Geographically Varying Coefficient Regression: GWR-Exit and GAM-On?

Alexis Comber
University of Leeds, UK

Paul Harris
Rothamsted Research, Harpenden, UK

Daisuke Murakami
Institute of Statistical Mathematics, Tokyo, Japan

Narumasa Tsutsumida
Saitama University, Japan

Chris Brunsdon
Maynooth University, Ireland

Abstract

This paper describes initial work exploring two spatially varying coefficient models: multi-scale GWR and GAM Gaussian Process spline parameterised by observation location. Both approaches accommodate process spatial heterogeneity and both generate outputs that can be mapped indicating the nature of the process heterogeneity. However the nature of the process heterogeneity they each describe are very different. This suggests that the underlying semantics of such models need to be considered in order to refine the specificity of the questions that are asked of data: simply seeking to *understand process spatial heterogeneity* may be too semantically coarse.

2012 ACM Subject Classification Information systems; Theory of computation

Keywords and phrases Geographically weighted regression, Spatial Analysis, Process Spatial Heterogeneity, Model Semantics

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.13

Category Short Paper

Funding This work was supported by ROIS-DS-JOINT (005RP2021), https://ds.rois.ac.jp/en/.

1 Introduction

Geographically varying regression models are those that estimate coefficients locally rather than globally. A key feature of such models is that they accommodate process spatial heterogeneity and support local understandings of how the relationship between different inputs and an outcome vary spatially. Geographically Weighted Regression (GWR) [2] is the best known method to calibrate spatially varying regression models. It uses a moving window (kernel) weighted regression centred on locations in the study area, to estimate local coefficients. The Geographically Weighted (GW) framework has been extended to other statistical methods as well as regression such as GW-PCA [10], GW Discriminant Analysis [3] and GW correspondence matrices [4]. The GWR framework has also been extended to accommodate parameter (predictor variable) specific bandwidths in Multi-Scale GWR (MS-GWR) [18, 8, 13].

Determining the kernel bandwidth (size) in any GW analysis is critical as this defines the variation in the local outputs (i.e. the degree of smoothing). Optimal bandwidths are identified through some measure of model fit. Thus calibrating GWR model bandwidth(s) in this way provides an indicator of the spatial scales over which heterogeneous processes operate, enhancing process understanding. There are some concerns about the identification of optimal bandwidths for GWR and MS-GWR using current search heuristics, as in some cases optimisation searches may return local rather than global optima [6]. The potential for this is particularly acute when the bandwidth search space is multi-dimensional (MS-GWR). This may be because the increased complexity / dimensionality of the bandwidth search space means that the potential for local optima to be identified by search heuristics is greater.

An alternative approach to calibrating local coefficient models can be constructed using Gaussian Processes (GPs) to model terms in Generalised Additive Models (GAMs) [17, 7]. A GP is a random process over *functions* and GAMs are a general approach to calibrating regression models with unspecified functions of the predictor variables, of the form:

$$y = \alpha + f_1(z_1) + f_2(z_2) + \cdots + f_m(z_m) + \epsilon$$

where z_j may be a vector.

These can be extended so that each $f_j(z_j)$ is a linear regression coefficient on another predictor x_j:

$$y = \alpha(z_0) + x_1 f_1(z_1) + x_2 f_2(z_2) + \cdots + x_m f_m(z_m) + \epsilon$$

Finally, if $z_0 = z_1 = \cdots z_m = z$ say, and z is a vector specifying spatial *locations* then this specifies a *geographically varying regression model*:

$$y = \alpha(z) + x_1 f_1(z) + x_2 f_2(z) + \cdots + x_m f_m(z) + \epsilon$$

One way of specifying $\alpha(z) \cdots f_m(z)$ is that each function is generated from a GP and each function estimate is an *a posteriori* estimate of a GPs with a zero mean. GPs also have a covariance function:

$$\kappa_m(\delta) = \mathrm{Cov}(f_m(z), f_m(z + \delta))$$

These control the “smoothness” of $f_m(z)$ - the more rapidly $\kappa_m(\delta)$ reduces as the magnitude of δ increases, the “smoother” $f_m(z)$ tends to be. These are similar to models based on Kriging as semivariogram functions are related to covariance functions. In a similar way to MS-GWR, the covariance function for each $f_m(z)$ is individually calibrated to optimise model fit. One task of the GAM is to estimate parameters in each $\kappa_j(\delta)$ and so estimate $f_m(z)$.

