Faculty of Engineering and Mathematical Sciences
School of Civil, Environmental and Mining Engineering
The University of Western Australia
M015
35 Stirling Highway
Perth, Western Australia 6009
AUSTRALIA

26 July 2020

Responses to reviewers' comments

We thank the reviewers for their detailed and excellent comments. In response to the comments, we have made substantial changes to the manuscript and believe that it is much improved. Specifically,

- The title and the abstract have been rewritten
- The discussion section was extended with more content of modelling results.
- A new figure was added to the main manuscript and several figures were updated.
- A new figure was added to the supplement

We have provided below our detailed responses to individual comments from the reviewers.

Please don't hesitate to contact me if you require any clarifications.

Yours sincerely

Benya Wang

Comments from reviewer #1 (Thu Huong Thi Hoang):

1. Does the paper address relevant scientific modelling questions within the scope of GMD? Does the paper present a model, advances in modelling science, or a modelling protocol that is suitable for addressing relevant scientific questions within the scope of EGU?

Comments: The paper demonstrated a hybrid model of RM and GBM, which was able to predict and explain accurately the historical missing data as well as different pathways of TN export from two distinct catchments. The scientific question is suitable for the aim of GMD.

Response: Thanks for the comment.

2. Does the paper present novel concepts, ideas, tools, or data?

Comments: In the present paper, a novel hybrid model has been developed and presented by the authors.

Response: Thanks for the comment.

3. Does the paper represent a sufficiently substantial advance in modelling science?

Comments: The novel model developed by the authors could also be applied in different research areas, demonstrating a huge significance of this paper.

Response: Thanks for the comment.

4. Are the methods and assumptions valid and clearly outlined?

Comments: Methods and data analysis were thoroughly presented.

Response: Thanks for the comment.

5. Are the results sufficient to support the interpretations and conclusions?

Comments: Overall, the results and discussion satisfied the major aim of this paper, though several results were not carefully presented. For instance, the RMSE of results in Figure 4 and the meaning of Figure 5 in contribution to the comparison of models that the authors could pay more attention to.

Response: Thanks for your comments. More content about Figure 4 and 5 results were added to the revised manuscript.

Changes in manuscript: Several paragraphs were updated in section 4.2 and please find the update content in the revised manuscript.

6. Is the description sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)? In the case of model description papers, it should in theory be possible for an independent scientist to construct a model that, while not necessarily numerically identical, will produce scientifically equivalent results. Model development papers should be similarly reproducible. For MIP and benchmarking pa-pers, it should be possible for the protocol to be precisely reproduced for an independent model. Descriptions of numerical advances should be precisely reproducible.

Comments: The method as well as code and data provided by the authors could be utilized to reproduce a similar work.

Response: Thanks for the comment.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

Comments: The novelty was shown in comparison with a wide range of previous reports.

Response: Thanks for the comment.

8. Does the title clearly reflect the contents of the paper? The model name and number should be included in papers that deal with only one model.

Comments: In the reviewer's point of view, the title could be improved to be more strength using the result of the discovery of pathways contribution of nutrient, not only prediction of concentration as its current state. The model name and version were provided.

Response: Thanks for the comment. The title is updated to include more result information.

Changes in manuscript: the title has been changed as "ML-SWAN-v1: a hybrid machine learning framework for the concentration prediction and discovery of transport pathways of surface water nutrients"

9. Does the abstract provide a concise and complete summary?

Comments: The content of the abstract is totally good, however, it may better if the authors reduce the introduction of models and add more results of their works.

Response: Thanks for the comment. The abstract is updated to include more result information.

Changes in manuscript: please find the updated abstract below.

Nutrient data from catchments discharging to receiving waters is monitoring for catchment management. However, nutrient data are often sparse in time and space and have non-linear responses to environmental factors, making it difficult to systematically analyse long- and short-term trends and undertake nutrient budgets. To address these challenges, we developed a hybrid machine learning (ML) framework that first separated baseflow and quickflow from total flow, generated data for missing nutrient species, and included pre-generated nutrient data as additional variables in a final simulation of tributary water quality. Hybrid random forest (RF) and gradient boosting machines

(GBM) models were employed and their performance compared with a linear model, a multivariate weighted regression model and stand-alone RF and GBM models that did not pre-generate nutrient data. The six models were used to predict six different nutrients discharged from two study sites in Western Australia: Ellen Brook (small and ephemeral) and the Murray River (large and perennial). Our results showed that the hybrid RF and GBM models had significantly higher accuracy and lower prediction uncertainty for almost all nutrient species across the two sites. The pre-generated nutrient and hydrological data were highlighted as the most important components of the hybrid model. The model results also indicated different hydrological transport pathways for TN export from the two tributary catchments. We demonstrated that the hybrid model provides a flexible method to combine data of varied resolution and quality, and is accurate for the prediction of responses of surface water nutrient concentrations to hydrologic variability.

10. Is the overall presentation well structured and clear?

Comments: The paper was well and logically organized.

Response: Thanks for the comment.

11. Is the language fluent and precise?

Comments: The authors used language precisely with clear meaning.

Response: Thanks for the comment.

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

Comments: Yes, it was accurately shown.

Response: Thanks for the comment.

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

Comments: The manuscript more focuses on modelling techniques, only a few ecological discussion was provided. The manuscript provided some discussion on the source of TN in Ellen Brook and Murray River, however, the discussion should be presented better to avoid subjective idea only reflect author assumption. Discussion should better follow results and references The main idea of the Ecological Modelling is not only a prediction tool but also an explanation of ecological significance and pattern of environmental variables. The paper will be greatly improved if the authors spent more discussion on temporal and spatial patterns of predicted variables. Main question can be - How different b/w patterns of DON, TN, NH-N. How results can be used to explain the source of nutrient, - Transformation of nitrogen (in different forms of NH4-N, TN, DON, etc.) from source to river water bodies. - Solution to improve eutrophication situation in river

Response: Thanks for the comment. It is exactly right that the purpose of environmental modelling is to have a better understanding of the ecosystem. The main purpose of this paper is to introduce the hybrid model and that's why we didn't put main focus on the application of this method before. In the updated manuscript, more contents about the sources of TN in Ellen Brook was added to section 5.1.

Changes in manuscript: Several paragraphs were updated. Please directly refer to section 5.1 for the changes.

14. Are the number and quality of references appropriate?

Comments: Yes.

Response: Thanks for the comment.

15. Is the amount and quality of supplementary material appropriate? For model description papers, authors are strongly encouraged to submit supplementary material containing the model code and a user manual. For development, technical, and benchmarking papers, the submission of code to perform calculations described in the text is strongly encouraged.

Comments: The authors provided sufficiently the code and data of the developed model

Response: Thanks for the comment.

Comments from reviewer #2:

Comments 1: Presentation of the relationships between in the inputs and outputs would be informative. As noted in the discussion (line 445) correlation tests can be used to assess the which parameters would be useful for pre-generation. Potentially a scatter plot matrix, with correlation values on the lower diagonal, would be useful to show the relationships concisely.

Response: Thanks for your comments. We did this analysis before the modelling phase. There are strong relationships between nutrients (see Figure SA.1). This is also one of the assumptions for this hybrid model that if we can pre-generate these nutrients and they can be further used in the final model to predict final target nutrient. In that case, model could have higher accuracy.

Changes in manuscript: please find the Figure SA.2-1 in the supplement document.

Comments 2: It would be useful for the authors to speculate in the discussion, or even undertake the modelling, if training the model on the observed values for the hybrid model, rather than the generated values, would further improve performance. Possibly this would allow the data driven model to fit the underlying relationships more accurately, even if when implemented less accurate predicted values must be used as the inputs. Or possibly the hybrid model is correcting for some of the errors in the generated inputs if there is a systematic bias, and this would degrade performance.

Response: Thanks for the suggestion. The hybrid model should have higher accuracy using observed values instead of the pre-generated data. However, there are only very limited number of data (less than 30 samples) that have completed nutrient species to do this test.

Changes in manuscript: no changes in the manuscript.

Comments 3 (Line 200): Please provide further information on how the 80:20 data split was implemented. Was the last 20% of the time series used for testing, or was a more complex method used? If this was the approach, does this explain why the DON, and DOC for Murray River, was the most important input, because this data is only present in the testing period? As noted in the manuscript, there appears to be a long term trend in the dataset, and only using this data that is more representative of the testing dataset may improve performance? Possibly data from the 1990s does not represent the responses experienced in the more recent testing period? The authors should further comment on these temporal and data length issues.

Response: The main aim of this research is to test the hybrid model, rebuild the historical nutrient data, and explore the short- and long-term nutrient changes. The first step is verifying the model performance. In that case, we randomly divided data into 80:20, built the model, tested on testing data (20%), and then repeated all steps for 30 times to further test model uncertainty and stability (Figure 3). After this, all data points including the testing data were then used to rebuild the historical nutrient data (Figure 4). The different feature importance of DOC and DON in Murray River and Ellen Brook may due to the different nutrient sources and water pathways. We think it is a really good idea to compare model performance on more recent data and old data but it may out of this paper's scope.

Changes in manuscript: The main aim of this research is to test the hybrid model, rebuild the historical nutrient data, and explore the short- and long-term nutrient changes. The first step is verifying the model performance. In that case, the data were randomly divided into 80:20. Different models were built and tuned on the training dataset (80%); the testing dataset (20%) was saved for the final test. To further test model uncertainty and stability, the divided and tested processes were repeated 30 times except WRTDS. After this, all data points including the testing data were then used to rebuild the historical nutrient data.

Comments 4 (Line 316): claims that the generated data enables the hybrid ML to capture long-term trends. Please elaborate on how this is possible, as it is not clear to me. A given set of inputs to the stand alone ML to predict the generated data will give the same result in 1990 as 2018, so how can long term trends be captured? As outlined above, possibly because using inputs only collected in the more recent period is more representative of the testing data?

Response: The stand-alone model and the hybrid model used same dataset to build and test model performance. If the pattern only exists in the recent samples, then both stand-alone and hybrid model should have similar fluctuation. The pre-generated nutrient is the only difference between stand-alone model and hybrid model. If there are long-term trends in the nutrient concentrations (e.g., TN), similar trends should also exist in the components of TN (either DON

or DIN). The pre-generated nutrients emphasise this impact. That is why we suggested the pre-generated data in the hybrid model helped model to capture long-term trends.

Changes in manuscript: The pre-generated nutrient is the only difference between stand-alone model and hybrid model. If there are long-term trends in nutrient concentrations (e.g., TN), similar trends should also exist in the components of TN (either DON or dissolved inorganic nitrogen). The pre-generated nutrients emphasise this impact on the hybrid model. This suggests that the generated nutrient data could provide additional information that allowed the hybrid ML to capture long-term trends; this information was not included in the seasonal components, but existed in the generated nutrient data.

Comments 5: Section 4.2 moves from considering all nutrients to TN only. Please add a justification why this was the case. Is this because TN is the more relevant for the application, or for the sake of brevity only one parameter is analysed in more detail.

Response: Thanks for the comment. TN was selected because TN is the most important and most frequently measured nutrient in many places. In addition, we want this paper to be more concise. That is why only TN was analysed in detail. This hybrid method can be used for other nutrients.

Changes in manuscript: Model performance for six nutrients was compared in last section. To make this section more concise, these six models were then compared in their ability to generate daily TN in Ellen Brook from 01/01/1989 to 16/07/2018 (Figure). TN was selected because TN is the most important and most frequently measured nutrient in many places. This hybrid method can also be used for other nutrients (see results in the supplement document).

Comments 6: Analysis of the relative importance of the different inputs is a valuable addition to the manuscript, which is often not considered in machine learning work. Overall, I consider this work of publishable merit subject to these minor additions and clarifications.

Response: Thanks for the comment.

Minor points

Comments 7: The first sentence of the abstract would benefit from being rewritten. Suggest making this two sentences and reworking, e.g. nutrient data is monitoring, not necessary for monitoring, for example. I found the term “temporal data” confusing, I was assuming this meant lagged data, Qb,t-1, for example. It wasn’t until I found Table 1 that this means Julian day of the year. It is suggested to be more specific, and call this “Julian day” or “seasonal component” or similar.

Response: thanks for the comment. This sentence has been rewritten and separated as two sentences. All “temporal data” has been changed as “seasonal component”.

Changes in manuscript: Nutrient data from catchments discharging to receiving waters is monitoring for catchment management. However, nutrient data are often sparse in time and space and have non-linear responses to environmental factors, making it difficult to systematically analyse long- and short-term trends and undertake nutrient budgets.

Comments 8 (Line 70): Another relevant study for hybrid modelling may be Hunter et al. (2018) https://doi.org/10.5194/hess-22-2987-2018

Response: thanks for sharing this paper. It's interesting to see that both this study and our study found the hybrid model could achieve the best performance than stand-along machine learning model and simple process-based model. We have included this paper in the discussion section.

Changes in manuscript: the following content was added.

Similar results were also found in Hunter et al. (2018) that a hybrid process-driven and ANN model was compared with the stand-alone ANN model and the process-driven model. In their study, the hybrid also achieved the best performance followed by stand-alone ANN. The process-driven benchmark model had significantly lower accuracy than other two models.

Comments 9 (Line 164): remove are from GBM model are generally have less …

Response: Thanks for the comment.

Changes in manuscript: this sentence has been removed from the manuscript.

Comments 10 (Line 214): Can be divided into there stages, add s to stages.

Response: Thanks for the comment.

Changes in manuscript: The overall processes of ML-SWAN can be divided into three stages (Figure 1).

Comments 11 (Line 268): I was confused by this sentence, in that lower RMSE and increased MEF are both improved performance, but opposite patterns. Suggest "overall, the scaled RMSE improved from LM ... And the same pattern was found form MEF …

Response: Thanks for the comment.

Changes in manuscript: Overall, the scaled RMSE reduced from LM, WRTDS, stand-alone ML, and hybrid ML for all nutrients except NH3, and the same pattern was found for MEF in both Ellen Brook and Murray River (Figure 3).

