Prediction of stream nitrogen and phosphorus concentrations from high-frequency sensors using Random Forests Regression

*Joel W. Harrison[a], Mark A. Lucius[a], Jeremy L. Farrell[a], Lawrence W. Eichler[a], Rick A. Relyea[a]

[a]*Darrin Fresh Water Institute, Rensselaer Polytechnic Institute, 110 Eighth St., Troy, NY USA 12180*

**Corresponding author: harrij18@rpi.edu*

Abstract

Stream nutrient concentrations exhibit marked temporal variation due to hydrology and other factors such as the seasonality of biological processes. Many water quality monitoring programs sample too infrequently (i.e. weekly or monthly) to fully characterize lotic nutrient conditions and to accurately estimate nutrient loadings. A popular solution to this problem is the surrogate-regression approach, a method by which nutrient concentrations are estimated from related parameters (e.g., conductivity or turbidity) that can easily be measured *in situ* at high frequency using sensors. However, stream water quality data often exhibit skewed distributions, nonlinear relationships, and multicollinearity, all of which can be problematic for linear-regression models. Here, we use a flexible and robust machine learning technique, Random Forests Regression (RFR), to estimate stream nitrogen (N) and phosphorus (P) concentrations from sensor data within a forested, mountainous drainage area in upstate New York. When compared to actual nutrient data from samples tested in the laboratory, this approach explained much of the variation in nitrate (89%), total N (85%), particulate P (76%), and total P (74%). The models were less accurate for total soluble P (47%) and soluble reactive P (32%), though concentrations of these latter parameters were in a relatively low range. Although soil moisture and fluorescent dissolved organic matter are not commonly used as surrogates in nutrient-regression models, they were important predictors in this study. We conclude that RFR shows great promise as a tool for modeling instantaneous stream nutrient concentrations from high-frequency sensor data, and encourage others to evaluate this approach for supplementing traditional (laboratory-determined) nutrient datasets.

Keywords: surrogate; regression; Random Forests; machine learning; nutrients; stream chemistry

1. Introduction

Nutrients are an important component of flowing waters. Nutrient availability limits ecosystem production and affects oxygen dynamics (Kalff, 2002), which influences food supply, habitat quality, and thus the community structure of aquatic invertebrates and fish (Miltner and Rankin, 1998). Moreover, toxic effects on stream biota can occur at high concentrations of nitrate (Camargo et al., 2005), nitrite (Eddy and Williams, 1987) and ammonia (Richardson, 1997). From a lake management perspective, flowing waters represent conduits by which nutrients are conveyed from catchments to receiving water bodies (Meybeck, 1982). The control of eutrophication requires a reduction in the catchment export of nutrients that limit aquatic primary production, namely nitrogen (N) and phosphorus (P) (Schindler, 1977; Elser et al., 2007). While a reduction in P loading alone may be the most efficient way to control eutrophication (Lewis and Wurtsbaugh, 2008; Schindler et al., 2008), this principle is not universally accepted (Paerl et al., 2016), and, in practice, information on both N and P loadings is generally desired by lake and watershed managers and scientists. Efforts to improve the water quality of the Laurentian Great Lakes, for example, involve large-scale tributary monitoring programs to quantify temporal and spatial variation in catchment export of both N and P (e.g., Robertson et al., 2018).

A challenge in monitoring stream nutrients is that the chemistry of flowing waters is highly dynamic (Blaen et al., 2016), exhibiting variation at diel, seasonal, and interannual time scales (Halliday et al., 2012; Pellerin et al., 2012). In addition to such periodic variability, large changes in the N and P concentrations of lotic ecosystems can occur rapidly due to episodic alterations in hydrology. Where predominantly supplied by a steady point source (e.g., a wastewater treatment plant) or a steady non-point source (e.g., groundwater), nutrient ion concentrations may decline markedly following rain events due to dilution (Bowes et al., 2015). Conversely, high-flow events are generally associated with large increases in nonpoint-source particulate nutrient inputs and concentrations (e.g., Correll et al., 1999; Bieroza and Heathwaite, 2015). For instance, urban stormwater (Rodak et al., 2019) and the erosion of agricultural soils (Cooper et al., 2015) often contribute P-rich particulate matter to streams.

Given that stream nutrients are highly dynamic, the accuracy of nutrient loading estimates may suffer due to the relatively low (weekly to monthly) frequency typical of stream nutrient data obtained by laboratory analysis of manually-collected water samples (Johnes, 2007; Cassidy and Jordan, 2011). As the use of continuous water quality sensors in streams increases (Rode et al., 2016), many studies have employed a “surrogate-regression” approach by

which nutrient concentrations are estimated from environmental surrogates (Horsburgh et al., 2010) that can be quantified *in situ* at high-frequency (hourly or less), such as discharge, turbidity, or specific conductance (Jones et al., 2011; Robertson et al., 2018; Villa et al., 2019). Most studies employing the surrogate-regression approach have used simple or multiple linear regression analysis (e.g., Christensen et al., 2002; Jones et al., 2011; Lannergård et al., 2019). Often, transformations must be applied to the (skewed) water quality data to normalize the error distributions to apply parametric statistical techniques (Christensen et al., 2002; Roberston et al., 2018). Additionally, issues with autocorrelation and covariation among predictor variables represent violations of the assumptions of these analyses (Zar, 1999) and have led researchers to explore alternative approaches. For instance, Lessels and Bishop (2014) employed mixed-effects models for TN and TP using turbidity and discharge as predictors, while Leigh et al. (2019) used mixed-effects modeling to predict NO_x from turbidity, conductivity, and water level.

Machine learning (ML) methods can be a useful alternative to parametric statistical models when the underlying assumptions of the latter are difficult to satisfy, such as when non-linear and/or non-monotonic relationships are present. ML is now widely used in many areas of water research; for instance, to predict flooding (Mosavi et al., 2018), stream flow (reviewed by Yaseen et al., 2015), groundwater pollution (Rahmati et al., 2019), phytoplankton abundance (Wilson and Recknagel, 2001) and photophysiology (Lucius et al., 2020), benthic invertebrate community structure (Jones et al., 2017), and lake trophic status (Hollister et al., 2016). In a comprehensive review, Maier et al. (2010) noted that of the 210 articles published between 1999 and 2007 that used artificial neural networks (ANNs; a popular class of ML methods) within the context of water resources, almost all were related to flow prediction and very few to water quality. Of such water quality studies, many have focused on constituents other than nutrients, such as suspended sediment (Demirci and Baltaci 2013; Afan et al., 2016), ions and related parameters such as total dissolved solids, salinity and/or conductivity (Maier and Dandy, 1996; Asadollahfardi et al., 2012; Orouji et al., 2013), dissolved oxygen (Wen et al., 2013; Ay and Kisi, 2017), biological and chemical oxygen demand (Dogan et al., 2009; Ay and Kisi 2014), toxic metals (Chang et al., 2014), and pH (Moatar et al., 1999; Ahmed et al., 2019).

Most studies using ML (often ANNs) to model lotic nutrients (N and P) have predicted daily mean concentrations (Suen and Eheart, 2003; Nour et al., 2006), catchment export/loadings (Schärer et al., 2006; Anctil et al., 2009), or spatial variation related to watershed features (Lek et al., 1999; Jones et al., 2017). We are aware of only one

existing study, that of Castrillo and García (2020), that has applied the flexible, robust, and relatively simple (i.e., compared with ANNs) ML technique known as Random Forests Regression (RFR; Breiman, 2001) to the surrogate-regression method for estimating instantaneous stream nutrient concentrations from continuous *in situ* sensor data.

RFR is an extension of regression trees (Breiman et al., 1984), which is a nonparametric technique that is highly flexible and robust; it can accommodate nonlinear relationships, high-order interactions, and missing values (De'ath and Fabricius, 2000). While useful in describing data, regression trees are prone to overfitting the training dataset and are thus poor at prediction (De'ath, 2007). As an ensemble method, RFR involves the creation of a large set of trees (typically hundreds or several thousand) that together constitute the model; predictions from the set of trees are averaged. The technique uses bootstrap aggregation (Breiman, 1996) and a random subset of predictors for splitting at each node of each tree to limit overfitting and thus improve predictive accuracy (Breiman, 2001).

In this study of a forested, temperate drainage basin in New York State, USA, we: (1) use RFR to estimate instantaneous lotic stream concentrations of N and P, based on high-frequency sensor data; (2) compare the predictive accuracy of RFR models for different N and P fractions; and (3) discuss the importance of various high-frequency sensor parameters (related to water quality and hydrology) as predictors in the models.

2. Methods

2.1. The Lake George drainage basin

Lake George (43°35′ N, 73°35′ W) is a large (114 km^2), deep (mean depth = 18 m; max. depth = 58 m; volume = 2.1 km^2) dimictic, oligotrophic lake in the Adirondack Mountain region of New York State, USA. The drainage basin geology is mostly shallow sandy till overlaying bedrock, with many granite outcrops and large boulders. The catchment is predominantly forested, though parts of the lake's shoreline are relatively well-developed (i.e., urbanized; Table 1). The lake has a relatively small surrounding drainage area of 492 km^2 and receives an estimated 57–75% of its water loading in the form of surface water inflow (Stearns and Wheler, 2001). There are 3 wastewater treatment plants (WWTPs) within the basin (Figure 1). The largest of these is the Village of Lake George WWTP. This facility services approximately 1,350 residents in winter and up to approximately 20,000 during the summer season; it is the only WWTP located within a monitored subwatershed (that of West Brook; Figure 1). The total annual P loading to the lake was calculated to be 10,163 kg/year, 83% of which is estimated to

be delivered via surface water inputs (Stearns and Wheler, 2001). Further details on Lake George and its drainage basin can be found in Boylen et al. (2014).

2.2. Stream monitoring

As part of a larger lake and watershed monitoring effort known as The Jefferson Project, we monitored 11 major tributaries of Lake George (Figure 1) during 2018–19. Using both manual and high-frequency monitoring, we quantified water chemistry, discharge, and soil characteristics. Monitoring stations (Table 1) are located close to the stream outlets (confluences with the lake) except for the station on Northwest Bay Brook (Figure 1) which is located farther upstream, at a historical United States Geological Survey (USGS) site (#04278300).

Baseflow and storm event samples were collected monthly at all sites and analyzed for a suite of water quality parameters at our water-testing laboratory. The storm event samples were collected using auto-samplers (Teledyne ISCO, Portable Sampler 3700): 12 samples (each comprising two 1-L bottles) were automatically collected at a 2–4 h interval over the course of each event. Of these, the 4–6 samples that corresponded to key phases of the storm hydrograph (i.e., pre-event, rising limb, peak flow, falling limb) were returned to the laboratory for analysis. In this study, the water quality parameters we focus on are total phosphorus (TP), total soluble phosphorus (TSP), particulate phosphorus (PP), soluble reactive phosphorus (SRP), nitrate (NO_3), and total nitrogen (TN).

