



# Chronic urban hotspots and agricultural drainage drive microbial pollution of karst water resources in rural developing regions

Sarah J. Buckerfield<sup>a,\*</sup>, Richard S. Quilliam<sup>a</sup>, Luc Bussiere<sup>a</sup>, Susan Waldron<sup>b</sup>, Larissa A. Naylor<sup>b</sup>, Siliang Li<sup>c</sup>, David M. Oliver<sup>a</sup>

<sup>a</sup> Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, United Kingdom

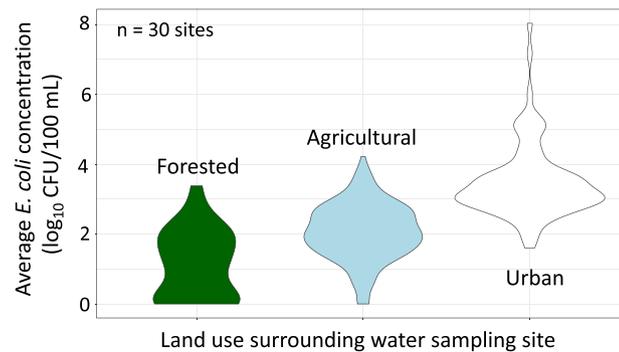
<sup>b</sup> School of Geographical and Earth Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom

<sup>c</sup> Institute of Surface-Earth System Science, Tianjin University, Tianjin 300072, China

## HIGHLIGHTS

- Karst water resources in paddy farming areas vulnerable to microbial contamination
- Land use is the most important control of average *E. coli* concentration in water.
- Water draining from urban land is consistently highly contaminated.
- Water draining from agricultural land is consistently moderately contaminated.
- Transient higher *E. coli* concentrations associated with rainfall and paddy drainage

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 1 May 2020

Received in revised form 16 June 2020

Accepted 9 July 2020

Available online 16 July 2020

Editor: José Virgílio Cruz

### Keywords:

Drinking water quality

Faecal contamination

Spatiotemporal controls

Karst water resources

Rural developing regions

Mixed-effects modelling

## ABSTRACT

Contamination of surface and groundwater systems with human and animal faecal matter leads to exposure of reliant populations to disease causing micro-organisms. This exposure route remains a major cause of infection and mortality in developing countries, particularly rural regions. To meet the UN's sustainable development goal 6: *Ensure availability and sustainable management of water and sanitation for all*, we need to identify the key controls on faecal contamination across relevant settings. We conducted a high-resolution spatial study of *E. coli* concentration in catchment drainage waters over 6 months in a mixed land-use catchment in the extensive karst region extending across impoverished southwest China. Using a mixed effects modelling framework, we tested how land-use, karst hydrology, antecedent meteorological conditions, agricultural cycles, hydrochemistry, and position in the catchment system affected *E. coli* concentrations. Land-use was the best predictor of faecal contamination levels. Sites in urban areas were chronically highly contaminated, but water draining from agricultural land was also consistently contaminated and there was a catchment wide pulse of higher *E. coli* concentrations, turbidity, and discharge during paddy field drainage. *E. coli* concentration increased with increasing antecedent rainfall across all land-use types and compartments of the karst hydrological system (underground and surface waters), but decreased with increasing pH. This is interpreted to be a result of processes affecting pH, such as water residence time, rather than the direct effect of pH on *E. coli* survival. Improved containment and treatment of human waste in areas of higher population density would likely reduce contamination hotspots, and further research is needed to identify the nature and distribution of sources in agricultural land.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\* Corresponding author.

E-mail address: [sarah.buckerfield1@stir.ac.uk](mailto:sarah.buckerfield1@stir.ac.uk) (S.J. Buckerfield).

## 1. Introduction

Faecal contamination of catchment drinking water sources increases the risk of human exposure to pathogenic micro-organisms. Consumption of faecally-contaminated water causes an estimated 1.8 million deaths annually and is the leading cause of waterborne disease (Brusseau et al., 2019). Due to reliance on untreated catchment water resources and a lack of sewage infrastructure and distribution networks, this impact is predominantly felt in developing countries, and in particular, rural regions (Bain et al., 2014; Bivins et al., 2017; UNICEF and WHO, 2019). Faecal contamination is often evaluated using faecal indicator organisms (FIOs), such as *E. coli*, which are easier and cheaper to quantify than specific pathogens and thus useful for initial assessment of the potential for public health risk from water sources (Edberg et al., 2000).

Faecal contamination of drinking water can originate from both point (e.g. sewage discharge) and diffuse (e.g. agricultural runoff) sources, making characterisation of controls on delivery of microbial contaminants to aquatic environments inherently difficult (Cho et al., 2016). Some studies have identified temporal variables, such as rainfall and source availability, to be major controls of *E. coli* delivery to drainage networks (Buckerfield et al., 2019a; McKergow and Davies-Colley, 2010; Sinclair et al., 2009), while others recognise a key role of spatially-distributed chronic point sources such as leaking septic tanks or livestock crossing points that are independent of rainfall/discharge conditions (Murphy et al., 2015; Neill et al., 2018). In karst drainage systems where point recharge via swallow or sink holes is a major source of recharge to the aquifer, the presence of sources surrounding the swallow or sink holes during recharge periods is generally a critical control on microbial water quality at discharge springs (Ender et al., 2018; Pronk et al., 2006). Existing evidence suggests catchments are often complex systems within which interactions between spatial and temporal controls ultimately govern the rate of *E. coli* delivery to receiving waters, often depending on climate and agricultural cycles, and hydrology. In larger and/or more complex catchments, where delineation of drainage networks and source distribution is more difficult, there is greater uncertainty in the relative importance of those controls, and how they might interact to influence microbial water quality, hindering assessment and modelling (Dymond et al., 2016).

Karst catchments are hydrologically complex and represent a key example of landscape systems that are less well understood with respect to microbial pollution (Buckerfield et al., 2019b). Karst is a unique geological terrain developed in highly-soluble rocks (Ford and Williams, 2007), where dissolution-developed cavities allow surface water and entrained contaminants to infiltrate directly into the groundwater system. Dissolution along existing planes of weakness such as fractures and faults results in extremely rapid flow pathways in underground conduits (White, 2018). Efficient hydrological pathways result in diminished capacity for *E. coli* die-off or removal by natural filtration processes before water re-emerges at springs, used worldwide for drinking and domestic purposes (Schiperski et al., 2016; Stevanović, 2018).

The southwest-China karst region is one of the most extensive in the world, characterised by the highest poverty rates in China (Cao et al., 2015). The rural population is at high risk of contracting infections due to living in impoverished conditions in close proximity with livestock (Yang et al., 2017), and high rates of acute diarrheal disease are observed in children (Zhang et al., 2016). Managing microbial water quality to help sustain ecosystem services for poor communities in this fragile karst environment is therefore a high priority and faecal contamination of drinking water supplies in China is receiving increased research attention (e.g. Hong et al., 2010; Ye et al., 2013). However, as is the case worldwide, and particularly for developing regions, most studies focus on points of human exposure, and very few studies in China have conducted high resolution spatial and temporal sampling at the catchment scale (Oliver et al., 2018; Xue et al., 2018).

The aim of this research, therefore, was to assess which spatiotemporal variables are most significant as predictors of *E. coli* concentration in an agricultural karst setting. This spatially high resolution study accommodates longer timeframe temporal changes in spatial controls, and complements a high temporal resolution study sampling high discharge conditions (Buckerfield et al., 2019a). The study site was a developing region of China, meaning results hold relevance for other developing regions of the world with similar conditions. Specifically, we generated repeated spatial and temporal measures of *E. coli* concentrations and relevant water quality parameters over a six month sampling campaign, and from this determined the significance of land-use, hydrological, and meteorological predictors of microbial water quality using a mixed modelling approach.

