



A data-based predictive model for spatiotemporal variability in stream water quality

Danlu Guo¹, Anna Lintern^{1,2}, J. Angus Webb¹, Dongryeol Ryu¹, Ulrike Bende-Michl³, Shuci Liu¹, and Andrew William Western¹

¹Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC, Australia

²Department of Civil Engineering, Monash University, Clayton, VIC, Australia

³Bureau of Meteorology, Parkes, ACT, Australia

Correspondence: Danlu Guo (danlu.guo@unimelb.edu.au)

Received: 5 July 2019 – Discussion started: 23 July 2019

Revised: 15 December 2019 – Accepted: 20 January 2020 – Published: 24 February 2020

Abstract. Our current capacity to model stream water quality is limited – particularly at large spatial scales across multiple catchments. To address this, we developed a Bayesian hierarchical statistical model to simulate the spatiotemporal variability in stream water quality across the state of Victoria, Australia. The model was developed using monthly water quality monitoring data over 21 years and across 102 catchments (which span over 130 000 km²). The modeling focused on six key water quality constituents: total suspended solids (TSS), total phosphorus (TP), filterable reactive phosphorus (FRP), total Kjeldahl nitrogen (TKN), nitrate–nitrite (NO_x) and electrical conductivity (EC). The model structure was informed by knowledge of the key factors driving water quality variation, which were identified in two preceding studies using the same dataset. Apart from FRP, which is hardly explained (19.9 %), the model explains 38.2 % (NO_x) to 88.6 % (EC) of the total spatiotemporal variability in water quality. Across constituents, the model generally captures over half of the observed spatial variability; the temporal variability remains largely unexplained across all catchments, although long-term trends are well captured. The model is best used to predict proportional changes in water quality on a Box–Cox-transformed scale, but it can have substantial bias if used to predict absolute values for high concentrations. This model can assist catchment management by (1) identifying hot spots and hot moments for waterway pollution; (2) predicting the effects of catchment changes on water quality, e.g., urbanization or forestation; and (3) identifying and explaining major water quality trends and changes. Further model improvements should focus on the following: (1) alternative statis-

tical model structures to improve fitting for truncated data (for constituents where a large amount of data fall below the detection limit); and (2) better representation of nonconservative constituents (e.g., FRP) by accounting for important biogeochemical processes.

1 Introduction

Deteriorating water quality in aquatic systems such as rivers and streams can have significant environmental, economic and social ramifications (e.g., Whitworth et al., 2012; Vörösmarty et al., 2010; Qin et al., 2010; Kingsford et al., 2011). Reducing these impacts requires effective management and mitigation of poor water quality; however, high variability in water quality across both space and time reduces our ability to accurately assess the status of water quality and develop effective management strategies. Thus, improved modeling frameworks to predict and interpret this variability would be useful for water quality management (Chang, 2008; Ai et al., 2015; Zhou et al., 2012).

Water quality conditions can vary across different events as well as at daily, seasonal and inter-annual scales at an individual location (Arheimer and Lidén, 2000; Kirchner et al., 2004; Larned et al., 2004; Pellerin et al., 2012; Saraceno et al., 2009). Water quality conditions also typically differ substantially across locations (Meybeck and Helmer, 1989; Chang, 2008; Varanka et al., 2015; Lintern et al., 2018a). These variabilities in stream water quality are driven by three key mechanisms: (1) the source, which defines the to-

tal amount of constituents available in a catchment; (2) the mobilization, which detaches constituents (both in the particulate and dissolved forms) from their sources via processes such as erosion and biogeochemical processing; and (3) the delivery of mobilized constituents from catchments to receiving waters via multiple hydrologic pathways including surface and subsurface flow (Granger et al., 2010).

Spatial variability in stream water quality is driven by human activities within catchments (e.g., land use and management, vegetation cover and so on; Lintern et al., 2018a; Carey and Migliaccio, 2009; Giri and Qiu, 2016; Heathwaite, 2010) as well as natural catchment characteristics such as climate, geology, soil type, topography and hydrology (Hrachowitz et al., 2016; Poulsen et al., 2006; Sueker et al., 2001; Onderka et al., 2012). At the same time, temporal shifts in water quality are also influenced by changes in pollutant sources such as land use and land management, including urbanization, agriculture and vegetation clearing (Ren et al., 2003; Smith et al., 2013; Ouyang et al., 2010). In addition, water quality can also vary in time with variations in the mobilization and delivery processes, which are largely driven by the hydroclimatic conditions in a catchment, such as streamflow (Ahearn et al., 2004; Mellander et al., 2015; Sharpley et al., 2002; Zhang and Ball, 2017), the timing and magnitude of rainfall events (Fraser et al., 1999; Miller et al., 2014), and temperature (Bailey and Ahmadi, 2014).

As mentioned above, we have a good understanding of the key controls of variations in water quality, albeit in an isolated, idealized context. However, we still lack a sound understanding of how relationships between specific landscape characteristics and water quality can shift with influences from other landscape characteristics, and how the drivers of temporal variability in water quality can interact and vary across large spatial scales (Musolff et al., 2015; Lintern et al., 2018a; Ali et al., 2017). Currently, the understanding of these factors has been primarily based on field studies at small scales with detailed information on specific temporal drivers ranging from hydrologic conditions to detailed management decisions such as fertilizer rates and application timing (Smith et al., 2013; Poudel et al., 2013; Adams et al., 2014). While operational weather observation networks, stream gauging networks and remote sensing can provide some of this information, developing a large-scale understanding of water quality patterns across catchments would ideally also involve an extensive suite of management information that substantially exceeds what is currently available.

Due to the limited understanding of large-scale water quality patterns, we currently lack the capacity to model spatiotemporal variabilities in water quality at large scales across multiple catchments. This hinders our ability to inform the development of effective policy and mitigation strategies over large regions. Specifically, conceptual or physically based water quality models are typically limited by the simplification of physical processes such as flow pathways (Hrachowitz et al., 2016). Furthermore, the practical implemen-

tation of these models can also be restricted by the intensive data requirements for calibration and validation, particularly for large regions with highly heterogeneous catchment conditions (Fu et al., 2018; Abbaspour et al., 2015). In contrast, when performed over large geographical regions, statistical water quality models are generally more capable of simulating water quality variability and require less detailed information and, thus, effort for implementation. However, existing statistical models often only focus on either the spatial variation of time-averaged water quality conditions (Tramblay et al., 2010; Ai et al., 2015) or the temporal variation at individual locations (Kisi and Parmar, 2016; Kurunç et al., 2005; Parmar and Bhardwaj, 2015), which often limits their value as practical management tools. Modeling the spatiotemporal variability simultaneously remains challenging over long time periods and large regions.

Accordingly, this research attempts to bridge the gap between fully distributed physically based water quality models and data-driven statistical approaches. We aim to develop a process-informed, data-driven model to predict spatiotemporal changes in stream water quality over a large region consisting of multiple catchments. Specifically, this model was established using long-term (21 years) stream water quality observations across 102 catchments in Australia, with an aggregate catchment area of more than 130 000 km². To obtain the necessary understanding of the process drivers required to develop this model, two preceding studies were conducted on the same dataset to identify the key drivers for the spatial and temporal variability of water quality, respectively (Lintern et al., 2018b; Guo et al., 2019). The aim of this study is to develop an integrated spatiotemporal model using the previously identified spatial and temporal predictors and to then assess the performance of this model. Spatiotemporal variability of water quality was modeled using a novel Bayesian hierarchical approach which can jointly consider both variability components, including accounting for varying temporal water quality dynamics between catchments. This modeling approach also has relatively low requirements for input data, which keeps the modeling detail commensurate with the level of data availability. During the model development, we also obtained an additional understanding of the patterns of spatial variations in the effects of each temporal predictor. The model can potentially provide useful information for large-scale catchment management, assessment and policy making, such as testing major changes in land use patterns, informing pollution hot spots, and identifying and attributing water quality trends and changes over time.

2 Method

We first discuss the process used to develop the integrated spatiotemporal model (Sect. 2.1). Section 2.1.1 and 2.1.2 introduce the statistical modeling framework and the data used for model development, respectively. The approaches used

to determine model structure were then introduced, which include the choice of key predictors (Sect. 2.1.3) and the calibration for model parameters (Sect. 2.1.4). Finally, the approaches utilized to evaluate model performance and robustness are described in Sect. 2.2.

2.1 Model development

2.1.1 Spatiotemporal modeling framework

A Bayesian hierarchical approach was used to model the spatiotemporal variability in stream water quality. The Bayesian approach enables the inherent natural stochasticity of water quality to be incorporated into the model (Clark, 2005). A key advantage of applying the hierarchical model structure to analyze spatiotemporal variability is that this structure enables the key controls of temporal variability in water quality to vary across locations (Webb and King, 2009; Borsuk et al., 2001). This variability has been found to be important in other study regions where the (temporal) solute export regime varies with catchment characteristics such as climate and land use (Musolff et al., 2015; Poor and McDonnell, 2007).

The structure of the Bayesian hierarchical model is presented in Eqs. (1)–(6). Equation (1) formulates the transformed constituent concentration (see Sect. 2.1.2 for justification) at time i and site j (C_{ij}) as a normal distribution with a mean μ_{ij} and standard deviation σ , representing inherent randomness:

$$C_{ij} \sim N(\mu_{ij}, \sigma). \quad (1)$$

To represent spatiotemporal variability, μ_{ij} is modeled as the sum of the site-level mean constituent concentration (\bar{C}_j) and the deviation from that mean at time i (Δ_{ij}):

$$\mu_{ij} = \bar{C}_j + \Delta_{ij}. \quad (2)$$

To describe spatial variability, the site-level mean concentration at site j (\bar{C}_j) is modeled as a linear function of a global intercept (intC) and the sum of m catchment characteristics $S_{1,j}$ to $S_{m,j}$ (e.g., land use and topography) weighted by their relative contributions to spatial variability (βS_1 to βS_m):

$$\bar{C}_j = \text{intC} + \beta S_1 \times S_{1,j} + \beta S_2 \times S_{2,j} + \dots + \beta S_m \times S_{m,j}. \quad (3)$$

The temporal variability, represented by the deviation from the mean (Δ_{ij}), is a linear combination of n temporal variables, $T_{1,ij}$ to $T_{n,ij}$ (e.g., climate condition, streamflow, vegetation cover) (Eq. 4), at time i and site j :

$$\Delta_{ij} = \beta T_{1,j} \times T_{1,ij} + \dots + \beta T_{n,j} \times T_{n,ij}. \quad (4)$$

The selection of key spatial and temporal predictors for the model was performed in our two preceding studies (Lintern et al., 2018b; Guo et al., 2019) and is briefly described

in Sect. 2.1.3. Equations (1)–(4) enable the model to separately represent the spatial and temporal variability in water quality; however, there is still a further step required to make the model fully spatiotemporal (i.e., able to predict over both time and location). Specifically, in Guo et al. (2019), clear spatial variation was observed in the relationships between water quality and its key temporal predictors (i.e., in the $\beta T_{N,j}$ in Eq. 4). To be able to model multiple catchments across a large spatial area simultaneously, we must account for differences in these temporal influences across sites. To do this, the effect of each temporal variable at site j ($\beta T_{N,j}$ with N in $1, 2, \dots, n$) is drawn from a distribution with a mean of $\mu\beta T_{N,j}$ (Eq. 5), which is then modeled with a linear combination of two additional catchment characteristics, $ST_{N1,j}$ and $ST_{N2,j}$ (Eq. 6). Details regarding the selection of these two additional predictors are presented in Sect. 2.1.3.

$$\beta T_{N,j} \sim N(\mu\beta T_{N,j}, \sigma\beta T), \text{ for } N \text{ in } 1, 2, \dots, n \quad (5)$$

$$\mu\beta T_{N,j} = \text{int}\beta T_N + \beta ST_{N1} \times ST_{N1,j} + \beta ST_{N2} \times ST_{N2,j}. \quad (6)$$

2.1.2 Data collection and processing

The Bayesian hierarchical model was developed using 21 years of monthly stream water quality observations from 102 catchments in the state of Victoria, Australia (aggregate catchment area greater than 130 000 km²). The collection and processing of the data are detailed in previous publications that worked with the same dataset (Lintern et al., 2018b and Guo et al., 2019). Briefly, stream water quality data were extracted from the Victorian Water Measurement Information System (Department of Environment, Land, Water and Planning Victoria, 2016), which contains monthly grab samples of water quality at approximately 400 sites across Victoria. From these, a total of 102 sites were used to develop the model (Fig. 1), with available water quality data at all sites that span over 1994–2014. These sites and the above-mentioned time period were chosen because they provided the longest consistent period of continuous records over the greatest number of monitoring sites. The catchments corresponding to these water quality monitoring sites were delineated using the Australian Hydrological Geospatial Fabric (Geofabric) tool (Bureau of Meteorology, 2012) and have areas ranging from 5 to 16 000 km². The water quality parameters of interest were total suspended solids (TSS), total phosphorus (TP), filterable reactive phosphorus (FRP), total Kjeldahl nitrogen (TKN), nitrate–nitrite (NO_x) and electrical conductivity (EC). These parameters represent sediments, nutrients and salts, which are some of the key concerns for water quality managers in Australia and around the world. These water quality samples were collected following standard Department of Environment, Land, Water and Planning protocols (Australian Water Technologies, 1999) and analyzed in National Association of Testing Authorities accredited laboratories. Note that FRP is defined as “Reactive Phosphorus for a filtered sample to a defined filter size (e.g.,

RP(<0.45 µm))” in the sampling protocol, which is equivalent to the more widely used terminology, SRP, i.e., soluble reactive phosphorus (Jarvie et al., 2002).

