https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions
Open Access

A predictive model for spatio-temporal variability in stream water quality

Danlu Guo[1], Anna Lintern[1,2], J. Angus Webb[1], Dongryeol Ryu[1], Ulrike Bende-Michl[3], Shuci Liu[1], Andrew William Western[1]

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

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

[3] Bureau of Meteorology, Parkes, ACT, 2600.

Corresponding author's email : danlu.guo@unimelb.edu.au

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

Abstract

Degraded water quality in rivers and streams can have large economic, societal and ecological impacts. Stream water quality can be highly variable both over space and time. To develop effective management strategies for riverine water quality, it is critical to be able to predict these spatio-temporal variabilities. However, our current capacity to model stream water quality is limited, particularly at large spatial scales across multiple catchments. This is due to a lack of understanding of the key controls that drive spatio-temporal variabilities of stream water quality. To address this, we developed a Bayesian hierarchical statistical model to analyse the spatio-temporal variability in stream water quality across the state of Victoria, Australia. The model was developed based on monthly water quality monitoring data collected at 102 sites over 21 years. The modelling 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). Among the six constituents, the models explained varying proportions of variation in water quality. EC was the most predictable constituent (88.6% variability explained) and FRP had the lowest predictive performance (19.9% variability explained). The models were validated for multiple sets of calibration/validation sites and showed robust performance. Temporal validation revealed a systematic change in the TSS model performance across most catchments since an extended drought period in the study region, highlighting potential shifts in TSS dynamics over the drought. Further improvements in model performance need to focus on: (1) alternative statistical model structures to improve fitting for the low concentration data, especially records below the detection limit; and (2) better representation of non-conservative constituents by accounting for important biogeochemical processes. We also recommend future improvements in water quality monitoring programs which can potentially enhance the model capacity, via: 1) improving the monitoring and assimilation of high-frequency water quality data; and 2) improving the availability of data to capture land use and management changes over time.

Keywords

stream water quality; spatio-temporal variability; sediments; nutrients; statistical modeling; Bayesian hierarchical model

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

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). However, our ability to manage and mitigate water quality impacts is hampered by the variability in water quality both across space and time, and our inability to predict this variability (Chang, 2008;Bengraïne and Marhaba, 2003;Ai et al., 2015). Water quality conditions can vary across individual 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 significantly across locations (Meybeck and Helmer, 1989). These variabilities in stream water quality are driven by three key mechanisms: (1) the source of constituents, which defines the total amount of constituents being available in a catchment; (2) the mobilization of constituents in particulate and dissolved forms, which detaches constituents 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 natural catchment characteristics (e.g., climate, geology, soil type, topography and hydrology) as well as by human activities within catchments (e.g., land use and management, vegetation cover etc.), all of which control the extent and magnitude of the three key mechanisms described above (Lintern et al., 2018a). At the same time, temporal shifts in water quality are influenced by changes in climatic, hydrological and other catchment conditions, such as temperature (Roberts and Mulholland, 2007), the timing and magnitude of rainfall events (Fraser et al., 1999), runoff generation and streamflow (Ahearn et al., 2004; Mellander et al., 2015; Sharpley et al., 2002), and vegetation cover changes over time (Kaushal et al., 2014;Ouyang et al., 2010).

Despite undertstanding of the basic mechanisms, we currently lack the ability to model these spatio-temporal variabilities at larger scales to inform the development of effective policy and mitigation strategies. Conceptual or physically-based water quality models are typically limited by the simplification of physical processes (Hrachowitz et al., 2016). Furthermore, practical implementation of these models can be also limited by the intensive requirements of data and calibration effort,

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

particularly for large regions with varying catchment conditions (Fu et al., 2018;Abbaspour et al., 2015). In contrast, whilst most statistical water quality models are easier to implement, they often 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). Consequently, it remains challenging to address spatio-temporal variability simulaneously over long time periods and large regions. This lack of integrated modelling of both spatial and temporal variability in water quality can not only limits our understanding of the key factors that affect water quality dynamics over both of these dimensions. It also hinders our ability to predict future water quality changes in un-monitored locations.

The aim of this research is to develop a data-driven model to predict spatio-temporal changes in stream water quality. This model was established using long-term (21 years) stream water quality observations across 102 catchments in the state of Victoria, Australia. The model built on two previous studies on the same dataset that identified the key drivers for water quality spatial and temporal variabilities, respectively (Lintern et al., 2018b; Guo et al., 2019). Our approach aims to bridge the gap between fully-distributed water quality models and statistical approaches to provide useful information for catchment managers, especially for large-scale water quality assessments.

2. Method

2.1 Spatio-temporal modelling framework

A Bayesian hierarchical approach was used to model the spatio-temporal variability in stream water quality. The Bayesian approach enables the inherent natural stochasticity of water quality to be incorporated into the model (Clark, 2005), and the hierarchical model structure enables the key controls of temporal variability in water quality to vary across locations (Webb and King, 2009;Borsuk et al., 2001).

The structure of this model is presented below in Eq. 1 to 6. The transformed concentration of a constituent (see Sect. 2.2 for justification) at time i and site j (C_{ij}) is assumed to be normally distributed with a mean μ_{ij} and standard deviation σ representing inherent randomness (Eq. 1).

$$\boldsymbol{C_{ij} \sim N(\mu_{ij}, \sigma)} \qquad (1)$$

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Open Access
Discussions

To represent spatio-temporal variability, μ_{ij} is modelled as the sum of the site-level mean constituent concentration ($\bar{C_j}$) and the deviation from that mean at time i (Δ_{ij}) (Eq. 2).

$$\boldsymbol{\mu_{ij}} = \boldsymbol{\overline{C}_j} + \Delta_{\boldsymbol{ij}} \tag{2}$$

To describe spatial variability, the site-level mean ($\bar{C_j}$) is modelled as a linear function of a global intercept (*int*), and the sum of the *m* catchment characteristics $S_{1,j}$ to $S_{m,j}$ (e.g. land use, topography) weighted by their relative contributions to spatial varaibility (β_S_1 to β_S_m) (Eq. 3).

$$\boldsymbol{\overline{C}_j = int + \beta\_S_1 \times S_{1,j} + +\beta\_S_2 \times S_{2,j} + \cdots + \beta\_S_m \times S_{m,j}} \tag{3}$$

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*.

$$\Delta_{\boldsymbol{ij}} = \boldsymbol{\beta\_T_{1,j} \times T_{1,ij} + \cdots + \beta\_T_{n,j} \times T_{n,ij}} \tag{4}$$

To account for differences in these temporal influences across sites, the effect of each temporal variable at site *j* ($\beta\_T_{N,j}$ with N in 1,2, … n) is drawn from a distribution with a mean of $N_{\beta\_T_{N,j}}$ (Eq. 5), which is then modelled with a linear combination of two additional chatchment characteristics, $S_{TN1,j}$ and $S_{TN2,j}$ (Eq. 6).

$$\boldsymbol{\beta\_T_{N,j} \sim N\left(\mu_{\beta\_T_{N,j}}, \sigma_{\beta\_T_N}\right), for\ N\ in\ 1, 2, \dots} \tag{5}$$

$$\boldsymbol{N_{\beta\_T_{N,j}} = int_{\beta\_TN} + \beta\_S_{TN1} \times S_{TN1,j} + \beta\_S_{TN2} \times S_{TN2,j}} \tag{6}$$

Section 2.2 introduces the data used to develop these Beyesian hierarchical models. Section 2.3 describes how the detailed model strucutre was determined, including the choice of key predictors for the spatial variability (i.e., catchment characteristics $S_{1,j}$ to $S_{m,j}$) and temporal variability (i.e. $T_{1,ij}$ to $T_{n,ij}$ and $S_{TN1,j}$ and $S_{TN2,j}$), and all their corresponding coefficient values. The approaches to evaluate model performance and robustness are described in Sect. 2.4.