## 2. Materials and methods

### 2.1. Study catchment

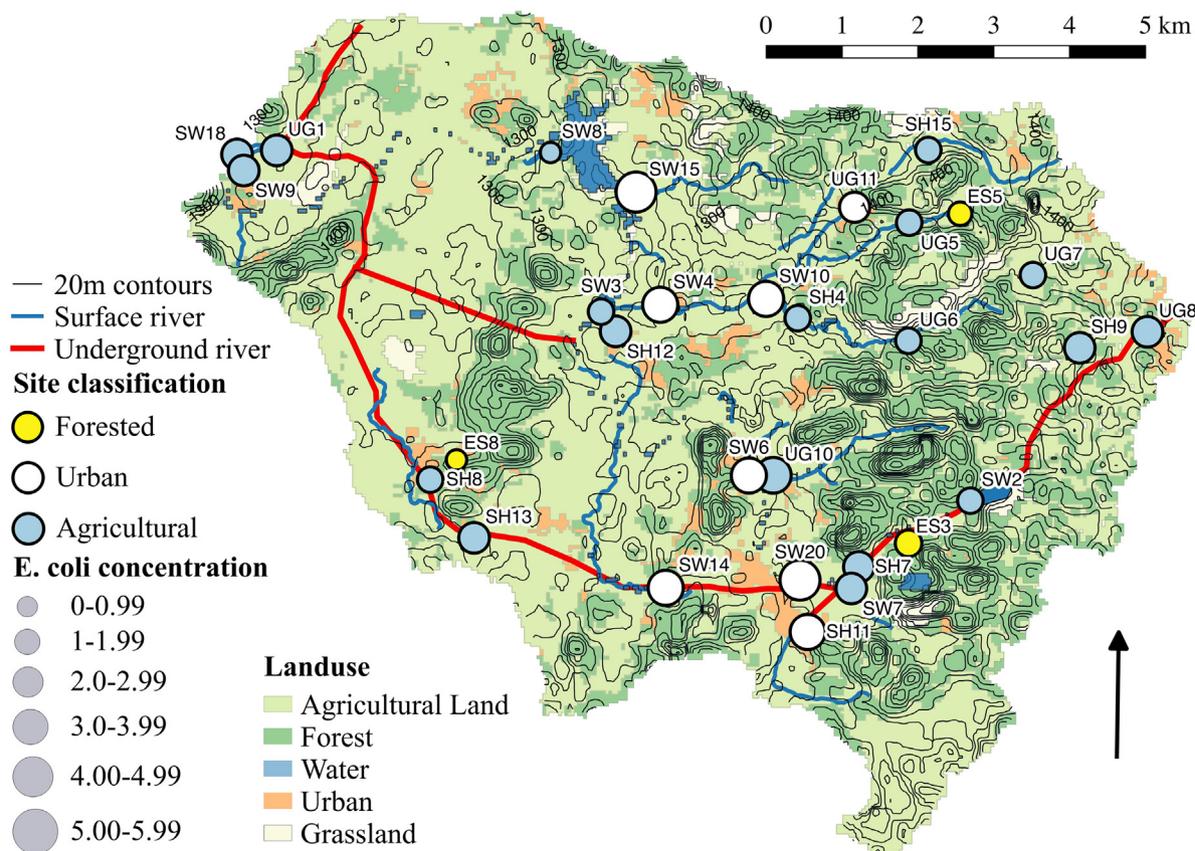
Water samples were collected from 30 sites distributed across the Houzhai (HZ) catchment, which drains a land area of 73.5 km<sup>2</sup> (Fig. 1). Samples were collected twice a month (an approximately 2 week interval) between 12/04/2018 and 08/10/2018 (total 11 samples/site). This six month period encompassed the end of the dry season (and harvest of dry season crops), the wet season (a full paddy rice crop), and transition into the following dry season. The sites were selected to: (i) provide a cross-section of the hydrological system from the headwaters to the outlet; (ii) represent the different land-uses; (iii) sample progressively down the major tributaries; and (iv) provide good representation of the types of water bodies and outlets within the surface and ground water compartments of the karst hydrological system. As such, sites were distributed as evenly as possible within the constraints of their abundance/accessibility in the catchment between epikarst springs (discharging water from hillslope epikarst, all forested in this study), valley outlet springs (higher order springs discharging water from entire valley systems), sinkholes (formed by bedrock collapse, providing surface connectivity with the groundwater system), surface rivers, and reservoir outlets. Paddy and dry land agriculture are practiced in the HZ catchment, and residents are often dependent on catchment water resources for drinking, domestic, and irrigation purposes (Buckerfield et al., 2019b). The hydrological distance of each sample site from the outlet of the catchment was calculated assuming shortest flow-path length using maps of topography, underground, and surface drainage systems in QGIS v.2.14.10-Essen, and field observations. The flow-path length for individual sampling sites was calculated as the longest possible flow path for water reaching each site.

### 2.2. Microbiological analysis

Grab water samples were collected aseptically, returned to the laboratory in a cool box, and processed within 14 h of collection to determine *E. coli* concentration. *E. coli* were enumerated using the standard method of membrane filtration (Environment Agency, 2009) following methods in Buckerfield et al. (Buckerfield et al., 2019a).

### 2.3. Water quality, discharge, and meteorological data

Sample temperature was recorded in situ, and electrical conductivity (EC) and pH were measured in the laboratory within 30 h of data collection using a WTW inoLab Multi 9430 IDS benchtop meter. Turbidity was measured in triplicate using a laboratory turbidity meter. V-notch weirs at four sites (UG5: 1.25 km<sup>2</sup>, UG6: 2.4 km<sup>2</sup>, SW14: 17.7 km<sup>2</sup>, and UG1: 73.5 km<sup>2</sup>, Fig. 1) provide infrastructure for water level loggers (GB/T3091-2008 pressure transducers) recording water depth at a 5 minute interval, from which stream discharge (Q) was derived using established rating curves (Zhang et al., 2017). In-stream temperature, turbidity, pH, and EC data was available for six sites (four shared with



**Fig. 1.** Catchment map showing location of sampling sites, land-use, location of major underground conduits and surface rivers, elevation, and average *E. coli* concentration across all sampling dates for all sites. Sites are coloured according to surrounding land-use and the size of the circles is proportional to the average *E. coli* concentration at that site ( $\log_{10}$  CFU/100 mL). Site descriptions: SW = surface water, UG = underground outlet (of sub-catchment), ES = epikarst spring, SH = sinkhole. Elevation ranges from 1552 m above sea level (masl) in the east, where forested karst cones delineate the headwaters, to 1212 masl at the outlet in the west (at the site marked SW18). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the Q data) at a 15 minute interval, from Aqua TROLL 600 multi-parameter sondes. Rainfall, temperature, and relative humidity data was obtained from a rain gauge at Lahoetain (site SW14, Fig. 1). Table 1 provides a summary of the rationale behind the parameters considered in this study, either as potential predictors of *E. coli* concentration or as a tracer of a relevant process.

#### 2.4. Data analysis and modelling procedure

All statistical analyses and modelling were conducted in R version 3.6.1 (R Core Team, 2019). *E. coli* concentration data was  $\log_{10}$  transformed, bringing the distribution closer to a normal distribution. A Wilcoxon signed-rank test was used to test whether the mean of samples taken on any given date or site was significantly different to the mean of all samples, as a means of identifying potential one-off influences that were not incorporated into the model as a predictor. Correlation between water quality variables was tested using a repeated measures correlation coefficient (Bakdash and Marusich, 2017).

##### 2.4.1. Random Forest algorithm: identifying relevant predictor variables

Numerous temporal and spatial predictors are potentially relevant to *E. coli* concentration (Table 1), and it is possible to represent a number of these predictors in different ways (e.g. categorical, continuous). To identify the most important predictors for testing in a linear mixed effects model, and the best representation of each predictor, the full suite of potential predictors was first tested using the Random Forest (RF) algorithm. Random Forest is a machine learning algorithm capable of handling a large number of predictor variables and ranking their importance, without being affected by variable collinearity (Liaw and

Wiener, 2007). Average antecedent meteorological conditions over a range of time frames prior to sampling, a range of continuous and categorical representations of land-use and position in the karst system, and a temporal predictor marking the timing of paddy field discharge were tested. A Mann-Whitney *U* test was used to test whether the mean *E. coli* and water quality parameters of groups defined by categorical splitting were significantly different from each other (significance threshold:  $p < 0.05$ ). A full list of predictors included in the random forest algorithm is given in SI (Supplementary Information) Table S1.

##### 2.4.2. Model selection with a mixed effects model structure

Mixed effects models estimate variance associated with group membership and account for pseudo-replication in hypothesis tests (McNeish and Kelley, 2018). We checked for multi-collinearity of predictors using variance inflation factors (VIFs), with  $VIF > 5$  considered as the threshold for predictors being collinear (Dormann et al., 2013). We assessed model quality by visual examinations of model diagnostics for each potential model. Residuals for the entire model, residuals for each continuous predictor, and residuals for each level of random effects were checked for non-linearity or non-constant variance, and quantile plots were checked for normality. Temporal autocorrelation was tested for by: (i) checking whether residuals followed any time related trend; (ii) inspecting residuals as a function of time; and (iii) correlation-lag plots.