To compile a dataset for the potential spatial explanatory variables (i.e., predictors to explain spatial variability in water quality), a comprehensive literature review was conducted (Lintern et al., 2018a) that summarized the key catchment landscape characteristics that are widely known to influence water quality. Furthermore, as part of Lintern et al. (2018b), a total of 50 potential explanatory catchment characteristics were selected, which included catchment land use, land cover, topographic, climatic, geological, lithological and hydrological catchment characteristics. These variables were derived using datasets obtained from Geoscience Australia (2004, 2011), the Bureau of Meteorology (2012), the Bureau of Rural Sciences (2010), the Department of Environment, Land, Water and Planning Victoria (2016) and the Terrestrial Ecosystem Research Network (2016) (see Table S1 in the Supplement for detailed variable names and data sources). We used a static set of land use data from 2005 to 2006 to represent the entire study period, because a preliminary analysis between 1996 and 2011 suggested less than 1 % changes in the key land uses in these catchments (i.e., agricultural, grazing and conservation).

A total of 19 potential temporal explanatory variables were included. Data of discharge (originally in ML d⁻¹) and water temperature (°C) corresponding to the same timestamps as for the water quality observations were also extracted for each monitoring site over the study period (Department of Environment, Land, Water and Planning Victoria, 2016). Discharge was converted to runoff depth (mm d⁻¹) for each catchment, and the average streamflows over 1, 3, 7, 14 and 30 d preceding the water quality sampling dates were calculated. In addition, we extracted a gridded dataset from the Australian Water Availability Project (AWAP; Frost et al., 2016; Raupach et al., 2009, 2012) and the Australian Water Resources Assessment Landscape (AWRA-L) model (Frost et al., 2016). These datasets were used to calculate catchment-averaged values of daily average temperature (°C), daily rainfall (mm), antecedent rainfall (1, 3, 7, 14 and 30 d preceding the sampling), dry spell (>0.1 mm rainfall) length of the antecedent 14 d, daily actual evapotranspiration (ET, mm), and soil moisture for the root zone and the deep zone (the averaged volumetric content for above and below a depth of 1 m from the surface, respectively). In addition, catchment-averaged monthly normalized difference vegetation index (NDVI) data were extracted from the Advanced Very High Resolution Radiometer (AVHRR) product (Eidenshink, 1992) and the Moderate Resolution Imaging Spectroradiometer MOD13A3 product (NASA LP DAAC, 2017). A summary of these datasets of temporal variables and their corresponding sources are given in Table S2, and details are provided in Guo et al. (2019).

The raw input data were filtered and transformed to increase the data reliability, continuity and symmetry, making

them more suitable for use in the linear spatiotemporal model structure (Eqs. 3, 4, 6). For the filtering process, we first removed all water quality records with flags indicating quality issues. We also removed any values below the detection limit (DL), which was defined as the “minimum concentration detected for which there is 95 % confidence of accuracy and therefore is accurate enough to report” in the monitoring protocols for this dataset (Australian Water Technologies, 1999). This was because the uncertainty in values below the DL would be amplified after transformation, which would influence the subsequent model fitting. Furthermore, these undetectable low concentrations were of less interest for management purposes. Water quality records corresponding to days with zero flows were also excluded from further analyses.

The transformation process was performed for each of the spatial catchment characteristics, temporal explanatory variables and water quality constituents to improve the symmetry of individual distributions. The log-sinh transformation (Wang et al., 2012; Eq. 7) was used for all catchment characteristics due to its ability to resolve the presence of zero values in several of the catchment characteristics (e.g., percentage area of individual land uses). The “GA” package in R (Luca Scrucca, 2019) was used to identify the log-sinh transformation parameters (a and b) for each spatial explanatory variable that minimized the data skewness (i.e., symmetry is maximized) across all 102 catchments:

$$y_{\log-\sinh} = \frac{1}{b} \log(\sinh[a + by_{\text{raw}}]). \quad (7)$$

In addition, all observed constituent concentrations and temporal explanatory variables were Box–Cox transformed (Box and Cox, 1964):

$$y_{\text{Box-Cox}} = \begin{cases} \frac{y_{\text{Raw}}^\lambda - 1}{\lambda}, & \text{for } \lambda \neq 0 \\ \log y, & \text{for } \lambda = 0 \end{cases}. \quad (8)$$

For each variable, the optimal Box–Cox transformation parameter λ was identified using the “car” R package and a maximum-likelihood approach. We first identified the optimal Box–Cox parameter λ using the data at each site (i.e., 21-year time series). The averaged λ across all sites was then used to transform the data across all catchments together. This transformation approach ensured that all sites used a consistent transformation parameter. All of the transformation parameters utilized are summarized in Tables S3 and S4. The transformation process greatly improved the data symmetry and, thus, the suitability for use in a linear model (the quality of the transformations was assessed via visual inspection in Lintern et al. (2018b) and Guo et al. (2019) and is summarized in Fig. S2, S4 and S6 in the Supplement).

2.1.3 Selection of key model predictors

Key predictors for the model were selected in a process-informed and data-driven manner based on our two preced-

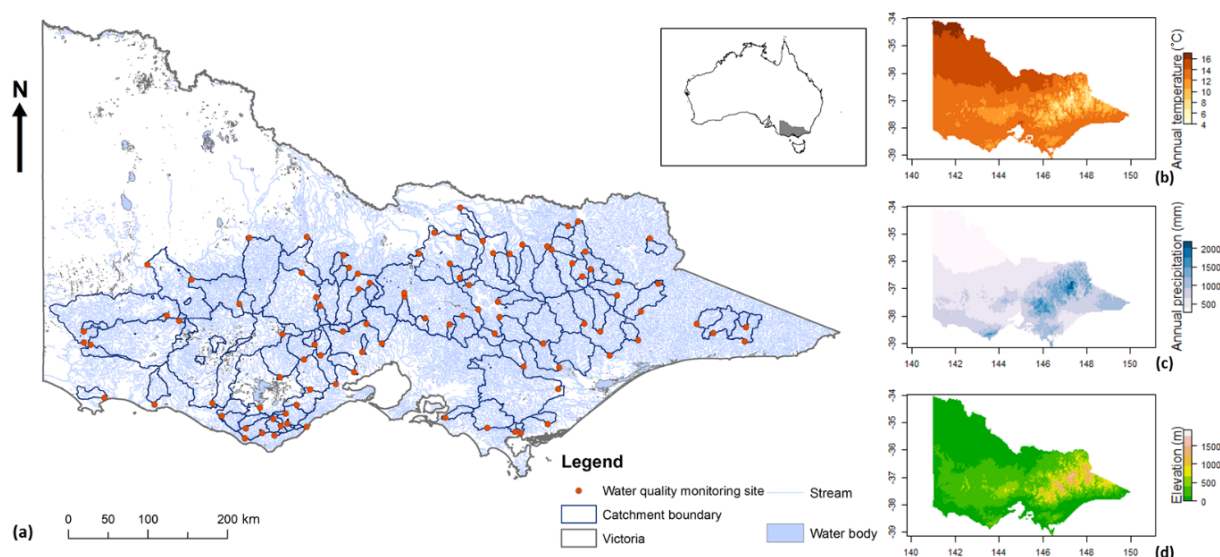


Figure 1. Map of (a) the 102 selected water quality monitoring sites and their catchment boundaries, with the inset showing the location of the state of Victoria within Australia. Maps of (b) the average annual temperature, (c) the annual precipitation and (d) the elevation across Victoria.

ing studies (Lintern et al., 2018b and Guo et al., 2019). Lintern et al. (2018b) identified the best spatial predictors (S_1 to S_m in Eq. 3) for the model, and the best temporal predictors across all sites (T_1 to T_n in Eq. 4) were identified in Guo et al. (2019). In both studies, the best predictors were selected using an exhaustive search approach (May et al., 2011; Saft et al., 2016) that considered all possible combinations of the potential predictors introduced earlier in this section. This selection approach required fitting an individual model to all possible candidate predictor sets, and then comparing all fitted models to select the single best set of predictors. Alternative models were evaluated based on the Akaike information criterion (AIC) (Akaike, 1974) and the Bayesian information criterion (BIC) (Schwarz, 1978) to ensure optimal balance between model performance and complexity.

The best predictors to explain the spatial and temporal variabilities in each constituent are listed in Table 1. Generally speaking, the key factors controlling the spatial variability in river water quality were land use and long-term climate conditions (Lintern et al., 2018b). Temporal variability was mainly explained by temporal changes in streamflow conditions, water temperature and soil moisture (Guo et al., 2019). The potential mechanisms via which these key drivers influence water quality are discussed in detail in the two above-mentioned studies.

While the previous studies, Lintern et al. (2018b) and Guo et al. (2019), identified the predictors for spatial and temporal variability, respectively, they did not provide guidance on the predictors for spatial variability in the relationships between drivers of temporal variability and temporal water quality response (i.e., βT in Eq. 4). As such, the final step of the predictor selection process to develop the combined spatiotem-

poral model was to identify the key catchment characteristics that affect spatial variability in the hydroclimatic parameters driving temporal changes in water quality (βT_1 to βT_n in Eq. (4) and the right column in Table 1). This is achieved by selecting two spatial characteristics that are most closely related to the coefficient for each temporal predictor (ST_{N1} and ST_{N2} , Eq. 6) across all sites, where only two spatial characteristics were used to avoid over-fitting. Selection of these two spatial characteristics were based on a Spearman correlation analysis between the fitted parameter values of each temporal predictor variable and the 50 potential spatial explanatory variables (as mentioned earlier in this section), following three steps:

1. from the 50 candidate spatial predictors, the one with the highest Spearman correlation with βT_N is selected as ST_{N1} , provided that the correlation is statistically significant ($p < 0.05$);
2. the subset of the remaining spatial predictors with a Spearman correlation of $ST_{N1} < 0.7$ is found; and
3. from this subset, the spatial predictor with the highest Spearman correlation with βT_N is selected as ST_{N2} , provided that the correlation is statistically significant ($p < 0.05$).

Steps 2 and 3 are intended to avoid cross-correlations between ST_{N1} and ST_{N2} . The selected spatial characteristics that influence the temporal relationships in our model are presented and interpreted in Sect. 3.1. Note that the entire selection process for ST_{N1} and ST_{N2} was performed with the fitted parameters for each predictor of the temporal variability, as obtained from Guo et al. (2019).

Table 1. Key factors affecting the spatial and temporal variability for each of six constituents, as identified in Lintern et al. (2018b) and Guo et al. (2019), respectively.

