Environmental Modelling & Software 60 (2014) 320–330

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Improving the predictive power of spatial statistical models of stream macroinvertebrates using weighted autocovariance functions

Jennifer C. Frieden [a], Erin E. Peterson [a, *], J. Angus Webb [b], Peter M. Negus [c]

[a] CSIRO Division of Computational Informatics, EcoSciences Precinct, PO Box 2583, Brisbane, Qld 4001, Australia
[b] Department of Resource Management and Geography, University of Melbourne, Parkville, Vic 3010, Australia
[c] Water Planning and Ecology, Queensland Department of Science, Information Technology, Innovation and the Arts, Dutton Park, Qld 4102, Australia

A R T I C L E I N F O

Article history:
Received 4 December 2013
Received in revised form
17 June 2014
Accepted 21 June 2014
Available online 17 July 2014

Keywords:
Spatial statistics
Spatial autocorrelation
Scale
Macroinvertebrate index
Aquatic monitoring

A B S T R A C T

Spatial statistical stream-network models are useful for modelling physicochemical data, but to-date have not been fit to macroinvertebrate data. Spatial stream-network models were fit to three macroinvertebrate indices: percent pollution-tolerant taxa, taxa richness and the number of taxalacking out-of-network movement (in-stream dispersers). We explored patterns of spatial autocorrelation in the indices and found that the 1) relative strength of in-stream and Euclidean spatial autocorrelation varied between indices; 2) spatial models outperformed non-spatial models; and 3) the spatial-weighting scheme used to weight tributaries had a substantial impact on model performance for the in-stream dispersers; with weights based on percent stream slope, used as a surrogate for velocity because of its potential effect on dispersal and habitat heterogeneity, producing more accurate predictions than other spatial-weighting schemes. These results demonstrate the flexibility of the modelling approach and its ability to account for multi-scale patterns and processes within the aquatic and terrestrial landscape.

Software availability

Name of software: SSN
Developer: Jay Ver Hoef and Erin PetersonDr Erin Peterson, CSIRO Division of Computational Informatics, EcoSciences Precinct, PO Box 2583, Brisbane, Qld, Australia, 4001.
Ph: +61 7 3833 5536, Email: support@spatialstreamnetworks.com
First year available: 2013
Hardware required: Standard laptop or desktop, with Windows or Linux OS
Software Required: R statistical software
Availability and Cost: Available for free online at http://cran.r-project.org/web/packages/SSN/index.html
Program language: R
Size: 5.72 MB
Name of software: STARS (Spatial Tools for the Analysis of River Systems)
Developer: Erin PetersonDr Erin Peterson, CSIRO Division of Computational Informatics, EcoSciences Precinct, PO Box 2583, Brisbane, Qld, Australia, 4001.
Ph: +61 7 3833 5536, Email: support@spatialstreamnetworks.com
First year available: 2013
Hardware required: Standard laptop or desktop, with Windows OS
Software Required: ESRI ArcGIS version 9.3.1, with ArcInfo License
Availability and Cost: Available for free online at http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml
Program language: Python version 2.5
Size: 135 KB
Name of software: FLoWS (Functional Linkage of Waterbasins and Streams)
Developer: David M. TheobaldConservation Science Partners, 11 Old Town Square, Suite 270, Fort Collins, CO, USA, 80524
Ph: +1 1 530 214 8905, Email: info@csp-inc.orgDr Erin Peterson, CSIRO Division of Computational Informatics, EcoSciences Precinct, PO Box 2583, Brisbane, Qld, Australia, 4001.
Ph: +61 7 3833 5536, Email: Erin.Peterson@csiro.au
First year available: 2006
Hardware required: Standard laptop or desktop, with Windows OS
Software Required: ESRI ArcGIS version 9.3.1, with ArcInfo License
Availability and Cost: Available for free online at http://www.fs.fed.us/rm/boise/AWAE/projects/SpatialStreamNetworks.shtml
Program language: Python version 2.5
Size: 556 kB

* Corresponding author.
E-mail addresses: jcfrieden@gmail.com (J.C. Frieden), Erin.Peterson@csiro.au (E.E. Peterson), angus.webb@unimelb.edu.au (J. Angus Webb), Peter.Negus@science.dsitia.qld.gov.au (P.M. Negus).

http://dx.doi.org/10.1016/j.envsoft.2014.06.019
1364-8152/© 2014 Published by Elsevier Ltd.

1. Introduction

Spatial autocorrelation represents the degree of spatial dependency in measurements collected in geographic space. It is an inherent characteristic of data collected in stream and river environments, where longitudinal and lateral connectivity, nested catchments, and broad-scale topographic and climatic gradients produce multiple, multi-scale patterns of spatial autocorrelation (Peterson et al., 2013). Spatial autocorrelation is often viewed as problematic; when traditional, non-spatial models are used to analyse spatially correlated data, it can lead to biased parameter estimates and invalid statistical inferences (Legendre, 1993). Alternatively, spatial statistical methods, such as geostatistical modelling (i.e. universal kriging) can be used to model spatially correlated data, account for influential covariates, and generate predictions with valid estimates of uncertainty at non-sampled locations (Cressie, 1993). These methods have recently been extended to represent the unique spatial relationships in stream networks (Ver Hoef et al., 2006; Ver Hoef and Peterson, 2010), which include the branching structure of the dendritic network, flow connectivity, the directionality of flow, and the 2-D terrestrial environment within which the network is embedded (Peterson et al., 2013). This provides a flexible modelling framework that can be used to account for *both* in-stream and Euclidean patterns of spatial autocorrelation in a single model (Peterson and Ver Hoef, 2010). Previous studies have been somewhat limited because proximity is based solely on Euclidean distance (e.g. Bonada et al., 2012; Shurin et al., 2009) or in-stream distance is used to study spatial relationships along a single, non-branching channel (e.g. Grenouillet et al., 2008).