Thus both MS-GWR and GAM GPs with a GP smooth construct spatially varying coefficient models: both require the degree of smoothing to be determined or specified, with this done through the optimisation of the bandwidth for each predictor variable via a back-fitting operation for MS-GWR, and in a GAM GP spline over geographic space, this is similarly determined through a smoothing parameters for each GP. Both provide a measure of the process heterogeneity specific to each predictor variable in a regression.

The aim of this paper is to explore the complementaries between MS-GWR and GAM GPs specified with observation spatial locations as different approaches for specifying geographically varying regression models in terms of the process understanding (the scale of spatial heterogeneity) they support. It will also reflect on how fit GA / GP and MS-GWR models can be optimised and issues around using them for prediction.

2 Background: GWR and GAM GP

2.1 GWR

Geographically Weighted Regression (GWR) [2] is spatially varying coefficient model, that uses a kernel based approach to create a series of local regression models, for which local coefficient estimates of the predictor variable can extracted and mapped. GWR attempts to calibrate regression models of the form:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{1i} + \cdots + \beta_m(u_i, v_i)x_{mi} + \epsilon_i$$

where y_i and $\{x_{1i}, \cdots, x_{m_i)}\}$ for $i = 1, \cdots, n$ are a set of observations with m predictor variables and a response, $\{\beta_0(.,.), \cdots \beta_m(.,.)\}$ are functions of two variables providing a mapping from location to a regression coefficient, (u_i, v_i) for $i = 1, \cdots, n$ are locations associated with each of the n observations, and ϵ_i is a random variable, typically from a Normal distribution. The functions $\beta_j(.,.)$ are usually of most interest, and after a model is calibrated these are typically illustrated cartographically. GWR estimates these functions using data subsets falling under a weighted kernel around a point (u, v) to calibrate a local weighted least squares regression. The kernel generally takes a distance decay function such as a Gaussian decay of the form:

$$w_i = \exp\left(-\frac{d_i^2}{2h^2}\right)$$

where d_i is the distance from (u, v) to (u_i, v_i) and h is a quantity termed the "bandwidth" that determines the size of the regression window. Bandwidths can be a fixed distance or a fixed number of nearest data points (i.e., an adaptive radius depending on the local density of points. If h is large, the functions $\beta_j(.,.)$ become smoother and thus determining the size of bandwidth in any GWR analysis is critical as this defines the variation in the local outputs (i.e. the degree of smoothing). Various approaches to finding an "optimal" h for fitting a model to a given data set exist and generally these try to optimise some measure of model fit and parsimony, such as AIC.

A standard GWR operates determines a single bandwidth and thus implicitly assumes that each input variable operates over the same scale with respect to its relationship to the response variable. In reality some relationships may operate over larger scales than others and a standard GWR finds a *best-on-average* scale of relationship non-stationarity [5]. Multi-Scale GWR (MS-GWR) can be used to address this [18, 8, 13]. It determines a bandwidth for each predictor variable plus the intercept individually, thereby allowing individual response-to-predictor variable relationships to vary. Recent thinking has suggested that because of this, MS-GWR should be the default GWR [5], and a standard GWR only used if there is evidence to support a single scale of relationship (and bandwidth).

