Article:
Comber, A., Harris, P. and Brunsdon, C. (Accepted: 2026) Multiscale space-time varying coefficient modeling using Generalized Additive Models: determining model form. International Journal of Geographical Information Science. ISSN: 1365-8816 (In Press)

ARTICLE TEMPLATE

Multiscale space-time varying coefficient modeling using Generalized Additive Models: determining model form

Alexis Comber[a], Paul Harris[b] and Chris Brunsdon[c]

[a]School of Geography, University of Leeds, UK; [b]Sustainable Agriculture Sciences, Rothamsted Research, North Wyke, UK; [c]National Centre for Geocomputation, Maynooth University, Ireland.

ARTICLE HISTORY
Compiled May 6, 2026

ABSTRACT
This paper describes an approach to multiscale space-time varying coefficient modelling using Generalized Additive Models (GAMs). This builds on GAMs that have been previously used to generate spatially varying coefficients, but now extended to the temporal domain. However, for this space-time extension, a key consideration concerns model specification - i.e. how best to specify space and time in the GAM smooths. This depends on the nature of the space-time dependencies in the predictor-to-response variable relationships for a given dataset, where a key issue is how to relax assumptions about the presence of space-time dependencies that are found. For this study, multiple competing GAMs are created, with their space-time interactions specified in different ways for each predictor variable. A final GAM is found by the lowest AIC. The proposed approach is applied to simulated data and an empirical case study, where in both cases the best fitting GAM was found to out-perform Multiscale Geographically and Temporally Weighted Regression, a popular alternative. Model specification / selection for GAMs undertaken in this way provides insights into the nature of the space-time dependencies present in the data's relationships, including the scale of these dependencies. The overall approach is novel and several areas of further work are identified.

KEYWORDS
stgam, Space-time relationships, Coefficient non-stationarity, Process heterogeneity, Varying parameter models

1. Introduction

Data are increasingly spatio-temporal with their collection typically automated and a key objective in data analysis, particularly in the geographical sciences, is to construct statistical models that support process understanding. These models allow inferences about space-time processes to be made, such as those about the strength of data relationships, in addition to simple prediction, forecasting or classification. In this respect and extending standard regression models where relationships are assumed to be spatially invariant, there is a rich history of adapting regressions for capturing spatially varying relationships (Casetti 1972, Jones 1991, Brunsdon *et al.* 1996, McMillen 1996, Gelfand *et al.* 2003, Griffith 2008, Murakami 2021, Comber *et al.* 2024a). These seek to determine the nature of any spatial heterogeneity in relationships captured by the

CONTACT Alexis Comber. Email: a.comber@leeds.ac.uk

data, and to estimate spatially varying coefficients to describe where this is found. Analogously, the spatial case has been extended to the space-time case, with early examples including Pace *et al.* (1998), Pace *et al.* (2000), Elhorst (2003), Gelfand *et al.* (2004), Di Giacinto (2006) and Crespo *et al.* (2007).

Space-time regressions themselves can be categorised into four broad sets (Griffith 2010): (i) space-time autoregressive integrated moving average (STARIMA) models, (ii) spatial panel regressions (SPR), (iii) geostatistical space-time models and (iv) models such as eigenvector space-time filtering (ESTF) and geographically and temporally weighted regression (GTWR). The first three have a focus on modelling heterogeneity via autocorrelation effects but assume stationary (fixed) data relationships, whilst only the last group is concerned with modelling heterogeneity in terms of non-stationary relationships. To achieve this, the ESTF models generate distributions of latent spatial and temporal autocorrelation in the variables (Patuelli *et al.* 2011, Chun 2014) where recent aggregation developments in the spatial only case (Murakami *et al.* 2024) have potential to be extended to the temporal domain. For GTWR, moving window / weighted kernel functions are used (Huang *et al.* 2010) and have recently been extended to capture the individual scale of each space-time relationship between target and predictor variables, to provide multiscale GTWR forms (Liu *et al.* 2017, Wu *et al.* 2019). All forms, across all four categories, face similar challenges, especially in terms of finding a suitable space-time distance or metric. For example, prediction inference in geostatistical approaches is often hindered by the difficulty in determining joint or separable space-time variograms (Gneiting *et al.* 2006, Ma 2008). A further challenge, one that is often overlooked, is that all approaches make strong assumptions about the presence and nature of the space-time interactions or relationships in the data, regardless of whether they should be fixed or varying.

This paper seeks to address both challenges but has a focus on the latter via a Generalized Additive Model (GAM) based space-time varying coefficient approach. Although many existing space-time GAMs undertake model selection and penalization, they do not examine the nature of the space-time dependencies present in the data relationships and adjust how individual predictor variables are specified in the model. For example, Dambon *et al.* (2022) consider the interactions of predictors with the target variable over space but with a focus on selection (penalization) and inference for target variable prediction and not on inference for data relationships and how each individual relationship is handled. By using GAMs (Hastie and Tibshirani 1987, 1990) this paper extends and revises the outline approach described in Comber *et al.* (2023b).

For model evaluation, simulated space-time coefficient surfaces with known space-time dependencies are created, and these true coefficients are used to generate 100 simulated datasets. The synthetic datasets are used to illustrate the process of determining the GAM specification. This is done by creating all possible space-time models in which each predictor variable is specified in different ways in GAM smooths for each simulated dataset, evaluating each model via their AIC value and selecting the best one. The best model is compared with a Multiscale Geographically and Temporally Weighted Regression (MGTWR) (Zhang *et al.* 2021). This approach of undertaking investigations of the nature of space-time variability in the predictor-to-response relationship emphasises the importance of user investigations of space-time interactions in order construct informed models (Comber and Wulder 2019). The approach is demonstrated empirically using data for the dry rain forest in the Chaco region of South America. The analysis uses the `stgam` (Comber *et al.* 2026b) and `mgcv` (Wood 2021) R packages.

2. Background: varying coefficient models with GAMs

GAMs (Hastie and Tibshirani 1987, 1990) extend linear models by using smooth functions to capture non-linear relationships in data. The type of smooth used can vary according to the problem or dataset (Hastie and Tibshirani 1990). Smooths estimate a continuous curve through the observations based on assumptions that values close to each other in attribute space tend to be related, and the overall fitted curve captures a broader correlated process. Technically, smooths are constructed from a set of basis functions of the predictor variables, which can be one- or multi-dimensional. This allows the GAM to represent complex relationships, with the amount of non-linearity in each smooth controlled by a smoothing parameter that manages the trade-off between flexibility and over-fitting.

GAMs are applicable to a wide range of response types (Wood 2017, Fahrmeir *et al.* 2021), can capture patterns in complex systems effectively (Hastie and Tibshirani 1990, Wood 2017), and in many cases perform comparably to or better than machine-learning (ML) methods (Friedman 2001, Wood 2017). In terms of interpretability, GAMs provide a modelling framework for understanding processes rather than just making predictions, as in many ML models. They have been described as providing the explanatory clarity of a *glass-box* model, producing results that are as technically robust as *black-box* ML models, while containing structures that are more readily understood and transparent (Zschech *et al.* 2022), thereby supporting critical perspectives to data science, and in context of this study, spatial data science (Brunsdon and Comber 2021).

Importantly, if GAM smooths are parameterised with observation location then they can be used to estimate spatially varying coefficients (Fan and Huang 2022, Comber *et al.* 2024a,b), and if they also include observation time then they can be used to estimate space-time varying coefficients (Comber *et al.* 2023b):

The model can be described formally by letting y_i denote the response observed at spatial location (u_i, v_i) and time t_i, with predictors $x_{i1}, x_{i2}, \ldots, x_{ip}$. The general space-time varying coefficient model is:

$$y_i = \sum_{j=1}^{p} x_{ij}\, \beta_j(u_i, v_i, t_i) + \varepsilon_i \tag{1}$$

where $x_{i1}, x_{i2}, \ldots, x_{ip}$ are the predictor variables, $\beta_j(u, v, t)$ is the coefficient estimate for predictor x_j varying over space and time, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the error term.

The use of GAMs to capture space-time varying processes builds on a series of key theoretical developments. Early approaches to space-time varying coefficient GAMs were introduced by Hastie and Tibshirani (1993). These allowed coefficient estimates to vary smoothly over another variable, such as the location or time of observation and provided a critical foundation that was further developed by modelling non-linear relationships via smooth functions (Hastie and Tibshirani 1990). Smooths (or splines - the terms are often used interchangeably) fit a smooth curve through response to predictor data points for which there may not be an obvious single function that describes their relationship. Smooths generate multiple fits which are combined, and are thus able to describe the relationship between variables better than a single function. Further work using GAMs to capture varying coefficients in a space-time context include Rigby and Stasinopoulos (2005); Stasinopoulos and Rigby (2008); Stasinopoulos *et al.* (2017); Umlauf *et al.* (2018) in which 'varying' refers to the distribution and form of the response variable and its properties.

Technically, the GAMs described are still in attribute space, where a subset of the predictors provide spatial or temporal information. However, other GAM-based approaches have explicitly considered how relationships as captured through coefficient estimates vary with observation location and / or time (i.e. within geographical and / or temporal space). It is this class of GAMs that are presented in this study. A key development in this respect is the use of Tensor Product (TP) smooths (Wood 2006) as these allow variables on different scales or with anisotropic structures (such as location and time) to be combined within a single smooth (i.e. they address the difficulty in finding a space-time metric). This is done by including separate smoothing penalties across different dimensions, allowing complex interactions, for example between spatial and temporal components, to be modelled. For example, Feng (2022) describes a space-time varying coefficient GAM in which local socio-economic factors were used to construct a two-dimensional spline smoother for COVID-19 mortality risk. Changes in the spatial pattern of relationships were evaluated over time through a spatio-temporal tensor.

Developments in the spatial only case are also noteworthy. Kim and Wang (2021) proposed a generalized spline-based approach for spatially varying coefficient modelling. Their model used approximate bivariate smooths to incorporate spatial information by constructing spline basis functions over a spatial triangulation, to handle irregularly shaped spatial domains with complex boundaries. Dambon *et al.* (2021) and Dambon *et al.* (2022) describe a spatially varying coefficient GAM based on Gaussian Processes (GPs) with a focus on variable selection, undertaken using a penalized maximum likelihood approach. Similarly, GAMs with GP smooths have been proposed as an approach to estimate spatially varying coefficients (Comber *et al.* 2024a,b) under assumptions of classic distance decay.

3. Methods

In overview, this study illustrates an approach for determining how to specify GAMs with smooths for multiscale space-time varying coefficient modelling. It seeks to determine the GAM form that best captures the space-time dependencies in the data relationships, and thus how best to specify the GAM's space-time smooths. This is illustrated using simulated data before being applied to an empirical case study. In both cases, results from a MGTWR fit, that similarly estimates multiscale space-time varying coefficients, are presented for context and discussion. All code and data used to generate the results are provided.

3.1. *Simulated data*

Simulated coefficients were created in a similar way to Comber *et al.* (2024a), but extended to have known space-time dependencies. Here four true coefficient surfaces β_0, β_1, β_2 and β_3 were created for a 15×15 regular square grid (u, v), with 15 time periods (t). These dimensions were chosen to manage the computational costs associated with the iterative back-fitting procedures required for the MGTWR model. The simulated (true) coefficients surface were defined as follows:

$$\beta_0(u,v,t) = 3 \tag{2}$$

where u and v are observation spatial coordinates, and t is observation time.

