Hydrological Processes

WILEY

RESEARCH ARTICLE

Stuck at Home: Machine-Learning Models Predicting Solute Concentrations of One Stream Failed to Predict Solute Concentrations in Other Streams

Hollis C. Harrington[1] | Mark B. Green[2,3] | John L. Campbell[3] | William H. McDowell[4,5] | Adam S. Wymore[4] | Ruth D. Yanai[6]

[1]Department of Chemistry, SUNY College of Environmental Science and Forestry, Syracuse, New York, USA | [2]Department of Earth, Environmental, and Planetary Sciences, Case Western Reserve University, Cleveland, Ohio, USA | [3]USDA Forest Service, Northern Research Station, Durham, New Hampshire, USA | [4]Department of Natural Resources and the Environment, University of new Hampshire, Durham, New Hampshire, USA | [5]Institute of Environment, Florida International University, Miami, Florida, USA | [6]Department of Sustainable Resources Management, SUNY College of Environmental Science and Forestry, Syracuse, New York, USA

Correspondence: Ruth D. Yanai (rdyanai@syr.edu)

Received: 27 November 2024 | **Revised:** 19 March 2025 | **Accepted:** 15 April 2025

Funding: This work was supported by the US Forest Service, US Department of Energy, Office of Experimental Program to Stimulate Competitive Research, Division of Environmental Biology, USDA McIntire-Stennis Cooperative Forestry Program, National Institute of Food and Agriculture, US Department of Agriculture, National Science Foundation.

Keywords: biogeochemistry | data interpolation | high-frequency environmental data | machine learning | model transferability | random forest models | stream solutes | water quality

ABSTRACT

Machine-learning models have been surprisingly successful at predicting stream solute concentrations, even for solutes without dedicated sensors. It would be extremely valuable if these models could predict solute concentrations in streams beyond the one in which they were trained. We assessed the generalisability of random forest models by training them in one or more streams and testing them in another. Models were made using grab sample and sensor data from 10 New Hampshire streams and rivers. As observed in previous studies, models trained in one stream were capable of accurately predicting solute concentrations in that stream. However, models trained on one stream produced inaccurate predictions of solute concentrations in other streams, with the exception of solutes measured by dedicated sensors (i.e., nitrate and dissolved organic carbon). Using data from multiple watersheds improved model results, but model performance was still worse than using the mean of the training dataset (Nash–Sutcliffe Efficiency <0). Our results demonstrate that machine-learning models thus far reliably predict solute concentrations only where trained, as differences in solute concentration patterns and sensor-solute relationships limit their broader applicability.

1 | Introduction

High-frequency stream solute data plays an important role in examining hydrological and biogeochemical dynamics in freshwater ecosystems and addressing water quality challenges. Streams are dynamic ecosystems that carry dissolved and particulate matter from headwaters to downstream receiving waters (Lintern et al. 2018; Raymond and Saiers 2010). Variations in solute concentrations occur on diel to interannual scales, necessitating both frequent and long-term sampling to fully understand stream dynamics (Hensley and Cohen 2016; Navrátil et al. 2010; Speir et al. 2024). Historically, samples from streams have been collected manually at discrete intervals (e.g., weekly or monthly), often missing hydrological events and diurnal variability (Fovet et al. 2018; Gu et al. 2012). Erroneous solute-flux estimates have been produced from studies excluding these brief, but important, hydrological events, highlighting the need for high-frequency data (Swistock et al. 1997; Worrall et al. 2013). However, more frequent manual sampling regimes are costly and time-consuming, which limits their widespread implementation. Thus, data interpolation methods are required to supplement grab samples when higher resolution concentration estimates are needed.

In situ stream chemistry sensors provide continuous data, which is an improvement over grab samples, allowing better estimates of fluxes and better linking chemical temporal trends to the hydrograph (Fazekas et al. 2020). Electrical and optical sensors, commonly deployed to measure parameters such as dissolved oxygen, nitrate, and specific conductance, have provided insight into stream solute dynamics (Rode et al. 2016). However, sensors are not yet available for all solutes of interest. For some solutes, a different solute can serve as a proxy, such as estimating mercury using fluorescent dissolved organic matter (FDOM) in forested watersheds (Vermilyea et al. 2017) or chloride concentrations from specific conductance in more urban watersheds (Shattuck et al. 2023). The direct utilisation of proxy data for solute prediction is limited to sensors and solutes with strong correlations. For most solutes, other analytical and statistical techniques are required for solute prediction.

Machine-learning models have been used to predict solute concentrations using sensor data. Such models have accurately predicted solute concentrations in single watersheds (Anmala and Venkateshwarlu 2019; Green et al. 2021; Zanoni et al. 2022) and in multi-watershed studies by pooling data from several watersheds (Harrison et al. 2021; Zhi et al. 2021). Random forest (RF) regression models create an ensemble of regression trees and classifications that develop multiple predictions of the dependent variable of interest. Random forest regression consistently outperformed linear regression and support vector machine regression for stream solute chemistry prediction across diverse datasets (Olson and Hawkins 2012). This strong performance is due in part to the resistance of RF models to overfitting as well as their ability to handle challenges such as limited sampling data and non-linear relationships (Breiman 2001; Green et al. 2021; Olson and Hawkins 2012; Tran et al. 2022). However, such studies have been confined to the watersheds in which the models were trained. To our knowledge, there has been no attempt to apply machine-learning models to watersheds outside of the dataset used for training. The ability to predict chemistry data with sensors in one watershed, using sensors and grab sample data from another watershed, could be a major breakthrough for predicting nutrient and solute fluxes in watersheds without grab sample data. Such an advance would reduce the labour and expense associated with manual stream sampling and justify investment in sensors.

In this study we assessed the capabilities of RF models to predict solute concentrations beyond the watershed in which they were trained. Models were trained and tested in multiple watersheds in temperate forests in New Hampshire, USA, with sizes ranging from small zero-order watersheds to large 6th-order rivers (Koenig et al. 2017). We tested whether a RF model trained with sensor and solute data from one watershed could be applied to sensor data from another watershed to predict solute concentrations. We expected models to perform better in watersheds with similar characteristics and comparable solute ranges. We also expected that including multiple watersheds in the training dataset would improve model predictive performance by exposing the models to a wider range of solute concentrations.

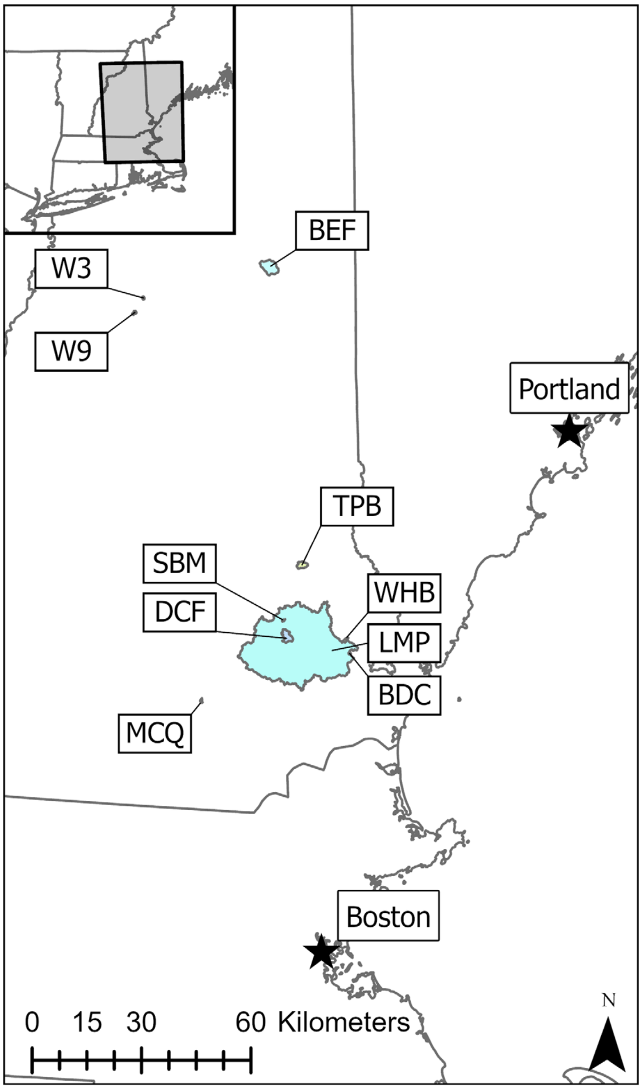


FIGURE 1 | Map of sampling locations in New Hampshire. Some watersheds are so small as to appear as points, and four are subcatchments of another (LMP) (ESRI 2011).

2 | Methods

2.1 | Site Description

Sensor and grab sample data were obtained from 10 watersheds in New Hampshire, USA (Figure 1 and Table 1). Data were collected by the New Hampshire EPSCoR High Intensity Aquatic Network distributed across the state (Koenig et al. 2017; Snyder et al. 2018) in conjunction with the US Forest Service at the Hubbard Brook Experimental Forest (HBEF) in Watersheds 3 and 9 (Campbell et al. 2021). The Lamprey River watershed is the largest in this study, with an area of 477 km^2 primarily forested, with interspersed regions

TABLE 1 | Characteristics of the sampled watersheds. Precise sampling locations are listed (latitude and longitude). Land cover types were determined from the USDA National Agricultural Statistics Service Cropland Data Layer (USDA-NASS 2024).