For the high-frequency sampling, we have deployed sondes (YSI EXO2), pressure transducers (WaterLOG H-3123 or Campbell Scientific CS451), and soil moisture probes (Stevens HydraProbe; absent from Finkle and Northwest Bay Brooks) on each stream. Sondes were equipped with sensors for water temperature (Temp), pH, specific conductance (SpCond), turbidity (Turb), and fluorescence dissolved organic matter (fDOM). Sondes and pressure transducers were housed within protective piping and mounted where water depth was sufficient for year-round data collection. At some locations (East and West Brooks), warmer water temperatures allowed for in-stream deployment year round. The remaining sites freeze during winter, preventing in-stream sonde deployment. Two streams (Indian and Sunset Brooks) were equipped with flow cell pumping systems (see Wagner et al., 2006) to pump water from beneath the ice to the sonde mounted in a heated shelter. Sensor data were not collected from the remaining sites during winter. Pressure transducers were left in place throughout the year at all sites. Soil moisture probes were installed horizontally just below the O-horizon, between 15 cm and 26 cm in depth.

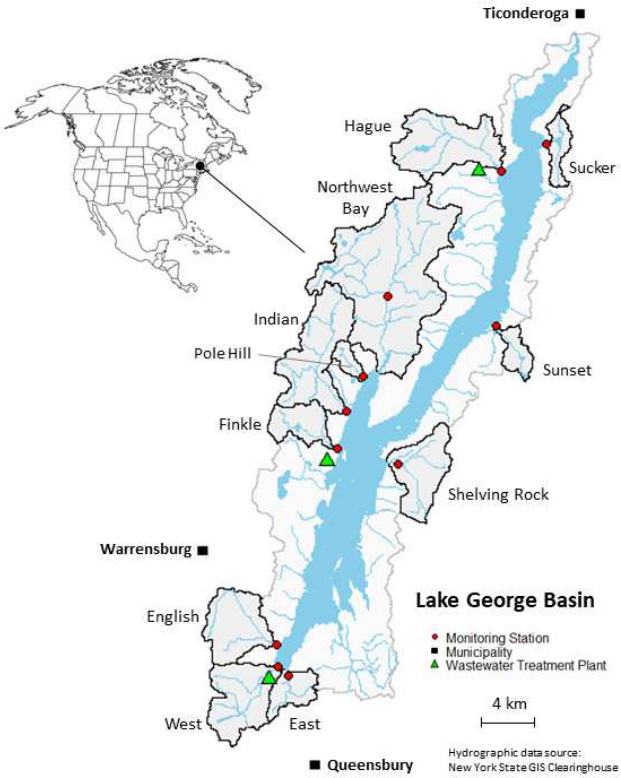


Figure 1. The Lake George drainage basin (light grey), drainage areas of monitored tributaries (darker grey), tributary monitoring stations (red dots), wastewater treatment plants (green triangles), and largest nearby municipalities (black squares).

Table 1. Stream drainage area, monitoring station and data availability information. Catchment characteristics are from U.S. Geological Survey (2016).

		Station Location		Drainage Area Upstream of Station					
Watercourse	**Data Availability[1]**	**Lat.**	**Long.**	**Area (ha)**	**% Total[2]**	**Forested (%)[3]**	**Urban (%)**	**Impervious (%)**	**Mean Basin Slope (m/km)**
East Brook	Jan–Dec (2018–19)	43.4130	-73.7001	790	92.4	89.9	16.1	3.61	185
English Brook	Jul–Dec (2018); Apr–Nov (2019)	43.4334	-73.7105	2,012	99.6	94.1	9.39	1.57	218
Finkle Brook	Jan, Mar–Dec (2018); Apr–Dec (2019)	43.5634	-73.6526	1,088	99.9	92.7	8.89	0.46	171
Hague Brook	Apr–Dec (2018–19)	43.7451	-73.4978	2,771	99.5	97.2	3.20	0.26	179
Indian Brook	Apr–Dec (2018); Jan–Dec (2019)	43.5875	-73.6438	3,004	99.5	97.3	3.34	0.28	160
Northwest Bay Brook	Mar–Oct (2018); Apr–Nov (2019)	43.6632	-73.6034	5,698	71.9	98.3	0.31	<0.01	225
Pole Hill Brook	Apr–Dec (2018–19)	43.6105	-73.6278	344	96.3	98.6	0.39	0.05	216
Shelving Rock Brook	Apr–Dec (2018); Apr–Nov (2019)	43.5521	-73.5964	1,707	89.2	98.4	1.17	0.02	218
Sucker Brook	Apr–Dec (2019)	43.7624	-73.4561	627	99.6	86.4	3.74	0.22	174
Sunset Brook	May–Dec (2018); Jan–Dec (2019)	43.6424	-73.5049	596	98.9	96.7	3.00	0.42	199
West Brook	Jan–Dec (2018–19)	43.4189	-73.7094	2,225	99.7	93.1	9.32	2.19	197
		Total or Average:		**20,862**	**95.1**	**94.8**	**5.35**	**0.83**	**195**

[1]Months for which both chemistry and sonde data were available; [2]area upstream of the monitoring station as a percent of the total drainage area for that stream; [3]includes treed urban areas.

144 *2.3. Chemical Analyses*

145 All chemical analyses of stream samples were performed at the Darrin Fresh Water Institute (DFWI) analytical
146 laboratory according to standard protocols. All calibration solutions were prepared in the laboratory from reagent-
147 grade materials and verified against certified reference solutions (SPEX CertiPrep, Metuchen, NJ). P concentrations
148 were determined colorimetrically (Method 365.2; U.S. Environmental Protection Agency, 1979). SRP was
149 determined on filtered samples (0.45-µm pore-size 47-mm diameter polycarbonate (Poretics) filters) within 24 h of
150 sample collection. TP was determined following persulfate digestion of whole-water samples. TSP was determined
151 after persulfate digestion of filtered water (0.45-µm pore-size 47-mm diameter polycarbonate (Poretics) filters). PP
152 was calculated as the difference between TP and TSP. TN was determined colorimetrically following persulfate
153 oxidation (Langner and Hendrix, 1982). NO_3 was determined by ion chromatography (Method 300; U.S.
154 Environmental Protection Agency, 1979).

155 *2.4. Data Quality Assurance & Quality Control*

156 Method detection limits (MDLs) for chemical analyses have been established by DFWI as part of its QA/QC
157 procedures and are equivalent to those described in Standard Methods (American Public Health Association, 2017).
158 The MDLs for the nutrient parameters included in this study are: 1 µg-P/L for TP, TDP, and SRP, 0.01 mg-N/L for
159 NO_3, and 0.05 mg-N/L for TN. Values below the MDL (non-detects) were replaced by half the MDL for all
160 analyses. The dataset used for this study had the following number of non-detects: 0 for TP, 2 for TSP (<1%), 4 for
161 TN (1%), 43 for NO_3 (11%), and 63 for SRP (15%).

162 The water quality sensors were calibrated on a monthly basis. All calibrations followed the manufacturer's
163 recommendations with one exception: EXO2 fDOM sensors were calibrated with a narrower calibrant range (0–40
164 RFU) than manufacturer recommendation (0–100 RFU) due to the relatively low fDOM levels observed in the
165 watershed (see Results). Sensors were inspected regularly, and maintenance or repair performed when required.

166 Data collection intervals ranged from 1–15 min depending on the location. Sensor data were flagged using an
167 automated Python flagging program and removed when necessary. Data from the entire sensor network were
168 reviewed monthly for quality via visual inspection of time-series plots. For each ambient temperature (T_{meas}),

fDOM data were corrected to values at a standard temperature (T_{ref}) of 22°C using the equation of Watras et al. (2011):

$$fDOM = fDOM_{uncorrected} \,/\, \left(1 + \left(k \times \left(T_{meas} - T_{ref}\right)\right)\right)$$

The temperature-specific coefficient of fluorescence (k) was determined to be -0.0075°C^{-1} based on laboratory studies with our fDOM sensors.

One-hour (centered) rolling medians were calculated at a 1-minute interval to fill data gaps in the time series and to reduce the influence of outliers. From these data, 1-h, 5-h, 24-h, and 120-h lags (e.g., $Temp_{t\text{-}5h}$ represents the temperature value at 5 hours before time *t*) and deltas (i.e., change in value over lag period; e.g., $Turb_{\Delta 24h}$ represents the change in turbidity over the past 24 hours) were calculated for use as additional predictor variables. The 1-h rolling medians that paired temporally with the nutrient chemistry data were used for all analyses and visualizations.

2.5 Data Exploration, Visualization, and Correlation Analysis

All data analyses and visualizations were performed using the R statistical software (R Core Team, 2019). Boxplots were created to display the nature of the data distributions of the nutrient parameters and how these varied by station and by month of year. A correlation matrix and biplots were produced to visualize the relationships between all parameters. Results of Spearman's test of rank correlation were reported because most of the variables had highly-skewed (non-normal) data distributions (Hollander and Wolfe, 1973).

2.6. Model Fitting and Cross Validation

Random Forests Regression (RFR) was used to predict manually collected nutrient concentrations (TN, NO_3, TP, PP, TSP, SRP) from high-frequency sensor data. The full set of predictors included time of day, soil moisture (SM), hydrostatic pressure (HP), Temp, Turb, fDOM, SpCond, pH and the 1-h, 5-h, 24-d, and 120-d lags and delta values of the sensor variables (e.g., $fDOM_{t\text{-}1h}$, $SpCond_{\Delta 5h}$, etc.).

2.6.1. Cross Validation

191 Model accuracy was assessed using a modification of leave-one-out cross validation (LOOCV; e.g. Stone, 1974;
192 Efron and Gond, 1983), the most classical exhaustive validation procedure (Arlott and Celisse, 2010). The simple
193 hold out validation method (splitting the dataset once into training and testing sets) was not used because it can
194 generate a wide range of results when applied to a relatively small dataset (e.g., several hundred cases), depending
195 on which data are used for model training vs. testing. We confirmed the existence of this variability using Monte
196 Carlo simulation, and note that the median results from 1000 random test/train splits (2/3 for testing, 1/3 for
197 training) using the simple hold out method were comparable to those obtained using LOOCV (i.e., mean absolute
198 error (MAE) differed by only 2–9%, depending on the nutrient parameter).