Define the normalized coordinates as $u_n = \frac{u}{15}$, $v_n = \frac{v}{15}$ and $t_n = \frac{t}{15}$, and let $\varepsilon_k \sim \mathcal{N}(0, 0.03^2)$ for $k = 1, 2, 3$.

$$\beta_1(u,v,t) = 1 + (u_n + v_n) \cdot \sin\left(2\pi t_n + \pi/4\right) + \varepsilon_1 \tag{3}$$

$$\beta_2(u,v,t) = 1 + \exp\left[-8 \cdot \left((u_n - c_u(t_n))^2 + (v_n - c_v(t_n))^2\right)\right] \cdot \Big(1 + 0.6, \sin(2\pi t_n)\Big) + \varepsilon_2 \tag{4}$$

where $c_u(t_n) = 0.3 + 0.4t_n$ and $c_v(t_n) = 0.7 - 0.3t_n$.

$$\beta_3(u,v,t) = 1 + 0.6\cos(2\pi u_n)\sin(2\pi v_n) + 0.5\sin(2\pi t_n) + \varepsilon_3 \tag{5}$$

A small amount of Gaussian noise was added to each of the true coefficient surfaces drawn from $\mathcal{N}(0, 0.03^2)$ and then they were re-scaled in the range 0 to 2.

The true coefficients have the following properties:

- β_0 is a constant intercept term.
- β_1 is a simple space–time process, with the spatial component a smooth linear gradient across the grid and the temporal component a sinusoidal oscillation, which scales the gradient over time. Although the space and time effects (a global gradient with temporal oscillation) interact, they are also separable.
- β_2 is a non-separable space–time process in which the spatial pattern is a localized Gaussian hotspot, whose centre and amplitude change over time. Here there are important space–time interactions.
- β_3 is a separable space-time process in which the shape of the spatial surface remains fixed and oscillates uniformly over time. Here the space–time coefficient surface is clearly additively separable.

The true regression coefficient surfaces used in the simulation are presented in Figures 1, 2, and 3 and the predictor variables x_1, x_2, and x_3 were independently drawn from a standard normal distribution, $N(0,1)$, while the error term ϵ was sampled from the same distribution and scaled by 0.25. A total of 100 datasets were generated. For each dataset, the response variable y was computed directly using the simulated coefficient surfaces, the corresponding predictor values, and the generated error term.

3.2. *Space-time GAM specification*

The simulation surfaces in Figures 1, 2 and 3 indicate mild space-time interactions for β_1, strong space-time interactions for β_2 and separate space-time interactions for β_3. Thus, it would be reasonable to assume that the most appropriate GAM specification should take the following form:

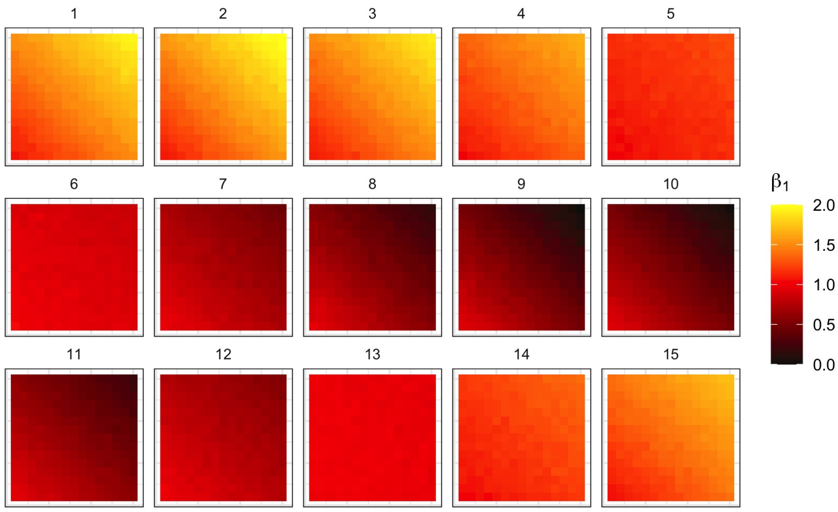

Figure 1. The simulated true regression coefficients over 15 time periods: β_1, with simple and seperable space-time interaction.

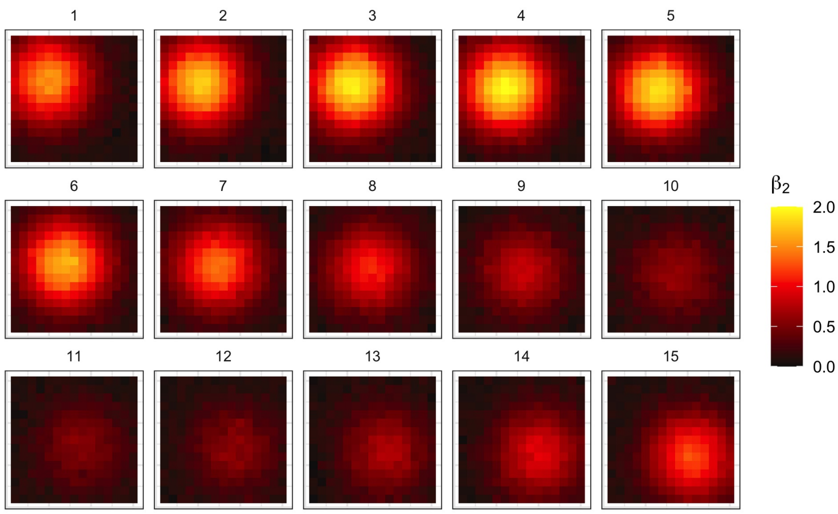

Figure 2. The simulated true regression coefficients over 15 time periods: β_2, with moderate (moving hotspot) space-time interaction.

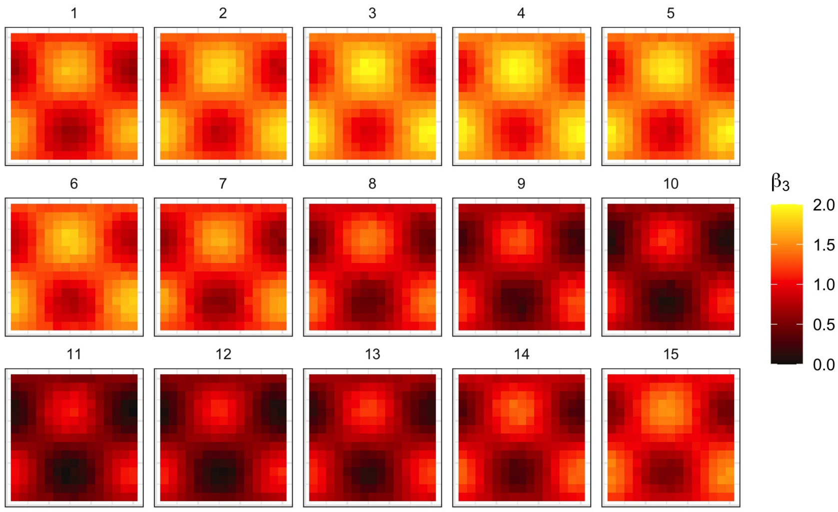

Figure 3. The simulated true regression coefficients over 15 time periods: β_3, with clearly sperable space-time interaction.

$$\begin{aligned}\mathrm{y}_i = \beta_0 \cdot \mathrm{Intercept}_i + \\ f_1(u_i, v_i, t_i) \cdot \mathrm{x1}_i + \\ f_2(u_i, v_i, t_t) \cdot \mathrm{x2}_i + \\ f_3(u_i, v_i) \cdot \mathrm{x3}_i + f_4(t_i) \cdot \mathrm{x3}_i + \epsilon_i\end{aligned} \tag{6}$$

where y_i is the response variable, β_0 is the global fixed intercept, combined with space-time TP smooths in $f_1(u, v, t)$, and $f_2(u, v, t)$ that are multiplied by x_1 and x_2 respectively, and $f_3(u, v)$ and and $f_4(t)$ are multiplied by x_3, and ϵ_i is a Gaussian error term.

However, the model could quite reasonably be specified several different ways. The β surface for x_1, for example, has interacting but multiplicably separable space-time relationships: each location follows the same temporal pattern. So, should the smooth specified by $f_1(u, v, t)$ be separated into $f(u, v)$ and $f(t)$ in same way as for x_3? The user does not know the true coefficient surfaces and thus does not know the nature of the space-time interactions in the relationships captured by the data. But they must make a choice. A classic approach would be for the user to specify the most complex model that is judged valid for the use-case and the system being investigated. This would rely heavily on the user perception and user experience of the data generating process. The size of the effects of the different GAM components (e.g. the smooths) could be investigated and the GAM refined heuristically. This is done in Comber *et al.* (2026a) but the final GAM arrived at heuristically was not one that best captured the space-time relationships in the data.

In much geographical analysis the aim is to develop and support understanding about the process and relationships captured in the data and to determine the nature of any space-time interactions. This is what is supported through kernel bandwidth optimization in the classic Geographically Weighted Regression (GWR) model (Brunsdon *et al.* 1996): the bandwidth provides a measure of the scale of the spatially varying relationships between the predictor and response variables. Larger bandwidths indicate more smooth (stationary, fixed) relationships while smaller bandwidths indicate more locally varying (nonstationary, varying) ones (Comber *et al.* 2023a). In GWR and its variants, bandwidths and thus model form, are optimised through a measure of model fit (AIC or cross-validation score).

A similar approach can be taken with this study's GAMs to determine *how* each variable is specified within the model. For a space-time varying coefficient GAM, each predictor variable can be specified in one of 6 different ways, with the usual parametric form also specified to provide a multiplicative constant and potentially an offset term. The 6 ways for a given predictor are:

i) It is omitted.
ii) It is included as a parametric response.
iii) It is included in parametric form and in a spatial smooth.
iv) It is included in parametric form and in a temporal smooth.
v) It is included in parametric form and in separate spatial and temporal smooths.
vi) It is included in parametric form and in a combined space-time smooth.

The Intercept can be treated in a similar way but without being omitted. The result is that for any space-time regression problem with k predictor variables, there are 5×6^k potential models to evaluate, when the response to predictor relationships are allowed to vary in some manner.

For this study's GAMs, the single space and time smooths (options iii), iv) and v) above) are specified using Thin Plate Regression Splines (TPRSs) while TP smooths are specified for the combined space-time smooths (option vi) above). This is because TPRSs are isotropic: they incorporate a single basis and assume that all the variables included in the smooth are on the same scale. By contrast, TP smooths are anisotropic and relax this assumption. A TP smooth can therefore be used to incorporate variables on different scales, such as space and time measurements. They construct smooths using separate basis functions, which are then combined and require the dimensions of each basis to be specified. These are two dimensions for spatial location and one dimension for time. The smooths in the individual GAMs are optimised using restricted maximum likelihood estimation (REML). The REML approach provides stable estimates of the smoothing parameters and reduces over-fitting, thereby generating more robust estimates of the variance. The over-arching approach is to create multiple models each with each predictor variable specified in a different way, and to evaluate each model fit via its AIC score. This was undertaken using tools in the `stgam` R package (Comber *et al.* 2026b) which provides a wrapper to the `mgcv` GAM package (Wood 2021). Observe that the modelled relationships in the space-time varying coefficient GAMs are multiscale by design.

3.3. *MGTWR*

In GWR, a moving window or kernel-based approach is used to create a series of local regressions, whose results are mapped. The spatial kernel extracts data subsets that are weighted by their distance to the kernel centre - the location at which a local regression is fitted. The bandwidth or size of the kernel is optimised through some measure of (global) model fit (e.g. AIC). By extension, GTWR models (Huang *et al.* 2010) optimally apply a space-time kernel, with data subsets now weighted by their space-time distance to the kernel centre. This requires a space-time interaction term to be defined (Huang *et al.* (2010) refer to it a ρ parameter) or in some instances, heuristics are used when the search space is too large (Fotheringham *et al.* 2015). For multiscale GWR (MGWR) models (Yang 2014, Lu *et al.* 2017, Fotheringham *et al.* 2017) a kernel bandwidth is optimally found for each predictor variable and the intercept. This simultaneous optimisation problem is undertaken through iterative partial regression (back-fitting), in which bandwidths are estimated one at a time, conditional on the others. For MGTWR models (Wu *et al.* 2019, Zhang *et al.* 2021, Yu and Fotheringham 2025) the model fit challenges for GTWR and for MGWR are combined.