2.2 Statistical Tensions with GWR

GWR has at its core the idea that "whole map" global statistical models ignore any process heterogeneity and implicitly assume it does not exist. GWR is attractive to geographers and GIScience because it shows how and where processes vary spatially and because it explicitly reflects Tobler's First Law of Geography [15]. This tension between advocates of global models in classic statistics and local models in spatial statistics can be observed in some of the critiques of GWR and other nonstationary models [16, 14]. On one side, the main critique is that GWR and nonstationary models provide only a collection of local models in order to model a non-stationary process (in this case the coefficients), whereas Bayesian models [9], for example, provide a full single model able to capture a non-stationary process. A further critique is that if the global model has locally clustered outliers (a key indicator of the potential suitability of a GWR model), then some explanatory variables are missing, or the process under investigation has been poorly represented by the model inputs or some theoretical understanding of the process is lacking (and not represented in the predictor variables).

On the other, advocates of GWR and local statistical analyses argue that whole map regression models may unreasonably assume stationary regression coefficients [12] and process heterogeneity (spatial variation in data relationships), whereas in reality the processes *do* vary spatially, for example the relationship between crime and unemployment levels is not the same everywhere. And potentially these arguments are more pertinent to socio-economic processes, which are more likely to be specifically concerned with *how* processes manifest themselves in different socio-economic (local) contexts local, while many environmental or physical processes have fixed global (mathematical) relationships - i.e. laws. An additional related argument is that socio-economic analyses are frequently in the situation where ideal data are never available for actual real-world analysis, as opposed to simulated data commonly used in theoretical statistics. Large swaths of GIScience and geography deal only in secondary data – data that was collected by someone else, usually for a different purpose – and frequently have to use proxies for the data they would really like to use. Very rarely are bespoke data, collected under experimental design and with full understanding, used in geographical analyses [1], especially socio-economic ones. In this sense, the use of local statistical models or global ones, and whether you believe in a global truth or local process understanding, are a bit like a religious belief: it either makes sense (conceptually or practically) or it does not, and no amount of logic will sway opinion.

2.3 GAMs, Bases and Gaussian Processes

A Generalized Additive Model (GAM) uses smooth functions of the predictor variables in which the values of y are assumed to be of an exponential distribution, such as a Gaussian one. If

$$y = f(x) + \epsilon$$

where f is the function being sought in the model, then in GAMs, rather than assuming y to be some linear function of x, a space of functions, or *basis*, is chosen of which f is some element. This allows the basic formula above to be expanded:

$$y = f(x) + \epsilon = \sum_{j=1}^{d} \beta_j(x)\gamma_j + \epsilon$$

where each β_j is a basis function of the transformed x and the γ are the corresponding regression coefficient estimates. One example of a basis is a Gaussian Process basis. If there are n distinct geographical locations in the data set, knowing the locations and the covariance function κ the variance covariance function of the values of β_j in each location can be found, giving a variance covariance matrix R. This can be translated into a set of n basis vectors $\beta_j(x)$ [11], and the GAM can be calibrated in this way. Thus, in contrast to standard linear models, the predictors in a GAM include smooth functions of some or all of the covariates, which allow for non-linear relationships between the predictors and the target variable.

