SECURE project final report – June 2016

Statistical software to identify spatiotemporal patterns and coherence over river networks

A collaborative project between Statistics at the University of Glasgow, the Environment Agency and the Scottish Environment Protection Agency.

1st November 2015 to 30th April 2016

Project partners

Dr Claire Miller	***Project lead,*** School of Mathematics and Statistics (M&S), University of Glasgow (UoG).
Prof. Marian Scott	School of M&S, UoG.
Mr John Douglass and Dr Linda Pope	Environment Agency (EA).
Dr Mark Hallard and Dr Graeme Cameron	Scottish Environment Protection Agency (SEPA).
Dr Robert Willows	Formerly of the EA, Honorary Senior Research Fellow, School of M&S, UoG.
Named researcher:	Ms Kelly Gallacher, School of M&S, UoG.

Summary of project

Environment agencies invest a wealth of resource in monitoring river networks across the UK. This is essential in order to protect water and ecological quality and for reporting to Europe under, for example, the European Water Framework and Nitrates Directives (European Parliament 2000; 1991). The Environmental Statistics group at The University of Glasgow have been developing statistical methods to describe, and to better understand possible causes of, spatial and long-term trends, and seasonal patterns in river water quality data (Miller et al. 2014). In particular, the methods enable flexible non-linear patterns to be identified and incorporate the river network structure within the modelling and inference (O'Donnell et al., 2014; Haggarty et al., 2015). The benefit of such a statistical modelling approach is improved power through modelling across space and time, and reliable inference through allowing flexibility in the patterns and incorporating the network structure. Recent work by Gallacher (2016) has extended the statistical approaches above to investigate reducing dimensionality in order to identify spatiotemporal patterns for river water quality within large hydrological areas, while accounting for the river network structure, catchment covariates and temporal correlation. This project funded by the SECURE network used the statistical methodology produced in Gallacher (2016) to create statistical software tools in the R programming language (R Core Team, 2014) for use within the environmental science partner organisations.

Additionally, associated help files and tutorial guides were developed, and methods were integrated into the software TIBCO Spotfire (http://spotfire.tibco.com/, commonly used software within SEPA), through the use of the TIBCO Enterprise Runtime for R engine.

Project Aims and Objectives

The main aim of the project was to develop novel statistical software tools, for use within environmental science partner communities, which enable identification of spatiotemporal and coherent patterns over river networks.

Specifically the *objectives* were:

- to identify the features of the statistical methodology in Gallacher (2016) which are of greatest relevance and utility to the environmental science partners;

- to develop R routines, interfaces (and associated documentation) to apply the currently developed statistical methodology to new data sets and demonstrate their utility and benefits to environmental science partners;

- to demonstrate the integration of these routines within TIBCO Spotfire.

Project Activities

- Project commenced 1st November 2015
- First meeting of all project partners – November 2015
- Interim report submitted to project partners – January 2016
- Second meeting of all project partners – January 2016
- Kelly Gallacher secondment to SEPA – 2 weeks Feb/Mar 2016
- Workshop 1 – EA, DEFRA London – April 2016
- Workshop 2 – SEPA, Stirling – April 2016

The activities listed above supported the funded aims in the following ways:

Project board meetings: The project board meetings and collaboration with project partners enabled statistical software to be developed that was in areas of specific relevance to the environmental science partner organisations, and ensured that documentation was produced that was appropriate to enable interdisciplinary colleagues within the EA and SEPA to be able to utilise the methods and interpret the results.

Secondment: Kelly Gallacher spent two weeks on secondment to SEPA working at their Stirling office in the Environmental & Spatial Informatics Unit. During the secondment Kelly developed a tool using TIBCO Spotfire software to allow users

within SEPA to implement statistical methods developed in Gallacher (2016). The tool enables users to identify dominant spatial and temporal patterns in data collected on river networks and can be applied across the whole of Scotland, as well as individual river catchments.

Workshops: In the final week of this 6-month project, Kelly Gallacher and Claire Miller delivered workshops for the EA and SEPA to launch the statistical software tools developed as part of this project. These workshops and demonstration sessions were essential to demonstrate the implementation of the software and benefits of the methods to the wider environmental science community within the EA and SEPA.

Project Outcomes

The project aims and objectives were fully met and the following outputs are available.

Outputs:

- **An R package 'stpca' with full help documentation and tutorial guides** - The R package contains demonstration river network data files along with full tutorial documentation and help files. The R package and demonstration data files are available at:http://dx.doi.org/10.5525/gla.researchdata.277 (open access).

- **A tool within TIBCO Spotfire** - The methods developed within the 'stpca' package were integrated within the TIBCO Spotfire software to allow users within SEPA to implement the methods within their standard modelling framework. The methods are accompanied by a user help guide.

- **Pending publication** - a draft paper 'Flow directed PCA for Monitoring Networks' with planned submission to the journal Environmetrics has been produced to describe the methodological developments in Gallacher (2016) in this area and the R package developed in this project is referenced and used to illustrate results.

