Author Contributions:
Conceptualization: Keira Johnson, Kathi Jo Jankowski, Joanna C. Carey, Diane McKnight, William H. McDowell, Adam S. Wymore, Amanda E. Poste, Pamela L. Sullivan
Data curation: Keira Johnson, Kathi Jo Jankowski, Joanna C. Carey, William H. McDowell, Adam S. Wymore, Nicholas

Climate, Hydrology, and Nutrients Control the Seasonality of Si Concentrations in Rivers

Keira Johnson[1], Kathi Jo Jankowski[2], Joanna C. Carey[3], Lienne R. Sethna[4], Sidney A. Bush[1], Diane McKnight[5], William H. McDowell[6], Adam S. Wymore[6], Pirkko Kortelainen[7], Jeremy B. Jones[8], Nicholas J. Lyon[9], Hjalmar Laudon[10], Amanda E. Poste[11], and Pamela L. Sullivan[1]

[1]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA, [2]U.S. Geological Survey, Upper Midwest Environmental Sciences Center, La Crosse, WI, USA, [3]Math, Analytics, Science & Technology Division. Babson College, Wellesley, MA, USA, [4]St. Croix Watershed Research Station, Marine on St. Croix, MN, USA, [5]Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder, Boulder, CO, USA, [6]Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, USA, [7]Finnish Environment Institute, Helsinki, Finland, [8]Institute of Arctic Biology & Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK, USA, [9]National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA, USA, [10]Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, Sweden, [11]Department of Arctic Ecology, Norwegian Institute for Nature Research, Tromsø, Norway & Norwegian Institute for Water Research, Oslo, Norway

Abstract The seasonal behavior of fluvial dissolved silica (DSi) concentrations, termed *DSi regime*, mediates the timing of DSi delivery to downstream waters and thus governs river biogeochemical function and aquatic community condition. Previous work identified five distinct DSi regimes across rivers spanning the Northern Hemisphere, with many rivers exhibiting multiple DSi regimes over time. Several potential drivers of DSi regime behavior have been identified at small scales, including climate, land cover, and lithology, and yet the large-scale spatiotemporal controls on DSi regimes have not been identified. We evaluate the role of environmental variables on the behavior of DSi regimes in nearly 200 rivers across the Northern Hemisphere using random forest models. Our models aim to elucidate the controls that give rise to (a) average DSi regime behavior, (b) interannual variability in DSi regime behavior (i.e., Annual DSi regime), and (c) controls on DSi regime shape (i.e., minimum and maximum DSi concentrations). Average DSi regime behavior across the period of record was classified accurately 59% of the time, whereas Annual DSi regime behavior was classified accurately 80% of the time. Climate and primary productivity variables were important in predicting Average DSi regime behavior, whereas climate and hydrologic variables were important in predicting Annual DSi regime behavior. Median nitrogen and phosphorus concentrations were important drivers of minimum and maximum DSi concentrations, indicating that these macronutrients may be important for seasonal DSi drawdown and rebound. Our findings demonstrate that fluctuations in climate, hydrology, and nutrient availability of rivers shape the temporal availability of fluvial DSi.

Plain Language Summary The amount of dissolved silicon (DSi) in rivers is an important control on numerous ecological and biogeochemical processes, such as types of algae that bloom and rates of carbon sequestration. Compared to our knowledge of other nutrients, such as nitrogen and phosphorus, we have limited understanding of what controls the timing and concentration of DSi in rivers. Previous work identified five distinct seasonal patterns of DSi concentrations in rivers across the Northern Hemisphere; here we look at the environmental variables that control these seasonal patterns. We found that rivers often have one to five seasonal patterns over time due to interannual shifts in temperature, evapotranspiration, and streamflow. In addition, we found that the average shape of the seasonal pattern for a given river, specifically minimum and maximum DSi concentrations, was related to nitrogen (N) and phosphorus (P) concentrations, highlighting linkages between N, P, and DSi cycling in rivers. This work identifies why river DSi concentrations exhibit both within and between year variability, highlighting that temperature, streamflow, and nutrient availability control the timing of river DSi availability for biological uptake.

J. Lyon, Hjalmar Laudon, Amanda
E. Poste
Formal analysis: Keira Johnson, Kathi
Jo Jankowski, Joanna C. Carey, Sidney
A. Bush, Nicholas J. Lyon, Pamela
L. Sullivan
Funding acquisition: Kathi Jo Jankowski,
Joanna C. Carey, Pamela L. Sullivan
Methodology: Keira Johnson, Kathi
Jo Jankowski, Joanna C. Carey,
Pirkko Kortelainen, Jeremy B. Jones,
Nicholas J. Lyon, Pamela L. Sullivan
Project administration: Pamela
L. Sullivan
Supervision: Kathi Jo Jankowski, Joanna
C. Carey, Pamela L. Sullivan
Validation: Keira Johnson
Visualization: Keira Johnson
Writing – original draft: Keira Johnson,
Kathi Jo Jankowski, Joanna C. Carey,
Lienne R. Sethna, Sidney A. Bush, Pamela
L. Sullivan
Writing – review & editing:
Keira Johnson, Kathi Jo Jankowski, Joanna
C. Carey, Lienne R. Sethna, Sidney
A. Bush, Diane McKnight, William
H. McDowell, Adam S. Wymore,
Pirkko Kortelainen, Jeremy B. Jones,
Nicholas J. Lyon, Hjalmar Laudon,
Amanda E. Poste, Pamela L. Sullivan

1. Introduction

Hydrologic regimes, and to a lesser extent biogeochemical regimes, have been characterized across many different environments (e.g., Berhanu et al., 2015; Pinay et al., 2007; Van Meter et al., 2020; Yang et al., 2002) and have provided a valuable tool for evaluating how rivers respond to environmental change (Bard et al., 2015; Schnorbus et al., 2014; Yu et al., 2018). Regimes provide a useful framework for describing the annual and seasonal behavior of various watershed processes, such as discharge, dissolved oxygen, or productivity (Bernhardt et al., 2022; Post & Jones, 2001; dos Reis Oliveira et al., 2019) and are characterized by the timing, magnitude, and rates of change in peaks and troughs. In this way, regime characterization provides a means for comparing seasonal patterns among catchments and years, elucidating when shifts in systems may be taking place due to changes in drivers, such as land use (Júnior et al., 2015; Thanapakpawin et al., 2007) and river flow regulation (Magilligan & Nislow, 2005; Peñas & Barquín, 2019; Wang et al., 2017).

Seasonal regimes are integral to understanding how river ecosystems function and can influence numerous biological processes as well as the overall condition, and development, and stability of aquatic communities (Van Meter et al., 2020; Seybold, Burgin, et al., 2022; Seybold, Fork, et al., 2022). Seasonal regimes reflect the integrated signal of a stream's hydroclimatic conditions and watershed characteristics, including discharge, climate, lithology, land use, and vegetation (Bernhardt et al., 2022; Bolotin et al., 2022). For example, the elemental regimes of nutrients such as nitrogen (N) and phosphorus (P) are influenced by upland sources and moderated by in-stream biological uptake, transformation and removal pathways (e.g., Van Meter et al., 2020). As a key control on aquatic primary production, nutrient availability directly influences energy availability for higher trophic levels. Thus, seasonal fluctuations of nutrient concentrations can drive shifts in ecosystem structure and function, influence the diversity and abundance of species, and control the overall condition and resilience of aquatic ecosystems both in the rivers themselves and in downstream freshwater and coastal ecosystems (Loewen et al., 2021; Van Meter et al., 2020; Williamson et al., 2021).

Silicon (Si) is an understudied nutrient in aquatic ecosystems, especially relative to N and P. Together these nutrient (Si, N, and P) concentrations respond to both terrestrial and aquatic processes, thereby reflecting how climate and land cover affect the terrestrial-aquatic continuum (Bouwman et al., 2013; Conley, 2002; Cornelis et al., 2011; Struyf et al., 2010). A variety of biogeochemical and hydrologic factors govern the concentration and seasonal regime of dissolved Si (DSi) in rivers (Cornelis et al., 2010; Phillips, 2020), including: (a) lithogenic mineral content (Meybeck, 1987; West et al., 2005; White & Blum, 1995), (b) chemical weathering rates (Drever, 1994; Gaillardet et al., 1999), (c) regional erosion rates (Hilley & Porder, 2008), (d) hydrologic regime (Carey et al., 2020; Godsey et al., 2009); and (e) terrestrial vegetation cycling (Carey & Fulweiler, 2012; Fulweiler & Nixon, 2005; Phillips, 2020). DSi has the potential to control phytoplankton and benthic algal species composition, as diatoms (ubiquitous autotrophs in freshwater and marine systems) require DSi for growth and are an important contributor to the base of many aquatic food webs. When streams are not DSi limited, diatoms can outcompete other species assemblages (Sommer, 1988), and diatom blooms can drawdown stream DSi concentrations (Casey et al., 1981; House et al., 2001; Wall et al., 1998). However, DSi limitation, especially in relation to N and P, can result in non-siliceous species, including cyanobacteria, outcompeting diatoms (Conley et al., 1993; Officer & Ryther, 1980; Teubner & Dokulil, 2002). DSi limitation of diatom growth has been widely documented in aquatic systems, including freshwaters and northern marine ecosystems, particularly during the summer and autumn (Giesbrecht & Varela, 2021; McNair et al., 2018; Schelske et al., 1986). As such, changes in DSi availability, especially how DSi concentration varies seasonally, have direct impacts on phytoplankton community and ecosystem structure and function.

Recent work by Johnson, Jankowski, Carey, Lyon, et al. (2024) showed substantial variability in seasonal dynamics of DSi across river systems. The authors identified five common DSi regimes across 201 streams in the Northern Hemisphere, predominately from the US and Scandinavia, that span multiple climates (Johnson, Jankowski, Carey, Lyon, et al., 2024; Figure 1). Many streams exhibited variability in DSi regime membership over time, displaying different seasonal patterns across years. While these analyses demonstrated that there are more diverse DSi regime patterns than previously recognized, the underlying drivers of the identified DSi regimes remain unclear. The objective of this follow-up study is to understand why different streams exhibit vastly different seasonal patterns in DSi and why these patterns change over time.

To meet our objective and address these knowledge gaps, we leverage an existing database that includes DSi concentrations and discharge from 201 rivers spanning the Northern Hemisphere Johnson, Jankowski, Carey,
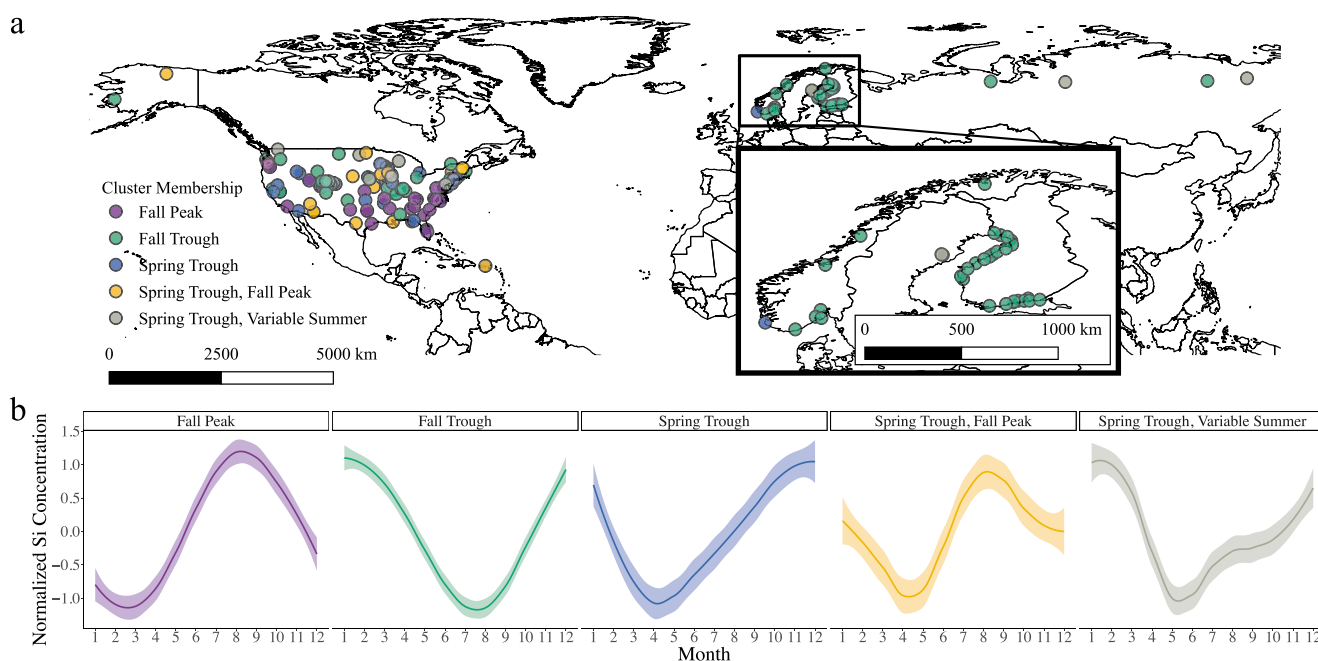
Figure 1. (a) Distribution of sites colored by modal DSi regime behavior and (b) the shape of the DSi five regimes that emerged from Johnson, Jankowski, Carey, Lyon, et al. (2024).

