ and the data point $j$ with $b_T$ being the temporal bandwidth. In this study, a time-decay temporal bandwidth is proposed, that is a temporal bandwidth in which data points located closer in time to the regression point have more influence on local estimates at the regression point $i$ than those located farther away in time. A description of a possible time-decay spatiotemporal bandwidth is given by Fig. 2, where the regression time period is $T$ and a temporal bandwidth 3 is considered ($T-1$, $T-2$, and $T-3$). As can be seen, the spatial bandwidth becomes smaller as the observations are further away from the regression point in time. In this way, the temporal bandwidth operates in a similar way to the spatial bandwidth in the sense that $b_T$ provides some control on the range of the "circle of influence"[3] in the geographical data over time.

$$w^t_{ijS,T} = k_S\left(d_{s_{ij}}, b_S\right)\times k_T\left(d_{t_{ij}}, b_T\right) \tag{3}$$

From this weighting scheme, a spatiotemporal version of the $\boldsymbol{W}$ diagonal weight matrix used for the GWR calibration can be generated. In this case $\boldsymbol{W}$ will be a ($n_{\Sigma_T} \times n_{\Sigma_T}$) diagonal matrix whose elements are given by equation (3), and where $n_{\Sigma_T}$ is the total number of data points from time periods $t, \ldots.., t-q$ used to calibrate the model. Once the weight matrix is obtained, local estimates at the regression point $i$ are derived using the traditional method specified in equation (2).

In equation (3), it is assumed that a unique spatial bandwidth $b_S$ is derived and applied to all data points in the data set. However, there is no reason to assume a priori a constant spatial bandwidth over time. In fact, other arrangements of spatial bandwidth are also possible to fit the data. For example, it seems reasonable to assume that spatial bandwidths become smaller as data points are located farther away in time from the regression point because such data points would induce greater bias into the model, as the case shown in Fig. 2. Alternatively, it can be though that spatial bandwidths become larger as data points are located farther away in time to the regression point in order to compensate for the lower temporal weight given to data points by the time-decay bandwidth. A next step in the model definition is therefore to use a set of segregated spatial bandwidths over time which must be estimated along with the temporal bandwidth to fit the data. Thus, equation (3) is extended to equation (4), in which different bandwidths for each time period are possible.

$$w^t_{ijS,T} = k_S\left(d_{s_{ij}}, b_{St}\right) \times k_T\left(d_{t_{ij}}, b_T\right) \tag{4}$$

where terms are defined as earlier and $b_{St}$ is the spatial bandwidth specific to time $t$. Under this weighting scheme, the $n_{\Sigma_T}$ elements of the diagonal weight matrix are obtained using a different spatial bandwidth ($b_{St}$) according to the time when the data points are collected. The $n_{\Sigma_T}$ diagonal elements can be arranged into $q + 1$ sets of $n_t$, $n_{t-1}$, . . .. $n_{t-q}$ elements where $n_t$ corresponds to the number of data points for time $t$. Thus, the elements of the first set, which make up the first $n_t$ elements of the diagonal, are derived from equation (4) using a spatial bandwidth $b_{St}$ and a temporal bandwidth $b_T$. Similarly, the elements of the second set, which make up the second $n_{t-1}$ elements of the diagonal, are derived from equation (4) using a spatial bandwidth $b_{S(t-1)}$ and the same temporal bandwidth $b_T$. This process is repeated in this same way to derive the elements of the other $q - 1$ sets. Thus the weight matrix for the case of the segregated spatial bandwidths over time is represented as expressed in equation (5):

$$W_i = \begin{bmatrix} w^t_{i1_{S,T}} & 0 & \cdots & & & & & & & & 0 \\ 0 & w^t_{i2_{S,T}} & & & & & & & & & 0 \\ 0 & & \ddots & & & & & & & & \vdots \\ \vdots & & & w^t_{in_tS,T} & & & & & & & \vdots \\ \vdots & & & & w^{t-1}_{i1_{S,T}} & & & & & & \vdots \\ \vdots & & & & & w^{t-1}_{i2_{S,T}} & & & & & \vdots \\ \vdots & & & & & & \ddots & & & & \vdots \\ \vdots & & & & & & & w^{t-1}_{in_{t-1}S,T} & & & \vdots \\ \vdots & & & & & & & & \ddots & & \vdots \\ \vdots & & & & & & & & w^{t-q}_{i1_{S,T}} & & \vdots \\ \vdots & & & & & & & & & w^{t-q}_{i2_{S,T}} & \vdots \\ \vdots & & & & & & & & & \ddots & \vdots \\ 0 & \cdots & & & & & & & & & w^{t-q}_{in_{t-q}S,T} \end{bmatrix} \tag{5}$$

where $w^t_{i1_{S,T}}, \ldots, w^t_{in_tS,T}$ are the weights given to $n_t$ data points located at time $t$ using a spatial bandwidth $b_{St}$ and a temporal bandwidth $b_T$ in the estimation.

**Bandwidth optimization**

As with the traditional GWR mechanism, the optimal spatial bandwidth ($b^*_S$) as well as the optimal temporal bandwidth ($b^*_T$) can be calculated by minimizing either the CV or the AIC function of the model in order to obtain a set of local estimates with optimal bias-variance trade-off. For example, the CV[4] function in equation (6) can be used to obtain $b^*_S$ and $b^*_T$:

$$CV(b_S, b_T) = \sqrt{\sum_{i=1}^{n} \left(y_i - \hat{y}_{-i}(b_S, b_T)\right)^2 / n} \tag{6}$$

where $\hat{y}_{-i}$ is the fitted value for $y_i$ with point $i$ excluded from the calibration process. To account for the various spatial bandwidths used in different time periods, equation (6) can be further rewritten as in equation (7) in which the $q + 1$ spatial bandwidths and the temporal bandwidth must be estimated simultaneously in order to minimize the CV function.

$$CV\left(b_{St}, b_{S(t-1)}, \ldots, b_{S(t-q)}, b_t\right) = \sqrt{\sum_{i=1}^{n} \left(y_i - \hat{y}_{-i}\left(b_{St}, b_{S(t-1)}, \ldots, b_{S(t-q)}, b_T\right)\right)^2 / n} \tag{7}$$

The remainder of this section will detail the procedure of selecting the optimal spatial/temporal bandwidth. A Gaussian spatiotemporal kernel function is taken as an example here. Accordingly, the weight can be defined as shown in equation (8).

$$w^t_{ijS,T} = exp\left(-\frac{d^2_{s_{ij}}}{b^2_{St}}\right) * exp\left(-\frac{d^2_{t_{ij}}}{b^2_T}\right) \tag{8}$$

Time period $t$ is used as an example to show by means of the nine steps listed later how to derive optimal temporal and spatial bandwidths, though the mechanism can be extended to all possible time periods in the regression model.

(1) Suppose the temporal bandwidth $b_T$ is set to 1 (e.g. year, month or day) temporal unit;

(2) The spatial bandwidth for time period $t$ ($b_{St}$) is calculated using data points only from time $t$. In this case, $d^2_{t_{ij}}$ is zero and the spatiotemporal Gaussian kernel becomes $w^t_{ijS,T} = exp\left(-\frac{d^2_{s_{ij}}}{b^2_{St}}\right)$. Next, using this weighting scheme GWR is utilized to calibrate the regression model on data points from time $t$. By minimizing the CV function, the optimal spatial bandwidth $b^*_{St}$ is obtained for the first $n_t$ diagonal elements of the weight matrix in equation (5) according to equation (8).