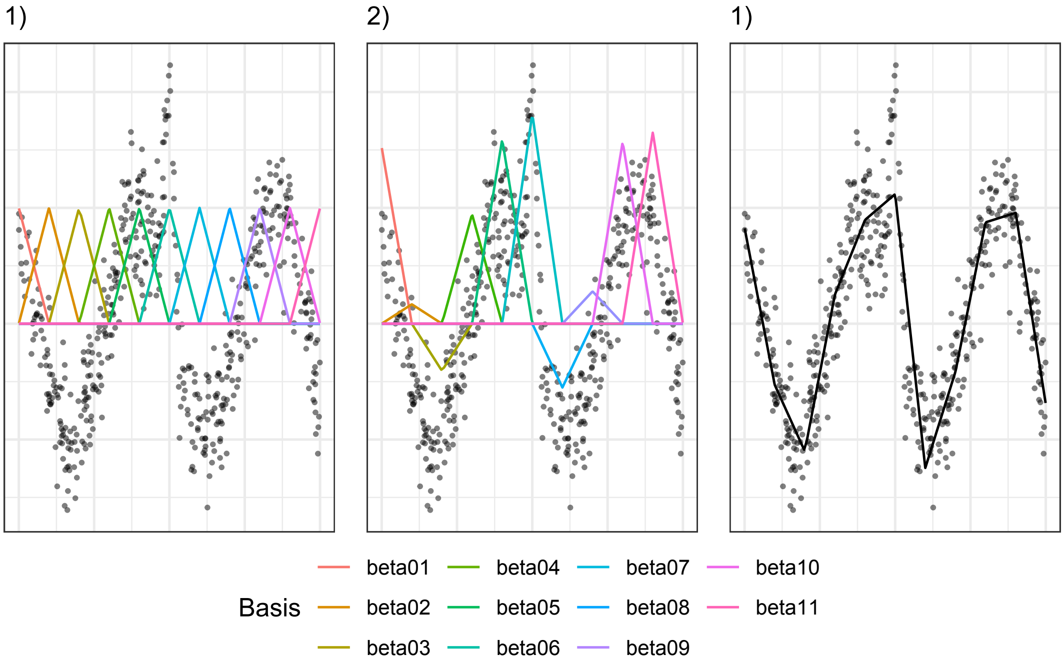


Figure 1 The working of a GAM spline, with simulated x and y data: the LHS plot shows a linear regression fitted between knots, the centre shows each basis multiplied by the corresponding piecemeal linear regression coefficient, and the RHS plot the sum of the basis functions.

GAMs have at their heart the fitting of a series of non-linear functions through the data as illustrated in Figure 1. In this, the various increasing and decreasing functions (centre panel) indicate the slope of coefficients from the data defined by a set of x-points called *knots* (LHS panel). The RHS panel shows the sum of the basis functions, and is equivalent to taking the fitted values from a regression on the basis expansion of x. Is this local fitting that starts to hint at how GAMs can be used with geographic data, where not only are the splines constructed in attribute space but over geographical space if that is included in the inputs to the spline. It suggests that GAMS have the potential to bridge between the need for local, spatial understanding such as is provided by a GWR analysis and for global models, as well as for the enhanced predictive power of non-linear statistical models.

3 Analysis

3.1 Overview and Data

Socio economic data from the `gw` R package is used to illustrate both spatial understanding from GWR and spatial prediction using GAMs. This has census data for the counties in state of Georgia in the USA from the 1990s. It has 159 observations and 6 variables of interest, median income, % of the population that is rural, % with degrees, % elderly, % foreign born and % black. The analyses below construct a MS-GWR model and a GP-derived GAM spline model of Median Income. Both generate local coefficient estimates, which can be mapped.

3.2 MSGWR vs GAM with GP splines

MS-GWR analyses require the explicit identification of the individual bandwidths for each covariate. These provide important information about spatial scales at which the relationship between the predictor variable and the target variables operate. Small bandwidths indicate a local scale and large ones a global scale. It is quite common for these to vary from highly local to highly global within a single MS-GWR analysis.

The MS-GWR model shows the covariates to have a range of adaptive bandwidths (Table 1), and the varying degrees of process heterogeneity are shown by the distributions of the coefficient estimates in Table 1 and by spatial distribution in Figure 2. In this case the bandwidths indicate that the variables for % rural (`PctRural`), % with degrees (`PctBach`) and % elderly (`PctEld`) are highly localised whereas % foreign born (`PctFB`) and % black (`PctBlack`) are global. These are also indicated by the variation in the coefficient estimates for the covariates in Figure 2.

Table 1 The MS-GWR bandwidths (BW) and distribution of the coefficient estimates.