Sethna, et al. (2024) and the five distinct DSi regime behaviors identified by Johnson, Jankowski, Carey, Lyon, et al. (2024). We also use newly acquired hydrologic, climatic, geologic, geomorphic, and biologic watershed characteristic data from these same rivers to address three fundamental questions: (a) What factors are associated with distinguishing Average DSi regime behavior across streams?, (b) What controls interannual variability in DSi regime membership (e.g., Annual DSi regime behavior)?; and finally, (c) Within Average DSi regimes, what processes control their overall shape? To address these questions, we apply a series of random forest machine learning models to identify the most important environmental variables to differentiate among DSi regimes. We hypothesized that average DSi regime behavior of a given river would be associated with metrics related to average seasonality (e.g., average precipitation, net primary productivity (NPP)). In addition, we hypothesized that interannual variability in DSi regime membership of a given river would be associated with changes in these seasonal patterns, such as shifts in precipitation and temperature over time. Finally, we hypothesized that minimum and maximum concentrations would respond to different factors across DSi regimes. Overall, this work aims to advance our ability to predict in-stream DSi availability in light of changing climate and land cover conditions and assess the potential implications for aquatic ecosystem structure and function.

2. Methods

2.1. Si Data Sets and DSi Seasonality Regimes

We used 189 of the 201 rivers published in the DSi data set (Johnson, Jankowski, et al., 2023) and in the regime classification established by Johnson, Jankowski, Carey, Lyon, et al. (2024) (Figure 1a) for this analysis. This data set includes monthly average DSi concentrations estimated from the Weighted Regression on Time, Discharge, and Season (WRTDS) model (Hirsch et al., 2010; Johnson, Jankowski, Carey, Sethna, et al., 2024). Data were sourced from published and/or publicly available data sets that: (a) ranged across eight climate zones, (b) occurred between 18 deg N and 70 deg N, (c) varied in drainage area from <1 km^2 to nearly 3 million km^2, (d) had a mean stream discharge from <0.01 m^3/s to nearly 20,000 m^3/s, (e) were perennial, and (f) had at least 5 years of DSi measurements and continuous streamflow.

Prior work found that five distinct regimes best explained the variability in average DSi seasonal concentration patterns in these rivers (Figure 1b; Johnson, Jankowski, Carey, Lyon, et al., 2024). The DSi regimes were characterized by the timing of their minimum and maximum: (a) fall peak; (b) fall trough; (c) spring trough; (d)

Table 1
Input Variables for Random Forest Model, Source, Scale, and Units

Parameter	Data source	Units	Spatial resolution	Temporal resolution used in models
Air Temperature	Global Historical Climatology Network and Climate Anomaly Monitoring System (GHCN_CAMS) Gridded 2 m Temperature (Land)	degrees Celsius	0.5°	Annual
Precipitation	National Center for Environmental Information Global Precipitation Climatology Project Monthly Precipitation Data Record	mm/day	2.5°	Annual
Evapotranspiration	MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500 m SIN Grid	kg/m^2	500 m	Annual
Maximum Snow-Covered Area	MODIS Snow and Ice mapping project 8-Day MOD10A2	proportion of watershed	500 m	Annual
Green-Up Day	MODIS/Terra + Aqua Land Cover Dynamics Yearly L3 Global 500 m SIN Grid	date	500 m	Annual
Net Primary Productivity (NPP)	MODIS/Terra Net Primary Production Gap-Filled Yearly L4 Global 500 m SIN Grid	$kgC/m^2/year$	500 m	Annual
Land Cover	USGS Earth Observation and Science Center Global Land Cover Characterization	percent watershed area	1 km	Static
Lithology	Global Lithological Map Database v1.0; PANGAEA	percent watershed area	0.5°	Static
Maximum Daylength	daylength function from R package chillR	hours	NA	Static
Median N	Chemistry Data	mg/L	NA	Static
Median P	Chemistry Data	mg/L	NA	Static
Coefficient of Variation in Q (CV(Q))	Discharge Data	NA	NA	Static
5th percentile Q	Discharge Data	cms	NA	Annual
95th percentile Q	Discharge Data	cms	NA	Annual
Day of minimum Q	Discharge Data	day of year	NA	Annual
Day of maximum Q	Discharge Data	day of year	NA	Annual

Note. See Tables S4 and S5 in Supporting Information S1 for details on land use and lithology classification. Parameters labeled as "Annual" in the Temporal Resolution Used in Models column were used in the Average, Annual, and Minimum/Maximum Si random forest models, whereas parameters labeled as "Static" were used only in the "Average" and Minimum/Maximum Si random forest models. MODIS = Moderate Resolution Imaging Spectroradiometer; SIN = Sinusoidal tile grid.

spring trough-fall peak; and (e) spring trough-variable summer. Of the 189 rivers included in this study, the fall trough regime contained the largest number of sites (62), whereas all other DSi regimes contained between 29 and 37 sites (Figure S1 in Supporting Information S1). Interestingly, all DSi regime types were observed across most of the major Köeppen-Geiger climate classification zones but some regimes, including the fall peak and fall trough regimes, exhibited higher stability than other regimes (Johnson, Jankowski, Carey, Lyon, et al., 2024).

2.2. Driver Acquisition and Harmonization

To evaluate which variables were associated with DSi regime classification, we sourced watershed and climate characteristics from globally available spatial data sets (Table 1, Table S1 in Supporting Information S1), hydrologic data from continuously monitored discharge (Table S2 in Supporting Information S1), and nutrient drivers from discretely sampled stream nitrogen (N) and phosphorus (P) concentrations (Table S2 in Supporting Information S1).

2.2.1. Spatial Data

We acquired globally available spatial data sets (Table 1) using the NASA Application for Extracting and Exploring Analysis Ready Samples portal (AρρEEARS), which includes geospatial data from a variety of federal data archives. We used data layers with global coverage in order to have consistent data sources across the extent of our data set. We used watershed boundaries to extract spatial data from the gridded data sources, and acquired watershed boundaries from existing data sources where possible (Table S3 in Supporting Information S1). When watershed shape files were not available, we used the HydroBASINS database (Lehner & Grill, 2013) to generate

a watershed boundary. HydroBASINS provides a global coverage of nested sub-basins at different scales. At its highest level of sub-basin breakdown, HydroBASINS divides a basin into two sub-basins at every location where two river branches meet given that each sub-basin has an upstream area of at least 100 km^2. To build watersheds for sites in our data set, we identified the basin intersecting the sample coordinates at the finest level of HydroBASINS delineation then iteratively included all upstream basins. Once all of these basin polygons were identified, we fused them into a single shape and used that as the watershed boundary. HydroBASINS are quite large relative to some of our streams (on average, the most granular delineation of basin results in polygons of ~100 km^2), but for this work HydroBASINS was not used to delineate any basins smaller than 2,000 km^2.

Data were available at different temporal resolutions (Table 1), which we then summarized for use in either the "Average regime" or "Annual regime" models (see methods below). Surface air temperature, precipitation, green-up day, and NPP were all available as annual mean values. We used the long-term mean of all annual values for the Average regime model and annual mean values for the Annual regime models. Evapotranspiration data were available on an 8-day time step, which we summarized to annual mean values. Snow-covered area (percent of watershed covered by snow) was also available at an 8-day time step, from which we extracted the maximum annual value for use in our models. Land cover and lithology and elevation data were all static (not time varying) and were only used in the Average regime and DSi minimum and maximum concentration models. Land cover data was sourced from the U.S. Geological Survey (USGS) Global Land Cover Classification product which classifies land cover into urban, agricultural, rangeland, forest, open water, wetland, barren, tundra, and perennial snow and ice and associated subclasses (e.g., forest cover includes deciduous, evergreen, and mixed classes) following the land cover classification in (Anderson et al., 1976) (Table S4 in Supporting Information S1). Lithology data was sourced from PANGEA data set and lumped into volcanic, sedimentary, plutonic, metamorphic, and carbonate/evaporite (Table S5 in Supporting Information S1).

Diatoms typically account for ~80% of periphyton biomass and can have significant effects on stream DSi concentrations (Casey et al., 1981; House et al., 2001; Wall et al., 1998). Processes such as stream velocity, and nutrient and light availability exert strong controls on diatom production, particularly benthic algal (Francoeur & Biggs, 2006; Hill & Dimick, 2002; Liess et al., 2009). Here we attempted to capture some of these factors through median N and P concentrations, maximum discharge, and green-up day but we were unable to include direct measures of diatom uptake, stream velocity, or subannual differences in nutrient and light availability due to lack of data availability.

2.2.2. Discharge Data

Daily discharge (Q) data were used to calculate annual discharge metrics to describe the timing and magnitude of low and high discharge and variability in discharge (Table 1). Discharge records spanned the period of record of DSi concentration. We used the annual coefficient of variation in discharge (CV(Q)) to describe the annual variability in discharge and the annual 5th and 95th percentiles of Q (Q5 and Q95), day of year of annual minimum Q, and day of year of annual maximum Q to describe the magnitude and timing of low and high discharge. Minimum and maximum values were evaluated using visual assessment to ensure they were representative of seasonal discharge conditions and not anomalously high or low values caused by instrumentation error. These metrics were calculated for each year to generate annual values and averaged over all the years to get a long-term average value for each site. Discharge source data information is provided in Table S2 in Supporting Information S1.

2.2.3. Nutrient Data

Median nitrogen (N) and phosphorus (P) concentrations were used as an index of the trophic state of each river (Dodds & Smith, 2016; Paerl et al., 2016). We used data reported as nitrate (NO_3) or nitrate and nitrite (NO_x) for N concentrations and data reported as phosphate (PO_4) or soluble reactive phosphorus (SRP) for P concentrations. All nutrient species represent inorganic dissolved forms available for biologic uptake. Median concentrations were calculated from discrete observations spanning the period of record of DSi observations. Values were adjusted to account for minimum detection limits and occurrences of spurious values (e.g., 0 or below) were filtered out. The median of all values within a given species (e.g., NO_3, NO_x, SRP, PO_4) was calculated for each site. Due to differences in sampling strategies year to year, annual median values of N and P were not calculated, and one median value of N and P across the period of record was used.

2.3. Random Forest Analysis

We used a series of 12 random forest models to address our hypotheses. Specifically, we evaluated which factors governed (a) membership in long-term average seasonal DSi regimes (1 model, "Average model"), (b) inter-annual DSi regime membership (1 model, "Annual model"), and (c) the shape of the DSi regimes by assessing controls on the minimum and maximum DSi concentration within each Average DSi regime (10 models). Random forest is a supervised machine learning technique that constructs many decision trees, and selects outputs based on the modal decision from all trees (Breiman, 2001; Cutler et al., 2007; Regier et al., 2023). Each tree is trained on a subset of the data, and uses a random selection of input variables, which leads to many uncorrelated models and allows for classification without overfitting (Ishwaran et al., 2011). Random forest models have been widely applied in hydrology and biogeochemistry because of their interpretability and flexibility (Yang & Olivera, 2023; Bolotin et al., 2022; Konapala et al., 2020; Larras et al., 2017; Harrison et al., 2021).

Random forest classification models were used to predict Average and Annual DSi regime membership (e.g., fall peak, fall trough, spring trough, spring trough-fall peak, spring trough-variable summer) and random forest regression models were used to predict minimum and maximum DSi concentration within each regime. The Average model and the minimum and maximum DSi concentration prediction random forest models were built using all variables listed in Table 1, specifically temporally variable drivers including climate (maximum snow-covered area, precipitation, temperature), productivity (NPP, green-up day, maximum daylength, N, and P), hydrology (low flow, high flow, CV(Q), day of min flow, day of max flow) as well as static basin characteristics including land cover and lithology. The Annual model was built using only temporally variable data (i.e., listed as "Annual" in Table 1). Environmental variables come from different sources with different periods of record. To ensure all variables would be included, the temporally variable model only includes data between 2002 and 2019, which is the range of dates where data for all temporally variable drivers were available (Table 1). Only sites with at least 10 years of overlapping driver and DSi data were used in the Annual model. Random forest models were built in the *randomforest* package in R.

2.3.1. Variable Selection

Recursive feature elimination is a backwards selection algorithm that is widely used in random forest modeling to select the optimal number of input features (or environmental variables in this study) that are most relevant to predicting the output (Darst et al., 2018; Das et al., 2022). Recursive feature elimination builds a base model using all features, ranks them by importance, and removes the feature with the least importance. The model is then rebuilt excluding that feature. This process is repeated iteratively for each combination of features. Model performance was evaluated at each interval using the root mean square error (RMSE), and the optimal combination of features was selected based on the model with the lowest RMSE for each random forest model. Recursive feature elimination was implemented using R package *caret* using five repeats of five-fold cross validation, which is a resampling procedure that reduces overfitting and improves performance of machine learning models built on small sample sizes.

2.3.2. Random Forest Model Parameterization

Model performance was evaluated using out-of-bag error, which describes the proportion of time a given stream was misclassified (e.g., wrong regime), for classification models, and mean squared error (MSE) for regression models. Both out-of-bag error and MSE are minimized in tuning and optimization. Each model was tuned independently to achieve the best classification accuracy or lowest MSE by adjusting the number of trees (ntree) in the forest and the number of variables tried at each node in a given tree (mtry). Values of 100–2000 were tested for the ntree parameter (similar to the ntree parameter optimization employed by Naghibi et al., 2017; Pham et al., 2021). The tuneRF function from the R package *randomforest* was used to tune the mtry parameter. Each base model was tuned before recursive feature elimination was employed, and then the model was re-tuned after recursive feature elimination was employed. For classification models, the sampling size was set to sample equally across all five DSi regimes despite unequal distribution of sites across DSi regimes.