Comments 12 (Line 277): suggest add to the end of this paragraph, "as such, the WRTDS results are not directly comparable to the other methods"

Response: Thanks for the comment.

Changes in manuscript: this sentence has been added to the end of this paragraph.

Comments 13 (Line 389): the temporal data was more useful for the perennial catchment, Murray River. Could this be because seasonal information is captured in other inputs, e.g. Q for the more ephemeral catchment?

Response: thanks for the comment. This could be a possible reason for the lower feature importance of seasonal components in Ellen Brook. But results in Figure 4-f and Figure SA.2-f suggest that TN in Murray River has higher seasonal changes than Ellen Brook. Quickflow and baseflow data in Ellen Brook didn't exhibit significantly higher features importance than Murray River (Figure 6).

Changes in manuscript: This may because seasonal information is captured in other inputs in Ellen Brook (e.g., quickflow and baseflow). But the main reason is the stronger seasonal TN signals in Murray River compared to Ellen Brook. This finding is supported by the generated daily TN data for Murray River (see results in supplement S.2).

Comments 14 (Line 399): Please elaborate further on the method used in this section, it is not clear what has changed. Should "gradually" be "sequentially"?

Response: Thanks for the comment. More contents were added to this section.

Changes in manuscript: The generated nutrient data provided additional information to enhance the hybrid model performance (Figure 3 and Figure 5). To assess the individual impact of a generated nutrient, we did a simple test that sequentially added generated TP, DOC, and DON data to the base GBM (only seasonal components and lagged hydrological data) and evaluated RMSE and MEF for TN prediction. This process was repeated 30 times and the results are presented in Figure 8.

ML-SWAN-v1: a hybrid machine learning framework for the concentration prediction and discovery of transport pathways ~~of~~ of ~~daily~~ surface water nutrients ~~concentrations~~

Benya Wang[1,2], Matthew R. Hipsey[2,3], Carolyn Oldham[1,2]

[1]Department of Civil, Mining and Environmental Engineering, The University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia
[2]Co-operative Research Centre for Water Sensitive Cities, Clayton, Australia
[3]UWA School of Agriculture and Environment, The University of Western Australia, 35 Stirling Highway, Crawley 6009, Australia

Correspondence to: Carolyn Oldham (carolyn.oldham@uwa.edu.au)

Abstract. Nutrient data from catchments discharging to receiving waters is ~~are necessary to~~ monitoring for ~~and manage catchment management water quality. ,~~ ~~H~~however, nutrient data ~~they~~ are often sparse in time and space and have non-linear responses to environmental factors, making it difficult to systematically analyse long- and short-term trends and undertake nutrient budgets. To address these challenges, we developed a hybrid machine learning (ML) framework that first separated baseflow and quickflow from total flow, ~~and then~~ generated data for missing nutrient species ~~, using relationships with hydrological data, rainfall, and temporal data~~, and then utilised ~~. The generated nutrient data were then~~ the ~~included~~ pre-generated nutrient data as additional variables in a final simulation of tributary water quality. Hybrid random forest (RF) and gradient boosting machines (GBM) models were employed and their performance compared with a linear model, a multivariate weighted regression model and stand-alone RF and GBM models that did not pre-generate nutrient data. The six models were used to predict ~~TN, TP, NH3, dissolved organic carbon (DOC), dissolved organic nitrogen (DON), and filterable reactive phosphorus (FRP)~~six different nutrients discharged from two study sites in Western Australia: Ellen Brook (small and ephemeral) and the Murray River (large and perennial). Our results showed that the hybrid RF and GBM models had significantly higher accuracy and lower prediction uncertainty for almost all nutrient species across the two sites. The pre-generated nutrient and hydrological data were highlighted as the most important components of the hybrid model. The model results also indicated different hydrological transport pathways for TN export from two tributary catchments. We demonstrated that the hybrid model provides a flexible method to combine data of varied resolution and quality, and is accurate for the prediction of responses of surface water nutrient concentrations to hydrologic variability.

Formatted: English (Australia)

Formatted: English (Australia)

1 Introduction

Surface water nutrient concentrations have been significantly increased by human activities (Forio et al., 2015) due to urbanisation, waste discharges and agricultural intensification (Liu et al., 2012; Kaiser et al., 2013; Li et al., 2013). Increased

nutrient concentrations and loads in streams alter the biogeochemical functioning and biological community structure in receiving estuaries (Jickells et al., 2014; Staehr et al., 2017), leading to increased incidence of harmful algal blooms (Domingues et al., 2011), anoxia and hypoxia (Li et al., 2016; Testa et al., 2017), and reduced water availability (Heathwaite, 2010). Analysis of tributary water quality data over time is therefore essential to compute incoming nutrient loads, support policy, and plan remediation measures.

Water quality data, however, often have constraints that make it challenging to analyse long- and short-term trends. Firstly, water quality data often have non-linear responses to environmental factors and show high-order interaction effects between different environmental variables. Moreover, nutrients can derive from different sources (point or non-point) in the landscape and are transported to receiving waters through different water pathways subject to varied catchment hydrological conditions and human intervention (Hirsch et al., 2010; Lloyd et al., 2014). Additionally, tributary nutrient datasets often are sparse in both space and time, due to the high cost of fieldwork and chemical analysis (Lamsal et al., 2006; Forio et al., 2015). Historical and current water quality monitoring programmes often use low-frequency sampling regimes on a weekly to monthly basis (Halliday et al., 2012). When monthly averaged concentrations are used, calculated nutrient loads to receiving environments such as lakes or estuaries may be poorly estimated (Cozzi and Giani, 2011), with high variability in the estimated loads (Jordan and Cassidy, 2011). It is also common to have patchy availability of nutrient species data across a study area, and combining datasets from different projects and analytical laboratories make the analysis of long-term trends fraught with uncertainty. For instance, total nitrogen (TN) and total phosphorus (TP) concentrations within catchment outflows, may have been monitored for decades, while dissolved organic nitrogen (DON) and dissolved organic carbon (DOC) concentrations may have only been monitored recently, with the increasing recognition of their ecological importance (Górniak et al., 2002; Petrone et al., 2009; Erlandsson et al., 2011). Given the hydrochemical correlation between different nutrient species and high analytical cost, there are benefits in extracting maximum information from all available nutrient data, particularly relating to changes in water quality over time (Hirsch et al., 2010). In summary, while high-quality nutrient data from tributaries are typically required as input to water quality modelling of receiving waters, the reliability and accuracy of trend analysis of tributary data are frequently restricted by data non-linearity, limited sample size, and variable nutrient availability.

Various models for constructing tributary water quality data have been developed. For example, linear models (LM) and generalised linear models (GLM) that use correlations between concentration (C) and flow (Q), have long played a central role in stream water quality analysis (Cohn et al., 1989; Chanat et al., 2002). Some multivariate regression models have been applied to analyse the long-term trend (Li et al., 2007; Tao et al., 2010; Greening et al., 2014) and seasonal patterns (Giblin et al., 2010; Chen et al., 2012) of surface water nutrients. For example, a weighted regression on time, discharge, and season (WRTDS) was introduced by Hirsch et al. (2010) and has been applied to a number of different water quality studies (Green et al., 2014; Zhang et al., 2016a, 2016b, 2016c).

Meanwhile, data-driven machine learning (ML) methods are increasingly being applied to quantify relationships between soil, water and environmental landscape attributes (Lintern et al., 2018; Wang et al., 2018; Guo et al., 2019). For instance, random forest (RF), a widely used ML method, was used to model the spatial and seasonal variability of nitrate concentrations in streams (Álvarez-Cabria et al., 2016). Gradient boosting machines (GBM) were used to quantify relationships between land-use gradients and the structure and function of stream ecology (Clapcott et al., 2012). In contrast to process-based conceptual models, ML methods simulate relationships purely from the data (Maier et al., 2014) and have the ability to incorporate different types of variables (e.g. numerical or categorised variables); this is particularly suitable for systems with complex variable interactions and non-linear response functions (Povak et al., 2014).

While both process-based and ML models can manage non-linear interactions and be used to explore long-term trends, they both have difficulty in fully extracting important hydrochemical information embedded in nutrient data. Hybrid methods have been proposed for flow forecasting, to enhance the performance of ML models by first using intermediate models to generate additional variables, which are then used for subsequent modelling. For instance, a neural network model is first applied to reconstruct surface ocean partial pressure of carbon dioxide (pCO_2) climatology, which is used as a input into another neural network to predict pCO_2 anomalies with other features (Denvil-Sommer et al., 2019). Similarly, Noori and Kalin (2016) used the soil and water assessment tool (SWAT) to generate baseflow and stormflow, which were then used as inputs to an Artificial neural network (ANN) model to improve daily flow prediction. Both studies used hybrid models to demonstrate that pre-generated variables provided additional information that was crucial to achieving higher prediction accuracy, compared with stand-alone ANN models.

Stream flow integrates water from multiple pathways resulting in a distribution of residence times. Stream nutrients are the product of over-lapping historical inputs and reaction rates, which are spatially distributed and temporally weighted within the catchment (Abbott et al., 2016). Therefore, it is beneficial to understand nutrient transport pathways from the source to receiving waters, to analyse the long- and short-term trends of stream nutrient data; this knowledge will improve management strategies to reduce nutrient transport (Tesoriero et al., 2009; Mellander et al., 2012). In the analysis of the streamflow hydrograph, separating baseflow (the long-term delayed flow from storage) and quickflow (the short-term response to a rainfall event) from total flow is a well-established strategy to better understand transport pathways (Tesoriero et al., 2009). To utilise all available nutrient data and assess the impact of different transport pathways on stream nutrient concentrations, we developed a hybrid machine learning framework for surface water nutrient concentrations (ML-SWAN) that first separated baseflow and quickflow from total flow, then built intermediate models to generate missing nutrient species within the total nutrient pool, using relationships with baseflow, quickflow, rainfall, and seasonal components~~temporal data~~. The generated nutrient data were included as additional variables for a final ML prediction. RF and GBM were employed and their performance compared in stand-alone mode and as a hybrid method.

This study aimed to compare model performance for nutrient concentration prediction, to generate accurate daily nutrient data, to assess the impacts of different water transport pathways on surface water nutrient concentrations, and to present a feasible framework for the application of the hybrid method for surface water nutrient prediction. It was hypothesised that the hybrid RF and hybrid GBM, which used pre-generated daily nutrient concentrations and the separated baseflow and quickflow as additional auxiliary inputs, would take advantage of the complementary strengths of hydrochemical and hydrological relationships to provide the most accurate and reliable nutrient predictions. To test this hypothesis, the hybrid RF and hybrid GBM were compared to a linear model, a multivariate weighted regression model (WRTDS), and stand-alone RF and GBM models, for the prediction of TN, TP, ~~ammonia (~~NH_{4}~~)~~, DOC, DON, and filterable reactive phosphorus (FRP) concentrations, at two different sites under varied hydrological conditions.

2 Model overview

Our modelling goal in this study was to minimise the sum of the overall loss function between the predicted nutrient concentrations and measured nutrient concentrations.

$$\sum_i L(y_i, F(X_i)) \qquad (1)$$

where L is a loss function (e.g., squared error), y_i are measured values, X_i are relevant variables, F is any approximation model, and $F(X_i)$ or $\hat{y}_i$ is the model-predicted value at X_i. The descriptions of different approximation models are described in the following sections.

2.1 Linear model and WRTDS model

Linear models (LM) are the most commonly used tool to describe concentration-discharge (C-Q) relationships (Hirsch et al., 2010). Typically, a *log* transformation is often applied to C and Q data (Crowder et al., 2007; Meybeck & Moatar, 2012; Herndon et al., 2015), with the linear model then described as:

$$\log(C) = \beta_0 + \beta_1 \log(Q) \qquad (2)$$

where C is nutrient concentration and Q is total flow. In this study, the linear model was used as a benchmark for other models. The fitted slope β_0 can represent the base nutrient concentration in a stream, while β_1 can describe relationships between hydrological and biogeochemical data. The WRTDS model was also used (Hirsch et al., 2010) and can be described as:

$$\log(C) = \beta_0 + \beta_1 \log(Q) + \beta_2 JD + \beta_3 \sin(JD) + \beta_4 \cos(JD) + \varepsilon \qquad (3)$$

where JD is the Julian day, ε is unexplained variation. $\beta_2 JD$ is used to represent the long-term trend from year to year, while $\beta_3 Cos(JD)$ and $\beta_4 Sin(JD)$ are used to describe the seasonal variation of stream nutrient concentrations. To calculate the Julian Day for use in Eq 3, days since 01/01/1970 was first calculated and then multiplied by 2π. WRTDS advances the simpler linear

model in two aspects. Firstly, the additional components in the equation allow consideration of seasonal and long-term patterns and make the WRTDS model more able to describe stream nutrient concentrations across the year. Secondly, unlike the linear model whose parameters are constant in time, WRTDS adjusts the parameters in a gradual manner throughout *Q*, *JD* space. This is accomplished by applying a weighted regression for the estimation of *log(C),* where the weights on each observation are based on three distances between the observation (Q_o, JD_o) and the estimation point (Q_i, JD_i) which are 1) the time distance between JD_o and JD_i, 2) the seasonal distance between the time of year at JD_o and the time of year at JD_i, and 3) the discharge distance between $log(Q_o)$ and $log(Q_i)$ (Hirsch et al., 2010; Green et al., 2014). Thus, *log(C)* is considered to be locally linearly related to *log(Q)*, *JD*, *Sin(JD)*, and *Cos(JD)*.

2.2 Random forest and gradient boosting machines

Random forest (RF) and gradient boosting machines (GBM) are ensemble models that combine multiple base learners inside the model to improve the prediction performance (Ishwaran and Kogalur, 2010; Singh et al., 2014). The ensemble methods are the main difference between RF and GBM. In RF, bootstrap aggregating is used to resample the original dataset with replacement. Hence, datasets with partial data are generated and then used to build individual base learners. Unlike bootstrap aggregating, GBM iteratively generates a sequence of base learners, where each successive base learner is built for residual prediction of the preceding base learner (Friedman, 2001, 2002). The probability with which data points are selected for the next training set is not constant and equal for all data points. The selection probability increases for data points that have been mis-estimated in the previous iteration; data points that are difficult to classify would receive higher selection probabilities than easily classified data points (Yang et al., 2010; Erdal & Karakurt, 2013).

