JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 32 (2026), pp. 37–66
doi:10.5311/JOSIS.2026.32.522

RESEARCH ARTICLE

Specification of multiscale space-time varying coefficient GAMs

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

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

Received: August 30, 2025; returned: November 6, 2025; revised: March 18, 2026; accepted: June 9, 2026.

Abstract: This paper demonstrates an approach to the application of generalized additive models (GAMs) with space-time smooths to model coefficient processes that vary over space and time. The approach is to create and evaluate multiple GAMs, each with the predictor variables specified in different ways. It emphasizes the need to determine the nature of the space-time dependencies present in the data relationships rather than to assume them, based on the perceived data generating process, especially if this is unknown. The approach is explored using simulated coefficient data with known space-time dependencies. The GAMs are compared with multiscale geographically and temporally weighted regression (MGTWR) models and are shown to have marginally weaker predictive performance and to be marginally better at coefficient recovery. The inferential costs of misspecifying the target-to-predictor variable relationships in the GAMs is quantified both for individual variable main effects and interacting misspecifications. The approach is then applied to an empirical case study of NDVI (as a proxy for forest productivity) informed by precipitation and temperature in the Chaco dry rainforest of South America. The best GAM is determined and its space-time varying coefficient estimates are investigated. The methods and results are discussed and several areas of further work and enhancements to the `stgam` R package used to undertake this analysis are identified.

Keywords: Space-time relationships, Coefficient non-stationarity, Process heterogeneity, Varying parameter models

1 Introduction

Geographical analyses are frequently concerned with understanding spatial, and increasingly space-time, processes. This can be done through regression analyses, the outputs of which can be used to support process inference (understanding) as well as prediction inference (confidence). In terms of the former, a number of methodological refinements have been developed to enhance the ability to capture main spatial and/or temporal trends in data relationships, which are reviewed below. However, a key issue concerns the embedded assumptions of process spatial, temporal, or space-time heterogeneity in many of these methods, and the way that regressions are parameterized, either with model terms or the data that are passed to them.

This paper outlines an approach for undertaking space-time regression analysis using generalized additive models (GAMs) [26, 27] for the specific case when the GAM smooths (or splines—the terms are used interchangeably) are parameterized with observation location and/or time so they capture variation in the relationships between target and predictor variables. A typical workflow would specify the most complex model that the user believed to be valid for the system being investigated, relying on their understanding of the data generating process, with investigation of the effects of different model components and iterative model refinement. In a GAM context this might involve specifying different smooth types and forms for each predictor variable. However, it is common for the final models to be misspecified—i.e., they provide good predictive accuracy but do not take account of potential important interactions between GAM smooths and thus fail to capture process understanding (an example of this is in [10]).

Central to the approach suggested in this paper is the explicit investigation of the space-time interactions between target and predictor variables with the aim of specifying the most appropriate model. Multiple GAMs are created with each predictor variable specified in different ways and each model is evaluated through its AIC. Searching through this "model space" to find the "best" model avoids making assumptions about the existence and nature of the space-time dependencies present in the data relationships. The proposed approach identifies the model with the lowest AIC (i.e., one that minimizes prediction error but at the same time penalizes for model complexity). This approach explicitly seeks to unpick the space-time data relationships, and potentially thereby support deeper understanding about the space-time process being considered—a key objective in much geographic analysis. It enables each data relationship to be characterized in a different form—ranging from a full interactive space-time form to one that is invariant to any space-time effects (i.e., stationary and global). The analysis draws from the functionality of the `stgam` R package [11], which provides space-time wrapper functions to the `mgcv` R package [43] for constructing GAMs. For context, the GAM-based approach is demonstrated and compared with a multiscale geographically and temporally weighted regression (MGTWR) approach [45,49] using simulated space-time data with known data relationships. The resultant GAM forms are unpicked before the GAM approach is applied empirically to Chaco dry rainforest data of South America for the space-time investigation of drivers (precipitation and temperature) of forest productivity using remotely sensed normalized difference vegetation index (NDVI). All code and data are provided.

2 Background

Over the years, a number of refinements to classic linear regression have been proposed to support deeper inference about the spatial properties of the process under investigation and any spatial dependencies in the data, the model error and/or data relationships. These can be grouped into a number of broad families whose methods are similar and which are described below (of course there are other ways that this grouping could be organized, for example see [24]).

Firstly, spatial lag and/or spatial error models incorporate some measure of proximity. Such models include spatial lags of the dependent variable or the spatial dependence of the error term, spatial lags of predictors, or combined lag and error specifications for when both observation spatial autocorrelation and spatial spillover effects are present [1, 32, 39, 40]. These approaches include spatial autoregressive models, spatial lag models, and spatial error models and are often used in spatial econometric applications. These have been extended to the temporal domain to include repeated observations over time (panel data) and may control for period-specific shocks by incorporating fixed time effects, spatially varying fixed time effects (i.e., unit fixed effects), or temporal lags to complement spatial lags or space–time autoregressive models [4, 14, 37, 48]. This group of models all assume fixed relationships and thus only estimate fixed spatially invariant coefficients.

For models that estimate spatially varying coefficients (SVC), moving window or kernel-based approaches are possible, the best known of which is the geographically weighted regression (GWR) family of models [2, 17]. These models construct and combine a series of local models, from distance weighted data falling under the kernel, and several refinements and extensions have been proposed (see [5,21]). Key to GWR is finding the optimal size of the moving window or kernel bandwidth and its multiscale extension (MGWR) optimally determines a bandwidth for each predictor variable [19, 46]. MGWR is now regarded as the default GWR form [5]. GWR has been extended to include time in geographically and temporally weighted regression (GTWR) via a temporal kernel [18, 30] and this in turn has been extended to the multiscale case, MGTWR [45, 49].

The second main group of SVC models are those based on eigenvector spatial filtering (ESF). These models add the eigenvectors from a spatial weights matrix as predictor variables in the regression where in fixed coefficient form they account for residual spatial autocorrelation and spatial structure, but can be extended to an SVC form [23, 35, 36]. Space–time varying coefficient eigenvector filtering has used distributions of variable space-time autocorrelation, Moran's eigenvectors, and space–time connectivity matrices to remove autocorrelation in both dimensions [3, 25, 38].

The third main group of models that estimate SVCs are those based on GAMs [26, 27]. In attribute space, GAMs use splines (smooths) to model non-linear relationships between target and predictor variables. Splines can have different forms [27] and if they are parameterized with location, then the target-to-predictor relationships are modelled over space [7, 9, 15]. Here [31] propose a spline-based SVC method, using approximate bivariate smooths defined over a spatial triangulation to accommodate irregular study areas. [12, 13] describe a GAM SVC framework built on Gaussian processes (GPs), incorporating penalized maximum likelihood estimation for variable selection. [7, 9] similarly specify GAMs with GP smooths where they are defined over spatial coordinates (longitude and latitude), specifying them as spatial basis functions for each predictor. The GP approach enables SVCs to be captured through the direct modeling of spatial covariance structures.

For the space-time varying coefficient case, [41] incorporated delayed (lagged) effects within a GAM, combining thin-plate regression splines (TPRS) for spatial surfaces, cubic regression (CR) smooths for environmental predictors, and smooth lag structures to represent delayed influences. By explicitly modeling spatial autocorrelation alongside lagged effects, their approach improved predictive performance in species distribution modeling. A critical development for any form of space-time GAM was the introduction of tensor product (TP) smooths [44] and recent TP smooth refinements allow purely smooth surfaces over space and time to be modeled without random effects [43]. For example, [16] present a space–time varying coefficient GAM for mortality risk in which local socio-economic variables were used to construct a spatial smoother, with temporal changes in spatial patterns assessed through a space-time TP. In their specification, TPRS were used to represent spatial location, CR smooths used to capture temporal trends, and TP smooths to model non-separable space–time dynamics, thereby capturing the evolving spatial pattern of drivers of mortality over time.

For estimating space-time varying coefficients, TP smooths enable anisotropic variables (i.e., those measured over different scales) such as spatial coordinates and time, to be combined within a single smooth term. Different smooth penalties are applied across each smooth dimension, making it possible to analyse complex interactions such as space-time dependencies in data relationships. The space-time varying coefficient GAM of this study is parameterized with observation location (spatial coordinates) and observation time using TP smooths, but unlike previous studies, the emphasis is on model misspecification. Specifically, misspecification in terms of the wrong type of smooth for a given predictor and in terms of focusing on GAM predictive accuracy (model fit) rather than coefficient accuracy (as required for process understanding).

3 Methods

3.1 Simulation experiment

Simulated coefficients were created in a similar way to the spatial-only study of [7] but now extended to have space-time dependencies. These were used to create 100 simulated datasets and for each of these, multiple varying coefficient space-time GAMs were evaluated and the best of these (via AIC) compared with MGTWR where an optimal set of bandwidths was found using AIC also.

Simulated data with specific space-time dependencies were created to provide 4 true, known coefficient surfaces β_0, β_1, β_2 and β_3, over a 15×15 regular square grid (u, v), over 15 time periods (t). These dimensions were chosen to accommodate the high computational overheads of running the space-time GAM selection (described below) and the back-fitting estimation method for MGTWR.

The simulated true coefficient surfaces were defined as follows:

$$\beta_0(u, v, t) = 3 \tag{1}$$

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

$$\beta_1(u, v, t) = 1 + (U + V) \cdot \sin\left(2\pi T + \frac{\pi}{4}\right) \tag{2}$$

where $(U + V)$ is a spatial term, a diagonal gradient across space, $\sin(2\pi T + \pi/4)$ is a temporal modulation that controls amplitude, and:

$$\beta_2(u, v, t) = 1 + \frac{1}{324}\left[36 - \left(2 - \frac{v}{2}\right)^2\right]\left[36 - \left(2 - \frac{u}{2}\right)^2\right] + 0.8\,\sin\left(\frac{2\pi t}{n_t} + \frac{\pi}{2}\right) \tag{3}$$

$$\beta_3(u, v, t) = 1 + 0.5 \cdot \sin(2\pi T) \cdot \cos(\pi U) \cdot \cos(\pi V) \tag{4}$$

where $\cos(\pi U) \cdot \cos(\pi V)$ determines a smooth bowl-like shape spatial pattern and $\sin(2\pi T)$ determines time-modulated amplitude. Each of the true coefficient surfaces were re-scaled in the range 0 to 2.

Note that the true coefficient surfaces have the following properties:

- β_0 is a constant intercept term.
- β_1 is a spatial gradient modulated by time, with simple space-time interaction.
- β_2 has one distinct spatial peak and time oscillates smoothly in a sinusoidal fashion, with no space-time interaction.
- β_3 is a space-time wave pattern (oscillating surface) with moderate space-time interaction.

The simulated true regression coefficient surfaces for β_1, β_2 and β_3 are shown in Figures 1, 2 and 3. Simulated values for predictor variables x_1, x_2 and x_3 were generated from a random normal distribution $N(0, 1)$ and ϵ generated from a similar distribution multiplied by 0.25. The 100 datasets were finalized with the value of the target variable y which was calculated directly from the simulated true coefficients, the simulated predictor variables and the simulated error term as follows:

$$y_i = \beta_0 + \beta_1(u_i, v_i, t_i)x_{i1} + \beta_2(u_i, v_i, t_i)x_{i2} + \beta_3(u_i, v_i, t_i)x_{i3} + \varepsilon_i \tag{5}$$

where y_i is the target variable observed at spatial location (u_i, v_i) and time t_i; x_{i1}, x_{i2}, x_{i3} are the simulated predictor variables; β_0, $\beta_1(u, v, t)$, $\beta_2(u, v, t)$, and $\beta_3(u, v, t)$ are the true intercept and true coefficients for the predictor $x_{1\ldots3}$ varying over space and time; and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ is the error term.

