A Bayesian approach to understanding the key factors influencing temporal variability in stream water quality: a case study in the Great Barrier Reef catchments

Shuci Liu\(^1\)*, Dongryeol Ryu\(^1\), J. Angus Webb\(^1\), Anna Lintern\(^1,2\), Danlu Guo\(^1\), David Waters\(^3\), Andrew W. Western\(^1\)

\(^1\)Department of Infrastructure Engineering, The University of Melbourne, Parkville, VIC, 3010.
\(^2\)Department of Civil Engineering, Monash University, VIC, 3800.
\(^3\)Queensland Department of Natural Resources, Mines and Energy, Toowoomba, QLD, 4350.

Correspondence to: Shuci Liu (shucil@student.unimelb.edu.au)

Abstract. Stream water quality is highly variable both across space and time. Water quality monitoring programs have collected a large amount of data that provide a good basis to investigate the key drivers of spatial and temporal variability. Event-based water quality monitoring data in the Great Barrier Reef catchments in northern Australia provides an opportunity to further our understanding of water quality dynamics in sub-tropical and tropical regions. This study investigated nine water quality constituents, including sediments, nutrients and salinity, with the aim of: 1) identifying the influential environmental drivers of temporal variation in flow event concentrations; and 2) developing a modelling framework to predict the temporal variation in water quality at multiple sites simultaneously. This study used a hierarchical Bayesian model averaging framework to explore the relationship between event concentration and catchment-scale environmental variables (e.g., runoff, rainfall and groundcover conditions). Key factors affecting the temporal changes in water quality varied among constituent concentrations, as well as between catchments. Catchment rainfall and runoff affected in-stream particulate constituents, while catchment wetness and vegetation cover had more impact on dissolved nutrient concentration and salinity. In addition, in large dry catchments, antecedent catchment soil moisture and vegetation had a large influence on dissolved nutrients, which highlights the important effect of catchment hydrological connectivity on pollutant mobilisation and delivery.

1 Introduction

In-stream water quality plays a vital role in influencing the health of freshwater ecosystems (Bhaduri et al., 2016; Pérez-Gutiérrez et al., 2017), which in turn underpins environmental, social and economic sustainability (McGrane, 2016; Ustaoglu et al., 2020). Pollution derived from agricultural land and urban development has led to water quality degradation in streams and lakes in many regions of the world (Novotny, 1999; Peters and Meybeck, 2000; Ren et al., 2003; Sharpley, 2016). Among these water quality issues, coastal regions with high agricultural production have been delivering large amounts of pollutants to the ocean, where marine ecosystems are vulnerable to the evaluated levels of nutrients and sediments.
(Carpenter et al., 1998; Gorman et al., 2009). It is estimated that 60% of coastal rivers in the USA have been moderately to severely degraded (Gorman et al., 2009; Howarth et al., 2002). Therefore, to protect both freshwater and marine ecosystems, better management of catchment-derived pollutants is needed.

Surface water quality is highly variable across spatial and temporal scales (Allan et al., 1997; Guo et al., 2019; Lintern et al., 2018a). These spatial and temporal variations are the result of complex interactions between three key pollutant processes in catchments, namely, sources (e.g., atmospheric deposition or anthropogenic inputs), mobilisation (e.g., detachment from the sources), and delivery (e.g., transport from sources to receiving waters) (Granger et al., 2010; Lintern et al., 2018a). Across different catchments, spatial differences in water quality concentration can vary markedly due, in part, to heterogeneity of natural landscapes in catchments (e.g., geology, topography and climate) and human-induced activities (e.g., agricultural and urban development) (Liu et al., 2018; Mainali and Chang, 2018; Mainali et al., 2019). At a site, water quality concentrations can also exhibit significant daily, event, seasonal and annual variability, driven by variations in climatic conditions, in-stream biogeochemical processes and hydrological transport (Hill, 1996; Pretty et al., 2006; Thompson et al., 2011). Thus, it can be challenging to design effective catchment water quality management strategies without a sound understanding of the spatial and temporal variation in water quality and the associated driving factors.

While it has been acknowledged that both spatial and temporal variations in water quality are of great importance for effective water resources management (Guo et al., 2020), this study focused on identifying key drivers of the temporal variability in water quality. It follows our previous study investigating spatial variation in water quality in the same region (Liu et al., 2018). A wide range of environmental factors may affect temporal changes in water quality. Runoff and rainfall have been considered as important factors and the most commonly used explanatory variables to describe temporal variation in water quality (Deletic and Maksimovic, 1998; Kim et al., 2007; Yang et al., 2009), for example early work by Hem (1948), Walling and Foster (1975) and Walling (1984). Studies considering hydrometeorological drivers have been typically related to the mobilisation and delivery of pollutants. Catchment soil moisture and evapotranspiration can also have an important role in determining the hydrological cycle (e.g., runoff generation), such as sediments (Bieger et al., 2014; Varanou et al., 2002), nutrients (Bouraoui et al., 2002; Lam et al., 2010) and salinity (Brevik et al., 2006; Tweed et al., 2007), thereby affecting the surface water quality. In addition, riverine water quality has been found to be strongly influenced by seasonal changes in vegetation cover (de Mello et al., 2018; Griffith et al., 2002; Shi et al., 2017). For instance, satellite-derived vegetation indices have provided an opportunity to explore the relationship between land cover and water quality temporal dynamics (Fu and Burgher, 2015; Griffith, 2002; Singh et al., 2013; Whistler, 1996). Even though significant research efforts have been made to explore the relationship between water quality and these environmental conditions, a comprehensive understanding of their relative importance in diverse environments and at large scales is still lacking.

Statistical modelling has been widely used to investigate water quality temporal dynamics in response to changes in the abovementioned environmental factors (Alexander et al., 2002; Fu et al., 2019; Miller et al., 2014; Singh et al., 2013; Zhang and Blomquist, 2018; Zhang et al., 2016; Zhang and Schilling, 2005). However, existing studies have limitations. Firstly,
water quality monitoring data have often been limited to low sampling frequencies, typically using monthly grab samples. This can result in a lack of information on water quality dynamics over runoff/storm events, which is when a significant proportion of nutrients and sediment loads are transported (Lloyd et al., 2016; Sherriff et al., 2015). Secondly, most studies on statistical water quality modelling have only investigated the relationship between water quality and explanatory variables in a single or limited number of catchments in small regions (Chang et al., 2015; Khan et al., 2020; Koci et al., 2020; Liu et al., 2008b; Noori et al., 2020; Wang et al., 2016; Zhang et al., 2016). Few studies have investigated water quality at multiple locations using the same modelling framework. Lastly, studies have usually relied on a single ‘best’ model with an assumption that it best approximated the true drivers of water quality (Paliwal et al., 2007; Zhang et al., 2009). This ignores the issue of selection uncertainty. Furthermore, relying on a single model structure might result in misleading conclusions or overconfidence in the results (Link and Barker, 2006; Wintle et al., 2003).

This study attempted to address these knowledge gaps, taking advantages of event-based water quality monitoring data from the Great Barrier Reef (GBR) catchments in northern Australia, where land-derived pollutants have posed threats to ecosystem of the GBR lagoon (Brodie et al., 2012; Hunter and Walton, 2008; McKergow et al., 2005b; Waterhouse et al., 2017). We targeted nine common water quality indicators, including sediments, nutrients and salinity. Bayesian hierarchical modelling was used to investigate water quality spatial and temporal variation. This allowed the prediction of water quality in multiple catchments, as well as simultaneously quantifying parameter uncertainty (Gelman et al., 2013; Rode et al., 2010; Webb and King, 2009). In addition, we used Bayesian model averaging (BMA) approaches to identify the relative importance of the different environmental factors and provide multi-model weighted predictions, which have been shown to better quantify the uncertainty arising from model selection (Höge et al., 2019; Raftery et al., 1997; Wang et al., 2012). Overall, this study aimed to: (1) identify the key drivers of temporal variation in water quality; and (2) predict water quality temporal variation using a Bayesian multi-model approach.

2 Materials and methods

2.1 Study area

The GBR catchments, situated in north-eastern Australia (Figure 1), consist of six natural resource management regions whose streams and rivers discharge into the Great Barrier Reef lagoon. These catchments cover a 437,354 km², approximately a quarter of the state of Queensland, and exhibit significant diversity in climatic, geological and topographical landscape characteristics, as well as in land use and land management (Bartley et al., 2018; Gilbert and Brodie, 2001). The GBR catchments range from small, steep, high-energy streams in the wet tropics, which are dominated by sugarcane crops and rainforest, to large inland catchments used for savannah grazing, and crops (e.g., grain) and with extensive low energy floodplains in the dry tropics (Table 1) (Davis et al., 2017; Koci et al., 2019; McKergow et al., 2005a). Spatial and temporal variations in rainfall in the GBR catchments are a major cause of the diversity in land use patterns. Annual rainfall ranges from less than 500 mm in the south-west to more than 8000 mm in the north-east (Figure 2 [c]) (Davis et al., 2017; Kuhnert
et al., 2009). Distinct wet (November to April) and dry (May to October) seasons result in high seasonal variation in runoff and El Nino-Southern Oscillation (ENSO) leads to high inter-annual variability (Day and McKeon, 2018; Murphy and Ribbe, 2004). In the dry tropics, a few large events in the wet season contribute the majority of annual runoff, and constant low flow dominates during the dry season (Jarihani et al., 2017).