For RF and GBM, the most commonly used base learner is a classification and regression tree (CART). A CART model is built to split the dataset into different nodes (Breiman et al., 1984), $\{X_1, x_i^a < v\}$ and $\{X_2, x_j^a \geq v\}$ for numeric variables or $\{X_1, x_i^d = c\}$ and $\{X_2, x_j^d \neq c\}$ for categorised variables, where *i* and *j* are the sample indices, *a* is a numerical variable, *v* is one of the values of *a* variable, *d* is a categorised variable, and *c* is one of the values of *d* variable. To split the dataset at *a* or *d*, the sum of least-square error of the two nodes are calculated for a regression task as:

$$error = \sum_{l=0}^{L}(y_l - \overline{y}_L)^2 + \sum_{r=0}^{R}(y_r - \overline{y}_R)^2 \qquad (4)$$

where y_l and y_r are observations in two split nodes, and $\overline{y_L}$ and $\overline{y_R}$ are the average *y* in that node. The split is chosen among all candidate variables and values to minimise this error. This splitting process is applied from the root to the terminal node, which creates a tree structure for the model (Erdal and Karakurt, 2013). A CART can be used both for classification and regression problems due to this tree-structure (Coops et al., 2011). However, a single CART can sometimes over-simplify variable interactions and may lead to low prediction performance (McBratney et al., 2000; Cutler et al., 2007; Coopersmith et

al., 2010). This drawback can be overcome by the ensemble method that generates many resampled datasets and creates various CARTs to achieve higher accuracy (Breiman, 2001) and more stable results when facing slight variations in input data (Martínez-Rojas et al., 2015). New data input is thus evaluated against all trees created in the ensemble model and each tree votes using the main class or the averaged values in the terminal node. The class with the maximum votes will be used for a classification model, and the averaged predicted value from all trees is used for a regression model (Singh et al., 2014; Belgiu & Drăgu, 2016;). It is found that ensemble methods in RF and GBM can significantly improve the prediction accuracy of CART (Ismail & Mutanga, 2010; Erdal & Karakurt, 2013). ~~GBM models are generally have less decision tree models than RF.~~

Compared to LM and WRTDS models, one drawback of RF and GBM, as well as many ML methods in general, is that there is no specific equation in GBM or RF to directly demonstrate model structures. However, GBM and RF do provide the relative importance of each variable, which is based on the empirical improvement in the loss function due to the split on the specific variable in a tree (Povak et al., 2014; Puissant et al., 2014). The improvement of a certain variable was averaged over all trees, and used as the relative importance of that variable for the final model. This relative importance serves as the key index to understand the model structure of RF and GBM (Makler-Pick et al., 2011).

2.3 Baseflow separation

Total flow is commonly conceptualised as including baseflow and quickflow components (Meshgi et al., 2015). Baseflow separation techniques use the time-series record of streamflow to extract the baseflow and quickflow signatures from the total flow. This can be done by using graphical methods to identify the intersection between baseflow and the rising and falling limbs of the quickflow response (Szilagyi and Parlange, 1998), or filtered methods which process the entire stream hydrograph to derive a baseflow hydrograph (Furey and Gupta, 2001). In this study, the three passes filtered method was applied for baseflow separation; the quickflow was first estimated as described below (Lyne and Hollick, 1979; Nathan and McMahon, 1990), and then baseflow calculated:

$$QF_i = \alpha QF_{i-1} + (Q_i - Q_{i-1})\frac{1+\alpha}{2} \qquad (5)$$

where QF_i is the filtered quickflow for the i^{th} sampling instant, QF_{i-1} is the filtered quickflow for the previous sampling instant to i, α is the filter parameter with a value of 0.925 for daily flow as recommended by Nathan and McMahon (1990). Baseflow is then calculated as $BF = Q - QF$.

2.4 Performance evaluation metrics

In this study, the root mean squared error (RMSE) and the Nash–Sutcliffe model efficiency coefficient (MEF) were used to compare model performance. The RMSE is a measure of overall error between the predicted and measured data and returns an error value with the same units as the data, which is given by the following equation:

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \qquad (6)$$

where n is the number of data samples. RMSE varies from 0 to + ∞, and a perfect model would have RMSE of 0. The MEF is a dimensionless ‘‘goodness-of-fit” measure which can vary from -∞ to 1, with a value of 1 indicating a perfect fit and 0 indicating that the mean of the measured values performs as well as the model. The MEF can be calculated as:

$$MEF = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \qquad (7)$$

where $\bar{y}_i$ is the mean of the measured values. Note that the predicted and measured nutrient values were normalised to [0, 1] in this study to compare model performance across different nutrient species.

2.5 Overview of modelling processes

The main aims of this research is to test the hybrid model, rebuild the historical nutrient data, and explore the short- and long-term nutrient changes. The first step is verifying the model performance. In that case, the data were randomly divided into 80:20. Different models were built and tuned on the training dataset (80%), and test on the~~; the~~ testing dataset (20%)~~was saved for the final test~~. To~~-~~ further test model uncertainty and stability, the divided and tested processes were repeated 30 times except WRTDS. After this, all data points including the testing data were then used to rebuild the historical nutrient data. ~~The data points were divided into the training dataset (80%) and the testing dataset (20%). Different models were built and tuned on the training dataset; the testing dataset was saved for the final test.~~ Five-fold cross-validation (CV) was done on the training dataset to tune the model parameters. Leave-one-out cross-validation (LOOCV) was used in WRTDS to predict daily nutrient concentrations; LOOCV is the default cross-validation method in the EGRET package. In that method, one data point was excluded at a time from the whole dataset, all other data points were used to build the model, and the excluded point was used for testing the model performance. This process was repeated for all data points. The performance of all six methods (LM, WRTDS, RF, GBM, hybrid RF, and hybrid GBM) was evaluated on the testing dataset. ~~WRTDS was run through the EGRET (Exploration and Graphics for RivEr Trends) package (Hirsch and De Cicco, 2015) in R to produce daily concentrations for six nutrient species (TP, TN, DON, DOC, NH_3, and FRP). The default settings specified by the user guide (Hirsch and De Cicco, 2015) were used. RF and GBM models were built through the H_2O package in R.~~

~~To assess the prediction uncertainty of the six models, the divided and tested processes were repeated 30 times (the process within the dashed line in Figure 1) except WRTDS. Leave-one-out cross-validation (LOOCV) was used in WRTDS to predict daily nutrient concentrations; LOOCV is the default cross-validation method in the EGRET package. In that method, one data point was excluded at a time from the whole dataset, all other data points were used to build the model, and the excluded point was used for testing the model performance. This process was repeated for all data points.~~ WRTDS was run through the EGRET (Exploration and Graphics for RivEr Trends) package (Hirsch and De Cicco, 2015) in R to produce daily concentrations for six nutrient species (TP, TN, DON, DOC, NH_4, and FRP). The default settings specified by the user guide (Hirsch and De Cicco, 2015) were used. RF and GBM models were built through the H_2O package in R.

The overall processes of ML-SWAN can be divided into three stages (Figure 1). The first stage was baseflow separation using the EcoHydRology package (Fuka et al., 2018). The generated baseflow, quickflow, total flow and rainfall were further transformed into lagged data (the averaged values over the previous 3, 7 and 15 days) to capture any short-term impacts of different water pathways and rainfall on stream nutrients. *JD*, *Cos(JD)*, and *Sin(JD)* were also calculated for RF and GBM to include seasonal and long-term impacts. A description of all the used variables is given in Table 1.

The second stage of ML-SWAN was to build intermediate RF and GBM models that generated daily nutrient concentrations. For the intermediate RF and GBM models, only lagged hydrological data (including total flow, baseflow, and quickflow), lagged rainfall, seasonal components~~temporal data~~ on the training dataset were used. Nutrients were not used as a predictor in the intermediate model. Note that, in this study TP, TN, DOC, and DON were selected to be generated in the second step. If one nutrient was considered as the final target, the other three nutrients were used to generate daily data. For instance, daily TP, DOC, and DON were generated as additional variables to predict TN. In that case, the missing TP, DOC, and DON were generated by the intermediate model for the training dataset and the testing dataset. Daily TN, TP, DOC, and DON data were generated and used for the final predictions. These nutrients were selected since they may be generated from similar sources, or are important components of the total nutrient load. For instance, DOC and DON may both be generated from DOM (Seitzinger et al., 2002; Bernal et al., 2005; Filep & Rékási, 2011). In the catchments studied here, DON can be a dominant component of TN (Nice et al., 2009; Petrone, 2010; Bourke et al., 2015). The selection of DOC and DON for pre-generation may not necessarily be appropriate for other catchments. The selection of nutrients for pre-generation depends on data availability in the dataset. The use of different species of the same nutrients (N or P) can generally improve model performance.

The third stage of ML-SWAN built an additional hybrid model using the training data, which has generated nutrient data by the intermediate models, lagged hydrological data, lagged rainfall data and seasonal components~~temporal data~~. Note that at this stage, the only difference between stand-alone ML and hybrid ML methods was that stand-alone ML did not use pre-generated daily nutrient data.

3. Site overview

To test the generalisability of the hybrid framework, two sites in Western Australia (Ellen Brook and Murray River) were selected as study areas. Ellen Brook and Murray River are key tributaries for Swan-Canning Estuary and Peel-Harvey Estuary (Figure 2), respectively, and have different hydrological conditions. The Swan-Canning Estuary is located adjacent to the Perth metropolitan area, with an area of approximately 40 km^2. The catchment comprises 30 catchments which drain approximately 2090 km^2 (Kelsey et al., 2010). Ellen Brook is the largest sub-catchment in the Swan-Canning catchment, comprising 34% (716 km^2) of the total catchment area. Ellen Brook is an ephemeral river with no flow recorded during summer and early autumn months (Table 2). The dominant land use in Ellen Brook is agricultural and grazing land. Ellen Brook is one of the highest contributors of TN and TP to the Swan-Canning Estuary (Swan River Trust, 2009). Bassendean sands and duplex Yanga (sand over clay) soils dominate the Ellen Brook catchment. Bassendean sands have very low phosphorus retention indices (PRI), while Yanga soils have low PRI in their upper horizon and become waterlogged in winter, promoting the release of retained nutrients to the stream (Kelsey et al., 2010).

The Peel–Harvey Estuary is located approximately 75 km south of the Swan-Canning Estuary and the Serpentine, Murray and Harvey Rivers drain into the estuary (Figure 2). The total catchment area of the estuary is approximately 11930 km^2. The Murray River catchment is dominated by deep grey sands, loams clay and peats (Ruibal-Conti et al., 2013), agricultural land use and natural reserves, and it contributes about 40% of annual TN loads and 7% of annual TP loads to the estuary (Kelsey et al., 2011).

Both Swan-Canning Estuary and Peel-Harvey Estuary experience a Mediterranean climate with cool, wet winters (June–August) and hot, dry summers (December–March). The long-term average annual rainfall varies from 1300 mm on the coast, to 800 mm in the southeast of the catchment area (1975–2009, Bureau of Meteorology station), and about 90% of the rain falls between April and October. Sample size and the first measurement year of six nutrients species are listed for the two study sites in Table 3. TN, TP, NH_4, and FRP have been monitored for decades, while DOC and DON have only been measured in recent years, with limited sample size. Several historical nutrient datasets were combined but significant changes occurred in water sampling devices and analytical instrumentation over the past decades. These changes can increase the complexity of nutrient data. For instance, auto-samplers sampled any time regardless of weather conditions (e.g., during the rainfall) while grab samples were typically collected under fine weather conditions due to safety concerns.

4. Results

4.1 Comparison of prediction accuracy between six methods

Overall, the scaled RMSE reduced from LM, WRTDS, stand-alone ML, and hybrid ML for all nutrients except NH_4, and the same~~opposite~~ pattern was found for MEF in both Ellen Brook and Murray River (Figure 3). The linear model had the worst performance: the scaled RMSE was significantly higher and MEF was significantly lower than the other models, for all six nutrients and across both sites. WRTDS generally had higher RMSE and lower MEF than the stand-alone ML, although it achieved similar results to stand-alone ML for FRP and NH_4 at both sites. LOOCV was used in WRTDS and only one set of results were generated, compared to 30 RMSE and MEF values for other methods. This results in a shortened line for WRTDS in Figure 3, instead of the inter-quartile ranges (IQR = 75^{th} percentile - 25^{th} percentile) presented for the other methods. LOOCV can sometimes over-estimate the model performance as only one sample was tested at a time; in contrast, 20% of the independent testing data were tested in the other five models. LOOCV can also have a higher variance than other CV methods (Li, 2016). As such, the WRTDS results are not directly comparable to the other methods.

Stand-alone ML achieved results that placed it between WRTDS and hybrid ML. Stand-alone GBM achieved the highest accuracy for NH_4 prediction in Murray River. Hybrid RF and hybrid GBM had the lowest RSME and highest MEF for all nutrients except NH_4, in Ellen Brook and Murray River (Figure 3). Compared to the stand-alone ML, the hybrid ML also had much lower prediction uncertainty, in that the RMSE and MEF had narrower IQR than that of the stand-alone ML, especially for DON and FRP prediction in Ellen Brook and DOC prediction in Murray River. The use of pre-generated daily nutrient data was the only difference between hybrid ML and stand-alone ML. This means that the generated nutrients provided additional information for the hybrid model that allowed more stable results. Interestingly, while the hybrid ML had better performance than the stand-alone ML, there was no significant difference in performance between the hybrid RF and hybrid GBM, though they showed differences between different nutrient species. For instance, hybrid RF achieved slightly better performance for DOC in Ellen Brook, while hybrid GBM had lower RMSE for DOC in Murray River. There was no significant performance difference between stand-alone RF and GBM.