2.3.3. Random Forest Model Variable Interpretation

The role of variables retained in each model was assessed using: (a) their importance in classifying DSi regime membership or predicting minimum/maximum DSi concentration and (b) their marginal impact on predictor

variables. Variable importance was evaluated using the mean decrease in accuracy (derived from out-of-bag error) for classification models and the mean decrease in MSE for regression models. In both cases, permutations of the model are built where one variable is removed from the input variable set and the performance (either classification accuracy or MSE) are evaluated without that variable included. Marginal impact of input variables was evaluated using partial dependence plots for the top six most important variables for minimum and maximum prediction models (Friedman, 2001; Greenwell, 2017). Partial dependence plots allow for evaluation of the relationship between input and predictor variables. They account for the average effect of other predictors in the model, and remove the effect of interactions with other features, thus removing the effect of collinearity between input variables. The effect of a given variable is evaluated using a partial dependence function which marginalizes the model output over the distribution of input variables, not including the variable of interest, so that the output function (yhat) shows the relationship between the variable of interest and the predictor variable.

3. Results

3.1. Model Parameters

A total of 30 environmental variables were included in random forest modeling to represent climate, primary productivity, discharge, lithology, and land cover characteristics at each site (Table 1). Correlations between all variables were examined; the strongest (e.g., highest R^2) correlations were observed between high flow and low flow and among climate variables (Figure S2 in Supporting Information S1). Land cover and lithology variables typically did not correlate strongly with all other drivers. Although multicollinearity generally does not impact the predictive power of random forest models because each tree only uses a subset of input features, high correlation between variables can skew the relative importance of included correlated drivers (Cammarota & Pinto, 2021; Strobl et al., 2007). The relative impact of important variables was assessed using partial dependence plots, which marginalize across all other variables and thus account for collinearity between variables.

3.2. Average DSi Regime Model Classification

The random forest model for predicting Average DSi regime behavior over the period of record ("Average" model) retained all 30 of the input variables and accurately classified sites into their DSi seasonality regime 59% of the time (out-of-bag error of 41%). Class error, which describes misclassification within a given DSi regime, ranged from 24% in the fall peak regime ($n = 37$) to 69% in the spring trough regime ($n = 29$) (Figure 2), with spring trough regimes (ST; STFP; STVS) predicted more poorly than the fall peak and fall trough regimes.

Climate and primary productivity variables emerged as most important in distinguishing among DSi regimes (Figure 3). The most important variables were maximum proportion of snow-covered area, maximum daylength, green-up day, and temperature. Additionally, some hydrologic variables including CV(Q), low flow discharge (q(5)), and high flow discharge (q(95)) emerged as important.

The range of observed values for the six most important variables differed across DSi regimes (Figure 4). Distributions of important variable values were typically most different between fall peak and fall trough regimes, with distributions of important variable values from regimes exhibiting a spring drawdown falling between the two. Specifically, values for maximum snow-covered area, temperature, and green-up day in the fall peak regime were the most different from other DSi regimes, indicating that sites within this regime were less snowy, warmer, and exhibited an earlier green-up day than the other DSi regimes. Values of variables in the fall trough regime were more often closely aligned with values for the three types of spring-trough regimes than the fall peak regime (e.g., similar maximum snow-covered area, green-up day, temperature, and CV(Q)), but not always (e.g., maximum daylength). In particular, the fall trough and spring trough-variable summer regimes showed similar distributions of all variables; both had higher proportions of maximum snow-covered area, longer maximum daylengths, colder temperatures, and lower evapotranspiration than other DSi regimes. The spring trough and spring trough-fall peak regimes had similar distributions of driver variables, which often fell between the fall peak and fall trough/spring trough-variable summer regime distributions. Finally, the fall peak and spring trough-fall peak regimes showed higher CV(Q) compared to other DSi regimes.
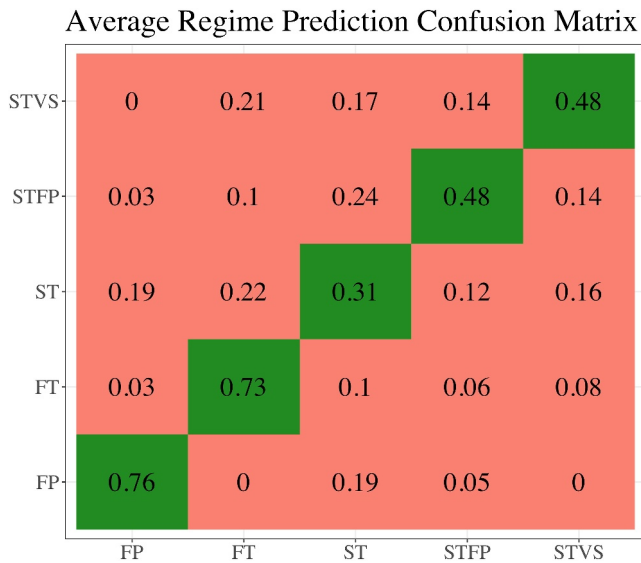
Average Regime Prediction Confusion Matrix

Figure 2. Confusion matrix showing the classification accuracy of the Average DSi regime model. Across a given row, each box indicates the proportion of time the sites within a given regime were classified into other regimes, with green boxes indicating the proportion of time sites were correctly classified and pink boxes indicating the proportion of time sites were inaccurately classified into each other regime. For example, in the fall peak (FP) row (bottom row), sites were correctly classified 76% of the time, and inaccurately classified into the spring trough (ST) regime 19% of the time and into the spring trough-fall peak (STFP) regime 5% of the time. Rows sum to 1. Regimes are fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS).

3.3. Annual DSi Regime Model Classification

The random forest model for predicting Annual DSi regime behavior over the period of record ("Annual" model) included 147 of the 189 sites included in the Average regime model. The number of sites was reduced to match a minimum of 10 years of DSi data between 2002 and 2019 (period where all driver variables were available). Between 2002 and 2019, many sites exhibited multiple DSi regime memberships (Figure 5). In general, the spring trough regimes showed lower modal membership proportions compared to the fall peak and fall trough regimes (Figure 5a). Additionally, the sites in the spring trough regimes showed membership across more regimes compared to sites in the fall peak and fall trough regimes (Figure 5b). Of the 147 sites included, 79 sites were in the same DSi regime each year, 46 sites had membership in two DSi regimes over their period of record, 17 sites had membership in three DSi regimes, and 5 sites had membership in four DSi regimes (Figure S3 in Supporting Information S1).

The Annual regime classification model performed better than the Average regime classification model. Sites were accurately classified 80% of the time (out-of-bag error = 20%), with a similar distribution of out-of-bag error to Average regime classification behavior (Figure 6). Fall trough and fall peak regimes showed lower out-of-bag error rates than the spring trough regimes. The largest increases in prediction accuracy between the Average and Annual models occurred for spring trough, spring trough-fall peak, and spring trough-variable summer regimes, with prediction accuracy increasing by 123%, 50%, and 78%, respectively (Figures 2 and 6) versus 13% and 19% for fall peak and fall trough regimes, respectively.

All variables except day of minimum flow and day of maximum flow were retained in the Annual regime model. A similar suite of variables emerged as most important compared to the Average model, with temperature, evapotranspiration, q5, q95, maximum snow-covered area, and precipitation in the top six (Figure 7). We interpret variables that emerge as important within the Annual model to exhibit interannual variability that drives interannual shifts in DSi regime membership (e.g., interannual variability in temperature may drive interannual variability in DSi seasonality).

3.4. Prediction of Average DSi Regime Minimum and Maximum

To better understand the drivers of DSi cycling that create the shapes of the DSi regime classifications, we predicted minimum and maximum DSi concentrations within each DSi regime using regime specific models (for a total of 10 models–one predicting minimum and one predicting maximum DSi concentration for each of the five DSi regimes). In general, models predicting minimum DSi concentration performed better (lower MSE) than models predicting maximum DSi concentration (Table S6 in Supporting Information S1). The best performing model was the minimum prediction model for the fall trough regime (MSE = 1.42). The worst performing model was the maximum prediction model for the spring trough-variable summer regime (MSE = 5.67).

Different variables and numbers of variables were retained for each model (Figure 8). The minimum DSi concentration prediction model for the fall peak, fall trough, and spring trough-variable summer regimes and the maximum DSi concentration prediction model for the fall peak and fall trough regimes retained the largest number of variables. The minimum DSi concentration prediction model for the spring trough and spring trough-fall peak regimes and the maximum DSi concentration prediction model for the spring trough and spring trough-variable summer regimes retained the fewest variables. Median N and P concentrations showed the strongest and most consistent importance across both the minimum and maximum prediction models. The average increase in MSE if N and P were removed across minimum models was 0.44 and 0.20, respectively and the average increase in MSE if N and P were removed across maximum models was 0.82 and 0.68, respectively. Additionally, in both the minimum and maximum prediction models, CV(Q), precipitation, proportion of land that is evergreen/needleleaf forest, q(95) and q(5) emerged as most important (Figure 8).

Figure 3. Variable importance, indicated by mean decrease in accuracy, for the Average regime classification model. Darker red indicates higher importance of a given variable. The *y*-axis shows variables retained in the model, ordered from most-to-least important for the overall regime classification. The "overall model" column (far right) indicates the importance of each variable in classifying regime membership across all five regimes, whereas the regime specific columns indicate variable importance in classifying sites within a given regime: fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS). A mean decrease in accuracy of 0.05 in the overall model column indicates that when a given variable was removed, the overall model performance declined by 5%. In contrast, when a variable shows a high mean decrease of 0.10 for a given regime (e.g., green-up day for fall peak), it indicates that removal of that variable would increase the out-of-bag error for that regime by 10% (i.e., sites within that regime would be more frequently misclassified).

Partial dependence plots were used to evaluate the relationship between input variables and minimum and maximum DSi concentrations generated by the random forest regression models (Figure 9). We evaluated these plots for variables that were consistently retained and important (as indicated in darker red, Figures 8a and 8b for minimum and maximum, respectively) in predicting DSi minimum and maximum concentrations. Similar relationships were observed for both the minimum and maximum with most variables. N and P had the strongest and most consistent relationships with minimum DSi concentrations. In addition, both variables exhibited threshold behavior, but in different ways; sometimes shifting from negative to positive (e.g., N concentration) or exhibiting abrupt upward shifts (e.g., P concentration) at certain concentrations of N and P. Strong, positive relationships between the maximum and minimum with CV(Q) were also observed in the fall trough and spring trough-fall peak regimes, with both DSi regimes exhibiting similarly threshold responses with step increases around a CV (Q) value of 2.5. Additionally, the minimum and maximum DSi values for fall peak, spring trough-fall peak, and spring trough-variable summer regimes increased with the proportion of evergreen needleleaf forest. Precipitation was the only variable for which a negative partial dependence was observed, specifically in the spring trough-fall peak regime.
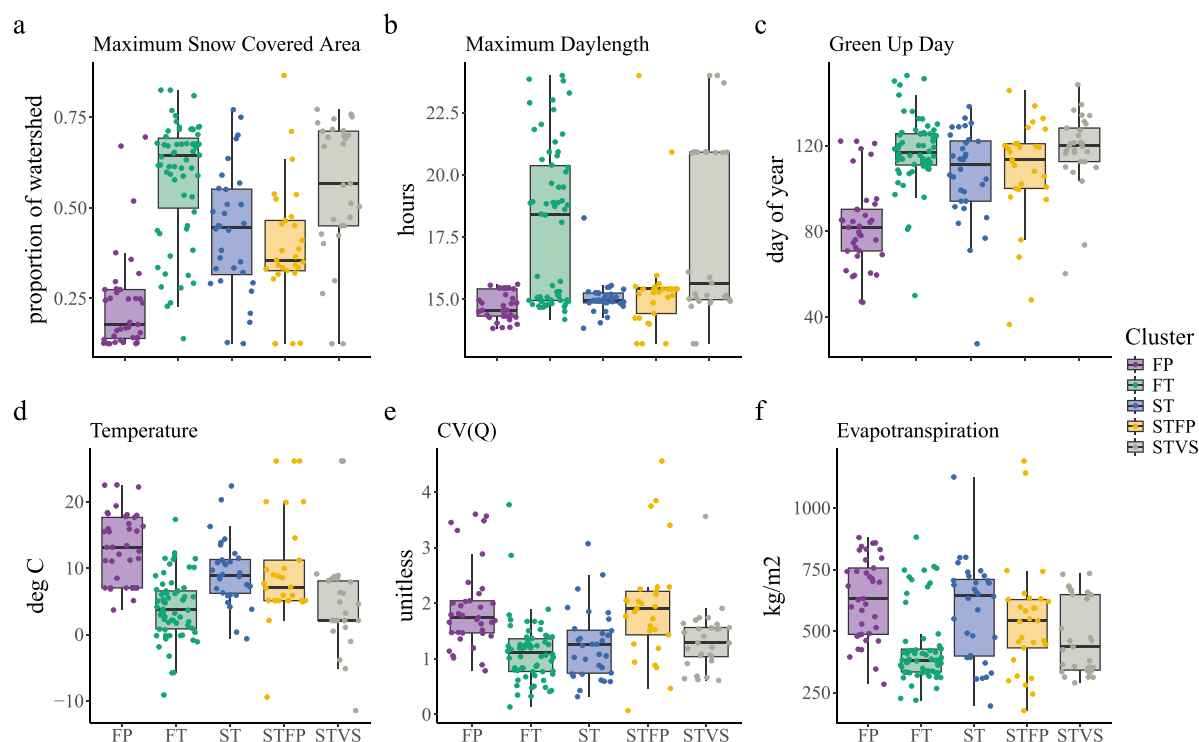
Figure 4. Distribution of the six most important variables for distinguishing among DSi regimes in the Average regime prediction model. The DSi regimes include: fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS). Each panels (a–f) shows a different driver, with each boxplot showing the distribution of that driver for a given DSi regime. Lower and upper lines of the boxplot box are quartile 1 and 3, respectively. The middle line is the median. Vertical lines indicate minimum and maximum, if less than ±1.5*interquartile range (IQR). Points outside ±1.5*IQR are considered outliers and are plotted above/below vertical lines. Colored dots show all values associated with the driver for each given DSi regime. The CV(Q) indicates the coefficient of variation for river discharge.