For this study, the approach taken for MGTWR was to create an array of space-time distance matrices and to pass these to a standard MGWR. The scaling parameter (ρ) was used to control the relationship between the distances in time and space. Here this was set to 1 as the data were simulated over hypothetical locations and time periods. Adaptive space–time Gaussian kernel bandwidths were optimally found for each predictor and intercept of the MGTWR model, with the potential to select up to 3375 (15^3) observations in each local regression when using a simulated dataset. The MGTWR models were constructed by passing the space-time distance matrices to the MGWR tools in the `GWmodel` R package (Lu *et al.* 2014).

3.4. *Model performance*

For the simulated data, 1080 GAMs were constructed for each of the 100 datasets with the intercept and each predictor specified in a different way in each GAM. From these, the GAM with the lowest AIC was selected for each of the 100 datasets. A MGTWR was fitted to each of the same 100 datasets, using the space-time distance array with bandwidths also found via AIC. The space-time varying coefficient estimates were extracted from the 100 GAMs and the 100 MGTWR models, as were the predicted responses, $\hat{y}$. The estimated coefficients were compared with the true coefficients using standard measures of model accuracy and fit (R^2, RMSE and MAE). The predicted $\hat{y}$ was also compared with the true y using the same fit diagnostics. Note that it is inappropriate to summarise AIC values across models.

3.5. *Space-time dependencies in GAM smooths*

It is also possible to investigate the space-time dependencies in the GAM space-time smooths. Each TP term typically introduces smoothing parameters for space and for time. These indicate the degree of complexity and flexibility of the smooth functions used to fit the model. The smoothing parameter determines the trade-off between fidelity to the data (i.e., closely following the observed data points), and the smoothness of the fitted function (i.e., avoiding overfitting or excessive ‘wiggliness’). In this way the smoothing parameters govern how variable (‘wiggly’) the estimated function is, with smaller values indicating less penalization and a model that captures finer detail, and larger values indicative of more smoothing and a model that ignores small-scale variation. By examining model smoothing parameters, it is possible to determine the nature of the structure found by the model. For example, a small temporal smoothing parameter indicates that the model has determined strong temporal dependence and has fitted a detailed time trend. Similarly, a small spatial smoothing parameter indicates the model has found strong spatial dependence (fine-grained spatial variation).

While this study's GAM smoothing parameters provide an indication of data structure and are indicative of space-time dependencies in the response-to-predictor relationships, there is no direct way to estimate distances from smoothing parameter values that are analogous to the measures of space-time heterogeneity in the response-to-predictor variable relationships provided by the MGTWR bandwidths. However, it is possible to estimate how far changes in the predictor variables propagate (spatially and / or temporally) before smoothing causes them to decay - i.e. the range of the effect. This can be done by varying each predictor while holding the others constant and measuring the correlation decay of the response over a space-time prediction grid. Effectively this uses the concept of decay or response change over a distance, to approximate declining autocorrelation in an interpretable manner.

Estimating the spatial and temporal length scales from GAM smooths can be done in the following way:

- Let the estimated smooth function be $\hat{f}(\mathbf{s}, t)$, where $\mathbf{s} = (u, v)$ denotes spatial coordinates and t denotes time.
- Let $\mathcal{P} = \{(\mathbf{s}_i, t_i, \hat{f}_i)\}_{i=1}^{N}$ be a set of predicted values from a fitted GAM surface, evaluated on a regular grid.
- Define:

$$\begin{aligned} \delta_{ij}^{\text{spatial}} &= \|\mathbf{s}_i - \mathbf{s}_j\| \\ \delta_{ij}^{\text{temporal}} &= |t_i - t_j| \\ \Delta_{ij} &= |\hat{f}_i - \hat{f}_j| \\ \Delta_{\max} &= \max_{i,j} \Delta_{ij} \end{aligned} \tag{7}$$

Given a threshold parameter $\tau \in (0, 1)$, the spatial and temporal length scales are estimated as the average distance (in space or time) over which the smooth changes by at least a proportion τ of its maximum variation:

$$\begin{aligned} \text{Spatial Length Scale} &= \frac{1}{N_{\text{sp}}} \sum_{i<j} \delta_{ij}^{\text{spatial}} \cdot \mathbb{I}(\Delta_{ij} \geq \tau \Delta_{\max}) \\ \text{Temporal Length Scale} &= \frac{1}{N_{\text{tmp}}} \sum_{i<j} \delta_{ij}^{\text{temporal}} \cdot \mathbb{I}(\Delta_{ij} \geq \tau \Delta_{\max}) \end{aligned} \tag{8}$$

where $\mathbb{I}(\cdot)$ is the indicator function and N_{sp}, N_{tmp} are the number of spatial or temporal pairs satisfying a threshold condition (here 0.5), respectively. Here N_{sp} and N_{tmp} denote the number of distinct point pairs for which the smooth difference exceeds the threshold proportion $\tau\Delta_{\max}$. They act as normalising constants so that the estimated length scales represent mean spatial or temporal distances among pairs exhibiting substantial change.

These length scales provide a summary of the effective range of spatial and temporal structural variation in the smooth components of the fitted GAM. They estimate length scales using the concept of autocorrelation decay or response change over a fixed distance, which is approximate but interpretable. The smooth's length scale is defined as the mean distance required for the estimated smooth to change by some proportion of its total range. Specifying this as 0.5 is analogous to the determination of practical ranges used by certain variogram model functions in Geostatistics and provides a robust, scale-invariant summary of the spatial (or temporal) rate of change that is not overly sensitive to local noise or extreme contrasts.

The approach is related to variograms, as both aim to quantify spatial or temporal dependence and provide a measure of dependence length but they differ in the way they are computed. A variogram quantifies the dependence of a variable by determining the average squared difference in values as a function of the distance in space or time between observations. Spatial smoothing by splines and by kriging are also strongly related, where Hutchinson and Gessler (1994) demonstrate that kriging and TPRs achieve comparable interpolation accuracy especially over larger at observations spacings. Both approaches seek to generate smooth surfaces while minimizing prediction error but differ in how assumptions about spatial variability are incorporated (Dubrule 1984).

The approach taken here estimates how the smooth effects vary with space and with time by examining differences in predicted smooth values over distances, with the length scale the average distance between observations and where the difference in the smooth effect exceeds a specified threshold (taken as a proportion of the maximum difference, in this case, 0.5). As such this approach is analogous to a variogram: rather than computing empirical variance of the data directly, it computes the smooth function's gradient across space or time.

4. Results

4.1. *Model performance*

Model performance diagnostics arising from the GAMs and MGTWR models applied to the 100 simulated datasets are summarised in Figure 4, where the best performing model has a lower MAE, a lower RMSE and a higher R^2. Examining each row in Figure 4 indicates that:

- MGTWR models are better at predicting the response variable y.
- There is little difference between GAMs and MGTWR models at estimating β_0.
- GAMs are better at estimating β_1.
- GAMs are better at estimating β_2.
- GAMs are better at estimating β_3.

Overall, there are marginal differences in model performance resulting from the two modelling frameworks. Of note are the better performance of the GAMs in estimating the true coefficients (β_1, β_2 and β_3), and the better performance in the prediction of y by the MGTWR models. In many situations, such bias-variance tradeoffs result in relatively accurate coefficient estimation, but relatively poor model prediction, as is the case with the GAMs here. This arises when a model has lower bias in coefficients but higher variance in predictions (Geman *et al.* 1992), resulting in better β identification but unstable $\hat{y}$, such that small increases in prediction variance inflate diagnostics such as RMSE, MAE and R^2. Here, the superior recovery of the true coefficients by the GAMs indicates stronger structural estimation of the data-generating process, whereas the better prediction of true y by the MGTWR models indicates superior predictive accuracy.

This outcome reflects the fact that coefficient recovery and predictive performance

Table 1. Summary of the GAM smooths used for the GAMs selected for each of the 100 simulated datasets.

	Intercept	x_1	x_2	x_3
Omitted	0	0	0	0
Parametric response	52	0	0	0
Spatial smooth	16	0	0	0
Temporal smooth	20	0	0	0
Separate space and time smooths	3	0	0	100
Combined space-time smooth	9	100	100	0

are two different things. Coefficient recovery (i.e. how close the estimated coefficients $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are to the true coefficients) relates to model estimation accuracy (i.e. bias-variance and mean standard error of coefficients) while predictive performance (i.e. how close $\hat{y}$ is to y) relates to model fit accuracy. Strong coefficient estimation may result in poor predictions for the following reasons. The expected prediction error decomposes into $bias^2 + variance + noise$, and thus it is possible for a model to estimate coefficients accurately yet produce less stable or higher-variance predictions (Geman *et al.* 1992). And, as prediction error necessarily includes irreducible stochastic noise, minimising coefficient error does not minimise predictive loss (Hastie *et al.* 2009). This reflects the broader methodological separation of explanation (accurate estimation of relationships) and prediction (accurate forecasting of observed outcomes), which are intrinsically related but fundamentally different statistical goals (Shmueli 2010). Consequently, the better approximation of the true coefficient surfaces by the GAMs but poorer model fits is entirely consistent, particularly in finite samples.

4.2. *Model form*

The study's GAM and MGTWR model construct different forms of space-time models. For the GAM this is through the specification of space-time smooths while for MGTWR this is through the space-time kernel specified (Gaussian) and its bandwidths (which also controls model smoothing). Recall that for each of the 100 simulated datasets, 1080 GAMs were evaluated and the one with the lowest AIC was selected. Table 1 summarises the nature of the GAM smooths used for the 100 selected GAMs. This indicates that the predictor variables x_1 and x_2 were always specified with combined space-time TP smooths, suggesting strong space-time interactions, while x_3 was always specified with separate space and time smooths. However, the results for the intercept are more variable, with a parametric form specified in 52% of the GAMs as would be expected, but with a relatively large number of spatial or temporal smooths also specified. This confirms that the space-time GAM should be specified as in Equation 6: with a fixed parametric intercept to estimate β_0 coupled with combined and separate space-time smooths for x_1, x_2 and x_3 to estimate β_1, β_2 and β_3.

The adaptive MGTWR kernel bandwidths can be investigated in a similar way. These indicate how many 'nearby' observations (in space-time) are included in each local regression for the intercept and for each of the three predictor variables in turn. They provide a measure of the space-time heterogeneity in each of the response-to-predictor relationships in the simulated data. For reference, a global (stationary and fixed) relationship would include all observations (n = 3375); and noting that a space-time scaling factor of $\rho = 1$ was specified to define the relationship of time to space.