2.2 Data collection and processing

The Bayesian hierarchical models were developed with 21-year stream water quality observations at 102 catchments in state of Victoria, Australia. The collection and processing of the data are detailed in previous publications that worked with the same dataset (Lintern et al., 2018b; Guo et al., 2019). Briefly,

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

however, stream water quality data were extracted from the Victorian Water Measurement Information System (Department of Environment Land Water and Planning Victoria, 2016b), which contains monthly grab samples of water quality at approximately 400 sites across Victoria. Water quality data sampled between 1994 and 2014 at 102 sites were used to develop the model (Fig. 1). This was because these sites and this time period 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 Geofabric tool (Bureau of Meteorology, 2012), and have areas ranging from 5 km^2 to 16,000 km^2. 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 data were sampled following standard DELWP protocols (Australian Water Technologies, 1999) and analysed in National Association of Testing Authorities accredited laboratories .

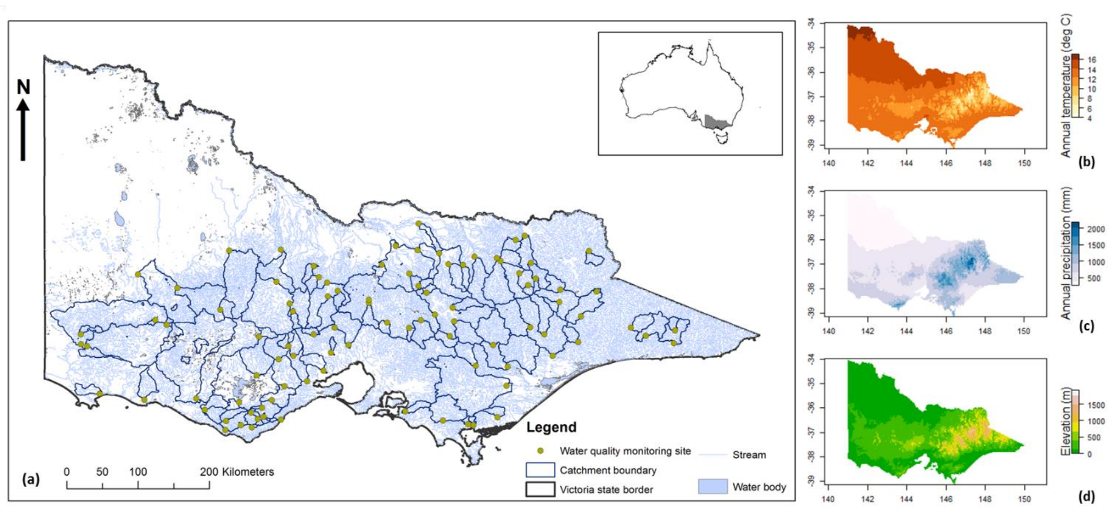

Figure 1. Map of (a) the 102 selected water quality monitoring sites and their catchment boundaries, with insert showing the location of the state of Victoria within Australia; (b) annual average temperature and (c) annual precipitation and (d) elevation across Victoria.

We selected potential spatial explanatory variables (i.e. predictors to explain spatial variability) based on catchment characterisitics that are widely known to influence water quality condition (Lintern et al., 2018a). Fifty potential explanatory catchment characteristics were selected based on a literature review.

These 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), Department of Environment Land Water and Planning Victoria (2016) and the Terrestrial Ecosystem Research Network (2016) (see Table S1 in the Supplementary Material for detailed variable names and data sources). We used a static set of land use data from 2005-2006 to represent the entire study period, as a preliminary analysis of land use data between 1996 and 2011 suggested less than 1% changes (i.e. agricultural, grazing, conservation).

Temporal explanatory variables of discharge (originally in ML d^{-1}) and water temperature (°C) corresponding to the same timestamps for water quality observations were also extracted for each site over the study period (Department of Environment Land Water and Planning Victoria, 2016). Discharge was converted to streamflow (mm d^{-1}) for each catchment, which allowed us to also calculate the average streamflows over 1, 3, 7, 14 and 30 days preceding the water quality sampling dates. In addition, gridded climate data (Frost et al., 2016;Raupach et al., 2009, 2012) and the normalized difference vegetation index (NDVI) data (NASA LP DAAC, 2017;Eidenshink, 1992) were also extracted to calculate the catchment average daily rainfall (mm), daily evapotranspiration (ET) (mm), daily average temperature (°C), daily root zone (shallower than 1m) and deep (deeper than 1m) soil moisture, as well as monthly NDVI. A summary of these data and their sources is in Table S2 in the Supplementary Material.

The raw input data were filtered and transformed to increase the data symmetry, making them more suitable for use in the linear spatio-temporal model structure (Eqs. 3, 4 and 6). For the filtering process, we first removed all water quality records with flags of quality issues and values below the limits of reporting (LOR). This was because that uncertainty of values below LOR may amplify after the transformation, posing large influence in the subsequent model fitting. Furthermore, those low concentrations were of less interest; poor water quality conditions (i.e., high constituent concentrations) were our primarily concerns to model. Water quality records corresponding to days with zero flows were also excluded from further analyses.

For the transformation process, we transformed the data of each of the spatial catchment characteristics,

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

temporal explanatory variables, as well aseach constituent to improve the symmetry of individual distributions. The log-sinh transformation (Wang et al., 2012) was used for all catchment characteristics due to the presence of zero values in several characteristics (e.g., percentage area of different types of land use). The best log-sinh transformation parameter was determined for each spatial explanatory variable across all 102 catchments. In addition, all observed constituent concentrations and temporal explanatory variables were Box-Cox transformed. For each variable, i.e., 21-year time-series data across all 102 sites, we first identified the optimal Box-Cox parameter at each site λ, and then the averaged λ across all sites to determine the final λ used to transform a respective variable. This ensured a consistent transformation for each variable across all sites. All log-sinh and Box-Cox transformation parameters used are summarized in Table S3 and S4 in the Supplementary Material.

