Author/s:
Liu, S;Ryu, D;Webb, JA;Lintern, A;Guo, D;Waters, D;Western, AW

Title:
A multi-model approach to assessing the impacts of catchment characteristics on spatial water quality in the Great Barrier Reef catchments

Date:
2021-11-01

Citation:
Liu, S., Ryu, D., Webb, J. A., Lintern, A., Guo, D., Waters, D. & Western, A. W. (2021). A multi-model approach to assessing the impacts of catchment characteristics on spatial water quality in the Great Barrier Reef catchments. Environmental Pollution, 288, https://doi.org/10.1016/j.envpol.2021.117337.

Persistent Link:
https://hdl.handle.net/11343/310875

A multi-model approach to assessing the impacts of catchment characteristics on spatial water quality 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.

Corresponding author: S. Liu (shucil@student.unimelb.edu.au)

Abstract

Water quality monitoring programs often collect large amounts of data with limited attention given to the assessment of the dominant drivers of spatial and temporal water quality variations at the catchment scale. This study uses a multi-model approach: a) to identify the influential catchment characteristics affecting spatial variability in water quality; and b) to predict spatial variability in water quality more reliably and robustly. Tropical catchments in the Great Barrier Reef (GBR) area, Australia, were used as a case study. We developed statistical models using 58 catchment characteristics to predict the spatial variability in water quality in 32 GBR catchments. An exhaustive search method coupled with multi-model inference approaches were used to identify important catchment characteristics and predict the spatial variation in water quality across catchments. Bootstrapping and cross-validation approaches were used to assess the uncertainty in identified important factors and robustness of multi-model structure, respectively. The results indicate that water quality variables were generally most influenced by the natural characteristics of catchments (e.g., soil type and annual rainfall), while anthropogenic characteristics (i.e., land use) also showed significant influence on dissolved nutrient species (e.g., NO_X, NH_4 and FRP). The multi-model structures developed in this work were able to predict average event-mean concentration well, with Nash-Sutcliffe coefficient ranging from 0.68 to 0.96. This work provides data-driven evidence for catchment managers, which can help them develop effective water quality management strategies.

Main finding of the work

A multi-model approach indicated that catchment natural characteristics have the greatest impact on water quality overall, but land use has an important impact on dissolved nutrients.

Keywords

Water quality; Catchment characteristics; Statistical model; Multi-model inference; Model averaging

1 Introduction

Fresh water resources are key to agricultural, industrial, and environmental activities (Kundzewicz et al., 2007). However, there is a worldwide concern that water quality in rivers and streams is deteriorating (Booth et al., 2016; Hiscock et al., 2002; Zhao et al., 2019). Elevated levels of pollutants in streams can lead to substantial economic and environmental losses, particularly in coastal and estuarine ecosystems (De Valck et al., 2018; Pickering et al., 1987). To address riverine water quality degradation, improved management in many coastal regions has been implemented in recent decades, e.g., Chesapeake Bay in the US (Noe et al., 2020; Preston et al., 1999; Zhang et al., 2018) and the Great Barrier Reef catchments in Australia (Brodie et al., 2017; Schaffelke et al., 2012; Waterhouse et al., 2017).

The effectiveness of improved management practices heavily depends on a sound understanding of pollutant processes (i.e., source, mobilization and delivery) in catchments (Granger et al., 2010). Riverine water quality is highly variable across space and time and is affected by a wide range of natural and anthropogenic factors in catchments (Chang, 2008; Zhang et al., 2016). Therefore, it is important to understand the key factors affecting spatial and temporal variability of riverine water quality.

Generally, spatial and temporal variation in water quality is driven by three key processes in catchments: 1) sources – the amount of pollutants available within a catchment; 2) mobilization – detachment of pollutants from the source by processes such as erosion and weathering; and 3) delivery – the transport of the detached pollutants to the receiving waters via surface or subsurface flow (Granger et al., 2010). Within a catchment, water quality exhibits substantial temporal variability, including at daily (Brainwood et al., 2004; Meybeck et al., 2012), seasonal (Ouyang et al., 2006; Xiaolong et al., 2010; Xu et al., 2019) and inter-annual (Fabricius et al., 2013; Zhuo et al., 2016) scales. Similarly, riverine water quality can vary markedly between catchments. Natural and anthropogenic characteristics of a catchment can influence the three key catchment processes, and thus lead to large spatial variation in water quality. The relationship between water quality and anthropogenic factors (e.g., land use) has been extensively studied and identified as one of the key controlling factors that affect spatial variation in water quality (Bramley et al., 2002; Jiang et al., 2015; Lintern et al., 2018a; Nash et al., 2011). For instance, land clearing and any associated intensification of agricultural activities post clearing can result in an increase in nutrient loads from fertilizer application, as well as suspended sediment caused by altering surface soil properties (e.g., tillage) and sediment budgets (Blevins et al., 2018; Smith et al., 2013). In addition, the natural conditions of catchments (climate, hydrology, vegetation cover, geology and topography) have a potential impact on the spatial variation in water quality (Donohue et al., 2006; Ye et al., 2009). For instance, catchment geology and soil type determine the source of sediment and naturally-derived nutrients in catchments (Grayson et al., 1997).

In this study, we focus on the spatial variability in long-term average water quality and how these spatial patterns vary with catchment characteristics, acknowledging that temporal dynamics are also important (Guo et al., 2019). Previous studies have highlighted a range of modeling techniques that can be used to explore the relationship between catchment characteristics and water quality spatial responses (Fu et al., 2019; Soranno et al., 1996). However, these studies have certain limitations. Firstly, these studies have mainly focused on a small number of catchment characteristics, mostly hydroclimatic and land uses characteristics (Afed Ullah et al., 2018; Young et al., 1996), and the relative importance of catchment natural and anthropogenic characteristics is rarely evaluated. More importantly, past investigations have often identified a single 'best model' using forward or backward stepwise variable selection to interpret complex processes (Juahir et al., 2011; Sangani et al., 2015). However, because multiple controlling factors can result in a number of plausible models with comparable predictive power (Whittingham et al., 2006), applying a single-best model provides: 1) limited understanding of key drivers affecting spatial variability in water quality; and 2) limited capacity for predicting water quality across space.

Multi-model inference overcomes the limitations of the single best model approach. This approach considers evidence from multiple plausible models by linearly combining these competing models' outputs into new averaged model predictions (Burnham et al., 2004; Cade, 2015; Parrish et al., 2012). Compared to the single model inference approach, the multi-model inference approach provides more robust predictions (Burnham et al., 2002; Poeter et al., 2005; Saft et al., 2016). The multi-model approach has gained increasing popularity in the water resources community in recent years, including ensemble hydrologic forecasting (Duan et al., 2007; Raftery et al., 2005), groundwater hydrology (Chen et al., 2006; Foglia et al., 2013), catchment functioning (Beck et al., 2013; Saft et al., 2016) and climate change impact assessment based on global climate model outputs (Deb et al., 2018; Stoll et al., 2011). Surprisingly, this approach has been rarely used to understand and predict riverine water quality responses. The only previous application of multi-model inference in evaluation of stream water quality is Lintern et al. (2018b), where multi-model inference was used to investigate key factors affecting water quality in temperate catchments. However, in that study, the predictions were still derived from a single best model structure.

To evaluate the utility of multi-model prediction for spatial variability in water quality, this study applies the approach to identify the important factors affecting riverine water quality and to predict water quality response across space. Specifically, the objectives of this study include: (1) identifying the influential catchment characteristics affecting spatial variability in different water quality constituents; and (2) developing a robust statistical modeling approach to predict the average water quality responses, using key catchment characteristics. We focused on the Great Barrier Reef region (Queensland, Australia), due to: (1) its high ecological and economic values (De Valck et al., 2018) that are threatened by water quality deterioration from inland catchments (Brodie et al., 2013b; Waterhouse et al., 2017); and (2) deficient understanding of water quality spatial variation in tropical and subtropical zones (Piazza et al., 2018). We used a long-term event-based water quality monitoring dataset of nine constituents, including: total suspended solids (TSS), particulate nitrogen (PN), oxidized nitrogen (NO_X), ammonium nitrogen (NH_4), dissolved organic nitrogen (DON), filterable reactive phosphorus (FRP), dissolved organic phosphorus (DOP), particulate phosphorus (PP) and electrical conductivity (EC). Monitoring data for all nine constituents were collected from the 32 GBR catchments. Fifty-eight catchment-scale natural and anthropogenic characteristics were investigated to assess their relative effects on water quality spatial variability. These characteristics came from six categories: catchment topography, land cover, land use, geology, climate and hydrology.

2 Materials and Methods

2.1 Study area

The Great Barrier Reef (GBR) is an iconic Australian coral reef ecosystem, with substantial environmental and economic value (De Valck et al., 2018; Whitten et al., 2004).

However, it has experienced a drastic decline in coral cover – 50% for the entire GBR – since 1985 (Brodie et al., 2013a; Kroon et al., 2016). This deterioration is thought to be driven in part by poor riverine water quality discharging from the adjacent catchments (Waterhouse et al., 2017).

The GBR catchments (an approximate total area of 432,000 km^2, Figure 1) consist of six Natural Resource Management (NRM) regions that discharge into the GBR lagoon. These NRM regions can be further divided into 35 major river basins (Pratchett et al., 2019). These catchments are the most climatically, geologically and topographically diverse natural landscapes on the Australian continent (Gilbert et al., 2001). Rainfall in GBR catchments occurs predominantly in the summer wet season (November to April), with an uneven spatial distribution of annual rainfall ranging from less than 500 mm to more than 8000 mm (Figure S1-a) (Davis et al., 2017; Kuhnert et al., 2009). As a result, runoff of most rivers in the GBR catchments exhibits high spatial and seasonal variations, with a few large events in the wet season contributing to the majority of annual runoff and low or no flow dominating during the dry season. Differences in geology (e.g., lithology, Figure S1-c) and biogeography (e.g., bioregion, Figure S1-d) across space also lead to the significant heterogeneity in the natural landscape characteristics in the GBR catchments. In this study, we used water quality monitoring data collected from 32 monitoring sites within the GBR catchments (Figure 1, detailed information of these catchments is in Table S1, in Supplementary Material). The selection of these catchments was based on: 1) sites that have continuous discharge monitoring; and 2) sites that have a water quality monitoring record of greater than 5 years, for the selected water quality variables (see Section 2.2.1), and with a good coverage of samples under both high and low flows. Detailed land use in the areas upstream of the sampling sites is included in Table S2 in Supplementary Material.

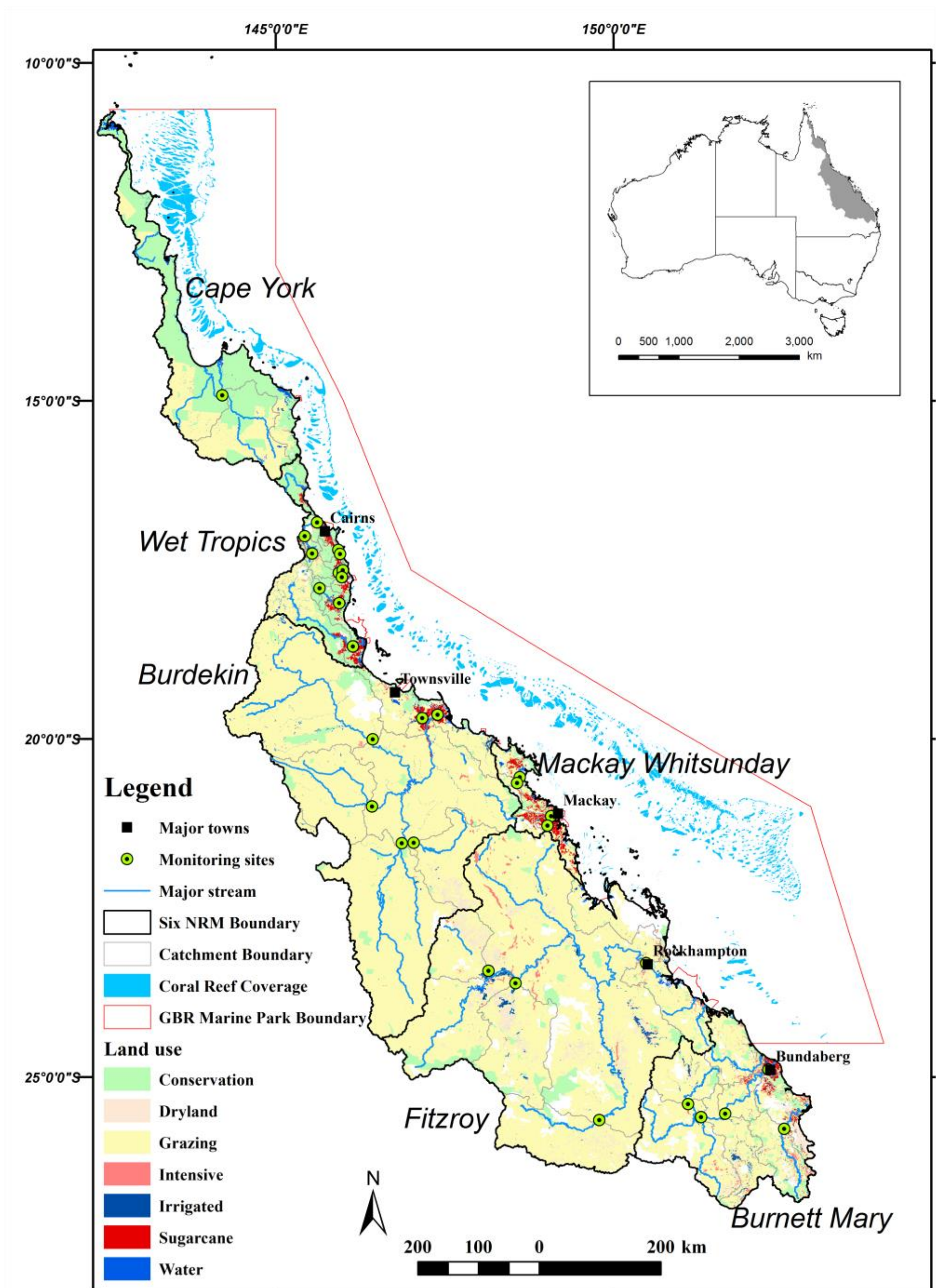

Figure 1. The Great Barrier Reef and GBR catchments, monitoring sites, land uses and the six NRM regions (Data sources - Great Barrier Reef Marine Park Authority (2004); Queensland Government (2017)). Land uses have the following characteristics: (1) Conservation: forest, woodland, savannah, etc for conservation purpose; (2) Dryland: rainfed agriculture but excluding

grazing and sugarcane; (3) Grazing: grazing native 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.