Model selection was performed using the widely-employed procedure of stepwise simplification from the most complex possible model containing predictors identified using Random Forest, as outlined in Harrison et al. (2018), and implemented in R (MuMIn package; Barton, 2019, lme4 package; Bates et al., 2015). Non-significant

Table 1

Parameter	Relevance to <i>E. coli</i> concentration	References
Land-use	Different land-use types are often characterised by different source types (e.g. urban land with point sources, agricultural with diffuse), which can result in contrasting contaminant export patterns from regions of different land-use.	Rochelle-Newall et al., 2016 Selvakumar and Borst, 2006 Schreiber et al., 2015
Karstic hydrology	Different compartments in the karst system may be characterised by different conditions for survival (e.g. lack of sunlight in the underground compartments, different native microbial communities), or susceptibility to contamination (e.g. sinkholes being used for rubbish disposal, direct connectivity of springs with sink holes).	Personné et al., 1998 John and Rose, 2005 Personné et al., 2004 He et al., 2010 Lindsey et al., 2010
Catchment characteristics	Characteristics related to catchment size and dimensions, such as flow-path length, contributing catchment area, stream morphology, and stream order can affect how <i>E. coli</i> is transported, stored, exported from, or attenuated in catchments.	Harmel et al., 2010 McGuire et al., 2005
Rainfall, discharge	Rainfall and associated increase in discharge can result in delivery of <i>E. coli</i> to drainage networks from surface stores via overland flow, re-suspension of stream-bed and conduit hosted <i>E. coli</i> , and flooding/failure of septic systems/waste infrastructure	Buckerfield et al., 2019a McKergow and Davies-Colley, 2010
Relative humidity and air temperature	Temperature and relative humidity affect the growth and persistence of <i>E. coli</i> in the environment	McCarthy et al., 2012
Water temperature	Typically, increasing water temperature is associated with increasing mortality in laboratory studies. Higher temperatures in Summer may provide more favourable conditions for growth on land in some settings. Tracer of water residence time	Nasser and Oman, 1999 Badgley et al., 2019 Chen and Chang, 2014 Martin and Dean, 1999
Electrical conductivity	<ul style="list-style-type: none"> <li>High electrical conductivity associated with contaminated water including human sewage and agricultural run-off</li> <li>Contrasting electrical conductivity associated with different residence times, due to increasing levels of carbonate/mineral dissolution with increasing residence time</li> </ul>	de Sousa et al., 2014 Harwell et al., 2008 Toran et al., 2006
pH	Contrasting pH associated with different compartments of karst/varying residence times, e.g. higher pCO <sub>2</sub> in soil waters results in lower pH than water residing in the rock matrix with high levels of dissolved carbonate	Liu et al., 2004 Liu et al., 2007
Turbidity (as a proxy for suspended sediment)	<ul style="list-style-type: none"> <li><i>E. coli</i> is commonly associated with sediment, both suspended and stored in stream-beds.</li> <li>Increased turbidity due to re-suspension of stream bed sediment or influx of water from overland flow pathways under high discharge may be associated with higher <i>E. coli</i> concentration</li> </ul>	Rügner et al., 2013 Garcia-Aljaro et al., 2017 Kim et al., 2010

parameters were sequentially removed in order of least significance, and at each stage the more complex model was compared with the simplified model using parametric bootstrapping (number of simulations = 1000) (Halekoh and Højsgaard, 2014). A simplified model was used for the four sites with discharge data to test for the significance of discharge at the time of sampling as a predictor. A threshold *p*-value of 0.05 was used to determine whether the simpler model provided a significantly poorer fit (Shang and Cavanaugh, 2008). Akaike's Information Criterion (AIC) was used for the one case of non-nested model comparison (testing land use versus karst system as the categorical predictor) because parametric bootstraps cannot be used to compare non-nested models. A threshold *delta* AIC value of 7 was considered as evidence that two models produced significantly different fit (Fabozzi et al., 2014).

Several methods of representing spatial-autocorrelation were tested. The approach yielding the best model fit is described here and details of all tested approaches are given in SI Table S2. The influence of upstream sites on downstream sites was calculated as a function of the inverse of their hydrological separation distance and the observed *E. coli* concentration at upstream sites, then weighted by average *Q* at the catchment outlet on the sample date. The discharge weighting on each sample date was calculated as a fraction of the maximum observed *Q* on all sample dates, which was given a weighting of 1. The result is a concentration 'contribution' to each site from all other sites for each time point (which will be zero for non-flow connected sites), and which increases with *Q*.

### 3. Results

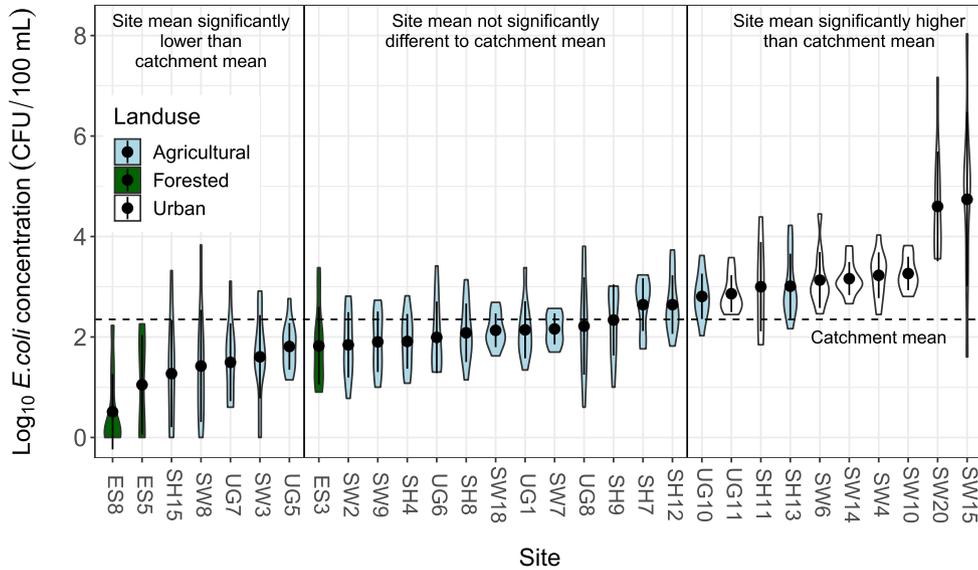
#### 3.1. Spatial and temporal variation in *E. coli* concentration

Variation in *E. coli* concentration was dominantly spatial; average variance by site was 0.57 log<sub>10</sub> CFU/100 mL (*n* = 30 sites), compared with 1.33 log<sub>10</sub> CFU/100 mL by date (*n* = 11 sample days). Mean values for individual sites varied by 4.29 log<sub>10</sub> CFU/100 mL (Figs. 1, 2), compared with variation between dates of 0.8 log<sub>10</sub> CFU/100 mL (Fig. 3). The catchment wide mean *E. coli* concentration taken across all sites

on the 19th September 2018 was significantly higher than the mean of all samples, and significantly lower on the 12th April. *E. coli* concentration ranged from below detection (<10 CFU/100 mL) at epikarst springs on forested hillslopes (e.g. ES8) to 8 log<sub>10</sub> CFU/100 mL at a contaminated site downstream of a sewage treatment plant (SW15). Samples from sites classified as 'urban' (*n* = 8) (see categorical representations of land use, SI Table S1) showed significantly higher *E. coli* concentrations than samples from agricultural sites (*n* = 19), and samples from forested sites (*n* = 3) were significantly lower than agricultural sites (SI Table S3). When classified according to compartment in the karst system (see categorical representation of position in the karst system, SI Table S1), epikarst springs were significantly lower in *E. coli* concentration than other compartments of the karst system, which were not significantly different.

#### 3.2. Meteorological, discharge, and water quality parameters

Meteorological conditions during the sampling period followed normal regional trends, with highest monthly precipitation in June (279 mm, 1982–2012 average 258 mm), and hottest average temperature in July (27.5, 1982–2012 average 23.6 °C) (Climate-Data.Org, 2012). The range in average temperature, relative humidity, and total rainfall during the 24 h prior to sampling captured for the sample dates in this study was 18.5–25.4 °C, 78–99.6%, and 0.0–9.4 mm, respectively. Total rainfall in the 24 h prior to sampling never exceeded 10% of the maximum daily rainfall that occurred during the study period (March–October 2018, SI Fig. S1). The maximum *Q* during the study period was also low relative to the range observed in this catchment (SI Fig. S2); *Q* at the four sites with pressure transducers was highest on the 19/9/2018 for all sites, reaching 57%, 11%, 54%, and 67% of maximum *Q* recorded during the study period at UG5, UG6, SW14, and UG1 respectively (Fig. 4). Discharge and turbidity were highest on 19/9/2018 for all sites (Fig. 4). There was a weak but significant negative correlation between pH and rainfall (repeated measures *r* = -0.27, *p* = 3.8 × 10<sup>-6</sup>), and a weak negative correlation between pH and EC (repeated measures *r* = -0.28, *p* = 0.015).



**Fig. 2.** Violin plots (which show data density), with mean and standard deviation (black circle, vertical bar within each violin) of *E. coli* concentration by site. Number of samples per site is 11.

Some water quality parameters showed clustering within the groups defined by categorical classification of sites according to surrounding land use or position in the karst system. Sample EC increased significantly in the order: forested sites < agricultural sites < urban sites, and epikarst springs < sinkholes/surface water sites < valley outlet springs (SI Table S3). Sample temperature was lowest at epikarst springs, higher in valley outlet springs, and highest in surface waters/sinkholes (all significantly different;  $p < 0.05$ ), and lower at forested sites than agricultural or urban sites. Turbidity was significantly lower at forested sites than urban or agricultural sites, and at epikarst springs than other components of the karst system ( $p < 0.05$ ).