Constituent	Key factors that affect spatial variability	Key factors that affect temporal variability
TSS	Hottest month maximum temperature Percentage area covered by grass Percentage area covered by shrub Percentage cropping area Maximum elevation Dam storage Percentage clay area	Same-day streamflow 7 d antecedent streamflow Water temperature Soil moisture root zone Soil moisture deep zone
TP	Erosivity Percentage area covered by grass Percentage area covered by shrub Percentage area made up of roads Percentage cropping area Average soil TP content	Same-day streamflow 30 d antecedent streamflow Water temperature Soil moisture root zone Soil moisture deep zone
FRP	Percentage area covered by shrub Percentage cropping area Catchment area Average soil TP content Mean channel slope	Same-day streamflow Water temperature Soil moisture deep
TKN	Percentage clay area Warmest quarter mean temperature Coldest quarter rainfall Percentage cropping area Percentage pasture area Average soil TP content	Same-day streamflow 30 d antecedent streamflow NDVI Water temperature Soil moisture root zone Soil moisture deep zone
NO _x	Annual radiation Warm quarter rainfall Hottest month maximum temperature Average soil TP content Mean channel slope	Same-day streamflow 30 d antecedent streamflow NDVI Water temperature Soil moisture root zone Soil moisture deep zone
EC	Annual radiation Annual rainfall Wettest quarter rain Hottest month maximum temperature Percentage agriculture area Percentage cropping area Percentage area covered by shrub Average soil TN content	Same-day streamflow 14 d antecedent streamflow Water temperature Soil moisture root zone Soil moisture deep zone

2.1.4 Model calibration

After identifying the spatial and temporal predictors for each constituent, as well as the spatial characteristics that affect the strengths of each temporal predictor, the Bayesian hierarchical spatiotemporal model was fitted for each constituent across all monitoring sites simultaneously. To achieve this, we used the “rstan” R package (Stan Development Team, 2018), which enabled both the sampling of parameter values from posterior distributions with Markov chain Monte Carlo

(MCMC) and model evaluation. The constituent standard deviation (σ) was assumed to be drawn from a minimally informative half-normal prior distribution of $N(0, 10)$ truncated to only positive values (Gelman, 2006; Stan Development Team, 2018). The regression coefficient of each spatial predictor ($\beta S_1, \beta S_2, \dots, \beta S_m$ in Eq. 3) was independently drawn from hyper-parameter distributions of $N(\mu\beta S_M, \sigma\beta S_M)$. The site-level regression coefficients of the temporal predictors ($\beta T_{1,j}, \beta T_{2,j}, \dots, \beta T_{n,j}$ in Eq. (4), respectively) were sampled from the corresponding hyper-parameter distribu-

tion of $N(\mu\beta T_N, \sigma\beta T_N)$. The hyper-parameters were further assumed to be drawn from minimally informative prior distributions, following recommendations in Gelman (2006) and Stan Development Team (2019): for all of the hyper-parameter means, a normal prior distribution of $N(0, 5)$ was used; for all of the hyper-parameter standard deviations, a half-normal prior distribution of $N(0, 10)$ was used, which was truncated to only positive values. In each model run there were four independent Markov chains. A total of 20 000 iterations were used for each chain. Convergence of the chains was ensured by checking the “Rhat” value (Sturtz et al., 2005), which is a summary statistic on the convergence of the Bayesian models from the four Markov chains used in model calibration (Stan Development Team, 2018). Specifically, a Rhat value much greater than 1 indicates that the independent Markov chains have not been mixed well, and a value of below 1.1 is recommended (Stan Development Team, 2018).

2.2 Model performance evaluation and sensitivity analyses

Performance evaluation of the model was undertaken on several aspects of the model results (Sect. 3.2). As the model was calibrated on a Box–Cox transformation scale (see justification in Sect. 2.1.2), the Box–Cox transformation scale was used for model evaluation to enable a clear investigation of the influences of a wide range of factors that can influence model performance. Detailed performance evaluations included the following:

1. *Ability to capture total spatiotemporal variability.* The simulations from the fitted model and the corresponding observed concentrations were compared at 102 sites altogether in order to understand how the overall spatiotemporal variabilities were captured. For each constituent, this evaluation was performed with (a) the above-DL data, to focus solely on data used for calibration (as detailed in Sect. 2.1.2); and (b) the full dataset including the below-DL data (set to half of the DL of the specific constituent), to understand how well the model represents the full distribution of constituent concentrations. A good model performance when including the below-DL data would suggest that the calibrated model is also transferable to below-DL data. All performance assessments were based on both visual inspection of model fitting as well as the Nash–Sutcliffe efficiency (NSE), which quantified the proportion of variability that was explained by the model (Nash and Sutcliffe, 1970).
2. *Proportions of spatial and temporal variability explained.* This involved a decomposition of the total observed variability, using Eq. (2), into proportions contributed by spatial variability (variations in all site-mean concentrations from the grand average of site-mean concentrations) and temporal variability (variations in all

concentrations from the corresponding site-mean concentrations). The corresponding modeled values were then used to calculate the NSE for each variability component of each constituent.

3. *Ability to capture the variation in ambient conditions across space and the temporal variation (including trends) across multiple catchments.* These model capacities were evaluated by (a) comparing all simulated and observed site-averaged long-term mean concentrations and (b) comparing the simulated and observed time series and long-term trends at representative sites. Further to (a), performance was also evaluated on a real measurement scale by back-transforming all modeled sample concentrations, calculating the back-transformed site-level means and then comparing the latter to the corresponding observations. A further analysis to (b) was also performed by comparing the estimated Sen’s slope (Akritas et al., 1995) for the observations and simulations at all sites and then computing the percentage of sites where the observed trends (as indicated by the Sen’s slope) have been correctly represented by the model.

Additional evaluations of model sensitivity were conducted with calibration and validation on subsets of the full data (Sect. 3.3) in order to understand model transferability and stability:

1. *Model sensitivity to the monitoring sites used for calibration.* We randomly selected 80 % of the sites for calibration and used the remaining 20 % for validation; we repeated this validation process 50 times. We compared the calibration and validation performance of all of these “partial models” with one another, as well as with the performance of the full model, in order to obtain a comprehensive evaluation of the sensitivity of model performance to calibration sites.
2. *Model sensitivity to calibration data period.* As the study region was greatly influenced by a prolonged drought from 1997 to 2009 – known as the “Millennium Drought” (van Dijk et al., 2013) – we also investigated model robustness for before, during and after this drought period. Specifically, we calibrated the model to each pre-, during- and post-drought period (1994–1996, 1997–2009 and 2010–2014, respectively) and carried out model validation on the remaining data. For example, when calibrating to the pre-drought period (1997–2009), validation was performed on the merged during and post-drought period (1994–1996 plus 2010–2014). The corresponding calibration and validation performances were compared with each other as well as against that of the full model in order to identify the potential impacts of the drought on model robustness.

3 Results

3.1 Spatial variation in the impact of temporal factors

The key controls of the spatial and temporal variations in water quality were identified in our two preceding studies (Lintern et al., 2018b; Guo et al., 2019) and are briefly summarized in Sect. 2.1.3; thus, they are not discussed here. As also detailed in Sect. 2.1.3, to achieve full spatiotemporal predictive capacity, the model developed in this study considers the spatial variation in the strength of each temporal predictor by using two additional catchment spatial characteristics ($ST_{N1,j}$ and $ST_{N2,j}$ in Eq. 6) based on the Spearman correlations. Here we focus on the most important temporal predictor for each constituent – streamflow – and Table 2 shows the two spatial characteristics that are identified as being most closely related to the spatial variation of the impact of streamflow on water quality. The full list of the selected key catchment characteristics for all of the temporal predictors of each constituent is summarized in Table S5 and visualized in Fig. S4.

TSS, TP and TKN show consistent patterns with respect to spatial variation in the effects of streamflow on water quality, which are strongly driven by the differences in average rainfall conditions across catchments. Specifically, streamflow generally has a larger effect on water quality in catchments with higher average annual rainfall. As the streamflow effects are positive for the majority of catchments (as shown in Fig. S5), these correlations indicate that a greater increase in transformed concentrations of TSS, TP and TKN will occur for the same increase in transformed streamflow at a catchment with a higher annual average rainfall. Given that the Box–Cox lambda values (Table S4) are close to zero, the transformation is log-like; therefore, changes in the transformed flow and concentration approximately correspond to proportional changes in the real values of flow and concentration. In contrast, for FRP, NO_x and EC, the spatial patterns of streamflow effects are specific to each constituent. This difference in the model results between TSS, TP and TKN and the other constituents might be related to the distinct transport pathways of particulate and dissolved constituents. The former is predominantly related to surface flow and, thus, relies heavily on rainfall contribution. Dissolved constituents are likely transported along the subsurface pathway. Apart from streamflow, the spatial patterns in other key temporal drivers of water quality (e.g., antecedent streamflow, soil moisture and so on) are less consistent across different constituents (Fig. S4).

3.2 Model performance evaluation

The spatiotemporal water quality models show varying performance between the constituents. When assessed with only the above-DL data (Fig. 2), the best performing models are those for EC and TKN, which capture 90.7 % and 65.8 %

of the total observed spatiotemporal variability, respectively. The modeling performance is lowest for FRP, NO_x and TSS, with NSE values of -1.92 , 0.216 and 0.225 , respectively. When evaluated against the entire dataset (i.e., including both below- and above DL data), the models explain 19.9 % (FRP) to 88.6 % (EC) of the spatiotemporal variability (Table 3). Model performances for FRP, NO_x and TSS improve notably compared with the previous evaluation of above-DL data; however, they remain the three constituents that are most difficult to predict. We further discuss the possible factors influencing their model performance in Sect. 4.1.

The model performance to predict spatial and temporal variability is summarized in Fig. 3, which compares the observed and explained variability for each of the spatial and temporal components (detailed in Sect. 2.1.4). Regarding the observed variability (lighter colors), EC is strongly dominated by spatial variability (91.8 %), highlighting that within-site variation in water quality is minimal compared with between-site variation. To a lesser extent, spatial variability also contributes to major proportions of total variability for TP and TKN (60.8 % and 66.6 %, respectively). TSS, FRP and NO_x are more influenced by temporal variability (57.4 %, 56.6 % and 60.5 %, respectively).

The explained variability (darker colors) shows that temporal variability is much more difficult to model than spatial variability across all catchments. It also appears that a substantial part of the model's overall performance is driven by its ability to capture spatial variability in ambient water quality conditions. For example, the models for TSS, FRP and NO_x show poorer overall performance (Fig. 2, with NSE values of 0.225 , -1.92 and 0.216 , respectively), because the total variability for each of these constituents is dominated by temporal variability (57.4 %, 56.6 % and 60.5 %, respectively), which remains largely unexplained by the model (Fig. 3). In contrast, the EC model shows a very good fit with 90.7 % of the total variability explained – 91.8 % of the total observed variability is due to spatial variability, of which 94.7 % is explained by the model. Therefore, although the EC model can only explain a small portion of temporal variability (20 % of the 8.2 % total variability), the overall model performance remains high.

As highlighted in Fig. 3, the model is capable of capturing spatial variability in water quality. This is further evaluated in Fig. 4 by comparing the simulated and observed site-level mean concentrations. The highest model performance is for EC and the lowest performance is for FRP (explaining 94.7 % and 44.2 % of the spatial variability, respectively). At the back-transformed scale, the model shows greater biases for sites with higher concentrations (approximately the highest 10 % sites for each constituent; see Fig. 5). This is not surprising, as the model was fitted to a Box–Cox-transformed space that reduces the focus on high values and increases the focus on low values. This compromised the model's ability to represent sites with unusually high concentrations. The im-

Table 2. The two key catchment landscape characteristics that were selected as predictors for the spatially varying streamflow effects in our model (see Sect. 2.3 for details regarding the selection method). The corresponding Spearman correlation (ρ , at $p < 0.05$) between the effect of streamflow and each catchment characteristic is presented.

Constituent	Key factors that affect spatial variability in temporal effects	Spearman's ρ ($p < 0.05$)
TSS	Annual rainfall	0.722
	Hottest month maximum temperature	−0.575
TP	Annual rainfall	0.695
	Percentage area used for cropping	−0.556
FRP	Percentage agriculture area	0.392
	Percentage area underlain by mixed igneous bedrock	0.314
TKN	Annual rainfall	0.713
	Hottest month maximum temperature	−0.618
NO _x	Total storage capacity of dams in catchment	−0.493
	Mean soil TN content	0.458
EC	Percentage area covered by grassland	−0.347
	Percentage area covered by woodland	−0.317