3.2 Space-time GAMs with varying coefficients

The simulated data were created with simple space-time interactions for β_1, no space-time interactions for β_2, and with moderate space-time interactions for β_3. This suggests *a priori* that the best space-time GAM would have the following specification:

$$\begin{aligned} \mathrm{y}_i = {} & \beta_0 \cdot \text{Intercept}_i + \\ & f_1(u_i, v_i, t_i) \cdot x_{i1} + \\ & f_2(u_i, v_i) \cdot x_{i2} + f_3(t_i) \cdot x_{i2} + \\ & f_4(u_i, v_i, t_i) \cdot x_{i3} + \epsilon_i \end{aligned} \tag{6}$$

where y_i is the target variable, β_0 is the global, fixed intercept coefficient estimate, $f_1(u, v, t)$ is a TP smooth of spatial and temporal coordinates, modulated by x_1, $f_2(u, v)$

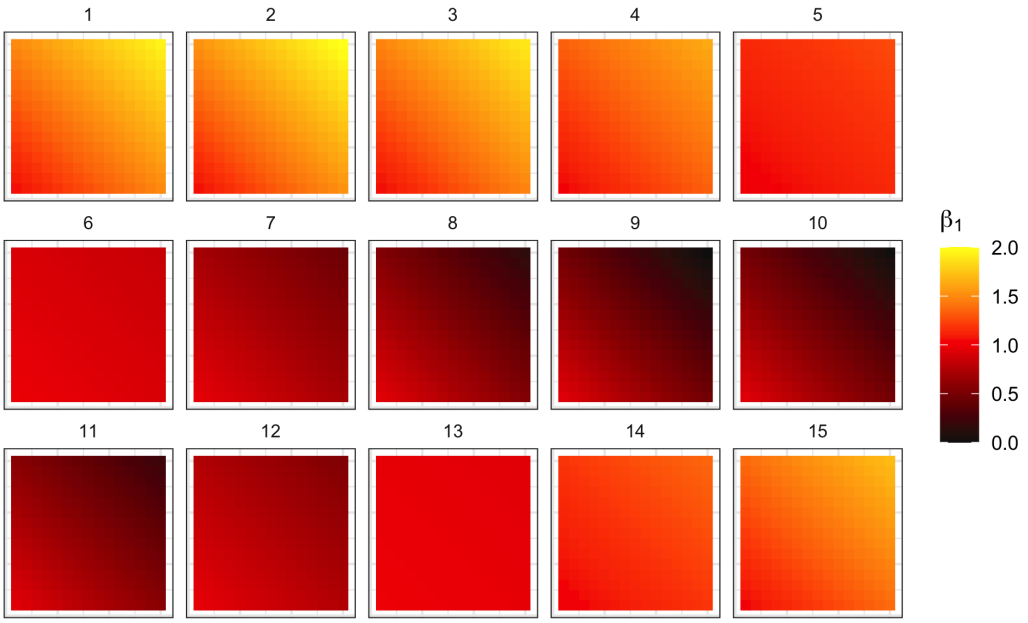

Figure 1: The simulated true regression coefficients over 15 time periods: β_1, with simple space-time interaction.

and $f_3(t)$ are the spatial and temporal smooths modulated by x_2, $f_4(u, v, t)$ is the TP space-time smooth modulated by x_3, with u, v the location coordinates, t the time variable, and ϵ_i the residual error term.

A common failing of much space-time modeling is to simply fit the most complex model that the analyst thinks is valid and assume the presence of specific space-time dependencies in data relationships (e.g., as done in MGTWR and related models). Here an important consideration was to explicitly investigate these, and to specify the most appropriate GAM. For space-time varying coefficient GAMs, each predictor can be specified in 6 different ways:

i) Omitted.
ii) Included as a parametric term.
iii) Included in parametric form and in a spatial smooth.
iv) Included in parametric form and in a temporal smooth.
v) Included in parametric form and in separate spatial and temporal smooths.
vi) Included in parametric form and in a combined space-time smooth.

Note that the parametric form is also specified to provide an offset term and a multiplicative constant. The intercept can be treated similarly but without being omitted. Thus,

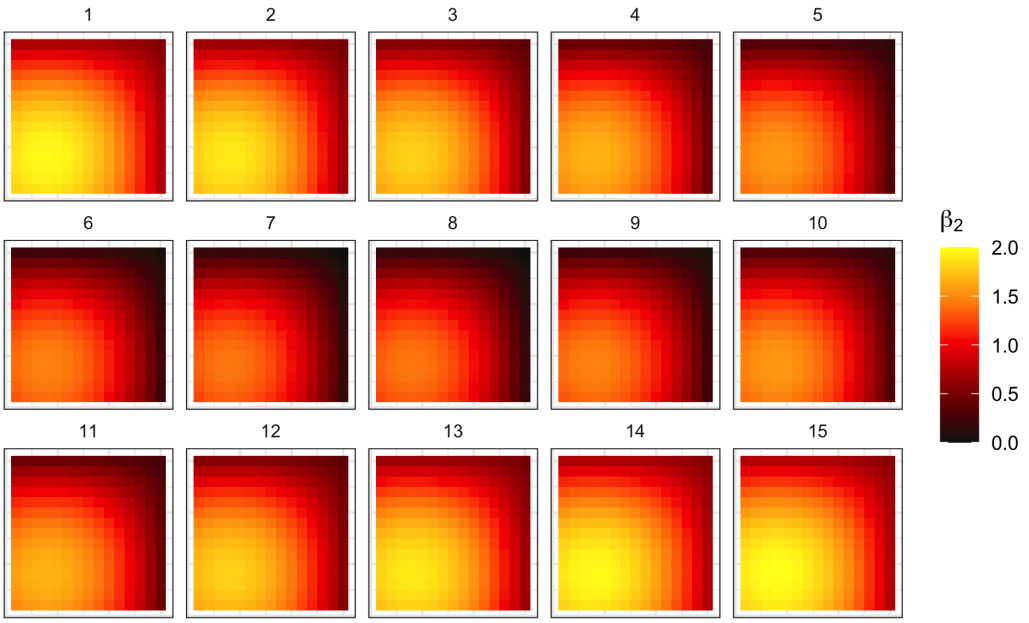

Figure 2: The simulated true regression coefficients over 15 time periods: β_2, with no space-time interaction.

for space-time GAMs with k predictor variables there are 5×6^k potential models to evaluate. For the simulated datasets described above, and with $k = 3$ variables (`x1`, `x2`, and `x3`), there are 1080 models to evaluate for each simulation. For model estimation, each GAM was specified with a restricted maximum likelihood (REML) approach which tends to reduce overfitting by providing more stable estimates of the smoothing parameters. Each GAM was evaluated through its AIC score with the "best" GAM having the lowest AIC. The intercept was specified as an addressable term in each GAM to allow it to vary over space and time.

The single GAM smooths (options iii), iv), and v) in the list above) were specified with TPRS and the combined space-time TP smooths (option vi) above) were constructed specifying TPRS to model location (u, v), and low-rank CR splines to model time, t. This approach allows for interactions between space and time margins, with location in a 2-dimensional smooth basis and time in a 1-dimensional one. These are combined via the TPs. The reason for the specification of different bases within the TP smooths in this way is that standard smooths (such as TPRS) are isotropic. They use a single basis under the assumption that all the variables specified in the smooth are on the same scale. The TP smooths are anisotropic (i.e., they relax this assumption) and can combine variables on different scales, like space and time, using separate basis functions, which are then combined.

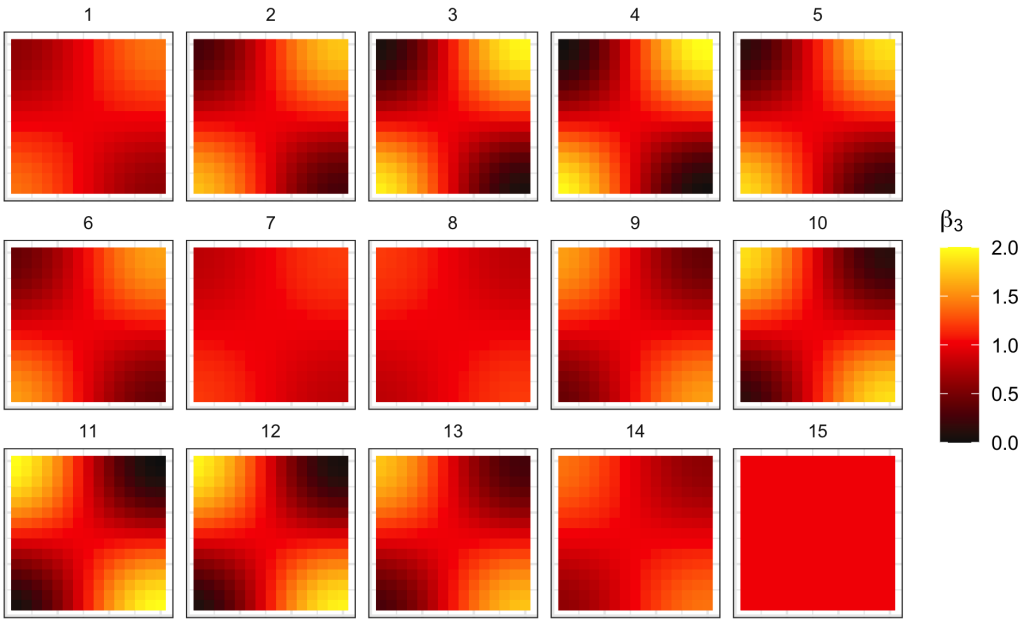

Figure 3: The simulated true regression coefficients over 15 time periods: β_3, with moderate space-time interaction.

Thus, a TP can be specified to investigate space-time effects in the target variable y for observation i.

$$\mathrm{y}_i = f(u_i, v_i, t_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \tag{7}$$

$$f(u, v, t) = \sum_{k=1}^{K_s} \sum_{\ell=1}^{K_t} \beta_{k\ell}, \phi_k^{(s)}(u, v), \phi_\ell^{(t)}(t) \tag{8}$$

where $\phi_k^{(s)}(u, v)_{k=1}^{K_s}$ are the spatial basis functions (e.g., TPRS), $\phi_\ell^{(t)}(t)_{\ell=1}^{K_t}$ are the temporal basis functions (e.g., low-rank CR splines). $\beta_{k\ell}$ are the TP coefficients and K_s and K_t are the numbers of spatial and temporal basis functions respectively.

