Ann Reg Sci (2015) 54:417–436
DOI 10.1007/s00168-015-0660-6

ORIGINAL PAPER

Exploring, modelling and predicting spatiotemporal variations in house prices

A. Stewart Fotheringham · Ricardo Crespo · Jing Yao

Received: 5 September 2014 / Accepted: 2 February 2015 / Published online: 4 March 2015

Abstract Hedonic price modelling has long been a powerful tool to estimate house prices in the real estate market. Increasingly, traditional global hedonic price models that largely ignore spatial effects are being superseded by models that deal with spatial dependency and spatial heterogeneity. In addition, many novel methods integrating spatial economics, statistics and geographical information science (GIScience) have been developed recently to incorporate temporal effects into hedonic house price modelling. Here, a local spatial modelling technique, geographically weighted regression (GWR), which accounts for spatial heterogeneity in housing utility functions is applied to a 19-year set of house price data in London (1980–1998) in order to explore spatiotemporal variations in the determinants of house prices. Further, based on the local parameter estimates derived from GWR, a new method integrating GWR and time series (TS) forecasting techniques, GWR–TS, is proposed to predict future local parameters and thus future house prices. The results obtained from GWR demonstrate variations in local parameter estimates over both space and time. The forecasted future values of local estimates as well as house prices indicate that the proposed GWR–TS method is a useful addition to hedonic price modelling.

Keywords Hedonic price models · GWR · Spatiotemporal modelling · GIS

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

R. Crespo
Facultad de Ingeniería y Administración, Universidad Bernardo O'Higgins, Santiago, Chile

J. Yao
Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow G12 8RZ, UK

JEL Classification C3 · R3

1 Introduction

Hedonic price modelling has long been a powerful tool to estimate the market value of properties and to investigate the relationships between property prices and their inherent attributes (Goodman 1978; Can 1992; Fotheringham et al. 2002; Páez 2009; Páez et al. 2011). Basically, hedonic price modelling is a statistical technique to model the price of goods based upon quantitative characteristics associated with those goods, often expressed in a traditional linear regression form (Goodman 1998). In terms of house price modelling, the attributes of properties usually include structural attributes (e.g. dwelling age and floor area, number of rooms and bedrooms), socio-economic characteristics of the surrounding neighbourhood (e.g. unemployed rate, racial diversity and occupations of the inhabitants) and locational attributes (e.g. accessibility to Central Business Districts (CBDs) and proximity to services and pleasant landscapes) (Basu and Thibodeau 1998). For example, studies have found that house prices are positively related to structural attributes such as property size and the number of rooms and negatively related to distance from the CBD, *ceteris paribus* (Fotheringham et al. 2002). Of interest in this paper is to use hedonic price models in a novel way to explore, model and predict spatial variations in house prices over time.

There is a strong evidence in the literature of the presence of both spatial dependence (also known as spatial autocorrelation) and spatial heterogeneity (also referred to as spatial non-stationary) in the housing market (Goodman and Thibodeau 2003). The former refers to the situation where the price of a house at a certain location is spatially correlated with the prices of nearby houses. House prices are usually spatially autocorrelated because neighbourhood properties share numerous features that influence house price. Also, most neighbourhoods are developed at about the same time, and proximity to both positive and negative externalities has similar effect on the market values of nearby properties. Furthermore, real estate agents usually use a comparative method to value properties based on neighbouring house prices. The latter refers to the relationships (measured by parameter estimates in regression models) or processes by which people evaluate the desirability of a property varying over space. For example, a two-car garage in an inner-city neighbourhood may not have the same marginal value as in a residential suburb area (Can 1990). Another source of spatial heterogeneity in the housing market is the presence of market segmentation caused by inelastic demand or supply functions under market conditions in disequilibrium, that is, a high demand in conditions of low supply or vice versa (Goodman and Thibodeau 1998). Therefore, hedonic house price models should address such spatial effects appropriately.

Two perspectives, one being global and the other being local, can be found in the hedonic modelling literature with global modals dominating. Various global models have been developed to deal with spatial dependence, such as the seminal work by Cliff and Ord (1981), Anselin (1988), Griffith (1988) and Haining (1990). Probably, the most widely utilized specification is the one proposed by Anselin (1988), which assumes spatial autocorrelation in either the response variable or the error terms. Local models, broadly speaking, are generalizations of standard global models where the model coefficients are allowed to vary over space in order to account for any spatial heterogeneity in the processes being modelled examples include moving window regression, multilevel modelling and geographically weighted regression (GWR). For example, Páez et al. (2008) applied a moving window method to a data set with over 30,000 recorded transactions in Toronto, Canada. Orford (2000) employed a multilevel analysis to investigate the variability of the housing valuation process across local sub-markets in Cardiff, while Fotheringham et al. (2002) used GWR to calibrate a hedonic price model for London.

The technique of interest in this research is GWR (Brundsdon et al. 1996; Fotheringham et al. 2002), which has been widely applied in modelling spatially heterogeneous processes across a variety of topics such as health (Nakaya et al. 2005), environment science (Harris et al. 2010b) and real estate (Bitter et al. 2007). GWR recognizes the potential spatial variations in the relationships between dependent and independent variables and also provides a way to measure them. The way GWR captures spatial varying relationships is to calibrate a series of local models, where the local neighbourhood is defined by spatial kernel functions with fixed or varying bandwidth. In addition to the classic GWR models introduced in Fotheringham et al. (2002), recent years have seen a growing number of extensions, such as ridge GWR (Wheeler 2007), GWR with non-Euclidean distances (Lu et al. 2014) and flexible bandwidth GWR (Yang et al. 2012), as well as diagnostics and testing techniques, such as tools for detecting local collinearity (Wheeler 2007) and multiple hypothesis testing (Byrne et al. 2009).

Also, noteworthy is that there has been increasing interest in incorporating the temporal dimension into hedonic house price modelling, both in global models (e.g. Pace et al. 1998) and in local models (Huang et al. 2010; Fotheringham et al. 2015), and also in a comparison of different models (Case et al. 2004). One reason for this is that the house pricing process evolves not only over space but also over time. Also, more spatiotemporal data are being made available which can be efficiently managed and processed in a geographical information system (GIS) environment. However, so far, most work dealing with spatiotemporal hedonic house price modelling has focused on incorporating prior prices in global/local parameter estimation. This study is particularly interested in predicting future values of local parameters in local hedonic house price models. To this end, the aim of this paper is twofold: (1) to explore and model the spatiotemporal variability of house prices and associated determinants and (2) to predict future values of local parameters as well as house prices based on the results obtained from (1).

The remainder of the paper is structured as follows. A 19-year set of house price data in London (1980–1998) is described in the next section. Methods of analysing these data including GWR and GWR–TS are then presented, followed by results from the analyses which include spatiotemporal variations in the London house prices and predictions of both local parameters and house prices. Finally, a discussion of the major findings and the implications for future research is provided.