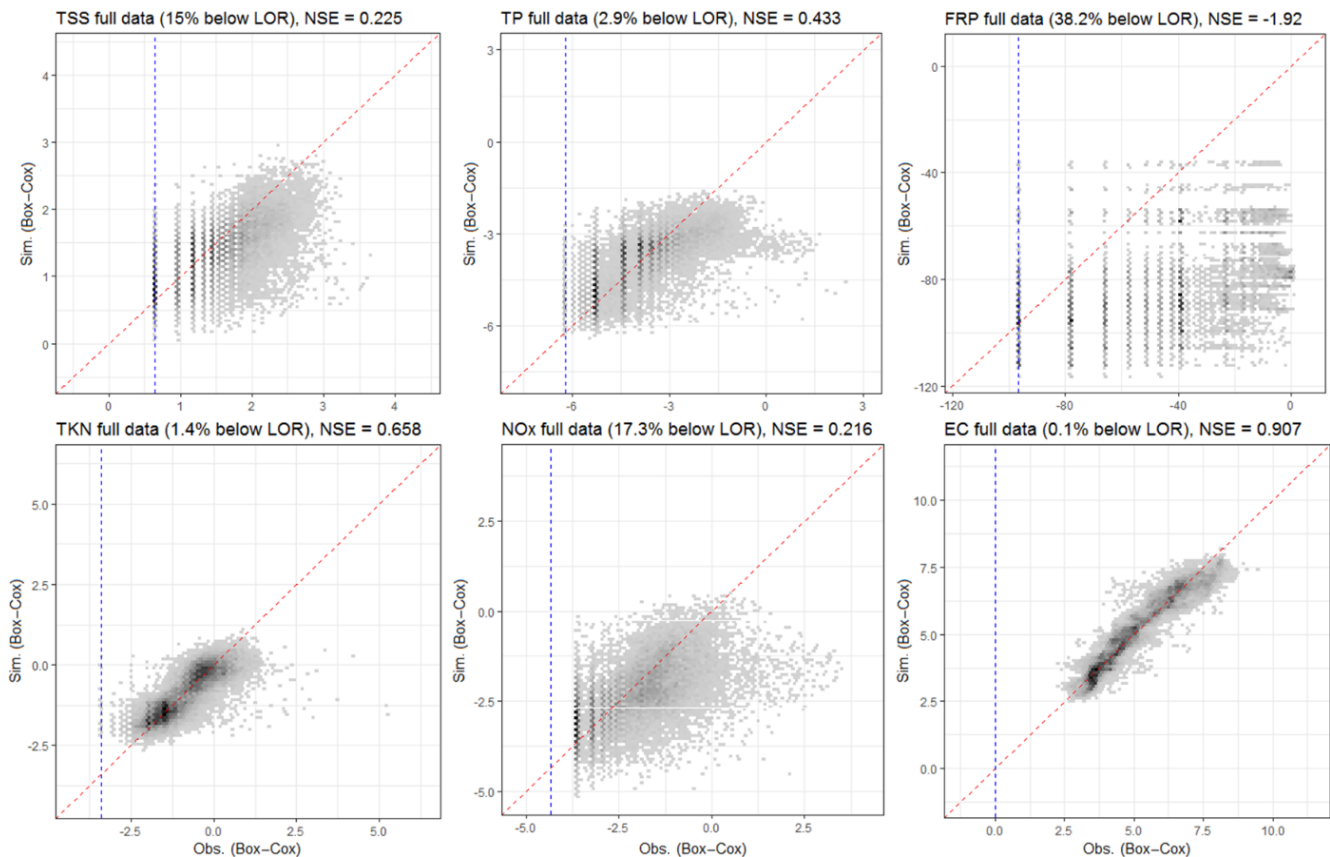


Figure 2. Performance of the spatiotemporal models for each of the six constituents, represented by the simulated median concentrations and corresponding observations of above-DL records across all 102 calibration sites, in Box–Cox-transformed space. Darker regions represent a denser distribution of simulation and observation points, dashed red lines show the 1 : 1 lines and dashed blue lines show the DL levels. For each constituent, the percentage of data below the DL and the model performance (NSE) are also specified.

Table 3. Comparison of model performance for all records and the above-DL records only for each constituent.

Constituent	Above-DL records only	All records
TSS	0.225	0.397
TP	0.433	0.445
FRP	−1.920	0.199
TKN	0.658	0.630
NO _x	0.216	0.382
EC	0.907	0.886

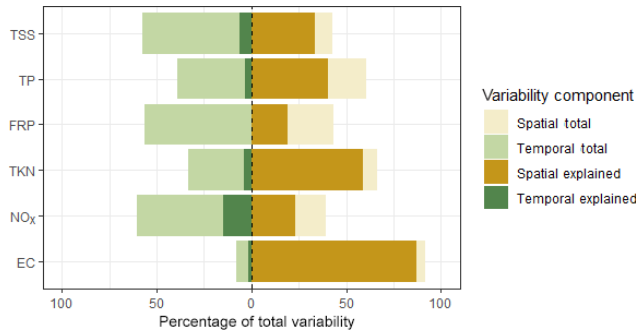


Figure 3. Observed spatial and temporal variabilities as proportions of the total variability (the total width of each bar is 100 %). The dashed line differentiates temporal variability (left) from spatial variability (right), and the darker colors highlight the proportions of the spatial and temporal variabilities that are explained by the model. All values were estimated in Box–Cox-transformed space.

plications of the model having a higher predictive capacity in the transformed scale is further discussed in Sect. 4.1.

As also noted in Fig. 3, the ability of the spatiotemporal model to explain temporal variability remains relatively limited. This is further explored in Fig. 6, where the observed and simulated time series are presented for the monitoring site for each constituent at which the model performance (NSE) was the highest. These results show that even for catchments where the model has the highest ability to capture temporal variability, the model consistently underestimated the temporal variability for all constituents.

Figure 6 also illustrates that, although the model shows substantial underestimation of temporal variability within site, long-term temporal trends in the time series are well captured at the best sites (except for FRP). Table 4 summarizes the ability of the model to capture observed trends across all 102 catchments for each constituent. In general, at most sites, the model is able to capture observed trends for NO_x and EC for both positive and negative trends. For TP and TKN, positive trends are well captured, whereas for TSS the negative trends are better captured.

Table 4. Model ability to capture observed water quality trends across all monitoring sites for each constituent. The percentages of sites where observed positive and negative trends are captured by the model are presented separately. Values in parentheses indicate the number of sites where corresponding positive or negative trends are observed. For detailed estimation of these percentages please refer to Sect. 2.2.

Constituent	Sites with positive trends captured (%)	Sites with negative trends captured (%)
TSS	33.3 (12)	85.0 (20)
TP	82.1 (28)	16.7 (12)
FRP	47.1 (17)	55.6 (9)
TKN	81.1 (37)	40.0 (10)
NO _x	68.6 (35)	66.7 (27)
EC	82.6 (23)	77.3 (22)

3.3 Model sensitivity analyses

We first compare the performance of each spatiotemporal model fitted with the full dataset with those obtained from the 50 corresponding “partial” models that were calibrated using only 80 % of the monitoring sites. Note that in this comparison, the FRP model was not assessed due to its poor performance (Sect. 3.2). The calibration and validation results for the 50 partial models are summarized in Table 5 along with the performance of the full model calibrated using all 102 sites (see Figs. S6 and S7 for a detailed comparison of the model residuals of the partial calibration/validation). Across constituents, the calibration performance of the full model was comparable to the 50 partial models. Note the slightly higher calibration performance for the partial models of NO_x compared with the full model. This seems to be related to the generally lower percentages of below-DL data in the 50 randomly chosen partial calibration datasets (14.1 %–17.9 %) compared with the full dataset (17.3 %) – we further discuss the impacts of below-DL data on model performance in Sect. 4.1. In addition, model performance is highly consistent between the corresponding calibration and validation, with most differences in NSEs being less than 0.1. This suggests that the spatiotemporal model performance is highly robust and unaffected by the choice of calibration sites.

The performance of the full model for each constituent is also compared with that of the three models calibrated using the pre-, during and post-drought periods. In general, we observe consistent performance for each constituent across calibrations using the three periods of contrasting hydrological conditions (Table 6, see Figs. S8 to S13 for detailed model fittings). One notable common pattern is that the performance for calibration and validation is more consistent during the drought period than during either the pre- and post-drought periods. However, this is most likely explained by the relative sizes of the calibration datasets, which are 3, 13 and 5 years for the pre-, during and post-drought periods, respectively.

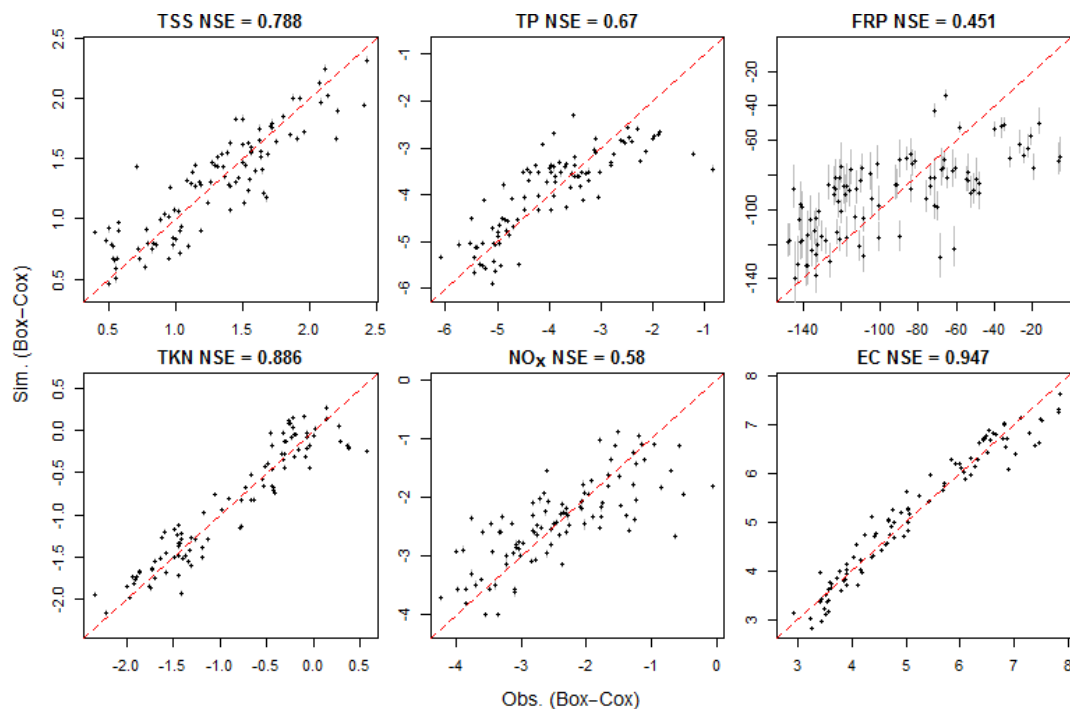


Figure 4. Model fit for the site-level mean concentration at the 102 calibration sites for six constituents, with the 95 % upper and lower bounds of the posterior simulations shown using vertical gray lines. All simulations and observations are presented in Box–Cox-transformed space. The NSE for each constituent is also shown, and red dashed lines show the 1 : 1 lines.

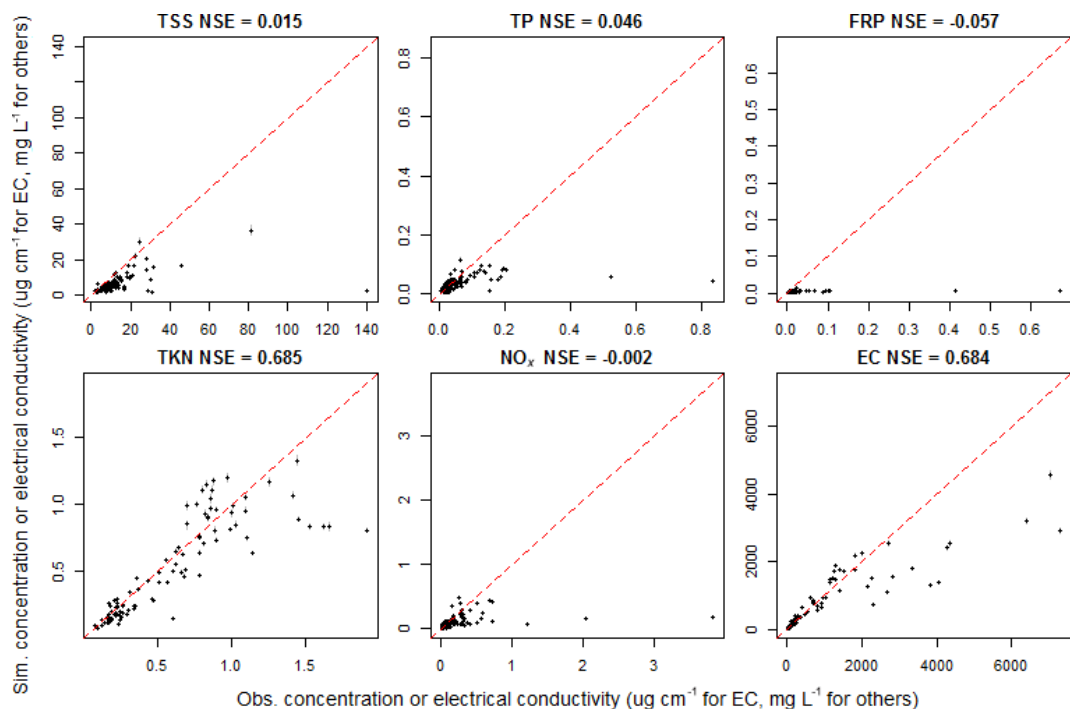


Figure 5. The back-transformation of the model simulations to the measurement scale emphasizes a lack of fit for the highest concentrations, as illustrated by the simulated against observed site-level mean concentrations of each constituent on a back-transformed scale. The 95 % upper and lower bounds of all posterior simulations are shown using vertical gray lines. The NSE for each constituent is also shown, and red dashed lines show the 1 : 1 lines.

