AGILE: GIScience Series, 3, 31, 2022. https://doi.org/10.5194/agile-giss-3-31-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Spatially Varying Coefficient Regression with GAM Gaussian Process splines: GAM(e)-on

Alexis Comber [1, 2], Paul Harris [3], and Chris Brunsdon [4]

[1] School of Geography, University of Leeds, Leeds, UK
[2] Leeds Institute for Data Analytics, University of Leeds, UK
[3] Sustainable Agriculture Sciences, Rothamsted Research, North Wyke, UK
[4] National Centre for Geocomputation, Maynooth University, Ireland

Correspondence: Alexis Comber (a.comber@leeds.ac.uk)

Abstract. This paper describes initial work exploring GAM Gaussian Process (GP) splines parameterised by observation location, as a geographical varying coefficient model. Similar to GWR, this approach accommodates process spatial heterogeneity and generates spatially distributed, local coefficient estimates. These can be mapped to indicate the nature of the heterogeneity. The paper investigates the effect of the smoothing parameters used in the splines and how they alter the nature of the modelled heterogeneity. It optimises these in the GAM GP and the tuned model has subtle but important differences with the initial model. This has impacts on the nature of the process understanding (inference) that can be extracted from the model. This in turn suggest the need examine the underlying semantics of the resultant models in relation to the scale of process suggested by the smoothing parameters. A number of areas of further work are identified.

Keywords. Geographically varying coefficient models, Spatial analysis, Process spatial heterogeneity, Model semantics

1 Introduction

Spatially varying regression models generate local rather than global coefficient estimates. These can be mapped and used to provide measures of process spatial heterogeneity – that is they describe how and where the relationship between the target variable and the predictor variables vary spatially. Probably the best known of these is Geographically Weighted Regression (GWR) (Brunsdon et al., 1996) which uses a moving window or kernel to extract data subsets from which a series of local regression models are constructed. In a GWR the scale of the predictor-to-response relationships is described by the kernel size or bandwidth, as this determines how many data points are included in each local subset. GWR has been extended to a multi-scale GWR (MS-GWR) (Yang, 2014; Fotheringham et al., 2017; Oshan et al., 2019) in which a bandwidth is determined for each predictor-to-response, thereby providing deeper understanding of the scales of the individual scales of relationship, rather than the best-on-average bandwidth provided by standard GWR. Because of this MS-GWR has recently been recommended as the default GWR (Comber et al., 2022). In GWR and MS-GWR models, the variation in the local outputs describes the nature of the process heterogeneity, and this is heavily influenced by the kernel bandwidth as this defines the degree of smoothing. Thus a key step in an GWR is determining or calibrating the bandwidth.

There are a number of criticisms of and contexts to the use of GWR models.

First, while GWR and MS-GWR models are useful tools for understanding process variation and directly accommodate both the spatial autocorrelation indicated Tobler's 1st law (Tobler, 1970) and the principle of spatial heterogeneity or non-stationarity in Goodchild's 2nd law (Goodchild, 2004), they are difficult to use for prediction. This is because the geography, the location, in the models is implicit rather than explicit, with prediction at a new location requiring weighted data around that location to determine the coefficient estimates at that location.

Second, critiques of GWR and other nonstationary models (Wolf et al., 2018; Sampson et al., 2001), argue that they are a collection of local models rather than a full single model able to capture a non-stationary process, as Bayesian models of (Gelfand and Dey, 1994), for example.

Third, there is a tension between global models in classic statistics. The presence of local clustering of residuals / outliers is a key indicator of the potential suitability of a GWR model. However, many in classic statistics argue that this indicates the that key explanatory variables are missing, that the process under investigation has been poorly represented by the model inputs, or the model has been constructed from an incomplete theoretical understanding of the process under consideration.

These critiques can be considered as being grounded in environmental or physical processes which have fixed global mathematical relationships - i.e. laws. Whereas many (most) socio-economic processes do not. For example the relationship between crime and unemployment is not the same everywhere, and whole map global models unreasonably assume stationary processes (Openshaw, 1996). Also many socio-economic analyses have to use less than perfect data: GIScience and spatial analyses generally use secondary data (i.e. data collected by someone else for a different purpose) and proxies for the variables of interest. It is rare for data describing socio-economic process to be collected under full experimental design, with full understanding (Brunsdon and Comber, 2021). In one senses, the choice of local statistical models or global ones will be informed by whether a global truth is believed to exists and the modelling aim is to find it, or whether the aim is to understand local nature of the process under consideration.