4. Discussion

Many stream ecosystems have strong and consistent seasonal patterns in their hydrological, chemical, and biological characteristics (Bard et al., 2015; Bernhardt et al., 2022; Bolotin et al., 2022; Crossman et al., 2016). We found that this is mostly true for DSi concentrations, but also that there is substantial variation in the nature of those seasonal signals. We also found variations in both the accuracy and the drivers of the Annual DSi model compared to the Average DSi model. The Annual DSi regime model (80% prediction accuracy) showed temperature, ET and flow were among the most important variables, while the Average model (59% prediction accuracy) indicated climate and primary productivity variables (i.e., maximum snow-covered area and maximum daylength) were key in predicting DSi regimes. Finally, predictions of DSi minimum and maximum that control the shape of DSi regime, were most sensitive to median nitrogen (N) and phosphorus (P) concentrations. This suggests the availability of these nutrients in rivers plays a key role in driving the seasonal variation in stream DSi concentrations. These findings indicate that accounting for spatial and temporal variability in the seasonality of stream biogeochemical signals and understanding the drivers of that variability will be important in better understanding how stream DSi availability will respond to environmental change.

4.1. Climate Variables Were Most Important in Distinguishing Among DSi Regimes, With Fall Peak and Fall Trough Regimes Predicted Most Accurately

In both the Average and Annual regime classification models, the fall peak and fall trough regimes were best predicted compared to DSi regimes exhibiting spring trough behavior (e.g., spring trough; spring trough-fall peak; and spring trough-variable summer). The fall peak and fall trough regimes also exhibited the highest stability over time (Figure 5, Johnson, Jankowski, Carey, Lyon, et al., 2024) and showed the most differentiation in the variables driving their behavior. In contrast, all three spring trough regimes were less stable and less well
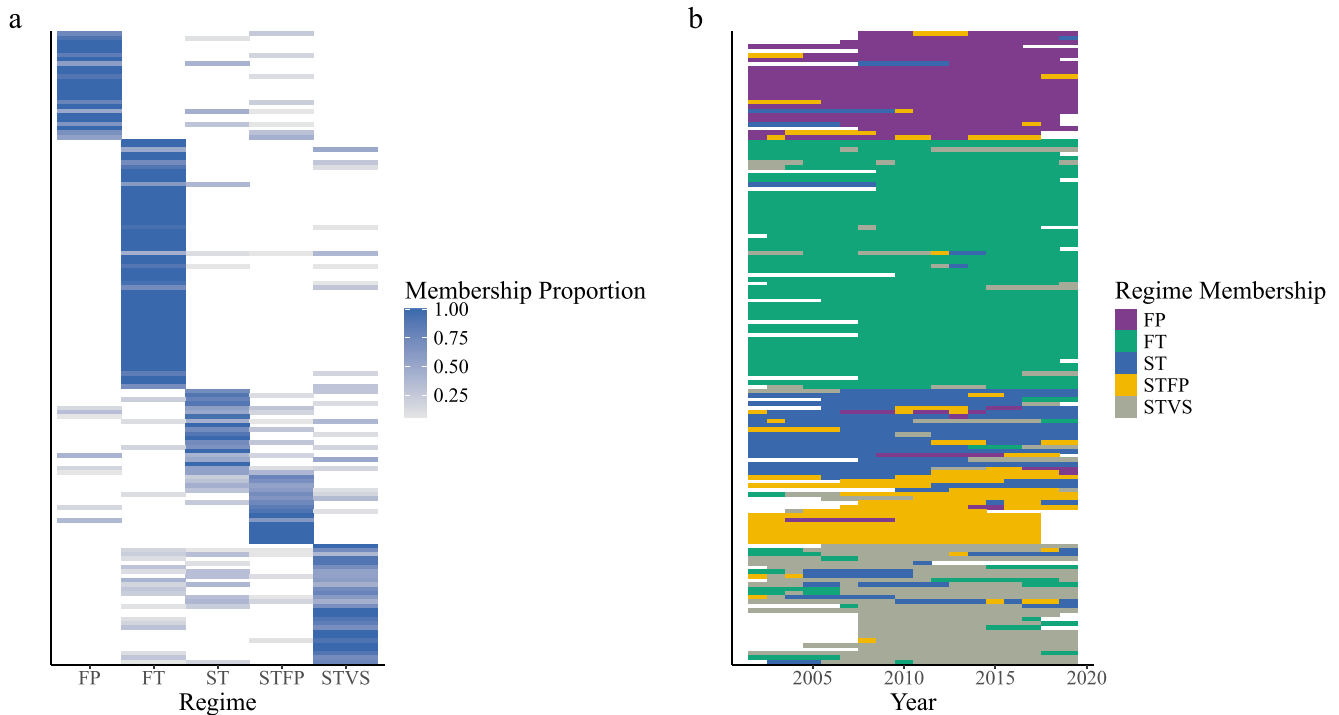
Figure 5. (a) Proportional membership of sites used in Annual regime analysis over the 18 years of record. Proportional membership was calculated as the number of years spent in each DSi regime (regimes include fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS)) divided by the number of years in the period of record between 2002 and 2019. Darker blue indicates a higher proportion of time spent in a given regime, where a value of 1 indicates 100% of the time the site was in the same regime. Sites are ordered on the *y*-axis alphabetically within Average regime membership groupings. (b) DSi regime membership of sites used in temporally variable regime analysis over the 18 years of record. Each column represents a year between 2002 and 2019 and each row represents a site. Sites are ordered on the *y*-axis alphabetically within Average regime membership groupings. The fill color of each grid cell (site-year combination) represents the DSi regime membership for that given site-year combination. Where fill color is white, it indicates that data do not exist for that site for that year.

differentiated across driver variables (Figures 4 and 5). Lower stability in spring trough regimes may reflect variability or shifts in winter and spring processes that drive DSi drawdown early in the calendar year. Drawdowns in stream DSi concentrations have been linked to the timing of processes that are sensitive to climate change and often exhibit a high degree of interannual variability, including dilution due to inputs of snow melt (Pulliainen, 2024; Vorkauf et al., 2021; Zheng et al., 2022), onset of spring and growing season length (Menzel et al., 2006; Monahan et al., 2016), and primary productivity, including DSi uptake by terrestrial plants and diatoms (Carey & Fulweiler, 2013; Fulweiler & Nixon, 2005; Sommer & Lengfellner, 2008). Lower stability in the spring trough regimes means that their average behavior is not consistently an accurate representation of their seasonal dynamics, which may make it harder to predict the timing of DSi delivery to downstream ecosystems.

Climate and productivity variables, including maximum proportion of snow-covered area, maximum daylength, green-up day, and temperature, emerged as most important in distinguishing among Average regime behavior. This indicates that DSi seasonality regimes are more strongly controlled by climate and ecosystem productivity than more static watershed characteristics such as lithology and land cover. Ecosystem productivity variables retained in the model, including daylength and green up day, influence photosynthesis and nutrient acquisition (including DSi) by terrestrial and aquatic primary producers thus DSi uptake on land and in the water. For example, fall peak regimes had earlier green-up timing than fall trough regimes, indicating that an earlier and longer period of uptake of DSi by primary producers may have contributed to lower concentrations earlier in the year in fall peak rivers. The relative difference in variable importance between static and dynamic (e.g., climate, productivity, hydrology) variables in the Average regime model is consistent with other large synthesis efforts of biogeochemical processes (Li et al., 2022; Phillips, 2020; White & Blum, 1995). This indicates that climate and productivity variables are more influential than static or slowly changing watershed characteristics (i.e., lithology,
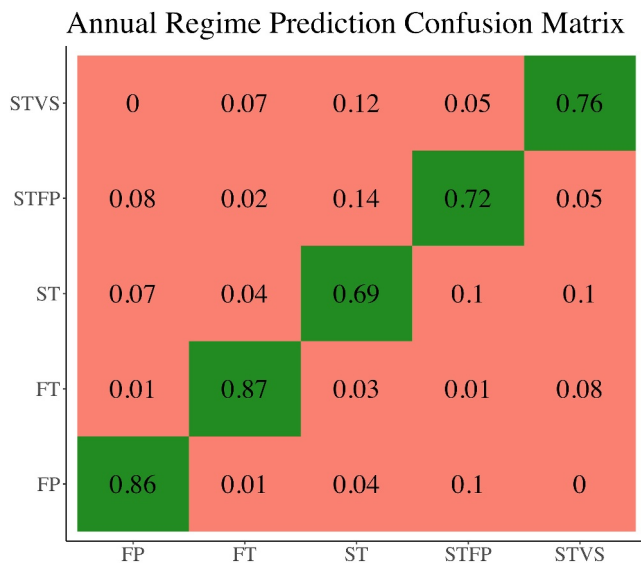
Annual Regime Prediction Confusion Matrix

Figure 6. Confusion matrix showing the classification accuracy of the Annual regime model. Across a given row, each box indicates the proportion of time the sites within a given DSi regime were classified into each other DSi regime, with green boxes indicating the proportion of time sites were correctly classified and pink boxes indicating the proportion of time sites were inaccurately classified into each other DSi regime. For example, in the fall peak (FP) regime (bottom row), sites were correctly classified 86% of the time, and inaccurately classified into the fall trough (FT) regime 1% of the time, into the spring trough (ST) regime 4% of the time and into the spring trough-fall peak (STFP) regime 10% of the time. Rows sum to 1. Regimes are fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS).

land cover) on seasonal DSi behavior. These static watershed characteristics may also mediate the degree to which climate and productivity affect DSi seasonality, but we did not evaluate these types of interactions in our model.

Compared to the Average regime model, different drivers emerged as important in the Annual regime model, including low and high discharge, while the importance of temperature and evapotranspiration became even more pronounced. We interpret drivers that emerge as important in the Annual regime model as not only capturing differences in environmental variables among DSi regimes but also accounting for how the interannual variability in those variables is related to shifts in DSi seasonal behavior over time. This suggests that while average distributions of variables such as maximum snow-covered area, daylength, and green-up day may differ among DSi regimes, shifts in DSi regime membership across years are driven mostly by differences in temperature, evapotranspiration, and discharge over time. Given shifts in the seasonal patterns of temperature, evapotranspiration, and hydrology due to climate change (Arnell et al., 2019; Bard et al., 2015; Li & Thompson, 2021; Mastrotheodoros et al., 2020; Schnorbus et al., 2014; Yang & Olivera, 2023; Yu et al., 2018), these results suggest that as warming continues, patterns in seasonal DSi availability will shift concomitantly with changes in hydroclimatic conditions.

4.2. Interannual Shifts in Annual DSi Regimes Suggest Changes in DSi Transport and Uptake

When we allowed DSi regimes to vary over time, Annual DSi regime membership classification accuracy increased by over 20% compared to Average DSi regime membership classification (Figures 2 and 6). This increase in predictive power indicates that there is substantial variability in Annual DSi regime membership. Movement among DSi regimes over time was not observed equally across DSi regimes (Figure S2 in Supporting Information S1). Spring trough regimes exhibited lower proportional membership in their modal regime and more shifting among other DSi regimes relative to the fall peak and fall trough regimes (Figure 5a). Additionally, the sites in the spring trough regimes showed membership across a broader range of DSi regimes compared to sites in the fall peak and fall trough regimes (Figure 5b), consistent with Johnson, Jankowski, Carey, Lyon, et al. (2024) who showed lower DSi regime membership stability in spring trough regimes. We found that there were patterns in the nature of the observed shifts among DSi regimes (Figure 5; Figure S3 in Supporting Information S1). For the fall peak regime, most sites shifted to the spring trough-fall peak regime over their period of record, while in the fall trough regime, most sites shifted to the spring trough-variable summer regime. However, the spring trough regimes showed much more variability in their regime shifting behavior. Both the spring trough and spring trough-fall peak regimes show shifts between all four other DSi regimes, whereas the spring trough-variable summer regime shows shifts between all DSi regimes except the fall peak. Taken together, shifts among DSi regimes represent three dominant patterns: (a) increases in fall DSi concentrations; (b) shifts and elongation of the timing of drawdown; and (c) dampening of the spring trough.