199 LOOCV involves omitting one case (row) at a time from the dataset while training (fitting) the model, then
200 predicting the value for the omitted case. This is done *n* times, where *n* is the number of values of the dependent
201 variable. We modified this approach slightly, omitting data from one sampling event at a time (i.e., 1 case from
202 baseflow dates, multiple cases from storm event dates when multiple observations were made on the same
203 day). Thus, data collected from a particular site during a storm event were never used in both the training and
204 testing datasets to eliminate potential autocorrelation between data used in model training and testing. We
205 compared the predicted values from LOOCV to the observed values using biplots and quantified model accuracy as
206 MAE, root mean squared error (RMSE), and Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970). NSE
207 expresses model error (the deviation between predicted and observed values) relative to the total variation in the
208 response variable; it ranges from 1 to -∞ with a value of 1 indicating no difference between predicted and observed
209 values, whereas lower values represent poorer predictive accuracy. NSE is calculated, similarly to the coefficient of
210 determination (R^2), as:

211 $$NSE = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

212 where y_i and $\hat{y}_i$ are the observed and predicted values, respectively, and $\bar{y}$ is the average of the observed values. The
213 above metric, referred to as NSE in the hydrology literature, is described as a "pseudo R-squared" in the
214 randomForest R package documentation (Liaw and Weiner, 2002), and has elsewhere been called the "mean square
215 deviation ratio" (Lessels and Bishop, 2013) or simply "efficiency" (Holmberg et al., 2006).

216 *2.6.2. Random Forests Regression (RFR)*

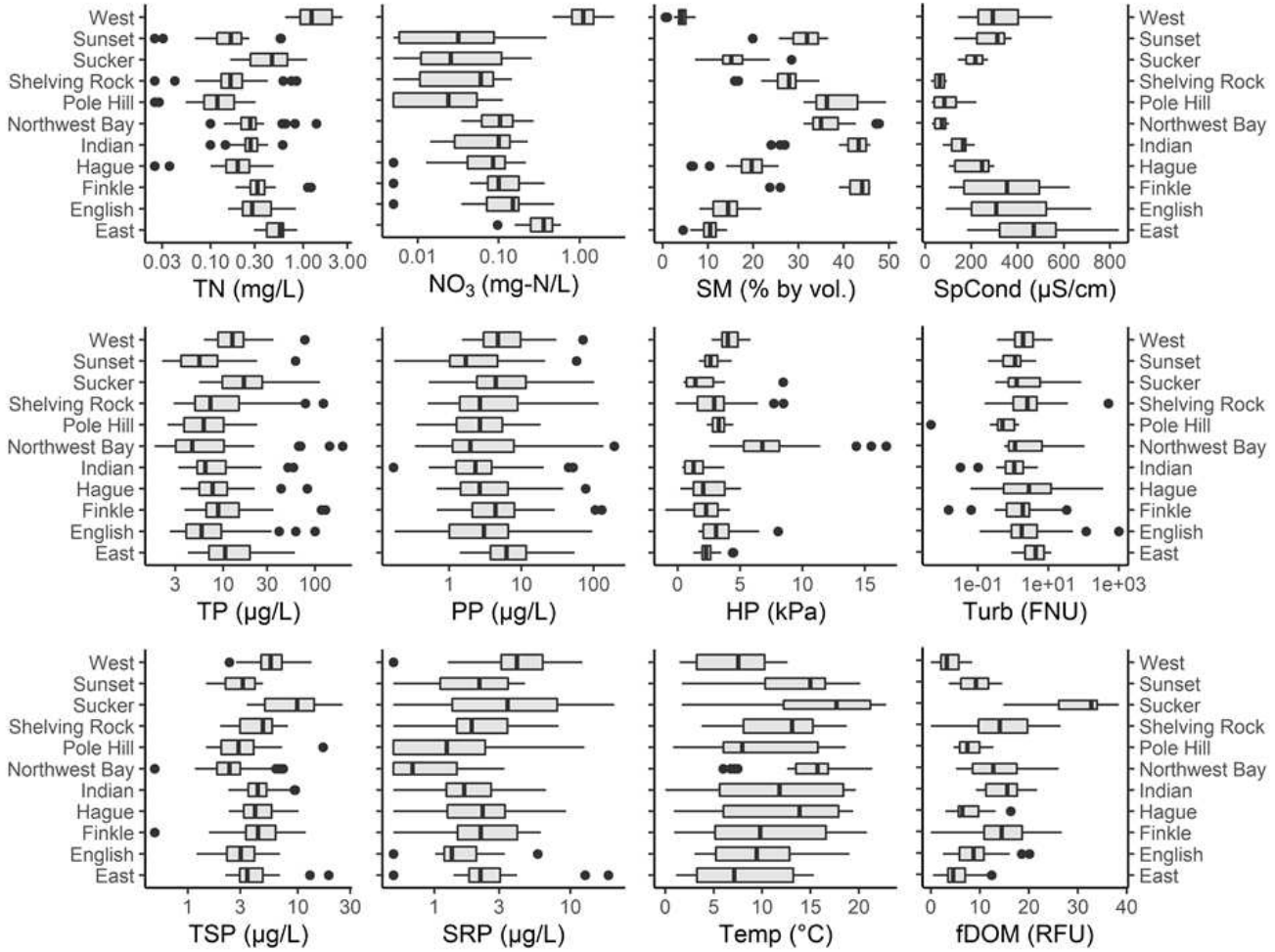
217 Exploratory analysis of the nutrient and sensor data revealed a high degree of correlation among the sensor variables
218 (i.e., multicollinearity among the potential predictors), nonlinear relationships between some of the sensor and
219 nutrient variables, and highly skewed data distributions (see Results). For these reasons, multiple linear regression
220 analysis was eschewed in favor of a more flexible and robust regression approach, RFR. Random Forests of 1,000
221 trees were created using the R package *randomForest* (Liaw and Weiner, 2002). Predictor importance was
222 quantified as the percentage increase in model mean squared error (MSE) when the predictor was permuted. The
223 first model fit for each nutrient included the full set of predictors (M = 57); the model was then simplified in a
224 stepwise fashion by removing the predictor of lowest importance after each fit. The best model (highest NSE
225 determined via LOOCV) was retained. If NSE was equivalent at 3 significant figures, the simpler model (i.e., the
226 model with fewer predictors) was retained. Model simplification was performed using the default value for F, the
227 number of predictors randomly chosen for splitting at each node (p/3, where p is the number of predictors included
228 in the model). Following model simplification, the optimal value for F was determined by varying F between 1 and
229 M-1. The effect of limiting the maximum tree depth (total number of splitting nodes) was assessed over a range of 1
230 to the maximum possible number of nodes for the best model for each nutrient. We chose not to constrain tree depth
231 because there was no evidence of a decrease in model accuracy with increasing tree depth.

232 ## 3. Results

233 *3.1. Stream water quality*

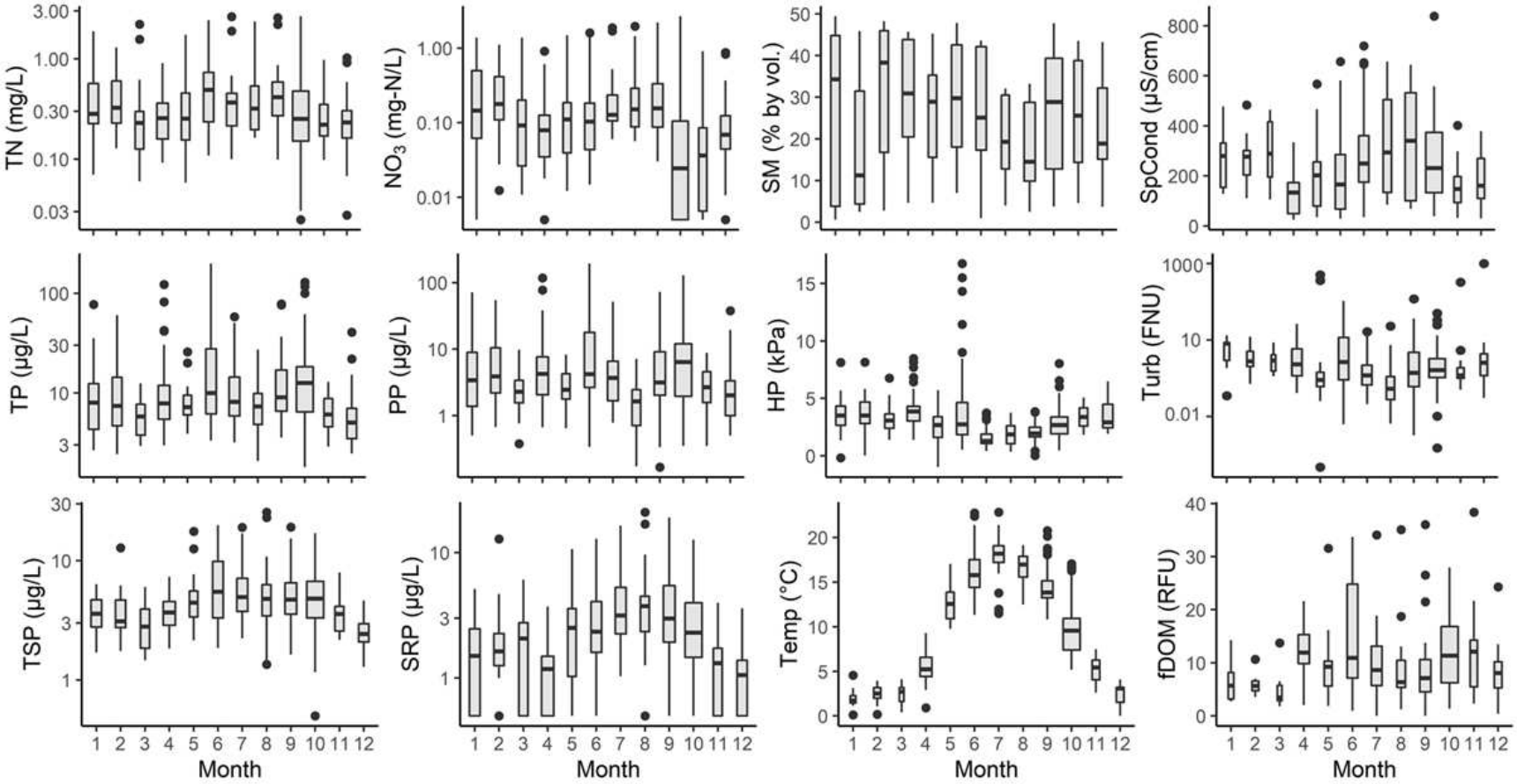
234 Nutrient concentrations were relatively low at all sites (Figure 2). Nitrogen (especially NO_3) concentrations were
235 generally higher in West Brook than in the other streams (Figure 2). SM and SpCond were highly variable among
236 sites. Average measured SM ranged from 4% at West Brook to 42% at Indian Brook. Average SpCond ranged
237 from 62 µS/cm at Shelving Rock Brook to 452 µS/cm at East Brook. Month-to-month variation in nutrient
238 concentrations was low relative to the magnitude of variation within months (Figure 3). Concentrations of TSP,
239 and, to a greater extent, SRP, exhibited some seasonality, with higher concentrations in the summer months (Figure
240 3). Water temperature exhibited pronounced seasonality; the monthly average was greatest in July, at 17.7°C. SM

241 decreased between Jun and Sep whereas SpCond increased over the same months. HP, Turb, and fDOM were all
242 relatively high and/or variable in Apr, Jun, and Oct.