2.2 Data collection and preparation

2.2.1 Water quality data collection

The water quality samples used in this study were obtained from the Great Barrier Reef catchment loads monitoring program, established and maintained by the Queensland Department of Natural Resources, Mines and Energy and Department of Environment and Science (Australian and Queensland governments, 2020; Bartley et al., 2017), who collect and analyze all samples according to American Public Health Association protocols (see details in Table S3) (APHA, 2005). The nine constituents studied covered a range of key indicators of stream sediments, nutrients and salinity, including: total suspended solids (TSS), particulate nitrogen (PN), oxidized nitrogen (NO_X), ammonium nitrogen (NH_4), dissolved organic nitrogen (DON), filterable reactive phosphorus (FRP), dissolved organic phosphorus (DOP), particulate phosphorus (PP) and electrical conductivity (EC), from 2006 to 2016. These nine water quality variables were selected because they pose a significant threat to the coral reef ecosystem (McCloskey et al., 2017), and there was sufficient data available to support our analyses. Analyses of these constituents will provide a useful comprehensive picture on the overall water quality status and its key spatial drivers. It is worth noting that other constituents of interest were not included because they are not monitored (organic carbon) or there was insufficient data for these analyses (pesticides).

Unlike many conventional operational water quality sampling programs, event-based samples were dominant in the water quality monitoring data. This dataset contained both the intensive 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) which were taken under baseflow conditions. The Great Barrier Reef catchment loads monitoring program (more details in Huggins et al. (2018)) was designed to capture the pollutant export during both high and low flow conditions, and was part of the Paddock to Reef Integrated Monitoring, Modelling and Reporting Program (Paddock to Reef program) (Shaw et al., 2014). In that program, the monitoring results were used to calibrate a conceptual catchment water quality model – Source Catchments, which enabled estimation of the annual pollutant loads to the reef lagoon (Orr et al., 2014; Waters et al., 2007; Waters et al., 2013).

2.2.2 Runoff event delineation and event-mean concentration calculation

Individual runoff events were delineated based on the continuous discharge record extracted from the Water Monitoring Information Portal (DNRME, 2018). To identify each event, we used an automated hydrograph analysis tool – *HydRun* (Tang et al., 2017). The start and end points of a specific event were determined by using a local-minimum method through calculating the first derivative of streamflow (first separated from baseflow). The event-mean concentration (EMC) was then calculated for each event that had at least two samples on both the rising limb and falling limbs of the hydrograph. This ensured sufficient samples taken over the runoff hydrograph, and the reliability of derived EMC (Waters et al., 2007). It is worth noting that uncertainty in baseflow estimation has a direct impact on pollutant load estimation (Binns et al., 2018), but a preliminary sensitivity analysis (Figure S2) indicated the influence on calculated event-mean concentration was small (only a 9.15% change in EMC under different filter coefficient β across all catchments). For each event, the EMC of a constituent was calculated as the total load per unit flow volume within the event (Joo et al., 2012), as follows:

$$EMC = \frac{Event\ Load}{Event\ Flow\ Volume} = \frac{\sum_{j=0}^{n} \frac{c_j + c_{j+1}}{2} \times q_{j+1/2} \times t_{j+1/2}}{\sum_{j=0}^{n} q_{j+1/2} \times t_{j+1/2}} \tag{1}$$

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. We identified the concentration at the start and end of all events (c_0 and c_{n+1}) by assuming they were the average concentration of all baseflow samples. The EMCs we developed were essentially a flow-weighted mean concentrations over different runoff events, which allowed us to compare water quality across catchments with contrasting flow regimes (Cooke et al., 2000).

The site-level average of constituent EMCs (i.e., average of all available EMCs for each constituent at each site) were calculated (summary statistics in Table S4) for use in statistical modeling. Prior to analysis, the site-level average EMCs for each constituent were Box-Cox transformed using the *car* package in *R* (Fox et al., 2012), to improve the data symmetry (Box et al., 1964). For each constituent, the Box-Cox parameter λ was estimated individually (λ can be found in Table S5, Supplementary Material). All the transformed variables were normally distributed based on the Shapiro-Wilk's test (Table S5) (Shapiro et al., 1965; Steinman et al., 2018). While the focus of this study is the spatial variability in water quality (between-site), we acknowledge that high temporal variability (within-site) would result in large uncertainty in time-averaged water quality. To test for this, we further decomposed the spatial and temporal variance in EMCs (Figure S3). The spatial component of the variance was derived from the deviation of site-level average EMC and globally-averaged EMC for specific constituents. The temporal component was derived from the deviation of individual EMCs and site-level average EMC for specific constituents. The partitioning of these two components (Figure S3) indicated that spatial variation accounts for the majority of variability for all constituents except PP and DOP. Given this result, we focused the analysis on analyzing spatial differences in water quality across different catchments and did not consider temporal variation further in this study.

2.2.3 Catchment characteristics data collection

Based on previous studies (Chang, 2008; Kleinman et al., 2004; Lintern et al., 2018a; Lintern et al., 2018b), there is a wide range of natural and anthropogenic catchment factors that affect spatial variability in water quality. Among these factors, catchment topography, land cover, land use, geology, climate and hydrology most commonly affect key processes (i.e., source, mobilization and delivery) in catchments (Table S6). Catchment boundaries of the 32 monitoring sites were delineated using the Geofabric tool provided by the Australian Bureau of Meteorology (Bureau of Meteorology, 2012). We obtained the 58 catchment characteristics that represent conditions upstream of the monitoring sites, from publicly available data sets (category and abbreviation are as detailed in Table 1). More detailed description, source and summary statistics can be found in Tables S7 and S8, Supplementary Material. ArcGIS 10.5 was used to extract the catchment average from gridded raster (e.g., catchment rainfall) data or proportion of catchment coverage for polygon (e.g., land use) data. Prior to the analyses, both response variables (i.e., EMCs in Section 2.2.1) and explanatory variables (i.e., 58 catchment characteristics) were standardized to a mean of 0 and standard deviation of 1 to enable comparison between explanatory variable regression coefficients (i.e., relative importance of predictors) (Cade, 2015). Some catchment characteristics were strongly cross-correlated (e.g., Spearman's rank correlation coefficient ρ = -0.77 for stream density and grazing in Figure S3). The collinearity between predictors will likely result in multiple models having similar performance. We will discuss the issue of collinearity between predictors in the analysis of the results (Section 4.1.1).

Table 1. Summary of 58 catchment characteristics and their abbreviations used in this paper. Among these categories, only land use characteristics are considered as anthropogenic influences. Detailed description, source and summary statistics of these characteristics can be found in Tables S7 and S8.

Catchment characteristic		Abbreviation	Catchment characteristic		Abbreviation
Topography	maximum catchment elevation (m)	Max_Elevation	**Land use**	conservation use (%)	PerConservation
	mean catchment elevation (m)	Mean_Elevation		dryland farming (%)	PerDrylandAgri
	catchment area (km^2)	Area		irrigated farming (%)	PerIrrigated
	stream density (km/km^2)	StreamDensity		water (%)	PerWater
	catchment relief ratio	CatRelifRatio		grazing (%)	PerGrazing
	valley bottoms (%)	Valley_Bottoms		sugar cane (%)	PerSugar
	mean catchment slope (%)	MeanCatSlope		intensive use (%)	PerIntensiveUses
	total catchment length (km)	CatLength		cropping (%)	PerCropping
	mean channel slope (%)	ChannelSlope		horticulture (%)	PerHorti
Land cover	grasses (%)	Grasses		catchment fertilized (%)	PerFertilized
	forest (%)	Forest		forestry (%)	PerForestry
	shrubs (%)	Shrubs		urbanized (%)	PerUrbanized
	woodland (%)	Woodland		maximum barrier free flow path length upstream (reservoirs) (km)	UpstreamReser
	bare (%)	Bare		maximum barrier free flow path length	UpstreamDam

				upstream (damwalls) (km)	
	average width of vegetated riparian zone	MeanVegW_m	**Climate**	average annual radiation ($MJ/m^2/day$)	AnnRad
	average catchment riparian zone fragmentation	FraRipaZone		average temperature (°C)	AnnTemp
Geology	catchment underlain by regolith (%)	PerUnconsolidated		minimum temperature of coldest month (°C)	ColdMonthTemp
	igneous rock (%)	PerIgneous		maximum temperature of hottest month (°C)	HotMonthTemp
	mixed igneous and sedimentary (%)	PerMixIgSed		annual average rainfall (mm)	AnnRain
	sedimentary rock (%)	PerSedimentary		average rainfall of the warmest quarter (mm)	WarmQRain
	cation exchange capacity	MeanCaExCap		average rainfall of the coldest quarter (mm)	ColdQRain
	acid sulphate level B (%)	PerAcidS_B		annual average catchment rainfall erosivity (MJ mm/ha hr yr)	Erosivity
	mean TN levels in soil (mg/kg)	MeanTN	**Hydrology**	average annual runoff (mm)	AnnRun
	mean TP levels in soil (mg/kg)	MeanTP		maximum annual runoff (mm)	MaxRun
	Clay Content (%)	Clay_30		pereniality of runoff (%)	RunPereniality
	mean soil erodibility	MeanSoilEro		mean base flow index	Mean_BFI
	soil pH	pH		runoff ratio	Mean_RR
	plant available water capacity (mm)	PAWC		mean number of days where there is no flow annually (days/year)	Cease_to_Flow
	bulk density (mg/m^3)	Bulk_density		mean 7-day low flow (ML/d)	Sevendaylowflow

2.3 Statistical modeling

We first mapped the time-averaged EMC values for the nine constituents across the GBR catchments. In addition, we undertook correlation analysis between each pair of water quality constituents that aimed to evaluate whether common patterns exist for different constituents. To assess the effect of each catchment characteristic, we used multi-model inference to investigate the key controlling catchment characteristics. Once the key catchment characteristics driving spatial differences in riverine water quality were identified, predictive models were built. Multi-model inference in this study involves three steps, namely, (1) identification of multiple plausible models using linear regression, (2) predictions using a model-averaging approach and (3) model assessment. Figure 2 shows the schematic diagram of methods that were used in these three steps. The analyses were performed in MATLAB version R2017b by MathWorks, Inc. (MATLAB and Statistics Toolbox, 2017).

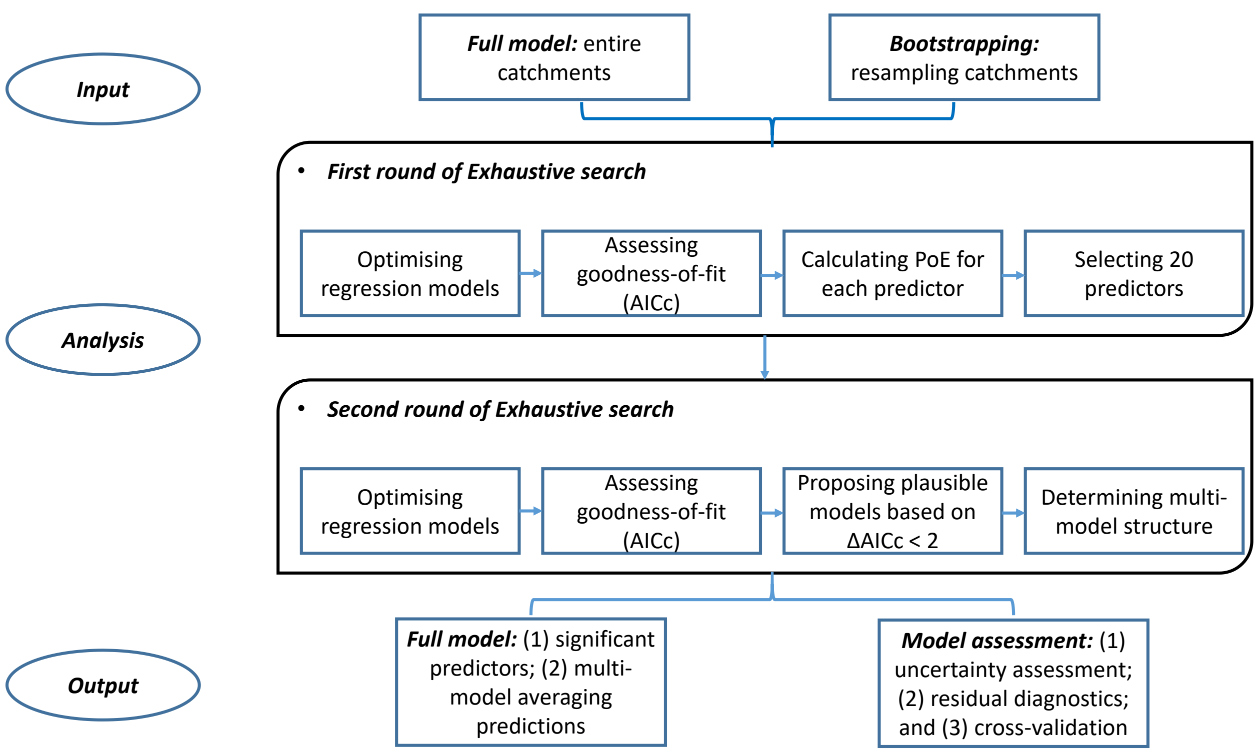

Figure 2. The schematic diagram of overall method used in this study. The two-round exhaustive search (Analysis) was applied to:1) entire catchments as full model results; and 2) resampling of

catchment with replacement (bootstrap) 500 times as uncertainty assessment results for model assessment.