(3) To obtain the second set of elements for the diagonal weight matrix, data points from time $(t - 1)$ are incorporated into the model. Thus, GWR is used to calibrate the regression model using data from $t$ and $t - 1$ for the regression points at time $t$. Data points from $t$ are weighted according to the first set of elements of the diagonal matrix obtained in equation (2), that is $b^*_{St}$ is kept fixed. Data points from $(t - 1)$ are weighted using the Gaussian spatiotemporal kernel defined by equation (8) from which the optimal spatial bandwidth $b^*_{S(t-1)}$ is to be estimated. As for time $(t - 1)$, $d^2_{t_{ij}} = 1$, the Gaussian spatiotemporal kernel used for data points from $(t - 1)$ becomes $w^{t-1}_{ijS,T} = exp\left(-\frac{d^2_{s_{ij}}}{b^2_{S(t-1)}}\right) * exp\left(-\frac{1}{1}\right)$. As with point 2), the optimal spatial bandwidth $b^*_{S(t-1)}$ is derived by minimizing the CV function for data from time $t$. Thus, the second set of $n_{t-1}$ diagonal elements of the weight matrix in equation (5) are derived by inputting $b^*_{S(t-1)}$ in equation (8) with $t = t - 1$.

(4) Similarly, the third set of elements of the diagonal weight matrix is obtained by incorporating data points from $t - 2$. In this case, data points from $t$ and $t - 1$ are weighted using the optimal spatial bandwidths $b^*_{S(t-1)}$ and $b^*_{St}$ as specified in equations (2)–(3), that is these spatial bandwidths are kept constant. The weighting scheme used for data points

from $t-2$ is derived from the equation (8) where $b^*_{S(t-2)}$ is to be estimated. As for time $t-2$, $d^2_{t_{ij}}=2$, the spatiotemporal Gaussian kernel for time $t-2$ becomes $w^{t-2}_{ij_{S,T}}=exp\left(-\frac{d^2_{s_{ij}}}{b^2_{S(t-2)}}\right)*exp\left(-\frac{2}{1}\right)$. Again, the optimal spatial bandwidth $b^*_{S(t-2)}$ is derived by minimizing the CV function for data from time $t$. Finally, once the optimal bandwidth for $t-2$ is obtained, the third set of the diagonal weight matrix is derived from equation (8) with $t=t-2$.

(5) The process described earlier is repeated incorporating one-by-one data points from time $t-3, t-4, \ldots, t-$ q to derive optimal spatial bandwidths for these time periods along with the corresponding sets of diagonal elements of the weight matrix.

(6) Once the diagonal matrix for data points from $t$ to $t-$ q is obtained, GWR is used to calibrate the regression model at points from time $t$ using the weighting scheme given by the diagonal weight matrix from equation (5). From this GWR calibration, a CV score is obtained which will be specific to the temporal bandwidth assumed to be 1 temporal unit in equation (1). This CV score is to be referred to as $\mathrm{CV}_{b_T=1}$.

(7) The process described from points (2) to (6) is repeated for the other $q-1$ possible temporal bandwidths according to the number of time lags in the model, that is for $b_T$ equal to 2, 3, 4, . . . , or $q$ temporal unit in the past. Thus, for each temporal bandwidth used to calibrate the model, a CV score is obtained, say, $\mathrm{CV}_{b_T=1}$, $\mathrm{CV}_{b_T=2}$, $\mathrm{CV}_{b_T=3}$, . . . , and $\mathrm{CV}_{b_T=q}$.

(8) The optimal temporal bandwidth ($b^*_T$) is the one for which the minimum CV score is obtained. The selection of an optimal temporal bandwidth yields the final set of optimal spatial bandwidths: $[b^*_{St}, b^*_{S(t-1)}, b^*_{S(t-2)}, \ldots, b^*_{S(t-q)}]_{b^*_T}$.

(9) Finally, local estimates for points located at time $t$ are estimated using equation (2) in which the diagonal elements of W are derived using the weighting scheme given by the optimal temporal and spatial bandwidth from point (8).

## Application of GTWR

In this section, the spatiotemporal GWR approach described earlier is employed to calibrate local hedonic price models in London using data from 1980 to 1998. First, the data and study area is presented. This is followed by a formal expression of the hedonic house price model. Further, the selection of optimal spatial and temporal bandwidths is detailed. Finally, model calibration including local parameter estimation and residual analysis is provided.

### Data and study area

The data selected for this study consist of a set of annual house prices and their determinants in London covering the time period from 1980 to 1998 obtained through the Nationwide Building Society. A sample of approximately 17,433 observations is utilized to develop and to calibrate a hedonic price model by a spatiotemporal GWR. To facilitate computation of this complex model, 1,000 observations were randomly selected from the original sample for each year (the only exception being in 1995 when only 433 observations exist).

In order to make the GWR/GTWR results comparable across time, house prices in each year are inflated to 1998 prices using the consumer price index in the U.K. obtained through the Office for National Statistics. Table 1 summarizes for each year the number of observations available, the sample size, and average of house prices and inflated house prices. The evolution of house prices in London over time can be observed in Fig. 3. As can be noticed, an increasing trend in

**Table 1.** Sample Size, and House Prices in London, 1980–1998

| Year | Number of observations | Sample size | House price average (£) | Inflated house price average (£) |
|---|---|---|---|---|
| 1980 | 14,233 | 1,000 | 30,475 | 58,896 |
| 1981 | 14,216 | 1,000 | 30,244 | 53,671 |
| 1982 | 17,728 | 1,000 | 30,194 | 49,023 |
| 1983 | 17,417 | 1,000 | 35,112 | 53,730 |
| 1984 | 18,803 | 1,000 | 40,490 | 59,520 |
| 1985 | 16,342 | 1,000 | 47,346 | 67,115 |
| 1986 | 19,990 | 1,000 | 55,526 | 76,345 |
| 1987 | 8,768 | 1,000 | 71,871 | 96,691 |
| 1988 | 13,617 | 1,000 | 79,226 | 103,581 |
| 1989 | 4,738 | 1,000 | 80,160 | 101,356 |
| 1990 | 4,844 | 1,000 | 93,453 | 113,730 |
| 1991 | 5,964 | 1,000 | 85,390 | 100,016 |
| 1992 | 5,545 | 1,000 | 77,028 | 87,255 |
| 1993 | 3,470 | 1,000 | 63,779 | 70,692 |
| 1994 | 3,901 | 1,000 | 62,070 | 66,988 |
| 1995 | 433 | 433 | 80,022 | 84,587 |
| 1996 | 11,365 | 1,000 | 87,159 | 90,236 |
| 1997 | 11,947 | 1,000 | 98,827 | 100,507 |
| 1998 | 11,282 | 1,000 | 119,703 | 119,703 |

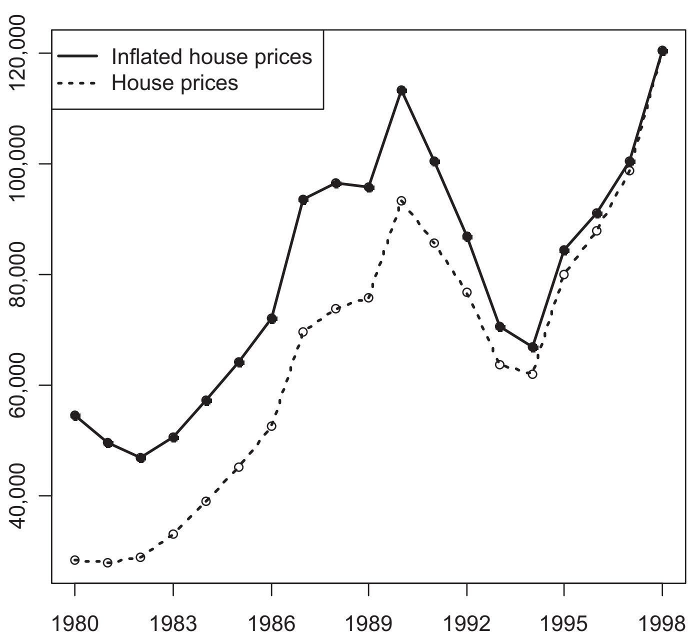


**Figure 3.** Time series of average house prices in London, 1980–1998.

house price from 1980 to 1990 is followed by a 4-year crash in the housing market. From 1994 house prices recovered reaching in the next 4 years approximately the same level as before the crash.