243

244 **Figure 2.** Stream nutrient concentrations (columns 1 & 2) and sensor readings (columns 3 & 4) for monitored Lake George tributaries (2018-19). Boxes span
245 the 1st to 3rd quartiles; data beyond 1.5 × the interquartile range from the 1st and 3rd quartiles are shown as points; vertical bars represent the medians. The
246 thickness of the boxes is proportional to the number of data for each station. Note the use of a logarithmic scale for the X axes of the nutrient and turbidity plots.



247

248 **Figure 3.** Seasonal (monthly) variation in stream nutrient concentrations (columns 1 & 2) and sensor readings (columns 3 & 4) during 2018-19 for monitored
249 Lake George tributaries (all 11 sites pooled). Boxplots represent data as in Figure 2. Note the use of a logarithmic scale for the Y axes of the nutrient and
250 turbidity plots.

3.2. Correlations among water quality parameters

Based on Spearman's test of rank correlation, there was a large number of statistically-significant ($p < 0.05$) associations (many of them highly significant; i.e., $p < 0.001$) among the nutrient variables and 1-h running medians of the sensor variables (Figure 4). SpCond was positively correlated with NO_3, TN, and SRP. Temp was positively correlated with SRP, TSP, and TP. Turbidity was positively correlated with TN, TSP, PP, and TP and negatively correlated with SRP. pH was positively correlated with NO_3, TN, SRP, and TSP and negatively correlated with PP. fDOM was positively correlated with TP, PP, and TSP, and both fDOM and SM were negatively correlated with SRP, TN, and NO_3. HP was positively correlated with PP and negatively correlated with SRP.

3.3. RFR models

Predictive accuracy of the RFR models varied by parameter (Table 2); NSE ranged from 0.89 for NO_3 to 0.32 for SRP. There was low systematic error in model estimates of the nutrient parameters; i.e., points were scattered relatively symmetrically around the 1:1 lines (Figure 5). However, the models did tend to slightly overestimate concentrations at the low ends of the distributions; this is particularly evident for SRP, which had the most values below the MDL (Figure 5).

The size and composition of the predictor set and the relative predictor importances varied considerably among models (Table 2). Antecedent soil moisture ($SM_{t\text{-}24h}$, $SM_{t\text{-}5h}$) was an important predictor of both NO_3 and TN. The change in HP over the 5 h preceding sample collection ($HP_{\Delta 5h}$) was the most important predictor of both PP and TP. Predictors based on fDOM (i.e., fDOM and its lags and deltas) were retained in the final models for all nutrients except for PP. Predictors based on turbidity were retained in the final models for all nutrients except for those in dissolved form (NO_3, TSP, SRP). Temp, as 1-h and 5-h lags, was retained as a predictor only in the SRP and TN models, respectively. SpCond was retained only as a predictor of NO_3. Time of day and pH were not predictors in any of the best models.

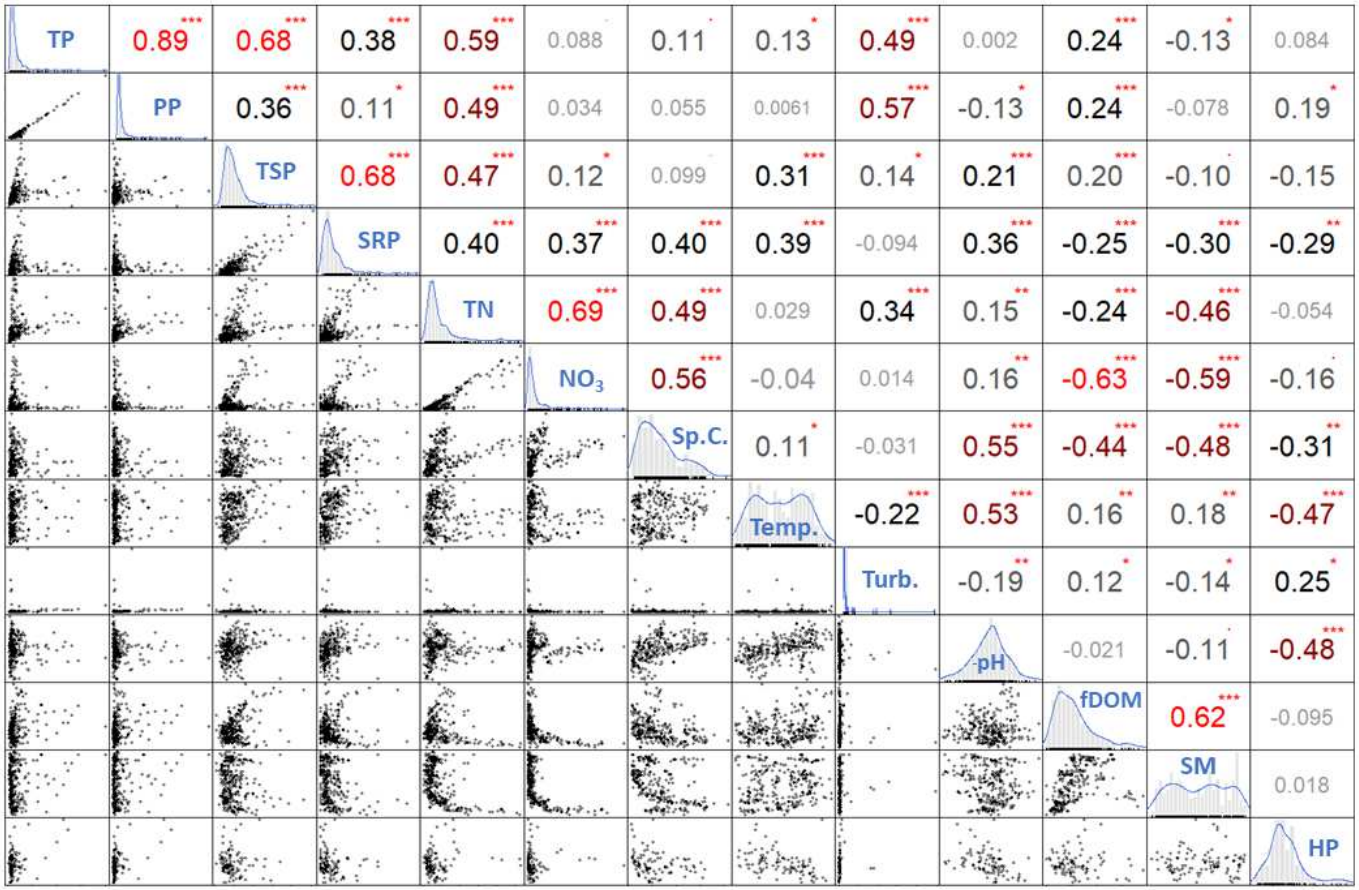


Figure 4. Spearman correlation matrix (upper-right), histograms (diagonal), and biplots (lower-left) for the sensor parameters (1-h rolling medians) and nutrients monitored in 11 Lake George tributaries (2018-19). Rank correlation coefficients (ρ) are color-coded according to magnitude as light grey ($|\rho| \leq 0.1$), dark grey ($0.1 < |\rho| \leq 0.2$), black ($0.2 < |\rho| \leq 0.4$), dark red ($0.4 < |\rho| \leq 0.6$), and bright red ($|\rho| > 0.6$). Statistical significance of rank correlations are denoted by a dot for $0.05 < p < 0.1$, and asterisks for $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

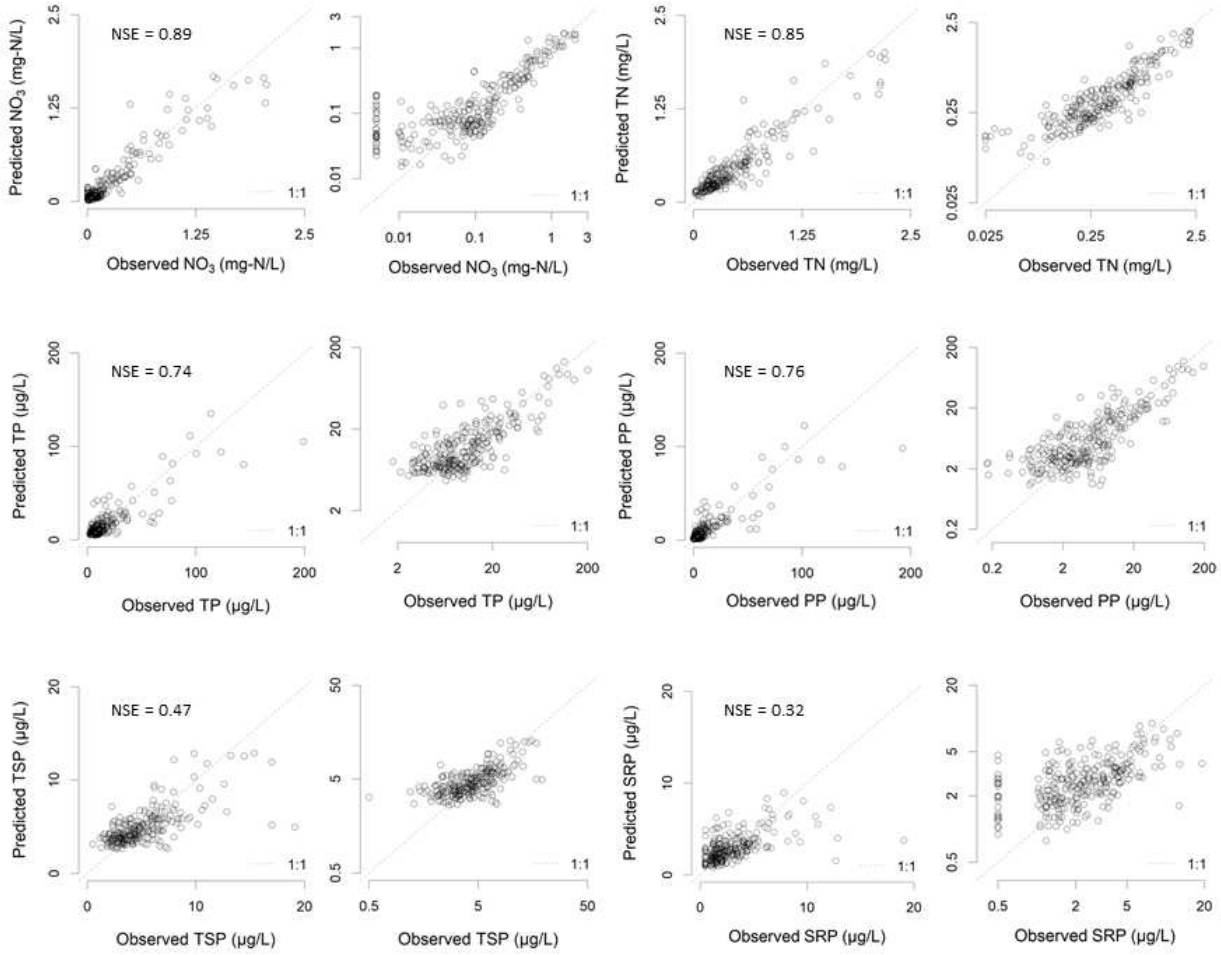


Figure 5. Biplots of leave-one-out cross validation of RFR models for nutrients. Model performance is quantified as NSE (a "pseudo-R^2"). Note the use of both linear scaling (columns 1 & 3) and logarithmic scaling (columns 2 & 4).