In summary, the hybrid ML had the best performance amongst the six methods, followed by stand-alone RF and GBM. WRTDS was better than the linear model but could only achieve results similar to stand-alone RF and GBM for NH_4 prediction in Ellen Brook, and for NH_4 and FRP prediction in Murray River.

4.2 Generated daily TN in Ellen Brook

Model performance for six nutrients was compared in last section. To make this section more concise, t~~T~~hese six models were then compared in their ability to generate daily TN in Ellen Brook from 01/01/1989 to 16/07/2018 (Figure 4). The daily TN in

Murray River, and daily TP in both sites were also generated (see results in the supplement document). TN was selected because TN is the most important and most frequently measured nutrient in many places. This hybrid method can also be used for other nutrients. ~~These six models were compared in their ability to generate daily TN in Ellen Brook from 01/01/1989 to 16/07/2018 (Figure 4).~~ Note that all data points (not just the 80% training dataset) were used to generate daily TN. ~~Similar results to those presented in the above section, were found for the generated daily TN in the Murray River, and for TP in both Ellen Brook and the Murray River (results in Appendix A).~~

Field Code Changed

The LM performed very poorly for TN prediction; low concentration samples (TN < 1.9 mg/L) were all under-estimated, and some extremely high concentrations were incorrectly generated due to the high flow (Figure 4-a). There were some seasonal patterns in the generated TN which come from the flow data. LM only used total flow to predict nutrient concentrations while other important hydrological processes were ignored. Thus the over-simplified LM had high errors in nutrient prediction (Figure 4), and this method might be more suitable for solutes that are not substantially bioactive (e.g., SiO_2, Ca^{2+}, Mg^{2+}, Cl^-) (Stallard and Murphy, 2014). The WRTDS captured some seasonal patterns of TN (from 2008 to 2018) but still had problems predicting TN between 1989 to 1996; some extremely high values were generated, and TN < 1.0 mg/L were over-estimated. Some high values (e.g., TN in 2008) were under-estimated (Figure 4-b). Stand-alone ML and hybrid ML generated similar daily TN data but varied in the detail. These models successfully captured the low concentration data and the seasonal pattern of TN. Unlike results by WRTDS, the generated TN by stand-alone ML and hybrid ML have more consistent seasonal pattern from 1989 to 2018. The RF and hybrid RF both under-estimated a few high concentration data (TN > 4.0 mg/L), compared to GBM and hybrid GBM, although hybrid RF still showed better performance than RF. For instance, high concentration data in 2007 and again from 2014 to 2017 were successfully predicted by hybrid RF but under-estimated by RF. Compared to stand-alone GBM, the hybrid GBM achieved lower errors for high concentration data.

Apart from the better performance for high concentration data, another difference between stand-alone ML and hybrid ML was that the long-term trend in TN was consistent in stand-alone ML, but this trend fluctuated in hybrid ML. For instance, hybrid GBM results fluctuated from 1989 to 1999 and then showed an increasing long-term trend from 2005~~0~~ to 2018, in addition to the seasonal fluctuation. The pre-generated nutrient is the only difference between stand-alone model and hybrid model.~~This suggests that the generated nutrient data provided additional information that allowed the hybrid ML to capture long-term trends; this information was not included in the temporal data, but existed in the generated nutrient data.~~ If there are long-term trends in nutrient concentrations (e.g., TN), similar trends should also exist in the components of TN (either DON or dissolved inorganic nitrogen). The pre-generated nutrients emphasise this impact on the hybrid model. This suggests that the generated nutrient data could provide additional information that allowed the hybrid ML to capture long-term trends; this information was not included in the seasonal components, but existed in the generated nutrient data.

The distribution of the TN data generated by the six models was compared to the distribution of the measured TN data (Figure 5). Similar to the results shown in Figure 4, hybrid GBM had the most similar distribution to the measured TN data. Only a few low and high concentration data ~~(TN> 4.0 mg/L)~~ were ~~under-estimated~~incorrectly predicted by the hybrid GBM. Hybrid RF also achieved a distribution similar to the measured data, but more extreme~~several~~ value~~high concentration~~ data were under-estimated compared to the hybrid GBM. Stand-alone GBM and RF showed similar distribution to the hybrid GBM and RF with less accuracy in the extreme data. Overall, GBM (either stand-alone model or hybrid model) could have a ~~n~~ better~~improved~~ distribution than~~compared to~~ RF~~, while~~ WRTDS generated some extremely high data and under-estimated many low concentration data which is also found in~~, as shown in~~ Figure 4-b. The linear model incorrectly predicted most of the TN data. The results in both Figure 4 and Figure 5 showed that hybrid GBM achieved the best simulated daily TN data, followed by hybrid RF, stand-alone GBM, and RF. WRTDS and LM generated large biases in TN prediction.

The hybrid ML models predicted most of the extreme concentrations (Figure 4 and Figure 5) and only a few points were under-predicted. The limited number of extreme data and the model structure that tried to balance the overall trend prediction with extreme data prediction can cause the under-prediction. For example, higher weights can be set up for extreme data during the model training process to force model to over-predict the value for extreme concentrations, which may reduce the accuracy for overall trend prediction. In this study, our target is to understand the long-term nutrient trend. Therefore, we did not use this technique during the model training process.

4.3 Comparison of variable importance in hybrid GBM for TN prediction

The daily data generated by the hybrid GBM showed lower RMSE and better distribution than stand-alone ML, WRTDS, and LM (Figure 4 and Figure 5). Compared to LM, WRTDS, and simple CART models, one drawback of RF and GBM, as well as many ML methods in general, is that there is no specific equation in GBM or RF to directly demonstrate model structures. However, GBM and RF do provide the relative importance of each variable, which is based on the empirical improvement in the loss function due to the split on the specific variable in a tree (Povak et al., 2014; Puissant et al., 2014). The improvement of a certain variable was averaged over all trees as the relative importance for the final model. This relative importance serves as the key index to understanding the model structure of RF and GBM (Makler-Pick et al., 2011).

The variable importance for TN prediction by hybrid GBM in Ellen Brook and Murray River is presented in Figure 6. The variable importance in the intermediate models is also included, and the length of coloured sections represents the importance of those variables in the hybrid GBM or intermediate GBM. The importance was scaled according to the most important variable. The generated DON and TP ranked as the first two critical variables in Ellen Brook, while all three generated nutrients were listed as the most important variables in Murray River. This suggests that the generated nutrients do provide critical information to the model and improve model performance. The quickflow was most important for the generated DON and TP, as well as the TN itself in Ellen Brook. The impacts of quickflow decreased, and baseflow, seasonal components ~~temporal data~~

and rainfall data become more important for TN prediction in Murray River. This difference in variable importance reflects different catchment characteristics across the two sites, and therefore different hydrological and hydrochemical processes controlling TN concentrations. The total flow was not of high importance in either site, which suggests that baseflow or quickflow had more impact on surface water TN. Moreover, TN concentrations were affected by more variables in Murray River than in Ellen Brook.

5. Discussion

5.1 Different sources of TN in Ellen Brook and Murray River

Hydrological conditions, specific sub-catchment characteristics and the chemical properties of nutrients can all impact surface water nutrient concentrations (Barron et al., 2009; Moatar et al., 2016), nutrient partitioning (Ruibal-Conti et al., 2013), and nutrient transport (Burt and Pinay, 2005; Tesoriero et al., 2009). TN prediction in Murray River was impacted by more variables than in Ellen Brook (Figure 6), suggesting more complex relationships in Murray River.

Quick flow is composed of runoff, interflow and direct precipitation (Brodie and Hostetler, 2005), and was shown to be important for TN prediction in Ellen Brook. Direct precipitation, however, did not have a large impact on TN (the green bars in Figure 6); this suggests that runoff and interflow were important for TN concentrations. Baseflow can account for (on average) 53% of annual stream discharge in Ellen Brook, but baseflow was not of high importance for TN prediction in this study. This may occur due to low TN concentrations in the baseflow (Barron et al., 2009), large areas of low nutrient-retaining sandy soils in Ellen Brook catchment, and high nutrient transport efficiency in quickflow and first flush. Mellander et al. (2012) quantified nutrient transport pathways in agricultural catchments and found that quickflow were only 2~8% of total flow, but it can transport up to 50% of TP. Gunaratne et al, (2017) found that the seasonal first flush were only 30% of runoff volume but contained 40~70% of the nutrient load.

Note that the median TN in Ellen Brook (2.1 mg/L) is significantly higher than that in Murray River (0.67 mg/L) which can be explained to some extent by the large area of grazing lands in Ellen Brook. Previous investigations in south-eastern Australia (Adams et al., 2014), New Zealand (Davies-Colley et al., 2004), and north-western Europe (Conroy et al., 2016) all suggested that livestock can increase TN discharge to the receiving water bodies. Most of the piggeries and poultry farms in the Swan-Canning catchment are located in Ellen Brook catchment (Kelsey et al., 2010), which has the highest TN and TP discharge loads. Thus the large grazing areas, piggeries and poultry farms, and low nutrient-retaining sandy soils may explain the importance of quickflow for TN prediction and high TN concentrations in Ellen Brook.

Baseflow is derived from groundwater discharge to streams and the slow drainage of water stored in local wetlands (Kelsey et al., 2010). Baseflow is highlighted as an important variable for TN prediction in Murray River. Murray River catchment has

large areas with high nutrient-retaining soils (high PRI) (Kelsey et al., 2011) and relative low TN concentrations, and it is likely that groundwater makes significant contributions to TN in Murray River. ~~Previous investigations in this study area found that DON was the dominant form of TN in both surface water and groundwater (Nice et al., 2009; Petrone, 2010; Bourke et al., 2015), and we assumed that groundwater TN contributed to surface water TN in the form of DON.~~ Ruibal-Conti et al. (2014) previously found that variability in TN is strongly associated with variability in flows in Murray River. Our results extend this finding, in that both baseflow and quickflow likely impact TN in the river.

It is noted that seasonal components ~~temporal data~~ including *Sin(JD)*, and *Cos(JD)* showed significantly higher importance in Murray River. This may because seasonal information is captured in other inputs in Ellen Brook (e.g., quickflow and baseflow). But the main ~~reason is the~~ ~~This indicates~~ stronger seasonal TN signals in Murray River compared to Ellen Brook. ~~;~~ ~~T~~this finding is supported by the generated daily TN data for Murray River (see results in supplement S.2). Natural reserves occupy large areas of the Murray River catchment, and this may increase seasonal signals. Additionally, the lagged quickflow, baseflow, and rainfall were generated (for the previous 3, 7 and 15 days) but only the lagged 15-day baseflow and quickflow were ranked as important variables for both Ellen Brook and Murray River. This suggests that a time-scale of nutrient transport in the sub-catchments, and likely reflects soil permeability and geology; long hydrochemical recessions from storm events may prolong their impact on the ecological status of receiving rivers (Mellander et al., 2012).

Six models were compared for nutrient predictions and the hybrid GBM model achieved the highest accuracy (Figure 3 and Figure 5). The long-term changes of TN have been discussed in previous sections. To understand the long-term changes of other nitrogen species across the year, the hybrid GBM was then applied to generate daily DON, NH_4, and NOx in Ellen Brook from 01/01/1989 to 16/07/2018 (Figure 7). The generated DON has much higher concentration than NH_4 and NOx. This is consistent with previous investigations in this study area that DON was the dominant form of TN in both surface water and groundwater (Nice et al., 2009; Petrone, 2010; Bourke et al., 2015). There is no clear long-term patterns in generated NH_4 and NOx, however, an increasing long-term trend in generated DON can be found from 2006 to 2018. There is also an increasing trend in TN from 2005 (Figure 4), suggesting DON was the main reason for the increasing TN concentrations. DON is often assumed to be relatively slow to react, but depending on the source of DON, it can turnover rapidly, thereby constituting an active contributor to the eutrophication of surface waters (Petrone et al., 2009).

Formatted: Subscript

Formatted: Subscript

Formatted: Subscript

5.2 Can we improve our understanding of historical nutrient conditions using a contemporary data?

The generated nutrient data provided additional information to enhance the hybrid model performance (Figure 3 and Figure 5). To assess the individual impact of a generated nutrient, we did a simple test that sequentially added generated TP, DOC, and DON data to the base GBM (only seasonal components and lagged hydrological data) and evaluated RMSE and MEF for TN prediction. This process was repeated 30 times and the results are presented in Figure 8.

The RMSE significantly decreased when generated TP was added as an additional variable. DOC and DON only have 297 and 129 data, respectively, and were only measured in recent years, while TP has more than a thousand data and has been measured since 1990 (Table 3). However, DOC and DON could still improve model performance (Figure 8), and the generated DON was ranked as the most important variable across both sites (Figure 6). The medium RMSE slightly decreased when both generated DOC and DON were added. Moreover, the generated DOC and DON also reduced the model uncertainty, such that the IQR ranges became narrower than model results without the generated nutrients.

Our results suggest that the recent DON and DOC data improved understanding of historical TN. It is not uncommon to have a similar data structure when several datasets are combined, or new measurements are added to a project. While there were no DON data prior to 2006 in Ellen Brook, daily DON can be generated back to 1990 with the help of generated TN, DOC, TP data; DON had the highest MEF among the six nutrients (Figure 6). This hybrid method provides a feasible process to fully utilise all available nutrient data to accurately fill gaps in either historical or recent nutrient datasets.

5.3 A comprehensive comparison of six models

Monitoring, modelling, and forecasting water quality inputs are essential to support the management of the quality of receiving waters while responding to current anthropogenic stressors (Holguin-Gonzalez et al., 2013; Schnoor, 2014). The performances of six models were comprehensively compared, in an exploration of historical and contemporary nutrient data across two study sites. LM had the highest error while stand-alone RF and GBM had similar error. This agrees with previous findings by Erdal and Karakurt (2013) that RF and GBM models achieved similar correlation coefficients (R) for streamflow forecasting. Ismail and Mutanga (2010) also reported that RF and GBM increased the R of a single CART by 10.01%, and 9.59%, respectively.