2.3.1 Identification of plausible models

To identify the plausible models to predict average EMCs for each constituent, and reduce the computational burden, we adopted a two-round exhaustive search approach (i.e., two stages in Analysis in Figure 2) (Guyon et al., 2003; Lintern et al., 2018b; May et al., 2011; Saft et al., 2016). Adopting a two-round exhaustive search approach had two advantages: 1) results were more interpretable; and 2) model complexity was controlled by information criterion, thus avoided overfitting (Gregorutti et al., 2017; Vatcheva et al., 2016). We used ordinary least squares to fit the all candidate models in both rounds of the exhaustive search.

In the first round of the Exhaustive Search, we aimed to truncate the number of predictors to a more manageable level to reduce computational overheads. All possible combinations of predictors - up to a maximum of five predictors - were used to construct linear additive models to predict EMCs. The selection of a maximum of five predictors was based on previous exhaustive studies of similar dimensions, to limit the computational demand and to avoid overfitting (Lintern et al., 2018b; Saft et al., 2016).

We assessed and compared all possible models derived from the first round using the Corrected Akaike Information Criterion (AIC_c), which is preferred for small sample applications (Hurvich et al., 1989). The model weights w_i are calculated as follows (Burnham et al., 2002; Ye et al., 2008),

$$w_i = e^{-0.5\Delta AIC_{ci}} \,/ \sum_{n=1}^{N} e^{-0.5\Delta AIC_{cn}} \tag{2}$$

where N is total number of models, and ΔAIC_c is the difference in AIC_c between model i and the minimum AIC_c. The AIC_c quantifies both the model performance and model complexity, which allowed us to compare and rank the candidate models. The model weight w_i is a transformation of AIC_c that provides the evidence/likelihood of i^{th} model being the best model with the minimum information loss (Poeter et al., 2005).

Model weights were used to estimate the relative importance of individual explanatory variables by summing w_i for each model in which that explanatory variable appeared. This is defined as the Proportion of Evidence (PoE) for each predictor (Mohan et al., 2018; Saft et al., 2016). If a predictor appears more frequently in models with small Δ_i (i.e., higher relative performance), then the PoE of that predictor is close to 1. This allowed us to consider the relative importance of individual exploratory variables across all models, and hence identify key predictors. We retained the 20 catchment characteristics with the highest PoE for the second round of the exhaustive search. This truncated number of potential predictors allowed us to explore more possible combinations of explanatory variables without excessive computational requirement.

In the second round of the exhaustive search, we aimed to establish the multi-model modeling averaging structure. We fitted all possible models with up to 10 predictors for each constituent. This ensured that the final models all had a minimum subjects per variable (SPV, ratio of number of observations to number of predictors) larger than 3. Previous studies have shown that this results in adequate estimates of regression coefficients (Austin et al., 2015).

2.3.2 Model averaging

We used the same information criterion (i.e., AIC_c) to assess each model (Bozdogan, 1987) and to identify the multi-model averaging structure to predict average water quality conditions. We identified all plausible models with AIC_c difference (ΔAIC_{ci}), compared to the best model (i.e., the lowest AIC_c). All models with ΔAIC_{ci} less than 2 were identified as plausible models as per previous recommendations (Burnham et al., 2002; Foglia et al., 2013).

The predictions of the individual plausible models were averaged using the weighting coefficients defined by ΔAIC_c, which was the same method we used to calculate the model weights w_i (Equation (2)). The model weights were then used to compute model-averaged predictions, defined as:

$$\bar{\boldsymbol{y}} = \sum_{i=1}^{N} w_i \boldsymbol{y}_{\boldsymbol{i}} \tag{3}$$

where $\boldsymbol{y_i}$ is a vector of the estimated EMCs across all sites, w_i is the weighting coefficients for model i, N is total number of plausible models ($\Delta AIC_{ci} < 2$) for each constituent and $\bar{y}$ is a vector of the weighted predictions.

To compare the effects of predictors selected in the *plausible* models on the response variable, the weighted parameter coefficients ($\bar{\beta}_j$, Equation (4)) can be calculated as follows,

$$\bar{\beta}_j = \sum_{i=1}^{N} w_i \beta_{i,j} \tag{4}$$

where $\beta_{i,j}$ is the fitted model coefficients of predictor j in a given model i, and w_i is the weight of model i. $\bar{\beta}_j$ was only averaged over the models that included the predictor of interest (Lukacs et al., 2010).

2.3.3 Model assessment

The averaged model predictions were evaluated using the Nash-Sutcliffe coefficient (NSE) (Nash et al., 1970). A residual assessment was performed to check: (1) normality of residuals; and (2) heteroscedasticity in residuals (i.e., no clear relationship between residual and predictors that were included in the model averaging structure).

Three additional assessments were performed. Firstly, we used a bootstrap approach to quantify the uncertainty in the relative importance of the catchment characteristics (i.e., PoE). We assessed this uncertainty based on the statistical assumption that individual models of the multi-model ensemble are from the same population and that the central limit theorem applies (Fischer, 2010). The entire set of catchments was sampled randomly with replacement to obtain a bootstrap data set, which was used as input for the identification of plausible models and model-averaging (Sections 2.3.1 and 2.3.2, Analysis in Figure 2). This was repeated 500 times for each constituent to ensure a sufficient number of bootstrap samples to obtain an estimate of the bootstrap distribution (Carpenter et al., 2000). We used two indices to assess the robustness of the selection of important catchment characteristics:

1) the distribution (median and 95% bootstrap confidence interval (CI)) of PoE of all 58 catchment characteristics derived from the first round; and

2) the significance of selected catchment characteristics, which was evaluated using the frequency of catchment characteristics selected in the final model-averaging structures, as well as the distribution of the weighted parameter coefficients across the 500 replicates of bootstrapping (i.e., 95% CI does not cross zero).

Only predictors with selected frequencies larger than 50% and with significant coefficients were designated as important factors to be discussed further. In addition, the model performance derived from the multi-model averaging approach was compared with that of the single best model. Specifically, for each bootstrap resampling run, we compared the difference in each pair of NSEs between the multi-model averaging structure and the single best model (i.e., the model with lowest *AICc* identified in the second round of the exhaustive search). We assigned a 1 if the NSE of the multi-model averaging structure was greater than the single best model, and zero otherwise. The probability of the multi-model approach providing an improved performance was estimated by averaging this statistic over the 500 bootstrap runs.

Secondly, to compare the relative importance of the natural and anthropogenic (i.e., land use) catchment characteristics, the averaged model structure was re-calibrated using only natural catchment characteristics (i.e., land use characteristics were excluded). This allowed us to understand the relative effect of land use when predicting the spatial variation in water quality.

Finally, to evaluate the robustness of model performance, a cross-validation was performed for the full model. Because this assesses the robustness of model performance rather than variable selection, we only focus on the final multi-model structure identified in Section 2.3.1 (i.e., full model). For each constituent, 80% of catchments were randomly selected to calibrate the multi-model structure, with all held-out catchments used for validation of how well the calibrated models performed (i.e., NSE). This was repeated 100 times to obtain an ensemble of model performance on independent datasets and eliminate the effect of sampling.

3 Results

3.1 Spatial pattern of averaged EMCs

The average EMC of the nine constituents showed different spatial patterns across the GBR catchments (Figure S5, Supplementary Material). Generally, the particulate constituents, i.e., TSS, PN and PP, exhibited a similar pattern, where averaged EMCs were lower in the northern region and increased towards the southern sites (Spearman's Rank cross-correlations among these three constituents $\rho > 0.76$, $p < 0.05$, Figure S6). A similar pattern was observed for the dissolved species (e.g., NH_4, DON, FRP, relationships $\rho > 0.45$, $p < 0.05$, Figure S6); however, averaged NO_X showed a contrasting spatial pattern, with sites in the coastal regions (e.g., Wet Tropics, Mackay-Whitsunday, Burdekin and Burnet Mary) having much higher averaged EMCs compared with other sites.

3.2 Multi-model inference on modeling of spatial variability in EMCs

Results presented below will focus mainly on three constituents (i.e., TSS, NO_X and FRP), due to their relatively high risk to the receiving marine environment (Bartley et al., 2017; Waterhouse et al., 2017). Results of the other six constituents are presented in the Supplementary Material.

3.2.1 Key factors identified in *plausible* models

To check the consistency between round 1 and round 2 variable selection, the PoE for each catchment characteristic derived from the full model was compared to the 95% bootstrap CI from the first round of the exhaustive search (Figures 3, S7 and S8), and was found to be within that CI. We note that the 95% CIs were much wider for the catchment characteristics with high PoEs than for those with lower PoEs. This indicated that while there was large uncertainty in relative importance derived from the first round of the exhaustive search, PoE from the two rounds were highly consistent (compare red dots to CI in Figure 3). We infer from this that the key drivers identified and the relative importance of those drivers are largely independent of input data. In addition, the distribution of weighted parameter coefficients and frequency of significant catchment characteristics in the final plausible models from the bootstrapping varied between different constituents (TSS, NO_X and FRP in Figure 4, for the other constituents see Figures S9 and S10, Supplementary Material). Note that significant predictors identified in Figures 4, S8 and S9 have all been identified based on the full models (i.e., after round 2) for individual constituents.

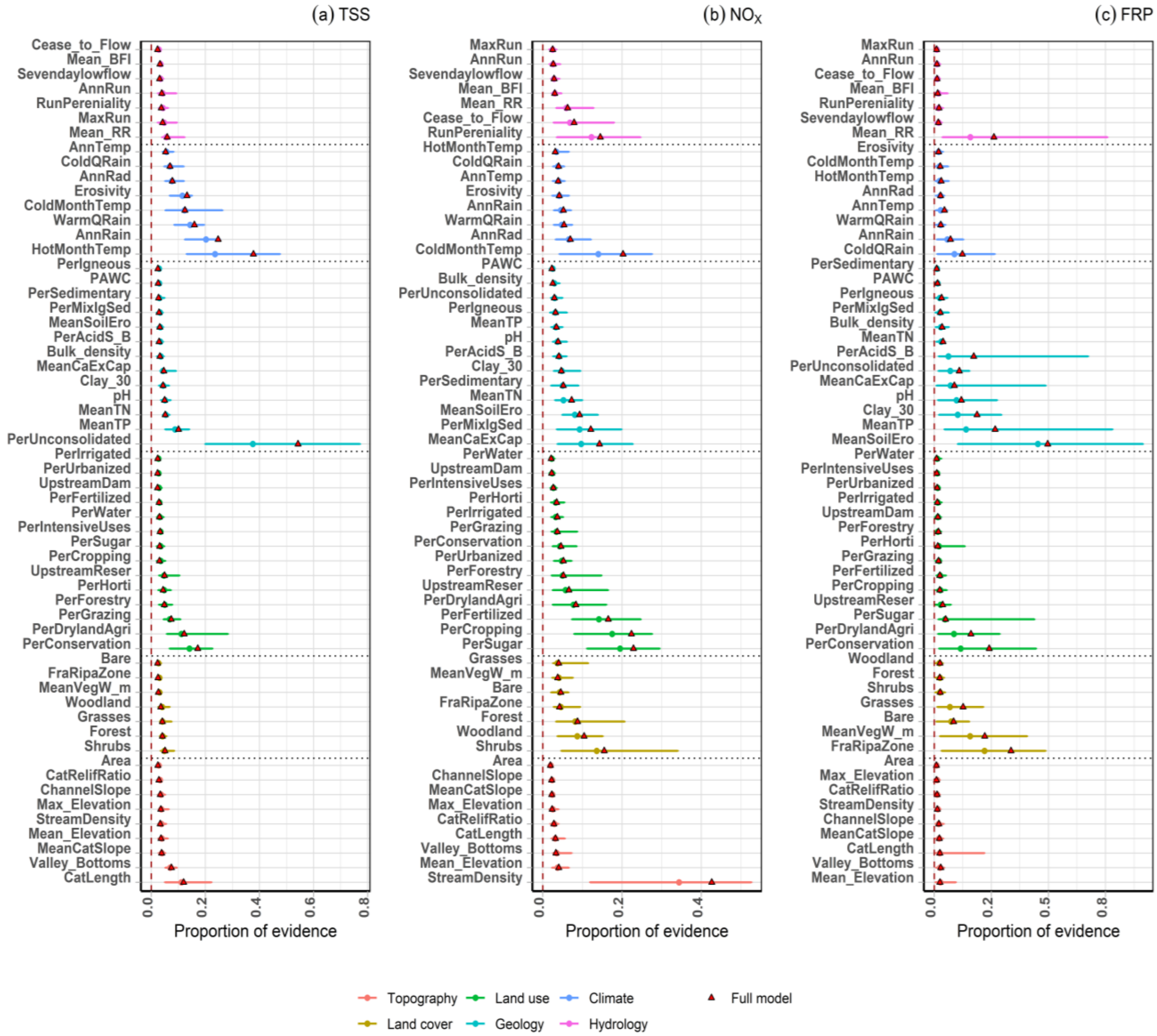

Figure 3. Proportion of Evidence (PoE) from the first-round of exhaustive search derived from full model (red triangle) and 500 runs of bootstrapping (colored dots and bars represent the median and 95% CI, respectively) of each catchment characteristics for: (a) total suspended solids (TSS); (b) oxidized nitrogen (NO_X) and (c) filterable reactive phosphorus (FRP). Catchment characteristics within each category (represented by six different colors) were plotted in descending order of the corresponding median PoE. The definition of abbreviation of each catchment characteristic can be found in Table 1.