2.3 Model fitting

Based on the general spatio-temporal modelling structure (Eqs. 2 to 6), we identified the best spatial predictors (S_1 to S_m in Eq. 3) and the best temporal predictors (T_1 to T_n in Eq. 4) in two sequential studies (Lintern et al., 2018b; Guo et al., 2019). The predictors were selected using an exhaustive search approach (May et al., 2011;Saft et al., 2016), which considered a large number of potential predictors and all possible combinations of these predictors. This selection approach required firstly fitting an individual model to each candidate predictor set, and then comparing all fitted models to select a single best set of predictors. Alternative models were evaluated based on the Akaike Information Criterion (AIC) (Akaike, 1974) and Bayesian Information Criterion (BIC) (Schwarz, 1978) to ensure optimal balance between model performance and complexity.

The key factors identified for the spatial and temporal variabilities in each constituent are listed in Tables S5 and S6 in the Supplementary Materials. General 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). We further modelled the spatial variation in each of these temporal relationships (β_T_1 to β_T_n in Eq. 4) with two spatial characteristics, S_{TN1} and S_{TN2} (Eq. 6), where a higher number of predictors was not used to avoid over-fitting. We found S_{TN1} and S_{TN2}

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Open Access
Discussions

via a Spearman correlation analyses ($p<0.05$) between the fitted parameter values of each temporal predictor variable (β_T_1 to β_T_n) and potential spatial explanatory variables as mentioned in Sect. 2.2. S_{TN1} and S_{TN2} were selected as the two catchment characteristics which had the highest correlations with the fitted parameter values of each temporal predictor, which were also summarized in Table S6 in the Supplementary Material.

After identifying the predictors for each constituent, the Bayesian hierarchical spatio-temporal model was fitted for each constituent across all monitoring sites. To achieve this, we used the R package *rstan* (Stan Development Team, 2018), which enabled both the sampling of parameter values from posterior distributions with Markov chain Monte Carlo (MCMC) and model evaluation. Constituent standard deviation (σ) was assumed to be drawn from a prior of minimally informative distribution of half-normal $N(0,10)$ that was truncated to only positive values (Gelman, 2006; Stan Development Team, 2018). The regression coefficient of each spatial predictor ($\beta\_S_1, \beta\_S_2, \ldots, \beta\_S_m$ in Eq. 3) was assumed to be drawn from an independent hyper-parameter normal distribution with mean of β_S and standard deviation of σ_S. The site-level regression coefficients of the temporal predictors ($\beta\_T_{1,j}, \beta\_T_{2,j}, \ldots, \beta\_T_{n,j}$ in Eq. 4, respectively) were sampled from the corresponding hyper-parameter normal distribution with means of $\mu.\beta\_T_1, \mu.\beta\_T_2, \ldots, \mu.\beta\_T_n$ and standard deviations of $\sigma.\beta\_T_1, \sigma.\beta\_T_2, \ldots, \sigma.\beta\_T_n$. The hyper-parameters were further assumed to be drawn from minimally informative normal distributions with $N(0,5)$ (for all the means) and minimally informative half-normal distribution of $N(0,10)$ that was truncated to only positive values (for all the standard deviations). 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 checked using the *Rhat* value (Sturtz et al., 2005).

2.4 Model performance and sensitivity analyses

The performance of the fitted model for each constituent was first evaluated by comparing the simulated and observed concentrations at 102 sites altogether to understand how the full spatio-temporal variabilties were captured (Sect. 3.1). As explained in Sect. 2.2, the model calibration for each constituent was performed with only the above-LOR data. Therefore, model performance was first evaluated with only these above-LOR data. Performance was then evaluated with the full dataset including the below-LOR data, to understand the model capacity to simulate the full distribution of

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

constituent concentration. In addition, the model performance for capturing spatial differences was assessed by comparing the simulated and observed long-term mean concentration at each site. The performance assessments were based on both visual inspection of model fitting as well as the Nash-Sutcliffe efficiency (NSE), which suggested the proportion of variability that can be explained by the models (Nash and Sutcliffe, 1970).

Additional evaluations of model sensitivity were conducted with calibration and validation on subsets of the full data (Sect. 3.2). Firstly, to understand the sensitivity of the model to the monitoring sites included for calibration, we randomly selected 80% of the sites for calibration and used the remaining 20% for validation, and repeated this validation process for five times for each constituent. The calibration and validation performance was compared to each other, as well as with the performance of the full model.

We also evaluated the model sensitivity to the periods of calibration. Since the study region was greatly influenced by a prolonged drought from 1997 to 2009 - known as the Millennium Drought, we focused on analysing the impact of this drought period. Specifically, we calibrated the model for each constituent to pre-, during- and post-drought periods (1994-1996, 1997-2009 and 2010-2014, respectively) and then validated the model on the remaining period which was not used for calibration. For example, when calibrating to the pre-drought period (1994-1996), validation was performed on both the during and post-drought data (1997-2014). Each corresponding calibration and validation performance was compared with each other as well as against that of the full model, to identify potential impacts of the drought on model robustness.

3. Results

3.1 Model performance

The spatio-temporal water quality models show varying performances among the constituents. When assessed with only the above-LOR data (Fig. 2), the best performing models are those for EC and TKN, which capture 90.7% and 65.8% of the total observed spatio-temporal variability. The modelling power is lowest for FRP (NSE = -1.92), which might be related to the large number of FRP records below the LOR (38%). Similar to FRP, poorer model performance is also observed for NO_x and TSS, with NSE

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Open Access
Discussions

values of 0.216 and 0.225, where the proportion of below-LOR samples were 17.3% and 15%, respectively. When evaluated against the entire dataset (i.e., including both below- and above LOR data), the models explain 19.9% (FRP) to 88.6% (EC) of spatio-temporal variability (Table 1). Model performances for FRP, NOx and TSS improve notably compared with previous evaluation on above-LOR data. However, FRP, NO_x and TSS remain as the three constituents that are most difficult to predict. We further discuss the possible factors influencing their model performance in Sect. 4.2.

