

# stgam: An R package for GAM-based varying coefficient models

Lex Comber<sup>1\*</sup>, Paul Harris<sup>2\*</sup>, and Chris Brunsdon<sup>3\*</sup>

<sup>1</sup> School of Geography, University of Leeds, UK <sup>2</sup> Rothamsted Research, North Wyke, UK <sup>3</sup> National Centre for Geocomputation, Maynooth University, Ireland \* These authors contributed equally.

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Open Journals](#) ↗

## Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

We are often interested in understanding how and where statistical relationships vary over space, and how they change over time. Quantifying such *process heterogeneity* (spatial and or temporal) can be done using *varying coefficient* models. The opportunity to undertake such space-time analyses are greater due to the increased generation and availability of data that include both spatial and temporal attributes (e.g. GPS coordinates and time-stamps). The **stgam** package provides a framework for creating regression models using Generalized Additive Models (GAMs) (Hastie & Tibshirani, 1990) in which the relationships between the response (dependent) variable and individual predictor (independent) variables are allowed to vary over space, time or both. It addresses a key question of the form space-time interaction to be specified in models. Frequently randomly guessed have to be made about the form of predictor-to-response relationships until an effective model form and associated model fits are found. To address this the **stgam** package includes functions to create multiple models, each specifying different relationships between the response variable  $y$  and the predictor variables  $x_1 \dots x_n$ . Each model is evaluated using the Bayesian Information Criterion (BIC) (Schwarz, 1978) which approximates the likelihood (probability) of the model being the correct model, given the data used in the analysis. Where multiple models are highly probable, then these can be combined using a Bayesian Model Averaging approach (Brunsdon et al., 2023; Fragoso et al., 2018). Finally the **stgam** package contains functions for creating spatially and / or temporally varying regression coefficient estimates. These can be mapped in the usual way to show how where and when the relationships between the response and the predictor variables vary over space and time.

## Reference

- Brunsdon, C., Harris, P., & Comber, A. (2023). *Smarter than your average model? Bayesian model averaging as a spatial analysis tool*. 12th International Conference on Geographic Information Science.
- Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86(1), 1–28.
- Hastie, T., & Tibshirani, R. (1990). Generalized additive models. Chapman hall & CRC. *Monographs on Statistics & Applied Probability*. Chapman and Hall/CRC, 1.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.