	BW	Min	1stQ	Median	Mean	3rdQ	Max
Intercept	25	27.98	38.85	46.12	45.60	52.45	59.97
PctRural	88	0.04	0.06	0.10	0.10	0.14	0.16
PctBach	26	0.02	0.58	0.95	0.87	1.15	1.82
PctEld	31	−2.46	−1.81	−1.42	−1.48	−1.15	−0.48
PctFB	157	−1.48	−1.46	−1.41	−1.37	−1.30	−1.09
PctBlack	157	−0.24	−0.23	−0.22	−0.22	−0.21	−0.20

In a similar way, Gaussian Process splines can be used in a GAM model, parameterised with observation location, in this case projected Easting and Northing. Each spline was specified with 7 knots in order to ensure sufficient degrees of freedom across the data and the splines. Splines optimise a smoothing parameter which controls the degree of smoothing of the data and as such indicates the locally varying nature of the coefficient. The GPs modelled in the GAM function all have a mean of zero, so for each covariate an extra fixed offset term is added. The fixed terms are shown in Table 2 and the spatially smoothed terms in Table 3. Here it can be seen that of the fixed terms, the Intercept, % with degrees (`PctBach`), % elderly (`PctEld`) and % black (`PctBlack`) are globally significant, while the Intercept, % with degrees (`PctBach`), % elderly (`PctEld`) and % black (`PctBlack`) are locally significant.

It is also possible to map locally significant predictors of Median Income arising from the GAM splines as in Figure 3. The trends in smoothed coefficients broadly show East-West gradients for the Intercept, % with Degree and % Elderly, and North-South ones for % Black.

Table 2 The coefficient estimates of the GAM fixed terms.

	Estimate	Std. Error	t-value	p-value
Intercept	46.398	4.004	11.588	0.000
PctRural	−0.541	0.882	−0.613	0.541
PctBach	0.356	0.179	1.985	0.049
PctEld	−0.482	0.112	−4.313	0.000
PctFB	−0.445	0.294	−1.513	0.133
PctBlack	−0.159	0.023	−6.840	0.000