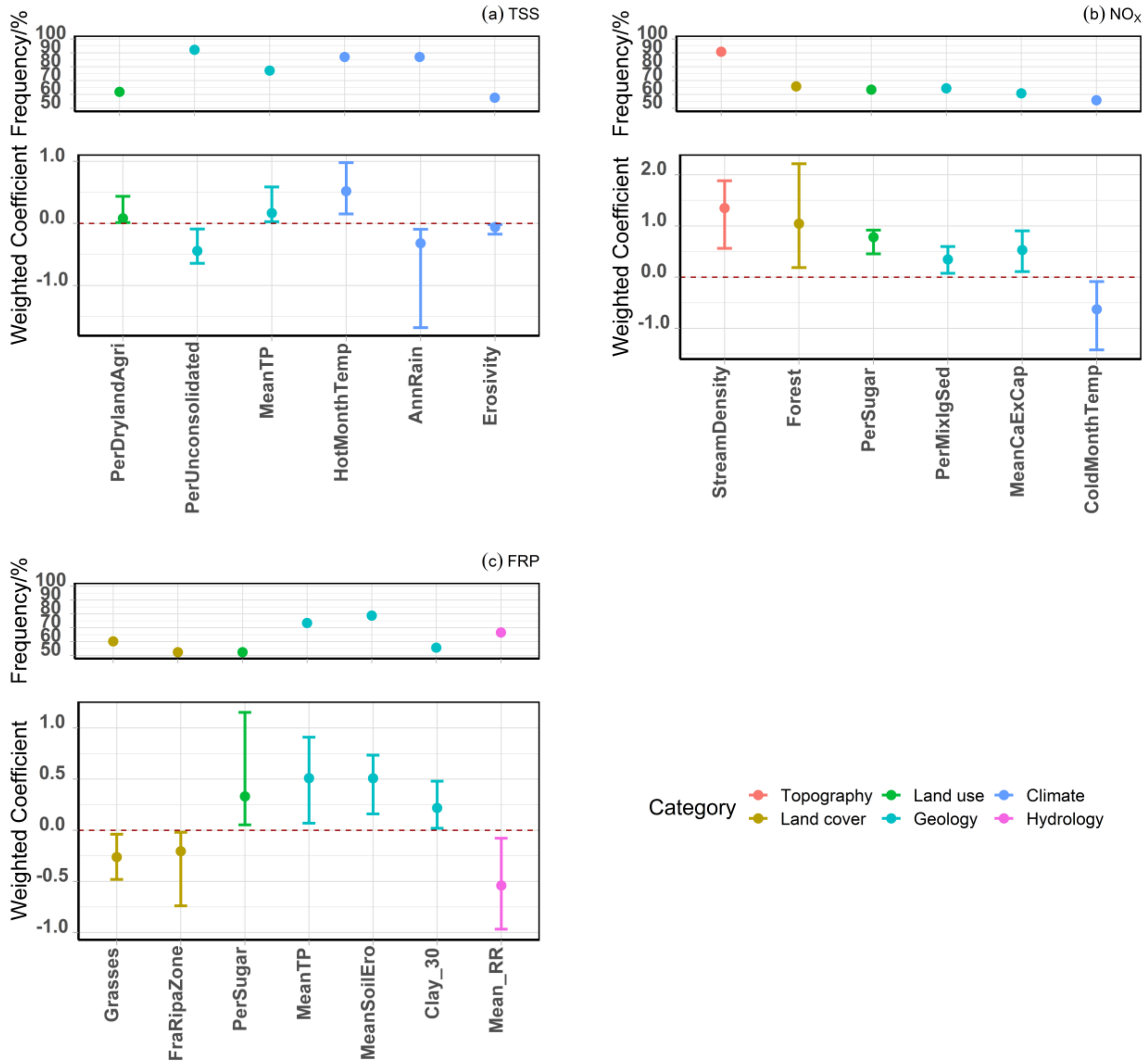

Figure 4. Results from 500 runs of bootstrapping: upper panel indicates the frequency of significant catchment characteristics selected in the final plausible models; lower panel indicates the median (dot) and 95% CI (error bar) of the weighted coefficient for each significant catchment characteristic for: (a) TSS; (b) NO_X and (c) FRP. Different colors represent categories. The definitions of abbreviations for each catchment characteristic can be found in Table 1.

3.2.2 Model-averaging performance

The number of likely models determined from the second round of the exhaustive search for each constituent ranged from 2 for EC to 34 for PN (a detailed summary of each likely model is available in Table S9). NSE ranged from 0.68 for PN to 0.96 for DON (Table 2). In addition, the exclusion of anthropogenic characteristics resulted in a large decrease in NSE for NO_X (-22%, Table 2) but had a small (<10%) to minimal effect on the other constituents. The residual analysis showed no clear heteroscedasticities in model residuals (Figures S11 to S19), and both visual check (Figure S20) and the Shapiro-Wilk's tests (all *p*-value > 0.05, Table 2) indicated that residuals were normally distributed.

Table 2. Model averaging performance (NSE) and weighted predictor coefficients. NSE is calculated based on averaged predictions across all identified plausible models. The third and fourth columns are averaged model performance when considering only natural catchment characteristics (i.e., land use characteristics are excluded), and the change in NSE compared to full model, respectively.

Constituent	NSE - full model	NSE - only natural catchment characteristics included	% change in NSE	Number of plausible models	Shapiro-Wilk's test *p*-value in residuals
TSS	0.75	0.73	-2%	21	0.45
PN	0.68	0.66	-2%	34	0.92
NO_X	0.82	0.64	-22%	3	0.13
NH_4	0.85	0.80	-6%	22	0.32
DON	0.96	0.96	0%	19	0.81
FRP	0.92	0.91	-1%	7	0.71
DOP	0.78	0.78	0%	5	0.29
PP	0.83	0.82	-1%	11	0.36
EC	0.95	0.87	-8%	2	0.67

Note: Shapiro-Wilk's test *p*-value > 0.05 indicates the acceptance of the null hypothesis that data comes from a normal distribution at the 5% significant level.

The averaged models demonstrated a good fit between the predictions and observations, for both full model and bootstrap analysis results (Figure S21, Supplementary Material). In addition, based on the bootstrap results, median NSEs from multi-model structure are higher than single-best models for all nine constituents, and the probability that the multi-model averaging structure performs better than the single best model ranged from 76% to 92% (with an average of 83% across nine constituents) (Table 3), indicating a higher predictive capacity of the multi-model averaging structure.

Table 3. Comparison of model performance between multi-model and single-best model structures for nine constituents, in terms of median NSE and probability of the multi-model structure providing better performance (i.e., higher NSE) across 500 bootstrap subsamples.

Constitute	Median NSE		Probability of the multi-model structure providing better performance
	Multi-model	Single-best model	
TSS	0.51	0.34	0.85
PN	0.32	0.28	0.92
NO_X	0.54	0.52	0.77
NH_4	0.71	0.62	0.89
DON	0.96	0.93	0.85
FRP	0.82	0.81	0.76
DOP	0.45	0.32	0.77
PP	0.50	0.44	0.83
EC	0.56	0.33	0.80

The calibration and validation results for 100 runs of cross-validation are summarized in Table 4. For all constituents, the performance of calibrated models was consistent with the full models. In addition, expect for NO_X, the validation runs showed only a slight decrease in median NSEs compared with calibration results (e.g., less than 0.2 difference in NSE). The larger drop in validation performance for NO_X might be explained by the strong effects of anthropogenic factors affecting the spatial variability in NO_X – we further discuss this impact in Section 4.1.1. Overall, the cross-validation analysis indicates that the multi-model structures are robust and less senstitve to the selection of input data.

Table 4. Comparison of NSEs of the full model and 100 runs of cross-validation models for each constituent.

Constitute	Full model	Median NSE - 100 Cross-validation	
		Calibration	Validation
TSS	0.75	0.74	0.63
PN	0.68	0.65	0.52

NO_X	0.82	0.84	0.47
NH_4	0.85	0.86	0.73
DON	0.96	0.97	0.89
FRP	0.92	0.93	0.73
DOP	0.78	0.74	0.61
PP	0.83	0.80	0.62
EC	0.95	0.91	0.82

4 Discussion

4.1 Influential factors affecting spatial variation in stream water quality

4.1.1 The relative importance of natural and anthropogenic landscape characteristics

For most constituents, natural catchment characteristics were more important than anthropogenic landscape factors for predicting spatial differences in water quality. Except for NO_X, exclusion of anthropogenic predictors in model building did not influence the prediction performance for any constituents markedly (Table 2). It is worth noting that the change in NSE measures the effect of anthropogenic variables relative to the background variation and that this background variation is large for some constituents, such as the particulate species (e.g., TSS, PN and PP), where there is strong natural variation in sediments due to climate, vegetation and soils (Lintern et al., 2018a).

In contrast, land use measures (e.g., proportion of sugarcane farming) had high predictive power for NO_X, NH_4 and FPR, demonstrating that in the GBR catchments specific land use was an important driving factor in spatial variability of EMCs of dissolved inorganic nutrients across catchments. These results are consistent with previous findings that land use changes are related to sources of dissolved forms of nutrients in the GBR catchments (Brodie et al., 2003; Hunter et al., 2008; Mitchell et al., 2009).

In general, natural catchment characteristics exhibited higher predictive power than anthropogenic catchment characteristics. Moreover, it is worth noting that natural catchment characteristics can have a direct impact on anthropogenic drivers, such as agricultural activities which are strongly influenced by climate (Hatfield et al., 2011; Thorburn et al., 2013). This leads to a cross-correlation between natural and anthropogenic drivers which might reduce the anthropogenic signal detected in our analysis because excluding anthropogenic effects might be compensated for through correlation with natural drivers. Indeed, we found that correlated predictors do not necessarily have similar proportions of evidence (e.g., stream density and grazing have -0.77 Spearman's ρ, but proportions of evidence are 0.43 and 0.04 for NO_X, respectively). This is because multi-model inference can handle the collinearity by shrinking the PoE of one of the correlated variables toward zero (Nakagawa et al., 2011; Posch et al., 2020; Walker, 2019). This shrinkage effect leads to a lower weight of the more complex model, if this more complex model only differs from a simpler model by having additional correlated variables. This is because the more complex model receives a higher penalty in AICc. The more complex model would only be favored if the benefit in higher predictive capability outweighs the cost in higher model complexity. This is often not the case as including correlated variables often does not increase the model predictive capacity (Daoud, 2017; Hinne et al., 2020; Kruschke, 2014). Collectively, the results demonstrate that spatial variation in water quality tends to be better explained by catchment natural landscape characteristics.

4.1.2 Climate

Climatic variables had high PoEs and consequently were included in the plausible models for the majority of constituents (Figures 3, S6 and S7). Air temperature was identified as a key factor among these climatic characteristics. It is not surprising since temperature affects almost all physio-chemical processes and biological reactions for nutrients (Huang et al., 2003; Lintern et al., 2018a; Sardans et al., 2008). The maximum temperature of hottest month (*HotMonthTemp*) was identified as significant predictor for TSS, PN and PP, with a strong positive effect on constituent concentration. On the other hand, the average lowest minimum temperature in each year (*ColdMonthTemp*) had a strong negative effect on NO_X. In our study region, climate and hydrology (e.g., discharge) have a clear seasonal pattern with most high flow events in the wet/hot summer (typical time when the EMCs are derived), and few events in the relatively cold/dry winter. High sediment and particulate nutrient concentrations in streams can be expected during hot/wet seasons due to high erosion (Kronvang et al., 1997; Sherriff et al., 2016). During the cold/dry periods, nutrients are more likely to accumulate in the soil, such that there are increases in nutrient availability for the subsequent summer wet season (Edwards et al., 2008; Houser et al., 2010; Pionke et al., 1999). It is worth noting that such seasonal patterns are not evident for the catchments in the Wet Tropics, due to continuous high flow events throughout the year, reducing the likelihood of nutrients building-up in the soil.