Spatial stream-network models have been successfully applied to a number of physicochemical indicators, including temperature (Isaak et al., 2010; Jones et al., 2013; Ruesch et al., 2012), nitrate (Gardner and McGlynn, 2009) and dissolved oxygen (Cressie et al., 2006), as well as *E. coli* measurements (Money et al., 2009) and a modelled fish index (Peterson and Ver Hoef, 2010). There is thus a growing body of evidence suggesting that these methods are useful for up-scaling site-based measurements collected on stream networks to provide a more continuous perspective of stream characteristics (Cressie et al., 2006; Isaak et al., 2010; Money et al., 2009; Peterson and Ver Hoef, 2010; Ruesch et al., 2012), which is crucial for the spatial prioritization of management actions (Fausch et al., 2002). In contrast to many physicochemical variables, macroinvertebrate community indices are often strongly related to a combination of local-scale physicochemical and biological conditions (Downes et al., 1993; Minshall, 1984; Sawyer et al., 2004), which suggests that spatial autocorrelation may not be as prevalent in these data. Yet, many of these local-scale characteristics are thought to be influenced by the interaction of broader-scale network structure, geomorphology, and disturbance regimes (Benda et al., 2004), as well as, water chemistry and land use (Kratzer et al., 2006). Thus, it remains unclear whether this relatively new family of spatial statistical models will be equally suitable for predicting biological variables, such as macroinvertebrate indices, commonly used in broad-scale monitoring programs (e.g. Munné and Prat, 2009; Smith et al., 2011).

Another important aspect of modelling spatial relationships in stream networks is allowing for potential disjunctions at stream confluences (Peterson et al., 2013). Confluence zones (i.e. stream junctions) are biologically important elements of streams (Illies, 1961; Rice et al., 2006; Statzler and Higler, 1986), and have been linked to changes in macroinvertebrate densities (Katano et al., 2009; Rice et al., 2001). For example, Kiffney et al. (2006) found that small streams funnel materials such as nutrients and woody debris into wider main stem channels, and that this produced peaks in macroinvertebrate densities downstream of confluences; likely due to increased productivity and habitat complexity. Small, steep headwater streams may also be important drivers of downstream food webs, through the entrainment of leaf litter, in northern hemisphere streams with deciduous riparian vegetation (Cummins, 1974; Vannote et al., 1980). However, Bunn et al. (1999) showed that algae, rather than inputs of leaf litter, were the main driver of macroinvertebrate food webs in northern Queensland, Australia, where riparian vegetation tends to be evergreen. Furthermore, while macroinvertebrates in the northern hemisphere are often productive in small, steep headwater streams and drift downstream (Meyer et al., 2007), Australian studies have found that drift is usually related to death or catastrophic events (e.g. flooding) and may not be important for dispersal (Kerby et al., 1995). Although there may be uncertainty about what is causing disjunctive biological conditions at confluences, it is clear is that those drivers may be substantially different than those influencing physicochemical discontinuities at confluences.

Spatial stream-network models account for potential disjunctions at confluences using a spatial-weighting scheme that determines the degree of influence that each converging stream segment has on downstream locations (Peterson and Ver Hoef, 2010). To date, a spatial-weighting scheme based on Shreve's stream order (Shreve, 1966) has been used to generate spatial stream-network models (Cressie et al., 2006; Garreta et al., 2010), as well as spatial weights based on catchment area (Gardner and McGlynn, 2009; Isaak et al., 2010; Peterson et al., 2006; Peterson and Ver Hoef, 2010; Ruesch et al., 2012). Note that, Shreve's stream order has been used because it is additive, but Strahler's stream order (Strahler, 1957) could also be used. Catchment area and stream order have been used as surrogates for flow volume, a conceptually intuitive approach for water quality, temperature and fish because of the strong effects of longitudinal connectivity on these variables. However, catchment area and Shreve's stream order may not be as relevant for macroinvertebrates, which are strongly affected by local characteristics (e.g., Downes et al., 2000). In addition, there a variety of macroinvertebrate metrics including trophic and dominance indices, diversity, richness, and composition metrics, as well as, indices designed to represent feeding strategies, pollution tolerances, and habitat measures (Barbour et al., 1999). There are also numerous ways to construct indices within these categories and each index will have a metric-specific response to environmental perturbation. It is therefore unlikely that a single spatial-weighting scheme will be suitable in all cases given the broad range of physicochemical and biological processes affecting macroinvertebrate distribution and the diversity of indices available.

In this analysis, we used spatial stream-network models to explore patterns of spatial autocorrelation in a suite of macroinvertebrate indices collected in the wet tropics of Queensland, Australia. In particular, we wanted to test whether 1) accounting for spatial autocorrelation improved the predictive power of the models fit to biological indices; 2) patterns of spatial autocorrelation differed depending on the macroinvertebrate index used; and 3) the choice of spatial-weighting scheme affected the predictive power of the spatial model.

2. Materials and methods

2.1. Data and study area

Macroinvertebrate data were collected at 60 sites in July and September 2009 (austral winter) in a sub-catchment of the Tully River Basin in the Wet Tropics bioregion of Queensland, Australia (Fig. 1). We collected data within a single season because there is little evidence of seasonal variability in Australian macroinvertebrate indices (Chessman et al., 1997; Marshall et al., 2001), which do not receive seasonal pulses of litterfall and subsequent increases in nutrients and productivity (Abelho and Graca, 1996; Boulton and Brock, 1999). The climate in the Tully