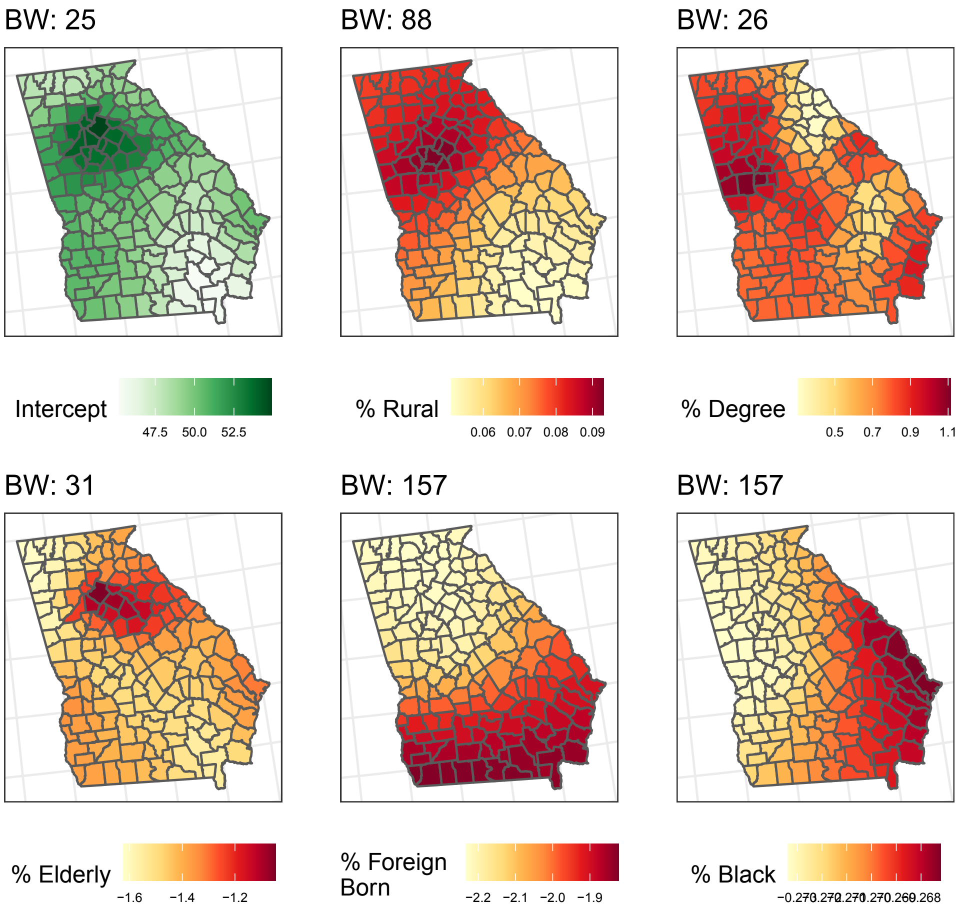


Figure 2 MS-GWR coefficient estimates.

Table 3 The GAM spatially smoothed terms from the GP splines with location.

	Effective df	Ref. df	F	p-value
s(X,Y):Intercept	15.538	18.932	2.093	0.008
s(X,Y):PctRural	5.616	6.044	2.081	0.075
s(X,Y):PctBach	2.517	2.529	2.780	0.031
s(X,Y):PctEld	2.500	2.500	4.812	0.015
s(X,Y):PctFB	2.500	2.500	0.797	0.614
s(X,Y):PctBlack	2.500	2.500	11.675	0.000

3.3 MSGWR bandwidth vs GAM Spline smoothing parameter

The MS-GWR and the smooth terms from the GAM GPs splines constructed with locations, both construct spatially varying coefficient models. They also both include some measure of the degree of local smoothing: in a MS-GWR this is specified through the optimisation of the bandwidth for each predictor variable via a back-fitting operation and in a GAP GP this is indicated through smoothing parameters for each spline. Importantly both methods provide a measure of the process heterogeneity that is specific to each predictor variable. However, it is evident from Figures 2 and 3 that the way that the spatial processes are being modelled by the 2 approaches is very different. The MS-GWR results have distinct but different locales