**Model specification**

In this study we follow the hedonic price model described in Fotheringham, Brunsdon, and Charlton (2002), in which house prices in London are regressed on three groups of explanatory variables: (1) *structural attributes*: floor area, type of the property, date of construction, number of bathrooms, provision of garage, and central heating; (2) *neighborhood attributes*: proportion of workforce in professional or managerial occupations and rate of unemployment at the census *output area*[5] where the property is located; and (3) *locational attributes*: distance to city centre.

The functional form of the model is represented as:

$$\begin{aligned} P_i = {} & \beta_0 + \beta_1 FLRAREA_i + \beta_2 BLDPWW1_i + \beta_3 BLDPOSTW_i + \beta_4 BLD60S_i \\ & + \beta_5 BLD70S_i + \beta_6 BLD80S_i + \beta_7 BLD90S_i + \beta_8 TYPDETCH_i + \beta_9 TYPTRRD_i \\ & + \beta_{10} TYPBNGLW_i + \beta_{11} TYPFLAT_i + \beta_{12} GARAGE_i + \beta_{13} CENHEAT_i + \beta_{14} BATH2_i \\ & + \beta_{15} PROF_i + \beta_{16} UNEMPLOY_i + \beta_{17} FLRDETCH_i + \beta_{18} FLRFLAT_i + \beta_{19} FLRBNGLW_i \\ & + \beta_{20} FLRTRRD_i + \beta_{21} \log_e(DISTCL_i) + \varepsilon_i \end{aligned} \tag{9}$$

where $P_i$ is the price in pounds sterling of the property; FLRAREA is the floor area of the property in $m^2$; BLDxxx is a set of dummy or indicator variables that depict the age of the property as follows:

BLDPWW1 is 1 if the property was built prior to 1914, 0 otherwise;
BLDPOSTW is 1 if the property was built between 1940 and 1959, 0 otherwise
BLD60S is 1 if the property was built between 1960 and 1969, 0 otherwise
BLD70S is 1 if the property was built between 1970 and 1979, 0 otherwise
BLD80S is 1 if the property was built between 1980 and 1989, 0 otherwise
BLD90S is 1 if the property was built between 1990 and 1999, 0 otherwise

TYPxxx is a set of dummy variables that depict the type of house as follows:

TYPDETCH is 1 if the property is detached (i.e. it is a stand-alone house), 0 otherwise;
TYPTRRD is 1 if the property is in a terrace of similar houses (commonly referred to as a "row house" in the USA), 0 otherwise;
TYPBNGLW is 1 if the property is a bungalow (i.e. it has only one floor), 0 otherwise;
TYPFLAT is 1 if the property is a flat (or "apartment" in USA), 0 otherwise;
GARAGE is 1 if the house has a garage, 0 otherwise;
CENTHEAT is 1 if the house has a central heating, 0 otherwise;
BATH2 is 1 if the house has two or more bathrooms, 0 otherwise;
PROF[6] is the proportion of the workforce in professional or managerial occupations in the census output area in which the house is located;
UNEMPLOY is the rate of unemployment in the census output area in which the house is located; and FLRxxx is a set of interaction terms where;
FLRDETCH = FLRAREA × TYPDETCH
FLRFLAT = FLRAREA × TYPFLAT
FLRBNGLW = FLRAREA × TYPBNGLW
FLRTRRD = FLRAREA × TYPTRRD

DISTCL is the straight-line distance from the property to the center of London (taken here to be Nelson's column in Trafalgar Square) measured in m; $\log_e$ denotes a natural logarithm; and $\beta$ denotes a parameter to be estimated.

**Bandwidth selection**

For the spatiotemporal version of GWR, adaptive spatial bandwidths are used in this study, with values ranging between 0 and 1 which specify the proportion of data points to be included in the local model calibration. The spatiotemporal weights are defined using a Gaussian kernel function as shown in (8). As mentioned, the data set contains a 19-year set of house price in London from 1980 to 1998. Thus, if the spatiotemporal GWR is to be used to calibrate the model at a regression point *i* located in year 1998, the CV function from equation (7) will consist of a set of 20 variables: $[b_{S1998}, b_{S1997}, \ldots, b_{S1980}, b_T]$ corresponding to the 19-year specific spatial bandwidths and the temporal bandwidth. However, the complexity of a 20-variable objective function makes the minimization process extremely computationally demanding because of the large number of possible combinations of the 19 spatial bandwidths and the temporal bandwidth to be inputted in the CV function. For example, if an adaptive type of bandwidth is to be used, each spatial bandwidth may take values between 0 and 1, i.e., the proportion of points to be included in the model calibration. Assuming, for simplification, that the spatial bandwidths may take values from the series [0.05, 0.1, 0.15, 0.20, 0.25, . . .. . . , 0.90, 0.95, 1], this yields 20 possible values for each spatial bandwidth in the model, which leads to $19^{20}$ combinations of possible spatial bandwidth to be tested in the CV function. Also, as there are 18 possible values for the temporal bandwidth (one for each time lag) the total amount of possible values of the CV function to be computed amounts at $18{*}19^{20}$; therefore the model must be constrained.

One way to constrain the model is by reducing the number of time lags. In this study, the number of time lags will be reduced from 18 to 5 years. Because of the data availability (from 1980 to 1998), local estimates for regression points located at years 1984, 1983, 1982, and 1981 are estimated using only 4, 3, 2, and 1 time lag periods respectively. This reduction in the time lags is based on the assumption that data points located more than five years away from the regression year[7] would have little or negligible influence on the estimation of local parameters at the regression point. The number of possible values of the CV function therefore drops to $5{*}6^{20}$. Nevertheless, this number of possible values for the CV function is still high enough to make the minimization process extremely computationally demanding. To cope with this, the estimation mechanism proposed in the section entitled *Bandwith optimization* is employed in which the optimal spatial bandwidths are derived one-by-one instead of simultaneously. For a given temporal bandwidth $b_T$, this mechanism firstly estimates the optimal spatial bandwidth $b^*_{St}$ for the regression year *t* using data points from the same year *t*. Next, by keeping $b^*_{St}$ constant, data points from the first time lag period *t*-1 are incorporated into the model. Weights for data points from year *t*-1 are given by Equation (4) in which the optimal spatial bandwidth $b^*_{S(t-1)}$ is to be derived. This procedure is repeated to estimate $b^*_{S(t-2)}$, $b^*_{S(t-3)}$, $b^*_{S(t-4)}$ and $b^*_{S(t-5)}$ by progressively incorporating data points from the corresponding time lag periods.