The performance of WRTDS, as well as many conceptual models, is often reliant on a prescribed set of input information, which can account for variance in nutrient concentrations but may miss some important processes for certain rivers (e.g., baseflow in this study). This can compromise the performance of WRTDS for nutrient prediction. Moreover, hydrological and chemical processes within the systems are typically ignored by many conceptual models, which may exclude important hydrochemical information. By contrast, some complex conceptual models may include these hydrochemical processes but are often constrained by insufficient nutrient data to calibrate and validate the models. Some simplifications may be made to account for lack of data, but the simplifications may often weaken model performance. The hybrid framework presented in this study has overcome the challenge caused by data paucity, by building intermediate models to generate missing nutrient data, and then using this additional hydrochemical information to improve final model performance.

The hybrid models developed in this study were able to take advantage of the complementary strengths of both hydrochemical (additionally generated nutrient data) and hydrological (lagged data) information. This was particularly the case for prediction of high nutrient concentrations, where the hybrid models were shown to outperform the stand-alone RF and GBM, in terms of accuracy, reliability and value distributions. Improved accuracy in the hybrid model was achieved by using intermediate models, although these intermediate models may also have a relatively high error (similar to stand-alone RF and GBM). However, if the improved model performance is higher than the introduced error, the results are manageable. Similar results were also found in (Hunter et al., (2018) that a hybrid process-driven and ANN model was compared with the stand-alone ANN model and the process-driven model. In their study, the hybrid also achieved the best performance followed by stand-alone ANN. The process-driven benchmark model had significantly lower accuracy than other two models.

A limitation of the hybrid modelling approach, however, is that it requires the time and expertise to develop intermediate models for generating additional nutrient data. Prior knowledge also plays an important role in identifying the variables for pre-generation. Some statistical methods (e.g. the correlation test, simple linear model) can be helpful to identify these variables if there is no clear theoretical or conceptual understanding on which to base selection of the important variables.

In this study, we tested the generalised performance of the hybrid model across six nutrient species and two tributaries. We also note that nutrients may not always be the critical variables targeted for pre-generation; the pre-generated DOC was ranked as having low importance for Ellen Brook, and produced only a slight improvement in the performance of the hybrid model for NH_4.

5.4 The application of ML methods for hydrological modelling

There were constraints in the nutrient datasets in this research, and similar constraints commonly exist in other study areas. Many nutrient datasets contain important information, but it sometimes can be challenging to directly combine or utilise them. ML methods provide a feasible approach to pre-process these datasets or combine them. In this study, the concentrations of missing nutrient species were first predicted by the intermediate ML method and then used as inputs for another ML method for final predictions. The pre-generation of missing data and pre-modelling hydrological analysis were critical components of the hybrid model and allowed identification of the impact of different hydrological transport pathways for TN export from the two tributary catchments. The hybrid ML methods were further applied to generate nutrient data for eight tributaries, and the generated data have since been used as inputs to an estuary prediction model, which simulates and forecasts nutrient concentrations in the previous and next five days in the Swan-Canning Estuary (Huang et al., 2019). The modelling methods and strategies developed in the work presented here, can be easily applied to other study areas. Overall, ML methods provide a flexible and feasible solution to explore the underlying relationships, re-construct spatial and temporal datasets, and combine different models.

6. Summary and conclusion

A hybrid machine learning model was developed, and its performance tested on six nutrients and two estuary tributaries and compared with alternative modelling approaches. The hybrid ML model exhibited higher prediction accuracy and lower prediction uncertainty than stand-alone ML, WRTDS, and LM, for almost all nutrients. The pre-generation of missing data and pre-modelling hydrological analysis were critical components of the hybrid model and allowed identification of the impact of different hydrological transport pathways for TN export from the two tributary catchments. The results of this study demonstrate the advantages of using hybrid models for high temporal resolution nutrient prediction; the results also demonstrate the use of the hybrid model for re-analysis of historical data in the light of contemporary data. Modelling strategies for different modelling targets and dataset structures have also been discussed. The modelling framework presented here can aid others to fully use all available nutrient data to generate accurate nutrient predictions.

Code and data availability. The data and the data sources used in this study are cited and explained in the text. ~~Readers can obtain all the necessary data and code from https://github.com/benyawang-uwa/daily-nutrient-prediction 10.5281/zenodo.3739611~~ The current version of model is available from the project website: https://github.com/benyawang-uwa/daily-nutrient-prediction under the MIT licence. The exact version of the model used to produce the results used in this paper is archived on Zenodo (doi:10.5281/zenodo.3739611).

Author contributions. Benya Wang, Matthew R. Hipsey, Carolyn Oldham contributed to the development of the methodology and designed the experiments, and Benya Wang carried them out. Benya Wang developed the model code and performed the simulations. Benya Wang prepared the paper with contributions from all coauthors.

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

Acknowledgements. The authors acknowledge Peisheng Huang and Brendan Busch for providing the historical nutrient data.

Financial support. Benya Wang was supported by a postgraduate scholarship provided by the CRC for Water Sensitive Cities. Matthew R. Hipsey received funding support from Australian Research Council project LP150100451.

References

Abbott, B. W., Baranov, V., Mendoza-Lera, C., Nikolakopoulou, M., Harjung, A., Kolbe, T., Balasubramanian, M. N., Vaessen, T. N., Ciocca, F., Campeau, A., Wallin, M. B., Romeijn, P., Antonelli, M., Gonçalves, J., Datry, T., Laverman, A. M., de Dreuzy, J. R., Hannah, D. M., Krause, S., Oldham, C. and Pinay, G.: Using multi-tracer inference to move beyond single-catchment ecohydrology, Earth-Science Rev., 160, 19–42, doi:10.1016/j.earscirev.2016.06.014, 2016.

Formatted: No underline, Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: Font color: Text 1
Formatted: English (Australia)

Adams, R., Arafat, Y., Eate, V., Grace, M. R., Saffarpour, S., Weatherley, A. J. and Western, A. W.: A catchment study of sources and sinks of nutrients and sediments in south-east Australia, J. Hydrol., 515, 166–179, doi:10.1016/j.jhydrol.2014.04.034, 2014.

Álvarez-Cabria, M., Barquín, J. and Peñas, F. J.: Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors, Sci. Total Environ., 545–546, 152–162, doi:10.1016/j.scitotenv.2015.12.109, 2016.

Barron, O., Donn, M., Furby, S., Chia, J. and Johnstone, C.: Groundwater contribution to nutrient export from the Ellen Brook catchment., 2009.

Belgiu, M. and Drăgu, L.: Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens., 114, 24–31, doi:10.1016/j.isprsjprs.2016.01.011, 2016.

Bernal, S., Butturini, A. and Sabater, F.: Seasonal variations of dissolved nitrogen and DOC:DON ratios in an intermittent Mediterranean stream, Biogeochemistry, 75(2), 351–372, doi:10.1007/s10533-005-1246-7, 2005.

Bourke, S., Hammond, M. and Clohessy: Perth Shallow Groundwater Systems Investigation: North Lake. [online] Available from: http://www.water.wa.gov.au/PublicationStore/first/91255.pdf, 2015.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, 2001.

Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A.: Classification and regression trees, CRC press, Boca Raton., 1984.

Burt, T. P. and Pinay, G.: Linking hydrology and biogeochemistry, Prog. Phys. Geogr., 3(29), 297–316, doi:10.1067/mva.2002.123763, 2005.

Chanat, J. G., Rice, K. C. and Hornberger, G. M.: Consistency of patterns in concentration-discharge plots, Water Resour. Res., 38(8), 10–22, doi:10.1029/2001WR000971, 2002.

Chen, Y., Liu, R., Sun, C., Zhang, P., Feng, C. and Shen, Z.: Spatial and temporal variations in nitrogen and phosphorous nutrients in the Yangtze River Estuary, Mar. Pollut. Bull., 64(10), 2083–2089, doi:https://doi.org/10.1016/j.marpolbul.2012.07.020, 2012.

Clapcott, J. E., Collier, K. J., Death, R. G., Goodwin, E. O., Harding, J. S., Kelly, D., Leathwick, J. R. and Young, R. G.: Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity, Freshw. Biol., 57(1), 74–90, doi:10.1111/j.1365-2427.2011.02696.x, 2012.

Cohn, T. A., Delong, L. L., Gilroy, E. J., Hirsch, R. M. and Wells, D. K.: Estimating constituent loads, Water Resour. Res., 25(5), 937–942, doi:10.1029/WR025i005p00937, 1989.

Conroy, E., Turner, J. N., Rymszewicz, A., O'Sullivan, J. J., Bruen, M., Lawler, D., Lally, H. and Kelly-Quinn, M.: The impact of cattle access on ecological water quality in streams: Examples from agricultural catchments within Ireland, Sci. Total Environ., 547, 17–29, doi:10.1016/j.scitotenv.2015.12.120, 2016.

Coopersmith, E. J., Minsker, B. and Montagna, P.: Understanding and forecasting hypoxia using machine learning algorithms, J. Hydroinformatics, 13(1), 64, doi:10.2166/hydro.2010.015, 2010.

Coops, N. C., Waring, R. H., Beier, C., Roy-Jauvin, R. and Wang, T.: Modeling the occurrence of 15 coniferous tree species throughout the Pacific Northwest of North America using a hybrid approach of a generic process-based growth model and decision tree analysis, Appl. Veg. Sci., 14(3), 402–414, doi:10.1111/j.1654-109X.2011.01125.x, 2011.

Cozzi, S. and Giani, M.: River water and nutrient discharges in the Northern Adriatic Sea: Current importance and long term changes, Cont. Shelf Res., 31(18), 1881–1893, doi:10.1016/j.csr.2011.08.010, 2011.

Crowder, D. W., Demissie, M. and Markus, M.: The accuracy of sediment loads when log-transformation produces nonlinear sediment load-discharge relationships, J. Hydrol., 336(3–4), 250–268, doi:10.1016/j.jhydrol.2006.12.024, 2007.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. and Lawler, J. J.: Random forests for classification in ecology, Ecology, 88(11), 2783–2792, doi:10.1890/07-0539.1, 2007.

Davies-Colley, R. J., Nagels, J. W., Smith, R. A., Young, R. G. and Phillips, C. J.: Water quality impact of a dairy cow herd crossing a stream, New Zeal. J. Mar. Freshw. Res., 38(4), 569–576, doi:10.1080/00288330.2004.9517262, 2004.

Denvil-Sommer, A., Gehlen, M., Vrac, M. and Mejia, C.: LSCE-FFNN-v1: A two-step neural network model for the reconstruction of surface ocean pCO2 over the global ocean, Geosci. Model Dev., 12(5), 2091–2105, doi:10.5194/gmd-12-2091-2019, 2019.

Domingues, R. B., Anselmo, T. P., Barbosa, A. B., Sommer, U. and Galvão, H. M.: Nutrient limitation of phytoplankton growth in the freshwater tidal zone of a turbid, Mediterranean estuary, Estuar. Coast. Shelf Sci., 91(2), 282–297, doi:10.1016/j.ecss.2010.10.033, 2011.

Erdal, H. I. and Karakurt, O.: Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms, J. Hydrol., 477, 119–128, doi:10.1016/j.jhydrol.2012.11.015, 2013.

Erlandsson, M., Cory, N., Fölster, J., Köhler, S., Laudon, H., Weyhenmeyer, G. A. and Bishop, K.: Increasing Dissolved Organic Carbon Redefines the Extent of Surface Water Acidification and Helps Resolve a Classic Controversy, Bioscience, 61(8), 614–618 [online] Available from: http://dx.doi.org/10.1525/bio.2011.61.8.7, 2011.

Filep, T. and Rékási, M.: Factors controlling dissolved organic carbon (DOC), dissolved organic nitrogen (DON) and DOC/DON ratio in arable soils based on a dataset from Hungary, Geoderma, 162(3–4), 312–318, doi:10.1016/j.geoderma.2011.03.002, 2011.

Forio, M. A. E., Landuyt, D., Bennetsen, E., Lock, K., Nguyen, T. H. T., Ambarita, M. N. D., Musonge, P. L. S., Boets, P., Everaert, G., Dominguez-Granda, L. and Goethals, P. L. M.: Bayesian belief network models to analyse and predict ecological water quality in rivers, Ecol. Modell., 312, 222–238, doi:10.1016/j.ecolmodel.2015.05.025, 2015.

Friedman, J.: Greedy Function Approximation: A Gradient Boosting Machine, Ann. Stat., 29(5), 1189–1232, doi:10.1214/009053606000000795, 2001.

Friedman, J. H.: Stochastic gradient boosting, Comput. Stat. Data Anal., 38(4), 367–378, doi:https://doi.org/10.1016/S0167-9473(01)00065-2, 2002.

Fuka, D., Walter, T., Archibald, J., Tammo, S. and Easton, Z.: R package ' EcoHydRology.', 2018.

Furey, P. R. and Gupta, V. K.: A physically based filter for separating base flow from streamflow time series, Water Resour. Res., 37(11), 2709–2722, doi:10.1029/2001WR000243, 2001.

Giblin, A. E., Weston, N. B., Banta, G. T., Tucker, J. and Hopkinson, C. S.: The Effects of Salinity on Nitrogen Losses from an Oligohaline Estuarine Sediment, Estuaries and Coasts, 33(5), 1054–1068, doi:10.1007/s12237-010-9280-7, 2010.

Górniak, A., Zieliński, P., Jekatierynczuk-Rudczyk, E., Grabowska, M. and Suchowolec, T.: The role of dissolved organic carbon in a shallow lowland reservoir ecosystem - A long-term study, Acta Hydrochim. Hydrobiol., 30(4), 179–189, doi:10.1002/aheh.200390001, 2002.

Green, C. T., Bekins, B. a, Kalkhoff, S. J., Hirsch, R. M., Liao, L. and Barnes, K. K.: Decadal surface water quality trends under variable climate, land use, and hydrogeochemical setting in Iowa, USA, , 2425–2443, doi:10.1002/2013WR014829.Received, 2014.

Greening, H., Janicki, A., Sherwood, E. T., Pribble, R. and Johansson, J. O. R.: Ecosystem responses to long-term nutrient management in an urban estuary: Tampa Bay, Florida, USA, Estuar. Coast. Shelf Sci., 151, A1–A16, doi:https://doi.org/10.1016/j.ecss.2014.10.003, 2014.

