Multiscale Spatially and Temporally Varying Coefficient Modelling Using a Geographic and Temporal Gaussian Process GAM (GTGP-GAM)

Alexis Comber[1]
School of Geography, University of Leeds, UK

Paul Harris
Sustainable Agriculture Sciences, Rothamsted Research, Harpenden, UK

Chris Brunsdon
National Centre for Geomcomputation, National University of Ireland, Maynooth, Ireland

Abstract

The paper develops a novel approach to spatially and temporally varying coefficient (STVC) modelling, using Generalised Additive Models (GAMs) with Gaussian Process (GP) splines parameterised with location and time variables - a Geographic and Temporal Gaussian Process GAM (GTGP-GAM). This was applied to a Mongolian livestock case study and different forms of GTGP splines were evaluated in which space and time were combined or treated separately. A single 3-D spline with rescaled temporal and spatial attributes resulted in the best model under an assumption that for spatial and temporal processes interact a case studies with a sufficiently large spatial extent is needed. A fully tuned model was then created and the spline smoothing parameters were shown to indicate the degree of variation in covariate spatio-temporal interactions with the target variable.

2012 ACM Subject Classification Information systems → Spatial-temporal systems

Keywords and phrases Spatial Analysis, Spatiotemproal Analysis

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.22

Category Short Paper

Funding This work was supported by the JSPS BRIDGE fellowship No. 220305.

1 Introduction

This paper describes a novel approach for spatially and temporally varying coefficient (STVC) modelling. It extends Geographical Gaussian Process GAMs (GGP-GAM) to include GP splines parameterised space and time. GGP-GAMs have been shown to be more accurate than Multiscale Geographically Weighted Regression (MGWR)[1] the effective SVC brand leader. GGP-GAMs explicitly accommodate process spatial heterogeneity and provide an alternative to assumptions of stationarity[4]. STVCs extend this to the temporal dimension.

Generalized Additive Models (GAMs) are general in that they can handle outputs with many types of distributions and not just linear relationships, polynomial or not. They are additive and because they generate multiple model terms which are added together to generate predictions. The advantages of a GAM-based approach to SVC and STVC modelling are because GAMs are flexible and able to handle different types of response[6, 2]. This is due to their additive nature which combines multiple sub-models, and the modelling of non-linear relationships using splines, the building blocks of GAMs. Splines are combination of functions (*basis functions*) which may be single or multi-dimensional, each of which is

[1] Corresponding author

12th International Conference on Geographic Information Science (GIScience 2023).
Editors: Roger Beecham, Jed A. Long, Dianna Smith, Qunshan Zhao, and Sarah Wise; Article No. 22; pp. 22:1–22:6
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

assigned a coefficient, which are combined to generate $\hat{y}$ and in this way complex relationships are modelled in GAMs. Splines parameterised with location form the basis of SVC modelling with GGP-GAMs and here this is extended to include time within the splines: the geographic and temporal Gaussian process GAM (GTGP-GAM).

The inclusion of temporal data can provide insight on the spatial process dynamics and a number of approaches that include time in spatial regressions exist. However, with the exception of GTWR, most of these are concerned with capturing autocorrelation effects, rathere than relationship heterogeneity. This paper extends GGP-GAMs to the temporal dimension. Temporal process are well described by GPs. However a key methodological consideration is *how* space and time should be analysed together. This paper uses a national case study to investigate the relative benefits of combining space and time into a single 3D GP spline against treating them separately in a 2D + 1D approach.

2 Case study

A case study of national data of livestock in Mongolia as reported in[5] was analysed. This reports livestock totals, here focussing on goats, over 341 soums (second-level administrative units) from 1991-2006 and these were considered to be a function of annual mean normalised difference vegetation index (NDVI), annual mean rainfall, the number of households working with livestock and the number of reported animal losses in the previous year. The choice of historical loss data and household working with livestock, as explanatory variables was because livestock losses have been found to play a critical role in livestock decisions and viability[3]. The environmental variables reflect the changes in biomass and their drivers over time. Figure 1 shows the spatial context of the soums and the trends over time of the variables. These indicate a steady increase in cattle numbers, which is associated with increased meat consumption (anecdotally increasingly concentrated around Ulaanbaatar). The goat losses indicate the dzud period 2001–2002 which are extreme weather events associated with deep snow, severe cold and conditions that make foraging difficult and results in livestock deaths. The number of households associated with livestock production increases and levels off as livestock management becomes more concentrated. The the median of mean monthly NDVI is relatively stable, and mean monthly precipitation shows some fluctuation.