Because of the tensions, there is interest in examining how alternative approaches could be used for spatial prediction **and** spatial inference (understanding), and Generalised Additive Models (GAMs) models with Gaussian Process (GP) splines parameterised with location offer a route to do this. The aim of this short paper is to explore the process understanding of these models and their predictive power.

2 GAMs with Gaussian Process splines

A Generalized Additive Model (GAM) uses smooth functions of the predictor variables. These assume the values of y have an exponential distribution, such as a Gaussian one and if

$y = f(x) + \epsilon$

where f is the function being modelled, then GAMs define a space of functions, or *basis*, of which f is some element (rather than assuming y to be some linear function of x).

In this way, GAMS fit of a series of non-linear functions through the data as illustrated in Figure 1. In this, the various functions in the centre graph show the coefficient slopes defined by a set of x-points called *knots* (left graph), with the sum of the basis functions in the right graph, equivalent to fitting values from a regression on the basis expansion of x. This within-the-data local fitting hints at how GAMs can be used with spatial data, where the splines are constructed over an attribute space that includes location, suggesting how they could bridge between local, spatial understanding and enhanced predictive power of non-linear statistical models.

Local coefficient models can be constructed using GPs to model terms in a GAM (Wood, 2006; Fahrmeir et al., 2021). A GP is a random process over *functions* and GAMs are a general approach for calibrating regression models with unspecified functions of the predictor variables. They have 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.

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

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

And, if $z_0 = z_1 = \cdots z_m = z$, and z is a vector specifying spatial **locations** then this specifies a **spatially 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:

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

These control the 'smoothness' of $f_m(z)$ - the more rapidly $\kappa_m(\delta)$ decreases with increases in δ, the 'smoother' $f_m(z)$ tends to be, in a similar way to models based on Kriging (because semivariogram functions are related to covariance functions). And in a similar way to MS-GWR, the covariance function for each $f_m(z)$ is individually calibrated as the GAM estimates parameters in each $\kappa_j(\delta)$ thereby estimating $f_m(z)$.

Thus in a similar way to MS-GWR, GAM GPs with a GP smooth, construct spatially varying coefficient models, requiring the degree of smoothing to be determined through the smoothing parameters for each GP and generating measures of process heterogeneity specific to each predictor variable in a regression.

3 Analysis

3.1 Overview and Data

Socio economic data from the `gw` 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 form the