### 3.3. Predictor variables selected using random Forest algorithm

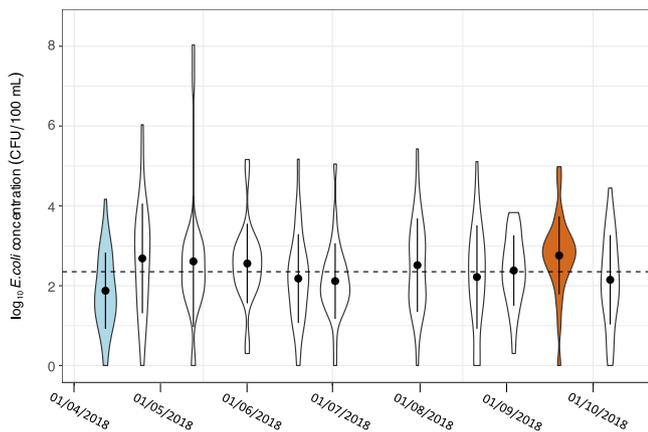
Land use variables produced the largest increase in mean squared error (MSE) with RF when randomly permuted (SI Fig. S3). Both continuous and categorical representations of land use produced a >90% increase in MSE. Variables describing the position of sites in the karst structure (categorical) and position in the catchment system (continuous) produced >50% increase in MSE, while water quality and rainfall

in the previous 24 or 48 h produced >20% increase. The remaining meteorological variables (temperature, relative humidity, solar radiation) resulted in a <20% increase. Variance Inflation Factors showed air temperature and solar radiation, and rainfall amount (over a 24 or 48 hour timeframe) and rainfall intensity to be collinear, so measures of these variables were not included simultaneously in any models.

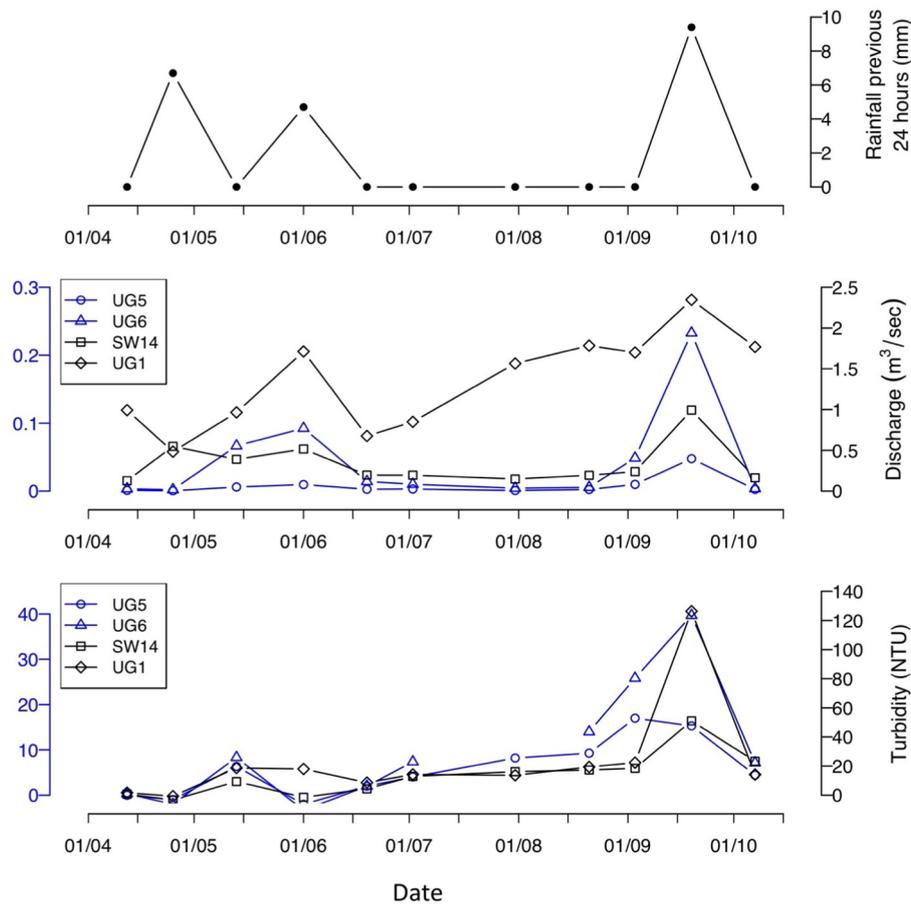
### 3.4. Linear mixed effects modelling results

The best performing model containing only significant predictors ( $p < 0.05$ ) incorporated land-use as a categorical predictor (with urban, forested, and agricultural land-use categories defined by dominant land-use in a 100 m radius), and rainfall and pH as continuous predictors (Table 2). Rainfall amount in the previous 24 and 48 h were both significant (not included simultaneously) and produced the same effect size. The effect size of rainfall (0.056) translates to a 1  $\text{log}_{10}$  unit increase in *E. coli* concentration per 18 mm rainfall, and the effect size of pH ( $-0.73$ ) translates to a 1.4  $\text{log}_{10}$  unit decrease in *E. coli* concentration per unit increase in pH. Site and date as random effects explained 27% and 5% of the variance not explained by fixed effects respectively. Including land use and karst hydrology simultaneously as categorical predictors resulted in non-convergent models. Although site type in the karst architecture (e.g. epikarst spring, sinkhole, etc.) was a significant predictor (categorical), models incorporating karst hydrology never performed as well as models incorporating land use as a predictor ( $\Delta \text{AIC} > 20$  for all karst models versus all land use models). Incorporating a fixed effect autocorrelation structure, where the contribution of *E. coli* from upstream sites to flow-connected downstream sites was calculated as a function of their hydrological separation distance, produced improvement on models without the spatial autocorrelation structure (SI Table S4). Weighting the autocorrelation structure by discharge at the outlet and scaling the whole contribution from upstream sites by a factor of 10 produced the best performing model, using parametric bootstrapping (threshold  $p < 0.05$ ) for model comparison (Halekoh and Højsgaard, 2014). For the four sites with pressure transducers, there was no relationship between *E. coli* concentration and Q but rainfall in the previous 24 h had a similar effect size as for the full dataset (0.06).

The predicted *E. coli* concentration, and 95% confidence interval, were calculated using the best performing model, and are visually depicted in Fig. 5 with the observational data. Observations are separated by land-use classification, and *E. coli* concentration is shown as a



**Fig. 3.** Violin plots (showing data density) and mean/standard deviation (black circle, vertical bar) for each sampling date. Number of samples per date is 30. Colouring indicates mean of all samples from given date is significantly different from mean of all samples (red: significantly higher, blue: significantly lower, dashed line: mean of all samples). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 4.** Selected meteorological and hydrological parameters through the sampling period (2018): (a) Rainfall, (b) discharge at four sites with pressure transducers, and (c) turbidity at same four sites as discharge in (b). Due to differences in scale, sites have been split between two axes for discharge and turbidity measurements.

function of rainfall in the previous 24 h, one of the two temporal predictors found to be significant.

#### 4. Discussion

The high-resolution spatial dataset collected in this karst terrain study spans a full paddy rice crop cycle and the monsoon, including the transitions between dry and wet seasons. The higher spatial than temporal variability in *E. coli* concentration suggests spatial controls are most relevant to average faecal contamination levels in this mixed land-use setting. Modelling identified land-use and individual site characteristics to be the most significant controls on average *E. coli* concentration, regardless of location of the sampled water source in the catchment. However, periods of higher catchment-wide *E. coli* concentrations corresponding with rainfall in the preceding days, and discharge of paddy fields, suggest these temporal controls (rainfall, and particular agricultural activities) can become more important controls of catchment-wide *E. coli* concentrations than immediate surrounding

land-use for transient periods. Faecal contamination of karst water resources is common worldwide and the *E. coli* concentrations observed in this study are comparable to concentrations observed in this karst region in China and other karst regions of the world with human or livestock inputs (He et al., 2016, Heinz et al., 2009, Howell et al., 1995, Lan et al., 2014, Sinreich et al., 2013). Drinking water guidelines set by the World Health Organization specify that zero faecal coliforms should be present in a 100 mL sample, although this is frequently not achieved and low concentrations can present a sufficiently low health risk (World Health Organization, 2017). The cause for concern lies in the high concentrations observed at selected sites, and the lack of access to an alternative supply, a problem still encountered across much of rural China (Liu, 2015).