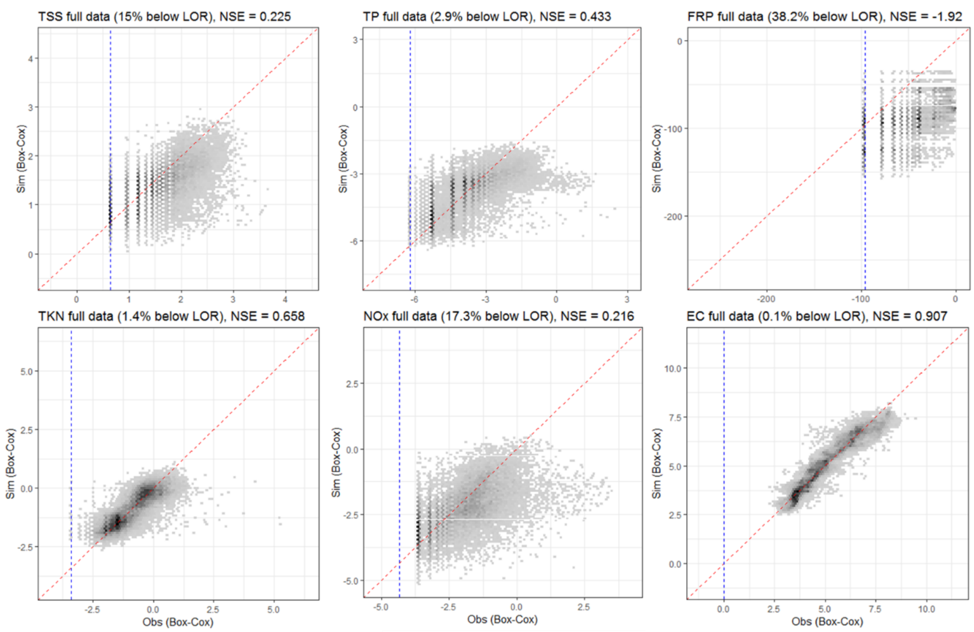

Figure 2. Performance of the spatio-temporal models for each of the six constituents, represented by the simulated and observed concentrations of above-LOR records across all 102 calibration sites, in Box-Cox transformed space. Darker regions represent denser distribution of simulation and observation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

Table 1. Comparison of model performance for all records and only the above-LOR records for each constituent.

When simulating the site-level mean concentrations, the spatio-temporal models generally show good abilities to capture variability across sites for all constituents (Fig. 3). The highest model performance is for EC (explaining 94.7% of spatial variability) and lowest performance is for FRP (explaining 44.2% spatial variability). The relative abilities of models to represent spatial variability in different

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions
Open Access

constituents are generally consistent with their capacities to capture spatio-temporal variability (Table 1).

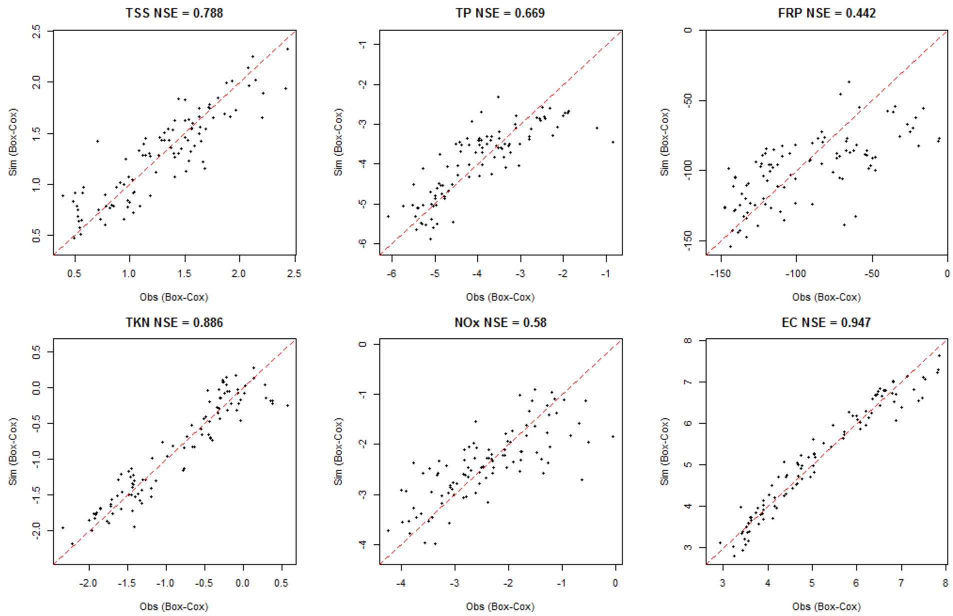

Figure 3. Model fit for site-level mean concentration at the 102 calibration sites, for the selected six constituents, in Box-Cox transformed space. The NSE for each constituent is also show and dashed red lines show the 1:1 lines.

3.2 Model sensitivity to calibration sites and periods

This section presents model sensitivity to different calibration sites and periods of record (as detailed in Sect. 2.4). Note that in these evaluations, the FRP model is not a focus due to the poor model performance observed in Sect. 3.1.

We first compare the performance of each spatio-temporal model fitted with the full dataset with those obtained from the five corresponding "partial" models that were calibrated to only 80% of the monitoring sites. Across constituents, the calibration performances obtained from the full dataset are comparable with the five models calibrated with 80% of the sites (calibration dataset). In addition, each pair of calibration and validation performance is highly consistent. In either comparison, the corresponding differences in NSE are within 0.1 (Table 2, see Figs. S1 to S6 in the Supplementary Material for detailed fitting plots for the partial calibration/validation). These suggest that the spatio-temporal model performance are highly robust and remain unaffected by the choice of calibration sites.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences Discussions
Open Access
EGU

Table 2. Comparison of model performances (as NSE) of the full model (Column 2) and the five partial models (Columns 3 to 7) with each calibrated to 80% randomly selected monitoring sites. In Columns 3 to 7, the top row showing calibration performance and the bottom row showing the validation performance (i.e. at the 20% sites that were not used for calibration).

The performance of the full model for each constituent is also compared with that of the three models calibrated to the pre-, during and post-drought periods. In general, we observe consistent performance for each constituent, across calibrations to the three periods of contrasting hydrological conditions (Table 3, see Figs. S7 to S12 in the Supplementary Material for detailed model fittings). One notable common pattern is that the performance for calibration and validation is more consistent for the drought period than either the pre- and post-drought periods. However, this is most likely explained by relative sizes of the calibration data sets, which are 3, 13 and 5 years for the pre-, during and post-drought periods respectively. Of all constituents (excluding FRP), TSS shows greater differences in model performances across periods – especially when comparing the pre-drought calibration with its validation. Fig. 4 shows the corresponding TSS model fit as represented by the site-level mean concentrations for the three calibration/validation datasets. Notably, when calibrated to the pre-drought period and validated on both the during- and post-drought periods, the model over-estimates a majority of the data (Fig. 4 (b)); and when calibrated to the during-drought period, it slightly under-estimates pre- and post-drought period TSS (Fig. 4 (d)).