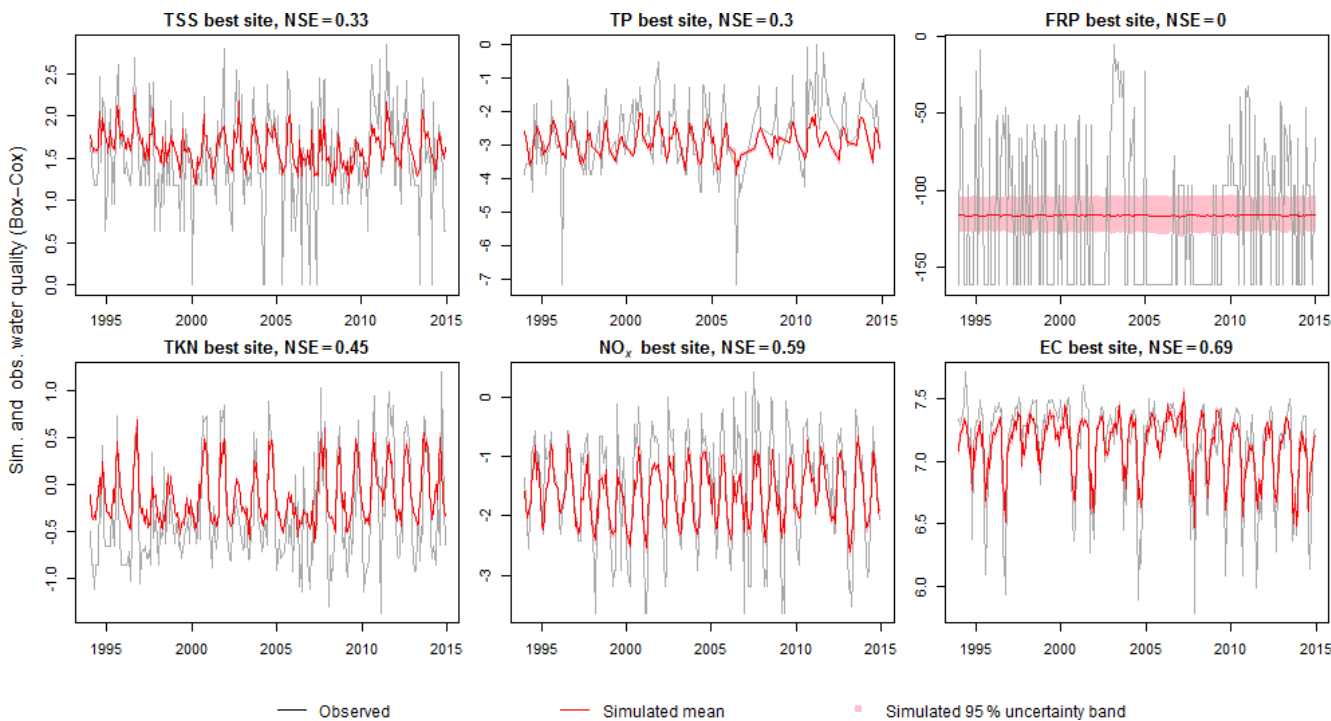


Figure 6. Model fit of the within-site (temporal) water quality variability, illustrated using the observed and simulated time series for the best-performing site for each constituent. All values are presented in Box–Cox-transformed space. The NSE for each constituent is also shown. The red line indicates the corresponding mean of all posterior simulations, and the pink shading shows the corresponding 95 % upper and lower bounds (only visible for FRP).

Table 5. Comparison of the model performance (NSE values) of the full model (column 2) and the 50 partial models (columns 3–5) that were each calibrated using 80 % of the monitoring sites, which were randomly selected. Columns 3 to 5 summarize the mean, minimum and maximum NSE values across the 50 runs of cross-validation (CV), where the top row shows the calibration performance and the bottom row shows the validation performance (i.e., at the 20 % sites that were not used for calibration) for each constituent.

Constituent	Full model	50 CV mean	50 CV min	50 CV max
TSS	0.397	0.413	0.376	0.439
		0.382	0.292	0.513
TP	0.445	0.461	0.427	0.501
		0.411	0.151	0.575
FRP	0.199	0.168	0.067	0.232
		0.129	−0.078	0.272
TKN	0.630	0.654	0.622	0.670
		0.622	0.468	0.691
NO _x	0.382	0.453	0.414	0.489
		0.397	0.258	0.563
EC	0.886	0.893	0.882	0.903
		0.875	0.809	0.924

Table 6. Comparison of model performance (NSE values) of the full model and the three models that were calibrated using the pre-drought (1994–1996), during drought (1997–2009) and the post-drought (2010–2014) periods. For each of the models, the calibration performance is shown in the top row and the validation performance (i.e., over the periods that were not used for calibration) is shown in the bottom row.

Constituent	Full model	Pre-drought calibration	During drought calibration	Post-drought calibration
TSS	0.397	0.495 0.208	0.399 0.402	0.499 0.390
TP	0.445	0.477 0.421	0.438 0.474	0.525 0.411
FRP	0.199	−1.336 −1.406	0.187 0.197	0.204 0.024
TKN	0.630	0.649 0.566	0.650 0.648	0.711 0.610
NO _x	0.382	0.443 0.394	0.426 0.471	0.509 0.393
EC	0.886	0.854 0.887	0.901 0.873	0.901 0.884

Of all of the constituents (excluding FRP), TSS shows greater differences in model performance across periods – especially when comparing the pre-drought calibration with its validation for the site-level mean concentrations (Fig. 7). Notably, when calibrated using the pre-drought period and validated with both the during- and post-drought periods, the validated model overestimates most of the data (Fig. 7b); when calibrated using the during-drought period, the validated model slightly underestimates the pre- and post-drought period TSS (Fig. 7d).

The potential impacts of drought on TSS dynamics are further illustrated by the performance of the spatiotemporal model (calibrated using the full dataset with all sites and all data from 1994 to 2014) over the pre-, during and post-drought periods (Fig. 8). Both the during- and post-drought periods have consistently good performance, whereas the model underestimates most sites for the pre-drought period. This is consistent with Fig. 7 in suggesting a systematic decrease in the TSS concentration since the beginning of the drought. The better performance of the full model during and the after drought (Fig. 8) may be a result of the calibration period of the full spatiotemporal model – between 1994 and 2014 – which was dominated by the during- and post-drought periods.

In summary, Figs. 7, 8 and S13–S17 suggest that while model performance for most constituents was not affected by the hydrological periods used for calibration and validation, the calibration period did have a notable impact on TSS. Some possible causes for this are discussed in Sect. 4.3.

4 Discussion

4.1 Implications for statistical water quality modeling

In this study, we developed the first process-informed statistical model that is capable of explaining a reasonable proportion of water quality variability for a large spatial area of over 130 000 km². Although the calibration data have a relatively low sampling frequency (i.e., monthly), our model generally performs satisfactorily with respect to explaining the total variability in water quality. This demonstrates the effectiveness of the Bayesian hierarchical modeling framework in predicting spatiotemporal variability in water quality across large scales. The Bayesian hierarchical model is more advantageous than other simpler statistical water quality models due to its more comprehensive and process-informed approach and capacity to represent varying temporal relationships across large-scale regions. Further, the model also has lower demand for input data compared with the requirements of fully distributed, processes-based models. From a practical perspective, this model has the potential to contribute to a number of management activities including catchment planning, management and policy-making activities. Specifically, the Bayesian hierarchical model can offer to the following:

1. The spatial predictive capacity can be used to identify pollution hot spots and the catchment conditions that are likely causes of high concentrations. This can be used to help identify target catchment(s) to prioritize future water quality monitoring and management (Figs. 4 and 5).
2. Further to (1), as water quality has been linked with catchment characteristics in this model, it can also be

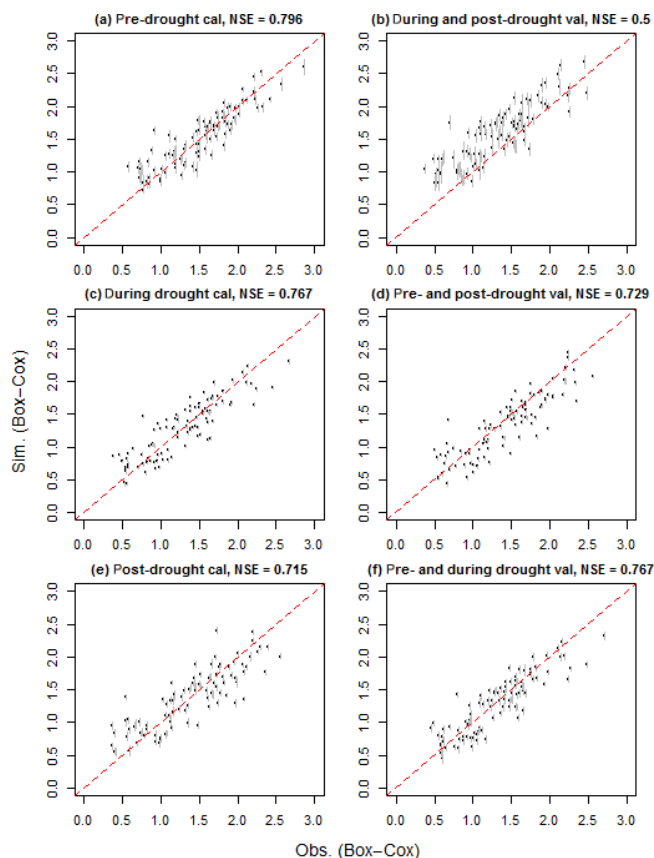


Figure 7. Comparison of the TSS model performance: the simulated against observed site-level mean concentrations in Box–Cox-transformed space. The left column shows the calibration performance for the model calibrated using the pre-drought (1994–1996), drought (1997–2009) and post-drought (2010–2014) periods, respectively; the right column shows the corresponding validation performance for each period. The 95 % upper and lower bounds of the simulations are shown using vertical gray lines, and red dashed lines show the 1 : 1 lines.

used to assess potential impacts of alternative options for land use and land cover change, as well as potential effects of climate change, on ambient water quality conditions.

3. The model's temporal predictive capacity can identify changes in water quality due to changes in hydroclimatic conditions and vegetation cover, thus enabling the attribution of detected trends. Conversely, any “unexpected” trends can be identified to prompt further investigation to identify causes (Fig. 6, Table 4). The model could also be used to assess the impacts of long-term catchment changes on water quality (Figs. 7, 8).

Despite the opportunities highlighted above, the model's performance also suggests some current limitations of the modeling framework in the following situations:

1. *High within-site temporal variability.* In Sect. 3.2 we identified a general lack of predictive power for temporal variability. The potential impacts of high temporal variability on model performance are particularly evident for TSS, NO_x and FRP in Fig. 3. As our model has already included hydroclimatic conditions and vegetation cover to explain temporal variability, the unexplained temporal variability is likely due to other uncaptured temporal drivers. These could be factors such as changes in land use and land management, biogeochemical processes or the transit time of water through catchments.

2. *The presence of high proportions of below-DL data.* The full datasets for the three poorly modeled constituents (FRP, TSS and NO_x) all have higher proportions of data below the detection limit (38.2 % 17.3 % and 15 % of all data, respectively) compared with other constituents. As illustrated in Fig. 2, for each of these constituents, removal of below-DL data before model calibration created a clear truncation on the left-hand side of the distribution. This substantially increases the degrees of skewness and discontinuity of the data, essentially violating the assumption of normally distributed residuals and, thus, limiting model performance. The model capacity to handle truncated data might be improved by model fitting approaches that are explicitly designed for this issue. For example, Wang and Robertson (2011) and Zhao et al. (2016) illustrated an approach to resolving the discontinuity of the likelihood estimation in model fitting to data with the presence of a lower bound such as zero rainfall values.

3. *The nonconservative nature of constituents.* The results indicate that the reactivity of the constituent is broadly associated with performance, which suggests that biogeochemical processes (e.g., phosphorus cycling and nitrification/de-nitrification) can make water quality dynamics more difficult for the model to capture. To better capture changes in reactive constituents, the model may require greater consideration of and more extensive spatial and temporal data to represent biogeochemical processes. Examples include improvements in the process representation for nitrogen cycling and the desorption and adsorption of phosphorus (Granger et al., 2010; Smyth et al., 2013; Tian and Zhou, 2007).