Watershed	Number of grab samples	Grab sample date range	Latitude	Longitude	Elevation (m above sea level)	Stream order	Drainage area (km^2)	Agricultural area (%)	Developed area (%)	Forested area (%)	Wetland area (%)
Burley-Demeritt Creek (BDC)	34	May ‘13—Nov ‘15	43.0933	−70.9890	25	0	0.3	48.8	7.5	32.5	11.2
Albany Brook (BEF)	32	Nov ‘13—Dec ‘16	44.0617	−71.2946	263	3	13.7	0	3.9	95.9	0.1
Dowst Cate Forest Stream (DCF)	194	Dec ‘12—Dec ‘16	43.1347	−71.1839	100	1	7.0	3.9	4.7	79.1	9.6
Lamprey River (LMP)	209	Dec ‘12—Dec ‘16	43.1040	−70.9626	15	6	476.7	4.7	9.6	72.1	10.3
McQuesten Brook (MCQ)	49	Mar ‘14—Sep ‘15	42.9648	−71.4780	39	1	0.4	0	95.4	2.8	1.8
Saddleback Mountain Stream (SBM)	105	Dec ‘12—May ‘16	43.1704	−71.2173	205	1	0.3	0	0	100	0
Trout Pond Brook (TPB)	40	Dec ‘13—Nov ‘16	43.3178	−71.1675	179	1	4.1	0.3	1.2	92.9	3.8
Watershed 3 at Hubbard Brook (W3)	342	Jun ‘13—Jul ‘17	43.9549	−71.7223	523	1	0.4	0	0	100	0
Watershed 9 at Hubbard Brook (W9)	140	Jul ‘18—Nov ‘21	43.9261	−71.7467	681	1	0.7	0	0	100	0
Wednesday Hill Brook (WHB)	144	Dec ‘12—Dec ‘16	43.1222	−71.0049	35	1	1.0	7.2	26.8	57.2	7.6

of agricultural and developed land (Wymore et al. 2021). Four subwatersheds of the Lamprey River were included in the study: Burley-Demeritt Creek (BDC), Dowst Cate Forest Stream (DCF), Saddleback Mountain (SBM), and Wednesday Hill Brook (WHB). DCF and SBM are predominantly forested watersheds with a higher proportion of wetland in DCF. BDC and WHB represent more disturbed watersheds, with BDC containing large proportions of agricultural land and WHB containing more developed areas with a high density of septic tanks (i.e., on-site waste disposal). McQuesten Brook (MCQ) is a heavily urbanised watershed and is a subwatershed of the Merrimack River. Trout Pond Brook (TPB) is predominantly forested and also drains to the Merrimack River. Three hardwood forest watersheds are located in the White Mountains: Albany Brook, located on an eastern slope in the Bartlett Experimental Forest (BEF); Watershed 3 (W3), on a southern slope in HBEF (W3); and Watershed 9 (W9), on a northern slope in HBEF. Across watersheds, the range in mean annual temperature is 6 to 8°C and precipitation is 1110 to 1400 mm (National Oceanic and Atmospheric Administration 2024). Additional watershed characteristics are listed in Table 1.

2.2 | Streamwater Sensor and Solute Data

Sensor data were collected between 2012 and 2022 for sampling frequencies and periods that varied by watershed (Table 1). At each watershed outlet, an EXO2 multiparameter sonde (Yellow Springs Instruments, Yellow Springs, Ohio) was installed to measure specific conductance, water temperature, dissolved oxygen, pH, and fluorescent dissolved organic matter (FDOM), an optical indicator of dissolved organic carbon (DOC) (Snyder et al. 2018; Wymore et al. 2018). A Satlantic Submersible Ultraviolet Nitrate Analyzer (SUNA; Sea-Bird Scientific, Bellevue, WA) was also installed for measuring nitrate concentrations (Snyder et al. 2018). Sensor measurements were obtained every 15 min. We did not include predictor variables such as soil moisture and discharge that are expensive to monitor because the value of this exercise was in evaluating the conditions in which a small investment in sensors could have a big benefit in understanding stream chemistry.

Measured solute concentrations from the same time period were used to train RF models, functioning as dependent variables in all models. These data are mostly weekly grab samples but include event-based sampling efforts. Laboratory measured solutes included calcium (Ca^{2+}), chloride (Cl^{-}), dissolved organic carbon (DOC), potassium (K^{+}), magnesium (Mg^{2+}), sodium (Na^{+}), nitrate (NO_3^{-}), and sulfate (SO_4^{2-}). Anions from W3 and W9 were analysed using inductively coupled plasma atomic emission spectrometry, while anions at all other sites and cations at all sites were measured via ion chromatography (Hubbard Brook Watershed Ecosystem Record (HBWatER) 2023; Pfaff and Hautman 1999; Wymore et al. 2021). Total dissolved nitrogen (TDN) and DOC were measured by high-temperature catalytic oxidation (Merriam et al. 1996; Potter and Wimsatt 2005). Dissolved organic nitrogen (DON) was estimated as TDN minus NH_4^{+} and NO_3^{-}.

Grab samples with solute concentrations below the detection limit were replaced with half the detection limit (Snyder et al. 2018; Wymore et al. 2021). Grab samples with any missing analyte were omitted from the dataset. Of the 1810 grab samples in the original dataset, 521 were omitted due to incomplete data for all solutes, leaving 1289 samples to be used for model generation.

2.3 | Random Forest Modelling

We used the RF algorithm to predict stream solute concentrations trained on stream sensor data. The RF approach produces an ensemble of regression trees, with each tree based on a bootstrap sample of the observations of stream chemistry data (grab and sensor data). For each split, a randomly chosen subset of the selected independent variables is used as predictors. The algorithm calculates the average prediction from the ensemble of regression trees as the best estimate (Breiman 2001). Random forests can be used for classification or regression; we used it as a regression model.

The randomForest package in R version 4.3.3 was used to generate RF models with 500 regression trees per forest, a minimum node size of five, and a third of the independent variables tried at each split (Liaw and Wiener 2001; R Core Team 2024). The independent variables used in each RF model included water temperature, sensor-based NO_3^{-} concentration, FDOM, specific conductance, pH, dissolved oxygen concentration, and the sine and cosine of the day of the year to account for seasonal patterns in the dataset. The Variable Selection Using RFs (VSURF) R package was used to select the variables used to make the RF models (Genuer et al. 2015).

The predictive capabilities of the models were evaluated using Nash–Sutcliffe model efficiency coefficients (NSE) for each solute across watersheds and modelling scenarios (Nash and Sutcliffe 1970). A model that perfectly predicts solute concentrations would have an NSE value of 1. A model with predictive performance equivalent to using the mean for solute concentration prediction would have an NSE value of 0. Negative NSE values indicate that the model predictions were worse than the mean of the observations. A model with NSE > 0.5 was considered satisfactory (Moriasi et al. 2007).

We evaluated the importance of predictor variables by excluding them when constructing a RF model and reporting the reduction in model accuracy. Variables more important for model accuracy result in a larger reduction. Partial plots were constructed to further evaluate how RF models estimated solute concentrations using predictor variables.

Three modelling scenarios were explored to assess the predictive capabilities of RF models: individual models, travelling models, and all-but-one models. The individual model scenario consisted of 10 models trained and tested with data from a single watershed. These models were generated to establish baseline performance of RF models. Models were validated using repeated k-fold cross validation ($k=5$, repetitions $=6$). NSE and R^2 values were calculated using observed solute concentrations from the $1/k$ of observations reserved for testing and the predicted solute concentration from the RF model for those same samples. Median NSE and R^2 values were reported. The travelling model assessed how well a model trained in one watershed could predict solute concentrations in a different watershed. Two travelling models were trained using the two largest datasets based on the size of the grab sample dataset, LMP and W3. In the all-but-one model scenario, RF models were trained on data from all watersheds except one. The excluded watershed was reserved for testing the model.

3 | Results

3.1 | Sensor and Solute Data Overview

Sensor values (Figure 2) and grab sample solute concentrations (Figure 3) exhibited qualitative trends that varied with the degree of human impact (agriculture and urban development) in the watershed. The most developed watersheds (MCQ, WHB, BDC, and LMP) had the highest specific conductance, nitrate, and pH sensor values (Figure 2), and they also had the highest concentrations of all cations and anions measured in the grab samples (Figure 3). In contrast, there was little intersite variability for dissolved oxygen or temperature. Sensor FDOM measurements (Figure 2) and concentrations of DOC (Figure 3) were highest at BDC and W9; DON showed similar patterns but was also high in DCF and LMP.

3.2 | Individual Models

Random forest models trained and tested in individual watersheds generally had NSE values greater than 0.5 (Figure 4), indicating success at predicting solute concentrations (Moriasi et al. 2007). DON at LMP and SBM had negative NSE values.