The above process is repeated for different values of the temporal bandwidth that is for $b_T$ equals 1, 2, 3, 4 or 5 years so that a specific set of optimal spatial bandwidths is obtained for each possible value of the temporal bandwidth, say, $[b^*_{St}, b^*_{S(t-1)}, b^*_{S(t-2)}, \ldots, b^*_{S(t-q)}]_{b_T}$. Thus, a CV score for the model calibration at the regression year *t* can be obtained for each set of optimal spatial bandwidths specific to each $b_T$. An obvious selection of the optimal temporal bandwidth $b_T^*$ is the value of $b_T$ for which the minimum CV score of the model is obtained.

*Selection of the temporal bandwidth*

The optimal temporal bandwidth for each year is given in Table 2. A temporal bandwidth of value 0 would mean only a spatial bandwidth is used. As with the spatiotemporal approach, data points from the regression year along with data points from the five time lag periods are used to calibrate the model for each regression year according to equation (8). When no temporal bandwidth is used to weight data points, $d_{t_{ij}} = 0$ in equation (8) and thus the term $exp\left(-d_{t_{ij}}^2 / b_{St}^2\right)$ becomes one. That is, data points are only weighted according to the distance from the regression point and a unique spatial bandwidth $b_S$ will be used to compute the weights from equation (8). It should be noted that in this case, the CV function would be evaluated at points located only at the regression year although data points from previous years are incorporated in the model.

With regard to the optimal temporal bandwidth, by a visual examination of Table 2, it can be observed that for all years apart from 1986, 1987, and 1995 the CV function is minimized with a temporal bandwidth equal to 1 year. For 1986 and 1987, the optimal temporal bandwidths is 2 years and for 1995 it is 3 years. Following the characteristics of temporal kernels described in the section entitled *GTWR*, a temporal bandwidth equal to 1 for a regressionyear $t$ yields a temporal weighting gradient in which data points located at year $t-1$ have significantly more influence on the regression points than data points located further away in time. Similarly, temporal bandwidths equal to 2 for years 1986 and 1987 indicate that the "circle of influence" is extended in time to data points located at year $t-2$. Such temporal bandwidths are reasonable as in the housing market, appraisers pay more attention to recent trends in the overall market to set house prices (Pace et al. 1998). Thus, data points located 1 or 2 years earlier than the regression year will have more influence on the house pricing process at the regression year than those located more distance in time. The only case in which a temporal bandwidth exceeds 2 years is year 1995 for which a temporal bandwidth equal to 3 years was obtained. However, it is worth noting that only 433 observations of house prices (Table 1) and their determinants are available for 1995 in contrast to the other years for which samples of 1,000 observations were selected. A larger temporal bandwidth for this year is thereby understandable as the inclusion of more data points with higher weights will likely yield a reduction in the variance of local estimates.

*Selection of the spatial bandwidth*

Table 2 summarize for each regression year the set of optimal spatial bandwidths specific to each optimal temporal bandwidth. As mentioned, once the optimal temporal bandwidth is obtained for a regression year $t$, the set of optimal spatial bandwidths $[b_{St}^*, b_{S(t-1)}^*, b_{S(t-2)}^*, b_{S(t-3)}^*, b_{S(t-4)}^*, b_{S(t-5)}^*]$ is simultaneously derived according to equations (1)–(9) provided in the section entitled *Bandwith optimization*. Adaptive spatial bandwidths are used here as a measurement of the proportion (between 0 and 1) of data points from years $t$, $t-1$, $t-2$, $t-3$, $t-4$, or $t-5$ used to calibrate GWR at a regression point $i$ located at the regression year $t$. In general, the set of optimal spatial bandwidths for each regression year $t$ in Table 2 exhibits a decreasing trend as data points are located farther away in time from the regression year, that is, $b_{St}^* > b_{S(t-1)}^* > b_{S(t-2)}^* > b_{S(t-3)}^* > b_{S(t-4)}^* > b_{S(t-5)}^*$ . This trend can probably be caused by the bias introduced to local estimates at a given location $i$ (located at the regression year $t$) by data points located at years $t-1, t-2, \ldots, t-5$ as the house pricing process may differ from year to year. As a result, it is expected that a data point located at the regression year $t$ and at a distance, say, $d$ from the

**Table 2.** Optimal Temporal and Spatial Bandwidths: 1981–1998

| | 1998 | 1997 | 1996 | 1995 | 1994 | 1993 | 1992 | 1991 | 1990 | 1989 | 1988 | 1987 | 1986 | 1985 | 1984 | 1983 | 1982 | 1981 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $b_T^* =$ | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| $b_{S1998}^* =$ | 0.106 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1997}^* =$ | 0.092 | 0.134 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1996}^* =$ | 0.018 | 0.021 | 0.557 | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1995}^* =$ | — | 0.020 | 0.383 | 0.961 | — | — | — | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1994}^* =$ | — | — | 0.042 | 0.070 | 0.637 | — | — | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1993}^* =$ | — | — | — | 0.018 | 0.303 | 0.616 | — | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1992}^* =$ | — | — | — | 0.227 | 0.054 | 0.013 | 0.760 | — | — | — | — | — | — | — | — | — | — | — |
| $b_{S1991}^* =$ | — | — | — | 0.012 | — | 0.011 | 0.082 | 0.142 | — | — | — | — | — | — | — | — | — | — |
| $b_{S1990}^* =$ | — | — | — | 0.001 | — | — | 0.080 | 0.016 | 0.078 | — | — | — | — | — | — | — | — | — |
| $b_{S1989}^* =$ | — | — | — | — | — | — | — | 0.235 | 0.036 | 0.705 | — | — | — | — | — | — | — | — |
| $b_{S1988}^* =$ | — | — | — | — | — | — | — | — | 0.028 | 0.603 | 0.237 | — | — | — | — | — | — | — |
| $b_{S1987}^* =$ | — | — | — | — | — | — | — | — | — | 0.236 | 0.148 | 0.346 | — | — | — | — | — | — |
| $b_{S1986}^* =$ | — | — | — | — | — | — | — | — | — | — | 0.025 | 0.166 | 0.999 | — | — | — | — | — |
| $b_{S1985}^* =$ | — | — | — | — | — | — | — | — | — | — | — | 0.014 | 0.781 | 0.228 | — | — | — | — |
| $b_{S1984}^* =$ | — | — | — | — | — | — | — | — | — | — | — | 0.010 | 0.005 | 0.364 | 0.485 | — | — | — |
| $b_{S1983}^* =$ | — | — | — | — | — | — | — | — | — | — | — | 0.000 | 0.005 | 0.015 | 0.216 | 0.999 | — | — |
| $b_{S1982}^* =$ | — | — | — | — | — | — | — | — | — | — | — | — | 0.002 | — | 0.024 | 0.768 | 0.082 | — |
| $b_{S1981}^* =$ | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | 0.806 | 0.001 | 0.349 |
| $b_{S1980}^* =$ | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | — | 0.000 | 0.292 |

regression point $i$ will have more influence on the estimation of local parameter at location $i$ than does a data point at year $t-1$ and located at the same distance $d$ from the regression point $i$.