Thirty-two sites within the GBR catchments were selected as case study catchments (Figure 1 and Table C1 in Appendix C). Previous multivariate analysis of the patterns of time-averaged concentrations indicated that there were two groups of sites (Table 1 and Figure 2 [a]), which was a result of spatial heterogeneity in catchment landscape characteristics (Figure 2 [b], [c] and [d]) (Liu et al., 2018).

Figure 1. The Great Barrier Reef catchments, monitoring sites, land uses and the six natural resource management (NRM) regions. Land uses have the following characteristics: (1) conservation (forest, woodland, savannah, etc for conservation
purposes); (2) dryland (rainfed agriculture including cereals but excluding grazing and sugar cane); (3) grazing (primarily cattle grazing of native and introduced vegetation); (4) intensive (urban areas, roads, etc); (5) irrigated (irrigated cropping excluding sugar cane); (6) sugar (rain-fed and irrigated sugar cane); and (7) water (water bodies, including lake, river, and marsh/wetland).

Figure 2: Spatial information of the GBR catchments in northeast of Australia: [a] site locations showing two groups based on clustering analysis of spatial variability in time-averaged water quality (Liu et al., 2018); [b] topographic elevation (250 m resolution) (Geoscience Australia, 2008); [c] annual average rainfall (Bureau of Meteorology, 2012), and [d] updated Köppen-Geiger climate zone classification (Peel et al., 2007).
Table 1: Summary of differences in landscape characteristics between the two clusters of sites (Liu et al., 2018).

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Climate</th>
<th>Hydrology</th>
<th>Land use/land cover</th>
<th>Topography</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wet tropics region with high annual rainfall</td>
<td>Perennial, high energy rivers</td>
<td>Dominated by conservation (e.g. rainforest), and cropping (e.g. sugar)</td>
<td>Small and steep</td>
</tr>
<tr>
<td>2</td>
<td>Mostly dry tropics, relatively dry with clear seasonal variability in rainfall</td>
<td>Ephemeral, low energy rivers, cease-to-flow in dry period</td>
<td>Dominated by brigalow native vegetation, and pastures for grazing,</td>
<td>Large and flat</td>
</tr>
</tbody>
</table>

2.2 Data collection and preparation

2.2.1 Water quality data

The nine studied constituents were total suspended solids (TSS), particulate nitrogen (PN), oxidized nitrogen (NOx), ammonium nitrogen (NH4), dissolved organic nitrogen (DON), filterable reactive phosphorus (FRP), dissolved organic phosphorus (DOP), particulate phosphorus (PP), and electrical conductivity (EC). Water quality monitoring data collected for the 32 GBR catchments over the 11-year period of 2006 to 2016 were obtained from the Loads Monitoring Program (Turner et al., 2012). This dataset contained both high-frequency event-based samples (e.g., daily or every few hours by automatic samplers) that were taken during runoff events, as well as grab samples (e.g., monthly) that were taken under baseflow conditions (Orr et al., 2014; Waters et al., 2013; Waters and Packett, 2007). As EC data from the Loads Monitoring Program were limited, we extracted additional EC data from the Water Monitoring Information Portal provided by the Department of Natural Resources, Mines and Energy of Queensland (DNRME, 2018) to complement the Loads Monitoring Program records.

2.2.2 Event mean concentration

We extracted continuous discharge records for each site from the Water Monitoring Information Portal (DNRME, 2018) to identify individual runoff events. An automated hydrograph analysis tool – HydRun (Tang and Carey, 2017) was used to delineate runoff events. The start and end points of a specific event were determined by using a local-minimum method that calculates the first derivative of the streamflow record (separated from baseflow). The event-mean concentration (EMC) was then calculated for each event that had at least two samples on each of the rising and falling limbs of the hydrograph. This ensured that the water quality dynamics over a runoff event were reasonably well-captured, and that the derived EMCs were reliable (Waters and Packett, 2007). For each event, the EMC of a constituent was calculated as the total load per unit flow volume within the event using (Bartley et al., 2012; Joo et al., 2012):
EMC = \frac{\text{EventLoad}}{\text{EventFlowVolume}} = \frac{\sum_{j=0}^{n} \frac{c_j + c_{j+1}}{2} \times q_{j+1/2} \times t_{j+1/2}}{\sum_{j=0}^{n} q_{j+1/2} \times t_{j+1/2}}

(1)

where \( n \) is the total number of samples for a given event, \( c_j \) is concentration of the \( j^{th} \) sample, \( q_{j+1/2} \) and \( t_{j+1/2} \) are the inter-sample mean discharge and time interval between \( j^{th} \) and \( (j+1)^{th} \) samples. The concentrations at the start and end of the event \((c_0 \text{ and } c_{n+1})\) are assumed to be the averaged value for samples during baseflow (with baseflow identified in the previous section). The EMCs were essentially flow-weighted mean concentrations over individual runoff events, which allowed the comparison of water quality across catchments with contrasting flow regimes (e.g., two clusters of sites in Figure 2) (Cooke et al., 2000; Richards and Baker, 1993). A total of 1412 events was identified across the 32 sites, and, depending on data availability, EMCs were calculated for between 21% (DOP) and 43% (TSS) of these identified runoff events (Table C2).

The derived EMCs were Box-Cox transformed to improve the symmetry of the response variable (Box and Cox, 1964) to improve model fitting (Hawkins and Weisberg, 2017; Lawrance, 1988; Zhang and Yang, 2017). The site-level Box-Cox transformation parameter \( \lambda \) for each constituent was first identified, using the \textit{car} package in R (Fox et al., 2012; R Core Team, 2013). Then, for each constituent, the average \( \lambda \) from the 32 sites was used to transform all available EMCs for that specific constituent. This ensured that an identical transformation parameter was applied across the different sites for each constituent (Guo et al., 2019).

### 2.2.3 Explanatory variables

This study investigated the effect of various hydrologic, climatic and vegetation cover characteristics for different events. These characteristics included runoff, catchment root zone soil moisture, actual evapotranspiration rainfall, air temperature, and vegetation cover. The continuous streamflow monitoring data, gridded weather and climatic products, and remotely sensed imagery were used to derive catchment average conditions for each event (Table 2).

**Table 2. Explanatory variables and their data sources**

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Unit</th>
<th>Spatial resolution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily rainfall</td>
<td>mm</td>
<td>5 km × 5 km</td>
<td>Australia Water Availability Project (AWAP) (Raupach et al., 2009). Available from <a href="http://www.csiro.au/awap/">http://www.csiro.au/awap/</a></td>
</tr>
<tr>
<td>Daily temperature</td>
<td>°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-day normalized difference vegetation index (NDVI)</td>
<td>-</td>
<td>1 km × 1 km</td>
<td>Moderate Resolution Imaging Spectroradiometer (MODIS) - MOD13A2v006 (Didan, 2015). Available from <a href="https://earthdata.nasa.gov/">https://earthdata.nasa.gov/</a></td>
</tr>
<tr>
<td>Daily soil moisture (root zone 0-100 cm)</td>
<td>mm</td>
<td>5 km × 5 km</td>
<td>Australia Landscape Water Balance model (AWRA-L) (Frost et al., 2016). Available from <a href="http://www.bom.gov.au/water/landscape">http://www.bom.gov.au/water/landscape</a></td>
</tr>
<tr>
<td>Daily actual ET</td>
<td>Mm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For individual runoff events identified in the previous section, three groups of event characteristics were prepared, characterising pre-event, during-event and post-event conditions (Table 3). Except for runoff, data for all explanatory variables were first extracted from gridded data using catchment boundaries were delineated using the Geofabric tool provided by the Australian Bureau of Meteorology (Bureau of Meteorology, 2012) (Figure 1). The catchment average time series data were then averaged over the specific time-window related to the event (Table 3).