As previously noted, our model was developed on a Box–Cox-transformed scale to ensure the validity of the statistical assumptions (see details on data transformation in Sect. 2.1.2), which shows limited performance for high constituent concentrations when simulations are back-transformed to the measurement scale (Figs. 4, 5). However, our model approximately represents proportional changes in

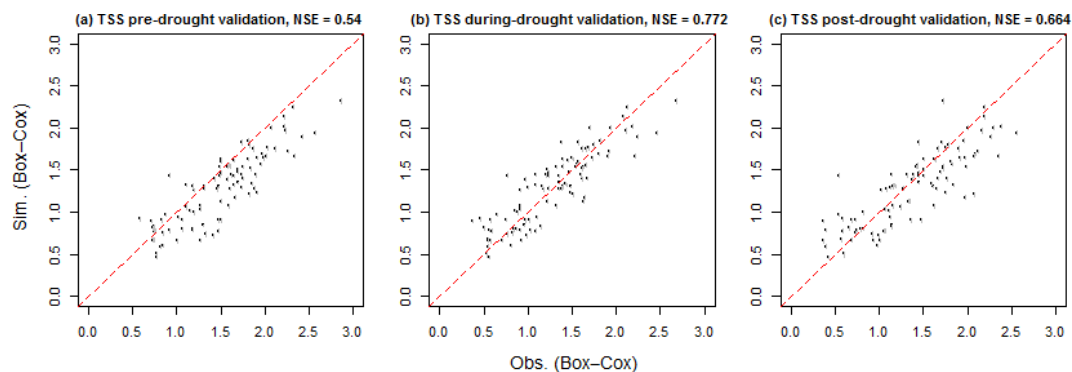


Figure 8. Comparison of the performance of the full spatiotemporal TSS model calibrated using all data across (a) the pre-drought (1994–1996), (b) drought (1997–2009) and (c) post-drought (2010–2014) periods, as represented by the simulated versus observed site-level mean concentrations in Box–Cox-transformed space. The 95 % upper and lower bounds of the simulations are shown using vertical gray lines, and red dashed lines show the 1 : 1 lines.

water quality¹, which, in turn, can help managers to understand proportional changes to inform practical catchment management.

For future implementations, the established model structure and parameterization would be best suited to within the study region. Before performing new simulations (e.g., for new monitoring sites or for current study sites over a different time period), the statistical properties of the new input datasets should be checked to ensure that they are similar to the calibration datasets. To model new catchments outside of the study region, a recalibration of the model is required. This would involve the extensive selection of key predictors and model calibration, which would have to be undertaken in much the same fashion as performed in this work and the two preceding studies (Lintern et al., 2018b; Guo et al., 2019). A sufficiently long record length (e.g., 20 years) is ideal for such modeling, as it ensures a reasonable understanding of the temporal variability to be obtained.

4.2 Implications for water quality monitoring programs

The current spatiotemporal model extracts water quality temporal variability from monthly data. Utilizing data with higher temporal resolution may further strengthen the model capacity to explain temporal variability, especially by capturing more information on water quality dynamics during flow events. This may be possible in the future; however, current high-frequency water quality sensors (Bende-Michl and Hairsine, 2010; Outram et al., 2014; Lannergård et al., 2019; Pellerin et al., 2016) still have very high resource requirements that limit widespread deployment in operational networks.

¹ All Box-Cox transformation parameters for water quality constituents are approximately zero (Table S4), which means that the transformations are similar to a log transformation.

Furthermore, changes in land use and management over time are currently not considered here as predictors of temporal variability in water quality, which include but are not limit to land clearing, urbanization, tillage, fertilizer application and irrigation. This is due to a complete lack or inconsistency in available data. However, changes in land use/land management practices can occur over short time periods, which can lead to increases in pollutant sources and changes in runoff generation processes (e.g., Tang et al., 2005; DeFries and Eshleman, 2004; Smith et al., 2013). Therefore, our modeling framework could potentially be improved by the addition of monitoring data on the temporal patterns of land use/land management, in order to better capture the impacts of these components on water quality.

4.3 Potential impacts of long-term drought on water quality dynamics

The results of model calibration and validation using different time periods suggest a systematic decrease in TSS concentrations during and after the prolonged drought compared with the pre-drought period under the same spatial and temporal conditions. Such a shift is not observed for any of the other five constituents analyzed (nutrients and salts; Sect. 3.3).

A further analysis of the calibrated model parameters for pre-, during- and post-drought periods suggests that the effects of key spatial predictors do not vary much across periods (Fig. S14). In contrast, the effects of key temporal predictors highlight a clear shift in the role of antecedent flow (7 d antecedent flow) across different time periods (Fig. 9). Specifically, the antecedent flow effects are mostly positive across catchments before the drought and shift to mostly negative during the drought. After the drought, the antecedent flow effects have mixed directions in different catchments. Considering the limited performance of the TSS model (i.e., the substantial underestimation of temporal variability in

Sect. 3.1), these changing relationships suggested in the calibrated parameters might be unreliable. However, this should not affect the reliability of the observed change in TSS since the drought (Sect. 3.3), which was based on the systematic differences in model fitting between different periods, revealing a broad-scale patterns across the state with respect to the drought influences.

In the literature, impacts of the Millennium Drought on the hydrology and runoff regimes of southeastern Australia are well understood (van Dijk et al., 2013; Leblanc et al., 2012; Saft et al., 2015). However, less is known about how this major and prolonged drought event has impacted water quality (Bond et al., 2008). Previous studies on other drought events around the world mainly focused on changes in water quality as responses to reduced streamflow during drought. For example, a reduction in sediment levels during drought has been reported and, in turn, attributed to lower erosion from the contributing catchment and lower rates of solid transport associated with reduced flows (Murdoch et al., 2000; Caruso, 2002). At a more local scale, increasing sediment concentrations during drought have also been observed in streams adjacent to land with high livestock and bushland densities, which both constantly contribute to sediment load during drought, leading to elevated concentrations with a lower dilution rate (Caruso, 2002). Similar to sediments, the impact of droughts on stream nutrient and salt concentrations have also commonly been understood to be responses to reduced runoff generation and streamflow. In catchments with no significant point-source pollution, nutrient concentrations typically decrease during droughts (Mosley, 2015) with less nutrient leaching and overland flow, but they may also increase due to increasing livestock inputs at more local scales (Caruso, 2002). In contrast, catchments with significant point-source pollution generally experience water quality deterioration during drought due to reduced dilution (van Vliet and Zwolsman, 2008; Mosley, 2015). With respect to salinity, the concentration often increases during drought with reduced dilution and increased evaporation (Caruso, 2002). This is particularly evident for catchments that are more influenced by saline groundwater input, as the relative contribution of groundwater increases during drought (Costelloe et al., 2005).

In contrast to these previous studies, our findings suggest additional possible pathways via which drought can affect stream water quality and that prolonged drought might alter the relationships between sediment and its predictors (Figs. 7, 8). In contrast to sediment, our model suggests no clear shifts in the dynamics of nutrients and salts at a regional scale. Our findings are in line with a few previous studies that reported temporal changes in the concentration–discharge relationships for sediments and nutrients, specifically when comparing high- and low-flow conditions (Zhang, 2018; Moatar et al., 2017) and drought and recovery periods (Burt et al., 2015). Our findings provide extra dimensions to what would be offered by simple trend analyses using ap-

proaches such as the Mann–Kendall test or the Sen's slope (e.g., Smith et al., 1987; Chang, 2008; Hirsch et al., 1991; Bouza-Deaño et al., 2008). These approaches are only capable of indicating the direction and magnitude of observed trends. In contrast, our model was able to attribute the consistent upward shift in the TSS concentration to the change in the relationships between water quality and its key driving factors since the start of drought.

In addition, we also acknowledge that our ability to represent the pre- and post-drought conditions in this study may be limited by the record length, as only 2 years of pre-drought and 4 years of post-drought data were available. Once longer records build up, they will enable us to update our understanding of the impact of this prolonged drought. We would also be able to conduct more sophisticated investigations, such as comparing the impacts of long-term droughts versus individual dry and wet years and events (e.g., Saft et al., 2015; Outram et al., 2014; Burt et al., 2015).

5 Conclusions

This study aims to address the current lack of water quality models that operate at large scales across multiple catchments. To achieve this, we used long-term stream water quality data collected from 102 sites in southeastern Australia and developed a Bayesian hierarchical statistical model to simulate the spatiotemporal variabilities in six key water quality constituents: TSS, TP, FRP, TKN, NO_x and EC. The choice of model predictors was guided by previous studies on the same dataset (Lintern et al., 2018b; Guo et al., 2019). The model generally captures the spatiotemporal variability in water quality well, where the spatial variability between catchments is much better represented than temporal variability. The model is best used to predict proportional changes in water quality on a Box–Cox-transformed scale and can have substantial bias if used to predict absolute values for high concentrations. Cross-validation shows that the spatiotemporal model can predict water quality in non-monitored locations under similar conditions to the historical period and the calibration catchments that we investigated. This can assist management by (1) identifying hot spots and key temporal periods for waterway pollution; (2) testing the effects of catchment changes, e.g., urbanization or afforestation; and (3) identifying and attributing major water quality trends and changes.

Based on the above model evaluations, we discussed potential ways to further enhance the model performance. In improving the modeling framework, alternative statistical approaches could be considered to reduce the impact of below-detection-limit data on model performance. In addition, the models could be extended to consider some key biogeochemical processes to improve the dynamics in nonconservative constituents (e.g., FRP or NO_x). Regarding data availability, the current models could potentially benefit from improved

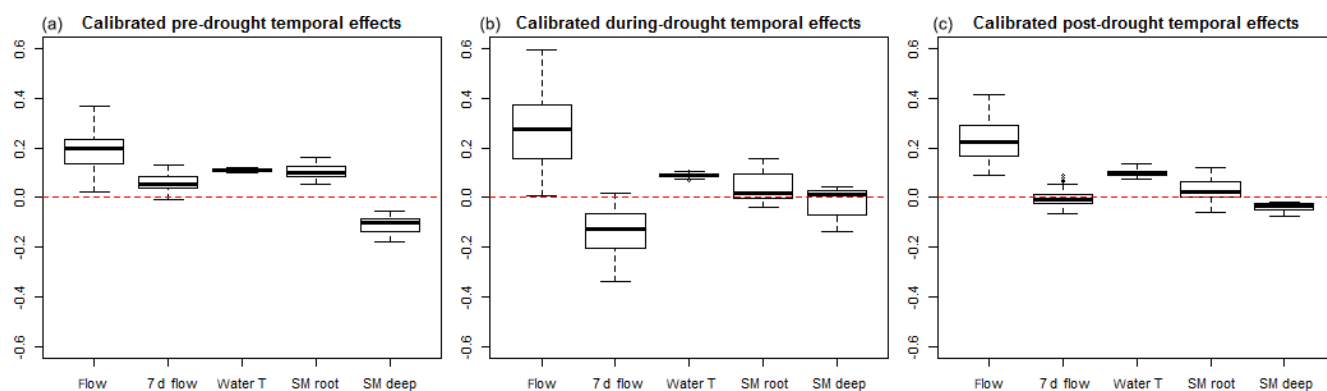


Figure 9. The effects of the five key predictors of the temporal variability in TSS across the 102 sites summarized by the posterior mean of the calibrated parameter values for each predictor (box shows values across all sites): flow, 7 d antecedent flow, water temperature, root zone soil moisture and deep soil moisture.

monitoring of changes in land use intensity and management in order to be able to include these drivers in the model. The inclusion of high-frequency water quality sampling data may also extend the model's ability to represent temporal variability. However, high-frequency water quality data are also typically highly variable with large noise. Therefore, the implication of such data for the spatiotemporal modeling framework remains an open question that requires further investigation in future applications of this modeling framework.

Data availability. All data used in this study were extracted from the public domain. All stream water quality data were extracted from the Victorian Water Measurement Information System (Department of Environment, Land, Water and Planning Victoria, 2016; <http://data.water.vic.gov.au/>). The catchments corresponding to these water quality monitoring sites were delineated using the Geofabric tool (<ftp://ftp.bom.gov.au/anon/home/geofabric/>; Bureau of Meteorology, 2012). All other data for the spatial and temporal predictors of our models are listed in Tables S1 and S2 in the Supplement.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/hess-24-827-2020-supplement>.

Author contributions. All authors contributed to the conceptualization the models and the design of methodology. AL and SL contributed to the data curation. DG carried out the formal analyses, visualization and validation. JAW, DR, UBM and AWW contributed to the funding acquisition. DG, AL, JAW, DR, SL and AWW contributed to the investigation. DG carried out project administration and coding to run the experiments. JAW, DR and AWW contributed to the supervision. DG prepared the paper with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors would also like to thank Matthew Johnson, Louise Sullivan, Hannah Sleeth and Jie Jian, for their assistance in the compilation and analysis of data. All water quality data used for this project can be found on the Water Measurement Information System (<http://data.water.vic.gov.au/>). The sources of other data are provided in Tables S1 and S2 in the Supplement.

Financial support. This research has been supported by the Australian Research Council via a Linkage Project (grant no. LP140100495), with contributions from the following industrial collaborators: the Victorian Environment Protection Authority; the Victorian Department of Environment, Land, Water and Planning; the Australian Bureau of Meteorology; and the Queensland Department of Natural Resources, Mines and Energy.

Review statement. This paper was edited by Christian Stamm and reviewed by Matthias Gassmann, Mark Honti and two anonymous referees.

