Science of the Total Environment 612 (2018) 840–852

Contents lists available at ScienceDirect

Science of the Total Environment

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

Using spatial-stream-network models and long-term data to understand and predict dynamics of faecal contamination in a mixed land-use catchment

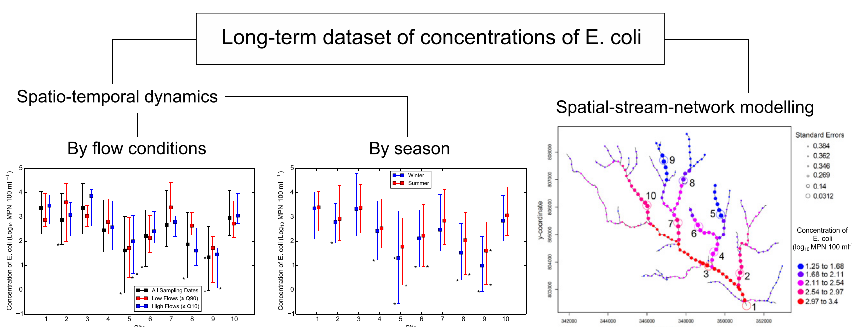
Aaron James Neill [a,b,*], Doerthe Tetzlaff [a,c,d], Norval James Colin Strachan [e], Rupert Lloyd Hough [b], Lisa Marie Avery [b], Helen Watson [b], Chris Soulsby [a]

[a] Northern Rivers Institute, School of Geosciences, St Mary's Building, Elphinstone Road, University of Aberdeen, Aberdeen AB24 3UF, Scotland, United Kingdom
[b] The James Hutton Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland, United Kingdom
[c] IGB Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
[d] Humboldt University Berlin, Berlin, Germany
[e] School of Biological Sciences, University of Aberdeen, Cruickshank Building, St Machar Drive, Aberdeen AB24 3UU, Scotland, United Kingdom

HIGHLIGHTS

- A key challenge is to understand and predict catchment-scale faecal contamination.
- We employ long-term *E. coli* data and novel use of spatial-stream-network models.
- Concentrations of *E. coli* not clearly associated with flow conditions or season
- A significant predictor of spatial patterns was an Anthropogenic Impact Index.
- Spatial-stream-network models helped predict potential "hot spots" of contamination.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 3 April 2017
Received in revised form 26 July 2017
Accepted 15 August 2017
Available online xxxx

Editor: D. Barcelo

Keywords:
E. coli
Faecal indicator organism
Microbial pollution
Spatio-temporal dynamics
Surface water
Water quality

ABSTRACT

An 11 year dataset of concentrations of *E. coli* at 10 spatially-distributed sites in a mixed land-use catchment in NE Scotland (52 km^2) revealed that concentrations were not clearly associated with flow or season. The lack of a clear flow-concentration relationship may have been due to greater water fluxes from less-contaminated headwaters during high flows diluting downstream concentrations, the importance of persistent point sources of *E. coli* both anthropogenic and agricultural, and possibly the temporal resolution of the dataset. Point sources and year-round grazing of livestock probably obscured clear seasonality in concentrations. Multiple linear regression models identified potential for contamination by anthropogenic point sources as a significant predictor of long-term spatial patterns of low, average and high concentrations of *E. coli*. Neither arable nor pasture land was significant, even when accounting for hydrological connectivity with a topographic-index method. However, this may have reflected coarse-scale land-cover data inadequately representing "point sources" of agricultural contamination (e.g. direct defecation of livestock into the stream) and temporal changes in availability of *E. coli* from diffuse sources. Spatial-stream-network models (SSNMs) were applied in a novel context, and had value in making more robust catchment-scale predictions of concentrations of *E. coli* with estimates of uncertainty, and in enabling identification of potential "hot spots" of faecal contamination. Successfully managing faecal

* Corresponding author at: Northern Rivers Institute, School of Geosciences, St Mary's Building, Elphinstone Road, University of Aberdeen, Aberdeen AB24 3UF, Scotland, United Kingdom.

E-mail addresses: aaron.neill@abdn.ac.uk (A.J. Neill), d.tetzlaff@abdn.ac.uk (D. Tetzlaff), n.strachan@abdn.ac.uk (N.J.C. Strachan), rupert.hough@hutton.ac.uk (R.L. Hough), lisa.avery@hutton.ac.uk (L.M. Avery), helen.watson@hutton.ac.uk (H. Watson), c.soulsby@abdn.ac.uk (C. Soulsby).

http://dx.doi.org/10.1016/j.scitotenv.2017.08.151

contamination of surface waters is vital for safeguarding public health. Our finding that concentrations of *E. coli* could not clearly be associated with flow or season may suggest that management strategies should not necessarily target only high flow events or summer when faecal contamination risk is often assumed to be greatest. Furthermore, we identified SSNMs as valuable tools for identifying possible "hot spots" of contamination which could be targeted for management, and for highlighting areas where additional monitoring could help better constrain predictions relating to faecal contamination.

1. Introduction

When faecal material is transferred to surface waters, the delivery of faecal pathogens including *Escherichia coli* O157, *Campylobacter* and *Cryptosporidium parvum* may also occur (Oliver et al., 2005a). Such pathogens can lead to gastrointestinal illness in humans if exposure to contaminated water occurs through, for example, recreational uses of water or consumption of drinking water from poorly-treated private supplies (Fewtrell and Kay, 2015; Strachan et al., 2006). In the European context, legislation such as the Drinking Water Directive (Council Directive 98/83/EC) and revised Bathing Water Directive (Council Directive 2006/7/EC) stipulate acceptable concentrations of faecal indicator organisms (FIOs), used as a proxy for faecal contamination, that should be complied with for different uses of water in order to safeguard public health. Such legislation has prompted increased recognition of the need to better understand the dynamics and drivers of faecal contamination in surface waters, so that effective management strategies can be devised that permit microbiological water quality standards to be met (Kay et al., 2008a).