Table 3: Three groups of event characteristics and averaging method

<table>
<thead>
<tr>
<th>Group</th>
<th>Explanatory variable</th>
<th>Abbreviation used in figures and tables in paper</th>
<th>Calculation method</th>
</tr>
</thead>
<tbody>
<tr>
<td>During-event</td>
<td>Average runoff</td>
<td>Event_ave_Q</td>
<td>Average of daily runoff during event</td>
</tr>
<tr>
<td></td>
<td>Maximum runoff</td>
<td>Event_max_Q</td>
<td>Maximum of daily runoff during event</td>
</tr>
<tr>
<td></td>
<td>Average rainfall</td>
<td>Event_ave_P</td>
<td>Average of daily rainfall during event</td>
</tr>
<tr>
<td></td>
<td>Maximum rainfall</td>
<td>Event_max_P</td>
<td>Maximum of daily rainfall during event</td>
</tr>
<tr>
<td></td>
<td>Average temperature</td>
<td>Event_T</td>
<td>Average of daily temperature during event</td>
</tr>
<tr>
<td></td>
<td>Average NDVI</td>
<td>Event_NDVI</td>
<td>Average of NDVI during event</td>
</tr>
<tr>
<td></td>
<td>Average soil moisture</td>
<td>Event_SM</td>
<td>Average of daily soil moisture during event</td>
</tr>
<tr>
<td></td>
<td>Average actual ET</td>
<td>Event_AET</td>
<td>Average of daily actual ET during event</td>
</tr>
<tr>
<td>Pre-event</td>
<td>Average runoff</td>
<td>Ante_Q</td>
<td>Average of daily runoff for 7 days prior to event</td>
</tr>
<tr>
<td></td>
<td>Average rainfall</td>
<td>Ante_P</td>
<td>Average of daily rainfall for 7 days prior to event</td>
</tr>
<tr>
<td></td>
<td>Average NDVI</td>
<td>Ante_NDVI</td>
<td>Average of NDVI for 3 months prior to event</td>
</tr>
<tr>
<td></td>
<td>Average soil moisture</td>
<td>Ante_SM</td>
<td>Average of daily soil moisture for 7 days prior to event</td>
</tr>
<tr>
<td></td>
<td>Average actual ET</td>
<td>Ante_AET</td>
<td>Average of actual ET for 7 days prior to event</td>
</tr>
<tr>
<td>Post-event</td>
<td>Average runoff</td>
<td>Post_Q</td>
<td>Average of daily runoff for 7 days after event</td>
</tr>
</tbody>
</table>

Note: Q – runoff; P – rainfall; T – temperature; NDVI – normalized difference vegetation index; SM – root zone soil moisture; ET – evapotranspiration.

The explanatory variables in the during-event conditions were averaged over the duration of the event. For the pre-event and post-event conditions, the 7 days prior to and after the event were used as the time-window (except NDVI). The 7-day period was the median of the time of concentration (i.e., the time for runoff to travel from the most remote point of the catchment to the monitoring site) across all catchments. These were estimated from catchment topography using the Bransby-William’s equation, following its wide application in Australian catchments for flood estimation (French et al., 1974; Pilgrim et al., 1987). The ground cover was quantified by NDVI, an indicator of the biophysical condition of the vegetation canopy (Griffith et al., 2002). Previous studies have also shown that there is a time-lag between water availability and a change in ground cover, which is typically three months for Australian catchments (De Keersmaecker et al., 2015; Papagiannopoulou et al., 2017). Therefore, to represent the pre-event ground cover condition, we averaged all available NDVI measurements.
for three months prior to an event. The runoff after the event (7 days) was also included as an indicator of catchment wetness at the end of the event, to assess if hydrologic condition towards the end of an event influences the temporal variation in water quality.

Similar to the EMCs, all the explanatory variables were Box-Cox transformed following the procedure described in Sect. 2.2.2. In addition, prior to the analyses, both transformed EMCs and explanatory variables were standardized to a mean of zero and standard deviation of one. As such, the magnitude of a coefficient indicates the effect of each predictor relative to other predictors (Wan et al., 2014). The cross-correlation (non-parametric Spearman’s Rank correlation coefficient) of all transformed predictors is provided in Figure B1, Appendix B.

2.3 Modelling: driver identification and water quality prediction using multi-model inference

The statistical analysis and modelling followed several steps (Figure 3). The Bayesian modelling framework was applied to catchments in Clusters 1 and 2 separately. This is because we assumed that the key drivers of temporal variability in water quality would differ between two Clusters due to their differences in land use and climate.

2.3.1 Bayesian variable selection

To investigate the relative importance of individual predictors, an indicator Bayesian variable selection method was used called Gibbs variable selection (GVS) (George and McCulloch, 1993; Ntzoufras, 2002). An auxiliary inclusion variable $I_n$ (Eq. (2)) for each predictor was introduced to indicate whether that predictor was ‘in’ or ‘out’ of an individual iteration of the hierarchical modelling structure.

$$
I_n = \begin{cases} 
1, & \text{if } n^{th}\text{ predictor present} \\
0, & \text{if } n^{th}\text{ predictor absent}
\end{cases}
$$

(Figure 3: Analyses steps; the detailed methods used in the hierarchical modelling framework and model prediction and evaluation are in the following sections.)
I_n was modelled at the top level of the hierarchy which enabled the use of identical model structures (i.e., combination of predictors) across different sites. The overarching hierarchical modelling framework was defined as follows:

\[ y_{i,j} \sim N(\mu_{i,j}, \sigma) \]  

\[ \mu_{i,j} = \text{mean}_{j} + \text{std}_{j} \times \Delta_{i,j} \]  

\[ \Delta_{i,j} = \sum_{n=1}^{N} \theta_{n,j} \times x_{n,i,j} \]  

\[ \theta_{n,j} = I_{n} \times \beta_{n,j} \]  

The data-level model (Eq. (3)) assumed that the EMC of a particular constituent (e.g., one of TSS, NOX, EC, etc) at \( i \)th time step in the \( j \)th sub-catchment, \( y_{i,j} \), followed a normal distribution (denoted as \( N(\cdot) \)), with mean \( \mu_{i,j} \) and a global standard deviation \( \sigma \). The mean value, \( \mu_{i,j} \), was modelled as the observed site-level averaged EMC plus \( \text{std}_{j} \times \Delta_{i,j} \), with the latter term being defined as the deviation from this averaged value (Eq. (4)) (Guo et al., 2019). The deviation term incorporated the site-level observed standard deviation \( \text{std}_{j} \), making \( \Delta_{i,j} \) a standardised measure that could be compared across sites. \( \Delta_{i,j} \) was further modelled as a linear additive function (Eq. (5) of all candidate predictors \( x_{n} \) in \( n = 1, 2, \ldots, N = 14 \) (e.g., event average runoff, rainfall and NDVI). Consequently, \( \Delta_{i,j} \) was defined as the temporal variability in water quality, and was the quantity of interest. The effect size (\( \theta_{n,j} \)) of individual predictors was another latent variable used in the GVS, and was estimated as the product of \( I_{n} \) and the regression coefficient \( \beta_{n,j} \) (Eq. (6)), such that \( \theta_{n,j} \) was either \( \beta_{n,j} (I_{n} = 1) \), or 0 (\( I_{n} = 0 \)).

### 2.3.2 Hierarchical prior specification and Bayesian inference of key drivers

Bayesian inference required specification of prior distributions for each model parameter. We used a hierarchical conditional prior specification for predictor coefficients, allowing the site-specific parameter values that describe the effects of each of the temporal predictors \( (\beta_{1,j}, \beta_{2,j}, \ldots, \beta_{n,j}) \) to be exchangeable between sites (Liu et al., 2008a; O’Hara and Sillanpää, 2009; Webb and King, 2009). The detail specification of priors for each model parameter can be found in Appendix A. In addition, to identify key drivers affecting temporal changes in water quality, the posterior inclusion probability (PIP - \( P(I_{n} = 1 | y) \), Eq. (A8) in Appendix A) of each predictor was used to compare the relative importance of individual predictors (i.e., how often the \( n^{th} \) predictor was ‘in’ the model).

### 2.3.3 Prediction from multi-model inference

We used Bayesian Model Averaging to generate an ensemble of predictions of temporal variation in EMC for individual constituents (Eq. (7)). The average posterior distribution of a quantity of interest (i.e., temporal variability in EMC) was
generated using the parameters (e.g., $\beta_{1,j}, \beta_{2,j}, ..., \beta_{n,j}$) sampled from the posterior distribution to simulate EMC values using the specific model, defined as follows:

$$[\hat{y}|y] = \sum_{x=1}^{L} [\hat{y}|y,M_x]P(M_x|y)$$

(7)

where $[\hat{y}|y,M_x]$ is the posterior distribution of a vector $\hat{y}$ of (prediction) derived from model $M_x$, and $P(M_x|y)$ is the posterior model probability (PMP, Eq. (A8), in Appendix A) (Hooten and Hobbs, 2015; O’Hara and Sillanpää, 2009).

2.3.4 Model evaluation and implementation

The proposed modelling framework was applied to the two clusters of sites independently. This allowed an investigation of whether the spatial heterogeneity in catchment landscapes led to differences in the key factors controlling temporal variation in water quality. The key drivers were determined as the predictors with a PIP above 0.8 (i.e., over 80% of the models included these predictors).

To further understand the reliability and robustness of the BMA framework, the consistency of the posterior inclusion probability of individual predictors was investigated by resampling subsets of the observations multiple times (Kohavi, 1995). For each cluster, 80% of events within one site were first randomly selected and the posterior inclusion probability for this subset of observations was estimated. This was repeated 1,000 times to produce a distribution of posterior inclusion probabilities for individual predictors, which was then used to assess the uncertainty in the posterior inclusion probability.