We evaluated the importance of each independent variable using the mean decrease in model accuracy. Independent variables were ranked by their importance in predicting Ca^{2+} and SO_4^{2-} in each watershed (Figure 5). While certain variables, such as specific conductance, were often important, few models depended on the exact same set of variables. The BEF model for Ca^{2+} and MCQ model for SO_4^{2-} were the only models that relied on a single predictor variable. The number of variables selected did not explain model performance. For example, the NSE values for SO_4^{2-} in MCQ and W9 differed by only 0.02, meaning that model performance was similar at these two sites. However, the MCQ SO_4^{2-} model used only NO_3^- data, whereas the W9 model used data from six sensors.

3.3 | Travelling Models

The travelling LMP and W3 models poorly predicted solute concentrations in other watersheds, with the vast majority of solutes having a NSE <0 (Figure 6). Exceptions with NSE values >0 were solutes with dedicated sensors, such as NO_3^-, and the solutes with a proxy sensor, namely FDOM for DOC and DON.

R^2 values for some solutes were >0.4 even though they had a NSE <0, suggesting model bias. For example, predicted Mg^{2+} concentrations in BEF by the travelling W3 model had a NSE value of -0.39 but a R^2 of 0.77. In some cases, no R^2 is reported

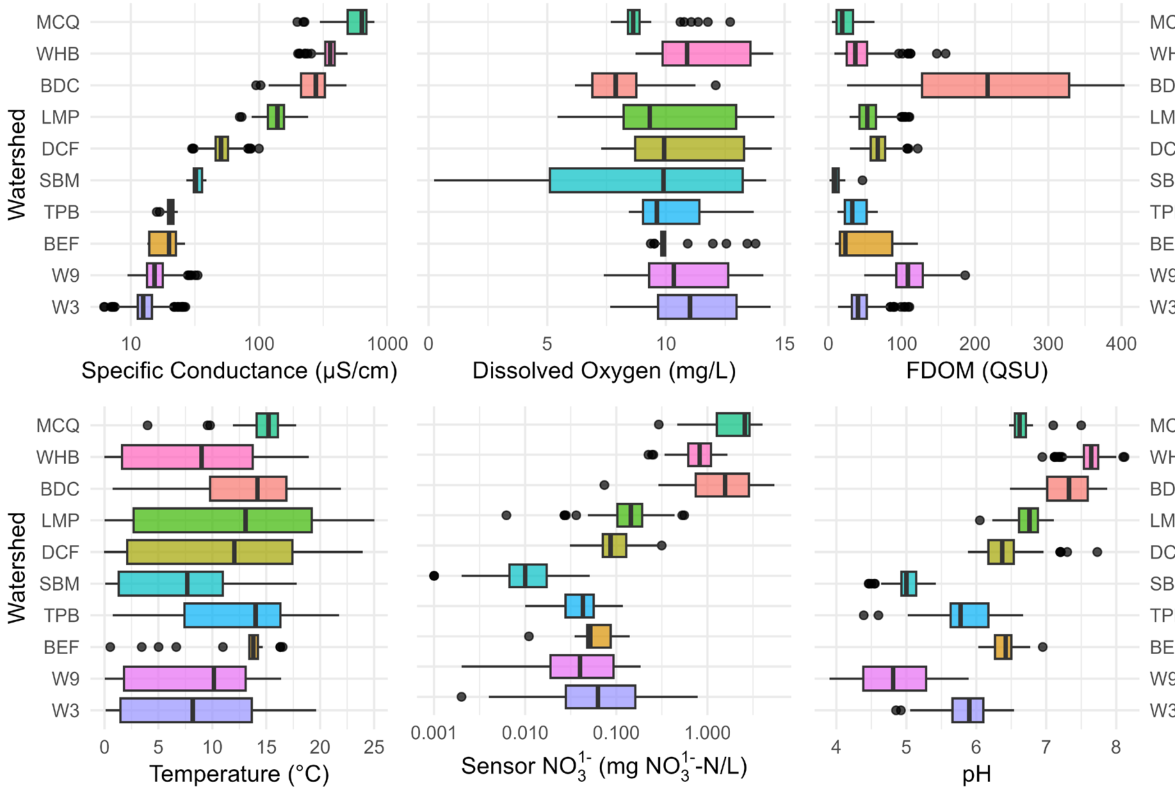


FIGURE 2 | Distribution of sensor values for each watershed ordered by median specific conductance. Specific conductance and sensor NO_3^- are logarithmically scaled. Outliers fall outside the interquartile range by at least 1.5 times the interquartile range.

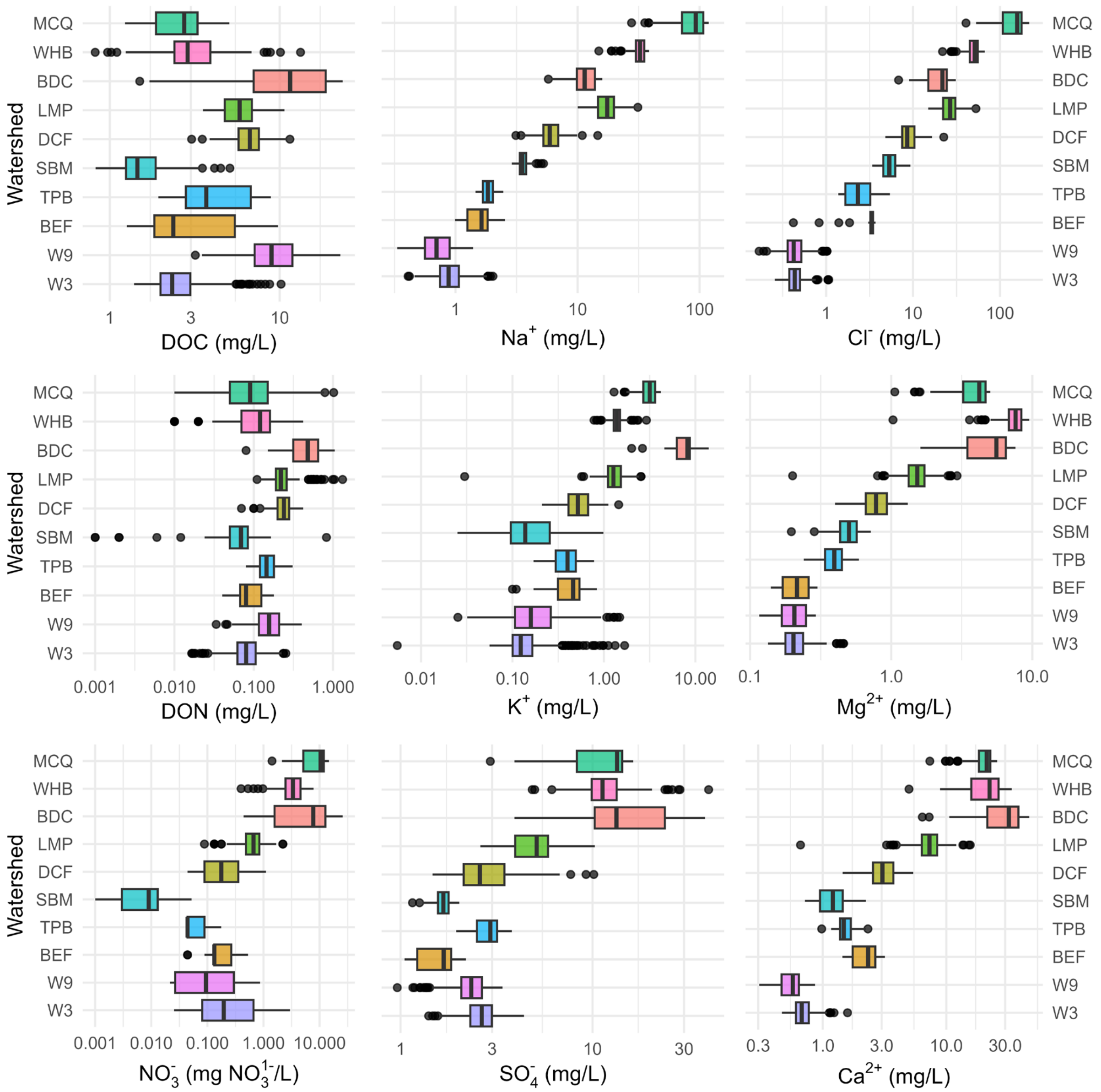


FIGURE 3 | Distribution of grab sample solute concentrations for each watershed ordered by median specific conductance. Na^+, Cl^-, DON, K^+, Mg^{2+}, NO_3^-, SO_4^{2-}, and Ca^{2+} are logarithmically scaled. Outliers fall outside the interquartile range by at least 1.5 times the interquartile range.