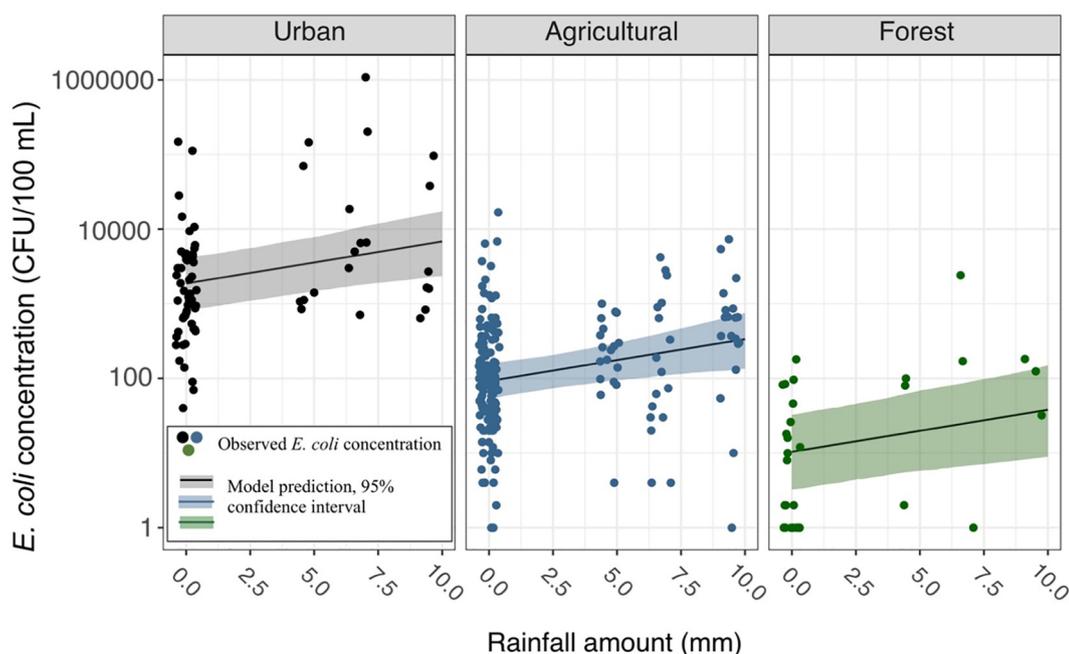
##### 4.1. Spatial controls

The consistently high *E. coli* concentrations observed at urban sites suggest urban land is a chronic source in this region. Urban land is a

**Table 2**

Parameters for the best performing model based on parametric bootstrapping (PB) model selection. The continuous predictors (rainfall amount and sample pH) were scaled in the model; the coefficient back-transformed to the original parameter space is given. The intercept is for an arbitrarily chosen level of the categorical predictor (agricultural land, in this model), and intercept deviations from this level are given for the other categories of land-use (forested and urban).

Fixed effects				Random effects			
	Value	Std. error	PB <i>p</i> -value	Groups	Variance	Std. dev	% variance explained
Intercept ( $\log_{10}$ CFU/100 mL)	2.07	0.12	<0.001	Site	0.17	0.41	27%
Land use (forest) ( $\log_{10}$ CFU/100 mL)	-0.95	0.28	0.003	Date	0.02	0.16	5%
Land use (urban) ( $\log_{10}$ CFU/100 mL)	1.31	0.20	<0.001	Residual	0.42	0.65	
Sample pH	-0.73	0.05	0.001				
Rain in previous 24 h (mm)	0.056	0.06	0.01				



**Fig. 5.** Observed *E. coli* concentration (coloured circles, jittered to show data density), predicted values from best performing model (solid lines), and lower and upper bounds of 95% confidence interval (shaded regions). Rainfall amount refers to rainfall in the proceeding 24 h, as implemented as a predictor in the model. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

major source of faecal contamination to catchment drainage systems under both low and high discharge conditions in many settings, although this may not be solely a result of human inputs (Sauer et al., 2011; Templar et al., 2016; Young and Thackston, 1999). Scavenging dogs are prevalent in residential areas, and water buffalo, used in farm labour, are typically housed in villages. Thus, multiple sources are likely to contribute to an *E. coli* signal in water draining urban environments (Ahmed et al., 2008; Suprihatin et al., 2003; Whitlock et al., 2002). Sites located in agricultural land were also consistently contaminated, implicating agricultural drainage waters are a source of *E. coli*. Crop fertilisation with organic fertiliser is widespread in this region, and is likely to be a major source of faecal contamination. Farmers are severely limited by labour availability, and encouragement to plan manure management for pathogen inactivation (such as use of digesters) may not occur. Buffalo and horses grazing in agricultural land or drinking near settlements may contribute a significant faecal source particularly if defecating in or around waterways (O'Callaghan et al., 2019), which could potentially be reduced by fencing off waterways (Muirhead, 2019). Containing and treating human waste would likely produce the single biggest improvement in microbial water quality in this region, but measures such as containment of waste away from watercourses would be necessary to address the problem of scavenging animals. Further research is needed to identify which *E. coli* sources in agricultural land are responsible for most export, and employing source tracking methods could provide valuable information on which species are relevant to faecal contamination across land-use types.

The lack of significant differences between sites located within different compartments of the karst system (except for epikarst springs) could be a result of the connectivity between the underground and surface water systems, a feature of well-developed karst hydrology (Ford and Williams, 2007). Relevant to both human and animal sources is the thin soil cover, rapid infiltration rates, and presence of sink holes that result from collapse of the karstic bedrock of this region. When sources are present around sinkholes, they present critical source areas (CSAs), where high sources of pollutants coincide with high potential for hydrological transfer (Heathwaite et al., 2005). Karst rocky desertification resulting in thin or absent soil cover, a major form of environmental

degradation in this region, may also enhance transfer of sediment and contaminants from the land surface to receiving waters by increasing runoff ratios (Dai et al., 2017). Buffers strips around sinkholes may reduce some of these enhanced risks posed by the karstic hydrology. In addition, all forested sites were classified as epikarst springs, meaning both factors related to land-use and hydrology may be relevant to the significantly lower *E. coli* concentrations at these sites. The low density or absence of people and livestock in such areas, and scarcity of wildlife more generally in the region (Zhang et al., 2008), would be expected to result in low risk of *E. coli* contamination (Porter et al., 2017), but the limited number of forested/epikarst sites (three) also places larger uncertainty around how representative the sites are of these categories.

The superior performance of models incorporating a spatial autocorrelation structure supports the need for continued improvement in representation of flow-connectivity between sample sites if we are to correctly attribute pollution source contributions. Discharge weighting of the spatial autocorrelation structure has been shown to improve model predictions of faecal coliforms in larger catchment systems (Jat and Serre, 2018). Indeed, tracer-tests have shown that flow connectivity is important for pollutant transfer at sites along more major tributaries in the lower reaches of this catchment under low  $Q$ , and becomes relevant to the whole catchment under conditions of high  $Q$  (Barna et al., 2020). Processes promoting *E. coli* die-off (e.g. U.V. radiation, predation) or sequestration (e.g. in sediments) during low discharge conditions will also counteract the progressive loading moving down drainage channels through the catchment.

#### 4.2. Temporal controls

Incidental events are likely to impact on short term *E. coli* concentration. The only date on which catchment wide *E. coli* concentration was significantly higher was during discharge of paddy fields, and combined with the elevated turbidity and  $Q$  at the four sites with pressure transducers, this suggests temporal changes in land-use due to agricultural cycles influence faecal contamination levels. Although there was rainfall in the 24 (and 48) hours prior to sampling, it was comparable to the 25th April, when paddy fields were not being discharged, and the rainfall produced minimal increase in  $Q$  and turbidity. Livestock manure is used

extensively to fertilise paddy (and other) crops, and *E. coli* may be capable of lengthy survival in the sediments. The faecal matrix would provide a protective niche from sunlight, and association with sediment once fields are flooded may provide nutrients and shelter from predation (Garzio-Hadzick et al., 2010; Jamieson et al., 2004). There is hence potential for *E. coli* remobilisation when rice is harvested and the crops are drained (Buckerfield et al., 2019b). Seasonality in *E. coli* export from agricultural land has been observed in similar agricultural – socioeconomic – climatic settings (Rochelle-Newall et al., 2016), and a range of other catchments in different agricultural-climatic settings (Kay et al., 2008; Tetzlaff et al., 2012). A multi-year sampling campaign, across multiple catchments and countries, would allow for this seasonality to be statistically tested and strengthen conclusions about the effects of agricultural cycles. Events in the cultural calendar, as well as agricultural, may also impact on short-term *E. coli* concentrations in rural settings where sewage infrastructure is lacking. Chinese New Year, for example, involves mass human migration from cities back to rural home-towns (Dou and Miao, 2017), which will result in a short term increase in sewage production.

The association of antecedent 24/48 h rainfall with significantly higher *E. coli* concentration across all land-use types is consistent with *E. coli* dynamics observed during a targeted study on rainfall-driven effects in this catchment (Buckerfield et al., 2019a), and most other studies assessing the effect of rainfall events, or high Q, on FIO export (Crowther et al., 2002; Olds et al., 2018). It is imperative to combine low and high Q sampling to improve catchment models and export estimates. Some studies in rural catchments that have targeted both low and high-flow conditions have estimated that over 90% of annual *E. coli* export occurs during high Q periods (Davies-Colley et al., 2008; McKergow and Davies-Colley, 2010).