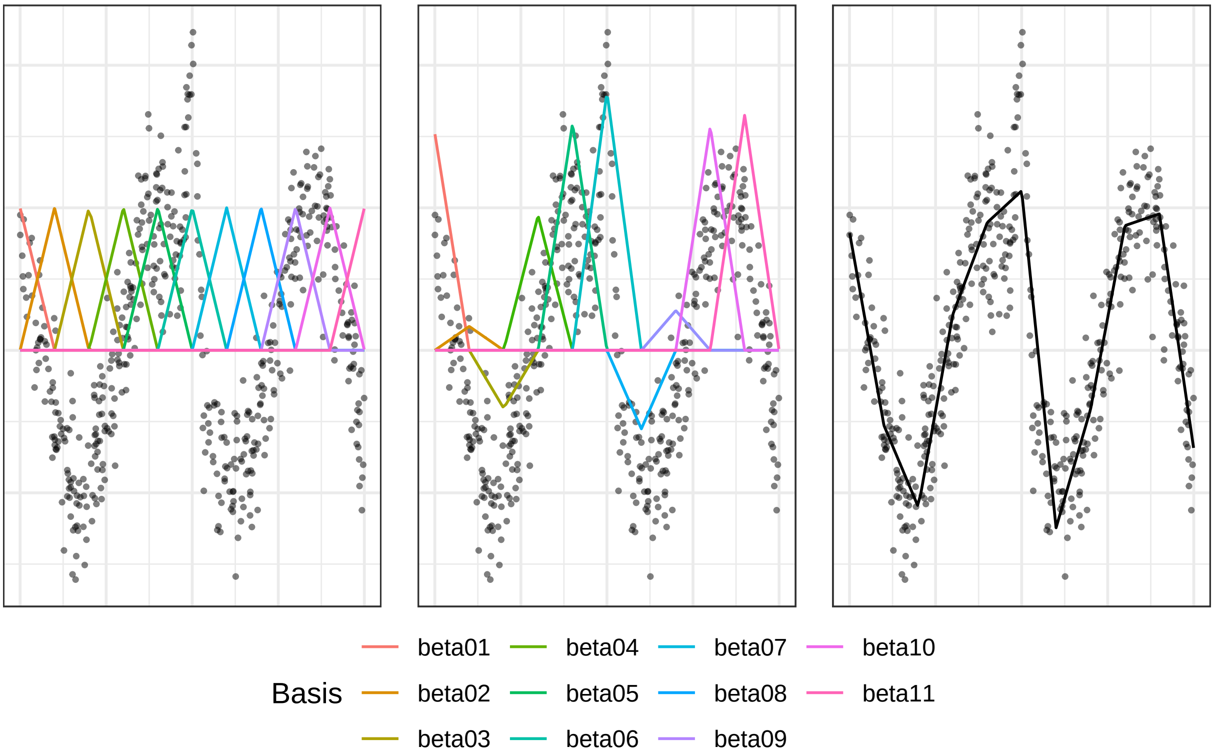

Figure 1. The working of a GAM spline, with simulated x and y data: the left graph shows a linear regression fitted between knots, the centre graph shows each basis multiplied by the corresponding piecemeal linear regression coefficient, and the right graph shows the sum of the basis functions, adapted from Clark (2017).

1990s. It has 159 observations and 6 variables of interest, median income (`MedInc`), % of the population that is rural (`PctRural`), % with degrees (`PctBach`), % elderly (`PctEld`), % foreign born (`PctFB`) and % black (`PctBlack`). The analyses below construct a series of GP-derived GAM spline models of Median Income, each with different knot and smoothing parameter specifications, generating local coefficient estimates, which are mapped.

3.2 GAM with GP splines

GAM GPs with a GP smooth, parameterised with observation location was constructed. The splines were specified with 7 knots to ensure sufficient degrees of freedom and a smoothing parameter was optimised by the spline function in the `mgcv` R package (S. Wood and Wood 2015). This controls the degree of smoothing and provides an indication of 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 (Table 1) along with the spatially smoothed terms (Table 2). 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 the predictors of Median Income arising from the GAM splines as in Figure 2. The trends in smoothed coefficients broadly show East-West gradients for the Intercept, % with Degree and % Elderly, and North-South ones for % Black. Notice also the smoothing parameters: these vary from 10^{-6} to 10^{+2}, giving an indication of the process spatial heterogeneity in a similar way to the bandwidths in GWR models.

Table 1. The coefficients 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

3.3 Tuning the GAM with splines

The results in Figure 2 are GWR like: they show the spatial distribution of the coefficients and the smoothing parameters indicate the scale of the relationship, although in less intuitive way than the GWR bandwidths. It is possible to tune the smoothing parameters of GAM model using a training and validation split of the data. The aim here is to train a series of models specified with different smoothing parameter values and to evaluate these against the validation data.

Table 2. The coefficients of the GAM smoothed terms

	Effective df	Ref. df	F	p-value	Smoothing Parameter
s(X,Y):Intercept	15.538	18.932	2.093	0.008	1.91e-06
s(X,Y):PctRural	5.616	6.044	2.081	0.075	1.93e-03
s(X,Y):PctBach	2.517	2.529	2.78	0.031	1.78e-01
s(X,Y):PctEld	2.5	2.5	4.812	0.015	5.40e+02
s(X,Y):PctFB	2.5	2.5	0.797	0.614	7.83e+01
s(X,Y):PctBlack	2.5	2.5	11.675	0	5.33e+02