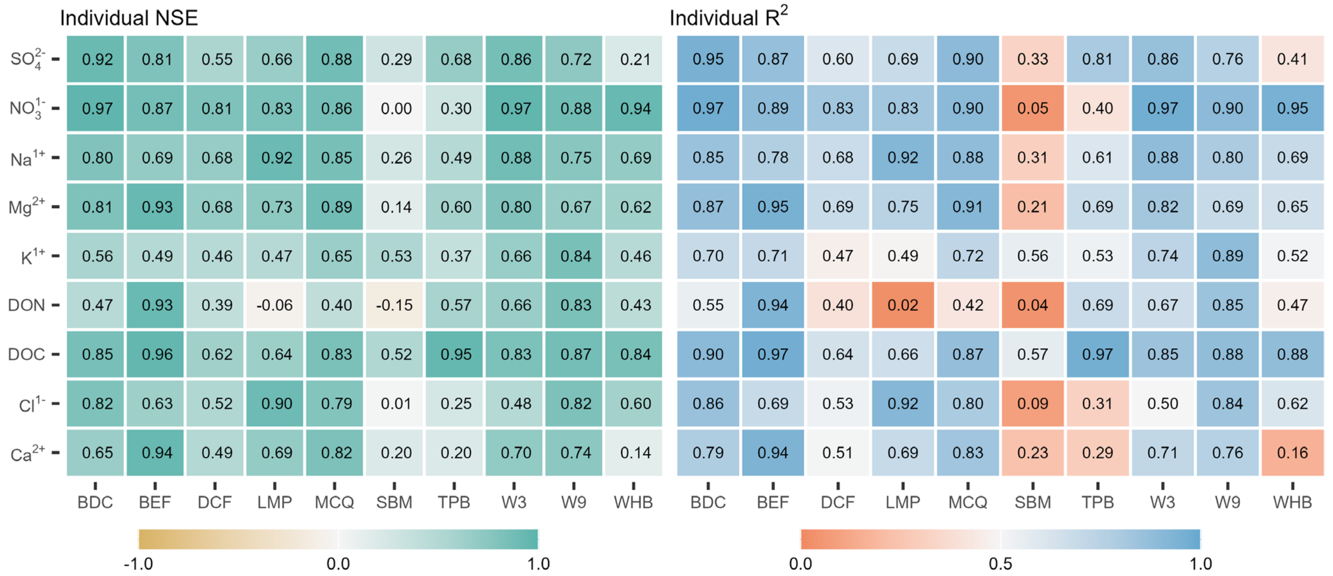
because the RF model predicted only the highest or lowest values observed in the training dataset. This situation occurred when a model was applied at a site with sensor measurements that were outside the range of the model training dataset, such as the travelling LMP model predicting Na^+ and Cl^- concentrations in many of the headwater forested watersheds with little human development.

3.4 | All-But-One Model

Results from all-but-one models were generally better than the travelling models. Model performance was still highest for solutes with dedicated sensors or clear sensor proxies (NO_3^-, DOC, and DON) (Figure 6). Highly negative NSE values were observed, especially for SO_4 and to lesser extents for Ca^{2+}, K^+, and Mg^{2+}. R^2 values were low, except for the same set of solutes (Figure 6). There were some exceptions, such as predicted Na^+ in LMP, for which NSE was highly negative and R^2 was 0.73, indicating a positive relationship between predicted and measured values despite poor model performance according to the NSE.

3.5 | Model Comparison

Ca^{2+} and SO_4^{2-} were further analysed to provide insight into RF model bias. They were selected because of their importance in acid–base chemistry and the lack of dedicated sensors for their measurement. Predicted vs. observed Ca^{2+} concentration plots for LMP illustrate model bias (Figure 7). The individual model performed well with a slope of 0.76 and an intercept of 1.50. The all-but-one model over-predicted Ca^{2+} concentrations, with a slope of 0.87 and an intercept of 2.06. The travelling W3 model at LMP exhibited even more bias, with a slope of 0.02 and an intercept of 0.85, as the model underpredicted Ca^{2+} concentrations.



Individual NSE

	BDC	BEF	DCF	LMP	MCQ	SBM	TPB	W3	W9	WHB
SO_4^{2-}	0.92	0.81	0.55	0.66	0.88	0.29	0.68	0.86	0.72	0.21
NO_3^{1-}	0.97	0.87	0.81	0.83	0.86	0.00	0.30	0.97	0.88	0.94
Na^{1+}	0.80	0.69	0.68	0.92	0.85	0.26	0.49	0.88	0.75	0.69
Mg^{2+}	0.81	0.93	0.68	0.73	0.89	0.14	0.60	0.80	0.67	0.62
K^{1+}	0.56	0.49	0.46	0.47	0.65	0.53	0.37	0.66	0.84	0.46
DON	0.47	0.93	0.39	-0.06	0.40	-0.15	0.57	0.66	0.83	0.43
DOC	0.85	0.96	0.62	0.64	0.83	0.52	0.95	0.83	0.87	0.84
Cl^{1-}	0.82	0.63	0.52	0.90	0.79	0.01	0.25	0.48	0.82	0.60
Ca^{2+}	0.65	0.94	0.49	0.69	0.82	0.20	0.20	0.70	0.74	0.14

Individual R^2

	BDC	BEF	DCF	LMP	MCQ	SBM	TPB	W3	W9	WHB
SO_4^{2-}	0.95	0.87	0.60	0.69	0.90	0.33	0.81	0.86	0.76	0.41
NO_3^{1-}	0.97	0.89	0.83	0.83	0.90	0.05	0.40	0.97	0.90	0.95
Na^{1+}	0.85	0.78	0.68	0.92	0.88	0.31	0.61	0.88	0.80	0.69
Mg^{2+}	0.87	0.95	0.69	0.75	0.91	0.21	0.69	0.82	0.69	0.65
K^{1+}	0.70	0.71	0.47	0.49	0.72	0.56	0.53	0.74	0.89	0.52
DON	0.55	0.94	0.40	0.02	0.42	0.04	0.69	0.67	0.85	0.47
DOC	0.90	0.97	0.64	0.66	0.87	0.57	0.97	0.85	0.88	0.88
Cl^{1-}	0.86	0.69	0.53	0.92	0.80	0.09	0.31	0.50	0.84	0.62
Ca^{2+}	0.79	0.94	0.51	0.69	0.83	0.23	0.29	0.71	0.76	0.16

FIGURE 4 | Nash–Sutcliffe efficiencies and coefficients of determination (R^2) for individually trained and tested watersheds.

Predictive time series plots for Ca^{2+} and SO_4^{2-} compare individual model predictions, all-but-one model predictions, and grab sample measurements in LMP (Figure 8). The time series of predicted Ca^{2+} concentrations from the individual model compared to the all-but-one model showed similar temporal variation, but the all-but-one model predicted greater extremes. Unlike the Ca^{2+} models, all-but-one SO_4^{2-} model predictions failed to follow temporal variations in LMP. The poor performance of the all-but-one model was expected for SO_4^{2-} in LMP as the model had an NSE value of −0.20 and R^2 of 0.06, suggesting the model was not merely biased but suffered from an overall inability to predict concentrations of SO_4^{2-}.

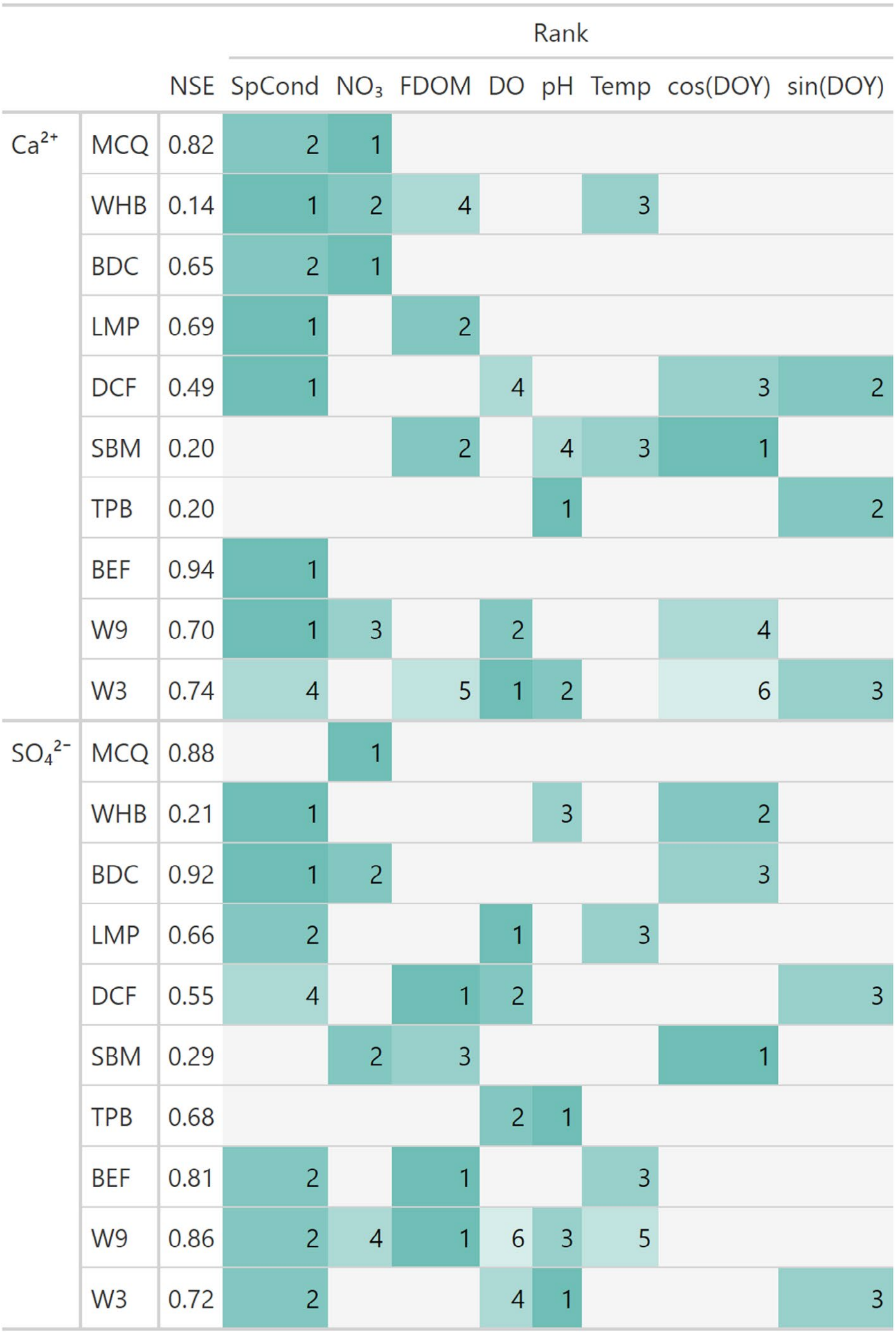
Partial dependence plots were made to illustrate Ca^{2+} model prediction of specific conductance for individual RF models (Figure 9). The same specific conductance value can result in a wide range of predicted Ca^{2+}. For example, a specific conductance value of 100 μS/cm can result in predicted Ca^{2+} concentrations of <1 mg/L in W3 and W9 to >15 mg/L in BDC and WHB. Additionally, RF models will predict a single value for concentrations beyond the range of the training data, as illustrated by the constant predictions for Ca^{2+} concentrations at high and low specific conductance values.