2 Data

The data utilized in this study are provided by the Nationwide Building Society in the UK and consist of house prices and associated attributes in London for the time period

Table 1 House price data for London (1980–1998)

Year	Number of observations	Sample size	House prices average (£)	Inflated house price average (£)
1980	14,233	1000	30,475	58,896
1981	14,216	1000	30,244	53,671
1982	17,728	1000	30,194	49,023
1983	17,417	1000	35,112	53,730
1984	18,803	1000	40,490	59,520
1985	16,342	1000	47,346	67,115
1986	19,990	1000	55,526	76,345
1987	8,768	1000	71,871	96,691
1988	13,617	1000	79,226	103,581
1989	4,738	1000	80,160	101,356
1990	4,844	1000	93,453	113,730
1991	5,964	1000	85,390	100,016
1992	5,545	1000	77,028	87,255
1993	3,470	1000	63,779	70,692
1994	3,901	1000	62,070	66,988
1995	433	433	80,022	84,587
1996	11,365	1000	87,159	90,236
1997	11,947	1000	98,827	100,507
1998	11,282	1000	119,703	119,703

1980–1998. In order to make subsequent computations more feasible, approximately 1000 non-overlapping observations were randomly sampled from the data set for each year—an exception being 1995 which only contains 433 observations. The house prices were inflated to the 1998 equivalent price using the inflation rate in the UK from the Office for National Statistics in order to make the results comparable over time. The number of observations available, the sample size and the average and the inflated house prices are described in Table 1.

In general, house prices experienced sustained increases between 1980–1990 and then decreased from 1991 due to the severe recession in the early 1990s. This downward trend in prices continued until 1995 when the market rapidly recovered so that by 1998 house prices reached approximately the same level as before the recession. According to the Nationwide Building Society, in spite of the various cycles of acceleration and downturn over the last 50 years, the long-term growth trend in real house prices in UK has been around 2.7 % annually since the 1950s.

3 Methodology

GWR is utilized to explore and model spatiotemporal variations in the determinants of house prices. A new approach, named GWR–TS, integrating GWR and time series forecasting techniques, is proposed to predict future local parameters and house prices.

Before applying GWR and GWR–TS; however, some exploratory analysis is carried out using GIS to detect general spatiotemporal patterns in the London house price data. Specifically, in order to illustrate the continuous spatial variations in house prices over time, surfaces of house prices are generated using spatial interpolation.

To further understand the underlying spatiotemporal processes leading to the variations in house prices, a hedonic price model is formulated to explore the relationships between house prices and associated determinants. Following the hedonic price model defined in Fotheringham et al. (2002), three types of explanatory variables are adopted in this research to estimate house prices: structural, neighbourhood and locational attributes associated with the properties. Using the notation defined in Table 2, the hedonic model for London house price data can be expressed as in (1):

$$\begin{aligned} P_i = {} & \beta_0 + \beta_1 \mathrm{FLRAREA}_i + \beta_2 \mathrm{BLDPWW1}_i + \beta_3 \mathrm{BLDPOSTW}_i + \beta_4 \mathrm{BLD60S}_i \\ & + \beta_5 \mathrm{BLD70S}_i + \beta_6 \mathrm{BLD80S}_i + \beta_7 \mathrm{BLD90S}_i \\ & + \beta_8 \mathrm{TYPDETCH}_i + \beta_9 \mathrm{TYPTRRD}_i \\ & + \beta_{10} \mathrm{TYPBNGLW}_i + \beta_{11} \mathrm{TYPFLAT}_i + \beta_{12} \mathrm{GARAGE}_i \\ & + \beta_{13} \mathrm{CENHEAT}_i + \beta_{14} \mathrm{BATH2}_i \\ & + \beta_{15} \mathrm{PROF}_i + \beta_{16} \mathrm{UNEMPLOY}_i + \beta_{17} \mathrm{FLRDETCH}_i \\ & + \beta_{18} \mathrm{FLRFLAT}_i + \beta_{19} \mathrm{FLRBNGLW}_i \\ & + \beta_{20} \mathrm{FLRTRRD}_i + \beta_{21} \log_e (\mathrm{DISTCL}_i) + \varepsilon_i \end{aligned} \tag{1}$$

The above model is calibrated using GWR which extends the global regression framework by allowing the parameters to vary over space (Fotheringham et al. 2002). In addition to the values of local parameters being obtained from GWR, a concern is whether the local estimates are statistically significant. This decision is generally based on t values which can be calculated by dividing each local estimate by its standard error. Given a significance level (e.g. 0.05 or 0.1), the local estimates are considered significant if their absolute values are larger than the pre-defined critical t value. Considering the multiple hypotheses testing problem in GWR, the critical t values used here are derived based on the adjusted family-wise error rate e as defined in (2) (Byrne et al. 2009):

$$e = \xi_m / p_e \tag{2}$$

where ξ_m is the original family-wise error rate (significance level), and p_e is the effective number of parameters as defined in Fotheringham et al. (2002).

Once local parameter estimates have been derived, future values of those local estimates as well as house prices can be predicted. This is achieved by a semi-parametric method referred to as GWR–TS, which is an integration of GWR and time series forecasting methods. With regard to the forecasting technique, Holt's exponential smoothing approach is adopted due to its simplicity in fitting non-seasonal data and its statistical robustness in modelling a very general class of state-space models, particularly for short-term forecasting (Chatfield 2004; Gardner 2006). As documented by Chatfield (2004), a simple exponential smoothed time series in recurrence form can be expressed as in (3):

Table 2 Variable definition for Eq. (1)

	Variable name	Type	Definition
Dependent variable	*P*	Numerical	Price in pounds sterling of the property
Independent variable	FLRAREA	Numerical	Floor area of the property in square metres
Structural attributes	BLDPWW1	Dummy	1: if the property was built prior to 1914; 0: otherwise
	BLDPOSTW	Dummy	1: if the property was built between 1940 and 1959; 0: otherwise
	BLD60S	Dummy	1: if the property was built between 1960 and 1969; 0: otherwise
	BLD70S	Dummy	1: if the property was built between 1970 and 1979; 0: otherwise
	BLD80S	Dummy	1: if the property was built between 1980 and 1989; 0: otherwise
	BLD90S	Dummy	1: if the property was built between 1990 and 1999; 0: otherwise
	TYPDETCH	Dummy	1: if the property is detached (i.e. it is a stand-alone house); 0: otherwise
	TYPTRRD	Dummy	1: if the property is in a terrace of similar houses (commonly referred to as a “row house” in the USA); 0: otherwise
	TYPBNGLW	Dummy	1: if the property is a bungalow (i.e. it has only one floor); 0: otherwise
	TYPFLAT	Dummy	1: if the property is a flat (or “apartment” in USA); 0: otherwise
	GARAGE	Dummy	1: if the house has a garage; 0: otherwise
	CENTHEAT	Dummy	1: if the house has a central heating; 0: otherwise
	BATH2	Dummy	1: if the house has 2 or more bathrooms; 0: otherwise
	FLRDETCH	Numerical	FLRAREA × TYPDETCH
	FLRFLAT	Numerical	FLRAREA × TYPFLAT
	FLRBNGLW	Numerical	FLRAREA × TYPBNGLW
	FLRTRRD	Numerical	FLRAREA × TYPTRRD
Neighbourhood attributes	PROF[a]		Proportion of the workforce in professional or managerial occupations in the census output area in which the house is located
	UNEMPLOY		The rate of unemployment in the census output area in which the house is located
Locational attributes	DISTCL		Straight-line distance from the property to the centre of London (taken here to be Nelson’s column in Trafalgar Square) measured in metres