Annual rainfall (*AnnRain*) was another controlling factor with a consistently negative impact on TSS, PN, and PP. This result contrasts with previous studies in other catchments (Cavelier et al., 1997; Granger et al., 2010; Perona et al., 1999). The contrast might be due to the high interaction between rainfall and land use/land cover in the GBR catchments. For example, an inverse relationship between annual rainfall and grazing land use (ρ = -0.89, $p < 0.05$, Figure S4) shows that grazing agriculture is mostly conducted in dry catchments (e.g., Fitzroy and Burdekin regions in Figure 1). Vegetation cover also tends to be lower in these catchments. These catchments contribute a large proportion of sediments to the GBR, due to a higher potential for gully and hillslope erosion (Bartley et al., 2018; Kuhnert et al., 2012).

4.1.3 Topography

Catchment topographic characteristics were identified as important factors for dissolved nutrients. Catchment elevation had a significant effect on DON and NH_4, with negative weighted regression coefficients. Meanwhile, there was a negative effect of catchment slope on DOP. These factors are relevant to pollutant mobilization and delivery. In the GBR catchments, low-elevation, small coastal catchments are relatively steeper with a high relief ratio (Figure S4). The topographic features also influence aquifer dynamics that control the fraction of groundwater in surface runoff (Skoulikidis et al., 2006). Lower-lying land is typically associated with higher groundwater contribution, leading to an increase in the concentration of dissolved nutrients in streams (Kratz et al., 1997; McKergow et al., 2003).

The large effects of stream density for NO_X (Figure 3(b)) can be explained by the catchment export processes for inorganic nitrogen, which are strongly controlled by density of stream network (Alexander et al., 2002; Prasad et al., 2005). Catchments with denser stream networks are more likely to have shorter runoff pathways to receiving waters. This may lead to more rapid delivery of NO_X with lower losses from denitrification (Young et al., 1996).

4.1.4 Geology

Catchment geology and soil type were important for all constituents, except for DON. Catchment lithology was one of the most frequently selected predictors. The percentage of catchment underlain by unconsolidated regolith materials (e.g., colluvium and alluvial deposits) (*PerUnconsolidated*) has a negatively weighted regression coefficient for particulate species, i.e., TSS, PN and PP. However, the positive regression coefficients of mixed sedimentary and igneous rock (*PerMixIgSed*) for PP demonstrate that sedimentary and igneous deposits may act as a source of phosphorus. Phosphate minerals can be derived from sedimentary and igneous deposits (mainly in forms of phosphorites and apatite, respectively), and the release of dissolved phosphorus in phosphate is enhanced due to weathering and hydrological transport (Holtan et al., 1988; Pufahl et al., 2017). During the wet season in the GBR catchments, the increased water availability is likely to enhance the erosion and chemical weathering of the bedrock (Bouchard et al., 2000). The weathered material tends to be transported via surface and subsurface runoff, leading to an increase in sediment-bound phosphorus (Hattanji et al., 2004; Pelletier et al., 2011). There is negative feedback between this process and the thickness of regolith (e.g., unconsolidated materials), such that catchments underlain by deeper unconsolidated materials might experience lower levels of particulate and dissolved nutrient species (Strudley et al., 2006).

4.1.5 Land cover

The impact of land cover characteristics in the GBR catchment were not as evident as geological and topographic characteristics. Land cover was only identified as an important predictor for dissolved nutrients (e.g., NO_X, NH_4, DON and FRP). We note that the land cover had a contrasting influence on NO_X (*Forest*, positive) compared to other constituents, e.g., FRP and DON (average catchment riparian zone fragmentation (*FraRipaZone*), negative). Generally, nutrients might be expected to be inversely correlated with riparian vegetation cover due to their reduction via biogeochemical processes (e.g., plant uptake and denitrification) and sedimentation of particulate forms (Johnson et al., 1997; Varanka et al., 2015). However, our results for the effect of riparian vegetation on oxidized nitrogen is the opposite to such findings. In the GBR catchments, a large proportion of NO_X export occurs in the wet regions, where rainfall and subsequent flow events are frequent. In such areas, the riparian trapping effect for NO_X is limited due to shorter residence time and greater contribution from subsurface flow (McKergow et al., 2003; Meynendonckx et al., 2006). One caveat on this result is the data source we used for riparian vegetation. The riparian vegetation information used in the study was measured using

Landsat satellite imagery. A buffer of 100m on both sides of rivers and riverine wetlands was considered to encompass the riparian area (Clark et al., 2015; Queensland Government, 2016). This might introduce bias when the actual width of riparian zone is less than the spatial resolution of the Landsat imagery (i.e., 30 meters) (Woodcock et al., 1994).

4.1.6 Land use

Agricultural activity (e.g., sugarcane farming and horticulture) had a significant and positive relationship with NO_X, NH_4 and FRP, indicating these anthropogenic land uses act as sources of instream pollutants. This is in accordance with previous studies that concluded sediment, nutrients and salts in rivers are likely to be sourced from agricultural activities (Afed Ullah et al., 2018; Liu et al., 2018; Teixeira et al., 2014).

The effect of land use is particularly evident for dissolved inorganic nitrogen (i.e., NO_X) and dissolved phosphorus (i.e., FRP). There is a clear association between dissolved inorganic nitrogen and sugarcane (*PerSugar*) production in the GBR catchments. The majority of the GBR sugarcane is concentrated in the coastal regions of the Wet Tropics, Mackay-Whitsunday, Burdekin and Burnet Mary (Figure 1) (Bainbridge et al., 2009; Hunter et al., 2008; Mitchell et al., 2009). Higher inorganic nitrogen (Figure S5) is linked with application of urea-based nitrogenous fertilizer, which transforms to NH_4 and NO_3 and enters waterways through surface and subsurface pathways (Connolly et al., 2015; Davis et al., 2016; Thorburn et al., 2017).

4.1.7 Hydrology

Catchment hydrological characteristics showed significant negative effects on most dissolved nutrient species (i.e., FRP, NH_4 and DON). This can be linked to the strong dilution effect of high flow on dissolved species, especially in the wet regions (e.g., Wet Tropics) (Orr et al., 2014). Runoff ratio and runoff pereniality (percentage contribution to mean annual discharge by the six driest months of the year) appear as predictors with strong explanatory power for DON and NH_4, respectively. The result is not surprising since catchment hydrology is highly correlated with other catchment characteristics (e.g., ρ = - 0.85 for runoff ratio and grazing agriculture, and ρ = 0.82 for baseflow index and soil TN level, Figure S4). Prathumratana et al. (2008) found that the inter-correlation between catchment runoff, temperature and rainfall could explain the spatial variation of sediments and nutrients in the lower Mekong River. A possible reason is that catchment hydrology is more likely to affect the temporal variation in water quality than spatial variability (Chen et al., 2007). However, the negative effect of the runoff ratio and runoff perennially on DON and NH_4 implies that EMCs of dissolved nitrogen are strongly affected by the temporal changes in runoff volume contributing to the annual runoff due to the dilution effect (Joo et al., 2012).

4.2 Predicting spatial variation in averaged EMCs

The statistical modeling framework proposed in this study was pragmatic and provided a simple approach for assessing average water quality conditions during runoff events across a large tropical region. The weighted prediction derived from the model averaging approach performed well and captured a large proportion of spatial variability in water quality. For all constituents we studied, we had better ability to predict the dissolved nutrient species than particulate pollutants. This is in contrast to an earlier study (Lintern et al., 2018b), which investigated the linkage between spatial variability in average water quality and catchment characteristics in 102 catchments in Victoria, Australia.

The contrasting results may be caused by difference in water quality data and study area characteristics. Firstly, the water quality monitoring data in our study focused on runoff events, rather than the monthly sampling used in Lintern et al. (2018b). We considered the variability in streamflow when developing EMCs using event-based water quality samples. This reduced the uncertainty associated with concentrations of monthly samples, and our samples are more reflective of high flow conditions. This might result in the averaged EMCs of dissolved nutrients being more strongly influenced by catchment natural characteristics (e.g., hydrology for NO_X and geology for FRP) than was the case for Lintern et al. (2018b). Secondly, these two studies also have the following specific differences in key processes driving dissolved and particulate pollutants: 1) land use (e.g., sugarcane) has a clear linkage to dissolved nutrients in our study area but there is no sugar cane and less intensive cropping in the Victorian catchments, and 2) the two study areas have contrasting climates (tropical and temperate), leading to, among other differences, flow regimes and rainfall being more variable in the GBR catchments compared to Victoria. Both differences can lead to higher sensitivity of dissolved nutrients to catchment characteristics and thus the higher predictive capacity of dissolved nutrients in the GBR catchments.

Our modeling was developed using a multi-model inference approach. Figures 3, S6 and S7 demonstrated that there is large uncertainty in relative importance of individual catchment characteristics (as indicated by the wide 95% CIs). Thus, we cannot only rely on the single best model structure when we aim to better understand the effect of catchment characteristics and provide predictions. Our results indicate that: 1) the multi-model approach was better able to consistently identify important factors among a large number of candidate predictors; and 2) compared to the conventional single best model approach (Ekholm et al., 2000; Sangani et al., 2015; Varanka et al., 2015), multi-model inference provided greater predictive ability (Table 3). This is due to the model (variable) selection uncertainty being considered inherently within the multi-model approach (Cade, 2015; Parrish et al., 2012; Ye et al., 2008).

This research was limited to catchments in the north-eastern part of Queensland, Australia. Even though the study catchments featured diverse landscapes (ranging from humid to semiarid tropical catchments) (Bell, 2001; Gilbert et al., 2001), there is still an issue of whether the understanding gained in this research could be transferred to other catchments in Australia and elsewhere. The controlling factors driving the spatial differences in water quality might vary region by region. These differences in catchment processes could be attributed to: 1) different baseline hydroclimatic and ecosystem conditions (Aubert et al., 2013; Nilsson et al., 2008); or 2) different processes (e.g., large floods or prolonged drought events) that lead to trends and changes in water quality conditions (Elchyshyn et al., 2018; Li et al., 2018; Richards et al., 2002). Hydrological conditions could be a more important factor in other catchments (e.g., the Lower Murray River catchments in South Australia) compared with the wetter GBR catchments considered in this study (Kingsford et al., 2011; Mosley et al., 2012). Therefore, the generality of findings from this research should be considered carefully, and the local geographic conditions of any new application should be taken into account. Nevertheless, the results of this study provide an indication of what might be important in other tropical regions such as northeastern coast of Brazil (de Arruda-Santos et al., 2018; Maciel et al., 2015), and the eastern coastal region of India (Damodharan et al., 2012; Govindaraj et al., 2011), where understanding of the drivers of water quality is currently lacking.

4.3 Management implications and future research

Catchment pollutant management requires us to recognize the importance of a wide range of catchment characteristics, especially how the effects of changes (e.g., in land use, land management, land cover or climatic condition under climate change) may alter pollutant export during runoff events. Our analyses show that catchment anthropogenic characteristics are more pertinent to dissolved nutrients (e.g., NO_X, NH_4 and FRP). Therefore, continuous monitoring of changes in land use and management, as well as water quality responses in these intensively used catchments would provide improved insight into managing nutrient sources. This is in line with the current 'best management practice' adopted in the GBR catchments (Lintern et al., 2020; Star et al., 2018; Thorburn et al., 2013). Our results also indicate that spatial variability in particulate pollutants is more directly influenced by a catchment's natural characteristics. This does not necessarily imply that land use is unimportant for particulate constituents. For example, grazing is the dominant land use in the GBR catchments. Hence, departures from this land use (e.g., presence of sugarcane) may be more useful in predicting water quality spatial variation. More detail on the grazing intensity (e.g., number of cattle per area) could potentially be useful in predicting variations in grazing dominated catchments (Smith et al., 2013); however, such detailed survey data are currently unavailable for this region, and so, future efforts are needed to quantify grazing intensity.

Our study focused solely on the averaged water quality conditions for runoff events. While this might be useful for the long-term planning of improved management practice, it does not provide insight into the temporal variation in water quality. Thus, future analyses will investigate the driving factors (e.g., changes in discharge and land cover) that influence the temporal variability in water quality (e.g., deviation of individual EMCs from site-level averaged EMC – the temporal variation component in Figure S3), which is also of great importance (Brodie et al., 2010; Guo et al., 2019). The incorporation of spatial and temporal modeling variability into a single modeling framework would provide a comprehensive understanding of how water quality changes across space and over time (Guo et al., 2020). In addition, it is noted that pesticides, which pose a direct threat to the GBR lagoon ecosystem (Haynes et al., 2000; Lewis et al., 2009), were not included in this study. Further investigation could extend our existing modeling framework to include these emerging chemicals.

5 Conclusions

In this study, a data-driven statistical approach was used to identify the important factors affecting the spatial differences in water quality in the Great Barrier Reef catchments. We used a multi-model approach to identify the influential characteristics for spatial variability in water quality and to make predictions more reliably and robustly, compared to the single best model that has been often used in previous statistical modeling of water quality. Our results indicate that natural catchment characteristics explain more variation in water quality than anthropogenic characteristics, although land use is strongly related to dissolved nutrient concentrations. The models developed were able to predict average event-mean concentrations well (NSE ranging from 0.68 to 0.96). The multi-model averaging framework could be used to identify potential hot spots of water quality concern, at unmonitored locations. This modeling framework also enables evaluation of water quality responses to future changes in climate or land use. With ongoing water quality monitoring data available at multiple GBR catchments, further investigations focusing on temporal variability in water quality are essential to advance our understanding of water quality dynamics.