The estimated MGTWR bandwidths from the 100 MGTWR models are summarised in Table 2. On average, the intercept is estimated as a global fixed model parameter as

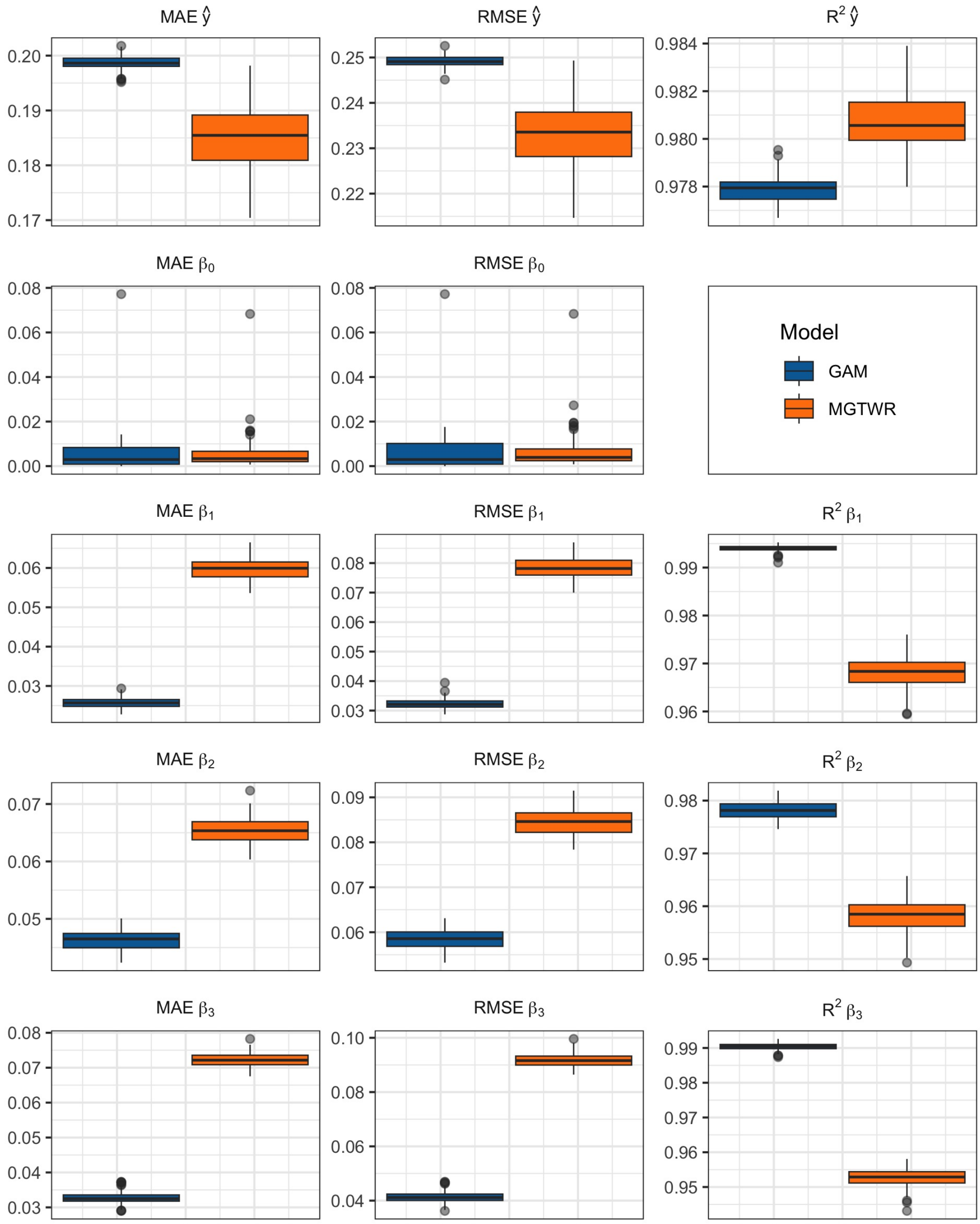

Figure 4. Distributions of model diagnostics (MAE, RMSE and R^2) for predicting the true response y and estimating the true coefficients (β_0, β_1, β_2,, β_1,) for the 100 GAMs and 100 MGTWR models.

Table 2. Summaries of the MGTWR adaptive bandwidths for each predictor variable, across 100 simulated datasets.

	Intercept	x_1	x_2	x_3
Min.	268.00	35	35	35
1st Q.	2123.75	58	73	36
Median	3355.00	63	73	41
3rd Q.	3372.00	83	83	43
Max	3372.00	94	87	83

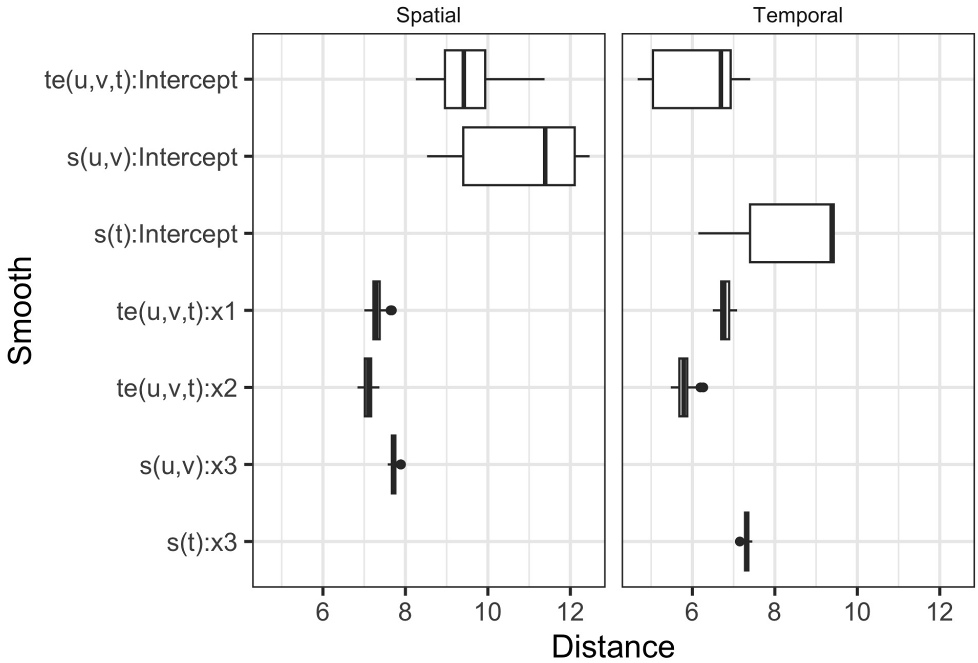

Figure 5. The distributions of the estimated GAM smooth spatial and temporal length ranges across 100 simulated datasets.

might be expected, coupled with highly localised space-time relationships for each of the predictors (x_1, x_2 and x_3). The most localised relationship is for x_3 (median bandwidth of 41 nearest neighbours) while the least localised is for x_2 (median bandwidth of 73 nearest neighbours).

4.3. *Space-time dependencies in GAM smooths*

The GAM smooth's spatial and temporal length ranges for the 100 GAMs of the simulated data are summarised in Figure 5. This reflects the heterogeneity of smooth form in Table 1 and suggests that the space-time TP smooths for x_1, x_2 and x_3 have similar spatial ranges (between 7 and 8 space-time units) but a wider range of temporal ones (from 5.5 to 7.5 units). The separate space and time smooths for x_3 behave similarly in space and in time. Overall, this suggests similar spatial dependencies in each of the response-to-predictor relationships but dissimilar temporal dependencies. The intercept term is much more volatile, as expected.

4.4. *Varying coefficient estimates for a single simulation*

So far only model summaries for the 100 simulated datasets have been presented. Greater insight is possible by choosing a single simulation and investigating coefficient estimation over space and time. Here, coefficient estimates were extracted from the GAM and MGTWR model of the 54th simulated dataset to explore the differences in coefficient surfaces over the 15 time periods. The best GAM for this dataset was one with a stationary intercept, combined TP space-time smooths for x_1 and x_2 , and separate space and time smooths for x_3. The MGTWR model had space-time adaptive bandwidths of 3368, 63, 83 and 43 observations for the intercept, x_1, x_2 and x_3, respectively.

The estimated space-time coefficient surfaces for β_1, β_2 and β_3 generated by the GAM and MGTWR model are shown in Figures 6, 7 and 8, with the values limited to the range [0, 2]. Spatial and temporal differences between the estimated coefficients and the true coefficients (β_1, β_2 and β_3) for the GAM and MGTWR model are shown in the Appendix to aid interpretation. In general, the GAM estimates surfaces that are closer to the true surfaces than the MGTWR model, as might be expected from Figure 4, but with some under-estimation of coefficients that are near zero. Several specific trends are evident in the comparisons:

- For β_1 (Figure 6), there are clear differences between the GAM and MGTWR coefficient estimations. The GAM captures the smooth surfaces of the true coefficients in Figure 1 well, whereas the MGTWR does this less reliably, with large underestimations in times 2 to 5 and 14 and 15. The GAM differences are small and the MGTWR differences are larger and more locally clustered (see Figure A1).
- For β_2 (Figure 7), there are strongly clustered under-estimations of the true coefficient surfaces by the MGTWR model. These follow the space-time pattern of the Gaussian hotspot in the true surfaces (see Figures 2 and A2), especially in times 2 to 7 and 14 and 15. There is less structure in the GAM differences, although greater under-estimation of coefficients that are near zero.
- For β_3 (Figure 8), the MGTWR differences have similarities with the space-time distribution of the true surfaces (see Figures 3 and A3) with large over-estimations in time 1 and large underestimations in time 15 . The GAM differences are smaller and more randomly distributed in relation to the true coefficient surfaces.

5. Empirical case study

The approach proposed in this study of determining the best GAM form for space-time varying coefficient modelling can be illustrated with an empirical case study. The `stgam` R package (Comber *et al.* 2026b) contains a dataset of a sample of 2000 observations of Normalised Difference Vegetation Index (NDVI) plus climate data for a small area (approximately 200 km by 200 km) of the Chaco dry rainforest in South America (Figure 9). The monthly NDVI data are from the PKU GIMMS NDVI v1.2 dataset from 2012 to 2022 (Li *et al.* 2023), with mean monthly maximum temperature and precipitation extracted from the TerraClimate dataset (IDAHO_EPSCOR/TERRACLIMATE) (Figure 10).

The study's GAM approach and MGTWR were both used to fit the following space-

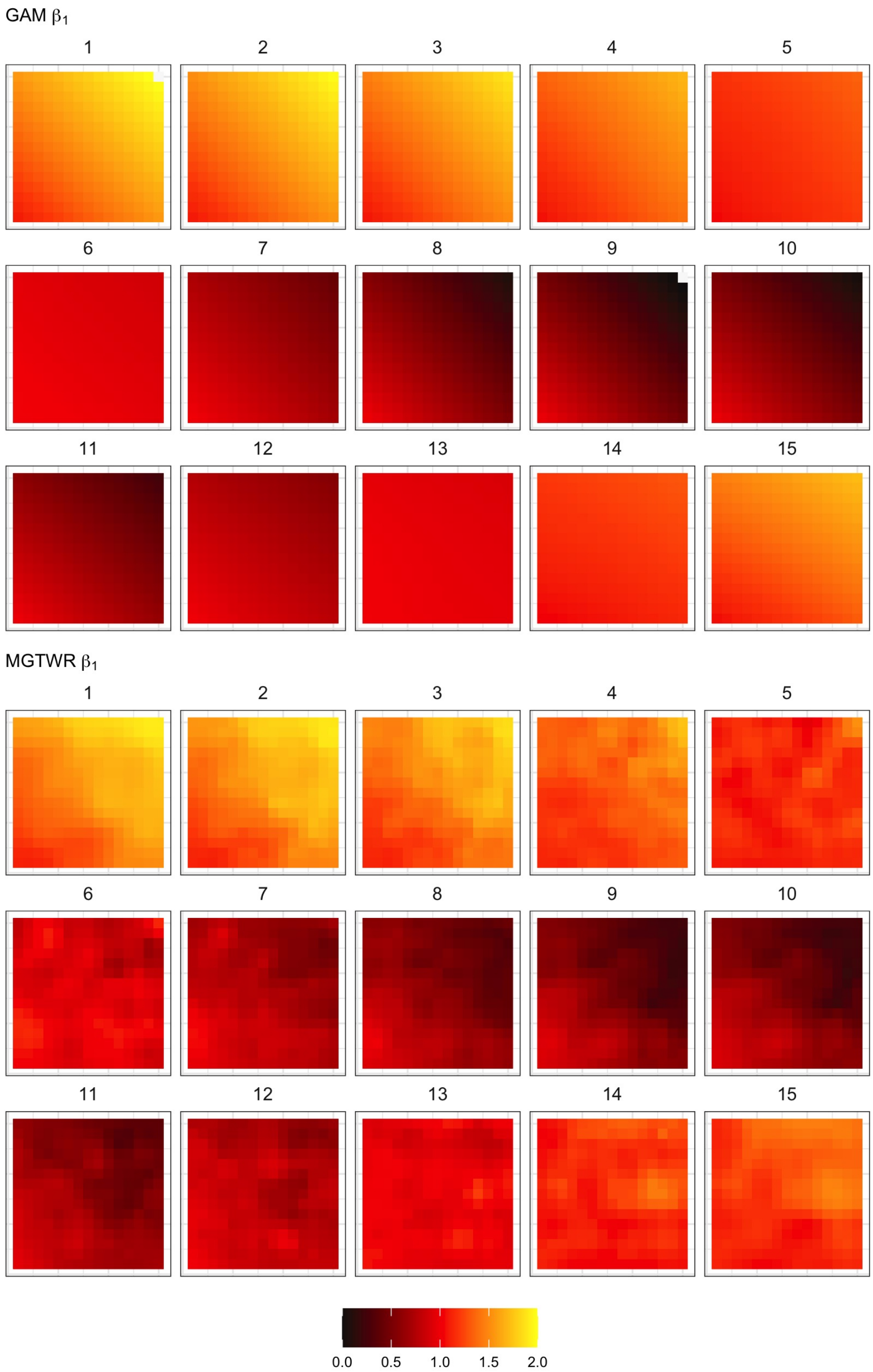

Figure 6. The GAM (top) and MGTWR (bottom) estimated true β_1 coefficients.