3.3 MGTWR

A key advance in GWR models has been the extension from a single bandwidth to predictor-specific bandwidths, in which an individual bandwidth is estimated for each target-to-predictor variable relationship and the intercept. This "flexible bandwidth" or "multiscale" approach in GWR, which is implicit in GAM smooths, has been extended

from spatial (MGWR) to space-time analysis [45, 47] with MGTWR. The approach taken here to construct such MGTWR models was to compute a space-time distance matrix between each observation's location in time and space, and then to apply a standard MGWR (specified with a Gaussian kernel). It is possible to specify a scaling factor (λ) in the space-time distance matrix to control the influence of time relative to space. Here this was simply set to 1 as a modeling choice because the data are simulated over hypothetical locations and hypothetical time periods, rather than implying the absence of anisotropy between space and time. In real-world applications, where spatial and temporal coordinates are physical units with dependence structures, the choice of λ is important and should be justified by properties of the study process or inferred from the data. In this simulation framework, however, λ was fixed to avoid introducing additional tuning parameters that are not central to the approach being proposed by this study. The resulting MGTWR bandwidths arising describe the space-time distance if they are fixed, or the number of nearest neighbors if they are adaptive, in a 3-dimensional space-time distance array, within which observations are selected for inclusion in each local regression. Here, fixed space-time kernel bandwidths were optimized for each MGTWR model across the 100 simulated datasets. For context in the resultant bandwidths found for MGTWR, the distances in the space-time distance matrix of the simulated data (a 3-dimensional array of $15 \times 15 \times 15$) are summarized in Table 1.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	7.14	9.95	9.9	12.65	24.25

Table 1: Summary of the distances in the space-time distance matrix used by MGTWR.

3.4 Analysis and evaluation

The analysis was undertaken as follows: for each of the 100 simulated datasets containing `y`, `x1`, `x2`, and `x3`, 1080 GAMs were constructed, each with the predictor variables specified in different ways. From these, the best GAM was selected (i.e., the one with the lowest AIC score). In theory, they should all have had the form specified as in Equation 6. For comparison with each of the 100 selected GAMs, a corresponding MGTWR was constructed from the same simulated data, using the space-time distance matrix. The space-time varying coefficient estimates were extracted from the GAM and GWR-based models and evaluated against the simulated (true) coefficients using standard measures of accuracy and fit: MAE, RMSE, and R^2. These measures were similarly used to compare y with $\hat{y}$. The two model forms were also compared by their AIC values.

A second section of analysis probed the GAMs in more detail, in order to understand the effects of parameter misspecification. Here the results of each of the 1080 models for each of the 100 simulated datasets are evaluated in terms of the predictive performance of the different smooths and the cost of the misspecification of individual predictor variables in any model. This was done through structured model comparisons using model AIC values and treating the 1080 specifications $\times$ 100 datasets as a factorial grid and undertaking an ANOVA.

4 Simulated study results

4.1 Comparisons of chosen GAM with MGTWR

In terms of AIC, the chosen GAM provided the lowest AIC in comparison to MGTWR on only 9 out of the 100 occasions—suggesting the MGTWR tended to provide the most parsimonious model. The other model performance measures are summarized in Figure 4 for the 100 GAMs and 100 MGTWR models. Examining each row in Figure 4, there are some clear trends, although the performance differences between the two models are marginal:

- MGTWR models are more accurate in predicting the target variable ($\hat{y}$) (with lower MAE and RMSE, higher R^2 values).
- GAMs are only slightly more accurate at estimating the true intercept, β_0 (with lower MAE and RMSE values).
- GAMs are more accurate at estimating the true coefficients, β_1, β_2, and β_3 (with lower MAE and RMSE, higher R^2 values).

Overall, the two modeling approaches yield only marginal differences in standard fit measures. However, a clearer distinction emerges if coefficient recovery is separated from predictive performance. The GAMs more accurately recover the true coefficients (β_1, β_2, β_3), whereas the MGTWR models provide more accurate predictions of y and lower AIC. These are not contradictory findings but reflect two different statistical objectives. Accurate parameter recovery refers to how closely the estimated coefficients ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$) approximate the true data-generating parameters. This concerns estimation properties such as bias, variance, and the standard errors of the coefficients. Predictive performance, by contrast, concerns how closely predicted values $\hat{y}$ match observed outcomes y. The intercept estimates $\hat{\beta}_0$ are also interesting, where MGTWR can produce relatively high inaccuracies, while the GAM's intercept estimation errors appear more stable. As emphasized in statistical literature, explanation (recovering true effects) and prediction (forecasting outcomes) are related but fundamentally distinct goals [42]. The difference observed here can be understood through the bias–variance decomposition of expected prediction error, which equals $bias^2 + variance + noise$ [20]. A model may exhibit low bias in coefficient estimation yet generate predictions with higher variance. In such cases, small increases in prediction variance can inflate RMSE, MAE, and related fit statistics for predictions of y, even when the underlying coefficients are well estimated. Because prediction error also contains irreducible stochastic noise, minimizing coefficient error does not necessarily minimize predictive loss [28].

Thus, the superior recovery of the true coefficient surfaces by the GAMs suggests stronger structural identification of the data-generating process, while the superior predictive accuracy of the MGTWR models not only reflects improved in-sample performance (as demonstrated here) but intuitively, out-of-sample forecasting performance also (not demonstrated). These results are consistent—particularly in finite samples—and illustrate the broader methodological distinction between process explanation and process prediction. These results may also reflect calibration of regression parameters at each location in MGTWR and its ability to capture sharp space-time heterogeneities and local variations in coefficient estimates, but also a greater risk of overfitting and instability. Whereas the GAMs estimate a single globally penalized smooth space-time surface, with smoothing

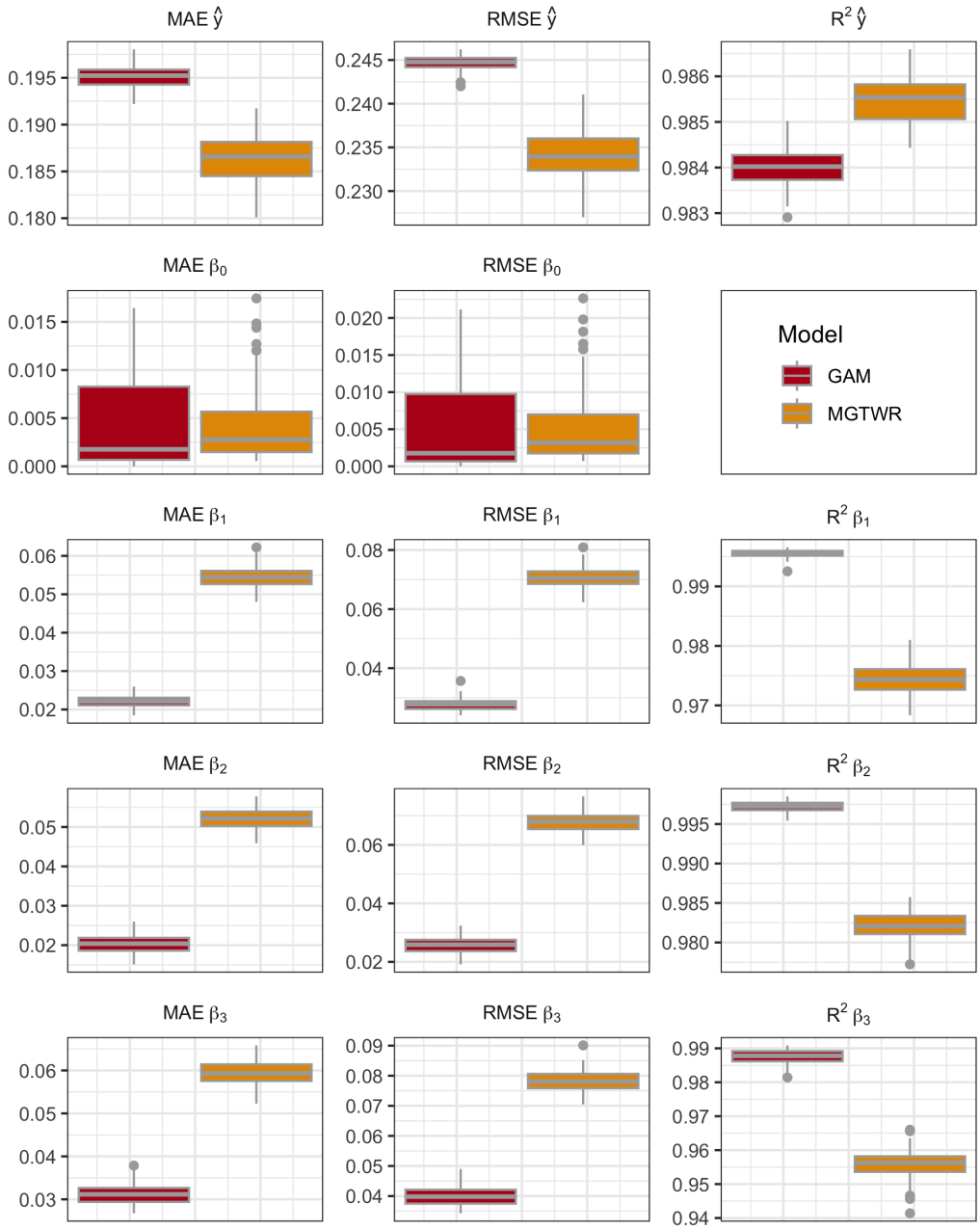

Figure 4: Distributions of model diagnostics (MAE, RMSE and R2) for predicting the true response y and estimating the true coefficients (β_0, β_1, β_2 and β_3) for the 100 chosen GAMs and 100 MGTWR models.

parameters selected to balance fit and complexity. Differences between the two modeling approaches may reflect these contrasting levels of local flexibility rather than fundamental differences in the representation of space-time structure.

It is instructive to consider the many space-time GAM forms. Recall that for each of the 100 simulated datasets, 1080 potential model forms were evaluated, each specifying each predictor variable in a different way, and the best GAM, by AIC, was chosen for the comparative analysis with the MGTWR model. Table 2 summarizes the model forms that were selected, and shows that each of the 100 models reflected the nature of the true coefficient surfaces, with β_0 mostly specified as a constant term, β_1 and β_3 with space-time interactions, and β_2 with no space-time interactions. This confirms that the space-time GAM should be specified as in Equation 6: with a fixed parametric intercept to estimate β_0 (see Equation 1), combined space-time smooths for `x1` and `x3` to estimate β_1 and β_3 (Equations 2 and 4), and a separate space-time smooth for `x2` to estimate β_2 (Equation 3).

	β_0	β_1	β_2	β_3
Omitted	0	0	0	0
Parametric form	51	0	0	0
Spatial smooth	17	0	0	0
Temporal smooth	17	0	0	0
Separate space-time smooths	7	0	100	0
Combined space-time smooth	8	100	0	100

Table 2: The different forms of each predictor variable in the best space-time GAM selected for each of the 100 simulated datasets.

In a similar way the MGTWR space-time kernel bandwidths can be investigated for the degree of space-time heterogeneity they capture in the target-to-predictor variable relationships in the simulated data. Recall that the MGTWR analyses optimized fixed space-time kernel bandwidths, and the space-time distance matrix between each observation was specified with a scaling factor ($\lambda = 1$) to control influence of time relative to space. The bandwidths are summarized in Table 3 and indicate a median space-time distance around 24 for β_0 with a large number of bandwidths lower than that, and bandwidths typically around 2.8, 2.9, and 2.7 for β_1, β_2, and β_3 respectively, with similar spreads around the median (interquartile ranges around 0.1). Comparing these with the distribution of space-time distances in Table 1 suggests that the space-time bandwidths for the intercept are typically global (i.e., nearly all observations are used in each local model, for most of the 100 MGTWR models), which is expected as this was simulated to be invariant to space-time effects, while the bandwidths for each of the predictor variables (`x1`, `x2`, and `x3`) are highly local, with a median of around 70 observations typically used in each local model.