Increases in fall DSi concentrations, as seen in the shifts from the spring trough regime to the fall peak and spring trough-fall peak regimes, may be interpreted as increased proportion of groundwater contributions in streamflow as snowpacks decline (Huntington & Niswonger, 2012) or summer dry-period base flows that persist longer into the fall (Demaria et al., 2016; Jefferson et al., 2008). In streams that exhibit diluting DSi concentration-discharge behavior, proportional increases in groundwater contribution may lead to higher in-stream DSi concentrations during low flow. Increases in groundwater contributions to streamflow in low-snow and low-flow years are well documented, and are expected to increase with climate-induced declines in snowpack and extended drought periods (Johnson, Harpold, et al., 2023; Segura et al., 2019).
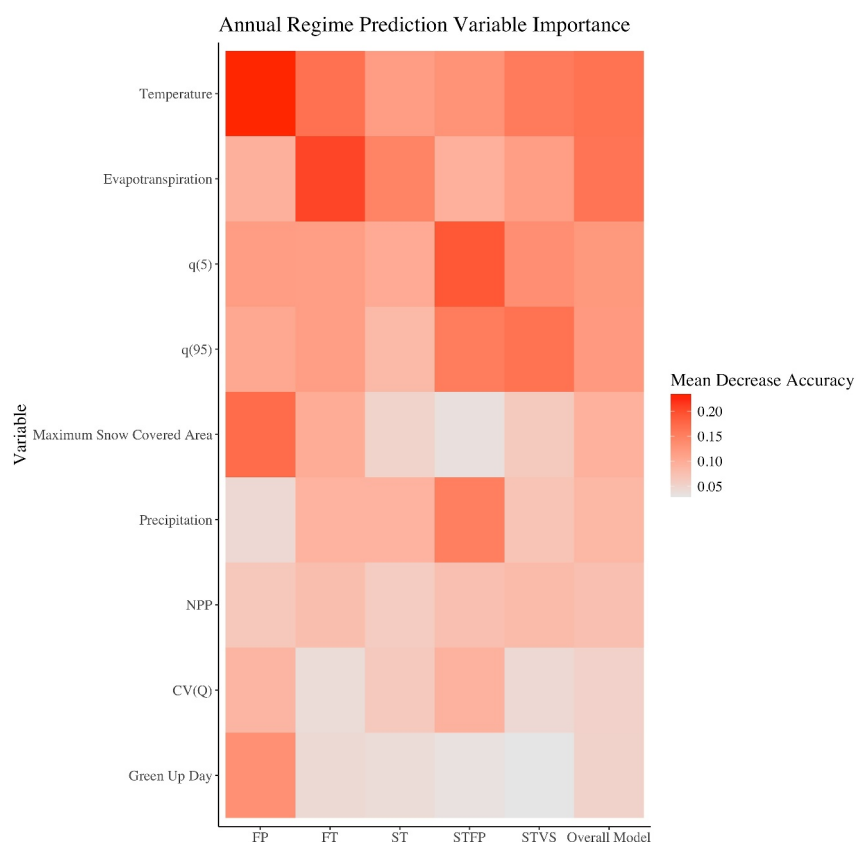
Figure 7. Variable importance, indicated by mean decrease in accuracy, for the Annual regime classification model. The *y*-axis shows variables retained in the model, ordered from most-to-least important for the overall regime classification. Individual regimes include: fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS). The "overall model" column (far right) indicates the importance of each variable in classifying regime membership across all five DSi regimes, whereas the columns labeled with regime acronyms (e.g., FP) indicate variable importance in classifying sites within a given DSi regime. For example, a mean decrease in accuracy (dark red) of 0.2 in the overall model column indicates that when a given variable was removed, the overall model performance declined by 20%. In contrast, when a variable shows a high mean decrease of 0.15 for a given regime (e.g., Maximum Snow-Covered Area for fall peak), it indicates that if that variable were to be removed the out-of-bag error for that regime would increase by 15% (e.g., sites in that regime would be more frequently misclassified).

Shifts in DSi drawdown timing, as well as elongation of DSi drawdown, may also be indicative of warming. This hypothesis is evidenced by shifts from the fall trough regime to spring trough-variable summer regime and the spring trough regime to spring trough-variable summer regime. Shifts in the timing of drawdown earlier in the year may indicate earlier snow and ice melt in high latitude and elevation sites leading to an earlier dilution signal in the stream (Carey et al., 2020; Vorkauf et al., 2021; Zheng et al., 2022). Additionally, warmer spring and summer temperatures may drive earlier uptake by aquatic primary producers (Monahan et al., 2016; Sommer & Lengfellner, 2008) and terrestrial vegetation (Carey & Fulweiler, 2013; Fulweiler & Nixon, 2005). Similarly, warmer, and longer, growing seasons may lead to an increase in biological productivity on land and in-stream and later autumn senescence, resulting in longer periods of DSi drawdown (Kienel et al., 2017; Michelutti et al., 2003; Wasmund et al., 2019).

Lastly, we observed a dampening of the spring trough at some sites, evidenced by shifts from the spring trough-fall peak regime to the fall peak regime and the spring trough to the fall peak regime. These shifts may be explained by two processes: (a) a reduction in the density of diatom communities present and/or shifting species assemblage from siliceous (e.g., diatom) toward non-siliceous species leading to less DSi uptake (Gobler et al., 2017; Schlüter et al., 2012; Visser et al., 2016); or (b) increased mobilization of biogenic DSi stored in the shallow subsurface (Mills et al., 2017; Scanlon et al., 2001), associated with shifts in infiltration and
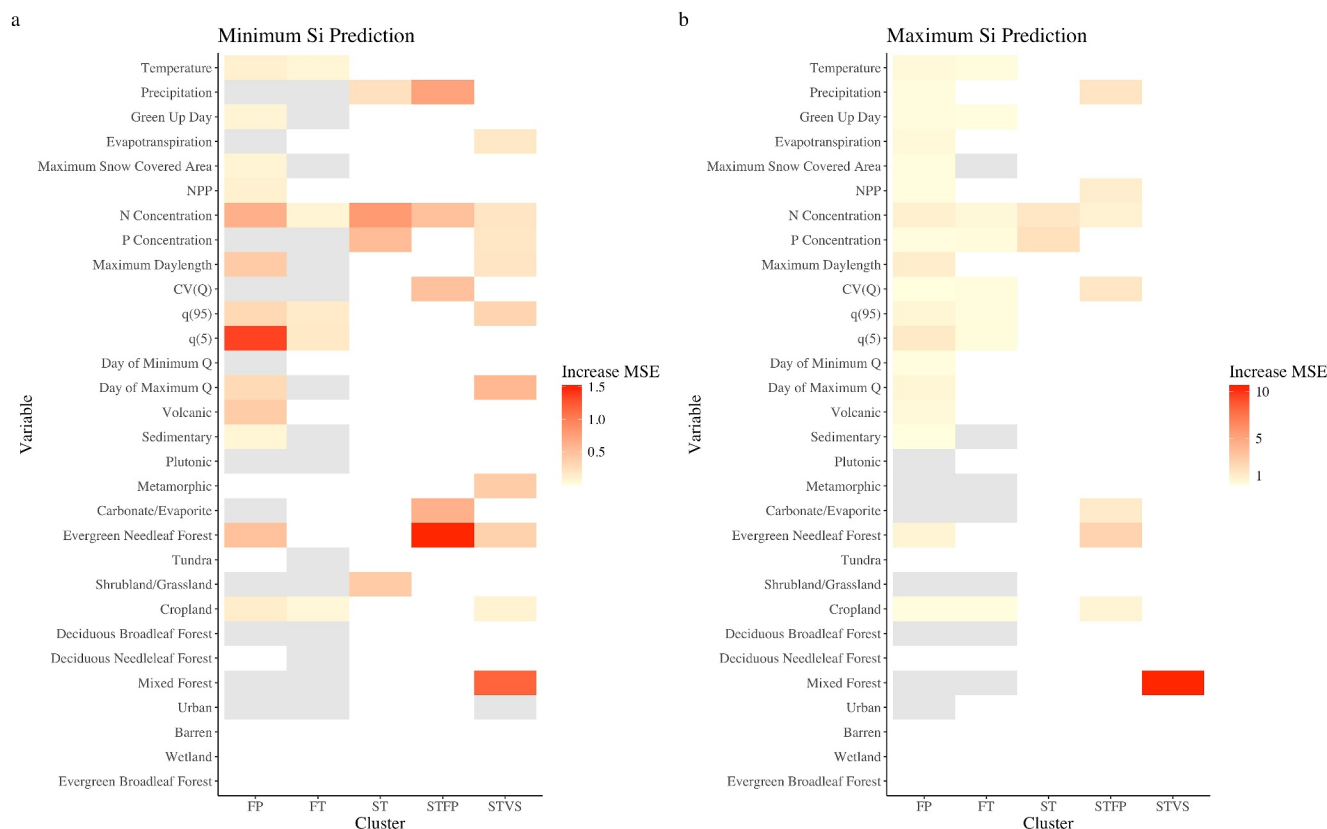
Figure 8. Variable importance, indicated by mean decrease in accuracy, for the (a) minimum DSi concentration prediction model and (b) maximum DSi concentration prediction model. The *y*-axis shows variables retained in the model (same variables on both panels a, b). Each column of the *x*-axis indicates the variables retained in the model (colored box present) for the distinct DSi regimes: fall peak (FP), fall trough (FT), spring trough (ST), spring trough-fall peak (STFP), and spring trough-variable summer (STVS). The absence of a colored box indicates that the variable was not retained in the model (e.g., in panel (b) within the ST (middle) column, N concentration and P concentration were the only variables retained). The color of the box indicates the importance of each variable in prediction of DSi minimum or maximum concentration given by the increase in mean squared error (MSE). A high importance indicates if that variable was removed, the MSE of the model would increase. If variables were retained but had an increase in MSE value less than 0.05, they are filled gray. Note that color scales in panels (a, b) are different.

flowpath activation driven by changes in precipitation phase and snowmelt rate and timing (Konapala et al., 2020; Musselman et al., 2017).

Observed shifts among DSi regimes align well with expected warming-induced changes in stream DSi concentrations. Climate-induced shifting of DSi regime is further supported by the fact that temperature and hydrology variables emerged as most important in the Annual regime membership model, both of which are rapidly changing associated with climate change (IPCC, 2023; Koutsoyiannis, 2020; Wu et al., 2022). Temperature and hydrology exert strong controls on the mobilization and export of DSi and on the density, composition, and productivity of algal communities (Bernhardt et al., 2022; Bondar-Kunze et al., 2021; Carey et al., 2020; Schneider, 2015; Struyf & Conley, 2012), and thus can be expected to alter DSi regime. However, it is not well understood whether shifts in DSi regime are associate with gradual changes in climate or extreme events (e.g., droughts or floods), but elucidating these driving mechanisms can help to characterize the resilience of ecosystems to ongoing change.

4.3. N and P Are Dominant Controls on DSi Minimum and Maximum Behavior Across DSi Regime Membership

We found evidence of strong relationships between N, P, and DSi concentrations across the rivers included in this study. This is likely because these elements often function together to support primary production in both aquatic and terrestrial environments (Abbott et al., 2018; Carey & Fulweiler, 2012; Dodds & Smith, 2016; Hill et al., 2012; LeBauer & Treseder, 2008). Median N and P concentrations were retained in both minimum and
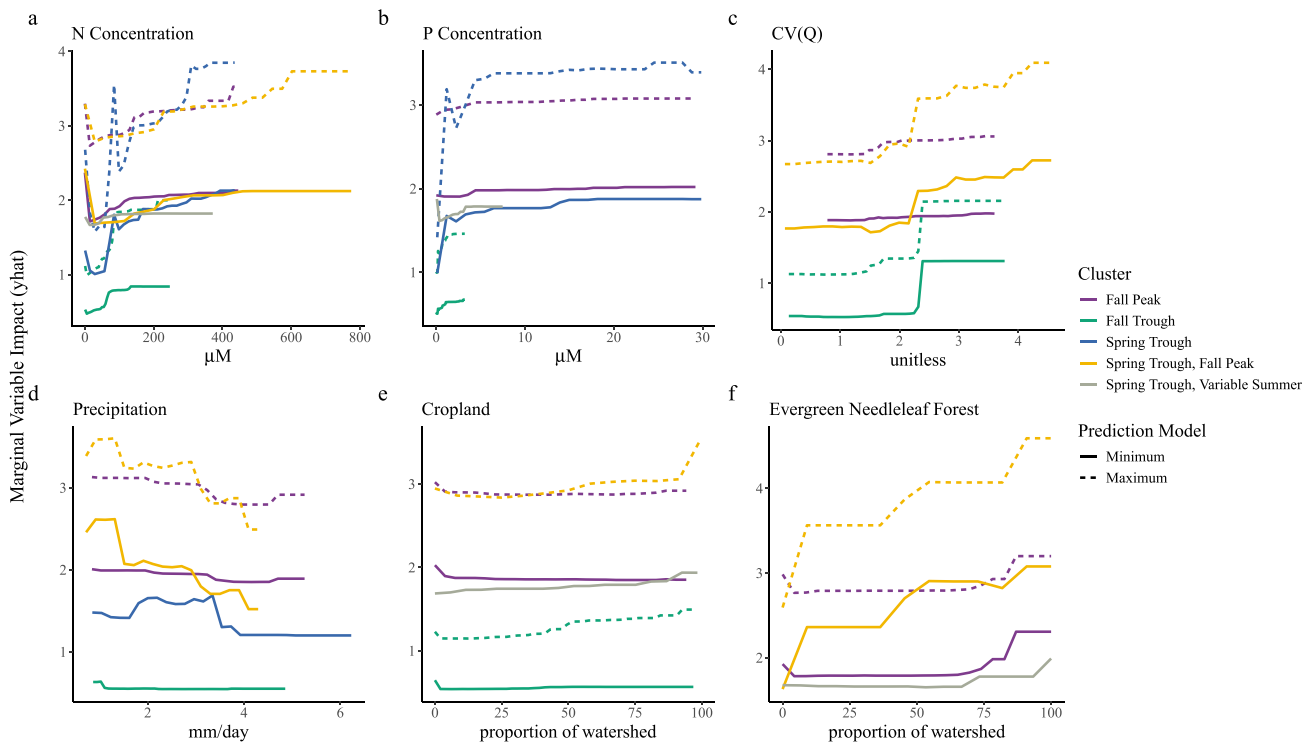
Figure 9. Partial dependence plots showing the relationship of the six most important variables for minimum (solid lines) and maximum (dashed lines) DSi concentration across the five DSi regimes (colored lines). Lines are only present in a given panel when that variable was retained in the model (see Figure 8). The x-axis (response) shows the actual values of each variable and the y-axis (yhat) shows the marginal impact of the variable on DSi minimum or maximum concentration. Marginal impact means that the effect of all other variables have been reduced such that yhat shows the response of DSi minimum/maximum concentration to a given variable independent of other variables included in the model. For example, the impact of N concentration on DSi minimum concentration in the spring trough regime (blue solid line) is initially negative (i.e., an increase in median N concentration is associated with a decrease in DSi concentration) before becoming positive (i.e., an increase in N concentration is associated with an increase in DSi concentration) over the rest of the response range. The difference in the extent of the lines within a given panel is due to differences in the range in the response variable between DSi regimes (e.g., median N concentration in the fall trough regime ranges from 0 to just over 200 while median N concentration in the spring trough-fall peak regime ranges from 0 to nearly 800 μM).

maximum DSi prediction models across most DSi regimes, with the exception of the spring trough-fall peak regime that only retained N in both models and the spring trough-variable summer regime that did not retain N or P in the maximum Si prediction model (Figure 8). Median N and P concentrations were the most important in predicting spring trough regime DSi minimum and maximum concentrations compared to other models (Figure 8). Across all regimes, the relationships between both minimum and maximum DSi concentrations with N and P were positive overall, although non-linearity and threshold behavior was observed in these relationships. At low concentrations of N and P, DSi concentrations often experienced a brief drawdown as N and P concentrations initially rose (Figure 9), indicating the importance of biological production in controlling DSi dynamics as the limiting nutrients N and P initially increase from very low levels. These relationships indicate that, above a certain concentration of N or P, DSi was independent of their concentration, suggesting that algal species assemblages may have shifted away from diatoms to other species. However, more eutrophic rivers generally exhibited higher DSi concentrations across DSi regimes, therefore small changes in DSi may be less notable relative to these higher concentrations.