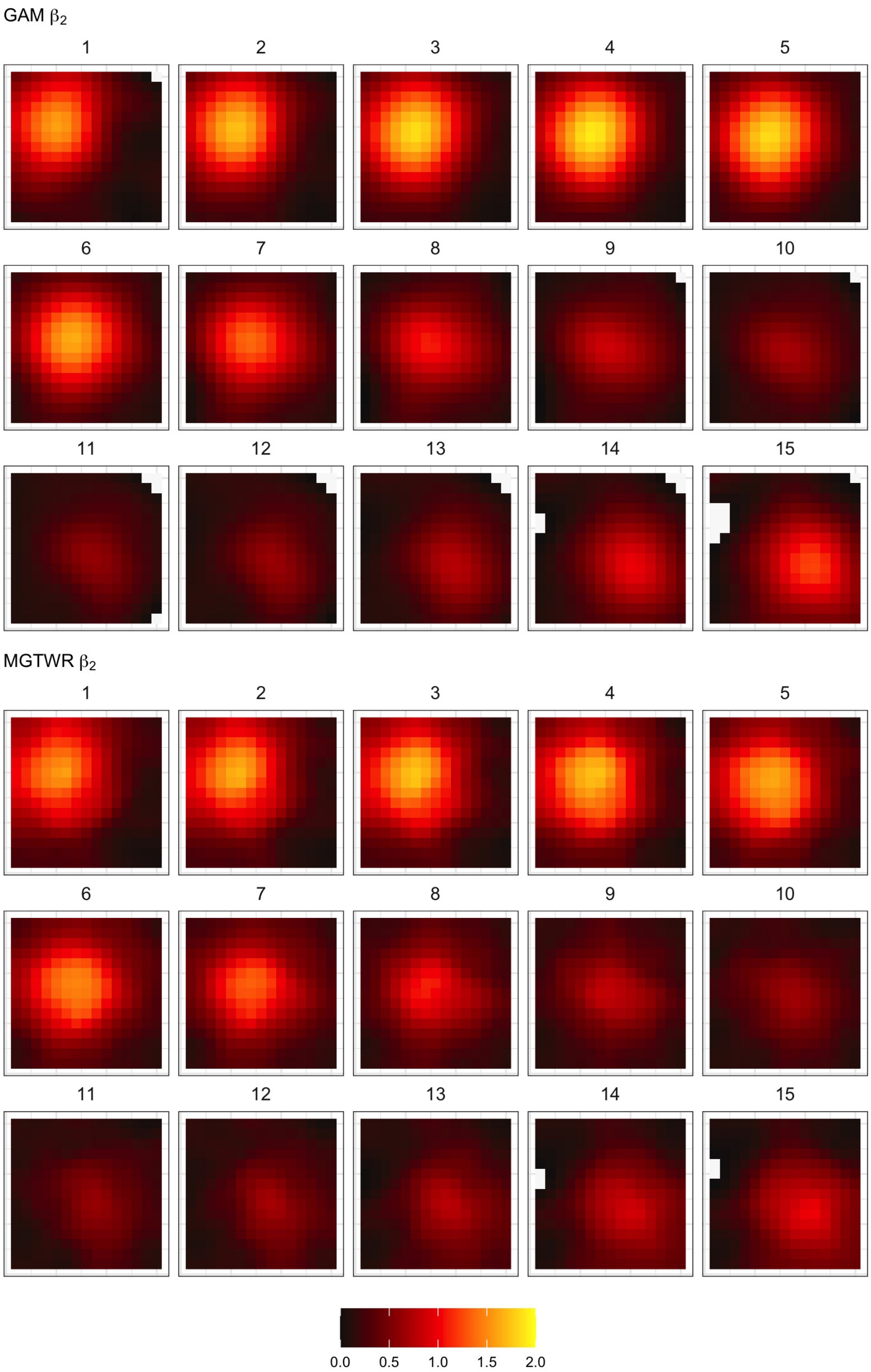

Figure 7. The GAM (top) and MGTWR (bottom) estimated true β_2 coefficients.

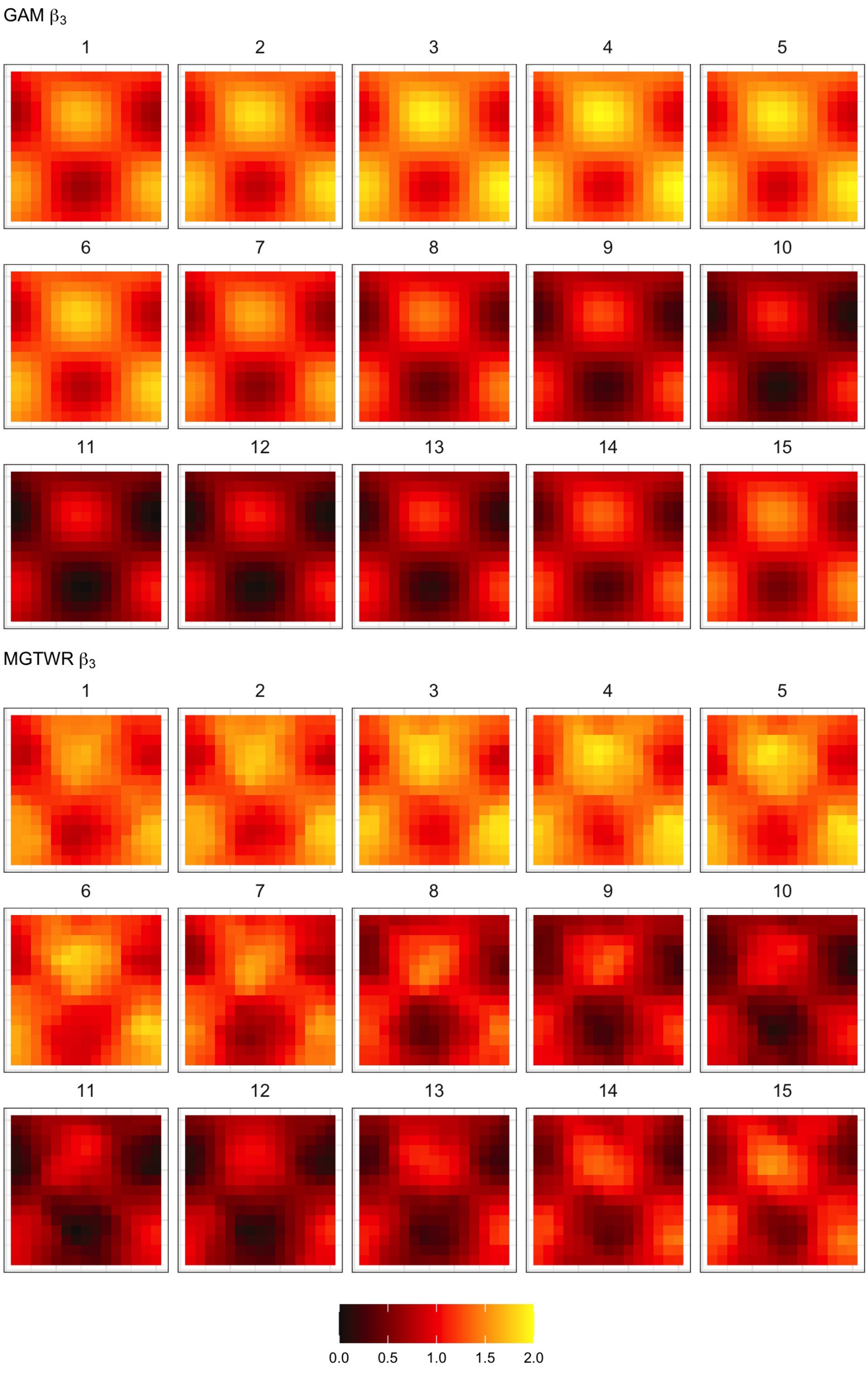

Figure 8. The GAM (top) and MGTWR (bottom) estimated true β_3 coefficients.

Figure 9. The location and context of the time series Chaco NDVI observations in Brazil, with a small transparency term and an OpenStreetMap backdrop.

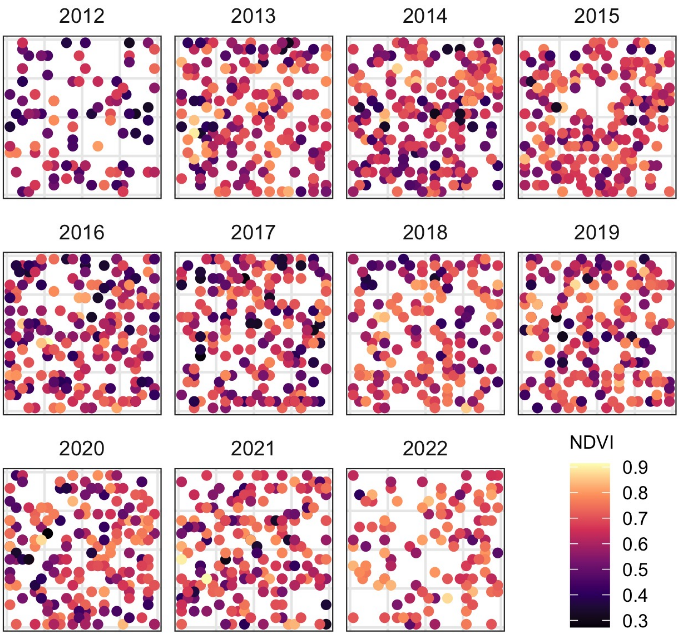

Figure 10. The Chaco NDVI case study data with monthly observations summarised over each year.

Table 3. Fit measures of the GAM and MGTWR model.

	R^2	RMSE	MAE	AIC
GAM	0.857	0.049	0.038	-5817.7
MGTWR	0.582	0.084	0.068	-4038.9

Table 4. The distributions of the GAM derived varying coefficient estimates, with smooth spatial and temporal length ranges, for the Chaco case study data.

	Intercept	Maximum temperature	Precipitation
Min.	0.64252	-0.02060	-0.00025
1st Qu.	0.92677	-0.01588	-0.00010
Median	1.02663	-0.01253	-0.00004
Mean	1.02734	-0.01264	-0.00004
3rd Qu.	1.13949	-0.00948	0.00001
Max.	1.32251	-0.00410	0.00017
Spatial length range	146.1	146.5	135.3
Temporal length range	39.9	NA	NA

time varying coefficient model:

$$\mathrm{NDVI}_i = \beta_0(u_i, v_i, t_i) + \beta_1(u_i, v_i, t_i)\, x_{1,i} + \beta_2(u_i, v_i, t_i)\, x_{2,i} + \varepsilon_i \tag{9}$$

where (u_i, v_i) denote the spatial coordinates of observation i, t_i represents the time of observation (in months), $\beta_0(u_i, v_i, t_i)$ are spatially and temporally varying intercepts $\beta_1(u_i, v_i, t_i)$ and $\beta_2(u_i, v_i, t_i)$ are the spatially and temporally varying coefficients associated with maximum temperature ($x_{1,i}$) and precipitation ($x_{2,i}$), respectively, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the random error term.