4.2 Quantifying the impact of GAM smooth mis-specification

It is possible to unpick the predictive performance of the different GAM smooths. Figure 5 shows the distribution of the model AIC values when each predictor variable and the intercept are specified in different ways in the GAM. As might be expected given the nature of the space-time interactions present in the true coefficient surfaces, the models with the lowest AIC values specified `x1` and `x3` in combined space-time TP smooths and `x2` in

	β_0	β_1	β_2	β_3
Min.	5.13	2.63	2.67	2.55
1st Q	11.89	2.76	2.84	2.64
Median	23.89	2.81	2.90	2.69
Mean	18.49	2.81	2.90	2.70
3rd Q	24.25	2.86	2.98	2.75
Max.	24.25	3.05	3.12	2.92

Table 3: Summary of the MGTWR bandwidths for the GAMs selected for each of the 100 simulated datasets.

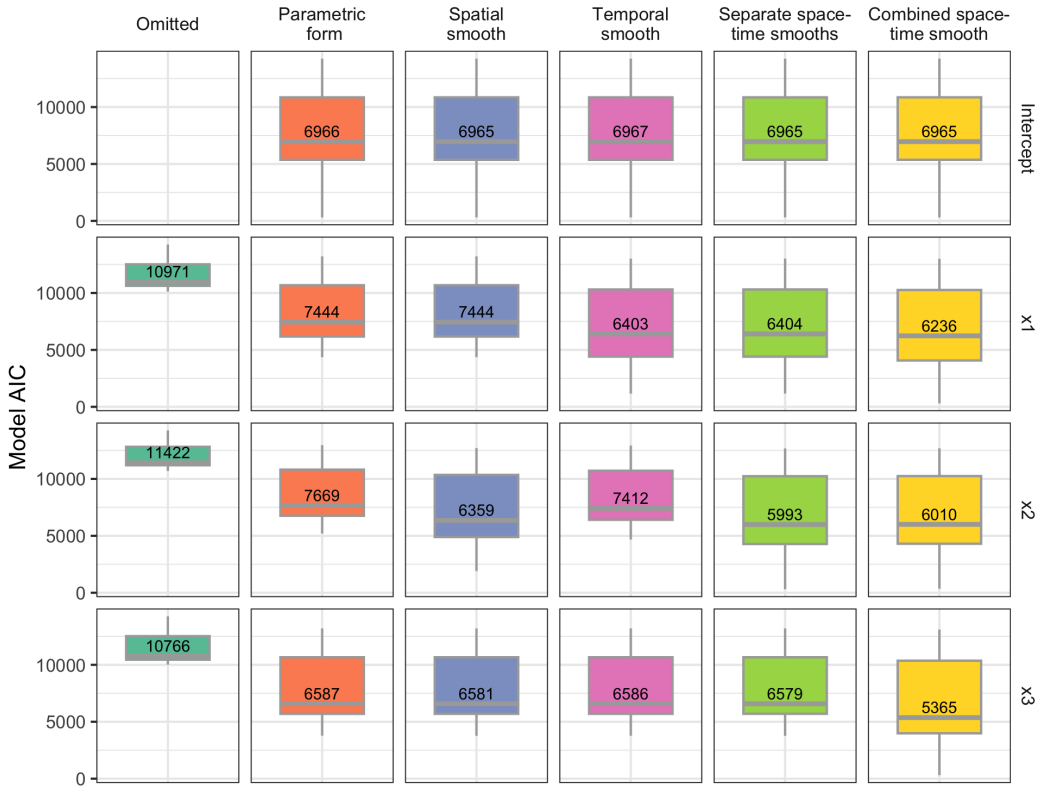

Figure 5: Distributions of the GAM AIC values, when each predictor variable is included in different forms, with the median values printed.

separate space and time smooths. It did not really matter how the intercept was specified. This potentially explains the distribution of forms for the intercept in Table 2.

It is also possible to systematically quantify the effect of smooth misspecification, through model AIC values. In this case there is a full factorial grid (1080 specifications $\times$ 100 datasets) from the comparison of GAM and MGTWR models, above. This was used

to undertake a structured model comparison, rather than, for example, treating the 1080 GAMs as unrelated.

Specifically, by retaining the way each predictor variable and the intercept were specified in the model and the model AIC, this factorial grid can be used to undertake an ANOVA. The basic approach was to, for each group i of the 100 datasets, compute the difference between the AIC of each GAM and the AIC of best GAM for that group, ΔAIC. This removes dataset-specific scale differences and the resulting values indicate how much worse each specification is when compared to the final (best) GAM chosen. Then a main effects ANOVA decomposition of ΔAIC was specified as follows:

$$Y_i = \beta_0 + \sum_{j=2}^{5} \beta_j^{(I)} \mathbf{1}(I_i = j) + \sum_{k=2}^{6} \beta_k^{(1)} \mathbf{1}(X_{1i} = k) + \sum_{m=2}^{6} \beta_m^{(2)} \mathbf{1}(X_{2i} = m) + \sum_{n=2}^{6} \beta_n^{(3)} \mathbf{1}(X_{3i} = n) + \varepsilon_i, \tag{9}$$

where $Y_i = \Delta\text{AIC}_i$, $I_i \in \{1, 2, 3, 4, 5\}$ is the level of the intercept factor for observation i, $X_{1i} \in \{1, 2, 3, 4, 5, 6\}$ is the level of factor `x1`, $X_{2i} \in \{1, 2, 3, 4, 5, 6\}$ is the level of factor `x2`, $X_{3i} \in \{1, 2, 3, 4, 5, 6\}$ is the level of factor `x3`, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. For `x1`, `x2`, and `x3` levels 1 to 6 indicate that the predictor variable was omitted, included as a parametric term, a spatial smooth (`s(S)`), a temporal smooth (`s(t)`), a separate space-time smooth (`s(S) + s(T)`), or as combined space-time smooths (`te(st)`), respectively. For the intercept, factor levels 1 to 5 relate to the last 5 of these.

A summary of this regression model is shown in Table 4. This indicates which GAM specification choices matter most and the relative magnitude of their contribution. The reference level specified a fixed, parametric intercept and the omission of `x1`, `x2`, and `x3` where the expected ΔAIC is 18,277.4. This is the baseline "worst" specification. The coefficient estimates indicate how much that specification reduces ΔAIC relative to the baseline specification. So for example `x1_te(ST)` lowers ΔAIC by about 4841 compared to if `x1` is omitted. This is a large systematic improvement in GAM predictive power.

Examining each predictor variable in turn:

- All intercept coefficients are small and non-significant. This means that the intercept specification has no systematic effect on GAM fit: it doesn't matter how it is parameterized.
- For `x1` there are large effects. If `x1` is specified in parametric form or in a spatial smooth (`s(S)`) it results in ~3400 decrease in ΔAIC, in a temporal smooth (`s(T)`) or separate space-time smooths (`s(S) + s(T)`) it lowers ΔAIC by ~4606, and if it is in a combined space-time smooth (`te(st)`) ΔAIC is lowered by ~4841. This indicates that the GAM is very sensitive to how `x1` is specified.
- For `x2` there are even stronger effects. The best result is when `x2` is specified in separate space-time smooths (`s(S) + s(T)`) (-5461 in ΔAIC). It has almost the same effect when specified in a combined space-time smooth (`te(st)`), which reduces ΔAIC by 5439. This indicates that the specification of `x2` matters even more than `x1`.
- The specification of `x3` has a large effect, with the greatest effect if it is specified in a combined space-time smooth, reducing ΔAIC by 5104, with other forms reducing it by ~3768. This indicates that only one form is clearly optimal for this predictor variable.

The regression model R^2 is high (0.905) due to the construction of a full factorial design and suggests that AIC differences are largely systematic across different GAM specifications. The residual standard error was 961. The ANOVA decomposition of ΔAIC, via an additive (main-effects only) regression model, indicated that specification of `x2` has the largest systematic impact on GAM fit, followed by `x1` and `x3`, while the intercept specification has a negligible influence.

	Estimate	Std. Error	t-value	p-value
(Intercept)	18277.433	13.077	1397.643	0.000
Intercept_s(S)	0.754	9.247	0.082	0.935
Intercept_s(S) + s(T)	1.364	9.247	0.148	0.883
Intercept_s(T)	0.614	9.247	0.066	0.947
Intercept_te(ST)	2.004	9.247	0.217	0.828
x1_Parametric	-3396.500	10.130	-335.303	0.000
x1_s(S)	-3397.763	10.130	-335.427	0.000
x1_s(S) + s(T)	-4606.123	10.130	-454.717	0.000
x1_s(T)	-4606.201	10.130	-454.725	0.000
x1_te(ST)	-4841.045	10.130	-477.908	0.000
x2_Parametric	-3575.259	10.130	-352.950	0.000
x2_s(S)	-4981.511	10.130	-491.775	0.000
x2_s(S) + s(T)	-5461.215	10.130	-539.132	0.000
x2_s(T)	-3832.361	10.130	-378.331	0.000
x2_te(ST)	-5439.679	10.130	-537.006	0.000
x3_Parametric	-3767.097	10.130	-371.888	0.000
x3_s(S)	-3769.890	10.130	-372.164	0.000
x3_s(S) + s(T)	-3770.383	10.130	-372.213	0.000
x3_s(T)	-3767.527	10.130	-371.931	0.000
x3_te(ST)	-5104.894	10.130	-503.956	0.000

Table 4: A summary of the regression model of the effect each term has on mean changes in AIC when specified in different ways (R^2 = 0.905), where '_Parametric' indicates that a parametric form was specified, '_s(S)' a spatial smooth, '_s(T)' a temporal smooth, '_s(S) + _s(T)' separate space-time smooths, and '_te(ST)' a combined space-time smooth in the GAM.

However, this is an additive model: it assumes that, for example, the penalty for misspecifying `x1` does not depend on how `x2` is specified. This might not be true as interaction between smooths is very plausible in GAMs. To expand this analysis a bit further a second-order factorial model was constructed. It is a three-factor fixed-effects ANOVA including all two-way interactions (but no three-way interactions), to allow the effect of one factor to depend on the level of another.

Let $Y_i = \Delta\text{AIC}_i$, $X_{1i}, X_{2i}, X_{3i} =$ factor levels for x_1, x_2, x_3, with baseline (reference) level for each factor equal to 1, then:

$$\begin{aligned} Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k \\ + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk}, \end{aligned} \tag{10}$$

where μ is the grand mean, α_i is the main effect of `x1` (level i), β_j is the main effect of `x2` (level j), and γ_k is the main effect of `x3` (level k). It includes all two-way interaction terms $(\alpha\beta)_{ij}$ etc. and $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$. This revised model estimates how ΔAIC (distance from

the best GAM specification) varies across all two-factor combinations of how the predictor variables are included in the GAMs.

The full summary of the resultant regression model is included in Table A1 in the Appendix. It has a very high R^2 (0.9927) indicating that almost all variation in ΔAIC is systematic and explained by the combination of `x1`, `x2`, and `x3` specification choices. This is not surprising due to the low effect of the intercept specification as described earlier. The residual standard error has reduced from ~961 in the additive model earlier to ~266, which is a very large improvement and strongly suggests that the additive model was severely misspecified.