However, there are three cases for which the decreasing trend of the optimal spatial bandwidths is interrupted. The first case corresponds to the optimal bandwidths $b^*_{S1993}$ and $b^*_{S1992}$ for the regression year 1995. As can be seen in Table 2, $b^*_{S1993} < b^*_{S1992}$, where the opposite order, $b^*_{S1993} > b^*_{S1992}$ is expected. This change suggests that the bias introduced to the model calibration for year 1995 by data points located at year 1992 and at a distance, say, $d$ from the regression point $i$ is lower than the bias introduced by data points located at year 1993 and at the same distance from the regression point. As a consequence, it can be inferred that there are more similarities in the house pricing process between years 1995 and 1992 than that between years 1995 and 1993. Such inference can be reinforced by examining the time series plot of average house prices in Fig. 3 where average house prices for years 1992 and 1995 are more similar to each other than those for years 1995 and 1993 are. The other two cases with regard to the interruption of the decreasing trend in the optimal spatial bandwidths are years 1991 and 1983. Specifically, for year 1991, $b^*_{S1990} < b^*_{S1989}$ (Table 2), and for year 1983, $b^*_{S1982} < b^*_{S1981}$ (Table 2), while the opposite order in the inequality is expected as with the case for year 1995. Again, the explanation for these two cases can also be found in the similarities of the house pricing process between the corresponding years. For example, Fig. 3 suggests that the average house prices for years 1991 and 1989 are more similar than those for years 1991 and 1990 are.

## Model calibration

*Estimation of local parameters*

As described in the section entitled *Model formulation*, a set of GTWR local estimates at regression point $i$ in regression year $t$ is derived according to equation (2) where the weight matrix $\boldsymbol{W}$ is obtained by equation (5) using the set of optimal temporal and spatial bandwidths specific for each regression year $t$. As GTWR is used to calibrate the London hedonic house price, matrix $\boldsymbol{X}$ in equation (2) represents house determinants, while $\boldsymbol{y}$ is a vector of house prices. As there are more than 20 parameters in the model, the results for the FLRAREA semi-detached parameter, which measures the relationship between house prices and the floor area for semi-detached properties, is selected as an example to explore the temporal and spatial variability of local parameters estimated by GTWR.

Fig. 4 depicts a time series plot for the FLRAREA semi-detached parameter estimated by GTWR. For comparison, the time series plot for the same parameter, but estimated by cross-sectional GWR is also added to Fig. 4. The GTWR time series plot closely follows the trend of the cross-sectional GWR plot. For some years such as 1982, 1988, and 1995 the value of the parameter from GTWR exceeds the value given by the cross-sectional GWR, while the opposite occurs for years such as 1985, 1989, 1991, and 1998. This is possibly attributed to the effects of earlier data points from years $t-1, \ldots, t-q$, where $q$ is the number of effective lag years used in the calibration of GTWR, on the parameter estimation at a regression year $t$. While for some years earlier data points down-weight the value of the parameter estimate at a regression year $t$, for others the effect is the opposite.

In addition, the nonstationarity of the parameter estimates can be assessed by comparing twice the standard errors of the global ordinary least squares (OLS) estimates with the interquartile of local estimates from GTWR, with larger values of the latter indicating significant spatial nonstationarity (Fotheringham, Brunsdon, and Charlton 2002). Although other techniques can be employed to assess nonstationarity, such as the Monte Carlo simulation, this method is

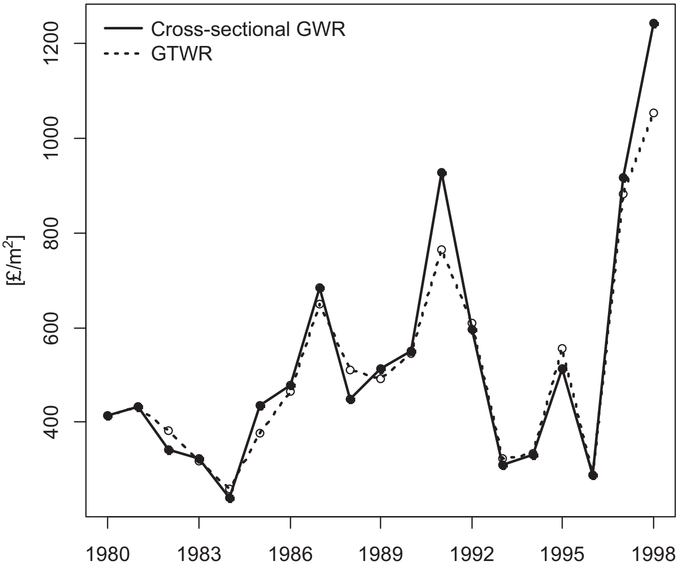


**Figure 4.** Temporal variation of FLRAREA Semi-detached estimate: cross-sectional geographically weighted regression (GWR) and geographical and temporal weighted regression (GTWR).

adopted here because it requires minimal computation effort. As an example, the results for the year 1998 are summarized in Table 3, from which it can be seen that all the parameter estimates exhibit extra variations beyond that expected from purely sampling.

Again, using the FLRAREA semi-detached parameter estimates as an example, the spatial variations in the local estimates can be examined through surface maps. Fig. 5 shows for a sample of years (1981, 1986, 1991, 1996, 1997, and 1998) the spatial distribution of local estimates of the FLRAREA semi-detached parameter obtained from GTWR. As the temporal weights assigned to data points in the regression year equal to 1, varying weights to such points are mainly derived by the spatial bandwidth for the regression year. Thus, spatial patterns of local estimates by GTWR are mostly controlled by the value of the spatial bandwidths for each regression year. It is thereby unlikely that there is any large spatial variation in years in which the spatial bandwidth is relatively large (1983, 1984, 1986, 1992, 1993, 1994 and 1995). Likewise, in the years when the spatial bandwidth is relatively small (1982, 1990, 1991, 1997, and 1998), a more spatially diverse set of local estimates of the parameter is observed.

Further, the spatial variation of local estimates over time is also driven by the temporal trend in house prices as depicted in Fig. 4. For example, the spatial distribution of the local estimates of the FLRAREA semi-detached parameter expands from 1997 to 1998 as the average value of semi-detached properties increased dramatically during this period. The increasing prominence of housing in central London, north of Thames, is clearly marked in these signs, which show the value/m$^2$ of semi-detached properties, holding all other factors constant.

*Analysis of residuals*

The selection of the temporal bandwidth that minimizes the CV score results from a bias-variance trade-off of local estimates. In order to examine the extent to which the drop in the CV score is due to a reduction in the bias or a reduction in the variance of the estimates, the boxplot of

**Table 3.** Spatial Nonstationarity Tests of Variables for Year 1998

| Variable | Interquartile (GTWR) | $2 \times$ SE (OLS) | Extra local variation |
|---|---|---|---|
| FLRAREA | 697.44 | 53.68 | YES |
| BLDPWW1 | 17,364.90 | 2,047.14 | YES |
| BLDPOSTW | 9,990.00 | 3,152.00 | YES |
| BLD60S | 12,350.00 | 3,388.32 | YES |
| BLD70S | 11,528.00 | 3,500.88 | YES |
| BLD80S | 15,473.00 | 3,922.90 | YES |
| BLD90S | 28,299.00 | 4,670.82 | YES |
| TYPDETCH | 85,806.00 | 14,459.62 | YES |
| TYPTRRD | 38,443.00 | 7,658.54 | YES |
| TYPBNGLW | 115,704.00 | 23,330.38 | YES |
| TYPFLAT | 39,220.00 | 7,339.86 | YES |
| GARAGE | 5,870.00 | 1,854.24 | YES |
| CENTHEAT | 8,030.00 | 2,576.52 | YES |
| BATH2 | 15,735.00 | 2,832.46 | YES |
| PROF | 74,190.00 | 7,977.50 | YES |
| UNEMPLOY | 92,541.00 | 24,655.34 | YES |
| FLRDETCH | 701.82 | 96.86 | YES |
| FLRFLAT | 400.70 | 72.00 | YES |
| FLRBNGLW | 363.53 | 230.88 | YES |
| FLRTRRD | 363.53 | 66.34 | YES |
| LOGDIST | 67,202.00 | 1,980.88 | YES |

GTWR, geographical and temporal weighted regression; OLS, ordinary least squares; SE, standard error.

residuals from GTWR is displayed in Fig. 6. In addition, two features are included in Fig. 6: (1) a point inside the rectangle indicating the mean of residuals, and (2) a line indicating the zero value for the residuals in each time period. As can be seen, the mean of the residuals from GTWR is approximately zero for all regression years. Also, greater variation in the residuals for the years 1988 and 1989 is evident, which might be caused by instability in the housing market prior to the crash in the early 1990s. In contrast, there is less variation in the residuals for the years 1993 and 1994, which might be relevant to the lower house prices for semi-detached properties during those years (see Fig. 5). The subsequent residuals kept increasing toward 1998, which preceded another crash in 2001.