Acknowledgments

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 provided valuable advice in the preparation of this manuscript. Water quality data (derived site-level averaged EMC) and catchment characteristics data used for the statistical analyses in this paper has been uploaded on the University of Melbourne research data repository (melbourne.figshare.com) with the following DOI: https://doi.org/10.26188/5de8d8f2da817. Sources of these catchment characteristics are provided in Table S7 in Supplementary Material.

References

Afed Ullah, K, Jiang, J, & Wang, P. (2018). Land use impacts on surface water quality by statistical approaches. *Global Journal of Environmental Science and Management, 4*(2), 231-250.

Alexander, Richard B, Elliott, Alexander H, Shankar, Ude, & McBride, Graham B. (2002). Estimating the sources and transport of nutrients in the Waikato River Basin, New Zealand. *Water Resources Research, 38*(12).

APHA. (2005). Standard methods for the examination of water and wastewater. *American Public Health Association (APHA): Washington, DC, USA*.

Aubert, Alice Helene, Gascuel-Odoux, Chantal, & Merot, Philippe. (2013). Annual hysteresis of water quality: A method to analyse the effect of intra-and inter-annual climatic conditions. *Journal of Hydrology, 478*, 29-39.

Austin, Peter C, & Steyerberg, Ewout W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of clinical epidemiology, 68*(6), 627-636.

Australian and Queensland governments. (2020). *Methods, Reef Water Quality Report Card 2019.* Brisbane: Retrieved from https://www.reefplan.qld.gov.au/__data/assets/pdf_file/0019/211672/report-card-2019-methods-combined.pdf.

Bainbridge, Zoe T, Brodie, Jon E, Faithful, John W, Sydes, Damon A, & Lewis, Stephen E. (2009). Identifying the land-based sources of suspended sediments, nutrients and pesticides discharged to the Great Barrier Reef from the Tully–Murray Basin, Queensland, Australia. *Marine and Freshwater Research, 60*(11), 1081-1090.

Bartley, Rebecca, Thompson, Chris, Croke, Jacky, Pietsch, Tim, Baker, Brett, Hughes, Kate, & Kinsey-Henderson, Anne. (2018). Insights into the history and timing of post-European land use disturbance on sedimentation rates in catchments draining to the Great Barrier Reef. *Marine pollution bulletin, 131*, 530-546.

Bartley, Rebecca, Waters, David, Turner, Ryan, Kroon, Frederieke, Garzon-Garcia, Alex, Kuhnert, Petra, Lewis, Stephen, Smith, Rachel, Bainbridge, Zoe, & Olley, Jon. (2017). 2017 Scientific Consensus Statement: land use impacts on the Great Barrier Reef water quality and ecosystem condition, Chapter 2: sources of sediment, nutrients, pesticides and other pollutants to the Great Barrier Reef.

Beck, Hylke E, van Dijk, Albert IJM, Miralles, Diego G, de Jeu, Richard AM, Bruijnzeel, LA Sampurno, McVicar, Tim R, & Schellekens, Jaap. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research, 49*(12), 7843-7863.

Bell, LC. (2001). Establishment of native ecosystems after mining—Australian experience across diverse biogeographic zones. *Ecological Engineering, 17*(2-3), 179-186.

Binns, P, & Waters, D. (2018). *Baseflow separation. Refinement of the Lyne & Hollick baseflow separation methodology using historical water quality data from Great Barrier Reef catchments.* Brisbane.

Blevins, RL, Lal, R, Doran, JW, Langdale, GW, & Frye, WW. (2018). Conservation tillage for erosion control and soil quality *Advances in soil and water conservation* (pp. 51-68): Routledge.

Booth, Derek B, Roy, Allison H, Smith, Benjamin, & Capps, Krista A. (2016). Global perspectives on the urban stream syndrome. *Freshwater Science, 35*(1), 412-420.

Bouchard, Mireille, & Jolicoeur, Serge. (2000). Chemical weathering studies in relation to geomorphological research in southeastern Canada. *Geomorphology, 32*(3-4), 213-238.

Box, George EP, & Cox, David R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological), 26*(2), 211-243.

Bozdogan, Hamparsum. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345-370.

Brainwood, Meredith A, Burgin, Shelley, & Maheshwari, B. (2004). Temporal variations in water quality of farm dams: impacts of land use and water sources. *Agricultural Water Management, 70*(2), 151-175.

Bramley, RGV, & Roth, CH. (2002). Land-use effects on water quality in an intensively managed catchment in the Australian humid tropics. *Marine and Freshwater Research, 53*(5), 931-940.

Brodie, J, Waterhouse, J, Schaffelke, B, Johnson, J, Kroon, F, Thorburn, P, Rolfe, J, Lewis, S, Warne, M, & Fabricius, K. (2013a). *Reef water quality scientific consensus statement 2013*.

Brodie, J, Waterhouse, J, Schaffelke, B, Kroon, F, Thorburn, P, Rolfe, J, Johnson, J, Lewis, S, Devlin, M, Warne, M, & McKenzie, L. (2013b). 2013 Scientific Consensus Statement: Land Use Impacts on the Great Barrier Reef Water Quality and Ecosystem Condition *Department of the Premier and Cabinet, Queensland Government, Brisbane*.

Brodie, Jon E, Lewis, Stephen E, Collier, Catherine J, Wooldridge, Scott, Bainbridge, Zoe T, Waterhouse, Jane, Rasheed, Michael A, Honchin, Carol, Holmes, Glen, & Fabricius, Katharina. (2017). Setting ecologically relevant targets for river pollutant loads to meet marine water quality requirements for the Great Barrier Reef, Australia: A preliminary methodology and analysis. *Ocean & Coastal Management, 143*, 136-147.

Brodie, Jon E, McKergow, Lucy A, Prosser, Ian P, Furnas, Miles, Hughes, Andrew O, & Hunter, Heather. (2003). Sources of sediment and nutrient exports to the Great Barrier Reef World Heritage Area.

Brodie, Jon, Schroeder, Thomas, Rohde, Ken, Faithful, John, Masters, Bronwyn, Dekker, Arnold, Brando, Vittorio, & Maughan, Mirjam. (2010). Dispersal of suspended sediments and nutrients in the Great Barrier Reef lagoon during river-discharge events: conclusions from satellite remote sensing and concurrent flood-plume sampling. *Marine and Freshwater Research, 61*(6), 651-664.

Bureau of Meteorology. (2012). Geofabric V2. Retrieved 02/09/2016 ftp://ftp.bom.gov.au/anon/home/geofabric/

Burnham, Kenneth P, & Anderson, David R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research, 33*(2), 261-304.

Burnham, KP, & Anderson, DR. (2002). *Model selection and multi-model inference: a practical information-theoretic approach.*: Springer-Verlag, New York, USA.

Cade, Brian S. (2015). Model averaging and muddled multimodel inferences. *Ecology, 96*(9), 2370-2382.

Carpenter, James, & Bithell, John. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine, 19*(9), 1141-1164.

Cavelier, Jaime, Jaramillo, María, Solis, Daniel, & de León, Doris. (1997). Water balance and nutrient inputs in bulk precipitation in tropical montane cloud forest in Panama. *Journal of Hydrology, 193*(1-4), 83-96.

Chang, Heejun. (2008). Spatial analysis of water quality trends in the Han River basin, South Korea. *Water research, 42*(13), 3285-3304.

Chen, Hua, Guo, Shenglian, Xu, Chong-yu, & Singh, Vijay P. (2007). Historical temporal trends of hydro-climatic variables and runoff response to climate variability and their relevance in water resource management in the Hanjiang basin. *Journal of Hydrology, 344*(3-4), 171-184.

Chen, Yen-chuan, & Ma, Hwong-wen. (2006). Model comparison for risk assessment: A case study of contaminated groundwater. *Chemosphere, 63*(5), 751-761.

Clark, Andrew, Tindall, Dan, & Healy, Al. (2015). *Riparian vegetation levels in the Queensland Murray-Darling Basin and Bulloo catchments for 2013.* Brisbane.: Retrieved from https://trove.nla.gov.au/version/244291037.

Connolly, NM, Pearson, RG, Loong, D, Maughan, M, & Brodie, J. (2015). Water quality variation along streams with similar agricultural development but contrasting riparian vegetation. *Agriculture, ecosystems & environment, 213*, 11-20.

Cooke, Sandra E, Ahmed, Said M, & MacAlpine, Neil. (2000). *Introductory guide to surface water quality monitoring in agriculture*: Conservation and Development Branch, Alberta Agriculture, Food and Rural ….

Damodharan, Usha, & Reddy, M Vikram. (2012). Impact of Sugar Industrial Treated Effluent on the Growth Factor in Sugarcane--Cuddalore, India. *Journal of Sustainable Bioenergy Systems, 2*(3), 43.

Daoud, Jamal I. (2017). *Multicollinearity and regression analysis.* Paper presented at the Journal of Physics: Conference Series.

Davis, Aaron M, Pearson, Richard G, Brodie, Jon E, & Butler, Barry. (2017). Review and conceptual models of agricultural impacts and water quality in waterways of the Great Barrier Reef catchment area. *Marine and Freshwater Research, 68*(1), 1-19.

Davis, Aaron M, Tink, Michelle, Rohde, Ken, & Brodie, Jon E. (2016). Urea contributions to dissolved 'organic'nitrogen losses from intensive, fertilised agriculture. *Agriculture, ecosystems & environment, 223*, 190-196.

de Arruda-Santos, Roxanny Helen, Schettini, Carlos Augusto França, Yogui, Gilvan Takeshi, Maciel, Daniele Claudino, & Zanardi-Lamardo, Eliete. (2018). Sources and distribution of aromatic hydrocarbons in a tropical marine protected area estuary under influence of sugarcane cultivation. *Science of the Total Environment, 624*, 935-944.

De Valck, Jeremy, & Rolfe, John. (2018). Linking water quality impacts and benefits of ecosystem services in the Great Barrier Reef. *Marine pollution bulletin, 130*, 55-66.

Deb, Proloy, Babel, Mukand S, & Denis, Anjelo Francis. (2018). Multi-GCMs approach for assessing climate change impact on water resources in Thailand. *Modeling Earth Systems and Environment, 4*(2), 825-839.

DNRME. (2018, 07/10/2016). Water Monitoring Information Portal. from https://water-monitoring.information.qld.gov.au/

Donohue, Ian, McGarrigle, Martin L, & Mills, Paul. (2006). Linking catchment characteristics and water chemistry with the ecological status of Irish rivers. *Water research, 40*(1), 91-98.

Duan, Qingyun, Ajami, Newsha K, Gao, Xiaogang, & Sorooshian, Soroosh. (2007). Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources, 30*(5), 1371-1386.

Edwards, AC, & Withers, PJA. (2008). Transport and delivery of suspended solids, nitrogen and phosphorus from various sources to freshwaters in the UK. *Journal of Hydrology, 350*(3), 144-153.

Ekholm, P, Kallio, K, Salo, S, Pietiläinen, O-P, Rekolainen, S, Laine, Y, & Joukola, M. (2000). Relationship between catchment characteristics and nutrient concentrations in an agricultural river system. *Water research, 34*(15), 3709-3716.

Elchyshyn, Leanne, Goyette, Jean-Olivier, Saulnier-Talbot, Émilie, Maranger, Roxane, Nozais, Christian, Solomon, Christopher T, & Gregory-Eaves, Irene. (2018). Quantifying the effects of hydrological changes on long-term water quality trends in temperate reservoirs: insights from a multi-scale, paleolimnological study. *Journal of paleolimnology, 60*(3), 361-379.

Fabricius, Katharina E, De'ath, Glenn, Humphrey, Craig, Zagorskis, Irena, & Schaffelke, Britta. (2013). Intra-annual variation in turbidity in response to terrestrial runoff on near-shore coral reefs of the Great Barrier Reef. *Estuarine, Coastal and Shelf Science, 116*, 57-65.

Fischer, Hans. (2010). *A history of the central limit theorem: from classical to modern probability theory*: Springer Science & Business Media.

Foglia, L, Mehl, SW, Hill, MC, & Burlando, P. (2013). Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, southern Switzerland. *Water Resources Research, 49*(1), 260-282.

Fox, John, Weisberg, Sanford, Adler, Daniel, Bates, Douglas, Baud-Bovy, Gabriel, Ellison, Steve, Firth, David, Friendly, Michael, Gorjanc, Gregor, & Graves, Spencer. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing*.

Fu, Baihua, Merritt, Wendy S, Croke, Barry FW, Weber, Tony, & Jakeman, Anthony J. (2019). A review of catchment-scale water quality and erosion models and a synthesis of future prospects. *Environmental Modelling & Software, 114*, 75-97.

Gilbert, M, & Brodie, JE. (2001). *Population and major land use in the Great Barrier Reef catchment area spatial and temporal trends.* Townsville.

Govindaraj, P, Sindhu, R, Balamurugan, A, & Appunu, C. (2011). Molecular diversity in sugarcane hybrids (Saccharum spp. complex) grown in peninsular and east coast zones of tropical India. *Sugar Tech, 13*(3), 206-213.