The regression model summary (Table A1) indicates the magnitude of both individual and interacting effects when compared to the reference level, which again has a fixed, parametric intercept plus the omission of `x1`, `x2`, and `x3` with a ΔAIC of 12,118.4. Here some relatively large main (individual) effects are present with, for example, the specification of `x1` in a combined space-time smooth (`te(ST)`) associated with a decrease in ΔAIC of ~1,090, *if* `x2` and `x3` stay at their reference levels (i.e., are omitted). The point being that because of the inclusion of interactions, the main effects are conditional on other factors being at reference levels, rather than overall effects as with the additive model, above.

Instead, the largest effects are seen in the interactions. Specifically, `x2_s(S) + s(T):x3_te(ST)` is associated with a 5,271 decrease in ΔAIC, `x1_te(ST):x3_te(ST)` with a 4,865 decrease in ΔAIC, and `x1_te(ST):x2_s(S) + s(T)` with a 4,691 decrease in ΔAIC. Interestingly, the individual forms of the variables in these interactions reflect the true coefficient surfaces. The interaction effects are much larger than the main effects and clearly indicate that the effect of one GAM choice depends very strongly on the other GAM choices, and that the performance impact of how `x1` is specified depends heavily on how `x2` and `x3` are specified. This is clearly non-additivity and indicates that certain combinations of GAM smooths interact synergistically to improve model fit or antagonistically to worsen it. This reinforces the importance of searching through the model space, rather than assuming model form.

5 Empirical case study using Chaco dry rainforest data

5.1 Data

A remote sensing dataset for the Chaco dry rainforest in South America (Figure 6) is used to illustrate the approach to GAM specification. It has 271,188 observations covering an approximately 200 km by 200 km area and records NDVI, which is coupled with climate temperature and precipitation data, with monthly observations from January 1982 to June 2022. The NDVI and climate data were extracted from Google Earth Engine [22]. The NDVI data is sourced from the PKU GIMMS NDVI v1.2 dataset, with observations at 1/12° spatial resolution [33]. The climate variables, maximum temperature (`tmax`) in degrees C and precipitation (`pr`) in millimeters, were extracted from the TerraClimate dataset (IDAHO_EPSCOR/TERRACLIMATE) and their monthly means were calculated. The data were projected to the SIRGAS 2000 / Brazil Mercator projection (EPSG: 5641).

The data covers the western Chaco region, which is drier than the eastern part with stronger seasonality. The western region has larger NDVI changes over the year and sharper green-up and senescence in the spring and autumn. The annual pattern is that NDVI is typically low in late dry season (August–September), followed by a rapid green-up

Figure 6: The Chaco dataset locations, with a small transparency term and an OpenStreetMap backdrop (map data © OpenStreetMap contributors, https://www.openstreetmap.org/copyright, available under the Open Data Commons Open Database License (ODbL)).

after the rains (October–November), with peak NDVI in the mid–late wet season (January–February), and then a gradual decline as soils dry (March–May). This suggests that although there may be an east-west gradient, the spatial pattern of NDVI within the case study area would not be expected to change over time (i.e., the annual pattern of changes in NDVI would not change over time). Thus, in this case any space-time effects would be expected to be independent of each other, indicating that a GAM incorporating separate space-time smooths of the following form would be appropriate:

$$\begin{aligned} \text{ndvi}_i = \beta_0 \cdot \text{Intercept}_i + f_1(u_i, v_i, t_i) \cdot \text{Intercept}_i + \\ f_2(u_i, v_i) \cdot \text{tmax}_i + f_3(t_i) \cdot \text{tmax}_i + \\ f_4(u_i, v_i) \cdot \text{pr}_i + f_5(t_i) \cdot \text{pr}_i + \epsilon_i \end{aligned} \tag{11}$$

where ndvi_i is the target variable, β_0 is the global, fixed intercept coefficient estimate, $f_1(u, v, t)$ is the TP space-time smooth of spatial and temporal coordinates, modulated by `Intercept`, $f_2(u, v), f_4(u, v)$ are the spatial smooths modulated by `tmax` (maximum temperature) and `pr` (precipitation) respectively, $f_3(t), f_5(t)$ are the temporal smooths modulated by `tmax` and `pr`, with u, v the location coordinates, t the time variable (month of observation, from 1 to 486), and ϵ_i is the residual error term.

5.2 Model specification

The process of model selection for the space-time varying coefficient GAM was undertaken in the same way as with the simulated data, above. Here 180 GAMs were created and evaluated. Each of these had the intercept and the predictor variables specified in different ways. The 5 best space-time GAMs (i.e., with the lowest AIC) are shown in Table 5. Interestingly the top-ranked model is specified with separate space-time smooths for the intercept and combined space-time smooths for the predictor variables—a different model specification from that in Equation 11—with the next ranked model specifying a separate space-time smooth for precipitation (`pr`). All of the top 5 ranked models contain combinations of combined (`te(ST)`) and separate (`s(S) + s(T)`) space-time smooths.

Rank	Intercept	tmax	pr	AIC
1	s(S) + s(T)	te(ST)	te(ST)	3202012
2	s(S) + s(T)	te(ST)	s(S) + s(T)	3202090
3	te(ST)	s(S) + s(T)	s(S) + s(T)	3202225
4	te(ST)	s(S) + s(T)	te(ST)	3202441
5	te(ST)	te(ST)	s(S) + s(T)	3202865

Table 5: The 5 best ranked space-time varying coefficient GAMs for the Chaco dry rainforest case study, showing how each predictor variable (tmax and pr) and the intercept were specified within each model, where 's(S) + s(T)' indicates separate space-time smooths, and 'te(ST)' a combined space-time tensor product smooth.

The effects of misspecifying the GAM terms can be quantified as before, by calculating the difference between the AIC of each GAM and the AIC of the best GAM, ΔAIC, and undertaking a two-factor factorial ANOVA with interaction of the 180 GAMs. This models how model fit varies depending on how the predictor variables are specified. Here, the target variable is again the difference from the best AIC and the predictor variables are how maximum temperature (`tmax`) and precipitation (`pr`) are specified in the GAM. The results of doing this are shown in Table 6.

The baseline GAM only specifies an intercept (i.e., `tmax` and `pr` are omitted). In this case the GAM performs ~105,281 AIC units worse than the best GAM. The main effects (i.e., when only one predictor variable is changed) show that if `pr` is omitted (`tmax_te(ST) = -12336`) then specifying `tmax` as a TP smooth reduces AIC by ~12,336. However, most `tmax` main effects are not statistically significant, which indicates that changing only `tmax` (while `pr` is omitted) does not dramatically improve GAM fit. The effect of changing `pr` is much greater, with all `pr` terms indicating a ΔAIC in the range ~48,000 to ~50,000 and highly significant (e.g., `pr_te(ST) = -58851`). This indicates that the specification of `pr` is more important than the specification of `tmax`. The interaction terms are more important than the main effects. All of the interaction terms are large (~30,000) and highly significant (e.g., `tmax_te(ST):pr_te(ST) = -32005`). This indicates that there is strong interaction between how `tmax` and `pr` are specified and that GAM performance depends on their joint specification. In this case it is evident that the specification of `pr` in the GAM is driving most of the differences in model AIC, with `tmax` having weaker independent effects. It is the combination of how the predictor variables are specified that is most important, and spatial and/or temporal smooths (especially TP smooths) tend to reduce AIC substantially. This again indicates that the choice of GAM

Rank	Term	Estimate	Std. Error	t-value	p-value
36	(Intercept)	105281.0	2888.9	36.4	0.000
35	tmax_Parametric	-245.1	4085.6	-0.1	0.952
34	tmax_s(S)	-4846.3	4085.6	-1.2	0.237
32	tmax_s(S) + s(T)	-11115.8	4085.6	-2.7	0.007
33	tmax_s(T)	-6410.0	4085.6	-1.6	0.119
31	tmax_te(ST)	-12336.5	4085.6	-3.0	0.003
5	pr_Parametric	-48725.2	4085.6	-11.9	0.000
3	pr_s(S)	-55528.8	4085.6	-13.6	0.000
2	pr_s(S) + s(T)	-58135.9	4085.6	-14.2	0.000
4	pr_s(T)	-51271.7	4085.6	-12.5	0.000
1	pr_te(ST)	-58851.2	4085.6	-14.4	0.000
23	tmax_Parametric:pr_Parametric	-31098.5	5777.9	-5.4	0.000
8	tmax_s(S):pr_Parametric	-34511.1	5777.9	-6.0	0.000
25	tmax_s(S) + s(T):pr_Parametric	-30977.6	5777.9	-5.4	0.000
30	tmax_s(T):pr_Parametric	-27622.7	5777.9	-4.8	0.000
10	tmax_te(ST):pr_Parametric	-33710.0	5777.9	-5.8	0.000
13	tmax_Parametric:pr_s(S)	-33313.0	5777.9	-5.8	0.000
20	tmax_s(S):pr_s(S)	-32002.0	5777.9	-5.5	0.000
29	tmax_s(S) + s(T):pr_s(S)	-28497.4	5777.9	-4.9	0.000
28	tmax_s(T):pr_s(S)	-29870.5	5777.9	-5.2	0.000
26	tmax_te(ST):pr_s(S)	-30730.2	5777.9	-5.3	0.000
12	tmax_Parametric:pr_s(S) + s(T)	-33348.0	5777.9	-5.8	0.000
18	tmax_s(S):pr_s(S) + s(T)	-32085.5	5777.9	-5.6	0.000
22	tmax_s(S) + s(T):pr_s(S) + s(T)	-31130.8	5777.9	-5.4	0.000
17	tmax_s(T):pr_s(S) + s(T)	-32384.4	5777.9	-5.6	0.000
14	tmax_te(ST):pr_s(S) + s(T)	-33227.7	5777.9	-5.8	0.000
24	tmax_Parametric:pr_s(T)	-31081.8	5777.9	-5.4	0.000
7	tmax_s(S):pr_s(T)	-34595.7	5777.9	-6.0	0.000
11	tmax_s(S) + s(T):pr_s(T)	-33609.2	5777.9	-5.8	0.000
27	tmax_s(T):pr_s(T)	-30006.2	5777.9	-5.2	0.000
6	tmax_te(ST):pr_s(T)	-36110.6	5777.9	-6.2	0.000
9	tmax_Parametric:pr_te(ST)	-33905.8	5777.9	-5.9	0.000
16	tmax_s(S):pr_te(ST)	-32825.8	5777.9	-5.7	0.000
21	tmax_s(S) + s(T):pr_te(ST)	-31926.7	5777.9	-5.5	0.000
15	tmax_s(T):pr_te(ST)	-32974.3	5777.9	-5.7	0.000
19	tmax_te(ST):pr_te(ST)	-32004.7	5777.9	-5.5	0.000

Table 6: A summary of the second-order factorial (regression) model (R^2 = 0.971) of the effect each term and each two-factor term has on mean changes in AIC when predictor variables are specified in different ways in the case study GAM.

smooth materially affects model performance and the interaction between predictor variables (and their specification) is crucial.