An ensemble of the averaged prediction in temporal variability of each event was obtained from each iteration of parameter updating using Markov chain Monte Carlo (MCMC). The model fit was evaluated using the Nash-Sutcliffe coefficient (NSE) (Nash and Sutcliffe, 1970) between the observed temporal variability and the median of ensemble predictions $\hat{y}$ derived from the BMA (Eq. (7)). The NSE was calculated at both the cluster- and site-levels. The model residuals were also checked for normality and heteroscedasticity (i.e., relationship between the residual and predictors). In addition, model performance was evaluated by providing the 50% and 95% credible interval (CI) of each prediction.

To compare the relative importance of the predictors that have been widely used in existing literature (i.e., runoff and rainfall) and other predictors (e.g., soil moisture, temperature, evapotranspiration, and vegetation cover), the modelling framework was re-calibrated using only the rainfall/runoff related predictors (including all pre-, during- and post-event predictors). This estimated the degree of improvement in the model’s explanatory power with the inclusion of environmental variables, such as catchment wetness and ground vegetation cover conditions.

The hierarchical modelling framework was implemented in *JAGS* (Plummer, 2003, 2013a), using the package *rjags* in *R* (Plummer, 2013b; R Core Team, 2013), which enabled both the estimation of parameter values from prior distributions with Markov chain Monte Carlo (MCMC) and the generation of model-averaged predictions. The MCMC sampling had three parallel chains with 25,000 iterations for each chain. The first 5,000 iterations were discarded as a ‘burn-in’ period to allow
convergence of the Markov chains, resulting in 60,000 values to estimate the posterior distribution for each model parameter and make model predictions.

3. Results

3.1 Key drivers of temporal variability in water quality

The three key measures that were used to quantify the effect of individual predictors are: (1) estimates of posterior inclusion probability (PIP), which quantifies relative importance of individual predictors; (2) posterior model probability (PMP), which estimates differences in plausible model structures; and (3) posterior distributions of coefficients for the key drivers (i.e., effect size, e.g., \( \theta_{1,j}, \theta_{2,j}, \ldots, \theta_{n,j} \) in Eq. (6)), which measures direction and magnitude of the effect of key predictors on water quality temporal variability.

Posterior inclusion probability (Figure 4 and Table C3 in Appendix C) from the Bayesian modelling results indicated that, in general, antecedent vegetation condition and antecedent soil moisture were key factors in explaining temporal variation in water quality, especially for Cluster 2 (warmer, drier) sites. Catchment runoff and rainfall were the second most important group of factors, especially for particulate pollutants (TSS, PN and PP; Clusters 1 and 2) and salinity. In addition, the three groups of predictors (pre-, during-, post-event) showed varying effects among the constituents. With regard to during-event conditions, event average runoff (\( \text{Event\_ave\_Q} \)), event maximum runoff (\( \text{Event\_max\_Q} \)) and event average rainfall (\( \text{Event\_ave\_P} \)) were three important factors with relatively high PIP. In contrast, among pre-event conditions, antecedent NDVI (\( \text{Ante\_NDVI} \)) and antecedent soil moisture (\( \text{Ante\_SM} \)) were driving factors for the majority of the constituents. Post-event runoff (\( \text{Post\_Q} \)) only affected a few constituents (e.g., on NO\(_X\) and FRP for Cluster 2), compared with the other two groups of predictors. Overall, there were notable differences in the important predictors for Clusters 1 and 2, and more important predictors were found for the Cluster 2 sites.
Results from here on will focus mainly on TSS, NO$_X$ and FRP, due to their impacts on the marine receiving environment. Results for the other six constituents are in the Supplementary Materials. Figure 5 shows the posterior model probabilities for TSS, NO$_X$ and FRP for the 100 models with highest PMP (Figures B2 and B3 in Appendix B show other constituents). Red indicates a negative influence and blue a positive influence. The difference in PIP between the two clusters resulted in quite different plausible model structures (models with relatively high posterior model probability). A stand-out difference between the results for the two Clusters was antecedent vegetation cover condition ($Ante_{NDVI}$), which tended to be a more important predictor of TSS for Cluster two, than for Cluster one (Figure 5 [a]). In addition, the plausible models for Cluster 2 were generally more complex (with a larger number of predictors), expect for DOP and EC (Figures B2 and B3).
Figure 5: Comparison of BMA model coefficients and cumulative model probabilities (only the first 100 models ranked according to the highest probability are shown) between Cluster 1 ("wet" - left) and Cluster 2 ("dry" - right) sites for [a] TSS, [b] NOX and [c] FRP. The order of predictors on the y-axis was ranked based on the posterior inclusion probability. Each column in the heatmap represents the one specific model (ranked from highest model probability from left to right) and the width of the column.
is normalised by the posterior model probability (i.e., the widest columns indicate models with the largest increase in probability compared to the next most probable model). The colour indicates the direction of the coefficients: red = negative; blue = positive. The coefficient value was averaged across the posterior median value of the site-specific coefficient within each cluster (effect size, $\theta_{nj}$, in Equation 6); the definition of the abbreviations of each predictor on the y-axis are in Table 3.

The distribution of posterior model coefficients for the key predictors (Figure 6, Figures B4 and B5) further demonstrated that the key drivers of temporal variability in water quality vary between catchments and between constituents. During-event runoff and rainfall tended to have a positive effect on sediment and particulate constituents and, a negative effect on NOX and EC. In addition, there was strong negative effect of antecedent vegetation condition on the majority of the constituents.
Figure 6: Distribution of median of site-level coefficients for all plausible models in BMA between Cluster 1 ("wet" - left) and Cluster 2 ("dry" - right) sites for: [a] TSS; [b] NOx and [c] FRP. Only predictors with PIP > 0.8 are included. For each specific model structure, the coefficient value of a predictor was the median of the site-specific coefficient across all sites (effect size, $\theta_{ij}$, in Eq. (6)). The distribution of this value thus represents the probability of the model (PMP), as well as variability in the same predictor across different sites; black dots = the median; grey vertical lines = 95% CI; blue coloured vertical lines = 50% CI; the definition of the abbreviation of each predictor on x-axis are in Table 3.

The uncertainty in PIP, derived from 1,000 subsampled BMA runs (Figure 7, Figures B6 and B7) highlighted that the BMA results were robust for most constituents, except for EC (Figure B7 [c]). BMA tends to identify important predictors and less sensitive to the input data which is evidenced by the relatively narrow range of interquartile ranges (IQR), when PIP for a specific predictors is large (e.g., antecedent soil moisture for FRP in Figure 7). It is also worth noting that large uncertainty in the PIP for EC was observed, indicating the BMA results were sensitive to the observations of EC. This might be related to data availability, which is further discussed in Sect. 4.2.

Figure 7: The comparisons of the distribution of posterior inclusion probabilities of the individual predictors derived from 1,000 subsampled BMA runs; the boxes are the interquartile ranges (IQR, 25th to 75th percentile), and the whiskers are the ranges between 1.5 IQR of the lower quartile and 1.5 IQR of the higher quartile; the vertical bar = median; blue = Cluster 1 ("wet"); red = Cluster 2 ("dry"); the definition of abbreviation of each predictor on y-axis are in Table 3.
3.2 Predictive performance

Moderate levels of temporal variability were explained by the BMA framework for the two independent clusters of sites (Figure 8, Figures B8 and B9). At the cluster level, the NSE ranged from 0.04 (DOP) to 0.68 (EC) and from 0.34 (NH₄) to 0.64 (NOₓ) for Clusters 1 and 2 (full model columns in Table 4), respectively. The comparison of the modelling performance (posterior median of BMA prediction) showed that the modelling framework performed better on the Cluster 2 sites than Cluster 1 (Figure 8, red 50% prediction CI – Cluster 2), except for NH₄ and EC (not shown). This was reflected in a better match to the 1:1 line within the 90% prediction CI for Cluster two catchments. It is also worth noting that the prediction interval for EC (Figure B9 [c]) was much wider than the rest of the constituents. Similar results were found in the site-level performance, with the average site-level NSE (Figure 9) for Cluster 2 models typical higher than for Cluster 1. The site-specific performance varied across sites, with the largest variation in EC (NSE for the Cluster 2 result ranged from approximately 0.20 to 0.90). The modelling performance of DOP in the Cluster 1 sites was poor (NSE = 0.04); all candidate covariates had low predictive power, resulting in the poor mixing of chains of the inclusion variable Iᵢ (i.e., posterior Iᵢ was around 0.5). The model residuals were normally distributed (Figure B10) and there was no clear heteroscedasticity within the residuals (Figures B11 to B19).
Figure 8: Performance of the BMA models of the temporal variability of three constituents across 32 sites, represented by prediction intervals from BMA and observed Box-Cox EMC across two clusters of sites for: (a) TSS; (b) NOX; and (c) FRP. Each bar shows a single event and all events at all sites in the cluster are included. The NSE values were calculated based on median predictions. Black dots show prediction median; grey vertical lines show 95% CI; coloured vertical lines show 50% CI; blue is Cluster 1 (“wet”); red is Cluster 2 (“dry”); and dashed black lines are the 1:1 relationship.
Figure 9: Distribution of site-level NSE for modelled the temporal variability of two clusters of sites. The interpretation of boxplot is the same as Figure 7. NSE values were calculated based on site-level predictions of event median EMC; blue is Cluster 1 (“wet”); and red is Cluster 2 (“dry”) (i.e., each boxplot is comprised of respective number of sites in each cluster, one for each catchment).