References

- Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., and Kløve, B.: A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model, *J. Hydrol.*, 524, 733–752, <https://doi.org/10.1016/j.jhydrol.2015.03.027>, 2015.
- Adams, R., Arafat, Y., Eate, V., Grace, M. R., Saffarpour, S., Weatherley, A. J., and Western, A. W.: A catchment study of sources and sinks of nutrients and sediments in south-east Australia, *J. Hydrol.*, 515, 166–179, <https://doi.org/10.1016/j.jhydrol.2014.04.034>, 2014.

- Ahearn, D. S., Sheibley, R. W., Dahlgren, R. A., and Keller, K. E.: Temporal dynamics of stream water chemistry in the last free-flowing river draining the western Sierra Nevada, California, *J. Hydrol.*, 295, 47–63, <https://doi.org/10.1016/j.jhydrol.2004.02.016>, 2004.
- Ai, L., Shi, Z. H., Yin, W., and Huang, X.: Spatial and seasonal patterns in stream water contamination across mountainous watersheds: Linkage with landscape characteristics, *J. Hydrol.*, 523, 398–408, <https://doi.org/10.1016/j.jhydrol.2015.01.082>, 2015.
- Akaike, H.: A new look at the statistical model identification, *IEEE T. Automat. Contr.*, 19, 716–723, 1974.
- Akritis, M. G., Murphy, S. A., and Lavalley, M. P.: The Theil-Sen estimator with doubly censored data and applications to astronomy, *J. Am. Stat. Assoc.*, 90, 170–177, 1995.
- Ali, G., Wilson, H., Elliott, J., Penner, A., Haque, A., Ross, C., and Rabie, M.: Phosphorus export dynamics and hydrobiogeochemical controls across gradients of scale, topography and human impact, *Hydrol. Process.*, 31, 3130–3145, <https://doi.org/10.1002/hyp.11258>, 2017.
- Arheimer, B. and Lidén, R.: Nitrogen and phosphorus concentrations from agricultural catchments—influence of spatial and temporal variables, *J. Hydrol.*, 227, 140–159, [https://doi.org/10.1016/S0022-1694\(99\)00177-8](https://doi.org/10.1016/S0022-1694(99)00177-8), 2000.
- Australian Water Technologies: Victorian water quality monitoring network and state biological monitoring programme manual of procedures, Australian Water Technologies, 68 Ricketts Rd, Mt Waverley VIC 3149, 1999.
- Bailey, R. T. and Ahmadi, M.: Spatial and temporal variability of in-stream water quality parameter influence on dissolved oxygen and nitrate within a regional stream network, *Ecol. Model.*, 277, 87–96, 2014.
- Bende-Michl, U. and Hairsine, P. B.: A systematic approach to choosing an automated nutrient analyser for river monitoring, *J. Environ. Monitor.*, 12, 127–134, 2010.
- Bond, N. R., Lake, P. S., and Arthington, A. H.: The impacts of drought on freshwater ecosystems: an Australian perspective, *Hydrobiologia*, 600, 3–16, <https://doi.org/10.1007/s10750-008-9326-z>, 2008.
- Borsuk, M. E., Higdon, D., Stow, C. A., and Reckhow, K. H.: A Bayesian hierarchical model to predict benthic oxygen demand from organic matter loading in estuaries and coastal zones, *Ecol. Model.*, 143, 165–181, [https://doi.org/10.1016/S0304-3800\(01\)00328-3](https://doi.org/10.1016/S0304-3800(01)00328-3), 2001.
- Bouza-Deaño, R., Ternero-Rodríguez, M., and Fernández-Espinoza, A. J.: Trend study and assessment of surface water quality in the Ebro River (Spain), *J. Hydrol.*, 361, 227–239, <https://doi.org/10.1016/j.jhydrol.2008.07.048>, 2008.
- Box, G. E. and Cox, D. R.: An analysis of transformations, *J. Roy. Stat. Soc. B*, 26, 211–243, 1964.
- Bureau of Meteorology: Geofabric V2, available at: <ftp://ftp.bom.gov.au/anon/home/geofabric/> (last access: 21 September 2016), 2012.
- Bureau of Rural Sciences: 2005/06 Land use of Australia, version 4, available at: <http://www.agriculture.gov.au/abares/aclump/land-use/data-download> (last access: 1 September 2016), 2010.
- Burt, T. P., Worrall, F., Howden, N. J. K., and Anderson, M. G.: Shifts in discharge-concentration relationships as a small catchment recover from severe drought, *Hydrol. Process.*, 29, 498–507, <https://doi.org/10.1002/hyp.10169>, 2015.
- Carey, R. O. and Migliaccio, K. W.: Contribution of wastewater treatment plant effluents to nutrient dynamics in aquatic systems: a review, *Environ. Manage.*, 44, 205–217, <https://doi.org/10.1007/s00267-009-9309-5>, 2009.
- Caruso, B. S.: Temporal and spatial patterns of extreme low flows and effects on stream ecosystems in Otago, New Zealand, *J. Hydrol.*, 257, 115–133, [https://doi.org/10.1016/S0022-1694\(01\)00546-7](https://doi.org/10.1016/S0022-1694(01)00546-7), 2002.
- Chang, H.: Spatial analysis of water quality trends in the Han River basin, South Korea, *Water Res.*, 42, 3285–3304, <https://doi.org/10.1016/j.watres.2008.04.006>, 2008.
- Clark, J. S.: Why environmental scientists are becoming Bayesians, *Ecol. Lett.*, 8, 2–14, <https://doi.org/10.1111/j.1461-0248.2004.00702.x>, 2005.
- Costelloe, J. F., Grayson, R. B., McMahon, T. A., and Argent, R. M.: Spatial and temporal variability of water salinity in an ephemeral, arid-zone river, central Australia, *Hydrol. Process.*, 19, 3147–3166, <https://doi.org/10.1002/hyp.5837>, 2005.
- DeFries, R. and Eshleman, K. N.: Land-use change and hydrologic processes: a major focus for the future, *Hydrol. Process.*, 18, 2183–2186, <https://doi.org/10.1002/hyp.5584>, 2004.
- Department of Environment Land Water and Planning Victoria: Victorian water measurement information system, available at: <http://data.water.vic.gov.au/> (last access: 1 January 2017), 2016.
- Eidenshink, J. C.: The 1990 Conterminous U.S. AVHRR Data Set, *Photogrammetric Engineering and Remote Sensing*, 58, 809–813, 1992.
- Fraser, A. I., Harrod, T. R., and Haygarth, P. M.: The effect of rainfall intensity on soil erosion and particulate phosphorus transfer from arable soils, *Water Sci. Technol.*, 39, 41–45, [https://doi.org/10.1016/S0273-1223\(99\)00316-9](https://doi.org/10.1016/S0273-1223(99)00316-9), 1999.
- Frost, A. J., Ramchurn, A., and Smith, A.: The bureau's operational AWRA landscape (AWRA-L) Model, Bureau of Meteorology Technical Report, Bureau of Meteorology, 2016.
- Fu, B., Merritt, W. S., Croke, B. F. W., Weber, T., and Jakeman, A. J.: A review of catchment-scale water quality and erosion models and a synthesis of future prospects, *Environ. Modell. Softw.*, 114, 75–97, <https://doi.org/10.1016/j.envsoft.2018.12.008>, 2018.
- Gelman, A.: Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), *Bayesian Anal.*, 1, 515–534, <https://doi.org/10.1214/06-BA117A>, 2006.
- Geoscience Australia: Dams and water storages, available at: <https://koordinates.com/layer/739-australian-dams-and-water-storages/> (last access: 1 September 2016), 2004.
- Geoscience Australia: Environmental Attributes Dataset, available at: <http://www.ga.gov.au> (last access: 5 February 2016), 2011.
- Giri, S. and Qiu, Z.: Understanding the relationship of land uses and water quality in Twenty First Century: A review, *J. Environ. Manage.*, 173, 41–48, <https://doi.org/10.1016/j.jenvman.2016.02.029>, 2016.
- Granger, S. J., Bol, R., Anthony, S., Owens, P. N., White, S. M., and Haygarth, P. M.: Chapter 3 – Towards a Holistic Classification of Diffuse Agricultural Water Pollution from Intensively Managed Grasslands on Heavy Soils, in: *Advances in Agronomy*, Academic Press, 83–115, 2010.
- Guo, D., Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Leahy, P., Wilson, P., and Western, A. W.: Key Factors Af-

- fecting Temporal Variability in Stream Water Quality, *Water Resour. Res.*, 55, 112–129, <https://doi.org/10.1029/2018wr023370>, 2019.
- Heathwaite, A. L.: Multiple stressors on water availability at global to catchment scales: understanding human impact on nutrient cycles to protect water quality and water availability in the long term, *Freshwater Biol.*, 55, 241–257, <https://doi.org/10.1111/j.1365-2427.2009.02368.x>, 2010.
- Hirsch, R. M., Alexander, R. B., and Smith, R. A.: Selection of methods for the detection and estimation of trends in water quality, *Water Resour. Res.*, 27, 803–813, <https://doi.org/10.1029/91wr00259>, 1991.
- Hrachowitz, M., Benettin, P., van Breukelen, B. M., Fovet, O., Howden, N. J. K., Ruiz, L., van der Velde, Y., and Wade, A. J.: Transit times – the link between hydrology and water quality at the catchment scale, *Wiley Interdisciplinary Reviews: Water*, 3, 629–657, <https://doi.org/10.1002/wat2.1155>, 2016.
- Jarvie, H. P., Withers, J. A., and Neal, C.: Review of robust measurement of phosphorus in river water: sampling, storage, fractionation and sensitivity, *Hydrol. Earth Syst. Sci.*, 6, 113–131, <https://doi.org/10.5194/hess-6-113-2002>, 2002.
- Kingsford, R. T., Walker, K. F., Lester, R. E., Young, W. J., Fairweather, P. G., Sammut, J., and Geddes, M. C.: A Ramsar wetland in crisis – the Coorong, Lower Lakes and Murray Mouth, Australia, *Mar. Freshwater Res.*, 62, 255–265, <https://doi.org/10.1071/MF09315>, 2011.
- Kirchner, J. W., Feng, X., Neal, C., and Robson, A. J.: The fine structure of water-quality dynamics: the (high-frequency) wave of the future, *Hydrol. Process.*, 18, 1353–1359, <https://doi.org/10.1002/hyp.5537>, 2004.
- Kisi, O. and Parmar, K. S.: Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution, *J. Hydrol.*, 534, 104–112, <https://doi.org/10.1016/j.jhydrol.2015.12.014>, 2016.
- Kurunç, A., Yürekli, K., and Çevik, O.: Performance of two stochastic approaches for forecasting water quality and streamflow data from Yeşilirmak River, Turkey, *Environ. Modell. Softw.*, 20, 1195–1200, <https://doi.org/10.1016/j.envsoft.2004.11.001>, 2005.
- Larned, S. T., Scarsbrook, M. R., Snelder, T. H., Norton, N. J., and Biggs, B. J. F.: Water quality in low-elevation streams and rivers of New Zealand: Recent state and trends in contrasting land-cover classes, *New Zeal. J. Mar. Fresh.*, 38, 347–366, <https://doi.org/10.1080/00288330.2004.9517243>, 2004.
- Lannergård, E. E., Ledesma, J. L., Fölster, J., and Futter, M. N.: An evaluation of high frequency turbidity as a proxy for riverine total phosphorus concentrations, *Sci. Total Environ.*, 651, 103–113, 2019.
- Leblanc, M., Tweed, S., Van Dijk, A., and Timbal, B.: A review of historic and future hydrological changes in the Murray-Darling Basin, *Global Planet. Chang.*, 80–81, 226–246, <https://doi.org/10.1016/j.gloplacha.2011.10.012>, 2012.
- Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Waters, D., Leahy, P., Wilson, P., and Western, A. W.: Key factors influencing differences in stream water quality across space, *Wiley Interdisciplinary Reviews: Water*, 5, e1260, <https://doi.org/10.1002/wat2.1260>, 2018a.
- Lintern, A., Webb, J. A., Ryu, D., Liu, S., Waters, D., Leahy, P., Bende-Michl, U., and Western, A. W.: What Are the Key Catchment Characteristics Affecting Spatial Differences in Riverine Water Quality?, *Water Resour. Res.*, 54, 7252–7272, <https://doi.org/10.1029/2017WR022172>, 2018b.
- Luca Scrucca: Package “GA”, The Comprehensive R Archive Network, available at: <https://cran.r-project.org/web/packages/GA/index.html> (last access: 1 November 2018), 2019.
- May, R., Dandy, G., and Maier, H.: Review of input variable selection methods for artificial neural networks, in: *Artificial neural networks-methodological advances and biomedical applications*, InTech, 2011.
- Mellander, P.-E., Jordan, P., Shore, M., Melland, A. R., and Shortle, G.: Flow paths and phosphorus transfer pathways in two agricultural streams with contrasting flow controls, *Hydrol. Process.*, 29, 3504–3518, <https://doi.org/10.1002/hyp.10415>, 2015.
- Meybeck, M. and Helmer, R.: The quality of rivers: From pristine stage to global pollution, *Palaeogeography, Palaeoclimatology, Palaeoecology*, 75, 283–309, [https://doi.org/10.1016/0031-0182\(89\)90191-0](https://doi.org/10.1016/0031-0182(89)90191-0), 1989.
- Miller, C., Magdalina, A., Willows, R. I., Bowman, A. W., Scott, E. M., Lee, D., Burgess, C., Pope, L., Pannullo, F., and Haggarty, R.: Spatiotemporal statistical modelling of long-term change in river nutrient concentrations in England & Wales, *Sci. Total Environ.*, 466–467, 914–923, <https://doi.org/10.1016/j.scitotenv.2013.07.113>, 2014.
- Moatar, F., Abbott, B. W., Minaudo, C., Curie, F., and Pinay, G.: Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions, *Water Resour. Res.*, 53, 1270–1287, <https://doi.org/10.1002/2016wr019635>, 2017.
- Mosley, L. M.: Drought impacts on the water quality of freshwater systems; review and integration, *Earth-Sci. Rev.*, 140, 203–214, <https://doi.org/10.1016/j.earscirev.2014.11.010>, 2015.
- Murdoch, P. S., Baron, J. S., and Miller, T. L.: POTENTIAL EFFECTS OF CLIMATE CHANGE ON SURFACE-WATER QUALITY IN NORTH AMERICA, *J. Am. Water Resour. As.*, 36, 347–366, <https://doi.org/10.1111/j.1752-1688.2000.tb04273.x>, 2000.
- Musolff, A., Schmidt, C., Selle, B., and Fleckenstein, J. H.: Catchment controls on solute export, *Adv. Water Resour.*, 86, 133–146, <https://doi.org/10.1016/j.advwatres.2015.09.026>, 2015.
- NASA LP DAAC: MOD13A3: MODIS/Terra Vegetation Indices Monthly L3 Global 1 km V005, available at: <https://lpdaac.usgs.gov/products/mod13a3v006/>, last access: 16 August 2017.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6), 1970.
- Onderka, M., Wrede, S., Rodný, M., Pfister, L., Hoffmann, L., and Krein, A.: Hydrogeologic and landscape controls of dissolved inorganic nitrogen (DIN) and dissolved silica (DSi) fluxes in heterogeneous catchments, *J. Hydrol.*, 450–451, 36–47, <https://doi.org/10.1016/j.jhydrol.2012.05.035>, 2012.
- Outram, F. N., Lloyd, C. E. M., Jonezyk, J., Benskin, C. McW. H., Grant, F., Perks, M. T., Deasy, C., Burke, S. P., Collins, A. L., Freer, J., Haygarth, P. M., Hiscock, K. M., Johnes, P. J., and Lovett, A. L.: High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in England, *Hydrol. Earth Syst. Sci.*, 18, 3429–3448, <https://doi.org/10.5194/hess-18-3429-2014>, 2014.