285 **Table 2**. RFR models of nutrient concentrations based on sensor data. Model performance is expressed as mean absolute error (MAE), root mean
286 squared error (RMSE), and Nash-Sutcliffe Efficiency (NSE). Predictors are colored according to their Spearman rank correlations with the
287 nutrients as boldface (positive correlation; $p < 0.05$), red (negative correlation; $p < 0.05$), and italicized (no significant rank correlation; $p \geq 0.05$)
288 text.

Nutrient	*n*	F	MAE*	RMSE*	NSE	Predictors (importance as % increase in model error when permuted)
NO_3	273	2	0.075	0.12	0.89	$SM_{t\text{-}24h}$ (23%), $SM_{t\text{-}5h}$ (21%), SM (21%), $HP_{t\text{-}1h}$ (21%), SpCond (20%), *HP* (20%), $fDOM_{t\text{-}5h}$ (19%)
TN	257	4	0.11	0.16	0.85	$SM_{t\text{-}24h}$ (21%), $\mathbf{HP_{\Delta 24h}}$ (20%), $fDOM_{t\text{-}5h}$ (20%), $SM_{t\text{-}1h}$ (18%), $\mathbf{HP_{t\text{-}5h}}$ (18%), $Temp_{t\text{-}5h}$ (18%), $SM_{t\text{-}5h}$ (18%), $HP_{t\text{-}1h}$ (17%), $\mathbf{Turb_{\Delta 24h}}$ (17%), $SM_{t\text{-}120h}$ (17%), $fDOM_{t\text{-}24h}$ (16%), $\mathbf{HP_{\Delta 5h}}$ (16%), $HP_{\Delta 120h}$ (16%)
PP	276	4	5.0	10	0.76	$\mathbf{HP_{\Delta 5h}}$ (41%), $\mathbf{HP_{\Delta 1h}}$ (20%), $\mathbf{Turb_{\Delta 24h}}$ (18%), **Turb** (16%), $HP_{\Delta 24h}$ (15%), $HP_{\Delta 120h}$ (11%)
TP	280	4	5.9	11	0.74	$\mathbf{HP_{\Delta 5h}}$ (37%), $\mathbf{fDOM_{\Delta 5h}}$ (20%), $\mathbf{HP_{\Delta 1h}}$ (18%), $\mathbf{Turb_{\Delta 24h}}$ (18%), $\mathbf{HP_{\Delta 24h}}$ (17%), **Turb** (16%)
TSP	267	1	1.3	2.0	0.47	$\mathbf{SM_{\Delta 24h}}$ (28%), $fDOM_{t\text{-}24h}$ (27%), $HP_{t\text{-}120h}$ (26%), $\mathbf{fDOM_{t\text{-}1h}}$ (25%), **fDOM** (24%)
SRP	285	1	1.2	2.0	0.32	$fDOM_{t\text{-}5h}$ (21%), fDOM (20%), $HP_{t\text{-}24h}$ (20%), $fDOM_{t\text{-}1h}$ (20%), $\mathbf{Temp_{t\text{-}1h}}$ (18%)

289 ** units are mg-N/L for NO_3 and TN and ug/L for TP, PP, TSP, and SRP*

290 ## 4. Discussion

291 We used RFR to predict instantaneous N and P concentrations in streams based on a suite of chemical and physical
292 parameters measured by high-frequency sensors. We found fDOM and SM, which have not (to our knowledge)
293 previously been included in surrogate-regression models, to be important predictors of total and dissolved nutrient
294 concentrations in streams within the Lake George catchment. Model accuracy was high for NO_3 and TN,
295 intermediate for TP, PP, and TSP, and lowest for SRP.

296 *4.1. Stream nutrient concentrations and model performance*

297 Nutrient concentrations in the study area were relatively low, as would be expected for streams draining a largely
298 unimpacted watershed (Kalff, 2002). Average TP concentrations were between 7 and 23 µg/L on a stream-specific
299 basis, comparable to values from other forested catchments in the temperate zone (e.g., Eimers et al., 2009; Viviano
300 et al., 2014). Most research using the surrogate-regression approach has been conducted in watersheds with much
301 higher mean TP concentrations. For example TP concentrations were 30 and 60 µg/L for two subcatchments of Lake
302 Burragorang, Australia (Lessels and Bishop, 2013), 72 µg/L for a mixed-land-use catchment in east central Sweden
303 (Lannergard et al., 2019), 79 and 256 µg/L for two sites on the irrigation-regulated Little Bear River in Utah, USA
304 (Jones et al., 2011), 9–358 µg/L for Laurentian Great Lakes tributaries (Robertson et al., 2018), and 102 µg/L for the
305 urbanized Labrone watershed in Italy (Viviano et al., 2014). When explaining variance in TP as a function of
306 turbidity and conductivity at a large number of sites, Villa et al. (2019) noted a positive correlation between
307 goodness of fit (R^2) and percent agricultural land use in the respective catchment, and a negative correlation between
308 R^2 and the proportion of forested land in the catchments. In another ML study of a forested catchment, Holmberg et
309 al. (2006) used ANNs to estimate TP concentrations in a low range (1–60 µg/L) based on meteorology, runoff, and
310 catchment characteristics, and obtained similar results (NSE = 0.75) to those we report. Working in boreal Canada,
311 Nour et al. (2006) estimated mean daily TP concentrations based on (modeled) streamflow using ANNs and
312 achieved comparable model accuracy (cross-validation R^2 = 0.78–0.82), though maximum TP concentrations were
313 much higher than in our study (i.e., frequently >200 µg-P/L). The other study which applied RFR to the surrogate
314 regression approach, that of Castrillo and García (2020), modeled TP of a highly urbanized tributary of the Thames
315 River, UK where the mean TP concentration was 640 µg/L. Using conductivity, Turb, Temp, pH, and flow as

316 predictors, their RFR model had a RMSE of 110 µg/L or 17% of the mean observed TP concentration; for
317 comparison, our TP model had a RMSE of 11 µg/L or 78% of a (much lower) observed mean TP of 14 µg/L. We
318 expect that the RFR approach would have higher predictive accuracy, especially for TSP and SRP, when applied to
319 a dataset with P concentrations that are higher and span a wider range, such as one from an agricultural or urbanized
320 watershed.

321 N concentrations in the Lake George catchment were in a moderate range (TN averaged 0.14–1.44 mg/L by site),
322 and are comparable to those observed in some other surrogate-regression studies (e.g., Ryberg, 2006; Kim and
323 Furumai, 2013; Robertson et al., 2018). As was the case for TP, Holmberg et al. (2006) obtained similar model
324 accuracy (NSE = 0.82) to that reported here (NSE = 0.85) when predicting TN within a similar range (0.1–1.4 mg/L)
325 using an ANN. Another study that used an ANN model to predict TN based on land use, hydrology, and
326 meteorology—that of He et al. (2011)—reported higher model accuracy (cross validation $R^2 = 0.93$) than we
327 observed, though TN concentrations were much higher (approximately 0–10 mg/L) than in this study. Using RFR,
328 Castrillo and García (2020) predicted NO_3 based on conductivity, pH, Temp, and stream flow and obtained a RMSE
329 of 0.194 mg-N/L or 4.8% of the mean concentration of 4.06 mg-N/L; RMSE of our NO_3 model was lower at 0.124
330 mg-N/L, though much higher as a percentage (52%) of a much lower mean concentration of 0.237 mg-N/L. The
331 somewhat higher accuracy of the N models relative to the P models presented here may reflect the contrasting
332 importance of dissolved vs. particulate nutrient fractions and their respective sources, as discussed below.

333 *4.2. Fluorescent dissolved organic matter as a predictor of nutrients*

334 fDOM varied temporally and spatially, and was strongly positively correlated with SM, suggesting that catchment
335 hydrology was a strong influence on fDOM concentration (Pellerin et al. 2012). fDOM was retained as a predictor
336 variable in all models except for the PP model; it was the best predictor of SRP, and the second-best predictor of TP
337 and TSP. The positive (rank) correlations between fDOM and P parameters (other than SRP), and between fDOM
338 and SM, suggest that catchment export of organic matter from forest soils (Schiff et al., 1997; Raymond and Saiers,
339 2010) is an important source of P to Lake George tributaries. Wastewater treatment plant effluent is another
340 potential source of P and of fDOM (Hudson et al., 2007); however, we expect this to be a negligible source in this
341 area because of the low population density and predominance of forested land. Moreover, the only subwatershed

342 with a WWTP within its boundaries (West Brook) had very low fDOM (see Figures 1 & 2). Finally, if wastewater
343 were an appreciable contributor of fDOM, a negative rather than positive relationship would be expected between
344 fDOM and SM because wastewater influence should be higher during dry (low SM) conditions due to less dilution.

345 NO_3 and SRP were both negatively correlated with fDOM and with HP, suggesting that anion concentrations were
346 diluted during periods of high runoff, and that dissolved inorganic nutrients (NO_3, SRP) may be predominantly
347 supplied by groundwater. This result is consistent with previous studies that have noted lower NO_3 concentrations
348 during or following high flows (Shrestha et al., 2007; Pellerin et al., 2012; Duncan et al., 2017) and negative
349 correlation between DOM and NO_3 (Frei et al., 2020). While the 1-h running medians of fDOM and HP were not
350 correlated (Figure 4), changes in fDOM over various periods (1, 5, 24, 120 h) were positively correlated with
351 changes in HP over the same periods (e.g., $\rho = 0.70$, $p < 0.001$ for $fDOM_{\Delta 24h}$ and $HP_{\Delta 24h}$); this result is consistent
352 with the tight coupling between flow and *in situ* fDOM noted by Pellerin et al. (2012), and suggests increased DOM
353 export during high-flow events, as would be expected within a forested watershed (Buffam et al., 2001; Raymond
354 and Saiers, 2010).