The GAM approach evaluated 180 potential models, with the two predictor variables specified in different ways in different space-time smooths, as described above. A space-time distance matrix specified with $\rho = 1$ was used to calibrate the MGTWR, and it was again specified with adaptive Gaussian kernels whose bandwidths were optimally found via AIC. The fit measures for the response (NDVI) of the two models are summarised in Table 3 and in this case, the GAM approach generated a final model with much lower prediction errors and AIC than MGTWR.

The top ranked GAM specified separate space and time smooths for the Intercept, a spatial smooth for maximum temperature and a spatial smooth for precipitation. This suggests that the size of the temporal effects is low and is effectively swept up in the temporal smooth for the intercept. This perhaps is not surprising as the data covers 132 months, and whilst temporal (e.g. seasonal) variation might be present, it evidently has a low effect. This is confirmed when the varying coefficient estimates are extracted and the GAM smooth spatial and temporal length ranges are examined (Table 4).

The MGTWR varying coefficients are summarised in Table 5 together with the optimised space-time bandwidths. Bandwidths were large for the intercept and precipitation, and small for maximum temperature, suggesting a highly localised, non-stationary relationship in space-time between maximum temperature and NDVI, and a global, stationary relationship between precipitation and NDVI. The intercept is also global and stationary. These results are somewhat contradicted by the summaries of the coefficient estimates which indicate only small variations over time and space in the relationship between maximum temperature and NDVI. This suggests a consequence

Table 5. The distributions of the MGTWR derived varying coefficient estimates, with kernel bandwidths, for the Chaco case study data.

	Intercept	Maximum temperature	Precipitation
Min.	1.52059	-0.03492	0.00050
1st Qu.	1.52080	-0.03157	0.00050
Median	1.52095	-0.03055	0.00050
Mean	1.52095	-0.03064	0.00050
3rd Qu.	1.52109	-0.02966	0.00051
Max.	1.52130	-0.02688	0.00051
Bandwidths	1997	42	1997

of over-fitting.

Without examining the spatial distributions of the coefficient estimates over time, there are evident differences in the two models:

- The GAM is a more accurate predictor of NDVI than the MGTWR model across a range of fit measures.
- The form of the GAM smooths suggests strong spatial interactions for NDVI predictors coupled with weak temporal interactions, with only the intercept specified with a temporal smooth. No combined space-time smooths, in which space and time interact, were specified.
- The MGTWR bandwidths suggest relationships with NDVI that are stationary with respect to time and space for the intercept and precipitation, and strongly nonstationary for maximum temperature.
- The coefficients estimated by the GAM indicate greater variation in the relationships of all predictors with NDVI than those estimated by the MGTWR model.

As there is no temporal variation in the response-to-predictor relationships (as indicated by the nature of GAM smooths in the final selected model, Table 4), it is possible to focus only on the spatial dimension for the GAM's varying coefficients. Here, the spatial variation in the relationship of maximum temperature with NDVI for 2020 is shown by mapping the corresponding coefficients over a gridded surface (Figure 11). The coefficients are mapped with their t-values, and in this case, coefficient estimates are highlighted where $|t| > 1.96$. Clearly most coefficients are significantly different from zero at this 95% level. This use of t-values for coefficient significance is considered highly informal and exploratory, and is similarly true when used in any GWR-based model (e.g Comber *et al.* (2023a)) such as MGTWR (see discussion).

6. Discussion

This paper describes an approach for space-time varying coefficient modelling using a GAM that specifies the model form that best captures the space-time dependencies in the data's relationships. It is a data driven approach to model selection rather than one based on user perceptions of the data generating process. This is important as data are increasingly from 3rd parties and of unknown provenance. Different GAM smooths are evaluated using AIC to determine the best model form, where the resultant GAM is taken to best capture data relationships through its coefficient estimates. The paper demonstrates the approach through simulated data with known coefficient space-time dependencies and an empirical case study. In both, the GAM approach was compared

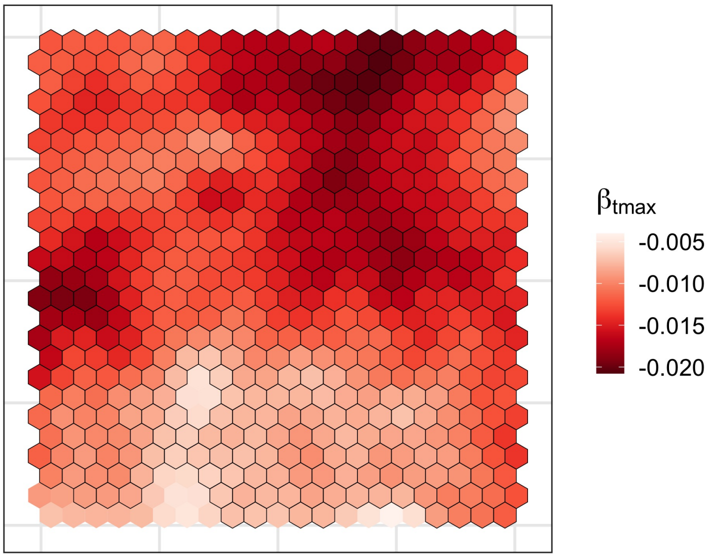

Figure 11. The GAM coefficient estimates for maximum temperature in 2020, with hexagon borders included for locations where the coefficient estimate is significant at the 95% level.

with MGTWR and found to out-perform MGTWR across a number of analytical aims and objectives.

In many applied studies of space-time GAMs, assumptions are made about the presence and nature of the space-time interactions and dependencies in the data. The typical workflow is to fit the most complex model that user believes best captures the data generating process. In this study's simulation experiment and empirical case study, each GAM may well have been specified with TP smooths for each predictor variable. In this context, an important extension in this paper is to explicitly examine different smooth forms for each predictor variable. These provide insight into the data's space-time dependencies and support an important aim in modelling urban and environmental systems: to understand the strength and nature of the interactions between an outcome and different individual drivers, and critically how they vary over space and / or time. The form of any space-time dependencies is usually unknown, with uncertainty about whether they are present, and if so whether they are independent of each other or whether they interact. Thus, a key consideration in this study's varying coefficient context is to determine the nature and strength of any space-time interactions in the data and to specify the GAM appropriately.

There is a rich history of space-time varying coefficient modelling and different types of models have been proposed. Unlike the GAM approach demonstrated in the study, most make implicit assumptions about the presence and nature of the space-time relationships present in the data; typically, assuming the same underlying model form for each response-to-predictor relationship. This is the case for ESTF and GTWR models (Griffith 2008, 2010, Murakami *et al.* 2017, 2024, Patuelli *et al.* 2011, Chun 2014, Huang *et al.* 2010, Elhorst 2003, Di Giacinto 2006, Deng *et al.* 2017). Although for GTWR, some recourse to such assumptions is possible through MGTWR, where a different bandwidth is found for each relationship (Liu *et al.* 2017, Wu *et al.* 2019, Zhang *et al.* 2021, Yu and Fotheringham 2025). Similar underlying assumptions similarly arise in alternative space-time varying coefficient GAMs, such those proposed in Kim and Wang (2021), Salazar *et al.* (2021) and Feng (2022), and also in methodologies that specify GP smooths (e.g. Anderson *et al.* (2022), Dambon *et al.* (2021), Dambon *et al.* (2022), Comber *et al.* (2024b), Comber *et al.* (2023c)). Many space-time varying coefficient models are a direct extension of a simpler spatially varying coefficient model where a key challenge is how to define the space-time distance metric (and hence the use of TPs, here).

The varying coefficient approach using GAM with smooths proposed in this study makes no such assumptions about the presence and nature of any space-time dependencies in the data relationships or the process being considered. Rather, it constructs multiple models, each specifying different permutations of the response-to-predictor relationship including in parametric form, in a spatial smooth, in a temporal smooth, in separate space-time smooths and in combined space-time smooths. It then identifies the 'best' model using AIC to rank them. This is the approach described in the `stgam` R package (Comber *et al.* 2026b). Evaluating different model forms in this way as part of the model specification process provides insight into how individual factors or drivers have different space-time dependencies with the outcome.

The empirical case study might, *a priori*, have been expected to exhibit interacting space-time effects due to its wide geographic extent and long time series. Whereas in a more local study over a shorter time period, any space-time effects might be expected to interact less strongly or be independent of each other, due to low variations over space and time of the process drivers. Undertaking investigations of the most appropriate model to specify allows interactions in the data relationships to be unpicked and supports deeper understanding about the nature of the space-time process being considered. Such explicit investigation of the space-time dependencies present in the data relationships avoids making assumptions about their presence and nature, by, for example, simply fitting the most complex model that the user believes is valid for the use-case and system. A key objective in much geographic analysis is to explicitly investigate space-time dependencies.

This study's approach to space-time varying coefficient modelling using GAMs is is very flexible, as demonstrated in recent related work (Klappstein *et al.* 2024). This study's approach extends GAMs with GP splines for spatially varying coefficient models (Comber *et al.* 2023c) into the temporal domain, with enhancements that consider not just predictor inclusion but also the form of predictor inclusion. A further contribution is the estimation of predictor-to-response variable relationship length ranges. These describe how quickly the GAM smooth function changes across space or time with, for example, a small spatial length range suggesting changes rapidly over small distances and a large temporal length range that the effect is stable over time. They provide estimates of the distances (in space or time) at which the smooth value changes by a threshold (here 50% of the maximum was used) and provide indications of the scale of the space-time dependencies in the data and data relationships. These measures are analogous to kernel bandwidths found in GWR and extensions, such as the MGTWR model employed here.

However there are some important considerations, that apply to any varying coefficient approach. First is the presence of structural discontinuities (for example, if the study area is divided by a large water body or national boundaries with different social systems). In such cases estimating a coefficient surface might be inappropriate. This applies to all methods for estimating spatially varying coefficients. Divisions can be masked (e.g. Comber *et al.* (2011)) or discrete analyses undertaken (e.g. Gosal *et al.* (2021)). Second, there is potential for identification errors to occur for intercepts that lack spatial or temporal dependencies (for example in Table 1 and Figure 5. Further work will investigate how the proposed approach infers the absence of such dependencies whether intercepts should always be specified with space-time dependencies in space-time problems. Finally, the analyses assumed an absence of any temporal lag in the response to predictor relationships. Some exploratory work has been undertaken to automatically specify appropriate lags for each predictor variable within space-time GAM smooths (Comber *et al.* 2025), which will be extended in the future.

Future work will continue in several areas, including the development of effect length scale and effect size in more depth. Some of these have already been included in `stgam` package updates. There are several approaches that could be used for the same task, such as that found with for variogram model parameter estimation (Hiemstra *et al.* 2009), where the variogram's range parameter could be used. Future work will also explore the impacts of some of the issues that have historically been problematic for GWR-based approaches such as local collinearity, and coefficient significance with inherent issues with multiple hypothesis testing. Finally, extrapolation and prediction beyond the spatial extent of the input data using GAM smooths parameterised with location can be difficult: unconstrained spline-based GAMs have been found to be unreliable for extrapolation beyond the observed data range. In models fitted with spline bases (e.g. TPRSs), predictions outside the data domain are largely determined by the penalty structure and the boundary behaviour of the basis functions rather than by the data. Future work will also examine how extrapolation — particularly in space–time TP smooths — can be interpreted as penalty-driven extensions (often approximately linear or flat) rather than statistically validated forecasts.

7. Conclusions

This paper proposes a geographically and temporally focused approach to multiscale varying coefficient modelling using GAMs to capture space-time interactions. It extends recent work describing spatially varying coefficient approaches (Comber *et al.* 2023c) into the space-time dimension and explicitly tests for the presence and nature of any space-time dependencies in order to inform model specification. This is done by constructing and evaluating multiple GAMs, each with space-time interactions defined in different ways for each predictor variable. Model evaluation was undertaken by AIC, as this allowed each GAM to be ranked and the best one specified (by the lowest AIC). The GAM approach was demonstrated through simulated data and an empirical case study (data from the Chaco dry rainforest in South America). In both, the GAM approach was compared with MGTWR and found to out-perform MGTWR across several analytical aims and objectives. The analyses were undertaken using tools in the stgam package (Comber *et al.* 2026b) and several areas of further work and package extension were identified.