[a] PROF and UNEMPLOY variables were created using the information provided by the UK Census of Population held in April 2001

$$S_t = \alpha x_t + (1 - \alpha)S_{t-1} \tag{3}$$

where S_t is the smoothed observation at time period t, x_t is the observed time series at time t, and α is a constant smoothing parameter such that $0 < \alpha < 1$. It can be seen that the higher the value of α, the more important is the selection of recent values to estimate the smoothed series. As with simple exponential smoothing, the trend T_t of a time series can be formulated in recurrence form as in (4):

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \tag{4}$$

where γ is a constant smoothing parameter such that $0< \gamma <1$. Similar to α, the higher the value of γ, the more important is the selection of recent values of S_t in the calculation of the smoothed trend T_t. Optimal values for α and γ are usually determined by minimizing the squared error term of observed values versus predicted values. Once those optimal values are found, future values of the series x_t for h time periods ahead is given by (5):

$$\hat{x}_t\,(h) = S_t + hT_t \tag{5}$$

By integrating GWR and Holt's exponential smoothing models, the method proposed in this research can be illustrated by the following five steps using London house price data. As a way to compare the performance of GWR–TS with a global modelling approach, parameter estimates from the hedonic price model in Eq. (1) are estimated by ordinary least squares (OLS) as well.

1. The sample of 1,000 observations for the year 1998 is used as a validation set for the forecasting procedure;
2. The hedonic price model defined in Eq. (1) is calibrated using both OLS and GWR for each year between 1980 and 1997. Taking β_1 as an example, by OLS, the global parameter estimates over time are expressed as $\hat{\beta}_{1,T} = [\hat{\beta}_{1,t}, \hat{\beta}_{1,t+1}, \ldots, \hat{\beta}_{1,t+n}]$, where t in this case corresponds to 1980. Similarly, the hedonic price model defined in Eq. (1) is calibrated using GWR for each year of the time period 1980–1997. Using the optimal bandwidths from these GWR calibrations, local estimates of β_1 for the same set of locations defined in 1) are derived for each time period. It should be noted that although house locations in 1998 are different from those in 1980–1997, the local estimates at the locations in the validation set can be interpolated by GWR (Harris et al. 2010a). Accordingly, 1,000 temporal sets specific to β_1 of the form $\hat{\beta}_{1_{i,T}} = [\hat{\beta}_{1_{i,t}}, \hat{\beta}_{1_{i,t+1}}, \ldots, \hat{\beta}_{1_{i,t+n}}]$ are obtained and locations remain constant over time, where t in this case corresponds to year 1980, n is equal to 18 and $i = 1, \ldots, 1{,}000$.
3. The next step is to determine the optimal values of α and γ for both OLS–TS and GWT–TS, which is carried out in the *forecasts* package of the open-source statistical software R (Hyndman and Khandakar 2008). Both parameters are selected to minimize the root mean square error of observed values for the validation set versus forecasted values.
4. Optimal values of α and γ are then employed to forecast futures values of $\hat{\beta}_{1,T}$ and $\hat{\beta}_{1_{i,T}}$ following Eq. (5), which is also implemented in the *forecast* R package. It

should be noted that a single set of 18 temporal observations (from 1980 to 1997) is used to forecast future values of $\hat{\beta}_{1,T}$, while in the case of $\hat{\beta}_{1_{i,T}}$, 1,000 sets of 18 temporal observations are needed to forecast future values of $\hat{\beta}_1$ at locations of the validation set.

5. Finally, steps 3 and 4 are repeated for the rest of the parameter estimates from the hedonic price model in Eq. (1), i.e. for $\beta_0, \beta_2, \ldots, \beta_{21}$. Forecasted values of global and local estimates can be input into the hedonic price models to predict future values of house prices at each location using Eqs. (6) and (7), respectively.

$$\hat{P}^{\text{OLS}}_{i,t+h} = \hat{\beta}_{0,t+h} + \hat{\beta}_{1,t+h}\text{FLRAREA}_i + \cdots + \hat{\beta}_{21,t+h} log_e(\text{DISTCL}_i) \quad (6)$$

$$\hat{P}^{\text{GWR}}_{i,t+h} = \hat{\beta}_{0_{i,t+h}} + \hat{\beta}_{1_{i,t+h}}\text{FLRAREA}_i + \cdots + \hat{\beta}_{21_{i,t+h}} log_e(\text{DISTCL}_i) \quad (7)$$

where $\hat{P}^{\text{OLS}}_{i,t+h}$ and $\hat{P}^{\text{GWR}}_{i,t+h}$ are the forecasted house prices at location i of the calibration set using estimates derived from OLS and GWR, respectively, h is the number of years ahead to be forecasted, and t is in this case the year 1997.

Thus, GWR–TS method is a combination of GWR and time series forecasting techniques, employing Holt's exponential smoothing models to temporally model each local parameter using the results obtained from the GWR models.