Table 4 compares the model performance using rainfall/runoff related predictors only and all candidate predictors (full model). A large increase in NSE was found for most dissolved nutrient species (e.g., NO$_X$, NH$_4$, DON, FRP and DOP) for the full model. Notably, for NH$_4$ in Cluster 1, factors other than rainfall and runoff explained almost all the variability that could be captured by the BMA.

Table 4: Comparison between BMA performance using rainfall/runoff predictors only and all candidate predictors (full models).

<table>
<thead>
<tr>
<th>Constituent</th>
<th>NSE for Cluster 1 (“wet”)</th>
<th>NSE for Cluster 2 (“dry”)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rainfall, runoff only</td>
<td>Full model</td>
</tr>
<tr>
<td>TSS</td>
<td>0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>PN</td>
<td>0.32</td>
<td>0.40</td>
</tr>
<tr>
<td>NO$_X$</td>
<td>0.23</td>
<td>0.49</td>
</tr>
<tr>
<td>NH$_4$</td>
<td>0.00</td>
<td>0.39</td>
</tr>
<tr>
<td>DON</td>
<td>0.20</td>
<td>0.37</td>
</tr>
<tr>
<td>FRP</td>
<td>0.27</td>
<td>0.45</td>
</tr>
<tr>
<td>DOP</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td>PP</td>
<td>0.29</td>
<td>0.36</td>
</tr>
<tr>
<td>EC</td>
<td>0.41</td>
<td>0.68</td>
</tr>
</tbody>
</table>
4 Discussion

4.1 Factors influencing temporal variability in stream water quality

4.1.1 Runoff and rainfall

Our results demonstrated that runoff and rainfall were important factors in explaining the temporal dynamics of particulate pollutants (i.e., TSS, PN and PP) and dissolved species (e.g., NO\textsubscript{X}, DOP and EC) in the GBR catchments. These results align with the findings of previous studies that have used these variables to understand changes in water quality over time (Beiter et al., 2020; Letcher et al., 2002; Liu et al., 2008b; McKergow et al., 2003; Schwarz et al., 2006; Tilburg et al., 2015).

Hydrologic and climatic variables (i.e. rainfall and runoff) showed distinct effects on different constituents, as well as different groups of catchments. The positive effect of event runoff and rainfall on sediment and particulate nutrients (i.e., PN, PP) revealed their underlying impacts on pollutant mobilisation and transport processes in catchments (Ballantine et al., 2009; Guo et al., 2010; Heathwaite et al., 2000; Hirsch et al., 2010; Lintern et al., 2018b; Musolff et al., 2015). In contrast, there were negative effects of during-event runoff on NO\textsubscript{X} (Cluster 1), DOP (Cluster 2) and EC (both clusters). For NO\textsubscript{X} and EC, this was most likely caused by hydrological transport processes; these constituents tend to be transported to receiving rivers via subsurface flows (Kratz et al., 1997; McKergow et al., 2003). For events with relatively low surface runoff, higher NO\textsubscript{X} and EC event concentrations could be expected in these catchments (Clow and Sueker, 2000; Skoulidakis et al., 2006; Young et al., 1996). In addition, for DOP, in-stream biogeochemical cycling was likely to have caused the negative effect of event runoff. The events with low runoff, coupled with high temperatures (positive effect of event temperature for DOP Cluster 2, Figure B3 [a]) may relate to increases in the rate of P releases from organic forms at higher temperatures (Verheyen et al., 2015).

Post-event runoff (Post\textsubscript{Q}) showed effects on specific constituents (e.g., NO\textsubscript{X}, FRP and EC). Two alternative reasons might explain this. First, high post-event runoff may be an indicator of large baseflow contribution during the events (Cuomo and Guida, 2016). Therefore, as discussed in the above paragraph, constituents that can be transported through subsurface flows tend to be influenced by amount of runoff after event. Alternatively, it was significantly and positively correlated with other event characteristics and catchment biophysical conditions (e.g., vegetation cover, Figure B1). These inter-correlated factors together could have influenced pollutant source, mobilisation and delivery (see discussions below) (Granger et al., 2010; Lintern et al., 2018a).

4.1.2 Vegetation cover

Vegetation cover was another driving factor that was found to have influenced water quality dynamics; antecedent NDVI (Ante\textsubscript{NDVI}) was included in the plausible models more frequently than event NDVI. The negative effect of antecedent
NDVI on particulate and dissolved nutrients (except DOP) was in line with previous studies that have found that NDVI was negatively correlated with these constituent concentrations in streams (Griffith et al., 2002; Liu et al., 2015; Masocha et al., 2017). An explanation for these results could be that high vegetation groundcover tended to stabilise the surface soil and reduce sediment losses by erosion (Meyer et al., 1997; Singh et al., 2008). In addition, vegetation nutrient assimilation and retention processes consumed nutrients in sediment and waterbodies, and these processes peaked in spring and early summer, typically before the wet season in the GBR catchments (Tabacchi et al., 2000; Uwimana et al., 2018; Vymazal, 2007).

The effect of antecedent NDVI varied among groups of constituents in Clusters 1 and 2. Specifically, it was a key predictor for NO$_X$, NH$_4$ and FRP for Cluster one, and almost all constituents for Cluster 2. This can be explained by the contrasting landscapes and climate of these two regions (Liu et al., 2018). In the dense, vegetation-covered catchments in Cluster 1 (i.e., the sites in the Wet Tropics), dissolved inorganic nutrient losses were likely due to more fertile soils (e.g., application of fertiliser on sugarcane) during the growing season (Hunter and Walton, 2008; McKergow et al., 2005a). Furthermore, denser natural vegetation cover (e.g., riparian vegetation and forest) could increase plant uptake and assimilation of dissolved nutrients compared to the sparse vegetation cover in the Dry Tropics (Cluster 2) region. Conversely, among Cluster 2 sites, vegetation coverage showed clear seasonal variation, which was linked closely to the seasonality in rainfall and grazing activity. Sediments and particulate pollutants were likely to be mobilized in grazed catchments (high rate of soil erosion) and delivered to streams via surface runoff (Ballantine et al., 2009; Neil et al., 2002; Turner et al., 2012). More importantly, high vegetation cover tended to mitigate mobilisation of pollutants, through stabilising the surface soil and such that reduces sediment losses from erosion (Meyer et al., 1997; Rey, 2004; Singh et al., 2008; Zorzal-Almeida et al., 2018).

### 4.1.3 Soil moisture and evapotranspiration

The results showed that soil moisture (SM) and actual evapotranspiration (AET) had a high impact on different constituents, particularly in the Cluster 2 catchments (e.g., antecedent soil moisture [DON and EC], antecedent AET [TSS and EC]). These two variables were inter-correlated and affect the hydrological cycle and vegetation cover (Correll, 1996; Correll and Weller, 1989; Legates et al., 2011). The results indicated that antecedent soil moisture had a negative effect on PN, NO$_X$, NH$_4$, DON, DOP and FRP. On one hand, this was expected as antecedent soil moisture was positively correlated with vegetation cover, and high soil moisture tends to reduce soil erosion and increase plant nutrient uptake. It may also be that soil water content affected soil microbial activity, influencing the biogeochemical processes in catchments, such as denitrification (Doran et al., 1988; Doran, 1980; Weier et al., 1993). The rate of denitrification was also enhanced under anoxic conditions, when soil moisture was high (Skopp et al., 1990; Zhu et al., 2018a). On the other hand, higher soil water can be associated with increased shallow subsurface flow and leaching of some constituents such as NO$_X$ (Zhu et al., 2018b). This appears not to occur to a sufficient extent for it to over-ride other impacts of soil moisture.
4.1.4 Temperature

Our results suggested that average event temperature (Event_T) had a positive effect on NO\textsubscript{X}, FRP, and DOP. This may be attributed to the strong negative cross-correlation between temperature and event runoff and antecedent vegetation condition (Figure B1). Rainfall during a warmer period might have been associated with less event runoff, resulting in higher event mean concentrations (Sect. 4.1.1). The effect of event temperature can be also attributed to the fact that the higher temperatures could lead to more recent mineralisation of nutrients, increasing readily transportable dissolved nutrient sources (Liu et al., 2017; Wang et al., 2020). In addition, higher event temperature might be associated with higher pre-event temperature, resulting in poor groundcover, potentially lowering the dissolved nutrients losses through plant assimilation/uptake (Sect. 0) (Muro et al., 2018).

