Assessing Temporal and Spatial Change in Nutrients for Large Hydrological Areas

Miller, C., Magdalina, A., Bowman, A.W., Scott, E.M. and Lee, D.
School of Mathematics and Statistics, University of Glasgow, UK
Claire.Miller@glasgow.ac.uk

Willows, R., Burgess, C., Pope, L and Johnson D.
Environment Agency, Evidence Directorate, UK

Abstract: Regulatory bodies, such as the Environment Agency of England & Wales, regularly monitor river surface water to assess quality. Maintaining and improving quality is important for society but is also a necessary requirement to comply with European directives. Spatiotemporal additive models for nutrients in hydrological areas in England & Wales are presented to assess and describe spatial and temporal trends over the past 20 to 40 years.

Keywords: spatiotemporal, smoothing, nutrients

1 Introduction

Previous modelling of nutrients within English & Welsh rivers has been carried out at individual monitoring locations in order to investigate trends over time and the effect of the contributing area at individual locations. However, for future nutrient policy decisions, there is a need to understand how historical patterns of water quality are described by catchment-scale influences rather than at individual sites. Spatiotemporal additive models have been developed for Large Hydrological Areas (LHAs) to investigate and describe nutrient trends on a catchment-wide basis.

2 Materials and Methods

2.1 The Data

There are 59 LHAs in England & Wales that contain independent river networks and their associated catchments. Monitoring locations within each LHA are associated with Water Framework Directive (WFD) waterbodies and each LHA consists of a number of such areal units. Orthophosphate (OP) and Total Oxidised Nitrogen (TON), mg/l, have been monitored on approximately a monthly basis by the Environment Agency of England & Wales over a period of 20 to 40 years at monitoring locations within each of the LHAs.

For this paper, the OP data in the Severn LHA, see Figure 1 (top left), will be investigated. These data span the time period 1971-2009 and have been aggregated within waterbodies. The OP data have been transformed using natural logs, to stabilise the variability throughout time, and measurements that were flagged as being below the limit of detection have been treated as censored observations and imputed (Helsel, D.R., 2005).

In order to investigate the effect of catchment covariates on nutrient levels, monitoring locations were selected which have no monitoring locations in further upstream waterbodies. This enables information from all local land area draining to the waterbody (contributing land) to be incorporated in the covariate value for an individual waterbody ensuring that contributed areas do not overlap and that there is little spatial correlation between measurements for a particular covariate.

The possible catchment covariates of interest are long-term (1961-1990) average base flow index (BFI) and discharge, total annual population, monthly total rainfall, fertiliser yearly application rates, land cover variables, livestock variables, crops, Agricultural Land Classification (ALC), slope of the land and soil type. All covariates have been aggregated within the 173 waterbodies of interest. Many of the continuous covariates have been log transformed to make their distributions more symmetric. Categorical variables were used for: slope (gentle to very steep), ALC (high quality agricultural to low quality grazing and non-agricultural land) and soil type (light to heavy), and for land cover, land use and population the hectares/counts have been standardised by the size of the contributing area.

2.2 Statistical Modelling

Model (1) was fitted to describe the relationships between the response of $\log_e(\mathrm{OP})$ and all possible covariates: spatial and temporal trend and seasonality and the catchment covariates. A model which excludes the first three smooth terms of Model (1) was also fitted to investigate the relationships, and the percentage of variability explained, using only the catchment covariates, Model (2).

$$\begin{aligned} y \;=\; & \alpha + s(\mathrm{Easting}, \mathrm{Northing}) + s(\mathrm{Year.month}) + s(\mathrm{month}) + s(\mathrm{discharge}) + s(\mathrm{BFI}) \\ & + s(\mathrm{land\,use}) + s(\mathrm{land\,cover}) + \beta_{\mathrm{ALC}_j} + \gamma_{\mathrm{slope}_k} + \delta_{\mathrm{soil}_l} + s(\mathrm{rainfall}) \\ & + s(\mathrm{population}) + s(\mathrm{fertiliser}) + \epsilon \end{aligned} \tag{1}$$

where y is $\log_e(\mathrm{OP})$, $s()$ is a smooth function, Year.month is decimal year and the errors (ϵ) are assumed to be $N(0, \sigma^2)$ and independent. For land use and land cover a series of different covariates are included individually such as potatoes, field vegetables, cows, etc. and the levels of the categorical variables are $j = 2, \ldots, 6$, $k = 2, \ldots, 4$, and $l = 2, \ldots, 8$. The degree of smoothing has been constrained to allow a maximum of 6 degrees of freedom for each univariate component to aid interpretation.