4 Results

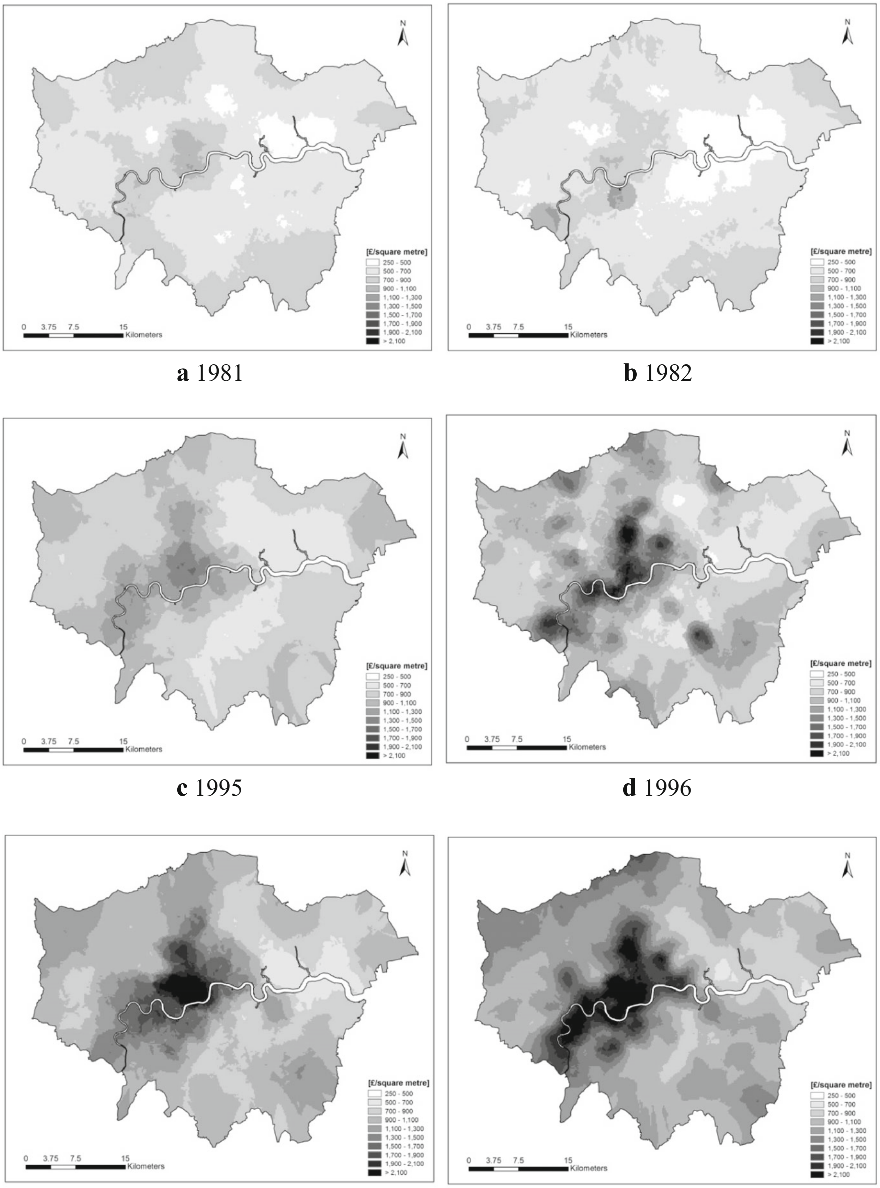
A set of continuous surfaces is first generated to explore the spatial variations in house prices over time. Further, the spatiotemporal variations in local parameter estimates obtained from GWR are presented, highlighting the non-stationary of the housing price process over both space and time. Finally, future values of local parameters are predicted by GWR–TS, and the spatiotemporal variations in the forecasted house prices are provided.

4.1 House price surfaces

Due to limited space, the house price (per square metre) surface maps are only presented for 6 years, 1981 and 1982 at the beginning of the sequence and 1995–1998 at the end of the sequence, as shown in Fig. 1 (the categories are the same on all the figures to ease comparison). As can be observed, higher valued housing tends to be generally located in the centre of London, and the price differential between core and periphery becomes increasingly noticeable over time. There is also a noticeable trend of higher house prices diffusing southwest from the centre along the river.

4.2 GWR results

The GWR calibrations of the hedonic price model in Eq. (1) were carried out using a Gaussian kernel with an adaptive bandwidth which is measured as the proportion of data points to be included in the local model calibrations. The optimal bandwidths with values ranging between 0 and 1 are derived by minimizing the cross-validation (CV) function as shown in (8):



a 1981 **b** 1982

c 1995 **d** 1996

e 1997 **f** 1998

Fig. 1 Surface map of observed house prices: 1981–1982 and 1995–1998

$$\mathrm{CV} = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_{-i})^2 / n} \tag{8}$$

Table 3 AIC values for GWR and global (OLS) models

	GWR	OLS
1980	21,121.8	21,395.8
1981	21,503.1	21,657.6
1982	21,159.16	21,244.3
1983	22,213.48	22,456.7
1984	22,256.92	22,408.54
1985	22,478.27	22,795.54
1986	25,357.06	25,812.66
1987	23,725.83	23,946.4
1988	23,820.12	23,924.49
1989	24,809.11	24,877.39
1990	24,558.77	24,790.51
1991	24,070.93	24,345.86
1992	24,159.46	24,290.81
1993	23,451.26	23,466.97
1994	23,698.28	23,726.65
1995	10,054.01	10,066.27
1996	25,520.37	25,543.95
1997	25,096.92	25,289.77
1998	25,392.11	25,734.49

where $\hat{y}_{-i}$ is the fitted value of y_i with observation i excluded in the local model calibration. Shown in Table 3 are the values of Akaike information criterion (AIC) for GWR and global models using OLS, with lower values indicating better model fit. It can be seen that GWR consistently outperforms global models for all the years throughout the study time period.

Figure 2 displays the box plots of residuals obtained from GWR and OLS regression. It can be seen that GWR and OLS yield broadly similar patterns of residuals, which tend to be less dispersed in the early 80s and early 90s. However, GWR shows smaller residuals in the first and fourth percentile than OLS. What is interesting to note is the relative inability of the model to predict house prices just prior to the crash of 1990, presumably when the market was becoming more chaotic. There then follows a period where residuals are relatively small but increasing towards 1998 which preceded another crash in 2001. The ability of the models to predict house prices thus appears to be linked to the state of the housing market with higher levels of error indicating the advent of a housing crash. Given the similar characteristic patterns of the residuals in the years prior to the two housing crashes in 1990 and 2001 (and we have noticed very similar behaviour on more local data set preceding the crash of 2007/8), this is perhaps worth investigating further.

As mentioned above, the bandwidth is optimized using the CV function (8) during the model calibration for each year. Thus, the optimal bandwidth can vary over time as shown in Fig. 3. It is worth observing how the underlying spatial process moves through cycles of high and low adaptive bandwidths from the GWR calibrations.

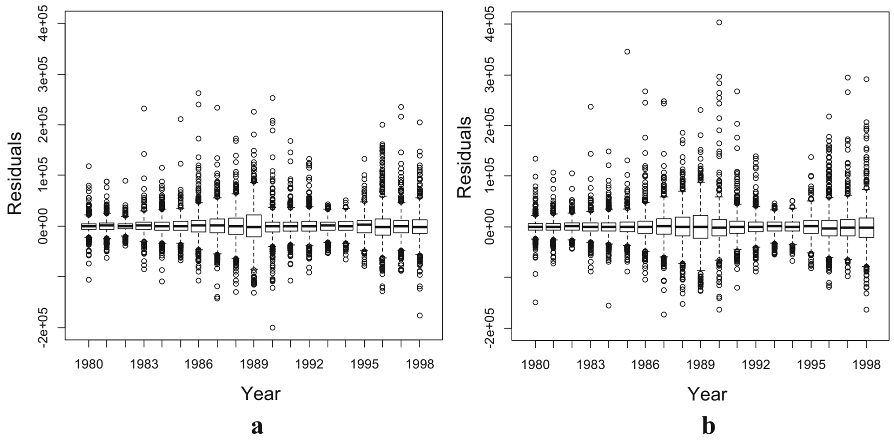


Fig. 2 Box plots of residuals from regression analysis, **a** GWR residuals, **b** OLS residuals

Values closer to zero imply that parameters are more locally estimated, while larger bandwidths indicate that relationships tend to be more global over space. Also worth highlighting is that these cycles tend to repeat every 3 years. For example, starting from 1980, the value of the adaptive bandwidth decreases during the next 2 years and abruptly increases in 1983 reaching a value close to one. Subsequently, the adaptive bandwidth decreases in 1984 and 1985 and then abruptly increases reaching a value close to one in 1986. This situation is repeated across the time spectrum and disrupted by the crash of 1990 and 1991. It should be noted that a maximum bandwidth of one does not relate to a global (OLS) regression as all the observations will be weighted according their distances to the regression point in model calibration. Again, this is something worth investigating further.