4.2 Predicting temporal variations in water quality

The Bayesian modelling framework in this study provided a useful tool to assess in-stream water quality dynamics. The models were able to explain more temporal variation in NO\textsubscript{X} and EC than in other constituents. This is related to the sources and delivery processes of these two constituents. Anthropogenic inputs (e.g. agriculture) for NO\textsubscript{X}, and large stores in groundwater together with limited geochemical transformation for EC (salts) suggested that temporal changes in event concentration could be well-captured by the changes in catchment hydroclimatic and vegetation conditions. In addition, NO\textsubscript{X} and EC tend to be transported in subsurface flow pathways. The dynamics of catchment soil wetness and vegetation cover have been previously linked to hydrological interactions between surface and subsurface flows (Ursino et al., 2004). The incorporation of soil moisture and vegetation cover into the Bayesian modelling framework more readily allowed the description of the main ecohydrological processes of these two constituents.

In contrast, model performance for DOP was poor in Cluster 1 catchments, which can be explained by two reasons. First, in the Wet Tropics catchments, DOP concentrations were generally stable, regardless of changes in flow, which can be explained by chemical exchange processes between water and sediment in stream (White et al., 1998). This means that the variability in DOP cannot be captured by the environmental variables considered here. Second, the poor performance might be attributed to the data set having fewer observations of DOP EMCS among Cluster 1 sites. There were only 66 observations, compared to the next lowest number of 167 (EC) among other constituents in the Cluster 1 catchments, which may not be sufficient to fully inform the model. This small sample size could have led to outcomes of: 1) poor mixing of MCMC chains for inclusion variables (Figure B7 [a]), where no predictors showed predictive power; and 2) the BMA failed to identify the plausible models, since none of the candidate models had enough predictive power to fit the data well (Guthke, 2017; Höge et al., 2019). Continuous DOP monitoring would be required to achieve a better understanding of the factors driving temporal variation in this constituent.

Statistical modelling in hydrology or water quality is affected by uncertainty, only some of which can be characterised within any particular modelling framework (Beck, 1987; Kavetski et al., 2006; Mantovan et al., 2006; Renard et al., 2010;
Yang et al., 2007). The Bayesian modelling framework used in this study incorporated the uncertainties in model selection (between-model), observations and model parameters (within-model) directly into the model predictions (Steel, 2019). This is a more comprehensive characterisation than in studies where model structures are assumed a priori. Reporting of predictive uncertainty of temporal variations in water quality also provided valuable information on the confidence in the averaged predictions. Nevertheless, limitations remain in the BMA approach which are important to understand. For example, for EC, there was a larger predictive uncertainty and larger uncertainty in posterior inclusion probability for each predictor from the robustness assessment than estimated in the fit to the complete data set. One limitation of BMA is that the posterior model probability could be sensitive to the specification of the parameter prior distribution (Fernandez et al., 2001). Specifying more informative priors on model parameters (i.e., inclusion variable $I_n$) would have the effect of restricting the set of candidate models (Eicher et al., 2011; Rockey et al., 2016). Indeed, several studies have compared different predictive performances of different prior specification of BMA coefficients and found that the choice of prior matters (Bayarri et al., 2012; Eicher et al., 2011; Liang et al., 2008). Future investigation of the sensitivity of prior distributions for BMA coefficients might achieve a reduction in predictive uncertainty and instability in posterior inclusion probabilities.

### 4.3 Management implications

The identification of key drivers of temporal variation in water quality can inform catchment water quality management. The results of this study showed that the effects of hydro-climatic drivers (e.g., rainfall and runoff) and vegetation cover varied among constituents and regions. This may allow funding bodies, such as government, regional natural resource management groups, to identify regions where land management and restoration would have a greater effect on mitigating sediments and nutrients export. The results suggested that, compared to wet catchments, maintaining vegetation ground cover in large dry grazed catchments (e.g., the Burdekin and Fitzroy catchments in Cluster 2) before the wet seasons could be an effective way of reducing sediment losses via erosion processes. These results are consistent with current, improved land management practices across the GBR catchments (Brodie et al., 2009; Brodie et al., 2012; Government, 2017; Hunter and Walton, 2008; Star et al., 2015). Management measures (e.g., establishment of wetlands, re-vegetation/rehabilitation of gully and stabilisation of river banks) can reduce sediment losses from hillslope and gully erosions (Koci et al., 2020; Loch, 2000; Sherriff et al., 2016; Wilkinson et al., 2018). In addition, catchment-specific management that accounts for temporal variation in catchment hydrological connectivity is required for the control of dissolved nutrients. Dominant flow pathways for dissolved nutrients can vary spatially and temporally. For example, subsurface flow in the Wet Tropics region have tended to transmit more dissolved nutrient, because prolonged wet conditions lead to this region that is more likely to be connected via lateral subsurface flow (Geng et al., 2017; Stieglitz et al., 2003). The enhanced mobilisation of leached dissolved nutrients from intensive cropping (e.g., sugarcane) from perched groundwater should be targeted in these catchments (Melland et al., 2012). Management practices, such as conservation tillage, and adaptation of ‘4R’ concept (right source, right rate, right time, right place) for fertiliser application may help to minimise dissolved nitrogen losses (Cestti et al., 2003; Lintern et al., 2020; Merriman et al., 2009; Snyder, 2017).
5. Conclusions

This study provides a data-driven understanding of key drivers influencing the temporal variation in water quality. A hierarchical Bayesian model averaging framework was used to identify the key environmental drivers and predict the water quality dynamics at multiple catchments. Results showed that the temporal dynamics of water quality can be predicted well using models considering the combined effects of hydroclimate and vegetation groundcover. The effects of key hydroclimatic and vegetation conditions varied among different constituents, and across regions. This study reinforces the importance of vegetation cover management as one key management response, especially for large grazed catchments. Future investigation could involve the development of a spatio-temporal modelling framework to fully capture the water quality dynamics. More importantly, it has continued to be challenging to prioritise management practices and evaluate the effectiveness of the improved management interventions. Consequently, with more land management surveys and continuous water quality monitoring data available, an extended temporal or spatio-temporal modelling framework could potentially be used to assess if the success of the restoration measures.

Data availability

Water quality data that supported this study was available upon request from the Great Barrier Reef Catchment Loads Monitoring Program (GBREvents@dsiti.qld.gov.au). Sources of explanatory variables were listed in Table 2.

Author contribution

All authors contributed to the design of the research. SL carried out data collation, performed the simulations and prepared the manuscript with contributions from all co-authors. All authors contributed to the interpretation of the results and provided feedback.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study was supported by the Australian Research Council (LP140100495), the Environment Protection Authority Victoria, the Victorian Department of Environment, Land, Water and Planning, Bureau of Meteorology and Queensland Department of Natural Resources, Mines and Energy. The author would like to acknowledge the efforts of the Queensland Department of Environment and Science who provided the water quality monitoring data. The authors would also like to
offer sincere gratitude to Ms. Jie Jian for her assistance in geospatial database compilation. Dr. Paul Leahy, Mr. Malcolm Watson, Dr. Ulrike Bende-Michl, Mr. Paul Wilson, and Ms. Belinda Thompson all of whom provided valuable advice in the preparation of this manuscript.

References


Bouraoui, F., Galbiati, L., and Bidoglio, G.: Climate change impacts on nutrient loads in the Yorkshire Ouse catchment (UK), Hydrology and earth system sciences discussions, 6, 197-209, 2002.


Hem, J. D.: Fluctuations in concentration of dissolved solids of some southwestern streams, Eos, Transactions American Geophysical Union, 29, 80-84, 1948.


Appendix A - Text

Hierarchical prior specification and Bayesian inference of key drivers

Bayesian inference required specification of prior distributions for each model parameter. A minimally-informative uniform prior (denote as $U(\cdot)$) between 0 and 10 was assigned to the global standard deviation ($\sigma$, Eq. A1) (Gelman, 2006). The prior of $I_n$ assumes that each indicator comes from an independent Bernoulli distribution, with a probability of 0.5 (Eq. A2) (Raftery et al., 1997). This vague prior results in each model structure having an equal prior model probability.

$$\sigma \sim U(0,10)$$  
$$I_n \sim \text{Bernoulli}(0.5)$$

We used a hierarchical conditional prior specification for predictor coefficients, allowing the site-specific parameter values that describe the effects of each temporal predictors ($\beta_{1,j}, \beta_{2,j}, \ldots, \beta_{n,j}$) to be exchangeable between sites (Liu et al., 2008; O’Hara and Sillanpää, 2009; Webb and King, 2009). The prior of $\beta_{n,j}$ was conditioned on $I_n$, resulting in a mixture distribution with ‘slab and spike’ prior, which was defined as follows,

$$\beta_{n,j} \mid (I_n = 1) \sim I_n N(0, \tau_n) + (1-I_n) N(0, \tau_{n,tune})$$

where $\beta_{n,j} \mid (I_n = 1)$ is the slab part of the mixture distribution. The $\beta_{n,j} \mid (I_n = 1)$ was estimated by including a higher-level distribution. The prior of $\beta_{n,j} \mid (I_n = 1)$ followed a normal distribution with random effect (Eq. A4), with the $\tau_n$ drawn from a common prior distribution, defined as a hyperparameter (i.e., uniform distribution between 0 to 20, Eq. A5) (Gelman, 2006; Kruschke, 2014).

$$\beta_{n,j} \mid (I_n = 1) \sim N(0, \tau_n)$$
$$\tau_n \sim U(0,20)$$

For the spike component, a data-dependent prior was specified for $\beta_{n,j} \mid (I_n = 0)$, drawing from a pseudo-prior (Eq. A6), that is, an prior distribution with no effect on the posterior distribution, but facilitating the mixing of the Gibbs sampler.

$$\beta_{n,j} \mid (I_n = 0) \sim N(0, \tau_{n,tune})$$

We estimated $\tau_{n,tune}$ from the standard deviations of the posterior of the $\beta_{n,j}$ in a global model structure (i.e., modelling structure using all predictors), as suggested by Carlin and Chib (1995) and Linden and Roloff (2015). The prior of $\beta_{n,j} \mid (I_n = 0)$ was near the posterior estimates to facilitate mixing in the MCMC (Hooten and Hobb, 2015).