5.3 Varying relationships for the chosen GAM

Using the GAM with the lowest AIC in Table 5 its varying coefficient estimates were extracted and explored. The temporal variations of the intercept (β_0) and the 2 coefficient estimates (β_{tmax} and β_{pr}) are summarized in Figure 7. These indicate that the relationships of

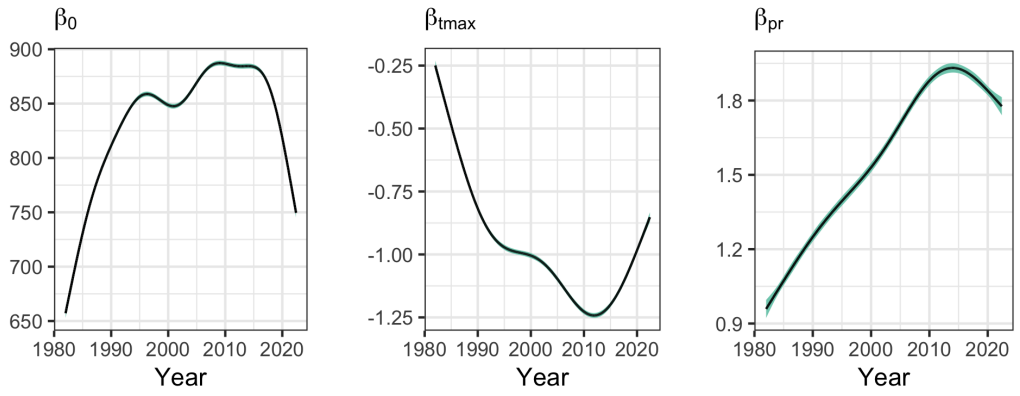

Figure 7: The temporal variation of the coefficient estimates (β_0, β_{tmax} and β_{pr}).

maximum temperature (`tmax`) and precipitation (`pr`) with NDVI change over time. Maximum temperature shows a steady and increasingly negative relationship with NDVI from 1980 to 2011 and an increase since then. By contrast the relationship of precipitation with NDVI increased from 1980 to 2016 (with a minor dip 1990 to 1996), leveled out to 2017, and decreased to 2022. The overall (i.e., all time periods) spatial variations of the 3 coefficient estimates are summarized in Figure 8. In some places these have clear spatial patterns that are potentially related (e.g., the northwestern corner) and worthy of further investigation. It is also possible to examine the spatial patterns over discrete time slices. Figure 9 shows a changing mixed relationship of the precipitation with the NDVI target variable over time, for which a combined space-time smooth was fitted.

6 Discussion

GAMs are increasingly being applied to spatial problems to create varying coefficient models [7, 9, 15] because of their theoretical and operational advantages over other commonly used approaches. GAMs are incredibly flexible. They use smooths to fit non-linear relationships between target and predictor variables. If these are parameterized with observation location, time, or location and time, then they can model target-to-predictor relationships that vary over space and/or time—i.e., moving from non-linear relationships in attribute space to non-stationary relationships over space and/or time. In this way, GAMs by default fit multiscale models. GAMs can handle different kinds of target variables (continuous, count, etc.), they have options for penalized terms to handle collinearity, can provide robust treatments of outliers, and allow for autocorrelated error terms. They also have a richer and more comprehensive theoretical background than GWR-based approaches (the current spatially varying coefficient model brand leader [7]).

This paper describes the extension of GAM-based approaches for spatially varying coefficient modeling into space-time. It proposes an approach that explicitly seeks to determine the nature of the space-time dependencies present in data relationships and to thereby inform model specification. Simply plugging and playing—making assumptions about the

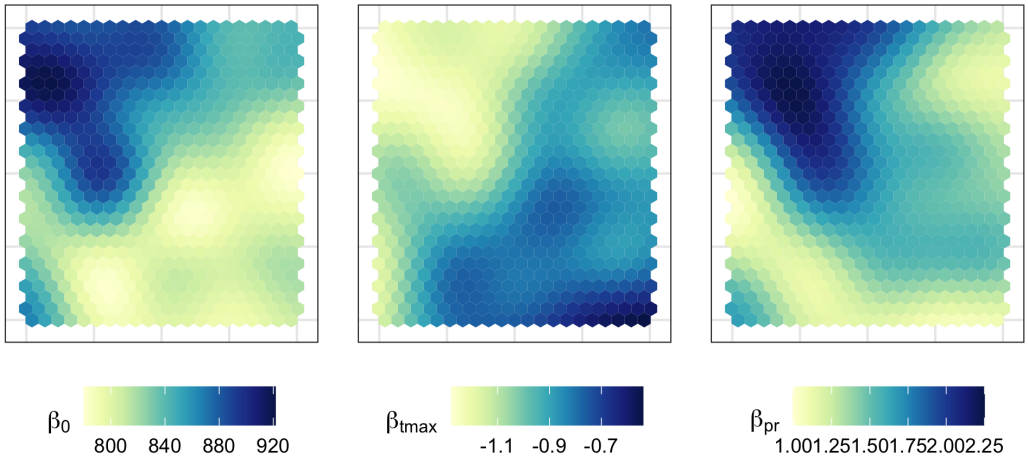

Figure 8: The spatial variation of the coefficient estimates (β_0, β_{tmax} and β_{pr}) over a gridded surface.

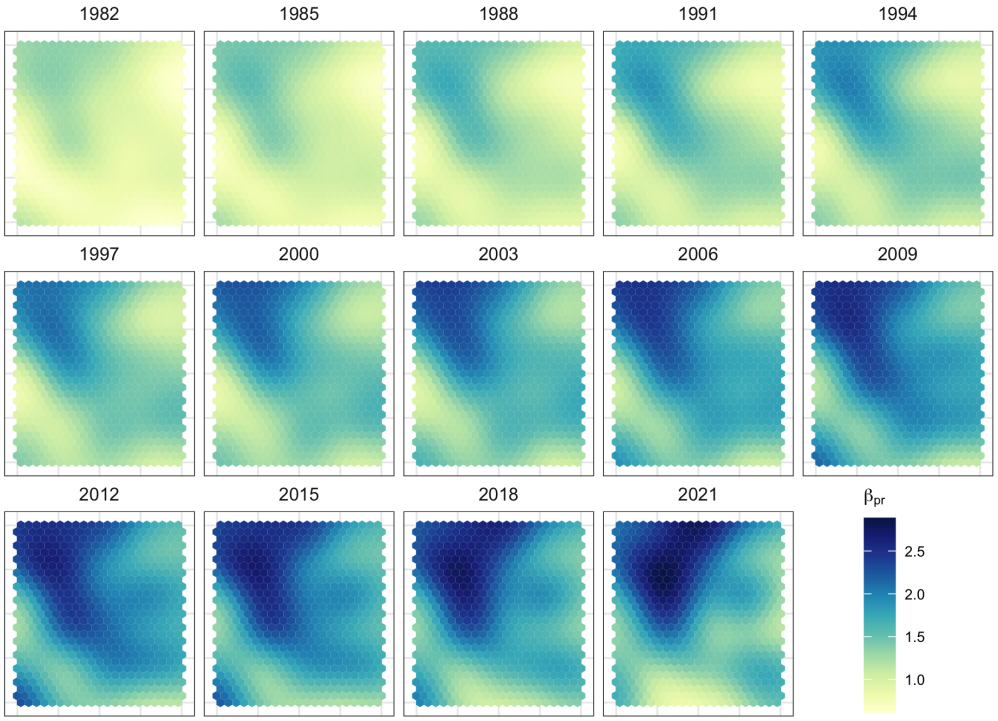

Figure 9: The spatial variation of significant precipitation coefficient estimates over time.

space-time dependencies arising from the data-generating process and constructing the most complex space-time model—risks model misspecification (as illustrated in [10]) and has a cost in terms of model predictive power and inference (as illustrated in Section 4.2 and Section 5.2). Understanding the nature of the space-time data relationships is important as data are increasingly second- or third-hand, with the user increasingly knowing less about the data recording process or the features or sensors generating the data, and therefore unsure about appropriate model structure and specification. The critical issue is how best to specify each predictor variable in the varying coefficient GAM. They can be omitted, included as a parametric term, included in a spatial smooth, included in a temporal smooth, included in separate space-time smooths, or in a combined space-time smooth (6 options). Inappropriate specification may result in incorrect inference about the target-to-predictor variable relationship. If the target-to-predictor relationships vary or interact over space and time simultaneously (i.e., the relationships have space-time dependencies), then a TP smooth should be specified as a combined space-time smooth rather than say a TPRS. Smooths such as TPRS are isotropic and assume that all of the variables included in the smooth are on the same scale, whereas TP smooths are anisotropic and can combine variables such as space and time that are measured over different scales. In essence, TP smooths allow for the unknown space-time distance metric to be captured in some manner, where early work on this problem failed to address this [6, 8].

To illustrate the importance of determining appropriate model specification, coefficient surfaces with known space-time dependencies were created, from which 100 simulated datasets containing simulated predictor and target variables were generated. These were then used to compare GAMs and MGTWR models and to quantify the impacts of misspecification. The GAM and MGTWR comparison (Figure 4) indicated that MGTWR models had better (in-sample) predictive performance while GAMs are better at coefficient recovery, supporting stronger understanding about the process being examined. Out-of-sample (forecasting) prediction performance was not assessed, where it is not clear if multiscale GWR forms, such as MGTWR, can do this given their back-fitting method of estimation, while the corresponding (single bandwidth) GTWR form can (e.g., [34]).

Through an analysis of predictor variable misspecification, the impacts on GAM fit of misspecifying individual variables was quantified (Table 4). This indicated that greatest improvements in model fit were when `x1` and `x3` were each specified in a combined space-time TP smooth (`te(ST)`) and when `x2` was specified in separate space-time smooths (`s(S) + s(T)`). It also showed that the cost of misspecifying `x2` was greater than for `x1` or `x3`: that is, which specification choices matter most. The analysis also examined the interacting misspecification costs to the GAM fit (Table A1) when predictor variables are specified in different ways. The best interactions (decreases in AIC) were seen in the interactions of `x1`, `x2`, and `x3` when they were correctly specified. Thus, although determining the most appropriate space-time GAM form adds a computational overhead to the analysis (here 1080 potential model forms were evaluated), it provides greater insight into the nature of the process being examined than MGTWR. Each of the chosen GAMs applied to the 100 simulated datasets identified the true nature of the space-time dependencies present in the data. In contrast, MGTWR models support process inference through the optimized space-time bandwidths for each predictor variable. In this case, these failed to describe the true space-time interactions in the data, generating highly localized bandwidths for each of the predictor variables (see Table 3), with little discernible difference for predictor variables with known joint and separable space-time dependencies.

Analysis of NDVI (as a proxy for forest productivity) in the South American Chaco dry rainforest case study identified the best space-time GAM to be one that included separate space-time smooths for maximum temperature (`tmax`) and precipitation (`pr`), as well as a combined space-time TP smooth for the intercept. This suggests that i) the effects of each predictor variable on NDVI are independent of each other, ii) that observations are subject to broadly the same environmental drivers at any given moment in time, and iii) that variations in the spatial distributions of those drivers do not change over time and do not interact over space and time simultaneously. However, an alternative potential explanation for this is that the case study examined a dataset that covered a relatively small geographical area (50,000 square kilometers) relative to the extent of the wider dry rainforest system (8.8 million square kilometers). A study of the full (eco-)system, where different pressures might be expected to have different effects in different regions, at different times, and across different scales, may indicate that space-time effects interact more strongly.