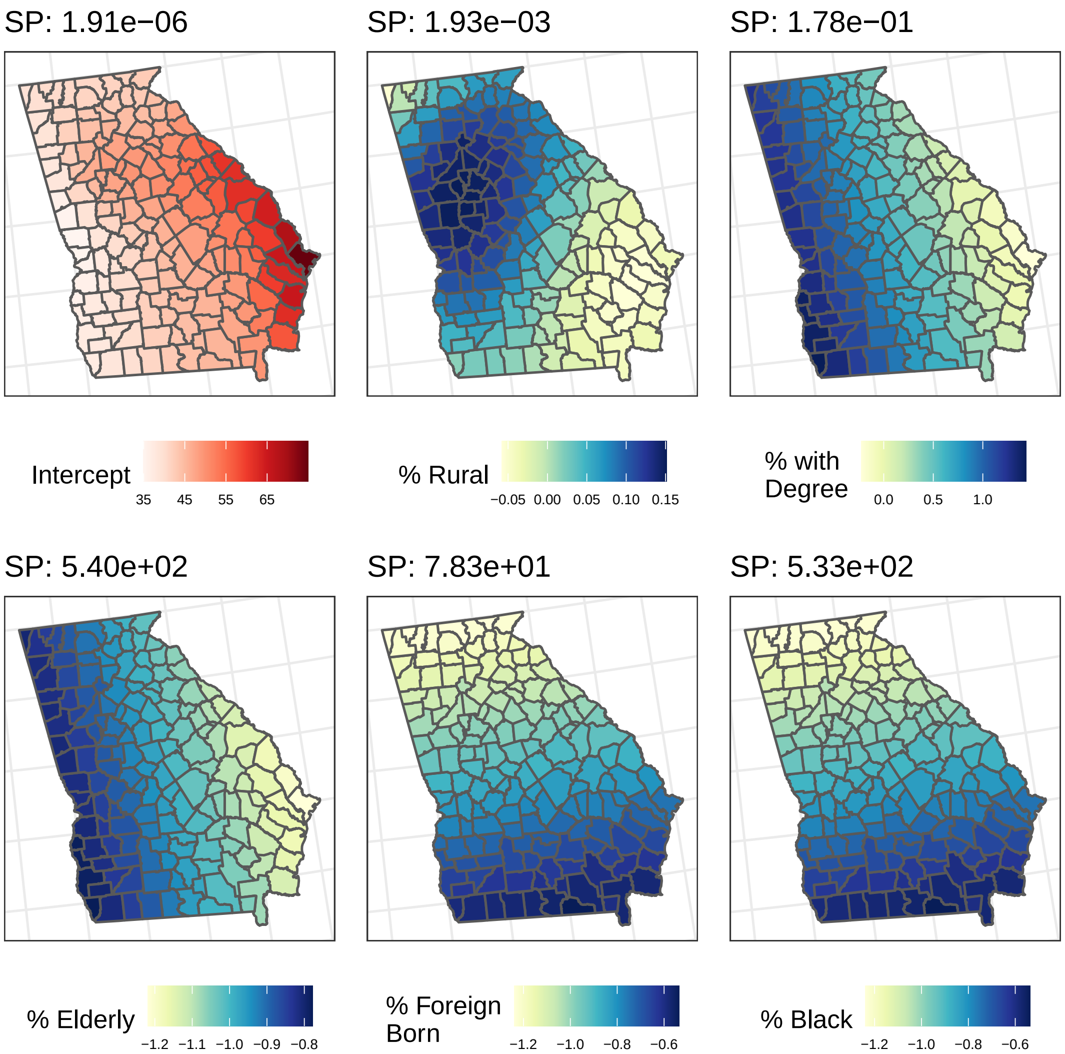

Figure 2. The local coefficient estimates from the GAM spline smoothed terms, with the smoothing parameter values (SP).

Here the Georgia data were split 80:20 into training and validation subsets using a bootstrap resampling approach to ensure equal distributions of the target variable (median Income) in both subsets. Then the a tuning grid of 531,441 different combinations of smoothing parameters was constructed. This was comprised of each of combination of 9 smoothing values in a log scale from 10^{-6} to 10^{+2} (i.e 0.000001 to 100) for the Intercept and 5 covariates. For each combination a GAM spline model was constructed and then used to predict the withheld values of Median Income in the validation subset. The combination with the best model fit (in this case R^2) was then used to fit a final model of the full dataset.