The significant decrease in *E. coli* concentration with increasing pH could be a result of increasing mortality or growth inhibition with high pH water (Parhad and Rao, 1974). However, the pH values observed in this catchment were only slightly alkaline (average across all sites = 8.01), and are unlikely to result in significant reduction in *E. coli* survival rates (McFeters and Stuart, 1972). The high pH may reflect water that has resided in the karst system for longer, thus more opportunity for *E. coli* concentrations to decline via processes such as die-off or sedimentation. The weak but significant negative correlation between pH and rainfall provides some evidence that pH decreases following rainfall, a relationship often observed due to the lower pH of rainfall than water that has been resident in karst (Yang et al., 2012), and the weak negative correlation with EC could be a result of fertiliser washed off during rainfall. Constraining water residence times using, for example, stable water isotopes would help elucidate whether pH is affecting *E. coli* survival, or solely a feature of water derived from particular hydrological pathways.

#### 4.3. Further factors affecting variation in *E. coli* concentration

The substantial variance explained by 'Site' as a random predictor in all models, and corresponding improvement in model fit, indicates that there are site characteristics not explained by the tested representations of spatial predictors. This could be a result of several factors, including: (a) variation in hydrological connectivity between *E. coli* sources and sampling sites due to variable levels of infrastructure, such as the presence of impermeable concrete surfaces built to facilitate access to sink-holes at some sites (Murphy et al., 2015); (b) use of sampling sites for different purposes (e.g. drinking water collection with clean vessels, or livestock watering points), and (c) inadequate representation of the heterogeneity of the karst system, a ubiquitous problem in modelling of karst hydrology (Hartmann et al., 2014). Representation of these attributes could be explored through mechanisms such as (a) introducing a metric describing the vulnerability of the surface surrounding the site to overland flow, (b) surveying what different sites are used for, and (c) extraction of aquifer properties (such as infiltration rate) from available hydrological models.

#### 4.4. Transferability across rural developing regions

This case study was conducted in one of the poorest regions of China, where the key drivers of poverty are resource scarcity, ecosystem degradation, mountainous topography and remoteness, and population pressure (Liu et al., 2017). Evidence suggests the most effective measures for improving microbial water quality and reducing microbial waterborne disease depends on several factors including the level of poverty (Ashbolt, 2004). Although the exact agricultural, hydrological, climatic, and socio-economic characteristics of this study area will not be shared with other developing regions, the general findings can be used to inform on the most likely causes of poor microbial water quality in rural farming districts under comparable development pressures. For example, the wider southwest China karst region, which consistently suffers from high rural poverty rates (Jiang et al., 2014), and neighbouring Vietnam, which shares the vulnerability to water contamination imposed by karstic bedrock, and similar land-use pressures from agriculture and population (Ender et al., 2018; Tuyet, 2001). Further, although China has achieved unprecedented rates of poverty alleviation, the story of economic growth bringing uneven prosperity, characteristically leaving behind rural populations, is one shared with other countries, such as Indonesia, Brazil, Mexico, and Bangladesh (Jalan and Ravallion, 2002; Yang et al., 2015). With the World Bank estimating that 736 million people still live in poverty, 79% of those people being in rural regions (World Bank, 2018), it is imperative that we focus on developing solutions that are practical for the characteristics of rural populations, with stronger mechanisms for co-producing science at the policy-practice interface (Zheng et al., 2019).

#### CRediT authorship contribution statement

**Sarah J. Buckerfield:** Conceptualization, Data curation, Methodology, Formal analysis, Investigation, Project administration, Writing - original draft, Writing - review & editing. **Richard S. Quilliam:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing. **Luc Bussiere:** Conceptualization, Formal analysis, Writing - review & editing. **Susan Waldron:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing. **Larissa A. Naylor:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing. **Siliang Li:** Conceptualization, Funding acquisition, Resources, Supervision, Writing - review & editing. **David M. Oliver:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

The authors would like to thank: Dr. Peng Tao, Guiyang Geochemistry Institute, Chinese Academy of Sciences for helping facilitate catchment access and accommodation, Dr. Fu-Jun Yue and Dr. Zhongjun Wang for assistance with laboratory access and logistics, Floyd Howard for tremendous help with sample collection and general support, and Cheng Qianyun and Wang Yanwei for general support.

#### Funding sources

This research was funded by the Natural Environment Research Council as part of the IAPETUS Doctoral Training Programme (NE/L002590/1) and has received additional assistance from NERC and the Chinese Academy of Sciences through the China-UK Critical Zone

project NE/N007425/1, and the National Natural Science Foundation of China (Grant No. 41571130072).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2020.140898>.