Another complementary way of examining the temporal variability of a spatial process is by time series plots of each regression parameter estimate. As examples, Fig. 4 shows the time series plots of the spatially averaged GWR local estimates for the parameters FLRAREA semi-detached and Log_e(DISTCL). Figure 4a depicts an increasing trend in house price per square metre for semi-detached houses over time, ranging from £400/m^2 in 1980 to £1200/m^2 in 1998. Although there are some declines in the value of semi-detached properties between 1982 to 1985, in 1989, and from 1990 to 1992, the market always recovers fairly quickly. Figure 4b depicts the price gradient of houses across London from the centre to the periphery. A value of 0 would indicate constant prices across the city, *ceteris paribus*. Positive values indicate the periphery is valued more highly than the centre, whereas negative estimates indicate the opposite. The general trend is that the centre has become relatively more expensive than the periphery with the trend very evident in the last 5 years covered by the data set. However, there were periods where prices in the periphery either rose faster or declined slower than in the centre: 1986–1989 and 1990–1993. Interestingly, the crash of 1990 falls in the middle of these two exceptional events.

The spatiotemporal variations in GWR results can be examined through surface maps of local parameter estimates. Again, due to limited space, the parameters

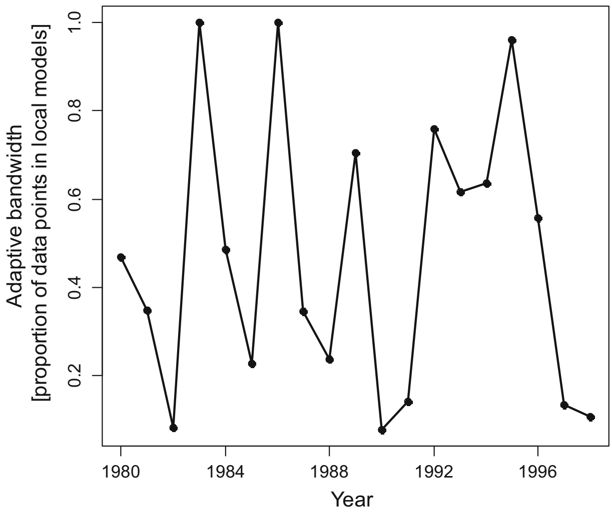


Fig. 3 Temporal variation of the optimal bandwidth: 1980–1998

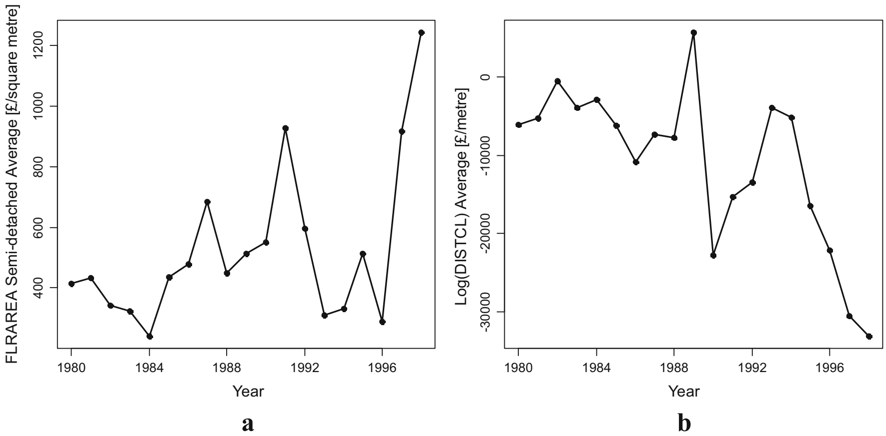


Fig. 4 Temporal variation of FLRAREA semi-detached and $\mathrm{Log_e}$(DISTCL) estimates: 1980–1998, **a** FLRAREA semi-detached, **b** $\mathrm{Log_e}$(DISTCL)

FLRAREA semi-detached and $\mathrm{Log_e}$(DISTCL), and the years 1987–1988 and 1997–1998 are chosen as examples for which the surface maps are displayed in Fig. 5. In general, it seems that the local parameter estimates exhibit a great deal of variation over space and time with increasing spatial evident variation over time. For instance, the local estimates in Fig. 5a (1987) depict a relatively uncomplicated variation in the value (*ceteris paribus*) of semi-detached housing with a simple north–south variation in which houses are slightly more expensive in the northern half of London. By 1988, this pattern becomes stronger with a core of low values merging in centred London just south of the river Thames. By 1998, this pattern has changed dramatically to one where there is a broad swathe of highly priced semi-detached housing/m^2 running

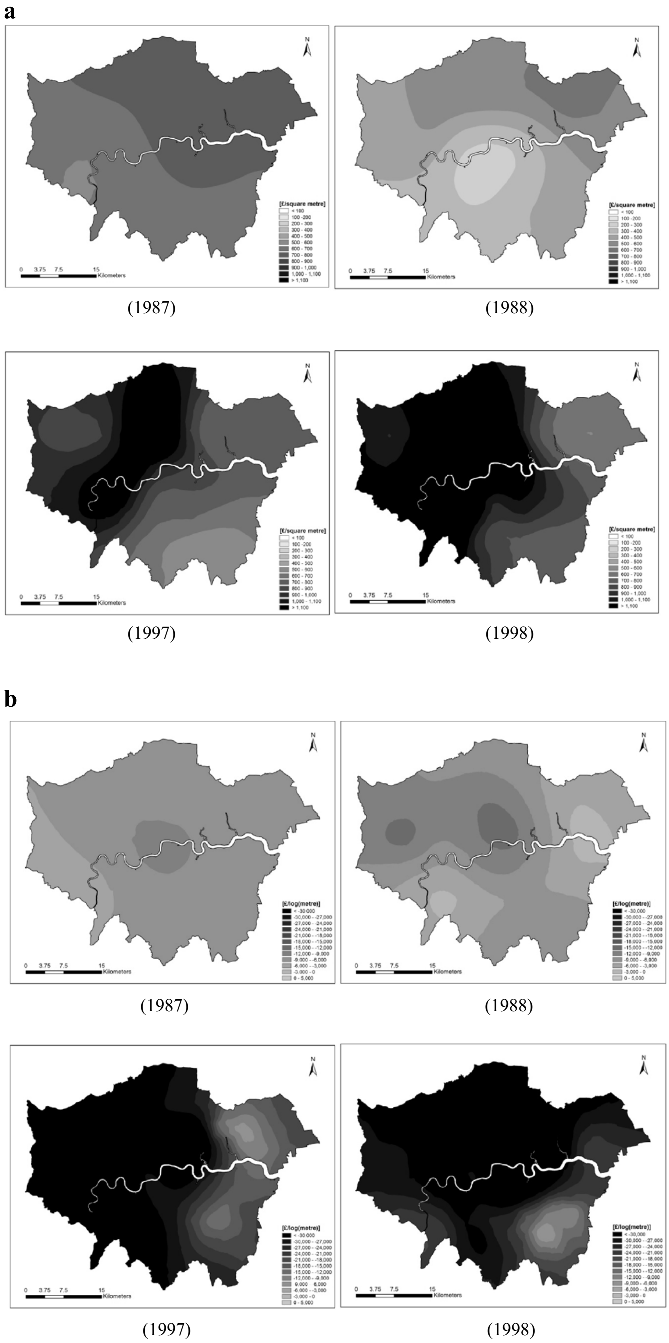
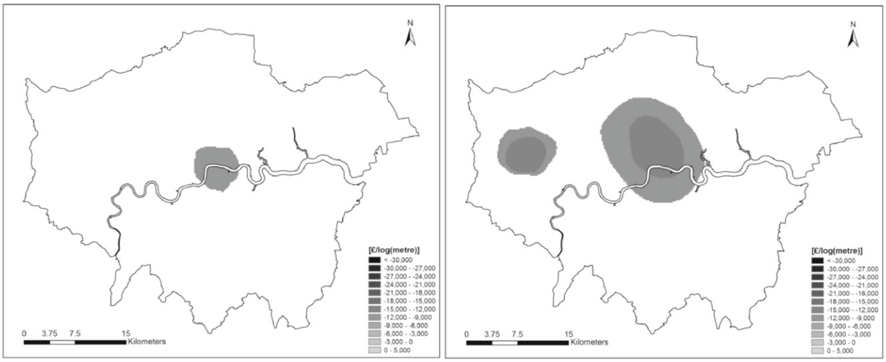