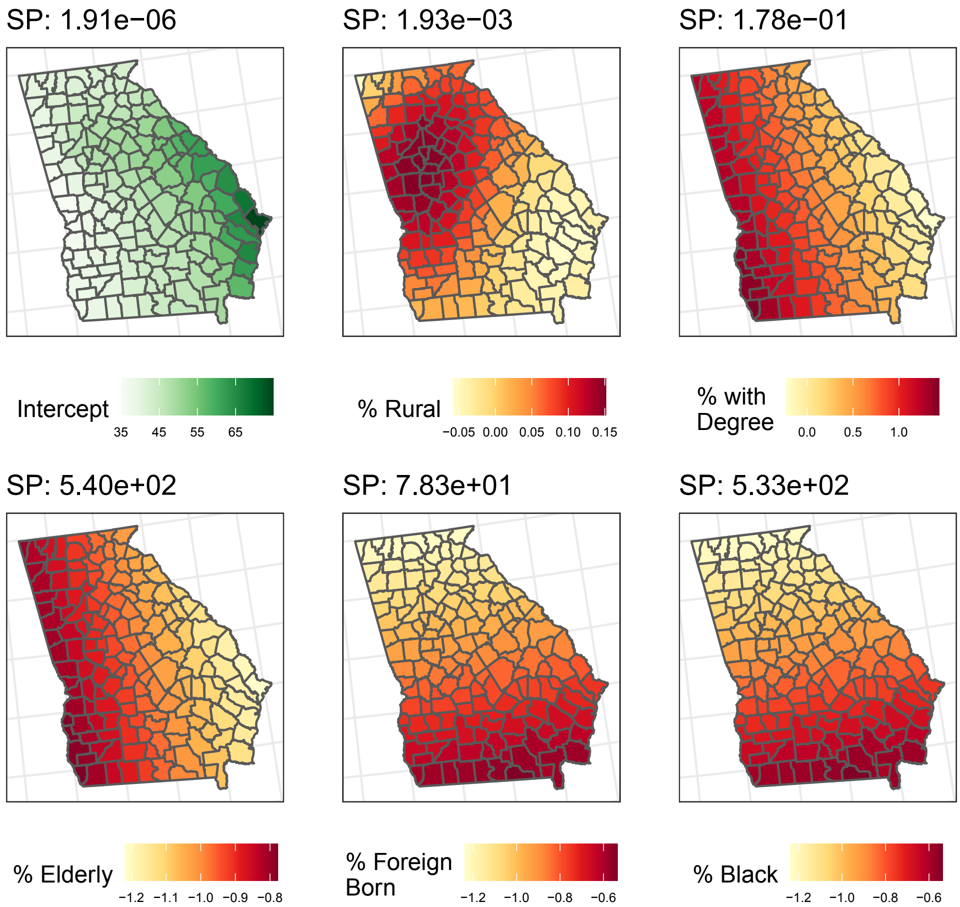
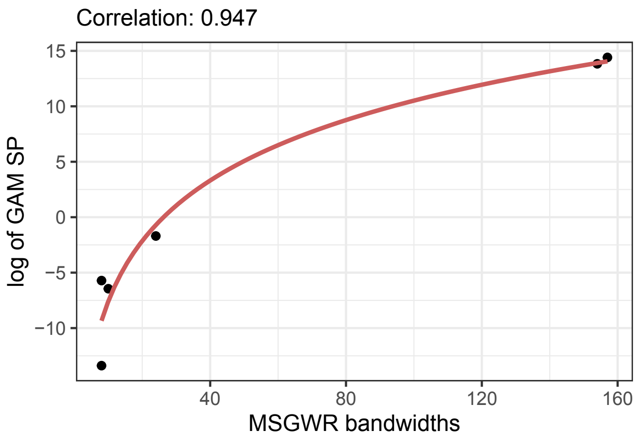


Figure 3 The local coefficient arising from the spatial GAM splines.

of high coefficient estimate value for the different covariates. For example, the association of % with Degree with Median Income is much greater in a region located in the centre of the Western part of Georgia under a MS-GWR, whereas under the GAM it is in the South West corner.

Evidently, despite using the same data, and generating spatially varying coefficient estimates, the 2 approaches are very different. It would be useful to better understand how their results relate, if at all, either to be able to link them, or to be able to infer the circumstances in which each approach may be most useful, under the assumptions that MS-GWR provides an intuitive understanding of the spatial of the relationship between target and individual predictor variables and GAMs probably have a stronger theoretical background.

To investigate how these different ways of capturing and modelling process spatial heterogeneity, relate if at all, the MSGWR bandwidths and GP spline smoothing parameters were examined. The smoothing parameters and the MSGWR coefficients were extracted from the models A correlation showed a high degree of fit for and when they were plotted as in Figure 4, there is a clear trend between them as shown in the tabular values and in the scatter plot. This shows, that at least in this example, there is some form of relationship between MSGWR bandwidth and smoothing parameters of Gaussian Process splines.



Term	MSGWR	GAM
Int	8	1.53e−06
PctBach	8	3.29e−03
PctEld	10	1.58e−03
PctRural	24	1.83e−01
PctBlack	154	1.02e+06
PctFB	157	1.81e+06

Figure 4 MSGWR bandwidths (x-axis) against the log of the GAM GP spline smoothing parameter (SP) (y-axis) with a log trend, and a table of the data included.

4 Final Comments

Spatially vary coefficient models are useful because they explicitly accommodate process spatial heterogeneity, where the relationships between an outcome and factors used to model or predict that outcome, may change with location. The locally varying coefficient estimates that are generated by such approaches provide an explicit representation of process spatial heterogeneity that can be easily communicated via maps. This features has underpinned the popularity of GWR and and other GW models.