4 | Discussion

Random forest models were not successful at predicting stream solute concentrations outside of the training dataset location. Poor model performance is perhaps not surprising, considering that different independent variables were important for solute prediction at each stream (Figure 5). Even when the same independent variable was used to predict a solute concentration, such as specific conductance for predicting Ca^{2+}, the relationships between the solute and sensed variable varied enough across streams that using one site to predict others was not effective (Figure 9). Despite the success of RF models predicting stream solutes at one location (e.g., this study, Anmala and Venkateshwarlu 2019; Green et al. 2021; Zanoni et al. 2022), these analyses suggest that grab sample campaigns are still necessary to develop models that accurately predict local solute concentrations. When there are many more sites with sufficient grab sample data to train a multi-site model, then including site characteristics as predictor variables should make RF models more transferable. The number of catchments in our study ($n=10$) was too limited for a RF model to be trained on catchment characteristics.

The travelling LMP and W3 models were rarely successful, showing few instances of NSE values >0, which was not surprising in cases where solute concentrations in the test stream were outside the solute concentrations in the training dataset. There were multiple cases where the travelling LMP model had higher solute concentrations than other sites and where W3 had lower solute concentrations than the other sites (Figure 3). However, the inability of the travelling W3 model to accurately predict solute concentrations with similar ranges at nearby sites, namely the two other steep mountainous watersheds (BEF and W9), was surprising (Figure 6). Similarly, the travelling LMP model poorly predicted solute concentrations in its subwatersheds BDC, DCF, SBM, and WHB (Figure 6), demonstrating that RF models are site-specific even within a single river network. The finding that RF models are site-specific is further supported by the differing variable importance ranking for the individual RF models (Figure 5). Cumulatively, our study suggests that it is unlikely that samples from one watershed can be used to predict solute concentrations in other watersheds even if the watershed is nearby or nested within the training watershed.

Exposing machine-learning algorithms to a larger number and diversity of samples often led to improved modelling predictions; this was the case with all-but-one models. Unfortunately, these improvements were not sufficient to yield many useful results (i.e., NSE <0) compared to the individual models (Figure 6). In most cases, the all-but-one model produced adequate predictions for sensors designed to indicate a

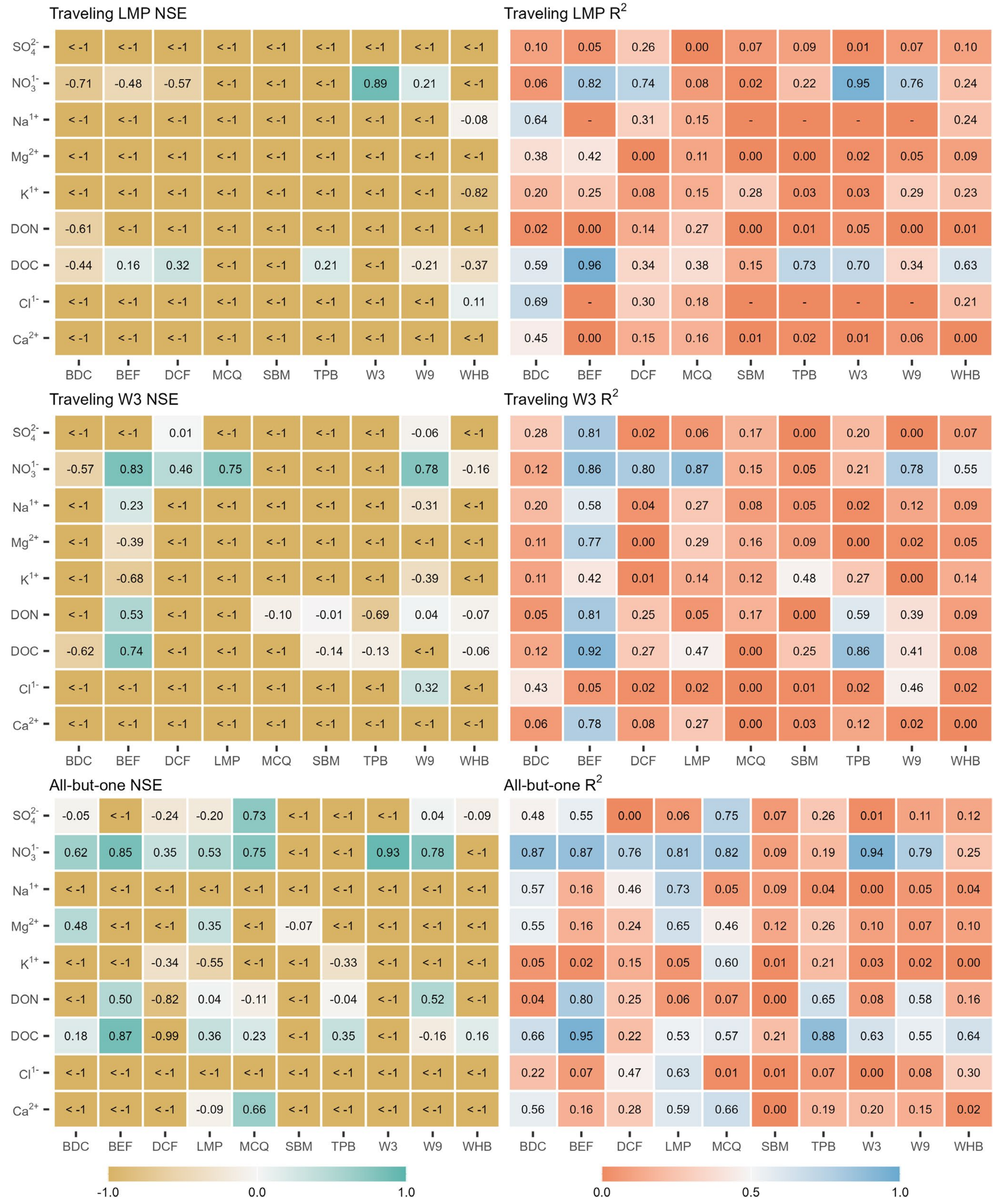


			Rank							
		NSE	SpCond	NO_3	FDOM	DO	pH	Temp	cos(DOY)	sin(DOY)
Ca^{2+}	MCQ	0.82	2	1						
	WHB	0.14	1	2	4			3		
	BDC	0.65	2	1						
	LMP	0.69	1		2					
	DCF	0.49	1			4			3	2
	SBM	0.20			2		4	3	1	
	TPB	0.20					1			2
	BEF	0.94	1							
	W9	0.70	1	3		2			4	
	W3	0.74	4		5	1	2		6	3
SO_4^{2-}	MCQ	0.88		1						
	WHB	0.21	1				3		2	
	BDC	0.92	1	2					3	
	LMP	0.66	2			1		3		
	DCF	0.55	4		1	2				3
	SBM	0.29		2	3				1	
	TPB	0.68				2	1			
	BEF	0.81	2		1			3		
	W9	0.86	2	4	1	6	3	5		
	W3	0.72	2			4	1			3

FIGURE 5 | Variable importance (1 being the most important) for Ca^{2+} and SO_4^{2-} prediction in each watershed using individual models.

single variable (e.g., NO_3^- or FDOM for DOC). However, the models in these cases are not better than the sensor estimate. The models also performed well for Ca^{2+} at MCQ, DON at BEF and W9, Mg^{2+} at BDC and LMP, and SO_4^{2-} at MCQ, likely because concentrations were high, thus producing a larger signal for the model to predict and strong individual sensor-solute relationships.