Fig. 5 Spatial variation of local parameter estimate for FLRAREA Semi-detached and $\mathrm{Log_e}$(DISTCL): 1987–1988 and 1997–1998, **a** FLRAREA semi-detached, **b** $\mathrm{Log_e}$(DISTCL)



a 1987 **b** 1988

Fig. 6 Significant local estimates for Log_e(DISTCL): 1987–1988

from north central London to south-west London. Prices/m^2 decline either side of this linear feature.

Similarly in Fig. 5b, the centre–periphery variation in house prices is a relatively simple one with a barely visible circle of higher values in central London. By 1998, the pattern is much more complicated with both a north vs. south and an east vs. west division evident with the south-east exhibiting relatively cheap semi-detached housing.

Finally, another aspect of the GWR results is the statistical significance over space. Consider the Log_e(DISTCL) parameter estimates as an example. In some instances, the distance to the centre of London has a negligible effect on local house prices, making the local estimate of Log_e(DISTCL) at the regression point i statistically non-significant due to the high standard errors associated with the corresponding local estimate. Again, the statistical significance of local estimates can be obtained using local t values. According to Byrne et al. (2009), the adjusted critical t value equated with an original significance level of 0.05 is approximately 3.7 in this case.

Figure 6 shows only the significant local estimates of Log_e(DISTCL) for 1987 and 1988, describing the areas where the relationship between house prices and the distance to the centre of London is statistically significant (i.e. t value < -3.7). It can be seen that in 1987, house prices in only a small part of the central district were significantly influenced by distance to the city centre, whereas in 1988, the area exhibiting a significant relationship expanded in the centre and also included some parts of western London.

4.3 GWR–TS results

As mentioned previously, the sample of 1000 observations from year 1998 was taken as a validation set, so both GWR–TS and OLS–TS methods can be used to predict house prices in 1998 at those sampling locations. To measure the overall performance of the forecasting methods, the predicted values are compared with the actual observed house prices using mean error (ME) and root mean squared error (RMSE) following the Eqs. (9) and (10).

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(P_{i,1998} - \hat{P}_{i,1998}\right) \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{i,1998} - \hat{P}_{i,1998})^2} \tag{10}$$

The ME for GWR–TS and OLS–TS are −5256 and 4967, respectively. Obviously, there are biases in both forecasting procedures with GWR–TS in general overestimating the observed house prices while OLS–TS underestimating the same prices. Nevertheless, the performance of both methods is quite satisfactory considering that the average house price is about £120,000 in 1998. The RMSE value is 34,991 for GWR–TS and 40,239 for OLS–TS. Moreover, the correlation coefficients between the observed and predicted prices are 0.87 for GWR–TS and 0.83 for OLS–TS (both statistically significant), respectively. Therefore, the results from GWR–TS appear to be more representative of the true values.

Again, spatial variations in forecasted house prices can be analysed through continuous surface maps. In Fig. 7a, c, the forecasted house prices for year 1998 based on GWR–TS and OLS–TS are mapped. When compared with Fig. 1f, it can be seen that the extent of higher prices along the river from GWR–TS is more accurate than that from OLS–TS although both methods overestimate values in central London and south-west along the river.

GWR–TS and OLS–TS can also be used to forecast house prices in London for 10 years in the future beyond 1998. In Fig. 8, the values of averaged forecasted house prices are plotted over time, along with the time series plot of the spatially averaged house prices of the observed data given in Table 1. For comparison purposes, the actual annual average house prices in London during 1998–2007 reported by the Nationwide Building Society are shown in Fig. 8 as well. As can be seen from Fig. 8, both GWR–TS and OLS–TS generally overestimate the average of house prices for the years 1998–2007, but the predictions obtained from GWR–TS are much closer to the actual time series.

In addition to the averaged prices, spatial variations in local estimates and house prices obtained from GWR–TS for future years are also examined using surface maps. Using the parameter BLD60S (houses built in the 1960s) as an example, Fig. 9 shows the surface maps of forecasted values for years 1998, 1999 and 2000 as well as the local estimates derived from GWR for year 1997. The predicted local parameter estimates are, to our knowledge, unique in the GWR literature in that they indicate spatial variations in predicted processes. They depict a strengthening of the spatial trends evident in 1997, i.e. an increasing premium on 1960s housing in north-west London and an increasingly negative view of such housing in south London. The trends suggest strongly that purchasing a 1960s built house in north-west London was a good idea; purchasing a similar house in south London was not. The intense gradient between these two parts of London is fascinating and the crossover point is at the river Thames, a discovery made without any geographical information on the river being supplied.