## References

- Ahmed, W., Powell, D., Goonetilleke, A., Gardner, T., 2008. Detection and source identification of faecal pollution in non-sewered catchment by means of host-specific molecular markers. *Water Sci. Technol.* 58, 579–586.
- Harmel, R.D., Karthikeyan, R., Gentry, T., Srinivasan, R., 2010. Effects of agricultural management, land use, and watershed scale on *E. coli* concentrations in runoff and streamflow. *Trans. ASABE* 53, 1833–1841.
- Ashbolt, N.J., 2004. Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology* 198, 229–238.
- Bain, R., Cronk, R., Wright, J., Yang, H., Slaymaker, T., Bartram, J., 2014. Faecal contamination of drinking-water in low- and middle-income countries: a systematic review and meta-analysis. *PLoS Med.* 11, e1001644.
- Badgley, B.D., Steele, M.K., Cappellin, C., Burger, J., Jian, J., Neher, T.P., et al., 2019. Faecal indicator dynamics at the watershed scale: Variable relationships with land use, season, and water chemistry. *Sci. Total Environ.* 134113.
- Bakdash, J.Z., Marusich, L.R., 2017. Repeated measures correlation. *Front. Psychol.* 8.
- Barna, J.M., Fryar, A.E., Cao, L., Currens, B.J., Peng, T., Zhu, C., 2020. Variability in groundwater flow and chemistry in the Houzhai Karst Basin, Guizhou Province, China. *Environ. Eng. Geosci.*
- Barton, K., 2019. *MuMIn: Multi-model Inference*. R Package Version 1.43.6. R Package Version 1.43.6.
- Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 1 (1), 2015.
- Bivins, A.W., Sumner, T., Kumpel, E., Howard, G., Cumming, O., Ross, I., et al., 2017. Estimating infection risks and the global burden of diarrheal disease attributable to intermittent water supply using QMRA. *Environmental Science & Technology* 51, 7542–7551.
- Brusseau, M.L., Ramirez-Andreotta, M., Pepper, I.L., Maximillian, J., 2019. Chapter 26 - environmental impacts on human health and well-being. In: Brusseau, M.L., Pepper, I.L., Gerba, C.P. (Eds.), *Environmental and Pollution Science*, Third edition Academic Press, pp. 477–499.
- Buckerfield, S.J., Quilliam, R.S., Waldron, S., Naylor, L., Li, S., Oliver, D.M., 2019a. Rainfall-driven *E. coli* transfer to the stream-conduit network observed through increasing spatial scales in mixed land-use paddy farming karst terrain. *Water Research X* 5, 100038.
- Buckerfield, S.J., Waldron, S., Quilliam, R.S., Naylor, L.A., Li, S., Oliver, D.M., 2019b. How can we improve understanding of faecal indicator dynamics in karst systems under changing climatic, population, and land use stressors? – research opportunities in SW China. *Sci. Total Environ.* 646, 438–447.
- Cao, J., Yuan, D., Tong, L., Mallik, A., Yang, H., Huang, F., 2015. An overview of karst ecosystem in Southwest China: current state and future management. *Journal of Resources and Ecology* 6, 247–256.
- Chen, H.J., Chang, H., 2014. Response of discharge, TSS, and *E. coli* to rainfall events in urban, suburban, and rural watersheds. *Environ. Sci.: Process. Impacts* 16, 2313–2324.
- Cho, K.H., Pachepsky, Y.A., Oliver, D.M., Muirhead, R.W., Park, Y., Quilliam, R.S., et al., 2016. Modeling fate and transport of fecally-derived microorganisms at the watershed scale: state of the science and future opportunities. *Water Res.* 100, 38–56.
- Climate-Data.Org, 2012. Climate Anshun, 2019.
- Crowther, J., Kay, D., Wyer, M.D., 2002. Faecal-indicator concentrations in waters draining lowland pastoral catchments in the UK: relationships with land use and farming practices. *Water Res.* 36, 1725–1734.
- Dai, Q., Peng, X., Yang, Z., Zhao, L., 2017. Runoff and erosion processes on bare slopes in the Karst Rocky Desertification Area. *CATENA* 152, 218–226.
- Davies-Colley, R., Nagels, J., Lydiard, E., 2008. Stormflow-dominated loads of faecal pollution from an intensively dairy-farmed catchment. *Water Sci. Technol.* 57, 1519–1523.
- de Sousa, D.N.R., Mozeto, A.A., Carneiro, R.L., Fadini, P.S., 2014. Electrical conductivity and emerging contaminant as markers of surface freshwater contamination by wastewater. *Sci. Total Environ.* 484, 19–26.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36, 27–46.
- Dou, J., Miao, S., 2017. Impact of mass human migration during Chinese New Year on Beijing urban heat island. *Int. J. Climatol.* 37, 4199–4210.
- Dymond, J.R., Serezat, D., Ausseil, A.G., Muirhead, R.W., 2016. Mapping of *Escherichia coli* sources connected to waterways in the Ruamahanga catchment, New Zealand. *Environ. Sci. Technol.* 50, 1897–1905.
- Edberg, S.C., Rice, Ew, Karlin, R.J., Allen, M.J., 2000. *Escherichia coli*: The Best Biological Drinking Water Indicator for Public Health Protection.
- Ender, A., Goepfert, N., Goldscheider, N.J.H.J., 2018. Hydrogeological Controls of Variable Microbial Water Quality in a Complex Subtropical Karst System in Northern Vietnam. 26 pp. 2297–2314.
- Environment Agency. *The Microbiology of Drinking Water (2009) – Part 4 – Methods for the isolation and enumeration of coliform bacteria and Escherichia coli (including E. coli O157:H7)*. Available online: <[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/755563/MoDWPpart4-223MAYH.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/755563/MoDWPpart4-223MAYH.pdf)>. Accessed 10/02/2020.
- Fabozzi, F.J., Focardi, S.M., Rachev, S.T., Arshanapalli, B.G., 2014. Appendix E Model Selection Criterion: AIC and BIC. *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. John Wiley & Sons, Inc.
- Ford, D., Williams, P., 2007. *Karst Hydrogeology and Geomorphology*. Chichester, West Sussex, John Wiley and Sons, Ltd.
- García-Aljaro, C., Martín-Díaz, J., Vinas-Balada, E., Calero-Caceres, W., Lucena, F., Blanch, A.R., 2017. Mobilisation of microbial indicators, microbial source tracking markers and pathogens after rainfall events. *Water Res.* 112, 248–253.
- Garzio-Hadzick, A., Shelton, D.R., Hill, R.L., Pachepsky, Y.A., Guber, A.K., Rowland, R., 2010. Survival of manure-borne *E. coli* in streambed sediment: effects of temperature and sediment properties. *Water Res.* 44, 2753–2762.
- Halekoh, U., Hojsgaard, S., 2014. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *J. Stat. Softw.* 59, 1–30.
- Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., et al., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6, e4794.
- Harwell, M.C., Surratt, D.D., Barone, D.M., Aumen, N.G., 2008. Conductivity as a tracer of agricultural and urban runoff to delineate water quality impacts in the northern Everglades. *Environ. Monit. Assess.* 147, 445–462.
- Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., Weiler, M., 2014. Karst water resources in a changing world: review of hydrological modeling approaches. *Rev. Geophys.* 52, 218–242.
- He, Q., Yang, P., Yuan, W., Jiang, Y., Pu, J., Yuan, D., et al., 2010. The use of nitrate, bacteria and fluorescent tracers to characterize groundwater recharge and contamination in a karst catchment, Chongqing, China. *Hydrogeol. J.* 18, 1281–1289.
- He, Q., Qiu, S., Jiang, Y., Wu, Z., Liu, Z., 2016. Land-use change caused microbial pollution in a karst underground river, Chongqing, China. *Environ. Earth Sci.* 75, 709.
- Heathwaite, A.L., Quinn, P.F., Hewett, C.J.M., 2005. Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. *J. Hydrol.* 304, 446–461.
- Heinz, B., Birk, S., Liedl, R., Geyer, T., Straub, K.L., Andresen, J., et al., 2009. Water quality deterioration at a karst spring (Gallusquelle, Germany) due to combined sewer overflow: evidence of bacterial and micro-pollutant contamination. *Environ. Geol.* 57, 797–808.
- Hong, H., Qiu, J., Liang, Y., 2010. Environmental factors influencing the distribution of total and fecal coliform bacteria in six water storage reservoirs in the Pearl River Delta Region, China. *J. Environ. Sci.* 22, 663–668.
- Howell, J.M., Coyne, M.S., Cornelius, P., 1995. Fecal bacteria in agricultural waters of the Bluegrass region of Kentucky. *J. Environ. Qual.* 24 (3), 411 (Print).
- Jalan, J., Ravallion, M., 2002. Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China. 17 pp. 329–346.
- Jamieson, R., Gordon, R., Joy, D., Lee, H., 2004. Assessing microbial pollution of rural surface waters: a review of current watershed scale modeling approaches. *Agric. Water Manag.* 70, 1–17.
- Jat, P., Serre, M.L., 2018. A novel geostatistical approach combining Euclidean and gradual-flow covariance models to estimate fecal coliform along the Haw and Deep rivers in North Carolina. *Stoch. Env. Risk A.* 32, 2537–2549.
- Jiang, Z., Lian, Y., Qin, X., 2014. Rocky desertification in Southwest China: impacts, causes, and restoration. *Earth Sci. Rev.* 132, 1–12.
- Kay, D., Crowther, J., Stapleton, C.M., Wyer, M.D., Fewtrell, L., Anthony, S., et al., 2008. Faecal indicator organism concentrations and catchment export coefficients in the UK. *Water Res.* 42, 2649–2661.
- Kim, J.-W., Pachepsky, Y.A., Shelton, D.R., Coppock, C., 2010. Effect of streambed bacteria release on *E. coli* concentrations: monitoring and modeling with the modified SWAT. *Ecol. Model.* 221, 1592–1604.
- Lan, J., Yang, P., Ren, K., Chen, X., Xu, X., Nu, N., 2014. Investigation of nitrogen, phosphorus and microbial contamination in Laolongdong underground river system of Chongqing. *Huan Jing Ke Xue* 35, 1297–1303.
- Liaw, A., Wiener, M.C., 2007. Classification and regression by randomForest. *R News* 2, 18–22.
- Lindsey, B.D., Katz, B.G., Berndt, M.P., Ardis, A.F., Skach, K.A., 2010. Relations between sink-hole density and anthropogenic contaminants in selected carbonate aquifers in the eastern United States. *Environ. Earth Sci.* 60, 1073–1090.
- Liu, Hongqiao, 2015. China's Long March to Safe Drinking Water: China Water Risk & Chinadiologue (Print).
- Liu, Z., Groves, C., Yuan, D., Meiman, J., 2004. South China karst aquifer storm-scale hydrochemistry. *Ground Water* 42, 491–499.
- Liu, Z., Li, Q., Sun, H., Wang, J., 2007. Seasonal diurnal and storm-scale hydrochemical variations of typical epikarst springs in subtropical karst areas of SW China: Soil CO<sub>2</sub> and dilution effects. *J. Hydrol.* 337, 207–223.
- Liu, Y., Liu, J., Zhou, Y., 2017. Spatio-temporal patterns of rural poverty in China and targeted poverty alleviation strategies. *J. Rural. Stud.* 52, 66–75.
- Martin, J.B., Dean, R.W., 1999. Temperature as a natural tracer of short residence times for groundwater in karst aquifers. In: Palmer, A.N., Palmer, V.P., Sasowsky, I.D. (Eds.), *Karst Modelling*. Karst Waters Institute, Charlottesville, Virginia, pp. 236–242 Special Publication 5.
- McCarthy, D.T., Hathaway, J.M., Hunt, W.F., Deletic, A., 2012. Intra-event variability of *Escherichia coli* and total suspended solids in urban stormwater runoff. *Water Res.* 46, 6661–6670.
- McFeters, G.A., Stuart, D.G., 1972. Survival of coliform bacteria in natural waters: field and laboratory studies with membrane-filter chambers. *Appl. Microbiol.* 24, 805–811.