One of the shortcomings of RF models is the inability to accurately predict values that are poorly represented in the training dataset. Observations beyond the range of the training dataset result in inaccurate predictions, as was the case with DOC and fDOM in SBM. SBM had the lowest fDOM values in this study, and DOC predictions for SMB by the all-but-one and travelling RF models were inaccurate. Similarly, RF models will not produce accurate concentration predictions if there are gaps in the training dataset. The training data for the all-but-one model for NO_3^- at WHB have NO_3^- concentrations above and below those observed in WHB, but very few in the range observed at WHB. Other modelling techniques, such as linear regression, would not have an issue with gaps in the middle of the data range as long as there were a consistent relationship between grab sample and sensor concentrations, as was pointed out by Snyder et al. (2018) for DOC and NO_3^- in this dataset. Additionally, RF models will fail at low concentrations if the relationship between sensor and grab sample concentrations is poor at low concentrations, as is the case with NO_3^- in SBM and TPB (Snyder et al. 2018). For all



Traveling LMP NSE

	BDC	BEF	DCF	MCQ	SBM	TPB	W3	W9	WHB
SO_4^{2-}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1
NO_3^{1-}	-0.71	-0.48	-0.57	< -1	< -1	< -1	0.89	0.21	< -1
Na^{1+}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	-0.08
Mg^{2+}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1
K^{1+}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	-0.82
DON	-0.61	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1
DOC	-0.44	0.16	0.32	< -1	< -1	0.21	< -1	-0.21	-0.37
Cl^{1-}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	0.11
Ca^{2+}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1

Traveling LMP R^2

	BDC	BEF	DCF	MCQ	SBM	TPB	W3	W9	WHB
SO_4^{2-}	0.10	0.05	0.26	0.00	0.07	0.09	0.01	0.07	0.10
NO_3^{1-}	0.06	0.82	0.74	0.08	0.02	0.22	0.95	0.76	0.24
Na^{1+}	0.64	-	0.31	0.15	-	-	-	-	0.24
Mg^{2+}	0.38	0.42	0.00	0.11	0.00	0.00	0.02	0.05	0.09
K^{1+}	0.20	0.25	0.08	0.15	0.28	0.03	0.03	0.29	0.23
DON	0.02	0.00	0.14	0.27	0.00	0.01	0.05	0.00	0.01
DOC	0.59	0.96	0.34	0.38	0.15	0.73	0.70	0.34	0.63
Cl^{1-}	0.69	-	0.30	0.18	-	-	-	-	0.21
Ca^{2+}	0.45	0.00	0.15	0.16	0.01	0.02	0.01	0.06	0.00

Traveling W3 NSE

	BDC	BEF	DCF	LMP	MCQ	SBM	TPB	W9	WHB
SO_4^{2-}	< -1	< -1	0.01	< -1	< -1	< -1	< -1	-0.06	< -1
NO_3^{1-}	-0.57	0.83	0.46	0.75	< -1	< -1	< -1	0.78	-0.16
Na^{1+}	< -1	0.23	< -1	< -1	< -1	< -1	< -1	-0.31	< -1
Mg^{2+}	< -1	-0.39	< -1	< -1	< -1	< -1	< -1	< -1	< -1
K^{1+}	< -1	-0.68	< -1	< -1	< -1	< -1	< -1	-0.39	< -1
DON	< -1	0.53	< -1	< -1	-0.10	-0.01	-0.69	0.04	-0.07
DOC	-0.62	0.74	< -1	< -1	< -1	-0.14	-0.13	< -1	-0.06
Cl^{1-}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	0.32	< -1
Ca^{2+}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1

Traveling W3 R^2

	BDC	BEF	DCF	LMP	MCQ	SBM	TPB	W9	WHB
SO_4^{2-}	0.28	0.81	0.02	0.06	0.17	0.00	0.20	0.00	0.07
NO_3^{1-}	0.12	0.86	0.80	0.87	0.15	0.05	0.21	0.78	0.55
Na^{1+}	0.20	0.58	0.04	0.27	0.08	0.05	0.02	0.12	0.09
Mg^{2+}	0.11	0.77	0.00	0.29	0.16	0.09	0.00	0.02	0.05
K^{1+}	0.11	0.42	0.01	0.14	0.12	0.48	0.27	0.00	0.14
DON	0.05	0.81	0.25	0.05	0.17	0.00	0.59	0.39	0.09
DOC	0.12	0.92	0.27	0.47	0.00	0.25	0.86	0.41	0.08
Cl^{1-}	0.43	0.05	0.02	0.02	0.00	0.01	0.02	0.46	0.02
Ca^{2+}	0.06	0.78	0.08	0.27	0.00	0.03	0.12	0.02	0.00

All-but-one NSE

	BDC	BEF	DCF	LMP	MCQ	SBM	TPB	W3	W9	WHB
SO_4^{2-}	-0.05	< -1	-0.24	-0.20	0.73	< -1	< -1	< -1	0.04	-0.09
NO_3^{1-}	0.62	0.85	0.35	0.53	0.75	< -1	< -1	0.93	0.78	< -1
Na^{1+}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1
Mg^{2+}	0.48	< -1	< -1	0.35	< -1	-0.07	< -1	< -1	< -1	< -1
K^{1+}	< -1	< -1	-0.34	-0.55	< -1	< -1	-0.33	< -1	< -1	< -1
DON	< -1	0.50	-0.82	0.04	-0.11	< -1	-0.04	< -1	0.52	< -1
DOC	0.18	0.87	-0.99	0.36	0.23	< -1	0.35	< -1	-0.16	0.16
Cl^{1-}	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1	< -1
Ca^{2+}	< -1	< -1	< -1	-0.09	0.66	< -1	< -1	< -1	< -1	< -1

All-but-one R^2

	BDC	BEF	DCF	LMP	MCQ	SBM	TPB	W3	W9	WHB
SO_4^{2-}	0.48	0.55	0.00	0.06	0.75	0.07	0.26	0.01	0.11	0.12
NO_3^{1-}	0.87	0.87	0.76	0.81	0.82	0.09	0.19	0.94	0.79	0.25
Na^{1+}	0.57	0.16	0.46	0.73	0.05	0.09	0.04	0.00	0.05	0.04
Mg^{2+}	0.55	0.16	0.24	0.65	0.46	0.12	0.26	0.10	0.07	0.10
K^{1+}	0.05	0.02	0.15	0.05	0.60	0.01	0.21	0.03	0.02	0.00
DON	0.04	0.80	0.25	0.06	0.07	0.00	0.65	0.08	0.58	0.16
DOC	0.66	0.95	0.22	0.53	0.57	0.21	0.88	0.63	0.55	0.64
Cl^{1-}	0.22	0.07	0.47	0.63	0.01	0.01	0.07	0.00	0.08	0.30
Ca^{2+}	0.56	0.16	0.28	0.59	0.66	0.00	0.19	0.20	0.15	0.02

FIGURE 6 | Nash–Sutcliffe efficiencies and coefficients of determination (R^2) for travelling LMP, travelling W3, and all-but-one RF models.

of these reasons, solutes with a dedicated sensor (e.g., DOC, NO_3^-) are unlikely to benefit from RF models.

It is likely that RF models will become more generalisable as more grab sample and sensor data become available. A notable challenge with biogeochemical machine-learning studies is having enough sampling data to construct robust models. While our dataset consisted of 1289 grab samples across 10 streams, an impressive effort for a biogeochemical dataset, it was relatively small compared to most machine-learning applications (Shen et al. 2020; Underwood et al. 2023). A more extensive array of watersheds (on the order of thousands of sites) would increase

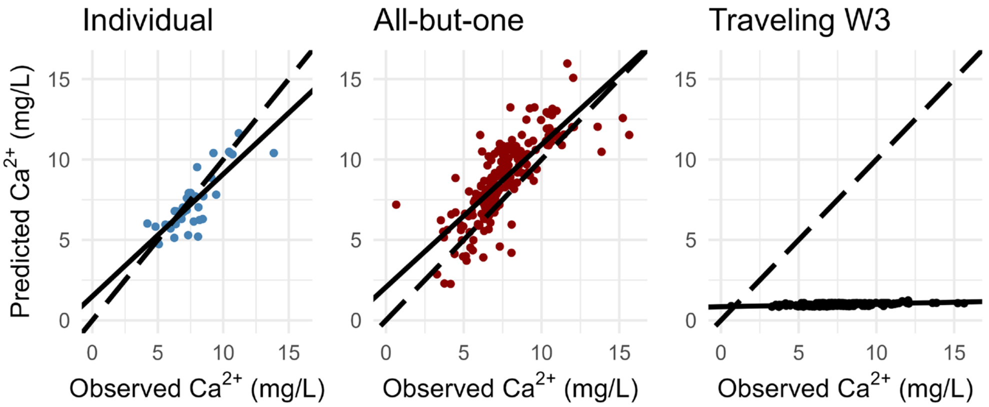


FIGURE 7 | Predicted vs. observed Ca^{2+} concentrations for the Lamprey river (LMP) using the individual, all-but-one, and travelling W3 models. The dashed line represents a 1-to-1 relationship. The solid line is the linear regression.