Finally, using the predicted local parameters, it is possible to generate predictions of house prices across London for future time periods. Predictions for 1999 and 2000

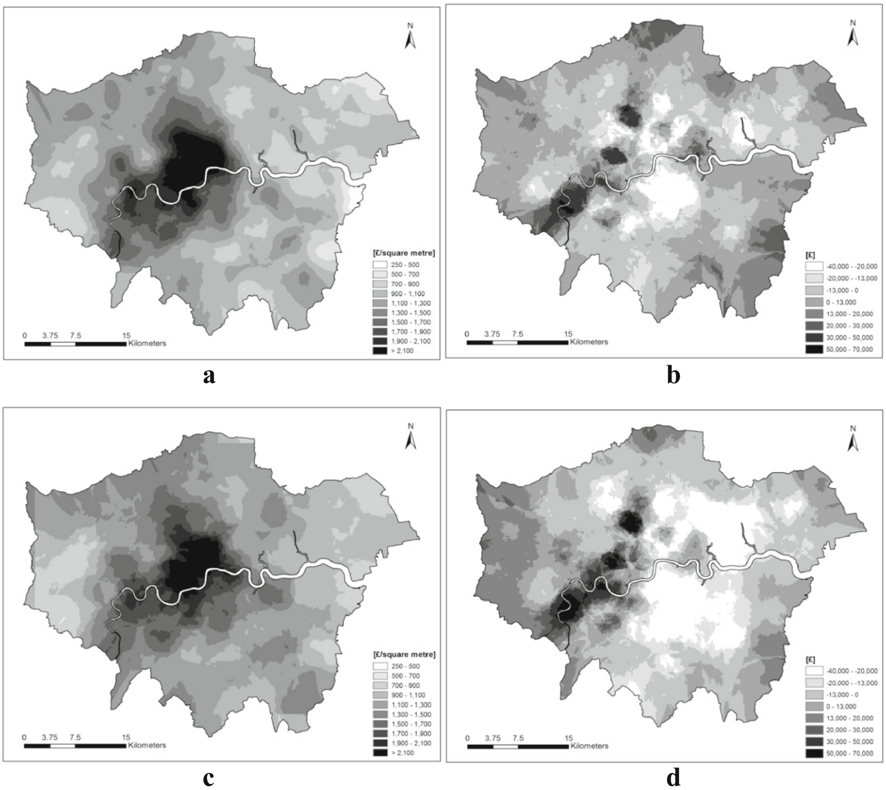


Fig. 7 Surface map of forecasted house prices: 1998, **a** surface map of forecasted house prices: 1998 (GWR), **b** surface map of residuals (GWR), **c** surface map of forecasted house prices: 1998 (OLS), **d** surface map of residuals (OLS)

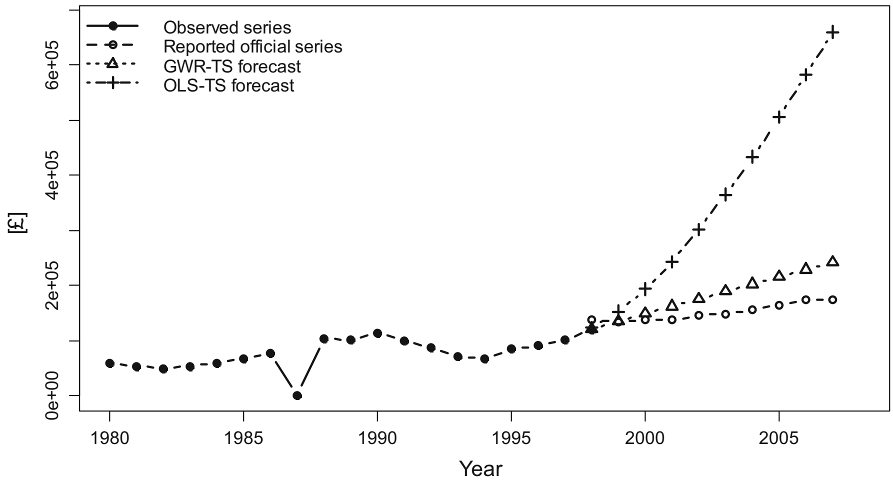
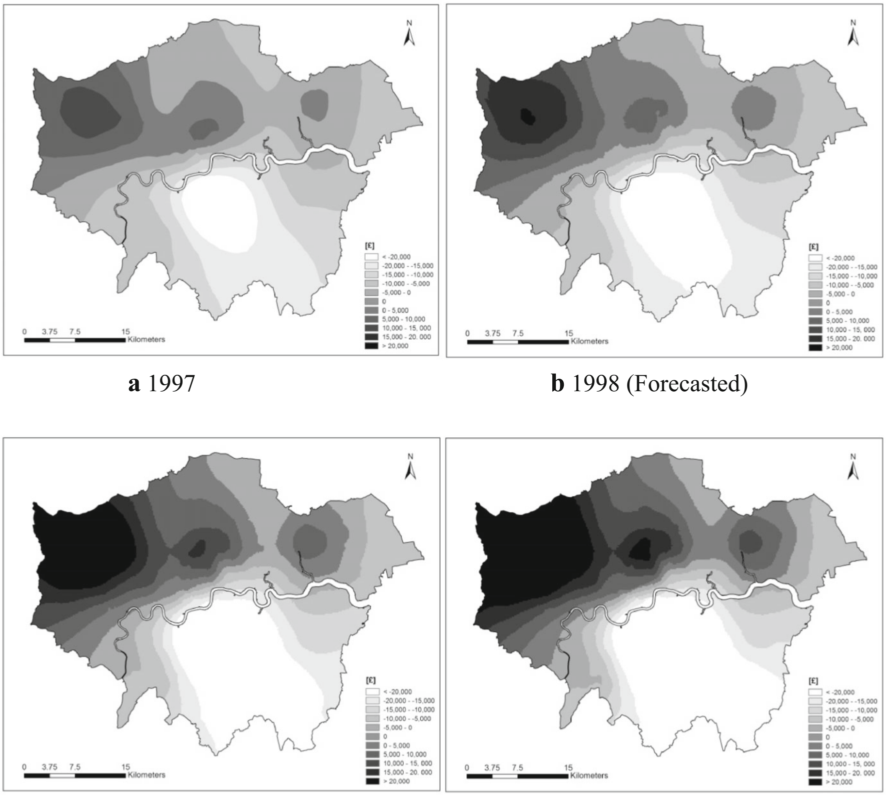


Fig. 8 Time series of forecasted and observed house prices



a 1997 **b** 1998 (Forecasted)

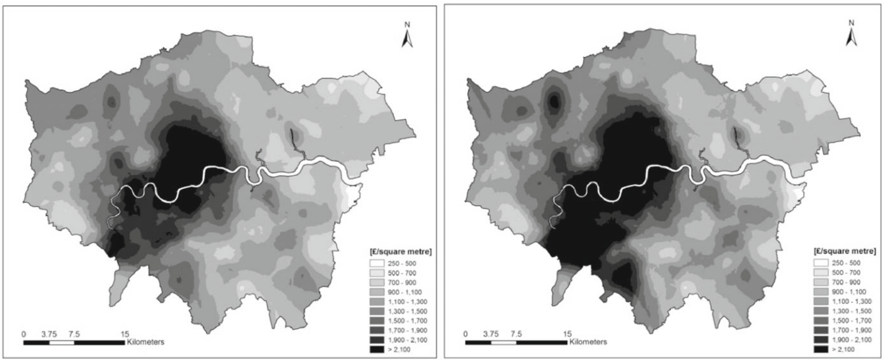
c 1999 (Forecasted) **d** 2000 (Forecasted)

Fig. 9 Spatial variation of local parameter estimate for BLD60S: 1997–2000

are shown in Fig. 10. The trend is one of a strengthening of house prices in central London and in a band towards the south-west. House prices rise elsewhere across London but not at such a high rate as in the centre and the southwest.