Granger, SJ, Bol, R, Anthony, S, Owens, PN, White, SM, & Haygarth, PM. (2010). Towards a holistic classification of diffuse agricultural water pollution from intensively managed grasslands on heavy soils *Advances in Agronomy* (Vol. 105, pp. 83-115): Elsevier.

Grayson, RB, Gippel, CJ, Finlayson, Brian L, & Hart, Barry T. (1997). Catchment-wide impacts on water quality: the use of 'snapshot'sampling during stable flow. *Journal of Hydrology, 199*(1-2), 121-134.

Great Barrier Reef Marine Park Authority. (2004). *Great Barrier Reef Marine Park Zoning*. Retrieved from: http://www.gbrmpa.gov.au/geoportal

Gregorutti, Baptiste, Michel, Bertrand, & Saint-Pierre, Philippe. (2017). Correlation and variable importance in random forests. *Statistics and Computing, 27*(3), 659-678.

Guo, Danlu, Lintern, Anna, Webb, J Angus, Ryu, Dongryeol, Bende-Michl, Ulrike, Liu, Shuci, & Western, Andrew William. (2020). A data-based predictive model for spatiotemporal variability in stream water quality. *Hydrology and Earth System Sciences, 24*(2), 827-847.

Guo, Danlu, Lintern, Anna, Webb, J Angus, Ryu, Dongryeol, Liu, Shuci, Bende-Michl, Ulrike, Leahy, Paul, Wilson, Paul, & Western, AW. (2019). Key Factors Affecting Temporal Variability in Stream Water Quality. *Water Resources Research, 55*(1), 112-129. doi: https://doi.org/10.1029/2018WR023370

Guyon, Isabelle, & Elisseeff, André. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182.

Hatfield, Jerry L, Boote, Kenneth J, Kimball, Bruce A, Ziska, LH, Izaurralde, Roberto C, Ort, Donald, Thomson, Allison M, & Wolfe, D. (2011). Climate impacts on agriculture: implications for crop production. *Agronomy Journal, 103*(2), 351-370.

Hattanji, Tsuyoshi, & Onda, Yuichi. (2004). Coupling of runoff processes and sediment transport in mountainous watersheds underlain by different sedimentary rocks. *Hydrological Processes, 18*(4), 623-636.

Haynes, David, Müller, Jochen, & Carter, Steve. (2000). Pesticide and herbicide residues in sediments and seagrasses from the Great Barrier Reef World Heritage Area and Queensland coast. *Marine pollution bulletin, 41*(7-12), 279-287.

Hinne, Max, Gronau, Quentin F., van den Bergh, Don, & Wagenmakers, Eric-Jan. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science, 3*(2), 200-215. doi: 10.1177/2515245919898657

Hiscock, Kevin M, & Grischek, Thomas. (2002). Attenuation of groundwater pollution by bank filtration. *Journal of Hydrology, 266*(3-4), 139-144.

Holtan, H, Kamp-Nielsen, L, & Stuanes, AO. (1988). Phosphorus in soil, water and sediment: an overview *Phosphorus in freshwater ecosystems* (pp. 19-34): Springer.

Houser, Jeffrey N, & Richardson, William B. (2010). Nitrogen and phosphorus in the Upper Mississippi River: transport, processing, and effects on the river ecosystem. *Hydrobiologia, 640*(1), 71-88.

Huang, XP, Huang, LM, & Yue, WZ. (2003). The characteristics of nutrients and eutrophication in the Pearl River estuary, South China. *Marine pollution bulletin, 47*(1-6), 30-36.

Huggins, R, Wallace, Rohan, Orr, David N, Smith, Rachael A, Taylor, O, King, Olivia C, Gardiner, Richard, Wallace, S, Ferguson, Ben, & Preston, S. (2018). Total suspended solids, nutrient and pesticide loads (2015–2016) for rivers that discharge to the Great Barrier Reef–Great Barrier Reef Catchment Loads Monitoring Program.

Hunter, Heather M, & Walton, Richard S. (2008). Land-use effects on fluxes of suspended sediment, nitrogen and phosphorus from a river catchment of the Great Barrier Reef, Australia. *Journal of Hydrology, 356*(1-2), 131-146.

Hurvich, Clifford M, & Tsai, Chih-Ling. (1989). Regression and time series model selection in small samples. *Biometrika, 76*(2), 297-307.

Jiang, Penghui, Cheng, Liang, Li, Manchun, Zhao, Ruifeng, & Duan, Yuewei. (2015). Impacts of LUCC on soil properties in the riparian zones of desert oasis with remote sensing data: a case study of the middle Heihe River basin, China. *Science of the Total Environment, 506*, 259-271.

Johnson, Lucinda, Richards, Carl, Host, George, & Arthur, John. (1997). Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology, 37*(1), 193-208.

Joo, Marianna, Raymond, Myriam AA, McNeil, Vivienne H, Huggins, Raethea, Turner, Ryan DR, & Choy, Satish. (2012). Estimates of sediment and nutrient loads in 10 major catchments draining to the Great Barrier Reef during 2006–2009. *Marine pollution bulletin, 65*(4-9), 150-166.

Juahir, Hafizan, Zain, Sharifuddin M, Yusoff, Mohd Kamil, Hanidza, TI Tengku, Armi, AS Mohd, Toriman, Mohd Ekhwan, & Mokhtar, Mazlin. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental monitoring and assessment, 173*(1-4), 625-641.

Kingsford, Richard T, Walker, Keith F, Lester, Rebecca E, Young, William J, Fairweather, Peter G, Sammut, Jesmond, & Geddes, Michael C. (2011). A Ramsar wetland in crisis–the Coorong, Lower Lakes and Murray Mouth, Australia. *Marine and Freshwater Research, 62*(3), 255-265.

Kleinman, Peter JA, Sharpley, Andrew N, Veith, Tamie L, Maguire, Rory O, & Vadas, Peter A. (2004). Evaluation of phosphorus transport in surface runoff from packed soil boxes. *Journal of Environmental Quality, 33*(4), 1413-1423.

Kratz, Timothy, Webster, Katherine, Bowser, Carl, Maguson, John, & Benson, Barbara. (1997). The influence of landscape position on lakes in northern Wisconsin. *Freshwater Biology, 37*(1), 209-217.

Kronvang, Brian, Laubel, Anker, & Grant, Ruth. (1997). Suspended sediment and particulate phosphorus transport and delivery pathways in an arable catchment, Gelbaek stream, Denmark. *Hydrological Processes, 11*(6), 627-642.

Kroon, Frederieke J, Thorburn, Peter, Schaffelke, Britta, & Whitten, Stuart. (2016). Towards protecting the Great Barrier Reef from land-based pollution. *Global change biology, 22*(6), 1985-2002.

Kruschke, John. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*: Academic Press.

Kuhnert, Petra M, Henderson, Brent L, Lewis, Stephen E, Bainbridge, Zoe T, Wilkinson, Scott N, & Brodie, Jon E. (2012). Quantifying total suspended sediment export from the Burdekin River catchment using the loads regression estimator tool. *Water Resources Research, 48*(4).

Kuhnert, Petra, Wang, You-Gan, Henderson, Brent, Stewart, Lachlan, & Wilkinson, Scott. (2009). Statistical methods for the estimation of pollutant loads from monitoring data. *Final Project Report. Report to the Marine and Tropical Sciences Research Facility, Reef and Rainforest Research Centre Limited, Cairns*.

Kundzewicz, Zbigniew W, Mata, Luis Jose, Arnell, NW, Doll, Petra, Kabat, Pavel, Jimenez, Blanca, Miller, Kathleen, Oki, Taikan, Zekai, S, & Shiklomanov, Igor. (2007). Freshwater resources and their management.

Lewis, Stephen E, Brodie, Jon E, Bainbridge, Zoë T, Rohde, Ken W, Davis, Aaron M, Masters, Bronwyn L, Maughan, Mirjam, Devlin, Michelle J, Mueller, Jochen F, & Schaffelke, Britta. (2009). Herbicides: a new threat to the Great Barrier Reef. *Environmental Pollution, 157*(8-9), 2470-2484.
Li, Tianyang, Li, Siyue, Liang, Chuan, Bush, Richard T, Xiong, Lihua, & Jiang, Yongjun. (2018). A comparative assessment of Australia's Lower Lakes water quality under extreme drought and post-drought conditions using multivariate statistical techniques. *Journal of Cleaner Production, 190*, 1-11.
Lintern, A, Webb, JA, Ryu, D, Liu, S, Bende-Michl, U, Waters, D, Leahy, P, Wilson, P, & Western, AW. (2018a). Key factors influencing differences in stream water quality across space. *Wiley Interdisciplinary Reviews: Water, 5*(1), e1260.
Lintern, A, Webb, JA, Ryu, D, Liu, S, Waters, D, Leahy, P, Bende-Michl, U, & Western, AW. (2018b). What are the key catchment characteristics affecting spatial differences in riverine water quality? *Water Resources Research, 54*(10), 7252-7272.
Lintern, Anna, McPhillips, Lauren, Winfrey, Brandon, Duncan, Jonathan, & Grady, Caitlin. (2020). Best Management Practices for Diffuse Nutrient Pollution: Wicked Problems Across Urban and Agricultural Watersheds. *Environmental science & technology, 54*(15), 9159-9174.
Liu, S, Ryu, D, Webb, JA, Lintern, A, Waters, D, Guo, Danlu, & Western, AW. (2018). Characterisation of spatial variability in water quality in the Great Barrier Reef catchments using multivariate statistical analysis. *Marine pollution bulletin, 137*, 137-151. doi: https://doi.org/10.1016/j.marpolbul.2018.10.019
Lukacs, Paul M, Burnham, Kenneth P, & Anderson, David R. (2010). Model selection bias and Freedman's paradox. *Annals of the Institute of Statistical Mathematics, 62*(1), 117.
Maciel, Daniele Claudino, de Souza, José Roberto Botelho, Taniguchi, Satie, Bícego, Márcia Caruso, & Zanardi-Lamardo, Eliete. (2015). Sources and distribution of polycyclic aromatic hydrocarbons in a an urbanized tropical estuary and adjacent shelf, Northeast of Brazil. *Marine pollution bulletin, 101*(1), 429-433.
MATLAB and Statistics Toolbox. (2017). The MathWorks, Inc., Natick, Massachusetts, United States.
May, Robert, Dandy, Graeme, & Maier, Holger. (2011). Review of input variable selection methods for artificial neural networks *Artificial neural networks-methodological advances and biomedical applications*: InTech.
McCloskey, G, Waters, D, Baheerathan, R, Darr, S, Dougall, C, Ellis, R, Fentie, B, & Hateley, L. (2017). Modelling reductions of pollutant loads due to improved management practices in the great barrier reef catchments: Updated methodology and results-technical report for reef report card 2015. *Queensland Department of Natural Resources and Mines, Brisbane, Queensland*.
McKergow, Lucy A, Weaver, David M, Prosser, Ian P, Grayson, Rodger B, & Reed, Adrian EG. (2003). Before and after riparian management: sediment and nutrient exports from a small agricultural catchment, Western Australia. *Journal of Hydrology, 270*(3-4), 253-272.
Meybeck, Michel, & Moatar, Florentina. (2012). Daily variability of river concentrations and fluxes: indicators based on the segmentation of the rating curve. *Hydrological Processes, 26*(8), 1188-1207.
Meynendonckx, J, Heuvelmans, G, Muys, Bart, & Feyen, Jan. (2006). Effects of watershed and riparian zone characteristics on nutrient concentrations in the River Scheldt Basin. *Hydrology and Earth System Sciences Discussions, 3*(3), 653-679.
Mitchell, A, Reghenzani, J, Faithful, J, Furnas, M, & Brodie, J. (2009). Relationships between land use and nutrient concentrations in streams draining a 'wet-tropics' catchment in northern Australia. *Marine and Freshwater Research, 60*(11), 1097-1108.
Mohan, Chinchu, Western, Andrew W, Wei, Yongping, & Saft, Margarita. (2018). Predicting groundwater recharge for varying land cover and climate conditions–a global meta-study. *Hydrology and Earth System Sciences, 22*(5), 2689-2703.
Mosley, Luke M, Zammit, Benjamin, Leyden, Emily, Heneker, Theresa M, Hipsey, Matthew R, Skinner, Dominic, & Aldridge, Kane T. (2012). The impact of extreme low flows on the water quality of the Lower Murray River and Lakes (South Australia). *Water resources management, 26*(13), 3923-3946.
Nakagawa, Shinichi, & Freckleton, Robert P. (2011). Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behavioral Ecology and Sociobiology, 65*(1), 103-116.
Nash, J Eamonn, & Sutcliffe, Jonh V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology, 10*(3), 282-290.
Nash, Maliha S, & Chaloud, Deborah J. (2011). Partial least square analyses of landscape and surface water biota associations in the Savannah River Basin. *ISRN Ecology, 2011*.
Nilsson, Christer, & Malm-Renöfält, Birgitta. (2008). Linking flow regime and water quality in rivers: a challenge to adaptive catchment management. *Ecology & society, 13*(2), 18.

Noe, Gregory B, Cashman, Matthew J, Skalak, Katie, Gellis, Allen, Hopkins, Kristina G, Moyer, Doug, Webber, James, Benthem, Adam, Maloney, Kelly, & Brakebill, John. (2020). Sediment dynamics and implications for management: State of the science from long-term research in the Chesapeake Bay watershed, USA. *Wiley Interdisciplinary Reviews: Water, 7*(4), e1454.

