Computers & Geosciences 36 (2010) 54–70

Contents lists available at ScienceDirect

Computers & Geosciences

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

Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics

Paul Harris [a,*], Chris Brunsdon [b]

[a] National Centre for Geocomputation, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland
[b] Department of Geography, University of Leicester, Leicester LE1 7RH, UK

ARTICLE INFO

Article history:
Received 20 October 2008
Received in revised form
29 March 2009
Accepted 1 April 2009

Keywords:
Local statistics
Geographical kernel weighting
Nonstationarity
Acidified surface waters
Catchment characteristics

ABSTRACT

In this study, geographically weighted summary statistics (GWSSs) are used to investigate spatial variation and spatial relationships in a freshwater acidification critical load data set covering Great Britain. This use of GWSSs not only provides valuable insight into the critical load process prior to a geographically weighted regression (GWR) calibration, but also helps in interpreting its output. GWSSs are similarly useful prior to the calibration of other spatial models, such as those used in geostatistics. Results agree with those of previous research, where relationships between critical load and contextual catchment data can vary across space. However the more sophisticated models used here are shown to be much more flexible and informative, allowing more spatial patterns to be revealed than before.

1. Introduction

Acid deposition is a major environmental threat to lakes and streams throughout large areas of upland Britain (Mason, 1993). Pollutants that contribute to freshwater acidification are generally emitted as sulphur dioxide and nitrogen oxides. The major sources of such acidifying compounds are from combustion of fossil fuels at power stations or from other industrial processes. Vehicle exhausts, agriculture, volcanoes and the oceans also contribute. For freshwater acidification most of the atmospheric deposition is to the terrestrial part of the catchment rather than open water. Therefore lake and stream acidification is a function of flow paths and the physical and chemical properties of catchment soils. Acidified freshwaters are a hostile environment for many forms of aquatic life and consequently of environmental concern. Continuous assessment and informed management strategies for freshwaters are fundamental for their protection.

One approach to protecting freshwaters focuses on the calculation of acid deposition critical load values at freshwater sites. Critical load values are calculated in such a way as to indicate a site's capacity to buffer the input of strong acid anions of sulphur and nitrogen. Critical load values are thresholds and can be compared directly to current and future deposition values. For sites where the deposition value exceeds the critical load value, acidification and associated environmental damage is expected. Spatial variability in critical load values should be considered jointly with spatial variability in deposition values. This approach allows for selectivity and for exceeded sites to be preferentially managed. For remediation of sites, two avenues are possible: (a) reduce (nearby) deposition rates or (b) physically neutralise freshwater acidity (e.g. by the addition of an alkali compound). In general, the susceptibility of freshwaters to acidification varies according to geology and land use. Waters situated on bedrocks with a high weathering rate are usually well buffered against rain-deposited acidity by the relatively rapid release of neutralising base cations (mainly Ca^{2+} and Mg^{2+}). However for areas of slowly weathering bedrocks the reverse is true, with acidifying compounds displacing H^{+} ions, which directly lead to acidification. For Great Britain, the granite regions of Scotland and Wales are particularly affected by acidification.

To calculate a critical load for any given freshwater site requires surface water chemistry data. Collecting such data for every site across Great Britain is prohibitively expensive. Therefore previous research has looked at ways of predicting critical loads at sites where water chemistry data are unavailable, as an alternative to a costly sampling programme. In this respect, research has endeavoured to link critical load variation with various catchment characteristics. This is useful as many catchment variables can be formulated from existing data sources and

* Corresponding author. Tel.: +353 1708 6208; fax: +353 1708 6456.
E-mail address: Paul.Harris@nuim.ie (P. Harris).

0098-3004/$ - see front matter
doi:10.1016/j.cageo.2009.04.012

are therefore relatively inexpensive. Catchment variables have been used to predict classes of critical load (Hall et al., 1995) or to explain critical load variation (Kernan et al., 1998, 2001) at Great Britain and similar spatial scales. Such studies found moderately strong relationships between critical load and catchment data, where the strength and nature of relationships could vary according to sample scale in both attribute-and geographic-space. However studies applied only basic methods, where any local regression modelling was fairly rudimentary in design using arbitrary aspatial or spatial partitions.

For this and companion studies (in preparation), it is taken that a critical load spatial process would be better investigated using more sophisticated techniques. In the first instance (this study), geographically weighted summary statistics (GWSSs; Brunsdon et al., 2002; Fotheringham et al., 2002) are used to explore the critical load data set spatially. This initial study acts as an informative precursor to an exploration with geographically weighted regression (GWR; Brunsdon et al., 1996; Fotheringham et al., 2002) and to other spatial models. With GWSSs (and GWR) nearby data are given more influence by weighting observations according to a distance-decay function. This use of spatially weighted data enables the calibration of numerous location-specific statistics (or regressions). This 'moving-window' approach (where weights follow a focal point around the map) allows statistics to be computed for regions not necessarily attainable in any partition-based approach. Such a use of sample information tends to provide a continuous and smooth model output, where a local statistic can be mapped and visually explored. The nonparametric techniques of GWSS and GWR are influenced by kernel density estimation (KDE) methods (Silverman, 1986; Wand and Jones, 1995), the attribute-space local regression (LR) models of Cleveland (1979) and Loader (2004), and the generalised additive/varying-coefficient models of Hastie and Tibshirani (1990, 1993). In addition to the modelling of continuous spatial processes (i.e. models used in this study), the kernel approach has been extensively adopted for the modelling of point spatial processes (see Diggle, 1985; Silverman, 1986).