## Comparison with basic GWR

As GTWR is an extension of traditional GWR into a temporal dimension, it might be of interest to know "is it statistically necessary to include data points from previous years $t-1, \ldots t-q$ to calibrate GWR at a regression point $i$ located at year $t$?" It is thereby worth comparing the statistical performance of GTWR against the cross-sectional GWR approach, that is when local parameters are estimated at each year by GWR using data exclusively from the year at which the model is being calibrated. Fig. 7 depicts for each year the CV score of the hedonic price model calibrated by the cross-sectional GWR and GTWR.[8] Three features of this plot are worth noting.

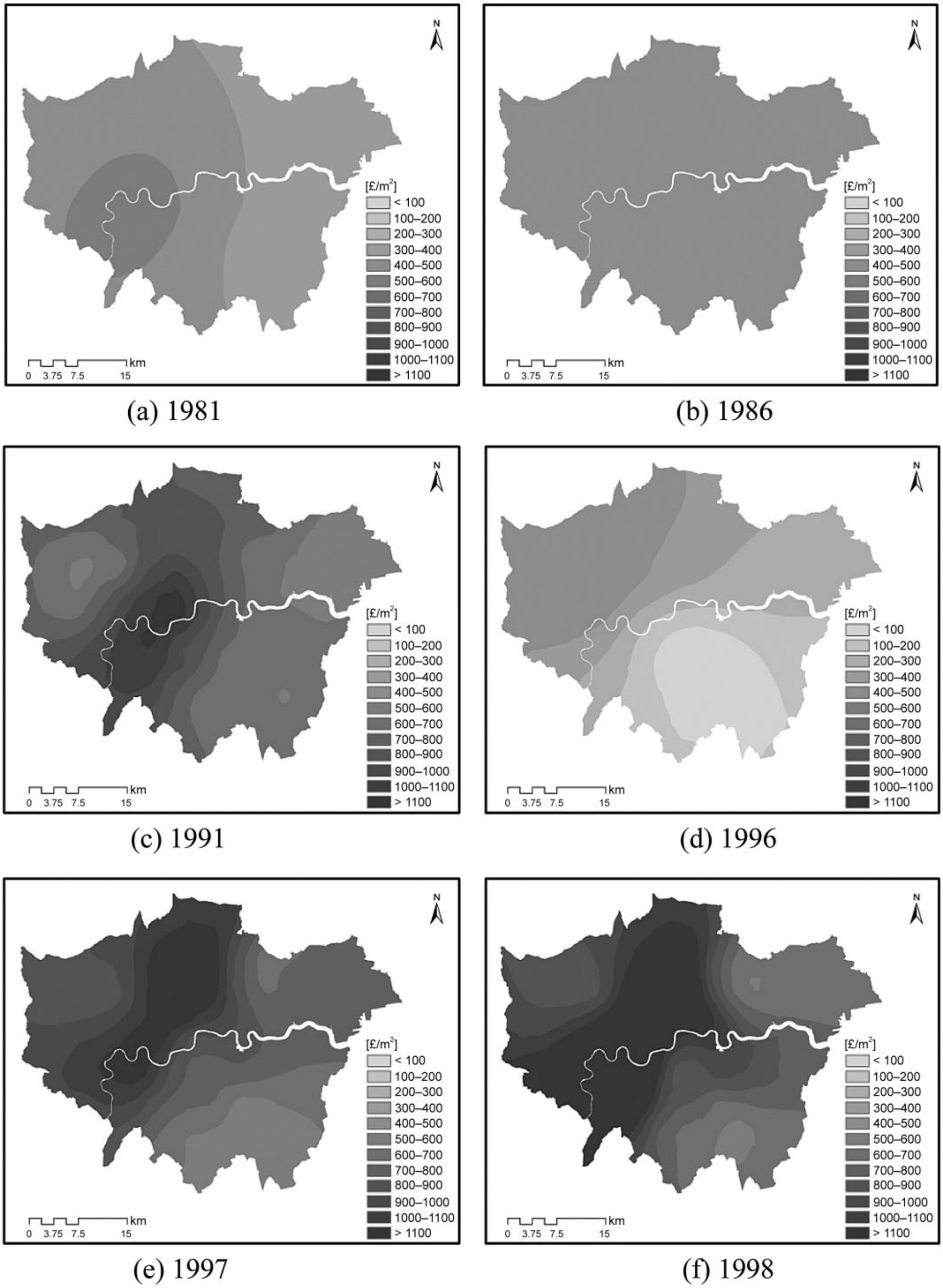


**Figure 5.** Spatial variation of local parameter estimate for FLRAREA Semi-detached from geographical and temporal weighted regression (GTWR): 1981–1998.