Functions from the `sm` library, see Bowman & Azzalini (1997) for details, and the `gam` function in the `mgcv` library, see Wood (2006) for full details, of `R` were used to fit these models. For the models that incorporate many covariates there is unlikely to be much spatiotemporal correlation remaining in the residuals and hence the assumption of independence appears appropriate. However, Moran's I Test in the `spdep` package of `R` and temporal variograms were used to check this assumption. If necessary, the methods of analysis can be modified to incorporate spatiotemporal correlation.

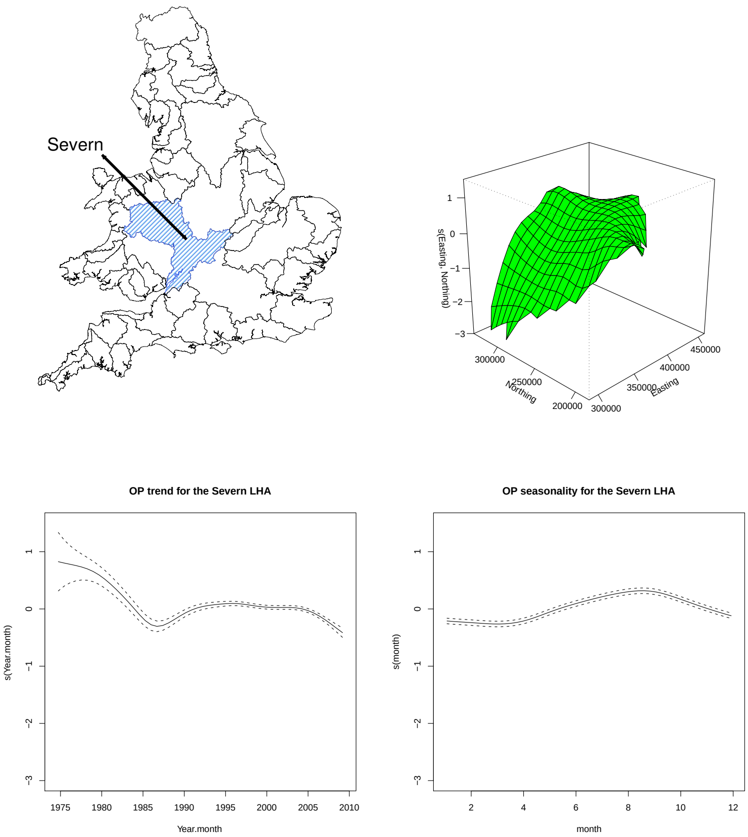

Figure 1: Map of Large Hydrological Areas in England & Wales with the Severn area highlighted (top left). For the Severn area: spatial trend (top right), temporal trend (bottom left), seasonal pattern (bottom right). The dashed lines indicate $\pm$ 2 standard errors.

3 Results

Figure 1 (top right and bottom panels) displays the spatial pattern, temporal trend and seasonality, respectively, for the Severn LHA. It highlights that the largest change for this area is spatial, followed by a smaller change over time and a small seasonal signal. Model (1) explains 70% of the variation in $\log_e(\mathrm{OP})$, see Table 1, with 66% of the variability explained using only the catchment covariates, Model (2). Therefore, the catchment covariates are usefully explaining the trends and seasonality in the area. Applying Moran's I and temporal variograms to the residuals from Model (1) suggests that there is very little evidence of spatial or temporal correlation remaining after incorporating all covariates.

In general, covariates are statistically significant as a result of the large amount of data. Many of the variables individually only explain a small proportion of the variability and hence it is difficult to reduce the number of covariates in the model. However, there are various possible combinations of a smaller subset of covariates that explain a reasonable amount of the variability. For example, reducing the number of catchment covariates from 23 to 8: ALC, soil, discharge, cattle, pigs, poultry, "other animals" and cumulative rainfall, still explains 46% of the variability. For this set of variables, relationships with discharge and rainfall appear to be curved with a decreasing relationship evident at higher values indicating a dilution effect. Relationships with the animal covariates are generally positive, especially for larger counts, indicating increases in measured OP with increasing animal waste. Higher grade agricultural land appears to contribute more to OP levels than lower quality grazing land with medium/heavy soils contributing more than chalk or light soils.

Model	Adjusted R^2	Number of covariates
1	70.1%	26
2	65.7%	23

Table 1: Adjusted R^2 values for the Severn LHA

4 Concluding Remarks

Trends and seasonality have been explored in all LHAs in England & Wales for OP, TON and Total Nitrogen with covariate information incorporated for a subset of these LHAs and a space-time interaction incorporated using p-splines for one example area. Future work will include exploring alternative approaches to dealing with large data dimensions and the hierarchical nature of the data.

References

Bowman, A.W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. OUP, Oxford.

Helsel, D.R. (2005). *Nondetects and data analysis: statistics for censored environmental data*. New York: Wiley-Interscience.

Wood, S.N. (2006). *Generalized Additive Models, An Introduction with R*. Chapman & Hall, Boca Raton.