Table 3. Comparison of model performances (as NSE) of the full model and the three models that were calibrated to the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods. For each of the models, the calibration performance is shown on the top row and the validation performance (i.e. over the periods that were not used for calibration) is shown on the bottom row.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

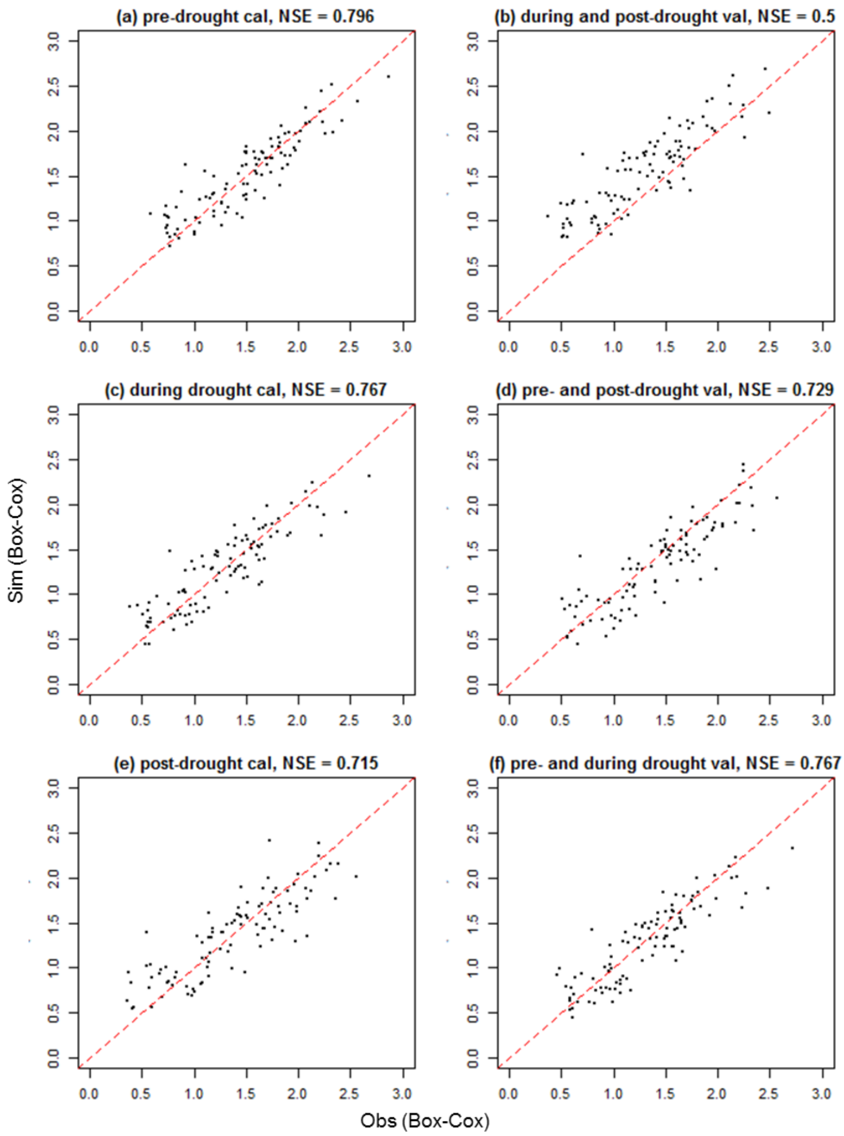

Figure 4. Comparison of the TSS model performance, as the simulated against observed site-level mean concentrations in Box-Cox transformed space. The left column shows calibration performance for the model calibrated to the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods, respectively; the right column shows the corresponding validation performance for each period. See Sect. 2.4 for details of the calibration and validation approach.

The potential impacts of drought on TSS dynamics are further illustrated with the performance of the full spatio-temporal model over the pre-, during and post-drought periods (Fig. 5). Both the during- and post-drought periods have consistently good performances, while the model underestimates the

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

majority of sites for the pre-drought period. This is consistent with Fig. 4 in suggesting a systematic decrease in TSS concentration since the drought. The better performance of the full model during and after drought (Fig. 5) can be a results of calibration period of the full spatio-temporal model – between 1994 and 2014 – which was dominated by the during- and post-drought periods; consequently, the full spatio-temporal model can be largely defined by observed TSS dynamics during and after the drought.

In summary, Figs. 4 and 5 suggest that since the drought, TSS concentrations experienced a large-scale downward shift compared to the pre-drought period, under otherwise identical spatial and temporal conditions. Such a shift indicates changes in the relationships between TSS and its key spatial and temporal controls since the start of the drought. Some possible causes are further discussed in Sect. 4.3.

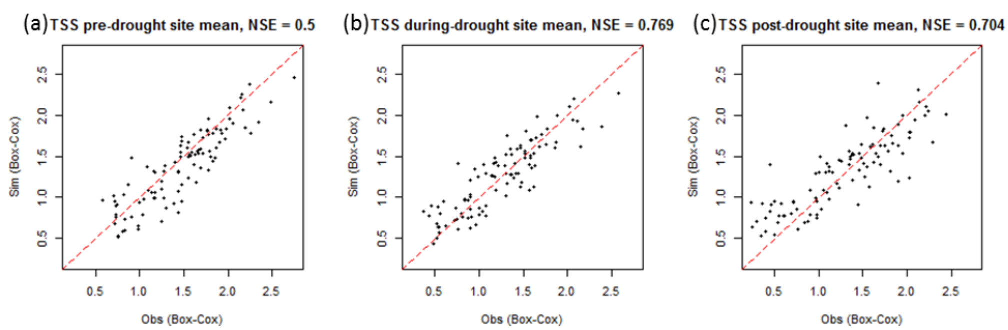

Figure 5. Comparison of the performance of the full spatio-temporal TSS model across a) pre-drought (1994-1996), b) during drought (1997-2009) and c) post-drought (2010-2014) periods, as represented by the simulated against observed site-level mean concentrations in Box-Cox transformed space.

4. Discussion

4.1 Implications for statistical water quality modelling

Our spatial-temporal water quality models are able to capture the majority of observed variability across the 102 sampling locations in Victoria (Sect. 3.1); the model performances also allow us to explore some limitations of the modelling framework. The greatest limiting factor for model performance seems to be when high proportions of LOR data are present. As shown in Fig. 2 and Table 1, model performance is best for TKN and EC, where proportions of below-LOR records are low. For constituents where the LOR records occupy greater proportions of the entire dataset, we observe poorer model performances (e.g. FRP). As illustrated in Fig. 2, the FRP calibration data have a clear left-truncation pattern resulted from removing a large proportion of below-LOR data. This substantially increased the degrees of skewness and discontinuity of the data, essentially violating the assumption of linear modelling of continuous data, and thus limiting the performance of the spatio-temporal model. It is worth noting that in this study, since we modelled spatial and temporal variabilities in an integrated manner, the model may compensate representation of the individual components of spatial and temporal variability to improve fitting to the overall variability during calibration. Consequently, in this spatio-temporal modelling framework, large presence of below-LOR data can limit the accurate representation of both variability components.

Figure 2 highlights another possible influence on model performance, which is a combination of our inability to analyse low concentrations and the limited resolution of these low-concentration measurements due to heavy transformation in data processing. This is evidenced by visually inspecting the fittings which show distinct "categorical" behaviour for low concentrations for some constituents. This "categorical" issue impacts the six constituents to different extents, ranking from the strongest as: FRP, TSS, TP, NO_x, TKN and EC – a ranking that is broadly aligned with the degree of lacking model performance for these constituents. Similar to the below-LOR records, when these categorical values are present in large proportions of the full records (e.g. TSS and FRP), they can also violate the linear model assumptions and cause performance deterioration. This issue could be overcome by alternative model structures that explicitly account for truncated data. For example, Wang and Robertson (2011) and Zhao et al. (2016) illustrated an approach to resolving the discontinuity of the likelihood estimation in modelling fitting to data with presence of zero values, which can be potentially extended to improve fitting for the categorical levels at low concentrations.