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science (BRIDGE fellowship No. 220305), by the UKRI (ES/Y006259/1) and by the Biotechnology and Biological Sciences Research Council (BBS/E/RH/23NB0008 and BBS/E/RH/230004C).

Data availability

The code and data used to undertake the analysis and to re-create all the figures in the paper are at `https://doi.org/10.6084/m9.figshare.29880935`.

AI Disclosure

ChaptGPT (GPT-5.3 Instant) was used to suggest some R code to generate β_2 with non-separable space-time effects and β_3 which has separate space and time effects.

Biographies and contribution

Lex Comber is Professor of Spatial Data Analytics at the University of Leeds. With Chris Brunsdon, he authored the first 'how to...' book for spatial analysis and mapping in R. His current research interests include methods for scale informed analyses and handling space-time data.

Professor Paul Harris is a Principal Research Scientist at Rothamsted Research. His work focuses on the development and application of data-driven and process-based models for the understanding and forecasting of spatio-temporal agroecological processes for sustainable and resilience farming.

Chris Brunsdon is Professor of Geocomputation at Maynooth University's National Centre for Geocomputation. His work focuses on spatial data analysis, geocomputation, and the development of statistical and computational methods for understanding geographical patterns and processes, with particular interests in crime and health data. He has published widely on spatial analysis and teaches data science, spatial modelling, and reproducible analytical workflows.

All authors contributed equally to the ideas in the paper. LC led the writing and developed the code. PH led the statistical positioning. CB provided the directional steer. All authors edited drafts along the way.

Appendix: Difference between estimated and true coefficients

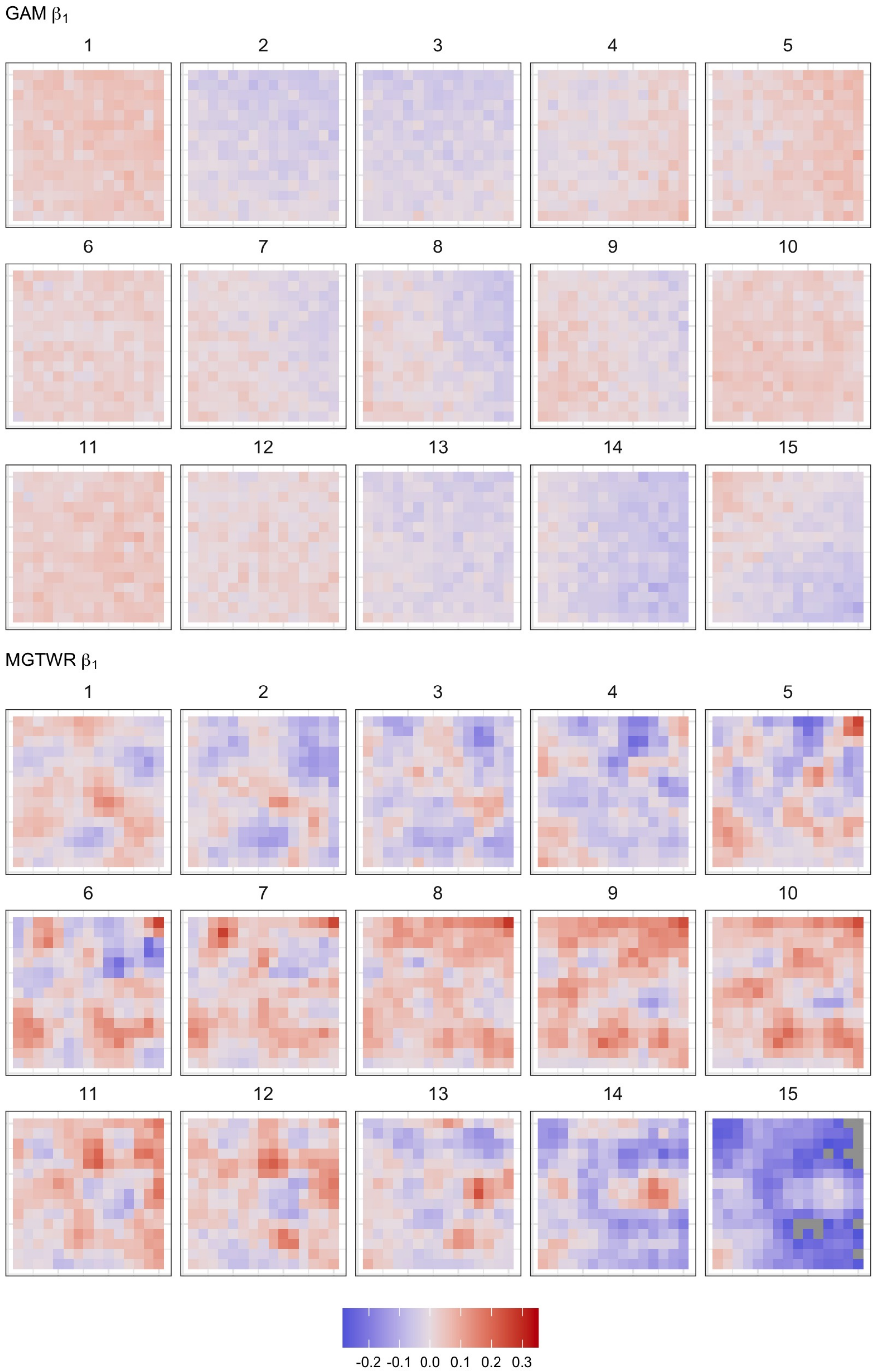

Figure A1. Difference between the GAM (top) and MGTWR (bottom) estimated coefficients and the true β_1 coefficients.

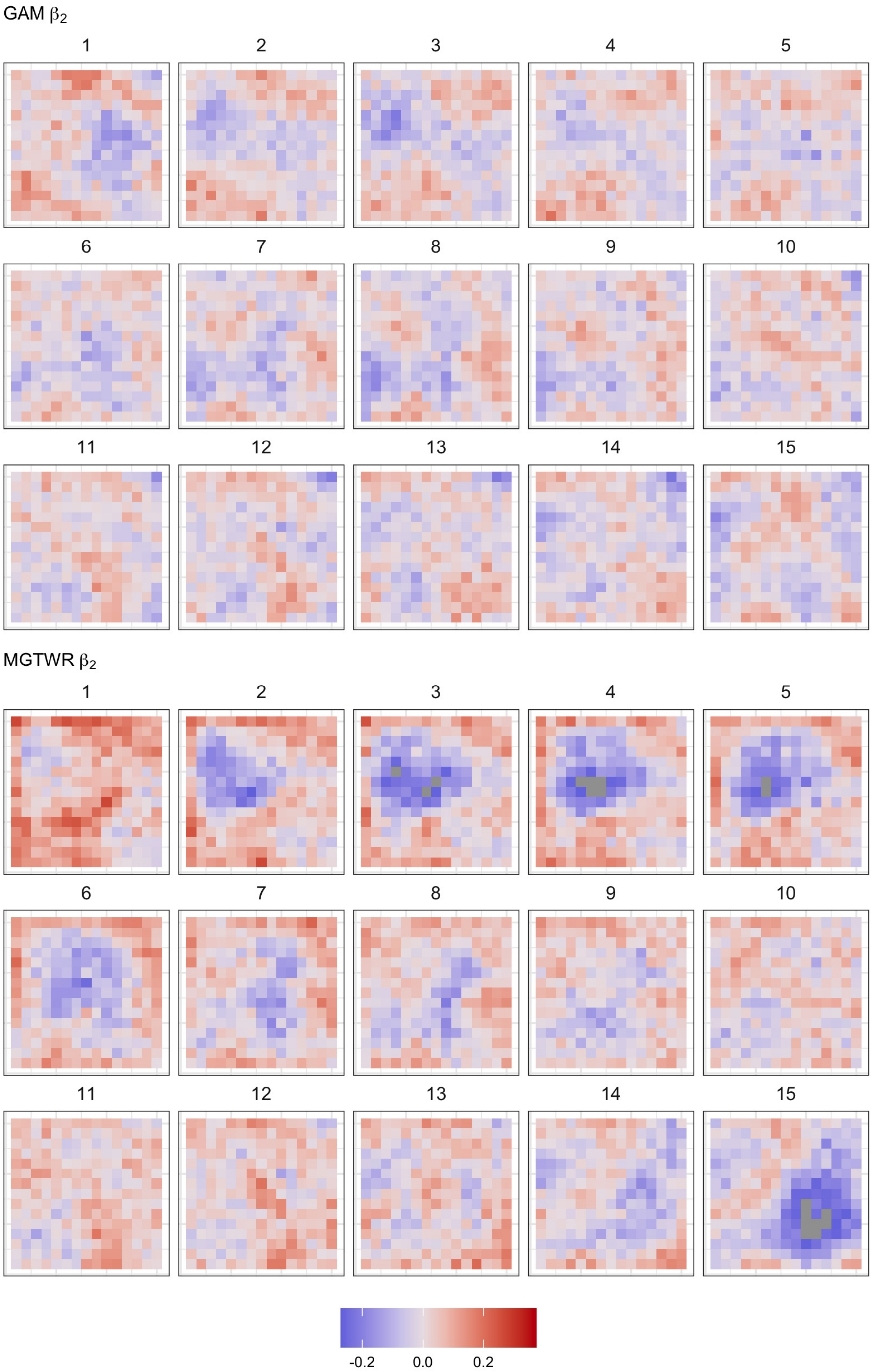

Figure A2. Difference between the GAM (top) and MGTWR (bottom) estimated coefficients and the true β_2 coefficients.

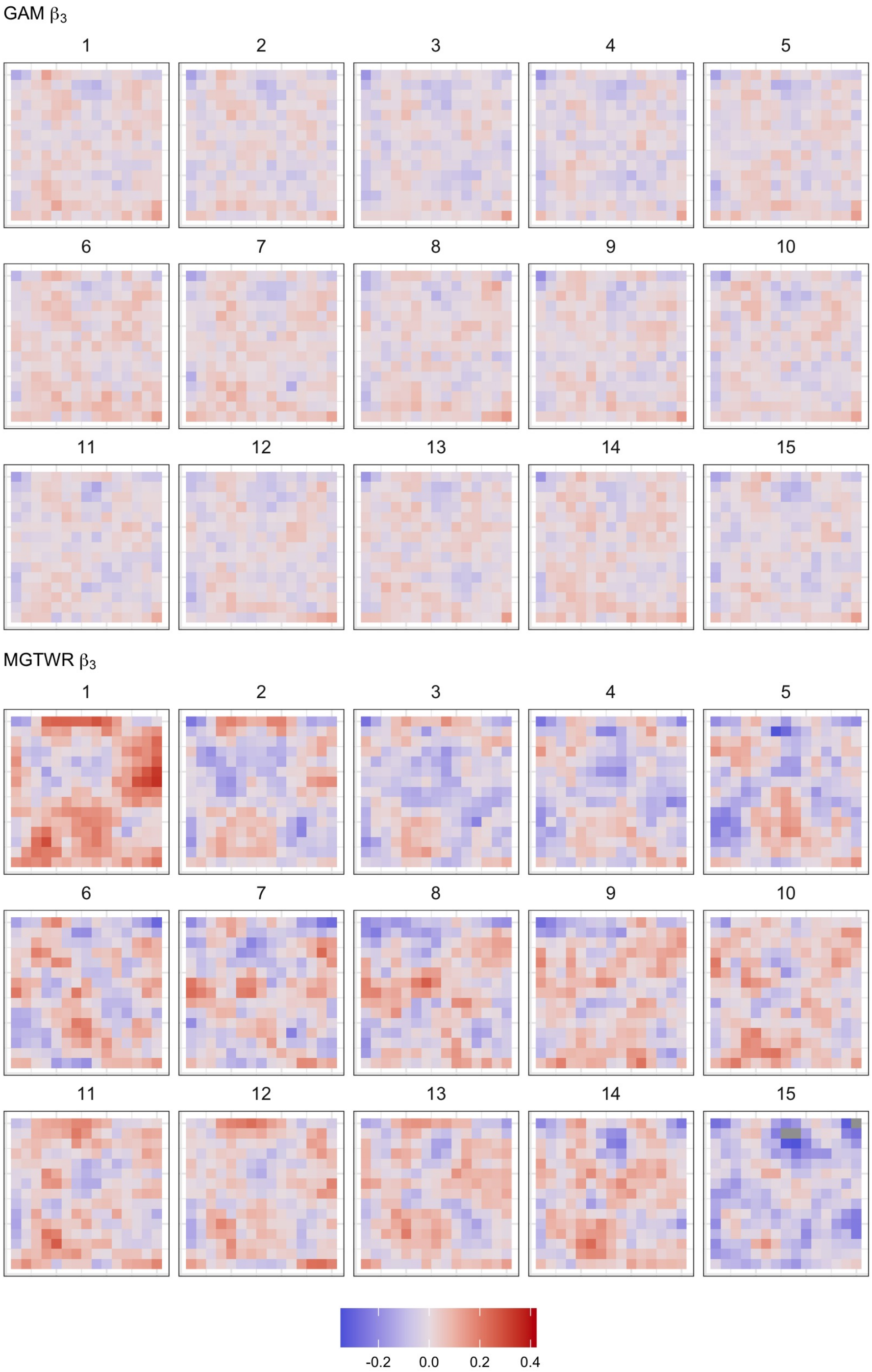

Figure A3. Difference between the GAM (top) and MGTWR (bottom) estimated coefficients and the true β_3 coefficients.