355 Paired measurements of *in situ* fDOM and nutrient concentrations such as DON (Snyder et al., 2018) and NO_3
356 (Pellerin et al., 2012) from streams have been presented in the literature, and DOM data from discrete samples have
357 been used in regression models of lotic nutrient concentrations (Mayora et al., 2018) and of watershed-scale nutrient
358 attenuation rates (Frei et al., 2020). However, *in situ* fDOM (i.e., recorded continuously by a sensor) has not, to our
359 knowledge, been used as a predictor of instantaneous riverine nutrient concentrations. This may be because sensors
360 for fDOM are not in widespread use relative to those for temperature, SpCond, and turbidity, or because linear
361 relationships between fDOM and other nutrients (e.g., NO_3, see Pellerin et al., 2012) are not readily apparent. A
362 third possibility is that where inorganic forms of N and P predominate (e.g., Frei et al., 2020), it may be expected *a*
363 *priori* that the effectiveness of fDOM as a surrogate for N or P would be low, that is, proportionate to the fraction of
364 the nutrients found in dissolved organic form (i.e., as DON or DOP). However, this was not the case in our study:
365 fDOM was in fact an important predictor of inorganic nutrients, presumably because it is generally at low
366 concentration in groundwater relative to surface water in forested catchments (Schiff et al., 1997; Buffam et al.,
367 2001), and thus reflective of the relative contributions of surface vs. groundwater inputs (Frei et al., 2020).

368 *4.3. Soil moisture as a predictor of nutrients*

We are not aware of other studies that have used soil moisture as a predictor of stream nutrient concentrations via the surrogate-regression approach. However, working in a small agricultural in France, Anctil et al. (2009) noted that soil moisture was an important predictor ('input') in a neural-network model for daily NO_3 flux (i.e., loading, not concentration) though it was much less important than discharge. The inverse relationships we observed between SM and N parameters in the Lake George drainage area are consistent with another study performed in the region (Burns et al., 1998), that noted an inverse relationship between NO_3 and runoff during summer baseflow conditions, when deep groundwater flow was an important contributor to stream discharge. There was a seasonal component to the covariation we observed between SM and NO_3: SM declined over the summer months (presumably due to the relatively low surface runoff and high air temperatures that are characteristic of this time of year in the region) while NO_3 increased over this period. However, most of the variation in these parameters was spatial rather than temporal. SM was much lower, and NO_3 much higher, in the southern portion of the Lake George catchment, at West and East Brooks. This is consistent with evidence that groundwater NO_3 in this area of the watershed is elevated due to seepage of treated municipal wastewater that is discharged into sand infiltration beds located within the West Brook drainage area (Aulenbach et al., 1974; Sutherland and Navitsky, 2015). Additionally, the inverse relationship between SM and NO_3 concentration may reflect variation in rates of nitrification by riparian soil microbes, as a higher rate of nitrification is generally associated with lower SM (Casson et al., 2014 and references therein).

Whereas spatial variation in SM appeared important in predicting TN and NO_3, the most important predictor of TSP was the change in SM over the preceding 24 h ($SM_{\Delta 24h}$), with which it was positively correlated. This suggests that episodic high runoff events driven by rainfall or rapid snowmelt (during which SM would be expected to increase) controlled most of the variation in TSP concentration, and that antecedent moisture conditions are likely important (Biron et al., 1999). The positive correlation between TSP and fDOM is consistent with increased catchment export of DOP during high-runoff periods, as noted in the preceding section.

4.4. Hydrostatic pressure and turbidity as a predictor of nutrients

Consistent with previous studies that have found turbidity and stream discharge to be useful proxies for suspended particles and associated nutrients (Jones et al., 2011; Viviano et al., 2014; Robertson et al., 2018), predictors based on HP and turbidity were important in our models of TN, TP, and PP. Lacking rating curves for some tributaries,

396 we used HP as a coarse approximation of water height (stage). We observed that changes in HP and in turbidity,
397 more so than the 1-hr running medians, were important predictors of the total (TN, TP) and particulate (PP) nutrient
398 fractions. It thus appears that relatively short-term hydrological variation (i.e., storm events) were of greater
399 importance than seasonal hydrological variation (e.g., related to snowpack melt in spring or to summer drought
400 periods) in controlling particulate nutrient concentrations. The fact that water temperature (which exhibited
401 pronounced seasonality) was not retained as a predictor in these models supports this idea.

402 It has been noted that turbidity is not solely a function of sediment mass by volume, but is influenced by sediment
403 texture and organic content (Slaets et al., 2014). Turbidity is an optical measure of suspended particles, and the
404 highly skewed distribution of its values in this study likely reflects the landscape of the Lake George drainage basin,
405 where export of particulate matter is generally low, but can at times include contributions of coarse particulate
406 organic matter (CPOM), such as leaf litter, wood debris, etc., that is highly effective at scattering light, but relatively
407 low in P (Kang et al., 2010). We expect that the application of this RFR method in an area where turbidity and
408 particulate P concentrations are higher, less skewed in distribution, and largely contributed by soil particles (vs.
409 CPOM), such as in agricultural watersheds (Cooper et al., 2015), may yield more accurate predictions. Indeed,
410 Jones et al. (2011) noted much stronger correlation between TP and turbidity at a high-turbidity site (max >800
411 NTU) than at a lower-turbidity site (max <60 NTU) located on the same watercourse.

412 *4.5. Other predictors*

413 Water temperature exhibited pronounced seasonality, peaking in July when monthly mean air temperature is highest
414 in the area (climate normal data not shown). Water temperature was retained as a predictor of SRP, with which it
415 was positively correlated. While variation in stream water temperature can reflect a suite of underlying causes
416 (Caissie, 2006), it was variation on a seasonal timescale that appeared to contribute to its importance as a predictor
417 of SRP in this study; the average SRP concentration was 4.4 µg/L during the mid-to-late summer (Jul to Sep) but
418 only 2.2 µg/L during the remainder of the year. The increase in SRP over the summer months is consistent with the
419 seasonal pattern in total reactive P noted by Bowes et al. (2015) who attributed this pattern to hydrological influence
420 (i.e., a lack of dilution during low flow periods).

Likewise, specific conductance generally increased over the summer months as stream temperatures increased and SM decreased, and it was negatively correlated with HP, all of which is consistent with a lack of solute dilution during periods of low flow (e.g., Clark et al., 2016; Moatar et al., 2016). SpCond was strongly positively correlated with TN, NO_3, and SRP, yet was retained as a predictor in only the NO_3 model, in which it was of moderate importance. This likely reflects the high degree of correlation between SpCond and other parameters that were important predictors, namely, fDOM, SM, and HP.

Diel variation in stream nutrient concentrations due to biological activity is a well-recognized phenomenon (e.g., Halliday et al., 2012; Cohen et al., 2013). However, time of day was not retained as a predictor (i.e., was not important) in any of our models. This is likely because diel variation can be relatively subtle, i.e., minor relative to other sources of variation (Bieroza and Heathewaite, 2016). Furthermore, steep watersheds, lack of headwater lakes, and limited wetland areas within most of the catchments may have limited the importance of biological processes.

5. Conclusions

5.1. Advantages & disadvantages of the RFR approach

We recommend the use of RFR for high-frequency estimates of stream nutrient concentrations using high-frequency sensor data. The technique is well suited to this application because it can be applied to multivariate datasets that include highly skewed data, multicollinearity between predictors, nonlinear relationships between predictors and response variables, and high-order interactions. Importantly, variables can be useful predictors in RFR models even if they do not have a monotonic relationship to the response variable (see italicized predictors in Table 2), which is a potential advantage over linear models. Furthermore, RFR is easily implemented in popular programming languages such as R or Python, is relatively simple from a conceptual standpoint (cf. ANNs), and there is no need to explicitly provide a theoretical model structure *a priori*. While we advocate for broader adoption of the technique within the aquatic sciences, we do recognize certain limitations of RFR, including the relative opacity of the resultant model structure (i.e., a set of hundreds of decision trees) and the inability to make predictions based on predictor data outside the ranges used for model calibration (i.e., to extrapolate).

5.2. Summary & Future Directions

While its use does not obviate the need for laboratory determination of nutrient concentrations, the surrogate-regression modeling approach can be used in a supplementary capacity to greatly augment the temporal resolution of nutrient datasets. It therefore has the potential to increase the accuracy of nutrient loading estimates and ~~can~~ further our understanding of the highly dynamic chemical conditions experienced by stream biota. To our knowledge, this is the first published study to use RFR to estimate stream nutrient concentrations based on sensor data in a relatively unimpacted (forested) catchment (cf. Castrillo and García, 2020), or to include instantaneous measurements of soil moisture and fDOM as predictors of lotic nutrient concentrations. The approach was highly successful for nutrient parameters that spanned a moderate range (TN, TP, NO_3, PP); it is expected that better results would be obtained for TSP and SRP in catchments with higher and more variable concentrations of these nutrient fractions. The potential for this approach to improve loading estimates could be further investigated by comparisons of RFR model estimates to continuous concentration data generated by field-deployable auto-analysers (e.g., the WIZ PO_4 sensor by Systea) and ion-selective electrodes (for NO_3).

5. Acknowledgements

We gratefully acknowledge assistance by personnel associated with the Jefferson Project at Lake George, a collaboration between Rensselaer Polytechnic Institute, IBM Research, and the FUND for Lake George. In addition to funding from the three partners, the sensor network has been partially supported by an NSF MRI grant (#1655168) and a New York State Higher Education Capital grant (#7290). We thank three anonymous reviewers of this this manuscript for providing helpful comments.

6. References

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., and Yaseen, Z.M. 2016. Past, present and prospect of an Artificial Intelligence (AI) based model for sediment transport prediction. *Journal of Hydrology,* 541: Part B, 902-913.

Ahmed, A.N., Othman, F.B., Afan, H.A., Ibrahim, R.K., Fai, C.M., Hossain, M.S., Ehteram, M., and Elshafie, A. 2019. Machine learning methods for better water quality prediction. *Journal of Hydrology*, 578: 124084.

471 Anctil, F., Filion, M., and Tournebize, J. 2009. A neural network experiment on the simulation of daily nitrate-
472 nitrogen and suspended sediment fluxes from a small agricultural catchment. *Ecological Modelling*, 220: 879–887.

473 Arlot, S. and Celisse, A. 2010. A survey of cross-validation procedures for model selection. *Statistical Surveys*, 4:
474 40–79.

475 Asadollahfardi, G., Taklify, A., and Ghanbari, A. 2012. Application of artificial neural network to predict TDS in
476 Talkheh Rud River. *Journal of Irrigation and Drainage Engineering*, 138: 363–370.

477 Aulenbach, D.B., Clesceri, N.L., and Tofflemire, T.J. 1974. Thirty-five years of studies of continuous discharge of
478 secondary treated effluent onto sand beds. Rensselaer Polytechnic Institute, Troy, New York; Rensselaer Fresh
479 Water Institute at Lake George. 33 pp + appendices. FWI 74-19.