Orr, D., Turner, R.D.R., Huggins, R., Vardy, S., & J., Warne. M. St. (2014). *Wet Tropics water quality statistics for high and base flow conditions.* Brisbane.

Ouyang, Y, Nkedi-Kizza, P, Wu, QT, Shinde, D, & Huang, CH. (2006). Assessment of seasonal variations in surface water quality. *Water research, 40*(20), 3800-3810.

Parrish, Mark A, Moradkhani, Hamid, & DeChant, Caleb M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research, 48*(3).

Pelletier, Jon D, & Baker, Victor R. (2011). The role of weathering in the formation of bedrock valleys on Earth and Mars: A numerical modeling investigation. *Journal of Geophysical Research: Planets, 116*(E11).

Perona, E, Bonilla, I, & Mateo, P. (1999). Spatial and temporal changes in water quality in a Spanish river. *Science of the Total Environment, 241*(1-3), 75-90.

Piazza, Gustavo Antonio, Dupas, Rémi, Gascuel-Odoux, Chantal, Grimaldi, Catherine, Pinheiro, Adilson, & Kaufmann, Vander. (2018). Influence of hydroclimatic variations on solute concentration dynamics in nested subtropical catchments with heterogeneous landscapes. *Science of the Total Environment, 635*, 1091-1101.

Pickering, AD, & Pottinger, TG. (1987). Poor water quality suppresses the cortisol response of salmonid fish to handling and confinement. *Journal of fish biology, 30*(3), 363-374.

Pionke, HB, Gburek, WJ, Schnabel, RR, Sharpley, AN, & Elwinger, GF. (1999). Seasonal flow, nutrient concentrations and loading patterns in stream flow draining an agricultural hill-land watershed. *Journal of Hydrology, 220*(1-2), 62-73.

Poeter, Eileen, & Anderson, David. (2005). Multimodel ranking and inference in ground water modeling. *Groundwater, 43*(4), 597-605.

Posch, Konstantin, Arbeiter, Maximilian, & Pilz, Juergen. (2020). A novel Bayesian approach for variable selection in linear regression models. *Computational Statistics & Data Analysis, 144*, 106881.

Prasad, V Krishna, Ortiz, Ariel, Stinner, Ben, Mccartney, David, Parker, Jason, Hudgins, Deana, Hoy, Casey, & Moore, Richard. (2005). Exploring the relationship between hydrologic parameters and nutrient loads using digital elevation model and GIS–a case study from Sugarcreek headwaters, Ohio, USA. *Environmental monitoring and assessment, 110*(1-3), 141-169.

Pratchett, Morgan S, Bridge, Tom CL, Brodie, Jon, Cameron, Darren S, Day, Jon C, Emslie, Michael J, Grech, Alana, Hamann, Mark, Heron, Scott F, & Hoey, Andrew S. (2019). Australia's Great Barrier Reef *World Seas: an Environmental Evaluation* (pp. 333-362): Elsevier.

Prathumratana, Lunchakorn, Sthiannopkao, Suthipong, & Kim, Kyoung Woong. (2008). The relationship of climatic and hydrological parameters to surface water quality in the lower Mekong River. *Environment international, 34*(6), 860-866.

Preston, Stephen D, & Brakebill, John W. (1999). *Application of spatially referenced regression modeling for the evaluation of total nitrogen loading in the Chesapeake Bay watershed*: USGS.

Pufahl, Peir K, & Groat, Lee A. (2017). Sedimentary and igneous phosphate deposits: formation and exploration: an invited paper. *Economic Geology, 112*(3), 483-516.

Queensland Government. (2016). *Great Barrier Reef Report Card 2016*. Brisbane: Retrieved from https://www.reefplan.qld.gov.au/tracking-progress/reef-report-card/2016.

Queensland Government. (2017). The Queensland Land Use Mapping Program (QLUMP). from Department of Science, Information Technology and Innovation https://www.qld.gov.au/environment/land/vegetation/mapping/qlump

Raftery, Adrian E, Gneiting, Tilmann, Balabdaoui, Fadoua, & Polakowski, Michael. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly weather review, 133*(5), 1155-1174.

Richards, R Peter, & Baker, David B. (2002). Trends in water quality in LEASEQ rivers and streams (Northwestern Ohio), 1975–1995. *Journal of Environmental Quality, 31*(1), 90-96.

Saft, Margarita, Peel, Murray C, Western, Andrew W, & Zhang, Lu. (2016). Predicting shifts in rainfall-runoff partitioning during multiyear drought: Roles of dry period and catchment characteristics. *Water Resources Research, 52*(12), 9290-9305.

Sangani, Mohammad Hasani, Amiri, Bahman Jabbarian, Shabani, Afshin Alizadeh, Sakieh, Yousef, & Ashrafi, Sohrab. (2015). Modeling relationships between catchment attributes and river water quality in southern catchments of the Caspian Sea. *Environmental Science and Pollution Research, 22*(7), 4985-5002.

Sardans, Jordi, Peñuelas, Josep, & Estiarte, Marc. (2008). Changes in soil enzymes related to C and N cycle and in soil C and N content under prolonged warming and drought in a Mediterranean shrubland. *Applied Soil Ecology, 39*(2), 223-235.

Schaffelke, Britta, Carleton, John, Skuza, Michele, Zagorskis, Irena, & Furnas, Miles J. (2012). Water quality in the inshore Great Barrier Reef lagoon: Implications for long-term monitoring and management. *Marine pollution bulletin, 65*(4-9), 249-260.

Shapiro, Samuel Sanford, & Wilk, Martin B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611.

Shaw, M, & Silburn, DM. (2014). *Paddock to Reef integrated monitoring, modelling and reporting program, Paddock scale modelling technical report*. Brisbane.

Sherriff, Sophie C, Rowan, John S, Fenton, Owen, Jordan, Philip, Melland, Alice R, Mellander, Per-Erik, & Huallachain, Daire O. (2016). Storm event suspended sediment-discharge hysteresis and controls in agricultural watersheds: implications for watershed scale sediment management. *Environmental science & technology, 50*(4), 1769-1778.

Skoulikidis, N Th, Amaxidis, Y, Bertahas, I, Laschou, S, & Gritzalis, K. (2006). Analysis of factors driving stream water composition and synthesis of management tools—a case study on small/medium Greek catchments. *Science of the Total Environment, 362*(1-3), 205-241. doi: https://10.1016/j.scitotenv.2005.05.018

Smith, Andrew P, Western, Andrew W, & Hannah, Murray C. (2013). Linking water quality trends with land use intensification in dairy farming catchments. *Journal of Hydrology, 476*, 1-12.

Soranno, PA, Hubler, SL, Carpenter, SR, & Lathrop, RC. (1996). Phosphorus loads to surface waters: a simple model to account for spatial pattern of land use. *Ecological Applications, 6*(3), 865-878.

Star, Megan, Rolfe, John, McCosker, Kevin, Smith, Rachael, Ellis, Robin, Waters, David, & Waterhouse, Jane. (2018). Targeting for pollutant reductions in the Great Barrier Reef river catchments. *Environmental Science & Policy, 89*, 365-377.

Steinman, Alan, Hassett, Michael, & Oudsema, Maggie. (2018). Effectiveness of Best Management Practices to Reduce Phosphorus Loading to a Highly Eutrophic Lake. *International journal of environmental research and public health, 15*(10), 2111.

Stoll, Sebastian, Hendricks Franssen, Harrie-Jan, Butts, Michael, & Kinzelbach, Wolfgang KH. (2011). Analysis of the impact of climate change on groundwater related hydrological fluxes: a multi-model approach including different downscaling methods. *Hydrology and Earth System Sciences, 15*(1), 21-38.

Strudley, Mark W, Murray, A Brad, & Haff, PK. (2006). Emergence of pediments, tors, and piedmont junctions from a bedrock weathering–regolith thickness feedback. *Geology, 34*(10), 805-808.

Tang, Weigang, & Carey, Sean K. (2017). HydRun: A MATLAB toolbox for rainfall–runoff analysis. *Hydrological Processes, 31*(15), 2670-2682.

Teixeira, Zara, Teixeira, Heliana, & Marques, João C. (2014). Systematic processes of land use/land cover change to identify relevant driving forces: Implications on water quality. *Science of the Total Environment, 470*, 1320-1335.

Thorburn, Peter J, Biggs, Jody S, Palmer, Jeda, Meier, Elizabeth A, Verburg, Kirsten, & Skocaj, Danielle M. (2017). Prioritizing crop management to increase nitrogen use efficiency in Australian sugarcane crops. *Frontiers in plant science, 8*, 1504.

Thorburn, PJ, Wilkinson, SN, & Silburn, DM. (2013). Water quality in agricultural lands draining to the Great Barrier Reef: a review of causes, management and priorities. *Agriculture, ecosystems & environment, 180*, 4-20.

Varanka, Sanna, Hjort, Jan, & Luoto, Miska. (2015). Geomorphological factors predict water quality in boreal rivers. *Earth Surface Processes and Landforms, 40*(15), 1989-1999.

Vatcheva, Kristina P, Lee, MinJae, McCormick, Joseph B, & Rahbar, Mohammad H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale, Calif.), 6*(2).

Walker, Jeffrey A. (2019). Model-averaged regression coefficients have a straightforward interpretation using causal conditioning. *BioRxiv*, 133785. doi: 10.1101/133785

Waterhouse, J, Schaffelke, B, Bartley, R, Eberhadr, R, Brodie, J, Star, M, Thorburn, P, Rolfe, J, Ronan, M, Taylor, B, & Kroon, F. (2017). *2017 Scientific Consensus Statement: A synthesis of the science of land-based water quality impacts on the Great Barrier Reef.* Brisbane.

Waters, D, & Packett, R. (2007). *Sediment and nutrient generation rates for Queensland rural catchments-an event monitoring program to improve water quality modelling.* Paper presented at the Proceedings of the 5th Australian Stream Management Conference. Australian rivers: making a difference. Charles Sturt University, Thurgoona, New South Wales.

Waters, DK, Carroll, C, Ellis, R, Hateley, L, McCloskey, J, Packett, R, Dougall, C, & Fentie, B. (2013). Modelling reductions of pollutant loads due to improved management practices in the Great Barrier Reef Catchments- Whole of GBR, Volume 1 Department of Natural Resources and Mines: Technical Report (ISBN: 978-1-7423-0999).

Whitten, S, & Bennett, J. (2004). *Economics for Natural Resources Management: Bioeconomic Modeling, Policy Threshold Analysis and Transaction Costs.* Paper presented at the Sixth Annual BioEcon Conference, Kings College, Cambridge.

Whittingham, Mark J, Stephens, Philip A, Bradbury, Richard B, & Freckleton, Robert P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of animal ecology, 75*(5), 1182-1189.

Woodcock, Curtis E, Collins, John B, Gopal, Sucharita, Jakabhazy, Vida D, Li, Xiaowen, Macomber, Scott, Ryherd, Soren, Harward, V Judson, Levitan, Jack, & Wu, Yecheng. (1994). Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model. *Remote Sensing of Environment, 50*(3), 240-254.

Xiaolong, Wang, Jingyi, Han, Ligang, Xu, & Qi, Zhang. (2010). Spatial and seasonal variations of the contamination within water body of the Grand Canal, China. *Environmental Pollution, 158*(5), 1513-1520.

Xu, Guoce, Li, Peng, Lu, Kexin, Tantai, Zhan, Zhang, Jiaxin, Ren, Zongping, Wang, Xiukang, Yu, Kunxia, Shi, Peng, & Cheng, Yuting. (2019). Seasonal changes in water quality and its main influencing factors in the Dan River basin. *Catena, 173*, 131-140.

Ye, Lin, Cai, Qing-hua, Liu, Rui-qiu, & Cao, Ming. (2009). The influence of topography and land use on water quality of Xiangxi River in Three Gorges Reservoir region. *Environmental Geology, 58*(5), 937-942.

Ye, Ming, Meyer, Philip D, & Neuman, Shlomo P. (2008). On model selection criteria in multimodel analysis. *Water Resources Research, 44*(3).

Young, William J, Marston, Frances M, & Davis, Richard J. (1996). Nutrient exports and land use in Australian catchments. *Journal of Environmental Management, 47*(2), 165-183.

Zhang, Qian, & Blomquist, Joel D. (2018). Watershed export of fine sediment, organic carbon, and chlorophyll-a to Chesapeake Bay: spatial and temporal patterns in 1984–2016. *Science of the Total Environment, 619*, 1066-1078. doi: https://doi.org/10.1016/j.scitotenv.2017.10.279

Zhang, Zhaoyong, Juying, Li, Mamat, Zulpiya, & QingFu, Ye. (2016). Sources identification and pollution evaluation of heavy metals in the surface sediments of Bortala River, Northwest China. *Ecotoxicology and Environmental Safety, 126*, 94-101.

Zhao, CS, Yang, Y, Yang, ST, Xiang, H, Wang, F, Chen, X, Zhang, HM, & Yu, Q. (2019). Impact of spatial variations in water quality and hydrological factors on the food-web structure in urban aquatic environments. *Water research*.

Zhuo, La, Mekonnen, Mesfin M, Hoekstra, Arjen Y, & Wada, Yoshihide. (2016). Inter-and intra-annual variation of water footprint of crops and blue water scarcity in the Yellow River basin (1961–2009). *Advances in Water Resources, 87*, 29-41.