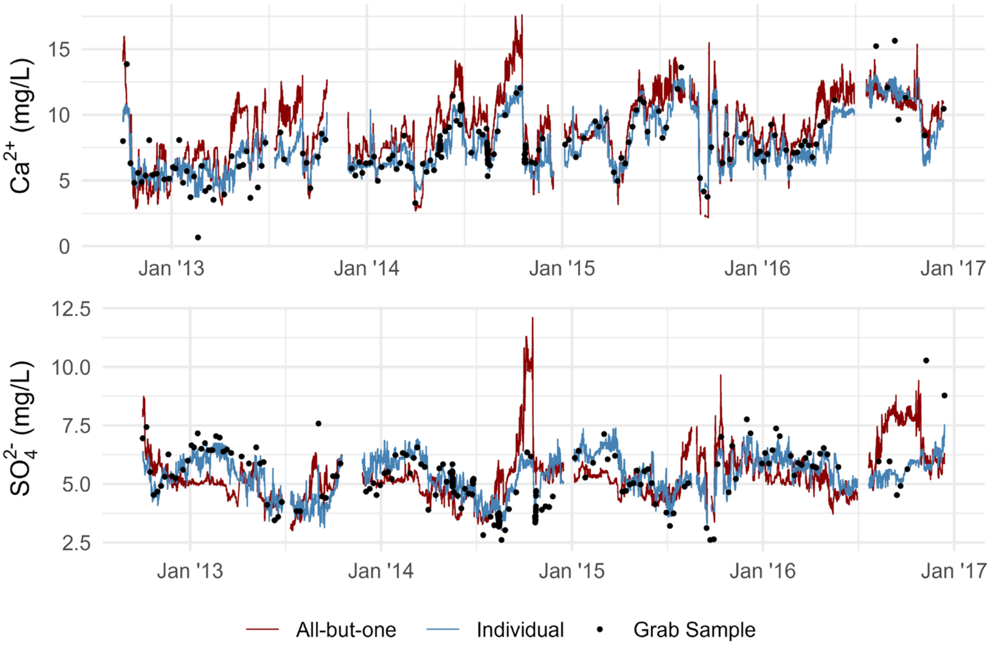


FIGURE 8 | Time series of predicted calcium (Ca^{2+}) and sulfate (SO_4^{2-}) concentrations in LMP.

the biogeochemical diversity of training samples and lead to more generalisable models, especially if watershed characteristics could be included as variables to predict the range of solute concentrations at a site (e.g., Müller et al. 2024). Production of larger stream sensing and solute concentration datasets may open avenues to more advanced machine-learning algorithms such as long short-term memory (LSTM) or neural network models (Anmala and Venkateshwarlu 2019; Zhi et al. 2021).

While our study failed to accurately predict solute concentrations in streams outside of the training set, there were still some successes. We were able to generate individual RF models with as few as 30–40 samples. Our individual RF models outperformed previous machine-learning models in W3 (Green et al. 2021), presumably because we used more stringent data filters, stratified k-fold cross validation, and variable selection via VSURF. Our travelling and all-but-one models were able to capture some of the characteristic temporal variability (Figure 8) even though there was some systematic error in the predictions (Figure 7).

5 | Conclusion

Random forest models were not successful at accurately predicting solute concentrations in watersheds in which they were not trained. This held true even when attempting to predict solute concentrations in subwatersheds or watersheds in close proximity with similar biogeochemistry. Results were improved, but not enough to be useful, by increasing the size of the training dataset by incorporating data from multiple watersheds. Despite these limitations, improvements were made in predicting solute concentrations in the same watershed in which the RF model was trained. Future studies can build upon this work by implementing other machine-learning models, including a more diverse selection of sensor data as more advanced sensors are developed, and incorporating more data as grab sampling campaigns continue. For now, grab sampling campaigns are still necessary in each watershed of interest to generate RF models for data interpolation and solute prediction.

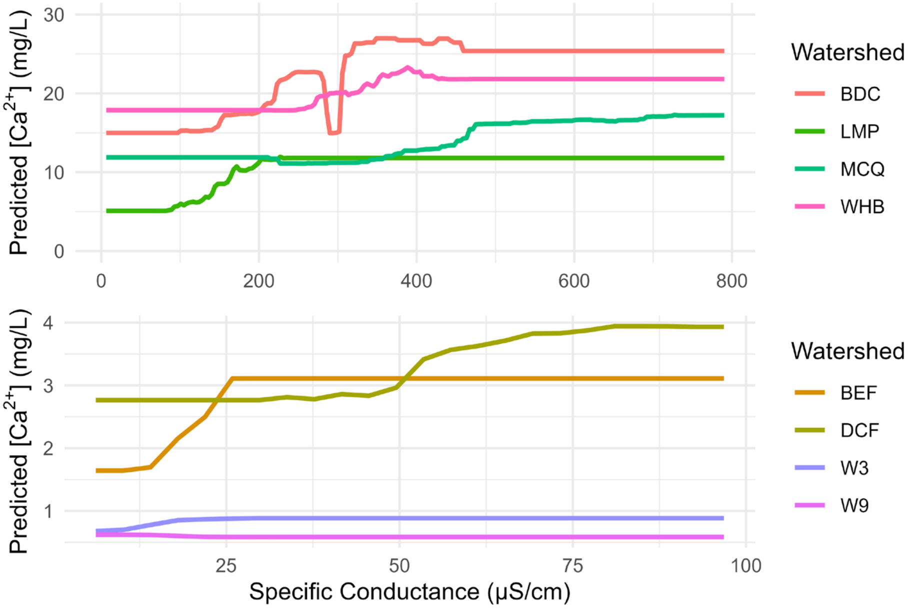


FIGURE 9 | Partial dependence plots visualising calcium (Ca^{2+}) model response to specific conductance for individual RF models. Every variable except specific conductance is held constant so the effect of specific conductance on model response can be observed. The two panels differ in the scale of both axes.

Acknowledgements

Jake Beidler, Erin Cornell, Kelley Gilhooly, and Ryan Ruggiero contributed to the initial exploration of machine learning applied to stream sensor data in a graduate seminar at SUNY ESF. Support for monitoring at the Hubbard Brook watersheds was funded by the National Science Foundation through the Long Term Ecological Research network (DEB-1637685) and Long Term Research in Environmental Biology (DEB-1907683) programs. The New Hampshire EPSCoR High Intensity Aquatic Network was funded by NSF EPS-1101245, the NH Water Resources Research Center, and the New Hampshire Agriculture Experiment Station (NH AES) via the United States Department of Agriculture National Institute of Food and Agriculture Hatch Multi-State Project 1022291, and McIntire-Stennis project 1019522. This is scientific contribution number 3037 of the New Hampshire Agriculture Experiment Station. Additional support was provided by the Department of Energy (DE-SC0024636) and the US Forest Service (19-JV-11242307-056).

Data Availability Statement

The long-term and grab sample data associated with this study are available through the Environmental Data Initiative and posted at the following resource available in the public domain: McDowell, W. 2021. Chemistry of stream water for the Lamprey River Hydrologic Observatory ver 1. Environmental Data Initiative. https://doi.org/10.6073/pasta/33595335d8dc149f8f9082e00045418e (Accessed 2024-11-26). The high-frequency sensor data are posted on CUAHSI Hydroshare: Potter, J. D., L. Koenig, W. H. McDowell, L. Snyder (2020). New Hampshire EPSCoR Intensive Aquatic Network continuous Discharge, Nitrate, fDOM, Temperature, and Specific Conductance Data, HydroShare, https://doi.org/10.4211/hs.8217eab0997d493782ff321ca5f95f28. The grab sample chemistry data for the Hubbard Brook watersheds are available at: Hubbard Brook Watershed Ecosystem Record (HBWatER). 2024. Continuous precipitation and stream chemistry data, Hubbard Brook Ecosystem Study, 1963 – ongoing. ver 10. Environmental Data Initiative. https://doi.org/10.6073/pasta/d2134b69e922988cb056a3c1a837e459 (Accessed 2024-11-26). The code necessary to recreate the machine-learning models is available via GitHub: https://github.com/hollis-harrington/StuckAtHome_RFModels.git

References

Anmala, J., and T. Venkateshwarlu. 2019. "Statistical Assessment and Neural Network Modeling of Stream Water Quality Observations of Green River Watershed, KY, USA." *Water Supply* 19, no. 6: 1831–1840. https://doi.org/10.2166/ws.2019.058.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45, no. 1: 5–32. https://doi.org/10.1023/A:1010933404324.

Campbell, J. L., L. E. Rustad, S. W. Bailey, et al. 2021. "Watershed Studies at the Hubbard Brook Experimental Forest: Building on a Long Legacy of Research With New Approaches and Sources of Data." *Hydrological Processes* 35, no. 1: e14016. https://doi.org/10.1002/hyp.14016.

ESRI. 2011. "ArcMap Desktop." Computer software.

Fazekas, H. M., A. S. Wymore, and W. H. McDowell. 2020. "Dissolved Organic Carbon and Nitrate Concentration-Discharge Behavior Across Scales: Land Use, Excursions, and Misclassification." *Water Resources Research* 56, no. 8: e2019WR027028. https://doi.org/10.1029/2019WR027028.

Fovet, O., G. Humbert, R. Dupas, et al. 2018. "Seasonal Variability of Stream Water Quality Response to Storm Events Captured Using High-Frequency and Multi-Parameter Data." *Journal of Hydrology* 559: 282–293. https://doi.org/10.1016/j.jhydrol.2018.02.040.