Gunaratne, G. L., Vogwill, R. I. J. and Hipsey, M. R.: Effect of seasonal flushing on nutrient export characteristics of an urbanizing, remote, ungauged coastal catchment, Hydrol. Sci. J., 62(5), 800–817, doi:10.1080/02626667.2016.1264585, 2017.

Guo, D., Lintern, A., Webb, J. A., Ryu, D., Liu, S., Bende-Michl, U., Leahy, P., Wilson, P. and Western, A. W.: Key Factors Affecting Temporal Variability in Stream Water Quality, Water Resour. Res., 55(1), 112–129, doi:10.1029/2018WR023370, 2019.

Halliday, S. J., Wade, A. J., Skeffington, R. A., Neal, C., Reynolds, B., Rowland, P., Neal, M. and Norris, D.: An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales, Sci. Total Environ., 434, 186–200, doi:10.1016/j.scitotenv.2011.10.052, 2012.

Heathwaite, A. L.: Multiple stressors on water availability at global to catchment scales: Understanding human impact on nutrient cycles to protect water quality and water availability in the long term, Freshw. Biol., 55(SUPPL. 1), 241–257, doi:10.1111/j.1365-2427.2009.02368.x, 2010.

Herndon, E. M., Dere, A. L., Sullivan, P. L., Norris, D., Reynolds, B. and Brantley, S. L.: Landscape heterogeneity drives contrasting concentration-discharge relationships in shale headwater catchments, Hydrol. Earth Syst. Sci., 19(8), 3333–3347, doi:10.5194/hess-19-3333-2015, 2015.

Hirsch, R. M. and De Cicco, L.: User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data, Tech. Methods B. 4, (February), 93, doi:http://dx.doi.org/10.3133/tm4A10, 2015.

Hirsch, R. M., Moyer, D. L. and Archfield, S. A.: Weighted regressions on time, discharge, and season (WRTDS), with an application to chesapeake bay river inputs, J. Am. Water Resour. Assoc., 46(5), 857–880, doi:10.1111/j.1752-1688.2010.00482.x, 2010.

Holguin-Gonzalez, J. E., Everaert, G., Boets, P., Galvis, A. and Goethals, P. L. M.: Development and application of an integrated ecological modelling framework to analyze the impact of wastewater discharges on the ecological water quality of rivers, Environ. Model. Softw., 48, 27–36, doi:10.1016/j.envsoft.2013.06.004, 2013.

Huang, P., Trayler, K., Wang, B., Saeed, A., Oldham, C., Busch, B. and Hipsey, M.: An integrated modelling system for water quality forecasting in an urban eutrophic estuary: The Swan-Canning Estuary virtual observatory, J. Mar. Syst., 199, 103218, doi:10.1016/j.jmarsys.2019.103218, 2019.

Hunter, J. M., Maier, H. R., Gibbs, M. S., Foale, E. R., Grosvenor, N. A., Harders, N. P. and Kikuchi-Miller, T. C.: Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems, Hydrol. Earth Syst. Sci., 22(5), 2987–3006, doi:10.5194/hess-22-2987-2018, 2018.

Ishwaran, H. and Kogalur, U. B.: Consistency of random survival forests, Stat. Probab. Lett., 80(13), 1056–1064, doi:https://doi.org/10.1016/j.spl.2010.02.020, 2010.

Ismail, R. and Mutanga, O.: A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa, Int. J. Appl. Earth Obs. Geoinf., 12, S45–S51, doi:https://doi.org/10.1016/j.jag.2009.09.004, 2010.

Jickells, T. D., Andrews, J. E., Parkes, D. J., Suratman, S., Aziz, A. A. and Hee, Y. Y.: Nutrient transport through estuaries: The importance of the estuarine geography, Estuar. Coast. Shelf Sci., 150(PB), 215–229, doi:10.1016/j.ecss.2014.03.014, 2014.

Jordan, P. and Cassidy, R.: Technical Note: Assessing a 24/7 solution for monitoring water quality loads in small river catchments, Hydrol. Earth Syst. Sci., 15(10), 3093–3100, doi:10.5194/hess-15-3093-2011, 2011.

Kaiser, D., Unger, D., Qiu, G., Zhou, H. and Gan, H.: Natural and human influences on nutrient transport through a small subtropical Chinese estuary, Sci. Total Environ., 450–451, 92–107, doi:https://doi.org/10.1016/j.scitotenv.2013.01.096, 2013.

Kelsey, P., Hall, J., Kitsios, A., Quinton, B. and Shakya, D.: Hydrological and nutrient modelling of the Swan-Canning coastal catchments, Water Science technical series, Department of Water, Western Australia., 2010.

Kelsey, P., Hall, J., Kretschmer, P., Quiton, B. and Shakya, D.: Hydrological and nutrient modelling of the Peel-Harvey catchment, Water Science Technical Series, Department of Water, Western Australia., 2011.

Lamsal, S., Grunwald, S., Bruland, G. L., Bliss, C. M. and Comerford, N. B.: Regional hybrid geospatial modeling of soil nitrate-nitrogen in the Santa Fe River Watershed, Geoderma, 135, 233–247, doi:10.1016/j.geoderma.2005.12.009, 2006.

Li, J.: Assessing spatial predictive models in the environmental sciences: Accuracy measures, data variation and variance explained, Environ. Model. Softw., 80, 1–8, doi:10.1016/j.envsoft.2016.02.004, 2016.

Li, M., Xu, K., Watanabe, M. and Chen, Z.: Long-term variations in dissolved silicate, nitrogen, and phosphorus flux from the Yangtze River into the East China Sea and impacts on estuarine ecosystem, Estuar. Coast. Shelf Sci., 71(1), 3–12, doi:https://doi.org/10.1016/j.ecss.2006.08.013, 2007.

Li, M., Lee, Y. J., Testa, J. M., Li, Y., Ni, W., Kemp, W. M. and Di Toro, D. M.: What drives interannual variability of hypoxia in Chesapeake Bay: Climate forcing versus nutrient loading?, Geophys. Res. Lett., 43(5), 2127–2134, doi:10.1002/2015GL067334, 2016.

Li, R., Liu, S., Zhang, G., Ren, J. and Zhang, J.: Biogeochemistry of nutrients in an estuary affected by human activities: The Wanquan River estuary, eastern Hainan Island, China, Cont. Shelf Res., 57, 18–31, doi:https://doi.org/10.1016/j.csr.2012.02.013, 2013.

Lintern, A., Webb, J. A., Ryu, D., Liu, S., Waters, D., Leahy, P., Bende-Michl, U. and Western, A. W.: What are the key catchment characteristics affecting spatial differences in riverine water quality?, Water Resour. Res., doi:10.1029/2017WR022172, 2018.

Liu, S. M., Li, L. W., Zhang, G. L., Liu, Z., Yu, Z. and Ren, J. L.: Impacts of human activities on nutrient transports in the Huanghe (Yellow River) estuary, J. Hydrol., 430–431, 103–110, doi:https://doi.org/10.1016/j.jhydrol.2012.02.005, 2012.

Lloyd, C. E. M., Freer, J. E., Collins, A. L., Johnes, P. J. and Jones, J. I.: Methods for detecting change in hydrochemical time series in response to targeted pollutant mitigation in river catchments, J. Hydrol., 514, 297–312, doi:10.1016/j.jhydrol.2014.04.036, 2014.

Lyne, V. and Hollick, M.: Stochastic Time-Variable Rainfall-Runoff Modeling., 1979.

Maier, H. R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L. S., Cunha, M. C., Dandy, G. C., Gibbs, M. S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D. P., Vrugt, J. a., Zecchin, A. C., Minsker, B. S., Barbour, E. J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M. and Reed, P. M.: Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions, Environ. Model. Softw., 62, 271–299, 2014.

Makler-Pick, V., Gal, G., Gorfine, M., Hipsey, M. R. and Carmel, Y.: Sensitivity analysis for complex ecological models - A new approach, Environ. Model. Softw., 26(2), 124–134, doi:10.1016/j.envsoft.2010.06.010, 2011.

Martínez-Rojas, M., Marín, N. and Vila, M. A.: The role of information technologies to address data handling in construction project management, J. Comput. Civ. Eng., 30(April), 1–10, doi:10.1061/(ASCE)CP.1943-5487, 2015.

McBratney, A. B., Odeh, I. O. A., Bishop, T. F. A., Dunbar, M. S. and Shatar, T. M.: An overview of pedometric techniques for use in soil survey, Geoderma, 97(3–4), 293–327, doi:10.1016/S0016-7061(00)00043-4, 2000.

Mellander, P. E., Melland, A. R., Jordan, P., Wall, D. P., Murphy, P. N. C. and Shortle, G.: Quantifying nutrient transfer pathways in agricultural catchments using high temporal resolution data, Environ. Sci. Policy, 24, 44–57, doi:10.1016/j.envsci.2012.06.004, 2012.

Meshgi, A., Schmitter, P., Chui, T. F. M. and Babovic, V.: Development of a modular streamflow model to quantify runoff contributions from different land uses in tropical urban environments using Genetic Programming, J. Hydrol., 525, 711–723, doi:10.1016/j.jhydrol.2015.04.032, 2015.

Meybeck, M. and Moatar, F.: Daily variability of river concentrations and fluxes: Indicators based on the segmentation of the rating curve, Hydrol. Process., 26(8), 1188–1207, doi:10.1002/hyp.8211, 2012.

Moatar, F., Abbott, B. W., Minaudo, C., Curie, F. and Pinay, G.: Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions, Water Resour. Res., 53, 1270–1287, doi:10.1002/2016WR019635, 2016.

Nathan, R. J. and McMahon, T. A.: Evaluation of automated techniques for base flow and recession analyses, Water Resour. Res., 26(7), 1465–1473, doi:10.1029/WR026i007p01465, 1990.

Nice, H., Foulsham, G., Bree, M. and Sarah, E.: A baseline study of contaminants in the sediments of the Swan and Canning estuaries., 2009.

Noori, N. and Kalin, L.: Coupling SWAT and ANN models for enhanced daily streamflow prediction, J. Hydrol., 533, 141–151, doi:10.1016/j.jhydrol.2015.11.050, 2016.

Paerl, H. W., Rossignol, K. L., Hall, S. N., Peierls, B. L. and Wetz, M. S.: Phytoplankton Community Indicators of Short- and Long-term Ecological Change in the Anthropogenically and Climatically Impacted Neuse River Estuary, North Carolina, USA, Estuaries and Coasts, 33(2), 485–497, doi:10.1007/s12237-009-9137-0, 2010.

Petrone, K. C.: Catchment export of carbon, nitrogen, and phosphorus across an agro-urban land use gradient, Swan-Canning River system, southwestern Australia, J. Geophys. Res., 115(G1), G01016, doi:10.1029/2009JG001051, 2010.

Petrone, K. C., Richards, J. S. and Grierson, P. F.: Bioavailability and composition of dissolved organic carbon and nitrogen in a near coastal catchment of south-western Australia, Biogeochemistry, 92(1–2), 27–40, 2009.

Povak, N. A., Hessburg, P. F., McDonnell, T. C., Reynolds, K. M., Sullivan, T. J., Salter, R. B. and Cosby, B. J.: Machine learning and linear regression models to predict catchment-level base cation weathering rates across the southern Appalachian Mountain region, USA, Water Resour. Res., 50(4), 2798–2814, doi:10.1002/2013WR014222.Received, 2014.

Puissant, A., Rougier, S. and Stumpf, A.: Object-oriented mapping of urban trees using Random Forest classifiers, Int. J. Appl. Earth Obs. Geoinf., 26, 235–245, doi:10.1016/j.jag.2013.07.002, 2014.

Quinlan, J. R.: Learning with continuous classes, Mach. Learn., 92, 343–348, doi:10.1.1.34.885, 1992.

Ruibal-Conti, A., Summers, R., Weaver, D. and Hipsey, M. R.: Hydro-climatological non-stationarity shifts patterns of nutrient delivery to an estuarine system, Hydrol. Earth Syst. Sci. Discuss., 10, 11035–11092, doi:10.5194/hessd-10-11035-2013, 2013.

Schnoor, J. L.: 4.1 - Water Quality and its Sustainability Introduction, edited by S. B. T.-C. W. Q. and P. Ahuja, pp. 1–40, Elsevier, Waltham., 2014.

Seitzinger, S. P., Sanders, R. W. and Styles, R.: Bioavailability of DON from natural and anthropogenic sources to estuarine plankton, Limnol. Oceanogr., 47(2), 353–366, doi:10.4319/lo.2002.47.2.0353, 2002.

Singh, K. P., Gupta, S. and Mohan, D.: Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches, J. Hydrol., 511, 254–266, 2014.

Staehr, P. A., Testa, J. and Carstensen, J.: Decadal Changes in Water Quality and Net Productivity of a Shallow Danish Estuary Following Significant Nutrient Reductions, Estuaries and Coasts, 40(1), 63–79, doi:10.1007/s12237-016-0117-x, 2017.

Stallard, R. F. and Murphy, S. F.: A Unified Assessment of Hydrologic and Biogeochemical Responses in Research Watersheds in Eastern Puerto Rico Using Runoff-Concentration Relations, Aquat. Geochemistry, 20(2–3), 115–139, doi:10.1007/s10498-013-9216-5, 2014.

Swan River Trust: Swan Canning Water Quality Improvement., 2009.

Szilagyi, J. and Parlange, M. B.: Baseflow separation based on analytical solutions of the Boussinesq equation, J. Hydrol., 204(1), 251–260, doi:https://doi.org/10.1016/S0022-1694(97)00132-7, 1998.

Tao, Y., Wei, M., Ongley, E., Li, Z. and Jingsheng, C.: Long-term variations and causal factors in nitrogen and phosphorus transport in the Yellow River, China, Estuar. Coast. Shelf Sci., 86(3), 345–351, doi:https://doi.org/10.1016/j.ecss.2009.05.014, 2010.