In addition, our current models are empirical relationships which are likely unable to represent complex biogeochemical processes. For example, performances for FRP and NO_x might be limited because: 1) the linear model structure can over-simplify constituent dynamics due to biogeochemical processes that are often highly non-linear; 2) the model may not include parameters that can adequately represent relevant biogeochemical processes (due to the lack of these data). To better capture changes in reactive constituents, greater consideration of and data representing biogeochemical processes may be required to address nutrient cycling including denitrification, ammonification and mineralisation (Granger et al., 2010). Therefore, possible ways to improve the statistical modelling of non-conservative constituents are: 1) alternative non-linear statistical model structures; or 2) inclusion of parameters to better represent biogeochemical processes.

Lastly, it is worth noting that our results are presented in the transformed scale for which the spatio-temporal models were developed and the statistical assumptions hold (see details on data transformation in Sect. 2.2). Model performance is heavily affected when simulated model outputs are back-transformed to the measurement scale (see Figs. S13 in the Supplementary Information). However, our models are very useful in representing and predicting proportional changes in concentrations, which adds important information for assessing and managing catchment water quality. For example, an increase of 1 mg L^{-1} in suspended solids would be alarming in pristine streams and/or periods of good water quality, while having much less impact on highly polluted conditions. The transformed models developed in this study can help managers to understand these proportional changes to identify critical locations and periods of key water quality concerns.

4.2 Implications for water quality monitoring programs

Within the current spatio-temporal models, water quality temporal variability is based on monthly monitoring data. This suggests potential oppourtunities to further strengthen the model capacity to explain temporal variability by utilizing data with higher temporal resolution. This approach can be supported by recent developments that significantly improved the accessibility of high frequency water quality monitoring data (Bende-Michl and Hairsine, 2010;Outram et al., 2014;Lannergård et al., 2019;Pellerin et al., 2016). Another potential development is to use remote sensing data to augment low frequency sampled data with higher frequency remotely sensed estimates e.g. for sediments and nutrients (Glasgow et al., 2004;Ritchie et al., 2003). Alternatively, where high frequency data are lacking for the target constituent, high frequency proxy data could also be utilized to enhance the understanding obtained from low frequency samples. For example, turbidity can be used as surrogate for sediments and nutrients (Schilling et al., 2017;Robertson et al., 2018;Lannergård et al., 2019). Currently, continuous turbidity data are available from Australia state agencies, such as the Victorian Water Quality Monitoring Network database (Department of Environment Land Water and Planning Victoria, 2016) and the NSW Water information database (WaterNSW, 2018), and collated at national level in the

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences Discussions

Bureau of Meteorology's Water Data Online portal (Bureau of Meteorology, 2019). These datasets may have great potential to enhance the temporal resolutions of records for other key water quality constituents (e.g. nutrients and sediments).

Changes in land management over time (e.g. tillage, fertiliser application, irrigation) are currently not considered as predictors of water quality temporal variability. This is due to a lack of availability and/or inconsistency of available data. However, changes in land use management practices can occur over short time periods, which can lead to increases in pollutant sources and changes to runoff generation processes (e.g. Tang et al., 2005;DeFries and Eshleman, 2004;Smith et al., 2013). Therefore, model performance can potentially be further improved by increased capacities in the monitoring of temporal patterns of land management.

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

Results of model calibration and validation to different time periods suggest a systematic decrease in TSS concentrations since the prolonged drought, in comparison with the pre-drought period under the same spatial and temporal conditions. Such a shift is not observed for any other five constituents analysed (nutrients and salts) (Sect. 3.2).

In the literature, impacts of the Millennium Drought on the hydrology and runoff regimes of south-eastern Australia are well understood (van Dijk et al., 2013;Leblanc et al., 2012;Saft et al., 2015). However, less is known about how this significant 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 the reduced streamflow during drought. For examples, reduction in sediment levels have been reported during drought, due 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 droguht have also been observed in streams adjscent to land with high densities of livestock and bushland, which both constantly contribute to sediment load during drought, leading to elevated concentrations due to lower dilution rate (Caruso, 2002). Similarly to sediments, the impact of droughts on stream nutrient and salt concentrations were also commonly understood as responses to reduced runoff generation and

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences Discussions

streamflow. Nutrient concentrations typically decrease during droughts in catchments with no significant point-source pollution (Mosley, 2015), as nutrient leaching and overland flow are reduced (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). For salinity, concentration often increases during drought with reduced dilution and increased evaporation (Caruso, 2002), particularly for catchments that are more influenced by saline groundwater input where drought can increase the relative contribution of saline groundwater input (Costelloe et al., 2005).

However, our findings highlight other possible pathways on how drought can affect stream water quality. The results suggest that the prolonged drought induced changes in sediment dynamics i.e. changes of relationships between sediments and its predictors (Figs. 4 and 5). In contrast to sediments, our models for nutrients and salts maintain consistent performance for different drought and non-drought periods, suggesting no clear shifts in dynamics. A few studies have also reported changes in the concentration-discharge relationships for sediments and nutrients. Specifically, these relationships changed from high- to low-flow conditions (Zhang, 2018;Moatar et al., 2017), as well as from drought to the recovery period (Burt et al., 2015). However, effects of extended multi-year droughts on the concentration-discharge relationships are less explored. Furthermore, there is also a lack of comprehensive assessments on the change of relationships between water quality and other relevant controls (e.g. water temperature, land cover etc.) during extended drought over large geographical regions. Our findings highlight great oppourtunities to use this dataset to further investigate the impacts of prolonged droughts on water quality dynamics, especially the changes in relationships between TSS and each of its key controls across multiple catchments.

In addition, we acknowledge that our ability to represent the pre- and post-drought conditions in this study may be limited by the record length, since 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 be also able to conduct more sophisticated investigations, such as comparing the impacts of long-term droughts versus individual dry and wet years. Addressing these research questions are particularly important in a changing climate that will be characterized by lower streamflows and possibly a shift towards more intermittent flows in many parts of the world (Saft et al., 2015;Chiew et al., 2014;Ukkola et al., 2015).

5. Conclusions

Using long-term stream water quality data collected from 102 sites in south-eastern Australia, we developed a Bayesian hierarchical statistical model to analyse the spatio-temporal variabilities in six key water quality constituents: TSS, TP, FRP, TKN, NO_x and EC. The spatio-temporal models are capable of predicting future water quality changes in non-monitored locations under similar conditions to the historical period we investigated. A notable shift in TSS dynamics is observed since the extended drought in the study region, which highlights potential oppourtunities for further research to better understand the impact of this significant drought event on water quality.