The posterior inclusion probability (PIP - $P(I_n = 1 \mid y)$, Eq.A7) of each predictor was used to compare the relative importance of individual predictors (i.e., how often the $n^{th}$ predictor was ‘in’ the model).

$$P(I_n = 1 \mid y) = \frac{1}{T} \sum_{t=1}^{T} I(I_n^{(t)} = 1)$$
where $T$ is the total number of iterations of Markov chains. The different combination of $I_n$ at each MCMC sampling represents a specific model structure. According to Bayes’ theorem, the posterior model probability ($P(M_k | y)$) can be estimated as,

$$P(M_k | y) = \frac{[y | M_k] P(M_k)}{\sum_{x=1}^{L} [y | M_x] P(M_x)}$$

where $L$ is the total number of possible models, and $P(M_k)$ is the prior probability of model $M_k$, among a group of models $M_x$, $x = 1, \ldots, X$. This posterior model probability can be obtained by assessing the frequency of a particular combination of $I_n$ during the MCMC sampling.

Reference


Gelman, A.: Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), Bayesian analysis, 1, 515-534, 2006.


Appendix B - Figure

![Spearman's Rank correlation between 14 candidate covariates.](image)

Figure B1: Spearman’s Rank correlation between 14 candidate covariates.
Figure B2: Comparison of BMA model coefficient and cumulative model probability (top 100 models) between two clusters for: (a) PN, (b) NH$_4$, and (c) DON. Left - cluster one sites and Right – cluster two sites. The order of predictors on the y-axis was ranked based on the posterior inclusion probability. Each column in the heatmap represents the one specific model (ranked from highest model probability) and the width of the column is normalised by the posterior model probability. The colour indicates the direction of the coefficients, red – negative and blue – positive. Note: the coefficient value was averaged across the posterior median value of site-specific coefficient within each cluster (effect size, $\theta_{n,j}$, in Eq. (6)).
Figure B3: Comparison of BMA model coefficient and cumulative model probability (top 100 models) between two clusters for: (a) DOP, (b) PP and (c) EC. Left - cluster one sites and Right – cluster two sites. The order of predictors on the y-axis was ranked based on the posterior inclusion probability. Each column in the heatmap represents the one specific model (ranked from highest model probability) and the width of the column is normalised by the posterior model probability. The colour indicates the direction of the coefficients, red – negative and blue – positive. Note: the coefficient value was averaged across the posterior median value of site-specific coefficient within each cluster (effect size, \( \theta_{n,j} \), in Eq. (6)).
Figure B4: Distribution of median of site-level coefficients for all plausible models in BMA. (a) PN, (b) NH$_4$ and (c) DON. Only predictors with PIP > 0.8 are included. For each specific model structure, the coefficient value of a predictor was the median of site-specific coefficient across all sites (effect size, $\theta_{n,d}$, in Eq. (6)). The distribution of this value thus represents the probability of the model (PMP), as well as variability in the same predictor across different sites. Note: black dots indicate the median; grey
vertical lines indicate 95% CI and blue coloured vertical lines indicates 50% CI. The definition of abbreviation of each predictor can be found in Table 3.

Figure B5: Distribution of median of site-level coefficients for all plausible models in BMA. (a) DOP, (b) PP and (c) EC. Only predictors with PIP > 0.8 are included. For each specific model structure, the coefficient value of a predictor was the median of
site-specific coefficient across all sites (effect size, $\theta_{n,j}$ in Equation 6). The distribution of this value thus represents the probability of the model (PMP), as well as variability in the same predictor across different sites. Note: black dots indicate the median; grey vertical lines indicate 95% CI and blue coloured vertical lines indicates 50% CI. The definition of abbreviation of each predictor can be found in Table 3.

Figure B6: The comparisons of distribution of posterior inclusion probability of individual predictors derived from 1,000 subsampled BMA runs. The interpretation of boxplot is the same as Figure 9. Note: colour represents different clusters: blue - Cluster one and red - Cluster two. The definition of abbreviation of each predictor can be found in Table 3.
Figure B7: The comparisons of distribution of posterior inclusion probability of individual predictors derived from 1,000 subsampling BMA. The interpretation of boxplot is the same as Figure 3. Note: colour represents different clusters: blue - Cluster one and red - Cluster two.
Figure B8: Performance of the BMA models of the temporal variability of nine constituents across 32 sites, represented by prediction intervals from BMA and observed Box-Cox EMC across two clusters of sites for: (a) PN; (b) NH\textsubscript{4} and (c) DON. The NSE values are calculated based on predictions within group- (cluster) level. Note: black dots are the prediction median; grey vertical lines are the 95% CI and coloured vertical lines indicates 50% CI: blue - Cluster one and red - Cluster two. The dashed black line is the 1:1 relationship.
Figure B9: Performance of the BMA models of the temporal variability of nine constituents across 32 sites, represented by prediction intervals from BMA and observed Box-Cox EMC across two clusters of sites for: (a) DOP; (b) PP and (c) EC. The NSE values are calculated based on predictions within group- (cluster) level. *Note:* black dots are the prediction median; grey vertical lines are the 95% CI and coloured vertical lines indicates 50% CI: blue - Cluster one and red - Cluster two. The dashed black lines is the 1:1 relationship.
Figure B10: Histograms showing distribution of residuals of nine constituents from BMA predictions. Red – Cluster one; Blue – Cluster two.
Figure B11: Relationship between residual in median of BMA prediction of TSS and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B12: Relationship between residual in median of BMA prediction of PN and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B13: Relationship between residual in median of BMA prediction of NO$\text{X}$ and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B14: Relationship between residual in median of BMA prediction of NH$_4$ and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B15: Relationship between residual in median of BMA prediction of DON and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B16: Relationship between residual in median of BMA prediction of FRP and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B17: Relationship between residual in median of BMA prediction of DOP and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B18: Relationship between residual in median of BMA prediction of PP and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
Figure B19: Relationship between residual in median of BMA prediction of EC and 14 candidate covariates in BMA. Note, difference colours indicate two clusters: Red – Cluster one; Blue – Cluster two.
### Appendix C - Table