Tesoriero, A. J., Duff, J. H., Wolock, D. M., Spahr, N. E. and Almendinger, J. E.: Identifying Pathways and Processes Affecting Nitrate and Orthophosphate Inputs to Streams in Agricultural Watersheds, J. Environ. Qual., 38(5), 1892, doi:10.2134/jeq2008.0484, 2009.

Testa, J. M., Clark, J. B., Dennison, W. C., Donovan, E. C., Fisher, A. W., Ni, W., Parker, M., Scavia, D., Spitzer, S. E., Waldrop, A. M., Vargas, V. M. D. and Ziegler, G.: Ecological Forecasting and the Science of Hypoxia in Chesapeake Bay, Bioscience, 67(7), 614–626 [online] Available from: http://dx.doi.org/10.1093/biosci/bix048, 2017.

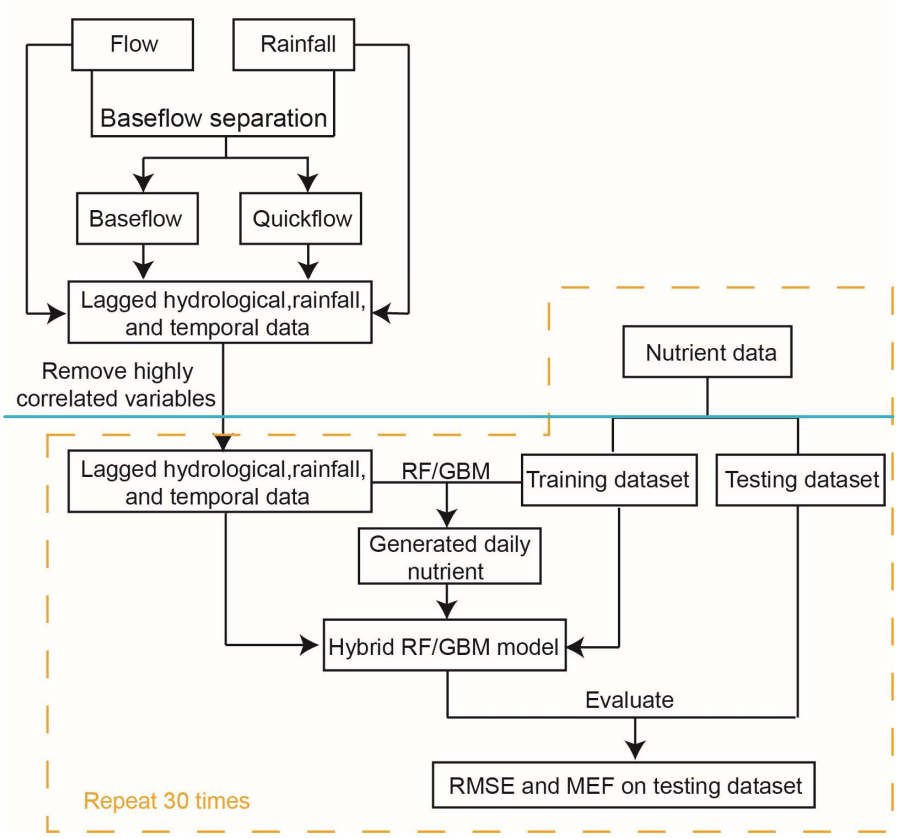
Wang, B., Hipsey, M. R., Ahmed, S. and Oldham, C.: The Impact of Landscape Characteristics on Groundwater Dissolved Organic Nitrogen: Insights From Machine Learning Methods and Sensitivity Analysis, Water Resour. Res., 54(7), 4785–4804, doi:10.1029/2017WR021749, 2018.

Yang, P., Yang, Y. H. and Zomaya, B. B. Z. and A. Y.: A Review of Ensemble Methods in Bioinformatics, Curr. Bioinform., 5(4), 296–308, doi:http://dx.doi.org/10.2174/157489310794072508, 2010.

Zhang, Q., Harman, C. J. and Ball, W. P.: An improved method for interpretation of riverine concentration-discharge relationships indicates long-term shifts in reservoir sediment trapping, Geophys. Res. Lett., 43(19), 10,215-10,224, doi:10.1002/2016GL069945, 2016a.

Zhang, Q., Ball, W. P. and Moyer, D. L.: Decadal-scale export of nitrogen, phosphorus, and sediment from the Susquehanna River basin, USA: Analysis and synthesis of temporal and spatial patterns, Sci. Total Environ., 563–564, 1016–1029, doi:10.1016/j.scitotenv.2016.03.104, 2016b.

Zhang, Q., Hirsch, R. M. and Ball, W. P.: Long-Term Changes in Sediment and Nutrient Delivery from Conowingo Dam to Chesapeake Bay: Effects of Reservoir Sedimentation, Environ. Sci. Technol., 50(4), 1877–1886, doi:10.1021/acs.est.5b04073, 2016c.



Formatted: Centered

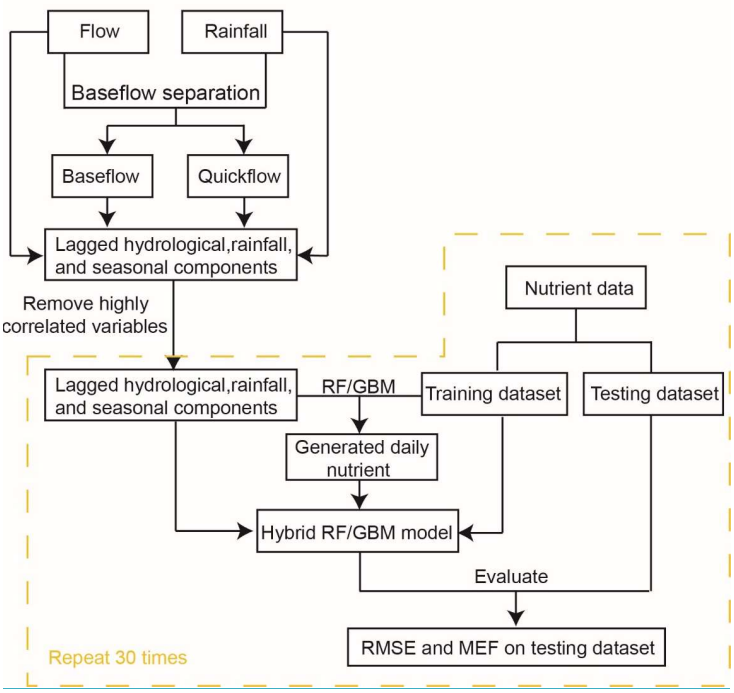


Figure 1. Overall modelling processes of ML-SWAN.

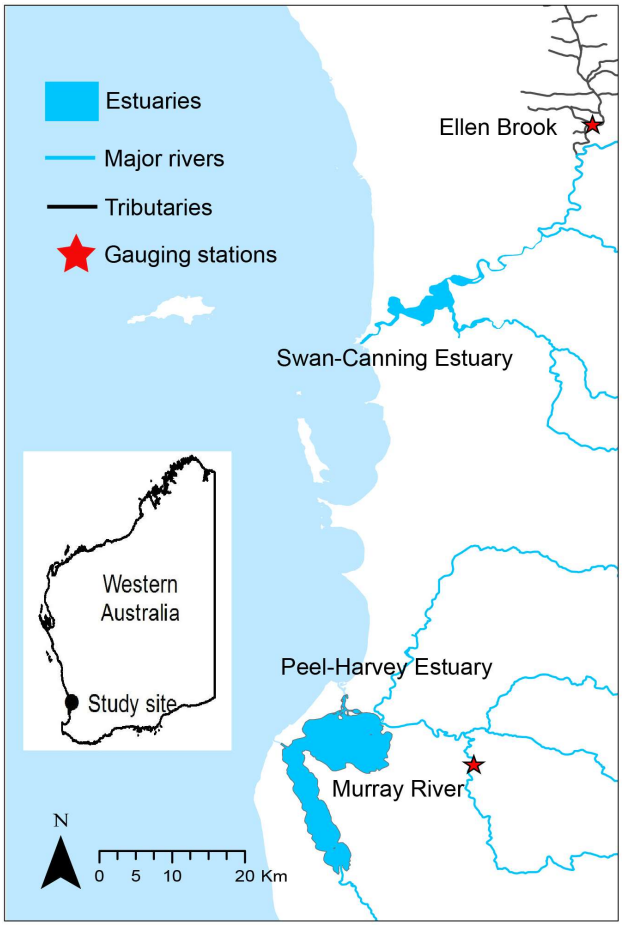


Figure 2. The location of Ellen Brook and Murray River.

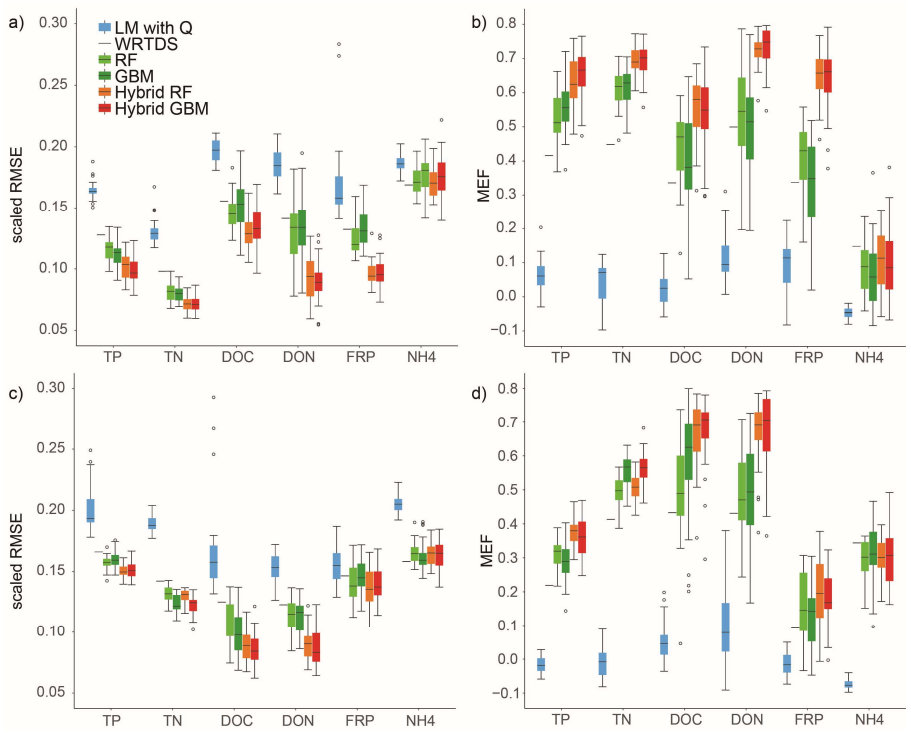


Figure 3. Model performance across six nutrients and the two sites: a) RMSE and b) MEF for Ellen Brook; c) RMSE and d) MEF results for Murray River.

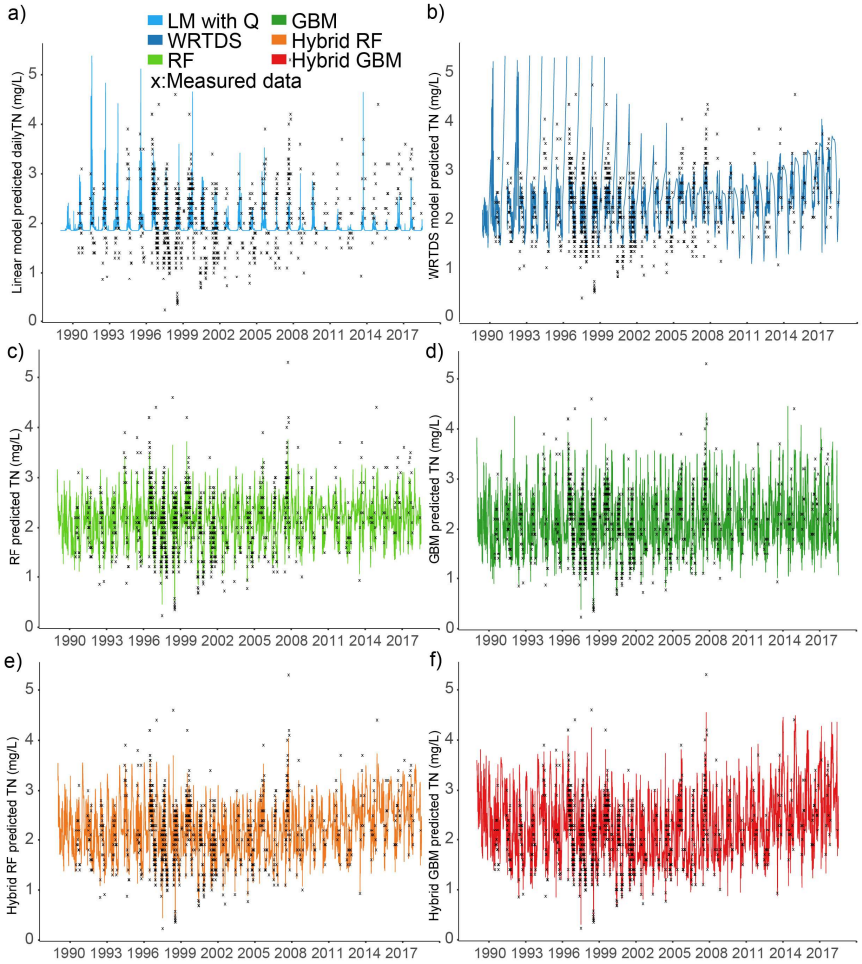


Figure 4. Daily TN generated by the six models, for Ellen Brook.