Genuer, R., J.-M. Poggi, and C. Tuleau-Malot. 2015. "VSURF: An R Package for Variable Selection Using Random Forests." *R Journal* 7, no. 2: 19. https://doi.org/10.32614/RJ-2015-018.

Green, M. B., L. H. Pardo, S. W. Bailey, et al. 2021. "Predicting High-Frequency Variation in Stream Solute Concentrations With Water Quality Sensors and Machine Learning." *Hydrological Processes* 35, no. 1: e14000. https://doi.org/10.1002/hyp.14000.

Gu, C., W. Anderson, and F. Maggi. 2012. "Riparian Biogeochemical Hot Moments Induced by Stream Fluctuations." *Water Resources Research* 48, no. 9: 2011WR011720. https://doi.org/10.1029/2011WR011720.

Harrison, J. W., M. A. Lucius, J. L. Farrell, L. W. Eichler, and R. A. Relyea. 2021. “Prediction of Stream Nitrogen and Phosphorus Concentrations From High-Frequency Sensors Using Random Forests Regression.” *Science of the Total Environment* 763: 143005. https://doi.org/10.1016/j.scitotenv.2020.143005.

Hensley, R. T., and M. J. Cohen. 2016. “On the Emergence of Diel Solute Signals in Flowing Waters.” *Water Resources Research* 52, no. 2: 759–772. https://doi.org/10.1002/2015WR017895.

Hubbard Brook Watershed Ecosystem Record (HBWatER). 2023. *Continuous Precipitation and Stream Chemistry Data, Hubbard Brook Ecosystem Study, 1963 – Ongoing*. Environmental Data Initiative.

Koenig, L. E., M. D. Shattuck, L. E. Snyder, J. D. Potter, and W. H. McDowell. 2017. “Deconstructing the Effects of Flow on DOC, Nitrate, and Major Ion Interactions Using a High-Frequency Aquatic Sensor Network.” *Water Resources Research* 53, no. 12: 10655–10673. https://doi.org/10.1002/2017WR020739.

Liaw, A., and M. Wiener. 2001. “Classification and Regression by RandomForest.” *Forest* 2, no. 3: 18–22.

Lintern, A., J. A. Webb, D. Ryu, et al. 2018. “Key Factors Influencing Differences in Stream Water Quality Across Space.” *WIREs Water* 5, no. 1: e1260. https://doi.org/10.1002/wat2.1260.

Merriam, J., W. H. McDowell, and W. S. Currie. 1996. “A High-Temperature Catalytic Oxidation Technique for Determining Total Dissolved Nitrogen.” *Soil Science Society of America Journal* 60, no. 4: 1050–1055. https://doi.org/10.2136/sssaj1996.03615995006000040013x.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. “Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations.” *Transactions of the ASABE* 50, no. 3: 885–900. https://doi.org/10.13031/2013.23153.

Müller, M., J. D'Andrilli, V. Silverman, et al. 2024. “Machine-Learning Based Approach to Examine Ecological Processes Influencing the Diversity of Riverine Dissolved Organic Matter Composition.” *Frontiers in Water* 6: 1379284. https://doi.org/10.3389/frwa.2024.1379284.

Nash, J. E., and J. V. Sutcliffe. 1970. “River Flow Forecasting Through Conceptual Models Part I — A Discussion of Principles.” *Journal of Hydrology* 10, no. 3: 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

National Oceanic and Atmospheric Administration. 2024. ”National Climatic Data Center.” https://www.ncei.noaa.gov/cdo-web/datatools.

Navrátil, T., S. A. Norton, I. J. Fernandez, and S. J. Nelson. 2010. “Twenty-Year Inter-Annual Trends and Seasonal Variations in Precipitation and Stream Water Chemistry at the Bear Brook Watershed in Maine, USA.” *Environmental Monitoring and Assessment* 171, no. 1–4: 23–45. https://doi.org/10.1007/s10661-010-1527-z.

Olson, J., and C. Hawkins. 2012. “Predicting Natural Base-Flow Stream Water Chemistry in the Western United States.” *Water Resources Research* 48, no. 2: 2011WR011088. https://doi.org/10.1029/2011WR011088.

Pfaff, J. D., and D. P. Hautman. 1999. *Method 300.1 Determination of Inorganic Anions in Drinking Water by Ion Chromatography*. USEPA.

Potter, B. B., and J. C. Wimsatt. 2005. *Method 415.3 Determination of Total Organic Carbon and Specific UV Absorbance at 254 Nm in Source Water and Drinking Water*. USEPA.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing [Computer software]*. R Foundation for Statistical Computing. https://www.R-project.org/.

Raymond, P. A., and J. E. Saiers. 2010. “Event Controlled DOC Export From Forested Watersheds.” *Biogeochemistry* 100, no. 1: 197–209. https://doi.org/10.1007/s10533-010-9416-7.

Rode, M., A. J. Wade, M. J. Cohen, et al. 2016. “Sensors in the Stream: The High-Frequency Wave of the Present.” *Environmental Science & Technology* 50, no. 19: 10297–10307. https://doi.org/10.1021/acs.est.6b02155.

Shattuck, M. D., H. M. Fazekas, A. S. Wymore, A. Cox, and W. H. McDowell. 2023. “Salinization of Stream Water and Groundwater at Daily to Decadal Scales in a Temperate Climate.” *Limnology and Oceanography Letters* 8, no. 1: 131–140. https://doi.org/10.1002/lol2.10306.

Shen, L. Q., G. Amatulli, T. Sethi, P. Raymond, and S. Domisch. 2020. “Estimating Nitrogen and Phosphorus Concentrations in Streams and Rivers, Within a Machine Learning Framework.” *Scientific Data* 7, no. 1: 1–11. https://doi.org/10.1038/s41597-020-0478-7.

Snyder, L., J. D. Potter, and W. H. McDowell. 2018. “An Evaluation of Nitrate, fDOM, and Turbidity Sensors in New Hampshire Streams.” *Water Resources Research* 54, no. 3: 2466–2479. https://doi.org/10.1002/2017WR020678.

Speir, S. L., L. A. Rose, J. R. Blaszczak, et al. 2024. “Catchment Concentration–Discharge Relationships Across Temporal Scales: A Review.” *WIREs Water* 11, no. 2: e1702. https://doi.org/10.1002/wat2.1702.

Swistock, B. R., P. J. Edwards, F. Wood, and D. R. Dewalle. 1997. “Comparison of Methods for Calculating Annual Solute Exports From Six Forested Appalachian Watersheds.” *Hydrological Processes* 11, no. 7: 655–669. https://doi.org/10.1002/(SICI)1099-1085(199706)11:7<655::AID-HYP525>3.0.CO;2-4.

Tran, Y. B., L. F. Arias-Rodriguez, and J. Huang. 2022. “Predicting High-Frequency Nutrient Dynamics in the Danube River With Surrogate Models Using Sensors and Random Forest.” *Frontiers in Water* 4: 894548. https://doi.org/10.3389/frwa.2022.894548.

Underwood, K. L., D. M. Rizzo, J. P. Hanley, et al. 2023. “Machine-Learning Reveals Equifinality in Drivers of Stream DOC Concentration at Continental Scales.” *Water Resources Research* 59, no. 3: e2021WR030551. https://doi.org/10.1029/2021WR030551.

USDA-NASS. 2024. U”SDA National Agricultural Statistics Service Cropland Data Layer.” https://nassgeodata.gmu.edu/CropScape.

Vermilyea, A. W., S. A. Nagorski, C. H. Lamborg, E. W. Hood, D. Scott, and G. J. Swarr. 2017. “Continuous Proxy Measurements Reveal Large Mercury Fluxes From Glacial and Forested Watersheds in Alaska.” *Science of the Total Environment* 599: 145–155. https://doi.org/10.1016/j.scitotenv.2017.03.297.

Worrall, F., N. J. K. Howden, and T. P. Burt. 2013. “Assessment of Sample Frequency Bias and Precision in Fluvial Flux Calculations – An Improved Low Bias Estimation Method.” *Journal of Hydrology* 503: 101–110. https://doi.org/10.1016/j.jhydrol.2013.08.048.

Wymore, A. S., J. Potter, B. Rodriguez-Cardona, and W. H. McDowell. 2018. “Using In-Situ Optical Sensors to Understand the Biogeochemistry of Dissolved Organic Matter Across a Stream Network.” *Water Resources Research* 54: 2949–2958. https://doi.org/10.1002/2017WR022168.

Wymore, A. S., M. D. Shattuck, J. D. Potter, L. Snyder, and W. H. McDowell. 2021. “The Lamprey River Hydrological Observatory: Suburbanization and Changing Seasonality.” *Hydrological Processes* 35, no. 4: e14131. https://doi.org/10.1002/hyp.14131.

Zanoni, M. G., B. Majone, and A. Bellin. 2022. “A Catchment-Scale Model of River Water Quality by Machine Learning.” *Science of the Total Environment* 838: 156377. https://doi.org/10.1016/j.scitotenv.2022.156377.

Zhi, W., D. Feng, W. P. Tsai, et al. 2021. “From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale?” *Environmental Science and Technology* 55, no. 4: 2357–2368. https://doi.org/10.1021/acs.est.0c06783.