3 Results

A key issue in space time analysis is to determine whether observation spatial and temporal variables should be handled separately or together, ie whether their covariances are separable or non-separable. One way to approach this is to construct models and compare their performance through some measure of model fit such as AIC, and prediction accuracy such as mean absolute error (MAE). Here the aim was to construct models of goat numbers. These have a classic Poisson distribution. One option is to construct Poisson regression models and another is to to transform the response variable and fit Gaussian regression models. A square root transform of the goat counts was undertaken here. Four GAM models were constructed with GP splines parameterised with location and temporal data separately and together, using normalised (z-scores) spatial and temporal data and the original spatial and temporal data. For each of these the model MAE and AIC are summarised in Table 1, with the "best" model determined from the AIC measure.

The best performing model was one which combined space and time, with normalised space and time variables. This indicates the interaction between the space and time effects in this case and their lack of independence. This is perhaps not surprising due to the large

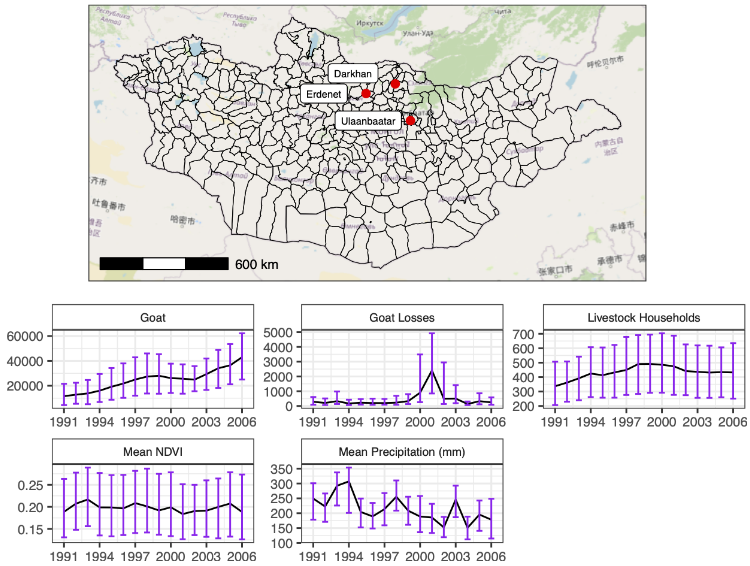

Figure 1 The Soums in Momgolia (n = 341) with the 3 largest cities (top), and the trends in the median values of the variables, with upper and lower quartiles indicated (bottom).

Table 1 Summaries of the model predictive preformance and fits, with separate and combined GP splines for locational and temporal variables, and with normalised and un-normalized data.

Splines	Normalized.Data	MAE	AIC
Separate	No	9593.17	-1800.22
Separate	Yes	9608.42	-1941.88
Combined	No	9405.98	-1812.12
Combined	Yes	9776.30	-1972.61

spatial extent of the case study and thus the spatial variation of the drivers of local goat numbers are likely to vary over space and over time. That is, different effects are more likely to be be experienced in different places at any given time and the pattern of these is more likely to change over time. Also it is important to note that although the results indicate that a better performing model is obtained by combining space and time, this is not to suggest that no distinct effects occur. Rather that the uncertainty in calibrating a more complex model with 3D splines leads to more reliable prediction.

The GTGP-GAM models were created using default parameters for the `gam` function in the `mgcv` package. The convergence of the spline smoothness optimisation of the best performing model was examined in detail, and specifically the effect fo the number of knots (k) used to construct the spline basis dimensions. Investigation indicated that k was potentially too low, with the effective degrees of freedom (EDF) for some splines close to k. The models was tuned by increasing k to 400 resulting in improved model fits (AIC) and convergence of the GP splines as the high k values ensured sufficient degrees of freedom in the splines.

The fixed parametric coefficient estimates are shown in Table 2. These show significant intercepts and generally insignificant covariates (except for mean NDVI in the Mongolian case study). The smooth terms for the combined spatial and temporal GP splines (ie the STVCs) are summarised in Table 3. The full set of coefficients are not printed because there many coefficients for each spline, one for each basis function. The `edf` (effective degrees of freedom) summarises the complexity of the spline smooths, with an `edf` value of 1 indicating

Table 2 The GTGP-GAM fixed parametric coefficients and their global significance.

Covariate	Estimate	Std. Error	t-value	p-value
Intercept	113.795	2.597	43.813	0.000
Goat Losses	0.209	0.413	0.507	0.612
Livestock Households	1.209	1.244	0.972	0.331
Mean Precipitation (mm)	-0.738	2.101	-0.351	0.725
Mean NDVI	-83.217	103.519	-0.804	0.422

Table 3 The GTGP-GAM smooth terms of the tuned model.