480 Ay, M. and Kisi, O. 2017. Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing
481 techniques. *KSCE Journal of Civil Engineering*, 21: 1631-1639.

482 Ay, M. and Kisi, O. 2014. Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering
483 techniques. *Journal of Hydrology,* 511: 279–289.

484 Bieroza, M.Z. and Heathwaite, A.L. 2015. Seasonal variation in phosphorus concentration-discharge hysteresis
485 inferred from high-frequency *in situ* monitoring. *Journal of Hydrology*, 524: 333–347.

486 Biron, P.M., Roy, A.G., Courschesne, F., Hendershot, W.H., Côté, B., and Fyles, J. 1999. *Hydrological Processes*,
487 13: 1541–1555.

488 Blaen, P.J., Khamis, K., Lloyd, C.E.M., Bradley, C., Hannah, D., and Krause, S. 2016. Real-time monitoring of
489 nutrients and dissolved organic matter in rivers: Capturing event dynamics, technological opportunities and future
490 directions. *Science of the Total Environment*, 569–570: 647–660.

491 Bowes, M.J., Jarvie, H.P., Halliday, S.J., Skeffington, R.A., Wade, A.J., Loewenthal, M., Gozzard, E., Newman,
492 J.R., and Palmer-Felgate, E.J. 2015. Characterising phosphorus and nitrate inputs to a rural river using high-
493 frequency concentration-flow relationships. *Science of the Total Environment*, 511: 608–620.

494 Boylen, C., L. Eichler, M. Swinton, S. Nierzwicki-Bauer, I. Hannoun, and J. Short. 2014. The state of the lake:
495 Thirty years of water quality monitoring on Lake George. Darrin Fresh Water Institute.
496 https://fundforlakegeorge.org/sites/default/files/site/default/files/lakegeorge/thestateofthelake-web-8-14-
497 2014_final_web_version.pdf

498 Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.G. 1984. Classification and regression trees. Wadsworth
499 International Group, Belmont, CA, USA.

500 Breiman, L. 1996. Bagging predictors. *Machine Learning*, 26: 123–140.

501 Breiman, L. 2001. Random Forests. *Machine Learning*, 45: 5–32.

502 Buffam, I., Galloway, J.N., Blum, L.K., and McGlathery, K.J. 2001. A stormflow/baseflow comparison of
503 dissolved organic matter concentrations and bioavailability in an Appalachian stream. *Biogeochemistry*, 53: 269–
504 306.

505 Burns, D.A., Murdoch, P.S., Lawrence, G.B. 1998. Effect of groundwater springs on NO_3^- concentrations during
506 summer in Catskill Mountain streams. *Water Resources Research*, 34: 1987–1996.

507 Caissie, D. 2006. The thermal regime of rivers: a review. *Freshwater Biology*, 51: 1389–1406.

508 Camargo, J.A., Alonso, A., and Salamanca, A. 2005. Nitrate toxicity to aquatic animals: a review with new data for
509 freshwater invertebrates. *Chemosphere*, 58: 1255–1267.

510 Cassidy, R. and Jordan, P. 2011. Limitations of instantaneous water quality sampling in surface-water catchments:
511 Comparison with near-continuous phosphorus time-series data. *Journal of Hydrology*, 405: 182–193.

512 Casson, N.J., Eimers, M.C., and Watmough, S.A. 2014. Controls on soil nitrification and stream nitrate export at
513 two forested catchments. *Biogeochemistry*, 121: 355–368.

514 Castrillo, M. and García, Á.L. 2020. Estimation of high frequency nutrient concentrations from water quality
515 surrogates using machine learning methods. *Water Research*, 172: 115490.
516

517 Chang, F.J., Chung, C.H., Chen, P.A., Liu, C.W., Coynel, A., and Vachaud, G. 2014. Assessment of arsenic
518 concentration in stream water using neuro fuzzy networks with factor analysis. *Science of the Total Environment,*
519 494–495: 202–210.

520 Christensen, V.G., Rasmussen, P.P., and Ziegler, A.C. 2002. Real-time water quality monitoring and regression
521 analysis to estimate nutrient and bacteria concentrations in Kansas streams. *Water Science and Technology*, 45: 205–
522 211.

523 Clark, E.V, Greer, B.M., Zipper, C.E., and Hester, E.T. 2016. Specific conductance–stage relationships in
524 Appalachian valley fill streams. *Environmental Earth Sciences*, 75: 1222.

525 Cohen, M. J., Kurz, M.J., Heffernan, J.B., Martin, J.B., Douglass, R.L., Foster, C.R., and Thomas, R.G. 2013. Diel
526 phosphorus variation and the stoichiometry of ecosystem metabolism in a large spring-fed river. *Ecological*
527 *Monographs*, 83: 155–176.

528 Cooper, R.J., Rawlins, B.G., Krueger, T., Lézé, B., Hiscock, K.M. and Pedentchouk, N. 2015. Contrasting controls
529 on the phosphorus concentration of suspended particulate matter under baseflow and storm event conditions in
530 agricultural headwater streams. Science of the Total Environment, 533: 49-59.

531 Correll, D.L., Jordan, T.E., and Weller, D.E. 1999. Transport of nitrogen and phosphorus from Rhode River
532 watersheds during storm events. *Water Resources Research*, 35: 2513–2521.

533 De'ath, G. and Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for
534 ecological data analysis. *Ecology*, 81:3178–3192.

535 De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology*, 88: 243–251.
536 Demirci, M. and Baltaci, A. 2013. Prediction of suspended sediment in river using fuzzy logic and multilinear
537 regression approaches. *Neural Computing Applications*, 23: 145–151.

538 Dogan, E., Sengorur, B., and Koklu, R. 2009. Modeling biological oxygen demand of the Melen River in Turkey
539 using an artificial neural network technique. *Journal of Environmental Management*, 90: 1229–1235.

Duncan, J.M., Band, L.E., and Groffman, P.M. 2017. Variable nitrate concentration–discharge relationships in a forested watershed. *Hydrological Processes*, 31: 1817–1824.

Eddy, F.B. and Williams, E.M. 1987. Nitrite and freshwater fish. *Chemistry and Ecology*, 3: 1-38.

Efron, B. and Gong, G. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37: 36-48.

Eimers, M.C., Watmough, S.A., Paterson, A.M., Dillon, P.J., and Yao, H. 2009. Long-term declines in phosphorus export from forested catchments in south-central Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 66: 1682–1692.

Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin, J.B., and Smith, J.E. 2007. Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*, 10: 1135–1142.

Frei, R.J., Abbott, B.W., Dupas, R., Gu, S., Gruau, G., Thomas, Z., Kolbe, T., Aquilina, L., Labasque, T., Laverman, A., Fovet, O., Moatar, F., and Pinay, G. 2020. Predicting Nutrient Incontinence in the Anthropocene at Watershed Scales. *Frontiers in Environmental Science*, 7: 200.

Halliday, S.J., Wade, A.J., Skeffington, R.A., Neal, C., Reynolds, B., Rowland, P., Neal, M., and Norris, D. 2012. An analysis of long-term trends, seasonality and short-term dynamics in water quality from Plynlimon, Wales. *Science of the Total Environment*, 434: 186–200.

Hollander, M. and Wolfe, D.A. 1973. Nonparametric Statistical Methods. John Wiley & Sons, New York, NY, USA.

Hollister, J.W., Milstead, W.B., and Kreakie, B.J. 2016. Modeling lake trophic state: a random forest approach. *Ecosphere*, 7: 1–14.

Holmberg, M., Forsius, M., Starr, M., and Huttunen, M. 2006. An application of artificial neural networks to carbon, nitrogen and phosphorus concentrations in three boreal streams and impacts of climate change. *Ecological Modelling*, 195: 51–60.

564 Horsburgh, J.S., Jones, A.S., Stevens, D.K., Tarboton, D.G., and Mesner, N.O. 2010. A sensor network for high
565 frequency estimation of water quality constituent fluxes using surrogates. *Environmental Modelling & Software*, 25:
566 1031–1044.

567 Hudson, N., Baker, A., and Reynolds, D. 2007. Fluorescence analysis of dissolved organic matter in natural, waste
568 and polluted waters — A review. *River Research and Applications*, 23: 631–649.

569 Johnes, P.J., 2007. Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation
570 methodology, sampling frequency, baseflow index and catchment population density. *Journal of Hydrology*, 332:
571 241–258.

572 Jones, A.S., Stevens, D.K., Horsburgh, J.S., and Mesner, N.O. 2011. Surrogate measures for providing high
573 frequency estimates of total suspended solids and total phosphorus concentrations. *Journal of the American Water*
574 *Resources Association*, 47: 239–253.

575 Jones, F.C., Plewes, R., Murison, L., MacDougall, M.J., Sinclair, S., Davies, C., Bailey, J.L., Richardson, M. and
576 Gunn, J. 2017. Random forests as cumulative effects models: A case study of lakes and rivers in Muskoka,
577 Canada. *Journal of Environmental Management*, 201: 407–424.

578 Kalff, J. 2002. Limnology: Inland Water Ecosystems. Prentice-Hall Inc., Upper Saddle River, NJ, USA.

579 Kang, H., Xin, Z., Berg, N., Burgess, P.J., Liu, Q., Liu, Z., Li, Z. and Liu, C. 2010. Global pattern of leaf litter
580 nitrogen and phosphorus in woody plants. *Annals of Forest Science*, 67: 811p1–811p8.
581
582 Kim J. and Furumai, H. 2013. Improved calibration of a rainfall-pollutant-runoff model using turbidity and electrical
583 conductivity as surrogate parameters for total nitrogen. *Water Environment Journal*, 27: 79–85
584

585 Langner, C.L. and Hendrix, P.F. 1982. Evaluation of a persulfate digestion method for particulate nitrogen and
586 phosphorus. *Water Research*, 16: 1451–1454.
587

588 Lannergård, E.E., Ledesma, J.L.J., Fölster, J., and Futter, M.N. 2019. An evaluation of high frequency turbidity as
589 a proxy for riverine total phosphorus concentrations. *Science of The Total Environment*, 651:103–113.
590
591 Leigh, C., Kandanaarachchi, S., McGree, J.M., Hyndman, R.J., Alsibai, O., Mengersen, K., and Peterson, E.E.
592 2019. Predicting sediment and nutrient concentrations from high-frequency water-quality data. PLoS ONE 14,
593 e0215503.
594
595 Lek, S., Guiresse, M., and Giraudel, J.-L. 1999. Predicting stream nitrogen concentration from watershed features
596 using neural networks. *Water Research*, 33: 3469–3478.
597
598 Lessels, J.S. and Bishop, T.F.A. 2013. Estimating water quality using linear mixed models with stream discharge
599 and turbidity. *Journal of Hydrology*, 498: 13–22.