The above tools enable users to identify dominant spatial and temporal patterns in data collected on river networks, and will help users to identify areas where water quality has remained stable over time or to identify groups of monitoring sites that exhibit the same temporal pattern. Such information is valuable to inform and improve the design of future water quality monitoring programmes.

Project participants

There were two main activities as part of the project which involved greater interaction with colleagues within the environmental science community.

Secondment
An essential part of Kelly Gallacher's secondment was the training and expert advice that she received from programmers and environmental modellers and scientists working in the team in which she was based at SEPA. This interaction was essential for the integration of the methods with TIBCO Spotfire.

Workshops

Workshop 1 (EA): 26th April 2016, a workshop was held for 15 EA employees at the DEFRA offices in London to illustrate the R package ('stpca') developed as part of this project. Attendees included modellers, environmental scientists and those involved in water quality management. The workshop was organised jointly with John Douglass and Linda Pope and also attended by Robert Willows.

Workshop 2 (SEPA): 28th April 2016, an information and demonstration session was held for 9 SEPA employees in Stirling to illustrate the TIBCO Spotfire tool, developed as part of this project. Attendees included modellers, environmental scientists and those involved in water quality management. The workshop was organised jointly with Graeme Cameron and Mark Hallard.

Future collaboration and funding opportunities

This project contributed to the remit and objectives of the SECURE network by bringing together statistical and environmental science partners through the key challenge area of water resource. Specifically, the software tools developed by this project will enable the environmental science community to 'provide fresh intelligence and new insights into environmental change and society's management of that change' in the context of river water quality. This collaboration ensured tools were developed that are of direct and timely relevance to the environmental science community. This was the first time that all partners: UoG, SEPA and EA had collaborated on such a project together.

The software and outputs produced were very well received by environmental science colleagues with the workshops receiving positive feedback. The software and help/tutorial guides developed enable implementation of the methodology within SEPA and the EA. However, all partners recognise the requirement for further collaboration to ensure that results from the methods are interpreted and used appropriately, and this will be the focus of discussions for future funding applications.

Quality, novelty and level of engagement with world-leading research

The project team carefully considered the novel cutting-edge statistical methodological developments proposed in the PhD thesis Gallacher (2016) in order to identify methods of specific relevance to the EA and SEPA. These methods were then implemented as statistical software tools in R and Spotfire, and associated help materials and training guides were produced for use within environmental science partner organisations. The statistical research and tools developed enable statistically sound, novel, world-leading research to be undertaken in order to investigate the presence of common spatiotemporal patterns in river water quality.

This is specifically evidenced in the development of a publication which is pending submission for the high-ranking, international Statistics journal, Environmetrics.

Engagement with or involvement with early career researchers

This project employed an early career researcher Kelly Gallacher to develop the statistical software and deliver workshops to environmental science colleagues in collaboration with the project partners. As part of this project, Kelly received excellent training and expert advice from the environmental science project partners and more widely from colleagues within SEPA and the EA, in addition to enhancing her statistical software development skills. Engagement with the environmental science community provided important additional training in interdisciplinary working and collaboration, and experience of the pressures of delivering and regulating environmental policy.

Inclusion of activities that promote socio-economic impact

The secondment and two workshop events organised by the project partners ensured wider engagement with environmental science colleagues within the EA and SEPA that went far beyond that of just the project partners. This was essential for the development of the Spotfire tool and to promote the benefits of potential use and applicability of the software within the agencies.

Specifically, the software developed will be useful for ongoing policy work within the agencies such as water quality assessments for the Nitrates/Water Framework directive and potentially for evaluation of new agri-environment schemes such as Countryside Stewardship (http://www.nfuonline.com/science-environment/agri-environment-schemes/).

References

European Parliament (1991). Council directive of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources (91/676/EEC). Off J Eur Communities 1991b:L375/1.

European Parliament (2000). Directive 2000/60/EC of the European parliament and of the council of 23 October 2000 establishing a framework for community action in the field of water policy. Off J Eur Communities 2000:L327/1**.**

Gallacher (2016). Using river network structure to improve estimation of common temporal patterns. PhD thesis, University of Glasgow

Haggarty, R.A., Miller, C.A., and Scott, E.M. (2015) Spatially weighted functional clustering of river network data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(3), pp. 491-506.(doi:10.1111/rssc.12082).

Miller, C., Magdalina, A., Willows, R., Bowman, A., Scott, E., Lee, D., Burgess, C., Pope, L., Pannullo, F., and Haggarty, R. (2014) Spatiotemporal statistical modelling of long-term change in river nutrient concentrations in England & Wales. *Science of the Total Environment*, 466-7, pp. 914-923.(doi:10.1016/j.scitotenv.2013.07.113).

O'Donnell D, Rushworth A, Bowman AW, Scott EM, Hallard M. (2014) Flexible regression models over river networks. J R Stat Soc Ser C Appl Stat Jan; 63(1): 47–63.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.