The fixed offset terms and the spatially smoothed terms are shown Tables 3 and 4, with the latter also showing the optimised smoothing parameters. The significant predictors for the fixed terms are the same except for `PctBlack` which is no longer significant. For the smoothed terms (Tables 4) the intercept is no longer significant and the effective degrees of freedom for this are much lower, but otherwise the numerical summaries are broadly the same.

The results of passing the best combination of tuning parameters for each covariate into a final model are shown in Figure 3. These show different patterns of spatial heterogeneity when compared with Figure 2:

- the Intercept is weaker and more localised to the South East corner (higher values);
- % Rural surface has a distinct gradient to the North West rather than a local focus around Atlanta;
- % Degree is similar in both models;
- % Elderly is localised to the South West Corner;
- % Foreign born and % Black have been pulled to the South East corner (high values).

Table 3. The coefficients of the GAM fixed terms

	Estimate	Std. Error	t-value	p-value
Intercept	47.036	4.224	11.136	0.000
PctRural	0.027	0.015	1.758	0.081
PctBach	0.409	0.272	1.504	0.135
PctEld	-0.614	0.119	-5.150	0.000
PctFB	-0.680	0.318	-2.137	0.034
PctBlack	0.865	1.026	0.844	0.400

4 Discussion

Geographically varying coefficient models are useful because they explicitly accommodate process spatial heterogeneity, where statistical relationships, as expressed trough coefficient estimates, may change with location. They provide an explicit representation of process spatial heterogeneity are are easily mapped.

Here an initial model was generated (Figure 2) and then the model was tuned in order to identify optimal smoothing parameters (here minimising a measure of model fit) and to explore how the scale of processes varies though the GAP GP smoothing parameter (Figure 3). The maps in Figures 2 and 3 show distinct differences, and provide an indication of the different underlying model semantics: very different process understandings would be are generated from these 2 models in terms of understanding how and where processes vary spatially. They indicate that the concept of 'process spatial heterogeneity', frequently referred to in the spatial analysis literature, may itself need to be refined to reflect these aspects of model semantics: process spatial heterogeneity with respect to fitted GAM GPs' for example. It also suggest the need examine the underlying semantics of the resultant models in relation to the scale of process suggested by the smoothing parameters.

Further work will explore a larger search space of smoothing operators, will apply GAM GP splines to simulated data with known spatial properties and may extend the splines to the temporal domain.

Acknowledgements. This work has received funding from the British Academy under the Urban Infrastructures of Well-Being programme [grant number UWB190190]. The data and code will be made available if issues around respondent privacy can be overcome.

References

Brunsdon, C. and Comber, A.: Opening practice: supporting reproducibility and critical spatial data science, Journal of Geographical Systems, 23, 477–496, 2021.

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E.: Geographically weighted regression: a method for exploring spatial nonstationarity, Geographical Analysis, 28, 281–298, 1996.

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., et al.: A route map for successful applications of Geographically Weighted Regression, Geographical Analysis, 0, 1–24, 2022.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. D.: Regression models, in: Regression, pp. 23–84, Springer, 2021.

Fotheringham, A. S., Yang, W., and Kang, W.: Multiscale geographically weighted regression (MGWR), Annals of the American Association of Geographers, 107, 1247–1265, 2017.

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

Goodchild, M. F.: The validity and usefulness of laws in geographic information science and geography, Annals of the Association of American Geographers, 94, 300–303, 2004.

Openshaw, S.: Developing GIS-relevant zone-based spatial analysis methods, Spatial analysis: modelling in a GIS environment, pp. 55–73, 1996.

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., and Fotheringham, A. S.: mgwr: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale, ISPRS International Journal of Geo-Information, 8, 269, 2019.

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

Tobler, W. R.: A computer movie simulating urban growth in the Detroit region, Economic geography, 46, 234–240, 1970.

Wolf, L. J., Oshan, T. M., and Fotheringham, A. S.: Single and multiscale models of process spatial heterogeneity, Geographical Analysis, 50, 223–246, 2018.

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

Yang, W.: An extension of geographically weighted regression with flexible bandwidths, Ph.D. thesis, University of St Andrews, 2014.