In rural areas, the potential for faecal contamination is often high due to potential for contributions from both point and diffuse sources. Sewage infrastructure is often more rudimentary in such areas, with septic tanks and combined sewer overflow waste water treatment works (WWTWs) being common, both of which represent important point sources of contamination (Kay et al., 2008b). Meanwhile, spread manure and faeces from grazing animals arising from intensive agriculture are examples of diffuse sources (Chadwick et al., 2008). The high potential for faecal contamination in rural areas can impact on a number of downstream water uses which, in turn, has implications for meeting legislative requirements and for public health. For example, exports of faecal contaminants from rural catchments have been suggested to account for large proportions of contamination observed in coastal bathing waters (Crowther et al., 2003). Furthermore, private water supplies are commonly relied upon to provide drinking water in rural areas, some of which may be drawn from surface waters. However, such supplies often employ only limited treatment mechanisms, meaning there is increased potential for human infection by faecal pathogens when the microbiological quality of the raw water of a private supply is poor (Kay et al., 2007). As such, there is a vital need to better manage faecal contamination in rurally-influenced catchments.

Compared with other types of water pollution, the evidence-base for understanding the behaviour and survival of faecal pathogens and FIOs in the environment has, historically, been more limited (Kay et al., 2008a). Significant knowledge gaps still persist in relation to understanding the spatio-temporal dynamics of faecal contamination, especially at the catchment scale where decisions regarding management of water quality need to be made (Oliver et al., 2016). In particular, understanding the response of concentrations of FIOs to hydrological conditions and season using datasets long enough to capture sufficient hydroclimatic variation, and developing models that can be used to infer potential sources of contamination from spatial patterns of FIOs and make robust predictions for unmeasured locations represent key challenges at this scale (Kay et al., 2010; Tetzlaff et al., 2012; Vitro et al., 2017).

Previous catchment-scale studies (e.g. Crowther et al., 2002, 2003; Kay et al., 2005, 2008b; McGrane et al., 2014; Tetzlaff et al., 2012) have offered important insights into the dynamics and controls of faecal contamination. In particular, high flow events and summer have often been identified as periods when concentrations of FIOs are likely to be elevated. In addition, multiple linear regression models (MLRMs) linking spatial patterns in concentrations to readily-available land-cover variables as proxies for different sources of contamination have generally identified intensive livestock farming and human sewage inputs as potentially important sources. Where more detailed datasets have been available, some studies have further identified physical, chemical and biological factors that can be significantly associated with spatial patterns of FIOs. For example, Dwivedi et al. (2013) found temperature, dissolved oxygen, phosphate, ammonia, suspended solids and chlorophyll to be important for estimating *E. coli* loads in Plum Creek, Texas. However, many past studies have generally been constrained by the availability of only short-duration (<1–2 year) datasets relating to concentrations of FIOs. Furthermore, many of the regression models based on land cover for FIOs are fairly simple in their implementation (Kay et al., 2010). For example, elevated concentrations of FIOs during high flow conditions are often attributed to increased hydrological connectivity between sources of contamination and the stream network, particularly via overland flow (Dwivedi et al., 2016; Kay et al., 2008b; Tyrrel and Quinton, 2003). However, conceptualisation of the connectivity potential of certain land covers within regression models is rare (an exception is Crowther et al., 2003, who showed that concentrations of FIOs during low flows were most influenced by land use within 1–2 km surface-flow distance of a sub-catchment outlet, whilst during high flows land use across the whole of a sub-catchment was important). Also rare is the recognition that concentrations of FIOs at flow-connected sampling sites along a stream network may not be independent of one another (although Vitro et al., 2017 successfully account for this with a spatial regression model). This may give rise to spatial autocorrelation between sampling sites, which, if not accounted for, may lead to significance being incorrectly assigned to the land-cover variables of the models (Isaak et al., 2014).

Whilst dataset length may be logistically constrained, representing hydrological connectivity in models is a possibility. A potential approach is the use of topographically-based indices, such as the Network Index (Lane et al., 2004). This is an extension of the topographic wetness index of Beven and Kirkby (1979) and accounts for the requirement that for a saturated area to be hydrologically connected to a stream via an overland flow path, the entire flow path must be saturated to prevent disconnection by processes such as re-infiltration (Lane et al., 2004, 2009). However, whilst this metric has potential in characterising the hydrological connectivity likelihood of diffuse sources of pollution, it has rarely been implemented in this context (Lane et al., 2009; an exception being SCIMAP outlined by Reaney et al., 2011).

Spatial-stream-network models (SSNMs) represent an advancement in geostatistical methods that mean it is also now possible to account for spatial autocorrelation between observations along stream networks (see Ver Hoef and Peterson, 2010 and Ver Hoef et al., 2006 for full details). Central to SSNMs is that, unlike traditional geostatistical methods, autocorrelation between observed locations is based on stream distance as opposed to Euclidian distance. Stream distance is the shortest distance between two points when following the stream network. Autocovariance functions based on stream distance are based on moving average constructions, and may be defined for sites