GTGP Spline	edf	Ref.df	F	p-value
s(X,Y,Ti):Intercept	10.688	11.515	22.090	0.000
s(X,Y,Ti):Goat Losses	206.717	248.831	2.704	0.000
s(X,Y,Ti):Livestock Households	220.063	256.766	7.805	0.000
s(X,Y,Ti):Mean Precipitation (mm)	143.082	168.939	4.311	0.000
s(X,Y,Ti):Mean NDVI	5.272	6.302	0.566	0.761

a straight line, 2 a quadratic curve etc. Higher `edf` values indicate increasing non-linearity in the relationship between the covariate and the response. The p-values relate to splines / smooths defined over these, and their significance can be interpreted as indicating whether they vary locally over space and time combined (i.e. spatio-temporally). That is, covariates with insignificant p-values (i.e. Mean NDVI) still have an effect, but these effects do not vary locally. In contrast to the fixed parametric coefficients, all are significant. That is, their relationship with the target variable y varies locally over space and serially over time with different temporal effects in different places.

It is possible to extract the spatially and temporally varying coefficient estimates. These describe how the relationship between y and the covariates varies over space and time. It is instructive to examine these alongside the GTGP-GAM smooth terms or smoothing parameters (SPs). The SPs indicate the scale of the relative spatial-temporal variation of the interaction between each covariate and the response. These and summaries of the STVCs over time are shown in Figure 2. This plots the median coefficient estimates for each year, describing how the coefficient estimates vary over time, with their variation over space summarised in their inter-quartile range (IQR). It shows that:

- The Intercept steadily increases over time and the the IQR gradually narrows towards the end of the sequence. It has a low SP indicating stable spatial relationships over time.
- The association of Goat Losses with goat numbers decreases to 2002 and then increases to 2006. However, the IQR shows high variation over time indicating, narrowing to 2002 and the increasing in later years. This has a high spline SP value, indicating a strong spatially and temporally varying relationship with the target variable.
- The relationship of Livestock Households with goat numbers steadily increases over time. This may indicate the impact of an increasing concentration of livestock within fewer households. The variation (IQR) over space also remains relative stable, with a small degree of variation, reflected in the moderate spline SP value.
- Mean NDVI and Mean Precipitation are both mostly negative in the association with goat numbers, with Mean NDVI decreasing in later years and Mean Precipitation increasing to zero and the decreasing before increasing in later years. There is more spatial variation over time in Mean Precipitation than Mean NDVI, as reflected in their SP values.

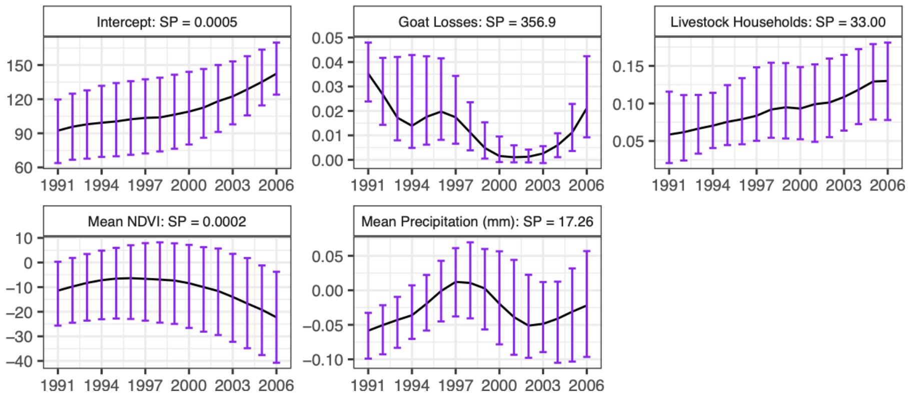

Figure 2 The temporal trends in the median values of the coefficient estimates over time (with upper and lower quartiles) indicated, with spline smoothing parameters (SP).

In summary, Figure 2 shows the temporal trend of the relationships between the covariates and the target variable. The changes in the IQRs over time indicate whether the relationship and thus the process is changing spatially as well as temporally. This interaction between space and time in the STVCs is is also reflected in the spline SP values.

It is also possible to confirm this interpretation of the spline SPs and the nature of the STVCs they indicate: broadly, larger SPs indicate greater spatio-temporal interaction of the covariate with the target variable. Figure 3 compares the coefficient estimates for the Goat Losses (high SP) and Mean NDVI (low SP). For the Goat Losses, this shows how the spatial relationship between the covariate and the target variable (Goat population) changes over time, with stronger relationship in and around the major population centres noticeable in 2006, for example. By contrast Mean NDVI has a spatially varying relationship with the target variable but this does not change over time.