References

Anderson, S.C., *et al.*, 2022. sdmtmb: an r package for fast, flexible, and user-friendly generalized linear mixed effects models with spatial and spatiotemporal random fields. *BioRxiv*, 2022–03.

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

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

Casetti, E., 1972. Generating models by the expansion method: applications to geographical research. *Geographical analysis*, 4 (1), 81–91.

Chun, Y., 2014. Analyzing space–time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geographical Analysis*, 46 (2), 165–184.

Comber, A., *et al.*, 2023a. A route map for successful applications of geographically weighted regression. *Geographical Analysis*, 55 (1), 155–178.

Comber, A., Harris, P., and Brunsdon, C., 2023b. Multiscale spatially and temporally varying coefficient modelling using a geographic and temporal gaussian process gam (gtgp-gam)(short paper). *In*: *12th International Conference on Geographic Information Science (GIScience 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Comber, A., Harris, P., and Brunsdon, C., 2023c. Multiscale spatially varying coefficient modelling using a geographical gaussian process gam. *International Journal of Geographical Information Science*, https://doi.org/10.1080/13658816.2023.2270285.

Comber, A., Harris, P., and Brunsdon, C., 2024a. Multiscale spatially varying coefficient modelling using a geographical gaussian process gam. *International Journal of Geographical Information Science*, 38 (1), 27–47.

Comber, A., Harris, P., and Brunsdon, C., 2025. How much time to include in multiscale space-time regressions? optimising predictor variable temporal lags. *AGILE: GIScience Series*, 6, 19.

Comber, A., *et al.*, 2024b. Encapsulating spatially varying relationships with a generalized additive model. *ISPRS International Journal of Geo-Information*, 13 (12), 459.

Comber, A. and Wulder, M., 2019. Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use.

Comber, A.J., Brunsdon, C., and Radburn, R., 2011. A spatial analysis of variations in health access: linking geography, socio-economic status and access perceptions. *International journal of health geographics*, 10 (1), 44.

Comber, L., Harris, P., and Brunsdon, C., 2026a. *A geographer's introduction to space-time regression with gams using 'stgam'*. Accessed: 2026-02-02, Available from: `https://cran.r-project.org/web/packages/stgam/vignettes/stgam_vc_models_chaco.html`.

Comber, L., Harris, P., and Brunsdon, C., 2026b. *stgam: Spatially and temporally varying coefficient models using generalized additive models*. R package version 1.2.0, Available from: `https://CRAN.R-project.org/package=stgam`.

Crespo, R., Fotheringham, S., and Charlton, M., 2007. Application of geographically weighted regression to a 19-year set of house price data in london to calibrate local hedonic price models. *In*: *Proceedings of the 9th International Conference on Geocomputation*. National University of Ireland Maynooth.

Dambon, J.A., Sigrist, F., and Furrer, R., 2021. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics*, 41, 100470.

Dambon, J.A., Sigrist, F., and Furrer, R., 2022. Joint variable selection of both fixed and random effects for gaussian process-based spatially varying coefficient models. *International Journal of Geographical Information Science*, 36 (12), 2525–2548.

Deng, M., Yang, W., and Liu, Q., 2017. Geographically weighted extreme learning machine: A method for space–time prediction. *Geographical Analysis*, 49 (4), 433–450.

Di Giacinto, V., 2006. A generalized space-time arma model with an application to regional

unemployment analysis in italy. *International Regional Science Review*, 29 (2), 159–198.
Dubrule, O., 1984. Comparing splines and kriging. *Computers & Geosciences*, 10 (2-3), 327–338.
Elhorst, J.P., 2003. Specification and estimation of spatial panel data models. *International regional science review*, 26 (3), 244–268.
Fahrmeir, L., *et al.*, 2021. Regression models. *In*: *Regression.* Springer, 23–84.
Fan, Y.T. and Huang, H.C., 2022. Spatially varying coefficient models using reduced-rank thin-plate splines. *Spatial Statistics*, 51, 100654.
Feng, C., 2022. Spatial-temporal generalized additive model for modeling covid-19 mortality risk in toronto, canada. *Spatial statistics*, 49, 100526.
Fotheringham, A.S., Crespo, R., and Yao, J., 2015. Geographical and temporal weighted regression (gtwr). *Geographical Analysis*, 47 (4), 431–452.
Fotheringham, A.S., Yang, W., and Kang, W., 2017. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107 (6), 1247–1265.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.
Gelfand, A.E., *et al.*, 2004. The dynamics of location in home price. *The journal of real estate finance and economics*, 29, 149–166.
Gelfand, A.E., *et al.*, 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98 (462), 387–396.
Geman, S., Bienenstock, E., and Doursat, R., 1992. Neural networks and the bias/variance dilemma. *Neural computation*, 4 (1), 1–58.
Gneiting, T., Genton, M.G., and Guttorp, P., 2006. Geostatistical space-time models, stationarity, separability, and full symmetry. *Monographs On Statistics and Applied Probability*, 107, 151.
Gosal, A.S., *et al.*, 2021. Do drivers of nature visitation vary spatially? the importance of context for understanding visitation of nature areas in europe and north america. *Science of The Total Environment*, 776, 145190.
Griffith, D.A., 2008. Spatial-filtering-based contributions to a critique of geographically weighted regression (gwr). *Environment and Planning A*, 40 (11), 2751–2769.
Griffith, D.A., 2010. Modeling spatio-temporal relationships: retrospect and prospect. *Journal of Geographical Systems*, 12, 111–123.
Hastie, T. and Tibshirani, R., 1990. Generalized additive models. chapman hall & crc. *Monographs on Statistics & Applied Probability. Chapman and Hall/CRC*, 1.
Hastie, T. and Tibshirani, R., 1987. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82 (398), 371–386.
Hastie, T. and Tibshirani, R., 1993. Varying-coefficient models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 55 (4), 757–779.
Hastie, T., *et al.*, 2009. *The elements of statistical learning: data mining, inference, and prediction.* vol. 2. Springer.
Hiemstra, P.H., *et al.*, 2009. Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network. *Computers & Geosciences*, 35 (8), 1711–1721.
Huang, B., Wu, B., and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International journal of geographical information science*, 24 (3), 383–401.
Hutchinson, M. and Gessler, P., 1994. Splines—more than just a smooth interpolator. *Geoderma*, 62 (1-3), 45–67.
Jones, K., 1991. Specifying and estimating multi-level models for geographical research. *Transactions of the institute of British geographers*, 148–159.
Kim, M. and Wang, L., 2021. Generalized spatially varying coefficient models. *Journal of Computational and Graphical Statistics*, 30 (1), 1–10.
Klappstein, N.J., *et al.*, 2024. Step selection functions with non-linear and random effects. *Methods in Ecology and Evolution.*

Li, M., *et al.*, 2023. Spatiotemporally consistent global dataset of the gimms normalized difference vegetation index (pku gimms ndvi) from 1982 to 2022. *Earth System Science Data*, 15 (9), 4181–4203.

Liu, J., *et al.*, 2017. A mixed geographically and temporally weighted regression: Exploring spatial-temporal variations from global and local perspectives. *Entropy*, 19 (2), 53.

Lu, B., *et al.*, 2017. Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 31 (5), 982–998.

Lu, B., *et al.*, 2014. The gwmodel r package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, 17 (2), 85–101.

Ma, C., 2008. Recent developments on the construction of spatio-temporal covariance models. *Stochastic Environmental Research and Risk Assessment*, 22, 39–47.

McMillen, D.P., 1996. One hundred fifty years of land values in chicago: A nonparametric approach. *Journal of Urban Economics*, 40 (1), 100–124.

Murakami, D., 2021. Gaussian spatial regression using the spmoran package: case study examples. *arXiv preprint arXiv:2004.06070*. Available from: `https://arxiv.org/abs/2004.06070`.

Murakami, D., *et al.*, 2024. Sub-model aggregation for scalable eigenvector spatial filtering: Application to spatially varying coefficient modeling. *Geographical Analysis*, 56 (4), 768–798.

Murakami, D., *et al.*, 2017. A moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, 19, 68–89.

Pace, R.K., *et al.*, 1998. Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics*, 17, 15–33.

Pace, R.K., *et al.*, 2000. A method for spatial–temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16 (2), 229–246.

Patuelli, R., *et al.*, 2011. Spatial filtering and eigenvector stability: space-time models for german unemployment data. *International Regional Science Review*, 34 (2), 253–280.

Rigby, R.A. and Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 54 (3), 507–554.

Salazar, J.E., *et al.*, 2021. Generalized additive models with delayed effects and spatial autocorrelation patterns to improve the spatiotemporal prediction of the skipjack (katsuwonus pelamis) distribution in the colombian pacific ocean. *Regional Studies in Marine Science*, 45, 101829.

Shmueli, G., 2010. To explain or to predict? *Statistical science*, 289–310.

Stasinopoulos, D.M. and Rigby, R.A., 2008. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23, 1–46.

Stasinopoulos, M.D., *et al.*, 2017. *Flexible regression and smoothing: using gamlss in r*. CRC Press.

Umlauf, N., Klein, N., and Zeileis, A., 2018. Bamlss: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, 27 (3), 612–627.

Wood, S., 2021. *mgcv: Mixed gam computation vehicle with automatic smoothness estimation*. R package version 1.9-1, Available from: `https://CRAN.R-project.org/package=mgcv`.

Wood, S.N., 2006. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62 (4), 1025–1036.

Wood, S.N., 2017. *Generalized additive models: an introduction with r*. CRC press.

Wu, C., *et al.*, 2019. Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, 33 (3), 489–511.

Yang, W., 2014. *An extension of geographically weighted regression with flexible bandwidths*. Thesis (PhD). University of St Andrews.

Yu, H. and Fotheringham, A.S., 2025. On the calibration of multiscale geographically and temporally weighted regression models. *International Journal of Geographical Information Science*, 39 (6), 1203–1222.

Zhang, Z., *et al.*, 2021. Multiscale geographically and temporally weighted regression with a unilateral temporal weighting scheme and its application in the analysis of spatiotemporal characteristics of house prices in beijing. *International Journal of Geographical Information Science*, 35 (11), 2262–2286.

Zschech, P., *et al.*, 2022. Gam (e) changer or not? an evaluation of interpretable machine learning models based on additive model constraints. *arXiv preprint arXiv:2204.09123*.