Reflecting on the methods and approach taken in this study, geographical and spatial analyses are frequently concerned with process understanding. The approach presented in this paper is that proposed and supported by the `stgam` R package [11]. It seeks to determine how each predictor variable should be included within the space-time varying coefficient GAM. This has computational overheads requiring for k predictor variables 5×6^k models to be evaluated. For the simulation experiment with its 3 predictor variables, this required 1080 potential models to be evaluated for each of the 100 simulated datasets with 3 predictor variables, while for the empirical rainforest study with 2 predictor variables, 180 potential models were evaluated. However, determining the most appropriate space-time GAM and how each predictor variable should be treated is an important step. For varying coefficient modeling, it avoids the misspecification costs to model predictive accuracy that arise when assumptions are made about the nature of space-time dependencies in data relationships. Many space-time varying coefficient models assume these to be simultaneous and simply fit the most complex model believed to be valid. This study's approach to determining the GAM (smooth) form was able to determine the appropriate GAM for each of the study's simulated and empirical datasets. In GWR-based approaches, the analogous specification is effectively determined by optimally found kernel bandwidth(s), as these control how much data is used in each localized weighted regression. For MGTWR, its predictor-specific bandwidths should enable inferences to be made about the spatial and temporal dependencies present in the data relationships. However, the bandwidths arising from the MGTWR fits of this study's simulated data did not capture the different space-time interactions present in the true coefficients of the simulated data: all of the bandwidths were similar for each predictor variable, for each of the 100 models (see Table 3), suggesting the same kind of space-time dependencies in each of the predictor variables. In fact, the simulated β_2 and derived `x2` were set up to have different space-time dependencies than β_1 and β_3 and `x1` and `x3`.

Future work will continue to refine the approach to develop and extend the space-time varying coefficient GAM-based approach described here and to update the functionality of the `stgam` R package. One immediate potential direction is to develop a method to estimate the scales of the target-to-predictor variable relationships. Initial work suggests that it may be possible to approximate these through measures of how quickly GAM smooth functions change over space or time, for example by determining the distance at which the smooth value changes by a threshold. This would provide an indication of the space-time dependency scale for each relationship, similar to kernel bandwidths in MGWR and

MGTWR approaches. It may also be possible to generate similar measures using the variogram range from an automatic fit of a suitable variogram model to the empirical variogram [29].

Finally, the obvious drawback of the approach presented here for determining GAM specification is the increase in the number of models to evaluate as the number of predictor variables increases. While search heuristics could be adapted to undertake this optimization, it may be more practical to undertake explorative investigations of the data, *a priori*, to inform on a simpler initial GAM specification rather than undertake the full search of all possible models as done here. Parallelization approaches can also help in this regard. However, we note that this is a dynamic field of research. As a result, it is likely that not all of the potential developments described above will be implemented. For example, some of the "future work" described in our early papers has been overtaken by developments here and by others. This is due to the ongoing development of our understanding of applying GAMs to space-time varying coefficient problems, coupled with ongoing developments in GAM functionality in general, principally in the `mgcv` R package [43].

7 Conclusions

This paper introduces a space–time extension of multiscale varying coefficient modeling, using GAMs with smooths to capture spatial and temporal interactions. Building on recent work on *spatially* varying coefficients [7], it investigates for both the presence and form of space–time data relationships by comparing alternative model specifications for each predictor variable. Different GAMs were evaluated using AIC. Simulated data with known space-time dependencies in data relationships were used to compare the GAM-based approach with MGTWR. The results were broadly similar, with MGTWR models having marginally better model fits, and the GAMs marginally better at estimating the true coefficients, supporting stronger inference about the process being examined. The GAM-based approach was then applied to a remote sensing and climate data case study for the Chaco dry rainforest in South America and distinct patterns of data relationship heterogeneity were identified with effects varying independently across space and time rather than jointly. The reasons for this are discussed and some directions for further methodological and software development in the `stgam` package [11] are described.

Acknowledgments

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). The authors would like to thank the anonymous reviewers whose suggestions and comments greatly improved the content, structure, and arguments of the final submission.

Data availability

The code and data used to undertake the analysis and to re-create all the figures in the paper are at https://figshare.com/s/94b07b0776062d03debb.

References

[1] ANSELIN, L. *Spatial econometrics: methods and models*, vol. 4. Springer Science & Business Media, 1988. doi:10.1007/978-94-015-7799-1.

[2] BRUNSDON, C., FOTHERINGHAM, A. S., AND CHARLTON, M. E. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis 28*, 4 (1996), 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x.

[3] CHUN, Y. Analyzing space–time crime incidents using eigenvector spatial filtering: an application to vehicle burglary. *Geographical Analysis 46*, 2 (2014), 165–184. doi:10.1111/gean.12034.

[4] CLIFF, A. D., AND ORD, J. K. *Spatial processes: models & applications*. Pion, London, 1981.

[5] 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 55*, 1 (2023), 155–178. doi:10.1111/gean.12316.

[6] COMBER, A., HARRIS, P., AND BRUNSDON, C. Multiscale spatially and temporally varying coefficient modelling using a geographic and temporal gaussian process gam (gtgp-gam). In *Proceedings of 12th International Conference on Geographic Information Science (GIScience 2023)* (2023), vol. 277, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, p. 22. doi:10.4230/LIPIcs.GIScience.2023.22.

[7] COMBER, A., HARRIS, P., AND BRUNSDON, C. Multiscale spatially varying coefficient modelling using a geographical gaussian process gam. *International Journal of Geographical Information Science 38*, 1 (2024), 27–47. doi:10.1080/13658816.2023.2270285.

[8] COMBER, A., HARRIS, P., AND BRUNSDON, C. How much time to include in multiscale space-time regressions? optimising predictor variable temporal lags. *AGILE: GIScience Series 6* (2025), 19. doi:10.5194/agile-giss-6-19-2025.

[9] COMBER, A., HARRIS, P., MURAKAMI, D., NAKAYA, T., TSUTSUMIDA, N., YOSHIDA, T., AND BRUNSDON, C. Encapsulating spatially varying relationships with a generalized additive model. *ISPRS International Journal of Geo-Information 13*, 12 (2024), 459. doi:10.3390/ijgi13120459.

[10] COMBER, L., HARRIS, P., AND BRUNSDON, C. A geographer's introduction to space-time regression with gams using 'stgam'. Vignette from the R package stgam, 2025. https://cran.r-project.org/web/packages/stgam/vignettes/stgam_vc_models_chaco.html. Accessed: 2025-06-02.

[11] COMBER, L., HARRIS, P., IRISARRI, G., AND BRUNSDON, C. *stgam: Spatially and Temporally Varying Coefficient Models Using Generalized Additive Models*, 2026. https://CRAN.R-project.org/package=stgam. R package version 1.2.0.

[12] DAMBON, J. A., SIGRIST, F., AND FURRER, R. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics 41* (2021), 100470. doi:10.1016/j.spasta.2020.100470.

[13] DAMBON, J. A., SIGRIST, F., AND FURRER, R. 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 (2022), 2525–2548. doi:10.1080/13658816.2022.2097684.

[14] ELHORST, J. P. Spatial panel data models. In *Spatial econometrics: From cross-sectional data to spatial panels*. Springer, 2013, pp. 37–93. doi:10.1007/978-3-642-40340-8.

[15] FAN, Y.-T., AND HUANG, H.-C. Spatially varying coefficient models using reduced-rank thin-plate splines. *Spatial Statistics 51* (2022), 100654. doi:10.1016/j.spasta.2022.100654.

[16] FENG, C. Spatial-temporal generalized additive model for modeling covid-19 mortality risk in toronto, canada. *Spatial statistics 49* (2022), 100526. doi:10.1016/j.spasta.2021.100526.

[17] FOTHERINGHAM, A. S., BRUNSDON, C., AND CHARLTON, M. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2002.

[18] FOTHERINGHAM, A. S., CRESPO, R., AND YAO, J. Geographical and temporal weighted regression (gtwr). *Geographical Analysis 47*, 4 (2015), 431–452. doi:10.1111/gean.12071.

[19] FOTHERINGHAM, A. S., YANG, W., AND KANG, W. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers 107*, 6 (2017), 1247–1265. doi:10.1080/24694452.2017.1352480.

[20] GEMAN, S., BIENENSTOCK, E., AND DOURSAT, R. Neural networks and the bias/variance dilemma. *Neural computation 4*, 1 (1992), 1–58. doi:10.1162/neco.1992.4.1.1.

[21] GOLLINI, I., LU, B., CHARLTON, M., BRUNSDON, C., HARRIS, P., ET AL. Gwmodel: An r package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software 63*, i17 (2015). doi:10.18637/jss.v063.i17.

[22] GORELICK, N., HANCHER, M., DIXON, M., ILYUSHCHENKO, S., THAU, D., AND MOORE, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment 202* (2017), 18–27. doi:10.1016/j.rse.2017.06.031.

[23] GRIFFITH, D. A. Spatial filtering. In *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization*. Springer, 2003, pp. 91–130. doi:10.1007/978-3-540-24806-4.

[24] GRIFFITH, D. A. Modeling spatio-temporal relationships: retrospect and prospect. *Journal of Geographical Systems 12* (2010), 111–123. doi:10.1007/s10109-010-0120-x.

[25] GRIFFITH, D. A. Some robustness assessments of moran eigenvector spatial filtering. *Spatial Statistics 22* (2017), 155–179. doi:10.1016/j.spasta.2017.09.001.

[26] HASTIE, T., AND TIBSHIRANI, R. Generalized additive models: some applications. *Journal of the American Statistical Association 82*, 398 (1987), 371–386. doi:10.1080/01621459.1987.10478440.

[27] HASTIE, T., AND TIBSHIRANI, R. Generalized additive models. chapman hall & crc. *Monographs on Statistics & Applied Probability. Chapman and Hall/CRC 1* (1990).

[28] HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J. H., AND FRIEDMAN, J. H. *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009. doi:10.1007/978-0-387-84858-7.

[29] HIEMSTRA, P. H., PEBESMA, E. J., TWENHÖFEL, C. J., AND HEUVELINK, G. B. Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network. *Computers & Geosciences 35*, 8 (2009), 1711–1721. doi:10.1016/j.cageo.2008.10.011.

[30] HUANG, B., WU, B., AND BARRY, M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International journal of geographical information science 24*, 3 (2010), 383–401. doi:10.1080/13658810802672469.

[31] KIM, M., AND WANG, L. Generalized spatially varying coefficient models. *Journal of Computational and Graphical Statistics 30*, 1 (2021), 1–10. doi:10.1080/10618600.2020.1754225.

[32] LESAGE, J., AND PACE, R. K. *Introduction to spatial econometrics*. Chapman and Hall/CRC, 2009. doi:10.1201/9781420064254.

[33] LI, M., CAO, S., ZHU, Z., WANG, Z., MYNENI, R. B., AND PIAO, S. Spatiotemporally consistent global dataset of the gimms normalized difference vegetation index (pku gimms ndvi) from 1982 to 2022. *Earth System Science Data 15*, 9 (2023), 4181–4203. doi:10.5194/essd-15-4181-2023.

[34] MAO, X., ZHENG, J., LU, B., WANG, R., HAN, W., AND HARRIS, P. Spatiotemporal drought forecasting in xinjiang's irrigated agriculture: Model comparison and multi-source data integration. *Journal of Hydrology 660* (2025), 133483. doi:10.1016/j.jhydrol.2025.133483.

[35] MURAKAMI, D., AND GRIFFITH, D. A. Eigenvector spatial filtering for large data sets: fixed and random effects approaches. *Geographical Analysis 51*, 1 (2019), 23–49. doi:10.1111/gean.12156.