Despite the promising performance of these models, the results also illustrate areas of further improvement, both in the modelling framework but also in the monitoring of water quality. In improving the modelling framework, alternative statistical approaches could be considered to reduce the impact of below detection limit and low concentration data on model performance. In addition, the models could be extended to take into account some key biogeochemical processes to better represent spatial-temporal variability in non-conservative constituents (e.g., FRP or NO_x). To further enhance the performance of the current models, we recommend that future water quality monitoring programs be enhanced with: 1) collection and assimilation of high-frequency sampling data to enhance the temporal resolution of water quality data; and 2) more frequent monitoring of changes in land use intensity and management to be able to include these parameters in the model. These improvements will be very helpful to operational catchment management and mitigation.

Data availability

All data used in this study were extracted from public domain. All stream water quality data were extracted from the Victorian Water Measurement Information System (via http://data.water.vic.gov.au/, provided by the Department of Environment Land Water and Planning Victoria). The catchments corresponding to these water quality monitoring sites were delineated using the Geofabric tool provided by the Bureau of Meteorology, via ftp://ftp.bom.gov.au/anon/home/geofabric/. We have listed the sources of all other data for the spatial and temporal predictors of our models in Tables S1 and S2 in the Supplementary Materials.

Author contribution

All authors contributed to the conceptualization the models and the design of methodology. A. Lintern and S. Liu contributed to the data curatiom. D.Guo carried out the formal analyses, visualization and validation. J.A. Webb, D. Ryu, U. Bende-Michl and A.W. Western contributed to the funding acquisition. D. Guo, A. Lintern, J.A. Webb, D. Ryu, S. Liu and A.W. Western contributed to the investigation. D. Guo carried out project administration and coding to run the experiments. J.A. Webb, D. Ryu, and A.W. Western contributed to the supervision. D.Guo prepared the manuscript with contributions from all co-authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgement

Australian Research Council and the Victorian Environment Protection Authority, the Victorian Department of Environment, Land Water and Planning, the Bureau of Meteorology and the Queensland Department of Natural Resources, Mines and Energy provided funding for this project through the linkage program (LP140100495). 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: Water Measurement Information System (http://data.water.vic.gov.au/monitoring.htm). Sources of other data are provided in Tables S1 and S2 of the Supplementary Materials.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions
Open Access

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, Journal of Hydrology, 524, 733-752, https://doi.org/10.1016/j.jhydrol.2015.03.027, 2015.

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, Journal of Hydrology, 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 transactions on automatic control, 19, 716-723, 1974.

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.

Bende-Michl, U., and Hairsine, P. B.: A systematic approach to choosing an automated nutrient analyser for river monitoring, Journal of Environmental Monitoring, 12, 127-134, 2010.

Bengraïne, K., and Marhaba, T. F.: Using principal component analysis to monitor spatial and temporal changes in water quality, Journal of Hazardous Materials, 100, 179-195, https://doi.org/10.1016/S0304-3894(03)00104-3, 2003.

Bond, N. R., Lake, P. S., and Arthington, A. H.: The impacts of drought on freshwater ecosystems: an Australian perspective, Hydrobiologia, 600, 3-16, 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, Ecological Modelling, 143, 165-181, https://doi.org/10.1016/S0304-3800(01)00328-3, 2001.

Bureau of Meteorology: Geofabric V2 2012.

Bureau of Meteorology: Water Data Online, 2019.

Bureau of Rural Sciences: 2005/06 Land use of Australia, version 4. . 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, Hydrological Processes, 29, 498-507, 10.1002/hyp.10169, 2015.

Caruso, B. S.: Temporal and spatial patterns of extreme low flows and effects on stream ecosystems in Otago, New Zealand, Journal of Hydrology, 257, 115-133, 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 Research, 42, 3285-3304, https://doi.org/10.1016/j.watres.2008.04.006, 2008.

Chiew, F. H. S., Potter, N. J., Vaze, J., Petheram, C., Zhang, L., Teng, J., and Post, D. A.: Observed hydrologic non-stationarity in far south-eastern Australia: implications for modelling and prediction, Stochastic Environmental Research and Risk Assessment, 28, 3-15, 10.1007/s00477-013-0755-5, 2014.

Clark, J. S.: Why environmental scientists are becoming Bayesians, Ecology Letters, 8, 2-14, doi: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, Hydrological Processes, 19, 3147-3166, 10.1002/hyp.5837, 2005.

DeFries, R., and Eshleman, K. N.: Land-use change and hydrologic processes: a major focus for the future, Hydrological Processes, 18, 2183-2186, doi:10.1002/hyp.5584, 2004.

Department of Environment Land Water and Planning Victoria: Victorian water measurement information system. . 2016.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

Eidenshink, J. C.: The 1990 Conterminous U.S. AVHRR Data Set, Photogrammetric Engineering and Remote Sensing, 58, 1992.

Frost, A. J., Ramchurn, A., and Smith, A.: The bureau's operational AWRA landscape (AWRA-L) Model, 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, Environmental Modelling & Software, https://doi.org/10.1016/j.envsoft.2018.12.008, 2018.

Geoscience Australia: Dams and water storages. . 2004.

Geoscience Australia: Environmental Attributes Dataset. 2011.

Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., and Kleinman, J. E.: Real-time remote monitoring of water quality: a review of current applications, and advancements in sensor, telemetry, and computing technologies, Journal of Experimental Marine Biology and Ecology, 300, 409-448, https://doi.org/10.1016/j.jembe.2004.02.022, 2004.

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 Affecting Temporal Variability in Stream Water Quality, Water Resources Research, 55, 112-129, 10.1029/2018wr023370, 2019.

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, doi:10.1002/wat2.1155, 2016.

Kaushal, S. S., Mayer, P. M., Vidon, P. G., Smith, R. M., Pennino, M. J., Newcomer, T. A., Duan, S., Welty, C., and Belt, K. T.: Land Use and Climate Variability Amplify Carbon, Nutrient, and Contaminant Pulses: A Review with Management Implications, JAWRA Journal of the American Water Resources Association, 50, 585-614, doi:10.1111/jawr.12204, 2014.

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, Marine and Freshwater Research, 62, 255-265, https://doi.org/10.1071/MF09315, 2011.

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, Journal of Hydrology, 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 Yeşilırmak River, Turkey, Environmental Modelling & Software, 20, 1195-1200, https://doi.org/10.1016/j.envsoft.2004.11.001, 2005.

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, Science of the Total Environment, 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 and Planetary Change, 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, doi: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 Resources Research, doi:10.1029/2017WR022172, 2018b.

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.

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, 1989.

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 Resources Research, 53, 1270-1287, 10.1002/2016wr019635, 2017.

Mosley, L. M.: Drought impacts on the water quality of freshwater systems; review and integration, Earth-Science Reviews, 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 AMERICA1, JAWRA Journal of the American Water Resources Association, 36, 347-366, 10.1111/j.1752-1688.2000.tb04273.x, 2000.