4 Brief Discussion and Conclusions

The paper explores STVC modelling through the application of GAMs with GP spline parameterised with location and time variables. Here these were combined into a single 3-dimensional spline for each predictor variable, under the assumption that the case study extent was sufficiently large for an assumption of the geographic and temporal process interacting over space and time to hold. In this model the temporal trends in the relationship between the predictor variable and the target variable were allowed to vary with location. The paper demonstrates STVC modelling using GAMs with GP spline parameterised with location and time variables. Different GP spline compositions were explored to determine whether space and time should be treated separately ar assumed to interact. In this case, exploring a national case study, the best fitting model was found to be one in which space and time measurements were re-scaled to z-scores and combined in 3-dimensional GP spline. This reflected *a priori* assumption that spatial and temporal processes would interact for case studies with a sufficiently large spatial extent, an assumption that proved to be true in this case. In other situations this might not be the case. Case studies with smaller spatial extents, or indeed with shorter runs over time, may require location and time to be treated separately, with separate GP splines for each predictor variable. In this situation, the assumption would be that the spatial and temporal trends in the data and their relationship with the outcome do not interact, and that any changes in the relationship with target variable over time would

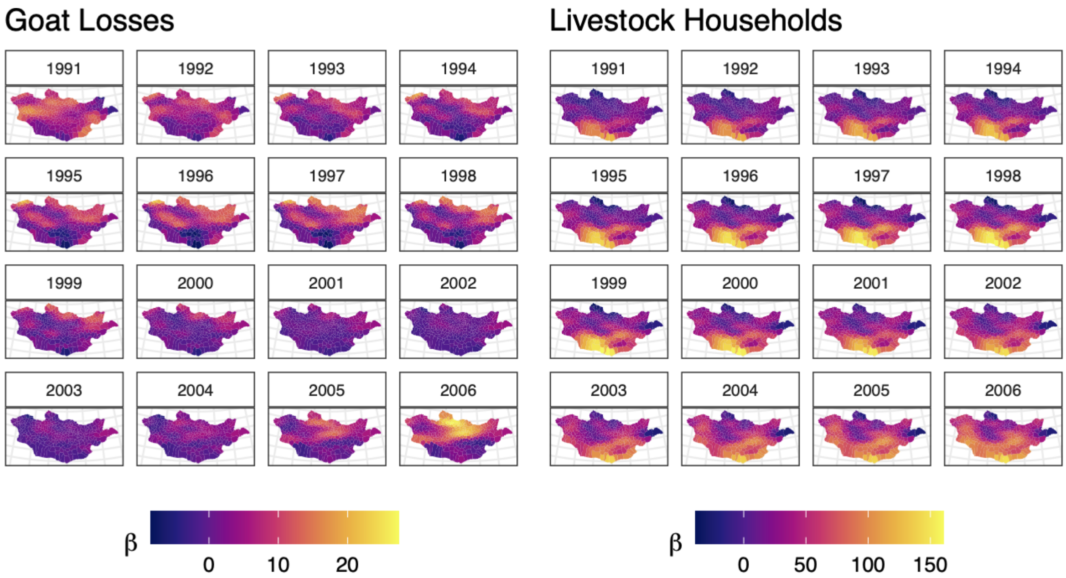

Figure 3 Examples of STVC estimates for Goat Losses and Livestock Households over a 16 year time period.

be independent of changes in location. These models exclude the possibility of different temporal trends in different locations. It also reflected an assumption that AIC as a measure of model fit and parsimony identified the "best" model. Other investigations (not reported here) has suggested that BICs (Bayesian Information Criteria) may be more appropriate for investigating and comparing space-time interacted models with independent space and time effects. Future work will explore these issues. The model was then tuned with large number of knots, allowing sufficient degrees of freedom for the model parameters. The relative values of the tuned model smoothing parameters provided an indication of the variation in the spatial and temporal interactions of the covariates with the target variable. Summaries over time of the median values of the coefficient estimates demonstrated the temporal trends, and the spatially varying nature of these was suggested by the interquartile ranges of these. These were confirmed by the smoothing parameters and through visual exploration.

References

1 Alexis Comber, Paul Harris, and Chris Brunsdon. Multiscale spatially varying coefficient modelling using a geographical gaussian process gam. *International Journal of Geographical Information Science*, submitted.

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

3 John McPeak. Confronting the risk of asset loss: What role do livestock transfers in northern kenya play? *Journal of Development Economics*, 81(2):415–437, 2006.

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

5 Narumasa Tsutsumida, Paul Harris, and Alexis Comber. The application of a geographically weighted principal component analysis for exploring twenty-three years of goat population change across mongolia. *Annals of the American Association of Geographers*, 107(5):1060–1074, 2017.

6 Simon N Wood. *Generalized additive models: an introduction with R*. Chapman Hall/CRC, 2006.