#### Table C.1. Description of 32 sites in the GBR catchments

<table>
<thead>
<tr>
<th>NRM</th>
<th>Site ID</th>
<th>River and site name</th>
<th>Latitude/°</th>
<th>Longitude/°</th>
<th>Catchment area / km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cape York</td>
<td>105107A</td>
<td>Normanby River at Kalpowar Crossing</td>
<td>-14.9185</td>
<td>144.2100</td>
<td>12934</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>110001D</td>
<td>Barron River at Myola</td>
<td>-16.7998</td>
<td>145.6121</td>
<td>1945</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>110002A</td>
<td>Barron River at Mareeba</td>
<td>-17.0022</td>
<td>145.4293</td>
<td>836</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>110003A</td>
<td>Barron River at Picnic Crossing</td>
<td>-17.2591</td>
<td>145.5386</td>
<td>228</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>1110056</td>
<td>Mulgrave River at Deeral</td>
<td>-17.2672</td>
<td>145.9544</td>
<td>524</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>1120049</td>
<td>North Johnstone River at Old Bruce Hwy Bridge (Goondi)</td>
<td>-17.5059</td>
<td>145.9920</td>
<td>959</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>112004A</td>
<td>North Johnstone River at Tung Oil</td>
<td>-17.5456</td>
<td>145.9325</td>
<td>925</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>112101B</td>
<td>South Johnstone River at Upstream Central Mill</td>
<td>-17.6106</td>
<td>145.9789</td>
<td>400</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>113006A</td>
<td>Tully River at Euramo</td>
<td>-17.9936</td>
<td>145.9411</td>
<td>1450</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>113015A</td>
<td>Tully River at Tully Gorge National Park</td>
<td>-17.7727</td>
<td>145.6507</td>
<td>482</td>
</tr>
<tr>
<td>Wet tropics</td>
<td>116001F</td>
<td>Herbert River at Ingham</td>
<td>-18.6328</td>
<td>146.1427</td>
<td>8581</td>
</tr>
<tr>
<td>Burdekin</td>
<td>119101A</td>
<td>Barratta Creek at Northcote</td>
<td>-19.6923</td>
<td>147.1688</td>
<td>753</td>
</tr>
<tr>
<td>Burdekin</td>
<td>120001A</td>
<td>Burdekin River at Home Hill</td>
<td>-19.6436</td>
<td>147.3958</td>
<td>129939</td>
</tr>
<tr>
<td>Burdekin</td>
<td>120002C</td>
<td>Burdekin River at Sellheim</td>
<td>-20.0078</td>
<td>146.4369</td>
<td>36290</td>
</tr>
<tr>
<td>Burdekin</td>
<td>120301B</td>
<td>Belyando River at Gregory Development Rd.</td>
<td>-21.5423</td>
<td>146.8656</td>
<td>35410</td>
</tr>
<tr>
<td>Burdekin</td>
<td>120302B</td>
<td>Cape River at Taemas</td>
<td>-20.9996</td>
<td>146.4271</td>
<td>16070</td>
</tr>
<tr>
<td>Burdekin</td>
<td>120310A</td>
<td>Suttor River at Bowen Developmental Road</td>
<td>-21.5375</td>
<td>147.0424</td>
<td>10760</td>
</tr>
<tr>
<td>Mackay Whitsunday</td>
<td>124001B</td>
<td>O’Connell River at Stafford’s Crossing</td>
<td>-20.6526</td>
<td>148.5730</td>
<td>342</td>
</tr>
<tr>
<td>Mackay Whitsunday</td>
<td>1240062</td>
<td>O’Connell River at Caravan Park</td>
<td>-20.5664</td>
<td>148.6117</td>
<td>825</td>
</tr>
<tr>
<td>Mackay Whitsunday</td>
<td>125013A</td>
<td>Pioneer River at Dumbarton Pump Station</td>
<td>-21.1441</td>
<td>149.0753</td>
<td>1485</td>
</tr>
<tr>
<td>Mackay Whitsunday</td>
<td>126001A</td>
<td>Sandy Creek at Homebush</td>
<td>-21.2831</td>
<td>149.0228</td>
<td>326</td>
</tr>
<tr>
<td>Fitzroy</td>
<td>1300000</td>
<td>Fitzroy River at Rockhampton</td>
<td>-23.3175</td>
<td>150.4819</td>
<td>139159</td>
</tr>
<tr>
<td>Fitzroy</td>
<td>130206A</td>
<td>Theresa Creek at Gregory Highway</td>
<td>-23.4292</td>
<td>148.1514</td>
<td>8485</td>
</tr>
<tr>
<td>Fitzroy</td>
<td>130302A</td>
<td>Dawson River at Taroom</td>
<td>-25.6376</td>
<td>149.7901</td>
<td>15850</td>
</tr>
<tr>
<td>Fitzroy</td>
<td>130504B</td>
<td>Comet River at Comet Weir</td>
<td>-23.6125</td>
<td>148.5514</td>
<td>16460</td>
</tr>
<tr>
<td>Burnett Mary</td>
<td>136002D</td>
<td>Burnett River at Mt Lawless</td>
<td>-25.5447</td>
<td>151.6549</td>
<td>29360</td>
</tr>
<tr>
<td>Burnett Mary</td>
<td>136004A</td>
<td>Jones Weir HW</td>
<td>-25.5948</td>
<td>151.2964</td>
<td>21700</td>
</tr>
<tr>
<td>Burnett Mary</td>
<td>136014A</td>
<td>Burnett River at Ben Anderson Barrage Head Water</td>
<td>-24.8896</td>
<td>152.2922</td>
<td>32891</td>
</tr>
<tr>
<td>Burnett Mary</td>
<td>136094A</td>
<td>Burnett River at Jones Weir (TW)</td>
<td>-25.5948</td>
<td>151.2974</td>
<td>21700</td>
</tr>
</tbody>
</table>
### Table C2. Number of EMCs for each constituent

<table>
<thead>
<tr>
<th>Cluster</th>
<th>TSS</th>
<th>PN</th>
<th>NO\textsubscript{x}</th>
<th>NH\textsubscript{4}</th>
<th>DON</th>
<th>FRP</th>
<th>DOP</th>
<th>PP</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>225</td>
<td>207</td>
<td>218</td>
<td>217</td>
<td>215</td>
<td>210</td>
<td>66</td>
<td>186</td>
<td>174</td>
</tr>
<tr>
<td>Two</td>
<td>381</td>
<td>370</td>
<td>372</td>
<td>370</td>
<td>373</td>
<td>372</td>
<td>231</td>
<td>366</td>
<td>354</td>
</tr>
<tr>
<td>% of event monitored</td>
<td>43</td>
<td>41</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>41</td>
<td>21</td>
<td>39</td>
<td>37</td>
</tr>
</tbody>
</table>

### Table C3. Posterior inclusion probability of individual predictor derived from BMA on two clusters of sites.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>TSS</th>
<th>PN</th>
<th>NO\textsubscript{x}</th>
<th>NH\textsubscript{4}</th>
<th>DON</th>
<th>FRP</th>
<th>DOP</th>
<th>PP</th>
<th>EC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event_ave_Q</td>
<td>0.93</td>
<td>0.47</td>
<td>0.87</td>
<td>0.49</td>
<td>0.97</td>
<td>0.56</td>
<td>0.21</td>
<td>0.68</td>
<td>0.47</td>
</tr>
<tr>
<td>Event_max_Q</td>
<td>0.73</td>
<td>1.00</td>
<td>0.76</td>
<td>1.00</td>
<td>0.33</td>
<td>0.58</td>
<td>0.14</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>Event_ave_P</td>
<td>0.92</td>
<td>0.92</td>
<td>0.98</td>
<td>0.92</td>
<td>0.32</td>
<td>0.07</td>
<td>0.25</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>Event_max_P</td>
<td>0.24</td>
<td>0.48</td>
<td>0.24</td>
<td>0.29</td>
<td>0.47</td>
<td>0.01</td>
<td>0.10</td>
<td>0.13</td>
<td>0.31</td>
</tr>
<tr>
<td>Event_T</td>
<td>0.07</td>
<td>0.27</td>
<td>0.50</td>
<td>0.21</td>
<td>0.19</td>
<td>0.98</td>
<td>0.16</td>
<td>0.58</td>
<td>0.88</td>
</tr>
<tr>
<td>Event_ND_VI</td>
<td>0.03</td>
<td>0.27</td>
<td>0.09</td>
<td>0.89</td>
<td>0.77</td>
<td>0.55</td>
<td>0.68</td>
<td>0.62</td>
<td>0.35</td>
</tr>
<tr>
<td>Event_SM</td>
<td>0.54</td>
<td>0.21</td>
<td>0.83</td>
<td>0.58</td>
<td>0.99</td>
<td>0.38</td>
<td>0.96</td>
<td>0.21</td>
<td>0.64</td>
</tr>
<tr>
<td>Event_AET</td>
<td>0.13</td>
<td>0.07</td>
<td>0.12</td>
<td>0.90</td>
<td>0.61</td>
<td>1.00</td>
<td>0.68</td>
<td>0.57</td>
<td>0.43</td>
</tr>
<tr>
<td>Ante_Q</td>
<td>0.23</td>
<td>0.18</td>
<td>0.76</td>
<td>0.76</td>
<td>0.15</td>
<td>0.98</td>
<td>0.30</td>
<td>0.37</td>
<td>0.56</td>
</tr>
<tr>
<td>Ante_P</td>
<td>0.20</td>
<td>0.05</td>
<td>0.22</td>
<td>0.06</td>
<td>0.16</td>
<td>0.03</td>
<td>0.25</td>
<td>0.70</td>
<td>0.22</td>
</tr>
<tr>
<td>Ante_ND_VI</td>
<td>0.23</td>
<td>1.00</td>
<td>0.56</td>
<td>1.00</td>
<td>0.86</td>
<td>1.00</td>
<td>0.99</td>
<td>0.89</td>
<td>0.47</td>
</tr>
<tr>
<td>Ante_SM</td>
<td>0.13</td>
<td>0.74</td>
<td>0.38</td>
<td>0.90</td>
<td>0.79</td>
<td>1.00</td>
<td>0.63</td>
<td>0.96</td>
<td>0.83</td>
</tr>
<tr>
<td>Ante_AET</td>
<td>0.09</td>
<td>0.81</td>
<td>0.27</td>
<td>0.31</td>
<td>0.31</td>
<td>0.60</td>
<td>0.20</td>
<td>0.72</td>
<td>0.33</td>
</tr>
<tr>
<td>Post_Q</td>
<td>0.41</td>
<td>0.07</td>
<td>0.21</td>
<td>0.27</td>
<td>0.18</td>
<td>1.00</td>
<td>0.16</td>
<td>0.77</td>
<td>0.66</td>
</tr>
</tbody>
</table>

**Note:** Posterior inclusion probability $\geq 0.8$ in italic.