In this work, the same process (Median Income in 159 counties of the state of Georgia) is clearly being modelled in different ways by two different spatially varying coefficient models, MS-GWR and GP splines with location in a GAM. The maps in Figures 2 and 3 show these differences, and provide an indication of the different model semantics. It suggests that the concept of "process spatial heterogeneity", which is frequently referred to in the GIScience and spatial analysis literature with a link to Tobler [15] needs to be refined. The results of the two methods used here indicate the need for more nuance in the way that we ask data for "process understanding".

There is evidently a relationship between MSGWR bandwidths and GP spline smoothing parameters. The primary role of GAMs is prediction, but here we have shown that when constructed over geographic space, they can be used to generate locally estimated coefficients. Further work will explore GAM GP splines over other spatial datasets including simulated data with known spatial properties. It may extend the splines to the temporal domain.

References

1 Chris Brunsdon and Alexis Comber. Opening practice: supporting reproducibility and critical spatial data science. *Journal of Geographical Systems*, 23(4):477–496, 2021.

2 Chris Brunsdon, A Stewart Fotheringham, and Martin E Charlton. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298, 1996.

3 Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted discriminant analysis. *Geographical Analysis*, 39(4):376–396, 2007.

4 Alexis Comber, Chris Brunsdon, Martin Charlton, and Paul Harris. Geographically weighted correspondence matrices for local error reporting and change analyses: mapping the spatial distribution of errors and change. *Remote Sensing Letters*, 8(3):234–243, 2017.

5 Alexis Comber, Christopher Brunsdon, Martin Charlton, Guanpeng Dong, Richard Harris, Binbin Lu, Yihe Lü, Daisuke Murakami, Tomoki Nakaya, Yunqiang Wang, et al. A route map for successful applications of geographically weighted regression. *Geographical Analysis*, 0:1–24, 2022.

6 Alexis Comber, Nick Malleson, Hang Nguyen Thi Thuy, Thanh Bui Quang, Minh Kieu, Hoang Huu Phe, and Paul Harris. Multiscale geographically weighted discriminant analysis. In *GIScience 2021 Short Paper Proceedings*, pages 17–32. Springer, 2021. doi:10.25436/E2PP4F.

7 Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian D Marx. Regression models. In *Regression*, pages 23–84. Springer, 2021.

8 A Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6):1247–1265, 2017.

9 Alan E Gelfand and Dipak K Dey. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):501–514, 1994.

10 Paul Harris, Chris Brunsdon, and Martin Charlton. Geographically weighted principal components analysis. *International Journal of Geographical Information Science*, 25(10):1717–1736, 2011.

11 Trevor J Hefley, Kristin M Broms, Brian M Brost, Frances E Buderman, Shannon L Kay, Henry R Scharf, John R Tipton, Perry J Williams, and Mevin B Hooten. The basis function approach for modeling autocorrelation in ecological data. *Ecology*, 98(3):632–646, 2017.

12 Stan Openshaw. Developing gis-relevant zone-based spatial analysis methods. *Spatial analysis: modelling in a GIS environment*, pages 55–73, 1996.

13 Taylor M Oshan, Ziqi Li, Wei Kang, Levi J Wolf, and A Stewart Fotheringham. mgwr: A python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information*, 8(6):269, 2019.

14 PD Sampson, D Damian, and P Guttorp. Advances in modeling and inference for environmental processes with nonstationary spatial covariance. In *GeoENV III—Geostatistics for Environmental Applications*, pages 17–32. Springer, 2001.

15 Waldo R Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1):234–240, 1970.

16 Levi John Wolf, Taylor M Oshan, and A Stewart Fotheringham. Single and multiscale models of process spatial heterogeneity. *Geographical Analysis*, 50(3):223–246, 2018.

17 Simon N Wood. *Generalized additive models: an introduction with R*. chapman and hall/CRC, 2006.

18 Wenbai Yang. *An extension of geographically weighted regression with flexible bandwidths*. PhD thesis, University of St Andrews, 2014.