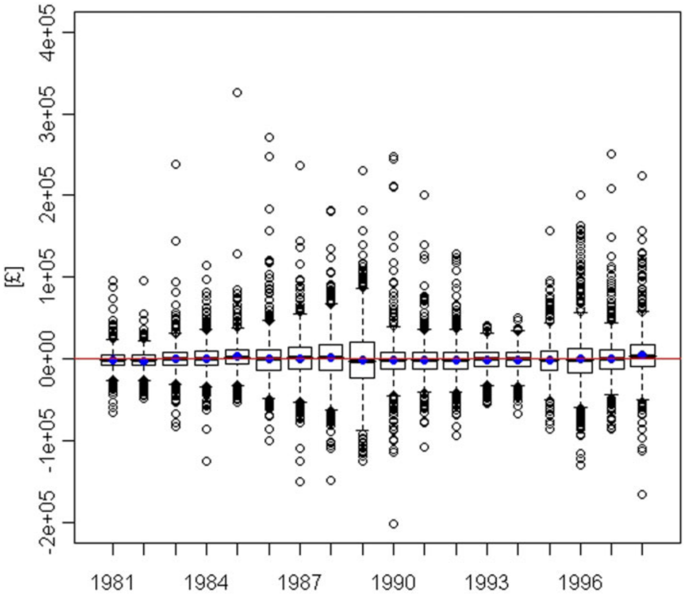


**Figure 6.** Boxplot of residuals from geographical and temporal weighted regression (GTWR).

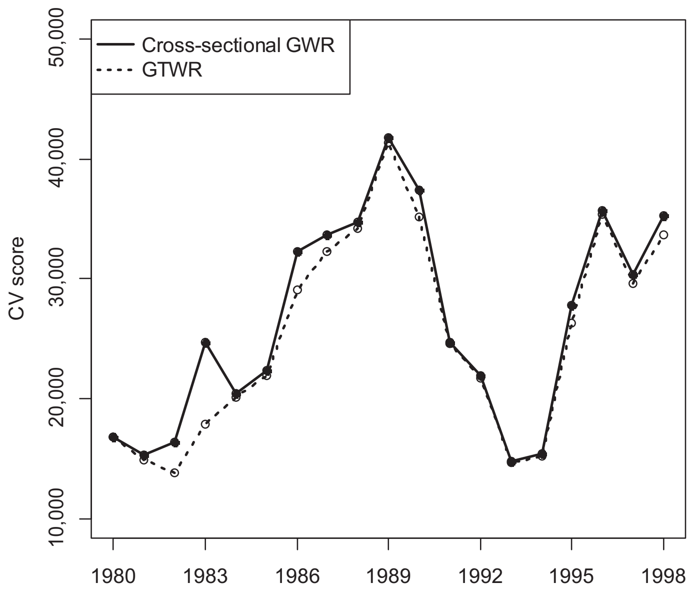


**Figure 7.** Cross-validation (CV) score from cross-sectional geographically weighted regression (GWR) and geographical and temporal weighted regression (GTWR): 1980–1998.

One is the similarity between the trend of the CV score of both techniques, and the time series plot of the average house prices shown in Fig. 3. The goodness-of-fit of the GTWR and the cross-sectional GWR approach decreases (i.e., the CV score becomes larger) as house prices rise. This would indicate that for periods of rapid acceleration in the housing market, for example between 1982 and 1989 and between 1995 and 1998, important changes in the housing pricing dynamics occur which might be modeled by the inclusion of additional variables in the hedonic price model. Another point to highlight is that the goodness-of-fit of GTWR is superior to the goodness-of-fit of the cross-sectional GWR for all years of the study, although for some years such as 1991, 1992, 1993, and 1994, the difference in the CV score between both techniques appears to be negligible.

Finally, it is worth noting that the goodness-of-fit of GTWR compared with the goodness-of-fit of the cross-sectional GWR is more superior during the 1980s than during the 1990s. In spite of a temporal bandwidth larger than zero being found for all years, the time series plots in Fig. 7 suggests that the influence of data points from earlier years contributed more to the calibration of the model in regression year *t* during the 1980s than during the 1990s. This finding implies that during the 1980s appraisals paid more attention to house prices from previous years than they did during the 1990s.

## Discussion and conclusions

As the importance of the time dimension in spatial phenomena and processes is increasingly recognized, spatiotemporal modeling has attracted extensive research interests in recent years in the field of GIS and analysis. The focus in this article is on local spatiotemporal modeling, that is a spatiotemporal version of GWR, accounting for the nonstationary processes in both space and time. A spatiotemporal extension of GWR, GTWR, is proposed, along with a method for optimal spatiotemporal bandwidth selection. The application of GTWR in hedonic price models using a 19-year data set in London, as well as the comparison with the traditional GWR approach, demonstrates the effectiveness and efficiency of the proposed model and calibration method.

One important characteristic of GTWR is the spatiotemporal kernel function, through which data points of interest are both spatially and temporally weighted. Further, the spatiotemporal bandwidth used by spatiotemporal kernel functions consists of the estimation of segregated bandwidths for each time period, allowing the size of individual bandwidths to vary over time. When compared with the cross-sectional GWR approach using the house price data in London, GTWR produced local parameter estimates with a better bias-variance trade-off, particularly for the house prices data from the 1980s, implying the significance of incorporating time dimension in local model calibration.

Closely related to the spatiotemporal bandwidth is the procedure through which the optimal spatial and temporal bandwidths are derived. In this study, spatial bandwidths are estimated one-by-one, which effectively addressed the computational infeasibility of deriving the spatial bandwidths simultaneously. However, although the CV score of the model estimated by GTWR in this way is smaller than that given by the cross-sectional GWR approach (see Fig. 7), there is no evidence that the one-by-one procedure yields a global minimum of the CV function. In other words, lower scores of the CV function might be obtained from other combinations of spatial bandwidths. Thus, the CV score given by the proposed procedure can be regarded as a particular solution for the minimization process of the CV score.

When compared with other local models dealing with space and time, the major difference in the operation of GTWR is the definition of the spatiotemporal bandwidth. Unlike the spatiotemporal distance proposed by Huang, Wu, and Barry (2010) and Yu (2014), which has the issue of dimensionality reduction, in GTWR proposed here the spatial and temporal distances are calculated separately while integrated through spatiotemporal kernel functions. In this way, such distances measured according to the underlying spatiotemporal scale can better reflect the spatiotemporal relationships among observations. However, GTWR has an increased computational complexity with the iterative optimal spatial and temporal bandwidth selection procedure given in the section entitled *GTWR*. Therefore, the choice of appropriate models is largely a function of the underlying data or problem of interest (Artelaris 2014).

As GWR is an important approach for local spatial modeling, the spatiotemporal version of GWR (GTWR) proposed in this study extends its capability in modeling spatiotemporal nonstationary processes. The application results have demonstrated that GTWR might be a useful and promising technique to calibrate spatiotemporal local models as well as to forecast future events. In terms of future research, it might be of interest to develop more efficient and sophisticated computational algorithms to derive the optimal spatial bandwidths simultaneously. Also, additional research can be conducted to investigate the statistical robustness of local estimates by GTWR, possibly by resampling methods such as bootstraping.

Another interesting avenue of research would be to explore the relationship between spatial and temporal kernels in GTWR and the use of separable space–time covariance structure in kriging. In summary, space–time modeling is complex and space–time local modeling is even more complex, but the increasing prevalence of large spatiotemporal data sets combined with ever more efficient algorithm development and faster computing resources, mean that models such as GTWR and subsequent development are now becoming practical to provide ever more details spatial and temporal processes.

## Notes

1 From this point onward, for simplification purposes, the version of GWR when data points are spatially and temporally weighted will be referred to as GTWR.
2 In GWR, a regression point corresponds to the point at which parameters are being estimated.
3 Brunsdon, Fotheringham, and Charlton (1998) claim that the spatial bandwidth $b_S$ provides some control on the range of the “circle of influence” in the geographical data.
4 In this study, the root mean square form of the CV function defined is used as a way of providing a more intuitive interpretation of the magnitude of the average error regardless of the error is either positive or negative.
5 Each census output area has an approximate population of 150 households.
6 PROF and UNEMPLOY variables were created using the information provided by the U.K. Census of Population held in April 2001.
7 The term *regression year* is used to refer to the year in which the regression point is located.
8 The CV score of GTWR for year 1980 is the same as the one of the cross-sectional GWR as no data points are available from previous years.

## References

Andrienko, G., N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Kraak, and H. Schumann. (2010). “Space, Time and Visual Analytics.” *International Journal of Geographical Information Science* 24, 1577–600.

Anselin, L. (1999). “Spatial Econometrics.” Working paper. Bruton Center, School of Social Sciences, University of Texas at Dallas. Accessed at: http://www.csiss.org/learning_resources/content/papers/baltchap.pdf.

Artelaris, P. (2014). “Local versus Regime Convergence Regression Models: A Comparison of Two Approaches.” *GeoJournal* 1–15.

Atkinson, P. M., S. E. German, D. A. Sear, and M. J. Clark. (2003). “Exploring the Relations between River Bank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression.” *Geographical Analysis* 35(1), 58–82.

Bitter, C., G. Mulligan, and S. Dall'erba. (2007). "Incorporating Spatial Variation in Housing Attribute Prices: A Comparison of Geographically Weighted Regression and the Spatial Expansion Method." *Journal of Geographical Systems* 9, 7–27.