Stream ecosystems can experience limitation of N and P from both instream and terrestrial processes. We interpret the emergence of median N and P concentrations as important in our models as indication that N and/or P availability controls seasonal variation in DSi concentrations. Specifically, high importance of P concentrations may be indicative of DSi uptake by instream processes, and high importance of N concentrations may be indicative of DSi uptake by terrestrial processes (Dodds & Smith, 2016; LeBauer & Treseder, 2008), although both terrestrial and aquatic systems can experience co-limitation by N and P (Elser et al., 2007; Harpole et al., 2011). For example, the high importance of P availability in the spring trough DSi minimum concentration

model suggests for in-stream diatom production is driving springtime DSi drawdown (Admiraal et al., 1990; Garnier et al., 1995; Wall et al., 1998). However, we do not see a strong linkage between P and DSi minimum concentrations in other regimes that exhibit a spring drawdown (e.g., spring trough-fall peak, spring trough-variable summer). There are several potential reasons for the lower importance of P concentration relative to N concentration on DSi minimums in other DSi regimes: (a) diatom uptake may be masked by other processes that control DSi minimum, such as high flows that flush DSi from the hyporheic zone or shifts in light availability (Hester et al., 2017; Hill et al., 2009); (b) the relationships between DSi with P concentrations may provide only limited insight into diatom uptake when P concentrations are very low (Trentman et al., 2021); or (c) processes other than in-stream diatom production may control in-stream DSi minimum concentrations.

The importance of terrestrial biological processing in controlling in-stream DSi dynamics is an outstanding research question. In our work, N concentrations emerge as more important in predicting DSi concentrations relative to P concentrations in the fall peak and the spring trough-fall peak regimes, which supports the hypothesis that terrestrial vegetation DSi uptake exerts influence on seasonal DSi dynamics in rivers (Carey & Fulweiler, 2013; Conley et al., 2008; Cornelis et al., 2010; Fulweiler & Nixon, 2005; Phillips, 2020). This hypothesis is further supported by the high importance of maximum daylength and green-up day as drivers in the Average regime model.

It is important to underscore the limitations in our modeling approach and the explanatory features included in the model. In our models, we use only median N and P concentrations across a given sites' period of record; given the dynamic nature of N and P in aquatic systems, future work would benefit from incorporating temporal dynamics when evaluating DSi behavior relative to N and P. In addition, there are several reasonable proxies for instream diatom uptake of N, P, and DSi such as seasonal channel width to flow ratios (proxy for light availability), leaf on and off timing, riparian area to stream width (proxy for hyporheic zone interaction), and stream velocity (i.e., scouring), which could be included in future work on DSi behavior (Francoeur & Biggs, 2006; Hill & Dimick, 2002; Julian et al., 2008).

4.4. Hydrologic and Land Cover Controls on Seasonal DSi Regimes

Changes in land cover and the hydrologic cycle are occurring at rapid rates across the globe (Koutsoyiannis, 2020; Song et al., 2018; Winkler et al., 2021; Wu et al., 2022). The emergence of land use and hydrologic variables across most minimum and maximum DSi models indicate that instream DSi cycling is sensitive to hydrologic and land use change. Specifically, CV(Q) emerged as important in predicting the minimum of the spring trough-fall peak regime and q(5) emerged as important in the predicting the minimum of the fall peak regime, suggesting that streamflow exerts stronger controls on DSi minimum concentrations than N and P in systems where biologic uptake is limited (Abbott et al., 2018; Moatar et al., 2017). Additionally, some land cover variables emerged as important, including evergreen needleleaf forest in the spring trough-fall peak regime and mixed forest in the spring trough-variable summer regime. Furthermore, cropland was retained in all models except the spring trough models. Both cropland and evergreen needleleaf forest show positive associations (i.e., increasing proportion of a given land cover leads to higher DSi concentrations) with both DSi minimum and maximum concentrations. This finding supports prior work demonstrating that land cover is one of the most important factors in DSi mobilization in temperate systems due to uptake of large quantities of Si by terrestrial vegetation, thus altering rates of mobilization and dissolution of Si from terrestrial to aquatic systems (Carey & Fulweiler, 2012; Chen et al., 2014; Conley et al., 2008; Struyf et al., 2010). The strong importance of green-up day as a driver of the Average regime model supplies further evidence that terrestrial vegetation uptake has a strong influence on stream DSi availability, almost regardless of the specific type of land cover. Given widespread changes in hydrology and land cover, we may expect to observe continued shifts in the minimum and maximum concentrations of fluvial DSi.

5. Conclusion

Understanding the processes that control variable seasonal regimes of DSi can help to predict how DSi availability will change with anthropogenic perturbations and increased occurrences of extreme events (Ingram, 2016; Satoh et al., 2022; Tabari, 2020). In this paper, we build upon prior work that identified five distinct seasonal patterns across 200 rivers in the Northern Hemisphere, and find that Average regime membership was controlled by climate and productivity variables, whereas minimum and maximum DSi concentrations were correlated with median N and P concentrations. Additionally, accounting for interannual variability in DSi regime membership

improved model classification by over 20%, with temperature, evapotranspiration, and discharge indicated as driving shifts in Annual DSi regime membership. These findings emphasize the importance in accounting for interannual variability in the timing and magnitude of nutrient export to downstream ecosystems, especially in systems that are, and continue to be, impacted by climate change.

DSi availability in streams directly supports algal communities that contribute to carbon sequestration and provide a high-quality food source for higher trophic levels (Brett et al., 2000; Street-Perrott & Barker, 2008; Wang et al., 2016). Shifts in the timing of DSi availability across the seasonal cycle likely impact the timing of DSi delivery to downstream ecosystems, including marine environments. This may result in shifts in algal community composition if the delivery of DSi availability is decoupled from N and P dynamics. Future work could quantify how stream DSi dynamics are shifting over time, how DSi regimes are shifting relative to other nutrient regimes, and evaluate of the implications of these shifts for downstream DSi load delivery. Additionally, future work would benefit from expanding study sites to include more representation from tropical systems, which transport nearly 75% of DSi from rivers to oceans (Tréguer et al., 1995, 2021).

Data Availability Statement

Discharge and DSi data are published in Johnson, Jankowski, et al. (2023). Watershed characteristics and nutrient data are published in Johnson, Jankowski, Carey, Sethna, et al. (2024). Scripts for random forest modeling are published in Johnson (2024). Scripts for spatial data extraction are published in Lyon et al. (2024).

References

Abbott, B. W., Gruau, G., Zarnetske, J. P., Moatar, F., Barbe, L., Thomas, Z., et al. (2018). Unexpected spatial stability of water chemistry in headwater stream networks. *Ecology Letters*, 21(2), 296–308. https://doi.org/10.1111/ele.12897

Admiraal, W., Breugem, P., Jacobs, D. M. L. H. A., & De Ruyter Van Steveninck, E. D. (1990). Fixation of dissolved silicate and sedimentation of biogenic silicate in the lower river Rhine during diatom blooms. *Biogeochemistry*, 9(2), 175–185. https://doi.org/10.1007/BF00692170

Anderson, J. R., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data (Report 964; Professional Paper). https://doi.org/10.3133/pp964

Arnell, N. W., Lowe, J. A., Challinor, A. J., & Osborn, T. J. (2019). Global and regional impacts of climate change at different levels of global temperature increase. *Climatic Change*, 155(3), 377–391. https://doi.org/10.1007/s10584-019-02464-z

Bard, A., Renard, B., Lang, M., Giuntoli, I., Korck, J., Koboltschnig, G., et al. (2015). Trends in the hydrologic regime of Alpine rivers. *Journal of Hydrology*, 529, 1823–1837. https://doi.org/10.1016/j.jhydrol.2015.07.052

Berhanu, B., Seleshi, Y., Demisse, S. S., & Melesse, A. M. (2015). Flow regime classification and hydrological characterization: A case study of Ethiopian rivers. *Water*, 7(6), 3149–3165. https://doi.org/10.3390/w7063149

Bernhardt, E. S., Savoy, P., Vlah, M. J., Appling, A. P., Koenig, L. E., Hall, R. O., et al. (2022). Light and flow regimes regulate the metabolism of rivers. *Proceedings of the National Academy of Sciences*, 119(8), e2121976119. https://doi.org/10.1073/pnas.2121976119

Bolotin, L. A., Summers, B. M., Savoy, P., & Blaszczak, J. R. (2022). Classifying freshwater salinity regimes in central and western U.S. streams and rivers. *Limnology and Oceanography Letters*, 8(1), 103–111. https://doi.org/10.1002/lol2.10251

Bondar-Kunze, E., Kasper, V., & Hein, T. (2021). Responses of periphyton communities to abrupt changes in water temperature and velocity, and the relevance of morphology: A mesocosm approach. *The Science of the Total Environment*, 768, 145200. https://doi.org/10.1016/j.scitotenv.2021.145200

Bouwman, A. F., Bierkens, M. F. P., Griffioen, J., Hefting, M. M., Middelburg, J. J., Middelkoop, H., & Slomp, C. P. (2013). Nutrient dynamics, transfer and retention along the aquatic continuum from land to ocean: Towards integration of ecological and biogeochemical models. *Biogeosciences*, 10(1), 1–22. https://doi.org/10.5194/bg-10-1-2013

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brett, M. T., Müller-Navarra, D. C., & Sang-Kyu, P. (2000). Empirical analysis of the effect of phosphorus limitation on algal food quality for freshwater zooplankton. *Limnology & Oceanography*, 45(7), 1564–1575. https://doi.org/10.4319/lo.2000.45.7.1564

Cammarota, C., & Pinto, A. (2021). Variable selection and importance in presence of high collinearity: An application to the prediction of lean body mass from multi-frequency bioelectrical impedance. *Journal of Applied Statistics*, 48(9), 1644–1658. https://doi.org/10.1080/02664763.2020.1763930

Carey, J. C., & Fulweiler, R. W. (2012). The terrestrial silica pump. *PLoS One*, 7(12), e52932. https://doi.org/10.1371/journal.pone.0052932

Carey, J. C., & Fulweiler, R. W. (2013). Watershed land use alters riverine silica cycling. *Biogeochemistry*, 113(1), 525–544. https://doi.org/10.1007/s10533-012-9784-2

Carey, J. C., Gewirtzman, J., Johnston, S. E., Kurtz, A., Tang, J., Vieillard, A. M., & Spencer, R. G. M. (2020). Arctic River dissolved and biogenic silicon exports—Current conditions and future changes with warming. *Global Biogeochemical Cycles*, 34(3), e2019GB006308. https://doi.org/10.1029/2019GB006308

Casey, H., Clarke, R. T., & Marker, A. F. H. (1981). The seasonal variation in silicon concentration in chalk-streams in relation to diatom growth. *Freshwater Biology*, 11(4), 335–344. https://doi.org/10.1111/j.1365-2427.1981.tb01265.x

Chen, N., Wu, Y., Wu, J., Yan, X., & Hong, H. (2014). Natural and human influences on dissolved silica export from watershed to coast in Southeast China. *Journal of Geophysical Research: Biogeosciences*, 119(1), 95–109. https://doi.org/10.1002/2013JG002429

Conley, D., Schelske, C., & Stoermer, E. (1993). Modification of the biogeochemical cycle of silica with eutrophication. *Marine Ecology Progress Series*, 101, 179–192. https://doi.org/10.3354/meps101179

Conley, D. J. (2002). Terrestrial ecosystems and the global biogeochemical silica cycle. *Global Biogeochemical Cycles*, 16(4), 68-1–68-8. https://doi.org/10.1029/2002GB001894

Conley, D. J., Likens, G. E., Buso, D. C., Saccone, L., Bailey, S. W., & Johnson, C. E. (2008). Deforestation causes increased dissolved silicate losses in the Hubbard Brook Experimental Forest. *Global Change Biology*, *14*(11), 2548–2554. https://doi.org/10.1111/j.1365-2486.2008.01667.x