The models of this and companion studies will use similar critical load and catchment variables to those used in the studies of Hall, Kernan and co-workers (from above) and therefore some limited comparison between studies will be possible. Furthermore, the use of GWSSs and GWR need not be confined to this particular data set as they are likely to be similarly applicable to data sets found from other environmental processes that are considered heterogenic. For example, applications to critical load data sets for regions of China where acidification is currently posing a major environmental problem (Brimblecombe, 2007) or to critical load data sets for other pollutants, such as those found for heavy metals (e.g. see Slootweg et al., 2007).

2. Data

The calculation of a critical load value for any freshwater site is itself a complex issue and competing models exist for their calculation. Steady-state approaches calculate values such that exceedances (critical load minus deposition) reflect potential future damage once steady-state is achieved. Steady-state models include the steady-state water chemistry (SSWC) model (Henriksen et al., 1992; Curtis et al., 2000), the Diatom model (Battarbee et al., 1996) and the First-order Acidity Balance model (Posch et al., 1997; Curtis et al., 2000). Models can be calibrated for sulphur deposition, for nitrogen deposition or for both (total acidity). For this study, critical load values from the SSWC model for total acidity are spatially modelled. Units for critical loads (and deposition data) are in keq. H^+ ha^{-1} $year^{-1}$.

Researchers at the Department of Geography, University College London (UCL), provided the critical load and the contextual catchment data. The critical load data stem from a water chemistry sampling programme for Great Britain as part of the UK Department of Transport and Regions critical loads mapping programme (Kreiser et al., 1993). Water chemistry samples were taken during the autumn or early spring over the period 1992–1994. Sites were chosen to represent the most sensitive water body within either a 10 km grid square (for medium- to high-sensitive areas) or within a 20 km grid square (for low- or non-sensitive areas) so that the minimum critical load could be calculated. Research teams within the Critical Loads Advisory Group (CLAG) then used the water chemistry data to calculate and map critical load values. Details of the sampling programme and mapping exercise are given in CLAG Freshwaters (1995).

The version of critical load and catchment data used for this study was provided in January 2002. At this time, the water chemistry data had been screened for problematic values by researchers at UCL, which resulted in a critical load (and catchment) data set of 1371 sites covering the whole of Great Britain. This data set was further manipulated for this and companion studies to avoid problems of preferential sampling (i.e. data in medium- to high-sensitive areas is over-represented) when calibrating GWSSs and other spatial models (not presented here). This data manipulation also removed sites with missing data. As a result, a spatially representative (declustered) data set of 497 sites for model calibration and a spatially representative (set-aside) data set of 189 sites for model validation (not used here) were found. The coverage of the resultant calibration data (Fig. 2b) is extensive and fairly regular, which is suitable for spatial modelling. Investigations (not presented) found this rather large loss of model calibration information to have a negligible effect on model interpretation or performance.

To explain critical load variation, four percentage-based class variables are used and manipulated. These catchment-specific variables are geological sensitivity (GSP), soil buffering capacity (SBCP), soil critical load (SCLP) and land cover (LCP). The first three of these variables relate to a freshwater site's ability to buffer acid loading and comprise of four, three or five ordinal classes for GSP, SBCP and SCLP, respectively. The twenty-five-class LCP variable is nominal. Such data were generated by over-laying digitised catchment areas for each sampled site on to existing digital maps. Full descriptions of the GSP and LCP data generation can be found in Kernan et al. (1998, 2001). For the SBCP and SCLP data generation, the reader is referred to Kernan et al. (1998) (where SBCP is termed soil sensitivity). If data reliability is considered an issue, then the following order of reliability is assumed: LCP, GSP, SBCP and SCLP (with the most reliable first). Other contextual variables were also available (e.g. site altitude, rainfall, etc.), but each variable added little to the variance explained of any regression fit; hence these variables were discarded.

To more easily facilitate the use of percentage-based class variables in this study's correlations (and a companion study's regressions), the three ordinal variables were re-formulated into single-value, weighted sensitivity data (Wt.GSP, Wt.SBCP and Wt.SCLP). This results in a continuous variable form with only a marginal loss of information. Table 1 summarises the range of values that the original percentage-based and corresponding weighted variable can take according to an expected acid buffering capacity (or acid sensitivity). Thus low critical load values would be expected to correspond to low Wt.GSP, Wt.SBCP and Wt.SCLP values (and vice versa). Twenty-five land cover