Table 4. The coefficients of the GAM spline smoothed terms

	Effective df	Ref. df	F	p-value	Smoothing Parameter
s(X,Y):Intercept	2	2	1.862	0.159	1.00e+02
s(X,Y):PctRural	2.507	2.514	1.115	0.2	1.00e+02
s(X,Y):PctBach	2.625	2.733	2.775	0.03	1.00e-01
s(X,Y):PctEld	2.5	2.5	9.157	0	1.00e+02
s(X,Y):PctFB	2.5	2.5	1.831	0.235	1.00e+02
s(X,Y):PctBlack	5.839	6.352	4.643	0.005	1.00e-03

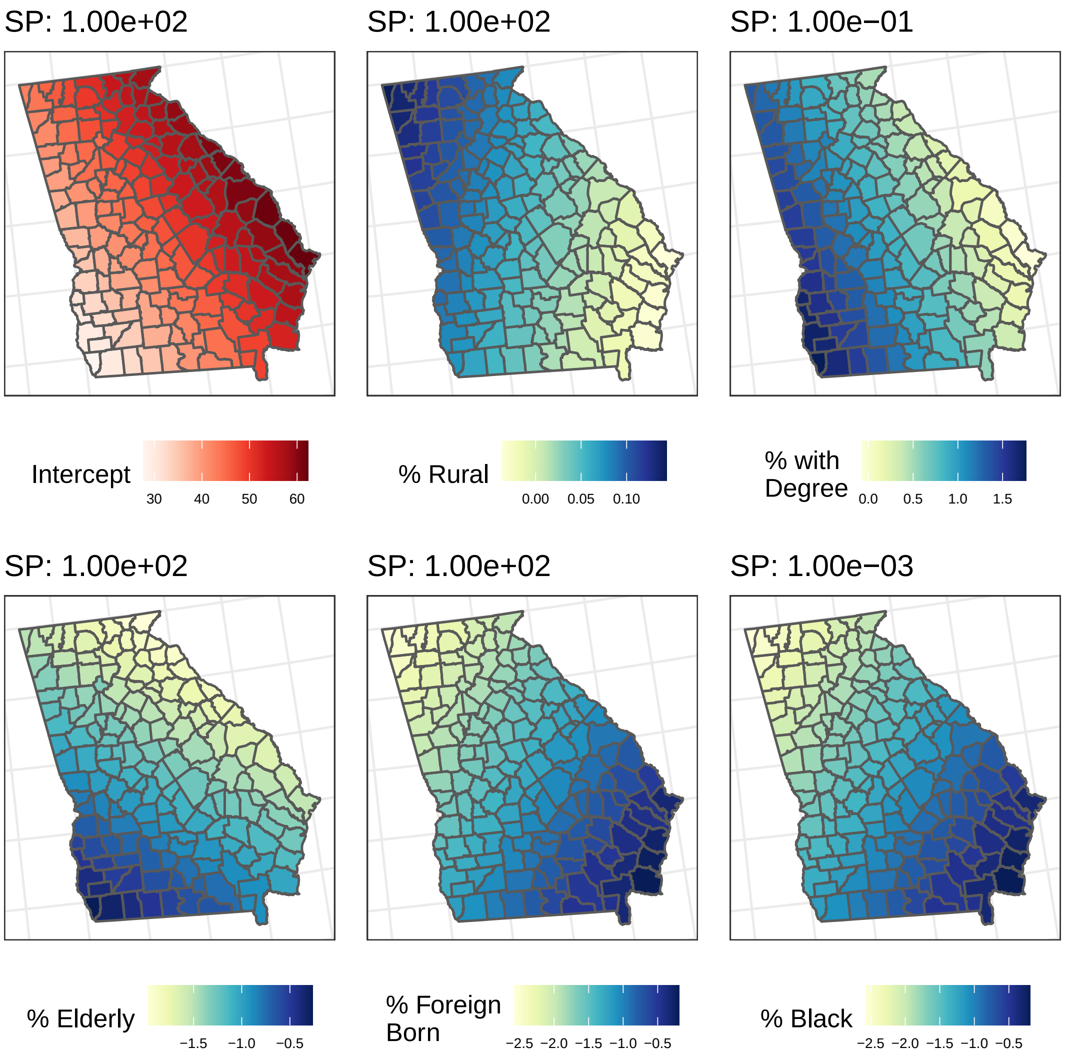

Figure 3. The local coefficient estimates from the tunes GAM spline smoothed terms, with the optimised smoothing parameter values (SP).