Cornelis, J.-T., Delvaux, B., Georg, R. B., Lucas, Y., Ranger, J., & Opfergelt, S. (2011). Tracing the origin of dissolved silicon transferred from various soil-plant systems towards rivers: A review. *Biogeosciences*, *8*(1), 89–112. https://doi.org/10.5194/bg-8-89-2011

Cornelis, J.-T., Ranger, J., Iserentant, A., & Delvaux, B. (2010). Tree species impact the terrestrial cycle of silicon through various uptakes. *Biogeochemistry*, *97*(2/3), 231–245. https://doi.org/10.1007/s10533-009-9369-x

Crossman, J., Catherine Eimers, M., Casson, N. J., Burns, D. A., Campbell, J. L., Likens, G. E., et al. (2016). Regional meteorological drivers and long term trends of winter-spring nitrate dynamics across watersheds in northeastern North America. *Biogeochemistry*, *130*(3), 247–265. https://doi.org/10.1007/s10533-016-0255-z

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

Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genetics*, *19*(1), 65. https://doi.org/10.1186/s12863-018-0633-8

Das, P., Sachindra, D. A., & Chanda, K. (2022). Machine learning-based rainfall forecasting with multiple non-linear feature selection algorithms. *Water Resources Management*, *36*(15), 6043–6071. https://doi.org/10.1007/s11269-022-03341-8

Demaria, E. M. C., Palmer, R. N., & Roundy, J. K. (2016). Regional climate change projections of streamflow characteristics in the Northeast and Midwest U.S. *Journal of Hydrology: Regional Studies*, *5*, 309–323. https://doi.org/10.1016/j.ejrh.2015.11.007

Dodds, W. K., & Smith, V. H. (2016). Nitrogen, phosphorus, and eutrophication in streams. *Inland Waters*, *6*(2), 155–164. https://doi.org/10.5268/IW-6.2.909

dos Reis Oliveira, P. C., van der Geest, H. G., Kraak, M. H. S., & Verdonschot, P. F. M. (2019). Land use affects lowland stream ecosystems through dissolved oxygen regimes. *Scientific Reports*, *9*(1), 19685. https://doi.org/10.1038/s41598-019-56046-1

Drever, J. I. (1994). The effect of land plants on weathering rates of silicate minerals. *Geochimica et Cosmochimica Acta*, *58*(10), 2325–2332. https://doi.org/10.1016/0016-7037(94)90013-2

Elser, J. J., Bracken, M. E. S., Cleland, E. E., Gruner, D. S., Harpole, W. S., Hillebrand, H., et al. (2007). Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*, *10*(12), 1135–1142. https://doi.org/10.1111/j.1461-0248.2007.01113.x

Francoeur, S. N., & Biggs, B. J. F. (2006). Short-term effects of elevated velocity and sediment abrasion on benthic algal communities. *Hydrobiologia*, *561*(1), 59–69. https://doi.org/10.1007/s10750-005-1604-4

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Fulweiler, R. W., & Nixon, S. W. (2005). Terrestrial vegetation and the seasonal cycleof dissolved silica in a southern New Englandcoastal river. *Biogeochemistry*, *74*(1), 115–130. https://doi.org/10.1007/s10533-004-2947-z

Gaillardet, J., Dupré, B., Louvat, P., & Allègre, C. J. (1999). Global silicate weathering and CO_2 consumption rates deduced from the chemistry of large rivers. *Chemical Geology*, *159*(1), 3–30. https://doi.org/10.1016/S0009-2541(99)00031-5

Garnier, J., Billen, G., & Coste, M. (1995). Seasonal succession of diatoms and Chlorophyceae in the drainage network of the Seine River: Observations and modeling. *Limnology & Oceanography*, *40*(4), 750–765. https://doi.org/10.4319/lo.1995.40.4.0750

Giesbrecht, K. E., & Varela, D. E. (2021). Summertime biogenic silica production and silicon limitation in the Pacific Arctic region from 2006 to 2016. *Global Biogeochemical Cycles*, *35*(1), e2020GB006629. https://doi.org/10.1029/2020GB006629

Gobler, C. J., Doherty, O. M., Hattenrath-Lehmann, T. K., Griffith, A. W., Kang, Y., & Litaker, R. W. (2017). Ocean warming since 1982 has expanded the niche of toxic algal blooms in the North Atlantic and North Pacific Oceans. *Proceedings of the National Academy of Sciences*, *114*(19), 4975–4980. https://doi.org/10.1073/pnas.1619575114

Godsey, S. E., Kirchner, J. W., & Clow, D. W. (2009). Concentration–discharge relationships reflect chemostatic characteristics of US catchments. *Hydrological Processes*, *23*(13), 1844–1864. https://doi.org/10.1002/hyp.7315

Greenwell, B. M. (2017). pdp: An R Package for constructing partial dependence plots. *The R Journal*, *9*(1), 421. https://doi.org/10.32614/RJ-2017-016

Harpole, W. S., Ngai, J. T., Cleland, E. E., Seabloom, E. W., Borer, E. T., Bracken, M. E. S., et al. (2011). Nutrient co-limitation of primary producer communities. *Ecology Letters*, *14*(9), 852–862. https://doi.org/10.1111/j.1461-0248.2011.01651.x

Harrison, J. W., Lucius, M. A., Farrell, J. L., Eichler, L. W., & Relyea, R. A. (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

Hester, E. T., Cardenas, M. B., Haggerty, R., & Apte, S. V. (2017). The importance and challenge of hyporheic mixing. *Water Resources Research*, *53*(5), 3565–3575. https://doi.org/10.1002/2016WR020005

Hill, B. H., Elonen, C. M., Seifert, L. R., May, A. A., & Tarquinio, E. (2012). Microbial enzyme stoichiometry and nutrient limitation in US streams and rivers. *Ecological Indicators*, *18*, 540–551. https://doi.org/10.1016/j.ecolind.2012.01.007

Hill, W. R., & Dimick, S. M. (2002). Effects of riparian leaf dynamics on periphyton photosynthesis and light utilisation efficiency. *Freshwater Biology*, *47*(7), 1245–1256. https://doi.org/10.1046/j.1365-2427.2002.00837.x

Hill, W. R., Fanta, S. E., & Roberts, B. J. (2009). Quantifying phosphorus and light effects in stream algae. *Limnology & Oceanography*, *54*(1), 368–380. https://doi.org/10.4319/lo.2009.54.1.0368

Hilley, G. E., & Porder, S. (2008). A framework for predicting global silicate weathering and CO2 drawdown rates over geologic time-scales. *Proceedings of the National Academy of Sciences*, *105*(44), 16855–16859. https://doi.org/10.1073/pnas.0801462105

Hirsch, R. M., Moyer, D., & Archfield, S. A. (2010). Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay River inputs. *Journal of the American Water Resources Association*, *46*(5)), 24–880. https://doi.org/10.1111/j.1752-1688.2010.00482.x

House, W. A., Leach, D. V., & Armitage, P. D. (2001). Study of dissolved silicon, and nitrate dynamics in a fresh water stream. *Water Research*, *35*(11), 2749–2757. https://doi.org/10.1016/S0043-1354(00)00548-0

Huntington, J. L., & Niswonger, R. G. (2012). Role of surface-water and groundwater interactions on projected summertime streamflow in snow dominated regions: An integrated modeling approach. *Water Resources Research*, *48*(11). https://doi.org/10.1029/2012WR012319

Ingram, W. (2016). Increases all round. *Nature Climate Change*, *6*(5), 443–444. https://doi.org/10.1038/nclimate2966

IPCC. (2023). *Climate Change 2023: Synthesis Report*. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team]. In H. Lee & J. Romero (Eds.), (pp. 35–115). IPCC. https://doi.org/10.59327/IPCC/AR6-9789291691647

Ishwaran, H., Kogalur, U. B., Chen, X., & Minn, A. J. (2011). Random survival forests for high-dimensional data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *4*(1), 115–132. https://doi.org/10.1002/sam.10103

Jefferson, A., Nolin, A., Lewis, S., & Tague, C. (2008). Hydrogeologic controls on streamflow sensitivity to climate variation. *Hydrological Processes*, *22*(22), 4371–4385. https://doi.org/10.1002/hyp.7041

Johnson, K. (2024). hydrokeira/Si_RandomForest: Si random forest analysis code (v1.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.12667698

Johnson, K., Harpold, A., Carroll, R. W. H., Barnard, H., Raleigh, M. S., Segura, C., et al. (2023). Leveraging groundwater dynamics to improve predictions of summer low-flow discharges. *Water Resources Research*, *59*(8), e2023WR035126. https://doi.org/10.1029/2023WR035126

Johnson, K., Jankowski, K. J., Carey, J., Lyon, N. J., McDowell, W. H., Shogren, A., et al. (2023). Monthly dissolved silicon concentrations from 198 rivers in the northern Hemisphere: U.S. Geological Survey data release [Dataset]. https://doi.org/10.5066/P9CXDEE8

Johnson, K., Jankowski, K. J., Carey, J., Lyon, N. J., McDowell, W. H., Shogren, A., et al. (2024). Establishing fluvial silicon regimes and their stability across the Northern Hemisphere. *Limnology and Oceanography Letters*, *9*(3), 237–246. https://doi.org/10.1002/lol2.10372

Johnson, K., Jankowski, K. J., Carey, J. C., Sethna, L. R., Bush, S. A., McKnight, D., et al. (2024). Average and annual driver data for rivers across the northern Hemisphere: U.S. Geological Survey data release [Dataset]. https://doi.org/10.5066/P14NBAYZ

Julian, J. P., Doyle, M. W., & Stanley, E. H. (2008). Empirical modeling of light availability in rivers. *Journal of Geophysical Research*, *113*(G3). https://doi.org/10.1029/2007JG000601

Júnior, J. L. S., Tomasella, J., & Rodriguez, D. A. (2015). Impacts of future climatic and land cover changes on the hydrological regime of the Madeira River basin. *Climatic Change*, *129*(1), 117–129. https://doi.org/10.1007/s10584-015-1338-x

Kienel, U., Kirillin, G., Brademann, B., Plessen, B., Lampe, R., & Brauer, A. (2017). Effects of spring warming and mixing duration on diatom deposition in deep Tiefer See, NE Germany. *Journal of Paleolimnology*, *57*(1), 37–49. https://doi.org/10.1007/s10933-016-9925-z

Konapala, G., Mishra, A. K., Wada, Y., & Mann, M. E. (2020). Climate change will affect global water availability through compounding changes in seasonal precipitation and evaporation. *Nature Communications*, *11*(1), 3044. https://doi.org/10.1038/s41467-020-16757-w

Koutsoyiannis, D. (2020). Revisiting the global hydrological cycle: Is it intensifying? *Hydrology and Earth System Sciences*, *24*(8), 3899–3932. https://doi.org/10.5194/hess-24-3899-2020

Larras, F., Coulaud, R., Gautreau, E., Billoir, E., Rosebery, J., & Usseglio-Polatera, P. (2017). Assessing anthropogenic pressures on streams: A random forest approach based on benthic diatom communities. *Science of the Total Environment*, *586*, 1101–1112. https://doi.org/10.1016/j.scitotenv.2017.02.096

LeBauer, D. S., & Treseder, K. K. (2008). Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, *89*(2), 371–379. https://doi.org/10.1890/06-2057.1

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Li, J., & Thompson, D. W. J. (2021). Widespread changes in surface temperature persistence under climate change. *Nature*, *599*(7885), 425–430. https://doi.org/10.1038/s41586-021-03943-z

Li, L., Stewart, B., Zhi, W., Sadayappan, K., Ramesh, S., Kerins, D., et al. (2022). Climate controls on river chemistry. *Earth's Future*, *10*(6), e2021EF002603. https://doi.org/10.1029/2021EF002603

Liess, A., Lange, K., Schulz, F., Piggott, J. J., Matthaei, C. D., & Townsend, C. R. (2009). Light, nutrients and grazing interact to determine diatom species richness via changes to productivity, nutrient state and grazer activity. *Journal of Ecology*, *97*(2), 326–336. https://doi.org/10.1111/j.1365-2745.2008.01463.x

Loewen, C. J. G., Vinebrooke, R. D., & Zurawell, R. W. (2021). Quantifying seasonal succession of phytoplankton trait-environment associations in human-altered landscapes. *Limnology & Oceanography*, *66*(4), 1409–1423. https://doi.org/10.1002/lno.11694

Lyon, N. J., Chen, A., Bush, S. A., Brun, J., Hwangbo, N., & Peach, K. (2024). lter/lterwg-silica-spatial: First Release (v1.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.12707913

Magilligan, F. J., & Nislow, K. H. (2005). Changes in hydrologic regime by dams. *Geomorphology*, *71*(1), 61–78. https://doi.org/10.1016/j.geomorph.2004.08.017

Mastrotheodoros, T., Pappas, C., Molnar, P., Burlando, P., Manoli, G., Parajka, J., et al. (2020). More green and less blue water in the Alps during warmer summers. *Nature Climate Change*, *10*(2), 155–161. https://doi.org/10.1038/s41558-019-0676-5

McNair, H. M., Brzezinski, M. A., & Krause, J. W. (2018). Diatom populations in an upwelling environment decrease silica content to avoid growth limitation. *Environmental Microbiology*, *20*(11), 4184–4193. https://doi.org/10.1111/1462-2920.14431

Menzel, A., Sparks, T. H., Estrella, N., & Roy, D. B. (2006). Altered geographic and temporal variability in phenology in response to climate change. *Global Ecology and Biogeography*, *15*(5), 498–504. https://doi.org/10.1111/j.1466-822X.2006.00247.x