5 Discussion and conclusions

This paper examines spatiotemporal variations in house prices in London for the time period 1980–1998 and predicts both local parameters and house prices using an innovative GWR–TS model. Given the potential spatial non-stationary in the housing price process, the local spatial modelling technique GWR is employed to calibrate a hedonic price model constructed for the London house price data. Results show that the local parameter estimates vary not only over space but also over time. The spatiotemporal variability of local estimates is then used to predict future values of local parameters and house prices. This is carried out using the proposed novel method, GWR–TS, a combination of GWR and time series forecasting techniques. GWR–TS outperforms OLS–TS in terms of both overall model fit and house price prediction.



a 1999 **b** 2000

Fig. 10 Surface map of forecasted house prices: 1999–2000

One interesting finding in this research is that the clusters of either high or low values of local estimates tend to gradually expand or contract over space from 1 year to the next as shown in Fig. 5. This type of spatial trend can be seen as a natural growth or contraction of sub-markets (measured through the spatial distribution of local estimates) caused by periods of acceleration or downturn in the housing market over time.

One concern here is to compare GWR–TS with other methods, particularly local spatiotemporal models. For example, Case et al. (2004) presented four models, and Huang et al. (2010) and Fotheringham et al. (2015) proposed temporal extensions of GWR, GTWR. It should be noted that all those approaches forecast real estate sales from different perspectives. Chatfield (2004) claims that there are two primary types of quantitative approaches to forecast future values of a time series: univariate and multivariate. The former is based on a model fitted only to present and past observations such as moving average, ARIMA and exponential smoothing, while the latter depends at least partially on one of more explanatory variables and includes such as classic econometric techniques such as those proposed by Case et al. (2004), Huang et al. (2010) and Fotheringham et al. (2015). Chatfield (2004) also pointed out that the choice of method largely depends on a variety of factors such as the properties of the time series under study and the purpose of forecasting. The GWR–TS proposed in this paper is a mix of multivariate and univariate procedures, where a local multivariate econometric technique (GWR) is used to estimate local parameters over time, followed by forecasting future house prices with a univariate procedure (exponential smoothing). Again, the main purpose here is to introduce this novel and hybrid approach of univariate and multivariate techniques. Admittedly, the absence of comparison of GWR–TS with other methods in existing literature in relation to house price prediction is a limitation of this research.

Finally, given the discussion above, future studies can be conducted to possibly improve the methodology by incorporating the effects of varying bandwidth sizes for different samples into the statistical inference. Possible ways include (1) using a logarithmic form of the response variable, (2) using averaged bandwidth sizes from different samples or (3) using the Bayesian GWR approach suggested by LeSage (2001) in which a posterior probability distribution of the bandwidth is derived so that inferences regarding how sensitive the GWR estimates are to alternative values of the bandwidth can be drawn.

Despite these issues, GWR–TS can be seen as a useful technique for forecasting house prices based on the novel approach which predicts spatial variations in local parameters from a hedonic model and uses them to forecast future prices. The technique generates very localized information on the processes influencing the complex housing market across a major city.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Basu S, Thibodeau TG (1998) Analysis of spatial autocorrelation in house prices. J Real Estate Finance Econ 17(1):61–85

Bitter C, Mulligan G, Dall'erba S (2007) Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. J Geogr Syst 9:7–27

Brundsdon C, Fotheringham AS, Charlton ME (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr Anal 28:281–298

Byrne G, Charlton M, Fotheringham AS (2009) Multiple dependent hypothesis tests in geographically weighted regression. Paper presented at the 10th international conference on geocomputation, Sydney, Australia, November 30–Decemeber 2, 2009

Can A (1990) The measurement of neighborhood dynamics in urban house prices. Econ Geogr 66(3): 254–272

Can A (1992) Specification and estimation of hedonic housing price models. Reg Sci Urban Econ 22(3): 453–474

Case B, Clapp J, Dubin R, Rodriguez M (2004) Modeling spatial and temporal house price patterns: a comparison of four models. J Real Estate Finance Econ 29(2):167–191

Chatfield C (2004) The analysis of time series, 6th edn. Chapman & Hall/CRC, Boca Raton

Cliff D, Ord J (1981) Spatial processes: models and applications. Pion, London

Fotheringham AS, Brunsdon C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester

Fotheringham AS, Crespo R, Yao J (2015) Geographical and Temporal Weighted Regression (GTWR). Geogr Anal (in press)

Gardner ES (2006) Exponential smoothing: the state of the art—part II. Int J Forecast 22(4):637–666

Goodman AC (1978) Hedonic prices, price indices and housing markets. J Urban Econ 5(4):471–484

Goodman AC (1998) Andrew Court and the invention of hedonic price analysis. J Urban Econ 44(2): 291–298

Goodman AC, Thibodeau TG (1998) Housing market segmentation. J Hous Econ 7(2):121–143

Goodman AC, Thibodeau TG (2003) Housing market segmentation and hedonic prediction accuracy. J Hous Econ 12(3):181–201

Griffith DA (1988) Advanced spatial statistics: special topics in the exploration of quantitative spatial data series. Kluwer, Dordrecht

Haining R (1990) Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge

Harris P, Fotheringham AS, Crespo R, Charlton M (2010a) The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. Math Geosci 42(6):657–680

Harris P, Fotheringham AS, Juggins S (2010b) Robust geographically weighted regression: A technique for quantifying spatial relationships between freshwater acidification critical loads and catchment attributes. Ann Assoc Am Geogr 100:286–306

Huang B, Wu B, Barry M (2010) Geographically and temporally weighted regression for modelling spatio-temporal variation in house prices. Int J Geogr Inf Sci 24(3):383–401

Hyndman R, Khandakar Y (2008) Automatic time series forecasting: the forecast package R. J Stat Softw 27(3):1–22

LeSage J (2001) A family of geographically weighted regressions. Lecture Notes. Department of Economics, University of Toledo. http://www.spatial-econometrics.com/html/bgwr.pdf

Lu B, Charlton M, Harris P, Fotheringham AS (2014) Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. Int J Geogr Inf Sci 28(4):660–681

Nakaya T, Fotheringham A, Brunsdon C, Charlton M (2005) Geographically weighted Poisson regression for disease association mapping. Stat Med 24:2695–2717

Orford S (2000) Modelling spatial structures in local housing market dynamics: a multilevel perspective. Urban Stud 37(9):1643–1671

Pace RK, Barry R, Clapp JM, Rodriquez M (1998) Spatiotemporal autoregressive models of neighborhood effects. J Real Estate Finance Econ 17(1):15–33

Páez A (2009) Recent research in spatial real estate hedonic analysis. J Geogr Syst 11(4):311–316

Páez A, Farber S, Wheeler D (2011) A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. Environ Plan Part A 43(12):2992–3010

Páez A, Long F, Farber S (2008) Moving window approaches for hedonic price estimation: an empirical comparison of modeling techniques. Urban Stud 45(8):1565–1581

Wheeler DC (2007) Diagnostic tools and a remedial method for collinearity in geographically weighted regression. Environ Plan A 39(10):2464–2481

Yang W, Fotheringham AS, Harris P (2012) An extension of geographically weighted regression with flexible bandwidths. In: Proceedings of the GISRUK 20th annual conference