600 Lewis, W.M. and Wurtsbaugh, W.A. 2008. Control of lacustrine phytoplankton by nutrients: Erosion of the
601 phosphorus paradigm. *International Review of Hydrobiology*, 93: 446–465.

602 Liaw, A. and Wiener, M. 2002. Classification and Regression by randomForest. R News, 2: 18–22.
603
604 Lucius, M.A., Johnston, K.E., Eichler, L.W., Farrell, J.L., Moriarty, V.W., and Relyea, R.A. 2020. Using machine
605 learning to correct for nonphotochemical quenching in high-frequency, in vivo fluorometer data. *Limnology and*
606 *Oceanography: Methods*, doi: 10.1002/lom3.10378.
607
608 Maier, H.R. and Dandy, G.C. 1996. The use of artificial neural network for the prediction of water quality
609 parameters. *Water Resources,* 32: 1013-1022.
610
611 Maier, H.R., Jain, A., Dandy, G.C., and Sudheer, K.P. 2010. Methods used for development of neural Networks for
612 the prediction of water resource variables in river systems: Current status and future directions. *Environmental*
613 *Modelling & Software,* 25: 891–909.

614 Mayora, G, Schneider, B., and Rossi, A. 2018. Turbidity and dissolved organic matter as significant predictors of
615 spatio-temporal dynamics of phosphorus in a large river-floodplain system. *River Research and Applications*,
616 34:629–639.
617
618 Meybeck, M. 1982. Carbon, nitrogen and phosphorus transported by world rivers. *American Journal of Science*,
619 282: 401–450.
620
621 Miltner, R.J. and Rankin, E.T. 1998. Primary nutrients and the biotic integrity of rivers and streams. *Freshwater*
622 *Biology*, 40: 145–158.
623
624 Moatar, F., Fessant, F., and Poirel, A. 1999. pH modelling by neural Networks. Application of control and validation
625 data series in the Middle Loire River. *Ecological Modelling,* 120: 141–156.
626
627 Moatar, F., Abbott, B.W., Minaudo, C., Curie, F., and Pinay, G. 2017. Elemental properties, hydrology, and biology
628 interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions, *Water Resources*
629 *Research*, 53: 1270–1287.
630
631 Mosavi, A., Ozturk, P., and Chau, K. 2018. Flood prediction using machine learning models: Literature review.
632 *Water*, 10: 1536.
633
634 Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models part I - A discussion of
635 principles. *Journal of Hydrology*, 10: 282–290.
636
637 Nour, M.H., Smith, D.W., El-Din, M.G., and Prepas, E.E. 2006. The application of artificial neural networks to
638 flow and phosphorus dynamics in small streams on the Boreal Plain, with emphasis on the role of wetlands.
639 *Ecological Modelling*, 191: 19–32.
640

641 Orouji, H., Haddad, O.B., Fallah-Mehdipour, E., and Mariño, M. 2013. Modeling of water quality parameters using
642 data-driven models. *Journal of Environmental Engineering*, 139: 947–957.
643
644 Paerl, H.W., Scott, J.T., McCarthy, M.J., Newell, S.E., Gardner, W.S., Havens, K.E., et al., 2016. It takes two to
645 tango: when and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems.
646 *Environmental Science and Technology*, 50: 10805–10813.
647
648 Pellerin, B. A., Saraceno, J.F., Shanley, J.B., Sebestyen, S.D., Aiken, G.R., Wollheim, W.M., and Bergamaschi,
649 B.A. 2012. Taking the pulse of snowmelt: in situ sensors reveal seasonal, event and diurnal patterns of nitrate and
650 dissolved organic matter variability in an upland forest stream. *Biogeochemistry*, 108: 183–198.
651
652 R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical
653 Computing, Vienna, Austria. URL https://www.R-project.org/.
654
655 Rahmati, O. Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J.,
656 Cipullo, S., Ahmad, B., and Bui, D. 2019. Predicting uncertainty of machine learning models for modelling nitrate
657 pollution of groundwater using quantile regression and UNEEC methods. *Science of the Total Environment*, 688:
658 855–866.
659 Raymond, P.A. and Saiers, J.E. 2010. Event controlled DOC export from forested watersheds. *Biogeochemistry*,
660 100: 197–209.
661 Richardson, J. 1997. Acute ammonia toxicity for eight New Zealand indigenous freshwater species. New Zealand
662 *Journal of Marine and Freshwater Research*, 31: 185–190.
663 Robertson, D.M., Hubbard, L.E., Lorenz, D.L., and Sullivan, D.J. 2018. A surrogate regression approach for
664 computing continuous loads for the tributary nutrient and sediment monitoring program on the Great Lakes. *Journal*
665 *of Great Lakes Research*, 44: 26–42.

Rodak, C.M., Moore, T.L., David, R., Jayakaran, A.D., and Vogel, J.R. 2019. Urban stormwater characterization, control, and treatment. *Water Environment Research*, 91: 1034–1060.

Rode, M., Wade, A.J., Cohen, M.J., Hensley, R.T., Bowes, M.J., Kirchner, J.W., Arhonditsis, G.B., Jordan, P., Kronvang, B., Halliday, S.J., Skeffington, R.A., Rozemeijer, J.C., Aubert, A.H., Rinke, K., and Jomaa, S. 2016. Sensors in the stream: The high-frequency wave of the present. *Environmental Science & Technology*, 50: 10297–10307.

Ryberg, K.R., 2006. Continuous Water-quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in the Red River of the North, Fargo, North Dakota, USGS Scientific Investigations Report 2006-5241, 35 pp. Available at: http://pubs.usgs.gov/sir/2006/5241/pdf/sir2006-5241.pdf.

Schärer, M., Page, T., and Beven, K. 2006. A fuzzy decision tree to predict phosphorus export at the catchment scale. *Journal of Hydrology*, 331: 484– 494.

Schiff, S.L., Aravena, R., Trumbore, S.E., Hinton, M.J., Elgood, R., and Dillon, P.J. 1997. Export of DOC from forested catchments on the Precambrian Shield of Central Ontario: Clues from ^{13}C and ^{14}C. *Biogeochemistry*, 36: 43–65.

Schindler, D.W. 1977. Evolution of phosphorus limitation in lakes. *Science*, 195: 260–262.

Schindler, D.W., Hecky, R.E., Findlay, D.L., Stainton, M.P., Parker, B.R., Paterson, M.J., Beaty, K.G., Lyng, M., and Kasian, S.E.M. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proceedings of the National Academy of Sciences*, 105: 11254–11258.

Shrestha, R.R., Bàrdossy, A., and Rode, M. 2007. A hybrid deterministic–fuzzy rule based model for catchment scale nitrate dynamics. *Journal of Hydrology*, 342: 143–156.

688 Slaets, J.I.F., Schmitter, P., Hilger, T., Lamers, M., Piepho, H.-P., Vien, T.D., and Cadisch, G. 2014. A turbidity-
689 based method to continuously monitor sediment, carbon and nitrogen flows in mountainous watersheds. Journal of
690 *Hydrology*, 513: 45–57.

691 American Public Health Association, 2017. Standard Methods for the Examination of Water and Wastewater, 23rd
692 Edition. Rice, E.W., Baird, R.B., Eaton, A.D. (Eds.). A joint publication of the American Public Health
693 Association, American Water Works Association, and Water Environment Federation. Washington, DC, USA.

694 Snyder, L., Potter, J.D., and McDowell, W.H. 2018. An evaluation of nitrate, fDOM, and turbidity sensors in New
695 Hampshire streams. *Water Resources Research*, 54: 2466–2479.

696 Stearns and Wheler Companies. 2001. Total phosphorus budget analysis Lake George Watershed New
697 York. Prepared for State of New York Lake George Park Commission. Prepared by Stearns & Wheler Companies
698 Environmental Engineers and Scientists. October 2001.

699 Stone, M. 1974. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical*
700 *Society, Series, B*, 36: 111–147.

701 Suen, J.-P. and Eheart, J.W. 2003. Evaluation of Neural Networks for Modeling Nitrate Concentrations in Rivers.
702 *Journal of Water Resources Planning and Management*, 129: 505–510.

703 Sutherland, J.W. and Navitsky, C. 2015. Village of Lake George Wastewater Treatment Plan - A Monitoring
704 Program to Document Current Treatment Efficiencies. Final Report to the Village of Lake George, December 2015.

705 U.S. Environmental Protection Agency. 1979. Methods for chemical analysis of water and wastes. Cincinnati, OH,
706 USA: EPA-600/4-79-020.

707 U.S. Geological Survey. 2016. The StreamStats program, online at http://streamstats.usgs.gov, accessed on 25
708 March 2020.

709 Villa, A., Fölster, J., and Kyllmar, K. 2019. Determining suspended solids and total phosphorus from turbidity:
710 comparison of high-frequency sampling with conventional monitoring methods. *Environmental Monitoring and*
711 *Assessment*, 191: 605.

712 Viviano, G., Salerno, F., Manfredi, E.C., Polesello, S., Valsecchi, S., and Tartari, G. 2014. Surrogate measures for
713 providing high frequency estimates of total phosphorus concentrations in urban watersheds. *Water Research*, 64:
714 265–277.

715 Wagner, R.J., Boulger, R.W., Jr., Oblinger, C.J., and Smith, B.A. 2006. Guidelines and standard procedures for
716 continuous water-quality monitors—Station operation, record computation, and data reporting: U.S. Geological
717 Survey Techniques and Methods 1–D3, 51 p. + 8 attachments; accessed April 10, 2020, at
718 http://pubs.water.usgs.gov/tm1d3

719 Watras, C.J., Hanson, P.C., Stacy, T.L., Morrison, K.M., Mather, J., Hu, Y.-H., and Milewski, P. 2011. A
720 temperature compensation method for CDOM fluorescence sensors in freshwater. *Limnology and Oceanography:*
721 *Methods*, 9: 296–301.
722
723 Wen, X., Fang, J., Diao, M., and Zhang, C. 2013. Artificial neural network modeling of dissolved oxygen in the
724 Heihe River, Northwestern China. *Environmental Monitoring and Assessment*, 185: 4361–4371.
725
726 Wilson, H. and Recknagel, F. 2001. Towards a generic artificial neural network model for dynamic predictions of
727 algal abundance in freshwater lakes. *Ecological Modelling,* 147: 69–84.
728

729 Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A., and Sayl, K.N. 2015. Artificial intelligence based models for
730 stream-flow forecasting: 2000-2015. *Journal of Hydrology,* 530: 829-844.

731 Zar, J.H. 1999. Biostatistical Analysis, 4th Edition. Prentice Hall, Upper Saddle River, NJ, USA.
732

Random Forests Regression

Nitrogen & Phosphorus concentrations