Meybeck, M. (1987). Global chemical weathering of surficial rocks estimated from river dissolved loads. *American Journal of Science*, *287*(5), 401–428. https://doi.org/10.2475/ajs.287.5.401

Michelutti, N., Douglas, M. S. V., & Smol, J. P. (2003). Diatom response to recent climatic change in a high arctic lake (Char Lake, Cornwallis Island, Nunavut). *Global and Planetary Change*, *38*(3), 257–271. https://doi.org/10.1016/S0921-8181(02)00260-6

Mills, T. J., Anderson, S. P., Bern, C., Aguirre, A., & Derry, L. A. (2017). Colloid mobilization and seasonal variability in a semiarid headwater stream. *Journal of Environmental Quality*, *46*(1), 88–95. https://doi.org/10.2134/jeq2016.07.0268

Moatar, F., Abbott, B. W., Minaudo, C., Curie, F., & Pinay, G. (2017). Elemental properties, hydrology, and biology interact to shape concentration-discharge curves for carbon, nutrients, sediment, and major ions. *Water Resources Research*, *53*(2), 1270–1287. https://doi.org/10.1002/2016WR019635

Monahan, W. B., Rosemartin, A., Gerst, K. L., Fisichelli, N. A., Ault, T., Schwartz, M. D., et al. (2016). Climate change is advancing spring onset across the U.S. national park system. *Ecosphere*, *7*(10), e01465. https://doi.org/10.1002/ecs2.1465

Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower snowmelt in a warmer world. *Nature Climate Change*, *7*(3), 214–219. https://doi.org/10.1038/nclimate3225

Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017). Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. *Water Resources Management*, *31*(9), 2761–2775. https://doi.org/10.1007/s11269-017-1660-3

Officer, C., & Ryther, J. (1980). The possible importance of silicon in marine eutrophication. *Marine Ecology Progress Series*, *3*, 83–91. https://doi.org/10.3354/meps003083

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 tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science & Technology*, *50*(20), 10805–10813. https://doi.org/10.1021/acs.est.6b02575

Peñas, F. J., & Barquín, J. (2019). Assessment of large-scale patterns of hydrological alteration caused by dams. *Journal of Hydrology*, *572*, 706–718. https://doi.org/10.1016/j.jhydrol.2019.03.056

Pham, L. T., Luo, L., & Finley, A. (2021). Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. *Hydrology and Earth System Sciences*, *25*(6), 2997–3015. https://doi.org/10.5194/hess-25-2997-2021

Phillips, A. K. (2020). *Modelling riverine dissolved silica on different spatial and temporal scales using statistical and machine learning methods* (Doctoral Dissertation). University of Toronto.

Pinay, G., Gumiero, B., Tabacchi, E., Gimenez, O., Tabacchi-Planty, A. M., Hefting, M. M., et al. (2007). Patterns of denitrification rates in European alluvial soils under various hydrological regimes. *Freshwater Biology*, *52*(2), 252–266. https://doi.org/10.1111/j.1365-2427.2006.01680.x

Post, D. A., & Jones, J. A. (2001). Hydrologic regimes of forested, mountainous, headwater basins in New Hampshire, North Carolina, Oregon, and Puerto Rico. *Advances in Water Resources*, *24*(9), 1195–1210. https://doi.org/10.1016/S0309-1708(01)00036-7

Pulliainen, J. (2024). Snow loss pinned to human-induced emissions. *Nature*, *625*(7994), 246–247. https://doi.org/10.1038/d41586-023-03993-5

Regier, P., Duggan, M., Myers-Pigg, A., & Ward, N. (2023). Effects of random forest modeling decisions on biogeochemical time series predictions. *Limnology and Oceanography: Methods*, *21*(1), 40–52. https://doi.org/10.1002/lom3.10523

Satoh, Y., Yoshimura, K., Pokhrel, Y., Kim, H., Shiogama, H., Yokohata, T., et al. (2022). The timing of unprecedented hydrological drought under climate change. *Nature Communications*, *13*(1), 3287. https://doi.org/10.1038/s41467-022-30729-2

Scanlon, T. M., Raffensperger, J. P., & Hornberger, G. M. (2001). Modeling transport of dissolved silica in a forested headwater catchment: Implications for defining the hydrochemical response of observed flow pathways. *Water Resources Research*, *37*(4), 1071–1082. https://doi.org/10.1029/2000WR900278

Schelske, C. L., Stoermer, E. F., Fahnenstiel, G. L., & Haibach, M. (1986). Phosphorus enrichment, silica utilization, and biogeochemical silica depletion in the Great lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *43*(2), 407–415. https://doi.org/10.1139/f86-051

Schlüter, M. H., Kraberg, A., & Wiltshire, K. H. (2012). Long-term changes in the seasonality of selected diatoms related to grazers and environmental conditions. *Journal of Sea Research*, *67*(1), 91–97. https://doi.org/10.1016/j.seares.2011.11.001

Schneider, S. C. (2015). Greener rivers in a changing climate? Effects of climate and hydrological regime on benthic algal assemblages in pristine streams. *Limnologica*, *55*, 21–32. https://doi.org/10.1016/j.limno.2015.10.004

Schnorbus, M., Werner, A., & Bennett, K. (2014). Impacts of climate change in three hydrologic regimes in British Columbia, Canada. *Hydrological Processes*, *28*(3), 1170–1189. https://doi.org/10.1002/hyp.9661

Segura, C., Noone, D., Warren, D., Jones, J. A., Tenny, J., & Ganio, L. M. (2019). Climate, landforms, and geology affect baseflow sources in a mountain catchment. *Water Resources Research*, *55*(7), 5238–5254. https://doi.org/10.1029/2018WR023551

Seybold, E. C., Burgin, A. J., Brown, C., Speir, S., Wheeler, C., & Zipper, S. C. (2022). Linking oxygen regimes to flow regimes in non-perennial streams. *2022*, H55L–H0729.

Seybold, E. C., Fork, M. L., Braswell, A. E., Blaszczak, J. R., Fuller, M. R., Kaiser, K. E., et al. (2022). A classification framework to assess ecological, biogeochemical, and hydrologic synchrony and asynchrony. *Ecosystems*, *25*(5), 989–1005. https://doi.org/10.1007/s10021-021-00700-1

Sommer, U. (1988). Phytoplankton succession in microcosm experiments under simultaneous grazing pressure and resource limitation. *Limnology & Oceanography*, *33*(5), 1037–1054. https://doi.org/10.4319/lo.1988.33.5.1037

Sommer, U., & Lengfellner, K. (2008). Climate change and the timing, magnitude, and composition of the phytoplankton spring bloom. *Global Change Biology*, *14*(6), 1199–1208. https://doi.org/10.1111/j.1365-2486.2008.01571.x

Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, *560*(7720), 639–643. https://doi.org/10.1038/s41586-018-0411-9

Street-Perrott, F. A., & Barker, P. A. (2008). Biogenic silica: A neglected component of the coupled global continental biogeochemical cycles of carbon and silicon. *Earth Surface Processes and Landforms*, *33*(9), 1436–1457. https://doi.org/10.1002/esp.1712

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Struyf, E., & Conley, D. J. (2012). Emerging understanding of the ecosystem silica filter. *Biogeochemistry*, *107*(1), 9–18. https://doi.org/10.1007/s10533-011-9590-2

Struyf, E., Smis, A., Van Damme, S., Garnier, J., Govers, G., Van Wesemael, B., et al. (2010). Historical land use change has lowered terrestrial silica mobilization. *Nature Communications*, *1*(1), 129. https://doi.org/10.1038/ncomms1128

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1), 13768. https://doi.org/10.1038/s41598-020-70816-2

Teubner, K., & Dokulil, M. T. (2002). Ecological stoichiometry of TN: TP: SRSi in freshwaters: Nutrient ratios and seasonal shifts in phytoplankton assemblages. *Archiv für Hydrobiologie*, *154*(4), 625–646. https://doi.org/10.1127/archiv-hydrobiol/154/2002/625

Thanapakpawin, P., Richey, J., Thomas, D., Rodda, S., Campbell, B., & Logsdon, M. (2007). Effects of landuse change on the hydrologic regime of the Mae Chaem river basin, NW Thailand. *Journal of Hydrology*, *334*(1), 215–230. https://doi.org/10.1016/j.jhydrol.2006.10.012

Tréguer, P., Nelson, D. M., Van Bennekom, A. J., DeMaster, D. J., Leynaert, A., & Quéguiner, B. (1995). The silica balance in the World Ocean: A reestimate. *Science*, *268*(5209), 375–379. https://doi.org/10.1126/science.268.5209.375

Tréguer, P. J., Sutton, J. N., Brzezinski, M., Charette, M. A., Devries, T., Dutkiewicz, S., et al. (2021). Reviews and syntheses: The biogeochemical cycle of silicon in the modern ocean. *Biogeosciences*, *18*(4), 1269–1289. https://doi.org/10.5194/bg-18-1269-2021

Trentman, M. T., Tank, J. L., Shepherd, H. A. M., Marrs, A. J., Welsh, J. R., & Goodson, H. V. (2021). Characterizing bioavailable phosphorus concentrations in an agricultural stream during hydrologic and streambed disturbances. *Biogeochemistry*, *154*(3), 509–524. https://doi.org/10.1007/s10533-021-00803-w

Van Meter, K. J., Chowdhury, S., Byrnes, D. K., & Basu, N. B. (2020). Biogeochemical asynchrony: Ecosystem drivers of seasonal concentration regimes across the Great Lakes Basin. *Limnology & Oceanography*, *65*(4), 848–862. https://doi.org/10.1002/lno.11353

Visser, P. M., Verspagen, J. M. H., Sandrini, G., Stal, L. J., Matthijs, H. C. P., Davis, T. W., et al. (2016). How rising CO_2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae*, *54*, 145–159. https://doi.org/10.1016/j.hal.2015.12.006

Vorkauf, M., Marty, C., Kahmen, A., & Hiltbrunner, E. (2021). Past and future snowmelt trends in the Swiss Alps: The role of temperature and snowpack. *Climatic Change*, *165*(3), 44. https://doi.org/10.1007/s10584-021-03027-x

Wall, G. R., Phillips, P. J., & Riva-Murray, K. (1998). Seasonal and spatial patterns of nitrate and silica concentrations in Canajoharie Creek, New York. *Journal of Environmental Quality*, *27*(2), 381–389. https://doi.org/10.2134/jeq1998.00472425002700020019x

Wang, B., Liu, C.-Q., Maberly, S. C., Wang, F., & Hartmann, J. (2016). Coupling of carbon and silicon geochemical cycles in rivers and lakes. *Scientific Reports*, *6*(1), 35832. https://doi.org/10.1038/srep35832

Wang, Y., Wang, D., Lewis, Q. W., Wu, J., & Huang, F. (2017). A framework to assess the cumulative impacts of dams on hydrological regime: A case study of the Yangtze river. *Hydrological Processes*, *31*(17), 3045–3055. https://doi.org/10.1002/hyp.11239

Wasmund, N., Nausch, G., Gerth, M., Busch, S., Burmeister, C., Hansen, R., & Sadkowiak, B. (2019). Extension of the growing season of phytoplankton in the western Baltic Sea in response to climate change. *Marine Ecology Progress Series*, *622*, 1–16. https://doi.org/10.3354/meps12994

West, A. J., Galy, A., & Bickle, M. (2005). Tectonic and climatic controls on silicate weathering. *Earth and Planetary Science Letters*, *235*(1), 211–228. https://doi.org/10.1016/j.epsl.2005.03.020

White, A. F., & Blum, A. E. (1995). Effects of climate on chemical_weathering in watersheds. *Geochimica et Cosmochimica Acta*, *59*(9), 1729–1747. https://doi.org/10.1016/0016-7037(95)00078-E

Williamson, T. J., Vanni, M. J., & Renwick, W. H. (2021). Spatial and temporal variability of nutrient dynamics and ecosystem metabolism in a hyper-eutrophic reservoir differ between a wet and dry year. *Ecosystems*, *24*(1), 68–88. https://doi.org/10.1007/s10021-020-00505-8

Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, *12*(1), 2501. https://doi.org/10.1038/s41467-021-22702-2

Wu, G., Chen, J., Shi, X., Kim, J.-S., Xia, J., & Zhang, L. (2022). Impacts of global climate warming on meteorological and hydrological droughts and their propagations. *Earth's Future*, *10*(3), e2021EF002542. https://doi.org/10.1029/2021EF002542

Yang, D., Kane, D. L., Hinzman, L. D., Zhang, X., Zhang, T., & Ye, H. (2002). Siberian Lena River hydrologic regime and recent change. *Journal of Geophysical Research*, *107*(D23). ACL 14-1-ACL 14-10. https://doi.org/10.1029/2002JD002542

Yang, M., & Olivera, F. (2023). Classification of watersheds in the conterminous United States using shape-based time-series clustering and Random Forests. *Journal of Hydrology*, *620*, 129409. https://doi.org/10.1016/j.jhydrol.2023.129409

Yu, Z., Gu, H., Wang, J., Xia, J., & Lu, B. (2018). Effect of projected climate change on the hydrological regime of the Yangtze River Basin, China. *Stochastic Environmental Research and Risk Assessment*, *32*(1), 1–16. https://doi.org/10.1007/s00477-017-1391-2

Zheng, L., Cheng, X., Chen, Z., Wang, S., Liang, Q., & Wang, K. (2022). Global snowmelt onset reflects climate variability: Insights from spaceborne radiometer observations. *Journal of Climate*, *35*(10), 2945–2959. https://doi.org/10.1175/JCLI-D-21-0265.1