Brunsdon, C., A. S. Fotheringham, and M. Charlton. (1996). "Geographically Weighted Regression: A Method for Exploring Spatial Non-Stationarity." *Geographical Analysis* 28(4), 281–98.

Brunsdon, C., A. S. Fotheringham, and M. Charlton. (1998). "Geographically Weighted Regression-Modelling Spatial Non-Stationarity." *The Statistician* 47(3), 431–43.

Brunsdon, C., J. Corcoran, and G. Higgs. (2007). "Visualizing Space and Time in Crime Patterns: A Comparison of Methods." *Computers, Environment and Urban Systems* 31, 52–75.

Buyantuyev, A., and J. Wu. (2010). "Urban Heat Islands and Landscape Heterogeneity: Linking Spatiotemporal Variations in Surface Temperatures to Land-Cover and Socioeconomic Patterns." *Landscape Ecology* 25(1), 17–33.

Chen, J., S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. (2011). "Exploratory Data Analysis of Activity Diary Data: A Space–Time GIS Approach." *Journal of Transport Geography* 19(3), 394–404.

Comber, A. J., C. Brunsdon, and R. Radburn. (2011). "A Spatial Analysis of Variations in Health Access: Linking Geography, Socio-Economic Status and Access Perceptions." *International Journal of Health Geographics* 10, 44. Accessed at: http://www.ij-healthgeographics.com/content/10/1/44.

Crespo, R., A. S. Fotheringham, and M. Charlton. (2007). "Application of Geographically Weighted Regression to a 19-Year Set of House Price Data in London to Calibrate Local Hedonic Price Models." In *Proceedings of the 9th International Conference on Geocomputation*, edited by U. Demšar. Maynooth, Ireland: NCG, NUI.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

Cressie, N., and C. K. Wilkie. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, NJ: John Wiley & Sons, Inc.

Demšar, U., and K. Virrantaus. (2010). "Space–Time Density of Trajectories: Exploring Spatio-Temporal Patterns in Movement Data." *International Journal of Geographical Information Science* 24, 1527–42.

Elhorst, J. (2003). "Specification and Estimation of Spatial Panel Data Models." *International Regional Science Review* 26(3), 244–68.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. (1996). "The Geography of Parameter Space: An Investigation of Spatial Non-Stationarity." *International Journal of Geographical Information Systems* 10, 605–27.

Fotheringham, A. S., M. Charlton, and C. Brunsdon. (1997). "Two Techniques for Exploring Nonstationarity in Geographical Data." *Geographical Systems* 4, 59–82.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley.

Gelfand, A. E., M. D. Ecker, J. R. Knight, and C. F. Sirmans. (2004). "The Dynamics of Location in Home Prices." *Journal of Real Estate Finance and Economics* 29, 149–66.

Giacinto, V. (2006). "A Generalized Space–Time ARMA Model with an Application to Regional Unemployment Analysis in Italy." *International Regional Science Review* 29(2), 159–98.

Goodchild, M. F. (2013). "Prospects for a Space–Time GIS." *Annals of the Association of American Geographers* 103(5), 1072–7.

Hägerstrand, T. (1970). "What about People in Regional Science?" *Papers of the Regional Science Association* 24, 1–12.

Harris, P., A. S. Fotheringham, and S. Juggins. (2010). "Robust Geographically Weighted Regression: A Technique for Quantifying Spatial Relationships between Freshwater Acidification Critical Loads and Catchment Attributes." *Annals of the Association of American Geographers* 100, 286–306.

Huang, B., B. Wu, and M. Barry. (2010). "Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices." *International Journal of Geographical Information Science* 24(3), 383–401.

Kwan, M. (2000). "Interactive Geovisualization of Activity Travel Patterns Using Three-Dimensional Geographical Information Systems: A Methodological Exploration with a Large Data Set." *Transportation Research Part* C(8), 185–203.

Kwan, M. P. (2004). “GIS Methods in Time-Geographic Research: Geocomputation and Geovisualization of Human Activity Patterns, Geografiska Annaler Series B.” *Human Geography* 86(4), 267–80.

Malczewskia, J., and A. Poetzb. (2005). “Residential Burglaries and Neighborhood Socioeconomic Context in London, Ontario: Global and Local Regression Analysis.” *The Professional Geographer* 57(4), 516–29.

Mennis, J., and L. Jordan. (2005). “The Distribution of Environmental Equity: Exploring Spatial Nonstationarity in Multivariable Models of Air Toxic Releases.” *Annal of the Association of Americans Geographers* 95(2), 249–68.

Nakaya, T., and K. Yano. (2010). “Visualising Crime Clusters in a Space–Time Cube: An Exploratory Data-Analysis Approach Using Space–Time Kernel Density Estimation and Scan Statistics.” *Transactions in GIS* 14(3), 223–39.

Nakaya, T., A. Fotheringham, C. Brunsdon, and M. Charlton. (2005). “Geographically Weighted Poisson Regression for Disease Association Mapping.” *Statistics in Medicine* 24, 2695–717.

Pace, K., R. Barry, J. Clapp, and M. Rodriquez. (1998). “Spatiotemporal Autoregressive Models of Neighbourhood Effects.” *Journal of Real Estate Finance and Economics* 17(1), 15–33.

Pace, R. K., R. Barry, O. W. Gilley, and C. F. Sirmans. (2000). “A Method for Spatial–Temporal Forecasting with An Application to Real Estate Prices.” *International Journal of Forecasting* 16(2), 229–46.

Rey, S. J., and M. V. Janikas. (2010). “STARS: Space–Time Analysis of Regional Systems.” In *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, 91–112, edited by M. M. Fischer and A. Getis. London: Springer.

Takahashi, K., M. Kulldorff, T. Tango, and K. Yih. (2008). “A Flexibly Shaped Space–Time Scan Statistic for Disease Outbreak Detection and Monitoring.” *International Journal of Health Geographics* 7, 14. Accessed at: http://www.ij-healthgeographics.com/content/7/1/14.

Tobler, W. R. (1970). “A Computer Movie Simulating Urban Growth in the Detroit Region.” *Economic Geography* 46, 234–40.

Wheeler, D., and L. Waller. (2009). “Comparing Spatially Varying Coefficient Models: A Case Study Examining Violent Crime Rates and Their Relationships to Alcohol Outlets and Illegal Drug Arrests.” *Journal of Geographical Systems* 11, 1–22.

Wrenn, D. H., and A. G. Sam. (2014). “Geographically and Temporally Weighted Likelihood Regression: Exploring the Spatiotemporal Determinants of Land Use Change.” *Regional Science and Urban Economics* 44, 60–74.

Wu, B., R. Li, and B. Huang. (2014). “A Geographically and Temporally Weighted Autoregressive Model with Application to Housing Prices.” *International Journal of Geographical Information Science* 28, 1186–204.

Wu, K., B. Liu, B. Huang, and Z. Lei. (2013). “Incorporating the Multi-Cross-Sectional Temporal Effect in Geographically Weighted Logit Regression.” In *Proceedings of the International Conference on Information Systems and Computing Technology (ISCT)*, 3–14, edited by L. Zhang and Y. Gu. London: CRC Press, Taylor & Francis Group.

Yu, D. (2014). “Understanding Regional Development Mechanisms in Greater Beijing Area, China, 1995–2001, from a Spatial–Temporal Perspective.” *GeoJournal* 79, 195–207.