- Ouyang, W., Hao, F., Skidmore, A. K., and Toxopeus, A. G.: Soil erosion and sediment yield and their relationships with vegetation cover in upper stream of the Yellow River, *Sci. Total Environ.*, 409, 396–403, <https://doi.org/10.1016/j.scitotenv.2010.10.020>, 2010.
- Parmar, K. S. and Bhardwaj, R.: Statistical, time series, and fractal analysis of full stretch of river Yamuna (India) for water quality management, *Environ. Sci. Pollut. Res.*, 22, 397–414, <https://doi.org/10.1007/s11356-014-3346-1>, 2015.
- Pellerin, B. A., Saraceno, J. F., Shanley, J. B., Sebestyen, S. D., Aiken, G. R., Wollheim, W. M., and Bergamaschi, B. A.: Taking the pulse of snowmelt: in situ sensors reveal seasonal, event and diurnal patterns of nitrate and dissolved organic matter variability in an upland forest stream, *Biogeochemistry*, 108, 183–198, <https://doi.org/10.1007/s10533-011-9589-8>, 2012.
- Pellerin, B. A., Stauffer, B. A., Young, D. A., Sullivan, D. J., Bricker, S. B., Walbridge, M. R., Clyde Jr., G. A., and Shaw, D. M.: Emerging Tools for Continuous Nutrient Monitoring Networks: Sensors Advancing Science and Water Resources Protection, *J. Am. Water Resour. As.*, 52, 993–1008, <https://doi.org/10.1111/1752-1688.12386>, 2016.
- Poor, C. J. and McDonnell, J. J.: The effects of land use on stream nitrate dynamics, *J. Hydrol.*, 332, 54–68, <https://doi.org/10.1016/j.jhydrol.2006.06.022>, 2007.
- Poudel, D. D., Lee, T., Srinivasan, R., Abbaspour, K., and Jeong, C. Y.: Assessment of seasonal and spatial variation of surface water quality, identification of factors associated with water quality variability, and the modeling of critical nonpoint source pollution areas in an agricultural watershed, *J. Soil Water Conserv.*, 68, 155–171, <https://doi.org/10.2489/jswc.68.3.155>, 2013.
- Poulsen, D. L., Simmons, C. T., Le Galle La Salle, C., and Cox, J. W.: Assessing catchment-scale spatial and temporal patterns of groundwater and stream salinity, *Hydrogeol. J.*, 14, 1339–1359, <https://doi.org/10.1007/s10040-006-0065-9>, 2006.
- Qin, B., Zhu, G., Gao, G., Zhang, Y., Li, W., Paerl, H. W., and Carmichael, W. W.: A Drinking Water Crisis in Lake Taihu, China: Linkage to Climatic Variability and Lake Management, *Environ. Manage.*, 45, 105–112, <https://doi.org/10.1007/s00267-009-9393-6>, 2010.
- Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., and Trudinger, C.: Australian water availability project (AWAP): CSIRO marine and atmospheric research component: final report for phase 3, 67, 2009.
- Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., and Trudinger, C.: Australian Water Availability Project, CSIRO Marine and Atmospheric Research, Canberra, Australia, 2012.
- Ren, W., Zhong, Y., Meligrana, J., Anderson, B., Watt, W. E., Chen, J., and Leung, H.-L.: Urbanization, land use, and water quality in Shanghai: 1947–1996, *Environ. Int.*, 29, 649–659, [https://doi.org/10.1016/S0160-4120\(03\)00051-5](https://doi.org/10.1016/S0160-4120(03)00051-5), 2003.
- Saft, M., Western, A. W., Zhang, L., Peel, M. C., and Potter, N. J.: The influence of multiyear drought on the annual rainfall-runoff relationship: An Australian perspective, *Water Resour. Res.*, 51, 2444–2463, <https://doi.org/10.1002/2014WR015348>, 2015.
- Saft, M., Peel, M. C., Western, A. W., and Zhang, L.: Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics, *Water Resour. Res.*, 52, 9290–9305, <https://doi.org/10.1002/2016WR019525>, 2016.
- Saraceno, J. F., Pellerin, B. A., Downing, B. D., Boss, E., Bachand, P. A. M., and Bergamaschi, B. A.: High-frequency in situ optical measurements during a storm event: Assessing relationships between dissolved organic matter, sediment concentrations, and hydrologic processes, *J. Geol. Res.*, 114, G00F09, <https://doi.org/10.1029/2009JG000989>, 2009.
- Schwarz, G.: Estimating the dimension of a model, *Ann. Stat.*, 6, 461–464, 1978.
- Sharpley, A. N., Kleinman, P. J. A., McDowell, R. W., Gitau, M., and Bryant, R. B.: Modeling phosphorus transport in agricultural watersheds: Processes and possibilities, *J. Soil Water Conserv.*, 57, 425–439, 2002.
- Smith, A. P., Western, A. W., and Hannah, M. C.: Linking water quality trends with land use intensification in dairy farming catchments, *J. Hydrol.*, 476, 1–12, 2013.
- Smith, R. A., Alexander, R. B., and Wolman, M. G.: Water-Quality Trends in the Nation's Rivers, *Science*, 235, 1607–1615, 1987.
- Smyth, A. R., Thompson, S. P., Siporin, K. N., Gardner, W. S., McCarthy, M. J., and Piehler, M. F.: Assessing nitrogen dynamics throughout the estuarine landscape, *Estuar. Coasts*, 36, 44–55, 2013.
- Stan Development Team: RStan: the R interface to Stan. R package version 2.18.1, 2018.
- Stan Reference Manual Version 2.20: available at: https://mc-stan.org/docs/2_20/reference-manual-2_20.pdf, last access: 28 September 2019.
- Sturtz, S., Ligges, U., and Gelman, A. E.: R2WinBUGS: a package for running WinBUGS from R, 2005.
- Sueker, J. K., Clow, D. W., Ryan, J. N., and Jarrett, R. D.: Effect of basin physical characteristics on solute fluxes in nine alpine/subalpine basins, Colorado, USA, *Hydrol. Process.*, 15, 2749–2769, <https://doi.org/10.1002/hyp.265>, 2001.
- Tang, Z., Engel, B. A., Pijanowski, B. C., and Lim, K. J.: Forecasting land use change and its environmental impact at a watershed scale, *J. Environ. Manage.*, 76, 35–45, <https://doi.org/10.1016/j.jenvman.2005.01.006>, 2005.
- Terrestrial Ecosystem Research Network: Soil and landscape grid of Australia, available at: <http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html>, last access: 7 July 2016.
- Tian, J. R. and Zhou, P. J.: Phosphorus fractions of floodplain sediments and phosphorus exchange on the sediment–water interface in the lower reaches of the Han River in China, *Ecol. Eng.*, 30, 264–270, <https://doi.org/10.1016/j.ecoleng.2007.01.006>, 2007.
- Tramblay, Y., Ouara, T. B. M. J., St-Hilaire, A., and Poulin, J.: Regional estimation of extreme suspended sediment concentrations using watershed characteristics, *J. Hydrol.*, 380, 305–317, <https://doi.org/10.1016/j.jhydrol.2009.11.006>, 2010.
- van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resour. Res.*, 49, 1040–1057, <https://doi.org/10.1002/wrcr.20123>, 2013.
- van Vliet, M. T. H. and Zwolsman, J. J. G.: Impact of summer droughts on the water quality of the Meuse river, *J. Hydrol.*, 353, 1–17, <https://doi.org/10.1016/j.jhydrol.2008.01.001>, 2008.
- Varanka, S., Hjort, J., and Luoto, M.: Geomorphological factors predict water quality in boreal rivers, *Earth Surf. Proc. Land.*, 40, 1989–1999, <https://doi.org/10.1002/esp.3601>, 2015.

- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R., and Davies, P. M.: Global threats to human water security and river biodiversity, *Nature*, 467, , 555, <https://doi.org/10.1038/nature09440>, 2010.
- Wang, Q. J. and Robertson, D. E.: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, *Water Resour. Res.*, 47, W02546, <https://doi.org/10.1029/2010wr009333>, 2011.
- Wang, Q. J., Shrestha, D. L., Robertson, D. E., and Pokhrel, P.: A log-sinh transformation for data normalization and variance stabilization, *Water Resour. Res.*, 48, W05514, <https://doi.org/10.1029/2011WR010973>, 2012.
- Webb, J. A. and King, L. E.: A Bayesian hierarchical trend analysis finds strong evidence for large-scale temporal declines in stream ecological condition around Melbourne, Australia, *Ecography*, 32, 215–225, <https://doi.org/10.1111/j.1600-0587.2008.05686.x>, 2009.
- Whitworth, K. L., Baldwin, D. S., and Kerr, J. L.: Drought, floods and water quality: Drivers of a severe hypoxic black-water event in a major river system (the southern Murray–Darling Basin, Australia), *J. Hydrol.*, 450–451, 190–198, <https://doi.org/10.1016/j.jhydrol.2012.04.057>, 2012.
- Zhang, Q.: Synthesis of nutrient and sediment export patterns in the Chesapeake Bay watershed: Complex and non-stationary concentration-discharge relationships, *Sci. Total Environ.*, 618, 1268–1283, <https://doi.org/10.1016/j.scitotenv.2017.09.221>, 2018.
- Zhang, Q. and Ball, W. P.: Improving riverine constituent concentration and flux estimation by accounting for antecedent discharge conditions, *J. Hydrol.*, 547, 387–402, <https://doi.org/10.1016/j.jhydrol.2016.12.052>, 2017.
- Zhao, T., Schepen, A., and Wang, Q. J.: Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach, *J. Hydrol.*, 541, 839–849, <https://doi.org/10.1016/j.jhydrol.2016.07.040>, 2016.
- Zhou, T., Wu, J., and Peng, S.: Assessing the effects of landscape pattern on river water quality at multiple scales: A case study of the Dongjiang River watershed, China, *Ecol. Indic.*, 23, 166–175, <https://doi.org/10.1016/j.ecolind.2012.03.013>, 2012.