- McGuire, K.J., McDonnell, J.J., Weiler, M., Kendall, C., McGlynn, B.L., Welker, J.M., et al., 2005. The role of topography on catchment-scale water residence time. *Water Resour. Res.* 41.
- McKergow, L.A., Davies-Colley, R.J., 2010. Stormflow dynamics and loads of *Escherichia coli* in a large mixed land use catchment. *Hydrol. Process.* 24, 276–289.
- McNeish, D., Kelley, K., 2018. Fixed effects models versus mixed effects models for clustered data: reviewing the approaches, disentangling the differences, and making recommendations. *Psychol. Methods* 24 (1), 20–35.
- Muirhead, R.W., 2019. The effectiveness of streambank fencing to improve microbial water quality: a review. *Agric. Water Manag.* 223, 105684.
- Murphy, S., Jordan, P., Mellander, P.E., O'Flaherty, V., 2015. Quantifying faecal indicator organism hydrological transfer pathways and phases in agricultural catchments. *Sci. Total Environ.* 520, 286–299.
- Nasser, A.M., Oman, S.D., 1999. Quantitative assessment of the inactivation of pathogenic and indicator viruses in natural water sources. *Water Res.* 33, 1748–1752.
- Neill, A.J., Tetzlaff, D., Strachan, N.J.C., Hough, R.L., Avery, L.M., Watson, H., et al., 2018. Using spatial-stream-network models and long-term data to understand and predict dynamics of faecal contamination in a mixed land-use catchment. *Sci. Total Environ.* 612, 840–852.
- O'Callaghan, P., Kelly-Quinn, M., Jennings, E., Antunes, P., O'Sullivan, M., Fenton, O., et al., 2019. The Environmental Impact of Cattle Access to Watercourses: A Review. 48 pp. 340–351.
- Olds, H.T., Corsi, S.R., Dila, D.K., Halmo, K.M., Bootsma, M.J., McLellan, S.L., 2018. High levels of sewage contamination released from urban areas after storm events: a quantitative survey with sewage specific bacterial indicators. *PLoS Med.* 15, e1002614.
- Oliver, D.M., Bartie, P.J., Louise Heathwaite, A., Reaney, S.M., Parnell, J.A.Q., Quilliam, R.S., 2018. A catchment-scale model to predict spatial and temporal burden of *E. coli* on pasture from grazing livestock. *Sci. Total Environ.* 616–617, 678–687.
- Oliver, D.M., Zheng, Y., Naylor, L.A., Murtagh, M., Waldron, S., Peng, T., 2020. How does smallholder farming practice and environmental awareness vary across village communities in the karst terrain of southwest China? *Agric. Ecosyst. Environ.* 288, 106715.
- Parhad, N.M., Rao, N.U., 1974. Effect of pH on survival of *Escherichia coli*. *Journal (Water Pollution Control Federation)* 46, 980–986.
- Personné, J.C., Poty, F., Mahler, B.J., Drogue, C., 2004. Colonization by Aerobic Bacteria in Karst: Laboratory and In Situ Experiments. *Groundwater* 42, 526–533.
- Personné, J.C., Poty, F., Vaute, L., Drogue, C., 1998. Survival, transport and dissemination of *Escherichia coli* and enterococci in a fissured environment. Study of a flood in a karstic aquifer. *J. Appl. Microbiol.* 84, 431–438.
- Porter, K.D.H., Reaney, S.M., Quilliam, R.S., Burgess, C., Oliver, D.M., 2017. Predicting diffuse microbial pollution risk across catchments: the performance of SCIMAP and recommendations for future development. *Sci. Total Environ.* 609, 456–465.
- Pronk, M., Goldscheider, N., Zopfi, J., 2006. Dynamics and interaction of organic carbon, turbidity and bacteria in a karst aquifer system. *Hydrogeol. J.* 14, 473–484.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rochelle-Newall, E.J., Ribolzi, O., Viguier, M., Thammahacksa, C., Silvera, N., Latschack, K., et al., 2016. Effect of land use and hydrological processes on *Escherichia coli* concentrations in streams of tropical, humid headwater catchments. *Sci. Rep.* 6, 32974.
- Rügner, H., Schwientek, M., Beckingham, B., Kuch, B., Grathwohl, P., 2013. Turbidity as a proxy for total suspended solids (TSS) and particle facilitated pollutant transport in catchments. *Environ. Earth Sci.* 69, 373–380.
- Sauer, E.P., VandeWalle, J.L., Bootsma, M.J., McLellan, S.L., 2011. Detection of the human specific *Bacteroides* genetic marker provides evidence of widespread sewage contamination of stormwater in the urban environment. *Water Res.* 45, 4081–4091.
- Schiperski, F., Zirlwagen, J., Scheytt, T., 2016. Transport and attenuation of particles of different density and surface charge: a karst aquifer field study. *Environmental Science & Technology* 50, 8028–8035.
- Schreiber, C., Rechenburg, A., Rind, E., Kistemann, T., 2015. The impact of land use on microbial surface water pollution. *Int. J. Hyg. Environ. Health* 218, 181–187.
- Selvakumar, A., Borst, M., 2006. Variation of microorganism concentrations in urban stormwater runoff with land use and seasons. *J. Water Health* 4, 109–124.
- Shang, J., Cavanaugh, J.E., 2008. Bootstrap variants of the Akaike information criterion for mixed model selection. *Computational Statistics & Data Analysis* 52, 2004–2021.
- Sinclair, A., Hebb, D., Jamieson, R., Gordon, R., Benedict, K., Fuller, K., et al., 2009. Growing season surface water loading of fecal indicator organisms within a rural watershed. *Water Res.* 43, 1199–1206.
- Sinreich, M., Pronk, M., Kozel, R., 2013. Microbiological monitoring and classification of karst springs. *Environ. Earth Sci.* 1–10.
- Stevanović, Z., 2018. Global Distribution and Use of Water From Karst Aquifers. 466 pp. 217–236.
- Suprihatin, I., Fallowfield, H., Bentham, R., Cromar, N., 2003. Determination of faecal pollutants in Torrens and Patawalonga catchment waters in South Australia using faecal sterols. *Water Sci. Technol.* 47, 283–289.
- Templar, H.A., Dila, D.K., Bootsma, M.J., Corsi, S.R., McLellan, S.L., 2016. Quantification of human-associated fecal indicators reveal sewage from urban watersheds as a source of pollution to Lake Michigan. *Water Res.* 100, 556–567.
- Tetzlaff, D., Capell, R., Soulsby, C., 2012. Land use and hydroclimatic influences on faecal indicator organisms in two large Scottish catchments: towards land use-based models as screening tools. *Sci. Total Environ.* 434, 110–122.
- Toran, L., Tancredi, J.H., Herman, E.K., White, W.B., 2006. Conductivity and sediment variation during storms as evidence of pathways to karst springs. In: Harmon, R.S., Wicks, C. (Eds.), Perspectives on karst geomorphology, hydrology, and geochemistry - A tribute volume to Derek C. Ford and William B. White. Geological Society of America, pp. 169–176.
- Tuyet, D., 2001. Characteristics of Karst Ecosystems of Vietnam and their Vulnerability to Human Impact. 75 pp. 325–329.
- UNICEF, WHO, 2019. Progress on Household Drinking Water, Sanitation and Hygiene 2000–2017: Special Focus on Inequalities, New York.
- White, W.B., 2018. Contaminant transport in karst aquifers: systematics and mechanisms. In: White, W.B., Herman, J.S., Herman, E.K., Rutigliano, M. (Eds.), Karst Groundwater Contamination and Public Health. Springer International Publishing, Cham, pp. 55–81.
- Whitlock, J.E., Jones, D.T., Harwood, V.J., 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. *Water Res.* 36, 4273–4282.
- World Bank, 2018. Poverty and Shared Prosperity 2018: Piecing Together the Poverty Puzzle. World Bank, World Bank, Washington D.C.
- World Health Organization, 2017. Guidelines for Drinking-water Quality. Fourth edition. Incorporating the First Addendum, Geneva, p. 149 (Web).
- Xue, F., Tang, J., Dong, Z., Shen, D., Liu, H., Zhang, X., et al., 2018. Tempo-spatial controls of total coliform and *E. coli* contamination in a subtropical hilly agricultural catchment. *Agric. Water Manag.* 200, 10–18.
- Yang, R., Liu, Z., Zeng, C., Zhao, M., 2012. Response of epikarst hydrochemical changes to soil CO<sub>2</sub> and weather conditions at Chenqi, Puding, SW China. *J. Hydrol.* 468–469, 151–158.
- Yang, G.-J., Utzinger, J., Zhou, X.-N., 2015. Interplay between environment, agriculture and infectious diseases of poverty: case studies in China. *Acta Trop.* 141, 399–406.
- Yang, Y., Zhou, Y.-B., Xiao, P.-L., Shi, Y., Chen, Y., Liang, S., et al., 2017. Prevalence of and risk factors associated with *Cryptosporidium* infection in an underdeveloped rural community of southwest China. *Infectious Diseases of Poverty* 6, 2.
- Ye, B., Yang, L., Li, Y., Wang, W., Li, H., 2013. Water sources and their protection from the impact of microbial contamination in rural areas of Beijing, China. *Int. J. Environ. Res. Public Health* 10, 879–891.
- Young, K.D., Thackston, E.L., 1999. Housing Density and Bacterial Loading in Urban Streams. 125 pp. 1177–1180.
- Zhang, L., Hua, N., Sun, S., 2008. Wildlife trade, consumption and conservation awareness in southwest China. *Biodivers. Conserv.* 17, 1493–1516.
- Zhang, S.-X., Zhou, Y.-M., Xu, W., Tian, L.-G., Chen, J.-X., Chen, S.-H., et al., 2016. Impact of co-infections with enteric pathogens on children suffering from acute diarrhea in southwest China. *Infectious Diseases of Poverty* 5, 64.
- Zhang, Z., Chen, X., Soulsby, C., 2017. Catchment-scale conceptual modelling of water and solute transport in the dual flow system of the karst critical zone. *Hydrol. Process.* 31, 3421–3436.
- Zheng, Y., Naylor, L.A., Waldron, S., Oliver, D.M., 2019. Knowledge management across the environment-policy interface in China: what knowledge is exchanged, why, and how is this undertaken? *Environ. Sci. Pol.* 92, 66–75.