1060

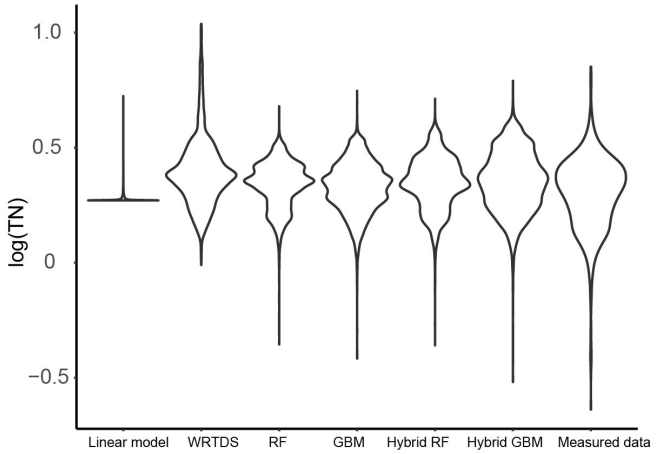
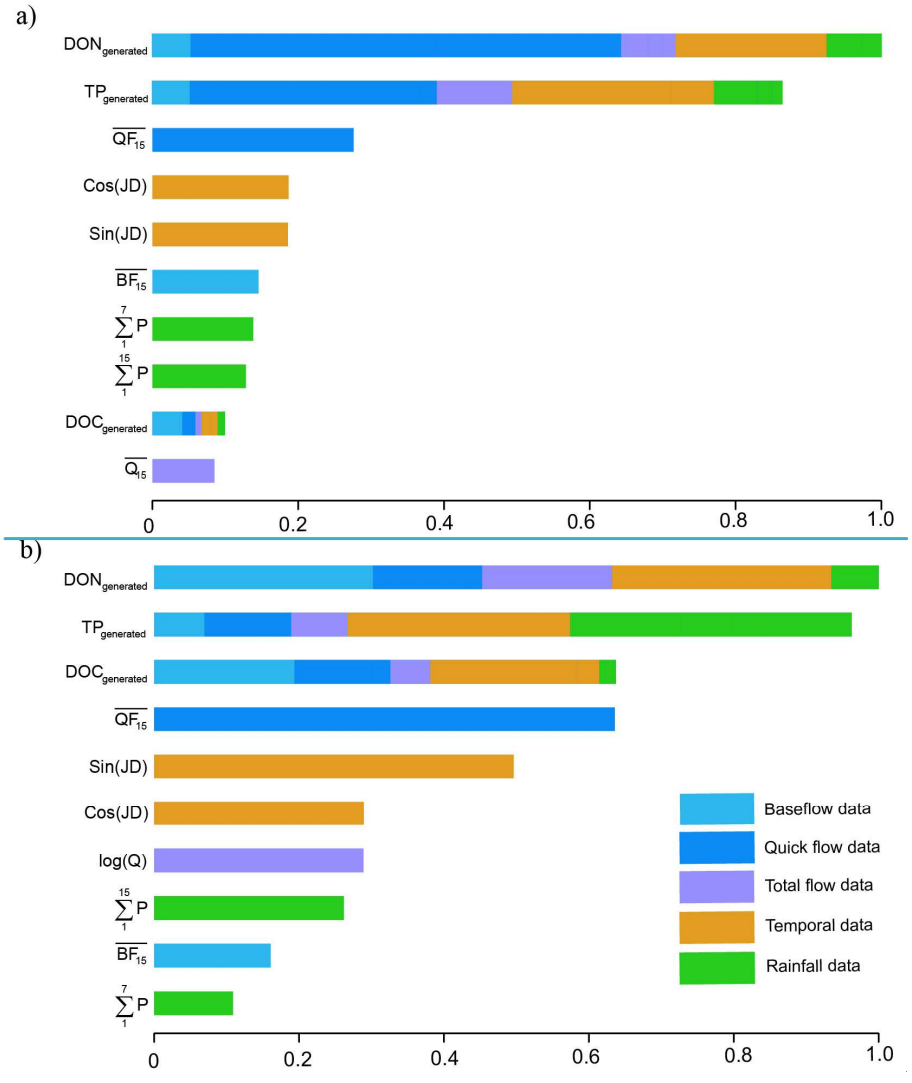


Figure 5. The distribution of the daily TN generated by the six models, and of the measured TN data in Ellen Brook.


a)
DON generated
TP generated
QF15
Cos(JD)
Sin(JD)
BF15
7 ∑ 1 P
15 ∑ 1 P
DOC generated
Q15
0
0.2
0.4
0.6
0.8
1.0
b)
DON generated
TP generated
DOC generated
QF15
Sin(JD)
Cos(JD)
log(Q)
15 ∑ 1 P
BF15
7 ∑ 1 P
Baseflow data
Quick flow data
Total flow data
Temporal data
Rainfall data
0
0.2
0.4
0.6
0.8
1.0

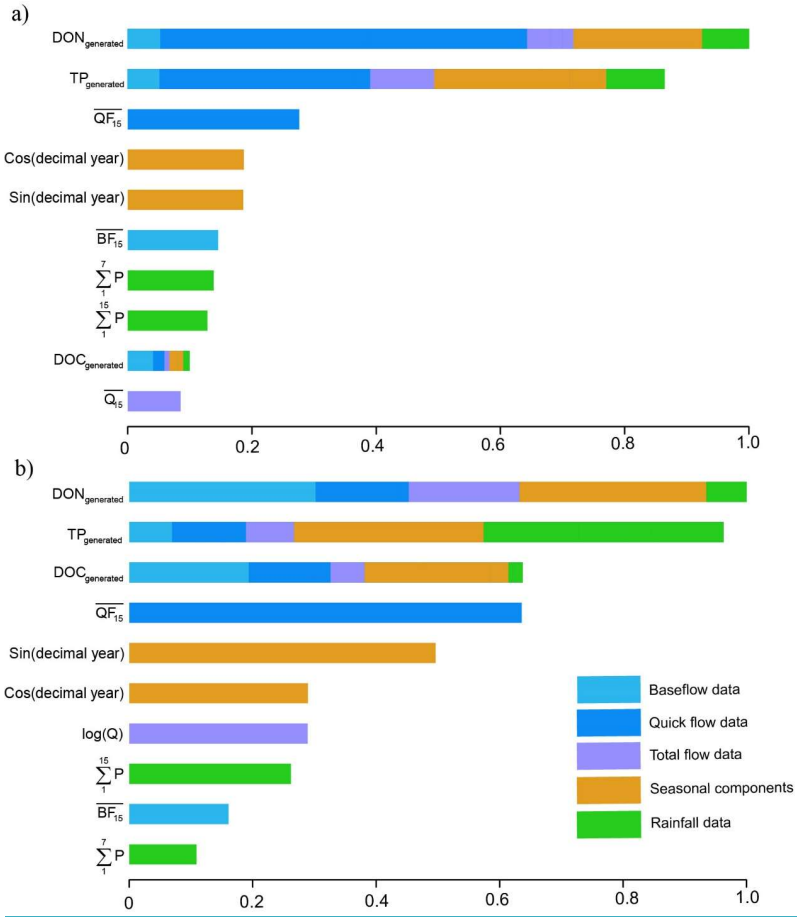


Figure 6. Variable importance in the hybrid GBM for TN prediction in a) Ellen Brook and b) Murray River.

Formatted: Normal

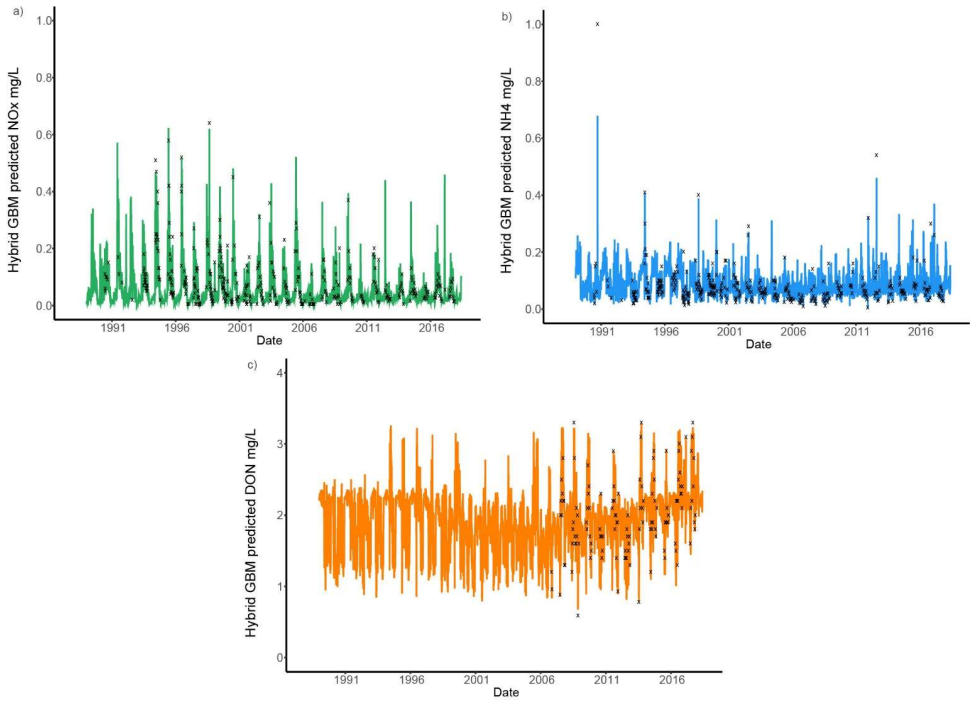
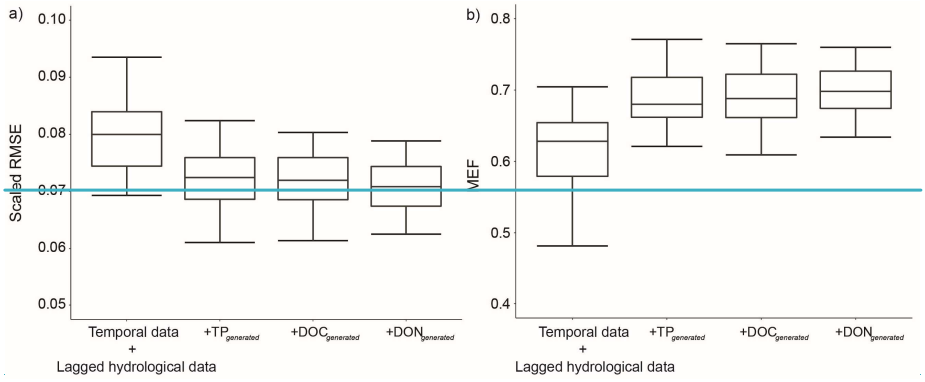


Figure 7. Generated daily DON, NOx, and NH_4 by the hybrid GBM for Ellen Brook.

Formatted: Subscript

Formatted: Normal



1070

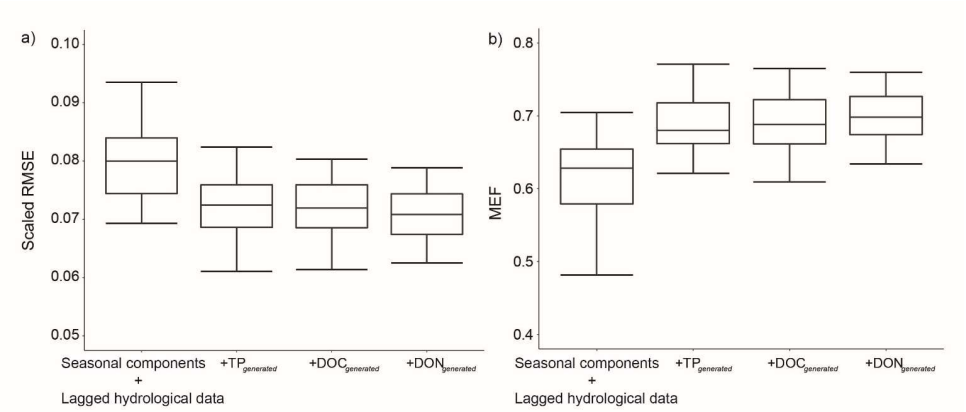


Figure 78. Model performance for TN prediction across different input variables, for Ellen Brook.

Table 1. Variable list and descriptions.

Variable type	Variable name	Abbreviation	Unit	Data source
Hydrological data	total discharge	Q	m³/s	data.wa.gov.au
	Average total discharge in last x days	$\bar{Q}_x$	m³/s	lagged average
	quickflow	QF	m³/s	equation 5
	Average quickflow in last x days	$\overline{QF}_x$	m³/s	lagged average

	baseflow	BF	m³/s	equation 5
	Average quickflow in last x days	$\overline{BF_x}$	m³/s	lagged average
Seasonal component ~~Temporal data~~ s	Julian day	JD		recorded
	cos (Julian day)	$Cos(JD)$		calculated
	sin (Julian day)	$Sin(JD)$		calculated
Metrological data	rainfall	P	mm	www.bom.gov.au
	cumulated rainfall in last x days	$\sum_{1}^{X} P$	mm	lagged sum
Nutrient data	total nitrogen	TN	mg/L	wir.water.wa.gov.au
	total phosphorus	TP	mg/L	wir.water.wa.gov.au
	dissolved organic carbon	DOC	mg/L	wir.water.wa.gov.au
	dissolved organic nitrogen	DON	mg/L	wir.water.wa.gov.au
	ammonia	NH_4	mg/L	wir.water.wa.gov.au
	filterable reactive phosphorus	FRP	mg/L	wir.water.wa.gov.au
	generated dissolved organic nitrogen	$DON_{generated}$	mg/L	generated by the intermediate model
	generated total phosphorus	$TP_{generated}$	mg/L	generated by the intermediate model
	generated dissolved organic carbon	$DOC_{generated}$	mg/L	generated by the intermediate model

Table 2. Hydrological characteristics of the two tributaries.

Site	Hydrological type	Annual flow (GL)	Area (km^2)	Land use
Ellen Brook	ephemeral	26.7	716	rural, agricultural, and grazing
Murray River	perennial	360	7855	agricultural and natural reserves

Table 3. Nutrient sampling time and sample size in Ellen Brook and Murray River

Site	Nutrient	First measurement	Sample size
	TN	1990	1057
	TP	1990	1022

Ellen Brook	DOC	1995	297
	DON	2006	129
	FRP	1990	404
	NH_4	1990	356
Murray River	TN	1983	1648
	TP	1983	1662
	DOC	2006	209
	DON	2006	207
	FRP	1990	300
	NH_4	1983	570