[36] MURAKAMI, D., SUGASAWA, S., SEYA, H., AND GRIFFITH, D. A. Sub-model aggregation for scalable eigenvector spatial filtering: Application to spatially varying coefficient modeling. *Geographical Analysis 56*, 4 (2024), 768–798. doi:10.1111/gean.12393.

[37] PACE, R. K., BARRY, R., CLAPP, J. M., AND RODRIQUEZ, M. Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics 17* (1998), 15–33. doi:10.1023/A:1007799028599.

[38] PATUELLI, R., GRIFFITH, D. A., TIEFELSDORF, M., AND NIJKAMP, P. Spatial filtering and eigenvector stability: space-time models for german unemployment data. *International Regional Science Review 34*, 2 (2011), 253–280. doi:10.1177/0160017610386482.

[39] PEBESMA, E., AND BIVAND, R. *Spatial data science: With applications in R*. Chapman and Hall/CRC, 2023. doi:10.1201/9780429459016.

[40] RÜTTENAUER, T. Spatial regression models: a systematic comparison of different model specifications using monte carlo experiments. *Sociological Methods & Research 51*, 2 (2022), 728–759. doi:10.1177/0049124119882467.

[41] SALAZAR, J. E., BENAVIDES, I. F., CABRERA, C. V. P., GUZMÁN, A. I., AND SELVARAJ, J. J. 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* (2021), 101829. doi:10.1016/j.rsma.2021.101829.

[42] SHMUELI, G. To explain or to predict? *Statistical science* (2010), 289–310. doi:10.1214/10-STS330.

[43] WOOD, S. *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*, 2025. R package version 1.9-3.

[44] WOOD, S. N. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics 62*, 4 (2006), 1025–1036. doi:10.1111/j.1541-0420.2006.00574.x.

[45] WU, C., REN, F., HU, W., AND DU, Q. Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science 33*, 3 (2019), 489–511. doi:10.1080/13658816.2018.1545158.

[46] YANG, W. *An extension of geographically weighted regression with flexible bandwidths*. PhD thesis, University of St Andrews, 2014.

[47] YU, H., AND FOTHERINGHAM, A. S. On the calibration of multiscale geographically and temporally weighted regression models. *International Journal of Geographical Information Science 39*, 6 (2025), 1203–1222. doi:10.1080/13658816.2024.2440600.

[48] YU, J., DE JONG, R., AND LEE, L.-F. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics 146*, 1 (2008), 118–134. doi:10.1016/j.jeconom.2008.08.002.

[49] ZHANG, Z., LI, J., FUNG, T., YU, H., MEI, C., LEUNG, Y., AND ZHOU, Y. 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 (2021), 2262–2286. doi:10.1080/13658816.2021.1912348.

Appendix 1: Model summary

Rank	Term	Estimate	Std. Error	t-value	p-value
88	(Intercept)	12118.4	7.7	1570.7	0.000
75	x1_Parametric	311.1	9.3	33.5	0.000
76	x1_s(S)	314.5	9.3	33.8	0.000
85	x1_s(S) + s(T)	929.7	9.3	100.1	0.000
84	x1_s(T)	927.3	9.3	99.8	0.000
86	x1_te(ST)	1089.8	9.3	117.3	0.000
73	x2_Parametric	-83.4	9.3	-9.0	0.000
77	x2_s(S)	518.6	9.3	55.8	0.000
83	x2_s(S) + s(T)	805.6	9.3	86.7	0.000
74	x2_s(T)	4.0	9.3	0.4	0.663
82	x2_te(ST)	803.1	9.3	86.4	0.000
78	x3_Parametric	622.3	9.3	67.0	0.000
80	x3_s(S)	627.6	9.3	67.5	0.000
81	x3_s(S) + s(T)	627.8	9.3	67.6	0.000
79	x3_s(T)	622.5	9.3	67.0	0.000
87	x3_te(ST)	1481.5	9.3	159.4	0.000
72	x1_Parametric:x2_Parametric	-1717.6	9.7	-177.0	0.000
71	x1_s(S):x2_Parametric	-1719.6	9.7	-177.2	0.000
54	x1_s(S) + s(T):x2_Parametric	-2349.6	9.7	-242.1	0.000
55	x1_s(T):x2_Parametric	-2348.7	9.7	-242.0	0.000
53	x1_te(ST):x2_Parametric	-2464.5	9.7	-253.9	0.000
52	x1_Parametric:x2_s(S)	-2547.0	9.7	-262.4	0.000
51	x1_s(S):x2_s(S)	-2549.6	9.7	-262.7	0.000
14	x1_s(S) + s(T):x2_s(S)	-3777.0	9.7	-389.2	0.000
15	x1_s(T):x2_s(S)	-3775.8	9.7	-389.0	0.000
13	x1_te(ST):x2_s(S)	-4035.4	9.7	-415.8	0.000
42	x1_Parametric:x2_s(S) + s(T)	-2809.0	9.7	-289.4	0.000
41	x1_s(S):x2_s(S) + s(T)	-2812.3	9.7	-289.8	0.000
9	x1_s(S) + s(T):x2_s(S) + s(T)	-4335.2	9.7	-446.7	0.000
10	x1_s(T):x2_s(S) + s(T)	-4333.5	9.7	-446.5	0.000
4	x1_te(ST):x2_s(S) + s(T)	-4690.5	9.7	-483.3	0.000
70	x1_Parametric:x2_s(T)	-1866.2	9.7	-192.3	0.000
69	x1_s(S):x2_s(T)	-1868.4	9.7	-192.5	0.000
49	x1_s(S) + s(T):x2_s(T)	-2587.0	9.7	-266.5	0.000
50	x1_s(T):x2_s(T)	-2586.0	9.7	-266.4	0.000
45	x1_te(ST):x2_s(T)	-2720.4	9.7	-280.3	0.000
44	x1_Parametric:x2_te(ST)	-2796.7	9.7	-288.2	0.000
43	x1_s(S):x2_te(ST)	-2800.1	9.7	-288.5	0.000
11	x1_s(S) + s(T):x2_te(ST)	-4316.4	9.7	-444.7	0.000
12	x1_s(T):x2_te(ST)	-4314.6	9.7	-444.5	0.000
5	x1_te(ST):x2_te(ST)	-4668.7	9.7	-481.0	0.000
66	x1_Parametric:x3_Parametric	-1953.8	9.7	-201.3	0.000
64	x1_s(S):x3_Parametric	-1956.3	9.7	-201.6	0.000
38	x1_s(S) + s(T):x3_Parametric	-2854.7	9.7	-294.1	0.000
40	x1_s(T):x3_Parametric	-2853.3	9.7	-294.0	0.000
27	x1_te(ST):x3_Parametric	-3031.9	9.7	-312.4	0.000
63	x1_Parametric:x3_s(S)	-1956.4	9.7	-201.6	0.000
60	x1_s(S):x3_s(S)	-1958.9	9.7	-201.8	0.000
34	x1_s(S) + s(T):x3_s(S)	-2860.5	9.7	-294.7	0.000
36	x1_s(T):x3_s(S)	-2858.9	9.7	-294.6	0.000

Rank	Term	Estimate	Std. Error	t-value	p-value
25	x1_te(ST):x3_s(S)	-3038.1	9.7	-313.0	0.000
61	x1_Parametric:x3_s(S) + s(T)	-1956.8	9.7	-201.6	0.000
59	x1_s(S):x3_s(S) + s(T)	-1959.3	9.7	-201.9	0.000
33	x1_s(S) + s(T):x3_s(S) + s(T)	-2861.0	9.7	-294.8	0.000
35	x1_s(T):x3_s(S) + s(T)	-2859.5	9.7	-294.6	0.000
24	x1_te(ST):x3_s(S) + s(T)	-3038.6	9.7	-313.1	0.000
65	x1_Parametric:x3_s(T)	-1954.1	9.7	-201.3	0.000
62	x1_s(S):x3_s(T)	-1956.7	9.7	-201.6	0.000
37	x1_s(S) + s(T):x3_s(T)	-2855.2	9.7	-294.2	0.000
39	x1_s(T):x3_s(T)	-2853.8	9.7	-294.0	0.000
26	x1_te(ST):x3_s(T)	-3032.4	9.7	-312.4	0.000
47	x1_Parametric:x3_te(ST)	-2688.2	9.7	-277.0	0.000
46	x1_s(S):x3_te(ST)	-2692.1	9.7	-277.4	0.000
7	x1_s(S) + s(T):x3_te(ST)	-4418.6	9.7	-455.3	0.000
8	x1_s(T):x3_te(ST)	-4416.9	9.7	-455.1	0.000
3	x1_te(ST):x3_te(ST)	-4864.7	9.7	-501.2	0.000
68	x2_Parametric:x3_Parametric	-1936.0	9.7	-199.5	0.000
31	x2_s(S):x3_Parametric	-2971.1	9.7	-306.1	0.000
19	x2_s(S) + s(T):x3_Parametric	-3333.9	9.7	-343.5	0.000
57	x2_s(T):x3_Parametric	-2122.7	9.7	-218.7	0.000
23	x2_te(ST):x3_Parametric	-3322.7	9.7	-342.3	0.000
67	x2_Parametric:x3_s(S)	-1939.4	9.7	-199.8	0.000
29	x2_s(S):x3_s(S)	-2976.5	9.7	-306.7	0.000
17	x2_s(S) + s(T):x3_s(S)	-3340.4	9.7	-344.2	0.000
56	x2_s(T):x3_s(S)	-2126.3	9.7	-219.1	0.000
21	x2_te(ST):x3_s(S)	-3329.5	9.7	-343.0	0.000
67	x2_Parametric:x3_s(S) + s(T)	-1939.4	9.7	-199.8	0.000
28	x2_s(S):x3_s(S) + s(T)	-2976.9	9.7	-306.7	0.000
16	x2_s(S) + s(T):x3_s(S) + s(T)	-3341.1	9.7	-344.2	0.000
56	x2_s(T):x3_s(S) + s(T)	-2126.3	9.7	-219.1	0.000
20	x2_te(ST):x3_s(S) + s(T)	-3330.2	9.7	-343.1	0.000
68	x2_Parametric:x3_s(T)	-1936.0	9.7	-199.5	0.000
30	x2_s(S):x3_s(T)	-2971.5	9.7	-306.2	0.000
18	x2_s(S) + s(T):x3_s(T)	-3334.4	9.7	-343.6	0.000
58	x2_s(T):x3_s(T)	-2122.6	9.7	-218.7	0.000
22	x2_te(ST):x3_s(T)	-3323.3	9.7	-342.4	0.000
48	x2_Parametric:x3_te(ST)	-2600.4	9.7	-267.9	0.000
6	x2_s(S):x3_te(ST)	-4420.0	9.7	-455.4	0.000
1	x2_s(S) + s(T):x3_te(ST)	-5270.5	9.7	-543.0	0.000
32	x2_s(T):x3_te(ST)	-2892.6	9.7	-298.0	0.000
2	x2_te(ST):x3_te(ST)	-5254.3	9.7	-541.4	0.000

Table A1: A summary of the second-order factorial model of the effect each term and each two-factor term has on mean changes in AIC when predictor variables are specified in different ways (R^2 = 0.993), where '_Parametric' indicates that a parametric form was specified, '_s(S)' a spatial smooth, '_s(T)' a temporal smooth, '_s(S) + _s(T)' separate space-time smooths, and '_te(ST)' a combined space-time smooth.