NASA LP DAAC: MOD13A3: MODIS/Terra Vegetation Indices Monthly L3 Global 1km V005. 2017.

Outram, F. N., Lloyd, C. E. M., Jonczyk, J., Benskin, C. M. 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, 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, Science of The Total Environment, 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, Environmental Science and Pollution Research, 22, 397-414, 10.1007/s11356-014-3346-1, 2015.

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, JAWRA Journal of the American Water Resources Association, 52, 993-1008, 10.1111/1752-1688.12386, 2016.

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, Environmental Management, 45, 105-112, 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.

Ritchie, J. C., Zimba, P. V., and Everitt, J. H.: Remote Sensing Techniques to Assess Water Quality, Photogrammetric Engineering & Remote Sensing, 69, 695-704, 10.14358/PERS.69.6.695, 2003.

Robertson, D. M., Hubbard, L. E., Lorenz, D. L., and Sullivan, D. J.: A surrogate regression approach for computing continuous loads for the tributary nutrient and sediment monitoring program on the Great Lakes, Journal of Great Lakes Research, 44, 26-42, https://doi.org/10.1016/j.jglr.2017.10.003, 2018.

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 Resources Research, 51, 2444-2463, doi:10.1002/2014WR015348, 2015.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences Discussions

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 Resources Research, 52, 9290-9305, doi:10.1002/2016WR019525, 2016.

Schilling, K. E., Kim, S.-W., and Jones, C. S.: Use of water quality surrogates to estimate total phosphorus concentrations in Iowa rivers, Journal of Hydrology: Regional Studies, 12, 111-121, https://doi.org/10.1016/j.ejrh.2017.04.006, 2017.

Schwarz, G.: Estimating the dimension of a model, The annals of statistics, 6, 461-464, 1978.

Smith, A. P., Western, A. W., and Hannah, M. C.: Linking water quality trends with land use intensification in dairy farming catchments, Journal of Hydrology, 476, 1-12, 2013.

Stan Development Team: RStan: the R interface to Stan. R package version 2.18.1, 2018.

Tang, Z., Engel, B. A., Pijanowski, B. C., and Lim, K. J.: Forecasting land use change and its environmental impact at a watershed scale, Journal of Environmental Management, 76, 35-45, https://doi.org/10.1016/j.jenvman.2005.01.006, 2005.

Terrestrial Ecosystem Research Network: Soil and landscape grid of Australia. 2016.

Tramblay, Y., Ouarda, T. B. M. J., St-Hilaire, A., and Poulin, J.: Regional estimation of extreme suspended sediment concentrations using watershed characteristics, Journal of Hydrology, 380, 305-317, https://doi.org/10.1016/j.jhydrol.2009.11.006, 2010.

Ukkola, A. M., Prentice, I. C., Keenan, T. F., van Dijk, A. I. J. M., Viney, N. R., Myneni, Ranga B., and Bi, J.: Reduced streamflow in water-stressed climates consistent with CO2 effects on vegetation, Nature Climate Change, 6, 75, 10.1038/nclimate2831

https://www.nature.com/articles/nclimate2831#supplementary-information, 2015.

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 Resources Research, 49, 1040-1057, 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, Journal of Hydrology, 353, 1-17, https://doi.org/10.1016/j.jhydrol.2008.01.001, 2008.

Victorian water measurement information system: http://data.water.vic.gov.au/monitoring.htm, 2016.

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, 10.1038/nature09440

https://www.nature.com/articles/nature09440#supplementary-information, 2010.

Wang, Q. J., and Robertson, D. E.: Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences, Water Resources Research, 47, 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 Resources Research, 48, doi:10.1029/2011WR010973, 2012.

Real-time water data: https://realtimedata.waternsw.com.au/, access: 12/07/2018, 2018.

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, doi: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 blackwater event in a major river system (the southern Murray–Darling Basin, Australia), Journal of Hydrology, 450-451, 190-198, https://doi.org/10.1016/j.jhydrol.2012.04.057, 2012.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Hydrology and Earth System Sciences
Discussions

Zhang, Q.: Synthesis of nutrient and sediment export patterns in the Chesapeake Bay watershed: Complex and non-stationary concentration-discharge relationships, Science of The Total Environment, 618, 1268-1283, https://doi.org/10.1016/j.scitotenv.2017.09.221, 2018.

Zhao, T., Schepen, A., and Wang, Q. J.: Ensemble forecasting of sub-seasonal to seasonal streamflow by a Bayesian joint probability modelling approach, Journal of Hydrology, 541, 839-849, https://doi.org/10.1016/j.jhydrol.2016.07.040, 2016.

https://doi.org/10.5194/hess-2019-342
Preprint. Discussion started: 23 July 2019

Tables

Table 1. Comparison of model performance for all records and only the above-LOR records for each constituent.

Constituent	Above-LOR 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

Table 2. Comparison of model performances (as NSE) of the full model (Column 2) and the five partial models (Columns 3 to 7) with each calibrated to 80% randomly selected monitoring sites. In Columns 3 to 7, the top row showing calibration performance and the bottom row showing the validation performance (i.e. at the 20% sites that were not used for calibration).

Constituent	Full model	80% sites split 1	80% sites split 2	80% sites split 3	80% sites split 4	80% sites split 5
TSS	0.397	0.406	0.431	0.390	0.428	0.423
		0.348	0.441	0.443	0.446	0.434
TP	0.445	0.440	0.422	0.440	0.472	0.456
		0.386	0.444	0.454	0.449	0.444
FRP	0.199	0.141	0.219	0.244	0.216	0.177
		0.041	0.350	0.337	0.356	0.344
TKN	0.630	0.664	0.643	0.630	0.658	0.669
		0.639	0.589	0.581	0.584	0.587
NO_x	0.382	0.410	0.464	0.438	0.476	0.466
		0.419	0.593	0.603	0.597	0.597
EC	0.886	0.895	0.894	0.875	0.900	0.892
		0.796	0.828	0.837	0.828	0.826

Table 3. Comparison of model performances (as NSE) of the full model and the three models that were calibrated to the pre-drought (1994-1996), drought (1997-2009) and the post-drought (2010-2014) periods. For each of the models, the calibration performance is shown on the top row and the validation performance (i.e. over the periods that were not used for calibration) is shown on the bottom row.

Constituent	Full model	Pre-drought calibration	During drought calibration	Post-drought calibration
TSS	0.397	0.495	0.399	0.499
		0.208	0.402	0.390
TP	0.445	0.477	0.438	0.525
		0.421	0.474	0.411
FRP	0.199	-1.336	0.187	0.204
		-1.406	0.197	0.024
TKN	0.630	0.649	0.650	0.711
		0.566	0.648	0.610
NOx	0.382	0.443	0.426	0.509
		0.394	0.471	0.393
EC	0.886	0.854	0.901	0.901
		0.887	0.873	0.884